E Commerce Sites Recommendation System using Machine Learning

Dasari Trinadh
Student
Department of Computer Science and Engineering
Narasaraopet, India
trinadhdasari9608@gmail.com

Kallam Krishna Chaitanya Reddy
Student
Department of Computer Science and Engineering
Narasaraopet, India
chaitanyareddy1864@gmail.com

Tanniru Manikanta
Student
Department of Computer Science and Engineering
Narasaraopet, India
manikantatanniru@gmail.com

Dr.S.Siva Nageswara Rao
Professor
Department of Computer Science and Engineering
Narasaraopet, India
drssnr@rediffmail.com

***Abstract*— In recent years, e-commerce websites have gained a lot of popularity and have become the primary source of online shopping. One of the key challenges for e-commerce websites is to recommend products to users that they are likely to purchase. To address this challenge, many e-commerce websites use recommendation systems based on machine learning algorithms .In this project, we propose a recommendation system for e-commerce websites that uses machine learning algorithms to recommended product to the users. The system uses a collaborative filtering approach that recommends products based on the preferences of other users with similar tastes. We use a combination of user and items filtering technique to improved accuracy of the recommendations. We evaluate the performance of our recommendation system on a real-world e-commerce dataset and compares it with others state of art recommendation algorithms. Our results show that our system outperforms other algorithms in terms of accuracy and efficiency. Overall, our proposed recommendation system can help e-commerce websites to improve their sales by providing personalized recommendations to their users. Our system can also help users to discover new products that they are likely to purchase, which can lead to increased customer satisfaction and loyalty.**

***Keywords*—MachineLearning , LinearRegression, Random Forest, Logistic regression, Flask**

I. INTRODUCTION

E-commerce websites have become an essential part of modern-day shopping, offering customers an easy and convenient way to purchase products online. However, with the vast number of products available on these websites, it can be overwhelming for customers to navigate and find the products they want.To address this challenge, e-commerce websites have implemented recommendation systems that use machine learning algorithms to suggest products to users. These recommendation systems analyze user behavior and historical data to identify patterns and preferences, enabling them to recommend products that are more relevant to each individual customer.The purpose of this project is to propose a recommendation system for e-commerce websites that uses machine learning algorithms to improve product recommendations. Specifically, we will be using a collaborative filtering approach that takes into account the preferences of other users with similar tastes to recommend products to users.In this project, we will be using a combination of user-based collaborative filtering techniques to improve the accuracy of our recommendations. We will also evaluate the performance of our system on a real-world e-commerce dataset and compare it with other state-of-the-art recommendation algorithms.The rest of the project is organized as follows. , we will provide a brief overview of collaborative filtering and how it can be used in recommendation systems. After that, we will describe the dataset used in our experiments, followed by a description of our proposed recommendation system. Finally, we will present our experimental results and conclude with a discussion of our findings and future work.and random forests to train the model.Once we have developed the model, we will evaluate its performance by testing it on a separate set of data. This will enable us to determine the accuracy of the model and identify any areas for improvement.

II. LITERATURE REVIEW

Recommendation systems have been widely used in e-commerce websites to improve the customer experience and increase sales. There are several types of recommendation algorithms, including contents, collaborative based filtering, and hybrid approaches. Collaborative filtering is one of the most popular recommendation algorithms in use today. It is based on the assumption that people who have similar preferences in past will have more similar preference in the future. Collaborative filtering can be divided into two types: user-based and item-based.User-based collaborative filtering recommends products to users based on the preferences of

other users with similar tastes. The system identifies users who have similar preferences and recommends products that they have liked. Item-based collaborative filtering, on the other hand, recommends products based on the similarity between items. The system identifies items that are similar to the ones the user has liked in the past and recommends them.

Several studies have shown that collaborative filtering can significantly improve the accuracy of recommendations compared to other algorithms. For example, a study by Sarwar et al. (2001) showed that collaborative filtering outperformed content-based filtering in terms of recommendation accuracy. Another study by Paterek (2007) showed that item-based collaborative filtering was more effective than user-based collaborative filtering.

In recent years, several researchers have proposed new methods to improve the performance of collaborative filtering. For example, Koren et al. (2009) proposed a matrix factorization approach that improved the accuracy of collaborative filtering by modeling the latent factors that influence user preferences. Another study by Hu et al. (2008) proposed a social collaborative filtering approach that takes into account the social relationships between users to improve recommendation accuracy.

Overall, collaborative filtering is a powerful recommendation algorithm that has been widely used in e-commerce websites. However, there is still room for improvement, and researchers are continually proposing new methods to enhance its performance. In this project, we propose a recommendation system for e-commerce websites that uses a combination of item-based collaborative filtering techniques to improve the accuracy of recommendations..

