Predicting Patients' Length of Stay in The Emergency Department at Mayo Clinic (MN)

Mohammed Baaoum

MAYO CLINIC

Research Background and Motivation

- EDs across the US facing longer patient wait times and an increase in the volume of patients visits (Di Somma et al., 2015; Higginson et al., 2011)

- Studies identified many adverse effects of overcrowding on quality of care and patient safety :
- Reduces timeliness of care and increases the likelihood of mortality and morbidity (Hoyle, 2013; Miro et al., 1999)
- Increases burnout among ED staff (Johnston et al., 2016)
- Increase the number of patients left without being seen (Rowe et al., 2006; Stock et al., 1994)

Research Background and Motivation

- Patient Length of Stay (LOS) is an important indicator of emergency department (ED) performance that correlates strongly with overcrowding and service quality.
- Prolonged LOS is viewed as evidence of poor hospital performance and has been associated with a higher mortality rate ((Pitts *et al.*, 2014, Singer, 2011).
- It is estimated that reducing ED LOS by 1 hour could decrease the number of deaths in high-risk patients by 6.5% and by almost 13% in lower-risk patients (Barish, & Arnold , 2012)
- Managers aim to predict the length of patient stays to address this.

Research Ideas Presented

1. The goal of this research is to predict patient stay duration in ED using ML model to prevent strain situations.

2. Train ML model to predict patient LOS (regression based)
    - **Regression model**

2. Train ML model to predict patient prolonged (LOS > 4)
    - **Classification model**

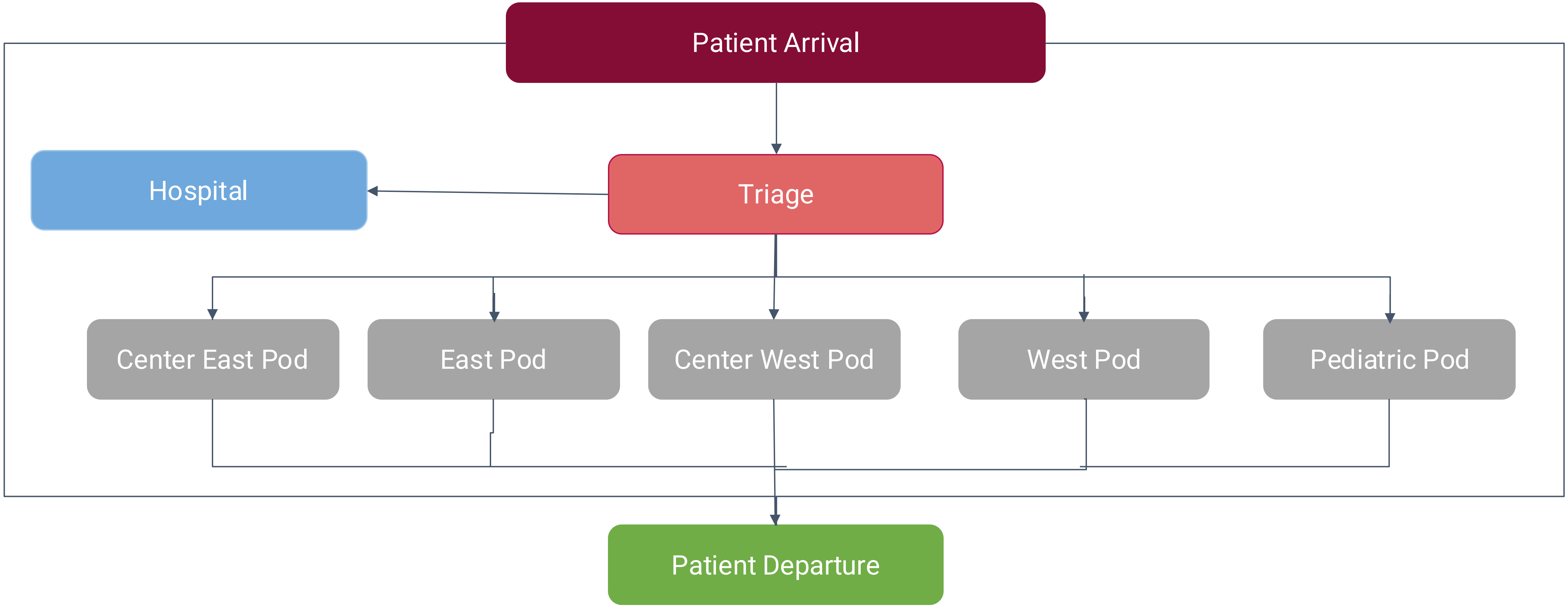
3. Train ML model to predict patient admission to hospital
    - **Classification model**

