



## PREDICTING COURSE SECTIONS REQUIRED FOR THE ACADEMIC SEMESTER

- GAYATHRI KRISHNAMOORTHY (STRATEGIC DATA ANALYST – ACADEMIC SCHEDULING ANALYTICS & PLANNING)
- DANIEL LE (EXECUTIVE MANAGING DIRECTOR OF ACADEMIC SCHEDULING ANALYTICS & PLANNING)

3/4/2025

1

1

# DANIEL LE



2

## Welcome to ASDA!

This site provides insight into the credit schedule build and related institutional processes that rely on the timely delivery of an accessible and equitable course schedule for students and instructors.

**Academic Scheduling and Data Analytics (ASDA)**, in collaboration with several Dallas College partners, develops scheduling management tools and processes to facilitate positive enrollment experiences for students and deliver relevant and reliable scheduling information to academic leaders and faculty to support their preparations for upcoming semesters. The result is a student-centric schedule that creates better outcomes for all stakeholders.

### Academic Scheduling Partners

• Academic Scheduling/Data Analytics (ASDA)	• Facilities Management
• Academic Leadership (Deans and Chairs)	• Human Resources
• Continuing Education (Workforce Advancement)	• Information Technology
• Curriculum Processing	• Special Programs ( <i>Emeritus, Honors, IncludED, YearUp, etc.</i> )
• Educational Partnerships	• Student Services

Together Academic Scheduling and its partners pursue the continued development and delivery of:

- A student-centric course schedule with the adaptability to pivot toward contingencies and the flexibility to accommodate situational anomalies.
- Forward-looking processes that are cognizant of and adaptable to approaching industry shifts and institutional evolutions.
- An environment of heightened responsiveness to both urgent and commonplace issues that routinely occur within interdependent processes.

3

3



# DANIEL LE



## Overview

- All Institutions strategically plans to increase the number of student enrollments by implementing course schedules that align with student demand.
- The number of sections needed for each course is important for the Academic Scheduling Department for offering right scheduling and allocating resources effectively.
- The outcome of this research will be beneficial for student success and potentially increase percentage of course enrollment.

4


A hand pointing at a circular graphic with the word "Agenda" in the center. The graphic contains various icons: a dollar sign, a person, a target, a bar chart, a gear, a building, and an envelope.

- Part I: Preliminary Study
- Part II: Creating an intuitive dashboard to monitor enrollment trend and make recommendations.

5

5

# DANIEL LE

A collection of colorful lightbulb icons, each containing a different symbol representing various fields: a magnifying glass, a classical building, a globe, a trophy, a target, a gear, a chess knight, and a shield with a key.

## PART I: PRELIMINARY STUDY

6

6



## Method

- Datasets includes 4,372 Unique Credit Courses offered at Dallas College for each Fall 2023 and Fall 2024. We excluded Dual Credit courses.
- Course Name, Modality, Academic Period, Section Count, Total seats used, Total section capacity, Total available seats, Average section utilization, Active student count and Enrollment Rate were collected.
- In this research project, multiple types of machine learning models are analyzed to identify significant factors in predicting course sections

7

# DANIEL LE



## Description of the Dataset

Variable	Description
<b>COURSE NAME</b>	Name of the Course offer at Dallas College
<b>MODALITY</b>	Mode of instruction or delivery of courses
<b>Sec Location Desc Cc</b>	Dallas College Campus Location
<b>Sec User16</b>	Academic Period
<b>FA24_Section_Count</b>	2024 Fall Course Section Count
<b>FA23_Section_Count</b>	2023 Fall Course Section Count
<b>FA23_Total_Seats_used</b>	2023 Fall Total number of seats used
<b>FA23_Total_section_capacity</b>	2023 Fall Total number of Section Capacity
<b>FA23_Total_Available_seats</b>	2023 Fall Total number of Available seats
<b>FA23_Avg_Section_Utilization</b>	2023 Fall Average percentage of the Section utilization
<b>FA23_Active_Student_Count</b>	2023 Fall number of active students
<b>FA23_Enrollment_Rate</b>	2023 Fall percentage of student students enrolled

