



# Improving the predictive accuracy of the cross-selling of consumer loans using deep learning networks

Noureddine Boustani<sup>1</sup> · Ali Emrouznejad<sup>2</sup> · Roya Gholami<sup>3</sup> · Ozren Despic<sup>1</sup> · Athina Ioannou<sup>2</sup>

Accepted: 19 January 2023  
© The Author(s) 2023

## Abstract

Traditionally most cross-selling models in retail banking use demographics information and interactions with marketing as input to statistical models or machine learning algorithms to predict whether a customer is willing to purchase a given financial product or not. We overcome with such limitation by building several models that also use several years of account transaction data. The objective of this study is to analysis credit card transactions of customers, in order to come up with a good prediction in cross-selling products. We use deep-learning algorithm to analyze almost 800,000 credit cards transactions. The results show that such unique data contains valuable information on the customers' consumption behavior and it can significantly increase the predictive accuracy of a cross-selling model. In summary, we develop an auto-encoder to extract features from the transaction data and use them as input to a classifier. We demonstrate that such features also have predictive power that enhances the performance of the cross-selling model even further.

**Keywords** Cross-selling · Deep learning · Artificial intelligence · Banking · Targeted marketing

## 1 Introduction

In many countries, the retail banking market is saturated and competition to acquire new costumers is fierce and costly. When a customer is lured by the offer of another bank, they may leave their initial bank for the new one, making it a zero-sum game, a phenomena referred to as churn (Kamakura et al., 2003). It is far easier and less expensive to sell more products to the existing customers than to acquire new ones, a tactic called cross-selling.

---

✉ Ali Emrouznejad  
a.emrouznejad@surrey.ac.uk

<sup>1</sup> Aston Business School, Aston University, Birmingham, UK

<sup>2</sup> Surrey Business School, University of Surrey, Guildford, UK

<sup>3</sup> Department of MIS, College of Business & Management, University of Illinois Springfield, Springfield, USA

Since cross-selling is crucial to banks, it needs to be done efficiently. When the bank contacts a customer to offer them a product, it must be sure that the given product fulfills their needs and the solicitation comes at the right time. Past experience shows when a customer is contacted unsuccessfully for offers too often, these solicitations can become a turn off and tend to weaken the relationship.

Financial institutions can use big data and artificial intelligence (advanced analytics) to identify cross-selling opportunities, pre-populate applications, and assess a client's coverage and risk management needs (Akter et al, 2022; Mahbobi et al, 2021; Wetzel, 2019). Unlike Amazon and other big tech companies, when the customer makes a bank transaction, they are unlikely to receive any recommendation from the bank. Banks struggle with customer analytics since they have a great amount of data, but they also have old databases and silos that do not want to share customer data with one another, and they have difficulty in identifying the best ways to use the data (Crosman, 2017). If banks and financial institutions want to provide the same type of personalized insight that big tech companies like Amazon do and compete with fintech companies, they need to adopt artificial intelligence since analytics and AI for marketing and cross-selling are key success factors (Crosman, 2017).

In order to implement efficient cross-selling campaigns, banks usually build models to identify the right product to offer to the right customer at the time with the hope of becoming efficient at the cross-selling effort. Statistical models and machine learning have been used extensively to that end. Most cross-selling models use the customer's demographic data and their financial products purchasing history to decide whether they will be targeted with a given offer or not. Demographic data encompasses age, gender, address, income, marital status, family size, etc. However, banks hold many other types of customer related information, among which the huge amounts of transactional data such as credit card transactions that contain a wealth of information on the customer's consumption behavior and if used properly, can potentially increase the accuracy of the cross-selling models.

Previous studies (see Mulyanegara 2009) have found that there is a link between consumption patterns and the customer personality traits, which has an impact on the customer's financial strategy and thus on the probability that they decide to purchase a given financial product. Brown and Taylor (2014) state that Individual risk attitudes has been a particular focus in recent years amongst both academics and policy makers. For example, existing research has found that openness to experience; neuroticism and agreeableness are all related to risk aversion. Their findings suggest that some personality traits are correlated with the amount of unsecured debt and financial assets held by households. Previous research has also found correlation between personality traits and consumption behavior.

Ładyżyński et al (2019) explored the use of Deep Learning and random forests for direct marketing campaigns in retail banking, while Zinovyeva et al (2020) used Deep Learning algorithm for detecting antisocial online behavior of customers. Deep LearningFor predicting online shopping behavior from clickstream data. On the topic of who will buy, Chaudhuri et al (2021) used Deep Learning to predict customers' purchase behavior and Kim et al (2021) proposed a deep hybrid learning model for customer repurchase behavior.

Furthermore, in the retail banking context, Deep Learning is also used for analyzing credit card fraud detection, for example Forough and Momtazi, (2021) ensembled of deep sequential models while Kriebel and Stitz (2021) proposed a Deep Learning algorithm for credit default prediction from user-generated text in peer-to-peer lending. In the same area, Zhang et al (2021) developed a novel feature engineering methodology for credit card fraud detection using Deep Learning architecture and Gunnarsson et al. (2021) proposed a Deep Learning method for credit scoring. Deep Learning is also applied in forecasting of the demand for advertising expenses and implications on sustainability (Birim et al., 2022) while van der

Gaast and Weidinger (2020) used Deep Learning approach for the selection of an order picking system and Jauhar et al (2022) proposed a Deep Learning-based approach for performance assessment and prediction in the pulp and paper industries.

Moreover, a wealth of previous studies has extracted consumption patterns from the transactional data in many ways. For example, using a clustering algorithm to group customers based on the percentage of income spent on different types of goods and services. Recently, Deep Learning algorithms have proved to be exceptionally powerful at extracting patterns from data. Najafabadi et al (2015) suggests that deep learning algorithms are a promising avenue of research on automated extraction of complex data representations (features) at high levels of abstraction while Karatzoglou et al. (2016) argue that Deep Learning is the next big things in Recommendation Systems technology.

