SMART LENDER-APPLICANT CREDIBILITY PREDICTION FOR LOAN APPROVAL

1.INTRODUCTION

1.1 PROJECT OVERVIEW

Loan Prediction is very helpful for employee of banks as well as for the applicant also. The aim of this Paper is to provide quick, immediate and easy way to choose the deserving applicants. Dream housing Finance Company deals in all loans. They have presence across all urban, semi urban and rural areas. Customer first apply for loan after that company or bank validates the customer eligibility for loan. Company or bank wants to automate the loan eligibility process (real time) based on customer details provided while filling application form. These details are Gender, Marital Status, Education, Number of Dependents, Income, Loan Amount, Credit History and other. This project has taken the data of previous customers of various banks to whom on a set of parameters loan were approved. So the machine learning model is trained on that record to get accurate results. Our main objective of this project is to predict the safety of loan. To predict loan safety, the SVM and Naïve bayes algorithm are used. First the data is cleaned so as to avoid the missing values in the data set.

1.2 PURPOSE

Loan approval is a very important process for banking organizations. The system approved or reject the loan applications. Recovery of loans is a major contributing parameter in the financial statements of a bank. It is very difficult to predict the possibility of payment of loan by the customer. Using Machine learning we predict the loan approval.

2. LITERATURE SURVEY

We start our literature review with more general systematic literature reviews that focus on the application of machine learning in the general field of Banking Risk Management. Since the global financial crisis, risk management in banks has to take a major role in shaping decision-making for banks. A major portion of risk management is the approval of loans to promising candidates. But the black-box nature of Machine learning algorithms makes many loan providers vary the result. Martin Leo, Suneel Sharma and k. Maddulety's [1] extensive report has explored where Machine Learning is being used in the fields of credit risk, market risk, operational risk, and liquidity risk only to conclude that the research falls short of extensive research is required in the field.

TITLE : 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.

2.1 EXISTING PROBLEM

Problem : **Home loan application rejected by the bank:**
Many times banks reject your home loan application at the beginning itself. This happens generally because of several discrepancies in your forms and documents. Banks may reject your application if your personal information such as age, income and qualification isn't matching the lender's terms and conditions.

How to overcome this problem :
Before submitting your home loan application to the bank, always check whether you meet the eligibility criteria set up by the bank. Computing tools such E.M.I. calculator prove to be of great help to check various eligibility conditions.

Problem : **Pressure of paying down payment:**
Any home loan borrower has to generally pay around 20% of the property value as down payment. If the down payment is not done on time, the bank can reject the loan. The real problem arises when the sanctioned amount of loan is way below the expectation of the borrower. In such situation, the down payment increases resulting in added burden to the borrower.

How to overcome this problem:
By valuing the property in advance, you can plan down payment accordingly. Also, banks do accept LIC policies, PF and national security certificates as down payments.

2.2 REFERENCES

Ashlesha Vaidya [2] used logistic regression as a probabilistic and predictive approach to loan approval prediction. The author pointed out how Artificial neural networks and Logistic regression are most used for loan prediction as they are easier comparatively develop and provide the most accurate predictive analysis. One of the reasoning behind this that that other Algorithms are generally bad at predicting from non-normalized data. But the nonlinear effect and power terms are easily handled by Logistic regression as there is no need for the independent variables on which the prediction takes place to be normally distributed.

Similar to PCA, Zaghdoudi, Djebali & Mezni [4] compared the use of Linear Discriminant Analysis versus Logistic Regression for Credit Scoring and Default Risk Prediction for foreseeing default risk on small and medium enterprises. Linear Discriminant Analysis (LDA) is like PCA for dimensionality reduction but instead of looking for the most variation, LDA focuses on maximizing the separability among the know categories. This subspace that well separates the classes is usually in which a linear classifier can be learned. The classification of those enterprises correctly in their original groups through both these methods was inconsequential with Logistic regression having a 0.3% better accuracy score than LDA.

2.3 PROBLEM STATEMENT DEFINITION

> Salary: Applicants with high income should have more chances of getting approval.

> Previous history: Applicants who have paid their historical debts have more chances of getting approval.

> Loan amount: Less the amount higher the chances of getting approval.

> Loan term: Less the time period has higher chances of approval.

> EMI: Lesser the amount to be paid monthly, the higher the chances of getting approval.

3. IDEATION AND PROPOSED SOLUTION

3.1 EMPATHY MAP CANVAS

Empathy are a powerful tool for helping terms better understand users.But that does not mean they are perfect.You should feel free to adapt them to your circumstances.

An empathy map is a collaborative tool that teams can use to better understand their customers. It consists of an image of the customer surrounded by six sections. These sections are Think and feel,Hear,See,Say and do,Pain,Gain.

Usage flexibility

You can use this personal loan amount for any of your needs. It can be education, home improvement, a vacation, buying a gadget and more. The loan amount can be used for flexible purposes.

Quick processing

You can get the loan amount within a day due to the quick processing of a pre-approved loan. This happens because, in most of the cases, the customer has passed the eligibility requirements of the loan offer. If all the given information matches, your loan will be processed very quickly.

3.2 IDEATION AND BRAINSTORMING

Ideation is the second stage of the design thinking process where participants in a design thinking workshop come up with ideas on how to solve a specific user problem. The design thinking process is made up of three phases: empathize, ideate, and prototype.

The ideation phase of design thinking is guided by the user problems that were defined during the empathize phase. Ideation is about the exploration and identification of potential solutions. Not all ideas will be viable solutions, and that's okay. The primary goal of ideation is to spark creativity and innovation.

If a bank promises to refund the processing fee, then, do take a written declaration from the bank for the same so that if your application gets rejected afterwards, you can still claim the refund of processing fee charged.

The loan amount sanctioned by the bank mostly varies then the expected amount of the buyer. Bank calculates the amount a buyer can repay taking into consideration his income, live loans, financial history etc. This calculated amount gets reflected in the sanctioned loan amount. The buyer can enhance his eligibility for a home loan by applying on a co-applicant basis.

3.3 PROPOSED SOLUTION

Insufficient Income.

Not having steady source of income. A steady source of income is the proof for any lender that you are capable of repaying your personal loan.

Banks have set a minimum monthly salary cap.

In the absence of a job, ensure that you have other sources of income. Avoid applying for a loan when you have just switched jobs.

Interest payouts or rental income, that will help you make EMI payments.

3.4 PROBLEM SOLUTION FIT

1. Refusal of a large insurance claim
2. Creating hindrance to the main source of income.
3. Diminishing deposit balance.
4. Inability to pay the debt of creditors other than the bank.
5. Non repayment of the loan installments.

Available solutions

1. First of all identify the solutions for their problems.
2. Customer wants to increase their income.
3. Make a budget to help you resolve their financial problems.
4. Avoid buying new
5. Customer meet their advisor to discuss their.

4. REQUIREMENT ANALYSIS

Solution requirements describe specific characteristics that a product must have to meet the needs of the stakeholders and the business itself. They fall into two large groups.

Functional requirements define what a product must do, what its features and functions are.

Nonfunctional requirements describe the general properties of a system. They are also known as quality attributes.

4.1 Functional Requirements: These are the requirements that the end user specifically demands as basic facilities that the system should offer. All these functionalities need to be necessarily incorporated into the system as a part of the contract. These are represented or stated in the form of input to be given to the system, the operation performed and the output expected. They are basically the requirements stated by the user which one can see directly in the final product, unlike the non-functional requirements.

FR No.	Functional Requirement	Sub Requirement
FR-1	User Registration	Registration through Form Registration through Gmail Registration through LinkedIN
FR-2	User Confirmation	Confirmation via Email Confirmation via OTP
FR-3	Loan Amount	Loan amount in thousands

FR-4	Property Area	Urban/ Semi-Urban/ Rural
FR-5	Loan Amount Term	Term of a loan in months
FR-6	Credit History	Credit history meets guidelines

4.2 Non-functional requirements: These are basically the quality constraints that the system must satisfy according to the project contract. The priority or extent to which these factors are implemented varies from one project to other. They are also called non-behavioral requirements.

They basically deal with issues like:

Portability, Security, Maintainability, Reliability, Scalability,Performance,Reusability.

FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	Application designed for online banking must be easy to use and on-line help should be incorporated.
NFR-2	**Security**	Application must be able to send or receive the information to or from the server and client in an encrypted way which ensures that system must be protected.
NFR-3	**Reliability**	It reflects the capacity of the software to maintain its performance over the time, how well the system performs in peak hours and how effortlessly it handles the bugs.
NFR-4	**Performance**	It alludes to the capacity of the system to process as many as transactions per second as submitted to it without failure.
NFR-5	**Availability**	It should be available round the clock and means for how long the system is available for its users or clients and for how long the system will be operational.
NFR-6	**Scalability**	It should be operational on any hand held gadgets and it ought to have the capacity to show the different types of currencies used in distinctive nations.

5 PROJECT DESIGN

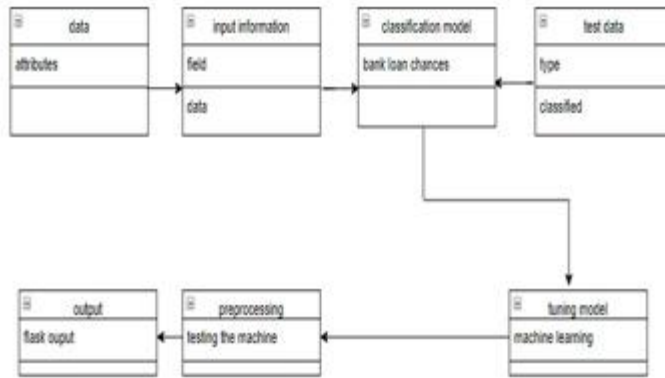## 5.1 DATA FLOW DIAGRAM
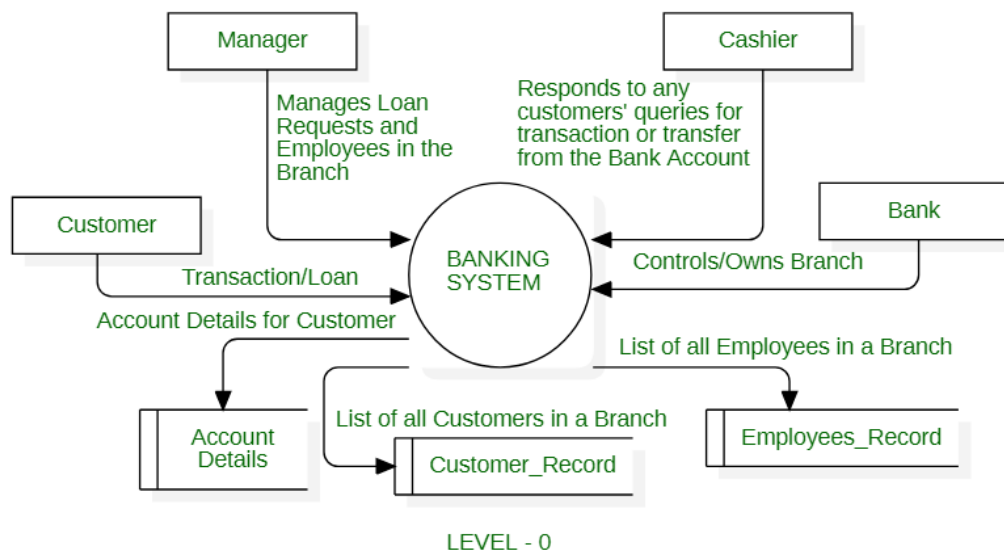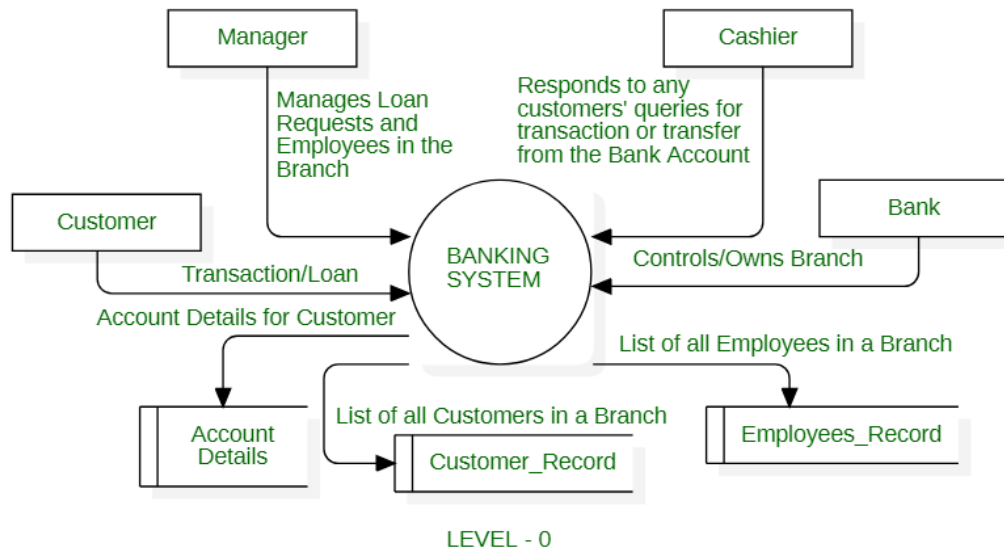
LEVEL - 0

LEVEL - 0

5.2 SOLUTION & TECHNICAL ARCHITECTURE
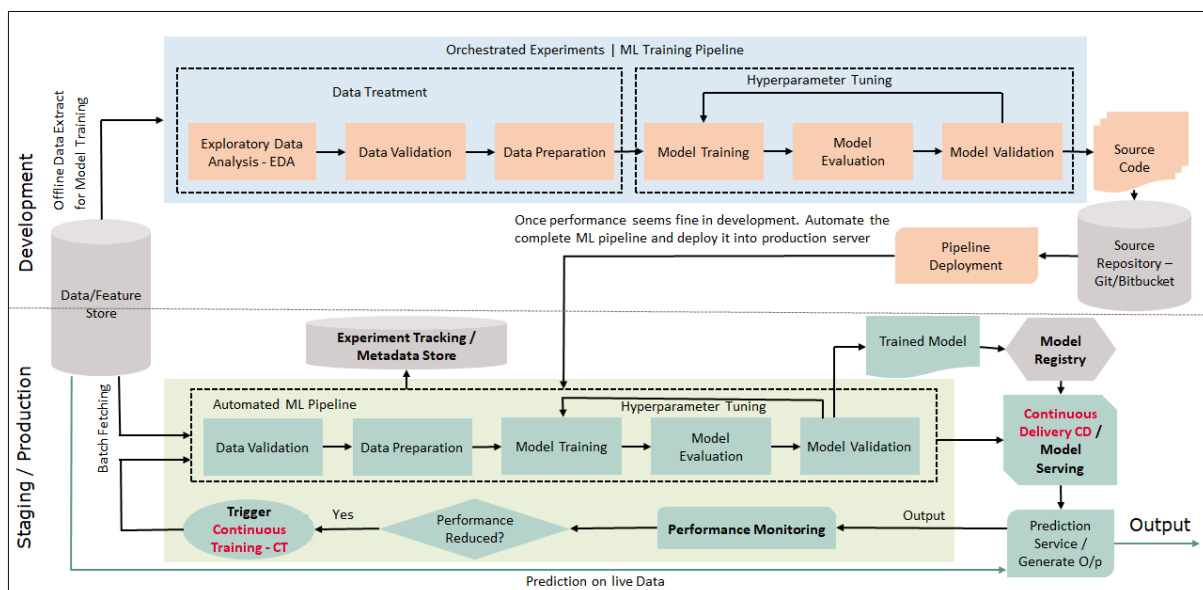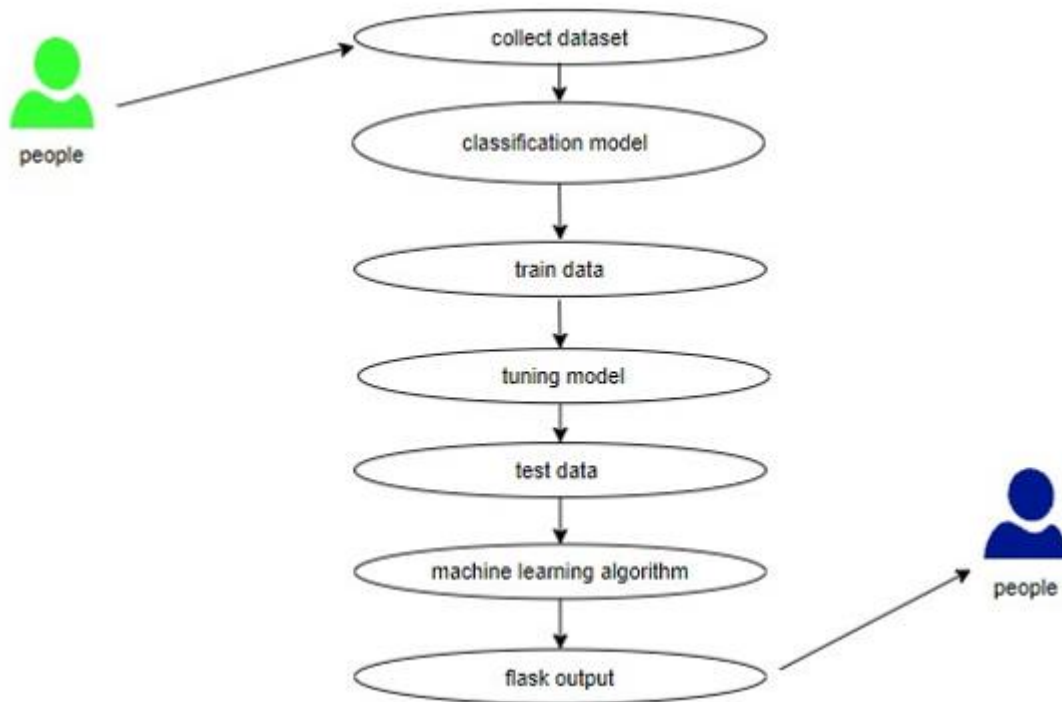
Use case diagrams are used for high level requirement analysis of a system. So, when analysing the requirements of a system, the functionalities are captured in use cases. So, uses cases are nothing, but the functionalities of the system written in an organized manner.
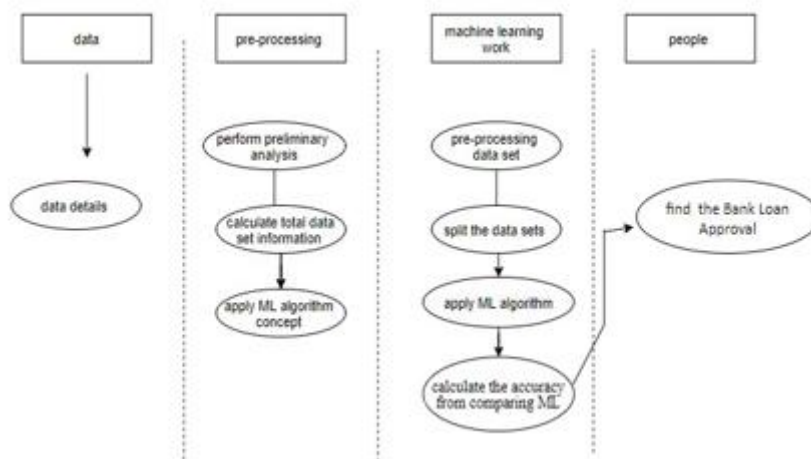
A. Class Diagram

Class diagram is generally a graphical representation of the static view of the system and represents different aspects of the application. A collection of class diagrams will represent the whole system.

B Activity Diagram

Activity diagrams not only visualize the dynamic nature of a system, but they are also used for constructing executable system by using forward and reverse engineering techniques. Activity diagram is some time considered as the flow chart, but it is not.

Sequence Diagram
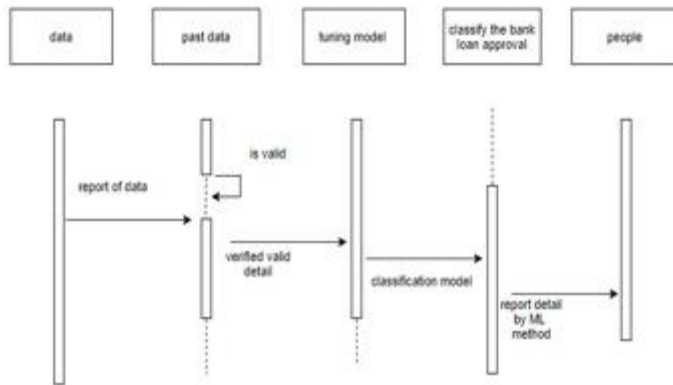
Sequence diagrams model the flow of logic within our system in a visual manner, enabling both to document and validate our logic, and are commonly used for both analysis and design purposes. Sequence diagrams are the most popular UML artifact for dynamic modelling, which focuses on identifying the behaviour within the system.

D. Entity Relationship Diagram
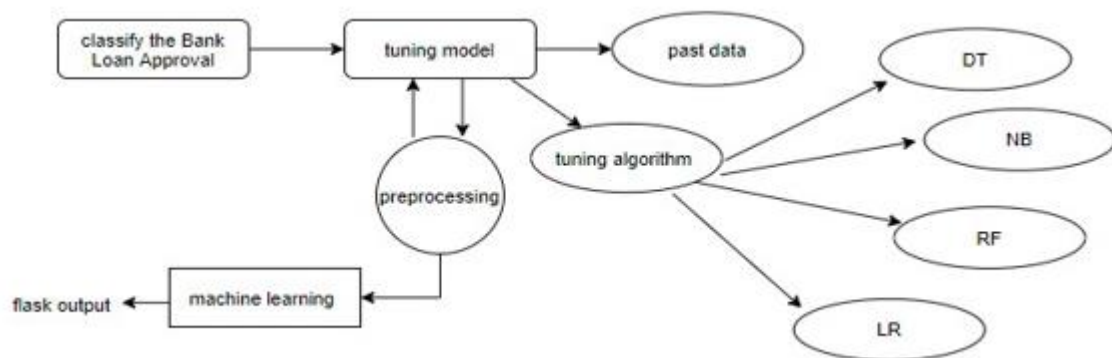
An entity relationship diagram (ERD) is a graphical representation of an information system that depicts the relationships among people, objects, places, concepts or events within that system. An Entity relationship model is a data modelling technique that helps define business processes and can be used as the foundation for a relational database.

6. PROJECT PLANNING AND SCHEDULING

6.1 SPRINT PLANNING AND ESTIMATION

Sprint planning is an event in scrum that kicks off the sprint. The purpose of sprint planning is to define what can be delivered in the sprint and how that work will be achieved. Sprint planning is done in collaboration with the whole scrum team.

In scrum, the sprint is a set period of time where all the work is done. However, before you can leap into action you have to set up the sprint. You need to decide on how long the time box is going to be, the sprint goal, and where you're going to start. The sprint planning session kicks off the sprint by setting the agenda and focus. If done correctly, it also creates an environment where the team is motivated, challenged, and can be successful. Bad sprint plans can derail the team by setting unrealistic expectations.

1. **The sprint goal:** A short written summary of what the team plans to accomplish in the next sprint.
2. **The sprint backlog:** The list of stories and other product backlog items the team has agreed to work on in the upcoming sprint.

6.2 SPRINT DELIVERY SCHEDULE

The objective of sprint planning is to work out the key details regarding the team's planned work during the next sprint. With that in mind, the sprint team should plan to address at least the following issues during this meeting.

1. Decide on the team's overall strategic objective for the next sprint. (This will be represented as the one- or two-sentence sprint goal.)
2. Review the product backlog and discuss which items belong on the next sprint backlog and why.
3. Call for a team consensus on the proposed sprint goal and backlog items (led by the scrum master).
4. Discuss team capacity.
5. Discuss known issues that could disrupt or slow progress on the sprint backlog.
6. Assign the new sprint backlog's tasks, according to skill sets, capacity, and other relevant criteria.
7. Estimate the timeframes for each of the tasks assigned and agree on what "done" will look like for each item.
8. Confirm the timeframe of the upcoming sprint.
9. Open the meeting to sprint-related questions. (The product owner should be responsible for coordinating this step, to ensure the discussion stays on track.)

6.3 REPORTS FROM JIRA

The Sprint Report is board-specific – that is, it will only include issues that match your board's saved filter.

Issues added after the sprint starts are indicated with an asterisk.

An issue is considered to be 'To Do' when it is in a status that has been mapped to the left-most column of your board. Similarly, an issue is considered to be 'Done' when it is in a status that has been mapped to the

right-most column of your board. See <u>Configuring columns</u> for more information.

The grey line is a guideline that's drawn from the total estimate of the issues at the start of the sprint, to 0 at the end of the sprint. The grey line stays flat during non-working days.

The red line represents actual work done during the sprint — it's drawn from the total estimate of the issues at the start of the sprint, and it goes higher or lower as your team works on the issues in the sprint. Note that the red line shows the current total estimate for unresolved issues at any point in the sprint. It also reflects issues that are added or removed from the sprint.

If your board is configured to <u>track remaining estimates and time spent</u>, a green line will display in the Sprint Report, indicating work logged on issues.

You can also see the Opsgenie alert chart for your Jira Software project. The project admin can enable **Display alert counts in sprint report** by navigating to **Project settings** > **Opsgenie**. Read more about <u>working with Opsgenie</u>.

7. CODING AND SOLUTIONING

7.1 FEATURE 1

1. Accessibility

An organization looking to build loan software may not have enough on-premise infrastructure capacities to ensure its non-disruptive operation, updates, and support. Scaling during peak workloads and handling an increase in the number of users and subscriptions may also be quite challenging. Using **cloud infrastructure** is best to ensure optimal scalability and availability.

2. Servicing different loan types

The more types of loans your money lending software is capable of servicing, the better. Lending applications that have a wide range of use cases, will surely attract more users than apps targeting only one specific loan type. A loantech software to create loan app estimation, for example, may have a broad range of applications from student loan tech calculations to estimating business loans and mortgages.

3. Centralized data storage

Every stage of the lending process involves working with customer data. The best loan servicing software stores this data in centralized storage accessible during every loan processing stage. A legacy loan management system, on the other hand, uses a siloed approach to data storage, which makes loan processing more laborious and lengthy.

1. Integrated credit assessment capabilities

Modern loan servicing software for private lenders should be able to instantly connect with credit bureaus and any other bodies responsible for credibility assessment. Such platforms should receive regular credit data updates and leverage big data analytics to assess the trustworthiness of applicants. The client's social media activity, for example, can be a valid source of alternative assessment of credibility.

2. Automation of routine processes

Using robotic process automation to streamline simple rule-based processes is another must-have feature of a loan management platform. Automation accelerates loan origination and processing and accounts for increasing client satisfaction. On top of that, it helps to avoid human error.

3. In-built analytic modules

Leveraging artificial intelligence (AI) and big data is another hallmark of excellent loan servicing software for lenders. Not only does it help to generate reports but also enables companies to evaluate market trends, detect patterns in customer behavior and come up with new products and offerings.
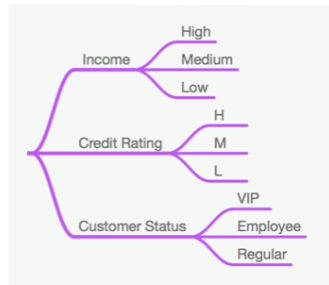
4. Third-party integration

Another feature that most organizations find especially attractive in a loan processing system, is its capability to integrate with other enterprise software. ERP and CRM solutions are capable of enriching the lending system with data and insights. Systems integrating lending modules with software for remote sales personnel are also enjoying high popularity among lenders.
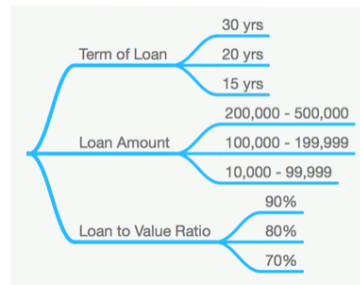
5. Security

Finance company software works with classified and highly sensitive data, and for both lenders and customers, security is a matter of paramount importance. An excellent lending system should possess advanced security capabilities, and ensure the highest level of customer, data, and network protection.
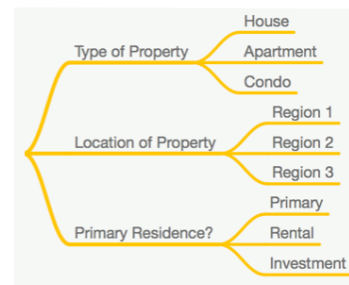
8. TESTING

Someone ...

- Income
  - High
  - Medium
  - Low
- Credit Rating
  - H
  - M
  - L
- Customer Status
  - VIP
  - Employee
  - Regular

... applies for a loan...

- Term of Loan
  - 30 yrs
  - 20 yrs
  - 15 yrs
- Loan Amount
  - 200,000 - 500,000
  - 100,000 - 199,999
  - 10,000 - 99,999
- Loan to Value Ratio
  - 90%
  - 80%
  - 70%

... to buy a property.

- Type of Property
  - House
  - Apartment
  - Condo
- Location of Property
  - Region 1
  - Region 2
  - Region 3
- Primary Residence?
  - Primary
  - Rental
  - Investment

Each time someone applies for a loan, nine different test conditions are included in the test scenario. Even in this over-simplified example, there are close to 20,000 possible scenarios. In this context, we want to test this loan application process relatively thoroughly - with a manageable number of tests.

We know that <u>testing each item in our system once is not sufficient;</u> we know that interactions between the different things in our system (such as a particular credit rating range interacting with a specific type of property, for example) could well cause problems. Similarly, we know that the written requirements document will be incomplete and will not identify all of those potentially troublesome interactions for us. As thoughtful test designers, we want to be smart and systematic about testing for potential problems caused by interactions without going off the deep end and trying to test every possible combination.

HEXAWISE	K) Mortgage Application Process [Mixed-Strength Testing] ▼		
Parameters			
Rules			
Scenarios	CUSTOMER DETAILS (1)	****	
Scripts	Income (3)	High	Medium	Low
Analysis	Credit Rating (3)	H	M	L
Review	Customer Status (3)	VIP	Employee	Regular
Export	LOAN DETAILS (1)	****	

Parameters table:

	Col 1	Col 2	Col 3
CUSTOMER DETAILS (1)	****		
Income (3)	High	Medium	Low
Credit Rating (3)	H ⁂ 3	M ⁂ 3	L ⁂ 3
Customer Status (3)	VIP	Employee	Regular
LOAN DETAILS (1)	****		
Term of Loan (3)	30 yrs	20 yrs	15 yrs
Loan Amount (3)	200,000 - 500,000	100,000 - 199,999	10,000 - 99,999
Loan-to-Value Ratio (3)	90%	80%	70%
PROPERTY DETAILS (1)	****		
Type of Property (3)	House	Apartment	Condo
Location of Property (3)	Region 1 ⁂ 4	Region 2 ⁂ 4	Region 3 ⁂ 4
Primary Residence? (3)	Primary	Rental	Investment

Asking the newspaper questions described above is useful to understand potential ways the system under test might behave.

Once we have decided which test conditions are important enough to include in this model (and excluded things - like "First Name" and "Last Name" in this example - that will not impact how the system being tested operates), Hexawise makes it quick and easy to systematically create powerful scenarios that will allow us to maximize our test execution efficiency.

Once we enter our parameters into Hexawise, we simply click on the "Scenarios" link in the left navigation pane.

9. RESULTS

	Loan_ID	Loan_Status
0	LP001015	Y
1	LP001022	Y
2	LP001031	Y
3	LP001035	N
4	LP001051	Y
5	LP001054	N
6	LP001055	Y
7	LP001056	N
8	LP001059	Y
9	LP001067	Y
10	LP001078	N
11	LP001082	Y
12	LP001083	Y
13	LP001094	N
14	LP001096	Y
15	LP001099	Y
16	LP001105	N
17	LP001107	Y
18	LP001108	Y
19	LP001115	N
20	LP001121	Y
21	LP001124	Y
22	LP001128	Y
23	LP001135	Y
24	LP001149	N

10. ADVANTAGES :

The loan is not repayable on demand and so available for the term of the loan - generally three to ten years - unless you breach the loan conditions.

Loans can be tied to the lifetime of the equipment or other assets you're borrowing the money to pay for.

At the beginning of the term of the loan you may be able to negotiate a repayment holiday, meaning that you only pay interest for a certain amount of time while repayments on the capital are frozen.

While you must pay interest on your loan, you do not have to give the lender a percentage of your profits or a share in your company.

Interest rates may be fixed for the term so you will know the level of repayments throughout the life of the loan.

There may be an arrangement fee that is paid at the start of the loan but not throughout its life. If it is an on-demand loan, an annual renewal fee may be payable.

DISADVANTAGES:

Larger loans will have certain terms and conditions or covenants that you must adhere to, such as the provision of quarterly management information.

Loans are not very flexible - you could be paying interest on funds you're not using.

You could have trouble making monthly repayments if your customers don't pay you promptly, causing cashflow problems.

In some cases, loans are secured against the assets of the business or your personal possessions, eg your home. The interest rates for secured loans may be lower than for unsecured ones, but your assets or home could be at risk if you cannot make the repayments.

There may be a charge if you want to repay the loan before the end of the loan term, particularly if the interest rate on the loan is fixed.

11. CONCLUSION

So here, it can be concluded with confidence that the Naïve Bayes model is extremely efficient and gives a better result when compared to other models. It works correctly and fulfills all

requirements of bankers. This system properly and accurately calculate the result. It predicts the loan is approve or reject to loan applicant or customer very accurately.

12. FUTURE SCOPE:

Provide more insights and data driven decision for stakeholders.

Clean and neat visuals.

Clean and process data properly.

Achieve higher accuracy.

Gain domain knowledge and deploy solution properly.

13. APPENDIX

https://github.com/IBM-EPBL/IBM-Project-12159-1659438917.git