

Implementing Big Data Analytics in Business Management

Dhiraj Lahoti, Atharva Panse

Computer Engineering Department, Padmabhooshan Vasantdada Patil Institute of Technology, Pune,
Maharashtra, India

ABSTRACT

Article Info

Volume 8, Issue 3

Page Number : 107-110

Publication Issue :

May-June-2022

Article History

Accepted: 20 April 2022

Published: 03 May 2022

The E-Commerce industry is facing a big challenge of handling huge amounts of data and profiting from it by analysing the data. This is caused by the Information Revolution. Big Data Analytics (BDA) are techniques used in order to manage and study this pile of data, which comprises of text messages, social media comments etc. In addition to this, Big Data Analytics can be used in the e-commerce sector to get increased revenues and draw in more customers. This project focuses on analysing and implementing these functionalities of BDA in e-commerce for the benefit of both, the seller and the consumer. E-commerce vendors as well as other enterprises such as general stores, malls etc use BDA techniques to get the competitive advantage by understanding consumer behavioural patterns. These patterns are then studied and used to achieve consumer loyalty and thus getting higher revenues through their businesses. Plus, the recommendation systems that are obtained through big data analytics help the customers shop better due to a personalized and tailor-made searching experience.

A business can experience tremendous growth if the customers and the customer reviews/patterns of the business are studied properly. The knowledge gained from this can be put to tackle the negatives about the business. Our project is based into two modules, Customer Segmentation and Sentimental Analysis. The project studies customers using the customer segmentation module. In Customer Segmentation, K-Means Algorithm is used to make clusters of customers based on their age, spending patterns etc. The second module is Sentimental Analysis which focuses on divide the positive reviews with the negative reviews. The business can then work upon tackling the negatives that the consumer thinks the business has.

Keywords: Segmentation Analysis, Sentimental Analysis, Naïve Bayes, K-mean Clustering, Machine Learning, Data Visualization, Data Preprocessing

I. INTRODUCTION

Customer Segmentation: -

Customer Segmentation is the procedure carried out by businesses or enterprises which categorizes the customers based on demographics or behavioural patterns so that marketing to those customers is easier. The customer segmentation groups hence formed could also be utilized in beginning discussions related to marketing persona building, since customer segmentation is mainly utilized to enlighten the business' message, market position and improving the reach and the sales of the business. Hence, for the technique to be efficient, the marketing personas have to be kept close to the customer segments.

A marketing persona could be defined as the embodiment of a customer segment. A business generally creates many personas which align with their variety of customer segments.

A huge set of customer segments is required to base our personas, which requires Sentimental Analysis.

Sentimental Analysis: -

Sentimental Analysis, also known as opinion mining, is a Natural language Processing (NLP) tool which recognises the emotional tone of a textual body. Many organizations use this to make out and categorize reviews about any service or idea, making it extremely popular. Artificial Intelligence, Data Mining and Machine Learning is used to mine the text out for sentiment and subjective information.

These Sentiment analysis systems help businesses gather insights from unorganized and unstructured data in forms of text from online sources such as mails, blogs, excerpts, comments, forms and social media etc. This data is not needed to be processes manually due to the employment of rule based, automatic or hybrid methods. Rule based systems work with sentiment analysis complying to predefined, lexicon based rules, while automatic systems gains insights from data with

machine learning techniques. A hybrid approach is the mixture of both.

II. METHODS AND MATERIAL

Tools: -

For Model development, the following Software Requirements are:

Operating System: Windows 7/10 or any Linux Ubuntu

Language: Python3

Tools: Jupyter Notebook, Tableau Desktop, Microsoft Excel (optional)