To our knowledge, the use of Deep Learning in Direct Marketing is limited. Therefore, the main objective of the current study is to show that Deep Learning techniques can be used to extract consumption patterns for the purpose of increasing the predictive accuracy of cross-selling models. Among Deep Learning algorithms, Auto-encoders are prominent, and so we use them in this study to extract the consumption patterns.

## 2 Literature review on cross-selling and use of deep learning

### 2.1 Cross-selling

The earliest studies (Hebden & Pickering, 1974) on cross-selling have focused on methodologies for identifying common patterns in acquisition products by customers based on ownership or usage data. Kamakura et al. (1991) describes and predicts purchase sequence based on the influence of the financial maturity of the customer (linked to life stage) and the acquisition difficulty of the service (such as resources required, level of risk and liquidity, information costs) on the acquisition sequence. While a substantial amount of research exists on the acquisition sequence for consumer durables, there is little research in the context of financial services (Jake et al., 2007).

Among the few studies focusing on financial products, Kamakura et al. (2007) is one of the most prominent which states that consumers must balance multiple financial objectives while they evolve through their life cycles. At the early stages, consumers finance current consumption with future income through credit. As they mature, their income surpasses their consumption and therefore they finance past and future consumption with current income through loan payments and savings/investments, respectively. Later in life, they finance current consumption with past income from retirement savings/investments.

Li et al. (2011) developed a multivariate customer-response model with Hidden Markov transition states to statistically capture the possibility that customer demand for various products is governed by latent financial states, during which customers have different preference priorities and responsiveness to cross-selling solicitations for various products.

Direct marketing literature also tries to answer the same type of question. For example “what is the probability that a given customer would accept to purchase a given product” provides an abundant source of models that are ready to use. Bose and Chen (2009) state that the analytics used by researchers for direct marketing fall in three categories: (1) the traditional statistical techniques. (2) The advanced statistical techniques that consist of two-stage models or latent class models. (3) The data mining techniques based on machine learning. Until recently, statistical models such as logistic regression and discriminant analysis

have dominated the modelling of consumer response to direct marketing. Although statistical models can be very powerful, they make strongest assumption on the types of data and their distribution and typically can handle only a limited number of variables. Most popular machine learning techniques used in direct marketing are Artificial Neural Networks (ANN), Bayesian Networks, CHAID, Decision Trees (DT), SVM, Genetic Programming (GP), Genetic Algorithms (GA), and Evolutionary Programming (EP).

Regarding more recent classification techniques, there are ensemble methods that ground on the principle of combining a large number of individual classification models to improve predictive accuracy. Representatives of this category differ mainly in terms of their approach to construct complementary classifiers. Therefore, the individual classifiers, commonly termed base models must show some diversity to improve the accuracy of the full model. We distinguish homogeneous ensembles, which create the base models using the same algorithm, and heterogeneous ensembles, which employ different algorithms.

Lessmann et al. (2015), in their comparative study of different classification techniques for credit scoring, observed that multiple classifier architectures predict credit risk with high accuracy. The strongest competitor outside this family is RF (Random Forest) which is often credited as a very strong classifier (e.g., Brown and Mues, 2012; Kruppa et al.; 2013). However, a comparison to heterogeneous ensemble classifiers reveals that such approaches further improve RF. Bose and Chen (2009) also state that the data used for direct marketing targeting fall into the following categories: Purchase History, Demographic, Geographic, Lifestyle, Socio-graphic, RFM data, Feedbacks from customer, Customer Web browsing log files, Product attributes, Marketing Solicitations.

Customer personality traits have the potential to be useful information for financial product cross-selling model because according to Brown and Taylor (2014) some personality traits are statistically significantly associated with the amount of unsecured debt and financial assets held by households. They also found that personality traits have different associations with the various types of debt. For example, extraversion is positively associated with the probability of holding credit card debt whilst conscientiousness is inversely associated with the probability of holding this type of debt. However, to our best of knowledge, there is no research that used transaction data such as the purchasing information made available to the bank when a customer uses his credit card to make a payment in a shop or on the internet. Such information can be a gold mine for the bank. There are ensemble methods that ground on the principle of combining a large number of individual purchasing habits and may yield information on his personality as shown by Mulyangegara et al. (2009) who found that some dimensions of The Big Five constructs (personality traits) are significantly related to preferences on particular dimensions of brand personality (brands purchased from). It was found, for example, that consumers who exhibit a Conscientious personality demonstrate preferences towards 'Trusted' brands. In contrast, those who are Extrovert in nature are motivated by 'Sociable' brands.

## 2.2 Deep learning

Consequently, credit card data, which contains purchasing information, which itself contains some information on the customer's personality traits, can be extremely useful input to our cross-selling model. However, handcrafting variables from credit card data to be used as proxy for personality traits is a risky enterprise. The act of engineering those variables, referred to as feature engineering, is time consuming and difficult and sometimes causes machine learning projects to fail. Performing feature engineering in a more automated and general

fashion can be done by Deep Learning algorithms which are a promising avenue of research into the automated extraction of complex data representations (features) at high levels of abstraction.

The knowledge learnt from Deep Learning algorithms has been largely untapped in the context of Big Data Analytics in general and cross-selling models, in particular. Certain Big Data domains, such as computer vision and speech recognition, have seen the application of Deep Learning largely to improve classification results. The ability of Deep Learning to extract high-level, complex abstractions and data representations from large volumes of data, especially unsupervised data makes it attractive as a valuable tool for extracting useful information from credit card data.

According to Karatzoglou et al. (2016), Deep Learning is one of the next big things in Recommendation Systems technology, however, to our knowledge, the use of Deep Learning in Direct Marketing is limited. The past few years have seen the tremendous success of deep neural networks in a number of complex tasks such as computer vision, natural language processing and speech recognition.

Despite this, limited work has been published on Deep Learning methods for Recommender Systems. Notable recent application areas are music recommendation, news recommendation, and session-based recommendation. van den Oord et al. (2013) showed that recent advances in Deep Learning translate very well to the music recommendation, with deep convolutional neural networks significantly outperforming a more traditional approach using bag-of-words representations of audio signals. Their results indicate that research in this domain could significantly benefit from using deep neural networks.

