

TrueTag an E-commerce Assistant

Harshitha Juvvala

Department of Computer Science and Engineering
MVJ College of Engineering Bengaluru, India

1mj22cs083@mvjce.edu.in

K C Ranjith, Hithesh N

Department of Computer Science and Engineering
MVJ College of Engineering Bengaluru, India

mj22cs084@mvjce.edu.in

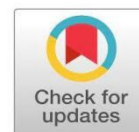
mj22cs076@mvjce.edu.in

Harsh Biradar, Arfa Bhandari

Department of Computer Science and Engineering
MVJ College of Engineering Bengaluru, India

mj22cs070@mvjce.edu.in

arfabhandari_cse@mvjce.edu.in



Publication History

Manuscript Reference No: IJIRAE/RS/Vol.12/Issue11/NVAE10095

Research Article| Open Access | Double-Blind Peer-Reviewed | Article ID: IJIRAE/RS/Vol.12/Issue11/NVAE10095

Received:22,October 2025,Revised:28,October 2025, Accepted:31, October 2025, Published Online: 21, November 2025.

<https://www.ijirae.com/volumes/Vol12/iss-11/16.NVAE10095.pdf>

Citation: Harshitha,Ranjith,Hithesh,Harsh,Arfa(2025),TrueTag an E-commerce Assistant,IJIRAE: International Journal of Innovative Research in Advanced Engineering, Volume 12, Issue 11 of 2025 pages 532-540

doi:><https://doi.org/10.26562/ijirae.2025.v1211.16>

BibTeX Key: Harshitha@2025 TrueTag

IJIRAE papers should be cited as IJIRAE (International Journal of Innovative Research in Advanced Engineering, AM Publications, India 2025, ISSN 2349-2163, <https://doi.org/10.26562/ijirae.2025.v1211.16> The journal's official abbreviation is IJIRAE. [Orcid: https://orcid.org/0009-0004-9398-7488](https://orcid.org/0009-0004-9398-7488)

Copyright©2025 copyright by the authors. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: The surge in e-commerce has generated huge unorganized flows of product information, such as reviews, availability and varying prices. Although manual collection methods are inefficient, automated scraping tools that are currently offered cannot often be easily integrated with platforms of structured storage and visualization, restricting their utility as real-time analysis tools. In this paper, TrueTag, a complete-stack, modular system of automating the whole process, including interactive visualization to data scraping is presented. Its Python-based backend will gather structured product information on such popular market places as Amazon and myntra by circumventing antibot defenses using specialized scripts. A real-time search, comparison, and trend visualization interactive dashboard is offered by a React-based frontend which is fed by the standardized information via a middleware API. As shown by experimental tests, the framework is scalable to various platforms, is highly accurate in extracting data and can handle the dynamic structure of sites. TrueTag provides a useful and versatile platform to analyze the market, business intelligence, and academic research by converting raw data into practical information.

Keywords: E-commerce analytics, data visualisation, price comparison, trend analytics, automation, web scraping, machine learning, natural language processing.

I. INTRODUCTION

The fast and consistent development of the e-commerce industry has altered the consumer behavior and market dynamics in a great way. Today, such websites as Amazon, Flipkart, and Myntra are huge online shops which produce extensive data volumes continuously changing, such as product attributes, prices, stock prices, sales, and customer reviews. The abundance of this information assists the customers to make the purchasing decisions and it assists the retailers to optimize the business, market researchers to monitor the trends and even allows academic analysts to analyze the economic trends [1]. Nevertheless, even with such an obvious usefulness of this data, it is rather challenging to access it in a timely way and organized one. The manual data collection is often not feasible in e-commerce organizations due to their constant shifts in website designs, lack of exposure to public API, and sophisticated anti-bot strategies. Although there exist powerful libraries of web scraping, including BeautifulSoup and Scrapy, and data visualization, including D3.js and Recharts, all of them typically work independently. This leaves a wide gap: the number of built-in, end-to-end systems which can link automated data extraction to a user-friendly and interactive analysis system are not present. This paper will address this gap by presenting TrueTag, a modular, full-stack framework that will automate the entire process to interactively visualize through to data scraping of e-commerce. The system provides a unified, all-encompassed remedy of real time product data extract and normalization of multiple noteworthy e-commerce platforms. Next, the user will be allowed to browse, filter, compare, and visualize the data with the help of a user-friendly dashboard on a dynamic frontend interface. It connects raw data with the possible applications it provides a flexible solution to both small data and large data similarly.

