Data Warehouse and Mining

Mini project report

Regression algorithm

Made by:

- **Surya Pratap – 8367**
- **Rachel Jose – 8361**
- **Nolita Rego – 8363**

Under the guidance of: **Mrs. Sujata Deshmukh**

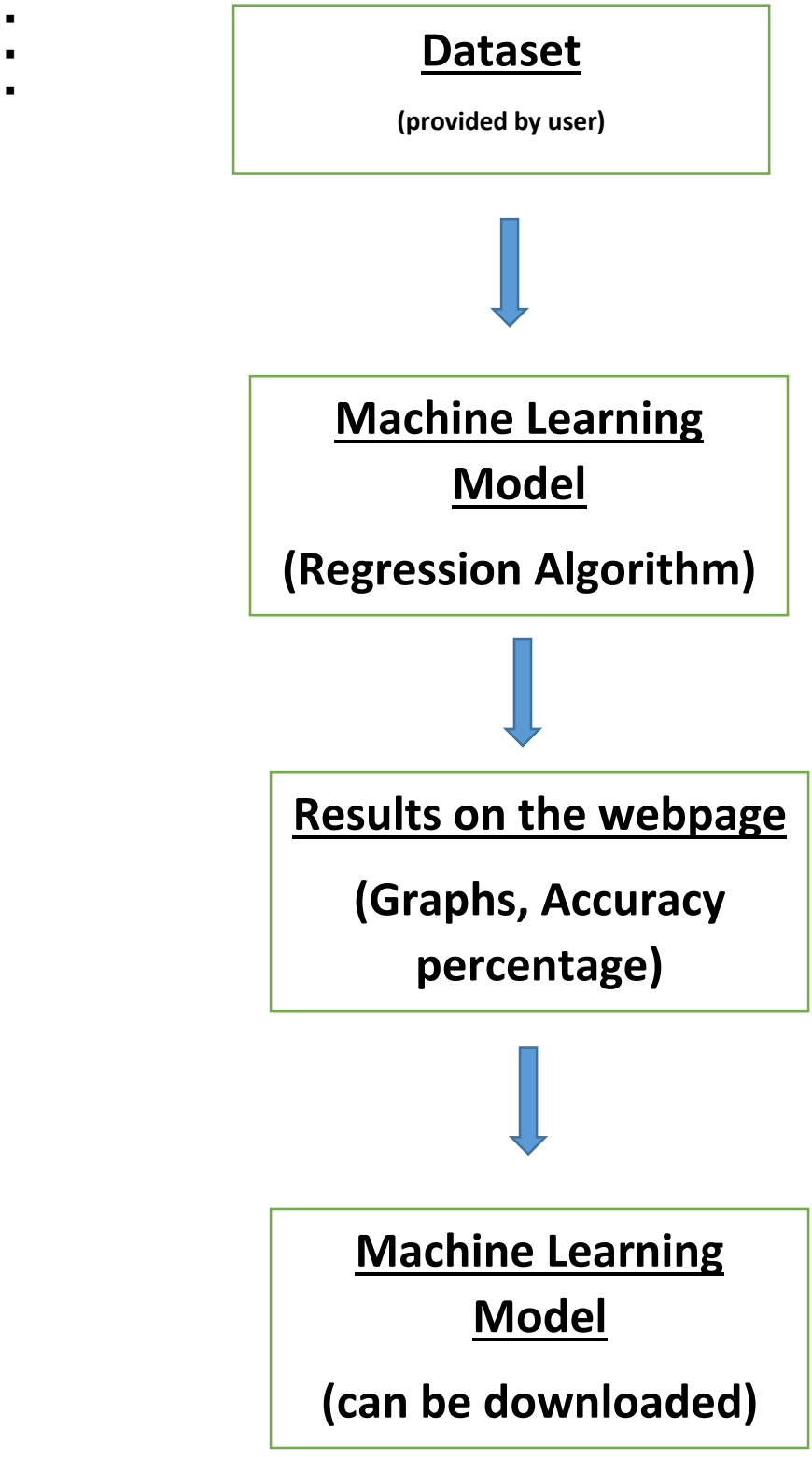
- **Problem Definition: -**

  The system takes in any dataset (for regression model) and evaluated the dataset of Linear Regression model after analyzing the dataset and displays the resulting graphs and other details. And finally gives the accuracy percentage of the model on that particular dataset.

