MDSAA

Master Degree Program in
Data Science and Advanced Analytics

Using NLP to prove the Unified theory of acceptance and use of technology (UTAUT2)

Francisco Manuel Carujo Tavares das Neves

Project Work

presented as partial requirement for obtaining the Master Degree Program in Data Science and Advanced Analytics

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

USING NLP TO PROVE THE UNIFIED THEORY OF ACCEPTANCE AND USE OF TECHNOLOGY (UTAUT2)

by

Francisco Manuel Carujo Tavares das Neves

Project Work report presented as partial requirement for obtaining the Master's degree in Data Science and Advanced Analytics

Advisor: Tiago Oliveira

Co Advisor: Sérgio Moro

November 2022

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration. I further declare that I have fully acknowledge the Rules of Conduct and Code of Honor from the NOVA Information Management School.

Francisco Carujo Neves

Lisbon, 16/11/2022

Acknowledgments

I would like to thank Professor Tiago Oliveira and Professor Sérgio Moro for accepting to supervise this thesis and for all the guidance and insight given throughout its development.

I am grateful to my parents and friends for all the support they have given me throughout the course of this work.

RESUMO

Com o presente trabalho pretende-se estudar se é possível aplicar as metodologias de “Natural Text Language Processing” e “Text Mining” a avaliações obtidas através da internet, usando a informação processada e avaliada, com recurso a “sentiment analysis” no modelo “Unified theory of acceptance and use of technology” (UTAUT2).

Refira-se que estudos anteriores focaram-se em usar questionários como fontes de dados para os modelos UTAUT e UTAUT2.

No presente trabalho foram utilizadas técnicas de “Natural Language Processing” para determinar a polaridade dos sentimentos reportados em avaliações on-line extraídas de um fórum de anime popular, designado “MyAnimeList” que também possui informação sobre uma grande variedade de séries e filmes de animação japonesa e a avaliação dos mesmos. Os dados respeitantes às informações pessoais e às avaliações foram extraídos e considerados mediante a sua polaridade, sendo analisados posteriormente utilizando métodos de “structural equation modelling” (SEM) e “partial least squares” (PLS).

Do estudo efetuado descobrimos que, embora o modelo UTAUT2 passe nos critérios do modelo de medição, possui valor explicativo baixo, necessitando de melhorias para ser completamente aplicável.

PALAVRAS CHAVE

UTAUT2, Unified theory of acceptance and use of technology, Sentiment Analysis, Text Mining, Natural Language Processing, Avaliações Online, SEM, PLS

Objetivos de Desenvolvimento Sustentável:

ABSTRACT

This study aims to discover whether it's possible to apply Natural Language Processing and Text Mining to unprocessed raw data available online and use the processed and scored information, using Sentiment Analysis, with the Unified theory of acceptance and use of technology (UTAUT2) model.

Previous studies have focused on using surveys and questionnaires as sources of data for the UTAUT and UTAUT2 model.

This study uses Natural Language Processing techniques to determine the polarity of sentiments found in online reviews extracted from the popular anime forum, "MyAnimeList", a website that also holds information on a multitude of animated shows and movies from Japan and reviews on said media. Data regarding user information and review content was extracted and scored, with it being analysed later using structural equation modelling (SEM) and partial least squares (PLS) methods.

We found that whilst the UTAUT2 model passed the measurement model criteria, it has a low explaining power, therefore meaning it would require further tweaking to be totally usable.

KEYWORDS

UTAUT2, Unified theory of acceptance and use of technology, Sentiment Analysis, Text Mining, Natural Language Processing, Online Reviews, SEM, PLS

Sustainable Development Goals (SGD):

INDEX

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

UTAUT - Unified Theory of Acceptance and Use of Technology

TRA - Theory of Reasoned Action

TAM - Technology Acceptance Model

MM - Motivational Model

TPB - Theory of Planned Behaviour

MPCU - Model of PC Utilization

IDT - Innovation Diffusion Theory

SCT - Social Cognitive Theory

GDPR - General Data Protection Regulation

NLP – Natural Language Processing

AI – Artificial Intelligence

SEM – Structural Equation Model

PLS – Partial Least Squares Regression

AVE - Average Variance Extracted

GoF – Goodness of Fit

SRMR - Standardized Root Mean Square Residual

NFI - Normed Fit Index

1. INTRODUCTION

In recent times, information has become a crucial part of our society. It plays a vital role in decision making, and from simple to key decisions, it aids us with a very good success ratio. However, information is also expensive money and resource-wise, since you need to setup ways to procure and to generate it, whether it is with market research, focus groups or questionnaires. So, what if we could use a resource available to everyone as a source of information? Nowadays, you can find everything on the internet, from educational videos to recipes. Our objective with this work is to see if it is possible to harness the power of the internet to gain insights and information that would aid in key decisions.

To study this possibility, we decided to try and use the Unified Theory of Acceptance and Use of Technology (UTAUT2) model to serve as the base for our analysis and to discover if it is possible to replace questionnaires with an automated model that would gather and analyse useful and meaningful information found on the internet.

In order to test the model, we procured on the internet for databases or review websites to collect data from. However, due to the strict General Data Protection Regulation (GDPR), our search was hindered.

GDPR, put into effect in 2018, "applies to the processing of personal data in the context of the activities of an establishment of a controller or a processor in the Union, regardless of whether the processing takes place in the Union or not."[1]. As such, due to obligations imposed, this meant that most of the websites that were possible candidates to serve as our datasource were cut out, as this regulation views to hide and keep personal information for as little as possible, as seen in the GDPR article 5 [2]. This meant that the only viable option we found was a forum about Japanese animation, which displayed both lengthy reviews and user information.

Animation has come leaps and bounds since the first animated short, "Humorous Phases of Funny Faces", back in 1906 [3]. Through several blockbusters, the industry has been an everlasting presence on its medium, with classics such as Snow White and the Seven Dwarfs, the first full length movie using cel animation, released in 1937 [4], earning over $184 million in box office in the United States [4]. Nowadays, the industry is thriving with blockbusters such as "Minions The Rise of Gru", released in 2022, earning over $369 million in the United States alone, $935 million worldwide, placing it in the top 100 of highest grossing movies, and still not being the highest grossing animation movie, with it being "Frozen 2", with over $1 billion in worldwide box office [6]. This has been an ever-growing industry, reaching global box office revenue of $42.3 billion in 2019 [7], only dropping in 2020 and 2021 due the Covid-19 pandemic.

Japanese animation also has a high impact on the market, although mainly focused on the Japanese market, with movies such as “Demon Slayer the Movie: Mugen Train”, released in 2020, earning over $447 million in worldwide box office [8]. Its market size is expected to grow from $26 billion to $48.3 Billion from 2022 to 2030, with a growth rate of 5.2% [9], thus overcoming the issues related to the Covid Pandemic that led the industry to stop growing after 10 years [10][9], reaching only in 2020 96.5% of the previous year’s profits [11], despite registering a growth in the overseas market.

In our approach, we will be extracting the information, through a web scrapper, regarding reviews and their writers, with a custom-built algorithm. This process would be automatic which would also reduce human input to a minimum thanks to some validation techniques. After this, the information would be analysed and further processed, retrieving all the reviews that were deemed to provide good enough insights. These reviews will then be scored based on a sentiment analysis of its composition, with the final scores being used to evaluate and confirm whether the model would succeed in this format and if the hypotheses proposed are correct.

2. REVIEW LITERATURE

2.1. ONLINE REVIEWS & SURVEYS

Data collection types can be divided into Primary and Secondary Data.

Primary data is a type of data collected by the researchers themselves through methods such as Surveys, Questionnaires or interviews. Because of this hands-on approach, this type of data is expensive due to the time and resources expended. In addition, it is limited by the user pool that the researcher can reach. This type of data is widely used in applied research.

Because this data is collected by the researcher, it means that the data itself will be curated and therefore richer, with it usually answering the questions prepared prior, which makes it widely used in applied research. However, because of this, it can be considered quite expensive due to the time and resources needed. Furthermore, this type of data will be limited by the user pool that the user can reach, leading to smaller and more similar sets of data.

Due to the large volume of data, during this work we will be collecting Secondary Data through web scraping, a technique to request data that would later be extracted, treated and analysed, since manually curating each individual review would be impossible. However, the lack of manual validation could result in a dataset filled with rich or poor reviews.

Online reviews can come in 3 formats: written reviews, audio reviews and video reviews.

