

# An Efficient Ad-Click Prediction System using Machine Learning Techniques



A.Lakshmanarao, A.Srisaila, T. Srinivasa Ravi Kiran

**Abstract:** Ad-click prediction is a learning problem that is highly related to the multi-billion-dollar ad-promoting the online advertising industry. As the number of internet users in India reached 525 million in 2019, online advertising companies are trying to influence internet usage users for promoting their business. Machine learning is a technique in which systems getting to act without any explicit programming. Currently, machine learning is pervasive today and we can use machine learning models in every research field. The accuracy of the ad-click prediction system impacts business revenue, so it is very important to build a prediction system with fewer false positives and false negatives. In this paper, we proposed an ad-click prediction system based on machine learning techniques. The dataset is taken from Kaggle. The dataset contains nine features. The goal of the model is to evaluate the probability of an online user to click on a given ad. We built a machine learning model based on these features. We applied a voting classifier on the dataset and achieved an accuracy of 98%. We used python language for implementation.

**Index Terms:** Ad-click prediction, machine learning, python.

## I. INTRODUCTION

The web is fantastically wealthy in the data that the individuals need on everyday premise. Consistently client get to a considerable lot of records on different points like news, travel, plans, diversion recordings and other subjects. Web search engines have income from sponsored search. Click events on ads produces financial benefits for search engines. Generally, ad-clicks highly depends on the order at which that ad is placed on the search page. During surfing of their preferred locales, they are given different of substance which they may be intrigued by. Also, since the clients are bound to tap on promotions that are important to their inquiries, it's of extraordinary significance that the correct advertisements are shown. Content disclosure stage that convey the client with their most loved destinations through online sponsors. This issue is essentially viewed as "Anticipating advertisement

navigate rates/Click Through Rates(CTR)". CTR expectations used in issues like logical notice, search ad motor, video commercial, and so on. Ad-click prediction supports expansion of an internet searcher's income.

The advertisements that are properly placed not only attracts the customers but also increases the likelihood of purchasing the products. If a user is searching for a particular issue and showing the Ad related to his/her issue greatly enhances the satisfaction of the user. The major sources of income also come from Search advertising for major companies that charge the advertiser click-per-ad policy. That is why the click rate prediction of the ads is very important for maximizing the revenue. As the internet users crosses 500 million in 2019 in india [1], it is important to determine correctly the click prediction of ads. Machine learning programming is not similar to traditional programming. Machine learning takes data samples and output as input. With these two, a machine learning algorithm builds a model that learns the given input. That learned model is used for the prediction or classification of new input samples.

## II. LITERATURE SURVEY

A various number of machine learning algorithms are applied for ad-click prediction. Stanley Jacob [2] et.al applied logistic regression, Naive Bayes, support vector machine, neural network models for ad-click prediction and achieved good accuracy with support vector machine. Qianqian Wang[3] et.al applied stacked autoencoder for feature selection process and then applied deep neural network models for ad-click prediction. Bin Liu[4] et.al applied convolution neural networks for ad-click prediction. They first applied feature generation techniques and then CNN is applied. They showed that their model achieved better results. Weiwei Deng [5] proposed an ad-click prediction system with recurrent neural network model with LSTM and achieved good results. S. Saraswathi [6] et.al applied four machine learning algorithms namely SVM, Naive Bayes classifier, Logistic Regression model, Decision Tree model. They achieved an accuracy of 96% with random forest. We also used the same dataset which is used in [6], but we achieved an accuracy of 98% with a voting classifier. M. Sree Vani [7] applied feature selection techniques and proposed supervised classification techniques for mobile ad-click prediction. With logistic regression, they achieved good results.

Revised Manuscript Received on February 05, 2020.