Research Environment and Roadmap

ED in Saint Mary's Hospital

- The department operates a 51-bed, 5-pod facility.
- 54 physicians provide care (Highest :Trauma Level I)
- The vast majority of patients come from Olmsted County and the six adjacent counties in southeastern Minnesota.
- On average, 20 percent of emergency department patients are children
- A high number of emergency department patients have acute or critical emergency needs.

ED Patient Flow



ED Process

1. **ESL prioritization**
   Patients enter triage and are prioritized by Emergency Severity Level (ESL) 1-5 by nurse. Vitals, pain level, medications, and more are recorded

2. **Initial work-up by nurse**
   Patient is taken to a pod, where the nurse may ask quesitons to gather additional information

3. **Care by physician**
   Physician may run tests, prescribe medication, order additional tests, and gather further details

4. **Follow-up by nurse**
   Nurse may administer medication and provide further care

5. **Patient discharge**
   The patient has received needed ED care and may be sent home, to an in-patient unit, to the ICU, or into surgery

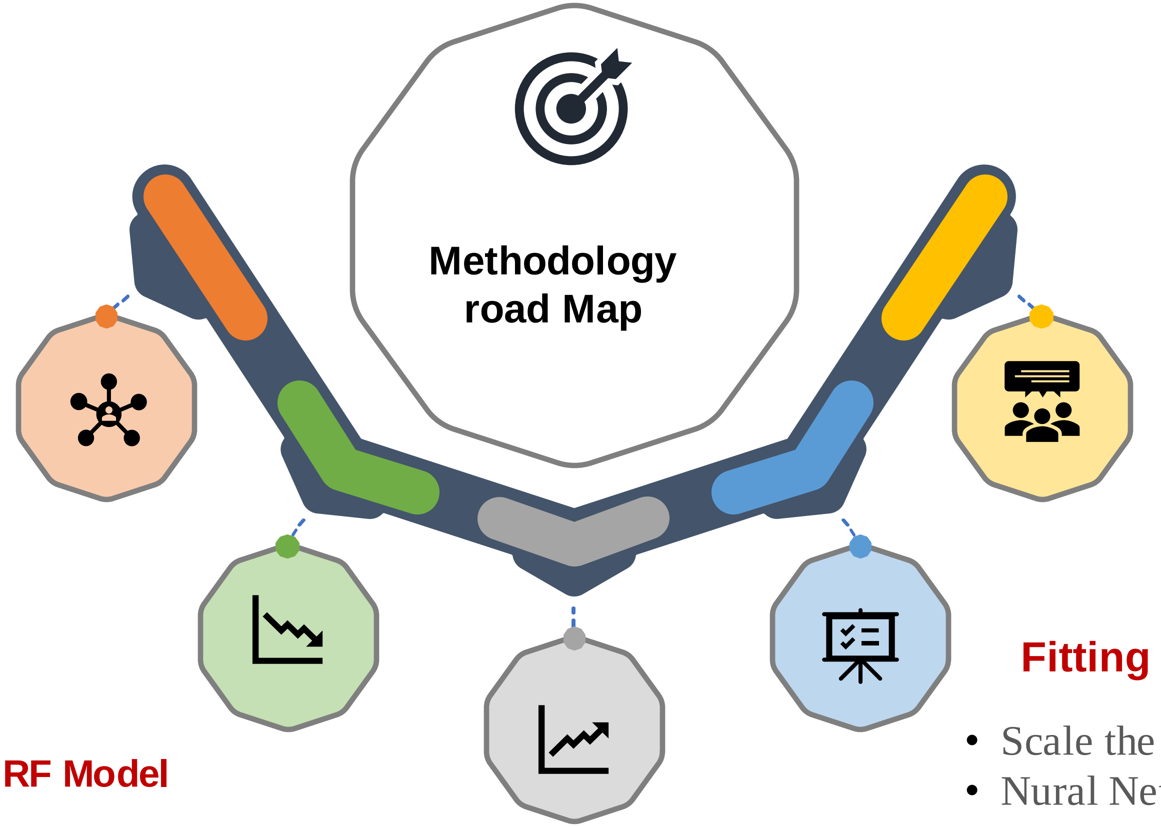
Research Roadmap

Model Evaluation,Hyperpramter Tunining, Feature importance

- Evaluated each model performance by calculating R^2 and MSE.
- Conducted hperpramter tuning.
- Plotted feature importee of each model.

Data Wrangling-Preprocessing

- Selct relvant variables
- Remove missing values, error , negative time
- Timestampt → daily shift
- Remove outliers (mental ilniess, direct admit, weekend)
- Hot encoding
- Disciptive Statstics, Visulaization
- Split the data into training and testing



Fitting RF Model

- Regression basedRandom Forest
- N-estimator =100

Fitting XG Boos

- Regresstionbased Xgboost
- N-estimator=100

Fitting NN

- Scale the feature
- Nural Network Reg model
- One layer with 100 neurons
-

Data Analysis Work

Dataset

- The dataset was gathered throughout the 2016–2017 fiscal year from Mayo Clinic Emergency Department in Rochester MN
- Over 79,000 patients from arrival to discharge from the ED.
- **Variables include** demographic of patients, organizational variables, patient variables, medical staff

Category	Variable	Description
	Gender	Patient gender: male or female
	Age	Patient age: below 21, 21–65, above 65
	Emergency severity index	Five-level ED triage algorithm, from 1 (most urgent) to 5 (least urgent), based on acuity and resource needs
Patients	Arrival mode	How patient arrived: walk-in, wheelchair, ambulance
	Discharge type	Discharge home, inpatient admission, hospital observation
	Arrival shift	Shift when a patient arrived: Shift 1: 12:00 AM to 8:00 AM, Shift 2: 8:00 AM to 4:00 PM, Shift 3: 4:00 PM to 12:00 AM
	Length of stay	The time a patient spent in the ED starting from registration
	First treatment ward	Initial ward where a patient was treated (Center 1, Center 2, East, West, Pediatric)
Service	Ward transfer	Whether a patient was transferred to another ward (yes or no)
	Care complexity (ED level)	Complexity of care provided to a patient, where a higher value represented higher complexity (Levels 1–5 and critical care)
	Patients per MD	Number of patients per MD when a patient arrived
Organizational	Patients per nurse	Number of patients per nurse when a patient arrived
	Patients per resident	Number of patients per resident when a patient arrived

Data Pre-Processing and Feature Engineering

1. Data Preprocessing Techniques

Temporal Feature Encoding

- Time variables converted to categorical features (month, daily shifts).
- Identified three shifts to analyze ED operational hours and patient flow.

Data Cleansing

- Removed anomalies like missing/negative patient arrival times to enhance data quality.
- Removing week-ends

2. Data Subset Selection

Exclusion of Mental Treatment Pods

- Removed mental health care pods from the dataset.
- Focused on general ED operations for typical patient experience analysis.

Data Pre-Processing and Feature Engineering

3. Outlier Analysis and Treatment

Visualization and Detection

- Identified outliers (LOS > 24 hours) using box plots and scatter plots.
- Found cases of misdirected mental health cases.

Removal of Outliers

- Eliminated 44 observations, including 37 misassigned mental health cases, to prevent data skew.

4. Feature Engineering

Age Categorization

- Divided patient ages into <21, 21-65, >65 groups.
- Aligns with ED studies and accounts for age-related healthcare usage variations.

