**Credit Score Classification**

**Python & Machine Learning Project**

**İsmail Aksu**

**Table of Contents:**

1. Introduction

2. Data Collection and Preprocessing

3. Variable Analysis

4. Model Development

5. Conclusions

6.Recommendations

7. References

8. Appendices

**1.Introduction**

Project Objective:

The aim of this project is to develop a classification model for credit score assessment. Credit score is a crucial metric used to evaluate the financial status of individuals or organizations. The primary objective of this project is to leverage data analysis and machine learning techniques to develop a classification model that assesses customers' credit risk.

The development of such a model plays a significant role in the lending process of financial institutions. An accurate credit score classification model can assist financial institutions in evaluating loan applications more accurately and identifying risky customers to mitigate credit risk. This, in turn, enables institutions to conduct more effective risk management and enhances financial stability.

The data analysis and machine learning techniques employed in this project are essential for understanding the complex relationships underlying credit score classification and predicting customers' credit risk. These techniques can identify patterns in the dataset, determine factors influencing credit risk, and select appropriate features to improve the model's accuracy.

In conclusion, this project aims to contribute to the improvement of credit score classification, making it more accurate and reliable. A robust credit score classification model will enhance both the risk management practices of financial institutions and provide customers with a fairer credit assessment process.

**2. Data Collection and Preprocessing**

During the data analysis phase of my project, various analyzes were performed on the data set obtained from "https://statso.io/credit-score-classification-case-study/".This dataset includes information about customers' incomes, ages, loan amounts, payment histories, and other financial and demographic attributes. Using exploratory data analysis (EDA), we examined the structural characteristics of the dataset and relationships between variables. Additionally, we addressed missing data, standardized numerical variables, and performed feature selection to identify important features. This step laid the foundation for subsequent model development and evaluation processes.

**Data:** <https://statso.io/credit-score-classification-case-study/>

**3.Variable Analysis (Python):**

#First, we import our Python libraries.

**metin, ekran görüntüsü, yazı tipi, beyaz içeren bir resim

Açıklama otomatik olarak oluşturuldu**

#upload data

****

#Let's examine the first 5 of our data.

**metin, ekran görüntüsü, yazı tipi, sayı, numara içeren bir resim

Açıklama otomatik olarak oluşturuldu**

#Let’s have a look at the information about the columns in the dataset:

metin, ekran görüntüsü, menü, doküman, belge içeren bir resim

Açıklama otomatik olarak oluşturuldu

#Before moving forward, let’s have a look if the dataset has any null values or not:

metin, ekran görüntüsü, doküman, belge, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

#The dataset doesn’t have any null values. As this dataset is labelled, let’s have a look at the Credit\_Score column values:

metin, ekran görüntüsü, yazı tipi, sayı, numara içeren bir resim

Açıklama otomatik olarak oluşturuldu

#I will start by exploring the occupation feature to know if the occupation of the person affects credit scores:

metin, ekran görüntüsü, diyagram, çizgi içeren bir resim

Açıklama otomatik olarak oluşturuldu

#There’s not much difference in the credit scores of all occupations mentioned in the data.

#Now let’s explore whether the Annual Income of the person impacts your credit scores or not:

metin, ekran görüntüsü, diyagram, çizgi içeren bir resim

Açıklama otomatik olarak oluşturuldu

#According to the above visualization, the more you earn annually, the better your credit score is.

#Now let’s explore whether the monthly in-hand salary impacts credit scores or not:

metin, ekran görüntüsü, diyagram, çizgi içeren bir resim

Açıklama otomatik olarak oluşturuldu

#like annual income, the more monthly in-hand salary you earn, the better your credit score will become.

#Now let’s see if having more bank accounts impacts credit scores or not:

metin, ekran görüntüsü, diyagram, çizgi içeren bir resim

Açıklama otomatik olarak oluşturuldu

#Maintaining more than five accounts is not good for having a good credit score .A person should have 2 – 3 bank accounts only. So having more bank accounts doesn’t positively impact credit scores.

#Now let’s see the impact on credit scores based on the number of credit cards you have:

metin, ekran görüntüsü, diyagram, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

#Justlike bank accounts, having more credit cards will not positively impact your credit scores.Having 3 – 5 credit cards is good for your credit score.

#Now let’s see the impact on credit scores based on how much average interest you pay on loans and EMIs:

metin, diyagram, ekran görüntüsü, plan içeren bir resim

Açıklama otomatik olarak oluşturuldu

#If the average interest rate is 4 – 11%, the credit score is good.Having an average interest rate of more than 15% is bad for your credit scores.

#Now let’s see how many loans you can take at a time for a good credit score:

metin, ekran görüntüsü, diyagram, çizgi içeren bir resim

Açıklama otomatik olarak oluşturuldu

#So you can delay your credit card payment 5 – 14 days from the due date. Delaying your payments for more than 17 days from the due date will impact your credit scores negatively.

#Now let’s have a look at if frequently delaying payments will impact credit scores or not:

metin, ekran görüntüsü, diyagram, çizgi içeren bir resim

Açıklama otomatik olarak oluşturuldu

#So, delaying 4 – 12 payments from the due date will not affect your credit scores. But delaying more than 12 payments from the due date will affect your credit scores negatively.

