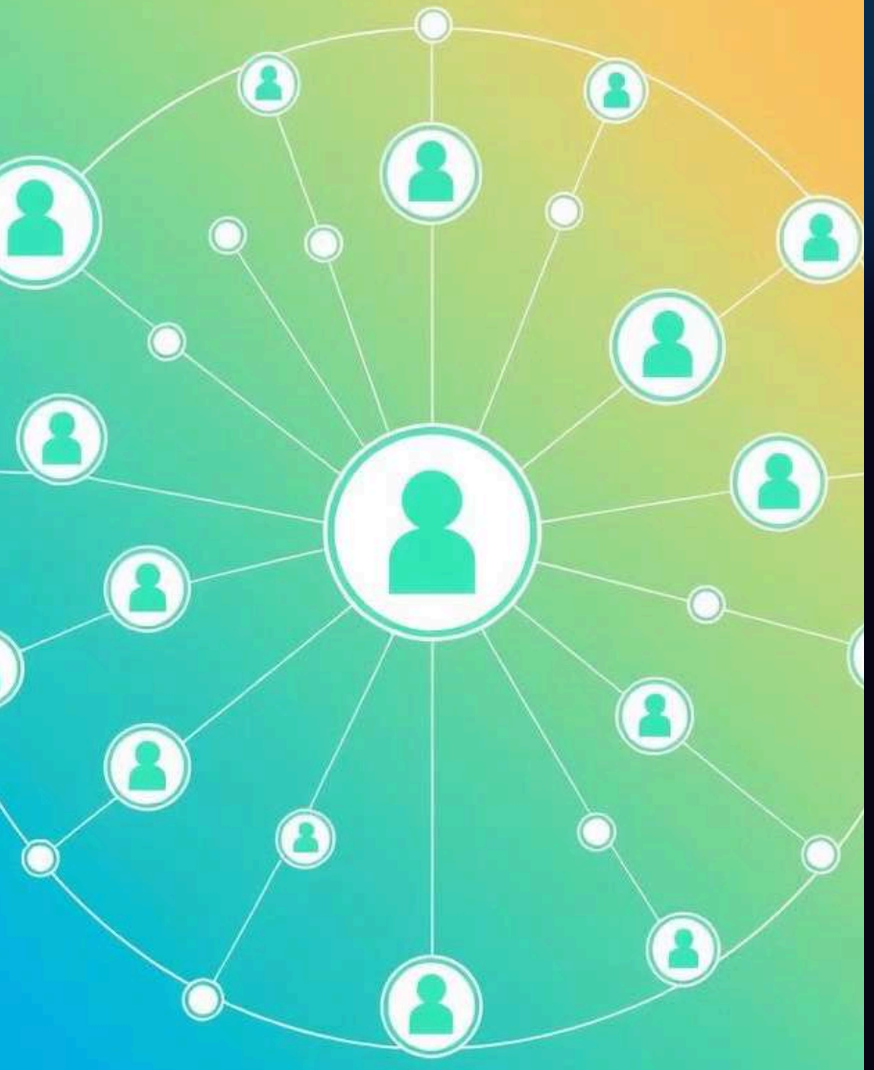
Project Report: Healthcare Data Analysis and Patient Readmission Prediction

This report details a comprehensive analysis of patient data to predict readmission risks. Utilizing various data mining techniques, we aim to identify key factors influencing readmission rates and provide actionable insights for healthcare providers.

OB **by Oindrilla Banerjee**



Executive Summary

1 Project Objective

To develop a predictive model capable of identifying patients at high risk of readmission based on their demographic, clinical, and treatment history data.

2 Data Source

We leveraged a large dataset encompassing patient demographics, medical records, diagnoses, procedures, medications, and readmission status.

3 Methodology

The project involved data preprocessing, exploratory analysis, feature engineering, model development, and performance evaluation.

4 Key Findings

The analysis revealed significant correlations between patient characteristics and readmission likelihood, allowing us to identify key factors contributing to readmission risk.

Data Collection and Preprocessing

1

Data Acquisition

The data was sourced from electronic health records (EHRs) and various healthcare databases.

2

Data Cleaning

We addressed missing values, inconsistencies, and outliers to ensure data integrity.

3

Data Transformation

Variables were transformed and standardized to prepare them for analysis and modeling.

Exploratory Data Analysis

Demographic Analysis

We examined patient age, gender, race, and socioeconomic factors to uncover patterns in readmission rates.

Clinical Analysis

We analyzed patient diagnoses, procedures, medications, and length of stay to identify key clinical predictors.

Readmission Patterns

We explored the timing of readmissions, the most common readmission diagnoses, and factors associated with readmission within specific timeframes.

Feature Engineering and Selection

1

Feature Engineering

We created new features based on existing data, such as interaction terms and derived metrics.

2

Feature Selection

We used statistical methods and machine learning algorithms to identify the most informative features for predicting readmissions.

3

Feature Importance

We analyzed feature importance scores to understand the relative contribution of each feature to the predictive model.

Predictive Modeling for Readmission Risk

Logistic Regression

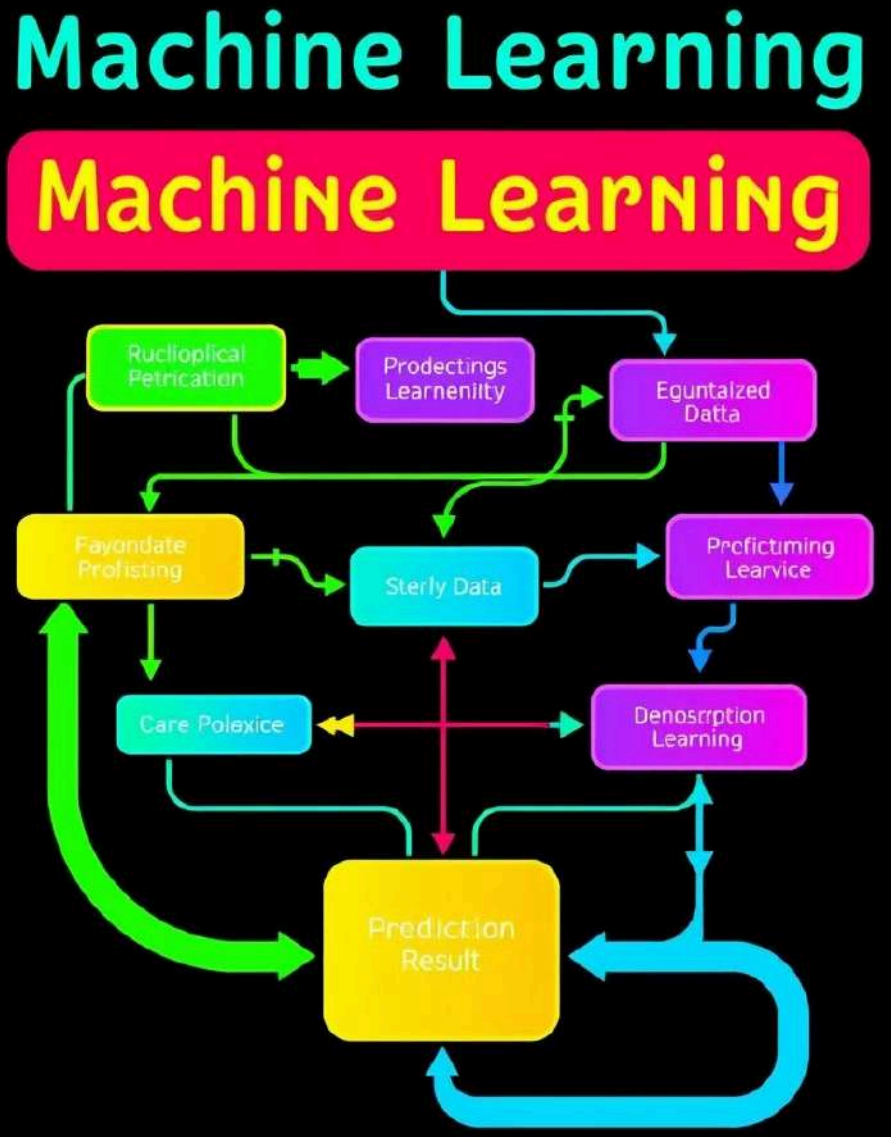
We trained a logistic regression model to predict the probability of readmission based on the selected features.

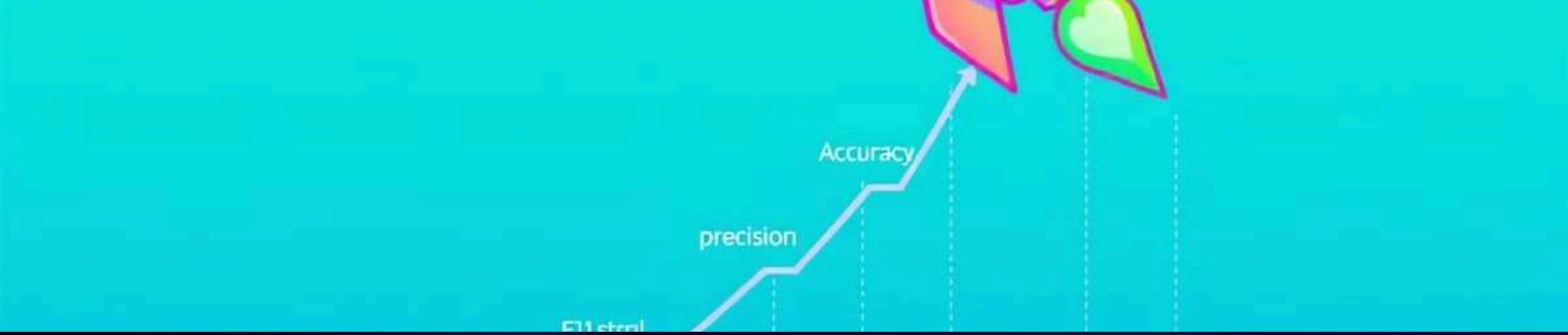
Random Forest

We developed a random forest model to enhance prediction accuracy and explore the influence of multiple features.

Support Vector Machine

We implemented a support vector machine model to address non-linear relationships and complex data patterns.





Evaluating Model Performance

Metric	Logistic Regression	Random Forest	Support Vector Machine
Accuracy	0.78	0.82	0.80
Precision	0.75	0.80	0.77
Recall	0.80	0.85	0.83
F1-Score	0.77	0.83	0.80

Insights and Recommendations

High-Risk Factors

Identified key factors influencing readmission risk, including chronic conditions, medication adherence, and socioeconomic factors.

Targeted Interventions

Proposed interventions to address identified risk factors, such as medication management programs and patient education initiatives.

Data-Driven Decisions

Emphasized the importance of using predictive models to guide clinical decision-making and improve patient care.

Next Steps and Future Considerations

1 Model Refinement

Continue to refine the model by incorporating new data sources, exploring advanced algorithms, and evaluating model performance over time.

2 Real-World Implementation

Collaborate with healthcare providers to implement the model in clinical settings and integrate it into existing workflows.

3 Patient Engagement

Develop strategies to empower patients by providing them with personalized insights and tools to manage their health and reduce readmission risk.