Technologies Used: Exploratory Data Analysis, Python

Software Requirements: -

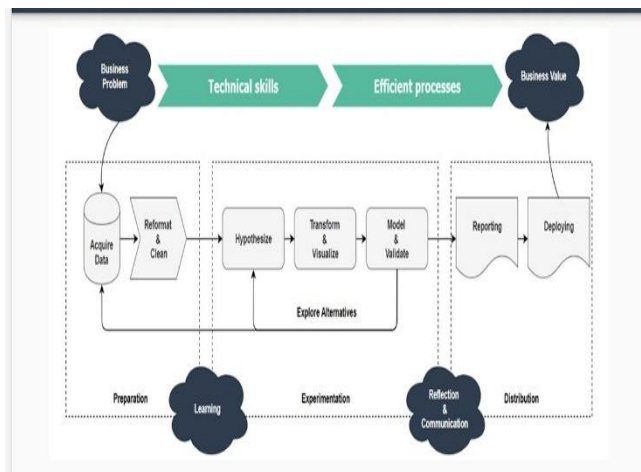
Operating System	Any OS with clients to access the internet
Network	Wi-Fi Internet or cellular Network
GitHub	Versioning Control
Software	Tableau Desktop
Platform	Jupyter Notebook
Google Chrome	Medium to find references to do system testing, Display and run model

Hardware Requirements: -

For Model development, the following Hardware Requirements are:

Processor:	Intel or AMD Ryzen or higher
RAM:	8 GB
Space on Disk:	512 SSD
Device	Any device with internet access

Approach: -



Naïve Bayes: -

$$P(A | B) = \frac{P(B | A)P(A)}{P(B)}$$

P(A|B) is Posterior probability: Probability of hypothesis A on the observed event B.

P(B|A) is Likelihood probability: Probability of the evidence given that the probability of a hypothesis is true.

P(A) is Prior Probability: Probability of hypothesis before observing the evidence.

P(B) is Marginal Probability: Probability of Evidence.

K-means Algorithm: -

K-means clustering is an unsupervised Machine Learning Algorithm that is used to solve segmentation and clustering problems in machine learning. K-mean clustering algorithm computes the centroids and iterates until we find the optimal centroid. The fundamental step for our algorithms is to determine

the optimal number of clusters into which data may be clustered.

III.RESULTS AND DISCUSSION

The first result we get is 3 distribution plots to show us the range of the age of our customers. Their annual income and their spending scores.

Next, we plot out a bar graph showing the number of male and female customers.

To simplify further, we draw out a violin plot for the density of the age, annual incomes and spending scores based on gender.

Another bar plot is drawn to bifurcate our customers based on their age ranges. The age ranges being studied were 18-25, 26-35, 36-45, 46-55, 55+.

A scatter plot for 2 criteria, the annual income and spending scores is drawn. This shows our customers in clusters to give us a clearer idea.

All of these graphs and plots help us gain deeper insights on the data that we're working upon.

Elbow method is used to find out the optimal number of centroids required to study our customers on specific criteria such as age and spending scores or annual income and spending scores.

A. Figures and Tables

Using elbow method, the optimal number of clusters in found out. This gives us the best number of clusters for a plot between age and spending score.

In our data, the spending scores range from 0 to 100. For our first plot, the Optimal number of clusters is 4, hence, there would be 4 centroids. Figure 1 shows the plot between the age and spending scores.

Similarly, the Elbow method is used again to find out the optimal number of clusters for a plot between the Annual Income of our customers and their spending scores. Here, the number of clusters for this plot is 5, which makes the number of centroids to be 5. Figure 2 shows the plot between Annual income and Spending Scores.

Figure 3 shows a combined plot between the Age, Annual Incomes and the Spending Scores of our customers.

