



IJIRCCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 7, July 2022

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.165



9940 572 462



6381 907 438



ijircce@gmail.com



www.ijircce.com

Product Recommendation System Using Sentimental Analysis

Priyanka K.R

4 th sem, M.Tech (CSE), Dayananda Sagar College of Engineering, Kumarswamy Layout, Bengaluru, India

ABSTRACT: As one of the fastest growing research areas in computer science, it can be difficult to keep up with everything that is happening in this field. As a result of user reviews of this system, the surrounding world changed its opinion on a particular product. Opinion mining, text mining and sentiment analysis are all used in this system. This research is based on internet product reviews from various sources. This system compares the sentiments expressed in the retrieved reviews. Other rating models are used Naive Base and Decision Tree to rank reviews. The "score" or "priority" that a user assigns to a particular item is evaluated by a recommendation engine belonging to the category of information filtering systems. Customer preferences are used to identify information designs in the database and provide results tailored to customer needs.

KEYWORDS: sentiment analysis, product reviews, reviews, recommendation, score, priority

I. INTRODUCTION

The mobile phone industry is growing at a tremendous rate as a result of the growing demand for smart phones. Keeping in mind the rapid development of the industry this research has to take a comprehensive look at the brand and smart phone model. There are many famous brands in the market, but only a few of them have a significant market share. For example, Samsung,

Apple and the like are famous brands worldwide. E-commerce has a significant impact on the growth of mobile phone sales and consumer buying habits. As a resource for making informed purchasing choices, customer reviews found on these e-commerce sites are invaluable. Reviewers on retail sites like Amazon.com have a variety of options on how to write their reviews. For example, a customer can give a product a numerical rating from 1 to 5 or write a review.

Since there are so many items from so many different manufacturers, it is important to provide helpful buyer reviews. Keeping track of all the reviews about a product or brand is becoming increasingly difficult, no less than dealing with big data. Biasing the review towards positive or negative sentiment improves judgment by rating reviews based on user sentiment.

Since there are many products in the market from many different companies, it is essential that customers have access to reviews that are relevant to their needs. The alarming increase in the number of reviews for a product or brand is nothing short of big data management. Categorizing reviews into positive and negative sentiments provides a better understanding of review sentiments, which leads to better decision making.

Sentimental Classification:

Customers may publish product reviews on e-commerce websites, which has increased the popularity of the industry. For product producers, keeping track of their customers' opinions is challenging given the millions of evaluations that are posted by consumers every day. In order to extract usable information from vast volumes of data, it is important to classify big and complicated data sets. Such issues need the use of classification methods. Data are divided

into multiple groups based on shared traits in a categorization process. Companies often worry about the difficulties of automating the categorization process when dealing with enormous data sets.

Finding and extracting personal information from text sources is referred to as sentiment analysis or opinion mining in natural language processing (NLP). Sentiment rating's objective is to analyze written user evaluations and classify them into favorable or unfavorable assessments so that the system does not need to precisely grasp each word's or document's meaning.

It's not that easy to designate words as good or negative. There are several challenges to be met in this situation. It may not always be successful to classify words and sentences according to their previous positive or negative polarity. Following a negative word like "no" will make it simple to identify the positive implications of a word like "great." Although unexpected cameras have a negative connotation, tourists like the feeling of uncertainty. This approach has been used in a variety of sectors, including vacation planning and movie reviews, to rate people's emotions. Lexical-based techniques and machine learning techniques are two of the most used techniques for emotion categorization.