Visualization for LOS

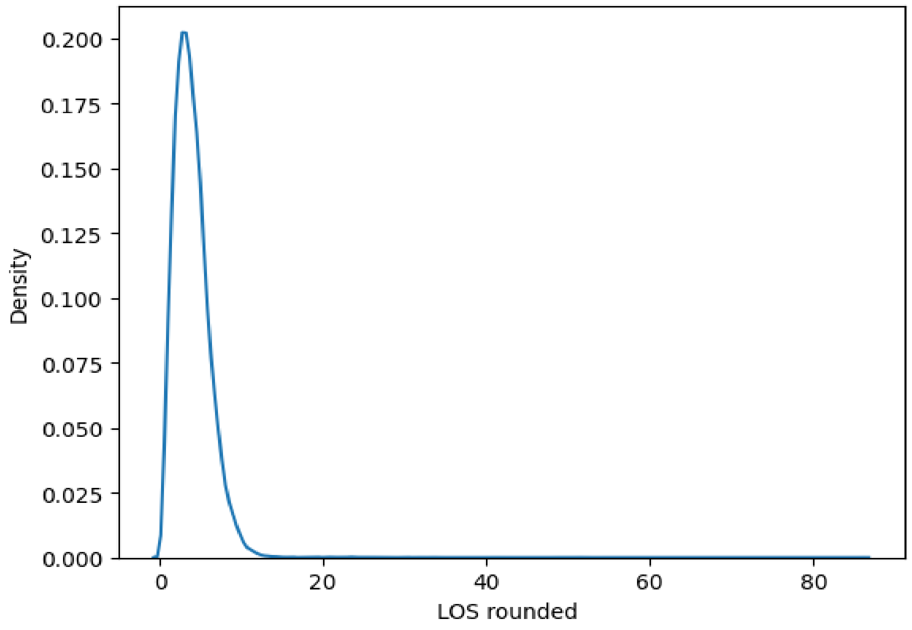


Figure 1: Length of Stay in Hours Before Cleaning Outliers

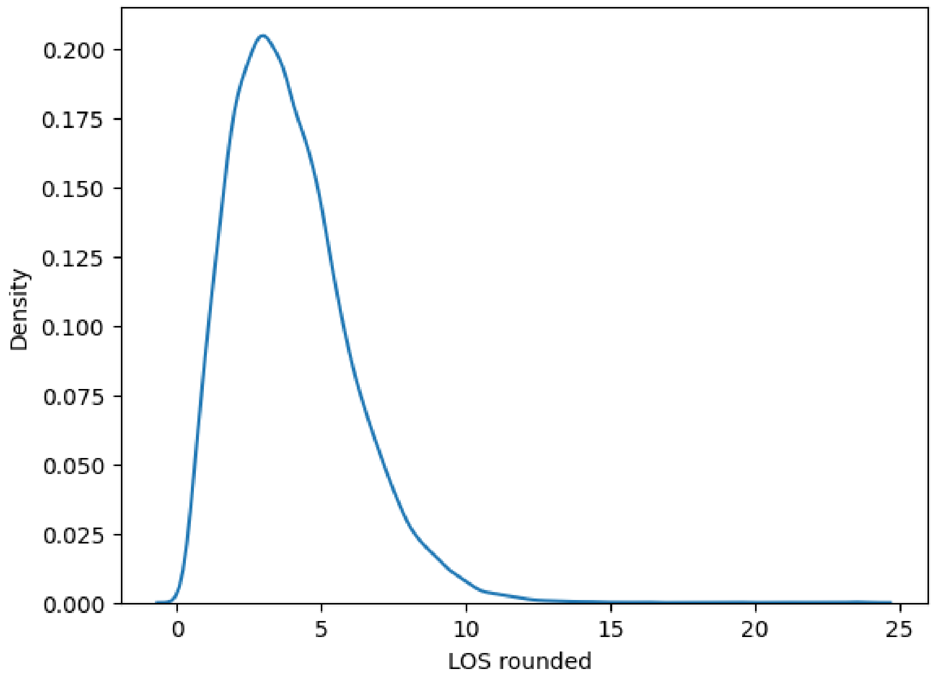
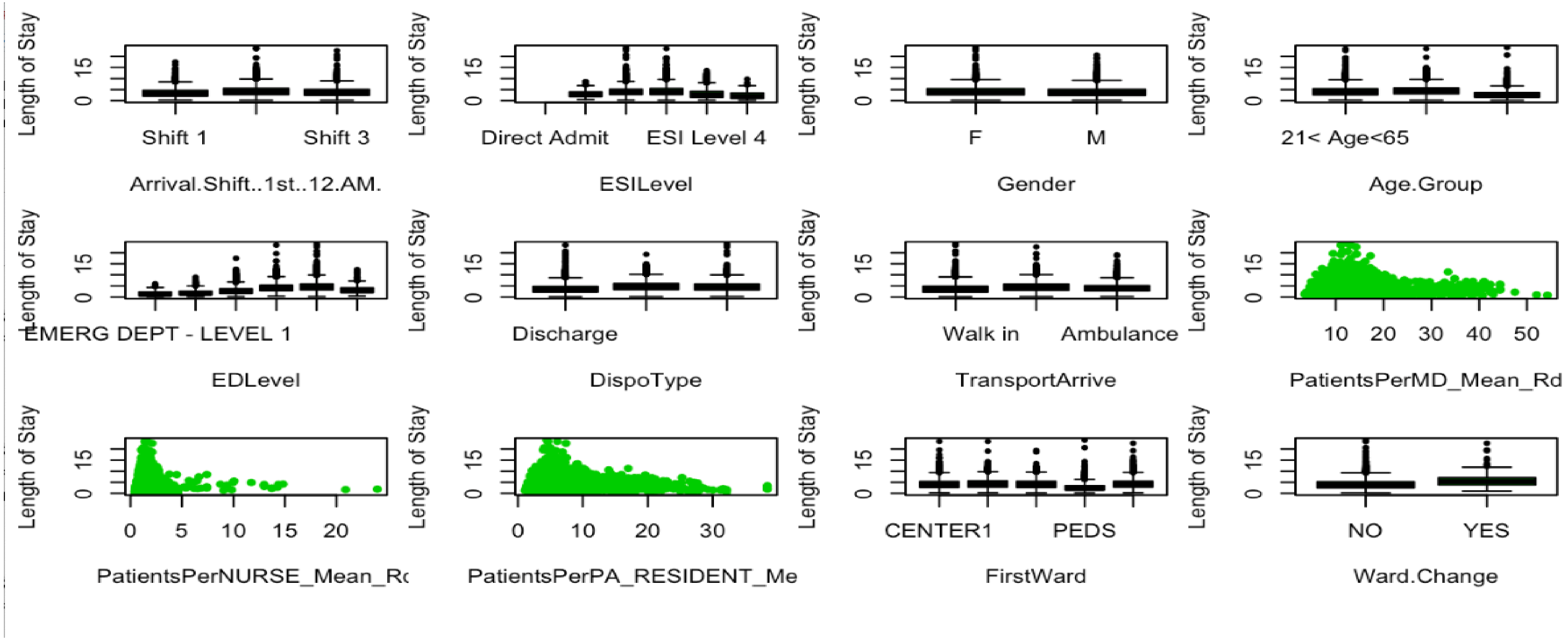


Figure 2: LOS After Cleaning Outliers

Visualization of Response and Predictor Variables



A total of 42,462 observations were analyzed for the entire year after data pre-proccing

