Loan Prediction

Trent

Problem

- In this project we will look how different supervised learning algorithms perform at predicting Loan Application Acceptance

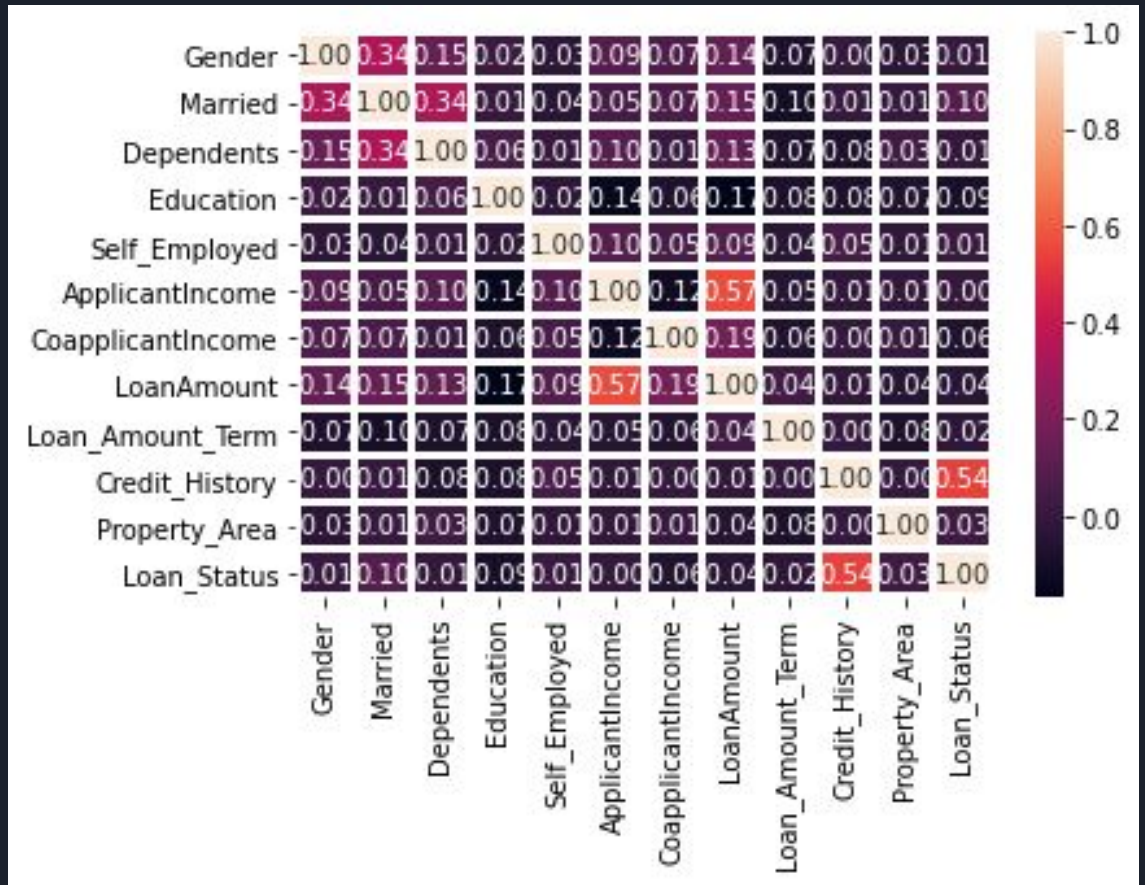
Data Pipeline

1. Find Data
2. Load Data
3. Clean Data
4. Explore Data
5. Model Data
6. Evaluate Models

Data

- Data includes multiple demographic and application data points
- Data includes 614 rows with 11 features
- Data does have nulls but can be filled through imputation
- Data is from Kagle

Data Summary



Models

- Random Forest
- K-nearest neighbors
- Logistic Regression
- Support Vector Machine

Train Data Results

RandomForestClassifier : Accuracy = 98.17

KNeighborsClassifier : Accuracy = 77.60

SVC : Accuracy = 70.26

Logistic Regression : Accuracy = 81.87

Test Data Results

RandomForestClassifier : Accuracy = 75.61

KNeighborsClassifier : Accuracy = 61.79

SVC : Accuracy = 65.04

LogisticRegression : Accuracy = 78.86

Analysis

- While the training data accuracy is quite high the test data have much lower accuracy
  - This points to overfitting of the data
- A larger dataset would probably help the overfitting and yield better results on the train data

Conclusion

- This data set shows promise, but needs more data points for better data