Reducing Unplanned Hospital Readmissions

94-881: Managing Analytics Projects (Fall 2023)
Final Project Report

Aditi Gupta
Dollaya Hirunyasiri
Greta Luo
Jorge Palacio

October 15, 2023

Table of Contents

Executive Summary

Unplanned readmitted diabetic patients within 30 days of their initial discharge create a financial burden to the US Health Care system, increase the health risks for patients, and negatively impact their experience. Using a public data set with 10 years of admission data for ~72,000 patients diagnosed with diabetes, we completed an initial analytics exercise with the following steps:

- Characterized the problem and identified its key features
- Performed an Exploratory Data Analysis (EDA) to identify and understand the data source's attributes
- Compared the performance of two widely used classification algorithms
- Refined the model behavior balancing the training data and achieving accuracy and precision close to 60%

For the next phase of the project, we propose a new iteration that refines the requirements and the proof of concept model, develop and test the new model's version, and deploy it to production process that integrates with the patients' journeys to support the health workers decision making.

Background

Unplanned hospital readmissions cause a significant financial obligation to the healthcare system in the United States. In 2011, American hospitals spent an estimated $41 billion on diabetic patients who were readmitted within 30 days of their initial discharge. According to medical studies, nearly 27% of these readmissions can be prevented through various measures, including improved communication among healthcare providers, improved decision-making regarding patients' readiness for discharge, enhanced disease monitoring, and better support for patients in managing their conditions [1]. Addressing these factors would help to improve the overall quality of care within the healthcare system and reduce the financial burden of preventable admissions.

Problem Framing

Hypothesis

The hypothesis is that the strongest predictors that contribute to hospital readmission of diabetic patients will be identified. These factors will help to improve the decision-making process in the hospital's day-to-day operations.

Decision to be improved

The decision that is being improved is whether a diabetic patient should be discharged from the hospital. This project aims to support the discharge definition using analytical tools. Specifically, we intend to answer what factors are the strongest predictors of hospital readmission in diabetic patients.

Decision maker

The actors who decide are the health services personnel responsible for discharging patients from hospitals, usually a team of healthcare professionals, including doctors, nurses, and other medical staff. These workers have to define the patients' readiness for discharge using the available information: diagnoses, procedures, medications, and clinical history.

Value of improved decision

The value of improving includes important qualitative and quantitative outcomes.

- The quantitative outcomes include a reduced diabetes patient readmission rate, translating into reduced health spending for readmitted patients. It's estimated that around 25 billion dollars are spent every year due to diabetes readmissions in the US [2]. Another quantitative result is the reduction in short-term mortality of diabetes patients after being initially admitted to hospitals because of the appropriate continuity of the treatment.
- The qualitative outcomes include a better treatment experience for diabetes patients with a reduced likelihood of readmission.

Initial choices for Phase 1

Data Sources

Dataset

The available information is a dataset representing 10 years (1999-2008) of clinical care for patients diagnosed with diabetes at 130 US hospitals and integrated delivery networks.

- 'Diabetic_data.csv' (18.2MB) with 101,766 data points [3]
    - Records of patients diagnosed with diabetes who underwent laboratory, medications, and stayed up to 14 days
    - Shared by BioMed Research International - a peer-reviewed open access scientific journal covering all aspects of biomedical sciences. BioMed's CiteScore is greater than 5.000 and is positioned in the 100-200 position rank of scientific journals, which qualifies as a reputational publication
    - Each record is an inpatient encounter (a hospital admission)
    - Considers any kind of diabetes as the diagnosis
    - Contains information about the age, gender, and race of the patients
    - This dataset is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license
    - Includes over 50 features representing patient and hospital outcomes - e.g., physician specialty, tests, treatments, medications, emergency visits the prior year
- 'IDS_mapping.csv' labels complement for 'Diabetic_data.csv' - Labels for 3 data columns

from diabetic data source

Other datasources such as medical institution or hospital characteristics, health professionals information, calendar events, and other patient's personal information were evaluated but not considered due to privacy concerns or the impossibility of joining the primary dataset because the lack of linking the datasets.

When handling medical data, there are several important considerations to ensure the privacy, security, and ethical use of this sensitive information:

- **Protecting the privacy** of patients is paramount. Medical data should be de-identified or anonymized whenever possible to prevent the identification of individual patients. Obtaining informed consent from patients before collecting or sharing their medical data is crucial. They should understand how their data will be used and can opt out if they choose.
- Researchers should ensure **compliance with healthcare regulations and laws**, such as the Health Insurance Portability and Accountability Act (HIPAA) in the United States. These regulatory frameworks and regulations not only consider privacy but also the security practices required to handle sensitive information such as medical data, including anonymized, but especially the data that could potentially identify individuals. Medical data breaches could generate significant emotional distress and harm with big consequences for individuals. Fraud, reputational damage, and terrorism are examples of bad behaviors that could take advantage of medical data.
- Ensuring **the accuracy and reliability** of medical data is important to avoid medical errors, misdiagnoses, incorrect treatments, and adverse patient outcomes. Inaccurate or tampered data can compromise the integrity of research findings, potentially leading to incorrect conclusions and misleading medical practices. Maintaining the trustworthiness of medical data is essential not only for the advancement of medical knowledge but also for ensuring the safety and well-being of patients who may benefit from evidence-based treatments and interventions.
- A less common consideration, but also important, is to define clear **policies for data retention and deletion.** Dispose of medical data in a secure and compliant manner when it is no longer needed should be planned in any related analytical project.
    - As per the data disposal policy, after analyzing the dataset, the raw files should remain in a private and secure repository. In this case, a Google Drive folder with public access deactivated is an appropriate tool.

