UNIVERSITY OF SUNDERLAND
ASSIGNMENT COVERSHEET

Student ID: 220075907	**Student Name/ Names of all group members:**
Programme: Computer System Engineering	**Module Code and Name: CET313 Artificial Intelligence**
Module Leader/ Module Tutor:	**Due Date:13th Oct 2023 Hand in Date:**
Assessment Title: Intelligent Prototype Development	
Learning Outcomes Assessed: (number *as appropriate*)	

	Mark
Areas for Commendation	
Areas for Improvement	
General Comments	

Assessor Signature:	**Overall mark (subject** to ratification by the assessment board**)**	**Moderator Signature**

………………………………………………………………………………………………

•

I confirm that in submitting this assignment that I have read, understood and adhered to the University's Rules and procedures governing infringements of Assessment Regulations.

PRINT Student Name: ____ Faculty Stamp *(date/time)*

Student Signature : ____

Module Code and Name: ____

Name of Module Tutor : ____

Table of Contents

Overview of E-Portfolio

Building upon the tutorial video presented in class proved to be a delightful venture. Our group's collaborative spirit enables us to convene and work on Python applications at our convenience, owing to the wealth of exercises and study resources at our disposal. Engaging in discussions and group study sessions enriched our knowledge of artificial intelligence, spanning diverse machine learning types, clustering algorithms, and search algorithms. Personally, I found great pleasure in utilizing Jupyter Notebook for our application development. Its introduction in a real-world context emphasized its adaptability and efficiency in code creation. Jupyter Notebook, being an open-source web application, facilitates the creation and sharing of live code documents. I have seamlessly integrated the Jupyter notebook code task into my e-portfolio, accessible through the provided link.

E-portfolio Link: https://canvas.sunderland.ac.uk/eportfolios/13610?verifier=VjrrPoWInFFhGeCkZC1km3R70IhB7yTZTk52zrKN

Introduction

Financial Risk Analysis

Financial institutions play a crucial role in the development of any country, and access to credit facilities is a key driver of progress in developing nations. Banks and other financial institutions are at the forefront of providing these credit options, and the credit business has seen significant growth in recent times. With numerous financial institutions offering services to customers, the decision to grant loans is a pivotal aspect of their operations, demanding careful consideration. Effectively managing the process of issuing loans is vital for a financial institution's stability and profitability, especially given the rapidly evolving business environment, heightened competition in lending strategies, evolving marketing approaches, and changing borrowing patterns of clients. The ability to accurately assess potential defaulters and non-defaulters is critical to ensure that the institution is not adversely affected by incorrect predictions. In this context, credit scoring, leveraging statistics and operations research, emerges as one of the most effective tools for evaluating credit risk. Machine Learning-based Credit Scoring aims to categorize customers as either reliable payers or risky ones. Reliable payers are those who are expected to repay their loans, while risky payers are those who may fail to meet their repayment obligations within the specified timeframe.

Credit scoring models, employed by credit bureaus, are instrumental in determining a customer's creditworthiness. These models generate credit scores by analyzing various statistical characteristics derived from an individual's credit payment history. Factors such as payment history, payment frequency, outstanding debt, credit charge-offs, and the number of credit cards held contribute to the calculation of a customer's credit score. Each of these factors is assigned a weight within the model's formula, ultimately resulting in a credit score ranging from 300 to 850, with 300 representing the lowest and 850 the highest score. Financial institutions use these credit scores to assess the risk associated with granting a loan, as well as to determine the loan terms and interest rates. Higher credit scores lead to more favorable loan terms for customers.

The primary objective of this project is to gain valuable insights into loan delinquency and creditworthiness among individual borrowers, as well as to understand bank lending practices. The ultimate aim is to reduce the number of nonperforming loans in commercial banks. To achieve this, the research endeavors to develop a comprehensive model for financial institutions, encompassing a wide array of factors with varying degrees of importance. The proposed credit scoring algorithms are intended to streamline the assessment of an individual's creditworthiness, effectively distinguishing between excellent and high-risk loan applications. Credit scoring models play a pivotal role in evaluating the risk posed by borrowers, as they calculate a credit score based on information extracted from loan applications, socio-demographic characteristics, and credit bureau reports.

Objectives

Develop a credit scoring model designed for individuals to assess their own creditworthiness.

Evaluate and compare the validity of the proposed credit scoring model against existing statistical credit scoring models in the field.

Develop a user-friendly credit assessment tool for individuals to gauge their credit risk and financial credibility.

Enhance financial literacy by providing individuals with a practical means to understand their credit standing through the utilization of a custom credit scoring framework.

