JISIB
Journal of Intelligence Studies in Business
Vol. 12 No. 2 (2022) pp. 54–65
Open Access: Freely available at: https://ojs.hh.se/

Stevan Didijer
1911-2004

Application of Business Intelligence in Decision Making for Credit Card Approval

Admel Husejinovic*

International Burch University, Bosnia and Herzegovina
Email: hadmel@hotmail.com

Nermina Durmić

International Burch University, Bosnia and Herzegovina
Email: nermina.durmic@ibu.edu.ba

Samed Jukić

International Burch University, Bosnia and Herzegovina
Email: samed.jukic@ibu.edu.ba

Received 23 November 2022 Accepted 16 December 2022

ABSTRACT This paper aims to show how business intelligence can be applied in the credit card approval process. More specifically, the paper investigates how information like an applicant's age, credit score, debt, income, and prior default can be used in credit card approval prediction. The dataset used for analysis is a publicly available dataset from the UCI machine learning repository. Logistic regression is used to make a prediction model with a reasonable number of attributes for a comprehensible business model. The Chi-square test of independence is used to test the dependence of credit card approval results with attributes. Research uncovers that prior default is supposed to be the most important attribute in the approval process. Finally, the authors propose several visualizations that could help make smarter decisions with effective credit risk assessment.

KEYWORDS: Business Intelligence, Chi-Square Test, Credit Card, Data Analysis, Linear Regression

1. INTRODUCTION

Starting in the United States of America in 2007 worldwide Global Financial Crisis (GFC) was the most serious economic crisis since the Great Depression (Grant and Wilson, 2012). Started with global housing market shocks and spread to other submarkets subsequently. It resulted in extreme credit losses in many countries worldwide (Uppal and Ullah Mangla, 2013). The credit decision process involves practice, judgment, and many analytical and risk-assessing techniques for determining the probability that money is going to be repaid in an equal amount and expected time (Brown and Moles, 2014). Although banks struggle with competition in meeting the set goals related to granting loans, they are obliged not to expose themselves unreasonably to credit risk. The credit card industry is rapidly rising with approximately 2.8 billion credit cards in use worldwide (Infographic, 2021).

* Corresponding Author

Business intelligence (BI) tools and applications are used as decision support in many bank activities (Ubiparipovi and Durkovic, 2011). Banks collect and store lots of data from different sources trying to use it in credit card approval decision-making. This study aims to provide insights into data used in credit card approval to provide answers such as whether information such as the client's gender, age, credit score, and debt balance can be used to predict whether a random client whose data we have will get credit card approved.

The paper is organized as follows: the next section presents the background of the business intelligence used in the credit card approval process and research questions. In section 3 research methodology with dataset and data analysis methods were used to answer the research questions. In section 4 data analysis methods with parts of python code are presented. Research results are introduced in section 5 with visualization graphics for easier understanding. The section is about the conclusions and the accomplished results of this research.

2. BACKGROUND AND LITERATURE REVIEW

2.1 Business Intelligence

The main goal of Business Intelligence in a company is to support management in making smarter business decisions (Rafi and S.M.K, 2012; Balachandran and Prasad, 2017; Alzeaideen, 2019). Business Intelligence (BI) indicates all activities which help collect, store, and analyses data produced by a company (Ćurko, Bach and Radonić, 2007). More specifically, BI refers to activities like data warehousing, descriptive analytics, data mining, performance benchmarking, predictive analytics, and reporting (Mehanović and Durmić, 2022). Data warehousing, as part of BI, incorporates different algorithms, tools, and architecture that bring together data and information from different sources into a specific repository (Widow, 1992). Performance benchmarking can be explained as gathering data about a company's products and services and comparing it with competitors' products and services, using different techniques and methods (Bogetoft, 2012). Data mining is a process of sorting through large data sources using different algorithms to find interesting data patterns useful for business needs (Bramer, 2016). Finally, reporting represents gathering and displaying data in charts and tables, while interpretation and giving context to data gained from reports is considered data analysis (King, 2022).

2.2 Credit Cards

Starting from 1949 and McNamara's "Forgotten Wallet" story, which is known as the beginning of the first credit card provider Diners Club, today there have been more than 2,8 billion credit cards issued worldwide (Brian, 2020). Nowadays, Visa, Mastercard, and American Express are among the most popular credit card brands (Shift, 2021). They represent credit card associations responsible for handling payment networks. These companies are responsible for processing the payment from the merchant. As Jason (2018) explains, the issuers of credit cards are usually banks or credit unions and they benefit from the interest rates, but they also accept the credit card owner's credit default risk. A credit card owner is usually a person or a small business. As shown in the credit card payment process (Tabul8tor, 2019), the credit card business involves different players like card issuers, card holders, merchants, acquirers, and card associations. A card issuer, usually a bank, after the credit card approval process issues a credit card to the customer with a certain credit limit. A credit cardholder uses a credit card for the payment of goods or services at the merchant's place. At the POS (point of sale) cardholder presents the card merchant inserts the card into the terminal and several authorization requests are sent through acquiring bank, payment network, and card issuer. The card issuer will validate whether the cardholder has an available credit limit for the payment. Upon receipt of the authorization message, the merchant issues a payment receipt to the credit cardholder. After, successful payment is realized card association calculates the settlement obligation fee to the card issuer. The card issuer takes money from the credit cardholder's account and sends it to the card association. Card association sends money to the acquiring bank after charging a fee. Acquiring bank finally sends money to the merchant after charging their fee.

