International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Enhancing Search Advertisement Performance with Effective Feature Engineering

Saravanan K^{1}, Sathish G^{1*}, Aadhityan S^{1*}, Srinidhi S^{2}

*[1]Computer Science and Engineering, Agni College of Technology

**[2]Assistant Professor, Computer Science and Engineering Department, Agni College of Technology

ABSTRACT:

One of the critical technologies in the study of search advertising recognition is feature engineering. Most search advertising strategies currently in use are chosen based on past information, which is too subjective to be widely adopted. A feature processing approach based on the preliminary analysis of a user and store data is proposed, and the conversion rate is then forecasted using XGBoost (eXtreme Gradient Boosting). This research uses the advertisements of Ali Search advertising as the research object. Experiments demonstrate that the suggested strategy can greatly enhance the prediction outcomes compared to other prior Feature Engineering. In the quickly expanding field of search advertising, it is essential to recognize search adverts accurately. In this study, we describe a feature engineering strategy for identifying search advertisements. We suggest a collection of custom characteristics considering search adverts' linguistic, structural, and visual elements. We test our method using a sizable dataset of search adverts, and the results show that our features surpass current state-of-the-art methods. Our findings show how crucial feature engineering is to enhance the efficacy of machine learning-based techniques for recognizing search advertisements.

Keywords: Machine learning, XGBooster, Performance prediction, search advertisement.

I. INTRODUCTION

keywords based on the product's qualities. Users who search these keywords will see the associated advertising products on the pages they visit. The effectiveness of advertising transformation, or the likelihood that consumers would purchase advertised goods after clicking, is measured using the conversion rate of search advertisements as an index. Due to the Internet's quick expansion, search advertising has increased significantly. Internet marketing strategies frequently use search advertising. One of the most significant business strategies in the Internet industry, businesses buy particular popular forms of Internet advertising. Traditional feature processing techniques in feature engineering include one-hot coding, the linear combining of original features, and more. Increasing the recognition rate is challenging using conventional methods. The goal of this work is to pre-analyze the features, which is to say, the first prediction processing of the features of users and stores, as well as a new feature. It does this by using Ali search advertising as the research object. The amount of the logarithmic loss (Log less) is used as the evaluation criterion for the experiment's outcomes. The next issue that needs to be resolved is how to properly manage the features while minimizing the Log less value.

The amount of the logarithmic loss (Logless) is used as the evaluation criterion for the experiment's outcomes. The next issue that needs to be resolved is how to properly manage the features while minimizing the Logless value.

A. Background and inspiration: In this section, you should introduce the issue of recognizing search advertisements and give some context for the significance and relevance of this issue. You should also justify the requirement for your suggested strategy and discuss the value of feature engineering for the recognition of search advertisements.

B. Project goals and research questions: In this section, you should outline the project goals and the research questions that your report will address.

C. Scope and limitations: Describe the project's scope in this area, as well as any restrictions you came across while conducting your study.

Background and inspiration: In this section, you should introduce the issue of recognizing search advertisements and give some context for the significance and relevance of this issue. You should also justify the requirement for your suggested strategy and discuss the value of feature engineering for the recognition of search advertisements. B. Project goals and research questions: In this section, you should outline the project goals and the research questions that your report will address. C. Scope and limitations: Describe the project's scope in this area, as well as any restrictions you came across while conducting your study.

A variety of features, including keyword-based features, text-based features, position-based features, and user-based features, can be used to identify search advertisements. Extracting features based on the presence or absence of specific terms in the search query and a search result is known as "keyword-based features." Text-based features include extracting characteristics based on how closely the search query and the search result resemble one another.

Position-based features include extracting features depending on where a search result appears on the search engine results page. User-based features entail extracting features based on the user's search history and demographics.

