**AD CLICK THROUGH RATE PREDICTION**

**KEY POINTS:**

1.Model used **Logistic Regression**

Logistic Regression is used to solve classification problem where we are trying to predict discrete categories. The reason why we are using Logistic regression is that this model gives high accuracy rate than other models then here we are gonna predict whether the ad is clicked / not.

2.Dataset name; Advertisements

3.This dataset contains **attributes** such as

* Daily Time Spent on Site
* Age
* Area Income
* Daily Internet Usage
* Ad Topic Line
* City
* Male / Female
* Country
* Timestamp: Time at which user clicked on an Ad or the closed window.
* Clicked on Ad: 0 or 1 is indicated clicking on an Ad.

4. **Steps** involved in prediction

* **Understanding the data**: Look at all the features by performing exploratory data analysis.
* **train\_test\_split:** Split the data into train and test datasets.
* **Building a Model:** Build a Logistic Regression model and fit it on the training set.
* **Predictions:** Predict values for the testing data.
* **Evaluations**

5.Necessary libraries

* Matplotlib 🡪 for data visualisation
* Pandas 🡪 for analysing the data
* Scikit learn 🡪 contains various ML algorithms

6. **confusion matrix** helps us to understand the behaviour of the model. Confusion matrix syntax

Predicted No Predicted Yes

Actual No TN FP

Actual Yes FN TP

TP- True Positive TN- True Negative FP- False Positive FN- False Negative

7. **classification report** gives us the precision and recall values accuracy,f1 core and support of the model.

8. **precision formula:** tp / (tp + fp).

9. **recall:** tp / (tp + fn).

10. **f1-score** harmonic mean of precision and recall best case value will be 1 and in worst case value will be 0

11. **support** - number of occurrences of each class in y\_test.

**Code explanation:**

1.Importing the libraries.i.e matplotlib,pandas

2.Read the data using read\_csv method

3.Get information about data using info()

4.Get statistical information about numerical columns using describe method

5.Plot graph for age attribute because age attribute plays a crucial role for tbis project.

6. split the data into training and testing set. Here for training we have taken 67% of data for training and 33% of data for testing.

7. Create an instance of logistic regression

8. Fit the data into model.

9. predict the output for the given test set.

1 represents the user clicked on the ad else 0

10 Then evaluate the model and the results of evaluation is as follows.

Explanation for Confusion matrix

[[156 6]

[ 24 144]]

**Confusion Matrix:**

User who were predicted to click on ad and the user who actually clicked on ad is 144.

Users who were predicted that they will not click on ad and the users who actually did not clicked on the ad is 156.

Users who were predicted to click on commercial and actually did not click on them is 6

Users who were not predicted to click on the commercials and actually clicked on them are 24.

**Accuracy 91%**