Section 1

Prototype Identification and Planning

Section 1.1 Literature Review on Prototype Identification

There have been numerous studies on financial risk management techniques. In order to convey a feel of the work which has been carried out in this field and, if any, to point out research gaps, an effort has been made to offer an overview of a few chosen works in this chapter. These studies have been organized chronologically so that a sound plan of action for the present endeavor may be developed. According to Thomas, Edelman, and Crook (2002), "Credit Scoring" is. Credit scoring involves a set of decision models and processes that assist lenders in approving consumer credit. These models determine who qualifies for credit, the approved credit amount, and strategies to optimize lender profitability. In summary, credit scoring helps lenders assess risk, set credit limits, and improve borrower profitability (ThomaCrook, 2002).

While the concept of credit has a history dating back 5000 years, credit ratings as we understand them today are a relatively recent development, emerging only in the last 50 years. Credit scoring became a practical way to categorize different groups when investors were faced with the challenge of assessing limited characteristics of an individual rather than all of them. Fisher, as noted by Robert and Halbert, was among the pioneers who tackled the issue of differentiating between individuals based on creditworthiness. It was also demonstrated that Durand's single method could effectively categorize creditors into good and weak categories. Credit scoring is often considered the precursor to data mining, thanks to its innovative use of customer behavior data (Halbert, 2010). In the credit approval process, the first crucial decision an investor must make is whether to extend credit to an existing debtor. The second key decision is how to grant credit to a new borrower. According to Crook, lenders in developed nations assess an applicant's creditworthiness by considering factors such as the borrower's income, experience, and credit history obtained from previous creditors.

When it comes to algorithms used to classify applicants based on their credit scores, they take into account various factors, including age, income, marital status, and payment history, as noted (ThomaCrook, 2002). Some banks categorize potential customers into binary groups, labeling them as "good" or "bad," while others introduce a third category, "refused," for applications that didn't meet initial criteria. Applications that are initially rejected may undergo further review by financial officers. Credit rating models, as highlighted by chon and hang can be approached quantitatively or qualitatively. Qualitative methods may suffer from bias and subjectivity, lacking a solid foundation for determining a candidate's creditworthiness. In contrast, quantitative approaches offer a systematic means to classify loans as performing or non-performing, proving to be a more reliable and precise model compared to emotional or judgment-based methods (Chen, 2010).

When loans go into default, both the lenders and borrowers can face financial consequences. For borrowers, this means they may not receive the loan amount or the interest payments they expected. Additionally, their credit score can take a hit as their name is added to a list of defaulters. This can have a cascading effect, making it difficult for them to secure further loans from the same lender and limiting their ability to use the borrowed funds for investments. To develop credit scoring models, financial institutions often analyze the credit histories of individuals who have recently been approved for loans. If these models do not account for failed loan applications, it can lead to less accurate and potentially biased credit assessments (Schreiner, 2018). .

Section 1.2 Reflection on the Prototype Identification

In the modern credit industry driven by data, major businesses rely heavily on customer loan data to create models aimed at reducing risks and increasing profits. Traditionally, consumers haven't had much of an edge in the market despite sharing their financial growth information.

The predominant method for assessing creditworthiness involves using evaluations generated through statistical models (Hussein A. Abdou, 2011). These algorithms assign scores to existing credit accounts. This approach is widely recognized and consistent across various systems, eliminating the need for businesses to invest heavily in creating new models since they can simply reuse existing code.

As customer data is analyzed and refined, new attributes are introduced. Subsequently, the significance of each attribute is determined, and data is transformed using the Weight of Evidence technique for each feature. This process is employed to gauge the predictive strength of each factor based on the information provided by the customer. It involves the use of Weight of Evidence and Information Value Formulas (Stackoverflow, 2020).

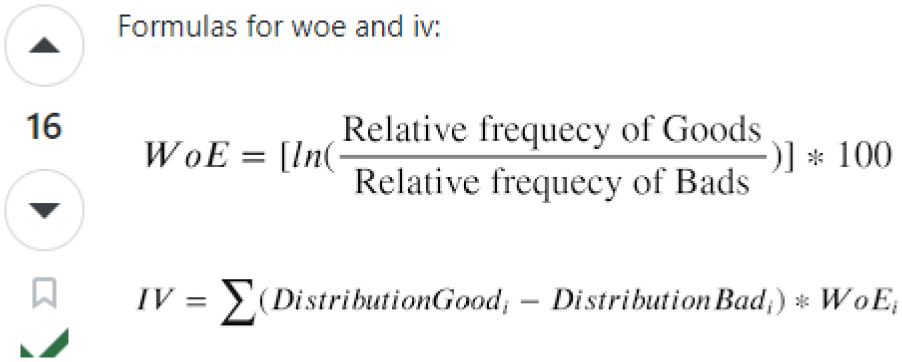


Figure.1

Determine the percentage of positive and negative events, compute the weight of evidence of Proof using the natural log, then substitute the Weight of Evidence values with the raw data in order to change the data factors that have been divided into bins.

