

Research Article

An Integrative Model to Predict Product Replacement Using Deep Learning on Longitudinal Data

Vinicius Andrade Brei¹
Leonardo Nicolao¹
Maria Alice Pasdiora¹
Rodolfo Coral Azambuja¹

¹ Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil

Received 29 October 2019. This paper was with the authors for two revisions. Accepted 19 July 2020. First published online 6 August 2020.

Paulo Henrique Muller Prado was the associate editor for this article.

Editor-in-chief: Carlo Gabriel Porto Bellini

Editorial assistants: Kler Godoy and Simone Rafael

ABSTRACT

Past research on product upgrades has focused either on understanding who and when will upgrade or on figuring out why consumers will upgrade, but seldom on all. It has also neglected the interplay between these matters with decision context and timing. This manuscript depicts a comprehensive approach where, for the first time, product characteristics, individual differences, process, and contextual variables are analyzed on a predictive model of real product upgrades, identified through the systematic collection of primary data from a panel of smartphone consumers. We tested one traditional linear logistic regression model and two types of non-linear, state-of-the-art machine-learning models (extreme gradient boosting and deep learning) to explain upgrading behavior. Results provide an integrative, yet parsimonious, product-upgrade model showing the importance of resources; news about the smartphone brand; sentimental value; predicted, current, and remembered enjoyment; update capacity; and how much the smartphone meets the user's current needs as the most relevant variables to determine which consumers are more prone to upgrade their smartphones. Our findings advance upgrade decision theory by taking a holistic approach to the phenomenon and bridging different theoretical accounts of the replacement decision literature.

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

JEL Code: M310.

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 there is a new version of their product available, an improved 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 a classic sandwich cookie. Periodically, new options abound. The intention is not only to generate sales through the acquisition of new customers but also 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 (Reisinger, 2018). They are not necessarily replacing faulty devices but shortening the lifespan of their functioning phones (Owen, 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 otherwise would remain active and in good conditions to be used for a few more years. This manuscript is concerned with understanding why many consumers prematurely replace or upgrade their products, leaving years of functionality on the table and, consequently, overspending.

Adoption and diffusion of new products are widespread and established topics in marketing (Mahajan, Muller, & Bass, 1990, 1995; Montaguti & Zammit, 2017, among others). However, upgrading products, a related 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 upgrade to a newer iPhone model (Munster, 2018). Yet, despite its ubiquity, few studies have tackled this phenomenon (Bayus, 1991; Bellezza, Ackerman, & Gino, 2017; Okada, 2001, 2006), much less comprehensively.

For example, Bayus (1991) has focused on individual traits more closely associated with those who replace a product in the early and late stages of its life cycle. His research has identified different profiles for different replacement timings, but was limited as to why the replacement was made or which contextual elements would influence such trade. Grewal, Mehta, and Kardes (2004) have also investigated which attitudinal functions (i.e., individual characteristics) lead to replacement decisions. Although the effects they reported are conditioned to product characteristics (public or private, for example), they do not include context and process measures. Okada (2001, 2006), on the other hand, proposed 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) described an interesting phenomenon where replacement is driven by a consumer's carelessness with her current product, an attempt to justify the desire to upgrade. Both Okada's and Belezza et al.'s explanations are insightful but do not include contextual or individual differences.

Taken together, the extant literature on product upgrades focuses on which and when customers will upgrade (e.g., Bayus, 1991; Grewal, Mehta, & Kardes, 2004), why they will upgrade (e.g.,

Okada, 2001, 2006; Bellezza et al., 2017), but seldom on all. These are not the only questions being left on the proverbial table. To the best of our knowledge, no previous work has combined these variables with context or timing either. For example, there is little about the role of media buzz – a contextual variable – and the natural decline of the enjoyment of a product – a process measure – and their effects on desire to upgrade. The present manuscript reflects a comprehensive research approach where, for the first time, product characteristics, individual differences, process measures, and contextual variables are taken into consideration on a predictive model of real product upgrades, identified through the systematic collection of primary, longitudinal data with smartphone consumers.

Contextual variables and their interplay with individual differences are particularly missed from previous research on replacement decisions. Mirroring trends in social psychology defining behavior as a function of both persons and situations, the consumer behavior literature has been reiterating the relevance of contextual features and individual traits from its very beginning (e.g., Lewin, 1946) to now (e.g., Wu, Moore, & Fitzsimons, 2019). One very established example is the EKB decision-making model (Engel, Kollatt, & Blackwell, 1968), which integrates individual and environmental (contextual) sources of influence over behavior. Under this model, social context (manifested through social influence) would shape need recognition, how much and what kind of information would be searched, the criteria for alternative evaluation, and the choice itself, all varying significantly from individual to individual, according to demographic and psychographic characteristics (Engel et al., 1968).

Contextual information and the characteristics of those who are upgrading are fundamental to the understanding of product upgrades. Bayus (1991) identified, for instance, that early replacers of automobiles – those who replace their automobiles early in the product life cycle – tend to earn higher incomes and have lower levels of educational achievement and occupational status than late replacement buyers. Grewal et al. (2004), on the other hand, focused on the nature of the product and of the decision, which they characterized as contextual factors, to explain the repurchase of consumer durables. Although these approaches tend to be exclusive, focusing either on individual differences or on contextual factors, realistic explanations of a product upgrade or replacement require both sets of variables. Consumers, regardless of how much self-control they can exert and how materialistic they are (all examples of individual traits), might only be aware of the brands mentioned by the media (context). Consumers deciding to upgrade their phones will do so within a context of incessant iPhone reviews, Galaxy unboxing and overall chatter on Twitter about these products.

Upgrade decisions are complex and research, up to this point, has chosen to focus on specific variables, a necessary simplification to tackle specific questions. Consequently, there is not a single piece of research that considers individual, product, contextual, and process variables enough to comprehend and predict upgrading behavior. Given the unknowns of how, why, and when individuals upgrade their current products and the lack of studies tackling this phenomenon in its entirety, we propose and test a comprehensive, integrative model of product upgrade, including personal differences, product characteristics, context, and psychological processes. We applied one type of traditional binary classification model (logistic regression) and

two types of non-linear, state-of-the-art machine-learning models (extreme gradient boosting and deep learning neural network) on a longitudinal data set of iPhone and Samsung smartphone consumers with six rounds of data collection during one year (one wave of data collection every two months) to explain upgrading behavior. 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). Smartphones 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, we review pertinent literature on replacement decisions. In the section that follows, we explore the data set and describe the procedures to assemble the longitudinal consumer panel. Next, we detail data preparation, followed by model choice strategy and, ultimately, model selection, tuning, and performance assessment. Finally, we discuss results from the explanatory upgrading model, limitations, and future research suggestions.

REPLACEMENT DECISION THEORY

The long-term ownership of a durable often involves decisions about its replacement. Consumers replace durables for two main reasons. The first is the poor performance of the status quo (the currently owned product), which causes forced replacements. The second is the innovations and enhancements in a product category, which stimulate unforced replacements (Grewal et al., 2004). For the 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 this sense, product replacement is an operational synonym for product upgrade and we use those terms interchangeably, respecting the original employment of the expression by its authors. Replacement decisions present some unique characteristics that have been explored by the marketing literature.

For instance, the literature has explored who and when replace their products. Bayus (1991) developed a model incorporating demographic characteristics, attitudes, and search behavior for differentiating consumers who replace a product when it is fairly new or when it is old. His results demonstrated that early replacers, those replacing a relatively new product, 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. Moreover, while early replacers usually upgrade for changes in preference, late replacers upgrade because of performance reasons. Huh and Kim (2008) demonstrated that the intention to upgrade depends not only on the propensity to adopt innovations but also on the usage of innovative features. According to their results, cell phone consumers who more frequently use innovative features are more prone to upgrade.

The literature on replacement behavior also explores the determinants and consequences of the length of purchase intervals. Grewal et al. (2004) reported results of a survey encompassing eight

product categories (cameras, cars, vacuum cleaners, wristwatches, DVD players, refrigerators, televisions, and golf bags) and suggested that attitude functions (knowledge, value expressive, social adjustive, and utilitarian) help explain and predict interpurchase intervals. This effect is contingent on the product nature (along with 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 for personal computers (PC) that accounts for replacement decisions when consumers are uncertain about future price and quality. His findings demonstrated that the length of replacement cycles, which is determined by the quality and technical features of the status quo, is a useful dimension for segmenting consumers. Consequently, it is useful for predicting which consumers are more likely to replace in the future. Although Grewal et al. (2004) and Gordon (2009) explored the role of product characteristics on replacement intervals, they acknowledged that they are not sufficient conditions, relying on consumer perceptions of uncertainty (Gordon, 2009) and attitudes toward the product and toward upgrading them (Grewal et al., 2004) for a more predictive model.

Other works draw attention to the importance of ownership time. According to Cripps and Meyer (1994), 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. In this sense, replacing a product is more about how long it has been in use than its perceived performance. Assuming that the decision to replace is not a binary choice, Miller, Wiles, and Park (2019) investigated how trade-in (i.e., the status quo given as part of the payment) and the marginal costs-benefits of the new purchase influence the degree of the upgrade. According to their findings, trade-in ownership time and brand loyalty enhance the replacement likelihood, reiterating Cripps and Meyer's (1994) rationale of time over performance. Timing is not a sufficient variable, though. There are also considerations of how much value, monetary or otherwise, consumers can derive from their status quo products and replacements.

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) explained that mental costs depend on the product's mental book value, i.e., 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 demonstrated 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) complemented her previous findings by showing that psychological costs become a lesser impediment to upgrading when consumers perceive the new product as dissimilar to the status quo. This barrier reduction happens because dissimilarity turns the sunk cost in the existing product less salient. Recent research shows that consumers unconsciously manage the status quo's value write-off when they want to make a justifiable replacement decision. Bellezza et al. (2017) examined the potential for consumers being careless with current possessions in the presence of appealing product upgrades. They show that consumers accidentally damage a

product or run out of it quickly, allowing them 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 explained 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 (Cripps & Meyer, 1994). However, those upgrade decisions may be influenced by agents as subtle as consumers' desire to differentiate themselves from dissimilar users of the same brand (Wang & John, 2019), inaccurate predictions about the 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).

