Clinical Decision Support System for OR Case Length Prediction

Brandon Gambrall
SAT5141 CDSS
12/5/2025

CONTENTS

01 INTRODUCTION

02 ML FRAMEWORK

03 FINDINGS

04 ANALYSIS & CONCLUSION

INTRODUCTION

The biggest cause of chaos:
OR schedule

Major cause of tension between staff, management, and amongst providers.

Block Time, Releases, and Add Ons

Tale of two practices:
-Overestimate case to preserve block time
-Underestimate case to squeeze more

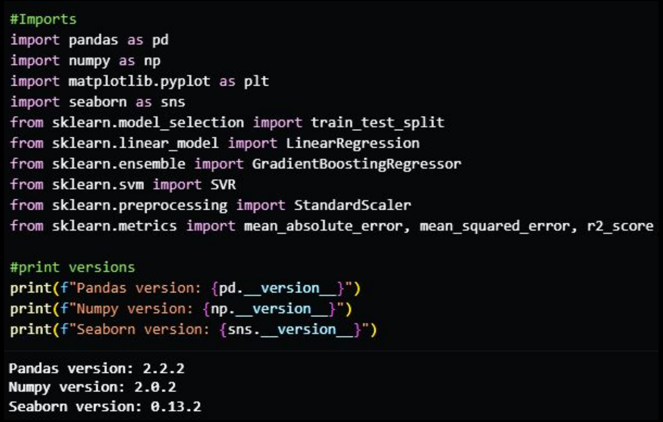
CDSS to Generate Predicted Case Time to support scheduler

Project started out as classification moved to regression

1

ML FRAMEWORK

IMPORTS
DATASET PATH/DF HEAD
PREPROCESSING
TARGET/FEATURE SELECTION
SPLIT
LR/SVR/GBR MODELS
MODEL EVALUATIONS/ANALYSIS
MODEL SELECTION

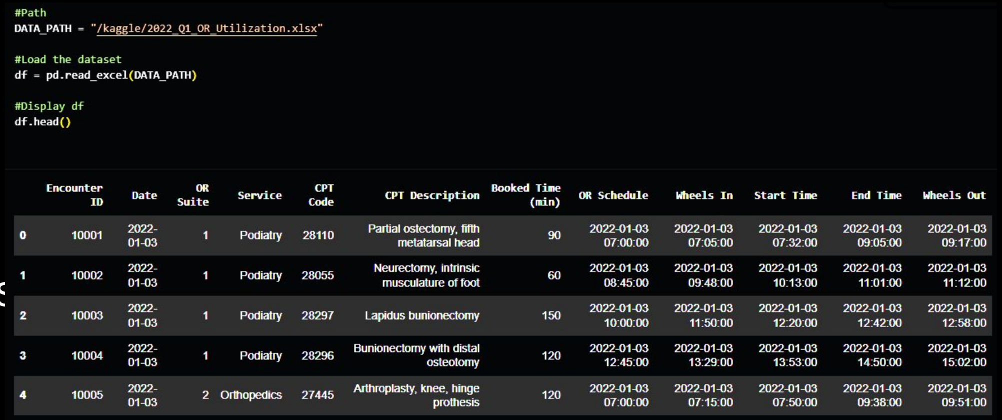


```
#Imports
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.ensemble import GradientBoostingRegressor
from sklearn.svm import SVR
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score

#print versions
print(f"Pandas version: {pd.__version__}")
print(f"Numpy version: {np.__version__}")
print(f"Seaborn version: {sns.__version__}")
```

```
Pandas version: 2.2.2
Numpy version: 2.0.2
Seaborn version: 0.13.2
```

ML FRAMEWORK

- IMPORTS
- DATASET PATH/DF HEAD
- PREPROCESSING
- TARGET/FEATURE SELECTION
- SPLIT
- LR/SVR/GBR MODELS
- MODEL EVALUATIONS/ANALYSIS
- MODEL SELECTION



```
#Path
DATA_PATH = "/kaggle/2022_Q1_OR_Utilization.xlsx"

#Load the dataset
df = pd.read_excel(DATA_PATH)

#Display df
df.head()
```

