THE BIG FIVE PERSONALITY TRAITS AS PREDICTORS OF FINANCIAL WELLBEING: A BIG DATA APPROACH

by

OSMAN CAN GENÇYÜREK

Submitted to the Graduate School of Management
in partial fulfillment of the requirements for the degree of
Master of Science in Business Analytics

Sabancı University
July 2019

THE BIG FIVE PERSONALITY TRAITS AS PREDICTORS OF FINANCIAL WELLBEING: A BIG DATA APPROACH

Approved by:

Prof. Dr. Burçin Bozkaya ....
(Thesis Supervisor)

Asst. Prof. Dr. Asuman Büyükcan Tetik ....
(Thesis Supervisor)

Assoc. Prof. Dr. Abdullah Daşcı ....

Assoc. Prof. Dr. Selim Balcısoy ....

Assoc. Prof. Dr. Mehmet Harma ....

Approval Date: July 19, 2019

ABSTRACT

THE BIG FIVE PERSONALITY TRAITS AS PREDICTORS OF FINANCIAL WELLBEING: A BIG DATA APPROACH

OSMAN CAN GENÇYÜREK

Master of Science Thesis, July 2019

Thesis Supervisors: Prof. Dr. Burçin Bozkaya,

Asst. Prof. Dr. Asuman Büyükcan Tetik

Keywords: Financial Wellbeing, Personality Traits, Predictive Modeling, Binary Classification

Research has posited credit card transactions as highly probable to be grounded on the personality of the card holder. In this research, we investigate whether the big five personality traits of customers derived from credit card transactions predict their financial wellbeing. Our approach uses real data from a private Turkish bank, which contain both the demographic and financial records of 10,172 consumers located in Istanbul with 911,280 transactions. We filter purchasing categories related to the big five personality traits from Matz, Gladstone, and Stillwell's study (2016). First, we link spending categories to the big five personality traits by considering Matz et al.'s study (2016). Then we calculate the big five factor scores of customers by monthly aggregating the individual big five scores of their transactions. Next, we investigate the relationship between the monthly big five personality scores and payment behavior of their credit card statements. In our main model, we estimated customers' on-time payment behavior of the full amount due 8.8 % better than a random prediction (with 54.4 % AUROC value) by using their monthly big five personality scores and yearly and six-month based trends as independent variables.

ÖZET

FİNANSAL REFAH TAHMİNLEYİCİSİ OLARAK BÜYÜK BEŞLİ KİŞİLİK ÖZELLİKLERİ: BÜYÜK VERİ YAKLAŞIMI

OSMAN CAN GENÇYÜREK

İş Analitiği Yüksek Lisans Tezi, Temmuz 2019

Tez Danışmanları: Prof. Dr. Burçin Bozkaya,

Yrd. Prof. Dr. Asuman Büyükcan Tetik

Anahtar Kelimeler: Finansal Refah, Kişilik Özellikler, Tahminsel Modelleme, İkili Sınıflandırma

Bu çalışma, kredi kartı işlemlerinin kart sahibinin kişilik özelliklerine dayandırılabileceği üzerine temellendirilmiştir. Bu araştırmada, kredi kartı işlemlerinden türetilen müşterilerin beş büyük kişilik özelliğinin finansal refahlarını öngörüp öngörmediğini araştırıyoruz. Yaklaşımımızda, 911.280 adet işlemle İstanbul'da yaşayan 10.172 tüketicinin demografik ve mali kayıtlarını içeren özel bir Türk bankasından elde edilen gerçek verileri kullanıyoruz. Verimizdeki satın alma kategorilerini Matz, Gladstone ve Stillwell'in çalışmasından (2016) büyük beş kişilik özelliği ile ilgili satın alma kategorileri ile eşleştirerek filtreliyoruz. Öncelikle, Matz ve çalışma arkadaşlarının çalışmasını göz önünde bulundurarak harcama kategorilerini beş kişilik kişilik özelliği ile ilişkilendirdik (2016). Ardından, müşterilerin büyük beş faktör puanını, işlemlerinin tekil büyük beş puanını aylık olarak toplayarak hesaplıyoruz. Daha sonra, aylık büyük beş kişilik puanı ile kredi kartı faturalarının ödenme davranışı arasındaki ilişkiyi araştırıyoruz. Ana modelimizde, aylık büyük beş kişilik puanını ve yıllık ve altı aylık trendlerini bağımsız değişkenler olarak kullanıp, müşterilerin fatura bedelinin tamamını zamanında ödeme davranışını rastgele tahminlemeden % 8,8 daha iyi (% 54,4 AUROC değeri ile) tahmin ettik.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

The primary purpose of this research is to investigate the behavioral roots of transactional banking data in relation to the personality traits of the card holder. Previous research indicates correlations between different aspects of personality traits and several aspects of financial wellbeing, for example, the relation of the big five personality traits with household saving behavior (Nyhus & Webley, 2001), money management (Donnelly, Iyer, & Howell, 2012), and shopping habits (Otero-López & Villardefrancos, 2013), as well as the association between financial wellbeing and self-control (Strömbäck, Lind, Skagerlund, & Västfjäll, 2017). As repositories of all of the financial actions of their customers, banks store these records including this enormous volume of transactional data reflects their day-to-day financial behavior. We hypothesize that such records of financial behavior are grounded on personality traits. Since customers in Turkey frequently use credit cards for purchases, we aimed to use a corresponding dataset to establish a link between financial behavior and the big five personality traits, and then investigate the dataset to demonstrate the relation with customers' financial wellbeing. We define financial wellbeing in four statuses of payment of the credit card statement: paying the *minimum* amount of the statement on time, paying the *minimum* amount of the statement within three days of the grace period, paying the *total* amount of statement on time and paying the *total* amount of statement within three days of the grace period. Our main research question, hence, is as follows: is it possible to predict an individual's financial wellbeing through his/her big five personality traits derived from transactional data? To explore this research question, we analyzed 8,138,525 credit card transactions and the associated spending categories (e.g., restaurants, hotels) for 103,209 customers of a private bank in Turkey and built prediction models for 22,401 customers using 911,280 credit card transactions, as we further detail below.

The main contributions of our research are two-fold:

- We combine two seemingly unrelated and different research cultures, namely behavioral personality science and (big) data science. We adopted our predictors based on big five personality traits from personality psychology in an empirical manner to apply in machine learning algorithms. In our empirical approach, we first transformed spending categories and amounts into big five personality scores of customers. Then we used those scores to predict future financial behavior and wellbeing of customers while we addressed the payment information on the credit card statement as indicators of financial wellbeing. Thus, this research unites behavioral personality science and (big) data science in the context of banking. While previous literature was mostly survey based, our contribution to this literature includes the big data perspective.

- From an applied perspective, our study provides a novel approach for banks to understand and predict the financial wellbeing of their customers by assessing customers' personality traits derived from spending categories.

1.1. Big Five Personality Traits

We used big five personality traits to measure personality of customers. These five traits represent the different dimensions of human personality. They are widely known with their acronym called "OCEAN". Explanations of those big five personality traits are below.

- Openness: The full name of this traits can be found as openness to new experiences. People who have high levels of openness tend to have more interest in different subject due to their will to explore new things. Hence, they tend to be more creative and engage with art. In Goldberg's study (1990) who is one of the pioneers of this notion, it was measured by the concepts of wisdom, originality and objectivity.

- Conscientiousness: This trait can be explained as being self-disciplined, goal-oriented and planning several future steps instead of having impulsive decisions. In Goldberg's study (1990), it was measured by the concepts of self-discipline, consistency and reliability.
- Extraversion: High levels of extraversion trait can be summarized by being outgoing and expressing emotions easily. Extravert people can socialize easily. Talkativeness, sociability and adventure are among the notions that Goldberg (1990) used them to measure extraversion trait.
- Agreeableness: People have high levels of agreeableness trait are good at having empathy with others. They tend to be supportive and compromising when other people in need. This trait is measured by trust, generosity and tolerance in Goldberg's (1990) study.
- Neuroticism: This trait is associated with mood swings and having instable emotions. It is related with self-pity, anxiety and insecurity in Goldberg's (1990) study.

CHAPTER 2

LITERATURE REVIEW

2.1. Personality Traits and Financial Wellbeing

Previous survey-based studies provide insight regarding the link between financial wellbeing and personality traits. For example, Nyhus and Webley (2001) investigate the effects of personality traits on saving and borrowing behavior. They study a Dutch dataset which includes detailed information of assets and debts of the subjects in their sample. Their results suggest that emotional instability (neuroticism) and extraversion are valuable predictors for saving and borrowing behaviors. Both neuroticism and extraversion are negatively related to saving while they are positively related to borrowing.

Brown and Taylor (2014) use the British Household Panel survey data to analyze the relation between personality traits and financial decision making. They use the big five personality traits to explore their effect on unsecured debt and financial assets. This dataset is collected over sequential waves from 1991 to 2008. Their results reveal that extraversion is positively related to unsecured debt in their sample of single individuals. However, in their sample of couples, agreeableness positively relates with unsecured debt. In the whole sample, conscientiousness has a negative, but other big five personality traits have a positive relation with unsecured debt. None of the big five personality traits have a significant association with financial assets.

Donnell et al. (2012) conducted four online surveys to explore the big five personality traits in relation to money management and compulsive buying behavior. In their study, conscientiousness has the strongest positive association with money management (e.g., budgeting, saving, investing Godwin & Koonce, 1992) and financial wellbeing.

Otero-López and Villardefrancos Pol (2013) underline the relation between both excessive and compulsive buying behavior as well as personality traits through the lens of the big five personality traits. Neuroticism, extraversion, openness, and agreeableness reveal a positive association with excessive buying while only conscientiousness negatively relates with excessive buying. Their additional study conducted 6 months later indicates another compelling correlation amongst the traits of neuroticism, agreeableness, and conscientiousness and compulsive buying (Otero-López & Villardefrancos Pol, 2013). The highest levels of neuroticism were observed for the high compulsive buying propensity group while the lowest levels of conscientiousness were observed for the same group. The high compulsive buying propensity group also has the lowest levels of agreeableness.

The literature discussed above forms the basis for our research to explore the relation between personality and financial wellbeing. Another related study that aims to deduce personality as a predictor out of credit card spending categories is by Matz, Gladstone, and Stillwell (2016). These authors analyze 76,000 bank transactions to explore the relation between personality and spending categories. They discuss the match between customer spending patterns and personality, and also the effect of this match on happiness. The researchers conclude by emphasizing the positive effect of spending that reflects an individual's personality on their wellbeing. In our study, we benefit from the big five personality ratings of spending categories Matz and his colleagues (2016) used. The significance of their research for our purposes is the big five personality scores for spending categories. As detailed in the following sections, we employ these scores in forming and testing our hypothesis that financial behavior is grounded on personality traits.

2.2. Financial Wellbeing through Big Data Lens

Built on the methods of previous research, our research is based on assessments of financial wellbeing and spending behaviors by using a big data approach. For example,

Singh, Bozkaya and Pentland (2015) study the association between financial wellbeing and foraging behavior of individuals by using a big data approach. These researchers indicate that the spatio-temporal data on customer transactions is an effective predictor of the future financial outcomes of customers. Their findings include the following observations : customers with regular mobility behavior tend to pay their bills on time; customers who manifest high levels of diversity (shopping behavior varying over space and time) and loyalty (a shopping behavior occurring frequently at the same or similar places and time slots) have less tendency to overspend but more tendency to miss payments.

Another study conducted by Singh, Freeman, Lepri and Pentland (2013) aims to predict the spending behavior of individuals through mobile phone-based social interaction data. Their results suggest that more social couples have a greater tendency to overspend.

Dong and his colleagues (2018) study common urban purchase behaviors stemming from individuals acting as “social bridges” between different communities, again with a big data perspective. These authors find that social bridges can influence the form of community purchase behavior. They show that the purchasing behaviors of consumers acts as social bridges, spreading out within their connections to their respective communities.

Khandani, Kim and Lo (2010) use machine learning approaches to predict consumer credit-risks. Their independent variables contain customer transactions credit bureau data. Their findings indicate that they successfully predict delinquencies and defaults of consumers.

Another study conducted by Kruppa, Schwarz, Arminger and Ziegler (2013) is designed to estimate the probability of default which is used by banks to decide credibility of customers. Their results indicate that using machine learning techniques can provide reliable outcomes about predicting the probability of default of customers.

Addo, Guegan and Hassani (2018) predict loan default probability in their study. Their data is a transactional banking data set including financial and income statements, balance sheets and cash flows. They perform binary classification prediction of default or no default. They used machine learning algorithms such as elastic net (an extension of linear regression), random forest, gradient boosting machine and deep learning. Their results indicate that tree-based models provide more consistent predictions instead of complex and non-transparent deep learning models.

2.3. The Present Study

In this study, we work on the assessment of financial wellbeing by using the big five personality traits inferred from transactional big data. As mentioned above, there are several similar studies in this vein. Our main contribution is at the intersection of two research fields: personality psychology and big data. We aim to connect personality traits to financial wellbeing by evaluating the indicators and associations of personality traits by using big data and machine learning methodologies. Our study relates personality traits and financial wellbeing by approaching the problem from the perspective of (big) data science.

CHAPTER 3

DESCRIPTIVE ANALYSIS AND DATA PREPROCESSING

3.1. Descriptive Analysis

In this research, we started to work on an anonymized dataset from a private Turkish bank, which contains both demographic and financial transaction records of 103,209 customers located in Istanbul, Turkey. This dataset has 8,138,525 credit card purchase activities during a period of 12 months from July 2014 to June 2015. We also have the payment history for each credit card account, along with monthly statements as well as the dates and amounts of payments. In Figures 1-4, we describe these customers with respect to their demographic data. We use monthly statements containing their corresponding payment information to assess the monthly financial well-being of customers based on the four measures described above.

The demographic structure of customers in our data is as follows:

74.2% of customers are male while 25.8%, female. In terms of level of education, 7.3% graduated from primary school; 8.5%, middle school; 45.4%, secondary school; 7.8%, college; 26%, university; 3.4%, masters; 0.2 % PhD; 1.3% uneducated, with 0.1% unknown in terms of education status. 30.5% were single, 63% married, 5% divorced, 0.5% widowed and 1.5% unknown in terms of marital status. 13.1% of customers were between the ages of 18 and 25, 38.2% between 26 and 35, 29.8% between 36 and 45, 14.3% between 46 and 55, 3.9% between 56 and 65, 0.6% between 66 and 75; and 0.1% were older than 75

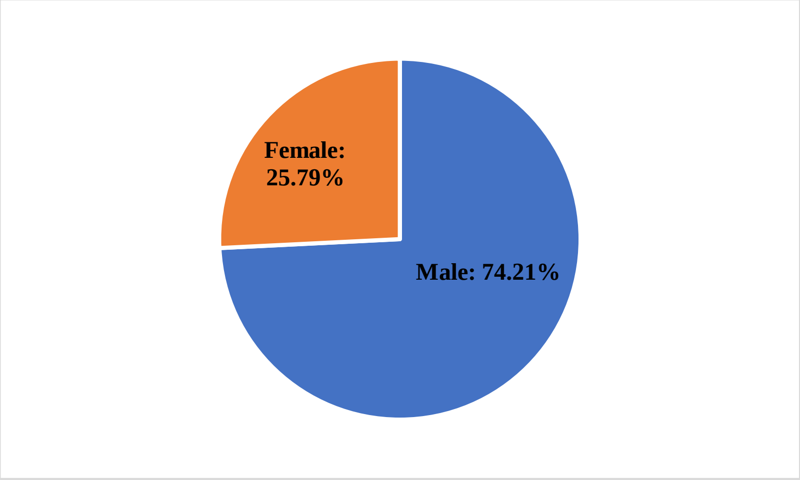


Figure 1. Pie Chart by Gender

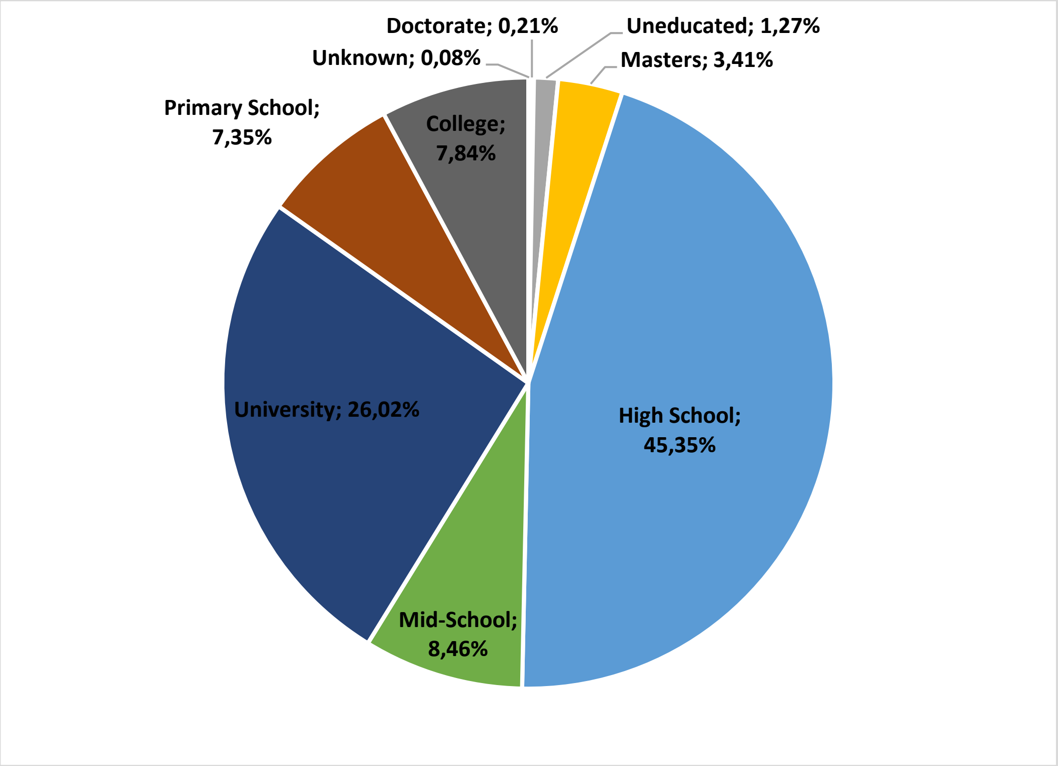


Figure 2. Pie Chart by Education Status

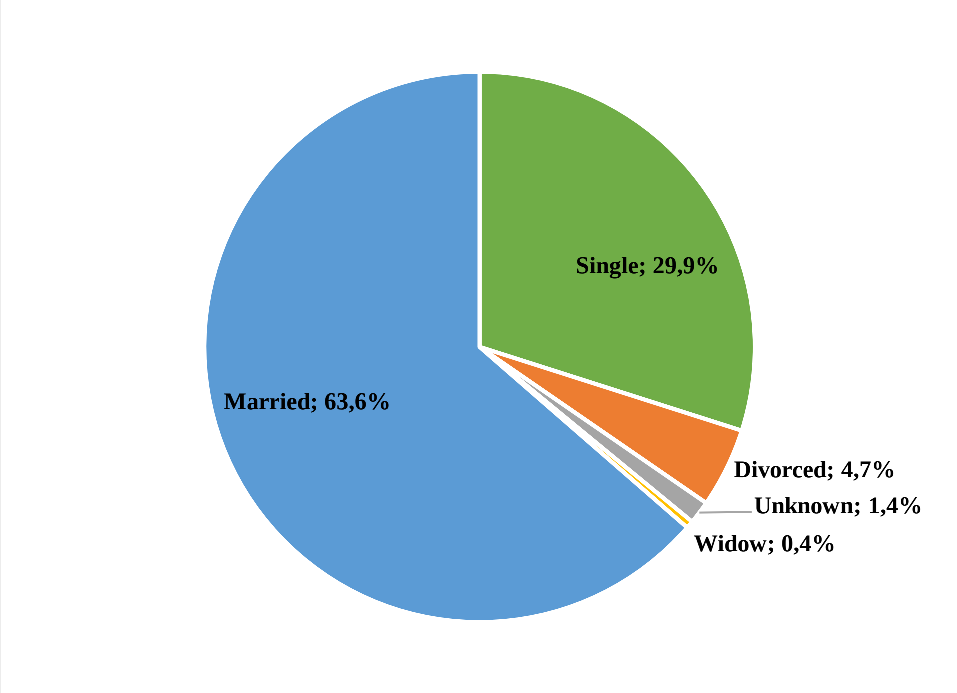


Figure 3. Pie Chart by Marital Status

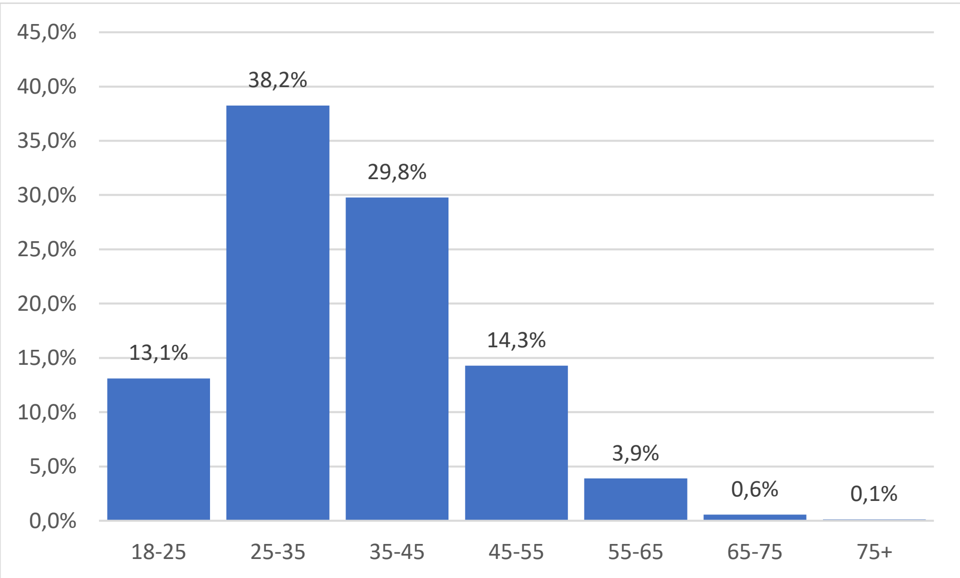


Figure 4. Histogram by Age (Average age = 36.5, *SD* = 10)

We use credit card purchase activities to generate monthly big five personality indices. The distribution of credit card spending categories in our initial data is as follows:

- 0.2% of credit card spending was marked for pubs and casinos, 7.9% fuel-oil,0.5% shopping malls, 0.1% car rentals, 1,.2% shoes, 0.5%, white goods, 4.2%, others, 0.3%, direct marketing, 0.3%education, 0.4%, fun and sports, 33.4%, food expenditure per household, 3.1% services, 0.6% airlines, 0.7% hotels, 1.5% cosmetic, 0.4% jewelry, 1.5% decoration, 0.7% music-market-stationary, 1.8% cash advance, 0.2% optic, 0.9% automotive, 0.4% toys, 15.5% restaurant, 1.9% insurance, 0.4% cinema-theatre-art, 3.9% health, 1.2% travel agencies-transportation, 0.8% sport wear, 2.3% technology, 8.1% textile, 4.2% telecommunications, 0.9% ironmongery (hardware store).

