

Elastic Cloud based implementation of Online Retail store Analytics using Machine Learning

Rahul Subramaniam, Arnab Jana, Kunal Bhandari
Department of Information Science and Engineering
R.V College of Engineering
Bengaluru, India
rahulsubramaniam0510@gmail.com

***Abstract**—In the highly competitive online retail sector, making decisions intuitively or by simply viewing selling history to improve overall sales is insufficient. Most of the online stores of today demonstrate a reliance on data analytics to make business, customer centric, operational as well as strategic decisions. This given work performs various kinds of data analysis on the data of online retail stores such as Flipkart, Amazon. The work also provisions the methodology to performs data analysis on the same. The dataset for this project was obtained by developing a web crawler and scraper that extracted relevant details with respect to every store item for the given retail stores. This unstructured data was then catalogued, and appended to an already existing database to create a custom database. The categorization process was followed by a graphical exploratory analysis, that would help data analysts, business analysts to make product specific discussions. Post this, Predictive analysis was done on the data, wherein the description of products was given as input, and clustering was done on the basis of frequency of words. This portion of the project made use of various Natural Language Processing techniques, and primarily serves the purpose of being a Recommender system. The code was run in the Amazon cloud (Amazon Machine Image in the Amazon Elastic cloud) to ameliorate the execution speed. The project also views similarity of products and buying pattern of customers by the application of existing Machine Learning algorithms on the product choices that are made by the customers. The accuracy of the various models are then compared to conclude which algorithm is most suited for prediction. A detailed analysis of this kind can help provide inferences and details to organizations to enhance their product sales and profits.*

***Keywords**— Clustering, Exploratory Analysis, Lemmatization, Machine Learning, Predictive analysis, Stemming.*