T.C.
TURKISH - GERMAN UNIVERSITY
SOCIAL SCIENCES INSTITUTE
INTERNATIONAL FINANCE DEPARTMENT

A CREDIT CLASSIFICATION APPLICATION WITH MACHINE LEARNING METHODS: GERMAN CREDIT DATASET EXAMPLE

MASTER'S THESIS

Egemen KAYABAŞI

ADVISOR
Doç. Dr. Semih Emre ÇEKİN

İSTANBUL, June 2022

T.C.

TURKISH - GERMAN UNIVERSITY

SOCIAL SCIENCES INSTITUTE

INTERNATIONAL FINANCE DEPARTMENT

A CREDIT CLASSIFICATION APPLICATION WITH MACHINE LEARNING METHODS: GERMAN CREDIT DATASET EXAMPLE

MASTER'S THESIS

Egemen KAYABAŞI

(198106006)

Thesis Submission Date to Institute: 24.06.2022

Thesis Defence Date : 08.09.2022

Thesis Advisor : Doç. Dr. Semih Emre ÇEKİN

Other Jury Members: Dr. Levent YILMAZ

Dr. Alpaslan AKAY

T.C.
TURKISH - GERMAN UNIVERSITY
SOCIAL SCIENCES INSTITUTE
INTERNATIONAL FINANCE DEPARTMENT

A CREDIT CLASSIFICATION APPLICATION WITH MACHINE LEARNING METHODS: GERMAN CREDIT DATASET EXAMPLE

MASTER'S THESIS

Egemen KAYABAŞI
(198106006)

ADVISOR
Doç. Dr. Semih Emre ÇEKİN

ISTANBUL, June 2022

ÖZET

MAKINE ÖĞRENMESI METOTLARI ILE KREDI SINIFLANDIRMA UYGULAMASI: ALMAN KREDI VERI SETI ÖRNEĞI

Yapay zekâ uygulamaları finans dâhil birçok iş ve akademik araştırma alanında kullanım alanı bulmuştur. Makine öğrenmesi algoritmaları, sunulan verileri insan zihninin öğrenme sürecine benzer bir şekilde analiz ederek, yüksek başarımla ve sürekli insan denetiminde olmaksızın bilgi kategorize etmek ve tahminler yapmak için kullanılabilecek karar verme mekanizmaları oluşturur. Makine öğrenmesi, pratik amaçlar için muazzam bir potansiyel vaat etse de, finansal uygulamaların son kullanıcılarına faydası sınırlı kalmıştır. Bu çalışma, tipik bir makine öğrenimi uygulamasının temel bileşenlerini tanımlamak, performansını etkileyen faktörleri analiz etmek ve kullanımını sınırlayan olası eksiklikleri gözlemlemek için finans alanındaki makine öğrenmesi uygulamalarının mevcut kullanımını bir kredi sınıflandırması örneğinde izlemeyi amaçlamaktadır. Kredi başvuru verileri üzerinde yapılan deneyler sonucunda, bir makine öğrenimi modelinin performansının, kullanılacak veri seti için özel olarak yapılması gereken model tasarımına fazlasıyla bağlı olduğu ve kullanıcı dostu bir arayüz olmadan güncel makine öğrenmesi araçlarının finans uzmanlarından çok bilgisayar profesyonellerine hitap ettiği gözlemlenmiştir.

Anahtar Kelimeler: Makine öğrenmesi, yapay zekâ, tüketici kredisi sınıflandırması

Tarih : 24.06.2022

ABSTRACT

A CREDIT CLASSIFICATION APPLICATION WITH MACHINE LEARNING METHODS: GERMAN CREDIT DATASET EXAMPLE

Artificial intelligence applications have found use in many business and academic research areas including finance. By analyzing existing data in a way similar to the learning process of a human mind, machine learning algorithms can create decision making mechanisms which can be used for categorizing and predicting new data with precision and without constant supervision. Although machine learning promises enormous potential for practical purposes, its presence to the end user of financial applications has been limited. This study aims to observe current uses of machine learning applications in the area of finance, specifically in a credit classification example, in order to analyze the basic components of a typical machine learning application, the factors affecting its performance and observe possible shortcomings that limit its use as a common tool. Experiments conducted on credit application data suggest that performance of a machine learning model is very much dependent on its design which has to be specifically made to match the dataset to be utilized and without a user friendly interface, machine learning tools address to a computing crowd instead of finance specialists at the moment.

Keywords: Machine learning, artificial intelligence, consumer credit classification

Date: 24.06.2022

TABLE OF CONTENTS

PAGE NO

LIST OF ABBREVIATIONS

AI	:	Artificial Intelligence
CPU	:	Central Processing Unit
DL	:	Deep Learning
GPU	:	Graphic Processing Unit
IT	:	Information Technology
ML	:	Machine Learning
SVM	:	Support Vector Machine
TPU	:	Tensor Processing Unit
XGB	:	Extreme Gradient Boosting

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

Digital transformation is inevitable. Companies are facing harsh competition in every business area and in order to survive and succeed maximizing efficiency is a must. When human skills could not cope with growing amount of calculations and storing and recalling data, business machines came to help. Use of computers made vast amounts of data available to decision makers and information systems made it possible to turn raw data for assisting managerial decisions. Now with the advances in technology a new era of automation and continuous improvement is upon us. Artificial intelligence (AI) has long took its place out of science fiction and became a reality in science, business and our daily lives. Therefore AI applications in finance are becoming more relevant every year.

Aim of this work is to observe and explain the mechanism of a typical machine learning (ML) model in the field of consumer finance. AI applications are already in use in financial corporations for some time yet there is no go to software packages or default methods to solving problems regarding finance sector. Due to private nature of the data being used and methods being specific to the available data, custom AI applications had to be made for company specific operations. For a researcher or an employee in the area of finance, utilization of AI tools can be considered a black box considering their software oriented nature. This work attempts to explain stages of building a ML model for a finance sector application without using complicated technical terms that could provide insights for people without computer science experience.

This thesis aims to investigate the basic structure of a ML model, current uses in finance sector, the tools and requirements for creating typical ML model and observe some of the factors that affect performance of ML algorithms such as sample size and feature selection. In order to conduct tests on a real world example, open source ML tools using Python programming language and a real life credit application dataset along with three randomly generated artificial datasets are used. By generating a ML model that evaluates credit applications the components and challenges to creating a practical model are observed. Conducting same tests on artificial datasets representing three scenarios provide insights on functioning mechanisms of ML methods on categorizing credit evaluation performances. In order to guide newcomers of the area that are interested in using ML functions, a step by step guide to a sample ML model building is added at the end as an appendix.

2. BANKING AND CREDITS

2.1 Evaluation of Credit Applications and Credit Score

2.1.1 Definition and Uses

Credit in its broader sense includes any form of deferred payment (O'Sullivan and Sheffrin, 2003). Credit would take in many forms and is categorized in many types according to their users, aim, payment periods and others. Scope of the experiment conducted in this work is limited to risk analysis of consumer credits provided by financial institutions to individual customers.

