RESEARCH ARTICLE OPEN ACCESS

E-Commerce Products Price Tracking and Prediction: A Survey

Mishra Pranav*, Patil Omkar**, Sharma Harsh***, Vaswani Naveen****
*(Student, Department of Computer Engineering, WIEECT Ulhasnagar (West), India
Email: pranavmishra1006@gmail.com)
** (Student, Department of Computer Engineering, WIEECT Ulhasnagar (West), India
Email: op09072000@gmail.com)
*** (Student, Department of Computer Engineering, WIEECT Ulhasnagar (West), India
Email: harsh.v23labs@gmail.com)
**** (Assistant Professor, Department of Computer Engineering, WIEECT Ulhasnagar (West), India
Email: vaswani.naveen@gmail.com)

----------**********----------

Abstract:

E-Commerce Products Price Tracking and Prediction is a system that offers an innovative solution for consumers and product sellers on any E-Commerce platform, using prediction techniques the consumers are offered an estimation of the future prices of products available on E-Commerce websites. This data can be used to make economic decisions regarding purchase and can also help sellers better understand the market and make financial decisions for the future. The following is a literature survey paper in which we go through various research papers to find suitable techniques to carry out the said system development.

Keywords — **Price Tracking, E-Commerce, Web Harvesting, Machine Learning**.

----------**********----------

I. INTRODUCTION

E-Commerce Products Price Tracking and Prediction is a revolutionary solution to multiple problems, Price Tracking and Prediction can help consumers to estimate at what point should they purchase a particular product and decide economically on the basis of the information provided, This system doesn't just help the consumers but also the sellers since the data that will be acquired by this system can also be used to analyse the sales and future decisions of the enterprise regarding financial aspects of products. This system requires predictive algorithm for using historical data and producing accurate predictions to display to the users. In this paper we study various papers from different sources that provide us with similar research studies that have used predictive algorithms, to find the techniques that best fit the description of the system requirement.

II. PREDICTION TECHNIQUES

A. K Nearest Neighbour

The K Nearest Neighbour is an instance-based learning example used for classification where the training data is stored so that a classification for a new unclassified record can be found simply by comparing it to the most similar records in the training set. The value of k determines how many more similar records are considered in the training record to classify a test record [3].

We calculate the distance between the two dependent and independent variables by applying function for distance dist(a,b), where a,b are scenarios composed Number of features, such that a={a1,a2,a3,…,aN}, b={b1,b2,b3,…,bN}.K nearest neighbours is a simple calculation that records all accessible cases and arranges new cases based on the k neighbours' majority votes. The case assigned to the class is in general normal among its K closest neighbours as determined by a separation work.

These separation capacities can be of different types (Euclidian, Hamming, etc.). If K is greater than one, the case is effectively relegated to the class of its nearest neighbour. Occasionally, selecting K becomes a test while performing kNN displaying. [6].

B. Long-Short Term Memory

Long short-term memory is a deep learning artificial recurrent neural network architecture. Unlike traditional feedforward neural networks, LSTM includes feedback connections. It is capable of processing not only single data points, but also entire data sequences.

To forecast the next time step, the weight values in the network must be updated, which necessitates the preservation of the initial time step data. An RNN can only learn a limited number of short-term affiliations; however, long-term time series, such as 1000-time steps, cannot be learned by RNNs; however, LSTMs can learn these long-term affiliations properly. If the length of the input vector is d and the number of neurons in the hidden layer is h, then the memory space required for an LSTM cell is O (d*h). The amount of memory required in LSTM is O(d*h) because the result of the next cell (t+1) is replaced in the same memory for the old values. Because computational models have an infinitely large memory, space complexity is not normally considered a primary issue; therefore, time complexity is of the utmost importance. [5]

C. Random Forest

A random forest is a machine learning technique for solving regression and classification problems. It makes use of ensemble learning, which is a technique that combines many classifiers to solve complex problems. A random forest algorithm is made up of numerous decision trees.

In more formal terms, each tree is determined by the values of a random vector sampled independently and with the same distribution across all trees in the forest. Each tree votes during classification, and the most popular class is returned.[3]

The random forest has the property that a small group of input features is randomly selected in each node of each tree, and for node division, the best feature with the highest information efficiency is selected for tree growth rather than searching all features. These features are fewer in number than the main features. Each tree in the RF grows to its maximum size using the CART decision tree algorithm, with no pruning.[5]

D. Decision Tree

Classification trees are decision trees that are based on answering questions and using this information to make a decision. A classification tree is one that determines whether a decision is fit or unfit by asking a series of questions and using the answers to arrive at a viable solution.

