

DATA ENGINEERING FINAL PROJECT

# TWAP FOREX CURRENCY TRADING

PRESENTATION BY PRIYA KUNDU (Management of Technology)

# In this Presentation

## Overview



- 01** Storytelling of Project understanding
- 02** How did I choose the currency pairs for trading?
- 03** Using the chosen currency pairs for ML based trading
- 04** Closing statement for the Project
- 05** My takeaways from this course!

# Tried using the best practices suggested throughout the course.

# Best Practice #1: Completely Understanding the Project Goal

- 01 Collected raw historical data for five separate currency pairs
- 02 Cleaned, Organized and Scaled the data for convenient regression.
- 03 Performed regression using pycaret and classified all the pairs as "Forecastable", "Partially Forecastable" and "Non-Forecastable" to choose two strong currency pair to execute the trade.
- 04 Ran four execution windows to get an understanding of the decision we made to trade long and short for the selected pairs.
- 05 Added an extended code snippet to reverse the transaction loss in order to maximize the profit which wasn't attained during first four executions.

# How did I choose the currency pairs for trading?

Generating clean and organized csv files for regression data:

01

**Best Practice #2:  
Collecting all fields  
that are relevant.**

In my case, I took only  
VWAP and Liquidity

02

**Best Practice #3:  
Maintaining the  
consistency of field  
values**

Standardized and  
reformated the values  
to make it coherent.

03

**Best Practice #4:  
Dealing with missing  
values and wrong data**

Replaced and  
interpolated the  
missing values with  
sums and means as per  
the category.

04

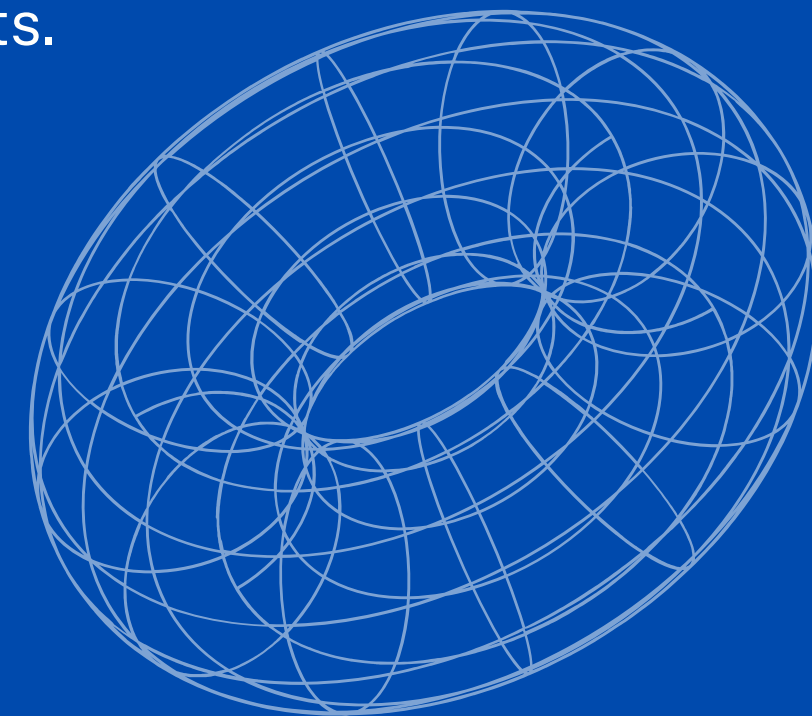
**Best Practice #5:  
Storing Large Scale  
Data**

Scaled down the whole  
dataset to hour level  
data points from  
minutes level data  
points.

Best Practice #6: Identifying Categorical Features with Numerical Values

Best Practice #7: Deciding on whether to Encode Categorical Features

Are not used in this project as there where no numeric values with categorical features involved in the dataset.



## Taking decisions while performing regression:

01

### **Best Practice #8: Deciding on whether to select features and if so, how to do it?**

This means choosing the correct input variable that will strengthen our predictive model. As per my understanding - this step was done by Professor himself, as he told us to consider VWAP value as a dependent variable for the regression.

02

### **Best Practice #9: Deciding on whether to reduce dimensionality, and if so, how to do it?**

The aim of this is to overcome the challenges by reducing the number of features while preserving as much information possible. One way of doing is by Feature Selection. This involves selecting a subset of original features based on their relevance to the target variables. Typically, this is done by correlation analysis, but in the case of this project, I did it on the basis of my fundamental understanding.

03

### **Best Practice #11: Saving, Loading and Reusing models**

By doing this, we can use the same models for similar predictions results for a new set of data.

Best Practice #10: Diagnosing and correcting over and underfitting the model is not used in this project as I used pycaret to perform the regression and it selected the best fit model for the further analysis.

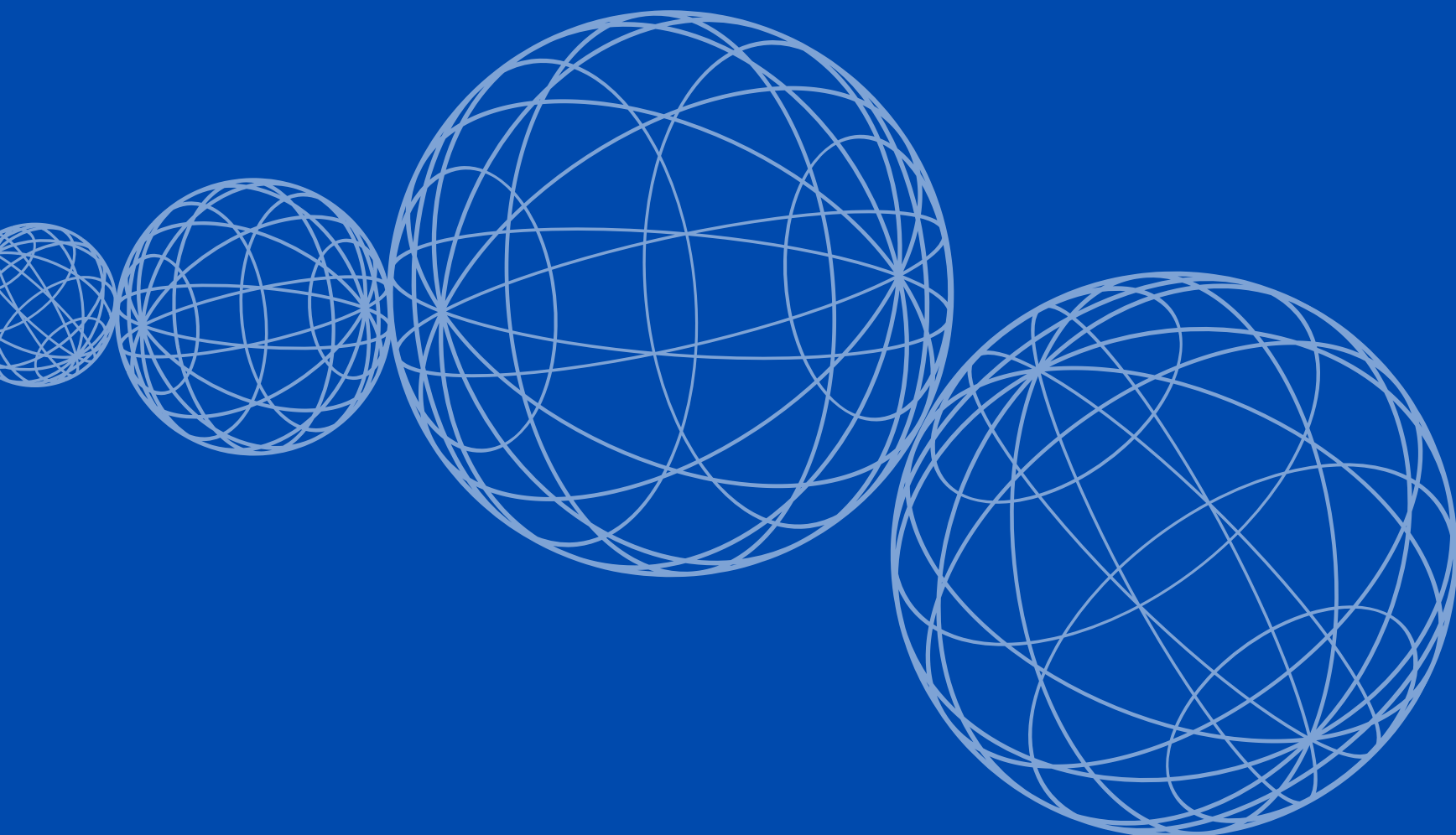


# Classifying the Currency Pairs:

01

## Best Practice #12: Deciding on whether to rescale features

In this, after getting the regression outputs, I used the multiclass SVM classifier method to separate my currency pairs as forecastable, non forecastable and partially forecastable where the RMSE acted as a hyperplane to classify



Currency Pair Classes

	Currency Pairs	Errors	Class
0	EURGBP	0.0008	Forecastable
1	USDCHF	0.001	Forecastable
2	USDMXN	0.0072	Partially Forecastable
3	EURJPY	0.0447	Non Forecastable
4	USDHUF	0.117	Non Forecastable

# Using the chosen currency pairs for ML based trading.



## Step 1

### Best Practice #13/14: Performing feature engineering without domain expertise

"Feature engineering" refers to the process of selecting, transforming, and extracting relevant features or characteristics from raw data to improve the performance of a machine learning model.

