Towards an interpretable Advertisement Click Prediction

Catarina Santos

Dissertation written under the supervision of professor Ana Guedes

Dissertation submitted in partial fulfillment of requirements for the MSc in Business Analytics, at the Universidade Católica Portuguesa, January 2023

Abstract

In the new era of computational advertising, Click prediction models are used to anticipate a click response and guide marketers' decisions about whom to target and how to personalize. The more prediction tasks achieve impressive performances, the more trust is put in black models to make important decisions in a business domain. Due to the complexity and lack of transparency of the models, posterior explanation methods are needed to identify features' contributions that envision a global explanation of the model. This thesis develops an advertisement Click prediction using a supervised machine learning framework and uses the KernelSHAP method to provide feature importance insights on the predictions made by the model. The thesis aims to answer: 1) how can marketers integrate Click prediction in their businesses; 2) which are the most impactful features for a click response; 3) how advertisement categories differ from an overall model. For that purpose, is used a publicly available ADS dataset (2016) to train a neural network. The results showed the overall model performance is not substantially different from a segmented categories performance. The output of KernelSHAP showed that even though the visual content is impactful for the likelihood of a Click for all models, each category has its own feature importance pattern influenced by the product the category promotes. The performance metrics presented a high ratio of true negative to true positive due to a class imbalance problem. To mitigate the cost of misclassification is suggested an individual analysis that better fit target business model.

Keywords: Computational Advertising, Advertising Click Prediction, Explainable AI, KernelSHAP, Neural Network.

Title: Towards an interpretable Advertising Click Prediction.

Author: Catarina Santos.

Resumo

Na era da publicidade computacional, os modelos de previsão Clique são utilizados para antecipar uma resposta de clique e guiar decisões-chave como a quem publicitar. Quanto mais preciso é o desempenho, mais confiança é depositada em modelos para tomar decisões importantes num domínio empresarial. Devido à complexidade e falta de transparência dos modelos, são necessários métodos de explicação posteriores para identificar como cada atributo contribuiu para a previsão. Esta tese desenvolve um modelo aprendizagem automática que prevê o Clique em anúncios e utiliza o método KernelSHAP que explica quais os atributos mais relevantes à previsão. A tese tem como objetivo responder a: 1) como podem os profissionais de publicidade integrar a previsão de clique nos seus negócios; 2) quais são as características mais impactantes para obter um Clique; 3) como as categorias de publicidade diferem de um modelo global. Para esse efeito, é utilizado um conjunto de dados publicamente disponível - ADS(2016)- para treinar uma rede neural. Os resultados mostraram que o desempenho global do modelo não é substancialmente diferente do desempenho segmentado de categorias. Os resultados do KernelSHAP mostraram que, embora o conteúdo visual seja impactante para a probabilidade de um Clique em todos os modelos, cada categoria tem o seu próprio padrão de atributos mais importantes para a classificação. Estes são influenciados pelo produto que a categoria promove. As métricas de avaliação apresentam uma discrepância entre previsões corretas entre classes. Para mitigar o custo de uma classificação errada, sugere-se uma análise individual que melhor se ajuste ao modelo de negócio.

Palavras-Chave: Publicidade computacional, Previsão de Click em Anúncios, Inteligência Artificial Explicável, KernelSHAP, Rede Neuronal.

Título: Towards an interpretable Advertising Click Prediction.

Autor: Catarina Santos.

Acknowledgments

Thank you first and foremost to my thesis supervisor, Professor Ana Guedes, for her guidance and availability throughout this process, for her recommendations, and for introducing me to much of the methodology that underlies this thesis.

Thank you to Católica University for creating this opportunity and challenge, which contributed to my personal and professional development. I am also grateful for all the professors whom I had the pleasure to meet during this Master's who impacted me with their kindness and knowledge.

Thank you to my parents, my brother, my colleagues, and my friends —all of whom provided the encouragement, support, and inspiration that made this thesis possible.

Table of Contents

List of Figures

List of Tables

List of Equations

1. Introduction

In today's highly competitive business environment, online advertising has become crucial for diverse industries to survive and thrive as it is a critical factor for business growth. Similarly to conventional advertising, the primary goal is to expand brand awareness, identify potential customers and encourage them to buy their product. However, the promotion is made through digital channels filled with clickable mechanisms on display advertisements. According to Statista Research Department (2022), the digital advertisement market has been growing rapidly worldwide throughout the years. Additionally, it is projected to continue that way reaching US$876.10 billion in 2026. WebFX (2022) also affirms that by 2023 "Digital advertisements will account for more than 66% of the total global advertising spend.". This colossal growth was imposed by the appearance of accessible technological devices and the Internet, which are constantly in a state of flux.

In recent years, large-scale data have become accessible allowing online advertising to go beyond. Mainly influenced by big tech companies, big data-driven innovations attracted more attention to how data could optimize advertising. With data promptly available along with machine learning-based methods, companies can now use gathered information on users' preferences and needs to transform it into valuable predictions of potential customers. As marketers adopted these technological developments, a new field of advertising called computational advertising arose (Dave and Varma 2014). Not only do advertisers benefit from having higher efficiency with consumer interactions, but also consumers are exposed to more relevant content. While companies are keen to build up complex machine learning models that secure a higher accuracy, the crescent complexity has a major limitation: compromised interpretability. For us humans, we can effortlessly understand the logic by which a decision is implemented. However, when considering machine learning as having all the decision-making power, there is no reasoning besides trained weights and biases parameters. Indeed, these models are called black box models due to the lack of transparency and trustworthiness of the prediction made. Black box models perform only based on data observations while taking the risk of undermining the estimated value with undesirable biases such as user interaction bias (Mehrabi et al. 2021). For example, a web-user is exposed to clickable advertisements design to save its historical click behavior and later use it on recommender systems. The user can be the source of the bias, known as presentation bias, whereby the webpage only recommends

previous seen or similar content (Baeza-Yates 2018). Some information is not seen displayed as a vicious cycle and consequently perpetuating bias and unfair results.

Companies want to avoid the risk of mistarget users when facing a large dependency on advertising outcomes with limited resources. Nonetheless, Explainable Artificial Intelligence (XAI) emerges to overcome the paradigm of black box models. It seeks to heighten transparency by deriving clear interpretations for deemed complex unexplainable models. The idea is to learn more about each factor's significance in the decision process in order to adapt the model for a more consistent and trustworthy outcome.

This thesis aims to address the interpretability concerns of an advertisement Click prediction. In the first stage, a supervised machine learning approach will be used to predict a click response based on gathered data about users' information, preferences, and their respective labeled rated response known as targets. In the second stage, an explainable machine learning technique will be applied envisioning a more understandable classification decision. Although computational advertisement is an emerging topic nowadays, it has a large margin for improvement motivating numerous studies. This thesis focuses on the following research questions:

1) To what extent can advertising marketers integrate computational advertising in their business models through the black box model?
2) How can advertising marketers measure the most impactful features contributions for a click response, and what are them?
3) How does advertisement category segmentation differ from a model that pools them all?

2. Literature Review

2.1. The evolution of Computational Advertisement

The advertisement business has been suffering from radical changes throughout the past years. In the last decades, the online advertisement ecosystem has been adopted by countless organizations which perceive Internet disruption as an opportunity to target more consumers. Once again, with the rise of big data availability and algorithms-driven approaches, new advertising possibilities have emerged and companies have shown a steadily growing interest in Computational Advertisement. The concept was born by intersecting the fields of marketing and computer science to better capture the target audience with pertinent and individually personalized advertisements. Consequently, enact adequate allocation of advertising budgets and resources (Huh and Malthouse 2020, 367-368). The purpose remains the same as the traditional advertising scope, from assessing internal and external market factors to the final target decision. Nevertheless, computational advertisement added value by establishing and controlling touch points essential to understanding customers' needs (Perlich et al. 2012). As a result, touchpoints are intended to ensure efficiency in whether or not to address the advertisement and adapt it to become more relevant (Blom 2000). Having all the data information on the customer side envisions a conceded way to satisfy both customers and advertisers. On the one hand, customers are entrusted with suitable advertisements. On the other hand, advertisers reduce the risk of applying resources to a worthless broad target.

2.2. Click on advertisement prediction

The purpose of online advertising is to discover the optimal combination of audience and advertisement given the subject's preferences and needs. From a computational view, this resembles to correctly anticipate whether users will respond favorably (Click) or negatively (Not Click) to an advertisement, given observed user data (Gharibshah and Zhu 2021). Click-Through Rate is a metric used to gauge the ratio of clicked advertisements to the number of impressions displayed. Indeed, it is a valuable indicator to capture user engagement with advertisements and evaluate the target performance.

$$CTR = \frac{Number\ of\ Clicks}{Number\ of\ Impressions}\ x\ 100$$

Equation 1 – Click-Through Rate formula

2.2.1. Related Work

This thesis is based on Roffo and Alessandro's (2016) paper and constructed the ADS dataset. They gathered a representative benchmark of actual displayed adverts to study how personality factors assessed by the Big Five Inventory questionnaire affected the advertising rating and Click prediction. The questionnaire intended to rate the study participants' degree of magnitude on each of the following personality factors: Openness to Experience; Conscientiousness; Extroversion; Agreeableness; Neuroticism (Gosling et al. 2003). Firstly, it was used Affinity Propagation algorithm to evaluate and compare data points' similarities given a binary design of Click/Not-Click to cluster the participants regarding their Big Five characteristics (B5). The algorithm identified eight clusters of personalities, each with its own set of individual characteristics. In further analysis, they perform the following models: Logistic Regressions, Support Vector Regression, and L2-regularized Support Vector Regression comparing the respective models with and without personality factors. The results showed a slight improvement in ROC-AUC, Precision, and Recall metrics on B5 models. Even though the personality traits contributed to the overall performance, the accretion was not substantial. Considering that the Logistic Regression model achieves the best performance, it only differs from the correspondent model not accounting for the Big Five Personality by only 1.5% on ROC-AUC, 0.9% on Precision, and 0.8% on Recall.

Many frameworks of Supervised Learning are used on Click classification problems from which are standout: Logistic Regression, Factorization Machines, and Deep Learning Methods (Gharibshah and Zhu 2021).

Logistic Regression

Logistic regression (LR) assumes a linear dependence of a target variable on a set of coefficient and input values and only infers its distinct impact on each class label. Sometimes LR suffers from underfitting as it is considered to be a high bias and low variance model, which means ignoring relevant relations between features and target outputs and also is not sensitive to small fluctuations in the dataset. A study by Olivier Chapelle (2014) executes a click and conversion prediction using an LR framework through a Maximum Entropy approach. A method that envisions feature selection by combining various contextual facts to calculate the likelihood that a specific class exists. However, the model is limited and reliant on the efficiency of feature engineering to capture valuable features. In addition, click intention can be supported by

multiple volatile interactions, so it is acceptable to assume that click and its predictive features have a non-linear click relationship.

