Capstone Session 5

Machine Learning for Modeling

Session 5: Dataset Description

FloridaBikeRentals.csv

Variable	Description	Variable	Description
Date	Date in year-month-day format	Windspeed	Speed of the wind in meters per sec (m/s)
Rented Bike count	Count of bikes rented each hour	Visibility	Visibility in meters
Hour	Hour of the day	Dew point temperature	Dew point temperature in Celsius
Temperature	Temperature in Celsius	Solar radiation	Radiant energy from the sun measured in MJ/m^2
Humidity	Humidity in percentage	Rainfall	Rainfall in mm
Snowfall	Snowfall in cm	Seasons	Season of the year – Winter, Summer, Spring, and Fall
Holiday	Is it a holiday? Holiday/ No holiday	Functional Day	Bike rented during functional (Fun) and nonfunctional hours (NoFunc)

Session 5: Predicting Bike Rental Demand

Task: Based on rented bike count, the hour of the day, the day's temperature, humidity, wind speed, rainfall, holidays, and many other factors, build a model to predict the bike count required for a stable supply of rental bikes.

- Load the dataset
  (Hint: use `encoding = "ISO-8859-1"` as a parameter in the `read_csv` function).

- Check for null values in any columns and handle the missing values

- Convert Date columns to Date format and extract day, month, day of week, and weekdays/ weekend from date column

- Check correlation of features using heatmap

- Plot the distribution plot of Rented Bike Count

- Plot the histogram of all numerical features

- Plot the box plot of Rented Bike Count against all the categorical features (Hint: Categorical features on X-axis and Rented Bike Count on Y-axis)

Session 5: Predicting Bike Rental Demand

- Plot the Seaborn catplot of Rented Bike Count against features like Hour, Holiday, weekdays, weekend, and seasons. Give your inferences.

- Build a violin plot for Seasons and Rainfall(mm). Provide your observation.

- Encode the categorical features into numerical features. (Hint: use get_dummies())

- Identify the target variable and split the dataset into train and test with a ratio of 80:20 and random state 1

- Perform Standard Scaling of the train dataset.

- Machine Learning:

  - Perform Linear Regression, Lasso Regression, and Ridge Regression to predict the bike count required and compare the results.

  - Run your evaluation using 2 regression metrics and provide your observations about the models above

  - **Bonus:** Build a Decision Tree model and evaluate it using cross-validation with the same metrics. Choose a value for max_depth and min_sample_split. Hint: you need to perform the cross-validation twice.

Thank You