

Observation:

1.Data Preprocessing:

- The given dataset(.csv file) doesn't contain any Null values and Outliers.
- Null values : can be identified using "df.isna().any()" python command
- Outliers : can be identified using "box plot" from seaborn library.
- While separating dependent and independent variables we must make sure that the independent variables should be of type "DataFrame" and dependent variable should be of type "Series"
- Encoding is done to change the object datatype of a particular variable to numerical.

2. Feature Engineering:

- For feature scaling we can make use of MinMaxScaler or StandardScaler.

3. Model Building:

- Used DecisionTreeClassifier

Using logistic Regression and decision tree, I got an accuracy score of 50%. Since there is very little correlation between each variables. The accuracy score is 50%.