8

# Exploratory Data Analysis

The objective of this section is to analyze and investigate the dataset to summarize their characteristics by employing various data visualization methods.




www.educba.com

9

# DANIELLE

## Descriptive Statistics of the Dataset

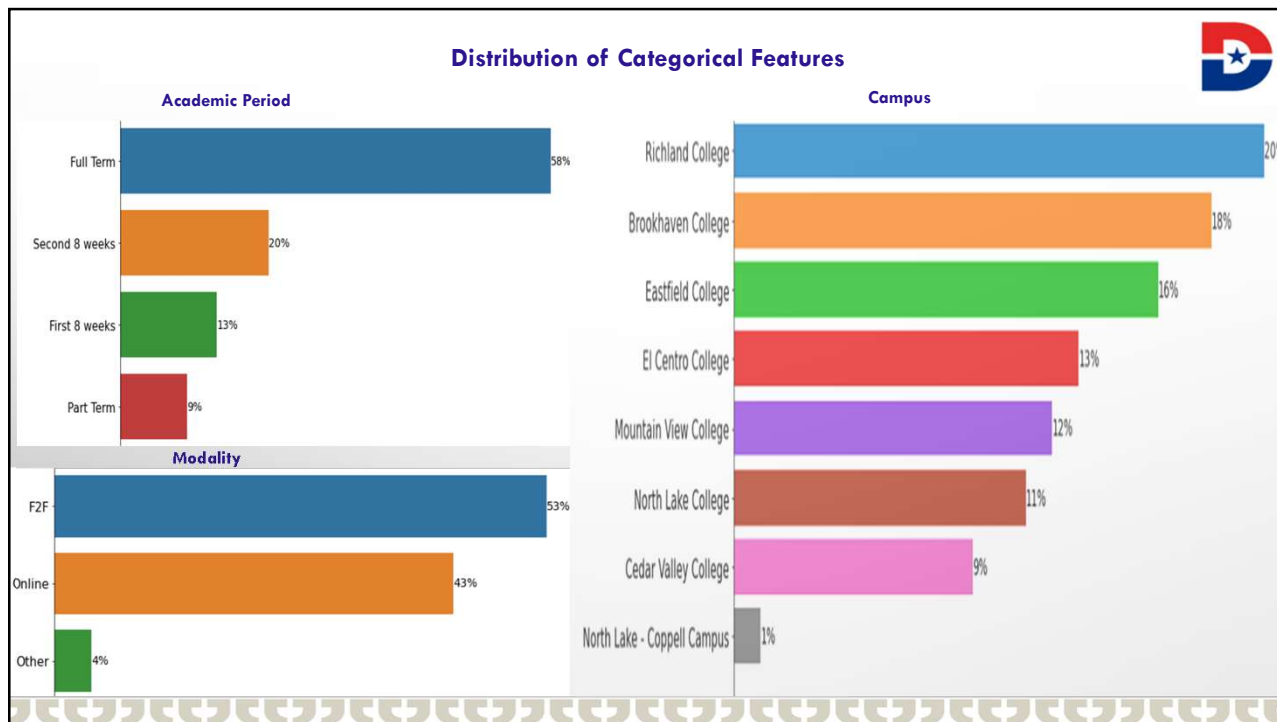


	count	mean	std	min	25%	50%	75%	max
FA24_Section_Count	4371.0	2.093571	3.116309	0.0	1.00	1.00	2.00	67.00
FA23_Section_Count	4371.0	1.834820	2.909652	0.0	1.00	1.00	2.00	66.00
FA23_Total_Seats_used	4371.0	34.105010	58.473626	0.0	6.00	19.00	37.00	1228.00
FA23_Total_section_capacity	4371.0	40.555937	63.113448	0.0	15.00	25.00	40.00	1268.00
FA23_Total_Available_seats	4371.0	9.881492	14.532805	-6.0	2.00	6.00	13.00	189.00
FA23_Avg_Section_Utilization	4371.0	58.070140	34.069455	0.0	32.00	70.00	86.00	126.67
FA23_Active_Student_Count	4371.0	30.750172	53.084694	0.0	6.00	17.00	33.00	1136.00
FA23_Enrollment_Rate	4371.0	75.020460	35.655797	0.0	76.47	90.91	98.25	100.00

