PREDICTING CREDIT RISK IN THE GERMAN CREDIT SYSTEM

CIDM 6355

FINAL PROJECT

TEAM MEMBERS:

Casey Stusynski	John Anokye
Danielle Burkett	Sanchi Srivastava
Evan Ringel	Vernice Tanquerido

EXECUTIVE SUMMARY

In the German banking sector, effective credit risk assessment is crucial for ensuring timely repayments and maintaining strong credit performance. Despite efforts in credit scoring, persistent loan defaults led to our focused data mining project on the German credit dataset.

Goal/Motivation: Our goal was to address challenges in credit risk assessment specific to the German financial landscape, considering cultural and regulatory nuances.

Method/Data: Utilizing the German Credit dataset, our structured data analytics process included preprocessing and preparation, model prediction, training, and evaluation, parameter tuning and feature engineering, and making predictions and model comparisons. Models such as Logistic Regression, Neural Networks, Decision Trees, and Naïve Bayes were employed with a 70/30 data split for accuracy.

Findings: Our analysis revealed that Neural Network, Logistic Regression, and Naïve Bayes models consistently demonstrated high performance, with Neural Network outperforming others even after data transformation, specifically, in terms of measuring recall. These models effectively classified credit applicants as "Good" or "Bad".

Recommendations/Conclusion: The discussion on implications for the banking sector highlights the staggering potential losses from credit risk, emphasizing the need for effective credit risk assessment. Our models, especially the Neural Network, offer a promising approach to reduce loan defaults and increase profits by identifying high-risk, or "bad credit," applicants.

The discussion further emphasizes key factors affecting loan default, such as checking account status, loan purpose, credit amount, and credit history. It suggests strategies for credit risk management, including incorporating cultural and regulatory factors into assessments and creating attribute-specific assessments.

The German financial landscape, influenced by cultural factors and regulatory frameworks, shapes credit risk assessment. Factors such as aversion to debt, emphasis on saving, stable employment, risk aversion, and a conservative financial culture contribute to the uniqueness of credit risk assessment in Germany (Folkerts-Landau D. et al, 2016).

The limitations and suggestions section acknowledges the Neural Network model's lack of transparency and proposes model stacking to mitigate this. It suggests obtaining a larger dataset for more robust analysis and recommends future research on evolving creditworthiness, economic cycles, and additional data sources, like, transaction history.

The conclusion summarized the project's significance in assessing credit risk in the German banking industry, underscoring the need for a tailored approach. Our models provide powerful tools, particularly the Neural Network, enabling financial institutions to make informed lending decisions and reduce their risk of loan defaults. This can ultimately contribute to a healthier financial landscape in Germany.

I. Introduction

Background: In our modern era, data mining and analytics have become prevalent tools across various industries, revolutionizing the way data is used in making informed decisions. Financial institutions are no exception as they continually seek to improve the loan application process. Data mining has evolved significantly in recent years with advancement in computing power and machine learning algorithms (Addo P., et al, 2018). This advancement has allowed for the extraction of more complex and subtle insights from data than was previously possible.

Financial institutions consider several factors before approving loans for any credit applicant, with one of the main considerations being the credit risk of the applicant. In the German banking system, the SCHUFA-Score is used to assist consumers and lenders in determining creditworthiness (SCHUFA Holding AG, n.d.). This score is based on similar criteria as our dataset. The motivation for this project is to develop a credit risk assessment model using data mining techniques that is more accurate and efficient than traditional methods. Other studies have been conducted and drawn their own conclusions from different data and have determined that loan amount and the duration of the loan have the greatest impact on the determination of creditworthiness (Bachmann, 2020). This project aims to aid financial institutions in reducing their losses and improving their profitability by enhancing decision-making processes and ensuring fair and accurate assessments of credit applicants. Instead of relying solely on credit scores, which are numerical representations of a person's creditworthiness, this project will categorize applicants' credit risks as “Good” or “Bad”. Data mining techniques such as logistic regression, decision trees, neural networks, and Naïve Bayes will be used to predict the creditworthiness of applicants and help reduce the risk of lending to those with high credit risks.

Business Problems to Answer: Credit risk refers to the possibility that a contractual party will fail to meet its obligations as agreed (Brown K. et al, 2014), and it is a critical concern for banks, as they are the custodians of public funds and must maintain their own financial stability. Extending credit to high-risk customers can lead to significant financial loss, reputation damage, and potential business loss. The economic consequences are severe, as evidenced by the 2008 global financial crisis, which was largely triggered by credit defaults on subprime mortgages in the United States. A study by the Federal Reserve Bank of St. Louis found that banks lost on average of $7.3 billion (about $22 per person in the US) per year due to credit defaults from subprime borrowers between 2015 and 2019 (FRED, 2023). Another study conducted by the Mercatus Center at the George Mason University found that banks lost an estimated $150 billion due to credit defaults between 2008-2015 (Miller, 2022).

Stakeholders: Traditionally, financial institutions try to mitigate these losses by employing a multifaceted approach in assessing credit risks of credit applicants. These often include determining the creditworthiness of the applicant by examining their credit history, past financial behavior, employment history, collateral, debt-to-income ratio and many more (Brown K. et al, 2014). Traditional credit risk assessment methods often rely on limited data points and heuristics, which can lead to inaccurate and biased decisions. This can result in financial institutions extending credit to risky borrowers and losing money on bad loans. Additionally, traditional methods may be unable to identify new and emerging credit risk factors, which can expose financial institutions to unexpected losses.

Motivation: The German financial landscape is different than the United States. Cash payments are very common and short-term debt is relatively uncommon (Santander, 2023). Also, there is generally little to no student debt in Germany because public universities are usually free (Push, 2022). **The general lack of debt provides an opportunity for financial institutions to create incentives for the best credit candidates to open new revenue streams. For the borrower, they can perhaps use a loan to maximize the growth in their business or their home.** There is untapped potential in the German financial system because debt is seen as a bad thing and Germans would rather rent than buy their residence because of this sentiment (Sawal, 2018). This prevents lenders and borrowers from maximizing their credit.

This project will leverage the German financial system as a case study, utilizing a dataset sourced from Professor Dr. Hans Hofmann of the University of Hamburg (1994). The dataset, obtained from the UCI Machine Learning Repository, consists of 1000 records. It also includes a cost matrix that underscores the higher cost of misclassifying "Bad" applicants compared to "Good" ones. The output model can be used to help banks determine which applicants are good borrowers versus ones that are not. Also, it can be used by potential borrowers to understand their creditworthiness and the factors that influence it. This will save time and effort from both the lender and the borrower.

Opportunity/Challenge: This dataset presents an opportunity and a challenge. First, the opportunity it provides is the ability for borrowers to see themselves as participants in the credit system. For lenders, this can be an opportunity to take a critical look at what the model produces as indicators for creditworthiness. There could be borrowers who are low risk but are being judged too harshly by the current system. This would result in the lenders missing out on profits from potentially good candidates. Lenders can use this model to determine how best to maximize the capital that they have available for lending. If they have money to lend, they can look at ways to lower the threshold for lending based on these models.

This dataset and resulting models also provide a challenge. These models are only as good as the inputs that we give them. There may be other factors that are linked to creditworthiness. Of the factors evaluated, these models might weigh certain factors higher than others based on observations but not lender policy. While our models will help to determine which factors are weighted higher than others, specifically, through the Logistic Regression and Neural Network models, we cannot determine with high degree of certainty how a change in lender policy will affect the creditworthiness decision without ongoing reevaluation.

Action Plan: Our action plan involves conducting a literature review on data mining techniques, selecting appropriate models for this specific domain, and preparing the data obtained from the UCI repository for analysis. Using tools like R Studio and Rapid Miner, various models will be developed, tuned, and evaluated to determine the most effective credit risk assessment model.

Summary of Research Process: Our research process begins with first understanding the business problem. Through this lens, we can better assess how to use the data at our disposal to model creditworthiness. Next, we must understand the structure, strengths, and limitations of the dataset. Factors such as missing values, coding/decoding, and basic statistics must be understood. Next, the data will be prepared. Since the categorical data in this dataset are coded, we must transform the data into something that is usable by all models and in both Rapid Miner and R. Next, during the modeling phase, we will use Logistic Regression, Decision Tree, Naïve

Bayes, and Neural Network frameworks to model the data using 70% of the existing dataset. In the evaluation phase, we will use the remaining 30% of the data to determine the accuracy, precision, and recall of each model and compare them. After selecting the model that best fits the data and business goals, we will deploy the model to predict creditworthiness in the German credit system.

Figure 1. Research Process

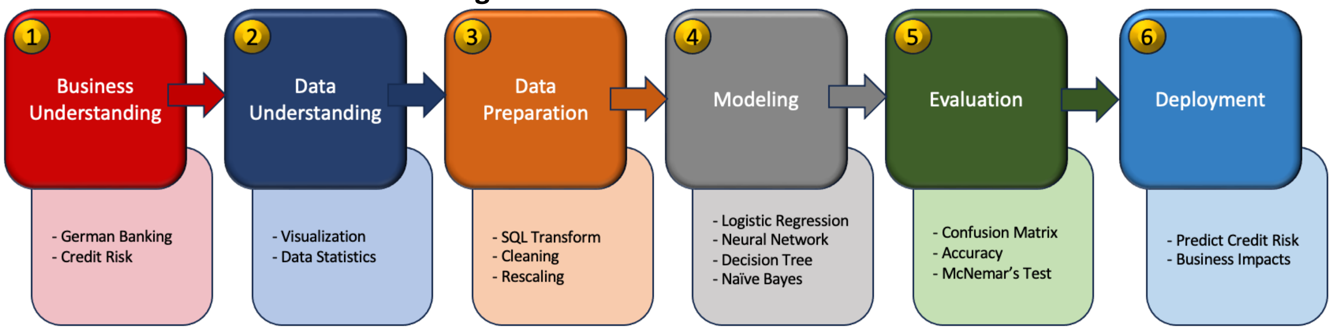


Figure developed by authors.

II. Data Description

Dataset Details

This dataset is from the University of California-Irvine Machine Learning Repository. It was accessed from: https://archive.ics.uci.edu/dataset/144/statlog+german+credit+data. The dataset was created in 1994 using German loan applicant information. It can be used to build a model to make predictions (a binary: "good" or "bad") of whether a customer will receive a loan. There are 1,000 records for every attribute in the dataset. The records have 20 predictor attributes and 1 target attribute, shown in the table in figure 1. These attributes are a mix of categorical and numerical variables. The categorical attributes are automatically coded by R and Rapid Miner (RM) to allow them to be used by models that cannot deal with strings. There are no missing attributes in this dataset. Therefore, this data is suitable for analysis.

The data is related to loan approval in the German banking system. It is designed to predict whether an applicant is a good or bad choice for a loan based on factors such as credit history, purpose of the loan, amount of the loan, employment status, property investments, and existing cash reserves (Hofmann, 1994). In the German banking system, however, a "good" or "bad" result can be over-fitted to a "bad" result. This is because there is a propensity to classify an applicant as bad versus good. If a bank classifies an applicant as bad and, therefore, does not give them a loan, they do not lose any money. If, instead, they classify an applicant as good and that individual defaults on their loan, it is detrimental to the bank. Therefore, it is better to classify an applicant as bad when they are good, and worse to classify an applicant as good when they are bad (Hofmann, 1994). This is presented in the model evaluation section.

The attributes and features of the dataset are described in figure 1 below. During our data preparation, we will determine if there is a strong correlation between any of the attributes and subsequently remove one from the dataset. One limitation of this dataset is that the 1994 German Banking system categories create bins in which each applicant is grouped. Therefore, there is potential that even though one might be a qualified applicant in most factors, a few factors may outweigh the others and sway the decision to "bad" for the applicant.