Considering the characteristics of the data set, the project team followed the next guidelines and premises:

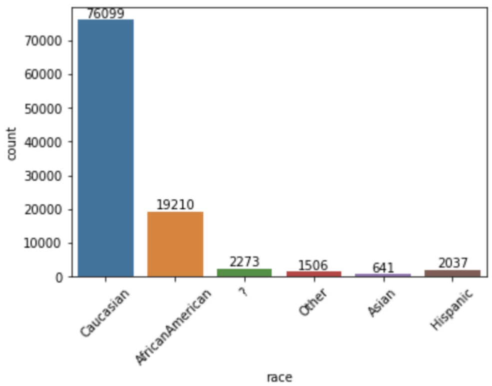
- **The granularity or precision of the available data** is sufficient to test the defined hypothesis - inpatient encounter/hospital admissions is granular enough
- As the dataset contains sensible information about gender, age, race and medical diagnosis, the data source should be **handled with security and confidentiality** - even if it's a public dataset.
- For the academic purposes of this exercise, and considering the difficulty of accessing the details for the data-gathering process, we **assumed that the medical data is accurate,**

reliable, and gathered using the appropriate privacy practices and user consent

- The data timeframe is from 1999 to 2008. For the academic purposes of this research, we assumed data between 15 and 23 years old **is still relevant** for the diabetes patient's hospital readmission problems. Still, when situated within the framework of other research, it is incumbent upon us to carry out more evaluations to establish the relevance of this data
- The license allows for the sharing and adapting of the datasets for any purpose, provided that the appropriate credit is given. Thus, **this data could be used for academic exercises such as this research.**
- As the dataset was donated in 2014, there are **no costs** associated with using the raw data in the research and it can be used timely.
- As per the data disposal policy, after analyzing the dataset, the raw files **should remain in a private and secure repository.** In this case, a Google Drive folder with public access deactivated is an appropriate tool.

Data Preparation and Initial Exploratory Data Analysis

- There are some missing values in the form of (?).
  - race has 2273 occurrences of '?'
  - weight has 98569 occurrences of '?'
  - payer_code has 40256 occurrences of '?'
  - medical_specialty has 49949 occurrences of '?'
  - diag_1 has 21 occurrences of '?'
  - diag_2 has 358 occurrences of '?'
  - diag_3 has 1423 occurrences of '?'
- Basic information from EDA
  - 50 variables
  - ten years (1999-2008) of clinical care at 130 US hospitals
  - 13 are integer type, 34 are categorical types, and 3 boolean type columns
  - 71518 unique patients
- Categorical data
  - Race: overwhelming majority of patients being Caucasian.
  - Gender: slightly larger number of female patients compared to male patients.
  - Age: The age group with the greatest percentage of patients falls within the range of 70 to 80 years old. There is a small percentage of patients under 20.
  - Readmission: The majority of patients had no record of readmission. There was a significantly higher number of patients (~24,000) readmitted in more than 30 days than less than 30 days.
- Numeric Variables
  - Hospital time: most patients stay in the hospital around 3-4 days.
  - Number of lab procedures: most patients receive around 40 lab procedures.
  - Number of medications: most patients receive about 10-15 medications.
- Emergency has some outliers but other variables don't have significant outliers.



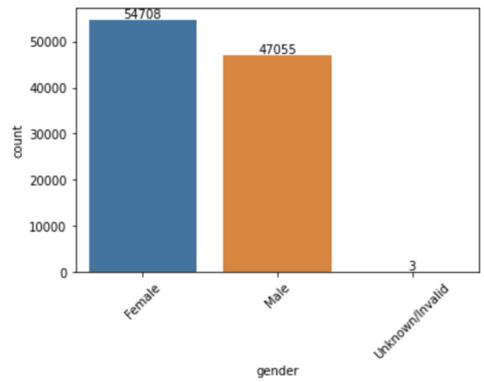
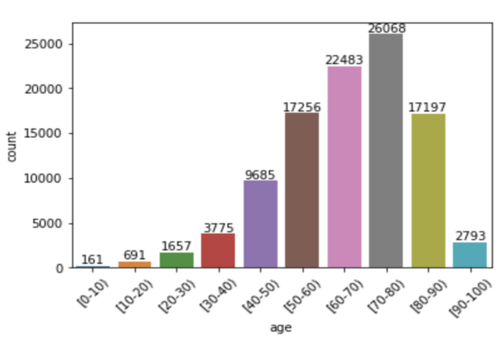


Figure 1: Counts for Race

Figure 2: Counts for Gender

Figure 1 displays counts for race with the overwhelming majority of patients being Caucasian.