- On a normal distribution, we expect to have approximately 68% of the values fall within one standard deviation of the mean; 95% within two standard deviations; and 99.7% within three standard deviations.
- The mean values of variables not same as the median values which are represented in the 50% (50<sup>th</sup> percentile) in the table. Hence, it shows the dataset do not have normal distribution.
- The non normal distribution is confirmed from the Standard deviation

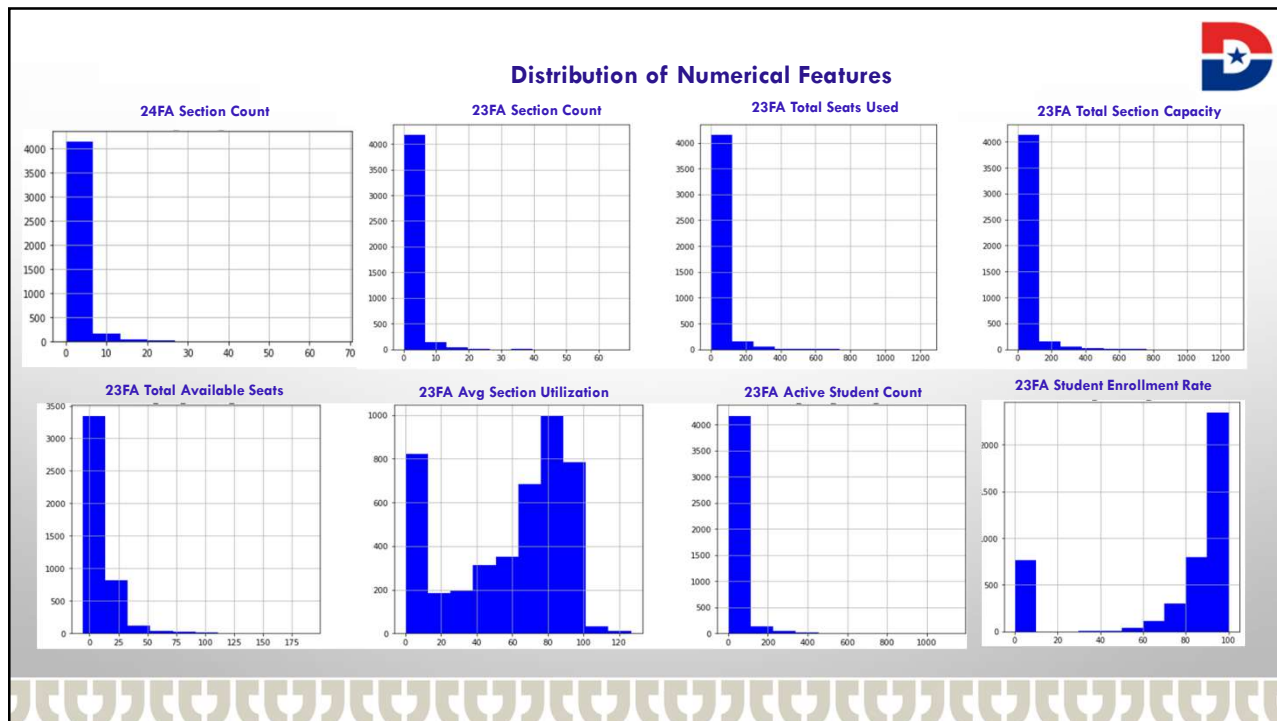
10



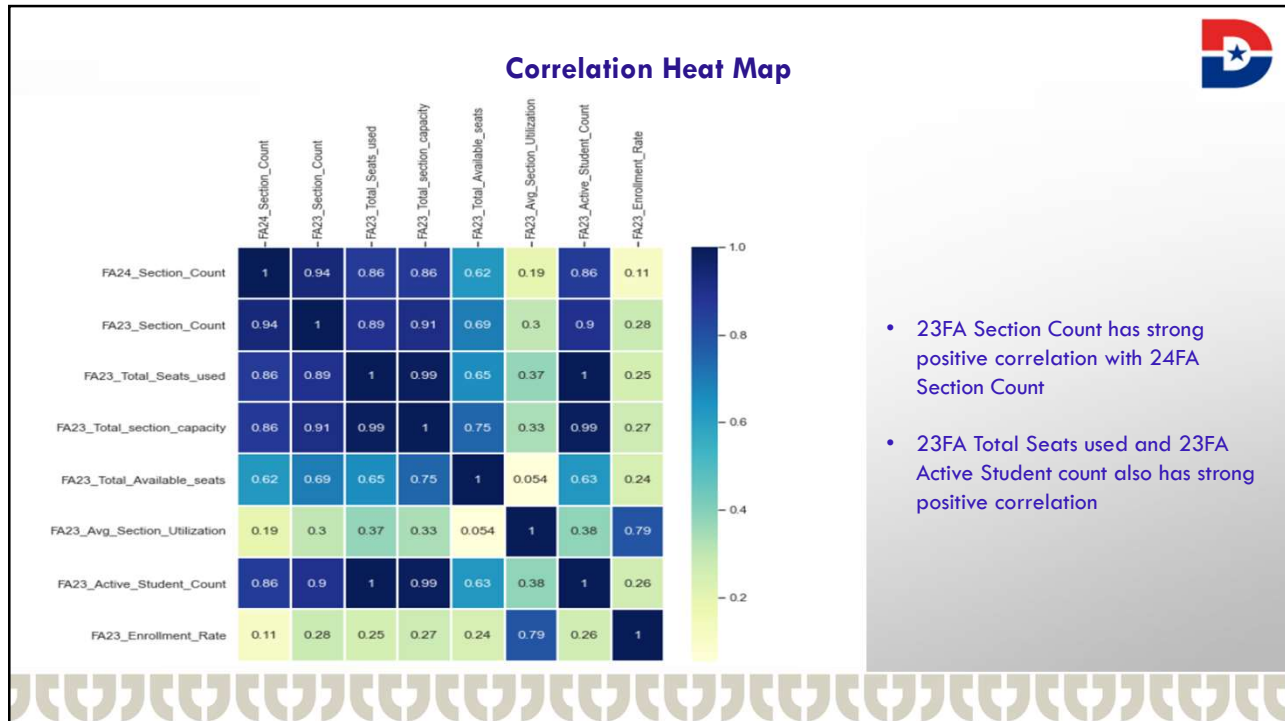


11

# DANIELLE





12



13

# DANIEL LE





## Modeling

---

The objective of this section is to build machine learning model to predict number of Course Sections required for 24 Fall Semester based on the 23 Fall Semester metrics

14

## Model Result – Predicting 24FA Section Count

Linear Regression Model Performance

- Developed multiple Ordinary Least Square models(OLS) models with various independent variable to predict 24FA Section Count.
- For developing the model, 80% of dataset is used for training and 20% for testing.
- The result of OLS Model with 23FA Section Count, 23FA Total Section Capacity and 23FA Active Student Count as independent variable is presented here.
- $R^2$  value of the model is 0.68 which shows that there is 68% of the variance in 24FA Section Count is explained by the independent variables.
- Even though the MAE and MSE is 0, residual plots violated Normal Distribution, and the theoretical quantile plot shows the Skewness after even log transformation of the dataset.

