Predictive Analysis in E-commerce Domain and Cloud Platforms – A Survey

[1] Nisha Jha, [2] Prof.Amit Saxena

M.Tech. Scholar

Computer Science Department

Truba Institute of Engineering & Information Technology, Bhopal

nishajha846@gmail.com

Associate Professor & Head

Computer Science Department

Truba Institute of Engineering &Information Technology, Bhopal

amitsaxena@trubainstitute.ac.in

ABSTRACT

For doing business in this communication era, web is the best medium. For business owners and consumers, online businesses broke down the barrier of time and space compared to the physical shop or office. Big companies around the world are realizing that E-commerce (EC) in not just buying and selling over Internet rather improving the competence than other giants in the market. E-Commerce has allowed businesses to offer more choices to consumers. Increasing choice, however, has also increased the amount of data and information that consumers must process before they are able to select which items meet their needs. To grow their potential markets, the big shopping platforms like Amazon, Flipkart and ebay etc. wants to utilize Machine Learning (ML) potential to build unmatched competitiveness in the market. ML has empowered businesses to analyze all queries, whether searched or abandoned, from all the users. Application of machine learning for Predictive Analytics can enhance business opportunities by analyzing customer's past click-through behaviour, purchases, preferences and history in real-time. To make fast real-time predictions from e-commerce data, the algorithm must be capable of processing huge volume of training data in reasonable time, and must be capable of handling large number of classes. So, the paper investigates the use of machine learning in E-commerce domain and its importance in predictive analysis. The need of Cloud platforms for analyzing E-commerce data is also established in this work. The paper concludes with exploration of potential areas of research in the field of E-commerce.

Keywords

Machine Learning, E-commerce, Predictive Analytics, Classification, Cloud

1. INTRODUCTION

E-Commerce (EC) has rapidly grown in past decade, enables consumers to purchase any product within few clicks. For the success of such online shopping platforms the key component is their ability to retrieve desired products for the consumers very quickly. The prospects in field of EC got boosted rapidly with the application of Machine Learning (ML) techniques such as Association Rule Mining and Classification..

A snapshot of a famous shopping website [1] suggesting or recommending frequently bought items with a mobile phone, is shown in Figure 1. The website is applying association principle and suggesting the frequently bought items like mobile covers and protective glass etc., along with a mobile.

FIGURE 1 COMES HERE

The factor such as ever growing amount of data for classification and constraints on response time, have made ML tasks a challenging job in the EC domain. Multiple choices of cloud computing models are available for different workload management, performance and computational requirements. The popular statistical tools and environments like Octave, R and Python are now embedded in the cloud as well [2].

Section-2 presents few papers that are reviewed to investigate the use of machine learning in E-commerce domain and its importance in predictive analysis. The need of Cloud platforms for analyzing E-commerce data is established in section-3. Finally the paper is

concluded in section-4 by exploring the research areas in e-commerce domain.

2. LITERATURE REVIEW

The important findings of work [3] indicate that the area of customer retention received most research attention. Among these, one-to-one marketing and loyalty programs are most popular research areas. Classification and Association Rule Mining based models are most commonly used models for data mining in Customer Relationship Management (CRM). Machine Learning is naturally a time consuming task, so Cloud computing paradigm turned out to be important alternatives to speed-up machine learning platforms [3]. The review provides a roadmap for future research in the field of application of data mining techniques in CRM.

Zornitsa Kozareva [4] of Yahoo labs, studied that the different taxonomies of organizing products are in use at various famous shopping platforms. Different taxonomies organize products making it hard and labour-intensive for sellers to categorize the products. To address the challenge an automatic product categorization mechanism is proposed, which assigns the correct product category from a taxonomy for a given product title. In work [4], 319 categories organized into 6 levels and performance evaluation is done for 445 product titles using multiple algorithms. The best f-score of 0.88 is obtained.

The work [5] focuses challenges presented by Big Data classification and prediction of intrusion in network traffic data, on the fly. It requires ML approaches that must capture global knowledge of the traffic patterns. Also, Big Data properties put significant challenges to implement machine learning frameworks. The paper also discusses the problems and challenges in handling Big Data classification.

The authors [6] presented an explanation of how recommender systems are related to some traditional database analysis techniques. They also analyze the recommender systems at six leading websites and examined how recommender systems help E-commerce sites increase sales. Based on the study, taxonomy of recommender systems is created, that includes the consumer inputs required, the knowledge required from the database, the ways the recommendations are offered to consumers, recommendation creation technologies, and personalization level of the recommendations. Also, several open research problems in the field of recommender systems and privacy implications of such systems are examined.

