A Work Project, presented as part of the requirements for the Award of a Master's degree in Finance from the Nova School of Business and Economics.

POTENTIAL FOR BANKS' PRIVATE MORTGAGE FINANCING BUSINESS THROUGH THE APPLICATION OF PREDICTIVE ANALYTICS

DOMINIK JAKOB LAUFS

Work project carried out under the supervision of

Professor João Pereira

31-12-2020

Abstract

Private mortgage financing accounts for a growing share of banks' lending business and, thus, their profits. Yet increasing competition and tightening market conditions are pressuring profitability. This work project analyzes predictive analytics as a means to alleviate these challenges, focusing on the German market and relying on guided expert interviews for primary research. In particular, it outlines 17 existing and potential fields of application, and examines corresponding types of predictive analytics models and data sources. Furthermore, 12 major challenges associated with the implementation of predictive analytics are identified.

Keywords: Predictive Analytics, Data Analytics, Private Mortgage Financing, Lending Business

Table of Contents

1 Introduction

Private mortgage loans are an essential component of the credit portfolios of German banks. The business has grown in recent years, reaching €263 billion in Germany in 2019 and constituting 15% of banks' total interest income (PWC 2020). Due to the Covid-19 crisis, opinions are mixed on the further development of the sector, with views ranging from a constant or increased demand for real estate (Engel&Völkers 2020; Rottwilm 2020), to surveys noting a deferment of demand and supply (Streit 2020; Zapf 2020). Either way, by binding customers to a bank for the long term and providing opportunities for cross-selling, private mortgage loans are one of the most important products banks possess (Douqué, Scharf, and Albrecht 2020; Hacikura and Seto 2014).

Market conditions in the lending business are tightening for a variety of reasons. While a low-interest environment is squeezing banks' profit margins (Klein 2020), an increase in regulatory requirements is compounding their expenses (Gaumert 2019). Furthermore, digitalization poses challenges to banks which still carry the burden of old processes (White 2020). Meanwhile, digital customers are demanding ever-more convenient, personalized services at low prices (Forest and Rose 2015; Bellens and Meekings 2020). Banks also face growing competition from financial technology startups (FinTechs) and existing technology companies that are entering the financial market with innovative services (Browne 2020; McWaters and Galaski 2017).

Given the importance of private mortgage financing, it is crucial for banks to address these challenges and differentiate themselves from competitors in the field. While there are different strategic and operational measures for this purpose, the use of innovative technologies for the standardization, automation, and finally digitalization of processes is one of the most important approaches available.

Predictive analytics (PA) is a tool that has the potential to substantially support these goals. The method consists of classification or regression models which both apply data mining (DM) techniques in order to identify data patterns and then use these to predict the likely occurrence of unknown and future events, such as the affinity for a product or the probability of payment defaults (Rehfisch 2019; Dölle 2018).

While some research exists on the application of PA in the general loan business, limited information is available on its use in private mortgage financing. This work project aims to fill that gap by asking the following research questions: What are the existing and potential fields of application of PA in the private mortgage financing process? What type of PA is applied and what data sources are used? Lastly, what risks and challenges are attached to the implementation of PA?

To answer these questions, guided expert interviews with experts from banks, building societies, FinTechs, and service providers were conducted. This work project seeks to derive fundamental recommendations for banks to tackle the challenging market environment by adopting PA tools in their private mortgage financing process.

2 Literature Review

2.1 Private Mortgage Financing

2.1.1 Market Developments

The lending business is one of the core revenue drivers of German banks (PWC 2017). Loans to private customers can be classified into three categories: overdraft facilities; installment or consumer loans; and mortgage loans[1] (BaFin 2020). In 2019, the private mortgage portfolio of German banks amounted to €1.3 trillion, an increase of 5.7% compared with 2018 and a share of 42.4% of their total credit volume[2] (Deutsche Bundesbank 2020; PWC 2020).

---

[1] Mortgage loans are used for purposes such as financing the construction, purchase, or modernization of a house, or as follow-up financing for a previous mortgage loan (Wuestenrot n.d.; drklein n.d.). Appendix 4 provides an overview of different types of German banks offering private mortgage loans.

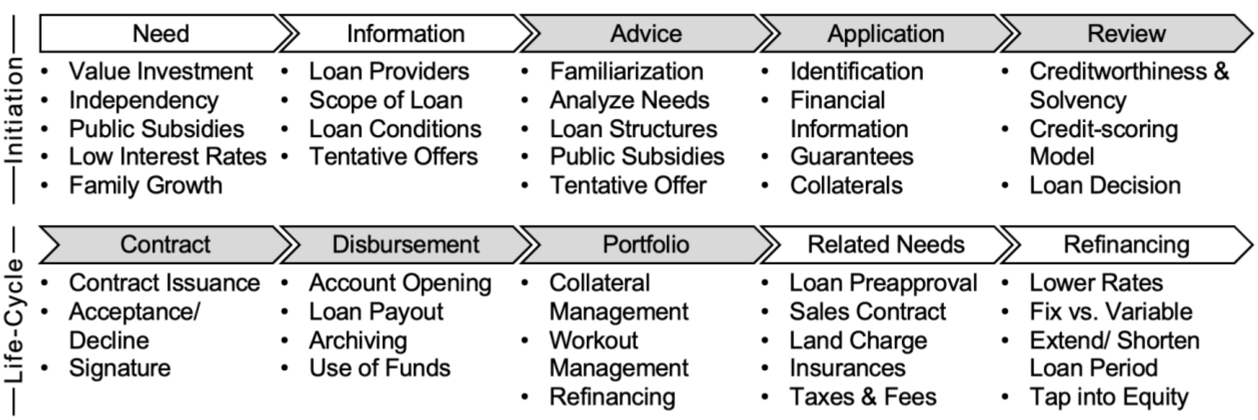
[2] Total credit volume does not include credit to banks, central banks, and public institutions.

While private mortgage loans only accounted for 5.5% of total interest income in 2012, this proportion had grown to 15% in 2019. This development is mainly due to the decline of interest rates on other loan types, as the margins in private mortgage loans fluctuated between 0.8% and 1.2% over the same period (PWC 2020). Additionally, due to their central importance to borrowers' lives and their long maturities, private mortgage loans function as anchor products and enable banks to realize cross-selling potential (Douqué, Scharf, and Albrecht 2020; Fleischer 2015).

2.1.2 Private Mortgage Financing Process

Within the private mortgage financing process, the perspectives of both bank and customer play a crucial role in forming a comprehensive "customer journey" (see Figure 1). The classical bank process, as described by PwC (2017) and Möst (2019) (see grey boxes in Figure 1), is extended with customer perspectives, as found in the relevant literature, to construct this more holistic approach. It is important to mention that the various steps of the journey are not clearly delimited from each other and have many interrelations. This is explained further in Appendix 5.

Figure 1: Customer Journey of the Private Mortgage Financing Process



Note: Boxes in grey represent process steps of the classical bank process

Source: Own illustration based on illustration adapted from Möst (2019) and PwC (2017)

2.2 Predictive Analytics

2.2.1 Relevance of Predictive Analytics

The amount of data available for computational analysis increases exponentially. It is expected that over the next three years, more data will be gathered than over the past three decades (IDC 2020; Mohr and Hürtgen 2018). With tightening market conditions, cost pressures, and growing competition, corporations see the need to leverage their rich pools of information and use PA rather than descriptive analytics to gain a competitive advantage (Selamat 2018). This development is supported by ever cheaper data storage capacities and increasing computational power (Henke et al. 2016; Mohr and Hürtgen 2018). In addition, there are significant advances in artificial intelligence (AI) that, applied to PA, enable more sophisticated analysis and make it accessible to a wider range of users (Henke et al. 2016; Singh 2019; SAS n.d.).

2.2.2 Definition of Predictive Analytics

PA can be categorized as a subcategory of data analytics, which is a general form of analyzing data to provide a basis for decision making (Gartner 2013; Ahonen et al. 2018; Tschakert et al. 2016). More detailed information on the different subcategories is included in Appendix 6.

Data mining (DM) is a major component of PA. By applying specific algorithms, it examines data to observe trends, patterns, and relationships which can be used to build a PA model (Dölle 2018; Finlay 2014; Lange 2006). DM identifies all types of relationships and correlations in the given data, regardless of their cause (Dölle 2018; Nyce 2007). PA then sets up and applies a model to analyze current and historical data and the patterns identified by the DM algorithms (Dölle 2018). These patterns are then applied to enable a nearest-possible prediction of unknown or future events (Dölle 2018).

Additionally, PA can incorporate machine learning (ML) techniques, whose recent developments allow for extensive analysis of more complex, unstructured data sets (TDWI 2014; Kelleher, Namee, and D'Arcy 2020). ML enables the software to learn without being explicitly programmed. Thus, PA models with ML can adapt to changes in data, while PA models without ML rely on input from humans to do so (Kumar 2018; Samuel 1959).

2.2.3 Predictive Analytics Models

The goal of PA models is to explain a dependent variable as a function of an explaining variable. This can be accomplished by a range of classification and regression models[3] (Dölle 2018; SAS n.d.). Some of the most commonly used PA models include decision trees, artificial neural networks (ANN), and linear or logistic regressions (SAS n.d.).

Decision trees aim at identifying variables which split a data set into the most different subsets in order to best possibly predict a target variable (Aunalytics 2015; Rokach and Maimon 2015). There are either regression trees, which forecast a continuous target value based on a set of input factors, or classification trees, which divide a data set into predefined classes of a categorical target value (Rokach and Maimon 2015; Kotu and Deshpande 2015). Such decision trees are particularly popular due to their simplicity and transparency (Rokach and Maimon 2015).

Regression models assess the cause-effect relationship of independent variables on a target variable (Ray 2015). While a linear regression can be used to describe the relationship between independent variables and a continuous target variable, a logistic regression can be used to show the relationship between independent variables and a categorical target variable (Larose 2015).

ANN models are "an attempt […] to imitate the type of nonlinear learning that occurs in the networks of neurons found in nature" (Larose, 2015, p.339). They reveal complex correlations and can be used, for example, to cluster unlabeled data based on similarities, or to classify data, if labeled data sets are provided for training (Team EA 2020; Bari, Chaouchi, and Jung 2016).

There is no model which is holistically better than others, as this depends on the objective of the model, the available data, and the context in which it will be used (Kuhn and Johnson 2013). An overview of different classification and regression models is displayed in Appendix 8.

---

[3] Appendix 7 provides a short overview about the setup and use of such PA models

2.2.4 Limitations of the Application of Predictive Analytics

Despite the benefits of using PA to improve the lending business, its use does not come without side effects. PA models run the risk of containing biases which could potentially discriminate against certain groups (BaFin 2018; Faggella 2020; Ryll et al. 2020). As such models are complex and do not necessarily allow their users to query the decision-making process, they are often seen as a black box (BaFin 2018; Ryll et al. 2020). This aspect is also criticized by regulators, who expect the decisions made by such models to be understandable, or in other words transparent and explainable (BaFin 2018; Rehfisch 2019).

Furthermore, integrating PA models into existing systems constitutes another challenge. With multiple operating platforms and developer languages in place, interoperable systems and language translations are required (Attaran and Attaran 2018; Eckerson 2007). For PA models to work, all relevant data needs to be accessible, up to date, in the same format, and free of errors (Hazen et al. 2014). If a model becomes too complex and memorizes training data, it is regarded as an "overfitting problem" which can yield unreliable interpretations for new data inputs. (Abbott 2014). Additionally, organizational limitations may include a lack of awareness within a firm, inadequate budgets, and insufficient in-house expertise to develop PA solutions. These can result in an inability to shift solutions from a development phase to an operational solution (Abbott 2014).

2.3 Fields of Application for Predictive Analytics

2.3.1 Predictive Analytics Across Industries

PA can be applied across numerous industries in a variety of ways. These include financial services, manufacturing, healthcare, utilities and logistics, among many others (Microstrategy 2018; SAS n.d.). Especially large technology companies such as Alphabet, Amazon, Facebook, and Apple are known for their deep expertise in and long-standing experience of PA (Zollinger 2020). They not only use the technology to target their customers more proficiently with ads, content, or purchase suggestions, but also offer their PA tools as a service for customers to apply to their own data (Britt 2015; Perrier 2017). Other technology companies such as IBM, Prediction IO, or BigML develop PA tools which can be applied across different industries (Perrier 2017).

2.3.2 Predictive Analytics in the Lending Business

Zollinger (2020) points out that financial institutions should learn from these technology companies, leverage their trust advantage over tech-companies among customers, and apply their valuable customer data in a targeted way. The fields of application for PA in the financial sector are diverse and range from the risk assessment of loans to product pricing, churn prevention, or product recommendations (Garg et al. 2017; Zollinger 2020). Often, however, the literature talks solely about AI without mentioning its connections to PA, or does not discuss private mortgage financing specifically but refers to a more general process. Some of the main areas of PA application related to the private mortgage financing process can be found in the process steps *need*, *advice*, *review* and *portfolio* and are briefly described below.

AI tools can be utilized in order to predict the communication channel most likely to yield a customer response and interaction (Tahedl 2019). Such tools are also able to predict the need for a loan, and to make recommendations as to what loan specifications are most appropriate for a customer (Rehfisch 2019; Garg et al. 2017).

A significant amount of research shows that different types of AI models perform more reliable credit risk assessments than traditional approaches (Byanjankar, Heikkilä, and Mezei 2015; van Thiel and van Raaij 2019). Additionally, they can enhance the speed and accuracy of credit decision making, work better the more data is available about an individual customer, and are particularly useful when no credit history is available (Dalela 2019; Sreedhar 2019; Faggella 2020). Data that can be incorporated in such analyses include social media activities, buying patterns from online shopping, demographic data, browsing preferences, purchasing habits, psychometric testing, or geo-location data (Dalela 2019; Ogbonna 2017; Ryll et al. 2020). However, depending on the availability and detail of some data, their added value can be questionable (Ryll et al. 2020).

With rich data available, AI tools can detect payment difficulties early on, before an actual default occurs. Consequently, banks can take preventive measures and a loan must not be classified as defaulted, ensuring the satisfaction of both banks and customers (Ogbonna 2017; Tahedl 2019).

2.3.3 FinTechs in the Business of Predictive Analytics

In recent years, FinTechs have upturned the financial industry, questioning the conventional business models of traditional financial institutions (Rode et al. 2018). Looking across borders, PA finds wide application among FinTechs (Zafra 2019). Alongside other areas, it is applied in the lending business and, to some extent also, in private mortgage financing. For example, the internationally operating FinTech *AdviceRobo* supports credit providers with PA-based risk management solutions which can be applied in small business lending and retail lending, including loans for mortgages (AdviceRobo n.d.). Looking at the German market, there is no FinTech offering PA solutions for the private mortgage financing process. The main challenge for traditional financial institutions in this market are comparison websites and online brokers which increase transparency among providers (Douqué, Scharf, and Albrecht 2020; Sparkassenzeitung 2018). Due to the ability of FinTechs to integrate AI solutions in a more cost-efficient manner, and in order to leverage their technology, banks are increasingly interested in collaborations (Ryll et al. 2020).

3 Methodology

3.1 Methodological Approach

As information is scarce on the specific topic of PA and its use in private mortgage financing, primary research is used as a methodology. This work project relies on guided expert interviews as a form of qualitative research. Such interviews are conducted with experts from the field and rely on open questions (Mayer 2013). This format gives the interviewer the freedom to orientate on a created questionnaire, deviate from it to dive deeper into certain topics, or follow up on newly arisen topics which are of relevance for the work project (Gibson and Brown 2009; Mayer 2013).

3.2 Structure of Guided Expert Interviews

All relevant interview questions are summarized in a questionnaire, which is divided into an informational phase, a main part, and a concluding phase, as proposed by Misoch (2015). The informational phase aims at gathering information about the interviewees and informing them about the rationale of the interview and the data usage. It is also used to discuss the private mortgage financing environment and financing process, as well as the definition of PA, in order to find common ground and demonstrate a high quality and comparability of interview results. Afterwards, the main part aims at answering the following research questions:

Research question 1: What are the existing and potential fields of application of PA in the private mortgage financing process?

Research question 2: What type of PA is applied and what data sources are used?

Research question 3: What risks and challenges are attached to the implementation of PA?

In order to explore the research questions in detail, a range of sub-questions accompany them. Here, particular focus was placed on the role of FinTechs, to consider their emerging role in the financial industry. The full questionnaire is shown in Appendix 9 and Appendix 10.

3.3 Selection and Evaluation of Interviewees

Since this empirical study focuses on the application of PA in private mortgage financing, the author firstly defined a target group likely to provide insights into this specific topic. The target group included banks and building societies offering private mortgage loans, service providers providing technological infrastructures and software solutions, consultancies advising in this area, and FinTechs. In the following list, references to “banks” include building societies and service providers as well. A total of 15 experts were interviewed in 12 interviews, which lasted an average of 61 minutes. Table 1 displays an overview over the interviewee sample.

Table 1: List of Interviewees

Expert ID	Firm ID	Type of Firm	Turnover in 2019 (in €M)	No. of Employees in 2019	Position of Interviewee
E1	1	Service Provider	100 - 500	500 - 1,000	Senior Manager
E2	2	FinTech	N.A.	< 500	Business Development Manager
E3	3	Service Provider	500 - 1,000	1,000 - 5,000	Chief Technology Officer
E4	4	Service Provider	100 - 500	500 - 1,000	Senior Manager
E5	5	FinTech	N.A.	< 500	Business Development Manager
E6	6	Building Society	1,000 - 5,000	5,000 - 10,000	Data Analyst
E7	7	Bank	100 - 500	500 - 1,000	Head of Innovation & Digital Division
E8	8	Bank	100 - 500	1,000 - 5,000	Member of the Board
E9	9	Building Society	1,000 - 5,000	500 - 1,000	Head of Mortgage Financing
E10	9	Building Society	1,000 - 5,000	500 - 1,000	Head of Process Mgmt. in Mortgage Financing
E11	10	Bank	100 - 500	1,000 - 5,000	Product Manager Mortgage Financings
E12	11	Bank	> 10,000	> 10,000	Head of Big Data & Advanced Analytics for Sales
E13	11	Bank	> 10,000	> 10,000	Product Owner for Loan Products
E14	12	Service Provider	< 100	< 500	Data Analyst
E15	12	Service Provider	< 100	< 500	Communicator between Developers and Banks

Source: Interviews, firms' websites and annual reports
Note: Service providers include organizations providing technology and/ or advisory services.

After conducting the interviews, the responses were paraphrased and all broader answers which did not come as a direct reply to a question from the questionnaire were assigned to respective questions. Content which was not relevant for the purpose of this work project was disregarded. All paraphrased interviews can be found in Appendix 11 through Appendix 22.

4 Results

All findings presented in this chapter are derived from the 12 interviews. Interviewees are identified by their corresponding Expert ID abbreviations (see Table 1).

4.1 Reaching a Common Ground

Ensuring the quality and comparability of the interview results, all interviewees agreed on the information presented on the private mortgage financing environment (see Chapter 2.1.1), the private mortgage financing process (see Chapter 2.1.2) and the definition of PA (see Chapter 2.2.2). Moreover, the statements regarding the relevance of the application of PA are also in accordance with the literature (see Chapter 2.2.1). The interviewees emphasized that the industry faces environmental challenges which include increased competition, higher cost pressures, and the further development of PA beyond the banking sector. Increased competition includes technological developments by banks as well as FinTechs entering the market. The use of PA is expected to overcome these challenges, building on the availability of larger amounts of data and advancements in technology. This includes better system infrastructure, higher computational power, cheaper and easier-to-use technologies, and advancements in underlying algorithms.

4.2 Predictive Analytics in the Private Mortgage Financing Process

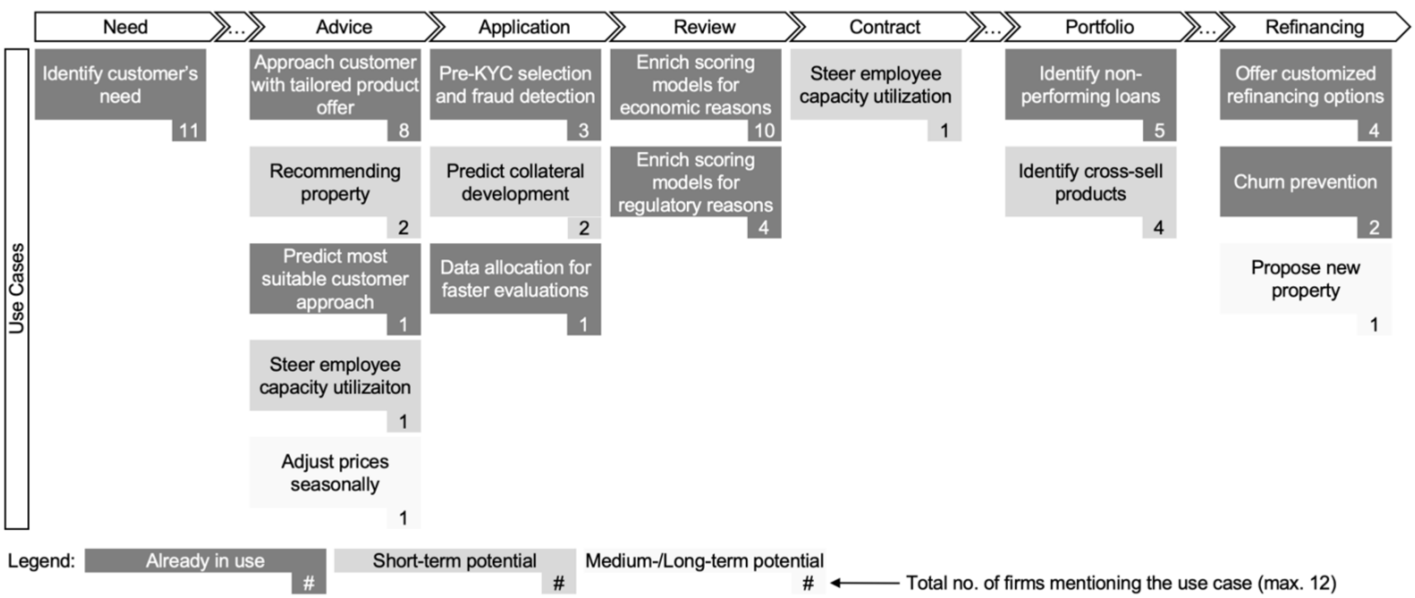
This section addresses the first research question and provides information on current and potential future use cases of PA in the private mortgage financing process. A total of 17 use cases were identified in seven of ten process steps[4] (see Figure 2). Here, the author emphasizes the most relevant of these examples, which are defined as those mentioned by at least a quarter of the interviewed firms. Appendix 23 reports on the remaining use cases, which had fewer mentions.

There is only one use case that exemplifies the process step *need*. However, 11 of the 12 interviewed firms mention it and of these, eight already see it in use. It involves PA drawing on customer data in order to "identify a customer's need" to finance a mortgage. It is highly relevant for banks, as they have realized that a targeted, proactive customer approach is essential for an early positioning in the customer journey. For this, variables of customer characteristics are analyzed to determine the probability of occurrence of certain events in a customer's life (e.g.,

[4] No use cases have been identified for the process steps *information, disbursement*, and *related needs*.

marriage, family growth, or start of work) which are associated with a higher probability of closing a mortgage loan. In addition to this, E4 and E13 mention that PA models also analyze interdependencies of available customer data and historical product completions to draw conclusions about other customers' needs. In this context, E1 emphasizes that such tools should aim to select customers who suit the business model of a bank and are most likely to complete a mortgage loan with it, and not just in general.

Figure 2: Predictive Analytics Use Cases in the Private Mortgage Financing Process



Note: A use case is considered to be "already in use" if at least one of the firms currently applies it. A use case is considered to have "short-term potential" if none of the firms currently applies it and at least one sees a short-term potential. A use case is considered to have "medium-/long-term potential" if it is seen as having medium- to long-term potential by all firms mentioning the use case.

Five use cases were identified in the process step *advice*, which is the most for any process step. One use case is mentioned eight times and is already seen in application in some form by six interviewees. It concerns "approaching customers with tailored product offers" and goes hand-in-hand with other use cases of PA in this process step. Due to increasing customer expectations for individually tailored solutions, this field of application is of special relevance as it enables customer-centric support, strengthening the customer-advisor relationship. In this use case, PA can determine the expected amount of loan required and indicate which product features a customer might need and can afford. It can also define a customer's price sensitivity and detect customer preferences, ranging from seeking the lowest price to the best service. This enables firms to classify customers with regard to their profitability and preselect which ones to offer a mortgage financing to, and which ones to refer to other providers. Consequently, interviewees mention a more efficient resource allocation and the possibility of tailoring marketing campaigns to individuals. After all, such tools support customer advisors in approaching their customers at the right time with a next-best, customized product. E3 and E7 emphasize that approaching a customer with such individualized solutions can occur when the customer is still unaware of his potential affinity for them.