Code to achieve this:

```
import numpy as np
import pandas as pd
np.random.seed(100)


df = pd.DataFrame({'grade': np.random.choice(list('ABCD'),size=(20)),
                   'pass': np.random.choice([0,1],size=(20))
})

feature,target = 'grade','pass'
df_woe_iv = (pd.crosstab(df[feature],df[target],
                      normalize='columns')
             .assign(woe=lambda dfx: np.log(dfx[1] / dfx[0]))
             .assign(iv=lambda dfx: np.sum(dfx['woe']*
                                           (dfx[1]-dfx[0]))))

df_woe_iv
```

In the process of WOE transformation, the focus is on establishing linear relationships, addressing anomalies, handling missing data, and preparing values that align with the model's requirements. Subsequently, attributes are selected based on the data values. The modified Weight of Evidence (WOE) from the initial dataset is then integrated into the scoring card, and a Logistic Regression model is applied. All of these steps collectively contribute to the calculation of a scorecard point for each attribute. These individual attribute scores are amalgamated to create the final scorecard. To determine each attribute's unique score, an algorithm is applied, taking into account the specific characteristics and their associated WOE values. This comprehensive approach results in the generation of the ultimate scorecard.

Scoring systems, while effective, have a limitation in that they require a substantial amount of data to achieve high accuracy. This means that a considerable amount of customer information needs to be incorporated into scorecards, and this may not provide a comprehensive understanding of an individual client. Rather than relying solely on diverse sets of data to assess different factors, an alternative approach involves utilizing a dataset that includes transactions involving both good and bad risks. This data can be used to instruct the model, and the same Logistic Regression model for binary categorization can be applied to identify patterns and determine whether a customer represents a good or bad risk. This approach requires fewer data points and leads to the development of a more personalized model.

Section 2 Development

Section 2.1 Developed code and planning documents for prototype

Requirement Analysis

Fig. Use Case Diagram

Scenarios of System Use Case

Use Case: Evaluate Loan Risk

This use case scenario highlights a critical system function. It outlines the specific use case where the system is required to determine whether a customer poses a high or low risk when it comes to loans.

Features Data

Name and Identification	**Determine the loan risk using UC 1**
Goal Context:	The client is given a rating indicating how safe or risky they are as a borrower.

Client:	Customer
Idea:	In this use case, it is assumed that the page will update correctly to show the client's latest rating.

Fundamental Acheivements Case

Methods	**Actor**	**Summary of the Procedures**
1	System	Completed loading of the calculate loan risk page.
2	Clients	User types up the necessary fields to calculate risk: Age Government issued id number Social Security Number Job Housing Deposits Fund Amount Checking Account Credit Account Time Period Purpose Contact Number
3	Customer	Client selects the compute buttons.
4	System	The framework for machine learning is given various data fields and the model determines the rating.

5	Systems	The evaluation provided by the machine learning model is displayed after the page is refreshed.

Apply for loans in the use case scenario.

The example use case below represents the loan application use case, in which the client requests a loan from the loan provider.

Features Data

Name and Identification	**Determine the loan risk using UC 3**
Goal Context:	The client submits an application for one of the specified loan packages.
Client:	Customer
Idea:	The expected result is that the page will reload, displaying an acknowledgment message, and removing the inputs once the page loads.

Fundamental Acheivements Case

Procedures	**Actor**	**Summary of the procedure**
1	System	Obtain a financial aid by completing the fields that are required.
2	Customer	Customers complete the following sections when applying for loans like age and objectives.

3	Client	The client enters their request by pressing the "Submit Key".
4	System	The inputs disappear as the website reloads.
5	System	The computer shows an affirmation notification.

Extended Case

Steps	**Events**	**Action Steps**
2a)	The request is accepted by the financial service provider.	The financing source selects "approve" from the menu. - When there is a favorable answer, the computer informs the client.
2b)	The financial source turns down the request.	The provider selects "decline" from the menu. An adverse reaction is communicated to the client via the software.

Response to the request for loan using the case scenario

The Response Utilization use case is expanded upon in the use case model. The following illustrates the way a finance source would react to a loan application that a user has filed.

