ML-PROJECT

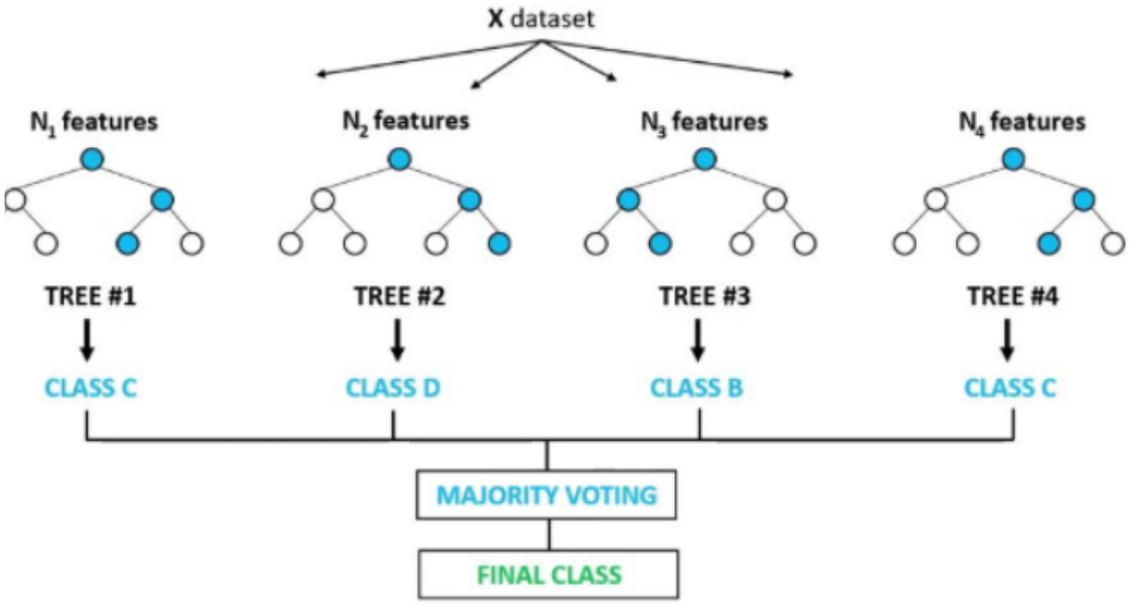
TRANSACTIONS OF TAXIS IN NYC

NAME: SNEHA T
REGISTER NUMBER: 21BDA30

INTRODUCTION

FOR A NEW YORK CITY TAXI DRIVER, BEING IN THE RIGHT PLACE AT THE RIGHT TIME IS OFTEN WHAT MAKES OR BREAKS A DAY.

TO ASSIST DRIVERS IN THIS DECISION, I EXPLORED WITH RANDOM FOREST CLASSIFIER TO FIGURE OUT WHICH WOULD BE THE BEST INPUT FEATURE TO PREDICT TOTAL AMOUNT GIVEN MANY INPUT FEATURES.



