TATA CONSULTANCY SERVICES

CREDIT RISK PREDICTION USING MACHINE LEARNING

An in-depth case study exploring machine learning techniques applied to the German Credit Dataset to enhance credit risk assessment and decision-making.

SHASHANK SHIVAM
CT20244445168

CHALLENGES IN ASSESSING CREDIT RISK

- **CREDITWORTHINESS ASSESSMENT ISSUES**
  Financial institutions struggle to accurately assess applicants' creditworthiness, impacting their decision-making.
- **RISK CLASSIFICATION ERRORS**
  Incorrect risk classifications lead to higher defaults and financial instability in institutions.
- **DEFAULTS AND INSTABILITY**
  Higher default rates result in significant instability for banks and lenders, affecting the economy.
- **NEED FOR PREDICTIVE MODELS**
  There is a strong need for effective predictive models to assess an applicant's risk accurately.
- **MACHINE LEARNING IN CREDIT RISK**
  Machine learning can enhance the prediction of good vs. bad credit risks, improving accuracy.
- **GOAL OF PREDICTION**
  The primary goal is to predict whether a loan applicant is a good or bad credit risk.

COMPREHENSIVE OVERVIEW OF DATASET

- **DATASET SOURCE**

  The dataset is sourced from the German Credit Dataset, a well-known resource for credit risk analysis.

- **NUMBER OF RECORDS**

  This dataset comprises a total of 1,000 records, which provide a substantial basis for analysis.

- **FEATURES INCLUDED**

  It includes various features such as personal information, financial history, and loan details.

- **TARGET VARIABLE**

  The target variable is credit risk, classified as Good (1) or Bad (0).

- **IMPORTANCE OF DATASET**

  This dataset is crucial for training machine learning models to predict credit risk effectively.

- **USE IN MACHINE LEARNING**

  It serves as a foundational dataset for implementing machine learning algorithms in finance.

EFFECTIVE DATA PREPROCESSING TECHNIQUES

- **HANDLING MISSING VALUES**
  Identify and impute or remove missing values to maintain data integrity.
- **REMOVING DUPLICATES**
  Eliminate duplicate records to ensure uniqueness in the dataset.
- **CATEGORICAL ENCODING**
  Utilize LabelEncoder and OneHotEncoder to convert categorical features into numerical formats.
- **SCALING NUMERICAL FEATURES**
  Apply scaling techniques like normalization or standardization for numerical attributes.
- **FEATURE ENGINEERING**
  Construct new features like credit-to-income ratio to enhance model performance.

COMPARATIVE ANALYSIS OF ML MODELS

- **LOGISTIC REGRESSION**

  A statistical method for binary classification, useful in predicting credit risk.

- **RANDOM FOREST**

  An ensemble learning method that combines multiple decision trees for improved

- **GRADIENT BOOSTING**

  A technique that builds models sequentially to improve prediction accuracy.

- **PERFORMANCE METRICS**

  Evaluation metrics include Accuracy, Precision, Recall, and F1-Score for model

COMPREHENSIVE ANALYSIS OF MODEL PERFORMANCE

Evaluating key metrics for model selection

MODEL	ACCURACY	PRECISION	RECALL	F1 SCORE
Random Forest	1.00	1.00	1.00	1.00
Gradient Boosting	1.00	1.00	1.00	1.00
Logistic Regression	0.99	0.99	0.99	0.99

INTERACTIVE STREAMLIT UI FOR PREDICTIONS

Current and Future Scope

STREAMLIT APP OVERVIEW

This demo showcases an interactive Streamlit application for credit risk predictions.

INPUT FIELDS

Users can input necessary data through user-friendly fields for accurate predictions.

PREDICTED OUTPUT DISPLAY

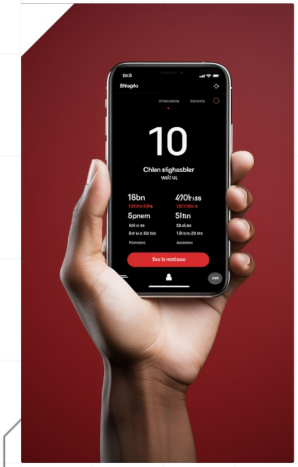
The app displays the predicted outcomes based on user inputs in real-time.

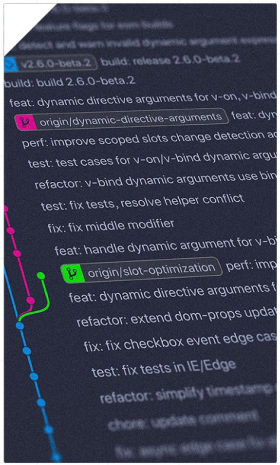
CSV UPLOAD FEATURE

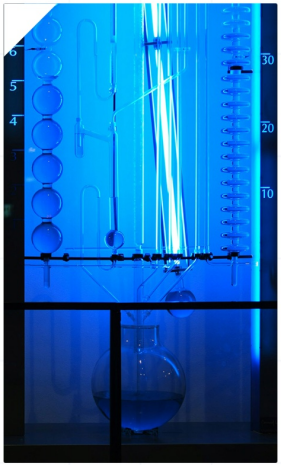
Users have the option to upload CSV files for batch predictions, enhancing usability.

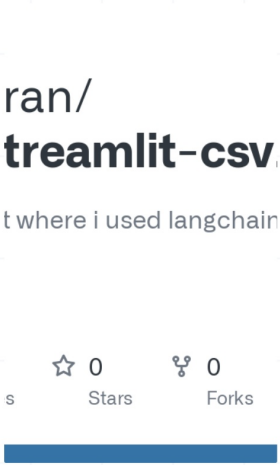
BATCH PREDICTION CAPABILITY

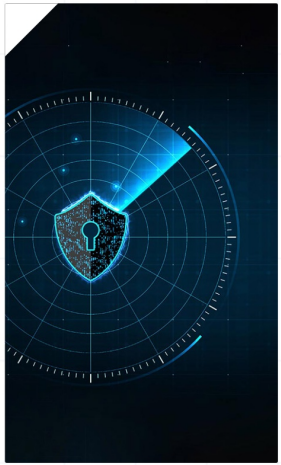
The app can process multiple records at once, making it efficient for large











NAVIGATING CHALLENGES IN CREDIT RISK PREDICTION

Addressing obstacles and enhancing predictive models

DATA IMBALANCE

Data imbalance can lead to biased model predictions, affecting accuracy and reliability.

FEATURE LIMITATIONS

Limited feature sets restrict the model's ability to capture complex patterns in data.

FUTURE WORK: DEEP LEARNING

Incorporating deep learning techniques may enhance model performance and prediction

FUTURE WORK: LARGER DATASETS

Utilizing larger datasets will help mitigate data imbalance issues and improve model robustness.

FUTURE WORK: MODEL EXPLAINABILITY

Enhancing model explainability is crucial for gaining stakeholder trust and regulatory

CONCLUSION AND DEMO OF CREDIT RISK SYSTEM

Key takeaways from our predictive model demo

- **CREDIT RISK PREDICTION SYSTEM**

  We successfully built a robust credit risk prediction system to assess creditworthiness.

- **KEY DRIVERS IDENTIFIED**

  The system identifies essential drivers that influence creditworthiness, enhancing decision-making.

- **INTERACTIVE WEB UI DEMONSTRATED**

  The model was showcased through an interactive web UI, allowing users to engage with the prediction process.

ENHANCE YOUR CREDIT RISK UNDERSTANDING

Thank you for your attention.