We did not include the income variable in our dataset since we learnt that the bank does not have the actual income data of its customers. Instead, income is a calculated variable based on the data owned by the bank. We decided that using this kind of estimated income variable would not produce a solid ground for our analyses. This is also the reason that we calculate monthly big five personality scores based on total spending on our selected categories instead of customer’s income nor using income as a predictor or control variable.

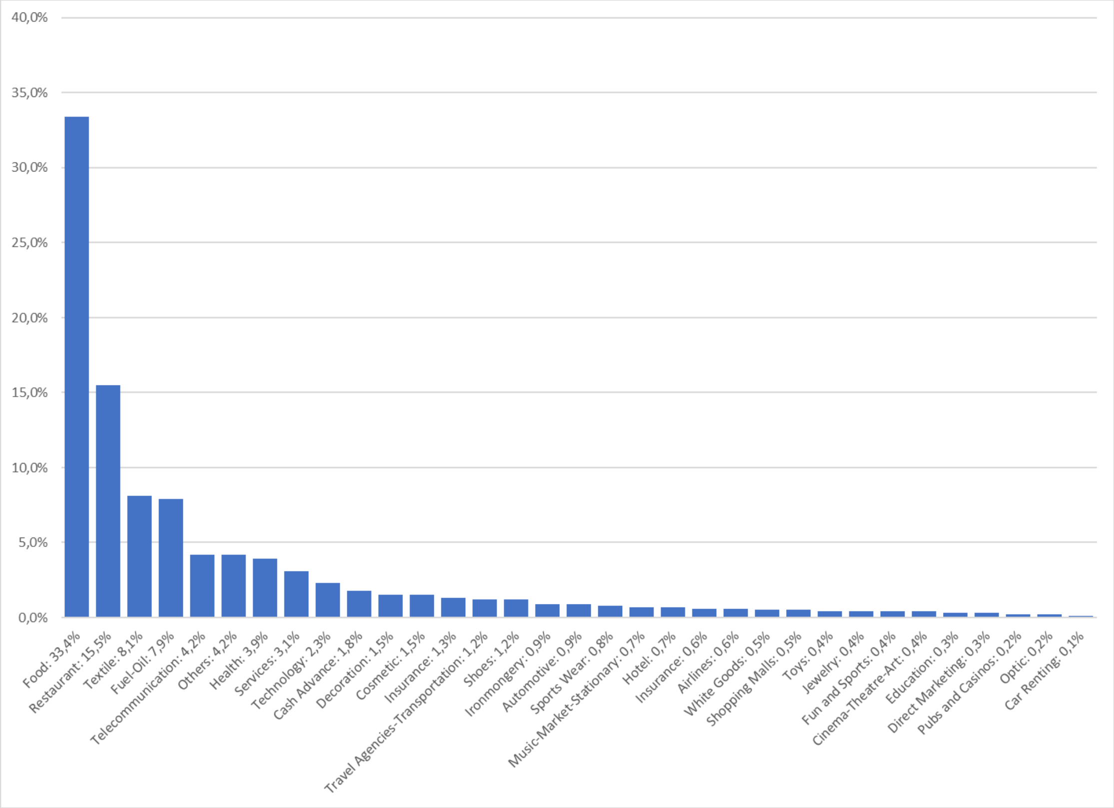


Figure 5. Bar Chart by Spending Categories Frequency

As we describe in the next chapter, we process the monthly statements with corresponding payment information to derive monthly financial well-being indicators in four separate measures.

3.2. Data Preprocessing

Our data pre-processing includes two main steps. In the first step, we apply the findings from personality psychology literature to generate the big five scores of customers in relation to their credit card spending categories. This first step is called translation step. The second step consists of traditional data cleaning and data manipulation operations to calculate the big five personality scores of each transaction and summarize them on a monthly basis for each customer. We use the resulting processed data in the prediction modeling phase. In the second phase, we model the payment behavior of customers in relation to their monthly big five personality scores and make predictions.

3.2.1. Translation Step

We use the findings of Matz et al. (2016) who have associated 59 spending categories with the big five personality traits. They hired 100 Amazon Mechanical Turk workers to score spending categories (from -3 to +3) as if they are real people to characterize based on big five personality traits. Our approach to benefit Matz et al.'s (2016) findings to form our spending categories' big five personality scores is summarized through depiction below.

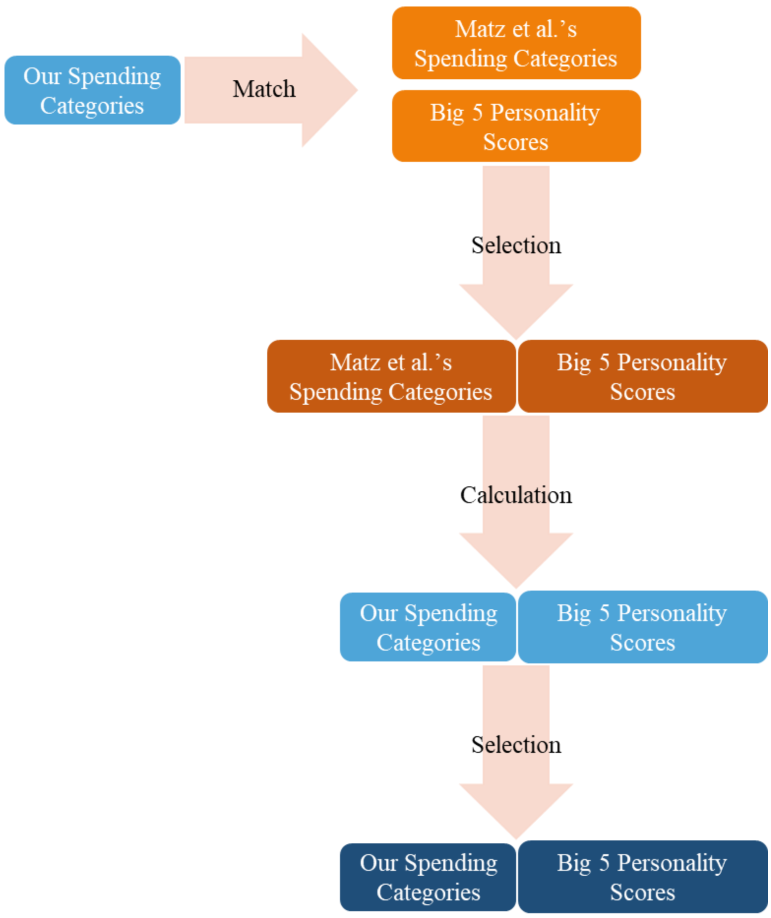


Figure 6. Flowchart of Translation Step Processes

In a nutshell, we first match the spending categories, then we select effective Matz et al.’s (2016) spending categories on big five personality traits among matched categories. Selected ones are represented by red boxes. After that selection, we calculate big five personality scores of our spending categories by transforming big five personality scores of Matz et al.’s (2016) study. Then we select our spending categories based on the consistency of their big five personality score calculation results. Final set of our spending categories are represented by dark blue boxes. The concepts of being effective and consistent and processes of matching and calculation are detailed below.

We match and merge those categories into our credit card transactions dataset and transform their ratings based on the big five personality indices between -3 and +3. We first take one-to-one exact match between the two sets of categories. Then we match our spending categories with one or more categories of Matz et al. from similar business areas. Categories matched and merged are in Table 1 below.

Table 1

Matched and Merged Spending Categories

Our Spending Categories	**Matz et al.'s Spending Categories**
Shopping Malls	Catalogue and bargain stores
Shopping Malls	Department stores
Shopping Malls	Discount stores
Shopping Malls	Supermarkets
Car Rentals	Car rentals
Shoes	Shoe shops
Fun and Sports	Entertainment
Fun and Sports	Sports
Food	Bakers and confectioners
Food	Takeout food
Hotel	Hotels
Pubs and Casinos	Eating out: pubs
Pubs and Casinos	Gambling
Cosmetic	Hair and beauty
Jewelry	Jewelry
Decoration	Home furnishing
Motorcycle	Motor sports
Music - Market - Stationary	Books
Music - Market - Stationary	Music
Toys	Toys and hobbies
Restaurant	Coffee shops
Restaurant	Eating out: restaurants
Health	Dental care

Health	Health and fitness
Travel Agencies-Transportation	Days out and tourism
Travel Agencies-Transportation	Foreign travel
Travel Agencies-Transportation	Travel
Insurance	Health insurance
Insurance	Home insurance
Insurance	Life insurance
Cinema - Theatre - Art	Arts and crafts
Cinema - Theatre - Art	Cinemas
Technology	Computers and technology
Technology	Digital
Technology	Hardware
Technology	Photography
Technology	Information technology
Technology	Mobile telephone
Textile	Clothes
Textile	Family clothes
Telecommunication	Cable and satellite TV
Telecommunication	TV license
Ironmongery (Hardware Store)	DIY projects
Ironmongery (Hardware Store)	Gardening

We compare each big five factor score of Matz et al.'s each spending category to that factor's mean and standard deviation of all matched categories. The reason of this comparison is to take only effective spending categories on big five personality traits in both positive and negative relation into account. These analyses enable us to eliminate Matz et al.'s categories which are not effective in increasing or decreasing the monthly big five personality scores of customers. We compare our categories among themselves and select based on four ways of comparisons to be sure about being selected effective spending categories on big five personality factors more precisely. The main reason for performing the four comparison analyses is to identify consistent spending categories based on the selection of most influential categories among Matz et al.'s categories. We define consistency as having the same big five personality scores as the result of each four analyses. We vary the selection procedures in four ways to have a more dependent set of categories. We hypothesize that the final set of our spending categories with their consistent big five personality scores formed through a selection based on those four comparisons would have more indicative power to reflect the monthly personality changes. Each selection way evaluates the scores of Matz et al.'s categories by comparing their scores to range of (-0.5, 0.5) or (-

1, 1) and range of (mean + 0.5*standard deviation, mean - 0.5*standard deviation) or (mean + 1*standard deviation, mean - 1*standard deviation). Each selection procedures result in some common categories with the same scores, common categories with different scores and different categories as well, based on being survivor after our selection criteria. Those selection ways and tables of selected categories with correspondent big five personality scores are listed below (red highlights indicate the values higher than upper comparison bound while yellow highlights indicate the values lower than lower comparison bound):

- There are 21 spending categories have at least one personality factor score which is not in the range of (-0.5, 0.5) or not in the range of (mean + 0.5*standard deviation, mean - 0.5*standard deviation)

Table 2
Big Five Factor Scores of each selected Matz et al.'s Spending Categories

Matz et al.'s Spending Categories	**O**	**C**	**E**	**A**	**N**
Supermarkets	-0.69	1.27	0.51	0.58	-0.73
Car rentals	-0.53	1.39	-0.06	0.31	-0.96
Entertainment	2.67	-0.43	2.51	0.31	0.49
Sports	1.44	1.30	2.24	-0.41	0.77
Bakers and confectioners	1.45	1.59	0.86	1.41	-0.80
Hotels	-0.16	1.69	0.31	1.55	-1.63
Eating out: pubs	1.35	-0.41	2.22	0.40	0.48
Gambling	1.55	-2.08	2.33	-1.81	1.98
Hair and beauty	1.91	0.31	1.49	0.85	0.22
Jewelry	1.60	0.73	1.43	0.96	-0.61
Home furnishing	0.63	1.48	0.17	1.38	-1.22
Motor sports	1.34	0.09	2.32	-0.55	0.82
Books	1.71	1.92	-0.82	1.53	-1.39
Music	2.61	0.12	2.33	0.94	0.15
Toys and hobbies	2.19	-0.90	1.94	0.78	-0.06
Coffee shops	0.89	1.24	0.45	1.79	-1.23
Eating out: restaurants	1.56	0.44	1.74	0.91	-0.39
Dental care	-1.25	1.79	-0.59	0.32	-0.59
Health and fitness	0.32	2.22	1.29	1.00	-0.93
Days out and tourism	2.19	0.57	2.25	1.10	-0.28
Foreign travel	2.54	0.65	2.15	0.85	-0.11
Travel	2.51	0.24	2.37	1.18	-0.20
Health insurance	-1.61	1.52	-1.11	-0.16	-0.50
Home insurance	-2.05	2.40	-1.46	0.33	-1.48

Life insurance	-1.30	2.21	-1.02	1.11	-1.25
Arts and crafts	2.51	0.20	1.05	1.71	-0.46
Cinemas	2.30	0.22	1.75	0.71	-0.02
Computers and technology	1.36	2.05	0.28	0.19	-1.00
Digital	1.55	1.05	0.77	0.02	-0.45
Hardware	-0.78	1.73	-0.61	0.04	-1.22
Photography	2.33	0.69	1.44	1.09	-0.33
Information technology	0.93	1.36	0.33	0.15	-0.80
Mobile telephone	1.02	1.33	1.65	0.33	-0.13
Family clothes	-0.28	0.43	0.00	1.16	-0.96
Cable and satellite TV	0.48	0.00	1.29	-0.17	0.14
DIY projects	2.22	1.37	1.20	0.98	-0.54
Gardening	0.59	1.75	-0.73	1.94	-1.59

O: Openness – C: Conscientiousness – A: Agreeableness N: Neuroticism – E: Extraversion

- There are 14 spending categories have at least one personality factor score which is not in the range of (-0.5, 0.5) or not in the range of (mean + 1*standard deviation, mean - 1*standard deviation)

Table 3

Big Five Factor Scores of each selected Matz et al.'s Spending Categories

Matz et al.'s Spending Categories	**O**	**C**	**E**	**A**	**N**
Entertainment	2.67	-0.43	2.51	0.31	0.49
Sports	1.44	1.30	2.24	-0.41	0.77
Hotels	-0.16	1.69	0.31	1.55	-1.63
Eating out: pubs	1.35	-0.41	2.22	0.40	0.48
Gambling	1.55	-2.08	2.33	-1.81	1.98
Home furnishing	0.63	1.48	0.17	1.38	-1.22
Motor sports	1.34	0.09	2.32	-0.55	0.82
Books	1.71	1.92	-0.82	1.53	-1.39
Music	2.61	0.12	2.33	0.94	0.15
Toys and hobbies	2.19	-0.90	1.94	0.78	-0.06
Coffee shops	0.89	1.24	0.45	1.79	-1.23
Dental care	-1.25	1.79	-0.59	0.32	-0.59
Health and fitness	0.32	2.22	1.29	1.00	-0.93
Days out and tourism	2.19	0.57	2.25	1.10	-0.28
Foreign travel	2.54	0.65	2.15	0.85	-0.11
Travel	2.51	0.24	2.37	1.18	-0.20
Health insurance	-1.61	1.52	-1.11	-0.16	-0.50
Home insurance	-2.05	2.40	-1.46	0.33	-1.48
Life insurance	-1.30	2.21	-1.02	1.11	-1.25

Arts and crafts	2.51	0.20	1.05	1.71	-0.46
Cinemas	2.30	0.22	1.75	0.71	-0.02
Computers and technology	1.36	2.05	0.28	0.19	-1.00
Hardware	-0.78	1.73	-0.61	0.04	-1.22
Photography	2.33	0.69	1.44	1.09	-0.33
DIY projects	2.22	1.37	1.20	0.98	-0.54
Gardening	0.59	1.75	-0.73	1.94	-1.59

O: Openness – C: Conscientiousness – A: Agreeableness N: Neuroticism – E: Extraversion

- There are 20 spending categories have at least one personality factor score which is not in the range of (-1, 1) or not in the range of (mean + 0.5*standard deviation, mean – 0.5*standard deviation)

Table 4

Big Five Factor Scores of each selected Matz et al.'s Spending Categories

Matz et al.'s Spending Categories	**O**	**C**	**E**	**A**	**N**
Car rentals	-0.53	1.39	-0.06	0.31	-0.96
Entertainment	2.67	-0.43	2.51	0.31	0.49
Sports	1.44	1.30	2.24	-0.41	0.77
Bakers and confectioners	1.45	1.59	0.86	1.41	-0.80
Hotels	-0.16	1.69	0.31	1.55	-1.63
Eating out: pubs	1.35	-0.41	2.22	0.40	0.48
Gambling	1.55	-2.08	2.33	-1.81	1.98
Hair and beauty	1.91	0.31	1.49	0.85	0.22
Jewelry	1.60	0.73	1.43	0.96	-0.61
Home furnishing	0.63	1.48	0.17	1.38	-1.22
Motor sports	1.34	0.09	2.32	-0.55	0.82
Books	1.71	1.92	-0.82	1.53	-1.39
Music	2.61	0.12	2.33	0.94	0.15
Toys and hobbies	2.19	-0.90	1.94	0.78	-0.06
Coffee shops	0.89	1.24	0.45	1.79	-1.23
Eating out: restaurants	1.56	0.44	1.74	0.91	-0.39
Dental care	-1.25	1.79	-0.59	0.32	-0.59
Health and fitness	0.32	2.22	1.29	1.00	-0.93
Days out and tourism	2.19	0.57	2.25	1.10	-0.28
Foreign travel	2.54	0.65	2.15	0.85	-0.11
Travel	2.51	0.24	2.37	1.18	-0.20
Health insurance	-1.61	1.52	-1.11	-0.16	-0.50
Home insurance	-2.05	2.40	-1.46	0.33	-1.48
Life insurance	-1.30	2.21	-1.02	1.11	-1.25
Arts and crafts	2.51	0.20	1.05	1.71	-0.46

Cinemas	2.30	0.22	1.75	0.71	-0.02
Computers and technology	1.36	2.05	0.28	0.19	-1.00
Digital	1.55	1.05	0.77	0.02	-0.45
Hardware	-0.78	1.73	-0.61	0.04	-1.22
Photography	2.33	0.69	1.44	1.09	-0.33
Information technology	0.93	1.36	0.33	0.15	-0.80
Mobile telephone	1.02	1.33	1.65	0.33	-0.13
Family clothes	-0.28	0.43	0.00	1.16	-0.96
Cable and satellite TV	0.48	0.00	1.29	-0.17	0.14
DIY projects	2.22	1.37	1.20	0.98	-0.54
Gardening	0.59	1.75	-0.73	1.94	-1.59

O: Openness – C: Conscientiousness – A: Agreeableness N: Neuroticism – E: Extraversion

- There 14 spending categories have at least one personality factor score which is not in the range of (-1, 1) or not in the range of (mean + 1*standard deviation, mean – 1*standard deviation)

