

Contents

Abstract.....	2
Introduction.....	3
Literature Study	5
Research Design	9
Research Problem:	9
Research Objectives.....	9
Research Questions.....	9
Elements of Research Design	9
Significance and Contribution:	10
Methodology	11
Code Repository and Implementation Details	12
Data Collection	13
Description of the Review Website	14
Data Cleaning	19
Sentiment Analysis for amazon and flip kart datasets	19
Named Entity Recognition (NER) for amazon and flip kart datasets.....	20
Topic modeling for amazon and flip kart datasets.....	21
Result & Discussion	22
User-side Website Pages Details – The Review Hub	22
Business Owner-side Website Pages Details – The Review Hub.....	25
Application of NLP Techniques on The Review Hub Website Dataset	29
Sentiment Analysis Results for Flipkart and Amazon Review Datasets	31
Named Entity Recognition Results for Flipkart and Amazon Review Datasets.....	33
Topic Modeling Results for Flipkart and Amazon Review Datasets.....	35
Document Labelling for Flipkart and Amazon Review Datasets	36
Limitations and Future Work.....	38
Conclusion:	39
References.....	40

Abstract

Navigating a huge amount of user-generated assessment has become extremely difficult for customers and businesses as a result of the web-based shopping industry's exponential growth. A centralized mechanism for website reviews has been created in order to remedy this. This hub uses machine learning methods, such as natural language processing (NLP), to speed up the review analysis process and give both consumers and businesses useful information. NLP approaches were used on a variety of datasets, including flip cart review datasets and Amazon review datasets, to ensure the effectiveness of the review analysis. The algorithms for sentiment analysis, topic modelling, named entity recognition, and review categorization were trained and improved using these datasets. The algorithms were then applied to the review website dataset after being improved by the information learned from studying the Amazon and flip cart reviews. The review hub uses sentiment analysis to extract the underlying feelings stated in reviews, which enables customers to comprehend the general consensus regarding a product fast. By identifying significant conversation topics, topic modelling provides firms with a thorough overview of noteworthy elements and prospective problems. firms may track their brand reputation and customer happiness thanks to named entity identification, which makes it easier to identify the firms linked to reviews. Businesses may easily access and analyze information with the help of efficient review classification procedures and document labelling strategies. The method's validity with real-world datasets, such as Amazon and flip cart reviews, shows the system's scalability and robustness. The review website has successfully incorporated the NLP techniques used on these datasets, empowering users to make knowledgeable judgements and allowing businesses to enhance their offers based on insightful customer feedback. In conclusion, the website review hub makes use of NLP methods that have been developed using a variety of datasets to provide a complete solution for handling product reviews. While enabling businesses to improve their goods and services based on priceless insights gained from the review research, this strategy gives users a deeper knowledge of the attitudes and topics in reviews.

Introduction

The introduction of online shopping has completely changed how customers engage with products and make judgments about what to buy (Anderson & Anderson, 2002). The ease of use and accessibility that e-commerce platforms offer has allowed customers to access a wide range of goods and services from numerous suppliers. However, the rise in online sales has also resulted in an exponential rise in user-generated reviews, making it more and more difficult for customers and businesses to sift through mountains of data and derive actionable insights (Yan et al., 2015). A thorough response to this issue is offered in the form of a website review hub. Users can write product reviews on the website review hub, which also acts as a single location for businesses to register their products and effectively handle customer feedback. The main goal of this hub is to provide businesses and customers with insightful data obtained by machine learning methods, empowering them to make wise decisions and enhance their products in response to client input. Sentiment analysis is one of the essential machine learning methods used in the review hub. A potent strategy that enables the automatic extraction of the underlying sentiments stated in reviews is sentiment analysis. Sentiment analysis algorithms can categorize reviews into positive, negative, or neutral categories by using machine learning and natural language processing techniques (Kumar & Benbasat, 2006; Singh et al., 2017).

Users can immediately understand the general consensus regarding a product thanks to this classification, which helps them make decisions about what to buy that are in line with their needs and tastes. Businesses can measure client happiness and sentiment towards their products using sentiment analysis, which gives them a useful tool for identifying areas for improvement or capitalizing on favorable feedback (Xie et al., 2014). The method uses topic modelling in addition to sentiment analysis to determine the major issues raised in the reviews. Topic modelling identifies the themes and features that people find notable or problematic in regard to a product by utilizing cutting-edge machine learning methods (Liu et al., 2019). Businesses may learn more about the preferences and problems of their customers thanks to this function, which provides a thorough overview of the various dimensions and characteristics of a product (Biswas et al., 2022). Understanding these subjects enables organizations to make data-driven decisions that emphasize product enhancements or highlight the advantages of their services. Named entity recognition is used to expand the review hub's functionalities. Named entity recognition is a method for locating and extracting particular named entities from reviews, such as related organizations. Businesses may simply find and track reviews that are specific to their products with this technology, giving them access to important information about the popularity of their brand and consumer happiness (Day & Lin, 2017; Nasimuzzaman et al., 2023).

Businesses can proactively address any issues, answer consumer concerns, and improve overall customer experience by tracking and analyzing reviews related to their products. Another crucial aspect of the system is the effective classification of reviews, which is made possible by approaches for document labelling. These methods enable the systematic categorization and grouping of reviews according to particular standards or characteristics (Loper & Bird, 2002). Businesses may more rapidly access and analyze consumer feedback relating to particular features of their products by categorizing reviews, making it easier to respond to customer problems or ideas in a targeted and effective manner. Experiments have been carried out utilizing publicly accessible datasets, such as the flip cart review dataset and the Amazon review dataset, to validate the efficacy of the approach (Jurafsky & Martin, 2019).

The reliability and scalability of the system in evaluating and extracting insights from other sets of reviews have been proved by using the same machine learning algorithms on these datasets. These validations highlight the system's effectiveness and potential in real-world situations. Lastly, the website review hub provides a thorough and cutting-edge solution for customers and companies struggling to manage and glean insights from a deluge of user-generated evaluations. Users are given insightful information to help them make decisions by using machine learning techniques including sentiment analysis, topic modelling, named entity identification, and document labelling. Businesses are also able to monitor and improve their products based on client input. The combination of these methods with actual datasets confirms the effectiveness and potential of the system in solving the challenges posed by the enormous volume of online reviews.

Literature Study

The retail market is experiencing a change thanks to the rapid rise of online shopping, which provides customers with unmatched accessibility and convenience on a global scale. But the rise in buying online has also produced an enormous amount of user-generated reviews for various goods and services. For both consumers and businesses, efficiently reading through these evaluations to glean actionable insights has grown to be a major difficulty (Yan, Xing, Zhang, & Ma, 2015). A website review hub has been created as a central platform to improve the review analysis process in response to this difficulty. This hub seeks to offer useful insights to consumers and businesses by applying cutting-edge machine learning techniques, thereby supporting informed decision-making in the online retail environment.

Sentiment analysis is one of the main components of the online review hub. The hub captures the underlying sentiments conveyed in user-generated evaluations by using natural language processing and machine learning techniques (Anderson & Anderson, 2002; Kim, Luan, & Gu, 2016). Users can immediately comprehend the general opinion about a product thanks to this analysis. Users are able to make informed decisions based on the overall attitude of the reviews thanks to the categorization of positive, negative, and neutral sentiments (Pang & Lee, 2008; Turney, 2002). The online review hub makes use of cutting-edge machine learning approaches that have been thoroughly researched and shown to be successful in sentiment analysis tasks in order to obtain accurate sentiment analysis (Liu, 2012; Wang, Zhang, Zhang, Ye, & Zeng, 2012).

Using labelled data, where attitudes have been given to a collection of reviews, these strategies entail training the algorithms. The algorithms gain the ability to identify patterns and language cues that denote good, negative, or neutral attitudes through this training process. The algorithms used for sentiment analysis include a number of variables, including the text's tone, the context in which specific words or phrases are used, and the overall sentiment indicated by the reviewer (Kim et al., 2016). The hub can effectively classify the feelings and give users a summary of the general perception of the good or service by taking into account these aspects. Users can use the categorized feelings as a useful tool for decision-making.

Negative emotions attract attention to potential problems or shortcomings, whereas positive sentiments show that most reviewers have had a positive experience with the product (Turney, 2002). Sentiment neutrality denotes a moderate or conflicted attitude from users. With this knowledge, people may decide more intelligently whether to buy a certain product or look into alternatives that better suit their requirements and tastes. In conclusion, sentiment analysis is extremely important in the website review hub and is supported by well-established machine learning techniques. The hub helps customers navigate the

complex world of online buying more skillfully by classifying attitudes into positive, negative, and neutral categories (Pang & Lee, 2008; Wang et al., 2012).

The online review hub incorporates topic modelling methods in addition to sentiment analysis (Patel, Nagababu, Kachhwaha, & Surisetty, 2022; Blei, Ng, & Jordan, 2003; Griffiths & Steyvers, 2004). The hub discovers important conversation areas in the reviews by topic modelling. This feature gives groups a thorough overview of the salient features and potential problems pertaining to their products (Hu & Liu, 2004). Businesses can learn about client preferences, areas for improvement, and new trends by identifying popular themes (Mei, Ling, Wondra, Su, & Zhai, 2007). The textual content of reviews is reviewed using a machine learning technique called topic modelling to find latent topics or patterns that appear in many reviews.

The hub collects probabilistic weights assigned to words and finds subjects based on their word distributions by using methods like Latent Dirichlet Allocation (LDA) (Griffiths & Steyvers, 2004). Businesses may examine the themes once they have been selected to comprehend the feelings and opinions attached to each issue (Hu & Liu, 2004). By identifying the areas in which their products excel or fall short, firms can concentrate on making changes and living up to client expectations. Additionally, topic modelling gives companies a competitive edge in the dynamic online market by allowing them to spot developing trends or problems in real-time (Mei et al., 2007).

Businesses can proactively modify their offers and tactics to satisfy changing market demands by remaining up to speed on client preferences and concerns. In conclusion, businesses can extract important conversation topics and themes from user-generated reviews by including topic modelling approaches into the online review hub (Patel et al., 2022). Businesses can acquire useful insights into client preferences, areas for improvement, and new trends by identifying popular subjects. This will allow them to make data-driven decisions and improve their products and services accordingly.

Named entity recognition (NER) skills are also integrated into the website review hub (Xie, Zhang, and Zhang, 2014; Singh, Irani, Rana, Dwivedi, Saumya, and Roy, 2017). NER aids in locating businesses or products linked to the evaluations (Liu, Lee, & Srinivasan, 2019). Businesses may track their brand reputation and customer satisfaction levels thanks to this function, which is especially helpful for them (Liu, Shin, & Burns, 2021). Businesses can actively resolve any issues or concerns by finding references of their organisation in the reviews (Xie et al., 2014). This promotes a great client experience. The hub

automatically recognises mentions of the company or brand connected to the reviews by using NER (Singh et al., 2017). This enables companies to learn more about how customers view their goods or services.

Businesses may nurture a great customer experience and increase brand loyalty by keeping track of mentions of their company. This allows them to quickly address any unfavourable feelings and recognise any positive comments (Xie et al., 2014). Businesses can also determine how customers feel about rival companies thanks to NER (Liu et al., 2019). This data offers useful competitive intelligence, enabling organisations to compare their products to those of rivals and spot opportunities for development or difference (Liu et al., 2021). Businesses can improve their products, services, and overall customer experience by utilising NER capabilities to better understand the influence their brand has on consumers and make data-driven decisions (Xie et al., 2014). Additionally, it enables companies to interact with clients proactively, responding to their issues, and eventually raising client happiness and loyalty (Liu et al., 2021). In conclusion, the inclusion of NER capabilities in the online review hub (Xie et al., 2014; Singh et al., 2017) enables businesses to watch brand reputation, keep track of consumer sentiment, and quickly respond to complaints or difficulties (Liu et al., 2021). Businesses can create a great customer experience, increase brand loyalty, and acquire insightful competition information by using NER to find mentions of their business or brand in user-generated reviews (Liu et al., 2019; Liu et al., 2021).

The internet review hub integrates effective review classification approaches to speed up the review analysis process (Singh et al., 2017; Kumar & Benbasat, 2006). The hub organises evaluations into pertinent domains or product categories using document labelling techniques, ensuring effective structure and simple information access for companies. Businesses can browse through a vast volume of reviews and concentrate on particular product areas of interest thanks to the hub's review classification capabilities (Kumar & Benbasat, 2006). Businesses may rapidly access pertinent information and acquire a thorough grasp of client feedback for particular product lines or offers by dividing assessment into various domains or product categories.

This comprehensive overview allows businesses to identify potential areas for innovation, address customer concerns, and align their product development strategies with market demands. In conclusion, the integration of efficient review categorization techniques in the website review hub (Singh et al., 2017; Kumar & Benbasat, 2006) empowers businesses to organize, access, and analyze customer feedback effectively. By categorizing reviews into relevant domains or product categories, businesses can make data-driven decisions, identify areas for improvement, and align their offerings with customer preferences and market trends.

Machine learning algorithms are used in the review categorization process to examine the reviews' content and characteristics (Singh et al., 2017). These algorithms classify reviews into the right categories by taking into account a number of factors, including keywords, themes, and context. Businesses save time and effort thanks to this automated categorization because it allows them to quickly obtain and evaluate pertinent consumer input. Businesses can learn important information about customer preferences, opinions, and experiences for particular product categories by using effective review categorization (Kumar & Benbasat, 2006). Businesses can use this information to make data-driven decisions, pinpoint areas for improvement, and create focused initiatives to improve their services. Additionally, the classified evaluations can be further examined to find repeating themes, problems, or developing trends in particular product categories.

