NYC TAXI TRIP FARE PREDICTIO

16_Harshpratap Singh
17_Akhila Anilkumar
13_Aman Sheikh

Problem Statement:

To Develop a robust and accurate taxi trip fare prediction system that can benefit both passengers and taxi drivers. By leveraging historical trip data and real-time information, we seek to create a model that can estimate the fare for a taxi journey with a high degree of precision. This
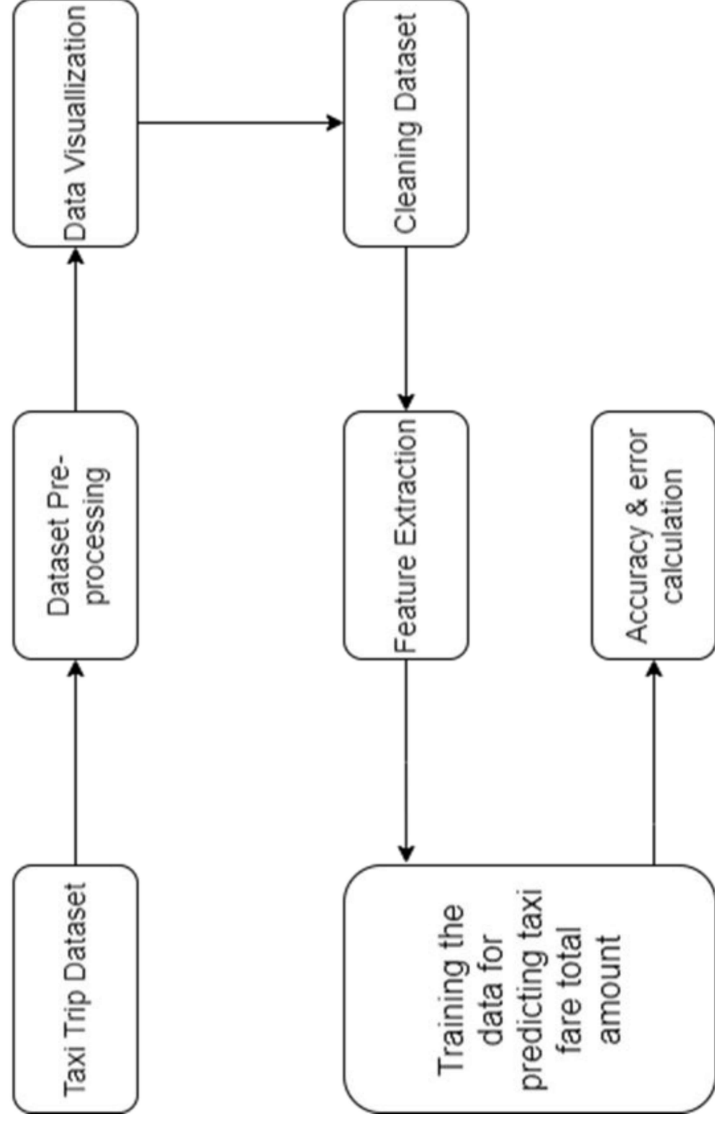
The Dataset

Field Name	Description
VendorID	A code indicating the TPEP provider that provided the record. **1= Creative Mobile Technologies, LLC; 2= VeriFone Inc.**
tpep_pickup_datetime	The date and time when the meter was engaged.
tpep_dropoff_datetime	The date and time when the meter was disengaged.
Passenger_count	The number of passengers in the vehicle. This is a driver-entered value.
Trip_distance	The elapsed trip distance in miles reported by the taximeter.
PULocationID	TLC Taxi Zone in which the taximeter was engaged
DOLocationID	TLC Taxi Zone in which the taximeter was disengaged
RateCodeID	The final rate code in effect at the end of the trip. **1= Standard rate** **2=JFK** **3=Newark** **4=Nassau or Westchester** **5=Negotiated fare** **6=Group ride**

The Dataset

Store_and_fwd_flag	This flag indicates whether the trip record was held in vehicle memory before sending to the vendor, aka "store and forward," because the vehicle did not have a connection to the server. **Y= store and forward trip** **N= not a store and forward trip**
Payment_type	A numeric code signifying how the passenger paid for the trip. **1= Credit card** **2= Cash** **3= No charge** **4= Dispute** **5= Unknown** **6= Voided trip**
Fare_amount	The time-and-distance fare calculated by the meter.
Extra	Miscellaneous extras and surcharges. Currently, this only includes the $0.50 and $1 rush hour and overnight charges.
MTA_tax	$0.50 MTA tax that is automatically triggered based on the metered rate in use.
Improvement_surcharge	$0.30 improvement surcharge assessed trips at the flag drop. The improvement surcharge began being levied in 2015.
Tip_amount	Tip amount – This field is automatically populated for credit card tips. Cash tips are not included.
Tolls_amount	Total amount of all tolls paid in trip.
Total_amount	The total amount charged to passengers. Does not include cash tips.
Congestion_Surcharge	Total amount collected in trip for NYS congestion surcharge.
Airport_fee	$1.25 for pick up only at LaGuardia and John F. Kennedy Airports

PROCESS DESIGN

```
Taxi Trip Dataset  →  Dataset Pre-
                       processing      →  Data Visuallization
                                                  │
                                                  ↓
                                           Cleaning Dataset
                                                  │
                                                  ↓
Training the                                Feature Extraction
data for
predicting taxi   →
fare total
amount
      │
      ↓
Accuracy & error
calculation
```

PROCESS DESIGN

Taxi Trip Dataset

Dataset Pre-processing

Data Visuallization

Cleaning Dataset

Feature Extraction

Training the data for predicting taxi fare total amount

Accuracy & error calculation

IMPLEMENTATION