
Amazon Marketplace: An Analysis of External Factors and Machine Learning Models - Survey

Muneer Hazaa Alsurori, Waheeb Abdo Almorhebi*

Computer Science and Information Technology, Science Faculty, Ibb University, Ibb, Yemen

Email address:

msurory@yahoo.com (Muneer Hazaa Alsurori), waheeba940@gmail.com (Waheeb Abdo Almorhebi)

*Corresponding author

To cite this article:

Muneer Hazaa Alsurori, Waheeb Abdo Almorhebi. (2023). Amazon Marketplace: An Analysis of External Factors and Machine Learning Models - Survey. *American Journal of Artificial Intelligence*, 7(2), 52-59. <https://doi.org/10.11648/j.ajai.20230702.13>

Received: September 21, 2023; **Accepted:** October 20, 2023; **Published:** November 29, 2023

Abstract: This research endeavors to utilize diverse machine learning algorithms to forecast product prices on the Amazon marketplace. The primary objective of the study is to examine the impact of external factors, such as Google Trends and customer reviews, on future product prices and demand. The research process involves gathering unstructured product information and pricing data from Amazon using APIs and crawlers, followed by preprocessing the data through techniques like tokenization and stopword removal. Machine learning algorithms, including decision trees, support vector regression, and random forests, are employed to predict product prices. The study also explores the challenges associated with web scraping and explores potential applications of web harvesting in e-commerce enterprises. To ensure a comprehensive analysis, the research draws upon relevant literature in the field, encompassing the use of machine learning models for stock price forecasting, time series forecasting, and sentiment analysis. By building upon and leveraging existing methodologies, the study aims to contribute to the understanding of price dynamics within the Amazon marketplace. The significance of this research lies in the growing reliance on e-commerce platforms like Amazon for product purchasing. By investigating the relationship between product prices and various influencing variables, this study can provide valuable insights to both sellers and consumers in the ever-evolving online market. Ultimately, the research seeks to predict product prices on the Amazon marketplace using machine learning algorithms and shed light on the dynamics of e-commerce, benefiting sellers and consumers alike.

Keywords: Machine Learning, Predicting Prices, Amazon Data, Google Trends

1. Introduction

The Amazon Marketplace Has Become a Popular Platform for Online Retail, And Predicting Product Prices It Has Become an Important Area of Research for Many Scholars. This Research Focuses on Understanding How Historical Information and External Factors Such as Google Trends or Customer Reviews May Affect the Future Price Direction of a Particular Product On the Amazon Marketplace. The Study Aims to Extract Unstructured Product Information and Pricing Data Using Amazon APIs and Crawlers, Pre-Process The Data Using Techniques Such as Tokenization, Stop Word Removal, And Feature Extraction, And Use Decision Trees, Support Vector Regression, And Random Forest machine Learning Calculations Such as Choice Rest to Predict Product Prices. The Research Also Explores the Challenges of Web Scraping

and The Potential Applications of Web Harvesting in E-Commerce Businesses. The Study the Utilize of the Machine Learning Models for Stock Price Forecast, Draws On Related Work in The Field, Which Includes Time Series Forecasting, And Sentiment Analysis. Overall, the research aims to provide insights into predicting product prices on the Amazon Marketplace and its potential applications in the e-commerce industry. The problem addressed in this study is how to predict the direction of product prices on the online selling platforms "Amazon Marketplace." This is done by utilizing historical information and external factors such as Google trends or customer reviews. The study aims to extract unstructured product information and pricing data using application programming interfaces (APIs) and web scraping tools. The data is then processed using techniques like tokenization, stop-word removal, and feature extraction. Machine learning models such as decision trees and support

vector machines are used to predict product prices. The study also explores the challenges of web data extraction and data mining applications in e-commerce. It is built upon related works in the field, such as time series forecasting and sentiment analysis, to leverage machine learning models in predicting product costs. Overall, the study aims to provide insights into predicting product prices on the Amazon Marketplace platform and its potential applications in the e-commerce industry.