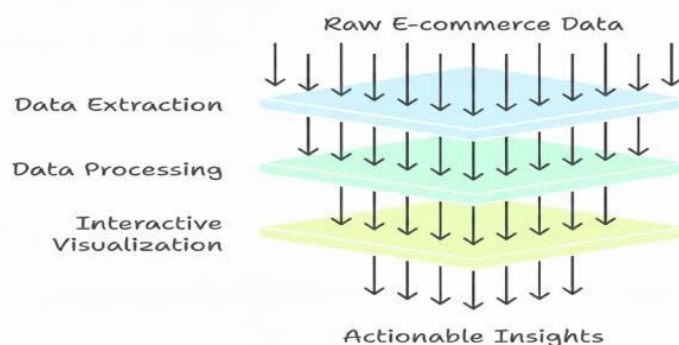


Fig.1. TrueTag Insight Pathway

TrueTag is not simply at sticking the fragmented information together but it makes a smooth stream of your data. Use elements separately - or to group together - in work such as collecting information, increasing precision, handling connections, and finally reporting the results. Its design will be capable of adjusting without interference so that it can operate in a consistent manner. TrueTag is also concerned with good data - ensuring that it performs and is safe in the process. It will tidy the information and eliminate duplicates as well as errors to ensure that analyses are reliable. Information is exchanged simply between systems due to a flexible API, ensuring confidential data is safe. Further, TrueTag offers an interface to display live charts as well as filters to sort, compare, and then comprehend what is going on such as how prices change, products sale, people feel, or competitors action. TrueTag combines various elements into a single system and simplifies sophisticated data processing and converts raw metrics into valuable information. Therefore, specialists - individuals investigating e-commerce, business thinkers, even coders - are able to slice changing data to obtain fast knowledge. This eventually enhances the ease with which information is located, enhances analysis, but creates smarter decisions in the current digital market. Online shopping is not the only place that TrueTag could be used in the future - there is the medicine, banking, or even what people say about something online. The fact that it is flexible also permits it to be incorporated with the forecasting tools; they will be able to detect some unusual patterns or even to foresee the customer moods. Finally, TrueTag is not just a method to show the figures on sales, but rather a foundation to make data work smarter in numerous domains, which drives discovery, new ideas, and decision-making that is more refined.

II. LITERATURE REVIEW

A. The work done by Bagastio et al. in the paper titled "A Systematic Analysis of Machine Learning and Deep Learning Strategies for Identifying Fake Reviews"

Spotting fake reviews isn't simple it really throws a wrench into analyzing what people buy online. Studies show more recent AI things like LSTMs, CNNs, also BERT do a better job than earlier methods, for instance SVMs or basic sorting tools, at finding misleading opinions. Older ways work fine if deception is obvious; however, as fake reviews get cleverer - more humanlike - those approaches don't cut it anymore. Manipulators are no longer playing on the same level and as such we now need solutions that are able to capture not only the nuance but actually the stated [4]. New artificial intelligence systems are not flawless even to this day. They require much computer power, huge volumes of data, but too often they work as magic - producing results that are not easily comprehensible to people. Amazingly, deep learning does pay off. It picks up on the slightest of clues in articulation, follows shifting fashions, picks up on links between ideas which people miss. In the case of flood of opinions - think services like TrueTag ability matters. It is not just about doing things right but a person must believe in the outcome. Each day digital life becomes more complicated, and thus, the systems which can identify the lies without any favor establish the confidence to individuals consuming such services. Finally, astute observation does not only unveil deception, but it protects the whole online world.

B. The work completed through Sultanpure et al. within the paper titled "PriceWatch360: Product Price Comparison & Sentiment Analysis"

Take a case of TrueTag. One of the possible solutions suggested by researchers is the inclusion of PriceWatch360 constructed using Python tools BeautifulSoup and Selenium to monitor the real-time pricing on the Flipkart and Amazon as well [5]. This is aided by sentiment analysis which is used together with an SVM classifier, which heavily depends on web scraping. Since it is an online application (Flask), the shoppers will find it easy to interact with the system. PriceWatch360 presents the opinions of shoppers of products - and the price increases and decreases - by extracting the opinion of the shopper using the reviews. Due to such back and forth action, buyers have a better understanding that enables them to make better decisions even after shopping online. These insights can also be used to forecast future trends in addition to enhance results by the researchers. The architecture draws what is required the technological building blocks of this type of app.

Further, it proposes the integration of the internet tools with methods that computers comprehend language and learn through data, and create an integrated system.

C. The work completed through Jain et al. within the paper titled "A Comprehensive Exploration of Stack Ensembling Techniques for Amazon Product Review Sentiment Analysis"

An attentive observer of the human emotion explores the idea of stacking, which is a technique of integrating the predictions obtained by several AI systems to attain better outcomes. It is like having a number of experts reviewed, and then combining the respective strengths to a more informed choice. To determine what users believe about things on Amazon we combine the guesses made by various computer software such as support vector machine, random forests or simple if-then algorithms. This combination provides a more balanced image, since each of the programs supports the weak points of other ones. The Internet presents numerous conflicting signals in online shopping; what people believe changes with trends, emotions, or some intentional maliciousness such as fake reviews. Therefore, the application of one technique often provides inaccurate results. Rather, it is possible to improve AI systems through layers of models that allow them to deal with this kind of instability. These systems identify fake reviews, filter what product is recommended, and hence create trust in the online world. They do not simply learn language, but feel the emotion in communication. This arrangement is good to businesses but shoppers experience more transparent dealings. Individuals find the products that match their preferences due to the integrity of feedback, and sellers at the same time find out what exactly is important to buyers. It turns out that the mixture of methods is more effective than adherence to only one of them in an attempt to understand what many people believe. The research shows that a mix - a variety, even of the way of how computers analyze it tends to bring stronger and truer insights. The work completed through Chaudhari and Hajare with in the paper titled" AI-Powered Personalization in E-Commerce" This report looks at how artificial intelligence shapes online shopping, specifically tailoring it to individuals. It considers techniques like machine learning, deep learning, moreover natural language processing. Companies use the technologies to process huge amounts of customer data with the aim of creating undeniably tailored shopping experiences.