\* Correspondence Author

**A. Lakshmanarao\***, Department of Computer Science & Engineering, Raghu Engineering College, Visakhapatnam, A.P, India (e-mail: laxman1216@gmail.com)

**Dr. A.Srisaila**, Department of Information Technology, V.R. Siddhartha Engineering College Vijayawada, India (e-mail: sr.saila@gmail.com)

**Dr. T. Srinivasa Ravi Kiran**, Department of Computer Science P.B.Siddhartha College of Arts & Science Vijayawada, India (e-mail : kirantsr1@gmail.com)

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

# An Efficient Ad-Click Prediction System using Machine Learning Techniques

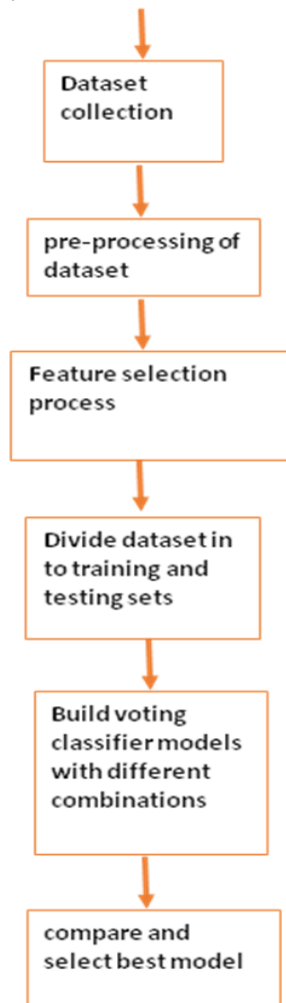
Rohit Kumar [8] et.al proposed logistic regression model for estimating Click through rate of Advertisements and achieved good results with their model.

Although most of the researchers applied deep learning techniques, we proposed machine learning models only. The reason for choosing machine learning models is that we achieved an accuracy of more than 97% accuracy without complex structures of deep learning models.

### III. RESEARCH METHODOLOGY

After collecting the dataset, we applied preprocessing, feature selection techniques. Later, machine learning model is built on the dataset.

#### Proposed model:

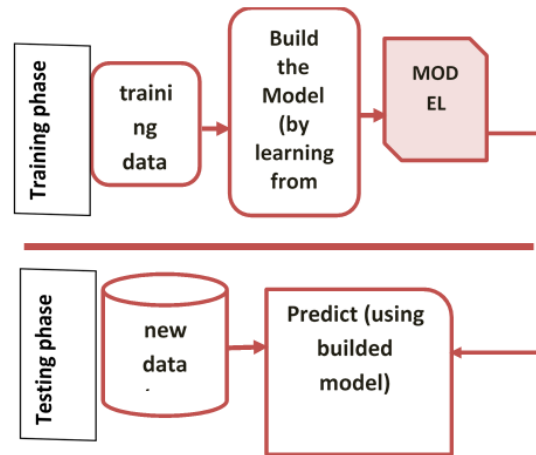


**Figure.1: Proposed Model**

We collected a dataset from Kaggle[9].The dataset contains 1000 samples with 9 different features namely user's daily spending time on Site, user's age, Income, Ad Topic Line, user's daily internet usage, user's city, user gender, Country, Timestamp. Based on these 9 features, we need to predict user can click an ad or not."Clicked on Ad" is a dependent variable(feature) which we need to predict.First,we applied data preprocessing step.As there are no missing values,there is no need to handle missing values. Next we applied feature selection process. In feature selection method, we eliminated four features namely Ad Topic Line, city, Country and Timestamp.ML models basically operates on mathematical operations. If there are

any categorical features, we need to apply label encoding and one hot encoding techniques to convert categorical features into numerical featured. But in our dataset, there are no important categorical features. So, we removed these 4 features..

#### Structure of supervised Machine Learning model:



**Figure.2: Structure of supervised ML model**

After identifying the best features, we divided the dataset as training and testing set using 5-fold cross validation technique. K-Fold cross validation is a method where given data samples are divided into k-no. of folds, where each fold can be used for testing part at some point. Let's take the situation of 5-Fold cross validation (here value of k is 5). Here, the dataset is part into 5 folds. In the primary emphasis, the main overlap is utilized to test the model and the rest are utilized to prepare the model. In the subsequent cycle, second overlay is utilized as the testing set while the rest fill in as the preparation set. This procedure is rehashed until each crease of the 5 folds have been utilized as the testing set. After applying 5-fold cross validation, we applied various machine learning models.

#### Support Vector Machine:

Support Vector Classification technique (SVM) is useful for classification as well as regression problems. Ad-click prediction is a classification problem. So here SVM is used for the classification task. In support vector classifier, an n-dimensional space is used to plot each data item. After that, a hyperplane is used which differentiates the classes to complete a classification task.

#### K- Nearest Neighbor Classification:

K-NN is one of the most widely used supervised learning classification method. In K-NN, K is the user-chosen number.