Credit cards are more suitable for various payment types, like online payments, than cash. Also, there are usually some benefits for certain purchases. On the other hand, there are APR (annual percentage rates) that refer to interest that will be applied from the issuer bank to the debit the credit card owner's account (Bridges, 2022). In America, the average debt per credit card is USD 5.221 (Constance, 2022). Credit cards fall into the revolving credit type of borrowing (Greg, 2021). That means that the bank gives access to individual limits of funds that are available to the credit card owner as he/she repays its monthly installments regularly (Dieker, 2022).

When a person applies for a credit card, the bank is going to check some important aspects of the applicant such as employment history, regular income details, and regularity of any other bank products. Some countries have national credit scoring like American FICO (Luthi and Karp, 2021). The authors explain that banks use FICO scores to check for credit card application approval. If your score is sufficient your application would be approved based on the FICO score result. Others, on the other hand, have developed their credit scoring models for credit card approval assistance. In modern times there are sufficient data gathered in databases, and those data are being used with different tools to show insides, patterns, and information to help business users make smarter decisions (Siddiqi, 2012).

2.3 Business Intelligence Application in the Credit Card Approval

The fundamental role of banks in the economic system is to collect money from economic units with money surplus and then lend it to the units that lack money (Gobat, 2012). In return, the bank charges interest and fees and thus makes money. By doing this work, banks expose themselves to credit risks (failure to collect money). Since 1970, the credit scorecard has been used in financial credit risk assessment (Peussa, 2016). A credit risk scorecard is a tool that uses predictive models to evaluate risks associated with applicants or customers (Siddiqi, 2012). In simple words, credit scorecards help bank clerks to identify the statistical probability of credit default. Scorecard attributes are selected from the applicant's personal information available to the bank. To collect the data for the scorecard model bank may use different sources. Personal data like gender, age, marital status, and zip code is collected from the applicant's application form data like time at a bank, the number of bank products used by the client, and payment performance are collected from records describing previous experience with the bank. Credit biro or Central bank also collects data about clients' debit history, trades, or public records. For a better assessment of credit risk, banks expanded datasets with possibilities of Big Data technologies (Ghobadi & Rohani, 2017; Pérez-Martín, Pérez-Torregrosa, & Vaca, 2018). Banks use data warehouses or data marts as they need reliable and clean data for credit scorecards (Siddiqi, 2012). Data marts are subject-oriented databases with a more narrow scope than a data warehouse, focusing on some application or company division work scope (Talend, no date; Watson and Wixom, 2007). As a result of the growing popularity of Artificial Intelligence, machine learning algorithms, such as Ensemble and Hybrid models with neural networks and SVM, are being implemented for credit scoring, a decrease in non-performing assets, and payment fraud (Bhatore, Mohan and Reddy, 2020).

Business intelligence helps many aspects of bank management like assets and liability management, risk management, performance management, and decision making (Ubiparipovi and Durkovic, 2011; Nithya and Kiruthika, 2021). Business intelligence with artificial intelligence and machine learning techniques provides better performance and more efficient solutions for decision-making (Aruldoss *et al.*, 2014).

2.4 Research questions

This paper tackles the relevance of certain attributes in credit card approval predictions using the Logistic regression method. More specifically, the question that this research aims to answer is: Are attributes, such as clients' age, gender, credit score, debt balance, and their occupation significant in the credit card approval process?

3. RESEARCH METHODOLOGY

3.1 Dataset

Dataset used in this research is a publicly available dataset on credit card applications downloaded from the www.kaggle.com website (Cortinhas, 2022). Dataset has 15 attributes and one class attribute. As Table 1 shows, attributes are presented with attribute names, detailed information, and data type columns. In the dataset, there are 690 entries of which 307 are in credit card approved status while 383 are in not approved status. Data types of dataset attributes are numerical and categorical.

Table 1. Dataset Attributes.

Attribute Name	Information	Data type
Gender	0 = Female, 1 = Male	Int64
Age	Age in years	float64
Debt	Outstanding debt (feature has been scaled)	float64
Married	0 = Single/Divorced/etc., 1 = Married	int64
Bank Customer	0 = does do not have a bank account, 1 = has a bank account	int64
Industry	job sector of current or most recent job	object
Ethnicity		object
Years employed	No. of Years Employed	float64
Prior default	0 = no prior defaults, 1 = prior default	int64
Employed	0 = not employed, 1 = employed	int64
Credit score	the feature has been scaled	int64
Driver's license	0 = no license, 1 = has license	int64
Citizen	either By Birth, By Other Means, or Temporary	object
Zip Code	(5-digit number)	int64
Income	the feature has been scaled	int64
Approved	0 = not approved, 1 = approved	int64

Note: Table 1 shows data attributes used in dataset. Tabe has tree columns: attribute name, information with possible values that attribute might have, and data type.

3.2 Data Analysis Methods

Python was used for data analysis in this research, as it offers stable numerical libraries with great quality in open-source documentation access. More precisely, the authors used the pandas' library for financial data manipulation and analysis (McKinney, 2009; Kibria and Sevkli, 2021).

Data from the dataset in this research is analyzed by exploring attributes, making meaningful categories, the testing relationship between them, and visualization. Attributes that have less significance for the model are removed to make the model less complex and easier for business use.

Binary logistic regression was used for application approval predictive model creation. Logistic regression describes the relationship between dependent variable Y and independent variables x-s as a function of $ln\left(\frac{p(x)}{1-p(x)}\right) = \beta_0 + \beta_1 * x_1 + \beta_2 * x_2 + \cdots + \beta_n * x_n$. Results ranging between values 0 and 1 are obtained after applying the sigmoid function $(x) = \frac{1}{1+e^{-x}}$. If the resulting probability is less than 0,5, the expected resulting dependent variable is predicted to be 0. On the other hand, if the result is greater than 0,5, the dependent variable is predicted to be 1 (Maalouf, 2011; Bramer, 2016). Logistic regression is performed to predict whether a client's credit card application is approved or not.

