ML-LAB-EXP-3 Date:5/08/2024

Lab Question (100 Marks)

Title: Predicting Fish Price Using Linear Regression and Data Analysis

Objective:

To perform various data analysis techniques, including handling missing values and building a linear regression model to predict the fish price based on their physical attributes.

Dataset:

You are provided with a dataset containing measurements of different fish species. The dataset includes the following columns:

- Price: The price of the fish in rupees (numerical).

- Weight: The weight of the fish in grams (numerical).

- Length: Vertical length in cm (numerical).

- Height: The height of the fish in cm (numerical).

- Width: The width of the fish in cm (numerical).

Tasks:

Use proper comments whenever required

1. Data Preprocessing and Missing Value Analysis (30 Marks)

- Check for missing values in the dataset and handle them without dropping any values appropriately. Provide a detailed explanation of the methods used for handling missing values.

- Normalize the dataset in the range 0 to 1 if necessary. Explain why the chosen method is appropriate for the given dataset.

2. Exploratory Data Analysis (EDA) (20 Marks)

- Provide a summary of the dataset, including the distribution of numerical features and the count of each species.

- Use visualizations (such as histograms, box plots, or scatter plots) to explore the relationships between different features and species.

- Discuss any patterns or trends observed from the visualizations.

3. Feature Selection and Engineering (10 Marks)

- Display the correlation among all features

4. Linear Regression Model (30 Marks)

- Build a linear regression model based on all the features to predict the fish price.

- Split the data into training and testing sets, ensuring an appropriate split ratio.

- Train the model on the training set and evaluate its performance on the testing set. Report the evaluation metrics (such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and R-squared (Coefficient of Determination)).

5. Discussion and Conclusion (10 Marks)

- Write a simple python function to do the following

-Discuss the performance of all the model. How well does the linear regression model predict the fish price?

- Identify any limitations of the model and suggest possible improvements.

- Reflect on the overall workflow and any challenges encountered during the lab.

Submission Requirements:

- Submit a Jupyter notebook with the complete code and analysis.

- Include detailed explanations for each step and any assumptions made.

- All visualizations should be appropriately labeled and described.

- The final notebook should be well-organized and formatted for readability.

Evaluation Criteria:

- Correctness and completeness of the analysis and model.

- Clarity of explanations and justifications.

- Quality and relevance of visualizations.

- Overall presentation and organization of the notebook.

Dr.Trilok Nath Pandey

SCOPE,VIT,Chennai