2. Literature Review

In this section, review relevant previous works collected in this field:

Websites are a sea of boundless data which anybody and everybody can get. The modern drift of innovation puts us to alter the way we our commerce ways. The Web presently is the modern put for commerce. Knowing how to utilize the Web and the distinctive openings that can bring us can be one of the keys to victory in e-commerce and e-marketing [11]. Online purchasing is gradually displacing traditional shopping methods in every manner. [9] Prediction is used in many types of products. The authors present an idea in Construction of a time analysis model in a hybrid approach for forecasting vegetable prices Agricultural e commerce platform that can be used by the government to derive Policies [3]. The high volatility underscores the significance of assessing the impact of external factors on the stock market, the stock market exhibits high volatility due to its reliance on political, financial, environmental, and various internal and external factors, in addition to historical stock data, the objective of this study is to create an intelligent system capable of predicting future trends in the stock market by analyzing the sentiments expressed in financial news articles [17]. Machine learning, a well-established algorithm in a wide range of applications, has been extensively studied for its potentials in prediction of financial markets [6]. Machine Learning empowers IT organizations to identify the patterns on the basis of currently available algorithms and data frames to cultivate acceptable solution concepts. [15]

The joint utilize of time arrangement, notoriety and estimation examination clearly speaks to one imperative approach to this investigations issues. In this paper we show Cost Test, a suite of computer program devices created to perform determining products' costs. Its essential point is to anticipate long-term cost slant of items producing a customized estimate through the misuse of autoregressive coordinates moving normal (ARIMA) demonstration. We experimented the adequacy of the proposed approach on one of the greatest E-commerce framework within the world of Amazon [5].

Content information preprocessing may be a key step in suspicion examination and high-quality data sources can be gotten after that. Most objects in characteristics of dialect, taking care of are sentences, so sentences ought to be divided into single words. in addition, the accentuations are removed since they have no reasonable meaning in sentences. In English substance, spaces between words are utilized as

separators, but for Chinese, because it was sentences and sections have self-evident separators, and there's no space between words. We utilize Jieba word division calculation to section unstructured information within the money related field. Jieba could be a capable word division library executed in Python dialect [1].

Houda Bakir, Ghassen Chniti, and Hédi Zaher, In this paper, them give a vigorous estimating show to anticipate phone cost in European markets utilizing Long Short-Term Memory (LSTM) neural arrange and Support Vector Relapse (SVR). We propose a comparison considering of time arrangement forecasting models for these two strategies LSTM, due to its engineering, is considered as the idealize arrangement to issue not resolvable by classic Repetitive Neural systems (RNNs), Support Vector Machines (SVMs) are a really capable machine learning strategy for both classification and relapse. After considering and comparing a few univariate models, SVR and LSTM neural networks show up to be the foremost exact ones. In expansion, we compared multivariate models for both these strategies Considering the multivariate approach, by displaying more variables, we get predominant estimate execution [8].

Artificial Neural Network (ANN), which is one brilliantly information mining procedure has been utilized for Stock Cost Expectation. It has been trusted as the foremost exact thought. This paper overview diverse machine learning models for stock cost forecast. We have prepared the accessible stock information of American Carriers for this venture. The programming dialect that we have utilizing in this paper is Python. The Machine Learning (ML) models utilized in this venture are Decision Tree (DT), Support Vector Relapse (SVR), Random Forest (RF), and Artificial Neural Network (ANN) [4].

Online obtaining is steadily uprooting conventional shopping strategies in each way from shoes to nourishment, everything is presently accessible on E-cart. Individuals like to buy online since there is so numerous conceivable outcome in each zone Client can take not the cost contrast between websites and, as a result, the larger part individual will select the benefit with the most reduced cost Our commitments incorporate three perspectives:

1. An S_I_LSTM system is outlined by joining different information sources and investors' opinion
2. Assumption investigation strategy based on CNN is proposed to calculate the investors' estimation list
3. LSTM arrange with consideration instrument is recommended stock price predication [1].

3. Methodology

The followed methodology includes the following step and as shown in the figure 1.

3.1. Reviewing Relevant Research in the Field

This step requires reviewing literature and previous research related to product price prediction and the use of machine learning algorithms in this context The aim of this step is to understand previous work, and extract useful ideas and methodologies for the current study.

3.2. Collecting Product Information and Pricing Data from the Amazon Marketplace

This involves gathering data on available products on the Amazon platform, with a focus on price information and available features. This can be done using application programming interfaces (APIs) and data scraping tools.