Table 1. Attributes and descriptions of the German Credit Data dataset

No.	Attribute Name	Type	Category	Description
1	Checking Account Status	Categorical	Financial	Status of existing checking account (DM = Deutsche Mark) - < 0 DM, 0 to < 200 DM - >= 200 DM/salary assignments for at least 1 year - no checking account
2	Credit History			History of credit repayments - no credits taken/all credits paid back duly - all credits at this bank paid back duly - existing credits paid back duly until now - delay in paying off in the past - critical account/other credits existing (not at this bank)
3	Savings Account/Bond			Total of applicant's cash holdings in savings and bonds (DM = Deutsche Mark) - < 100 DM - 100 to < 500 DM - 500 to < 1000 DM - >= 1000 DM - unknown/no savings account)
4	Other Installment Plans			Existing installment commitments - bank - stores - none
5	Purpose			Type of loan applied for - car (used) - car(new) - furniture/equipment - radio/television - domestic appliances - repairs - education - vacation - retraining - business - others
6	Property			Types of property owned - real estate - if not real estate: building society savings agreement/life insurance - if not real estate or building society savings agreement/life insurance: car or other not in attr. 6 - unknown/no property
7	Other Debtors/ Guarantors			Additional debtors, complaints, or guarantors - none - co-applicant - guarantor
8	Present Employment		Employment	Duration of current employment - Unemployed - < 1 year - 1 to < 4 years - 4 to < 7 years - >= 7 years

9	Job	Categorical	Employment	Status and level - unemployed/unskilled - non-resident - unskilled – resident - skilled employee/official - management/self-employed/highly-qualified employee/officer
10	Foreign Worker			Foreign worker status - yes - no
11	Personal Status		Demographic	Gender and marital status application - male: divorced/separated - female: divorced/separated/married - male: single - male: married/widowed - female: single
12	Housing			Housing status - rent - own - free
13	Telephone			Status of telephone - none - yes, registered under the customer's name
14	Duration	Numerical	Financial	Duration of loan payments in months
15	Credit Amount			Amount of credit applied for in DM (DM = Deutsche Mark)
16	Existing Credits at a Bank			Number of existing credit accounts at bank
17	Installment Rate			Installment rate as % of disposable income
18	Present Residence Since		Demographic	Duration of years in current residence
19	Age			Age in years
18	Present Residence Since			Duration of years in current residence
19	Age			Age in years
20	Number of People to Provide Maintenance for			Number of dependents
21	Good or Bad Credit	Categorical/Binomial		1 = Good, 2 = Bad

Figure developed by authors from Hofmann, 1994.

III. Data Preprocessing and Preparation

Descriptive Analysis

We will start by inspecting our dataset, cleaning the data as needed, and addressing any missing values. Any irrelevant attributes will be removed, and we will check for highly correlated features. SQL Server, Excel, R and Rapid Miner will be our tools of choice for data preprocessing. When we use classification models, we will split our dataset into 70% for training and 30% for testing, providing us with 700 records for training and 300 records for testing.

The data provided has 1000 records per attribute and 20 attributes (7 numerical, 13 categorical, excluding the target attribute). The data from the UCI repository provided by Prof. Hofmann (1994) comes in two forms: the first one contains categorical/symbolic attributes. Since some algorithms need numerical attributes, Strathclyde University produced the file "german.data-numeric" which has been edited and several indicator variables added to make it suitable for algorithms which cannot cope with categorical variables. Several attributes that are ordered categorical (such as attribute 17) have been coded as integer.

For the purposes of our project, we loaded the first dataset (containing categorical and symbolic attributes) into SQL Server for transformation and exploratory analysis. We created a legend/mapping from the Word doc provided in the UCI dataset that maps each attribute category and symbols to attribute values. While performing this analysis, we found that there were 12 records in the Purpose column that had a typo in their categorical values – possibly due to fat-fingering /error in copying and pasting the data. This has been corrected for the final dataset and is depicted in Figure 3. In addition, codes on the personal status and sex are not equal since females can either be single or divorced/separated/married (2 categories). In contrast, males can be divorced/separated, single, or married/widowed (3 categories). To tackle this, we have collapsed the 3 male categories into 2 categories - married/divorced/separated/widow.

Categorical Data:

This dataset has many categorical variables. Decision Tree and Naïve Bayes models can use categorical attributes while Neural Network and Logistic Regression cannot. To mitigate this issue, we utilized the built-in features that RM and R provided to automatically convert the categorical attributes to dummy variables.

Handling Missing Data

There is no missing data in this dataset.

Handling Outliers/Anomalies

No anomalies were found in this dataset.

Data Normalization

As a general best practice and to avoid model issues such as overfitting and bias, we have normalized the numeric attributes to ensure equal weights are used during prediction.

An important note about the dataset is that 70% of the applicants are marked as good credit risk. Therefore, there is a tendency for the model to be biased for that result. One way to overcome this is through stratification, and since we are using the 70/30 testing/training holdout method, we should be able to combat this potential issue.

Data Correlation

The highest correlation between any two attributes is 0.625. This correlation is between the duration of the loan and the amount of the loan. We feel that this correlation is not high enough to warrant removing one of these attributes. The duration of the loan as well as its amount are important factors to keep in our model. This correlation matrix was produced using R but a nearly identical matrix was also produced using RM for comparison.

Table 2. German Credit Data Correlation Matrix

Attribute	Duration	Credit amount	Installment rate in percentage of disposable income	Present residence since	Number of existing credits at this bank	Number of people being liable to provide maintenance for
Duration	1.0000	0.6250	0.0747	0.0341	-0.0113	-0.0238
Credit amount	0.6250	1.0000	-0.2713	0.0289	0.0208	0.0171
Installment rate in percentage of disposable income	0.0747	-0.2713	1.0000	0.0493	0.0217	-0.0712
Present residence since	0.0341	0.0289	0.0493	1.0000	0.0896	0.0426
Number of existing credits at this bank	-0.0113	0.0208	0.0217	0.0896	1.0000	0.1097
Number of people being liable to provide maintenance for	-0.0238	0.0171	-0.0712	0.0426	0.1097	1.0000

Results in this table produced by the authors in R.

Removing Attributes

Since age and personal status (male/female and married/divorced/separated/widowed) are not part of the SCHUFA credit scoring in Germany (SCHUFA Holding AG, n.d.), we decided to remove these attributes before designing all four of our models. The remaining attributes aligned more closely with what is included in the SCHUFA credit scoring process.

Train and Test Sets

We will use the holdout method for training and testing our data. In this method, we will use 70% of the data to train the model, then use the remaining 30% for the testing. With 1000 records, this equates to 700 records that will be used for training and 300 that will be used for testing.

IV. Model Results and Interpretation

The models that we will build will be a Logistic Regression (LR), Neural Network (NN), Decision Tree (DT), and Naïve Bayes (NB). Through these models, we are predicting one of two choices, either "good" or "bad". These models are appropriate given the data and the desired outputs.

Logistic Regression Analysis in RM and R

First, we will use logistic regression, which is a regression of a binary variable and appropriate given the binary nature of our target attribute. Logistic regression is sensitive to outliers which can strongly influence coefficients. Fortunately for our dataset, the attributes that are numerical will be normalized to lessen the influence of these strong influences.

Figure 4 depicts attribute weights derived from the dataset. The "tallest" bar represents the attribute "foreign worker = no" and "purpose = business" as well as "purpose = vacation" and "savings account or bonds < = 100 DM." This means that customers who reported not being a foreign worker, utilizing the loan for business or vacation, and stating they had a savings account or bond worth equal to or less than 100 Deutschmarks were the attributes being the strongest contributing factors to the results of the model.

Figure 4. Attribute Weights

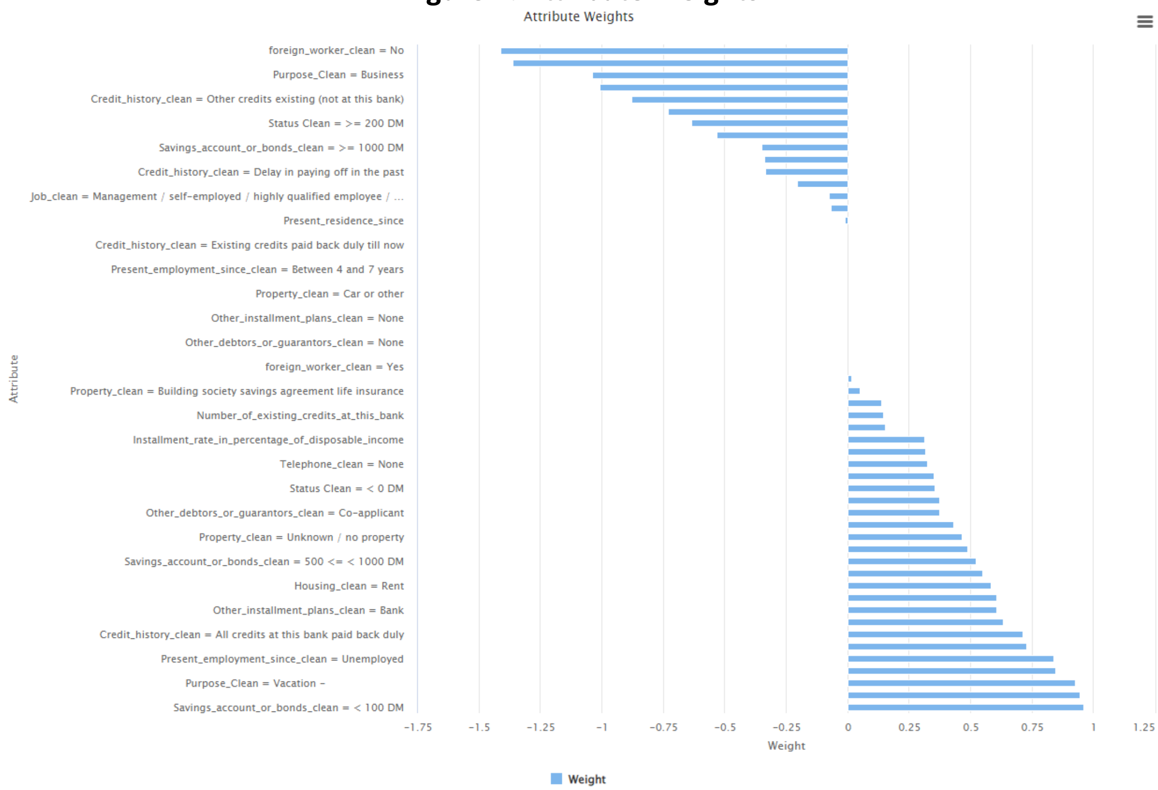


Table 4 shows the Logistic Regression model generated in R. To account for the 13 categorical variables in our dataset, we relied on the nominal to numerical conversion feature in R and RM for conversion of the categorical variables.

Table 4. Logistic Regression Model Results in R

Logistic Regression Confusion Matrix		
Prediction	0	1
0	187	46
1	29	38
Model Statistics		
Accuracy	0.75	
95% CI	(0.697, 0.798)	
No Information Rate	0.72	
P-Value [Acc>NIR]	0.13674	
Kappa	0.3391	
McNemar's Test P-Value	0.06467	
Sensitivity	0.8657	
Specificity	0.4524	

Pos Pred Value	0.8026
Neg Pred Value	0.5672
Prevalence	0.7200
Detection Rate	0.6233
Detection Prevalence	0.7767
Balanced Accuracy	0.6591
‘Positive’ Class	0

Results from this table were produced in R using the same threshold as the default 0.5 in RM.

Neural Network in RM and R

The second model used to predict credit worthiness is the Neural Network Model. The NN model creates a network of nodes to understand the training set given through the node’s activation function. Each connection is evaluated for its relative weight in the model. This model is useful when the inputs and outputs are well understood and the goal is prediction, not understanding of the model.

Due to the nature of the neural network model, nominal to numerical conversion was used on the categorical variables to remove the ordinality of the categories or the model’s possible assignment of non-existent relationships between the categories. If we take the attribute purpose, for example, domestic appliances do not necessarily rank higher than furniture/equipment.

Below is the model generated through RM. A total of 37 variables were fed to the model. There is one hidden layer with 21 nodes and two output nodes. The plot of the model is shown below.

Figure 5. Neural Network Model in Rapid Miner

The same dataset was passed through R. There is one hidden layer with 8 nodes, two bias nodes, and one output node.

