

Project X3: Predicting loan repayment with Python Random Forest Classifier

In this project a simple model was developed to lower the risk of default on credit cards by analyzing applicant's attributes such as age, years at current address, years at current employer, amount of debt, car debt and income using Random Forest classifier with Python. #most important attributes #Example based on my case

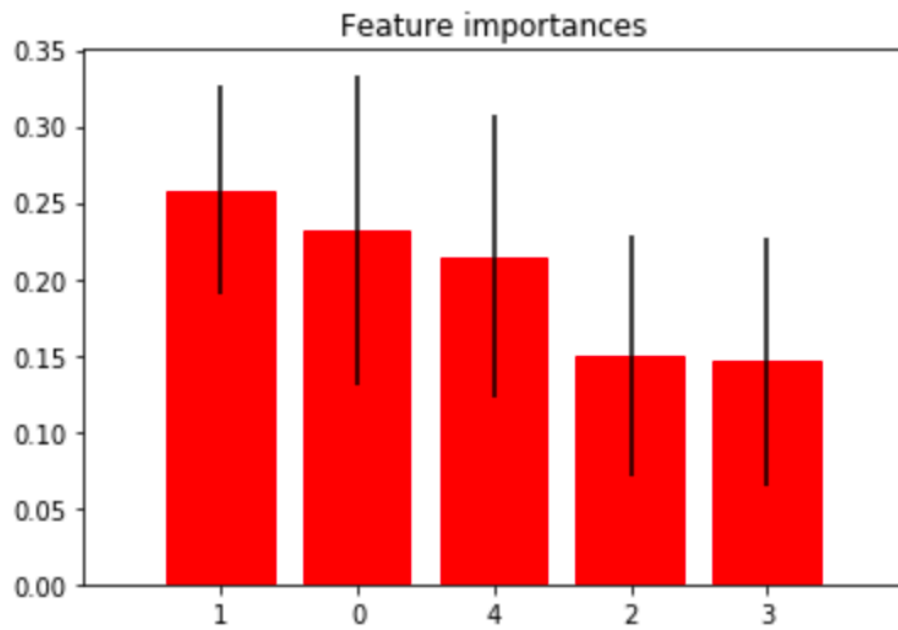
- The training and test datasets used in this project were located in AWS RDS (Check Ref. [1] for the datasets source).
- MYSQL workbench was used to retrieve data into csv files from AWS RDS.
- The Random Forest classifier was used to build a binary classification.
- The importance of applicant's attributes such as age, years at address and ... was investigated.
- The chance of sample persona called Mary to get a loan from the bank, if they use the model developed in this project was checked. #Curiositytest

Keywords: Loan repayment, Risk, Random Forest classifier, Python, AWS RDS

Age, years at current address, years at current employer, amount of debt, car debt and income are numbered from 0 to 4 and the feature ranking details are shown below and in Fig. 1. The most important features are years at address, age and amount of debt.

Feature ranking:

1. feature 1 (0.257920)
2. feature 0 (0.232189)
3. feature 4 (0.214635)
4. feature 2 (0.149483)
5. feature 3 (0.145774)



Ref [1]: Mastering data analysis in Excel course, Duke University course @ Coursera.

Ref [2] https://scikit-learn.org/stable/auto_examples/ensemble/plot_forest_importances.html