3.3. Data Preprocessing

This includes cleaning the data and removing unnecessary or duplicated information. It is also necessary to identify relevant features related to product price prediction, which will be used in applying machine learning algorithms.

3.4. Selecting Appropriate Machine Learning Models

This step relies on identifying and choosing suitable models for predicting product prices, such as decision trees, vector regression, and random forests. The selection of models is based on data characteristics and study requirements.

3.5. Training and Testing the Models and Verifying Their Accuracy

This involves splitting the data into training and testing sets, Training the models using the preprocessed data, and testing them on the test set. The accuracy of the models should be evaluated and verified according to appropriate performance metrics, such as accuracy, standard deviation, and error rate.

3.6. Interpreting the Results and Providing Insights on Product Price Prediction on the Amazon Platform

After training and testing the models, the results need to be interpreted and analyzed.

Insights and conclusions can be provided regarding the models' ability to predict product prices and their application in the context of the Amazon platform.

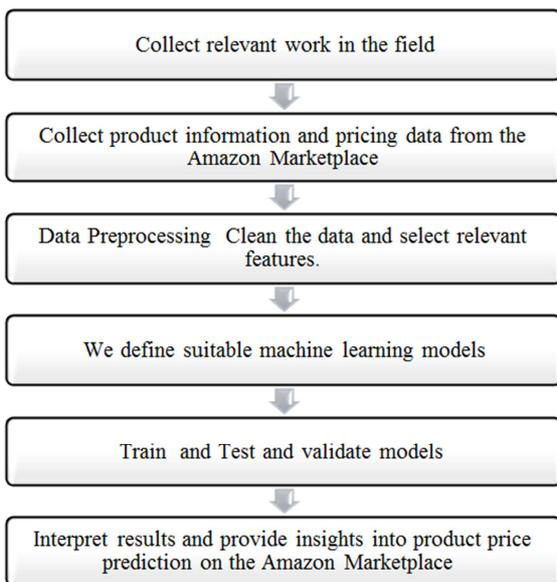


Figure 1. Methodology.

4. Prediction Techniques

4.1. Decision Tree

Classification trees are choice trees that are based on replying questions and utilizing this data to create a choice. A classification tree is one that decides whether a choice is fit or unfit by inquiring an arrangement a set of arranged questions and utilizing the answers to reach at a reasonable arrangement. The classification and relapse tree (cart) calculation produces twofold choice trees with precisely two branches for each choice hub. Cart recursively segments the preparing information set into subsets of records with comparative target quality values. The cart calculation builds the trees by performing a comprehensive look on each hub for all available factors and all conceivable part values, and after that selecting the ideal part based on a few productive part criteria [7].

Decision trees offer a valuable set of models and tools that provide high accuracy and predictive performance, enabling farmers to make informed decisions. Farmers can greatly benefit from these models by improving resource allocation, developing efficient irrigation schedules, and designing agricultural practices tailored to the specific requirements of different crops, ultimately enhancing crop productivity [16].

4.2. Support Vector Regression

It may be an Administered Machine learning calculation utilized for relapse investigation. It finds the work that makes a difference on an us inexact mapping based on the preparing test from an input space to genuine numbers [4].

4.3. Random Forest

Random Forest Is a Machine Learning Algorithm Used for Regression Analysis. With This Decision Tree It Was Overcome the Problem of the Obvious Oversimplification of the Ensemble Learning Method. The Prediction Steps Are First a Random K Data Point Selected from The Training Set and A Decision Tree Is Built Accordingly. Then Select the Tree We Want to Build and Follow the Previous Steps Again. From Each New Data Point, Predict N-Trees for The Y Value of the Data Points and Assign the New Data Points to All Y Predicted Y Values. [4].

4.4. Artificial Neural Network

A manufactured Neural organize is an interconnection of hubs that's just like the natural neuron in our body but not comparable. For the final few decades, ANN has been utilized for Stock Cost Forecast It contains three layers, to begin with is the Input Layer – this layer takes diverse inputs variable from the client at that point, the covered up layer-This layer is a display between 166 the input layer which distinguishes all covered up highlights and designs and the final layer is the Yield layer- This layer gives the ultimate yield ANN takes diverse inputs and duplicates them with the desired weights for each with an actuation work for the actuation of neurons [4].

Machine learning models, particularly Artificial Neural

Networks (ANN), have demonstrated the ability to learn input-output relationships and provide accurate forecasts of daily closing prices when trained on the same data [17].