In written reviews, the users can write what they want without any restrictions, barring some impositions by the website's holder such as a character limit or a dictionary of banned words. However, because of this, the quality of the data is unknown, since written reviews can be plagued with issues such as bad spelling and grammatical errors.

Audio or Video Reviews are usually found in social media websites or applications such as Youtube or TikTok and their purpose is to illustrate the uploaders' feelings towards a certain product, whilst also showing the product to the viewer.

Whilst written reviews can be prone to grammatical errors and misspellings, audio and video reviews are not free of issues either. Just like its written counterpart, they still fall flat to grammatical errors and alongside this they are subject to external factors, such as the user's voice. If an automated transcript program has difficulty understanding a user's voice, whether it is due to a thick accent, how clear is the user's speech or poor audio quality, it will not fully transcribe the review, thus making it a poor review. In addition, there is a further issue in video reviews: due to them being a content with visual aids, the reviewer can sometimes refer to and praise a product's characteristic using visual pointers or motions, and since we cannot analyse this type of data, the review would, just like audio, be poorer all around when converted to its written form.

Due to an error margin of between 1.75% to 18.75% [12] using the Trint's AI transcription software, and the vast space required to save the video or audio reviews, we felt that this approach would be better suited for written reviews.

In addition to these issues, there is also the issue of fake reviews. Due to a significant number of fake reviews out there (in 2021 on average 4% of all reviews were fake) [13], the results could be affected.

Despite this, we still feel that, due to this last issue also occurring with primary data, along with the sheer number of reviews, the written reviews data can work as a good data source for the model in question.

2.2. Research Model

The original UTAUT model was formed taking into consideration the theories and models most prevalent at the time (year 2003), namely the Theory of Reasoned Action (TRA), the Technology Acceptance Model (TAM), the Motivational Model (MM), the Theory of Planned Behaviour (TPB), a model resulted by the combination of TAM and TPB (C-TAM-TPB), the Model of PC Utilization (MPCU), the Innovation Diffusion Theory (IDT) and the Social Cognitive Theory (SCT), therefore unifying these models with the aim of advancing individual acceptance research.

Through research, Venkatesh found that 7 constructs had direct determinants of intention [14], having chosen the following 4 to integrate into his model: Performance Expectancy, Effort Expectancy, Social Influence and Facilitating Conditions, discarding attitude toward using technology, self-efficacy and anxiety due to them not being theorized of being determinants of intention.

Performance Expectancy, Effort Expectancy, Social Influence influenced Behavioural Intention, with Facilitating Conditions influencing Use Behaviour. In addition to this, every relationship was moderated through external factors such as Age, Gender, Experience and Voluntariness of Use as seen in Figure 1.

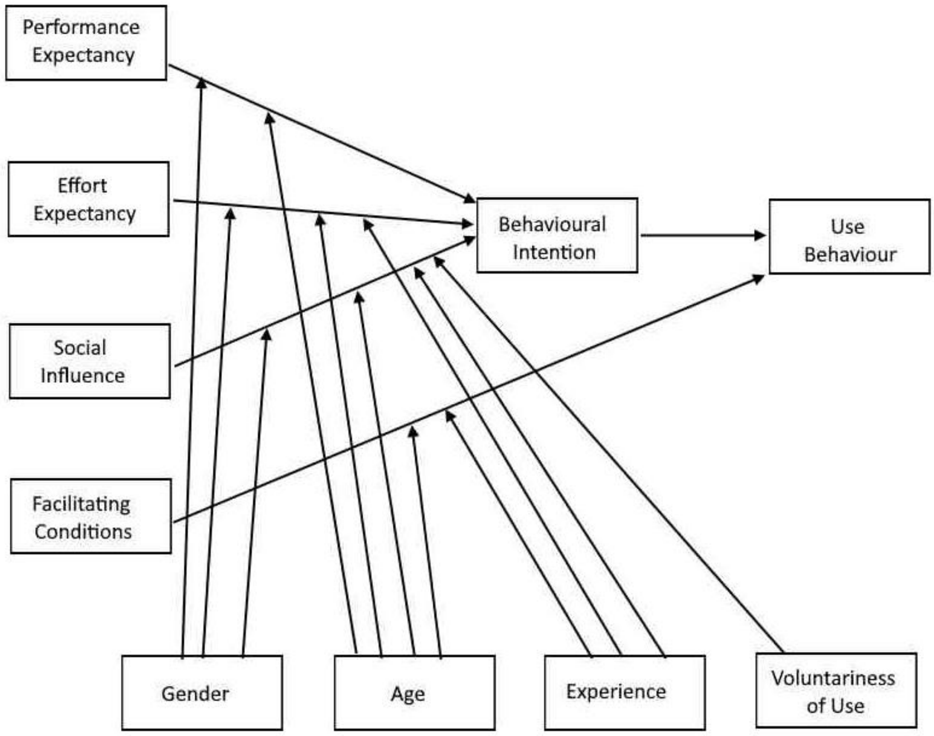

Figure 1 - Original UTAUT Model

In 2012 a new updated version of UTAUT was designed, dubbed UTAUT2, with the goal of adapting to a consumer technology use context. This new version of UTAUT adds 3 new constructs to the original model in the form of Hedonic Motivation, Price Value and Habit, with the first 2 constructs being only related to Behavioural Intention whilst the latter is related to both Behavioural Intention and Use Behaviour. In addition to this, Facilitating Conditions gained a new relationship with Use Behaviour. Furthermore, in UTAUT2, the moderator Voluntariness of Use is dropped to make the model applicable in the context of voluntary behaviour.

It is a model with high versatility, having had a wide range of applications, from cases in the field of medicine to more recent technologies such as blockchain [15] or automatic cars [16] and even tourism [17] and restaurants [18].

As such, the model ends up having Performance Expectancy, Effort Expectancy, Social Influence, Facilitating Conditions, Hedonic Motivation and Price Value influencing Behavioural Intention. Moreover, it has Facilitating Conditions, Habit and Behavioural Intention affecting Use Behaviour.

All these relationships will be moderated by at least one external factor in the form of Age, Gender and Experience as seen in Figure 2.

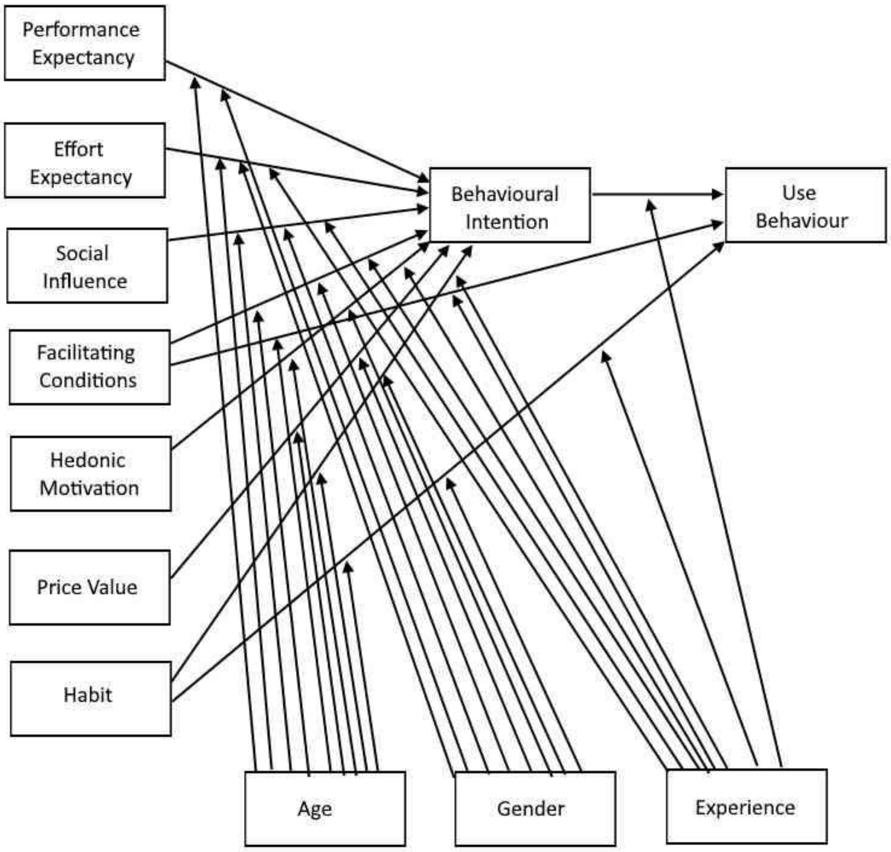

Figure 2 - UTAUT2 Model

2.3. HYPOTHESIS

2.3.1. Performance Expectancy

Venkatesh describes Performance Expectancy as the degree to which an individual believes that using a certain system would be more beneficial to the individual, improving the performance of the task, with it also being considered one of the main predictors to a user's intention to adopt a certain technology [14]. Due to this concept not fitting the theme of the reviews in our dataset, watching animated movies or shows, we adapted this construct to be the expectancy a user had about the performance of a piece of media in key areas such as visual effects, soundtrack composition or its characters.

