**Credit Score Classification**

The primary goal of this project is to develop an algorithm capable of accurately segregating individuals into credit score brackets. This algorithm aims to reduce manual efforts involved in credit score determination by leveraging machine learning techniques.

Through the utilization of advanced algorithms and predictive modeling, the algorithm endeavors to analyze the complex interplay of financial and personal attributes to predict credit scores effectively.

By automating this process, financial institutions can streamline their operations, minimize risks, and offer better-tailored financial products and services to their customers. Ultimately, the classification of credit scores into distinct brackets [Poor, Standard, and Good] enables lenders to make informed decisions, mitigating potential credit risks while facilitating access to credit for deserving individuals.

Before delving into the insights gleaned from the dataset, it's imperative to recognize the critical role that accurate credit score classification plays in financial decision-making. With the proliferation of machine learning technologies, financial institutions are increasingly turning to intelligent systems to streamline credit assessment processes.

The dataset sourced from Kaggle offers a comprehensive array of financial and personal attributes, providing a rich tapestry of information for analyzing creditworthiness. Through meticulous data cleaning and preprocessing, the raw data has been refined into a format conducive to machine learning model training. Here's a breakdown of the key attributes:

* ID: Unique identifier for each entry.
* Customer\_ID: Unique identification of an individual.
* Month: Month of the year the data was recorded.
* Name: Name of the individual.
* Age: Age of the individual.
* SSN: Social Security Number of the individual.
* Occupation: Occupation of the individual.
* Annual\_Income: Annual income of the individual.
* Monthly\_Inhand\_Salary: Monthly base salary of the individual.
* Num\_Bank\_Accounts: Number of bank accounts held by the individual.
* Num\_Credit\_Card: Number of other credit cards held by the individual.
* Interest\_Rate: Interest rate on the credit card.
* Num\_of\_Loan: Number of loans taken from the bank.
* Type\_of\_Loan: Types of loan taken by the individual.
* Delay\_from\_due\_date: Average number of days delayed from the payment date.
* Num\_of\_Delayed\_Payment: Average number of payments delayed by the individual.
* Changed\_Credit\_Limit: Percentage change in credit card limit.
* Num\_Credit\_Inquiries: Number of credit card inquiries.
* Credit\_Mix: Classification of the mix of credits.
* Outstanding\_Debt: Remaining debt to be paid.
* Credit\_Utilization\_Ratio: Utilization ratio of credit card.
* Credit\_History\_Age: Age of credit history of the individual.
* Payment\_of\_Min\_Amount: Whether only the minimum amount was paid by the individual.
* Total\_EMI\_per\_month: Monthly EMI payments.
* Amount\_invested\_monthly: Monthly amount invested by the individual.
* Payment\_Behaviour: Payment behavior of the individual.
* Monthly\_Balance: Monthly balance amount of the individual.
* Credit\_Score: Bracket of credit score (Poor, Standard, Good) - Target variable.

Now, let's explore some key insights extracted from this dataset, shedding light on the intricate factors influencing credit score classification and paving the way for enhanced decision-making in the realm of financial lending.

A graph of different colored squares

Description automatically generated with medium confidence

The analysis of the provided graphs reveals a noteworthy trend: a considerable proportion of customers with good credit scores demonstrate a proactive approach towards loan repayment, as evidenced by their reluctance to opt for minimum payments. Conversely, customers with poor credit scores exhibit a contrasting behavior, showing a higher inclination towards making only the minimum payments for their loans. This suggests a potential correlation between payment behavior and creditworthiness, highlighting the importance of conscientious repayment habits in achieving and maintaining a favorable credit standing.

A graph of credit and credit card

Description automatically generated with medium confidence

Insights gleaned from the graph indicate a striking correlation between credit scores and monthly in-hand salaries among customers. Specifically, the majority of individuals with poor credit scores tend to have notably lower monthly in-hand salaries in comparison to their counterparts with standard and good credit scores. This disparity underscores the significance of income levels in influencing creditworthiness, suggesting that individuals with higher incomes may be better positioned to manage their finances responsibly, thus leading to improved credit profiles.

A screenshot of a graph

Description automatically generated

Insights drawn from the data indicate that the annual income of customers does not exert a significant influence on their credit scores. This observation is supported by the fact that despite considerable variance in annual income levels, individuals can still maintain a good credit score irrespective of their earnings. This suggests that while income is undoubtedly a factor in financial health, other aspects such as payment behavior, credit history, and debt management play more pivotal roles in determining creditworthiness.

A screen shot of a graph

Description automatically generated

An insightful trend emerges from the analysis, revealing that individuals who invest between $200 to $350 per month tend to exhibit a propensity for maintaining a good credit score. Conversely, those with a standard credit score typically fall within the range of investing between $170 to $200 per month. This suggests a potential link between investment behavior and creditworthiness, highlighting the importance of prudent financial planning and management in achieving favorable credit standings.

A graph of different colored bars

Description automatically generated with medium confidence

Insights derived from the data reveal distinct patterns in credit score distribution based on the presence and composition of a credit mix among individuals.

Remarkably, those without a credit mix predominantly fall within the standard credit score category, with the second-largest group displaying a bad credit score. Conversely, individuals boasting a diversified credit mix predominantly achieve a good credit score, closely followed by a substantial portion maintaining a standard credit score.

