REC QUESTION PAPER GENERATOR

# MINI PROJECT REPORT

***Submitted by***

## PRAVINESH H 210701194

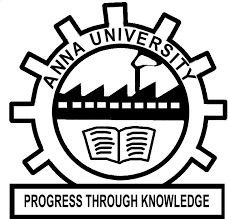
**RAM PRAKASH L** **210701208**

***in partial fulfillment for the award of the degree of***

# BACHELOR OF ENGINEERING

***in***

**COMPUTER SCIENCE AND ENGINEERING**



**RAJALAKSHMI ENGINEERING COLLEGE, CHENNAI ANNA UNIVERSITY:: CHENNAI 600 025**

**MAY 2024**

**BONAFIDE CERTIFICATE**

Certified that this Report titled “**REC QUESTION PAPER GENERATOR**” is the bonafide work of **“PRAVINESH H (210701194) and RAM PRAKASH (210701208)”**who carried out the work under my supervision. Certified further that to the best of my knowledge the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

# SIGNATURE

## Dr. [Rakesh Kumar M](mailto:rakeshkumar.m@rajalakshmi.edu.in), M.E, PhD.,

Assistant Professor,

Department of Computer Science and Engineering, Rajalakshmi Engineering College,

Chennai – 602015

Submitted to Mini Project Viva-Voce Examination held on

## Internal Examiner External Examiner

**ABSTRACT**

Disease prediction using machine learning is the system that is used to predict the diseases from the symptoms which are given by patients or any other user. The system processes the symptoms provided by the user as input and gives the output as the probability of the disease. Naive bayes, KNN, Decision tree, Random forest classifiers is used in the prediction of the disease which is supervised machine learning algorithm. With an increase in biomedical and healthcare data, accurate analysis of medical data benefits early disease detection and patient healthcare. By using linear regression and decision tree we are predicting diseases like Diabetes, Malaria, Jaundice, Dengue and Hypertension etc.., In conclusion, machine learning has the potential to revolutionize disease prediction by leveraging large-scale healthcare data and advanced algorithms. By addressing the challenges and harnessing the strengths of machine learning, healthcare professionals can better predict and manage diseases, ultimately leading to more efficient and personalized healthcare. This abstract serves as a foundation for further exploration into the promising field of disease prediction in machine learning.

**ACKNOWLEDGEMENT**

Initially we thank the Almighty for being with us through every walk of our life and showering his blessings through the endeavour to put forth this report. Our sincere thanks to our Chairman **Mr. S.MEGANATHAN, B.E, F.I.E.**, our Vice Chairman **Mr. ABHAY SHANKAR MEGANATHAN, B.E., M.S.,** and our

respected Chairperson **Dr. (Mrs.) THANGAM MEGANATHAN**, **Ph.D.,** for providing us with the requisite infrastructure and sincere endeavoring in educating us in their premier institution.

Our sincere thanks to **Dr. S.N. MURUGESAN, M.E., Ph.D.,** our beloved Principal for his kind support and facilities provided to complete our work in time. We express our sincere thanks to **Dr. P. KUMAR, Ph.D.**, Professor and Head of the Department of Computer Science and Engineering for his guidance and encouragement throughout the project work. We convey our sincere and deepest gratitude to our internal guide, **Dr. Rakesh Kumar M, PhD.,** Assistant Professor, Department of Computer Science and Engineering. Rajalakshmi Engineering College for his encouragement and guiding us throughout the project to

build our project.

