Measuring the Influence of Mere Exposure Effect of TV Commercial Adverts on Purchase Behavior based on Machine Learning Prediction Models

Elisa Claire Alemán Carreón[a,*], Hirofumi Nonaka[a], Asahi Hentona[a], Hirochika Yamashiro[a]

[a] *Nagaoka University of Technology, Nagaoka, Japan*

Abstract

Since its introduction, television has been the main channel of investment for advertisements in order to influence customers purchase behavior. Many have attributed the mere exposure effect as the source of influence in purchase intention and purchase decision; however, most of the studies of television advertisement effects are not only outdated, but their sample size is questionable and their environments do not reflect reality. With the advent of the internet, social media and new information technologies, many recent studies focus on the effects of online advertisement, meanwhile the investment in television advertisement still has not declined. In response to this, we applied machine learning algorithms SVM and XGBoost, as well as Logistic Regression, to construct a number of prediction models based on at-home advertisement exposure time and demographic data, examining the predictability of ~~Purchase Intention and~~ Actual Purchase and Purchase Intention behaviors of 3000 customers across ~~38~~ 36 different products during the span of 3 months. If we were able to predict purchase behaviors with models based on exposure time ~~only~~more reliably than with models based on demographic data, the obvious strategy for businesses would be to increase the number of adverts. On the other hand, ~~unpredictability would put doubts~~

*Corresponding author

Email addresses: s153400@stn.nagaokaut.ac.jp (Elisa Claire Alemán Carreón), nonaka@kjs.nagaokaut.ac.jp (Hirofumi Nonaka), s173348@stn.nagaokaut.ac.jp (Asahi Hentona), s173358@stn.nagaokaut.ac.jp (Hirochika Yamashiro)

Preprint submitted to Journal of Information Processing and Management December 5, 2018

~~in~~ if models based on exposure time had unreliable predictability in contrast to models based on demographic data, doubts would surface about the effectiveness of the hard investment in television advertising. ~~With our user-based predictability analysis~~Based on our results, we found that ~~only a fourth of the population was predictable in regards to their Purchase Intention, and that exposure to advertisements doesn't relate to Actual Purchase behaviors in any observable way. With our product-based analysis, only a few products produced predictability in Purchase intention, and none were able to influence Actual Purchase predictions. This has immense implications for the advertisement industry, since the return of investment in advertisement cannot be predicted accurately, and the effectiveness of television advertisements in increasing sales is now doubtful~~models based on advert exposure time were consistently low in their predictability in comparison with models based on demographic data only, and with models based on both demographic data and exposure time data. We also found that there was not a statistically significant difference between these last two kinds of models. This suggests that advert exposure time has little to no effect in the short-term in increasing positive actual purchase behavior.

Highlights.

- ~~Exposure time to adverts was not found to induce predictability of purchase behavior~~Models based on exposure time to television adverts have significantly lower predictability of actual purchase in comparison with models using demographic data.
- ~~Purchase Intention is only slightly predictable, and no link to Actual Purchase behavior was found.~~
- ~~Actual Purchase behavior was completely unpredictable, regardless of exposure to advertisement~~Actual Purchase behavior predictability was not significantly different in models including both demographic data and exposure time data as opposed to those with only demographic data.
- ~~Based on mere exposure effect only, effectiveness of television advertisement in increasing sales was not observed~~Results suggest that advert exposure

time has little to no effect in the short-term in increasing positive actual purchase behavior.

Keywords: Television Adverts, Purchase Behavior, SVM, XGBoost, Machine Learning

---

1. Introduction

It is generally thought that in order for companies to increase sales, they must somehow increase the purchase intention of their potential customers [1, 2]. Historically this has been approached through many channels, but since the successful introduction of the television to the general public, it has been largely attempted via television commercial advertisements, and many companies invest heavily on these efforts. However, most studies to prove the effectiveness of these advertisements have been conducted on small sample groups, usually introducing a customer to a commercial advertisement and measuring their intentions to purchase a product before and after watching the advertisement with a survey [e.g. 3]. Studies on the predictability of purchase behavior from purchase intention data have pointed out that many of these analyses have very different results [2, 4, 5], presumably because of small and non-representative samples, and controlled environments that do not reflect reality.

With the advent of Big Data and new methodologies in the field of information technology, there is a new and improved lens for advertisement research ~~in real~~ on the effects it can have on people outside controlled environments; however, its focus is mostly on similarly new advertisement online and in social media [6, 7, 8, 9], leaving behind the study of more traditional advertisement which has not declined in use since the increase of online advertisement. In response to this lack of current research in the field of television advertisement, we propose a machine learning approach to this problem, with a large database of the household television usage timelines of surveyed individuals and their answers regarding recent purchase intentions and actual purchase recalls at two points in time separated by 3 months, provided by the Nomura Research Institute, Ltd.

Now, following the traditional train of thought of the effects of mere exposure [10], we propose collecting the accumulated number of seconds that a user has viewed a commercial advert related to a certain product and observe its effects on the users. With this data we propose training a ~~model~~ number of models to predict the purchase intention and purchase recall of users based on the amount of accumulated seconds of being exposed to the advertisement for the related product in the survey, and then compare it to models that use demographic data of the users as a control. We propose to do this by unit of product, to observe the difference in marketing success from product to product, and by unit of user, to observe the rate of population that was potentially influenced by advertisements. This introduces both granularity, as we are using precise television viewing time and observing effects over time, and the potential of generalizing our prediction model to unknown new users or products after training.

2. Research Objective

The objective of this study is to provide an updated methodology and a larger scale database to measure the mere exposure effect and perceptual fluency effect of television adverts on purchasing behavior. For a long time, psychology based studies have been widely performed on small groups of people in very controlled environments that do not reflect customers in real life accurately, and they have been traditionally thought effective without criticism. We aim to measure the predictability in purchase behavior based on the time spent exposed to adverts of specific products in ~~a real environment~~ household televisions during the duration of 3 months, ~~and~~ then compare it to the predictability in purchase behavior when using demographic data to provide a clearer answer to whether the heavy investment into TV advertising is actually having an effect on customers to purchase more. As control, we will also measure the predictability in purchase behavior based on demographic data and combining the two sources of data. In the case that the predictability is high enough compared to models that don't include exposure time, this methodology could be used as a measure for future sales. On the other hand, a low predictability in comparison to the control would create doubts on whether the mere exposure to advertisements on television is being effective.

3. Related Work

In previous research, there have been attempts to analyze the effects of adverts via mere exposure [11], and many studies have replicated the original experiment by Zajonc unrelated to adverts in the field of psychology [12, 13]. Now, in addition to the focus on the mere exposure effect, there have been attempts to measure the effects of advertisements on brand recognition and perception fluency [14], as well as its effects on the perception of the product [15]. Fluency is defined as the level of ease or difficulty with which external information is processed [16]. Previously it has been proven that it can produce bias, and it has been shown to affect the judgement of truth [17]. For a long time, the perceptual fluency model has stated that repeated exposure leads to a more readily accessibility of the target brand in memory, which in turn must have an effect on the ability to recognize a brand in the future [e.g. 18]. Most of the older research had arrived to a consensus that there is a positive influence [19, 20]. More recent research, however, explores further whether these effects in memory are strictly related to positive emotional judgment on the brands or if they can also imply negative judgements based on the main objective of a product. [21].

Research of the direct effects of television advertisement has also been attempted. One study focuses on child obesity by using weight measurements [22]. An even more direct approach has been made in another study which has used brain imaging in order to explore the short-term and long-term memory effects of TV commercials [23]. It should be noted that, as is to be expected in a brain imaging experiment, the participants observed the advert directly and more consciously than in mere exposure experiments.

Now, two of the main issues with these studies and others in television advertisement effects are that, not only is the size of the samples in these experiments questionably small, but the environment is limited in that it becomes extremely controlled, to the point where it doesn't reflect the reality of customers watching daily television in their homes and making purchases anymore, and the observations environment itself could affect the results.

In order to solve this limitation, our research is based in data science analysis methodology, such as machine learning algorithms trained from large samples of data. Current big data analysis on advertising is mostly focused on online advertisements [9, 24] , where, with the advance of current technology, a user is exposed to adverts placed near to the the content they are currently consuming which are specifically targeting their interests [25, 26] , catching their attention (which is no longer mere exposure, but direct interaction), or a user is incentivized to watch an advertisement by blocking completely the content they were consuming until the advertisement is finished playing on screen. Most of the research in this area is focused on new ways to create online advertisements in social media [27] and suggestions or recommendations targeted to a user's interests [e.g. 28, 29, 30, 31] reducing the need of mere exposure advertisement while online. In addition to this, some studies have focused on testing the effects of online advertisement on customers [32, 33, 34].

While these new technologies made possible the analysis of online advertisement and social media, the focus has shifted and there is no research using these technologies to test the effectiveness of the mere exposure effect based advertisements which are still in use in other traditional means, such as billboards, or as we analyze in our study, television advertisement. Our study is unique in that, using data from television advertisement in household environments and not online ads, we apply data science methodology to explore with a larger sample and a ~~real life environment~~household environment, if there is an effect caused by mere exposure advertisement, and to what extent this effect happens. Our study is also unique in that comparing the results of prediction models based on exposure time to those using demographic data used in previous literature, we can determine if there is an effect caused by exposure, or if purchase behavior is better decided by other external factors, such as income of each individual and their marital and parental statuses.

4. Methodology

As explained above, our approach is to train machine learning models based on the number of seconds of advertisement exposure and demographic data, to predict the effect on the customers purchase decisions measure their predictability. A high predictability based on exposure time would be useful for measuring and predicting sales in any industry. On the other hand, a low predictability in comparison to that of demographic data models would create doubts that the current advertisement based on mere exposure is effective.

Our proposed method is explained in detail in the following sections.

4.1. ~~Survey Data Analysis~~Experiment Design Overview

First we will explain the general design of the experiments. Each experiment consists in creating a prediction model based on a dataset comprised of input features and previously known output labels. After the model is trained, it is able to make predictions of new output labels of unknown data if a given new input values. In this study, we created many models by variating the training input features and output labels and compare their results. For the input data, we prepared datasets based on advertisement viewing time and demographic data. For the prediction targets, we prepared datasets for purchase intention and actual purchase behaviors. We also measured the predictability of each purchase behavior target either by unit of product, to observe the difference in marketing success from product to product, and by unit of user, to observe the rate of population that was potentially influenced by advertisements. Finally we utilized 3 different prediction models, Support Vector Machine [35], XGBoost [36] and Logistic Regression [37] in order to compare performance.

Table 1: Experiment Variations

Experiment Contents	**Variations**
Prediction Model Bases	• Product Based Models • User Based Models
Prediction Targets	• Purchase Intention • Actual Purchase
Input Data Variants	• Advert Viewing Time • Advert Viewing Time, Demographics, (and Purchase Intention) • Demographics (and Purchase Intention)
Prediction Models	• Support Vector Machine • XGBoost • Logistic Regression

These variations for each experiment are shown in Table 1 and each item will be explained in detail in the following sections. It is important to note that Purchase Intention is described to be used in the input vectors in our experiments in Table 1, but this was of course removed for the experiments in which it was the Prediction Target to avoid redundancies.

4.2. Prediction Model Bases

In this study, we measured the predictability of each purchase behavior target either by unit of product, to observe the difference in marketing success from product to product, and by unit of user, to observe the rate of population that was potentially influenced by advertisements. After extracting the commercial advert viewing data using these parameters from the 3000 users that answered the survey, which includes purchase behavior questions from 200 products at two different points in time, only 36 products from those in the survey were linked to commercial adverts that were actually viewed by those same users. Thusly, we performed our experiments using the viewing data of 3000 users for

Table 2: Prediction Model Bases

Prediction Model Base	**Description**
Product Based Prediction Models	For each product from 36 available in the survey, data from 3000 users was collected and paired with their labels.
User Based Prediction Models	For each user from 3000 available, data corresponding to the 36 products available in the survey was collected and paired with their labels.

these 36 products in the configurations explained before in Table 2.

4.3. Prediction Targets

4.3.1. Purchase Intention and Actual Purchase

From the survey data provided by Nomura Research Institute Ltd., we can examine 3000 customer samples, of which we can extract the Purchase Intention and Actual Purchase answers at two points in time, one in January 2017, and another in March 2017, for 200 different products, 36 of which had advertisemnts in the same time period. Each time, the surveys inquire the customer if they have recently had an intention or desire to purchase a certain product (regardless of action on this desire), which corresponds to Purchase Intention; likewise, it inquires if they have recently had purchased a product, corresponding to the Actual Purchase element. We will inspect the effect of adverts on these two elements of a customer's purchase decisions and observe their change with time on the span of three months.

4.3.2. Prediction Target Data Categorization

In order to explore the different effects commercial adverts may have on the purchase decisions of customers based on their answers from two different points in time, we have labeled each user in regard to each product with 6 categories (from 0 to 5), describing several patterns of behavior. For example, let's examine customers who answered they had purchased a product in January and then not in March, corresponding to category 0, in comparison to customers who purchased the product in March, corresponding to category 4: It is possible that, had category 0 customers were exposed to adverts in greater quantity than other users who still purchased the product and weren't exposed to as many adverts on the span of 3 months, this could mean that the advert was at least not effective, or in a worse scenario, off-putting. On the other hand, if the amount of advert exposure was minimal with category 0 customers and at the same time, customers in category 4 who actually recall having purchased the product in the March survey had been exposed to a large amount of adverts, it would prove to be an effective commercial advert campaign.

Although our approach for analysis is different, the above is a simple example of the importance of this distinction between behavior categories. The six categories for each element are explained in detail in Table 3 and Table 4.

Table 3: Category definition for Actual Purchase element

Category	January Actual Purchase	March Actual Purchase
0	Yes	No
1	No	No
2	No	Yes
3	Yes	Yes
4	~~Not Considered~~ Yes/No	Yes
5	~~Not Considered~~ Yes/No	No

4.4. ~~Advert Viewing Time Calculation~~Input Data

~~Now, on the other side of our analysis, observing the relation between the previously explained data and the effectiveness of commercial adverts on television, we extract the~~

Table 4: Category definition for Purchase Intention element

Category	January Purchase Intention	March Purchase Intention
0	Yes	No
1	No	No
2	No	Yes
3	Yes	Yes
4	~~Not Considered~~ Yes/No	Yes
5	~~Not Considered~~ Yes/No	No

4.4.1. Advert Viewing Time

We extracted the viewing time for adverts of each product for each customer from the household television viewing data collected and provided by Nomura Research Institute Ltd. Now the data provided tells us if a user had ~~the television~~ their personal television turned on at the moment of a certain show. Using the information provided of which commercial advert was shown during which television show and how long they lasted, we extracted the number of accumulated seconds a user had the television on for the adverts of each product, and organized them into different weekdays~~and~~. We called this the Weekday data configuration. For comparison, in a different model, we separated each weekday into two time slots. We did this to further analyze whether the time period regularly described as "Primetime" (19:00 to 23:00) had any different influence than other time slots. ~~The final data for each user and product included the elements described in~~ We called this the Weekday Time Slot data configuration. We show the detailed features in Appendix A in Table A.18. ~~We created different datasets so that the viewing timeof one element never repeated in a later one. Because of this, the total viewing time was analyzed on its own.~~

4.4.2. Demographic Data

In order to perform control experiments, in which the prediction is either aided by, or designed only to be based on external factors from the advert exposure time, we performed experiments using the demographic information of each user collected at the time of the survey by Nomura Research Institute, Ltd. We used the age, sex, marital status, parental status and income bracket reported by each user. The answers and consequently the vector features are shown in detail in Appendix A in Table A.19.

4.5. Prediction Models

In this study we chose 3 prediction models: Support Vector Machine, XGBoost and Logistic Regression. SVM and XGBoost are considered well performing supervised machine learning models in the machine learning field considering the size of the data available for this study. Logistic Regression is a statistical model commonly used for binary prediction that is also appropriate for the size of our data. We explain each of those models in more detail in the following sections.

4.5.1. Support Vector Machine

Support Vector Machines (later abbreviated SVM) are supervised machine learning models used in regression and classification problems [35]. Supervised learning meaning that the model trains on previously labeled data, and establishes a way to match the labels as accurately as possible for new unlabeled data to be analyzed. In a binary classification problem, also called a Support Vector Classifier (SVC), previously established binary labels are matched with a p-dimensional vector of input data. Each column or dimension in the vector expresses a feature in the input data, and each row of the vector is a different data point. After each data point is matched with a label, an SVM uses an algorithm to determine a (p-1)-dimensional hyperplane that separates the p-dimensional space in a way that minimizes error in classification, by maximizing the distance between the hyperplane and the nearest point in either classification. ~~A two-dimensional example is shown in Figure ??.~~

~~Two dimensional example of an SVM classification problem~~

In our study we used the linear kernel for our SVC~~, defined by the formula (??) below, where x is the input vector, w is the weight vector, and b is the bias vector. The dimensions of these vectors are such that $f(x)$ and b are the size of the sample size, and w is the size of the amount of features. The sign of the value of $f(x)$ determines which classification label y is applied, as shown in the formulas (??) and (??)~~.

$$\underline{f(x) = w^{\top} x + b}$$

$$\underline{f(x) \geq 0 \rightarrow y = +1}$$

$$\underline{f(x) \leq 0 \rightarrow y = -1}$$

~~The algorithm consists of, starting with a weight and bias vector comprised of zeroes, a randomly placed hyperplane is drawn. Each data point is tested for correct classification, and if the classification fails, the value of w is changed by a value of α as follows (??) . Finally the distance to the nearest points, the support vectors, in either classification, called the margin, is calculated~~

$$\underline{w \leftarrow w + \alpha sign(f(x_i))x_i}$$

~~This process is repeated so that the margin is maximized and the number of erroneous classifications are minimized.~~

4.5.2. XGBoost

Originally started as a research project by Tianqi Chen [36], XGBoost is an improved and optimized application of a Gradient Boosting Machine, or GBM, also called gradient tree boosting, or gradient boosted regression tree. A Gradient Boosted Regression Tree (GBRT) works by building an ensemble model from several weak learning machines which are just above random guessing in accuracy, in this case using Decision Trees. The misclassified results from these weak predictions are then weighted and added to a final strong learning machine. This process iteratively optimizes the misclassification cost in a functional gradient descent so that the final learning machine focuses on important factors from the training data for a stronger prediction model.

4.5.3. Logistic Regression

The logistic model [37] uses a logistic function to model a binary dependent variable. It is a form of regression in which the probability of the dependant variable being one of two possible values (0 or 1) is estimated from the independent variables.

4.6. Model Evaluation Metrics

~~Now, in~~ In order to measure the effectiveness of the training process and data, we ~~perform~~ performed what is called a K-fold cross validation. This means that after randomly shuffling and splitting our training data in k equal parts, k-1 of those parts are used for training, while the remaining one part is used in validation. Using the trained ~~SVM~~models, a prediction is made, and it is decided if such a prediction is correct or not, and counted and grouped as a True Positive, True Negative, False Positive or False Negative prediction. This is explained in Table 5.