Factorization Machines

Factorization machines (FM) are similar to Support Vector Machines frameworks but utilize a factorized parametrization to capture pairwise interactions. A standard factorization machine approach can typically estimate model parameters accurately when considering sparse data and train with linear complexity, allowing them to scale to vast datasets (Rendle 2010). A study of CTR prediction task (Juan et al. 2017) compared LR and FM models in terms of utility and concluded that FM had a better result by an increment of 3.71%. Nevertheless, the FM model neglects feature importance order and by considering every pairwise interaction with the same weight, the model could learn noise from insignificant relations. As a matter of fact, another study (Huang et al. 2019) developed a CTR model combining feature importance and bilinear interaction that outperformed Factorization Machine models.

Deep Learning Methods

Deep learning methods are the most complex frameworks and nowadays consist of one of the most popular models used for Click prediction. A study proposed a multi-head self-attentive neural network to attribute relationships in the low-dimensional space and predict CTR based on user profile and item attribute features (Yang et al. 2019). Also, another study implemented two joint deep-learning models to predict click and conversion rates (Xu et al. 2020). The architecture of a Deep Neural Network relies on an input layer, hidden layer(s), and an output layer connected through neurons. The input layer receives the training data as input and randomly assigns weights for all the neurons present in the next hidden layer. Also, the weighted sum of each neuron passes through a non-linear activation function. The role of the activation function is to determine whether the neuron should be activated. This process continues for as long as there are hidden layers. While reaching the output layer, the final output is compared with the actual label target using a loss function to measure that margin error. Hence, the model performs backpropagation that aims to improve the final output based on the given error to recalculate the weights. This architecture enhances many advantages due to its compound layers that can exploit non-linear interactions and execute better importance feature selection.

From all aforementioned click prediction models, there is a tendency towards deeper learning architectures since they are more effective at learning about feature relations.

2.3. Interpretability

From a machine learning perspective, performance metrics are the only existing tools that a model relies on when making a decision. Nevertheless, performance metrics might not be enough to evaluate how well a black model performs and be misleading in a multifaceted real world. The more complex a model becomes, the harder it is to understand the reasoning process behind the predicted output. Many links between inputs and outputs can be broken during the training process, as well as having different biases leading to incorrect and unfair predictions (Zablocki et al., 2022). As the deployment of machine learning systems became omnipresent in almost every decision-making process, explainable AI gained visibility in order to overcome bias and incompleteness pitfalls (Doshi-Velez and Kim 2017). According to Doshi-Velez et al. (2017), an explanation is a human-interpretable logic of the steps taken by a decision-maker to reach a specific result after considering a set of inputs. Nowadays, General Data Protection Regulation has developed a crucial role in Explainable AI, requiring a transparent approach when processing data and having human supervision on complex black models (Greco 2019, 49). Having an explainable model is essential to achieve an equilibrium among trust that the predictions are correct, predict an understandable behavior and finally improve potential mistakes. As depicted in Figure 1, interpretability can be reached by two approaches depending on the model's transparency. Either Model Transparency in which the model is intrinsically interpretable as a product of model training, or Post-hoc Interpretability in which interpretability is achieved via external methods.

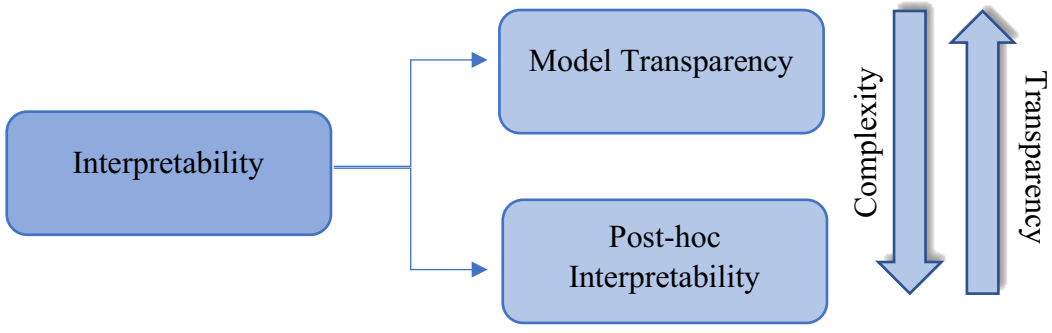


Figure 1 – Taxonomy of interpretability showing the balance between model's complexity and transparency.

According to Freitas (2014), decision trees and linear regressions are considered interpretable since the low complexity of the model is associated with its transparency. In comparison, neural networks are considered black-box models with a high level of complexity. In an advertisement click prediction context, marketers complain about the lack of trust and reliability black box frameworks have to fundament their operational decisions. Consequently, it is increasingly common to use post-hoc explainer algorithms to address these challenges. On the scope of interpretability, the explanations can be local or global-centered. Local explanations aim to explain a specific decision or output, whereas global explanations focus on understanding the general patterns for all outcomes of the model (Doshi-Velez and Kim 2017).

LIME

Locally Interpretable Model-Agnostic Explanations known as LIME is a local explainer algorithm useful for understanding selective data points. After the modeling process, LIME examines the changes in predictions when different versions of data are fed into the model. Then, it creates a brand-new dataset made up of altered samples and the respective predictions given by the model. The explainer develops a model with the created dataset, which measures the local fidelity based on how close the sampled examples are to the related actual predictions (Molnar 2019). Equation 2 clarifies the intuition of LIME that locally explains the data point *x* by minimizing the loss function and adding the model complexity.

$$explanation(x) = arg \min_{g \in G} L(f, g, \pi_x) + \omega(g)$$

Equation 2 – LIME explanation for a single data point.

Shapley Values (SHAP)

SHAP is a model-agnostic method that produces feature summary statistics and visualization used for global and local scale explanations. In its origins, Shapley value was used in cooperative game theory to determine how to fairly distribute the benefits and costs among a group of cooperating agents. Nowadays, when applying to model predictions, agents are the features, and outcomes of the coalition are the predictions. So, Shapley is a feature attribution method that explains the contribution of each variable to the outcome called Shapley value (Molnar 2019).

To calculate the Shapley value of a single variable (*i*) is computed the average marginal attribution of the variable to all potential coalitions. When considering all the variables (*N*), is computed the weighted average of all marginal attributions to all potential coalitions (*S*). (Jong 2021, 15)

$$\emptyset_i(\gamma) = \frac{1}{|N|} \sum_{S \subseteq N \setminus \{i\}} \binom{|N| - 1}{|S|}^{-1} (\gamma(S \cup \{i\}) - \gamma(S))$$

Equation 3 – The Shapley value as the contribution of player (*i*) to the value averaged of all possible coalitions *S*.

The major advantage of SHAP is that guarantees a solid foundation because is the only method that fulfills the properties of Dummy, Efficiency, Symmetry, and Additivity (Jong 2021, 15).

- Dummy: A variable that does not change the predicted value must have a Shapley value of 0.
- Efficiency: The variable contributions must total the difference between the prediction and the average prediction.
- Symmetry: The contributions of two variables' values should be the same if they contribute equally to all possible coalitions.
- Additivity: Considering more than one prediction function, the sum of Shapley values can be calculated either in each prediction or using both prediction functions.

As machine learning models are trained on all features is worthy to estimate a complete distribution of the prediction among all feature values.

3. Methodology

3.1. Dataset Description

ADS Dataset is a publicly available benchmark for computational advertising design by Giorgio Roffo and Alessandro Vinciarelli in 2016. The data consists of an advertising portfolio containing twenty advertisement categories (Table 1). Each category contains fifteen advertisements equally divided between Rich Media Ads, Image Ads, and Text Ads. Text ads

refer to only text format, Image Ads refer to image format, whilst Rich Media Ads are a combination of both.

CategoryID	Category Name
1	Clothing & Shoes
2	Automotive
3	Baby
4	Health & Beauty
5	Media
6	Consumer Electronics
7	Console & Video Games
8	Tools & Hardware
9	Outdoor Living
10	Grocery
11	Home
12	Betting
13	Jewelery & Watches
14	Musical instruments
15	Stationery & Office Supplies
16	Pet Supplies
17	Computer Software
18	Sports
19	Toys & Games
20	Social Dating Sites

Table 1 – The distinct advertisement' categories present in the dataset with their respective category ID.

The raw dataset is also composed of personal information, individuals' preferences, and the Big Five Inventory questionnaire of 120 unacquainted individuals. The participants were selected through a public platform on a first come basis. Regarding personal information was collected their age, gender, type of job, weekly working hours among full-time and part-time status, level of monetary well-being, and CAP/Zip-Code. Some variables such as the name and e-mail could compromise the privacy of participants or cause response bias, therefore were ensured to be kept anonymous. The participants revealed their favorite choices on music, websites visited, watched films, watched TV programs, books, sports, and pastime activities. In addition, they disclose their past travel destinations.

All the participants were submitted to Big Five Inventory (BFI-10) questionnaire, in which they completed a 10-item scale from -2 (Strongly Disagree) to 2 (Strongly Agree). The test is based on BFI-44 items retaining all the significant levels of trustworthiness and validity (Rammstedt and John 2007). It intends to measure the following five dimensions of personality: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism (Gosling et al. 2003).

The participants submitted some images that they like (Positives) and some others that have aversion to (Negatives) from a repository gallery, along with some labels to describe the content of each image.

Finally, each user was exposed to each of the 300 advertisements with the aim of rating on a numerical scale from 1 (worst feedback) to 5 (best feedback).

3.2. Proposed Methodology

The proposed methodology for click prediction and interpretation is illustrated in Figure 2. Firstly, the ADS dataset was subject to a data cleaning process to ensure the reliability and quality of the model. Generally, this stage involves transforming data when it is not structured to be processed, removing data when there are non-random null values, outliers, or even extraneous variables and exploring pertinent connections among data. The final consolidated data was randomly partitioned into three subsets of data: Training, Validation and Test through the *train_test_split* package on an 80:10:10 ratio.

Considering that the data structure has two indexes - for each individual ID, there are 300 Advertisement IDs rated - the split was made by grouping the individuals as one. Thus no individual information is duplicated in more than one dataset. The training dataset is composed of 96 random individuals linked with their 300 rate responses, while both the validation and testing datasets are composed of 12 random individuals each linked with their 300 rate responses. Training comprises the largest slice of data used to train the model and define appropriate weights and biases. Validation employed a small proportion of data to the trained model fit in order to deliver an unbiased evaluation valuable to tune hyperparameters and avoid overfitting. The remaining portion of the data is on the Test dataset used to predict unlabeled observations to compare them with the actual values later.

The training dataset was fed to a fully-connected neural network during the modeling process. In the second stage, the trained model is passed to an explainer method. The use of KernelSHAP method was intended to obtain global explanations of the model and learn the feature importance and contribution to the final Click classification.

