Project Synopsis

on

Smart Health Prediction System

Submitted as a part of course curriculum for

Bachelor of Technology
in
Computer Science

Submitted by
Deepak Goswami
(1900290120033)
Mohd. Zubair
(1900290120065)
Under the Supervision of
Prof. Shivani
Assistant Professor

KIET Group of Institutions, Ghaziabad
Department of Computer Science
Dr. A.P.J. Abdul Kalam Technical University

2021-2022

DECLARATION

We hereby declare that this submission is our work and that, to the best of our knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgement has been made in the text.

Signature of Students
Name:
Deepak Goswami
(1900290120033)
Mohd. Zubair
(1900290120065)
Date: 13/12/2021

CERTIFICATE

This is to certify that Project Report entitled "**Smart health prediction system**" which is submitted by **Mohd Zubair and Deepak Goswami** in partial fulfilment of the requirement for the award of degree B. Tech. in Department of Computer Science of Dr A.P.J. Abdul Kalam Technical University, Lucknow is a record of the candidates own work carried out by them under my supervision. The matter embodied in this report is original and has not been submitted for the award of any other degree.

Date:13/12/2021

Supervisor Signature

Prof.Shivani
(Assistant Professor)

ACKNOWLEDGEMENT

It gives us a great sense of pleasure to present the synopsis of the **B.Tech** Mini Project undertaken during B.Tech. Third Year. We owe a special debt of gratitude to **Prof. Shivani** Department of Computer Science**,** KIET Group of Institutions, Delhi- NCR, Ghaziabad, for her constant support and guidance throughout the course of our work. Her sincerity, thoroughness and perseverance have been a constant source of inspiration for us. It is only her cognizant efforts that our endeavours have seen the light of the day.

We also take the opportunity to acknowledge the contribution of **Dr. P. K Singh**, Head of the Department of Computer Science, KIET Group of Institutions, Delhi- NCR, Ghaziabad, for his full support and assistance during the development of the project. We also do not like to miss the opportunity to acknowledge the contribution of all the faculty members of the department for their kind assistance and cooperation during the development of our project.

Last but not the least, we acknowledge our friends for their contribution to the completion of the project.

Signature:
Date : 13/12/2021
Name : Deepak Goswami
(1900290120033)
Mohd. Zubair
(1900290120065)

ABSTRACT

To increase of age average led to an increase in the demand of providing and improving the service of healthcare. The advancing of the information and communication technology (ICT) led to the development of smart cities which have a lot of components. One of those components is Smart Health (s-Health), which is used in improving healthcare by providing many services such as patient monitoring, early diagnosis of diseases and so on. Nowadays there are many machine learning techniques that can facilitates s-Health services. The results show that the ML approach is used in many s-Health applications such as Glaucoma diagnosis, Alzheimer's disease, bacterial sepsis diagnoses, the Intensive Care Unit (ICU) readmissions, and cataract detection. The Artificial Neural Network (ANN), Support Vector Machine (SVM) algorithm, Decision tree classifier , Naïve Bayes classifier and deep learning models especially the Convolutional Neural Network (CNN) are the most commonly used machine learning approaches where they proved to get high evaluation performance in most cases.

TABLE OF CONTENTS

Content **Page no.**

CHAPTER 1

INTRODUCTION

1.1 Introduction to project

The core idea behind the project is to propose a system that allows users to get instant guidance on their health issues. This system is fed with various symptoms and the disease/illness associated with those systems. This system allows user to share their symptoms and issues It then processes user's symptoms to check for various illnesses that could be associated with it if the system is not able to provide suitable results, it informs the user about the type of disease or disorder it feels user's symptoms are associated with and also suggest the doctor to whom he or she can contact.