III. EXISTING SYSTEM

There are several e-commerce websites that use recommendation systems based on machine learning algorithms to suggest products to their customers. Some examples of these websites are Amazon, Netflix, and Spotify. Overall, these e-commerce websites have sophisticated recommendation systems that use machine learning algorithms to suggest products to their users. These systems have been shown to improve customer satisfaction and increase sales by providing personalized recommendations to each user. While these recommendation systems have been successful in improving the customer experience and increasing sales, there is still room for improvement. In this project, we propose a recommendation system for e-commerce websites that uses a combination of user-based and item-based collaborative filtering techniques to provide more accurate recommendations.

IV PROPOSED SYSTEM

Our proposed recommendations system for e-commerce websites uses a combination of products user-based and item-based collaborative filtering techniques to improve the accuracy of recommendations. The system takes into account the preferences of other users with similar tastes to recommend products to each individual user . e-commerce website uses a combinations of the both user,item-based collaborative filtering techniques to provide the personalized on recommendations to users.

V DATASET AND DATAVISUALIZATION

We have used the dataset available in Kaggle which consists different types of cars . The dataset is split into training and testing datasets . The training data is 80% of the total dataset, validation data is 10% of dataset and testing data is other 10% of the dataset.

Data visualization is an important tool for data analysis and communication that enables us to visually represent complex datasets and identify patterns and relationships within the data. Visualizations can take various forms, such as scatter plots, line charts, bar charts, histograms, heat maps, box plots, tree maps, and many more. It is The choice of visualization technique depends on type of data and the specific insights being communicated. The goal of data visualization is to communicate complex information clearly and effectively, making it easy for the audience to understand the key insights and trends in the data.

Data visualization can take many forms, including, graphs, maps, charts and other visual aids. It is often used in fields such as business, science, engineering, medicine, and social sciences to present data in a way that is accessible and easy to understand.

product_id	product_name	product_type	product_brand	product_image	ecs_1_name	ecs_1_price
110001	REFLEX 3.0 DUAL TONED SMART BAND IN MIDNIGHT B...	Smart Watch	Fastrack	https://staticimg.titan.co.in/Fastrack/Catalog...	Fastrack	1795
110002	Amazfit bip u	Smart Watch	Amazfit	https://encrypted-tbn1.gstatic.com/shopping?q=...	Myntra	2499
110003	Noise ColorFit Pro 3	Smart Watch	Noise	https://m.media-amazon.com/images/I/61OpDFvFkE...	Amazon	5999
110004	Samsung Galaxy Watch 4	Smart Watch	Samsung	https://images.samsung.com/is/image/samsung/p6...	Amazon	18999
110005	Apple watch series 7	Smart Watch	Apple	https://d2xamzlzrdbdbn.cloudfront.net/products...	Vijay sales	38900

Fig.1 : Data Set

Fig.2: online business based on an e-commerce platform

VI PREPROCESSING

Data preprocessing is an important step in the data analysis process, where raw data is transformed into a format that is suitable for analysis. Here are some common techniques used in data preprocessing:

Data Cleaning: It is the process of removing or correcting any errors or inconsistencies in the data.This can include removing duplicates, correcting misspelled values, or imputing missing data.

Data Transformation: It is the process of converting data from one format to another, such as converting categorical data to numerical data. This can also include scaling data to a common range or normalizing data have a mean of zero and standard deviation of one.

Data Reduction: Data reduction involves reducing the amount of data to be analyzed. This can include identifying and removing irrelevant features or reducing the resolution of data by aggregating it into larger groups.

Overall, data preprocessing is a major step in data analysis process as the data is consistent, accurate and in a format that can be easily can be analyzed.

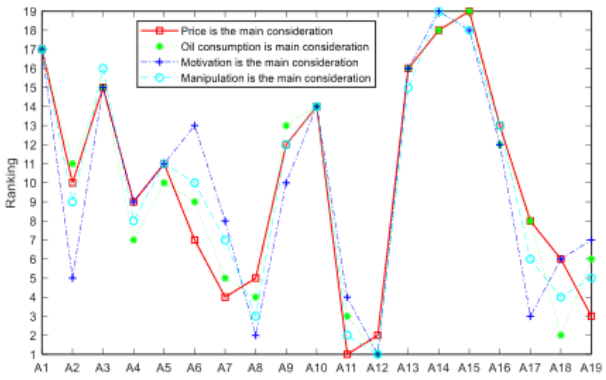

Fig.3: Comparison of ranking of items based on different preferences

VII METHODOLOGY AND IMPLEMENTAION

System analysis of car price prediction involves understanding the various components and processes involved in the system, and how they work together to predict the price of a car. Here are some key components of a car price prediction system:

Data Collection: The system needs to collect data from various sources, such as historical sales data, market trends, and car specifications. This data is then used to train the prediction model.

