

Predictive Modeling of Patient Readmission in EHR Data

MAHE DDS 6106: Statistical Research Methodology | Exemplar Reading Guide

· Classification: Instructional Master's Thesis Teaching Model (5-Chapter Baseline)

.. Reading Time: 20-25 mins | · Difficulty: · Beginner | · Audience: Postgraduate Master's Degree Candidates

? Why Selected: The baseline 5-chapter Master's thesis model. Teaches how problem formulation, literature gap, methodology, empirical results, and clinical discussion connect.

STRUCTURAL COLOR LEGEND & ANNOTATION GUIDE

Title & Abstract: High-impact summary, study rationale, & indexing keywords

Introduction & Literature: Research problem, background literature, & study hypothesis

Methods & Study Design: Study design, participants, sample size, SAP, & ethics approval

Results & Data Analysis: Empirical findings, statistical tables, p-values, & figures

Discussion & Conclusion: Interpretation, comparison with literature, limitations, & conclusion

.. THESIS STRUCTURAL ROADMAP

· Front Matter (Title, Abstract, Declaration, TOC)

· Chapter 1: Introduction & Research Problem

· Chapter 2: Literature Review & Gap

· Chapter 3: Research Methodology & SAP

· Chapter 4: Empirical Results & Analysis

· Chapter 5: Discussion & Limitations

· Back Matter: References & Appendices

· MILESTONE ICONS TO LOCATE

· Research Problem

High 30-day hospital readmission rates

· Research Question

Can machine learning predict readmission risk?

· Hypothesis

Random Forest outperforms logistic regression

· Literature Gap

Lack of temporal EHR validation in regional cohorts

· Methodology

Retrospective cohort, 70/30 train-test split

· Analysis & Results

AUC-ROC = 0.84, sensitivity = 0.79

· Discussion

Clinical utility & hospital deployment limits

· HOW A THESIS BECOMES A JOURNAL PAPER (IMRaD)

Students often ask how their 150-page Master's thesis becomes a paper. Chapter 4 (Results) and Chapter 3 (Methods) are extracted and condensed into an 8-page IMRaD manuscript for journal submission.

Full Monograph (150-200 Pages)

Contains literature review, raw tables, complete proofs, & code

· Select One Primary Question

Extract single core empirical finding or novel theoretical contribution

· Condense to 8-10 Pages (IMRaD)

Introduction (Ch1) -> Methods (Ch3) -> Results (Ch4) -> Discussion (Ch5)

· Submit to Peer-Reviewed Journal

Peer review, revision, and open-access journal publication

PREDICTIVE MODELING OF PATIENT READMISSION IN EHR DATA

A Dissertation Presented for the Degree of Thesis Blueprint 1: Traditional Master's Archetype

Department of Applied Statistics and Data Science

Prasanna School of Public Health, MAHE, Manipal

DECLARATION & CERTIFICATE

I hereby declare that this dissertation is my own original work and has not been submitted for any other degree.

ABSTRACT

ABSTRACT SUMMARY:

This dissertation presents a comprehensive investigation into predictive modeling of patient readmission in ehr data. We detail theoretical foundations, execute simulation studies, and validate results on real health data.

CHAPTER 1: INTRODUCTION

BACKGROUND & CONTEXT:

Hospital readmissions represent a critical burden on public health systems. Identifying high-risk patients early allows targeted discharge interventions.

RESEARCH PROBLEM:

Existing clinical risk scores fail to incorporate non-linear temporal interactions present in Electronic Health Records (EHR).

RESEARCH OBJECTIVES:

1. Construct a standardized EHR feature matrix.
2. Train gradient boosted trees and random forest classifiers.
3. Evaluate predictive performance using cross-validation.

CHAPTER 2: LITERATURE REVIEW

REVIEW OF PRIOR WORK:

Early studies by LACE and HOSPITAL scores relied on simple linear point scales. Recent biostatistical literature has transitioned to machine learning classifiers.

RESEARCH GAP:

Prior studies did not validate models on regional cohorts with missing data imputation.

CONCEPTUAL FRAMEWORK:

Input EHR Features -> Data Preprocessing & Imputation -> ML Classifier -> Risk Stratification.

CHAPTER 3: METHODOLOGY

STUDY DESIGN:

Retrospective observational cohort study utilizing 500 patient EHR records from 2021-2024.

VARIABLES & MEASURES:

Predictors: Age, BMI, comorbid conditions, prior admissions. Outcome: 30-day readmission (0/1).

STATISTICAL ANALYSIS PLAN:

Multivariable logistic regression baseline compared against Random Forest and XGBoost using 5-fold cross-validation. Metrics: AUC-ROC, Sensitivity, Specificity.

ETHICS APPROVAL:

Approved by MAHE Institutional Ethics Committee (Protocol #2024-889).

CHAPTER 4: RESULTS

DESCRIPTIVE STATISTICS:

The cohort comprised 500 patients (mean age = 62.4 years, 52% female). 30-day readmission rate was 18.4% (n=92).

CLASSIFIER PERFORMANCE:

Random Forest achieved the highest discrimination (AUC-ROC = 0.84, 95% CI: 0.80-0.88), significantly outperforming logistic regression (AUC-ROC = 0.74, $p < 0.001$).

FEATURE IMPORTANCE:

Prior 12-month admission frequency and BUN levels were the strongest predictors.

CHAPTER 5: DISCUSSION & CONCLUSION

INTERPRETATION:

Machine learning classifiers successfully leverage complex EHR feature interactions to stratify 30-day readmission risk.

STUDY LIMITATIONS:

1. Single-center retrospective design.
2. Potential unmeasured residual confounding.

RECOMMENDATIONS & FUTURE WORK:

Prospective clinical trial validation before embedding model predictions into electronic health record dashboards.

INTERACTIVE STUDENT CHECKLIST: CAN YOU LOCATE...?

MAHE DDS 6106: Statistical Research Methodology | Active Self-Assessment

Before closing this paper, test your ability to locate these 10 core research components:

· 1. Research Problem & Gap ·

Where do the authors explicitly state what previous studies failed to solve?

· 2. Research Question & Objectives ·

Where is the primary study aim or formal statistical hypothesis stated?

· 3. Study Design & Target Population ·

Is this an RCT, Cohort, Case-Control, Cross-Sectional, or Methodological design?

· 4. Variables & Primary Outcome ·

What is the primary dependent variable / clinical endpoint?

· 5. Sample Size Calculation ·

Did the authors provide a formal power calculation or sample size justification?

· 6. Statistical Analysis Plan (SAP) & Software ·

Which software packages (R, SAS, Python) and tests (ANOVA, Cox, Logistic) were used?

· 7. Primary Empirical Finding & Effect Size ·

Locate the primary result, confidence interval, p-value, or Odds/Hazard Ratio.

· 8. Study Limitations & Confounding ·

Where do the authors acknowledge unmeasured bias, attrition, or design constraints?

· 9. Ethics Approval & IRB Number ·

Locate the institutional ethics committee protocol approval code.

· 10. Funding & Conflict of Interest ·

Who funded the study, and did the authors declare any financial conflicts?

· Completed this checklist? Excellent work mastering scientific paper navigation!