Table 5: Prediction outcomes

	Prediction is Correct	**Prediction is Incorrect**
Prediction is Positive	True Positive	False Positive
Prediction is Negative	True Negative	False Negative

Measures of accuracy are determined from these prediction outcomes. This process is then repeated k times and the measures taken are averaged. In this study we used the F_1 score, which measure is a harmonic mean between precision and recall. Precision, described in formula (~~??~~1), lets us observe the rate of correct positive predictions from all the positive predictions, while Recall, detailed in formula (~~1~~2), observes the rate of correct positive predictions from the total of actual positive data. The F_1 score in formula (~~2~~3) then can only be high when both of these measures are high simultaneously, and will lower substantially if they are not consistent. We use this score as it allows us to avoid overlooking data while maintaining accurate predictions.

$$Precision = \frac{TruePositives}{TruePositives + FalsePositives} \tag{1}$$

$$Recall = \frac{TruePositives}{TruePositives + FalseNegatives} \tag{2}$$

$$F_1 = 2\frac{Precision * Recall}{Precision + Recall} \tag{3}$$

5. Experiments

5.1. *Model Training*

As explained in section 4.1, we designed the experiment by training variations of models depending on the input and output values. In the Tables **??** and **??**, we describe the total of different variations that we performed on each prediction model. The combinations of configurations for the input data shown were explained in Table 1, and they give us a total of 30,360 possible inputs for experiment variations. The possible targets explained in Tables 3 and 4 give us a total of 12 possible prediction targets. Together, we performed a total of 364,320 experiments per prediction model. Since we used 3 kinds of prediction models (SVM, XGBoost and Logistic Regression), we performed a total of 1,092,960 experiments in this study.

5.2. *Experiment Parameters*

Each prediction model, SVM, XGBoost and the Logistic Regression function can have different parameters when fitting the data to the model. In this study the parameters were chosen broadly to make a general approach (not very specialized) to all the different configurations of the experiment that could take place. Because of the number of experiments explained in section 5.1, to choose parameters in a specific manner could unbalance one experiment in favor of the other. As such, we chose simple parameters that can apply to many cases.

The SVM experiments were performed with a linear kernel and a C value of 1. The C parameter allows for misclassification in exchange of a larger margin at small values, and it becomes stricter for larger values, perhaps causing overfitting if large enough. The XGBoost experiments were performed with a learning rate of 0.1, a maximum tree depth of 3, and 100 estimators. The Logistic Regression experiments were performed with unit weight per individual sample. A 5-Fold cross validation was performed for all of the models.

5.3. ~~Prediction Models~~

~~Based on the input vector that is used, we have proposed investigating the two different configurations possible for our prediction models , as explained in Table 2.~~

~~Prediction Model BasesPrediction Model Base Description For each product from 200 available in the survey, data from 3000 users was collected and paired with their labels. For each user from 3000 available, data corresponding to all 200 products available in the survey was collected and paired with their labels.~~

~~After extracting the commercial advert viewing data using these parameters from the 3000 users that answered the survey, which includes purchase behavior questions from 200 products at two different points in time, only 38 products from those in the survey were linked to commercial adverts that were actually viewed by those same users. Thusly, we performed our experiments using the viewing data of 3000 users for these 38 products in the configurations explained before in Table 2~~

5.3. ~~Prediction Model Targets~~

~~Using the previously explained bases, we performed analysis for both Actual Purchase and Purchase Intention, and each variation in change explained before as categories in section **??** of this paper. With this categorization, we can observe the difference in relation between the number of seconds of advert viewing and a change in Purchase Behavior between January 2017 and March 2017. In total we performed experiments with 24 different prediction models . This can be visualized in Table **??**~~

6. Results

Because of the large number of experiments performed in this study, we analyze the average performances for different variations of the model input and prediction output. In order to compare the performance across different variations, we performed t-tests and examined the p-values for statistical significance. The average performance results are detailed in section 6.1. The t-test comparisons are shown in section 6.2.

6.1. *Prediction Score Averages*

The F_1 scores for the SVM product based model for all 36 products were averaged for each variation of the experiment. The results are shown in Table 6. The average F_1 scores for the SVM user based model for all 3000 users are shown in Table 7.

Similarly, the XGBoost product based models average F_1 scores are shown in Table 8, and the user based models average F_1 scores are shown in Table 9.Lastly, the Logistic Regression product and user based models average results are shown in Tables 10 and 11.

6.2. *Statistical Analysis*

In this study we performed a series of experiments where we trained different prediction models based on either advert viewing time or demographic data to

Table 6: SVM Product Based Models Average F_1 scores

Prediction Target	Category	Advert Viewing Weekday Time Slots	Advert Viewing Weekday Only	Demographics	Advert Viewing Weekday Time Slots and Demographics	Advert Viewing Weekday Only and Demographics	Total Average
Actual Purchase	General Average	**0.293**	**0.292**	**0.489**	**0.495**	**0.497**	**0.413**
	0	0.000	0.000	0.211	0.225	0.218	0.131
	1	0.856	0.852	0.876	0.878	0.875	0.867
	2	0.000	0.000	0.327	0.343	0.343	0.204
	3	0.000	0.000	0.161	0.171	0.171	0.102
	4	0.000	0.000	0.446	0.441	0.441	0.267
	5	0.901	0.900	0.910	0.911	0.911	0.907
Purchase Intention	General Average	**0.252**	**0.248**	**0.273**	**0.275**	**0.276**	**0.265**
	0	0.000	0.000	0.000	0.000	0.000	0.000
	1	0.570	0.558	0.590	0.590	0.596	0.581
	2	0.000	0.000	0.000	0.000	0.000	0.000
	3	0.115	0.110	0.139	0.139	0.138	0.129
	4	0.166	0.156	0.226	0.226	0.233	0.202
	5	0.662	0.666	0.685	0.685	0.689	0.678
Both Targets	Total Average	**0.273**	**0.270**	**0.381**	**0.385**	**0.386**	**0.339**

Table 7: SVM User Based Models Average F_1 scores

Prediction Target	Category	Advert Viewing Weekday Time Slots	Advert Viewing Weekday Only	Demographics	Advert Viewing Weekday Time Slots and Demographics	Advert Viewing Weekday Only and Demographics	Total Average
Actual Purchase	General Average	**0.317**	**0.315**	**0.391**	**0.359**	**0.373**	**0.351**
	0	0.055	0.049	0.095	0.085	0.087	0.074
	1	0.745	0.755	0.854	0.789	0.812	0.791
	2	0.074	0.070	0.140	0.114	0.126	0.105
	3	0.076	0.071	0.108	0.107	0.121	0.096
	4	0.140	0.126	0.254	0.221	0.239	0.196
	5	0.812	0.822	0.893	0.840	0.855	0.844
Purchase Intention	General Average	**0.297**	**0.289**	**0.242**	**0.296**	**0.290**	**0.283**
	0	0.036	0.032	0.006	0.036	0.033	0.029
	1	0.553	0.553	0.534	0.548	0.553	0.548
	2	0.048	0.043	0.007	0.049	0.042	0.038
	3	0.217	0.195	0.093	0.217	0.198	0.184
	4	0.301	0.279	0.174	0.299	0.280	0.267
	5	0.629	0.634	0.639	0.626	0.632	0.632
Both Targets	Total Average	**0.307**	**0.302**	**0.316**	**0.327**	**0.331**	**0.317**

Table 8: XGBoost Product Based Models Average F_1 scores

Prediction Target	Category	Advert Viewing Weekday Time Slots	Advert Viewing Weekday Only	Demographics	Advert Viewing Weekday Time Slots and Demographics	Advert Viewing Weekday Only and Demographics	Total Average
Actual Purchase	**General Average**	**0.293**	**0.294**	**0.307**	**0.309**	**0.308**	**0.302**
	0	0.001	0.000	0.027	0.030	0.029	0.017
	1	0.847	0.849	0.756	0.755	0.752	0.792
	2	0.001	0.000	0.048	0.054	0.052	0.031
	3	0.003	0.003	0.043	0.044	0.045	0.028
	4	0.008	0.010	0.136	0.136	0.133	0.085
	5	0.898	0.900	0.835	0.835	0.833	0.860
Purchase Intention	**General Average**	**0.257**	**0.257**	**0.266**	**0.266**	**0.268**	**0.263**
	0	0.000	0.000	0.000	0.002	0.001	0.001
	1	0.574	0.571	0.573	0.565	0.572	0.571
	2	0.000	0.001	0.001	0.001	0.002	0.001
	3	0.121	0.122	0.136	0.136	0.142	0.131
	4	0.175	0.175	0.211	0.220	0.219	0.200
	5	0.670	0.674	0.673	0.669	0.674	0.672
Both Targets	**Total Average**	**0.275**	**0.275**	**0.287**	**0.287**	**0.288**	**0.282**

