ARUNAI ENGINEERING COLLEGE

COLLEGE CODE: 5104

**DATA ANALYTICS WITH CONGOS**

**PRESENTED BY : M.VIJAYASARATHI**

**REGISTER NO : 510421205059**

**DEPARTMENT : INFORMATION TECHNOLOGY**

**DATA EXPLORATION & VISUALIZATION**

**INTRODUCTION:**

* **Data visualization in data exploration leverages familiar visual cues such as shapes, dimensions, colors, lines, points.**
* **Angles so that data analysts can effectively visualize and define the metadata, and then perform data cleansing.**

**DEFINITION DATA EXPLORATION:**

* **Data exploration is an approach similar to initial data analysis, whereby a data analyst uses visual exploration to understand what is in a dataset and the characteristics of the data, rather than through traditional data management systems.**
* **These characteristics can include size or amount of data, completeness of the data, correctness of the data, possible relationships amongst data elements or files/tables in the data.**

**DEFINITION OF DATA VISUALIZATION:**

* **Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data.**
* **Additionally, it provides an excellent way for employees or business owners to present data to non-technical audiences without confusion.**
* **This is called as Data Visualization.**

**TOOLS:**

* **“ Data Exploration “ tools make data analysis easier to present and understand through interactive, visual elements, making it easier to share and communicate key insights.**
* **Data exploration tools include data visualization software and business intelligence platforms, such as Microsoft Power BI, Qlik and Tableau.**
* **“ Data Virtualization “ is an approach to integrating data from multiple sources of different types into a holistic, logical view without moving it physically.**
* **In simple terms, data remains in original sources while users can access and analyze it virtually via special middleware.**

**WORKING:**

* **Data visualization is the practice of translating information into a visual context, such as a map or graph, to make data easier for the human brain to understand and pull insights from.**
* **The main goal of data visualization is to make it easier to identify patterns, trends and outliers in large data sets.**
* **Data Visualization tools include Google Charts, Tableau, Grafana, Chartist, FusionCharts, Datawrapper, Infogram, and ChartBlocks etc.**
* **These tools support a variety of visual styles, be simple and easy to use, and be capable of handling a large volume of data.**

**SUPERVISED LEARNING-REGRESSION :**

* **Regression is a supervised machine learning technique which is used to predict continuous values.**
* **The ultimate goal of the regression algorithm is to plot a best-fit line or a curve between the data.**

**TOOLS**:

* ***Regression is a supervised machine learning technique which is used to predict continuous values.***
* ***The ultimate goal of the regression algorithm is to plot a best-fit line or a curve between the data.***
* ***The three main metrics that are used for evaluating the trained regression model are variance, bias and error.***

**REGRESSION EXAMPLE:**

1. **Linear Regression.**
2. **Ridge Regression.**

**3) Neural Network Regression.**

**4) Lasso Regression.**

**5) Decision Tree Regression.**

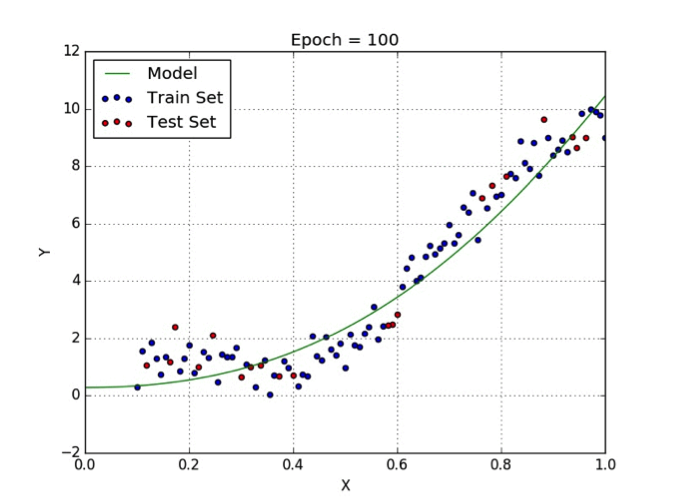
**6) Random Forest.**

**7) KNN Model.**

**8) Support Vector Machines (SVM)**

**SUPERVISED LEARNING-REGRESSION WORKING:**

* **Regression is a supervised machine learning technique which is used to predict continuous values.**
* **The ultimate goal of the regression algorithm is to plot a best-fit line or a curve between the data.**
* **The three main metrics that are used for evaluating the trained regression model are variance, bias and error.**

****

* **Supervised learning, also known as supervised machine learning, is a subcategory of machine learning and artificial intelligence.**
* **It is defined by its use of labeled datasets to train algorithms that to classify data or predict outcomes accurately.**

**CONCLUSION:**

* **Supervised learning is the most commonly utilized machine learning algorithm, as it is easy to understand and use. The model helps form accurate results using labeled information and variables as inputs.**
* **Regression is a supervised machine learning technique which is used to predict continuous values. The ultimate goal of the regression algorithm is to plot a best-fit line or a curve between the data.**