	Encounter ID	Date	OR Suite	Service	CPT Code	CPT Description	Booked Time (min)	OR Schedule	Wheels In	Start Time	End Time	Wheels Out
0	10001	2022-01-03	1	Podiatry	28110	Partial ostectomy, fifth metatarsal head	90	2022-01-03 07:00:00	2022-01-03 07:05:00	2022-01-03 07:32:00	2022-01-03 09:05:00	2022-01-03 09:17:00
1	10002	2022-01-03	1	Podiatry	28055	Neurectomy, intrinsic musculature of foot	60	2022-01-03 08:45:00	2022-01-03 09:48:00	2022-01-03 10:13:00	2022-01-03 11:01:00	2022-01-03 11:12:00
2	10003	2022-01-03	1	Podiatry	28297	Lapidus bunionectomy	150	2022-01-03 10:00:00	2022-01-03 11:50:00	2022-01-03 12:20:00	2022-01-03 12:42:00	2022-01-03 12:58:00
3	10004	2022-01-03	1	Podiatry	28296	Bunionectomy with distal osteotomy	120	2022-01-03 12:45:00	2022-01-03 13:29:00	2022-01-03 13:53:00	2022-01-03 14:50:00	2022-01-03 15:02:00
4	10005	2022-01-03	2	Orthopedics	27445	Arthroplasty, knee, hinge prothesis	120	2022-01-03 07:00:00	2022-01-03 07:15:00	2022-01-03 07:50:00	2022-01-03 09:38:00	2022-01-03 09:51:00

ML FRAMEWORK

IMPORTS
DATASET PATH/DF HEAD
PREPROCESSING
TARGET/FEATURE SELECTION
SPLIT
LR/SVR/GBR MODELS
MODEL EVALUATIONS/ANALYSIS
MODEL SELECTION

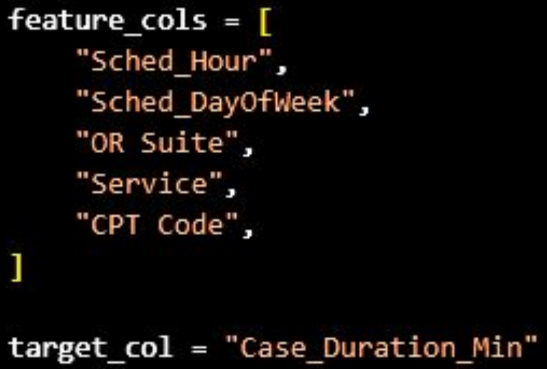
Removed Columns that added no value, redundant value/cause data leak:

Wheels In/Out, Start/End, OR Schedule Start Time, Date, Booked Time

Checked for Null Values or 0 values (Cancelled cases.

ML FRAMEWORK

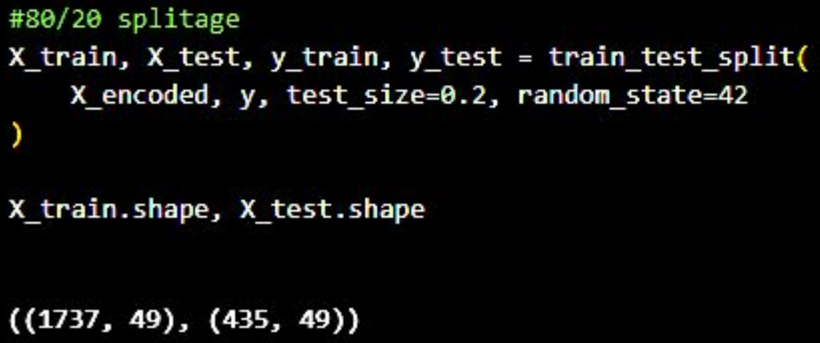
IMPORTS
DATASET PATH/DF HEAD
PREPROCESSING
TARGET/FEATURE SELECTION
SPLIT
LR/SVR/GBR MODELS
MODEL EVALUATIONS/ANALYSIS
MODEL SELECTION



```
feature_cols = [
    "Sched_Hour",
    "Sched_DayOfWeek",
    "OR Suite",
    "Service",
    "CPT Code",
]

target_col = "Case_Duration_Min"
```