4. DATA ANALYSIS

4.1 Data Preparation

Initially, the dataset had 16 fields and all fields are used in preprocessing stage. Fields like ethnicity, industry, and citizen are string values and those are normalized across the dataset. Missing values are replaced with the attribute's mode for gender, marital status, ethnicity, industry, and Zip code shown in Table 2. Below is a piece of python code used for removing some attributes and checking for missing values.

```
# Missing values
mv_df=pd.DataFrame(columns = ['Attribute', 'No.Missing'])
for col in df2.columns:
    mv_df=mv_df.append({'Attribute':col, 'No. Missing':(df2[col]=='?').sum()}, ignore_index=True)
# Show data
mv_df
```

Table 2. Attributes with missing values.

	Attribute	No. Missing values
1	Gender	12
2	Age	12
3	Debt	0
4	Married	6
5	Bank Customer	6
6	Industry	9
7	Ethnicity	9
8	Years Employed	0
9	Prior Default	0
10	Employed	0
11	Driver License	0
12	Citizen	0
13	ZipCode	13
14	Credit Score	0
15	Income	0
16	Approved	0

Note: Table 2 shows number of missing values for the dataset attributes.

Missing values for gender are replaced using the attribute's mode value.

```
#Gender
df ['Gender'].value_counts()
# Replace missing value with attribute mode
value
df['Gender']. replace ('?', 1, inplace=True)
```

Missing values for the age variable are replaced using the attribute's median value.

```
# find median age
age_medean=df.loc[data['Age']!='?','Age'].
median()
print('age median:' age_medean)

# Set missing values with the median value
df.loc[df['Age']=='?','Age']= age_medean
```

4.2 Data Analysis

This section presents the analysis of data conducted to check the insides of collected applications. First, the Age attribute is grouped into 4 age groups: teens (up to 10 years old), adults (20–39 years old), middle-aged adults (40–59 years old), and senior adults (over 60 years old). Below is a block of code used to generate the age categories and create a bar chart.

```
for i, row in df1.iterrows():
    if df1.loc[i, 'Age']< 20:
        df1.loc[i, 'Age'] = 'teens'
    elif df1.loc[i, 'Age']>=20 and df1.loc[i,
'Age']<40:
        df1.loc[i, 'Age'] = 'adults'
    elif df1.loc[i, 'Age']>=40 and df1.loc[i,
'Age']<60:
        df1.loc[i, 'Age'] = 'middle age adults'

    elif df1.loc[i, 'Age']>=60:
        df1.loc[i, 'Age'] = 'senior adults'
#visualizing bar chart
(df1.groupby('Age')['Approved'].value_
counts(normalize=True)
  .unstack('Approved').plot.bar(stacked=True))
```

Next, the credit score groups of applicants are analyzed. Groups are based on the credit score results attribute. Since credit score is numerical variable categories are performed by grouping data based on credit score distribution. Below is a piece of code used to generate Credit score categories and visualize the categories' distribution for credit card approvals.

```
for i, row in df1.iterrows():
    if df1.loc[i, 'CreditScore']>=0 and df1.loc[i,
'CreditScore']<2:
        df1.loc[idx, 'CreditScore'] = 'Bad'
    elif df1.loc[i, 'CreditScore']>=2 and df1.loc[i,
'CreditScore']<4:
        df1.loc[i, 'CreditScore'] = 'Fair'
    elif df1.loc[i, 'CreditScore']>=4 and df1.loc[i,
'CreditScore']<6:
        df1.loc[i, 'CreditScore'] = 'Good'
    elif df1.loc[i, 'CreditScore']>=6:
        df1.loc[i, 'CreditScore'] = 'Excellent'
# approval cases with group by credit score
category
(df1.groupby('CreditScore')['Approved'].value_
counts(normalize=True)
  .unstack('Approved').plot.bar(stacked=True))
```

Next, the client's debt groups are analyzed. Groups are based on the debt balance attribute. Debt is also a numerical variable and categories are grouped by debt balance distribution. Below is a piece of code used to generate debt balance categories and a bar chart that visualize the credit card approval grouped by debt categories.

```
for i, row in df1.iterrows():
   if df1.loc[i, 'Debt']>=0 and df1.loc[i,
'Debt']<1:
       df1.loc[i, 'Debt'] = 'Very Low'
   elif df1.loc[i, 'Debt']>=1 and df1.loc[i,
'Debt']<2.75:
       df1.loc[i, 'Debt'] = 'Low'
   elif df1.loc[i, 'Debt']>=2.75 and df1.loc[i,
'Debt']<7:
       df1.loc[i, 'Debt'] = 'Medium'
   elif df1.loc[i, 'Debt']>=7:
       df1.loc[i, 'Debt'] = 'High'
 # According to Boxplot distribution
# approval cases with a group by credit score
category
(df1.groupby('CreditScore')['Approved'].value_
counts(normalize=True)
  .unstack('Approved').plot.bar(stacked=True))
```

4.3 Chi-Square Independence Test

A Chi-square test (x^2) used in statistics for hypothesis testing when variables are nominal (Dahiru, 2013; Mchugh, 2013). Test measures how far the model frequency outcomes vary from expected results if the null hypothesis were correct (Bramer, 2016). The Chi-square test depends on the size of the difference between actual and observed values, the degrees of freedom, and the highest number of possible independent values outcomes (Hayes, 2022). The formula for the chi-square test is as follows:

$$X_c^2 = \sum \frac{(O_i - E_i)}{E_i} \quad (1)$$

Where:
c = degree of freedom
O = Observed values
E = Expected values

Hypothesis

$H_0: Variables\ are\ independent$

$H_1: Variables\ are\ not\ independent$

If calculate p-value < 0,05 H_0 is rejected. If p-value > 0,05 we fail to reject H_0 and H_1 is accepted.