WORKFLOW

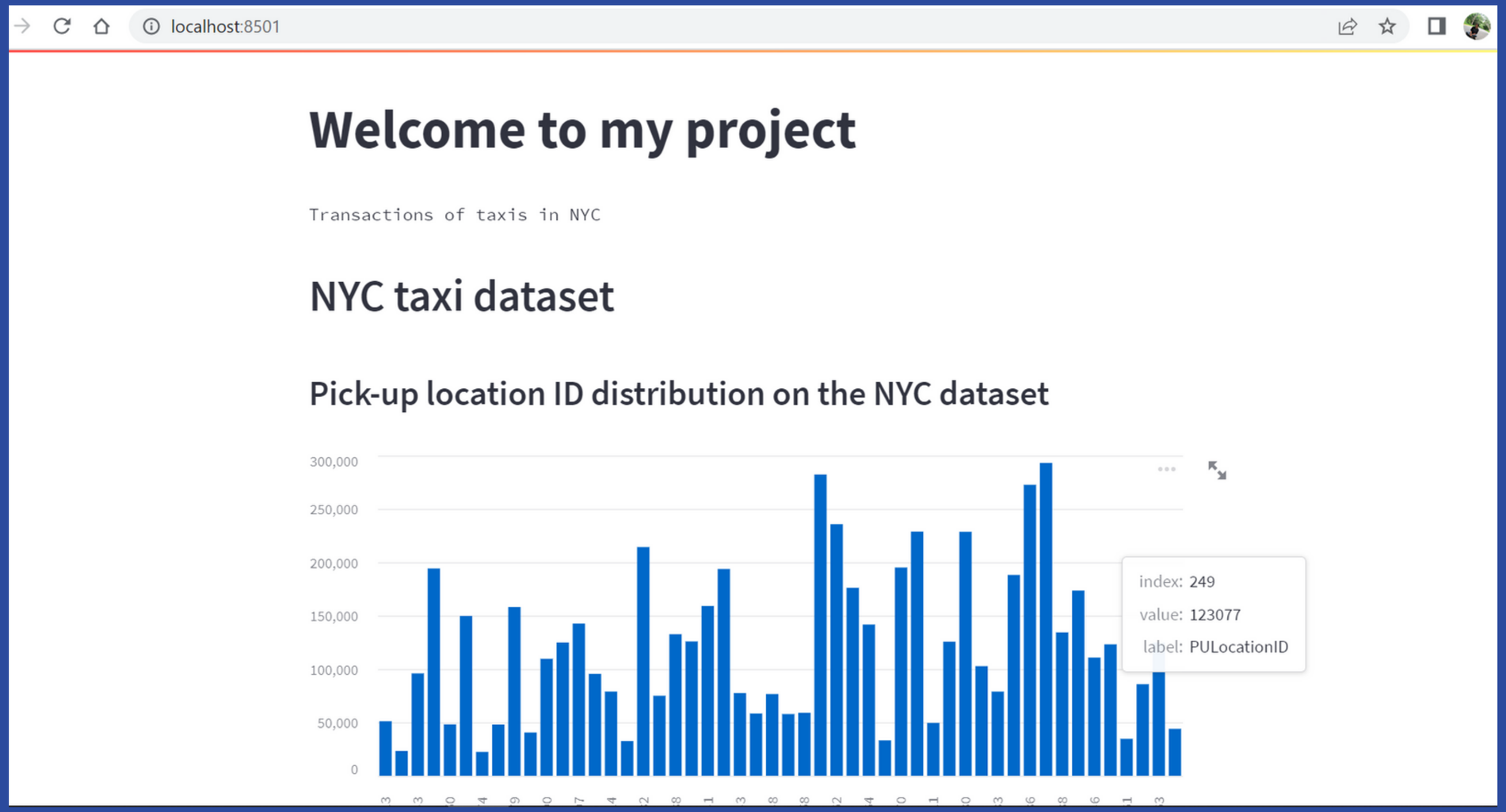
- DOWNLOAD AND INSTALL STREAMLIT
- DECIDE ON DESIGN
- BRING IN THE DATA
- COLLECT USER INPUT
- MODEL TRAINING
- INTERPRETING THE RESULTS OF OUR CLASSIFIER
- OPTIMIZE THE APP'S RUNTIME
- PERSONALIZE THE APP

VARIABLES IN THE DATASET

- VENDOR ID
- PICKUP DATE/TIME
- DROP DATE/TIME
- NO OF PASSENGERS

- TRIP DISTANCE
- PICKUP LOCATION ID
- DROP LOCATION ID
- FARE AMOUNT
- TIP AMOUNT
- TOTAL AMOUNT

WHAT DOES THE UI LOOK LIKE?



Let's train the model

e we can choose the hyperparameters of the model and see how the performance changes

What should the max_depth of the model be?

20

10 100

How many trees should there be?

300

Here is a list of input features in the d

	0
6	store_and_fwd_flag
7	PULocationID
8	DOLocationID
9	payment_type
10	fare_amount
11	extra
12	mta_tax
13	tip_amount
14	tolls_amount
15	improvement_surcharge

Which feature should be used as the input feature?

PULocationID

Mean absolute error of the model:

21.45808583738046

Mean sqauared error of the model:

726.4638499822195

R sqaured score of the model:

-2.3357730262997043

FUTURE SCOPE

-THE APP IS ONLY RESTRICTED TO FINDING THE BEST INPUT PARAMETR TO PREDICT OUR TARGET VARIABLE. WE CAN EXTEND THE IDEA TO MAKE ACTUAL PREDICTIONS

-DEPLOY THE APP

THANKYOU