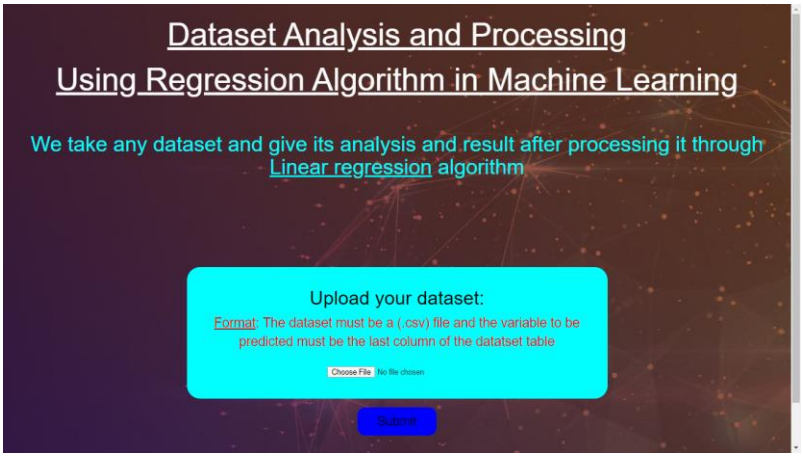
- **Requirements: -**

  - **Dataset** – The dataset must be .csv file and the output variable to be predicted must be at the last column of the dataset table.
  - **Python libraries** – pandas, numpy, matplotlib, seaborn, scikit-learn, flask, werkzeug, os, pathlib, joblib and basic libraries.

▪ Flow Diagram: -



- **Screen Shots: -**



Dataset Analysis and Processing
Using Regression Algorithm in Machine Learning

Results

Statistical Values for the model

Accuracy of the model: 83.10500898073398

Mean Absolute Error: 0.6159600772681728

Mean Squared Error : 0.7465379364740276

Root Mean Squared Error: 0.7465379364740276

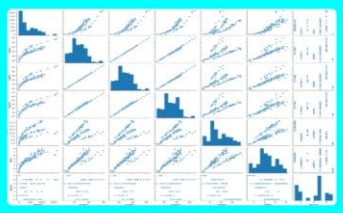
Dataset analysis

Weights Table

	Coefficient
Weight	0.002745055117725262
Length1	0.4641466994687056
Length2	1.0360836437251668
Length3	-1.350983348652849
Height	0.027444602879829 81

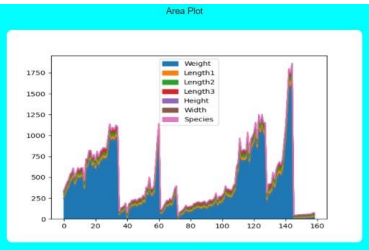
	Coefficient
Weight	0.002745055117725262
Length1	0.4641466994687056
Length2	1.0360836437251668
Length3	-1.350983348652849
Height	0.027444602879829 81
Width	-0.913909228744412

