Credit Score Prediction

Can Berk Durak-20190702043

Decoding Credit Scores: A Comprehensive Analysis

The project revolves around developing a Credit Score Predictor, leveraging machine learning techniques to forecast individuals' credit scores. Credit scores play a pivotal role in financial decision-making, influencing loan approvals, interest rates, and overall access to credit. By accurately predicting credit scores, this project aims to enhance risk assessment for lenders, streamline financial processes, and contribute to informed decision-making in the financial domain

Relevance of the Chosen Dataset

Credit scores are pivotal in shaping financial decisions for both lenders and borrowers.

An individual's credit score directly impacts their access to credit and the terms offered (interest rates, credit limits).

Financial institutions and credit reporting agencies continually seek accurate predictive models for credit scores to streamline their operations

Credit scores play a role in economic activities, influencing spending patterns, investment decisions, and overall economic health.

Data Cleaning and Preparation Steps

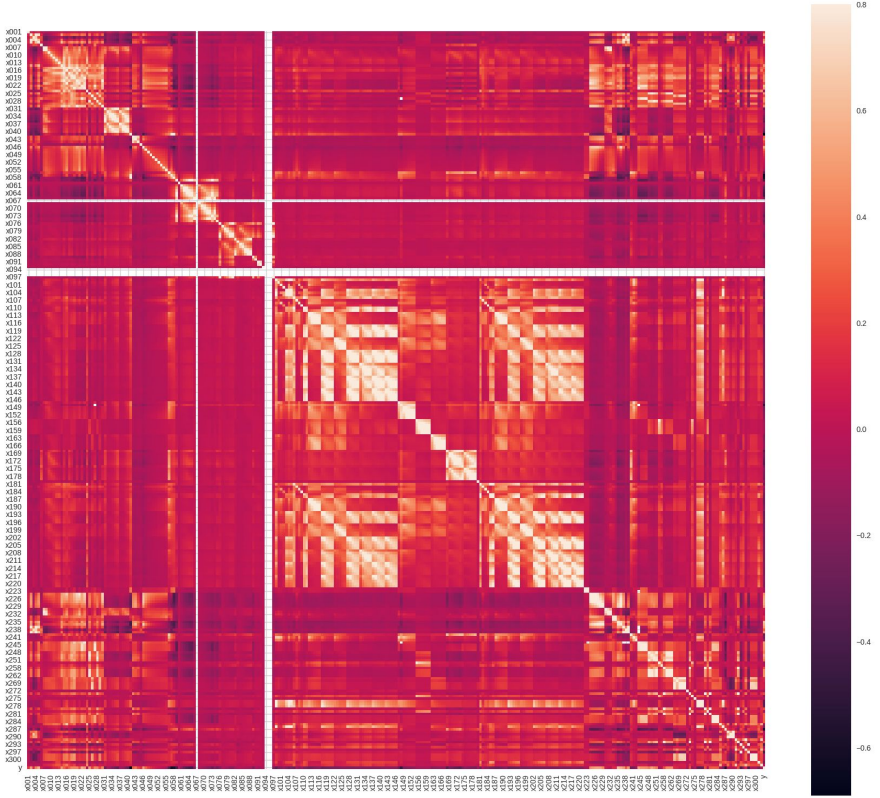
Handling Missing Values: dataset = dataset.dropna(): Removed rows with missing values. This step ensures that the dataset contains only complete records

Removing Duplicates: dataset.drop_duplicates(): Eliminated duplicate rows from the dataset. This ensures that each data point is unique

Standardization: scaler = StandardScaler(): Initialized a StandardScaler instance for feature standardization. Standardization ensures that all features have a mean of 0 and a standard deviation of 1.