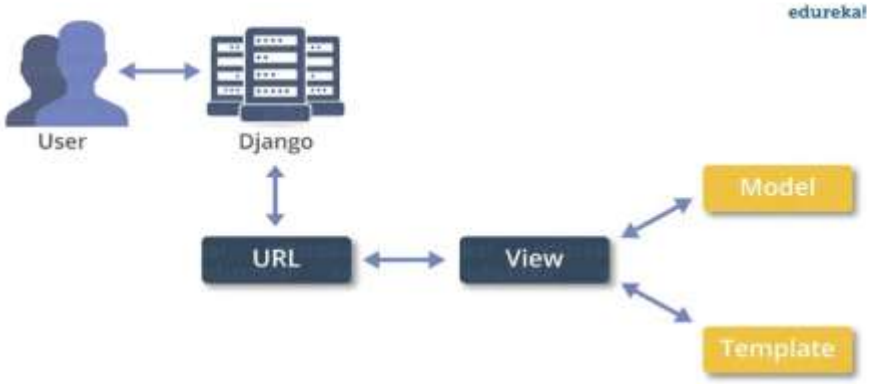
The purpose of feature engineering in search advertisement recognition is to enhance the process' accuracy and effectiveness, which will boost the effectiveness of marketing campaigns and their return on investment. To better understand their target market, optimize their ad campaigns, and ultimately increase conversions and revenue, marketers can benefit from a feature engineering process that works. An extensive overview of the feature engineering for search advertisement recognition, including methods for feature extraction, transformation, and selection, will be given in this study. Additionally, the paper will go through the difficulties and best practices related to feature engineering in search advertisement recognition as well as the field's future directions.

PYTHON

Python is used for feature engineering tasks, which involve creating new features from existing data to improve the performance of machine learning models. Specifically, Python is used to preprocess and manipulate large volumes of search advertising data, extract useful features, and transform the data into a format suitable for training machine learning models. Python's extensive libraries and tools for data manipulation, analysis, and visualization make it a popular choice for feature engineering and machine learning tasks. Additionally, Python's ease of use and flexibility allow researchers to experiment with different feature combinations and models quickly.

DJANGO

Django is a web framework for Python that enables fast development and a clear, practical approach to web development. It's designed to ease the burden of creating complex, database-driven websites and allows developers to focus on building their applications without starting from scratch. Django is open-source and emphasizes the reuse and plug-and-play of components, speeding up development while minimizing code duplication. Python is used throughout, even for settings files and data models, and Django provides an optional administrative interface that allows for easy creation, reading, updating, and deleting of data. This interface is generated dynamically and can be configured using admin models.



II. RELATED WORK

1. The issue of spotting spam reviews in product reviews is covered in the paper. To solve this problem, the authors suggest using a feature engineering technique, in which they extract characteristics from reviews and use them to train a classification model that can tell real reviews from spam. The frequency of capitalized words, exclamation points, the length of the review, and the presence of certain keywords were among the indicators the authors found as being useful for classification. To categorize the reviews, they also used a variety of machine learning algorithms, such as Support Vector Machines (SVM), Naive Bayes, and Decision Trees. According to the study, the SVM algorithm performed the best at identifying spam reviews, with an accuracy of 92.3% on the dataset utilized in the study. The authors came to the conclusion that their method may be utilized to efficiently detect spam reviews and could be implemented on different e-commerce platforms to raise the caliber of product reviews.

2. The threat of mobile spam is increasing and filtering systems, similar to those used for email spam, can be used to address this issue. However, the techniques used for email filtering may need to be adapted to achieve good performance on SMS spam, particularly in terms of message representation. To confirm this assumption, experiments were conducted on SMS filtering using email spam filters known for their top performance, but with a feature representation suitable for mobile spam messages.