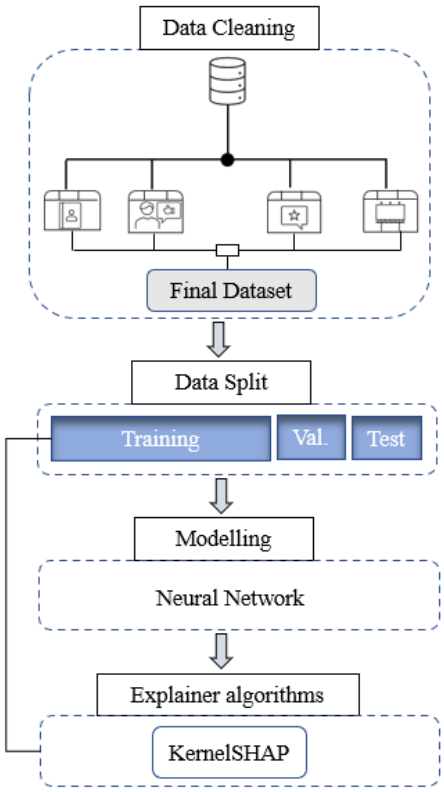


Figure 2 – Outline of the proposed methodology. [1] The data, separated into 4 subsets, was subject to data cleaning process before merging to the final dataset. [2] The final dataset was split into 3 subsets of data on an 80:10:10 ratio. [3] Modeling process of a Neural Network framework to predict *Click* on advertisements using the split data to train and assess its performance. [4] Use KernelSHAP explainer method to learn feature importance on *Click* classification.

3.2.1. Data Preparation

Given the raw data description presented, four datasets were formed: Individuals' information, Individuals' preferences, Advertisements and Ratings. The four mentioned datasets were preprocessed separately and merged at the end when suitable for further analysis with all the variables presented in Table 2.

Dataset	Variables	Description
Individuals' information	Age	Indicate individual's age. [Numeric]
	Gender	Indicate individual's gender. [Dummy]
	Type of Job	Individual's job. [Categorical]
	Weekly Working hours	Individual's working schedule. [Categorical]
	Income	Individual's income level. [Categorical]
	Timepass	Individual's favorite hobby. [Categorical]
	Number of Countries visited	Indicate the number of countries the individual visited. [Numeric]
	O_Score	Indicate the score level of individual Openness. [Numeric]
	C_Score	Indicate the score level of individual Conscientiousness. [Numeric]
	E_Score	Indicate the score level of individual Extraversion. [Numeric]
	A_Score	Indicate the score level of individual Agreeableness. [Numeric]
	N_Score	Indicate the score level of individual Neuroticism. [Numeric]
Individuals' preferences	Fave Sports	Individual's favorite sports. [Text]
	Most listened musics	Individual's favorite type of music. [Text]
	Most read books	Individual's favorite books. [Text]
	Most watched movies	Individual's favorite movies. [Text]
	Most watched tv programs	Individual's favorite tv programs. [Text]
	Most visited websites	Individual's favorite websites. [Text]
	Pos[1-5]	Variables that indicate the description of images submitted as positives. [Text]
	Neg[1-5]	Variables that indicate the description of images submitted as negatives. [Text]
Advertisements	Category Name	Category of advertisement. [Categorical]
	Type of Ad	Type of advertisement. [Categorical]
Ratings	Rating	Scale feedback on advertisements [Ordinal]
	Click	Indicate a 'Click' response. [Dummy]

Table 2 – Data dictionary of the final dataset. Identification and description of the variables used, classified according to domain dataset and datatype.

Individuals' information dataset

On the Individuals dataset, all the variables that undermine the participant's anonymity were deleted [Name, Last Name, PayPal, Cap/Zip-Code]. To clarify the influence categorical variables have on the click response and assuming no natural ordering relationship among the categories within, these were converted to dummy variables through a one-hot encode to construct their attributes [Type of Job, Weekly working hours, Timepass]. In addition, the categorical variable countries visited was converted to a numeric one that captures instead the number of countries visited. To do so was used a word counting approach on the original variable.

Apart from the categorical variables, the responses provided to the 10-item personality questionnaire were handled following the official measurement score, and later grouped by personality trait. The final score ranges from 0 to 10, which was respectively allocated to the five personality dimensions: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism.

Individuals' preferences dataset

Mislabeled columns were detected, for both positive and negative image labels, with multiple missing values. These columns were irrelevant to the classification problem, thus were deleted from the dataset.

Word Embedding

The individuals' preferences are described in text form, as well as the positive and negative image tags chosen by each user. As unstructured data has no linear pattern, it is consequently difficult to be interpreted by a machine learning model. A way to transform unstructured to structured data is to use Natural Language Understanding (NPU), a subset of Natural Language Processing. NPU considers the input text of individuals' preferences and image tags to capture key understandings into machine format. The technique used to translate text into numerical vectors is called word embedding representation. This work used an open source pre-trained model GloVe: Global Vectors created by Stanford University. GloVe is trained to understand the context of words by observing word-word cooccurrence probabilities ($X_{i,\,j}$) on a broad corpus with around 6 billion words of English vocabulary, including punctuation, to create an encoding featurized representation for each word of the vocabulary (Greco 2019, 18-22). This

method enables the model to automatically learn analogies and discrepancies between different words by varying the representation of similar words (Drozd et al. 2016). A simplification of the model is denoted in Equation 3 where *w* are word vectors, $\widetilde{w}$ are context word vectors and *b* represents the bias needed to restore the symmetry in both word and context word vectors (Pennington et al. 2014).

$$J = \sum_{i,j=1}^{V} f(X_{ij})\ (w_i^T \widetilde{w}_j + b_i + \tilde{b}_j - logX_{ij})^2$$

Equation 4 – GloVe embedding model.

The first stage of the word embedding process is Tokenization – taking a string and break it down into *n* tokens. Followed by the creation of the vocabulary containing all the unique words present in the considered variable. From the GloVe library, was used a file with a 6Billion corpus size and with a 100-dimension dense vector, meaning that each word in the file has a correspondent vector of size 100. Then, an embedding matrix was constructed with the vectors from GloVe that had a corresponding word match on the unique variable vocabulary at their respective position. To complete the text processing was used pre-trained word embedding model to train the data having the Click variable as the target direction. The transformation resulted in a vector-based dataset that aims to capture words' semantical and contextual significance for the Click variable.

Advertisements Dataset

The advertisement dataset combined all 300 advertisements displayed for the participants. The dataset was enriched by adding their numeric corresponding category and information regarding the type of ad (Text, Rich, Image) was also included in dummy input form.

Rating Dataset

At last, Rating Dataset gathers the feedback from the participants rated on a scale of 1 to 5, being the best possible feedback. Considering Giorgio Roffo and Alessandro Vinciarelli's paper (2016), they have stipulated that values 4 and 5 represent a Click on the exhibited advertisement, and all the remaining rates represent a 'No Click' response. Accordingly, a dummy variable Click was added where 1 indicates the existence of Click and 0 the opposite effect. This variable is the dependent variable for the classification problem prediction.

In the end, all the datasets were merged into a singular data frame composed of two index variables: the userId and the AdId, totaling 36000 observations. All variables present non-missing values, except for one participant's description of positive/negative image tags. The data were missing at random which would not affect the prediction results, so its elimination was not justified. Once an in-depth examination of the variables using *pandasProfileReport* package, no relevant outliers and no duplicated rows were identified. The collection of data was static in time and consequently is not required to control for the unobserved fixed effects.

3.2.2. Exploratory Data Analysis

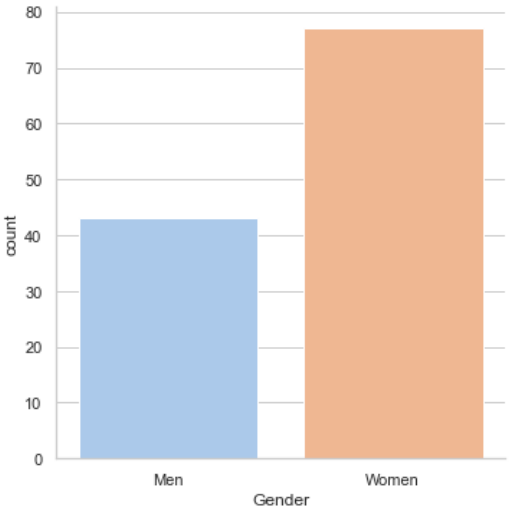
From all the 120 participants, 77 are women (64,17%) and 43 are men (35.83%) with an age range between 18 and 68 (Figure 3-A). The age distribution is skewed to the left, as observable in Figure 3-B, meaning that the majority of the participants were young adults with an age range between 18 and 35 in both genders.

There are eight types of jobs identified from which Contract Employee (32.5%), Student (30.83%) and Temporary (15%) categories amount to 78.33% (Figure 3-C). Similarly, there is an uneven distribution for the working hours whereby approximately 66% of the participants work full-time and the remaining on a part-time schedule (Figure 3-D).

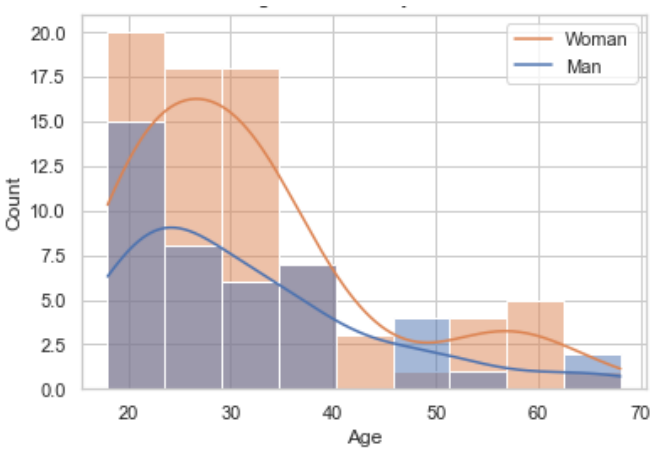
Regarding participants' time pass, the Internet is the general favorite choice followed by much lower interest in reading, sports, music, and movies (Figure 3-F). Also, men demonstrate more interest in sports and women in reading when compared to the opposite gender.

Almost half of the individuals have an annual income from USD 11k to USD 50k. The remaining percentage of income level is distributed by 23,33% for yearly income lower than USD 11k, 21,67% for salaries from USD 50k to USD 85k, and lastly 7,5% for income greater than USD 85k ((Figure 3-E). Regarding their personality traits, as presented in Figure 3-G, no substantial difference was observed among them. The medium scores showed that, on average, the participants stood out in all the traits studied, especially Openness to experience, having a median score of 8.

