Predicting Smartphone Upgrades through Deep Learning Classification Models

Autoria

Vinicius Andrade Brei - brei@ufrgs.br
Prog de Pós-Grad em Admin/Esc de Admin - PPGA/EA/UFRGS - Universidade Federal do Rio Grande do Sul

Leonardo Nicolao - leonardo.nicolao@ufrgs.br
Prog de Pós-Grad em Admin/Esc de Admin - PPGA/EA/UFRGS - Universidade Federal do Rio Grande do Sul

Maria Alice Pasdiora - mariaalice.pasdiora@gmail.com
Prog de Pós-Grad em Admin/Esc de Admin - PPGA/EA/UFRGS - Universidade Federal do Rio Grande do Sul

Rodolfo Coral Azambuja - rodolfoazambuja@gmail.com
Prog de Pós-Grad em Admin/Esc de Admin - PPGA/EA/UFRGS - Universidade Federal do Rio Grande do Sul

Agradecimentos

The first author expresses his gratitude for the post-doc scholarship 200863/2018-5 received from CNPq - National Council for Scientific and Technological Development ? Brazil. All the authors express their gratitude to the Cátedra Tramontina Eletrik for the funding that allowed the data collection.

Resumo

This manuscript explains which variables are more relevant to predict which consumers will replace their products (smartphones) in the future. Data was collected through a longitudinal consumer panel, which measured upgrade decisions and independent variables classified into eight main groups: ownership, enjoyment, desire, perceptions about the smartphone, context, individual traits, and demographics. We tested two types of non-linear, state-of-the-art machine-learning models to explain upgrade behavior: Extreme Gradient Boosting (decision-tree) and two integrative Deep Learning models. Results provide a comprehensive, yet parsimonious model showing ownership, enjoyment, and context variables as the most relevant to determine which consumers are more prone to replace smartphones. Our findings enhance previous understanding of upgrade decision theory by taking a holistic approach and bridging different theoretical accounts. Further, results contribute to marketing theory and practice, shedding new light on the understanding of consumer decision making when upgrading products. Managers may apply findings to identify which variables are more relevant to influence consumer decision process regarding new products. Therefore, they can adapt their marketing strategy. Consumers, in turn, may be aware of how their behavior can be shaped by marketing actions and, then, react accordingly, making better decisions.

Predicting Smartphone Upgrades through Deep Learning Classification Models

ABSTRACT
This manuscript explains which variables are more relevant to predict which consumers will replace their products (smartphones) in the future. Data was collected through a longitudinal consumer panel, which measured upgrade decisions and independent variables classified into eight main groups: ownership, enjoyment, desire, perceptions about the smartphone, context, individual traits, and demographics. We tested two types of non-linear, state-of-the-art machine-learning models to explain upgrade behavior: Extreme Gradient Boosting (decision-tree) and two integrative Deep Learning models. Results provide a comprehensive, yet parsimonious model showing ownership, enjoyment, and context variables as the most relevant to determine which consumers are more prone to replace smartphones. Our findings enhance previous understanding of upgrade decision theory by taking a holistic approach and bridging different theoretical accounts. Further, results contribute to marketing theory and practice, shedding new light on the understanding of consumer decision making when upgrading products. Managers may apply findings to identify which variables are more relevant to influence consumer decision process regarding new products. Therefore, they can adapt their marketing strategy. Consumers, in turn, may be aware of how their behavior can be shaped by marketing actions and, then, react accordingly, making better decisions.

Keywords: upgrade, product replacement, longitudinal panel, deep learning, machine learning.

INTRODUCTION

Every year, like clockwork, companies line up at their conferences, trade shows, through press releases to announce their latest and greatest. They want their consumers to know that there is a new version of their product available. A better option of their beloved pickup truck, a new version of the well-established productivity software, an updated smartphone with better specs and better cameras, or even a new flavor of the classic sandwich cookie. Periodically, new options abound. The idea is not only to generate sales through the acquisition of new customers, but to maintain current customers engaged with the brand, through upgrading opportunities.

In fact, Apple believes that many of its already existing iPhone users will update to new iPhone models between February 2018 and July 2019 (Fortune, 2018). They are not necessarily replacing faulty devices, but shortening the lifespan of their functioning phones (AppleInsider, 2018). This allows us to infer that most of the potential upgrade sales for the next months will correspond to the substitution of handsets that remain in active use and in good conditions to be used for a few more years. This manuscript is concerned with understanding why many consumers prematurely upgrade their products, leaving years of functionality on the table and, consequently, overspending.

The impetus to understand customer's adoption and diffusion of new products is widespread and established (Rogers, 1976; Lenk & Rao, 1990; Mahajan, Muller & Bass, 1990, 1995; Montaguti & Zammit, 2017). However, upgrading products, a ubiquitous behavior where consumers acquire an updated version of a product already in their possession, is less documented. For instance, it is estimated that 42% of current iPhone users will upgrade to a newer iPhone model (Munster, 2018). Yet, few studies have tackled this phenomenon (Bayus, 1991; Okada, 2001, 2006; Bellezza, Ackerman & Gino, 2017; Miller, Wiles & Park, 2019), much less comprehensively.

For example, Bayus (1991) has focused on the individual traits more associated with those who replace a product in the early and late stages of its lifecycle. His effort has identified different profiles for different replacement timings, but it was limited as to why the replacement was made or which contextual elements would influence such trade. Grewal, Mehta & Kardes (2004) have also investigated which attitudinal functions (i.e., individual characteristics) lead to replacement decisions. Although these effects are conditioned to product characteristics (public or private, for example), they do not include context and process measures. Okada (2001, 2006) proposes an explanatory process leading to replacements, where consumers consider the "mental book value", the weighing of the initial monetary value of a product and the cumulative enjoyment derived from this product. Finally, Bellezza, Ackerman, and Gino (2017) describe an interesting phenomenon where replacement is driven by a consumer's carelessness with her current product, an attempt to justify their desire to upgrade. Both Okada's and Belezza et al.'s explanations are insightful, but do not include contextual or individual differences.