Pandas crosstab and scipy.stats.chi2_contingency libraries (Sphinx, 2008; Pandas, 2022) are used to perform the chi-square test in python. Below is a piece of python code that generates a cross-tabulation table for the chi-square test of independence with scipy.stats python library.

```
crosstabAge = pd.crosstab(df1["Age"], df1["Ap-
proved"])
crosstabGender = pd.crosstab(df1["Gender"],
df1["Approved"])
crosstabDebt = pd.crosstab(df1["Debt"],
df1["Approved"])
crosstabCreditScore = pd.crosstab(df1["Cred-
itScore"], df1["Approved"])
import scipy.stats as stats
stats.chi2_contingency(crosstabAge)
stats.chi2_contingency(crosstabGender)
stats.chi2_contingency(crosstabCreditScore)
stats.chi2_contingency(crosstabDebt)
```

4.4 Logistic Regression

The logistic regression is applied in a predictive model using statsmodels (*statsmodels*, no date) python library. The attribute "approved status" with possible results 0 (not approved) and 1 (approved) is dependent variable Y. Independent variables are the client's: gender, age, debt balance, marital status, being already a bank customer or new applicant status, working job industry classification, ethnicity, work experience in years, the record of previous default, employment status, credit score, the position of driver license, citizenship status, resident address zip code, and income amount. In the resulting predictive model variables with a probability value greater than 0,05 are excluded, as those values are not significant for the model (Dahiru, 2013). After the data analysis process for the creation of the predictive model 6 attributes ('Gender', 'Driver License', 'Zip Code', 'Ethnicity', 'Citizen', and 'Industry') are removed as irrelevant for the model. Below is a python code for the creation of the Logistic regression model and the removal of irrelevant attributes using statsmodels python library.

```
#Remove irrelevant data attributes
df2.drop(['Gender', 'DriversLicense', 'ZipCode',
'Ethnicity', 'Citizen', 'Industry'], axis=1,
inplace=True)
#x variable(features),y variable(dependent
variable)
x=df.drop('Approved',axis=1)
y=df['Approved']
import statsmodels.api as sm
logit_model=sm.Logit(y,x)
```

```
result=logit_model.fit()
print(result.summary2())
```

After removing the insignificant predictors, we run again our predictive model with significant predictors. And generate a new Logistic regression model.

```
df.drop(['Debt', 'Married', 'BankCustomer',
'Employed'], axis=1, inplace=True)
x=df.drop('Approved',axis=1)
y=df['Approved']
logit_model=sm.Logit(y,x)
result=logit_model.fit()
print(result.summary2())
```

Generated logistic regression model's coefficients are presented as $\log odds$ $\log odds$ probability for proper interpretation should be done after exponentiating values of coefficients (Jankovic, 2021). Below is a python code that generates odds for logistic regression coefficients.

```
odds = np.exp(logreg.coef_[0])
pd.DataFrame(odds,
        x.columns,
        columns=['coef'])\
       .sort_values(by='coef', ascending=False)
```

So, if independent variable X increases by one unit, the odds that the credit card application will be approved [coefficient value] are as large as the odds that it won't be approved (Benton, 2020).

After the removal of attributes dataset is split into training (30%) and testing (70%) datasets and the model is tested on its prediction accuracy.

```
from sklearn.model_selection import train_
test_split
x_train,x_test,y_train,y_test = train_test_
split(x,y,test_size=0.30,random_state=0)

from sklearn.linear_model import
LogisticRegression
text_classifierLR=LogisticRegression(random_
state=0)
text_classifierLR.fit(x_train,y_train)
predLR = text_classifierLR.predict(x_test)

from sklearn.metrics import accuracy_score
print(accuracy_score(y_test, predLR))
```

5. RESULTS

After data pre-processing missing values were replaced with suitable values and all instances are kept in the dataset for further analysis.

5.1 Data analysis

After making categories from client age attributes in the data analysis section Figure 1 demonstrates the distribution of credit card approvals according to four age groups. Data shows that middle-aged adults and senior adults groups have more approved, while adults and teenage groups have more not approved applications on average.

To obtain some insides from data regarding the correlation between credit score and approval results credit score attribute was categorized into four groups (bad, fair, good, and excellent) and Figure 2 visualizes approval results based on credit score results categories. Visual shows that approved applications more appear in the *excellent* and *good* credit score category while the *bad* credit score category has more not approved on average.

Another piece of information that was tested about the relationship with the client application approval is the client's debt balance. Figure 3 shows the distribution of approval results grouped by the client's debit balance groups. It shows that the *high* debt group has more approved while the *low* debt group has more not approved applications on average.

To confirm these hypotheses a Chi-square test of independence was performed to see whether credit card approval was dependent on the client's age, credit score, and debt amount. Results are presented in the Table 3.

Table 3. Variable dependency using Chi-square test.

Variable 1	Variable 2	Dependency
Credit Card Approved	Age	Yes
Credit Card Approved	Gender	No
Credit Card Approved	Credit Score	Yes
Credit Card Approved	Debt	Yes

Note: Table 3 shows chi-square dependency test results for different variables.

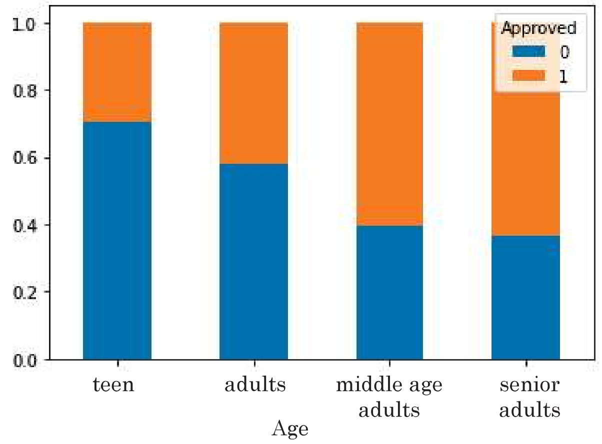

Figure 1. Distribution of applications by age groups.

