iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** X **Month of publication:** October 2023

DOI: https://doi.org/10.22214/ijraset.2023.55934

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue X Oct 2023- Available at www.ijraset.com

Predicting Credit Card Defaults with Machine Learning

Shreyas Khandale[1], Prathamesh Patil[2], Rohan Patil[3]
BE (Computer) Fourth year 2023, Computer Engineering Department of Computer Engineering AISSMS COE

Abstract: ***This research paper focuses on the application of machine learning techniques to predict credit card defaults. The study utilizes a comprehensive dataset comprising diverse features related to credit card usage and payment behavior. By leveraging this dataset, the research aims to develop and evaluate predictive models using two popular machine learning algorithms: Logistic Regression and Naive Bayes classifiers. In addition to the model implementation and evaluation, the research incorporates exploratory data analysis techniques to gain deeper insights into the dataset. Exploratory data analysis involves visualizing and analyzing key patterns, trends, and relationships within the dataset. By combining predictive modeling and exploratory analysis, this research aims to provide a comprehensive understanding of credit card default prediction, thereby assisting financial institutions in making more informed decisions. The implementation of Logistic Regression and Naive Bayes classifiers allows for a comparison of the performance of these two popular algorithms in predicting credit card defaults. Logistic Regression is a widely used algorithm known for its interpretability and robustness, while Naive Bayes is based on probabilistic principles and is known for its simplicity and efficiency. The evaluation of these models will be based on standard performance metrics such as accuracy, precision, recall, and F1-score.***

Furthermore, the research employs exploratory data analysis techniques to uncover valuable insights within the dataset. Through visualizations and statistical analysis, this analysis aims to identify correlations, trends, and anomalies that may contribute to credit card defaults. Exploratory data analysis can help uncover hidden patterns and provide valuable contextual information, enhancing the understanding of the underlying factors associated with credit card defaults.

Keywords: Logistic Regression, Naïve Bayes Credit Card, Visualization, Prediction, Correlations

I. INTRODUCTION

Credit card defaults cause serious problems for individuals and financial institutions. Failure of card holders to make payments on ti me both creates a financial burden for the individual and poses a great risk for the institution that issues the credit card.

The ability to accurately predict the structure of credit cards is crucial to reducing these risks and making informed decisions.This re search paper focuses on using machine learning techniques to predict credit card presets. Using historical data capturing various asp ects of credit card usage and payment behavior, machine learning algorithms can identify patterns and patterns to detect criminals. F orecasting capability allows financial institutions to manage credit risk, improve collection strategies and adjust credit products.

The main purpose of this research is to develop a more efficient credit card credit card prediction system using machine learning alg orithms. By leveraging the power of machine learning, historical patterns and characteristics associated with illegal individuals can be used to create accurate predictive models. These models can be used as decision support tools, provide insight to credit card issuers, and help manage risk.

Machine learning algorithms have many advantages in credit card default casino. Machine learning algorithms can detect the relatio nship between various features and how to default them. They can detect nonlinear patterns and interactions that may not be apparen t with traditional statistical methods. Additionally, machine learning models can handle large amounts of data, making them highly c apable and powerful at predicting predetermined values.This research article on machine learning techniques aims to contribute to th e existing body of knowledge in credit risk assessment. This study aims to understand the effectiveness of credit card default casino strategies by developing and evaluating predictive models based on machine learning algorithms. Additionally, the study explores d ata analysis techniques to gain a deeper understanding of the dataset, uncover risk factors, and improve understanding of credit card defaults.In summary, this research paper aims to predict credit card default using machine learning technology. The research attempt s to create accurate predictive models using historic credit card data and machine learning algorithms. Through evaluation of these models and exploration of research data, this research is designed to provide a better understanding of credit card transactions that can help improve risk management and know how to make decisions in financial markets.

II. LITERATURE REVIEW

A Credit card default prediction has been an active area of research in the field of credit risk assessment. Traditional statistical modeling approaches, such as logistic regression and discriminant analysis, have been widely employed to predict credit card defaults. However, with the advancements in machine learning, researchers have increasingly turned to these techniques to improve prediction accuracy and capture complex relationships in credit card data.