INTEGRATIVE MODEL FOR UPGRADE DECISIONS

Based on the literature, the most frequent determinants of replacement can be classified in five main groups: factors related to ownership of status quo (e.g., Cripps & Meyer, 1994; Huh & Kim, 2008), perception about the status quo (e.g., Meyer et al., 2008; Sela & LeBoeuf, 2017), context variables (e.g., Gordon, 2009; Miller, Wiles, & Park, 2019), individual traits (Wang & John, 2019), and demographic characteristics (Bayus, 1991). A thorough list of publications, along with the nature of their investigated variables and how they relate to this research, is shown in Table 1.

Research described in Table 1 focuses either on one group of variables or, at most, two, given that authors are often faced with trade-offs when choosing which variables to investigate. For instance, Bayus (1991) focuses on demographic characteristics to the detriment of process measures. Okada (2006) explained how companies can mitigate the psychological costs of upgrading, but did not consider the role of context or of product enjoyment. Others relied on the nature of the data to which they had access. For instance, Gordon (2009) constructed a model of consumer product replacement based on a sales data set for PC processors. Consequently, the questions he asked were necessarily limited to the variables at his disposal. These variables are the individual building blocks to our understanding and the consequent prediction of product upgrading behavior. However, purchasing, including upgrades, is a complex phenomenon and, as such, can benefit from the simultaneous consideration of a multitude of variables (MacInnis & Folkes, 2010). To the best of our knowledge, our work is the first to combine groups of variables from past research to provide a comprehensive, integrative model of the factors explaining upgrading decisions and to test it on a purposefully built longitudinal data set. Beyond the factors already examined by past research, we also investigated whether the enjoyment derived from the status quo and the desire for the upgrade affect the decision to replace.

Our motivation to explore the influence of enjoyment and desire on upgrading decisions comes from findings of Okada (2001) and Bellezza et al. (2017). Results from Okada (2001) highlighted

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 lead to a decreased hedonic response in the form of less desire and less ongoing enjoyment (Galak & Redden, 2018). Simply put, with repeated consumption comes a decrease in enjoyment for a product.

Although Bellezza et al. (2017) did not present any measure of desire, their work implies that the desire for the upgrade influences how consumers treat their status quo. For the authors, it is because consumers desire a new version of their current product that they become careless. Therefore, we believe desire is central to upgrading behavior and follow 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 the ownership of the status quo product, the 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 (see Figure 1). It enhances prior understanding by considering multiple constructs and showing which variables are more relevant to predict which consumers will replace. An overview of the relevant literature on upgrading and how this manuscript differs from it is shown in Table 1.

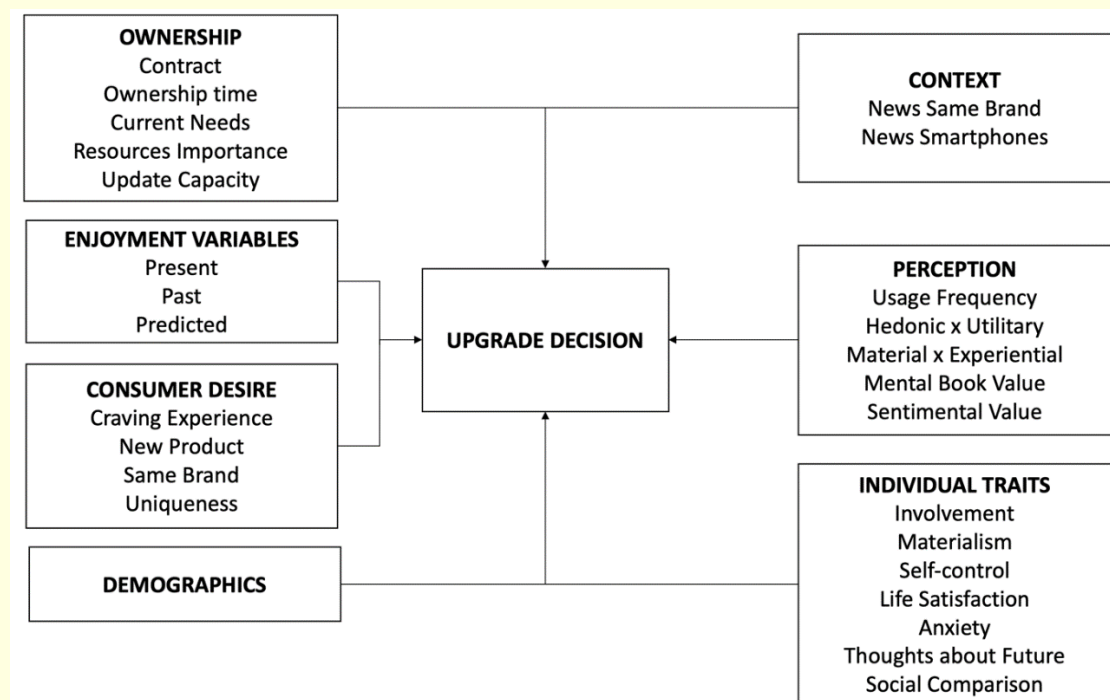


Figure 1. Integrative model for explaining upgrade decisions

Table 1.

Review of 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	-	-	-	-	-	Yes	-
Wang & John (2019)	Scenario	-	-	-	-	-	-	Yes
Sela & LeBoeuf (2017)	Scenario	-	-	-	-	Yes	-	-
Bellezza et al. (2017)	Real Transactions Scenario	Yes	-	-	-	-	-	-
Shih & Schau (2011)	Scenario	-	-	-	-	-	-	-
Gordon (2009)	Real Transactions	Yes	-	-	-	-	Yes	-
Meyer et al. (2008)	Scenario	-	-	-	-	Yes	Yes	-
Huh & Kim (2008)	Purchase Intentions	-	Yes	-	-	-	-	-
Okada (2006)	Scenario	-	-	-	-	Yes	Yes	-
Grewal et al. (2004)	Real Transactions	-	-	-	-	-	Yes	-
Okada (2001)	Scenario	-	-	Yes	-	-	-	-
Cripps & Meyer (1994)	Scenario	-	Yes	-	-	Yes	-	-
Bayus (1991)	Real Transactions	-	Yes	-	-	Yes	Yes	-

To understand what drives a consumer to upgrade her status quo product, we created and maintained a consumer panel measuring behavioral outcomes and psychological traits of its participants for one year. This panel provided answers to when, how, and why people decided to upgrade, capturing actual product upgrades. The next section describes the panel and the data collection procedures in detail.

LONGITUDINAL CONSUMER PANEL

We designed a longitudinal consumer panel to develop a deeper, temporal understanding of when, how, and why consumers replace their smartphones. Through the panel, we collected data from Apple and Samsung smartphone owners for one year. We have chosen smartphones for several reasons. First, they are largely consumed worldwide. In 2018, about 39% of the world population had a smartphone (Takahashi, 2018). Second, their market is competitive and dominated by two major global players, Apple and Samsung (Chowdhry, 2015). Third, both major players have been releasing upgrades for their best-selling smartphones systematically in September (Apple) and in April (Samsung), suggesting a regular upgrade cycle. Finally, these releases are accompanied by substantial marketing investments, media coverage, and social media buzz. Taken together, these reasons make up the foundation of a research environment that is richer and more insightful than that of the launch of other widespread durables (e.g., refrigerators and televisions).

Procedures

We collected data every two months through six survey rounds on Amazon Mechanical Turk (MTurk). The first round was in August 2015, the second in October 2015, the third in December 2015, the fourth in February 2016, the fifth in April 2016, and the last in June 2016. We chose MTurk because it allows for broad geographical reach and control, easy communication with participants (to avoid sample attrition), and fast data collection. Moreover, it is a platform with minimal non-response bias and consistent longitudinal responses, in terms of both demographics and personality traits (Daly & Natarajan, 2015). MTurk also allowed us to control many important variables, such as respondent IDs, to ensure the same individual was answering each of the survey rounds, if she changed the brand or model of her smartphone (through the smartphone's International Mobile Equipment Identity – IMEI), and even the consistency of the information provided about the smartphone and its use. This consistency was assessed through the repeated measure of some of the variables of interest. Others, such as time-invariant psychological traits, were distributed along the six rounds of surveys. To avoid possible confounds that could be associated with specific cultural traits and different market conditions, we restricted the sample to English-speaking participants that lived in the US.

Previous studies have also relied on longitudinal consumer panels to study either binary decisions like ours or continuous dependent variables. For example, Castaño, Suján, Kacker, and Suján (2008) developed a two-wave lab study to analyze student decisions to enroll in a virtual course. Ittersum and Feinberg (2010) carried out a two-year longitudinal study with 143 consumers to evaluate the probability of technology adoption. In a 10-month field study, Haisley and Lowenstein (2011) analyzed the impact of gifts on deposit balances and customer satisfaction. Hasegawa, Terui, and Allenby (2012) conducted a longitudinal lab experiment for seven weeks to evaluate purchase quantities. Luo, Ratchford, and Yang (2013) created an MTurk panel and collected consumer data every week, for six months, to evaluate allocation decisions among

leisure activities of 287 consumers. Finally, Botner, Mishra, and Mishra (2015) conducted a five-week lab study to examine whether people would donate to charity and how much.

In our longitudinal panel, we were especially concerned about sample size and data quality. Previous longitudinal studies have shown that the final sample was about 24% of the initial size for studies with two data collections in six months (Kim, Wohl, Salmon, Gupta, & Derevensky, 2015) and 21% with three rounds in the same period (Douglass, Duffy, & Autin, 2016). Thus, we departed from a sample of 730 MTurkers that owned either an iPhone or Samsung Galaxy models, the market leaders when the data collection started (Chowdhry, 2015). We have taken different measures to avoid sample attrition, the loss of respondents through the rounds of data collection. For example, we gradually increased the payment per round, which started at USD 0.30 per responded survey and ended at USD 2.00 plus a one-dollar bonus for those individuals who participated in all rounds.

Although Daly and Nataraajan (2015) have not directly tested the effects of different payment strategies to MTurk participants, they argued that escalating the payments may have the benefit of encouraging users to track the requester and look for opportunities to continue participating. To the best of our knowledge, no study has ever shown that increasing the payment would result in lower data quality, including papers that have successfully used similar procedures as ours. For example, Botner et al. (2015) offered payments beyond school credits in each round of data collection, and Ittersum and Feinberg (2010) paid a USD 20.00 bonus for individuals who participated in all rounds of data collection. Our final sample was of 144 participants, an attrition rate consistent with previous longitudinal data collection (e.g., Douglass et al., 2016; Kim et al., 2015), and a sample size that allows the robust estimation of machine learning methods (e.g., Climent, Momparler, & Carmona, 2019).

