NOVA IMS Information Management School

MGI

Mestrado em Gestão de Informação
Master Program in Information Management

# Recognition and classification:

## The use of computer vision in the retail industry.

Pedro Miguel Cabrita Duarte

Project work presented as a partial requisite to obtain the Master Degree in Information Management with specialization in Knowledge Management and Business intelligence

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

**NOVA Information Management School**
**Instituto Superior de Estatística e Gestão de Informação**
Universidade Nova de Lisboa

# RECOGNITION AND CLASSIFICATION: THE USE OF COMPUTER VISION IN THE RETAIL INDUSTRY

by

Pedro Miguel Cabrita Duarte

Project Work presented as a partial requirement for obtaining the Master’s degree in Information Management with specialization in Knowledge Management and Business intelligence

**Advisor:** Mauro Castelli

September 2019

# ABSTRACT

Automatic recognition of text and classification of it, using image processing techniques such as optical character recognition and machine learning, are indicating new ways of capturing information on fast-moving consumer goods. Such systems can play an important role in market research processes and operations, in being more efficient and agile. The necessity is to create a system that is able to extract all text available on the packaging and quickly arrange it into attributes. The goal of this investigation is to use a combination of optical character recognition and machine learning to achieve a satisfactory level of efficiency and quality. In order for such a system to be introduced to the organization, it needs to be faster and more effective than currents process. One of the advantages of using such a system is the independence of the human factor, which leads to a higher probability of error.

# KEY-WORDS

Market Research, Coding, Retailer, Computer Vision, Optical Character Recognition, Neural Networks

# INDEX

1. Introduction ........................................ 1
2. Literature review ........................................ 4
3. Methodology ........................................ 7
   3.1. Conceptual model ........................................ 7
   3.2. Research Hypothesis ........................................ 9
   3.3. Research Method ........................................ 9
   3.4. Sample ........................................ 10
4. Results ........................................ 14
   4.1. Optical Character Recognition ........................................ 14
   4.2. Data Preparation ........................................ 17
   4.3. Machine Learning Classification ........................................ 19
      4.3.1 Brand Owner International ........................................ 21
      4.3.2 Brand 1 ........................................ 24
      4.3.3 Global Chocolate Type ........................................ 26
5. Discussion ........................................ 28
6. Conclusion ........................................ 29
7. Bibliography ........................................ 31

# IMAGE CONTENTS


Figure 1 - Two *Toblerone* products, with the same weight and type of chocolate, but different packaging. .......... 1
Figure 2 – Example Receipt of a Retailer .......... 1
Figure 3 – Illustrative Coding of Products .......... 2
Figure 4 - Computer Vision and Optical Character Recognition example in car license plates. 4
Figure 5 - Two identical products, with different lettering, unlike logos and different positions of characters. .......... 4
Figure 6 - Workflow of Google's Vision API for Image Search .......... 5
Figure 7 – Schematic overview of our proposed system. .......... 7
Figure 8 – Example of Google's OCR on a Cadbury Heroes Party Bag .......... 7
Figure 9 – Workflow of Google's AutoML .......... 8
Figure 10 – Combination of examples of different picture angles available. .......... 10
Figure 11 – Python script build to run OCR on the sample .......... 14
Figure 12 – Three examples of product images fed to the script, Thornton's Irresistibles on the left, Mars Milk Chocolate Egg in the middle, and Lindt Lindor on the right. .......... 15
Figure 13 – Cadbury Mini Eggs two picture angles, one with a 360º view of the packaging as seen on the left, and a second one with just the front packaging on the right. .......... 16
Figure 14 – Google AutoML Tables Schema tab .......... 17
Figure 15 – Google AutoML Tables Analyze tab .......... 18
Figure 16 – Schematic of how AutoML works .......... 19
Figure 17 – Google AutoML Tables Train Tab .......... 20
Figure 18 – Model Results for Brand Owner International, where precision and recall are generated using a score threshold of 0.5 .......... 21
Figure 19 – Brand Owner International results comparison between a threshold of 0.5 (on the right side) and a threshold of 0.8 (on the left side) .......... 22
Figure 20 - Model Results for Brand 1, where precision and recall are generated using a score threshold of 0.5 .......... 24
Figure 21 – Brand 1 results comparison between a threshold of 0.5 (on the right side) and a threshold of 0.8 (on the left side) .......... 25
Figure 22 - Model Results for Global Chocolate Type, where precision and recall are generated using a score threshold of 0.5 .......... 26
Figure 23 – Confusion matrix for Global Chocolate Type .......... 26
Figure 24 – Global Chocolate Type results comparison between a threshold of 0.5 (on the right side) and a threshold of 0.8 (on the left side) .......... 27

Figure 25 – Nutrition Facts on the back of a product ............................................................ 29
Figure 26 – Block Identification of products ......................................................................... 30

# LIST OF TABLES

Table 1 – Sample collected across the various modules........................................................ 10
Table 2 – 28 Variables harvested, with its definitions and two examples.............................. 12
Table 3 – Results summary of OCR Detection & Extraction.................................................... 15
Table 4 – Example result of predicted brand owner international with the score, where Ferrero has the highest predicted score ......................................................................... 23

# LIST OF ACRONYMS AND ABBREVIATIONS

| | |
|---|---|
| **API** | Application Programming Interface |
| **AUTOML** | Auto Machine Learning |
| **CSV** | Comma Separated Values |
| **FMCG** | Fast Moving Consumer Goods |
| **OCR** | Optical Character Recognition |

# 1. INTRODUCTION

The *Fast-Moving Consumer Goods* (**FMCG**) players resort to market research companies to acquire data regarding their own sales, alongside with their competitors' information, within their playground environment. The latter obtain this data through partnerships with retailers and or consumer panels. Manufacturers rely on this data to assess several metrics, such as sales, brand loyalty, distribution of products, and share of the market. Knowledge of the state of the market is crucial for their day-to-day business, in order to stay relevant, assess opportunities and threats. Although, most of the times, retailers send this type of information without any harmonization or classification, meaning, as they are displayed within their stores.

Naturally, each retailer has their own descriptions and their own method of seeing the same product. International companies also have adopted such methodology as they adapt to each market or time they sell in, as seen in Figure 1. Additionality, products are displayed and advertised according to each countries culture and laws. Therefore, any qualitative and quantitative analysis made on this data is useless, when compared to another country or region, as the data is neither normalized nor compared to the same standards. This is where market research companies come in. It's up to them to classify and recognize merchandise across the entire world.


Figure 1 - Two *Toblerone* products, with the same weight and type of chocolate, but different packaging.

Companies, such as Nielsen, often recur to two ways to classify and recognize products. The first one is by using the descriptions sent by retailers or by collaborating with consumers which then provide receipts from their purchases (see Figure 2). Such descriptions can be as long as a whole paragraph or as short as two words. The second approach is for a field associate to go to each store of retailer partner and photograph, from most angles, the products that the organization has received in the past. Afterwards, the data is harvested by each picture and coded considering definitions and consistent guidelines of the product category.


TESCO
BICESTER 2 0845 677 9063

| | £ |
|---|---|
| FRESH MILK | 0.81 |
| S/BERRY CONSV | 0.96 |
| VIC PLUM CON | 1.65 |
| B/BERRY CONSVE | 1.38 |
| ISB LOAF | 0.70 |
| ORG AVOCADO'S | 1.69 |
| BLUEBERRIES | 1.49 |
| CARROTS 1KG | 0.59 |
| ORG SWT PEPPRS | 1.69 |
| ONION SHALLOTS | 0.63 |
| AUBERGINES | |
| 0.290 kg @ £2.99/ kg | 0.87 |
| GRAPE W.SDLESS | |
| 0.595 kg @ £2.48/ kg | 1.48 |

Figure 2 – Example Receipt of a Retailer

Retailers and manufacturers must then be able to retain all the information displayed on their product packaging, and competitors. Consequently, the market research organization must be apt to answer the important of these elements to consumer perception and how they affect sales. However, before, they need to harmonize and normalize the data before conducting any study. It's crucial that this process is fast, but at the same time with extreme quality.

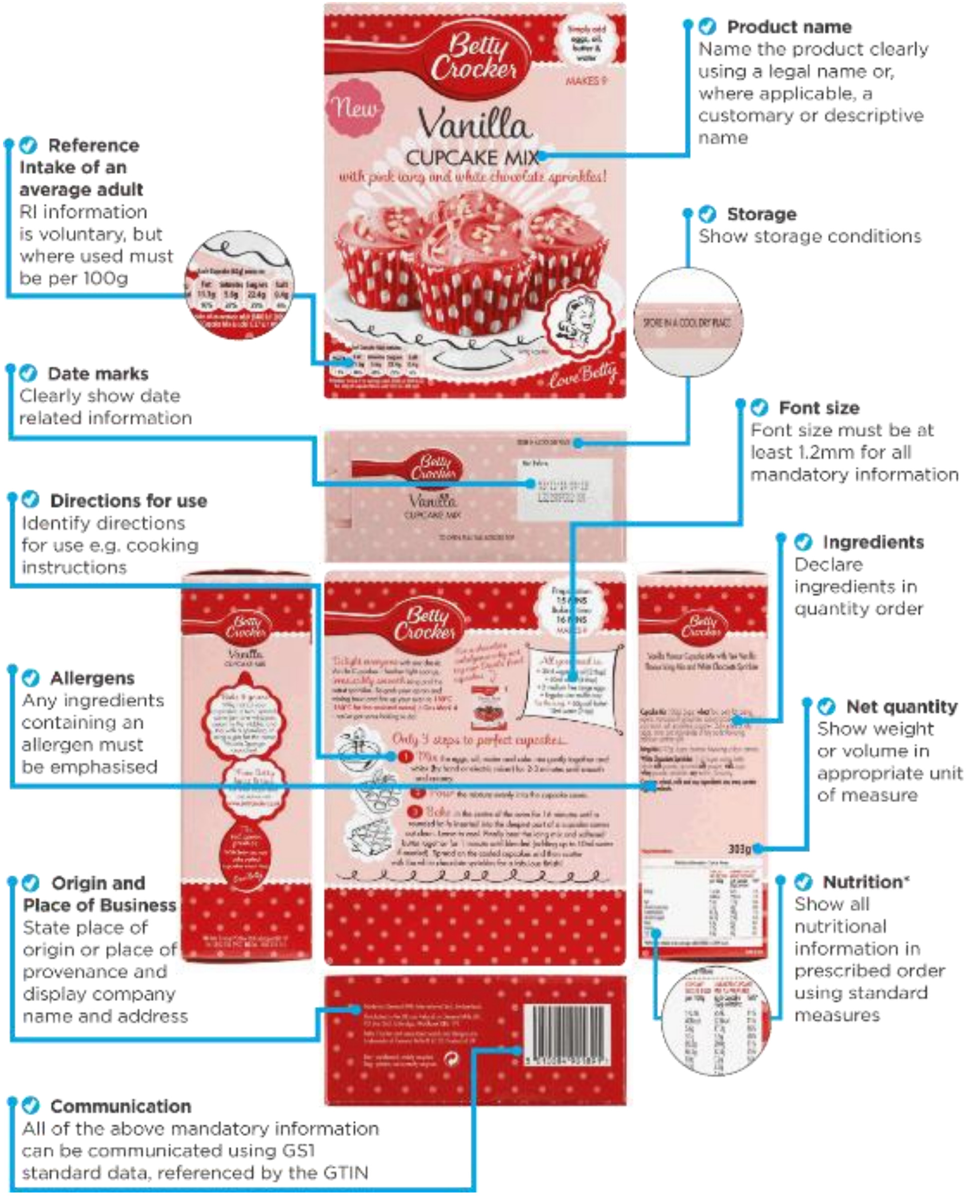



Figure 3 – Illustrative Coding of Products

However, circumstances have changed with the increased adoption of online sale channels in the FMCG industry. For example, with the introduction of Amazon into most households, the variety of products has been amplified exponentially. On average, three hundred thousand new products are introduced to the consumer, every month worldwide. On top of that, more than two million existing products suffer alterations either by evolution or innovation (Worton, 2017). Consumers have also changed their shopping habits. Instead of just looking at price, for example, coffee consumers pay more importance to flavour (Loo, et al., 2015), which makes it even more important to have a consistent way to analyse sales with a standard view. Dealing with a high number of new products and new consumer patterns requires developments of the technologies associated with such processes.

At the same pace, other organizations are also looking to take advantage of recent results of neural networks and deep convolutional networks in aiding visual recognitions. In 2015 C-Discount, a French e-commerce retailer, challenged the data science community to predict the correct category of its products and attributes. Out of the five winners, the contributions were able to predict, on average, the correct category of 66-68% products, whilst using linear classifiers (Jiao, Goutorbe, Grauer, Cornec, & Jakubowicz, 2017). In 2017 C-Discount held once again another challenge aimed at building an image categorization system based on their owned supplied data, where the final results were able to correctly classify close to 80% on the test set (Bellétoile, 2019).

Taking into consideration the state of the FMCG market and the recent developments in neural networks, we are proposing a system that will correctly recognize and classify each product, in their own attributes and categories. We aim to use a combination of the two ways that market research companies use to classify and recognize products: a description, and a picture.

## 2. LITERATURE REVIEW

Computer vision exists and its uses never stop to amaze the human being. This technology has been used in several industries and cases, such as banking, to identify and classify if the signature on a check is genuine. For example, text localization algorithms have been proven to correctly capture 99% of characters in printed documents (Lin, 2002). In another instance, which relates similarly the problem in question, is the use of neural networks alongside with Optical Character Recognition (**OCR**) in car license plates identification (Fahmy, 1994), where a self-organizing neural network was designed and trained to recognize the characters in license plates.


Figure 4 - Computer Vision and Optical Character Recognition example in car license plates.

However, the requirement and the complexity of the problem here at hand are different. Whereas a car's license plate is throughout its country of origin, mostly standard, with the same background, or with slight variations, FMCG products are not. On top of this, in the practicality of the example shown in Figure 4, there is a finite number of combinations of alphanumeric strings. There are many factors that contribute to these major differences such as, but not limited to, non-uniform backgrounds and designs, different fonts, text positioning, diverse languages and even non-identical logos, as seen in Figure 5. Taking this into consideration, we can assert that the methodology cannot single-handedly rely on only character recognition.


Figure 5 - Two identical products, with different lettering, unlike logos and different positions of characters.

An end-to-end deep neural network supported by OCR has been applied to the FMCG industry, by automatically verifying if a product has expired, by using the use-by dates in package photos, yielding an accuracy of close to 90% (Ribeiro, et al., 2018). The use by dates is one of the attributes, that vary from manufacturer to manufacturers, such as the font or even the placement taking into consideration the type of product. Drawing inspiration from such, we idealize that the methodology would require a neural network of sorts to support the data harvested by OCR.

Automated Machine Learning (**AutoML**), has become one of the go-to concepts when talking about machine learning and neural networks, as it challenges norms by building models without human assistance and with limited computational budgets. (Yao, et al., 2019). This is derived from the features they provide, such as automated feature engineering, model selection and algorithm selection. As importantly, the introduction of automation to machine learning aims to systematize the time spent on data preparation and ingestion, on tasks for example as column type detection, task detection, leakage detection and handling of skewed data and or missing values (Lee, 2018).

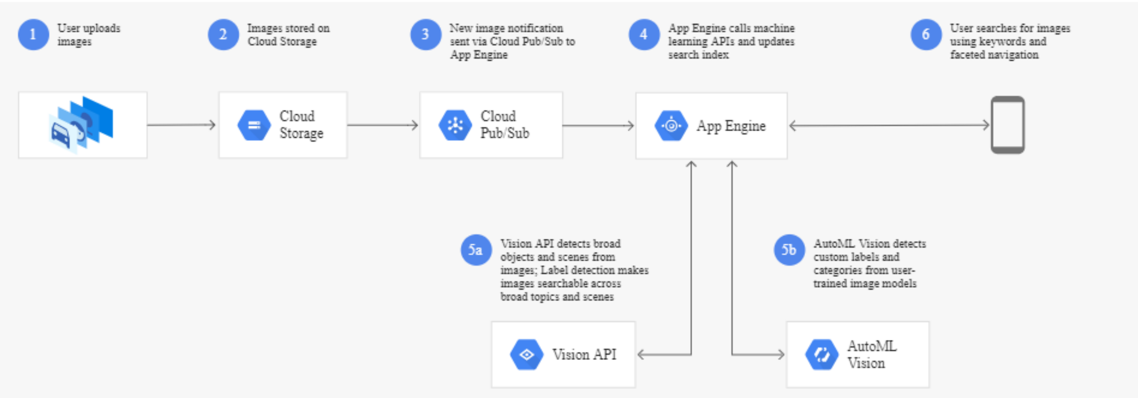



Figure 6 - Workflow of Google's Vision API for Image Search

Nielsen has several partnerships with technology companies, with one of them being Google. Cloud Vision's Application Programming Interface (**API**) has many uses as it enables developers to understand the content of an image by encapsulating powerful machine learning models (AI & Machine Learning Products, 2019). Apart from this, it allows users to detect labels, web detection, OCR, logo detection, and product search. Equally important is the ability to be used with any operating system. For instance, other OCR systems such as Tesseract (Tesseract, 2019), ABBYY Fine Reader (Abbyy, 2019) or Transym (Transym, 2019), are only available on a handful of systems. Furthermore, the latter two are not available as service (Tafti, et al., 2016), hence it would not be viable for scalability for all the companies' business. Additionally, being a web service and with the unique ability to have API's, it enables the organization, in the future, to link between existing systems Google's Visions services.

Google has also their own service of AutoML, called Google Cloud AutoML, which separates itself into three pillars: Sight, Language and Structured Data (Cloud AutoML, 2019). By using these products, they allow users to build models with the capability of transferring learnings and neural structure search technology alongside with natural language processing and translation with an advantage of automated data preparation and ingestion. (Lee, 2018). Simultaneously, the services described above by Google are integrated between each other, making an even step further with automation and linking between each task easier.

In view of the requirements, we will be using Google’s Cloud service, such as Vision API and AutoML to understand the efficiency of these two existing services on the problem at hand.

# 3. METHODOLOGY

## 3.1. CONCEPTUAL MODEL

To tackle the concept we have at hand, we propose a methodology that can be described as a two-layer system. Objectively, the first layer will recognize the text present on the packaging and extract it into a string, via image as an input. Afterwards, the output obtained will be used as an input to the second layer, where we will classify each string collected into attributes with a level of confidence for each instance, as seen in Figure 7.

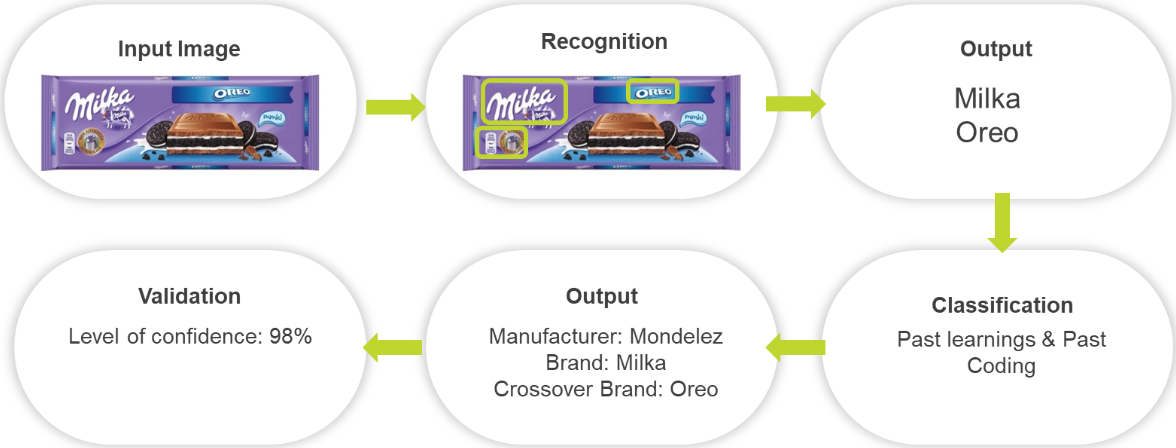



Figure 7 – Schematic overview of our proposed system.

On the first layer, we will be using Google's Detect Text API, Vision API and OCR API in order to detect and extract text from images. It is important to reference that we will be only analyzing data that is considered physical, meaning, presented on the packaging and that is clear and objective. In other words, we will not be focused on trying to determine, for instance, if the packaging is luxury due to the fact is made of metal. Such attributes are focused on a consumer perspective, that is mainly subjective. These three APIS can all be used on Google's user interface, as seen in Figure 8, or by recurring to a programming language. Currently, it is supported by Node.JS, Python and Go (Compute Products, 2019). For this project, we will rely on Python's supported classes in order to perform the first layer of our methodology. Afterwards, still, on the same stage, we will be analyzing the results to understand what conclusions can be made. Should any data points produce no result, these will not be assessed and will not be included on the second stage.