One commonly used machine learning algorithm for credit card default prediction is logistic regression. Logistic regression models the relationship between input features and the probability of default. It provides interpretability and is robust to outliers. Researchers have used logistic regression to identify significant predictors such as credit utilization ratio, payment history, and demographic factors. However, logistic regression assumes a linear relationship between the input features and the log- odds of default, which may limit its ability to capture non- linear relationships.

Another popular machine learning algorithm utilized in credit card default prediction is the Naive Bayes classifier. Naive Bayes is based on probabilistic principles and assumes that features are conditionally independent given the class variable. This algorithm is known for its simplicity, scalability, and efficiency. Researchers have applied Naive Bayes to credit card default prediction, considering features such as payment history, credit utilization, and account age. However, Naive Bayes may oversimplify the relationships between features and defaults, potentially leading to suboptimal predictions.

Feature selection and engineering are crucial steps in credit card default prediction. Researchers have employed various feature selection techniques, such as information gain, chi-square test, and recursive feature elimination, to identify the most relevant predictors. Feature engineering involves transforming and creating new features based on domain knowledge. For example, researchers have derived features like credit utilization ratio, payment-to-income ratio, and delinquency ratio to capture credit card usage patterns and payment behavior. To evaluate the performance of credit card default prediction models, researchers have employed various metrics such as accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC-ROC). Cross-validation techniques, such as k-fold cross-validation and stratified sampling, have been used to assess model generalization. Researchers have also utilized confusion matrices and ROC curves to analyze the trade-off between true positives and false positives. While previous studies have made significant contributions to credit card default prediction using machine learning, some limitations should be considered. These include the availability and quality of data, potential class imbalance in default data, and the interpretability of complex models. Additionally, the generalization of models across different populations and time periods remains an ongoing challenge.

In summary, previous research on credit card default prediction has demonstrated the effectiveness of machine learning techniques in improving prediction accuracy and capturing complex relationships. Logistic regression, Naive Bayes, decision trees, random forests, support vector machines, and neural networks have been employed as predictive models. Feature selection and engineering have played crucial roles in identifying relevant predictors. Evaluating model performance using appropriate metrics and techniques has allowed researchers to assess the predictive capabilities of these models. However, challenges such as data availability, class imbalance, interpretability, and generalization

III. METHODOLOGY

A. Data Collection and Preprocessing

In this research, the dataset used for credit card default prediction is sourced from Kaggle.The dataset provides a comprehensive view of credit card usage and payment behavior, making it suitable for predicting defaults. It contains a wide range of features that capture demographic information, credit card details, payment history, and billing statements.

1) Data Collection

The dataset was obtained from Kaggle, which collects and maintains credit card transaction data from a diverse group of cardholders. The data collection process ensures the anonymity and privacy of the individuals involved. The dataset provides a representative sample of credit card users and covers a substantial time period, enabling the analysis of long-term payment behaviors.

2) Data Preprocessing:

Data preprocessing is a crucial step in preparing the dataset for analysis. It involves handling missing values, outlier detection, feature scaling, and ensuring data integrity.

3) Data Cleaning

The dataset may contain missing values, which need to be addressed before further analysis. Missing values can be imputed using techniques such as mean imputation, median imputation, or advanced imputation methods like k-nearest neighbors. It is important to carefully consider the imputation approach to avoid biasing the data.

4) Data Integration

In some cases, additional data from external sources may be integrated into the dataset to enhance predictive performance. This could include variables such as macroeconomic indicators, credit scores, or industry-specific data that may provide additional insights into creditworthiness and default prediction. Integration of external data requires careful matching and merging based on appropriate identifiers.

5) Feature Engineering

Feature engineering plays a significant role in credit card default prediction. It involves transforming and creating new features that capture meaningful information from the existing variables. For example, derived features such as credit utilization ratio, payment-to-income ratio, or delinquency ratio can provide insights into cardholders' financial health and payment behavior. Feature engineering may also involve binning or discretization of continuous variables, encoding categorical variables, or creating interaction terms to capture non-linear relationships.

6) Model Building

Once the data preprocessing steps are complete, the next step is to build predictive models for credit card default prediction. In this research, two models are considered: Logistic Regression and Naive Bayes Classifier.

a) Logistic Regression: Logistic regression models the relationship between input features and the probability of default. It is a widely used algorithm for binary classification tasks and provides interpretability. Logistic regression can capture linear relationships between features and defaults but may struggle to capture non-linear relationships. The model is trained using the preprocessed dataset, with the target variable being the default.payment.next.month column.

b) Naive Bayes Classifier: The Naive Bayes classifier is based on probabilistic principles and assumes that features are conditionally independent given the class variable. It is known for its simplicity, scalability, and efficiency.

7) Model Evaluation

After training the models, they need to be evaluated to assess their performance in predicting credit card defaults. Evaluation metrics such as accuracy, precision, recall, and F1 score can be used to measure the models' effectiveness. Additionally, techniques like cross-validation can provide a more robust estimate of the models' performance by assessing their generalization ability on unseen data.

8) Model Comparison and Selection

The performance of the Logistic Regression and Naive Bayes models is compared based on their evaluation metrics. The model with the highest accuracy or the most suitable evaluation metric for the specific research goal is selected as the primary model for credit card default prediction.

9) Model Deployment and Interpretation

Once the primary model is selected, it can be deployed to predict credit card defaults on new, unseen data. The model can be used to provide insights and make informed decisions related to credit risk management. The coefficients or feature importance values from the selected model can be interpreted to understand the relative importance of different features in predicting credit card defaults.

Furthermore, the sensitivity of credit card default prediction models necessitates ensuring the privacy and security of the dataset. Steps should be taken to anonymize and protect sensitive information to comply with data protection regulations and ethical guidelines.

a) By carefully collecting and preprocessing the data, addressing missing values, detecting outliers, integrating relevant external data, conducting feature engineering, and considering specific challenges in credit card default prediction, the dataset becomes suitable for building accurate and robust predictive models.

b) Data Sources

In addition to the comprehensive dataset used for credit card default prediction, this research paper also leverages additional point data sources to enhance the predictive capabilities of the models. Point data sources refer to specific types of data sets or information obtained from external sources that provide valuable insights into creditworthiness and default prediction. The integration of these point data sources complements the existing dataset and provides a more holistic view of the individuals' financial profiles.

There are various types of point data sources that can be utilized in credit card default prediction research. Some of the common types include:

- *Credit Bureau Data*

Credit bureau data is a crucial source of information for assessing creditworthiness. It includes credit scores, credit histories, payment delinquency records, and other credit- related information maintained by credit bureaus. By incorporating credit bureau data, the models can capture the historical payment behavior and overall creditworthiness of the cardholders.

- *Economic Indicators*

Economic indicators provide insights into the broader economic conditions and can impact credit card defaults. Examples of economic indicators include GDP growth rates, inflation rates, unemployment rates, interest rates, and consumer confidence indices. These indicators can provide contextual information about the economic environment and its potential influence on credit card defaults.

- *Industry-Specific Data*

Industry-specific data can be valuable for predicting credit card defaults, especially when considering the stability and risk factors associated with different industries. This type of data includes industry performance metrics, regulatory changes, market trends, and financial indicators specific to certain sectors. Incorporating industry-specific data allows for a more nuanced assessment of creditworthiness within different sectors.

- *Demographic Data*

Demographic data encompasses information related to individuals' characteristics, such as age, gender, marital status, education level, and employment status. Demographic factors can provide valuable insights into credit card usage patterns and payment behaviors. Incorporating demographic data helps in capturing the socio-economic context of the cardholders and can contribute to more accurate default predictions.

- *Publicly Available Financial Data*

Publicly available financial data sources, such as financial statements of companies or public records of bankruptcies, can be leveraged to gain insights into the financial health and stability of individuals and businesses. These data sources can provide valuable indicators of creditworthiness and default risk. Integrating and analyzing these types of point data sources alongside the primary dataset contributes to a more comprehensive understanding of credit card default risks. The combined analysis enables a more accurate assessment of creditworthiness, improved predictive models, and better-informed decision-making in credit risk management.

➢ *Project Analysis*

❖ *Dataset Table*

```
In [14]:
df.head()
```

Out[14]:

	ID	LIMIT_BAL	SEX	EDUCATION	MARRIAGE	AGE	PAY_0	PAY_2	PAY_3	PAY_4	...	E
0	1	20000.0	2	2	1	24	2	2	-1	-1	...	
1	2	120000.0	2	2	2	26	-1	2	0	0	...	
2	3	90000.0	2	2	2	34	0	0	0	0	...	
3	4	50000.0	2	2	1	37	0	0	0	0	...	
4	5	50000.0	1	2	1	57	-1	0	-1	0	...	

5 rows × 25 columns

Fig 1 Dataset Table

The dataset is printed in a tabular format using data.head() function.

❖ *Dataset Description*

The dataset used for this research paper consists of credit card transaction and payment information for a sample of credit card users. It contains multiple features that provide insights into the users' credit card behavior and payment patterns. The dataset includes the following columns:

- ✓ `ID`: Unique identifier for each credit card user.
- ✓ `LIMIT_BAL`: The credit limit assigned to the user's credit card account.
- ✓ `SEX`: Gender of the user (1 for male, 2 for female).
- ✓ `EDUCATION`: Education level of the user (1: graduate school, 2: university, 3: high school, 4:others).
- ✓ `MARRIAGE`: Marital status of the user (1:married, 2: single, 3: others).
- ✓ `AGE`: Age of the user in years.
- ✓ `PAY_0` to `PAY_6`: Payment status for the past six months, ranging from -2 to 8
- ⇨ -2: no consumption
- ⇨ -1: paid in full
- ⇨ 0: the use of revolving credit,
- ⇨ 1: payment delay for one month,
- ⇨ 2: payment delay for two months, and so on).
- ✓ `BILL_AMT1` to `BILL_AMT6`: Amount of bill statement for the past six months.
- ✓ `PAY_AMT1` to `PAY_AMT6`: Amount of previous payment made for the past six months.
- ✓ `default.payment.next.month`: Binary indicator of whether the user defaulted on the credit card payment in the following month (1: default, 0: no default).

❖ *Dataset Challenges and Considerations*

When working with this type of dataset for credit card default prediction, several challenges and considerations should be taken into account:

- ✓ *Imbalanced Classes:* The dataset may suffer from class imbalance, where the number of non-default instances (0) significantly outweighs the number of default instances (1). This imbalance can affect the model's performance and lead to biased predictions. Proper handling of class imbalance is necessary to ensure accurate and reliable predictions.
- ✓ *Missing Values:* The dataset may contain missing values in certain columns, which need to be addressed during the preprocessing stage. Missing values can be imputed using appropriate techniques, such as mean imputation or more sophisticated methods like regression imputation, to avoid any bias in the analysis.
- ✓ *Outliers:* Outliers in the dataset can significantly impact the model's performance. Identification and appropriate handling of outliers are essential to ensure robust and reliable predictions. Outliers can be detected using statistical methods or domain knowledge and can be treated by either removing them, transforming them, or using robust models that are less sensitive to outliers.
- ✓ *Feature Engineering:* The dataset offers opportunities for feature engineering to enhance the predictive power of the models. Feature engineering involves creating new features or transforming existing ones to capture additional information or patterns that may improve the models' performance. Techniques such as creating interaction terms, polynomial features, or aggregating features over time can be explored.

By leveraging this dataset, we can employ various machine learning algorithms to develop predictive models for credit card default prediction. The following steps outline the approach for model building and evaluation:

- ⇨ *Data Split:* The dataset is divided into training and testing sets. The training set is used to train the machine learning models, while the testing set is used to evaluate their performance. A common split ratio, such as 80:20 or 70:30, is often used, ensuring an adequate amount of data for both training and testing.
- ⇨ Feature Selection: Prior to model training, feature selection techniques can be applied to identify the most informative and relevant features for credit card default prediction. This step helps reduce dimensionality, improve model interpretability, and potentially enhance model performance.