Therefore, with this construct we hypothesize that:

H1. Performance Expectancy has a positive influence on Behavioural Intention.

2.3.2. Effort Expectancy

Effort Expectancy is defined as the amount of effort a consumer puts in to use a certain technology [14].Due to substantial similarities between the amount of effort a user would put into a technology and into watching animated content, we felt this construct would not need further changes.

As such, we present the following hypothesis:

H2. Effort Expectancy has a positive influence on Behavioural Intention.

2.3.3. Social Influence

Venkatesh defines Social Influence as the extent to which consumers perceive that their important ones believe they should use a particular technology [14].

As such due to the nature of reviews of recommending a certain technology, in this scenario an animated show, this construct should, in theory, be a good a fit.

Hence, we hypothesize that:

H3. Social Influence has a positive influence on Behavioural Intention.

2.3.4. Facilitating Conditions

Facilitating Conditions is defined as the level of belief from a user that technical or organization support exists for the use of a technology [14].

To better fit the theme in question, this construct was changed to whether the conditions given by the piece of media consumed were enough to fully understand and keep up it.

Venkatesh initially defined that this construct would influence Use Behaviour, however later updated it to also influence Behavioural Intention during the model's second iteration [19].

As such, we hypothesize:

H4a. Facilitating Conditions has a positive influence on Behavioural Intention.

H4b. Facilitating Conditions has a positive influence on Use Behaviour.

2.3.5. Hedonic Motivation

Venkatesh defined Hedonic Motivation as the fun or enjoyment a consumer had from using a certain technology, whilst also detailing that this construct has influence over technology acceptance and use, in addition to it being a key predictor of a lot of consumer behaviour research [19]. Thanks to

Sentiment analysis being the analysis of a user's emotions in text form, we feel like this will be a construct that will interact well and be expressive with the approach of this study.

Currently, we hypothesize that:

H5. Hedonic Motivation has a positive influence on Behavioural Intention.

2.3.6. Price Value

Price Value refers to the trade-off between the expenditure of the usage of a certain technology and its perceived benefits [19]. However, this construct is not applicable, in this form, to reviews of an animated show or movie, since these will mostly focus on the show as content and not as a product, thus not evaluating price aspects such as streaming services or retail releases. Because of this, we considered the cost to be the time used watching the animated content and its budget.

With this construct, we hypothesize that:

H6. Price Value has a positive influence on Behavioural Intention.

2.3.7. Habit

Habit has been defined as the degree to which users tend to perform behaviours automatically due to learning [19]. Whilst constructing the keyword dictionary, we considered this to be the patterns exhibited by the user when watching a show or movie.

In addition, this construct is expected to be an influence to Use Behaviour and Behavioural Intention[19].

As such, we hypothesize that:

H7a. Habit has a positive influence on Behavioural Intention.

H7b. Habit has a positive influence on Use Behaviour.

2.3.8. Behavioural Intention

One of the key constructs of the UTAUT2 model, due to it being heavily influenced by other constructs, Behavioural Intention is the user's future plans for a certain technology, in this scenario being whether a user intends to keep following a certain show.

Given this, we propose that:

H8. Behavioural Intention has a positive influence on Use Behaviour.

2.3.9. Use Behaviour

The other key construct of the model, Use Behaviour is considered as the user's behaviour measured from the frequency of use of a certain tool, in this case the frequency of watching.

It can be considered as one of the hardest constructs to work alongside an online data extraction-based approach due to combining an action with a frequency.

Furthermore, there isn't any hypothesis originated from this construct, due to it not influencing any other construct or moderator, instead being the influenced construct.

2.3.10. Moderators

As mentioned previously, during the jump to UTAUT2 [19], the Voluntariness of Use was dropped, keeping only Age, Gender and Experience as moderators.

However, as we later verified, due to the low sample size that survived the algorithm filtering, we had to avoid adding new variables to the model and as such scrapped the hypotheses related to the moderation of the relationships between constructs, using the moderators as filtering criteria during the filtering stage.

2.4. PRIOR RESEARCH

Although there has been a great deal of documented studies on animation, its history and techniques, through careful research we only managed to find one work dedicated to the usage of UTAUT alongside animation [20].

Whilst the previous work is based on the original UTAUT model and uses questionnaires as a data source, we can still take comparisons between both studies and can, as such, say that, based on the results from that study in question, the animation theme can, in theory, work alongside the original UTAUT model.

With that being said, we didn't find any studies related to using online generated content with the UTAUT2 model, with that in mind our main goal becomes to see if it is possible to combine both these concepts successfully, thus proving that with the revisions made to the UTAUT model, in the form of the addition of the Hedonic Motivation, Price Value and Habit Constructs in UTAUT2, will not prove a deterrent to the functioning of the enhanced model alongside the animation theme, and that information found online can also be useful with the research model.

3. METHODOLOGY

3.1. METHODS USED

3.1.1. Web Scrapping

Web Scrapping is a process of utilizing bots to extract content from a certain webpage [21].

By going through several webpages, it will extract all the relevant information from one page before moving to the next one and repeating this process as many times as desired.

3.1.2. Text Mining

Currently, data comes in 3 different formats:

- Structured Data – Data standardized into a format fit for tables and, as such, ready to analyse and used with machine learning algorithms.
- Unstructured Data – Raw data that can come from several sources and that was not yet cleaned and treated.
- Semi-Structured Data – A mix of Structured and Unstructured Data.

Text Mining is the process of extracting insights from unstructured data and, as such, is of key importance to the development of this work, since all the data extracted from the internet will not be processed and will instead be raw and unstructured, which is something normal, as 80% of data in the world is currently in an unstructured format [22].

3.1.3. Natural Language Processing

Natural language processing (NLP) is a field of Artificial Intelligence (AI) that aims to give computers the ability to understand natural language, written and spoken human language, allowing it to detect figures or speech or detect a user's feelings through Sentiment Analysis.

3.1.4. Sentiment Analysis

Sentiment analysis is a method that detects a customer's feelings toward a product, services, individuals and/or themes within a string of text, which can be a single phrase, an entire paragraph or even a whole document. As such, the aim is to measure an individual's polarity (positive, neutral or negative opinion) based on computational treatment of subjectivity in a text.

3.2. DATA

The data used on this project is divided into 4 sets: Review Data, UTAUT Data, User Data and Review Score Data

The Review data contains 193596 entries, being composed of the review's contents, the year they were written and the writer's username, later adapted to be an ID. This dataset has reviews that span from 2005 to 2021.

The User data has 70078 records, with its components being the user's username, later adapted to be an ID, their gender, date of birth (with the last two previously mentioned being optional fields), the day they created an account on the website and the total number of reviews they had written at the point of extraction.

The aforementioned dataset greatly suffers from the UTAUT2 requirements, and as such ends up being the dataset that is compromised the most. Since, as mentioned previously, the fields of gender and birthday are not mandatory in the website, there is a significant number of users that only have one of these fields or none at all. After further analysis to the User Data, we found that only 38.71% of the users had all the required fields associated to their profile, which represented 44.27% of the total reviews. As such due to this requirement alone, we wouldn't be able to use over half of the reviews contained in the dataset.

Furthermore, there exist many users with false information that is easily detectable, such as odd or even impossible dates of birth. For such a reason, we set an age threshold to attempt to elude such accounts, only taking into consideration users which ages between 16 and 90 at the time of their review, using 16 as this is the oldest age of digital consent defined between all European member states through GDPR [23].

Using these guidelines and the maximum and minimum ages a user could get, we found that our users would have been born between 1915 and 2006 to fit the criteria imposed above; 1915 so the user would be 90 years old in the first year of reviews (2005) and 2005 so the user would be 15 in the last year of reviews (2021). With this information, we filtered for Users with all the mandatory information and between the age gap defined. Through this, we found that our users were predominantly male (18996 users, corresponding to 71.14%), with female users only amounting to 26.93% (7193 users). Users that displayed their gender only corresponded to 1.92% (512 users).