Table 5

Big Five Factor Scores of each selected Matz et al.'s Spending Categories

Matz et al.'s Spending Categories	**O**	**C**	**E**	**A**	**N**
Entertainment	2.67	-0.43	2.51	0.31	0.49
Sports	1.44	1.30	2.24	-0.41	0.77
Hotels	-0.16	1.69	0.31	1.55	-1.63
Eating out: pubs	1.35	-0.41	2.22	0.40	0.48
Gambling	1.55	-2.08	2.33	-1.81	1.98
Home furnishing	0.63	1.48	0.17	1.38	-1.22
Motor sports	1.34	0.09	2.32	-0.55	0.82
Books	1.71	1.92	-0.82	1.53	-1.39
Music	2.61	0.12	2.33	0.94	0.15
Toys and hobbies	2.19	-0.90	1.94	0.78	-0.06
Coffee shops	0.89	1.24	0.45	1.79	-1.23
Dental care	-1.25	1.79	-0.59	0.32	-0.59
Health and fitness	0.32	2.22	1.29	1.00	-0.93
Days out and tourism	2.19	0.57	2.25	1.10	-0.28
Foreign travel	2.54	0.65	2.15	0.85	-0.11
Travel	2.51	0.24	2.37	1.18	-0.20
Health insurance	-1.61	1.52	-1.11	-0.16	-0.50
Home insurance	-2.05	2.40	-1.46	0.33	-1.48
Life insurance	-1.30	2.21	-1.02	1.11	-1.25
Arts and crafts	2.51	0.20	1.05	1.71	-0.46
Cinemas	2.30	0.22	1.75	0.71	-0.02

Computers and technology	1.36	2.05	0.28	0.19	-1.00
Hardware	-0.78	1.73	-0.61	0.04	-1.22
Photography	2.33	0.69	1.44	1.09	-0.33
DIY projects	2.22	1.37	1.20	0.98	-0.54
Gardening	0.59	1.75	-0.73	1.94	-1.59

O: Openness – C: Conscientiousness – A: Agreeableness N: Neuroticism – E: Extraversion

For instance, we match our "Travel Agencies – Transportation" category with Matz et al.'s "Days out and Tourism", "Foreign Travel" and "Travel". Then we take the weighted average of the big five personality traits of Matz et al.'s categories by considering their absolute values as weights. Equation (1) that represents this translation process calculates the big five personality scores of our *i*'th spending category (denoted by A) by using one or more (up to *n*) categories of Matz et al. (denoted by M and indexed by *j*). The left side of the equation is the big five personality trait value we calculate: 'O' for openness, 'C' for conscientiousness, 'E' for extraversion, 'A' for agreeableness and 'N' for neuroticism. The right side of the equation calculates the combined scores of Matz et al. Big Five Personality Traits:

- O: Openness
- C: Conscientiousness
- E: Extraversion
- A: Agreeableness
- N: Neuroticism

Ownership of category:

- A: Our categories
- M: Matz et al.'s categories

Indices:

- i: Our i'th category
- j: Matz et al.'s j'th categories
- n: Number of correspondent categories from Matz et al.'s study to our spending categories

A numerical example of this translation process is given below as well as in Figure 6. The openness (O) value of our i'th category, which is "Travel Agencies-Transportation", is calculated by using the openness scores of $n = 3$ categories (indexed as j) from Matz et al., which are "Day out and tourism", "Foreign Travel" and "Travel", and is expressed as:

$$O_i^A = \frac{\sum_{j=1}^{n} O_j^M |O_j^M|}{\sum_{j=1}^{n} |O_j^M|}$$

	A	B	C	D	E	F	G
1	**Our Spending Category**	**Matz's Spending Categories**	**Openness**	**Conscientiousness**	**Extraversion**	**Agreeableness**	**Neuroticism**
2	TRAVEL AGENCIES-TRANSPORTATION	Days out and tourism	2,19	0,57	2,25	1,10	-0,28
3	TRAVEL AGENCIES-TRANSPORTATION	Foreign travel	2,54	0,65	2,15	0,85	-0,11
4	TRAVEL AGENCIES-TRANSPORTATION	Travel	2,51	0,24	2,37	1,18	-0,20
5		C6	**=(C2*ABS(C2) +C3* ABS(C3) +C4* ABS(C4))/(ABS(C2) + ABS(C3) + ABS(C4))**				
6		TRAVEL AGENCIES-TRANSPORTATION	2,42	0,55	2,26	1,06	-0,22

Figure 7. Excel Spreadsheet Example (Values are from Matz et al.'s research)

The remaining personality traits of the same category as well as other categories are calculated using the same approach.

We combined the processes of spending category matching results, selection of effective Matz et al.'s spending categories' big five factor scores in four ways and merging Matz et al.'s spending categories' big five factor scores into our spending categories' big five factor scores. This combination demonstrated that we need further filtering among our spending categories due to following reasons:

- Some of our categories do not have values of big five personality scores due to not selecting a spending category from Matz et al.'s study according to our four-way selection approach (e.g. Shopping Malls).
- Some of our categories have different values of big five personality scores due to selecting different spending categories of Matz et al.'s study according to our four-way selection approach (e.g. Restaurant).

To summarize, we calculate big five factor scores of our spending categories on each selection criteria based on formulation above by using matched categories of Matz et al.'s study. The results of this calculations can be found in Table 6 below:

Table 6
The Big Five Factor Scores of Our Spending Categories for each Selection Approach for Matz et al.'s Spending Categories

Our Spending Categories	**Selection Approaches**	**O**	**C**	**E**	**A**	**N**
Shopping Malls	(-1,1) v (mean +/-1 sd)					
	(-1,1) v (mean +/- 0.5 sd)					
	(-0.5,0.5) v (mean +/-1 sd)					
	(-0.5, 0.5) v (mean +/- 0.5 sd)	-0.69	1.27	0.51	0.58	-0.73
Car Rentals	(-1,1) v (mean +/-1 sd)					
	(-1,1) v (mean +/- 0.5 sd)	-0.53	1.39	-0.06	0.31	-0.96
	(-0.5,0.5) v (mean +/-1 sd)					
	(-0.5, 0.5) v (mean +/- 0.5 sd)	-0.53	1.39	-0.06	0.31	-0.96
Fun and Sports	(-1,1) v (mean +/-1 sd)	2.24	0.87	2.38	-0.10	0.66
	(-1,1) v (mean +/- 0.5 sd)	2.24	0.87	2.38	-0.10	0.66
	(-0.5,0.5) v (mean +/-1 sd)	2.24	0.87	2.38	-0.10	0.66
	(-0.5, 0.5) v (mean +/- 0.5 sd)	2.24	0.87	2.38	-0.10	0.66
Food	(-1,1) v (mean +/-1 sd)					
	(-1,1) v (mean +/- 0.5 sd)	1.45	1.59	0.86	1.41	-0.80
	(-0.5,0.5) v (mean +/-1 sd)					
	(-0.5, 0.5) v (mean +/- 0.5 sd)	1.45	1.59	0.86	1.41	-0.80
Hotel	(-1,1) v (mean +/-1 sd)	-0.16	1.69	0.31	1.55	-1.63
	(-1,1) v (mean +/- 0.5 sd)	-0.16	1.69	0.31	1.55	-1.63
	(-0.5,0.5) v (mean +/-1 sd)	-0.16	1.69	0.31	1.55	-1.63
	(-0.5, 0.5) v (mean +/- 0.5 sd)	-0.16	1.69	0.31	1.55	-1.63
Pubs and Casinos	(-1,1) v (mean +/-1 sd)	1.46	-1.81	2.28	-1.41	1.69
	(-1,1) v (mean +/- 0.5 sd)	1.46	-1.81	2.28	-1.41	1.69
	(-0.5,0.5) v (mean +/-1 sd)	1.46	-1.81	2.28	-1.41	1.69
	(-0.5, 0.5) v (mean +/- 0.5 sd)	1.46	-1.81	2.28	-1.41	1.69
Cosmetic	(-1,1) v (mean +/-1 sd)					
	(-1,1) v (mean +/- 0.5 sd)	1.91	0.31	1.49	0.85	0.22
	(-0.5,0.5) v (mean +/-1 sd)					
	(-0.5, 0.5) v (mean +/- 0.5 sd)	1.91	0.31	1.49	0.85	0.22
Jewelry	(-1,1) v (mean +/-1 sd)					
	(-1,1) v (mean +/- 0.5 sd)	1.60	0.73	1.43	0.96	-0.61
	(-0.5,0.5) v (mean +/-1 sd)					
	(-0.5, 0.5) v (mean +/- 0.5 sd)	1.60	0.73	1.43	0.96	-0.61

Decoration	(-1,1) v (mean +/-1 sd)	0.63	1.48	0.17	1.38	-1.22
	(-1,1) v (mean +/- 0.5 sd)	0.63	1.48	0.17	1.38	-1.22
	(-0.5,0.5) v (mean +/-1 sd)	0.63	1.48	0.17	1.38	-1.22
	(-0.5, 0.5) v (mean +/- 0.5 sd)	0.63	1.48	0.17	1.38	-1.22
Motorcycle	(-1,1) v (mean +/-1 sd)	1.34	0.09	2.32	-0.55	0.82
	(-1,1) v (mean +/- 0.5 sd)	1.34	0.09	2.32	-0.55	0.82
	(-0.5,0.5) v (mean +/-1 sd)	1.34	0.09	2.32	-0.55	0.82
	(-0.5, 0.5) v (mean +/- 0.5 sd)	1.34	0.09	2.32	-0.55	0.82
Music - Market - Stationary	(-1,1) v (mean +/-1 sd)	2.25	1.81	1.51	1.31	-1.24
	(-1,1) v (mean +/- 0.5 sd)	2.25	1.81	1.51	1.31	-1.24
	(-0.5,0.5) v (mean +/-1 sd)	2.25	1.81	1.51	1.31	-1.24
	(-0.5, 0.5) v (mean +/- 0.5 sd)	2.25	1.81	1.51	1.31	-1.24
Toys	(-1,1) v (mean +/-1 sd)	2.19	-0.90	1.94	0.78	-0.06
	(-1,1) v (mean +/- 0.5 sd)	2.19	-0.90	1.94	0.78	-0.06
	(-0.5,0.5) v (mean +/-1 sd)	2.19	-0.90	1.94	0.78	-0.06
	(-0.5, 0.5) v (mean +/- 0.5 sd)	2.19	-0.90	1.94	0.78	-0.06
Restaurant	(-1,1) v (mean +/-1 sd)	0.89	1.24	0.45	1.79	-1.23
	(-1,1) v (mean +/- 0.5 sd)	1.32	1.03	1.47	1.49	-1.03
	(-0.5,0.5) v (mean +/-1 sd)	0.89	1.24	0.45	1.79	-1.23
	(-0.5, 0.5) v (mean +/- 0.5 sd)	1.32	1.03	1.47	1.49	-1.03
Health	(-1,1) v (mean +/-1 sd)	-0.93	2.03	0.70	0.84	-0.80
	(-1,1) v (mean +/- 0.5 sd)	-0.93	2.03	0.70	0.84	-0.80
	(-0.5,0.5) v (mean +/-1 sd)	-0.93	2.03	0.70	0.84	-0.80
	(-0.5, 0.5) v (mean +/- 0.5 sd)	-0.93	2.03	0.70	0.84	-0.80
Travel Agencies - Transportation	(-1,1) v (mean +/-1 sd)	2.42	0.55	2.26	1.06	-0.22
	(-1,1) v (mean +/- 0.5 sd)	2.42	0.55	2.26	1.06	-0.22
	(-0.5,0.5) v (mean +/-1 sd)	2.42	0.55	2.26	1.06	-0.22
	(-0.5, 0.5) v (mean +/- 0.5 sd)	2.42	0.55	2.26	1.06	-0.22
Insurance	(-1,1) v (mean +/-1 sd)	-1.71	2.11	-1.23	0.82	-1.24
	(-1,1) v (mean +/- 0.5 sd)	-1.71	2.11	-1.23	0.82	-1.24
	(-0.5,0.5) v (mean +/-1 sd)	-1.71	2.11	-1.23	0.82	-1.24
	(-0.5, 0.5) v (mean +/- 0.5 sd)	-1.71	2.11	-1.23	0.82	-1.24
Cinema - Theatre - Art	(-1,1) v (mean +/-1 sd)	2.41	0.21	1.49	1.42	-0.44
	(-1,1) v (mean +/- 0.5 sd)	2.41	0.21	1.49	1.42	-0.44
	(-0.5,0.5) v (mean +/-1 sd)	2.41	0.21	1.49	1.42	-0.44
	(-0.5, 0.5) v (mean +/- 0.5 sd)	2.41	0.21	1.49	1.42	-0.44
Technology	(-1,1) v (mean +/-1 sd)	1.49	1.72	0.76	0.93	-1.02
	(-1,1) v (mean +/- 0.5 sd)	1.38	1.51	1.02	0.75	-0.88
	(-0.5,0.5) v (mean +/-1 sd)	1.49	1.72	0.76	0.93	-1.02
	(-0.5, 0.5) v (mean +/- 0.5 sd)	1.38	1.51	1.02	0.75	-0.88

Textile	(-1,1) v (mean +/-1 sd) (-1,1) v (mean +/- 0.5 sd)	-0.28	0.43	0.00	1.16	-0.96
	(-0.5,0.5) v (mean +/-1 sd) (-0.5, 0.5) v (mean +/- 0.5 sd)	-0.28	0.43	0.00	1.16	-0.96
Telecommunication	(-1,1) v (mean +/-1 sd) (-1,1) v (mean +/- 0.5 sd)	0.48	0.00	1.29	-0.17	0.14
	(-0.5,0.5) v (mean +/-1 sd) (-0.5, 0.5) v (mean +/- 0.5 sd)	0.48	0.00	1.29	-0.17	0.14
Ironmongery (Hardware Store)	(-1,1) v (mean +/-1 sd)	1.88	1.58	0.47	1.62	-1.32
	(-1,1) v (mean +/- 0.5 sd)	1.88	1.58	0.47	1.62	-1.32
	(-0.5,0.5) v (mean +/-1 sd)	1.88	1.58	0.47	1.62	-1.32
	(-0.5, 0.5) v (mean +/- 0.5 sd)	1.88	1.58	0.47	1.62	-1.32

O: Openness – C: Conscientiousness – A: Agreeableness N: Neuroticism –E: Extraversion

This translation process leads us to a final set of 12 spending categories by considering the filtering reasons stated above. This dataset filtering based on these 12 categories leads us to removing the customers who do not have any spending from these categories. This process results in 80,250 unique customers with 911,280 credit card transactions. Table 7 reflects the frequency, total amount, and average amount of remaining categories (amounts are in Turkish Lira). The resulting big five personality scores calculated per spending category are provided in Table 8.

Table 7
Frequency, Total Amount and Average Amount of Spending Categories of Remained Dataset

Spending Categories	**Frequency**	**Total Spending Amount (TL)**	**Average Spending Amount (TL)**
Fun and Sports	35,677	4,950,835	138.77
Hotels	55,633	23,448,988	421.49
Pubs and Casinos	19,368	2,822,597	145.74
Decoration	125,634	41,364,031	329.24
Motorcycle	1,836	923,183	502.82
Music-Market-Stationary	60,060	4,804,215	79.99
Toys	35,983	3,725,555	103.54
Health	318,789	36,670,361	115.03
Travel Agencies-Transportation	101,372	34,418,481	339.53
Insurance	46,887	16,931,095	361.10
Cinema-Theatre-Art	34,539	1,635,559	47.35
Ironmongery (Hardware Store)	75,502	328,796	4.35
Total	911,280	172,023,696	2584.6

In addition, table of big five personality scores calculated per spending categories is below:

Table 8
Big Five Personality Values Per Spending Category

Spending Categories	**O**	**C**	**A**	**N**	**E**
Fun and Sports	2.24	0.87	-0.10	0.66	2.38
Hotels	-0.16	1.69	1.55	-1.63	0.31
Pubs and Casinos	1.46	-1.81	-1.41	1.69	2.28
Decoration	0.63	1.48	1.38	-1.22	0.17
Motorcycle	1.34	0.09	-0.55	0.82	2.32
Music-Market-Stationary	2.25	1.81	1.31	-1.24	1.51
Toys	2.19	-0.90	0.78	-0.06	1.94
Health	-0.93	2.03	0.84	-0.80	0.70
Travel Agencies-Transportation	2.42	0.55	1.06	-0.22	2.26
Insurance	-1.71	2.11	0.82	-1.24	-1.23
Cinema-Theatre-Art	2.41	0.21	1.42	-0.44	1.49
Ironmongery (Hardware)	1.88	1.58	1.62	-1.32	0.47

O: Openness – C: Conscientiousness – A: Agreeableness N: Neuroticism – E: Extraversion

3.2.2. Data Manipulation and Cleaning

Data manipulation phase links the big five personality scores of spending categories to big five personality scores of customers by simply aggregating the contributions of each transaction in a month to form the personal monthly big five personality scores. We calculate monthly big five personality scores of customers in two steps. First, we multiply each big five personality score of a spending category with the ratio of that category's amount to the total monthly spending of all 12 categories. Then we add these scores to aggregate them by month, resulting in a monthly big five personality score for individuals. We further filter the customers who have at least six months of scores calculated for the first 11 months to predict the probability of payment in the 12th month. This filtering has resulted in 22401 customers. The formulation of monthly big five personality scores is explained below:

Transaction Amounts:

- T: Transaction amount
- S: Sum of monthly transaction amount for selected categories.

Indices:

- i: Customer index
- j: Month index
- t: Transaction index
- k: Number of transactions for a i'th customer in j'th month.

As an example, the formulation of monthly openness scores is given in Equation (2), where O_i^j is the openness score of customer *i* in month *j* and is equal to the openness score O^t of category *t* multiplied by the ratio of customer *i*'s spending T_i^{jt} in category *t* in month *j* to his/her total spending S_i^j in month *j*, summed over all categories $t = 1, \dots, k$. The monthly score for the remaining big five traits for each individual is calculated similarly.

$$O_i^j = \sum_{t=1}^{k} \frac{T_i^t}{S_i^j} O^t$$

3.2.3. Target Variables

We also extract our dependent variables, or labels, by processing the credit card statement and payment data tables as well as by comparing the payment date and amounts with the matching statement due dates and amounts. We produce labels for four types of payments, which are:

- The minimum due amount is paid or not by the due date (i.e. without any grace period)
- The minimum due amount is paid or not within 3 days after the due date (i.e. within a 3-day grace period);
- The total due amount is paid or not by the due date
- The total due amount is paid or not within 3 days after the due date.

Distributions of our four types of target variables per month is presented in Table 9 below.

Table 9

Distributions of Target Variables per Year and Payment Behavior Type

	On-time payment of the minimum amount due		Payment before grace period of 3 days of the minimum amount due	
Predicted Month	**Percentage of Customers Paid**	**Percentage of Customers Did Not Pay**	**Percentage of Customers Paid**	**Percentage of Customers Did Not Pay**
1	82%	18%	90%	10%
2	78%	22%	89%	11%
3	81%	19%	90%	10%
4	73%	27%	88%	12%
5	84%	16%	91%	9%
6	80%	20%	90%	10%
7	75%	25%	88%	12%
8	88%	12%	91%	9%
9	83%	17%	91%	9%
10	78%	22%	90%	10%
11	79%	21%	89%	11%
12	82%	18%	91%	9%
	Payment of full amount due without delay		**Payment before grace period of 3 days of the full amount due**	
Predicted Month	**Percentage of Customers Paid**	**Percentage of Customers Did Not Pay**	**Percentage of Customers Paid**	**Percentage of Customers Did Not Pay**
1	51%	49%	57%	43%
2	47%	53%	55%	45%
3	49%	51%	55%	45%
4	45%	55%	56%	44%
5	50%	50%	56%	44%
6	48%	52%	55%	45%
7	44%	56%	54%	46%
8	53%	47%	55%	45%
9	50%	50%	55%	45%
10	51%	49%	59%	41%
11	47%	53%	55%	45%
12	48%	52%	53%	47%

Numbers of available customers by months are 15103, 15816, 15466, 16623, 16886, 17567, 17347, 16372, 15029, 16689, 16428, 10172.

CHAPTER 4

PREDICTIVE ANALYTICS METHODOLOGY

At this stage, our goal is to determine or predict customers who are possibly in financial trouble, where financial trouble is signaled by the four labels described above. We choose to train main model for predicting the 12^{th} month's payment behavior, using all available information of customers from the first 11 months. That is, we use the monthly big five scores and demographic variables for the first 11 months as inputs for predicting the last month's payment behavior. We also calculate and use as input the linear trends of each big five personality score over 12 months and the most recent 6 months to account for possible score changes throughout the year.

Before we train our final model, we exclude the customers who did not have any statement amount due in the 12^{th} month as these customers do not have any labels calculated. This filtering results in a final set of 10,172 customers. In our methodology, we use several machine learning algorithms to predict our four labels; and evaluate their performance with widely used classification metrics and validation methods from the big data literature (cf. Bleidorn & Hopwood, 2018; see below for details). Thus, our methodology has five components which are the big five personality scores as input and payment behavior indicators as dependent variables, algorithms, validation methods, and classification metrics.