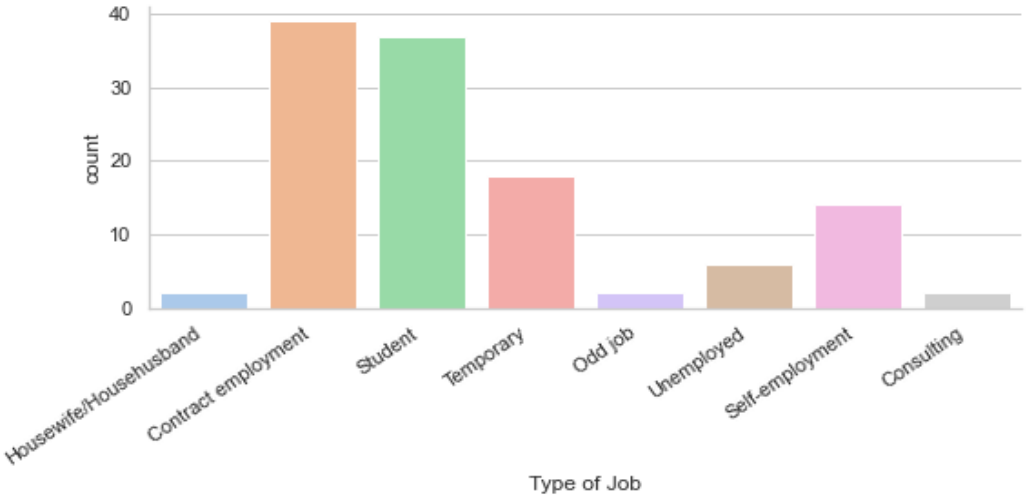
(A) Gender distribution



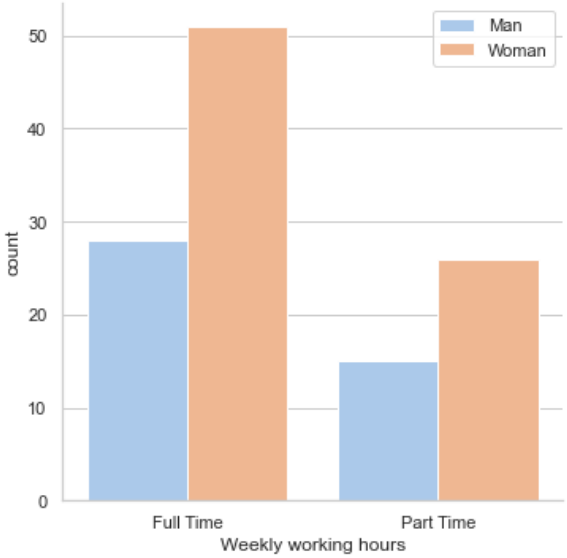
(B) Age distribution by gender



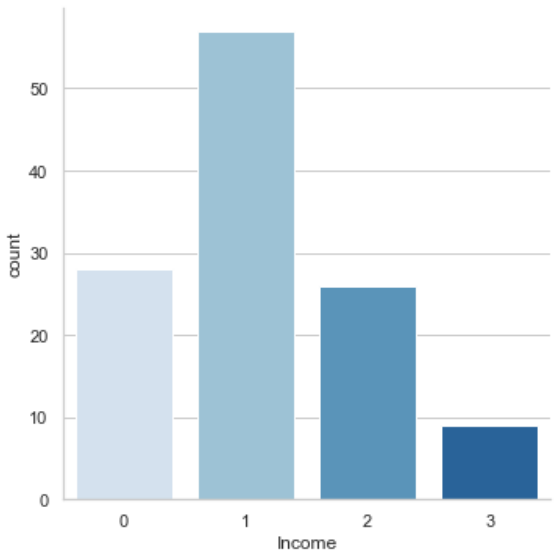
(C) Distribution of type of job



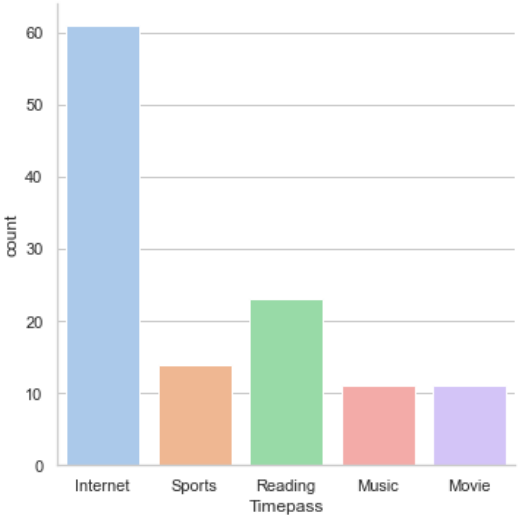
(D) Working schedule by gender



(E) Level of income distribution



(F) Distribution of favorite timepass



(G) Distribution of personality score

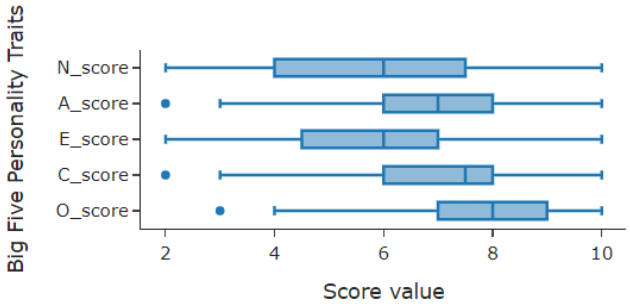


Figure 3 – Exploratory analysis of participants' personal information and preferences. (A) Bar plot showing uneven gender class distribution whereas women are predominant. **(B)** Histogram of age distribution showing a skewed left tendency. **(C)** Bar plot showing that contract employee and student are the type of job that most participants have. **(D)** Bar plot discriminated by gender that full-time working schedule in predominant among participants. **(E)** Bar plot showing that level 1 of income which represents an annual salary range from USD 11k to USD 50k is the prevalent among participants. **(F)** Internet is the favorite timepass. **(G)** Boxplot representing the score distribution of each big five personality trait.

As observable in Figure 4, there is an uneven distribution of the target class's observations. Having an imbalanced dataset is common in click classification problems known for low CTR. In this scenario, the overall CTR is 13,45% meaning that there were registered 4841 clicks. Considering the type of advertisement, there is no meaningful percentual difference among them. Image advertisement stands in front with a CTR of 15,8%, followed by rich and text advertisements with a CTR of 12,9% and 12,1% respectively.

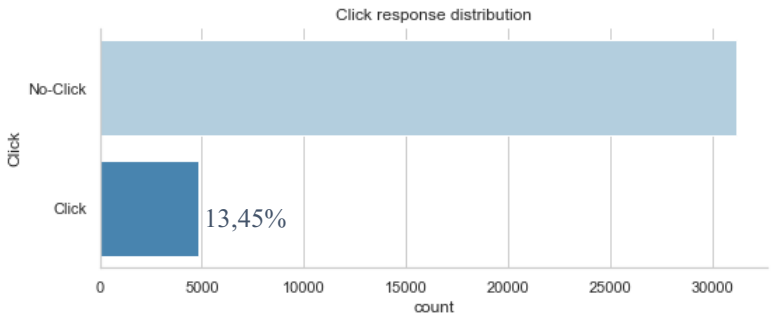


Figure 4 – Bar plot of click response distribution showing a sharp class imbalance.

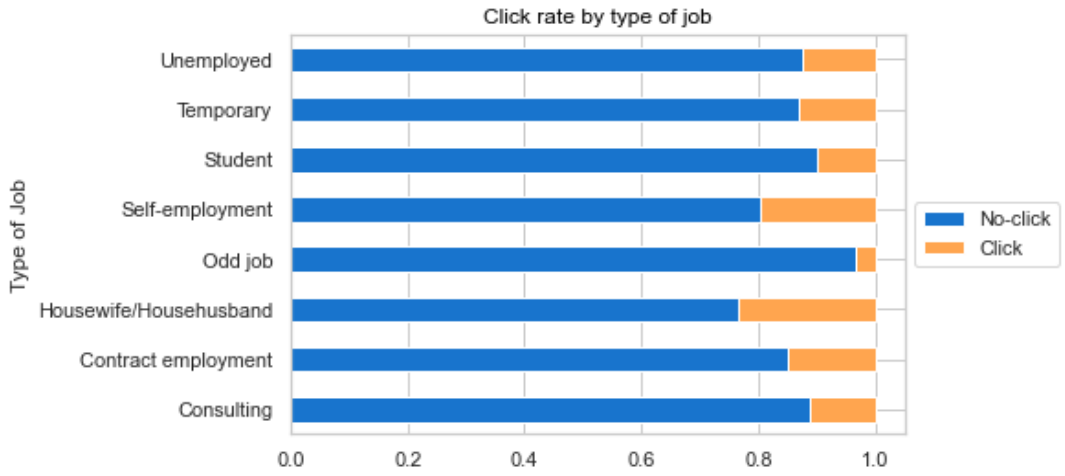
Even though the ads were uniformly distributed in all categories, the clicks were more prominent in the following categories: Consumer Electronics, Console & Videogames, and Groceries, as noted in Table 3. On the contrary, Social Dating Sites, Betting, Baby, and Tools & Hardware are the category where the participants click the least.

Category Name	Category CTR
Clothing & Shoes	17%
Automotive	10,6%
Baby	8,1%
Health & Beauty	14,8%
Media	19,3%
Consumer Electronics	20,6%
Console & Video Games	19,9%
Tools & Hardware	9%
Outdoor Living	14,7%
Grocery	18,8%
Home	12,1%
Betting	4,6%
Jewelery & Watches	15,8%
Musical instruments	9,5%
Stationery & Office Supplies	15,3%
Pet Supplies	9,3%
Computer Software	16,7%
Sports	14%
Toys & Games	14%
Social Dating Sites	4,5%

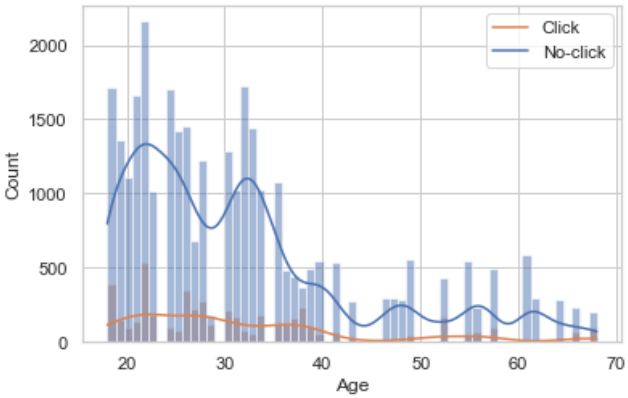
Table 3 – Click-through rate per advertisement's category in the dataset. Consumer Electronics has the higher percentage of clicks, contrasting with Social Dating Sites that has the lowest rate.