II. RELATED WORK

Due to the many online evaluations, sentiment analysis has lately gained a lot of interest. As a consequence, several studies have been conducted on this field of study. This section presents some of the research findings that are pertinent to this thesis. SVM was put to the test for text classification by Joachims (1998), who found that it performed better than competing approaches in all trials with low error rates. Pang, Lee, and Vaithyanathan (2002) investigated supervised learning on categorizing movie reviews into two categories, positive and negative, using SVM, naïve basis, and maximum entropy classification. The accuracy of all three methods gave very good results. Several features were investigated in this study, and it was discovered that the machine learning algorithms performed better when the bag of words was employed as a feature in those classifications. In a recent research by Ye et al., three supervised machine learning algorithms—naive base model, SVM, and N-gram—were evaluated using Internet reviews of different tourism sites worldwide. According to this study, appropriately trained machine learning algorithms excel in evaluating travel destinations.

Additionally, they demonstrated that in terms of outcomes, SVM and N-gram models beat the simplistic base strategy. However, the disparity between the algorithms rapidly diminished when additional training data sets were included. The supervised machine learning technique outperformed the unsupervised method in Chaovalit and Zhou's 2005 comparison of semantic orientation, an unsupervised approach to movie appraisal. The two approaches that are reportedly utilized the most commonly to address sentiment classification issues are naive base and SVM (Joachims 1998; Pang et al. 2002; Ye et al. 2009). As a result, this thesis tries to apply the supervised machine learning algorithms Naive Bayes and SVM to Amazon beauty product evaluations

III. THE METHODOLOGY

1. Data collection

Based on Amazon dataset information, the internet the most popular site is amazon.com. The website has unusually large amount of product reviews. From the site, 28,000 reviews were randomly selected. On amazon.com, reviews are rated from 1 to 5. For each rating system, 1,000 ratings were selected. Later, interpret the numbers 1 and 2 as negative and the numbers 4 and 5 as positive in both experiments to categorize polar feelings. Neutral comments are collected as data with a three-point rating. Then, ranked 1 and 2 as negatively biased, 3 as neutral, and 4 and 5 as positively biased for the three polarized emotion rating tests. Finally, give 1 to indicate more, 2 to still be negative, 3 to be neutral, 4 to be even positive, and 5 to be very positive in the context of an empirical study of emotion 5 polarity.

2. Data preparation

Real-world data is often incomplete because it lacks attribute values, contains only some desirable feature, or all of the information. Noisy: Presence of errors or outliers. Inconsistency: Presence of inconsistencies in names or symbols.

To effectively deal with data, it is necessary to understand the idea of missing values. If missing values are not appropriately addressed the researcher risks making incorrect conclusions about the data. A researcher's results differ when there are missing data due to improper handling. Null values are often processed using data preprocessing technology.

In this instance, a specific row with a null value for a specific feature and a specific column with more than 75% of missing data must be eliminated. Only when the data set has a sufficient number of samples is this approach advised. It should be assured that no bias is applied once the data has been removed. Information is lost when data are removed, which hinders output estimate from producing the required results.

3. Remove missing values

This technique can be used in a feature that contains numerical information, such as a year column or a major team goal column. To fill in missing data, we can calculate the mean, median, or mode of a feature. This is an estimate that leads to variance in the data set. But this approach, which produces better results than removing rows and columns, can reduce data loss. The above three assumptions can be replaced by a statistical method to address missing variables. This method is also known as training data leakage. It can also be estimated using the difference of nearest values. The simpler the data, the better it works.

4. Choosing of learning algorithms

In this step the optimal learning algorithm is investigated. Depending on the type of data we have and the type of problem to be solved. Classification algorithms are used when there is a classification problem and the data is classified. Regression analysis is a work in progress and regression algorithms are used if the data is categorical. Grouping methods are used when the goal is to create groups and the data is unlabeled.

FEW LEARNING ALGORITHMS:

Logistic regression:

One of the most popular machine learning algorithms in the supervised learning space is logistic regression. They are used to forecast the categorical dependent variable using a specified set of independent variables. Logistic regression is used to estimate the output for dependent variables that are categorical. The outcome must thus be a discrete or categorical value. It delivers probability values between 0 and 1 rather than absolute numbers between 0 and 1. They may be true or false, 0 or 1, yes or no, etc. Except for how they are used, logistic regression and linear regression are quite similar. Classification and regression issues are solved using logistic regression and linear regression, respectively.