The use case "Pre-KYC selection and fraud detection" applies to the process step *application*. This relates only to new customers and was mentioned by three interviewees. E5 mentions that the KYC-process is costly for banks; yet it is important to prevent avoidable effort and reject unsuitable customers in advance. PA models can support banks in detecting fraudulent behavior and conducting brief creditworthiness checks based on little available data. However, a negative pre-check must not necessarily result in a customer's rejection but might signal a customer advisor to take a more thorough look at the request.

The process step *review* contains a use case dealing with the "enhancement of credit scoring models" through the application of PA. It was mentioned by 10 interviewees, of whom five currently see it in application, three see a short-term potential, and two see a medium- to long-term potential. In this case, a PA model predicts potential future events in a customer's life which could affect their repayment ability. Thus, more informed decisions about giving out mortgage loans can be made. E1 and E8 point out that the collateral provided is normally able to cover most of the losses in case of a default, so the need to apply advanced models in this review step is not urgent.

Such a model is currently used for economic decisions, meaning whether or not to grant a loan. Applying such PA models to enable “regulatory” decisions, for example, determining how much capital is needed to back a mortgage loan, constitutes a further field of application in the *review* process step which was mentioned by four interviewees. Most agreed that PA cannot yet be applied to regulatory decisions due to the models’ inability to meet strict regulatory requirements. Only E12 mentioned an active usage of PA models for both economic and regulatory purposes.

In the process step *contract*, only one expert sees a use case with short-term potential. This deals with “steering employee capacity utilization” through the application of PA, and functions in the same way as the same use case in the process step *advice* (see Appendix 23).

In the *portfolio* process step, the “identification of non-performing loans” is mentioned by five firms, with two seeing a current application, two seeing short-term potential, and one seeing a medium- to long-term potential. The use case entails the prediction of upcoming events in a customer’s lifetime, or the identification of anomalies through the loan life-cycle, in order to identify customers needing more intensive help. This prevents potential payment delays, mitigates default risks, enables an organization to efficiently allocate its resources and, ultimately, strengthens the customer-advisor relationship. Nevertheless, as E8 emphasizes, this use case has even greater relevance in the general lending business, as in private mortgage financing the underlying collateral protects banks from defaults.

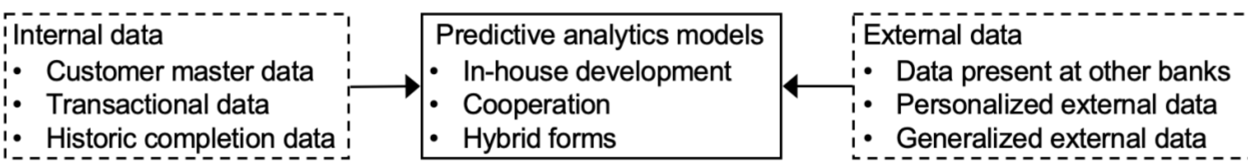
With two mentions each for short-term and medium- to long-term potential, the “proposal of cross-selling products” to a customer is another use case in the *portfolio* process step. PA models can propose the most suitable products for a customer from a portfolio of cross-selling products. These can range from insurance contracts to additional loans which might be required, for example, for household goods or house repairs. Based on experiences with the customer and their respective peer group, these products can be customized, and the collateral of the initial financed property can be reused as a security for the bank.

In the last process step *refinancing*, four interviewees mention the use case "offering customized refinancing options", with one seeing it already in use, two seeing a short-term and another seeing a medium- to long-term potential. Here, PA can suggest the best, customized refinancing options following the expiration of the initial loan. E13 highlights the similarities to "approaching customers with tailored products" – the use case identified in the *advice* process step – the only difference being the availability of better data as a basis for prediction, as the whole peer group now shares their mortgage financing history with the firm.

4.3 Types of Predictive Analytics Models and Data Sources Used

This section follows up on the second research question by outlining which types of PA models and which data sources are used to conduct predictive analyses in the private mortgage financing business. A summarized overview is displayed in Figure 3.

Figure 3: Data Sources for Predictive Analytics Models



The interviews revealed that both in-house developed PA solutions and solutions subject to charge are applied in organizations. In-house development is conducted with open-source programming languages, including ready-made packages. While both FinTechs and two banks rely entirely on in-house development, other firms cooperate with large technology companies, such as IBM, SAP, Salesforce, or Microsoft, and license and individualize their solutions. In addition, some firms not only cooperate with large technology companies, but also develop solutions themselves in parallel, and thus, work in a hybrid form. E4 mentions that smaller banks in particular have less internal know-how about data analytics and are therefore more dependent on the support of third parties. Consequently, they often enter into cooperative agreements and eventually complement these with in-house developments.

Regarding the type of models underlying the PA tools, interviewees revealed that they are experimenting with all types of classification and regression models for different use cases. This is because there is no one type of PA which is best suited for a particular use case; rather, it is dependent on the data available. For example, E14 mentions that they utilize gradient-boosted trees for the analysis of customer master data in order to identify the affinity of customers for mortgage financing. However, if a sequence of time-dependent data, such as transactional data, is available, they apply ANN-based PA models. E6 mentions that they use random forest and logistic regression models for the same use case. Additionally, E6 and E8 state the use of a range of regression-based PA tools for fraud detection and for the identification of influencing factors on (non-) defaulted loans. E4 emphasizes that decision tree or random forest models can easily be visualized which makes such models explainable for users without knowledge of the software. Furthermore, E14 talks about PA models based on gradient boosted trees which are used for credit scoring models. Once such PA models are developed and in application, they complement conventional tools and do not directly replace them.

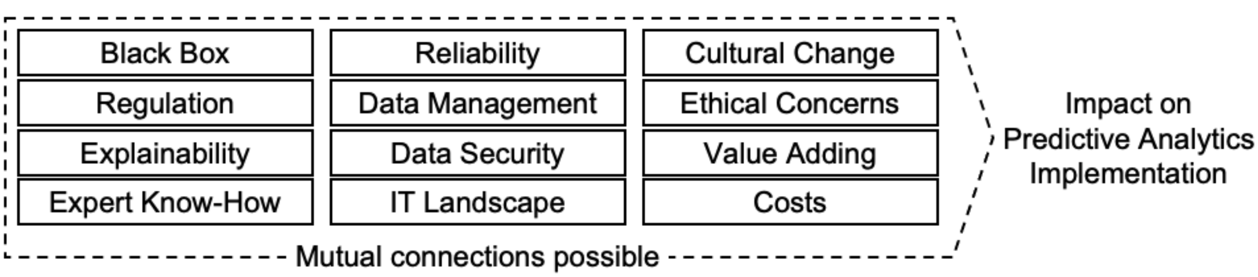
The data utilized varies from use case to use case and is also dependent on the availability of information, which can differ between potential and existing customers. E9 emphasizes that with good data management, the more access to data points is provided, the more accurately PA models can predict unknown events. For this reason, PA tools applied in later stages of a loan life cycle are able to make more precise predictions because they can be fueled with more data points as the customer and the peer-group have a history with the bank.

The interviews reveal that internal data used might include customers' master data, their transactional data, and historical data based on the behavior of existing portfolios. Due to regulatory requirements, however, firms require their customers' approval to analyze transactional data by machine. For external data, a variety of sources can be used. For example, E9 mentions that the Revised Payment Service Directive of the European Union enables banks to analyze a customer's data held by other banks, if the customer consents. E6 further emphasizes that combining data from a person's building society with data from local banks and comparison websites has the potential to enrich the data pool for more precise predictions in the medium- to long-term. Once a customer has applied for a loan, banks can also draw on personalized information from external credit rating agencies for their analysis. Additionally, the interviews reveal the use of tracking data from websites, micro-geographic data or Google Analytics data for analysis. Some interviewees also see medium- to long-term potential in incorporating data from social networks and other information freely available on the internet. The FinTech represented by E5 bases its PA models exclusively on psychometric data. Other experts also see a medium- to long-term potential in using such data points and can imagine that this could potentially make the collection of conventional documents, such as salary documents, obsolete.

4.4 Risks and Challenges Attached to the Application of Predictive Analytics

In the context of the third research question, this section explores the concerns accompanying the implementation of PA and identifies a total of 12 risks and challenges (see Figure 4).

Figure 4: Risks & Challenges of Predictive Analytics



Depending on the exact type of PA used, some models are considered to be "black boxes," as the cause-effect relationships inherent to the algorithms cannot always be explained. This can be problematic for customer advisors who want to know the reasons behind a tool's recommendations. Additionally, the issue impinges on processes such as regulatory credit scoring, for which regulators require a fully comprehensible decision process. E14 mentions that there are certain tools on the market explaining the input-output relations of algorithms. However, they cannot yet explain the interdependencies of initial and new features within a model; therefore, these are not applicable for use cases subject to regulation. Without going into details, E12 states that PA tools are actively used in their regulatory credit scoring process, solving the black box issue in some way. Attached to the development of PA models, there is the risk that data scientists modeling the PA models leave the firm. In this event, no-one would be able to understand and explain the functioning of the model, either to regulatory authorities or to customer advisors using the tools. Therefore, involving people in the development process is essential, E4 mentions.

A further challenge mentioned is the reliability of models. E4 stresses that the outputs should be consistent over time and become more precise with more data and more training. Although, as E8 mentions, users must be able to trust their models, but also accept that a small margin of error remains. In addition, there are concerns regarding the need for proper data management. The amount of data is no guarantee of quality; rather the aim is to gather the right data for evaluation. Yet not all data can be used for reasons of data security. Not only is customer consent required for the analysis of certain information, but also, as E12 emphasizes, meticulous attention needs to be paid to who has access to this information and who is allowed to process it.

Above all, PA models need to be effectively integrated in the existing system landscape. The often complex and historically grown systems are not easy to dock onto, not only for in-house developments but also for third-party solutions. E6 and E12 mention that creating a more adaptable platform or following an open banking strategy are possible solutions for this.

Along with these technical challenges, a firm also faces a cultural change with the introduction of PA. PA involves deviations from secure paths that have been used in the past, works in new ways, and as such is often not directly comprehensible to users. Also new is the emergence of ethical concerns attached to the application of PA models, as the models' operators need to ensure that no customer group is discriminated against in any way. Furthermore, E12 highlights that PA should not be applied blindly for the sake of it, but rather a firm should assess whether it actually adds value compared to more conventional solutions. Once suitable use cases are identified for follow-up, it is important to invest in the technology and the people responsible for its development as this enables mitigation of the above-mentioned concerns. However, the costs attached to this constitute a further challenge for the firms.

4.5 Potential for Cooperation with FinTechs

The interviews reveal that cooperation with large technology providers is seen as an option to counteract a lack of internal expertise, enable a timely implementation of PA models, and alleviate some associated challenges. Nevertheless, most interviewees also see potential in cooperating with smaller FinTechs offering suitable PA solutions for the private mortgage financing process. They associate the following advantages with FinTechs:

- Flexibility: No existing customer portfolio and historically grown systems to take care of.
- Customer-oriented: Solution-oriented thinking from a customer-centric perspective.
- Regulation: Starting from scratch, dealing with regulation and data security afterwards.
- Innovation: Focusing on specific products with all resources enables strong innovation.
- Agility: Lean processes and agile development enable quick testing of solutions on the market.

If a solution provided by a FinTech is promising, banks or service providers can either copy the idea, cooperate with the FinTech, or acquire it to integrate the solution, says E5. Cooperation or acquisition can also mitigate the weaknesses of FinTechs such as less financial power, absence of an initial customer base, and, accordingly, no access to valuable customer data for their PA models. Furthermore, E7 sees white-labelling the solution of FinTechs as an approach that includes innovative solutions without handing the customer to a third-party.

Nevertheless, no interviewee is yet aware of a FinTech in the German market with a PA solution for any steps in the private mortgage financing process. E5 mentions that his FinTech already cooperates with banks to integrate their PA solution in different product processes, but says the German market is slow to adopt the technology. Banks and regulators are conservative with regard to using psychometric data in PA models and to trusting in fully online processes. The other FinTech, represented by E2, has developed a ML solution to enhance their comparison platform for mortgage financing and installment credits. Several interviewees stress that they cooperate with such comparison websites, as they play an important role in the customer journey and are currently the only FinTechs in the private mortgage financing business attacking their profitability.

5 Discussion

In summary, all interviewees agree on the need to engage with PA in the private mortgage financing process in order to counteract the tightening market conditions described in Chapter 2.1.1. They emphasize that technological advancements and rich data availability enhance the development of PA solutions, as also mentioned by Mohr & Hürtgen (2018).

The first research question regards existing and potential fields of application of PA in the private mortgage financing process. While research on this subject is scarce and refers broadly to AI use in certain fields of application (see Chapter 2.3.2), this work project confirms the application of PA in these areas and also explores further use cases. All of them are presented in Figure 2, with their relevance being displayed based on topicality and number of mentions.

While only little research focuses on analyzing customer data to identify a customer's need for a specific product, for instance Rehfisch (2019) and Garg et al. (2017), it is the most prominent application of PA. Along with other less mentioned use cases such as providing customized products, recommending properties, cross-selling products, or proposing refinance options, this use of PA focuses on a higher product completion, which increases a bank's revenue. Interviewees point out that detecting a customer's demands as early as possible and subsequently tailoring solutions to them is essential in a competitive market. Different types of PA models, including ML-based gradient boosted tree or ANN-based models, are able to pursue such tasks in a more efficient way than conventional approaches, which rely on a more hands-on customer assessment.

Advancing credit models through the application of PA is also important. While most research deals with such advances in the lending business in general and discusses its potential for customers with no credit history (see Chapter 2.3.2), the interviews confirm that PA is also applied in the credit-scoring process of private mortgage financings. Although PA has been used in this field for some time, more advanced PA models are emerging due to the use of broader data sources, the availability of more computing power, and the inclusion of ML and DM elements. Nevertheless, this field of application is contentious, due to the differentiation between economic and regulatory models. While advanced PA models can certainly be used for economic decisions, most interviewees mention that the "black box" issue prevents them from applying it for regulatory decisions. One large bank currently does claim to use PA models for such decisions, although a different perception of PA characteristics cannot be excluded in this case. While most interviewees might refer only to advanced PA models, this particular respondent might refer to rather conventional PA models without DM and ML elements that face black box issues. Also, the application of PA in this use case aims to mitigate the default rate of mortgage loans, similar to the use case on identifying non-performing loans. Although these use cases are already in use, some experts play down the urgency of employing more sophisticated analytics as, in the event of a mortgage loan default, the collateral provided reduces a bank's ultimate loss.

Taking a holistic view, it is important to emphasize the structural parallels between the use cases mentioned and the levels of interaction. Thus, they should not only be seen as individual solutions but as interconnected links in a chain. Their application leads not only to more efficient resource allocation, higher completion, and lower default rates, but also to improved customer and customer advisor satisfaction. This in turn enables banks to establish a favorable customer relationship in the long run. This is especially relevant for a comprehensive product such as a private mortgage loan, particularly given the challenging market environment and cost pressures.

The second research question addresses the types of PA applied and data sources used by the interviewed firms. The interviews reveal that there is no right or wrong PA type for a specific use case but rather a lot of experimentation with different approaches is required. This is because the topic is constantly evolving, different data sources become available, and specific objectives can vary, as also stated by Kuhn & Johnson (2013). While we can observe examples of conventional regression models in use, there is also considerable ongoing development in terms of ANN-based and ML-based PA models. Such advanced models can provide more comprehensive insights into the same data sets addressed by rather conventional PA models.

While some PA use cases seem less relevant, either because they have fewer mentions or are not yet applied in practice, they should not be ignored. A low mention by the interviewees can be due to different possibilities. On the one hand, banks have limit limited resources at their disposal, which can lead to the prioritization of certain use cases. On the other hand, as the technology and its fields of application are constantly evolving, interviewees cannot be aware of all potential use cases and related PA model types. At the same time, some interviewees may not have disclosed all of their firms' use cases or types of PA models for competitive reasons. Consequently, it is critical to observe the further development of existing use cases and the emergence of new ones over the next few years, as well as to further explore different technical specifics of PA types used.

The empirical analysis shows that PA tools provided by technology companies find wide application among the interviewed firms. Especially for companies with smaller data analytics teams and limited know-how, third-party cooperation provides a solid starting point. With more experience available, firms can go beyond modifying third-party solutions to develop their own solutions in parallel. It should be critically evaluated, therefore, whether in-house developments or modified third-party solutions add more value. For this, the complexity of each use case, availability of relevant know-how, and individualization options of third-party solutions, among others, should be considered.

Most fields of application identified use internal data, meaning there is still considerable potential for the use of external data. While generalized external data is already frequently analyzed, recent developments in regulations enable an increase in the usage of data present at other banks. In comparison, personalized data, such as social media activities, browsing preferences, or psychometric testing, as mentioned by Dalela (2019), Ogbonna (2017), and Ryll et al. (2020), are rather expected to be used in the medium- to long-term.

Answering the third research question, the analysis confirms the main risks and challenges attached to the implementation of PA in a firm and identifies a total of 12 main concerns (see Figure 4). In line with the presented research in Chapter 2.3.3, the interviewees emphasize the potential of FinTechs for the financial industry. While no FinTechs with PA solutions are active yet in the German private mortgage financing market, most interviewees would embrace their emergence. On the whole, firms seek proactive collaborations with them as they value their agility and focus on specific process features, which is in line with research from Ryll et al. (2020). In order for this to work, an openness towards the use of more unconventional data sources, an open, integrable system landscape, and a more compatible regulatory framework, appear to be essential.

6 Conclusion

This work project identifies 17 use cases for the application of PA in the private mortgage financing process with current, short-term, and medium- to long-term potential. Thus, it provides banks with indications of where to advance their development of PA. The most relevant use cases include: identifying a customer's needs; offering customized products; and advancing scoring models. The application of PA enables banks to increase their product completion, reduce default rates, and allocate resources more efficiently. This is an essential prerequisite for building a long-term customer relationship and asserting oneself in the increasingly competitive environment of private mortgage financing. Additionally, this study finds that there is considerable experimentation with regard to types of PA models, as there is no best type available for a specific use case. It further outlines that while some firms develop PA models by themselves, other firms cooperate with large technology companies, individualize their solutions, and, depending on their available know-how, develop PA models by themselves as well. The data used for analysis differs depending on the use case and includes different types of both internal and external data. Furthermore, 12 major risks and challenges associated with the implementation of PA are ascertained, ranging from regulatory concerns over technical implementation hurdles to organizational questions. All of those should be carefully addressed when preparing an implementation roadmap. Finally, this paper demonstrates the awareness of banks with regard to the activity of FinTechs and elaborates on the potential for cooperation with them.

The primary research conducted for this study raises valuable implications not only for practitioners, but also for theorists. Focusing on the application of PA in the private mortgage financing business, the findings of this work project contribute to a specific field in the existing academic research. Other researchers can build on these findings, dive deeper into the technical specifics of the use cases outlined, and transfer this knowledge to other lending processes.

References

Abbott, Dean. 2014. *Applied Predictive Analytics: Principles and Techniques for the Professional Data Analyst*. Wiley.

AdviceRobo. n.d. "Retail Lending | AdviceRobo." Accessed November 21, 2020. https://advicerobo.com/retail-lending/.

Aggarwal, Charu C. 2018. *Neural Networks and Deep Learning: A Textbook*. Springer International Publishing.

Ahonen, Toni, Jyri Hanski, Teuvo Uusitalo, Henri Vainio, Susanna Kunttu, Pasi Valkokari, Helena Kortelainen, and Kari Koskinen. 2018. "Smart Asset Management as a Service - Deliverable 2.0." VTT Technical Research Centre of Finland. https://projectsites.vtt.fi/sites/smartadvantage/files/SMACC_SmartAssetManagement_julkaisu_032018_web.pdf.

Attaran, Mohsen, and Sharmin Attaran. 2018. "Opportunities and Challenges of Implementing Predictive Analytics for Competitive Advantage." *International Journal of Business Intelligence Research.* 9 (July).

Aunalytics. 2015. "Decision Trees: An Overview." *Aunalytics* (blog). January 31, 2015. https://www.aunalytics.com/decision-trees-an-overview/.

BAFA. n.d. "BAFA - Für Privatpersonen." Accessed July 18, 2020. https://www.bafa.de/DE/Home/Zielgruppeneinstiege/privatpersonen_node.html.

BaFin. 2018. "Big Data Trifft Auf Künstliche Intelligenz - Herausforderungen Und Implikationen Für Aufsicht Und Regulierung von Finanzdienstleistungen." https://www.bafin.de/SharedDocs/Downloads/DE/dl_bdai_studie.html;jsessionid=18135DE06E5B60C3471BCBBFB7D44938.1_cid383.

———. 2020. "Kredite und Verbraucherdarlehen." September 22, 2020.

https://www.bafin.de/DE/Verbraucher/Bank/Produkte/KrediteDarlehen/kredite_node.html
.

banken-auskunft. n.d. "Liste Privatbanken in Deutschland." Accessed July 18, 2020. https://www.banken-auskunft.de/privatbanken/deutschland.

Bari, Anasse, Mohamed Chaouchi, and Tommy Jung. 2016. *Predictive Analytics für Dummies*. 1st ed. Weinheim: Wiley-VCH.

Bauratgeber-Deutschland. n.d. "Welche Sicherheiten müssen Bauherren für ein Darlehen geben?" Accessed July 18, 2020. https://www.bauratgeber-deutschland.de/baufinanzierung/baufinanzierung-faq-zur-baufinanzierung/welche-sicherheiten-muessen-bauherren-fuer-ein-darlehen-geben/.

Bellens, Jan, and Karly Meekings. 2020. "Banking in the New Decade: Three Big Bets to Boost Profitability and Free up Capital to Invest in Transformation." Ernst & Young. February 10, 2020. https://www.ey.com/en_gl/banking-new-decade/why-global-banking-profitability-will-remain-a-challenge-in-2020.

Berger, Rob. 2016. "4 Smart Reasons To Refinance A Mortgage." Forbes. November 29, 2016. https://www.forbes.com/sites/robertberger/2016/11/29/4-smart-reasons-to-refinance-a-mortgage/.

Britt, Kyle J. 2015. "Apple Takes Predictive Analytics Mainstream." Medium. September 24, 2015. https://medium.com/@kylejbritt/apple-takes-predictive-analytics-mainstream-e34aaffa6e95.

Browne, Ryan. 2020. "Big Tech Will Push Deeper Into Finance This Year — But Avoid the 'Headache' of Being a Bank." CNBC. January 3, 2020. https://www.cnbc.com/2020/01/03/big-tech-will-push-into-finance-in-2020-while-avoiding-bank-regulation.html.

BVR. n.d. “Die Idee Der Genossenschaftsbanken.” Accessed July 18, 2020. https://www.vr.de/privatkunden/was-wir-anders-machen/genossenschaftsbank.html.

Byanjankar, Ajay, Markku Heikkilä, and Jozsef Mezei. 2015. “Predicting Credit Risk in Peer-to-Peer Lending: A Neural Network Approach.” In , 719–25. Cape Town, South Africa: IEEE. https://doi.org/10.1109/SSCI.2015.109.

Dalela, Megha. 2019. “Disruptive Technologies Shaping the Lending of Tomorrow.” *Nucleus Software* (blog). April 25, 2019. https://www.nucleussoftware.com/blog/lending/disruptive-technologies-shaping-the-lending-of-tomorrow/.

Deutsche Bundesbank. 2020. “Wohnungsbaukredite an private Haushalte / Hypothekarkredite auf Wohngrundstücke.” December 2, 2020. https://www.bundesbank.de/de/statistiken/geld-und-kapitalmaerkte/zinssaetze-und-renditen/wohnungsbaukredite-an-private-haushalte-hypothekarkredite-auf-wohngrundstuecke-615036.

Discover. n.d. “Why Refinance - Why Should I Refinance My Mortgage.” Accessed July 20, 2020. https://www.discover.com/home-loans/articles/top-three-reasons-to-refinance/.

Dölle, Christian. 2018. “Projektsteuerung in der Produktentwicklung mittels Predictive Analytics.” PhD diss., Aachen, Germany: RWTH Aachen University.

Douqué, Sandra, Hans-Jürgen Scharf, and Johannes Albrecht. 2020. “Baufinanzierungen über Vermittlerplattformen.” *BankingHub* (blog). February 9, 2020. https://bankinghub.de/banking/operations/baufinanzierungen-ueber-vermittlerplattformen.

drklein. n.d. “Immobiliendarlehen Ablauf: Wie Geht Baufinanzierung?” Accessed July 18, 2020a. https://www.drklein.de/ablauf-baufinanzierung.html.

———. n.d. “Versicherungen Für Immobilienbesitzer Bei Dr. Klein.” Accessed July 20, 2020b. https://www.drklein.de/immobilien-versicherung.html.

Eckerson, Wayne W. 2007. “Predictive Analytics - Extending the Value of Your Data Warehousing Investment.” *TDWI Best Practices Report* 1st Quarter.

