**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| Name= Krushna Chaure  Email id= [chaurekrushna143@gmail.com](mailto:chaurekrushna143@gmail.com)  Contribution>   * Doing project individually. * Introduction. * Connecting and Uploading dataset to the google colab notebook. * Understand and Analyze the data. * Data Cleaning & Data Wrangling * Exploratory Data Analysis.  1. Univariate. 2. Bivariate 3. Multivariate  * Hypothesis Testing * Feature Engineering & Data Preprocessing * Challenges Faced * ML Model Implementation  1. Logistic Regression 2. Decision Tree Classifier 3. Support Vector Machine 4. Random Forest Classifier 5. XGBoost Classifier  * Conclusion. * Technical writeup. * Power Point presentation. * Project Summary. |
| **Please paste the GitHub Repo link.** |
| Github Link:- https://github.com/KrushnaChaure/Credit-Card-Default-Prediction.git |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)**  PROBLEM  The credit card default prediction problem is a machine learning task that aims to predict whether a customer is likely to default on their credit card payments. Credit card default occurs when a customer fails to make the minimum payment required on their credit card for a specified period of time.  The problem is of significant importance to the banking and financial industry as it can help institutions identify high-risk customers and prevent potential losses. The objective of the problem is to build a predictive model that can accurately classify customers as default or non-default.  To address the problem, historical data of customers' credit card usage and payment behavior is collected and used to train a machine learning model. The model is trained on a labeled dataset where each record represents a customer and their credit card usage history. The model is then evaluated using a testing dataset to determine its accuracy and performance.  The main challenges in credit card default prediction include handling imbalanced data, dealing with missing values, selecting relevant features, and choosing an appropriate machine learning algorithm. The success of the predictive model depends on the quality and quantity of the data, the accuracy of the model, and the interpretability of the results.  Overall, the credit card default prediction problem is a critical task for financial institutions, as it can help them identify high-risk customers and take necessary steps to prevent potential losses. |
| APPROACH  Credit card default prediction is a machine learning project that involves predicting whether a customer is likely to default on their credit card payments. This is an important task in the banking and financial industry, as it can help financial institutions identify high-risk customers and prevent potential losses.  The project involves several steps, including data preparation, feature selection, model training, model evaluation, and hyperparameter tuning. The data is preprocessed by cleaning, encoding, and normalizing it. The most important features that impact credit card defaults are selected using techniques such as correlation analysis and feature importance analysis. The predictive model is trained using algorithms such as Logistic Regression, Decision Tree Classifier, Support Vector Machine(SVM), Random Forest Classifier and XGBoost Classifier. The performance of the model is evaluated using metrics such as accuracy, precision, recall, and F1 score. Finally, the model is fine-tuned by adjusting its hyperparameters.  Credit card default prediction models can be used to improve the decision-making process of financial institutions by identifying high-risk customers and preventing potential losses. These models can also help customers understand their credit card usage behavior and take necessary steps to avoid defaulting on their payments.  Overall, credit card default prediction is a critical task that can benefit both financial institutions and customers. By building accurate and reliable predictive models, financial institutions can make informed decisions and reduce the risk of losses, while customers can better manage their credit card payments and avoid defaulting.  CONCLUSION  In conclusion, credit card default prediction is an important task in the banking and financial industry, as it can help financial institutions identify high-risk customers and prevent potential losses. The task involves building a predictive model that can accurately classify customers as default or non-default based on their credit card usage and payment behavior.  To build an accurate and reliable predictive model, the approach involves several steps, including data preparation, feature selection, model training, model evaluation, and hyperparameter tuning. The most important features that impact credit card defaults are selected using techniques such as correlation analysis and feature importance analysis.   1. Credit history is a significant factor in predicting default. Those with a poor credit history are more likely to default on their credit card payments. 2. Other factors such as income level, age, and employment status also play a role in predicting default. For example, those with low income or unstable employment may be more likely to default. 3. In order to improve the accuracy of credit card default prediction models, it is important to continually update the models with new data and adjust the weighting of different factors based on their predictive power. 4. **Logistic Regression:-** We have implemented logistic regression and we getting scores nearly 69% for all and f1-sore approx 69% which say that model is not as much accurate that we want. As we have imbalanced dataset, F1- score is better parameter.The logistic regression model did a good job accurately predicting when a client would not default, but did not do very well predicting when a client would default. 5. **Decision Tree Classifier:-** We have implemented Decision Tree Classifier and we getting scores nearly 80% for all and f1-sore approx 79%. As we have imbalanced dataset, F1- score is better parameter.also accuracy score is 99% on training dataset which leads to say that Decision Tree Classifier is overfitted.The Decision Tree Classifier model did a good job accurately predicting when a client would not default, also do very well predicting when a client would default. 6. **Support Vector Machine(SVM):-** We have implemented Support Vector Machine(SVM) and we getting scores nearly 77% all and f1-sore approx 75% which is averagely good. As we have imbalanced dataset, F1- score is better parameter.The Support Vector Machine(SVM) model did a good job accurately predicting when a client would not default, but did not do very well predicting when a client would default. 7. **Random Forest Classifier:-** We have implemented Random Forest Classifier and we are getting around 99% train accuracy and 86% for test accuracy which depicts that model is overfitting. However our f1-score is around 86%, which is not bad.Then we implemented hyperparameter tuning and we are getting around 99% train accuracy and 86% for test accuracy even after we use hyperparameter gridsearchcv the model will not get accuracy which depicts that model is overfitting. However our f1-score is still around 86% after using hyperparameter gridsearchcv which is not bad.The Random Forest Classifier model did a good job accurately predicting when a client would not default, also do very well predicting when a client would default. 8. **XGBoost Classifier:-** We have implemented XGBoost Classifier and we are getting around 91% train accuracy and 85% for test accuracy which depicts that model is very good. However our f1-score is around 84%, which is not bad.Then we implemented hyperparameter tuning and we are getting around 99% train accuracy and 87% for test accuracy after we use hyperparameter gridsearchcv the model will not get accuracy which depicts that model is overfitting, so hyperparameter tuning is not helpfull in XGBoost Classifier to get good model. However our f1-score is still around 86% after using hyperparameter gridsearchcv which is not bad.The XGBoost Classifier model did a good job accurately predicting when a client would not default, also do very well predicting when a client would default. 9. Credit card default prediction is a complex problem that requires careful data cleaning, exploratory data analysis, feature engineering, and modeling. By using various statistical and machine learning techniques we found that XGBoost is most accurate among all,hence we can build accurate models that help credit card issuers minimize their financial risks and improve their profitability. |