In total, it meant that our pool of valid users would be, at maximum, of 26701 users, a big decline from the initial 70078 users (the pool of valid users corresponds to 38.1% of this value).

The UTAUT Data is the dataset used to filter the reviews. It has the UTAUT constructs and their respective categories, alongside with the questions and keywords to said categories. Through a basic question and a lot of trial and error, we devised the final question to each category and from there we researched various words and synonyms that would fit that specific question. This resulted in all the words currently present in the UTAUT dataset.

Besides the data currently presented, we also attempted to create new variables to use in the model and changed others.

To match the experience moderator in the UTAUT model, we created a new variable named Number Of Reviews, which was the sum of all reviews present in the Review Data from each user. When deciding on the experience moderator, we felt there were only 2 suitable choices: the number of reviews or a variable already present when scrapping, which is the day the user created the account. Whilst the age of the account could present us with some knowledge of when the user started to follow the forum and the media, it could also mean absolutely nothing, as the website did not have an admission fee to register and, as such, one could easily create a new account and start posting reviews there. For this reason, we felt that the number of reviews was the only variable usable for this moderator.

In a bid to further improve user data, we originally added a variable Country as we thought this would give more depth to the user variables. This was done using the country or city stated in the user profile and matching it to a database of countries, with the cities being converted to countries based on their location. However, due to not a lot of profiles having this information, or having false or invalid information, such as misspelled city or country names or non-English names, we ended up scrapping this variable, as it would reduce our original user pool from 70078 to 16787, a reduction of ≈76.05%.

As such, we felt that it would not end up adding more value to the output than what we would be giving up. The variable Birthday suffered alterations due to the way it was presented when scraping: the date was inputted in segments of day, month and year. That led to having a date of birth array with missing values, such as having the day and year but not the month. To standardize and avoid losing data because of this issue, we only took into consideration the year of birth, essentially defaulting all birthdays to January 1st of the users' year of birth.

To conduct a further analysis to the review dataset, we took all the unique instances of words in the reviews, and we considered words to be a group of letters separated by empty spaces, filtering out special characters like dots or comas. In addition to this, we counted all the occurrences of said words, thus creating a Word dataset. Along with this dataset, we also developed a script that, using the package *english-words*, filtered the word set occurrences based on if they were a valid word in the English dictionary or if they were one of the following instances: nonsensical words, words that did not exist, words in another language or words with typos. The main goal was to try and find whether a good percentage of our words were usable with our algorithm. In total, we had 465306 unique words, with only 15429 (≈3.32%) words being real and the overwhelmingly majority, 449877 unique words (≈96.68%), being considered invalid. However, we can see a different scenario when looking at the total occurrences, where we can find a total of 88999207 words. The previously great majority of unique instances have a total of 21274908 (≈23.9%) words, which translates to roughly a quarter of the total word count. Despite having a surprising low count of unique words, the total number of real words is 67724299 (≈76.1%). This is due to the possibility of having a lot of variations of errors, with a single different character creating another word, thus creating the illusion of a large failure, when it is the other way around.

3.3. ALGORITHM

The algorithm was broken into 3 different stages: Data Collection, Data Filtering and Data Scoring.

During the data collection stage, we built a web scrapping algorithm that accessed every show or movie and its review pages, retrieving all the important review information such as the contents of the review, the date it was written and the user who wrote it. After going through over 40000 entries, with multiple review webpages each, we had scrapped all available reviews. After this, we went through all the individual users who had written reviews and extracted all of their available information, again through another custom scrapper, thus gaining information such as gender, date of birth and the date they registered with the website. In this stage, we also calculated the number of reviews of a certain user and their country of origin, but the latter variable ended up not being used.

To avoid straining the website with multiple GET requests and to reduce time between each run, we decided to save all of the previously mentioned data locally, thus increasing the speed of the entire process.

Both these web scrapping algorithms were also built with scalability in mind, thus allowing the database to be updated whenever necessary.

During the data validity stage, we took all the reviews from our dataset, and using all the keywords from our UTAUT dataset, proceeded to filter the reviews.

In the UTAUT model, there exist 9 constructs, with each, barring Use Behaviour, being composed of 2 or more Categories. The main goal of this stage was to have the review match with a single word from every category. In this stage, we ran into some issues when developing the algorithm.

The main issue originally was getting the reviews to match with all the items on our constructs. With some thorough analysis, we noted that several reviews only had a couple items as NULL, which meant that the system could not detect the item's words in the review. To try and surpass this issue, we initially decided that the review would only have to answer to one item of a category, so for example if the construct Performance Expectancy had 3 items named PE1, PE2 and PE3, only PE's words would need to be detected in the review for the construct to be valid, still requiring the other constructs to be valid. However, this led to a lot of categories having sentiment scores of 0, usually reserved for neutral values, thus tampering with our results as this would lead to a high correlation.

As such, we decided to go for a different approach: reducing the number of categories per construct to 2 or more (in the original test we were always using 3 per construct).

Because several constructs on the original model had a lot of categories, it was difficult to populate enough questions and unique keywords that would only fit one question. The low number of keywords made it hard to find matches without resorting to the previous idea. By reducing our categories to only 2 or 3 per construct, we allowed the algorithm to have more words per category and, as such, to find more reviews within our criteria. Whilst this could also lead to different results, since we would sometimes only have 2 categories, we found that it would not be as harmful as the previous approach.

After this, it would then proceed to filtering the users. The algorithm first checks if the review was made by a valid user, who is someone with personal information assigned and correct. The conditions defined for correct personal information were having an age between 15 and 90 at the time of the review and a non-null value for gender, defined as Male, Female or Non-Binary.

With all the filtering done, we entered the Data Scoring stage. In this stage, we used rule-based sentiment analysis tool to score each sentence, defined by the words between periods, with a keyword, assigning a value to the category represented in that sentence based on the sentiment output. After all the relevant sentences were scored, it would average the score per category based on how many times the category was present throughout the review.

We opted to use a rule-based sentiment analysis tool rather than a machine learning classifier due to the nature of the data, since because we do not have any scores to classify our data, we could not create a testing and training set to classify all the reviews. While the rule-based classifier is more generic, it also allows for a thorough analysis of the data taking into consideration factors like the sentence strength.

The chosen rule-based sentiment analysis tool used was VADER [24] (Valence Aware Dictionary and Sentiment Reasoner), a lexicon and rule-based sentiment analysis tool introduced in 2014.

VADER calculates the scores of each sentence on a lexicon of sentiment features (a combination of words, expressions and emoticons) and through some rules such as punctuation, capitalization and degree modifiers. Therefore, with this sentiment analysis tool we'll be able to score each sentence from a review, with the attributes mentioned previously, based on how positive or negative the chosen text is, normalizing the score between -1 and 1, with 0 representing a neutral reaction.

VADER is, according to its creator, "*specifically attuned to sentiments expressed in social media*" [25]. With this in mind, we felt it would fit well with the reviews in our dataset, due to the more informal language used, such as using emoticons and high rate of punctuation marks, features highly rated by VADER.

Finally, after having all reviews rated and scored, they will be imported through a final filtering process to create a final score dataset with the UTAUT2 model format, with the sentiment scores of each category and user information for each user. In the event that a user has more than one review in this stage, all of its values, including age at the time of review, will be averaged, thus only having one row per user.

In figure 3 we can see the outline of the entire process described above.