Given the unknowns of how, why and when individuals upgrade their current products and the lack of studies tackling the phenomenon in its entirety, we propose a comprehensive, integrative model of product upgrade, including personal differences, product characteristics, context, and psychological processes. We tested two types of non-linear machine-learning models, settling on a Deep Learning neural network model, on a longitudinal data set of iPhone and Samsung cellphone consumers with six rounds of data collection during one year (one wave of data collection every two months). This data set provided natural occurrences of product upgrades (dependent variable) as well as individual characteristics (e.g., materialism, involvement with the category), product and usage characteristics (e.g., frequency of usage, hedonic value), context (e.g., news about the brand, buzz) and process measures (e.g., hedonic adaptation, desire). Cell phones were chosen for their widespread adoption in the market and a relatively fast upgrade cycle (both Apple and Samsung release new models at least annually).

The manuscript is structured as follows. First, pertinent literature on replacement decisions is reviewed. In the following section, dataset is explored and procedures to assembly longitudinal consumer panel are described. Next, an explanation of data preparation is followed by the model choice strategy and, ultimately, model selection. Finally, results from the explanatory upgrading model are explored.

REPLACEMENT DECISIONS

The long-term ownership of a durable often involves decisions on its replacement. Consumers replace durables for two main reasons. The first is the poor performance of the status quo, which causes forced replacements. The second are the innovations and enhancements in a product category, which stimulate unforced replacements (Grewal, Mehta & Kardes, 2004). For purposes of this study, "product replacement" will be considered as an unforced replacement decision that encompasses the substitution of a good for its upgrade, which is an enhanced version in the same category. In comparison to regular purchases, replacement decisions present some unique properties that have been explored by the literature on marketing.

One example is the work from Bayus (1991), who developed a model incorporating demographic characteristics, attitudes and perceptions, and search behavior for differentiating consumers who replace a product during the early and late parts of its lifetime. His results demonstrate that early replacers usually have higher educational achievement and occupational status. They are also more concerned with styling, while late replacers are more concerned with cost-related attributes and engage in more search activity before replacing. Finally, early replacers usually upgrade due to reasons such as changes in preference, and late replacers upgrade because of performance reasons.

Research stream on replacement behavior also investigated the determinants and consequences of the length of purchase intervals. Grewal, Mehta, and Kardes (2004) suggested that attitude functions (i.e., knowledge, value expressive, social adjustive, and utilitarian) help explain and predict interpurchase intervals. This effect is contingent on the product nature (along public-private and luxury-necessity dimensions) and the nature of the decision (forced or unforced purchase decision). Complementarily, Gordon (2009) built a model of consumer demand that accounts for replacement decisions when consumers are uncertain about future price and quality. His findings demonstrate that the length of replacement cycles, which is determined by quality and technical features of the status quo, is a useful dimension for segmenting consumers and, consequently, for predicting which consumers are more likely to replace in the near future.

Other works draw attention to the importance of ownership time. According to Cripps and Meyer (1994), consumers' decisions to replace are more influenced by the time elapsed since the last substitution than by the lag between the expected and delivered performance of the status quo. Assuming that the decision to replace in not a binary choice, Miller, Wiles, and Park (2019), investigated how trade-in (i.e., the status quo given as part of the payment) characteristics and the marginal costs-benefits of the new purchase influence the degree of upgrade. According to their findings, trade-in ownership time and brand loyalty enhance the replacement degree of upgrade.

Willingness to upgrade is also influenced by the mental costs of retiring the old model before consumers have gotten their full money's worth out of it. Okada (2001) explains that mental costs depend on product's mental book value, which refers to the difference between the initial purchase price and the cumulative enjoyment up to the potential replacement point. When the cumulative enjoyment from consumption increases to a point where it equals the purchase price, the net entries in the account become zero and consumers feel they have gotten their money's worth from the old reusable. Her results demonstrate that replacement decisions may be more sensitive to this mental cost than to any attribute of the new model itself, such as price and quality.

These mental costs, however, can be alleviated by product features and consumer perceptions. Okada (2006) complements her previous findings by showing that, because dissimilarity turns the sunk cost in the existing product less salient, psychological costs become a lesser impediment to upgrading when consumer perceives the enhanced product as dissimilar to the status quo. Recent research shows that consumers unconsciously manage the status quo's write-off when they want to make a justifiable replacement decision. Bellezza, Ackerman, and Gino (2017) examined the potential for consumers being careless with current possessions in the presence of appealing product upgrades. Because "accidentally" damaging a product or running out of it quickly allows consumers to write off the residual value of the product and upgrade without recording a loss or appearing wasteful, the authors suggest that such careless tendencies are intended to promote the acquisition of upgrade products by helping consumers to justify the new purchase.

Previous research also concentrated its efforts on explaining how replacement decisions are guided rather by subjective factors than by optimal replacement principles. According to these principles, a replacement should occur when the decision maker perceives that the lag between expected and delivered performance exceeds some threshold. Literature demonstrates, however, that upgrade decisions may be influenced by agents as subtle as consumers' desire to differentiate themselves from dissimilar users of the same brand (Wang & Roedder John, 2019), inaccurate predictions about future use of additional capabilities (Meyer, Zhao & Han, 2008), and anticipated regret derived from the fear of prematurely adopting the current best technology and missing out on the future technology when it becomes available (Shih & Schau, 2011).

To sum up, the most frequent determinants of replacement investigated by the literature can be classified in five main groups: factors related to ownership of status quo, perception about the status quo, context variables, individual traits, and demographic characteristics. To the best of our knowledge, our work is the first to combine these groups of variables to provide a comprehensive, integrative model of the factors explaining the decision to upgrade. Beyond these factors already examined by past research, we investigate whether the enjoyment derived from the status quo and the desire for the upgrade affect the decision to replace.

Motivation to explore the influence of enjoyment and desire on replacement decisions comes from findings of Okada (2001) and Bellezza, Ackerman, and Gino (2017). Results from Okada (2001) highlight the relevance of enjoyment for the decision to replace. While her work focused on expected enjoyment, we rely on hedonic adaptation literature to propose that consumers' current enjoyment with the status quo influences replacement decisions. Hedonic adaptation refers to a reduction in the affective intensity of favorable and unfavorable circumstances and, in its broadest sense, corresponds to any action, process, or mechanism that reduces the effects of a constant, repeated stimulus (Frederick & Loewenstein, 1999). In the consumption context, ongoing ownership and repeated usage leads to a decreased hedonic response in the form of less desire and less ongoing enjoyment (Galak & Redden, 2018).

