CHAPTER 1

INTRODUCTION

1.1   PROJECT OVERVIEW

One of the most important factors which affect our country's economy and financial condition is the credit system governed by the banks. The process of bank credit risk evaluation is recognized at banks across the globe. "As we know credit risk evaluation is very crucial, there is a variety of techniques are used for risk level calculation. In addition, credit risk is one of the main functions of the banking community.

The prediction of credit defaulters is one of the difficult tasks for any bank. But by forecasting the loan defaulters, the banks definitely may reduce their loss by reducing their non-profit assets, so that recovery of approved loans can take place without any loss and it can play as the contributing parameter of the bank statement. This makes the study of this loan approval prediction important. Machine Learning techniques are very crucial and useful in the prediction of these types of data.

We will be using classification algorithms such as Decision tree, Random forest, KNN, and xgboost. We will train and test the data with these algorithms. From this best model is selected and saved in pkl format. We will be doing flask integration and IBM deployment.

1.2   PURPOSE

The enhancement in the banking sector lots of people are applying for bank loans but the bank has its limited assets which it has to grant to limited people only, so finding out to whom the loan can be granted which will be a safer option for the bank is a typical process. we try to reduce this risk factor behind selecting the safe person so as to save lots of bank efforts and assets. This is done by mining the Big Data of the previous records of the people to whom the loan was granted before and on the basis of these records/experiences the machine was trained using the machine learning model which give the most accurate result. The main objective of our project is to predict whether assigning the loan to particular person will be safe or not.

CHAPTER 2

LITERATURE SURVEY

2.1 EXISTING PROBLEM

It is no secret that the loan industry is a multi-billion dollar industry. Lenders make money by charging interest on loans, and borrowers want to get the best loan terms possible. In order to qualify for a loan, borrowers are typically required to provide information about their income, assets, and credit score. This process can be time consuming and frustrating for both lenders and borrowers. AI / machine learning can be used to predict loan eligibility. As data scientists, it is of great importance to understand some of challenges in relation to loan eligibility and how machine learning models can be built to address those challenges.

2.2 REFERENCES

[1] *B. P. Lohani, M. Trivedi, R. J. Singh, V. Bibhu, S. Ranjan and P. K. Kushwaha*, "Machine Learning Based Model for Prediction of Loan Approval," *2022 3rd International Conference on Intelligent Engineering and Management (ICIEM)*, 2022, pp. 465-470, doi: 10.1109/ICIEM54221.2022.9853160.

[2] *H. Ramachandra, G. Balaraju, R. Divyashree and H. Patil*, "Design and Simulation of Loan Approval Prediction Model using AWS Platform," 2021 International Conference on Emerging Smart Computing and Informatics (ESCI), 2021, pp. 53-56, doi: 10.1109/ESCI50559.2021.9397049.

[3] *V. Singh, A. Yadav, R. Awasthi and G. N. Partheeban*, "Prediction of Modernized Loan Approval System Based on Machine Learning Approach," 2021 International Conference on Intelligent Technologies (CONIT), 2021, pp. 1-4, doi: 10.1109/CONIT51480.2021.9498475.

[4] *S. Barua, D. Gavandi, P. Sangle, L. Shinde and J. Ramteke*, "Swindle: Predicting the Probability of Loan Defaults using CatBoost Algorithm," 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), 2021, pp. 1710-1715, doi: 10.1109/ICCMC51019.2021.941827

[5] *D. P. Rajesh, M. Alam, M. Tahernezhadi, C. Vikram and P. N. Phaneendra*, "Real Time Data Science Decision Tree Approach to Approve Bank Loan from Lawyer's Perspective," 2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA), 2020, pp. 921-929, doi: 10.1109/ICMLA51294.2020.00150.