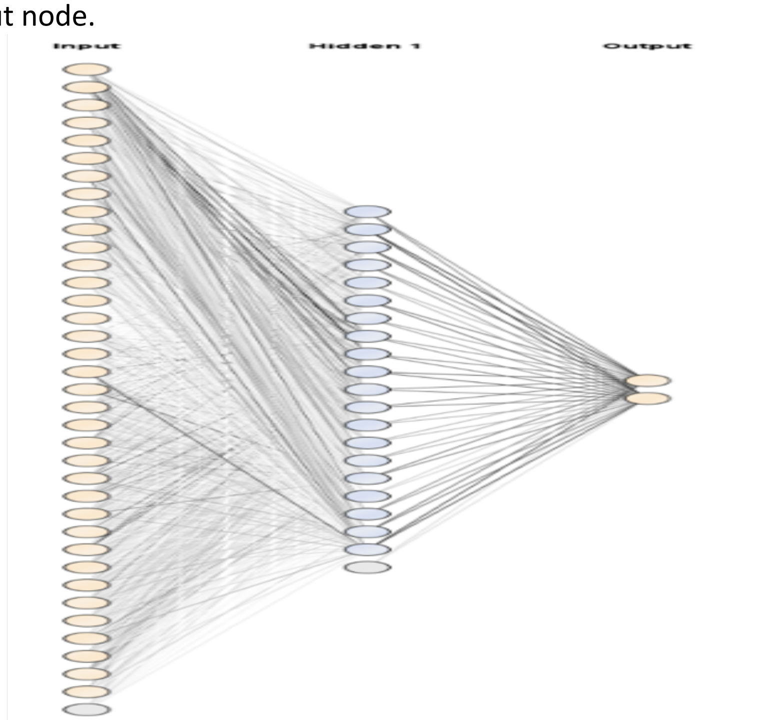


Figure 6. Neural Network Model in R

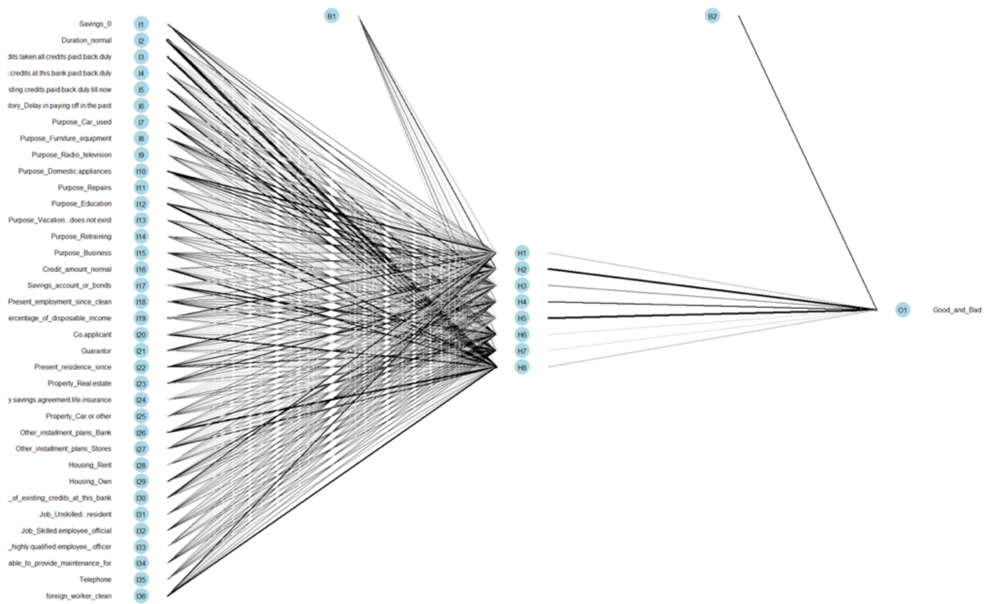


Table 5. Neural Network Model Results in R

Neural Network Confusion Matrix		
Prediction	1	2
1	173	46
2	35	46
Model Statistics		
Accuracy	0.7300	
95% CI	(0.676, 0.7794)	
No Information Rate	0.6933	
P-Value [Acc>NIR]	0.09323	
Kappa	0.3432	
McNemar’s Test P-Value	0.26652	
Sensitivity	0.5000	
Specificity	0.8317	
Pos Pred Value	0.5679	
Neg Pred Value	0.7900	
Prevalence	0.3067	
Detection Rate	0.1533	
Detection Prevalence	0.2700	
Balanced Accuracy	0.6659	
‘Positive’ Class	0	

Results in this table were produced in R using the same threshold as the default 0.5 in RM.

Decision Tree Analysis in RM and R

Decision Tree Models are useful for creating a set of rules by which to classify the data in a tree format to predict the outcome. When designing our Decision Tree model, we selected "gain ratio" to be the selection criterion so that the information gain for each attribute would be adjusted for the model to consider the variations in the attributes. In both R and RM, we applied pruning and pre-pruning by setting a maximal depth of 10, a confidence level of 0.1, a minimal gain of 0.05, a minimal leaf size of 4, a minimal size for split of 5, and the number of pre-pruning alternatives to 3. The confidence level was higher than the 0.05 setting for the other models. We recognized that this confidence level was not ideal as it allowed a wider confidence interval allowing more values to "pass" for a node to split. We were concerned with acquiring accuracy, precision, and recall values that were similar to the other models and chose to sacrifice the confidence level to achieve this.

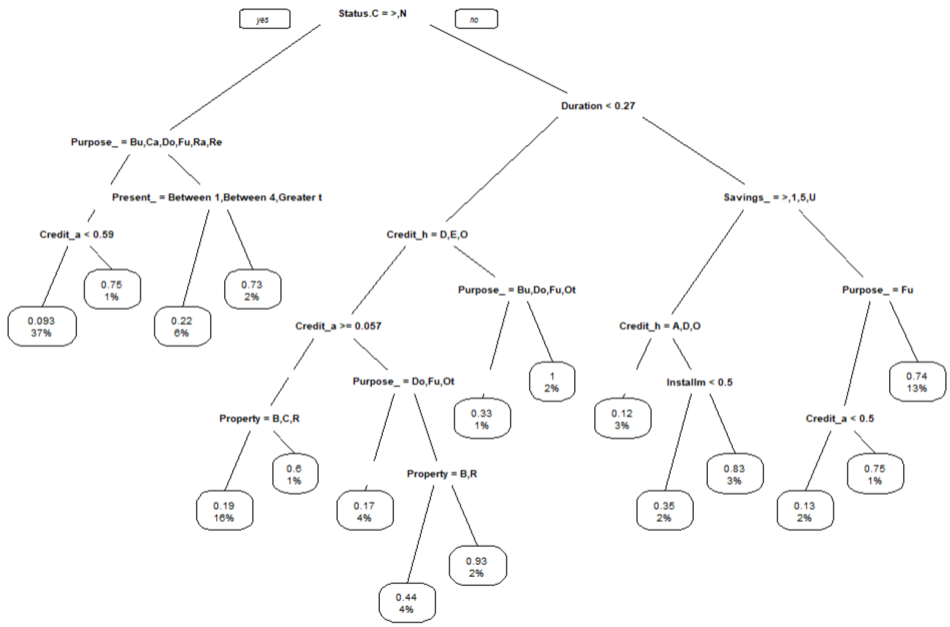
Figure 7. Decision Tree Results in Rapid Miner

Status = 0 <= < 200 DM
| Credit_amount > 0.663: 2 {1=0, 2=12}
| Credit_amount ≤ 0.663: 1 {1=164, 2=93}
Status Clean = < 0 DM
| Credit_amount > 0.010
| | foreign_worker = No: 1 {1=13, 2=2}
| | foreign_worker = Yes: 2 {1=121, 2=133}
| Credit_amount ≤ 0.010: 1 {1=5, 2=0}
Status= >= 200 DM
| Duration > 0.051
| | Number_of_people_being_liable_to_provide_maintenance_for > 0.500: 2 {1=1, 2=4}
| | Number_of_people_being_liable_to_provide_maintenance_for ≤ 0.500: 1 {1=40, 2=10}
| Duration ≤ 0.051: 1 {1=8, 2=0}
Status = No checking account: 1 {1=348, 2=46}

Decision Tree results from this figure were produced in RM.

The decision tree produced in R was built similarly to the decision tree built in RM. A visual of the decision tree is below. For clarity of the nodes, the specific criterion contributing to each attribute was abbreviated.

Figure 8. Decision Tree Results in R



R Decision Tree Key

Purpose		Credit History	Property
CA – car (used) **FU –** furniture/equipment **RA** – radio/television **DO –** domestic appliances	**RE –** repairs **BU –** business **OT –** other	**D –** delay paying in past **E –** existing credits paid back duly **O** – other credits not at this bank **A** – all credits at this bank paid back duly	**B –** building/society savings agreement **R –** real estate **U** – unknown

Table 6. Decision Tree Model Results in R

Decision Tree Confusion Matrix		
Prediction	0	1
0	193	49
1	23	35
Model Statistics		
Accuracy	0.76	
95% CI	(0.7076, 0.8072)	
No Information Rate	0.72	
P-Value [Acc>NIR]	0.067970	
Kappa	0.3426	
McNemar’s Test P-Value	0.003216	
Sensitivity	0.8935	
Specificity	0.4167	
Pos Pred Value	0.7975	
Neg Pred Value	0.6034	
Prevalence	0.7200	

Detection Rate	0.6433
Detection Prevalence	0.8067
Balanced Accuracy	0.6551
'Positive' Class	0

Results in this table were produced using R.

Predictor Rankings (determinants in the DT model)

The attribute that had the most influence on the decision tree model was the status of the customer's checking account. Status was followed by the purpose of the loan, the credit amount of the loan, and the customer's credit history. The attribute importance, from largest to smallest, is listed in table 7. The highest importance is 0.0541 for that attribute "Status" referring to the status of existing checking account. This means that this attribute was the first attribute used to split a note in our decision tree. It follows that our decision tree shown previously in figure 8 has the attribute "Status" as its root note. Furthermore, attributes such as "Credit_history," "Savings_account_or_bonds," and "Purpose" were the next highest deciding factors to split leaf nodes

Table 7. Information Gain in DT Model

Attribute	**attr_importance**
Status	0.0541
Credit_history	0.0246
Savings_account_or_bonds	0.0215
Purpose	0.0212
Property	0.0163
Present_employment_since	0.0146
Housing	0.0081
Job	0.0068
Other_installment_plans	0.0062
foreign_worker	0.0053
Other_debtors_or_guarantors	0
Telephone	0
Duration	0
Credit_amount	0
Installment_rate_in_percentage_of_disposable_income	0
Present_residence_since	0
Number_of_existing_credits_at_this_bank	0
Number_of_people_being_liable_to_provide_maintenance_for	0

Results in this table were produced in R

Naïve Bayes Analysis in RM and R:

The Naïve Bayes method is a simple conditional probability method. A limitation of this model for this dataset is that the model assumes features are independent, which might not be the case given that an individual's sound financial practices are indicative of other sound financial practices and therefore may influence each other.

The Naïve Bayes model showed that approximately 71.5% of the features in our dataset were classified as having "Good" credit (Class 1) while 28.5% were classified as having "Bad" credit (Class 2). The following charts show the distribution of each attribute depending on their classification. The distribution of the conditional probabilities for each attribute and their classifications can be found in Figure A2 and Table A1 in the appendix.

Table 8. Distribution Model for Good and Bad Attributes

Class 1 (Good)	Class 2 (Bad)
0.715	0.285
19 distributions	19 distributions

Results in this table were produced in R.

Table 9. Naïve Bayes Model Results in R

Naïve Bayes Confusion Matrix		
Prediction	0	1
0	184	43
1	32	41
Model Statistics		
Accuracy	0.75	
95% CI	(0.697, 0.798)	
No Information Rate	0.72	
P-Value [Acc>NIR]	0.1367	
Kappa	0.3541	
McNemar's Test P-Value	0.2482	
Sensitivity	0.8519	
Specificity	0.4881	
Pos Pred Value	0.8106	
Neg Pred Value	0.5616	
Prevalence	0.7200	
Detection Rate	0.6133	
Detection Prevalence	0.7567	
Balanced Accuracy	0.6700	
'Positive' Class	0	

Results in this table were produced in R

V. Model Evaluation

Model Benchmark (confusion matrix)

Since this objective is assessing risk, the focal class observed is "bad credit" when comparing accuracy, recall, and precision between all four models built in R and RM.

Table 10: Confusion Matrix for Monitoring "Bad" Credit

	DT_R		DT_RM		LR_R		LR_RM		NN_R		NN_RM		NB_R		NB_RM	
	BAD	GOOD	BAD	GOOD	BAD	GOOD	BAD	GOOD	BAD	GOOD	BAD	GOOD	BAD	GOOD	BAD	GOOD
BAD	35	23	38	38	38	29	43	28	46	35	58	45	41	32	48	33
GOOD	49	193	52	172	46	187	47	182	46	173	32	165	43	184	42	177
Accuracy	76.00%		70.00%		75.00%		75.00%		73.00%		74.33%		75.00%		75.00%	
Recall	41.67%		42.22%		45.24%		47.78%		61.23%		64.44%		48.81%		53.33%	
Precision	60.34%		50.00%		56.72%		60.56%		56.79%		56.31%		56.16%		59.26%	

Simulation Results and Interpretation

The original dataset contained 30% of applicants with the status of having bad credit. Thus, that is a benchmark metric to base assessment and evaluation of models. Based on the outputs above for the Confusion Matrix, the goal of the problem is to predict if a customer is good or bad from a credit-risk standpoint. Also, as mentioned in the problem statement – it is worse to classify a customer as good when they are bad than it is to classify a customer as bad when they are good. Thus, the quality of results is very important for this problem statement.

Given the confusion matrix, we will base our results on the following key metrics:

- **Recall -** What proportion of actual positives ("bad" credit) was identified correctly?
- **Precision -** What proportion of positive identifications ("bad" credit) was actually correct?
- **Overall accuracy -** Overall, how often is the classifier correct?

Assessing the above for each of the models, we can find that the Neural Network performs the best. Looking at the results in Table 10, we can see that the Precision from RM and R is about 56% while recall is 64.44% from RM and 61.23% from R. The overall accuracy of the model is 74.33% from RM and 73% from R, which is lower than Logistic Regression and Naïve Bayes models from both RM and R but split for the DT model (the accuracy from R was higher whereas the accuracy from RM was lower). All these metrics are also comparable to the baseline evaluation metric. However, it is more critical to ensure that the applicants marked as bad are in fact bad (highest recall) which is why we chose the Neural Network to be the best model based on the dataset.