Consumer credits are a tool for individuals to buy the things they want without paying cash at the time of purchase (USA Federal Trade Commission). Mortgages, credit cards, car loans and education loans are examples of credits that are offered to consumers. Apart from specialized products that are offered for a specific reason, loans can be offered for any means of use. Sometimes referred to as personal loans, these type of credits are less in amount and payment periods compared to mortgages and their applications take less time to process. The dataset used for the experiments in this work was taken from personal loan applications with a range of proposed reasons such as home renovations and holidays.

Along with the type of credit product, the loaner makes up a very important part of the loan agreement. In order to maximize their profits, institutions have to collect information and asses the risk connected to the loaner. This is the part where credit risk determination tools are utilized.

2.1.2 Consumer Credits and Risk

Interest payments made from customers to the company make up a significant portion of the income of a bank. Financial Institutions must use a review process in order to maximize their profits and minimize the risk of loss due to its debtors' inability to pay back the loan. Even with the advanced technology evaluating consumers creditworthiness individually would require tremendous effort and time, therefore automated decision making tools are put into action such as credit scores.

2.2 Credit Approval Process

When processing a loan application, a financial institution will use models tailored for its customer base, product variety and risk preferences. Most of the time, when a consumer applies for a loan, the application will go through a pre-processing phase, where the basic information about the customer and the loan, such as consumer income, payback period and existing loans, will be put through a mathematical equation that will approve or reject the application based on equations. Pre-approving process does not involve an employee or specialist decision making, thus it should eliminate unwanted loan applications easily within a very short period of time. A credit score is a powerful tool for selecting pre-approved loans, since it readily defines the loan applicant for a mathematical model, using his income, loan and payment information.

Depending on the type of the loan, a pre-approval process could directly lead to acceptance or refusal of an application. For typically high amount and longer payment period loans such as mortgages, additional processing is required to asses an application. This is where a loan underwriting process comes into action, where a professional loan agent assesses the truthfulness of the information provided by the customer and other factors that could affect the payback of a loan to comment on the risk involved. Such a review process requires trained personnel and intuition also could take days. AI applications like machine learning models can be used for underwriting processes for faster and more precise reviews (Swati et al., 2019).

2.3 Credit Scores

A credit score is a tool that a financial institution uses to assess the probability of its potential customers' paying a loan back on time. It is prepared using the information based on a customer's credit report (Consumer Financial Protection Bureau, 2020). Although there are several decision support companies offer many credit scoring products, the basic principles of credit scoring are similar among them.

2.3.1 FICO

First introduced by Fair Isaac Corporation in 1989, a FICO score is a number between 300 and 850 which calculated by using credit information of an individual. Five categories of information are used in them payment history and current debt level being of highest importance with having 35 and 30 percent impact on score respectively. A higher score can lead to approval of higher amounts of loans and lower interest rates for an individual while scores lower than average may seriously hurt chances of taking out a loan. Fair Isaac Corporation claims that approximately 90 percent of all credit decisions in USA use FICO scores (Fair Isaac Corporation, 2018).

2.3.2 SCHUFA

Schufa Holding AG is a private German financial corporation that specializes in providing information on creditworthiness of companies and individuals. The company produces credit scores that are expressed in percentages where a 100 percent score means the payback on time probability is the highest. Schufa claims to have processed information from over 68 million individuals and 6 million companies and produces 510 thousand credit reports daily (Schufa Holding AG, 2022). The SCHUFA score is a benchmark for individuals, and used frequently for renting agreements and other loan applications. The company does not disclosure its scoring methods.

2.3.3 Findeks

Findeks is a credit score product offered by Kredi Kayıt Bürosu A.Ş, a private company founded in 1995 by 9 banks in Turkey for assessing credit risk. Today the company's services are used by 46 banks, 14 consumer financing companies, and 53 factoring companies among others. Findeks is a credit score for individuals expressed in numbers between 1 and 1900. Its four calculation factors are disclosed the company website, debt payment habits and current account and debt information being the two most important factors. (Kredi Kayıt Bürosu, 2020).

2.4 Big Data

While credit scores are largely based on previous loans, payments and financial strength, additional consumer information can influence an individual's risk levels in the eyes of debtors. Companies can monitor and use information such as a customer's shopping habits and social media activity to develop a model that can assess creditworthiness of an individual when traditional credit information is not available or limited (European Data Protection Supervisor, 2021). While in theory, with the use of excess amount of information and computing capabilities, such models could carry significant potential, transparency of individual information used and fairness in evaluations arises new challenges for companies (Culnan and Armstrong, 1998).

3. ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

3.1 Definitions and Basics

3.1.1 Artificial Intelligence and Machine Learning

Humankind uses tools for all kinds of tasks to make life easier. Computers brought tremendous speed to calculation tasks and when calculations could be used for tasks where human cognitive functions such as learning and problem solving the term artificial intelligence came into existence.

Although the term AI immediately brings to the idea of machines mimicking and replacing people thanks to science fiction and movies, actual AI tasks could be as simple as recognizing a letter or a number written on a piece of paper. Text recognition though a miniscule task for many people, is a very difficult job for a programmer to teach to a computer through a software. Traditional programming model for such a task could mean digitalizing and loading each and every possible image of a letter to a database and coding an algorithm that compares and matches an image to that database in order to find a letter. In such an arrangement most of the time programmers would have to teach a computer every single image possible every time for a recognition task. Yet the human mind operates in a much more efficient way for seeing and defining. While looking at a piece of text, even when the print is unclear or letters are written in a form different than usual, a person would most likely define a letter and number instantly.

A person uses his or her past experiences, judgment and memory to define a seen object. Similar to the cognitive functions of a human, if we could load pictures of letters and numbers to a computer and use a code that does not define exactly how a number should be written but instead tell which image corresponds to which number, we would be sort of teaching a computer. And when a computer using the given images, starts predicting the letters and numbers of a given image using that code, it does an intelligent task that is artificially coded. This sort of image recognition functions are now in common place in our everyday lives and consists some of the most common uses of AI functions.

Deep learning is a subset of machine learning which itself is a subset of artificial intelligence. Deep learning utilizes artificial neural networks, which mimic biological neural networks found in human and animal brains.

3.1.2 Supervised and Unsupervised Learning

Machine Learning is categorized according to learning types. In supervised learning data has labels. Without a specified category or label, every data is non-labeled. For example a long list of numbers do not represent any information by themselves but when they are listed under a table as prices of an item, they became labeled and become main inputs of supervised learning. The main goal in supervised learning is creating a framework using existing data that will be used for predictions in the future. In semi-supervised learning, labeled and non-labeled data are used at the same time. Using a scarce amount of labeled data can vastly improve a machine learning model that has to deal with readily available non-labeled data.

Data without labels are usually not used for predictions. Non-Supervised learning uses data without labels, instead of predictions it is focused on clustering and grouping existing data. A computer cannot differentiate a dog from a cat when it is presented with non-labeled animal images. Yet it can still divide those images into groups and decide if a new image presented to it belongs to a cat-like animals group when it is trained with a non-supervised ML algorithm.

In reinforcement learning, an intelligent agent is placed in an interactive environment, where the goal is to optimize its actions that lead to cumulative gains. Each iteration of the algorithm uses reinforcement or a feedback to plan the next iteration either from the environment or from an administrator.