Figure 8 – Example of Google's OCR on a Cadbury Heroes Party Bag

On the second layer, we will be constructing neural networks that will classify the text gather earlier and classify it based on past coding of that exact picture. We will be using Google's AutoML Tables, which is meant to create structured machine learning models based on structured data (Cloud AutoML, 2019). By being fully integrated with other Google services, it allows to store, train and generate predictions in one single dataset. AutoML automatically understands which model architecture fits best for the data its fed, such as Linear, Feedforward deep neural network, Gradient Boosted Decision tree, as examples. On top of this, as seen before, we will be able to leverage Google's AutoML capabilities to our project.

Lastly, with the outputs of the two layers we will investigate whether the system is able to fully produce an output with a high level of quality but at the same time with the appropriate response time.

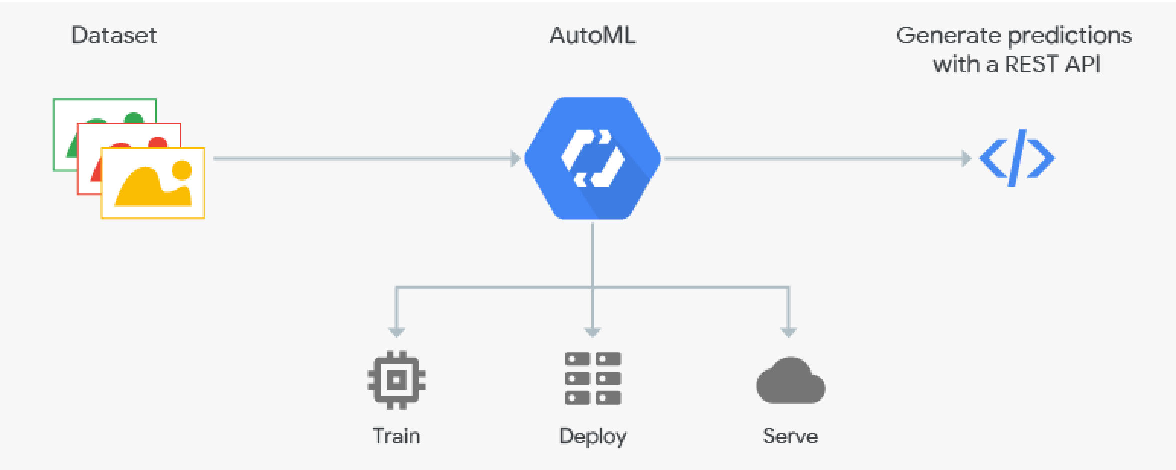



Figure 9 – Workflow of Google's AutoML

In order to support the first stage, and consequently, the second stage, we will be using a sample of data, and pictures, associated with two countries for a specific category of products, as it is described later, on section 3.4.

### 3.2. RESEARCH HYPOTHESIS

Like any organization, Nielsen needs to be agile when facing new products within the industry, although always maintaining quality as one of the main goals. Yet, the organization understands that within the digital transformation which the technology atmosphere faces, they can benefit from the introduction of new processes or methodologies within their business.

With this work, we intend to understand what the organization can gain from adopting computer vision, with an emphasis on automated processes, within their current processes. In order to measure this, we can divide the hypothesis taking into considerations the two layers, recognition and classification.

- **Recognition**
    a. Is the proposed methodology able to identify and capture physical text within each image?
- **Classification**
    a. Is the proposed methodology able to classify the extracted text into attributes?

Apart from this, we also aim to understand:

1. Is the proposed new system more efficient than the current one?
    a. What is the time, end-to-end, of a product being coded by the system introduced here versus the previous method?
    b. What is the quality that this system can purpose?
    c. What is the behaviour of this work when introducing new data?

### 3.3. RESEARCH METHOD

A research method is applied considering a different kind of problems. We used empirical type research in order to answer our hypothesis. Merriam-Webster defines the word empiric as one who relies on experience, dating the first use of the word in the $17^{th}$ century (Merriam-Webster.com, 2019). Moreover, an Assessment Study will be employed in order to test our established baselines and ranges.

Our main goal is to apply this empiric methodology to understand the effects of existing working models on this research problem. Although, the algorithms we intend to use throughout are almost impossible to look within, and the ones there are, yield no real value. Accordingly, we will investigate gathering as much data as we can and embody the hypothesis in the algorithms.

Under these circumstances, we will be collecting data from the company's internal systems, using real examples and real data. Next, we will validate the data to understand if we can reduce if found, any inconsistencies. After the model is deployed, it's important to understand the results of the two stages identified in the conceptual model: recognition and classification. With the help of previous data, we will investigate the results and the statistical value it produces to understand if the current working algorithms can be applied and sustain our hypothesis to the identified problem or if a new, or adapted algorithm, is needed.

### 3.4. SAMPLE

To conduct the methodology described throughout the document, we are using a sample provided by the company. The sample can be split into two pillars: the first we have text data that describes products; the second it's the picture of the referred product. We have identified one category of products that show the highest level of diversity within the market: Chocolate. Additionally, inside the chocolate category, we refer to nine sub-categories, which we will refer to "modules". Besides this, we have also selected two countries to analyze, Great Britain and Portugal.

As can be seen in the below table, from roughly 59000 unique products, from various manufacturers, only 13% of the items have an attributed picture which has a good quality resolution.

| MODULE | NUMBER OF ITEMS FOUND | NUMBER OF PICTURES EXTRACTED | NUMBER OF PICTURES OF ITEMS EXTRACTED |
|---|---|---|---|
| NOVELTIES EGGS BITESIZE | 2040 | 348 | 758 |
| NOVELTIES EGGS HANDHELD | 6146 | 1278 | 3495 |
| NOVELTIES MOULDED BITESIZE | 320 | 54 | 201 |
| NOVELTIES MOULDED HANDHELD | 7367 | 1090 | 2464 |
| ASSORTED INCLUDING MULTIPACK HANDHELD | 518 | 86 | 361 |
| ASSORTMENT BITESIZE | 6020 | 1040 | 3039 |
| SINGLE VARIETY BITESIZE | 17877 | 1673 | 6084 |
| SINGLE VARIETY HANDHELD | 18778 | 2167 | 8677 |
| VARIABLE WEIGHT | 85 | 0 | 0 |
| TOTAL | 59151 | 7736 | 25079 |

Table 1 – Sample collected across the various modules.

Each product has multiple pictures, each from different angles. These unique angles, which can be, for example, the front, back, and sides, or in cases all of them together in one single picture (See Figure 10). The diverse images will help capture physical elements on the packaging that might be found in different locations rather than just the front. In its totality, we have 25 thousand pictures that we can utilize.

Figure 10 – Combination of examples of different picture angles available.

Apart from the pictures, we also harvested 28 different variables that characterized the product. These variables have diverse definitions and are acquired by different data on the product packaging, as can see below (see Table 2).

| CHARACTERISTIC NAME | OPEN/ CLOSED | VARIABLE TYPE | DEFINITION |  |  |
|---|---|---|---|---|---|
| **BRAND OWNER INTERNATIONAL** | OPEN | TEXT | This refers to the Senior Holding Company identified as being the company who can exercise the right to re-brand the product. | MONDELEZ INTERNATIONAL | NESTLE / NESTLE |
| **BRAND 1** | OPEN | TEXT | This refers to the Brand Name which appears on the pack itself and by which the consumer would recognise the product. | CADBURY | ROWNTREES |
| **GLOBAL BRAND EXTENSION** | OPEN | TEXT | The Brand Extension serves to refine the brand coded in Brand where different varieties of that Brand exist. The Brand Extension often defines a specific variant within a range of products under a common Brand name. Consequently, the Brand Extension will usually be found across a narrower range of modules than Brand. | CADBURY DAIRY MILK | ROWNTREES |
| **GLOBAL CANDY PRODUCT TYPE** | CLOSED | TEXT | Indicates, with reference to the product branding, labelling or packaging, the type of candy product. All candy products are split into five categories - Solid, Hollow, Filling, Ingredient and Filling & Ingredient. | FILLING & INGREDIENT | HOLLOW |
| **GLOBAL CHOCOLATE TYPE** | CLOSED | TEXT | Indicates, with reference to the product branding, labelling or packaging, the descriptive term that is used by the product manufacturer to identify a particular type of chocolate. | MILK | MILK |
| **GLOBAL COLLECTION CLAIM** | OPEN | TEXT | Indicates, with reference to the product branding, labelling or packaging design, the descriptive term that is used by the product manufacturer to identify the theme of a product. | NO CLAIM | NO CLAIM |
| **GLOBAL CONSUMER LIFESTAGE CLAIM** | CLOSED | TEXT | Indicates, with reference to the product branding, labelling or packaging, the descriptive term that is used by the product manufacturer to identify the period or stage in the consumer's life during which the product is considered to be suitable for consumption or use. | NO CLAIM | NO CLAIM |
| **GLOBAL CROSSOVER BRAND** | OPEN | TEXT | Crossover branded products show an additional brand on the label in addition to the primary brand. The Crossover Brand is a brand usually present in another category and used by the manufacturer to add a specific benefit/property to the product. | OREO | RANDOMS |
| **GLOBAL FILLING** | OPEN | TEXT | This refers to the filling inside the chocolate as stated on the packaging. | VANILLA MOUSSE | NOT STATED |
| **GLOBAL FLAVOUR** | OPEN | TEXT | Indicates, with reference to the product branding, labelling or packaging, the descriptive term that is used by the product manufacturer to identify the flavour of a product. | BISCUIT & CREAM | NOT STATED |
| **GLOBAL FLAVOUR GROUP** | OPEN | TEXT | Indicates, with reference to the product branding, labelling or packaging, the descriptive term that is used to identify the generic flavour of a product. | BISCUIT & CREAM | NOT STATED |
| **GLOBAL IF WITH COOKING CLAIM** | CLOSED | TEXT | Indicates, with reference to the product branding, labelling or packaging, whether the product makes claim to being intended for cooking purposes. | WITHOUT COOKING CLAIM | WITHOUT COOKING CLAIM |
| **GLOBAL IF WITH DAIRY FREE CLAIM** | CLOSED | TEXT | Indicates, with reference to the product branding, labelling or packaging, whether the product makes claim to be manufactured without any dairy content. The claim found on the packaging is typically Dairy Free. | WITHOUT DAIRY FREE CLAIM | WITHOUT DAIRY FREE CLAIM |
| **GLOBAL IF WITH ORGANIC CLAIM** | CLOSED | TEXT | Indicates, with reference to the product branding, labelling or packaging, whether the product makes claim to the use of organic ingredients or production methods, irrespective of the product's compliance with any mandatory industrial or legislative criteria. | WITHOUT ORGANIC CLAIM | WITHOUT ORGANIC CLAIM |