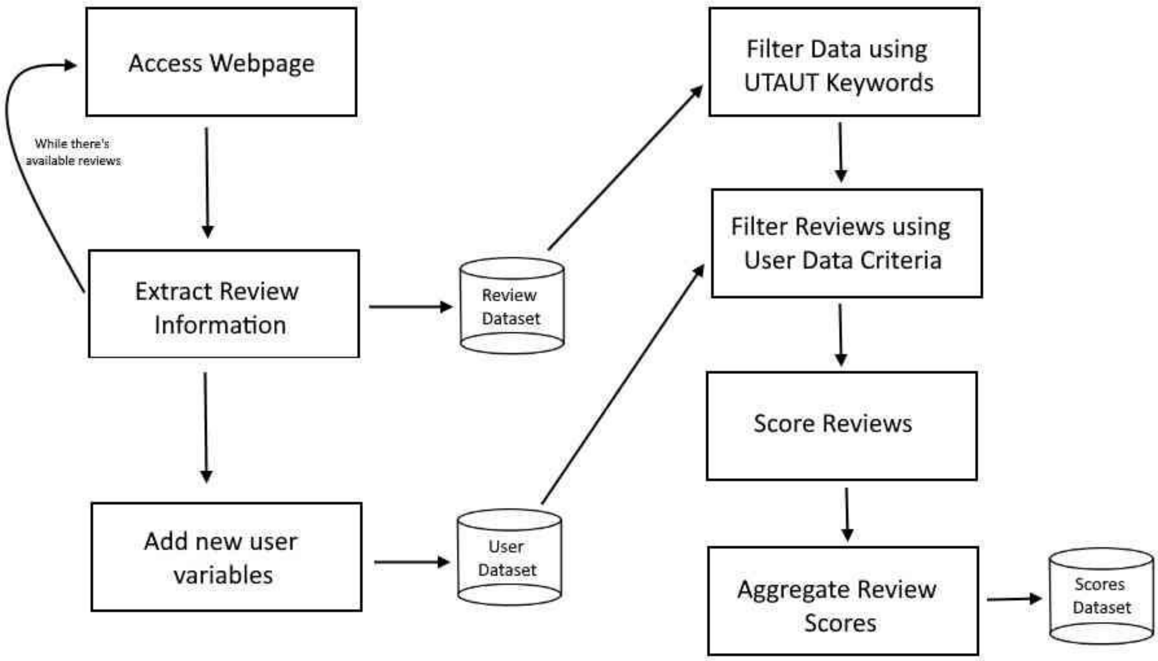

Figure 3 - Algorithm Methodology

4. RESULTS & DISCUSSION

To analyse the results obtained in the scoring stage, we imported them to SMART-PLS [26]. The software was at version 4.0.8.4 at the time of this study. SMART-PLS is an interface software for variance based structural equation modelling (SEM) using the partial least squares method (PLS).

PLS-SEM is a modelling approach aimed to maximize the explained variance of the dependant latent constructs [27], by calculating the construct scores for the latent variables[28].

As mentioned previously after scoring the reviews, they are then aggregated by user, with this being done since in a regular survey you would normally only have 1 answer per user, hence us trying to simulate that through an aggregation of all the user's reviews. After running the algorithm, we found that we had a total of 407 valid reviews, that fit the criteria defined previously, and after the aggregation we only had a total of 219 different users.

4.1. MEASUREMENT MODEL

The evaluation of the measurement model is dependent on the construct reliability, convergence and discriminant validity of the model.

Construct reliability is measured by the results of the composite reliability, with the results requiring a value higher than 0.708 to represent cohesion[29], a result seen in all constructs as described in Table 1.

The next step is to verify the convergence validity of each construct. We can analyse this by using the metric Average Variance Extracted (AVE), the sum of the squared loadings divided by the number of indicators, of each construct. To describe over 50% of the indicator's variance, we require that AVE is over 0.50 [29], something that is verified in all constructs as seen in Table 1, thus indicating convergence.

Constructs	CR	AVE	PE	EE	SI	FC	PV	HM	HT	BI
Performance Expectancy (PE)	0.847	0.649	**0.806**							
Effort Expectancy (EE)	0.754	0.606	0.456	**0.778**						
Social Influence (SI)	0.723	0.566	0.426	0.291	**0.752**					
Facilitating Conditions (FC)	0.738	0.590	0.420	0.361	0.323	**0.768**				
Price Value (PV)	0.788	0.650	0.591	0.405	0.393	0.329	**0.806**			
Hedonic Motivation (HM)	0.768	0.632	0.637	0.440	0.390	0.353	0.544	**0.795**		
Habit (HT)	0.750	0.599	0.453	0.293	0.342	0.301	0.448	0.446	**0.774**	
Behavioural Intention (BI)	0.807	0.677	0.389	0.395	0.388	0.313	0.422	0.388	0.287	**0.823**

Table 1 - Correlations, Reliability and Validity Measures (Composite Reliability and AVE) of Latent Variables

Finally, we must assess discriminant validity. To do this we'll be using the Fornell-Larcker criterion and Cross loadings. With the Fornell-Lacker criterion, we require that the square root of the AVE in a construct is higher than the relations between the other constructs. Through Table 1, it's possible to check that all the values of the square root of AVE, represented in bold, are higher than the relations between the other constructs, thus passing the criterion.

Constructs	**PE**	**EE**	**SI**	**FC**	**PV**	**HM**	**HT**	**BI**	**UB**
Performance Expectancy (PE)									
PE1	**0.804**	0.400	0.352	0.324	0.492	0.525	0.427	0.300	0.321
PE2	**0.747**	0.333	0.266	0.338	0.415	0.429	0.292	0.227	0.127
PE3	**0.862**	0.371	0.390	0.359	0.512	0.568	0.369	0.383	0.197
Effort Expectancy (EE)									
EE1	0.383	**0.810**	0.238	0.319	0.353	0.401	0.296	0.326	0.148
EE2	0.323	**0.746**	0.214	0.239	0.274	0.277	0.152	0.287	0.058
Social Influence (SI)									
SI1	0.264	0.172	**0.754**	0.141	0.214	0.225	0.204	0.293	0.090
SI2	0.377	0.266	**0.750**	0.346	0.379	0.362	0.311	0.290	0.161
Facilitating Conditions (FC)									
FC1	0.232	0.175	0.147	**0.647**	0.254	0.238	0.185	0.185	0.095
FC2	0.392	0.353	0.322	**0.873**	0.261	0.301	0.269	0.285	0.155
Price Value (PV)									
PV1	0.589	0.386	0.351	0.335	**0.775**	0.500	0.432	0.315	0.236
PV2	0.381	0.276	0.289	0.206	**0.836**	0.387	0.301	0.363	0.267
Hedonic Motivation (HM)									
HM1	0.626	0.429	0.389	0.378	0.548	**0.928**	0.466	0.388	0.257
HM2	0.332	0.237	0.191	0.118	0.258	**0.635**	0.175	0.187	0.075
Habit (HT)									
HT1	0.402	0.302	0.366	0.217	0.464	0.415	**0.778**	0.224	0.295
HT2	0.299	0.151	0.163	0.249	0.227	0.276	**0.771**	0.221	0.291
Behavioural Intention (BI)									
BI1	0.366	0.339	0.345	0.268	0.400	0.364	0.317	**0.883**	0.277
BI2	0.265	0.314	0.290	0.250	0.283	0.264	0.129	**0.758**	0.104
Use Behaviour (UB)									
UB1	0.271	0.135	0.166	0.168	0.312	0.238	0.379	0.246	**1.000**

Table 2 - Loadings and Cross-Loadings

Furthermore, we need to ensure that a category item has higher loadings than cross-loadings. By going through Table 2, it is possible to observe that all categories fulfil these requirements, with the loadings represent in bold.

With these 2 factors ensured, so is discriminant validity.

As such, all the previous results confirm that there is an analytical difference between all the constructs allowing us to advance to the testing of the structural model.

4.2. STRUCTURAL MODEL

Since all the measurement results met all their criteria, we could then move on to the structural analysis.

During this stage, we will be assessing collinearity issues, the significance of path coefficients and their relevance, along with the model's variation.

To assess collinearity issues, we will be calculating the variance inflation factor (VIF), which should be below the threshold of 5. The VIF range in the current model ranged between 1.018 and 1.467, thus meaning that collinearity isn't an issue, which could have still been a potential problem if the values were between 3 and 5.

In order to evaluate the significance of the path coefficients and the model's variation, we applied bootstrapping, a procedure that tests the statistical significance of several results, including path coefficients and R^2, with 5000 subsamples.

In order to validate if the path was significant, we checked for the p-values generated during the bootstrapping process, looking for those values to be below either 0.1, 0.05 or 0.01 to prove significant validity.

Effort Expectancy ($\beta = 0.186$, $p < 0.05$), Social Intention ($\beta = 0.189$, $p < 0.05$) and Price Value ($\beta = 0.179$, $p < 0.05$) are statistically significant towards explaining Behavioural Intention, therefore proving H2 and H5 respectively.

In relation to Use Behaviour, we have Habit ($\beta = 0.330$, $p < 0.01$) and Behaviour Intention ($\beta = 0.144$, $p < 0.5$) who achieve statistical significancy, thus proving H7b and H8 respectively.