**PRAVINESH H (210701194)**

**RAM PRAKASH L (210701208)**

**TABLE OF CONTENTS**

**CHAPTER NO.**

**TITLE PAGE**

**NO.**

|  |  |
| --- | --- |
| **ABSTRACT** | **iii** |
| **ACKNOWLEDGEMENT** | **iv** |
| **LIST OF FIGURES** | **vii** |
| **LIST OF TABLES** | **viii** |
| **LIST OF ABBREVIATIONS** | **ix** |
| **1. INTRODUCTION** | **10** |
| 1.1 GENERAL | 10 |
| 1.2 OBJECTIVE | 10 |
| 1.3 EXISTING SYSTEM | 10 |
| 1.4 PROPOSED SYSTEM | 10 |
| **2. LITERATURE SURVEY** | **11** |
| **3. SYSTEM DESIGN** | **15** |
| 3.1 DEVELOPMENT ENVIRONMENT | 15 |
| 3.1.1 HARDWARE SPECIFICATIONS | 15 |
| 3.1.2 SOFTWARE SPECIFICATIONS | 15 |
| 3.2 SYSTEM DESIGN | 16 |
| 3.2.1 ARCHITECTURE DIAGRAM | 16 |

1. PROJECT DESCRIPTION 18
   1. MODULES DESCRIPTION 18
2. IMPLEMENTATION AND RESULTS 19
   1. [IMPLEMENTATION 19](#_TOC_250002)
   2. [OUTPUT SCREENSHOTS 21](#_TOC_250001)

6. 22

CONCLUSION AND FUTURE ENHANCEMENT

* 1. [CONCLUSION 22](#_TOC_250000)
  2. FUTURE ENHANCEMENT 22

**REFERENCES** **2**

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **S.NO** | **NAME** | **PAGE NO** |
| 3.3.1 | ARCHITECTURE DIAGRAM | 16 |

**LIST OF TABLES**

|  |  |  |
| --- | --- | --- |
| **S.NO** | **NAME** | **PAGE NO** |
| 3.2.1 | HARDWARE SPECIFICATIONS | 15 |
| 3.2.2 | SOFTWARE SPECIFICATIONS | 15 |

**LIST OF ABBREVIATIONS**

|  |  |
| --- | --- |
| **OC** | Organized Crimes |
| **SVM** | Support Vector Machines |
| **CMS** | Crime Monitoring System |
| **KNN** | K Nearest Neighbours |
| **FIR** | First Information Report |
| **JSON** | Java Script Object Notation |

**CHAPTER 1**

**INTRODUCTION**

Machine Learning is the domain that uses past data for predicting. Machine Learning is the understanding of a computer system under which the Machine Learning model learns from data and experience. The machine learning algorithm has two phases: 1) Training & 2) Testing. To predict the disease from a patient’s symptoms and from the history of the patient, machine learning technology is struggling from past decades. Healthcare issues can be solved efficiently by using Machine Learning Technology. We are applying complete machine learning concepts to keep track of patient’s health.

ML model allows us to build models to get quickly cleaned and processed data and deliver results faster. By using this system doctors will make good decisions related to patient diagnoses and according to that, good treatment will be given to the patient, which increases improvement in patient healthcare services. To introduce machine learning in the medical field, healthcare is the prime example. To improve the accuracy of large data, the existing work will be done on unstructured or textual data. For the prediction of diseases, the existing will be done on linear, KNN, Decision Tree algorithm.

With the help of developed big data analytics technology, more attention has been paid to patients disease prediction from the perspective of big data analysis, various researches have been researched and conducted by selecting the characteristics automatically from a larger number of data to improve the efficiency and accuracy of risk classification and reduction rather than the previously selected characteristics. However, those existing works are mostly considered structured data

# CHAPTER 2

**LITERATURE SURVEY**

( Dr C K Gomathy et al,2021)[1] In disease prediction, machine learning (ML) processes people's symptoms to predict the likelihood of various diseases. Naive Bayes classifier is a supervised machine learning algorithm that calculates disease incidence to aid in early diagnosis and patient care. Combining regression and decision trees can improve the prediction of diseases such as diabetes, malaria, jaundice, dengue fever, and tuberculosis and use the physical growth of biomedical information to improve health outcomes.

( Sethi, R. S et al ,2019)[2] Today, more mobile phones are used in the world than ever before. Cell phones are everywhere and mobile technology is growing exponentially. The functionality of mobile phones allows them to provide us with services that make people's lives better. One of the services that mobile phones can provide us is digital therapy. Additionally, mobile apps are recognized to provide cost-effective healthcare solutions. Such applications provide easy and portable healthcare services for everyone. Apps like these provide users with rich experiences where they can learn more about their health and fitness. Mobile digital health apps can use a patient's symptoms to diagnose their illness. Doctors can use this information for further consultation.

(Geluvaraj et.al., 2022)[3] Research approaches disease prediction based on the prediction models and using classifiers and classification techniques. These approaches using the KNN , RFC , NB demonstrate the high accuracy of comparing one another and making them promising for healthcare applications requiring precise predictions.

(Singh, R et.al., 2019 )[5] Precise early diagnosis is crucial for improving prognosis and survival rates,spurring advancement in ai for interpreting cardiovascular data to identify risk factors and manifestation of disorders.

(S Vijayarani et al 2020)[6]Their study elaborates the data have become important for predicting clinical outcomes and extracting useful information from general medical data. Disease prediction from big medical data enables researchers to perform tasks such as classification and policies. This study focuses on the prediction of liver disease using extraction methods, specifically naive Bayes and Support Vector Machine (SVM). The performance of this method, testing time and testing time showed that the vector material was before gambling.

(S Mohan et al, 2019)[7]Cardiovascular disease is still the leading cause of death worldwide, so its prediction is important in the analysis of medical data. Machine learning (ML) has proven useful in predictive analytics using big medical data and has also been incorporated into Internet of Things (IoT) applications. Current research is only a fraction of the machine learning research into predicting heart disease. This paper presents a new method to improve the accuracy of prediction by identifying key features and using various classification methods. Our hybrid model combines random forests and methods to increase the accuracy of heart disease detection by 88.7%.

(Kumar, Y., et al,2023)[10]Artificial Intelligence (AI) improves patient care and health by using machine learning and deep learning to perform tasks such as disease diagnosis and drug discovery. The research examined AI tools used to diagnose Alzheimer's, cancer, diabetes, heart disease, tuberculosis, stroke, high blood pressure, and more, using different clinical data such as MRI and CT scans. Articles up to October 2020 are selected from archives such as Web of Science, Scopus and PubMed. The survey compares studies on metrics such as accuracy, sensitivity, specificity, AUC, precision, recall and F1 score, demonstrating the effectiveness of clinical skills and patient treatment.

**CHAPTER 3**

**SYSTEM DESIGN**

* 1. **DEVELOPMENT ENVIRONMENT**
     1. **HARDWARE SPECIFICATIONS**

This project uses minimal hardware but in order to run the project efficiently without any lack of user experience, the following specifications are recommended

**Table 3.1.1 Hardware Specifications**

|  |  |
| --- | --- |
| **PROCESSOR** | Intel Core i5 |
| **RAM** | 4GB or above (DDR4 RAM) |
| **GPU** | Intel Integrated Graphics |
| **HARD DISK** | 6GB |
| **PROCESSOR FREQUENCY** | 1.5 GHz or above |

# SOFTWARE SPECIFICATIONS

The software specifications in order to execute the project has been listed down in the below table. The requirements in terms of the software that needs to be pre- installed and the languages needed to develop the project has been listed out below.

**Table 3.1.2** Software Specifications

|  |  |
| --- | --- |
| **FRONT END** | Streamlit |
| **BACK END** | Python |
| **FRAMEWORKS** | LangChain, Streamlit, Groq, Ollama |
| **SOFTWARES USED** | Python IDLE, Google Chrome |

# SYSTEM DESIGN

* + 1. **ARCHITECTURE DIAGRAM**



## Fig 3.2.1 Architecture Diagram

**ALGORITHM:**

**NAIVE BAYES:**

Naive Bayes is an easy however amazingly powerful rule for prognosticative modeling. The independence assumption that allows decomposing joint likelihood into a product of marginal likelihoods is called 'naive'. This simplified Bayesian classifier is called naive Bayes. The Naive Bayes classifier assumes the presence of a particular feature in a class is unrelated to the presence of any other feature. It is very easy to build and useful for large datasets. Naive Bayes is a supervised learning model. Bayes theorem provides some way of calculative posterior chance P(b|a) from P(b), P(a) and P(a|b).

Look at the equation :

P(b v a)= P(a v b)P(b)/P(a) Above,

P(b|a) is the posterior chance of class (b,target) given predictor (a, attributes)

.

P(b) is the priori probability of class.

P(a|c) is the chance of a predictor given the class. P(a) is the priori probability of a predictor.

In our system, Naïve Bayes decides which symptom is to put in the classifier and which is not. 8.3 LOGISTIC REGRESSION Logistic regression could be a supervised learning classification algorithm accustomed to predict the chance of a target variable that is Disease.

**DECISION TREE**

A decision tree is a structure that can be used to divide up a large collection of records into successfully smaller sets of records by applying a sequence of simple decision tree. With each successive division, the members of the resulting sets become more and more similar to each other. A decision tree model consists of a set of rules for dividing a large heterogeneous population into smaller, more homogeneous (mutually exclusive) groups with respect to a particular target. The target variable is usually categorical and the decision tree is used either to: Calculate the probability that a given record belongs to each of the categories and, To classify the record by assigning it to the most likely class (or category). In this disease prediction system, the decision tree divides the symptoms as per its category and reduces the dataset difficulty.

**RANDOM FOREST ALGORITHM**

Random Forest is a supervised learning algorithm. It is an extension of machine learning classifiers which include the bagging to improve the performance of Decision Tree. It combines tree predictors, and trees are dependent on a random vector which is independently sampled. The distribution of all trees are the same. Random Forests splits nodes using the best among of a predictor subset that are randomly chosen from the node itself, instead of splitting nodes based on the variables. The time complexity of the worst case of learning with Random Forests is O(M(dnlogn)) , where M is the number of growing trees, n is the number of instances, and d is the data dimension.It can be used both for classification and regression. It is also the most flexible and easy to use algorithm. A forest consists of trees. It is said that the more trees it has, the more robust a forest is. Random Forests create Decision Trees on randomly selected data samples, get predictions from each tree and select the best solution by means of voti pretty good indicator of the feature importance. Random Forests have a variety of applications, such as recommendation engines, image classification and feature selection. Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model.As the name suggests, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output.The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

**K NEAREST NEIGHBOR:**

k-Nearest Neighbors (kNN) is a simple machine learning method. This article introduces some basic concepts of the kNN algorithm and then focuses on the use of R for kNN models. After predicting the results using the kNN algorithm, the diagnosis of the model should be checked. True mean is the most commonly used method regarding the kNN algorithm. Factors such as the k value, distance calculation, and selection of the appropriate variable value all have an impact on the performance of the model and The nearest neighbor (KNN) algorithm is a general machine learning algorithm used for classification tasks, including disease pr

**CHAPTER 4**

## IMPLEMENTATION

### **MACHINE LEARNING:**

In machine learning, classification refers to a predictive modeling problem where a class label is predicted for a given example of input data.

* Supervised Learning:

Supervised learning is the type of machine learning in which machines are trained using well "labeled" training data, and on the basis of that data, machines predict the output. The labeled data means some input data is already tagged with the correct output. In supervised learning, the training data provided to the machines work as the supervisor that teaches the machines to predict the output correctly. It applies the same concept as a student learns in the supervision of the teacher. Supervised learning is a process of providing input data as well as correct output data to the machine learning model. The aim of a supervised learning algorithm is to find a mapping function to map the input variable(x) with the output variable(y).

* Unsupervised learning:

Unsupervised learning cannot be directly applied to a regression or classification problem because unlike supervised learning, we have the input data but no corresponding output data. The goal of unsupervised learning is to find the underlying structure of the dataset, group that data according to similarities, and represent that dataset in a compressed format.