Product classification for E-commerce sites is a necessity for successful business and product sales. It is vital that the products are listed in accurate categories so that users find their products in appropriate categories. The paper [7] explores the experimental results that were conducted with various feature classification methods in combination with three main classifiers Naïve Bayes, SVM, K-Nearest Neighbors, along with LDA an unsupervised document topic classifier.

Authors [8] presented a method for classifying products into a set of known categories by using Naïve Bayes supervised learning and its improved version. For building features for classifier the product catalog information from different distributors on Amazon.com is used. The purpose is to show the improvement in automation of product categorization.

Table 1, showing comparison of various works using different flavours of Machine Learning for Classifying and predicting the E-commerce product category.

Table 1: Comparison: ML in E-commerce domain

Reference	Dataset	Machine Learning Algorithms Compared
[4]	Manually annotated product titles from Yahoo's shopping platform	One-Against-All (OAA)and Error Correcting Tournament (ECT)
[8]	Multiple datasets from Ingram Micro, the leading IT distributor	Naïve Bayes, KNN and Tree Classifiers
[7]	Information on 35,000 products and attributes for 45 categories from Amazon site.	Naïve Bayes, SVM, K-Nearest Neighbors, and Latent Dirichlet Allocation (LDA is an unsupervised document topic classifier).
[13]	Real-world industrial data from eBay	Fat-Naïve Bayes, Flat K-Nearest Neighbour (KNN) and KNN-SVM

3. NEED OF CLOUD PLATFORMS FOR ANALYZING E-COMMERCE DATA

Classification and regression analysis on very large amounts of training data can require a great deal of computer memory and processing power. Especially with data representing complex non-linear behaviours, as with text, speech, handwriting, face recognition, stock price prediction, and financial forecasting, the computing bill can be quite large. However, the emergence in recent years of cloud computing changes everything. IaaS providers like Amazon Web Services (AWS) [9] and 'Google Cloud' [10] platform now offer access to virtually unlimited

computing power on demand, in the form of clustered parallel servers that can be used for an hourly fee. The shopping platforms like Amazon, e-Bay, Walmart and Yahoo etc. organize products into different product taxonomies which make it hard for sellers to categorize. The categorization of goods is major challenge identified in E-Commerce.

Machine Learning is inherently a time consuming task, thus plenty of efforts were conducted to speed-up the execution time [2]. Cloud computing paradigm and cloud providers turned out to be valuable alternatives to speed-up machine learning platforms. The authors [2] have investigated how cloud computing model impacted the field of ML.

Cloud computing platforms for ML enables developers of all skill levels to use machine learning technology. It enables users:

- With visualization tools and wizards that guide them through the process of creating machine learning (ML) models without having to learn complex ML algorithms and technology

- To build sophisticated, large scale ML models including sophisticated regression models for classification.
- To auto scales the ML based applications.To generate predictions by finding patterns in our existing data.

