RMK ENGINEERING COLLEGE





(An Autonomous Institution)

R.S.M. Nagar, Kavaraipettai-601 206, Gummidipoondi Taluk, Thiruvallur District.

PROJECT

Smart Lender - Applicant Credibility Prediction for Loan Approval

DONE BY

TEAM ID: PNT2022TMID15600

ALLEN DEVADHANAM (111719104009)

DEVA B (111719104040)

DHAMODHARAN (111719104042)

BHARATH KUMAR C (111719104024)

BHARATH S (111719104021)

CONTENTS

1. INTRODUCTION

- a. Project Overview
- b. Purpose

2. LITERATURE SURVEY

- a. Existing problem
- b. References
- c. Problem Statement Definition

3. IDEATION & PROPOSED SOLUTION

- a. Empathy Map Canvas
- b. Ideation & Brainstorming
- c. Proposed Solution
- d. Problem Solution fit

4. REQUIREMENT ANALYSIS

- a. Functional requirement
- b. Non-Functional requirements

5. PROJECT DESIGN

- a. Data Flow Diagrams
- b. Solution & Technical Architecture

c. User Stories

6. PROJECT PLANNING & SCHEDULING

- a. Sprint Planning & Estimation
- b. Sprint Delivery Schedule
- c. Reports from JIRA

7. CODING & SOLUTIONING

- **a**. Feature 1
- b. Feature 2
- c. Database Schema (if Applicable)
- 8. TESTING
 - a. Test Cases
 - b. User Acceptance Testing
- 9. **RESULTS**
 - a. Performance Metrics
- 10. ADVANTAGES & DISADVANTAGES
- 11. CONCLUSION
- 12. FUTURE SCOPE
- 13. APPENDIX

Source Code

GitHub & Project Demo Link

1. INTRODUCTION

1.1 Project Overview

Despite the fact that our banking system has many products to sell, the main source of income for a bank is its credit line. So, they can earn from interest on the loans they credit. Commercial

loans have always been a big part of the banking industry, and lenders are always aiming to reduce their credit risk. Nowadays in the market economy banks play a very crucial role. The profit or loss of a bank is largely influenced by loans, i.e., whether the customers repay the loans or default on them. The banks need to decide whether he/she is a good(non-defaulter) or bad(defaulter) before giving the loans to the borrowers. Among the most important problems to be addressed in commercial loan lending is the borrowers' creditworthiness. The credit risk is defined as the likelihood that borrowers will fail to meet their loan obligations. To predict whether the borrower will be good or bad is a very difficult task for any bank or organization. The banking system uses a manual process for checking whether a borrower is a defaulter or not. The system predicts on the basis of the model that has been trained using machine learning algorithms. We have even compared the accuracy of different machine learning algorithms. We got a percentage of accuracy ranging from 75-85% but the best accuracy we got was from random forest algorithm i.e., 88.70% The system includes a user interface web application where the user can enter the details required for the model to predict. The drawback of this model is that it takes into consideration many attributes but in real life sometimes the loan application can also be approved on a single strong attribute, which will not be possible using this system.

1.2 Purpose

Data Science is being used in the Finance Industry reduces a lot more effort and time. Data Science is a field that is used for many finance areas such as algorithmic trading, fraud detection, customer management, risk analytics and many more. It is through the use of data science for finance that firms are able to have a better bonding with the consumers by knowing their choices, which in turn leads to an increase in their sales, thus increasing their profit margin. It also helps to identify risks and fraud and protect the firm. Thus, a data scientist is the most important asset to a firm without which an organization cannot simply perform.

2. LITERATURE SURVEY

2.1 Existing problem

The core business part of every bank is the distribution of loans. The main objective of the banking sector is to give their assets in safe hands. But the banks or the financial companies take a very long time for the verification and validation process and even after going through such a regress process there is no surety that whether the applicant chosen is deserving or not. To solve this problem, we have developed a system in which we can predict whether the applicant chosen will be a deserving applicant for approving the loan or not

2.2 References

[1]Golak Bihari Rath. Debasish Das. BiswaRanja n Acharya. ,"Modern Approach for Loan Sanctioning in Banks Using Machine Learning ",As the needs of people are increasing, the demand for loans in banks is also frequently getting higher every day.

[2]Anuja Kadam. Pragati Namde. Sonal Shirke. Siddhesh Nandgaonkar. Dr.D.R. Ingle.,"Loan Credibility Prediction System using Data Mining Techniques",As we know that now - a -days there is a rapid growth in the banking sector, resulting in lots of people applying for bank loans. Finding out the applicant to whom the loan will be approved is a difficult process.

[3]Kshitiz Gautam. Arun Pratap Singh. Keshav Tyagi. Mr. Suresh Kumar. ,"Loan Prediction using Decision Tree and Random Forest",In India, the number of people or organizations applying for loans increases every year.

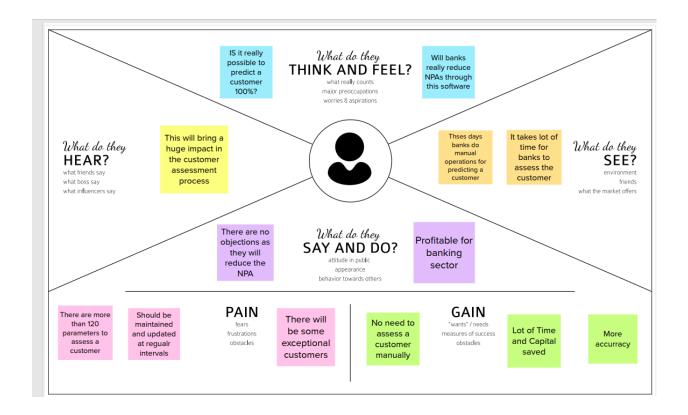
2.3 Problem Statement Definition

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.

3. IDEATION & PROPOSED SOLUTION

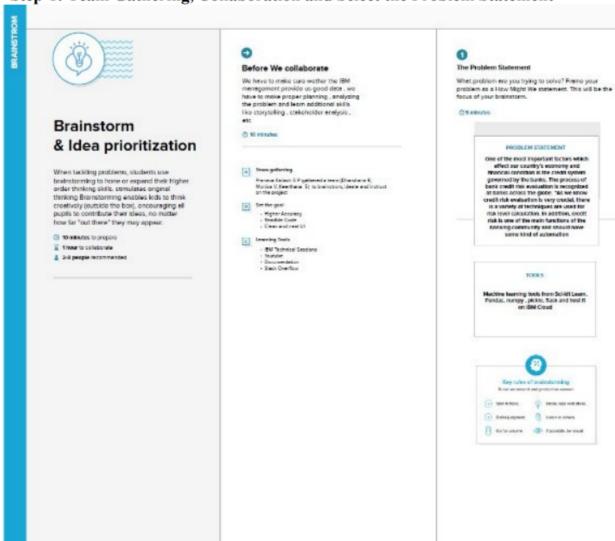
3.1 Empathy Map Canvas



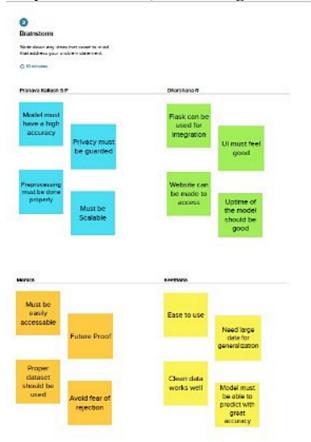
3.2 Ideation & Brainstorming

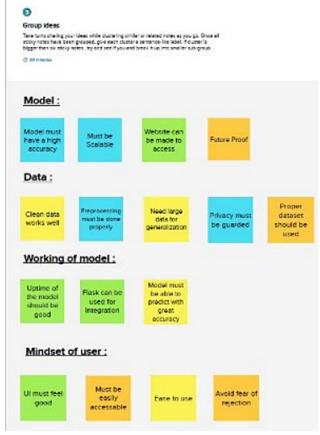
Brainstorming provides a free and open environment that encourages everyone within a team to participate in the creative thinking process that leads to problem solving. Prioritizing volume over value, out-of-the-box ideas are welcome and built upon, and all participants are encouraged to collaborate, helping each other develop a rich amount of creative solutions.