Features Data

Name and Identica6ion	**Determine the loan risk using UC 4**
Goal Context:	A borrower submits a financing application, and the loan company responses.

Client:	Customer
Idea:	When a loan application is submitted, the financier answers, and the borrower is informed when the loan request has been reviewed.

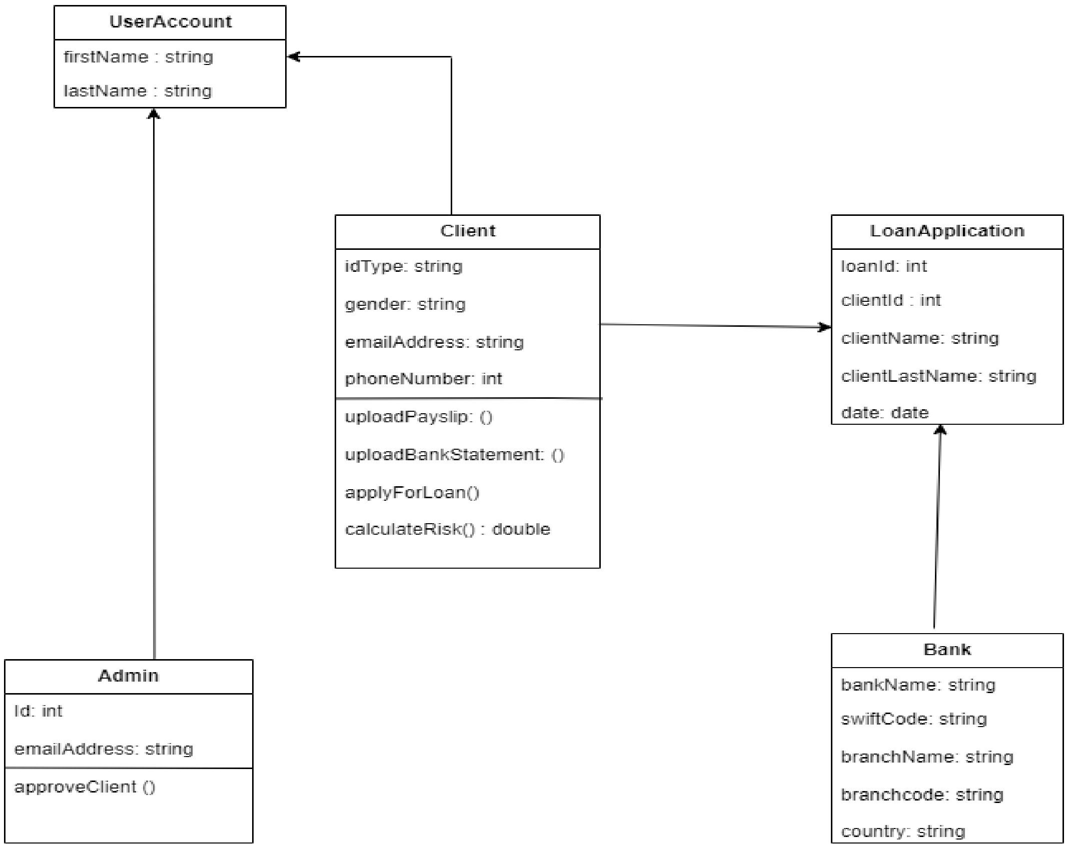
Fundamental Acheivements Case

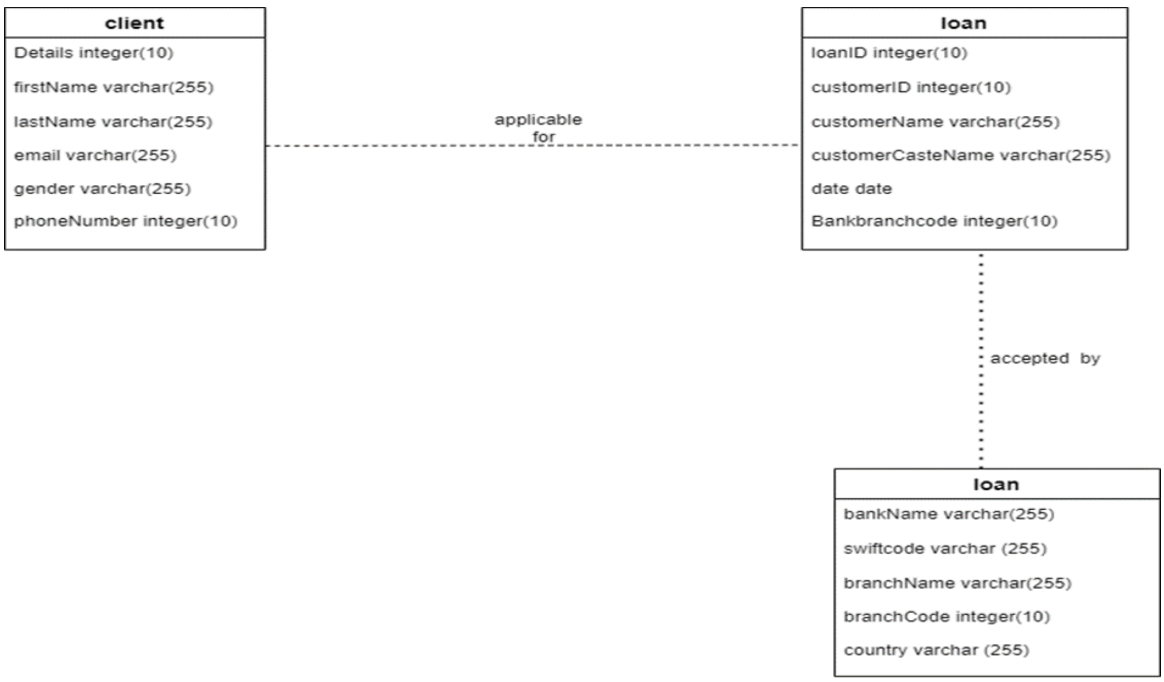
Procedures	Actor	Summary of the Procedures
1	Loan Company	Obtain a loan by completing the necessary fields.
2	Loan Company	Based on their assessment of the loan application they can choose between the following: Approve the application Decline the application
3	System	When the client's request for financing is handled, the system notifies them by sending them an email.

ER Diagram

An Entity-Relationship Diagram (ERD) is a visual representation used in database design to model and depict the structure of a database system. ERDs are used to illustrate how different entities or objects in a database are related to each other and how data flows between them. They are a fundamental tool in designing and understanding the structure of relational databases (Biscobing, 2019).