[6]*Loan Approval Prediction Machine Learning*,

from https://www.analyticsvidhya.com/blog/2022/02/loan-approval-prediction-machine-learning/

[9] *How to predict Loan Eligibility using Machine Learning Models*. (2020, September). Towards Data Science, Medium from https://towardsdatascience.com/predict-loan-eligibility-using-machine-learning-models-7a14ef904057/

2.3   PROBLEM STATEMENT DEFINITION

Banks, Housing Finance Companies and some NBFC deal in various types of loans like housing loan, personal loan, business loan etc in all over the part of countries. These companies have existence in Rural, Semi Urban and Urban areas. After applying loan by customer these companies validates the eligibility of customers to get the loan or not. This project provides a solution to automate this process by employing machine learning algorithm. So the customer will fill an online loan application form. This form consist details like Sex, Marital Status, Qualification, Details of Dependents, Annual Income, Amount of Loan, Credit History of Applicant and others

CHAPTER 3

IDEATION AND PROPOSED SOLUTION

3.1 EMPATHY MAP CANVAS

3.2   IDEATION AND BRAINSTORMING
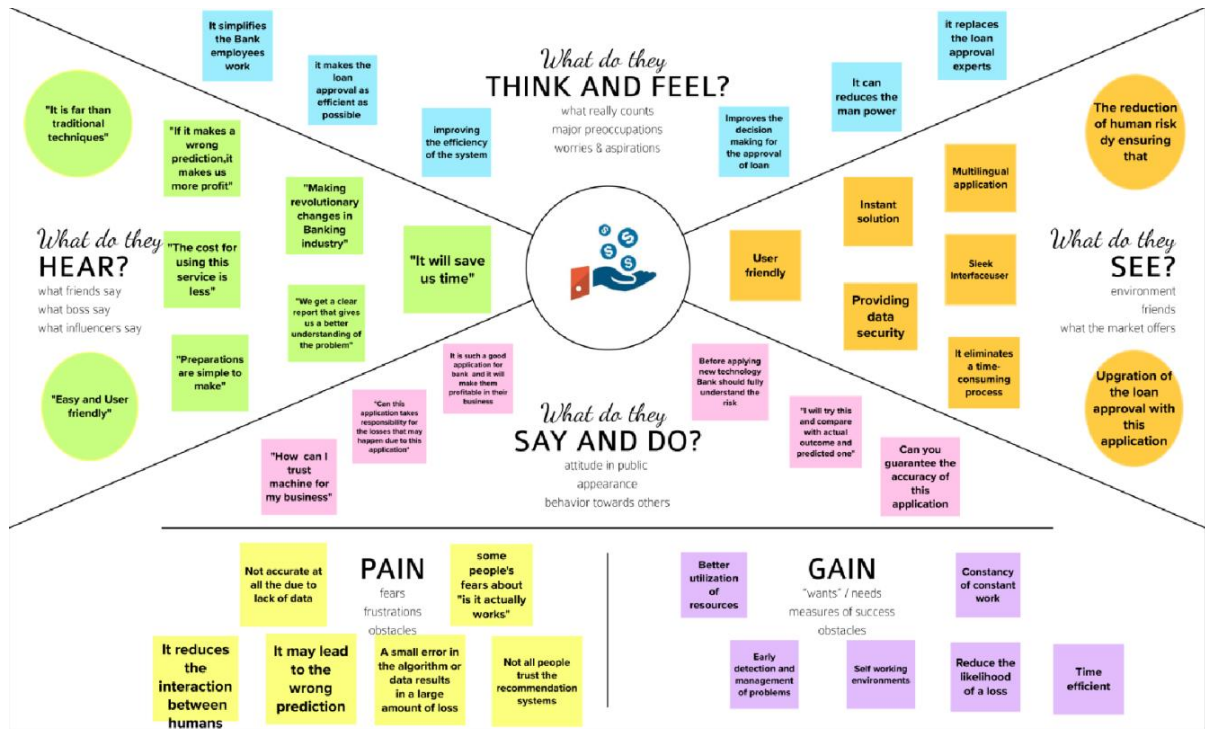
Brain storming and idea prioritization

In this Template share ideas and further ideas can be written here to modify accordingly , leader will modify these chart based on mentor feedback

🕒 2 month to prepare
⧗ 1 month to collaborate
👤 4 members

Created in partnership with 〰️Meta 〰️Meta

A   **Team Gathering**
Sanjay(team leader) will gather group and instruct , ask idea and lead the group further

B   **Set Goal**
Setting up our team goals

Higher Accuracy.
Clean Visuals.
Clean Code.
More Insights

C   **Facilitation tools**
1. Youtube and IBM sessions to learn concepts.
2. Use documentation to code new concepts.
3. use discord , stackoverfow to clear doubts.

