PROJECT TITLE:

ADVERTISING BASED ON

USAGE

PROJECT SUBMITTED BY:

* R.MARIO VISHAL
* SAI KALYAN AYYAGARI
* RAJKUMAR CHIRRA
* P S V G SAI AMRUTH
* KRISHNA TEJA

INTRODUCTION:

ARTIFICAL INTELLIGENCE

Artificial Intelligence is an approach to make a computer, a robot, or a product to think how smart human think. AI is a study of how human brain think, learn, decide and work, when it tries to solve problems. And finally this study outputs intelligent software systems. The aim of AI is to improve computer functions which are related to human knowledge, for example, reasoning, learning, and problem-solving.

The intelligence is intangible. It is composed of

* Reasoning
* Learning
* Problem Solving
* Perception
* Linguistic Intelligence

The objectives of AI research are reasoning, knowledge representation, planning, learning, natural language processing, realization, and ability to move and manipulate objects. There are long-term goals in the general intelligence sector.

Approaches include statistical methods, computational intelligence, and traditional coding AI. During the AI research related to search and mathematical optimization, artificial neural networks and methods based on statistics, probability, and economics, we use many tools. Computer science attracts AI in the field of science, mathematics, psychology, linguistics, philosophy and so on.

MACHINE LEARNING

Machine learning is a subfield of artificial intelligence (AI). The goal of machine learning generally is to understand the structure of data and fit that data into models that can be understood and utilized by people.

Although machine learning is a field within computer science, it differs from traditional computational approaches. In traditional computing, algorithms are sets of explicitly programmed instructions used by computers to calculate or problem solve. Machine learning algorithms instead allow for computers to train on data inputs and use statistical analysis in order to output values that fall within a specific range. Because of this, machine learning facilitates computers in building models from sample data in order to automate decision-making processes based on data inputs.

Any technology user today has benefitted from machine learning. Facial recognition technology allows social media platforms to help users tag and share photos of friends. Optical character recognition (OCR) technology converts images of text into movable type. Recommendation engines, powered by machine learning, suggest what movies or television shows to watch next based on user preferences. Self-driving cars that rely on machine learning to navigate may soon be available to consumers.

Machine learning is a continuously developing field. Because of this, there are some considerations to keep in mind as you work with machine learning methodologies, or analyze the impact of machine learning processes.

MACHINE LEARNING METHODS

* **Supervised machine learning algorithms**can apply what has been learned in the past to new data using labeled examples to predict future events. Starting from the analysis of a known training dataset, the learning algorithm produces an inferred function to make predictions about the output values. The system is able to provide targets for any new input after sufficient training. The learning algorithm can also compare its output with the correct, intended output and find errors in order to modify the model accordingly.
* **Unsupervised machine learning algorithms**are used when the information used to train is neither classified nor labeled. Unsupervised learning studies how systems can infer a function to describe a hidden structure from unlabeled data. The system doesn’t figure out the right output, but it explores the data and can draw inferences from datasets to describe hidden structures from unlabeled data.
* **Reinforcement machine learning algorithms**is a learning method that interacts with its environment by producing actions and discovers errors or rewards. Trial and error search and delayed reward are the most relevant characteristics of reinforcement learning. This method allows machines and software agents to automatically determine the ideal behavior within a specific context in order to maximize its performance. Simple reward feedback is required for the agent to learn which action is best; this is known as the reinforcement signal.

PYTHON FOR MACHINE LEARNING

Python is a widely used high-level programming language for general-purpose programming. Apart from being open source programming language, python is a great object-oriented, interpreted, and interactive programming language. Python combines remarkable power with very clear syntax. It has modules, classes, exceptions, very high level dynamic data types, and dynamic typing. There are interfaces to many system calls and libraries, as well as to various windowing systems. New built-in modules are easily written in C or C++ (or other languages, depending on the chosen implementation). Python is also usable as an extension language for applications written in other languages that need easy-to-use scripting or automation interfaces.

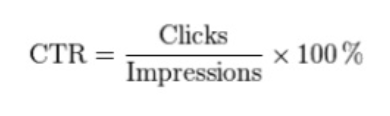
Popular Libraries Used

* NUMPY
* MATPLOTLIB
* PANDAS
* SIKIT-LEARN
* SCIPY
* TENSORFLOW
* SEABORN

OBJECTIVES OF RESEARCH:

The main Objective of this research is to predict the Click-Through-Rate.

Basically CTR is the ratio of users who click on an ad to the number of total users who view the ad.



Used to measure the success of an online advertising campaign. So in this research we take potential inputs which affect the target variable in one or the other way.

And creating a model and applying a suitable algorithm for a prediction.

PROBLEM STATEMENT:

To predict whether a user clicks on an advertisement or not based on Click-through-rate (CTR) prediction.

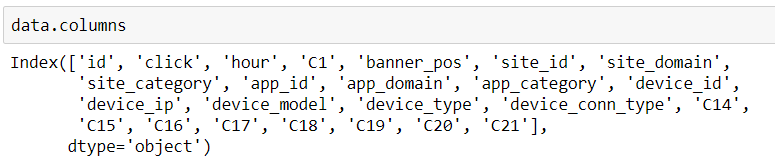
Review of Literature

Data set has 24 features and 10k observations.