4.5. Convolutional Neural Network

Convolutional neural network The CNN show was proposed by and was broadly utilized in picture acknowledgment structure of CNN primarily incorporates input layer, convolution layer, pooling layer, completely associated layer and yield layer. The CNN Basically Chooses the Classified Highlight Set Through the Convolution Bit Within the Convolution Layer. Firstly, Connected CNN into Content Classification and Accomplished Great Comes About. Utilized Three Diverse Profound Learning Structures, Specifically LSTM, RNN and CNN for The Cost Forecast and Compared Their Exhibitions, And The Tests Appeared CNN-Sliding Window Show Perform Way Better and Had Less Rate Blunder [1].

4.6. K-Nearest Neighbor

The K- Closest Neighbor Is an Illustration of Instance-Based Learning Utilized for Classification in Which the Preparing Information Is Put Away, So That a Classification for A Unused Unclassified Record May Be Found Essentially by Comparing It to The Foremost Comparative Records Within the Preparing Set. The Esteem of K Decides How Numerous Closest Comparable Records Within the Preparing Information Set Is Considered for Classifying a Test Information Set Record [8].

4.7. Long-Short Term Memory

Long-Short Term Memory could be a profound learning manufactured repetitive neural organized engineering. Not at all like conventional feedforward neural systems, LSTM incorporates critical associations, it is able to prepare not as it was single information focuses, but moreover whole information groupings. To estimate the next time step, the weight values within the organize must be upgraded, which requires the conservation of the introductory time step information. An RNN can as it was learning a constrained number of short-term affiliations; be that as it may, a long-term time arrangement, such as 1000-time steps, cannot be learned by RNNs; in any case, LSTMs can learn these long-term affiliations legitimately. On the off chance that the length of the input vector is d and the number of neurons within the covered up layer is h , at that point the memory space required for an LSTM cell is $O(d \cdot h)$. The sum of memory required in LSTM is $O(d \cdot h)$ since the result of the following cell ($t+1$) is supplanted within the same memory for the ancient values. Since computational models have a boundlessly huge memory, space complexity isn't ordinarily considered an essential issue; hence, time complexity is of the most extreme significance [10].

A Long Short-Term Memory (LSTM) model based on Recurrent Neural Network (RNN) architecture was employed to capture long-term patterns and learn the stock market trend.

This model exhibited superior performance compared to previous models and successfully predicted the opening price [17].

5. Web Harvesting

There are numerous strategies utilized for scratching information on web (1) Conventional duplicate and glue (2) Content grapping and standard expression (3)- Hypertext Exchange Convention (HTTP) Programming (4) Hyper Content Markup Dialect (HTML) Parsing (5) Record Protest Show (DOM) Parsing (6) Web Scratching Computer program (7) Vertical conglomeration stages (8) Semantic explanation recognizing (9) Computer vision web-page analyzers [11].

5.1. Practices [11]

1. Price-monitoring and Product Research
2. Online price comparison
3. Better Customer analysis
4. Market Analysis

5.2. Potential Challenge of Web Scraping

One of the challenges we discover whereas scratching data from different websites is that these websites have different compositions. Meaning the formats of websites are diverse and special and therefore extreme to circular [11].

6. Preprocessing Data for Amazon Product Price Prediction

6.1. Collection Products Data

Today, Amazon has become the number one online retailer for numerous shoppers. There's a part of inquiries about It Amazon administrations and cost expectations on Amazon items to is one of the bearings that numerous analysts are heading. The costs of items sold by Amazon are influenced by numerous variables, most of which are obscure. It can be influenced by swelling, and deals volume, its ubiquity, the ubiquity of its producer, by the ubiquity of its lesson. A few data approximately the items are obscure, such as the number of their deals, since Amazon does not uncover this sort of information for vital reasons. A few of Amazon's information related to its items may be collected utilizing Amazon APIs, or it may be mined utilizing crawlers, scrubbers, and other advances. [5] The point of our work is to consider and analyze how verifiable data and outside factors may influence long haul cost heading of a specific item and the effect of each of the outside variables based on the cost desires of each item. We utilized Amazon as a stage to extricate unstructured item data and estimating with the objective of executing Amazon item cost expectations on the extricated information utilizing Amazon APIs and crawlers to get it whether outside data, such as that provided by Google Patterns or client audits (the last mentioned is as of now confirmed to be gotten) Impacts on future item prices and demand, which may bring benefits to

aggregate forecasting [5].