Hypotheses		β	Support	R^2
Behavioural Intention				0.294
H1	Performance Expectancy → Behavioural Intention	0.023	No	
H2	Effort Expectancy → Behavioural Intention	0.186**	Yes	
H3	Social Intention → Behavioural Intention	0.189**	Yes	
H4a	Facilitating Conditions → Behavioural Intention	0.082	No	
H5	Price Value → Behavioural Intention	0.179**	Yes	
H6	Hedonic Motivation → Behavioural Intention	0.085	No	
H7a	Habit → Behavioural Intention	0.015	No	
Use Behaviour				0.164
H4b	Facilitating Conditions → Use Behaviour	0.023	No	
H7b	Habit → Use Behaviour	0.330***	Yes	
H8	Behavioural Intention → Use Behaviour	0.144**	Yes	

Note : *significant at p < 0.10; ** significant at p < 0.05; significant at p < 0.01

Table 3 - Results of hypotheses testing

As the other relationships did not attain statistical significance their related hypotheses are invalid.

As seen in table 3, 5 of our 10 hypotheses are valid, we believe that these values are also resultant of the format of open questions required to apply the filtering software to the entire review. Whilst reviews hold a method of linear data gathering, with the results directly tied to the questions, in extracting review information the results of our questions are subject to the user's perception thus not as reliable as closed answers.

The variation R^2 of Behavioural Intention and Use Behavioural was explained by 29.4% and 16.4% respectively. This is likely due to a lot of variances in the dependant data and the lack of control in the data, something expected as we can't validate all the data extracted nor fully convert it to the standards of a survey or questionnaire.

As goodness of fit (GoF), a measure of fit for PLS-SEM, can't reliably distinguish between valid and invalid models [26], with researchers being advised to avoid this measure, despite this, we turned to Standardized Root Mean Square Residual (SRMR) criterion and to Normed Fit Index (NFI) to prove model fitness.

According to previous studies, to prove model fitness the value of SRMR should be below the threshold of 0.05, with values below 0.1 considered acceptable [30]. With the SRMR value in our model being of 0.085, a value below one of the thresholds defined.

However, the same success can't be seen with NFI, boasting a value of 0.275. NFI ranges from 0 to 1, with values closer to 1 suggesting a better fit [30]. Despite this, NFI is known to have issues with sample size [30],which leads us to think that the low value might not be an issue as problematic as initially thought.

While we can't completely prove model fitness with the dataset in question, we can't disregard that the model passed all the measurement criteria and the SRMR criterion, suggesting that it doesn't fully fail, leading to the possibility of the model fully working in the future.

As there are not, to our knowledge, any studies on using this specific model, without discarding any constructs or moderators, with online reviews using sentiment analysis, this is still an area in an infancy stage, and as such we believe that with further iterations to surpass the limitations that will be described in the next chapter, it might be possible to fully apply the model at hand alongside online content, thus possibly generating a fully reliable source of information and removing the need of surveys or questionnaires.

5. LIMITATIONS & FUTURE RESEARCH

5.1. LIMITATIONS

To preserve the authenticity of the UTAUT2 model, nor the constructs nor the moderators could be heavily tampered with, as this would put into question the integrity of the study, which resulted in some big limitations for this work, some of which were expected during the planning phase. However, as the development of the algorithm progressed, we found that more and more limitations were arising in several areas. Although the limitations are closely connected and depend on each other for the most part, they can be divided into limitations with the UTAUT2 model and limitations with the algorithm. Website Limitations, Wordset Limitations and Theme Limitations fall uniquely into the UTAUT2 limitations umbrella, whilst Algorithm Limitations is, as suggested by the name, the set of limitations exclusive to the algorithm. The Review limitation is composed of two issues, both related to the UTAUT2 and algorithm issues.

5.1.1. Website

The first limitation comes from the moderators related to personal information: Age and Gender.

Both the original UTAUT model and UTAUT2 require personal information in the form of Age, Gender and Experience. While collecting this information in a survey or questionnaire is not usually an issue, it suddenly becomes a big problem while scrapping information on the internet due to the General Data Protection Regulation (GDPR).

As mentioned previously, a lot of multinational corporation's websites are GDPR compliant or abide to federal or local data protection laws, with the same occurring in a large number of forums or other websites with user information. As such, a lot of websites with usable information become suddenly unavailable, since personal information is a key aspect and moderator of the UTAUT2. We think that this is the biggest limitation of the entire project, as this issue in the chosen website needing to have these features visible and available for scrapping drastically reduces the pool of usable websites, since that is our main way to get information, which was a problem, as websites with readily available User information were few and far between, making it a big issue for data collection.

5.1.2. Wordset

Another construct limitation lies in the wordset itself. In order to differentiate which sentences pertain to which wordset question, we must have a certain number of keywords assigned to each question.

By having a unique set of keywords for each category, we will decrease the chance that 2 categories are present in the same sentence, thus somewhat reducing the high correlation between categories since there will be a lower number of sentences with the same sentiment score.

However, having a unique set of words will mean that the category will have to be represented by at least one word, which makes it harder to find complete matches, thus hindering the number of correct reviews, which is something we saw earlier. Whilst using the constructs themselves instead of their categories would counteract this issue, it would also fall prey to the high correlation problem.

With a normal survey or questionnaire, all the questions would be answered separately, eliminating the data correlation issue from the start, making the results more reliable. In addition, all the questions would be represented, thus ensuring that the data would be valid and usable.

5.1.3. Review

This limitation is, in fact, composed of two large flaws: one in the choice of website to extract information from, and the other in the analysis portion of the algorithm.

The first issue – choosing an extraction point – comes from the wordset problem. As mentioned previously, we need to have every category of every construct represented on a review, and as such we will require reviews that are large enough to have all these requirements present. This, combined with the already described website issue, made it nearly impossible to find a suitable candidate in an already reduced pool of websites, due to many having either short reviews but user data or the other way around. This ended up resulting in poorer reviews, which in turn damaged the effectiveness of the model.

The second issue – the review analysis portion – is riddled with flaws and limitations, due to the nature and quality of the reviews themselves, with the most important one being the lack of context on the review. The algorithm that was setup for the project essentially takes in each sentence and analysis if it contains any of our keywords, which makes it extremely vulnerable to information out of context, since for example a word can be mentioned but not evaluated on. This type of review is also susceptible to synopsis. A synopsis, in this type of review, is a brief outline of the story and its characters, and as such it is prone to trigger the filtering algorithm with words that are part of the keywords set, but in the context of the review should not trigger them, as they do not represent the user's thoughts and feelings about the categories and constructs that contain the keywords used, and are merely used to describe the show itself rather than its attributes. This results in false positives, impossible to differentiate without manual validation, which is out of question due to the mass number of reviews. Another limitation would be the existence of mistakes in words, with this resulting in some words not being picked up by the algorithm and, as such, missing out on valid reviews often by a single character.

Other more common limitations would be the bad grammar, sentence structuration and poor punctuation, which would affect the classifier by having shorter and incomplete sentences that would make no sense.

Lastly, there are reviews in other languages, and although we can translate them into English, some meaning would be lost, which could result in subpar classification.

Like we mentioned with the first issue, some of these limitations could be solved by manual validation, but we deemed it impossible due to the sheer amount of data at hand.

5.1.4. Theme

One more limitation is the existence of constructs unfit for the theme chosen.

The UTAUT Model is used to evaluate the user's acceptance of a certain technology, though it can be extended to a product or service. Although a show or a movie can be part of these latter categories, they are not in this scenario, since the reviews are only about the Show and Movie themselves, evaluating only the piece of animation itself and the user's thoughts about it.

In a regular product review, some key points it could target would be the quality of the product or its pricing, and whether the writer thought it was good, bad or somewhere in between.

In this scenario a construct like Price Value is not originally a good fit, due to the fact that we cannot evaluate a characteristic such as pricing, since the reviews are mostly focused about the show or movie themselves and, as such, do not take into account details relating to its release, whether it is for home release or cinemas, thus requiring for the construct to be adapted as best as possible to fit the scenario at hand. This heavily limits a construct and its keywords, which in turn holds back the filtered dataset by having strangle points in some constructs.

Whilst this is definitely the easiest limitation to fix, its problems are rooted in limitations such as the Websites and Reviews Limitations, which dwindle down the number of websites to extract information from, leading to a scenario like this one, where it was the only acceptable open-source pool of information.