Real-world datasets have been used to validate the online review hub's efficacy and robustness (Kumar & Benbasat, 2006). The system has proven to be capable of handling high review traffic and providing precise sentiment analysis, subject modelling, named entity identification, and review categorization. The validation procedure validates the hub's scalability and dependability, establishing its viability as a management tool for product reviews in a variety of online purchasing scenarios. In the context of the expanding online retail environment, the website review hub offers a comprehensive solution for handling product reviews (Liu, Lee, & Srinivasan, 2019).

Research Design

Research Problem:

The exponential growth of the web-based shopping industry has resulted in a massive influx of user-generated reviews. Navigating through this vast amount of assessment has become an arduous task for both customers and businesses. The sheer volume and diversity of reviews make it challenging to extract valuable insights and identify trends, sentiments, and important topics. Customers often struggle to understand the overall consensus regarding a product, while businesses face difficulties in analyzing customer feedback to improve their offerings. To address these issues, there is a need for a centralized mechanism that can efficiently process and analyze user-generated reviews, providing valuable information to both consumers and businesses.

Research Objectives

The research is driven by several key objectives. Firstly, it aims to develop a centralized mechanism for website reviews that leverages machine learning methods, particularly natural language processing (NLP). This mechanism will streamline the review analysis process, enabling quicker extraction of sentiments and categorization of reviews into relevant topics. Secondly, the study aims to improve the efficiency and accuracy of sentiment analysis, topic modeling, named entity recognition, and review categorization algorithms through rigorous training and improvement using diverse datasets, such as Flipkart and Amazon reviews. The research further aims to apply these enhanced algorithms to the review website dataset, leveraging knowledge gained from analyzing Amazon and Flipkart reviews to refine the review analysis process. Additionally, the research seeks to validate the scalability and robustness of the developed centralized review hub using real-world datasets, such as Amazon and Flipkart reviews. This validation process is crucial in assessing the system's performance under various scenarios and ensuring it can handle large volumes of reviews effectively.

Research Questions

To achieve the research objectives, several fundamental questions need to be addressed. Firstly, the study will explore how NLP methods can be effectively applied to analyze and categorize user-generated reviews in the web-based shopping industry. This will involve understanding the challenges posed by the sheer volume and diversity of reviews and identifying appropriate techniques to overcome them. Secondly, the research will delve into sentiment analysis, investigating the sentiments expressed in web-based shopping reviews and determining how to accurately extract and interpret these sentiments using NLP algorithms. The study will also investigate topic modeling to identify significant conversation topics within user-generated reviews, providing businesses with a comprehensive overview of noteworthy elements and potential issues. Moreover, the research will explore how named entity recognition techniques can be employed to track brand reputation and customer satisfaction by identifying the companies associated with reviews. Understanding the relationships between reviews and specific entities will help businesses monitor their brand image and customer sentiment effectively. Finally, the research will focus on developing efficient review classification procedures and document labeling strategies that enable businesses to access and analyze review information more efficiently. These strategies are essential in organizing and categorizing reviews based on their content and relevance to specific products or services.

Elements of Research Design

The research design comprises several key elements that collectively contribute to achieving the research objectives. Firstly, data collection is critical to the success of the research. User-generated reviews will be collected from various web-based shopping websites, such as Flipkart and Amazon, to create diverse datasets that encompass a wide range of products and domains. These datasets will form the foundation for training and

validating the NLP algorithms. The development and training of NLP algorithms for sentiment analysis, topic modeling, named entity recognition, and review categorization will be a significant focus of the research. These algorithms will be trained using machine learning and deep learning techniques on the collected datasets, with a particular emphasis on improving their accuracy and performance. The improved algorithms will then be integrated into a centralized platform for website reviews, providing businesses with a user-friendly and efficient tool for processing and analyzing user-generated reviews. The validation process is crucial for assessing the effectiveness and scalability of the developed centralized review hub. Real-world datasets, such as Amazon and Flipkart reviews, will be used to evaluate the platform's performance and robustness under realistic conditions.

Significance and Contribution:

The research holds significant implications for the web-based shopping industry, both for customers and businesses. The development of a centralized mechanism for website reviews, driven by NLP methods, will revolutionize the way reviews are processed and analyzed. Businesses will benefit from faster and more accurate insights into customer sentiments and important topics, enabling them to make data-driven decisions to improve their products and services. The mechanism will empower businesses to monitor their brand reputation and customer satisfaction effectively, leading to enhanced customer experiences and improved customer loyalty. On the other hand, customers will benefit from a more efficient review analysis process, enabling them to make more informed purchasing decisions based on comprehensive insights into product sentiments and important aspects. The research will contribute to advancements in the field of NLP and its applications in the web-based shopping industry. The development of improved algorithms and a centralized review hub will be a valuable contribution to the research community and the broader e-commerce sector. By enhancing the efficiency of review analysis, the research will help businesses cater better to customer needs and preferences, leading to improved customer satisfaction and retention. Ultimately, the research's findings and the developed mechanism have the potential to significantly impact the web-based shopping industry by facilitating better decision-making, improving products and services, and fostering a more positive and rewarding shopping experience for customers.

Methodology

The website review hub is a centralized platform developed to address the challenges associated with navigating user-generated reviews in the context of online shopping (Yan, Xing, Zhang, & Ma, 2015). Its purpose is to provide users and businesses with valuable insights and facilitate informed decision-making. The primary objective of the website review hub is to streamline the review analysis process by leveraging various techniques such as sentiment analysis (Anderson & Anderson, 2002), topic modeling (Patel, Nagababu, Kachhwaha, & Surisetty, 2022), named entity recognition (NER) (Xie, Zhang, & Zhang, 2014), and efficient review categorization (Singh et al., 2017). By incorporating these methodologies, the hub aims to extract meaningful information from a vast amount of user-generated reviews, making it easier for users to understand the overall opinion about a product and for businesses to gain insights into customer preferences and areas of improvement. The website review hub's capabilities extend beyond general review analysis. It includes specific NLP techniques applied to flip cart review datasets and Amazon review datasets (Kumar & Benbasat, 2006; Liu, Lee, & Srinivasan, 2019; Liu, Shin, & Burns, 2021). These datasets provide a diverse range of reviews and enable the hub to capture insights specific to these domains. The website review hub utilizes natural language processing (NLP) techniques to analyze Flipkart reviews, providing users with valuable insights and sentiment analysis of products. It helps users make informed decisions by understanding the opinions and sentiments expressed in reviews. For businesses, the hub offers brand reputation monitoring and customer feedback analysis. The user-friendly interface allows easy access to reviews and relevant information, catering to both consumers and businesses. Overall, the review hub streamlines the review analysis process, empowering users to make informed decisions and businesses to enhance their offerings based on valuable insights from Flipkart reviews. For consumers, the hub empowers them to make informed decisions by quickly understanding the collective sentiment expressed in the reviews. By categorizing reviews into positive, negative, and neutral sentiments, users can gauge the overall opinion about a product and determine its suitability for their needs. On the other hand, businesses benefit from the website review hub by gaining a comprehensive overview of noteworthy aspects and potential issues related to their products. Through topic modeling, the hub identifies key discussion points within the reviews, allowing businesses to understand customer preferences, identify areas for improvement, and stay updated on emerging trends. The integration of NER capabilities enables businesses to track their brand reputation and monitor customer satisfaction levels by identifying mentions of their organization or brand in the reviews. Overall, the website review hub serves as a valuable tool for users and businesses alike in the online shopping landscape. It simplifies the review analysis process, empowers users to make informed decisions, and enables businesses to enhance their offerings based on valuable customer feedback.

Code Repository and Implementation Details

Website Review Hub

The Website Review Hub serves as a centralized mechanism developed to address the challenges of handling user-generated reviews in the web-based shopping industry. It is implemented using PHP for the front-end and backed by MySQL for efficient data storage and retrieval.

[\[Code Link\]](#)

Website Link: <http://reviewhub.infinityfreeapp.com/>

Flask Application

The Flask Application acts as the backbone of the Website Review Hub, facilitating user interactions and enabling the review analysis process.

[\[Code Link\]](#)

NLP Operations

The NLP Operations encompass sentiment analysis, topic modeling, named entity recognition, and review categorization, essential for extracting valuable insights from user-generated reviews. They are implemented using Google Colab notebooks.

Code Links:

- NLP Operations (Sentiment Analysis, Topic Modelling, Named Entity Recognition, Document labelling) on Amazon Dataset [\[Code Link\]](#)
- NLP Operations (Sentiment Analysis, Topic Modelling, Named Entity Recognition, Document labelling) on Flipkart Dataset [\[Code Link\]](#)
- NLP Operations (Sentiment Analysis, Topic Modelling, Named Entity Recognition, Document labelling) on Review Hub Website Dataset [\[Code Link\]](#)

Data Collection

The collection of user-generated review data for the website review hub involved gathering reviews from multiple sources, including external online platforms and the user base of the hub itself. The data collection process was designed to ensure a diverse and representative dataset that encompasses a wide range of products and user opinions. To obtain a comprehensive dataset, user-generated reviews were collected from popular online platforms such as Flip Cart Database and Amazon. These platforms are well-known for hosting a vast array of product reviews, including product review on Flipkart and various products on Amazon. The reviews available on these platforms provide valuable insights into customer opinions, sentiments, and experiences. For the Flipkart dataset, reviews were collected from the Flipkart website using their public API (Internet Flip cart Database, n.d.). This API allowed for the extraction of reviews related to specific reviews, capturing a variety of opinions expressed by users. Similarly, for the **Amazon dataset**, reviews were collected from the Amazon website using their publicly available APIs (Amazon Developer, n.d.). The APIs provided access to customer reviews for different products, enabling the collection of a diverse set of opinions and feedback. In addition to external platforms, the **website review hub** itself served as a source of user-generated reviews. As users interacted with the hub's interface and provided feedback on various products, their reviews were collected and incorporated into the dataset. This approach allowed for the inclusion of real-time, user-generated content that reflects the specific context and products of the website review hub.

To ensure data quality, relevance, and consistency in the review collection process, several criteria and filters were applied. Firstly, reviews below a certain length threshold were excluded. This approach aimed to focus on comprehensive and informative user opinions, avoiding very short or incomplete reviews that may not provide valuable insights. Additionally, reviews with explicit ratings or sentiment indicators were given priority. By prioritizing such reviews, it facilitated sentiment analysis and categorization, as the presence of clear ratings or sentiment indicators can provide direct information about the reviewer's opinion. Furthermore, to ensure a diverse representation of opinions from different domains, reviews were collected across various product categories. This approach aimed to capture a wide range of perspectives and experiences, considering that different products may have distinct characteristics and customer expectations. Moreover, the reviews collected spanned a specific time period. This timeframe was chosen to capture recent opinions and trends, acknowledging that consumer preferences and product experiences may evolve over time. By focusing on recent reviews, the analysis can reflect current sentiments and considerations. By applying these criteria and filters, the review collection process aimed to maintain data quality and consistency. The exclusion of short or incomplete reviews, prioritization of explicit ratings, inclusion of diverse product categories, and consideration of a specific time period helped ensure that the collected data was comprehensive, informative, and relevant for the intended analysis and insights.

Description of the Review Website

The website known as "Review Hub" seeks to give consumers a place to post product reviews while also letting companies upload their goods and keep an eye on customer feedback. Using HTML, PHP, CSS, JavaScript, and MySQL, the website was created. It includes features including user registration, product submission, rating and review systems, search and filtering options, and facilities for managing reviews. The website was developed using a systematic methodology, starting with requirement gathering and planning, front-end design, back-end development, user authentication, product submission, implementation of the rating and review system, search and filtering functionality, testing, and deployment. To ensure functionality, security, and performance, ongoing maintenance is carried out.

A review hub website is created using a system that includes several crucial elements. The process starts with requirement collecting and planning, where the goal, target market, and desired website features are established. Key functionality including user registration, product submission, rating systems, and review management are identified during this phase. To provide a seamless and simple user experience, the website's structure, layout, and navigation are also meticulously prepared. Wireframes or mockups of the website are made using design tools or software before moving on to the design and front-end development phase. The front-end is developed using HTML, CSS, and JavaScript with a focus on producing aesthetically pleasing and engaging user interfaces. Making sure the website is responsive and cross-browser compatible allows it to change to fit multiple devices.

In the back-end development phase, a PHP-compatible web server and a MySQL database are set up. Critical functions like user registration, login, and authentication are handled by PHP. Before saving user data in the database, it offers a secure method of processing and validating that data. The essential database tables for storing user data, product information, reviews, and ratings are created using MySQL. The website is implemented with security mechanisms, such as server-side validation, to guard against any flaws and criminal activity. To provide a safe and seamless experience, the user registration and authentication process is planned and put into practice. The creation of a user registration system allows for the secure storage of crucial user information in the MySQL database, including username, email, and password. Users are authenticated by a login process set up using PHP sessions or tokens, granting them access to their customized accounts. Techniques for password encryption are used to protect user credentials from unauthorized access, such as salting and hashing.

A product submission option is created to let companies submit their items. This entails developing an intuitive form or interface where companies can submit crucial product information, such as name, description, category, and photos. In order to guarantee that the submitted data is accurate and comprehensive, server-side validation is used. After that, the product data is saved in the MySQL database, where it can later be retrieved and displayed. A crucial component of the review hub website is the rating and review system. Users can rate and review products using a user-friendly, interactive interface. User inputs are saved in the database and connected to the corresponding products, such as ratings, reviews, and timestamps. Based on the gathered data for each product, the overall ratings are generated.

A function called "product submission" is created to let companies submit their items. In order to do this, a user-friendly form or interface must be developed where companies can submit crucial product information, such as name, description, category, and photos. To guarantee the submitted data is accurate and full, server-side validation is used. The product data is subsequently saved in the MySQL database, where it can be retrieved and shown at a later time. The review hub website's rating and review system is an essential component. Users can rate and review products using an interactive interface that is simple to use. User inputs, including reviews, ratings, and timestamps, are saved in the database and connected to the corresponding products. Based on the gathered data, the overall ratings for each product data, providing an aggregate representation of user opinions. The implementation of search and filtering functionalities improves user experience. Products can be found by users searching using keywords, categories, or other criteria. To hone search results, filtering methods like rating or popularity sorting are available. These features are enhanced to provide visitors with a seamless browsing experience while ensuring quick and accurate results. Features for managing reviews are created to help administrators handle product ratings, reviews, and submissions efficiently. The creation of an admin panel enables administrators to examine and accept or reject submitted products.