Table 9: XGBoost User Based Models Average F_1 scores

Prediction Target	Category	Advert Viewing Weekday Time Slots	Advert Viewing Weekday Only	Demographics	Advert Viewing Weekday Time Slots and Demographics	Advert Viewing Weekday Only and Demographics	Total Average
Actual Purchase	**General Average**	**0.291**	**0.291**	**0.297**	**0.301**	**0.302**	**0.296**
	0	0.008	0.008	0.018	0.021	0.021	0.015
	1	0.771	0.769	0.752	0.758	0.760	0.762
	2	0.014	0.014	0.030	0.032	0.033	0.025
	3	0.027	0.029	0.042	0.048	0.049	0.039
	4	0.082	0.084	0.119	0.121	0.123	0.106
	5	0.841	0.842	0.825	0.827	0.825	0.832
Purchase Intention	**General Average**	**0.267**	**0.269**	**0.239**	**0.267**	**0.268**	**0.262**
	0	0.008	0.008	0.001	0.008	0.007	0.006
	1	0.538	0.543	0.533	0.542	0.542	0.540
	2	0.013	0.013	0.001	0.013	0.013	0.010
	3	0.154	0.161	0.085	0.155	0.159	0.143
	4	0.249	0.250	0.174	0.248	0.252	0.235
	5	0.638	0.637	0.643	0.636	0.636	0.638
Both Targets	**Total Average**	**0.279**	**0.280**	**0.268**	**0.284**	**0.285**	**0.279**

Table 10: Logistic Regression Product Based Models Average F_1 scores

Prediction Target	Category	Advert Viewing Weekday Time Slots	Advert Viewing Weekday Only	Demographics	Advert Viewing Weekday Time Slots and Demographics	Advert Viewing Weekday Only and Demographics	Total Average
Actual Purchase	**General Average**	**0.293**	**0.293**	**0.500**	**0.513**	**0.508**	**0.421**
	0	0.000	0.000	0.226	0.234	0.237	0.139
	1	0.851	0.853	0.874	0.874	0.875	0.865
	2	0.000	0.000	0.343	0.368	0.355	0.213
	3	0.002	0.000	0.181	0.208	0.195	0.117
	4	0.006	0.004	0.462	0.480	0.475	0.286
	5	0.899	0.901	0.914	0.914	0.914	0.909
Purchase Intention	**General Average**	**0.259**	**0.256**	**0.289**	**0.294**	**0.292**	**0.278**
	0	0.000	0.000	0.000	0.000	0.000	0.000
	1	0.575	0.572	0.608	0.611	0.611	0.595
	2	0.000	0.000	0.000	0.001	0.000	0.000
	3	0.125	0.118	0.159	0.170	0.163	0.147
	4	0.184	0.174	0.270	0.280	0.278	0.237
	5	0.671	0.668	0.696	0.703	0.701	0.688
Both Targets	**Total Average**	**0.276**	**0.274**	**0.394**	**0.404**	**0.400**	**0.350**

Table 11: Logistic Regression User Based Models Average F_1 scores

Prediction Target	Category	Advert Viewing Weekday Time Slots	Advert Viewing Weekday Only	Demographics	Advert Viewing Weekday Time Slots and Demographics	Advert Viewing Weekday Only and Demographics	Total Average
Actual Purchase	**General Average**	**0.312**	**0.309**	**0.347**	**0.333**	**0.341**	**0.328**
	0	0.046	0.040	0.037	0.056	0.054	0.047
	1	0.740	0.750	0.846	0.771	0.794	0.780
	2	0.064	0.058	0.070	0.083	0.082	0.071
	3	0.074	0.066	0.064	0.089	0.091	0.077
	4	0.139	0.124	0.175	0.173	0.185	0.159
	5	0.806	0.815	0.888	0.827	0.844	0.836
Purchase Intention	**General Average**	**0.298**	**0.295**	**0.242**	**0.298**	**0.293**	**0.285**
	0	0.037	0.035	0.007	0.038	0.034	0.030
	1	0.555	0.557	0.535	0.554	0.556	0.551
	2	0.052	0.047	0.009	0.052	0.044	0.041
	3	0.220	0.204	0.093	0.220	0.207	0.189
	4	0.298	0.288	0.175	0.300	0.286	0.269
	5	0.628	0.635	0.636	0.628	0.634	0.632
Both Targets	**Total Average**	**0.305**	**0.302**	**0.294**	**0.316**	**0.317**	**0.307**

predict purchase behaviors of actual purchase and purchase intention across 3000 users and 36 products. If we were able to predict purchase behaviors with models based on exposure time more reliably than with models based on demographic data, obvious strategy for businesses would be to increase the number of adverts. On the other hand, if models based on exposure time had unreliable predictability in contrast to models based on demographic data, doubts would surface about the hard investment in television advertising.

In order to analyze the change in predictability of purchase behavior we averaged the results ~~for SVM and XGBoost experiments in a later section~~of predictions across different variations of the experiments, detailed in the previous section, and then performed t-tests to observe the difference in performance between sets of results. We established 3 hypotheses to test for, explained below.

Hypothesis 1: *Advert viewing time based models perform differently from demographics based models.*

For this hypothesis, we performed a t-test bewteen the results from models that include advert viewing time and the models that only include demographic data. More specifically, we tested the Weekday Time Slot model results against the Demographics models, and the Weekday Only models against the Demographics models. The p-values for each t-test are shown in Table 12 for Actual Purchase predictions and in Table 13 for Purchase Intention predictions. With these tests, we will examine the changes in predictability against demographic data, which we are using as the control data for our experiments. This will allow us to determine whether the advert viewing time based models are performing better or worse than the demographic models, and therefore conclude whether the advert viewing time is having an effect on customers purchase behavior or if it is decided by external factors.

Hypothesis 2: *Demographic and advert viewing based models perform differently from demographic based models.*

Table 12: Hypothesis 1 t-test: p-values for Actual Purchase behavior

Model	**Base**	**Configuration**	**Actual Purchase Categories**					
			0	**1**	**2**	**3**	**4**	**5**
SVM	product	Weekday Time Slot	0.000	0.328	0.000	0.001	0.000	0.567
		Weekday Only	0.000	0.284	0.000	0.001	0.000	0.527
	user	Weekday Time Slot	0.000	0.000	0.000	0.000	0.000	0.000
		Weekday Only	0.000	0.000	0.000	0.000	0.000	0.000
XGBoost	product	Weekday Time Slot	0.000	0.005	0.000	0.015	0.000	0.013
		Weekday Only	0.000	0.004	0.000	0.015	0.000	0.011
	user	Weekday Time Slot	0.000	0.004	0.000	0.000	0.000	0.004
		Weekday Only	0.000	0.010	0.000	0.000	0.000	0.003
Logistic Regression	product	Weekday Time Slot	0.000	0.282	0.000	0.000	0.000	0.348
		Weekday Only	0.000	0.327	0.000	0.000	0.000	0.414
	user	Weekday Time Slot	0.012	0.000	0.236	0.032	0.000	0.000
		Weekday Only	0.374	0.000	0.007	0.757	0.000	0.000

Table 13: Hypothesis 1 t-test: p-values for Purchase Intention behavior

Model	**Base**	**Configuration**	**Purchase Intention Categories**					
			0	**1**	**2**	**3**	**4**	**5**
SVM	product	Weekday Time Slot	nan	0.803	nan	0.700	0.405	0.767
		Weekday Only	nan	0.694	nan	0.644	0.329	0.808
	user	Weekday Time Slot	0.000	0.026	0.000	0.000	0.000	0.234
		Weekday Only	0.000	0.028	0.000	0.000	0.000	0.534
XGBoost	product	Weekday Time Slot	0.324	0.983	0.041	0.804	0.598	0.969
		Weekday Only	0.324	0.982	0.203	0.811	0.606	0.992
	user	Weekday Time Slot	0.000	0.533	0.000	0.000	0.000	0.585
		Weekday Only	0.000	0.246	0.000	0.000	0.000	0.509
Logistic Regression	product	Weekday Time Slot	nan	0.673	0.803	0.582	0.222	0.724
		Weekday Only	nan	0.655	0.324	0.514	0.175	0.704
	user	Weekday Time Slot	0.000	0.017	0.000	0.000	0.000	0.364
		Weekday Only	0.000	0.010	0.000	0.000	0.000	0.947

Table 14: Hypothesis 2 t-test: p-values for Actual Purchase behavior

Model	Base	Configuration	Actual Purchase Categories					
			0	**1**	**2**	**3**	**4**	**5**
SVM	product	Weekday Time Slot	0.923	0.953	0.748	0.777	0.968	0.980
		Weekday Only	0.850	0.936	0.839	0.868	0.942	0.956
	user	Weekday Time Slot	0.056	0.000	0.000	0.837	0.000	0.000
		Weekday Only	0.139	0.000	0.033	0.038	0.078	0.000
XGBoost	product	Weekday Time Slot	0.758	0.973	0.581	0.936	0.991	0.993
		Weekday Only	0.811	0.929	0.703	0.897	0.924	0.956
	user	Weekday Time Slot	0.074	0.318	0.413	0.057	0.625	0.655
		Weekday Only	0.069	0.187	0.173	0.045	0.388	0.993
Logistic Regression	product	Weekday Time Slot	0.894	0.997	0.674	0.627	0.772	0.971
		Weekday Only	0.856	0.955	0.846	0.808	0.831	0.990
	user	Weekday Time Slot	0.000	0.000	0.007	0.000	0.823	0.000
		Weekday Only	0.000	0.000	0.017	0.000	0.187	0.000

For this hypothesis, we performed a t-test bewteen the results from models that include both advert viewing time and demographich data, and the models that only include demographic data. More specifically, we tested the Weekday Time Slot and Demographics model results against the Demographics models, and the Weekday and Demographics models against the Demographics models. The p-values for each t-test are shown in Table 14 for Actual Purchase predictions and in Table 15 for Purchase Intention predictions. With these tests, we will examine if adding the advert viewing data to the demographic data causes any major changes, to determine if the predictions are being improved, worsened, or if they stay the same regardless of advert viewing.