* Unsupervised learning is helpful for finding useful insights from the data.
* Unsupervised learning is much similar to how a human learns to think by their own experiences, which makes it closer to the real AI.
* Unsupervised learning works on unlabeled and uncategorized data which make unsupervised learning more important.
* In real-world, we do not always have input data with the corresponding output so to solve such cases, we need unsupervised learning.
* Reinforcement learning

Reinforcement learning is an area of Machine Learning. It is about taking suitable action to maximize reward in a particular situation. It is employed by various software and machines to find the best possible behaviour or path it should take in a specific situation. Reinforcement learning differs from supervised learning in a way that in supervised learning the training data has the answer key with it so the model is trained with the correct answer itself whereas in reinforcement learning, there is no answer but the reinforcement agent decides what to do to perform the given task. In the absence of a training dataset, it is bound to learn from its experience.

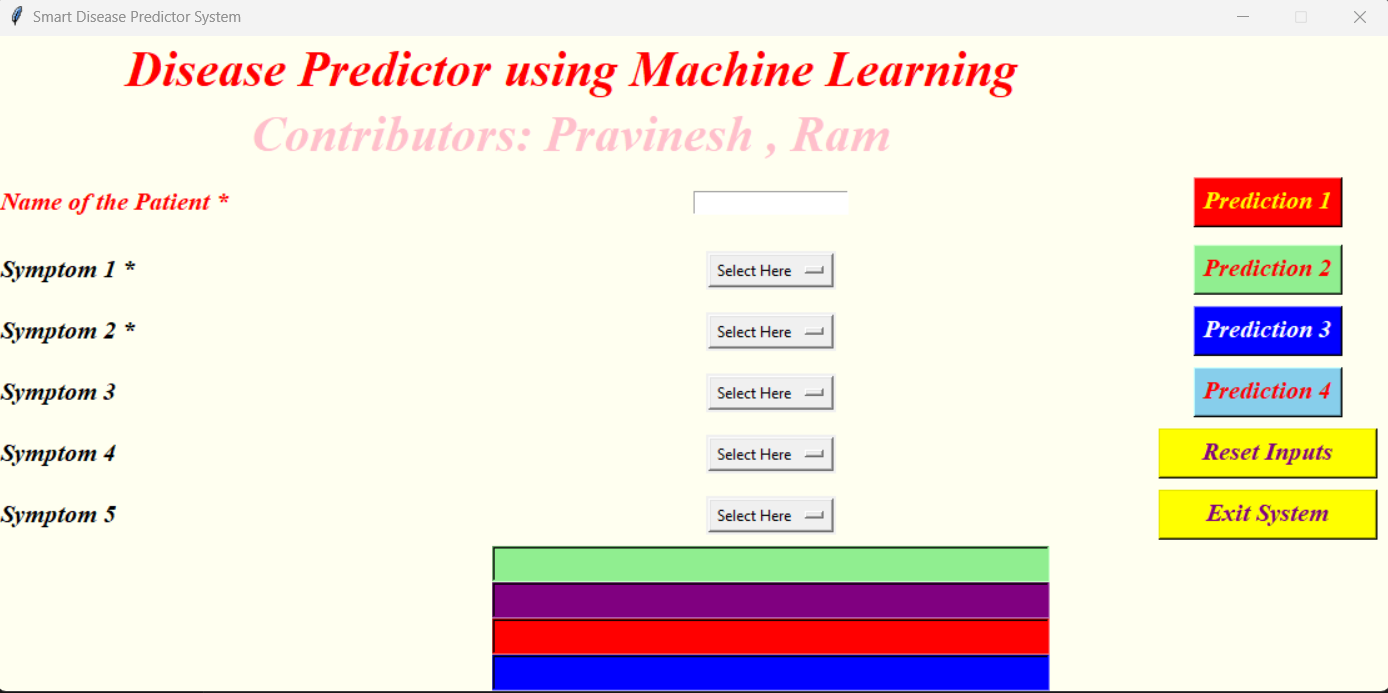
**CHAPTER 5**

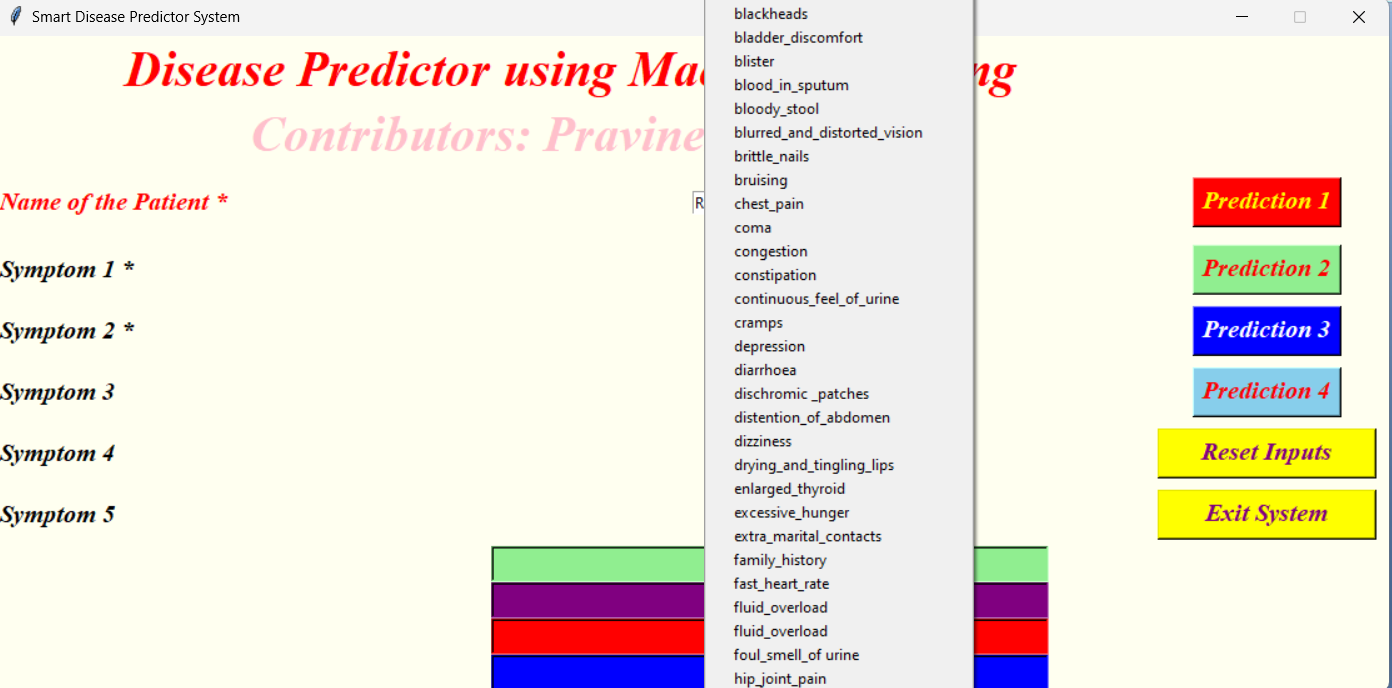
**IMPLEMENTATION AND RESULTS**

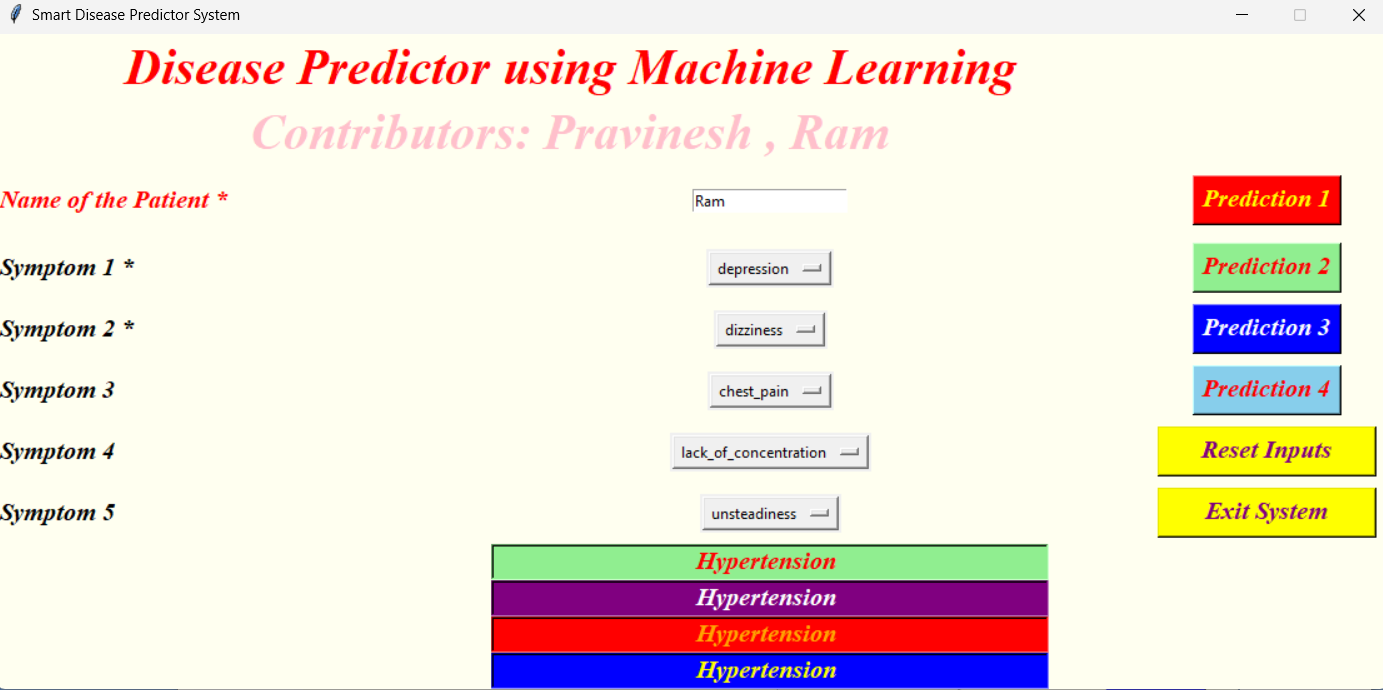
# IMPLEMENTATION

1. Analyze the problem in terms of what we want to predict and what kind of observation data we have to make those predictions. Predictions are generally a label or a target answer; it may be a yes/no label (binary classification) or a category (multiclass classification) or a real number (regression).
2. Collect and clean the data and Identify what kind of historical data we have for prediction modeling. The next step is to collect the data from datasets or from any other data sources.
3. Prepare data for ML application Transform the data in the form that the Machine Learning system can understand.
4. Prepare the Graphical User Interface (GUI) of the model Graphical User Interface (GUI) is designed for taking input and displaying output.
5. Train the model: Before training the model, it is essential to split the data into training and evaluation sets, as we need to monitor how well a model generalizes to unseen data. Now, the algorithm will learn the pattern and mapping between the feature and the label.
6. Evaluate and improve model accuracy: Accuracy is a measure to know how well or bad a model is doing on an unseen validation set. Based on the current learning, evaluate the model on validation sets.

# OUTPUT SCREENSHOTS







**CHAPTER 6**

**CONCLUSION AND FUTURE ENHANCEMENTS**

# CONCLUSION

In conclusion, this application not only accelerates the question paper generation process but also improves the quality and relevance of the papers generated. Automating the monotonous aspects of question paper creation enables professors to spend more time focusing on other aspects of teaching and engaging their students. Furthermore, this application creates a level playing field for all assessment aspects by ensuring that all subjects and topics are well-represented in the question papers. Moreover, institutions can customize the question papers to align with their individual curriculum demands and assessment objectives. Be it based on the difficulty of questions, some topics that should be given weight, or the types of questions to include, the educators will have the freedom to adapt question papers to themes that suit their various categories. By leveraging technologies such as natural language processing, large language models, and retrieval augmented generation, the application empowers educators to create high-quality question papers efficiently, benefiting both students and institutions.

Through NLP, LLM, and RAG, the application facilitates educators in generating the best question papers with high efficiency, benefiting students and the institution in various ways. With the integration of NLP, the application delves into the contents of subject notes and thus derives questions that are contextually relevant and meet the educational criteria. The application aims to create appropriate and meaningful questions that will test the comprehension and critical thinking skills of students with paramount effectiveness.Furthermore, the RAG increases precision and relevance by integrating the information retrieval capability. This approach not only makes the assessment process much more enriching but also includes the capacity for the educator to draw from a wide range of sources while formulating questions. The application saves time by automating tasks that an educator has to do, like formatting and arranging questions in a specified pattern while generating question papers. This allows professors to spend more time on interactive teaching methods and student support.Most importantly, the application tailors question papers according to specific demands from the curriculum and objectives of the assessment, thus ensuring that educators develop assessments for their specific needs and educational goals. Whether it is question difficulty, topic weighting, or picking the right type of question, the educator has the freedom to design an assessment that will be best at measuring understanding and progress.This is a leap in educational technology because it makes assessment practices easier, faster, and individualistic. This application automatically removes

tedious administrative jobs from educators' tasks and enhances the quality and relevance of assessments. Ultimately improving student learning outcomes and contributing to the excellence of education at the institution.

# FUTURE ENHANCEMENTS

Adaptive Learning Integration: Use adaptive learning strategies to change and revise questions to reflect on individual students’ performance. It can also manipulate the type of questions and the level or area of concentration which relate to previous accomplishment by the learner.

Advanced Analytics and Reporting: Leakage – Use advanced analytical tools to issue the reports on quality and difficulty level of the questions covered in the question paper. The use of past assessments can shed light on overall trends on the performance of students.

Integration with Learning Management Systems (LMS):Integration with Learning Management Systems (LMS): Integrate with general LMS platforms – such as Moodle, Blackboard e. g. Instructors should advocate for an active adoption of LMSs such as Moodle, Blackboard, or Canvas for the ease of transferring quiz and assessment questions.

# REFERENCES

**[1]. The prediction disease using machine learningß || Dr C K Gomathy , Mr.A.Rohith Naidu,”Volume:05” “Issue:10”||Oct 2021||IJSREM.**

**[2]. Sethi, R. S., Thumar, A., Jain, A. V., & Chavan, S. (2019). Disease prediction application based on symptoms. International Journal of Scientific Research in Computer Science Engineering and Information Technology, 641-646.**

**[3]. Geluvaraj , B., Santhosh, K., Prabhu, N., Reddy, A., Sandhya, T., & Bhaskar, S. V. (2022, January). A hybrid approach for predicting diseases using clustering and classification techniques. In 2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI) (pp. 1-6). IEEE\**

**[4]. S Vijayarani , S Dhayanand Liver disease prediction using SVM and Naive bayes algorithms International Journal of Science, Engineering and Technology Research (IJSETR) , volume 4 , issue 4 Posted: 2015**

**[5]. S Mohan , C Thirumalai , G Srivastava Effective heart disease prediction using hybrid machine learning techniques IEEE Access , volume 7 Posted: 2019**

**[6]. Kumar, Y., Koul, A., Singla, R., & Ijaz, M. F. (2023). Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda. Journal of ambient intelligence and humanized computing, 14(7), 8459-8486**



