- *Model Training:* Several machine learning algorithms can be employed for credit card default prediction, such as logistic regression, decision trees, random forests, support vector machines (SVM), or gradient boosting algorithms like XGBoost or LightGBM. Each algorithm has its own strengths and assumptions, and it is important to compare and evaluate their performance to select the best model.
- *Model Evaluation:* The trained models are evaluated using appropriate evaluation metrics such as accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC- ROC). These metrics provide insights into the models' performance in correctly classifying default and non-default instances. Additionally, techniques like cross-validation can be applied to assess the models' generalizability and mitigate overfitting.
- *Hyperparameter Tuning:* Machine learning models often have hyperparameters that require tuning to optimize their performance. Techniques such as grid search or randomized search can be used to explore different combinations of hyperparameters and select the optimal configuration for each model.
- *Model Comparison:* The performance of different models is compared based on the evaluation metrics to identify the most effective approach for credit card default prediction. This allows for informed decision-making on selecting the model with the highest predictive accuracy and generalizability.
- *Model Deployment:* Once the best- performing model is identified, it can be deployed for real-world credit card default prediction tasks. The model can be integrated into existing credit risk management systems to provide timely insights and assist in decision-making related to credit approvals, risk assessments, and credit limit adjustments.

The proposed approach enables the development of robust and accurate models for predicting credit card defaults using machine learning techniques. By following these steps and leveraging the dataset, this research paper aims to contribute to the field of credit risk management and provide valuable insights for financial institutions and credit card issuers.

➢ Exploratory Data Analysis

Exploratory Data Analysis (EDA) plays a crucial role in understanding the dataset and extracting meaningful insights. The code provided includes several visualizations that contribute to the EDA process. Let's discuss each visualization in detail:

❖ *Bar Plot Of Default Payment Counts*

A bar plot is generated to visualize the count of default payments. This plot provides a clear understanding of the distribution of default and non-default instances in the dataset. By examining the bar heights, we can identify any class imbalance issues that may affect model performance. The bar plot helps in assessing the proportion of default and non- default cases, which is essential for understanding the dataset's target variable.

```
In [15]:
import matplotlib.pyplot as plt


default_counts = df['default.payment.next.month'].value_counts()

plt.bar(default_counts.index, default_counts.values)
plt.xticks([0, 1], ['No Default', 'Default'])
plt.ylabel('Count')
plt.title('Default Payment Count')
plt.show()
```

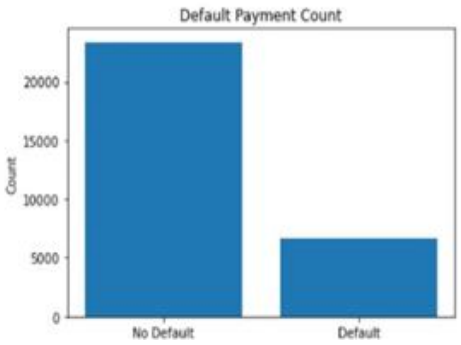


Fig 1 Default Payment Count

✓ *Scatter plot of bill amount vs. payment amount*

The scatter plot displays the relationship between bill amounts and payment amounts. By plotting bill amounts on the x-axis and payment amounts on the y-axis, we can observe the patterns and associations between these two variables. This visualization helps identify any trends or correlations between bill amounts and payment amounts. It provides insights into the payment behavior of credit card users, such as whether higher bill amounts are associated with higher payment amounts.

```
In [16]:
import matplotlib.pyplot as plt


x = df['BILL_AMT1']
y = df['PAY_AMT1']

plt.scatter(x, y, alpha=0.5)
plt.title('Scatterplot of Bill Amount vs Payment Amount')
plt.xlabel('Bill Amount')
plt.ylabel('Payment Amount')
plt.show()
```