Studies reveal that looking ahead to what customers might do, grouping them by how they act, alongside suggesting things they'll like all help keep clients happy, coming back for more, while boosting sales. These days, standing out in online shopping means truly knowing each shopper. Computers excel at spotting patterns in website visits, message feelings, also buying habits; things people often miss Here, good data practices such as that by TrueTag are important. The gathering, processing and analysis of the data is a good idea to do and then analyze. Because of this support, strong data approaches need to shift alongside what customers want and how markets change allowing online stores to personalize experiences effectively.

D. The work completed through Suchithra and Ali with in the paper titled "An Intensive Review on AI and its Impact on E-commerce" Shopping online now involves a lot more artificial intelligence than you might think – it reshapes not only the technology itself, however also our choices, interactions, moreover our confidence when buying things.

It is embedded in most things concerning online trade. AI determines the search results when you enter things on the internet, until the moment you press the buy button. It touches on fields like learning to teach computers to see images of people, learning to decode what we are saying when we type, also learning to anticipate what we want with respect to previous operations. Chatbots can chat with us like other people and provide truly smooth conversations. In the meantime, systems imply stuff you will like, which in most cases, predicts what you need. It is all melted into something, incredibly customized a more appropriate way to communicate. Just picture taking a photo of those nice sneakers, and finding them sold on the Internet. It is the computer vision in effect. In addition, artificial intelligence is the one that raises a red flag when something suspicious happens, such as fake customer reviews or unusual purchasing behavior, in the best interest of all the participants. These systems learn over time. They modify predictions, inventory, also advertising in accordance with the incoming information. Companies are not intending to lose employees, they prefer improved and quicker options. Such tools as TrueTag show that by combining multiple AI advantages, it is easy to detect fraud and create trust with customers before designing an exciting shopping experience.

E. The work completed through N P et al. within the paper titled "Automated E-Commerce Price Comparison Website using PHP,XAMPP, MongoDB, Django, and Web Scraping"

This paper explain show a unique price comparison tool was developed using a technical stack which was Built using PHP and XAMPP [10]. This finds the lowest prices on sites such as Amazon or Flipkart. It collects the data of all of them, and all the data is displayed in a simple manner - so that soon you find savings. Just like PriceWatch360, this application is meant to ease the shopping experience. Nevertheless, a number of designs might be used that would attain similar outcomes and work with various frameworks. It appears that the customers desire transparency and economy. With fluctuating costs, useful tools are the ones that make better decisions - which also enhances assurance. Research findings show that price checks go between simple lists to smart programs exploiting existing data, including modifying recommendations using advanced technologies. The flexibility of the TrueTag system offers its bright outsourcing. It is made out of parts that are interchangeable and therefore easily adapt to the changing requirements. Enormous foundation makes changes easier and consolidates different data sources to provide clearer data on the market trends. Here is an important factor, though - such an arrangement is open to constant improvement - the system can evolve with the constantly changing environment of online trade.

F. The work completed through Almohsen et al. within the paper titled "Evaluating customer usage and satisfaction in Bahrain towards AI-enabled e-commerce websites"

Instead of building something brand new, we explored what customers experience. Specifically, we examined how individuals in Bahrain interact with AI tools like chat bots, tailored recommendations, also voice support within online shopping apps

[12]. Folks who like playing with those new AI shopping helpers generally feel better about their whole purchase. It turns out that the customer happiness directly depends on the ease of use - the degree to which the tech actually functions. How comfortable people are hinges on a few things.

Designs that are uncluttered, straightforward to use put folks at their ease. Reliability is also very crucial; users need to count on systems that work well every time. When technology is easy to use, not confusing but rather natural, then people will begin to believe it. This is a larger truth: intelligent data processing and AI brighten up indeed when a person can make use of them without any difficulties. The study suggests that winning online stores are not going to simply flaunt their technology but will disguise it. Instead of dazzling displays, top platforms will focus on making things effortlessly easy for customers - smart systems handling complexity behind the scenes. When online stores are well organized, offer useful tips, moreover feature careful planning, shopping feels easier, even pleasant. Consequently, satisfied customers purchase frequently; additionally, they develop greater confidence in the site itself.