In logistic regression, we fit an "S"-shaped logistic function that approximates two peak values instead of the regression line (0 or 1). A fundamental strategy in machine learning, logistic regression may categorize new data using both continuous and discrete data sets.

Decision Tree Algorithm

An approach to supervised learning called a decision tree is mostly used to solve classification problems, but it is also used to solve regression problems. Both categorical and continuous variables are suitable for this. It exhibits a tree-like structure with nodes and branches growing on other branches starting at the root node and ending at the leaf node. The base node is used to represent the characteristics of the data set, the branches show the decision-making procedures and the leaf nodes are the solutions to the problem.

Decision tree algorithms are used in real-world settings to distinguish between malignant and noncancerous cells, make recommendations for car purchases to customers, and other things.

Support Vector Machine Algorithm

Support vector machines, sometimes referred to as SVMs, are supervised learning techniques that may be used to classification and regression problems. But its primary use is classification problems. Building a hyper plane or decision boundary that can classify datasets is the goal of SVM. The support vectors, which are the data points that aid in defining the hyper plane, are what give the support vector machine approach its name.

Random Forest:

Popular machine learning algorithm RandomForest is an element of the supervised learning process. It may be applied to ML difficulties requiring both classification and regression. It is founded on the notion of ensemble learning, which is a way of merging several classifiers to handle complex difficulties and boost model performance. Random Forest, as the name indicates, is a classifier that utilizes a number of decision trees on distinct subsets of the input dataset and averages them to boost the dataset's prediction accuracy. Instead of relying on a single decision tree, the random forest takes forecasts from each tree and guesses the conclusion based on the votes of the majority of predictions.

Higher accuracy and overfitting are avoided by the increased number of trees in the forest.

5. Flask - web framework

A microweb framework built on Python is called Flask. Due to the lack of specialized tools or libraries, it qualifies as a micro framework. It lacks any elements, such as a database abstraction layer, form validation, or other elements, that pre-existing third-party libraries already offer for common functionality. However, Flask enables for the addition of app feature extensions. Several open authentication protocols, relational mappers, form validation, load management, and numerous utilities are all extensions for well-known frameworks.

Flask requires the Werkzeug WSGI toolkit and the full-featured Python template engine Jinja. completely. The Web Server Gateway Interface, or WSGI for short, describes how a Web server interacts with Web applications and how they can be linked together to execute a single request. Werkzeug is a complete framework for WSGI web application. Activate the appropriate environment before start working on project, set up the flask in an active atmosphere. The following combinations are registered in the Flask app, which is called as "the app":

- First import the Flask type. WSGI application will be a subcategory of this category.
- After that, make a class instance. The name of the application module or package is the first argument. For this, a useful abbreviated name that generally works. This is crucial since Flask needs to know where to search for resources like templates and static files.
- Next, point Flask to the URL where method should be called using the path designator ().
- This method returns a message to the user to be displayed in their browser. HTML is the default content type and is rendered by the browser.



- flask or python -m flask command is used to run the application. Before that, export the FLASK APP environment variable to point our terminal to the program that will interact with it.

IV. THE MODULES USED

Guide the flask apps

It is used to associate a specific URL with a function designed to perform a specific operation. In our project, we used the path() constructor to set various URLs, including: /home, /about, /recommend, /register, /login and /search.

URL Building

To construct a URL for a particular function, use the url() function. It accepts the function name as its first argument along with any number of keyword arguments, each of which corresponds to a separate variation section of the URL rule.

HTTP protocols

Web applications employ a variety of HTTP techniques to go to URLs. Different HTTP methods may be handled using the decorator's method parameter path(). If GET is available, Flask automatically adds support for the HEAD method and processes HEAD requests in accordance with the HTTP RFC. In our program, we've utilized the GET and POST methods to visit several urls, including the login, register, about, and refer ones.