Hypothesis 3: *Advert viewing time based models perform differently from demographic and advert viewing based models.*

For this hypothesis, we performed a t-test bewteen the results from models that include both advert viewing time and demographich data, and the models that only include advert viewing data. More specifically, we tested the Weekday Time Slot and Demographics model results against the Weekday Time Slot models, and the Weekday and Demographics models against the Weekday Only models. The p-values for each t-test are shown in Table 16 for Actual Purchase predictions and in Table 17 for Purchase Intention predictions. With these tests, we will examine if adding the demographic data to the advert viewing data causes any major changes, to determine if the predictions are being improved, worsened or if they stay the same regardless of demographic variances. By performing this last test, as well as the differences tested by Hypothesis 1 and Hypothesis 2, we can assume significant differences across all major 3 groups of data.

6.3. ~~*Product Based Models Results*~~

6.2.1. ~~*Product Based CM View Time -Actual Purchase Category 2*~~

~~For all 38 products, we collected the data of actual purchase behavior from 3000 customers, labeling those who changed their behavior from not having purchased in January 2017 to having made a purchase in March 2017 as the positive classification, and any customer other than those as the negative classification for our SVM training data.~~

~~The results of the F_1 score for the prediction models for each of the 38 products resulted in 0. This means that any predictions were failed and that the SVM could not find a separating (p-1)-dimensional hyperplane for the viewing data. In other words, people who changed their purchasing behavior from January to March, and people who did otherwise, were exposed to similar amounts of advert time, or that viewing time was spread indiscriminately for all users regardless of purchase decision for any of the products that we investigated separately.~~

6.2.1. ~~*Product Based CM View Time -Actual Purchase Category 4*~~

~~Now, the results for the product based model labeling positively those who regardless of change, had actually purchased the product by March 2017; and labeling negatively any customers who did not purchase the product by March are presented in this section~~

Table 15: Hypothesis 2 t-test: p-values for Purchase Intention behavior

Model	Base	Configuration	Purchase Intention Categories					
			0	**1**	**2**	**3**	**4**	**5**
SVM	product	Weekday Time Slot	nan	0.942	nan	0.985	0.930	0.969
		Weekday Only	nan	0.986	nan	0.991	0.950	0.955
	user	Weekday Time Slot	0.000	0.099	0.000	0.000	0.000	0.118
		Weekday Only	0.000	0.029	0.000	0.000	0.000	0.402
XGBoost	product	Weekday Time Slot	0.002	0.924	0.850	0.996	0.900	0.960
		Weekday Only	0.115	0.990	0.364	0.916	0.916	0.986
	user	Weekday Time Slot	0.000	0.290	0.000	0.000	0.000	0.409
		Weekday Only	0.000	0.303	0.000	0.000	0.000	0.400
Logistic Regression	product	Weekday Time Slot	nan	0.957	0.331	0.856	0.884	0.916
		Weekday Only	nan	0.957	0.871	0.943	0.903	0.943
	user	Weekday Time Slot	0.000	0.028	0.000	0.000	0.000	0.313
		Weekday Only	0.000	0.016	0.000	0.000	0.000	0.869

Table 16: Hypothesis 3 t-test: p-values for Actual Purchase behavior

Model	Base	Configuration	Actual Purchase Categories					
			0	**1**	**2**	**3**	**4**	**5**
SVM	product	Weekday Time Slot	0.000	0.354	0.000	0.000	0.000	0.553
		Weekday Only	0.000	0.253	0.000	0.000	0.000	0.488
	user	Weekday Time Slot	0.000	0.000	0.000	0.000	0.000	0.000
		Weekday Only	0.000	0.000	0.000	0.000	0.000	0.000
XGBoost	product	Weekday Time Slot	0.000	0.005	0.000	0.009	0.000	0.013
		Weekday Only	0.000	0.003	0.000	0.013	0.000	0.009
	user	Weekday Time Slot	0.000	0.051	0.000	0.000	0.000	0.011
		Weekday Only	0.000	0.181	0.000	0.000	0.000	0.002
Logistic Regression	product	Weekday Time Slot	0.000	0.287	0.000	0.000	0.000	0.367
		Weekday Only	0.000	0.303	0.000	0.000	0.000	0.400
	user	Weekday Time Slot	0.005	0.000	0.000	0.001	0.000	0.000
		Weekday Only	0.000	0.000	0.000	0.000	0.000	0.000

~~The F_1 scores for all but one of the 38 prediction models in this experiment were 0. The remaining product was the japanese bottled tea "IEMON", with an F_1 score of 0.018. Now, for a score to describe a relatively accurate prediction model, it must be at least above 0.5. This means that even though this particular product had a score slightly above 0, there is no particular relation found by the SVM that could determine the purchase behavior of customers based on the time spent viewing adverts for the related product.~~

6.2.1. ~~*Product Based CM View Time -Purchase Intention Category 2*~~

~~Similar to our results for Actual Purchase behavior, the experiment labeling positively customers whose Purchase Intention changed from not having any intention at the first point in time to having a purchase intention at the latter, and labeling negatively those who did otherwise resulted in an F_1 score of 0 for all 38 products. This means that for any of the 38 products, advert viewing time did not have any relation to their purchase intentions changing or not in the vector space.~~

6.2.1. ~~*Product Based CM View Time -Purchase Intention Category 4*~~

~~The results for Purchase Intention behavior, labeling positively customers whose Purchase Intention was positive at the latest point in time in the survey, regardless of their behavior before that point; and labeling negatively those who did otherwise, resulted in an F_1 score greater than 0.5 for 8 of the 38 products available in our viewing data.These prediction accuracy results are presented in Table **??**. There were other 3 products which had an F_1 score between 0 and 0.5, not listed, and the remaining 27 products had an F1 score of 0.~~

6.3. ~~*User Based Models Results*~~

~~This section shows the results from the user based experiments. These experiments calculate the number of specific users that are predictable in their Actual Purchase and Purchase Intention behaviors based on their advert viewing time.~~

6.2.1. ~~User Based CM View Time -Actual Purchase Category 2~~

~~This section presents the results for the Actual Purchase behavior prediction of Category 2 users (those who didn't purchase in the first survey and changed their behavior in the second survey 3 months later). Only 30 of 3000 users had an F_1 score greater than 0.5, meaning they were fairly predictable in their behavior based on their advert viewing time. The remaining 99% was completely unpredictable. This result is shown in Figure **??**.~~

6.2.1. ~~User Based CM View Time -Actual Purchase Category 4~~

~~The results for the user based model in Actual Purchase Category 4, however, presents a larger amount of users, with 253 of 3000 users having been predictable, obtaining an F_1 score greater than 0.5. It is easier to predict a single purchasing behavior in their last survey, based on their viewing time from the previous 3 months, than to predict a specific change in behavior like the last experiment. However, there is still a 92% of users whose behavior was unpredictable with only advertisement viewing time considered. This result is shown in Figure **??**.~~

6.2.1. ~~User Based CM View Time -Purchase Intention Category 2~~

~~Similarly to the experiment for Actual Purchase behavior the results for Purchase Intention prediction for Category 2 users, who changed their behaviorfrom not purchasing to purchasing between surveys, was extremely low. Only 40 of the 3000 users had a prediction score greater than 0.5. This result is shown in Figure **??**.~~

6.2.1. ~~User Based CM View Time -Purchase Intention Category 4~~

~~In contrast to all previous results, the user based prediction model for Purchase Intention in the last survey was more successful. 753 of 3000 users, roughly 25% of users were predictable based on their advert viewing time regarding their Purchase Intention behavior. This result is shown in Figure **??**.~~

~~Prediction Results for User Based CM View Time -Purchase Intention Category 4~~

6.3. ~~*XGBoost Comparison to SVM*~~

~~As the results shown above are all from the experiments done with SVM machine learning algorithms, we show a comparison of the experiments done with XGboost in Table **??** and Table **??**, for product based models and user based models respectively.~~

~~The results for SVM prediction models held a slightly higher percentage of predictable elements, but overall they both show similar results.~~

7. Discussion

7.1. *Influence of TV adverts on Actual Purchase and Purchase Intention*

~~In this paper, we have obtained results on the predictability of Actual Purchase and Purchase Intention customer behaviors~~ Observing our results across models in the Tables of section 6.1, in general, we can observe that SVM models perform relatively better than XGBoost models and Logistic Models, and that the differences and directional change in averages between Advert Viewing Time based models, Demographics based models, and Advert Viewing Time and Demographics based models stay consistent across SVM, XGBoost and Logistic Regression models. That is to say, low predictability in Advert Viewing Time based models compared to Demographic data models stays constant regardless of the changes in performance accross prediction techniques.

In Tables 6 and 7 we can observe this more closely. In general for Actual Purchase behavior, predictions using Advert Viewing Time only have a lower performance than the other models. Specially in categories 2 and 4 of the purchase behavior, we can see that the average predictabiity rises from 0 or close to 0, to a higher predictability in every case that demographic data is used. The exception to this rule is in category 1 of the data, where customers consistently answered "NO" in their purchase recall or purchase intention questions of the survey both in January 2017 and March 2017. Predictions seem to be high across all models, but this is irrelevant since the customer is not changing their negative purchase behavior.