6.2. Pre-processing the Data

Pre-processing connected to the information is Avery imperative figure in deciding the victory of declining this application of neural systems, the essential reason of information pre-processing is to adjust the input factors so they can coordinate superbly the anticipated yield [12]. The primary purpose of data pre-processing is to modify the input variables so they can better match the predicted output [13].

1. Data Cleanup Data cleaning will be done by expelling any clear spaces from the information, at the pointe changing all the content to lowercase to dodge disarray and expelling halt words from the information. Halt words are those words which don't play a vital part in sentence arrangement, such as "are", "we", "is", etc. Cleaned information is put in a separate dataset containing 10,000 records with two main categories "question" and "description" [14].
2. Tokenization: In this step, each segment inside the corpus i e., each segment inside the file will be broken down into a set of words. To begin the tokenization to get ready, we look for concepts or words that make up a character gathering. This will be basic since we are going be able to induce meaning from the primary substance gathering utilizing these terms. Tokenization is the strategy of confining tremendous chunks of substance into small r pieces known as tokens. Usually frequently satisfied by deleting or isolating characters like complement and. Tokens are expressions that are isolated into individual words after being tokenized out of sections. We may get information such as the number of words in a substance, the repetition of a specific term inside the substance, and much more by doing Tokenization. Strategies, such using the Natural Language Toolkit [NLTK], he extensive library etc. Tokenization may be a mandatory step in ongoing content preparation, such as the completion of word exclusion, derivation, and lemmatization [14].
3. Stemming and Lemmatization: These run the appeared based calculations that examine certain words underneath a collection of scenarios and then select how to shorten it based on a list of recognized increases. It is worth taking note that the root made after stemming may not be the same as the word's morphological root. Stemming is slanted to underneath and over-stemming. Doorman-Stemmer, Snowball stemmer, and Lancaster

stemmer are a numerous common stemming calculations. Lemmatization, on the other hand, is the strategy of accurately reducing words to root words utilizing a lingo dictionary. Lemmatization, as contradicted to Stemming, which fair chops out tokens by essential design coordinating, may be a more advanced procedure that utilizes dialect lexicon and Morphological thinking about words to supply phonetically appropriate lemmas. This induces that lemmatization makes utilize of setting information and may subsequently differentiate between words with distinctive suggestion based on parts of talk. For the English tongue, our system businesses the NLTK python package's WordNet Lemmatizes (based on the Word present day Database). The concept of stemming and lemmatization can be caught on by a coordinate case. The word "loving" will turn into "lov" after stemming which has no meaning whereas in case lemmatized, "loving" will turn into "love" which straightforwardly consolidates a true meaning and utilize case [14].

7. Amazon Market Price Prediction

Supervised machine learning employments calculations for cases to deliver common theories, and after that makes expectations almost future cases. and utilizing straight relapse to anticipate the design or cost of an item. To precisely decide cost or design estimation, a number of regularization strategies must be connected in straight relapse [2]. We will collect data about the specifications, features, and pricing of each electronic product in the Amazon market and store it in databases, and then carry out a pre-processing process for this data before using one of the appropriate forecasting techniques that were mentioned earlier and training the models on it, which will analyze the name and price Characteristics, features, and specifications of each product using Natural Language Processing (NLP) technology These steps are done in order to have adjusted machine learning methods to discover the forecast. In this case, show preparing is completely vital. Arbitrary timberland, choice tree, and back vector relapse models were utilized to do the expectation work [4], The plausibility to alter the cost by determining competitors' items costs is comparable to anticipating the monetary stock costs after you have satisfactory competitor's time to arrange information, which essentially is the collection of the whole shown information [8].

8. Comparison Table

Table 1. Comparison Table.