The machine learning algorithms we use include logistic regression, decision tree, linear discriminant analysis, quadratic discriminant analysis (QDA), naïve-Bayes classifier, support vector machines, and random forest considering the performance metrics findings (James, Witten, Hastie, & Tibshirani, 2017).

The validation methods we use are the train-test set split approach and a 30-fold cross-validation. We first set aside a random 30% percent of the data as the test set for our models. Thus, we use the dataset of 7,121 customers to train our model for prediction of payment behavior using the big five personality scores and trends as predictors. We test our trained model on a dataset of 3,051 customers and compare these results with the actual payment behavior indicators. We also use 30-fold cross validation during the model training process each time we use a different algorithm. Here, the dataset is randomly divided into 30 equal partitions for model validation purposes (James et al., 2017). This method works similarly to the train-test split method, but it uses 29 partitions of the train set to train the model and validate it by evaluating with the last partition. This process is done 30 times and the best model out of 30 models is selected as a result of the training process. The selected model is tested on the test set of 3,051 customers for the prediction of their payment behavior.

There are several metrics to evaluate the performances of our classification models. The most common is *accuracy* (James et al., 2017). Different metrics have different aspects to evaluate true positives, true negatives, false positives, and false negatives. Hence, different classification metrics are good at reflecting and comparing different types of distributions of these elements of true positives and true negatives. There is a noticeable difference between the distributions of different payment labels (e.g., there are more customers who pay the minimum due of the statement within the grace period than there are customers who pay the full amount on time). We have selected a metric to evaluate the performances due to different and imbalanced payment label distributions of customers. The evaluation and selection of our models is based on the area under the receiving operating characteristic curve (AUROC) (Fogarty, Baker, & Hudson, 2005). We choose AUROC because it balances the effect of such differences by handling false positive and negatives. In evaluating AUROC values of prediction models, values greater than 50% and closer to 100% indicate increasingly better performance of a model.

We replicate the process above for a number of alternatives with three (input dataset, machine learning algorithms and target variables per month) of the five components of our methodology, then compare these alternatives and set benchmarks for our main model. Further paragraphs describe these folds below starting from the second fold. Our predictive analytics fold is listed below:

- Main Model
- Monthly Models with Only Demographic Information as Predictors
- Monthly Models with Correspondent Monthly Big Five Personality Scores and Demographic Information as Predictors
- Main Model with also Demographic Information as Predictors
- Main Models with Lookback Periods
- Main Models with Step-back Dependent Variables

After our main model, we start by using only the demographic information of customers to predict the four labels of payment using all twelve months' data. We manipulate the demographic information into one-hot encoded format. It results in separate columns for each level of categorical variables which are gender, education and marital status variables. We also manipulate numeric variables which are customer age and banking age, into categorical variables to represent the age intervals of the customers. These categorical age interval variables are manipulated into one-hot encoded format as well. This pipeline produces slightly better AUROC results to predict the targets represent paying the full amount of statements. Its reason might be proposed as the contribution of one-hot encoding and using those variables only which all of them have the same logic and range (0.1).

We repeat the same pipeline above, but with one difference: adding the big five scores of the corresponding months. In this pipeline we both use one-hot encoded demographic information and correspondent monthly big five personality scores of customers to predict payment behaviors on each month.

We also add one-hot encoded demographic information to our main model to see their effects when they are combined with our features related with big five personality factors. Their contribution affects negatively our models in terms of predictive power in our structure according to AUROC values we calculated as Table 16 suggests.

Then we shorten the lookback period, which by default was eleven months for our main model, considering periods of 1, 2, 3, 4, 5, and 6 months. We use the seven machine learning algorithms stated above while replicating the process. Then we chose the best algorithm and lookback period to continue our analysis, which are the naïve-Bayes classifier and the four-month lookback period. Their results are generally worse than the previous pipelines as it can be presumed due to decreasing the number of independent variables.

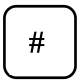
We train models to predict 11th, 10th, 9th, 8th, 7th, 6th and 5th month's payment labels by using the naïve-Bayes classifier and the preceding four-month input data.

We also train five more models by excluding one type of the big five personality scores at a time to evaluate the total contribution of each personality trait. We endeavor this method since excluding variable importance produce monthly big five personality scores individually and it may result in different directions for the same big five personality factor from different month.

4.1. Illustrated Steps of Predictive Modelling

We depict our predictive analytics approach below per step, starting with the description of icons:


Demographic information of customers as inputs

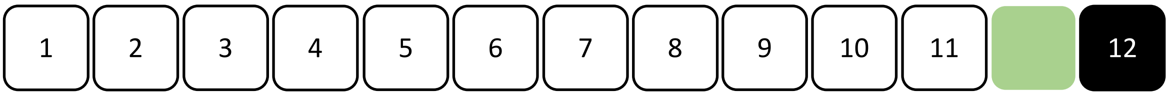


Big five personality scores of correspondent months as inputs

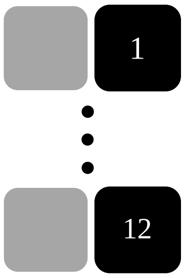

Trends of big five personality scores as inputs

Payment behaviors of correspondent months as targets to predict

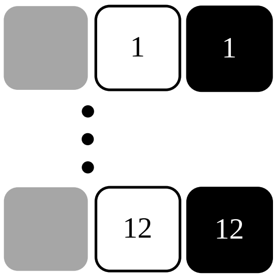
Thus, our main model can be depicted as follows:



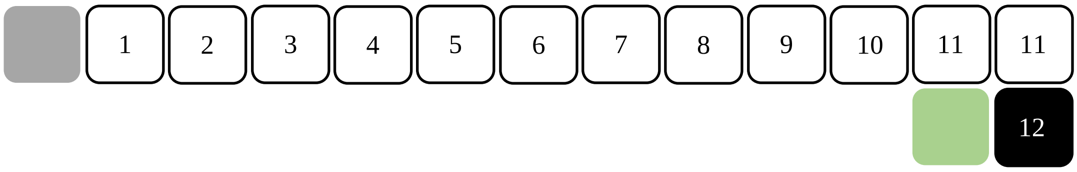
Models that were built having only demographic information as inputs to predict targets of each month can be depicted as below:



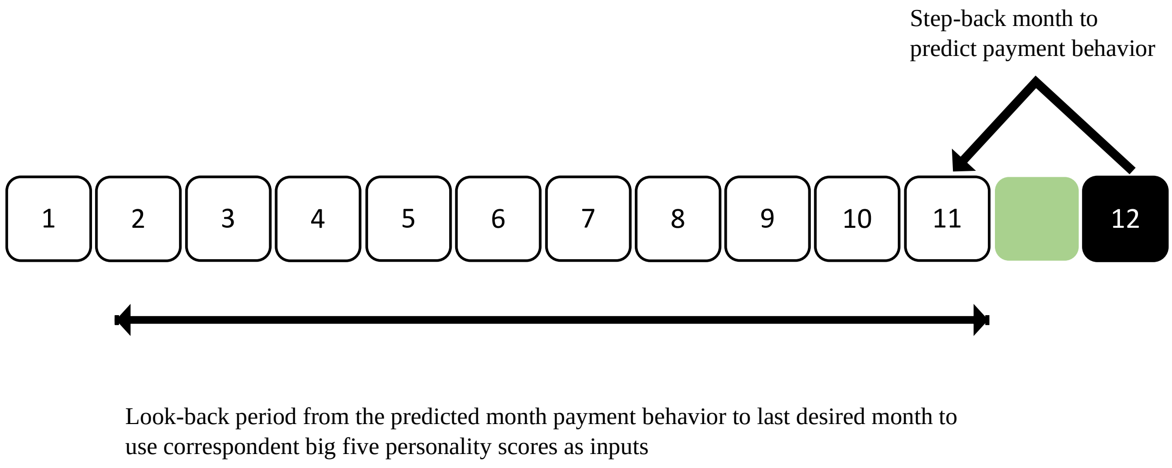
Models that were built having demographic information with correspondent monthly big five personality scores as inputs to predict targets of each month can be depicted as:



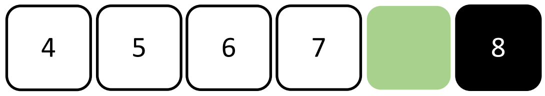
The model that is formed when demographic information is attached to our main model can be depicted as follows:



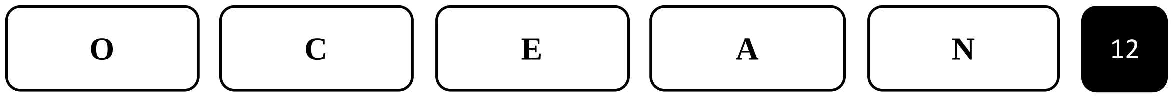
The shortened lookback period and predicting the 12th month to the prior months to by step-back months can be illustrated with our main model as below:



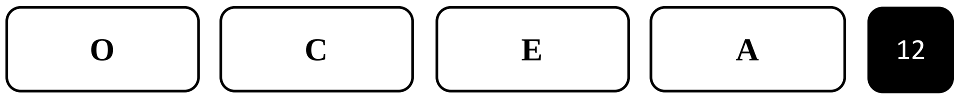
Hence, we can depict the model in Table 22 for the behavior of completely paying the bill on the 8th month with the 4-month lookback period by using naïve-Bayes algorithm:



Since we use/apply the big five personality scores monthly, we developed 5 more models by removing all monthly scores and trends of a big five personality factor to assess the contributions of each of the big five personality traits as whole. This process can be depicted on our main model as below:



Removing the neuroticism personality trait would be seen as following:



The same operation was conducted for each of the remaining big five personality traits.

O	**C**	**E**	**N**	12
O	**C**	**A**	**N**	12
O	**E**	**A**	**N**	12
C	**E**	**A**	**N**	12

CHAPTER 5

COMPUTATIONAL RESULTS

Our results indicate noticeable correlations between big five personality traits and on-time (no grace period) payment behavior. Table 10 provides the results of main model combinations of target variables and machine learning algorithms.

Table 10
List of Main Model Combinations with Four Types of Dependent Variable and Machine Learning Algorithms

Dependent Variable	**Machine Learning Algorithm**	**Accuracy**	**AUC**
On-time payment of the minimum amount due	Decision Tree	81%	50.00%
On-time payment of the minimum amount due	Logistic Regression	81%	50.00%
On-time payment of the minimum amount due	QDA	76%	51.12%
On-time payment of the minimum amount due	SVM	81%	50.00%
On-time payment of the minimum amount due	Random Forest	81%	49.98%
Payment of full amount due without delay	Decision Tree	54%	53.22%
Payment of full amount due without delay	Logistic Regression	55%	54.40%
Payment of full amount due without delay	QDA	53%	53.37%
Payment of full amount due without delay	SVM	53%	52.89%

Payment of full amount due without delay	Random Forest	55%	53.43%
Payment before grace period of 3 days of minimum amount due	Decision Tree	90%	50.00%
Payment before grace period of 3 days of minimum amount due	Logistic Regression	90%	50.00%
Payment before grace period of 3 days of minimum amount due	QDA	86%	50.14%
Payment before grace period of 3 days of minimum amount due	SVM	90%	50.00%
Payment before grace period of 3 days of minimum amount due	Random Forest	90%	49.98%
Payment before grace period of 3 days of the full amount due	Decision Tree	52%	50.75%
Payment before grace period of 3 days of the full amount due	Logistic Regression	54%	52.80%
Payment before grace period of 3 days of the full amount due	QDA	53%	53.04%
Payment before grace period of 3 days of the full amount due	SVM	53%	52.67%
Payment before grace period of 3 days of the full amount due	Random Forest	55%	53.75%

Table 11 summarizes the best algorithms for our main approach for which we have included all big five personality scores for the prior 11 months and their trends as predictors to predict the 12^{th} month payment labels (i.e., dependent variables), and their corresponding AUROC performances.

Table 11
The Best Performing Machine Learning Algorithm and AUROC Values for Each Dependent Variable

Dependent Variable	**Algorithm**	**AUROC**
Payment of full amount due without delay	Logistic Regression	54.4
Payment before grace period of 3 days of the full amount due	Random Forest	53.75
On-time payment of the minimum amount due	QDA	51.1
Payment before grace period of 3 days of minimum amount due	QDA	50.1

AUC curves of the first two main models are provided below.

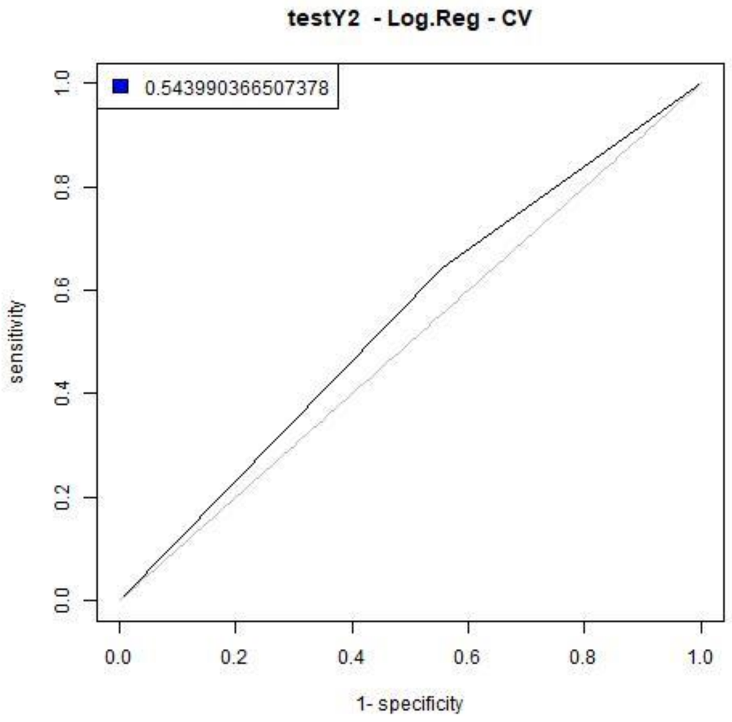


Figure 8: AUC Curve of the 1st model in Table 11

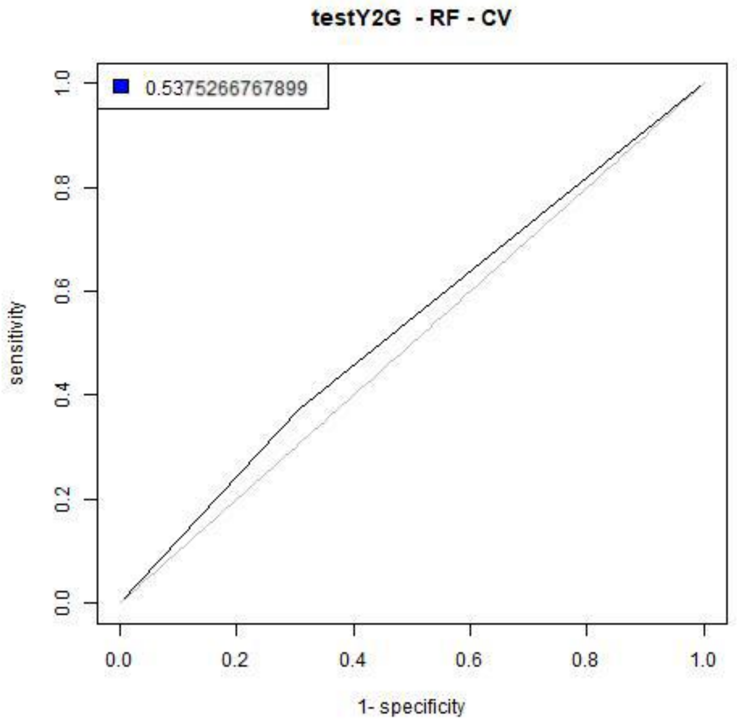


Figure 9: AUC Curve of the 2nd model in Table 11

This finding indicates that the best performing machine learning algorithm to predict full on-time payment of a credit card statement in month 12 was logistic regression with 54.40% AUROC value, which means that for this case our model performed 8.8% better than random guessing or classifying the payment labels.

Table 12 provides the summary of the best results of models using demographic information only as independent variables to predict four types of target variables by using our selected machine learning algorithms. The prediction months for which the best performance is achieved are also shown in Table 12:

Table 12
The Best Performing Machine Learning Algorithm, Predicted Month and AUROC Values for Each Dependent Variable

Dependent Variable	**Algorithm**	**Month**	**AUROC**
Payment of full amount due without delay	Logistic Regression	3	56.2%
Payment before grace period of 3 days of the full amount due	LDA	12	57.9%
On-time payment of the minimum amount due	LDA	2	50.2%
Payment before grace period of 3 days of the minimum amount due	Logistic Regression	1	50.0%

Similarly, the model that predicts full on-time payment best is the one with logistic regression and for the 3rd month with a performance 12.4% better than a random guess or classification.

Table 14 provides the summary of the results of models that belong to same data science pipeline dealing with demographic information but also add the big five personality scores of the corresponding month to the set of predictor variables.

Table 14
The Best Performing Machine Learning Algorithm, Predicted Month, and AUROC Values for Each Dependent Variable

Dependent Variable	**Algorithm**	**Month**	**AUROC**
Payment of full amount due without delay	Decision Tree	8	54.33%
Payment before grace period of 3 days of the full amount due	Decision Tree	2	54.34%
On-time payment of the minimum amount due	Logistic Regression	6	50.07%
Payment before grace period of 3 days of the minimum amount due	Logistic Regression	6	50.09%

Table 15 presents the results when we add the demographic information of customers to our main model with the big five personality scores for 11 months and their trends as predictors to predict payment labels in month 12.

Table 15
List of Demographic Information added to Our Main Models Combinations with Four Types of Dependent Variable and Machine Learning Algorithms

Dependent Variable	**Algorithm**	**Accuracy**	**AUC**
On-time payment of the minimum amount due	Logistic Regression	81%	49.91%
On-time payment of the minimum amount due	Decision Tree	81%	50.00%
On-time payment of the minimum amount due	SVM	81%	50.00%
On-time payment of the minimum amount due	Random Forest	81%	50.00%
Payment of full amount due without delay	Logistic Regression	54%	50.58%
Payment of full amount due without delay	Decision Tree	53%	50.28%
Payment of full amount due without delay	Random Forest	51%	50.13%
Payment of full amount due without delay	SVM	51%	50.04%
Payment before grace period of 3 days of the minimum amount due	Logistic Regression	90%	49.98%
Payment before grace period of 3 days of the minimum amount due	Decision Tree	90%	50.00%
Payment before grace period of 3 days of the minimum amount due	SVM	90%	50.00%
Payment before grace period of 3 days of the minimum amount due	Random Forest	90%	50.00%
Payment before grace period of 3 days of the full amount due	Logistic Regression	51%	49.92%
Payment before grace period of 3 days of the full amount due	Decision Tree	52%	50.00%
Payment before grace period of 3 days of the full amount due	SVM	51%	50.18%
Payment before grace period of 3 days of the full amount due	Random Forest	52%	50.33%

The best results of Table 15 are summarized in Table 16 below.

Table 16
The Best Performing Machine Learning Algorithm, and AUROC values for Each Dependent Variable

Dependent Variable	**Algorithm**	**AUROC**
Payment of full amount due without delay	Logistic Regression	50.6%
Payment before grace period of 3 days of the full amount due	Random Forest	50.3%
On-time payment of the minimum amount due	Decision Tree	50.0%
Payment before grace period of 3 days of the minimum amount due	Decision Tree	50%

The results in Table 16 suggest that logistic regression is the best model to predict the full on-time payment label of month 12 with a performance only 1.2% better than random guess or classification.

As we stated in the Methodology section, we have built models for the prediction of the 12th month's payment behavior with look-back periods of 1, 2, 3, 4, 5, and 6 months. Table 17 demonstrates the best results regarding this stage of our prediction analytics methodology:

Table 17
Results belong to The Best Performing Models with Different Lookback Periods and Machine Learning Algorithms

Dependent Variable	**Algorithm**	**Lookback Period**	**Accuracy**	**AUC**
Payment of full amount due without delay	Logistic Regression	6	55%	51.8%
Payment before grace period of 3 days of the full amount due	Naïve-Bayes	4	52%	51.9%
On-time payment of the minimum amount due	Naïve-Bayes	6	79%	50.9%
Payment before grace period of 3 days of the minimum amount due	QDA	4	76%	51.7%

We make our selection of amongst look-back periods of 1 to 6 months and various algorithms based on the comparison of their AUROC values through the t-test. While, table 19 presents the t-test results for machine learning algorithm selection, table 18 presents the t-test results is performed by comparing their correspondent AUROC values for lookback period selection. The first t-test above presented in Table 18 is performed by comparing the AUROC values of different models with respect to their lookback period only (i.e. by not separating them in terms of their correspondent machine learning algorithm). Vice-versa for the second t-test above presented in Table 19.