Measurements

Aiming at similar answering times for all rounds, we distributed time-invariant psychological variables along the six rounds. This kept the duration of each survey similar, at around three minutes. Each round had four sets of variables: the dependent variable, contextual variables, psychological variables, and demographic profile measures. The dependent variable of the study was how many times each participant upgraded her smartphone. We asked the IMEI (unique identity) of the smartphone and other questions related to the use of the smartphone in every round. For example, for how long had the person owned the product, for how long had she desired a new smartphone, etc. A summary of the independent variables, their operationalization, and their correlation with the dependent variable is presented in Table 2. The full correlation table with all the research variables is shown in the Appendix (see Table 5).

Table 2.

Independent variables

Ownership Variables	Description	Operationalization	Measurement Time (Rounds)	Correlation with Upgrade
Contract	Whether consumers signed up for a contract when they purchased the smartphone	Dummy coded variable (yes or no)	T ₁ — T ₆	0.022
Ownership time — T ₁	Number of months consumers owned their smartphone before exchanging	Number of months before acquiring the last smartphone	T ₁ — T ₆	0.104
Status quo meets current needs	How much the smartphone meets consumers' current needs	Rating in T ₁ minus rating in the moment consumers indicated the exchange (scale: 0–100)	T ₁ — T ₆	0.386*
Importance of resources	Importance assigned by consumers to resources of the smartphone	Rating in T ₁ minus rating in the moment consumers indicated the exchange (scale: 1–7)	T ₁ — T ₆	0.374*
Update capacity	Update capacity of the apps installed in consumers' smartphone	Rating in T ₁ minus rating in the moment consumers indicated the exchange (scale: 1–7)	T ₁ — T ₆	0.340*
Enjoyment Variables	Description	Operationalization	Measurement Time (Rounds)	Correlation with Upgrade
Enjoyment	How much consumers enjoy their smartphone	Rating in T ₁ minus rating in the moment consumers indicated the exchange (scale: 0–100)	T ₁ — T ₆	0.316*
Enjoyment — purchase time	How much consumers enjoyed their smartphone in the day of purchasing	Rating in T ₁ minus rating in the moment consumers indicated the exchange (scale: 0–100)	T ₁ — T ₆	0.370*
Enjoyment — 1 month ago	How much consumers enjoyed their smartphone one month ago	Rating in T ₁ minus rating in the moment consumers indicated the exchange (scale: 0–100)	T ₁ — T ₆	0.206
Enjoyment — 1 month ahead	How much consumers believe they will enjoy their smartphone in one month	Rating in T ₁ minus rating in the moment consumers indicated the exchange (scale: 0–100)	T ₁ — T ₆	0.436*
Enjoyment — 1 year ahead	How much consumers believe they will enjoy their smartphone in one year	Rating in T ₁ minus rating in the moment consumers indicated the exchange (scale: 0–100)	T ₁ — T ₆	0.311*
Desire Variables	Description	Operationalization	Measurement Time (Rounds)	Correlation with Upgrade
Craving experience	Craving Experience Questionnaire (May et al., 2014)	Rating in T ₂ minus rating in the moment consumers indicated the exchange (scale: 0–100)	T ₂ — T ₆	-0.011
Desire — feelings and thoughts	Feelings and thoughts related to consumer desire (Boujbel & d'Astous, 2015)	Index created from factorial loads of the first PCA component (scale: 0–100)	T ₃	0.196
Desire — new smartphone	How much consumers desire to exchange their smartphone	Rating in T ₁ minus rating in the moment consumers indicated the exchange (scale: 0–100)	T ₁ — T ₆	0.075
Desire — new smartphone of same brand	How much consumers desire to exchange their smartphone for a newer model of the same brand	Rating in T ₁ minus rating in the moment consumers indicated the exchange (scale: 0–100)	T ₁ — T ₆	0.125
Desire — unique products	Desire for Unique Products Scale (Lynn & Harris, 1997)	Index created from factorial loads of the first PCA component (scale: 0–100)	T ₆	0.062

Continues

Table 2 (continued)

Perceptions about the Smartphone	Description	Operationalization	Measurement Time (Rounds)	Correlation with Upgrade
Usage frequency	How much consumers currently use their smartphone	Scale: 0 — Extremely low; 100 — Extremely high	T ₆	0.108
Usage frequency — 1 month ago	How much consumers used their smartphone one month ago	Scale: 0 — Extremely low; 100 — Extremely high	T ₆	0.084
Usage frequency — purchase time	How much consumers used their smartphone in the purchase day	Scale: 0 — Extremely low; 100 — Extremely high	T ₆	0.121
Hedonic x Utilitary	How consumers classify their smartphones in a continuum between hedonic and utilitarian	Rating in T ₁ minus rating in the moment consumers indicated the exchange (scale: 1–7)	T ₁ — T ₆	0.249
Material x Experiential	How consumers classify their smartphones in a continuum between experiential and material	Rating in T ₁ minus rating in the moment consumers indicated the exchange (scale: 1–7)	T ₁ — T ₆	0.239
Mental book value	Feeling of having gotten money's worth from the smartphone (Okada, 2001)	Rating in T ₁ minus rating in the moment consumers indicated the exchange (scale: 1–7)	T ₁ — T ₆	0.722*
Sentimental value	How much sentimental value consumers assign to the smartphone	Rating in T ₁ minus rating in the moment consumers indicated the exchange (scale: 1–7)	T ₁ — T ₆	0.294*
Context Variables	Description	Operationalization	Measurement Time (Rounds)	Correlation with Upgrade
News about Galaxy	Consumers' degree of contact with news about Galaxy in the last month	Rating in T ₁ minus rating in the moment consumers indicated the exchange (scale: 1–7)	T ₁ — T ₆	0.286*
News about iPhone	Consumers' degree of contact with news about iPhone in the last month	Rating in T ₁ minus rating in the moment consumers indicated the exchange (scale: 1–7)	T ₁ — T ₆	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	Rating in T ₁ minus rating in the moment consumers indicated the exchange (scale: 1–7)	T ₁ — T ₆	0.347*
Individual Traits	Description	Operationalization	Measurement Time (Rounds)	Correlation with Upgrade
Involvement	Involvement Scale (Zaichowsky, 1994)	Index created from factorial loads of the first principal component analysis (PCA) component (scale: 1–7)	T ₂	0.091
Materialism	Materialism Scale (Richins, 2004)	Index created from factorial loads of the first PCA component (scale: 1–7)	T ₂	0.100
Self-control	Self-Control Scale (Tangney, Baumeister, & Boone, 2004)	Index created from factorial loads of the first PCA component (scale: 1–5)	T ₄	0.039
Life satisfaction	Life Satisfaction Scale (Pavot & Diener, 1993)	Index created from factorial loads of the first PCA component (scale: 1–7)	T ₄	-0.043
Anxiety	Anxiety Scale (Lau-Gesk & Meyers-Levy, 2009)	Index created from factorial loads of the first PCA component (scale: 1–7)	T ₅	0.011

Continues

Table 2 (continued)

Thoughts about future	How much consumers think about future incomes	Index created from factorial loads of the first PCA component (scale: 1–7)	T ₅	0.194
Social comparison	Social Comparison Scale (Gibbons & Buunk, 1999)	Index created from factorial loads of the first PCA component (scale: 1–7)	T ₆	0.051
Demographic Variables	Description	Operationalization	Measurement Time (Rounds)	Correlation with Upgrade
Gender — Female	Gender — only female	Dummy coded variable indicating female or not	T ₁	0.068
Gender — Male	Gender — only male	Dummy coded variable indicating male or not	T ₁	-0.068
Age	Age in years	Age indicated in T ₁	T ₁	0.035
Education	Education level	Education level indicated in T ₁	T ₁	-0.008
Annual household income — T ₁	Annual household income	Annual household income indicated in T ₁	T ₁	0.052
Annual household income — T ₂	Annual household income	Annual household income indicated in T ₂	T ₂	0.137
Discretionary income	Percentage of household income considered as discretionary	0–100%	T ₁	0.051

Note. *Correlation significant at $p < 0.05$.

DATA PREPARATION AND ANALYTICAL STRATEGY

We divided the analytical strategy into four stages: (1) data preparation; (2) choice of estimation models and algorithms; (3) simulation of different regularization, cross-validation, and tuning parameters; and (4) model performance comparison. All analyses were intended to determine which variables explain upgrade decisions. In other words, we strived to build the best classification model to predict who has and who has not upgraded their smartphone.

Data preparation

Analysis required a substantial amount of data preparation. Such preparation involved the selection, preprocessing, and transformation of the data, which was composed of a complex set of 39 different variables and constructs, including 33 behavioral 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 various procedures for data preparation.

The first step of data preparation was data selection, which involved the decision of which participants to consider for the final sample. As previously described in this and in previous research (e.g., Douglass et al., 2016; Kim et al., 2015), sample attrition is a considerable problem in long-lasting longitudinal research, such as our one-year panel. As expected, there was a substantial amount of missing values. Although we considered imputing data for the missing

variables, the attrition from Round 1 (730 respondents) to Round 6 (144 respondents) drove us to decide to analyze only the full data set, composed of the 144 individuals who participated in all the six rounds.

In the data processing and transformation stage, we took into account the different requirements for the classification models and created two other data sets, one with all variables standardized (z-scores) and another with all variables normalized (from zero to one). For each of the estimated classification models, we tested its performance using the original and the two transformed data sets. This conservative approach was chosen to deal with the requirements and limits of each type of model. While linear models, such as logistic regression, and nonlinear models, such as decision trees, do not necessarily require scaling or normalizing inputs, other models require such transformations. This is especially the case for deep learning models that cannot operate on label data directly. They require all input and output variables to be numeric (Chollet & Allaire, 2018). For these models, we also converted categorical and nominal data (e.g., gender) to numerical data using the one-hot encoding strategy. This encoding strategy is used when a natural ordering between categories may result in poor performance or unexpected results (for example, predictions halfway between categories). That is, we have created an additional binary column (0 or 1) for each categorical or nominal category class (see DelSole, 2018 for more details).