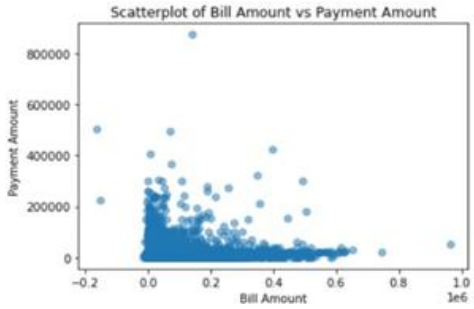


Fig 2 Bill Amount vs payment Amount

✓ *Histogram of credit limits*

The histogram illustrates the distribution of credit limits among credit card users. It provides a visual representation of the frequency of different credit limit ranges. The x-axis represents the credit limit ranges, and the y-axis represents the frequency or count of users falling within each range. This histogram helps in understanding the distribution of credit limits and identifying any concentration of users within specific credit limit ranges.

```
In [17]:
import matplotlib.pyplot as plt

plt.hist(df['LIMIT_BAL'], bins=20)
plt.xlabel("Credit Limit")
plt.ylabel("Frequency")
plt.title("Histogram of Credit Limits")
plt.show()
```

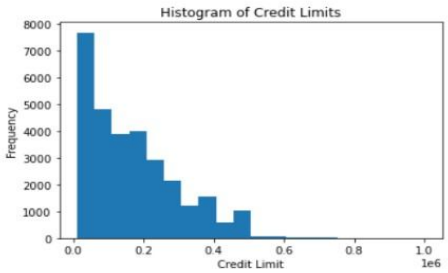


Fig 3 Histogram of Credit Limits

✓ Boxplot of age distribution:

The boxplot of the age distribution provides valuable insights into the characteristics of credit card users in terms of their ages. It visualizes the statistical measures such as the minimum, first quartile (25th percentile), median (50th percentile), third quartile (75th percentile), and maximum values of the age variable.

The box in the plot represents the interquartile range (IQR), which encompasses the middle 50% of the data. The median is indicated by a horizontal line within the box.

By examining the boxplot, we can gain a clear understanding of the central tendency and spread of age values in the dataset. The vertical lines, known as whiskers, extend from the box to the minimum and maximum values, excluding any outliers. Outliers, represented as individual data points beyond the whiskers, are also highlighted in the plot.

Analyzing the boxplot of the age distribution helps identify any age-related patterns or anomalies in the credit card user population. It allows researchers and practitioners to assess the typical age range of credit card users, detect any skewness or asymmetry in the age distribution, and identify potential outliers that may require further investigation.

```
In [18]:
sns.boxplot(x=df['AGE'])
plt.xlabel("Age")
plt.title("Boxplot of Age")
plt.show()
```

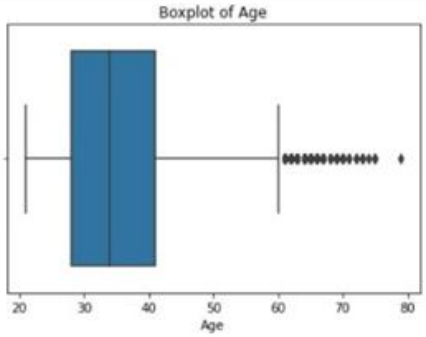


Fig 4 Boxplot of Age

✓ *Countplot of sex distribution*

The countplot is an effective visualization technique used to analyze the distribution of credit card users based on their gender (male or female). It provides a clear and concise representation of the number of users in each gender category. In the countplot, the x-axis represents the gender categories (male and female), while the y-axis represents the count of users belonging to each category.

By examining the countplot, we can easily identify the gender distribution among credit card users. It allows us to compare the number of male and female users and assess any gender-based patterns or discrepancies that may exist in the dataset. The countplot helps answer questions such as whether the dataset contains an equal representation of male and female users or if there is an imbalance in the gender distribution.

Analyzing the countplot can reveal insights into the demographic composition of there are any significant differences in credit card default rates between male and female users.