Applicant Credibility Prediction for Loan Approval

This data science project will help fnance and banking people who give 100's of loan to their applicant and this group project will help stakeholder will come to the number if applicant who are eligible and not eligible by using data visualization , machine learning algorithms and stakeholder will make data driven decisions from this project

Problem

We are gonna solve this problem by using machine learning algorithms using sci-kit learn and other conventional libraries like spark to handle big data, numpy and pandas for reshaping ,cleaning data,etc

3.3   PROPOSED SOLUTION

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	• Not enough verification process • Fake Documentations • Miscalculations of candidates capacity
2.	Idea / Solution description	People's Need Increased, so demand for loans in banks also increased. Loan approval is a time consumption process, in order to reduce the time consumption we are going to create a chatbot with audio facility. The loan approval can be predicted using any of the following machine learning algorithms like Linear Regression, Decision Tree, Random Forest. Since the prediction is based on a machine learning algorithm, the loan approval process can be accurate
3.	Novelty / Uniqueness	Web application platform that will select the appropriate banks for the users eligibility. And banks can also find the right users based on their requirements.

4.	Social Impact / Customer Satisfaction	Our platform will be helpful not only for the banks to identify the eligible people but also benefit lenders to find next bank which provides loan to user. It will save time and effort of both banks and customers.
5.	Business Model (Revenue Model)	If the loan approval process time consumption is less, more loan prediction can be easily done, it will help for the bank to raise their business growth.
6.	Scalability of the Solution	Every business sector and normal people are able to use this system.

3.4   PROBLEM SOLUTION FIT

CHAPTER 4

REQUIREMENT ANALYSIS

4.1   FUNCTIONAL REQUIREMENTS

Following are the functional requirements of the proposed solution.

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Input	• Selecting the type of loan • Filing the necessary details
FR-2	Eligibility of Loan	• Loan Approval • Loan Rejection
FR-3	Check for approvable amount of loan	• Maximum Eligible loan Prediction
FR-4	Chatbot	• Clarifying user's doubts.

4.2   NON-FUNCTIONAL REQUIREMENTS

Following are the non-functional requirements of the proposed solution.

FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	• Easy to use and self-explaining website. • Easy navigation between pages. • Simple structure for faster access.
NFR-2	**Reliability**	• Ensembling the outputs of various ML models.
NFR-3	**Performance**	• Web Based Application. • Ability to indicate user inputs of erroneous datatypes.
NFR-4	**Availability**	• Application is available 24 / 7 as it is hosted onIBM cloud. • Simple web browser is enough to access thewebsite.
NFR-5	**Scalability**	• Can be extended for other types of loans. • Aadhar and PAN verification can also beimplemented

CHAPTER 5

PROJECT DESIGN

5.1   DATA FLOW DIAGRAMS

Data Flow Diagrams:

A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data is stored.

DFD LEVEL 0
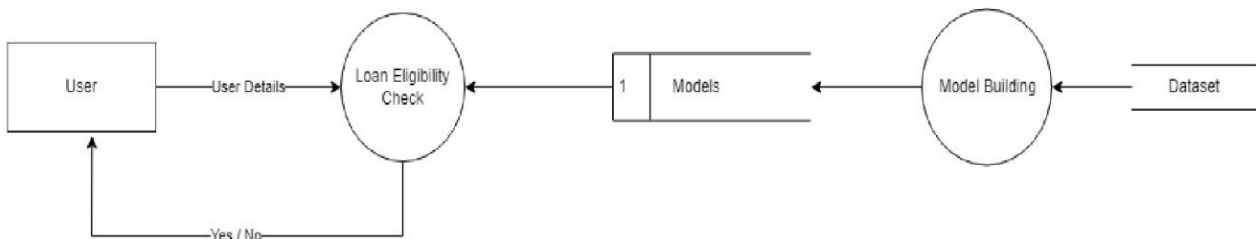
DFD LEVEL 1
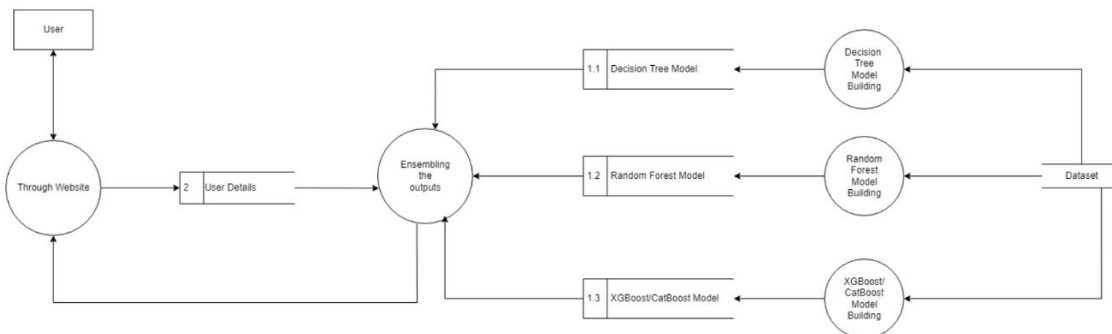
5.2   SOLUTION AND TECHNICAL ARCHITECTURE