1.2 Problem statement

- Smart Health Care System has great potential to improve existing health care sector by using technology. Smart health care system will be of great use for providing solutions at home with ease. Homes are becoming centers of modern health care; therefore, by using technology create easy to use Smart Health Care System.
- There is some sort of resources available to predict smart health. However, chronic diseases have been studied in particular and a level of risk has been identified. However, these methods are not widely used for disease prediction in general disease. Smart health prediction helps in the diagnosis of multiple diseases by analysing patient symptoms using a perfect fitting Machine Learning Algorithm technique.

1.3 Objective

In our day to day life we must have faced a situation where we are worried about our health, we are feeling ill but do not have enough time to visit a doctor. Or we need to consult a good specialist but due to some circumstances we can't visit them. Smart health prediction system is an online consultation application. Here we offer an application that allow user to get instant guidance on their health status using intelligent health care online system. Our application is trained with various symptoms of diseases using Naïve Bayes algorithm. When user shares their symptoms, on the basis of those symptoms our application checks the possibility of diseases which could be associated with those symptoms, suggest nearby doctor to cure it and provide basic dos and don'ts.

1.4 Scope

This system can be accessed 24*7 using smart phones/ laptop/PC and is capable of giving common advice of dos and don’ts (prescribed by doctors in past). User can book appointment sitting in any remote area, i.e. rural people don’t need to visit separately for appointment and then for treatment.
Also user can get able to detect a disease at a very early age and can go through the online appointment with the best doctor. By maximizing the work efficiency and production the system will meet the patient’s needs while remaining easy to understand and use. More specifically, this system is designed to allow a patient to manage and communicate with doctors to a public website. The software will facilitate communication between patient, and doctor. The system also contains a relational database containing a list of Patients and Doctors.

CHAPTER 2

LITERATURE REVIEW

Smart Health Prediction Using Machine Learning

Machine learning is used in predictive modelling using certain instances. There are two stages of the machine learning algorithm: preparation and research. The signs and symptom logs of the user/patient are used to predict the illness. Machine Learning technology offers a strong application forum in the medical sector to address health disease prediction concerns based on the user/patient experience.We use machine learning to keep track of all signs and diseases. Machine learning technology helps predictive models to rapidly analyse data and produce meaningful results more quickly. With the aid of technology, the user/patient may make an informed decision to see a doctor about their particular symptoms, resulting in improved patient health services.The Naïve Bayes Classifier technique is used to analyse a large amount of data obtained. For each sub-field of Disease Predictions, we also demonstrated how symptom data storage combined with data classification can assist the administrative, clinical, academic and educational aspects of Disease Prediction from Symptoms. There are a host of data collection issues that can be discussed in terms of health prediction.

An Introduction to Logistic Regression Analysis and Reporting

In this it is demonstrate that logistic regression can be a powerful analytical technique for use when the outcome variable is dichotomous. The effectiveness of the logistic model was shown to be supported by
(a) significance tests of the model against the null model
(b) the significance test of each predictor
(c) descriptive and inferential goodness-of-fit indices
(d) and predicted probabilities.
During the last decade, logistic regression has been gaining popularity. The trend is evident in the JER and higher education journals. Such popularity can be attributed to researchers' easy access to sophisticated statistical software that performs comprehensive analyses of this technique. It is anticipated that the application of the logistic regression technique is likely to increase. This potential expanded usage demands that researchers, editors, and readers be coached in what to expect from an article that uses the logistic regression technique. What tables, charts, or figures should be included? What assumptions should be verified? And how comprehensive should the presentation of logistic regression results be? It is hoped that this article has answered these questions with an illustration of logistic regression applied to a data set and with guidelines and recommendations offered on a preferred pattern of application of logistic methods.