Table 18
T-Test Results for Lookback Period Selection

Selected LookBack	**Comparing LookBack Periods Against Selected Lookback Period**				
LB1	**LB2**	**LB3**	**LB4**	**LB5**	**LB6**
p-value	0.0078	0.3637	7.77E-08	0.04414	1.87E-06
LB2	**LB3**	**LB4**	**LB5**	**LB6**	
p-value	0.01487	3.37E-16	0.0012	3.7E-06	
LB3	**LB4**	**LB5**	**LB6**		
p-value	4.94E-10	0.0741	0.0002		
LB4	**LB5**	**LB6**			
p-value	8.32E-05	0.464			
LB5	**LB6**				
p-value	0.0004				

Table 19
T-Test Results for Machine Learning Algorithm Selection

Selected Algorithm	**Comparing Algorithms Against Selected Algorithm**					
Decision Tree	**LDA**	**Logistic Regression**	**Naïve-Bayes**	**QDA**	**Random Forest**	**SVM**
p-value	2.27E-11	2.70E-11	1.15E-14	7.7E-20	2.4E-05	1.3E-02
LDA	**Logistic Regression**	**Naïve-Bayes**	**QDA**	**Random Forest**	**SVM**	
p-value	3.46E-01	9.28E-03	4.50E-03	1.4E-11	9.8E-11	
Logistic Regression	**Naïve-Bayes**	**QDA**	**Random Forest**	**SVM**		
p-value	8.89E-03	4.40E-03	1.94E-11	7.4E-11		
Naive-Bayes	**QDA**	**Random Forest**	**SVM**			
p-value	4.09E-01	1.80E-11	9.96E-10			
QDA	**Random Forest**	**SVM**				
p-value	1.65E-20	3.32E-17				
Random Forest	**SVM**					
p-value	9.11E-07					

Furthermore, 4 months of lookback period and the Naïve-Bayes algorithm have produced significantly better AUROC values than do other configurations. Although there is not significant difference between AUROC values related with 4 months and 6 months of lookback period, analysis is continued with 4 months of lookback period since it has the highest value among averages and maximum AUROC values. In addition, although there is not significant difference between AUROC values related with naïve-Bayes and QDA, analysis is continued with naïve-Bayes algorithm since it has the highest value among averages and maximum AUROC values. Table 20 depicts the best results using naïve-Bayes algorithm for predicting payment labels in month 12 with 4 months of the look-back period:

Table 20
AUROC Values for Each Dependent Variable When Using the Naive-Bayes Algorithm and 4-month look-back period

Dependent Variable	AUROC
Payment of full amount due without delay	51.4
Payment before grace period of 3 days of the full amount due	52.6
On-time payment of the minimum amount due	50.7
Payment before grace period of 3 days of the minimum amount due	50.0

Considering the best performance reported in Table 20, the Naïve-Bayes algorithm with the 4-month look-back period predict the 12^{th} month's payment of the full amount within the 3 days grace period with 52.6% AUROC, a result which, is 5.2% better than a random guess or classification.

The final step of our method is using the four-month look-back period and naïve-Bayes algorithm to predict customer payment behavior on months11, 10, 9, 8, 7, 6, and 5. Table 21 lists the results of this data science pipeline.

Table 21
Summary of the Results belong to Models with Different (Step-back) Months to Predict with 4-months Lookback Period by using Naïve-Bayes Algorithm

Dependent Variable	Predicted Month	Accuracy	AUC
On-time payment of the minimum amount due	5	83%	50.16%
Payment of full amount due without delay	5	49%	49.37%
Payment before grace period of 3 days of the minimum amount due	5	91%	49.94%
Payment before grace period of 3 days of the full amount due	5	50%	47.90%
On-time payment of the minimum amount due	6	78%	50.09%
Payment of full amount due without delay	6	50%	50.59%
Payment before grace period of 3 days of the minimum amount due	6	90%	50.00%
Payment before grace period of 3 days of the full amount due	6	53%	49.97%
On-time payment of the minimum amount due	7	75%	50.18%

Payment of full amount due without delay	7	55%	50.93%
Payment before grace period of 3 days of the minimum amount due	7	88%	49.98%
Payment before grace period of 3 days of the full amount due	7	53%	49.99%
On-time payment of the minimum amount due	8	88%	50.00%
Payment of full amount due without delay	8	53%	51.71%
Payment before grace period of 3 days of the minimum amount due	8	91%	50.00%
Payment before grace period of 3 days of the full amount due	8	53%	50.59%
On-time payment of the minimum amount due	9	83%	49.51%
Payment of full amount due without delay	9	49%	49.14%
Payment before grace period of 3 days of the minimum amount due	9	90%	50.13%
Payment before grace period of 3 days of the full amount due	9	53%	50.32%
On-time payment of the minimum amount due	10	77%	50.32%
Payment of full amount due without delay	10	49%	49.51%
Payment before grace period of 3 days of the minimum amount due	10	89%	50.04%
Payment before grace period of 3 days of the full amount due	10	57%	50.47%
On-time payment of the minimum amount due	11	78%	50.12%
Payment of full amount due without delay	11	51%	50.18%
Payment before grace period of 3 days of the minimum amount due	11	88%	49.63%
Payment before grace period of 3 days of the full amount due	11	53%	50.37%

Table 22 depicts the best results for each payment label with the best month of prediction:

Table 22
AUROC Values for Each Dependent Variable When Using the Naive-Bayes Algorithm and 4-month lookback period with the Best Predicted Step-Back Month

Dependent Variable	**Month**	**AUROC**
Payment of full amount due without delay	8	51.7
Payment before grace period of 3 days of the full amount due	8	50.6
On-time payment of the minimum amount due	10	50.3
Payment before grace period of 3 days of the minimum amount due	9	50.1

Again, the best performance listed in this table tells us that the Naïve-Bayes algorithm with the big five personality factors of the prior 4 months as input predictors can predict full payment within the 3 days grace period with 51.7% AUROC.

We build five more models by excluding one type of the big five personality scores at a time to evaluate the total contribution of each personality trait. However, we acquired unexpected results. These results indicate some of the big five personality scores deteriorate our models through increasing AUROC values when they are excluded presented in Appendix 4 with their correspondent dependent variables, machine learning algorithms, lookback periods, excluded big five personality factors, AUROC values before and after exclusion big five personality factors.

Moreover, we performed following analysis to extract the effects of independent variables on our predictions. We extract the logistic regression coefficients and apply partial dependency methods on random forest machine learning algorithm for the models stated in Table 11. In this analysis green highlighted values are positively related coefficients while red highlighted values are negatively related coefficients with our predictions.

Table 23
Logistic Regression Coefficients and Random Forest Partial Dependency Scores for the Models Described in Table 11

Big Five Factor	Predictor Month	Logistic Regression Coefficients	Random Forest Partial Dependency Score
(Intercept)	(Intercept)	0.8109	-
O	1	0.0344	0.0006
O	2	0.0015	-4.04E9
O	3	0.0600	-0.0003
O	4	0.0058	0.0003
O	5	-0.0431	-0.0007
O	6	-0.0231	0.0002
O	7	-0.0393	0.0013
O	8	0.1090	0.0015
O	9	-0.0356	0.0009
O	10	0.0336	-0.0007
O	11	-0.0037	0.0011
E	1	-0.0319	0.0002
E	2	-0.1307	-0.0032
E	3	-0.1872	-0.0018
E	4	-0.0508	-0.0023
E	5	0.0678	-0.0005
E	6	0.0508	0.0004
E	7	-0.0130	-0.0001
E	8	-0.2058	-0.0012
E	9	0.0657	-0.0002
E	10	0.0271	-0.0014
E	11	-0.0986	-0.0003
A	1	-0.0334	-0.0016
A	2	0.2452	-0.0010
A	3	0.1824	-0.0048
A	4	0.0875	-0.0046
A	5	-0.1337	-0.0030
A	6	-0.1732	-0.0031
A	7	0.0085	-0.0041
A	8	0.1397	-0.0006
A	9	-0.0491	-0.0021
A	10	-0.3125	-0.0036
A	11	0.2348	-0.0043
N	1	0.0412	0.0034
N	2	0.1829	0.0021
N	3	0.4798	0.0045
N	4	0.2453	0.0058

N	5	-0.0826	0.0037
N	6	-0.1167	0.0034
N	7	0.0514	0.0058
N	8	0.4489	0.0029
N	9	-0.1534	0.0023
N	10	-0.2961	0.0016
N	11	0.4200	0.0052
C	1	0.0501	-0.0020
C	2	-0.0768	-0.0039
C	3	0.1785	-0.0012
C	4	0.0134	-0.0018
C	5	-0.0256	-0.0017
C	6	-0.0334	-0.0027
C	7	-0.0879	-0.0072
C	8	0.1799	-0.0017
C	9	-0.1210	-0.0046
C	10	-0.0103	-0.0013
C	11	0.0608	-0.0016
O	12-month trend	-0.1067	-0.0032
O	6-month trend	-0.2314	-0.0092
E	12-month trend	1.0492	0.0135
E	6-month trend	-0.0401	0.0085
A	12-month trend	-0.8541	-0.0125
A	6-month trend	-0.0213	-0.0166
N	12-month trend	-2.1379	0.0019
N	6-month trend	0.6961	0.0022
C	12-month trend	-0.2156	0.0053
C	6-month trend	0.0600	0.0113

Logistic regression results in Table 23 show that monthly big five personality traits do not show a consistent pattern in terms of their direction of effect across months (e.g., openness has both positive and negative coefficients in different months), which prevents us from making a clear conclusion about their effects. Logistic regression analyses using trend variables, however, revealed that changes in openness and agreeableness are both negatively related to our dependent variables, which means that customers who experienced an increase in their agreeableness or openness levels throughout the year and the last six months are more likely to pay their credit card statements on time.

Random forest partial dependency scores in Table 23 show that agreeableness and conscientiousness are negatively associated with our dependent variables in all months. These results mean that high levels of agreeableness or conscientiousness increase the possibility that those customers will pay their statements without any delay. On the other hand, revealing the tendency of customers with high levels of neuroticism to not pay their statements on time, monthly scores of neuroticism are positively related with our dependent variables in all months. We also found that the trends of openness and agreeableness are negatively related, while trends of extraversion, neuroticism and conscientiousness are positively related with our dependent variables in the random forest analysis. Thus, increases in openness and agreeableness and decreases in extraversion, neuroticism and conscientiousness will have a positive impact on the possibility of paying on time.

CHAPTER 6

CONCLUSION

Our research brings together two research fields: (big) data science, personality psychology, to investigate financial well-being. In this paper, we have used a data science approach to predict customer payment behavior as a signal for financial well-being by using big five personality scores derived from spending categories.

Our empirical approach with predictive modeling can allow banks to identify customers who will potentially experience financial trouble before they do, by monitoring their big five personality scores and their trends on a monthly basis. Thus, this research can be considered as possessing a novel behavioral personality aspect to evaluate the financial well-being of bank customers in addition to their financial and behavioral models.

Since we could acquire signals at all steps of our research by using big five personality traits to predict financial wellbeing, the presence of a significant finding new alternative to assess people's financial risks. Among the prediction models we have built, there are models that use the big five personality scores as reported in the literature.

Our results show that customers with any of the following factors are more likely to pay their credit card statements on time: Customers who have high levels of agreeableness, who show an increase in their agreeableness or openness levels over time, who have low levels of neuroticism, or who show a decrease in their neuroticism or extraversion levels. Among these, our findings on openness contradicts whereas findings on neuroticism and extraversion matches with previous studies, which have shown the influence of these three traits on high levels of debt and overspending (Brown & Taylor, 2014; Nyhus & Webley, 2001; Otero-López & Pol, 2013). About agreeableness, however, there are mixed results in the literature. Although two studies (Brown & Taylor, 2014; Otero-López & Villardefrancos, 2013) report a positive association between agreeableness and overspending; similar to our study, Otero-López and Pol (2013) found a negative association. Thus, our findings about openness and agreeableness should be interpreted with caution until they are replicated and validated in future studies.

Despite the consistent findings in the literature on the negative effect of conscientiousness on excessive buying behaviors and debt (Brown & Taylor, 2014; Donnelly et al., 2012; Otero-López & Pol, 2013; Otero-López & Villardefrancos, 2013), we had different results using monthly conscientiousness levels versus trends in conscientiousness. In line with the literature, customers with high levels of conscientiousness were more likely to pay their credit card statements on time compared to customers with low levels of conscientiousness. Nevertheless, an increase in conscientiousness level over time was positively associated with late payment. To our knowledge, no study has yet tested and demonstrated the effect of *change* in conscientiousness on financial wellbeing. Why does the effect of conscientiousness on financial wellbeing turn out to be negative when its change over time is considered? This result would be an indicator of a nonlinear association between conscientiousness and payment behavior. Perhaps, the link between conscientiousness and payment behavior over time varies across levels of conscientiousness. Future research should investigate whether our finding about the change in conscientiousness could be replicated and further examine the possibility of a nonlinear association.

5.1. Limitations and Future Research

Despite our promising findings, our research has some limitations and future directions that we must address. First, in our research, we inferred the big five personality traits scores from the spending categories as reported in the study of Matz et al. (2016). The participants in the Matz et al. (2016) study, however, were customers of a UK-based multinational bank. Thus, the scores derived from the Matz et al. (2016) study may not perfectly reflect the personality traits of a Turkish bank's customers. Customers in the UK and Turkey might have different demographic backgrounds and perceptions regarding shopping habits, which in turn might affect their spending patterns. Hence, future research can aim to conduct a survey with the customers themselves to assess their big five personality traits. Such an approach could also take the local socio-economic conditions and cultural context into account. There is also one more limitation that stems from Matz et al.'s (2016) study. In that research, big five personality scores of spending categories are determined by Amazon's Mechanical Turk workers' ratings. Those workers rated spending categories as if they are people to describe by big five personality traits. This approach might reflect the bias of workers from different socio-cultural environment. To prevent this kind of bias and avoid the effort and customer friction of conducting a survey, extraction of big five personality scores of customers from different banking records might be considered. For instance, determining the extraversion levels of customers based on their spatio-temporal distributions of their credit card spending. Following such an approach would also put the future research on the safe zone. Because big five personality traits are challenged in terms of representation power of people's personality across the globe. The study of Laajaj and his colleagues (2019) is one example of these kind of challenges. In their study, they concluded that assessing big five personality traits by using traditional measures does not result well with populations which are not western, educated, industrialized, rich and democratic (WEIRD), unlike WEIRD populations. Although their main point of critique is relying on content of the surveys, it is worth considering the re-evaluation of the validity of big five personality traits while measuring the personality of non-WEIRD populations or following the suggested approach before.

Another limitation of our research might stem from the translation step in which we translate Matz et al.'s (2016) spending categories to our spending categories. This step forms the basis of the big five personality scores belonging to our spending categories. There were ambiguous translations such as translating "mobile telephone" from Matz et al.'s (2016) categories into our "technology" category that might end up in our "telecommunication" category as well. This kind of translation shifts will change the final big five personality scores belonging to our spending categories. Thus, avoiding the translation step whenever possible would put the research into a safer process. Future studies should also include data with longer time periods to replicate our studies. For example; although our data included 12-month transactions of customers, we do not know whether those transactions can also be used to predict the financial wellbeing levels 2 years later. Last, we used only one type of measure, i.e. paying the minimum or total amount due on the credit card statement, as an indicator of financial wellbeing. The big data science approach we used in this research can also be used to predict other financial wellbeing indicators such as probability of default, overspending or loan/mortgage default, or even in other service domains (e.g. utility, telecommunications) to predict monthly invoice payment collection rates or probabilities.

REFERENCES

Addo, P. M., Guegan, D., & Hassani, B. (2018). Credit Risk Analysis Using Machine and Deep Learning Models. *Risks*, 6(2), 38.

Bleidorn, W., & Hopwood, C. J. (2018). Using Machine Learning to Advance Personality Assessment and Theory. *Personality and Social Psychology Review*, 23(2) 1-14.

Brown, S., & Taylor, K. (2014). Household finances and the 'Big Five' personality traits. *Journal of Economic Psychology*, *45*, 197-212.

Buckland, M., & Gey, F. (1994). The relationship between Recall and Precision. *Journal of the Association for Information Science and Technology*, 45(1), 12-19.

Dong, X., Suhara, Y., Bozkaya, B., Sing, V. K., Lepri, B., & Pentland, A. S. (2017). Social Bridges in Urban Purchase Behavior. *ACM Transactions on Intelligent Systems and Technology*, 9(3).

Donnelly, G., Iyer, R., & Howell, R. T. (2012). The Big Five personality traits, material values, and financial well-being of self-described money managers. *Journal of Economic Psychology*, 33(6),1129-1142.

Fogarty, J., Baker, R. S., & Hudson , S. E. (2005). Case studies in the use of ROC curve analysis for sensor-based estimates in human computer interaction. *Graphics Interface* (pp. 129-136). Canadian Human-Computer Communications Society School of Computer Science, University of Waterloo, Waterloo, Ontario, Canada ©2005.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2017). *Introduction to Statistical Learning: with Applications in R.* Springer.

Khandani, A. E., Kim, A. J., & Lo, A. W. (2010). Consumer credit-risk models via machine-learning algorithms. *Journal of Banking and Finance*, 34(11), 2767-2787.

Kruppa, J., Schwarz, A., Arminger, G., & Ziegler, A. (2013). Consumer credit risk: Individual probability estimates using machine learning. *Expert Systems with Applications*, 40(13), 5125-5131.

Laajaj, R., Macours, K., Hernandez, D. A., Arias, O., Gosling, S. D., Potter, J., . . . Vakis, R. (2019). Challenges to capture the big five personality traits in non-WEIRD populations. *Science Advances*, 5(7).

Matz, S. C., Gladstone, J. J., & Stillwell, D. (2016). Money Buys Happiness When Spending. *Psychological Science*, 27(5), 715-725.

Nyhus, E. K., & Webley, P. (2001). The Role of Personality in Household Saving and Behaviour. *European Journal of Personality*, 15(S1), 85-103.

Otero-López, J. M., & Villardefrancos Pol, E. (2013). Compulsive buying and the Five Factor Model of personality: A facet analysis. *Personality and Individual Differences*, 55(5), 585-590.

Otero-López, J. M., & Villardefrancos, E. (2013). Five-Factor Model personality traits, materialism, and excessive buying: A mediational analysis. *Personality and Individual Differences*, 54(6), 767-772.

Singh, V. K., Bozkaya, B., & Pentland, A. (2015). Money Walks: Implicit Mobility Behavior and Financial Well-Being. *PLOS One*, 10(8).

Strömbäck, C., Lind, T., Skagerlund, K., & Västfjäll, D. (2017). Does self-control predict financial behavior and financial well-being? *Journal of Behavioral and Experimental Finance*, 14, 30-38.