Engel&Völkers. n.d. “Fallen jetzt die Immobilienpreise durch die Corona-Krise?” Accessed August 18, 2020. https://www.engelvoelkers.com/de-de/immobilienpreise/corona-krise/.

Faggella, Daniel. 2020. “Artificial Intelligence Applications for Lending and Loan Management.” Emerj. April 3, 2020. https://emerj.com/ai-sector-overviews/artificial-intelligence-applications-lending-loan-management/.

Finlay, Steven. 2014. *Predictive Analytics, Data Mining and Big Data: Myths, Misconceptions and Methods*. UK: Palgrave Macmillan.

Fleischer, Klaus. 2015. “Wandel Der Vertriebswege in Der Baufinanzierung – Siegeszug Unabhängiger Vermittler.” *Immobilien & Finanzierung* 9: 16–18.

Fontinelle, Amy. 2020. “First-Time Homebuyer’s Guide.” Investopedia. March 20, 2020. https://www.investopedia.com/updates/first-time-home-buyer/.

Forest, Helena, and Donya Rose. 2015. “Delighting Customers and Democratising Finance: Digitalisation and the Future of Commercial Banking.” Deutsche Bank. https://cib.db.com/docs_new/GTB_Digitalisation_Whitepaper_(DB0388)_v2.pdf.

Garg, Amit, Davide Grande, Gloria Macías-Lizaso Miranda, Christoph Sporleder, and Eckart Windhagen. 2017. “Analytics in Banking: Time to Realize the Value.” McKinsey. https://www.mckinsey.com/industries/financial-services/our-insights/analytics-in-banking-time-to-realize-the-value.

Gartner. 2013. “Magic Quadrant for BI Platforms. Analytics Value Escalator.” https://www.gartner.com/en/documents/2326815/magic-quadrant-for-business-intelligence-and-analytics-p.

Gaumert, Uwe. 2019. “Ergebnisse der BdB-Auswirkungsstudie zur europäischen Umsetzung von

Basel IV-Auswirkungen auf ausgewählte Kreditportfolios." Bankenverband. https://bankenverband.de/newsroom/presse-infos/studie-zeigt-negative-auswirkungen-von-basel-iv/.

Gibson, William, and Andrew Brown. 2009. *Working with Qualitative Data*. London, UK: SAGE Publications.

Goldschmidt, Andreas. 2020. "Baufinanzierung in 7 Schritten: Die ultimative Anleitung." *bauherren.net* (blog). January 25, 2020. https://bauherren.net/baufinanzierung/.

Hacikura, Ahmed, and Sayako Seto. 2014. "Mortgage Cross-Sell." Oliver Wyman. https://www.oliverwyman.com/content/dam/oliver-wyman/v2/publications/2014/jun/Oliver_Wyman_Mortgage_Cross_Sell_The_Elusive_Opportunity.pdf.

Hazen, Benjamin T., Christopher A. Boone, Jeremy D. Ezell, and L. Allison Jones-Farmer. 2014. "Data Quality for Data Science, Predictive Analytics, and Big Data in Supply Chain Management: An Introduction to the Problem and Suggestions for Research and Applications." *International Journal of Production Economics* 154 (August): 72–80.

Henke, Nicolaus, Jacques Bughin, Michael Chui, James Manyika, Tamim Saleh, Bill Wiseman, and Guru Sethupathy. 2016. "The Age of Analytics: Competing in a Data-Driven World." McKinsey. https://www.mckinsey.com/~/media/mckinsey/industries/public%20and%20social%20sector/our%20insights/the%20age%20of%20analytics%20competing%20in%20a%20data%20driven%20world/mgi-the-age-of-analytics-full-report.pdf.

Hilmes, Christian. 2018. "Beratungs- und Servicequalität: Das sind die 10 besten Baufinanzierer Deutschlands." Das Investment. July 20, 2018. https://www.dasinvestment.com/beratungs-und-servicequalitaet-das-sind-die-10-besten-baufinanzierer-deutschlands/.

IDC. 2020. “IDC’s Global DataSphere Forecast Shows Continued Steady Growth in the Creation and Consumption of Data.” May 8, 2020. https://www.idc.com/getdoc.jsp?containerId=prUS46286020.

immonet. n.d. “10 Gute Gründe Für Bauen Statt Mieten.” Accessed July 20, 2020. https://www.immonet.de/service/bauen-statt-mieten.html.

immoverkauf24. n.d. “Bonitätsprüfung.” immoverkauf24 GmbH. Accessed July 18, 2020. /baufinanzierung/baufinanzierung-a-z/bonitaetspruefung/.

interhyp. n.d. “Sicherheiten bei der Baufinanzierung » Baufinanzierungsrechner.net.” Accessed July 18, 2020. https://baufinanzierungsrechner.net/baufinanzierungslexikon/sicherheiten/.

Investopedia. 2011. “9 Reasons To Buy A House Now.” Forbes. June 3, 2011. https://www.forbes.com/sites/investopedia/2011/06/03/9-reasons-to-buy-a-house-now/.

———. 2020. “When (and When Not) to Refinance Your Mortgage.” August 26, 2020. https://www.investopedia.com/mortgage/refinance/when-and-when-not-to-refinance-mortgage/.

Kamp, Meike, and Thilo Weichert. 2005. “Scoringsysteme Zur Beurteilung Der Kreditwürdigkeit - Chancen Und Risiken Für Verbraucher.” Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein. https://www.datenschutzzentrum.de/uploads/projekte/scoring/2005-studie-scoringsysteme-uld-bmvel.pdf.

Kelleher, John D., Brian Mac Namee, and Aoife D’Arcy. 2020. *Fundamentals of Machine Learning for Predictive Data Analytics*. 2nd ed. MIT Press.

KfW. n.d. “KfW-Förderprodukte für Ihren Neubau.” Accessed July 18, 2020. https://www.kfw.de/inlandsfoerderung/Privatpersonen/Neubau/Förderprodukte/Förderprodukte-PB-Neubau.html.

Klein, Melanie. 2020. “Implications of Negative Interest Rates for the Net Interest Margin and

Lending of Euro Area Banks." *Deutsche Bundesbank Discussion Paper* 10: 38.

Kotu, Vijay, and Bala Deshpande. 2015. *Predictive Analytics and Data Mining: Concepts and Practice with RapidMiner*. Elsevier.

Kreditrechner. 2015. "Wie eine Baufinanzierung abläuft." November 30, 2015. https://www.kreditrechner.com/ratgeber/ablauf-baufinanzierung/.

Kuhn, Max, and Kjell Johnson. 2013. *Applied Predictive Modeling*. New York: Springer.

Kumar, Shaily. 2018. "The Differences Between Machine Learning And Predictive Analytics." March 15, 2018. https://www.digitalistmag.com/digital-economy/2018/03/15/differences-between-machine-learning-predictive-analytics-05977121/.

Lange, Kathy. 2006. "Differences Between Statistics and Data Mining - ProQuest." *Digital Medievalist Review* 16 (12): 32.

Lantz, Brett. 2019. *Machine Learning with R: Expert Techniques for Predictive Modeling*. 3rd ed. Packt Publishing Ltd.

Larose, Daniel T. 2015. *Data Mining and Predictive Analytics*. 2nd ed. John Wiley & Sons.

Leichsenring, Hansjörg. 2018. "10 gute Gründe für den Kauf einer Wohnung oder eines Hauses." *Der Bank Blog* (blog). June 5, 2018. https://www.der-bank-blog.de/ratgeber/immobilienkauf/1833/.

Mayer, H. O. 2013. "Interview Und Schriftliche Befragung: Grundlagen Und Methoden Empirischer Sozialforschung: Grundlagen Und Methoden Empirischer Sozialforschung." De Gruyter Oldenbourg. https://www.degruyter.com/view/title/316955.

McWaters, R. Jesse, and Rob Galaski. 2017. "Beyond Fintech: A Pragmatic Assessment Of Disruptive Potential In Financial Services." World Economic Forum & Deloitte. http://www3.weforum.org/docs/Beyond_Fintech_-_A_Pragmatic_Assessment_of_Disruptive_Potential_in_Financial_Services.pdf.

Microstrategy. 2018. “Predictive Modeling: The Only Guide You Need.” 2018. http://www.microstrategy.cn/us/resources/introductory-guides/predictive-modeling-the-only-guide-you-need.

Misoch, Sabina. 2015. *Qualitative Interviews*. De Gruyter Oldenbourg.

Mohr, Niko, and Holger Hürtgen. 2018. “Achieving Business Impact with Data.” McKinsey. https://www.mckinsey.com/~/media/mckinsey/business%20functions/mckinsey%20analytics/our%20insights/achieving%20business%20impact%20with%20data/achieving-business-impact-with-data_final.ashx.

Möst, Fidelius. 2019. “‚Time to Credit’: Schnelligkeit Als Ein Erfolgsfaktor in Der Baufinanzierung.” *Berg Lund & Company* (blog). June 17, 2019. https://berg-lund.de/blog/time-to-credit-schnelligkeit-als-ein-erfolgsfaktor-in-der-baufinanzierung?backLink=blog.

Nyce, Charles. 2007. “Predictive Analytics White Paper.” American Institute for CPCU. https://www.the-digital-insurer.com/wp-content/uploads/2013/12/78-Predictive-Modeling-White-Paper.pdf.

Ogbonna, Prince. 2017. “Application of Artificial Intelligence in the Lending and Loan Management Process.”

Perrier, Alexis. 2017. *Effective Amazon Machine Learning*. Packt. https://hub.packtpub.com/predictive-analytics-with-amazon-ml/.

PWC. 2017. “Effizienz Der Kreditprozesse 2017.” https://www.pwc.de/de/finanzdienstleistungen/banken/pwc-studie-effizienz-der-kreditprozesse-2017.pdf.

———. 2020. “Der große Baufi-Boom - 2019: Rekordwachstum & Margenerholung.” https://www.pwc.de/de/finanzdienstleistungen/pwc-der-grosse-baufi-boom.pdf.

Ray, Sunil. 2015. “Regression Techniques in Machine Learning.” *Analytics Vidhya*, August 13, 2015. https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/.

Rehfisch, Michael. 2019. “Smarter Kreditprozess dank künstlicher Intelligenz.” Knowis. October 28, 2019. https://www.knowis.com/de/blog/smarter-kreditprozess-dank-kuenstlicher-intelligenz.

Rode, Matthias, Thomas Kurth, Jennifer Rabener, and Antonia Föhse. 2018. “FinTechs Auf Der Überholspur.” *Compliance Business*, no. 4 (November): 15–18.

Rokach, Lior, and Oded Maimon. 2015. *Data Mining with Decision Trees: Theory and Applications*. 2nd ed. Hackensack, New Jersey: World Scientific.

Rottwilm, Christoph. 2020. “Immobilienbranche erwartet Preisanstieg trotz Corona - manager magazin - Finanzen.” Manager Magazin. May 19, 2020. https://www.manager-magazin.de/finanzen/immobilien/immobilienbranche-erwartet-preisanstieg-trotz-corona-a-1307172.html.

Ryll, Lukas, Mary Emma Barton, Bryn Zheng Zhang, and Jesse McWaters. 2020. “Transforming Paradigms - A Global AI in Financial Services Survey.” University of Cambridge, World Economic Forum. http://www3.weforum.org/docs/WEF_AI_in_Financial_Services_Survey.pdf.

Samuel, Arthur L. 1959. “Some Studies in Machine Learning Using the Game of Checkers.” *IBM Journal of Research and Development* 3 (3): 210–29.

Sapountzis, Ted. 2019. “IT Decision-Making - Gartner’s Analytics Maturity Model.” *Numerify* (blog). September 3, 2019. https://www.numerify.com/blog/it-decision-making-through-the-lens-of-gartners-analytics-maturity-model/.

SAS. n.d. “Predictive Analytics: What It Is and Why It Matters.” Accessed June 30, 2020. https://www.sas.com/en_us/insights/analytics/predictive-analytics.html.

Scherff, Dyrk. 2017. "Ehrlich durchgerechnet: Wie viel Haus kann ich mir leisten?" *Faz.net*, January 5, 2017. https://www.faz.net/1.4603036.

Schulze, Eike, and Anette Stein. 2013. *Immobilien- und Baufinanzierung*. 2nd ed. Haufe-Lexware.

Selamat, Siti Aishah Mohd. 2018. "Survey on the Emergence of Predictive Analytics." Department of Management and Business Systems at University of Bedfordshire. https://www.researchgate.net/publication/323116271_Survey_on_the_Emergence_of_Predictive_Analytics.

Singh, Aarushi. 2019. "How AI Is Taking Predictive Analytics to the Next Level." *MarTech Vibe* (blog). August 20, 2019. https://www.martechvibe.com/insights/staff-articles/how-ai-is-improving-predictive-analytics/.

Sparkassenzeitung. 2018. "Baufinanzierung: Makler Wollen Die Plattform - SparkassenZeitung." April 17, 2018. https://www.sparkassenzeitung.de/vertrieb/baufinanzierung-makler-wollen-die-plattform.

Sreedhar, Brinda. 2019. "Vereinfachung von Hypothekenanträgen mit KI und RPA." Automation Anywhere. April 25, 2019. https://www.automationanywhere.com/de/blog/automation-anywhere-news/how-artificial-intelligence-and-rpa-can-help-lenders-process-mortgage-applications-faster.

Streit, Matthias. 2020. "Marktprognose: Immobilienbranche fürchtet Coronafolgen – Stimmung bricht ein." June 25, 2020. https://www.handelsblatt.com/finanzen/immobilien/marktprognose-immobilienbranche-fuerchtet-coronafolgen-stimmung-bricht-ein/25947020.html.

Tahedl, Christoph. 2019. "Potential Künstlicher Intelligenz im Kredit- und Forderungsmanagement." *IT Finanzmagazin* (blog). February 19, 2019. https://www.it-finanzmagazin.de/ki-forderungsmanagement-85509/.

Tax, Niek, Ilya Verenich, Marcello La Rosa, and Marlon Dumas. 2017. "Predictive Business Process Monitoring with LSTM Neural Networks." In *29th International Conference on Advanced Information*, 10253:477–92.

TDWI. 2014. "Predictive Analytics:Revolutionizing Business Decision Making." https://www.sas.com/content/dam/SAS/en_us/doc/whitepaper2/tdwi-predictive-analytics-107459.pdf#Find%20out%20more%20about%20predictive%20analytics.

Team EA. 2020. "Prediction Using Neural Networks." *Express Analytics* (blog). February 27, 2020. https://expressanalytics.com/blog/prediction-using-neural-networks/.

Thiel, Diederick van, and Willem Frederik van Raaij. 2019. "Artificial Intelligence Credit Risk Prediction: An Empirical Study of Analytical Artificial Intelligence Tools for Credit Risk Prediction in a Digital Era." Henry Stewart Publications. June 1, 2019. https://www.ingentaconnect.com/content/hsp/jrmfi/2019/00000012/00000003/art00008.

Tschakert, Norbert, Julia Kokina, Stephen Kozlowski, and Miklos Vasarhelyi. 2016. "The next Frontier in Data Analytics." *Journal of Accountancy* 222 (August): 58.

Verbraucherzentrale. 2020. "Fördermittel vom Staat." February 27, 2020. https://www.verbraucherzentrale.de/wissen/geld-versicherungen/bau-und-immobilienfinanzierung/foerdermittel-vom-staat-5815.

WellsFargo. n.d. "Why Refinance a Mortgage." Accessed July 20, 2020. https://www.wellsfargo.com/mortgage/mortgage-refinance/why-refinance/.

White, Danni. 2020. "Digital Banking Challenges and Opportunities For the Banking Industry." *Techfunnel* (blog). February 3, 2020. https://www.techfunnel.com/fintech/digital-banking-challenges-opportunities/.

Winters, Ralph. 2017. *Practical Predictive Analytics*. Packt Publishing Ltd.

Wirtschaftslexikon24. n.d. "Sparkassen - Wirtschaftslexikon." Accessed July 18, 2020.

http://www.wirtschaftslexikon24.com/d/sparkassen/sparkassen.htm.

Wollseifen, Christina. 2017. "Mini-Zinsen, Zulagen, Familien-Bonus: So bezahlt der Staat Ihr Haus." Focus Online. September 9, 2017. https://www.focus.de/immobilien/finanzieren/foerderung-fuers-eigenheim-mini-zinsen-zulagen-familien-bonus-so-hilft-der-staat-beim-hauserwerb_id_4053338.html.

Wuestenrot. n.d. "Baufinanzierung mit Top-Zinsen." Accessed July 18, 2020. https://www.wuestenrot.de/de/baufinanzierung/uebersicht.html.

Zafra, Sonia. 2019. "How Predictive Analytics Is Transforming Fintech and Financial Services." *Fintech News* (blog). July 9, 2019. https://www.fintechnews.org/how-predictive-analytics-is-transforming-fintech/.

Zaleo Digital. n.d. "Bausparkassen in Deutschland." Accessed July 18, 2020. https://bausparvertrag.com/bausparkassen/.

Zapf, Marina. 2020. "Immobilien: Corona-Krise verunsichert Käufer wie Verkäufer." Capital. June 19, 2020. https://www.capital.de/immobilien/immobilienmarkt-corona-krise-verunsichert-kaeufer-wie-verkaeufer.

Zollinger, Gery. 2020. "Predictive Analytics in Banking and Wealth Management." *it-daily.net* (blog). June 10, 2020. https://www.it-daily.net/it-management/big-data-analytics/24433-predictive-analytics-in-banking-und-wealth-management.

List of Appendices

Appendix 1: List of Tables

Appendix 2: List of Figures

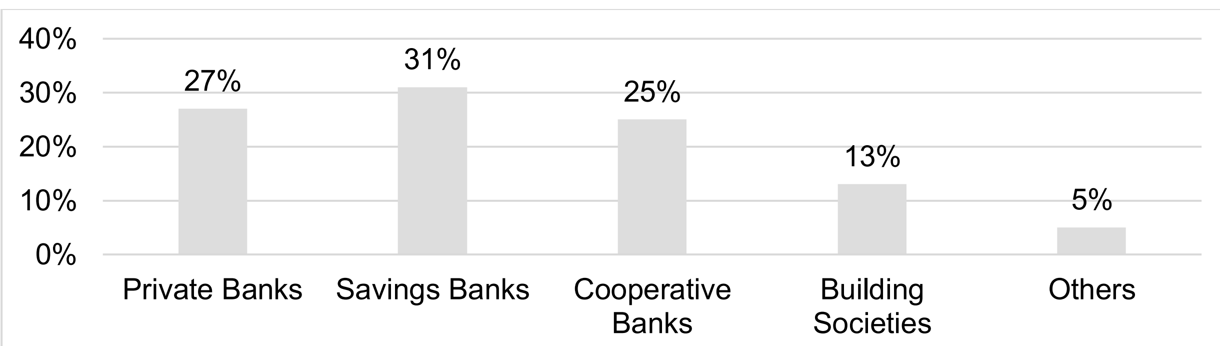
Appendix 3: List of Abbreviations

AI	Artificial Intelligence
ANN	Artificial Neural Network
DM	Data Mining
FinTech	Financial Technology (Startup)
ML	Machine Learning
PA	Predictive Analytics

Appendix 4: Bank Types Offering Mortgage Loans

Figure 5 gives an overview about the market share of several types of banks offering mortgage loans in the German market in 2019. Private banks include banks such as Deutsche Bank, Commerzbank or Unicredit Bank (banken-auskunft n.d.). The public savings banks, which are legal persons under public law, and cooperative banks, which are owned by their customers who hold cooperative shares, are situated all over Germany (BVR n.d.; Wirtschaftslexikon24 n.d.). Building societies serve the sole purpose to offer mortgage loans, belong to private or public institutions and include institutions such as LBS Bausparkasse, Bausparkasse Schwäbisch Hall or Wüstenrot Bausparkasse (Zaleo Digital n.d.). Further institutions include insurances among others, who under certain circumstances also give out mortgage loans (Hilmes 2018).

Figure 5: Market Share of Mortgage Loans in 2019



Source: PWC (2020)

Appendix 5: Interrelations and Functions within the Customer Journey of the Private Mortgage Financing Process

In the beginning of the process is the *need* of the customer to own an own property which can arise due to different reasons. Owning a property is often seen as a value investment or pension provision as there will be no need to pay rent in the long-term and there is no dependency from a landlord leaving space for personal individualization of the property (immonet n.d.; Leichsenring 2018; Scherff 2017). Further factors convincing people to take on mortgage loans are favorable interest rates in a low interest environment, the availability of public subsidies and the need for a larger home due to family growth (Investopedia 2011; Leichsenring 2018; Wollseifen 2017; immonet n.d.).

Once the customer identified the need for an own property, it is essential to gather *information* about private mortgage loan providers and the required scope of the loan. This process step partly overlaps with the following process steps of the traditional bank process, as such information can be gathered during advisory meetings with banks. Nevertheless, also intermediary agents or comparison websites can give a first overview over various market players, their mortgage loan conditions and help customers define the scope of their loan (Fontinelle 2020; Kreditrechner 2015). Once the exact need is defined, customers should collect a few non-binding offers from different providers, in order to then decide on one of them and proceed with the traditional bank process (drklein n.d.).

In the process step *advise*, the bank advisor and the customer get to know each other and analyze the customer's needs. This can occur online or on site (Möst 2019; PWC 2017). Different mortgage loan configurations are discussed[5], possibilities for government grants are evaluated[6] and

[5] Such configurations include maturity, interest rate, amortization or equity contribution among others (drklein n.d.).

[6] In Germany, such grants are provided by the Bank of Reconstruction (KfW) and Federal Office of

a tentative offer is presented (drklein n.d.; PWC 2017; Schulze and Stein 2013).

The next process step *application* includes the collection of necessary documents for the full identification of the customer and the creditworthiness and solvency check[7] which are required by the bank to proceed with their loan decision (immoverkauf24 n.d.; PWC 2017). In this step, the bank also checks for possible guarantees and collects data on the object to be financed[8] which banks' usually use as collateral for the mortgage loan[9] (interhyp n.d.).

In the *review* step, banks use the previously collected information to conduct the creditworthiness and solvency check – often, via an internal credit-scoring model (Kamp and Weichert 2005). In this context, the available guarantees and collateral including its respective insurances are examined (Bauratgeber-Deutschland n.d.; interhyp n.d.). In the end, the relevant department issues a vote for the mortgage loan decision.

Based on the loan decision in the previous process step, a contract is or is not issued in the *contract* process step (PWC 2017). The customer can either reject the proposed contract or accept it with his/her signature.

The signing of the contract initiates the next process step *disbursement*. After an eventual

---

Economics and Export Control (BAFA) for the use of regenerative energies, for families with children or for building energy-efficiently, among others (BAFA n.d.; KfW n.d.). Additionally, the government provides subsidies for certain people via Wohn-Riester and federal states also often provide own subsidy programs (Verbraucherzentrale 2020). Furthermore, the KfW provides promotional loans of up to € 100,000 for the purchase or the building of any privately used mortgage (KfW n.d.).

[7] Documents on the financial conditions of the customer are collected, including information on the customer's living conditions, job situation, income and wealth among others. Also, their prior payment behavior is analyzed by checking prior business with the bank and collecting data from neutral credit agencies, such as for example SCHUFA or Creditreform Boniversum. Such information include the number of further ongoing or past credits, credit cards, insolvencies, past defaults or deferred payments, fraudulent behavior and a score ascertained by the credit agency (immoverkauf24 n.d.).

[8] Documents on an object to be financed may vary due to different types of mortgage loans as. Such documents include a land register entry, building plans, insurance and documents on construction progress among others (drklein n.d.).

[9] Usually, a mortgage or land charge gets entered in the land register plan of the property which serves the bank as collateral (interhyp n.d.). However, also further mortgages, securities, precious metals and savings can be called upon as collateral for the mortgage loan (interhyp n.d.).

new account opening, the mortgage loan can be transferred in agreement with the customer (PWC 2017). Depending on the contract, it can be disbursed either at once or in installments according to the progress of construction. Instead of directly being paid to the borrower, the loan can also be paid out directly to the respective third parties which are connected to the building, modernization or purchase of the property (drklein n.d.).

Following this step, documents are archived and the mortgage loan is transferred to the banks' loan *portfolio* (PWC 2017). This step includes various business with existing customers, including the management of available collateral, the checking of reliability of repayments, the conduction of workout managements, the sale of connected products and the refinancing of mortgage loans (PWC 2017).

While negotiating with different mortgage loan providers and waiting for their offers, there are further concerns, customers have to take care of. Consequently, these *related needs* are also partly overlapping with the prior steps of the traditional bank process. Before the bank will payout a mortgage loan, it is required that the property sales contract is concluded, a land charge is registered in the land register and required property insurances are existent (drklein n.d.; Goldschmidt 2020; Kreditrechner 2015). Before the conclusion of the sales contract, however, the seller often already requires a pre-approval for the mortgage loan from the bank (Goldschmidt 2020). In case customers do not use the mortgage loan to buy a property but to develop one, they might have to coordinate with craftsmen, architects and construction companies among others in order to gather relevant documents required by banks before loan payout (Goldschmidt 2020). Furthermore, customers have to pay land transfer taxes and eventually brokerage fees (drklein n.d.). Once the mortgage loan has been paid out and the property is in the customer's own possession or is being built, other factors arise, such as moving, change of registration, purchase of furniture and consumer goods, and the consideration to craft certain things by oneself.

