Share-Bike Demad Predictio

Detailed Project Report

WRITTEN BY SADASHIV NANDANIKAR

Objective

With over 500 programs and 500 thousand bicycles worldwide, bike-sharing systems are becoming increasingly important for traffic, environmental and health issues. The aim is to create the best performing hyper-tuned machine learning model to predict the number of bike rentals based on various features such as time, location, weather, and holidays,

The developed model can be utilized to gain insights into the bike-sharing system's usage patterns and its impact on traffic, environmental, and health issues. Additionally, the model can help in identifying significant events in the city by monitoring the data generated by the bike-sharing system.

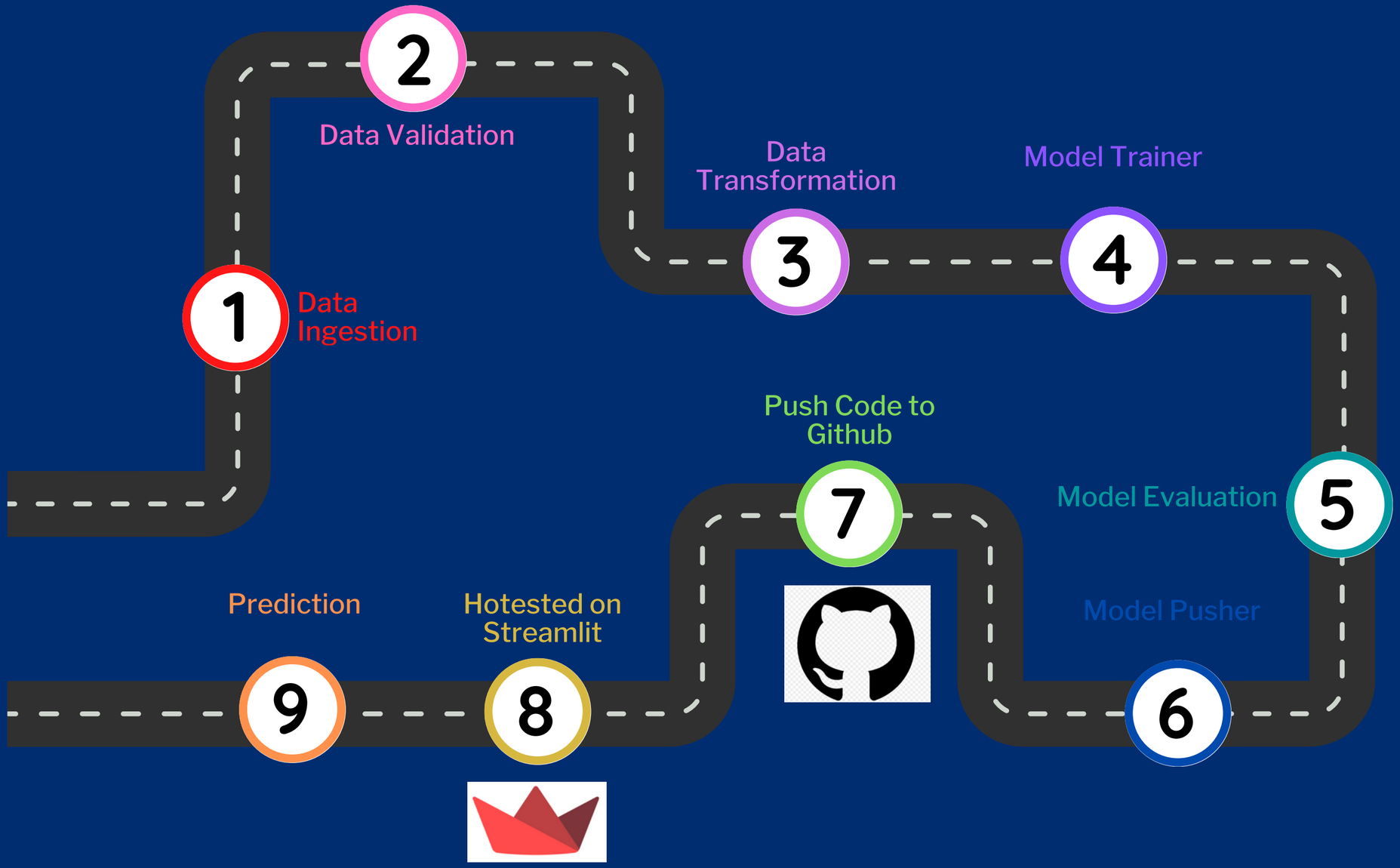
Benifits

- Accurate demand prediction improves bike distribution and user experience.
- Usage pattern identification leads to efficient resource allocation and lower operational costs.
- Precise demand prediction enables bike-sharing system operators to optimize bike distribution and improve user experience.

Data Sharing Agreement

- **Number of columns**: 17
- **Name of the columns** : instant, dteday, season, yr, mnth, hr , holiday , weekday , workingday , weathersit , temp , atemp, hum, windspeed, casual, registered, cnt
- **Data type of the columns** : 17 numerical features,

Architecture



DATA INGESTION

- Loading the data from Mongodb Database
- Dividing the dataset into training and testing sets.
- Storing both the dataset and the train-test split into the DataIngestion Artifact.

DATA VALIDATION

- Ensure that all columns are included in the dataset.
- Verify if there is any data drift between the base and current dataset.
- Store the report file (.yaml) within the Data Validation Artifact.

DATA TRANSFORMATION

- Eliminating outliers from the dataset.
- Converting categorical features into numerical values through encoding.
- Rescaling numerical features.
- Saving the transformer object, the transformed training and testing dataset into the Datatransformation artifact.

MODEL TRAINING

- The model is trained with the lightgbm regressor algorithm.
- The resulting model is saved within the ModelTrainerArtifacts.

MODEL EVALUATION

- The current model undergoes an evaluation process against the previous model, if one exists. If no prior model exists, the current model is stored within the "saved models" directory.
- The evaluation process is conducted between the current model and the prior model, if one exists. The optimal model is then saved within the "saved models" directory.

MODEL PUSHER

- Storing the most recent transformer object and model into the "saved models" directory.

PREDICTION

- To generate predictions, the Streamlit application requires user inputs, which can then be displayed on the screen as output.

Conclusion

Successfully completed the end-to-end machine learning share-bike demand project. The project's predictive model can help to forecast the count of rental bikes and identify usage patterns, resulting in efficient resource allocation and reduced operational costs. With this demand prediction capability, bike-sharing system operators can optimize bike distribution and enhance the user experience.

Thank You

SLIDE PRESENTATIONS DESIGN