Study and Analysis of Decision Tree Based Classification Algorithms

Machine learning is to learn machine on the basis of various training and testing data and determines the results in every condition without explicit programmed. One of the

techniques of machine learning is Decision Tree. A normal tree includes root, branches and leaves.The same structure is followed in Decision Tree. It contains root node branches, and leaf nodes. Testing an attribute is on every internal node, the outcome of the test is on branch and class label as a result is on leaf node. A root node is parent of all nodes and as the name suggests it is the topmost node in Tree. A decision tree is a tree where each node shows a feature (attribute), each link (branch) shows a decision and each leaf shows an outcome (categorical or continues value) . As decision trees mimic the human level thinking interpretations. The whole idea is to create a tree like this for the entire data and process a single outcome at every leaf. The Decision Tree algorithms ID3 C4.5 and CART were applied on the dataset. Decision tree outperforms others in terms of accuracy, time and precision. It quite relies on the algorithm used for recommendation to find interesting resources. At last, the comprehensive study is done about decision tree algorithms and this paper concludes that CART is the algorithm for this dataset is very precise and most accurate among the others

An empirical study of the naive Bayes classifier

The naive Bayes classifier greatly simplify learning by assuming that features are independent given class. ich affect the performance of naive Bayes. Our approach uses Monte Carlo simulations that allow systematic study classification accuracy for several classes of randomly generated problems. We analyze the impact of the distribution entropy on the classification error, showing that certain almost deterministic, or low entropy, dependencies yield good performance of naive Bayes. Particularly, we demonstrate thatnaive Bayes works best in two cases: completely independent features (as expected) and functionally dependent features (which is surprising). Naive Bayes has its worst performance between these extremes. inally, a better understanding of the impact of independence assumption on classification can be used to devise better approximation techniques for learning efficient Bayesian net classifiers, and for probabilistic inference, e.g., for finding maximum-likelihood assignments.

A Review on Various Aspects of MongoDb Databases

MongoDB is an open-source document database that provides high performance, high availability, and automatic scaling. A record in MongoDB is a document, which is a data structure composed of field and value pairs. MongoDB documents are similar to JSON object.
MongoDB has document oriented storage; data is stored in the form of JSON style documents. It can be indexed on any attributes. We can also explain where to use MongoDB:

- Big Data

- Content Management and Delivery
- Mobile and Social Infrastructure
- User Data Management
- Data Hub

MongoDB is an open-source document database and leading NoSQL database. MongoDB is written in C++. MongoDB is a cross-platform, document oriented database that provides high performance, high availability and easy scalability. MongoDB works on concept of collection and document.

- Database - Database is a physical container for collections. Each database gets its own set of files on the file system.
- Collection - Collection is a group of MongoDB documents. It is the equivalent of a RDBMS table. Collections do not enforce a schema. Documents within a collection can have different fields. Typically, all documents in a collection are of similar or related purpose.
- Document - A document is a set of key- value pairs. Documents have dynamic schema. Dynamic schema means that documents in the same collection do not need to have the same set of fields or structure, and common fields in a collection's documents may hold different types of data. MongoDB is currently the most popular Document oriented DB, but it is hardly the most robust or performant implication because of its ability to fully utilize so called "shared-nothing cluster architecture."

SUPPORT VECTOR MACHINES: THEORY AND APPLICATIONS

Support Vector Machines (SVM) have been recently developed in the framework of statistical learning theory and have been successfully applied to a number of applications, ranging from time series prediction to face recognition to biological data processing for medical diagnosis Their theoretical foundations and their experimental success encourage further research on their characteristics, as well as their further use. In statistical learning theory (SLT) the problem of supervised learning is formulated as follows. We are given a set of l training data {(x1,y1)...(xl,yl)} in Rn × R sampled according to unknown probability distribution
P(x,y), and a loss function V(y,f(x)) that measures the error done when, for a given x, f(x) is "predicted" instead of the actual value y. The problem consists in finding a function f that minimizes the expectation of the error on new data, that is, find a function f that minimizes the expected error:

$$\int V(y,f(x))\, P(x, y)\, dx\, dy$$

An application of SVM regression was discussed in .The problem was time series prediction. The approach taken was the use of SVM regression to model the dynamics of the time series and subsequently predict future values of the series using the constructed model. Instead of using the standard SVM regression formulation described above, a variation developed in was used. Using this variation the ε parameter of the SVM regression loss function (see above) is automatically estimated.

Furthermore, used an approach to learning which is different from the standard one: instead of developing one global regression model from all the available training data, develops a number of SVM regression models, each one trained using only part of the initial training data

MongoDB – a comparison with NoSQL databases

Web based applications and their data management needs have changed dynamically in the past few years. Variety of features and strict data consistency is provided by the relational databases. Due to massive cost of storing and manipulating data in classical relational database systems, NoSQL databases have been developed. NoSQL databases provide more scalability and heterogeneity when compared to RDBMS. MongoDB, a NoSQL database provides high scalabilty, performance and availability. MongoDB is a document based NoSQL database designed for Internet and web based applications. Data model of MongoDB is easy to build on due to its inherent support for unstructured data.

Data models are designed to be quite flexible in order to support the storage needs arising from applications dealing with highly heterogeneous data. Also, the wide-spread use of dynamically typed scripting languages has made less strictly structured background storage system favourable. While a highly generic data model looks reasonable from the aspect of the client, efficient server-side processing makes certain restrictions on the data model necessary. As a result, many NoSQL systems offer semistructured models and list-like data types.

No SQL Data Models The main idea here is using a hash table where there is a unique key and a pointer to a particular item of data. The Key/value model is the simplest and easiest to implement. But it is inefficient when you are only interested in querying or updating part of a value, among other disadvantages. Column Family Stores were created to store and process very large amounts of data distributed over many machines. There are still keys but they point to multiple columns. The columns are arranged by column family.

Graph Databases ,Instead of tables of rows and columns and the rigid structure of SQL, a flexible graph model is used which, again, can scale across multiple machines. NoSQL databases do not provide a highlevel declarative query language like SQL to avoid overtime in processing. Rather, querying these databases is data-model specific. Many of the NoSQL platforms allow for RESTful interfaces to the data, while other offers query API.

MongoDB is a database management system designed for the Internet and web-based applications. It is a document based NoSQL database. The data model and persistence strategies are built for high read and write throughput and the ability to scale easily with automatic failover. MongoDB's document data model makes it easy to build on, since it has inherent support for unstructured data and does not require costly and time consuming migrations when application requirements change.. MongoDB uses BSON as network transfer format for documents.. MongoDB also supports rich queries and full indexes. This distinguishes it from other document databases in which a separate server layer is used to handle complex queries. Its other features include automatic

sharding, replication, and easy storage provided by these systems. When MongoDB was initially designed, security was not a primary concern of its designer.

Machine Learning- for web page adapatation

World Wide Web is a collection of varied documents constituting a huge repository of information available; the massive nature of the data imposes difficulty in extracting knowledge. Analyzing such large amount of data is a challenge and provides a fertile ground for further research. Web mining aims to discover useful information or knowledge from the web hyperlink structure, page content and usage data. The web page contents can also be organized in a tree like structure, using various HTML and XML tags within the page. This tree like structure is referred as Document Object Model (DOM) Tree. Fig 1: Classification of web mining Web content is analyzed by using machine learning and rule based algorithm to provide useful information regarding web data that can be utilized for various applications. Some of the domains are: 1. News Filtering. 2. Web Content Adaptation. 3. Spam Detection. 4. Noise Removal. 5. Search Result Mining. The web page segmentation module partitions the web page based on semantic and visual characteristic to provide various visual blocks Gi, where the value of i range from 1 to n, where n is the total number of blocks extracted from the web page S. Union of all these blocks combined together must provide the complete set of information from the web page. that describes the relevance of each block in context to the web page. Once the previous module has extracted the visual blocks, the machine learning (ML) algorithm is applied to obtain the block importance. Each ML technique requires specific feature set, which is prepared from the data received by the web page segmentation module. Block identification can be utilized for topic specific search where user is interested in finding the useful content related to any topic from different web site. The main content from different web sites can be clubbed and displayed to the user. In this paper we proposed a system to extract visual blocks from web page, classify the blocks using machine learning techniques and to segregate informative and non informative content efficiently. Finally the system filters the informative content and provide accurate and faster user perceive adaptation of content on the mobile devices. Further research work can extend this to incorporate evolutionary algorithms to web page layout optimization.

A Step-by-Step Guide To Build ML Models For Research

Building machine learning models for research can be challenging, especially if you are developing them from scratch. Several factors come into play before you successfully deploy ML models for production. Some of these aspects include data collection, training data in a limited amount of time, nonrepresentative training data, poor quality of data, irrelevant/unwanted features, overfitting the training data, underfitting the training data and offline learning, development of the model, etc. Building models is one of the most important parts of machine learning. With modern ML frameworks, it is easy to throw all techniques at your data and see what works. This, sometimes, leads to a disorganised mess of experiments that is hard to justify and record. Lones believes that approaching model building in an organised manner is the way forward. Comparing models is one of the most common practices in AI/ML

research work, but it is also challenging to get it right. Lones says that if a researcher carries out a comparison unfairly and publishes it, other researchers may subsequently be led astray. Therefore, he suggested that researchers evaluate different models within the same context, explore multiple perspectives, and utilise statistical tests appropriately. Here are some of the critical things to consider before reporting the results of your model:

- Be transparent
- Do report performance in multiple ways
- Do not generalise beyond the data
- Be careful when reporting statistical significance

CHAPTER 3

PROPOSED METHODOLOGY

3.1 METHODOLOGY

Inside the application there will be three modules :

1> User Module : This module will be responsible for all user end tasks. Such as giving symptoms, fixing an appointment, searching nearby specialist for their disease diagnosis, searching doctors profile and experience etc.
2> Doctor's Module : This module will be responsible for all doctor related tasks. Such as keeping record of his online Users, their appointment conformation and reminder etc.
3> Admin Module : This module will be responsible for all type administration which are needed for smooth working and to update model time to time.

Algorithm used are:

1> Naïve Bayes Classifier
2> Decision tree Classifier
3> Support Vector Machine
4> Logistic Regression

Key Features :

• **Disease Prediction** : After asking certain questions regarding illness application will predict the disease, give common tips of dos and don'ts and finally will suggest doctors in nearby location.
• **Search Doctor** : User can directly search doctor and fix their appointment.
• **Registration** : Both User and Doctor have to register before accessing any services.
• **User's Login** : After login using User's login, User's user interface will be accessible.
• **Doctor's Login** : After login using doctor's login, Doctor's user interface will be accessible.
• **Profile Feedback** : This will enable user (both User and user to rate behavior of other after diagnosis).

• **Application Feedback** : If user want to give any suggestion or any bug occur, they can give feedback about problem they are facing here.

3.2 FLOW CHART

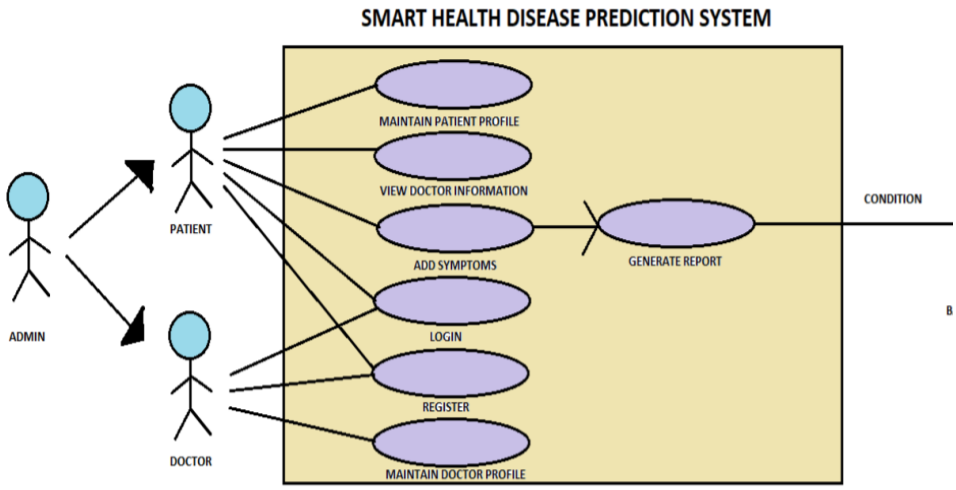

Fig 3.1.Depicting smart health prediction system

CHAPTER-4

PROPOSED ALGORITHM

NAÏVE BAYES ALGORITHM:- Naive Bayes algorithm is a classification algorithm based on Bayes' theorems use in predictive modeling and this algorithm uses Bayesian techniques .This algorithm is less computationally intense then other and therefore is useful for quickly generating mining models to discover relationships between input columns and predictable columns.

DECISION TREE ALGORITHM:- Decision Tree algorithm belongs to the family of supervised learning algorithms. Unlike other supervised learning algorithms, the decision tree algorithm can be used for solving regression and classification problems too.
The goal of using a Decision Tree is to create a training model that can use to predict the class or value of the target variable by learning simple decision rules inferred from prior data(training data).
In Decision Trees, for predicting a class label for a record we start from the root of the tree. We compare the values of the root attribute with the record's attribute. On the basis of comparison, we follow the branch corresponding to that value and jump to the next node.