Although Bellezza, Ackerman, and Gino (2017) did not present any measure of desire, their work implies that the desire for the upgrade influences how consumers treat their status quo. We take the approach of Boujbel and d'Astous (2015), who proposed that the experience of consumption desire must be understood as an aggregation of affective and cognitive psychological events.

In summary, the present work combines factors related to ownership of status quo, perception about the status quo, context variables, individual traits, demographic characteristics, enjoyment with the status quo, and desire for the upgrade to develop a model that explains the decision to upgrade. It enhances prior understanding by showing which variables are more relevant to predict which consumers will replace. An overview of the relevant literature on upgrading and how this paper differs from it is shown on Table 1.

Table 1. Relevant Literature on Upgrade Decision

	Data	**Longitudinal Data**	**Ownership Variables**	**Enjoyment**	**Desire**	**Perception**	**Context**	**Individual Traits**
This Research	Real Transactions	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Miller et al. (2019)	Real Transactions	No	No	No	No	No	Yes	No
Wang & Roedder John (2019)	Scenario	No	No	No	No	No	No	Yes
Sela & LeBoeuf (2017)	Scenario	No	No	No	No	Yes	No	No
Bellezza, Ackerman & Gino (2017)	Real Transactions Scenario	Yes	No	No	No	No	No	No
Shih & Schau (2011)	Scenario	No	No	No	No	No	No	No

Gordon (2009)	Real Transactions	Yes	No	No	No	No	Yes	No
Meyer, Zhao & Han (2008)	Scenario	No	No	No	No	Yes	Yes	No
Okada (2006)	Scenario	No	No	No	No	Yes	Yes	No
Grewal, Mehta & Kardes (2004)	Real Transactions	No	No	No	No	No	Yes	No
Okada (2001)	Scenario	No	No	Yes	No	No	No	No
Cripps & Meyer (1994)	Scenario	No	Yes	No	No	Yes	No	No
Bayus (1991)	Real Transactions	No	Yes	No	No	Yes	Yes	No

To understand what drives someone to upgrade her status quo product, we created and mantained a consumer panel measuring its participant's behavioral and psychological traits during one full year. This panel provided answers to when, how, and why people decide to upgrade and actually upgraded their products. The next section describes the panel in detail.

LONGITUDINAL CONSUMER PANEL

A longitudinal consumer panel was designed to reach a deeper understanding of replacement behavior and its underlying reasons. Through the panel, we collected data from Apple and Samsung smartphone owners for one year. Smartphones were chosen for several reasons. First, they are largely consumed worldwide. In 2018, about 39% of the world population had a smartphone (VentureBeat, 2018). Second, its market is competitive and dominated by two major global players, Apple and Samsung (Forbes, 2015). Third, both major players have been launching upgrades for their best-selling smartphones systematically in September (Apple) and in April (Samsung), suggesting a regular upgrade cycle. Finally, these launches are accompanied by strong marketing investments, substantial media coverage, and social media buzz. Taken together, these features create a research environment that is richer and more insightful than the launch of other widespread durables (e.g., refrigerators and televisions).

Procedures. Due to market characteristics, we collected data every two months through six survey rounds on Amazon Mechanical Turk (MTurk). The first round was in August 2015 and the last in June 2016. MTurk was chosen because it allows a substantial geographical reach, easy contact with participants, fast data collection, and control of whether the same person is participating in the six rounds (through MTurk worker id). To avoid influence of cultural traits and different market conditions, the sample was restricted to English speakers, U.S.-based participants.

Initial sample size was 730 participants. Different measures to avoid sample attrition were taken. For example, payment was gradually increased as the study progressed (from USD 0.30 to USD 2.00), and a one-dollar bonus was paid for those who completed all rounds. Aiming to keep the answer time similar for all rounds, time-invariant psychological variables were distributed along the six rounds. Final sample, however, was of 144 participants. This attrition rate is consistent with previous longitudinal data collection (Kim et al., 2015; Douglass, Duffy & Autin, 2016). Of the 144 participants, 22 upgraded their smartphones. The dependent variable was upgrade behavior. The independent variables, their operationalization, and their correlation with the dependent variable are described in Table 2.

Table 2. Independent Variables

	Description	Measure	Correlation with Upgrade
Ownership Variables			
Contract	Whether consumers signed up for a contract when they purchased the smartphone	T1 – T6	0.022
Ownership time - T1	Number of months consumers owned their smartphone before exchanging	T1 – T6	0.104
Status quo meets current needs	How much the smartphone meets consumers' current needs	T1 – T6	0.386*
Importance of resources	Importance assigned by consumers to resources of the smartphone	T1 – T6	0.374*
Update Capacity	Update capacity of the apps installed in consumers' smartphone	T1 – T6	0.340*
Enjoyment Variables			
Enjoyment	How much consumers enjoy their smartphone	T1 – T6	0.316*
Enjoyment - purchase time	How much consumers enjoyed their smartphone in the day of purchasing	T1 – T6	0.370*
Enjoyment - 1 month ago	How much consumers enjoyed their smartphone one month ago	T1 – T6	0.206
Enjoyment - 1 month ahead	How much consumers believe they will enjoy their smartphone in one month	T1 – T6	0.436*
Enjoyment - 1 year ahead	How much consumers believe they will enjoy their smartphone in one year	T1 – T6	0.311*
Desire Variables			
Craving experience	Craving Experience Questionnaire (May et al., 2014) –	T2 – T6	-0.011
Desire - feelings and thoughts	Feelings and thoughts related to consumer desire (Boujbel & d'Astous, 2015)	T3	0.196
Desire - new smartphone	How much consumers desire to exchange their smartphone	T1 – T6	0.075
Desire - new smartphone of same brand	How much consumers desire to exchange their smartphone for a newer model of the same brand	T1 – T6	0.125
Desire - unique products	Desire for Unique Products Scale (Lynn & Harris, 1997)	T6	0.062
Perceptions about the Smartphone			
Usage frequency	How much consumers currently use their smartphone	T6	0.108

