

College of Business Administration - Information Systems Department

IS699: Information Systems Project (Fall 2022 Term)

Building a Loan Sanction Predictor
&
Embedding it in a Financial Consulting Website

Katja Crusius

Ruchi Isaac

Ansuya Patel

Professor Abraham Asher

California State University, Long Beach

Table of Contents

1. Executive Summary	4
2. Identification / Selection of Project / Opportunity Evaluation	4
3. Initiation and Planning of System Project	6
4. System Requirements	19
5. Models	20
6. System Development	30
7. Implementation	32
8. System Documentation	33
9. Application State / Developed Features / To-Do / Prototype	37
10. Lessons Learned / References	38
11. Closing Comments	39
12. Software Delivery	40
13. Application Delivery	40

Table of Figures

Fig 1: Gantt Chart	12
Fig 2: Backlog	18
Fig 3: Velocity	18
Fig 4: Sprint Plans	19
Fig 6: System Requirements	20
Fig 7: System Architecture	21
Fig 8: Class Diagrams	22
Fig 9: Use Case Diagrams	22
Fig 10: Activity Diagrams	26
Fig 11: Sequence Diagrams	28
Fig 12: State Diagram	29
Fig 13: ER Diagram	30
Fig 14: Data Dictionary	30
Fig 15: Development Approach	32
Fig 16: Website Screenshots	35
Fig 17: Result Comparison	37

1. Executive Summary

We plan to establish a financial consulting website that includes a loan sanction predictor as its main application. This loan sanction predictor will be able to identify if a customer's loan application will likely be accepted or denied based on a list of inputs determined in the process of establishing it. Customers can create an account on the website and use the application once for free, then they will be charged for every further use. In addition to that, they can look at demographics and financial tips to get an idea of how their chances are and how to improve them.

2. Identification / Selection of Project / Opportunity Evaluation

a. Background

In the area of personal finance, it is essential to maintain a good credit score. Credit card and loan applications that are declined can lead to a decrease in someone's credit score, but it is nearly impossible to accurately predict the chances of a loan application getting accepted or declined without having knowledge in that area.

As college students and young professionals transition from being in school to the workforce and to a different phase in their lives which might include buying real estate or other assets it is important that they know about their financial situation and can take advantage of it. A lot of younger people are intimidated by the concept of loan applications and credit scores even though they would be good candidates in a lot of cases, so encouraging young people to build their assets and invest would support businesses as well as the economy.

b. Strategic Planning

i. Mission

The purpose of the project is to create an accessible website that helps college students and young professionals to determine whether they should file a loan application based on certain metrics that they put into the system.

ii. Objectives

The goal is to create a fully functional website that includes a loan sanction predictor which is based on machine learning. While the software itself will be more complex coding wise, it should be accessible and easy to use which makes it important to create a user friendly website design.

iii. Competitive Strategies

The software engineers and developers on the project will work on the creation of the machine learning algorithm to determine the chances of getting a loan approved together. They will coordinate with the graphic designer and project manager to embed this software on the website and make it user friendly.

The project team is not responsible for the marketing and distribution of the website and the software and will outsource that to the university, which will market the website to students and recent alumni as well as bring it to the attention of other universities in the area.

c. Constraints

We are predicting data that may or may not reflect the actual outcome since we are not the bank that the applicant is actually going to apply to for a loan grant and factors personal to the institute of application such as budget cuts or number of loan grant quota, could provide unforeseeable results.

We are constricted to the parameters we are using in our system to derive the fate of our user's loan application.

d. Overall System Need

The final system needs to be able to predict the possibility of someone being able to repay their loan after taking into consideration various factors like their gender, marital status, dependants, credit history, educational background, the requested loan amount, co applicant income, applicant income etc. The more accurate the system the better.

e. Strategic Fit

Our target group are students as well as young professionals, which is a demographic group that is starting to build a credit score and file loan applications, but has not previously been exposed to this field. Therefore, it can be hard to determine whether a loan application has a good chance of being accepted, if certain credit scores can be considered good, and lastly how to improve chances of getting loans.

We are planning to utilize that demand by offering a solution to those problems by building a reliable application that helps users to determine whether their loan application is promising or not.

3. Initiation and Planning of System Project

a. Charter

GENERAL PROJECT INFORMATION

PROJECT NAME		PROJECT MANAGER	PROJECT SPONSOR
Loan Sanction Predictor		Katja Crusius	CSULB - College of Business Administration
EMAIL	PHONE	ORGANIZATIONAL UNIT	
katja.crusius@student.csulb.edu	562-714-2642	Project Management	

PROJECT OVERVIEW

PROBLEM OR ISSUE	In the area of personal finance, it is essential to maintain a good credit score. Credit card and loan applications that are declined can lead to a decrease in someone's credit score, but it is nearly impossible to accurately predict the chances of a loan application getting accepted or declined without having knowledge in that area.
PURPOSE OF PROJECT	The purpose of the project is to create an accessible website that helps college students and young professionals to determine whether they should file a loan application based on certain metrics that they put into the system.
BUSINESS CASE	As college students and young professionals transition from being in school to the workforce and to a different phase in their lives which might include buying real estate or other assets it is important that they know about their financial situation and can take advantage of it. A lot of younger people are intimidated by the concept of loan applications and credit scores even though they would be good candidates in a lot of cases, so encouraging young people to build their assets and invest would support businesses as well as the economy. This system is for them and everyone else who would like to use it.
GOALS / METRICS	The goal is to create a fully functional website that includes a loan sanction predictor which is based on machine learning. While the software itself will be more complex coding wise, it should be accessible and easy to use which makes it important to create a user friendly website design.

**EXPECTED
DELIVERABLES**

The deliverables are the source code and the commands to run the loan predictor, as well as the website with text explanations and graphics which will be created for the purpose of this project.

PROJECT SCOPE**WITHIN
SCOPE**

The software engineers and developers on the project will work on the creation of the machine learning algorithm to determine the chances of getting a loan approved together. They will coordinate with the graphic designer and project manager to embed this software on the website and make it user friendly.

**OUTSIDE
OF SCOPE**

The project team is not responsible for the marketing and distribution of the website and the software and will outsource that to the university, which will market the website to students and recent alumni as well as bring it to the attention of other universities in the area.

TENTATIVE SCHEDULE

KEY MILESTONE	START	FINISH
Form Project Team / Preliminary Review / Scope	08/23/22	08/30/22
Finalize Project Plan / Charter / Kick Off	09/06/22	09/13/22
Define Phase	09/20/22	09/27/22
Measurement Phase	10/4/22	10/11/22
Analysis Phase	10/25/22	11/01/22
Improvement Phase	11/08/22	11/15/22
Control Phase	11/29/22	12/04/22
Project Summary Report and Close Out	12/04/22	12/07/22

RESOURCES

PROJECT TEAM	Katja Crusius- Project Manager Ruchi Isaac- Chief Engineer Ansuya Patel- Chief Product Officer
SUPPORT RESOURCES	Operations, Sales, Project Management, Engineering
SPECIAL NEEDS	None

COSTS

COST TYPE	VENDOR / LABOR NAMES	RATE	QTY	AMOUNT
Computers	Apple Inc.	\$1500.00	3	\$4,500.00
Other Hardware setup and maintenance	Spectrum	\$1000.00	3	\$3000.00
Tech Support	Spectrum	\$350.00	1	\$350.00
Developers cost	Self	\$10000.00	2	\$20,000.00
Management Costs	Self	\$4,850.00	3	\$14,550.00
Miscellaneous	Self	\$100.00	3	\$300.00
TOTAL COSTS				\$42,700.00

BENEFITS AND CUSTOMERS

PROCESS OWNERS	Katja Crusius- Project Manager Ruchi Isaac- Chief Engineer Ansuya Patel- Chief Product Officer
KEY STAKEHOLDERS	Ruchi Isaac Katja Crusius Ansuya Patel
FINAL CUSTOMER	Students and young professionals all across the world hoping to find out the fate of their loan requests.
EXPECTED BENEFITS	For people who have uncertainty in regards to their loan applications, we built the Loan Sanction Predictor system, to provide them a perspective on their applications. Every loan application takes a long amount of time, effort and money and this system helps you save all of that.

RISKS, CONSTRAINTS, AND ASSUMPTIONS

RISKS	We are predicting data that may or may not reflect the actual outcome since we are not the bank that the applicant is actually going to apply to for a loan grant and factors personal to the institute of application such as budget cuts or number of loan grant quota, could provide unforeseeable results.
CONSTRAINTS	We are constricted to the parameters we are using in our system to derive the fate of our user's loan application.
ASSUMPTIONS	One of our assumptions is that the data submitted by the people using our system is accurate and correct. We also assume that users will be able to identify their requirements with the help of our data visualization.

PREPARED BY	TITLE	DATE
Katja Crusius	Project Manager	09/13/2022

b. Statement of Work

The software engineers and developers on the project will work on the creation of the machine learning algorithm to determine the chances of getting a loan approved together. They will coordinate with the graphic designer and project manager to embed this software on the website and make it user friendly.

The project team is not responsible for the marketing and distribution of the website and the software and will outsource that to the university, which will market the website to students and recent alumni as well as bring it to the attention of other universities in the area.

c. Introduction

i. Project Overview

There are various reasons as to why some individuals are people who have a lower chance of getting a loan approved compared to individuals that are more “settled”. This necessitates loan approval checking before submitting one’s details to the bank for the actual process to begin. A simulative process could help determine some, if not all, flaws or incompletions in an individual’s application and considerably increase the success rate of their loan application approval.

The approach we chose was to find a dataset of users with their information regarding loans and run different algorithms on it to find out whether the specified user would be approved or denied for the loan they applied for.