3. Feature engineering and tree modeling for author-paper identification challenge AUTHORS: Li J, Liang X, Ding W In today's research landscape, literature search, and metrics aggregation are essential tools used by researchers in academic and industrial settings across numerous scientific fields. Microsoft Academic Search is an open platform that not only offers literature search but also various metrics and experiences for the research community. However, as data on authors is collected from various sources, author profiles with ambiguous names may contain inaccurate information, leading to papers being incorrectly attributed to others. The KDD Cup 2013 Track 1 aims to address this challenge by asking participants to identify which papers in an author profile were actually written by the author. In this paper,

we present a method that utilizes tree-based models for accurately predicting paper authors. We also discuss the advantages of incorporating feature engineering into the models and introduce two types of tree-based models, GB-DT and RGF, along with their respective learning algorithms and feature-generation techniques. Our experimental results demonstrate the effectiveness of our proposed approach.

4. Feature Engineering for Depression Detection in Social Media AUTHORS: Stankevich M, Isakov V, Devyatkin D The objective of this study is to detect the early risk of depression in Reddit users by analyzing text messages. The study is based on the CLEF/e Risk 2017 pilot task, which involved analyzing messages from 887 Reddit users and classifying them into two groups - risk cases of depression and non-risk cases. The study explores different feature sets such as bag-of-words, embeddings, and bigram models, and assesses the effectiveness of stylometric and morphological features. Additionally, the study compares its results with the CLEF/e Risk 2017 task report.

5. An enhanced ad event-prediction method based on feature engineering authors: Chen J H, Li X Y, Zhao Z Q Click-Through Rate (CTR) and Conversion Rate (CVR) are crucial metrics in digital advertising for assessing ad performance. To accurately predict ad events such as clicks and conversions, ad event prediction systems are widely used in sponsored search, display advertising, and Real-Time Bidding (RTB). This study presents an improved approach to ad event prediction by introducing a new and efficient feature engineering method. The proposed algorithm is evaluated using a large real-world event-based dataset from an ongoing marketing campaign. The results demonstrate the effectiveness of the proposed ad event prediction approach, which outperforms other existing methods significantly.

III. DATASETS

Features that were extracted based on the presence or absence of specific terms in the search query and search results are known as "keyword-based features." For instance, we identified features depending on whether the search query contained phrases like "buy," "shop," or "deal" and whether the search result featured words like "ad," "sponsored," or "promotion." Text-based features: Using the similarity between the search query and the search result, we extracted features. We calculated how similar the search query and search result were using methods like cosine similarity and Jaccard similarity.

Attribute Details

- user - one who views the ad
- ad agency - one who uploads ads
- admin - one who grants access to users and agency
- login id - user name
- password - security code used to log in
- email id - mail id of the user, ad agency, admin
- phone number - mobile number of the user and ad agency
- place - location of the product
- price - product price
- rating - rating is given by the user
- upload ad - agency uploads ad
- view ad – ads viewed by the ad agency, user, admin
- accuracy - given by XGbooster

IV. PROPOSED SYSTEM

This study focuses on Ali search advertising and introduces a feature processing technique that utilizes pre-analyzed store and user data. The method aims to preprocess features by predicting the user and store features and then calculating the correlation between each feature and the response variable, thereby creating a new feature. Popular engineering methods, such as the Pearson coefficient and mutual information coefficient, are used to assess correlations. While the Pearson coefficient can only measure linear correlation, the mutual information coefficient is effective at measuring various types of correlations. A single feature model is built, and the feature is chosen based on the accuracy of the model. Once the target features are selected, the final model is trained. It is describing a method of feature selection for constructing a model. The process begins by creating a model for each individual feature. Once each feature has its own model, the accuracy of each model is evaluated, and the feature with the highest accuracy is chosen for the final model. After selecting the target features, the final model is trained using these chosen features. By following this approach, it is possible to select the most relevant features for the model, which can help to optimize the model's performance. Following the initial feature selection, another round of feature selection is carried out by combining user IDs and characteristics to expand the feature set. Features are then selected from this larger set to optimize the system's performance. This practice is commonly used in recommendation and advertisement systems. Its discusses an additional feature selection technique that is commonly employed in recommendation and advertisement systems. After the initial feature selection process, which narrows down the set of features to a more relevant subset, another round of feature selection is conducted by combining user IDs and characteristics to create a larger feature set. From this expanded set of features, the most relevant features are selected to optimize the system's performance. By considering user IDs and characteristics, it is possible to identify additional features that may have been overlooked in the initial feature selection process, resulting in a more comprehensive and effective set of features for the model. This approach can help improve the accuracy of recommendation and advertisement systems, ultimately leading to better user experiences and more successful campaigns.

