Website Rating Using Opinion Mining

Term Paper Submission for the Course
IT414 - Data Warehousing and Data Mining
Even 2022-23

by

Yuvraj Singh Jadon (201IT268)
SYP Madesh (201IT254)
K SRITEIA (201IT227))
Akki Sai Kumar Reddy(201IT101)
Madhava Chary Ravula (201IT235)

under the guidance of

Dr. Shrutilipi Bhattacharjee

DEPARTMENT OF INFORMATION TECHNOLOGY
NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA
SURATHKAL, MANGALORE - 575025

April, 2023

ABSTRACT

Abstract-The rise of the internet and its ever-expanding presence in our daily lives have led to the proliferation of websites serving various purposes. However, not all websites are created equal, and users must be able to evaluate a website's quality before engaging with it.

In this paper, we build upon the methodology proposed in the "Website Evaluation using Opinion Mining" research paper by P Keerthana, B Meghana, and P Akshaya, and introduce several enhancements to further improve the accuracy and effectiveness of website evaluation using opinion mining techniques.

We extend our approach to support multiple languages, implement topic modelling to extract specific insights, develop user profiles to identify trends, visualisation of data, user feedback collection and use natural language generation to generate evaluation reports.

Opinion mining, also known as sentiment analysis, is a natural language processing technique that aims to extract subjective information from text. By analyzing the opinions expressed in website reviews and comments, we can identify the strengths and weaknesses of a website and determine its overall quality. extract key aspects of the website that are mentioned in the reviews. We will evaluate our enhanced methodology on a dataset of website reviews and expect to see improved performance over the original methodology. Our approach will provide valuable insights for website owners and users, aiding in informed decisions about website quality.

Keywords— Opinion mining, Websites, Data, Sentiment analysis, Topic modelling

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 Introduction

Website evaluation has become a critical duty for website administrators and visitors alike as the internet becomes an increasingly important medium for commerce and communication. The goal of website evaluation is to rate a website's overall quality, including its usability, content, and user experience. Website evaluation has traditionally relied on professional manual examination, which can be expensive, time-consuming, and biassed.

Opinion mining, commonly referred to as sentiment analysis, has become a viable method for website evaluation in recent years. Opinion mining is the practise of analysing user evaluations and gaining insights into user sentiment and opinions using natural language processing and machine learning techniques. This method can offer insightful information about a website's advantages and disadvantages, enabling website owners to enhance their services and visitors to choose well-informed judgements.

Several research have investigated the use of opinion mining for website evaluation. But, there is still a lot that can be done to increase precision, effectiveness, and scalability. With the help of opinion mining, we hope to expand upon earlier studies on website evaluation and make a number of improvements.

Our approach will be expanded to accommodate more languages, subject modelling will be used to extract precise insights, user profiles will be created to spot patterns, and natural language generation will be used to create evaluation reports automatically. Using a dataset of reviews of e- commerce websites, we will test our methodology and assess its effectiveness in comparison to other methods. Our method will give website owners and consumers insightful information that will help them make wise judgements about the calibre of their websites.

We describe our proposed methodology, including the enhancements we introduce.We present our evaluation results and compare our approach to existing methods. Finally, we conclude the paper and discuss directions for future research.

2 Literature Survey

As the online world continues to grow, the importance of using mining techniques for website evaluation cannot be overemphasized. Website evaluation using strategy mining has received a lot of attention in recent years, as websites are widely used for many purposes such as e-commerce, education, and entertainment. In this section, we provide an overview of the existing literature on web evaluation using theory mining and other methods. A branch of natural language processing called "emotion mining" attempts to extract thoughts, feelings, and emotions from data. Sentiment analysis, the most popular emotion mining technique, classifies text as positive, negative, or neutral.

Another good way to determine opinion about certain features or products of a good or service is called appearance-based sensitivity analysis (ABSA). A method called thought summarization condenses the thoughts expressed in the document to provide a clear summary of behavior.

During our research we found some exiting work being done in the field:

Bing Liu [**4**] used a distance based approach to extract opinion words and phrases after extracting aspects. WordNet, which offered a set of methods for mining and summarising product reviews based on data mining and natural language processing techniques, was used to calculate the polarity of each retrieved opinion word. The goal was to present a feature-based overview of several customer reviews of an internet product.

Richa Sharma, et al. [**3**] proposed an aspect based opinion mining system to classify the reviews as positive, neutral or negative for each feature. In their system negation is also handled.

Madhavi Kulkarni and Mayuri Lingayat [**2**] proposed a technique which ranks efficiently the products by mining the genuine reviews of the product. System provides a method which allows only those users to write a review about a product who have purchased from the website. Other users are not allowed to give review.

Aashutosh Bhatt, et al. [**1**] proposed a method that first classifies customer feedback before determining the sentiment of those evaluations. The feature for the supplied product is extracted using rules. Their system illustrates the reviewer's

mood, which is subsequently displayed as a chart. The method became more accurate because to the classification of reviews and sentimental analysis, which in turn gave the user accurate reviews.

Anusha Sinha[5] proposes a methodical structure for identifying fake reviews. Sentiment analysis, often known as opinion mining, is the scientific study of how individuals feel about certain things. They take on a basic issue with sentiment analysis and categorising sentiment polarity. Every product has millions of reviews, and the majority of the time, consumers base their purchasing decisions on these reviews. Some businesses post false reviews of real products, leaving users in a difficult situation.

Once the analysis is complete, the admin user can restrict any IP address that routinely posts false reviews according to their software. Moreover, it notifies the user through email about blacklisted IP addresses. In this manner, it keeps track of any fabricated product reviews. And user can be sure about the products availability on that application and reviews too.

2.1 Motivation

In today's world, businesses rely heavily on an online presence to reach customers and increase sales. That's why it's important for businesses to have a good website that provides a good user experience. Getting feedback can help businesses understand how users interact with their websites and identify areas for improvement.

Traditional web testing methods such as focus groups or surveys can be time consuming and expensive and may not provide a complete picture of the user experience. Emotion mining can help overcome these limitations by extracting insights from a variety of user-generated content, including comments and social media posts.

The user experience on the website can be improved to increase engagement, increase customer loyalty and ultimately increase revenue. By using feedback to evaluate websites, businesses can gain a competitive advantage and position themselves for long-term success.

3 Problem Statement(s)

Despite the growth of online businesses and the increasing importance of website performance, effective website evaluation remains a challenge. Website owners and other stakeholders find it difficult to understand and respond to user complaints and criticisms due to the volume and variety of data generated by users.
Due to the increasing frequency of fake reviews and spam, the process of evaluating a website has become more difficult.
Traditional website evaluation methods can be successful up to a point, but they are often time-consuming and limited.
To solve this problem, this study aims to investigate the use of opinion mining techniques such as topic models, aspect-based sentiment analysis, and sentiment analysis in order to collect and analyze user feedback and evaluate the website in detail. It will also look at the use of this method to identify fake reviews or spam.
To improve the accuracy and usability of the evaluation results, this study will also explore the use of multilingual help, user profiles, and user feedback collection methods.
The main purpose of this research is to develop a method to evaluate websites that is more efficient and effective and can inform stakeholders and website owners about the strengths and weaknesses of their websites, leading to a better user experience and more engagement. fake reviews and spam.

4 Implementation

The plan for implementation of the proposed website evaluation method includes the following steps:

Data Collection: The Sentiment140 dataset was used for this project, which consists of 1.6 million tweets extracted using the Twitter API. The tweets have been manually annotated as negative (0) or positive (4) and can be used for sentiment analysis. The dataset includes six fields, including the polarity of the tweet, tweet ID, date, query, user, and text. This dataset was chosen due to its large size and

manually labeled sentiment, making it suitable for training machine learning models for sentiment analysis. The dataset was originally used in a Stanford CS224N project, and the official link and paper for the dataset can be found for further information.

Data preprocessing: In the machine learning project, the text data was preprocessed and vectorized successfully. Firstly, the dataset was loaded into a pandas DataFrame and the columns were renamed. The sentiment values of 0 and 4 were mapped to negative and positive, respectively. After that, unnecessary columns were removed and the comments were preprocessed by removing special characters, punctuation, and stop words. Then, the preprocessed dataset was saved to a new file.
A function was written to clean the text data by removing punctuation, numbers, extra whitespaces, and converting the text to lowercase. This function was applied to the 'comment' column of the DataFrame. The text data was vectorized using CountVectorizer, which counts the frequency of words in the text and creates a matrix of word counts. Finally, the dataset was split into training and testing sets.

Sentiment Analysis: To perform sentiment analysis, a logistic regression model was used to classify the comments into either positive or negative sentiment. The comments were vectorized using TF-IDF, which weights the word frequency by their importance in the text. The logistic regression model was trained on the vectorized training data.
It's important to note that while preprocessing and vectorizing the data are important steps, they are not the only ones in sentiment analysis. The data should also be explored and visualized, different vectorization techniques should be experimented with, different classification models should be tried, and the hyperparameters of the chosen model should be fine-tuned to improve its performance. Additionally, the model's performance on the testing set should be evaluated using metrics such as accuracy, precision, recall, F1 score, and ROC AUC score. These metrics help to assess the effectiveness of the model in classifying the sentiment of the comments.

Deployment: Deploy the website evaluation system on a cloud-based platform such as Amazon Web Services or Google Cloud to make it easily accessible to website

owners and stakeholders.

5 Future Work

Aspect-based Sentiment Analysis: Extract specific aspects of the website mentioned in the reviews, such as customer service, product quality, and pricing.Use natural language processing techniques to preprocess the reviews and extract sentiment-bearing phrases related to the identified aspects. Conduct aspect-based sentiment analysis using machine learning or rule-based approaches to determine the sentiment of each aspect mentioned.

Topic Modelling: Preprocess the reviews using techniques such as tokenization, stop word removal, and stemming. Implement topic modelling techniques such as Latent Dirichlet Allocation (LDA) to identify common topics and themes present in the reviews. Analyze the topics to gain more specific insights into the strengths and weaknesses of the website.

Fraud Detection: Use machine learning algorithms such as clustering, outlier detection, or anomaly detection to identify suspicious reviews. Develop rules and thresholds to classify reviews as fraudulent or not based on features such as review length, rating consistency, and language complexity. Implement fraud detection mechanisms to identify fraudulent reviews and spamming reviews.

Visualization: Develop interactive dashboards and visualizations to present the website evaluation results in an easy-to understand format for different stakeholders

6 Results

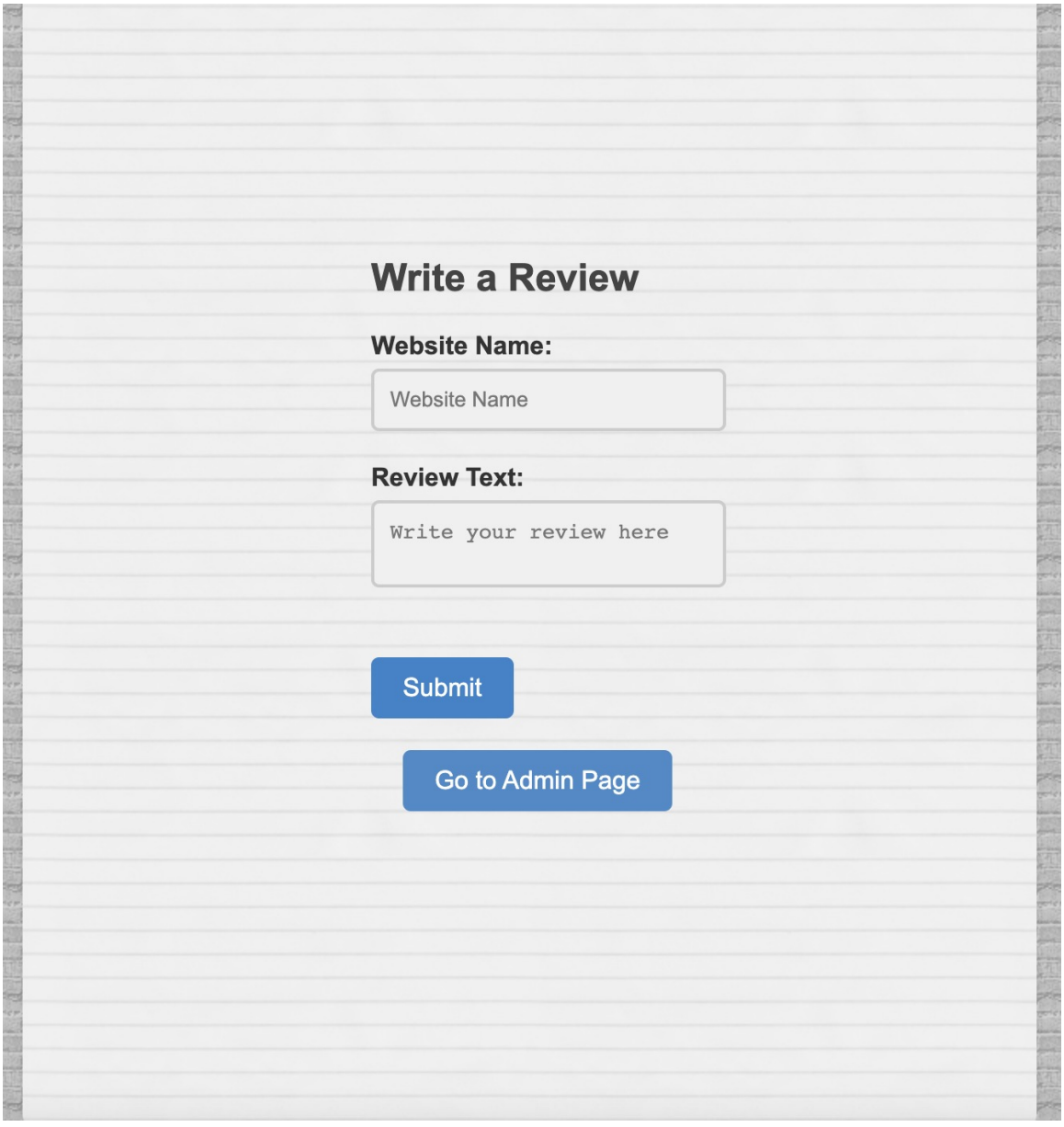

Figure 6.1: Home Page

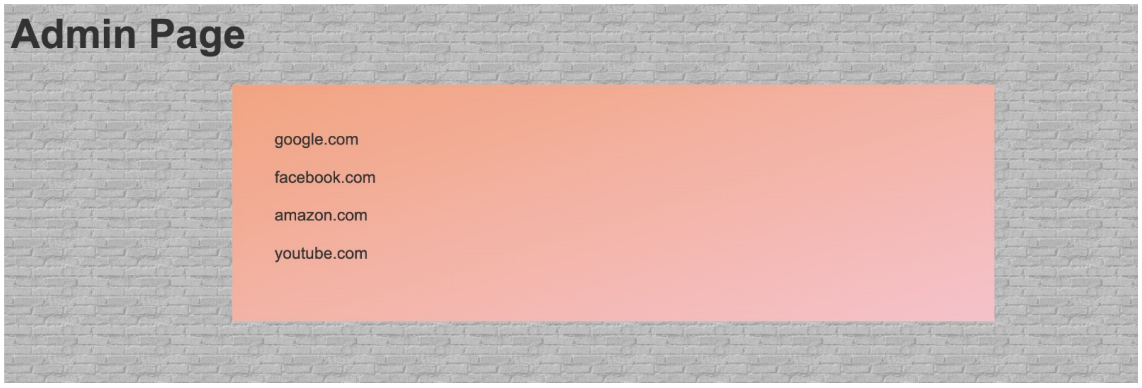

Figure 6.2: Admin Page

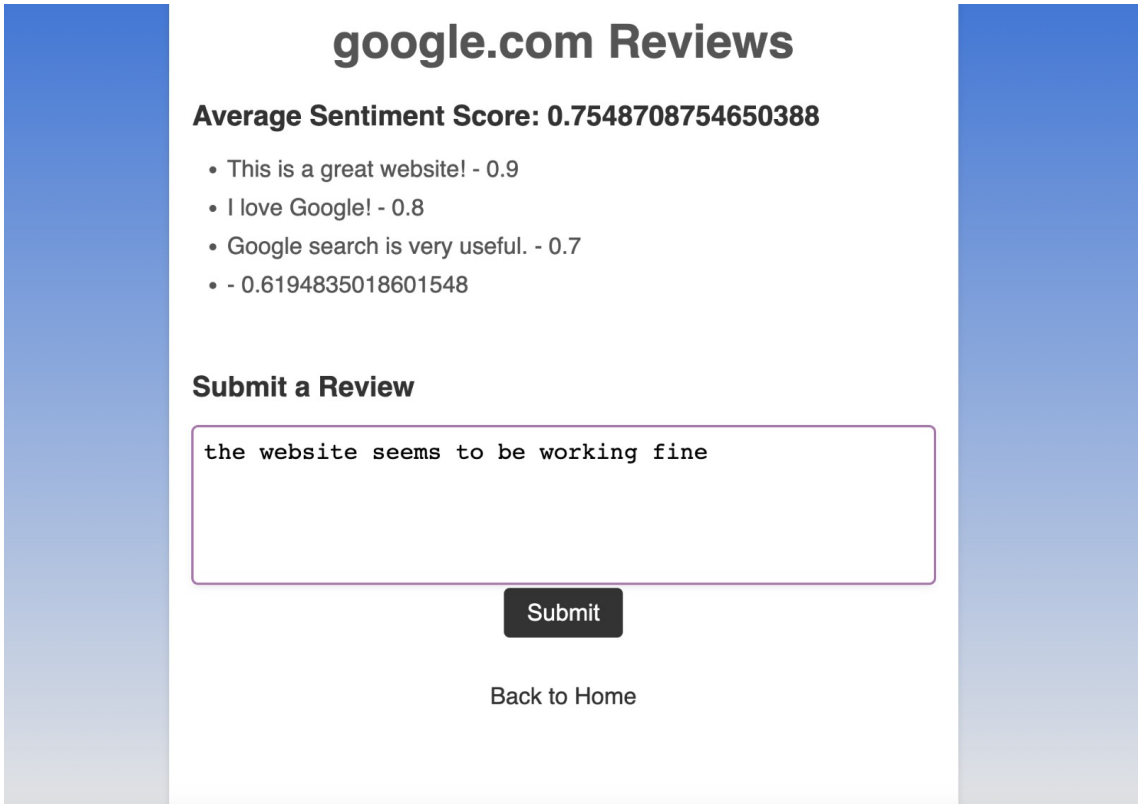

Figure 6.3: Website Details Page

7 References

As the basis for our research, we utilized the paper 'Website Evaluation Using Opinion Mining' by Ganesh R, Saketh D, Sai Kumar, and M Ramesh (year of publication), which proposed a method for evaluating websites using sentiment analysis and opinion mining techniques.
link to the papaer

References

[1] Aashutosh Bhatt, Ankit Patel, Harsh Chheda, Kiran Gawande, 2015, "Amazon Review Classification and Sentiment Analysis", International Journal of Computer Science and Information Technologies, Vol. 6.

[2] Madhavi Kulkarni,Mayuri Lingayat, 2015,"Effective Product Ranking Method based on Opinion Mining", International Journal of Computer Applications (0975 – 8887),Volume 120 – No.18.

[3] Richa Sharma,Shweta Nigam and Rekha Jain, 2014,"Mining Of Product Reviews At Aspect Level", International Journal in Foundations of Computer Science , Technology , Vol.4, No.3.

[4] Bing Liu, 2012, Sentiment analysis and opinion mining, Morgan and Claypool Publishers

[5] Anusha Sinha, Nishant Arora, Shipra Singh, Mohita Cheema, Akthar Nazir,"Fake Product Review Monitoring Using Opinion Mining",International Journal of Advances Engineering and Management (IJAEM)