Solution Architecture Diagram:
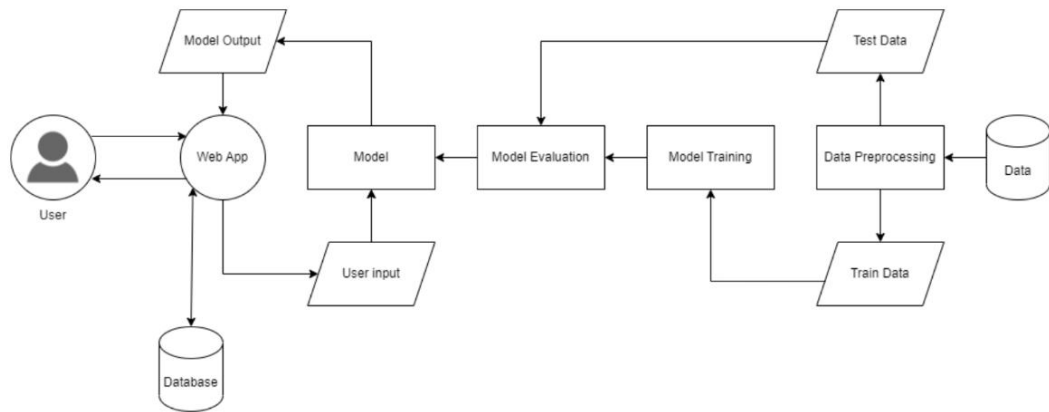
Technical Architecture:
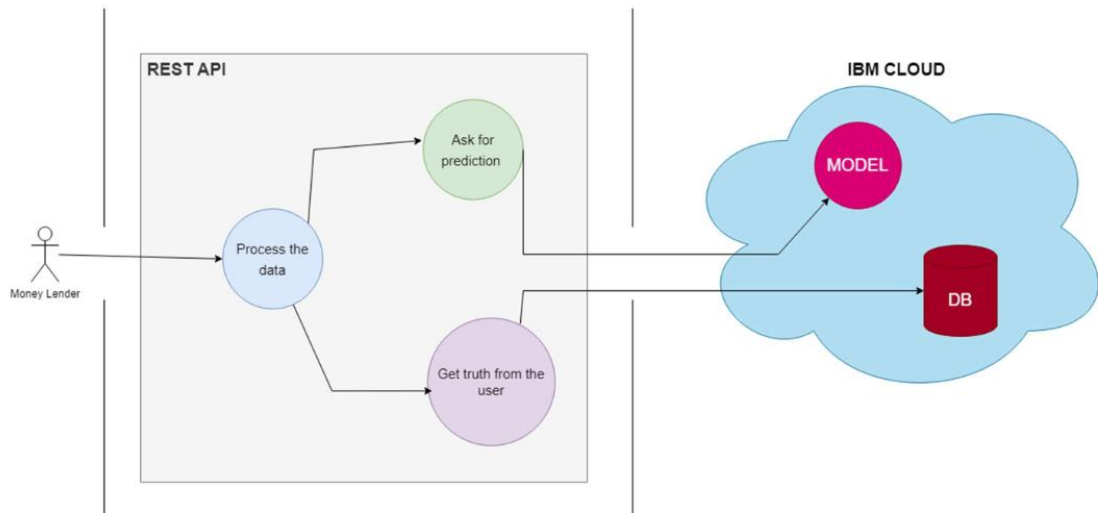
5.3   USER STORIES

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Money Lender (Web user)	Dashboard	USN-1	As a user, I should be able to access the dashboard.	Access the dashboard	Low	Sprint-3
		USN-2	Select the type of loan	Select the type of loan	Medium	Sprint 3

