

Panacea

A Project Synopsis Report

Submitted by
**PANNAG SHAH (C066),
DEEP SHAH (C070),
SAMIT SHAH (C072),
DEV DHAWAN(C088)**

Under the Guidance of

Dr. Seema Shah

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

**BACHELOR OF TECHNOLOGY (Integrated)
COMPUTER ENGINEERING
At**



**MUKESH PATEL SCHOOL OF TECHNOLOGY
MANAGEMENT AND ENGINEERING**

August 2021

Table of Content

Chapter No.	Title	Page No.
1	Introduction	3
2	Problem Statement	3
3	Motivation	4
4	Purpose of the Project	4
5	Project Scope	4
6	Project Objectives	5
7	Plane of Action	5
8	Impact	5
-	References	6

Chapter 1

Introduction

Panacea, is a Greek word that means to cure all diseases and prolong life indefinitely. Our project aims to enhance the existing healthcare system and help in solving post recovery side effects and symptoms helping the medical professionals to help cure patients' symptoms at the onset of a sever disease. We aim to work using Machine Learning and Graph Database using Python and Neo4j.

Chapter 2

Problem Statement

Rather than the knowledge-rich data hidden in the data set and databases, decisions are frequently made purely on doctors' intuition and expertise. Machine Learning has a lot of promise in the healthcare business since it allows health organizations to systematically use data and analytics to uncover inefficiencies and best practices in order to enhance care. If the condition can be predicted ahead of time, several diseases can be effectively controlled with a combination of lifestyle changes, medicine, and, in some circumstances, surgery. People nowadays suffer from a variety of diseases because of the environment and their lifestyle choices.^[1]

As a result, predicting sickness at an early stage becomes crucial. However, doctors find it challenging to make precise predictions based on symptoms. The most difficult challenge is correctly predicting sickness. To solve this problem, data mining plays a critical role in disease prediction. Medical science generates a lot of data every year. The accurate analysis of medical data has been benefited from early patient care due to an increased amount of data growth in the medical and healthcare area. Data mining uses disease data to uncover hidden pattern information in massive amounts of medical data.

We developed a broad disease prediction based on the patient's symptoms. The projected outcomes can be utilized to avoid and thereby lower the cost of future surgical treatment and other expenses.

Chapter 3

Motivation

One of the most serious consequences of disorganized data is that it jeopardizes patient safety. Because they are being treated based on the medical records of a completely different patient, one patient will receive erroneous and potentially harmful treatment.

Chapter 4

Purpose of the project

With the rapid advancement of technology and data, the healthcare domain is one of the most significant study fields in the contemporary era. The enormous amount of patient data is tough to manage. Big Data Analytics makes it easier to manage this information.^[2] There are numerous ways for treating various ailments all throughout the world. Machine Learning is a new approach that aids in disease prediction and diagnosis. This research shows how machine learning can be used to predict disease based on symptoms. On the presented dataset, machine learning algorithms such as Naive Bayes, Decision Tree, and Random Forest are used to forecast the disease. The python programming language is used for its implementation. Based on their accuracy, the research shows which algorithm is the best. The performance of an algorithm on a given dataset determines its accuracy. Our goal is to apply Machine Learning to develop an accurate model that can predict future symptoms/diseases with greater accuracy and confidence score.

Chapter 5

Project Scope

The use of deep learning and machine learning (ML) in medical science is increasing, particularly in the visual, audio, and language data fields. We aimed to build a new optimized ensemble model by blending a DNN (deep neural network) model with two ML models for disease prediction using laboratory test results. 86 attributes (laboratory tests) were selected from datasets based on value counts, clinical importance-related features, and missing values. We collected sample datasets on 5145 cases, including 326,686 laboratory test results. We investigated a total of 39 specific diseases based on the International Classification of Diseases, 10th revision (ICD-10) codes. These datasets were used to construct light gradient boosting machines (LightGBM) and extreme gradient boosting (XGBoost) ML models and a DNN model using TensorFlow. The optimized ensemble model achieved an F1-score of 81% and prediction accuracy of 92% for the five most common diseases. The deep learning and ML models showed differences in predictive power and disease classification patterns. We used a confusion matrix and analyzed feature importance using the SHAP value method. Our new ML model achieved high efficiency of disease prediction through classification of diseases. This study will be useful in the prediction and diagnosis of diseases. Integration of clinical decision support with

computer-based patient records could reduce medical errors, enhance patient safety, decrease unwanted practice variation, and improve patient outcome. This suggestion is promising as data modelling and analysis tools, e.g., data mining, have the potential to generate a knowledge-rich environment which can help to significantly improve the quality of clinical decisions

Chapter 6

Project Objectives

1. Integrate multiple datasets (RxNorm, MED-RT, SemMedDB, MeSH, etc) that specify relationships among UMLS concepts to build a knowledge graph.
2. Using machine learning algorithms on this dataset to yield more accurate output.
3. Create a highly accurate system for medicinal practitioners and the general public as well to provide better healthcare service and prevent deaths.

Chapter 7

Plan of action to complete the project

1. Gain a general understanding of what each dataset is.
2. Use documentation and the data from each dataset to develop a data model that specifies how each dataset connects to the UMLS.
3. Import the UMLS concepts as nodes into a graph.
4. Import each of the other databases to create relationships among UMLS concepts in the graph.
5. Test various graph visualization tools to display the knowledge graph in the most convenient way for clinicians.

Chapter 8

Impact

1. AI increases the ability for healthcare professionals to better understand the day-to-day patterns and needs of the people they care for, and with that understanding they can provide better feedback, guidance and support for staying healthy.
2. AI can lead to better care outcomes and improve the productivity and efficiency of care delivery. It can also improve the day-to-day life of healthcare practitioners, letting them spend more time looking after patients and in so doing, raise staff morale and improve retention.
3. The proliferation of consumer wearables and other medical devices combined with AI is also being applied to oversee early-stage heart disease, enabling doctors and other caregivers to better monitor and detect potentially life-threatening episodes at earlier, more treatable stages.

References

[1] World Health Organization

[2] Centres for Disease Control and Prevention