4.1 ER DIAGRAM

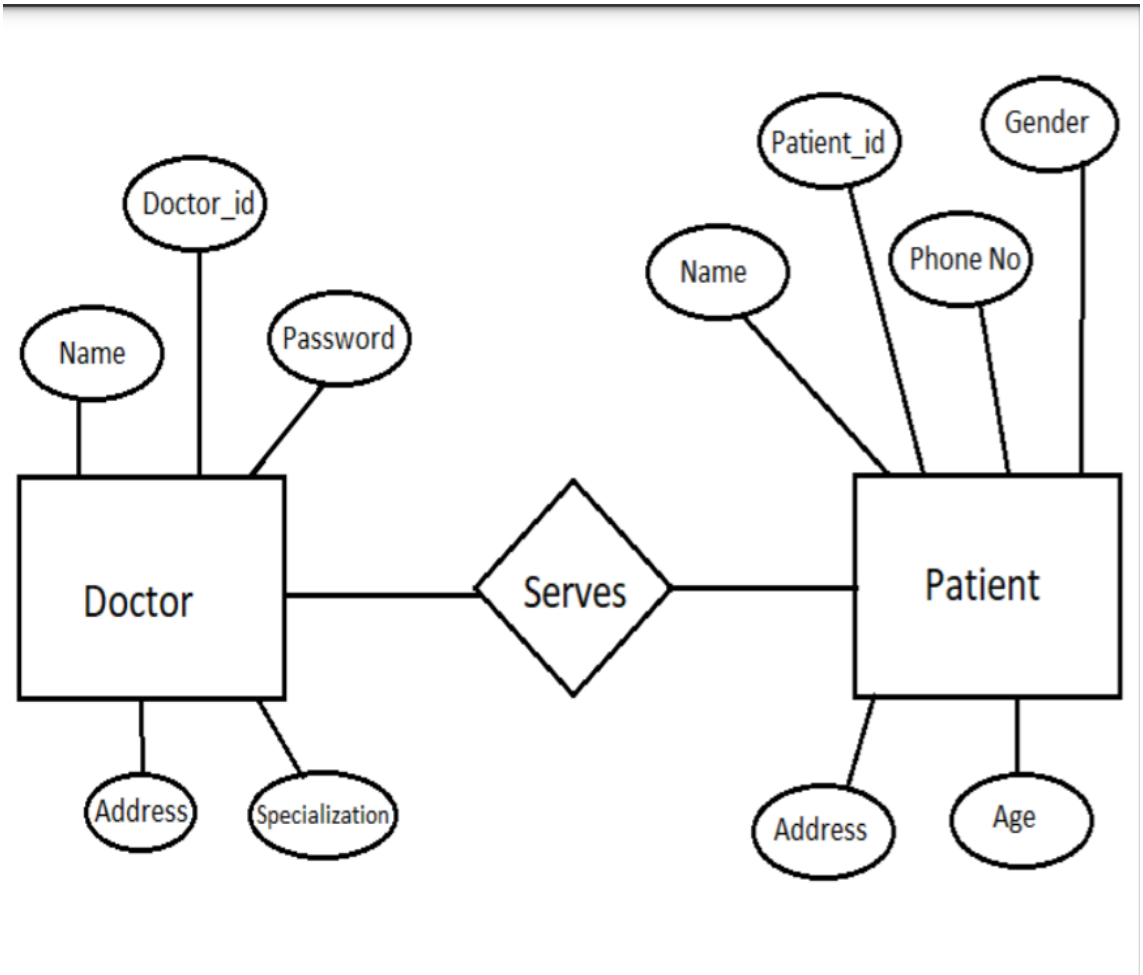

Fig 3.1. ER Diagram

CHAPTER- 5

CONCLUSION

Predict the amount of possibility of have a particular type of disease.Provide the best doctors for consultation as per the symptoms given by the user.Create the database of the user medical history.

REFERENCES

[1] Naveen Kumar and Kirubhakaran “smart health prediction using machine learning”, procedia-International Research Journal on Advanced Science Hub, 2021, Volume 3,2015.Availaible :10.1016/j.sbpro.2014.12.12.573.

[2] CHAO-YING JOANNE and PENGUK LIDA LEEGARY “An introduction to logistic regression analysis and reporting” procedia-computer science, vol. 14 pp .28-35 ,2012 avilaible 10.1016/j.sbspro.2014.12.573.

[3] Harsh Patel and Purvi Prajapati” Study and Analysis of Decision Tree BasedClassification Algorithms”procedia computer science ,vol. 96, pp. 446-455 2016. availaible: 10.1016/j.procs.2016.08.009.

[4] Rishi T.J” an emperical study of naïve bayes classifier”, Procedia science , vol. 173 , pp, 264-271 ,2020. Available : 10.1016/j.procir.2021.01.124.

[5] Anjali Chauhan “ a review on various aspects of mongo db” ,procedia Computer science , vol , 183, , pp, 364-371 ,2020, 10.1016/j.procir.2021.01.124.

[6] Theodorus Evegeniou “ Support vector machines :theory and applications “ Procedia computer science , vol 101 , pp,446-455 2018. Available : 10.1016/j.procir.2021.01.124.

[7] Hema Krishnan and Neetu Narwal “ mongo db- a comparison with no sql Database” procedia computer science ,vol 7 ,no. 3, p.e06277,2021, available :10.1016/j.heliyon.2021.e06277

[8] Neetu Narwal abd dr. Sanjay kumar sharma “machine learning for web page Adpatation” procedia technology ,vol. 11 .pp . 903-911,2013.

[9] Anjali Chauhan “ A Step-by-Step Guide To Build ML Models For Research” procedia technology , vol 8 ,no. 5, p.e062899 , 2020, avalaible:10.1016/j.procir.2020.01.124

RELATED WORK

In this section research efforts and machine learning technique has been discussed and algorithms such as Naive Bayes algorithm which is a classification algorithm based on Bayes' theorems and is use in predictive modeling and this algorithm uses Bayesian techniques as well as DECISION TREE ALGORITHM:- Decision Tree algorithm belongs to the family of supervised learning algorithms.Several research paper on mongo db and a comparsion of mongo db with no sql has been discussed in detail. Apart from this research paper related to machine learning for web page adapation is also discussed.Objective of the research is to compare machine learning algorithms on the basis of performance and also comparison of databases is done in detail .