15

# DANIEL LE

## Model Result – Predicting 24FA Section Count

Linear Regression Model Performance

$R^2$	MAE	MSE
0.68	0	0

Actual Vs Predicted 24FA Section Count

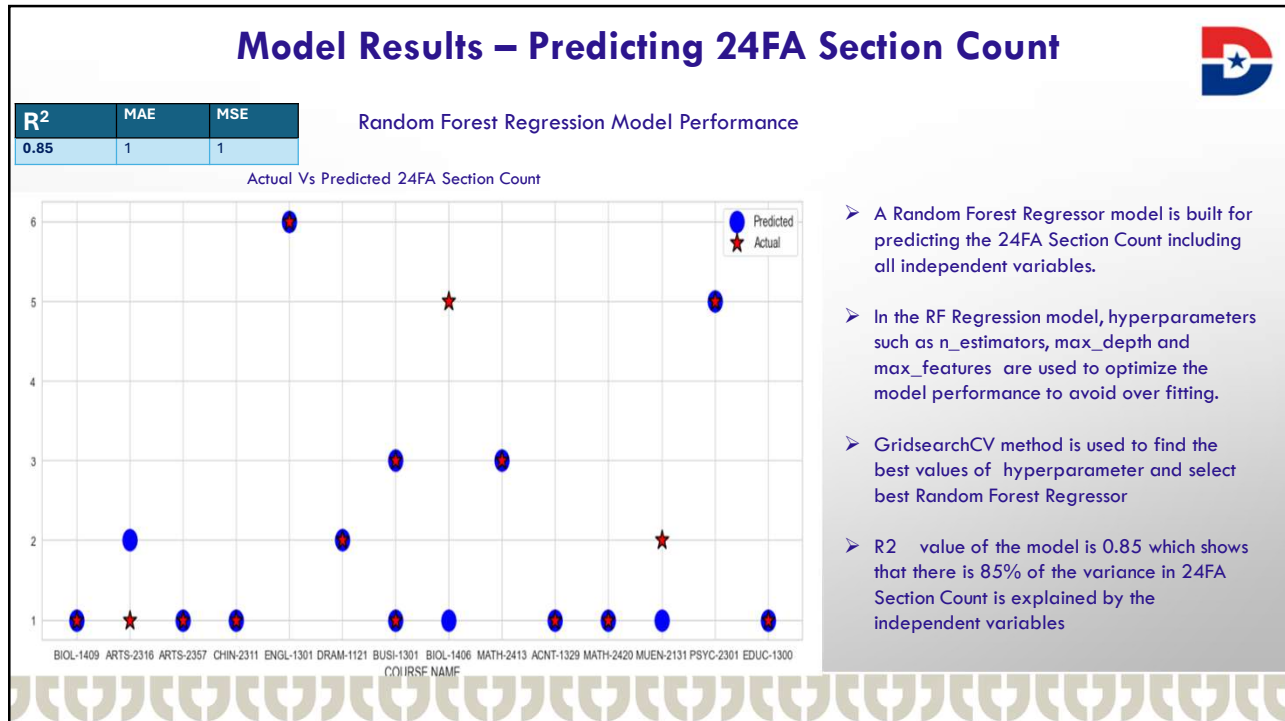
COURSE NAME

Plotting Residual Distribution

Theoretical Quantiles

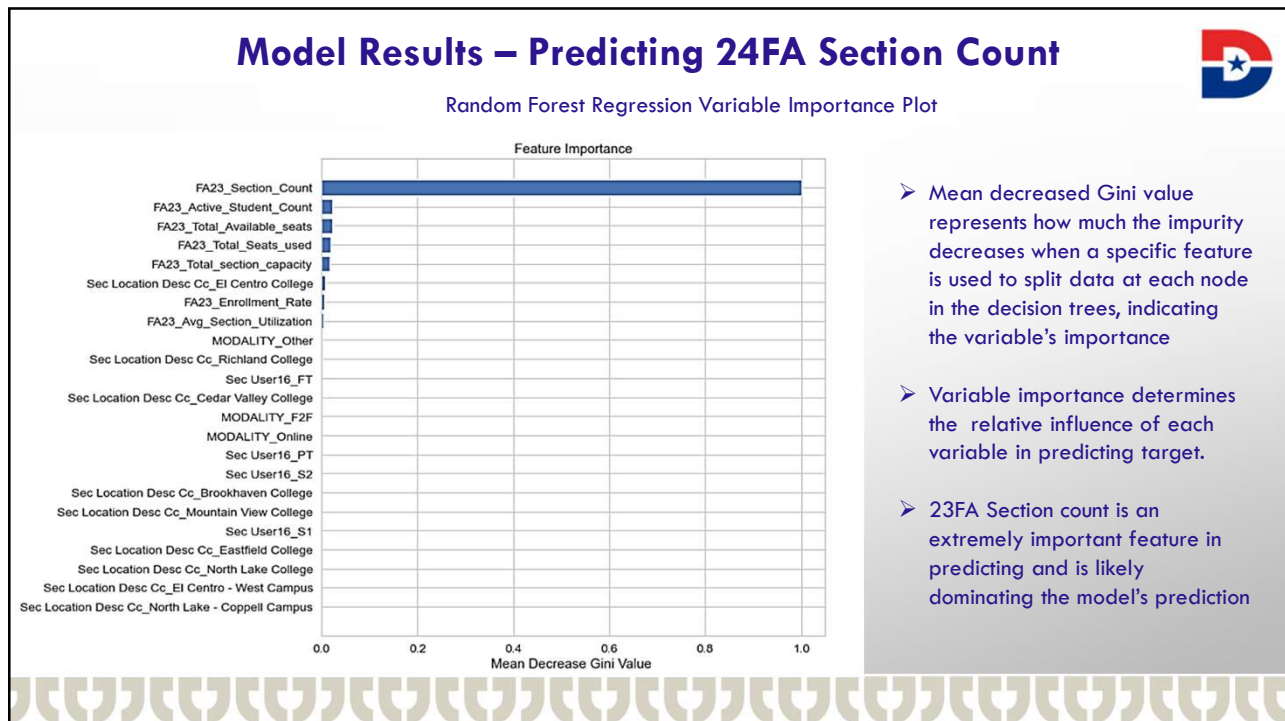
16





17

# DANIEL LE



18

## Model Results – Predicting 24FA Section Count



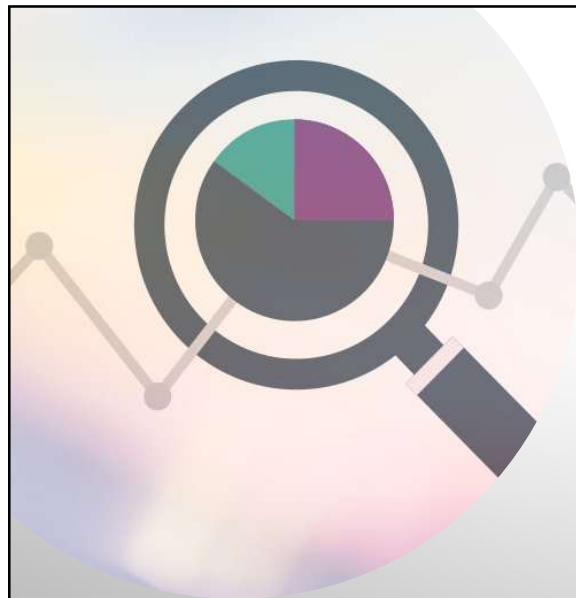
### Model Comparison

Model	R <sup>2</sup>	MAE	MSE
Linear Regression	0.68	0	0
Random Forest Regressor	0.85	1	1
Support Vector Regressor	0.80	0	1
KNN Regressor	0.40	1	1

- R<sup>2</sup> value and the MAE/MSE of Random Forest and Support vector Regressor models are almost similar.
- Random Forest Regressor performed well with both training and testing dataset compared to Support vector Regressor
- R<sup>2</sup> value for Linear Regression is .68 and the model indicate overfitting as training value is greater than testing value.

19

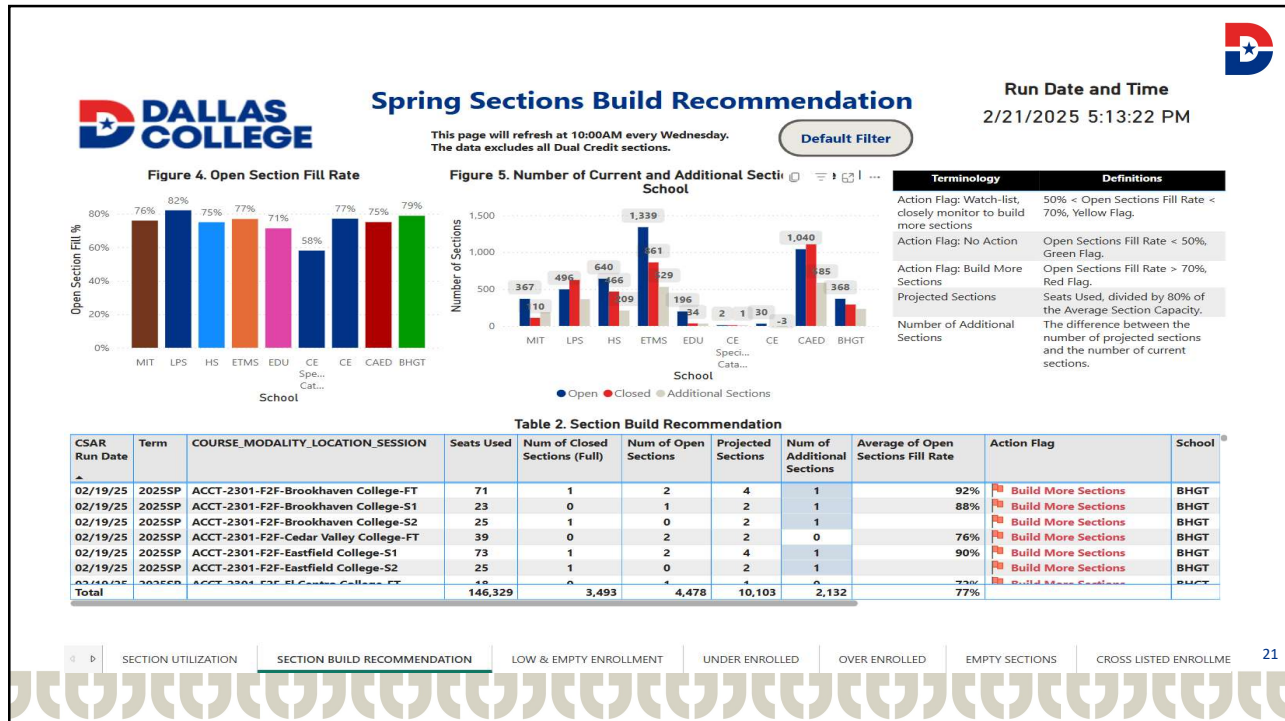
# DANIEL LE



- PART II: CREATING AN INTUITIVE DASHBOARD TO MONITOR ENROLLMENT TREND AND MAKE RECOMMENDATIONS.

20

20



21

# DANIEL LE



22



## Conclusion

- We have used several models to ascertain the main predictors as well as predict 24FA Section Count.
- Random forest can handle complex and non-linear dataset without any prior insights and, robust to outliers.
- Random Forest provides importance of individual features within each decision tree. According to Random Forest model, 23FA Section count is an extremely important variable in predicting 24FA Section Count.
- Random Forest and SVR models have similar results; however, SVR can be harder to interpret due to the kernel function.
- For future studies on Section Count prediction, we can include additional features such when timeline on when the section is opened, faculty teaching the course and information of students enrolling the course
- Also, we can use other machine learning model such as Xgboost to compare the prediction and feature importance.

23

# DANIEL LE



## Contact Info

- Feel free to send your question/suggestion/discussion to:

Gayathri Krishnamoorthy  
gkrishnamoorthy@dallascollege.edu

Daniel Le  
dle@DallasCollege.edu



24

24

Scan the QR code to complete the session survey.



Texas Association for Institutional Research  
Annual Conference: February 25-28, 2025  
Omni Hotel in Corpus Christi, TX





25

25

# DANIELLE

THANK YOU FOR LISTENING!



26

26