5.1.5. Algorithm

Currently, the algorithm only holds two limitations, with both being present in the filtering phase: having words that trigger the algorithm when in reality they hold no value to the sentence, whether it is due to a word being mentioned but not scored or words taken out of context and being impossible to differentiate the sentiment analysis score of two words in one sentence, with it taking the overall value of the sentence. The first one was already touched upon in the review limitation, and the latter one can be easily explained with the following example of a sentence that would be picked up by the filtering algorithm: “Mikey liked the artwork but hated the music”.

Currently, as there is no way to determine which sentiment word relates to which keyword, we take the sentiment score of the entire sentence. If we consider “liked” as a score of 1 and “hated” as a score of -1, we should have a final sentence score of 0. The main issue is that to the algorithm “Mikey” was neutral to both the artwork and to the music, something that, after manual human analysis, is possible to disprove, as both are vastly different opinions. This leads to inconsistencies in our data which would not be present on a regular “rate this from 1-5” question in a questionnaire.

In addition to this, in this scenario, since we are grading the entire sentence, both categories would use the unique score associated to the sentence, which will also have an interference on the discriminant validity tests, represented by a higher correlation, something that would not happen with a survey approach as overlapping would not be an issue.

Although an option could be to only use a certain number of words surrounding our keyword, we could risk not covering enough words to get the sentiment analysis of the user regarding the topic, or even covering too much and leaking back into other keywords in the same sentence, which would lead to the current scenario. As such, this ends up causing some patterns in the data that should not exist, which ends up damaging our analysis.

5.2. FUTURE RESEARCH AND IMPROVEMENTS

Whilst the limitations mentioned hinder the UTAUT2 model from explaining well the insights obtained from raw data obtained from online reviews, we think that with some future improvements there is a possibility of successfully proving what we currently could not. These are the following:

1. Creating a new UTAUT revision, to better fit the more modern internet age. By swapping personal information moderators such as Age and Gender it will be possible to have access to more and better datasets, as it wouldn’t be bound to GDPR regulations. In addition, changing

some of the more incompatible constructs could also lead to better results as they would better fit scenarios where technology isn't being evaluated.

2. Another approach could be the existence of a dedicated database. A dedicated database would mean that the data in it would be curated, and therefore richer and better suited for data analysis. Thanks to this, we could avoid the major issue of badly written reviews.
3. Finally, we think that with further improvements to either the sentiment analysis package or a filtering and scoring software, it could be possible to differentiate which sentiment words relate to a certain keyword, thus fixing the issue of all categories represented in a single sentence having the same score, which in turn could fix the correlation problem.

6. CONCLUSIONS

The main goal of this work was to find a system that would replace the need for questionnaires with the analysis of data already available online, spanning from 2016 to 2021, obtained using text mining and processed through sentiment analysis.

From the results, we found the model had passed all the criteria of the measurement model. However, only a handful of our hypotheses were valid, with the biggest issue coming from the hypotheses related to the moderators. With this, we discovered that Effort Expectancy, Social Intention and Price Value were the only constructs to play positively influential roles in Behavioural Intention, whilst Habit and Behavioural Intention were the only ones to positively influence Use Behaviour.

In addition, the Animation reviews were only explained through Behavioural Intention and Use Behaviour by 29.4% and 16.4% respectively, thus not explaining a lot, this is likely due to the nature of the approach of this study, namely the collection of uncurated data.

As mentioned previously, data protection laws lead to most websites not displaying user information publicly, although they still collect it most of the times, which means that web-scrapping becomes an unviable method, as it could only capture the reviews themselves and not the user's information. This is a major roadblock for the UTAUT2 model, since it requires some user information to analyse results (gender, age, experience), with this requirement leading to a lot of websites that include review entries being useless for data extraction, thanks to the lack of user information, which alongside datasets with small and poor reviews, greatly reduces the pool of viable datasets. This ended up also affecting the dataset we chose, with personal information displayed being an optional feature, leading to a lot of reviews having to be written off and affecting the results.

Not having a dedicated database was also an issue. As mentioned previously we had to dig deep through the confinements of the web to find a website with a pool of reviews that could initially fit to our requirements, which means that we had to scrape and process these reviews without knowing if they would end up being good and matching up to the expectations set previously, and also that we had to work with what we could find. Having a dedicated database already curated, or at least with credible reputation, would lead to more choices of themes for the analysis and, as such, to a better choice of filtering words. The main issue is that, since these databases would likely be owned by the representatives responsible for the website, it would be hard to get access to this information.

Lastly, there's the issue of a lack of user control. The main issue with taking information from reviews is that there can be anything in them, including information that's unrelated to the topic at hand, spelling errors, slang or reviews that don’t make sense, and that manage to hit all the correct keywords at chance but don’t elaborate on them. In questionnaires or surveys, there are closed questions which allow the results to be within what’s expected of the questions, only needing to grade a certain response with a closed meaning. With the reviews, there’s a bigger uncertainty that can’t be easily curated as you could be going through thousands of reviews, rather than a checkmark or a number. Because of this, the results could deviate from those you would get if you made the users take a questionnaire, which ends up skewing the analysis and damaging the credibility of the results.

In conclusion, we can say that while the model is valid, due to the nature of the UTAUT2 model and because its heavily dependent on the data at end, there would need to be a lot of improvements on the issues mentioned previously to have a more efficient result and the values we were expecting upon taking on this research.

7. BIBLIOGRAPHY

[1] European Parliament, 2022, General Data Protection Regulation (GDPR) Compliance Guidelines (URL: https://gdpr.eu) [Last Accessed in 05/10/2022]

[2] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation

[3] MotionCue, 2020, History of Animation and its Evolution in Film - MotionCue (URL: https://motioncue.com/history-and-evolution-of-animation-in-film/) [Last Accessed in 16/09/2022]

[4] Vectornator, 2022, What Were the First Animated Films? (URL: https://www.vectornator.io/blog/first-animated-films/) [Last Accessed in 13/08/2022]

[5] The Numbers, 2022, Snow White and the Seven Dwarfs (1937) - Financial Information (URL: https://www.the-numbers.com/movie/Snow-White-and-the-Seven-Dwarfs) [Last Accessed in 10/11/2022]

[6] The Numbers, 2022, All Time Worldwide Box Office (URL: https://www.the-numbers.com/box-office-records/worldwide/all-movies/cumulative/all-time) [Last Accessed in 10/11/2022]

[7] Statista, 2022, Global box office revenue 2021 | Statista (URL: https://www.statista.com/statistics/271856/global-box-office-revenue/) [Last Accessed in 12/10/2022]

[8] Box Office Mojo, 2022, Demon Slayer the Movie: Mugen Train (URL: https://www.boxofficemojo.com/title/tt11032374/) [Last Accessed in 11/11/2022]

[9] Precedence Research, 2022, Animation Market Size to Hit over USD 587.1 Bn by 2030 (URL: https://www.precedenceresearch.com/animation-market) [Last Accessed in 13/11/2022]

[10]PRTIMES, 2022, Anime production market declines for the first time in 10 years, down 2% year-on-year, 40% loss, new record high｜TDB press release [Translated Title, Website in Japanese] (URL : https://prtimes.jp/main/html/rd/p/000000328.000043465.html) [Last Accessed in 02/11/2022]

[11] The Association of Japanese Animations, 2022, Anime Industry Report 2021 Summary_02 (URL: https://aja.gr.jp/download/anime-industry-report-2021-summary_02) [Last Accessed in 02/07/2022]

[12] Trint, 2018, AI speaks out: How 'politically correct' are today's top politicians? (URL: https://trint.com/blog/ai-speaks-out-how-politically-correct-are-todays-top-politicians) [Last Accessed in 05/11/2022]

[13]World Economic Forum, 2022, Fake online reviews cost $152 billion a year. Here's how e-commerce sites can stop them | World Economic Forum (URL: https://www.weforum.org/agenda/2021/08/fake-online-reviews-are-a-152-billion-problem-heres-how-to-silence-them/) [Last Accessed in 01/11/2022]

[14]VENKATESH, VISWANATH & MORRIS, MICHAEL & DAVIS, GORDON & DAVIS, FRED. (2003). User Acceptance of Information Technology: Toward a Unified View. MIS Quarterly. 27. 425-478. 10.2307/30036540. [page 446-455]

[15]HANDOKO, BAMBANG & TJANDRA, REYNER & MOZES, LAVIE & GIOK), GIOVANI. (2020). Model for Predicting Auditor Intention to Adopt Blockchain. 44-50. 10.1145/3446922.3446930

[16]NORDHOFF, SINA & LOUW, TYRON & INNAMAA, SATU & LEHTONEN, ESKO & BEUSTER, ANJA & TORRAO, GUILHERMINA & BJORVATN, AFSANEH & KESSEL, TANJA & MALIN, FANNY & HAPPEE, RIENDER & MERAT, NATASHA. (2020). Using the UTAUT2 model to explain public acceptance of conditionally automated (L3) cars: A questionnaire study among 9,118 car drivers from eight European countries.