| CHARACTERISTIC NAME | OPEN/ CLOSED | VARIABLE TYPE | DEFINITION |  |  |
|---|---|---|---|---|---|
| **GLOBAL INGREDIENTS** | OPEN | TEXT | This refers to the ingredient added to the product as stated on the packaging. | BISCUIT | NOT STATED |
| **GLOBAL ITEM IDENTIFIER** | CLOSED | TEXT | Refers to which type of product is it and what is identifies and the consumer can clearly understand. | CHOCOLATE EGG | CHOCOLATE EGG |
| **GLOBAL LEVEL OF SUGAR CLAIM** | CLOSED | TEXT | Indicates, with reference to the product branding, labelling or packaging, the descriptive term that is used by the product manufacturer to identify the relative level of sugar that is contained within the product. | NO CLAIM | NO CLAIM |
| **GLOBAL NUMBER IN MULTIPACK BASE** | OPEN | INTERVAL | Indicates, with reference to the product labelling or packaging, the numerical term that is used by the product manufacturer to identify the number of individual items that are typically contained within a single product, excluding any additional quantity that results from a promotional offer. | 1 | 1 |
| **GLOBAL NUMBER IN MULTIPACK ACTUAL** | OPEN | INTERVAL | Indicates, with reference to the product labelling or packaging, the numerical term that is used by the product manufacturer to identify the number of individual items that are actually contained within a single product, including any additional quantity that results from a promotional offer. | 1 | 1 |
| **GLOBAL OCCASION SEASON** | CLOSED | TEXT | Indicates, with reference to the product branding, labelling or packaging, the descriptive term that is used by the product manufacturer to identify the specific or generic name of the occasion during which it is intended that the product be consumed or used. | EASTER | EASTER |
| **GLOBAL PACKAGING** | CLOSED | TEXT | Indicates, in descriptive terms, the form of packaging that has been used by the product manufacturer to directly enclose the product contents that are to be consumed or used and [normally] to provide the consumer with all necessary branding, labelling and product information. | WRAPPED | BOX |
| **GLOBAL PACKAGING MATERIAL** | CLOSED | TEXT | Indicates, in descriptive terms, the generic name of the primary material that has been used in the manufacture of the product packaging. | FOIL | CARDBOARD |
| **GLOBAL PRODUCT TYPE** | CLOSED | TEXT | This refers to the type of chocolate product. | EGG | EGG |
| **GLOBAL PRODUCT VARIANT CLAIM** | OPEN | TEXT | Indicates, with reference to the product branding, labelling or packaging design, the descriptive term that is used by the product manufacturer to identify the particular variant of a product. | NO CLAIM | NO CLAIM |
| **GLOBAL TOTAL MASS WEIGHT GROSS** | OPEN | INTERVAL | This refers to the total weight of the entire product, including any additional weight due to promotional activity and multipacks. | 31 | 250 |
| **GLOBAL TOTAL PACKS IN MULTIPACK** | OPEN | INTERVAL | This refers to the total number of packs in the entire product, including any additional packs due to promotional activity. | 1 | 1 |
| **GLOBAL WEIGHT VOLUME PER MEASURED UNIT ACTUAL** | OPEN | INTERVAL | Indicates, with reference to the product branding, labelling or packaging, the numerical term that is used by the product manufacturer to identify the actual metric weight or volume of a product, including that of any additional quantity that results from a promotional offer. | 31 | 250 |
| **GLOBAL WEIGHT VOLUME PER MEASURED UNIT BASE** | OPEN | INTERVAL | Indicates, with reference to the product branding, labelling or packaging, the numerical term that is used by the product manufacturer to identify the typical metric weight or volume of a product, excluding that of any additional quantity that results from a promotional offer. | 31 | 250 |

Table 2 – 28 Variables harvested, with its definitions and two examples

At the end of the two layers of the system, ideally, the text produced by the OCR will populate the characteristics as seen in Table 2. Although, this will not be achieved by this work. Google's APIs are not free to use (Auto ML Tables Price, 2019), and there is a cost by model training and deployment. Taking this into consideration, we will choose three attributes as described in Section 4.3.

# 4. RESULTS

## 4.1. Optical Character Recognition

The first step to our proposed system is to feed the images from our sample, in order to understand if Google's existing tools can extract text, by using the OCR API, from the pictures. A Python script was created to call Google Vision, as can be seen in Figure 11, as it is one of the programming languages that the API works with.  Each image was run once, one after the other, extracting one label which contained all the information it was able to extract from the picture. This label was then stored back onto a data dump for later analysis, alongside with the image name and file path. At this stage, this is still a linear process, as the second layer of the proposed methodology does not have yet the necessary data to train the neural network. In other words, after each picture is run, it is not going forward immediately to a prediction of a value. Instead, all text is extracted and then stored back. Once the second layer is built, trained and deployed, it is possible to automate this process, relying once again on capabilities of Google Cloud, where each service can be integrated.

```python
import io
import os
import xlrd
import xlwt
import time
#start timer to assess how much time it takes
start = time.time()
#counter to assess progress
countertimes = 0
#import API Key needed to use Google's Services
os.environ["GOOGLE_APPLICATION_CREDENTIALS"] = "app/apikey-final.json"
# Imports the Google Cloud client library
from google.cloud import vision
from google.cloud.vision import types

#Opens and stores in memory the sample, which has the file path, file name.
loc =(r"C:\Users\Pedro\Desktop\Cloud Vision\Data_Dump.xlsx")
wb_in = xlrd.open_workbook(loc)
sheet_name = wb_in.sheet_names()[0]
ws_in = wb_in.sheet_by_name(sheet_name)
nrows = ws_in.nrows

#Creates a new excel file to store our output, with the File Path, File Name and the output, named as Label 1
wb_out = xlwt.Workbook()
ws_out = wb_out.add_sheet(sheet_name, cell_overwrite_ok=True)
ws_out.write(0,0,"File Path")
ws_out.write(0,1,"File Name")
ws_out.write(0,2,"Label 1")

#Loop for as much images found in the sample
for x in range (1,nrows):
#Read workbook image
    imagename = ws_in.cell_value(x, 0)
    imagefilename = ws_in.cell_value(x, 1)
#Instantiates a client
    client = vision.ImageAnnotatorClient()
#The name of the image file to annotate
    file_name = os.path.join(
    os.path.dirname(__file__),
    imagename)
#Loads the image into memory
    with io.open(file_name, 'rb') as image_file:
        content = image_file.read()
        image = types.Image(content=content)

#Performs label detection on the image file
        response = client.text_detection(image=image)
        texts = response.text_annotations

#Prints all Labels within output Excel File
        ws_out.write(0,0,"File Path")
        ws_out.write(0,1,"File Name")
        y=2
        ws_out.write(x, 0, imagename)
        ws_out.write(x, 1, imagefilename)
        for text in texts:
            ws_out.write(x, 2, text.description)
            break
    countertimes += 1
#Print Progress Report for User Knowledge
    print("Progress Report " + str(countertimes) + " out of " + str(nrows))

#Give your name to the new file
wb_out.save("Output_OCR_Time.xls")
#Stops timer
end = time.time()
#Prints alert to the user that the task has been done
print("Done. File has been created for " + str(nrows-1) + " picture(s)")
print(end - start)
```

Figure 11 – Python script build to run OCR on the sample

| MODULE | NUMBER OF PICTURES WHERE OCR RETRIEVED TEXT | NUMBER OF PICTURES WHERE OCR RETRIEVED NO TEXT | TOTAL TIME IN DURATION HH:MM:SS |
|---|---|---|---|
| ASSORTED INCLUDING MULTIPACK HANDHELD | 361 | 25 | 00:25:58 |
| SINGLE VARIETY BITESIZE | 6084 | 415 | 05:30:14 |
| NOVELTIES EGGS HANDHELD | 3495 | 611 | 03:38:39 |
| NOVELTIES EGGS BITESIZE | 758 | 73 | 00:39:19 |
| ASSORTMENT BITESIZE | 3039 | 221 | 03:36:29 |
| NOVELTIES MOULDED HANDHELD | 2464 | 309 | 02:29:59 |
| NOVELTIES MOULDED BITESIZE | 201 | 15 | 00:13:39 |
| SINGLE VARIETY HANDHELD | 8677 | 168 | 06:05:11 |
| TOTAL | 23242 | 1837 | 23:29:28 |

Table 3 – Results summary of OCR Detection & Extraction

Out of the 25079 inputs fed to the script, 1837 failed to produce any output from Google's optical character recognition API. In other words, 7,32% of the sample for this project were unsuccessful. Despite this, the API took 3,7 seconds, on average, to upload an image from a local machine, detect and extract text on Google's cloud server and feedback once again to the local machine running the script.

For example, as can be seen in Figure 12 below, we have three examples that were chosen due to particularities they present. As a first illustration, the first product on the left in Figure 12, we have Thornton's Irresistibles which is described by having a full black packaging with pink and gold lettering coupled with different fonts and sizes. However, the API was successful in perceiving the text and overcoming this task, resulting in the following output: *"Thorntons Irrestistibles GOOEY CARAMEL, GOOEY CARAMEL WRAPPED IN SUMPTUOUS MILK CHOCOLATE NET WT.200G E 7oz".*


Figure 12 – Three examples of product images fed to the script, Thornton's Irresistibles on the left, Mars Milk Chocolate Egg in the middle, and Lindt Lindor on the right.

Secondly, Mars Milk Chocolate Egg has different shades of colouring throughout the box. Additionally, it also has partial text, which is not completely seen on the pack, with two Mars bars being covered by a black and white triangle. Yet again, we were able to extract all the relevant text, excluding our concern: "*MARS MILK CHOCOLATE EGG FULL SIZE MARS BARS 2"*. Although, as can be seen from the end result, we were not able to decipher the order of the text, with the numerical text 2 being put at the end, instead of in the middle.

