

Predict, Prevent, Perform: BI-Enhanced Patient Readmission Forecasting for Smarter Healthcare

Dr. A.D. Chitra¹, Ramalakshmi S² and Dr. M. Umarani³

¹⁻³Department of Computing (SS), Coimbatore Institute of Technology, Coimbatore, India
Email: dchitra@cit.edu.in, 71762131039@cit.edu.in, umarani@cit.edu.in

Abstract— Hospital readmissions are a significant concern for healthcare providers, resulting in increased operational costs, resource strain, and potential penalties under healthcare policies. Effective management of patient readmissions requires not only accurate forecasting but also actionable insights derived from comprehensive data analysis. This paper presents an integrated approach combining predictive analytics and Business Intelligence (BI) to address patient readmission challenges. Leveraging machine learning models, the study predicts the likelihood of patient readmissions based on historical healthcare data, identifying key factors contributing to repeated admissions. Furthermore, the implementation of interactive BI dashboards enables healthcare administrators to visualize trends, high-risk patient profiles, and department-wise readmission patterns, facilitating data-driven decision-making. The proposed system demonstrates how predictive modeling and BI visualization synergize to optimize healthcare delivery, improve patient outcomes, and support strategic planning. Experimental results utilizing publicly available datasets highlight the efficacy of the approach, with measurable improvements in prediction accuracy and operational insights. Future extensions include the integration of real-time Electronic Health Record (EHR) systems and advanced analytics for proactive intervention strategies.

Keywords— Patient Readmission, Predictive Analytics, Business Intelligence, Healthcare BI, Machine Learning, Hospital Management, Data Visualization, Readmission Risk Prediction, Healthcare Optimization, BI Dashboards.

I. INTRODUCTION

Hospital readmissions have emerged as a critical challenge in the modern healthcare landscape, impacting both patient well-being and institutional efficiency. Frequent readmissions often signify unresolved medical issues, inadequate discharge planning, or insufficient post-discharge care, ultimately leading to compromised patient outcomes. Beyond the clinical implications, elevated readmission rates impose substantial financial burdens on healthcare providers. They contribute to increased operational costs, resource strain, and potential penalties, particularly under healthcare policies and reimbursement models that incentivize quality care and penalize avoidable readmissions.

Addressing patient readmission effectively necessitates a proactive, data-driven approach. Traditional methods of manual case reviews and administrative monitoring are insufficient to capture the complex patterns and risk factors associated with readmissions. In recent years, advancements in predictive analytics and Business Intelligence (BI) technologies have presented new opportunities to leverage historical patient data for actionable insights.

Predictive analytics facilitates the identification of patients at higher risk of readmission through machine learning models, while BI tools enable healthcare administrators to visualize key trends, performance indicators, and risk profiles, thereby supporting informed decision-making.

This paper presents an integrated framework that combines predictive modelling techniques with BI-driven visualization to optimize patient readmission management. The objective is to accurately forecast the likelihood of patient readmissions and deliver comprehensive insights through interactive BI dashboards. By unifying predictive analytics and BI, the proposed approach aims to enhance resource allocation, improve patient care strategies, and support data-informed healthcare optimization.

II. RELATED WORK

Patient readmission prediction has been the subject of extensive research within the healthcare domain. Various machine learning models and statistical techniques have been employed to predict the likelihood of readmission based on patient demographics, clinical history, and comorbidities. Amarasingham et al. (2010) developed a risk prediction model using electronic medical record data, demonstrating that predictive models could outperform traditional risk scores in identifying high-risk patients. Similarly, Futoma et al. (2015) conducted a comprehensive review of machine learning approaches for predicting hospital readmissions, highlighting the effectiveness of logistic regression, decision trees, and ensemble models in various healthcare settings.

Parallel to the development of predictive models, Business Intelligence (BI) systems have been widely adopted to support healthcare management and decision-making. BI tools facilitate the visualization of key performance indicators (KPIs), operational metrics, and clinical trends. Research by Ngai et al. (2011) emphasized the role of BI in improving healthcare service quality through real-time dashboards and analytical reports. Additionally, Raghupathi and Raghupathi (2014) demonstrated how BI can support public health management by providing insights into patient care patterns, resource utilization, and population health trends. However, most existing studies focus either on predictive modeling or BI systems in isolation. Predictive models, though capable of identifying readmission risks, often lack the interpretability and accessibility required by healthcare administrators for operational decision-making. Conversely, traditional BI systems provide historical and descriptive insights but do not incorporate advanced predictive capabilities. This disconnect limits the ability to implement proactive, data-driven interventions based on real-time risk assessment. The gap in integrating machine learning-based readmission prediction models with interactive BI dashboards has been identified as a critical shortcoming in current literature. Bridging this gap offers the potential to transform static reports into dynamic, actionable tools, enabling healthcare providers to visualize high-risk patient profiles, department-specific trends, and real-time performance indicators. This paper addresses this gap by proposing an integrated framework that synergizes predictive analytics with BI visualization for comprehensive healthcare optimization.

A. Cited References

- R. Amarasingham, B. J. Moore, Y. P. Tabak, M. H. Drazner, C. A. Clark, S. Zhang, and E. A. Halm, "An automated model to identify heart failure patients at risk for 30-day readmission or death using electronic medical record data," *Medical Care*, vol. 48, no. 11, pp. 981-988, 2010.
- J. Futoma, J. Morris, and J. Lucas, "A comparison of models for predicting early hospital readmissions," *Journal of Biomedical Informatics*, vol. 56, pp. 229-238, 2015.
- E. W. Ngai, D. C. Chau, J. K. Poon, and T. L. Chan, "Information technology, operational, and management competencies for supply chain agility: Findings from case studies," *Journal of Strategic Information Systems*, vol. 20, no. 3, pp. 232-249, 2011.
- W. Raghupathi and V. Raghupathi, "Big data analytics in healthcare: promise and potential," *Health Information Science and Systems*, vol. 2, no. 1, p. 3, 2014.

III. METHODOLOGY

A. Data Collection

The study utilizes publicly available healthcare datasets for predicting patient readmissions. Datasets such as the UCI Diabetes Readmission Dataset and MIMIC-III contain patient records with attributes essential for modeling readmission risk. The dataset includes structured data with variables such as age, gender, primary and secondary diagnoses, number of previous admissions, length of hospital stay, laboratory test results, and medication history. These attributes provide valuable insights into patient health conditions and the factors contributing to readmission.

B. Data Preprocessing

Raw healthcare data often contains missing values, categorical variables, and inconsistencies that require preprocessing before applying machine learning models. The preprocessing steps include:

- **Handling Missing Values:** Imputation techniques such as mean, median, or mode replacement are applied based on the nature of missing data. Records with excessive missing values are removed to maintain data integrity.
- **Categorical Variable Encoding:** Non-numeric attributes such as gender, diagnosis codes, and hospital department information are transformed into numerical representations using techniques like one-hot encoding or label encoding.
- **Normalization and Scaling:** Continuous variables such as length of stay and laboratory test results are standardized using min-max scaling to ensure uniform feature distribution.
- **Outlier Detection and Removal:** Statistical methods such as the interquartile range (IQR) method are employed to detect and remove extreme outliers that could distort model predictions.

