IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 4, May 2022
Impact Factor: 6.252

Machine Learning in E-Commerce WebApp

Madeeha Quazi, Aniket Krushanaraj Pimpalkar, Parag Janardhan Gajbhiye, Riyaz Kadar Sayyad, Shrirang Shailesh Chopade
Final Year Students, Department of Computer Science and Engineering
Sipna College of Engineering & Technology, Amravati, Maharashtra, India

Abstract: *This research highlights the machine learning algorithm used in e-commerce website and how it helps in product recommendation of the e-commerce website used nowadays on a large scale basis.*

Keywords: Machine Learning, Algorithm, KNN, E-Commerce

I. INTRODUCTION

E-commerce interfaces are elemental place for people to attain, compare, and ultimately buy wares. They operate Machine Learning (ML), Business Intelligence (BI), mathematical formalism, and artificial intelligence (AI) to beget marketable erudition about client deportment, bringing compensations for both clients themselves and dealers. Level of development in this area does not include inclusive and updated survey that examines most prevailing aims of e- commerce-related studies and acceptable ML proficiencies and frameworks for selective sheaths. In this context, we inaugurate methodical literature review that revisits modern enterprises to practice ML techniques on different e- commerce situations. Participations to level of development are bifold: (I) Inclusive summary of ML procedures and their connection with main aim of e-commerce interfaces, along with effect on gain; (ii) novel taxonomy to reorient ML-based e-commerce capabilities, which aids researchers to relate and arrange endeavors in this developing area. This inclusive literature review authorizes researchers and e-commerce administrators to maintain research-based projects better and divert expenses as well as human resource chores. [1] With the explosive growth of data, it is one of the most important challenges for modern enterprises to develop data-based infrastructures. The scientific discovery of Artificial Intelligence (IA) is an open opening for a wide range of applications, which allows large quantities of real time business data to be obtained. As sales of electronic electronics continue to grow, the question arises whether e-commerce will overcome or completely replace sales of physical stores. Although nature of technological advances and consumer behavior are increasingly unpredictable, there are some key trends that I believe will mark both the medium and long term growth of e-commerce.

II. MACHINE LEARNING

Machine learning is a science found as a branch of artificial intelligence in the 1950s. The first steps of machine learning go back to the 1950s but there were no compelling research and developments on this science. It is a science that will upgrade more in the future. Machine learning is based on the fundamental of finding the prime model for the new data amid the preceding data. Therefore, machine learning research will go on side by side with the increasing data and information. This research includes the algorithm used for websites and how machine learning is a part of eCommerce websites. This study aims to convey the awareness of machine learning, which has become very demanding nowadays, and its applications to the developers. [2]

Machine learning is the process of providing computers with the ability to learn by using data and experience like a human brain. The main aim of machine learning is to create models, which can train themselves to improve, perceive complex patterns, and find solutions/predictions to new problems by using the previous data. Machine learning (ML) is used to teach machines how to handle data more accurately. Sometimes after viewing the data, we cannot predict the exact information from the data. This is why we apply machine learning. With the huge number of datasets available, the demand for machine learning is at its peak. Many industries apply machine learning to extract appropriate data. The use of machine learning is to learn from the data. Many studies have been done on how to make machines learn by themselves without being explicitly programmed. Many mathematicians and programmers apply several approaches to find the solution to this problem which is having a big amount of data sets. [3]

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 4, May 2022
Impact Factor: 6.252

Understanding the machine learning workflow
We can define the machine learning workflow in 3 stages.

2.1. Gathering Data

The process of gathering data depends on the type of project we desire to make, if we want to make an ML project that uses real-time data, then we can build an IoT system that using different sensors data. The data set can be collected from various sources such as a file, database, sensor and many other such sources but the collected data cannot be used directly for performing the analysis process as there might be a lot of missing data, extremely large values, unorganized text data or noisy data. Therefore, to solve this problem Data Preparation is done.

2.2 Data pre-processing

Data pre-processing is one of the most important steps in machine learning. It is the most important step that helps in building machine learning models more accurately. In machine learning, there is an 80/20 rule. Every data scientist should spend 80% time for data pre-processing and 20% time to actually perform the analysis.

2.3 Researching the Model that is best for the Data

Machine Learning can be Supervised or Unsupervised. If you have very little data and are accurately titled as data for training, choose Supervised Learning. While Unsupervised Learning would generally give better performance and results for big data sets. Today every person is using machine learning knowingly or unknowingly. From getting a recommended product in online shopping to updating photos on social networking sites all are applications of machine learning.