Wangperawong et al. (2016) used Customer temporal behavioral data and represented it as images in order to perform churn prediction by leveraging Deep Learning architectures prominent in image classification. Deep convolutional neural networks and auto-encoders proved useful for predicting and understanding churn in the telecommunications industry, outperforming other simpler models such as decision tree modelling.

### **3 Research methods and data collection**

Typically, there are two kinds of data on customers characteristics that are used in quantitative models for direct marketing. The first type of data includes customers' geographic, lifestyle and or Socio-demographic characteristics, the second type of data contains interactive behavior with marketers (Bose & Chen, 2009). The objective of this study is to provide empirical evidence in the context of banking products, that account transaction data contains valuable information which can enhance the accuracy of these predictive models.

For this purpose, two experiments have been set. The objective of the first experiment is to compare the AUC (Area Under Curve) of a model using only demographic data and product ownership with another model that also uses transaction data. We will demonstrate that the AUC of the latter is significantly higher. Many types of classifiers will be used to that end in order to reach a robust conclusion.

The objective of the second experiment is to prove not only the predictors within the transaction data have an intrinsic value, but also their combination contains information that has predictive power too. This will be done by using an auto-encoder. Auto-encoders are a representation learning technique using unsupervised pre-training to learn good representations of the data transform and reduce the dimensionality of the data in order to facilitate the supervised learning stage (Vieira, 2015).

### 3.1 Experiment 1

The objective of first stage (experiment 1) of this study is to measure the predictive accuracy of a model using only demographic data and product ownership with another model that uses both demographic data transaction data. The stage consists of the following steps. 1. Select some machine learning (ML) algorithms that are known for their strong performance. 2. Feed them with three data sets: A. Demographic and Product ownership data; B. Transaction Data; C. Both data set A and data set B. 3. For each ML algorithm, select Consumer Loan as a target and run a k-fold cross-validation with each one of the data sets above (A, B and C). Measure AUC (Area under Curve) for each run. 4. Average AUCs for data sets A, B and C over all the selected ML algorithms and compare them.

The ML algorithms that have been chosen for this study are: 1. An Artificial Neural Network. In this case a multilayer perceptron with a tanh activation function using a cross-entropy loss as a cost function. The algorithm uses a simulated annealing method to determine the architecture of the NN. 2. A CHAID decision tree with a maximum depth of 5. 3. A Support Vector Machine with an RBF kernel. 4. A Random Forest with a maximum depth of 20 and number of trees varying from 50 to 100. 5. A Gradient Boosting Machine with a maximum depth of 5 and number of trees ranging from 50 to 100.

A heterogeneous ensemble made of an ensemble of  $n$  Bagged Neural Networks, an ensemble of  $n$  bagged CHAID trees and an SVM. Number  $n$  will take the values 5, 10 to 25. The Neural Network ensemble and the CHAID ensemble use a majority vote mechanism, while at the level of the heterogeneous ensemble, a confidence weighted vote is used. All cross-validations are 6 folds, in order to have test folds of approximately 1000 samples each.

### 3.2 Experiment 2

In this section (experiment 2), we prove not only transaction data has a significant predictive power, but also there are some patterns within the data that can be extracted and used as predictors, adding extra predictive power to the cross-selling model. We will use an auto-encoder for that. To achieve this, the following steps will be taken: 1. From the set of ML algorithms used in experiment 1, select the one with the highest predictive power, based on the AUC measure. We call this algorithm X. 2. Run an auto-encoder on the data set B (transaction data) to extract hidden features from it. We obtain a data set made of 8 features that we will call data set D. 3. Run algorithm X with the following data sets A, B, C and C + D (i.e., all raw data plus the features extracted by the auto-encoder). 4. Vary some parameters of algorithm X (the size of the ensemble, for example) and compare AUCs obtained. On average, algorithm X should have a higher AUC when using data set C + D than when using only data set C.

### 3.3 Data collection

A sample of 6127 individual customers of a Middle Eastern bank has been used in the present study.<sup>1</sup> SMEs and liberal professionals have been excluded from the sample since the type of financial products they usually hold are of a different nature and they use their credit cards for different purposes and thus it was assumed that their consumption patterns which are the focus of this study, would be different from those of households. The selected customers are

<sup>1</sup> For confidentiality, the name of the bank is not revealed.

individuals who satisfy the following criteria: (1) They have at least one credit card; and (2) They have an active account.

While both internal and external data such as social network activity can be used to estimate the propensity of a customer to purchase a given financial product, only internal data has been used here. Three categories of internal data including (1) demographic data, (2) product ownership and (3) transaction data have been gathered. Demographic data and product ownership history are known to be the best predictors of future purchases in general (Bose & Chen, 2009). The objective of this study is to evaluate the extra predictive accuracy that can be obtained by using transaction data, and especially credit card data.

Other types of internal data that could have been used are customer contact history, marketing solicitation history, and web browsing activity. Unfortunately, these data sets are not available in an exhaustive manner, and so are not used in this study. External data such as social network activity have not been considered because of the inherent difficulties in collecting such data. Nevertheless, the purpose of this study is to provide empirical evidence in support of the hypothesis that credit card data adds extra accuracy to the cross-selling model, so there is no need to use a plethora of categories of input.

## 4 Summary of demographic data

In the banking business, demographic data usually refers to age, gender, family status, family size, age of children, relationship lifetime, profession, and income. However, in the datasets that have been made available, only age, gender and relationship lifetime were considered reliable enough to be used (see Table 1). The other types of demographic data either suffered from completeness problems or were not accurate enough and thus have been discarded.

