3 DATASETS

8 continuous valued attribute
1 categorical valued attribute
Target: weather condition

DATASET 1

4 continuous valued attribute
Target: weather condition

DATASET 2

5 continuous valued attribute
Target: weather condition

DATASET 3

DATA PREPARATION

02

1

MISSING DATA

Check for missing value in each attribute and delete the record if null value is found

2

DUPLICATES

Check for duplicate records and delete if duplicates are present

3

OUTLIERS

Check for outlier and eliminate it

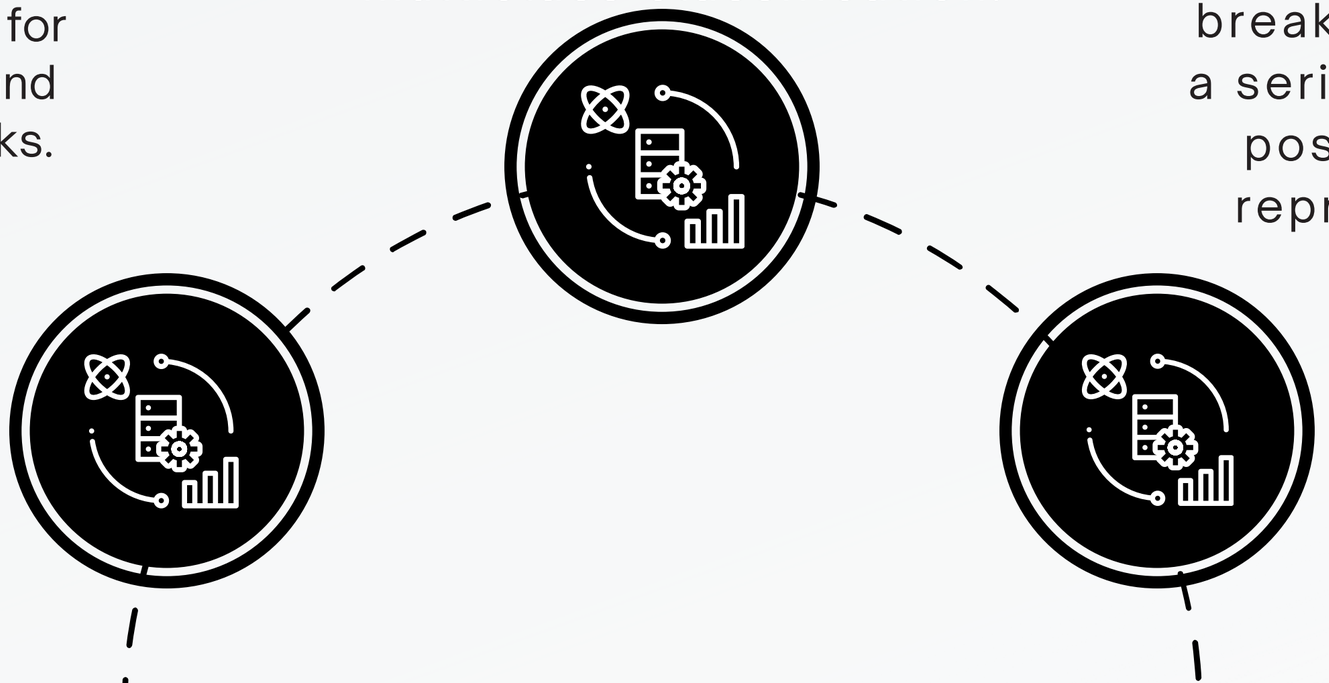
ALGORITHMS

KNN

K-Nearest Neighbors (KNN) is a supervised machine learning algorithm used for classification and regression tasks.

LOGISTIC REGRESSION

Logistic Regression is a supervised machine learning algorithm used for binary and multiclass classification.



DECISION TREE

A decision tree is a tree-like model used for decision-making in various fields. It breaks down a decision into a series of choices and their possible consequences, represented as branches and nodes.