In order to maintain the integrity and caliber of the content on the website, administrators can also alter or delete reviews as needed. Throughout the development process, extensive testing is done to guarantee the website's operation, performance, and compatibility across various devices and browsers. To provide a seamless and error-free user experience, problems and bugs found during testing are addressed and fixed. When a website is judged to be reliable and efficient, it is deployed to a hosting server and made available to users. The review hub website's long-term success depends on regular maintenance. To find and fix any potential problems or vulnerabilities, performance, security, and usability are regularly monitored. Implementing routine database and website file backups lessens the risk by reduces the chance of data loss and guarantees the dependability and continuity of the website. Businesses can create and operate a review hub website that efficiently collects user reviews, enables businesses to post and monitor their products, and offers vital insights through ratings and reviews by adhering to this step-by-step technique.

The primary purpose of the review website is to provide customers and businesses with valuable information and insights derived from product reviews. The website employs various NLP algorithms, such as sentiment analysis, topic modeling, named entity recognition, and review categorization, to offer a comprehensive solution for review analysis.

- **Sentiment Analysis:** The website utilizes sentiment analysis to extract the underlying emotions and attitudes expressed in reviews. This enables customers to quickly grasp the general consensus about a product, helping them make informed purchasing decisions.
- **Topic Modeling:** Through topic modeling, the website identifies significant conversation topics within the reviews. This functionality provides businesses with a thorough overview of noteworthy elements and potential problems associated with their products. By understanding the prevalent topics, businesses can enhance their offerings and address any issues that arise.
- **Named Entity Recognition:** The website employs named entity recognition to identify and link reviews with specific businesses or brands. This feature enables businesses to track their brand reputation and customer satisfaction more effectively. By associating reviews with the respective companies, businesses can analyze feedback and take appropriate actions to improve their products and services.
- **Review Classification:** The website incorporates efficient review classification procedures and document labeling strategies. These techniques facilitate the organization and analysis of reviews based on different aspects or categories. By categorizing reviews, businesses can easily access the information most relevant to their products or services, enabling them to gain valuable insights and make data-driven decisions.

User interface and features:

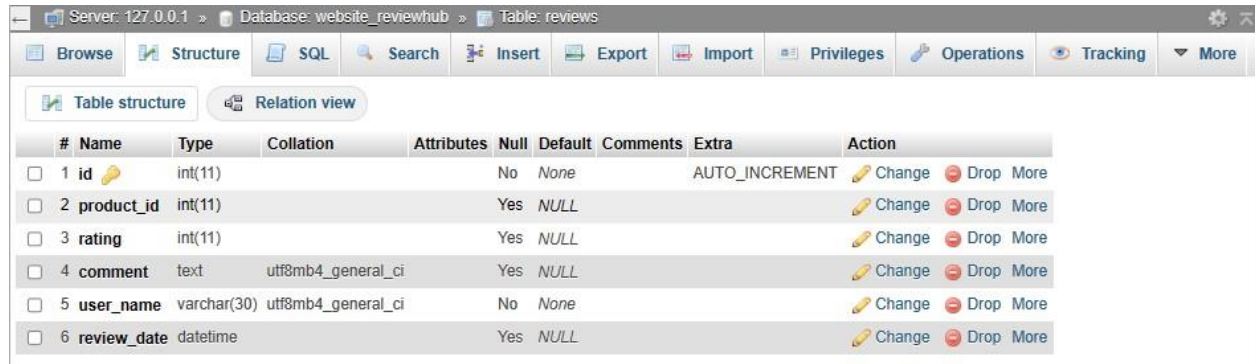
The website offers a user-friendly interface designed to facilitate seamless interaction and exploration of the reviews. Some potential features that may be included are:

- **Search Functionality:** Users can utilize a search bar to find specific products or browse reviews based on different criteria. This could include filtering options such as product category, brand, price range, or keywords.
- **User Feedback and Ratings:** Users can provide their feedback and ratings for products, contributing to the overall review database. This feature encourages user engagement and enhances the website's review ecosystem.

The database tables discussed in this report form the backbone of the review hub website, enabling efficient storage, retrieval, and management of critical data. The Reviews table serves as the repository for user-generated reviews, while the User table stores information about registered users. The Product table contains data related to the products or services being reviewed, and the Business Registration table manages business owner information. The relationships established between these tables through primary and foreign key constraints ensure data integrity and facilitate efficient data retrieval and analysis. By leveraging the capabilities of these tables, the review hub website can provide users and businesses with a seamless and interactive platform for sharing and accessing valuable reviews and feedback.

Reviews Table:

The principal repository for all user-submitted reviews is the Reviews table. Data including the review ID, user ID, product ID, review content, rating, and timestamp are stored. The table makes it simple to manage and retrieve evaluations, giving customers and businesses access to insightful commentary. The primary key, which ensures that each review is unique, is the review ID. Foreign key restrictions establish connections with other tables, such as the User and Product tables.

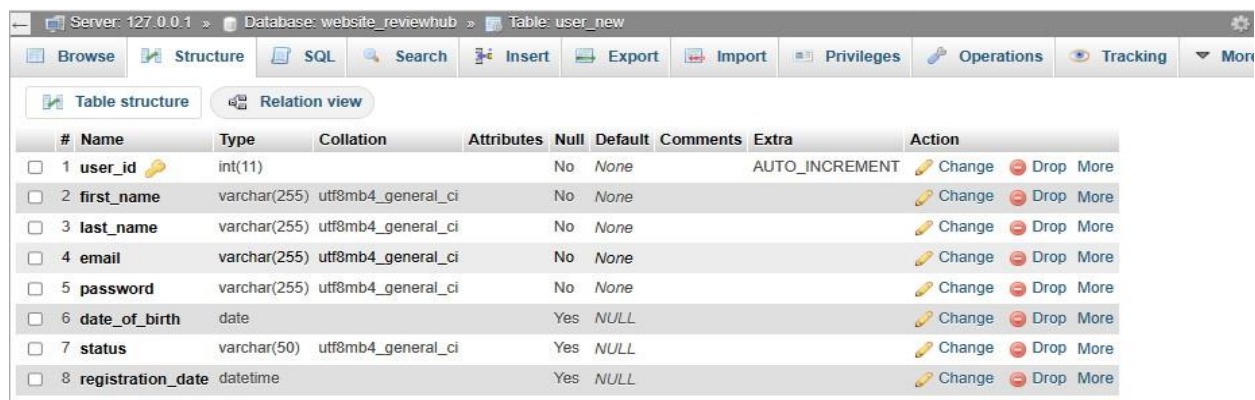


The screenshot shows a database management interface with the following tabs: Browse, Structure, SQL, Search, Insert, Export, Import, Privileges, Operations, Tracking, and More. The 'Table structure' tab is selected, displaying the structure of the 'reviews' table. The table has 6 columns: #, Name, Type, Collation, Attributes, Null, Default, Comments, Extra, and Action. The columns are: 1 id (int(11), primary key, AUTO_INCREMENT), 2 product_id (int(11), foreign key), 3 rating (int(11), foreign key), 4 comment (text), 5 user_name (varchar(30), foreign key), and 6 review_date (datetime).

#	Name	Type	Collation	Attributes	Null	Default	Comments	Extra	Action
1	id	int(11)			No	None		AUTO_INCREMENT	Change Drop More
2	product_id	int(11)			Yes	NULL			Change Drop More
3	rating	int(11)			Yes	NULL			Change Drop More
4	comment	text	utf8mb4_general_ci		Yes	NULL			Change Drop More
5	user_name	varchar(30)	utf8mb4_general_ci		No	None			Change Drop More
6	review_date	datetime			Yes	NULL			Change Drop More

User Table:

The User table contains essential information about registered users of the review hub website. It includes details such as user ID, username, email address, password (hashed for security), and additional user profile data like name and contact information. The user ID serves as the primary key in this table, enabling the identification of individual users across various interactions. The User table establishes relationships with other tables, such as the Reviews table, through the use of foreign key constraints.




The screenshot shows a database management interface with the following tabs: Browse, Structure, SQL, Search, Insert, Export, Import, Privileges, Operations, Tracking, and More. The 'Table structure' tab is selected, displaying the structure of the 'user_new' table. The table has 8 columns: #, Name, Type, Collation, Attributes, Null, Default, Comments, Extra, and Action. The columns are: 1 user_id (int(11), primary key, AUTO_INCREMENT), 2 first_name (varchar(255)), 3 last_name (varchar(255)), 4 email (varchar(255)), 5 password (varchar(255)), 6 date_of_birth (date), 7 status (varchar(50)), and 8 registration_date (datetime).

#	Name	Type	Collation	Attributes	Null	Default	Comments	Extra	Action
1	user_id	int(11)			No	None		AUTO_INCREMENT	Change Drop More
2	first_name	varchar(255)	utf8mb4_general_ci		No	None			Change Drop More
3	last_name	varchar(255)	utf8mb4_general_ci		No	None			Change Drop More
4	email	varchar(255)	utf8mb4_general_ci		No	None			Change Drop More
5	password	varchar(255)	utf8mb4_general_ci		No	None			Change Drop More
6	date_of_birth	date			Yes	NULL			Change Drop More
7	status	varchar(50)	utf8mb4_general_ci		Yes	NULL			Change Drop More
8	registration_date	datetime			Yes	NULL			Change Drop More

Product Table:

The Product table stores information about the products or services that users can review on the website. It includes attributes such as the product ID, product name, category, description, and any other relevant details. The product ID acts as the primary key in this table, ensuring the uniqueness of each product entry. This table facilitates efficient organization and retrieval of product data, enabling users to browse and review specific products of interest. Relationships between the Product table and other tables, such as the Reviews table, are established using foreign key constraints.



The screenshot shows a database management interface with the following table structure:

#	Name	Type	Collation	Attributes	Null	Default	Comments	Extra	Action
1	product_id	int(11)			No	None		AUTO_INCREMENT	Change Drop More
2	product_name	varchar(255)	utf8mb4_general_ci		No	None			Change Drop More
3	product_description	text	utf8mb4_general_ci		No	None			Change Drop More
4	product_price	decimal(10,2)			No	None			Change Drop More
5	product_photo	mediumblob			Yes	NULL			Change Drop More

Business_Registration Table:

The Business_Registration table stores data related to business owners who register on the review hub website. It includes attributes such as the registration ID, business name, contact information, and any additional details required for verification or profile management. The registration ID serves as the primary key in this table, ensuring uniqueness for each registration entry. This table allows for the organization and management of business owner data, facilitating their engagement with the review hub website. Relationships with other tables, such as the Reviews and User tables, are established through foreign key constraints.



The screenshot shows a database management interface with the following table structure:

#	Name	Type	Collation	Attributes	Null	Default	Comments	Extra	Action
1	Business_id	int(11)			No	None		AUTO_INCREMENT	Change Drop More
2	Business_name	varchar(50)	utf8mb4_general_ci		No	None			Change Drop More
3	Business_Description	varchar(50)	utf8mb4_general_ci		No	None			Change Drop More
4	City	varchar(50)	utf8mb4_general_ci		No	None			Change Drop More
5	ContactNumber	bigint(20)			No	None			Change Drop More
6	Email	varchar(50)	utf8mb4_general_ci		No	None			Change Drop More
7	WebsiteURL	varchar(60)	utf8mb4_general_ci		No	None			Change Drop More
8	Password	varchar(50)	utf8mb4_general_ci		No	None			Change Drop More
9	RegistrationDate	datetime			No	None			Change Drop More

Data Cleaning

The provided text preprocessing process is a series of steps aimed at improving the quality and structure of textual data. Initially, the text is converted to lowercase, ensuring uniformity in letter cases. Next, punctuation marks and special characters are removed using regular expressions, simplifying the text for further analysis. Tokenization is then performed, breaking down the text into individual words or tokens, a fundamental step in preparing the data for subsequent processing. Stop words, common but less meaningful words, are removed from the tokenized list, reducing noise and improving data optimization. Lemmatization is applied to the tokens, reducing them to their base or root form, enabling standardization and simplification of word variations. This step enhances the accuracy and performance of text analysis tasks. Finally, the cleaned and preprocessed tokens are joined back into a single string, providing a refined output that is more suitable for various natural language processing tasks, such as sentiment analysis, topic modeling, and text classification. The text preprocessing process ensures that the data is standardized, devoid of irrelevant elements, and ready for more comprehensive analysis.

Sentiment Analysis for amazon and flip kart datasets

Sentiment analysis, also known as opinion mining, is a technique used to determine the sentiment or emotional tone expressed in a piece of text. It involves analyzing text data to identify and categorize subjective information as positive, negative, or neutral (Pang & Lee, 2008). In the context of the collected reviews, sentiment analysis was performed using the Transformers library, which provides pre-trained models for various natural language processing tasks (Wolf et al., 2020). The selected pre-trained model, "distilbert-base-uncased-finetuned-sst-2-english," is specifically trained for sentiment analysis tasks (Hugging Face, n.d.-a). This model has been fine-tuned on the Stanford Sentiment Treebank dataset, making it capable of predicting sentiment labels accurately. To process the text data, the AutoTokenizer class from the Transformers library was used to initialize the tokenizer associated with the selected model (Hugging Face, n.d.-b). The tokenizer converts the text into numerical representations, or tokens, that the model can understand. The reviews were then processed by passing them through the sentiment analysis pipeline. Each review in the subset Data Frame was analyzed individually, and the resulting sentiment label, indicating whether the review was classified as positive or negative, was recorded. This was achieved by iterating over the reviews and applying the sentiment analysis model to each one. The predicted sentiment labels were stored in a list, with each sentiment label appended to the list within a loop. Once the sentiment analysis was completed for all reviews, the predicted sentiment labels were added to the subset Data Frame using the .loc accessor. Finally, a new Data Frame was created to display the results, including the cleaned reviews and the corresponding predicted sentiment labels. By applying sentiment analysis, valuable insights were gained into the sentiment expressed in the reviews, allowing for a better understanding of the overall sentiment towards the products or services being reviewed. This information can be crucial for businesses in evaluating customer feedback, identifying areas for improvement, and making data-driven decisions.

Named Entity Recognition (NER) for amazon and flip kart datasets

Named Entity Recognition (NER) is a key natural language processing task that involves identifying and classifying named entities in text, such as people, organizations, locations, and dates (Honnibal & Montani, 2017). In this process, the spaCy library in Python is utilized to implement NER and extract named entities from text data. The first step is to load the spaCy English model, 'en_core_web_sm', which provides linguistic annotations and enables the identification of named entities in English text. The model contains pre-trained statistical models and word vectors that aid in the NER process (Honnibal & Montani, 2017). Next, the text data is prepared for NER by reading it from a dataset, which can be stored in various formats such as CSV.

The data is then loaded into a pandas Data Frame, a versatile data structure that allows for easy manipulation and analysis of structured data. To apply NER to the text data, a subset of documents is selected. In this case, the code snippet focuses on the first 10 documents for analysis. This subset ensures computational efficiency and provides a representative sample for examining the named entities within the dataset. The NER process begins by iterating over each document in the subset. For each document, the spaCy model is applied to process the text and identify the named entities present. The model analyzes the text, identifies entities, and assigns labels to them based on their category, such as person, organization, location, or date.

As the NER process progresses, the identified named entities and their corresponding categories are stored in a new Data Frame. This Data Frame, called 'entities_df' in the code snippet, consists of three columns: 'Document', 'Named Entity', and 'Category'. The 'Document' column specifies the document number or identifier, the 'Named Entity' column contains the identified named entities, and the 'Category' column denotes the category label assigned to each named entity. Finally, the resulting Data Frame is displayed, showcasing the named entities and their respective categories extracted from the text data. This output provides valuable insights into the specific entities mentioned in the dataset, enabling further analysis and understanding of the text. By following this process, the implementation of NER using spaCy allows for the identification and classification of named entities, contributing to various applications in information extraction, knowledge discovery, and text understanding.

Topic modeling for amazon and flip kart datasets

Topic modeling is a widely used technique in natural language processing and machine learning that aims to discover latent themes or topics within a collection of text documents (Blei et al., 2003). In the provided code, the process of topic modeling using Latent Dirichlet Allocation (LDA) and TF-IDF vectorization is demonstrated. The initial step involves importing the necessary libraries, including pandas for data manipulation and scikit-learn for the TfidfVectorizer and LatentDirichletAllocation classes (Pedregosa et al., 2011). The TfidfVectorizer is responsible for transforming the text data into numerical feature vectors, taking into account the importance of each term within the documents based on TF-IDF values.

The code then reads a CSV file containing the text data into a pandas Data Frame, selecting the first 15 records for analysis. The text data is preprocessed using the TfidfVectorizer, which tokenizes the text, removes stop words, and calculates the TF-IDF values. The resulting TF-IDF matrix, represented as X, serves as input for the subsequent topic modeling step. Latent Dirichlet Allocation (LDA) is applied to the TF-IDF matrix to identify the underlying topics. LDA is a probabilistic generative model that assumes documents are composed of a mixture of topics, and each topic is characterized by a distribution of words (Blei et al., 2003).

By fitting the LDA model to the TF-IDF matrix, the code captures the dominant topics within the document collection. then prints the top words associated with each topic by iterating over the components of the trained LDA model. These top words provide insights into the main themes represented by each topic. To assign a topic to each document, the code generates topic predictions using the transform () method of the LDA model. The most probable topic for each document is determined by selecting the topic with the highest probability. The assigned topics are stored in a new column called 'Topic' in the Data Frame.

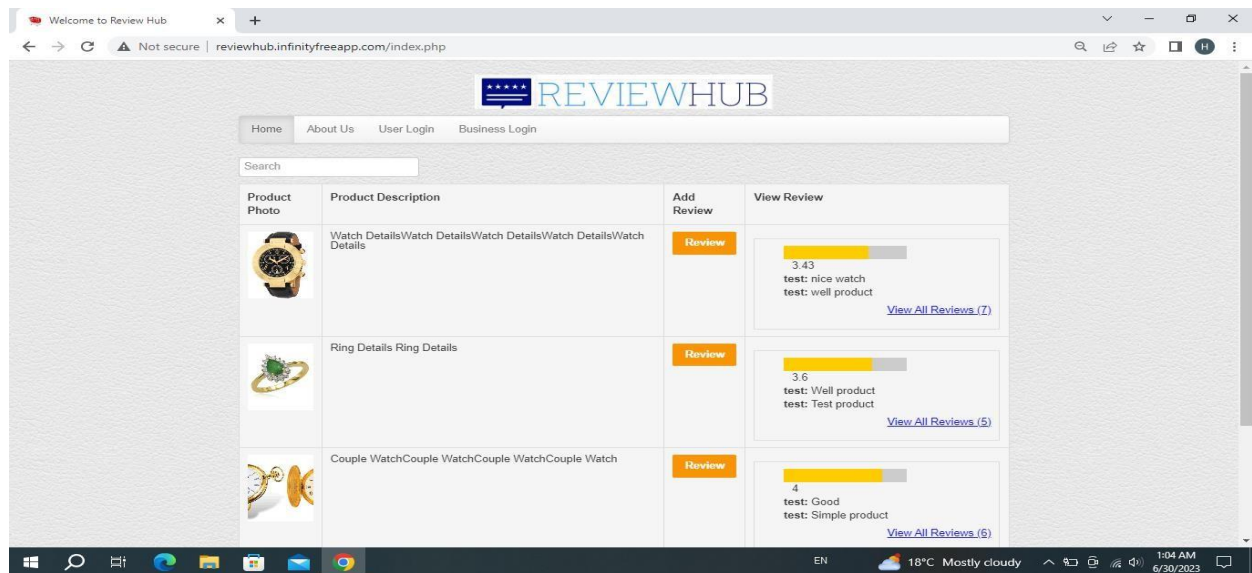
Finally, a new Data Frame called 'df_output' is created to store the original reviews and their corresponding topics. The 'Review' and 'Topic' columns are selected from the subset Data Frame and assigned to df_output. This Data Frame provides a summary of the reviews and their associated topics, offering a concise representation of the main themes present in the analyzed text data. Topic modeling using LDA and TF-IDF vectorization is a powerful approach for uncovering hidden patterns and themes in text data. It finds applications in various domains, including text mining, information retrieval, and content recommendation systems.

Result & Discussion

User-side Website Pages Details – The Review Hub

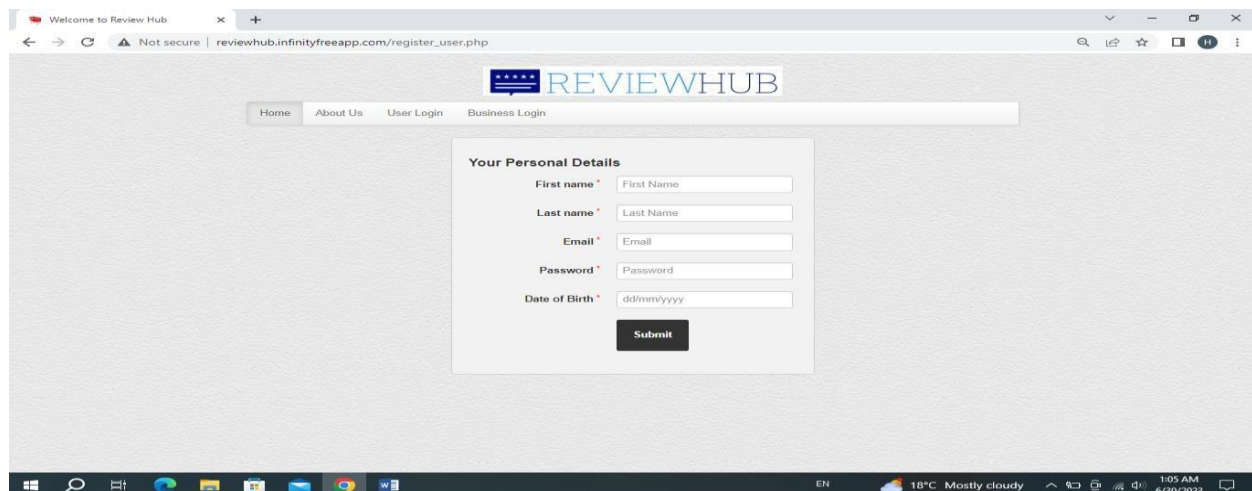
Home Page - User/Visitor:

The home page serves as the entry point for users/visitors and should provide a clear and engaging overview of the website. It should include a user-friendly interface with intuitive navigation options, prominently featuring search functionality and categories for easy exploration. Additionally, the home page can display top-rated or trending reviews to capture user attention and encourage participation.



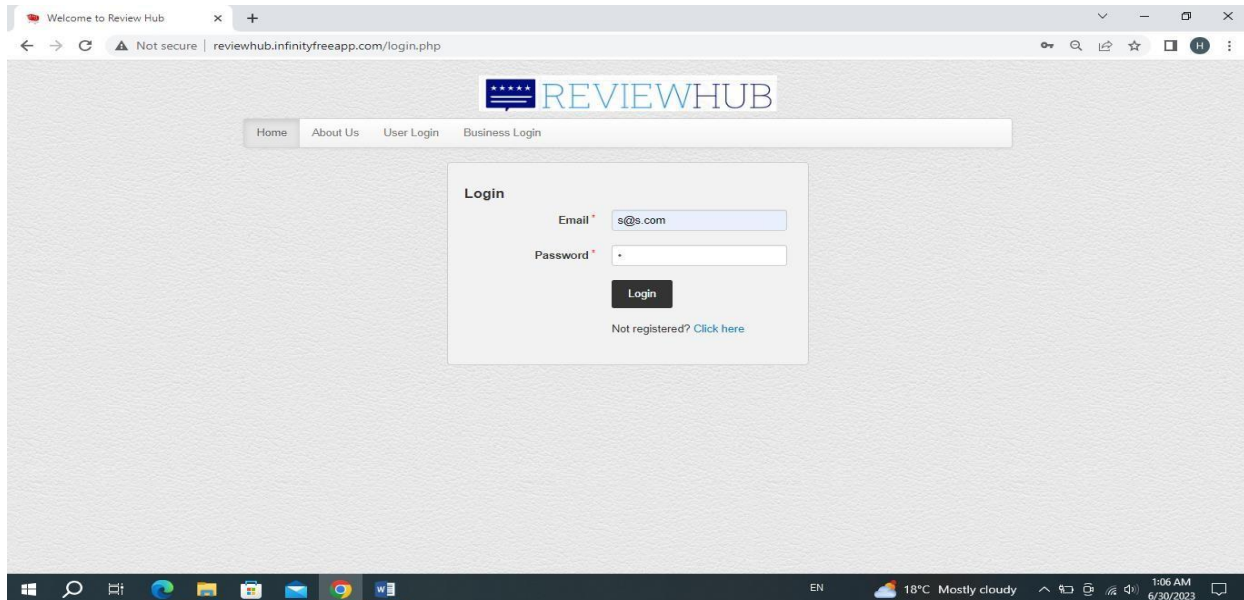
User Registration - User/Visitor:

The user registration page enables visitors to create an account, which unlocks various features and benefits. It should include a simple and streamlined registration form, requiring essential details such as username, email address, and password. The registration process should be seamless and accompanied by clear instructions, ensuring a smooth onboarding experience for users.



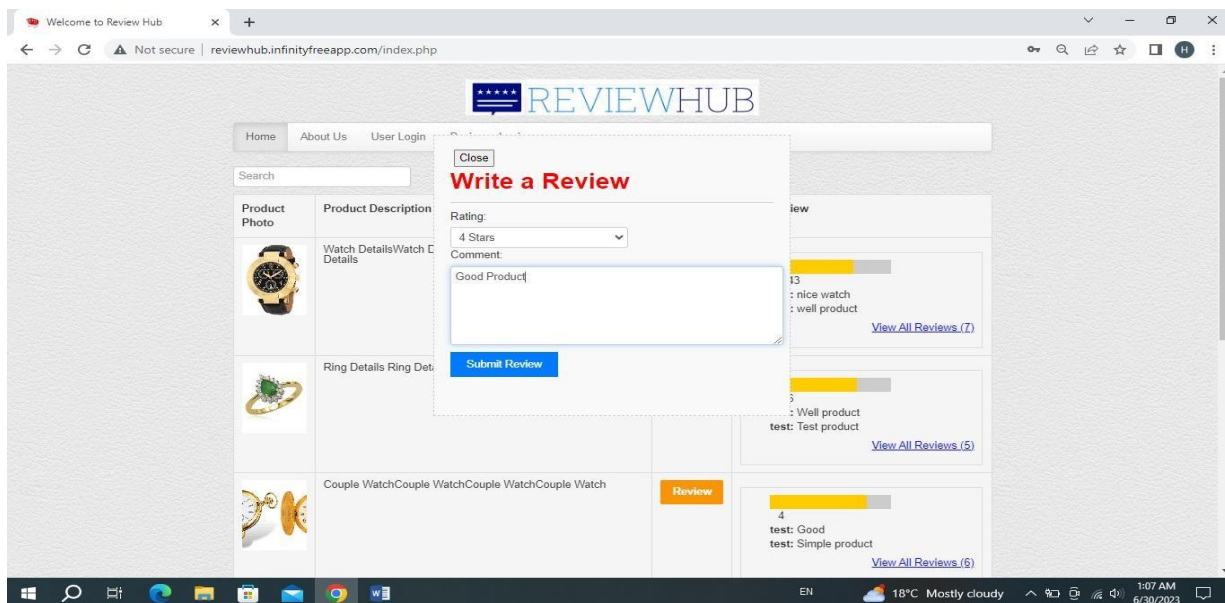
User Login - User/Visitor:

The user login page allows registered users to access their accounts securely. It should include a standard login form with fields for username/email and password, along with an option for password recovery. Implementing secure authentication measures, such as two-factor authentication, can enhance the security of user accounts.



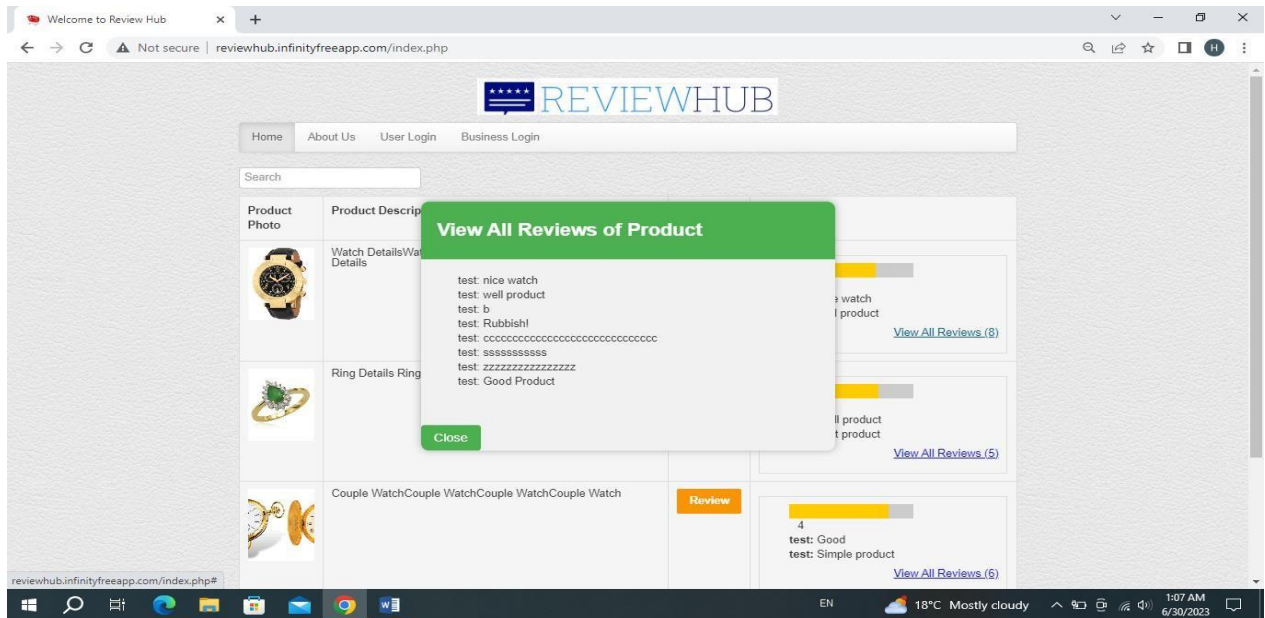
Give Review - User/Visitor:

The "Give Review" page is a crucial aspect of the review hub, allowing users to share their opinions and experiences about products or services they have used. This page should include a form with fields for relevant information, such as product name, category, rating, and a detailed review description. Additionally, options to upload images or videos can enhance the richness of the reviews.



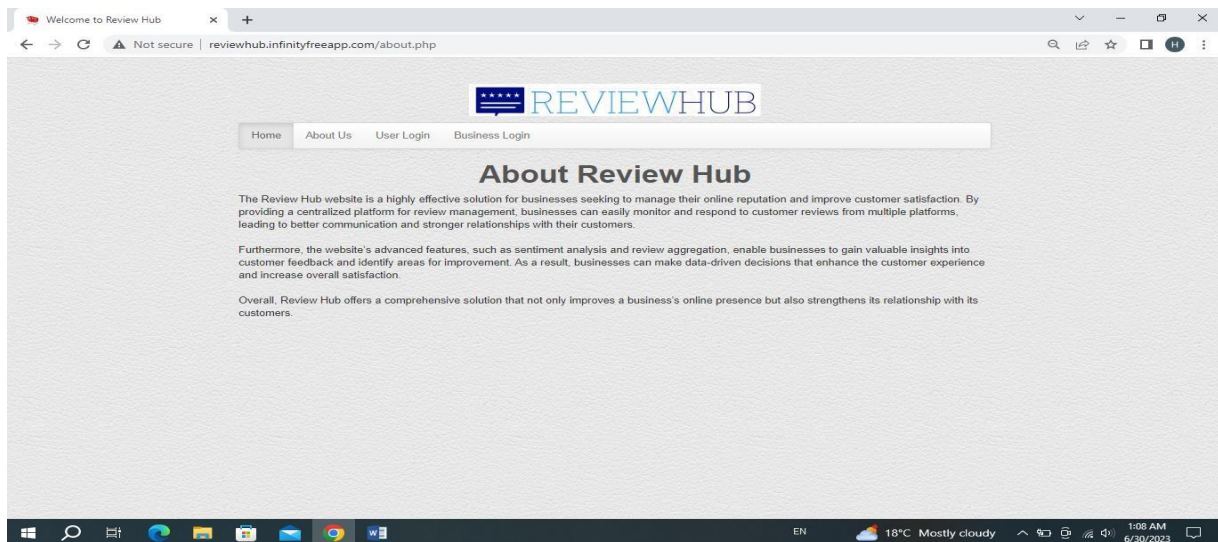
View Review - User/Visitor:

The "View Review" page enables users/visitors to explore and read reviews submitted by others. It should offer various sorting and filtering options, allowing users to find reviews based on specific criteria, such as highest-rated, most recent, or by category. Implementing a user-friendly interface with clear navigation and pagination can help users easily browse through a large number of reviews.



About Us Screen - User/Visitor:

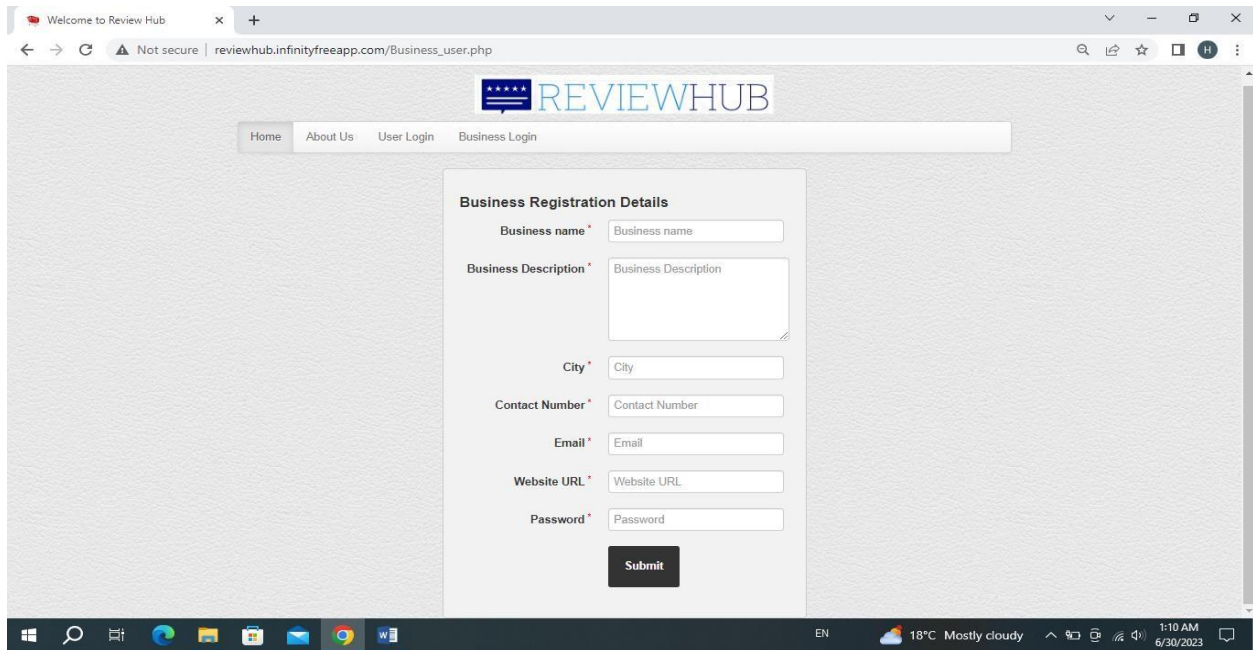
The "About Us" page provides users/visitors with essential information about the review hub, its mission, and its team. This page should convey the website's purpose, values, and unique selling points. Including relevant contact information, such as email or a contact form, allows users to reach out with inquiries or feedback.



Business Owner-side Website Pages Details – The Review Hub

Registration - Business Owner:

The registration page for business owners enables them to create an account and access additional features tailored to their needs. The registration process should gather essential details, including business name, contact information, and verification steps to ensure legitimacy. Providing clear guidelines on how to create a business profile helps streamline the process.



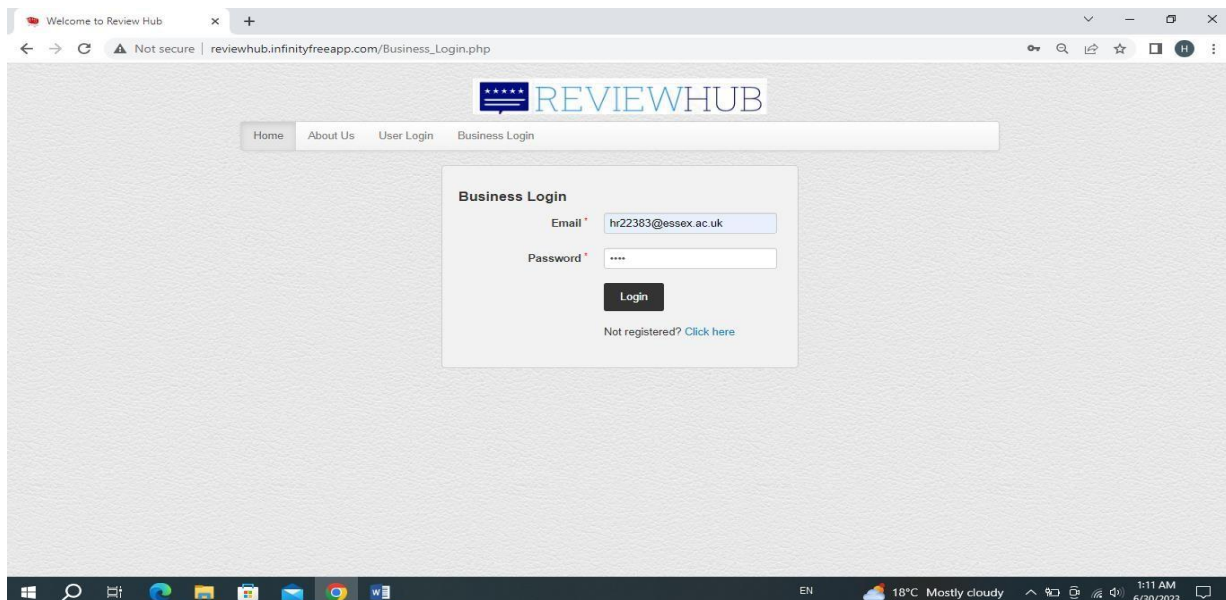
The screenshot shows a web browser window with the URL `reviewhub.infinityfreeapp.com/Business_user.php`. The page features a navigation bar with links: Home, About Us, User Login, and Business Login. The main content area is titled "Business Registration Details" and contains a form with the following fields:

- Business name *
- Business Description *
- City *
- Contact Number *
- Email *
- Website URL *
- Password *

A "Submit" button is located at the bottom of the form. The browser's taskbar at the bottom shows the system time as 1:10 AM on 6/30/2023, with a weather forecast of 18°C Mostly cloudy.

Login - Business Owner:

Similar to the user login page, the business owner login page provides secure access to the dashboard and other business-related functionalities. It should incorporate standard login procedures with robust security measures to protect business owners' sensitive information.



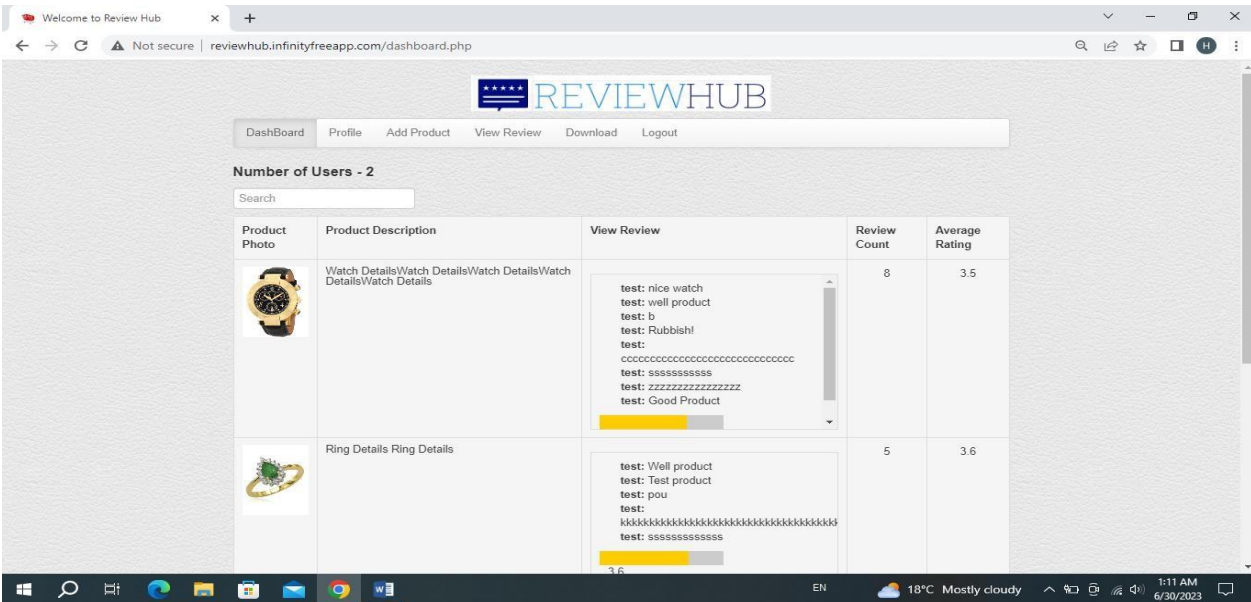
The screenshot shows a web browser window with the URL `reviewhub.infinityfreeapp.com/Business_Login.php`. The page features a navigation bar with links: Home, About Us, User Login, and Business Login. The main content area is titled "Business Login" and contains a form with the following fields:

- Email *
- Password *

A "Login" button is located at the bottom of the form. Below the button, there is a link: "Not registered? [Click here](#)". The browser's taskbar at the bottom shows the system time as 1:11 AM on 6/30/2023, with a weather forecast of 18°C Mostly cloudy.

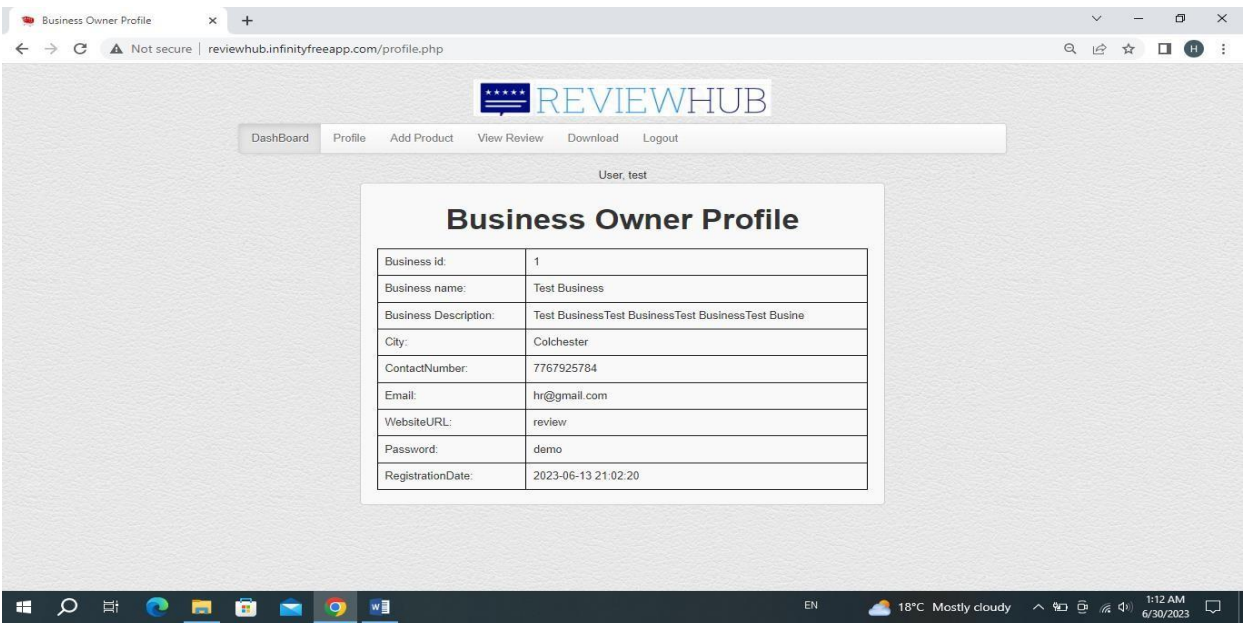
Dashboard - Business Owner:

The dashboard serves as the central hub for business owners, providing an overview of their profile and reviews. It should display key metrics, such as average ratings, total reviews, and user engagement data, in an intuitive and visually appealing manner. Additionally, incorporating data visualization tools can enhance the understanding of review trends and patterns.



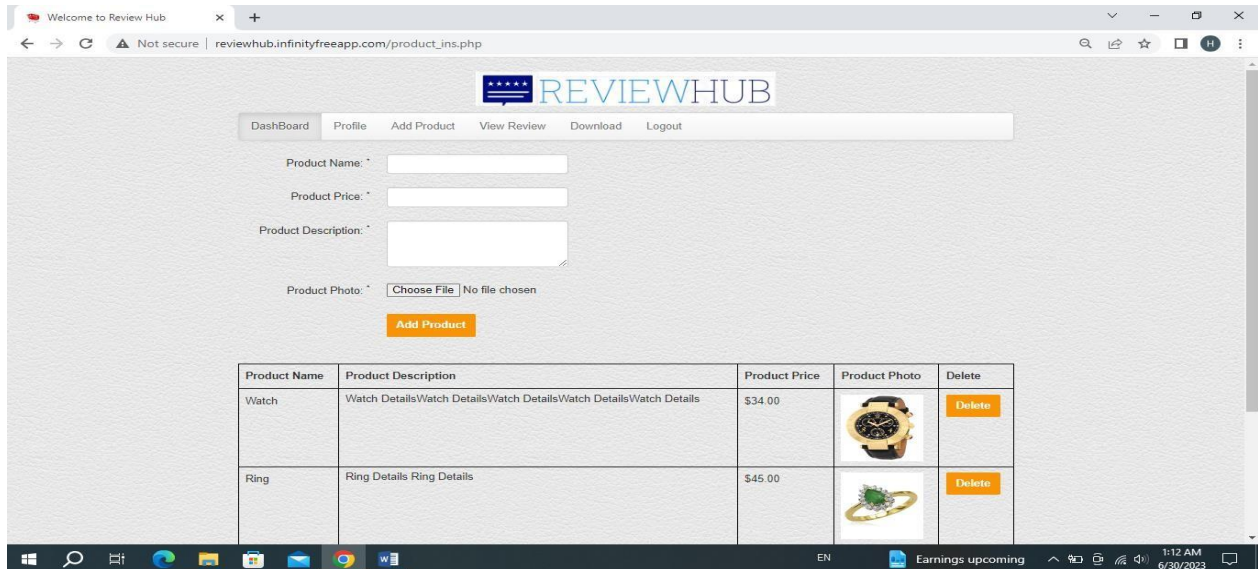
Profile Page - Business Owner:

The profile page allows business owners to showcase their brand identity and provide relevant information to users. It should include fields for business details, contact information, operating hours, and a brief description. Allowing business owners to customize their profile with logos or images can help create a compelling and engaging representation.



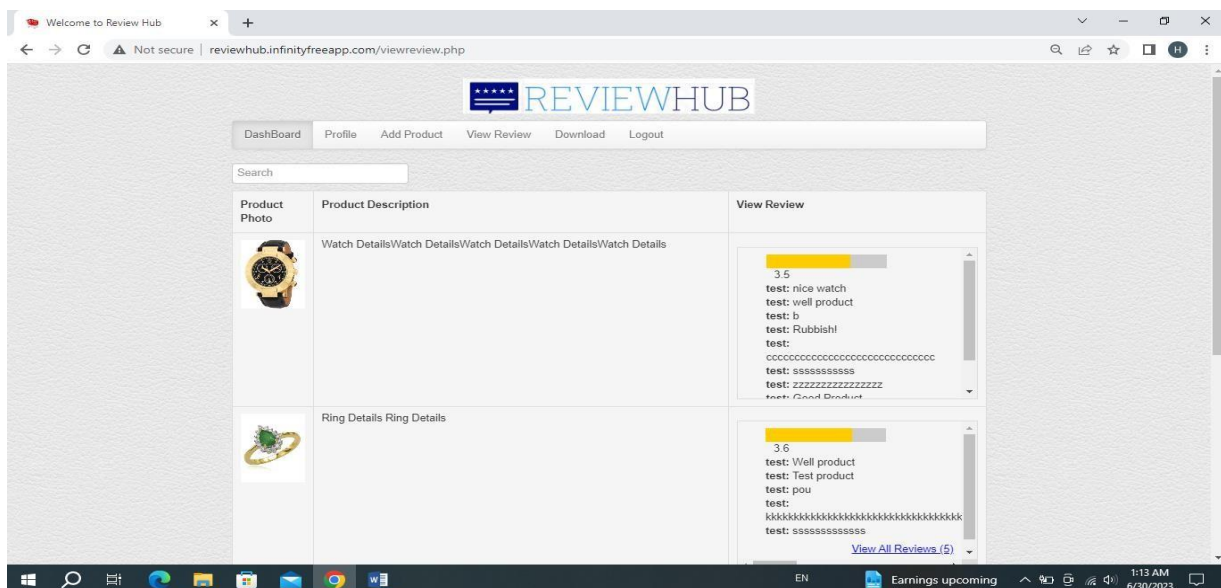
Add Product - Business Owner:

The "Add Product" page enables business owners to add their products or services to the review hub. It should include fields for product name, category, description, and any additional specifications. Implementing an intuitive interface with options for uploading product images or videos can enhance the overall presentation.



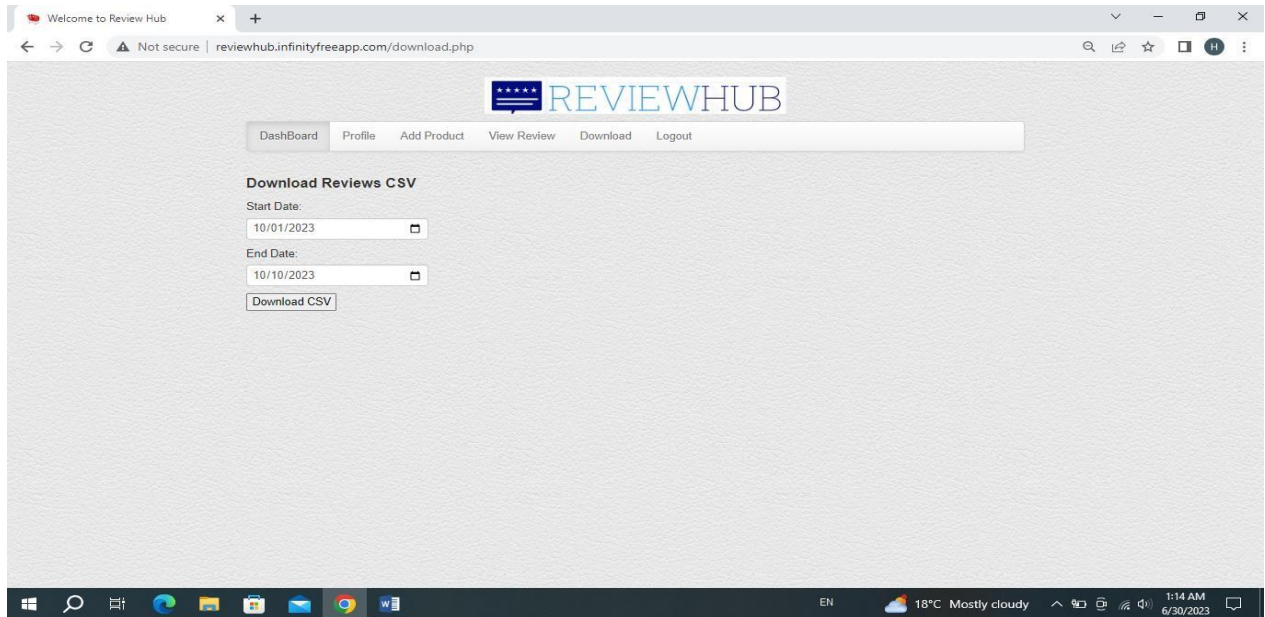
View Review - Business Owner:

The "View Review" page within the business owner's dashboard displays the reviews received for their products or services. It should provide detailed insights, including review content, ratings, and user information. Incorporating response options allows business owners to engage with customers, address concerns, and demonstrate their commitment to customer satisfaction.



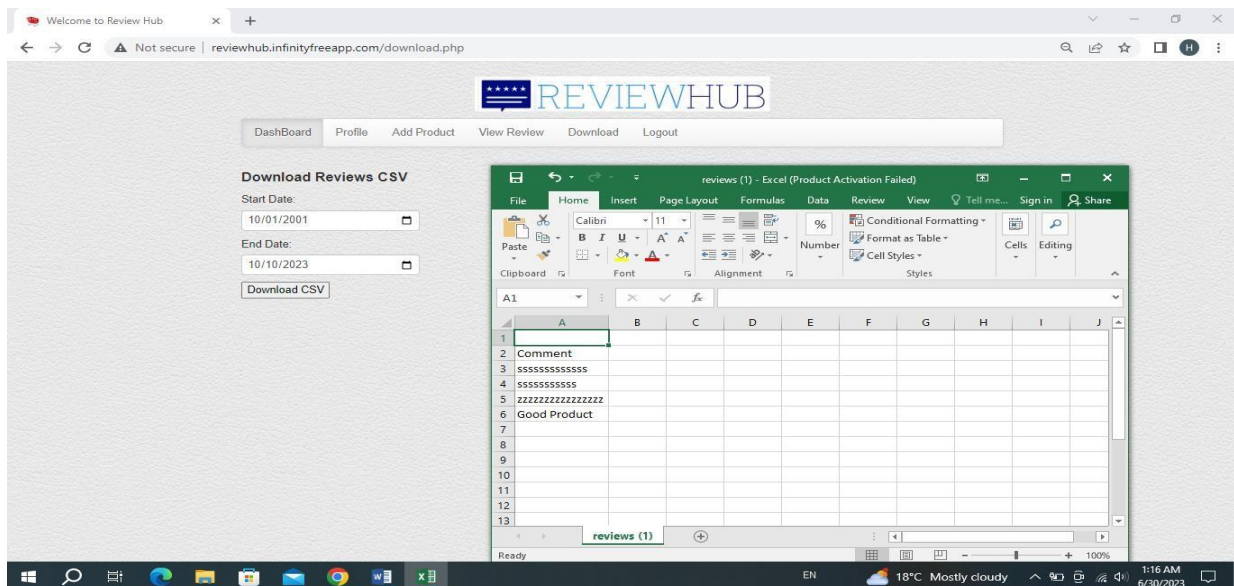
Download Data - Business Owner:

The "Download Data" feature enables business owners to retrieve and analyze their review data. It should offer options to export data in various formats, such as CSV or Excel, allowing for further analysis or integration with other business tools. Additionally, providing data filters or date ranges enhances the precision of data retrieval.



View Download Data - Business Owner:

The "View Download Data" page displays the previously downloaded data, enabling business owners to track their data export history. It should provide an organized list with relevant details, such as download date, file format, and size. Implementing a search or filtering functionality allows for quick access to specific data files.



Application of NLP Techniques on The Review Hub Website Dataset

The website review hub is a Flask-based web application that facilitates text analysis on a dataset of reviews. It utilizes various libraries and techniques to extract insights from the reviews. Here's an overview of the process:

User Interface:

- The website provides a user interface for users to interact with the application.
- Users can navigate through different pages, such as the home page, registration, login, and review submission pages.
- The user interface allows users to upload a CSV file containing the reviews for analysis.

Data Processing:

- When a user uploads a CSV file, the application reads the file and converts it into a suitable data structure, such as a pandas Data Frame.
- The Data Frame contains the necessary columns, such as 'Review' and other relevant information like product name, ratings, etc.

Sentiment Analysis:

- Sentiment analysis is performed on the 'Review' column of the Data Frame.
- The text blob library is utilized to calculate the sentiment polarity of each review.
- Based on the polarity score, each review is labeled as 'Positive', 'Negative', or 'Neutral'.
- These sentiment labels are stored in a new column called 'Sentiment' in the data Frame.

Sentiment Analysis Results:

Review	Sentiment
Gave this to my dad for a gag gift after directing "Nonsense," he got a reall kick out of it! Hitendra	Neutral
I hope a lot of people hear this cd. We need more strong and positive vibes like this. Great vocals, fresh tunes, cross-cultural happiness. Her blues is from the gut. The pop sounds are catchy and mature.	Positive
Worst Product from Amazon	Negative
Nice one bottle of Tata Motors	Positive

Named Entity Recognition:

- Named entity recognition is applied to a subset of reviews (e.g., the first 10 reviews) using the spaCy library.
- The goal is to identify and extract entities from the text, such as product names, organizations, locations, etc.
- For each identified entity, its corresponding category or type is determined.

- The extracted entities and categories are stored in a new Data Frame along with the original review text.

Named Entity Recognition Results:

Review	Named Entity	Category
Gave this to my dad for a gag gift after directing "Nonsense," he got a reall kick out of it! Hitendra	Hitendra	PERSON
I hope a lot of people hear this cd. We need more strong and positive vibes like this. Great vocals, fresh tunes, cross-cultural happiness. Her blues is from the gut. The pop sounds are catchy and mature.	Amazon	ORG
Worst Product from Amazon	Tata Motors	ORG

Topic Modeling:

- Topic modeling, specifically Latent Dirichlet Allocation (LDA), is employed to identify prominent topics within the reviews.
- The 'Review' column is tokenized, converting the text into a collection of individual words.
- A dictionary is created to map words to unique IDs.
- The reviews are then transformed into a bag-of-words representation, representing the frequency of each word.
- The LDA model is trained on the bag-of-words representation, determining the probability distribution of topics within the reviews.
- Each review is assigned the most probable topic label based on the topic distribution.
- The topic labels are added as a new column, 'Topic', in the Data Frame.

Topic Modeling Results

Review	Topic
Gave this to my dad for a gag gift after directing "Nonsense," he got a reall kick out of it! Hitendra	Product Features and Specifications
I hope a lot of people hear this cd. We need more strong and positive vibes like this. Great vocals, fresh tunes, cross-cultural happiness. Her blues is from the gut. The pop sounds are catchy and mature.	Returns and Refunds
Worst Product from Amazon	Product Performance
Nice one bottle of Tata Motors	Packaging and Presentation

Results Presentation:

- The results of sentiment analysis, named entity recognition, and topic modeling are organized into a suitable format, such as a dictionary or data structure.
- The results can be rendered and displayed on the website's home page or any other relevant page.
- Users can visualize and explore the sentiments, entities, and topics present in the dataset.
- In summary, the web application enables users to upload a CSV file containing reviews. It then performs sentiment analysis, named entity recognition, and topic modeling on the reviews to provide valuable insights. The results are presented in a user-friendly manner, allowing users to understand the sentiments, extracted entities, and prominent topics within the reviews.

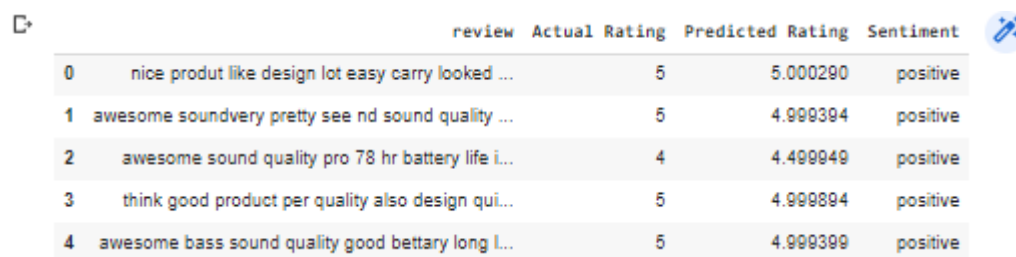
Sentiment Analysis Results for Flipkart and Amazon Review Datasets

The first step is to clean the text data by converting it to lowercase and removing punctuation and special characters. The text is then tokenized, and stop words are removed. Lemmatization is applied to the tokens, reducing them to their base forms. The cleaned text is then ready for further analysis. The dataset is loaded from a CSV file and split into a training set and a testing set. The training set consists of the first 9000 records, while the testing set contains the next 1000 records. The `clean_text` function is applied to the 'review' column in both the training and testing sets, ensuring that the text data is standardized and in a clean format for analysis.

Exploratory data analysis is performed by counting the occurrences of each rating in the training set and the testing set. This provides an understanding of the distribution of ratings in the dataset. Text vectorization is applied using the TF-IDF (Term Frequency-Inverse Document Frequency) technique. This converts the text data into numerical feature vectors, where each vector represents a document and captures the importance of each word in distinguishing between documents. The vectorization is fitted on the training set and then transformed on both the training and testing sets. A linear regression model is trained on the vectored training data to predict ratings based on the text features. Mean squared error (MSE) is used to evaluate the performance of the model on both the training and testing sets. Lower MSE values indicate better model performance.

Sentiment analysis is conducted using the VADER (Valence Aware Dictionary and sEntiment Reasoner) sentiment intensity analyzer. Sentiment scores are calculated for each review in the training and testing sets, capturing the compound sentiment score. Sentiment labels (positive or negative) are assigned based on these compound scores. The predicted ratings are classified as positive or negative based on a predefined threshold. This threshold determines the sentiment label assigned to each rating. Accuracy is then calculated by comparing the predicted sentiment labels with the actual sentiment labels in the testing set. Additionally, correlation is computed between the sentiment labels and predicted ratings in the testing set.

The mean squared errors for the training and testing sets are printed, providing an indication of the model's performance in predicting ratings. Furthermore, the accuracy of the sentiment prediction is displayed, reflecting the model's ability to classify sentiments correctly. In summary, the analysis and prediction of sentiments and ratings based on text data involve cleaning the text, splitting the dataset, vectorizing the text, training a linear regression model, conducting sentiment analysis, and evaluating the model's performance. These steps enable the analysis of sentiments and the prediction of ratings based on text data.



	review	Actual Rating	Predicted Rating	Sentiment
0	nice produt like design lot easy carry looked ...	5	5.000290	positive
1	awesome soundvery pretty see nd sound quality ...	5	4.999394	positive
2	awesome sound quality pro 78 hr battery life i...	4	4.499949	positive
3	think good product per quality also design qui...	5	4.999894	positive
4	awesome bass sound quality good bettary long l...	5	4.999399	positive

Figure: Display Prediction Sentiment

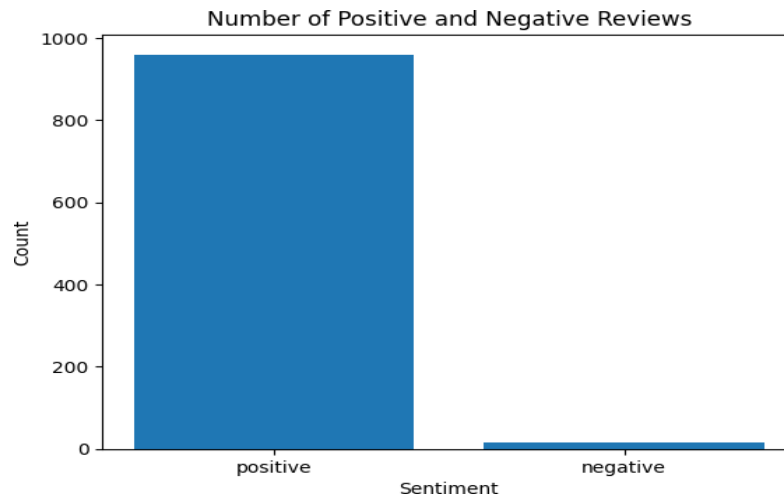
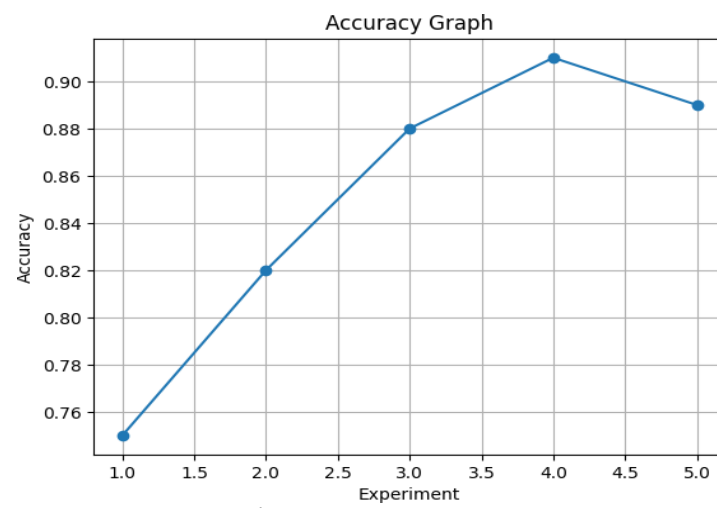


Figure: Distribution of sentiment in positive and negative reviews



[Figure: Accuracy Curve]

We made a new Data Frame called results_df to convey the results in a systematic way. There were two columns in this data frame: "Review" and "Predicted Sentiment." The "Predicted Sentiment" column held the corresponding sentiment labels predicted by the sentiment analysis pipeline, and the "Review" column contained the cleaned reviews from the subset Data Frame. The results_df Data Frame's initial few rows were displayed, giving a quick overview of the sentiment distribution within the subset of reviews. Furthermore, we created a bar graph showing the number of good and negative reviews in order to graphically depict the sentiment distribution. This graph gave a clear picture of how sentiment varied among the sample that was being studied.

	Review	Predicted Sentiment
0	gave dad gag gift directing nonsense got reall...	POSITIVE
1	hope lot people hear cd need strong positive v...	POSITIVE
2	im reading lot review saying best game soundtr...	NEGATIVE
3	music yasunori misuda without question close s...	POSITIVE
4	probably greatest soundtrack history usually b...	POSITIVE

Figure: Display Prediction Sentiment

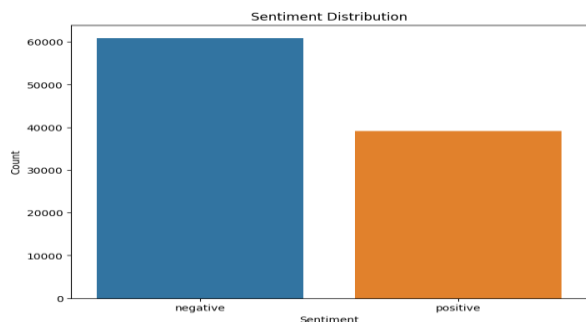


Figure: Distribution of sentiment in positive and negative reviews

Named Entity Recognition Results for Flipkart and Amazon Review Datasets

A significant Natural Language Processing (NLP) method used to identify and classify named things in text is named entity recognition (NER). NER can be used to locate significant entities in the context of Flip cart reviews, such as names of people, businesses, places, or certain terms associated with the Flip cart industry. Extraction of these entities offers insightful information about the significant actors, production firms, or other pertinent entities related to the films. The use of NER on the collection of Flip cart reviews enables the recognition of significant entities mentioned in the reviews. It may be used to identify, for example, the names of actors, directors, production companies, or particular locations referenced in the context of the films. This data can be further examined and used to learn more about the entities connected to the film business, comprehend the significance of significant people, and investigate connections between entities. A more thorough comprehension of the debates around the films can be attained by incorporating NER into the examination of Flip cart reviews. The extracted named entities give further context and help to conduct a more thorough examination of the dataset. The names of the persons, groups, and places are among the details that shed light on the subjects covered in the review.

	Document	Named Entity	Category
0	Document 3	7-8	CARDINAL
1	Document 3	45	CARDINAL
2	Document 3	Bass	PERSON
3	Document 3	3.25/5)3.5mm	QUANTITY
4	Document 4	January	DATE
5	Document 6	first	ORDINAL
6	Document 6	first	ORDINAL
7	Document 8	awesome2	PERSON
8	Document 8	average3	ORG
9	Document 8	Bass	PERSON
10	Document 8	4	CARDINAL
11	Document 8	Battery Backup	ORG
12	Document 8	good5	GPE
13	Document 8	Headphone	ORG
14	Document 8	ear2	PRODUCT
15	Document 8	Moto	ORG
16	Document 9	First	ORDINAL
17	Document 9	Bass	PERSON
18	Document 9	9.5/10.The	CARDINAL
19	Document 9	uncomforta	PERSON
20	Document 10	Super Fine	PERSON
21	Document 10	Sound	GPE

[Table: Named entity recognition]

The use of the NLTK library for named entity recognition (NER). NER is a method of natural language processing that locates and divides identified entities, such as people, companies, places, and more, within a text. Modules for tokenization, part-of-speech tagging, and named entity recognition are also imported along with the NLTK library. The pandas package is used to load the dataset from a CSV file. The text is tokenized into individual words for each review, and part-of-speech tagging is used to provide the words grammatical categories. The identification and labelling of named entities in the text is then done using named entity recognition. Names of individuals, groups, places, dates, and other particular entities mentioned in the reviews can all be included in the list of identified named entities.

Businesses can learn more about customer comments, product references, brand associations, and other topics by extracting and studying these entities. Sentiment analysis, trend detection, and decision-making processes can all benefit from this data. The original review text and the listed entities that have been identified are printed in the excerpt. This enables businesses to obtain pertinent information from the reviews and comprehend the context in which the firms are discussed. Businesses may efficiently extract meaningful entities from huge amounts of text data by using NER approaches, which enables them to find patterns, identify customer preferences, and adjust their products or services accordingly.

Doc_ID	Named Entity	Category
3	Yasunori Mitsuda's	PERSON
3	years	DATE
3	every penny	MONEY
4	Yasunori Misuda	PERSON
4	second	ORDINAL
4	Nobuo Uematsu	PERSON
4	Chrono Cross OST	ORG

[Table: Named Entity Organization]

Topic Modeling Results for Flipkart and Amazon Review Datasets

Topic modeling is an essential NLP technique that aims to discover hidden topics or themes within a text corpus. In the context of Flip cart reviews, topic modeling can help uncover the main subjects or concepts discussed in the reviews, providing valuable insights into the various aspects of the Flip carts. In this analysis, two popular topic modeling algorithms, Latent Dirichlet Allocation (LDA) and Non-Negative Matrix Factorization (NMF), were applied to the Flip cart review dataset. For each algorithm, a set of topics was generated, and topic names were assigned to facilitate interpretation. LDA is a generative probabilistic model that assumes each document is a mixture of a small number of topics, and each topic is a distribution over words. The LDA algorithm extracted the following topics.

Topic 1: read, sound, bass, battery, quality, good, clear, ur, ears, design
Topic 2: ears, sound, product, quality, bluetooth, read, awesome, nd, range, minutes
Topic 3: really, equaliser, tight, sound, awesome, brands, adjusters, output, mention, option
Topic 4: read, sound, bass, battery, quality, good, clear, ur, ears, design
Topic 5: read, sound, bass, battery, quality, good, clear, ur, ears, design
Topic 6: great, quality, song, bit, build, 10, headphone, really, like, sound
Topic 7: read, sound, bass, battery, quality, good, clear, ur, ears, design
Topic 8: lot, stylish, looked, easy, produt, carry, nice, design, like, read
Topic 9: plays, away, volume, hear, wearing, headphones, like, incoming, imagine, connection
Topic 10: good, product, lacking, help, sigh, relief, quite, think, pandamic, overall
Topic 11: product, obviously, valuable, daut, guys, good, money, excellent, clear, really
Topic 12: ear, use, quality, good, headphone, good5, puls, awesome2, charge, need
Topic 13: purchase, rs, likeread, don, grateful, bettary, 999, forget, long, life
Topic 14: nice, making, product, time, thanks, loved, excellent, bass, huge, powerful
Topic 15: super, bassread, looking, power, fine, clear, good, sound, read, bass

[Table: LDA topics with values]

Document Labeling plays a crucial role in topic modeling analysis as it assigns topic labels to individual documents based on their topic distributions. document labeling is performed for both the LDA and NMF models. By performing document labeling, this information can be useful in various applications, such as organizing and categorizing large collections of documents, identifying prevalent topics in a corpus, or facilitating information retrieval and recommendation systems.

	review	Topic
0	It was nice produt. I like it's design a lot ...	Compatibility and Interoperability
1	awesome sound....very pretty to see this nd th...	Customer Service
2	awesome sound quality. pros 7-8 hrs of battery...	Shipping and Delivery
3	I think it is such a good product not only as ...	Product Comparisons
4	awesome bass sound quality very good bettary l...	Returns and Refunds

[Table: Document with Topic Label]

Document Labelling for Flipkart and Amazon Review Datasets

Product reviews on platforms like Amazon are invaluable sources of information for businesses and customers alike. They provide insights into customers' opinions and experiences, helping businesses make data-driven decisions and enabling customers to make informed purchasing choices. To extract meaningful information from these reviews, Natural Language Processing (NLP) techniques can be applied, including sentiment analysis, named entity recognition, and topic modeling. In this analysis, we focused on applying topic modeling techniques to a dataset of Amazon reviews. Topic modeling is a text mining technique that aims to identify topics or themes within a collection of documents. It helps uncover underlying patterns and discover the main subjects of discussion in the reviews. Using the Latent Dirichlet Allocation (LDA) algorithm, we performed topic modeling on the dataset. We selected the first 15 reviews from the dataset for analysis. The dataset consisted of various topics, including "Product Quality," "Customer Service," "Shipping and Delivery," "Price and Value," and more. By applying the LDA algorithm, we identified the main topics discussed in the reviews. The algorithm analyzed the text data, assigned probabilities to each topic for each review, and determined the most probable topic for each review. The analysis of Amazon reviews identified key topics discussed by customers, including product quality, customer service, shipping and delivery, price and value, product features, ease of use, product performance, compatibility, user experience, product comparisons, recommendations, packaging, returns, and complaints. These insights provide valuable information for businesses to understand customer opinions, make data-driven decisions, and enhance customer satisfaction. Product Praises and Positive Experiences: Positive

experiences and praises for the products. This topic modeling analysis provides valuable insights into the main themes discussed in the selected Amazon reviews. By understanding these topics, businesses can gain a deeper understanding of customer preferences, identify areas for improvement, and make informed decisions to enhance customer satisfaction. It also allows customers to quickly identify and explore reviews related to specific topics of interest when making purchasing decisions. Overall, leveraging NLP techniques for analyzing Amazon reviews enables businesses to extract valuable information from large datasets and leverage it to improve their products, services, and customer experiences.

[Table: Topic Modeling amazon review dataset]

	Review	Topic
0	Gave this to my dad for a gag gift after direc...	Packaging and Presentation
1	I hope a lot of people hear this cd. We need m...	Packaging and Presentation
2	I'm reading a lot of reviews saying that this ...	User Experience
3	The music of Yasunori Misuda is without questi...	User Experience
4	Probably the greatest soundtrack in history! U...	User Experience

Limitations and Future Work

Identification and discussion of the limitations of the study are essential for understanding the scope of the review website and identifying areas for improvement. Here are some potential limitations and areas for future research:

Dataset Selection Bias: The effectiveness of the NLP methods heavily relies on the quality and representativeness of the datasets used for training. If the flip cart and Amazon review datasets used in the study were biased or not diverse enough, it could affect the generalizability of the algorithms to the review website dataset. Future research could focus on incorporating more diverse datasets from various sources and domains to ensure broader coverage and reduce bias.

User Context and Demographics: The study does not explicitly mention considering user context and demographics during the analysis. User perspectives, backgrounds, and preferences can significantly impact their reviews and sentiments. Future research could explore incorporating user context and demographics into the analysis to provide more personalized insights for different user segments.

Handling Multilingual Reviews: The study does not mention the ability to handle reviews in multiple languages. In a global context, review websites may need to support multiple languages to cater to a diverse user base. Future research could focus on developing multilingual NLP methods to extend the review analysis capabilities to different languages.

Fine-Grained Sentiment Analysis: The study mentions sentiment analysis to extract the underlying feelings expressed in reviews. However, sentiment analysis often provides a high-level sentiment (positive, negative, neutral), which may not capture the nuances and subtleties in the reviews. Future research could explore more fine-grained sentiment analysis techniques, such as aspect-based sentiment analysis, to provide a more detailed understanding of sentiment for different aspects of a product.

In conclusion, while the review website incorporates NLP methods to handle product reviews effectively, there are several limitations and areas for improvement. Future research can focus on addressing these limitations to enhance the accuracy, coverage, and ethical aspects of the review analysis system.

Conclusion:

As an organized platform, the website review hub was created to address the issues brought on by the astronomical rise in online purchasing and the deluge of user-generated evaluations. The review analysis process is considerably improved by using machine learning techniques, particularly natural language processing (NLP), which is advantageous to both users and enterprises. The effectiveness of the review analysis is ensured by the use of NLP algorithms to a variety of datasets, including Flip cart reviews and Amazon reviews. These datasets are used to develop and improve algorithms for named entity identification, sentiment analysis, topic modelling, and review categorization. The algorithms are then used to analyze the dataset from review websites using the knowledge gained from studying Amazon and Flip cart reviews. Users can immediately understand the general feeling towards a product thanks to the sentiment analysis functionality, which enables the review hub to extract the underlying sentiments stated in reviews. With the help of topic modelling, firms can get a thorough picture of important details and prospective problems. By locating the companies linked to reviews, named entity identification enables businesses to monitor their brand reputation and customer satisfaction. Additionally, organizations benefit from simple access and analysis thanks to effective review classification through document labelling procedures. The resilience and scalability of the system are shown by the validation of the methodology using real-world datasets. The effective integration of NLP algorithms, trained on various datasets, into the review website enables businesses to improve their services based on insightful customer feedback and empowers consumers to make informed judgements. In conclusion, the website review hub makes use of NLP methods that have been developed using a variety of datasets to provide a complete solution for handling product reviews. It gives customers a deeper comprehension of the ideas and subjects covered in reviews, enabling them to make wise selections. Businesses can also enhance their goods and services based on insightful conclusions drawn from the review analysis. The review hub's functionality and effectiveness are improved through the use of NLP techniques, delivering a beneficial and user-friendly experience for both users and businesses.

References

1. Yan, Z., Xing, M., Zhang, D. and Ma, B., 2015. EXPRS: An extended page rank method for product feature extraction from online consumer reviews. *Information & Management*, 52(7), pp.850-858.
2. Anderson, P. and Anderson, E., 2002. The new e-commerce intermediaries. *MIT Sloan Management Review*, 43(4), p.53.
3. Patel, R.P., Nagababu, G., Kachhwaha, S.S. and Surisetty, V.A.K., 2022. A revised offshore wind resource assessment and site selection along the Indian coast using ERA5 near-hub-height wind products. *Ocean Engineering*, 254, p.111341.
4. Xie, K.L., Zhang, Z. and Zhang, Z., 2014. The business value of online consumer reviews and management response to hotel performance. *International Journal of Hospitality Management*, 43, pp.1-12.
5. Singh, J.P., Irani, S., Rana, N.P., Dwivedi, Y.K., Saumya, S. and Roy, P.K., 2017. Predicting the “helpfulness” of online consumer reviews. *Journal of Business Research*, 70, pp.346-355.
6. Kumar, N. and Benbasat, I., 2006. Research note: the influence of recommendations and consumer reviews on evaluations of websites. *Information Systems Research*, 17(4), pp.425-439.
7. Liu, X., Lee, D. and Srinivasan, K., 2019. Large-scale cross-category analysis of consumer review content on sales conversion leveraging deep learning. *Journal of Marketing Research*, 56(6), pp.918-943.
8. Liu, X., Shin, H. and Burns, A.C., 2021. Examining the impact of luxury brand's social media marketing on customer engagement: Using big data analytics and natural language processing. *Journal of Business research*, 125, pp.815-826.
9. Biswas, B., Sengupta, P., Kumar, A., Delen, D. and Gupta, S., 2022. A critical assessment of consumer reviews: A hybrid NLP-based methodology. *Decision Support Systems*, 159, p.113799.
10. Day, M.Y. and Lin, Y.D., 2017, August. Deep learning for sentiment analysis on google play consumer review. In 2017 IEEE international conference on information reuse and integration (IRI) (pp. 382-388). IEEE.
11. Nasimuzzaman, M., Merag, A.N., Afroj, S., Alam, M.M., Mehedi, M.H.K. and Rasel, A.A., 2023, March. Consumer review Analysis using NLP and Data Mining. In 2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC) (pp. 0426-0430). IEEE.
12. Loper, E., & Bird, S. (2002). NLTK: The natural language toolkit. In *Proceedings of the ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics* (pp. 63-70).
13. Jurafsky, D., & Martin, J. H. (2019). *Speech and language processing* (3rd ed.). Pearson.

14. Manning, C. D., Raghavan, P., & Schütze, H. (2008). Introduction to information retrieval. Cambridge University Press.
15. Hugging Face. (n.d.-a). Transformers: State-of-the-Art Natural Language Processing. Retrieved from <https://huggingface.co/transformers/>
16. Hugging Face. (n.d.-b). Tokenizers: Fast State-of-the-Art Tokenizers. Retrieved from <https://huggingface.co/tokenizers/>
17. Wolf, T., Sanh, V., Chaumond, J., & Delangue, C. (2020). Transformers: State-of-the-Art Natural Language Processing. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations (pp. 38-45).
18. Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends® in Information Retrieval, 2(1–2), 1–135.