International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-8 Issue-6, August 2019

Machine Learning Based Ad-click Prediction System

S. Saraswathi, Vallidevi Krishnamurthy, D. Venkata Vara Prasad, Tarun R K, S Abhinav, D. Rushitaa

Abstract:
Online advertising is a gargantuan commerce and has potential for rapid growth. This paper presents novel approach of solving the advertisement prediction problem. The aim of this research is to predict an ad-click through various machine learning techniques and to compare their accuracy rates. This, would help the advertisers use the appropriate technique to increase their overall revenue through targeted advertising. The combination of features used, makes this research unique.

Keywords: Machine Learning, Ad-click, Prediction, Ad-serving

I. INTRODUCTION

Online advertising is a colossal business worth more than 50 billion dollars. The revenue that advertisers earn is increasing due to their practice of targeted advertising. A lot of analysis has been done in the previous years, from ad-click prognosis to ad serving. The ad-click prognosis has been used in history in every type of advertisement format such as, search engines, textual and contextual advertisements, video advertisements etc. With the rapid increase of advertising network, click prediction needs huge data analysis. The advertisement prediction is considered to be one of the most lucrative stories in the domain of machine learning. Also, the advance in ad serving machinery have brought a real time bidding answer where ads are short listed based on the attributes of the publishers and the viewers. In general, the methodology and the technique followed in machine learning, brings a change to the ad-click prediction method. The methods include data assortment, attribute extrication, ad-click prognosis and ad serving.

Revised Manuscript Received on August 30, 2019.
* Correspondence Author
S. Saraswathi*, Associate Professor, SSN College of Engineering, Kalavakkam, Chennai, TamilNadu, India
Vallidevi Krishnamurthy, Associate Professor, SSN College of Engineering, Kalavakkam, Chennai, TamilNadu, India
D. Venkata Vara Prasad, Professor, Department of CSE, SSN College of Engineering, Kalavakkam, Chennai, TamilNadu, India
Tarun R K, Under Graduate Students, Department of CSE, SSN College of Engineering, Kalavakkam, Chennai, TamilNadu, India
S Abhinav, Under Graduate Students, Department of CSE, SSN College of Engineering, Kalavakkam, Chennai, TamilNadu, India
D. Rushitaa, Under Graduate Students, Department of ECE, SSN College of Engineering, Kalavakkam, Chennai, TamilNadu, India

The features used in this work are Frequent Time Spent on Website, Lifetime, Field Revenue, Frequent Internet Usage and Gender. The features used in this work are all human based which contributes a lot to developed ad-click prediction system.

This research study explains the entire procedure of ad click prognosis with various steps which are briefed in the further sections. Section 2 deals with the literature survey on ad-click prediction and serving. Section 3 deals with feature extraction and selection. Section 4 describes the research mode and approach used to predetermine ad-click. Empirical outcomes are discussed in section 5. Conclusion is drawn in section 6.

II. LITERATURE SURVEY

Several research works to find solution to the ad-click prediction have already been done using various machine learning techniques like, Logistic Regression, Naive Baye's Classifier, Support Vector Machines and Decision Trees. Logistic Regression has a pivotal role in the early analysis in this domain. The relevance of logistic regression in discovering the users behaviour can be observed in [6], [2]. Naive Baye's has played a fundamental role in the construction of ad-click prognosis, which can be noticed in [3], [4]. Due to the accelerated growth of online activity, analysts have been procuring different procedures to display relevant advertisements which suits the viewers interest. The assessment of how viewers acknowledge to the display advertisements is given in [5]. The research showed in [10] has concentrated on aiming display advertisements using large-scale machine learning. The recent advancement in social networking on the internet has formed some fabulous illustrations of ad serving machinery such as Twitter and Facebook. Twitter advertising prototype is also based on logistic regression as presented in [8]. Another top social networking site Facebook has also used decision trees with logistic regression shown in [9] to serve their ads to their agglomeration of customers frequently.

III. FEATURE EXTRACTION AND SELECTION

The data gathered by the data scientist have several features that may or may not be relevant to the topic of interest. Also it may not be in a suitable format. The first and foremost task to the data scientist is to extract the appropriate collection of attributes which preferably suits the learning algorithm.

Before processing, it needs to be transformed to prevent relapse problems like over fitting and under fitting as presented in [7]. The following Table 1 shows the list of features present in the dataset.