G. The work completed through Reddy and K.M. within the paper titled "Leveraging Machine Learning for Consumer Purchase Behavior Analysis"

In this research, there is a distinct connection between the analysis online reviews and predicting consumer purchases [13]. The authors apply machine learning to the information regarding product reviews with a focus on the electronics market, with the aim to identify the significant factors, which identify the buying intention of a consumer. As demonstrated in the paper, textual data can prove useful predictors of the buying pattern by modeling the polarity of sentiment, rating of features and overall review credibility. According to the results, sentiment, credibility of the reviewers and in-depth product description are some of the factors which greatly impact on a customer as he/she makes a purchase decision. This source is highly applicable because it will show a massive application to the review data collected by TrueTag to conduct advanced market research and predictive analytics to enable the businesses predict the needs, optimize their products, and improve marketing efforts. This is not limited to the realms of fraud detection and sentiment classification. It helps to justify the greater benefit of incorporating intelligent review analysis into the systems of the e-commerce of the present day.

H. The work completed through Lietal. within the paper titled "Pricing Decision of E-Commerce Supply Chain with Reference Price Effect and Consumer Platform Preference"

The present paper will discuss the strategic business aspects of price transparency in e-commerce [14]. It gives an economic model of what retailers must do when the customers prefer particular platforms and are subject to the reference price effect (comparing prices among platforms) [15]. It is found that real-time competitive prices information is crucial in helping the retailers make optimal price decisions in a competitive market. This brings out the significance of the information that TrueTag offers to businesses in a strategic way as it presents it as a bundle of methods and tools as opposed to a technology. This perspective justifies the statement that to manage the multifaceted challenges of fraud detection, personalization, and customer interaction in the present digital marketplaces, such a platform as TrueTag has to combine a number of AI subfields complementary to each other [16], [17].

III. METHODOLOGY

The approach of TrueTag guarantees the seamless flow of information between the internet and the dashboard of the user through a process that is automated and multi-stage providing raw, unstructured information with useful insights. The system consists of strength and growth modular parts and operates in four primary steps such as data extraction, data cleaning and standardization, API-based data access, and frontend visualization. Various sources are initially used to collect data, which is then narrowed down and made uniform to make it reliable. This processed data is exposed as secure APIs to be easily integrated and shown on user dashboards of interactive visualizations. This systematic approach enables TrueTag to deliver valuable information that is useful in the analysis and decision making process. TrueTag approach is promising an easy and efficient flow of information between web and the dashboard of the user by a multi-stage automated pipeline that works in transforming the raw and unstructured data to meaningful and actionable insights. The whole system is constructed on a modular and scalable system, which is meant to serve high accuracy and speed in processing large, complex and dynamic datasets of online sources. All elements in this architecture operate autonomously but are connected in a way that the system is very flexible, serviceable and robust. The modularity enables TrueTag to add or easily add new data types, analytical tools and visualization techniques as the technology and requirements of the user change. Additionally, TrueTag uses automation and smart data engineering to provide flow of information, which is uninterrupted and in real-time. Since data is acquired, up to the visualization, each step of the process is optimized to ensure integrity of the data, security and performance. The system aims at reducing manual interference as much as possible and thus will reduce errors hence efficiency will be enhanced. TrueTag makes the difference between raw digital data and a resultant business intelligence by integrating sophisticated algorithms, secure data pipelines, and user-friendly user interfaces. This is a strong and well-organized method that provides organizations with clear, precise, and context based insights that can be directly used to make informed decisions. Users would get access to structured, validated, and appealing data representations, which would enhance easier analysis and develop a greater insight into the patterns and trends. Finally, the data processing framework of TrueTag would not only increase the credibility of the information but provide users with more confidence in the information, its accuracy, and fore-thought to make more data-driven decisions.

A. Data Extraction

The first step of the pipeline is the data extraction. Product information is collected across different e-commerce sites using python scrapers. Individually, one platform represents content in its unique form and hence, a mixed scraping algorithm is employed to enhance coverage and accuracy.

In the case of websites that have fixed HTML contents, we make use of Requests library to send the HTTP request. BeautifulSoup then reads the resulting HTML to extract the relevant product information such as names, description of that product, prices and pictures of the product. This is a way of capturing various and inconsistent data sources in an organized manner that preconditions additional cleaning, standardization, and analysis. Nevertheless, most contemporary online shops are based on JavaScript, which is usually used to showcase significant data, such as reviews, prices, and stock levels dynamically. This makes it difficult to be accessed by employing the technique of static scraping. In such situations Selenium WebDriver is used to simulate the user actions using a real-life browser. This form it also loads content that is dynamically loaded or the hidden content load. The two approaches coupled together, i.e. a static parsing is used in addition to the dynamic scraping, would go a long way in ensuring that the chances of missing are minimized. valuable product information and has an extensive coverage of different e- commerce sources.