Demographic Information of Nominal Variables

Category	Variable	N	%
Patient Gender	Male	20,404	48.05
	Female	22,058	51.95
Patient Age	< 21	7,241	17.05
	21-65	22,060	51.95
	> 65	13,161	31.00
Patient Severity Index	Level 1	360	0.85
	Level 2	8,210	19.33
	Level 3	27,140	63.92
	Level 4	6,529	15.38
	Level 5	223	0.52
Arrival Shift	Shift 1	6,408	15.09
	Shift 2	19,502	45.93
	Shift 3	16,552	38.98
Patient Arrival Mode	Walk-in	21,362	50.31
	Wheelchair	10,465	24.65
	Ambulance	10,635	25.04
Patient Discharge	Discharge home	27,859	65.61
	Inpatient admission	9,314	21.93
	Hospital observation	5,289	12.46
Ward	Center 1	11,328	26.68
	Center 2	9,911	23.34
	East	4,927	11.6
	West	10,475	24.67
	Pediatric	5,821	13.71
Ward Transfer	Yes	433	1.02
	No	42,029	98.98
ED Care Level	Level1	112	0.26
	Level2	1409	3.32
	Level3	10106	23.8
	Level4	12939	30.47
	Level5	17294	40.73
	Critical Care	602	1.42

Descriptive Statistics of Continuous Variables

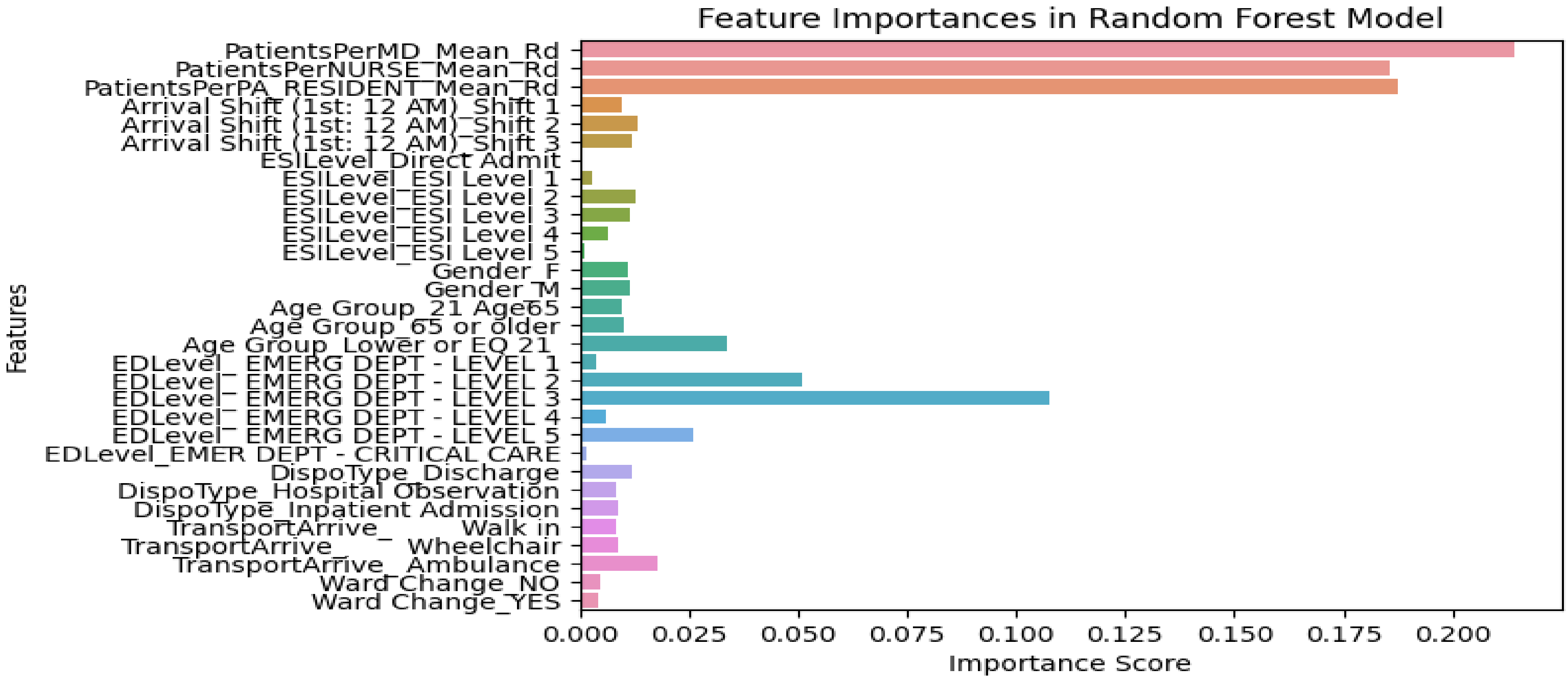
Continuous Variable	Mean	SD	Minimum	Q1	Median	Q3	Maximum
LOS	4.06	2.09	0.18	2.52	3.77	5.25	23.97
Patients per MD	11.99	3.188	3.57	9.94	11.64	13.73	54.3
Patients per nurse	1.59	0.45	0.44	1.34	1.57	1.81	23.97
Patients per resident	5.035	1.839	1.083	4.066	4.066	5.69	38.43