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
	Check for loan eligibility	USN-3	Fill the application with the details of the borrower.	Check the eligibility of the borrower.	High	Sprint-4
Borrower	Dashboard	USN-4	Should be able to access the dashboard.	Access the dashboard.	Low	Sprint-3
		USN-5	Choose the type of loan	Choose the type of loan	Medium	Sprint-3
	Prediction of loan approval	USN-6	Fill the application to check if the loan will be approved or not.	Check for the chances of loan approval. If the loan will not be approved for the current amount, find the amount that will be approved by the lender	High	Sprint-4
	Chatbot	USN-7	Ask doubts to the chatbot	Getting clarified on the doubts	Medium	Sprint-2

CHAPTER 6

PROJECT PLANNING AND SCHEDULING

6.1 SPRINT PLANNING AND ESTIMATION

Product Backlog, Sprint Schedule, and Estimation (4 Marks)

Product Backlog, Sprint Schedule, and Estimation (4 Marks)

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority
Sprint-1	Dataset	USN-4	Downloading the dataset	1	High
Sprint-1		USN-5	Visualizing the dataset	2	Low
Sprint-1		USN-6	Pre-process the dataset	3	Medium
Sprint-1	Machine Learning Model	USN-7	KNN model building	5	High
Sprint-2		USN-8	Decision Tree model building	5	High

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority
Sprint-2		USN-9	Naive Bayes model building	5	High
Sprint-2		USN-10	Fine Tuning the model	3	Low
Sprint-2		USN-11	Evaluation and saving of the models	5	High
Sprint-3	Customer User Interface	USN-12	Model Integration with flask	5	High
Sprint-3		USN-1	As a user, I should be able to access the dashboard.	3	Medium
Sprint-3		USN-2	Select the type of loan	3	Low
Sprint-3		USN-3	Fill the application and check the eligibility of loan approval	5	High
Sprint-4	Deployed the website	USN-13	Register on IBM Cloud	3	Low
Sprint-4		USN-14	Train the ML model on IBM Cloud	5	Medium

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority
Sprint-4		USN-15	Deploy the website on IBM Cloud	8	High

6.2   SPRINT DELIVERY SCHEDULE

Sprint	Total StoryPoints	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Release Date (Actual)
Sprint-1	11	6 Days	24 Oct 2022	29 Oct 2022	11	29 Oct 2022
Sprint-2	18	6 Days	31 Oct 2022	05 Nov 2022	18	05 Nov 2022
Sprint-3	16	6 Days	07 Nov 2022	12 Nov 2022	16	12 Nov 2022
Sprint-4	16	6 Days	14 Nov 2022	19 Nov 2022	16	19 Nov 2022