The popular statistics tools and libraries are already deployed in cloud. The existing platforms also allow users to create a Hadoop [11] cluster in the cloud and run jobs on it. Several libraries are now available for of distributed implementations for ML algorithms for on-premise deployments of complex systems for data mining and data analytics.
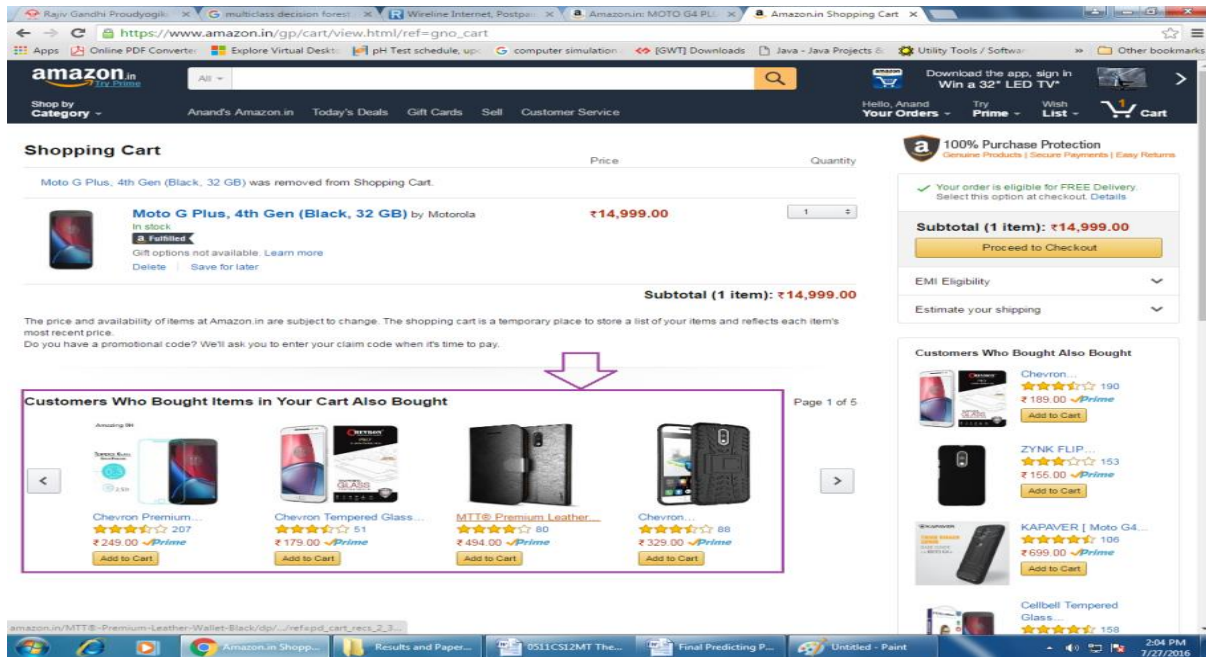

Fig 1: Screenshot of Cart at Shopping Website www.amazon.in

4. CONCLUSION AND FUTURE DIRECTIONS

The potential areas of research in the field e-commerce include correct classification of products to improve customer experience with e-commerce website. Other potential areas are: finding abandoned customer, enhancing sale probability using recommendation systems, finding reason for customer churn, customer purchase behaviours for promotions and offers.

The machine learning process isn't especially simple. To make life easier for people doing machine learning, Cloud provides several different components. Several Cloud platforms like Azure [12] Machine Learning was designed for applied Machine Learning and enables the process of creating ML models. Cloud platform uses best-in class algorithms and a simple drag-and-drop interface along with easy deployment service. In future we propose a dynamic ML model at cloud platform, for product categorization of e-commerce website.

5. ACKNOWLEDGMENTS

REFERENCES

[1]   www.amazon.in

[2]   Daniel Pop, "Machine Learning and Cloud Computing: Survey of Distributed and SaaS Solutions", https://www.researchgate.net/publication/257068169.

[3]   ] E.W.T. Ngai ,, Li Xiu, D.C.K. Chau, "Application of data mining techniques in customer relationship management: A literature review and classification", Expert Systems with Applications 36 (2009) 2592–2602, Elsevier.

[4]   Zornitsa Kozareva, "Everyone Likes Shopping! Multi-class Product Categorization for e-Commerce", Human Language Technologies: The 2015 Annual Conference of the North American Chapter of the ACL, pages 1329–1333, Denver, Colorado, May 31 – June 5, 2015, Association for Computational Linguistics.

[5]   Suthaharan, S., "Big Data Classification: Problems and Challenges in Network Intrusion Prediction with Machine Learning. Performance Evaluation Review", 41(4), ACM 2014

[6]   J. BEN SCHAFER, JOSEPH A. KONSTAN, JOHN RIEDL, "E-Commerce Recommendation Applications", Data Mining and Knowledge Discovery, 5, 115–153, 2001, Kluwer Academic Publishers, Netherlands.

[7]   Srinivasu Gottipati and Mumtaz Vauhkonen, "E-Commerce Product Categorization",

[8]   Sushant Shankar and Irving Lin, "Applying Machine Learning to Product Categorization",

[9]   https://aws.amazon.com/machine-learning/

[10] https://cloud.google.com/ml/

[11] Apache Hadoop Website http://hadoop.apache.org/

[12] David Chappell, "INTRODUCING AZURE MACHINE LEARNING", A GUIDE FOR TECHNICAL PROFESSIONALS, Sponsored by Microsoft Corporation, 2015.

[13] Dan Shen, Jean-David Ruvini, Badrul Sarwar, "Large-scale Item Categorization for e-Commerce", CIKM'12, 2012, Maui, HI, USA. ACM 978-1-4503-1156-4/12/10