The relationship within the Client, Loan, and Bank entity sets out on the below ERD diagram. It demonstrates the relationships among the data generated by this system.

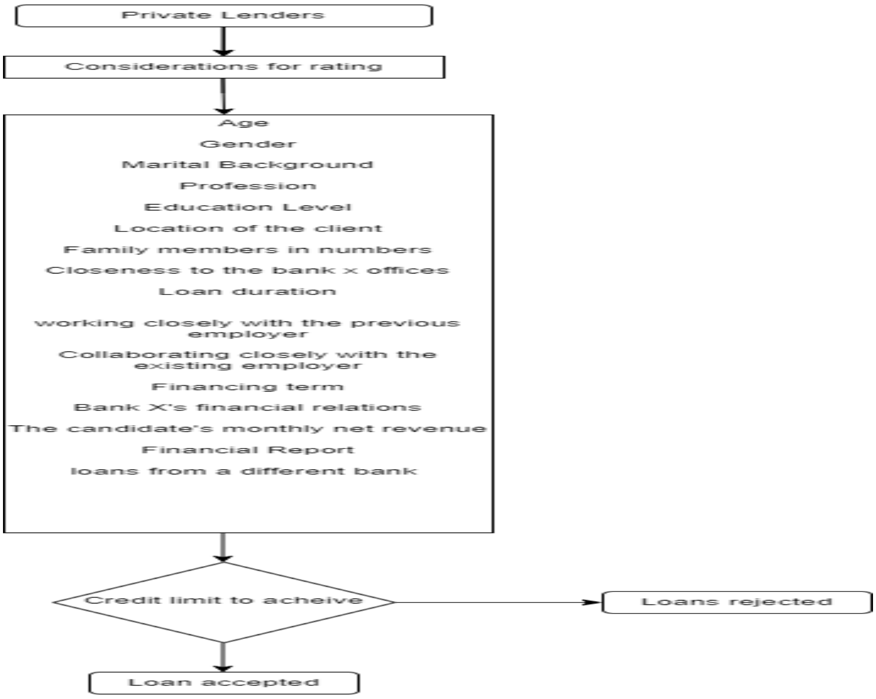




Construction of a Financial Rating System.

The Credit Score System for Anyone prototype, a novel and possibly better rating model, is being designed and developed in this part. Finding the different elements that affect a candidate's trustworthiness was the first step in developing credit rating algorithms. These standards were established after a thorough investigation of numerous customer loan-related articles and websites.

Credit Scoring Process



The financial standing of every applicant was evaluated based on socioeconomic variables. There are multiple attributes with various scores for each SCMI element. The credit score of an individual is calculated using an aggregate of 16 factors.

