SOCIAL NETWORK ADS

C R DEEPAK KUMAR

1833009

M.Sc DECISION & COMPUTING SCIENCES (VI SEM)

1. IMPORT ALL LIBRARIES
   * pandas
   * numpy
   * matplotlib.pyplot
   * maplotlib.colors
   * seaborn
2. READ THE DATASET (LOAD)
   * Using pd.read\_excel()
3. DISPLAY THE COLUMNS PRESENT
   * df.columns
4. PREPROCESS THE DATASET
   * Using label encoder transform gender Categorical variable to Numerical variable
5. DEFINE RANDOM FOREST ALGORITHM
   * Using Library sklearn.ensemble import the Random Forest Algorithm
6. FIND THE IMPORTANT FEATURES
   * Find important features using Random Forest Algorithm
   * Important features found was Age & EstimatedSalary
7. SEPARATE INDEPENDENT AND DEPENDENT VARIABLES
   * X=INDEPENDENT VARIABLES
   * Y=DEPENDENT VARIABLES
8. ANALYSE THE DATASET
   * Analyze the distribution of the data according to the Age and EstimatedSalary
   * Count of People who bought and people who didn’t buy
   * Purchase factor analyzed through Age
   * Purchase factor analyzed through EstimatedSalary
9. DEFINE ADABOOST CLASSIFIER
   * Using Library sklearn.ensemble import the AdaBoost Algorithm
   * Keep base estimator as Random Forest Algorithm
10. DEFINE XGBCLASSIFIER ALGORITHM
    * Using xgboost library import the XGBClassifier
11. DEFINE DECISIONVIZ AND TRAIN\_TEST\_SPLIT
    * Using yellowbrick import DecisionViz
    * Using sklearn.model\_selection library import train\_test\_split
12. USED TRAINING AND TESTING SET MODELS
    * These sets will help me determine if my model overfits or not.
13. XGBCLASSIFIER RESULT
    * Training: 98.7 %
    * Testing: 98.1 %
14. ADABOOST RESULT
    * Training: 100 %
    * Testing: 99.4 %
15. CONCLUSION
    * BEST MODEL IS ADABOOST ALGORITHM