6.3   REPORTS FROM JIRA

Sprint : 1
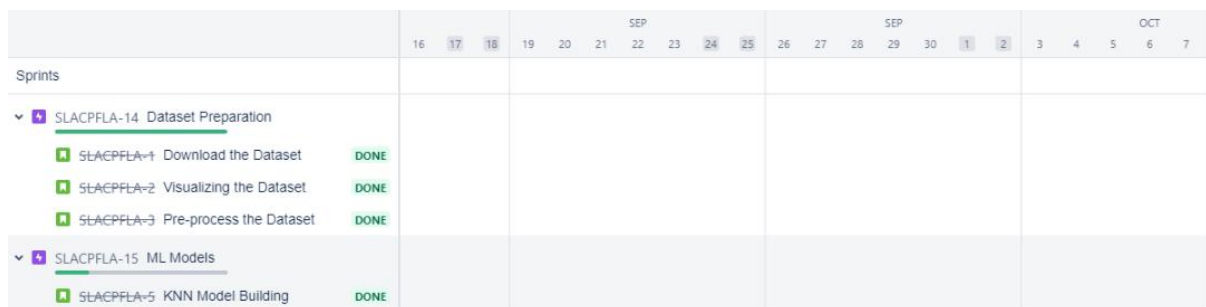
Sprint 2 :
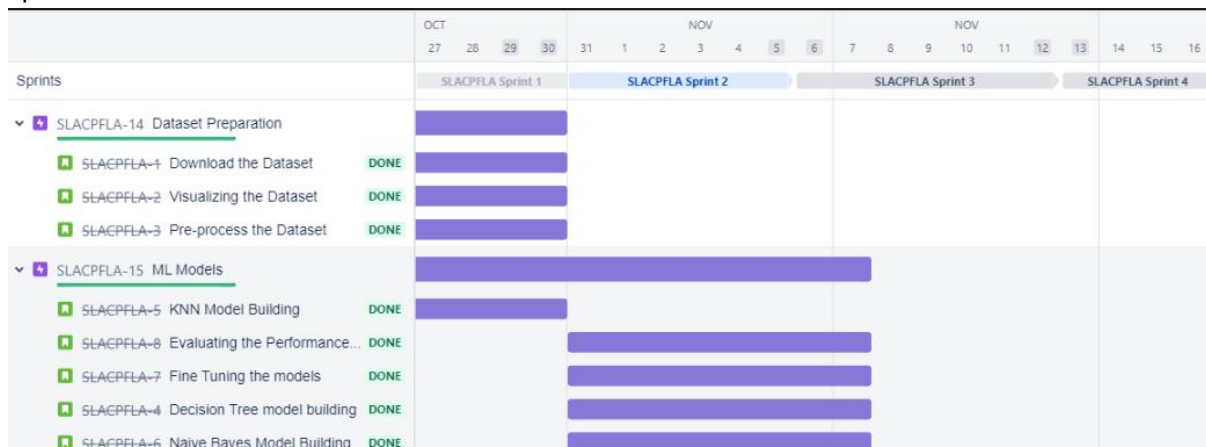
Sprint 3 :
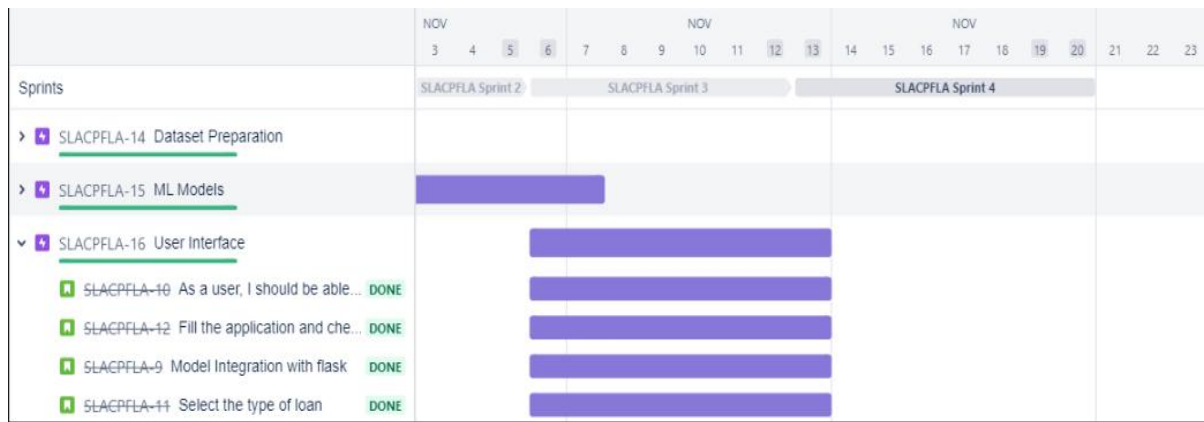
Sprint 4:
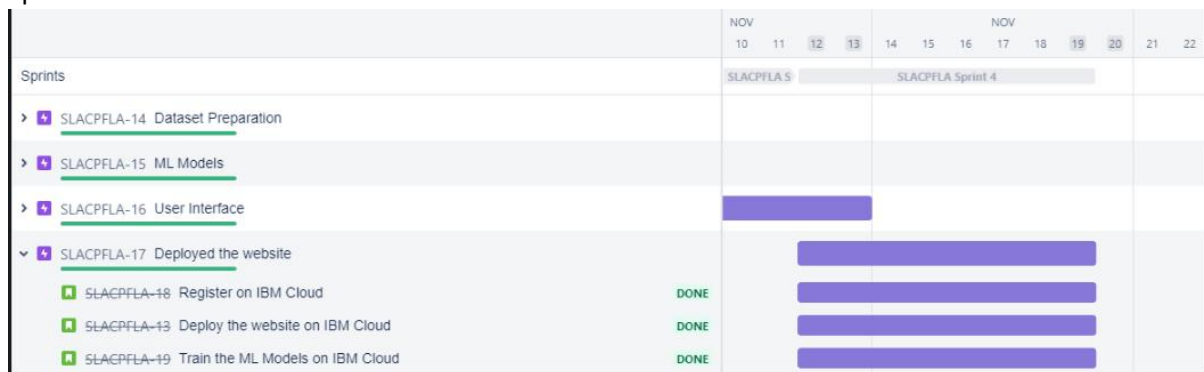
CHAPTER 7

CODING AND SOLUTIONING

7.1   LOAN ELIGIBILITY PREDICTION

K-NEIGHBORS CLASSIFIER:

The k neighbours classification or KNN classification is a non-parametric algorithm where the model arrives at a distribution based on an assumed probability distribution. Here all parameters of the dataset and several latent parameters are used.