Usage frequency - 1 month ago	How much consumers used their smartphone one month ago	T6	0.084
Usage frequency - purchase time	How much consumers used their smartphone in the purchase day	T6	0.121
Hedonic x Utilitary	How consumers classify their smartphones in a continuum between hedonic and utilitarian	T1 – T6	0.249
Material x Experiential	How consumers classify their smartphones in a continuum between experiential and material	T1 – T6	0.239
Mental book value	Feeling of having gotten money's worth from the smartphone (Okada, 2001)	T1 – T6	0.722*
Sentimental value	How much sentimental value consumers assign to the smartphone	T1 – T6	0.294*
Context Variables			
News about Galaxy	Consumers' degree of contact with news about Galaxy in the last month	T1 – T6	0.286*
News about iPhone	Consumers' degree of contact with news about iPhone in the last month	T1 – T6	0.314*
News about same brand	Consumers' degree of contact with news about smartphones of the same brand of their own in the last month	T1 – T6	0.347*
Individual Traits			
Involvement	Involvement Scale (Zaichowsky, 1994)	T2	0.091
Materialism	Materialism Scale (Richins, 2004)	T2	0.100
Self-control	Self-control Scale (Tangney, Baumeister & Boone, 2004)	T4	0.039
Life satisfaction	Life Satisfaction Scale (Pavot & Diener, 1993)	T4	-0.043
Anxiety	Anxiety Scale (Lau-Gesk & Meyers-Levy, 2009)	T5	0.011
Thoughts about future	How much consumers think about future incomes	T5	0.194
Social comparison	Social Comparison Scale (Gibbons & Buunk, 1999)	T6	0.051
Demographic Variables			
Gender - Female	Gender - only female	T1	0.068
Gender - Male	Gender - only male	T1	-0.068
Age	Age in years	T1	0.035
Education	Education level	T1	-0.008
Annual household income - T1	Annual household income	T1	0.052
Annual household income - T2	Annual household income	T2	0.137
Discretionary income	Percentage of household income considered as discretionary	T1	0.051

*Correlation significant at $p < 0.05$.

DATA PREPARATION AND ANALYTICAL STRATEGY

Analytical strategy was composed of four stages: (1) data preparation, (2) choice of models and algorithms to be estimated; (3) simulation of different regularization, cross-validation, and tuning parameters; and (4) performance comparison. All analyses carried out were intended to determine which variables explain upgrade decision. In other words, this was a classification problem of who have and who have not upgraded.

Data preparation. Preparation involved data selection, preprocessing, and transformation. Data set was composed of 33 behavioral and psychological independent variables and 6 demographic variables (shown in Table 2). Among these variables, there were cross-sectional and longitudinal, as well as multi- and single-item scales. Such heterogeneity required varied procedures for data preparation.

Longitudinal variables were operationalized through the difference between the rating in the first round (T1) and the rating in the round of the upgrade (Tupgrade). For participants who did not upgrade, the difference was calculated between T1and T6. This approach was taken because of our interest in the magnitude of behavioral change during the panel, and not in the serial correlation of each variable. Further, multi-item scales were summarized through factorial loads of the first PCA component. Because of requirements of one of tested models (Chollet & Allaire, 2018), all variables were standardized (z-scores) and categorical data were converted to numerical through one-hot encoding.

Model choice. Two types of non-linear, supervised Machine Learning (ML) models were run: a decision-tree model (eXtreme Gradient Boosting, hereafter XGB) and a Deep Learning neural network model, to date the most advanced algorithms for classification (Becker, 2018; Chollet & Allaire, 2018). These models were preferred instead of traditional linear classification models, such as Logistic Regression (LR) and Linear Discriminant Analysis (LDA), for two reasons. First, because of the substantial likelihood of non-linearity and multiple interactions among the variables in dataset. Second, traditional classification models, such as LR and LDA, are usually estimated using the full data set, a practice that does not allow the test of accuracy with data not "seen" by the model. ML models, in contrary, split the data into training and test sets. Thus, the model is estimated in a set and evaluated in another one. In this sense, the prediction ability of the model is tested with data which was not "seen" in the estimation step (Goodfellow et al., 2016). Finally, these models were chosen because they provide not only performance indices, but also ways of identifying which variables are more relevant for prediction. XGB model was tested using the XGB package (He, 2018) and the Deep Learning model using the Keras package (Allaire & Chollet, 2018), both in their versions for R. The code for reproducing our analysis is available at request.

Regularization, cross-validation, and model tuning. Because of large dataset and relatively reduced sample size, we were concerned with the relationship between sample size/number of independent variables and overfitting. To overcome these problems, we applied regularization techniques early stopping, dropout, and Ridge and Lasso shrinkage while running XGB and Deep Learning models. By imposing a size constraint on the coefficients, we reduced the issues of including non-relevant variables on the model and of having a large positive coefficient on one variable canceling a similarly large negative coefficient on another correlated variable, improving model performance (Trevor, Robert & JH, 2009).

When estimating XGB and Deep Learning models, the difference between test error rate and the training error rate could also impose a problem. Since dataset was not large, both training and test samples were likely to generate low power estimates. To minimize such a risk, we estimated the test error rate (i.e., accuracy) by applying K-fold cross-validation (Trevor, Robert & JH, 2009) through the caret R package (Kuhn, 2018). The data was divided into K=10 roughly equal parts. For each kth part, the model was fit to the other K – 1 parts of the dataset, and calculated the prediction error of the fitted model when predicting the kth part of the data. This was done for k = 1, 2…, K, and then combined the K estimates of prediction error. More formally, the cross-validation estimate of prediction error was calculated as follows.

$$CV(\hat{f}) = \frac{1}{N}\sum_{i=1}^{N} L\left(y_i, \hat{f}^{-k(i)}(x_i)\right) \tag{1}$$

Where k: {1, ..., N} ↦ {1, ..., K} is an indexing function that indicates the partition to which observation i is allocated by the randomization; and $\hat{f}^{-k}(x)$ the fitted function, computed with the *k*th fold removed. Each model was estimated in 90 randomly selected training datasets and tested in 10 randomly selected datasets. Next, we estimated parameters' tuning of XGB and Deep Learning. Thus, aiming to improve model accuracy and prediction capacity, we tested different hyperparameters that optimize each model architecture. Best hyperparameters were found by applying Automated Machine Learning through the H2O platform (H2O, 2019).