In a further analysis, it was concluded that Housewife/Househusband and Self-employment jobs are associated with a higher percentage of clicks and are more occupied by women (Figure 5 -A). In addition, Figure 5-B shows that age bucks behave differently concerning clicks on an advertisement. The tendency is a higher click rate in young adults but accompanied by considerable volatility, whilst older participants seem to be more consistent and give less positive feedback responses. Considering their personality traits, participants who click seem slightly more associated with Agreeableness and Conscientiousness. Comparing to the remaining personality traits appear to have a medium score and minimum slightly lower than, on average, people that do not click.

(A) Click rate by types of job



(B) Click distribution by age



(C) Personality score by Click response

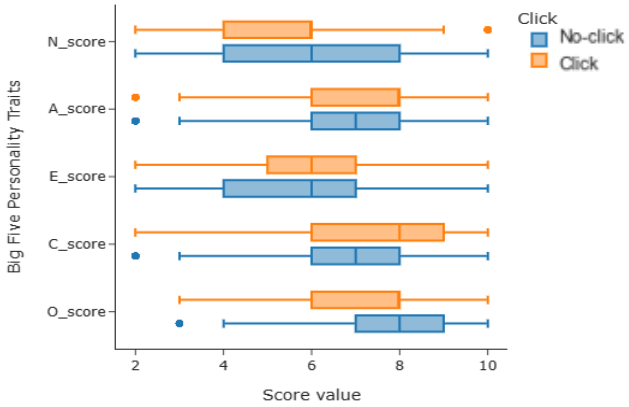


Figure 5 – Exploratory analysis of click rate. (A) Bar plot of click rate distribution by type of job that indicating that housewife/househusband is the job with higher click rate. **(B)** Histogram of age showing volatility in click response and young adults have, on average, a higher click percentage. **(C)** Boxplot showing how personality scores are distributed according to click response.

3.2.3. Dataset limitations

One of the most relevant limitations is associated with the sample size. When not having a representative sample, small connections among data might be neglected as well as placing too much weight on others. The research being conducted is based on only 120 individuals and 300 different advertisements, and working with a small sample could be more challenging to induce significant relationships and statistical inferences.

The lack of a positive click response is one of the biggest hurdles for Click predictions. As a positive reaction is so scarce, it raises a class imbalance concern. In fact, statistics reveal that CTR is no greater than two percent across all exhibiting advertisements (Gharibshah and Zhu 2021). Hence, is no different for the ADS dataset being employed in this work.

3.2.4. Modelling

In the modeling process was implemented an Artificial Neural Network (ANN) with the *Keras Tensorflow package* to model the Click prediction properly. As illustrated in Figure 6, both training and test datasets have similar click balance distributions.

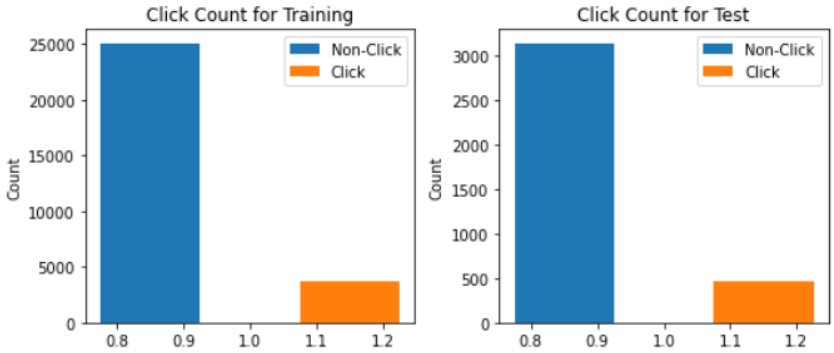


Figure 6- Click distribution on Training and Test datasets. The plots show there is a proportional ratio among the two subsets of data.

In practice, the best model design is composed of layers. Firstly, the input layer was fed with 55 standardized and flattened predictor features of the Training dataset. The weights were randomly assigned to 150 neurons latent in the first hidden layer and were passed through ReLu activation function. Posteriorly to the second hidden layer with 30 neurons activated again by

ReLu function. ReLu function ranges between zero attributed to neurons with negative values and positive values to the neurons that are not negative. When using ReLu, saturation and vanishing gradient only occur for negative values. Lastly, the weights and biases were passed to the output layer with the sigmoid activation function. The sigmoid function, common in classification problems, will normalize the output neuron to a probability score between [0, 1] interval of the input predicted click label. At the end of every epoch, a Binary Crossentropy function was used to measure the suitability of the prediction. Indeed, the function determines how distant from the actual value the prediction is, and also contributes to adjusting the weights on the backpropagation process. Allied with the loss function, Adam optimizer was used to update the weights in the right direction based on the learning rate. Adam algorithm accumulates the gradient of the past steps to determine the direction to go. This process was repeated for 30 epochs. The metrics chosen to evaluate the model performance were accuracy and, most importantly, Recall, Precision and F1-Score, represented in Table 4.

Performance Metric	**Formula**
Accuracy	$\frac{Correct\ Predictions}{Total\ Predictions}$
Recall	$\frac{True\ Positive}{(True\ Positive\ +\ False\ Negative)}$
Precision	$\frac{True\ Positive}{(True\ Positive\ +\ False\ Positive)}$
F1 Score	$2 * \frac{Recall\ *\ Precision}{Recall\ +\ Precision}$

Table 4 - Metrics used to evaluate the model performance.

As previously noted, the advertisement categories with higher click-through rates were Consumer electronics, Console & video games, and Grocery. From the final dataset, three sub-datasets were created by filtering the category ID with the corresponding highest click-through rate categories. For each category mentioned above, a neural network model was performed to observe similarities and differences between categories and the overall model. The data was split in the same 80:10:10 proportions. In Table 5 is presented the model architecture for every category dataset.

Category Model	
Model	Input Layer: 55, ReLu activation function Hidden Layer: 250, ReLu activation function Output Layer: 1, Sigmoid activation function
Loss Function	Binary Crossentropy
Optimizer	Adam

Table 5 – Neural network architecture of the 3 highest rate categories: Consumer electronics, Console & video games, and Grocery.

3.2.5. KernelSHAP

In the second stage, was proposed to interpret the probability scores given by the model using a post-hoc method to understand which features have the most impact on classification labels. To do so, SHAP provided explanations of feature importance and contributions. The challenge with computing Shapley values for all features based on a whole model is the requirement of sampling the coalition values for each possible feature permutation, which in a model explainability setting with 55 features means having to evaluate the model millions of times. To overcome this issue, Lundberg and Lee devise the Shapley kernel, a mean of approximating Shapley values through much fewer samples. Kernel Explainer SHAP is a model-agnostic implementation that combines linear LIME and Shapley values (Lundberg and Lee 2017). Also, among every possible forms of using SHAP, Kernel is the one that has universal applicability to any machine learning framework. The process consists in passing samples of various feature permutations through the model for a particular data point. As a neural network will not omit features, instead it will define a background dataset that contains a set of representative data points the model was trained over. Then, fill the omitted feature/features with values from the background dataset while holding the features that are included in the premutation fixed to their original values. Afterward, is computed the average of model output over all of these new

synthetic data points. Once, having several samples computed in this way, these are fitted in a weighted linear regression with each feature assigned a coefficient – the Shapley value.

In practice, KernelSHAP can be used to make local and global explanations. Local explanations focus on estimating feature contributions for a singular individual, represented as forces in a plot that either increase or decrease to balance the prediction value. Whilst global explanations put together the feature contributions for all the individuals and can be visualized through a summary feature importance plot. In the light of the work, was only valuable to exploit global explanations wherefore were performed not only a feature importance plot but also a summary plot to evaluate its magnitude (Molnar 2019). To generate the mentioned plots were used 100 samples from the Training dataset, previously randomized to collect as much differentiated information as possible.

4. Results

As illustrated in Figure 7, there are two types of errors (I and II) in statistical predictions. Type I error – falsely predict a positive class – happens when the model decides to target as click an interaction that was not a click, whilst Type II error – falsely predict a negative class – happens when the model decides not to target an interaction that was a click. These errors are inverse measures dependent on the threshold value, meaning there is a tradeoff between errors I and II. The choice of threshold value must be defined by maximizing the type of error that is less impactful in the overall problem domain.

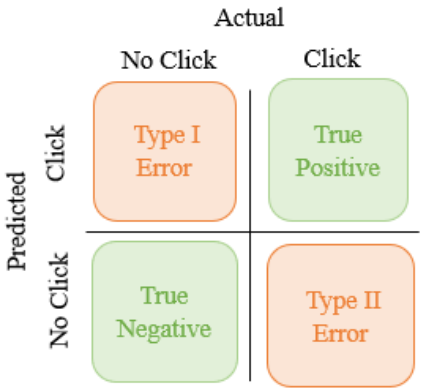


Figure 7 - Confusion matrix representing the two types of error. Type I error refers to false positive labeled class and type II to false negative labeled class.

The probabilities scores obtained were assigned to Click or No Click labels by the default threshold of 0.5. Considering that the primary goal of a marketer is to correctly predict the minority class of Click because a wrong prediction results in a lost consumer target. Thus, the cost of a false negative is higher than the cost of a false positive, which lead to optimizing the threshold by maximizing the recall metric. In Figure 8, a graph of threshold tunning aids in visualizing the minimum error value is a threshold of 0.15.

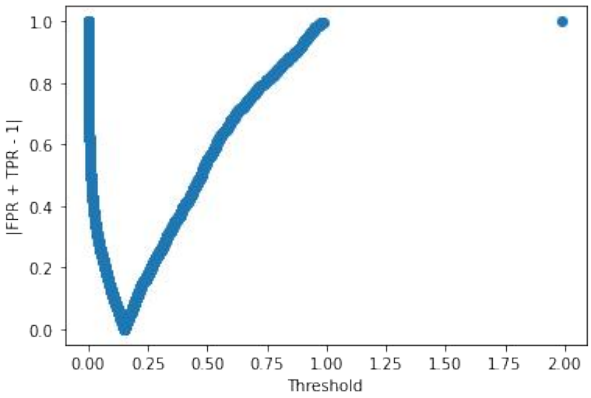


Figure 8 – Threshold tunning plot. The threshold value that minimizes the false positive rate is 0.15.

The plots of ROC curve and Precision-Recall Curve, depicted in Figure 9-A and 9-B respectively, represent visually the performance metrics used with the tuned threshold.

