# COMPARISON OF DIFFERENT MACHINE LEARNING ALGORITHMS

TEAM : EAST BU

AKASH, BHAVANA, KARTHIK MADHUHASA

# COMPARISON TABLE - TYPE

| Algorithm | Type |
|---|---|
| Linear Regression | Supervised Learning |
| Logistic Regression | Supervised Learning |
| SVM | Supervised Learning |
| K-means | Unsupervised Learning |
| Naïve Bayes | Supervised Learning |
| Decision Tree | Supervised Learning |
| Random Forest | Supervised Learning |

Supervised Learning : A **supervised learning** algorithm analyzes the **training** data and produces an inferred function, which can be used for mapping new examples

Unsupervised Learning : **Unsupervised learning** is a type of **machine learning** algorithm used to draw inferences from datasets consisting of input data without labeled responses

# COMPARISON TABLE - APPLICATION

| Algorithm | Application |
|---|---|
| Linear Regression | Regression |
| Logistic Regression | Classification |
| SVM | Both Classification and Regression |
| K-means | Clustering |
| Naïve Bayes | Both Classification and Regression |
| Decision Tree | Both Classification and Regression |
| Random Forest | Both Classification and Regression |

Regression : The output is usually a numerical value.
Eg: Predicting height of a person given weight and age

Classification : The output is the name of the class to which the data-point belong. Eg: Yes or No

Clustering : It is similar to classification but with no labels.
Eg: Clustering different groups of people

# COMPARISON TABLE – ADVANTAGES AND DISADVANTAGES

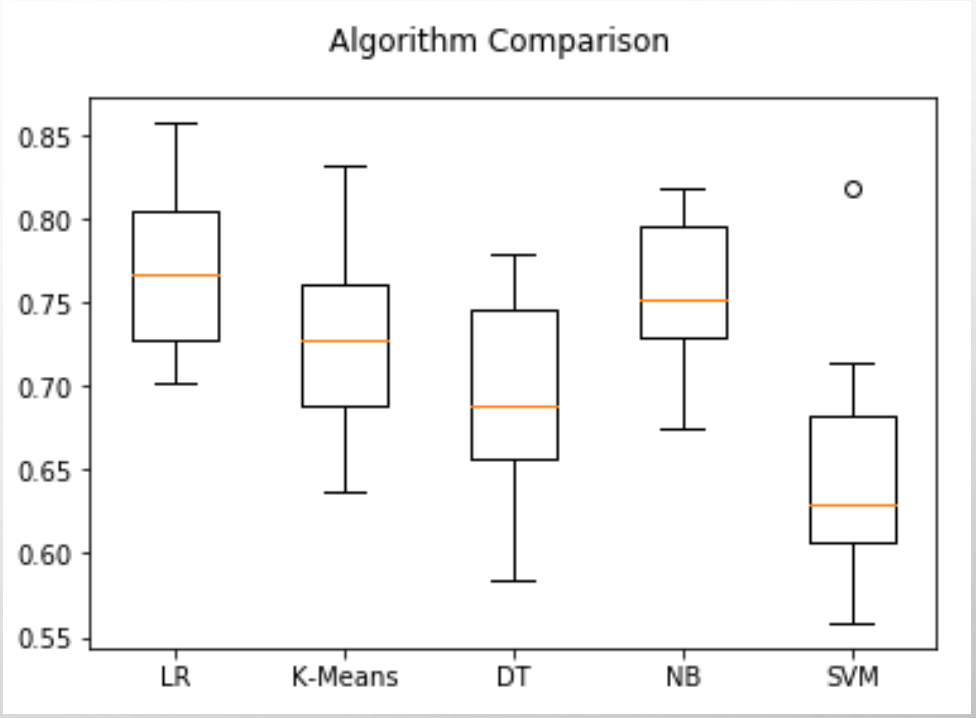
| Algorithm | Advantages | Disadvantages |
|---|---|---|
| Linear Regression | Easy to implement and understand | Efficiency is very low as most of the problem are non-linear in nature |
| Logistic Regression | Easy to scale up the model | The performance reduces with increased decision boundaries |
| SVM | High Accuracy, Mainly used in NLP | Selecting the kernel is difficult. High-Dimensional spaces difficult to visualize, Memory intensive |
| K-means | Faster Computation | Difficult to predict the value of K Initial centroid can affect the accuracy |
| Naïve Bayes | It works well with small datasets | It assumes features to be independent Unbalanced dataset leads to poorer accuracy. |
| Decision Tree | Good visualization | Need to build the tree each time new examples comes |
| Random Forest | Fast and highly scalable . Does not overfit | Its very complex and time consuming |

# DIABETES DATASET

PREDICT THE OUTCOME, WHETHER THE PATIENT IS DIABETIC OR NOT BASED ON THEIR AGE, BMI, INSULIN LEVEL, NO. OF PREGNANCIES AND SO ON

ALGORITHMS : RA (MEAN, STD)

- LOGISTIC REGRESSION : 0.769515 (0.048411)
- K – MEANS : 0.726555 (0.061821)
- DECISION TREES : 0.700427 (0.063340)
- NAÏVE BAYES : 0.755178 (0.042766)
- SVM : 0.651025 (0.072141)



# THANK YOU