	Title of Paper	Place of publication	Title of Paper
1	S_I_LSTM: stock price prediction based on multiple data sources and sentiment analysis	Taylor & Francis Group CONNECTION SCIENCE 14 Jun 2021	S_I_LSTM: stock price prediction based on multiple data sources and sentiment analysis
	PriceCop – Price Monitor and Prediction Using Linear Regression and LSVM-ABC Methods for	(Modern Education and Computer Science Mecs) 08 February 2021	

	Title of Paper	Place of publication	Title of Paper
	E-commerce Platform		
	A Hybrid Approach for Product Price Prediction	European Journal of Engineering and Technology Research (71-75 Shelton Street, Covent Garden, London, WC2H 9JQ, United Kingdom)	
2	Stock Market Prediction Using Machine Learning Techniques	October 4, 2022 Savannah, United States Workshop on Advances in Computation Intelligence, its Concepts & Applications 2022 May 17-19	PriceCop – Price Monitor and Prediction Using Linear Regression and LSVM-ABC Methods for E-commerce Platform
3	Forecasting E-Commerce Products Prices by Combining an Autoregressive Integrated Moving Average (ARIMA) Model and Google Trends Data	a publisher of open access scientific journals MDPI 24 December 2018	A Hybrid Approach for Product Price Prediction
4			Stock Market Prediction Using Machine Learning Techniques
5	Stock Market Forecasting Using Machine Learning Algorithms	Department of Electrical Engineering, Stanford University, Stanford, CA 2012	Forecasting E-Commerce Products Prices by Combining an Autoregressive Integrated Moving Average (ARIMA) Model and Google Trends Data
6	Stock Price Prediction Using Machine Learning and Deep Learning Frameworks	Proceedings of the 6th International Conference on Business Analytics and Intelligence, Bangalore, India 2018/12/20	
7	E-Commerce Price Forecasting Using LSTM Neural Networks,	International Journal of Machine Learning and Computing, Vol. 8 No. 2, April 2018	Stock Market Forecasting Using Machine Learning Algorithms
8			Stock Price Prediction Using Machine Learning and Deep Learning Frameworks
9	Stock Prediction by Integrating Sentiment Scores of Financial News and MLP-Regressor: A Machine Learning Approach	Procedia Computer Science 31 January 2023 Volume 218	E-Commerce Price Forecasting Using LSTM Neural Networks,

Table 1. Continued.

	Dataset	Accuracy	Weaknesse.	Future work
1	1- Traditional data (data of historical transactions of stocks and technical indicators) consisting of (3377) rows, and each row contains six columns 2- Non-traditional data (securities and financial news) It consists of (2351) news articles and (33,500) posts in forums For the period from 01 July 2017 to 30 April 2020. CITIC Securities Co., Ltd.-Bank of China Co., Ltd. - Industrial and Commercial -Bank of China Limited Co., Ltd -China Construction Bank Co., Ltd - Agricultural Bank of China Co., Ltd	Tests appear that the anticipated closing cost of the stock is closer to the genuine closing cost than the person information source, and the supreme blunder implies that it accomplishes 2.386835 which is way better than the conventional strategies	Needs more optimization eg there is less intrigued in preparing labelled data.	Improvement of the proposed demonstrate by paying more consideration to preparing the labelled data
2	Between (365 and 372) daily price for ten electronic products collected during the period from June 1, 2019 to July 6, 2020	Pricing Prediction Accuracy Using LSSVMABC is 84%	The data set is small and the products in it are few	The data set is large and contains many products
3	Collection of 535 records of California homes And the second group is a collection of more than 12 million records of Brazilian housing advertisements	The proposed show has accomplished a Cruel Squared Logarithmic Mistake (MSLE) of 0.05 and a R ² of 0.91.	The accessible datasets appropriate for this sort of explore are not constrained to fair two sets of information	Moving forward the demonstrate by including more information that contains the characteristics of the items, through which the capacity to anticipate item costs can be expanded

	Dataset	Accuracy	Weaknesse.	Future work
4	Each house is represented by four pictures: a front view of the house, the bedroom, the kitchen, and the bathroom. and numerical characteristics: the number of bedrooms, the number of bathrooms, the square footage of the home, the zip code of the home's location, and the list price	(Mean Absolute Percentage Error)		
	Stock market for US airlines from 2-2013 to 2-07-2018 This dataset contains 1258 rows and 7 columns. Feature description	Random Forest: 0.36 Decision Tree: 1.60 Artificial Neural Network: 0.37 Support Vector Regression: 3.56		in the long run, it is aiming to work on progressed ANN developmental strategies like Calculation Algorithm to diminish the MAPE values for superior implementations
	Amazon 9 millions of products and 96 millions of tuples collected our data was Between 2015 and 2016	MAPE	No more outside highlights have been chosen (For illustration, the ubiquity of a specific items, on Twitter, surveys on Google items, or Another site on items notoriety over time)	Dissect stock markets and their design and perform figure of long and brief operations by tallying data coming from social networks (e. g., Facebook, Twitter, Stock Twits, etc.)
	Google Trends history from 2015 to the middle of 2016	None		news and any other kind of information that can be removed and associated with the stock markets values
5	Test 10%	2.31%		
	Test 20%	9.56%		
	Test 30%	9.25%		
		Google Trends		
	Test 10%	1.98%		
	Test 20%	5.69%		
	Test 30%	6.64%		
		Sentiment		
	Test 10%	1.83%		
	Test 20%	7.69%		
Test 30%	7.90%			
	Google Trends, Sentiment			
Test 10%	3.59%			
Test 20%	11.37%			
Test 30%	10.65%			
6	Daily price of NASDAQ from 04-Jan2000 to 25- Oct -2012	Support Vector Machine: 74.4%	The precision of SVM utilized in this can be not up to the desires, other calculations have to be be investigated for testing.	explore other creative and practical methodologies that might abandon in fact better performance on stock promote forecasting.
		Random Forest Classification: Tata-95%, Hero-70%	The time period of the information is reasonable brief term, the exactness will be superior measured with long term information as this as it were appears the execution on a negligible level.	The utilized of classification based approaches and other relapse procedures +for predicting
7	NSE Data of: Tata Steel Hero Moto	SVM: Tata-95%, Hero-76% KNN: Tata-83.33%, Hero-80% Decision Tree: Tata-96.4%, Hero-78.5%		price movement
8	Data series available on Amazon France from 3 April 2016 to March 23, 2017, in Prices related to a telephone product, Available here. Samsung Galaxy S7 (32 GB)	SVR: 39.912 (RMSE) LSTM: 23.640 (RMSE)	Data in cost related to one phone item he Samsung World S7 (32 GB)	Dataset can be extended for more item to consider what sort of items are anticipated accurately
	The historical time series stock data of four companies (Reliance, Tata Motors, Tata Steel, HDFC) And sentiments of financial new	0.90 and shows a high correlation between stock price and financial news.	The lack of a comprehensive, detailed analysis of previous studies and their results related to predicting stock prices using machine learning algorithms.	models can be employed LSTM, can be used to remember the trend and More sentiments can be added

9. Conclusion

The Amazon Marketplace has become a popular platform for online retail, and predicting product prices It has become an important area of research. This study aims to predict the future price direction of a particular product on the Amazon Marketplace using historical information and external factors such as Google Trends and customer reviews. The study uses Amazon APIs and crawlers to collect product information and pricing data, preprocesses the data using techniques such as tokenization and feature selection, and uses various machine learning algorithms such as decision trees, support vector regression, and random forest to predict product prices. The study also explores the challenges of web scraping and the potential applications of web harvesting in e-commerce businesses. The findings of this study provide insights into predicting product prices on the Amazon Marketplace and its potential applications in the e-commerce industry.

Acknowledgments

I would like to express my heartfelt gratitude and appreciation to all the faculty members of the Department of Computer and Information Technology at Ibb University who taught me. They provided me with exceptional education and guidance, contributing to the development of my skills and knowledge in the field of computer and information technology.

