Predict the Relevance of Search Results from E-Commerce Sites

Mukul Singhal[1], Dikshant Awasthi[2], Parshvi Verma[3], Sumit Kumar Malik[4]
[1,2,3,4] *Department of Computer Engineering, Raj Kumar Goel of Institute of Technology*

Abstract **- In today era Ecommerce is very essential for all of us. Large Ecommerce Sites such as amazon, flipkart etc. plays an important role like we want to purchase a product from amazon then we can search a product that we want to buy. So, search engines play an important role in ecommerce sites so that customer can see their products easily. In this review paper we can give a robust way of predicting scores given a search query and a product using different techniques like machine learning algorithms, deep learning algorithms and natural language processing and information retrieval.**

Index Terms **- E-commerce, Online Shopping, Products, Query-Based Search, Search Engines.**

I.INTRODUCTION

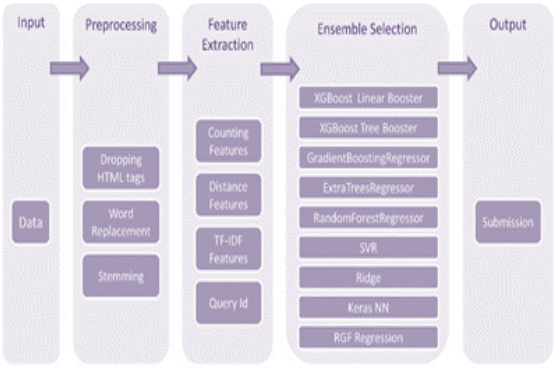
Large Ecommerce Sites typically used query-based search to find information from the websites. They are able to use technology to provide user with better experience. Because they understand the importance of search relevance and search engines provides a better way to reach the product easily. So, our project aims to rank those product items which are not only relevant to user query but also aims those items which has higher probability to be purchased by the user. Here is where we are going today: Give only as raw text as input, our goal is to predict the relevancy of the results from ecommerce sites. As a result, help them to improve their shopping experience.

In this work we can use some machine learning algorithms and implementation of preprocessing methods to predict the text search relevance. The paper describes some basic concepts about data preprocessing towards feature extraction. At the end we can show some results that come out from machine learning algorithms and feature extraction as well [2].

1.1 PROPOSED WORK

In proposed work model data is taken as input.

After input part, for preprocessing we mainly perform some html tags dropping a word replacement and stemming part. After preprocessing, Feature extraction is used for counting features, distance features, Tf-Idf features and query id. After feature extraction, we can use some ensemble selection methods. In Ensemble Selection Method, XGBoost Linear Booster, XGBoost Tree Booster, Gradient Boosting Regressor, Extra Trees Regressor, Random Forest Regressor, SVR, Ridge, Keras NN, RGF Regression. Python language is used as a programming language in this project [1].



Flow Chart of the Work The measurement of the separate model and the ensemble model [2]

II. REVIEW SURVEY

Many of our daily activities are carried out by many search algorithms. Whether we were trying to find a stream of any reality TV show or shopping on eCommerce site for a new set of things, the relevance of search results is often responsible for your happiness. Currently, small online businesses do not know how to increase the performance of their search algorithms, which were downgrading their customer experience. many of retailers use query-based search to help consumers find information/products on their websites. They make use technology to provide users with a better experience. As they understand the

importance of search relevance, and that long or unsuccessful searches can turn their users away because users are accustomed to and expect instant, relevant search results like they get from Google and Amazon. While search is critical to the success of any eCommerce business, it is not always as easy as it seems, in particular, for middle or small online retailers, because it requires huge amount of manually labelled data and machine learning techniques. The improvement to searching does not limited to the exact match. Machine learning help in displaying the related result which may increase the sale of the products.

Older, ancient web site searches in area referred to as "recommender" or "product recommendation" searches. They need very little imagination and deliver solely results centered on the keyword. Some search apps cannot perceive misspelled words returning no results in the last. The client should try hard once more or go else wherever. In the SLI study, seventy-three of shoppers left a web site once in two minutes if they had not found what the searches increased with machine learning came back a wider selection of results to every question. They map merchandise and interconnect them in new ways in which. for example, a search for "cat food" returns pet-food wet, dry, mat, bowls, container, dispenser, were checking out. Mobile users were u even less patient. Product lid and canned. Adding one among those additional keywords can yield additional connected decisions. The program improves the search results supported the preferences clicked by customers [3].

III. INFERENCES DRAWN FROM LITERATURE

In this paper, we tend to square measure to showing the cycle and steps concerned for activity a data analysis on planet knowledge, from knowledge pre-processing till feature prediction. As delineate in the last section, such approaches may be effectively applied on great deal of datasets. Throughout this study, we tend to may testify scikit-learn package, similarly as different built-in Python packages, saves substantial development time. With our benchmark on C check set, we tend to may attest that Random Forest is associate degree efficient machine-learning algorithmic rule. It shown higher accuracy compared to a simple S.V.M implementation. this is often because of the ensemble nature of Random Forest, that permits multiple learning algorithms to be run. Beside this fact, the character of the data set is additionally another excuse for S.V.M disadvantage, since such algorithms offer poor result once the amount of feature is a lot of. Finally, we are able to check that the Pre-processing step is crucial for the full knowledge analysis. It additionally consumes most of the time and implementation efforts required on the full analysis. Moreover, our study additionally shown that Pre-processing is also a lot of essential for preciseness than the machine-learning algorithmic rule itself. Earlier within the analysis papers the accuracy that was calculated was concerning fifty-seven however in our model we are going to try and increase the accuracy by a minimum of 10-15 percent [4].

IV. CONCLUSION

1. When Benchmark is set on CrowdFlower test set, Random Forest can be attested in an efficient ML algorithm.
2. When compared to SVM implementation Random Forest has shown better accuracy. This in part is due to:
    a. The ensemble nature of Random Forest allows multiple algorithms to be run.
    b. Beside this fact, the nature of the dataset is also another reason for SVM disadvantage, since such algorithm is likely to provide poor performances when the number of features is much greater than the number of samples.
3. tf-idf is employed for preprocessing features and Random Forest is a powerful and effective approach for predicting and measuring the relevance text search in E-commerce scenarios.
4. Such approach suits small e-commerce businesses which have emerged in big data necessities.
5. The preprocessing is necessary for the whole data analysis because:
    a. It consumes most of the time and implementation effort which is needed in the whole analysis.
    b. Preprocessing is more critical for precision than the machine learning algorithm itself.

REFERENCES

Proceeding Papers:

[1] Predicting the relevance of search results for Ecommerce Systems Mohammed Zuhair Al-Taiel, Siti Mariyam Shamsuddin1, and Joel Pinho Lucas.
[2] Qiqi Wang E-commerce Sites search results relevance Prediction Based on Ensemble Approach.

Journal Papers
[1] https://towardsdatascience.com/predict-search-relevance-using-machine-learning-for-online-retailers-5d3e47acaa33
[2] https://www.eventige.com/blog/machine-learning-ecommerce-search