Static files

Dynamic web applications also need static files. CSS and JavaScript files often originate from there. Although Flask can also achieve this during testing, the web server should be set up to serve us in an ideal scenario. Add a folder named static adjacent to package or module and accessible via the application/static directory. Under static folder saved main10.css and several jpeg files.

The request object

It must first be imported using a Flask unit. The method property may then be used to get the current request method. Utilize form property to get form data after that (information sent in POST or PUT request). afterwards utilized the sqlite3 database to assess the matching input data values after posting requests utilizing the request methods on the registration and login pages.

Errors and redirects

Use the redirect() and abort() functions to send the user to a different endpoint or terminate the request early with an error code.

Flashing indicator

Feedback is the foundation of all effective software and user interfaces. A user is likely to hate a program if it doesn't get enough feedback. With Flask's blink mechanism, users can get a really straightforward feedback. Use the flash() function to flash the message. This software uses flashing messages when a new user successfully registers on the website, when login is successful or even when login or registration fails.

WT Form validation

When user have to interact with the form data passed through the browser view the code becomes very difficult tounderstand very quickly. To facilitate this process, libraries are available. WTForms is one of them .When using WTForms, first user need to define forms as classes. As a result,create a login form, registration form and class homepage.

- If the data is submitted using the HTTPPOST method, construct the form from the value of the request form, and if it is submitted using the GET method, construct it from the args.
- To verify the data, use the verify() method. If the data is checked, it returns true; Otherwise it returnsfalse.
- Access form.NAME > .data to get specific values from the form.

The following steps in our proposed application describe how to process a proposed web application:

- 1) Start the procedure
- 2) The system is queried by an activeuser by entering search parameters and visitor types (single, even,business, etc.).
- 3) The system then searches the system database for information from previous users (ratings, ratings,reviews) that match the web query.
- 4) The system filters the query data by searching for matches in accessible external web sources.
- 5) Collect information if the result matches the query.
- 6) Metadata must be saved in a no SQL database.
- 7) Ignore if both do not match.
- 8) Continue until all matching metadata is found.
- 9) Conclusion.

V. RESULTS AND DISCUSSION

Results obtained using Logistic Regression:

Sentimentreviews	Precision	recall	F1-score
-1(negative)	0.67	0.02	0.04
0(neutral)	0.50	0.00	0.01
1(positive)	0.94	1.00	0.97

Results obtained using Random Forest Algorithm:

Sentiment reviews	Precision	Recall	F1-score
-1(negative)	0.69	0.16	0.26
0(neutral)	0.56	0.10	0.17
1(positive)	0.94	1.00	0.97

Results obtained using Decision TreeClassifier:

Sentiment reviews	Precision	Recall	F1-score
-1(negative)	0.51	0.09	0.16
0(neutral)	0.50	0.01	0.01
1(positive)	0.94	1.00	0.97

The accuracy score achieved using random forest tree is: 93.53 %

The accuracy score achieved using logistic regression is: 94.02 %

The accuracy score achieved using decision tree is: 93.56 %

As completed with User interface of this project with user login, website creation and Recommendation page for the Product view.

Home Page and Search Page:

This page of the web application, is the main page of the website, providing navigation to all other tabs of the website, such as the about page, register page, login page and recommendation page. User has to login using email and password, search for the product of their interest. The result is displayed using sentimental analysis and recommends product to user based on the result.