References

- [1] Shengting Wu, Yuling Liu, Ziran Zou & Tien-Hsiung Weng (2021). S_I_LSTM: stock price prediction based on multiple data sources and sentiment analysis, Taylor & Francis Group Connection Science. doi: 10.1080/09540091.2021.1940101.
- [2] Mohamed Zaim Shahre, Sofanita Mutalib and Shuzlina Abdul-Rahman (2021) "PriceCop – Price Monitor and Prediction Using Linear Regression and LSVM-ABC Methods for E-commerce Platform", Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia nformation Engineering and Electronic Business, doi: 10.5815/ijieeb.2021.01.01.
- [3] Rola M. Elbakly, Magda M. Madbouly, and Shawkat K. Guirguis (2022) "A Hybrid Approach for Product Price Prediction", European Journal of Engineering and Technology Research doi: <http://dx.doi.org/10.24018/ejeng.2022.7.5.2883>.
- [4] Jagruti Hota 1, Sujata Chakravarty 2, Bijay K. Paikaray 3 and Harshvardhan Bhojar, (2022)." Stock Market Prediction Using Machine Learning Techniques", Savannah, United States Workshop on Advances in Computation Intelligence, its Concepts & Applications., 2022 May 17-19.
- [5] Salvatore Carta, Andrea Medda, Alessio Pili, Diego Reforgiato Recupero and Roberto Saia (2018)"Forecasting E-Commerce Products Prices by Combining an Autoregressive Integrated Moving Average (ARIMA) Model and Google Trends Data, doi: 10.3390/fi11010005.
- [6] Shunrong Shen, Haomiao Jiang and Tongda Zhang,(2021) "Stock Market Forecasting Using Machine Learning Algorithms", Department of Electrical Engineering Stanford University.
- [7] Jaydip Sen, Department of Analytics and Information Technology Praxis Business School Kolkata, INDIA,(2018) "Stock Price Prediction Using Machine Learning and Deep Learning Frameworks" 6th International Conference on Business Analytics and Intelligence (ICBAI) At: Indian Institute of Science, Bangalore, INDIA.
- [8] Houda Bakir, Ghassen Chniti, and Hédi Zaher (2018)," E-Commerce Price Forecasting Using LSTM Neural Networks", International Journal of Machine Learning and Computing, doi: 10.18178/ijmlc.2018.8.2.682.
- [9] Dr. G Madhusudhan, 2Nitin Gopalakrishna Bhat, 2Sahana Venkatraman Patgar, 2Chandan N A, 2Bharath S V, (2021)"E-COMMERCE PRODUCT PRICE TRACKER", 1Assistant. Professor, 2Student 1,2Computer Science and Engineering 1,2JSS Science and Technology University, Mysuru, India Journal of Emerging Technologies and Innovative Research (JETIR) June, Volume 8, Issue6.
- [10] Mahla Nikoul | Gholamreza Mansourfar1 | Jamshid Bagherzadeh2, (2019) "Stock price prediction using DEEP learning algorithm and its comparison with machine learning algorithms", 1 Faculty of Economics and Management, Urmia University, Urmia, Iran 2 Faculty of Electrical and Computer Engineering, Urmia University, Urmia, Iran waileyonlinelipray, 20 September, doi: 10.1002/isaf.1459.
- [11] Henrys Kasereka "Importance of web scraping in e-commerce and e-marketing (2020) SSRN Electronic Journal doi: 10.6084/m9.figshare.13611395.v1.
- [12] Krystyna Kuźniar, Maciej Zajac (2015). Some methods of pre-processing input data for neural networks Institute of Fundamental Technological Research, Polish Academy of Science Computer Assisted Methods in Engineering and Science, 22: 141–151.
- [13] Shilpi Kulshrestha, M. L. Saini. (2020). Study for the Prediction of E-Commerce Business Market Growth Using Machine Learning Algorithm. 5th IEEE International Conference on Recent Advances and Innovations in Engineering- ICRAIE 2020 (IEEE Record#51050).
- [14] Haishan Gao, Jingqian Li, Zhaoqiang Bai (2017). Sales Prediction Based On Product Titles and Images with Deep Learning Approaches CS230: Deep Learning, Fall 2021, Stanford University.
- [15] Naeem Ahmed Mahoto1, Rabia Iftikhar1, Asadullah Shaikh2, Yousef Asiri2, Abdullah Alghamdi2 and Khairan Rajab (2021) Tech Science Press Intelligent Automation & Soft Computing doi: 10.32604/iasc.2021.018944.
- [16] Isaac Kofi Nti a, Adib Zaman a, Owusu Nyarko-Boateng a, Adebayo Felix Adekoya b, Frimpong Keyeremeh (2023), A predictive analytics model for crop suitability and productivity with tree-based ensemble learning, Decision Analytics Journal doi.org/10.1016/j.dajour.2023.100311
- [17] Junaid Maqbool, preeti Aggarwal, Ravreet Kaur, Ajay Mittal, Ishfaq Ganie (2023), Stock Prediction by Integrating Sentiment Scores of Financial News and MLP-Regressor: A Machine Learning Approach, Procedia Computer Science 31 -1 202.