Model Performance. Classification models are evaluated by the proportion of events they correctly predict. In the present work, that reflects how much upgrades and not upgrades the model accurately classifies. XGB and Deep Learning were compared in relation to three performance parameters: Accuracy, Receiver Operating Characteristic (ROC) curve, and F1 score (Murphy & Bach, 2012) (see Table 3). First, we created Confusion Tables (see Table 4) comparing the true positive rate (TPR) and the false positive rate (FPR) of each model. These results were useful to the calculation of accuracy (calculated as the number of corrected predictions divided by the total number of predictions) and the generation of the ROC curve, which evaluates the capacity of the model as a binary classifier system. We created the ROC curve by plotting the true positive rate (TPR) against the false positive rate (FPR) at a variety of threshold settings. ROC curve allows to compare the predictive power as a function of the Type I Error (MathWorks, 2018) through a single value indicating the area under the curve (AUC). Complementarily, we calculated F1 score, which is the harmonic mean of precision and recall (Murphy & Bach, 2012).

$$F \triangleq \frac{2}{\frac{1}{P}+\frac{1}{R}} = \frac{2PR}{R+P} \tag{2}$$

Where PR is the positive rate, P is *precision*, defined as $TP/\hat{N}_{+p(y=1 \vee \hat{y}=1)}$ and the *recall* is defined as $TP/N_{+p(\hat{y}=1|y=1)}$. F1 is a particularly relevant indicator when one seeks a balance between precision and recall in a model estimated from a dataset with uneven class distribution (e.g., larger number of non-upgraders) (Shung, 2018).

Table 3. Models performance comparison

	Decision-Tree	**Deep Learning Models**	
	XGBoost	Full Model	Short Model
Accuracy	0.932	0.955	0.977
AUC	0.874	0.979	0.990
F1	0.727	0.833	0.889

In comparison to the XGB (Accuracy = 0.932, AUC = 0.874, F1 = 0.727), performance of de Deep Learning (Accuracy = 0.955, AUC = 0.979, F1 = 0.833) was better. Thus, we adopted the Deep Learning model to further explore the results. This model presented the following architecture: one input layer, two densely connected hidden layers with 50 neurons each, activated by a rectified linear unit (ReLU) function, with shrinkage L1 (Lasso) and L2 (Ridge) both of 0.001, a layer dropout rate of 0.1, and a final output dense layer with one neuron (possible outputs: 1 - Yes or 0 - No) activated by a sigmoid function. The model was compiled through Adam Optimization Algorithm, and the loss function was a binary cross-entropy loss (log loss). On total, the Deep Learning model estimated a total of 4,701 parameters.

Next, the Deep Learning model cited above (full Deep Learning model), composed by all variables, was compared with the short Deep Learning model, composed by the variables that presented a statistically significant correlation with the Upgrade dependent variable (Table 2). The short Deep Learning model presented better performance (Accuracy = 0.977, AUC = 0.990, F1 = 0.889) than the full Deep Learning model.

Table 4. Confusion Tables

		XGBoost		**Deep Learning Full Model**		**Deep Learning Short Model**	
		Truth					
		No	Yes	No	Yes	No	Yes
Prediction	No	37	1	37	0	39	1
	Yes	2	4	2	5	0	4

RESULTS AND DISCUSSION

All three estimated models achieved high levels of precision to estimate both upgraders and non-upgraders, despite the unbalanced distribution of the two categories. Taking into account the classification data, we calculated accuracy, AUC, and F1 scores. Even considering the small difference in performance among the decision-tree XGBoost (XGB) and the two Deep Learning neural networks models, accuracy increased from the XGB (Accuracy = 0.932) compared to the Deep Learning full model (Accuracy = 0.955) and the Deep Learning short model (Accuracy = 0.977). The quality of the ROC, summarized by the area under the curve (AUC), followed the same pattern as the accuracy. The F1 Score is a better measure when there is an uneven class distribution (Shung, 2018). Again, the short Deep Learning model showed better results compared with the XGB and the full Deep Learning model. Considering the evident prominence of the short Deep Learning model in all performance indicators and its simplicity (12 variables) compared to the two full models (33 behavioral and psychological variables, plus 6 demographic covariates), hereafter the results of the short Deep Learning model are discussed.

To evaluate the contribution of each variable to the model predictive power, the model-agnostic interpretation method of Feature Importance was applied. Importance of each variable (or feature) was measured by calculating the increase in the model's prediction error after permuting the feature. To be considered as "important", the exclusion of one variable should increase the model error or reduce its predictive capacity. If a variable exclusion leaves the model error unchanged, it means that the model ignored the feature for the prediction (Molnar, 2019). Feature Importance calculation (Fisher, Rudin & Dominici, 2018) was implemented through the following algorithm:

Algorithm 1: Feature Importance

Input: Trained model f, feature matrix X, target vector y, error measure $L(y, f)$

1. Estimate the original model error $e^{orig} = L(y, f(X))$ (e.g., mean squared error)
2. For each feature j = 1, …, p do:

o Generate feature matrix X^{perm} by permuting feature j in the data X. This breaks the association between feature j and true outcome y.

