Team Presentation Report

Info 7390

Advanced Data Science

**Default credit card client Analysis**

Presented By:

Birwa Galia

(001824469)

galia.b@husky.neu.edu

Milony Mehta

(001869360)

mehta.mil@husky.neu.edu

Shantanu Deosthale

(001851612)

deosthale.s@husky.neu.edu

This report summarizes the analysis of the clients which would be able to pay credit card which would be generated next month based on the data present of previous month bills, person’s personal details and credit card limit.

We are getting the data from:

<https://s3.amazonaws.com/assignment3datasets/default+of+credit+card+clients.xls>

1. **Data Wrangling**

* Exploratory Data Analysis on Python.
* Manipulating the data necessary

1. **Dockerizing the process:**

* Created docker image to pickle the various models and get the accuracy metrics
* Airflow pipeline

1. **Building and Evaluating models:**

* Classification using Random forest, Logistic Regression, KNN Classifier, Extra Trees Classifier, Bernoulli Navies Bayesian

1. **Creating User interface for the Bank**

* **Bank:**  The bank will be able to predict whether user will be able to pay the next month bill or not.
* **The output will be either 0 or 1 where 0 is NO and 1 is Yes**

**PART 1: DATA WRANGLING**

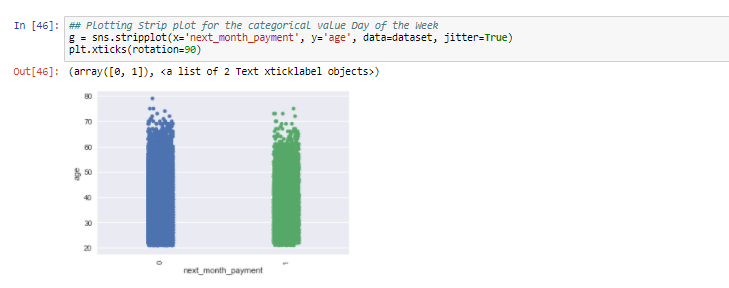
**1. Exploratory Data Analysis**



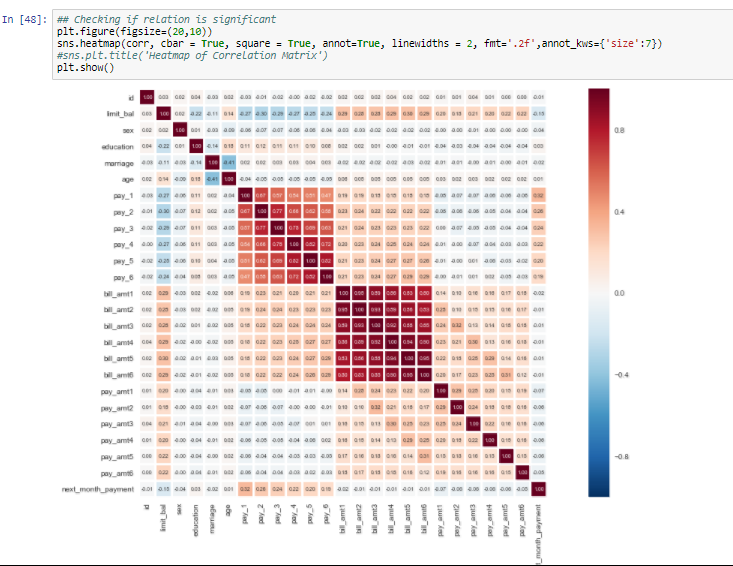
**The** first graph shows that the no. of female that cannot pay the next\_month \_payment is more than that of male.

The second graph shows that the single people wont be able to pay next month payment as compared to the others

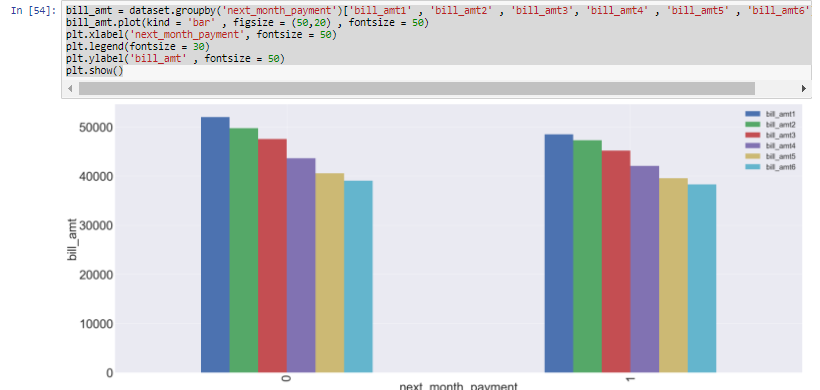
The third graph shows the person who has completed high school will be able to pay.