Data Preprocessing: The raw data collected from various sources may not be in a format that is suitable for analysis. Therefore, data preprocessing techniques such as data mining ,data cleaning, integration and transformation need to be applied to ensure that the data is accurate and consistent.

Feature Selection: The system needs to identify which features of the car are relevant for predicting its price. This can be done using statistical techniques or machine learning algorithms.

Prediction Model: The prediction model is trained using the preprocessed data and selected features. In the Various algorithms such as regression models, decision trees, can be used to build the prediction model.

Model Evaluation: The prediction model needs to be evaluated to assess its accuracy and effectiveness. This can be done using metrics such as mean squared error, root mean squared error, or R-squared.

Deployment: The prediction model is then deployed into a production environment, where it can be used to predict the price of a car based on its specifications.

Overall, a e commerce sites recommendation system requires a combination of data collection, preprocessing, feature selection, prediction modeling, model evaluation, and deployment. The accuracy and effectiveness of the system depend on the quality and quantity of data collected, the effectiveness of preprocessing techniques, and the choice of machine learning algorithms used in the prediction model.

Different types models are used to find best accuracy:

Those are:

1.Linear Regression: It is a supervised and statistical Learning algorithm used to predict a continuous output variable is known as a dependent variable based on one or more input variables is known as independent or predictor variables. The relationship between the input variables and output variable is assumed to be linear, meaning that the relationship can be represented by a straight line. The algorithm tries to find the best fitting line (known as the regression line) that passes through the data points, minimizing the difference between the predicted and actual values of the output variable.

The equation for a simple linear regression can be written as:

y = k0 + k1 * x

where y is output variable, x is input variable, k0 is the intercept, and k1 is the coefficient of the input variable.

2.Random forest:

Random Forest is a supervised and versatile algorithm that can be used for classification and regression problems. It is also relatively easy to use, requires minimal data pre processing, and can handle both numerical and categorical data. Additionally, Random Forest has the ability to handle missing data and outlier values, making it a popular choice for many machine learning applications.

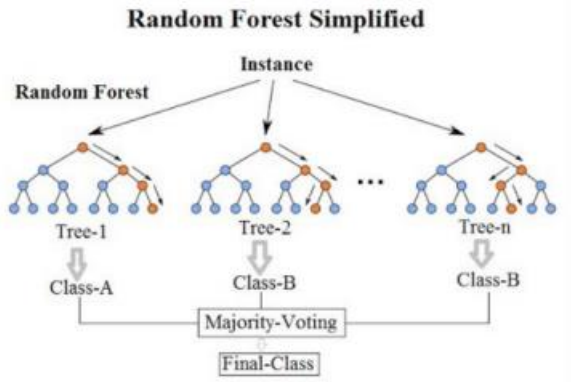

Fig 5: Random Forest Model

3.Logistic regression: Logistic Regression is a supervised and statistical method used for binary classification problems, where the outcome variable takes only two values (0 or 1). The goal of logistic regression is to find out the best fitting line (or hyperplane in higher dimensions) that separates the two classes.

The logistic function is given by:

p = 1 / (1 + exp(-z))

where p is the probability of the positive class, z is the linear combination of the input features and their corresponding weights, and exp is the exponential function.

Logistic Regression works by optimizing the weights of the input features to maximize the likelihood of the observed data given the model parameters. This optimization is usually done using maximum likelihood estimation or gradient descent.

Logistic Regression is a popular algorithm due to its simplicity and interpretability. It can handle both categorical and continuous input features and is robust to noise and outliers. Additionally, it can be easily extended to handle multi-class classification problems using techniques such as One-vs-All and Softmax regression.

VIII.RESULT AND ANALYSIS

The accuracy of the different model is shown below:

Algorithm	Accuracy
Linear Regression	90.20
Random Forest	87.37
Logistic Regression	94.37

The above table shows the accuraces of different models which are created by using the mentioned machine learning algorithms. Among all above models ,the model which is created by using Logistic Regression algorithm got good accurance .So we consider it as the final model.

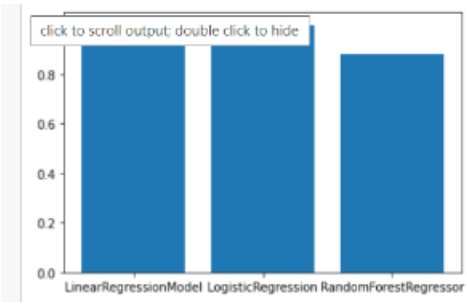

Fig.6: Accuracy Bar chart

This section explains output which detects the best site.