APPENDIX

1) Table of Demographic Information Only Models Combinations with Four Types of Dependent Variable and Machine Learning Algorithms

Dependent Variable	Predicted Month	Algorithm	Accuracy	AUC
On-time payment of the minimum amount due	1	Logistic Regression	82%	50,00%
Payment of full amount due without delay	1	Logistic Regression	56%	56,02%
Payment before grace period of 3 days of the minimum amount due	1	Logistic Regression	90%	50,00%
Payment before grace period of 3 days of the full amount due	1	Logistic Regression	59%	56,31%
On-time payment of the minimum amount due	1	LDA	82%	49,96%
Payment of full amount due without delay	1	LDA	56%	55,88%
Payment before grace period of 3 days of the minimum amount due	1	LDA	90%	49,95%
Payment before grace period of 3 days of the full amount due	1	LDA	59%	56,08%

On-time payment of the minimum amount due	1	Decision Tree	82%	50,00%
Payment of full amount due without delay	1	Decision Tree	55%	54,72%
Payment before grace period of 3 days of the minimum amount due	1	Decision Tree	90%	50,00%
Payment before grace period of 3 days of the full amount due	1	Decision Tree	59%	54,36%
On-time payment of the minimum amount due	1	SVM	82%	49,97%
Payment of full amount due without delay	1	SVM	55%	55,20%
Payment before grace period of 3 days of the minimum amount due	1	SVM	90%	50,00%
Payment before grace period of 3 days of the full amount due	1	SVM	58%	54,99%
On-time payment of the minimum amount due	1	Random Forest	82%	50,00%
Payment of full amount due without delay	1	Random Forest	55%	55,12%
Payment before grace period of 3 days of the minimum amount due	1	Random Forest	90%	50,00%
Payment before grace period of 3 days of the full amount due	1	Random Forest	58%	52,58%
On-time payment of the minimum amount due	2	Logistic Regression	78%	50,10%
Payment of full amount due without delay	2	Logistic Regression	57%	55,44%
Payment before grace period of 3 days of the minimum amount due	2	Logistic Regression	88%	50,00%

Payment before grace period of 3 days of the full amount due	2	Logistic Regression	59%	57,85%
On-time payment of the minimum amount due	2	LDA	78%	50,16%
Payment of full amount due without delay	2	LDA	57%	55,36%
Payment before grace period of 3 days of the minimum amount due	2	LDA	88%	50,00%
Payment before grace period of 3 days of the full amount due	2	LDA	59%	57,82%
On-time payment of the minimum amount due	2	Decision Tree	78%	50,00%
Payment of full amount due without delay	2	Decision Tree	57%	54,05%
Payment before grace period of 3 days of the minimum amount due	2	Decision Tree	88%	50,00%
Payment before grace period of 3 days of the full amount due	2	Decision Tree	57%	55,56%
On-time payment of the minimum amount due	2	SVM	78%	50,06%
Payment of full amount due without delay	2	SVM	57%	55,14%
Payment before grace period of 3 days of the minimum amount due	2	SVM	88%	49,99%
Payment before grace period of 3 days of the full amount due	2	SVM	58%	56,26%
On-time payment of the minimum amount due	2	Random Forest	78%	50,00%
Payment of full amount due without delay	2	Random Forest	56%	54,23%

Payment before grace period of 3 days of the minimum amount due	2	Random Forest	88%	50,00%
Payment before grace period of 3 days of the full amount due	2	Random Forest	57%	55,08%
On-time payment of the minimum amount due	3	Logistic Regression	80%	50,00%
Payment of full amount due without delay	3	Logistic Regression	57%	56,24%
Payment before grace period of 3 days of the minimum amount due	3	Logistic Regression	90%	50,00%
Payment before grace period of 3 days of the full amount due	3	Logistic Regression	58%	55,97%
On-time payment of the minimum amount due	3	Decision Tree	80%	50,00%
Payment of full amount due without delay	3	Decision Tree	56%	55,12%
Payment before grace period of 3 days of the minimum amount due	3	Decision Tree	90%	50,00%
Payment before grace period of 3 days of the full amount due	3	Decision Tree	57%	55,42%
On-time payment of the minimum amount due	3	SVM	80%	49,97%
Payment of full amount due without delay	3	SVM	51%	50,03%
Payment before grace period of 3 days of the minimum amount due	3	SVM	90%	50,00%
Payment before grace period of 3 days of the full amount due	3	SVM	55%	49,99%

On-time payment of the minimum amount due	3	Random Forest	80%	50,00%
Payment of full amount due without delay	3	Random Forest	56%	55,27%
Payment before grace period of 3 days of the minimum amount due	3	Random Forest	90%	50,00%
Payment before grace period of 3 days of the full amount due	3	Random Forest	56%	53,19%
On-time payment of the minimum amount due	4	Logistic Regression	74%	49,99%
Payment of full amount due without delay	4	Logistic Regression	56%	53,11%
Payment before grace period of 3 days of the minimum amount due	4	Logistic Regression	88%	50,00%
Payment before grace period of 3 days of the full amount due	4	Logistic Regression	57%	55,23%
On-time payment of the minimum amount due	4	Decision Tree	74%	50,00%
Payment of full amount due without delay	4	Decision Tree	56%	52,60%
Payment before grace period of 3 days of the minimum amount due	4	Decision Tree	88%	50,00%
Payment before grace period of 3 days of the full amount due	4	Decision Tree	57%	55,72%
On-time payment of the minimum amount due	4	SVM	74%	50,00%
Payment of full amount due without delay	4	SVM	55%	50,11%
Payment before grace period of 3 days of the minimum amount due	4	SVM	88%	50,00%

Payment before grace period of 3 days of the full amount due	4	SVM	55%	49,92%
On-time payment of the minimum amount due	4	Random Forest	74%	50,00%
Payment of full amount due without delay	4	Random Forest	56%	52,05%
Payment before grace period of 3 days of the minimum amount due	4	Random Forest	88%	50,00%
Payment before grace period of 3 days of the full amount due	4	Random Forest	56%	52,65%
On-time payment of the minimum amount due	5	Logistic Regression	84%	50,00%
Payment of full amount due without delay	5	Logistic Regression	58%	57,95%
Payment before grace period of 3 days of the minimum amount due	5	Logistic Regression	91%	50,00%
Payment before grace period of 3 days of the full amount due	5	Logistic Regression	59%	57,45%
On-time payment of the minimum amount due	5	Decision Tree	84%	50,00%
Payment of full amount due without delay	5	Decision Tree	55%	55,36%
Payment before grace period of 3 days of the minimum amount due	5	Decision Tree	91%	50,00%
Payment before grace period of 3 days of the full amount due	5	Decision Tree	58%	54,48%
On-time payment of the minimum amount due	5	SVM	84%	50,00%
Payment of full amount due without delay	5	SVM	57%	57,15%

Payment before grace period of 3 days of the minimum amount due	5	SVM	91%	50,00%
Payment before grace period of 3 days of the full amount due	5	SVM	58%	55,68%
On-time payment of the minimum amount due	5	Random Forest	84%	50,00%
Payment of full amount due without delay	5	Random Forest	56%	55,82%
Payment before grace period of 3 days of the minimum amount due	5	Random Forest	91%	50,00%
Payment before grace period of 3 days of the full amount due	5	Random Forest	58%	54,72%
On-time payment of the minimum amount due	6	Logistic Regression	80%	50,00%
Payment of full amount due without delay	6	Logistic Regression	57%	55,55%
Payment before grace period of 3 days of the minimum amount due	6	Logistic Regression	90%	50,00%
Payment before grace period of 3 days of the full amount due	6	Logistic Regression	57%	56,43%
On-time payment of the minimum amount due	6	Decision Tree	80%	50,00%
Payment of full amount due without delay	6	Decision Tree	55%	54,15%
Payment before grace period of 3 days of the minimum amount due	6	Decision Tree	90%	50,00%
Payment before grace period of 3 days of the full amount due	6	Decision Tree	57%	54,61%

On-time payment of the minimum amount due	6	SVM	80%	50,02%
Payment of full amount due without delay	6	SVM	57%	55,93%
Payment before grace period of 3 days of the minimum amount due	6	SVM	90%	50,00%
Payment before grace period of 3 days of the full amount due	6	SVM	57%	55,35%
On-time payment of the minimum amount due	6	Random Forest	80%	50,00%
Payment of full amount due without delay	6	Random Forest	57%	55,46%
Payment before grace period of 3 days of the minimum amount due	6	Random Forest	90%	50,00%
Payment before grace period of 3 days of the full amount due	6	Random Forest	56%	52,99%
On-time payment of the minimum amount due	7	Logistic Regression	75%	49,97%
Payment of full amount due without delay	7	Logistic Regression	57%	54,07%
Payment before grace period of 3 days of the minimum amount due	7	Logistic Regression	88%	50,00%
Payment before grace period of 3 days of the full amount due	7	Logistic Regression	58%	57,34%
On-time payment of the minimum amount due	7	Decision Tree	75%	50,00%
Payment of full amount due without delay	7	Decision Tree	57%	53,13%
Payment before grace period of 3 days of the minimum amount due	7	Decision Tree	88%	50,00%

Payment before grace period of 3 days of the full amount due	7	Decision Tree	56%	56,20%
On-time payment of the minimum amount due	7	SVM	75%	49,95%
Payment of full amount due without delay	7	SVM	57%	53,40%
Payment before grace period of 3 days of the minimum amount due	7	SVM	88%	50,00%
Payment before grace period of 3 days of the full amount due	7	SVM	58%	57,11%
On-time payment of the minimum amount due	7	Random Forest	75%	50,00%
Payment of full amount due without delay	7	Random Forest	56%	51,27%
Payment before grace period of 3 days of the minimum amount due	7	Random Forest	88%	50,00%
Payment before grace period of 3 days of the full amount due	7	Random Forest	58%	56,52%
On-time payment of the minimum amount due	8	Logistic Regression	88%	50,00%
Payment of full amount due without delay	8	Logistic Regression	59%	58,41%
Payment before grace period of 3 days of the minimum amount due	8	Logistic Regression	91%	50,00%
Payment before grace period of 3 days of the full amount due	8	Logistic Regression	59%	57,06%
On-time payment of the minimum amount due	8	Decision Tree	88%	50,00%
Payment of full amount due without delay	8	Decision Tree	57%	56,40%

Payment before grace period of 3 days of the minimum amount due	8	Decision Tree	91%	50,00%
Payment before grace period of 3 days of the full amount due	8	Decision Tree	57%	56,20%
On-time payment of the minimum amount due	8	SVM	88%	50,00%
Payment of full amount due without delay	8	SVM	58%	57,46%
Payment before grace period of 3 days of the minimum amount due	8	SVM	91%	50,00%
Payment before grace period of 3 days of the full amount due	8	SVM	58%	56,94%
On-time payment of the minimum amount due	8	Random Forest	88%	50,00%
Payment of full amount due without delay	8	Random Forest	58%	56,70%
Payment before grace period of 3 days of the minimum amount due	8	Random Forest	91%	50,00%
Payment before grace period of 3 days of the full amount due	8	Random Forest	58%	54,95%
On-time payment of the minimum amount due	9	Logistic Regression	84%	50,00%
Payment of full amount due without delay	9	Logistic Regression	57%	56,45%
Payment before grace period of 3 days of the minimum amount due	9	Logistic Regression	91%	50,00%
Payment before grace period of 3 days of the full amount due	9	Logistic Regression	57%	55,83%

On-time payment of the minimum amount due	9	Decision Tree	84%	50,00%
Payment of full amount due without delay	9	Decision Tree	55%	55,15%
Payment before grace period of 3 days of the minimum amount due	9	Decision Tree	91%	50,00%
Payment before grace period of 3 days of the full amount due	9	Decision Tree	57%	55,15%
On-time payment of the minimum amount due	9	SVM	84%	49,97%
Payment of full amount due without delay	9	SVM	56%	55,73%
Payment before grace period of 3 days of the minimum amount due	9	SVM	91%	50,00%
Payment before grace period of 3 days of the full amount due	9	SVM	56%	54,80%
On-time payment of the minimum amount due	9	Random Forest	84%	50,00%
Payment of full amount due without delay	9	Random Forest	56%	55,55%
Payment before grace period of 3 days of the minimum amount due	9	Random Forest	91%	50,00%
Payment before grace period of 3 days of the full amount due	9	Random Forest	56%	52,63%
On-time payment of the minimum amount due	10	Logistic Regression	78%	50,01%
Payment of full amount due without delay	10	Logistic Regression	56%	55,66%
Payment before grace period of 3 days of the minimum amount due	10	Logistic Regression	89%	50,00%

Payment before grace period of 3 days of the full amount due	10	Logistic Regression	59%	52,14%
On-time payment of the minimum amount due	10	Decision Tree	78%	50,00%
Payment of full amount due without delay	10	Decision Tree	55%	55,00%
Payment before grace period of 3 days of the minimum amount due	10	Decision Tree	89%	50,00%
Payment before grace period of 3 days of the full amount due	10	Decision Tree	58%	50,00%
On-time payment of the minimum amount due	10	SVM	78%	49,99%
Payment of full amount due without delay	10	SVM	55%	54,91%
Payment before grace period of 3 days of the minimum amount due	10	SVM	89%	50,00%
Payment before grace period of 3 days of the full amount due	10	SVM	59%	52,48%
On-time payment of the minimum amount due	10	Random Forest	78%	50,00%
Payment of full amount due without delay	10	Random Forest	54%	53,91%
Payment before grace period of 3 days of the minimum amount due	10	Random Forest	89%	50,00%
Payment before grace period of 3 days of the full amount due	10	Random Forest	58%	50,06%
On-time payment of the minimum amount due	11	Logistic Regression	78%	50,00%
Payment of full amount due without delay	11	Logistic Regression	57%	55,80%

Payment before grace period of 3 days of the minimum amount due	11	Logistic Regression	89%	50,00%
Payment before grace period of 3 days of the full amount due	11	Logistic Regression	59%	57,81%
On-time payment of the minimum amount due	11	Decision Tree	78%	50,00%
Payment of full amount due without delay	11	Decision Tree	55%	53,02%
Payment before grace period of 3 days of the minimum amount due	11	Decision Tree	89%	50,00%
Payment before grace period of 3 days of the full amount due	11	Decision Tree	57%	56,35%
On-time payment of the minimum amount due	11	SVM	78%	50,02%
Payment of full amount due without delay	11	SVM	53%	50,00%
Payment before grace period of 3 days of the minimum amount due	11	SVM	89%	50,00%
Payment before grace period of 3 days of the full amount due	11	SVM	55%	49,89%
On-time payment of the minimum amount due	11	Random Forest	78%	50,00%
Payment of full amount due without delay	11	Random Forest	56%	53,93%
Payment before grace period of 3 days of the minimum amount due	11	Random Forest	89%	50,00%
Payment before grace period of 3 days of the full amount due	11	Random Forest	58%	55,37%

On-time payment of the minimum amount due	12	Logistic Regression	81%	50%
On-time payment of the minimum amount due	12	LDA	81%	50%
On-time payment of the minimum amount due	12	Decision Tree	81%	50%
On-time payment of the minimum amount due	12	SVM	81%	50%
On-time payment of the minimum amount due	12	Random Forest	81%	50%
Payment of full amount due without delay	12	Logistic Regression	57%	55%
Payment of full amount due without delay	12	LDA	57%	55%
Payment of full amount due without delay	12	Decision Tree	56%	53%
Payment of full amount due without delay	12	SVM	56%	56%
Payment of full amount due without delay	12	Random Forest	56%	55%
Payment before grace period of 3 days of the minimum amount due	12	Logistic Regression	90%	50%
Payment before grace period of 3 days of the minimum amount due	12	LDA	90%	50%
Payment before grace period of 3 days of the minimum amount due	12	Decision Tree	90%	50%
Payment before grace period of 3 days of the minimum amount due	12	SVM	90%	50%
Payment before grace period of 3 days of the minimum amount due	12	Random Forest	90%	50%

Payment before grace period of 3 days of the full amount due	12	Logistic Regression	58%	58%
Payment before grace period of 3 days of the full amount due	12	LDA	58%	58%
Payment before grace period of 3 days of the full amount due	12	Decision Tree	55%	54%
Payment before grace period of 3 days of the full amount due	12	SVM	57%	56%
Payment before grace period of 3 days of the full amount due	12	Random Forest	57%	56%

2) Table of Models Built with Demographic Information and Correspondent Big Five Personality Scores Combinations with Four Types of Dependent Variable and Machine Learning Algorithms for all Months

Dependent Variable	**Predicted Month**	**Machine Learning Algorithm**	**Accuracy**	**AUC**
On-time payment of the minimum amount due	1	Decision Tree	82%	50,00%
On-time payment of the minimum amount due	1	Logistic Regression	82%	50,00%
On-time payment of the minimum amount due	1	naive-Bayes	82%	50,00%
On-time payment of the minimum amount due	1	SVM	82%	50,00%
On-time payment of the minimum amount due	1	Random Forest	82%	50,00%
Payment of full amount due without delay	1	Decision Tree	54%	53,46%
Payment of full amount due without delay	1	Logistic Regression	53%	52,51%
Payment of full amount due without delay	1	naive-Bayes	50%	50,00%
Payment of full amount due without delay	1	SVM	53%	53,42%
Payment of full amount due without delay	1	Random Forest	53%	52,55%
Payment before grace period of 3 days of the minimum amount due	1	Decision Tree	90%	50,00%
Payment before grace period of 3 days of the minimum amount due	1	Logistic Regression	90%	50,00%
Payment before grace period of 3 days of the minimum amount due	1	naive-Bayes	90%	50,00%
Payment before grace period of 3 days of the minimum amount due	1	SVM	90%	50,00%
Payment before grace period of 3 days of the minimum amount due	1	Random Forest	90%	50,00%
Payment before grace period of 3 days of the full amount due	1	Decision Tree	57%	50,00%
Payment before grace period of 3 days of the full amount due	1	Logistic Regression	56%	51,33%

Payment before grace period of 3 days of the full amount due	1	naive-Bayes	57%	50,00%
Payment before grace period of 3 days of the full amount due	1	SVM	57%	49,93%
Payment before grace period of 3 days of the full amount due	1	Random Forest	57%	49,96%
On-time payment of the minimum amount due	2	Decision Tree	78%	50,00%
On-time payment of the minimum amount due	2	Logistic Regression	78%	49,97%
On-time payment of the minimum amount due	2	naive-Bayes	78%	50,00%
On-time payment of the minimum amount due	2	SVM	78%	49,97%
On-time payment of the minimum amount due	2	Random Forest	78%	50,00%
Payment of full amount due without delay	2	Decision Tree	53%	50,00%
Payment of full amount due without delay	2	Logistic Regression	54%	52,05%
Payment of full amount due without delay	2	naive-Bayes	53%	50,00%
Payment of full amount due without delay	2	SVM	54%	51,29%
Payment of full amount due without delay	2	Random Forest	53%	50,16%
Payment before grace period of 3 days of the minimum amount due	2	Decision Tree	88%	50,00%
Payment before grace period of 3 days of the minimum amount due	2	Logistic Regression	88%	49,98%
Payment before grace period of 3 days of the minimum amount due	2	naive-Bayes	88%	50,00%
Payment before grace period of 3 days of the minimum amount due	2	SVM	88%	49,98%
Payment before grace period of 3 days of the minimum amount due	2	Random Forest	88%	50,00%
Payment before grace period of 3 days of the full amount due	2	Decision Tree	55%	54,34%

Payment before grace period of 3 days of the full amount due	2	Logistic Regression	55%	52,69%
Payment before grace period of 3 days of the full amount due	2	naive-Bayes	55%	52,71%
Payment before grace period of 3 days of the full amount due	2	SVM	55%	54,32%
Payment before grace period of 3 days of the full amount due	2	Random Forest	54%	50,00%
On-time payment of the minimum amount due	3	Decision Tree	80%	50,00%
On-time payment of the minimum amount due	3	Logistic Regression	80%	49,97%
On-time payment of the minimum amount due	3	naive-Bayes	80%	50,00%
On-time payment of the minimum amount due	3	SVM	80%	50,00%
On-time payment of the minimum amount due	3	Random Forest	80%	50,00%
Payment of full amount due without delay	3	Decision Tree	52%	51,76%
Payment of full amount due without delay	3	Logistic Regression	52%	51,57%
Payment of full amount due without delay	3	naive-Bayes	51%	50,34%
Payment of full amount due without delay	3	SVM	51%	51,73%
Payment of full amount due without delay	3	Random Forest	52%	51,51%
Payment before grace period of 3 days of the minimum amount due	3	Decision Tree	90%	50,00%
Payment before grace period of 3 days of the minimum amount due	3	Logistic Regression	90%	50,09%
Payment before grace period of 3 days of the minimum amount due	3	naive-Bayes	90%	50,00%
Payment before grace period of 3 days of the minimum amount due	3	SVM	90%	50,00%
Payment before grace period of 3 days of the minimum amount due	3	Random Forest	90%	50,00%

Payment before grace period of 3 days of the full amount due	3	Decision Tree	55%	50,00%
Payment before grace period of 3 days of the full amount due	3	Logistic Regression	54%	49,87%
Payment before grace period of 3 days of the full amount due	3	naive-Bayes	55%	50,00%
Payment before grace period of 3 days of the full amount due	3	SVM	55%	49,92%
Payment before grace period of 3 days of the full amount due	3	Random Forest	55%	49,97%
On-time payment of the minimum amount due	4	Decision Tree	74%	50,00%
On-time payment of the minimum amount due	4	Logistic Regression	74%	50,00%
On-time payment of the minimum amount due	4	naive-Bayes	74%	50,00%
On-time payment of the minimum amount due	4	SVM	74%	50,00%
On-time payment of the minimum amount due	4	Random Forest	74%	50,00%
Payment of full amount due without delay	4	Decision Tree	55%	50,00%
Payment of full amount due without delay	4	Logistic Regression	55%	50,40%
Payment of full amount due without delay	4	naive-Bayes	55%	50,00%
Payment of full amount due without delay	4	SVM	55%	50,02%
Payment of full amount due without delay	4	Random Forest	55%	49,92%
Payment before grace period of 3 days of the minimum amount due	4	Decision Tree	88%	50,00%
Payment before grace period of 3 days of the minimum amount due	4	Logistic Regression	88%	50,05%
Payment before grace period of 3 days of the minimum amount due	4	naive-Bayes	88%	50,00%
Payment before grace period of 3 days of the minimum amount due	4	SVM	88%	50,00%