The Classification and Regression Tree (CART) algorithm generates binary decision trees with exactly two branches for each decision node. CART recursively partitions the training data set into subsets of records with similar target attribute values. The CART algorithm builds the trees by performing an exhaustive search on each node for all available variables and all possible splitting values, and then selecting the optimal split based on some efficient splitting criteria.[3]

E. Support Vector Machine

The SVM is another nonlinear model that has seen widespread use in financial markets in recent years and has produced promising results. The SVM has been used for time series prediction in non-stationary variables, unjustifiability of classic methods, and time series complexity. SVM models are divided into two types: SVM and SVR. SVRs are a type of SVM that is used to forecast future prices. In the prediction process, the SVM has the ability to eliminate irrelevant and scattered data and improve prediction precision. The SVM is based on statistical training theory's minimization of structural risk.

The basis for the SVM is the linear classification of data, where we choose a line with higher reliability. Solving the equations of the best line for data by using quadratic coding techniques, which are known for solving constraining problems. [5]

III. WEB HARVESTING

Web scraping is the automated retrieval (and processing) of data from websites. Web scraping is the process of creating a semi-structured document from the internet, typically in the form of web pages like HTML or XHTML and analysing the document to extract specific data from the page to be used for other purposes.
Web scraping has emerged as an important strategy for e-commerce businesses, particularly in terms of providing rich data-based insights.[7]

A. Practices [7]

- Price-monitoring and Product Research
- Online price comparison
- Better Customer analysis
- Market Analysis

B. Potential challenge of web scraping

One of the difficulties we find while scraping information from various websites is that these websites have various structures. That means the templates of websites are different and unique and therefore tough to generalize.

Other difficulty could be longevity. Since the developers update their websites frequently, you cannot be reliant on a scraper for a long period of time. Although the webpages might not have major differences it can still hinder the process of acquiring data. [11]

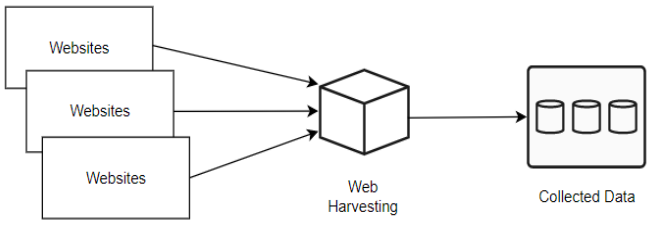

Fig.1 Web Scraping Schema

IV. LITERATURE REVIEW

For prediction of price of e-commerce products, we require a dataset that we acquire from the ecommerce website and we need an algorithm which we will train using our dataset for implementing it in real life applications.
Through this we will be able to predict the price of various products on e-commerce websites. Listed below are a few papers that have been published in various journals related to web scraping and various machine learning algorithms that we can use for our system development.

Table 1

Comparison of various algorithms for price prediction and tracking

Ref. No.	Title of Paper	Algorithm	Dataset	Accuracy	Improvement
[1]	Stock Market Forecasting Using Machine Learning Algorithms.	Support Vector Machine	Daily price of NASDAQ from 04-Jan-2000 to 25-Oct -2012	Support Vector Machine :74.4%	The accuracy of SVM used in this is not upto the expectations, other algorithms need to be explored for testing.
[2]	Stock Market Data Prediction Using Machine Learning Techniques.	Random Tree	Stock price of Apple.inc	Random Tree-0.16(RMSE)	Use a lot of other machine learning and deep learning techniques are not considered it is hard to say random forest would be applicable for other such predictive algorithms.

[3]	Stock Price Prediction Using Machine Learning and Deep Learning Frameworks.	Random Forest Classification SVM KNN Decision Tree	NSE Data of: Tata Steel Hero Moto	Random Forest Classification: Tata-95%, Hero-70% SVM: Tata-95%, Hero-76% KNN: Tata-83.33%, Hero-80% Decision Tree: Tata-96.4%, Hero-78.5%	The time period of the data is fairly short term, the accuracy will be better measured with long term data as this only shows the performance on a minimal level.
[4]	An Adaptive SVR for High-Frequency Stock Price Forecasting,	SVR	The data set is from Shanghai Stock Exchange, including SH600006, SH600016, SH600026, SH600036, SH600056.	SVR: SH06-0.75(RMSE) SH16-1.04(RMSE) SH26-0.79(RMSE) SH36- 1.81(RMSE) SH56-6.17(RMSE)	Instead of SVR models the best SVR model should have been studied against other Machine Learning algorithms to show true performance of SVR.
[5]	Stock price prediction using DEEP learning algorithm and its comparison with machine learning algorithms.	Random Forest SVR	Daily data of iShares MSCI United Kingdom from January 2015 to June 2018	Random Forest-0.391(RMSE) SVR-0.341(RMSE)	Various LSTM models can be used to determine which model performs the best.
[6]	A Research on Bitcoin Price Prediction Using Machine Learning Algorithms.	KNN Random Forest	Historical value of Bitcoin value	Training: KNN- 92.99% Random Forest- 89.78% Test RSS: KNN-90.5% Random Forest-71.56%	Known techniques such as SVR that are know for such being efficient can be used.
[7]	Importance of Web Scraping in E-Commerce and E-Marketing,	Web Scraping	Amazon.com	----------	----------