KNN

DATA SET 1
Test Accuracy: 0.99

DATA SET 2
Test Accuracy: 0.78

DATA SET 3
Test Accuracy: 0.53

LOGISTIC REGRESSION

DATA SET 1
Test Accuracy: 0.92

DATA SET 2
Test Accuracy: 0.76

DATA SET 3
Test Accuracy: 0.44

DECISION TREE

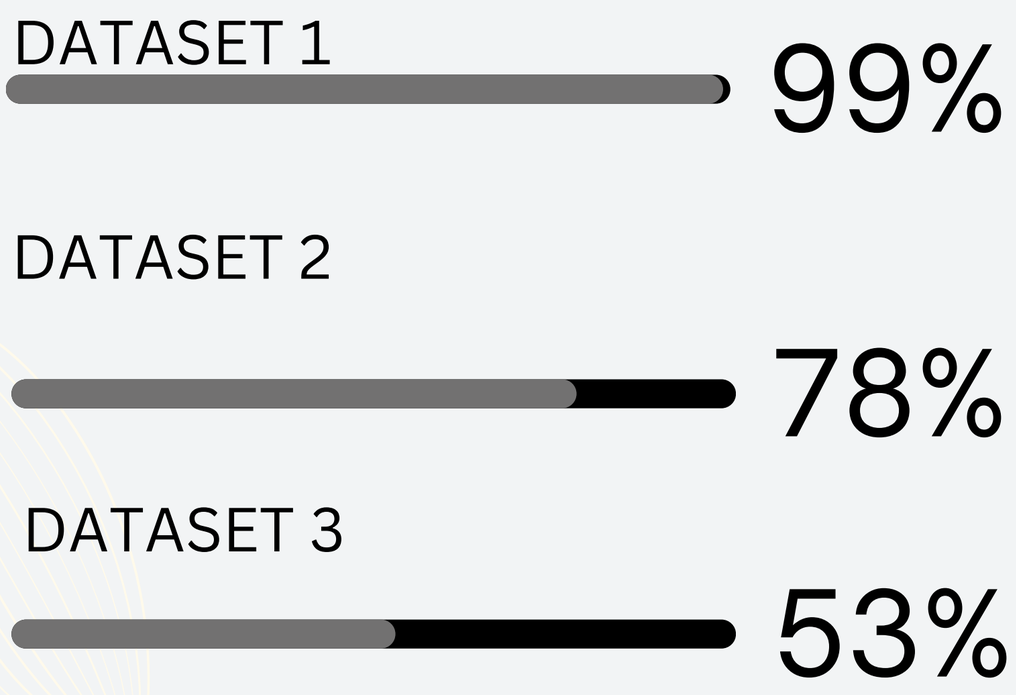
DATA SET 1
Test Accuracy:0.98

DATA SET 2
Test Accuracy: 0.75

DATA SET 3
Test Accuracy: 0.46

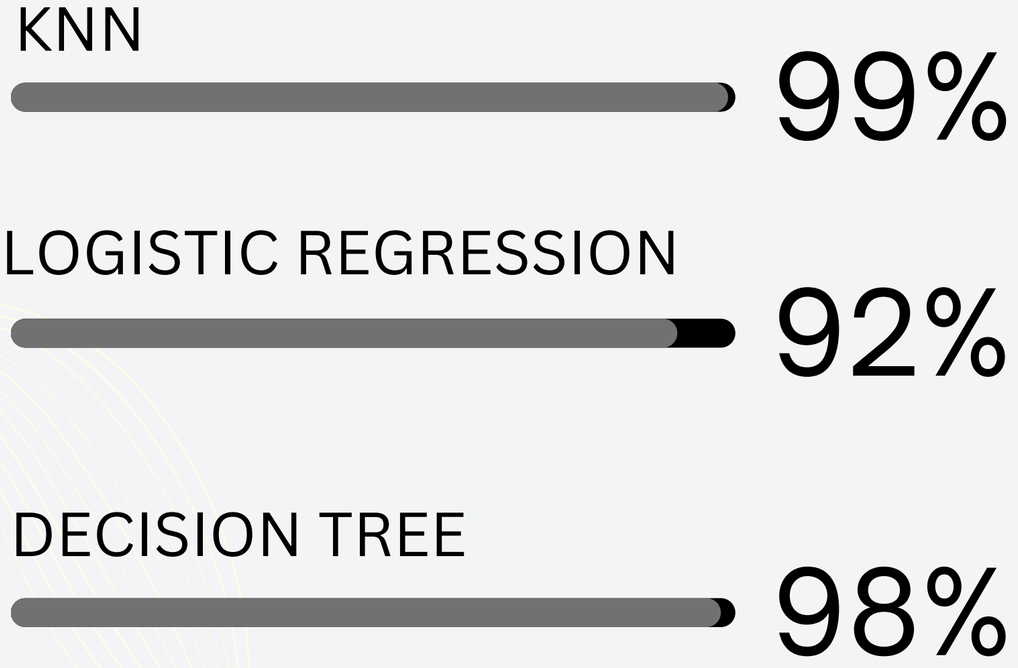
ALGORITHM ANALYSIS

KNN gives the best accuracy for all 3 datasets



DATASET ANALYSIS

DATA SET 1 has the best accuracy



CONCLUSION

Through the analysis we found that the algorithm that gives the most accurate predictions is KNN

Among the datasets used, the one which is most suited for our project is dataset 1 which had the highest accuracy as compared to the other two datasets.