Lastly, we have the Lindt Lindor. Admittedly this type of packaging is not seen often by the typical consumer, although it is frequently used by Cash & Carry stores, or in other words, business to business stores. On the one hand, the Lindt Lindor, in comparison to the latter two products has fewer colours. On the other hand, it has text scattered across the card box. Still, the API was able to extract nearly the entirety of existing text: *"LINDT LINDOR GOLD SELECTION BOX 10 x e 500g PRODUCT CODE 598222 THIS WAY UP THIS WAY UP LINDL LINDOR LINDT & SPRUNGLI (UK) LTD, TOP FLOOR, 4 NEW SQUARE, FELTHAM, MIDDLESEX, TW148HA".* For instance, due to the tab overlaying the Lindor word, the output was compromised. In addition to this, we also see that the font used in Lindt unquestionably impacted the result, with swapping a T for an L.

For the most part, we were able to extract data successfully from our sample. Although, as analysed before, there were occasions where we were unsuccessful in mining any text. These were caused by images where a 360 view was provided to the script. In result, the API was unable to effectively detect any text, since the image was compressed to have multiple angles of the packaging resulting in lower font size. In comparison, when the front packaging is the only viewpoint visible in the input, Google Vision was able to harvest the following information: *"Cadbury 1 Large Egg Bag Mini Eggs Cadbury"*.


Figure 13 – Cadbury Mini Eggs two picture angles, one with a 360º view of the packaging as seen on the left, and a second one with just the front packaging on the right.

### 4.2. DATA PREPARATION

Data preparation and clean-up were needed to comply with the second layer of the system. AutoML Tables by Google supports data importation through either comma-separated-values, also known as CSV, or by using BigQuery (Auto ML Tables Data Types, 2009), another cloud service provided by Google which enables the analysis of big data sets. Since we were already using Excel-based format on the first layer, it was opted to upload our data extract via CSV. Although, before importing the data into a storage location, we flagged any data point which was not active, or in other words, data points that are no longer in distribution in the market, and therefore Nielsen no longer supports coding activities. These were automatically included on a test run on a later stage. Identically, the same methodology was applied to any data point which had an attributed coded as not determined, hence the organization did not have enough information to code a specific attribute. In short, these 78 data points were also automatically inserted into the test data set.

Afterwards, the sample was brought to AutoML tables. From there, the first stage is determining the schema. This is where AutoML, as seen before, shows one of the objectives it aims to overcome, automating steps that were before manual. Instantly, all columns were given a variable type, as for example, categorical, numeric, text. The system was also able to understand if a certain characterization of the data should be nullable or not, as can be seen in Figure 14.

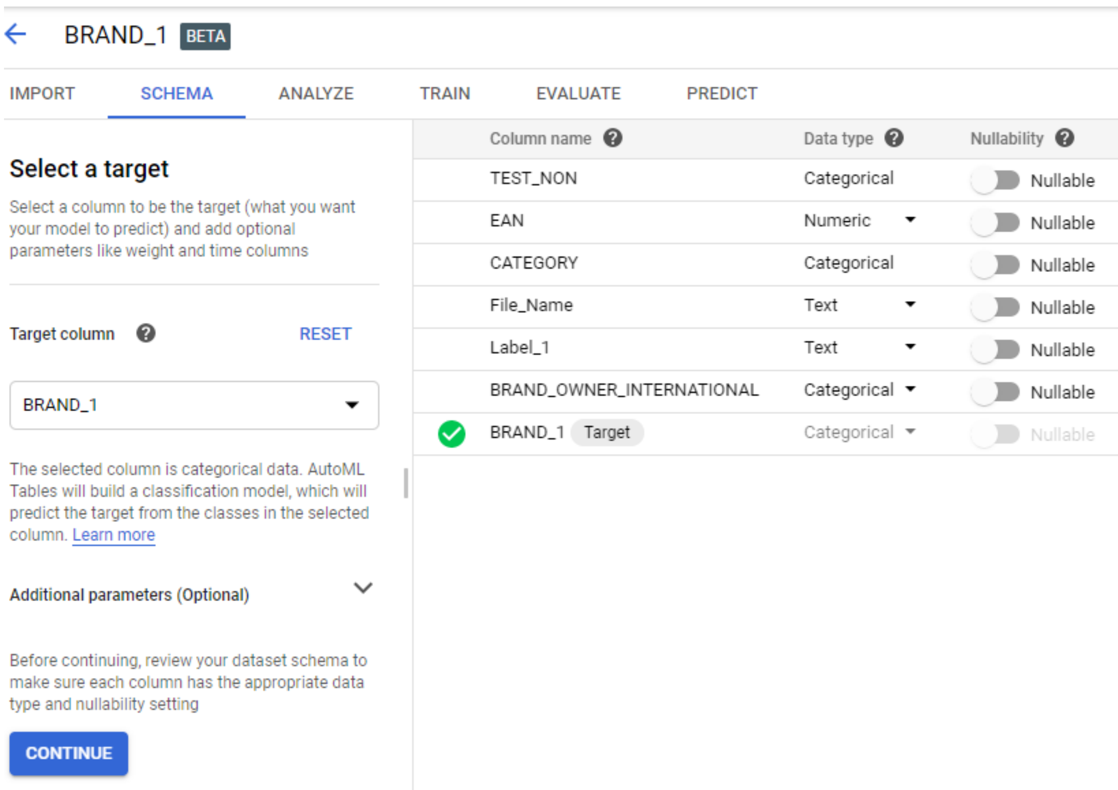
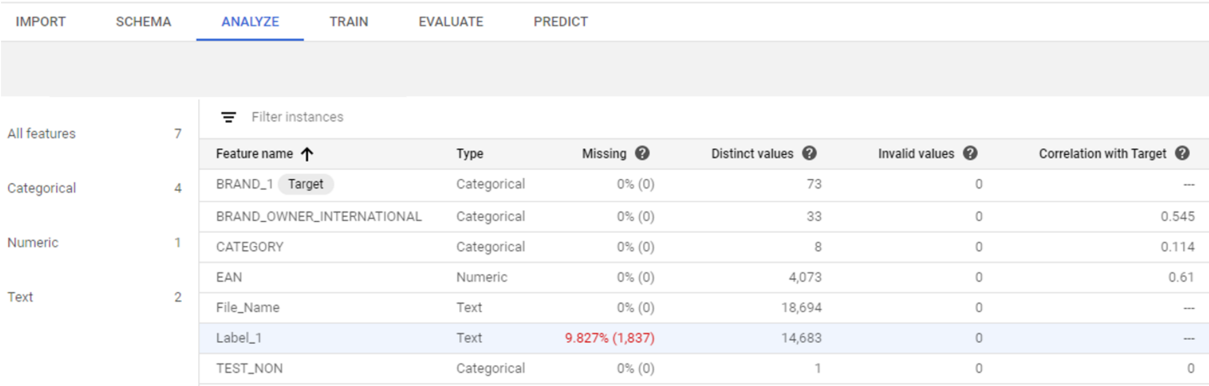



Figure 14 – Google AutoML Tables Schema tab

After selecting the target column, where we select the values which the model should predict, we move on to the next tab, the analyse tab. In here, the user is prompted to analyse stats in regard to the features selected on the previous step.



IMPORT SCHEMA ANALYZE TRAIN EVALUATE PREDICT

Filter instances

| | |
|---|---|
| All features | 7 |
| Categorical | 4 |
| Numeric | 1 |
| Text | 2 |

| Feature name | Type | Missing | Distinct values | Invalid values | Correlation with Target |
|---|---|---|---|---|---|
| BRAND_1 Target | Categorical | 0% (0) | 73 | 0 | — |
| BRAND_OWNER_INTERNATIONAL | Categorical | 0% (0) | 33 | 0 | 0.545 |
| CATEGORY | Categorical | 0% (0) | 8 | 0 | 0.114 |
| EAN | Numeric | 0% (0) | 4,073 | 0 | 0.61 |
| File_Name | Text | 0% (0) | 18,694 | 0 | — |
| Label_1 | Text | 9.827% (1,837) | 14,683 | 0 | — |
| TEST_NON | Categorical | 0% (0) | 1 | 0 | 0 |

Figure 15 – Google AutoML Tables Analyze tab

Unquestionably, all the units that did not have any result on the first layer of the system were removed, as they would provide null correspondence to the machine learning algorithm, leading to a reduction of 1837 fewer pictures to the sample. This was achieved by Google's AutoML capabilities, as seen in Figure 15, as it automatically treats missing values (Lee, 2018). Taking into consideration this is a classification problem, there is a minimum of 50 rows by each distinct value you have on your target class (AutoML Tables Preparing your training data, 2019). Consequently, each target variable we assess will have different sample sizes, in order to be compliant with this requirement. This will be highlighted throughout section 4.3. AutoML also recommends using spaces when using text, as it tokenizes text strings. In this case, the OCR API had already done this job for us, so no action was required.

Finally, AutoML automatically normalizes and buckets numeric features, creates one-hot encoding and embeddings for categorical features, performing basic processing for text features. Therefore, no activity was done for this type of exercise of data preparation.

## 4.3. Machine Learning Classification

Lastly, the final layer of the proposed system is used to run all the data collected as described in the previous chapters through the AutoML platform by Google.

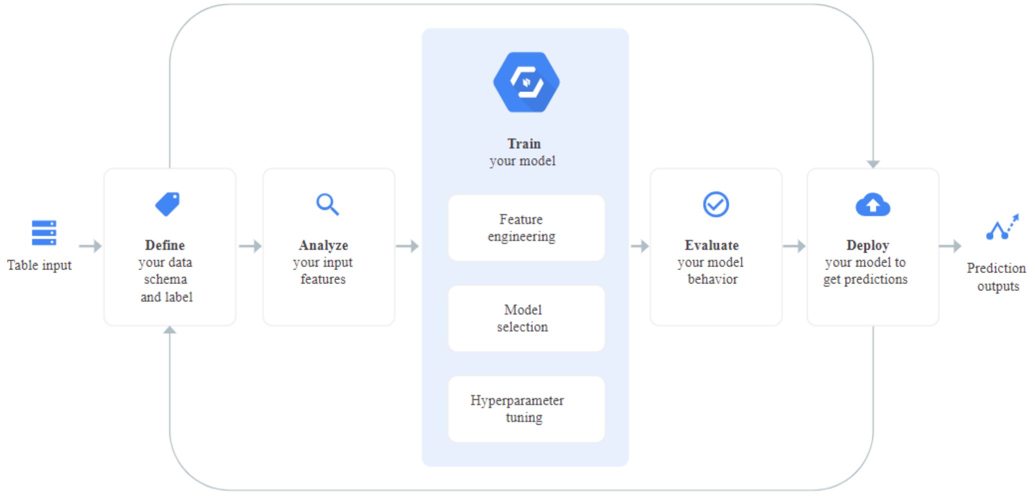



Figure 16 – Schematic of how AutoML works.

Although, as mentioned before, this service is not entirely free. Consequently, we have chosen a set of three characteristics to run the prototype against. Firstly, we have Brand Owner International, as this is one of the more basic variables to introduce the module, even though most of the times is not present on the packaging, but instead being derived from Brand 1. Secondly, Brand 1 as this describes, as mentioned in table 2, the name of which consumers typically recognize immediately. Lastly and thirdly, we have Global Chocolate Type, a characteristic which groups types of chocolate by a pre-defined list of values where we will aim to understand if the model is able to infer a correlation rule between what the manufacturer adversities on the product packaging to what Nielsen defines as a type of chocolate.

As mentioned before, AutoML takes each dataset and starts training for multiple model architectures at the same time, which enables to determine the best model architecture. By default, AutoML Tables randomly selects 80% of the data rows for training, 10% for validation and 10% for testing. Taking into consideration that our data set does not change over time we chose to continue with Google's recommendation and default values. Moreover, the user also has the ability to give different weight to each row. As a default, Google trains each model with the same weight for every single input row, which we used in order to not deviate any of the training exercises.

After the data preparation, defining the schema and analysing the input features, we move to the bigger stage, which is to train the model. Once on this tab, the user has the ability to give a name to the model, enter a training budget, select which features are to be used for training and early stopping objective. In regard to the budget, the user can select between 1 to 72 hours of node hours in training the model. Although, the user is also able to select, early stopping, which effectively ends the training when Google's service detects no more improvements can be made. Google recommends between 1-3 hours if your data source has less than 100,000 rows, (AutoMl Tables Managing Models, 2019). However, in order to test the early stopping mechanism, and even though we have less than 100,00 rows, we have deployed each model with 7 node hours as a budget. Moreover, this service automatically defines the model optimization objective. Taking into consideration the attributes we are trying to classify are multi-class problems, AutoML Tables automatically recommended to optimize by Log Loss, in order to keep the prediction probabilities as accurate as possible.

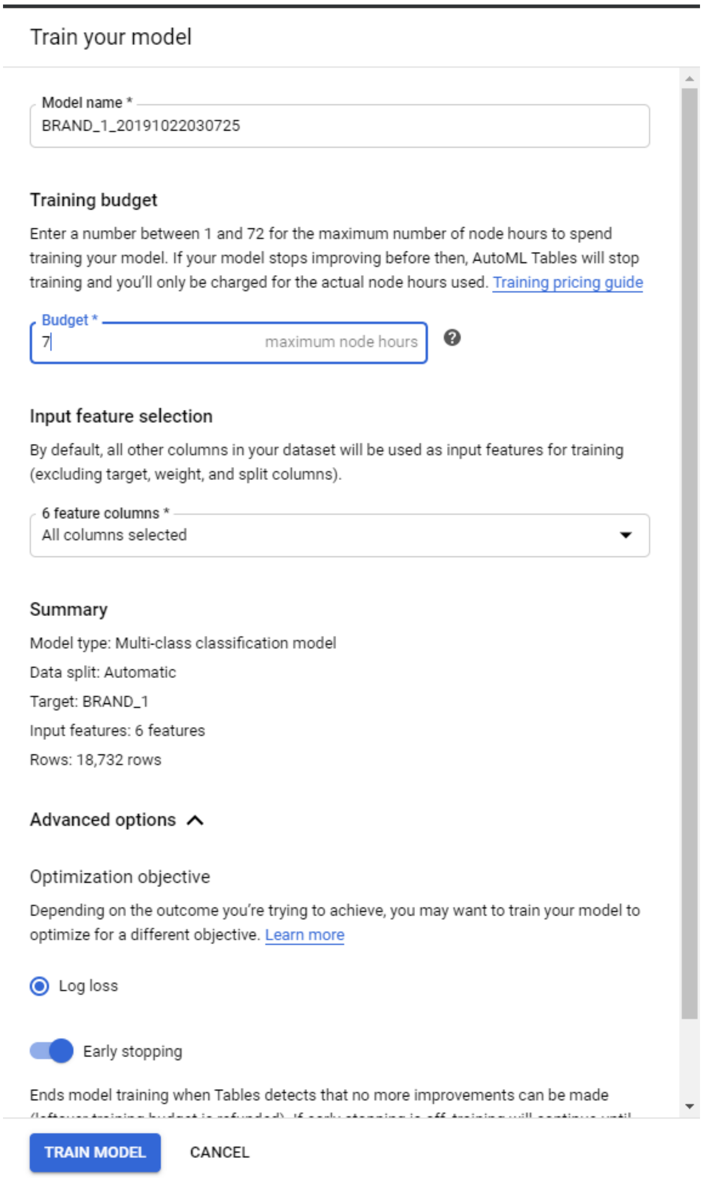



Figure 17 – Google AutoML Tables Train Tab

This methodology will be used throughout the three different models, one per each chosen attribute, described from this point forward.

### 4.3.1 Brand Owner International

The sample for this target value was 20074 data points, which excludes 1837 data points from non-OCR text, 78 coming out of non-active/not determined items and finally 1253 rows due to the fact of having less than 50 entries for a specific target value.

Considering the main purpose of this prototype is to understand if by using, solely, the text extracted on the first layer if we can establish a target variable, the only column used as input would be the label from OCR. Moreover, it also avoids target leakage, which is when we use predictive information that is not available when we will be asking for a prediction.

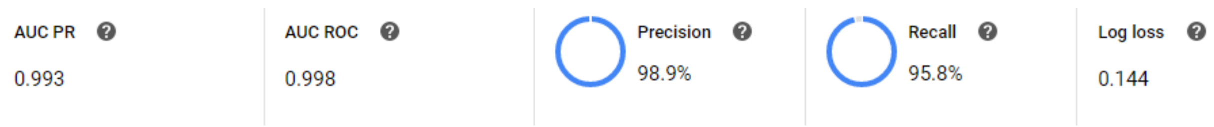



Figure 18 – Model Results for Brand Owner International, where precision and recall are generated using a score threshold of 0.5

Overall, the model took 7 hours until the best result, as assessed by the layer, was reached. Globally the model performs well, with the precision-recall curve, or AUC PR, having a high value with conjunction with the receiver operating characteristic curve, or AUC ROC. At the same time, we have a log loss of 0.144, or in other words, a low cross-entropy between the model predictions and the label values. However, these are results with a score threshold of 0.5, so it is important to understand how to model behaves if the score-threshold increases.

As can be seen in Figure 18, once we intensify the score threshold, we see a decrease of the F1 score, which calculates the harmonic mean of the precision and recall. Taking into consideration that we have an imbalance in the observations for each class, the F1 score reduces from 0.907 from a 0.5 score threshold to a 0.881 F1 where the threshold is 0.8. At the same time, we see a growth of precision of the model in contrast to a decrease of true positive rate and false positive rate, where we understand that 71 out of 14287 entries are false.

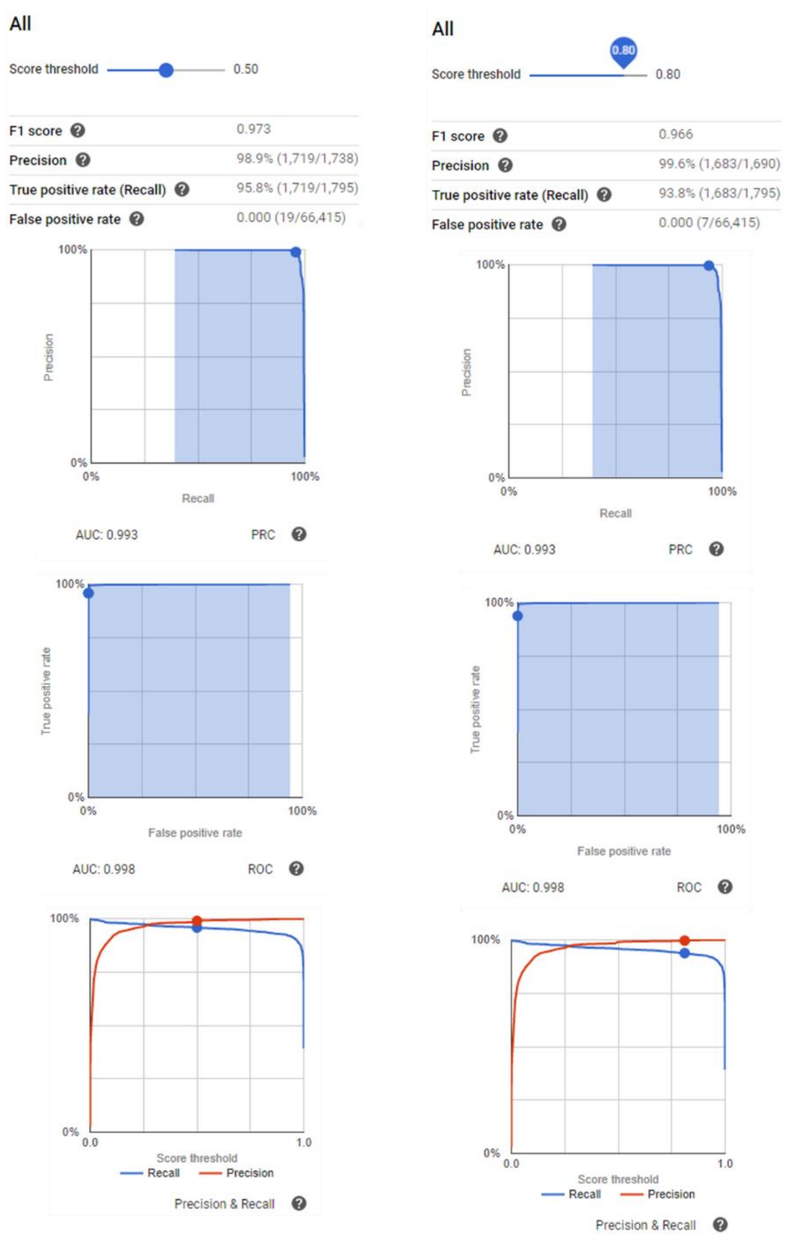



Figure 19 – Brand Owner International results comparison between a threshold of 0.5 (on the right side) and a threshold of 0.8 (on the left side)