The range of the credit score is 2 to 49. A person's credit score can range from 2 to 49, with 49 being the maximum. As a result of the greater likelihood of failure and perceived creditworthiness, people with lower credit ratings are seen as less trustworthy compared to those with better credit ratings.

Quality	**Riskclass**	**Credit Score Range**	**Credit Score %**
Highest	A	45 - 48	90% - 100%
Good	B	36 - 42	75% - 89%
Average	C	25 - 35	50% - 74%
Below Average	D	< 25	Below 50%

When assessing loan applications, credit rating is a crucial consideration. The risk class A in this approach, which denotes the lowest risk and the highest application quality, is assigned to applicants with credit scores between 44 and 48. This group includes the top ten percent of students who scored between 90 and 100 percent. Risk class B, which includes credit scores between 36 and 42, denotes good application quality. Risk class C, which denotes medium application quality, is assigned to candidates with scores greater than 25 but lower than 35.

The model sets a cutoff score of 25, which is equal to one-half of the maximum credit score of 48. Any applicant who receives a score below 25 will not be approved for a loan.

The cutoff score essentially denotes the minimum credit score that is acceptable. Applications with scores below 25 won't be accepted, while those with scores over 25 will be given loan approval. Depending on their allocated risk class, applications below the cutoff score carry a significant risk, while those above it have a moderate risk. The lowest risk of default is indicated by risk class A, which also has the highest credit score. The risk of default is lowest for risk class B, which is supported by a high credit score. Due to an average credit score, risk class C indicates a moderate amount of default or credit risk. Risk class D denotes a high risk environment and subpar creditworthiness.

Implementation

In the subsequent section of this chapter, we will delve into the practical aspects of implementing the system discussed earlier. This will involve providing insights into the technology employed and offering illustrative code snippets. The primary programming language utilized for building the system was Python, primarily to execute the machine learning model. A significant component of the system is the integration of a machine learning model, which plays a crucial role in assessing whether a user poses a high or low credit risk. We will also provide a comprehensive explanation of the model's inception and development, with detailed code listings for reference.

Financial Risk Analysis System

```
In [1]: import pandas as pd

In [2]: df = pd.read_csv("D:\Personal\Diksoochi\Projects\Scorecardpy Tutorial\german_credit_data.csv")

In [3]: df.head()
```

Out[3]:

	Age	Sex	Job	Housing	Saving accounts	Checking account	Credit amount	Duration	Purpose	Risk
0	67	male	2	own	NaN	little	1169	6	radio/TV	good
1	22	female	2	own	little	moderate	5951	48	radio/TV	bad
2	49	male	1	own	little	NaN	2096	12	education	good
3	45	male	2	free	little	little	7882	42	furniture/equipment	good
4	53	male	2	free	little	little	4870	24	car	bad

```
In [5]: import scorecardpy as sc

In [ ]: pip install scorecardpy

In [6]: y = 'Risk'

In [7]: bins = sc.woebin(df,y)

[INFO] creating woe binning ...

C:\Users\Aditya.Nathireddy\Anaconda3\lib\site-packages\scorecardpy\condition_fun.py:113: UserWarning: The positive value in "Ri
sk" was replaced by 1 and negative value by 0.
  warnings.warn("The positive value in \"{}\" was replaced by 1 and negative value by 0.".format(y))

Binning on 1000 rows and 10 columns in 00:00:12

In [8]: sc.woebin_plot(bins)
```

In the provided code listing, we observe the process of loading data from a CSV file that comprises the dataset used for training the credit risk evaluation model. This data is pivotal for the model's learning process.

On line 17, we identify the selection of specific features from the dataset that the model will utilize to make predictions regarding the user's creditworthiness. These features serve as crucial inputs for the model's decision-making.

Moving on to line 28, the code demonstrates the incorporation of the sickie-learn library, specifically leveraging the linear model package housing the Logistic Regression algorithm. This algorithm is instrumental in training the model to assess a user's credit status based on the given features.

At line 32, the code assigns the Logistic Regression algorithm to the variable "lr," and subsequently, on line 33, the model undergoes training through the "fit" function. This function employs the data from the previously loaded CSV file, as established in line 9, to instruct the model and enhance its ability to evaluate credit standing accurately.

Section 3: Testing

Section 3.1 Reports on the Testing

We evaluated each person's creditworthiness using a variety of credit scoring techniques as part of our credit evaluation procedure. This involved the application of Discriminant Analysis (DA), Logistic Regression (LR), and a model of credit rating specifically designed for individuals. We performed a comparative analysis to assess the correctness of the credit scoring model generated by LR and DA in order to determine the efficacy of these approaches. We looked closely at each credit rating model's output and made comparisons between them.