3.2 Why Machine Learning has Recently Became More Relevant than Ever

Compared to some of the IT technologies used today, Machine Learning techniques and algorithms are not new. In fact invention and first use of ML algorithms go back as early as 1960s (Fradkov, 2020). Yet it took decades to fully utilize the practical uses of it in real life. Since 1960s the amount of big data and computing speed has improved significantly. Although algorithms and learning techniques have also been improved, now with the vast amount of available data to the analysts, with fast interconnected devices and relatively cheap and efficient computing power, ML can be used for numerous functions in a variety of environments, which was not possible before.

3.3 ML Hardware

As long as it is able to run the open source ML software libraries, any type of personal computer can run ML algorithms. As some common machine learning algorithms tend to perform numerous small tasks repetitively, a graphic processing unit (GPU) which is normally designed to handle graphic generation for a computer and consists of more processing unit than a central processing unit (CPU), can perform better in some ML tasks. While advantage of GPUs in ML is often credited to their higher computational power and parallel processing abilities, a lot of this performance difference can be attributed to their higher memory bandwidth. Although they were not designed with ML tasks in mind, their ability to access more data than CPUs in a given time frame make GPUs perform better in ML tasks (Ongsulee, 2017).

With the expansion of application areas of ML technology, the need for specifically designed AI processing units has emerged. By introducing its AI open source software platform Tensorflow, introducing its application specific computing chip Tensor processing unit (TPU) and offering cloud access to its ML specialized hardware with its Vertex AI platform, Google has been investing heavily on the technology. Tensor chips has the advantage of being designed with mathematical and matrix operations used for AI applications, therefore they can more efficiently handle such tasks (Google Cloud, 2022).

Google has also added a TPU unit to its mobile phone Pixel 6. The mobile ML chip handles many tasks in photography features such as image sharpening and removing objects in the background. The chip will also provide faster voice and text recognition by handling these tasks along within the device in addition to cloud computing (Molloy, 2021). With the availability of more capable mobile devices and fast internet connections mobile devices are capable of handling machine learning tasks that can enrich the services provided to end customers. Some examples of such services include Microsoft’s Seeing AI application on IOs for visually impaired (Microsoft, 2022) or medical services used through mobile phones in order to analyze health metrics that assist in diagnose and patient monitoring (Khan and Alotalibi, 2020).

3.4 Current Trends and Applications

Finance sector is one of the many that enjoyed several practical uses of ML. Some examples of ML services include fraud detection, where algorithms are able to detect subtle hints of fraudulent transactions within a dataset of millions, regulatory compliance where software keeps companies up to new regulations and policies efficiently with minimized risks, customer service where bank customers are offered the most fitting services and products without excessive personnel costs, and investment advice where the optimum portfolio for each customer can be created with the highest performance and no personal bias (Emerson et al., 2019).

Apart from popular uses such as autonomous driving, x-ray analyzing, image processing, voice recognition and social media detections ML techniques has been used for several different research and practical use areas such as discovering influences in paintings among artists (Saleh et al., 2014), predicting green behavior of tourists (Rezapouraghdam et al., 2021) and COVID-19 diagnosis (Alballa and Al-Turaiki, 2021).

4. LITERATURE REVIEW

With the increased computing power and availability of data, capabilities of machine learning became more apparent and AI applications started to appear more within many research topics including finance. Among with the total number of articles produced, the number of patent applications can illustrate this point. He, Du and Qiao (2020) pointed out this situation in their paper, in which they show that global number patent applications in the field of intelligent finance, which includes AI and ML applications, increased from 10 in 1995 to 2505 in 2019 and almost 70 percent of all applications within this time frame were recorded in the last five years. The article also draws attention to the fact that China and USA were the top two countries in intelligence finance patent applications.

Master’s theses and PhD dissertations exist where similar subject matters in machine learning and AI techniques in finance and especially credit risk assessment are investigated. In one such works, Alperen (2019) added a risk factor based on location of loan applicants to a dataset based on Turkey and compared performances of machine learning models in categorizing risky and non-risky applications. Pham (2021), demonstrated capabilities of ML and deep learning methods in defining risk factors and prediction of risky loan applications. His work suggested a single model could not perform best in maximizing all performance metrics.

The effects of ML and AI techniques in credit scoring functions has been investigated by many researchers. A literature survey by Dastile, Celik and Potsane (2020), not only explains the typical framework and structure of such articles, it also evaluates the results and possible fruitful areas of future research. The article points out that ensemble models, consisting of more than one classifier tend to perform better in credit scoring and categorization and special attention should be given to working with imbalanced datasets, where categories are not equally represented. Junior et al. (2019), stressed the importance of handling imbalanced datasets in their article and proposed some techniques of ML to improve model performance, when approaching skewed credit scoring data. When dealing with credit, financial institutions are also evaluated in terms of their riskiness. Li et al. (2020), tested several ML models on evaluating of bank credit ratings.

Investment decision support is one area of finance that ML can find practical use. Culkin and Das (2017), produced a model in their article where deep learning techniques were used to accurately calculate option prices. Their findings suggest that AI models can be trained to simulate actions of traders and offer ease of use to investment specialist. Shanmugananthan (2020), addressed the use of AI in behavioral finance areas and use of robo-advisors in investment which are defined as portfolio generating algorithms. The paper suggests ML technology can offer cheaper and efficient investment advice, without emotional factors observed in human advisors.

Qu et al. (2019) published a review of bankruptcy prediction models using AI techniques. In the article by Hamori et al. (2018) performance of ML and DL techniques are compared. The study concluded in the tested areas boosting came up as the superior method compared to ML models and DL techniques. A deep neural network application to default prediction classification model study was presented in an article by Bayracı and Susuz (2019). Their model was tested using Python language on a Turkish dataset. Their conclusion suggests deep learning performed better than regression models in complex datasets.

In the area of fraud detection function of ML, Shah and Metha (2021) compared machine learning based classification techniques in credit card identification. While many algorithms performed well in their experiments, a combined method was advised for improved metric figures. Another work that offered a similar advice had come from Kahn et al. (2020). Their tests on credit card fraud detection with deep learning and ensemble learning techniques concluded a combined algorithm would perform the best.

An important part of machine learning and finance research is based on its ability to work efficiently with big data. An article by Wang (2021) investigated the use of machine learning techniques that are applied to big data of Chinese customers. Another article by Gao and Xiao (2021) focused on credit risk management on big data. Their proposed model performed better than traditional risk score based categorization in their experiments. The article from Sanz and Zhu published in 2021 focuses on scalability aspect of AI practices in finance. Some of the challenges of using ML in finance sector are listed as secrecy and privacy of data, problems with sharing datasets among institutions, models being too dependent on specific datasets and long processing times.

Finally a subject that should not be forgotten is the ease of use of AI applications in finance by the professionals in the sector. The article by Černevičienė and Kabašinskas (2022) pointed out that ML applications could operate as a black box and their outputs cannot really be explained or interpreted by decision makers. They highlighted the term "Explainable Artificial Intelligence" and proposed a model framework that could be used to produce transparent and interpretable AI based methods for users of the technology. Gambhir and Bhattacharjee (2021) pointed to the fact that the importance of AI in accounting and finance functions are already affecting human resource decisions in corporations. With the increase of AI applications in everyday tasks, finance professionals will have to work regularly with relevant algorithms, should be able to adopt and be ready to develop themselves in required areas.

5. APPLICATION OF ML TECHNIQUES TO SOUTH GERMAN CREDIT DATASET

5.1 The Scope and Aims

The aim of this work is to examine the use and effectiveness of machine learning techniques in a real world example where an anonymous real dataset, non-specialized consumer computer hardware, open source software and very limited coding skills are used. Current events and research indicates that ML applications will continue to expand their scope and variety of uses in the future. While advancements in ML algorithms or data engineering could expected from experts in respective fields, people from outside computing and software development areas are likely to use some part AI applications one way or another.

This work demonstrates if and how can a researcher without specialized software skills use existing ML algorithms in order to find a solution to a real world problem in finance area. Several works have been published that compare ML algorithms' and models' performance in assessing credit risk and categorizing loan applications. Instead of taking a currently created, very detailed dataset and shaping it into a form where ML algorithms would perform best, in this research a dataset with limited number of examples and variables is used with little to no feature engineering. In addition, models are not configured to work best with the given dataset.

In accordance with this aim a standard ML model is created to assess loan applications. Five ML algorithms are integrated to this model and three performance measures of tested algorithms are presented.

Due to characteristics of the business, an automation of loan application assessment may not refer to a process that is completely handled by computers without any human intervention. However, precise and timely responses to credit demand would require use of all possible computing power of an organization. Definition and current research suggest the more the data the more precise the AI performance. To examine if use of data that is seemingly unserviceable to assessing credit risk could still improve performance of an algorithm, four subsets of dataset is created with each one having fewer number of variables for each application. Testing the data with these subsets could demonstrate the effects of removing seemingly irrelevant data to model performance.

5.2 The Data

Dataset used for this work titled “German Credit Dataset” is taken from University of California Irvine Machine Learning Repository website (Dua and Graff, 2019). The website is particularly developed for providing example data for researchers and students of machine learning. According to the website the database was donated in 1994, consists of 1000 anonymous instances of credit applications with 20 attributes of which all of them are expressed in integers while some of them are categorical and some of them are numerical. Dataset includes 300 bad or in this case risky and 700 good or acceptable loan applications.

Apart from “credit_amount”, “duration” and “age” all attributes are represented in categories. Although this arrangement simplifies the dataset, it also limited the possibility of modifying and gathering information from the data. For example instead of gender information, dataset combined gender and marital status as “personal_status”, in a way that gender cannot be inferred from the dataset. Also the reason for the loan application is categorized with ten possible types, which caused some types of having no or very little number of instances, making this attribute less effective in determining approval process.

The database, since it was most likely modified in such a way by its creator for educational purposes, is not missing any values for any attributes therefore no imputation or deletion is necessary. Also since all variables are presented in numbers, dataset is ready to be used as is without conversion of data types. Credit application approval ratio numbers from Federal Reserve Bank of New York’s Credit Access Survey in 2022 shows an average of credit denial rate below 20% percent within a timeframe of 10 years (New York Fed SCE, 2022). Credit denial rate of 30% in German Credit Dataset may represent a set that leaned to declining application more than average but considering economic environment of the particular time frame when the dataset was created and policies of the financial institution that received the applications, such a ratio may be realistic. Also a denial rate closer to 50% will discourage the algorithms from evaluating applications more forgiving and showing results closer to training dataset approval rates, producing higher number of false positives.

In addition to the German dataset, three artificial datasets are created that include randomly generated data to test and analyze further the mechanism and characteristics of categorization algorithms in different settings. These three datasets each contain 10.000 examples of credit applications, generated randomly by simulating the same distribution types using “fitter” function of Python. Since the category of these randomly generated instances are also random, there will not be a real learning situation for the algorithms, and the software will just try to fit the existing information to a random determination of riskiness. That being said this non learning random data algorithm can still provide information on performance of ML algorithms by providing differences in real world and artificial random data settings.

Scenario 1 of the artificial datasets has the exact proportion for each type of independent variable values with the same distribution and simulates the exact same settings with German dataset, differing only in the number of instances, i.e. 10.000 instead of 1000. Scenario 2 simulates a setting where economic expansion is in effect and only 5 percent of applications are deemed risky and approved credits correspond to 95 of all applications while keeping all other data value at their means and distributions are kept the same. In scenario 3, a recession is simulated where only 50 percent of all credit applications are approved. Also instead of all the other scenarios where the majority of applicants request a credit amount closer to the theoretical minimum, in scenario 3 most applicants apply for a theoretical maximum amount of credit. To simulate such a situation, instead of randomly assigning the credit amount, a series of amount data is created to fit a beta distribution that is symmetrical to the true probability distribution of the amount data in the original dataset. This way the amount probability distribution is set to be skewed to the left instead of right.

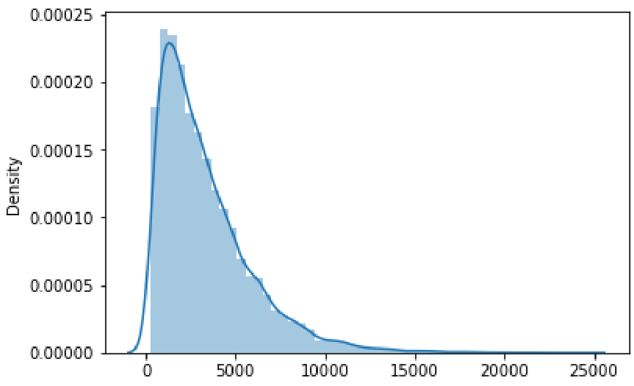

Figure 5.1 Probability Distribution Graph of “amount” in Scenarios 1 and 2

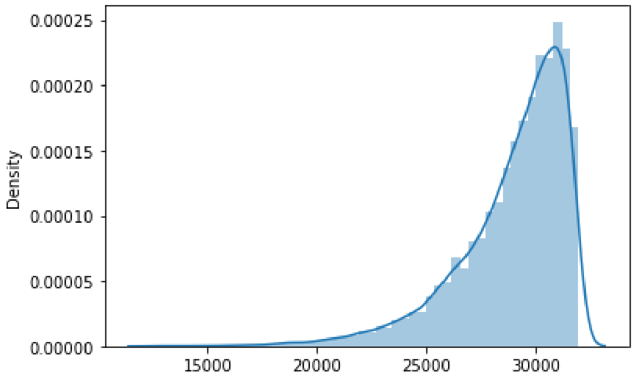

Figure 5.2 Probability Distribution Graph of “amount” in Scenario 3

5.3 The Environment and Algorithms

In order to conduct the tests, Python programming language is used through Conda Environment, and Jupyter Notebook platform. Machine learning algorithms and other necessary tools for measuring performance are handled through codes from scikit-learn library. All programing tools used for ML applications in this work are available for free. Software and codes are run in a mobile computer with Windows 10 operating system.

5.4 Models and Testing

ML categorization processes follow training, validation and testing steps. These steps require separate datasets to train the model, calibrate its hyper parameters and measure its performance. In order to simplify these processes and improve performance of the model where an imbalanced dataset is used a cross sampling method can be used. Cross sampling takes the entire dataset and separates it into a given number of subsets, which are used for training, calibrating and testing at the same time with subsets containing the same ratio of each class without contamination. This way a need for separate datasets testing and validation is eliminated. In this experiment, a Stratified K-Fold cross validation with K being equal to 5, which means 4 folds of data are used for training and 1 is used for testing for a total of 5 times. The performance measures of models are then calculated by taking average performance of these folds.

To test if removing less relevant information from dataset improves model performance, 4 data subsets are created, the first one being the complete dataset and each iteration with 5 less features. Feature definitions, correlation scores and subset information is given in Appendices A and B. Tests with artificial datasets are conducted with complete 20 attributes included and compared to the original German dataset test results for an appropriate comparison.

Three performance measures are calculated to compare performances of selected models with chosen data subsets: accuracy score, F1 score and precision score.

5.4.1 Logistic Regression

Logistic regression is a method used in classification problems. Unlike logistic regression which calculates a numeric value in a given estimation problem, logistic regression results are expressed in binary as zeroes and ones. Instead of ordinary least squares, parameters are selected in maximum likelihood method and given a threshold model categorizes an input in to 1 or 0.

5.4.2 Decision Tree

A decision tree is a machine learning technique that separates the data into categories which are then categorized into further categories until no further gain is available for decision making. It somewhat mimics human decision methods and can be used with both continuous and categorical variables.

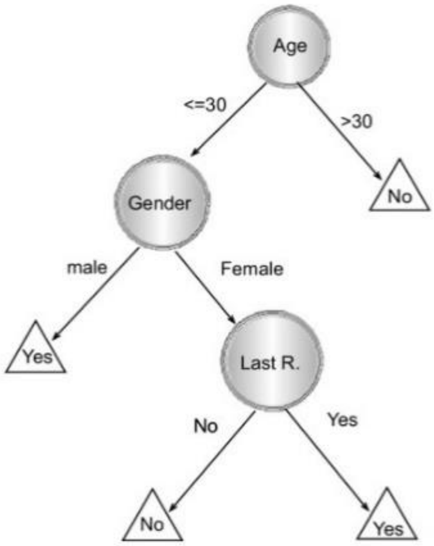

Figure 5.3 - A decision tree example (Rokach and Maimon, 2005)

5.4.3 Random Forest

A random forest combines several decision trees and uses subsets of the dataset to improve predictive performance. It proposed many advantages to decision trees since it can work with large datasets with multiple dimensions efficiently and do not tend to overfit even when significant amount of data is missing (Cutler et al., 2012).

5.4.4. Support Vector Machine

Support vector machines (SVM) uses hyperplanes to distinguish classes in an N-dimensional space where the number of features are represented by N (Boulanger et al., 2015). In order to separate classes it aims to find a plane in space that has the furthest distance from data points of classes. It works well with non-linear data and can be used for both regression and classification tasks. SVM can perform even when number of features is greater than number of samples.

5.4.5 XGBoost

XGBoost or Extreme Gradient Boosting is an open source gradient boosting algorithm. Gradient boosting is a machine learning techniques that combines several weaker algorithms, usually decision trees, in order to create a robust prediction model.

XGBoost is a flexible solution that operates very fast and works well with missing data. It is widely used and became the choice of many winners of machine learning competitions (Chen and Guestrin, 2016).

5.5 Performance Metrics

To assess the performance of the ML models several metrics can be used. A financial institution would like to maximize its profits by both giving out loans to customers that is most likely to pay back on time and avoid loaning to customers that are considered more likely to fail paying back on time.

According to definitions of the model not loaning to a customer with good credit would mean a false negative and loaning to a customer with bad credit would mean a false positive. A false positive is likely to cost more to a financial institution since there is the possibility of not recovering majority of the capital loaned in the first place. A false negative also means a loss but it is limited to possible servicing and interest profits from a turned down customer. Therefore a metric that illustrates false positives would be most useful from the perspective of the loaning institution.

The experiment includes performance figures presented in three metrics. Accuracy score shows percentage performance of true and false predictions. A perfect accuracy score of 1 would mean all instances are predicted correctly while a score of 0 would mean that all predictions were wrong. Precision score would show a ratio of all true positives to all positives, thus the models with higher precision scores can be more useful in minimizing false positives and profit maximization. F1 score is a combination of precision and recall, recall being the ability of the model to detect positives. Instead of using recall by itself, combining it with precision and using F1 score as a metric can provide a better insight of model performance in case of credit categorization.

5.6 Results and Interpretation

5.6.1 South German Dataset

In terms of accuracy highest score was obtained by XGBoost algorithm when every attribute was available. Except random forest, all models suffered some form of decreased performance with decreased data, yet random forest seems to perform better with each removed feature set.

Table 5.1 Accuracy Scores

Accuracy Score	**DATA SUBSETS**			
Algorithms	**X**	**X2**	**X3**	**X4**
Logistic Regression	0.744	0.738	0.737	0.728
Decision Tree	0.693	0.683	0.685	0.683
Random Forest	0.716	0.713	0.733	0.739
SVM	0.755	0.735	0.744	0.744
XGB	0.758	0.744	0.757	0.700

In comparison to other metrics, models get higher F1 scores in general. With the use of alternative data subsets, random forest model F1 performance did not change at all while logistic regression and decision tree models showed very little differences. Highest F1 scores were achieved by SVM and XGB algorithms using the entire dataset. Logistic regression algorithm consistently scored among best across all subsets.

Table 5.2 F1 Scores

F1 Score	**DATA SUBSETS**			
Algorithms	**X**	**X2**	**X3**	**X4**
Logistic Regression	0.826	0.823	0.825	0.820
Decision Tree	0.777	0.770	0.774	0.770
Random Forest	0.783	0.783	0.783	0.783
SVM	0.834	0.825	0.820	0.824
XGB	0.834	0.822	0.832	0.789

Precision score is the most likely measure an institution would try to maximize since it would shape the decision process in a way that can minimize loans delinquency and default and therefore maximize profits. For this metric SVM performs the best with X3 subset which includes the most relevant ten variables. Similarly XGB gets a score close to SVM with X3 subset. Apart from logistic regression all models showed an improvement in performance when less extensive subsets are used instead of the whole database. Logistic regression seems to suffer in precision when attributes are removed from samples, while random forest model performance got with every decrease in data variety.

Table 5.3 Precision Scores

Precision	**DATA SUBSETS**			
Algorithms	**X**	**X2**	**X3**	**X4**
Logistic Regression	0.78	0.772	0.767	0.759
Decision Tree	0.789	0.781	0.776	0.780
Random Forest	0.713	0.715	0.731	0.747
SVM	0.792	0.767	0.810	0.796
XGB	0.801	0.799	0.805	0.774

5.6.2 Artificial Dataset Scenarios

Exact same tests are implemented to the three datasets that represent three scenarios and results are listed below. Results of complete German dataset is included in the tables to better view the differences between a real and randomly generated situations.

Table 5.4 Scenario Accuracy Scores

Accuracy Score	DATASETS			
Algorithms	**German**	**Scenario 1**	**Scenario 2**	**Scenario 3**
Logistic Regression	0.744	0.701	0.952	0.503
Decision Tree	0.693	0.573	0.895	0.496
Random Forest	0.716	0.701	0.952	0.504
SVM	0.755	0.699	0.952	0.510
XGB	0.758	0.652	0.951	0.504

Taking into account that credit approval rates were set at 70 percent in South German dataset and Scenario 1, 95 percent in Scenario 2 and 50 percent in Scenario 3, it is obvious that with artificial data where an approval is totally random instead of consistent factors, the algorithms cannot really learn categorization instead they try to fit existing data to the random examples they were given. All algorithms except decision tree, fit their settings in a way that would produce the exact approval rate given in the training dataset. This situation also proved even with a very limited sample size and no feature engineering, machine learning algorithms are able to learn and perform better in a real world situation compared to working with randomly generated data.

Table 5.5 Scenario F1 Scores

F1 Score	DATASETS			
Algorithms	**German**	**Scenario 1**	**Scenario 2**	**Scenario 3**
Logistic Regression	0.826	0.824	0.975	0.401
Decision Tree	0.777	0.691	0.944	0.496
Random Forest	0.783	0.824	0.975	0.508
SVM	0.834	0.823	0.975	0.487
XGB	0.834	0.780	0.975	0.503

With F1 scores Scenario 1 and 2 performed better than their training dataset approval percentages yet 3 algorithms in Scenario 1 and 4 algorithms in Scenario 2 produced the exact same results. The interesting find in this setting was that algorithms still produced the best performance in a real world situation compared to an artificial one regarding Scenario 1; however highest performances of F1 scores were much closer compared to accuracy scores. Also XGBoost algorithm failed to produce highest F1 scores for artificial datasets and all algorithms failed worse in F1 than accuracy for dataset Scenario 3.

Table 5.6 Scenario Precision Scores

Precision Score	**DATASETS**			
Algorithms	**German**	**Scenario 1**	**Scenario 2**	**Scenario 3**
Logistic Regression	0.780	0.701	0.952	0.403
Decision Tree	0.789	0.703	0.952	0.496
Random Forest	0.713	0.701	0.952	0.504
SVM	0.792	0.701	0.952	0.510
XGB	0.801	0.701	0.952	0.504

Comparing precision scores further confirms that algorithms cannot really make a difference when they are not properly taught by the existing data. In order machine learning to define risky credit applications and prevent false positives, the algorithm has to be supplied by real world information that has genuine correlations. This situation also showed that users of machine learning algorithms should be cautious of the fact that performance levels very close or lower than the training data's success ratio could indicate that the algorithm used may not be performing as planned, instead it might be producing results that would only perform for the given training dataset.

6. CONCLUSION

Although ML techniques used for the experiments in this work proved to be significantly higher than random predictions, given the nature of credits and institutions performance metrics are far from being acceptable by any business format. With the use of very limited database and basic methods creating a model that could be used in a real world environment is out of question. That being the case, performance results calculated by the models can still give the reader insights on how machine learning in financial area functions and how it could be utilized.

Test results demonstrates no single model is the best throughout all variations of the dataset and metric type. Depending on the characteristics of the data and the metric that is to be maximized, the best option for a model changes. Also presuming that every extra attribute a dataset has would improve the prediction performance could be wrong. In the credit dataset used, limiting the data dimensions by removing some less relevant variables improved model performance. Although, keeping in mind the limited nature of the dataset, it can be claimed there is still a possibility that with a database with higher instances and more relevant personal information that could affect credit payments such as employer, residence and personal health data, models could have performed better. In this case it all comes to testing and deciding which information is relevant and which is not. ML models will inherently help with this decision.

Testing the algorithms with artificial datasets provided interesting insights. Since artificial datasets are randomly generated and the riskiness decision in those datasets are not logically bounded to the credit application information in the datasets, the algorithms cannot really learn from them. However the performance results acquired by using artificial scenarios proved that the original German dataset was indeed able to teach algorithms on the essence of risk assessment and the user should be aware of the fact that one can still accomplish high metric scores with completely random data, which in turn would generate models that would not perform as expected in the real world.

Gathering data for a better performing model inevitably brings with it the question of legal use and privacy of personal data. In order to improve their prediction performance and maximize profits for a loan categorization application a financial institution may want to collect and use every detail of an individual's life. Even with the consent of the individual, making use of such data would raise ethical questions. In a way, privacy laws protecting the customer is likely to hinder true performance of ML algorithms in the area of individual's customer financing.

Another part of the discussion on using ML in finance would be utilizing the hardware, power and time. SVM model experiments that were carried out in this work took 100 to 60 minutes, about 600 to 300 times more than other models. In a server environment with specialized hardware, processing times could be significantly diminished yet in an environment where seconds count, minimal differences in processing time could be crucial. This situation brings another decision to be made by the users of ML, if extra efficiency is worth the additional time needed to achieve it.

By bringing AI precision and computing power to decision making in areas of business that were solely under human jurisdiction, ML is likely to find more uses in finance in the future. However, for the end user of the technology, it may not be in a stage to become an everyday tool. Before ML models can achieve their best performance they require the data to be used to be engineered and require testing and calibrating. At least until standardized techniques are simplified and put into everyday use, advancements and real life practices, development of ML models will stay more of a computing and software subject rather than a business and finance one.

LIST OF REFERENCES

Alballa, N., & Al-Turaiki, I. (2021). Machine Learning Approaches in COVID-19 Diagnosis, Mortality, and Severity Risk Prediction: A review. Informatics in Medicine Unlocked, 24, 100564. https://doi.org/10.1016/j.imu.2021.100564.

Alperen, N. (2019). Tüketici Kredisi Taleplerinin Yapay Öğrenme Modelleriyle Değerlendirilmesi, MSc Thesis, Istanbul Technical University, Management Engineering Department, Istanbul.

Bayraci, S., & Susuz, O. (2019). A Deep Neural Network (DNN) Based Classification Model in Application to Loan Default Prediction. Theoretical and Applied Economics, 75-84.

Boulanger, D., Mercier-Laurent, E., Owoc, M., & Alsqour, M. (2015). Artificial Intelligence for Knowledge Management: Second IFIP WG 12.6 International Workshop, AI4KM 2014, Warsaw, Poland, September 7-10, 2014.

Černevičienė, J., & Kabašinskas, A. (2022). Review of Multi-Criteria Decision-Making Methods in Finance Using Explainable Artificial Intelligence. Frontiers in artificial intelligence, 5, 827584. https://doi.org/10.3389/frai.2022.827584.

Consumer Financial Protection Bureau. (2020, September). “What Is a Credit Score?” https://www.consumerfinance.gov/ask-cfpb/what-is-a-credit-score-en-315/.

Culkin, R., & Das S. R. (2017). Machine Learning in Finance: The Case of Deep Learning for Option Pricing. Journal of Investment Management, v15(4), 92-100.

Culnan, M. & Armstrong, P. (1998). Information Privacy Concerns, Procedural Fairness and Impersonal Trust: An Empirical Investigation. Organization Science. 10. 10.1287/orsc.10.1.104.

Cutler, A., Cutler, D.R., & Stevens, J.R. (2012). Random Forests. In: Zhang, C. and Ma, Y.Q., Eds., Ensemble Machine Learning, Springer, New York, 157-175. http://dx.doi.org/10.1007/978-1-4419-9326-7_5.

Dastile, X., Çelik, T., & Potsane, M.M. (2020). Statistical and Machine Learning Models in Credit Scoring: A Systematic Literature Survey. Applied Soft Computing Journal, 91, 106263.

Dua, D., & Graff, C. (2019). UCI Machine Learning Repository, http://archive.ics.uci.edu/ml, Irvine, CA: University of California, School of Information and Computer Science.

Emerson, S., Kennedy, R., O'Shea, L., & O'Brien, J.R. (2019). Trends and Applications of Machine Learning in Quantitative Finance. Machine Learning eJournal.

European Data Protection Supervisor. (2021). Opinion 11/2021 on the Proposal for a Directive on Consumer Credits. https://edps.europa.eu/system/files/2021-08/opinion_consumercredit-final_en.pdf.

Fair Isaac Corporation. (2018). "Frequently Asked Questions about Fico® Scores". https://www.ficoscore.com/ficoscore/pdf/Frequently-Asked-Questions-About-FICO-Scores.pdf.

Federal Reserve Bank of New York. (2022). Sce Credit Access Survey. https://www.newyorkfed.org/microeconomics/sce/credit-access#/experiences-credit-applications1, 08.05.2022.

Fradkov, A. (2020). Early History of Machine Learning. IFAC-PapersOnLine. 53. 1385-1390. 10.1016/j.ifacol.2020.12.1888.

Gambhir, B., & Bhattacharjee, A. (2021). Embracing the Role of Artificial Intelligence in Accounting and Finance: Contemplating the Changing Skillset Expectations. Development and Learning in Organizations, Vol. 36 No. 1, pp. 17-20. https://doi.org/10.1108/DLO-01-2021-0016.

Gao, L., & Xiao, J. (2021). Big Data Credit Report in Credit Risk Management of Consumer Finance. Wireless Communications and Mobile Computing. 2021. 1-7. 10.1155/2021/4811086.

Google Cloud. (2022). Cloud TPU. https://cloud.google.com/tpu.

Hamori, S., Kawai, M., Kume, T., Murakami, Y., & Watanabe, C. (2018). Ensemble Learning or Deep Learning? Application to Default Risk Analysis. Journal of Risk and Financial Management, 11(1), 12. https://doi.org/10.3390/jrfm11010012.

He, J., Du, M., & Qiao, L. (2020). Competitive Landscape and Patent Activity of Artificial Intelligence Technology in Finance. 330-337. 10.1109/ICAIBD49809.2020.9137450.

Sanz, J. L. C., & Zhu Y. (2021). Toward Scalable Artificial Intelligence in Finance. IEEE International Conference on Services Computing (SCC), pp. 460-469, doi: 10.1109/SCC53864.2021.00067.

Li, J., Mirza, N., Rahat B., & Xiong D., (2020). Machine Learning and Credit Ratings Prediction in the Age of Fourth Industrial Revolution. Technological Forecasting and Social Change, Volume 161, 2020, 120309, ISSN 0040-1625, https://doi.org/10.1016/j.techfore.2020.120309.

Junior, L., Nardini, F. M., & Renso, C., & Macedo, J. (2019). An Empirical Comparison of Classification Algorithms for Imbalanced Credit Scoring Datasets. 747-754. 10.1109/ICMLA.2019.00133.

Khan Z.F., & Alotaibi S.R. (2020). Applications of Artificial Intelligence and Big Data Analytics in m-Health: A Healthcare System Perspective. J Healthc Eng. 2020 Aug 30;2020:8894694. doi: 10.1155/2020/8894694. PMID: 32952992; PMCID: PMC7481991.

Kredi Kayıt Bürosu. (2020). “Findeks Kredi Notu”. https://www.findeks.com/urunler/findeks-kredi-notu.

Microsoft Corporation. (2022). Seeing AI. https://www.microsoft.com/en-us/garage/wall-of-fame/seeing-ai/.

Molloy D. (2021). Google's Pixel 6 Processor Brings AI Photo Features, BBC News, 19 October 2021. https://www.bbc.com/news/technology-58955304.

Ongsulee, P. (2017). Artificial Intelligence, Machine Learning and Deep Learning. 15th International Conference on ICT and Knowledge Engineering (ICT&KE). pp. 1-6. DOI: 10.1109/ICTKE.2017.8259629.

O'Sullivan, A., & Sheffrin, S. M. (2003). Economics: Principles in Action. Needham, Mass: Pearson Prentice Hall. p. 512. ISBN 0-13-063085-3.

Pham, H. P., (2021), Risk Detection In Consumer Credit Loan - A Comparative Study Of Machine Learning Algorithms And Artificial Neural Network In Detecting Risk With Loan Application Data, MSc Thesis, Tilburg University, School of Humanities and Digital Sciences ,Department of Cognitive Science & Artificial Intelligence, Tilburg, The Netherlands.

Rezapouraghdam, H., Akhshik, A. & Ramkissoon, H. (2021). Application of Machine Learning To Predict Visitors’ Green Behavior in Marine Protected Areas: Evidence from Cyprus. Journal of Sustainable Tourism. DOI: 10.1080/09669582.2021.1887878.

Rokach, L., & Maimon, O. (2005). Top-Down Induction Of Decision Trees Classifiers - A Survey. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), Systems, 35(4), 476–487. https://doi.org/10.1109/TSMCC.2004.843247.

Sachan, S., Yang, J., Xu, D., Benavides, D., & Li, Y. (2019). An Explainable AI Decision-Support-System to Automate Loan Underwriting. Expert Systems with Applications. 144. 113100. 10.1016/j.eswa.2019.113100.

Saleh, B., Abe, K., Arora, R., & Elgammal, A. (2014). Toward Automated Discovery of Artistic Influence. Multimedia Tools and Applications, 75, 3565-3591.

SCHUFA Holding AG. (2022). “So Funktioniert Die Schufa”. https://www.schufa.de/ueber-uns/schufa/so-funktioniert-schufa/.

Shanmuganathan, M. (2020). Behavioural Finance in an Era of Artificial Intelligence: Longitudinal Case Study of Robo-Advisors in Investment Decisions. Journal of Behavioral and Experimental Finance, 27, 100297.

Chen, T., & Guestrin, C. (2016) XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16). Association for Computing Machinery, New York, NY, USA, 785–794. https://doi.org/10.1145/2939672.2939785.

United States of America Federal Trade Commission. Consumer Credit Law & Practice in the U.S. https://www.ftc.gov/sites/default/files/attachments/training-materials/law_practice.pdf.

Wang H. (2021). Credit Risk Management of Consumer Finance Based on Big Data. Mobile Information Systems. vol. 2021, Article ID 8189255, https://doi.org/10.1155/2021/8189255.

Qu, Y., Quan, P., Lei, M., & Shi, Y. (2019). Review of Bankruptcy Prediction Using Machine Learning and Deep Learning Techniques, Procedia Computer Science, Volume 162, Pages 895-899, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2019.12.065.

APPENDICES

APPENDIX A: VARIABLE NAMES AND EXPLANATIONS

`laufkont = status`

1 : no checking account

2 : ... < 0 DM

3 : 0<= ... < 200 DM

4 : ... >= 200 DM / salary for at least 1 year

`laufzeit = duration`

`moral = credit_history`

0 : delay in paying off in the past

1 : critical account/other credits elsewhere

2 : no credits taken/all credits paid back duly

3 : existing credits paid back duly till now

4 : all credits at this bank paid back duly

`verw = purpose`

0 : others

1 : car (new)

2 : car (used)

3 : furniture/equipment

4 : radio/television

5 : domestic appliances

6 : repairs

7 : education

8 : vacation

9 : retraining

10 : business

`hoehe = amount`

`sparkont = savings`

1 : unknown/no savings account

2 : ... < 100 DM

3 : 100 <= ... < 500 DM

4 : 500 <= ... < 1000 DM

5 : ... >= 1000 DM

`beszeit = employment_duration`

1 : unemployed

2 : < 1 yr

3 : 1 <= ... < 4 yrs

4 : 4 <= ... < 7 yrs

5 : >= 7 yrs

`rate = installment_rate`

1 : >= 35

2 : 25 <= ... < 35

3 : 20 <= ... < 25

4 : < 20

`famges = personal_status_sex`

1 : male : divorced/separated

2 : female : non-single or male : single

3 : male : married/widowed

4 : female : single

`buerge = other_debtors`

1 : none

2 : co-applicant

3 : guarantor

`wohnzeit = present_residence`

1 : < 1 yr

2 : 1 <= ... < 4 yrs

3 : 4 <= ... < 7 yrs

4 : >= 7 yrs

<u>`verm = property`</u>

1 : unknown / no property

2 : car or other

3 : building soc. savings agr./life insurance

4 : real estate

<u>`alter = age`</u>

<u>`weitkred = other_installment_plans`</u>

1 : bank

2 : stores

3 : none

<u>`wohn = housing`</u>

1 : for free

2 : rent

3 : own

<u>`bishkred = number_credits`</u>

1 : 1

2 : 2-3

3 : 4-5

4 : >= 6

<u>`beruf = job`</u>

1 : unemployed/unskilled - non-resident

2 : unskilled - resident

3 : skilled employee/official

4 : manager/self-empl./highly qualif. employee

<u>`pers = people_liable`</u>

1 : 3 or more

2 : 0 to 2

<u>`telef = telephone`</u>

1 : no

2 : yes (under customer name)

<u>`gastarb = foreign_worker`</u>

1 : yes

2 : no

`kredit = credit_risk`

0 : bad

1 : good

APPENDIX B: CORRELATION COEFFICIENTS AND TRAINING SUBSETS

Feature	Correlation	Absolute Value	Group
account_status	0.350847	0.350847	1
credit_history	0.228785	0.228785	1
duration	-0.214927	0.214927	1
savings	0.178943	0.178943	1
amount	-0.15474	0.15474	1
property	-0.142612	0.142612	2
employment_duration	0.116002	0.116002	2
other_payment_plans	0.109844	0.109844	2
age	0.091272	0.091272	2
personal_status	0.088184	0.088184	2
foreign_worker	-0.082079	0.082079	3
installment_rate	-0.072404	0.072404	3
number_of_credits	0.045732	0.045732	3
telephone	0.036466	0.036466	3
job	-0.032735	0.032735	3
other_debtors	0.025137	0.025137	4
housing	0.018119	0.018119	4
purpose	-0.017979	0.017979	4
people_liable	-0.003015	0.003015	4
present_residence	-0.002967	0.002967	4

Training Subsets:

Subset X4 includes all features in Group 1

Subset X3 includes all features in Group 1 and 2

Subset X2 includes all features in Group 1, 2 and 3

APPENDIX C: DESCRIPTIVE STATISTICS OF GERMAN DATASET

	account status	duration	credit history	purpose	amount
Count	1000	1000	1000	1000	1000
Mean	2.577	20.903	2.545	2.828	3271.248
Std. Dev.	1.257637727	12.05881445	1.083119637	2.74443946	2822.75176
Min.	1	4	0	0	250
25%	1	12	2	1	1365.5
50%	2	18	2	2	2319.5
75%	4	24	4	3	3972.25
Max.	4	72	4	10	18424
	savings	**employment duration**	**instalment rate**	**personal status**	**other debtors**
Count	1000	1000	1000	1000	1000
Mean	2.105	3.384	2.973	2.682	1.145
Std. Dev.	1.580022617	1.208306254	1.118714674	0.708080064	0.477706189
Min.	1	1	1	1	1
25%	1	3	2	2	1
50%	1	3	3	3	1
75%	3	5	4	3	1
Max.	5	5	4	4	3
	present residence	**property**	**age**	**other payment plans**	**housing**
Count	1000	1000	1000	1000	1000
Mean	2.845	2.358	35.542	2.675	1.928
Std. Dev.	1.103717896	1.050208998	11.35267013	0.705601072	0.530185908
Min.	1	1	19	1	1
25%	2	1	27	3	2
50%	3	2	33	3	2
75%	4	3	42	3	2
Max.	4	4	75	3	3
	number of credits	**job**	**people liable**	**telephone**	**foreign worker**
Count	1000	1000	1000	1000	1000
Mean	1.407	2.904	1.845	1.404	1.963
Std. Dev.	0.577654468	0.653613962	0.362085772	0.490942996	0.188856206
Min.	1	1	1	1	1
25%	1	3	2	1	2
50%	1	3	2	1	2
75%	2	3	2	2	2
Max.	4	4	2	2	2

	credit risk
Count	1000
Mean	0.7
Std. Dev.	0.45848687
Min.	0
Max.	1

APPENDIX D: A SAMPLE GUIDELINE FOR BUILDING A MACHINE LEARNING MODEL FOR BEGINNERS

1. **<u>Define your problem</u>**

   What will be the ultimate goal of your project? Will you be using the model to categorize items or calculate a certain number or percentage? Answers to these questions will also define the tools you may use.

2. **<u>Set your computing environment</u>**

   Which hardware and software environment fits best to your needs and capabilities? Download and set up necessary software packages accordingly.

3. **<u>Collect the data</u>**

   Keep in mind that size and precision of a dataset is likely to impact the performance of your model.

4. **<u>Format the data for your environment</u>**

   In order to observe the characteristics of your data, and use it with common machine learning algorithms you should format it the way algorithms are programmed to operate.

5. **<u>Check for missing values and eliminate them using appropriate techniques</u>**

   You can either delete items that contain missing data or fill the missing part with compensating techniques, which in this case are called imputations. Either to impute or which techniques to use would be based on the characteristics of your dataset.

6. **<u>Select your algorithms best fitting to your model</u>**

   There are several algorithms available for each type of ML problems. Select algorithms that performs best in models resembling yours.

7. **Train the model**

   Feed your data to the model and run performance metrics.

8. **Observe performances**

   Compare performance of your model with different algorithms.

9. **Optimize parameters**

   After running the data, fine tune your model with hyper-parameter and algorithm settings.

10. **Update when necessary**

    Since your model will do predictions best when it is fed with data which resembles your training dataset, you should train and tune your model once your dataset to predict gets large and displays different characteristics compared to your training dataset.