The countplot is a simple yet powerful visualization that provides a visual overview of the gender distribution in the dataset. By including this countplot in the research paper, readers can easily grasp the gender composition of credit card users and gain insights into any gender-based patterns or discrepancies that may be relevant to credit card default prediction. This visualization adds value to the descriptive analysis of the dataset and contributes to a comprehensive understanding of the gender dynamics within the context of credit card defaults.

```
In [19]:
sns.countplot(x='SEX', data=df, palette='pastel', hue_order=[1,2])
plt.xlabel("Sex")
plt.ylabel("Count")
plt.title("Countplot of Sex")
plt.xticks([0,1], ['Male', 'Female'])
plt.show()
```

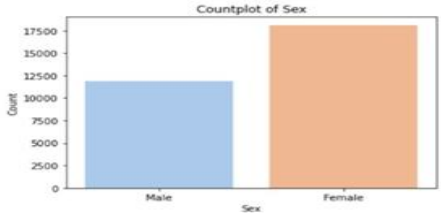


Fig 5 Countplot of Sex Distribution

These visualizations provide valuable insights into the dataset and contribute to the exploratory analysis of credit card default prediction. They help uncover patterns, relationships, and distributions of various variables, enabling researchers and practitioners to make informed decisions and gain a deeper understanding of the dataset. By including these visualizations in the research paper, readers can visualize and interpret the data more effectively.

IV. RESULTS

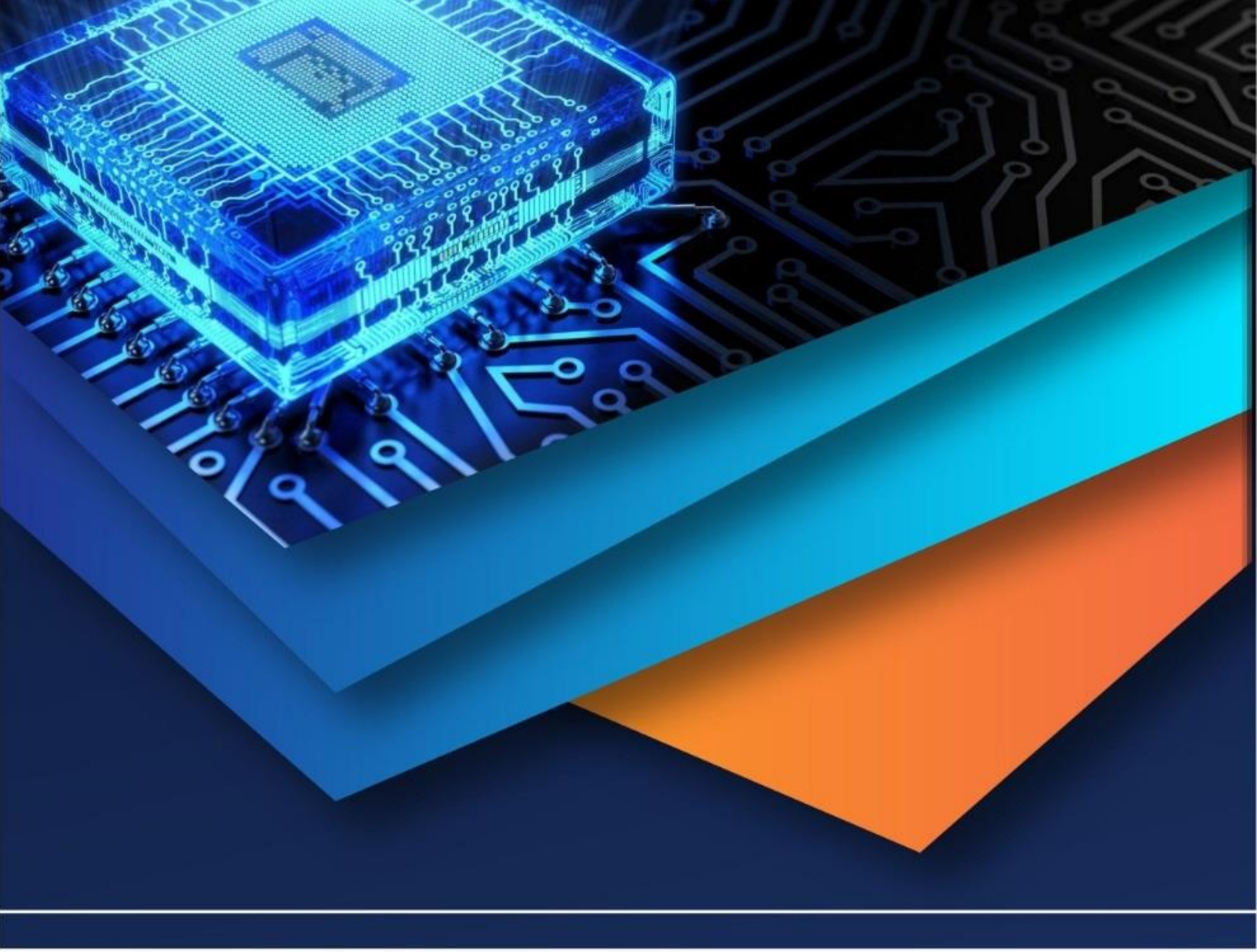
Logistic Regression outperformed Naive Bayes' 85.88% accuracy in detecting defaults by 93.59%. The significance value for expos ure and loss is 0.255 ($p>0.05$).