We operationalized longitudinal variables through the difference between the rating in the first round (T_1) and the rating in the round of the upgrade (T_{upgrade}), as described in Table 2. For participants who did not upgrade, we calculated the difference between T_1 and T_6 , the last round of data collection. We have taken this approach because we were interested in the magnitude of behavioral change during the panel and not in the serial correlation of each variable. Further, we summarized multi-item scales through factorial loads of the first principal component analysis (PCA) component. Because we used established scales for our constructs, we did not perform this PCA with the goal of abbreviating these scales through factor or item reduction. Rather, this analysis summarized the whole scale on a single composite measure, allowing non-additive effects to be expressed (DeVellis, 2016). For instance, Boujbel and d'Astous' (2015) desire scale contains items that, if summed, could cancel each other out (e.g., the great pleasure a consumer experiences could offset the simultaneous low discomfort). Finally, we estimated the models using all three types of variables (original, normalized, and standardized), including the covariates. However, because of the requirements of deep learning models – and following the recommendations of Chollet and Allaire (2018) –, we are reporting only models estimated with standardized z-scores.

Model choice

We tested the most traditional type of linear model for binary classifications (logistic regression, hereafter LR) and two types of non-linear, state-of-the-art supervised machine learning (ML) models: a decision-tree (eXtreme gradient boosting, hereafter XGB) and a deep learning neural network model (hereafter DL), to date the most advanced algorithms for classification (Becker, 2018; Chollet & Allaire, 2018). As we will show later in the Model Performance section (see Table 3 for performance comparison), the ML, non-linear models have superior prediction

performance compared to the traditional linear classification models (i.e., LR) for a number of reasons.

First, it is very likely that our data is non-linear and that multiple interactions among the variables in the dataset would also be non-linear. Consequently, linear models would not be able to capture the richness of the investigated phenomena. Second, DL completely automates the feature engineering process (Chollet, 2017). That is, it can transform the data into refined representations compared to the original variables. Third, traditional classification models, such as LR, 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 contrast, split the data into training and test sets, with the model being estimated in a set and evaluated in another one. In this sense, the prediction ability of the model is tested with data that was not seen in the estimation step (Goodfellow, Bengio, & Courville, 2016), which ultimately increases the accuracy of the predictions. Finally, we have estimated the ML non-linear models because they provide not only performance indices, but also ways of identifying which variables are more relevant for prediction. Previous research endorses the efficacy of these variable selection procedures. For instance, similar techniques were employed for classification problems concerning the prediction of bank failures (Climent et al., 2019), categorization of online content (Salminen, Yoganathan, Corporan, Jansen, & Jung, 2019), prediction of consumer purchase intention (Bag, Tiwari, & Chan, 2019), and prediction of the helpfulness of online reviews (Singh, Irani, Rana, Dwivedi, Saumya, & Roy, 2017).

We estimated the XGB model using the XGB package (He, 2018) and the DL model using the Keras package (Allaire & Chollet, 2018), both in their versions for R. The code for reproducing our analysis is available upon request.

Regularization, cross-validation, and model tuning

Regularization. Prediction accuracy is compromised when the number of samples (n) is similar or inferior to the number of variables (p). By constraining or shrinking the estimated coefficients, we tried to reduce the variance of the models at the cost of a negligible increase in bias. Such a strategy leads to improvements in accuracy (James, Witten, Hastie, & Tibshirani, 2013). We were also concerned with *model interpretability*, that is, with avoiding estimating a model where some or many of the variables were not associated with the upgrade decision. Such inclusion of unnecessary variables increases model complexity and, consequently, reduces model interpretability. Considering the large number of variables in our data set and the sample size, we were especially attentive to fitting procedures that could yield better prediction *accuracy* and *model interpretability*: subset selection and regularization (or shrinkage).

In the subset selection, we tested a stepwise procedure, sequentially adding into the estimated models the predictors that most improved the classification accuracy. Next, we tried the backward selection, starting with the full model, and sequentially excluding the predictors that had the least impact on accuracy. However, because subsetting is a discrete process, that is, variables are either retained or discarded, it results in high variance, which does not reduce the prediction error of

the full models. Considering that shrinkage methods are more continuous and do not suffer as much from high variability, we preferred the more modern regularization (i.e., shrinkage) approaches (Hastie, Tibshirani, & Friedman, 2009).

In all models, we tested both Ridge and Lasso shrinkage methods, by imposing a penalty on the coefficient size. This is important when there might be many correlated variables, which is the case for our data, because a large positive coefficient on one variable can be canceled by a similarly large negative coefficient on another correlated variable. By imposing a size constraint on the coefficients, this problem is reduced, improving model performance (Hastie et al., 2009). With such a strategy, we tried to shrink the coefficients toward zero. Similarly, with 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. However, Ridge does have one obvious disadvantage. Contrary to the previous subset strategy, it maintains all p predictors in the final model. Considering that the number of predictors in our data set is large, we also tested a Lasso approach, shrinking the coefficient estimates toward zero, therefore reducing the number of variables in the model, resulting in a more parsimonious set of variables in each model (James et al., 2013).

Cross-validation. When estimating the ML models, the difference between the test error rate and the training error rate could also impose a problem. Since the data set 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 (Hastie et al, 2009) through the caret R package (Kuhn, 2019). The data was divided into $K = 10$ roughly equal parts. For each k th part, the model was fit to the other $K - 1$ parts of the data set and calculated the prediction error of the fitted model when predicting the k th part of the data. This was done for $k = 1, 2, \dots, K$, and then combined the K estimates of prediction error. More formally, the cross-validation estimate of prediction error was calculated as follows.

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

where $k: \{1, \dots, N\} \mapsto \{1, \dots, K\}$ is an indexing function that indicates the partition to which observation i is allocated by the randomization; and $\hat{f}^{-k}(x)$ is the fitted function, computed with the k th fold removed. At the end of the K-fold cross-validation process, each model was estimated in 10 randomly selected training datasets, and tested in another 10 randomly selected test sets. It is also important to note that the split of a dataset between test and train sets consists of picking instances randomly and setting them aside (Géron, 2019). Following this procedure, for each of the folds we assigned 30% of the dataset to the test set and 70% to the train set.

Cross-validation provides a more precise estimation of the true prediction value through the assessment of the test error rate. Instead of using the mean square error measure (as in traditional regression models), we have calculated the average number of misclassified observations, the recommended approach for classification problems (James et al., 2013). The advantage of cross-validation is that the predicted (test error) value is not the result of just one random sample of the data, but of multiple systematic and ordered random samples of the same data (for more

details about cross-validation methods, see Browne, 2000). In short, each time one of the models was estimated using a cross-validation strategy, a prediction score was generated, as shown in Table 3.

Model tuning. The last step of the model fitting process was to tune the estimated parameters of the XGB and DL models. We tested different hyperparameters to optimize model architecture, a necessary step to improve model accuracy and prediction capacity. To discover the most efficient hyperparameters, we applied an automated machine learning strategy using the H2O platform in R (H2O, 2019). The best hyperparameters were then carried out to the XGB and DL models. As shown in Table 3, the performance of the DL models was better than that of the XGB model, with the short-DL model showing better performance than the full-DL model. The short-DL model had the following architecture and hyperparameters: Keras sequential model; one input layer; one first hidden layer (units = 50, kernel_regularizer = regularizer_l1_l2 [l1 = 0.001, l2 = 0.001], activation = rectified linear unit [ReLU]), with a dropout hyperparameter to prevent overfitting (layer_dropout [rate = 0.1]); a second densely connected hidden layer (units = 50, kernel_regularizer = regularizer_l1_l2 [l1 = 0.001, l2 = 0.001], activation = rectified linear unit [ReLU]), with another dropout hyperparameter to prevent overfitting (layer_dropout [rate = 0.1]); followed by a densely connected binary (possible outputs: 1 – Yes or 0 – No) output layer (units = 1, kernel_regularizer = regularizer_l1_l2 [l1 = 0.001, l2 = 0.001], activation = “sigmoid”). Finally, the model compilation hyperparameter was Adam optimization algorithm, and the loss function a binary cross-entropy loss (log loss). In total, the DL model estimated a total of 4,701 parameters.

Model performance

We evaluated the linear and the non-linear classification models by the proportion of events they correctly predicted, i.e., how many upgrades and not upgrades the model accurately classified. We compared the LR, XGB, and DL models using three performance parameters: Accuracy, Area Under the Curve (AUC), and F_1 score (Murphy, 2012) (see Table 3 for a summary of the performance parameters and the Appendix [Figures 5 to 9] for the receiver operating characteristic [ROC] curves of the estimated models). For performance comparison purposes, we tested the best linear LR models and the best non-linear ML models (the DL model) in a full and a short version. We did not test the XGB in a full and a short version because its performance was inferior to that of the DL models. The full versions were estimated including all variables and the short versions including only the variables that presented a statistically significant correlation with the Upgrade dependent variable (as shown in Table 2).

To calculate the performance indices, we first 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 estimate the Accuracy (calculated as the number of corrected predictions divided by the total number of predictions) and the generation of the ROC curves, which evaluate the capacity of the models as binary classifier systems. We created the ROC curves by plotting the true positive rate (TPR) against the false positive rate (FPR) at a variety of threshold settings. The ROC curves allow the comparison of 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 the F_1 score (see Equation 2), which is the harmonic mean of precision and recall. P is *precision*, defined as $TP/\widehat{N}_+ = p(y = 1 | \hat{y} = 1)$, and the R is *recall*, defined as $TP/\widehat{N}_+ = p(\hat{y} = 1 | y = 1)$. While precision measures the proportion of the detected instances that are actually positive, recall measures the proportion of the positive instances that were actually detected (Murphy, 2012). F_1 is a relevant metric for different reasons. For one, in a fixed threshold, one can compute a single precision and recall value. Also, it is a widely used metric to measure the success of a binary classifier when one class is rare (Lipton, Elkan, & Narayanaswamy, 2014).

$$F_1 \triangleq \frac{2}{\frac{1}{P} + \frac{1}{R}} = \frac{2PR}{R+P} \quad (2)$$

This uneven class distribution is frequent in upgrade decisions for consumer goods, such as smartphones. Considering the one-year duration of our longitudinal panel, and the average product life cycle of a smartphone, it would be very unlikely that a substantial part of the sample would upgrade their smartphones in such a short period.

Table 3.

Models performance comparison

	LR Full	LR Short	XGBoost	DL Full	DL Short
Accuracy	.841	.909	.932	.955	.977
AUC	.728	.801	.874	.979	.990
F_1	.526	.696	.727	.833	.889

Compared to the linear LR Full (LR Accuracy = 0.841, AUC = 0.728, F_1 = 0.526), the linear LR Short (LR Accuracy = 0.909, AUC = 0.801, F_1 = 0.696), and the XGB (Accuracy = 0.932, AUC = 0.874, F_1 = 0.727), the performance of the deep learning models was superior (DL full Accuracy = 0.955, AUC = 0.979, F_1 = 0.833; DL short Accuracy = 0.977, AUC = 0.990, F_1 = 0.889). Given the superiority of their performance and the space constraints of this manuscript, we will focus on the discussion of the non-linear ML models results. We will pay close attention to the results of the DL short model, due to its overall best performance and parsimony.