Payment before grace period of 3 days of the minimum amount due	4	Random Forest	88%	50,00%
Payment before grace period of 3 days of the full amount due	4	Decision Tree	55%	50,00%
Payment before grace period of 3 days of the full amount due	4	Logistic Regression	55%	52,34%
Payment before grace period of 3 days of the full amount due	4	naive-Bayes	51%	52,69%
Payment before grace period of 3 days of the full amount due	4	SVM	55%	49,90%
Payment before grace period of 3 days of the full amount due	4	Random Forest	55%	50,00%
On-time payment of the minimum amount due	5	Decision Tree	84%	50,00%
On-time payment of the minimum amount due	5	Logistic Regression	84%	49,98%
On-time payment of the minimum amount due	5	naive-Bayes	84%	50,00%
On-time payment of the minimum amount due	5	SVM	84%	50,00%
On-time payment of the minimum amount due	5	Random Forest	84%	50,00%
Payment of full amount due without delay	5	Decision Tree	54%	53,63%
Payment of full amount due without delay	5	Logistic Regression	53%	53,03%
Payment of full amount due without delay	5	naive-Bayes	49%	50,00%
Payment of full amount due without delay	5	SVM	54%	53,67%
Payment of full amount due without delay	5	Random Forest	54%	53,61%
Payment before grace period of 3 days of the minimum amount due	5	Decision Tree	91%	50,00%
Payment before grace period of 3 days of the minimum amount due	5	Logistic Regression	91%	49,98%
Payment before grace period of 3 days of the minimum amount due	5	naive-Bayes	91%	50,00%

Payment before grace period of 3 days of the minimum amount due	5	SVM	91%	50,00%
Payment before grace period of 3 days of the minimum amount due	5	Random Forest	91%	50,00%
Payment before grace period of 3 days of the full amount due	5	Decision Tree	56%	50,00%
Payment before grace period of 3 days of the full amount due	5	Logistic Regression	56%	52,12%
Payment before grace period of 3 days of the full amount due	5	naive-Bayes	55%	49,99%
Payment before grace period of 3 days of the full amount due	5	SVM	55%	49,82%
Payment before grace period of 3 days of the full amount due	5	Random Forest	56%	49,98%
On-time payment of the minimum amount due	6	Decision Tree	80%	50,00%
On-time payment of the minimum amount due	6	Logistic Regression	80%	50,07%
On-time payment of the minimum amount due	6	naive-Bayes	80%	50,00%
Payment of full amount due without delay	6	Decision Tree	53%	53,09%
Payment of full amount due without delay	6	Logistic Regression	53%	51,85%
Payment of full amount due without delay	6	naive-Bayes	54%	51,47%
Payment before grace period of 3 days of the minimum amount due	6	Decision Tree	90%	50,00%
Payment before grace period of 3 days of the minimum amount due	6	Logistic Regression	90%	50,09%
Payment before grace period of 3 days of the minimum amount due	6	naive-Bayes	90%	50,00%
Payment before grace period of 3 days of the full amount due	6	Decision Tree	54%	50,00%
Payment before grace period of 3 days of the full amount due	6	Logistic Regression	54%	51,01%

Payment before grace period of 3 days of the full amount due	6	naive-Bayes	54%	50,00%
On-time payment of the minimum amount due	7	Decision Tree	75%	50,00%
On-time payment of the minimum amount due	7	Logistic Regression	75%	50,00%
On-time payment of the minimum amount due	7	naive-Bayes	75%	50,00%
Payment of full amount due without delay	7	Decision Tree	55%	50,00%
Payment of full amount due without delay	7	Logistic Regression	55%	50,44%
Payment of full amount due without delay	7	naive-Bayes	55%	50,06%
Payment before grace period of 3 days of the minimum amount due	7	Decision Tree	88%	50,00%
Payment before grace period of 3 days of the minimum amount due	7	Logistic Regression	88%	50,00%
Payment before grace period of 3 days of the minimum amount due	7	naive-Bayes	88%	50,00%
Payment before grace period of 3 days of the full amount due	7	Decision Tree	55%	54,19%
Payment before grace period of 3 days of the full amount due	7	Logistic Regression	54%	52,21%
Payment before grace period of 3 days of the full amount due	7	naive-Bayes	54%	50,00%
On-time payment of the minimum amount due	8	Decision Tree	88%	50,00%
On-time payment of the minimum amount due	8	Logistic Regression	88%	50,00%
On-time payment of the minimum amount due	8	naive-Bayes	88%	50,00%
Payment of full amount due without delay	8	Decision Tree	55%	54,33%
Payment of full amount due without delay	8	Logistic Regression	54%	53,27%
Payment of full amount due without delay	8	naive-Bayes	54%	50,00%

Payment before grace period of 3 days of the minimum amount due	8	Decision Tree	91%	50,00%
Payment before grace period of 3 days of the minimum amount due	8	Logistic Regression	91%	50,00%
Payment before grace period of 3 days of the minimum amount due	8	naive-Bayes	91%	50,00%
Payment before grace period of 3 days of the full amount due	8	Decision Tree	56%	50,00%
Payment before grace period of 3 days of the full amount due	8	Logistic Regression	54%	51,43%
Payment before grace period of 3 days of the full amount due	8	naive-Bayes	56%	50,00%
On-time payment of the minimum amount due	9	Decision Tree	84%	50,00%
On-time payment of the minimum amount due	9	Logistic Regression	27%	48,41%
On-time payment of the minimum amount due	9	naive-Bayes	84%	50,00%
Payment of full amount due without delay	9	Decision Tree	53%	53,32%
Payment of full amount due without delay	9	Logistic Regression	52%	52,63%
Payment of full amount due without delay	9	naive-Bayes	49%	50,00%
Payment before grace period of 3 days of the minimum amount due	9	Decision Tree	91%	50,00%
Payment before grace period of 3 days of the minimum amount due	9	Logistic Regression	91%	50,00%
Payment before grace period of 3 days of the minimum amount due	9	naive-Bayes	91%	50,00%
Payment before grace period of 3 days of the full amount due	9	Decision Tree	54%	50,00%
Payment before grace period of 3 days of the full amount due	9	Logistic Regression	54%	50,55%
Payment before grace period of 3 days of the full amount due	9	naive-Bayes	54%	50,00%

On-time payment of the minimum amount due	10	Decision Tree	78%	50,00%
On-time payment of the minimum amount due	10	Logistic Regression	78%	50,00%
On-time payment of the minimum amount due	10	naive-Bayes	78%	50,00%
Payment of full amount due without delay	10	Decision Tree	53%	53,03%
Payment of full amount due without delay	10	Logistic Regression	53%	53,18%
Payment of full amount due without delay	10	naive-Bayes	53%	52,59%
Payment before grace period of 3 days of the minimum amount due	10	Decision Tree	89%	50,00%
Payment before grace period of 3 days of the minimum amount due	10	Logistic Regression	89%	50,00%
Payment before grace period of 3 days of the minimum amount due	10	naive-Bayes	89%	50,00%
Payment before grace period of 3 days of the full amount due	10	Decision Tree	58%	50,00%
Payment before grace period of 3 days of the full amount due	10	Logistic Regression	58%	49,94%
Payment before grace period of 3 days of the full amount due	10	naive-Bayes	58%	50,00%
On-time payment of the minimum amount due	11	Decision Tree	78%	50,00%
On-time payment of the minimum amount due	11	Logistic Regression	78%	49,99%
On-time payment of the minimum amount due	11	naive-Bayes	78%	50,00%
Payment of full amount due without delay	11	Decision Tree	53%	50,74%
Payment of full amount due without delay	11	Logistic Regression	53%	51,06%
Payment of full amount due without delay	11	naive-Bayes	53%	50,00%
Payment before grace period of 3 days of the minimum amount due	11	Decision Tree	89%	50,00%

Payment before grace period of 3 days of the minimum amount due	11	Logistic Regression	89%	50,00%
Payment before grace period of 3 days of the minimum amount due	11	naive-Bayes	89%	50,00%
Payment before grace period of 3 days of the full amount due	11	Decision Tree	55%	50,00%
Payment before grace period of 3 days of the full amount due	11	Logistic Regression	54%	51,50%
Payment before grace period of 3 days of the full amount due	11	naive-Bayes	55%	50,00%
On-time payment of the minimum amount due	12	Decision Tree	81%	50,00%
On-time payment of the minimum amount due	12	Logistic Regression	81%	50,00%
On-time payment of the minimum amount due	12	naive-Bayes	81%	50,00%
Payment of full amount due without delay	12	Decision Tree	54%	51,92%
Payment of full amount due without delay	12	Logistic Regression	53%	50,90%
Payment of full amount due without delay	12	naive-Bayes	55%	50,41%
Payment before grace period of 3 days of the minimum amount due	12	Decision Tree	90%	50,00%
Payment before grace period of 3 days of the minimum amount due	12	Logistic Regression	90%	50,00%
Payment before grace period of 3 days of the minimum amount due	12	naive-Bayes	90%	50,00%
Payment before grace period of 3 days of the full amount due	12	Decision Tree	52%	50,00%
Payment before grace period of 3 days of the full amount due	12	Logistic Regression	52%	51,16%
Payment before grace period of 3 days of the full amount due	12	naive-Bayes	52%	50,00%
On-time payment of the minimum amount due	6	SVM	80%	50,07%

On-time payment of the minimum amount due	6	Random Forest	80%	50,00%
Payment of full amount due without delay	6	SVM	53%	53,34%
Payment of full amount due without delay	6	Random Forest	53%	51,27%
Payment before grace period of 3 days of the minimum amount due	6	SVM	90%	50,00%
Payment before grace period of 3 days of the minimum amount due	6	Random Forest	90%	50,00%
Payment before grace period of 3 days of the full amount due	6	SVM	54%	50,12%
Payment before grace period of 3 days of the full amount due	6	Random Forest	54%	50,13%
On-time payment of the minimum amount due	7	SVM	75%	50,00%
On-time payment of the minimum amount due	7	Random Forest	75%	50,00%
Payment of full amount due without delay	7	SVM	55%	50,17%
Payment of full amount due without delay	7	Random Forest	55%	50,00%
Payment before grace period of 3 days of the minimum amount due	7	SVM	88%	50,00%
Payment before grace period of 3 days of the minimum amount due	7	Random Forest	88%	50,00%
Payment before grace period of 3 days of the full amount due	7	SVM	55%	54,20%
Payment before grace period of 3 days of the full amount due	7	Random Forest	54%	50,50%
On-time payment of the minimum amount due	8	SVM	88%	50,00%
On-time payment of the minimum amount due	8	Random Forest	88%	50,00%
Payment of full amount due without delay	8	SVM	55%	54,33%
Payment of full amount due without delay	8	Random Forest	54%	50,75%

Payment before grace period of 3 days of the minimum amount due	8	SVM	91%	50,00%
Payment before grace period of 3 days of the minimum amount due	8	Random Forest	91%	50,00%
Payment before grace period of 3 days of the full amount due	8	SVM	56%	50,00%
Payment before grace period of 3 days of the full amount due	8	Random Forest	56%	50,09%
On-time payment of the minimum amount due	9	SVM	84%	50,00%
On-time payment of the minimum amount due	9	Random Forest	84%	50,00%
Payment of full amount due without delay	9	SVM	53%	52,98%
Payment of full amount due without delay	9	Random Forest	52%	52,39%
Payment before grace period of 3 days of the minimum amount due	9	SVM	91%	50,00%
Payment before grace period of 3 days of the minimum amount due	9	Random Forest	91%	50,00%
Payment before grace period of 3 days of the full amount due	9	SVM	54%	50,00%
Payment before grace period of 3 days of the full amount due	9	Random Forest	54%	50,00%
On-time payment of the minimum amount due	10	SVM	78%	50,00%
On-time payment of the minimum amount due	10	Random Forest	78%	50,00%
Payment of full amount due without delay	10	SVM	53%	52,99%
Payment of full amount due without delay	10	Random Forest	52%	52,33%
Payment before grace period of 3 days of the minimum amount due	10	SVM	89%	50,00%
Payment before grace period of 3 days of the minimum amount due	10	Random Forest	89%	50,00%

Payment before grace period of 3 days of the full amount due	10	SVM	58%	49,98%
Payment before grace period of 3 days of the full amount due	10	Random Forest	58%	50,00%
On-time payment of the minimum amount due	11	SVM	78%	50,00%
On-time payment of the minimum amount due	11	Random Forest	78%	50,00%
Payment of full amount due without delay	11	SVM	53%	50,77%
Payment of full amount due without delay	11	Random Forest	53%	50,53%
Payment before grace period of 3 days of the minimum amount due	11	SVM	89%	50,00%
Payment before grace period of 3 days of the minimum amount due	11	Random Forest	89%	50,00%
Payment before grace period of 3 days of the full amount due	11	SVM	55%	50,00%
Payment before grace period of 3 days of the full amount due	11	Random Forest	55%	49,98%
On-time payment of the minimum amount due	12	SVM	81%	50,00%
On-time payment of the minimum amount due	12	Random Forest	81%	50,00%
Payment of full amount due without delay	12	SVM	55%	52,11%
Payment of full amount due without delay	12	Random Forest	54%	50,98%
Payment before grace period of 3 days of the minimum amount due	12	SVM	90%	50,00%
Payment before grace period of 3 days of the minimum amount due	12	Random Forest	90%	50,00%
Payment before grace period of 3 days of the full amount due	12	SVM	53%	51,52%
Payment before grace period of 3 days of the full amount due	12	Random Forest	52%	50,20%

3) Table of Results belong to Models with Different Lookback Periods and Machine Learning Algorithms

Dependent Variable	Machine Learning Algorithm	Lookback Period	Accuracy	AUC
On-time payment of the minimum amount due	Decision Tree	1	81%	50,0%
On-time payment of the minimum amount due	Decision Tree	2	81%	50,0%
On-time payment of the minimum amount due	Decision Tree	3	81%	50,0%
On-time payment of the minimum amount due	Decision Tree	4	81%	50,0%
On-time payment of the minimum amount due	Decision Tree	5	81%	50,0%
On-time payment of the minimum amount due	Decision Tree	6	81%	50,0%
Payment of full amount due without delay	Decision Tree	1	53%	50,3%
Payment of full amount due without delay	Decision Tree	2	53%	50,3%
Payment of full amount due without delay	Decision Tree	3	53%	50,3%
Payment of full amount due without delay	Decision Tree	4	53%	50,3%
Payment of full amount due without delay	Decision Tree	5	53%	50,3%
Payment of full amount due without delay	Decision Tree	6	53%	50,3%
Payment before grace period of 3 days of the minimum amount due	Decision Tree	1	90%	50,0%
Payment before grace period of 3 days of the minimum amount due	Decision Tree	2	90%	50,0%
Payment before grace period of 3 days of the minimum amount due	Decision Tree	3	90%	50,0%
Payment before grace period of 3 days of the minimum amount due	Decision Tree	4	90%	50,0%
Payment before grace period of 3 days of the minimum amount due	Decision Tree	5	90%	50,0%

Payment before grace period of 3 days of the minimum amount due	Decision Tree	6	90%	50,0%
Payment before grace period of 3 days of the full amount due	Decision Tree	1	52%	50,0%
Payment before grace period of 3 days of the full amount due	Decision Tree	2	52%	50,0%
Payment before grace period of 3 days of the full amount due	Decision Tree	3	52%	50,0%
Payment before grace period of 3 days of the full amount due	Decision Tree	4	52%	50,0%
Payment before grace period of 3 days of the full amount due	Decision Tree	5	52%	50,0%
Payment before grace period of 3 days of the full amount due	Decision Tree	6	52%	50,0%
On-time payment of the minimum amount due	LDA	1	81%	50,0%
On-time payment of the minimum amount due	LDA	2	81%	50,0%
On-time payment of the minimum amount due	LDA	3	81%	50,0%
On-time payment of the minimum amount due	LDA	4	81%	50,0%
On-time payment of the minimum amount due	LDA	5	81%	50,0%
On-time payment of the minimum amount due	LDA	6	81%	50,0%
Payment of full amount due without delay	LDA	1	55%	51,2%
Payment of full amount due without delay	LDA	2	54%	51,0%
Payment of full amount due without delay	LDA	3	54%	51,2%
Payment of full amount due without delay	LDA	4	54%	51,3%
Payment of full amount due without delay	LDA	5	54%	51,2%
Payment of full amount due without delay	LDA	6	55%	51,7%

Payment before grace period of 3 days of the minimum amount due	LDA	1	90%	50,0%
Payment before grace period of 3 days of the minimum amount due	LDA	2	90%	50,0%
Payment before grace period of 3 days of the minimum amount due	LDA	3	90%	50,0%
Payment before grace period of 3 days of the minimum amount due	LDA	4	90%	50,0%
Payment before grace period of 3 days of the minimum amount due	LDA	5	90%	50,0%
Payment before grace period of 3 days of the minimum amount due	LDA	6	90%	50,0%
Payment before grace period of 3 days of the full amount due	LDA	1	52%	50,1%
Payment before grace period of 3 days of the full amount due	LDA	2	51%	49,5%
Payment before grace period of 3 days of the full amount due	LDA	3	52%	50,6%
Payment before grace period of 3 days of the full amount due	LDA	4	52%	50,4%
Payment before grace period of 3 days of the full amount due	LDA	5	51%	49,8%
Payment before grace period of 3 days of the full amount due	LDA	6	52%	50,3%
On-time payment of the minimum amount due	Logistic Regression	1	81%	50,0%
On-time payment of the minimum amount due	Logistic Regression	2	81%	50,0%
On-time payment of the minimum amount due	Logistic Regression	3	81%	50,0%
On-time payment of the minimum amount due	Logistic Regression	4	81%	50,0%
On-time payment of the minimum amount due	Logistic Regression	5	81%	50,0%

On-time payment of the minimum amount due	Logistic Regression	6	81%	50,0%
Payment of full amount due without delay	Logistic Regression	1	55%	51,2%
Payment of full amount due without delay	Logistic Regression	2	54%	51,0%
Payment of full amount due without delay	Logistic Regression	3	54%	51,2%
Payment of full amount due without delay	Logistic Regression	4	54%	51,3%
Payment of full amount due without delay	Logistic Regression	5	54%	51,1%
Payment of full amount due without delay	Logistic Regression	6	55%	51,8%
Payment before grace period of 3 days of the minimum amount due	Logistic Regression	1	90%	50,0%
Payment before grace period of 3 days of the minimum amount due	Logistic Regression	2	90%	50,0%
Payment before grace period of 3 days of the minimum amount due	Logistic Regression	3	90%	50,0%
Payment before grace period of 3 days of the minimum amount due	Logistic Regression	4	90%	50,0%
Payment before grace period of 3 days of the minimum amount due	Logistic Regression	5	90%	50,0%
Payment before grace period of 3 days of the minimum amount due	Logistic Regression	6	90%	50,0%
Payment before grace period of 3 days of the full amount due	Logistic Regression	1	52%	50,1%
Payment before grace period of 3 days of the full amount due	Logistic Regression	2	51%	49,6%
Payment before grace period of 3 days of the full amount due	Logistic Regression	3	52%	50,6%
Payment before grace period of 3 days of the full amount due	Logistic Regression	4	52%	50,3%

Payment before grace period of 3 days of the full amount due	Logistic Regression	5	51%	49,8%
Payment before grace period of 3 days of the full amount due	Logistic Regression	6	52%	50,2%
On-time payment of the minimum amount due	naive-Bayes	1	72%	50,1%
On-time payment of the minimum amount due	naive-Bayes	2	80%	50,2%
On-time payment of the minimum amount due	naive-Bayes	3	79%	50,3%
On-time payment of the minimum amount due	naive-Bayes	4	79%	50,8%
On-time payment of the minimum amount due	naive-Bayes	5	79%	50,3%
On-time payment of the minimum amount due	naive-Bayes	6	79%	50,9%
Payment of full amount due without delay	naive-Bayes	1	48%	50,8%
Payment of full amount due without delay	naive-Bayes	2	52%	51,5%
Payment of full amount due without delay	naive-Bayes	3	52%	51,1%
Payment of full amount due without delay	naive-Bayes	4	53%	51,4%
Payment of full amount due without delay	naive-Bayes	5	53%	50,1%
Payment of full amount due without delay	naive-Bayes	6	53%	50,0%
Payment before grace period of 3 days of the minimum amount due	naive-Bayes	1	78%	49,9%
Payment before grace period of 3 days of the minimum amount due	naive-Bayes	2	90%	50,0%
Payment before grace period of 3 days of the minimum amount due	naive-Bayes	3	90%	50,0%
Payment before grace period of 3 days of the minimum amount due	naive-Bayes	4	90%	50,0%
Payment before grace period of 3 days of the minimum amount due	naive-Bayes	5	90%	50,0%

Payment before grace period of 3 days of the minimum amount due	naive-Bayes	6	90%	50,0%
Payment before grace period of 3 days of the full amount due	naive-Bayes	1	52%	50,2%
Payment before grace period of 3 days of the full amount due	naive-Bayes	2	51%	50,7%
Payment before grace period of 3 days of the full amount due	naive-Bayes	3	51%	51,0%
Payment before grace period of 3 days of the full amount due	naive-Bayes	4	52%	51,9%
Payment before grace period of 3 days of the full amount due	naive-Bayes	5	51%	51,4%
Payment before grace period of 3 days of the full amount due	naive-Bayes	6	50%	49,4%
On-time payment of the minimum amount due	QDA	1	68%	50,7%
On-time payment of the minimum amount due	QDA	2	66%	49,8%
On-time payment of the minimum amount due	QDA	3	65%	49,9%
On-time payment of the minimum amount due	QDA	4	66%	50,2%
On-time payment of the minimum amount due	QDA	5	65%	49,9%
On-time payment of the minimum amount due	QDA	6	68%	50,5%
Payment of full amount due without delay	QDA	1	47%	50,2%
Payment of full amount due without delay	QDA	2	48%	50,6%
Payment of full amount due without delay	QDA	3	49%	50,9%
Payment of full amount due without delay	QDA	4	50%	51,4%
Payment of full amount due without delay	QDA	5	49%	50,2%
Payment of full amount due without delay	QDA	6	50%	51,2%