As an example, we can see on Table 4 a particular product by Thorntons which was chosen to understand if, even if not available on the packaging, would the model be able to determine that Thornton’s is a brand from Ferrero taken into consideration the data fed through the layer. Google AutoML Tables determined that the predicted brand owner international would be Ferrero, with the highest score of close to one, on a scale of 0 to 1. Google AutoML Tables enables the analysis of every single row tested by using BigQuery.

| BRAND OWNER INTERNATIONAL | INPUT IMAGE | OCR LABEL | PREDICTED BRAND OWNER INTERNATIONAL SCORE | OUTPUT BRAND OWNER INTERNATIONAL |
|---|---|---|---|---|
| FERRERO |  | Thorntons MILK CHOCOLATE EGG WITH 6 CHOCOLATES ehorntor TOFFEE, FUDGE & CARAMEL COLLECTION A DELIGHTFUL COLLECTION OF TOFFEE, FUDGE AND CARAMELS SMOTHERED IN MILK AND DARK CHOCOLATE | 0.9999546409 | FERRERO |
| | | | 0.0000133517 | TESCO |
| | | | 0.0000092653 | WALKERS CHOCOLATES |
| | | | 0.0000042346 | JOHN LEWIS PARTNERSHIP |
| | | | 0.0000024428 | DIAGEO |
| | | | 0.0000023426 | WALMART |
| | | | 0.0000014138 | COLIAN |
| | | | 0.0000013703 | BUTLERS CHOCOLATES |
| | | | 0.0000011645 | ELIZABETH SHAW |
| | | | 0.0000011369 | ZERTUS |

Table 4 – Example result of predicted brand owner international with the score, where Ferrero has the highest predicted score

Summarizing, Brand Owner International has a great performance on all the rankings observed, with low false-positive rates, a high true positive rate alongside with ideal precision.

### 4.3.2 BRAND 1

The sample for this target value was 18732 data points, which excludes 1837 data points from non-OCR text, 78 coming out of non-active/not determined items and finally 2595 rows due to the fact of having less than 50 entries for a specific target value.

In this case, we can consider Brand Owner International as an input, as we can establish that this second layer will be linear, meaning, it will only try to predict Brand 1 only and only if it already has predicted Brand Owner International.  Similar to the Brand Owner International model, Brand 1 also took 7 hours to perform the multi-class model.

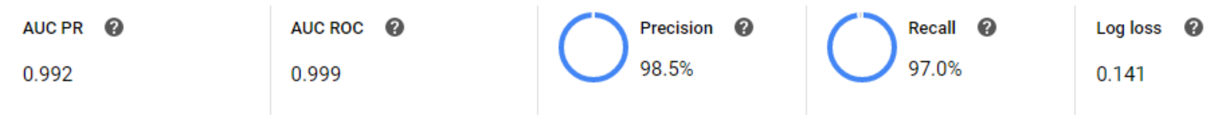



Figure 20 - Model Results for Brand 1, where precision and recall are generated using a score threshold of 0.5

In total, and in conjunction with the Brand Owner International results, Google took 7 hours to create and train the data, until it reached the optimum point. Altogether the accomplishes the task at hand, having a high AUC PR alongside with AUC ROC. Whilst looking and relating the log loss to the one from Brand Owner International, we can see an improved result of 0.141, translating into a better probability of converging to the actual label on the training data.  Whilst observing how the model reacts when the score threshold is amplified to 0.8, as can be seen in Figure 20, we can see a slight reduction on F1 scores, which is normal as we have enlarged the confidence level of when the result is to be taken. Alongside, we detect a better false-positive rate, which increases the quality of the model, from 26 to 14 out of 125,712.

Looking towards the recall when the confidence level is set at 0.8, 95% of the testing data was label correctly. On the whole, the model is appropriate for the purpose of Brand 1 and clarifying the usefulness of this layer.

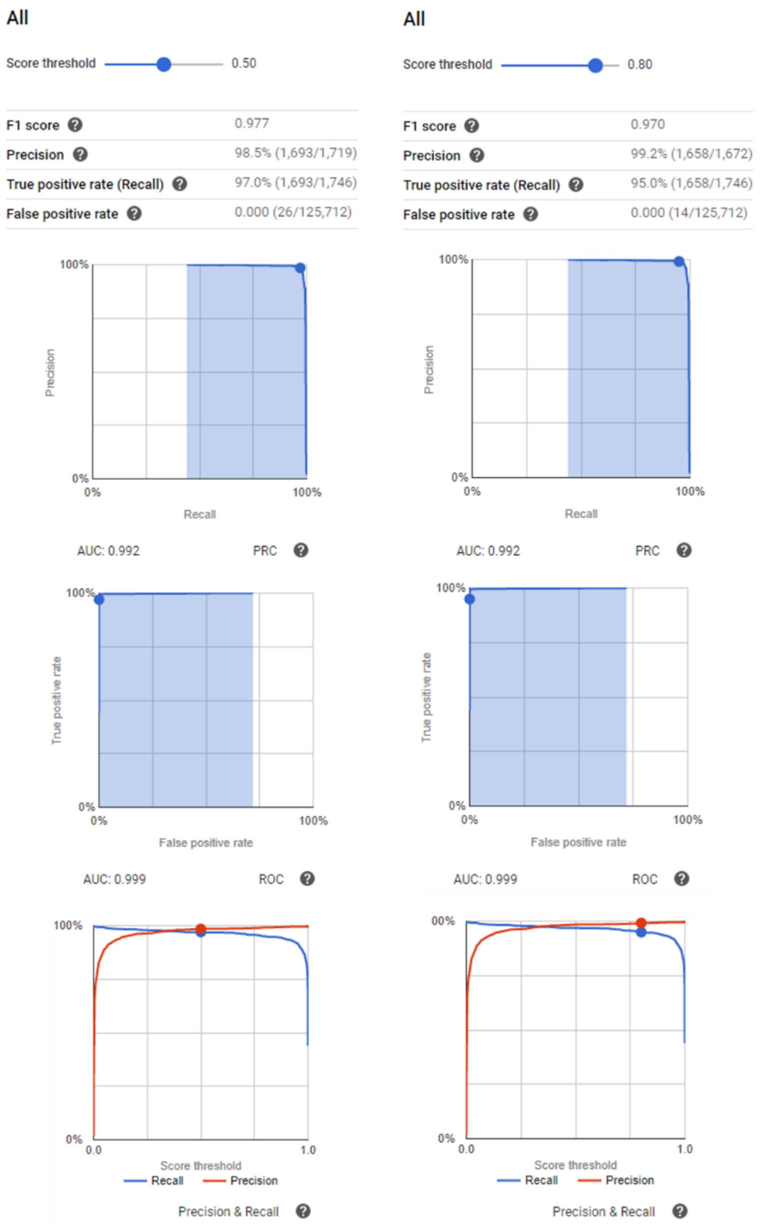



Figure 21 – Brand 1 results comparison between a threshold of 0.5 (on the right side) and a threshold of 0.8 (on the left side)

### 4.3.3 Global Chocolate Type

The sample for this target value was 21327 data points, which excludes 1837 data points from non-OCR text, 78 coming out of non-active/not determined items. In this situation, we had at least 50 raw inputs for each distinct value, so no further reduction of the sample was needed.

In this case, we can consider Brand Owner International alongside with Brand 1 as an input, in resemblance of Brand 1 methodology, which will help us understand if there is any correlation between a brand and chocolate types.

| AUC PR | AUC ROC | Precision | Recall | Log loss |
|---|---|---|---|---|
| 0.964 | 0.992 | 91.9% | 89.5% | 0.312 |

Figure 22 - Model Results for Global Chocolate Type, where precision and recall are generated using a score threshold of 0.5

In this situation, the marks provided by the performance indicators were not as similar to the one reached before on Brand Owner International and Brand 1. For the most part, it achieved good scores, but with a recall and precision percentage lower than the ones observed with the other characteristics. Granted, this type of characteristic has a degree of subjectivity, particularly when we diverge between languages. In fact, the model has particular lower performance when dealing with multi-case situations, as illustrated with the confusion matrix in Figure 23. A trend is seen on values when a combination of values appears, such as DARK & MIL, DARK & WHITE and MILK & WHITE, where only 14%,33% and 31% respectively where predicted accurately.

| True labels \ Predicted labels | ASSORTED | DARK | DARK & MILK | DARK & MILK & WHITE | DARK & WHITE | MILK | MILK & WHITE | WHITE |
|---|---|---|---|---|---|---|---|---|
| ASSORTED | 71% | 2% | - | 2% | - | 23% | - | 1% |
| DARK | 0% | 88% | - | 1% | - | 10% | - | 0% |
| DARK & MILK | 5% | 24% | 14% | 24% | - | 33% | - | - |
| DARK & MILK & WHITE | 3% | 4% | - | 77% | - | 13% | 1% | 1% |
| DARK & WHITE | - | 22% | - | 33% | 33% | 11% | - | - |
| MILK | 0% | 1% | - | 1% | - | 97% | - | 0% |
| MILK & WHITE | - | - | - | 8% | - | 58% | 31% | 4% |
| WHITE | 1% | 4% | - | 2% | - | 17% | - | 77% |

Figure 23 – Confusion matrix for Global Chocolate Type

However, Global Chocolate Type still yields high results when we compare a confidence level of 0.5 to 0.8. For instance, we see a precision score of 95.9% coupled with an F1 score of 0.881 on a confidence level of 0.8. Moreover, the false positive rate is decreased slightly by increasing the confidence level, lowering from 0.11 to 0.005.