#Now let’s see if having more debt will affect credit scores or not:

metin, ekran görüntüsü, diyagram, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu#In outstanding debt of $380 – $1150 will not affect your credit scores. But always having a debt of more than $1338 will affect your credit scores negatively.

#Now let’s see if having a high credit utilization ratio will affect credit scores or not:

metin, diyagram, ekran görüntüsü, öykü gelişim çizgisi; kumpas; grafiğini çıkarma içeren bir resim

Açıklama otomatik olarak oluşturuldu

# Credit utilization ratio means your total debt divided by your total available credit. According to the above figure, your credit utilization ratio doesn’t affect your credit scores.

#Now let’s see how the credit history age of a person affects credit scores:

metin, diyagram, ekran görüntüsü, çizgi içeren bir resim

Açıklama otomatik olarak oluşturuldu

#So, having a long credit history results in better credit scores.

#Now let’s see if your monthly investments affect your credit scores or not:

metin, ekran görüntüsü, diyagram, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

#The amount of money you invest monthly doesn’t affect your credit scores a lot.

**4.Model Development:**

#As the Credit\_Mix column is categorical, I will transform it into a numerical feature so that we can use it to train a Machine Learning model for the task of credit score classification:

metin, ekran görüntüsü, sayı, numara, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

#Now I will split the data into features and labels by selecting the features we found important for our model:

metin, yazı tipi, ekran görüntüsü, cebir içeren bir resim

Açıklama otomatik olarak oluşturuldu

#Now, let’s split the data into training and test sets and proceed further by training a credit score classification model:

metin, yazı tipi, çizgi, ekran görüntüsü içeren bir resim

Açıklama otomatik olarak oluşturuldu

#Now, let’s make predictions from our model by giving inputs to our model according to the features we used to train the model:

metin, ekran görüntüsü, yazı tipi, sayı, numara içeren bir resim

Açıklama otomatik olarak oluşturuldu

**5. Conclusions**

This project aims to develop a model for credit score classification. The model developed using data analysis and machine learning techniques could be a valuable tool for assessing customers' credit risk. The results of the project are summarized below:

- The developed classification model has achieved high accuracy and reliability.

- The model has successfully identified significant factors affecting credit scores, ensuring accurate prediction of credit risk.

- The use of this model in evaluating credit applications could enhance the risk management processes of financial institutions and enable more effective credit decisions.

- Future work could focus on improving the model's performance through different feature selection techniques and tuning model hyperparameters.

These findings underscore the importance and impact of this study on credit score classification. Future work could involve further refinement of the model and its integration into real-world applications.

**6.Recommendations**

1. Further Model Refinement: Continuously refine the classification model by exploring additional features and experimenting with different machine learning algorithms. This iterative process can lead to enhanced model performance and better predictive accuracy.

2. Real-time Implementation: Consider implementing the developed model into real-time credit assessment systems used by financial institutions. This would allow for automated credit decision-making processes, resulting in faster and more efficient customer service.

3. Continuous Monitoring: Regularly monitor the model's performance and update it as needed to ensure its effectiveness in adapting to changing market dynamics and customer behaviors. This proactive approach will help maintain the model's relevance and accuracy over time.

4. Ethical Considerations: Pay close attention to ethical considerations related to data privacy, fairness, and transparency when deploying the model in real-world settings. Ensuring fairness and transparency in the credit assessment process is essential for maintaining trust and credibility with customers.

By following these recommendations, future efforts in credit score classification can build upon the foundation laid by this project and further improve the accuracy, efficiency, and fairness of credit assessment processes.

**7.References**

1. Peterson, T. (2022). "Understanding Credit Score Classification." Investopedia. Retrieved from: [https://www.investopedia.com/understanding-credit-score-classification-5194446](<https://www.investopedia.com/understanding-credit-score-classification-5194446>)

2. "Credit Score Classification Case Study." Statso. Retrieved from: [https://statso.io/credit-score-classification-case-study/](<https://statso.io/credit-score-classification-case-study/>)

3. Kaggle. (2023). "Credit Score Classification Notebook." Retrieved from: [https://www.kaggle.com/code/saloni1712/credit-score-classification](<https://www.kaggle.com/code/saloni1712/credit-score-classification>)

**8.Appendices:**

Appendix A: Project Code

This appendix contains the code used in the project. Python code was utilized for data analysis, model development, and obtaining results.

Appendix B: Dataset Description

This appendix provides a detailed description of the dataset used in the project. It includes information about the columns, sample records, and variables in the dataset.

Appendix C: Model Performance

This appendix presents a detailed analysis of the performance metrics evaluating the developed model. Metrics such as accuracy, precision, recall, and F1 score are examined.

Appendix D: Additional Graphs and Tables

This appendix includes additional graphs, tables, or visuals not included in the main text of the project. These supplementary materials contribute to understanding the project and visually presenting the results.

Appendix E: Used Model - Random Forest

This appendix provides a detailed description of the model used in the project. It outlines the implementation of the Random Forest algorithm, the hyperparameters used, and information about the model's performance.