[17]AHMED, MOHAMMEDJUD & BOGALE, ADINA & TILAHUN, BINYAM & KALAYOU, MULUGETA & KLEIN, JOERN & SHEGAW, MENGISTE & ENDEHABTU, BERHANU. (2020). Intention to use electronic medical record and its predictors among health care providers at referral hospitals, north-West Ethiopia, 2019: using unified theory of acceptance and use technology 2(UTAUT2) model. BMC Medical Informatics and Decision Making. 20. 207. 10.1186/s12911-020-01222-x.

[18]PALAU-SAUMELL, RAMON & FORGAS, SANTIAGO & SÁNCHEZ-GARCÍA, JAVIER & ROBRES, EMILIO. (2019). User Acceptance of Mobile Apps for Restaurants: An Expanded and Extended UTAUT-2. Sustainability. 11. 1210. 10.3390/su11041210.

[19]VENKATESH, VISWANATH & THONG, JAMES & XU, XIN. (2012). Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. MIS Quarterly. 36. 157-178. 10.2307/41410412. [page 161-162]

[20]MOHD SUKI, NORBAYAH & MOHD SUKI, NORAZAH. (2017). Determining students' behavioural intention to use animation and storytelling applying the UTAUT model: The moderating roles of gender and experience level. The International Journal of Management Education. 15. 528-538. 10.1016/j.ijme.2017.10.002.

[21]Imperva, 2019, What Is Scraping | About Price & Web Scraping Tools | Imperva (URL: https://www.imperva.com/learn/application-security/web-scraping-attack/) [Last Accessed in 19/08/2022]

[22]IBM, 2020, What is Text Mining? | IBM (URL: https://www.ibm.com/cloud/learn/text-mining) [Last Accessed in 03/10/2022]

[23]Privo, 2022, GDPR Age of "Digital" Consent https://www.privo.com/blog/gdpr-age-of-digital-consent [Last Accessed in 05/10/2022]

[24]HUTTO, C.J. & GILBERT, E.E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014

[25]GitHub, 2022, GitHub - cjhutto/vaderSentiment: VADER Sentiment Analysis. https://github.com/cjhutto/vaderSentiment [Last Accessed in 30/07/2022]

[26] RINGLE, CHRISTIAN M., WENDE, SVEN, & BECKER, JAN-MICHAEL. ,2022, SmartPLS 4. Boenningstedt: SmartPLS (URL: https://www.smartpls.com/) [Last Accessed in 01/11/2022]

[27]JR, HAIR, & RINGLE, CHRISTIAN & SARSTEDT, MARKO. (2011). PLS-SEM: Indeed a silver bullet. The Journal of Marketing Theory and Practice. 19. 139-151. 10.2753/MTP1069-6679190202. [page 141]

[28]SARSTEDT, MARKO & RINGLE, CHRISTIAN & HAIR, JOSEPH. (2021). Partial Least Squares Structural Equation Modeling. 10.1007/978-3-319-05542-8_15-2. [page 3]

[29]SARSTEDT, MARKO & RINGLE, CHRISTIAN & HAIR, JOSEPH & HULT, THOMAS & DANKS, NICHOLAS & SOUMYA RAY. (2021). Partial Least Squares Equation Modeling (PLS-SEM) Using R : A Workbook [page 77-93]

[30]SCHERMELLEH-ENGEL, KARIN & MOOSBRUGGER, HELFRIED & MÜLLER, HANS. (2003). Evaluating the Fit of Structural Equation Models: Tests of Significance and Descriptive Goodness-of-Fit Measures. Methods of Psychological Research Online. 8. 23–74. [page 37-41]

8. APPENDIX

Constructs	Survey question	KeywordsEnglish
Performance Expectancy	PE1 - I found the animation and art for this show good	animation, artwork, art, cgi, model, design, special effect, portrayal, drawing, effect, 3D, cartoon, anime, 2D, shading, color, colour
	PE2 - I liked the use of music, sound effects and voice acting on this show	music, sound, soundtrack, arrangement, sound effect, tune, singing, song, insert, opening, ending, theme, ost, orchestra, melody, noise, voice, actor, instrument, instrumental, dialogue, speech
	PE3 - I enjoyed the characters for this show	character, cast, depth, protagonist, deuteragonist, antagonist, crew, hero, heroes, villain, role, figure, main, heroine, villainess, boss
Effort Expectancy	EE1 - I found the show easy to understand	understand, keep up, explain, figure, grasp, learn, perceive, realize, tolerate, accessible, effortless, obvious, simple, straightforward, apparent, manageable
	EE2 - I thought the show was hard to follow	follow, accept, comprehend, complex, fathom, grasp, catch on, realize, difficult, hard, challenging, demanding, arduous, problematic, tough, troublesome, pace
Social Influence	SI1 - A lot of friends told me this was a must watch	friend, family, boyfriend, girlfriend, someone, buddy, classmate, colleague,

		partner, associate, companion, cousin, wife, husband, roommate, mate, pal
	SI2 - People on the internet raved about this show	reviewer, expert, weeb, influencer, youtuber, specialist, experienced, recommended, recommend, recommendation, suggest, suggested, suggestion, told, urge, urged, also like, others
Facilitating Conditions	FC1 - I have to knowledge necessary to understand this show	knowledge, ability, information, intelligence, expertise, weird, confusion, confusing, odd, confused, unclear, expert, understanding, wisdom, education, know-how
	FC2 - I am able to easily find this show	find, procure, obtain, get, secure, purchase, rent, buy, stream, crunchyroll, netflix, discover, locate, recover, torrent, nyaa, download, unhearth
Hedonic Motivation	HM1 - I found this show fun	entertaining, entertainment, humor, hilarious, exciting, delight, delightful, engaging, inspiring, fun, funny, amusing, enjoyable, enjoyment, joy, love, loved, like, liked, lovable, pleasant, adore, adored, fond, inspiration
	HM2 - I thought this experience was dull	dull, abhorred, awful, distaste, appalling, depressing, disgusting, dreadful, gruesome, hideous, horrendous, horrible, horrific, horrifying, horror, nasty, ugly,

		unpleasant, boring, dumb, stupid, slow, meh, nasty, displeasing
Price Value	PV1 - I felt my time was well spent	time, well spent, rate, appreciate, cherish, esteem, treasure, respect, season, while, beneficial, effecting, rewarding, worthwhile, useful
	PV2 - The budget was well allocated	budget, price, money, allocated, allocation, resource, value, cost, spend, spent, expenditure, fee, worth, cash grab, money, cash cow, euro, yen, dollar, cashgrab, cashcow
Habit	HT1 - I usually watch shows like this	usually, commonly, frequently, generally, routinely, sometimes, normally, consistently, often, regularly, repeatedly, systematically, consistently
	HT2 - I must always know what happens next	must know, know, intrigued, captivated, interested, interest, curious, curiosity inquisitive, peculiar, puzzling, mysterious, unique
Behavioural Intention	BI1 - I intend to keep watching more things like this	intend, intention, attempt, effort, endeavour, want, find more, plan, planned, prepare, prepared, plan, mean, aim, objective
	BI2 - I'll try to keep up with this show	try, weekly, routine, seek, keep up, hold on, continue, persevere, carry on, carry, go on, stay, standard, everyday, habitual, periodically, daily, week, day

Use Behaviour	UB1 - I *often* watch shows like this	watch, watching, view, viewing, observe, observing, witness, witnessing, see, seeing, check, check out, checking out, checking, visualize

*User Behavioural Complement Words – These words work alongside the User Behaviour construct to represent the number of times or how often an action is made:

many, frequently, abounding, countless, frequent, plentiful, prevalent, several, often, periodically, regularly, usually, oftentimes, recurrently, essentially, largely, particularly, predominantly, generally, repeatedly, over, commonly, ordinarily

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa