PATIENT CHARACTERISTICS TO PREDICT THE TYPE OF HEALTHCARE SERVICE

Optimizing Healthcare Services through Predictive Analytics

PRESENTED BY

Rabiya Fatima | Srilakshmi Mallipudi | Gautam Reddy | Barkha Sharma

AGENDA

INTRODUCTION

Overview: Analyzing Patient Attributes for Healthcare Optimization.

Background Context: In an era of evolving healthcare demands, our analysis empowers data-driven strategies

Relevance: Enhancing Healthcare Efficiency through Predictive Insights.
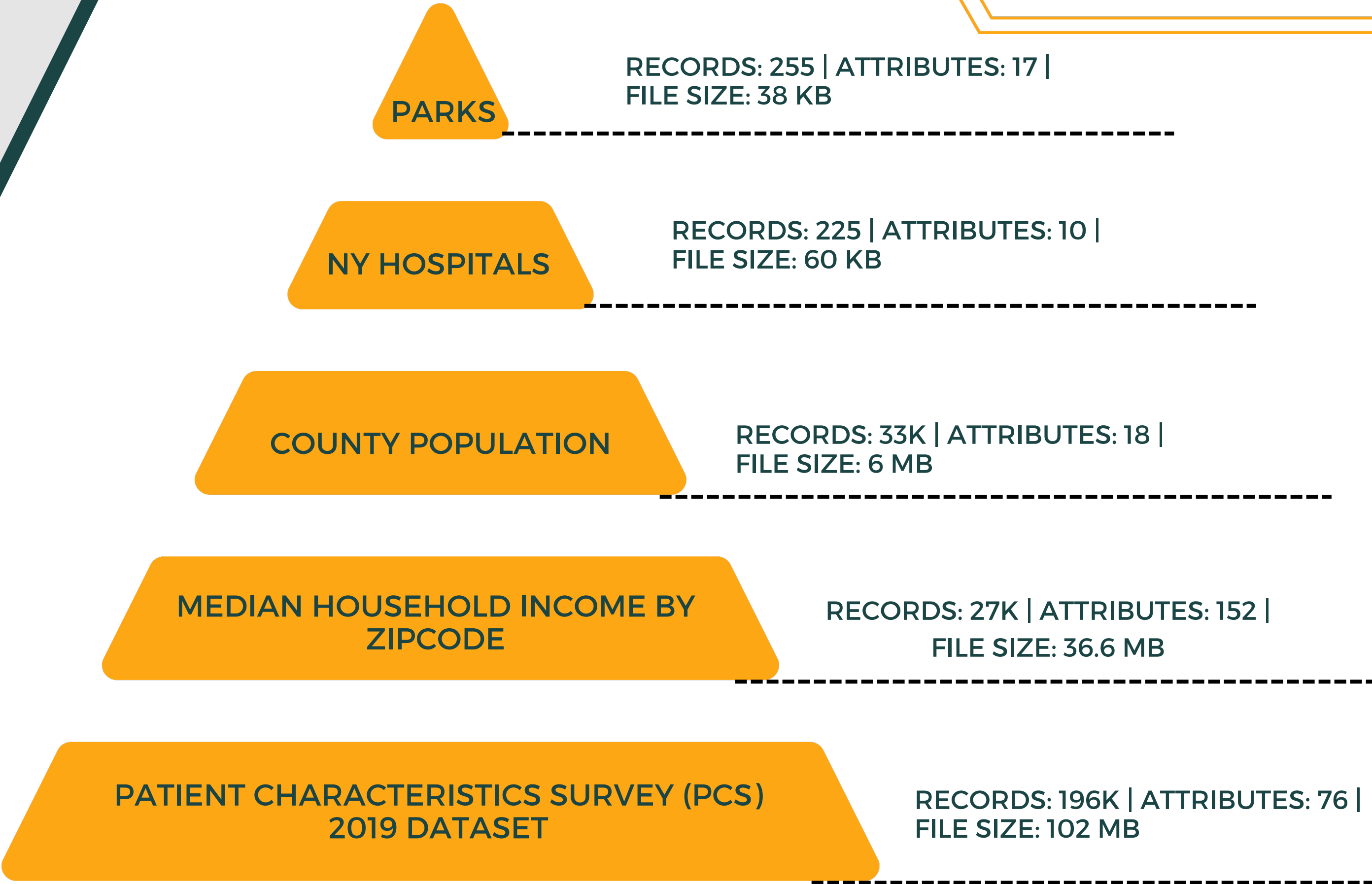
PROBLEM STATEMENT

To predict the type of healthcare service (inpatient, outpatient, residential, emergency) patients are likely to use based on their characteristics and medical history

OBJECTIVES

Revolutionize healthcare operations by deploying predictive models for
proactive patient care and
cost-effective service delivery.

DATA OVERVIEW

PARKS
RECORDS: 255 | ATTRIBUTES: 17 | FILE SIZE: 38 KB

NY HOSPITALS
RECORDS: 225 | ATTRIBUTES: 10 | FILE SIZE: 60 KB

COUNTY POPULATION
RECORDS: 33K | ATTRIBUTES: 18 | FILE SIZE: 6 MB

MEDIAN HOUSEHOLD INCOME BY ZIPCODE
RECORDS: 27K | ATTRIBUTES: 152 | FILE SIZE: 36.6 MB

PATIENT CHARACTERISTICS SURVEY (PCS) 2019 DATASET
RECORDS: 196K | ATTRIBUTES: 76 | FILE SIZE: 102 MB
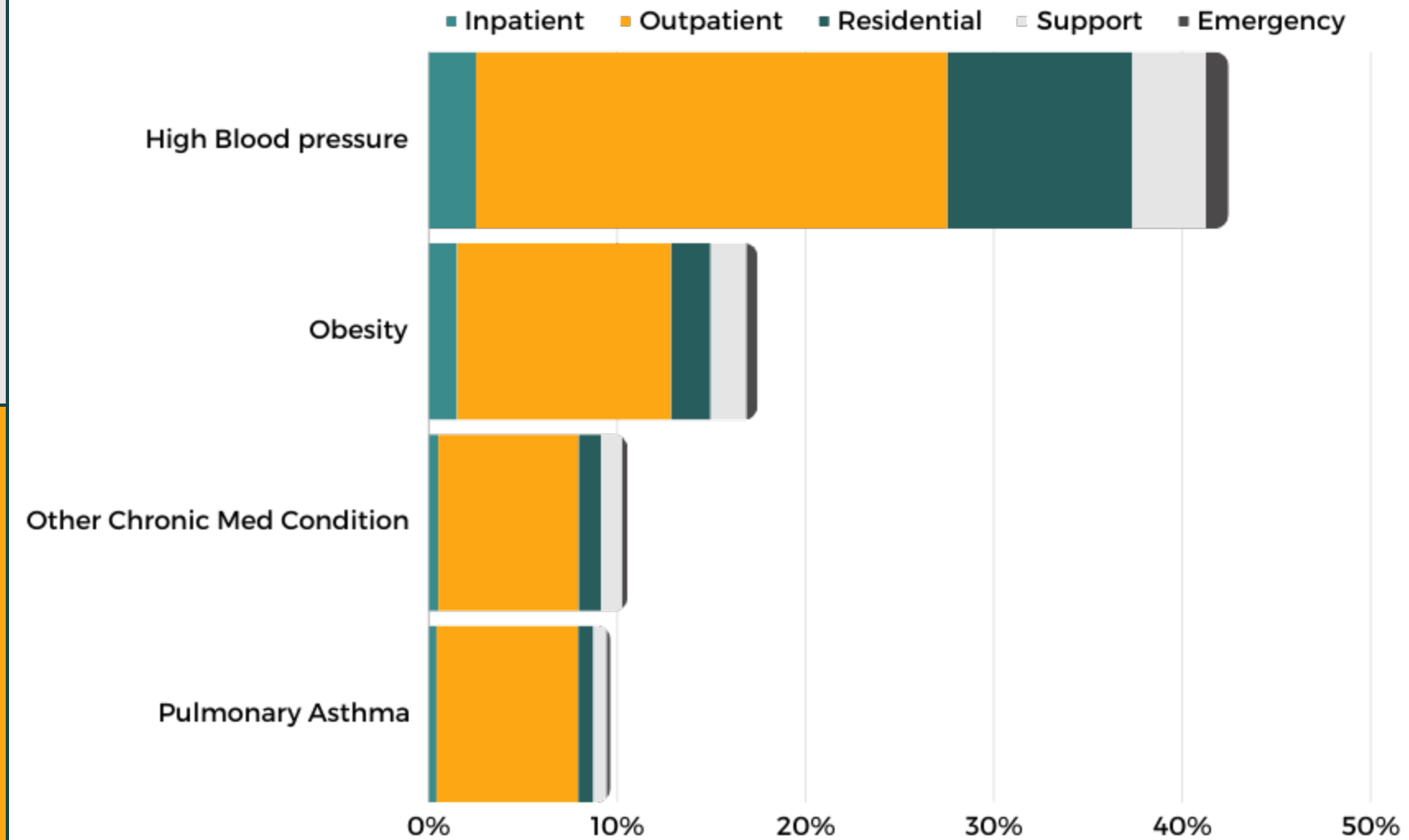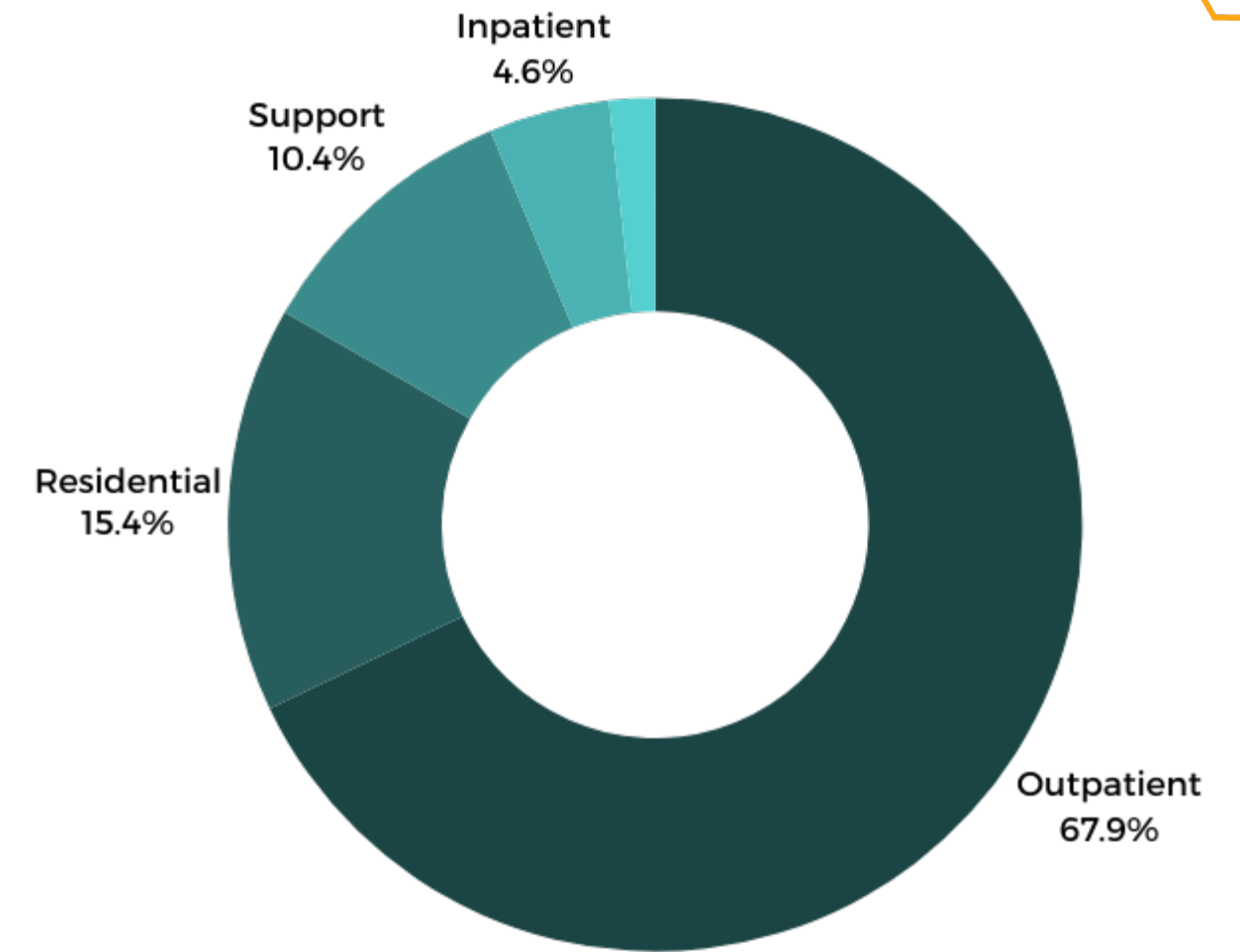
EDA INSIGHTS

Primary Service →

Outpatient care

Among 196K patient records, approx. 78% of patients are adults.

Outpatient services are more commonly chosen by patients with private insurance.

Disease Trends →

Inpatient 4.6%
Support 10.4%
Residential 15.4%
Outpatient 67.9%

■ Inpatient ■ Outpatient ■ Residential ☐ Support ■ Emergency

- High Blood pressure
- Obesity
- Other Chronic Med Condition
- Pulmonary Asthma

0% 10% 20% 30% 40% 50%

DATA PREPROCESSING

Data preparation included vital steps to refine and optimize the dataset for thorough analysis

Integration

Integrated external datasets using zip codes

Cleaning

Handled missing values by eliminating rows (where infrequent), assigning new labels, and replacing unknown values.

Correlation Analysis

Removed highly correlated columns

Converting nominal/ordinal values to numerical values

Mapped ordinal values to numeric scale and used one-hot encoding for nominal attributes.

METHODOLOGY

NORMALIZATION

- Z-score normalization

FEATURE SELECTION

- Importance values using Random Forest Classifier

CLUSTERING

- K-means Clustering

PRINCIPAL COMPONENT ANALYSIS (PCA)

- n = 20

OVERSAMPLING

- Naive Random Oversampler

HYPERPARAMETER TUNING

- RandomSearchCV
- GridSearchCV

MODEL PERFORMANCE EVALUATION

Datasets used

- Baseline Dataset
- Oversampled Dataset
- Oversampled PCA Dataset

Dataset Partitioning

- 70% for training
- 30 % for testing
- Stratification

Model Training Approach

- Trained for each dataset separately
- Evaluated using test data without oversampling

Key Metric

- Recall

Classification Model	Precision	Recall	Accuracy	Computation Time (in secs)
Random Forest	0.98	0.97	0.97	215.25 secs
Decision Tree	0.96	0.96	0.96	22.84 secs
Gradient Boosting	0.78	0.79	0.79	2322.39 secs

PREDICTIVE INSIGHTS

SSI Cash Assistance

Diagnosis Class: Mental Illness

Have No Insurance
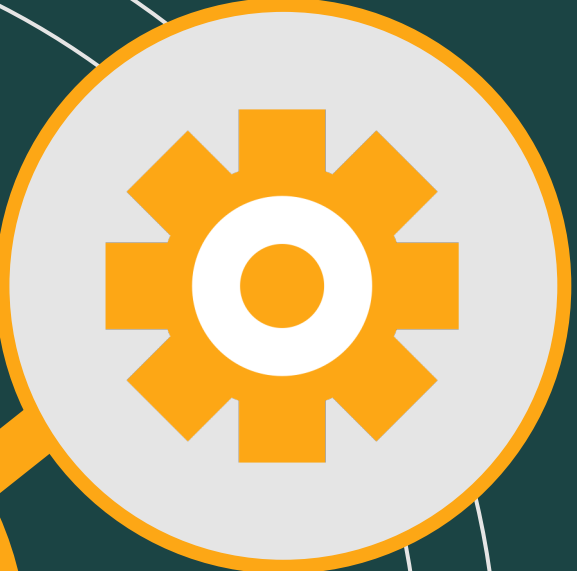
Outpatient Inpatient Services

Have Hospital in vicinity

From New York City

Education Status

USE CASE AND STAKEHOLDER

Hospital Administrators

◆ **Resource Allocation Challenge**
Hospitals struggle to efficiently allocate resources across services, resulting in operational inefficiencies and cost concerns.

◆ **Uncertain Service Demands**
Determining exact service needs without predictive insights leads to staffing and resource mismatches.

RECOMMENDATIONS

◆Implement predictive models for proactive resource allocation in healthcare settings.

◆ Foster proactive interventions using predictive insights for regional patient care plans

◆Lead proactive healthcare approaches via early identification and intervention for prevention strategies.

◆ Tailor care plans based on data insights to enhance patient engagement and outcomes.

FUTURE DIRECTIONS

- Expanded Data Integration
- Cross-Regional Validation
- Collaborative Data Collection
- Specialized Model Development
- Disparity Analysis

CONCLUSIONS

Modelling Success and Limitations

- Successful modeling observed for non-emergency services.
- Random Forest, Decision Tree and Gradient Boosting models showcased notable performance.

Challenges in Emergency Service Prediction

- Accurately forecasting emergency services remains a challenge due to their unpredictable nature.

Data Validation and Model Refinement

- Pioneering advancements in healthcare service prediction models.
- Urging continuous refinement for enhanced real-world applicability.

Thank You

Do you have any questions?