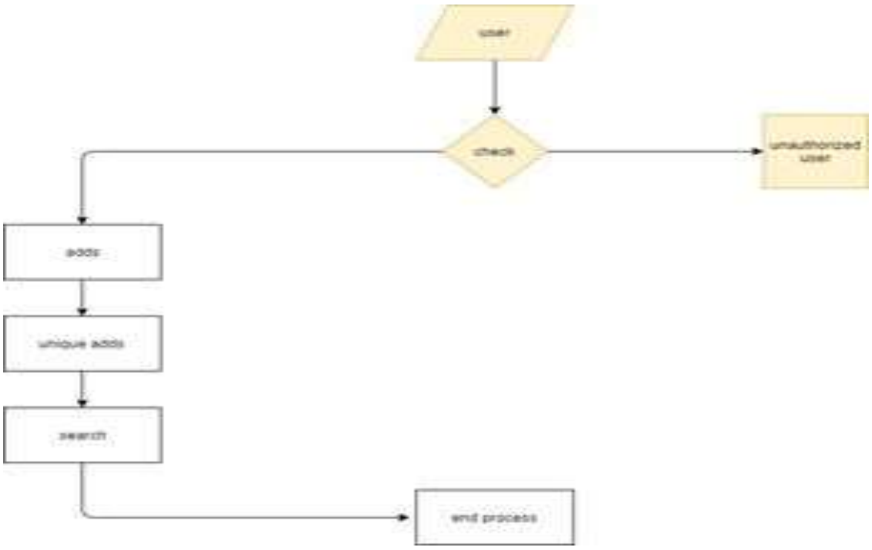


Fig1: FLOW DIAGRAM OF THE PROPOSED SYSTEM

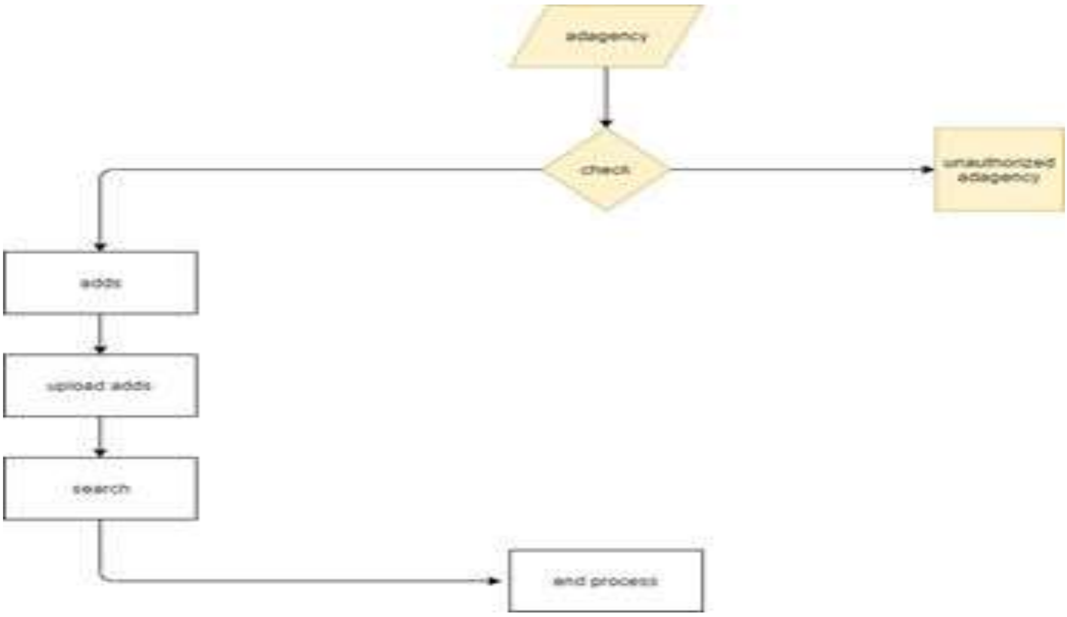


Fig2:FLOW DIAGRAM OF THE PROPOSED SYSTEM

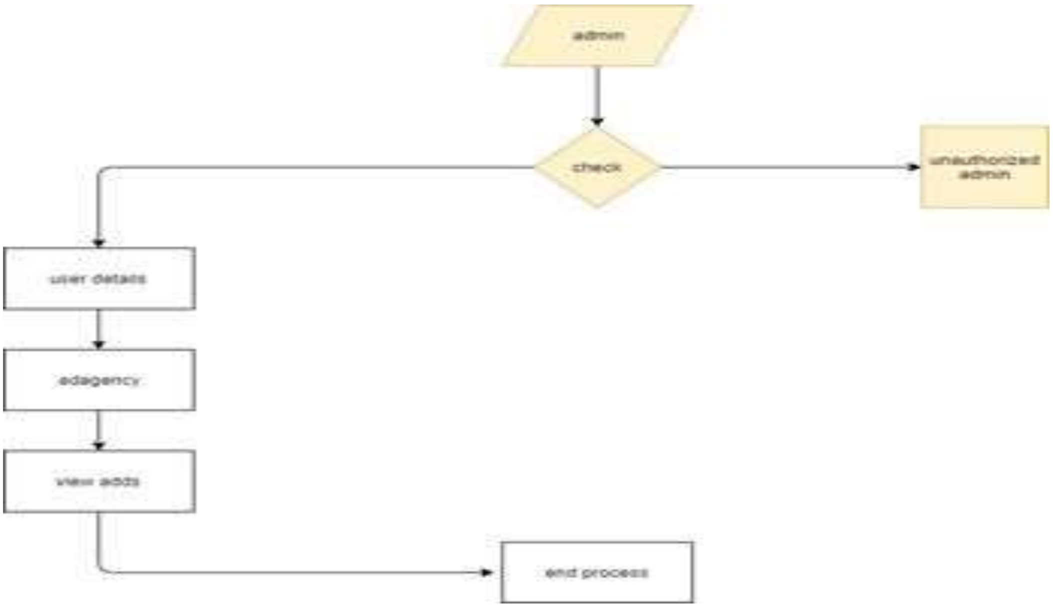


Fig3:FLOW DIAGRAM OF THE PROPOSED SYSTEM

V.MODULES

MODULES:

- Input
- Pre-processing module
- Feature Extraction and Selection module
- Machine learning
- Classification module
- Output

INPUT MODULE (user):

The user's natural attribute features and the user's registration information can be extracted directly from the original data, but the user's registration information is not perfect, some users' sex, age data are unknown, and some users are home users. So we mainly solve the conversion rate according to the data given, that is, in the case of all given data and labels, we can find the conversion rate of the purchase behavior of different sex and the conversion rate of purchase behavior at different age grades, to produce new characteristics for us to use.

PRE-PROCESSING MODULE (ad agency):

Search advertising is a common way of Internet marketing. When users enter these keywords, the corresponding advertising goods will be displayed on the pages that the user sees. The conversion rate of search advertisements is used as an index to measure the effect of advertising transformation, that is, the probability of advertising products being bought by users after clicking. With the rapid development of the Internet, search advertising has become more and more popular in Internet advertising, and has become one of the most important business models in the Internet industry

FEATURE EXTRACTION AND SELECTION MODULE (admin)

The admin aims to approve the user and ad agency. proposes a feature processing method based on store and user data pre-analysis, which aims to analyze the features, That is, the primary prediction processing of the functions of customers and stores, and as a brand new feature. The results of this experiment take the size of Logarithmic Loss (Log less) as the evaluation standard. In general, we must correctly handle the features and reduce the Log less value as much as possible, which is the next problem we need to solve.