[8]	Stock Price Prediction Using Machine Learning and LSTM-Based Deep Learning Models.	LSTM-Based Deep Learning Models	The historical index values of NIFTY 50	LSTM (univariate time series ,1 Week): 0.0311(RMSE) LSTM (univariate time series ,2 Week): 0.0353(RMSE) LSTM (univariate encoder -decoder, 2 Week): 0.0369(RMSE) LSTM (multivariate encoder -decoder ,2 Week): 0.1711(RMSE)	Other deep learning models could be mentioned to show how LSTM performs the best.
[9]	E-Commerce Price Forecasting Using LSTM Neural Networks,	Long-Short Term Memory Support Vector Regression	Amazon.fr	SVR: 39.912(RMSE) LSTM: 23.640(RMSE)	Dataset could be expanded for more products to study what type of products are predicted accurately

V. CONCLUSION

In this literature survey we studied papers on price prediction using Machine learning algorithms and Web Harvesting. We have found various suitable algorithms that could be used for the development of the system. After thoroughly studying all the research papers it was observed that the best results were acquired by using Long-Short Term Memory deep learning algorithm. Since LSTM outperforms every other algorithm for outcome of the provided datasets, we have concluded that it is the best algorithm for price prediction given historical datasets.

REFERENCES

[1] Shunrong Shen, Haomiao Jiang and Tongda Zhang, "Stock Market Forecasting Using Machine Learning Algorithms", Department of Electrical Engineering Stanford University.

[2] Edgar P. Torres P.1(&), Myriam Hernández-Álvarez1, Edgar A. Torres Hernández2 , and Sang Guun Yoo1,"Stock Market Data Prediction Using Machine Learning Techniques", Facultad de Ingeniería de Sistemas, Escuela Politécnica Nacional, Quito, Ecuador.

[3] Jaydip Sen, "Stock Price Prediction Using Machine Learning and Deep Learning Frameworks", Department of Analytics and Information Technology Praxis Business School Kolkata, INDIA.

[4] YANHUI GUO, SIMING HAN, CHUANHE SHEN, YING LI, XIJIE YIN and YU BAI, (Member, IEEE), "An Adaptive SVR for High-Frequency Stock Price Forecasting", 1. School of Information Technology, Shandong Women's University, Jinan 250300, China ,2. School of Computer Science, Shaanxi Normal University, Xi'an 710062, China 3.School of Engineering and Computer Science, California State University, Fullerton, CA 90831, USA,IEEE Access (Volume: 6).

[5] Mahla Nikou1, Gholamreza Mansourfar1, Jamshid Bagherzadeh, "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, Intelligent Systems in Accounting Finance & Management 26(2).

[6] Lekkala Sreekanth Reddy, Dr.P. Sriramya, "A Research on Bitcoin Price Prediction Using Machine Learning Algorithms", INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH VOLUME 9, ISSUE 04, APRIL 2020 ISSN22778616.

[7] Kasereka Henrys, "Importance of Web Scraping in E-Commerce and E-Marketing, Bugema University", January 19, 2021.

[8] Sidra Mehtab, Jaydip Sen, Abhishek Dutta, "Stock Price Prediction Using Machine Learning and LSTM-Based Deep Learning Models", Department of Data Science and Artificial Intelligence Praxis Business School KolkataIndia, 06 February 2021.

[9] Houda Bakir, Ghassen Chniti, and Hédi Zaher, "E-Commerce Price Forecasting Using LSTM Neural Networks", International Journal of Machine Learning and Computing, Vol. 8, No. 2, April 2018.

[10] Chayan Bhattacharya," Web Scraping E-commerce sites using Selenium & Python", https://medium.com.

[11] Aditya Sharma," Web Scraping with Python: Beautiful Soup", https://datacamp.com.

[12] Arun Kumar, " Price Prediction using Machine Learning Regression - a case study", https://towardsdatascience.com.

