**Credit scoring using social media analytics.**

Here we use social media analytics data and the customer personal information to predict whether the customer who has applied for loan is a potential candidate to return it or might come in the defaulter list.

Here we are not only focusing on the financial institutions lending loans but also P2P transactions which help people with no credit information can supply social media information to get loans

**Data Preparation:**

So here we basically use the twitter information and the network information of the customer to actually figure out their personality and based on the information.

These are the features being used in our model:

Age 21–50yrs

Gender -{male, female}

Marriage -{married, unmarried, divorced}

Children -{have, don’t have}

Family -{Number of family members the borrower has: 1, 2, 3, 4, 5}

Education -{Junior high school, masters, bachelors}

Income - The annual income of the borrower

Car -{yes, no}

Income\_type -Income type, six types

House -{own, parent’s house, rent}

Work\_days 3-7days

Registration\_time -Numbers of minutes before the application the borrower started registration in this platform

Document\_change -Numbers of minutes before the application the borrower changed his/her document

Jobtitle -Job title, six types

Amount -Loan amount

Rate -Loan rate

Period -8–45 months

Time -When the borrower applies for a loan during the day

Social\_stability:- As to how long the the person has used the same network {1, 1–3 years, >3 years}

Social\_exposure: as to how many person’s data has been shared by the customer {1, 2, 3, 4}

Social\_quality : how many defaulters include in that shared information {0, 1, 2 or above}

Sentiment {0-neutral,1-positive}

Here we have not only included the traditional credit information but also the social parameters which we use to make the final prediction.

**Steps to execute the solution:**

Here since we have not deployed the code so,i would be showing manual way to execute the code and reach the final output

1) Data preparation & Exploratory Data Analysis

the first thing is to get the customer's twitter information since this is the part we are majorly using for all the social information and it will also help us to get an insight to how that person's personality would be

We use Twitter API in order to fetch the tweets based on some hashtags which would be related to his lifestyle personality etc

i)Go to the tweepy crawler folder

ii)open tweepy1.py file >download it or copy it to the local IDE

iii)Replace the API keys( u can ask for this if it changes) and the search path in order to fetch tweets

iv)once the tweets appear in the console, copy them to a csv file

v) perform some preprocessing to make it structured as this would be used as input in the next step.

**SENTIMENT ANALYSIS & OTHER SOCIAL PARAMETERS**

After we receive the tweets in the csv format we actually divide the tweets into train and test tweets as we need to find the sentiment analysis on the basis of that, we also take

these tweets and try to analyze them in order to find few other social parameters which are social network quality, social exposure & social stability

i)Go to the sentiment analysis folder

ii)place the structured tweets which we fetched in the previous steps, i.e the train and test tweets in the path where we need to run the file

iii)open the main.py file and run the file

iv)we have divided the sentiments into neutral mode- 0, positive mode-1

v) we get a csv file of the predicted sentiments based on the tweets from the predicted model

vi)use this output as one of the input with the customers information

vii) based on the tweets received and the sentiment predicted we analyze both and try to create the other social parameters like social exposure, social network quality, social stability

**Final prediction:**

After we have done the EDA and prepared the data we now need to integrate everything with the original csv file and give it as an input to our model which we will use to find out how and which of the features are basically affecting the dependent feature and we select those in order to predict the output.

Steps:

1. Go to the python folder and you would find the dataset and the jupyter notebook file for the use case
2. Download the files in your local and run it in either the jupyter notebook local or in Google Colab
3. Running line by line would give you the best results and would fetch you the best features required to model the solution
4. At last when we need to predict, the last tab uses the data in order to predict and give you the output, 1- defaulter, 0 – non defaulter
5. Based on this prediction the financial institution or the p2p lender can think of giving a loan to the individual.

I hope this helps, if there is any help required u can always reach out!