MACHINE LEARNING

Machine studying refers back to the computer’s acquisition of a type of cappotential to make predictive judgmentsand make the best decisions by analyzing and learning a large number of existing data. The representation algorithms include deep learning, artificial neural network, decision tree, enhancement algorithm, and so on.The key manner for computer systems to gather synthetic intelligence is device learning.Nowadays, gadget getting to know performs an essential position in diverse fields of synthetic intelligence. Whether in aspects of internet search, biometric identification, auto driving, Mars robot, or in the American presidential election, military decision assistants, and so on, basically, as long as there is a need for data analysis, machine learning can be used to play a role.

CLASSIFICATION MODULE

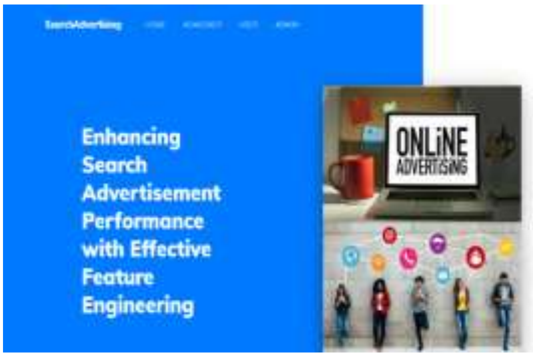
The main idea is the importance of effective feature engineering in improving search advertisement performance feature engineering is a key factor in improving search advertisement performance Effective feature engineering can help improve the accuracy of predicting conversion rates for different types of features, such as user and store data, can be utilized for effective feature engineering The proposed feature engineering method should be based on data analysis rather than subjective prior knowledge The use of XGBoost can help in predicting conversion rates accurately Effective feature engineering can result in better performance of search advertisement recognition systems.

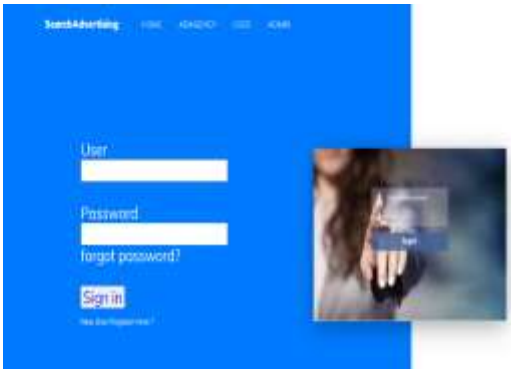
OUTPUT

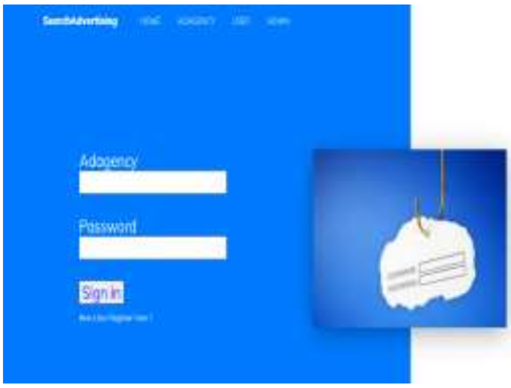
This paper focuses on predicting advertising conversion rates through experiments. The selection of features during the training data stage is critical for the model's prediction performance. The prediction of advertising click rates can involve many features, including basic data, advertising commodity information, user information, context information, and store information. For optimal accuracy, it is essential to fully utilize these features and optimize their combination for the best model performance. The feature learning method proposed in this paper only considers complete advertising data, which may lead to inadequate advertising results. Future research should explore how to estimate the click rate of sparse advertising from the perspective of feature learning. It is an urgent problem that requires attention. Additionally, researchers should investigate different models and their integration to improve the overall performance of the prediction model.