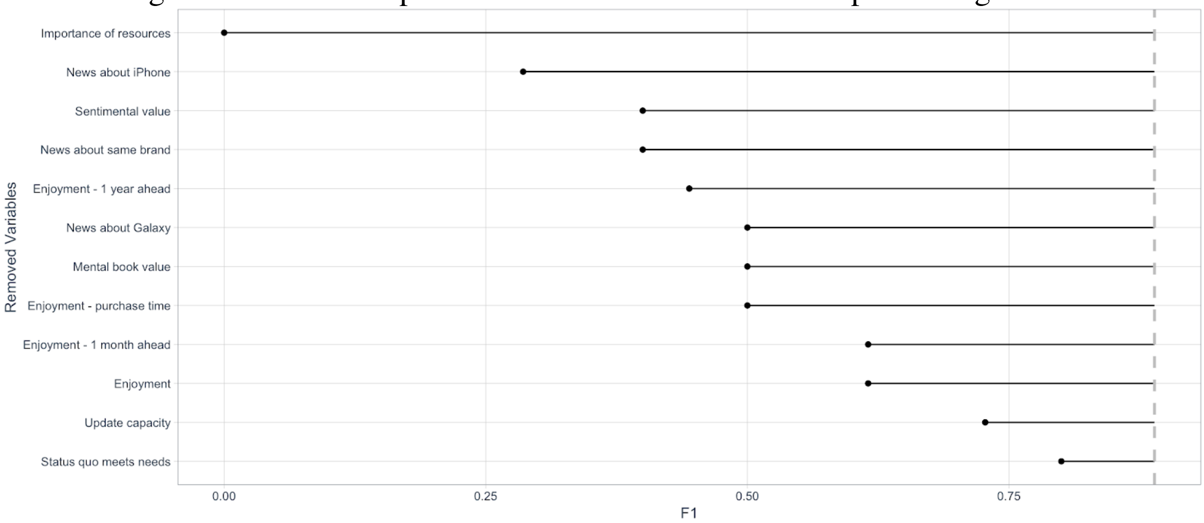
o Estimate error $e^{perm} = L(y, f(X^{perm}))$ based on the predictions of the permuted data.

o Calculate permutation feature importance $FI^j = e^{perm}/e^{orig}$. Alternatively, the difference can be used: $FI^j = e^{perm} - e^{orig}$

3. Sort features by descending FI

Feature Importance algorithm was calculated for the three performance indicators. The three performance results were similar, with minor changes in the relevance ranking of the predictor variables. Figure 1 below presents results of the Feature Importance algorithm. From the set of variables correlated with the upgrade decision (the short Deep Learning model), the most relevant variable to predict the upgrade decision was resources importance and the least important was how much the current smartphone meet one's needs today.

Figure 1. Variable Importance for F1 Scores – Short Deep Learning Model



Considering the variable factors presented on Table 2 and on Figure 1 above, the factors enjoyment, ownership, context, and perception about the smartphone were the more relevant predictors to upgrade decision, in this order of relevance. Throughout this discussion, we will take into account both the contribution of each factor, manifested through their F1 scores, and their correlation to the upgrade decision.

The most important factor to contribute to an explanatory model of upgrading behavior was the change in importance consumers assigned to the resources and features of their current smartphone, from the first to the last round of data collection. As consumers in the sample believed the resources of their current smartphone were less and less important (a positive difference between T1 and T6), the likelihood of upgrade increased. There seems to be a process by which consumers perceive what is being delivered by their current option as of a lesser value than available upgrading options. In fact, this sentiment is corroborated by other ownership variables as whether the product meets its owner's current needs and update capacity. As consumers perceive their device to become less able to update their apps over time, their likelihood to upgrade also increases. Unable to extend the usability of the product and, consequently, its life-cycle, consumers might feel compelled to replace their current options. Overall, upgrading decisions are marked by a sense of a product's suitability to satisfy its consumer needs. This declining trend is captured by a change in whether the status quo (current phone) meets its consumer's needs, followed closely by a decline in the enjoyment associated with consuming or using the product. These findings not only corroborate previous research showing the relevance of perceptions on performance for upgrade decisions (Bayus, 1991; Gordon, 2009), but also enhance this research stream. For example, Bayus (1991) differentiated early and late replacers, giving weight for product lifetime cycle in his conceptualization, and Gordon (2009) emphasized the relevance of the length of replacement cycles for segmenting consumers and for predicting which consumers are more likely to replace. We advance the previous theory showing the impact of the variation in the importance of smartphone resources along ownership time. Thus, our results demonstrate how the fluctuation on resources valuation shapes replacement decisions.

Enjoyment refers to a class of effects closely related to hedonic adaptation. Each time participants answered the questionnaire, they were asked to estimate how much they enjoyed their phones in the past, at purchase time and one month before answering time, and how much they were currently enjoying and how much they would enjoy their phones in the future (one month and one year into the future). For all the enjoyment measures, the score at time 6 (or at time before upgrade, for those participants that bought new smartphones during the panel) was subtracted from the first measure (T1 – Tupgrade). In effect, a hedonic adaptation measure is generated, a picture of how much enjoyment has declined (if positive) or increased (if negative). The differential score of enjoyment at purchase time, of current enjoyment, enjoyment a month and a year into the future were relevant predictors of the upgrade decision, as evidenced by the F1 scores.

As current enjoyment and prospective enjoyment decreased over time, individuals in sample were more likely to upgrade to a new phone. Interestingly, this effect was also evident for enjoyment at purchase time. As individuals in sample recalled their enjoyment at time of purchase as being reduced over time, their likelihood of upgrade increased. There might be motivated reasoning effects in play (Kunda, 1990; Reczek et al., 2018), causing individuals to misremember past enjoyment to justify the upgrade behavior, along the lines of those reported by Bellezza, Ackerman and Gino (2017). Motivated reasoning is a desire to think about and evaluate information in a way that supports a particular directional conclusion. Motivation comprises any wish, desire, or preference that concerns the outcome of a reasoning task (Kunda, 1990). One example of motivated reasoning is the willfully ignorant memory effect reported by Reczek et al. (2018), whereby consumers present better memory for an ethical attribute when a product performs well on the attribute versus when a product performs poorly on the attribute. Specifically, consumers systematically forget or misremember negative ethical attribute information when they face the conflict between their "in the moment" desire to avoid negative ethical information and their general long-term belief that they should be morally good. Their results suggest that consumers forget ethical attributes when this kind of information engenders mental conflicts between the want/should selves by allowing the want self to prevail, just as they do in choice contexts not related to ethicality.