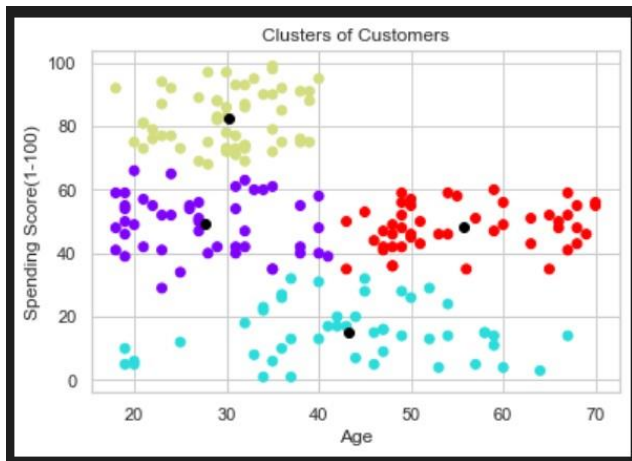


Figure 1. Clustering based on Age and Spending Score

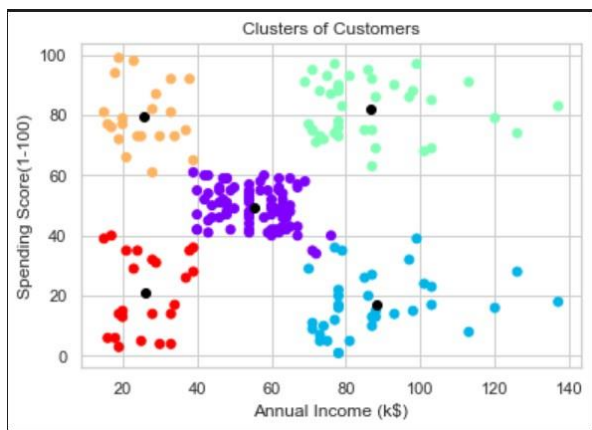


Figure 2. Clustering based on Annual Income and Spending Score

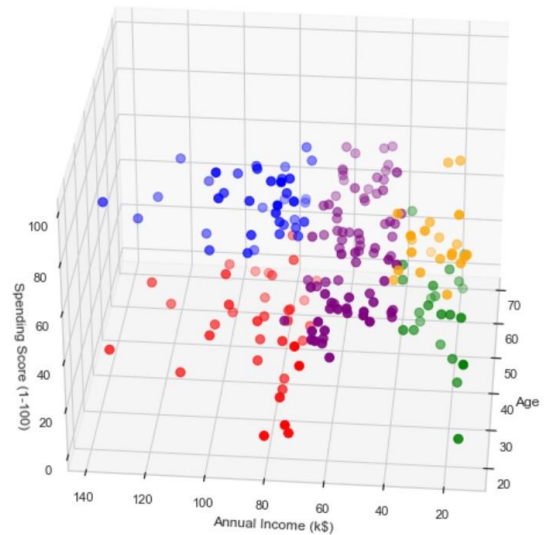


Figure 3. Final Clustering based on Age, Annual Income and Spending Score

IV. CONCLUSION

The usage of big data analytics is aimed to to get data about the customer buying patterns, study the data and make educated predictions about what a customer will buy, for an instance if a customer buys bread, there is a high chance that he even buy's butter, thus can be placed nearby: which will help the company make higher profits in the long run. This gives us an instance of customer segmentation. Reviews about this move, if available, can be gathered, categorized and studied using sentimental analysis

V. REFERENCES

- [1]. Sarah S. Alrumiah, Mohammed Hadwan., "Implementing Big Data Analytics in E-Commerce : Vendor and Customer View" 2021 Institute of Electrical And Electronics Engineers Access. (Mar 2021), DOI:10.1109/ACCESS.2021.306135
- [2]. Zeenia Singla, Sukhchandan Randhawa, Sushma Jain "Sentiment Analysis of Customer Product Reviews Using Machine Learning" 2017

- International Conference on Intelligent Computing and Control, (2017)
- [3]. Z. Singla, S. Randhawa and S. Jain, "Sentiment analysis of customer product reviews using machine learning," 2017 International Conference on Intelligent Computing and Control (I2C2), 2017, pp. 1-5, doi: 10.1109/I2C2.2017.8321910.
- [4]. Juni Nurma Sari, Lukito Edi Nugroho, Ridi Ferdiana, P. Insap Santosa "Review on Customer Segmentation Technique on Ecommerce" 2011 American Scientific Publishers. (2011)

Cite this article as :

Dhiraj Lahoti, Atharva Panse, "Implementing Big Data Analytics in Business Management", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 8 Issue 3, pp. 107-110, May-June 2022. Available at doi : <https://doi.org/10.32628/CSEIT228324>
Journal URL : <https://ijsrcseit.com/CSEIT228324>