DECISION TREE:

A decision support tool known as a decision tree employs a tree-like model to represent options and their potential outcomes, including utility, resource costs, and chance event outcomes.

SUPPORT VECTOR CLASSFIER:

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning. The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.

XGBOOST MODEL:

XGBoost is a supervised algorithm for classification and regression. It is an ensembling technique that is an optimized gradient boosting algorithm through parallel processing, tree pruning, handling missing values, etc to avoid overfitting & bias.

7.2   CONTACT

The user can contact the bank incase of any queries through this contact feature. They should fill their names, emailed, mobile number and the message they want to send.

CHAPTER 8
TESTING

8.1   TEST CASES

Test Scenario	Pre-Requisite	Steps To Execute	Test Data	Expected Result	Actual Result	Status
Verify the user is able to see the	URL of the website	1.Enter URL and click go		Form of the loan approval	Working as	Pass
Verify the UI elements are		1. Enter URL and click enter.		Application should be able to :	Working as	Pass
Verify user is able to correctly		1.Enter URL and click enter.	Name: Rohith	User should able to:	Working as	Pass
Verify user is able to obtain the		1.Enter URL and click enter.	Name: Rohith	Website should display "Hi Rohith,	Working as	Pass

8.2   USER ACCEPTANCE TESTING

Defect Analysis

Resolution	Severity 1	Severity 2	Severity 3	Severity 4	Subtotal
By Design	2	1	1	0	4
Duplicate	0	0	0	0	0
External	0	1	0	0	1
Fixed	2	2	1	0	5
Not Reproduced	0	0	0	0	0
Skipped	0	0	0	0	0
Won't Fix	0	0	0	0	0
Totals	4	4	2	0	10
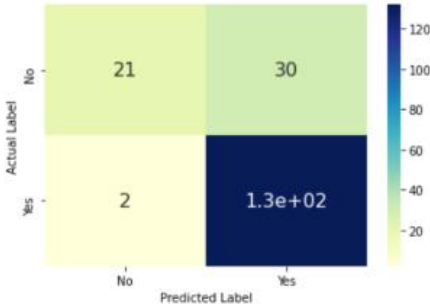
TestCase Analysis

Section	Total Cases	Not Tested	Fail	Pass
Print Engine	6	0	0	6
Client Application	6	0	0	6
Security	1	0	0	1
Outsource Shipping	2	0	0	2
Exception Reporting	1	0	0	1
Final Report Output	2	0	0	2
Version Control	1	0	0	1

CHAPTER 9

RESULTS

9.1 PERFORMANCE METRICS

S.No.	Parameter	Values	Screenshot
1.	Metrics	**Classification Model -** **Support Vector Classifier**	**Confusion Matrix:**

```
import seaborn as sn

data = confusion_matrix(y_test, y_pred)
df_cm = pd.DataFrame(data, columns=['No', 'Yes'], index = ['No', 'Yes'])

ax = sn.heatmap(df_cm, annot=True, cmap="YlGnBu", annot_kws={"size": 16})
ax.set(ylabel="Actual Label", xlabel="Predicted Label")
```

`[Text(33.0, 0.5, 'Actual Label'), Text(0.5, 15.0, 'Predicted Label')]`

Accuracy Score:

```
print('Accuracy: ', accuracy_score(y_test, y_pred))
```

```
Accuracy:  0.827027027027027
```