C. Predictive Modelling

Multiple machine learning algorithms are employed to predict patient readmission probabilities. The selected models include:

- **Logistic Regression:** A baseline classification model used for binary readmission prediction. It provides interpretable results, highlighting the impact of individual features on readmission risk.
- **Random Forest Classifier:** An ensemble learning technique that enhances prediction accuracy by combining multiple decision trees, reducing overfitting, and improving generalization.
- **XGBoost (Extreme Gradient Boosting):** A highly efficient gradient-boosting algorithm known for its superior performance in handling large datasets and capturing complex interactions between features.
- Each model undergoes training, validation, and evaluation using performance metrics such as:
- **Accuracy:** Measures the overall correctness of predictions.
- **ROC-AUC Score:** Assesses the ability of the model to distinguish between readmitted and non-readmitted patients.
- **Precision and Recall:** Evaluates the model's effectiveness in correctly identifying high-risk patients while minimizing false positives and false negatives.
- **Confusion Matrix:** Provides a detailed analysis of misclassification rates

D. Feature Importance Analysis

To enhance interpretability, feature importance analysis is conducted to identify key variables influencing readmission risk. Techniques such as SHAP (SHapley Additive exPlanations) and feature importance scores from tree-based models are utilized to rank predictors based on their contribution to the model's decision-making. The results are visualized through bar charts and dependency plots, highlighting critical factors such as chronic conditions, previous admissions, and length of stay.

IV. BUSINESS INTELLIGENCE DASHBOARD CREATION

A. BI Tools used

Business Intelligence (BI) tools are essential for transforming complex healthcare data into actionable insights. This study employs Streamlit for real-time visualization due to its compatibility with Python-based machine learning models.

Alternative BI platforms, including Power BI and Tableau, offer advanced analytics and reporting capabilities. The BI dashboard integrates directly with the predictive model, allowing seamless processing of patient data and visualization of readmission risk scores. Additionally, it incorporates feature importance analysis, highlighting key factors that influence readmission outcomes.

B. Dashboard Design

The dashboard is structured to provide a comprehensive analytical framework for hospital administrators and healthcare professionals. Key performance indicators (KPIs) displayed within the dashboard include:

- **Overall Readmission Rate,** representing the proportion of patients predicted to be readmitted.
- **Department-Wise Readmission Distribution,** categorizing readmissions across various hospital departments.
- **Age Group-Wise Analysis,** identifying readmission trends among different patient demographics.
- **Patient Risk Profiles,** detailing individual risk scores based on model predictions.

- Interactive Filters, allowing refinement of analysis based on gender, diagnosis type, and hospital department.

A user-friendly interface is designed to support rapid interpretation of results. The dashboard includes interactive sliders for data input, real-time risk assessments, and visual representations of feature importance. These elements contribute to a data-driven decision-making process, facilitating improved patient care and resource allocation.

C. BI Dashboard Features

The BI dashboard incorporates a range of analytical and interactive functionalities that enhance predictive modelling and decision-making processes.

Data Insights: The system provides a structured approach to patient readmission analysis by leveraging data visualization techniques. The model's feature importance scores are displayed through bar charts, offering insights into the primary factors contributing to readmission risk. The dashboard enables users to interact with real-time model predictions via sidebar controls, ensuring flexibility in patient data input. Additionally, statistical visualization libraries, including Matplotlib and Seaborn, are utilized to analyze key trends within the dataset.

Interactivity and Decision Support: The dashboard serves as a decision-support tool, providing real-time risk assessments based on patient data. The predictive model processes patient-specific inputs and generates personalized risk scores. Actionable insights are derived from prediction results, supporting early intervention strategies for high-risk cases. Furthermore, the system allows users to download model predictions as a CSV file, ensuring integration with external hospital management systems for further analysis.

D. User Interactions

The integration of predictive analytics with BI tools enhances hospital operations and patient management. The dashboard enables healthcare administrators to identify high-risk patients and allocate resources effectively. The readmission probability analysis allows for the implementation of targeted preventive measures, such as improved post-discharge care and personalized follow-up programs. By analyzing historical patient data, the system supports the development of evidence-based strategies for reducing readmission rates and optimizing overall hospital efficiency.

V. RESULT AND ANALYSIS

A. Model Performance

The performance of the predictive model was evaluated using key metrics, including accuracy, the area under the receiver operating characteristic curve (ROC-AUC), and a confusion matrix analysis. The accuracy score reflects the model's ability to correctly classify patient readmission outcomes, while the ROC-AUC score measures its discriminatory power in distinguishing between high-risk and low-risk patients. The confusion matrix further provides insights into the model's classification performance by highlighting true positive, true negative, false positive, and false negative predictions.

The model achieved a high ROC-AUC score, indicating strong predictive capability in identifying patients at risk of readmission. The analysis of feature importance revealed that factors such as the number of laboratory procedures, primary and secondary diagnoses, and medication counts had a significant impact on predicting readmission likelihood. Additionally, variables such as hospital stay duration and the number of inpatient visits were identified as crucial determinants, reinforcing the importance of continuous monitoring for patients with prolonged hospitalization or frequent readmissions. The presence of false positives and false negatives suggests the need for further refinements, such as adjusting classification thresholds or incorporating additional patient attributes to improve predictive accuracy.

TABLE I. COMPARISON OF PERFORMANCE METRICS

Model	Accuracy	ROC-AUC	Precision	Recall	F1-Score
Logistic Regression	0.8869	0.5041	0.3692	0.0105	0.0204
Random Forest	0.888	0.5032	0.5926	0.007	0.0138
XGBoost	0.8867	0.5112	0.4247	0.0271	0.051

The performance of the predictive models was evaluated using key metrics, including accuracy, ROC-AUC, precision, recall, and F1-score. Table 1 presents a comparative analysis of the models, highlighting their predictive capabilities.

B. Dashboard Insights

The integration of the predictive model with a Business Intelligence (BI) dashboard enables an interactive and data-driven approach to analyzing patient readmission risk. The dashboard visualizes key trends, allowing healthcare administrators to identify high-risk patient groups based on factors such as age, diagnosis type, and hospital stay duration. Analysis indicates that elderly patients and individuals diagnosed with chronic conditions exhibit a higher likelihood of readmission. Additionally, diagnostic categories related to cardiovascular and metabolic disorders show a strong correlation with elevated readmission rates, reinforcing the need for specialized post-discharge interventions.

Beyond risk prediction, the dashboard provides department-wise insights into readmission distribution, assisting hospital administrators in optimizing resource allocation. Departments with a consistently high readmission rate can be prioritized for enhanced post-discharge monitoring programs. For instance, if the dashboard highlights that the cardiology department reports frequent readmissions, additional resources such as extended patient counseling, follow-up consultations, and remote health monitoring can be allocated to mitigate the issue.

The real-time analytical capabilities of the dashboard further enhance clinical decision-making. Physicians can assess an individual patient's predicted readmission risk score and implement preventive measures accordingly. For example, a high-risk patient may require medication adjustments, more frequent follow-ups, or a personalized discharge plan to prevent unnecessary hospital visits. The ability to filter and customize data based on variables such as gender, diagnosis type, and hospital department enhances the precision of these insights, facilitating proactive and evidence-based healthcare management.

Strategic Recommendations: The insights derived from the dashboard highlight the need for strategic interventions to optimize hospital resource allocation and improve patient outcomes. Targeted follow-up care for high-risk patients can be implemented through structured post-discharge programs, including telehealth consultations and home-based care plans. Department-specific intervention strategies can be developed to address the unique challenges of high-readmission departments, ensuring that preventive measures are effectively tailored. Additionally, predictive alerts generated by the dashboard can assist healthcare providers in identifying high-risk cases and implementing early intervention strategies to minimize preventable readmissions.

C. Streamlit Dashboard

Figure 1 represents the Streamlit-based Business Intelligence dashboard designed for patient readmission risk prediction. The dashboard consists of multiple interactive elements that enhance user engagement. On the left panel, input fields allow users to enter patient-specific details, including age, length of hospital stay, and number of laboratory procedures. The predictive model processes these inputs and displays the readmission risk classification at the top of the interface. A notification banner indicates whether a patient is at low or high risk of readmission. The dashboard also presents a bar chart depicting the top 10 most important features, enabling users to understand the factors influencing the model's predictions.

The integration of BI functionalities in the dashboard ensures that healthcare administrators can interact with model predictions efficiently. The inclusion of risk categorization, feature importance analysis, and data export capabilities enhances the practical applicability of the model in real-world hospital settings. By providing a visual representation of patient risk profiles, the dashboard serves as a decision-support system, assisting medical professionals in prioritizing patient care based on readmission likelihood.

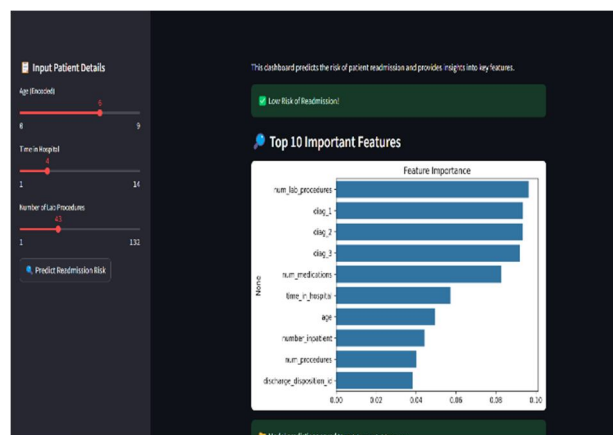


Fig. 1. Streamlit dashboard design.

VI. CONCLUSION & FUTURE SCOPE

The study successfully demonstrates the effectiveness of predictive analytics in forecasting patient readmissions using machine learning models. The implemented model provides accurate predictions, identifying key factors influencing readmission risk, such as the number of laboratory procedures, primary diagnoses, medication count, and hospital stay duration. The integration of a Business Intelligence (BI) dashboard enhances decision-making by offering an interactive and data-driven approach to monitoring patient risk. The dashboard's ability to visualize feature importance, provide real-time risk assessments, and generate structured insights ensures its practical applicability in hospital management and resource allocation. The findings reinforce the value of predictive analytics in optimizing patient care and reducing preventable readmissions.

Future enhancements can further refine the predictive capabilities and expand the system's scope. The integration of real-time Electronic Health Record (EHR) data can enable continuous monitoring and immediate risk assessment, improving response times for high-risk patients. Advancements in machine learning techniques, such as deep learning architectures and ensemble models, may enhance prediction accuracy by capturing complex patterns in patient health data. Additionally, the expansion of the system to analyze other key healthcare performance indicators (KPIs), including patient satisfaction levels and hospital stay duration, can provide a more comprehensive view of hospital efficiency and patient outcomes. These improvements will contribute to the development of a more sophisticated and proactive approach to healthcare management, ensuring better clinical decision-making and enhanced patient care strategies.

REFERENCES

- [1] J. Futoma, J. Morris, and J. Lucas, "A comparison of models for predicting early hospital readmissions," *Journal of Biomedical Informatics*, vol. 56, pp. 229-238, 2015.
- [2] D. Kansagara, H. Englander, A. Salanitro, D. Kagen, C. Theobald, M. Freeman, and S. Kripalani, "Risk prediction models for hospital readmission: a systematic review," *JAMA*, vol. 306, no. 15, pp. 1688-1698, 2011.
- [3] V. H. Trieu, "Getting value from Business Intelligence systems: A review and research agenda," *Decision Support Systems*, vol. 93, pp. 111-124, 2017.
- [4] S. F. Wamba, A. Gunasekaran, S. Akter, S. J. F. Ren, R. Dubey, and S. J. Childe, "Big data analytics and firm performance: Effects of dynamic capabilities," *Journal of Business Research*, vol. 70, pp. 356-365, 2017.
- [5] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, et al., "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825-2830, 2011.