Individual Finance Scoring Model (IFSM)

We created the "Finance Scoring Model for Individual (IFSM)" finance scoring model, which took into account all of the significant aspects such as socio demographic data, credit history, loan tenure, age, and occupation.

Classification results using Individual Finance Scoring Model

Monitoring Group		Expected Team		
		Financial Score		
		0 Bad	1 Good	Percentage
Financial Score	0 Bad	90	0	100.0
	1 Good	0	148	100.0
Total Percentage				100.0

A fifty percent cutoff score was used to compare each person's overall credit score to after the whole credit score was calculated by combining individual credit scores from different predictors. An applicant for a loan would have their application approved if their credit score was higher than this threshold than denied if it was below. The finance analysis team looked into the applications whose credit ratings were exactly the same as the cutoff score before accepting them.

The Credit Scoring Model for Individuals predicts that 90 applicants, or 38.4% of the total application pool, have credit ratings below the cutoff point, which puts them at risk of being bad applicants or defaulting on their loans. This is based on a sample of 238 applicants. Conversely, 148 candidates, or 61.6 percent of the total, have finance scores higher than the cutoff point. These people are thought to be excellent clients since they have excellent creditworthiness and a minimal default risk. This model is incredibly accurate—it gets a perfect 100 percent accuracy rate, misclassifying zero applications as good or bad. Consequently, misclassification has no related penalties and no Type I or Type II errors.

We calculated a total credit score by putting the credit scores of all predictors together. An individual's overall credit score was compared to a 50 percent cutoff score. When the credit score was above the cutoff, the decision was made to approve and give the loan, and when it was below the cutoff, it was rejected. All candidates who lied exactly at the cut-off score are allowed, but credit analysis will look into them further

According to the Credit Scoring Model for Individuals, out of 238 applicants, 90 are anticipated to be bad or defaulters, accounting for 38.4 percent of the overall population, and these defaulter applicants have credit scores below the cutoff score.

There are 148 applicants, or 61.6 percent of the overall population, who have a credit score higher than the cutoff, indicating that they are likely to be good customers with solid creditworthiness and a lower risk of default. This model is 100 percent accurate overall. There are 0 applications who are classified as bad as good and 0 applicants who are classified as good as bad. As a result, because there are no Type I and Type errors, there is no penalty associated with misclassification

Logistic Regression

Logistic regression is a statistical model used for binary outcome prediction based on input variables. It's commonly used for classification tasks where the goal is to determine one of two categories. This model calculates the probability of the binary outcome and is employed in various fields, including machine learning, statistics, and social sciences, for tasks like risk assessment and medical diagnosis (David, 2010). Using the sixteen factors and the logistic regression credit scoring model (LR), the classification results are as follows:

Classification results of Logistic Regression

Monitoring Group		Expected Team Financial Score Bad	Expected Team Financial Score Good	Accurate Percentage
Financial Score	Bad	80	2	98
	Good	3	137	97.5
Total Percentage				97.75

The LR classification results show that 137 applicants (63 percent) out of 217 are above the cut point 0.5 and thus acceptable for loan grant, indicating that they are good applicants, while 80 applicants (37 percent of the total population) are predicted to be bad or defaulters.

Because the P-value of LR was less than 0.01, it was determined that default predictors are statistically related at the 80 percent confidence level. The correct classification rate of LR with a cut value of 0.5 was 97.75.

The two types of errors that need to be discussed are Type I and Type II errors. When a good credit application is misidentified as a bad credit application, it is called a Type I error; similarly, when a good credit application is misidentified as a horrible credit application, it is called a Type II error. Our results show that Type II error is 1.3 percent, while Type I error is 1.1 percent.

Discriminant Analysis

Discriminant analysis was used to determine the credit score, which was the dependent variable in our study. We were able to build a strong credit scoring model with sixteen independent variables by using this advanced technique. These independent variables cover a variety of personal and financial aspects, allowing us to evaluate a person's creditworthiness in a thorough manner. In order to determine the correlation between these independent characteristics and the credit score and to make well-informed decisions about credit approval, discriminant analysis, a potent statistical technique, was utilized.