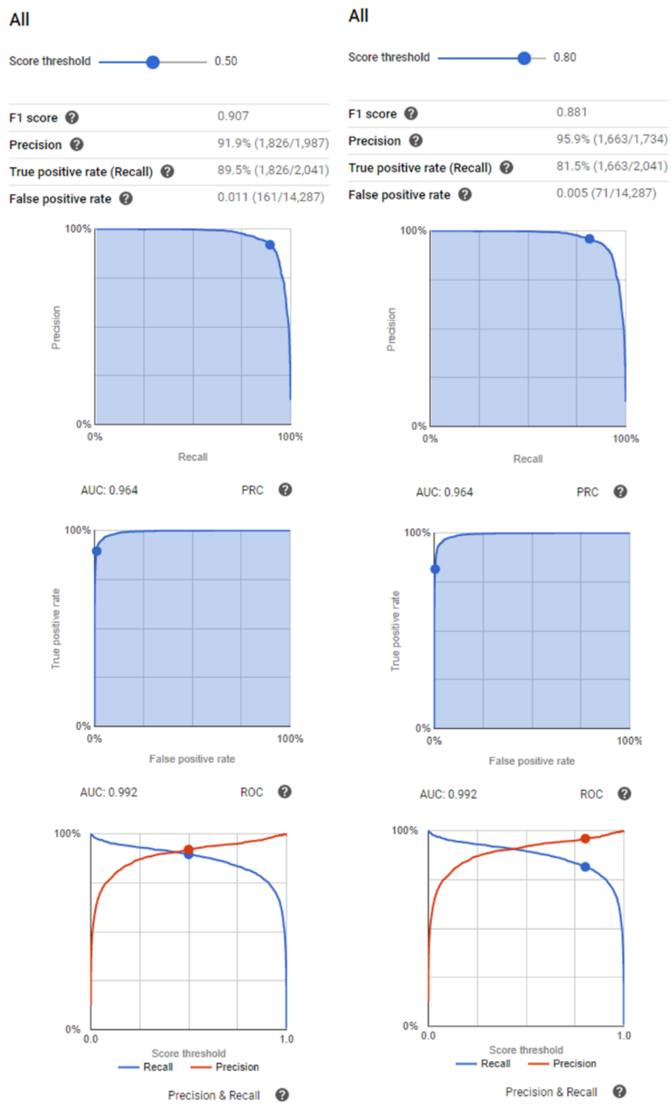



Figure 24 – Global Chocolate Type results comparison between a threshold of 0.5 (on the right side) and a threshold of 0.8 (on the left side)

Unquestionably, this variable does not perform as well as the others. In any case, it still has immense results with an excellent AUC PR and AUC ROC, even though the log loss is superior as expected from the other characteristics.

# 5. DISCUSSION

In line with the hypothesis, the results indicate, generally, that the proposed methodology can identify and capture physical text within each image. However, there are cases where it fails. These can be overcome by only feeding images where only one angle of the product is seen. Additionally, in this layer, not all data provided by the API was taken into consideration. The optical character recognition by Google also provides as an output the font size of the text extracted with the correspondent weight and height of where it was found. For instance, Brand 1 is often associated with the largest word on the packaging. Introduction of such variables could potentially help the second layer of the model to achieve better results.

On our first layer, the OCR API was able to generate a single description of all the text displayed on an image, for 92.68% of the sample provided. Moreover, it took, on average 3.7 seconds to create such an output, taking already in consideration the output saving into an excel based format. Overall, all the entries where we were unable to create any description were cases where more than one angle was provided in a single input. Whereas we before thought this would prove to be beneficial, it proved to be where the API had some drawbacks.

On the second layer, it was proved that the text provided by the API was easily transformed into attributes, using AutoML by Google. Once test data was introduced to the system, the behaviour of the system did not diverge from previous learning of the model, reaching high results and accomplishing low false-positive results.

Taking into consideration the time on an end-to-end product to run through our proposed model, it takes on average 3,7 seconds to pass through the OCR layer and an extra 19 seconds for each characteristic. Considering that we have 28 attributes within the chocolate category, it would take, usually, 9.3 minutes per product, in a study conducted in 43 countries. In contrast, the current model deployed by Nielsen, which is human-based, takes 14 minutes to code new items, meaning, the item cannot be matched to any other previous coded item, and 6 minutes to code formerly seen items.

When relating to quality, they have similar aspects, with any model generated here having at least a precision of 90%, whereas human coding activities guaranteeing 95% quality.

In relation to Google's AutoML abilities, we can see a big effort to reduce and eliminate some human tasks in comparison to traditional machine learning engines. Data ingestion, task detection, handling of missing values, model and algorithm selected were all automated, with no need for user interaction. Moreover, the time spent on data preparation and jobs such as column type and attribute classification was reduced, with validation only step required from the end-user.

Summarizing, this end to end system is faster when compared to Nielsen's human-based approach when coding new, never seen, products. Our proposed workflow takes on average, 4.7 minutes less than the organization's. When comparing to products that already exist, or already have been coded in another country, our process is slightly slower, with a difference of 3.3 minutes more on average. When taking into single pieces, we can see that the big driver of this time is the classification layer, as the recognition is almost instantaneous, with a mean time of 3.7 seconds, whereas the second layer takes 19.79 seconds per attribute.

## 6. CONCLUSION

Recognition, classification and machine learning are themes that are gaining relevance throughout our society and are embedding into organizations that wish to compete in their market. In the research market environment is important to get information quickly and with quality. The model proposed aimed to introduce this type of technology to the process of coding products and at the same time improve current efforts and efficiency.

On a sample of 25079 pictures, the first layer of the system was able to retrieve text for over 23242 images, with a success rate of 92.68%, with an average of 3.7 seconds per product. Taking into consideration the instances where no text was successfully obtained, it was identified that the number of angles of pictures in the same input can prejudice Google's OCR API. Although, this is generally easy to overcome. Although, this API can produce an output other relevant information, such as text size, font and positioning in the packaging. Normally, in the FMCG industry, there is a pattern to designing products, with the Brand Owner International information being the biggest text, the Brand 1 attribute the second-highest and Global Brand Extension being the third. Additionally, Brand Owner International usually sits at the top, whereas Global Brand Extension usually is placed at the bottom. Such information could provide to be useful to feed as an input to the second layer of the system, to also try to understand, how relevant the position on the packaging has on a certain attribute.

One of the other advantages that the recognition layer provides is the ability to automatically generate when fed a picture with a single angle, a string with the complete text available on the image. This will lead to a higher detail of information available for each associate when coding products. At the same time, with this information, the company can also choose to audit variables that they chose not to. As an example, the organization, most of the times, does not collect information in respect to nutrition facts, as it is often on the back of each product, as can be seen in Figure 25. The effort to gather it is superior to basic characteristics such as Brand Owner International. Although, with this API we can reduce that effort.


Figure 25 – Nutrition Facts on the back of a product

Moving to the second layer of the system, three models were created based on three different characteristics. Each one of them took 7 hours to train all available data, and for the most part, produced high results on each single key performance indicator. Overall, we can reach the desired output within 9.7 minutes, for all 28 attributes within this certain category. This threshold is beneficial when comparing to new items, which have never been seen in the market. Taking into consideration the environment of fast-moving consumer goods and the trends in relation to innovation, the introduction of this system proves to reduce the cycle time of coding a product. Moreover,

Nevertheless, should the company introduce such a model, it needs to take into consideration the cost associated with it. Google's AutoML is not free, and it does have costs. The organization will need to assess if the cost supported by this operation is feasible.

In future work, products from other categories apart from chocolate, and with a more diverse and bigger sample should be put to the test, in order to have an overall, more robust performance measure, which can be applied worldwide. Moreover, we have only used multi-class problems in our attributes, whereas we expect a higher performance on numerical attributes, such as weight. Lastly, the introduction of data points in relation to font, size and colour can be a beneficial output coming from our suggested model, as can be seen on Figure 26.


Figure 26 – Block Identification of products

# 7. BIBLIOGRAPHY

*Abbyy*. (2019). Retrieved from http://www.abbyy.com

*AI & Machine Learning Products*. (2019, January). Retrieved from Google Cloud: https://cloud.google.com/vision/

*Auto ML Tables Data Types*. (2009, September). Retrieved from Google Cloud: https://cloud.google.com/automl-tables/docs/data-types

*Auto ML Tables Price*. (2019, September). Retrieved from Google Cloud: https://cloud.google.com/automl-tables/pricing

*AutoMl Tables Managing Models*. (2019, September). Retrieved from Google Cloud: https://cloud.google.com/automl-tables/docs/models

*AutoML Tables Preparing your training data*. (2019, September). Retrieved from Google Cloud: https://cloud.google.com/automl-tables/docs/prepare

Bellétoile, A. (September de 2019). Large e-retailer image dataset for visual search and product classification.

*Cloud AutoML*. (2019, September). Retrieved from Google Cloud: https://cloud.google.com/automl/

*Compute Products*. (2019, September). Retrieved from Google Cloud: https://cloud.google.com/functions/docs/tutorials/ocr#functions-prepare-environment-python

Fahmy, M. M. (1994). *Computer Vision Application To Automatic Number Plate Recognition.*

Jiao, Y., Goutorbe, B., Grauer, C., Cornec, M., & Jakubowicz, J. (Dezembro de 2017). The categorization challenge organized by Cdiscount on datascience.net in 2015: analysis of the released dataset and winning contributions.

Lee, Y. S. (2018). Analysis on Trends of Automated Machine Learning. *International Journal of New Innovations in Engineering and Technology*, 32-35.

Lin, X. (2002). Reliable OCR solution for digital content re-mastering. *SPIE Conference on Document Recognition and Retrieval.* California.

Loo, E. J., Caputo, V., Nayga Jr., R. M., Seo, H.-S., Zhang, B., & Verbeke, W. (2015). Sustainability labels on coffee: Consumer preferences, willingness-to-pay.

*Merriam-Webster.com*. (2019, January 1). Retrieved from Merriam-Webster.com: https://www.merriam-webster.com/dictionary/empiric

Ribeiro, F. d., Gong, L., Calivá, F., Swainson, M., Gudmundsson, K., Yu, M., . . . Kollias, S. (2018). An End-to-End Deep Neural Architecture for Optical Character Verification and Recognition in Retail Food Packaging. *Image Processing, 2018 IEEE International Conference.* Athens, Greece.

Tafti, A. P., Baghaie, A., Assefi, M., Arabnia, H. R., Yu, Z., & Peissig, P. (2016). OCR as a Service: An Experimental Evaluation of Google Docs OCR, Tesseract, ABBYY FineReader, and Transym. In *International Symposium on Visual Computing* (pp. 735-746). Springer, Cham.

*Tesseract*. (2019). Retrieved from https://github.com/tesseract-ocr.

*Transym*. (2019). Retrieved from http://www.transym.com/

Worton, K. (2017, February 22). *Nielsen*. Retrieved May 25, 2018, from Nielsen: http://www.nielsen.com/ie/en/insights/news/2017/perspectives-product-reference-data-the-oil-for-for-the-digital-business-engine.html

Yao, Q., Wang, M., Chen, Y., Dai, W., Hu, Y.-Q., Li, Y.-F., . . . Yu, Y. (January de 2019). Taking the Human out of Learning Applications: A Survey on Automated Machine Learning.