We can confirm this increase is statistically significant by observing the results of Hypothesis 1 in Table 12. For the most part, excluding negative purchase behavior, the data is significantly different at a 95% confidence level ($p < 0.05$) between models that use advert viewing time as a base for prediction and models that use demographic data as a base for prediction. Moreover, we can confirm that the changes in predictability between models that include both advert viewing time and demographic data, and models that only include demographic data are not statistically significant by observing the results of Hypothesis 2 in Table 14. In most cases, ($p > 0.05$), proving that the advert viewing time data did not influence the prediction scores significantly, and that whatever correct predictions were made were most likely based on the ~~time spent exposed to TV adverts. These results would not only be useful in predicting the purchasing behavior of customers, but it would be useful as a direct measure of the effectiveness of mere exposure to advertisements. However, these results are low enough that it can be said~~ coefficients and weights of the demographic data. Finally, observing Hypothesis 3 in Table 16 also confirms the difference and increase of performance between advert viewing time models and those that combine advert data with demographic data.

With these results in mind, it could be said that TV adverts are not a main factor in predicting whether a customer will change their purchasing behavior or not, specially their actual purchase behavior. While the research based on the mere exposure effect would suggest otherwise, customers are observed to decide on their purchase without much predictability, except for their demographic data. It could be said that while there ~~is~~ might be influence in the customer's knowledge of the brand, the data suggests that the amount of time exposed to TV adverts ~~seems to have an indirect influence on purchasing behaviorat most, if not none at all~~has no effect in the customers Actual Purchase behavior.

Other studies, using a controlled environment, have linked mere exposure with bias in consumer choice [38]. However, there is a possible explanation for these discrepancies in results. While controlled experiments show the TV adverts to their sample audience directly in most cases, in an uncontrolled environment of a customer's home, the customer is left free to ignore the advert and do something unrelated in the meanwhile [39]. In the United Kingdom, there is a widely documented phenomenon involving TV advert timing and a surge in electricity caused by the use of electric kettles for preparing tea. This phenomenon is commonly called TV pickup, and has been documented for long [40, 41]. Similar to these cases, if the customers whose data were actively ignoring the adverts, the sample for training the prediction models would contain noise, altering the results. It stands to reason that without the influence of this active aversion would have on our learning model, it might correctly predict purchasing behavior as expected. However, this is more of a problem with the current TV advertisement model than with the methodology of this study. We will discuss this further in section 7.3 of this paper.

7.2. Influence of TV adverts based on Primetime

In our prediction model experiments, we used data from advertisement exposure during different time periods, days of the week and weekends. While we did this in order to observe differences in predictability for different time schedules available to different kinds of customers, especially during prime-time television hours, we arrived to ~~similarly low~~ similar results for all time ~~categories~~data configurations. We did not observe any difference in predictability based on Primetime television watching compared to other time periods~~, as well as differences in weekdays and weekends~~. This could be interpreted as there being little influence in time periods and changes in purchasing behavior. ~~However, as we stated in section 7.1, the common problem with advertisements being actively ignored by customers has existed for long. Taking this problem into consideration, our results imply that the problem is constant over all the time periods, and that there is not a particular time slot that results in customers being more attentive to adverts.~~

7.3. Implications for the TV advert industry

Based on the low results of predictability of purchase behavior by advert exposure, it can be observed that TV adverts have a low probability of achieving their main purpose: to increase sales. As was stated in section 7.1 of this paper, there could be a large influence on this study's results from customers actively ignoring the adverts although they are being broadcast to their TVs. It is left to further discussion and research if adverts actually have the intended effect on customers when watched properly, or if this effect is not achieved anyway. In [14] it is proposed that while the mere exposure of banner advertisement increases perceptual fluency, it doesn't have an effect on actual brand recognition compared to the control groups, for example. The existence or absence of influence by perceptual fluency on a customer's purchase decision hasn't been fully explored, but the consensus in the processing fluency model is that perceptual fluency influences brand judgement on some level, although it depends on the concept if the reception is positive or not [21]. The problem with these studies and the current consensus, as has been said previously in this paper, is both that most experiments are done with relatively small sample sizes, and that there is a factor of uncertainty that comes with the physical avoidance of adverts in a ~~customers home environment~~customer's household.

With these things in mind, we consider both possibilities: either customers are attentive and the adverts have the expected influence in their short and long term memories in the case of repeated exposures [23]; or the customers are inattentive of the advert and there might be some level of unconscious effect of mere exposure in their perception fluency [14]. We observed however in our results that ~~the~~ there is no effect on Actual Purchase behavior~~is minimal~~. While it may be true and out of the reach of our data that the customers would have influence in their memory, there was no link observed between the time of advert exposure and the purchase decisions. This raises a concern for the TV advert industry. Regardless of the cause of our results, the main implication of our paper is that currently, TV adverts are shown to have little to no effect on changes in Actual Purchase behavior~~, and only some observable effect in~~

~~Purchase Intention~~. While thousands of billions of japanese yen are spent on TV advertisements each year [1], the effects observed in this study are negligible. Because of this, changes are necessary in the current TV advertisement model.

8. Limitations

In comparison with previous research regarding this topic, our study presents a much larger database, a sample of 3000 users for 36 different products and the previously unavailable household television viewing data increases the possibilities for studying the effects of advert exposure more realistically. In accordance to this size of data, we used SVM and XGBoost, which are considered well performing machine learning algorithms in this level of magnitude. However, while we propose using machine learning algorithms as an effective method, we are still limited by the magnitude of the data. Top performing and state of the art models, such as Deep Neural Networks, with their variations and advancements, use datasents with orders of magnitude much larger (for example, 3,000,000 users instead of 3,000, and thousands more products instead of 36) in order to perform to the level that they are praised for. The calculation time also increases greatly, and is more appropriate for single models being trained, instead of a comparison of a large array of models.

Another limitation of this study is the nature of the prediction targets collected by survey. While a person can be asked directly in a survey whether they would purchase an item (purchase intention) or if they had already in the recent past (actual purchase), research based on online shopping has access to the actual purchase data, and to the number of times a person looks at a products description page, or searches terms related to it. Television advertisement research, by its nature, is harder to connect to the actual behavior of the customers and can only be assumed to be equal to their reported behavior.

[1] Dentsu, inc. 2017 Advertising Expenditures in Japan. Retrieved on May 2018 from `http://www.dentsu.com/knowledgeanddata/ad_expenditures/pdf/expenditures_2017.pdf`

There is also a limitation of the number of questions that a person might answer, and how honestly they might answer them with a survey of this magnitude.

In addition to this, because of the timing of the surveys being 3 months appart between January and March 2017, we can only examine the short-term effect of advertisements, and not the long-term effect across different years of constant advertisement exposure.

Furthermore, much of the data that could be used to inspect this matter further belongs to private institutions and in many cases, is treated as a company secret.

However, with the measurements of short-term effects of advertisement in a field where not much new research is done, we can start to shed light on problems that could be having a large impact on the costs of many industries.

9. Conclusion and Future Work

In this paper we analyzed the ability to predict purchasing behavior, namely ~~Purchase Intention and~~ Actual Purchase and Purchase Intention, based on the customers' time spent exposed to television adverts using machine learning algorithms~~. We analyzed the data by product and by customer, determining which specific products and which specific customers provided a better prediction model~~, and compared it to the ability to predict the same behavior by using demographic data on its own and in combination with the exposure time data. Based on ~~our low results for any prediction~~ the low prediction results of Actual Purchase ~~, we concluded that there must be other factors that are more strongly tied to the customer's purchasing behavior. The results for Purchase Intention were relatively higher but still low enough that only a few products, mostly tea and chocolate snacks, could be predicted, and only one fourth of the customers were predictable in their purchase intention.~~ by exposure time models and the relatively high prediction results for demographic based models, as well as a non-significant difference between the demographic models and the combined models, we concluded that advertisement exposure has little to no effect in short-time Actual Purchase behavior.

We discussed possible influence by deliberate avoidance of advert cuts to prepare food or tea, and while some studies focus on the effect of attentive watching of adverts, other studies focus on the mere exposure effects, which would be achieved despite physical avoidance because of advert audio and simple proximity of the television. Both scenarios are in strong contrast with the results of our study, which shows little to no predictability in purchase behavior. Points left to research in future work are a deeper analysis of the predictable customers, looking for similarities or clusters within this class, as well as using ~~different machine learning algorithms , which weren't considered because of requiring bigger datasets~~ newer and better performing deep learning algorithms when larger datasets are available.

10. Acknowledgements

This paper was made possible by the data provided by Nomura Research Institute, Ltd. for their yearly Marketing Analysis Contest.

Declarations of interest: none

References

[1] J. Armstrong, V. Morwitz, V. Kumar, Sales forecasts for existing consumer products and services: Do purchase intentions contribute to accuracy?, International Journal Of Forecasting 16 (3) (2000) 383–397. `doi:10.1016/s0169-2070(00)00058-3`.

[2] V. Morwitz, J. Steckel, A. Gupta, When do purchase intentions predict sales?, International Journal Of Forecasting 23 (3) (2007) 347–364. `doi:10.1016/j.ijforecast.2007.05.015`.

[3] M. Khuong, T. Nguyen, The effects of television commercials on customers purchase intention – a study of milk industry in ho chi minh city, vietnam, Journal Of Economics, Business And Management 3 (9) (2015) 851–857.

[4] B. Sun, V. Morwitz, Stated intentions and purchase behavior: A unified model, International Journal Of Research In Marketing 27 (4) (2010) 356–366. doi:10.1016/j.ijresmar.2010.06.001.

[5] C. Newberry, B. Klemz, C. Boshoff, Managerial implications of predicting purchase behavior from purchase intentions: a retail patronage case study, Journal Of Services Marketing 17 (6) (2003) 609–620. doi:10.1108/08876040310495636.