ML FRAMEWORK

IMPORTS
DATASET PATH/DF HEAD
PREPROCESSING
TARGET/FEATURE SELECTION
SPLIT
LR/SVR/GBR MODELS
MODEL EVALUATIONS/ANALYSIS
MODEL SELECTION

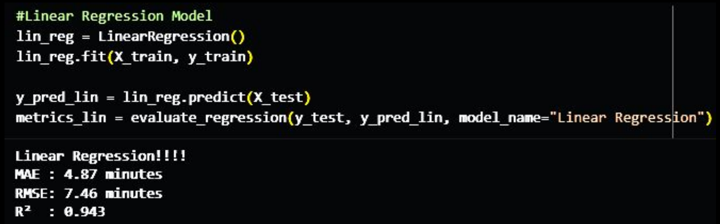


```
#80/20 splitage
X_train, X_test, y_train, y_test = train_test_split(
    X_encoded, y, test_size=0.2, random_state=42
)

X_train.shape, X_test.shape
```

```
((1737, 49), (435, 49))
```

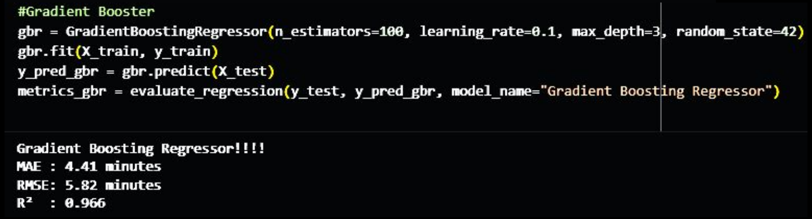
FINDINGS

LR/SVR/GBR MODELS



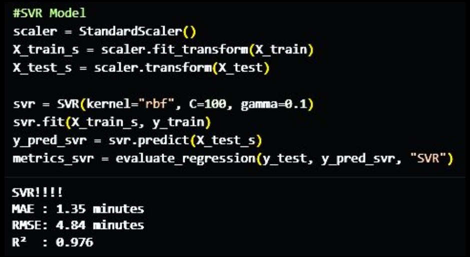
```
#Linear Regression Model
lin_reg = LinearRegression()
lin_reg.fit(X_train, y_train)

y_pred_lin = lin_reg.predict(X_test)
metrics_lin = evaluate_regression(y_test, y_pred_lin, model_name="Linear Regression")
```

```
Linear Regression!!!!
MAE : 4.87 minutes
RMSE: 7.46 minutes
R²  : 0.943
```



```
#Gradient Booster
gbr = GradientBoostingRegressor(n_estimators=100, learning_rate=0.1, max_depth=3, random_state=42)
gbr.fit(X_train, y_train)
y_pred_gbr = gbr.predict(X_test)
metrics_gbr = evaluate_regression(y_test, y_pred_gbr, model_name="Gradient Boosting Regressor")
```

```
Gradient Boosting Regressor!!!!
MAE : 4.41 minutes
RMSE: 5.82 minutes
R²  : 0.966
```

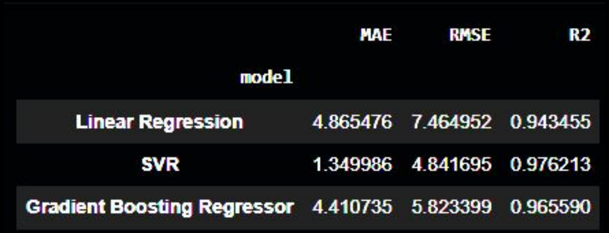


```
#SVR Model
scaler = StandardScaler()
X_train_s = scaler.fit_transform(X_train)
X_test_s = scaler.transform(X_test)

svr = SVR(kernel="rbf", C=100, gamma=0.1)
svr.fit(X_train_s, y_train)
y_pred_svr = svr.predict(X_test_s)
metrics_svr = evaluate_regression(y_test, y_pred_svr, "SVR")
```

```
SVR!!!!
MAE : 1.35 minutes
RMSE: 4.84 minutes
R²  : 0.976
```