Classification Report:

```
----------------------------------------------------------------
Support Vector Classifier
Accuracy Score :  0.827027027027027
              precision    recall  f1-score   support

           0       0.91      0.41      0.57        51
           1       0.81      0.99      0.89       134

    accuracy                           0.83       185
   macro avg       0.86      0.70      0.73       185
weighted avg       0.84      0.83      0.80       185

----------------------------------------------------------------
```

| 2. | Tune the Model | Hyperparameter Tuning -

Validation Method - **K Fold Cross Validation** | **Hyperparameter Tuning -**

```python
from sklearn.model_selection import RandomizedSearchCV

svm_tun = RandomizedSearchCV(svc, svm_params, scoring='roc_auc', cv=5)
svm_tun.fit(X_train, y_train)

print(svm_tun.best_params_)

{'kernel': 'rbf', 'gamma': 'scale', 'C': 1}
```

```python
model.get_params()
```

```python
{'C': 1.0,
 'break_ties': False,
 'cache_size': 200,
 'class_weight': None,
 'coef0': 0.0,
 'decision_function_shape': 'ovr',
 'degree': 3,
 'gamma': 'scale',
 'kernel': 'rbf',
 'max_iter': -1,
 'probability': False,
 'random_state': None,
 'shrinking': True,
 'tol': 0.001,
 'verbose': False}
```

Validation:

```python
kfold = model_selection.KFold(
    n_splits=5,
    shuffle=True
```|

| | | | ```python
cv_results = model_selection.cross_validate(
    model, X_train, y_train,
    cv=kfold,
    scoring=scoring
)
``` |

CHAPTER 10

ADVANTAGES AND DISADVANTAGES

 Keep Control of the Company A bank loans money to a business based on the value of the business and its perceived ability to service the loan by making payments on time and in full. Unlike with equity finance where the business issues shares, banks do not take any ownership position in businesses. Bank personnel also do not get involved in any aspect of running a business to which a bank grants a loan. This means you ghet to retain full management and control of your business with no external interference. Bank Loan is Temporary Once a business borrower has paid off a loan, there is no more obligation to or involvement with the bank lender unless the borrower wishes to take out a subsequent loan. Compare this with equity finance, where the company may be paying out dividends to shareholders for as along as the business exists.Interest is Tax Deductible The interest on business bank loans is tax-deductible. In addition, especially with fixed-rate loans, in which the interest rate does not change during the course of a loan, loan servicing payments remain the same throughout the life of the loan. This makes it easy for businesses to budget and plan for monthly loan payments. Even if the loan is an adjustable-rate loan, business owners can use a simple spreadsheet to compute future payments in the event of a change in rates

Disadvantages:

Tough to Qualify One of the greatest disadvantages to bank loans is that they are very difficult to obtain unless a small business has a substantial track record or valuable collateral such as real estate. Banks are careful to lend only to businesses that can clearly repay their loans, and they also make sure thatthey are able to cover losses in the event of default. Business borrowers can be required to provide personal guarantees, which means the borrower's personal assets can be seized in the event the business fails and is unable to repay all or part of a loan. High Interest Rates Interest rates for small-business loans from banks can be quite high, and the amount of bank funding for which a business qualifies is often not sufficient to completely meet its needs. The high interest rate for the funding a business does receive often stunts its expansion, because the business needs to not only service the loan but also deal with additional funding to cover funds not provided by the bank. Loans guaranteed by the U.S. Small Business Administration offer better terms than other loans, but the requirements to qualify for these subsidized bank loans are very strict.

CHAPTER 11

CONCLUSION

This application is working properly and meeting to all Banker requirements. This component can be easily plugged in many other systems. It works correctly and fulfils all requirements of bankers and can be connected to many other systems. There were multiple malfunctions in the computers, content errors and fixing of weight in computerized prediction systems. In the near term, the banking software could be more reliable, accurate, and dynamic in nature and can be fit in with an automated processing unit. There have been numbers cases of computer glitches, errors in content and most important weight of features is fixed in automated prediction system more secure, reliable and dynamic weight adjustment. The system is trained on old training dataset in future software can be made such that new testing date should also take part in training data after some fix time. Machine learning helps to understand the factors which affect the specific outcomes most. Other models like neutral network and discriminate analysis can be used individually or combined for enhancing reliability and accuracy prediction.

CHAPTER 12

FUTURE SCOPE

The system is trained on old training dataset in future software can be made such that new testing date should also take part in training data after some fix time. Machine learning helps to understand the factors which affect the specific outcomes most. Other models like neutral network and discriminate analysis can be used individually or combined for enhancing reliability and accuracy prediction

In order to analyse the risk associated for the bank, credit evaluation largely involves gathering information about the customer and examining the project's technical, financial, and economic viability and this process developed a lot.

CHAPTER 13

APPENDIX

Github link :https://github.com/IBM-EPBL/IBM-Project-30305-1660143809