2.4. Training and Testing the Model on Data

For training a model we initially split the model into 3 three sections which are 'Training data'
'Validation data' and 'Testing data'.
You train the classifier using 'training data set', tune the parameters using 'validation set' and then test the performance of your classifier on unseen 'test data set'. An important point to note is that during training the classifier only the training and/or validation set is available. The test data set must not be used during training the classifier. The test set will only be available during testing the classifier.

Training set: The training set is the material through which the computer learns how to process information. Machine learning uses algorithms to perform the training part. A set of data used for learning, that is to fit the parameters of the classifier.

Validation set: Cross-validation is primarily used in applied machine learning to estimate the skill of a machine learning model on unseen data. A set of unseen data is used from the training data to tune the parameters of a classifier.

Test set: A set of hidden data used only to assess the performance of a fully-specified classifier. In a data set, a training set is implemented to create a model, while a test set is to validate the model built. Usually, a data set is divided into a training set, a validation set in each iteration, or divided into a training set, a validation set and a test set in each iteration.

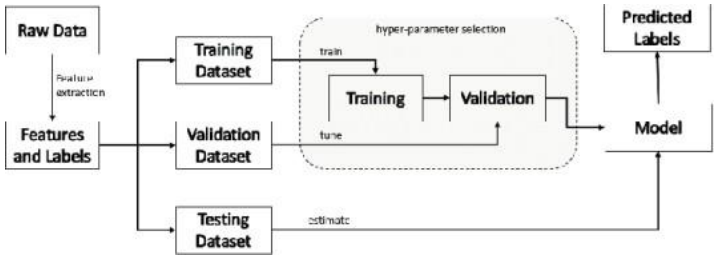
The model uses any one of the models that we had chosen in step 3/ point 3. Once the model is trained we can use the same trained model to predict using the testing data i.e. the unseen data. Once this is done we can develop a confusion matrix, this tells us how well our model is trained. A confusion matrix has 4 parameters, which are 'True positives', 'True Negatives', 'False Positives' and 'False Negative'. We prefer that we get more values in the True negatives and true positives to get a more accurate model. The size of the Confusion matrix completely depends upon the number of classes.

2.5 Evaluation

Model Evaluation is an integral part of the model development process. It helps to find the best model that represents our data and how well the chosen model will work in the future.

To improve the model we might tune the hyper- parameters of the model and try to improve the accuracy and also looking at the confusion matrix to try to increase the number of true positives and true negatives.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 4, May 2022
Impact Factor: 6.252



III. Literature Review

Analyzing the most globally cited documentsFootnote2 in the dataset (those with 100 citations) reveals that recommender systems are the main topic of interest in this research area. Recommender systems are software agents that make recommendations for consumers by implicitly or explicitly evoking their interests or preferences (Bo et al., 2007). The topic has been investigated in many flavors, including hybrid recommender systems (Burke, 2002), personalized recommender systems (Cho et al., 2002), collaborative recommender systems (Lin et al., 2002) and social recommender systems (Li et al., 2013). The central concept of interest is personalization, specifically leveraging recommender systems to offer more personalized product/service recommendations to customers using e-commerce platforms. Thus, designing recommender systems that surpass existing ones is the leading orientation of AI in e- commerce research. Researchers have mostly adopted experimental rather than theory-driven research designs to meet this overarching research objective. Research efforts focus more on improving the performance of recommendations using advanced AI algorithms than on understanding and modelling the interests and preferences of individual consumers. Nevertheless, the advanced AI algorithms developed are trained primarily using customer product reviews [4]

Research on AI in e-commerce is published in two main journal subject areas: computer science & AI and business & management. This result confirms the multidisciplinary nature of this research area, which has both business and technical orientations. Four themes characterize research on AI in e-commerce: sentiment analysis, trust & personalization, optimization, AI concepts, and related technologies. The sentiment analysis theme represents the stream of research focused on interpreting and classifying emotions and opinions within text data in e-commerce using AI techniques like ML and natural language processing (NLP). The trust and personalization theme represents research that focuses on establishing trust and making personalized recommendations for consumers in e-commerce using AI techniques like collaborative filtering, case-based reasoning, and clustering algorithms. The optimization theme represents research that focuses on using AI algorithms like genetic algorithms to solve optimization problems in e- commerce. Finally, the AI concepts and related technologies theme represent research that focuses on using different techniques and concepts used in the research area.

Collaborative filtering, recommender systems, social information filtering, latent Dirichlet allocation, and matrix factoring techniques are the foundational topics in research on AI in e- commerce.[5] They were identified by analyzing the most cited references in the dataset. These references were mostly literature reviews and documents that discussed the basic ideas and concepts behind specific technologies or techniques used in recommender systems.