The output variable or the target variable is the feature “click”.

The data types of the features include int,float and object values.

The data type of “click” is int, and it has the output of either 0s or 1s.



Data Collection

Source of Data Collection:

Kaggle: Avazu CTR Prediction Dataset

Pictures/Plots: Jupyter Notebook

**Methodology**

**Exploratory Data Analysis**

Following are the Input Variables of the dataset:

* id: ad identifier
* click: 0 for non-click, 1 for click
* hour: in the format of YYMMDDHH
* C1: some anonymized categorical variable e.g. 1002
* banner\_pos: where a banner is located, 1 and 0
* site\_id: site identifier
* site\_domain: hashed site domain
* site\_category: hashed site category e.g. 28905ebd
* app\_id: mobile app identifier
* app\_domain
* app\_category
* device\_id: mobile device identifier
* device\_ip: IP address
* device\_model: hashed model e.g. iPhone 6, Samsung
* device\_type: hashed device type e.g. tablet, smartphone
* device\_conn\_type: hashed type of connection e.g. Wi-Fi, 4G
* C14-C21: some more anonymized categorical variables

Python Libraries Used:

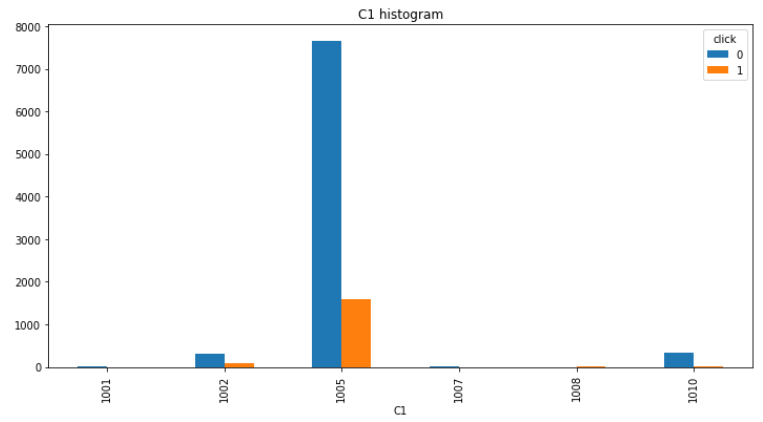
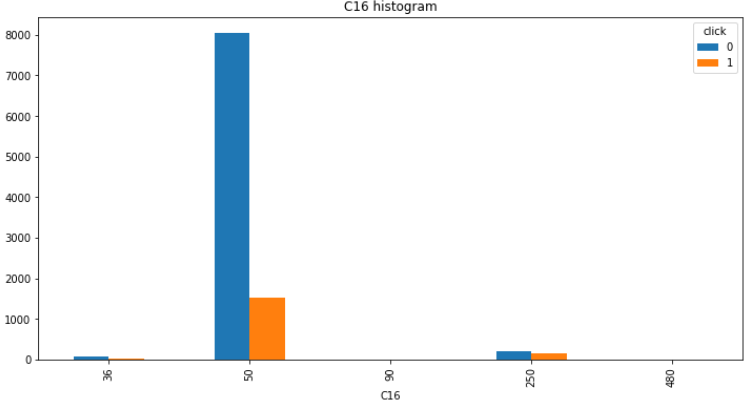
Numpy

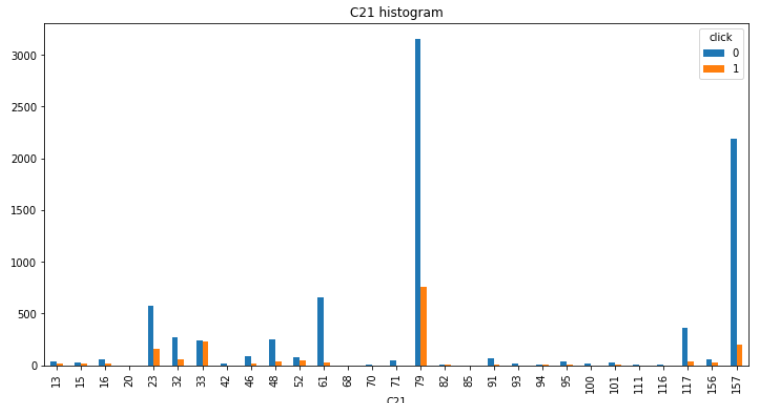
Pandas

Scikit-learn

Matplotlib

Seaborn



FINDINGS AND SUGGESTIONS :

We have found that only 17% of the users have clicked the ad which is appeared on the screen.

As from the previous slides the accuracy of Random Forest,Logistic Regression,Support Vector machine(SVM) are equal i.e 83.6%

And the ROC Scores are all equal i.e 50.

So we considered Logistic regression as the best suited algorithm for the model.

CONCLUSION :

Through data analysis ,we have concluded that only 17% of the users have clicked the ad which is appeared on the screen.

The highest accuracy obtained is 83% for Random Forest Classifier.

So we considered Random Forest Classification as the best suited algorithm for the model.