Figure 2 displays counts for gender. There is a slightly larger amount of female patients compared to male patients. There are 3 genders that were encoded as invalid that we will further investigate.



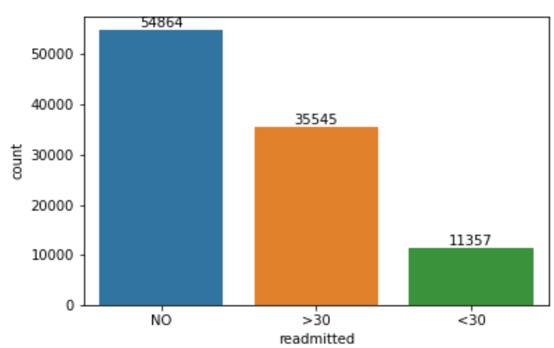


Figure 3: Counts for Age

Figure 4: Counts for Readmission

Figure 3 displays counts for age. The age group with the greatest percentage of patients falls within the range of 70 to 80 years old. There is a small percentage of patients under 20.

Figure 4 displays the counts for readmission. The majority of patients had readmission marked as "No", meaning that there was no record of readmission. There was a significantly higher number of patients (~24,000) readmitted in more than 30 days than in less than 30 days.

Numeric Variables

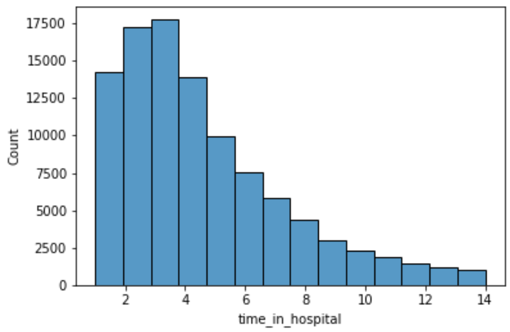


Figure 5: Histogram of Patient Time in Hospital (Days)

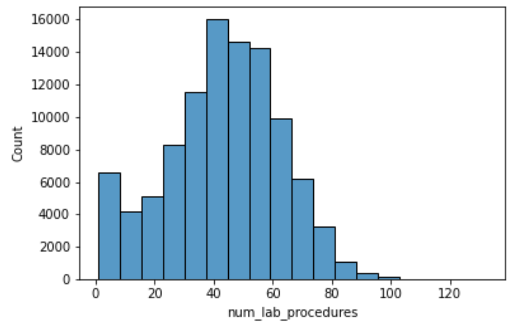


Figure 6: Histogram of Number of Lab Procedures

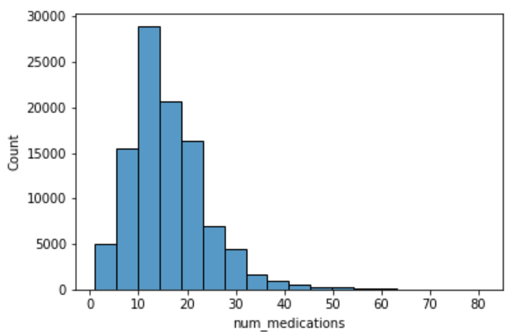


Figure 7: Histogram of Number of Medications

Figure 5 displays a histogram of hospital time, number of days between admission and discharge. This histogram is skewed right with a peak around 3-4, meaning that most patients stay in the hospital around 3-4 days.

Figure 6 displays a histogram of the number of lab procedures. The histogram does not display a normal distribution as there's a significant count approximately below 10 lab procedures as seen in the left-most bar. The peak of the histogram is around 40, meaning that most patients receive around 40 lab procedures.

Figure 7 displays a histogram of the number of medications. The histogram is not heavily skewed. The peak is at around 10-15, meaning that most patients receive about 10-15 medications.

Analytics

Our goal is to determine the most influential factors predicting hospital readmission for diabetic patients. The appropriate analytics approach for this goal is **classification**, as we aim to estimate the probability that a diabetic patient will be readmitted to the hospital within 30 days.

The problem is better characterized as a Classification one because of several reasons.

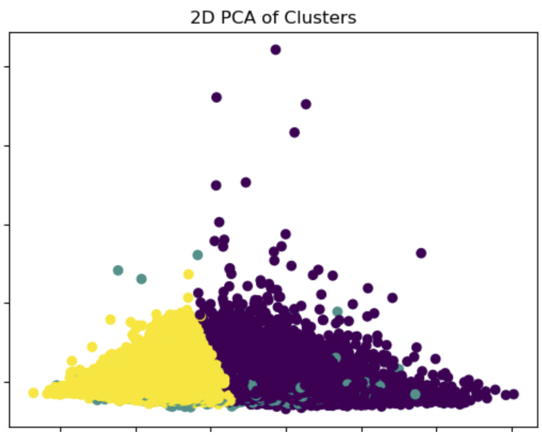
- **Binary Outcome:** The problem is inherently binary in nature: a patient is either readmitted within 30 days (class 1) or not (class 0). Classification algorithms are specifically designed for predicting distinct classes.
- **Probability Estimates**: Classification models like logistic regression provide not just a binary outcome but also the probability of an event occurring. This can help hospitals in risk stratification. For example, patients with a 90% probability of readmission might need different interventions compared to those with a 60% probability.
- **Performance Metrics**: Classification tasks offer a variety of performance metrics such as accuracy, precision, recall, F1-score, and AUC-ROC, among others. This allows healthcare institutions to choose a metric that aligns best with their goals (e.g., maximizing true positives while minimizing false negatives).
- **Imbalance Handling**: In our situation, the number of patients readmitted within 30 days is much lower than those who aren't. Many classification algorithms have techniques to handle class imbalances, such as oversampling, undersampling, or using anomaly detection methods.

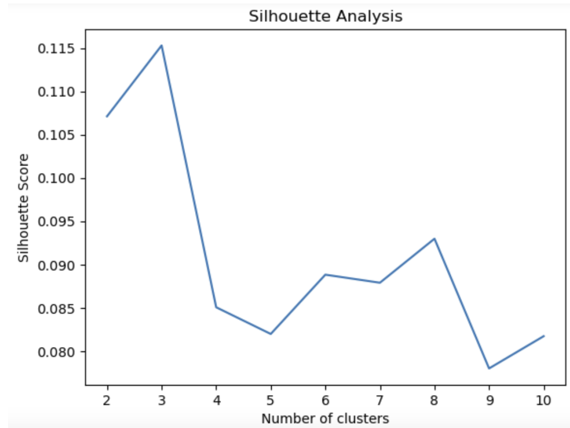
Data Preparation

1. Profiling the data to understand the data set we have.
2. Dropping columns that are not relevant for analysis.
3. Checking for outliers and dropping the values that are out of range.
4. Dropping duplicates from the data set.
5. Replacing null values and "?" with appropriate values.
6. Dividing the whole data set into two groups. The two labels used as dependent variable are:
    a. Readmitted within <30 days
    b. No readmission or readmitted within >30 days
7. Converting categorical values into numerical data using one-hot encoding or dummy variables.
8. Replacing age ranges with mean of the range for each column.
9. Standardizing the data for model preparation using StandardScaler().

Data Clustering

Clustering is used for customer/patient segmentation to understand the specifications of each segment. It helps in identifying unique features/ characteristics of each segment. For deciding on number of clusters we first used K-mean clustering which wasn't able to give good result. The elbow method was not useful in K-mean clustering as there was no clear elbow. So we shifted to Silhouette Method. This method was able to give a clear elbow at K=3. So we formed three clusters.





Data Modeling

There were some models we considered for classification:

- Binary logistic regression as classification doesn't completely satisfy the problem because most of the variables are categorical
  - ✔️Works well to predict the probability that the observation matches a category. In this case, the category is being readmitted
  - ❌The input variables are numeric or mostly numerical
  - ✔️Allows to understand the relative impact of features

- Decision trees satisfy the constraints of the problem. However, this technique doesn't have the best accuracy of the viable alternatives
  - ✔️used to find meaningful subgroups of sample related optimally to a dependent variable – readmitted is the dependent variable
  - ✔️When many types of input variables are present – we have many numeric and a few categorical
  - ✔️When large data sets are available to cover the space – more than 100K data points
  - ✔️Where relationships are not linear - categorical variables are not linear
  - ✔️When algorithms can be run multiple times on subsets of data to obtain robust results - a data snapshot is available and is adequate for multiple runs
  - ✔️Easy to interpret - the resulting three has a clear set of factors
  - ❌Accuracy depends on the training data and can be less accurate than other techniques
  - ✔️Processing times suitable for real time applications

- Random forest produces good results with better accuracy, but it's difficult to interpret
  - ✔️Used to find meaningful subgroups of sample related optimally to a dependent variable – readmitted is the dependent variable
  - ✔️When many types of input variables are present – We have many numeric and a few categorical
  - ✔️When large data sets are available to cover the space – more than 100K data points
  - ❌Difficult to interpret as the combination of random trees usually produces a

complex logic

- ✔️Processing times aren't suitable for real-time applications, but they can provide classification predictions in human-readable time

We decided to choose the *decision tree and random forest* models since these are non-linear models that can capture more complex relationships.

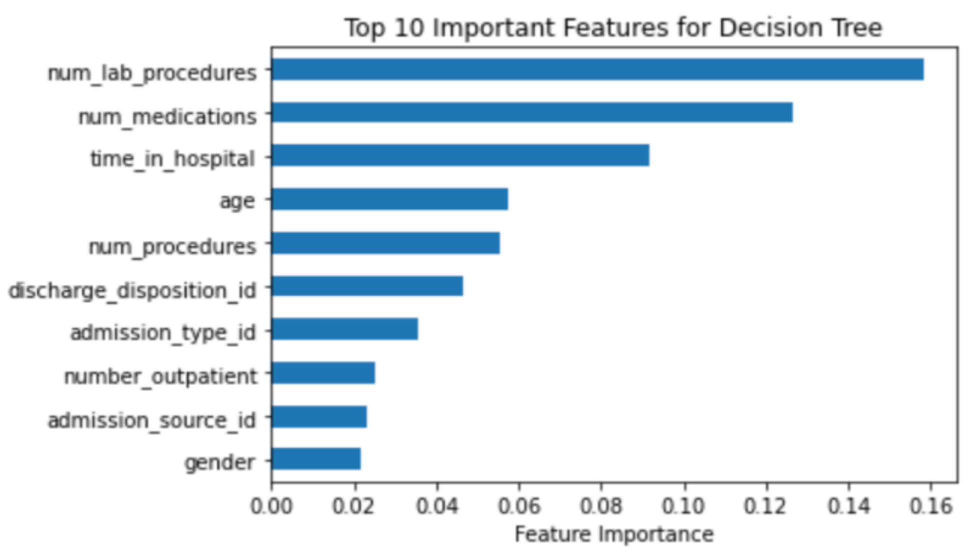
After separating our dataset into 70% train and 30% validation, we trained and tested our models using Python with scikit-learn.

We provide our results in an understandable manner using visualizations in the next section.

Visualizations and Results

We will then evaluate our classification performance (accuracy, precision, recall, F-1 score). Accuracy is the fraction of time the classifier is correct. Precision is the fraction of the time positive indications are correct (minimizing false positives). Recall is the fraction of the time items in the class are detected (minimizing false negatives). The F-1 score combines the precision and recall scores. Moreover, from our model, we can obtain feature weights. These feature weights help assess the significance of each feature in predicting the outcome variable.

If the value of the feature weight is larger, this indicates that the feature is more important in predicting the outcome variable in our model. A positive weight indicates that as the value of the feature increases, the probability of our target variable being 1 (patient being readmitted) increases. A negative weight indicates that the probability of our target variable being 0 (patient not being readmitted) increases.



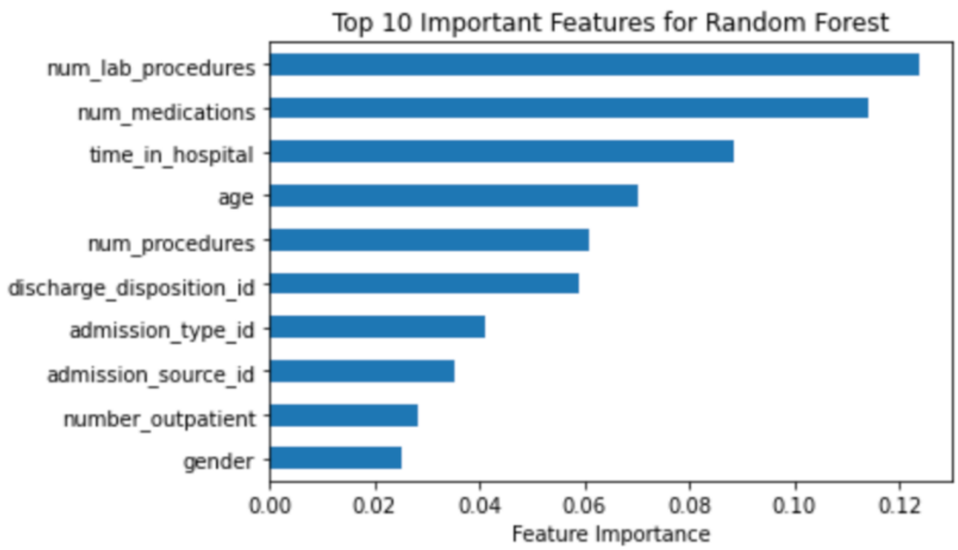
Accuracy	54.51%
Precision	55.02%
Recall	55.88%
F-1 Score	55.45%

Decision Tree Metrics

	Predicted Label = 0 (patient not readmitted < 30 days)	Predicted Label = 1 (patient readmitted < 30 days)
Label = 0 (patient not readmitted < 30 days)	1786	1577
Label = 1 (patient readmitted < 30 days)	1523	1929

Decision Tree Confusion Matrix

Decision tree had metric scores around 55% for accuracy, precision, recall, and f-1 score obtained from the confusion matrix. The top 5 important features for the decision tree model included number of lab procedures, number of medications, time in hospital, age, and number of procedures.



Accuracy	59.18%
Precision	60.16%
Recall	57.44%
F-1 Score	58.77%

Random Forest Metrics

	Predicted Label = 0 (patient not readmitted < 30 days)	Predicted Label = 1 (patient readmitted < 30 days)
Label = 0 (patient not readmitted < 30 days)	2050	1313
Label = 1 (patient readmitted < 30 days)	1469	1983

Random Forest Confusion Matrix

Random forest had slightly higher metric scores around 60% for accuracy, precision, recall, and f-1 score obtained from the confusion matrix. The top 5 important features for the random forest model are the same as those from the decision tree, including number of lab procedures, number of medications, time in hospital, age, and number of procedures. The decision tree and random forest models shared the same top 10 features with the exception of random forest weighing the admission source of the patient higher than number of outpatient visits.

Recommendations

Lessons Learned

We learned that our data set needs to be balanced properly. Prior to addressing this issue, we achieved accuracies of up to 90% with our models but precision and recall rates were only 1%. Our models were doing very poorly at predicting true positives since there were many more negative labels than positive, meaning that the models tended to predict negative more often than not. After balancing, our models' accuracies were lower but the rest of the scores were much higher.

Recommendations for next phase

Random Forest would be the slightly better model to choose with metric scores around 60%. Since these scores are quite low, we strive to enhance the quality of our data and refine the model's parameters to strive for improved performance.

Project Plan Phase 2

Scope

The project plan for Phase 1 consisted of two key steps: Step 1, which involves finalizing requirements, and Step 2, which focuses on developing a Proof of Concept analytics model. After establishing the accuracy and reliability of this model, our next milestone is Phase 2.

In Phase 2, our primary goal is to **transition from the Proof of Concept stage (phase 1) to constructing a fully functional analytics model for production use.** This phase will encompass various activities, including design (step 3), development (step 4), testing (step 5), and, ultimately, the delivery of the data analytics project (step 6). The projected timeline for Phase 2 is four months.
We have created two visual representations: **the first one is the analytics service blueprint, and the second one is the timeline of the analytics project.**

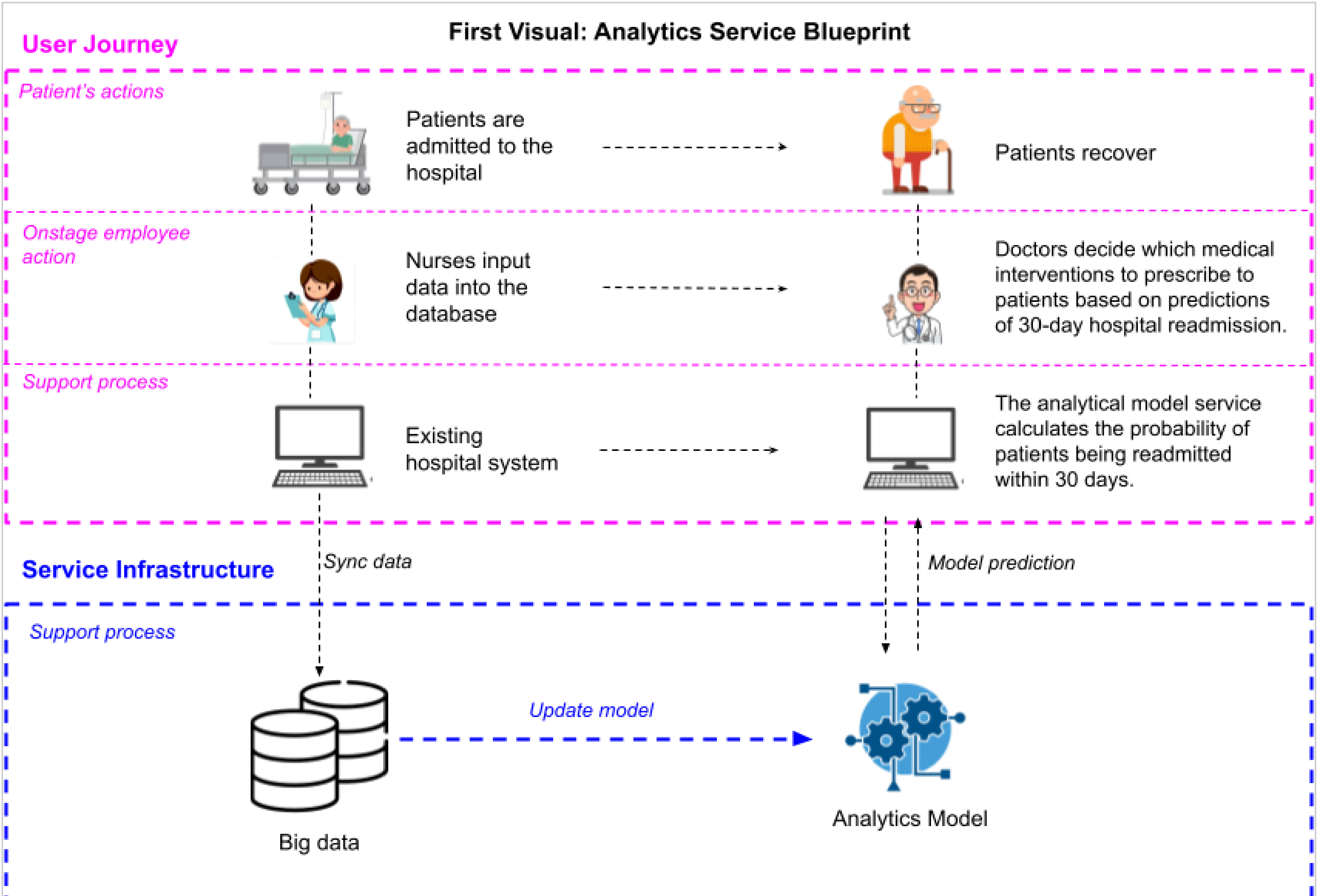
First Visual: Analytics Service Blueprint