IV. Collaborative Filtering

One of the potent personalization technologies powering the adaptive web is collaborative filtering. Collaborative filtering (CF) is the process of filtering or evaluating items through the opinions of other people. CF technology brings together the opinions of large interconnected communities on the web, supporting filtering of substantial quantities of data. In this chapter we introduce the core concepts of collaborative filtering, its primary uses for users of the adaptive web, the theory and practice of CF algorithms, and design decisions regarding rating systems and acquisition of ratings. We also discuss how to evaluate CF systems, and the evolution of rich interaction interfaces collaborative filtering is the most commonly used recommender system. The products are recommended based on the opinions of other customers. This opinion includes the trends of a particular customer on several products and several customers on a particular product. These systems try to find the neighbor of an item. They showed two approaches for representation; aggregate and center based respectively. Then the respective algorithm is used for neighborhood formation and can be reduced to low dimension from high dimension.[6] Finally recommendation generation is performed by observing most frequent item and applying appropriate association rule.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 4, May 2022
Impact Factor: 6.252

V. K NEAREST NEIGHBOUR (KNN)

K-Nearest Neighbor is one of the simplest Machine Learning algorithms based on Supervised Learning technique. It assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories. It also stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm. K-NN algorithm can be used for Regression as well as for Classification but mostly it is used for the Classification problems.[7] As it is a nonparametric algorithm, which means it does not make any assumption on underlying data[5]. It is also called a lazy learner algorithm because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset.

K-nearest-Neighbours (kNN) is a simple but an effective classification method. The value of k is automatically determined, is varied for different data, and is optimal in terms of classification accuracy. Experiments were carried out on some public datasets collected from UCI machine- learning repository to test.[7]

Machine learning classifier is a crucial part of pattern recognition system and it is also an important research field of machine learning[9]. KNN algorithm at the training phase just stores the dataset and when it gets new data, then it classifies that data into a category that is much similar to the new data. [8]The K-NN working can be explained on the basis of the below algorithm:

- Step-1: Select the number K of the neighbors
- Step-2: Calculate the Euclidean distance of K number of neighbors
- Step-3: Take the K nearest neighbors as per the calculated Euclidean distance.
- Step-4: Among these k neighbors, count the number of the data points in each category. Step-5: Assign the new data points to that category for which the number of the neighbor is maximum.
- Step-6: Our model is ready.

VI. CONCLUSION

With the advancement of digitization in marketing, the web applications are using more smarter techniques to improve customer experience. Machine Learning techniques have already proven to be an indispensable part in filling the gaps in this experience. It has been applied in various fields like Product recommendation thereby reducing the need for human intervention by substituting it with machine efficiency. But it has not yet completely been Mastered for automation, since it stumbles when it comes to more complex queries, answering a variety of questions and detecting human emotion. In the future, these limitations might be attended to, bringing forth a better ecommerce experience.

REFERENCES

[1]. Micu, Adrian & Geru, Marius & Capatina, Alex & Constantin, Avram & Rusu, Robert & Andrei Alexandru, Panait. (2019). Leveraging e- Commerce Performance Through Machine Learning Algorithms. Annals of Dunarea de Jos University of Galati. Fascicle I. Economics And Applied Informatics. 25. 162- 171.10.35219/eai1584040947.

[2]. Zohuri, B. and Zadeh S., Artificial Intelligence Driven by Machine Learning and Deep Learning, Nova Science Pub Inc (October 22, 2020), first edition

[3]. Zohuri, B. and Mossavar Rahmani, F., "Artificial Intelligence Driven Resiliency with Machine Learning and Deep Learning Components" Int J Nanotechnol Nanomed, 2019, Volume 4 | Issue 2 | 1 of 8

[4]. https://link.springer.com/article/10.1007/s12525-022-00537-z

[5]. M. Bkassiny, Y. Li, and S. K. Jayaweera, "A survey on machine learning techniques in cognitive radios," IEEE Communications Surveys & Tutorials, vol. 15, no. 3, pp. 1136–1159, Oct. 2012.

[6]. Sohail, Shahab & siddiqui, jamshed & Ali, Rashid. (2012). Product Recommendation Techniques for Ecommerce - past, present and future. Ijarcet. 1. 219-225

[7]. Mitchell, T.M., 2006. The discipline of machine learning (Vol. 3). Carnegie Mellon University, School of Computer Science, Machine Learning Department.

[8]. P. Harrington, "Machine Learning in Action", Manning Publications Co., Shelter Island, New York, ISBN 9781617290183, 2012