We have not focused on a specific type of loan but on a data set of a part of the general public that have applied for loans. We also made a website to show a more professional front to any user who might be interested in our services.

ii. Recommendation

Our recommendation is to use a complex model which may take time to implement but would definitely increase the likelihood of getting more accurate results since they would have the capacity to handle a wider variety of data.

d. System Description

We build a system to find out whether an applicant will be able to repay the loan or not. If the applicant is able to repay the role, they should be granted the loan; otherwise they should not. We build the system by cleaning the data set we acquired of various applicants. The training data set shows various parameters along with the status of an applicant being able to repay the loan or not. The testing dataset has all those various factor however this time, the loan status is the target variable to be predicted. We use three different models to do this. Logistic Regression, Decision Tree and Random Forest. We also build a website which the applicant can use to access our services.

e. Feasibility Assessment**i. Economic feasibility**

This product is economically feasible as it is a simple system that takes an input of a few collectible factors in a database and projects the possibility of loan approval or denial. The functionality is simple and online hence, so is the economic feasibility.

ii. Technical feasibility

Technically, we had the knowledge to construct the backend system and the front end system separately; however, we lacked the knowledge to integrate both these systems. The system also works on static data instead of a dynamic input by a user; however, it runs well with significant accuracy. We are experienced in using the pandas library thanks to the previous subject and our ongoing classes like Machine learning or Deep Learning.

iii. Operational feasibility

The backend is fully operable and so is the front end. The only thing is that they are not operating together as a unit. The business plan would be to operate the system on our back end where Loan Analysis Experts would consider the data submitted by a user and run it by our system to check whether the person would qualify or not.

Our business plan would offer one free service as a trial and then for other future services, they would be charged. Their accounts would be linked via their SSN's hence making them unique and eliminating the risk of making multiple accounts for one person.

iv. Schedule Feasibility

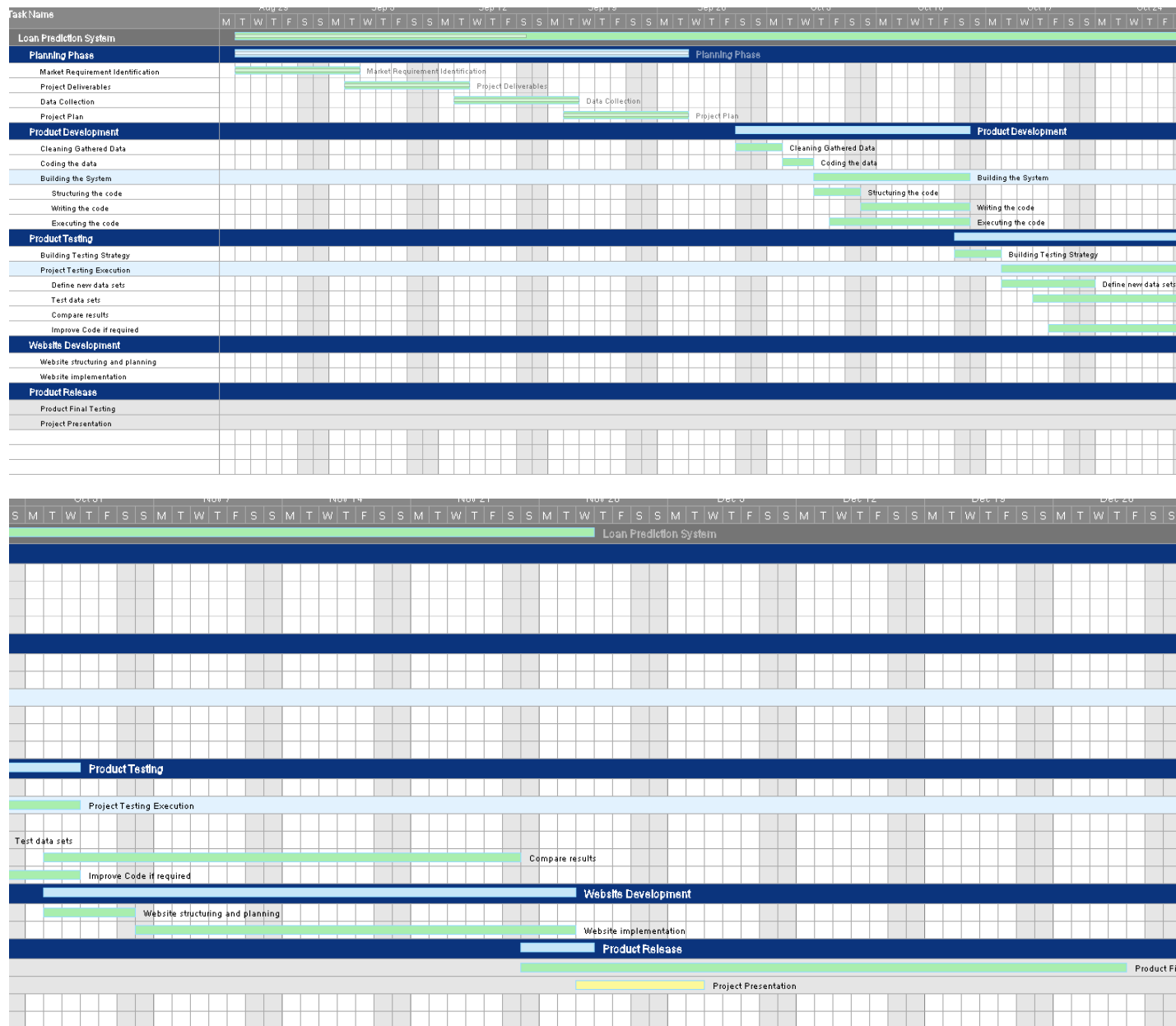
We finished the work we wanted to complete in time however if we had more time we would have liked to integrate both the systems and perhaps presented it more dynamically than working on static data sets. The system functions well even though the models are separate.

f. Project Plan

i. Implementation Schedule – Gantt Chart

Fig 1: Gantt Chart

Task Name	Status	Health	Start Date	End Date	Assigned To	Description	Duration
Loan Prediction System	In Progress	Green	08/30/22	11/30/22			92d
Planning Phase			08/30/22	09/27/22			28d
Market Requirement Identification	Complete	Green	08/30/22	09/06/22	Ruchi I	Expect minor changes as we go.	7d
Project Deliverables	Complete	Green	09/06/22	09/13/22	Katja Crusius	Could be modified further	7d
Data Collection	Complete	Green	09/13/22	09/20/22	Ansuya Patel	compiling a database	7d
Project Plan	Complete	Green	09/20/22	09/27/22	Ruchi I	defining a structured plan	7d
Product Development			10/01/22	10/15/22			14d
Cleaning Gathered Data	Complete	Green	10/01/22	10/03/22	Katja Crusius	cleaning data to fit model	3d
Coding the data	Complete	Green	10/04/22	10/05/22	Ansuya Patel	encoding data as models only understand numbers	2d
Building the System	Complete	Green	10/06/22	10/15/22	Ruchi I		10d
Structuring the code	Complete	Green	10/06/22	10/08/22	Katja Crusius	defining how the code would start and end	3d
Writing the code	Complete	Green	10/09/22	10/15/22	Ansuya Patel	implementing the code	7d
Executing the code	Complete	Green	10/07/22	10/15/22	Ruchi I	running the code to view results at every step	8d
Product Testing			10/15/22	11/02/22			18d
Building Testing Strategy	Complete	Green	10/15/22	10/17/22	Katja Crusius	since its RAD we test after ever step	3d
Project Testing Execution	Complete	Green	10/18/22	11/02/22	Ansuya Patel		15d
Define new data sets	Complete	Green	10/18/22	10/23/22	Ruchi I	split train and test dataset	6d
Test data sets	Complete	Green	10/20/22	10/29/22	Katja Crusius		10d
Compare results	Complete	Green	11/26/22	11/01/22	Ansuya Patel	check for scope of improvement in results	7d
Improve Code if required	Complete	Green	10/21/22	11/02/22	Ruchi I		12d
Website Development			11/01/22	11/29/22			28d
Website structuring and planning	Complete	Green	11/01/22	11/05/22	Katja Crusius	defining most important pages in the website	4d
Website implementation	Complete	Green	11/06/22	11/29/22	Katja Crusius	coding and implementing webste	23d
Product Release			11/27/22	11/30/22			4d
Product Final Testing	Complete	Green	11/27/22	12/29/22	All	check if all models are running perfectly	3d
Project Presentation	In Progress	Yellow	11/30/22	12/06/22	All	present the model	1d



g. Communication Plan

We communicated twice a week to check on each other's progress in our respective parts. We communicated using Email and Whatsapp. We also communicated using comments and the chat feature on google docs to give feedback to someone about their progress if required.

h. Quality Management Plan

The quality of accuracy our product has reached is 80%. In all three models the output comes out to be similar but the accuracy keeps increasing with every model we test it on. Hence it is a well working model. The website that we created is also graphically clean and user friendly.

i. Change Control Plan

We planned to construct only a model that would run and show our analysis for it; however, we also decided to include a website which could portray our business operational side of skills. This change was done since it was beneficial to depict more skills we gained from this course and combine them with our past experiences to build a whole system instead of just a model.

j. Risk Plan

i. Identify at least three Project Risks (risks in developing the system, not system risks)

1. A team member gets sick and is not able to meet.
2. The time for developing the software is underestimated.
3. Team members cannot find the right resources to support the development.

- **Assess each risk: use a risk matrix (likelihood vs. consequence)**

1. likelihood: high, consequence: serious → high risk
2. likelihood: moderate, consequence: serious → moderate risk
3. likelihood: moderate: consequence: serious → moderate risk

- **What are the consequences?**

1. Project gets delayed, and work cannot be completed
2. Deadlines cannot be met
3. Parts of the project cannot be completed

- **Provide risk mitigation plan**

1. If someone gets sick, there is a plan in place to work from home (everything is on Microsoft Teams, all documents are on Google Documents, the team is using Azure DevOps, Gitub, and Zoom)
2. To prevent issues with the time given for the semester, every single step is planned out and included in AzureDevOps, therefore the project is broken down and with smaller steps to complete the likelihood of estimating the completion time wrong is decreased
3. There are resources given by the professor, and in the case that we cannot find resources it is possible to go to the office hour or ask peers taking similar courses, further there are more resources in the book as well as on Github, so there will be a solution to possible problems appearing during the development process

ii. Identify at least three System Risks (once the system is built, what risks does it have?)

1. The system does not yield accurate results
2. Users are abusing the Free Trial function and not paying for their use
3. The input by the user is incorrect, therefore the output is incorrect
 - **Assess each risk: use a risk matrix (likelihood vs. consequence)**
 1. likelihood: high, consequence: serious
 2. likelihood: moderate, consequence: serious
 3. likelihood: high, consequence: tolerable
 - **What are the consequences?**
 1. The results are not accurate, therefore users are not satisfied
 2. The system is not being profitable, and there are no resources to maintain and improve the system
 3. Users are not satisfied, because the system is not yielding correct results
 - **Provide risk mitigation plan**
 1. We plan to run different analyses and then choose the one with the highest accuracy, after testing and adjusting the system it will have a high accuracy
 2. Users will only be able to use the Loan Sanction Predictor when creating an account on the website, for creating an account they have to put their passport or

ID number. Therefore, it is not possible to create multiple accounts and to take advantage of the free trial somehow.

3. While the loan sanction predictor would yield wrong results based on wrong inputs, this risk is tolerable as it is the user's responsibility to put in the correct data in order to get useful results. We will remind users to be as accurate as possible when putting their inputs.

k. Microsoft Teams

i. Pros / Advantages

- Microsoft Teams is a holistic tool used for communication between different members in the practical world. With a bump from 13 million users in July 2019 to 145 million users in April 2021, Microsoft teams is a leading tool for communication in organizations.
- One of the best advantages we gained by using this tool was getting an idea about the kind of communication portals we need to be comfortable with before we venture out in the technical world to perform our job duties.

ii. Cons / Disadvantages

- As a group of 3 students, making a project on a small scale, we simply didn't see the need to use it since it was introduced mid semester onwards and by then, most of our communication was set up on different platforms. It is an effective tool no doubt, however, we also don't have the kind of manpower required to even fully utilize it.

iii. Potential Usage / Lessons learned

- **If you were to start a new project, how would you use it differently?**
If we were to start a new project then we definitely would have incorporated the use of teams right from the beginning to correctly get an idea about its importance and be skilled at using its various features.

- **What features would you take advantage of?**

We could have our conversations over teams and post all our updates over Microsoft Teams and upload our documents over the cloud so that we could all have the same updates files.

iv. Overall assessment

Microsoft Teams is a powerful tool for professional communication and everyone should know its features and how to use it effectively. We have also had experience during our internships and other classes in using Microsoft Teams and everyone should be exposed to such an omnipresent tool through their courses or jobs.

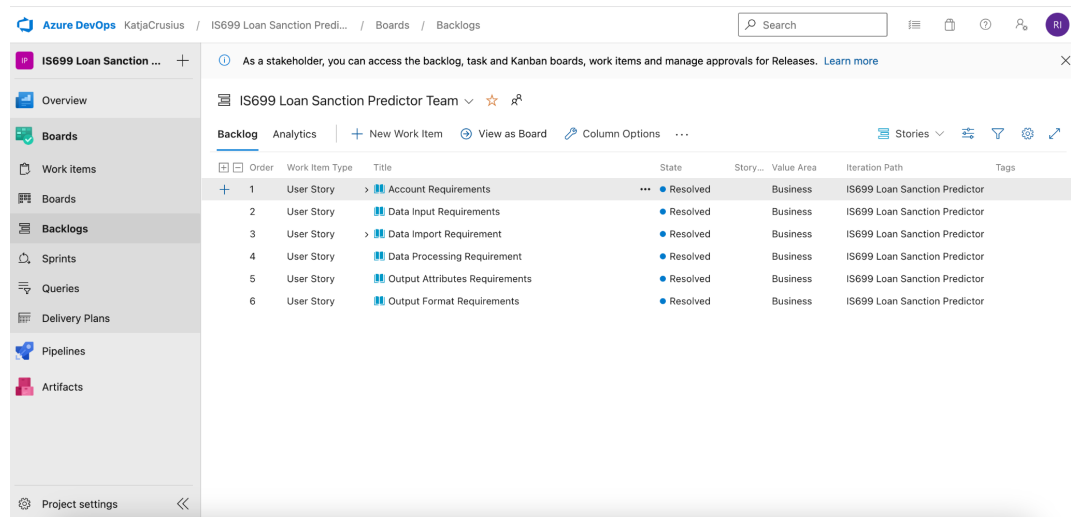
I. Azure DevOps

i. How did you team use Azure DevOps to manage your work?

Our team used Azure DevOps to track our basic tasks like input requirements, output requirements and the system requirements. We condense all of that under the epic- project requirements. Our Azure board is shared amongst the three of us and the professor.

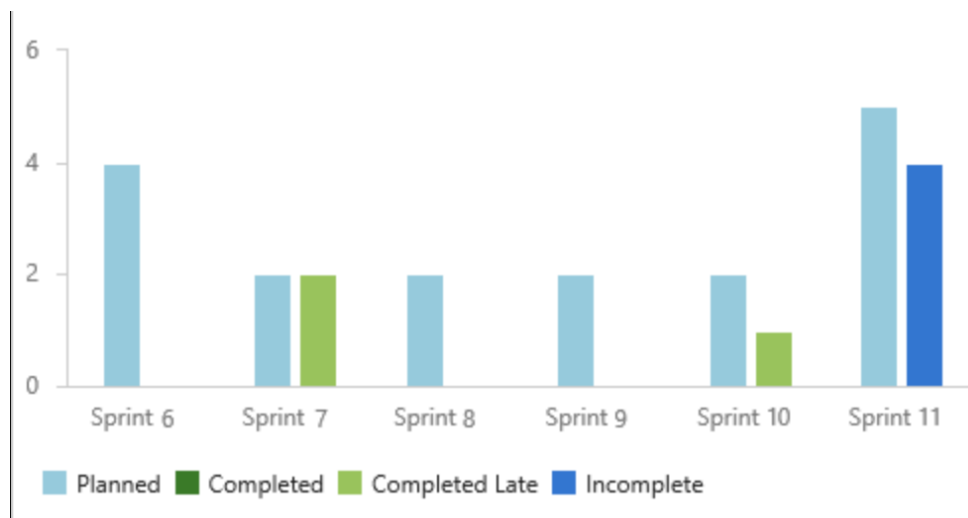
- **Backlog**

Fig 2: Backlog



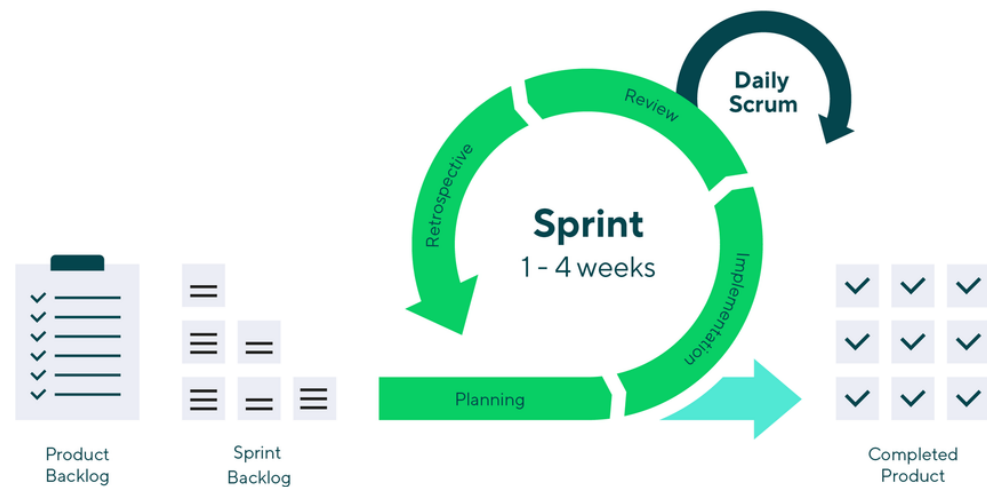
- **Velocity**

Fig 3: Velocity



- **Sprint Plans**

Fig 4: Sprint Plans



This is the sprint plan we have followed.

We would plan our tasks for the week, implement it and have a daily scrum to review the progress. This helped us keep each other's tasks in check. We would repeat this process till everyone was satisfied with the result.

ii. Pros / Advantages

- Azure DevOps helps organize the entire flow of the project into categories as required. This helps any existing or new developer to get a clear idea of the flow and understand how to carry on from where we left off.

iii. Cons / Disadvantages

- The only issue was that it was a bit hard to organize every little thing since it was also our first time working with the software.

iv. Potential Usage / Lessons learned

- **If you were to start a new project, how would you use it differently?**

From the beginning we would keep a record of a whole chain of activities and tasks that need to be completed. Checking off the state of these activities would also help make the system documentation better.

- **What features would you take advantage of?**

We would have used the features such as flexibility. Since it is cloud independent we would also have support in terms of continuous integration.

v. Overall assessment

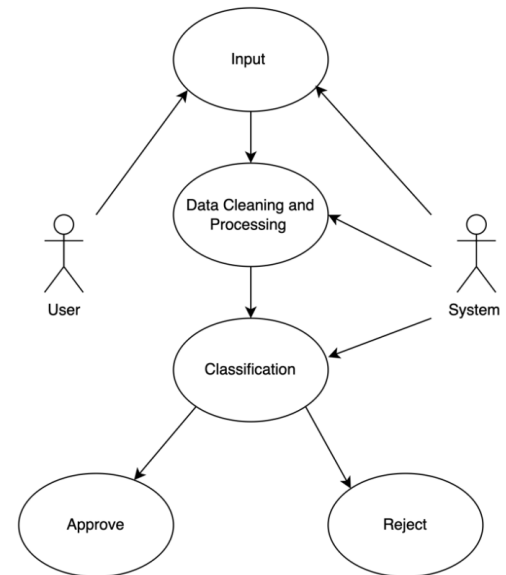
We think Azure DevOps is a good way to explore our organizational skills and learn in to use tools that are used in the industry.

4. System Requirements

a. Approach to identifying requirements

We started identifying system requirements by identifying the different types of people interacting with our system and figuring out what activities the system needs to perform. Further, we looked on websites with similar elements (such as pricing plans, or login options) what the requirements could be.

Fig 6: System Requirements



b. Identify at least six functional systems requirements

1. A user shall be able to create an account on the website.
2. A registered user shall be able to complete the login process and access their account.
3. The Loan Sanction Predictor should be able to access data from the database of data from different users.
4. The system shall generate the likelihood of the loan being approved based on the dataset.
5. The user shall be able to see the different components the system took into consideration
6. Each user shall be identified by a unique user ID to make it possible to create user statistics to improve the website

c. Identify at least three non-functional requirements

1. Reliability
2. Performance
3. Security

i. How will you address these requirements?

1. We are planning to run our system on several servers so that in case that one server is encountering difficulties, it can still operate.
2. We have tested our system with several different methods to ensure it delivers accurate results.
3. Users create an account with their passport number, we are also planning to enable Multi Factor Authentication in order to increase security.

5. Models

a. System Architecture Model

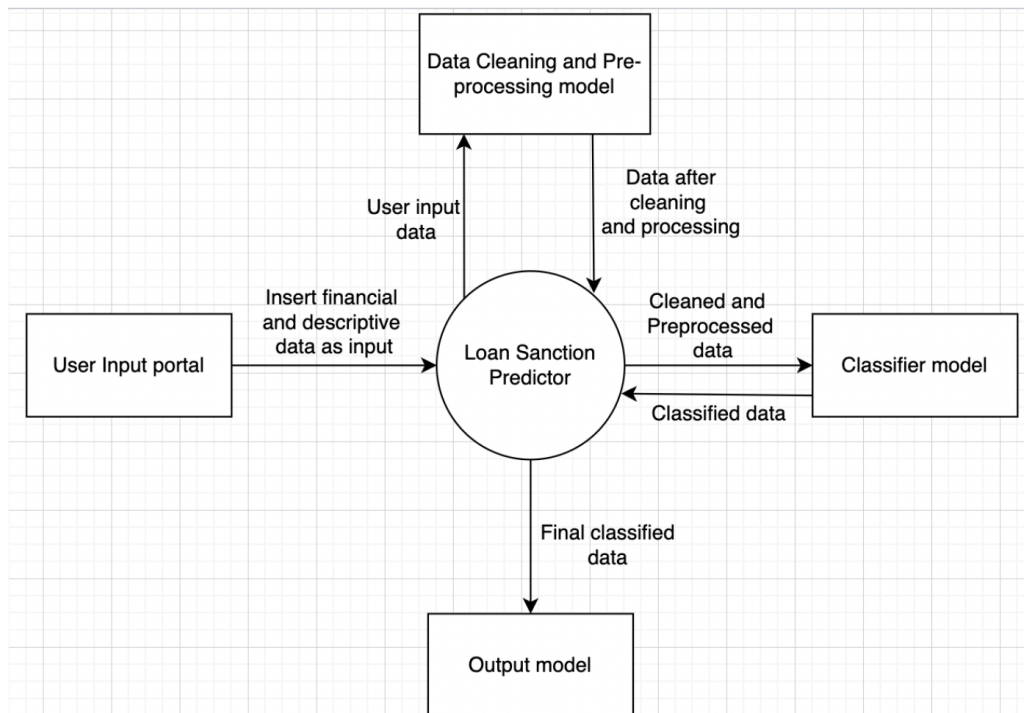


Fig 7: System Architecture

b. Class Diagram

i. Create at least three class diagrams for main objects in your system

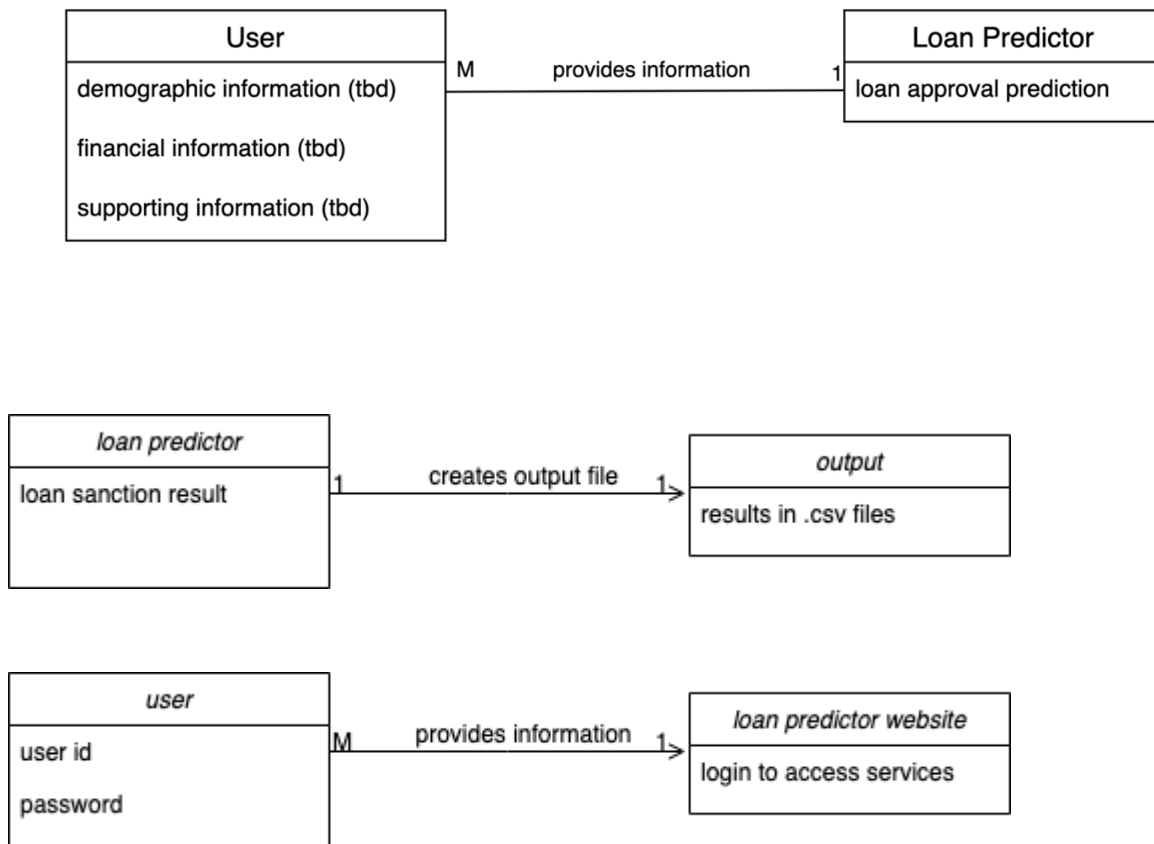
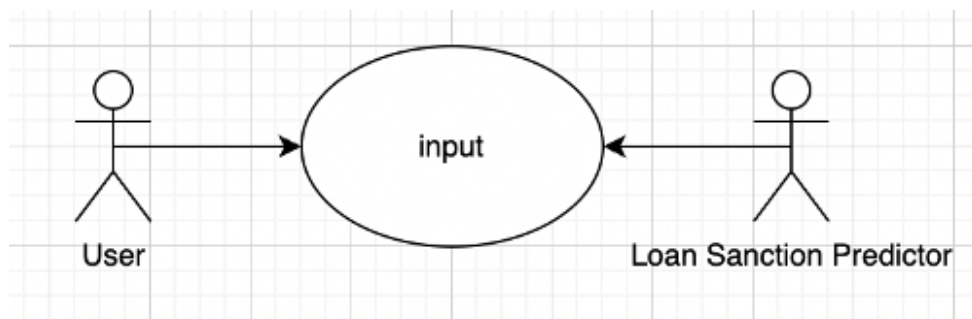


Fig 8: Class Diagrams

c. Use Case Diagrams

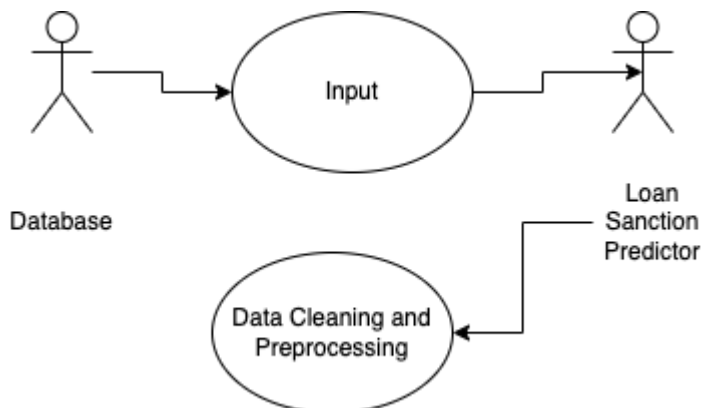
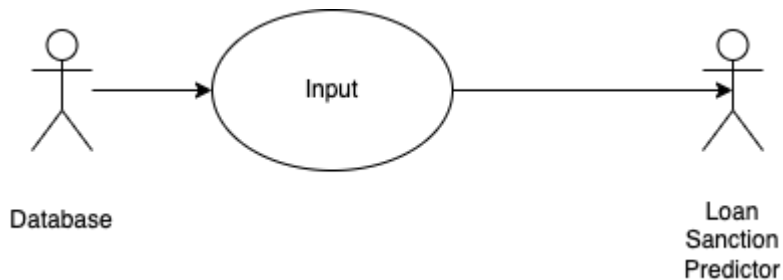
1. A user shall be able to input their info into the Loan Sanction Predictor database which accesses the data.

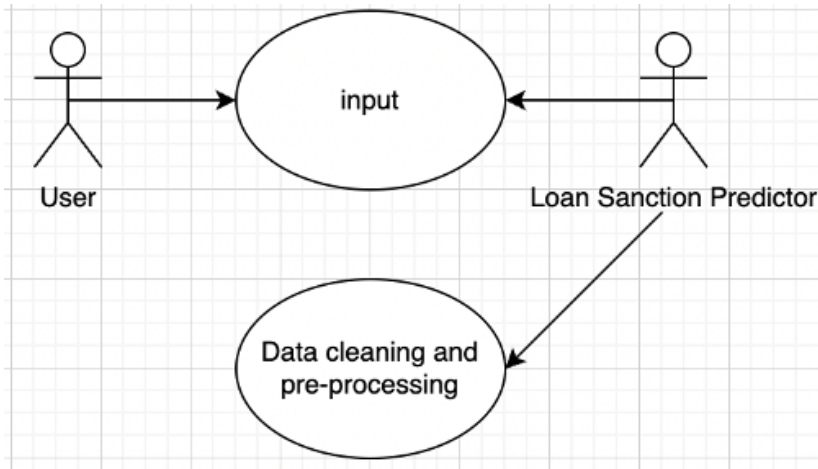
Fig 9: Use Case Diagrams



Actors	User, Loan Sanction Predictor
Description	A user who wants to check if their loan application will be approved or not, would input details of different parameters as requested by the system.
Data	A variety of financial data requested by the user.
Stimulus	The need to check the fate of a loan application before filling the application with a real bank.
Response	The system accepts this output as the data it needs to clean and process on.
Comments	The data entered by the user should be accurate, otherwise the analysis would be incorrect.

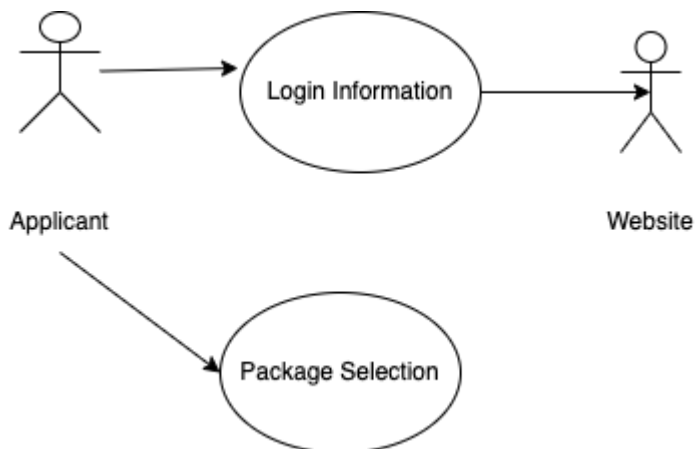
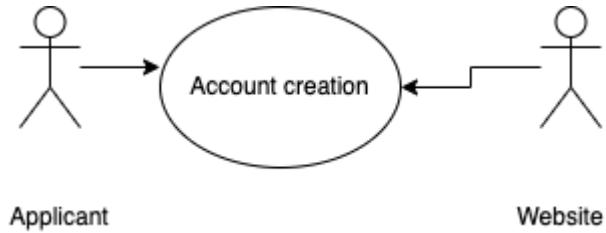
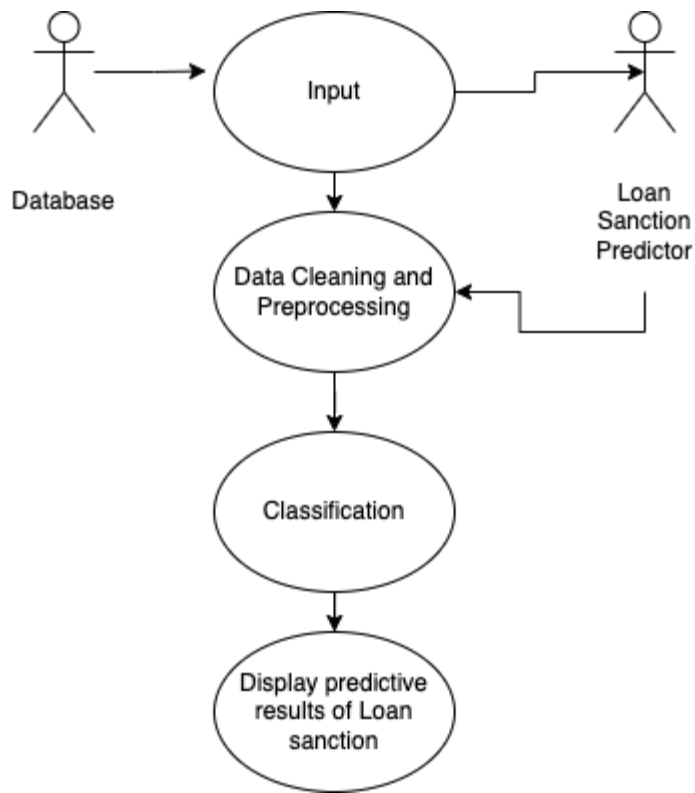
2. The system will clean the data and pre-process it.





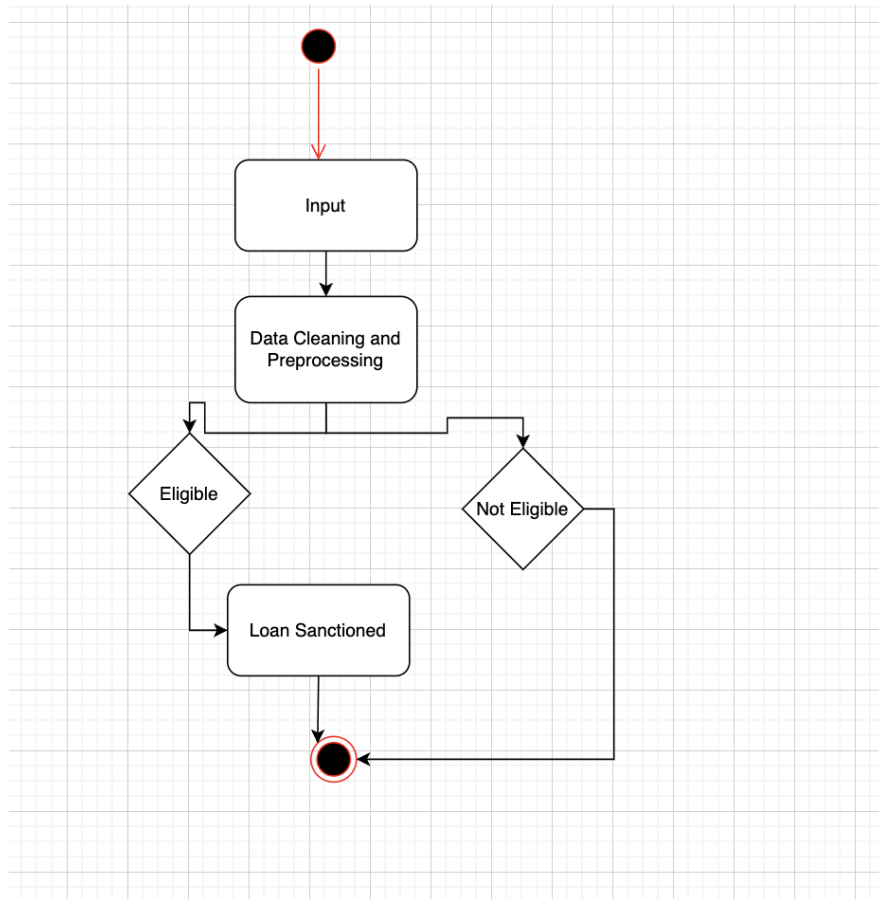
Actors	User, Loan Sanction Predictor
Description	The data entered needs to be cleaned and preprocessed before being given to the model as an input, this would help make the analysis more reliable.
Data	A variety of financial data requested by the user.
Stimulus	The data taken as input may need to be cleaned off null values or misspellings. I also needs to be preprocessed before being given as an input to the model.
Response	The data is cleaned and preprocessed and converted into a format acceptable to the model.
Comments	The data entered by the user should be accurate, otherwise the analysis would be incorrect.

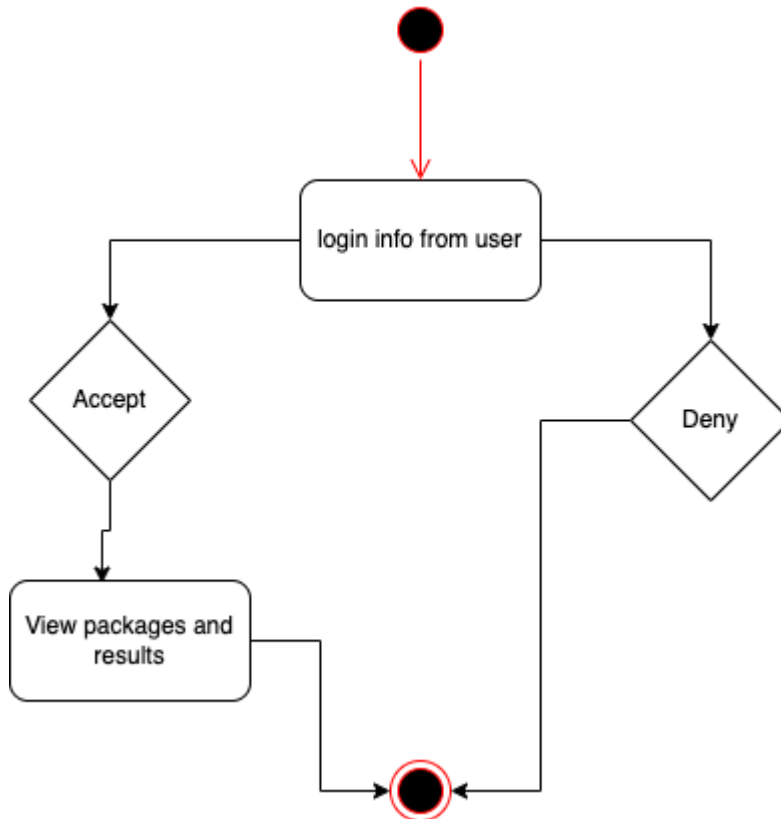
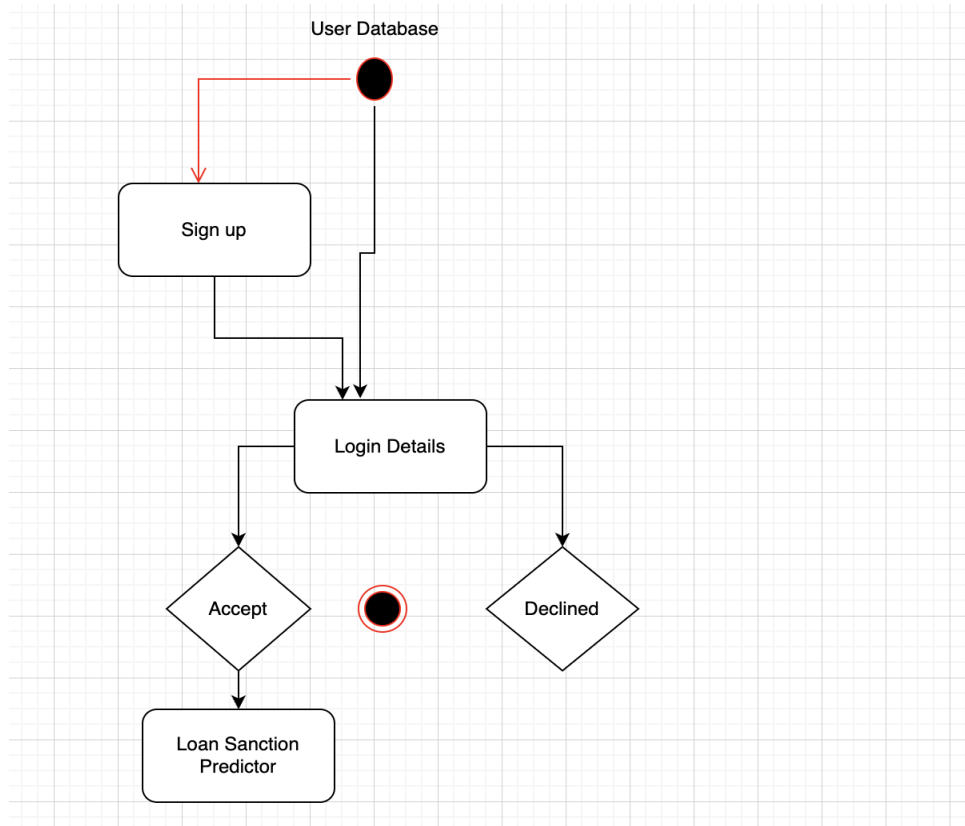
3. The System shall generate the likelihood of the loan being approved based on the information put in by the user



d. Activity Diagram

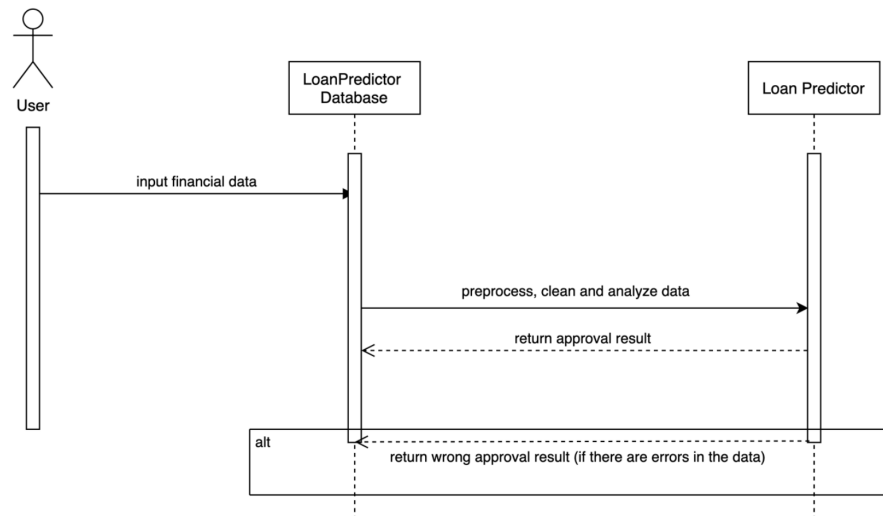
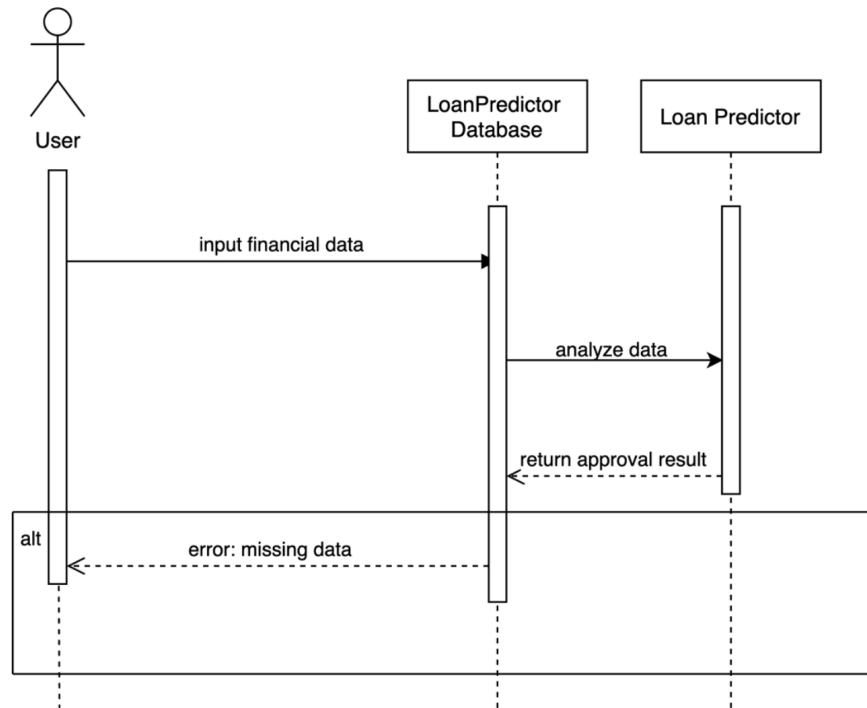
Fig 10: Activity Diagrams

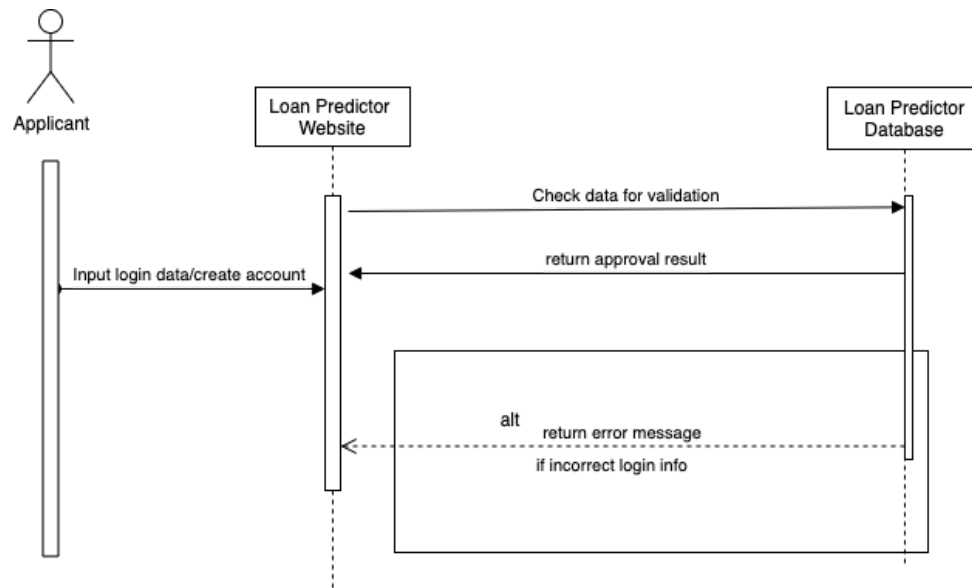




e. Sequence Diagram

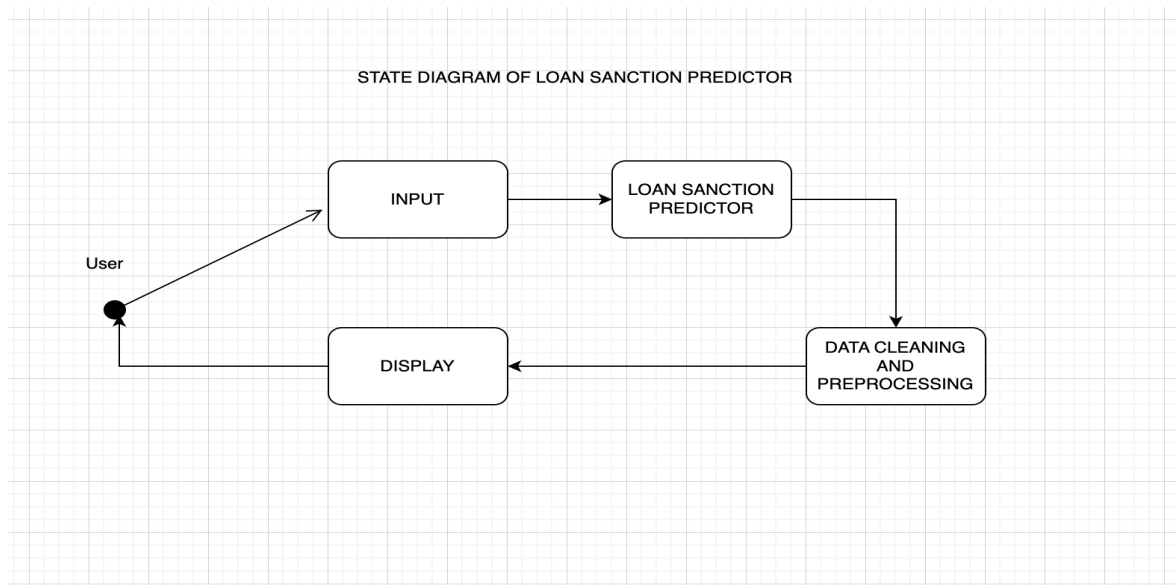
Fig 11: Sequence Diagrams





f. State Diagram

Fig 12: State Diagram

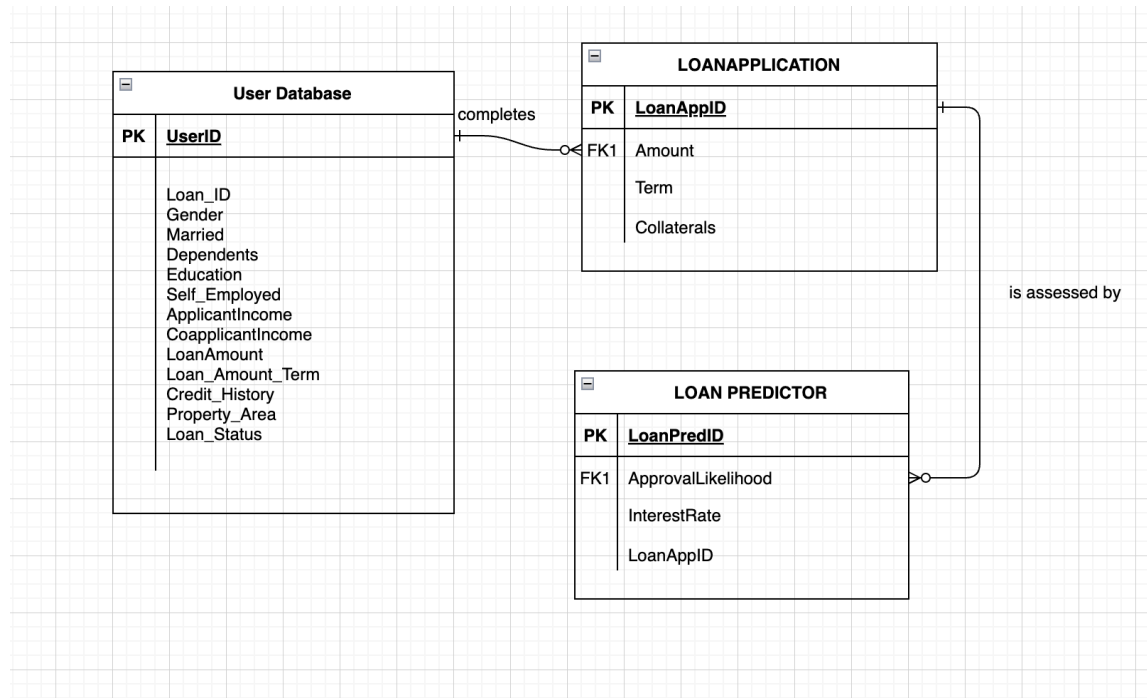


g. ER Diagram

i. Detailed ER diagram for your database tables

- Show table attributes
- Identify Primary Keys and Foreign Keys

Fig 13: ER Diagram



h. Data Dictionary

Fig 14: Data Dictionary

TABLE 1:

No.	Field Name	Data Type	Constraints	Description
1	UserID	int	PK	Unique ID by which user is identified.
2	Loan_ID	int	not null	Unique ID loan application.
3	Gender	cat	not null	Gender of the user.
4	Married	cat	not null	Marital status of the user.
5	Dependants	int	not null	Dependants of the user.
6	Education	int	not null	Educational status of the user.

7	Self_Education	cat	not null	Self educated or not.
8	ApplicantIncome	int	not null	Annual income of the applicant.
9	CoapplicantIncome	int	not null	Annual income of the co applicant.
10	LoanAmount	int	not null	Amount of loan requested.
11	Loan_Amount_Term	int	not null	Number of years of loan.
12	Credit_History	int	not null	Good or bad credit history.
13	Property_Area	cat	not null	The property area is divided into urban, semi-urban and rural.
14	Loan_Status	cat	not null	status of loan which is predicted

TABLE 2:

No.	Field Name	Data Type	Constraints	Description
1	LoanAppID	int	PK	ID to identify the processing request
2	Loan_ID	int	FK	Link to the user database
3	Term	var	not null	Term for the loan
4	Collaterals	var		Collaterals against the loan if any

6. System Development

a. Hardware requirements

Since this is a heavily software based prediction project, the only hardware we require is our basic devices like laptops and mobile phones.

b. Software requirements

We used python to code for the backend of the project. To run the codes we used Jupyter Notebook over the data science platform Anaconda. We used wix.com to build our website. We used .csv files to store our datasets. We also used Azure Devops to mention our different stages of the projects using epics, features and tasks.

c. Version control systems

We have uploaded our software files on GitHub which is one of the best internet based hosting services for software development and version control. Please find the link below:
<https://github.com/ruchi3/Loan-Sanction-Predictor>

d. Cloud Usage?

- i. We used OneDrive and Google Drive to store our backup files, only so we all could download and access the original files or work on them simultaneously while any other person always has the most recent version of any file.

e. Development approach

Our development approach is using specified data instead of dynamic. The reason we decided to do it this way is because we decided to give an analysis to the user after they input their financial and descriptive data; instead of directly displaying if they are approved or not.

Our approach was like Rapid Application Development. It comprised of defining the project requirements, designing the model by testing and redesigning it; then getting feedback on it amongst the three of us and finally implementing it.

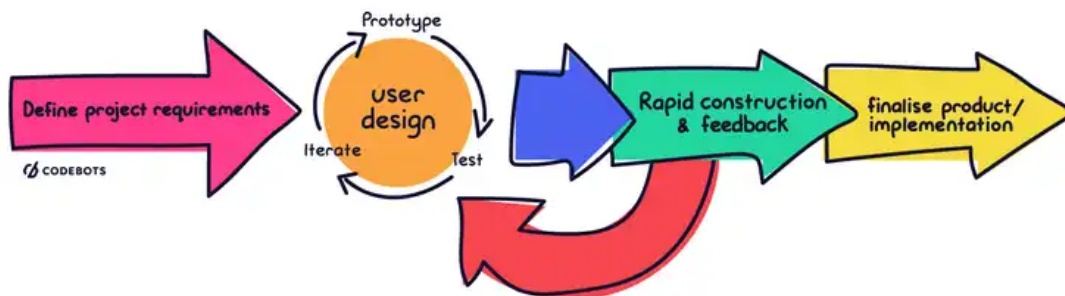


Fig 15: Development Approach

f. New developer Instructions

- i. The hardware needed for this is a normal laptop that has satisfactory memory and is coding compatible.
- ii. Make sure that the device has Jupyter notebook installed either directly or through a virtual machine. This environment would help run the code more smoothly.
- iii. Download the code or import it from GitHub and save it on your laptop.
- iv. Open the code.ipynb file. This file consists of the code for the system.

- v. The Output folder consists of all the outputs we generated.
- vi. The Dataset folder consists of the data sets we use for training and test as well as a sample of what the output data should look like.
- vii. To make any modifications, it would be recommended to make a copy of this folder and store it so that the original code remains untouched just in case of any errors.
- viii. For editing the code further keep in mind that the three forms of algorithms are performed using validation methods which you can read more about in the links cited in references.
- ix. You will run the code.ipynb file to test it and to execute it. The result will be as it displays.
- x. Now for the website, make sure you have access to the editing rights.
- xi. Open a wix.com account so that you can view the website as a developer rather than a user.
- xii. Edit the website if required and for more instructions follow the guide cited in references.

7. Implementation

a. Testing

We simply tested the code by running it and trying to increase its accuracy. Our first approach was Logistic Regression however, the accuracy was lesser than we expected and hence we decided to implement our logic using other models as well. This worked out well for us since Random Forest met our expectations while testing different types of models. We would simply run the code, change certain factors, and run it again to see the mean accuracy.

b. Deployment

This product would be privately deployed instead of being a software accessible to users that pay for it. This would be a service request based product rather than a periodic payment based one. It would be available to the developers and analysts on our side. The customer would be able to access our service of predicting the status of their loan and for everytime they request it, we will be able to provide them with their answers. The website is deployed on wix.com which is a cloud based web development company that can help design extremely user friendly websites.

c. Training

The training required for this model would be to learn how to edit the code that is written in python. A good knowledge of the models used would also be welcome. Apart from that, the

developer should also be made familiar with the guide to website development using wix.com. The link to everything required is cited in our references. Vetting the data accepted in the database is more important than anything else. Hence the new developers would have to be trained to take care of these minute details. Overall, the functionality is simple and reading the new developer instructions should suffice.

d. Support

Any support for any kind of issues would come straight to us. The model could need to be updated based on the type of open ended data it could get from the applicants. That is why data collection for various users and vetting the data collected before actually accepting it as an input to the system is the main task which only skilled individuals would be able to achieve.

8. System Documentation

a. Forms / User Interface

1. The user can click on “Login” in the upper hand right corner to access their account
2. The user can scroll to the bottom of the page to get to the contact information
3. The user can access a form to create an account by clicking on “Create an Account”

i. For at least three User Interface Screens

- Provide a screenshot

The screenshot shows the footer of a website for 'CRUSIUS ISAAC PATEL FINANCIAL CONSULTING'. The footer includes a navigation menu with links: Home, About, Loan Sanction Predictor, Plans & Pricing, More, and Log In. Below the navigation menu is a contact form with three columns. The first column is titled 'Let's Connect' and contains text about questions, doubts, and feedback, along with a 'Reach out now!' section and a 'We look forward to hearing from you!' message. The second column is titled 'Address' and contains the address '1250 Bellflower Boulevard, Long Beach, CA 90840', an email 'katja.crusius@student.csulb.edu', and a phone number '562-714-2642'. The third column is titled 'Contact Us' and contains input fields for 'First Name', 'Last Name', 'Email *', and 'Message', along with a 'Submit' button. A 'Back to Top' link is located at the bottom left of the contact form area.

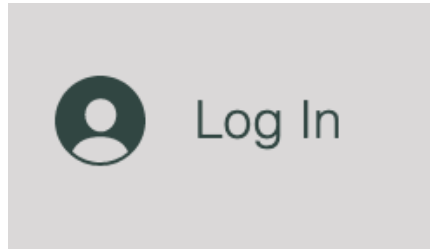


Fig 16: Website Screenshots

- Provide end user instructions
 - Purpose
 1. Entering the user account to access personal information
 2. Getting the contact information to reach out in case of questions
 3. Creating an account to use the loan sanction predictor
 - What values must be entered
 1. Username and password
 2. None required, first name / last name / email / message to reach out
 3. Personal information, passport number, username and password
 - What is the outcome
 1. Users can access their user account
 2. Users can contact the administrators for questions
 3. Users can use the loan sanction predictor

b. Reports

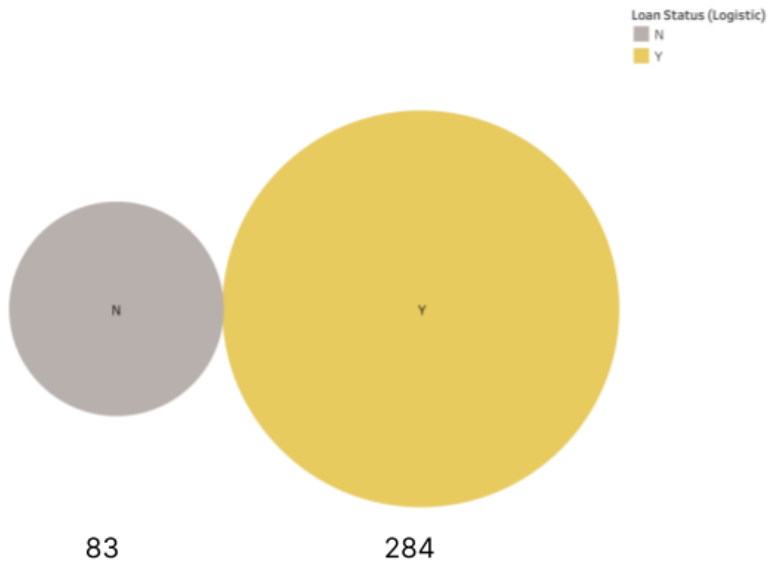
The outputs for the different models have been documented in data visualization for. This is done so that we have a clear understanding of exactly how different the different models namely Logistic Regression, Decision Tree and Random Forest are affecting the output.

We made the data visualizations using Tableau. The imported files for output were in .csv format.

Logistic Regression

<https://github.com/ruchi3/Loan-Sanction-Predictor/blob/main/Output/Log1.csv>

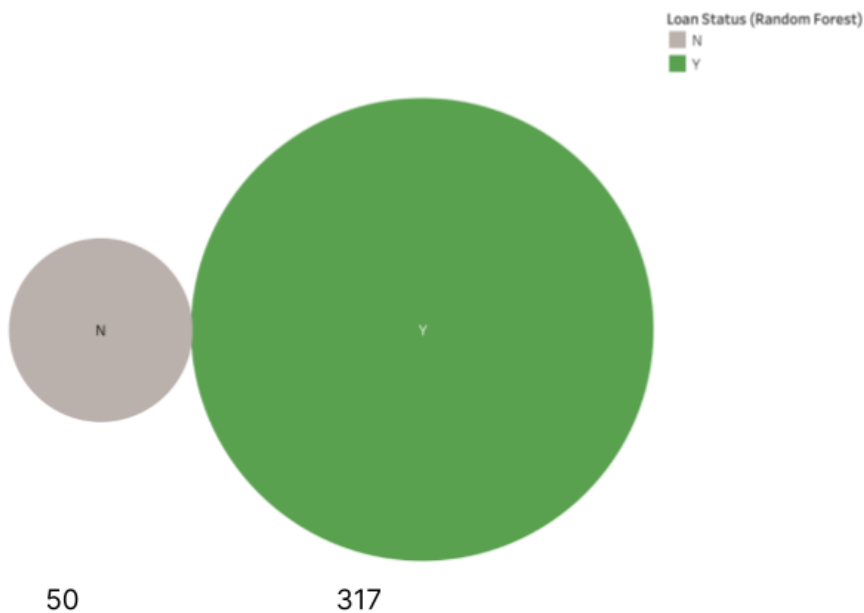
Logistic Regression



Decision Tree

<https://github.com/ruchi3/Loan-Sanction-Predictor/blob/main/Output/DecisionTree.csv>

Random Forest



Random Forest

<https://github.com/ruchi3/Loan-Sanction-Predictor/blob/main/Output/RandomForest.csv>

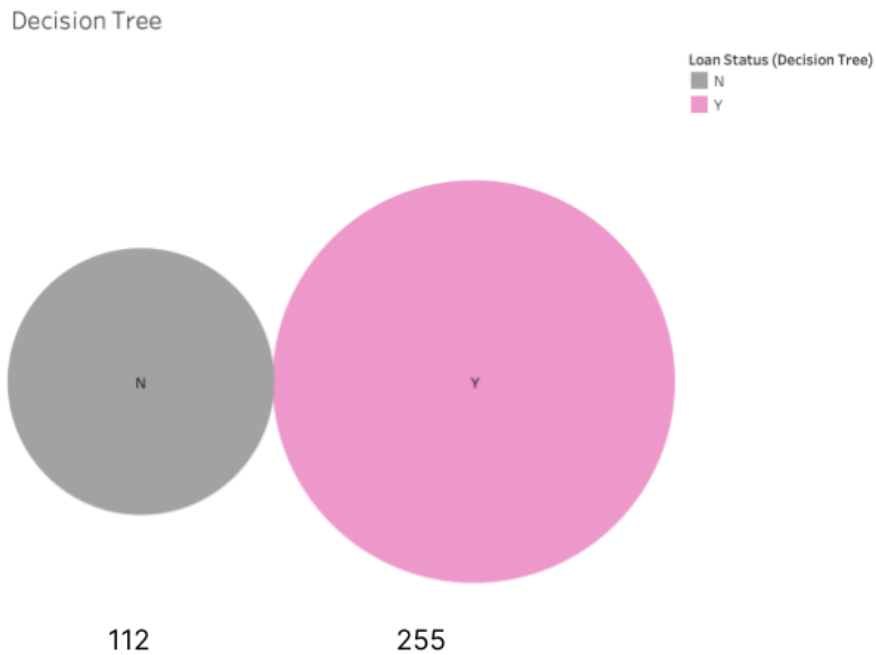
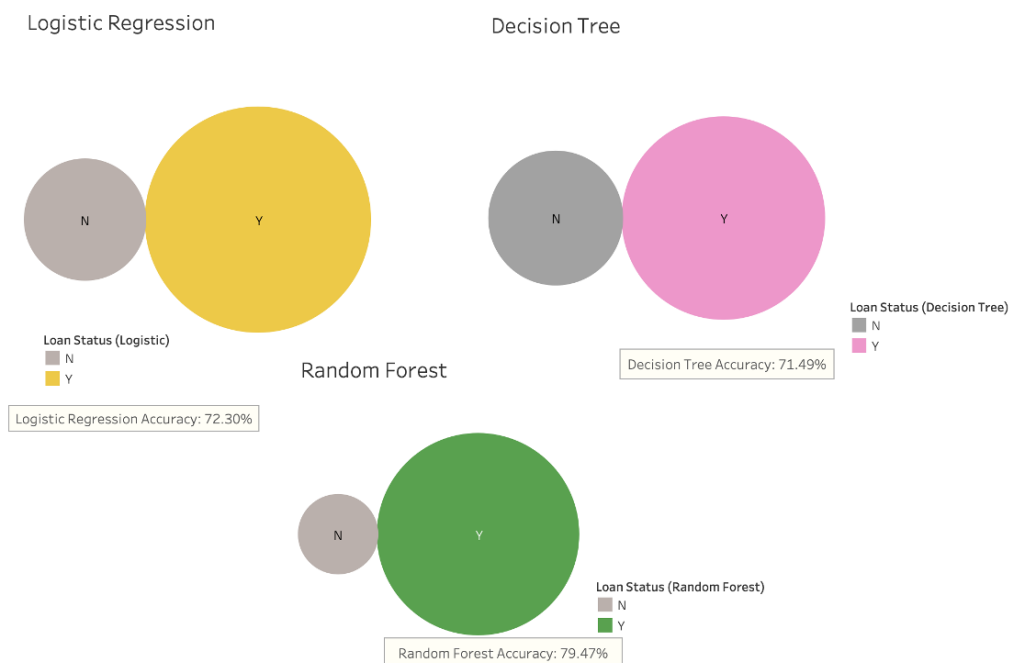


Fig 17: Result Comparison

The final result we document is the comparison of accuracy for our models

Accuracy Comparison of different models

The comparison of these different algorithms brings about a conclusion that Random Forest has the best accuracy out of all different models.



9. Application State / Developed Features / To-Do / Prototype

a. Describe what requirements/features have been developed and are fully functional

We have developed the loan sanction predictor and tested it, so on the basis of demographic and financial information we can make a prediction on whether a person's loan application would be approved or denied. The website is also almost fully functional, with different sections that allow the user to learn about the loan sanction predictor.

We are still missing the dynamic input feature on the website and will need to connect the loan sanction predictor code to the website in the next step so that the user can input their data and get an output that way.

b. What functions/features are in the backlog

i. Depending on what you were able to develop and what remains, provide a prototype for a few missing features

- making it possible for the user to input their data on the website itself to
- having the loan sanction predictor give a result to the user on the website

10. Lessons Learned / References

a. What went well?

We completed the backend of the project entirely and successfully designed a website to relatively demonstrate the results of our work to the users. The process of designing the project from scratch with the help of online resources like the pandas python library or stackoverflow for error solving, worked out well for us. We wanted to demonstrate our visualization and analytical skills by making a website and an analysis from different data points to be depicted on the website.

b. What challenges did you encounter?

One of the challenges was to use different algorithms since we wanted to know which model was giving a good accuracy.. There were a variety of resources as cited in our references and different models that we could work with. We gained 80% accuracy in this project using the Random Forest model and a basic Logistic Regression model. We had to try different models taught to us during the duration of our program since we wanted to compare and depict how different algorithms would affect our results.

To overcome this challenge, we definitely had to refer to various Q&As like stack overflow resources online and try different types of codes before the whole system ran perfectly like clockwork.

c. What would you do differently?

We would have researched a bit more about existing models to learn from them, instead of building the model around information we have been taught and a few additions. We did use some models as the basis of our planning however, if we looked around a bit more, we could have found methods to make the model more accurate.

d. What software/approach did you use and decide against?

The programming language used here is python.

We import libraries like sklearn, seaborn, pandas etc. These libraries are essential to implementing the analysis and visualization responsible for providing us a desired outcome.

We condense the data through our model of three different types namely, Logistic Regression, Random Forest and the Decision Tree model.

We decided against using XGBoost since it is a model we are not very familiar with and choosing XGBoost is not as necessary as predicting the outcome of a loan application.

e. Anything that may be helpful for future development (by you or somebody else)

The future development for this would be using different models for predictions and getting the accuracy higher. If we could pool in more factors of the users and plan out on how to analyze them well enough to collectively feed it to a model that could use the smallest features of a user and generate our required output, that would be an improvement to our current model. It is a very possible future scope.

f. References (if applicable) ONLY if you had any otherwise no need.

- i. <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.html>
- ii. https://scikit-learn.org/stable/user_guide.html
- iii. https://pandas.pydata.org/docs/user_guide/visualization.html
- iv. <https://www.wix.com/blog/2021/02/how-to-make-a-website/>

11. Closing Comments

a. The team's closing comments

i. About the specific application

We are overall very happy with the application we have developed. As we have encountered the issue of building a credit score and having to apply for certain things based on it ourselves, we recognized the importance of it and chose to build this application for it. On the technical side, we were able to complete almost everything - we built the system, the system gives recommendations, and we designed the website around it. The only thing that we did not have time for was integrating the system into the website, which is something we could work on in the future.

ii. About the class

We all enjoyed the class a lot, because it provided a good balance between learning about software engineering and completing our own project at the same time. It was really good learning or repeating certain concepts in a class and then completing the corresponding homework as well as a project deliverable for it, it made the development of the project much easier for us and helped a lot.

iii. Recommendations for future classes

For future classes we would recommend potentially allowing a little more time in class at the beginning to figure out the project deliverables. While we were mostly fine with the deadlines and time limits and completed our work from home, it would be helpful allowing 10 to 15 minutes at the beginning of every class for the teams to regroup and make a list of things to do as well as distribute tasks. Communicating with other team members can sometimes be a challenge especially with people being busy with other classes, so allowing just a few minutes in the beginning of each class would be really helpful.

12. Software Delivery

We have uploaded our software files on GitHub which is one of the best internet based hosting services for software development and version control. Please find the link below:

<https://github.com/ruchi3/Loan-Sanction-Predictor>

13. Application Delivery

a. Place your code in a PaaS (Heroku, AWS, or similar)

We chose to use WIX to build our website as we have previously used it with other professors and it has proven to be a reliable service to build an interactive website without having to buy a domain.

b. Provide a link to the application.

This is the link to the website: <https://katjacrusius.wixsite.com/cipc>

If you encounter any issues, please let us know.