Pair Plots

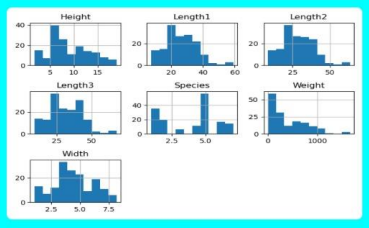


Area Plot

Area Plot



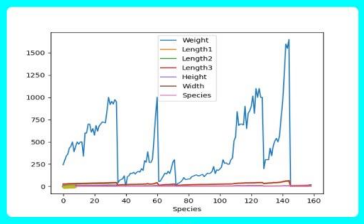
Histogram



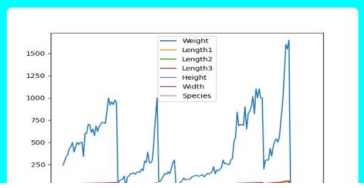
Line Plot

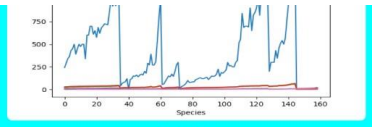
Results of Linear Regression

Regression Plot

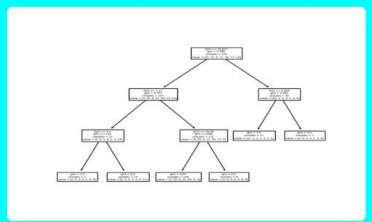


Line Plot



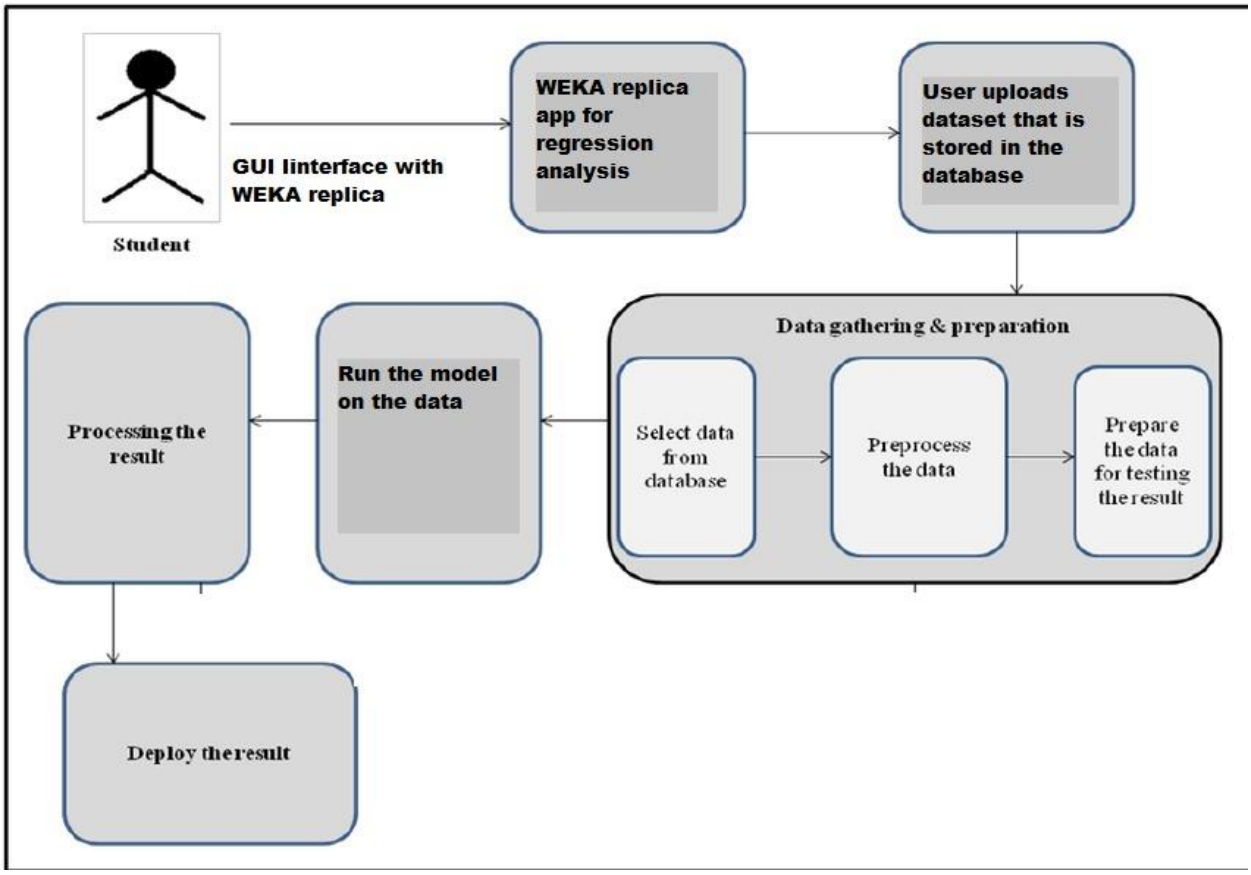


Decision Tree



Download your model file below

- **Architecture Diagram:**



- **<u>Code</u>: Included in the zip file**

- **<u>Conclusion</u>**: We have successfully made a linear regression model that takes in any dataset and displays the result in the form of graphs, plots and performance in terms of accuracy and performance of the data. The results displayed are statistical analysis, dataset analysis and results of linear regression which can be saved.

Thank you!!