# **MINOR PROJECT-1**

**SYNOPSIS**

**ON**

## **HealthCare Data Analysis**

**Submitted By**

Student Name-1: Rishabh Yadav Student Name-2: Kirit Somani

SAP-ID-1: 500122500 SAP-ID-2: 500123413

***Under the guidance of***

**Dr. Neeraj Chugh**

**Associate Professor**

**Data Science Cluster**



**School of Computer Science**

### **UNIVERSITY OF PETROLEUM AND ENERGY STUDIES**

**Dehradun**

**January-May, 2025**

**Index**

**Contents Page Number**

Chapter 1: Abstract 3

Chapter 2: Introduction 3

Chapter 3: Problem Statement 3

Chapter 4: Literature 3

Chapter 5: Objective 3

Chapter 6: Methodology 4

Chapter 7: System Requirements 4

Chapter 8: PERT Chart 4

Chapter 9: SWOT Analysis 6

Chapter 10: References 6

**Chapter 1: Abstract**

This project aims to develop a data-driven healthcare analysis system that takes patient symptoms as input and provides a risk assessment of possible diseases, recommended medications, and suggested treatment plans. The system is designed to assist medical professionals and patients in early diagnosis and treatment planning through a web-based application.

**Chapter 2: Introduction**

Advancements in artificial intelligence and machine learning have enabled predictive analytics in healthcare. This project leverages data science to enhance diagnosis, optimize treatment recommendations, and improve healthcare accessibility.

**Chapter 3: Problem Statement**

Many individuals experience difficulty in recognizing potential health risks based on symptoms. Delayed diagnosis and incorrect medication can lead to serious health complications. This project aims to bridge this gap by using AI models to predict disease risks and suggest the necessary treatments efficiently.

**Chapter 4: Literature**

* Overview of AI in healthcare
* Machine learning techniques for disease prediction
* Previous research on symptom-disease mapping and diagnosis automation

**Chapter 5: Objective**

* Develop a machine learning model for disease prediction.
* Provide a risk assessment based on symptom severity.
* Recommend appropriate medications and treatment.
* Develop a web application for easy user interaction.

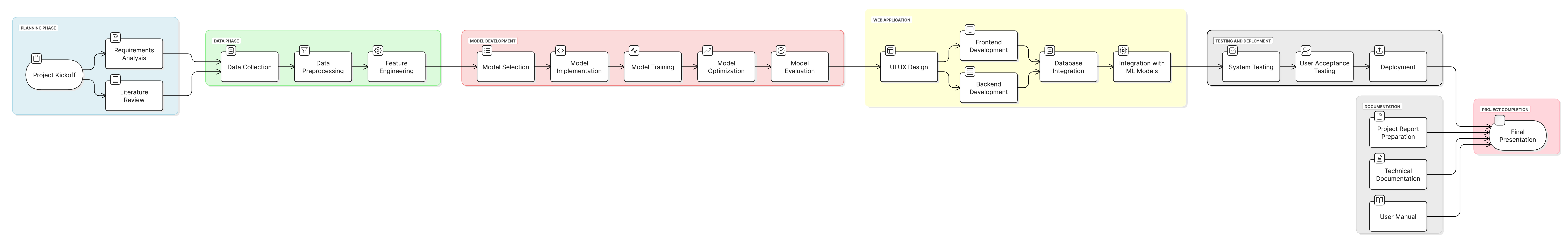
**Chapter 6: Methodology**

1. **Data Collection & Preprocessing:** Cleaning and structuring healthcare data.
2. **Feature Engineering:** Identifying significant health indicators.
3. **Model Development:**
   * Implementing supervised learning models (Decision Trees, Random Forest, Neural Networks).
   * Utilizing ensemble techniques for accuracy improvement.
4. **Model Evaluation & Optimization:** Using accuracy, precision, recall, and F1-score.
5. **Web Application Development:** Creating an intuitive user interface for diagnosis.

**Chapter 7: System Requirements**

* **Programming Languages:** Python, Java
* **Frameworks:** Flask/Django (backend), HTML, CSS, JavaScript (frontend)
* **Database:** MySQL, Kaggle

**Chapter 8: Pert Chart**

****

**Planning Phase**

* Project Kickoff: February 2, 2025
* Requirements Analysis: February 2-6, 2025 (5 days)
* Literature Review: February 2-9, 2025 (8 days, concurrent with requirements)

**Data Phase**

* Data Collection: February 10-16, 2025 (7 days)
* Data Preprocessing: February 17-21, 2025 (5 days)
* Feature Engineering: February 22-26, 2025 (5 days)

**Model Development**

* Model Selection: February 27-28, 2025 (2 days)
* Model Implementation: March 1-5, 2025 (5 days)
* Model Training: March 6-10, 2025 (5 days)
* Model Optimization: March 11-15, 2025 (5 days)
* Model Evaluation: March 16-18, 2025 (3 days)

**Web Application**

* UI/UX Design: March 19-23, 2025 (5 days)
* Frontend Development: March 24-30, 2025 (7 days)
* Backend Development: March 24-30, 2025 (7 days, parallel with frontend)
* Database Integration: March 31 - April 4, 2025 (5 days)
* Integration with ML Models: April 5-9, 2025 (5 days)

**Testing & Deployment**

* System Testing: April 10-14, 2025 (5 days)
* User Acceptance Testing: April 15-17, 2025 (3 days)
* Deployment: April 18-20, 2025 (3 days)

**Documentation** (compressed and overlapping)

* Technical Documentation: March 15-22, 2025 (8 days)
* User Manual: March 20-25, 2025 (6 days)
* Project Report Preparation: April 1-15, 2025 (15 days)

**Project Completion**

* Final Presentation: April 25, 2025

**Chapter 9: SWOT ANALYSIS**

**Strengths:**

* AI-driven diagnosis for faster healthcare decisions.
* Accessible via a web application.

**Weaknesses:**

* Dependency on data accuracy.
* Challenges in patient data privacy.

**Opportunities:**

* Integration with wearable devices for real-time monitoring.
* AI-driven chatbot for medical guidance.

**Threats:**

* Regulatory compliance challenges.
* Potential user mistrust in AI-generated diagnoses.

**Chapter 10: References**

1. Jackins V, Vimal S, Kaliappan M, Lee MY (2021) AI-based smart prediction of clinical disease using random forest classifier and Naive Bayes. J Supercomput 77(5):5198–5219. <https://doi.org/10.1007/s11227-020-03481-x>
2. Qayyum A, Qadir J, Bilal M, Al-Fuqaha A (2021) Secure and robust machine learning for healthcare: a survey. IEEE Rev Biomed Eng 14:156–180. <https://doi.org/10.1109/RBME.2020.3013489>
3. Siddique S, Chow JCL (2021) Machine learning in healthcare communication. Encyclopedia 1(1):220–239. <https://doi.org/10.3390/encyclopedia1010021>