## Step 2

### Best Practice #15: Choosing the appropriate algorithm

I have used TWAP Trading Algorithm in this project so that there's no significant impact of my trading on forex market.

## Step 3

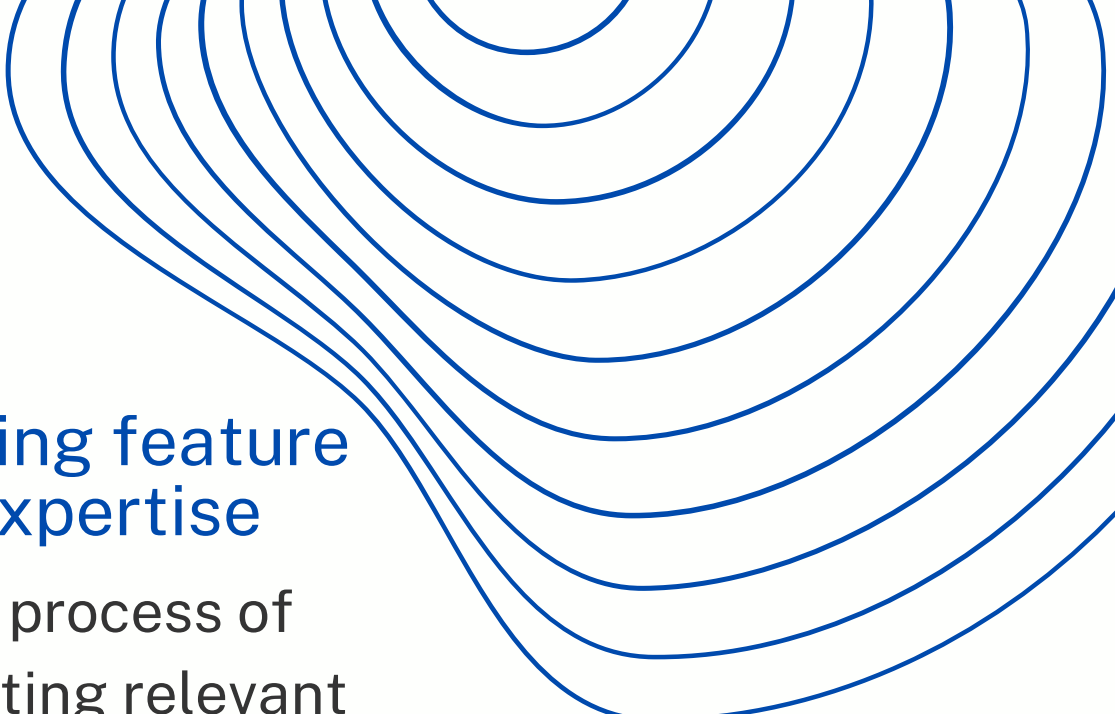
### Best Practice #16: Modeling on large-scale datasets

This was not required as per the project needs.

## Step 4

### Best Practice #17: Monitoring model performance

This was done throughout the run time. Additionally, specific print statements have been added in the script to avoid any overlooks.



# Closing statement for the Project

In this TWAP trading strategy, we traded a total of 100,000 USD for Long EURGBP and 100,000 USD for Short USDCHF

Final Project - Long - Execution Output Vectors

Interval	Executed Quantity	Percentage of Executed Quantity	Non Executed Quantity	Percentage of Non Executed Quantity	Average Execution Price
Interval 1	20000	100	0	0	0.8725595000000000
Interval 2	18000	60	12000	40	0.8725902631578950
Interval 3	3200	16	16800	84	0.87259825
Interval 4	0	0	30000	100	0.87259825

Final Project - Short - Execution Output Vectors

Interval	Executed Quantity	Percentage of Executed Quantity	Non Executed Quantity	Percentage of Non Executed Quantity	Average Execution Price
Interval 1	200000	100.0	0	0.0	0.8940465000000000
Interval 2	16000	53.33	14000	46.67	0.8938025
Interval 3	20000	100.0	0	0.0	0.8934060714285710
Interval 4	27000	90.0	3000	10.0	0.8934060714285710



---

# My takeaways from this course!

---