Furthermore, individuals with a standard credit mix are inclined to have a standard credit score, with the second-largest segment falling into the bad credit score category. Conversely, those with a poor credit mix primarily exhibit a bad credit score, with the second-largest group aligning with a standard credit score.

These findings underscore the importance of credit mix diversity in influencing creditworthiness, emphasizing its role as a key determinant alongside other factors such as payment history and debt management.

A graph of different colored bars

Description automatically generated with medium confidence

A graph of different colored squares

Description automatically generated with medium confidence

Insights drawn from the data reveal intriguing trends in outstanding debt distribution across different age groups. Notably, customers aged between 30 and 45 emerge as the most significant category with substantial outstanding debts. This pattern suggests that individuals in their prime working years tend to leverage their higher purchasing power, resulting in increased indebtedness. Conversely, customers aged between 45 and 55 exhibit lower outstanding debt levels, indicating a more conservative financial approach as they approach retirement.

Furthermore, it's noteworthy that customers aged between 30 and 45 also represent the highest category in terms of annual income. This underscores their financial capacity and potentially higher debt accumulation. Interestingly, the second-largest age group falls between 14 and 25, indicating that individuals in this age range are capable of generating substantial income despite their younger age. However, despite their earning potential, it's notable that the two largest age groups predominantly exhibit standard or poor credit scores.

In contrast, individuals aged 45 to 55 exhibit a higher proportion of good credit scores compared to the younger age group (14 to 25). This suggests a possible correlation between age, financial behavior, and creditworthiness, highlighting the importance of considering demographic factors in credit risk assessment and financial planning strategies.

Having developed a machine learning model using the refined dataset, our aim was to predict credit scores based on user-input financial parameters. The model underwent training on carefully selected features, utilizing historical data to make accurate predictions.

We achieved an acceptable accuracy rate, considering the implementation of a relatively simple quantile-based approach and the absence of parameter optimization. Despite these constraints, our model demonstrates promising capabilities in credit score prediction, paving the way for potential enhancements and optimizations in future iterations.

#1. Training RandomForestClassifier

Model cross-validation accuracy on train set: 85.51%

Test set Classification Report - RandomForestClassifier

precision recall f1-score support

0 0.87 0.89 0.88 10167

1 0.86 0.78 0.82 10166

2 0.87 0.94 0.90 10167

accuracy 0.87 30500

macro avg 0.87 0.87 0.87 30500

weighted avg 0.87 0.87 0.87 30500

=======================================================

#2. Training ExtraTreesClassifier

Model cross-validation accuracy on train set: 86.57%

Test set Classification Report - ExtraTreesClassifier

precision recall f1-score support

0 0.88 0.90 0.89 10167

1 0.88 0.79 0.83 10166

2 0.89 0.95 0.92 10167

accuracy 0.88 30500

macro avg 0.88 0.88 0.88 30500

weighted avg 0.88 0.88 0.88 30500

=======================================================

#3. Training AdaBoostClassifier

Model cross-validation accuracy on train set: 73.69%

Test set Classification Report - AdaBoostClassifier

precision recall f1-score support

0 0.75 0.70 0.73 10167

1 0.77 0.62 0.68 10166

2 0.71 0.89 0.79 10167

accuracy 0.74 30500

macro avg 0.74 0.74 0.73 30500

weighted avg 0.74 0.74 0.73 30500

=======================================================

#4. Training GradientBoostingClassifier

Model cross-validation accuracy on train set: 77.73%

Test set Classification Report - GradientBoostingClassifier

precision recall f1-score support

0 0.81 0.75 0.78 10167

1 0.78 0.69 0.73 10166

2 0.76 0.90 0.82 10167

accuracy 0.78 30500

macro avg 0.78 0.78 0.78 30500

weighted avg 0.78 0.78 0.78 30500

=======================================================

#5. Training SVC

Model cross-validation accuracy on train set: 74.10%

Test set Classification Report - SVC

precision recall f1-score support

0 0.75 0.68 0.71 10167

1 0.64 0.44 0.52 10166

2 0.68 0.95 0.79 10167

accuracy 0.69 30500

macro avg 0.69 0.69 0.68 30500

weighted avg 0.69 0.69 0.68 30500

=======================================================

#6. Training KNeighborsClassifier

Model cross-validation accuracy on train set: 82.62%

Test set Classification Report - KNeighborsClassifier

precision recall f1-score support

0 0.83 0.92 0.88 10167

1 0.87 0.68 0.76 10166

2 0.86 0.95 0.90 10167

accuracy 0.85 30500

macro avg 0.85 0.85 0.85 30500

weighted avg 0.85 0.85 0.85 30500

=======================================================

In conclusion, our exploration encompassed multiple advanced machine learning algorithms, each meticulously evaluated through essential metrics such as precision, recall, F1-score, and accuracy. These metrics serve as critical indicators of the model's performance and its suitability for addressing the given problem.

Among the algorithms assessed, Random Forest, ExtraTrees, Gradient Boosting Classifier, and KNeighbours Classifier emerged with the highest test set accuracy. Recognizing the significance of fine-tuning model parameters to optimize performance further, our next step involves conducting hyperparameter tuning on these top-performing algorithms.

This iterative process aims to enhance the accuracy and robustness of our predictive model, ultimately bolstering its efficacy in accurately determining credit scores based on user-input financial parameters. By leveraging advanced machine learning techniques and continuous refinement, we aspire to deliver a reliable and insightful tool for financial decision-making in the realm of credit assessment.