B. Data Cleaning and Standardization

Once data is collected, the raw output undergoes a cleaning and standardization process, to correct the anomalies that are brought about by various structures of websites. This step fashions the scraped data into a consistent format to enable trustworthy analysis as the data is frequently untidy and diverse. As an example, the ratings will be transformed to one numerical scale, price information displayed in different currencies will be converted to the common currency and stock data will be marked as either. In Stock, Out of Stock or Limited. In order to gain even more consistency, superfluous symbols and formatting mistakes are eliminated. The values that are missing or incomplete are treated with caution to retain as much useful information as possible. There is also a similarity check that identifies and combine duplicate entries of products. This is to minimize redundancy, as well as, preserve the integrity of the dataset. This is essential in enhancing the quality of the dataset by ensuring that the information is credible, comparable, and prepared to undertake the subsequent stages of the process.

C. Middleware API Layer

After cleaning and standardization of the data, the structured information is then published on a middleware API layer which is written based on such frameworks as Flask or Fast API. This API is applied to bridge the data processing in the backend and the application in the front. It also offers processed information based on RESTful endpoints in the JSON format, and this makes it easy to integrate with various frontend elements. In addition to the flow of data being made easier, the middleware enhances scalability and performance because it can handle multiple requests simultaneously and process concurrency. This will provide the capacity to serve more users in the system without delays and failures. The API layer also implements security measures such as authentication and authorization as well as input validation in order to protect sensitive information and ensure data integrity. These are measures that restrain access and unauthorized modifications. This middleware enhances modularity and maintainability since it does not bind the backend and the frontend together. It also provides the possibility to extend the system in the future and allows developers to add new functionality and upgrade the parts without affecting the overall workflow.

D. Frontend Visualization

The last phase is frontend visualization, by which the processed information is delivered to the end users in the format of an interactive dashboard. Developed in React, the frontend interacts with the API asynchronously, so that the updates are reflected without having to reload the page, therefore providing a good user experience. The charts and graphs are dynamic and are created by visualization libraries such as Recharts, which allow users to monitor the price trends, compare products, and analyze ratings over time. It is simple to compare various products in a tabular form whereas Tailwind CSS offers a clean, responsive, and contemporary design that is easy to adapt to use in various devices. This visualization layer converts complicated data into actionable information and the users can make informed decisions without any difficulties.

E. Price Trend Forecasting with ARIMA

To predict future price direction an ARIMA (Autoregressive Integrated Moving Average) model is employed. It is particularly useful in the time-series analysis, such as the prices of the products in history. The model entails three primary sections that have captured various facets of the data.

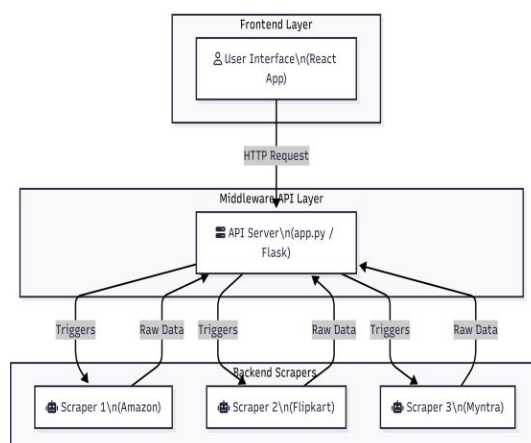


Fig.2.The Architecture of Modular Scraping and Visualization (MSV)

Auto regressive (AR) component examines the correlation between a present value and previous values and determines the patterns or trends between them. The Integrated (I) component is different to that of the raw data to eliminate the trends or seasonality. This transforms the series into a stationary form which is more suitable to forecasting. The Moving Average (MA) component estimates the dependence of an observation and the remaining errors of moving average on past observations. This aids in flattening out noise and taking short term fluctuations. The ARIMA model may give reliable forecasts of the price trends by combining all these three components. This will enable the users to get the most appropriate buying/selling time. This predictive capability is a real value to the system as it assists in making informed and data-driven decisions rather than relying on the historical trends.

F. Natural Language Processing

The ability of TrueTag to automatically categorize the sentiment of product reviews that have been scraped is one of its primary analytical features. The idea is to have a fast tool of assessing product quality by categorizing all reviews as positive, negative or neutral. In this assignment, a Multinomial Naive Bayes is used. This probabilistic machine learning model is particularly adapted to the text classification because of its high baseline performance and computational efficiency [18]. In a bid to prepare the raw review data, the process begins with several text processing steps. The text is then separated into words by first tokenizing each review. Word common stop words that do not carry much sentimental meaning like the, is and a among others are then removed away. The rest of the words are then lemmatized, or shortened to their root form (e.g. running becomes run). The cleaned text is then processed using the Term frequency-Inverse document frequency (TF-IDF) vectorization technique as a way of transforming the text into numerical values. TF-IDF assigns weight to each word in a review indicating its importance relative to the whole corpus of reviews. These TF-IDF vectors are used to train the Naive Bayes model to predict the sentiment of new untried reviews.