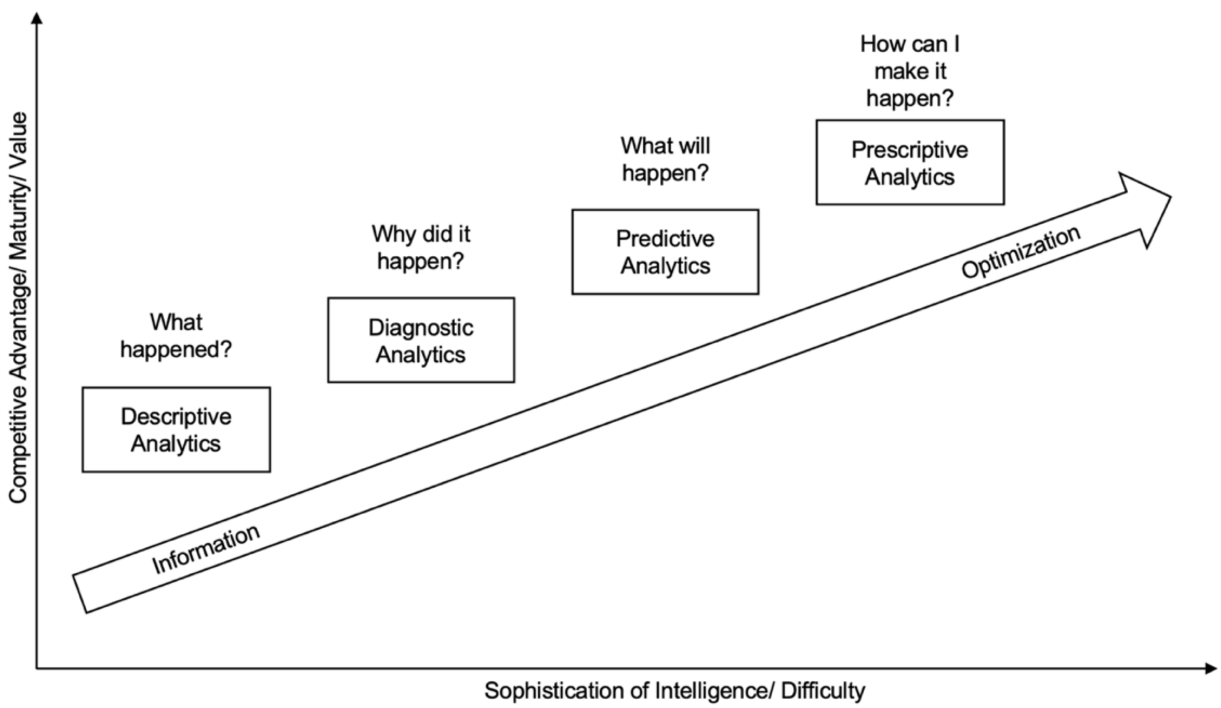
At some point in time, customers might consider a *refinancing* of the initial mortgage loan. This can for example be due to the availability of more favorable interest rates or the wish to switch from fixed interest rates to variable interest rates and vice versa (Berger 2016; Investopedia 2020). Furthermore, the customer might want to lower the monthly interest payment and therefore, extend the loan period (Berger 2016). In case enough capital is available, refinancing might lead to a shortened loan period with higher monthly interest payments (Investopedia 2020). Another reason is that the customer might want to tap into equity and get a higher mortgage loan than the remaining one in order to get some cash out of the refinancing (Discover n.d.; Investopedia 2020; WellsFargo n.d.).

After all, it is to be mentioned, that customers require a quick and qualitative process with cheap conditions which challenges the financial institutions to continuously improve their process performance and still offer competitive conditions (PWC 2017).

Appendix 6: Gartner's Maturity Model for Analytics

According to Gartner's Maturity Model for Analytics (2012), data analytics can be broken down into backward looking descriptive and diagnostic analytics, and forward looking predictive and prescriptive analytics (see Figure 6). While descriptive analytics provides insights into data by applying metrics on them, diagnostic analytics examines the root cause of past results. Compared to PA, which predicts the likelihood of occurrence of certain future events, prescriptive analytics recommends the best course of action to take (Gartner 2013; Ahonen et al. 2018; Sapountzis 2019; Tschakert et al. 2016).

Figure 6: Gartner's Maturity Model for Analytics

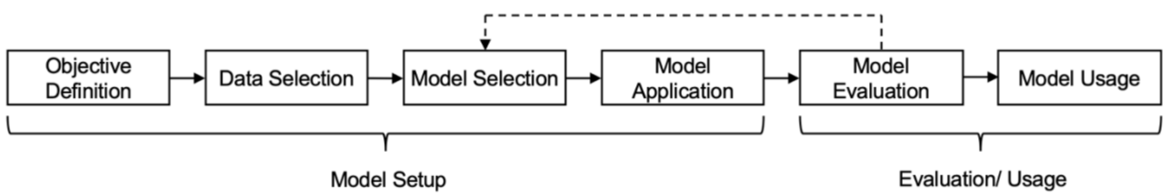


Source: Illustration adapted from Gartner (2012)

Appendix 7: Setup and Use of Predictive Analytics Models

The setup of PA models occurs in six steps (see Figure 7). In a first step, the business objective and desired outcomes for the project are defined and converted into PA tasks. In a second step, source data gets analyzed and the most appropriate data for the project gets selected and extracted upon which the model can be created. Afterwards, a suitable methodology or a mix of such gets selected based on the objective of the project and the available data. Once the model is built with an appropriate method, the underlying data need to be processed and its results need to be applied on processes or business decisions. This includes embedding the model into applications for automation or sharing data insights with business users. Once a model is implemented, a continuous evaluation of the model enables an improved performance, controlled access, reduction of redundant activities, standardization of toolsets and promotion of reuse.

Figure 7: Setup of a Predictive Analytics Model



Source: Illustration adapted from Eckerson (2006), Winters (2017) and Dölle (2018)

Appendix 8: Classification and Regression Models

Table 2: Classification and Regression Models

Model	Classification models	Regression models	Classification and Regression models
Linear regression		X	
Partial least squares			X
Ridge regression			X
Elastic net/ lasso			X
Neural networks			X
Support vector machines			X
MARS/ FDA			X
K-nearest neighbors			X
Single trees			X
Model trees/ rules		X	
Bagged trees			X
Random forest			X
Boosted trees			X
Cubist		X	
Logistic regression	X		
(LGRM)DA	X		
Nearest shrunken centroids	X		
Naïve Bayes	X		
C5.0	X		

Source: Kuhn & Johnson (2013)

For detailed information on these techniques, reference is made to the corresponding literature (Dölle 2018; Kuhn and Johnson 2013; Lantz 2019; SAS n.d.; Tax et al. 2017; Aggarwal 2018; Rokach and Maimon 2015; Kotu and Deshpande 2015; Bari, Chaouchi, and Jung 2016; Larose 2015).

Appendix 9: Guided Expert Interview Framework (English)

Questionnaire Work Project	
Title	*Potential for Banks' Private Mortgage Financing Business through the Application of Predictive Analytics*
University	Nova School of Business and Economics
Supervisor	Professor João Pereira
Author	Dominik Laufs
Year	2020
Structure	
Part 1	General Introduction
Part 2	Topic Introduction
Part 3	Introductory Questions
Part 4	1st Research Question
Part 5	2nd Research Question
Part 6	3rd Research Question
Part 7	Conclusion
Research Questions of the Work Project	
1st Research Question	What are the existing and potential fields of application of predictive analytics in the private mortgage financing process?
2nd Research Question	What type of predictive analytics is applied and what data sources are used?
3rd Research Question	What risks and challenges are attached to the implementation of PA

Part 1: General introduction	Topics	Details	Notes
Introduction Interviewer	Thank you!	*Thanks for participating in the interview and contributing to my work project*	
	Personal background	*Dominik Laufs, International Master in Finance, Nova SBE, 3. Semester, Work Project*	
	Work project: title	*"Potential for Banks' Private Mortgage Financing Business through the Application of Predictive Analytics"*	
	Work project: research questions	*Introduction to research questions*	
	Work project: goal	*Close research gap (see research questions)*	
		Give banks and fintechs an overview about best practices and outline which adjustments can be made	
	Function of interview	*Best practices and technical information of leading institutions*	
Information on interview	Audio recording of Interview	*In order to paraphrase the interview. Provided to supervisor and correctors*	
	Anonymization	*Anonymizing the institution's and interviewee's name in the work project in order to protect the data. Unless an explicit reference is desired*	
	Procedure	*Introduction to the topic and queries along the research questions*	
Introduction interviewee	Name		
	Corporation		
	Business model of corporation		
	Position/ Function in the corporation		
	others		

Part 2: Topic introduction	Topic	Questions	Notes
Mortgage financing	Relevance	*Highlight relevance of mortgage financing and outline the current market environment*	
		(Present graphs & bullets on the market - see Tableau 1)	
	Process definition	*(Present the defined process - see Figure 1)*	
Predictive Analytics	Definition	*Outline definition of predictive analytics*	
		(Present bullets on predictive analytics definition & distinction - see Tableau 2)	

Tableau 1: Mortgage financing market environment	
High share in total credit volume	Private mortgage loan is highly important for German banks, making out 42.4% of their credit volume in 2019
Comparatively constant interest rates	Interest rates of private mortgage loans are relatively constant around 1% over the last 4-5 years and did not decrease such as interest rates of other loan types
Anchor product	Private mortgage loan function as an anchor product to banks due to long maturities and the financing of an important object in the borrowers' lives.
Provided by all sorts of banks	Market share of mortgage loans in 2019 in Germany: see Figure 5
Tableau 2: Predictive Analytics Definition	
As part of data analytics	Following Gartner's Maturity Model for Analytics (see Figure 6)
Data mining	Data mining as a major component of predictive analytics models Predictive analytics applies the identified relationships and patterns in a predictive model
Machine learning	Can or cannot be included in predictive analytics models. If yes, the model can adapt itself to new data input. If not, the model needs input by humans to develop
Definition	*"Predictive analytics as the setup and use of a model which identifies data patterns via data mining algorithms and applies these patterns to enable a nearest possible prediction of future events"*
Types of predictive analytics models	Classification models or regression models (see Table 2) Often, clustering and association models described as other categories of data mining next to them. However, these are only descriptive and not predictive

Part 3: Introductory questions	Topic	Questions	Notes
Mortgage Financing	Market environment	*How do you assess the current market environment for financial institutions? Do you agree with the presented perspective or would you add something?*	
	Process	*Do you agree with the presented private mortgage financing process?*	
Predictive Analytics	Definition	*How would you define the term predictive analytics? Do you agree with the presented definition or would you emphasize or add something?*	

Part 4: 1st research question	Topic	Questions	Notes
What are the existing and potential fields of application of predictive analytics in the private mortgage financing process?			
	Introductory question/ inefficiencies	*Which aspects are the least efficient in the private mortgage financing process?*	
		Why is predictive analytics currently so relevant in the private mortgage financing business?	
	Fields of application	*At which points in the private mortgage financing process do you see potentials for the application of predictive analytics?*	
	Status quo	*Which fields of applications of predictive analytics in the private mortgage financing process are known to you (in your company)?*	
	Fields of application: short-term	*At which points in the private mortgage financing process do you see the most potentials to apply predictive analytics in the short-term?*	
	Fields of application: medium- to long-term	*At which points in the private mortgage financing process do you see the most potentials to apply predictive analytics medium- to long-term?*	

Part 5: 2nd research question	Topic	Questions	Notes
What type of predictive analytics is applied and what data sources are used?			
	Type	*Which type of predictive analytics is used in the described fields of application?*	
		Development: Open-source, purchased, in-house development	
		Underlying techniques: AI/ ML/ Data mining/ …	
		Method: Regressions/ neural networks/ …	
	Data sources	*Which data sources are used in the described fields of application?*	
	Data collection	*Which additional data could be generated in order to increase the predictability of the respective use cases?*	
	Requirements	*Which role does the legacy system and its architectural structure play for such cooperation?*	

Part 6: 3rd research question	Topic	Questions	Notes
What risks and challenges are attached to the implementation of PA			
	Risks/ Challenges	*Which risks and challenges are considered when applying predictive analytics?*	
	Data security	*With regard to data security, how do you assess the evaluability of the available data or the feasibility of data processing?*	
	FinTechs as competitors	*At which points in the private mortgage financing process do you consider the potential for FinTechs higher to create a more efficient process? Why?*	
		To what extent could banks anticipate and counteract this development?	
	FinTechs for cooperation	*Are cooperation between FinTechs and banks in place? (or are considered?)*	
	Potential of PA in comparison	*Compared to other measures or technologies, how do you assess the potential of predictive analytics in the private mortgage financing process?*	

Part 7: Conclusion	Notes
Are there further aspects you consider as relevant for this work project that we did not cover until now?	
Thanks for participating. Offer to share the results of the work project.	
Conclusion	

Appendix 10: Guided Expert Interview Framework (German)

Fragebogen Work Project	
Titel	Potential for Banks' Private Mortgage Financing Business through the Application of Predictive Analytics
Universität	Nova School of Business and Economics
Supervisor	Professor João Pereira
Autor	Dominik Laufs
Jahr	2020
Struktur	
Part 1	Generelle Einleitung
Part 2	Themen Einführung
Part 3	Einleitende Fragen
Part 4	1. Forschungsfrage
Part 5	2. Forschungsfrage
Part 6	3. Forschungfrage
Part 7	Abschluss
Forschungsfragen des Work Projects	
1. Forschungsfrage	Welche aktuellen und zukünftigen Anwendungsfelder von Predictive Analytics gibt es im privaten Immobilienfinanzierungsprozess?
2. Forschungsfrage	Welche Art von Predictive Analytics wird verwendet und welche Datenquellen werden herangezogen?
3. Forschungfrage	Welche Herausforderungen entstehen bei der Implementierung von Predictive Analytics?

Part 1: Generelle Einleitung	Themen	Details	Notizen
Vorstellung Interviewer	Vielen Dank!	*Bedanken für Verfügbarkeit zum Interview*	
	Persönlicher Hintergrund	*Dominik Laufs, International Master in Finance, Nova SBE, 3. Semester, Work Project*	
	Work Project: Titel	*"Potential for Banks' Private Mortgage Financing Business through the Application of Predictive Analytics"*	
	Work Project: Forschungsfragen	*Forschungsfragen vorstellen*	
	Work Project: Ziel	*Forschungslücke füllen (vgl. Forschungsfragen 1-3)*	
		Banken und FinTechs Anhaltspunkte über Best Practices der Praxis geben und aufzeigen, an welchen Stellschrauben gedreht werden kann	
	Funktion des Interviews	*Best Practices und fachliche Informationen aus der Praxis von führenden Instituten*	
Informationen zum Interview	Tonaufnahme des Gesprächs	*Zur späteren Paraphrasierung des Interviews. Wird Korrektoren zur Kontrolle zur Verfügung gestellt.*	
	Anonymisierung	*Anonymisierung des Instituts und des Interviewten aus Datenschutzgründen, es sei denn, explizite Nennung ist gewünscht*	
	Ablauf	*Themeneinführung. Fragen entlang der drei Forschungsfragen*	
Vorstellung des Interviewten	Name		
	Unternehmen		
	Geschäftsmodell des Unternehmens		
	Position/ Funktion im Unternehmen		
	Sonstiges		

Part 2: Themen Einführung	Themen	Details	
Immobilienfinanzierung	Relevanz	*Relevanz von privaten Immobilienfinanzierungen hervorheben und Umreißen des aktuellen Marktumfeldes*	
		(Vorlegen von Grafik & Bullets über Markt - siehe Tableau 1)	
	Prozessdefinition	*(Vorlegen des definierten Prozesses - siehe Figure 1)*	
Predictive Analytics	Definition	*Vorstellung der Definition von Predictive Analytics*	
		(Vorlegen der Hauptbullets über Predictive Analytics Definition & Abgrenzung - siehe Tableau 2)	

Tableau 1: Mortgage financing market environment	
High share in total credit volume	Private mortgage loan important is highly important for German banks, making out 42.4% of their credit volume in 2019
Comparatively constant interest rates	Interest rates of private mortgage loans are relatively constant around 1% over the last 4-5 years and did not decrease such as interest rates of other loan types
Anchor product	Private mortgage loan function as an anchor product to banks due to long maturities and the financing of an important object in the borrowers' lives.
Provided by all sorts of banks	Market share of mortgage loans in 2019 in Germany: siehe Figure 5

Tableau 2: Predictive Analytics Definition	
As part of data analytics	Following Gartner's Maturity Model for Analytics (siehe Figure 6)
Data mining	Data mining as a major component of predictive analytics models Predicitve analytics applies the identified relationships and patterns in a predictive model
Machine learning	Can or cannot be included in predictive analytics models. If yes, the model can adapt itself to new data input. If not, the model needs input by humans to develop
Definition	*"Predictive analytics as the setup and use of a model which identifies data patterns via data mining algorithms and applies these patterns to enable a nearest possible prediction of future events"*
Types of predictive analytics models	Classification models or regression models (siehe Table 2) Often, clustering and association models described as other categories of data mining next to them. However, these are only descriptive and not predictive

Part 3: Einleitende Fragen	Themen	Fragen	Notizen
Immobilienfinanzierung	Marktumfeld	*Wie schätzen Sie das aktuelle Marktumfeld für Finanzinstitute ein? Teilen Sie die eingeführte Sichtweise oder würden Sie etwas ergänzen?*	
	Prozess	*Teilen Sie die Sichtweise auf den privaten Immobilienfinanzierungsprozess?*	
Predictive Analytics	Definition	*Wie würden Sie den Begriff Predictive Analytics aus Praxissicht definieren? Teilen Sie die eingeführte Sichtweise oder würden Sie etwas ergänzen?*	

Part 4: 1. Forschungsfrage	Themen	Fragen	Notizen
Welche aktuellen und zukünftigen Anwendungsfelder von Predictive Analytics gibt es im privaten Immobilienfinanzierungsprozess?			
	Einleitende Frage/ Ineffizienzen	*Welche Punkte sind im Immobilienfinanzierungsprozess aktuell am ineffizientesten?* *Warum ist Predictive Analytics aktuell so relevant in der privaten Immobilienfinanzierung?*	
	Anwendungsfelder	*An welchen Punkten sehen Sie im privaten Immobilienfinanzerungsprozess Potentiale für die Anwendung von Predictive Analytics?*	
	Status Quo	*Welche konkreten Anwendungsfelder von Predictive Analytics gibt es im privaten Immobilienfinanzierungsprozess (in ihrem Unternehmen)?*	
	Anwendungsfelder: kurzfristig	*An welchen Punkten sehen Sie kurzfristig im privaten Immobilienfinanzierungsprozess die meisten Potentiale für die Anwendung von Predictive Analytics?*	
	Anwendungsfelder: mittel- bis langfristig	*An welchen Punkten sehen Sie mittel- bis langfristig im privaten Immobilienfinanzierungsprozess die meisten Potentiale für die Anwendung von Predictive Analytics?*	

Part 5: 2. Forschungsfrage	Themen	Fragen	Notizen
Welche Art von Predictive Analytics wird verwendet und welche Datenquellen werden herangezogen?			
	Art	*Welche Art von Predictive Analytics wird bei den von Ihnen beschriebenen Anwendungsfeldern verwendet?*	
		Fertigung: Open-Source, eingekauft, selbst entwickelt	
		Unterliegende Techniken: AI/ ML/ Data Mining, etc.	
		Analysmethoden: Regressionen, neuronale Netze, etc.	
	Datenquellen	*Welche Datenquellen werden in den von Ihnen beschriebenen Anwendungsfeldern verwendet?*	
	Datensammlung	*Welche zusätzlichen Daten können gewonnen/ ergänzt werden, um die Vorhersagewahrscheinlichkeit für die entsprechenden Use Cases zu erhöhen?*	
	Voraussetzungen	*Welche Rolle spielt aus Ihrer Sicht hierbei das Kernbanksystem und die architektonische Struktur für Kooperationen?*	

Part 6: 3. Forschungsfrage	Themen	Fragen	Notizen
Welche Herausforderungen entstehen bei der Implementierung von Predictive Analytics?			
	Risiken/ Herausforderungen	*Welche Risiken werden bei der Verwendung von Predictive Analytics genauer betrachtet?*	
	Datenschutz	*Wie schätzen Sie mit Blick auf Datenschutz die Auswertbarkeit der vorhandenen Daten, bzw. Data Processing ein?*	
	FinTechs als Konkurrenz	*An welchen Punkten sehen Sie besonders für FinTechs Potentiale den privaten Immobilienfinanzierungsprozess zukünftig effizienter zu gestalten? Weshalb?*	
		Inwiefern sehen Sie Möglichkeiten, dass Banken dieser Entwicklung zuvorkommen können?	
	FinTechs zur Kooperation	*Bestehen Kooperationen zwischen FinTechs und Banken? (oder sind solche in Überlegung?)*	
	Potential von PA im Vergleich	*Im Vergleich zu anderen Maßnahmen oder Technologien, wie schätzen Sie das Potenzial von Predictive Analytics im Immobilienfinanzierungsprozess ein?*	

Part 7: Abschluss	Notizen
Gibt es weitere Aspekte, die Sie für das Work Project als relevant erachten und bisher nicht angesprochen wurden?	
Bedanken und Angebot die Ergebnisse zuzusenden	
Verabschiedung	

Appendix 11: Paraphrased Interview Expert 1

11. August 2020, 1h 24min, translated from German to English

Part 1-3: Introductions

I: Do you agree with the presented perspective on the current market environment for financial institutions or would you add something?

Expert 1: Absolutely, I'd like to emphasize that the interest rates for mortgage financings have been constant over about the last four years and did a significant decrease the years before that. That's especially important if you consider that this is the interest rate charged by banks and does not represent the margin which banks receive in the end. Over the last years, the pressure on the banks' profit margins has increased due to the lower total interest rates. The increase in competition, especially caused by FinTechs, has increased the margin pressure further.

I also agree that we deal with an anchor product which has a relatively high volume of the total credit volume. Nevertheless, we cannot say that it is a product which has contributed to the profit of banks in a constant way. The product has developed as a challenging product leading to an increase in digitalization and the question of how to approach the customers. All this in order to keep their market share and not be replace by the fierce competition.

I: Do you agree with the presented private mortgage financing process?

Expert 1: I'd like to point out that *use of funds* can be labelled a sub-process of *Disbursement*. This is due to the fact that initial loan payments are often made after the structural work of the building is done and further loan payments are conducted step after step and not at once – in case of a construction financing. In case of a real-estate purchase, the loan payout is conducted via notary escrow accounts in order ensure that the money is correctly used.

Furthermore, I'd like to emphasize that we definitely do not have a linear sequence of events. Especially the *related needs* and the *refinancing* steps run parallelly to the process steps of the *bank* *process*. Thus, refinancing can also be considered from a bank's perspective and allocated to the *portfolio* step of the bank process.

Additionally, it is important to mention that while the customer journey prior to the bank process may take several months or even years, the actual bank process is normally processed over the course of a few weeks.

I: Do you agree with the presented definition of the term predictive analytics or would you emphasize or add something?

Expert 1: This definition is fine by me.

Part 4-6: Research Questions

I: Which aspects are the least efficient in the private mortgage financing process?

Expert 1: While the advice process step is currently very well developed in German banks, the collateral management is currently the least efficient process step. This is due to the fact that it involves a notary and the registration in the land register among others. Here, documents are mostly processed via paper and are not digitalized. Nevertheless, this is only a question of efficient and digitalized processes and cannot be solved with predictive analytics alone.

I: At which points in the private mortgage financing process do you see potentials for the application for predictive analytics?

Expert 1: We especially see fields of application for predictive analytics when the question of creditworthiness of the customer arises. Predictive analytics helps to answer the question whether a customer will be able to pay back the loan at any point over the whole course of the loan period and not only as of the date of the loan payout. Foreseeable events may or may not change the ability to pay back, such as the omission of one of two salaries from a couple due the arrival of a child or the intertemporal omission of the salary due to limited working contracts.

Furthermore, there is potential for the application of Predictive Analytics in the credit-scoring models of banks. Predictive analytics can advance the current models which base on historical data and try to make best guesses about future events. The current process is highly standardized and needs a regulatory approval. While the inclusion of traditional historical data may not lead to a positive score, the inclusion of further data may enrich the data pool, such as for example, social media information. Thus, a customer may receive a bad score if the bank only looks at the current salary. However, if further available data is analyzed and a positive future is projected, the customer may now receive a better score and the bank may take the risk and still give out the loan. The German FinTech Kreditech tried to advance such a technology in the German market. However, German customers do not feel comfortable with their banks analyzing such data which are out of their ordinary business. Kreditech now faces insolvency. This technology especially sees application in emerging and Asian countries, where customers seem to care less about data privacy. There, the final interest rate of the loan even depends on the amount of provided data. The more access to data points is given, the more secure the predictive analytics model can predict futures states and thus, a lower interest rate may be provided.

However, this is also a special feature in Germany where we initially have fixed interest rates. In other countries, mostly variable interest rates are used. Such rates do not only adapt to the interest rate curve but also to the individual creditworthiness of the customers. Here, banks are inclined to lower the interest rate in case of a good customer behavior, because otherwise, the customer might switch to another bank with lower interest rates as they are able to cancel their variable interest rates loans every three months – of course depending on the bank, the country and the individual contracts. Consequently, a bank may also increase the interest rates due to the short interest adaption intervals in order to move the customer to another bank and take risk of their books.

Regarding the application of Predictive Analytics in the Workout Management, there is the chance to apply it. However, I currently do not see potential for the German market. This is especially due to the reason that banks almost do not have problems with loan defaults. In case a loan still defaults, the provided collateral is mostly able to cover the respective losses. Thus, we have losses of 0.2% to 0.5% in normal market conditions. Nevertheless, Predictive Analytics may be used in Workout Management for an active portfolio management by identifying non-performing loans early on and to take preventive actions. Referring to my point before, a segregation between such non-performing loans and performing loans could be conducted and measures be taken to reduce the risk of the total portfolio by adjusting interest rates. Such measures being the adjustment of interest rates or the respective trade of credit derivates.

In Germany, such a portfolio management could be feasible after the maturity of the fixed interest period which is around 12.5 years in average in Germany. However, with the option for customers and banks to cancel the contract after 10 years due to strong consumer rights.

What's more is a potential in the collateral management of banks. The central question is how the value of the object is developing in future. On the one hand, with a higher collateralization ratio, the bank's risk is lower. On the other hand, the object is an investment object which might be collateralized for a second or third time. Having a range of older objects in their portfolio, banks can realize that certain replacement or repair investments might be necessary via applying predictive analytics. Consequently, a bank could think forward, use the collateralized object and creditworthiness of the existing customer and offer him an underwritten installment credit. This has the advantage that no new information needs to be gathered and the existing collateral can be used (Tap into equity). This also refers to the point that a mortgage financing is an anchor product and yields the possibility for future cross-selling with a customer.

Compared to that, there are not many use cases in the process steps contract, disbursement or portfolio.

Most fields of application are currently in the process step need. First of all, it's a challenge to receive a full pipeline of customer engagements and to identify suitable customers at the right point in time in order to get into advisory meetings or make them suitable offers also referred to as the next-best-offer. This is due to the emerging ecosystems and platform business over the last years. Many queries are not initiated at the customers' house bank but rather online where customers get a good overview about everything that is necessary, ranging from YouTube Videos to comparison websites. Consequently, banks have the pressure to get new customers and thus, ask themselves how they can realize that a customer might be interested in a mortgage financing in the near future and how advanced is the customer in his/her customer journey. It is a bank's goal to position themselves in the customer journey as early as possible by identifying suitable customer patterns in the data, as the customer will parallelly start checking other providers once he has the need for a mortgage financing. This field of application is already in use and banks are exploring it further. While small- to medium-sized banks have to rely on their centralized data centers to advance these approaches, larger companies can push this development more quickly with departments allocated for this purpose.

I: At which points in the private mortgage financing process (of your company) do you see the most potentials to apply predictive analytics in the short-term?

Expert 1: I especially see potentials at the customer interaction. Referring to your process steps, this is somewhere in between the initial customer journey steps and the beginning of the bank process. For this field of application, various banks are making considerations and invest already.

Furthermore, we often talk about an ecosystem around mortgage financing. There is huge potential in cross-selling further products after financing a mortgage. Banks are not only interested in providing the mortgage financing itself but also in selling the mortgage as a brokerage product.

This, however, is not related to predictive analytics itself.

I: At which points in the private mortgage financing process (of you company) do you see the most potentials to apply predictive analytics medium- to long-term?

Expert 1: In the medium-term, I especially see the application of predictive analytics models in the credit-scoring process of banks. Especially due to the fact that such models might take time to be validated before they comply with all rules and regulations and can be applied in practice in the end. Especially for larger banks, such model change processes need the validation of regulators.

In the long-term, we see topics such as the workout management which I previously already mentioned.

I: At which points in the private mortgage financing process do you consider the potential for FinTechs higher to create a more efficient process? Why?

Expert 1: FinTechs have the great advantage that they can build their ideas from scratch from a greenfield site. Especially in the mortgage financing area, we have to divide FinTechs into two categories. On the one hand side, we’ve got FinTechs which work at the customer intersection points (e.g., interhyp, Check24). Such data collecting companies have the advantage in their business model that they can focus with their IT-infrastructure on pushing predictive analytics forwards, gather and process customer data in a very fast way without carrying big regulatory burdens such as established banks. We assume that around 50-60% of mortgage financings start via mortgage financing platforms which makes the process more expensive for banks as they have to pay margins to such platform providers. Consequently, the customer intersection is provided by such online comparison providers with the bank only being involved in the process by offering their product in a banking as a service style. This disrupts the connection between banks and customers and removes the function of the anchor product as customers are not bound as a whole but only via a single product.

On the other hand, we've got FinTechs which do not try to acquire banks' customers, but rather support banks in their processes. However, this often does not necessarily contain predictive analytics.

I: To what extent could banks anticipate and counteract this development?

Expert 1: Building up cooperation with such platforms is a more feasible solution than fighting such providers.

Partly, single bank groups aim to build up such ecosystems or platforms on their own. E.g., ING owns interhyp and makes profit with each mortgage financing which is processed via this platform. Furthermore, we see regional cooperation of banks which also try to build up an own ecosystem. Such an ecosystem builds up on further cooperation with regional craftsmen to promote the concept "from the region for the region". This must not necessarily yield the cheapest price for the customer, however, the after all cheapest price is not the ultimate criteria for all customers. Also, other factors, such as regionality or sustainability, play a role.

For banks, it is not only necessary to identify customers that potentially need a mortgage financing, but it becomes more important to identify customers that fit into the business model of the bank and are more likely to take a mortgage financing at the respective bank. E.g., because they require certain standards and do not only look for the cheapest price. For such customers, it pays off to invest into the beginning of a relationship in order to stay in a longer-term relationship.

I: Are cooperation between FinTechs and banks in place? (or are considered?)

Expert 1: We know FinTechs which offer technologies for cooperation, however, not including predictive analytics. Up to now, I am not aware of banks cooperating with FinTechs or other companies that offer predictive analytics solutions to integrate it into their processes. This is also for banks to maintain their reputation as such cooperation for the processing of sensible data might not be well viewed at in the German market.

I: Which role does the legacy system and its architectural structure play for such cooperation?

Expert 1: Banks often still operate with old legacy systems. Thus, they are often not able to offer open-source architectures for other companies to dock on to. However, this will change in future as banks are trying to open up their systems and make it easier for others to connect to their systems. This also happens due to regulatory reasons such as the introduction of PSDII which obliges banks to make certain data points accessible to outsiders if the customer agrees to it. In general, banks rather refuse to give access to their customers' data, even if they agree to it, because it's kind of their gold treasure which is the basis for their business.

I: Which type of predictive analytics is used in the respective use cases?

Expert 1: Most predictive analytics tools are based on DM methods. However, with regard to artificial intelligence, most banks are still in the developing process as such techniques require extensive computing power and it is very challenging to integrate it into the existing IT-infrastructure. Thus, it mostly is not yet ready to be applied in practice.

For the technology itself, IBM is well renown to offer such technology solutions and this company has a trust-advantage over smaller-sized or younger corporations. Also, solutions from Microsoft, SAP and Salesforce are applied occasionally, but also Amazon's solutions are more and more considered. Nevertheless, I am not 100% sure about the inclusion of predictive analytics itself in the cooperation of such companies.

I: Which risks and challenges are considered when applying predictive analytics?

Expert 1: Currently, we still lack regulations with regard to such technologies which still challenges the BaFin. E.g., data security is very challenging, especially when banks rely on services of American companies, such as Amazon, which do not have their servers in Germany but e.g., in Ireland and may or may not comply with European data security standards due to their proximity to America and American institutions.

For many years, banks are aware of the fact that their business model will need less employees especially in the operations departments. However, this is not necessarily due to such predictive analytics tools which advance the processes and also require employees focusing on the development of such technologies. We expect that workplaces decrease in different departments, such as operations, but will increase in other increasingly relevant departments, such as the IT-department.

Another challenge is that banks will be required to explain their applied algorithms and the basis of their decisions at any time to regulatory authorities. This requires that banks themselves (and not their technology/ code providers) need to understand, analyze and explain the code lines of their models. The exact requirements are not yet predictable. Nevertheless, new jobs will be created for this as a bank will have to do a similar job as IT companies.

Furthermore, the costs related to the development of predictive analytics tools and over which time-horizon its application pays off are further relevant questions. For this, it might need banks which pioneer with such technologies before other banks follow and the technology sees a wide application.

I: With regard to data security, how do you assess the evaluability of the available data or the feasibility of data processing?

Expert 1: Banks are able to evaluate the available data of their customers. However, only when they initially agreed upon the analysis via computational evaluations due to DSGVO regulations. If customers did not agree to such regulations, e.g., because DSGVO didn’t exist yet when the customers got their credit or in case they did not agree to the computational evaluation of their data, banks are not allowed to computationally evaluate their data but only manually by customer advisors looking at it.

Regarding the evaluation of further information of the customer which he provides e.g., via cookies from web-surfing, social media entries or in general other data which exist outside the ordinary banking business, German customers are rather restrained and do not wish their bank to pursue such procedures. Thus, I am not aware of German banks following such an approach although with regard to regulators, it would be allowed. As it is especially used in emerging countries and other countries in general, the attitude of the German market may change as well at some point and the inclusion of such external data may be considered.

I: Which additional data could be generated in order to increase the predictability of the respective use cases?

Expert 1: Even without utilizing public personal data of their customers, due to the aforementioned not-readiness of German customers, banks can include other freely available data which are not personalized. This especially includes micromarketing with static information about certain districts, streets, etc. which cannot be traced back to a single person. This static information can be combined with the information available about customers in the bank's archive and thus, lead to new insights or marketing opportunities – even without relying on freely available personalized online-data, such as social media.

I: Compared to other measures or technologies, how do you assess the potential of predictive analytics in the private mortgage financing process?

Expert 1: Another hot topic is the digitalization in general, including the digitalization of data and processes. Furthermore, the modernization of banks' legacy systems and open-source architectures are currently very relevant. Also, how different streams of data can be related and processed with each other.

Appendix 12: Paraphrased Interview Expert 2

14. August 2020, 56min, translated from German to English

Part 1-3: Introductions

I: Do you agree with the presented perspective on the current market environment for financial institutions, the presented private mortgage financing process and the presented definition of the term predictive analytics or would you add something?

Expert 2: Yes, this is fine by me.

Part 4-6: Research Questions

I: At which points in the private mortgage financing process do you see potentials for the application for predictive analytics?

Expert 2: One field of application is to check how much a customer can afford and to use targeted advertising to promote offers to take loans of a certain amount. This is based on several information about the customer, including demographic and transactional data.

Theoretically, it would also be possible to predict the need for a mortgage based on a changed consumer behavior. E.g., having a baby or a growing family, is one of the most common reasons to expand one's property in size. Thus, if the buying pattern changes to buying at certain stores, e.g., baby stuff in baby stores, it would be an indication for family growth and the possibility of interest in a mortgage financing.

Unfortunately, such types of products are not yet offered by FinTechs or Challenger Banks in the private mortgage financing process. For us, this is not covered by our business model and for others, I suppose, it's due to the complexity and amount of mortgage financings.

In case the customer gives a bank or a comparison website access to all accounts and there is sufficient data available, a machine based on predictive analytics will be able to define the most suitable mortgage financing product for the customer. In such an extreme case, no bank and no advisors would be required any more – only someone providing the capital in the end. Currently, brokers try to collect as many information about customers as they can to offer them the best possible product based on the collected information. Their product fitting is based on this information and their experience, not on big data technologies.

For our company, we want to improve our systems to such a point that it can best possibly predict the most suitable products. Nevertheless, we believe in the human in this process, a broker respectively. Also due to the reason that such brokers are currently our customers which use our platform for the end customers.

The next field of application is in the review step. Different banks create their own rating systems to predict the default probability of customers. Also, the Schufa, a credit rating agency, defines such scores based on predictive analytics and some banks use their score as one data-point within their own rating system.

I: At which points in the private mortgage financing process do you consider the potential for FinTechs higher to create a more efficient process? Why?

Expert 2: The mortgage financing sector in Germany is kind of in a golden cage without many companies challenging traditional banks. I am convinced that more competition, including FinTechs and comparison platforms, will revitalize this business for customers.

I: Are cooperation between FinTechs and banks in place? (or are considered?)

Expert 2: We are partnering up with banks as they are offering their product on our platform. They approach us and we include their products on the brokerage platform. We stay neutral in this process and do not feature any kind of products of any kind of bank.

I: Which role does the legacy system and its architectural structure play for such cooperation?

Expert 2: We have an open-API structure that institutions can dock on to for free and we earn money at the point when you complete with our technology. Consequently, we ourselves do not have problems with old legacy systems of banks as our technology is not integrated in their IT-infrastructure.

I: Which type of predictive analytics is used in the respective use cases?

Expert 2: Our software is self-developed. We include machine learning to optimize our platform and solve complex financing questions.

I: Which risks and challenges are considered when applying predictive analytics?

Expert 2: It's our goal to take as much workload off the broker as possible to make the process more efficient for him, based on modern technologies. Thus, we do not see the risk of replacing humans in this business.

I: Which additional data could be generated in order to increase the predictability of the respective use cases?

Expert 2: For us as a FinTech which strives to create the least complex private mortgage financing process for customers and brokers, we would like to see a very lean overall process in mortgage financing. Currently, every bank conducts a different process and thus, a unified process does not exist which increases the complexity for us, brokers and after all, the end-consumers. At some banks you easily get a loan, while at others, you will not get one at all. We, as a FinTech, try to reduce this complexity, but actually it should not exist at all.

I: Compared to other measures or technologies, how do you assess the potential of predictive analytics in the private mortgage financing process?

Expert 2: I am convinced that technology alone is not the one solution for the problem we have in the German banking environment. Banks need to start moving and adapting before such progressive technologies can be applied.

Appendix 13: Paraphrased Interview Expert 3

18. August 2020, 1h, translated from German to English

Part 1-3: Introductions

I: Do you agree with the presented perspective on the current market environment for financial institutions or would you add something?

Expert 3: Yes, this is fine by me.

I: Do you agree with the presented private mortgage financing process?

Expert 3: I agree with you. It is important to say that several process steps are overlapping, especially between the bank process and the customer journey itself.

I: Do you agree with the presented definition of the term predictive analytics or would you emphasize or add something?

Expert 3: I'd like to emphasize that Predictive Analytics must not necessarily only base on historical data, but also on current, unstructured data. E.g., new information which are not necessarily related to the topic and are either structured or unstructured. With such data, machine learning tools could identify a correlation, or the inclusion of new parameters might lead to new insights. While historical data include data points such as the development of credit loans, vacancies of houses, etc., non-historical data might include social network data and other forms of unstructured data which are not necessarily obvious to include in such analyses.

Part 4-6: Research Questions

I: Which aspects are the least efficient in the private mortgage financing process?

Expert 3: Looking at the process defined by you, the market is relatively efficient with regard to receiving different offers from different banks through the existence of comparison websites. Compared to that, the transparency throughout the bank process is not very good. All banks use different scoring systems, and the end-consumer does not know what variables led to the disapproval of a loan request or to certain conditions of a loan. Consequently, the end-consumer does not know how to improve to get his loan approved or to receive lower rates, because he could be classified in another risk-group.

I: At which points in the private mortgage financing process do you see potentials for the application for predictive analytics?

Expert 3: One field of application is predicting the occurrence of certain events in a customer's life. E.g., predicting the need of a customer for a mortgage finance, based on e.g., marriage, children, etc., in order to increase the probability of a positive completion. It can also answer other questions, such as how good a loan will be repaid or how often will a refinancing be required. This would be in the area of credit-scoring and creditworthiness, but also in the area of collateral management, use of funds or refinancing.

Currently, this is a very static process, where a customer asks for a credit and receives an offer based on the provided information. Including predictive analytics in this process step would analyze the behavior of a similar customers and include these insights in the product offering.

Furthermore, another field of application is to predict the need for mortgage financing. Currently, this e.g., happens regularly in standard meetings between customers and bank-advisors or when the customer proactively approaches a bank and asks for a credit. Predictive analytics is able to help the bank identify the customers' needs and show the next-best-product. Eventually, a bank could even point out that he/she is eligible to receive a mortgage financing at his current status or if he/she would have a little more capital or save a certain amount of money, without the customer initially thinking about it.

In collateral management, predictive analytics could enrich the available data to show further collateral options. Consequently, a bank could offer to use other data-points as collateral for the loan before they would start with margin calls, etc. This also creates the opportunity to cross-sell further products, such as giving out loans based on the existing collateral.

Our credit-engine is very variable in its offerings, with different kinds of layers which come on top of the one for the customers. We, as a service-provider, have a pre-configuration how certain processes might look like and offer workflows and settlements for these processes. In addition, we developed new products based on analytic AI and data science which can simply be adopted to the customers' needs as a ready-to-go platform which does not require developing skills to use. Consequently, they do not have to develop their own data platform and not care about the selection of different tools, the provision of computing power, the hiring of qualified employees, etc. This tool can be fully integrated in our provided systems, is available on mobile and complies with all security standards. It also includes a distributed learning which means that you can benefit from the insights of other banks using the tool.

This tool can be applied for a variety of use cases as required by the customers. The tool will always be adapted to their needs. However, it is important that the banks keep their competitive advantage and only use such tools as an enriching factor within their processes. This tool can also be applied for the use cases mentioned before in the private mortgage financing process, such as the prediction of events happening in the customers' lives based on the analysis of a peer-group and the possibility to predict the likeliness of a default.

I am convinced that there will come up future fields of application that we currently do not even think about.

I: At which points in the private mortgage financing process do you see the most potentials to apply predictive analytics in the short-, medium- and long-term?

Expert 3: We always evaluate the use cases with our customers. Either we approach them and tell them what's possible or they approach us with a new strategy and ask us how we think we might apply e.g., predictive analytics to enhance this process.

The use cases I mentioned before are current, and I am not aware of other predictive analytics use cases in the private mortgage financing process we are currently investigating.

I: At which points in the private mortgage financing process do you consider the potential for FinTechs higher to create a more efficient process? Why?

Expert 3: We define ourselves as a FinTech due to our product offerings.

Either FinTechs are crowd-funding platforms or companies offering comparisons of different loan providers. What's always going to be missing for such companies is a balance sheet. Either you do this via crowdfunding, or you are a bank. However, for mortgage financing, there is no crowdfunding company yet.

In general, the advantage of a FinTech is that it's working with apps and technologies and can quickly test them in the market. Nevertheless, in the end, it needs financial power to be successful in the market. Consequently, it rather ends in cooperation than in a cannibalization.

I: Which role does the legacy system and its architectural structure play for such cooperation?

Expert 3: When there is a working platform and the field of application is chosen well, then it should not be a problem to integrate new solutions here. Our advantage is, that our products are all coming from us – we offer the IT-infrastructure and as part of this, we offer the possibility to integrate predictive analytics tool in the infrastructure.

I: Which type of predictive analytics is used in the respective use cases?

Expert 3: The software we provide is a mixture of self-development (based on open source) and other providers, such as IBM Watson, etc. which get adapted to the respective use case.

Our predictive analytics tools partly include machine learning.

I: With regard to data security, how do you assess the evaluability of the available data or the feasibility of data processing?

Expert 3: Banks themselves are in charge to make sure that data security standards are fulfilled. We then implement it and only work with the customer data the bank gives our systems access to.

I: Compared to other measures or technologies, how do you assess the potential of predictive analytics in the private mortgage financing process?

Expert 3: We currently pursue three main topics. The first one is our platform which provides big data, machine learning or predictive analytics tools for many departments, including the private mortgage financing process. Furthermore, we check, how we can improve ways to contact customers in an efficient way and to have a digital exchange with them. In addition, we look for new ways how to manage the wealth of customers. All this in addition to our banking offering.

All this is conducted with a very agile approach in close exchange with our customers.

We also look at the opportunities Blockchain plays for our infrastructures. E.g., you are able to trade cryptocurrencies and manage wallets on our platform.

Appendix 14: Paraphrased Interview Expert 4

18. August 2020, 1h 25min, translated from German to English

Part 1-3: Introductions

I: Do you agree with the presented perspective on the current market environment for financial institutions or would you add something?

Expert 4: Next to being an anchor product, mortgage financing product is the only product with which banks can still make money. Of course, also next to other products based on provisions. All other products do not have relevant margins for German banks anymore.

I: Do you agree with the presented private mortgage financing process?

Expert 4: Yes.

I: Do you agree with the presented definition of the term predictive analytics or would you emphasize or add something?

Expert 4: To clarify this even further, we also often differentiate between supervised and unsupervised learning. The more untransparent a method is in the end for an expert, the less likely it is that it will actually be applied. Consequently, our customers prefer tree-like decision making to be able to follow what's happening in the model.

Part 4-6: Research Questions

I: Which aspects are the least efficient in the private mortgage financing process?

Expert 4: This is the creditworthiness check within the bank process. It needs various documents which are required. As of today, the collection, validation and processing of these documents needs the biggest manual effort. While the private mortgage financing process is very standardized, it unfortunately is everything else but automated. Including the customer journey as well, a best possible identification of current or new customers for a mortgage financing is crucial.

Identifying and approaching these customers is often still very inefficient and not necessarily a targeted, efficient process.

I: At which points in the private mortgage financing process do you see potentials for the application for predictive analytics?

Expert 4: Identifying customers who could potentially be interested in a mortgage financing is one field of application. We call this initial point the ‘magic moment’ in the process. Predictive analytics helps creating these magic moments or identifying interested customers for a product campaign. While for current customers this is relatively easy based on the existent data, for new customers this is rather difficult as you do not know much about them.

Checking the creditworthiness of customers is another main field of application. Machine learning can even be applied just before this step, namely when the provided documents have to be sorted and checked. Also, customers can send pictures and other characteristics about their homes to the bank and based on machine learning and available data about this and other houses, they can say how much it could be worth. While this is not based on predictive analytics itself, it enriches the available data about the customer and consequently, with predictive analytics tools, further needs of these customers can be identified.

Furthermore, predictive analytics can be applied in the monitoring of outstanding loans. It can answer the question of what the likeliness of a credit defaulting is. Following that, the bank is prepared for this case and can get active early on. This enables an early approach to the customer and a close cooperation between bank and customer. Here, we can differ between an economical and regulatory point of view. While the bank can use such a tool to take action early on and approach the customer, it yet cannot be used from a regulatory perspective, e.g., in order to determine early on by how much capital the credit should be backed (based on the output of such predictive analytics tools).

The creditworthiness-check includes the initial information about the customer, the object and the required financing and checks whether the customer has a debt capital capability and that he is not indebted. This is included in the final rating. For regulatory purposes, such a rating requires traceable rules. Compared to that, for economical purposes, the bank is free to use other tools, e.g., based on predictive analytics tools which are not yet approved by regulatory authorities. In such a case, a customer could have medium rating based on the regulatory-approved rating system, but based on the predictive analytics rating, he/she could have a good rating. Consequently, a bank could still favor this customer above another one with a worse regulatory-approved rating as it is up to them which customer they want to work with or not. This customer still needs to match the minimum requirements of the regulatory-approved model and the amount of capital for the backing of the credit to this customer will be determined by this model. To sum up, you can choose freely which customer you serve and what prices you offer based on your predictive analytics model, while still complying with the regulatory-approved model.

I am convinced that there are many further, smaller fields of application for predictive analytics in the private mortgage financing process if you look more closely. However, the largest and most current ones are the ones I just mentioned.

I: At which points in the private mortgage financing process do you see the most potentials to apply predictive analytics in the short-, medium- and long-term?

Expert 4: Currently, and in the very short-term, banks try to make product campaigns and next-best-offers in a very targeted way with their available budget. This helps banks to focus on customers that are actually interested in a mortgage financing, not annoyed by such targeted advertisements and have a higher probability to complete such a financing with their bank. For current customers, this can already be conducted with the available data of banks. It’s rather rare that they also include external data for this which is due to the complexity of including such data.

There are different levels of data that can be included. Firstly, it can be done by including basic customer data and account data. Secondly, transaction data of the customers' accounts can be utilized. However, it is questionable whether the bank is allowed to use them for data security reasons. This needs to be initially approved by the customers. Another step would be the inclusion of external data, such as aggregated information about the place of residence, its environment and the expected future development of these. Other external data include the purchase of customer data by companies which collect information about people in terms of how many mobile phone contracts are connected to a house, what's the DSL line and speed there or what's the traffic on that road, etc. This more tailored information aims to categorize households into the 'Sinus-Milieus' and are sold to advertising companies and is also available for banks. So, this data is not collected on an individual level, but on a household level. This data can be connected with the information available about individual customer data if they are a current customer of the bank. Including social networks activities would again be on another level. However, there is no current application of that in the German banking market that I am aware of. The challenge that arises here for banks is that they will have to somehow offer customers an advantage or a gimmick in order for them to be allowed to use and process these data points. And banks do not have something that customers are really looking for.

Concerning the risk-assessment of loans, predictive analytics is not yet applied. This is because of regulatory reasons, responsible people mistrusting such new analysis methods and traditional methods still working fine. Such risk experts also refer to the regulatory necessity to conduct the risk-assessment with such existing tools.

Applying predictive analytics in the monitoring of loans to predict their current probability of default finds its first applications. This topic can be developed in two directions: it can either be used to predict real defaults of loans or to predict cases where fraud takes place. Consequently, banks might have a well working traditional model for credit defaults while they don't have such a tool for frauds and might thus, develop such a predictive analytics model with real time data for potential fraud cases only.

I am convinced that these topics and especially the regulatory topics will play a significant role. However, rather in the longer-term. Regulatory authorities are already in exchange with banks and ask them to what extent they are active with these topics so that they can consider these topics in their future regulatory developments. The regulatory authorities are aware of the fact that they have to engage with this topic and cannot prohibit these technological advancements in general.

I: At which points in the private mortgage financing process do you consider the potential for FinTechs higher to create a more efficient process? Why?

Expert 4: FinTechs do not have the rat-tail that large companies have. They are freer in their thinking and in their doing. Looking at traditional banks, they take years to conduct prototypes and bring it to the market. Compared to that, FinTechs' Go-to-market is much shorter. Also, the questions these corporations are posing vary fundamentally. While startups might ask data security questions in the very end and think customer-centric from the very beginning, traditional banks often think differently here.

Most FinTechs do not offer an all-around solution for products or services but rather deal with a very specific topic. Consequently, banks can get attacked in single products or services and also lose money here. However, the FinTechs themselves are not that big that they could replace them totally. This, however, looks different for large tech giants, such as Alphabet, Apple or Amazon.

I: To what extent could banks anticipate and counteract this development?

Expert 4: Up to now, everything FinTechs are doing is copyable. Thus, banks look at successful FinTechs and can copy and integrate their solutions. The danger is that this often comes along with a cultural change in banks.

Nevertheless, it will not be able to make a large, traditional bank as quick-witted as FinTechs. Consequently, banks have to double check what they can copy from FinTechs or whether they might think about acquiring a FinTech after all. For this, also the IT-infrastructure of the traditional bank needs to be able to cope with these new services or products. Many legacy systems of large banks are currently not yet able to do this and firstly, need to be updated or renovated.

With regard to the private mortgage financing market, there are several comparison websites in the market who enter this market and lower the banks' margins. In the end, consumers still make their credit with a bank, however, a small margin will be paid to these comparison websites for initiating the loan. Nevertheless, such institutions might also integrate vertically. E.g., one comparison website now has banking license. We do not know yet what exactly they want to do with this, whether they only want to do payment transactions, but it also enables them to give out standardized loans in future. To sum it up, such FinTechs have the potential to enter the market at one point and then expand in further market points if successful.

I: Are cooperation between FinTechs and banks in place? (or are considered?)

Expert 4: Yes, they are existing. E.g., the house valuation app that I mentioned earlier is a FinTech cooperating with a group of several banks and is integrated in their processes. Even, if this use case does not necessarily use predictive analytics tools, but rather machine learning alone, this product still creates data points which have the potential to be used by predictive analytics tools for other use cases.

There is also the possibility that FinTechs provide certain process steps for banks to integrate in their processes. Banks have to think about what exactly their USP is. Most banks won't say that the creditworthiness-check is their USP which differs them from other banks. Consequently, a FinTech cooperation might create such a USP or lets a bank focus on other parts of their processes.

Partly, this is already the case if we look at credit rating agencies which provide scores about the customers which are used by banks. It might not be the only point which decides about the approval or disapproval, but it certainly plays a role in the banks' processes.

Nevertheless, regulatory authorities will always require banks to explain their models and why it made certain decisions.

I: Which role does the legacy system and its architectural structure play for such cooperation?

Expert 4: Old legacy systems are a blocker if the system is not approachable via APIs or web-services. Without this, banks are not able to react agile. This is in general a wish that comes along with data analytics, namely that you want to be able to quick and agile in analyzing available data and also quickly remove and exchange models that are not working as planned. For this agility, you need systems which are easily integrated, and you need high-quality data. Consequently, technical conditions and the architectural structures are very important enablers.

I: Which type of predictive analytics is used in the respective use cases?

Expert 4: Normally, there is no 100% self-development of such use cases where organizations select a programming language and program the whole model by themselves. Especially, because there is a variety of offered open-source products and products subject to charges on the market. These solutions may have different preconfigured models and a substantially different user guidance. Thus, a different technical know-how may be required to use certain products. Nevertheless, there is a variety of modules which can be applied here, ranging from IBM, H2O or university spin-offs.

Now, there might be a big difference between smaller banks and larger banks. On the one hand, smaller banks rather open up a small 1-2-person department under control of corporate steering or corporate controlling. They conduct such data analytics for smaller use cases with pilots being created every few weeks and may have support from companies such as IBM or consultancies such as ours. On the other hand, medium- to larger-sized banks may have larger departments which often cooperate with larger tech-companies such as IBM in order to buy-in more know-how and maintenance.

From my experience, banks did not make good experiences with tools aiming to achieve a specific goal as the available data often does not fit 100% and the tool is not individualized enough. This is especially the case when a solution, as for example next-best-offer, comes from another industry. E.g., a next-best-offer tool from libraries cannot be used in mortgage financing as the characteristics of the products, such as amount of occurrence, fundamentally differs.

In practice, unsupervised models are not used, because users want to understand what exactly the input-question is and what can be expected as an output.

The thing of e.g., neural networks or similar is that they are not able to necessarily explain the cause-effect relationship. Consequently, it is very limited in its application. Banks are currently of the opinion that they still want to explain the factors why a decision was made, so that it can also be approved by regulatory authorities which requires them to prove how a decision was made. But such an explainable model is also important in order to convince the employees to use the tool in the end, because if he or she does not understand the cause-effect relationship anymore, he or she might rather refrain from using such a tool.

Thus, mainly classification and regression models are currently used. For this, very often decision trees or random forests trees are applied, which – if wanted – even enables users to graphically display the decision-making process.

Furthermore, k-means models are often applied for clustering. There are also tests with neural networks, but often only in order to make comparisons with other models.

I: Which risks and challenges are considered when applying predictive analytics?

Expert 4: A very big risk is the disadvantage of some models being a black box and not letting the users realize cause-effect relationships.

Furthermore, the quality of models needs to be reliable and not output different results every other day.

In addition, regulation topics are also very critical. A bank would rather pick topics which are not yet regulated, or they would differ between regulatory and economical models to be on the safe side.

Concerning workplaces, there is no major risk. Banks are aware of the fact that they will shrink the number of customer advisors as less customer visit banks. However, this is not due to such predictive analytics tools.

Another risk is to lose the employees who are in charge of the predictive analytics models. They might be the ones being able to explain the functioning of the models and such employees leaving would mean that there is the risk for the bank of not being able any more to explain their models to regulatory authorities. Thus, employees should be included in the development process and also salespeople's opinions during pilot testing phases should be processed.

I: With regard to data security, how do you assess the evaluability of the available data or the feasibility of data processing?

Expert 4: Data security is a very big topic. The only solution for this is to convince the customer of an added value that comes along with the processing of their data in order to get their confirmation to process their data. They won't just do this in order to get more advertising. Such an added value could be e.g., in the area of multi-banking where different accounts of customers are merged and clustered in a useful way. This might also include special offers for certain product categories.

I: Which additional data could be generated in order to increase the predictability of the respective use cases?

Expert 4: One topic is the transaction analysis, for which banks need the confirmation of their customers to use their data. The data is existent, but without the agreement of their customers, banks are not allowed to use it.

Furthermore, there is the opportunity to buy-in data as said in the beginning. This, however, just costs money. Basically, there is already a lot of free information for banks. Especially for the private mortgage financing, there are information available such as living location, working location, frequency of ATM usage at certain places, how often a customer is on the road, how the location where he or she works has developed over the past or are information about his/her household available such as whether there are frequent transactions to another person. E.g., maybe a couple is not living together and have a lot of transactions with each other. Such data might be very interesting to analyze, both with internal data only but also with added external data.

I: Compared to other measures or technologies, how do you assess the potential of predictive analytics in the private mortgage financing process?

Expert 4: Data analytics is currently the most important technology for banks to apply, including predictive analytics as a part of it. Other technologies, such as robotics might help to reduce costs or blockchain might help to automize or clear up processes. However, this is not comparable to predictive analytics which will introduce the most change to the private mortgage financing business.

Also, the mortgage financing business is one of the most important processes as it is one of the largest earnings drivers and a high credit volume is attached to it. Wealth management is another product area that might be of high relevance.

Appendix 15: Paraphrased Interview Expert 5

19. August 2020, 54min

Part 1-3: Introductions

I: Do you agree with the presented perspective on the current market environment for financial institutions or would you add something?

Expert 5: In more competitive and developed markets and sometimes in undeveloped markets, the mortgage application process is growing a lot, supported by videoconferences between banks and potential customers. Consumers are more or less shopping their products. Furthermore, the mortgage product has been simplified a lot over the last 10 years throughout the banking crisis, as no too complex products are allowed to be sold any more. The mortgage products are easy understandable and easy to compare. In addition, the regulators demand from banks that the consumer is well assessed at the KYC process and that the banks are asking money for this step. Thus, we see that customers are building their own file in which they prepare their financial information, etc. in order to get a reduction in fees.

Currently, mortgages are the only product with which banks are making profit on due to the low interest environment. In comparison, saving products only cost money due to the ECB policies. Nevertheless, the competition is accelerating as more and more insurance, investment and banking companies are entering the mortgage market.

Consequently, the mortgage process should be designed as a complete digital process with a lot of DIY functionality included. This is partly already the case in some markets, such as in the UK, Netherlands or Denmark. In terms of retail banking, the German market is lagging a bit behind in terms of their processes and digitalization.

In general, the expertise within banks about the potentials of AI, machine learning, etc. are still very thin. Over the last 20-30 years, the IT departments within banks focused on controlling everything. The problem with AI is that it’s not necessarily explainable and rather a black box. Consequently, banks are afraid of implementing such new technologies as it is hard to understand it. Often, the IT departments are rather conservative and try to keep such technologies out of the organization. Next to this, another hurdle is that implementing AI is not about developing software but analyzing data. The programming part itself is very small, but gathering the required data is becoming more and more important. For banks, it’s hard to change their mind after having worked for a certain way for 20-30 years and then change it from one day to the other.

The bottom line is, that this kind of technology needs a different mindset and that change still has to come in most of the IT departments. I am not only talking about the credit-score here, but the technology in general.

I: Do you agree with the presented private mortgage financing process?

Expert 5: Yes.

I: Do you agree with the presented definition of the term predictive analytics or would you emphasize or add something?

Expert 5: An AI model itself can also add different insights in an analysis and is not necessarily based on historical data.

Part 4-6: Research Questions

I: Which aspects are the least efficient in the private mortgage financing process?

Expert 5: Automation should be used to read all the documents the consumer is exchanging with the bank, meaning the whole process of reviewing the documents and gathering all the data is getting more and more automated, as this – as of today – is the least efficient part in the process. This automation also helps to get a richer pool of information for the predictions.

I: At which points in the private mortgage financing process do you see potentials for the application for predictive analytics?

Expert 5: Our fields of application focus on the risk-management.

One field of application is in the credit-scoring process which is our base use case.

Another field of application is the pre-qualification of customers for the bank which is mainly applied in more developed markets. So that the banks can save costs by not pursuing the KYC process.

Furthermore, we use the model to do default predictions and setting up risk profiles of consumers.

The model can also be used to predict churn.

Our model is based on all type of behavioral aspects (psychometric data) and we are asking the consumer to answer between 7-10 questions. Based on the answers, the response time, the software of the mobile phone, the device used and further aspects, we can determine the prequalification of the credit score. The data provided on the mobile phone varies between countries. With PSDII, we are also able to include the transactional data of the consumer in our analysis.

Unfortunately, this model is not applied in the German market, as banks are more conservative, and consumers are more sensitive about using personal data. I am convinced it will come at some point, but it will be a little bit later than in the other developed countries.

For the credit-scoring process, we have done this already with more than two million customers and on average we can see a reduction in defaults by 20%.

I: At which points in the private mortgage financing process do you see the most potentials to apply predictive analytics in the short-, medium- and long-term?

Expert 5: The fields of applications I mentioned before are currently all offered by our company. As we focus on the risk management, we currently do not pursue further use cases with our model.

I: At which points in the private mortgage financing process do you consider the potential for FinTechs higher to create a more efficient process? Why?

Expert 5: I am convinced that the whole process for mortgage applications will be completely digitalized and automated. All this, using new technologies such as AI, Big Data, automatization procedures and so on.

In general, FinTechs are faster to implements such technologies compared to banks.

I: To what extent could banks anticipate and counteract this development?

Expert 5: There are two models. Either they integrate Fintechs' services, as FinTechs often offer their products as software as a service which can be integrated as a white-labelled product in the banks' processes. The other opportunity is that they acquire FinTechs and implement their products in their processes. There is definitely a trend that large, international banks are pursuing such strategies and build networks with FinTechs. For many steps in the banks' processes, there are FinTechs focusing on such a special process step.

I: Which role does the legacy system and its architectural structure play for such cooperation?

Expert 5: Our service is cloud-based, and APIs are used to connect to the banks' infrastructures. Most banks are also starting with a cloud-environment within their own domain.

By the way, for the use of machine learning in our cases, it is essential that the banks share their historical data. The amount of historical data is very important to enrich the model and the algorithm. If the machine learning capabilities are not shared with other banks and is only used within the own environment, then a bank would be lagging behind with regard to the usability of the model. Nevertheless, there is a decoupling between the model and the data. Thus, banks need to share the enriched model and not their customer data. We can then further develop the models in each bank based on the advances in the model of a single bank. It is important to have high volumes in order to have a very mature volume. That's why large American and Chinese banks have a competitive advantage. Consequently, it is of even higher importance for European banks to share their models and advances among each other to push the quality of the models even further.

I: Which type of predictive analytics is used in the respective use cases?

Expert 5: For our model, we use a regression model which uses parts of machine learning, artificial intelligence and DM. This is also accepted by all necessary regulators.

As our technology is still very young, we will change and adapt our model over time. Thus, we possibly might integrate a different kind of model than regression model in future as well.

Around 80% of our customers do not individualize our product at all and just integrate it into their services. Some, however, want to integrate not only one product step but the whole risk management. In such cases, there is more customization possible.

I: Which risks and challenges are considered when applying predictive analytics?

Expert 5: The black box issue is a very important one. In practice, regulators are more sensitive with this issue than banks themselves. Nevertheless, this means that banks have to build up expertise within their own bank in order to understand these models to review everything by themselves and not rely on FinTechs for this. With regard to building up those capacities, most banks are lagging behind as it is hard to find that kind of capabilities and banks are often not the first address as an employer for people with such capabilities.

I: Which additional data could be generated in order to increase the predictability of the respective use cases?

Expert 5: We think that our model based on psychometric data is sufficient. Nevertheless, banks have the desire to include even more data as they're convinced this will improve the outcome. However, the more data, the more complex it gets and more interpretation issues, data type issues, etc. arise.

I: Compared to other measures or technologies, how do you assess the potential of predictive analytics in the private mortgage financing process?

Expert 5: Custom Automation Technologies are very important as well. This helps to digitalize unstructured documents.

Appendix 16: Paraphrased Interview Expert 6

20. August 2020, 1h 17min, translated from German to English

Part 1-3: Introductions

I: Do you agree with the presented perspective on the current market environment for financial institutions or would you add something?

Expert 6: For building societies, the anchor product itself is the saving process for a mortgage. The actual mortgage loan is rather a cross-sold product. Nevertheless, when talking about banks in general, it is correct to define a mortgage loan as an anchor product.

Part 4-6: Research Questions

I: Which aspects are the least efficient in the private mortgage financing process?

Expert 6: The collection and processing of application data is currently the least efficient process step. Here, OCR, is able to advance the process much further and there is still a lot to be done. However, this is not a field of application for predictive analytics. Currently, a lot needs to be evaluated by hand and only a pre-selection is done with OCR although it would be more efficient to have OCR technology which could analyze everything with much more detail.

I: At which points in the private mortgage financing process do you see potentials for the application for predictive analytics?

Expert 6: In the beginning, we have the prediction of customers needing a mortgage loan. This helps customer-advisors to approach customers in time via e.g., e-mail, phone or postal letters. For this, we use different kinds of data, including contract data, website data from our own website when e.g., a customer plays around with a financing-calculator to check different conditions. Furthermore, we buy leads from comparison website such as Immobilien-Scout which include all data the customer provides to this platform. However, Immobilien-Scout does not apply predictive analytics by themselves, as they only generate leads and sell them to loan providers. They do not know which customer actually reacts on the providers' offers. For me, a next step would be to combine these different data points and apply predictive analytics here. Here, an early categorization of customers who are likely to get a financing in the end would be useful. This, however, necessarily needs a combination of the comparison website provider's data and the bank's data on the type of customer who actually receives a financing in the end.

Another opportunity would be to build a predictive analytics tool based on the data we own ourselves. Such data include the customer data we have from the customers saving their money at our institution for a mortgage financing and predicting, how likely they are to redeem their opportunity and ask for a mortgage loan. Unfortunately, this does not include the account transaction of the customers, as we do not get these data from the bank itself – we are only the building society.

Our challenge is that we do not hear about the latest changes in the customers' lives. For this, a closer cooperation to their house-bank would be necessary. Instead, we use website-tracking data and micro-geographic data, such as age distribution, etc. Using social network activities can unfortunately not be used as it still represents an uncommon thing to do for German institutions. Customers would rather be shocked if we would predict everything about them or especially, what they could potentially be interested in. Nevertheless, there are already institutions who process such data and collect loads of information about customers. In the end, this can be anonymized and shared with institutions such as ours.

A further field of application is risk controlling. However, here, the application is limited due to the regulation by BaFin – but I think, we will dive deeper into this at a later point in the interview.

In addition, we see a potential use case in fraud detection. This is included in the application step of you process where our institution decides whether to move forward with a customer or not.

I: At which points in the private mortgage financing process do you see the most potentials to apply predictive analytics in the short-term and medium- to long-term?

Expert 6: Most use cases we talked about are currently applied or in the making. I see medium-to long-term potential in the utilization of social media activities, customer (transaction-) data from our partnering cooperative banks and a combination of data pools from comparison websites with our institution.

I: At which points in the private mortgage financing process do you consider the potential for FinTechs higher to create a more efficient process? Why?

Expert 6: I especially see potentials in the initial phase of a bank-customer relationship. As said before, the collection of personal information could be much leaner and more efficient. However, this does not necessarily make use of predictive analytics.

FinTechs' advantage is that they do not have an old-complex administration and processes. They are able to automate their processes from the very beginning compared to banks with their older, traditional processes.

Apart from that, I do not see a big advantage of FinTechs over traditional financial institutions with regard to the application of predictive analytics. Especially in the customer acquisition, where banks have the know-how and the customer contacts to work with. FinTechs still rely on banks as an actual loan provider in the end. In case FinTechs would get a banking license, similar to Check24, this might become an increasing threat as they are in direct customer contact with their comparison website and would not necessarily rely on banks anymore.

I: To what extent could banks anticipate and counteract this development?

Expert 6: I would not say that we are under pressure by FinTechs. Our institution didn't start their lab because of them but rather because of the general development of technologies and the advancements of competitors. E.g., we make use of micro-geographic data for many years already and the obvious next step is to try to enhance this data even further with the latest technologies or methods.

I: Are cooperation between FinTechs and banks in place? (or are considered?)

Expert 6: No, not with regard to predictive analytics. Our institution is part of a large group where we have a lot of cooperation and where we are providing services to each other.

I: Which type of predictive analytics is used in the respective use cases?

Expert 6: In cooperation with our service provider Fiducia AG, we self-developed a model based on random forest (classification models) which should be available for the whole group and not only our institution. This model is mainly used for marketing purposes or product recommendations which is not a regulated area by BaFin. This model has the advantage that it will be used throughout the whole group, meaning that it can also include transaction data of customers which are normally not available to us as a building society. Furthermore, google analytics data and micro-geographic data will be considered. In such a case, the building society might realize that the customer needs mortgage financing and the actual mortgage loan might be given out by the cooperative bank and/ or by the building society. This also depends on the amount of loan requested.

Logistic regression is also currently used by the building society in their customer approaching models. Their models are licensed. Thus, if they wanted to use e.g., random forest modules, this would not work as it's not included in their product licensing. However, for more licensing fees, it could get included. After all, one can say that we self-adjusted and self-developed the models based on licensed platform and provided packages. But programming the full algorithm by us is not conducted.

We do not apply machine learning in this model yet. Initially, we use clustering methods to segment customers into clusters in order to approach them in a more targeted way.

Currently, we are in a transition phase in which we want to shift from the licensed program to more open-source software. Here, we use Azure by Microsoft. This has the advantage that we can include further programming languages by ourselves and are not restricted such as in the licensed model which I described you before. Thus, we are more flexible on what type of algorithms we want to utilize, such as Artificial Intelligence, Machine Learning, etc. which can quickly and individually be adapted. Here, we can use integrated packages of Azure itself, but it is also possible to integrate tools of other companies, such as DataIQ, SAS, etc. The infrastructure by Azure offers computing power and a set of packages which can be used and further adapted, added on or developed. A lot is based on the programming language "Python" and consequently, all types of predictive analytics tools can be included that are available in packages. However, also other programming languages, such as R, or self-developed packages by e.g., DataIQ can be integrated. Consequently, one can say that we develop the models ourselves based on open-source and paid for algorithms and data.

For the fraud detection use case, we use logistic regressions although it is not strictly regulated by BaFin and consequently, we could also use another type of predictive analytics. It is not regulated by BaFin, as it is used in the application process and our institution then decides whether to proceed or whether not to proceed with the customer.

I: Which risks and challenges are considered when applying predictive analytics?

Expert 6: One challenge for sure is the regulation. But this highly depends on the process step and use case we are looking at. Credit approval or disapproval process needs to be approved by BaFin and needs to comply with their rules. Thus, no black box algorithms may be used which only give a yes or no as a result. E.g., in the marketing department, there is more freedom with regard to what algorithms can be used as the algorithm only decides whether a customer is or is not approached. Here, it is no problem to use e.g., random forest models which are kind of a black box and whose results cannot yet be explained by our employees.

In comparison, the logistic regression models can be used in such BaFin regulated areas.

Another very big challenge is data security. Here questions arise such as whether we are allowed to use the available data, whether we are allowed to score certain customers, whether we need to get the allowance for all such procedures, etc.

The transition to using the platform for our models is another challenge we currently have.

I: Which additional data could be generated in order to increase the predictability of the respective use cases?

Expert 6: Such data would be transaction data from the customers' house banks, in all cases cooperative banks respectively. Furthermore, making use of social network activities will be very interesting in future. However, as I told you before, we currently cannot use it.

I: Compared to other measures or technologies, how do you assess the potential of predictive analytics in the private mortgage financing process?

Expert 6: Mainly, OCR for the application process would be interesting to apply in this process as well. Furthermore, the application of machine learning or artificial intelligence in all kind of processes will be relevant in future – not only in the private mortgage financing process.

In addition, a very personalized approach to customers will be an idea. E.g., using the information available about the style and age of the house. This would base on classification models and not on predictive analytics.

Appendix 17: Paraphrased Interview Expert 7

3. September 2020, 44min, translated from German to English

Part 1-3: Introductions

I: Do you agree with the presented perspective on the current market environment for financial institutions, the presented private mortgage financing process and the presented definition of the term predictive analytics or would you add something?

Expert 7: Yes, this is all fine by me.

Part 4-6: Research Questions

I: Which aspects are the least efficient in the private mortgage financing process?

Expert 7: Regarding the creditworthiness check, there are the largest inefficiencies. Here, all kind of documents need to be collected, processed and created. So, these steps until the creation of the contract are the largest inefficiencies.

I: At which points in the private mortgage financing process do you see potentials for the application for predictive analytics?

Expert 7: The best step for the bank would be to enter the customer journey as early as possible, namely exactly at the point or shortly before the point where the customer thinks about his need for a mortgage. For this, google analytics, transactional data, comparison website data, etc. could be utilized. Meaning, to identify the need for a mortgage financing as early as possible.

Furthermore, I see a field of application in the area of credit approvals. Meaning the creditworthiness check in the review step.

Additionally, there is a potential in the portfolio step, e.g., how the creditworthiness is developing over time and what peer customers are doing.

Looking at the refinancing phase, based on repayment patterns and other behavioral patterns so far, one could conclude suggestions for future developments and suggestions could be made for a custom-made refinancing.

But also, in the informational phase, suggestions could be made to the customer based on predictive analytics. This must not necessarily be with regard to the financing but can also refer to what kind of property would most likely suit the needs of the customer.

Unfortunately, we are not yet working on such use cases for predictive analytics in our institution.

I: At which points in the private mortgage financing process do you see the most potentials to apply predictive analytics in the short-term?

Expert 7: Especially generating more efficiencies in the area of review and credit assessment. These processes are often very manual and will be digitalized as soon and as good as possible.

Also, the use cases in the informational phase and to identify their needs as early as possible are currently already happening and are likely to be developed further in the short- to medium-term.

I: At which points in the private mortgage financing process do you see the most potentials to apply predictive analytics medium- to long-term?

Expert 7: The portfolio use case, will rather be in the medium- to long-term or rather after the other previously mentioned use cases have been implemented. Such a use case is rather a nice to have on top, once the other things are done.

Similar to this, the use cases in the refinancing phase are also rather in the medium- to long-term.

I: At which points in the private mortgage financing process do you consider the potential for FinTechs higher to create a more efficient process? Why?

Expert 7: In my opinion, the danger through FinTechs is often exaggerated. There is a huge market for them, and they are active in all sorts of business fields and services. They often have similar business models which might ultimately lead to consolidations between them or to the domination of a few. Some FinTechs are also only able to survive due to cooperating with large banks. Consequently, FinTechs are not necessarily a threat by themselves.

Nevertheless, they are able to quickly develop with an agile approach, do not have thinking and bureaucratic barriers. This is compared to traditional banks, where they want to develop something and always need to embed all sorts of departments in the process and need to check for available resources, etc. All this is more flexible in the environment of FinTechs. They are also able to think everything more from a customer's perspective. With regard to technical aspects, also FinTechs are able to test a lot around.

In principle, banks and FinTechs are able to nurture from each other. FinTechs look for banks in order to acquire customers as this is their major problem. They have nice front ends and well-developed technologies and interesting products but lack customers. On the other hand, banks have the customers, but do not have such innovative developments and products. Thus, I am convinced that there is a lot of potential with regard to cooperation. However, not in all business fields. Especially in e.g., data analytics or in back-end processes of banks, I see a lot of potential for such cooperation.

I: Are cooperation between FinTechs and banks in place? (or are considered?)

Expert 7: We have cooperation with a few FinTechs, however, not yet in the private mortgage financing process with regard to predictive analytics. Nevertheless, there are a few cooperation in place in other business divisions and we are not afraid to enter further cooperation, we embrace them.

As previously mentioned, for banks, I especially see potential for cooperation in back-end processes or technology. If you cooperate with FinTechs that are directly operating at the customer-interface, there is always the danger that they kind of take away the banks' customers in the long run, by convincing them with their lean processes, nice front ends, etc. Thus, white-label solutions would be another way to have FinTechs included at the customer interface.

I: Which role does the legacy system and its architectural structure play for such cooperation?

Expert 7: Our legacy systems are provided by a service provider. They have been rather restrictive with regard to opening their APIs. However, in future, it is planned to cooperate better und and open such APIs. For this, an open product architecture is a success factor for the future.

I: Which type of predictive analytics is used in the respective use cases?

Expert 7: Most likely, this has going to happen with cooperation and know-how that will be bought in. Being a smaller bank, such specialized know-how is not available and thus, we either employ specialists to carry those topics forward or we have specialists who work together with external provides to integrate their know-how.

Being part of the cooperative bank network, we also appreciate solutions coming from this network. However, we would rather initiate this ourselves.

I: Which risks and challenges are considered when applying predictive analytics?

Expert 7: The algorithm can only be as clever as the team that has developed it. Thus, if market development or risks are not correctly considered, this might lead to wrong outputs and thus, losses or not conducted business which might have ended in a profit.

Data security is also a major point, such as the compliance with DSGVO and the evaluation of personal data. Manually, advisors are allowed to analyze customer data but not in an automated way. Consequently, according to DSGVO, you need the agreement of the customer to evaluate their data for further purposes. Also, the risk of discriminating certain customers is existent, e.g., when based of automated evaluations someone receives a refusal. I think, this could especially begin in the introduction phase of such algorithms. Normally, you never start with the 100%-

solution but rather, the technology gets adapted over time. Here, the problem is that the organization starting off with such solutions will also be the one making first mistakes and being criticized for it.

I: Which additional data could be generated in order to increase the predictability of the respective use cases?

Expert 7: We also work a lot with Sinus-milieus, for which customers are segmented into different cluster in a matrix. Such external data helps to analyze different target groups and can be purchased from providers. Furthermore, social media activities and geo-data can be very interesting to analyze. E.g., there is one bank which can track the location of their customers when they use their app and also agreed to it. Once the customer has been to an empty property of the bank's portfolio, they would send them a letter offering the customer a mortgage financing.

I: Compared to other measures or technologies, how do you assess the potential of predictive analytics in the private mortgage financing process?

Expert 7: This is currently RPA. They are applied in areas such as automating processes which have been processed manually before. Both at the interface to customers, but also for internal processes.

Appendix 18: Paraphrased Interview Expert 8

7. September 2020, 1h 1min, translated from German to English

Part 1-3: Introductions

I: Do you agree with the presented perspective on the current market environment for financial institutions, the presented private mortgage financing process and the presented definition of the term predictive analytics or would you add something?

Expert 8: Yes, well described.

Part 4-6: Research Questions

I: Which aspects are the least efficient in the private mortgage financing process?

Expert 8: Checking the creditworthiness is a very time consuming and complex process step. However, the results resulting from this step are essential for the bank to move forward with the customer and thus, it's worth the effort as you get the full insights.

However, the successive process step can be called the least efficient process step. Once the approval is done, a bulk of paperwork is necessary to fix the collateral, enter everything into systems, consider additional agreements, etc. All this requires a lot of manual work which is not too complex but needs to be done. Here, many data and quality issues may arise. As the credit is already approved, this does not bring much added value to the product itself anymore and thus, the contract step can be considered the least efficient process step.

I: At which points in the private mortgage financing process do you see potentials for the application for predictive analytics?

Expert 8: In the advice and application step, predictive analytics cannot be applied that much.

Definitely in the review step. Here, it's about automated pattern recognition, correlation effects and statistical analyses, etc.

Also, in the creditworthiness review, predictive analytics helps giving an insight into possible future events, e.g., based on recognized patterns of other customers. Currently, our institution does not use this technique for private people. We make a creditworthiness check of the customer based on the fact that the debt service capability can be generated from the customer's income (this is also required by the regulator). The scoring of customers may include predictive analytics as it is a mini-rating procedure which is based on the behavior of the customer's account, including e.g., number of direct debit returns, income, saving quota and further factors.

Currently, factors for the default of private mortgage loans are the death of a customer, divorce, etc. Here, developing a model based on predictive analytics which predicts the likeliness of the occurrence of health problems, a divorce or a job loss of a customer, might be possible. For this, data would have to be collected first. However, with DSGVO, the agreement of customers would be required and e.g., for health data, a cooperation with health insurance companies is needed. Thus, it would theoretically be possible, but in practice there are no saving banks currently applying it.

For credit institutions in the private mortgage financing business, they can rely on the collateral of their financings. Thus, if the buying price of the property is appropriate and the institution is in first rank for the collateral, everything is fine, and the financing can be done. Even if in a large financing there would be minor doubts, the financing could be done due to the appropriate collaterals and there would be no need to make an effort and apply advanced models to approve or disapprove customers.

Looking at large projects and real estate developers, which generate cash flows from the object itself, we also use rating procedures which partly include predictive analytics. There, simulations are made over the lifetime of the project, including Monte Carlo simulations to predict whether the cash flows are able to service the debt under different scenarios. For such simulations, I see big potential for a further development based on predictive analytics tools.

Furthermore, predictive analytics can be used in other parts of the life cycle of a credit loan. Such a use case would be to recognize certain econometric anomalies in the credit's life cycle and to get a feeling whether the project is running as planned or whether there are indicators which show e.g., repayment difficulties and the need for a workout management. Until now, our institution does not conduct this use case. Rather, the credit loan is given out to private costumers and we only get aware of it again when there are actual payment difficulties. Again, this is mainly due to the reason that we have the collateral to secure the credit.

In the application process, predictive analytics could be applied. E.g., in order to initially classify customers as more or less price sensitive and risk averse or risk seeking. Consequently, such a customer could be treated differently to a customer nothing is known about.

There are many use cases which seem to be feasible. However, it's necessary to differentiate between theory and practice. While in theory use cases seem straight forward, the reality is much more complex and the legacy systems of banks, which are the basis for such analyses, are outdated. Thus, data is not that well accessible in the banks' systems in order to be able to be examined by such tools. Adding further data to these data bases would be a next step after the systems are once decently structured.

With regard to approaching customer, predictive analytics can be used to identify the needs of customers and support advisors in approaching them accordingly. We developed such systems which use transactional data of customers that agreed to it. Here, the result shows that a very experienced advisor without technical support who knows his customer is able to identify such needs in the same way as the model. Consequently, the system outruns advisors with less experience.

Another interesting use case is in the contract and disbursement step and focuses on the capacity utilization of employees. Predictive analytics can more precisely predict how many customers might approach for mortgage financings than rough manual estimations. Consequently, internal processes and the capacity utilization of employees in the settlement or advisory steps can be adapted to the expectations. This could be developed even further to the point that prices are adjusted seasonally to adjust the demand for mortgage financings. E.g., when the capacity utilization is expected to be low, prices can be decreased to increase the capacity utilization.

Predicting customer churn to prevent them from leaving, is another potential field of application. Which customer might think about leaving due to which factors? However, it is really difficult to predict this, as customers behave very differently and not all data is available to us. While some customers break off their relationship abruptly, other customers reduce their turnovers over time before cancelling their account. For us, there is no easy way to build such a model as you cannot look into the minds of customers. If there existed a system to successfully predict this churn, this would be a very good progress and sell very well. But the lack of data availability, currently, seems to be too big of an issue. Such data also includes e.g., price adjustments of competitors, competitive landscape, own price developments or the place of the customer’s friends’ accounts, etc.

Like this, there are many ideas where predictive analytics could be applied in theory but usually, this lacks the availability of data and a correct model which has been validated.

I: At which points in the private mortgage financing process do you consider the potential for FinTechs higher to create a more efficient process? Why?

Expert 8: I believe FinTechs have an advantage for two reasons. Firstly, they are newly founded and can build their infrastructure from ground up. Secondly, they are focusing on one business product or service alone and not on a full range such as traditional institutions. Thus, they are able to relatively put more effort into this one business field and are able to include further data to this in order to develop a certain model which can predict potential future events related to this specific business field.

This yields an advantage to FinTechs in such models. Nevertheless, they are most likely to offer this product to banks to support their processes as banks have the financing power and the customer contact.

I: To what extent could banks anticipate and counteract this development?

Expert 8: As mentioned before, cooperation between traditional banks and FinTechs will play an important role. This can also be the case in the private mortgage financing process, e.g., in the approach of customers or in the prediction of potential events based on customer data when there exists a FinTech which has a good, well-functioning product.

This is also the current case in our group, where we have one service provider which experiments with different projects and provides us with working solutions for our processes.

I: Which role does the legacy system and its architectural structure play for such cooperation?

Expert 8: The current IT-infrastructure is very complex, but on the other hand very robust which is a very important point which should not be neglected. This leads to a loss of speed.

Our service provider created an extra data warehouse center for their portfolio and rating models. For this, they extract the data they need from existing data systems and put them into a totally new data structure which is the basis for their latest projects.

I: Which type of predictive analytics is used in the respective use cases?

Expert 8: In order to identify influencing factors which are typical for defaulted and non-defaulted cases, partial least square is utilized among other types of regression methods and statistical methods. Neural networks are currently not used. Looking at the path dependent methods, such as the tree-models, those models are used for large-sized real estate investment. Such methods are always a part of a whole rating system and never used as a standalone. The final rating itself checks different given factors, weighs them and creates a rating out of them. And as previously mentioned, they are currently rather used for commercial and not private mortgage financings.

I: Which risks and challenges are considered when applying predictive analytics?

Expert 8: Everyone is aware of the progress of digitalization and that there will be an increased support by machine learning, artificial intelligence, etc. Thus, this should rather happen today than tomorrow and is mostly viewed as a big chance for future development, especially as with a big enough portfolio the statements are more and more reliable.

Years ago, when rating procedures came up, such systems were seen very critically by the best analysts which only believed in humans assessing the customers and not letting weird algorithms do this task. However, nowadays, this attitude has changed especially in the medium- and younger-aged generation. It is accepted that there is always a probability of error, but it is also known that the tools are able to give good hints to analysts or advisors to support them in their tasks. Thus, such algorithms should not be trusted blindly but rather reviewed with a healthy common sense and used as a support. However, when looking at the volume business, such as consumer loans, it is widely accepted to trust the output of such tools. The bigger the business gets, the more individual you need to work and the less chances you have to argument with the law of big numbers.

With regard to the acceptance or refusal of customers, banks are free to choose a model which does not need an approval of the regulatory authorities. A regulatory approval is only relevant for the creditworthiness-check where banks define how much equity they have to provide for a loan. This is relevant for the steering of the organization and thus, falls under the supervision of the regulator. E.g., based on such a model, general value adjustments could be made.

I: Which additional data could be generated in order to increase the predictability of the respective use cases?

Expert 8: Socio-economic data would be very interesting to include in such models. This is partly already used outside of such models, e.g., based on the living-location of a customer they might be approached with product offers or not.

I: Compared to other measures or technologies, how do you assess the potential of predictive analytics in the private mortgage financing process?

Expert 8: Currently, the biggest challenge is that you do not only need to predict the borrowers over the next years, but also the credit institutions themselves over the next years. This includes what is potentially going to happen under the condition of certain interest rate developments, default rates, economic conditions etc. These overall bank simulations, which can only base on a simulation such as Monte Carlo, are required by regulators and are still improvable as they require a lot of data and consider various assumptions and interrelated relationships. For this, enormous computing power and advanced systems are required which are able to process the parameters in a way they do not output contradicting results. Setting up all this kind of procedures is one of the main challenges for banks.

Furthermore, data analytics is highly relevant for the sales team of banks. Here, it is already processed but getting meaningful statements are still restricted.

I believe that we currently evaluate many ideas on the application of predictive analytics, but that it might take another century or so in order to get all sorts of relevant data together and to successfully apply such models.

Appendix 19: Paraphrased Interview Expert 9 and Expert 10

15. September 2020, 45min, translated from German to English

Part 1-3: Introductions

I: Do you agree with the presented perspective on the current market environment for financial institutions, the presented private mortgage financing process and the presented definition of the term predictive analytics or would you add something?

Expert 9: Yes, fine from my side.

Expert 10: Fine from my side too.

Part 4-6: Research Questions

I: Why is predictive analytics currently so relevant in the private mortgage financing business?

Expert 9: Predictive Analytics gets more and more relevant for us, as our IT-infrastructure is constantly getting better and processes are increasingly digital. There is a large information budget which now needs to be meaningfully analyzed. It needs to be checked where we can get even more efficient in our processes. This development has increased drastically over the last 10-15 years, so that a point is reached where banks can review the possibilities and make a next step.

Expert 10: I share this view.

I: At which points in the private mortgage financing process do you see potentials for the application for predictive analytics?

Expert 9: We had a thorough look at the customer journey of online banking. Probably the most interesting field of application seems to be in the beginning with the business initiation and the connected sales service. This would be to identify customers as early as possible who could potentially be interested in a mortgage financing.

We also offer customers a buyer-certificate which confirms their financial soundness to the seller when visiting a mortgage. We point this possibility out via advertisements on real estate comparison websites. Customers enter their private and financial data and we can make a prior creditworthiness check which the customer can use to show it to the seller. We also process this data and, if the value of the property is reliable, can offer the customer to give him/her a mortgage financing for this object. Here, predictive analytics could even further improve this business by e.g., checking their behavior in the online search-process.

In addition, as conditional offers are mostly based on the creditworthiness-situation and situation of the property, the more information is available for processing, and the quicker and the more precise one can evaluate a customer (for which predictive analytics can be used), the better this conditional offer can be in order to make it as custom-made as possible to stand out from the competition.

Predictive analytics could also enhance fraud detection before credits are given out. If enough data was available on other defaults, etc. such a model could point out a possible issue and ask for a more thorough investigation before giving out the loan.

Expert 10: In the risk analysis department, predictive analytics plays an important role. In the surrounding of customer ratings, very standardized methods have been used in the past. However, with all new analysis methods, the customer behavior can be analyzed much more precisely. The information which is available about customers online is immense and exceeds the information which are currently in possession of a bank about the customer. Thus, shopping behavior, website tracking, tracking of other special interests, or whether someone’s children use their parents web-shop-account or not, etc. can be analyzed and conclusions can be drawn. The whole risk assessment will significantly change in future and I believe that we will not ask for salary statements in future anymore. Rather, our models will be better to make credit decisions about the information which we gathered internally about the customer and the information which are freely available by the customer to us.

Expert 9: Here, also the inclusion of transactional data plays a big role and can be included in the risk assessment of customers. With PSDII, banks have to give access to such data about their customer if they agree to it. Thus, banks can aggregate several accounts, and this could potentially also be used in order to approve or disapprove a mortgage financing for a customer. Here, also the outputs from credit rating agencies should be acquired and included in the model assessing the customer risk.

Furthermore, predictive analytics sees application in the real estate evaluation. Provider analyze and compare market data and include historical data in order to identify short- or medium-term price adaptions and conclude what the realistic value should be. This can be used in order to predict the used collateral and its future development.

What's more, is the application of predictive analytics in the prediction of potential cross-selling products to the customer, including e.g., the proposal for a prolongation of the loan in advance with customized conditions or an additional consumer loan. This might even include the proposal for another mortgage after let's say 20 years, after the children have moved out and there might be interest in moving into a smaller house.

I: At which points in the private mortgage financing process do you consider the potential for FinTechs higher to create a more efficient process? Why?

Expert 10: Speed and the possibility to experiment without the regulatory framework of traditional banking institutions are major advantages of FinTechs.

Expert 9: Furthermore, IT-specialists with focus on such technologies are rather found in FinTechs than in banks. This has the advantage that in a cooperation, a bank or building society brings in the technical know-how and the FinTech can focus on technical aspects.

I: Are cooperation between FinTechs and banks in place? (or are considered?)

Expert 10: There are several FinTechs active and we also closely work together with these. However, they do not necessarily primarily focus on predictive analytics. Such FinTechs include e.g., Hypoport or Interhyp which provide platforms for the initial steps in a customer journey. As a bank, we use their platform and use our competency as a credit financing institution and use their competency with regard to customer approaching.

Our institution does not cooperatively analyze customers with such FinTechs.

If there was a FinTech advancing with predictive analytics tools in e.g., creating creditworthiness checks on customers, we would willingly cooperate with them which gives them the platform to test their product and advertise their company.

I: Which role does the legacy system and its architectural structure play for such cooperation?

Expert 10: This is still a development. Updating APIs and further interfaces is an essential point which also poses a current challenge.

I: Which risks and challenges are considered when applying predictive analytics?

Expert 10: Data management is a major point. Eventually, the data is wrongly evaluated. You can collect as much data as you want, and it will not bring you an advantage in the end when you wrongly interpret it. Consequently, this might lead to a cut off of business which you eventually would have conducted before using the new analyzing methods or you all at once include customers in your portfolio that you would not have included priorly and vice versa.

I: Which additional data could be generated in order to increase the predictability of the respective use cases?

Expert 9: All information that give insights about the financial situation of a customer would be nice to use, including transactional data through APIs coming into place through PSDII.

It is possible to make a credit risk assessment based on various doings of a customer. This can be seen as a similar thing to the rating of large companies by rating-companies. Such doings include all sorts of doings connected to the finance-behavior of the customer, such as credit cards, number of credit cards, usage of credit cards, credit limits, PayPal transactions, education, shopping behavior, etc. If all these data were available, this would advance the customer assessment. However, with regard to data security, this is critical.

It is important to mention, that such a thorough assessment is not only good for the bank but also for the customer as he/she might be held back from overestimating himself/herself in terms of credit capacity.

I: Compared to other measures or technologies, how do you assess the potential of predictive analytics in the private mortgage financing process?

Expert 9: Currently, predictive analytics is the core of new technologies that can be applied in the private mortgage financing process.

Appendix 20: Paraphrased Interview Expert 11

24. September 2020, 55min, translated from German to English

Part 1-3: Introductions

I: Do you agree with the presented perspective on the current market environment for financial institutions or would you add something?

Expert 11: Yes.

I: Do you agree with the presented private mortgage financing process?

Expert 11: In general, it has to be said that not all mortgage financings are conducted the same way. E.g., sometimes, customers already need a financing approval when visiting and applying for the purchase of a property. Consequently, such a pre-approval can but must not necessarily be placed much earlier in your described process. Such pre-approvals can be binding or non-binding depending on the needs of the customer or their brokers. Non-binding offers before an existing sales-contract are pretty much offered by all banks, sometimes even online.

I: Do you agree with the presented definition of the term predictive analytics or would you emphasize or add something?

Expert 11: Yes, fine from my side.

Part 4-6: Research Questions

I: Which aspects are the least efficient in the private mortgage financing process?

Expert 11: The decision process of the customer takes the longest time in the process. Compared to traditional consumer loans, where customers are keener to make quick decisions and do not make comparison of many providers, for private mortgage loans, customers do the opposite here. While our bank process could directly offer a credit contract after making the due diligence, the customer will take time to compare it with other banks' offers and conditions.

In the bank process itself, the escalation of credit requests to the next hierarchy level, due to insufficient budget statement, takes too much time. In a consultation meeting the employee would defer such a decision to a next meeting or would inform the customer in the hindsight about the result e.g., via e-mail.

Many financial services and IT providers try to standardize the private mortgage financing process as good as possible. This is especially useful for non-complex products such as consumer loans. However, for more complex mortgage financing products where there is a strong competition with a variety of financing models and financing structures. Consequently, it is close to impossible to standardize such financing product in a reasonable way. To some extent, something will always need to be adapted and there will always be an exception which requires special treatment or e.g., the involvement of higher hierarchy levels which might ultimately lead to e.g., special conditions, higher special repayment options or less debt service capability requirements.

Furthermore, the financing approval enriches the data pool and can be used for further predictive analytics.

I: At which points in the private mortgage financing process do you see potentials for the application for predictive analytics?

Expert 11: One potential field of application would be to suggest customers a certain property based on the selection of other customers in similar life situations and with a similar budget. This might especially be useful, as we sometimes give a customer a credit approval and then they do not receive the property they initially intended to buy.

Furthermore, predictive analytics might support in identifying customers who are still paying rent, are in a point in their lifecycle to have potentially interest, have payment surpluses or saved capital and thus, are eligible for a mortgage financing. Such customers would also be able to enter a more standardized process as they'd be chosen based on their positive transaction profile.

I: At which points in the private mortgage financing process do you consider the potential for FinTechs higher to create a more efficient process? Why?

Expert 11: In the private mortgage financing process, I would not say they are able to display the process in a more efficient way. Rather, FinTechs are able to add value by complementing the process of a bank.

Looking at our customers, we realize that they still rely on physical consultation and mostly do not seek to finalize their mortgage financing completely online. They might utilize FinTechs such as comparison websites to get a feeling for conditions of different providers but would still pursue the process on site.

Nevertheless, we see an increase in the utilization of private mortgage financing brokers, such as Interhyp or Dr. Klein. They have several banks in their portfolio and guide their customers to the ones offering the customers the best and most suitable conditions. Such (online) brokers differ from comparison websites to the extent that such brokers pursue a consultation and make most suitable offers based on the outcome.

I: To what extent could banks anticipate and counteract this development?

Expert 11: We cooperate with such brokers. However, for a regional bank such as ours, it is difficult to come on top of their comparison listings due to different cost structures etc. Thus, it is important to cooperate especially with larger brokers and to convince with quality and not necessarily the cheapest conditions. E.g., when we are not on the first places in the listing, the broker would still mention our offer, emphasize good prior experiences, and quick response times. Of course, the conditions will still be in a competitive manner.

I: Compared to other measures or technologies, how do you assess the potential of predictive analytics in the private mortgage financing process?

Expert 11: Nowadays, it is really important that the customer is able to do as much online as possible. Still, a lot of actions are done analog. Thus, digitalizing and automating process steps is very essential and would save time for both, banks and customers.