Table 4.

Confusion tables

	LR Full		LR Short		XGB		DL Full		DL Short		
	Truth										
		No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Prediction	No	35	3	34	3	37	1	37	0	39	1
	Yes	1	5	2	5	2	4	2	5	0	4

Due to the random nature of the selection process, we did not have any influence on which instances were assigned to the train and the test sets. Table 4 presents the results for five different machine-learning models. Each of them performed its own random split between test and train sets. As a consequence, there might be differences in the proportion of positive/negative instances.

RESULTS

All three ML models achieved high levels of precision at estimating both upgraders and non-upgraders, despite the unbalanced distribution of the two categories. Even considering the small difference in performance among the decision-tree XGB and the two DL neural networks models, accuracy was higher for the DL short model (Accuracy = 0.977) when compared to the DL full model (Accuracy = 0.955) and the XGB (Accuracy = 0.932). The quality of the ROC, summarized by the area under the curve (AUC), followed the same pattern as the Accuracy (see Figures 5 to 9 in the Appendix). The F_1 score, a better measure when one of the predicted classes is rare (Lipton et al., 2014), is also higher for the DL short model, when compared to the XGB and the DL full model. Considering the evident predominance of the DL short 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), only the results of the DL short model are discussed in the following paragraphs.

To evaluate the contribution of each variable to the predictive power of the model, the model-agnostic interpretation method of *Feature Importance* was applied. We measured the importance of each variable (or feature) by calculating the increase in the model's prediction error after permuting that 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). We implemented the *Feature Importance* calculation through the following algorithm (Fisher, Rudin, & Dominici, 2019) (Figure 2).

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 $j = 1, \dots, p$ do:
 - Generate feature matrix X^{perm} by permuting feature j in the data X . This breaks the association between feature j and true outcome y .
 - Estimate error $e^{perm} = L(y, f(X^{perm}))$ based on the predictions of the permuted data.
 - Calculate permutation feature importance $FI^j = \frac{e^{perm}}{e^{orig}}$. Alternatively, the difference can be used: $FI^j = e^{perm} - e^{orig}$
3. Sort features by descending FI

Figure 2. Feature Importance algorithm.

We calculated the *Feature Importance* algorithm. Figure 3 presents its results. Figure 4 presents the difference between the F_1 score of the DL short model when all variables are included and the F_1 score when the variables were removed from the model. From the set of variables correlated with the upgrade decision (the DL short model), the most relevant variable to predict the upgrade decision was *resources importance* and the least important was *how much the current smartphone meets one's needs today*.

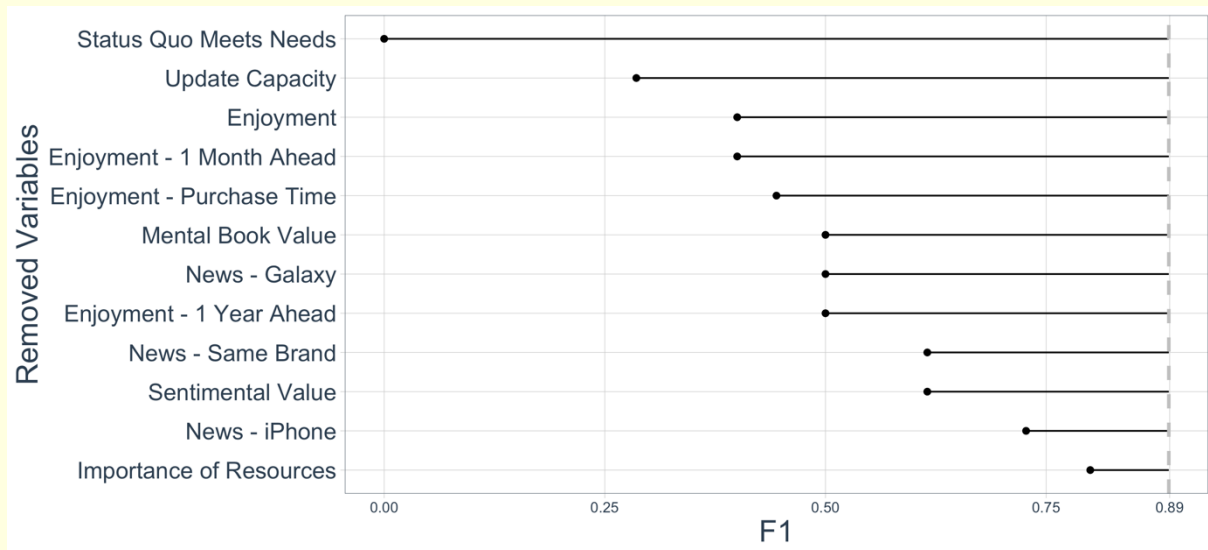


Figure 3. Variable importance based on F_1 scores — short deep learning model

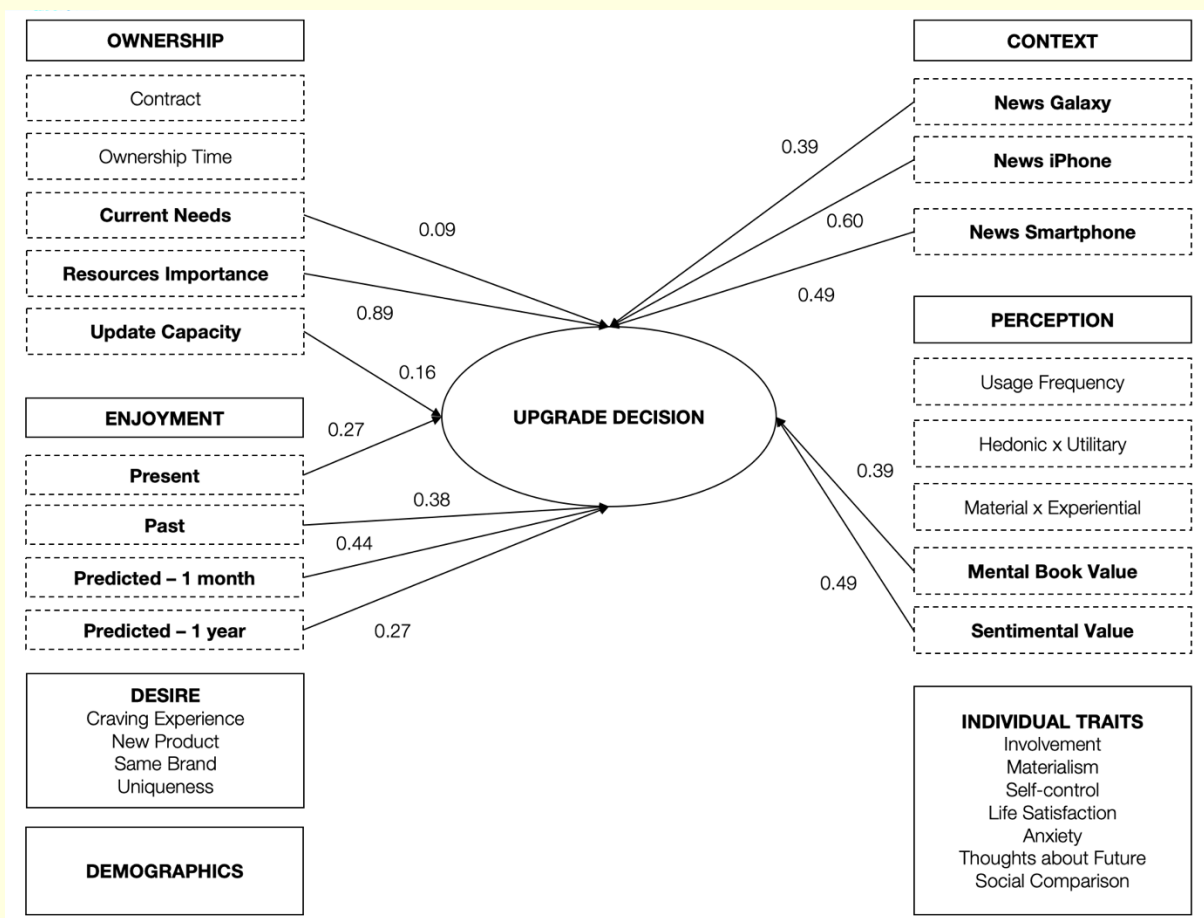


Figure 4. Most relevant variables for explaining upgrade decisions.

Numerical values correspond to the difference between the F_1 score of the short deep learning model and the model's F_1 score when the variable is removed.

DISCUSSION

Considering the sets of variables presented in Table 2 and on Figure 1 introduced earlier in the document, *enjoyment*, *ownership*, *context*, and *perception about the smartphone* were the more relevant predictors of the upgrade decision, in this order of relevance. Throughout this discussion, we will take into account both the contribution of each variable, manifested through their F_1 scores, and their correlation to the upgrade decision.

The most important variable 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 T_1 and T_6), 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 like 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 performance perceptions for upgrade decisions (Bayus, 1991; Gordon, 2009), but also contribute to this research stream. For example, Bayus (1991) differentiated early and late replacers, giving weight for product lifetime cycle in his conceptualization, while Gordon (2009) emphasized the relevance of the length of replacement cycles for segmenting consumers and for predicting which consumers are more likely to replace. In both cases, the authors took into account the resources of the status quo product and its replacement. However, they did so statically, where the *importance of product resources* is measured cross-sectionally and, consequently, ignored the dynamics of product usage and consumption and the reconsideration of product attributes that follows. For example, a consumer buys a smartphone for its long-lasting battery, revealing that this is an important resource to be considered. Over the course of usage, however, she might realize that it was not so important after all, leading her to realize other resources from competitive options are more prominent. This dynamic in the importance of resources is driving upgrading decisions. Therefore, we advance the previous theory showing the impact of the variation in the importance of smartphone resources along ownership time, not just of specific product attributes. In doing so, our results demonstrate how the fluctuation in 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*, 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 of the *enjoyment* measures, the score at time six (or at the time before the upgrade, for those participants that bought new smartphones during the panel duration) was subtracted from the first measure ($T_1 - T_{\text{upgrade}}$). 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, and of enjoyment a month and a year into the future were all relevant predictors of the upgrade decision, as evidenced by the F_1 scores.