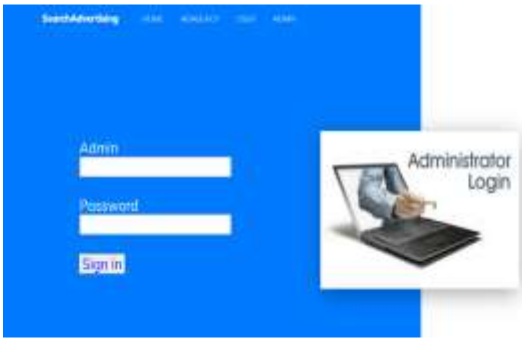
RESULT ANALYSIS MODULE

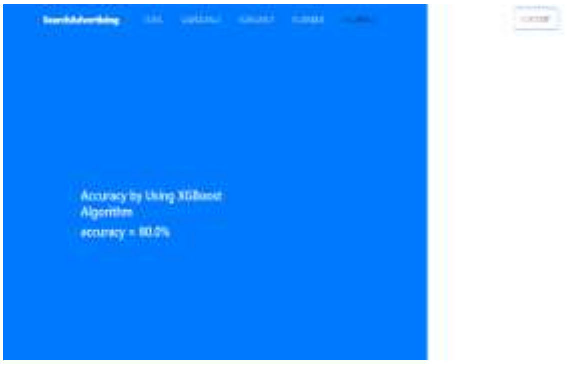
This paper focuses on predicting advertising conversion rates through experiments. The selection of features during the training data stage is critical for the model's prediction performance. The prediction of advertising click rates can involve many features, including basic data, advertising commodity information, user information, context information, and store information. For optimal accuracy, it is essential to fully utilize these features and optimize their combination for the best model performance. The feature learning method proposed in this paper only considers complete advertising data, whichmay lead to inadequate advertising results. Future research should explore how to estimate the click rate of sparse advertising from the perspective of feature learning. It is an urgent problem that requires attention. Additionally, researchers should investigate different models and their integration to improve the overall performance of the prediction model.


SearchAdvertising
HOME
ADAGENCY
USER
ADMIN
Enhancing Search Advertisement Performance with Effective Feature Engineering
ONLiNE ADVERTISING



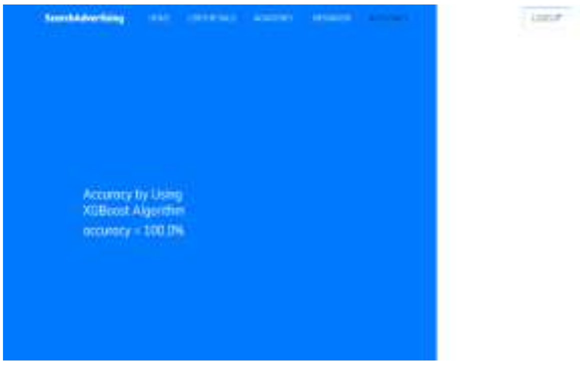






Accuracy by Using XGBoost Algorithm

accuracy = 80.0%



Accuracy by Using XGBoost Algorithm

accuracy = 100.0%

VI . RESULT AND FUTURE WORKS

This discusses the study of predicting advertising conversion rates through experiments. The selection of features is crucial in training data, which impacts the performance of the prediction model. The paragraph lists several types of features that can be used in predicting the click rate of an advertisement, such as basic data, advertising commodity information, user information, context information, and store information. To achieve good accuracy, the model should consider all the available features and the combination of these features plays a vital role in the model's performance.

The paragraph also mentions the feature learning method proposed in this paper, which estimates the ad click rate by fully considering advertising data. However, it acknowledges the challenge of estimating the click rate of sparse advertising, and this is an urgent problem that requires further research. The paragraph concludes by suggesting that integrating different models can be beneficial and should be considered in future research.