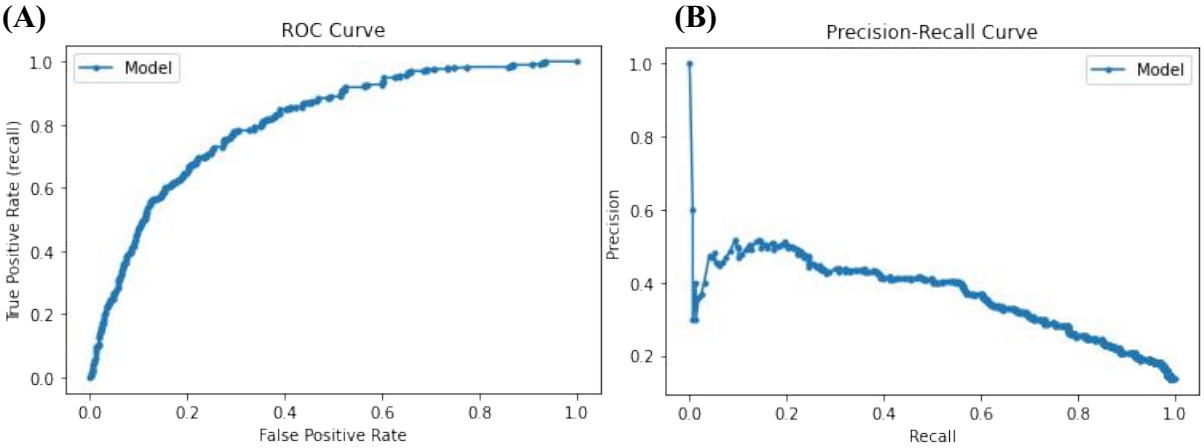


Figure 9 - Plots representations of ROC Curve and Precision-Recall Curve. (A) A ROC Curve summarizes the performance of the true and false positive. As the baseline curve is approximating the top-left corner means that the model has an overall good performance. The closer the curve is to the top-left the better performance. **(B)** A Precision Recall Curve summarizes the performance of the precision and recall, focusing on the minority class.

Having the optimized threshold value (= 0.15), the performance metrics results are represented in Table 6, and are detailed below:

Performance Metric	Value
Accuracy	83,94%
Precision	41,78%
Recall	60,39%
Specificity	87,46%
F1 Score	49,39%

Table 6 – Results of performance metrics for the global model.

- The model can correctly classify whether the participant clicked or not on the advertisements in around 84% of Test observations.
- From all the predicted click observations, the model could correctly predict 41,78% of those observations.
- From all the actual clicks, the model could predict correctly 60,39% of those observations.
- From all the actual no clicks, the model could predict correctly 87,46% of those observations.
- F1 Score consider both Precision and Recall to measure how well the model is performing accurately. Given the low Precision and moderate recall level, the F1 Score is also low 49,39%.

The performance of the proposed categories prediction models are described in Table 7:

Performance Metric	Consumer Electronics	Console & Video games	Grocery
Threshold	0.3	0.25	0.32
Accuracy	83,88%	86,11%	78,33%
Precision	57,14%	50%	44,74%
Recall	58,82%	60%	48,57%
Specificity	89,73%	90,32%	93,1%
F1 Score	57,97%	54,54%	46,58%

Table 7 – Results of performance metrics for the categories models.

The metrics indicate that models of Consumer electronics and Console & Video games categories perform better than the overall model. Indeed, these categories present slightly better results for accuracy and moderate better results F1 score metric, concluding that the percentage of correct predictions increased due to a decrease in both false positive and false negative rates. Despite minor differences and considering the click rate variations among categories, Consumer electronics and Console & Video games models present similar performance metrics as the categories sell similar products. Regarding Grocery category, it was noticed a decline in accuracy and F1 Score metrics comparing to all other mention models. Grocery's model capture better the true No Click responses as demonstrated in the Specificity metric.

The output of KernelSHAP involves the following plots (Figure 10-A, Figure10-B) reflecting each feature's contribution to the prediction made by the neural network.

(A) **(B)**

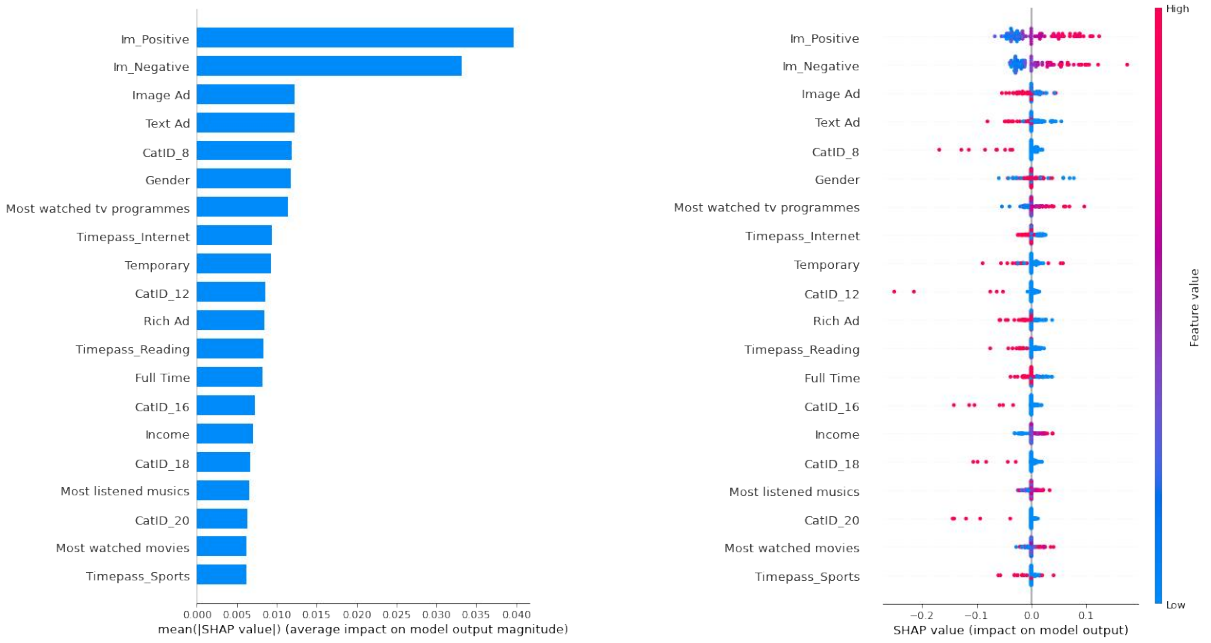


Figure 10 – SHAP values output on the overall neural network model by using KernelSHAP method. (A) Summary plot displaying the importance each feature has on changing the *Click* prediction. **(B)** Summary plot displaying the magnitude each feature importance has on the *Click* prediction.

As implied in the plots, both positive and negative image description labels are the most important features changing the predicted absolute click probability. On average, the variable Im_Positive change the prediction outcome by 4 percentage points, followed by Im_Negative which changes, on average, by 3.3 percentage points. Besides these variables, the type of

advertisement (image or text) has also a change impact on the outcome by around 1.3 percentage points each. When considering the magnitude of the variables, high standardized values of image positive and negative labels suggest an increase in the predicted probability score. Regarding the type of advertisement, their influence is equally distributed, yet rich and image advertisements present a lower concentration of negative values for a positive contribution to the prediction.

Moreover, it is clear that categories 8, 12, 16, and 20 reduce the predicted probability score, and were the categories observed as the least clicked: Tools & Hardware, Betting, Pet Supplies, and Social Dating Sites, respectively. On the contrary, if the advertisement was not associated with the mentioned categories, the mean Shapley value is positive, reflecting an increased contribution to the prediction output.

Evaluating KernalSHAP for categories' models, the following plots were obtained:

Consumer Electronics

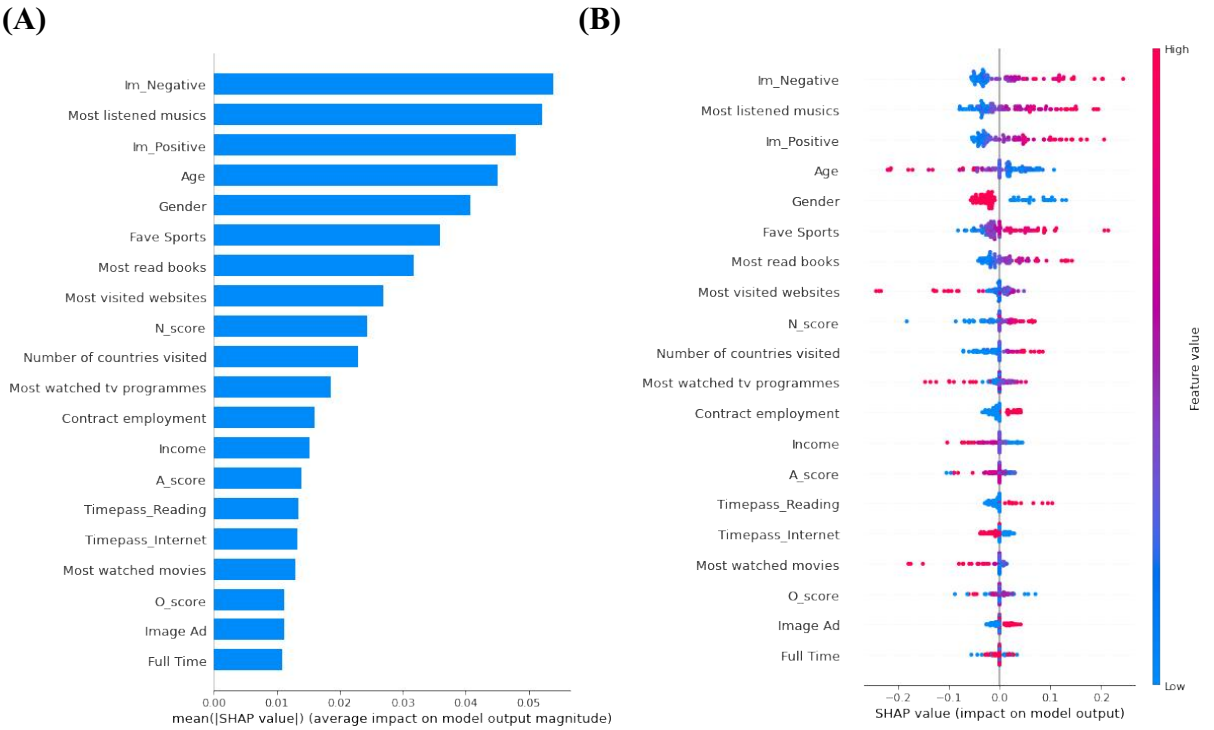


Figure 11 - SHAP values output on the Consumer Electronics neural network model by using KernelSHAP method. (A) Summary plot displaying the importance each feature has on changing the *Click* prediction. **(B)** Summary plot displaying the magnitude each feature importance has on the *Click* prediction.