Feature Engineering: as part of model evaluation and benchmarking, we looked at performing feature engineering through the following:

1. Reassessing the role of attributes – removed certain features that were either obscure in their definition and / or had low contribution to the model. We made this assessment based on the pairwise correlation table. We chose the following variables in our model - Age, Sex, Job, Housing, Saving accounts, Checking account, Credit amount, Duration, Purpose, Risk.
2. Combining categorical levels - we combined certain choices in purpose values as well as housing values, since some choices had very data points associated with them.

Based on the above reruns through feature engineering, the initial NN model still performed the best. Even though the model accuracy was higher with feature engineering, precision and recall values deteriorated across these reruns. Thus, based on our initial criteria of model selection focusing on precision and recall being given higher weightage than accuracy, we concluded that the initial NN model is still the best performing overall.

As a comparison, the confusion matrices for all the models with feature engineering are presented in the Appendix.

Lift Charts

Lift charts are displayed in figure A4 in the Appendix. The Logistic Regression model lift chart showed the following results:

% of Population	% correct in Confidence Segment	% cumulative coverage
10%	79%	26%
20%	63%	47%
30%	46%	62%
40%	37%	75%
50%	30%	85%
60%	16%	90%
70%	14%	95%
80%	9%	98%
90%	7%	100%
100%	0%	100%

This shows the largest lift of the model at 20% level of the population, indicating a relatively good model.

The Decision Tree model lift chart showed the following results:

% of Population	% correct in Confidence Segment	% cumulative coverage
10%	63%	21%
20%	53%	39%
30%	54%	57%
40%	27%	66%
50%	41%	80%
60%	19%	86%
70%	10%	89%
80%	11%	93%
90%	16%	98%
100%	6%	100%

This model has lower overall confidence segments than the LR model. The largest lift of this model is at the 30% level of the population, indicating a lower performance than the LR model.

The Naïve Bayes model lift chart showed the following results:

% of Population	% correct in Confidence Segment	% cumulative coverage
10%	70%	23%
20%	63%	44%
30%	50%	61%
40%	39%	74%
50%	31%	84%

60%	17%	90%
70%	9%	93%
80%	7%	95%
90%	11%	99%
100%	3%	100%

This model shows relatively good performance. The highest lift is at the top 20% of the population.

The Neural Network model lift chart showed the following results:

% of Population	% correct in Confidence Segment	% cumulative coverage
10%	77%	26%
20%	50%	42%
30%	47%	58%
40%	17%	63%
50%	30%	73%
60%	20%	80%
70%	23%	89%
80%	13%	92%
90%	13%	97%
100%	10%	100%

This model shows the best overall performance. The highest lift is generated at the top 10% of the population. Based on lift charts, the highest performing models were the Logistic Regression, Neural Network, Naïve Bayes models. The Decision Tree model had the worst performance relative to the others.

VI. Discussion

Implications of the findings for the banking sector.

The assessment of credit risk is of utmost importance in the financial industry. Potential losses from credit risk can be staggering. In the quarter three report of Deutsche Bank for 2023, the provision for credit losses was € 245 million *(Deutsche Bank, 2023)*. Effectively assessing credit risk can substantially minimize these losses and protect the bank's financial health. One tool that banks can use to make better informed lending decisions are data driven models.

The use of data driven models is a promising approach for achieving financial stability and reducing loan defaults in the German banking sector. Cultural and regulatory nuances of the German financial landscape can be incorporated in credit risk assessments and can help financial institutions make informed lending decisions.

Our models, especially the Neural Network model, had the best success in classifying applicants as "Good" or "Bad". Consequently, this and the other models can help identify those who are more likely to default on their loans. By applying our models, banks can improve their loan approval process, reduce their credit defaults, and increase their profits. The efficiency of this type of credit risk management can also provide significant cost savings for banks.

Highlight of the key factors affecting loan default.

In our analysis we found specific attributes that influenced the creditworthiness of loan applicants. The status of a customer's checking account was a dominant factor, then the purpose of the loan, the credit amount, and the customer's credit history. The weight of such attributes can help banks place greater emphasis on them during their banking credit assessment process. For example, a strong checking account status is an indicator of a lower risk for defaulting on a loan.

Suggested strategies for credit risk management.

Banks can consider the following strategies to help them manage credit risk:

- Include <u>cultural and regulatory factors</u> (due to the unique financial culture and regulatory environment in Germany) into their credit risk assessments. This can help them have more accurate predictions and manage risk proactively.
- Create <u>attribute specific assessments</u> by prioritizing the attributes which had the most influence on the models, such as checking account status, loan purpose, and credit amount. This can enhance their loan approval process.
- Offer <u>financial education</u> to customers with higher credit risk to empower them to make informed decisions and reduce their likelihood of experiencing financial difficulties.
- <u>Update and improve</u> the credit risk assessment models to reflect changing conditions in the economy or in customers' behavior.
- <u>Develop risk-based offers</u> and pricing to compensate for the increased risk in the borrowers' assessment.

Limitations/Suggestions of the study.

While the Neural Network Model performed the best, it is not the most transparent of the models. Pairing neural network with another model, such as Naïve Bayes, via model stacking will help harness the strengths of both models and mitigate individual weaknesses of each and mitigate risk in an industry with strict regulatory requirements such as banking and helping to overcome black box nature of NN for regulators.

While our dataset was comprehensive, it might not capture all relevant attributes that influence credit risk. Further research may uncover additional attributes, such as customer behavior or economic indicators, that may be factors which are impactful on credit risk. In addition, the size of our dataset and its timeliness may limit our models' ability to predict trends in credit risk over the long-term. Given the many categorical attributes, some needing dummy variables for specific methods, our sample size may not adequately support numerous attributes. This could affect our model's ability to capture credit risk nuances, so we recommend obtaining a larger dataset for future analysis. As time progresses, new trends and patterns emerge, especially in a rapidly changing industry like finance. More recent data would lead to more relevant, accurate, and complete insight into the modern financial landscape.

Proposed future research.

Future research involving assessment of credit risk can delve into the evolving nature of creditworthiness and consider the impact of economic cycles and customer financial behaviors over time. Also, adding other data sources, such as transaction history, can increase the robustness of predictive data mining models.

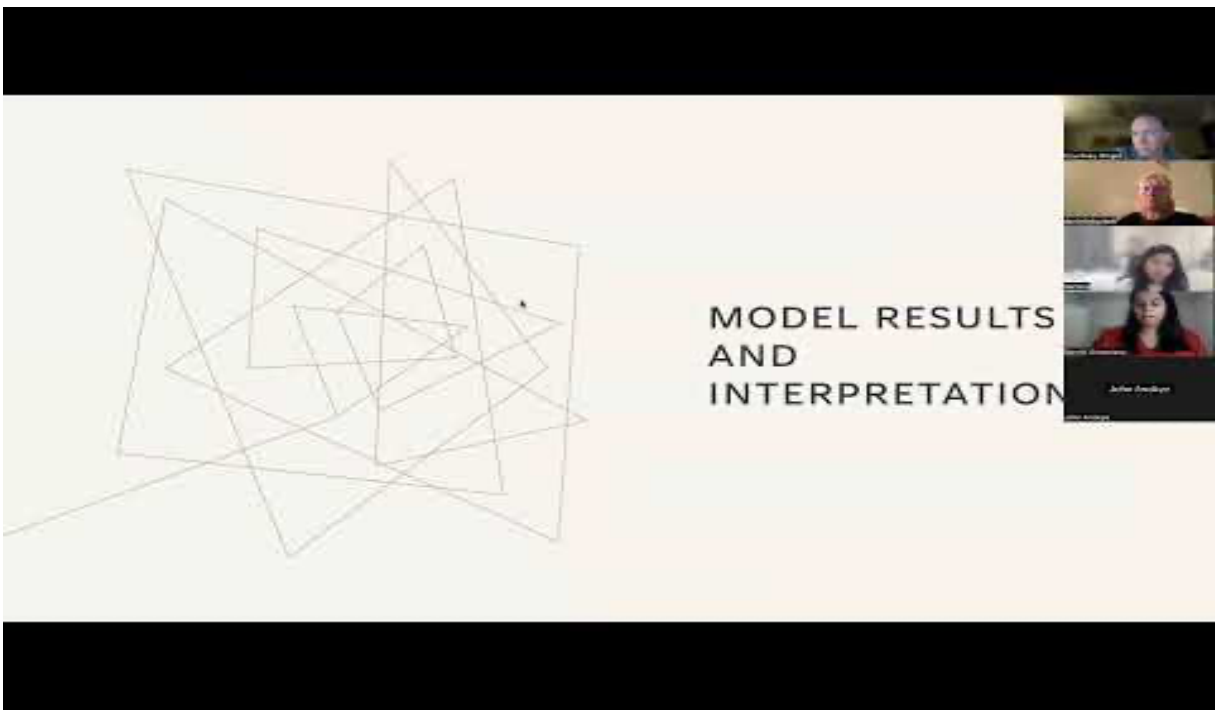
Conclusion

Our project highlights the significance of assessing credit risk within the German banking industry. The impact of cultural and regulatory factors requires a tailored approach to managing credit risk. Our models provide powerful tools for financial institutions and empower them to make more informed lending decisions and reduce their risk of loan defaults.

Through our analysis, we have demonstrated that machine learning models, particularly the Neural Network Model, can provide strong credit risk assessments. By understanding, incorporating, and identifying key attributes that affect creditworthiness, German banks can improve their operational efficiency, make better informed data-driven decision, and reduce their credit risk, ultimately contributing to a healthier financial landscape in Germany.

VII. Link to Video

https://youtu.be/1gu8h_IMABU



REFERENCES

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

Bachmann, J. (2020, December 21). *German credit analysis || A Risk Perspective*. Kaggle. https://www.kaggle.com/code/janiobachmann/german-credit-analysis-a-risk-perspective/notebook

Brown K., Moles P. (2014), *Credit Risk Management*. Edinburgh Business School. ebs.online.hw.ac.uk

Deutsche Bank (2023, October 25). *Deutsche Bank reports nine-month 2023 profit before tax of € 5.0 billion and raises capital outlook*. Retrieved November 4, 2023, from https://www.db.com/news/detail/20231025-deutsche-bank-reports-on-the-third-quarter-results-of-2023?language_id=1

Folkerts-Landau, D., & Schneider, S. (2016, December 15). *Beacon of stability: The foundations of Germany's financial landscape.* Deutsche Bank Research. https://www.dbresearch.com/PROD/RPS_EN-PROD/PROD0000000000441807/Beacon_of_stability:_The_foundations_of_Germany%E2%80%99s_.pdf?undefined&realload=bUDDDnjeFePriGvaVnaxi9TulrTHaD3EdRM/mLpfA2EBFDgffAk0KJ1/pe/4pbrh

Hofmann, Hans. (1994, November 16). *Statlog (German credit data)*. UCI Machine Learning Repository. https://archive.ics.uci.edu/dataset/144/statlog+german+credit+data

Miller M. S. (2022), *Bank Instability Concerns Asset Risk and Leverage Rather Than Size.* Mercatus Center – George Mason University. https://www.mercatus.org/research/public-interest-comments/bank-instability-concerns-asset-risk-and-leverage-rather-size

Push, A. (2022, July 1). *Student debt by country: College costs, student loans around the world*. LendingTree. https://www.lendingtree.com/student/student-debt-by-country/

SCHUFA Holding AG. (n.d.). *Scoring at SCHUFA: explaining the score procedure*. SCHUFA. https://www.schufa.de/schufa-en/scores-data/scoring-at-schufa/

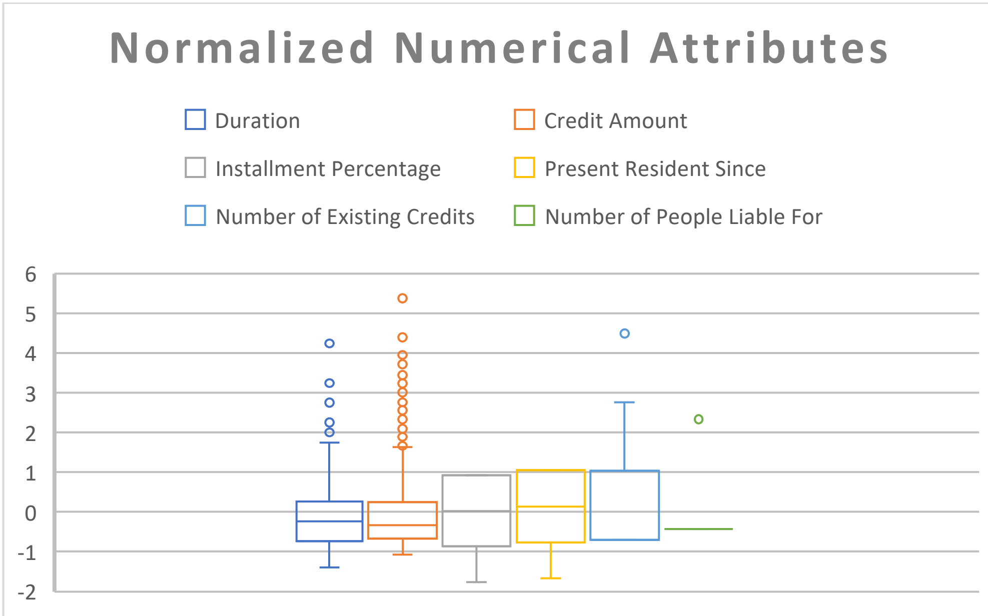
Sawala, G. (2018, December 29). *Germany vs USA: 7 things that are different*. The Honest Blog. https://thehonest.blog/germany-vs-usa-7-things-that-are-different/

