**HEALTH PREDICTOR**

**Submitted for**

**Artificial Intelligence and Machine Learning (CSET301)**

Submitted by:

**(E23CSEU1723) ISHIKA**

Submitted to

**DR. NITIN ARVIND SHELKE**

**Jan-Apr 2025**

**SCHOOL OF COMPUTER SCIENCE AND ENGINEERING**

A close-up of a logo

Description automatically generated

**1. Abstract**

In recent years, healthcare has seen a significant transformation due to advances in Artificial Intelligence (AI) and Machine Learning (ML). These technologies are now being applied to various domains, especially in early diagnosis and preventive healthcare. The Health Predictor project is an intelligent system designed to predict potential diseases based on user-input symptoms using machine learning algorithms. The tool allows individuals to receive basic health insights, increasing health awareness and enabling early medical consultation. This system is a step toward reducing dependency on immediate medical consultations for every symptom and enhancing self-awareness among users.

**2. Introduction**

With increasing access to technology, more people are seeking preliminary medical advice online. However, information overload and unverified content often lead to confusion. Health Predictor offers a reliable solution using supervised ML models trained on curated medical datasets. The user enters a list of symptoms, and the system predicts the most likely disease or health issue. This prediction, while not a professional diagnosis, gives the user direction and guidance toward appropriate action. The tool is especially useful in rural or remote areas where immediate access to healthcare professionals might not be available.

**3. Related Work**

There are existing health analysis systems and applications such as:

* **IBM Watson Health**: Uses AI for cancer treatment recommendations.
* **Ada Health**: Mobile app that guides users based on symptom input.
* **WebMD Symptom Checker**: Web-based tool offering potential causes of symptoms.

However, most of these tools are either proprietary, expensive, or lack transparency in prediction logic. Our project differentiates itself by being open-source, lightweight, and focused on educational use, while still being able to deliver reliable predictions for basic symptoms.

**4. Methodology**

The process for developing the Health Predictor system involved several phases:

**4.1 Data Collection**

* Dataset sourced from Kaggle containing records with symptoms and diagnosed diseases.
* Each row in the dataset corresponds to a set of symptoms mapped to a specific disease.

**4.2 Data Pre-processing**

* Handling missing values and duplicates.
* Converting symptoms into numerical features using label encoding.
* Applying standardization where necessary.

**4.3 Model Training**

* Splitting the dataset into training and testing sets (80:20 ratio).
* Algorithms implemented:
  + **Random Forest Classifier**
  + **Decision Tree Classifier**
  + **Naive Bayes**
  + **Support Vector Machine (SVM)**
* Performance metrics such as accuracy, precision, and recall were calculated.

**4.4 Model Evaluation**

* Random Forest achieved the best results with:
  + Accuracy: **94%**
  + Precision: **0.93**
  + Recall: **0.94**

**4.5 Deployment**

* The trained model was deployed using **Flask**.
* Users enter symptoms on a web form, and the server returns a predicted disease.

**5. Hardware/Software Requirements**

**Hardware:**

* Laptop or Desktop with minimum 4 GB RAM
* Internet connection for deployment and testing

**Software:**

* **Python 3.10+**
* **Flask** for backend integration
* **scikit-learn, pandas, numpy** for data science
* **VS Code/Jupyter Notebook** for development
* **HTML/CSS/JavaScript** for the frontend
* **Git/GitHub** for version control and project management

**6. Experimental Results**

After training the model, we tested it with real-world inputs (symptom combinations) and observed the following:

| **Test Case** | **Input Symptoms** | **Predicted Disease** | **Accuracy** |
| --- | --- | --- | --- |
| TC-01 | fever, headache, chills | Malaria | 94% |
| TC-02 | chest pain, short breath | Heart Disease | 92% |
| TC-03 | fatigue, weight loss, thirst | Diabetes | 95% |
| TC-04 | cough, sore throat, runny nose | Common Cold | 93% |

The system successfully predicted most common illnesses, and its accuracy varied slightly depending on symptom similarity across diseases.

**7. Conclusion**

The Health Predictor project demonstrates the usefulness of machine learning in the healthcare domain. While it is not a replacement for clinical diagnostics, it is a powerful tool for raising health awareness and enabling early action. By providing preliminary health predictions, users can make informed decisions, seek medical help when necessary, or track patterns in recurring symptoms.

**8. Future Scope**

To improve the Health Predictor, the following future developments are proposed:

* **Deep Learning Integration**: Use of neural networks for better generalization and learning from large datasets.
* **Voice-Based Symptom Input**: Users can speak their symptoms instead of typing.
* **Mobile App Version**: Making the tool accessible on Android/iOS.
* **Medical Chatbot Integration**: Conversational interface to ask symptom-based questions.
* **Real-Time Feedback**: Suggest immediate actions such as drinking fluids or resting, depending on the symptoms.
* **User Data Storage (With Consent)**: Allow users to track their health records over time.

**9. GitHub Link of the Complete Project**

**https://github.com/ishikabeniwal**