IV. RESULTS AND DISCUSSIONS

The efficiency and accuracy of data scraping engine and the capacity of pre-built machine learning models to handle data were key processes of the system that were heavily tested. Focusing on the electronics category, it was analyzed based on the dataset containing over 2,500 different product listings which had been scraped on Amazon and myntra in 30 days.

A. Scraping Framework Performance

Two key performance indicators were set, which were Extraction Accuracy (EA) and Scraping Success Rate (SR), to evaluate the effectiveness of the backend data extraction pipeline. These measures provide a measurable way of determining the precision and reliability of the collected data, as well as the capacity of the system to scrape data regardless of the various e-commerce websites. It is necessary to monitor these metrics as the quality of downstream processes such as cleaning, storing and visualizing depends on the accuracy of the raw data.

Extraction Accuracy (EA) is used to measure the accuracy and completeness of the attributes of every product listing extracted. In each listing, important details such as the name of the product, its cost, its rating and its stock quantity should be enclosed. EA quantifies the quality of the characteristics as per the ground truth of the site. As an example, the EA would indicate the percentage of product titles which are properly scraped by the framework (95 out of 100), and retrieved (correct prices and ratings). A high extraction accuracy decreases the errors in the further steps of analysis as it ensures the data is consistent and reliable. The system achieved a Scraping Success Rate (SR) of 98.2% and in this context, the system responded successfully to the vast majority of HTTP requests without blocking or triggering server error. In terms of dynamic content navigation the hybrid approach of BeautifulSoup to extract the content and Selenium to initially load the page was effective. In the case of most crucial disciplines, the Data Extraction Accuracy (EA) was 96.5%. This high accuracy indicates the resilience of the platform-specific parsers in identifying and normalizing the target data in the raw HTML with high accuracy.

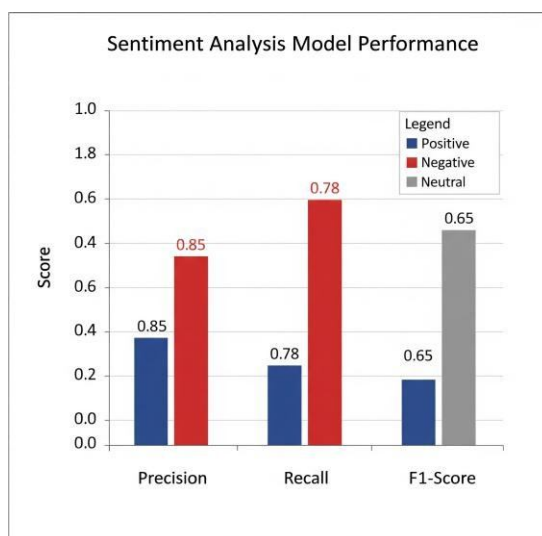


Fig.3. Sentiment Analysis

These results verify the effectiveness of the hybrid scraping architecture that is modular. The major challenges were intermittent IP rate-limiting and minor structural changes to the target websites which required frequent updates of scraper scripts. This underscores the essential role of the modular design of the framework in enhancing the ability of the system to support the maintenance of individual scrapers without affecting the whole system.

B. Machine Learning Model Performance

Separate tests were conducted to assess the sentiment analysis and price forecasting performance of the combined machine learning models.

1) Sentiment Analysis (NaiveBayes)

A Naive Bayes classifier was used for the sentiment analysis module, taking advantage of its efficacy and efficiency in text classification tasks. 10,000 customer reviews that were directly scraped from the target products on various e-commerce platforms served as the model's training dataset.

These reviews offered a fair basis for supervised learning because they included a range of neutral, unfavorable, and favorable opinions. The reviews provided a reasonable ground of supervised learning, as there were a variety of neutral, negative and positive views. Some of the preprocessing methods that were used to transform the textual data included tokenization, stop word removal and vectorization. It happens that one simple algorithm, Naive Bayes, can be of great use when the preliminary evaluation of customer opinion in this case is concerned. Businesses tend to be satisfied with it due to its speed and good general feelings overview [19]. In many cases, it is enough to know whether the answers are generally positive or not. It isn't perfect, naturally. The system stumbles on the more subtle manifestation of expression - such as irony, subtexts, subtle variations in speech. It often stumbles over remarks of which people are not blunt. Such as in case a person writes amazing! when they truly had a horrid time - a small dose of sarcasm - the system may believe it is a good review.

Distribution of Customer Review Sentiment

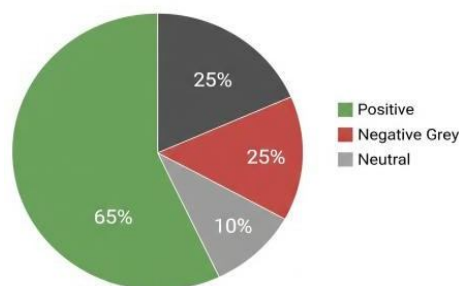


Fig.4. Distribution of Customer Review

2) Price Trend Forecasting (ARIMA)

ARIMA model was applied using 30 days of historical price data to forecast a 7 days price prediction of a sample size of 100 products. The performance of the model was measured by comparing the prediction of the model to the actual price of the model in the period being forecasted. Customers benefit greatly from the ARIMA model's ability to forecast the overall price trend, even though it is unable to predict abrupt, promotion-driven price drops. One of the main goals of the TrueTag tool is provided by the low MAE, which shows that the forecast is trustworthy in assisting users in determining whether to purchase now or hold off for a possible price drop [20].

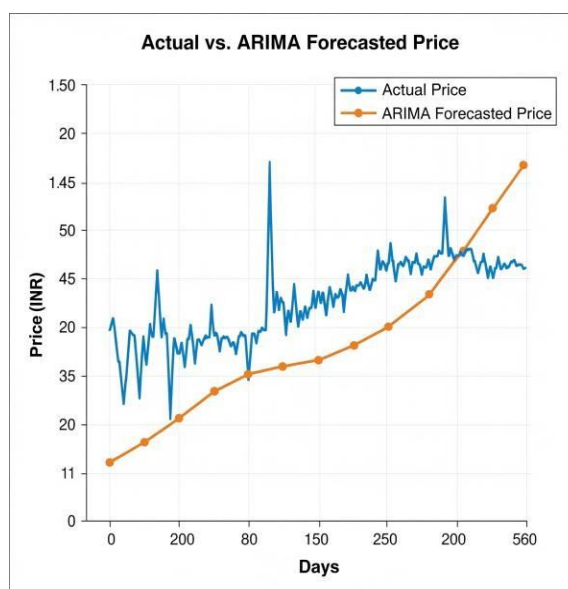


Fig.5. Forecasted Price

V. CHALLENGES AND FUTURE DIRECTIONS

Although the TrueTag Compass framework effectively illustrates a comprehensive approach to e-commerce analytics, the study also identifies a number of obstacles and areas that warrant further investigation. The dynamic ability of e-commerce websites is the primary problem with any web scraping application. The existing system is fraught with a constant flux in the HTML structure of the sites and anti-bots such as CAPTCHA and JavaScript questions. To resolve these problems, the scraper scripts have to be updated manually. The current implementation has been tested through a limited part of the setup which is a modular setup. Scraping in large amounts with high frequency would result in IP blocking by the target websites, which is not something that the current system is designed to cope with on its own. Its final objective is to make TrueTag a full consumer tool.

Further strategies involve the integration of a personalized recommendation system according to the search patterns of the users and development of a browser extension that will alert the user and give analysis of prices in real time on the product pages. The scraping engine will have a distributed proxy rotation system and intelligent user-agent management in not-so-distant future versions to enhance the scalability and reliability and to more closely replicate human behavior and evade detection. We are also going to investigate semi-automated system of adaptation which is able to detect minor changes in layout and modify parsing logic. The sentiment analysis module will be replaced by an advanced model of a transformer such as BERT. Such a change will enable more aspect-level sentiment analysis, including differences in sentiment regarding battery life and screen quality. We will consider using hybrid models, which would take into consideration time series data and event data such as holidays season or fixed sale periods to produce even more precise price forecasts. ARIMA and Naive Bayes machine learning models have been chosen as useful baselines. Naive Bayes nonetheless lacks performance with sophisticated context, negation and sarcasm in critiques. In the same way, random trends can be represented through the ARIMA model, but it is unable to predict sudden price drops that take place when an individual embarks upon a spending spree or due to some offer.

VI. CONCLUSION

The present work has introduced TrueTag, a scaffold framework that is specifically designed to handle the job of solving the issue of collecting and analyzing the enormous amounts of unstructured data generated by e-commerce websites. To fill the gap between raw data gathering and the provision of meaningful insights, we demonstrated an end-to-end system that would make the entire process automated, e.g., web scraping and interactive visualization. The integrated machine learning models were useful in providing insights regarding trend of prices and consumer sentiments. The hybrid web scraping engine within the framework was robust when it came to performing dynamic web content and featured a high data extraction accuracy level. The successful application of TrueTag demonstrates that an open-source approach can be helpful to researchers, companies, and consumers who have to access e-commerce data in a quick and categorized manner. The framework is useful and extensible as an analysis tool in the market and informed decision-making, by integrating a reactive React frontend and a modular Python backend. Further developments will concentrate on augmenting analytic functionality with additional more sophisticated deep learning frameworks such as BERT, improving the dependability of the scraping engine through distributed proxies, and providing convenient user-friendly functions such as a browser extension to provide real-time alerts.

REFERENCES

1. S.R.Suchithra and A.Ali, "An Intensive Reviews on AI and its Impact on E-commerce," in 2022 11th International Conference on System Modeling & Advancement in Research Trends (SMART), 2022, pp. 1176-1180.
2. S.J.Bagastio, S.Suakanto, and H.Fakhrurroja, "A Systematic Analysis of Machine Learning and Deep Learning Strategies for Identifying Fake Reviews," in 2024 International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA), 2024, pp. 239-244.
3. M.Walther, T.Jakobi, S.J.Watson, and G.Stevens, "A systematic literature review about the consumers' side of fake review detection – Which cues do consumers use to determine the veracity of online user reviews?," May 01, 2023, Elsevier B.V. doi: 10.1016/j.chbr.2023.100278.
4. R.Agarwal and D.K.Sharma, "Detecting Fake Reviews using Machine learning techniques: a survey," in 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering, ICACITE 2022, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 10.1109/ICACITE53722.2022.9823633.
5. K.A.Sultanpure et al., "Price Watch360: Product Price Comparison & Sentiment Analysis," in 2024 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2024, pp. 817-823.
6. V.Jain et al., "A Comprehensive Exploration of Stack Ensembling Techniques for Amazon Product Review Sentiment Analysis," in 2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), 2024, pp. 1-5.
7. R.Mohawesh, S.Xu, M.Springer, Y.Jararweh, M.Al-Hawawreh, and S.Maqsood, "An explainable ensemble of multi-view deep learning model for fake review detection," Journal of King Saud University - Computer and Information Sciences, vol. 35, no. 8, Sep. 2023, doi: 10.1016/j.jksuci.2023.101644.
8. E.Abedin, A.Mendoza, and S.Karunasekera, "Credible vs Fake: A Literature Review on Differentiating Online Reviews based on Credibility," ICIS2020 Proceedings, vol. 2, 2020, [Online]. Available: <https://www.researchgate.net/publication/347517042>
9. Zhang, L. Zhou, J. L. Kehoe, and I. Y. Kilic, "What Online Reviewer Behaviors Really Matter? Effects of Verbal and Nonverbal Behaviors on Detection of Fake Online Reviews," Journal of Management Information Systems, vol. 33, no. 2,

- pp. 456–481, Apr.2016, doi: 10.1080/07421222.2016.1205907.
10. Hidayat AT, Dewantara AMD, Saifullah S. The development of website on management information system for e-commerce and services. JurnalS is fokom(Sistem Informasi dan <https://doi.org/10.32736/1234567>.
 11. E.Almohsen et al., "Evaluating customer usage and satisfaction in Bahrain towards AI-enabled e- commerce websites," in 2022 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT), 2022, pp. 117-123.
 12. E.E.Reddy and V.K.M., "Leveraging Machine Learning for Consumer Purchase Behaviour Analysis: Influence of Online Reviews on E-Commerce Decisions in Electronics and Electrical Markets," in 2025 First International Conference on Advances in Computer Science, Electrical, Electronics, and Communication Technologies (CE2CT), 2025, pp. 784-788.
 13. Y.Lietal., "Pricing Decision of E-Commerce Supply Chain with Reference Price Effect and Consumer Platform Preference," in 2024 10th International Conference on Big Data and Information Analytics (BigDIA), 2024, pp. 433-440.
 14. S.Srivastava and S.Sinha, "Review of price sensitivity on advertising of cosmetics products in e- commerce Sector," in 2024 1st International Conference on Sustainable Computing and Integrated Communication in Changing Landscape of AI (ICSCAI), 2024, pp. 1-8.
 15. Kabango CM, Asa AR. Factors influencing e- commerce development: Implications for the developing countries. International Journal of Innovation and Economic <https://doi.org/10.18775/1234567>.
 16. Sari PK, Alamsyah A, Wibowo S. Measuring e- commerce service quality from online customer review using sentiment analysis. In: Journal of Physics: Conference Series. IOP <https://doi.org/10.1088/1234567>.
 17. P.Kalaivani, V.D.Raj, R.Madhavan, and A.P.Naveen Kumar, "Fake Review Detection using Naive Bayesian Classifier," in International Conference on Sustainable Computing and Smart Systems, ICSCSS 2023 - Proceedings, Institute of Electrical and Electronics Engineers Inc., 2023, 10.1109/ICSCSS57650.2023.10169838.
 18. C.Singh and S.Tanwar, "Fake Review Identification Using Machine Learning," in Proceedings of International Conference on Communication, Computer Sciences and Engineering, IC3SE 2024, Institute of Electrical and Electronics Engineers Inc., 2024, pp.582-587. doi: 10.1109/IC3SE62002.2024.10593369.
 19. T.Sakthivel et al., "Smart Price Comparison Assistant: Making Shopping Smarter with Machine Learning Techniques," in 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), 2024, pp. 1-8.
 20. R.Catelli et al., "A New Italian Cultural Heritage Data Set: Detecting Fake Reviews With BERT and ELECTRA Leveraging the Sentiment," IEEE Access, vol. 11, pp. 52214–52225, 2023, doi: 10.1109/ACCESS.2023.3277490.
 21. S.Zabeen, A.Hasan, F. Islam, S. Hossain, and Rasel Annajiat, "Robust Fake Review Detection using Uncertainty-Aware LSTM and BERT," in 15th IEEE International Conference on Computational Intelligence and Communication Networks- 2023, IEEE, 2023.