The relationship between a consumer and her smartphone, described through perceptions of the importance of resources, update capacity and overall suitability to satisfy her needs, and her declining (current and prospective) enjoyment with it, is joined by product characteristics to determine upgrading behavior. The sentimental value consumers attach to their products also changes over time. In the sample, a decline in sentimental value between the first and last rounds of data collection is positively associated with a greater likelihood to upgrade. At a similar rate, the changes in mental book value, a perceptual measure of getting one's money worth from ownership and usage of the product (Okada, 2001), also influence upgrading behavior. As mental book value decline, upgrading likelihood increases. Taken together, these two variables describe a discounting process that either leads or, at least, is associated with upgrading decisions.

Contextual factors also carry importance in predicting upgrading behavior. News about both brands (iPhone and Samsung) and about the brand of the smartphone the respondent owns influence the likelihood of upgrading. As time progresses, a reduction in exposure to news about smartphones seems to anticipate its upgrade. This reduction might signal a decision that had already been made and further search is not needed, or further evidence of motivated reasoning processes (Kunda, 1990; Reczek et al., 2018) that might prevent consumers from learning more about their chosen product.

Taken together, these variables comprise a parsimonious upgrading model that considers context, product characteristics, and their relationship with their users through ownership and enjoyment. As time passes and new options are available, consumers perceive changes in functionality (importance of resources, whether it meets one's needs, update capacity) that are associated with changes in value perception (mental book value and sentimental value), which, in turn, are associated with changes in the enjoyment derived and expected from the device. These factors, along with the disconnect between users and news about devices, explain upgrading behavior for this product category.

CONCLUSIONS, LIMITATIONS, AND FUTURE STUDIES

From a naive perspective, long lines of anxious shoppers in front of Apple stores every September for buying the new iPhone model seems a curious or even amusing mass behavior. From a consumer behavior perspective, however, it is an intriguing phenomenon. As for any human behavior, it is hard to explain replacement through just one psychological or contextual variable. Yet, most of the literature on replacement decision tries such a demarcation (Okada, 2001; Bellezza, Ackerman & Gino, 2017; Sela & LeBoeuf, 2017). Although valuable, this conceptualization oversimplifies upgrade behavior.

The present work enhances previous understanding by taking a holistic, integrative approach. The panel created for this manuscript provided a detailed, longitudinal account of consumers' characteristics, their perception about products, and contextual elements. Data collected every two months through a year allowed us to test a parsimonious Deep Learning model explaining upgrade behavior. The consequent short Deep Learning model presents high performance indices and evidences which variables are more relevant to explain and predict upgrade behavior. Besides the overall model performance, results also identify and rank the twelve most relevant variables, representing four groups explored in the model (enjoyment, ownership behavior, contextual variables, and perception about the smartphone). Together, they provided a ranked explanation of which variables are more relevant to determine who will and will not upgrade.

Contrary to previous works that have studied upgrade decision process based on longitudinal data (e.g., Bellezza, Ackerman & Gino, 2017; Gordon, 2009), we provide an empirical perspective that estimated the interactive and non-linear combined effects of the most relevant variables previously studied on the literature. Our perspective provides a new theoretical stance on how to understand upgrade behavior, taking into consideration the combined effects of those variables.

Although not central to its the intended contribution, this manuscript also presents a road map for researchers interested in using machine learning and artificial intelligence models to predict behavior or classify customers. These tools allow researchers to accommodate a myriad of distributions, beyond more traditional linear models. They also allow for model practice and validation within the same sample, increasing the external validity of the findings and the robustness of the measures.

This work, however, is not without limitations. Although negative consequences of attrition were mitigated through K-fold cross-validation, larger samples and the inclusion of other product categories would further validate results. Another limitation is the idiosyncratic structure of the market at research time. Although Apple and Samsung keep similar levels of market share since 2015 (Forbes, 2018), new competitors have gained market share, and iPhone and Samsung Galaxy models may have lost part of their dominance. Future research should extend the data collection for longer periods to avoid short-term market conditions, as well as include new smartphone models to reflect the actual market scenario.

REFERENCES

Allaire, J. J. & Chollet, F. (2018) *R Interface to 'Keras'*. Available at: https://keras.rstudio.com/index.html (Accessed: 16 January 2019).

AppleInsider (2018) *Analyst estimates average lifespan for all Apple devices at over four years, AppleInsider*. Available at: https://appleinsider.com/articles/18/03/01/analyst-estimates-average-lifespan-for-all-apple-devices-at-over-four-years (Accessed: 3 May 2018).

Bayus, B. L. (1991) 'The consumer durable replacement buyer', *The Journal of Marketing*, pp. 42–51.

Becker, D. (2018) *XGBoost | Kaggle*. Available at: https://www.kaggle.com/dansbecker/xgboost (Accessed: 16 January 2019).

Bellezza, S., Ackerman, J. M. & Gino, F. (2017) '"Be Careless with That!" Availability of Product Upgrades Increases Cavalier Behavior Toward Possessions', *Journal of Marketing Research*, 54(5), pp. 768–784.

Boujbel, L. & d'Astous, A. (2015) 'Exploring the feelings and thoughts that accompany the experience of consumption desires', *Psychology & Marketing*, 32(2), pp. 219–231.

Chollet, F. & Allaire, J. J. (2018) *Deep Learning with R*. Manning Publications Company.

Cripps, J. D. & Meyer, R. J. (1994) 'Heuristics and biases in timing the replacement of durable products', *Journal of Consumer Research*, 21(2), pp. 304–318.

Douglass, R. P., Duffy, R. D. & Autin, K. L. (2016) 'Living a calling, nationality, and life satisfaction: A moderated, multiple mediator model', *Journal of Career Assessment*, 24(2), pp. 253–269.

Fisher, A., Rudin, C. & Dominici, F. (2018) *'All Models are Wrong but many are Useful: Variable Importance for Black-Box, Proprietary, or Misspecified Prediction Models, using Model Class Reliance'*, arXiv:1801.01489 [stat]. Available at: http://arxiv.org/abs/1801.01489 (Accessed: 6 March 2019).

Forbes (2015) *Apple Surpassed Samsung As Global Phone Market Leader, Says Report, Forbes*. Available at: https://www.forbes.com/sites/amitchowdhry/2015/03/04/apple-passes-samsung/ (Accessed: 7 March 2019).