3

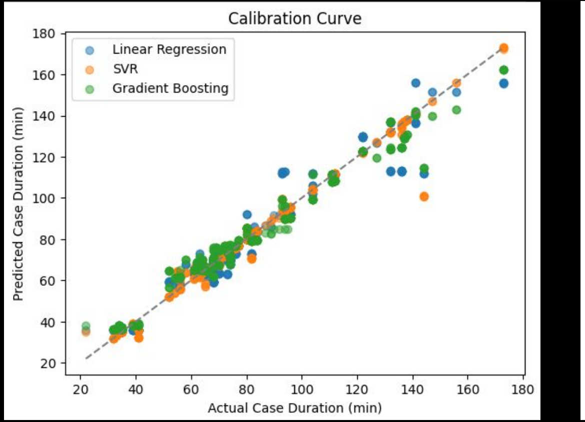
ANALYSIS

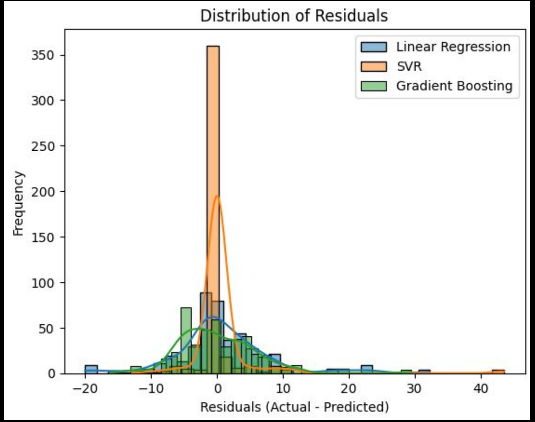


model	MAE	RMSE	R2
Linear Regression	4.865476	7.464952	0.943455
SVR	1.349986	4.841695	0.976213
Gradient Boosting Regressor	4.410735	5.823399	0.965590

MODEL EVALUATIONS/ANALYSIS

MODEL SELECTION &
CONCLUSION



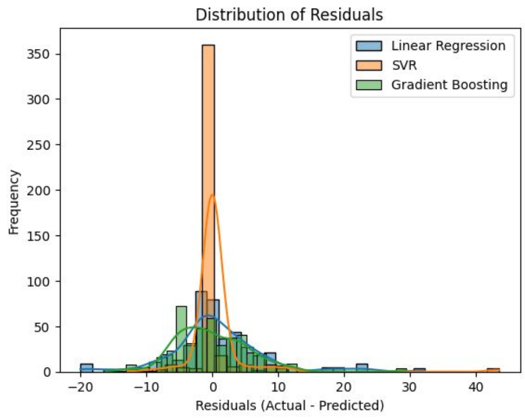


4

ANALYSIS

MODEL EVALUATIONS/ANALYSIS
MODEL SELECTION &
CONCLUSION

SVR showed the strongest model:
Tighter Residuals
Higher R2 value
One really erroneous



Honorable Mentions:
GBR did well but had some outliers that were +/- 20 to 30 min
Lin model was fair but had a long positive tail and underestimated long cases.

WORK CITED

B, Arunachalam. “How to Build a Linear Regression Model – Machine Learning Example.” *freeCodeCamp.Org*, freeCodeCamp.org, 6 Sept. 2023, www.freecodecamp.org/news/build-a-linear-regression-model-with-an-example/.

Falk, J. (2022). Operating Room Utilization. Kaggle. https://www.kaggle.com/datasets/thedevastator/optimizing-operating-room-utilization

Hembroff, Guy. “SAT5141 Disease Progression.” *CAS*, SAT5141, mtu.instructure.com/courses/1575642/pages/part-i-disease-progression-modeling-dpm. Accessed 20 Nov. 2025.

Saraswat, Manish. “Beginners Guide to Regression Analysis and Plot Interpretations Tutorials & Notes: Machine Learning.” *HackerEarth*, www.hackerearth.com/practice/machine-learning/machine-learning-algorithms/beginners-guide-regression-analysis-plot-interpretations/tutorial/. Accessed 5 Dec. 2025.

“Support Vector Regression (SVR) Using Linear and Non-Linear Kernels in Scikit Learn.” *GeeksforGeeks*, GeeksforGeeks, 6 Aug. 2025, www.geeksforgeeks.org/machine-learning/support-vector-regression-svr-using-linear-and-non-linear-kernels-in-scikit-learn/.

THANK YOU

SAT5141