Appendix 21: Paraphrased Interview Expert 12 and Expert 13

1. Oktober 2020, 56min, translated from German to English

Part 1-3: Introductions

I: Do you agree with the presented private mortgage financing process?

Expert 13: Looking at predictive analytics in the private mortgage financing process, it is essential to consider what purpose the data analysis pursues, at which point of the process it is applied and which data points are used for it. For example, before the actual review is conducted, there is no permission to score a customer and gather credit information about him/her. Consequently, e.g., Schufa data are not allowed to be used for initial process steps such as marketing.

Expert 12: Exactly, also for existing customers, Schufa is only allowed to be used when there is a specific loan application request by a customer, not without that.

Expert 13: Thus, there are significant differences when looking at existing customers or new customers, as for existing customers, we can include certain existing business data in our analyses.

I: Do you agree with the presented definition of the term predictive analytics or would you emphasize or add something?

Expert 13: Yes, that's kind of how it is applied in practice. It is always to be evaluated whether it adds value to develop such predictive analytics tools or whether it is possible to generate the same results with other tools. This always depends on the type of question and what the tools' final purpose should be.

Part 4-6: Research Questions

I: Why is predictive analytics currently so relevant in the private mortgage financing business?

Expert 13: Looking at the banking sector, we can observe a shift away from many people looking after a few customers which resulted in a very close relationship between the two parties. Once a customer chose a bank, it was the main bank to go to for all sorts of products. In comparison, nowadays, competition has arisen and e.g., FinTechs increasingly consult the customers and only forward the product to a bank in the end. Thus, a bank does not necessarily have a rich pool of information about an individual customer anymore. Not having this thorough overview over a customer, statistical approaches are one way to derive solutions or next-best-offers for customers. Furthermore, it has been realized that data analyses based on single rules are not the ultimate solution and that analyses based on pattern recognition etc. can derive much more precise solutions.

Additionally, technology has become faster, cheaper and easier to use.

However, predictive analytics itself is not a new technology and has been used already for many decades, as for example for scoring models. With the technological advancements and the evolvement of the competitive environment, a further focus and investment in this technology became more essential which is why it also is a clear part of our latest corporate strategy and there are whole departments dedicated to data analytics.

Expert 12: I agree with my colleague. On top, the general market environment of banks is constantly progressing. We have a low interest rate environment, a competitive environment with new, strong competitors and thus, also face a high-cost pressure. Predictive analytics offers a chance to lower costs and improve processes' accuracy. This is especially possible due to higher computational power and new dimensions of available data which are, at least in theory, available for analyses. In practice, data protection regulations might impede certain data pools from being used for such analyses.

I: At which points in the private mortgage financing process do you see potentials for the application for predictive analytics?

Expert 12: First of all, predictive analytics can help to identify customers who actually have the need for a mortgage financing.

Furthermore, in the application process it can be applied in order to increase its efficiency by introducing a structure to the available data, assign the data accordingly and in the end, support to create capital surplus calculations from this. This especially happens via analyzing the transactional data of customers, if he/she gives us the Opt-In to do so, which also gives insights into what kind of product characteristics a customer can actually afford. It also has the advantage that a pre-selection of customers is possible. At the same time, it increases the convenience of the customer experience, as once the customer tells the bank about his/her needs, such a tool just needs to be applied to his/her data and a Schufa request needs to be made. This enables a fast credit decision instead of a long process where the gathered data is sent to a back-office for manual analysis. Consequently, even in the very first consultation meeting, the customer can be told whether he/she is able to receive a mortgage financing of a certain volume. Thus, the credit decision process is speeded up.

Predictive analytics tools are also utilized in scoring models. Here, predictive analytics tools are used for economic and regulatory purposes, meaning by how much capital a loan needs to be backed and whether it is economically reasonable to pursue a certain mortgage financing or not.

Furthermore, there are customers who receive “intensive care” due to an increasingly high default risk during the lifetime of their loan. To identify customers who need more or less attention or individual care of our bank, predictive analytics can help. This yields in a more efficient assignment of human resources to topics where they’re most needed. Such models would mainly be based on historical data from the behavior of the existing portfolio and a few other parameters and not transaction data itself.

Concerning the analysis of transaction data, we try to avoid it as much as possible as this data enjoys a special data protection and there is the necessity of an "opt-in", where we need the agreement of a customer to be allowed to analyze his/her transactional data.

All use cases are different as they pursue different goals. For each one of them, different data input is required and also, different data is allowed to be utilized from a data protection view.

Expert 13: In the process step "need", it's about identifying the need of customers based on the analysis of certain parameters via applying predictive analytics. Here, patterns of customers who already received a mortgage financing would be compared to other customers who haven't received a mortgage financing yet. All this without the utilization of risk-data, such as credit rating, Schufa, etc., and data whose utilization require the consent of a customer.

For the renewal of the contract, when the initial contract has expired, there is a much better peer group to compare the customer with, compared to the initial loan decision. This helps to predict the best possible refinancing options.

For the typical rating process, there are two possibilities. First, there is the debt service capability and the degree of over-indebtedness. Second, there is the check whether principal and interest payments are made in time once the loan is ongoing, resulting in a rating about the probability of default. This rating procedure is also based on predictive analytics and is actually one of the oldest predictive analytics applications in banks.

I: At which points in the private mortgage financing process (of your company) do you see the most potentials to apply predictive analytics in the short-term?

Expert 12: The fields of application we mentioned before, are currently explored and also applied.

I: At which points in the private mortgage financing process (of you company) do you see the most potentials to apply predictive analytics medium- to long-term?

Expert 13: A very interesting point is that FinTechs use behavioral scoring models to score customers. Here, we face the question whether in the long-term debt service capability is the one measure to stick to with or whether rather rich data pools, including behavioral and payment patterns will be the basis in decision processes.

There is a wide range of applications and data pools which we are currently not really aware of or cannot utilize due to regulatory requirements. E.g., back then, receiving Schufa data about how many loans etc. a customer has in his/her portfolio would be the most interesting for a new credit decision. Compared to that, nowadays, it might be more interesting to analyze the motion profile of a customer and get insights whether someone is e.g., often in a casino or has other interesting transaction patterns.

As a bank, we have strict regulatory requirements and thus, changing from data sources which we have used for decades for the analysis of customers, is not possible from one day to another, because regulatory compliance is very important to us.

I: At which points in the private mortgage financing process do you consider the potential for FinTechs higher to create a more efficient process? Why?

Expert 13: FinTechs mainly solve specific problems for the customer. This is also their main advantage as they can solely focus on these topics. A bank could do the same, however, a bank also has to take care of their existing customers, their portfolio management and has to prevent technical breakdowns and other risks.

Expert 12: With regard to the consumer loans, FinTechs also apply predictive analytics already. Looking at private mortgage financings, we mainly see FinTechs entering the market with comparison websites. However, it could theoretically be possible that they generate their leads also based on predictive analytics.

I: Are cooperation between FinTechs and banks in place? (or are considered?)

Expert 12: We also cooperate with FinTechs which e.g., mediate loans to us.

Expert 13: There is also a FinTech which supports us with regard to the customers' account turnover analysis. Such cooperation is detached from our credit business. Nevertheless, they are an essential part of our bank. Thus, we support FinTechs at different stages in their life cycle and check whether we can include their solutions in our processes or whether we can directly offer it to our customers.

I: Which role does the legacy system and its architectural structure play for such cooperation?

Expert 13: This always depends. The data environment and connected systems have grown in the past and consequently, IT infrastructure has become more complex. Consequently, we always need to find and create individual solutions. With regard to APIs, our bank follows an API strategy.

Expert 12: It is not going to happen that person-related data are easily exchanged via APIs. Being a bank and holding a customer's account means that we keep our bank secret and do not tell anyone about it. This also means that the customer needs to proactively allow us to share his/her person-related data. Furthermore, with regard to risk profiling, we are not allowed to share these assessments with others and only use it for ourselves. It is one of our highest reputations, because we really stick to the data security and privacy of our customers. Consequently, all use cases are pursued after getting the approval from the data protection department.

Expert 13: In order to get a customer's allowance to use his/her data, it needs to add value for his/her purposes as well. We need to make sure that nothing bad happens to the data and that no negative consequences are feared by the customers because of these analyses, such as e.g., a negative grade which might impede the customer from receiving further credits.

I: Which type of predictive analytics is used in the respective use cases?

Expert 12: We kind of experiment with and utilize all sorts of available methods. Whether it's self-developed, in cooperation with 3rd party providers or whether it includes different types of regression or classification models.

E.g., in the marketing of private mortgage loans, we use a SAS-based application to run rule-based models.

Looking at other current research and development use cases, we do not use standardized software of third parties but rather single modules and then build everything by ourselves in order to find the correct path for the best development. There is a variety of ways of how to create such models and we prefer to use self-developed models in the analysis phase so that we do not rely on ready-made applications. That's also why we utilize all sorts of models ranging from neural networks over regression models, some being more complex, some being more standardized.

There is no clear answer to what predictive analytics model is the best for a certain use case. There is a huge toolbox, and you want to test out all sorts of tools until you find the most suitable one for a certain use case.

Also, with regard to cloud computing, we look at different cloud service providers such as Azure or Google cloud where we can apply our tools. For this, we follow a clear cloud strategy. However, I'd like to emphasize again that this is use case dependent and always in agreement with the data protection department to ensure the integrity and security of our customers' data. Nevertheless, it is important to pose the question, whether the utilization of a cloud is essential for every use case or not, as there are also a range of other possibilities to pursue use cases without a cloud.

Furthermore, I would like to emphasize that stand-alone, all the development and use cases do not help us. It is essential that it adds value to the customer and thus, adds value to our institution as well.

I: Which risks and challenges are considered when applying predictive analytics?

Expert 13: Next to the obvious challenges which you just mentioned, the topic of ethical questions is also very important for us. You need to take care that you do not discriminate anyone because of the decisions made by such tools.

Expert 12: Furthermore, we face the challenge to ensure the true value of such tools. The issue with modern technologies can be that its true added value is not questioned, and it is blindly applied. It is important that such a technology adds value for all stakeholders. Otherwise, it is only a job creation measure.

I: Compared to other measures or technologies, how do you assess the potential of predictive analytics in the private mortgage financing process?

Expert 12: There are many other technologies which are of importance for our bank which is why we defined key technologies for our bank. E.g., robotics is one of them which also has several applications in the private mortgage financing process.

We closely keep track of the market and the general development of these identified technologies and check for the development of potential applications for our products, services and processes.

Appendix 22: Paraphrased Interview Expert 14 and Expert 15

3. November 2020, 1h 10min, translated from German to English

Part 1-3: Introductions

I: Do you agree with the presented perspective on the current market environment for financial institutions or would you add something?

Expert 14: Interest rate margins can mainly be received via savings, transformations and giving out loans. Due to the monetary policies of ECB, there are low interest rates for banks in savings and transformations. Compared to that, small margins are possible in the loan market. A lot of offers and low prices lead to low margins.

With regard to mortgage financings, there is a lot of competition, ranging from banks, insurances and building societies.

I: Do you agree with the presented private mortgage financing process?

Expert 15: Yes.

I: Do you agree with the presented definition of the term predictive analytics or would you emphasize or add something?

Expert 14: Yes, it is important to mention that predictive analytics itself does not only predict future events but unknown values in general. Those might as well have happened in the past or are currently ongoing and must not necessarily lie in the future. E.g., to predict the current or past efficiency of a machine based on different inputs.

Part 4-6: Research Questions

I: Why is predictive analytics currently so relevant in the private mortgage financing business?

Expert 14: It enables us to make forecasts for our sales teams which customers have a certain need or a certain affinity for specific financial products. This is especially important, because we currently see a digital transformation in many areas, also beyond the banking sector which kind of leads to a pulling effect. In the banking sector, FinTechs are advancing a lot and are emphasizing the strategic importance of using agglomerated data. This especially holds for banks who have gathered data over many decades and where this “data-gold” needs to be “mined”. Some banks have realized that other banks, FinTechs and other industries have made a lot of progress here and now need to and want to catch up in order to not miss the train which already left the station.

Expert 15: From a sales perspective, it has always been very effortful to individually analyze customers, find e.g., whether they use third-party products and identify their potential needs before consultation meetings. With predictive analytics, you can now see customers who have e.g., third-party products at the push of a button. This facilitates the workday of salespeople and enables them to conduct a more tailored consultation. Also, because the affinity of costumers to certain products is more known, these products can better be offered. This is in line with our goal to approach customers with products which they actually need most at a certain point in time. E.g., if there is an inheritance case, it is advisable to directly approach a customer with solutions and not a year to late.

I: At which points in the private mortgage financing process do you see potentials for the application for predictive analytics?

Expert 14: Looking at the process step “need”, we use predictive analytics to identify customers’ affinity for a mortgage financing and to predict a probability with which the customer has a need for such a product.

Expert 15: Exactly, we have a sales process, ranging from customer selection over customer approach to the finalization of a product. In this process, we support the customer selection and customer approach.

Expert 14: For this, we use transaction data in order to identify relevant events in the customers' lives, such as e.g., a growing family or customers having started to work. We analyze this data based on expert-rules with regard to certain events and third-party product relations. However, these transaction data are very critical with regard to the automated analysis due to regulatory requirements and the need of a positive declaration of consent of the customers.

Also, we use the master data of our customers, such as demographics, income situation, etc., which are the main part of our scores predicting the customers' affinity for certain products. The predictive model learns interdependencies of customer master data and historic product completions and can apply this knowledge.

Furthermore, we have tools which show the salesperson why their customers are identified to have an affinity for certain products.

As an institution, we also provide credit scoring solutions for our customers, the saving banks. Here, we also include machine learning methods in order to identify patterns within our portfolio about e.g., default rates of customers which then influences the economical decision whether a customer receives a credit or not. For the topic of credit scoring, there are much stronger regulatory requirements compared to the sales topics.

I: At which points in the private mortgage financing process do you see the most potentials to apply predictive analytics in the short-term?

Expert 14: The use cases we mentioned before are currently conducted and in development in our institution.

I: At which points in the private mortgage financing process do you see the most potentials to apply predictive analytics medium- to long-term?

Expert 14: As we are developing data analytics solutions for the whole product range of saving banks, the private mortgage financing use cases are only one product focus of many.

Currently, we see a development away from specializing on such individual products to focusing more on topics such as system integration where data analytics is not necessarily in the foreground anymore.

I: At which points in the private mortgage financing process do you consider the potential for FinTechs higher to create a more efficient process? Why?

Expert 14: FinTechs have totally different legal requirements than banks, e.g., with regard to the usage of third-party data or data security.

Expert 15: With regard to the fields of application, FinTechs are also able to create a more efficient process. For this, banks and FinTechs have different processes running with different requirements for documents that are needed. This is also due to the reason, that FinTechs operate online compared to some bank processes which are currently only able to run analog in a bank branch. This is currently the case for loans and mortgage financings whose process saving banks cannot display online but only in a bank branch.

Furthermore, FinTechs have a lower cost block compared to traditional banks as they mainly operate online and have leaner processes underlying. Thus, they can also offer lower interest rates due to a leaner cost structure.

I: Which role does the legacy system and its architectural structure play for such cooperation?

Expert 14: Some FinTechs follow the minimum legal requirements which we as a bank cannot integrate as such in our processes. Also, because we do not have the technical interfaces available. Firstly, to dock FinTechs to the systems costs a lot of money and second, we are not interested in proactively forwarding customer data to FinTechs.

I: Which type of predictive analytics is used in the respective use cases?

Expert 14: We utilize open source to develop our models. Mainly, we program in R and Python and also apply the available packages to build our models. We use specific packages to develop our scores and also work with packages to apply neural networks on different use cases. In general, most packages and frameworks used by corporations around machine learning with neural networks are written in Python. Compared to R, where rather traditional approaches are used, such as regressions, tree methods or nearest neighbor approaches.

With regard to analysis methods, we use DM and feature engineering among others. Feature engineering kind of belongs to DM and is about building predictive features from the available raw data. This is because we do not have ready-made features which we can train on, and thus, we use our raw data sets to define suitable features.

This is especially required when classical methods are applied as the data scientists need to provide best possible predictive features for the models. On this basis, a model can learn predictive relations within the data. Compared to that, looking at e.g., neural networks, feature engineering by data scientists themselves is not required any more, as neural network models are set up in a way that they automatically conduct feature engineering. This means, that a neural network model can learn predictive features on basis of its raw data.

Looking at the customer master data analysis, we use classical approaches and thus, we as data scientists, are required to provide suitable features on basis of the raw data. For our scores, we use gradient boosted trees.

Almost all machine learning methods can be used to solve classification or regression problems. Whether it is neural networks, regression models or tree algorithms, such as gradient boosted trees. We also use gradient boosted trees to solve both kind of problems.

With regard to the private mortgage financing process, we basically have a classification problem in the beginning. It’s whether a customer will conduct a mortgage financing or not. The regression model is about the possible volume of the contract. Meaning, in case a customer conducts a mortgage financing, what volume do we expect him/her to finance. Consequently, banks can either approach customers which are most likely to conduct a mortgage financing (classification problem) or which have the highest volume (regression problem) in case they conduct a mortgage financing.

Neural networks are especially applicable for unstructured data, such as pictures, language, etc. or when there is an especially complex data problem. E.g., for the payment transaction data, there is a data problem which is about the transactions of a customer. One data set is the sequence of different payments. Each single transaction in a transaction history is one data set as a whole and additionally, is also interdependent among each other, meaning that one transaction is on condition of a previous transaction. Applying predictive analytics here, it is important to consider the temporal interdependence between the individual payments for which especially neural networks are useful as they have the architecture to also learn such interdependencies. We are still researching this topic and plan to integrate it at some point.

For structured data, you can use both neural networks and classical approaches. For non-linear approaches, there is the problem of how to make predictions explainable. Nevertheless, there are first open-source solutions which solve this problem, such as “shap” in Python, and make tree models or neural networks explainable. This solution shows the relationship between inputs and outputs of a model with the architectonical structure within the black box of such a model not being relevant. It can be applied for all sorts of algorithms, including classical models, such as different tree models as for example gradient boosted trees. It kind of solves the black box issue connected to advanced models. Thus, we can e.g., explain for what reasons a customer is or is not affine to a certain product which is especially useful for consultation meeting where it enables our customer (the saving banks) to understand the reasoning behind the outputs of models.

Also, the use cases in the credit scoring are mainly based on gradient boosted trees. However, in a slightly different approach compared to the sales-process due to regulatory reasons.

The problem with neural networks in the credit scoring is that the regulations and the credit scoring itself are not yet at this point of development that it could be applied in practice. Basically, these models could be made explainable. However, this can only be done on the level of the applied features which are the initial input of the model. It cannot be explained what implicitly happens to new features within such a neural network. This is the point where it gets complicated with regard to its explainability: It cannot fully be explained how these different features interact with each other and how further features are built out of initial features within a neural network.

I: Which additional data could be generated in order to increase the predictability of the respective use cases?

Expert 14: Including external data in our analyses is one of our current main focuses we are working on. External data which do not come from our group organization itself and external data which describe the customer behavior, such as log data from the internet websites or mobile app. This could further be enriched with Google Analytics data.

Furthermore, including regional data, coming from the Bundesamt as an open data source, can be used in order to consider regional differences between different saving banks. This is because customers are not only dependent on their own characteristics but also on the region which they live in. E.g., customers' income might be the same, but it might state something totally different about their income-level depending on the region they are working in.

Analyzing Schufa data could give insights into contractual relationships of the customers which might not be observable in the transaction data.

Expert 15: With regard to Schufa, a bank also needs the agreement of the customer to receive this data. Furthermore, Schufa gives information about what other credits or credit cards the customer has. However, it does not give information about the counterparty of such relationships.

I: Compared to other measures or technologies, how do you assess the potential of predictive analytics in the private mortgage financing process?

Expert 14: One other main topic is the data and system integration. Meaning, how can we integrate our offerings with other key systems in the banks so that they can interact with each other. It is a challenge to put all this into one system without having several system-breaks within it.

Furthermore, the overall topic of Big Data is very relevant for us. It is not only about having vast amounts of data but rather about exploiting a range of different data sources so that a 360-degree view on a customer and his/her behavior is possible which can be useful throughout the full customer journey.

E.g., the customer master data are very nice to have and analyze, however, it is only a historical view on the customer. Big Data enables to exploit further data sources and make them usable.

With regard to predictive analytics, it is not always about predicting certain conditions or future events. It can also be about being a bit more explorative. This means that the technology would be used to get a better understanding of the customer and his behavior. Consequently, this would be a focus on descriptive analytics. This is also what we are currently already doing. As I mentioned before, we analyze transaction data to get a better understanding of the customer's life situation which is, up to this point, rather descriptive than predictive.

Appendix 23: Use Cases with Less Than Three Mentions

In the process step *advice*, four use cases were mentioned by fewer than three interviewees. The use case “recommending property” was mentioned by two experts who both see a short-term potential in it. It goes beyond the advice for individualized loan structures and deals with advising the customer with regard to what kind of property would suit him/her. This is especially possible if a bank is a real estate agent itself or cooperates with such. E11 mentions that this use case is not only relevant for the case that a customer is willed to buy an own property but does not have one in mind, but it could especially come into play, if a customer already got the mortgage financing approval of a bank but in the meantime, his/her preferred property is sold to another interested party. Based on the available historic portfolio of product completions and properties sold, PA tools could assist in finding another mortgage which is best suited for the characteristics of this customer.

Furthermore, “Predicting the most suitable customer approach” is another use case in the *advice* process step and is currently applied by one interviewed firm. Here, E14 mentions the application of a reinforcement learning model in order to predict via which channel a customer wants to be approached for a desired product. This can, for example, include an initial approach via phone or e-mail. After all, such tools learn how to strategically act in all kind of situations in an optimal way in order to maximize the expected profit.

“Steering the employee capacity utilization” is a further use case in the *advice* process step which is seen with a short-term potential by one interviewee. E8 sees the potential of the same use case in the process step *contract* as well in a slightly deviated form. The use case is based on the fact that advising properly advising customers takes a lot of time and binds resources. While currently, the supply of customer advisors is estimated based on past experiences and rule of thumb, PA models can support in specifying the exact need for customer advisors over the months of a year.

The next use case “adjust prices seasonally” in the *advice* process step, spins the previous use case even further and is seen with a medium- to long-term potential. While the previous use case approaches the supply-demand topic from the supply side, this use case approaches it from the demand side and suggests that based on a model’s predictions on the demand for private mortgage financings over the course of a year, the prices of the loans could be adjusted in order to straighten out peaks and troughs in the expected demand. E8 further mentions that this could go hand in hand with the “steering of employee capacity utilization” use case or fully compensate it.

The *application* process step contains two further use cases with less than three mentions. The use case “predicting collateral development” is mentioned by two interviewees, one seeing a short-term and one seeing a medium- to long-term potential in it. Based on ML or PA tools, a valuation of properties can be conducted. E9 emphasizes that this information can be drawn upon in order to determine the current collateral value of the property and also has the potential to determine the expected collateral development over the years.

E12 mentioned that his firm currently applies a further use case in the *application* process step, dealing with the “allocation of data for faster evaluations”. He highlights that PA models can introduce a structure into the available data and assign it accordingly to the respective documents. If all required data is available and in place, capital surplus calculations, etc. can be conducted and consequently, the application process can be accelerated. Furthermore, this also enables an analysis of which product characteristics a customer can actually afford.

In the process step *contract*, only one expert sees a use case with short-term potential. It deals with “steering employee capacity utilization” through the application of PA and is supposed to function in the same way as the same use case in the process step *advice*. The use case is based on the fact that processing customer documents in the application process takes a lot of time and ties up resources. Currently, the deployment of employees processing private mortgage financing documents over the months of a year are based on the demand experience for the same time periods of the previous years. Thus, E8 mentions that PA models could specify the upcoming need for document processors at certain times of a year with more precision.

The *refinancing* process step contains the use case to “prevent churn”, is currently applied by one firm and is seen with short-term potential by another one. While E5 highlights that their PA model is able to predict the churn of customers based on collected psychometric data, E8 mentions the difficulty of precisely predicting the potential churn of customers due to the non-availability of necessary data. Such data can range from a changed turnover behavior and a bank’s own service developments to competitors’ price adjustments and the place of a customer’s friends’ accounts.

Lastly, E9 mentions a use case to “propose a new property” in the *refinancing* process step and sees it with medium- to long-term potential. This use case is based on the initial motivation to buy a property in the first place. For example, if it was bought due to family growth, a model could consider that the children move out around 20 years later and the property owners could potentially seek a new property in a more suitable size.

Text of Certification

I hereby confirm that the Work Project presented by me has been prepared independently, using no other sources, resources and other aids than those mentioned. All parts – literally or by their meaning – taken from published or non-published sources are credited as such. The Work Project in its current or similar form has never been submitted as a graded assignment.

31 December 2020