Forbes (2018) *Samsung Beats Apple In The Global Smartphone Market As Chinese Brands Close In, Forbes*. Available at: https://www.forbes.com/sites/panosmourdoukoutas/2018/09/13/samsung-beats-apple-in-the-global-smartphone-market-as-chinese-brands-close-in/ (Accessed: 8 March 2019).

Fortune (2018) *Apple iPhone Owners Will Buy 350M New Handsets in 2018, 2019* | Fortune. Available at: http://fortune.com/2018/02/27/apple-iphone-upgrades/ (Accessed: 3 May 2018).
Frederick, S. & Loewenstein, G. (1999) '16 Hedonic Adaptation', Well-Being. *The foundations of Hedonic Psychology*/Eds. D. Kahneman, E. Diener, N. Schwarz. NY: Russell Sage, pp. 302–329.
Galak, J. & Redden, J. P. (2018) 'The Properties and Antecedents of Hedonic Decline', *Annual Review of Psychology*, 69.
Goodfellow, I. et al. (2016) *Deep learning*. MIT press Cambridge.
Gordon, B. R. (2009) 'A dynamic model of consumer replacement cycles in the PC processor industry', *Marketing Science*, 28(5), pp. 846–867.
Grewal, R., Mehta, R. & Kardes, F. R. (2004) 'The timing of repeat purchases of consumer durable goods: The role of functional bases of consumer attitudes', *Journal of Marketing Research*, 41(1), pp. 101–115.
H2O (2019) *H2O Driverless AI, Open Source Leader in AI and ML*. Available at: https://www.h2o.ai/products/h2o-driverless-ai/ (Accessed: 28 February 2019).
He, T. (2018) *xgboost v0.71.2*, R Documentation. Available at: https://www.rdocumentation.org/packages/xgboost/versions/0.71.2 (Accessed: 15 January 2019).
Kim, H. S. et al. (2015) 'Do social casino gamers migrate to online gambling? An assessment of migration rate and potential predictors', *Journal of Gambling Studies*, 31(4), pp. 1819–1831.
Kuhn, M. (2018) *The caret Package*. Available at: http://topepo.github.io/caret/index.html (Accessed: 16 January 2019).
Kunda, Z. (1990) 'The case for motivated reasoning.', *Psychological Bulletin*, 108(3), p. 480.
Lenk, P. J. & Rao, A. G. (1990) 'New models from old: Forecasting product adoption by hierarchical Bayes procedures', *Marketing Science*, 9(1), pp. 42–53.
Mahajan, V., Muller, E. & Bass, F. M. (1990) 'New product diffusion models in marketing: A review and directions for research', *Journal of Marketing*, 54(1), pp. 1–26.
Mahajan, V., Muller, E. & Bass, F. M. (1995) 'Diffusion of new products: Empirical generalizations and managerial uses', *Marketing Science*, 14(3_supplement), pp. G79–G88.
MathWorks (2018) *Detector Performance Analysis Using ROC Curves - MATLAB & Simulink*. Available at: https://www.mathworks.com/help/phased/examples/detector-performance-analysis-using-roc-curves.html (Accessed: 15 January 2019).
Meyer, R. J., Zhao, S. & Han, J. K. (2008) 'Biases in valuation vs. usage of innovative product features', *Marketing Science*, 27(6), pp. 1083–1096.
Miller, C. J., Wiles, M. A. & Park, S. (2019) 'Trading on Up: An Examination of Factors Influencing the Degree of Upgrade: Evidence from Cash for Clunkers', *Journal of Marketing*, 83(1), pp. 151–172.
Molnar, C. (2019) *Interpretable Machine Learning*. Available at: https://christophm.github.io/interpretable-ml-book/ (Accessed: 26 February 2019).
Montaguti, E. & Zammit, A. (2017) 'Being the first entrant and getting stuck in the middle: The risks of becoming the intermediate pioneer', *European Journal of Marketing*, 51(7/8), pp. 1178–1196.
Munster, G. (2018) *iPhone Intent to Upgrade Survey Suggests More Predictable Demand, Loup Ventures*. Available at: https://loupventures.com/iphone-intent-to-upgrade-survey-suggests-more-predictable-demand/ (Accessed: 4 March 2019).
Murphy, K. P. & Bach, F. (2012) *Machine Learning: A Probabilistic Perspective*. MIT Press.

Okada, E. M. (2001) 'Trade-ins, mental accounting, and product replacement decisions', *Journal of Consumer Research*, 27(4), pp. 433–446.
Okada, E. M. (2006) 'Upgrades and new purchases', *Journal of Marketing*, 70(4), pp. 92–102.
Reczek, R. W. et al. (2018) 'That's Not How I Remember It: Willfully Ignorant Memory for Ethical Product Attribute Information', *Journal of Consumer Research*, 45(1), pp. 185–207.
Rogers, E. M. (1976) 'New product adoption and diffusion', *Journal of Consumer Research*, 2(4), pp. 290–301.
Sela, A. & LeBoeuf, R. A. (2017) 'Comparison neglect in upgrade decisions', *Journal of Marketing Research*, 54(4), pp. 556–571.
Shih, E. & Schau, H. J. (2011) 'To justify or not to justify: the role of anticipated regret on consumers' decisions to upgrade technological innovations', *Journal of Retailing*, 87(2), pp. 242–251.
Shung, K. P. (2018) *Accuracy, Precision, Recall or F1?, Towards Data Science*. Available at: https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9 (Accessed: 7 March 2019).
Trevor, H., Robert, T. & JH, F. (2009) *The elements of statistical learning: data mining, inference, and prediction*. New York, NY: Springer.
VentureBeat (2018) '*Newzoo: Smartphone users will top 3 billion in 2018, hit 3.8 billion by 2021'*, VentureBeat, 11 September. Available at: https://venturebeat.com/2018/09/11/newzoo-smartphone-users-will-top-3-billion-in-2018-hit-3-8-billion-by-2021/ (Accessed: 7 March 2019).
Wang, Y. & Roedder John, D. R. (2019) 'Up, Up, and Away: Upgrading as a Response to Dissimilar Brand Users', *Journal of Marketing Research*, 56(1), pp. 142–157.