Features	Description
Frequent Time Spent on Website	User time spent on Website in minutes
Lifetime	User lifetime in years
Field Revenue	Average revenue of the geographic location of the user
Frequent internet usage	User time spent on the internet in minutes
Ad topic line	Caption of the ad
Place	Place where the user lived
Male	User's gender
Country	Country of the user
TimeStamp	Time on which the user clicked the advertisement
Clicked on Ad	1 or 0 indicates whether the advertisement is clicked or not

Table 1: List of features

The proposed ad-click prediction model is based on human features. To adapt to this, certain human related features like Frequent Time Spent on Website, Lifetime, field Revenue, Frequent Internet Usage, and Gender are alone considered in this model. These attributes are extricated from the dataset to efficiently develop the prototype. Some features such as Advertisement Topic Line, City, Country, Time-stamp are not human features, so they are ignored from consideration. The features that are taken into consideration are shown in Table 2. All extracted attributes have been indoctrinated into a convenient form to make study easy.

Features	Description
Frequent Time used on Website	User time spent on Website in minutes
Lifetime	User lifetime in years
Field Revenue	Average revenue of the geographic location of the user
Frequent internet usage	User time spent on the internet in minutes
Male	User's gender
Clicked on Ad	1 or 0 indicates whether the advertisement is clicked or not

Table 2: Features taken into consideration

IV. PROPOSED RESEARCH METHODOLOGIES AND TECHNIQUES

The aim of this work is to predetermine the Click Through Rate (CTR) of a particular user for a particular advertisement. The CTR prediction is used to predict whether the web-site viewer will be interested in a particular advertisement (ad) or not. When an observer visits a publisher's web-site, in a period of few milliseconds the ad is being furnished established on the maximal CTR. The human attributes that have been picked through feature selection phase are now passed to learning algorithm to predict CTR. The analysis has been carried out using different learning algorithms like i) Naive Bayes Classifier, ii) Logistic Regression, iii) Support Vector Machine, iv) Decision Tree. Among these SVM is the supervised learning model, which is shown in Figure 1. Grid Search algorithm was implemented and the results were tested.

Naive Baye's Classifier works under the consideration that each feature relies only on the class. This becomes the main drawback as it does not consider features that are independent of class. The Naive Baye's classifier is implemented using the scikit learning technique. Here CTR is predicted using the assumption that the human features such as Frequent Time Spent on Website, Age, Field Revenue, Frequent Internet Usage, and Gender are tentatively self-reliant of one and another, given the class. An accuracy rate of 85 percent was obtained.

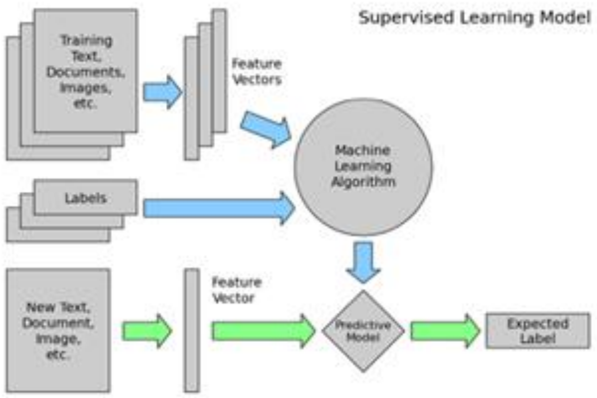

Figure 1: Supervised Learning Model

Logistic Regression is used for regression analysis when the classification is binary. Since the aim is to predict whether the person clicked on an ad or not, that is either 1 or 0, logistic regression was considered. A logistic regression model was advanced using the scikit learn technique and it's accuracy rate was 92 percent.

Support Vector Machines (SVM) is an administrated machine learning algorithm used for classification and regression. SVM was used to predict the CTR of the ad click prediction problem as in [1]. SVM was implemented using the scikit learn technique. An accuracy rate of 92 percent was obtained.

Decision Trees (DTs) is a non-parametric administered learning technique used for classification and regression. The ad prediction was done by learning simple decision rules that are inferred from the selected human features. The DT based model was created using the scikit learn technique. The accuracy rate of 92 percent was obtained.

All machine learning algorithms depends upon the parameters used for training. The Grid search algorithm was to select the hyper parameters that were contributing highly for the better accuracy. The Grid Search and cross validation in the context of Random Forest Classifier and Stratified K-Fold methods were used to boost the efficiency of the prognosis and have obtained the accuracy of 96 % for the same.

V. EXPERIMENTAL RESULTS

This model was utilised in ad-click prediction for an iOS application which is shown in Figure 2. The improvised results are shown below in Figure 3.