In summary, the data cleaning and preparation steps involved removing missing values and duplicates, standardizing the features to a common scale, and selecting a subset of the most influential features for further analysis and modeling

Key visualizations from exploratory data analysis
Exploratory analysis

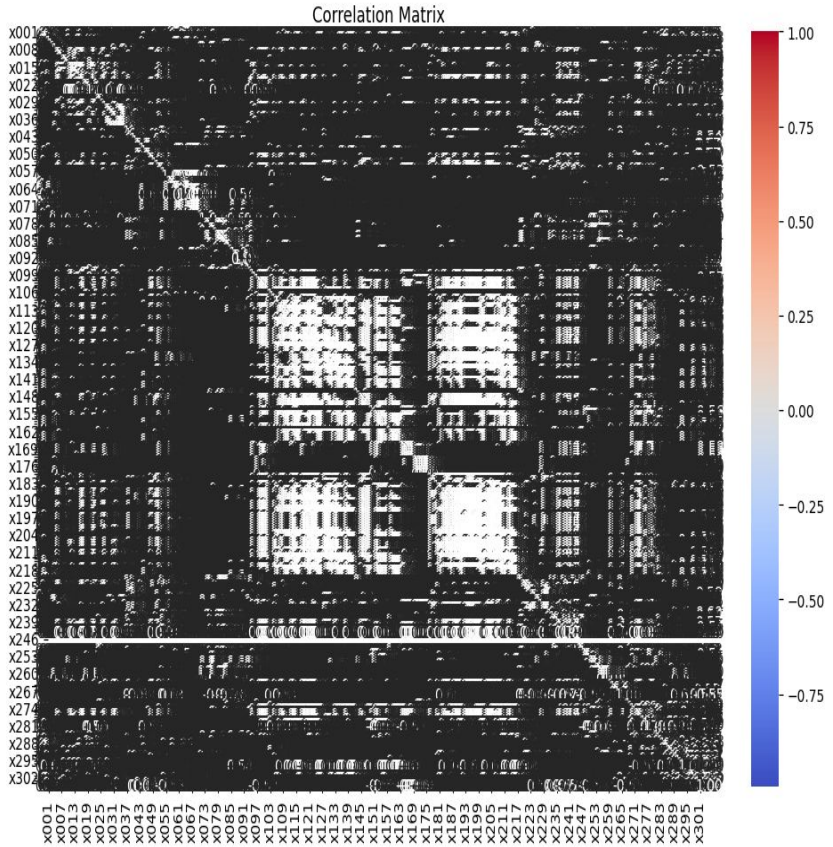


Key visualizations from exploratory data analysis

The central line in the box is the median of the entire data distribution.

The right and left edges in the box are the medians of data distribution to the right and left from the central median, respectively

Key visualizations from exploratory data analysis
Correlation matrix of train data



Models Used for Credit Score Prediction

In the pursuit of accurate credit score prediction, we employed a variety of regression models

Linear Regression, Lasso Regression, Ridge Regression, Elastic Net Decision Tree Regressor, KNN Regressor, Random Forest Regressor, AdaBoost Regressor, Gradient Boosting Regressor, XGBoost Regressor

Then every model's MAPE were calculated and compared and the results showed the best two models are the Gradient Boosting Regressor and Random Forest Regressor

Comparison of models

```
model = RandomForestRegressor()
model.fit(X_test,Y_test)

#Predicting TEST & TRAIN DATA
train_predict = model.predict(X_train)
test_predict = model.predict(X_test)

error_percent = np.mean(np.abs((Y_test - test_predict) / Y_test)) * 100
print("MAPE - Mean Absolute Percentage Error (TEST DATA): ",error_percent )
Y_test, test_predict = np.array(Y_test), np.array(test_predict)
```

```
MAPE - Mean Absolute Percentage Error (TEST DATA):  1.6990274121736826
```

+ Code + Markdown

```
model = GradientBoostingRegressor()
model.fit(X_test,Y_test)

#Predicting TEST & TRAIN DATA
train_predict = model.predict(X_train)
test_predict = model.predict(X_test)

error_percent = np.mean(np.abs((Y_test - test_predict) / Y_test)) * 100
print("MAPE - Mean Absolute Percentage Error (TEST DATA): ",error_percent )
Y_test, test_predict = np.array(Y_test), np.array(test_predict)
```

```
MAPE - Mean Absolute Percentage Error (TEST DATA):  4.575102092073961
```

Key results and insights gained during the analysis

The analysis provided a comprehensive understanding of the dataset, enabled effective feature selection, and highlighted the Random Forest Regressor as the model of choice for accurate credit score prediction. The insights gained contribute to informed decision-making in the realm of credit assessment and financial risk management.