Early Disease Detection and Resource Optimization Using Machine Learning

Final Thesis  
  
In Partial Fulfillment of the Requirements for the Degree of  
Master in Computer Science  
  
Student Name: Roshan Shrestha  
Student ID: 2598090  
Supervisor: [To be Filled]

# Abstract

The integration of machine learning (ML) in healthcare has opened new avenues for predictive analytics, particularly in the early detection of critical medical conditions and the strategic management of hospital resources. This dissertation presents a data-driven framework that applies supervised and unsupervised machine learning models to real-world hospital patient data to achieve two core objectives: (1) early identification of patients with critical illnesses and (2) optimization of hospital operations through risk-based clustering and forecasting. A structured dataset containing patient demographics, medical conditions, admission types, medication, billing, and clinical test results was used for model training and evaluation.  
  
The supervised learning component utilizes a Random Forest classifier to distinguish between critical and non-critical conditions, achieving an accuracy exceeding 70% after implementing class balancing with SMOTE and feature engineering strategies. SHAP (SHapley Additive exPlanations) was used to interpret model predictions, ensuring transparency and clinical trust. For unsupervised learning, K-Means clustering was employed to stratify patients based on risk indicators such as age, billing amount, and length of stay. The performance of K-Means was benchmarked against Agglomerative Clustering and DBSCAN using Silhouette Scores and Davies-Bouldin Index, with K-Means outperforming the alternatives in cohesion and separation.  
  
To support resource planning, an ARIMA-based time series model was developed to forecast hospital admission trends, offering actionable insights for managing capacity. The results demonstrate that the proposed framework can effectively support clinical decision-making and hospital management by delivering interpretable, accurate, and scalable predictions. This dissertation concludes that machine learning holds significant potential for enhancing patient care and operational efficiency in hospital environments.

# Acknowledgements

I would like to express my deepest gratitude to my academic supervisor, [Supervisor’s Name], whose guidance, feedback, and encouragement were invaluable throughout the course of this dissertation. Their expertise in data science and healthcare analytics has greatly enriched the direction and depth of this research.  
  
I also extend sincere thanks to the faculty and staff of the [Your Department Name], [Your University Name], for creating a supportive academic environment and providing access to the computational tools and resources necessary for this project.  
  
Special appreciation goes to my family and friends for their unwavering support, patience, and motivation during challenging moments. Their encouragement helped me stay focused and committed to delivering high-quality research.  
  
Lastly, I thank the open-source data science community and all researchers whose work provided a foundation for the development of this dissertation. Without their contributions, this research would not have been possible.

# Table of Contents

TO BE UPDATED IN WORD

# Chapter 1 Introduction

TO BE INSERTED

# Chapter 2 Literature Review/Related Work

TO BE INSERTED

# Chapter 3 Methodology

TO BE INSERTED

# Chapter 4 Experimental Results

TO BE INSERTED

# Chapter 5 Conclusion and Future Work

TO BE INSERTED

# References

TO BE INSERTED

# Appendix A. Title of Appendix

TO BE INSERTED

# Appendix B. Title of Appendix

TO BE INSERTED