Step-1: Team Gathering, Collaboration and Select the Problem Statement



Step-2: Brainstorm, Idea Listing and Grouping





Step-3: Idea Prioritization



Prioritize

The key is not to prioritize what's on your schedule, but to schedule your priorities

② 20 minutes



3.3 Proposed Solution

| S.No. | Parameter | Description |
|-------|--|--|
| 1. | Problem Statement (Problem to besolved) | Not enough verification process Fake Documentations Miscalculations of candidatescapacity |
| 2. | Idea / Solution description | People's Need Increased, so demand for loans in banks also increased. Loan approval is a time consumption process, inorder to reduce the time consumption we are going to create a chatbot with audio facility. The loan approval can be predictedusing any of the following machine learning algorithms like Linear Regression, Decision Tree, RandomForest. Since the prediction is based on amachine learning algorithm, the loan approval process can beaccurate |
| 3. | Novelty / Uniqueness | Web application platform that will select the appropriate banks for the users eligibility. And banks can also find the rightusers based on their requirements. |
| 4. | Social Impact / Customer Satisfaction | Our platform will be helpful not only for thebanks 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.

If the loan approval process time

consumption is less, more loan prediction can be easily done, it will help for the bankto raise their business growth.

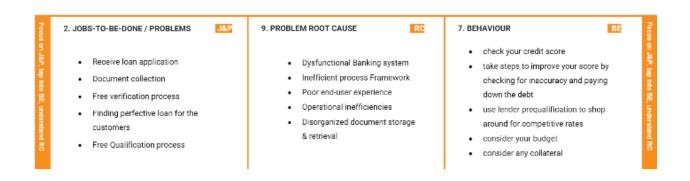
5.

Business Model (Revenue Model)

| 6. | Scalability of the Solution | Every business sector and normal |
|----|-----------------------------|----------------------------------|
| | | peopleare |
| | | able to use this system. |

3.4 Problem Solution fit





4. REQUIREMENT ANALYSIS

4.1 Functional requirement

| S.No. | Parameter | Description |
|-------|---|--|
| 1. | Problem Statement (Problem to besolved) | Not enough verification process Fake Documentations Miscalculations of candidatescapacity |
| 2. | Idea / Solution description | People's Need Increased, so demand for loans in banks also increased. Loan approval is a time consumption process, inorder to reduce the time consumption we are going to create a chatbot with audio facility. The loan approval can be predictedusing any of the following machine learning algorithms like Linear Regression, Decision Tree, RandomForest. Since the prediction is based on amachine learning algorithm, the loan approval process can beaccurate |

| 3. | Novelty / Uniqueness | Web application platform that will select the appropriate banks for the users eligibility. And banks can also find the rightusers based on their requirements. |
|----|----------------------|---|
| | | |

| 4. | Social Impact / Customer Satisfaction | Our platform will be helpful not only for thebanks 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 bankto raise their business growth. |
| 6. | Scalability of the Solution | Every business sector and normal peopleare able to use this system. |

4.2 Non-Functional requirements

| S.No. | Parameter | Description |
|-------|---|---|
| 1. | Problem Statement (Problem to besolved) | Not enough verification process Fake Documentations Miscalculations of candidatescapacity |

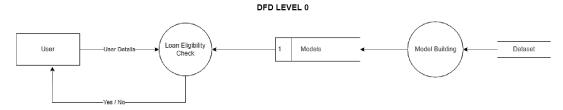
| 2. | Idea / Solution description | People's Need Increased, so demand for loans in banks also increased. Loan approval is a time consumption process, inorder to reduce the time consumption we are going to create a chatbot with audio facility. The loan approval can be predictedusing any of the following machine learning algorithms like Linear Regression, Decision Tree, RandomForest. Since the prediction is based on amachine learning algorithm, the loan approval process can beaccurate |
|----|--|--|
| 3. | Novelty / Uniqueness | Web application platform that will select the appropriate banks for the users eligibility. And banks can also find the rightusers based on their requirements. |
| | | |
| 4. | Social Impact / Customer Satisfaction | Our platform will be helpful not only for thebanks to identify the eligible people but also benefit lenders to find next bank |

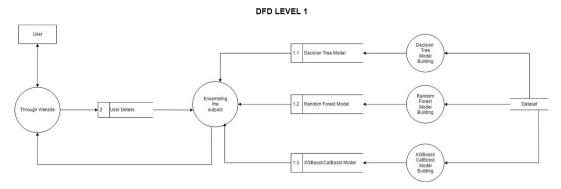
| 4. | Social Impact / Customer Satisfaction | Our platform will be helpful not only for thebanks to identify the eligible people but also benefit lenders to find next bank which provides loan to user. It will save time and effort of both banks and customers. |
|----|--|--|
| 5. | Business Model (Revenue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

5. PROJECT DESIGN

5.1 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



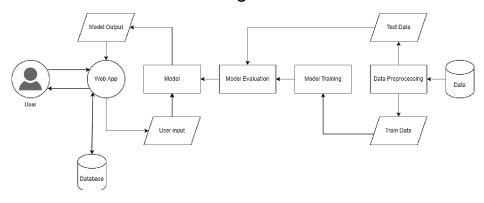


5.2 Solution & Technical Architecture

Solution architecture is a complex process – with many sub-processes – that bridges the gap between business problems and technology solutions. Its goals are to

- Find the best tech solution to solve existing business problems.
- Describe the structure, characteristics, behaviour, and other aspects of the software to project stakeholders.
- Define features, development phases, and solution requirements.
- Provide specifications according to which the solution is defined, managed, and delivered.

Solution Architecture Diagram:



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. | Checkfor 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 |

6.PROJECT PLANNING & SCHEDULING

6.1 Sprint Planning & Estimation

| Sprint | Functional Requirement (Epic) | User Story Number | User Story / Task | Story Points | Priority | Team Members |
|----------|------------------------------------|----------------------|---|--------------|----------|-----------------------------|
| Sprint-1 | Pre – processing | USN-1 | Need for the data to be clean enough for Model Prediction | 5 | High | Allen Deva |
| Sprint-1 | Web UI | USN-2 | As a user, I would need a place to enter my data to predict my results | 3 | High | Bharath kumar |
| Sprint-2 | Model Creation | USN-3 | As the data is clean now, the data can be used to Train and Evaluate the results | 4 | Medium | Dhamodha ran, Bharath |
| Sprint-3 | Integration of Model and Web UI | USN-4 | Using Flask, now we can integrate the Model with the input given by the user | 2 | Medium | Allen Bharath kumar |
| Sprint-4 | Deployment in the Cloud | USN-5 | After Complete integration, now the model should be deployed in IBM Cloud and put for use | 1 | Medium | Deva,dhamo , Bharath S |

6.2 Sprint Delivery Schedule

| Sprint | Total Story Points | Duration | Sprint Start Date | Sprint End Date (Planned) | Story Points Completed (as on Planned End Date) | Sprint Release Date (Actual) |
|----------|-----------------------|----------|-------------------|------------------------------|---|---------------------------------|
| Sprint-1 | 8 | 6 Days | 24 Oct 2022 | 29 Oct 2022 | 8 | 29 Oct 2022 |
| Sprint-2 | 4 | 6 Days | 31 Oct 2022 | 05 Nov 2022 | 4 | 05 Nov 2022 |
| Sprint-3 | 2 | 6 Days | 07 Nov 2022 | 12 Nov 2022 | 2 | 12 Nov 2022 |
| Sprint-4 | 1 | 6 Days | 14 Nov 2022 | 19 Nov 2022 | 1 | 19 Nov 2022 |

6.3 Reports from JIRA

Velocity:
Imagine we have a 10-day sprint duration, and the velocity of the team is 20 (points per sprint). Let's calculate the team's average velocity (AV) per iteration unit (story points per day)

$$AV = \frac{sprint\ duration}{velocity} = \frac{20}{10} = 2$$

$$AV = Sprint duration / Velocity = 15/6 = 2.5$$

Burndown Chart:

A burn down chart is a graphical representation of work left to do versus time. It is often used in agile software development methodologies such as Scrum. However, burn down charts can be applied to any project containing measurable progress over time.

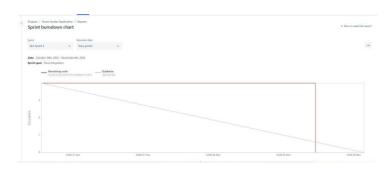
SPRINT-1



SPRINT-2



SPRINT-3

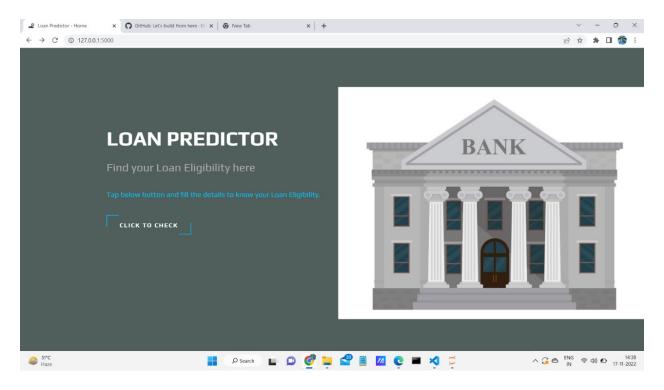


SPRINT-4



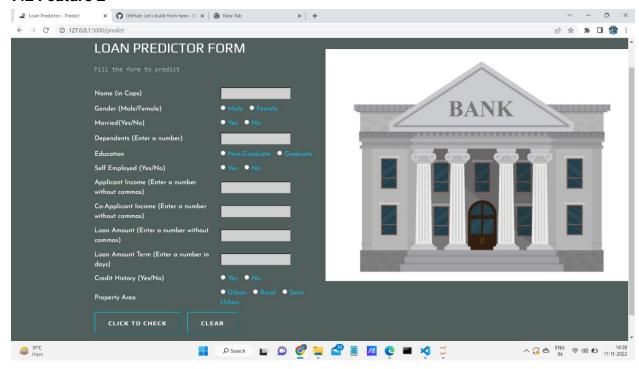
7. CODING & SOLUTIONING (Explain the features added in the project along with code)

7.1 Feature 1

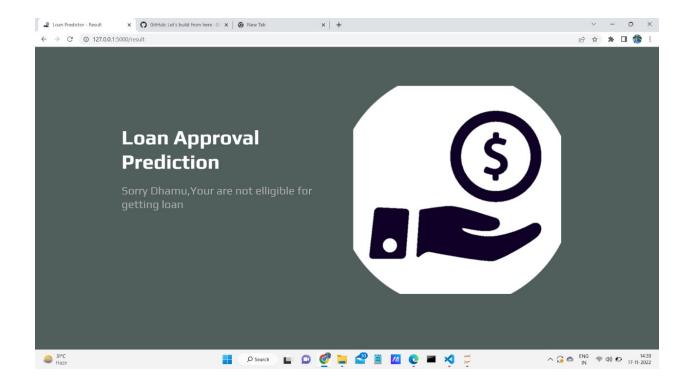


```
<body> <!--background="../STATIC/IMAGES/Lending.jpg"-->
   <main>
      <div class="container">
         <div class="row">
             <div class="col-lg-6" style="align-items:center">
                <h1>LOAN PREDICTOR</h1>
                   <h3>Find your Loan Eligibility here</h3><br>
                    <h5>Tap below button and fill the details to know your Loan Eligibility.</h5><br>
                   <div class="portfolio_btn">
                       <a href="{{ url_for('predict') }}" class="tm-nav-link primary-btn" data-hover="Loan Predictor">
                             Click to Check
                    </div>
             </div>
             <div class="col-lg-6">
                <img src="static/images/Lending_1.jpg">
             </div>
             <div class="col-lg-12"><br></div>
      </div>
      </div>
   </main>
</body>
```

7.2 Feature 2



```
<form action="/result" method="POST" enctype="'multipart/form-data">
      Name (in Caps)
      <input name='name' type="text" required="">
Gender (Male/Female)
       <input id="gender" name="gender" type="radio" value=1 required>&nbsp;Male&ensp;&ensp;<input id="gender" name="gender" type="radio" val</pre>
       Married(Yes/No)
       <input id="married" name="married" type="radio" value=1 required>&nbsp;Yes&ensp;&ensp;<input id="married" name="married" type="radio"</p>
Dependents (Enter a number)
       <input name='dep' type="number" min="0" step='1' placeholder="" required="">
      Education
       <input id='edu' name='edu' type="radio" value=1 required>&nbsp;Non-Graduate&ensp;&ensp;<input id="edu" name="edu" type="radio" value=8</p>
      Self Employed (Yes/No)
      <input id="se" name="se" type="radio" value=0>&nbsp;Yes&ensp;&ensp;<input id="se" name="se" type="radio" value=0>&nbsp;No
      Applicant Income (Enter a number without commas)
      <input id='AI' name='ai' type="number" min='0' required="">
       Co-Applicant Income (Enter a number without commas)
      <input id='CAI' name='cai' type="number" min='0' required="">
Loan Amount (Enter a number without commas)
       <input id='la' name='la' type="number" min='0' required="">
       Loan Amount Term (Enter a number in days)
      <input id='lat' name='lat' type="number" min='0' step="1" required="">
      Credit History (Yes/No)
       <input id="ch" name="ch" type="radio" value=1 required>&nbsp;Yes&ensp;&ensp;<input id="ch" name="ch" type="radio" value=0>&nbsp;No
      Property Area
      <input id="pa" name="pa" type="radio" value=2 required>&nbsp;Urban&ensp;&ensp;<input id="pa" name="pa" type="radio" value=0>&nbsp;&ur
```



9. RESULTS

9.1 Performance Metrics

The accuracy score of model using Random forest classifier is 0.77(77%) (f1-score)

10. ADVANTAGES & DISADVANTAGES

Advantages:

- 1. As the Model was build using the Random Forest Algorithm it is immunity to overfitting.
- 2. More Efficient on large dataset.
- 3. providing user friendly web interface.
- 4. Low consumption of time, which lead to decrease in NPA(Non- Profitable Assets).

Disadvantages:

The drawback of this model is that it takes into consideration many attributes but in real life

sometimes the loan application can also be approved on a single strong attribute, which will not be possible using this system.

11. CONCLUSION

For the purpose of predicting the loan approval status of the applied customer, we have chosen the machine learning approach to study the bank dataset. We have applied various machine learning algorithms to decide which one will be the best for applying on the dataset to get the result with the highest accuracy. Following this approach, we found that apart from the logistic regression, the rest of the algorithms performed satisfactory in terms of giving out the accuracy. The accuracy range of the rest of the algorithms were from 75% to 85%. Whereas the logistic regression gave us the best possible accuracy (88.70%) after the comparative study of all the algorithms. We also determined the most important features that influence the loan approval status. These most important features are then used on some selected algorithms and their performance accuracy is compared with the instance of using all the features. This model can help the banks in figuring out which factors are important for the loan approval procedure. The comparative study makes us clear about which algorithm will be the best and ignores the rest, based on their accuracy.

12. FUTURE SCOPE

The future scope of the project is to use different Algorithm to train the model to handle large and any type of data value so that the model try to give correct results. The next step is to increase the accuracy of the model by preprocessing more data and train the model using the existing random forest classification algorithm.

13. APPENDIX

Source Code

```
# NOTE: you must manually set API_KEY below using information retrieved from your IBM Cloud account.
API_KEY = "W5zK1ki_mpb4VeCgGLak0aqQVf16wY4y4ZyfHw9F3501"
token_response = requests.post('https://iam.cloud.ibm.com/identity/token', data={"apikey":
    API_KEY, "grant_type": 'urn:ibm:params:oauth:grant-type:apikey'})
mltoken = token_response.json()["access_token"]
header = {'Content-Type': 'application/json', 'Authorization': 'Bearer ' + mltoken}
app = Flask(__name__)
```

```
@app.route('/')
def index():
    return render_template('home.html')
@app.route('/predict')
def predict():
    return render_template('predict.html')
@app.route('/result', methods=['GET', 'POST'])
def upload():
    if request.method == "POST":
        name= request.form.get('name')
        gender=request.form.get('gender')
        married=request.form.get('married')
        dependency=request.form.get('dep')
        education=request.form.get('edu')
        self employed=request.form.get('se')
        applicant_income=request.form.get('ai')
        coapplicant_income=request.form.get('cai')
        loan_amount=request.form.get('la')
        loan_amount_term=request.form.get('lat')
        credit_history=request.form.get('ch')
        property_area=request.form.get('pa')
        dependency=int(dependency)
        if(str(gender)=="Male"):
            gender=1
        else:
            gender=0
        if (str(married)=="Yes"):
            married=1
        else:
            married=0
        if (str(education)=="Graduate"):
            education=0
        else:
            education=1
```

GitHub:https://github.com/IBM-EPBL/IBM-Project-25357-1659960296

Project Demo: