# Chapter 1

**INTRODUCTION TO PREDICTION OF LIVER DISEASE**

This chapter describes about the Machine learning, problem statement of the project, objective of the project, Existing System, Proposed System and Advantages.

# Introduction to Prediction of Liver Disease

The liver plays an important role in many bodily functions from protein production and bloodclotting to cholesterol, glucose (sugar), and iron metabolism. It has a range of functions,including removing toxins from the body, and is crucial to survival. The loss of those functions can cause significant damage to the body. When liver is infected with a virus injured by chemicals, or under attack from own immune system, the basic danger is the same that liver will become so damaged that it can no longer work to keep a person alive.

The point of this task is to some degree diminish the time delay caused because of the superfluous forward and backward transporting between the healing centre and the pathology lab. For this situation, a machine learning calculation will be prepared to foresee a liver ailment in patients. This is by all accounts a great case of managed learning.Our point will be to prepare an assortment of supervised learning and unsupervised Learning calculations on this information, so that, when another information point emerges, our best performing classifier can be utilized to classify the information point as a positive case or negative.To analyze the conceivable issue in view of the test report esteems. The specialists can enter the patient's report as information. Utilizing the framework we are foreseeing if the patient is having a liver infection or not by examining the accessible dataset.We are catching the test report contribution from UI and store it in a table. We have made a table with the current dataset which has every one of the records. We have utilized Naïve Bayes, Support Vector Machine, to make a model. In light of the model and the test information we are anticipating if the patient can have Liver Disease and the precision of the ailment.

## Introduction to Machine Learning

Machine learning is a branch of Artificial Intelligence(AI) which is heavily used in the field of data science. It hasa strong potential in health-related data analysis for automated disease prediction. Machine learning technology is widely used these days

in various fields such as WiFi-hot spot detection, diagnosis of heart disease, diagnosis of cancer, diagnosis of tumor detection, prediction of type of cancer, etc. Machine learning is a field of artificial intelligence (AI) that enables the system to learn by itself without being explicitly programmed.According to, it is a field of study that gives computers the ability to learn without programming them explicitly. Unsupervised machine learning is the method of inferring a function that defines the structure of unlabeled data. The main objective of unsupervised learning is to draw inferences from input data and model the structure based on that data.

# Overview of Project

The work focuses on three different machine learning techniques, i.e., Naïve Bayes, K- means, and Support Vector Machine, propagation to compare their prediction accuracy and computational complexity. The study concentrates on liver disease-related health care data set and used for comparative performance measurement of the three techniques.

The utilization of medicinal datasets has pulled in the consideration of specialists around the world. Machine Learning methods have been broadly utilized as a part of creating choice emotionally supportive networks for ailments forecast through an arrangement of therapeutic datasets. Grouping systems have been broadly utilized as a part of the restorative field for exact order than an individual classifier. Liver malady is a sort of harm to or illness of the liver. In this task, I have taken the datasets of general Indian liver ailment patient's records to help basic leadership. Indian Liver Patient's datasets demonstrate that proposed technique amazingly enhances the illnesses expectation precision.

# Problem Statement

The Liver Disease Detection Problem includes modeling past liver disease patient data with the knowledge of the ones that has liver disease or not. This model is then used to identify whether a new person is prone to get liver disease or not. Our aim here is to detect 100% of the liver disease prediction.

# Key Features of Machine Learning

1. **Classification**: A set of numeric features can be conveniently described by a feature vector. An example of reaching a two-way classification from a feature vector (related to the perceptron) consists of calculating the scalar product between the feature vector

and a vector of weights, comparing the result with a threshold, and deciding the class based on the comparison.

1. [**Character Recognition**](https://en.wikipedia.org/wiki/Character_recognition): features may include [histograms](https://en.wikipedia.org/wiki/Histogram) counting the number of black pixels along horizontal and vertical directions, number of internal holes, stroke detection and many others.
2. [**Speech Recognition**](https://en.wikipedia.org/wiki/Speech_recognition)**:** features for recognizing [phonemes](https://en.wikipedia.org/wiki/Phonemes) can include noise ratios, length of sounds, relative power, filter matches and many others.
3. [**Spam**](https://en.wikipedia.org/wiki/Spam_(electronic)) **Detection:** features may include the presence or absence of certain email headers, the email structure, the language, the frequency of specific terms, the grammatical correctness of the text.
4. **Selection and Extraction:** [Extracting](https://en.wikipedia.org/wiki/Feature_extraction) or [selecting](https://en.wikipedia.org/wiki/Feature_selection) features is a combination of art and science developing systems to do so is known as [feature engineering.](https://en.wikipedia.org/wiki/Feature_engineering) It requires the experimentation of multiple possibilities and the combination of automated techniques with the intuition and knowledge of the [domain expert](https://en.wikipedia.org/wiki/Domain_expert). Automating this process is [feature learning](https://en.wikipedia.org/wiki/Feature_learning), where a machine not only uses features for learning, but learns the features itself.

# Aim of the Project

The main objective of this research is to use classification algorithms to identify the liver patients from healthy individuals. In this study,Three classification algorithms Support Vector Machines (SVM), K-mean algorithm and Naive Bayes have been considered for comparing their performance based on the liver patient data. Further, the model with the highest accuracy is implemented as a user friendly Graphical User Interface (GUI). The GUI can be readily utilized by doctors and medical practitioners as a screening tool for liver disease.

# Objective of the Project

To solve the problems facing physicians in diagnosis of liver diseases. Experience has shown that many patients suffering from liver disorder die daily as a result of misdiagnosis of the diseases.

Early prediction of liver disease is very important to save human life and take proper steps to control the disease. This research work explores the early prediction of liver disease using deep learning techniques. The liver disease dataset which is select

for this study is consisting of attributes like total bilirubin, direct bilirubin, age, gender, total proteins, albumin and globulin ratio.

Following are the objectives performed by the system:

* The application must provide user interface for doctors input object of the prescription.
* The application should have the capability for preprocessing of the given input.
* The System should be capable to detect the chances of lever disease using past patient data.

# Existing System

In existing system to analyze the conceivable issue in view of the test reportesteems. The specialists can enter the patient's report as information. Utilizing the framework we are foreseeing if the patient is having a liver infection or not by examining the accessible dataset. Some degree diminish the time delay caused because of the superfluous forward and backward transporting between the healing centre and the pathology lab. **Disadvantages of Existing System:**

* There is no real time identity verification.
* Not using data-mining technique to detect chances of getting lever disease.
* Time intense.

# Proposed System

Work has been done in recognizing the beginning of sicknesses like coronary illness. For this situation, a machine learning calculation will be prepared to fore see a liver ailment in patients. This is by all accounts a great case of managed learning. We have been given a settled number of highlights for every datum point, and our point will be to prepare an assortment of supervised learning and unsupervised Learning calculations on this information Three classification algorithms Support Vector Machines (SVM), K-mean algorithm and Naive Bayes have been considered for comparing their performance based on the liver patient data. Further, the model with the highest accuracy is implemented as a user friendly Graphical User Interface (GUI). The GUI can be readily utilized by doctors and medical practitioners as a screening tool for liver disease.

### Advantages of Proposed System:

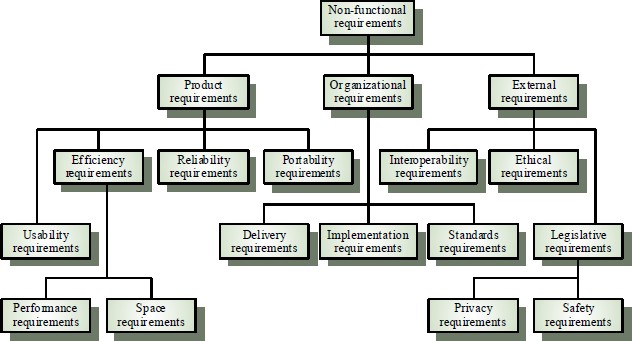
* Proposed system will allow people to get early prediction of getting chances of lever disease.
* It didn’t require medical expertise.
* Early prediction of liver disease can save human life and take proper steps to control the disease.

# Functional Requirements

A function of software system is defined in functional requirement and the behavior of the system is evaluated when presented with specific inputs or conditions which may include calculations, data manipulation and processing and other specific functionality.

# Non-functional Requirements

Nonfunctional requirements describe how a system must behave and establish constraints of its functionality.Attributes such as performance, security, usability, compatibility are not the feature of the system, they are a required characteristic. They are "developing" properties that emerge from the whole arrangement and hence we can't compose a particular line of code to execute them. Any attributes required by the customer are described by the specification. We must include only those requirements that are appropriate for our project.



### Figure 1.1 Non Functional Requirements

Figure 1.2 Non functional requirements shows details of all types of requirements in a hirachical form.

**Some Non-Functional Requirements are as follows:**

* **Reliability:** The structure must be reliable and strong in giving the functionalities. The movements must be made unmistakable by the structure when a customer has revealed a couple of enhancements. The progressions made by the Programmer must be Project pioneer and in addition the Test designer.
* **Maintainability:** The system watching and upkeep should be fundamental and focus in its approach. There should not be an excess of occupations running on diverse machines such that it gets hard to screen whether the employments are running without lapses.
* **Performance:** The framework will be utilized by numerous representatives all the while. Since the system will be encouraged on a single web server with a lone database server outside of anyone's ability to see, execution transforms into a significant concern. The structure should not capitulate when various customers would use everything the while. It should allow brisk accessibility to each and every piece of its customers. For instance, if two test specialists are all the while attempting to report the vicinity of a bug, then there ought not to be any irregularity at the same time.
* **Portability:** The framework should to be effectively versatile to another framework. This is obliged when the web server, which s facilitating the framework gets adhered because of a few issues, which requires the framework to be taken to another framework.
* **Scalability:** The framework should be sufficiently adaptable to include new functionalities at a later stage. There should be a run of the mill channel, which can oblige the new functionalities.
* **Flexibility:** Flexibility is the capacity of a framework to adjust to changing situations and circumstances, and to adapt to changes to business approaches and rules. An adaptable framework is one that is anything but difficult to reconfigure or adjust because of diverse client and framework prerequisites. The deliberate division of concerns between the trough and motor parts helps adaptability as just a little bit of the framework is influenced when strategies or principles change.

# Summary

This chapter gives the introduction to how the Liver disease is diagnosed, overview of the project, problem statement of the project, objective of the project, Existing system, proposed system, functional requirements and non-functional requirements.

# Chapter 2

**LITERATURE SURVEY ON LIVER DISEASE PREDICTION**

This chapter gives a clear history of machine learning techniques which are used in this project.

# Introduction

To analyze the conceivable issue in view of the test report esteems. The specialists can enter the patient's report as information. Utilizing the framework, we are foreseeing if the patient is having a liver infection or not by examining the accessible dataset. The datasets are taken from UCI vault. We are catching the test report contribution from UI and store it in a table. We have made a table with the current dataset which has every one of the records. We have utilized , Naïve Bayes, Support Vector Machine, k-Means to make a model.

# Machine Learning Techniques

## K-Means Algorithm

K-means is a method of clustering observations into a specific number of disjoint clusters. The K” refers to the number of clusters specified. Various distance measures exist to determine which observation is to be appended to which cluster. The algorithm aims at minimizing the measure between the centroid of the cluster and the given observation by iteratively appending an observation to any cluster and terminate when the lowest distance measure is achieved.

* + The sample Dataset is initially partitioned into K clusters and the observations are randomly assigned to the clusters.

For each sample:

* + Calculate the personnel details of customer to the centroid of the cluster.
  + IF the sample data is closer to its own cluster THEN Accept ELSE or reject
  + Repeat steps 1 and 2 until no observations are moved from one cluster to another
  + When step 3 terminates the clusters are stable and each sample is assigned a cluster which results in the lowest possible distance to the centroid of the cluster.

A Person or Medical expert wants to analyze the data in order to know which person is prone to get liver disease or not. There is an enormous amount of data which one can actually retrieved from the health sector. The data can actually be used to detect the chance of liver disease. The proposed system categorizes chance of getting liver disease based on some personal information. Detection is then carried out to know the liver disease. The system takes all this information first as the input and points out person who are prone to get liver disease. A training dataset is being prepared beforehand based upon the person previous history. The system uses the concept of k-means clustering to cluster the person based upon the similarities or the patterns they share among each other naïve baye’s and SVM used to classify and detect the chance of getting liver disease with the training dataset and testing dataset and after classifying it will detect the liver disease.

K-means is a method of clustering analysis into a precise number of disjoint clusters. The K-refers to the total of clusters specified. Different distance events survive to resolve which observation is to be appended to which cluster.

* + A cluster of information items can be deal with as one gathering.
  + While doing bunch examination, we first parcel the arrangement of information into bunches in view of information comparability and afterward dole out the names to the gatherings.
  + The principle preferred standpoint of bunching over arrangement is that, it is versatile to changes and helps single out valuable highlights that recognize distinctive gatherings

## Distance measures

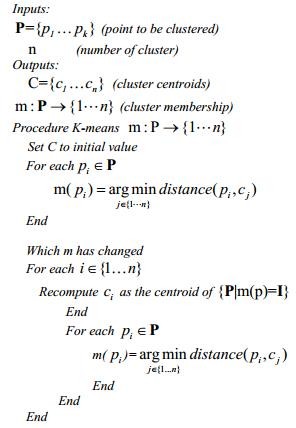
Common distance measures include the Euclidean distance, the Euclidean squared distance and the Manhattan or City distance.

The Euclidean measure corresponds to the shortest geometric distance between to points.

A faster way of determining the distance is by use of the squared Euclidean distance whichcalculates the above distance squared, i.e.



The specific k-Means algorithm is defined as follows



## Naïve Bayes

Classification techniques are maximum ideal for predicting or describing facts sets with binary or nominal classes. They are less effective for ordinal classes because they do now not consider the implicit order among the classes. Other kinds of relationships, which includes the subclass–super class relationships among classes also are unnoticed. The remainder of this bankruptcy focuses best on binary or nominal magnificence labels.

A class technique is a scientific method to constructing type fashions from an enter information set. Examples encompass choice tree classifiers, rule-based totally classifiers, neural networks, aid vector machines, and Naive Bayes classifiers. Each approach employs a mastering set of rules to pick out a model that quality fits the connection among the attribute set and sophistication label of the enter records. The version generated by using a studying algorithm ought to each suit the input records properly and efficiently are expecting the magnificence labels of facts it has never visible earlier than. Therefore, a key goal of the gaining knowledge of set of rules is to construct fashions with correct generalization capability, models that accurately expect the elegance labels of previously unknown facts. The tree has 3 styles of nodes:

* + A root node that has no approaching edges and at least zero active edges
  + Internal node, each of which has precisely one approaching edge and at least two active edges.
  + Leaf or terminal node, each of which has precisely one approaching edge and no friendly edges.

Naive Bayes is a simple technique for constructing classifiers: models that assign class labels to problem instances, represented as vectors of [feature](https://en.wikipedia.org/wiki/Feature_vector) values, where the class labels are drawn from some finite set. It is not a single [algorithm](https://en.wikipedia.org/wiki/Algorithm) for training such classifiers, but a family of algorithms based on a common principle: all Naive Bayes classifiers assume that the value of a particular feature is [independent](https://en.wikipedia.org/wiki/Independence_(probability_theory)) of the value of any other feature, given the class variable.

* In this progression the grouping calculations assemble the classifier.
* The classifier is worked from the preparation set made up of database tuples and their related class marks.
* Each tuple that constitutes the preparation set is alluded to as a classification or class. These tuples can likewise be alluded to as test, question or information focuses.
* above diagram shows that the liver disease is detected based on person personal details.

## Support Vector Machine

Support Vector Machine was first found by Vapnik in 1979. It was again recommended by Vapnik in 1995 for regression and classification. Support vector can be used for pattern classification which has multilayer perceptron and radial-basis function networks. The SVM is the advanced technology with maximum classification algorithms embedded in statistical learning theory. SVM methods are used in classification of linear and non-linear data. It transforms the original training data into higher dimension using non-linear mapping. Within this new dimension it searches for linear optimal separating hyperplane. Data from two classes can be separated by hyperplane with an appropriate nonlinear mapping to a sufficiently high dimension. Using support vectors and margins the SVM finds these hyperplane. SVM implements the classification task by maximizing the margin classifies both class while minimizing the classification errors. Although the SVM can be applied to various optimization problems such as regression, the classic problem is that of data classification. The data

points are identified as being positive or negative, and the problem is to find a hyperplane that separates the data points by a maximal margin.

* + Support Vector Machine” (SVM) is a supervised [machine learning](https://courses.analyticsvidhya.com/courses/introduction-to-data-science-2?utm_source=blog&utm_medium=understandingsupportvectormachinearticle)  [algorithm](https://courses.analyticsvidhya.com/courses/introduction-to-data-science-2?utm_source=blog&utm_medium=understandingsupportvectormachinearticle) which can be used for both classification or regression challenges. However, it is mostly used in classification problems. In this algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiate the two classes very well (look at the below snapshot).
  + Support Vectors are simply the co-ordinates of individual observation. Support Vector Machine is a frontier which best segregates the two classes (hyper-plane/ line)
  + An SVM generates parallel partitions by generating two parallel lines. For each category of data in a high-dimensional space and uses almost all attributes. It separates the space in a single pass to generate flat and linear partitions. Divide the 2 categories by a clear gap that should be as wide as possible. Do this partitioning by a plane called hyperplane.
  + An SVM creates hyperplanes that have the largest margin in a high-dimensional space to separate given data into classes. The margin between the 2 classes represents the longest distance between closest data points of those classes.
  + STEP1: Select the feature sets from different classes of data
  + STEP2: Calculate the intersection points of each class of feature and plot, repeat for all the features of data.
  + STEP3: Remove the features which are intersecting and data of all the classes.
  + STEP4: Plot the hyper planes for the remaining points.
  + STEP5: Calculate the distance of the hyper planes in different class of objects.
  + STEP6: Select the hyper plane which is consistent for each class of data

### Pros

* + Accuracy
  + Works well on smaller cleaner datasets
  + It can be more efficient because it uses a subset of training points

### Cons

* + Isn’t suited to larger datasets as the training time with SVMs can be high
  + Less effective on noisier datasets with overlapping classes

# Scope of the Project

* + - The main intention of our project is to solve the problems faced by physicians in diagnosis of liver diseases by using Machine Learning technique.
    - To early prediction of chance of getting Liver disease in health sector using K- Means and naive Bayesian and SVM algorithm in data mining.
    - To save human life and take proper steps to control the disease by early prediction of disease.

# Related Work

**Ashwani Kumar., [1] “Categorization of Liver Disease Using Classification Techniques”** It is one of the great techniques to dissect liver and non-liver patient in awesome exactness. They utilized the C4.5, Random Forest, CART, Random Tree and REP tree as classifier for order of liver information and accomplished 79.22% exactness in Random Forest utilizing 80-20% information segment with 6 highlights. They suggested Random woodland is better classifier for characterization of liver illness among all.

**Chandrasegar Thirumalai., [2] “Cost Optimization using Normal Linear Regression Method for Breast Cancer Type I Skin**” Pearson strategy has broken down its relationship coefficient with different characteristics, by utilizing Pearson strategy. From these coefficient esteems, the properties having values which are more prominent than 0.55, are mulled over. Further, machine learning technique is utilized for discovering the minimum cost capacity and its relating theta esteem. By utilizing straightforward direct relapse strategy on the diminished properties, we have gotten direct conditions. Their proposed display one can take astute choices with the assistance of convenient electronic gadgets like advanced cells and tabs.

**Harsha Pakhale.,[3] “Development of an Efficient Classifier for Classification of Liver Patient with Feature Selection”** has examined the individual model does not fulfil the arrangement exactness of model so they have gathering the individual models to build up the hearty classifier. They have additionally connected the data pick up include determination strategy to computationally expand the execution of model.

**Dr. S. Vijayarani., [4] “Liver Disease Prediction using SVM and Naïve Bayes Algorithms”** finished up the SVM classifier is considered as a best calculation due to its most elevated grouping exactness. Then again, while looking at the execution time, the Naïve Bayes classifier needs least execution time.

**Anju Gulia., [5] “Liver Patient Classification Using Intelligent Techniques”** presents an approach that will be utilized for half breed display development of group wellbeing administrations. These arrangement calculations can be actualized for other overwhelming sicknesses likewise like cardiovascular and diabetes forecast and grouping. Another extension is to seeing climate by applying new calculations will made any upgrades over strategies which are utilized as a part of this paper in future.

**Kalyan Nagaraj., [6] “NeuroSVM: A Graphical User Interface for Identification of Liver Patients”** say an endeavour is made for forecast of liver infection in patients utilizing information mining systems. A cross breed Neuro-SVM show was produced for characterization of liver patients in view of their organic parameters utilizing fake neural system. The cross-breed display is sent as a graphical UI (GUI) in R. The GUI can be utilized as a screening instrument by specialists for expectation of liver illness in patients in future.

**Jankisharan Pahareeya., [7]** , **“Liver Patient Classification using Intelligence Techniques”** closes Throughout the examination ten times cross approval is performed. The proposed Random Forest over testing (200%) show beat every single other method. The Random Forest over examining (200%) display is the generally best indicator among every single other strategy.

**Bendi Venkata Ramana.,[8] “A Critical Study of Selected Classification Algorithms for Liver Disease Diagnosis”** has broken down Modified pivot timberland calculation was proposed with multi-layer observation grouping calculation and

irregular subset include determination technique for UCI liver informational collection. Changed pivot woodland calculation was proposed with closest neighbour arrangement calculation and connection-based element determination strategy for INDIA liver dataset.

**Kotsiantis. S.B., [9] “A Critical Study of Selected Classification Algorithms for Liver Disease Diagnosis**” advises to build the forecast exactness of the straightforward Bayes demonstrate. The idea of joining classifiers is proposed as another bearing for the change of the execution of individual classifiers. He utilized a discretization strategy and expelled excess highlights utilizing a channel include determination technique. At long last, he played out a substantial scale correlation with different endeavours that have attempted to enhance the precision of the basic Bayes calculation and additionally other best in class calculations and he took better exactness as a rule utilizing less time for preparing, as well.

# Summary

This chapter describes the machine learning techniques used in the project and also explaines the scope of project.

# Chapter 3

**LIVER DISEASE PREDICTION SYSTEM DESIGN AND ARCHITECTURE**

Design is a meaningful engineering representation of something that is to be built. It is the most crucial phase in the developments of a system. Software design is a process through which the requirements are translated into a representation of software. Design is a place where design is fostered in software Engineering. Based on the user requirements and the detailed analysis of the existing system, the new system must be designed. This is the phase of system designing.

# Introduction to System Architecture

System Architecture design-identifies the overall hypermedia structure for the Application. Architecture design is tied to the goals establish for a Application, the content to be presented, the users who will visit, and the navigation philosophy that has been established. Content architecture, focuses on the manner in which content objects and structured for presentation and navigation. Application architecture, addresses the manner in which the application is structure to manage user interaction, handle internal processing tasks, effect navigation, and present content. Application architecture is defined within the context of the development environment in which the application is to be implemented.

# System Requirements

Software Requirement Specification (SRS) is a fundamental document, which forms the foundation of the software development process. SRS not only lists the requirements of a system but also has a description of its major features. These recommendations extend the IEEE standards. The recommendations would form the basis for providing clear visibility of the product to be developed serving as baseline for execution of a contract between client and the developer. SRS constitutes the agreement between clients and developers regarding the contents of the software product that is going to be developed. SRS should accurately and completely represent the system requirements as it makes a huge contribution to the overall project plan. The software being developed may be a part of the overall larger system or may be a complete standalone system in its own right.

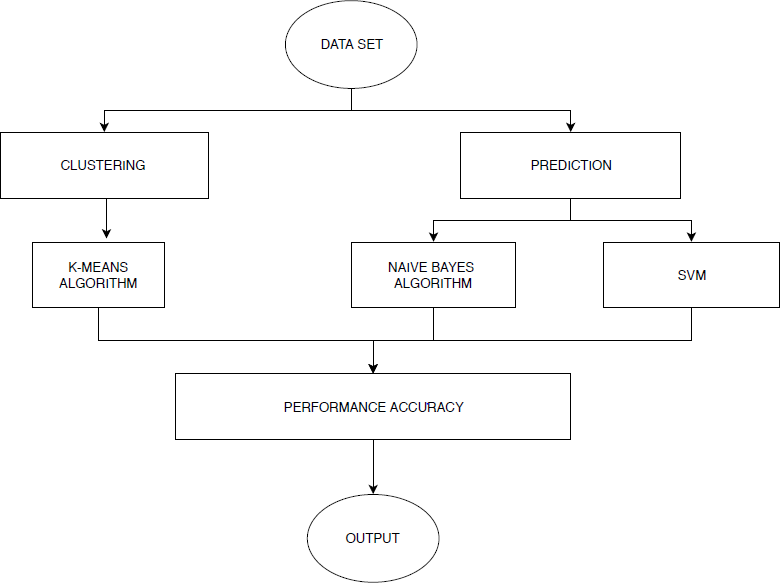
## Hardware Requirements

* + - * Processor : Intel Core 2 Duo
      * RAM : 2 GB
      * Hard Disk : 80 GB

## Software Requirements

* + - * Operating System : Windows (Any Version)
      * Programming Language : JAVA, AWT, SWING
      * IDE : Eclipse

# System Architecture

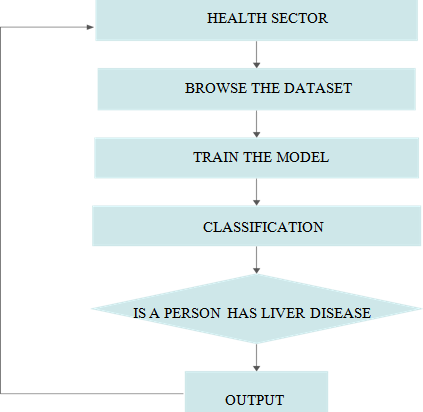
System Architecture design-identifies the overall structure for the Liver disease as shown in below diagram.

### Figure 3.1: System Architecture

Figure 3.1 System Architecture for Liver disease ,Compares Naïve Bayes and SVM algorithms are done and it is based on the performance factors classification accuracy and execution time.

# Data Flow Diagram

Data flow diagram is a graphical portrayal of the flow of information through a data framework, displaying its procedure angles. A DFD is regularly utilized as a preparatory advance to make an outline of the framework without broadly expounding, which can later be explained. A DFD demonstrates what sort of data will be contribution to and yield from the framework, how the information will progress through the framework, and where the information will be put away. It doesn't indicate data about process timing or whether procedures will work in succession or in parallel, dissimilar to a customary organized flowchart which centers around control stream, or an UML movement work process outline, which presents both control and information streams as a brought together model.



### Figure 3.2: Data flow diagram

Figure 3.2 is the data flow diagram the output of the results is feedback to the dataset the flow of data presented above.A DFD demonstrates what sort of data will be contribution to and yield from the framework

# Use Case Diagram

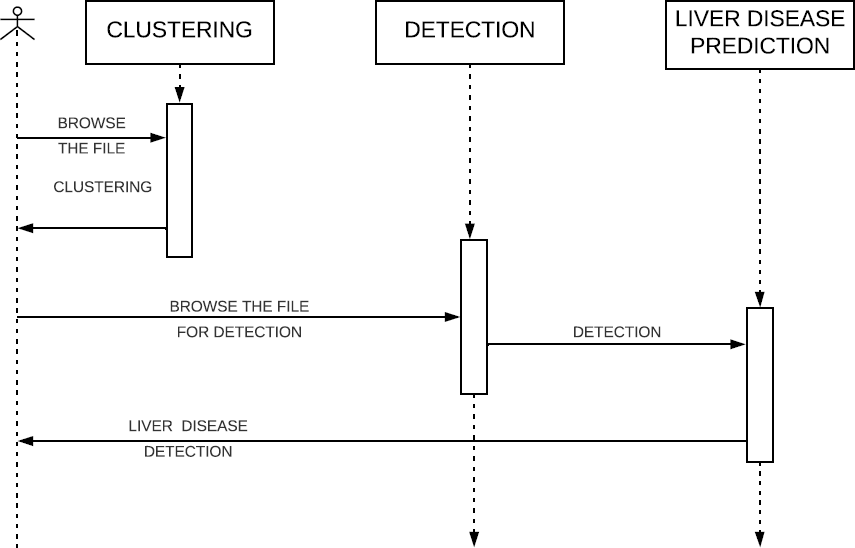
Use case diagrams are utilized to accumulate the necessities of a framework including inside and outside impacts. These prerequisites are generally plan necessities. Thus, when a framework is investigated to assemble its functionalities, utilize cases are readied and on-screen characters are recognized. At the point when the underlying assignment is finished, utilize case outlines are demonstrated to exhibit the outside view. To somethings up, the motivations behind utilize case graphs can be said to be as per the given

* Used to assemble the prerequisites of a system.
* Used to get an outside perspective of a system.
* Distinguish the outer and interior components affecting the system.
* Demonstrate the collaboration among the necessities are on-screen characters.



### Figure 3.3: Use Case diagram

* 1. **Sequence diagram**

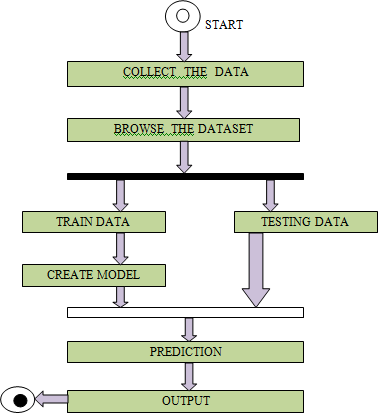


**Figure 3.4: Sequence diagram**

In Figure 3.4 a sequence diagram demonstrates protest collaborations organized in time arrangement.

It delineates the articles and classes engaged with the situation and the succession of messages traded between the items expected to complete the usefulness of the situation. Arrangement charts are regularly connected with utilize case acknowledge in the Logical View of the framework a work in progress.

### Activity diagram



**Figure 3.5: Activity diagram**

In Figure 3.5 Activity diagram is essentially a flowchart to speak to the spill out of one action to another movement.

The action can be portrayed as an activity of the framework. The control stream is attracted starting with one task then onto the next. This stream can be consecutive, fanned, or simultaneous. Action graphs manage all sort of stream control by utilizing diverse components, for example, fork, join, and so on .The fundamental reasons for action charts is like other four graphs. It catches the dynamic conduct of the framework. Other four charts are utilized to demonstrate the message spill out of one protest another however action graph is utilized to indicate message spill out of one movement to another. Movement is a specific task of the framework. Movement graphs are not just utilized for envisioning the dynamic idea of a framework, yet they are additionally used to build the executable framework by utilizing forward and figuring out procedures. The main missing thing in the movement chart is the message part. It doesn't demonstrate any message spill out of one movement to another. Movement graph is some of the time considered as the flowchart. In spite of the fact that the graphs

resemble a flowchart, they are definitely not. It indicates diverse streams, for example, parallel, fanned, simultaneous, and single. The reason for a activity diagram can be depicted as ,

* Draw the movement stream of a framework.
* Depict the succession starting with one action then onto the next.
* Depict the parallel, extended and simultaneous stream of the framework.

# Summary

This chapter gives the description about the system architecture and system design. It gives description about the flow diagrams and use case diagrams,sequence diagrams and activity diagram.

# Chapter4

**LIVER DISEASE PREDICTION SYSTEM IMPLEMENTATION AND TESTING**

The implementation phase involves more than just writing code. Code also needs to be tested and debugged as well as compiled and built into a complete executable product. We usually need to utilize configuration management in order to keep track of different version of code. This is the stage of the project where the theoretical design is turned into a working system. If the implementation is not carefully planned and controlled, it can cause chaos and confusions. It is always a good idea to keep in mind that some characteristics that should be found in a good implementation like Readability- our code is written in MVC Architecture ,JAVA to achieve the objective of the project that is to introduce a novel scheme of mechanism design for balancing the resource consumptions .

Our implementation stage requires the following tasks:

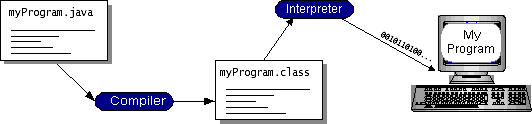
* Careful planning
* Investigation of system and constraints
* Design of methods to achieve the changeover
* Evaluation of the changeover method
* Correct decisions regarding selection of the platform
* Appropriate selection of the language for application development
* Java technology is both a programming language and a platform.

# The Java Programming Language

The Java programming language is a high-level language that can be characterized by all of the following buzzwords: Simple, Architecture neutral, Object oriented, Portable, Distributed, High performance, Interpreted, Multithreaded, Robust, Dynamic and Secure.

With most programming languages, you either compile or interpret a program so that you can run it on your computer. The Java programming language is unusual in that a program is both compiled and interpreted. With the compiler, first you translate a program into an intermediate language called Java byte codes —the platform- independent codes interpreted by the interpreter on the Java platform. The interpreter parses and runs each Java byte code instruction on the computer. Compilation happens

just once; interpretation occurs each time the program is executed. The following figure illustrates how this works.



### Figure 4.1: Java Virtual Machine

You can think of Java byte codes as the machine code instructions for the Java Virtual Machine (Java VM). Every Java interpreter, whether it’s a development tool or a Web browser that can run applets, is an implementation of the Java VM. Java byte codes help make “write once, run anywhere” possible. You can compile your program into byte codes on any platform that has a Java compiler. The byte codes can then be run on any implementation of the Java VM. That means that as long as a computer has a Java VM, the same program written in the Java programming language can run on Windows 2000, a Solaris workstation, or on an iMac.

# Java Platform

A platform is the hardware or software environment in which a program runs. We’ve already mentioned some of the most popular platforms like Windows 2000, Linux, Solaris, and MacOS. Most platforms can be described as a combination of the operating system and hardware. The Java platform differs from most other platforms in that it’s a software-only platform that runs on top of other hardware-based platforms.

The Java platform has two components:

* + - The Java Virtual Machine (Java VM)
    - The Java Application Programming Interface (Java API)

You’ve already been introduced to the Java VM. It’s the base for the Java platform and is ported onto various hardware-based platforms.

The Java API is a large collection of ready-made software components that provide many useful capabilities, such as graphical user interface (GUI) widgets. The Java API

is grouped into libraries of related classes and interfaces; these libraries are known as packages. The next section, What Can Java Technology Do? Highlights what functionality some of the packages in the Java API provide.

The following figure depicts a program that’s running on the Java platform. As the figure shows, the Java API and the virtual machine insulate the program from the hardware.

Native code is code that after you compile it, the compiled code runs on a specific hardware platform. As a platform-independent environment, the Java platform can be a bit slower than native code. However, smart compilers, well-tuned interpreters, and just-in-time byte code compilers can bring performance close to that of native code without threatening portability.

## The Importance of Java Technology

The most common types of programs written in the Java programming language are appletsand applications. If you’ve surfed the Web, you’re probably already familiar with applets. An applet is a program that adheres to certain conventions that allow it to run with in a Java-enabled browser. However, the Java programming language is not just for writing cute, entertaining applets for the Web. The general-purpose, high-level Java programming language is also a powerful software platform. Using the generous API, you can write many types of programs.

An application is a standalone program that runs directly on the Java platform. A special kind of application known as aserver serves and supports clients on a network. Examples of servers are Web servers, proxy servers, mail servers, and print servers. Another specialized program is a Servlets. A Servlets can almost be thought of as an applet that runs on the server side. Java Servlets are a popular choice for building interactive web applications, replacing the use of CGI scripts. Servlets are similar to applets in that they are runtime extensions of applications. Instead of working in browsers, though, Servlets run within Java Web servers, configuring or tailoring the server.

**AWT** (Abstract Window Toolkit) is an application programming interface to develop graphical user interface such as objects, such as buttons, scroll bars, and windows. Java AWT components are platform-dependent or components are displayed according to the view of operating system.

**SWING** was developed to provide a more sophisticated set of GUI [components](https://en.wikipedia.org/wiki/Software_component) than the earlier [Abstract Window Toolkit (AWT)](https://en.wikipedia.org/wiki/Abstract_Window_Toolkit). Swing provides a [look and feel](https://en.wikipedia.org/wiki/Look_and_feel) that emulates the look and feel of several platforms, and also supports a [pluggable look and](https://en.wikipedia.org/wiki/Pluggable_look_and_feel) [feel](https://en.wikipedia.org/wiki/Pluggable_look_and_feel) that allows applications to have a look and feel unrelated to the underlying platform. It has more powerful and flexible components than AWT. In addition to familiar components such as buttons, check boxes and labels, Swing provides several advanced components such as tabbed panel, scroll panes, trees, tables, and lists

**SRS (**Software requirement specification) might be a entire description of the meant reason and surroundings for software program program this is below development. The SRS describes without a doubt what the software program application machine can do and the way it's miles going to be anticipated to perform. An SRS permits minimizing the time and energy wished via builders to attain their preferred desires and moreover minimizes the rate of development. A brilliant SRS defines how an software program ought to interact with the device hardware components, special programs and human customers in a really wide selection of actual-international subjects. SRS is a report created by means of tool analyst after the necessities are accrued from numerous stakeholders. The parameters like operation velocity, c language, availability, portability, maintainability, foot print, protection and velocity of restoration from negative sports are evaluated. The necessities received from customer are written in natural language. It is the responsibility of machine analyst to record the necessities in technical language in order that they will be comprehended and useful with the aid of the software development group. The requirement machine is achieved for enhancement of the present gadget, in addition the requirement evaluation which is likewise deals with shooting the useful and non-practical necessities from various users.

Functional necessities may be calculations, technical details, statistics manipulation and processing and different particular capability that define what a device is supposed to accomplish. Behavioural necessities describing all the instances where the machine uses the purposeful requirements are captured in use instances. Functional necessities are supported via non-purposeful necessities additionally known as best necessities, which impose constraints on the layout or implementation along with overall performance necessities, protection, or reliability. Generally, functional requirements are expressed in the for"machine have to do requirement", at the same time as non-functional necessities are "device shall be requirement". The plan for

imposing functional requirements is detailed inside the machine design. The plan for enforcing non-purposeful necessities is special inside the device structure.

As defined in necessities engineering, practical requirements specify unique outcomes of a device. This should be contrasted with non-useful requirements which specify ordinary traits inclusive of fee and reliability. Functional requirements force the utility structure of a device, even as non-practical necessities drive the technical architecture of a device.

* **Get started quickly**: Although the Java programming language is a powerful object-oriented language, it’s easy to learn, especially for programmers already familiar with C or C++.
* **Write less code**: Comparisons of program metrics (class counts, method counts, and so on) suggest that a program written in the Java programming language can be four times smaller than the same program in C++.
* **Write better code**: The Java programming language encourages good coding practices, and its garbage collection helps you avoid memory leaks. Its object orientation, its JavaBeans component architecture, and its wide-ranging, easily extendible API let you reuse other people’s tested code and introduce fewer bugs.
* **Develop programs more quickly**: Your development time may be as much as twice as fast versus writing the same program in C++. Why? You write fewer lines of code and it is a simpler programming language than C++.
* **Avoid platform dependencies with 100% Pure Java**: You can keep your program portable by avoiding the use of libraries written in other languages. The100% Pure JavaProduct Certification Program has a repository of historical process manuals, white papers, brochures, and similar materials online.
* **Write once, run anywhere**: Because 100% Pure Java programs are compiled into machine-independent byte codes, they run consistently on any Java platform. **Distribute software more easily**: You can upgrade applets easily from a central server. Applets take advantage of the feature of allowing new classes to be loaded “on the fly,” without recompiling the entire program.

# Implementation of K-Means Algorithm

K-means is a method of clustering observations into a specific number of disjoint clusters. The K” refers to the number of clusters specified. Various distance measures exist to determine which observation is to be appended to which cluster. The algorithm

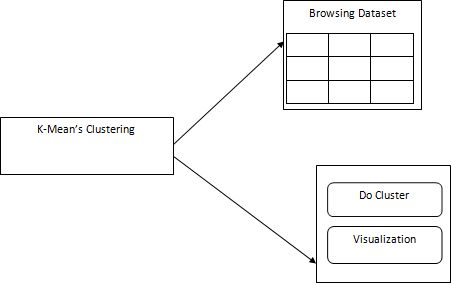
aims at minimizing the measure between the centroide of the cluster and the given observation by iteratively appending an observation to any cluster and terminate when the lowest distance measure is achieved.

* + - The sample Dataset is initially partitioned into K clusters and the observations are randomly assigned to the clusters.

For each sample:

* + - Calculate the personnel details of customer to the centroide of the cluster.
    - IF the sample data is closer to its own cluster THEN Accept ELSE or reject
    - Repeat steps 1 and 2 until no observations are moved from one cluster to another
    - When step 3 terminates the clusters are stable and each sample is assigned a cluster which results in the lowest possible distance to the centriode of the cluster.

A Person or Medical expert wants to analyze the data in order to know which personis prone to get liver disease or not. There is an enormous amount of data which one can actually retrieved from the healthsector. The data can actually be used to detect the chance of liver disease. The proposed system categorizes chance of getting liver diseae based on some personal information. Detection is then carried out to know the liver disease. The system takes all these information first as the input and points out person who are prone to get liver disease. A training dataset is being prepared beforehand based upon the person previous history. The system uses the concept of k-means clustering to cluster the person based upon the similarities or the patterns they share among each other naïve bayes and SVM used to classify and detect the chance of getting liver disease with the training dataset and testing dataset and after classifying it will detect the liver disease.



### Figure 4.2: Clustering

Figure 4.2 illustrates,step for clustering using K-Means are first browse the required training dataset and then do cluster and visualization of cluster.

K-means is a method of clustering analysis into a precise number of disjoint clusters. The K-refers to the total of clusters specified. Different distance events survive to resolve which observation is to be appended to which cluster.

* A cluster of information items can be deal with as one gathering.
* While doing bunch examination, we first parcel the arrangement of information into bunches in view of information comparability and afterward dole out the names to the gatherings.
* The principle preferred standpoint of bunching over arrangement is that, it is versatile to changes and helps single out valuable highlights that recognize distinctive gatherings.

### Distance measures

Common distance measures include the Euclidean distance, the Euclidean squared distance

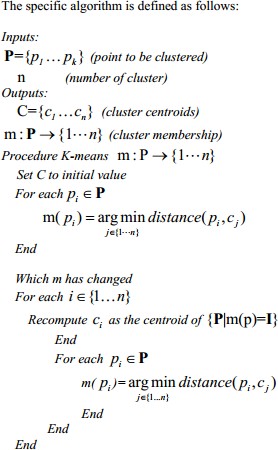
and the Manhattan or City distance.

The Euclidean measure corresponds to the shortest geometric distance between to points.

A faster way of determining the distance is by use of the squared Euclidean distance which

calculates the above distance squared, i.e.





# Implemention of Naïve Bayes

Classification techniques are maximum ideal for predicting or describing facts sets with binary or nominal classes. They are less effective for ordinal classes because they do now not consider the implicit order among the classes. Other kinds of relationships,

which includes the subclass–super class relationships among classes also are unnoticed. The remainder of this bankruptcy focuses best on binary or nominal magnificence labels.

A class technique is a scientific method to constructing type fashions from an enter information set. Examples encompass choice tree classifiers, rule-based totally classifiers, neural networks, aid vector machines, and Naive Bayes classifiers. Each approach employs a mastering set of rules to pick out a model that quality fits the connection among the attribute set and sophistication label of the enter records. The version generated by using a studying algorithm ought to each suit the input records properly and efficiently are expecting the magnificence labels of facts it has never visible earlier than. Therefore, a key goal of the gaining knowledge of set of rules is to construct fashions with correct generalization capability, models that accurately expect the elegance labels of previously unknown facts. The tree has 3 styles of nodes:

* A root node that has no approaching edges and at least zero active edges
* Internal node, each of which has precisely one approaching edge and at least two active edges.
* Leaf or terminal node, each of which has precisely one approaching edge and no friendly edges.

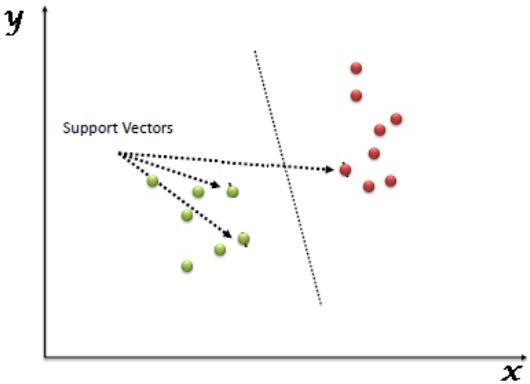
Naive Bayes is a simple technique for constructing classifiers: models that assign class labels to problem instances, represented as vectors of [feature](https://en.wikipedia.org/wiki/Feature_vector) values, where the class labels are drawn from some finite set. It is not a single [algorithm](https://en.wikipedia.org/wiki/Algorithm) for training such classifiers, but a family of algorithms based on a common principle: all Naive Bayes classifiers assume that the value of a particular feature is [independent](https://en.wikipedia.org/wiki/Independence_(probability_theory)) of the value of any other feature, given the class variable.

* + In this progression the grouping calculations assemble the classifier.
  + The classifier is worked from the preparation set made up of database tuples and their related class marks.
  + Each tuple that constitutes the preparation set is alluded to as a classification or class. These tuples can likewise be alluded to as test, question or information focuses.
  + The above diagram shows that the liver disease is detected based on person personal details.

# Implemention of Support Vector Machine

Support Vector Machine was first found by Vapnik in 1979. It was again recommended by Vapnik in 1995 for regression and classification. Support vector can be used for pattern classification which has multilayer perceptron and radial-basis function networks. The SVM is the advanced technology with maximum classification algorithms embedded in statistical learning theory. SVM methods are used in classification of linear and non-linear data. It transforms the original training data into higher dimension using non-linear mapping. Within this new dimension it searches for linear optimal separating hyperplane. Data from two classes can be separated by hyperplane with an appropriate nonlinear mapping to a sufficiently high dimension. Using support vectors and margins the SVM finds these hyperplane [6]. SVM implements the classification task by maximizing the margin classifies both class while minimizing the classification errors. Although the SVM can be applied to various optimization problems such as regression, the classic problem is that of data classification.The data points are identified as being positive or negative, and the problem is to find a hyperplane that separates the data points by a maximal margin.

Support Vector Machine (SVM) is a supervised [machine learning](https://courses.analyticsvidhya.com/courses/introduction-to-data-science-2?utm_source=blog&utm_medium=understandingsupportvectormachinearticle) [algorithm](https://courses.analyticsvidhya.com/courses/introduction-to-data-science-2?utm_source=blog&utm_medium=understandingsupportvectormachinearticle) which can be used for both classification or regression challenges. However, it is mostly used in classification problems. In this algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiate the two classes very well (look at the below snapshot).



### Figure 4.3: SVM Model

In figure 4.3 Support Vectors are simply the co-ordinates of individual observation. Support Vector Machine is a frontier which best segregates the two classes (hyper- plane/ line)

An SVM generates parallel partitions by generating two parallel lines. For each category of data in a high-dimensional space and uses almost all attributes. It separates the space in a single pass to generate flat and linear partitions. Divide the 2 categories by a clear gap that should be as wide as possible. Do this partitioning by a plane called hyperplane.

An SVM creates hyperplanes that have the largest margin in a high-dimensional space to separate given data into classes. The margin between the 2 classes represents the longest distance between closest data points of those classes.

STEP1: Select the feature sets from different classes of data

STEP2: Calculate the intersection points of each class of feature and plot, repeat for all the features of data.

STEP3: Remove the features which are intersecting and data of all the classes. STEP4: Plot the hyper planes for the remaining points.

STEP5: Calculate the distance of the hyper planes in different class of objects. STEP6: Select the hyper plane which is consistent for each class of data

### Pros

* + - Accuracy
    - Works well on smaller cleaner datasets
    - It can be more efficient because it uses a subset of training points

### Cons

* + - Isn’t suited to larger datasets as the training time with SVMs can be high
    - Less effective on noisier datasets with overlapping classes

# Software Testing Introduction

Software testing is defined as the process of investigation that is done on the software or software products to express or provide the customer with the details of the software product. Testing is defined as the process of development where developer or programmer create different test cases as they develop or build the software as per defined user requirement. These cases are small and simple which checks the functions and workability of particular part of system of their code, such as class of functions. The developer uses different type of open source libraries function which runs and tests automatically and helps in checking problems quickly. With the test case been developed parallel with different source unit test express its correctness. An implementation which has been evaluated using corporate testing samples gives us the confident that can detect chance of getting liver disease better than manual detection. The same applies with personal testing samples and personal liver disease detection.

* **Unit Testing:** Unit testing is defined as the first basic phase or level of the software testing process where single components of the software in elaborated. It is done to check whether all the components are working as per the description done during design. It increases the confidence for in elaborating code. It is basically followed with the help of white box testing method.
* **Integration Testing**: Integration testing is a phase that comes after unit testing where the combination of units are done in a group and tested to check the compatibility and adoptability to different units of the whole system. It is done to correct or the check the faults that are appeared due to the combination or integration of different testing units. Any method such as black box, grey box or white box testing method is helpful for integration testing.
* **System testing**: It is the third phase of testing level where it is done on the complete integrated software system. It is basically done to check the software obedience as per described requirements. This process is done with the help of black box testing method and it is carried out by the separate software testers.
* **Acceptance Testing**: This is the final phase of testing system which is done on the complete system to check whether it is complete with business requirement and is in a phase of delivery to the customers.

### Table 4.1: Test Case For The Software

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Test**  **Case ID** | **Test**  **Objective** | **Precondition** | **Steps** | **Expected**  **result** | **Post**  **Condition** |
| TC\_001 | Clustering | Browse the file | Add the dataset which include personnel  details | User details are clustered | Success |
| TC\_002 | Detection | Browse Train & Testing  Dataset | Files are browsed  successfully | Data should be tested | Success |
| TC\_003 | Liver diseaase status | Based on sample data set | Dataset should be checked | Chance of getting liver disease detected on basis of  health data | Success |

* 1. **Explanation for SDLC**

**SDLC**: The software development life cycle (SDLC) is a conceptual model used in project management that describes the stages involved in an information system development project, from an initial feasibility study through maintenance of the completed application.

### Phases Of Software Development

* + - Requirement Analysis
    - Software design
    - Development or Coding
    - Testing
    - Maintenance

### Requirement Analysis

The requirements of a desired software product are extracted. Based the business scenario the SRS (Software Requirement Specification) document is prepared in this phase.

### Software Design

Plans are laid out concerning the physical construction, hardware, operating systems, programming, communications, and security issues for the software. Design phase is concerned with making sure the software system will meet the requirements of the product.

### Testing

Testing is evaluating the software to check for the user requirements. Here the software is evaluated with intent of finding defects.

### Maintenance

Once the new system is up and running for a while, it should be exhaustively evaluated. Maintenance must be kept up rigorously at all times. Users of the system should be kept up-to-date concerning the latest modifications and procedure

# SDLC Models

* **Water fall model:** It will be executing one by one of the SDLC process. The design Starts after completing the requirements analysis coding begins after design. It is a traditional model It is a sequential design process, often used in SDLC, in which the progress is seen as flowing steadily downwards ( like a waterfall ), through the different phases.
* **Prototype model:** Developed from the sample after getting good feedback from the customer. This is the Valuable mechanism for gaining better understanding of the customer needs
* **Rapid application development model(RAD):**This mechanism will develop from already existing one. If The New requirement is matching in already existing requirement, will develop from that**.**
* **Spiral model:** This mechanism is update the application version by version. All the SDLC process will update version by version**.**

# Types of Testing

1. White Box Testing
2. Black Box Testing
3. Grey box testing

### White Box Testing

[White box testing](http://www.buzzle.com/editorials/4-10-2005-68350.asp) as the name suggests gives the internal view of the software. This type of testing is also known as structural testing or glass box testing as well, as the interest lies in what lies inside the box.

### Black Box Testing

Itsalso called as behavioral testing. It focuses on the functional requirements of the software. Testing either functional or non functional without reference to the internal structure of the component or system is called black box testing.

### Grey Box Testing

Grey box testing is the combination n of black box and white box testing. Intention of this testing is to find out defects related to bad design or bad implementation of the system.

# Summary

This chapter describes about implementation of java platform, Implementation of k- means,Naïve Bayes,SVM algorithms and Types of testing.

# Chapter 5

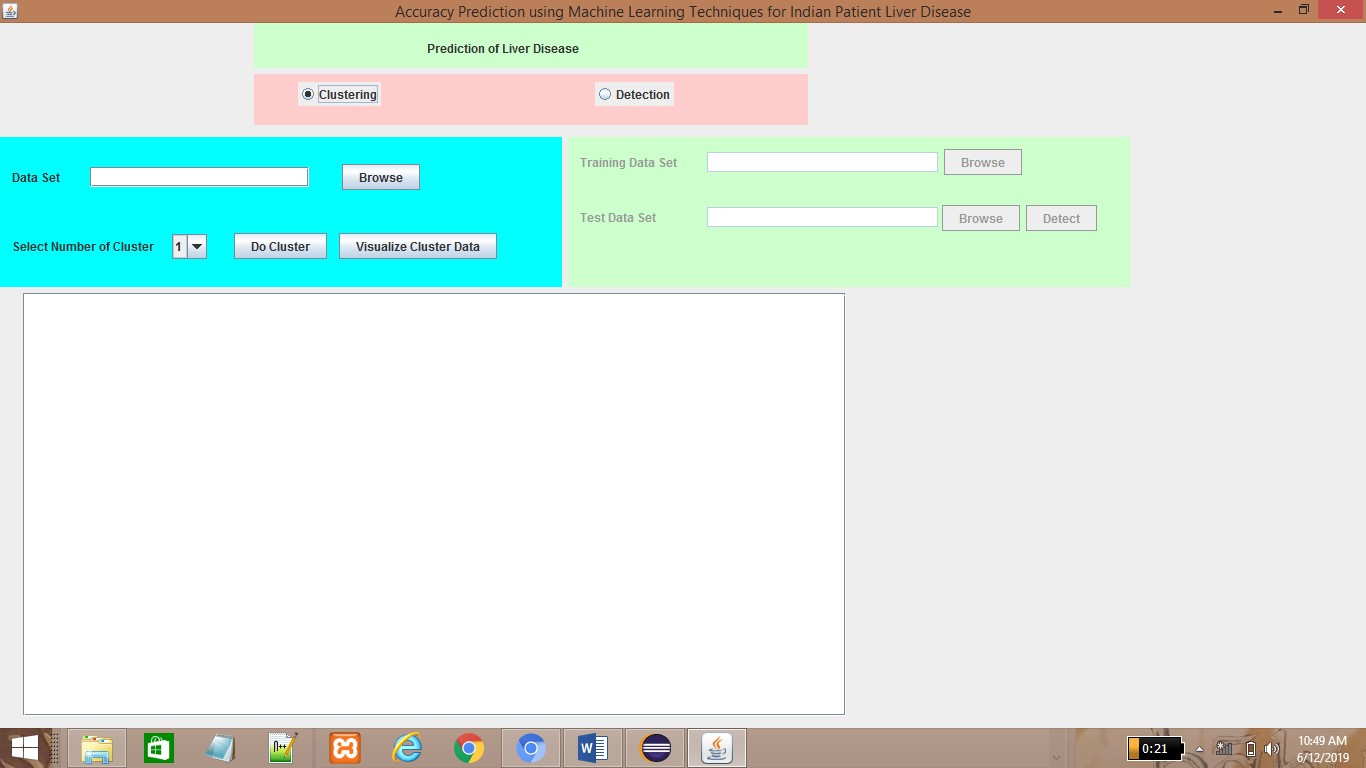
**CONCLUSION AND FUTURE ENHANCEMENT**

Classification is the major data mining technique which is primarily used in healthcare sectors for medical diagnosis and predicting diseases. This research work used classification algorithms namely Naïve Bayes and Support Vector Machine (SVM) for liver disease prediction. Comparisons of these algorithms are done and it is based on the performance factors classification accuracy and execution time. From the experimental results, this work concludes, the SVM classifier is considered as a best algorithm because of its highest classification accuracy. On the other hand, while comparing the execution time, the Naïve Bayes classifier needs minimum execution time

It was shown via experiments that the proposed method outperforms state-of- the-art methods for liver disease prediction of classification accuracy. However, machine learning is not only important for increasing the classification accuracy, but it also reduces training time and leads to more robust systems because it exploits the high variability of large-scale datasets.The sytem should handle the big data with enhanced user interface.

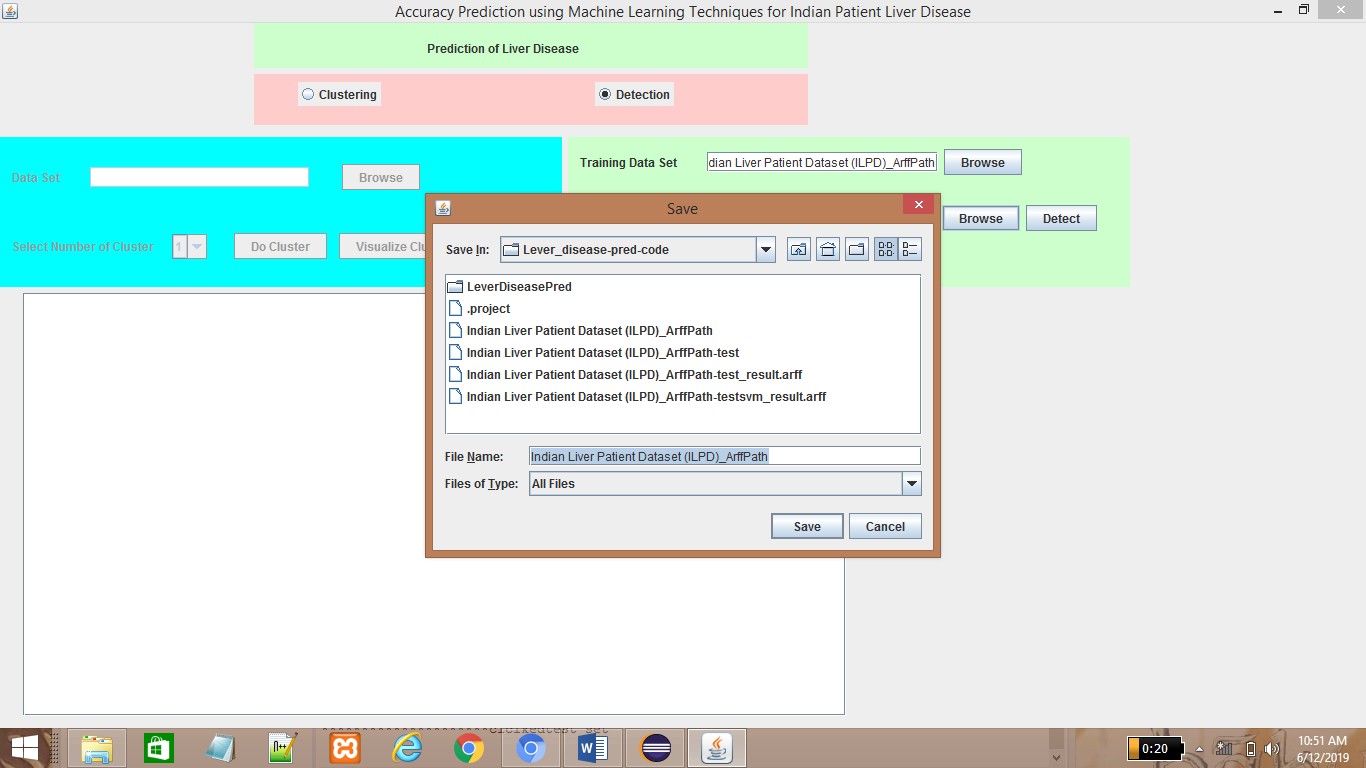
# APPENDIX - A

A result is the final consequence of actions or events expressed qualitatively or quantitatively. Performance analysis is an operational analysis, is a set of basic quantitative relationship between the performance quantities.



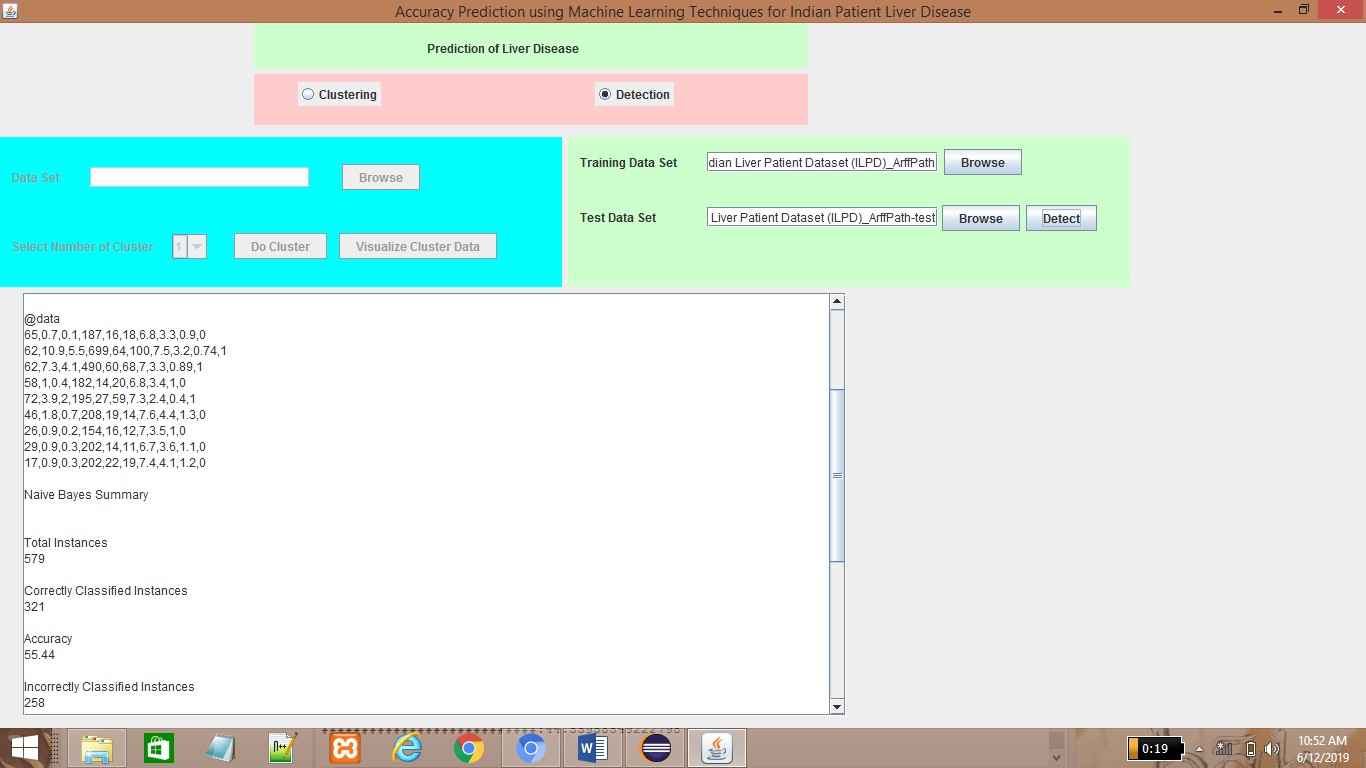
### Figure A1: Front Page

Figure A1: Front page containes two options that is Clustering and Detection. User can select the option based on their requirements either Clustering or Detection.



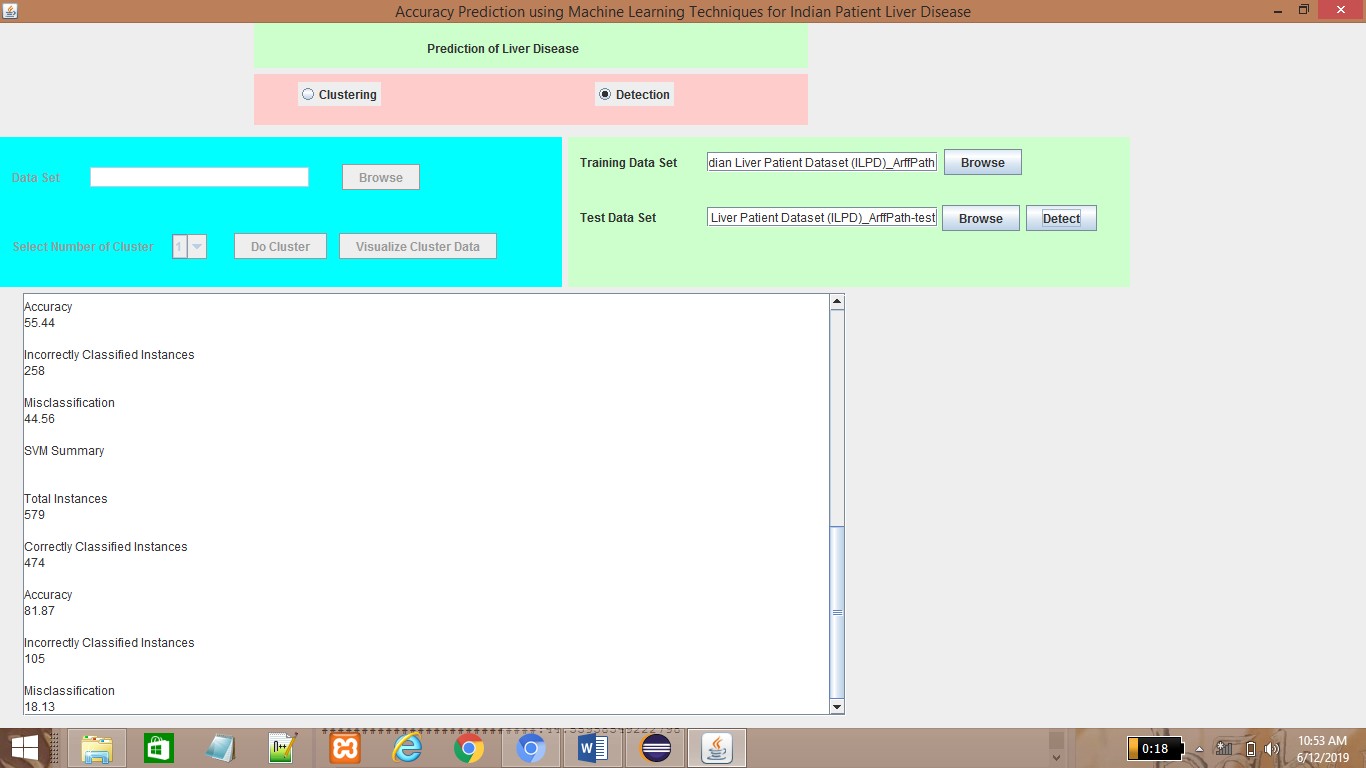
### Figure A2: Data Set Selection

Figure A2: User can browse both training and testing data set. The training data and testing data are stored in .arff extension file. In Detection user has to browse both training and testing data set. In Clustering user has to browse only training data.



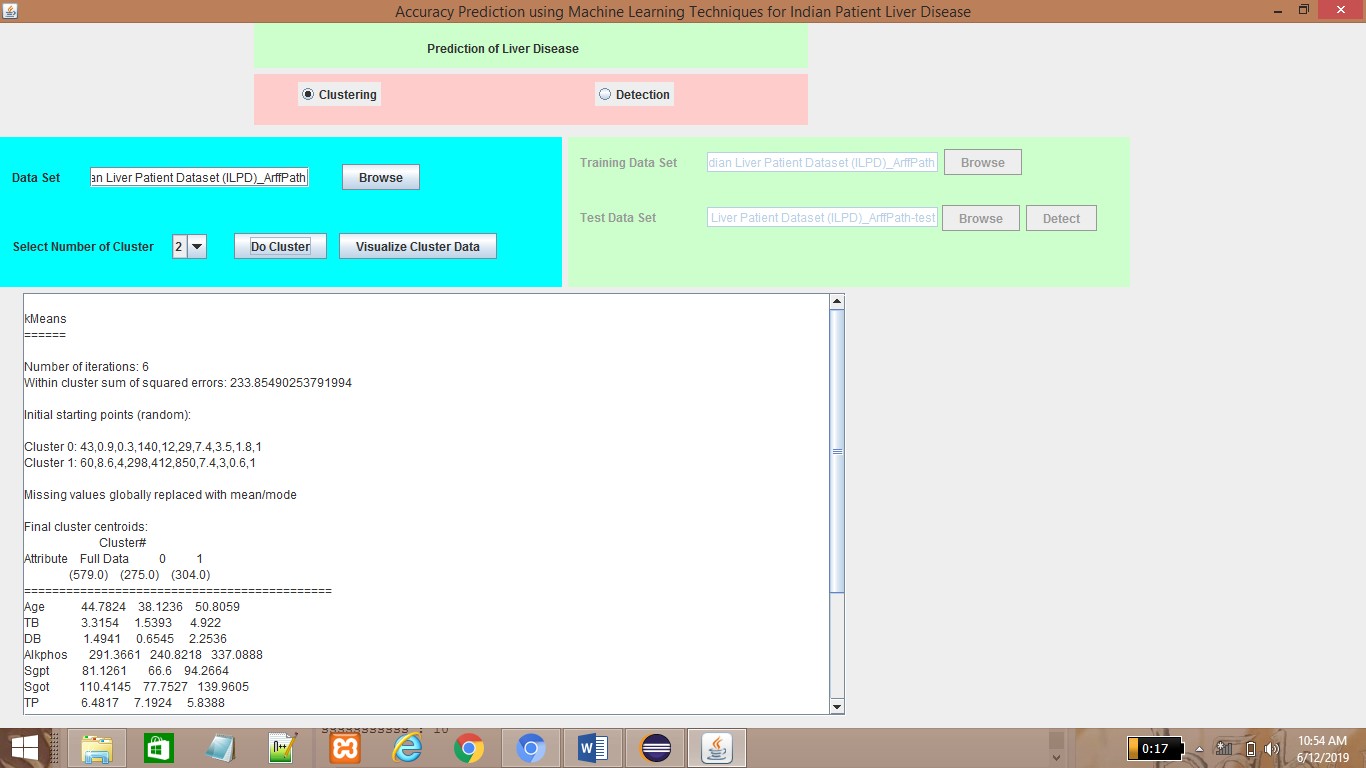
### Figure A3: Summary of Naïve Bayes

Figure A3: User can view the summary of Naïve Bayes. It contains Total instances, Correctly classified instances, Accuracy, Incorrect classified instances, Miss classifications.



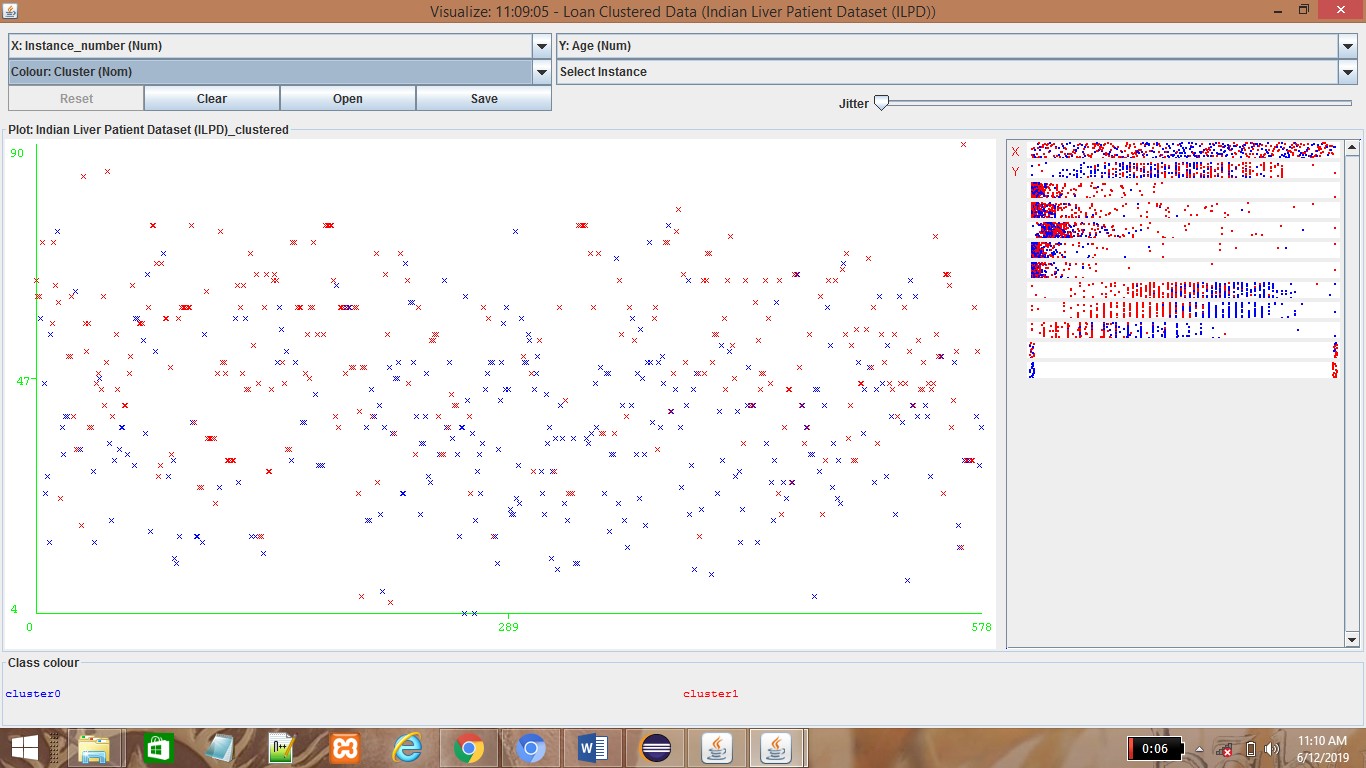
### Figure A4: Summary of SVM

Figure A4: User can view the summary of SVM. It contains Total instances, Correctly classified instances, Accuracy, Incorrect classified instances, Miss classifications.



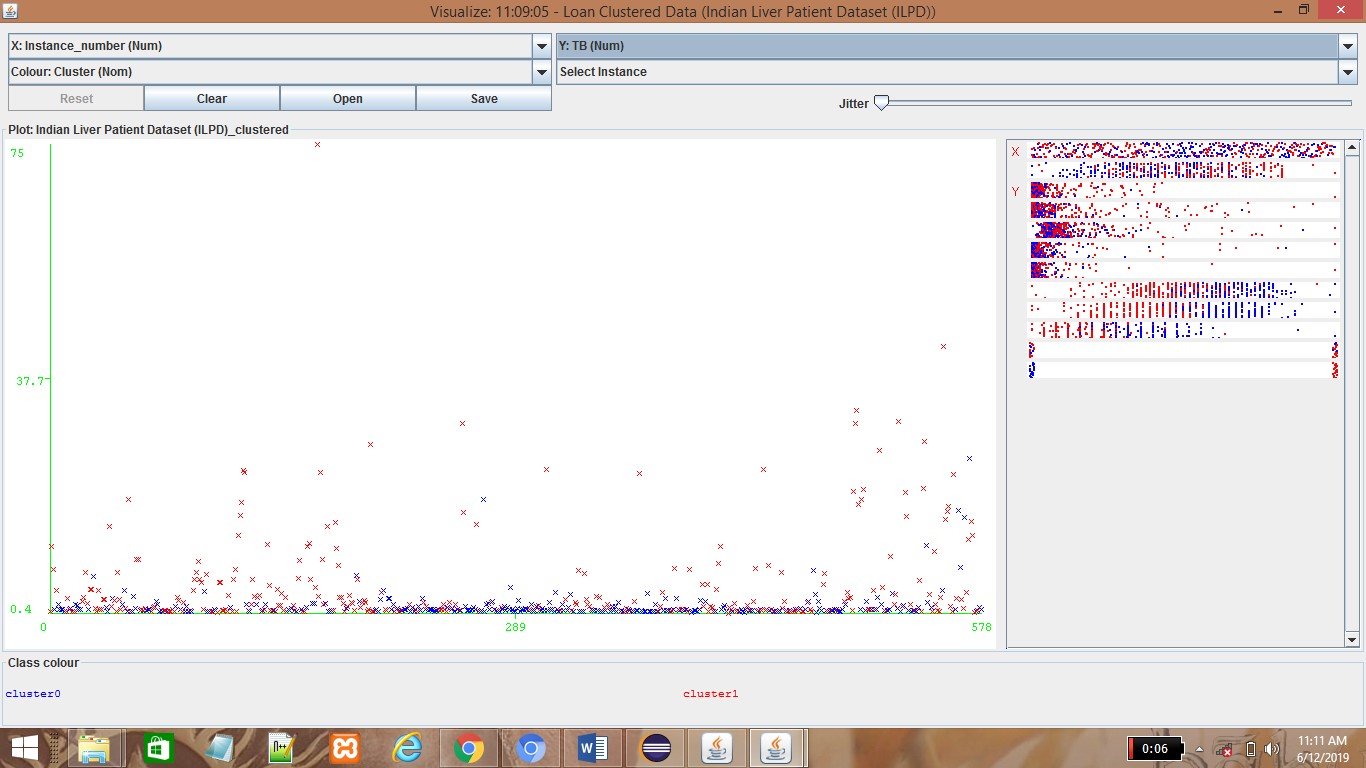
### Figure A5: Summary of K-Means

Figure A5: In summary of K-means algorithm it will contain number of iterations, within cluster sum of squared errors, initial starting point, final cluster centroid. Finally it produces the percentage of clustered instances.



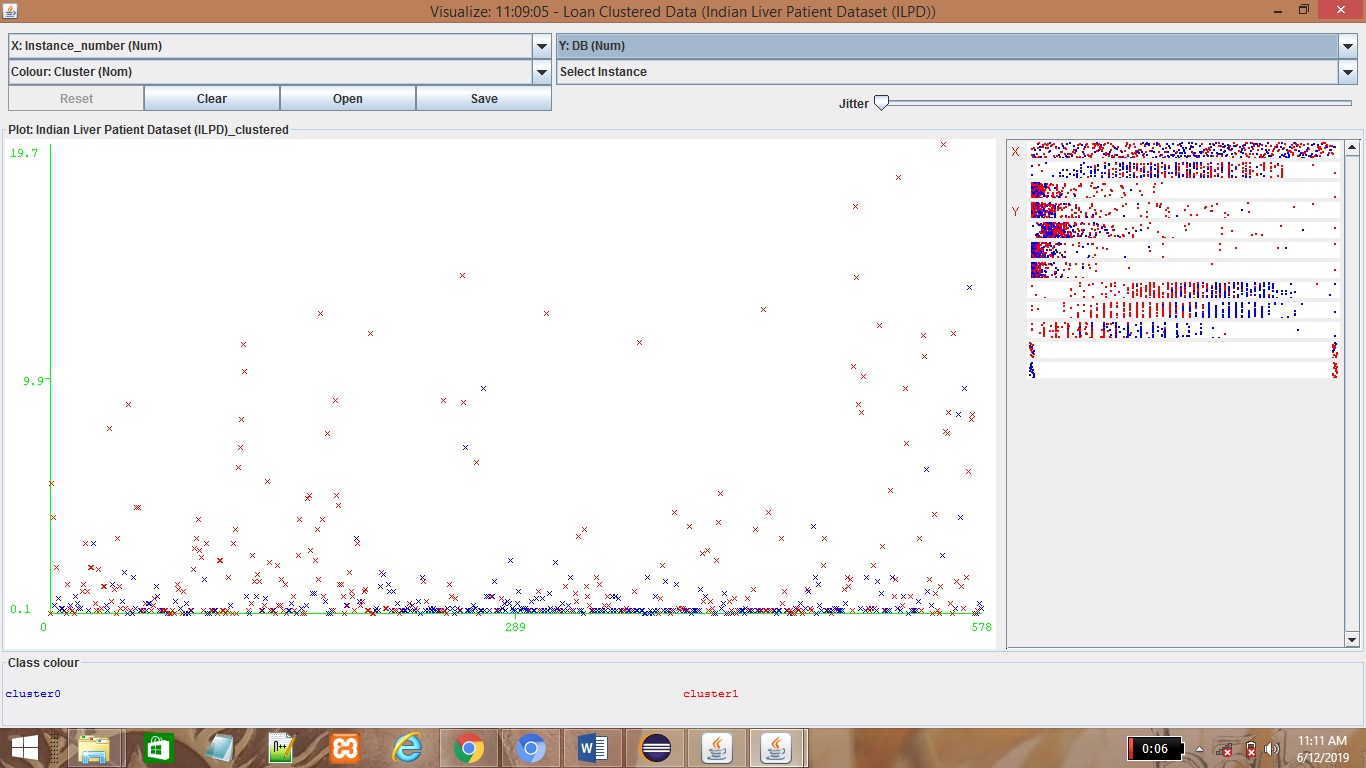
### Figure A6: Visualization of Cluster with respect to Age

Figure A6: This graph represents the Visualization of instances with respect to Age. The clustering of instances represents the presence of liver disease in patients.



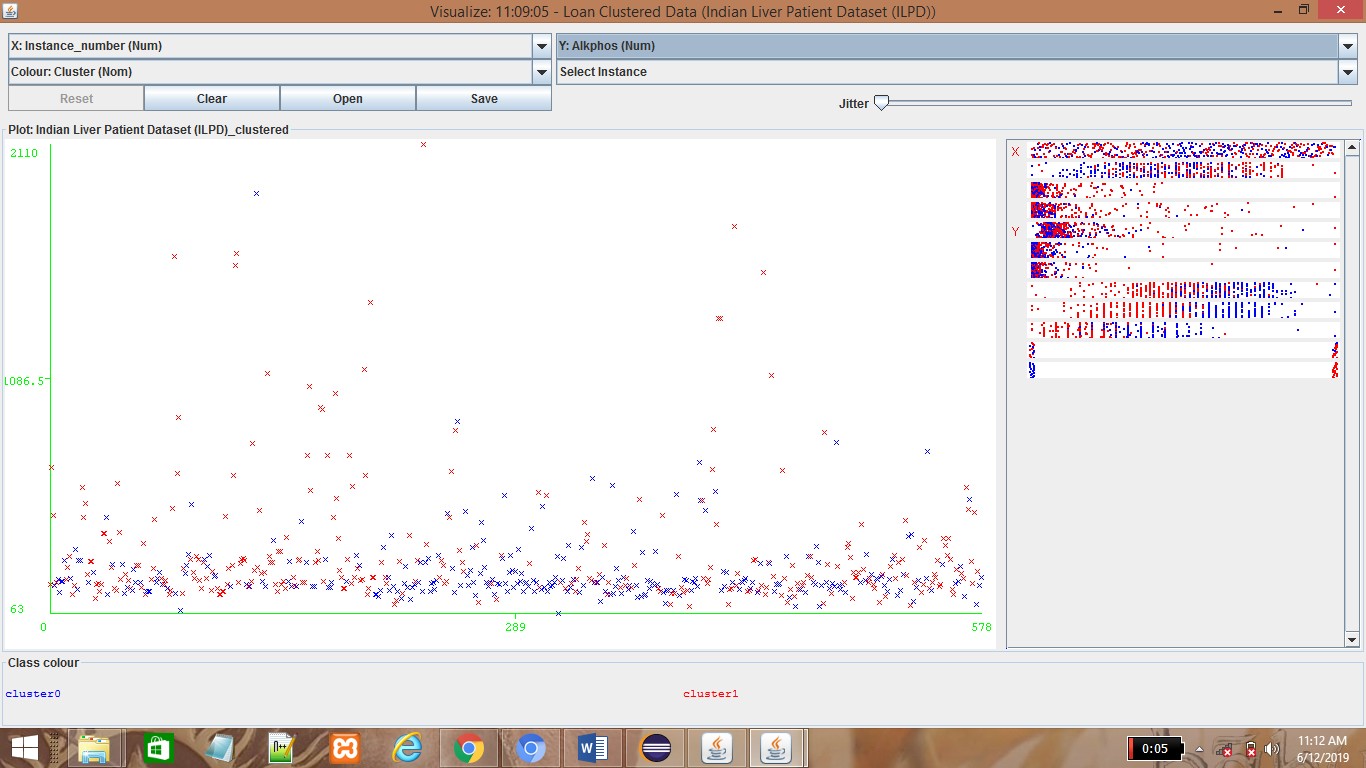
### Figure A7: Visualization of Cluster with respect to TB

Figure A7: This graph represents the Visualization of instances with respect to Total Bilirubin.The clustering of instances represents the presence of liver disease in patients with attribute Total Bilirubin.



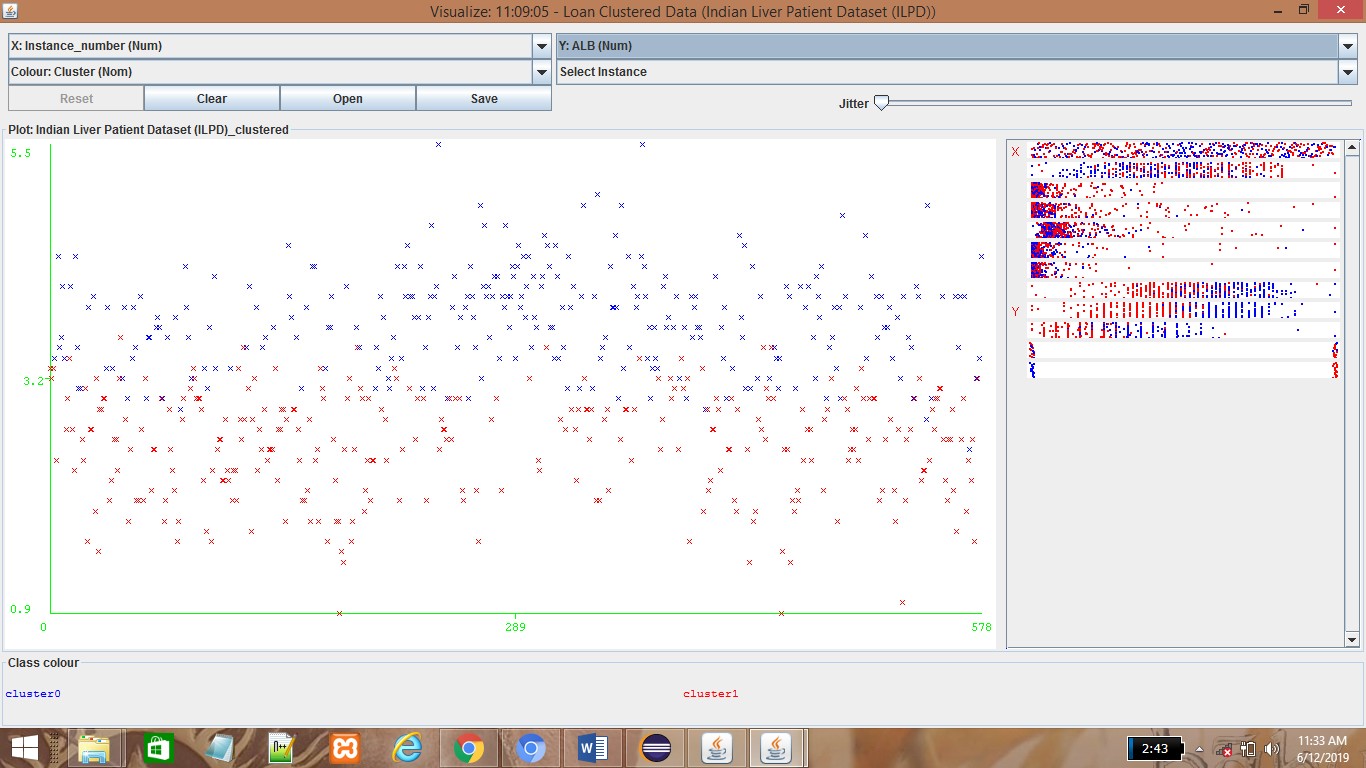
### Figure A8: Visualization of Cluster with respect to DB

Figure A8: This graph represents the Visualization of instances with respect to Direct Bilirubin .The clustering of instances represents the presence of liver disease in patients with respect to the attribute Direct Bilirubin.



### Figure A9: Visualization of Cluster with respect to Alkphos

Figure A9: This graph represents the Visualization of instances with respect to Alkaline Phosphatese. The clustering of instances represents the presence of liver disease in patients with respect to the attribute Alkphos. In the above graph blue color instances represents absence of liver disease and red color instances represents presence of liver disease.



### Figure A10: Visualization of Cluster with respect to ALB(num)

Figure A10: This graph represents the Visualization of instances with respect to Albumin. The clustering of instances represents the presence of liver disease in patients with respect to the attribute Albumin. In the above graph blue color instances represents absence of liver disease and red color instances represents presence of liver disease.

# REFERENCES

1. Ashwani Kumar, Neelam Sahu, “Categorization of Liver Disease Using Classification Techniques”, International Journal for Research in Applied Science & Engineering Technology (IJRASET), Volume 5 Issue V, May 2017, IC Value: 45.98 ISSN: 2321-9653.
2. Chandrasegar Thirumalai, IEEE Member, Rashad Manzoor, “Cost Optimization using Normal Linear Regression Method for Breast Cancer Type I Skin”, International Conference on Electronics, Communication and Aerospace Technology ICECA 2017.
3. Harsha Pakhale, Deepak Kumar Xaxa, “Development of an Efficient Classifier for Classification of Liver Patient with Feature Selection”, International Journal of Computer Science and Information Technologies, Vol. 7 (3), 2016,

1541-1544.

1. Dr. S. Vijayarani, Mr.S.Dhayanand, “Liver Disease Prediction using SVM and Naïve Bayes Algorithms”, International Journal of Science, Engineering and Technology Research (IJSETR) Volume 4, Issue 4, April 2015.
2. Anju Gulia, Dr. Rajan Vohra , Praveen Rani “Liver Patient Classification Using Intelligent Techniques”, International Journal of Computer Science and Information Technologies, Vol. 5 (4), 2014, 5110-5115.
3. Kalyan Nagaraj and Amulyashree Sridhar, “NeuroSVM: A Graphical User Interface for Identification of Liver Patients”, IJCSIT. 5(6): 8280-8284 (2014).
4. Jankisharan Pahareeya, Rajan Vohra, Jagdish Makhijani , Sanjay Patsariya, “Liver Patient Classification using Intelligence Techniques”, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 2, February 2014.
5. Bendi Venkata Ramana, Surendra. Prasad Babu. M, Venkateswarlu. N.B, “A Critical Study of Selected Classification Algorithms for Liver Disease Diagnosis”, International Journal of Database Management Systems (IJDMS), Vol.3, No.2, May 2011 page no 101-114.
6. Kotsiantis. S.B, “Increasing the Classification Accuracy of Simple Bayesian Classifier”, AIMSA, pp.198- 207, 2004.
7. V.Kirubha, S.Manju Priya, “Survey on Data Mining Algorithms in Disease Prediction”, International Journal of Computer Trends and Technology (IJCTT) – Volume 38 Number 3 - August 2016.
8. Bendi Venkata Ramana, Prof. M. S. Prasad Babu and Prof. N. B. Venkateswarlu, “Critical Comparative Study of Liver Patients from USA and INDIA: An Exploratory Analysis, International Journal of Computer Science Issues, ISSN: 1694 -0784, May 2012.
9. A.S.Aneeshkumarand C.JothiVenkateswaran, “Estimating the Surveillance of Liver Disorder using Classification Algorithms”, International Journal of Computer Applications (0975 – 8887), Volume 57– No.6, November 2012.
10. S. Karthik A, Priyadarishini, J. Anuradha and B. K. Tripathy (2011). “Classification and Rule Extraction using Rough Set for Diagnosis of Liver Disease and its Types”, Advances in Applied Science Research.
11. Esraa M Hashem, Mai S Mabrouk (2014). “A Study of Support Vector Machine Algorithm for Liver Disease Diagnosis”, American Journal of Intelligent Systems. 4(1): 9-14.