[6] M. Shareef, B. Mukerji, M. Alryalat, A. Wright, Y. Dwivedi, Advertisements on facebook: Identifying the persuasive elements in the development of positive attitudes in consumers, Journal Of Retailing And Consumer Services 43 (2018) 258–268. doi:10.1016/j.jretconser.2018.04.006.

[7] L. Gonzalez Camacho, S. Alves-Souza, Social network data to alleviate cold-start in recommender system: A systematic review, Information Processing & Management 54 (4) (2018) 529–544. doi:10.1016/j.ipm.2018.03.004.

[8] K. Ramaboa, P. Fish, Keyword length and matching options as indicators of search intent in sponsored search, Information Processing & Management 54 (2) (2018) 175–183. doi:10.1016/j.ipm.2017.11.003.

[9] C. Wu, S. Kao, C. Wu, S. Huang, Location-aware service applied to mobile short message advertising: Design, development, and evaluation, Information Processing & Management 51 (5) (2015) 625–642. doi:10.1016/j.ipm.2015.06.001.

[10] R. Zajonc, Attitudinal effects of mere exposure, Journal Of Personality And Social Psychology 9 (2 Pt.2) (1968) 1–27. doi:10.1037/h0025848.

[11] P. Hekkert, C. Thurgood, T. Whitfield, The mere exposure effect for consumer products as a consequence of existing familiarity and controlled exposure, Acta Psychologica 144 (2) (2013) 411–417. doi:10.1016/j.actpsy.2013.07.015.

[12] Y. Huang, P. Hsieh, The mere exposure effect is modulated by selective attention but not visual awareness, Vision Research 91 (2013) 56–61. doi:10.1016/j.visres.2013.07.017.

[13] A. Dechêne, C. Stahl, J. Hansen, M. Wänke, Mix me a list: Context moderates the truth effect and the mere-exposure effect, Journal Of Experimental Social Psychology 45 (5) (2009) 1117–1122. doi:10.1016/j.jesp.2009.06.019.

[14] X. Fang, S. Singh, R. Ahluwalia, An examination of different explanations for the mere exposure effect, Journal Of Consumer Research 34 (1) (2007) 97–103. doi:10.1086/513050.

[15] A. Gmuer, M. Siegrist, S. Dohle, Does wine label processing fluency influence wine hedonics?, Food Quality And Preference 44 (2015) 12–16. doi:10.1016/j.foodqual.2015.03.007.

[16] N. Schwarz, Metacognitive experiences in consumer judgment and decision making, Journal Of Consumer Psychology 14 (4) (2004) 332–348. doi:10.1207/s15327663jcp1404_2.

[17] R. Silva, T. Garcia-Marques, R. Reber, The informative value of type of repetition: Perceptual and conceptual fluency influences on judgments of truth, Consciousness And Cognition 51 (2017) 53–67. doi:10.1016/j.concog.2017.02.016.

[18] L. Jacoby, M. Dallas, On the relationship between autobiographical memory and perceptual learning, Journal Of Experimental Psychology: General 110 (1981) 306–304.

[19] R. Reber, P. Winkielman, N. Schwarz, Effects of perceptual fluency on affective judgments, Psychological Science 9 (1) (1998) 45–48. doi:10.1111/1467-9280.00008.

[20] J. Seamon, P. Williams, M. Crowley, I. Kim, S. Langer, P. Orne, D. Wishengrad, The mere exposure effect is based on implicit memory: Effects of

stimulus type, encoding conditions, and number of exposures on recognition and affect judgements, Journal Of Experimental Psychology: Learning, Memory And Cognition 21 (3) (1995) 711–721.

[21] A. Lee, A. Labroo, The effect of conceptual and perceptual fluency on brand evaluation, Journal Of Marketing Research 41 (2) (2004) 151–165. doi:10.1509/jmkr.41.2.151.28665.

[22] E. Boyland, J. Halford, Television advertising and branding: Effects on eating behaviour and food preferences in children, Appetite 62 (2013) 236–241. doi:10.1016/j.appet.2012.01.032.

[23] J. Rossiter, R. Silberstein, P. Harris, G. Nield, Brain-imaging detection of visual scene encoding in long-term memory for tv commercials, Journal Of Advertising Research 41 (2) (2001) 13–21. doi:10.2501/jar-41-2-13-21.

[24] O. Stitelman, B. Dalessandro, C. Perlich, F. Provost, Estimating the effect of online display advertising on browser conversion, in: Proceedings of the Fifth International Workshop on Data Mining and Audience Intelligence for Advertising (ADKDD 2011), San Diego, California, USA, 2011, pp. 8–16.

[25] C. Perlich, B. Dalessandro, T. Raeder, O. Stitelman, F. Provost, Machine learning for targeted display advertising: transfer learning in action, Machine Learning 95 (1) (2013) 103–127. doi:10.1007/s10994-013-5375-2.

[26] E. Schwartz, E. Bradlow, P. Fader, Customer acquisition via display advertising using multi-armed bandit experiments, Marketing Science 36 (4) (2017) 500–522. doi:10.1287/mksc.2016.1023.

[27] M. Shareef, B. Mukerji, Y. Dwivedi, N. Rana, R. Islam, Social media marketing: Comparative effect of advertisement sources, Journal Of Retailing And Consumer Servicesdoi:10.1016/j.jretconser.2017.11.001.

[28] B. Jansen, K. Moore, S. Carman, Evaluating the performance of demographic targeting using gender in sponsored search, Information Processing & Management 49 (1) (2013) 286–302. doi:10.1016/j.ipm.2012.06.001.

[29] Y. Zhang, W. Zhang, B. Gao, X. Yuan, T. Liu, Bid keyword suggestion in sponsored search based on competitiveness and relevance, Information Processing & Management 50 (4) (2014) 508–523. doi:10.1016/j.ipm.2014.02.004.

[30] R. Kannan, G. Ghinea, S. Swaminathan, What do you wish to see? a summarization system for movies based on user preferences, Information Processing & Management 51 (3) (2015) 286–305. doi:10.1016/j.ipm.2014.12.001.

[31] I. Choi, M. Oh, J. Kim, Y. Ryu, Collaborative filtering with facial expressions for online video recommendation, International Journal Of Information Management 36 (3) (2016) 397–402. doi:10.1016/j.ijinfomgt.2016.01.005.

[32] A. Alalwan, Investigating the impact of social media advertising features on customer purchase intention, International Journal Of Information Management 42 (2018) 65–77. doi:10.1016/j.ijinfomgt.2018.06.001.

[33] J. Lee, I. Hong, Predicting positive user responses to social media advertising: The roles of emotional appeal, informativeness, and creativity, International Journal Of Information Management 36 (3) (2016) 360–373. doi:10.1016/j.ijinfomgt.2016.01.001.

[34] M. Shareef, Y. Dwivedi, V. Kumar, U. Kumar, Content design of advertisement for consumer exposure: Mobile marketing through short messaging service, International Journal Of Information Management 37 (4) (2017) 257–268. doi:10.1016/j.ijinfomgt.2017.02.003.

[35] C. Cortes, V. Vapnik, Support-vector networks, Machine Learning 20 (3) (1995) 273–297. doi:10.1007/bf00994018.

[36] T. Chen, C. Guestrin, Xgboost: A scalable tree boosting system, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'16), San Francisco, California, USA, 2016, pp. 785–794. doi:10.1145/2939672.2939785.

[37] S. Walker, D. Duncan, Estimation of the probability of an event as a function of several independent variables, Biometrika 54 (1-2) (1967) 167–179. doi:10.2307/2333860.

[38] C. Janiszewski, Preattentive mere exposure effects, Journal Of Consumer Research 20 (3) (1993) 376–392.
URL http://www.jstor.org/stable/2489354

[39] A. Abernethy, Television exposure: Programs vs. advertising, Current Issues And Research In Advertising 13 (1-2) (1991) 61–77. doi:10.1080/01633392.1991.10504959.

[40] D. Bunn, Audience presence during breaks in television programs, Journal Of Advertising Research 22 (5) (1982) 35–39.

[41] P. Boait, R. Rylatt, A. Wright, Exergy-based control of electricity demand and microgeneration, Applied Energy 84 (3) (2007) 239–253. doi:10.1016/j.apenergy.2006.09.001.

Table 17: Hypothesis 3 t-test: p-values for Purchase Intention behavior

Model	Base	Configuration	Purchase Intention Categories					
			0	1	2	3	4	5
SVM	product	Weekday Time Slot	nan	0.750	nan	0.713	0.359	0.738
		Weekday Only	nan	0.682	nan	0.633	0.300	0.766
	user	Weekday Time Slot	0.874	0.494	0.853	0.934	0.835	0.655
		Weekday Only	0.822	0.991	0.621	0.658	0.806	0.805
XGBoost	product	Weekday Time Slot	0.031	0.910	0.071	0.805	0.514	0.993
		Weekday Only	0.197	0.992	0.051	0.731	0.532	0.994
	user	Weekday Time Slot	0.946	0.634	0.803	0.838	0.885	0.757
		Weekday Only	0.586	0.886	0.941	0.764	0.880	0.841
Logistic Regression	product	Weekday Time Slot	nan	0.634	0.427	0.464	0.169	0.647
		Weekday Only	nan	0.618	0.324	0.468	0.139	0.652
	user	Weekday Time Slot	0.904	0.823	0.934	0.968	0.746	0.905
		Weekday Only	0.556	0.848	0.315	0.708	0.743	0.907

Appendices

Appendix A. Input Data Details

As explained in section 4.4.1, two data configurations were used for advert viewing time. The detailed features used are shown in Table A.18.

Similarly, as explained in section 4.4.2, demographic data was used as input for predictions in a number of our experiments. The detailed features used in the input vectors are described in Table A.19.

Appendix B. Data Distribution

In this section we will describe the data we received from the Nomura Research Institute, Ltd., and the distribution and nature of products, adverts and prediction targets.

Appendix B.1. Surveyed Products

The surveys of purchase behavior taken in January 2017 and March 2017 included 200 products, from which only 36 were matched to television adverts during the period between both surveys. Because most of the products are sold only in Japan, a general description of their nature and distribution is explained in Figure B.1.

Appendix B.2. Advert Exposure and Broadcasting Data

The data we received from the Nomura Research Institute, Ltd. included the surveyees' household television viewing times and the program that was displayed when television was on. By matching this data with the adverts that were in between breaks from those programs for the products that were surveyed, we obtained the advert exposure time for each user for each product. In this study we explore the possibility of there being some difference in effect depending on the time slot, particularly the Primetime (19:00 to 23:00) time slot. In Figure B.2 we show the broadcasting time distribution for the programs

Table A.18: Viewing time analysis elements

Data Configuration	Advert Viewing Time in seconds Data Features
Weekdays	• Monday • Tuesday • Wednesday • Thursday • Friday • Saturday • Sunday
Weekday Time Slots	• Monday Primetime • Monday Non-Primetime • Tuesday Primetime • Tuesday Non-Primetime • Wednesday Primetime • Wednesday Non-Primetime • Thursday Primetime • Thursday Non-Primetime • Friday Primetime • Friday Non-Primetime • Saturday Primetime • Saturday Non-Primetime • Sunday Primetime • Sunday Non-Primetime

Table A.19: Demographic data used in input vectors

Survey Data	Possible Answers
Age	• 18 to 25 years old • 26 to 35 years old • 36 to 45 years old • 46 to 55 years old • 56 or older
Sex	• Male • Female
Marital Status	• Single • Married • Divorced or Widowed
Parental Status	• Parent • Not a Parent
Income Bracket	• Not disclosed • No Income • Under 1,000,000 yen • From 1,000,000 yen to 2,000,000 yen • From 2,000,000 yen to 3,000,000 yen • From 3,000,000 yen to 4,000,000 yen • From 4,000,000 yen to 5,000,000 yen • From 5,000,000 yen to 6,000,000 yen • From 6,000,000 yen to 7,000,000 yen • From 7,000,000 yen to 10,000,000 yen • From 10,000,000 yen to 15,000,000 yen • From 15,000,000 yen to 20,000,000 yen • Over 20,000,000 yen

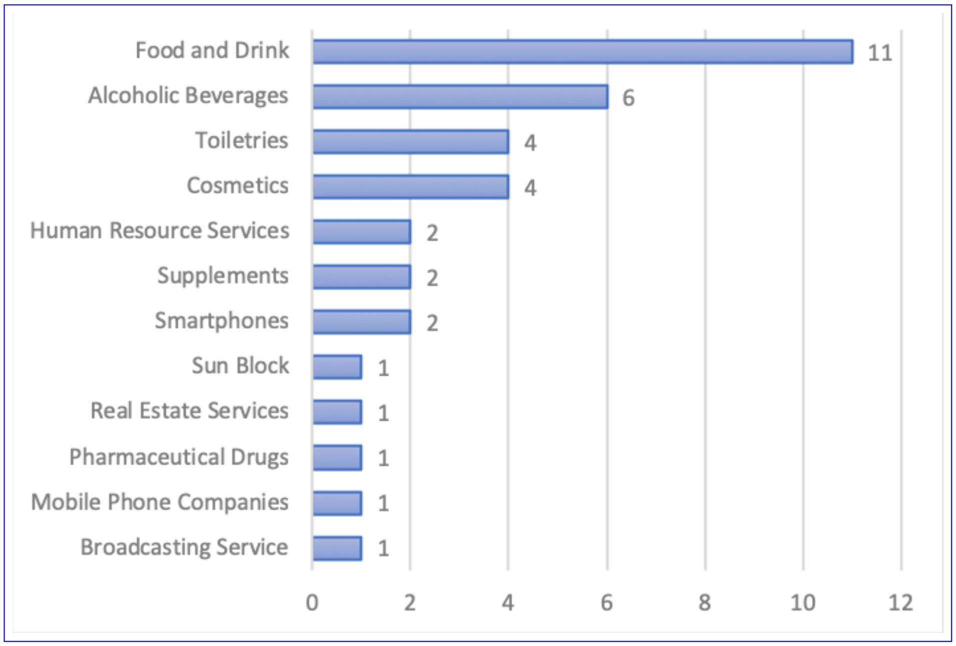

Figure B.1: Products Matched with Advert Viewing Data

that displayed these 36 products during the period of time in between the survey in January 2017 and the survey in March 2017. In Figure B.3 we show the total sum of exposure time in seconds across all users and products for the Primetime and otherwise time slots for each day of the week.

Appendix B.3. Prediction Target Categories

The surveys included data of purchase intention and actual purchase at the times of January 2017 and March 2017 for 200 products, 36 of which were matched with television advert viewing data. As was explained in section 4.3.2 and in Tables 3 and 4, we divided the data in 6 categories (0 to 5) in order to observe the changes in time for these purchase behaviors. The distributions of these categories are shown in Table B.20. Note that categories 4 and 5, by their nature, are a sum of categories 2 and 3, and 0 and 1 respectively.

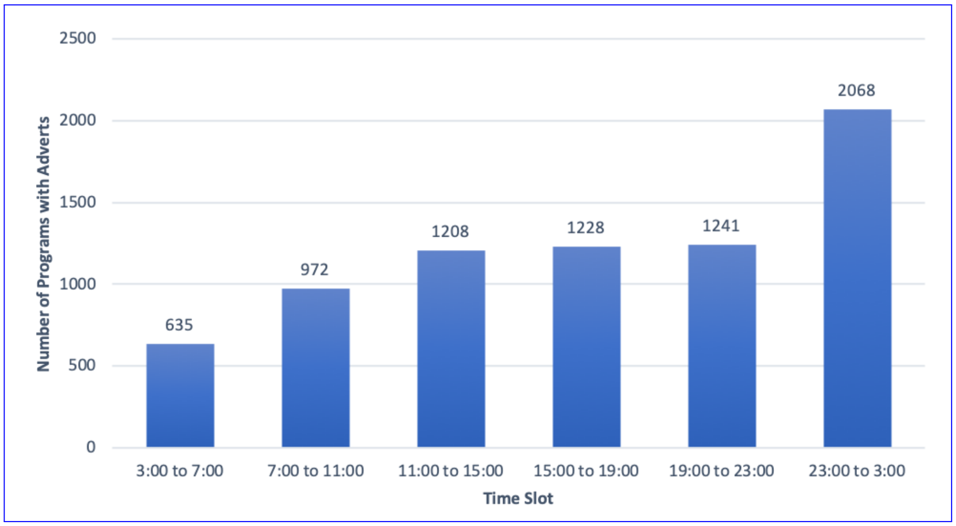

Figure B.2: Programs including adverts broadcast time distribution

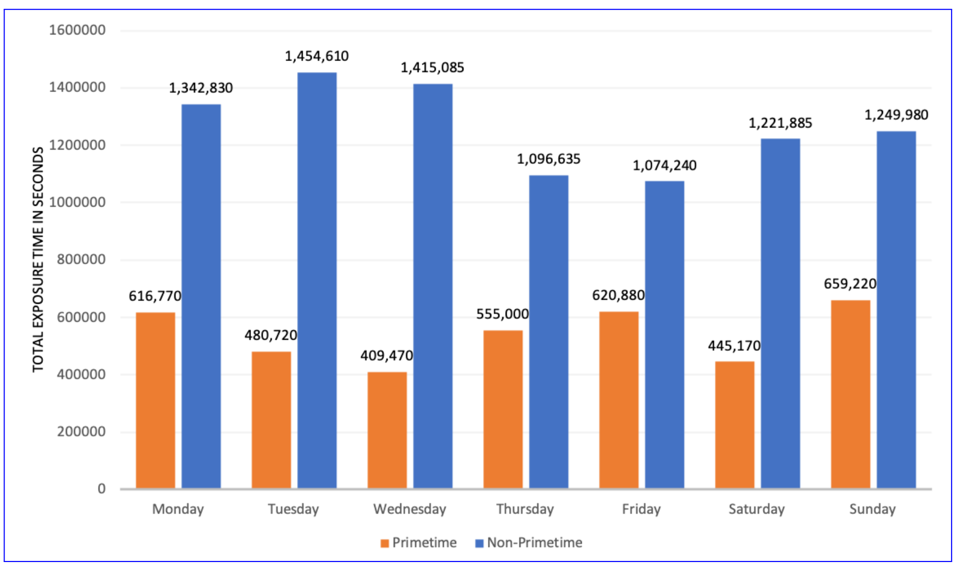

Figure B.3: Advert Exposure Time for all users and products by Weekday and Time Slot

Table B.20: Prediction Target Categories Distribution

	All products (200)		Advert matched products (36)	
Category	Actual Purchase	Purchase Intention	Actual Purchase	Purchase Intention
0	6%	8%	6%	8%
1	73%	59%	76%	58%
2	8%	9%	7%	8%
3	13%	24%	10%	26%
4	21%	33%	17%	35%
5	79%	67%	83%	65%