International Journal of Creative and Open Research in Engineering and Management
ISSN: 3108-1754 (Online)
Volume 02 Issue 11 November-2026 | Impact Factor: 3.5



Healthcare Risk Analytics: An Interactive Dashboard and Machine Learning Model for Patient Risk Prediction

Paras Sharma
Master of Computer Applications (M.C.A)
Jagan Institute of Management Studies (JIMS), Rohini, New Delhi

Paarth Kumar
Master of Computer Applications (M.C.A)
Jagan Institute of Management Studies (JIMS), Rohini, New Delhi

Lakshay
Master of Computer Applications (M.C.A)
Jagan Institute of Management Studies (JIMS), Rohini, New Delhi

How to Cite this Article:

Sharma, P., Kumar, P. & Lakshay, (2026). Healthcare Risk Analytics: An Interactive Dashboard and Machine Learning Model for Patient Risk Prediction. International Journal of Creative and Open Research in Engineering and Management, 02(11), 1-9. https://doi.org/10.55041/ijcope.v2i11.115

License:

This article is published under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

© The Author(s). Published by International Journal of Creative and Open Research in Engineering and Management.







https://doi.org/10.55041/ijcope.v2i11.115

ABSTRACT

A huge amount of information about patients is collected in hospitals, however, the data is stored separately in tables, and the risk assessment is done manually. In this paper we introduce the Healthcare Risk Analytics – a light-weight web-application developed with Python and Streamlit that combines data about patients, diagnoses, outcomes, and laboratories results in a single interactive dashboard and also implements a supervised machine learning classifier to determine whether a patient is High Risk or Low Risk based on his/her age, length of stay, treatment cost, and abnormal laboratory count with probability output. The dashboard includes filter options, KPI, cost and length of stay analysis per diagnoses, distribution of laboratory results, patient table for search purposes, and a page with predictions. The system was evaluated through such metrics as accuracy, precision, recall, F1-score and also tested on a set of functional test cases. The paper demonstrates that a small team of a hospital can easily analyze its records and assess the risk for each patient using free tools and a small dataset.

KEYWORDS

Healthcare Analytics, Predicting Patient Risk, Machine Learning, Streamlit, Dashboard, Power BI.

INTRODUCTION

Daily, the hospital documents the patient's identity, the diagnosis made, the types of tests conducted, the patient's length of stay, the cost of treatment, and the outcome of the case. This data is usually scattered throughout different databases, making it difficult to obtain a full view of the situation. The management wants to know where funds are being allocated and which diagnoses require additional bed space. The physicians are also interested in which patients require more attention.

Machine learning can be used to answer the second problem by identifying patterns in previous data and predicting the risk for a new patient. Extensive research conducted on EHRs has found that machine learning models were able to distinguish between high-risk and low-risk patients more effectively than most existing risk scores [1]–[3]. Yet, all of these approaches have been using huge amounts of data and complex machine learning models and have not provided a simple solution for an average small group.

The following paper provides an insight into Healthcare Risk Analytics, created as a Minor Project in the MCA program of Jagan Institute of Management Studies, Delhi. The project involves exploratory data analysis, reports generation, and risk prediction through a single Streamlit dashboard with Power BI report.

LITERATURE REVIEW

Machine Learning versus Traditional Risk Scores

In comparing four different machine learning methods with the ACC/AHA guideline algorithm using routine data of 378,256 UK patients, Weng et al. [3] showed that all of the four algorithms performed better, with the best one giving an AUC value of 0.764 compared to 0.728. In a study by Alaa et al. [6], a similar conclusion was drawn after the comparison of the performance of automatic machine learning algorithm ensemble with the Framingham score, where the former gave an AUC value of 0.774 compared to 0.724 using the Framingham score on 423,604 UK Biobank participants. In predicting hospital ward clinical deterioration, Churpek et al. [2] revealed that random forest



and boosted trees performed with an AUC value of 0.80 against 0.70 of the MEWS early warning score.

Readmission and Large-Scale Record Models

According to Kansagara et al. [4], there are 26 admission models, but none of them is capable of effectively distinguishing between high-risk and low-risk patients. On the other hand, according to Mortazavi et al. [5], random forests, boosting, and hybrid techniques performed better in predicting readmission after heart failure compared to logistic regression. Furthermore, Rajkomar et al. [1] implemented deep learning in the comprehensive data of 216,221 adults of two hospitals and achieved an AUROC of 0.93–0.94 for in-hospital mortality prediction.

Research Gap

However, all the studies cited above are oriented towards predicting the accuracy and are performed using large data sets consisting of many variables. None of these researches give an idea about a simple interface where the patient's data, costs and outcomes summary, as well as risk predictions, can be presented altogether on a smaller data set.

RESEARCH OBJECTIVES

1. For cleaning and consolidating patient records (patient names, diagnosis, outcomes and lab results) from hospitals.

2. For constructing the machine learning classifier that will categorize the patient as either High-Risk or Low-Risk.

3. A dashboard with filters, KPIs, and charts to analyze cost, stay, and outcome would be required.

4. In order to offer a prediction page that gives the risk and probability of one patient.

5. In order to examine the system and evaluate it using traditional metrics.

METHODOLOGY

Data

The database has two tables. The first table named 'patients' has PatientID, Name, Age, Gender, DiagnosisName, OutcomeName, LengthOfStay, TreatmentCost and RiskLabel. Another table named 'labs' has PatientID, TestName, and Result, which has relation with patients through PatientID. The sample database has around 200 patient data records. To conduct test on the dashboard and the pipeline on a larger data set, an additional 500 records for 1,500 lab data records have been created using scripts. These are not actual patients.

Preprocessing and Features

The column names have been standardised ("Length of stay" was changed to LengthOfStay), all numerical columns have been transformed with any invalid values made blank, and the blank textual values have been replaced by "Unknown". AbnormalLabCount is a feature engineered to count the number of abnormal laboratory values per each individual. Table 1 describes the input variables used in the models.

Table 1: Model input features

Feature	Type	Description
Age	Numeric	Patient age in years
LengthOfStay	Numeric	Days in hospital
TreatmentCost	Numeric	Total cost in rupees
AbnormalLabCount	Numeric	Lab results outside the normal range

Model

Risk Model: This is a supervised classification model that uses four features to classify patients in the two classes (Low risk and High risk). The algorithm used in this model is Logistic Regression. The training set has been divided into 80% training data and 20% testing data. The risk model is then saved as Risk_model1.joblib using joblib. For prediction of a new patient, the dashboard creates a table having one row and the same columns as the training data.

System Architecture

The app consists of one single file written in Python. Pandas processes and cleans CSV files while scikit-learn and joblib deliver the model. The visualizations in the app are created using Plotly while Streamlit creates the web app itself with custom CSS. There is a sidebar with four filters with multiselect option and age slider. One filtered table delivers all the KPIs, visualizations and tables that update everything in unison. The app has four tabs: Overview, Diagnosis & Labs, Patient Explorer and Predict Risk. Users can also upload their own patient and labs CSV files. Power BI report with the same measures was created for reporting purposes with relationship between PatientID and DAX measures like High Risk % and Recovery Rate.

RESULT & DISCUSSION

Dashboard Output

The dashboard includes five KPI cards (patient, high risk patient, average age, treatment cost and recovery patient). The Overview tab is represented by the risk-level pie chart, outcome pie chart, treatment cost by diagnosis, average age by risk level histogram and cost-stay scatter plot. The Diagnosis and Labs tab contains patients' information, average treatment cost and average stay by diagnosis. Also there are histograms of laboratory results. The prediction tab is made of the red/green result and gauge chart divided into three categories (0-35, 35-65 and 65-100).

Model Evaluation

The performance of the model was assessed using the test portion of the data. Recall is the most critical metric to use in this instance, as a false negative can be very detrimental.

Table 2: Performance of the risk model on test data

Measure	Value
Accuracy	73%
Precision	50%
Recall	56%
F1-score	0.53



Testing

14 test cases were used to test the functionality of startup, missing files, filters, age slider, patients' search, CSV export, prediction page with normal, high, and zero KPI, as well as all four tabs. Issues identified and solved while developing included hard-to-read labels for the KPI due to the conflict of browser dark mode with custom CSS, KPI value truncation, and unwanted text cursor on drop-down menus.

Discussion and Limitations

The dashboard provides one location to visualize cost, stay, outcome, and risk, which have not been described in the literature with regard to small dataset. The main drawbacks of the pipeline include the small dataset size, usage of four features only, and lack of clinical validation of the predictive model. Due to the presence of the synthetic data in the test dataset, the performance results obtained with the help of this data prove the validity of the pipeline but not the model.

CONCLUSION AND FUTURE SCOPE

The Healthcare Risk Analytics dashboard was introduced in this paper to demonstrate the use of streamlit as a single platform for patient data, analysis, and prediction through machine learning models. The demonstration revealed that exploration, reporting, and risk prediction can be provided together using open-source software and small datasets.

Future studies can train the algorithm on a larger real-world dataset with more features like vital signs and medications, unlike others that use cross-validation, and can incorporate model interpretability techniques like SHAP. Additionally, it can be put online with users' login and role-based access and can connect directly to the hospital's database. The privacy of patients must be ensured through anonymization of their information and non-public release of any actual data.

REFERENCES

[1] A. Rajkomar et al., "Scalable and accurate deep learning with electronic health records," npj Digital Medicine, 2018, doi: 10.1038/s41746-018-0029-1.

[2] M. M. Churpek, T. C. Yuen, C. Winslow, D. O. Meltzer, M. W. Kattan, and D. P. Edelson, "Multicenter comparison of machine learning methods and conventional regression for predicting clinical deterioration on the wards," Critical Care Medicine, vol. 44, no. 2, pp. 368–374, 2016.

[3] S. F. Weng, J. Reps, J. Kai, J. M. Garibaldi, and N. Qureshi, "Can machine-learning improve cardiovascular risk prediction using routine clinical data?," PLOS ONE, vol. 12, no. 4, e0174944, 2017.

[4] D. Kansagara et al., "Risk prediction models for hospital readmission: a systematic review," JAMA, 2011.

[5] B. J. Mortazavi et al., "Analysis of machine learning techniques for heart failure readmissions," Circulation: Cardiovascular Quality and Outcomes, 2016.

[6] A. M. Alaa, T. Bolton, E. Di Angelantonio, J. H. F. Rudd, and M. van der Schaar, "Cardiovascular disease risk prediction using automated machine learning: a prospective study of 423,604 UK Biobank participants," PLOS ONE, vol. 14, no. 5, 2019.

[7] T. Desautels et al., "Prediction of sepsis in the intensive care unit with minimal electronic health record data: a machine learning approach," JMIR Medical Informatics, vol. 4, no. 3, e28, 2016.

[8] B. Shickel, P. J. Tighe, A. Bihorac, and P. Rashidi, "Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis," IEEE Journal of Biomedical and Health Informatics, 2018, doi: 10.1109/JBHI.2017.2767063.

[9] Streamlit documentation, https://docs.streamlit.io

[10] scikit-learn documentation, https://scikit-learn.org