Payment before grace period of 3 days of the minimum amount due	QDA	1	80%	50,5%
Payment before grace period of 3 days of the minimum amount due	QDA	2	79%	50,4%
Payment before grace period of 3 days of the minimum amount due	QDA	3	75%	50,0%
Payment before grace period of 3 days of the minimum amount due	QDA	4	76%	51,7%
Payment before grace period of 3 days of the minimum amount due	QDA	5	76%	50,8%
Payment before grace period of 3 days of the minimum amount due	QDA	6	79%	50,5%
Payment before grace period of 3 days of the full amount due	QDA	1	52%	50,5%
Payment before grace period of 3 days of the full amount due	QDA	2	52%	50,0%
Payment before grace period of 3 days of the full amount due	QDA	3	52%	50,1%
Payment before grace period of 3 days of the full amount due	QDA	4	52%	50,2%
Payment before grace period of 3 days of the full amount due	QDA	5	52%	50,3%
Payment before grace period of 3 days of the full amount due	QDA	6	53%	50,9%
On-time payment of the minimum amount due	SVM	1	81%	50,0%
On-time payment of the minimum amount due	SVM	2	81%	50,0%
On-time payment of the minimum amount due	SVM	3	81%	50,0%
On-time payment of the minimum amount due	SVM	4	81%	50,0%
On-time payment of the minimum amount due	SVM	5	81%	50,0%

On-time payment of the minimum amount due	SVM	6	81%	50,0%
Payment of full amount due without delay	SVM	1	54%	50,0%
Payment of full amount due without delay	SVM	2	54%	50,0%
Payment of full amount due without delay	SVM	3	54%	50,4%
Payment of full amount due without delay	SVM	4	54%	50,5%
Payment of full amount due without delay	SVM	5	54%	50,5%
Payment of full amount due without delay	SVM	6	55%	51,4%
Payment before grace period of 3 days of the minimum amount due	SVM	1	90%	50,0%
Payment before grace period of 3 days of the minimum amount due	SVM	2	90%	50,0%
Payment before grace period of 3 days of the minimum amount due	SVM	3	90%	50,0%
Payment before grace period of 3 days of the minimum amount due	SVM	4	90%	50,0%
Payment before grace period of 3 days of the minimum amount due	SVM	5	90%	50,0%
Payment before grace period of 3 days of the minimum amount due	SVM	6	90%	50,0%
Payment before grace period of 3 days of the full amount due	SVM	1	52%	50,0%
Payment before grace period of 3 days of the full amount due	SVM	2	52%	50,0%
Payment before grace period of 3 days of the full amount due	SVM	3	52%	50,0%
Payment before grace period of 3 days of the full amount due	SVM	4	52%	50,0%

Payment before grace period of 3 days of the full amount due	SVM	5	52%	50,0%
Payment before grace period of 3 days of the full amount due	SVM	6	52%	50,0%
On-time payment of the minimum amount due	Random Forest	1	81%	50,2%
On-time payment of the minimum amount due	Random Forest	2	81%	50,2%
On-time payment of the minimum amount due	Random Forest	3	81%	50,0%
On-time payment of the minimum amount due	Random Forest	4	81%	50,0%
On-time payment of the minimum amount due	Random Forest	5	81%	50,0%
On-time payment of the minimum amount due	Random Forest	6	81%	50,0%
Payment of full amount due without delay	Random Forest	1	50%	49,6%
Payment of full amount due without delay	Random Forest	2	51%	50,0%
Payment of full amount due without delay	Random Forest	3	50%	49,1%
Payment of full amount due without delay	Random Forest	4	51%	50,1%
Payment of full amount due without delay	Random Forest	5	51%	50,5%
Payment of full amount due without delay	Random Forest	6	51%	50,5%
Payment before grace period of 3 days of the minimum amount due	Random Forest	1	90%	50,0%
Payment before grace period of 3 days of the minimum amount due	Random Forest	2	90%	50,0%
Payment before grace period of 3 days of the minimum amount due	Random Forest	3	90%	50,0%
Payment before grace period of 3 days of the minimum amount due	Random Forest	4	90%	50,0%
Payment before grace period of 3 days of the minimum amount due	Random Forest	5	90%	50,0%

Payment before grace period of 3 days of the minimum amount due	Random Forest	6	90%	50,0%
Payment before grace period of 3 days of the full amount due	Random Forest	1	50%	49,2%
Payment before grace period of 3 days of the full amount due	Random Forest	2	49%	48,1%
Payment before grace period of 3 days of the full amount due	Random Forest	3	49%	48,1%
Payment before grace period of 3 days of the full amount due	Random Forest	4	50%	49,0%
Payment before grace period of 3 days of the full amount due	Random Forest	5	51%	50,4%
Payment before grace period of 3 days of the full amount due	Random Forest	6	52%	51,0%

4) Table of Results belong to Models with Increasing AUROC values after Big Five Personality Traits Excluded from Independent Variables.

Dependent Variable	Algorithm	Lookback Period	Excluded B5 Trait	Base AUC	Excluded AUC
Payment before grace period of 3 days of the full amount due	Naive-Bayes	4	A	51.90%	52.16%
Payment before grace period of 3 days of the full amount due	Naive-Bayes	4	N	51.90%	52.90%
Payment before grace period of 3 days of the full amount due	Naive-Bayes	5	E	51.44%	51.56%
Payment before grace period of 3 days of the full amount due	Naive-Bayes	5	A	51.44%	52.57%
Payment before grace period of 3 days of the full amount due	Naive-Bayes	5	N	51.44%	52.44%
Payment before grace period of 3 days of the full amount due	Naive-Bayes	5	C	51.44%	51.52%
Payment of full amount due without delay	QDA	4	O	51.44%	51.98%
Payment of full amount due without delay	Naive-Bayes	4	A	51.39%	51.74%
Payment of full amount due without delay	Logistic Regression	3	O	51.20%	51.27%
Payment of full amount due without delay	LDA	3	E	51.17%	51.19%
Payment of full amount due without delay	LDA	5	C	51.16%	51.19%
Payment of full amount due without delay	Logistic Regression	5	C	51.13%	51.19%
Payment of full amount due without delay	Naive-Bayes	3	O	51.08%	51.18%
Payment of full amount due without delay	Naive-Bayes	3	E	51.08%	51.51%
Payment of full amount due without delay	Naive-Bayes	3	A	51.08%	51.48%
Payment of full amount due without delay	Naive-Bayes	3	N	51.08%	51.53%
Payment of full amount due without delay	Naive-Bayes	3	C	51.08%	51.25%
Payment before grace period of 3 days of the full amount due	Naive-Bayes	3	E	51.01%	51.32%
Payment before grace period of 3 days of the full amount due	Naive-Bayes	3	A	51.01%	52.04%

Payment before grace period of 3 days of the full amount due	Naive-Bayes	3	N	51.01%	52.29%
Payment before grace period of 3 days of the full amount due	Naive-Bayes	3	C	51.01%	52.00%
Payment before grace period of 3 days of the minimum amount due	QDA	5	C	50.84%	51.97%
Payment of full amount due without delay	Naive-Bayes	1	A	50.77%	51.80%
Payment before grace period of 3 days of the full amount due	Naive-Bayes	2	N	50.71%	51.04%
Payment before grace period of 3 days of the full amount due	Naive-Bayes	2	C	50.71%	51.03%
Payment before grace period of 3 days of the full amount due	LDA	3	A	50.62%	50.85%
Payment before grace period of 3 days of the full amount due	LDA	3	N	50.62%	51.00%
Payment before grace period of 3 days of the full amount due	Logistic Regression	3	A	50.62%	50.98%
Payment before grace period of 3 days of the full amount due	Logistic Regression	3	N	50.62%	50.79%
Payment of full amount due without delay	QDA	2	E	50.62%	50.84%
On-time payment of the minimum amount due	QDA	6	O	50.54%	50.97%
On-time payment of the minimum amount due	QDA	6	C	50.54%	51.36%
Payment of full amount due without delay	SVM	5	E	50.54%	50.64%
Payment of full amount due without delay	SVM	5	N	50.54%	50.96%
Payment of full amount due without delay	SVM	5	C	50.54%	50.63%
Payment before grace period of 3 days of the full amount due	QDA	1	O	50.52%	50.71%
Payment of full amount due without delay	Random Forest	5	A	50.51%	50.83%
Payment of full amount due without delay	Random Forest	5	N	50.51%	50.67%
Payment of full amount due without delay	Random Forest	5	C	50.51%	50.60%

Payment before grace period of 3 days of the minimum amount due	QDA	1	O	50.49%	50.87%
Payment before grace period of 3 days of the minimum amount due	QDA	1	C	50.49%	50.82%
Payment of full amount due without delay	SVM	4	N	50.49%	50.67%
Payment of full amount due without delay	Random Forest	6	E	50.47%	51.43%
Payment of full amount due without delay	Random Forest	6	A	50.47%	51.10%
Payment of full amount due without delay	Random Forest	6	N	50.47%	50.92%
Payment of full amount due without delay	Random Forest	6	C	50.47%	50.76%
Payment before grace period of 3 days of the full amount due	Random Forest	5	A	50.44%	50.45%
Payment before grace period of 3 days of the full amount due	Random Forest	5	C	50.44%	51.19%
Payment before grace period of 3 days of the minimum amount due	QDA	2	O	50.38%	50.52%
Payment before grace period of 3 days of the full amount due	LDA	4	O	50.35%	50.39%
Payment before grace period of 3 days of the full amount due	LDA	4	A	50.35%	50.59%
Payment before grace period of 3 days of the full amount due	Logistic Regression	4	O	50.32%	50.36%
Payment before grace period of 3 days of the full amount due	Logistic Regression	4	A	50.32%	50.55%
Payment before grace period of 3 days of the full amount due	Logistic Regression	4	N	50.32%	50.36%
Payment of full amount due without delay	Decision Tree	1	O	50.28%	50.78%
Payment before grace period of 3 days of the full amount due	LDA	6	A	50.27%	50.64%
Payment before grace period of 3 days of the full amount due	LDA	6	N	50.27%	50.47%
On-time payment of the minimum amount due	Naive-Bayes	3	O	50.27%	50.45%
On-time payment of the minimum amount due	Naive-Bayes	3	E	50.27%	50.50%

On-time payment of the minimum amount due	Naive-Bayes	3	N	50.27%	50.40%
On-time payment of the minimum amount due	Naive-Bayes	3	C	50.27%	50.56%
On-time payment of the minimum amount due	Naive-Bayes	5	O	50.26%	50.43%
On-time payment of the minimum amount due	Naive-Bayes	5	E	50.26%	50.62%
On-time payment of the minimum amount due	Naive-Bayes	5	N	50.26%	50.28%
Payment before grace period of 3 days of the full amount due	QDA	5	O	50.26%	50.47%
Payment before grace period of 3 days of the full amount due	QDA	5	E	50.26%	50.31%
On-time payment of the minimum amount due	Random Forest	1	A	50.22%	50.32%
On-time payment of the minimum amount due	Random Forest	1	C	50.22%	50.33%
Payment of full amount due without delay	QDA	5	O	50.21%	51.57%
Payment of full amount due without delay	QDA	5	E	50.21%	50.72%
Payment of full amount due without delay	QDA	5	N	50.21%	50.25%
Payment of full amount due without delay	QDA	5	C	50.21%	50.52%
Payment before grace period of 3 days of the full amount due	Logistic Regression	6	A	50.20%	50.77%
Payment before grace period of 3 days of the full amount due	Logistic Regression	6	N	50.20%	50.66%
Payment before grace period of 3 days of the full amount due	Logistic Regression	6	C	50.20%	50.21%
Payment before grace period of 3 days of the full amount due	Naive-Bayes	1	E	50.20%	50.52%
Payment before grace period of 3 days of the full amount due	Naive-Bayes	1	A	50.20%	51.26%
Payment before grace period of 3 days of the full amount due	Naive-Bayes	1	N	50.20%	51.00%
Payment before grace period of 3 days of the full amount due	QDA	4	O	50.19%	50.81%

Payment before grace period of 3 days of the full amount due	QDA	4	A	50.19%	50.90%
On-time payment of the minimum amount due	Naive-Bayes	2	E	50.17%	50.28%
On-time payment of the minimum amount due	Naive-Bayes	2	N	50.17%	50.45%
On-time payment of the minimum amount due	Naive-Bayes	2	C	50.17%	50.28%
On-time payment of the minimum amount due	QDA	4	O	50.16%	50.88%
On-time payment of the minimum amount due	QDA	4	A	50.16%	50.44%
On-time payment of the minimum amount due	QDA	4	N	50.16%	50.32%
On-time payment of the minimum amount due	QDA	4	C	50.16%	51.43%
Payment of full amount due without delay	Random Forest	4	N	50.15%	50.33%
Payment of full amount due without delay	Random Forest	4	C	50.15%	50.96%
Payment before grace period of 3 days of the full amount due	LDA	1	A	50.14%	50.94%
Payment before grace period of 3 days of the full amount due	LDA	1	N	50.14%	50.30%
Payment before grace period of 3 days of the full amount due	Logistic Regression	1	O	50.14%	50.20%
On-time payment of the minimum amount due	Naive-Bayes	1	O	50.10%	50.13%
On-time payment of the minimum amount due	Naive-Bayes	1	E	50.10%	50.26%
On-time payment of the minimum amount due	Naive-Bayes	1	A	50.10%	50.17%
On-time payment of the minimum amount due	Naive-Bayes	1	N	50.10%	50.30%
On-time payment of the minimum amount due	Naive-Bayes	1	C	50.10%	50.28%
Payment of full amount due without delay	Naive-Bayes	5	O	50.10%	51.48%
Payment of full amount due without delay	Naive-Bayes	5	E	50.10%	51.30%
Payment of full amount due without delay	Naive-Bayes	5	A	50.10%	51.64%
Payment of full amount due without delay	Naive-Bayes	5	N	50.10%	51.16%

Payment of full amount due without delay	Naive-Bayes	5	C	50.10%	51.36%
Payment before grace period of 3 days of the full amount due	QDA	3	O	50.10%	50.67%
Payment before grace period of 3 days of the full amount due	QDA	3	E	50.10%	50.52%
Payment before grace period of 3 days of the full amount due	QDA	3	A	50.10%	50.13%
On-time payment of the minimum amount due	Random Forest	4	O	50.00%	50.05%
On-time payment of the minimum amount due	Random Forest	6	O	50.00%	50.09%
On-time payment of the minimum amount due	Random Forest	6	A	50.00%	50.07%
Payment of full amount due without delay	Naive-Bayes	6	O	49.98%	51.34%
Payment of full amount due without delay	Naive-Bayes	6	E	49.98%	51.60%
Payment of full amount due without delay	Naive-Bayes	6	A	49.98%	51.61%
Payment of full amount due without delay	Naive-Bayes	6	N	49.98%	51.17%
Payment of full amount due without delay	Naive-Bayes	6	C	49.98%	51.97%
Payment of full amount due without delay	Random Forest	2	A	49.98%	50.41%
Payment of full amount due without delay	Random Forest	2	N	49.98%	50.49%
On-time payment of the minimum amount due	Random Forest	3	O	49.98%	50.00%
On-time payment of the minimum amount due	Random Forest	3	E	49.98%	50.00%
On-time payment of the minimum amount due	Random Forest	3	A	49.98%	50.07%
On-time payment of the minimum amount due	Random Forest	3	N	49.98%	50.09%
On-time payment of the minimum amount due	Random Forest	3	C	49.98%	50.07%
On-time payment of the minimum amount due	Random Forest	5	O	49.98%	50.07%
On-time payment of the minimum amount due	Random Forest	5	C	49.98%	50.00%
Payment before grace period of 3 days of the minimum amount due	QDA	3	O	49.97%	50.28%

Payment before grace period of 3 days of the full amount due	QDA	2	O	49.96%	50.52%
Payment before grace period of 3 days of the full amount due	QDA	2	C	49.96%	50.00%
Payment before grace period of 3 days of the minimum amount due	Naive-Bayes	1	O	49.88%	50.00%
Payment before grace period of 3 days of the minimum amount due	Naive-Bayes	1	E	49.88%	50.00%
Payment before grace period of 3 days of the minimum amount due	Naive-Bayes	1	A	49.88%	50.00%
Payment before grace period of 3 days of the minimum amount due	Naive-Bayes	1	N	49.88%	50.00%
Payment before grace period of 3 days of the minimum amount due	Naive-Bayes	1	C	49.88%	50.00%
On-time payment of the minimum amount due	QDA	3	O	49.87%	50.52%
On-time payment of the minimum amount due	QDA	3	A	49.87%	49.98%
On-time payment of the minimum amount due	QDA	3	N	49.87%	50.32%
On-time payment of the minimum amount due	QDA	3	C	49.87%	50.48%
On-time payment of the minimum amount due	QDA	5	O	49.86%	50.93%
On-time payment of the minimum amount due	QDA	5	A	49.86%	50.56%
On-time payment of the minimum amount due	QDA	5	N	49.86%	50.23%
On-time payment of the minimum amount due	QDA	5	C	49.86%	50.07%
On-time payment of the minimum amount due	QDA	2	A	49.83%	50.50%
On-time payment of the minimum amount due	QDA	2	N	49.83%	50.48%
On-time payment of the minimum amount due	QDA	2	C	49.83%	51.71%
Payment before grace period of 3 days of the full amount due	Logistic Regression	5	A	49.81%	50.08%
Payment before grace period of 3 days of the full amount due	LDA	5	O	49.78%	50.41%

Payment before grace period of 3 days of the full amount due	LDA	5	C	49.78%	49.84%
Payment of full amount due without delay	Random Forest	1	E	49.60%	50.24%
Payment before grace period of 3 days of the full amount due	Logistic Regression	2	O	49.58%	49.81%
Payment before grace period of 3 days of the full amount due	Logistic Regression	2	E	49.58%	50.06%
Payment before grace period of 3 days of the full amount due	Logistic Regression	2	A	49.58%	50.55%
Payment before grace period of 3 days of the full amount due	Logistic Regression	2	C	49.58%	49.59%
Payment before grace period of 3 days of the full amount due	LDA	2	O	49.55%	50.06%
Payment before grace period of 3 days of the full amount due	LDA	2	E	49.55%	49.92%
Payment before grace period of 3 days of the full amount due	LDA	2	A	49.55%	50.94%
Payment before grace period of 3 days of the full amount due	LDA	2	N	49.55%	50.30%
Payment before grace period of 3 days of the full amount due	LDA	2	C	49.55%	49.83%
Payment before grace period of 3 days of the full amount due	Naive-Bayes	6	O	49.43%	50.58%
Payment before grace period of 3 days of the full amount due	Naive-Bayes	6	E	49.43%	51.69%
Payment before grace period of 3 days of the full amount due	Naive-Bayes	6	A	49.43%	49.91%
Payment before grace period of 3 days of the full amount due	Naive-Bayes	6	N	49.43%	49.86%
Payment before grace period of 3 days of the full amount due	Random Forest	1	A	49.23%	49.33%
Payment of full amount due without delay	Random Forest	3	O	49.10%	49.31%
Payment of full amount due without delay	Random Forest	3	A	49.10%	49.59%
Payment of full amount due without delay	Random Forest	3	N	49.10%	49.23%

Payment before grace period of 3 days of the full amount due	Random Forest	4	A	48.96%	49.91%
Payment before grace period of 3 days of the full amount due	Random Forest	4	C	48.96%	49.53%
Payment before grace period of 3 days of the full amount due	Random Forest	2	O	48.09%	48.67%
Payment before grace period of 3 days of the full amount due	Random Forest	2	E	48.09%	49.54%
Payment before grace period of 3 days of the full amount due	Random Forest	2	A	48.09%	49.24%
Payment before grace period of 3 days of the full amount due	Random Forest	2	N	48.09%	49.19%
Payment before grace period of 3 days of the full amount due	Random Forest	2	C	48.09%	48.72%
Payment before grace period of 3 days of the full amount due	Random Forest	3	O	48.06%	48.69%
Payment before grace period of 3 days of the full amount due	Random Forest	3	E	48.06%	48.29%
Payment before grace period of 3 days of the full amount due	Random Forest	3	A	48.06%	49.09%
Payment before grace period of 3 days of the full amount due	Random Forest	3	N	48.06%	48.81%
Payment before grace period of 3 days of the full amount due	Random Forest	3	C	48.06%	49.88%