This graph that there is high possibility of the people whose age is from 20 – 50 will be able to pay.

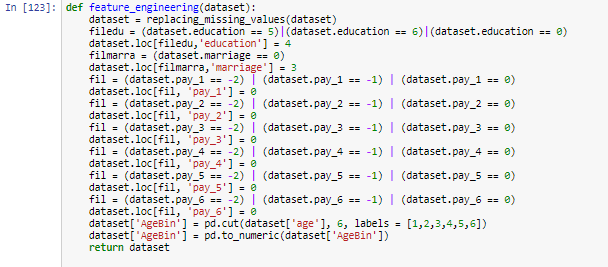


In the above correlation, the pay\_n columns is highly correlated to target value while limit\_bal is correlated to bill\_amt and pay\_amt



This graph shows that bill\_amt is dependent on each other

**Feature Engineering**



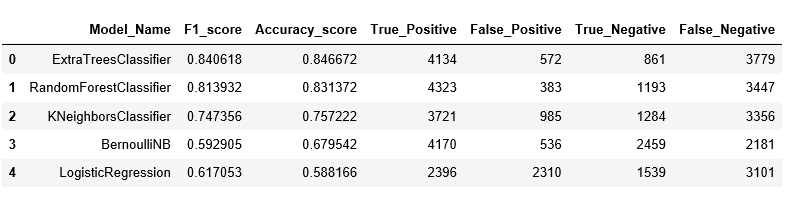
The value which is undefined have been manipulated

**Model Performance**

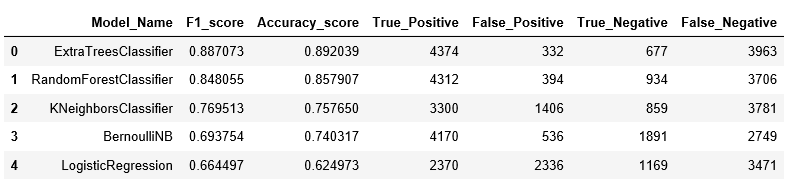
We tried to balance the data using SMOTE and NeverMiss, the sensitivity and specificity was getting better by SMOTE . We performed oversampling using SMOTE to get better accuracy.