As *current enjoyment* and *prospective enjoyment* decreased over time, individuals in the sample were more likely to upgrade to a new phone. Interestingly, this effect was also evident for *enjoyment at purchase time*. As individuals in the sample recalled their *enjoyment at the time of purchase* as being reduced over time, their likelihood of upgrade increased. There might be motivated reasoning effects at play (Kunda, 1990; Reczek, Irwin, Zane, & Ehrich, 2018), causing individuals to misremember past enjoyment to justify the upgrading behavior, along the lines of those reported

by Bellezza et al. (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, Irwin, Zane, and Ehrich (2018), whereby consumers present better memory for an ethical attribute when a product performs well on that attribute versus when a product performs poorly on that same 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. With this in mind, the desire to upgrade their current smartphones might be leading individuals in our sample to misremember the enjoyment at the time of purchase.

The relationship between a consumer and her smartphone, described so far 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 our 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* declines, upgrading likelihood increases. Taken together, these two variables describe a discounting process that either leads to 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 that further search is not needed. It can also signal 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 theoretical 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 disconnection 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 to buy the new iPhone model seems to be 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 decisions tries such a demarcation (Bellezza et al., 2017; Okada, 2001; Sela & LeBoeuf, 2017). However valuable, this conceptualization oversimplifies upgrading behavior.

The present work enhances previous understanding by taking a holistic, integrative theoretical and empirical approach. The panel created for this manuscript provided a detailed, longitudinal account of consumer characteristics, of their perception about products, and of elements of the context. Data collected every two months for a year allowed us to test a parsimonious deep learning model explaining upgrading behavior. The resulting short deep learning model presents high performance indices and evidences which variables are more relevant to explain and predict upgrading behavior. Apart from 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 who will not upgrade.

Contrary to previous works that have studied the upgrade decision process based on longitudinal data (e.g., Bellezza et al., 2017; Gordon, 2009) and either focused on a single variable or relied on linear effects, 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 upgrading behavior, taking into consideration the combined effects of those variables.

Although not central to its 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 usual and 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 and past research obtained robust results by applying similar machine learning methods to equivalent sample sizes (e.g., Climent et al., 2019), larger samples and the inclusion of other product categories would further validate our 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 (Mourdoukoutas, 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 current market scenario.

Acknowledgments

The authors acknowledge funding from the National Council for Scientific and Technological Development (Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPQ) and from the Chair Tramontina Eletrik. The authors also acknowledge that this article is an improved version of the paper awarded Best Paper at the 43rd Annual Meeting of the Brazilian Academy of Management (ANPAD), in São Paulo, Brazil, October 2-5, 2019.

REFERENCES

- Allaire, J. J. & Chollet, F. (2018). R Interface to Keras. Retrieved from <https://keras.rstudio.com/index.html>
- Bag, S., Tiwari, M. K., & Chan, F. T. S. (2019). Predicting the consumer's purchase intention of durable goods: An attribute-level analysis. *Journal of Business Research*, 94, 408-419. <https://doi.org/10.1016/j.jbusres.2017.11.031>
- Bayus, B. L. (1991) The consumer durable replacement buyer. *The Journal of Marketing*, 55(1), 42-51. <https://doi.org/10.1177%2F002224299105500104>
- Becker, D. (2018) XGBoost. Retrieved from <https://www.kaggle.com/dansbecker/xgboost>
- 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), 768-784. <https://doi.org/10.1509%2Fjmr.15.0131>
- Botner, K. A., Mishra, A., & Mishra, H. (2015). What's in a message? The longitudinal influence of a supportive versus combative orientation on the performance of nonprofits. *Journal of Marketing Research*, 52(1), 39-55. <https://doi.org/10.1509%2Fjmr.13.0211>
- Boujbel, L. & d'Astous, A. (2015). Exploring the feelings and thoughts that accompany the experience of consumption desires. *Psychology & Marketing*, 32(2), 219-231. <https://doi.org/10.1002/mar.20774>
- Browne, M. W. (2000). Cross-validation methods. *Journal of Mathematical Psychology*, 44(1), 108-132. <https://doi.org/10.1006/jmps.1999.1279>
- Castañó, R., Suján, M., Kacker, M., & Suján, H. (2008). Managing consumer uncertainty in the adoption of new products: Temporal distance and mental simulation. *Journal of Marketing Research*, 45(3), 320-336. <https://doi.org/10.1509%2Fjmk.45.3.320>
- Chollet, F. (2017). *Deep learning with python*. Greenwich, CT: Manning Publications CO.
- Chollet, F. & Allaire, J. J. (2018). *Deep learning with R*. Shelter Island, NY: Manning Publications
- Chowdhry, A. (2015, May). Apple surpassed Samsung as global phone market leader, says report. *Forbes*. Retrieved from <https://www.forbes.com/sites/amitchowdhry/2015/03/04/apple-passes-samsung/>
- Climent, F., Momparler, A., & Carmona, P. (2019). Anticipating bank distress in the Eurozone: An extreme gradient boosting approach. *Journal of Business Research*, 101, 885-896. <https://doi.org/10.1016/j.jbusres.2018.11.015>
- Cripps, J. D., & Meyer, R. J. (1994). Heuristics and biases in timing the replacement of durable products. *Journal of Consumer Research*, 21(2), 304-318. <https://doi.org/10.1086/209399>
- Daly, T. M., & Natarajan, R. (2015). Swapping bricks for clicks: Crowdsourcing longitudinal data on Amazon Turk. *Journal of Business Research*, 68(12), 2603-2609. <https://doi.org/10.1016/j.jbusres.2015.05.001>
- DelSole, M. (2018) What is one hot encoding and how to do it. Retrieved from <https://medium.com/@michaeldelsole/what-is-one-hot-encoding-and-how-to-do-it-f0ae272f1179>
- DeVellis, R. F. (2016). *Scale development: Theory and applications*. Los Angeles: Sage Publications.

- 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), 253-269. <https://doi.org/10.1177%2F1069072715580324>
- Engel, J. F., Kollat, D. T., & Blackwell, R. D. (1968). *Consumer behavior*. New York: Holt, Rinehart & Winston.
- Fisher, A., Rudin, C., & Dominici, F. (2019). All Models are wrong, but many are useful: learning a variable's importance by studying an entire class of prediction models simultaneously. *Journal of Machine Learning Research*, 20(177), 1-81. Retrieved from <http://www.jmlr.org/papers/v20/18-760.html>
- Frederick, S., & Loewenstein, G. (1999). Hedonic adaptation. In D. Kahneman, E. Diener, N. Schwarz (Eds.), *Well-being: The foundations of hedonic psychology* (pp. 302-329). New York: Russell Sage Foundation.
- Galak, J., & Redden, J. P. (2018). The properties and antecedents of hedonic decline. *Annual Review of Psychology*, 69, 1-25. <https://doi.org/10.1146/annurev-psych-122216-011542>
- Géron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems* (2nd ed.). Sebastopol, CA: O'Reilly Media.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. Cambridge, MA: MIT Press.
- Gordon, B. R. (2009). A dynamic model of consumer replacement cycles in the PC processor industry. *Marketing Science*, 28(5), 846-867. <https://doi.org/10.1287/mksc.1080.0448>
- 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), 101-115. <https://doi.org/10.1509%2Fjmk.41.1.101.25090>
- H2O. (2019). H2O Driverless AI. Retrieved from <https://www.h2o.ai/products/h2o-driverless-ai/>
- Haisley, E., & Loewenstein, G. (2011). It's not what you get but when you get it: The effect of gift sequence on deposit balances and customer sentiment in a commercial bank. *Journal of Marketing Research*, 48(1), 103-115. <https://doi.org/10.1509%2Fjmk.48.1.103>
- Hasegawa, S., Terui, N., & Allenby, G. M. (2012). Dynamic brand satiation. *Journal of Marketing Research*, 49(6), 842-853. <https://doi.org/10.1509%2Fjmr.11.0036>
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. New York, NY: Springer.
- He, T. (2018). XGBoost v0.71.2. R Documentation. Retrieved from <https://www.rdocumentation.org/packages/xgboost/versions/0.71.2>
- Huh, Y. E., & Kim, S.-H. (2008). Do early adopters upgrade early? Role of post-adoption behavior in the purchase of next-generation products. *Journal of Business Research*, 61(1), 40-46. <https://doi.org/10.1016/j.jbusres.2006.05.007>
- Ittersum, K. V., & Feinberg, F. M. (2010). Cumulative timed intent: A new predictive tool for technology adoption. *Journal of Marketing Research*, 47(5), 808-822. <https://doi.org/10.1509%2Fjmk.47.5.808>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: With Applications in R*. New York: Springer.
- Kim, H. S., Wohl, M. J. A., Salmon, M. M., Gupta, R., & Derevensky, J. (2015). Do social casino gamers migrate to online gambling? An assessment of migration rate and potential predictors. *Journal of Gambling Studies*, 31(4), 1819-1831. <https://doi.org/10.1007/s10899-014-9511-0>
- Kuhn, M. (2019). *The caret package*. Retrieved from <http://topepo.github.io/caret/index.html>
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108(3), 480-498. <https://psycnet.apa.org/doi/10.1037/0033-2909.108.3.480>
- Lewin, K. (1946). Behavior as a function of total situation. In D. Cartwright (Ed.), *Field theory in social science: Selected theoretical papers* (pp. 238-304). New York: Harper & Row.
- Lipton, Z. C., Elkan, C., & Narayanaswamy, B. (2014). Thresholding classifiers to maximize F1 score. ArXiv, 1402.1892 [stat.ML]. Retrieved from <http://arxiv.org/abs/1402.1892>