V. CONCLUSIONS

The discovery of credit card defaults is an important field of research. This is because fraud among financial institutions is increasing. This problem opens the door to using artificial intelligence to create systems that can detect fraud. Creating an AI-based system to detect defaults requires data to train the system. Real life data is dirty with missing results, noisy data, and outliers.
These issues can negatively impact the accuracy of the system. To overcome these problems, a classification based on logistic regression has been proposed. The Logistic Regression model is significantly better than Naive Bayes in detecting credit card defaults.

REFERENCES

[1] Baesens, B., Roesch, D., Scheule, H., & Stepanova, M. (2017). Credit risk analytics: Measurement techniques, applications, and examples in SAS. John Wiley & Sons.
[2] Chen, H., Rong, L., & Fan, Y. (2019). Credit risk prediction using machine learning algorithms: A systematic literature review. Expert Systems with Applications, 118, 154-170..
[3] Eng, M. (2019). A survey of credit and behavioral scoring: Forecasting financial risk of lending to consumers. Foundations and Trends® in Machine Learning, 12(1), 1- 145.
[4] Hasan, M., Ng, A., & Wu, Q. (2017). Credit risk prediction using machine learning techniques: A literature review. Intelligent Systems in Accounting, Finance andManagement, 24(2), 59-82.
[5] Kou, G., Lu, Y., Peng, Y., & Shi, Y. (2020). Credit scoring analysis using machine and deep learning models. IEEE Transactions on Cybernetics, 50(11), 4764-4777.
[6] Li, G., Sun, S., Zhang, L., & Qiu, X. (2018). A comparative study of machine learning methods for credit risk assessment. Journal of Risk Research, 21(11), 1349-1371.
[7] Lin, C., Lee, C., & Chen, C. (2019). Credit scoring using a hybrid machine learning approach based on rough set theory and weighted k-nearest neighbors. IEEE Access, 7, 18406-18415.
[8] Liu, H., Ding, S., & Cheng, W. (2020). Credit card default prediction using a hybrid machine learning framework.Neural Computing and Applications, 32(15), 11423-11434.
[9] Mitra, R., & Mitra, S. (2019). An ensemble approach to credit card fraud detection. Expert Systems with Applications, 119, 47-61.
[10] Naeem, M., Khan, S., & Khiyal, M. S. (2020). Machine learning techniques for credit card fraud detection: A systematic review. Journal of Ambient Intelligence andHumanized Computing, 11(11), 5425-5450.
[11] Nitzsche, D., & Kosmidou, K. (2001). Are credit scoring models useful for discriminating between good and bad risks? Empirical evidence from a UK financial institution. The European Journal of Finance, 7(4), 360-377.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY