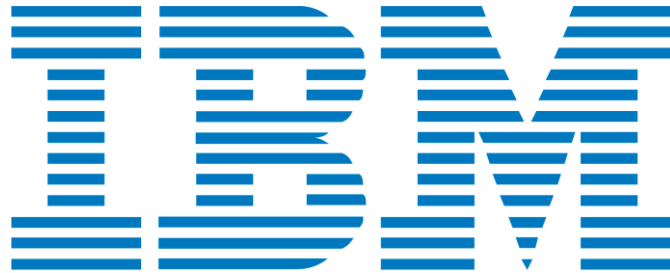


# **NAALAIYA THIRAN PROJECT**



## **SMART LENDER – APPLICANT CREDIBILITY PREDICTION FOR LOAN APPROVAL**

**( BASED ON : APPLIED DATA SCIENCE (ADS) )**

**TEAM ID: PNT2022TMID36990**

### **TEAM MEMBERS:**

**RESHMA P (212019104030)**

**REVATHI A (212019104031)**

**KAVIYA S (212019104014)**

**SRI SWETHA (212019104033)**

## **1.1 INTRODUCTION:**

A loan is the major source of income for the banking sector of financial risk for banks. Large portions of bank's assets directly come from the interest earned on loans given. The Activity of lending loans carry great risks including the inability of borrower to pay back the loan by the stipulated time. It is referred as "credit risk". A candidate's worthiness for loan approval or rejection was based on a numerical score called "credit score". Therefore, the goal of this paper is to discuss the application of different Machine Learning approach which accurately identifies whom to lend loan to and help banks identify the loan defaulters for much-reduced credit risk.

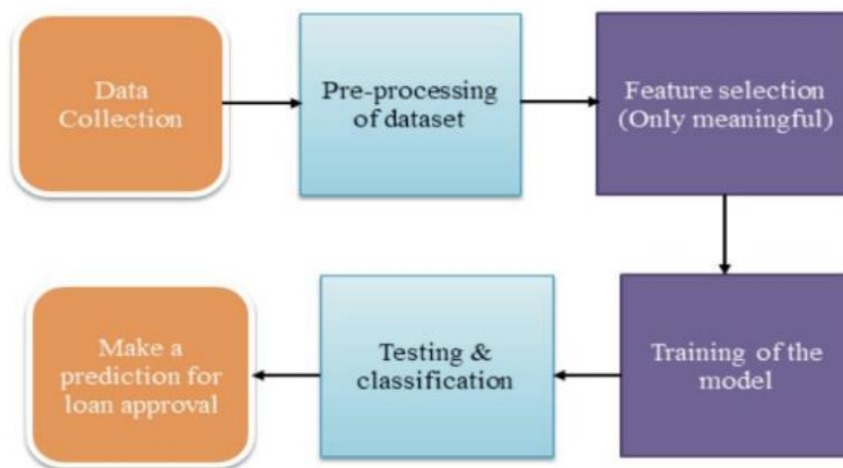
## **2.1 LITERATURE SURVEY:**

A recent development of data science techniques and data mining has led to an interest of implementing these techniques in various fields. The banking sector is no exclusion and the increasing requirements towards financial institutions to have robust risk management has led to an interest of developing current methods of risk estimation. Potentially, the implementation of data science techniques could lead to better quantification of the financial risks that banks are exposed to.

Within the credit risk area, there has been a continuous development of the Basel accords, which provides frameworks for supervisory standards and risk management techniques as a guideline for banks to manage and quantify their risks. From Basel II, two approaches are presented for quantifying the minimum capital requirement such as the standardized approach and the internal ratings based approach (IRB).

There are different risk measures banks consider in order to estimate the potential loss they may carry in future. One of these measures is the expected loss (EL) a bank would carry in case of a defaulted customer. One of the components involved in EL estimation is the probability if a certain customer will default or not.

Customers in default means that they did not meet their contractual obligations and potentially might not be able to repay their loans.



Thus, there is an interest of acquiring a model that can predict defaulted customers. A technique that is widely used for estimating the probability of client default is Logistic Regression. In this thesis, a set of machine learning methods will be investigated and studied in order to test if they can challenge the traditionally applied techniques. A prediction is a statement about what someone thinks will happen in the future.

People make predictions all the time. Some are very serious and are based on scientific calculations, but many are just guesses. Prediction helps us in many things to guess what will happen after some time or after a year or after ten years. Predictive analytics is a branch of advanced analytics that uses many techniques from data mining, statistics, modeling, machine learning, and artificial intelligence to analyze current data to make predictions. This literature review helps us carry out our work and propose a reliable bank loan prediction model. Manjeet et al (2018) there are seven types of variables that may influence consumer loan default; consumer's annual income, debt-income ratio, occupation, home ownership, work duration and whether or not consumer possesses a saving/checking account.

In a work by Steenackers and Goovaerts, the key factors that may influence loan default are borrower's age, location, resident/work duration, owner of phone, monthly income, loan duration, whether or not applicant works in a public sector, house ownership and loan numbers. Another study by Ali Bangher pour on a large dataset within the period of 2001-2006 indicated that loan age was the most important factor when predicting loan default while market loan-to-value was the most effective factor for mortgage loan applications. In addition to identifying factors that may influence loaned fault, there is also a need to build robust and effective machine learning models that can help capture important patterns in credit data. The choice of model so great importance as the chosen model plays a crucial role in determining accuracy, precision and efficiency of a prediction system. Numerous models have been used for loan default prediction and although there is no one optimal model, some models definitely do better than others. In 2019, Vimala and Sharmili proposed a loan prediction model using and Support Vector Machines(SVM)methods.

Naïve Bayes, an independent speculation approach, encompasses probability theory regarding the data classification. On the other hand, SVM uses statistical learning model for classification of predictions. Dataset from UCI repository with 21 attributes was adopted to evaluate the proposed method. Experimentations concluded that, rather than individual performances of classifiers (NB and SVM), the integration of NB and SVM resulted in an efficient classification of loan prediction. In 2019, Jency, Sumathi and Shiva Sri proposed an Exploratory Data Analysis(EDA) regarding the loan prediction procedure based on the client's nature and their requirements.

The major factors concentrated during the data analysis were annual income versus loan purpose, customer's trust, loan tenure versus delinquent months, loan tenure versus credit category, loan tenure versus number of years in the current job, and chances for loan repayment versus the house ownership. Finally, the outcome of the present work was to infer the constraints on the customer who are applying for the loan followed by the prediction regarding the repayment. Further, results showed that, the customers were interested more on availing short-tenure loans rather than long-tenure loans. In 2019, Supriya, Pavani, Saisushma, Vimala Kumari and Vikas presented a ML based loan prediction model.

The modules in the present approach were data collection and pre-processing, applying the ML models, training followed by testing the data. During the pre-processing stage, the detection and removal of outliers and imputation removal processing were carried out. In the present method, SVM, DT, KNN and gradient boosting models were employed to predict the possibilities of current status regarding the loan approval process. The conventional 80:20 rule was adopted to split the dataset into training and testing processes. Experimentation concluded that, DT has significantly higher loan prediction accuracy than the other models. In 2017, Goyal and Kaur presented a loan prediction model using several Machine Learning (ML) algorithms. The dataset with features, namely, gender, marital status, education, number of dependents, employment status, income, co applicant's income, loan amount, loan tenure, credit history, existing loan status, and property area, are used for determining the loan eligibility regarding the loan sanctioning process. Various ML models adopted in the present method includes, Linear model, Decision Tree (DT), Neural Network (NN), Random Forest (RF), SVM, Extreme learning machines, Model tree, Multivariate Adaptive Regression Splines, Bagged Cart Model, NB and TGA.

When evaluated these models using Environment in five runs, TGA resulted in better loan forecasting performance than the other methods. In 2016, Aboobyda Jafar Hamid and Tarig Mohammed Ahmed presented a loan risk prediction model based on the data mining techniques, such as Decision Tree (J48), Naïve Bayes (NB) and BayseNet approaches. The procedure followed was training set preparation, building the model, Applying the model and finally. Evaluating the accuracy. This approach was implemented using Weka Tool and considered a dataset with eight attributes, namely, gender, job, age, credit amount, credit history, purpose, housing, and class. Evaluating these models on the dataset, experimental results concluded that, J48 based loan prediction approach resulted in better accuracy than the other methods.

Variable	Description
Loan_ID	Unique Loan ID
Gender	Male/ Female
Married	Applicant married (Y/N)
Dependents	Number of dependents
Education	Applicant Education (Graduate/ Under Graduate)
Self_Employed	Self employed (Y/N)
ApplicantIncome	Applicant income
CoapplicantIncome	Coapplicant income
LoanAmount	Loan amount in thousands
Loan_Amount_Term	Term of loan in months
Credit_History	credit history meets guidelines
Property_Area	Urban/ Semi Urban/ Rural
Loan_Status	Loan approved (Y/N)

In 2016, Kacheria, Shivakumar, Sawkar and Gupta suggested a loan sanctioning prediction procedure based on NB approach integrated with K-Nearest Neighbor (KNN) and binning algorithms. The seven parameters considered were income, age, profession, existing loan with its tenure, amount and approval status. The sub-processes include, Preprocessing (handling the missing values with KNN and data refinement using binning algorithm), Classification using NB approach and Updating the dataset frequently results in appropriate improvement in the loan prediction process.

Experimentation put-forth the conclusion that, integration of KNN and binning algorithm with NB resulted in improved prediction of loan sanctioning process. In 2016, Goyal and Kaur suggested an ensemble technique based loan prediction procedure for the customers. The sub processes in the present method includes, data collection, filtering the data, feature extraction, applying the model, and finally analysis the results. The various loan prediction procedures implemented in the present method were Random Forest (RF), SVM and Tree model with Genetic Algorithm (TGA). The parameters considered for evaluating the models were accuracy, Gini Coefficient, Area Under Curve (AUC), Receiver Operating Curve (ROC), Kolmogorov - Smirnov (KS) Chart, Minimum Cost - Weighted Error Rate, Minimum Error Rate, and K-Fold Cross Validation parameters.

Experimentation outcome concluded that the integration of three methods (RF, SVM and TGA) resulted in improved loan - prediction results rather than individual method 's prediction. In 2006, Sudhamathy suggested a risk analysis method in sanctioning a loan for the customers using R package. The various modules include data selection, pre-processing, feature extraction and selection, building the model, prediction followed by the evaluation. The dataset used for evaluation in this method was adopted from UCI repository. To fine tune the prediction accuracy, the preprocessing operation includes the following sub-processes: detection, ranking and removal of outliers, removal of imputation, and balancing of dataset by proportional bifurcation regarding testing and training process.

Further, feature selection process improves the prediction accuracy. When evaluated, the DT model resulted in 94.3% prediction accuracy. The process of analyzing data from different perspectives and extracting useful knowledge from it. The core of knowledge discovery process. The various steps involved in extracting knowledge from raw data. Different data mining techniques include classification, clustering, association rule mining, prediction and sequential patterns, neural networks, regression etc. Classification is the most commonly applied data mining technique, which employs a set of pre-classified examples to develop a model that can classify the population of records at large.

Fraud detection and credit risk applications are particularly well suited to classification technique. This approach frequently employs Decision tree based classification Algorithm. In classification, a training set is used to build the model as the classifier which can classify the data items into its appropriate classes. A test set is used to validate the model.

Further works on loan approval prediction models are given as follows:

TITLE 1:

Improving Information Quality in Loan Approval Processes for Fair Lending and Fair Pricing AUTHOR: M. Cary Collins YEAR: 2013 DESCRIPTION: Bank data management on loan approval processes has great room for improvements of information quality and data problems prevention especially with regards to fair lending and fair pricing practices. They first reviewed briefly typical data collection protocols deployed at many financial institutions for loan approval and loan pricing. Federal regulations mandate portions of these data protocols. While discussing the data capture and analysis for fair lending, they illustrated some initial key steps currently needed for improving information quality to all parties involved.

TITLE 2:

Loan Credibility Prediction System Based on Decision Tree Algorithm AUTHOR: Sivasree M S, Rekha Sunny T YEAR: 2015 DESCRIPTION: Data mining techniques are becoming very popular nowadays because of the wide availability of huge quantity of data and the need for transforming such data into knowledge. Data mining techniques are implemented in various domains such as retail industry, biological data analysis, intrusion detection, telecommunication industry and other scientific applications. Techniques of data mining are also be used in the banking industry which help them compete in the market well equipped. In this paper, they introduced a prediction model for the bankers that will help them predict the credible customers who have applied for a loan. Decision Tree Algorithm is being applied to predict the attributes relevant for credibility. A prototype of the model has been described in this paper which can be used by the organizations for making the right decisions to approve or reject the loan request from the customers.

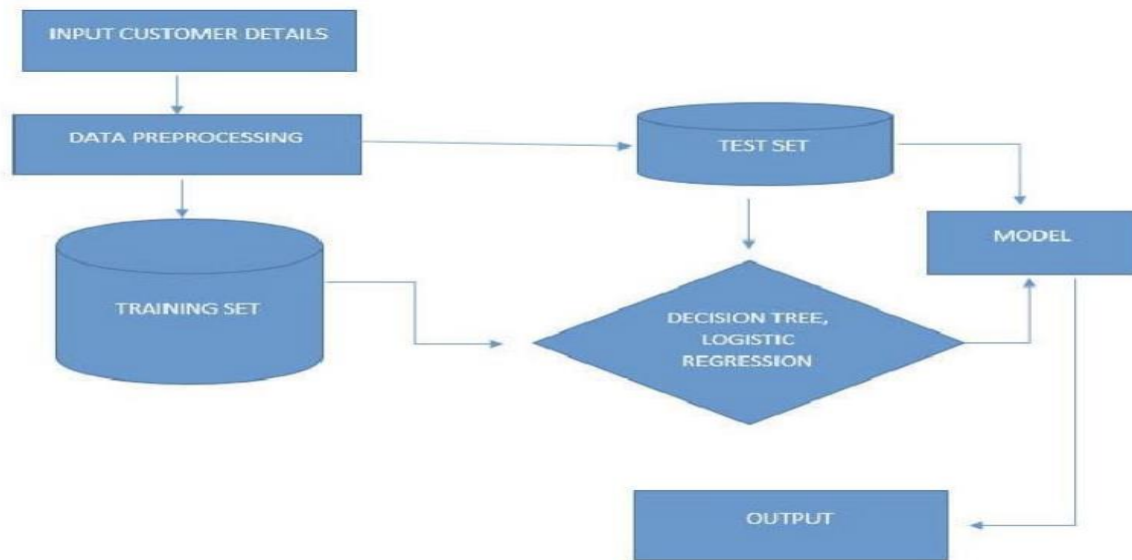
TITLE 3:

Loan Approval Prediction based on Machine Learning Approach AUTHOR: Kumar Arun, Garg Ishan, Kaur Sanmeet .

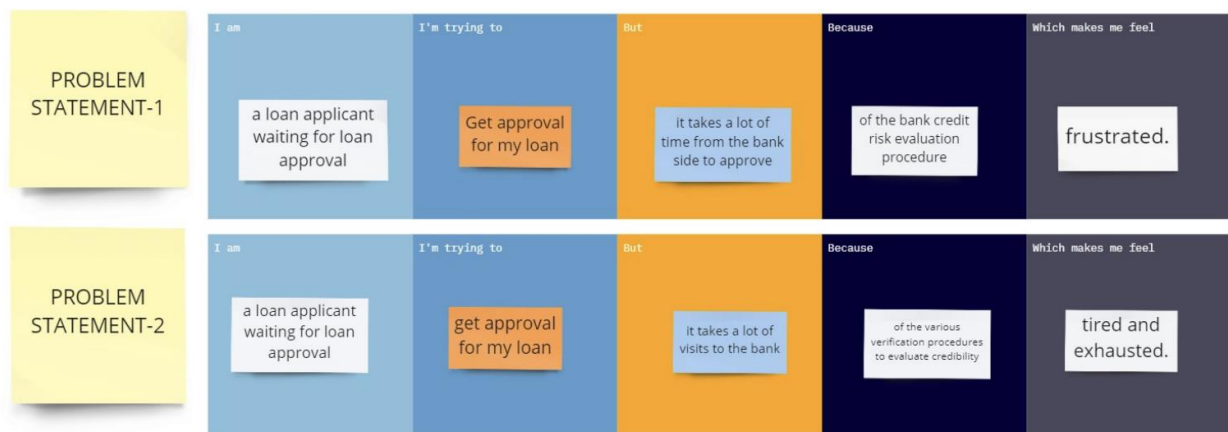
## **2.2 DESCRIPTION:**

With the enhancement in the banking sector, lots of people apply for bank loans but the bank has its limited assets which it grants to only limited people , so finding out to whom the loan can be granted is a typical process for the banks. So, in this paper , they tried to reduce this risk by selecting the safe person so as to save lots of bank efforts and assets. It was done

by mining the previous records of the people to whom the loan was granted before and on the basis of these records the machine was trained using the machine learning model which gave the most accurate result. The main goal of this paper is to predict if loan assignment to a specific person will be safe or not. This paper has into four sections (i) Collection of data (ii) Comparing the machine learning models on collected data (iii) Training the system on most promising model (iv) Testing the system.



## 2.3 PROBLEM STATEMENT DEFINITION:



## IDEATION AND PROPOSED SOLUTION

### 3.1 EMPATHY MAP CANVAS:





S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	How might we build a model to predict applicant credibility of a customer who applies for loan in order to make the process easy and simple for the customer?
2.	Idea / Solution description	Building a prediction model using Machine Learning Algorithms like Decision tree, Random Forest, KNN, Xgboost and comparing them to choose the most accurate one. Training the model and integrating flask with scoring endpoint.
3.	Novelty / Uniqueness	Using various prediction models, then evaluating and comparing them to choose the best makes the solution more accurate. Ease of customer is enhanced using simple and easy to use User Interface.
4.	Social Impact / Customer Satisfaction	Customers are satisfied with the model as it saves a lot of time taken for processing during loan approval and makes them free from visiting the bank often for evaluations.
5.	Business Model (Revenue Model)	Almost every bank is digitalized nowadays hence this prediction model to compute applicant's credit risk evaluation is of much use in the business and revenue point of view.
6.	Scalability of the Solution	The model is scaled such that it can handle any amount of data and perform computations in a cost-effective and time saving way.

### 3.4 PROBLEM SOLUTION FIT:

Define CS, fit into CC	<div>1. CUSTOMER SEGMENT(S)<div>CS</div></div> <div>Who is your customer? i.e. working parents of 0-5 y.o. kids</div> <div>Applicants who apply for loan and wait for approval.</div>	<div>6. CUSTOMER CONSTRAINTS<div>CC</div></div> <div>What constraints prevent your customers from taking action or limit their choices of solutions? i.e. spending power, budget, no cash, network connection, available devices.</div> <div>Web phishing which makes the customer worry about privacy to personal details Network issues.</div>	<div>5. AVAILABLE SOLUTIONS<div>AS</div></div> <div>Which solutions are available to the customers when they face the problem or need to get the job done? What have they tried in the past? What pros &amp; cons do these solutions have? i.e. pen and paper is an alternative to digital notetaking</div> <div>Applicants credibility prediction may be done in minutes for loan approval which usually takes days and multiple visits before using ML algorithms. It makes the procedure for approval easy and simple for the customer. The cons include the network issues etc.</div>	Explore AS, differentiate
	<div>2. JOBS-TO-BE-DONE / PROBLEMS<div>J&amp;P</div></div> <div>Which jobs to be done (or problems) do you address for your customers? There could be more than one; explore different sides.</div> <div>Wait for the loan to be approved. Multiple visits to the bank for credit risk evaluation. Multiple visits to the bank for document verification etc.</div>	<div>9. PROBLEM ROOT CAUSE<div>RC</div></div> <div>What is the real reason that this problem exists? What is the back story behind the need to do this job? i.e. customers have to do it because of the change in regulations.</div> <div>Former bank procedures were all manual and took days or even months to process these loan applications which involves complex process like bank credit risk evaluation. Customers can adapt to this predicting site in order to save time and as banks are also gradually digitalized as well.</div>	<div>7. BEHAVIOUR<div>BE</div></div> <div>What does your customer do to address the problem and get the job done? i.e. directly related: find the right solar panel installer, calculate usage and benefits; indirectly associated: customers spend free time on volunteering work (i.e. Greenpeace)</div> <div>Go to the loan applicant credibility site, view the details required for approval then apply. Search for the procedure and details to be provided for application</div>	
Focus on J&P, tap into BE, understand RC	<div>3. TRIGGERS<div>TR</div></div> <div>What triggers customers to act? i.e. seeing their neighbor installing solar panels, reading about a more efficient solution in the news.</div> <div>Knowing that a procedure exists for loan approval with applicant credibility prediction in minutes which eliminates the time taken and multiple visits in the former one.</div>	<div>10. YOUR SOLUTION<div>SL</div></div> <div>If you are working on an existing business, write down your current solution first, fill in the canvas, and check how much it fits reality. If you are working on a new business proposition, then keep it blank until you fill in the canvas and come up with a solution that fits within customer limitations, solves a problem and matches customer behavior.</div> <div>An Applicant Credibility Prediction Model for performing the credit risk evaluation for loan approval- Built using ML algorithm to predict the credibility of an loan applicant using the details provided like Loan ID, Applicant Income, Loan Amount and its term, Credit History, Property area etc.</div>	<div>8. CHANNELS of BEHAVIOR<div>CH</div></div> <div>8.1 ONLINE What kind of actions do customers take online? Extract online channels from #7</div> <div>8.2 OFFLINE What kind of actions do customers take offline? Extract offline channels from #7 and use them for customer development.</div> <div><ul style="list-style-type: none"><li>• ONLINE: Visit the credibility prediction site and apply by filling the necessary details.</li><li>• OFFLINE: Verify their documents to ensure that proper details are provided for prediction.</li></ul></div>	Focus on J&P, tap into BE, understand RC
Identify strong TR & EM	<div>4. EMOTIONS: BEFORE / AFTER<div>EM</div></div> <div>How do customers feel when they face a problem or a job and afterwards? i.e. lost, insecure &gt; confident, in control - use it in your communication strategy &amp; design.</div> <div>Tired of waiting &amp; multiple visits &gt; Instant prediction</div>			Identify strong TR & EM

## REQUIREMENT ANALYSIS

#### 4.1 FUNCTIONAL REQUIREMENT:

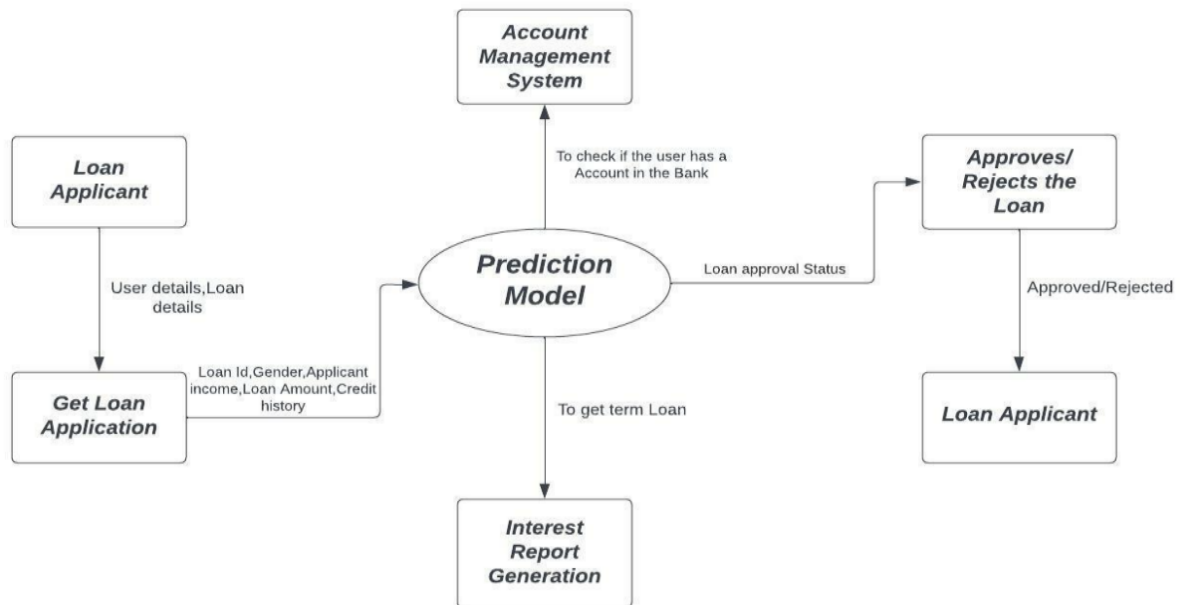
FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Registration through Form
FR-2	User Confirmation	Confirmation via Email Confirmation via SMS
FR-3	Detail collection	Details provided by user should be taken as input for the model
FR-4	Prediction Model	The prediction model predicts the applicant credibility prediction for loan approval
FR-5	User Review	Review through forms Quick review through ratings

#### 4.2 NON-FUNCTIONAL REQUIREMENT:

FR No.	Non-Functional Requirement	Description
NFR-1	<b>Usability</b>	The website is made readily available and is useful to the applicants who wish to predict their credibility for loan approval.
NFR-2	<b>Security</b>	The user may have a doubt on the privacy of their loan details hence a Well defined mechanism is used to maintain all the details provided securely.
NFR-3	<b>Reliability</b>	The reliability is ensured by the accuracy of our prediction model. Since we build 4 models and choose the best and accurate one, Accuracy could be guaranteed up to a great extent.
NFR-4	<b>Performance</b>	The prediction model takes the necessary details from the user and does the prediction within minutes and with accuracy.
NFR-5	<b>Availability</b>	The website ensures availability to every user and the server may handle multiple requests at the same time without any delays.
NFR-6	<b>Scalability</b>	The website scales the load itself according to the usage of applicants.

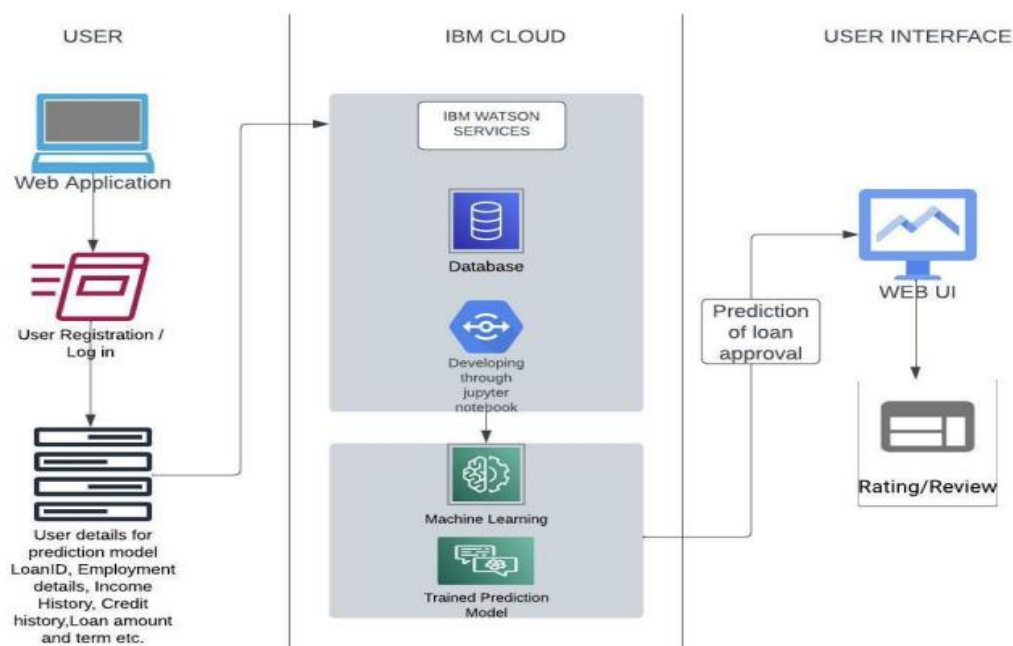
## PROJECT DESIGN

## 5.1 DATA FLOW DIAGRAM:



## 5.2 SOLUTION AND TECHNICAL ARCHITECTURE

### TECHNICAL ARCHITECTURE



## 5.3 USER STORIES:

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (Mobile user)	Registration	USN-1	As a user, I can register for the loan application by entering my email, password, and confirming my password.	I can access my account / dashboard	High	Sprint-3
		USN-2	As a user, I will receive confirmation email once I have registered for the loan application	I can receive confirmation email & click confirm	High	Sprint-3
	Login	USN-3	As a user, I can log into the loan application by entering email & password	I can receive the message that your ID is get login	High	Sprint-3
	Dashboard	USN-4	As a User, I can use the dashboard it will Display the summary of the total loan process	I can access my dashboard to view entire summary of the loan application	Medium	Sprint-3
	User detail form	USN-5	As a User, I provide the required details to get the applicant credibility prediction for loan approval	I can get the prediction for loan approval for prediction details	High	Sprint-4
Customer (Web user)	Registration	USN-6	As a User, I can register for loan website by entering my email, password and confirming my password	I can receive my acceptance mail	High	Sprint-3
	User detail form	USN-7	As a User, I provide the required details to get the applicant credibility prediction for loan approval	I can get the prediction for loan approval for prediction details	High	Sprint-4
Customer Care Executive	Queries	USN-8	As a new user how can I create my account, as a old user how can I resolve my login issues	Clarify queries through via phone calls or by email	Medium	Sprint-3
Bank Administrator	Referring the credit process	USN-9	Referring the loan approval status of the particular loan ID	Loan approval	High	Sprint-4

## PROJECT PLANNING AND SCHEDULING

## 6.1 SPRINT PLANNING AND ESTIMATION:

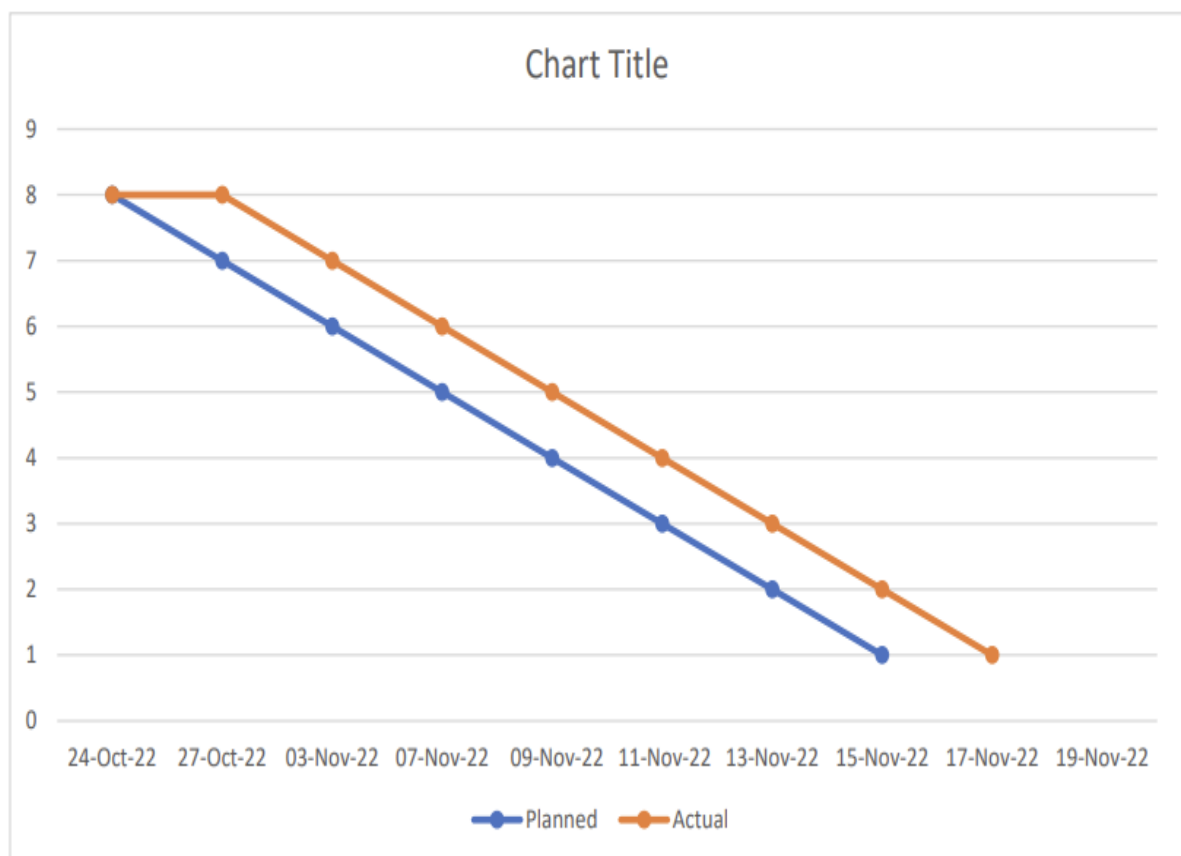
Activity No.	Activity	Activity Description	Duration
1.	Understanding the project	Understanding the problem statement, project objective, flow and structure.	1 week
2.	Attending Training sessions	To gain technical knowledge on the technology required to complete the project.	4 weeks
3.	Ideation and design	Brainstorming solution ideas for the problem statement and designing the solution.	2 weeks
4.	Data collection, Visualizing and analyzing, Data Preprocessing	Downloading the dataset. Importing libraries and reading the dataset Performing Univariate, Bivariate, Multivariate Visualizations and description analysis. Data Preprocessing.	1 week
5.	Model building	Decision Tree, Random Forest, KNN, XgBoost. Comparing Models. Evaluating performance.	1 week
6.	Application Building	Building HTML Webpage. Python for prediction model and running application.	1 week
7.	Training Model on IBM	Training ML model on IBM. Integrating flask with scoring end point.	1 week

## 6.2 SPRINT DELIVERY SCHEDULE:

Sprint	Functional Requirement	Story Points	Priority	Team Members
Sprint-1	Data collection	5	High	Reshma, Revathi, Sri Swetha, Kaviya
	Visualizing and analyzing	5	High	Reshma, Revathi, Sri Swetha, Kaviya
	Data preprocessing	10	High	Reshma, Revathi, Sri Swetha, Kaviya
Sprint-2	Model building		High	Reshma, Kaviya
	decision tree	2	High	Reshma, Revathi, Sri Swetha
	Random forest	2	High	Reshma, Sri Swetha, Kaviya
	knn	2	High	Reshma, Kaviya
	Xgboost	2	High	Revathi, Kaviya
	Compare model	6	High	Reshma, Revathi, Sri Swetha
	Evaluate performance & save	6	High	Reshma, Revathi, Kaviya
Sprint-3	Application building		High	
	Building HTML	5	High	Revathi, Sri Swetha
Sprint	Functional Requirement	Story Points	Priority	Team Members
	Python	5	High	Reshma, Kaviya
	Run application	10	High	Reshma, Revathi, Sri Swetha, Kaviya
Sprint-4	Train model on IBM		High	
	Register for IBM	2	High	Reshma, Revathi, Sri Swetha, Kaviya
	Train ML model	10	High	Reshma, Kaviya
	Integrate flask with	8	High	Reshma, Revathi
	Scoring end point			

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	20	11 Days	24 Oct 2022	03 Nov 2022	20	03 Nov 2022
Sprint-2	20	4 Days	04 Nov 2022	07 Nov 2022		
Sprint-3	20	5 Days	08 Nov 2022	12 Nov 2022		
Sprint-4	20	7 Days	13 Nov 2022	19 Nov 2022		

$$AV = \frac{\text{sprint duration}}{\text{velocity}} = \frac{27}{20} = 1.35$$





## **7.1 SOLUTION:**

To search multiple repositories more systematically, you can use a cross-repository search. A cross-repository search can be used to search across IBM Content Manager, FileNet® P8, IBM Content Manager OnDemand, and CMIS repositories. The cross-repository search can be saved in IBM Content Manager or FileNet P8 repositories.

With a cross-repository search, you can specify how different properties are related to each other across a selection of repositories. You build a cross-repository search by specifying the repositories that contain the documents that you need and then mapping properties that you want to use in your search. You can associate properties from different repositories by mapping properties to a representative property name, which is then displayed as an available search criterion. When the cross-repository search is created and saved, you can gather information this way for each future loan applicant.

When the scope for the cross-repository search is set up, you then enter your search criteria for the mapped properties, for example, the client's identification number and a date range for the documents and records that you are searching for.

## **7.2 SECNARIO:**

In this scenario, the bank uses the following general steps to research and process a loan for the client:

1. Before a loan can be approved, the loan officer must gather all pertinent documents to confirm that the client can be approved for a loan. The underwriter runs five separate searches to find everything that he needs.
2. The bank asks a senior business analyst to create a cross-repository search to find these types of documents more efficiently. The search can be used whenever a loan officer needs to find and review these types of documents.
3. The business analyst creates a cross-repository search that searches transaction records that are stored on Content Manager OnDemand repositories (contains transaction records over a specific time, for example, three months), and any documents in a IBM Content Manager or FileNet P8 repositories that might explain any suspicious transactions.
4. An email with a link to the saved cross-repository search is sent to the team of loan officers, and they are informed that the saved search is also available in the “Loans” teamspace, where other loan resources can be found. The loan officers can individually add the search to their favorites.
5. A loan officer processes the customer's bank records and financial history to determine whether a loan can be approved.
6. The loan officer approves the bank loan.
7. The bank sets up a meeting to discuss the terms of the loan contract.



Gerard is the manager of the central branch for Example Bank Corporation. The central branch receives 50 loan applications per week on average. Each time that a loan application is received, a loan officer must review the applicant's transaction records and check that information with information provided in loan application forms and other data provided by the applicant. The loan officer must search for transaction records in one repository and find other documents in another repository. The loan officer enters the same search criteria to the saved searches used in both repositories.

Gerard calls Priya, a business analyst in the bank's internal information technology (IT) team, to create a cross-repository search so that the redundancy in searching for information about the applicant can be reduced.

Priya creates a cross-repository search that can search all the repositories that are necessary to search during the loan application approval process. Priya names the search, **Applicant Data Search**, and saves the search to a FileNet P8 repository and adds the search to the “Loans” teamspace so that the team of loan officers can easily access the search.

Hassan, a loan officer at the bank, receives a loan application for his new client, Corinne. In the “Loans” teamspace, Hassan opens the **Applicant Data Search** saved cross-repository search.

Hassan enters search criteria to find Corinne's transaction records by document type and by selecting a date range and Corinne's ID to scope the search results.

## TESTING:

IBM Watson Studio interface showing a deployed model named "loan-demo" in the "Test" tab. The "Enter input data" section is active, displaying a table with 10 rows and 7 columns. The table contains numerical data for various attributes like Self\_Employed, ApplicantIncome, CoapplicantIncome, LoanAmount, Loan\_Amount\_Term, Credit\_History, and Property\_Area. A "Predict" button is visible at the bottom right of the input section.

	Self_Employed (float64)	ApplicantIncome (float64)	CoapplicantIncome (float64)	LoanAmount (float64)	Loan_Amount_Term (float64)	Credit_History (float64)	Property_Area (float64)
1		6589.0	1569.0	125.0	180.0	1.0	1.0
2		5613.0	2565.0	1632.0	360.0	0.0	0.0
3		9985.0	2587.0	1250.0	360.0	0.0	1.0

IBM Watson Studio


Deployments / loan-demo / loan /

### Prediction results

Prediction type: ☒ Table view ☐ JSON view

#### Binary classification

Prediction percentage



10 Records

	Prediction	Confidence
1	1	55%
2	0	52%
3	0	52%
4	0	53%
5	0	74%
6	0	53%
7	0	52%
8	0	56%
9	0	53%
10	1	55%

Download

## RESULT

A **Loan Approval** is an unsecured loan and is one of the easiest ones to get if you are pre-qualified. It involves minimal documentation, and if you are pre-qualified, then the process is even more straightforward. Let's take a look at the whole process end-to-end.

### **Getting a loan approval involves four steps:**

**Loan application and eligibility:** Today, we can apply for a loan either online or offline. It is preferable to apply online because it is quicker and safer (no need for physical interaction with anyone outside). An online application is approved within minutes with some information like Load\_ ID, Gender, Education, Applicant Income, Loan Amount such like the images given below.

The loan predication is done successfully by the loan approval using applied data science. Apply online for loan approval makes the process quicker and simpler.

**Lender meeting and discussion:** Once the application has gone through to the next stage; the lender sends someone to meet the applicant. This usually takes place at the office or home of the applicant. Questions asked pertain to the stability of employment and regular income. Once the meeting is done, the application is either approved or rejected. If one applies online, sometimes a bank representative asks questions over the phone or video call. If you are pre-approved, a meeting generally doesn't occur, and you skip straight to step three.

**Documentation:** After approval, documents like income proof, employment details, bank statements, age proof, address proof (Aadhaar card/passport), and PAN card must be submitted to the bank. If all the documents match the approved criteria, the application proceeds to the next stage. If the details don't match, the loan application process can be delayed or even rejected.

**Approval and disbursement:** Once all the documents are submitted, verified and approved, the personal loan is disbursed. Online applications generally complete the process within days. An offline process might take two weeks to get done. A pre-

approved applicant can get the loan disbursed into their account within 72 hours if all documentation is in order.

The loan Approval is done successfully by using applied data science. Applying online for loan approval makes the process quicker and simpler.

**Loan Prediction**

**Using Applied Data Science**

Gender

Married

Dependents

Education

Self\_Employed

ApplicantIncome

CoapplicantIncome

LoanAmount

Loan\_Amount\_Term

Loan Prediction

1.0

Dependents

1.0

Education

1.0

Self\_Employed

1.0

ApplicantIncome

4583.0

CospplicantIncome

1506.0

LoanAmount

128.0

Loan\_Amount\_Term

360.0

Credit\_History

1.0

Property\_Area

1.0

Submit

28°C Haze

Search

ENG IN

15:01 20-11-2022

Loan Prediction Category

127.0.0.1:5000/predict

ENG IN

13:01 20-11-2022

The predicted loan status is

1

[Goback](#)

27°C Haze

Search

ENG IN

13:01 20-11-2022

# **ADVANTAGES AND DISADVANTAGES**

## **ADVANTAGES:**

### **1.Flexibility:**

A bank loan allows one to repay as per convenience as long as the instalments are regular and timely. Unlike an overdraft where all the credit is deducted in go. Or a consumer credit card where the maximum limit cannot be utilised in one go.

### **2.Cost Effectiveness:**

When it comes to interest rates, bank loans are usually the cheapest option compared to overdraft and credit card.

### **3.Profit Retention:**

When you raise funds through equity you have to share profits with shareholders. However, in a bank loan raised finance you do not have to share profits with the bank.

### **4.Benefit of Tax:**

Government makes the interest payable on the loan a tax-deductible item when the loan has been taken for business purpose.

## **DISADVANTAGES:**

### **1.Hard Prerequisite:**

Since big finance from a bank is based on collateral, most young businesses will find it hard to finance the operations based on bank loan.

### **2.Irregular Payment Amounts:**

Over a long duration payback via monthly instalment might witness variation in the rate of interest. This means that the EMI will not be constant, rather it will change as per the influence of the market on the interest applicable.

## **CONCLUSION**

The analysis starts from data cleaning and processing missing value, exploratory analysis and finally model building and evaluation of the model. The best accuracy on public test set is when we get higher accuracy score and other performance metrics which will be found out. This paper can help to predict the approval of bank loan or not for a candidate.

## **FUTURE SCOPE**

### **ZERO PHYSICAL DOCUMENTATION:**

When you decide to borrow using an online loan app, you do away with visiting a bank or any other office. You can upload the documents required to avail of a loan on the loan app or website.

### **HASSLE-FREE APPLICATION:**

Many reputable banks, institutions, and companies offer instant loan with a hassle-free and time-efficient loan application process. PaySense offers instant personal loans with easy application. You just have to download the app or go to the PaySense website, sign up, upload the required documents, and your loan will get approved. You will get notified when your loan request is processed.

### **QUICK PROCESSING:**

Traditionally loan application used to take at least a few days. In some cases, the loan approval used to get stretched to over a month. But the development in the fintech sector and the onset of online loan apps have entirely changed it. The average loan approval time today is measured in minutes. With the PaySense instant loan app, you can get your loan approved within a few hours.

### **FLEXIBILITY OF LOAN AMOUNT AND TENURE:**

You get flexibility in choosing loan amount and loan duration when you take an instant personal loan. You may take a loan of Rs. 5,000 to Rs. 5,00,000 with online money loan apps. You can avail of any loan amount from this bracket depending upon your eligibility. The same thing goes with the term of the loan. You may take a loan for a term ranging from 3 months to 60 months with online loan apps.

### **ALL-TIME ACCESSIBILITY:**

Online loan apps offer a feature of all time accessibility. You just need to download an app from the Google play store or Apple app store, and you can apply for a loan any time of the day. You get access to money anytime you want with money loan apps.

### **CONVENIENCE:**

You don't need to go to a bank anymore and wait for hours to get a loan. With simple, easy-to-use online loan apps compatible with all the devices, you can conveniently get approval of a loan while sitting in your home. You just have to download the PaySense app and apply for a loan, which hardly takes any time or effort.

### **FLEXIBLE AND USAGE:**

Another benefit of availing of a loan from online loan apps is that you can use the money for your personal or business needs. You don't have to specify a reason for taking the loan. You can use the money of your instant loan towards buying a property or spend the

money for urgencies like a medical emergency. You don't have to give a reason to the lender for taking the loan.

#### **DIRECT TRANSFER TO YOUR BANK ACCOUNT:**

Often, even after getting your loan approved, the process of getting the loan amount transferred to you can take time and get complicated. But that is not the case with online loan apps that offer a direct transfer option.

## **APPENDIX**

#### **SOURCE CODE:**

<https://drive.google.com/drive/folders/1FsM1zdVXMHYIVacKpbF3ATWgFCHZjVtT?usp=sharing>

#### **GITHUB:**

<https://github.com/IBM-EPBL/IBM-Project-30793-1660189780>