Results

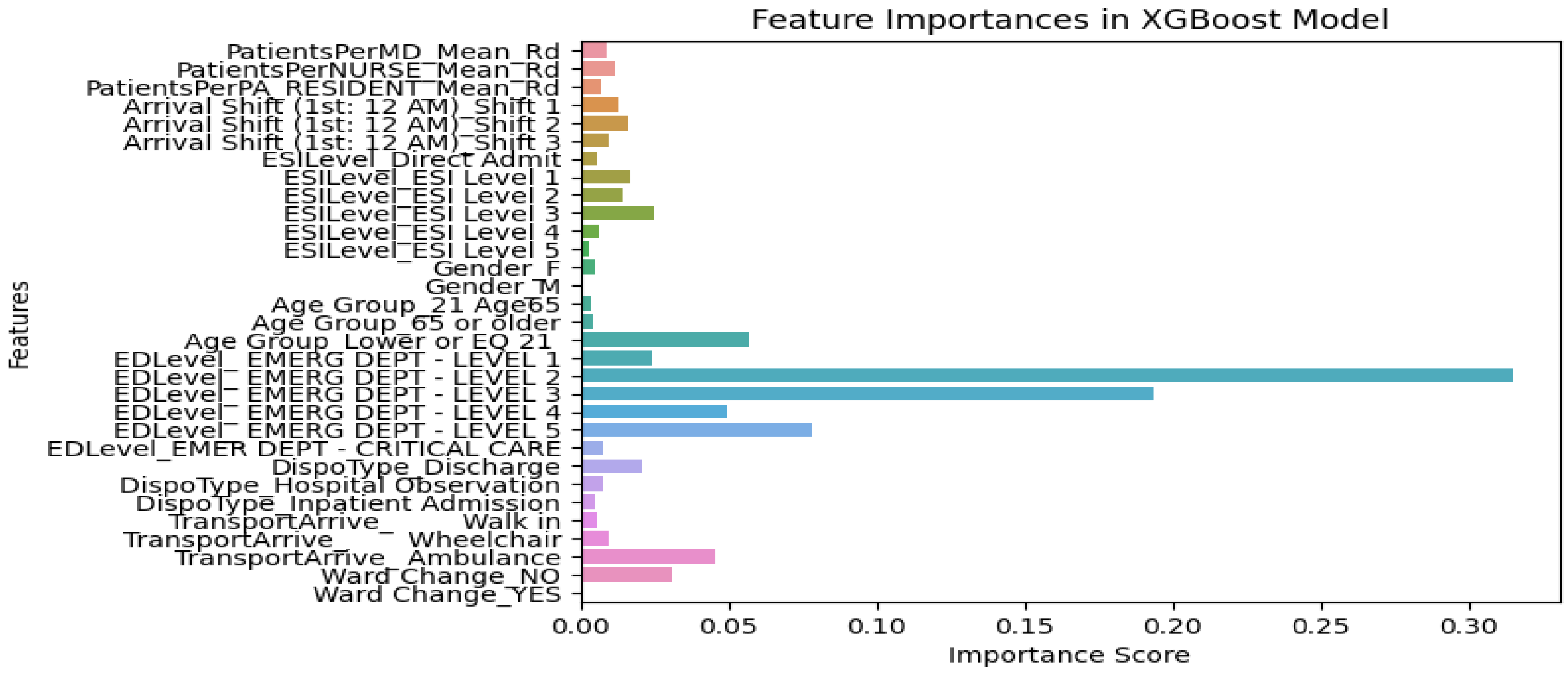
Results Before And After Tunning

	Before Hyperparameter Turning		After Hyperparameter Turning	
ML-Model	R^2	MSE	R^2	MSE
RF	0.37	2.6	0.38	2.5
XG-Boost	0.39	2.5	0.4	2.4
NN	0.3878	2.6	0.385	2.5