Algorithms identifier similar things nearer to K nearest neighbors. For finding distance, Euclidean distance measure is used. Finally,all the items in the given dataset are classified.

**Decision Tree Classification:**

It is one of the famous classifiers used in machine learning. A decision tree can also be useful for classification, regression tasks. It breaks the given data samples into smaller parts and a decision tree is constructed incrementally. The child nodes of a tree are used for classification.

**Random Forest classification:**

Random Forest is one of the ensemble methods used in machine learning. Ensemble means combining more than one classification technique. RF combines several decision trees to make decisions. RF is the most powerful classification technique.

**Adaboost:**

Adaboost is another ensemble approach used for the classification task. After training each classifier, the classifier is assigned with weight-based on accuracy. Higher weight is assigned to the best classifier so that it will get a better outcome.

**Naive Bayes classifier:**

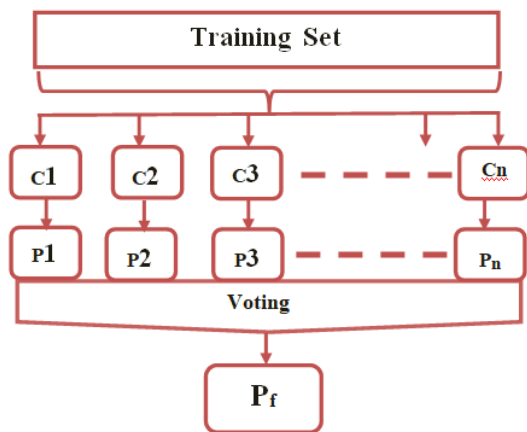
A Naive Bayes classifier is based on probability. This technique is used only for classification method. This algorithm is totally based on bayes theorem.

**Logistic Regression:**

The name include regression, but it is a classification technique. It is a statistical supervised learning approach. It is used for binary classification tasks.

**Voting classifier:**

The Ensemble Voting Classifier is also called as meta-classifier. It combines similar or conceptually different types of machine learning algorithms for classification task through majority voting.



**Figure.3: Structure of voting classifier**

Here C1,C2,C3... represents classifiers and p1,p2,p3... represents classifier probabilities. Pf is final probability. A voting classifier apply majority voting from these class probability values.

**IV. EXPERIMENTS & RESULTS**

All the experiments are implemented in python language. Python’s scikit-learn package contains vast number of functions for implementing machine learning algorithms. Scikit-learn provides different modules for different tasks.

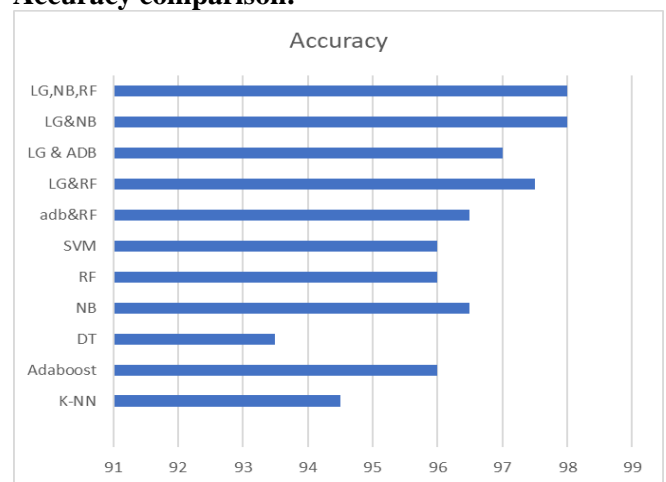
With five features which are selected in feature selection method, we divide the dataset into training and testing sets. Later normalization technique is used using python’s standard scalar package. After that, we applied seven machine learning algorithms. All the machine learning models achieved more than 90% accuracy. But we achieved best accuracy of 97% with Logistic Regression model. For enhancing efficiency, we implemented various combinations of voting classifier models and achieved an accuracy of 98%, which is greater than accuracy with single model. The voting classifier implements either soft voting or hard voting. In hard voting, it predicts the final class label that has been predicted most frequently by classifier models, where as in soft voting, it predicts the class labels by averaging the class probabilities. We applied soft voting method. We tried different combinations of above seven algorithms. We tried combination of two algorithms as well as combination of three algorithms. We achieved best accuracy with combination of logistic regression and naïve bayes. All the results are tabulated for comparisons.

