Mid-Program Project - 1

PG Program in AI ML – NITW

TABLE OF CONTENTS

BACKGROUND

Cab booking system is the process where renting a cab is automated through an app throughout a city. Using this app, people can book a cab from one location to another location. Being a cab booking app company, exploiting the understanding of cab supply and demand could increase the efficiency of their service and enhance user experience by minimizing waiting time.

Objective of this project is to combine historical usage pattern along with the open data sources like weather data to forecast cab booking demand in a city.

PROCESS FLOW

You will be provided with hourly renting data span of two years. Data is randomly divided into **train** and **test** set. You must predict the total count of cabs booked in each hour covered by the **test** set, using the information available prior to the booking period. You need to append the **train_label** dataset to train.csv as 'Total_booking' column.

Please find the descriptions of the columns present in the dataset as below.

datetime - hourly date + timestamp

season - spring, summer, autumn, winter

holiday - whether the day is considered a holiday

workingday - whether the day is neither a weekend nor holiday

weather - Clear , Cloudy, Light Rain, Heavy temp - temperature in Celsius

atemp - "feels like" temperature in Celsius

humidity - relative humidity

windspeed - wind speed

Total_booking - number of total booking

DATASET

The recommended datasets will be shared. You can download them from the LMS.

TARGET ENVIRONMENT

You can use Edureka's CloudLab, a cloud-based Jupyter Notebook, which is pre-installed with Python and other required packages to work on this Project. It is offered by Edureka as a part of the course, where you can execute all the demos and work on the projects hassle-free.

TASKS

Following are the tasks, which need to be developed while executing the project:

Task 1:

1. Visualize data using different visualizations to generate interesting insights.
2. Outlier Analysis
3. Missing value analysis
4. Visualizing Total_booking Vs other features to generate insights
5. Correlation Analysis

Task 2:

1. Feature Engineering
2. Grid search
3. Regression Analysis
4. Ensemble Model