Regarding age, in the country where the bank operates, the minimum legal age for a bank customer is 18 years. Because the bank operates in a developing country, the customer base has a large share of young adults (i.e., mean 41 years with min 19 years, see Table 1). Regarding income, it is evident that the average income in the sample is 60,179, with a max value of around 7 million US\$, showing that most customers have a medium to low income. Moreover, the average relationship lifetime of customers is 9.42 years; about 40% of the sample has a 5-year relationship lifetime or less, which reflects the aggressive customer acquisition plans run by the bank in the last years. Also, ATM withdrawals range from 10 to 800 transactions.

**Table 1** Descriptive statistics of demographic data

Variable	N	Mean	Std	Min	Max
Age	9676	41.69	14.23	19	96
Income	9834	60,179	178,499	0	7,361,014
Relationship lifetime	9676	9.42	8.03	0.25	74.33
ATM withdrawals	9175	1147	84	10	800

## 4.1 Summary of financial products held per customer and demographics

The main financial products are listed in Fig. 1. As mentioned earlier, they are considered to be good predictors of which products a customer will buy in the future.

### 4.1.1 Checking accounts

Checking accounts are the most widely known financial products. They can be split into two broad categories of Joint accounts and Individual accounts. This account can hold both local and foreign currencies. They also may have sub-accounts. While most customers hold one checking account, some customers maintain two or more accounts dedicated to different types of usage. For example, a customer may have a joint account with his spouse, an individual account for his own use and a sub-account for each one of his children.

### 4.1.2 Savings accounts, credit cards, loans

A saving account is an interest yielding account that cannot be used for check writing. Unlike savings plans, deposits and withdrawals can be made at any time on these accounts. Also, in this study, the term credit card is used in its widest sense and may even include what is known as debit cards. Credit cards come with different types and usage. However, in the country where the bank operates, the most common ones are: (1) Credit cards attached to checking accounts which can hold local or foreign currency; (2) Prepaid cards, which can be topped up with local currency or foreign currency (Euro or Dollar); (3) Credit cards attached to two accounts. These are mainly credit cards that use a local currency account when used locally and a foreign currency account when used abroad. Moreover, the main loans are Mortgage loan and Consumer loan. A mortgage loan is an instrument by which the bank lends money

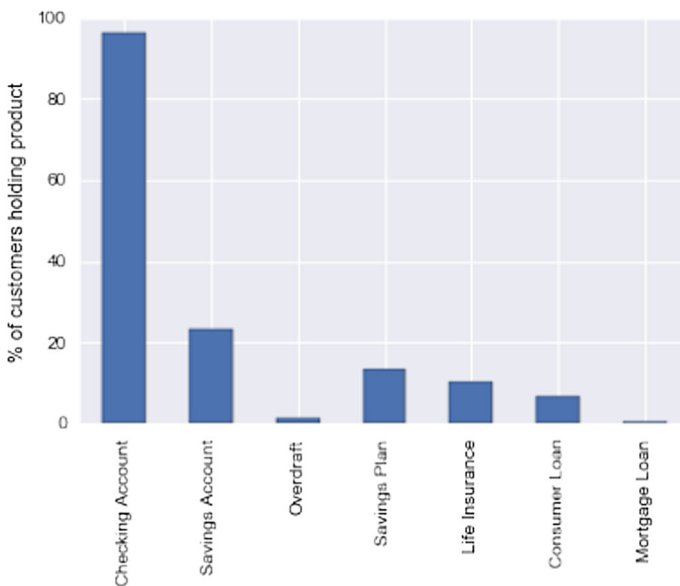


Fig. 1 The main financial products



to a customer for the purpose of purchasing a property. A special mechanism is set in place to allow the bank to recover the amount landed by selling the property (foreclosure) in case the customer defaults on the loan. Only a small proportion of the customer base in general holds a mortgage loan. But in the case of our sample, this ratio is even smaller (~ 0.3%). Consumer loans are contracted for a period of up to 7 years and do not require a collateral. However, in order to secure the amount landed, the bank requires that the customer buys a life insurance contract in which the bank is the beneficiary. Other account facilities include life insurance, savings plan, and overdraft. A life insurance is a product for which the insurance company, for which the bank is acting as an intermediary, is paid a premium on a regular basis in exchange of a lump sum that is paid to a designated person when the insured person either dies or has a major disability. Life insurance is a mandatory product for every loan holder, in which case, the insured person is the customer himself and the designated beneficiary is the bank issuing the loan. A savings plan is an investment instrument in which a customer contributes a fixed amount on a regular basis, typically through an automatic transfer from his checking account. The saved amount is paid an interest rate that is defined on a yearly basis with a guaranteed minimum. An overdraft is a form of credit where the bank allows the customer to continue withdrawing money even if the account has no funds up to a certain limit. The customer usually pays an interest rate for the total amount withdrawn. Tables 2 and 3 presents the descriptive statistics of customers' age and relationship lifetime per bank product, respectively.

Regarding gender, the percentage of male and female credit card users are 60.5% and 39.5%, respectively (see Fig. 2). However, while females are under-represented in the bank's customer base, they are on par with males when it comes to holding consumer loan. On average, a female has a higher probability of holding a consumer loan than a male. This fact makes females a better target for cross-selling consumer loans. The same phenomenon holds true for life insurance and is even more striking for overdraft and mortgage loans).

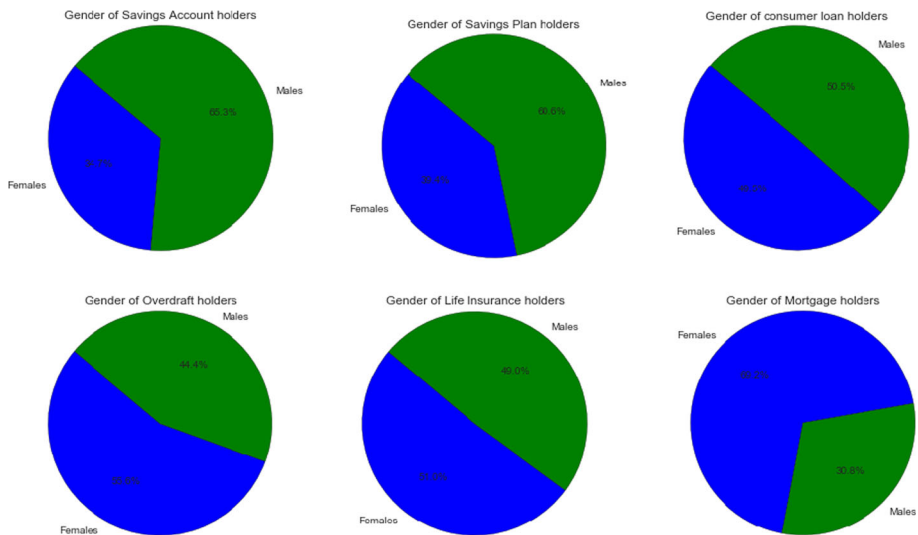
Focusing on income, the average number of cards held by a customer increases as compared to average income. It is also evident that the higher the income, the higher the number of credit cards. Customers with high income and only one credit card could be targeted