Note: Figure 1 shows credit card applicants' approval status distribution by age group.

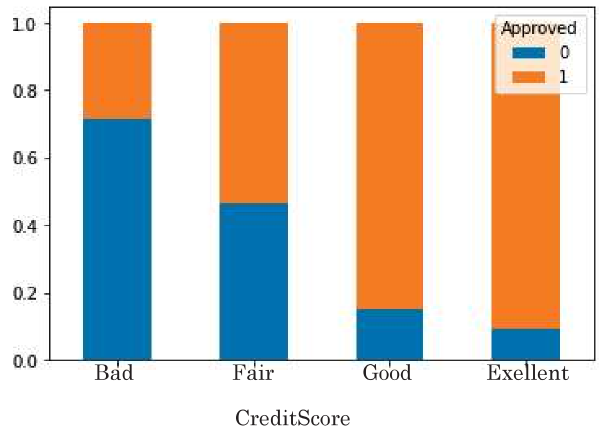

Figure 2. Distribution of applications by credit score groups.

Note: Figure 2 shows credit card applicants' approval status distribution by credit score group.

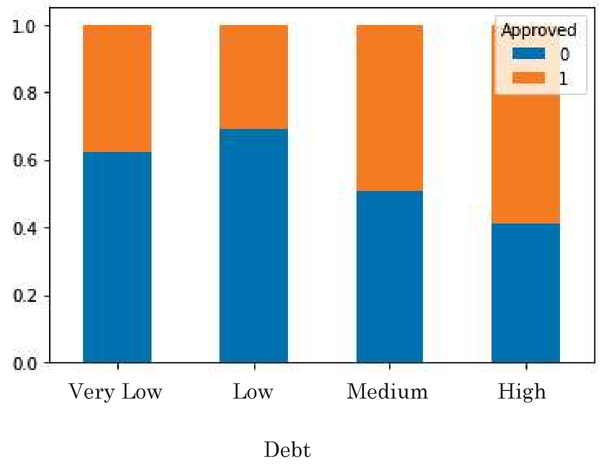

Figure 3. Distribution of applications by debt balance groups.

Note: Figure 3 shows credit card applicants' approval status distribution by debt balance group.

5.2 Logistic Regression Model

Figure 4 shows the Implementation of the Logistic regression model with 9 predictors. Predictors with a p-value less than 0,05 are insignificant and removed from the model. Attributes like debt, married, bank customer, and employed have a p-value greater than 0,05.

After the removal of insignificant predictors new model with 5 predictors is presented in Figure 5. A final predictive model is created using age, years of employment, prior defaults, and income attributes. Finally, the logistic regression coefficients are analyzed for understanding the significance of the independent variables on the target variable.

We have three attributes (prior default, credit score, and income) with positive coefficients and the age attribute with a negative coefficient. Results shown in Table 4 indicate that prior default is the most important attribute for the Logistic regression model, an applicant without prior default has 18,39 times the odds of the applicant with prior default to have a credit card approved.

Table 4. Logistic regression model's coefficients log odds and exponentiating values

Variable	Coefficient	exp Coefficient
Age	- 0,0802	0,922932
Years Employed	0,1363	1,146026
Prior Default	2,9120	18,393549
Credit Score	0,1908	1,210217
Income	0,0004	1,000400

Note: Table 4 shows regression model coefficients for log odds and exponentiating values.

The final stage of the research was creating a classification model with the data that have been prepared and trying to predict whether a client would get a credit card application approved or not. The logistic regression model predicted approval outcomes with 86% accuracy.

5.3 Proposed Dashboards

Many companies use Microsoft Excel for daily reports preparation for measurement of the performance of their sales departments (Beltran *et al.*, 2021) which is time-consuming and not always the best solution for visual presentations. The authors propose a dashboard to concentrate on the performance of the credit

```
                            Results: Logit
=========================================================================
Model:              Logit               Pseudo R-squared:  0.449
Dependent Variable: Approved            AIC:               542.6008
Date:               2022-08-29 22:31    BIC:               587.9677
No. Observations:   690                 Log-Likelihood:    -261.30
Df Model:           9                   LL-Null:           -474.08
Df Residuals:       680                 LLR p-value:       4.8009e-86
Converged:          0.0000              Scale:             1.0000
No. Iterations:     35.0000
-------------------------------------------------------------------------
               Coef.    Std.Err.      z    P>|z|      [0.025      0.975]
-------------------------------------------------------------------------
Gender        -0.7833      0.2249 -3.4832 0.0005      -1.2241     -0.3426
Age           -0.0582      0.0086 -6.7417 0.0000      -0.0751     -0.0412
Debt          -0.0343      0.0234 -1.4644 0.1431      -0.0802      0.0116
Married      -22.1653 65747.8886 -0.0003 0.9997 -128885.6590 128841.3283
BankCustomer  21.9233 65747.8886  0.0003 0.9997 -128841.5704 128885.4169
YearsEmployed  0.1328      0.0417  3.1864 0.0014       0.0511      0.2145
PriorDefault   3.0662      0.2585 11.8637 0.0000       2.5596      3.5728
Employed       0.1647      0.2974  0.5537 0.5798      -0.4182      0.7476
CreditScore    0.1642      0.0500  3.2861 0.0010       0.0663      0.2621
Income         0.0004      0.0001  3.2768 0.0010       0.0002      0.0007
=========================================================================
```

Figure 4. Logistic Regression first model with all predictors.