- Luo, L., Ratchford, B. T., & Yang, B. (2013). Why we do what we do: A model of activity consumption. *Journal of Marketing Research*, 50(1), 24-43. <https://doi.org/10.1509%2Fjmr.10.0416>
- MacInnis, D. J., & Folkes, V. S. (2010). The disciplinary status of consumer behavior: A sociology of science perspective on key controversies. *Journal of Consumer Research*, 36(6), 899-914. <http://dx.doi.org/10.1086/644610>
- 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), 1-26. <https://doi.org/10.1177%2F002224299005400101>
- Mahajan, V., Muller, E. & Bass, F. M. (1995). Diffusion of new products: Empirical generalizations and managerial uses. *Marketing Science*, 14(3 sup.), 79-88. <https://doi.org/10.1287/mksc.14.3.G79>
- MathWorks (2018). Detector performance analysis using ROC curves. Retrieved from <https://www.mathworks.com/help/phased/examples/detector-performance-analysis-using-roc-curves.html>
- Meyer, R. J., Zhao, S., & Han, J. K. (2008). Biases in valuation vs. usage of innovative product features. *Marketing Science*, 27(6), 1083-1096. <https://doi.org/10.1287/mksc.1080.0367>
- 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), 151-172. <https://doi.org/10.1177%2F0022242918809380>
- Molnar, C. (2019). Interpretable machine learning: A guide for making black box models explainable. Retrieved from: <https://christophm.github.io/interpretable-ml-book/>
- 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), 1178-1196. <https://doi.org/10.1108/EJM-09-2016-0521>
- Mourdoukoutas, P. (2018). Samsung beats Apple in the global smartphone market as Chinese brands close in. *Forbes*. Retrieved from <https://www.forbes.com/sites/panosmourdoukoutas/2018/09/13/samsung-beats-apple-in-the-global-smartphone-market-as-chinese-brands-close-in/>
- Munster, G. (2018, March). iPhone intent to upgrade survey suggests more predictable demand. *Loup ventures*. Retrieved from <https://loupventures.com/iphone-intent-to-upgrade-survey-suggests-more-predictable-demand/>
- Murphy, K. P. (2012). *Machine learning: A probabilistic perspective*. Cambridge, MA: MIT Press.
- Okada, E. M. (2001). Trade-ins, mental accounting, and product replacement decisions. *Journal of Consumer Research*, 27(4), 433-446. <https://doi.org/10.1086/319619>
- Okada, E. M. (2006). Upgrades and new purchases. *Journal of Marketing*, 70(4), 92-102. <https://doi.org/10.1509%2Fjmk.70.4.092>
- Owen, M. (2018). Analyst estimates average lifespan for all Apple devices at over four years. Retrieved from <https://appleinsider.com/articles/18/03/01/analyst-estimates-average-lifespan-for-all-apple-devices-at-over-four-years>
- Reczek, R. W., Irwin, J. R., Zane, D. M., & Ehrich, K. R. (2018). That's not how I remember it: Willfully ignorant memory for ethical product attribute information. *Journal of Consumer Research*, 45(1), 185-207. <https://doi.org/10.1093/jcr/ucx120>
- Reisinger, D. (2018). Why Apple will sell hundreds of millions of iPhone upgrades in the next 18 months. *Fortune*. Retrieved from <http://fortune.com/2018/02/27/apple-iphone-upgrades/>
- Salminen, J., Yoganathan, V., Corporan, J., Jansen, B. J., & Jung, S.-G. (2019). Machine learning approach to auto-tagging online content for content marketing efficiency: A comparative analysis between methods and content type. *Journal of Business Research*, 101, 203-217. <https://doi.org/10.1016/j.jbusres.2019.04.018>
- Sela, A., & LeBoeuf, R. A. (2017). Comparison neglect in upgrade decisions. *Journal of Marketing Research*, 54(4), 556-571. <https://doi.org/10.1509%2Fjmr.14.0394>
- 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), 242-251. <https://doi.org/10.1016/j.jretai.2011.01.006>

- Singh, J. P., Irani, S., Rana, N. P., Dwivedi, Y. K., Saumya, S., & Roy, P. K. (2017). Predicting the “helpfulness” of online consumer reviews. *Journal of Business Research*, 70, 346-355. <https://doi.org/10.1016/j.jbusres.2016.08.008>
- Takahashi, D. (2018, September). Newzoo: Smartphone users will top 3 billion in 2018, hit 3.8 billion by 2021. *Venture Beat*. Retrieved from <https://venturebeat.com/2018/09/11/newzoo-smartphone-users-will-top-3-billion-in-2018-hit-3-8-billion-by-2021/>
- Wang, Y., & John, D. R. (2019). Up, up, and away: Upgrading as a response to dissimilar brand users. *Journal of Marketing Research*, 56(1), 142-157. <https://doi.org/10.1177%2F0022243718820572>
- Wu, E. C., Moore, S. G., & Fitzsimons, G. J. (2019). Wine for the table: Self-construal, group size, and choice for self and others. *Journal of Consumer Research*, 46(3), 508-527. <https://doi.org/10.1093/jcr/ucy082>

Authors' contributions

1st author: conceptualization (lead); methodology (lead); writing – original draft (equal); writing – review & editing (supporting); project administration (lead); supervision (lead); formal analysis (equal); software (equal); validation (supporting); funding acquisition (lead).

2nd author: conceptualization (supporting); methodology (supporting); writing – original draft (equal); writing – review & editing (lead); validation (supporting).


3rd author: methodology (supporting); writing – original draft (equal); writing – review & editing (supporting); formal analysis (equal); data curation (lead); software (equal); validation (lead).

4th author: methodology (supporting); investigation (lead); writing – original draft (supporting); data curation (supporting).

Authors


Vinicius Andrade Brei

Universidade Federal do Rio Grande do Sul, Escola de Administração
Rua Washington Luis, 855, Centro Histórico, 90010-460, Porto Alegre, RS, Brazil
brei@ufrgs.br

 <https://orcid.org/0000-0002-0502-4533>


Leonardo Nicolao

Universidade Federal do Rio Grande do Sul, Escola de Administração
Rua Washington Luis, 855, Centro Histórico, 90010-460, Porto Alegre, RS, Brazil
leonardo.nicolao@ufrgs.br

 <https://orcid.org/0000-0001-8726-2326>


Maria Alice Pasdiora

Universidade Federal do Rio Grande do Sul, Escola de Administração
Rua Washington Luis, 855, Centro Histórico, 90010-460, Porto Alegre, RS, Brazil
alice.pasdiora@ufrgs.br

 <https://orcid.org/0000-0001-9561-757X>

Rodolfo Coral Azambuja

Universidade Federal do Rio Grande do Sul, Escola de Administração
Rua Washington Luis, 855, Centro Histórico, 90010-460, Porto Alegre, RS, Brazil
rodolfoazambuja@gmail.com

 <https://orcid.org/0000-0001-6656-2549>

Peer review is responsible for acknowledging an article's potential contribution to the frontiers of scholarly knowledge on business or public administration. The authors are the ultimate responsible for the consistency of the theoretical references, the accurate report of empirical data, the personal perspectives, and the use of copyrighted material.



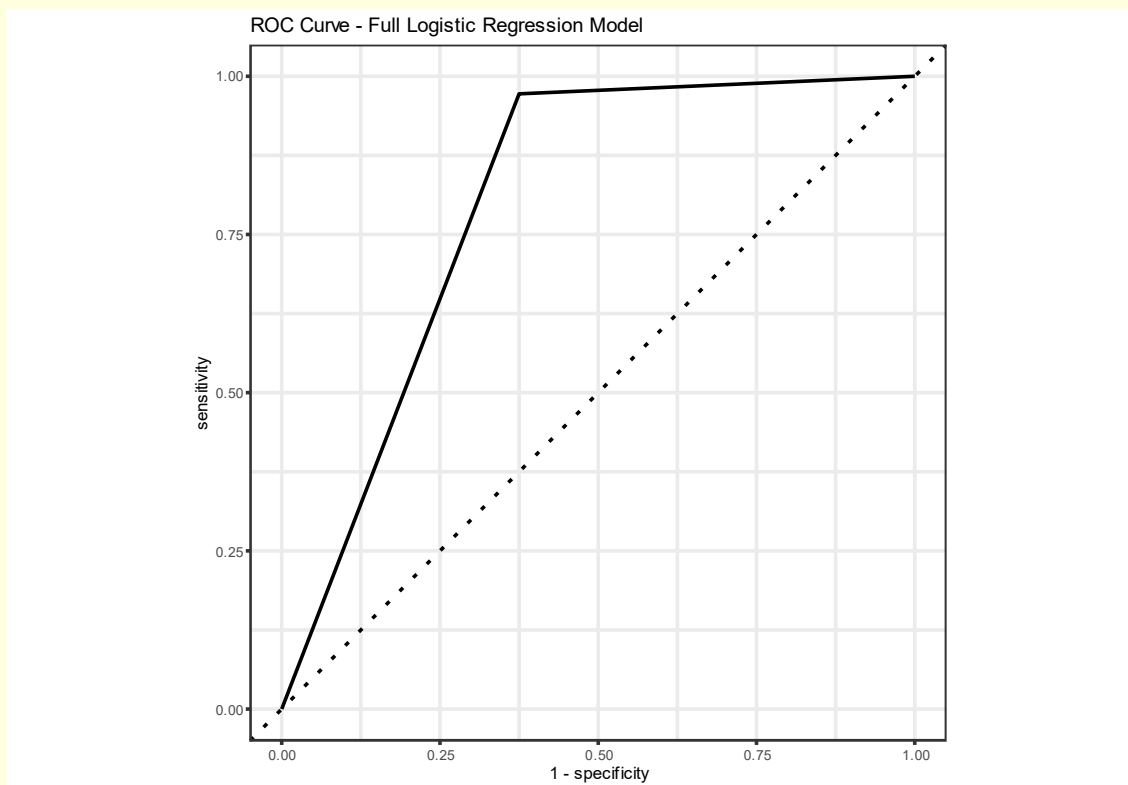


Figure 5. ROC curve – full logistic regression model

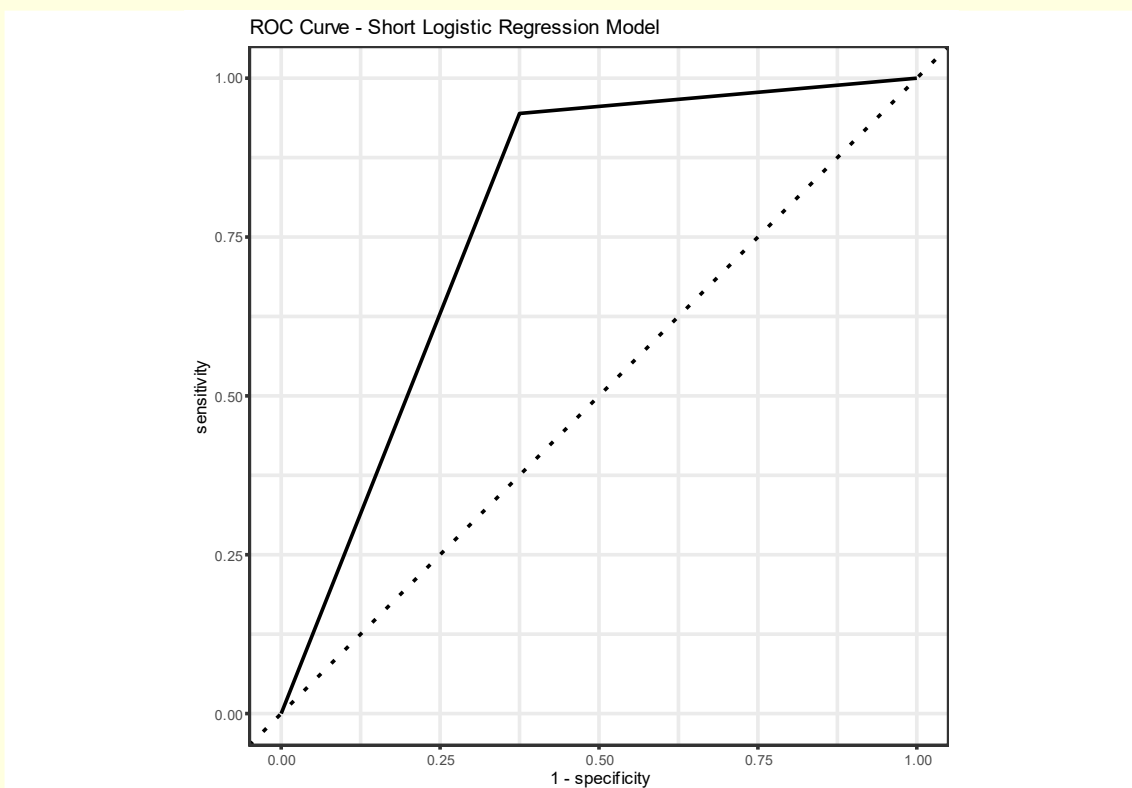


Figure 6. ROC curve – short logistic regression model

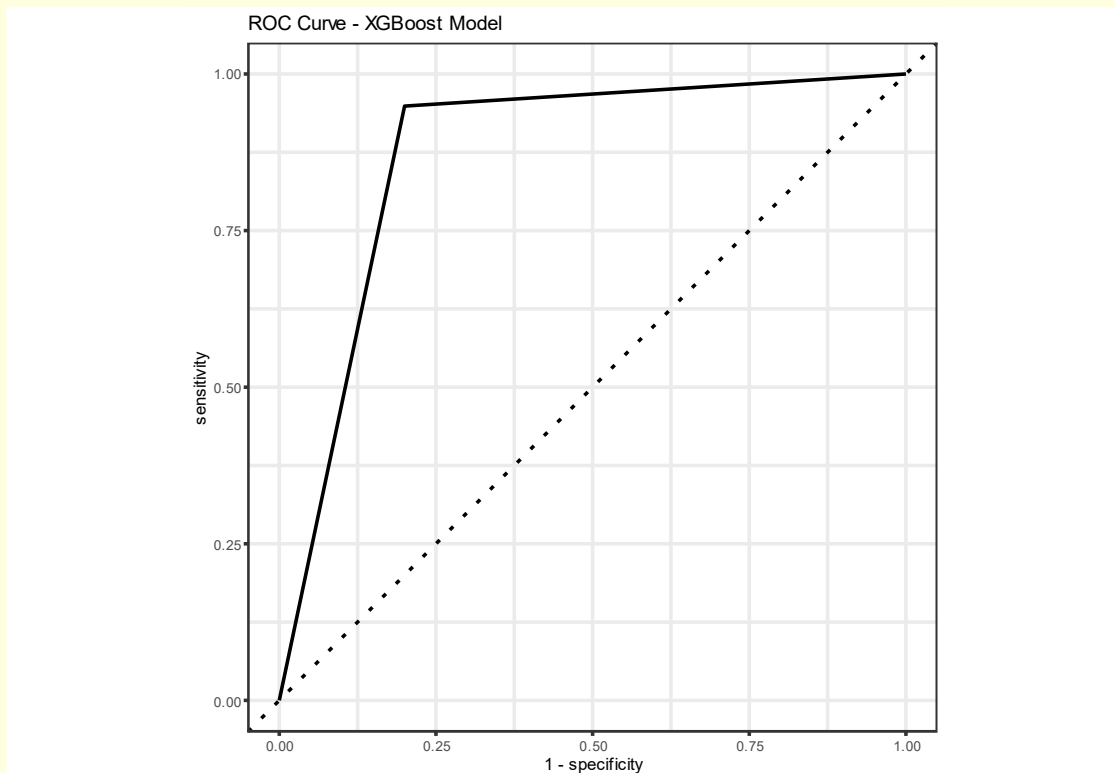


Figure 7. ROC Curve – XGBoost model

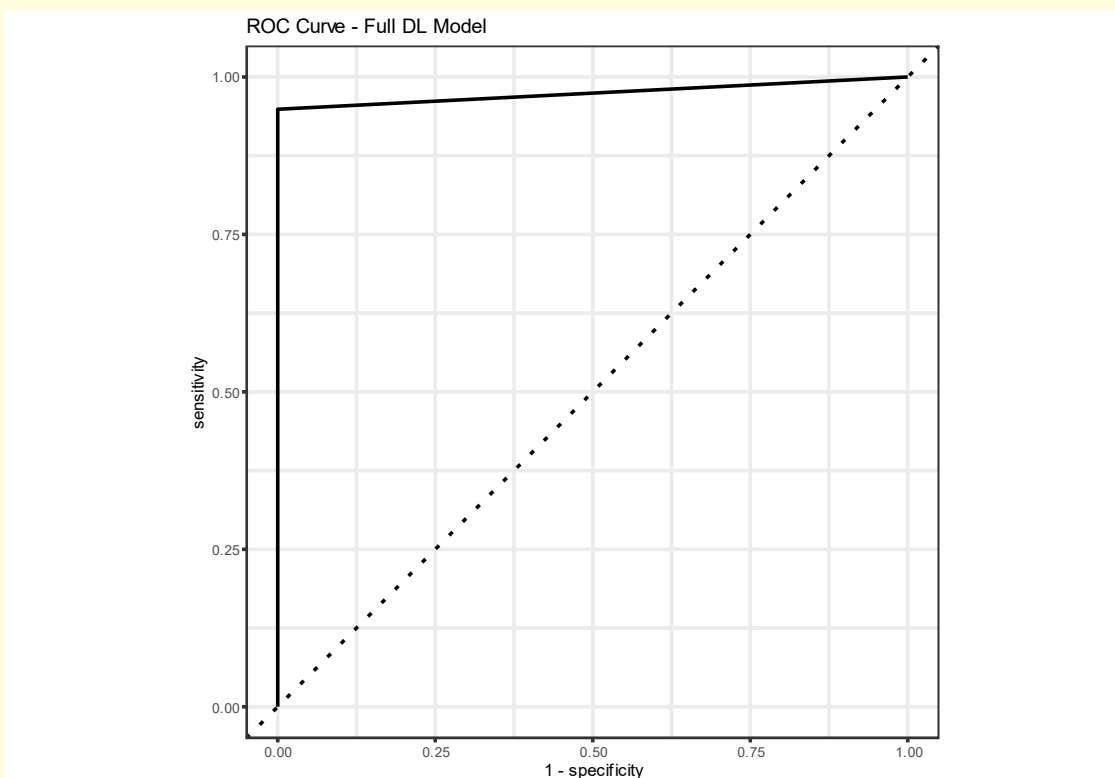


Figure 8. ROC curve – full DL model

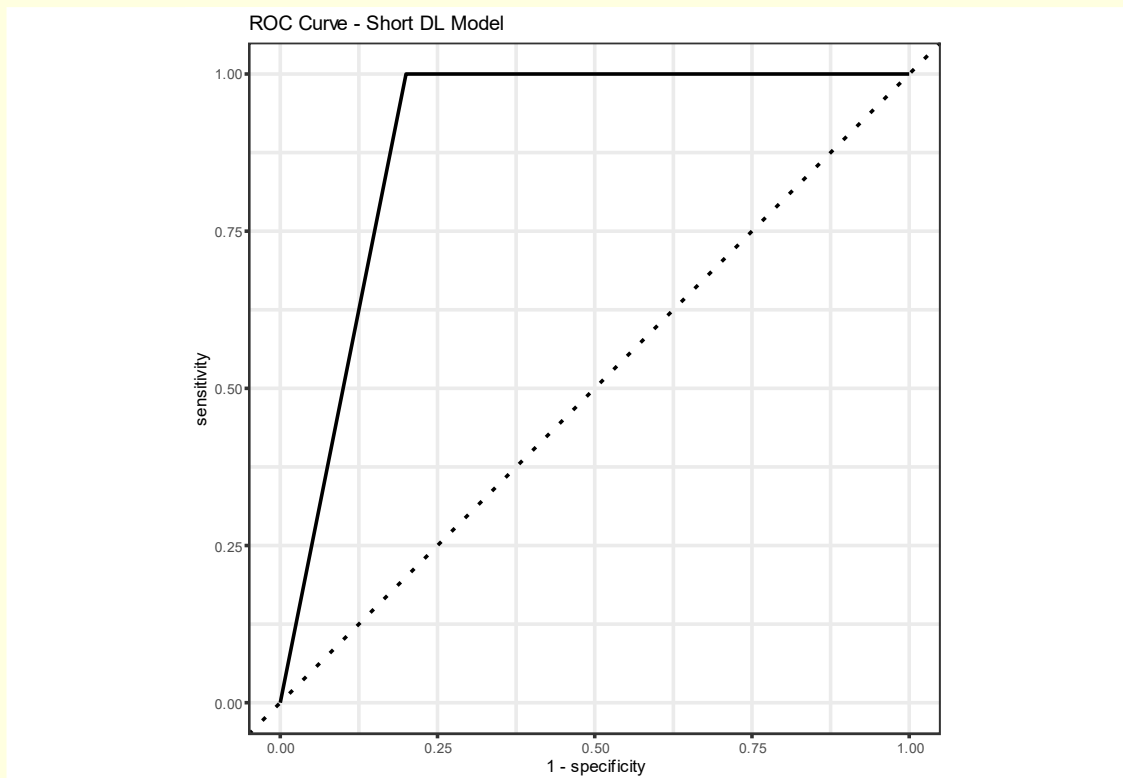


Figure 9. ROC curve – short DL model