Console & Video Games

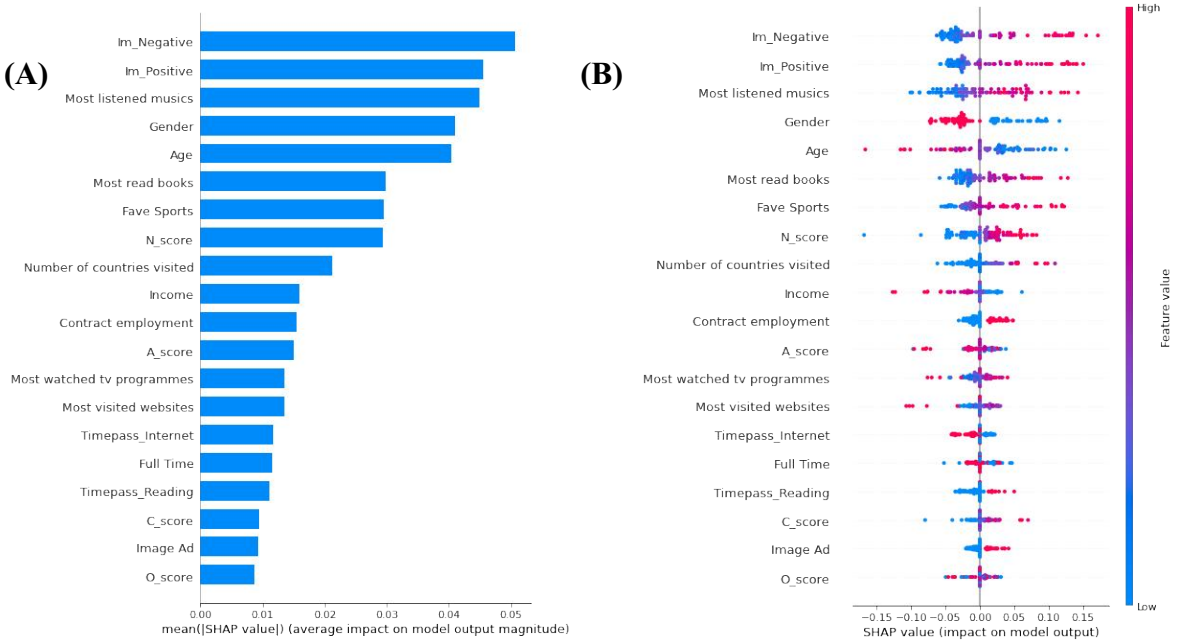


Figure 12 - SHAP values output on the Console & Video Games neural network model by using KernelSHAP method. (A) Summary plot displaying the importance each feature has on changing the *Click* prediction. **(B)** Summary plot displaying the magnitude each feature importance has on the *Click* prediction.

Grocery

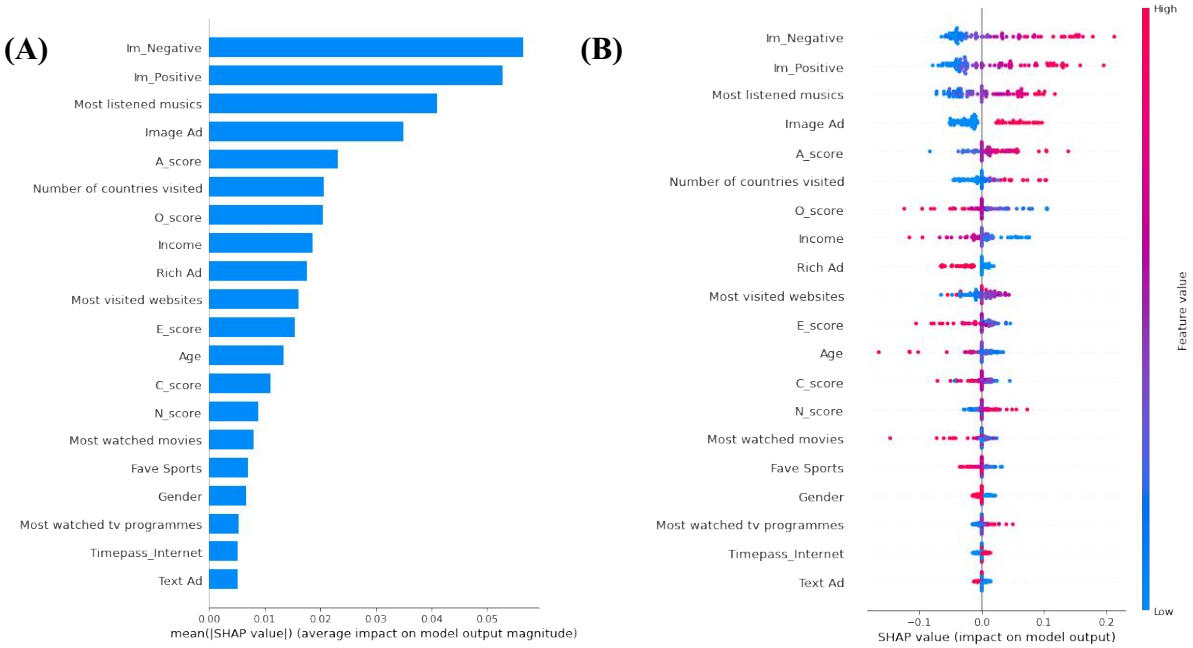


Figure 13 - SHAP values output on the Grocery neural network model by using KernelSHAP method. (A) Summary plot displaying the importance each feature has on changing the *Click* prediction. **(B)** Summary plot displaying the magnitude each feature importance has on the *Click* prediction.

Similarly to what was perceived in the overall model, the variables that have the most impact contributions on changing the predicted outcome are the negative and positive labels attributed to some images. Conversely, the most listened music gained relevance in all category models. In general, high standardized values of negative and positive labels as well as most listened music contributed to an increase in the predicted probability score. It is also noted a similar distribution of Shapley values among consumer electronics and console and video games categories that consider age, gender and favorite sports as important features. Whereas the Grocery category standout the importance of image advertisements, Agreeableness and Openness personality traits, number of countries visited, and income level. In both electronics and Console & video games, being from the male gender and having a younger age increases the predicted probability of clicking on the advertisement. Regarding the grocery category, participants who consider themselves more Agreeableness but less open to new experiences increase the probability of click. Also, the plot indicates that there is a higher probability of clicking if the advertisements have an image and if the participant has low levels of income.

5. Discussion

This thesis aimed to build a black-box model for advertisement click prediction whereby marketers could trust and rely on it as a decision-making tool to personalize and target their audience better. To answer the proposed research questions, a neural network model was developed, as well as KernelSHAP explainer to interpret the model. Global feature importance statistics were extracted, and the key variables determinants of the 3 highest CTR advertisement categories were analyzed.

Firstly, the performance metrics of the model suggest the model is predicting correctly a much higher percentage of No Click responses than Click ones. Naturally, due to class imbalance the algorithm has a greater margin of learning by the majority class when compared to much fewer observations on the click class. Thus, the model was mainly trained to recognize No Click patterns leading to a high false positive rate. A technique that would allow to mitigate the class imbalance problem is resampling the target class. The idea behind resampling is to balance both classes of the variable, so no class is dominant over the other. In general, advertising Click prediction problems implement under-sampling, a technique that consists in randomly remove data observations from the majority class. Even though, is one of the most common strategies

to overcome class imbalance it could cause loss of information when some observations are eliminated. One of the ADS dataset limitations is the Click imbalance, however due to the nature of the dataset structure composed of two indexes, the use of the under-sampling technique would not improve the model's performance. Randomly removing data observations of the negative class, will lead to loss of information for specific users when splitting the data through grouping the participant's ID.

As prior mention, false positive classification error is inversely coexistent with the identification of false negative observations. As a result, there is a tradeoff between them to maximize the business' profit. From a business perspective, the cost of misclassification can be perceived in two scenarios: minimize type error I or II depending on the company's resources, strategy, and the price per displayed advertisement. On the one hand, considering a company with low advertisement price or simply a company that recently entered the market and want to spread brand awareness, these type of business want to target as many clicks as possible. Therefore, for them is more important to minimize type II error and guarantee that the rate of false negatives is low. On the other hand, if a company has a limited budget and a high advertisement price is more beneficial to minimize type I error because a false positive would be costly to the company's profitability. In this thesis, the developed model's metrics were tuned to a threshold that minimizes false positive errors generalizing a unique advertising business model. In general, the neural network framework proved a higher performance than Logistic Regression, Support Vector Regression with radial basis function and L2-regularized L2-loss Support Vector Regression present in Roffo and Vinciarelli's paper (2016) using the same ADS dataset. Even though this thesis did not focus on developing additional machine learning frameworks to compare their performances, this piece of evidence suggests that neural network models are more efficient at learning about feature interactions with non-linear patterns.

Overall, marketers should integrate computational advertisement through a machine learning model for advertisement Click prediction having into account their business strategy and resources to choose the minimum cost of misclassification. Also, marketers should adopt the necessary resampling techniques to overcome the class imbalance problem.

Secondly, to get the magnitude and importance of the predictor variables was used KernelSHAP method which provides an accurate approximation of Shapley values based on samples. The results indicate that positive and negative chosen image tags, image and text advertisements are

the features that mainly impact changing the predicted click probability score. Nevertheless, it is important to recall that computational complexity is time costly and seen as a weakness since the efficiency of SHAP relies on a high number of samples and the number of features allocated.

Once examining these features with the respective magnitude, the findings suggest that images and visual stimulus induce reactions that likely increase the probability of click response. As a matter of fact, theoretical background shown evidence that images on advertisements have a positive influence on advertisement effectiveness since they capture the user's attention (Liu and Yu 2022). Nowadays, users spend hours on digital platforms consuming all types of instant information. As a result, their attention is immediately dispersed and it gets more difficult to obtain a Click reaction from the user side. As the SHAP results shown, the relationship between visual content and advertisement attractiveness, measured by the increase likelihood of a click response, is meaningful. Images have the power of influence purchase behavior and disputing emotions, which can be the key difference to have a successful advertisement.

Contrary to what was supported in Giorgio Roffo and Alessandro Vinciarelli's paper, personality traits do not have relevant Shapley values compared to other features. In general, the features that have more impact are individuals' preferences and advertisement characteristics such as their type or category. The results suggest that the Click trigger is not only influenced by user characteristics and personality traits, but also on the content the advertisement delivers. Nonetheless, the data used disregard specific content annotations that each advertisement have, which could be more relevant to measure an advertisement performance than just accounting for its type. For instance, the use of pre-trained machine learning models, such as Cloud Vision API or Microsoft Computer Vision, can boost the performance of the model by capturing visual features and annotations on advertisements.

Finally, the categories model performed slightly better than the overall model. Even though it was only considered the three highest click-through rate categories, the false positive and negative rates were reduced, suggesting that segmentation is important since each category has its feature contribution values. The KernelSHAP statistics indicate that Consumer electronics and Console and video games categories have similar feature importance and magnitude. Besides negative image tags and most listened music variables, both have an increased likelihood of clicking if the individual interacting with the advertisement is a male and belongs in a young adult buck. The similar pattern of feature importance derives from the fact that the two categories sell similar products, hence the desire to target the same audience. In the grocery

category, stands out the importance for image advertisements. A plausible explanation relies on the fact that image advertisements trigger positive stimulation of senses, namely thoughts of previous tastes that enhance the likelihood of a click response.