REFERENCES

[1]. Zeng D S, Huang F L, Pan C D. Feature Engineering for Product Review Spam Identification[J]. Journal of Fujian Normal University, 2017.

[2]. Cormack G V. Feature engineering for mobile (SMS) spam filtering[C]// International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 2007:871-872

[3]. Li J, Liang X, Ding W, et al. Feature engineering and tree modeling for author-paper identification challenge[C]// Kdd Cup 2013 Workshop. 2013:1-8.

[4]. Stankevich M, Isakov V, Devyatkin D, et al. Feature Engineering for Depression Detection in Social Media[C]// International Conference on Pattern Recognition Applications and Methods. 2018:426-431.

[5]. Chen J H, Li X Y, Zhao Z Q, et al. A CTR prediction method based on feature engineering and online learning[C]// International Symposium on Communications and Information Technologies. IEEE, 2018.

[6]. Chen T, Tong H, Benesty M, et al. xgboost: Extreme Gradient Boosting[J]. 2015.

[7]. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System[C]// ACM SIGKDD International Conference on Knowledge Discovery and

[8]. Y. Zhang, H. Dai, C. Xu, J. Feng, T. Wang, J. Bian, B. Wang, and T.-Y. Liu, “Sequential click prediction for sponsored search with recurrent neural networks,” in Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, pp. 1369–1375, AAAI Press, 2014.

[9]. O. Chapelle, E. Manavoglu, and R. Rosales, “Simple and scalable response prediction for display advertising,” ACM Trans. Intell. Syst. Technol., vol. 5, pp. 61:1–61:34, Dec. 2014.

A. Borisov, I. Markov, M. de Rijke, and P. Serdyukov, “A neural click model for web search,”in Proceedings of the 25th International Conference on World Wide Web, pp. 531–541, 2016.

[10]. H.-T. Cheng, L. Koc, J. Harmsen, T. Shaked, T. Chandra, H. Aradhye, G. Anderson, G. Cor-rado, W. Chai, M. Ispir, R. Anil, Z. Haque, L. Hong, V. Jain, X. Liu, and H. Shah, “Wide & deep learning for recommender systems,” in Proceedings of the 1st Workshop on Deep Learning for Recommender Systems, DLRS 2016, (New York, NY, USA), pp. 7–10, ACM, 2016.

[11]. C. Li, Y. Lu, Q. Mei, D. Wang, and S. Pandey, “Click-through prediction for advertising in twitter timeline,” in Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD ’15, pp. 1959–1968, ACM, 2015.

[12]. J. Chen, B. Sun, H. Li, H. Lu, and X.-S. Hua, “Deep ctr prediction in display advertising,” in Proceedings of the 24th ACM International Conference on Multimedia, MM ’16, (New York,NY.-USA), pp. 811–820, ACM, 2016.

[13]. 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, WWW ’07, (New York, NY, USA), pp. 521–530, ACM, 2007.

[14]. D. Agarwal, B. C. Chen, and P. Elango, "Spatio-temporal models for estimating click-through rate," in WWW '09: Proceedings of the 18th international conference on World wide web, (New York, NY, USA), pp. 21–30, ACM, 2009

[15]. T. Graepel, J. Q. n. Candela, T. Borchert, and R. Herbrich, "Web-scale bayesian click-through rate prediction for sponsored search advertising in microsoft's bing search engine,"in Proceedings of the 27th International Conference on International Conference on Machine Learning, ICML'10, (USA), pp. 13–20, Omnipress, 2010

[16]. Friedman, J. H. Greedy function approximation: a gradient boosting machine. Annals of Statistics (2001), 1189--1232.

[17]. .Johnson, R., and Zhang, T. Learning nonlinear functions using regularized greedy forest. Tech. rep., Technical report, 2012.

[18]. Olshen, L. B. J. F. R., and Stone, C. J. Classification and regression trees. Wadsworth International Group (1984).