**Feature Selection:**

**1. Recursive Feature Elimination**

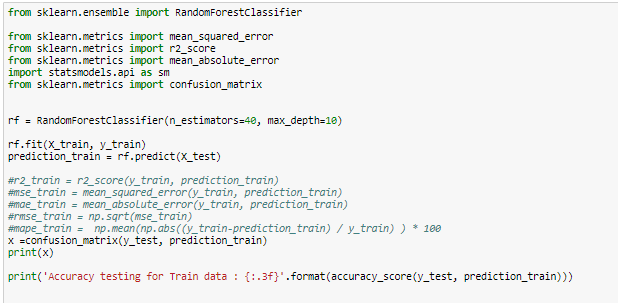


**2. Stepwise Selection**

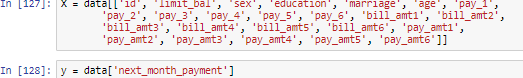


**Building and evaluating the models:**

**1. RandomForest Classifier**

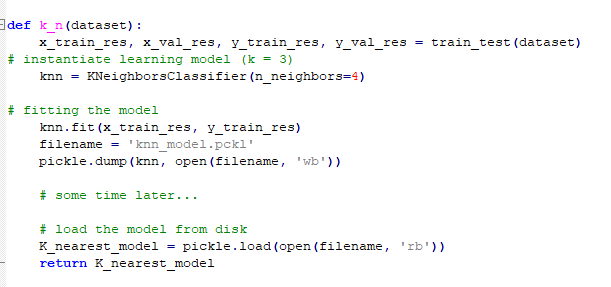


Similarly, for other models the X and Y has been fitted



This X are independent variables and Y is target variable

**Pickling the models**

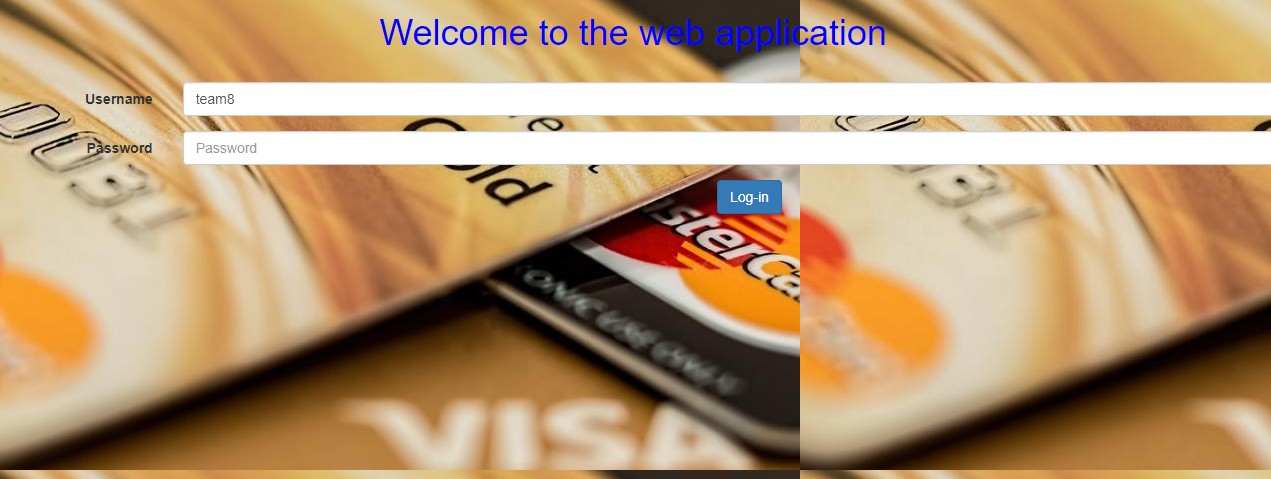


This is the just the pickling of KNN model.

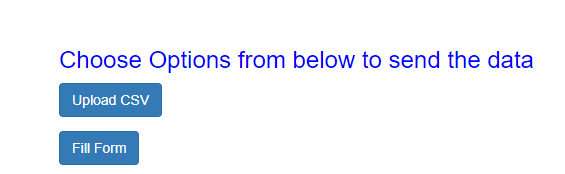
The app for determining the default next month prediction is developed in

* Python Flask
* HTML
* CSS (BootStrap)

**Web Application using Flask**

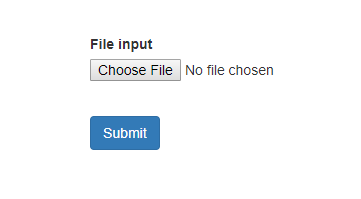


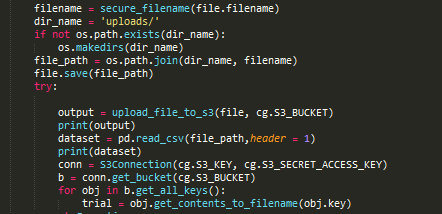
**This is our login page where user of the bank will put his/her credentials for login purpose. The input is static where Username = “team8” and password = “team8”.**



**This are the two ways user can input the data to analyze the result**

1. **Uploads The CSV FIle**





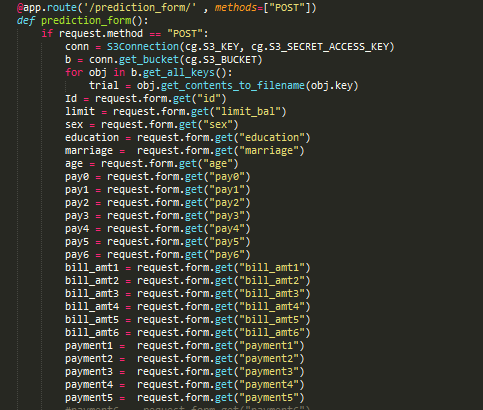
**Takes files from the choose file and save it into local and uploads directed into the S3.**

**This is the page after user clicks upload csv till will only accept that format and downloads the pickle models from S3.**

**2.) Takes User Inputs In HTML Form**



**This is the page after user chooses fill form option where user puts the data manually**

****

**A black and silver text on a screen

Description generated with very high confidence**

**Takes user input values and performs prediction models by fetching Pickle models from S3. Save the prediction into CSV and saves it to the local directory.**

**A screenshot of a social media post

Description generated with very high confidence**

**After user clicks on the submit button it gets the predicted value in 0 and 1 format where 0 is user won’t be able to pay for next month and 1 would be he will be able to pay**