Discriminant Analysis Classification Outcomes

Discriminant Analysis Classification Outcomes

		Finance or Credit Score	Bad	Good	Total
Initial	Count	Bad	78	2	80
		Good	3	134	137
	Percent %	Bad	97.5	2.5	100
		Good	2.1	97.9	100
Corelated Validation	Count	Bad	90	7	97
		Good	4	116	120
	Percent %	Bad	90.1	9.9	100
		Good	3.3	96.7	100

Cross-validation is a vital step performed in the analysis, specifically for the cases under examination. During this process, each case is classified using functions derived from all available cases, ensuring a comprehensive assessment. Notably, the results are highly promising, with an impressive 97.7% of the original group cases being correctly classified. Even during correlated validation, a remarkable 93.4% of the cases are accurately classified, demonstrating the robustness of the analysis. The cutoff value set at 0.5 serves as a crucial threshold, assisting in the classification process and contributing to the overall accuracy of the assessment.

The results of the discriminant analysis (DA) categorization reveal that 97.5% of applicants were projected to belong to the higher credit risk category, while 97.9% were anticipated to be classified as good applicants. Notably, the Type I error rate stands at 2.1%, with a slightly higher Type II error rate of 2.5%. Interestingly, the larger Type II error rate doesn't pose significant concerns because it implies that some strong candidates may be erroneously categorized as potentially risky, which is a more manageable error in terms of cost. The original group cases achieved a commendable 97.7% correct classification rate. Subsequent to cross-validating the DA results, it is projected that 90.1% of candidates are likely to fall within the higher risk category, while 96.7% are expected to be categorized as good applicants. However, during cross-validation, the Type I error rate increases to 3.3%, with the Type II error rate rising to 9.9%.

The significance of the DA model is supported by a p-value of less than 0.01, signifying a strong association among default predictors at a 95% confidence level. This reinforces the reliability of the model in making accurate credit risk assessments. Furthermore, the entire model is considered significant due to the p-value being less than 0.01. In the cross-validation process, using a cutoff value of 0.5, the DA model achieves a correct classification rate of 95.5%. These findings underscore the effectiveness and credibility of the credit assessment model.

Conclusion

This prototype aims to evaluate the creditworthiness of individuals who have obtained personal loans, with the goal of streamlining the credit approval process and reducing the incidence of non-performing loans within financial institutions. Within the framework of the Finance Scoring Model for Individuals (IFSM), it was identified that out of 217 applicants, 80 were categorized as high credit risk or potential defaulters, as their credit scores fell below the established cutoff. Conversely, 137 applicants were deemed strong, reliable customers. The IFSM, prior to any defaults, exhibited an impeccable 100% accuracy in assessing individual borrowers' creditworthiness and effectively distinguishing between high and low-risk loan applications.

To reinforce the outcomes of our credit rating model, we harnessed the power of logistic regression and discriminant analysis. The individual credit scoring model achieved an outstanding 100% accuracy rate. In comparison, logistic regression (LR) demonstrated a commendable accuracy of 97.75%, while the credit scoring model for individuals using discriminant analysis reached an accuracy level of 94.5%. Notably, the IFSM emerged as the most accurate and efficient model among the three, surpassing logistic regression and discriminant analysis in credit scoring.

References

Biscobing, J., 2019. *Techtarget.* [Online] Available at: https://www.techtarget.com/searchdatamanagement/definition/entity-relationship-diagram-ERD
[Accessed 2019].

Chen, H., 2010. *jstor.* [Online] Available at: https://www.jstor.org/stable/23324408
[Accessed 2010].

David, M. a., 2010. *books.google.com.* [Online] Available at: https://books.google.com.np/books/about/Logistic_Regression.html?id=i3IDuAAACAAJ&redir_esc=y
[Accessed 2010].

Halbert, R. a., 2010. *Econpapers.* [Online] Available at: https://econpapers.repec.org/article/eeeeconom/v_3a157_3ay_3a2010_3ai_3a1_3ap_3a110-119.htm
[Accessed 2010].

Hussein A. Abdou, J. P., 2011. *onlinelibraray.* [Online] Available at: https://onlinelibrary.wiley.com/doi/10.1002/isaf.325
[Accessed 2011].

Schreiner, M., 2018. *Microfinance.* [Online] Available at: https://www.microfinance.com/English/Papers/Schreiner_CV.pdf

Stackoverflow, 2020. *Stackoverflow.* [Online] Available at: https://stackoverflow.com/questions/60892714/how-to-get-the-weight-of-evidence-woe-and-information-value-iv-in-python-pan
[Accessed 2020].

ThomaCrook, E., 2002. *Google.com.* [Online] Available at: https://books.google.com.np/books/about/Credit_Scoring_and_Its_Applications.html?id=GMWcWuBDJZUC&redir_esc=y
[Accessed 2002].