Retrieval Number: F9366088619/19©BEIESP
DOI: 10.35940/ijeat.F9366.088619
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-8 Issue-6, August 2019

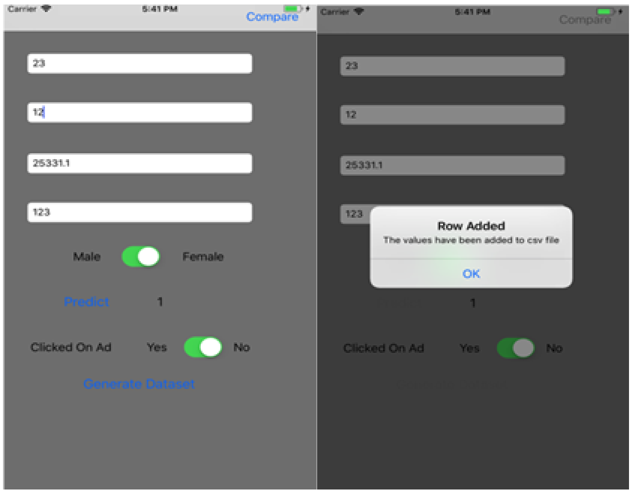

Figure 2: Ad-Click prediction system in an iOS application

	mean_fit_time	mean_score_time	mean_test_accuracy_score	mean_test_precision_score	mean_test_recall_score	mean_train_accuracy_score	mean_train_prec
14	0.867128	0.423854	0.960000	0.971569	0.949333	0.993630	
13	2.647535	0.708878	0.960000	0.971569	0.949333	1.000000	
3	3.890655	0.778030	0.964000	0.971565	0.957333	0.972889	
18	0.929687	0.416271	0.964000	0.971565	0.957333	0.999704	
6	0.920267	0.428986	0.961333	0.971565	0.952000	0.986666	

5 rows × 83 columns

Figure 3: Results with Improvised Accuaracy

VI. CONCLUSION

The implementation of various machine learning algorithms for the ad-click prediction system development is explained in detail. Their accuracy rates were recorded and compared with that of the custom model. The accuracy rate of the custom model was found to be ninety six percent which is much higher than that of the standard models. As a future work, the custom model developed can further be enhanced by addition of new features. Grid Search techniques allows the user to tweak the parameter values giving the user multiple ways to improve the existing model. The ML model was trained using Scikit-Learn. The Coremltools framework was used with Python to generate a model file. This file was utilised within the iOS app, to read the input features and gave an output of 0 or 1. The app allowed the user to further click ad or not and stored the result in the .csv file, which was further used to retrain the model. As the data keeps on growing with this system, the accuracy of the ad-click prediction system, which can predict if or not the user clicks any ad or not has increased to 96 percentage where the existing models gave an accuracy rate of only 92 percentage.

REFERENCES

1. N. Kwak, S.-I. Choi, and C.-H. Choi, Feature extraction for regression problems and an example application for pose estimation of a face, in Interna-tional Conference Image Analysis and Recognition, pp. 435-444. 2008.
2. H. B. McMahan et al., Ad click prediction: a view from the trenches, in Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining" pp. 1222-1230, 2013.
3. M. Richardson, E. Dominowska, and R. Ragno, Predicting clicks: estimating the click-through rate for new ads, in Proceedings of the 16th international conference on World Wide Web, pp. 521-530, 2007.
4. T. Graepel, J. Q. Candela, T. Borchert, and R. Herbrich, Webscale bayesian click-through rate predic-tion for sponsored search advertising in microsofts bing search engine, in Proceedings of the 27th International Conference on Machine Learning (ICML 10), pp. 13-20. 2010.
5. O. Chapelle, E. Manavoglu, and R. Rosales, Simple and scalable response prediction for display advertising, ACM Trans. Intell. Syst. Technol. TIST, Vol. 5, No. 4, p. 61, 2015.
6. A. H. Karp, Using logistic regression to predict customer retention, in Proceedings of the Eleventh Northeast SAS Users Group Conference. http://www. lexjansen. om/nesug/nesug98/solu/p095. pdf, 1998.
7. M. A. Babyak, What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models, Psychosom. Med., Vol. 66, No. 3, pp. 411-421, 2004.
8. C. Li, Y. Lu, Q. Mei, D. Wang, and S. Pandey, Click-through Prediction for Advertising in Twitter Timeline, pp. 1959-1968, 2015.
9. S. Kotsiantis and D. Kanellopoulos, Association rules mining: A recent overview, GESTS Int. Trans. Comput. Sci. Eng.,Vol. 32, No. 1, pp. 71-82, 2006.
10. B. Dalessandro, F. Provost, T. Raeder, C. Per-lich, and O. Stitelman, Machine learning for targeted display advertising: Transfer learning in action, 2013.