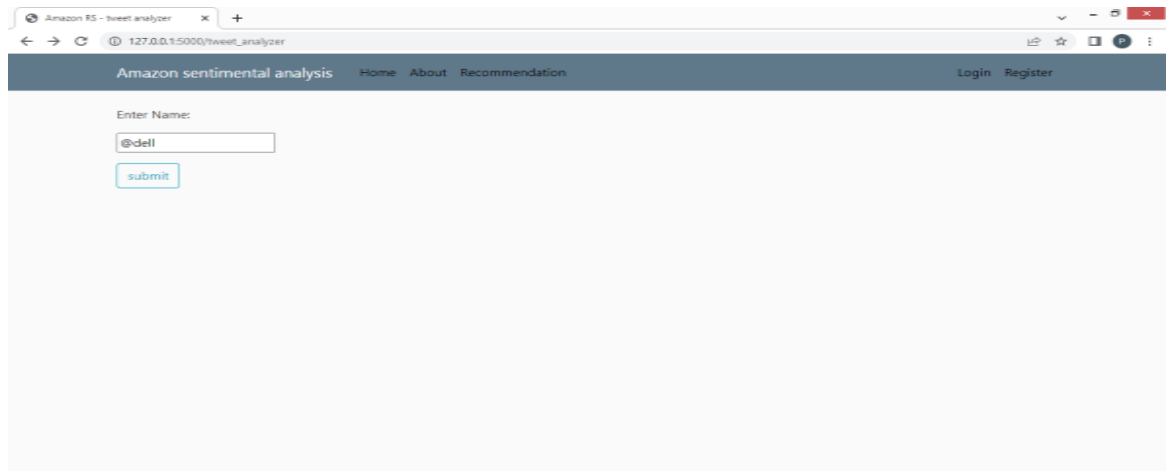


Fig 1: Search Page



Fig 2: Sentimental Analysis

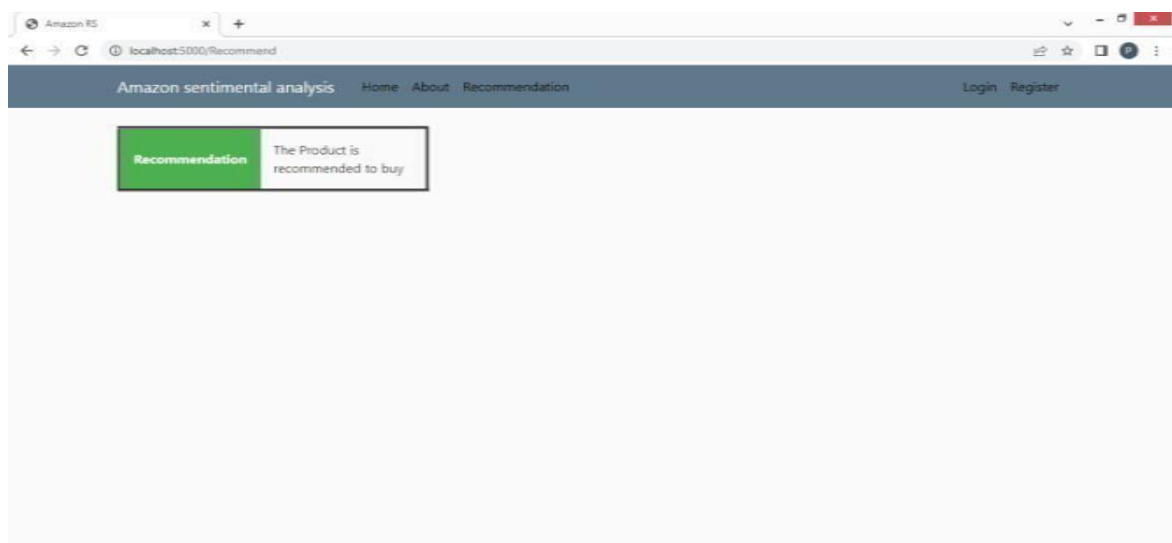


Fig 3: Recommendation Page

V. CONCLUSIONS

Sentiment analysis is the process of classifying works according to the emotions they express. In this project, discussion of representative strategies used in the three primary components of the standard sentiment analysis model: data preparation, review analysis, and sentimentrating is done. Research interest in sentiment analysis has increased significantly in recent years, an emerging field in text mining and computational linguistics. Capturing complex opinions and product attributes, as well as developing new classification models that can take into account the label attribute required in classification inference and product recommendation systems, are all topics of future study.