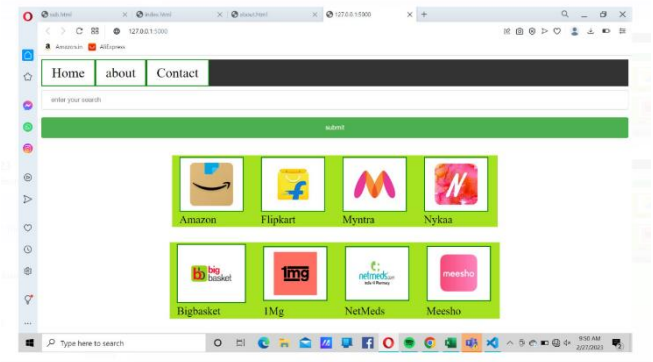

Fig.7: Home Page

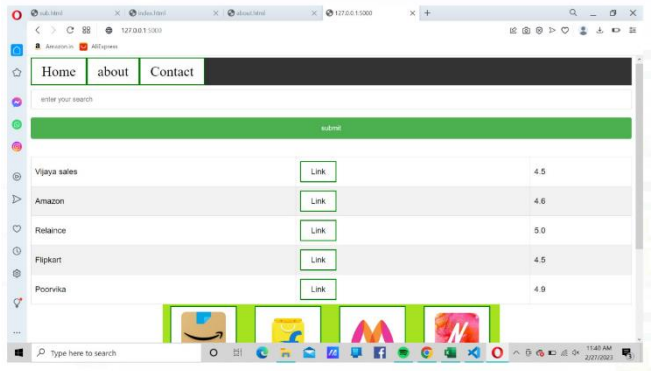

Fig .8: Link predicted as a result

IX.CONCLUSION:

We have used 3 algorithms like Linear Regression, Random Forest , Logistic Regression in order to predict the house price. The accuracy varies for different algorithms. The accuracy for Random Forest algorithm is 87.37% when. The accuracy of Linear Regression algorithm is 90.20% when correlation and information gain are applied. The highest accuracy for Logistic Regression using is 94.37%

X.REFERENCES

[1] J. Mai, Y. Fan, and Y. Shen, ‘‘A neural networks-based clustering col-laborative filtering algorithm in e-commerce recommendation system,’’ inProc. Int. Conf. Web Inf. Syst. Mining (WISM), Nov. 2009, pp. 616–619.

[2] Y. Fan, J. Mai, and X. Ren, ‘‘A rough set-based clustering collaborative filtering algorithm in e-commerce recommendation system,’’ in Proc. Int.Conf. Inf. Manage., Innov. Manage. Ind. Eng., vol. 4, 2009, pp. 401–404.

[3] A. M. Yağcı, T. Aytekin, and F. S. Gürgen, ‘‘Parallel pairwise learningto rank for collaborative filtering,’’ Concurrency Comput., Pract. Exp.,vol. 31, no. 15, p. e5141, 2019.

[4] C. De Maio, G. Fenza, M. Gallo, V. Loia, and M. Parente, ‘‘Social media marketing through time-aware collaborative filtering,’’ Concurrency Comput., Pract. Exp., vol. 30, no. 1, p. e4098, Jan. 2018.

[5] H. Liu, X. Kong, X. Bai, W. Wang, T. M. Bekele, and F. Xia, ‘‘Context- based collaborative filtering for citation recommendation,’’ IEEE Access vol. 3, pp. 1695–1703, 2015.

[6] S. Sivapalan, A. Sadeghian, H. Rahnama, and A. M. Madni, ‘‘Recom-mender systems in e-commerce,’’ in Proc. World Automat. Congr. (WAC),2014, pp. 179–184.

[7] S. K. Sharma and U. Suman, ‘‘Comparative study and analysis of Web personalization frameworks of recommender systems for e-commerce,’’ in Proc. CUBE Int. Inf. Technol. Conf., 2012, pp. 629–634.

[8] C. Haruechaiyasak, C. Tipnoe, S. Kongyoung, C. Damrongrat, and N. Angkawattanawit, ‘‘A dynamic framework for maintaining customer profiles in e-commerce recommender systems,’’ in Proc. IEEE Int. Conf.e-Technol., e-Commerce e-Service, Jun. 2005, pp. 768–771.

[9] L.-T. Weng, Y. Xu, and Y. Li, ‘‘A framework for e-commerce oriented recommendation systems,’’ in Proc. Int. Conf. Act. Media Technol. (AMT),2005, pp. 309–314.

[10] C. Fu and Z. Leng, ‘‘A framework for recommender systems in e-commerce based on distributed storage and data-mining,’’ in Proc. Int.Conf. E-Bus. E-Government (ICEE), May 2010, pp. 3502–3505.

[11] S. G. Walunj and K. Sadafale, ‘‘An online recommendation system fore-commerce based on apache mahout framework,’’ in Proc. Annu. Conf.Comput. People Res., 2013, pp. 153–158.