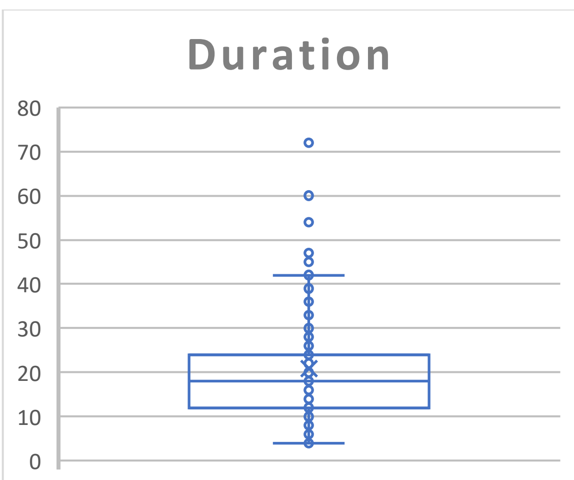
Santander. (n.d.). *Germany: Reaching the consumer*. Reaching the German consumer - Santandertrade.com. https://santandertrade.com/en/portal/analyse-markets/germany/reaching-the-consumers

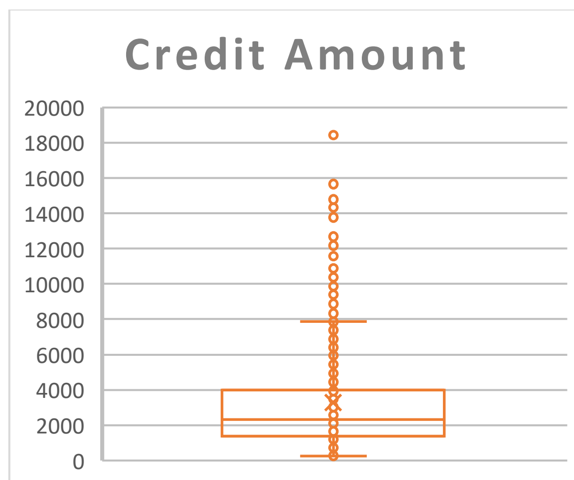
St. Louis FRED. (2023). *Asset Quality Measures, Delinquencies on All Loans and Leases, Secured by Real Estate, Single-Family Residential Mortgages, Booked in Domestic Offices, All Commercial Banks*. St. Louis FED. https://fred.stlouisfed.org/series/DALLSRESFRMACBEP

APPENDICES

Figure A1: Boxplots of Numerical Attributes







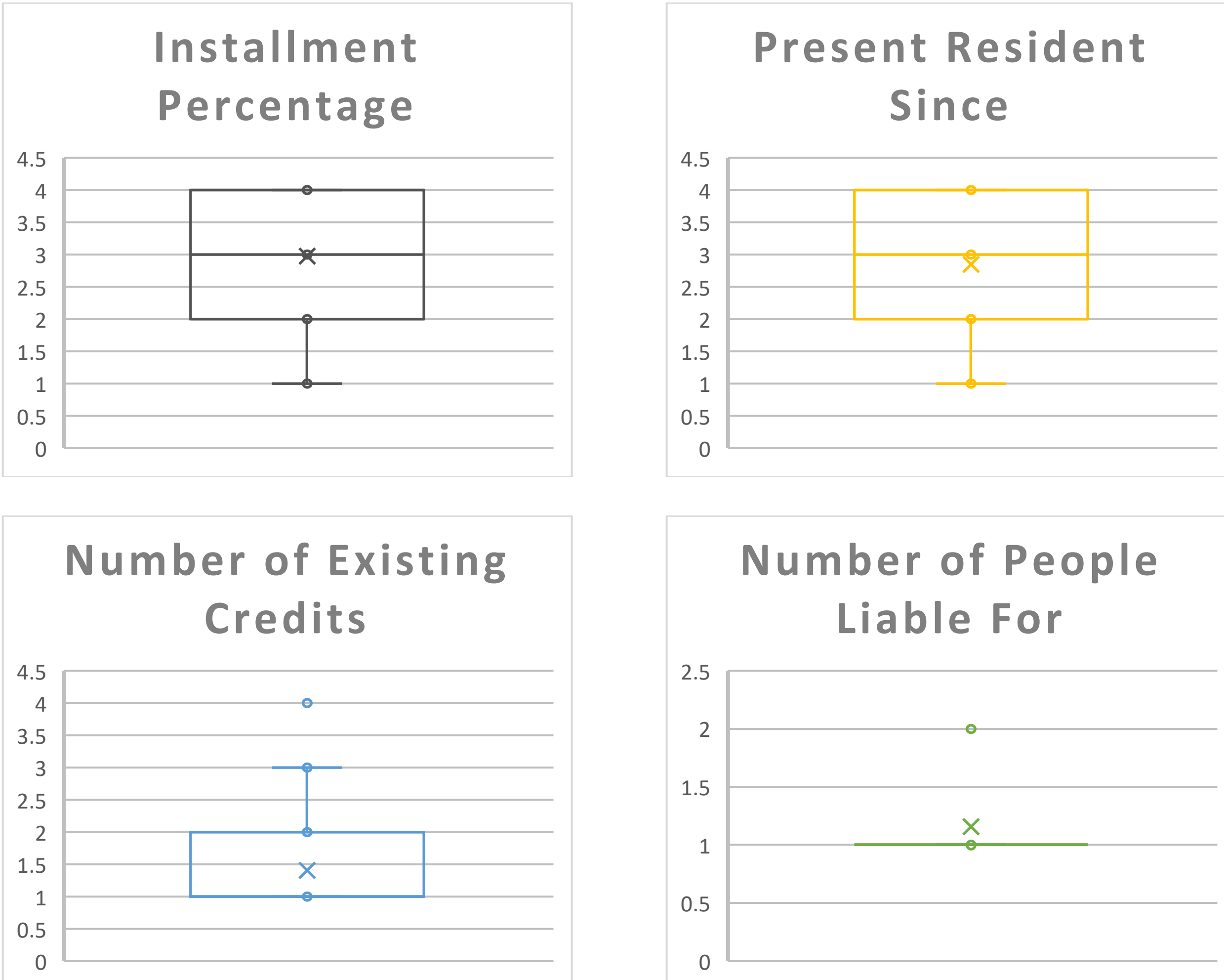


Figure A2: List of codes for categorical attributes combined paired with their respective meanings from the dataset.

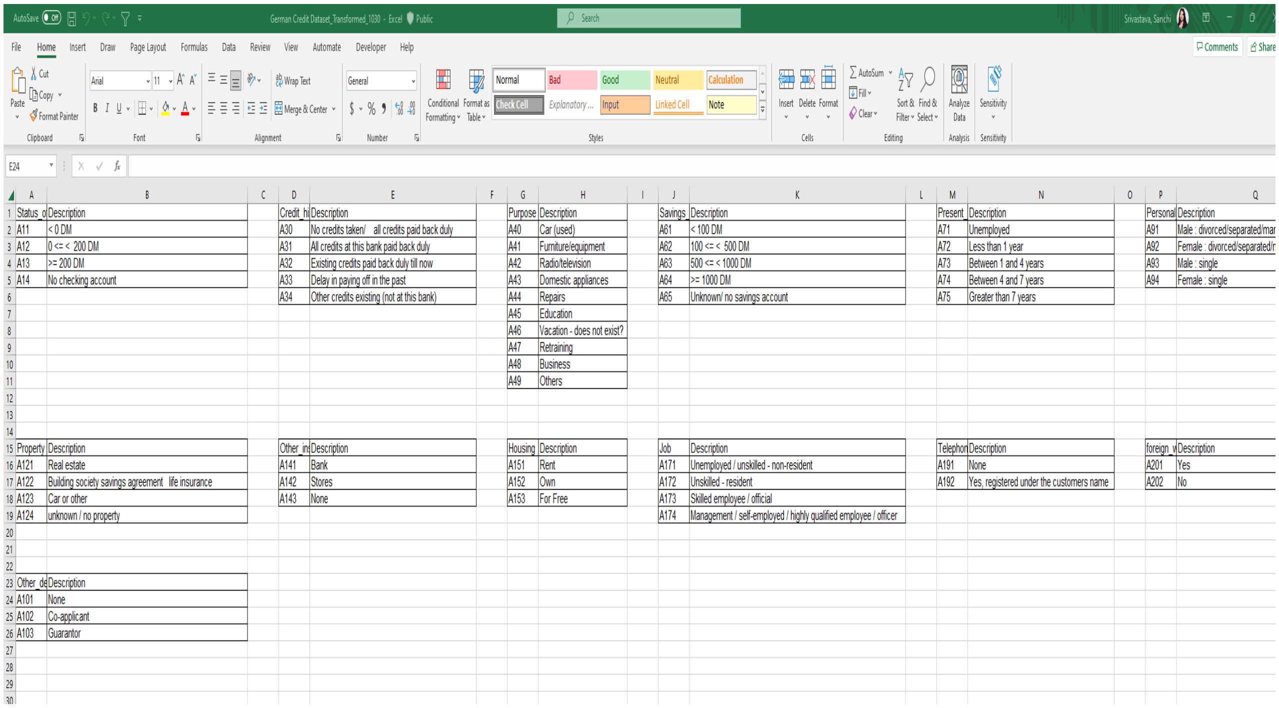


Figure A3: A glance at the dataset.

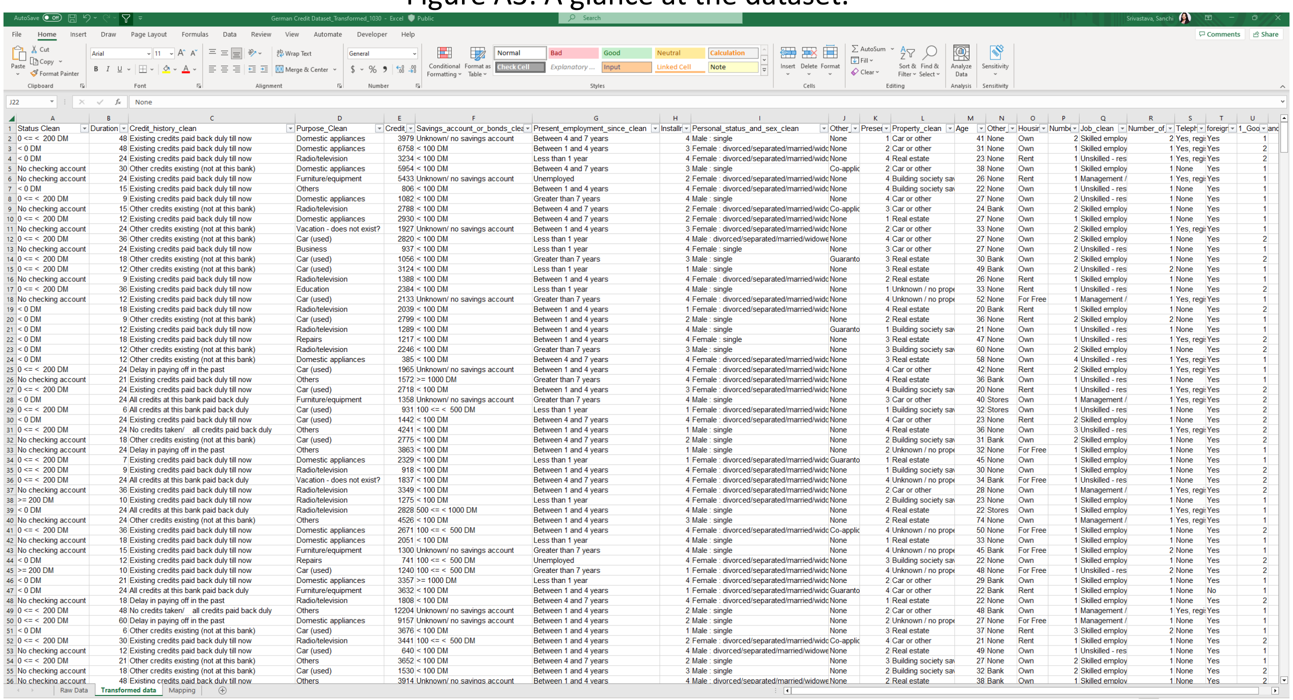


Table A1: Logistic Regression Results in Rapid Miner

Attribute	Coefficient	Std. Coefficient	Std. Error	z-Value	p-Value
Purpose_Clean.Radio/television	0.137	0.137	0.262	0.524	0.601
Purpose_Clean.Furniture/equipment	-0.731	-0.731	0.362	-2.016	0.044

Purpose_Clean.Others	0.154	0.154	0.332	0.464	0.643
Purpose_Clean.Vacation - does not exist?	0.925	0.925	0.402	2.299	0.022
Purpose_Clean.Car (used)	0.848	0.848	0.244	3.480	0.001
Purpose_Clean.Business	-1.037	-1.037	1.191	-0.871	0.384
Purpose_Clean.Education	0.607	0.607	0.559	1.086	0.278
Purpose_Clean.Repairs	0.374	0.374	0.758	0.494	0.622
Credit_history_clean.Other credits existing (not at this bank)	-0.878	-0.878	0.255	-3.436	0.001
Credit_history_clean.Delay in paying off in the past	-0.334	-0.334	0.315	-1.061	0.289
Credit_history_clean.All credits at this bank paid back duly	0.714	0.714	0.381	1.874	0.061
Credit_history_clean.No credits taken/ all credits paid back duly	0.548	0.548	0.427	1.284	0.199
Savings_account_or_bonds_clean.< 100 DM	0.960	0.960	0.260	3.692	0.000
Savings_account_or_bonds_clean.>= 1000 DM	-0.350	-0.350	0.560	-0.625	0.532
Savings_account_or_bonds_clean.100 <= < 500 DM	0.635	0.635	0.346	1.837	0.066
Savings_account_or_bonds_clean.500 <= < 1000 DM	0.521	0.521	0.446	1.170	0.242
Present_employment_since_clean.Between 1 and 4 years	0.729	0.729	0.267	2.732	0.006
Present_employment_since_clean.Less than 1 year	0.944	0.944	0.299	3.153	0.002
Present_employment_since_clean.Unemployed	0.838	0.838	0.439	1.908	0.056
Present_employment_since_clean.Greater than 7 years	0.488	0.488	0.293	1.666	0.096
Status Clean.< 0 DM	0.355	0.355	0.216	1.648	0.099
Status Clean.No checking account	-1.360	-1.360	0.230	-5.916	0.000
Status Clean.>= 200 DM	-0.633	-0.633	0.368	-1.723	0.085
Property_clean.Real estate	-0.203	-0.203	0.234	-0.870	0.384
Property_clean.Building society savings agreement life insurance	0.052	0.052	0.229	0.227	0.820
Property_clean.Unknown / no property	0.465	0.465	0.395	1.177	0.239
Job_clean.Unskilled - resident	-0.068	-0.068	0.227	-0.300	0.764
Job_clean.Management / self-employed / highly qualified employee / officer	-0.076	-0.076	0.282	-0.269	0.788
Job_clean.Unemployed / unskilled - non-resident	-0.532	-0.532	0.643	-0.828	0.408
Other_installment_plans_clean.Bank	0.608	0.608	0.236	2.577	0.010
Other_installment_plans_clean.Stores	0.431	0.431	0.370	1.164	0.244

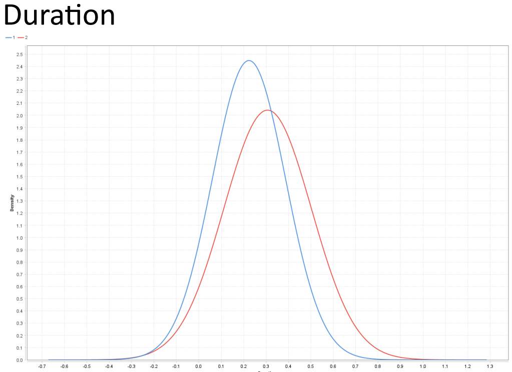
Housing_clean.Rent	0.582	0.582	0.227	2.563	0.010
Housing_clean.For Free	-0.339	-0.339	0.441	-0.769	0.442
Other_debtors_or_guarantors_clean.Co-applicant	0.375	0.375	0.403	0.932	0.351
Other_debtors_or_guarantors_clean.Guarantor	-1.009	-1.009	0.418	-2.413	0.016
Telephone_clean.None	0.325	0.325	0.198	1.644	0.100
foreign_worker_clean.No	-1.412	-1.412	0.613	-2.304	0.021
Duration	1.987	0.352	0.629	3.161	0.002
Credit_amount	2.052	0.319	0.800	2.566	0.010
Installment_rate_in_percentage_of_disposable_income	0.844	0.315	0.260	3.253	0.001
Present_residence_since	-0.031	-0.011	0.255	-0.122	0.903
Number_of_existing_credits_at_this_bank	0.759	0.146	0.560	1.356	0.175
Number_of_people_being_liable_to_provide_maintenance_for	0.046	0.017	0.239	0.191	0.848
Intercept	-3.723	-2.242	0.541	-6.878	0.000

Results from this table were produced by the authors in RM.

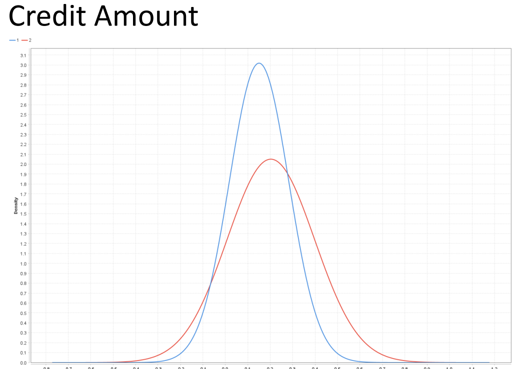
Figure A2: Simple Charts Showing the Distribution of Classification for Every Attribute from the Naïve Bayes Model in RM (blue = good credit, red = bad credit)

Simple Charts

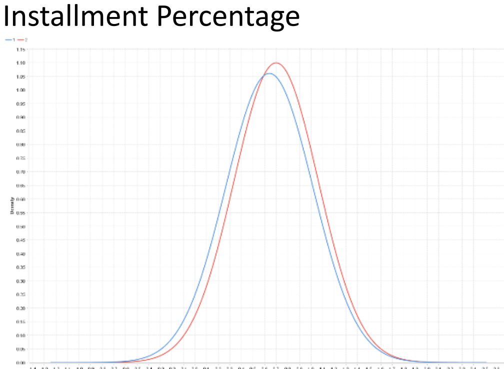
Duration



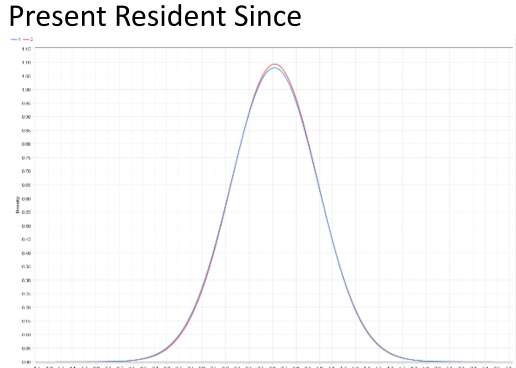
Credit Amount



Installment Percentage

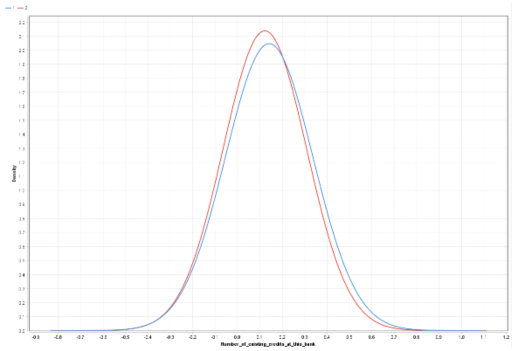


Present Resident Since

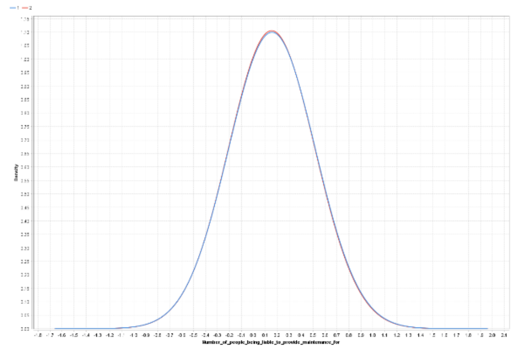


Number of Existing Credit Accounts

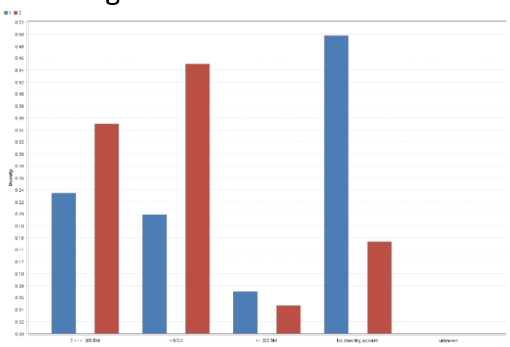
Number of People Liable For



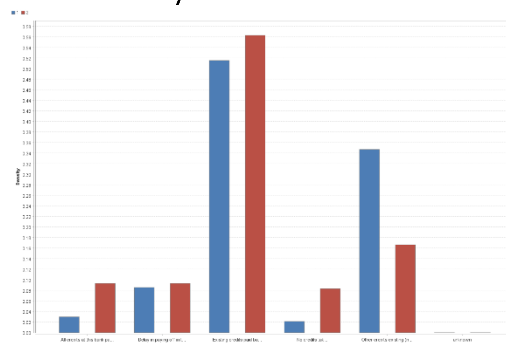
Checking Account Status



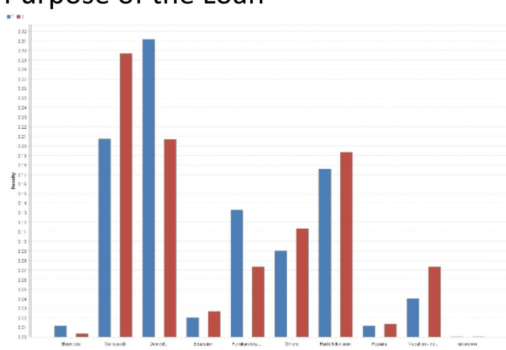
Credit History



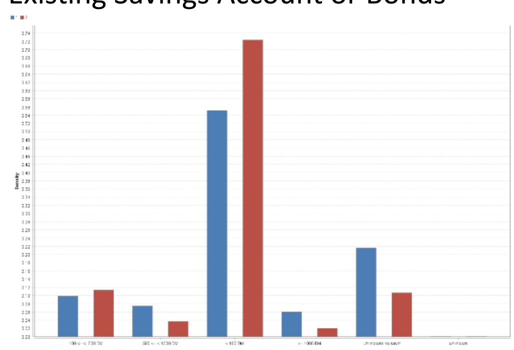
Purpose of the Loan



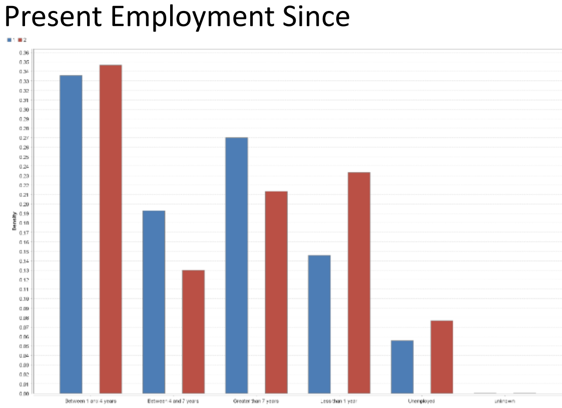
Existing Savings Account or Bonds



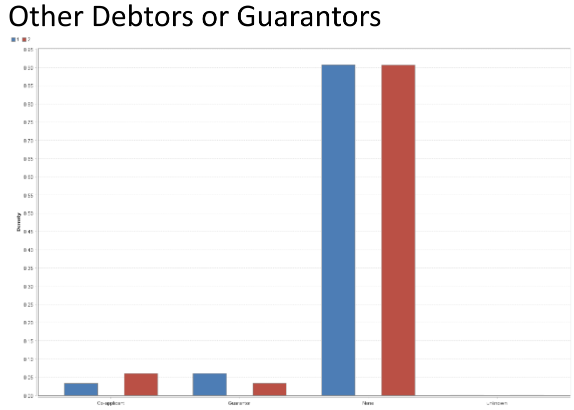
Present Employment Since



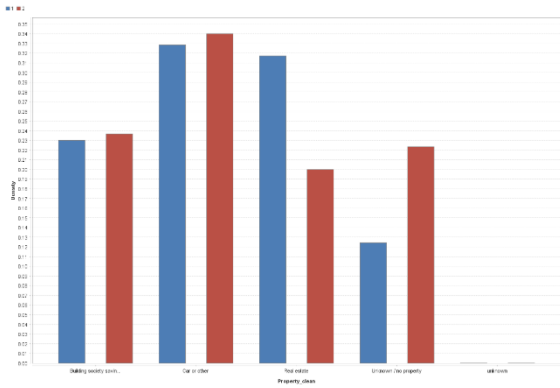
Other Debtors or Guarantors

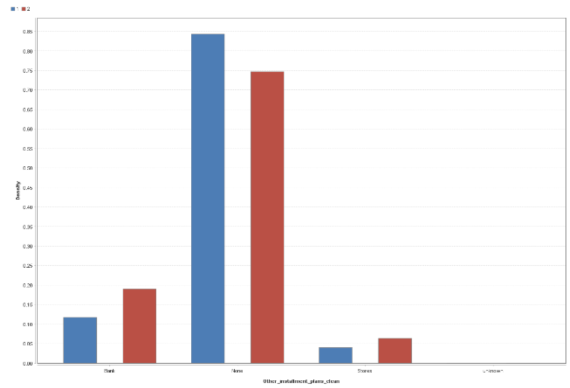


Property

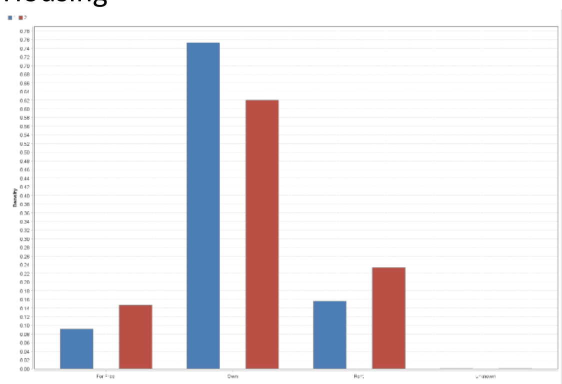


Other Installment Plans

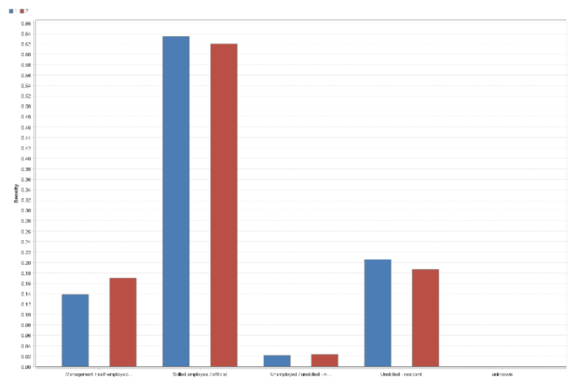




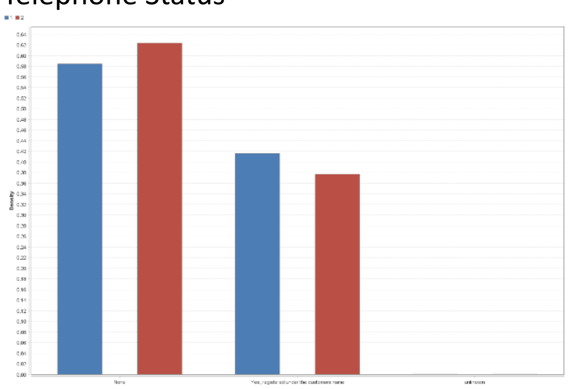
Housing



Job Status



Telephone Status



Foreign Worker

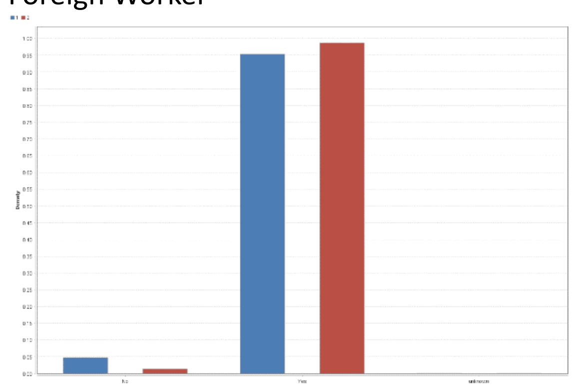


Table A2: Distribution of Classification for Every Attribute from the Naïve Bayes Model in RM

Attribute	Parameter	1 (Good Credit)	2 (Bad Credit)
Duration	mean	0.224	0.307
Duration	standard deviation	0.163	0.195
Credit_amount	mean	0.151	0.203
Credit_amount	standard deviation	0.132	0.195
Installment_rate_in_percentage_of_disposable_income	mean	0.640	0.699
Installment_rate_in_percentage_of_disposable_income	standard deviation	0.376	0.363
Present_residence_since	mean	0.614	0.617

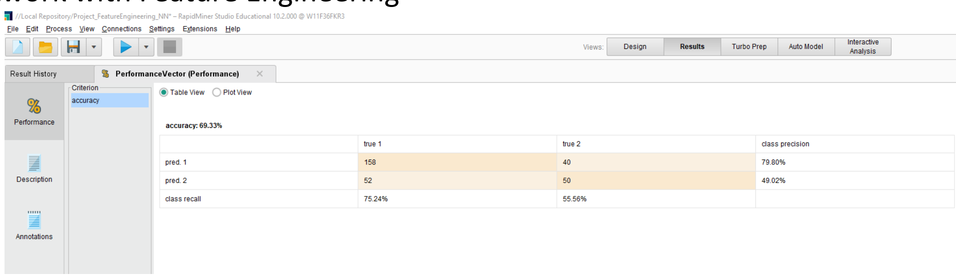
Present_residence_since	standard deviation	0.369	0.365
Number_of_existing_credits_at_this_bank	mean	0.141	0.122
Number_of_existing_credits_at_this_bank	standard deviation	0.195	0.187
Number_of_people_being_liable_to_provide_maintenance_for	mean	0.156	0.153
Number_of_people_being_liable_to_provide_maintenance_for	standard deviation	0.363	0.361
Status Clean	value=0 <= < 200 DM	0.234	0.350
Status Clean	value=< 0 DM	0.199	0.450
Status Clean	value=No checking account	0.497	0.153
Status Clean	value=>= 200 DM	0.070	0.047
Status Clean	value=unknown	0.000	0.000
Credit_history_clean	value=Existing credits paid back duly till now	0.516	0.563
Credit_history_clean	value=Other credits existing (not at this bank)	0.347	0.167
Credit_history_clean	value=Delay in paying off in the past	0.086	0.093
Credit_history_clean	value=All credits at this bank paid back duly	0.030	0.093
Credit_history_clean	value=No credits taken/ all credits paid back duly	0.021	0.083
Credit_history_clean	value=unknown	0.000	0.000
Purpose_Clean	value=Domestic appliances	0.311	0.207
Purpose_Clean	value=Radio/television	0.176	0.193
Purpose_Clean	value=Furniture/equipment	0.133	0.073
Purpose_Clean	value=Others	0.090	0.113
Purpose_Clean	value=Vacation - does not exist?	0.040	0.073
Purpose_Clean	value=Car (used)	0.207	0.297
Purpose_Clean	value=Business	0.011	0.003
Purpose_Clean	value=Education	0.020	0.027
Purpose_Clean	value=Repairs	0.011	0.013
Purpose_Clean	value=unknown	0.000	0.000
Savings_account_or_bonds_clean	value=Unknown/ no savings account	0.216	0.107
Savings_account_or_bonds_clean	value=< 100 DM	0.551	0.723
Savings_account_or_bonds_clean	value=>= 1000 DM	0.060	0.020
Savings_account_or_bonds_clean	value=100 <= < 500 DM	0.099	0.113

Savings_account_or_bonds_clean	value=500 <= < 1000 DM	0.074	0.037
Savings_account_or_bonds_clean	value=unknown	0.000	0.000
Present_employment_since_clean	value=Between 4 and 7 years	0.193	0.130
Present_employment_since_clean	value=Between 1 and 4 years	0.336	0.347
Present_employment_since_clean	value=Less than 1 year	0.146	0.233
Present_employment_since_clean	value=Unemployed	0.056	0.077
Present_employment_since_clean	value=Greater than 7 years	0.270	0.213
Present_employment_since_clean	value=unknown	0.000	0.000
Other_debtors_or_guarantors_clean	value=None	0.907	0.907
Other_debtors_or_guarantors_clean	value=Co-applicant	0.033	0.060
Other_debtors_or_guarantors_clean	value=Guarantor	0.060	0.033
Other_debtors_or_guarantors_clean	value=unknown	0.000	0.000
Property_clean	value=Car or other	0.329	0.340
Property_clean	value=Real estate	0.317	0.200
Property_clean	value=Building society savings agreement life insurance	0.230	0.237
Property_clean	value=Unknown / no property	0.124	0.223
Property_clean	value=unknown	0.000	0.000
Other_installment_plans_clean	value=None	0.843	0.747
Other_installment_plans_clean	value=Bank	0.117	0.190
Other_installment_plans_clean	value=Stores	0.040	0.063
Other_installment_plans_clean	value=unknown	0.000	0.000
Housing_clean	value=Own	0.753	0.620
Housing_clean	value=Rent	0.156	0.233
Housing_clean	value=For Free	0.091	0.147
Housing_clean	value=unknown	0.000	0.000
Job_clean	value=Skilled employee / official	0.634	0.620
Job_clean	value=Unskilled - resident	0.206	0.187
Job_clean	value=Management / self-employed / highly	0.139	0.170

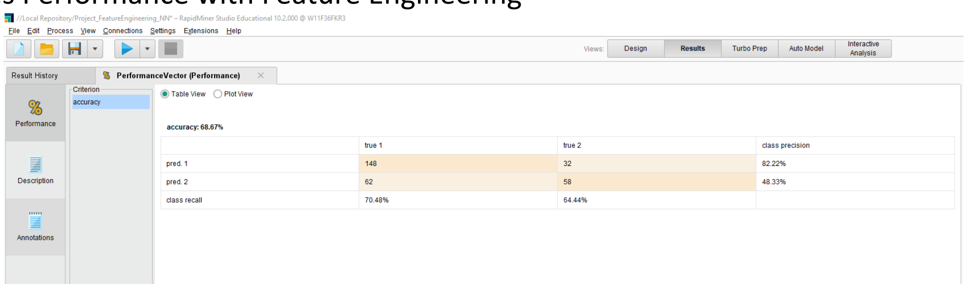
	qualified employee / officer		
Job_clean	value=Unemployed / unskilled - non-resident	0.021	0.023
Job_clean	value=unknown	0.000	0.000
Telephone_clean	value=Yes, registered under the customers name	0.416	0.377
Telephone_clean	value=None	0.584	0.623
Telephone_clean	value=unknown	0.000	0.000
foreign_worker_clean	value=Yes	0.953	0.987
foreign_worker_clean	value=No	0.047	0.013
foreign_worker_clean	value=unknown	0.000	0.000

Figure A3: Confusion Matrices for the Models with Feature Engineering

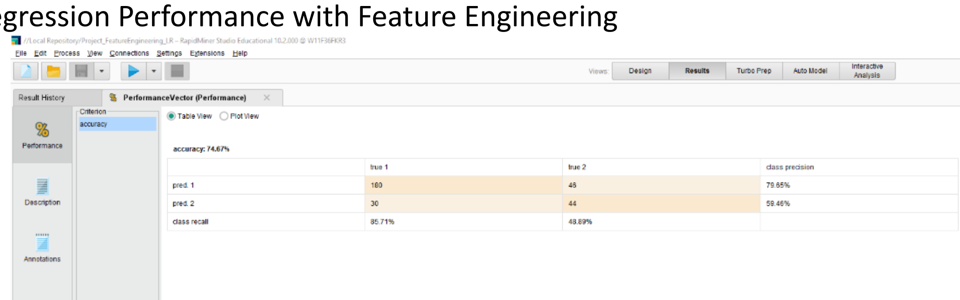
Neural Network with Feature Engineering



Naive Bayes Performance with Feature Engineering



Logistic Regression Performance with Feature Engineering



Decision Tree Performance with Feature Engineering

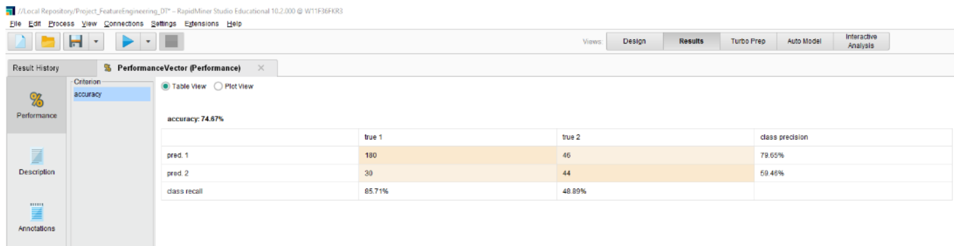
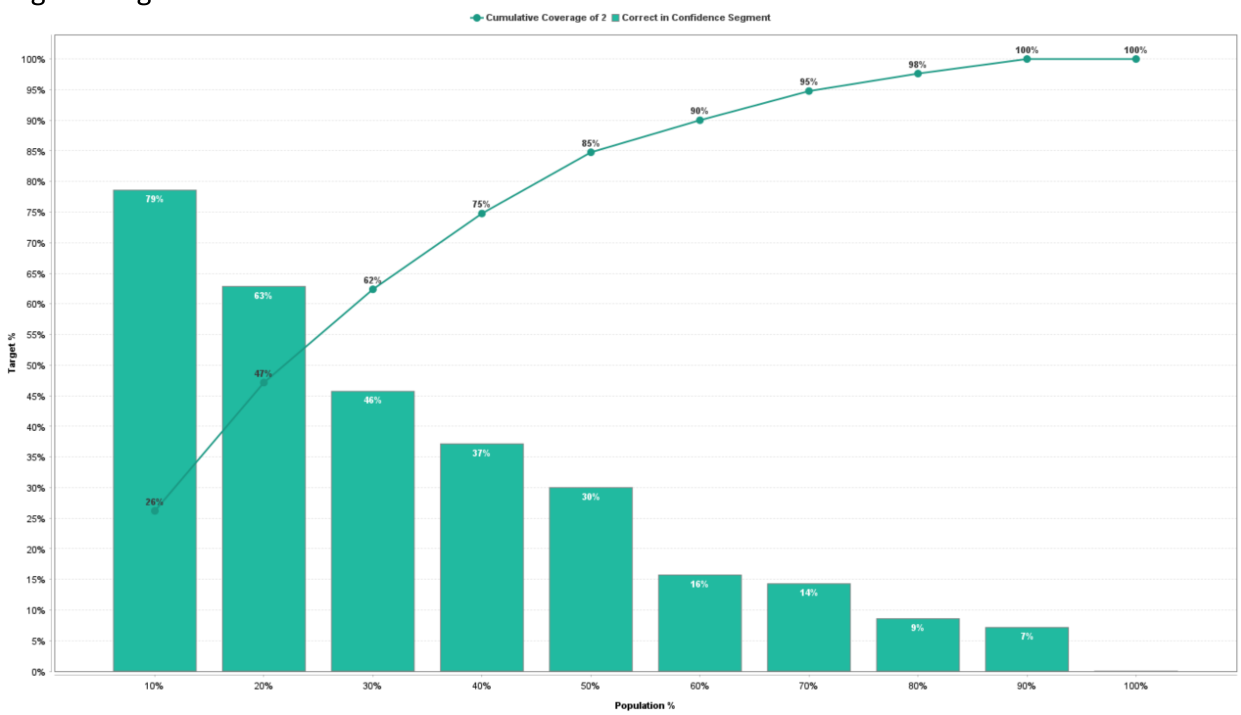
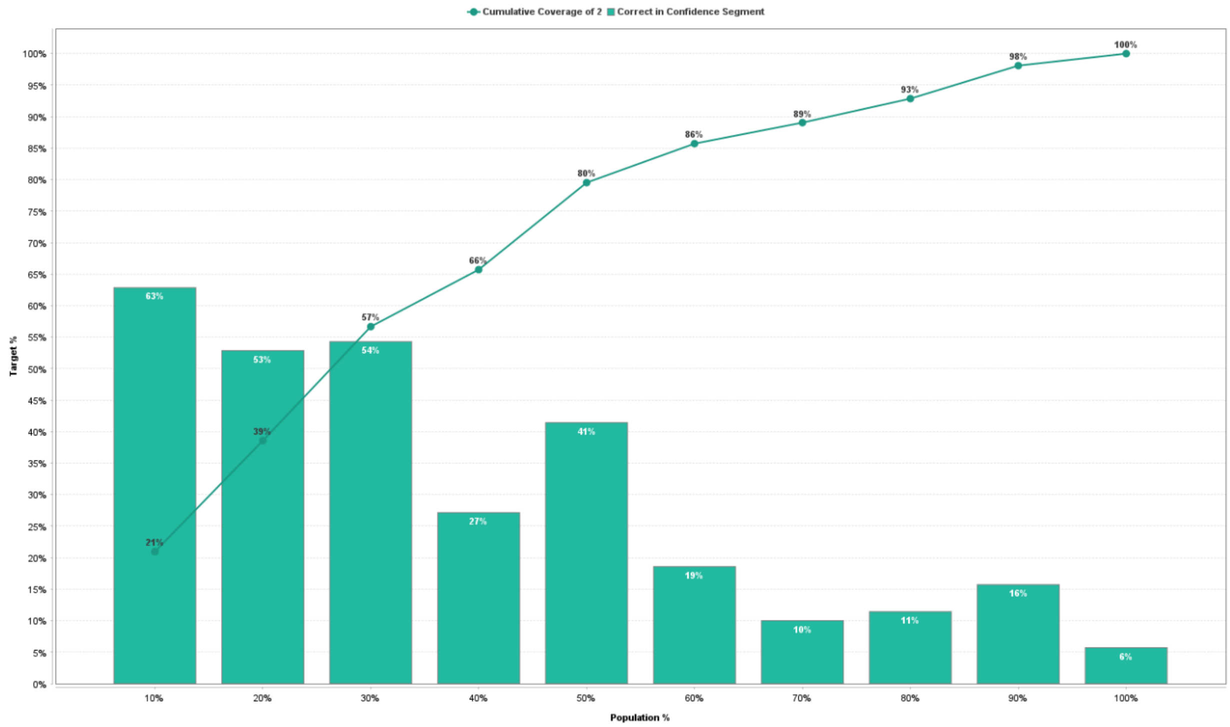


Figure A4: Lift Chart for Logistic Regression and Decision Tree Models in RM

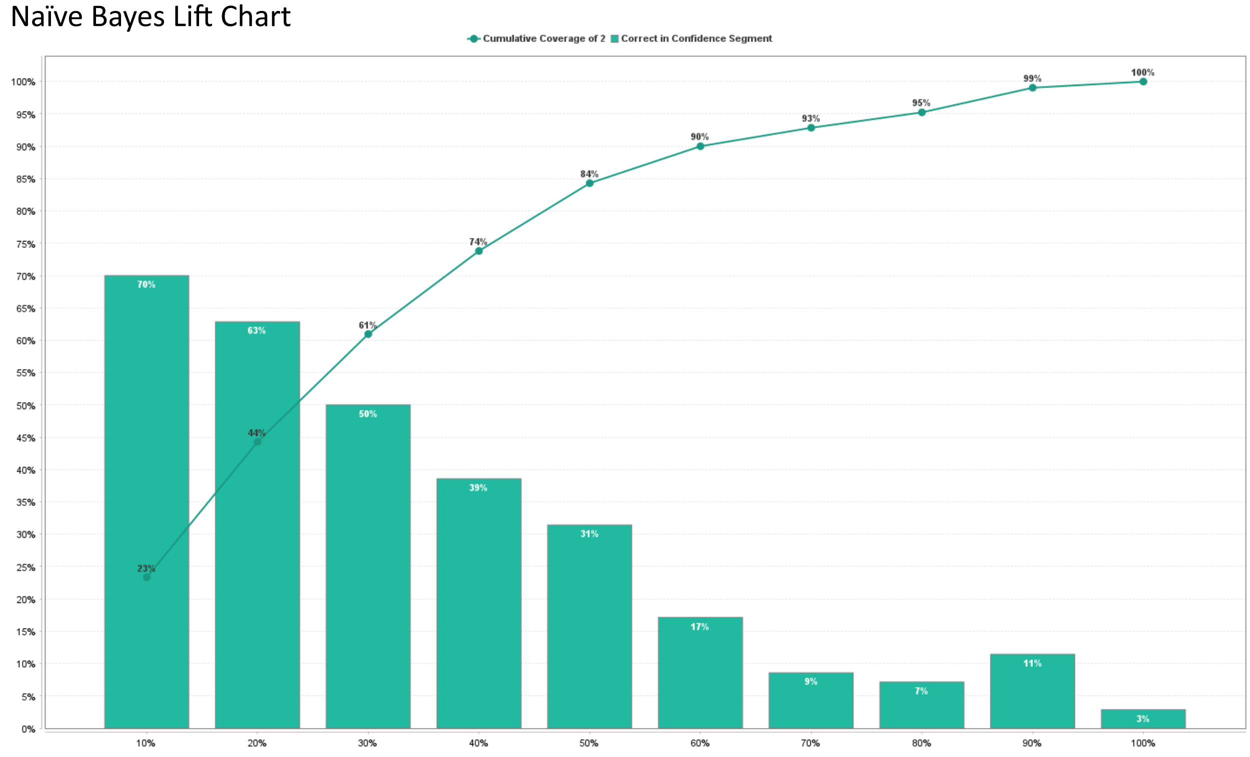
Logistic Regression Lift Chart



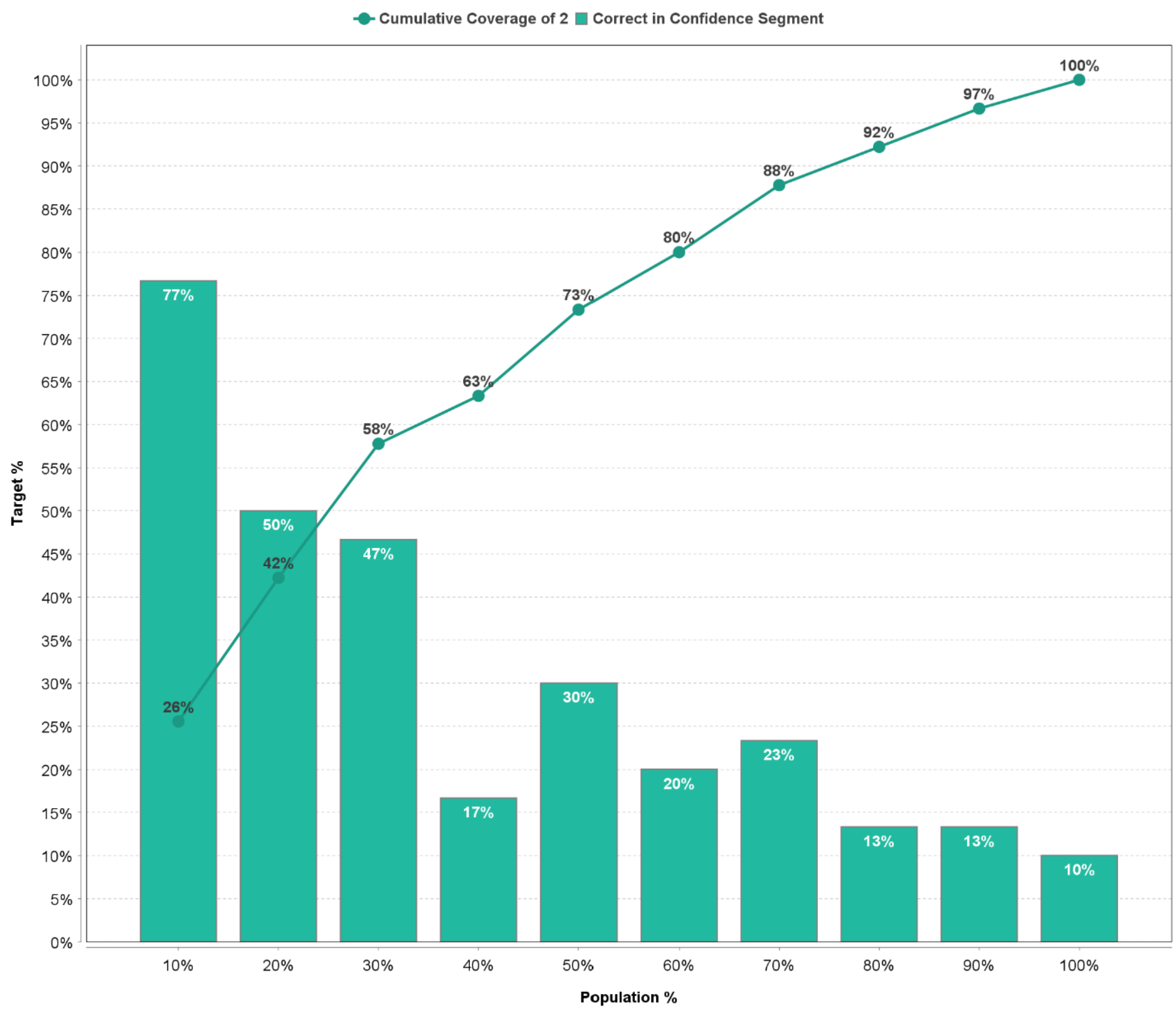
Decision Tree Lift Chart



Naïve Bayes Lift Chart



Neural Network Lift Chart



Cumulative Coverage of 2
Correct in Confidence Segment
Target %
Population %
100%
95%
90%
85%
80%
75%
70%
65%
60%
55%
50%
45%
40%
35%
30%
25%
20%
15%
10%
5%
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
77%
50%
47%
17%
30%
20%
23%
13%
13%
10%
26%
42%
58%
63%
73%
80%
88%
92%
97%
100%