Note: Figure 4 shows Logistic regression model approved status as dependent variable and 10 independent variables.

```
                         Results: Logit
=================================================================
Model:              Logit            Pseudo R-squared: 0.429
Dependent Variable: Approved         AIC:              551.7480
Date:               2022-08-30 08:20 BIC:              574.4315
No. Observations:   690              Log-Likelihood:   -270.87
Df Model:           4                LL-Null:          -474.08
Df Residuals:       685              LLR p-value:      1.1479e-86
Converged:          1.0000           Scale:            1.0000
No. Iterations:     8.0000
-----------------------------------------------------------------
                Coef.  Std.Err.    z      P>|z|   [0.025  0.975]
-----------------------------------------------------------------
Age            -0.0802   0.0064 -12.5147 0.0000 -0.0928 -0.0677
YearsEmployed   0.1363   0.0408   3.3384 0.0008  0.0563  0.2163
PriorDefault    2.9120   0.2490  11.6952 0.0000  2.4240  3.4000
CreditScore     0.1908   0.0398   4.7922 0.0000  0.1128  0.2689
Income          0.0004   0.0001   3.1705 0.0015  0.0001  0.0006
=================================================================
```

Figure 5. Logistic Regression first model with significant predictors.

Note: Figure 5 shows Logistic regression model approved status as dependent variable and 5 independent variables significant variables.

cards sales department with a focus on monitoring the following visualizations shown on Figure 6:

– **Count of approved/rejected credit card applications** shows the number of approved and rejected applications for monitoring of sales department performance,
– **Average credit score by approval outcome** monitors average credit score attribute, since credit score is an important attribute,
– **Average debt by approval outcome** monitors average debt attribute, since debt is an important attribute,
– **The average income balance for approved applicants** monitors average income attributes, since income is an important attribute,
– **The average of prior default by approval** monitors the average prior default attribute since prior default is an important attribute,

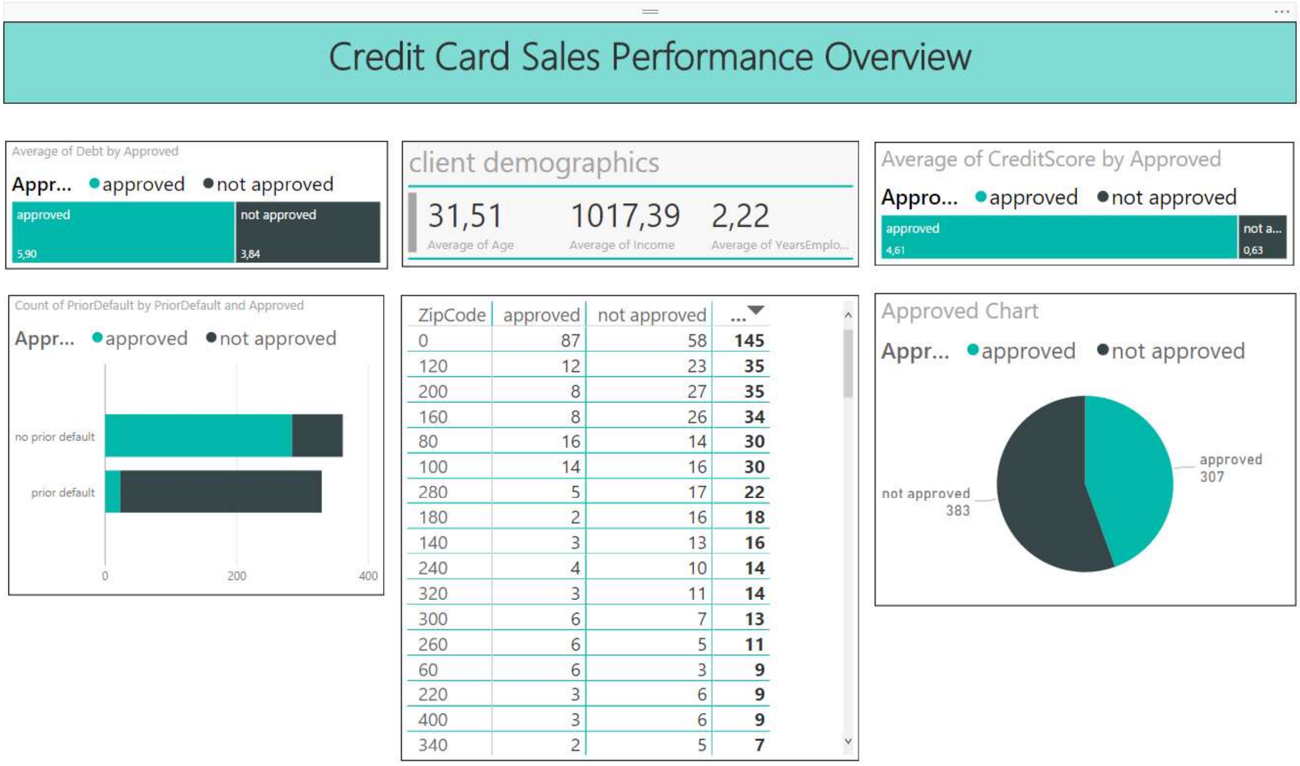

Figure 6. Proposed Credit Card Sales Performance Dashboard in Power BI.

Note: Figure 6 shows Power BI dashboard with suggested key sales performance indicators.

- **The average of prior default by approval** monitors average prior default attributes since prior default is an important attribute,
- **Median of the client age by approval**: monitors Median of the client age attribute, for focusing on targeting population for marketing campaigns,
- **Count of approved/rejected credit card applications by client address zip code** shows geographic areas where marketing strategies should take place.

6. CONCLUSION

Data analysis conducted in this research confirmed that it is possible to predict the outcome of the credit card approval process using information like prior defaults, credit score, years of employment, and income balance. The Chi-square independence test and logistic regression model showed that credit card approval is dependent on the client's prior defaults, credit score, and years employed while it is not dependent on gender. Furthermore, the logistic regression model showed that: (1) The change in the prior default variable increases the odds that the credit card application will be approved 18,39 times; (2) An increase in the credit card score variable by 1 increases the probability that the application will be approved by 21%; (3) An increase in the years employed variable by 1 increases the probability that the application will be approved by 14%. As for the log odds with a value less than 1, it is shown that when the customer age increases by 1, the probability that the credit card application won't be approved increases by 1/0.922932.

Corresponding to this research paper's question testing, the results show that credit card approval is expected to depend on the age of the client, their credit score, and debt, but not on the client's gender. The prediction model that was created using the logistic regression algorithm was tested with an accuracy rate of 86%. For a simpler business model number of features was decreased to 5 (age, years employed, prior default, credit score, and income).

It can be concluded that, as the credit card industry is still growing, banks need business intelligence solutions to increase credit card approval process time and prediction accuracy. For a better sales performance dashboard for monitoring important attributes, Key Performance Indicators (KPIs) are proposed. Certainly, dashboard visuals could help sales departments increase marketing campaigns.

REFERENCES

Alzeaideen, K. (2019) 'Credit risk management and business intelligence approach of the banking sector in Jordan', *Cogent Business & Management*. Cogent, 6(1). doi: 10.1080/23311975.2019.1675455.

Aruldoss, M. *et al.* (2014) 'A survey on recent research in business intelligence', *Journal of Enterprise Information Management*. doi: 10.1108/JEIM-06-2013-0029.

Balachandran, B. M. and Prasad, S. (2017) 'Challenges and Benefits of Deploying Big Data Analytics in the Cloud for Business Intelligence', *Procedia Computer Science*. Elsevier B. V., 112, pp. 1112–1122. doi: 10.1016/J.PROCS.2017.08.138.

Beltran, D. J. *et al.* (2021) 'Credit card sales performance dashboard', in *Proceedings of the International Conference on Industrial Engineering and Operations Management*, pp. 1–12.

Benton, J. (2020) *Interpreting Coefficients in Linear and Logistic Regression, Towards Data Science*. Available at: https://towardsdatascience.com/interpreting-coefficients-in-linear-and-logistic-regression-6ddf1295f6f1 (Accessed: 29 August 2022).

Bhatore, S., Mohan, L. and Reddy, Y. R. (2020) 'Machine learning techniques for credit risk evaluation: a systematic literature review', *Journal of Banking and Financial Technology 2020 4:1*. Springer, 4(1), pp. 111–138. doi: 10.1007/S42786-020-00020-3.

Bogetoft, P. (2012) 'Performance Benchmarking', in *Performance Benchmarking*. Boston, MA: Springer US (Management for Professionals). doi: 10.1007/978-1-4614-6043-5.

Bramer, M. (2016) *Introduction to Data Mining*. doi: 10.1007/978-1-4471-7307-6_1.

Brian, O. (2020) *History of the Credit Card: Origins, Laws and Timeline - TheStreet, History of the Credit Card: Origins, Laws and Timeline*. Available at: https://www.thestreet.com/personal-finance/credit-cards/history-of-credit-cards (Accessed: 21 July 2022).

Bridges, B. (2022) *What Is APR On A Credit Card?, Credit Cards*. Available at: https://www.bankrate.com/finance/credit-cards/what-is-credit-card-apr/(Accessed: 19August 2022).

Brown, K. and Moles, P. (2014) 'Credit risk management', in *Credit risk management*, pp. 105–138.

Constance, S. (2022) *Average Credit Card Debt In The U.S. | Bankrate, Credit Cards*. Available at: https://www.bankrate.com/finance/credit-cards/states-with-most-credit-card-debt/ (Accessed: 21 July 2022).

Cortinhas, S. (2022) *Credit Card Approvals | Kaggle*. Available at: https://www.kaggle.com/datasets/samuelcortinhas/credit-card-approval-clean-data (Accessed: 11 August 2022).

Ćurko, K., Bach, M. P. and Radonić, G. (2007) 'Business intelligence and business process management in banking operations', in *Proceedings of the International Conference on Information Technology Interfaces, ITI*, pp. 57–62. doi: 10.1109/ITI.2007.4283744.

Dahiru, T. (2013) 'P-Value', *Encyclopedia of Radiation Oncology*, 6(1), pp. 692–692. doi: 10.1007/978-3-540-85516-3_649.

Dieker, N. (2022) *How Your Credit Card Limit Is Determined, Bankrate*. Available at: https://www.bankrate.com/finance/credit-cards/how-issuers-determine-credit-card-limits/ (Accessed: 19 August 2022).

Ghobadi, F. and Rohani, M. (2017) 'Cost sensitive modeling of credit card fraud using neural network strategy', *Proceedings - 2016 2nd International Conference of Signal Processing and Intelligent Systems, ICSPIS 2016*. Institute of Electrical and Electronics Engineers Inc. doi: 10.1109/ICSPIS.2016.7869880.

Gobat, J. (2012) *Banks: At the Heart of the Matter*. Available at: https://www.imf.org/external/pubs/ft/fandd/basics/bank.htm (Accessed: 11 August 2022).

Grant, W. and Wilson, G. K. (2012) 'Consequences of the Global Financial Crisis', *The Consequences of the Global Financial Crisis: The Rhetoric of Reform and Regulation*. Oxford University Press, p. 287. doi: 10.1093/ACPROF:OSO/9780199641987.001.0001.

