Statement of Work V1

AIDI 1002-02 AI Algorithms 1

Professor – Marcos Bittencourt

Email: - marcos.bittencourt@dcfaculty.mycampus.ca

Submitted by:

Biraj Prajapati (100766706)

1) Rational Statement: <u>Loan Prediction Problem</u>

2) Business Problem in Brief:

Find out the person is eligible to acquire loan based on their qualifications, employment, earning, dependent, their dependent's income, credit history, their loan amount, and loan term. Create a machine learning model to generate loan approval from person's information.

3) Data Source: https://www.kaggle.com/altruistdelhite04/loan-prediction-problem

```
Loan_ID : Unique Loan ID

Gender : Male/ Female

Married : Applicant married (Y/N)

Dependents : Number of dependents

Education : Applicant Education (Graduate/ Under Graduate)

Self_Employed : Self employed (Y/N)

ApplicantIncome : Applicant income

CoapplicantIncome : Coapplicant income

LoanAmount : Loan amount in thousands of dollars

Loan_Amount_Term : Term of loan in months

Credit_History : credit history meets guidelines yes or no

Property_Area : Urban/ Semi Urban/ Rural

Loan_Status : Loan approved (Y/N) this is the target variable
```

Image 1 (Source: towardsdatascience)

4) Data Requirement:

   Gender, married, dependents, education, self-employed, applicant income, co-applicant income, loan amount, loan term, credit history, property area and loan status.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 614 entries, 0 to 613
Data columns (total 13 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   Loan_ID            614 non-null    object
 1   Gender             601 non-null    object
 2   Married            611 non-null    object
 3   Dependents         599 non-null    object
 4   Education          614 non-null    object
 5   Self_Employed      582 non-null    object
 6   ApplicantIncome    614 non-null    int64
 7   CoapplicantIncome  614 non-null    float64
 8   LoanAmount         592 non-null    float64
 9   Loan_Amount_Term   600 non-null    float64
 10  Credit_History     564 non-null    float64
 11  Property_Area      614 non-null    object
 12  Loan_Status        614 non-null    object
dtypes: float64(4), int64(1), object(8)
memory usage: 62.5+ KB
```

5) Data Assumption:
   - Decision tree will have high accuracy than random forest, and logistic regression
   - Loan status is extremely reliant on credit history.
   - Linear relationship between columns.

6) Data Limitations and Constraints:

   Dataset only got 613 entries in train dataset and 366 entries in test dataset.

   Dataset got some missing values in some columns, which need to be cleaned.

7) Test Process:

   Data Cleaning → EDA → Feature Engineering → Preprocessing → Modeling → Model testing

References:

Image 1 source: https://towardsdatascience.com/ml-basics-loan-prediction-d695ba7f31f6

Dataset source: https://www.kaggle.com/altruistdelhite04/loan-prediction-problem