**Table 2** Descriptive Statistics of Age per product

Variable: age	N	Mean	Std	Min	Max
Customers without a Savings Plan	8339	41.22	14.61	19	96
Customers with a Savings Plan	1337	44.63	11.14	19	86
Customers without life insurance	8630	41.35	14.13	19	96
Customers with life insurance	1046	44.53	14.75	19	90
Customers Without overdraft	9503	41.58	14.26	19	96
Customers With overdraft	173	47.98	11.25	21	77
Customers Without Savings account	7534	40.91	14.24	19	96
Customers With savings account	2142	44.46	13.87	19	86
Customers Without consumer loan	9208	41.67	14.40	19	96
Customers With consumer loan	468	42.11	10.30	22	70
Customers Without mortgage	9656	41.70	14.24	19	96
Customers With mortgage	20	36.95	11.09	24	73

**Table 3** Descriptive Statistics of relationship lifetime per product

Variable: Relationship lifetime	N	Mean	Std	Min	Max
Customers without a Savings Plan	8339	8.73	7.95	0.25	74.33
Customers with a Savings Plan	1337	13.72	7.12	0.75	53.83
Customers without life insurance	8630	9.11	7.84	0.25	74.33
Customers with life insurance	1046	11.98	9.06	0.42	62.33
Customers Without overdraft	9503	9.31	8.00	0.25	74.33
Customers With overdraft	173	15.50	7.15	1.00	36.33
Customers Without Savings account	7534	8.39	7.69	0.25	74.33
Customers With savings account	2142	13.05	8.12	0.25	53.83
Customers Without consumer loan	9208	9.31	8.06	0.25	74.33
Customers With consumer loan	468	11.49	7.09	0.58	37.42
Customers Without mortgage	9656	9.42	8.03	0.25	74.33
Customers With mortgage	20	7.97	8.00	0.33	31.50

**Fig. 2** Percentage of male and female customers per product

for cross-selling credit cards. However, it is an interesting to see credit card usage tends to decrease with income (see Table 4).

#### 4.2 Account movement and credit card transactions data

Transaction data that has been made available for this study is of two types: (1) Monthly account movement data; and (2) Credit Card transaction data. These are the total monthly creditor and debtor movements. They have been used as approximations of monthly income

**Table 4** Number of credit cards per income range

Income range (\$)	Average number of cards held
(0, 20000]	1.248979592
(20,000, 30000]	1.338433843
(30,000, 40000]	1.350125945
(40,000, 50000]	1.443686007
(50,000, 60000]	1.439922481
(60,000, 70000]	1.453216374
(70,000, 80000]	1.490566038
(80,000, 90000]	1.502262443
(90,000, 100000]	1.61849711

and monthly spending. When summed over a period of a year, they yield the yearly income and the yearly spending which will be used as predictors in the rest of the paper.

#### 4.2.1 Credit card transaction data

The list of all credit card transactions (approximately 800,000) of all the customers in the sample, spanning a period of a year have been made available for this study. Each transaction is a record that is transmitted by the credit card company to the bank whenever a customer uses his credit card to purchase an item from a merchant (either a physical merchant such as a restaurant or an e-commerce merchant) or when they use their card to withdraw money from an ATM.

Every such transaction record is made of customer ID, date of transaction, merchant name (when applicable), amount of transaction both in local and foreign currency, transaction type (local or foreign ATM withdrawal, local merchant payment, foreign merchant payment...) and MCC (Merchant Category Code) is a universal classification code used by major credit card companies to encode the merchant activity type. It is a 4-digit code. Examples of MCC codes are: 5812 for eating places and restaurants; 5411 for supermarkets; and 5651 for clothing stores.

MCCs are a valuable and rich source of information because they give a very detailed description of the customer consumption behavior. There are thousands of MCC codes, but in order to make the data easier to interpret in this study, they have been grouped into 22 categories and 146 sub-categories.

#### 4.3 Variables extracted from the credit card dataset

From the list of transactions mentioned above the following variables have been computed for every customer, for the year under study: Credit card usage, Total money spent through the credit card, Total number of shops used, Total ATM withdrawals, Mean ATM withdrawal, Number of ATM withdrawals, Standard deviation of ATM Withdrawals, ATM usage skewness; Number of sub-categories purchased from; For every category of goods and services: The total amount of money spent on the category, and The number of shops used by the customer to purchase those goods and services. For every sub-category: The amount of money spent on the sub-category.

### 4.3.1 Credit card usage and ATM ratio

Credit card usage is defined as the yearly amount of credit card payments made by the customer divided by the customer's total yearly expenses (total yearly debit movements). It is the proportion of payments for which a credit card is used. Moreover, some customers use cash more than others for their daily purchases. This customer characteristic can be described by using two variables of ATM withdrawal frequency and ATM withdrawal amount. The ratio of the mean ATM withdrawal to the yearly income is called 'ATM ratio' and will be used as a predictor in this study.

## 4.4 The main variables

In this section, the impact of the main variables on the propensity to hold some banking products are discussed. The products that are considered in this case are Checking accounts, Saving accounts, Saving plans, Life Insurance, Overdraft, Consumer loan and Mortgage Loan. However, a special focus has been made on Consumer Loan.

### 4.4.1 Number of sub-categories used

Customers buy goods and services of different types. However, some tend to buy from a wider variety of goods and services than others, which can be related to purchasing power, lifestyle and personality traits. In the present study, the diversity of goods and services purchased is represented by the number of sub-categories (maximum of 146) for which the customer used his credit card. This variable represents a facet of the customer consumption habits and thus will be used as a predictor. Figure A1 in Appendix shows the distribution of the number of sub-categories used by customers in the sample. Figure A2 shows in Appendix that when income increases, the number of categories of goods and services the customer spends money on increases too.

For the first half of the curve, the number of sub-categories purchased from varies almost linearly, so it makes sense to find out whether this linearity is the same for consumer loan holders and non-holders. This is accomplished by computing the ratio 'Nbr Sub-Categories Used'/'Yearly Income' that we will call from now on "Consumption ratio". Figure A3 in Appendix shows how this variable behaves for consumer loan holders and for the rest of the sample. It appears that on average, at the same level of income, customers who have a consumer loan purchase from a smaller number of sub-categories than those who don't have a consumer loan. This ratio will also be used as a predictor.

### 4.4.2 Spending per category

It is obvious that customers spend more on some categories of products and services. The percentage of income spent on a given category of goods is an important component of the consumption behavior of a customer and will be used as a predictor. Figure A5 in Appendix presents the percentage of credit card spending for the main categories within the sample. Figure A7 in Appendix shows how the yearly food expenses change with income. The curve seems to be linear up to about \$80,000 and then plateaus. The ratio Food Expenses / Yearly Income which we will call Food Ratio is not distributed in the same manner for both consumer loan holders and the rest of the sample. Consumer loan holders tend to spend a smaller percentage of their income on food than the rest of the customer sample. So, the Food

Ratio will be used as a predictor. The same observation holds true for clothing and beauty related expenses.

#### 4.4.3 Number of shops used

Customers also have different habits when it comes to the number of shops, they use to purchase a given product or service. For example, some customers go to the same small group of restaurants while others try to constantly discover new ones. Such consumption habits may yield information on a customer personality and thus on his propensity to buy a given financial product. Figure A1 in Appendix shows the distribution of the number of shops used by the customers in the sample. As can be seen from the curve in Figure A1, the number of shops used by a customer increase with income almost in a linear manner up to a certain point and then saturates. However, at the same level of revenue, consumer loan holder tends to shop from a smaller number of shops than the rest of the sample as shown in Figure A3 in Appendix. The ratio 'Number of shops'/'Yearly Income' will be referred to as the shops ratio and will be used as a predictor. Savings plans predictors.

Figure A8 shows the relationship between age and the probability of having a savings plan. It is not surprising to see there is a sharp fall after age 60 as most customers tap into their savings after retirement. The curve shows how the probability of having a savings plan increases sharply with income.

## 5 Discussion of findings and implications

### 5.1 Results of experiment 1

As mentioned above, we will try to prove that transaction data adds a significant amount of predictive accuracy to the cross-selling model by measuring the AUC of the classifiers when they are fed with data sets A, B and C. This is first done with the Random Forest (Table 5) and the Gradient Boosting Machine (Table 6) for which the number of trees is varied manually from 50 to 100. This was done in order to prove that the results are stable when the configurations of the classifiers are changed.

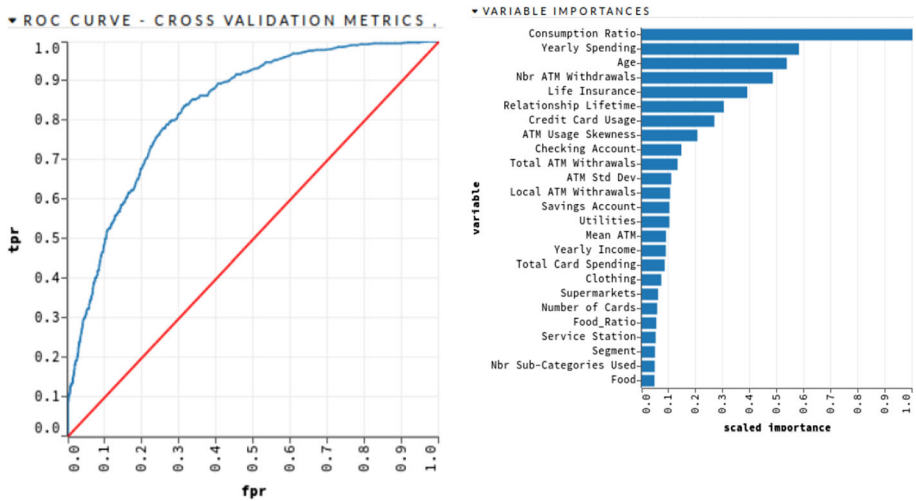
The numbers in red are the maximum values of AUC when using a given data set. They show that the predictive power obtained by using transaction data is about the same as the one obtained when using demographics and product ownership. It is even more interesting to see by combining both transaction data, demographics, and product ownership, we get a 4% increase of AUC in the case of Random Forests and 5.4% in the case of Gradient Boosting Machine. The graph on the right-hand side shows the relative importance of each

**Table 5** AUCs of Random Forests

	Number of trees					
	50	60	70	80	90	100
A = Demographics and Products	0.7779	0.7758	0.7764	0.7785	0.784	<b>0.7866</b>
B = Transaction data	0.7769	0.775	0.7726	0.7737	<b>0.7816</b>	0.781
C = A + B	0.8121	0.8146	0.8127	<b>0.8176</b>	0.8078	0.8148

**Table 6** AUCs of Gradient Boosting Machines

	Number of trees					
	50	60	70	80	90	100
A = Demographics and Products	0.7803	0.7854	0.7828	0.7841	<b>0.7895</b>	0.7814
B = Transaction data	<b>0.7877</b>	0.7789	0.7807	0.7743	0.7825	0.7789
C = A + B	0.823	0.8291	0.8272	<b>0.8323</b>	0.8266	0.8204

**Fig. 3** Relative importance of each predictor for the Gradient Boosting Machine

predictor for the Gradient Boosting Machine (Fig. 3). The most important predictor is the ‘Consumption Ratio’, which is the number of sub-categories used by the customer divided by his yearly income. It gives an idea about how varied are his purchases compared to his level of income.

There is a negative correlation between the Consumer Loan dependent variable and this predictor, which means that customers who buy highly varied types of products relatively to their income, most likely won’t acquire a consumer loan. It is also noticeable that the number of ATM withdrawals per year, ATM usage skewness within a month, the yearly amount of ATM withdrawals and ATM withdrawals standard deviation are among the top predictors. Hence, ATM usage contains valuable information on the propensity of a customer to acquire a consumer loan. For ANN, CHAID and SVM, the results are summarized in the Table 7.

CHAID provides the highest values of AUC for each dataset. With both CHAID and SVM, there is more predictive value in transaction data than in Demographics and Product Ownership. However, the heterogeneous ensemble depicted above yields far more interesting results. Table 8 presents a summary of the ensemble performance for all three datasets. The number of bags is the number of instances of Neural Networks and CHAID trees in the ensemble.

**Table 7** AUC for ANN, CHAID and SVM

	ML Algorithm		
	ANN	CHAID	SVM
A = Demographics and Products	0.727	0.799	0.624
B = Transaction data	0.688	0.81	0.774
C = A + B	0.711	0.826	0.824

**Table 8** AUC for Heterogeneous ensembles

	Number of bags					Average
	5	10	15	20	25	
A = Demographics and Products	0.847	0.843	0.849	0.849	0.848	0.847
B = Transaction data	0.906	0.907	0.91	0.91	0.908	0.908
C = A + B	0.914	0.932	0.927	0.925	0.925	0.925

The results suggest two interesting points. (1) With the heterogeneous ensemble, the AUC obtained from the transaction data is far higher than that obtained from demographics and product ownership data. This was not the case with the other classifiers; and (2) The AUC obtained when the ensemble of ten bags (let's call it algorithm X) is fed with all data sets is 0.932, that is 0.100 point higher than the highest AUC obtained so far, which is the one obtained from the Gradient Boosting Machine with 80 trees (AUC = 0.832). This is a huge leap in performance that is due to both the information contained in the transaction data and to the strong predictive performance of the heterogeneous ensemble.

## 5.2 Results of experiment 2

Now that it is clear that the heterogeneous ensemble with ten bags (Algorithm X) is the most powerful classifier for the subject at hand, let's add the 8 features produced by the auto-encoder to the list of inputs and see how the AUC behaves. In order to confirm that there is valuable information in these 8 features, a sixfold cross-validation has been run with algorithm X with only those features as input. It produced an AUC of 0.777 (Table 9).

**Table 9** AUC of Heterogeneous Ensemble with features from Auto-encoder

	Number of bags					Average
	5	10	15	20	25	
C = Demographics + Products + Transactions	0.914	0.932	0.927	0.925	0.925	0.925
D = C + 8 Features from auto-encoder	0.925	0.943	0.93	0.934	0.933	0.933

Adding the 8 features produced by the auto-encoder adds on average 0.008 to the AUC. This is a modest but valuable enhancement.

### 5.3 Practical implications

As mentioned earlier, unlike Amazon and other big tech companies, when the customer makes a bank transaction, they are unlikely to receive any recommendation from the bank. Banks struggle with customer analytics since they have a great amount of data, but they also have old databases and silos that do not want to share customer data with one another, and they have difficulty in identifying the best ways to use the data (Crosman, 2017). If banks and financial institutions want to provide the same type of personalized insight that big tech companies do and compete with fintech companies, they need to adopt artificial intelligence since analytics and AI for marketing and cross-selling are key success factors (Crosman, 2017).

Since banks have made cross-selling a strategic choice and cross-selling is crucial to banks, it needs to be done efficiently. When the bank contacts a customer to offer them a product, it must be sure that the given product fulfills their needs and the solicitation comes at the right time. Past experience shows when a customer is contacted unsuccessfully for offers too often, these solicitations can become a turn off and tend to weaken the relationship.

Hence there are two main advantages of using the proposed model: (1) Banks can contact customers periodically for offers is not only a way to accumulate information on them but it is also known that these solicitations tend to strengthen the relationship between the customer and the bank and thus decreasing the probability of churn. (2) When a customer purchases many products from their bank, it becomes such a hassle to cancel all of them and negotiate new contracts with another bank and most of the time customers do not do it. Hence, this improved version of cross-selling prediction can be used as a way to reduce churn. The limitation of this analysis is that we need to access real-time data which sometime banks limit us to access real-time data.

All models used in both experiments in our study strongly support the hypothesis that there is significant predictive power embedded in the customer transaction data and especially in the credit card data. The use of Credit Card data can significantly enhance the predictive accuracy of a cross-selling model in retail banking. Heterogeneous ensembles seem to have a significantly higher predictive accuracy than single classifiers and the size of the ensemble seem to have a modest impact on performance beyond a certain point. Deep Learning techniques such as auto-encoders can contribute to the overall model performance.

## 6 Conclusion

While most cross-selling models in retail banking use demographics and interactions with marketing as predictors, this paper introduced the use of account transaction data and credit card transactions in particular and demonstrated their ability to increase the performance of such model. Some ratios were extracted from credit card transactions and their correlation with the propensity to hold a consumer credit was established. We showed how these derived variables contain valuable information on the customer's consumption behavior, as their importance as predictors was among the highest. The analysis of these variables suggests customers who acquire a consumer loan are most likely to be on a tight budget. This hypothesis



needs to be confirmed by future research and a closer look at the whole set of account transactions may prove to be helpful for that matter.

Besides the credit card transactions, no other types of transactions (for example transfers, check) were made available to this study. These other types, after being aggregated, could be used in the future as predictors and may increase the predictive accuracy of the model. Several types of classifiers were used. The relative importance of transaction data as a predictor depended on the type of classifier, but overall proved to be at least as good as that of demographics and product ownership together. The heterogeneous ensemble had, by far, the highest performance and showed that transaction data could have an even higher predictive power.

Finally, an auto-encoder was used to extract some features from the transaction data, which then were used as predictors by the heterogeneous ensemble. We showed that they contain relevant information and that they can increase the performance of the classifier even further. In this study, only one auto-encoder was used. However, deep architectures such as stacked auto-encoders and convolutional networks have already been used successfully in other customer relationship management areas such as churn prediction (Wangperawong et al. 2016), so their use in the case of cross-selling could be beneficial and represent an interesting direction for future research.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s10479-023-05209-5>.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Akter, S., Michael, K., Uddin, M. R., McCarthy, G., & Rahman, M. (2020). Transforming business using digital innovations: the application of AI, blockchain, cloud and data analytics. *Annals of Operations Research*, 308, 7–39.
- Birim, S., Kazancoglu, I., Mangla, S. K., Kahraman, A., & Kazancoglu, Y. (2022). The derived demand for advertising expenses and implications on sustainability: A comparative study using Deep Learning and traditional machine learning methods. *Annals of Operations Research*. <https://doi.org/10.1007/s10479-021-04429-x>
- Bose, I., & Chen, X. (2009). Quantitative models for direct marketing: A review from systems perspective. *European Journal of Operational Research*, 195, 1–16.
- Brown, S., & Taylor, K. (2014). Household finances and the 'Big Five' personality traits. *Journal of Economic Psychology*, 45, 197–212.
- Chaudhuri, N., Gupta, G., Vamsi, V., & Bose, I. (2021). On the platform but will they buy? Predicting customers' purchase behavior using Deep Learning. *Decision Support Systems*, 149, 113622.
- Crosman, P. (2017). Can AI make banks as good as Amazon at knowing customers? *American Banker*, 182(18), 1.
- Forough, J., & Momtazi, S. (2021). Ensemble of deep sequential models for credit card fraud detection. *Applied Soft Computing*, 99, 106883.
- Gunnarsson, B. R., van den Broucke, S., Baesens, B., Óskarsdóttir, M., & Lemahieu, W. (2021). Deep learning for credit scoring: Do or don't? *European Journal of Operational Research*, 295, 292–305.

- Hebden, J. J., & Pickering, J. F. (1974). Patterns of acquisition of consumer durables. *Oxford Bulletin of Economics & Statistics*, 36, 67–94.
- Jauhar, S. K., Raj, P. V. R. P., Kamble, S., Pratap, S., Gupta, S., & Belhadi, A. (2022). A Deep Learning-based approach for performance assessment and prediction: A case study of pulp and paper industries. *Annals of Operations Research*. <https://doi.org/10.1007/s10479-022-04528-3>
- Kamakura, W. A. (2007). Cross-selling: Offering the right product to the right customer at the right time. *Journal of Relationship Marketing*, 6, 41–58.
- Kamakura, W. A., Ramaswami, S. N., & Srivastava, R. K. (1991). Applying latent trait analysis in the evaluation of prospects for cross-selling of financial services. *International Journal of Research in Marketing*, 8, 329.
- Kamakura, W. A., Wedel, M., De Rosa, F., & Mazzon, J. A. (2003). Cross-selling through database marketing: A mixed data factor analyzer for data augmentation and prediction. *International Journal of Research in Marketing*, 20, 45–65.
- Kim, J., Ji, H., Oh, S., Hwang, S., Park, E., & del Pobil, A. P. (2021). A deep hybrid learning model for customer repurchase behavior. *Journal of Retailing and Consumer Services*, 59, 102381.
- Koehn, D., Lessmann, S., & Schaal, M. (2020). Predicting online shopping behaviour from clickstream data using Deep Learning. *Expert Systems with Applications*, 150, 113342.
- Kriebel, J., & Stitz, L. (2021). Credit default prediction from user-generated text in peer-to-peer lending using Deep Learning. *European Journal of Operational Research*. <https://doi.org/10.1016/j.ejor.2021.12.024>
- Ładyżyński, P., Żbikowski, K., & Gawrysiak, P. (2019). Direct marketing campaigns in retail banking with the use of Deep Learning and random forests. *Expert Systems with Applications*, 134, 28–35.
- Lessmann, S., Baesens, B., Seow, H.-V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research*, 247, 124–136.
- Li, S., Sun, B., & Montgomery, A. L. (2011). Cross-selling the right product to the right customer at the right time. *Journal of Marketing Research*, 48, 683–700.
- Mahbobi, M., Kimiagari, S., & Vasudevan, M. (2021). Credit risk classification: an integrated predictive accuracy algorithm using artificial and deep neural networks. *Annals of Operations Research*. <https://doi.org/10.1007/s10479-021-04114-z>
- Mulyanegara, R. Z., Tsarenko, Y., & Anderson, A. (2009). The Big Five and brand personality: Investigating the impact of consumer personality on preferences towards particular brand personality. *Journal of Brand Management*, 16, 234–247.
- Najafabadi, M. M., Villanustre, F., Kkshgofaar, T. M., Seliya, N., Wald, R., & Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. *Journal of Big Data*, 2, 1–21.
- van der Gaast, J. P., & Weidinger, F. (2020). A deep learning approach for the selection of an order picking system. *European Journal of Operational Research*. <https://doi.org/10.1016/j.ejor.2022.01.006>
- Wetzel, T. (2019). Going digital and building relationships. *Insurance Journal*, 97(18), 54.
- Zhang, X., Han, Y., Xu, W., & Wang, Q. (2021). HOBA: A novel feature engineering methodology for credit card fraud detection with a Deep Learning architecture. *Information Sciences*, 557, 302–316.
- Zinovyeva, E., Härdle, W. K., & Lessmann, S. (2020). Antisocial online behavior detection using Deep Learning. *Decision Support Systems*, 138, 113362.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.