6. Conclusions

A critical evidence that motivated this thesis was acknowledging how the advertisement process evolved into a multidisciplinary field that comprises a computational perspective. The availability of data regarding consumers' personal information and preferences as well as their historical purchase behavior, enables marketers to understand the choices and needs of these consumers. Therefore, with the support of machine learning algorithms, businesses develop more accurate target strategies to manage their resources better. Furthermore, the interpretability concern was raised among the black box models used as a decision tool in those new target strategies. As a matter of fact, this thesis aimed to design a machine learning framework, and afterward provide explanations to support the classification results for the users' advertisement clicks. The results were obtained through the Neural Network model with the help of KernelSHAP explainer to consistently designate the features which are more significant to decide whether the user clicked or not on the advertisement.

The model correctly predicted 84% of the tested observations. Nonetheless, due to target class imbalance problem, the rate of false positive is high reflecting a low F-Score accuracy of 49,39%. As expected, the tradeoff cost between misclassifying one of the classes suggests that the decision makers should maximize the less harmful error for the business. Moreover, results show that image advertisements and label descriptions of selected positive and negative images from an image repository have high Shapley values meaning that their average contribution to the overall model was higher when compared to other features. An in-depth analysis was done on three advertisement categories (Consumer electronics, Console & Video games, and Grocery) to monotonize the changes in Shapley values. The results showed that the more impactful features remain the same among categories and the general model. Nevertheless, each advertisement category had its own feature importance statistics highlighting different variable contributions. Moreover, categories within the same scope of products present a similar structure of feature importance, as demonstrated in the electronics and console categories that both suggested an increased probability of clicks in the male gender and younger age.

Overall, the thesis introduces an approach toward an interpretable click advertisement prediction model based on the principles of transparency, trust and reliability of the results. Even though a complete interpretation may still be unclear, KernelSHAP provided valuable insights to explain the predictions made globally.

6.1. Future Work

There are a few recommendations to include in future work: (1) to enrich the dataset with advertisement particularities or/and user past click behavior. For example, using Google Cloud Vision API enables gathering information such as the identification of notable entities' logotypes, face recognition to detect emotion likelihood scores and text detection which is converted to machine-coded text. All these feature collections could determine and influence a participant's act of Click. Also, as suggested in Xingquan and Zhabiz's (2021) paper past user behavior is commonly used when trying to achieve an accurate Click prediction. The past user interactions with certain advertisements can be interpretated as a feedback source for the relevance target. The association between web pages and advertisements is recorded in order to use it for collaborative filtering techniques. (2) increase the number of participants to achieve more representative results and mitigate discrepancies across both target classes; (3) explore categories individualities and ensemble in a whole model. According to Shapley values, each category has a different feature importance pattern which reflects the product characteristics as well as the main target it wants to attract. Amid that context, consider an ensemble model with all the twenty distinct categories would synthesize the individual categories' results into a single model. (4) use Contextual Importance and Utility (CIU) interpretability algorithm to explore explanations for a particular result. CIU can provide useful explanations of each feature contribution and its utility value (Malhi et al. 2021). The contextual utility supports the classification inference of the participants who click on an advertisement.

All in all, the thesis enhances the importance of incorporating explanations in a Click prediction model and encourages marketers to integrate these approaches, which will help them improve their results and user experience through more personalized and targeted advertisements for their audience.

References

Baeza-Yates, Ricardo. 2018. “Bias on the Web.” Communications of the ACM 61 (6). Association for Computing Machinery: 54–61. doi:10.1145/3209581.

Blom, Jan. 2000. “Personalization - A Taxonomy.” In Conference on Human Factors in Computing Systems - Proceedings, 313–14. doi:10.1145/633292.633483.

Chapelle, Olivier. 2014. “Modeling Delayed Feedback in Display Advertising.” In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1097–1105. Association for Computing Machinery. doi:10.1145/2623330.2623634

Dave, Kushal, and Vasudeva Varma. 2014. “Computational Advertising: Techniques for Targeting Relevant Ads.” Foundations and Trends in Information Retrieval 8 (4–5). Now Publishers Inc: 263–418. doi:10.1561/1500000045.

Doshi-Velez, Finale, and Been Kim. 2017. “A Roadmap for a Rigorous Science of Interpretability.” ArXiv Preprint ArXiv:1702.08608v1, 1–13. http://arxiv.org/abs/1702.08608.

Doshi-Velez, Finale, Mason Kortz, Ryan Budish, Christopher Bavitz, Samuel J. Gershman, David O’Brien, Stuart Shieber, Jim Waldo, David Weinberger, and Alexandra Wood. 2017. “Accountability of AI Under the Law: The Role of Explanation.” *SSRN Electronic Journal*, November. Elsevier BV. doi:10.2139/ssrn.3064761.

Drozd, Aleksandr, Anna Gladkova, and Satoshi Matsuoka. 2016. “Word Embeddings, Analogies, and Machine Learning: Beyond King - Man + Woman = Queen.” In COLING 2016 - 26th International Conference on Computational Linguistics, Proceedings of COLING 2016: Technical Papers, 3519–30. Association for Computational Linguistics, ACL Anthology.

Freitas, Alex A. 2014. “Comprehensible Classification Models.” ACM SIGKDD Explorations Newsletter 15 (1). Association for Computing Machinery (ACM): 1–10. doi:10.1145/2594473.2594475.

Gharibshah, Zhabiz, and Xingquan Zhu. 2021. “User Response Prediction in Online Advertising.” *ACM Computing Surveys*. Association for Computing Machinery. doi:10.1145/3446662.

Gosling, Samuel D., Peter J. Rentfrow, and William B. Swann. 2003. “A Very Brief Measure of the Big-Five Personality Domains.” Journal of Research in Personality 37 (6). Academic Press Inc.: 504–28. doi:10.1016/S0092-6566(03)00046-1.

Greco, Salvatore. 2019. “Explaining black-box models in the context of Natural Language Processing”. Politecnico di Torino, Master degree course in Computer Engineering

Huang, Tongwen, Zhiqi Zhang, and Junlin Zhang. 2019. “Fibinet: Combining Feature Importance and Bilinear Feature Interaction for Click-through Rate Prediction.” In RecSys 2019 - 13th ACM Conference on Recommender Systems, 169–77. Association for Computing Machinery, Inc. doi:10.1145/3298689.3347043.

Huh, Jisu, and Edward C. Malthouse. 2020. “Advancing Computational Advertising: Conceptualization of the Field and Future Directions.” *Journal of Advertising* 49 (4). Routledge: 367–76. doi:10.1080/00913367.2020.1795759.

Jong, J. T. de. 2021. “Shapley Values A Comparison of Definitions and Approximation Methods” Thesis of Delft University of Technology Repository, Master of Science. http://repository.tudelft.nl/.

Juan, Yuchin, Damien Lefortier, and Olivier Chapelle. 2017. “Field-Aware Factorization Machines in a Real-World Online Advertising System.” In 26th International World Wide Web Conference 2017, WWW 2017 Companion, 680–88. International World Wide Web Conferences Steering Committee. doi:10.1145/3041021.3054185.

Liu, Tong and Yu Zhengdong. 2022. “Is something out of reach more attractive? The effectiveness of visual distance in computational advertising.” Front. Psychol. 13:994573. doi: 10.3389/fpsyg.2022.994573

Lundberg, Scott M., and Su In Lee. 2017. “A Unified Approach to Interpreting Model Predictions.” In Advances in Neural Information Processing Systems, 2017-December:4766–75. Neural information processing systems foundation.

Malhi, Avleen, Manik Madhikermi, Yaman Maharjan, and Kary Framling. 2021. “Online Product Advertisement Prediction and Explanation in Large-Scale Social Networks.” In *2021 8th International Conference on Social Network Analysis, Management and Security, SNAMS 2021*. Institute of Electrical and Electronics Engineers Inc. doi:10.1109/SNAMS53716.2021.9732145.

Mehrabi, Ninareh, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. 2021. “A Survey on Bias and Fairness in Machine Learning.” ACM Computing Surveys. Association for Computing Machinery. doi:10.1145/3457607.

Molnar, Christoph. 2019. “Interpretable Machine Learning. A Guide for Making Black Box Models Explainable.” Book, 247. https://christophm.github.io/interpretable-ml-book

Pennington, Jeffrey, Richard Socher, and Christopher D. Manning. 2014. “GloVe: Global Vectors for Word Representation.” In EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference, 1532–43. Association for Computational Linguistics (ACL). doi:10.3115/v1/d14-1162.

Perlich, Claudia, Brian Dalessandro, Rod Hook, Ori Stitelman, Troy Raeder, and Foster Provost. 2012. “Bid Optimizing and Inventory Scoring in Targeted Online Advertising.” In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 804–12. doi:10.1145/2339530.2339655.

Rammstedt, Beatrice, and Oliver P. John. 2007. “Measuring Personality in One Minute or Less: A 10-Item Short Version of the Big Five Inventory in English and German.” Journal of Research in Personality 41 (1): 203–12. doi:10.1016/j.jrp.2006.02.001.

Rendle, Steffen. 2012. “Factorization Machines with LibFM.” ACM Transactions on Intelligent Systems and Technology 3 (3). doi:10.1145/2168752.2168771.

Roffo, Giorgio, and Alessandro Vinciarelli. 2016. “Personality in Computational Advertising: A Benchmark.” In CEUR Workshop Proceedings, 1680:18–25. CEUR-WS.

Statista Research Department. “Digital Advertising Report 2022.” Statista. 2022. URL https://www.statista.com/study/42540/digital-advertising-report/.

WebFX. “The Ultimate List of Impressive Digital Advertising Statistics in 2022”. Digital Advertising Statistics. 2022. URL https://www.webfx.com/digital-advertising/statistics/.

Xu, Yong, Jiahui Chen, Chao Huang, Bo Zhang, Hao Xing, Peng Dai, and Liefeng Bo. 2020. “Joint Modeling of Local and Global Behavior Dynamics for Session-Based Recommendation.” In Frontiers in Artificial Intelligence and Applications, 325:545–52. IOS Press BV. doi:10.3233/FAIA200137.

Yang, Xiao, Tao Deng, Weihan Tan, Xutian Tao, Junwei Zhang, Shouke Qin, and Zongyao Ding. 2019. “Learning Compositional, Visual and Relational Representations for CTR Prediction in Sponsored Search.” In International Conference on Information and Knowledge Management, Proceedings, 2851–59. Association for Computing Machinery. doi:10.1145/3357384.3357833.

Zablocki, Éloi, Hédi Ben-Younes, Patrick Pérez, Matthieu Cord. 2022. “Explainability of Deep Vision-Based Autonomous Driving Systems: Review and Challenges.” Int J Comput Vis 130, 2425–2452. https://doi.org/10.1007/s11263-022-01657-x