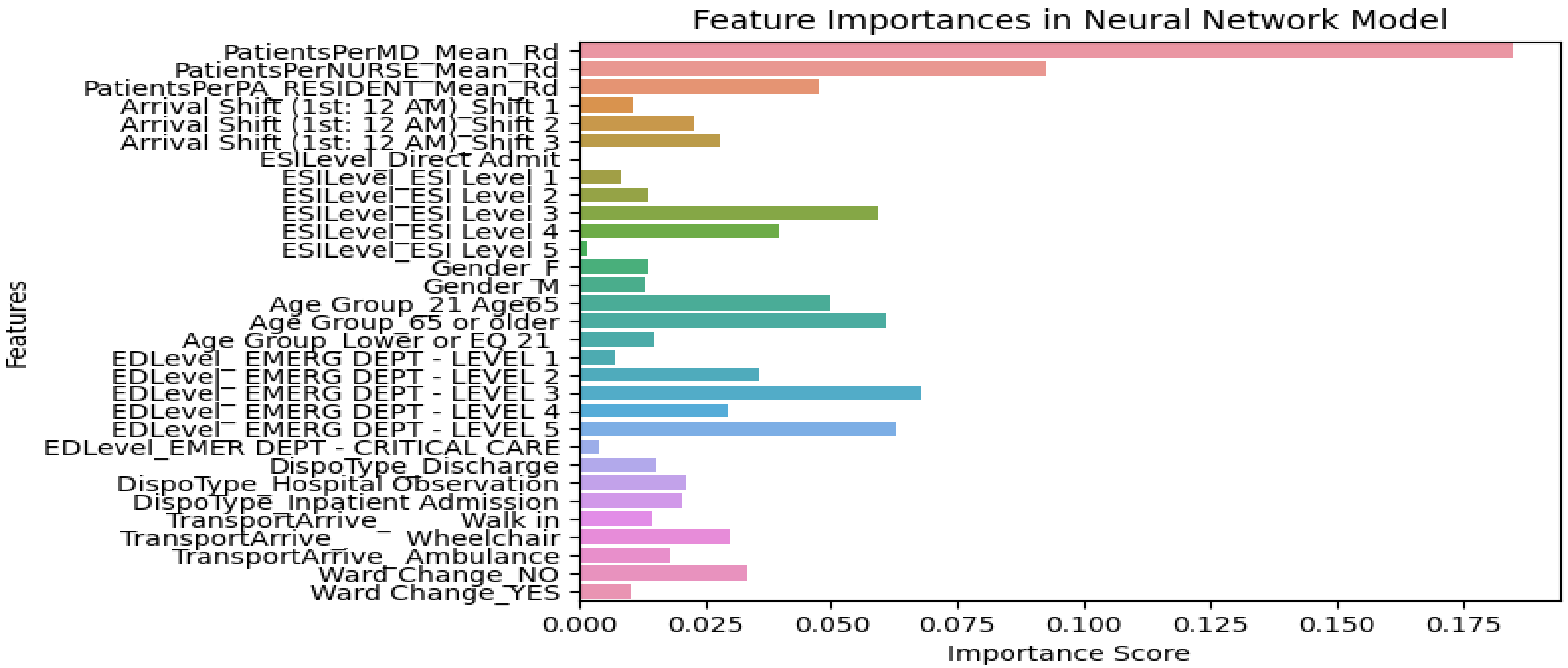
Feature Importance RF



XG- Boost Feature Importance



NN Feature Importance



MAYO CLINIC

Q&A