Greg, M. (2021) *Personal Loans vs. Credit Cards: What's the Difference?, Personal Loans vs. Credit Cards: What's the Difference?* Available at: https://www.investopedia.com/articles/personal-finance/041415/pros-cons-personal-loans-vs-credit-cards.asp (Accessed: 23 July 2022).

Hayes, A. (2022) *Chi-Square (χ2) Statistic*. Available at: https://www.investopedia.com/terms/c/chi-square-statistic.asp (Accessed: 27 August 2022).

Infographic (2021) *Credit Card Statistics: Global Facts, Data, and Figures*. Available at: https://rcbcbankard.com/blogs/credit-card-statistics-global-facts-data-and-figures-16 (Accessed: 29 August 2022).

Jankovic, D. (2021) *A Simple Interpretation of Logistic Regression Coefficients, Towards*

Data Science. Available at: https://towardsdatascience.com/a-simple-interpretation-of-logistic-regression-coefficients-e3a40a62e8cf (Accessed: 29 August 2022).

Kibria, M. G. and Sevkli, M. (2021) 'Application of Deep Learning for Credit Card Approval: A Comparison with Two Machine Learning Techniques', *International Journal of Machine Learning and Computing*, 11(4), pp. 286–290. doi: 10.18178/ijmlc.2021.11.4.1049.

King, M. (2022) *6 Key Differences Between Data Analysis and Reporting*, *6 Key Differences Between Data Analysis and Reporting*. Available at: https://databox.com/data-analysis-reporting (Accessed: 28 July 2022).

Luthi, B. and Karp, G. (2021) *How to Apply for a Credit Card So You'll Get Approved, NerdWallet*. Available at: https://www.nerdwallet.com/article/credit-cards/apply-for-a-credit-card (Accessed: 19 August 2022).

Maalouf, M. (2011) 'Logistic regression in data analysis: An overview', *International Journal of Data Analysis Techniques and Strategies*, 3(3), pp. 281–299. doi: 10.1504/IJDATS.2011.041335.

Mchugh, M. L. (2013) 'The Chi-square test of independence Lessons in biostatistics', *Biochemia Medica*, 23(2), pp. 143–9. Available at: http://dx.doi.org/10.11613/BM.2013.018.

McKinney, W. (2009) 'pandas: a Foundational Python Library for Data Analysis and Statistics', *Python for high performance and scientific computing*, 14(9), pp. 1–9. doi: 10.1002/mmce.20381.

Mehanović, D. and Durmić, N. (2022) 'Case Study Application of Business Intelligence in Digital Advertising', *International Journal of E-Business Research*. IGI Global, 18(1), pp. 1–16. doi: 10.4018/IJEBR.293294.

Nithya, N. and Kiruthika, R. (2021) 'Impact of Business Intelligence Adoption on performance of banks: a conceptual framework', *Journal of Ambient Intelligence and Humanized Computing*. Springer Berlin Heidelberg, 12(2), pp. 3139–3150. doi: 10.1007/s12652-020-02473-2.

Pandas (2022) *pandas.crosstab — pandas 1.4.3 documentation, API reference*. Available at: https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.crosstab.html (Accessed: 27 August 2022).

Pérez-Martín, A., Pérez-Torregrosa, A. and Vaca, M. (2018) 'Big Data techniques to measure credit banking risk in home equity loans', *Journal of Business Research*. Elsevier, 89, pp. 448–454. doi: 10.1016/J.JBUSRES.2018.02.008.

Peussa, A. (2016) 'Credit Risk Scorecard Estimation By Logistic Regression', *Faculty of Science Tekijä, Författare.*, (May), p. 33.

Rafi, K. A. and S.M.K, Q. (2012) 'Business Intelligence : An Integrated Approach', *Business Intelligence Journal*, 5(1), pp. 64–70.

Shift (2021) *Credit Card Statistics - Shift Processing, Credit Card Statistics*. Available at: https://shiftprocessing.com/credit-card/ (Accessed: 21 July 2022).

Siddiqi, N. (2012) *Credit Risk Scorecards, Credit Risk Scorecards*. doi: 10.1002/9781119201731.

Sphinx (2008) *scipy.stats.chi2_contingency — SciPy v1.9.1 Manual, API reference*. Available at: https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.chi2_contingency.html (Accessed: 27 August 2022).

statsmodels (no date). Available at: https://pypi.org/project/statsmodels/ (Accessed: 15 August 2022).

Tabul8tor (2019) *How do Credit Cards work?. Anyonewhoownsacreditordebitcard,Howdo Credit Cards work?* Available at: https://medium.com/@tabul8tor/how-do-credit-cards-work-a15596e14860 (Accessed: 21 July 2022).

Talend (no date) *What is a Data Mart? (vs a Data Warehouse), Talend*. Available at: https://www.talend.com/resources/what-is-data-mart/ (Accessed: 28 August 2022).

Ubiparipovi, B. and Durkovic, E. (2011) 'Application of Business Intelligence in the Banking Industry', *Management Information Systems,* 6, pp. 23–30. Available at: http://www.ef.uns.ac.rs/mis/archive-pdf/2011 - No4/MIS 2011_4_4.pdf.

Uppal, J. Y. and Ullah Mangla, I. (2013) 'Extreme loss risk in financial turbulence – evidence from the global financial crisis', *Managerial Finance*. Emerald Group Publishing Ltd., 39(7), pp. 653–666. doi: 10.1108/03074351311323446/FULL/XML.

Watson, H. J. and Wixom, B. H. (2007) 'The current state of business intelligence', *Computer*, 40(9), pp. 96–99. doi: 10.1109/MC.2007.331.

Widow, J. (1992) 'Research problems in data warehousing', in *Proceedings of the fourth international conference on Information and knowledge management*, pp. 25–30.