REFERENCES

1. Sahraoui Dhelim , Member, IEEE, Huansheng Ning , Senior Member, IEEE, Nyothiri Aung,Runhe Huang , Senior Member, IEEE, and Jianhua Ma, Member, IEEE “Personality-Aware Product Recommendation System Based on User Interests Mining and Metapath Discovery” IEEE transactions on computational social systems, vol. 8, no. 1, february 2021
2. Sunkuru Gopal Krishna Patro¹, Brojo Kishore Mishra¹, Sanjaya Kumar Panda², Raghvendra Kumar¹, Hoang Viet Long^{3,4,*}, David Taniar⁵, Ishaani Priyadarshini⁶ “A Hybrid Action-Related K- Nearest Neighbour (HAR-KNN) Approach for Recommendation Systems”
3. Anitha Anandhan^{1*}, Liyana Shuib^{1*}, Maizatun Akmar Ismail^{1*}, and Ghulam Mujtaba¹, “Social Media Recommender Systems: Review and Open Research Issues”
4. Babak Maleki Shoja, And Nasseh Tabrizi “Customer Reviews Analysis with Deep Neural Networks for E-Commerce Recommender Systems”
5. Yuzhou Liu^{1,2}, Lei Liu^{1,2}, Huaxiao Liu^{1,2*}, Suji Li³ “Information Recommendation based on Domain Knowledge in App Descriptions for Improving the Quality of Requirements”
6. Zhenyong Wu¹, Lu Li², And Haotian Liu³ “Process Knowledge Recommendation System for Mechanical Product Design” date of publication June 16, 2020,
7. Arta Iftikhar^{1,*}, Mustansar Ali Ghazanfar¹, Mubbashir Ayub^{1,*}, Zahid Mehmood², Muazzam Maqsood³ “An improved product recommendation method for collaborative filtering”
8. Debashis Das, Laxman Sahoo “A Survey on Recommendation System” Article in International Journal of Computer Applications · February 2017.
9. Ekant Khushhal Kagate, Rakesh Kannan, A Venkata Harshvardhan, Aprajita Singh, Nisha Ahuja “Online Product Recommendation Engine” 2020 IEEE International Conference on Computing, Power and Communication Technologies (GUCON) Galgotias University, Greater Noida, UP, India. Oct 2-4, 2020
10. Abhiraj Biswas, Kaza Sai Vineeth, Ayush Jain, Mohana “Development of Product Recommendation Engine By Collaborative Filtering and Association Rule Mining Using Machine Learning Algorithms” Proceedings of the Fourth International Conference on Inventive Systems and Control (ICISC 2020).
11. Gayatri Khanvilkar, Deepali Vora “Product Recommendation using Sentiment Analysis of Reviews: A Random Forest Approach” International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-8, Issue-2S2, January 2019
12. Rahul Pradhan, Vedant Khandelwal, Ankur Chaturvedi, Dilip Kumar Sharma “Recommendation System using Lexicon Based Sentimental Analysis with collaborative filtering” 2020 International Conference on Power Electronics & IoT Applications in Renewable Energy and its Control (PARC) GLA University, Mathura, UP, India. Feb 28-29, 2020
13. Kunal Shah¹, Akshaykumar Salunke², Saurabh Dongare³, Kisandas Antala⁴ “Recommender Systems: An overview of different approaches to recommendations” 2017 International Conference on Innovations in Information Embedded and Communication Systems (ICIIECS) Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-8, Issue-2S2, January 2019



INNO  SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.165

 **doi**[®]
cross **ref**

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 9940 572 462  6381 907 438  ijircce@gmail.com



www.ijircce.com

Scan to save the contact details