**Results of Experiments:**

Algorithm	Precision	Recall	Accuracy
Logistic Regression	97	97	97
K-NN	95	94	94.5
AdaBoost	96	96	96
Decision Tree	94	94	93.5
Naïve Bayes	97	96	96.5
Random Forest	96	96	96
SVM	96	96	96
ADB&RF	96	96	96.5
LG&RF	98	97	97.5
LG & ADB	97	97	97
LG&Naïve Bayes	98	98	98
LG, NB, RF	98	97	98

**Table 1: Accuracy values of ML models**

**Accuracy comparison:**



**Figure.4: Accuracy comparison of ML models**

## V.CONCLUSION

In this paper, we build a model for Ad-click prediction using machine learning techniques. We first evaluated seven machine learning algorithms SVM, decision tree, random forest, adaboost, K-NN, logistic regression, Naïve Bayes. We achieved an accuracy of 97% with Logistic Regression model. Next, we applied several voting classifier combinations and achieved an accuracy of 98% with Logistic Regression-Naive Bayes voting classifier.

## REFERENCES

1. Statista, "The statistics portal." [Online]. Available: [https://www.statista.com\(https://www.statista.com/statistics/255146/number-of-internet-users-in-india/\)](https://www.statista.com(https://www.statista.com/statistics/255146/number-of-internet-users-in-india/).
2. Stanley Jacob, Lingjie Kong, "Progress Report: Predicting Which Recommended Content Users Click", Stanford Journal, 2015
3. Qianqian Wang, Xing, Xiaohui Zhao, Fang'ai Liu, Shuning, "A New Approach for Advertising CTR Prediction Based on Deep Neural Network via Attention Mechanism", Research Article, Hindawi, Computational and Mathematical Methods, Volume 2018, Article ID 8056541, <https://doi.org/10.1155/2018/8056541>.
4. Bin Liu, Yuzhou Zhang, Yingzhi Chen, Jinkai Yu, Huifeng Guo, Ruiming Tang, "Feature Generation by Convolutional Neural Network for Click-Through Rate Prediction", rXiv:1904.04447v1 [cs.IR] 9 Apr 2019.
5. Weiwei Deng, Yang Qi, Tunzi Tan, Xiaoliang LingEren Manavoglu, Qi Zhang, "Ad Click Prediction in Sequence with Long Short-Term Memory Networks, Externality-aware Model", SIGIR'18, July 8-12, Ann Arbor, MI, USA, 2018.
6. S. Saraswathi, Valli Devi Krishnamurthy, D. Venkata Vara Prasad, Tarun R K, S Abhinav, D.Rushitaa, "Machine Learning Based Ad-click Prediction System", International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-8, Issue-6, August 2019.
7. M. Sree Vani, "Prediction of Mobile Ad Click Using Supervised Classification Algorithms", International Journal of Computer Science and Information Technologies, Vol. 7 (2), 623-625, 2016.
8. Rohit Kumar, Vani D Naik, Sneha Manjunath Naik, Smita Shiralli, Sunil V.G, Moula Husain, "Predicting Clicks: CTR Estimation of Advertisements using Logistic Regression Classifier", Advance Computing Conference, IACC, IEEE International, 2015.
9. Available online: <https://www.kaggle.com/fayomi/advertising>

## AUTHORS PROFILE



**Mr. A. Lakshmanarao**, Assistant Professor, CSE Dept, Raghu Engineering College, completed Bachelor of Technology in CS&IT and M. Tech in Software Engineering. Currently pursuing Ph.D. in Machine Learning in Andhra University, Visakapatnam. His areas of interest are Machine Learning, Cyber Security and Deep Learning.



**Dr. A. Srisaila**, Assistant Professor, IT Dept, P.B. Siddhartha Engineering College completed M.Tech (CSE) from Bapatla Engineering College and Ph.D in CSE from Acharya Nagarjuna University. Her research areas are Software Reliability Engineering, Graphical Passwords, Human Computer Interaction and Big Data Analytics.



**Dr. T. Srinivasa Ravi Kiran**, Assistant Professor & HOD, Dept of Computer Science, P.B. Siddhartha College of Arts & Science, Vijayawada completed M.Tech(CSE) from ANU and Ph.D. from ANU. His research areas are Graphical Password Authentication, Cryptography, HCI, Software Reliability Engineering Cyber Security.