- **Objective:** This visualization outlines the functioning of the analytics service within a hospital setting for diabetic patients.
- **Components:**
    - **User Journey:**

        - Patients are admitted to the hospital, subsequently discharged, and given recommendations based on the doctor's advice.
        - Nurses input patient data into the system and monitor patients based on doctors' decisions.
        - Doctors rely on our analytics service to assist in making medical interventions.
    - **Service Infrastructure:** This component includes the predictive analytics model and data storage for maintaining different versions of analytics.

Second Visual: Timeline of the Analytic Project

- **Objective:** This visual represents the timeline for Phase 2 of the analytics project.
- **Four Main Steps:** (Note that Step 1 & Step 2 are for the past first phase)
    - **Step 3 - Design:** This step involves planning where to store the analytics model, how data will flow between different components, and how the model will seamlessly integrate into the user workflow.
    - **Step 4 - Develop:** This step focuses on constructing the system based on the design created in step 3.
    - **Step 5 - User Acceptance Testing:** This step ensures that the service aligns with the subject matter's intentions and expectations.
    - **Step 6 - Deliver:** This final step entails testing in a production environment and delivering the completed project.


First Visual: Analytics Service Blueprint
User Journey
Patient's actions
Patients are admitted to the hospital
Patients recover
Onstage employee action
Nurses input data into the database
Doctors decide which medical interventions to prescribe to patients based on predictions of 30-day hospital readmission.
Support process
Existing hospital system
The analytical model service calculates the probability of patients being readmitted within 30 days.
Sync data
Model prediction
Service Infrastructure
Support process
Update model
Big data
Analytics Model

Second Visual: Timeline of the Analytic Project (Phase 1: Step 1-2 || Phase 2: Step 3-6)																									
Milestones (including Risk Mitigation Plans)	**Deliverable**	**Month 1**				**Month 2**				**Month 3**				**Month 4**				**Month 5**				**Month 6**			
		W1	**W2**	**W3**	**W4**	**W1**	**W2**	**W3**	**W4**	**W1**	**W2**	**W3**	**W4**	**W1**	**W2**	**W3**	**W4**	**W1**	**W2**	**W3**	**W4**	**W1**	**W2**	**W3**	**W4**
Step 1: Finalize requirement																									
Finalize requirement	Approved requirement																								
Step 2: Develop a Proof Of Concept analytics model																									
Use CRISP-DM for developing a data analytics model	Approved Proof of Concept (POC) model																								
Step 3: Design																									
Design a data governance framework and protocol	Data governance framework and protocol																								
Design user journey with the proposed service	Design documents & Prototypes																								
Design data integration flow																									
Design infrastructure																									
Design data pineline																									
Design deployment flow	Deployment flow																								
Step 4: Develop																									
Develop infrastructure	Infrastructure architecture																								
Develop data integration flow	API Documentation																								
Develop data pipeline & analytics model	Analytics model																								
Step 5: User Acceptance Testing (UAT)																									
Test full system integration (Data pipeline, Service Interface, Model, User Journey)	Test results reports																								
Test service performance	Performance test results																								
Test model results	Model evaluation reports																								
Step 6: Deliver																									
Run full loop test on the production system	User Acceptance Testing (UAT) feedback & Test summary																								
Deliver	All software components and infrastructure																								

Resources

Resources crucial for the project's success can be categorized into two main areas: human resources and technological resources.

Human Resources:

- **Client Team:** This team comprises decision-makers, subject matter experts, and the operational working team.
- **Project Leadership:** The project is led by the analytics manager (1 person).
- **Design Team:** A system analyst (1 person) forms the backbone of this team.
- **Development Team:** This team comprises software developers (4 people) and infrastructure developers (2 people).
- **Testing Team:** This team comprises a tester lead (1 person) and an additional tester (1 person).
- **Data Team:** This team comprises a data analyst (1 person), a data scientist (1 person), and a data engineer (1 person).

Technological Resources:

- **Cloud Services:** For this project, we employ cloud computing solutions, such as Amazon Web Services (AWS).
- **Databases:** A robust SQL database system.
- **Programming Tools:** The project employs both R and Python for various computational tasks.
- **Version Control:** Tools specifically designed to manage versions of software models are put in place.

Critical Success Factors

To ensure the success of the project, it's essential that the following critical success factors are established.

- **Clear and well-defined business objectives**: For this project, the objective is to predict the 30-day readmission rates for diabetic patients. Such objectives should address critical challenges, ensuring sustained motivation from all stakeholders throughout the project's duration.
- **Stakeholder engagement and prioritization**: It's crucial that primary stakeholders dedicate sufficient time to the project. For instance, in scenarios where challenges arise, analytics managers should be able to quickly access decision-makers.
- **Expertise of team members:** We operate under the assumption that all team members, especially those on the technical side, are experts in their respective fields. For example, the data scientists are anticipated to have prior experience in constructing analytical models, especially within the healthcare sector. Their expertise ensures the creation of accurate and reliable models.

- **Adherence to health regulations:** It's imperative that the analytics model complies with all relevant health laws and regulations to avoid legal implications and ensure the model's ethical deployment.

Risks and Mitigation Plans

Major risks have been identified for each project step, along with their mitigation plans.

Step	Risks	Mitigation Plans
Step 3: Design	Potential misuse of patients' data.	● Establish a data governance framework and protocol. ● Regularly check to ensure stakeholders are adhering to this framework.
	Designing technical systems that are either overly complex or inadequate.	● Include technical experts in the design process: system analysts, software developers, infrastructure developers, data engineers, data scientists, and testers.
	The user journey design may not align with user needs or expectations.	● Conduct user research and develop prototypes early in the design phase. ● Gather feedback and make continuous iterations to the user journey.
Step 4: Develop	Delays in developing infrastructure may impact the overall project timeline.	● Allocate sufficient resources and prioritize infrastructure development early.
	The hospital might provide low-quality data, affecting the accuracy of data analytic models.	● Identify key variables, data sources, and devise data collection and preparation plans. ● Monitor the quality of data closely.
Step 5: Test	Integration of different components might cause compatibility problems.	● Emphasize continuous integration. ● Utilize automated testing.
	Analytic models may be biased.	● Conduct rigorous model testing and validation, particularly in sensitive areas like race or socio-economic status, to ensure model equity.
Step 6: Deliver	There may be risks not detectable even during a full-loop test.	● Identify crucial errors vital to the system and prioritize their testing.

Communication plans to stakeholders

Communication is vital to ensure alignment with various stakeholders. We've established the following communication plans:

- **Internal Working Team**: In this context, the internal working team refers to the external service provider responsible for building this analytic service for the hospital. The composition of this team may vary depending on the project phase. Nevertheless, they should communicate daily to synchronize all updates and changes.
- **Collaboration Between Our Working Team and the Hospital's Operational Working Team**: The primary objective here is to remain updated and ensure that both teams are aligned toward achieving the client's goals. Regular updates should be conducted either weekly or bi-weekly to maintain this alignment.
- **Reporting to Decision Makers and Key Stakeholders:** It's essential to liaise with decision-makers either bi-weekly or when significant project milestones are reached. This ensures that any necessary changes can be promptly implemented, and alignment with project goals is maintained.
- **Other Stakeholders:** It's important to periodically update other stakeholders to keep them informed about the project's progress. By doing so, we facilitate smoother collaboration when their involvement becomes necessary. For instance, when we plan to deploy the system to medical staff like nurses and doctors, they should be clearly briefed about the project's objectives, solution usage, the origins of the solution (to foster trust), and the channels available to contact the working teams should issues arise.

Division of labor

Aditi Gupta	EDA and Analytics
Dollaya Hirunyasiri	Project Plan Phase 2
Greta Luo	EDA and Analytics
Jorge Palacio	Executive Summary, Problem Framing, and Data sources

References

[1] Auerbach AD, Kripalani S, Vasilevskis EE, et al. Preventability and Causes of Readmissions in a National Cohort of General Medicine Patients. JAMA Intern Med. 2016;176(4):484–493. doi:10.1001/jamainternmed.2015.7863. https://jamanetwork.com/journals/jamainternalmedicine/fullarticle/2498846

[2] Soh JGS, Wong WP, Mukhopadhyay A, et al. Predictors of 30-day unplanned hospital readmission among adult patients with diabetes mellitus: a systematic review with meta-analysisBMJ Open Diabetes Research and Care 2020;8:e001227. https://drc.bmj.com/content/8/1/e001227.citation-tools

[3] Beata Strack, Jonathan P. DeShazo, Chris Gennings, Juan L. Olmo, Sebastian Ventura, Krzysztof J. Cios, and John N. Clore, "Impact of HbA1c Measurement on Hospital Readmission Rates: Analysis of 70,000 Clinical Database Patient Records," BioMed Research International, vol. 2014, Article ID 781670, 11 pages, 2014. https://archive.ics.uci.edu/dataset/296/diabetes+130-us+hospitals+for+years+1999-2008