WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

A Study on Product Data Analytics on Big Data

Arugula Deepthi[1], Ms. P. Kanchanamala[2]

[1]UG Student, [2]Associate Professor
GMR Institute of Technology, Rajam, India deepthiarugula2805@gmail.com

ABSTRACT:

Big data is a new driver of the world economic and societal changes. The world's data collection is reaching a tipping point for major technological changes that can bring new ways in decision making, managing our health, cities, finance and education. While the data complexities are increasing including data's volume, variety, velocity and veracity, the real impact hinges on our ability to uncover the `value' in the data through Big Data Analytics technologies. Big Data Analytics poses a grand challenge on the design of highly scalable algorithms and systems to integrate the data and uncover large hidden values from datasets that are diverse, complex, and of a massive scale. Potential breakthroughs include new algorithms, methodologies, systems and applications in Big Data Analytics that discover useful and hidden knowledge from the Big Data efficiently and effectively. Data can be of any type i.e., Structured data, Unstructured data and semi structured data where the size varies from TBs to ZBs. In this process, we present Apache Spark as a better ecosystem to handle big data efficiently and vigilantly. Amazon is a growing company today. So, the product Development from year to year is analyzed by Big Data. Here we using some machine learning algorithms like Support Vector Machine(SVM),Decision tree(DT),Random Forest(RF) etc, for analeyzing the product data analytics on big data. This study is to investigate new product development on Big Data Analytics.

Keywords: Big Data, Apache spark, Product Reviews, Data Analytics, Machine Learning Algorithm

INTRODUCTION:

Product data analytics on big data refers to the process of analyzing large volumes of data generated by products or product-related activities to gain insights,make informed decisions, and optimize various aspects of product development, marketing, and performance. This field has gained significant importance in recent years as businesses recognize the value of leveraging data to enhance their products and services. Here's an introduction to the key concepts and aspects of product data analytics on big data.Big data refers to the massive amounts of structured and unstructured data that organizations collect from various sources, including customer interactions, sensors,social media, and more. Big data is characterized by the 5Vs: volume, velocity, variety, Veracity and Value.Analytics involves the use of statistical, mathematical, and computational techniques to uncover patterns, trends, and insights within data. In product data analytics, these techniques are applied to product-related data to extract valuable information.Product data encompasses information related to a company's products or services, such as specifications, usage data, customer feedback, and sales data. This data can come from multiple sources, including e-commerce platforms, IoT devices, and customer reviews.Many products are now equipped with sensors that collect real-time data on usage and performance.Direct feedback from customers can be a rich source of data.

LITERATURE SURVEY:

Rameshwar Dubey1•David J.et.Al.(2021) New products play an important role in gaining competitive advantage and can significantly boost organizational performance.In this research they suggests that organizations should invest in building alliance management capability (AMC), big data analytics capability (BDAC) and information visibility (IV) to achieve their desired NPD success.In this paper they stated seven research hypotheses.Introducing new ideas to the market quickly, with a focus on high customer satisfaction, and on developing products which are easier to manufacture, use and repair than existing ones. For every seven new product ideas, about four enter development, one and a half are launched, and only one succeeds.The Review of literature suggests that the role of alliances among partners engaged in developing the NPD capability has not received significant attention. In this study, therefore, we aim to investigate how the alliances between large, well-established firms and small firms influence their joint NPD capability under the influence of visibility.

Tomohiko Sakao and Abhijna Neramball et.Al.(2020) In this survey they introducing the broad scope of environmental sustainability and the role of Product/Service Systems (PSSs).And they explain the significance of new technologies, especially Big Data Analytics (BDA), in improving PSSs.In this paper they discussing the challenges and the growing importance of PSSs in addressing environmental concerns.And explain the basic concepts and importance of PSSs in the context of sustainability.Here they exploring the potential of BDA in improving or enabling PSSs.They discussing the key technological advancements and applications of BDA in industry for sustainability and PSSs.They discuss the lack of clear understanding regarding past

research in the intersection of PSS design and BDA.And they Identify gaps in existing literature that demonstrate insufficient exploration or connection between the two fields.The methodologies used in the article (literature synthesis and systematic literature review) to bridge the gap between PSS design and BDA.They Analyze how the two areas of research intersect and complement each other, emphasizing specific areas where BDA can enhance PSS design and delivery .

Saqib Ali , Petra Poulova et. Al. (2020)have contextualize the global challenge of increasing haze pollution, its adverse effects on human health, and the urgent need for sustainable environmental solutions.They highlighting the emergence of smog as a significant problem in Pakistan, particularly stemming from hazardous vehicle emissions.In this paper they discuss the Pakistan's environmental challenges, particularly focusing on air pollution and the significant impact of smog.They elaborate on repercussions of smog on public health and its recognition as a fifth season. And they investigate how the booming automobile industry in Pakistan is affected by environmental challenges, specifically focusing on sustainable product development and organizational performance. The review existing literature on the application of big data analytics in environmental sustainability efforts, particularly in the context of the automobile industry and emissions reduction. They exploring the concepts of sustainable development programs and their integration with big data analytics. They analyze literature related to the development of conceptual models exploring the impact of big data analytics on organizational performance within the context of sustainable product development. And highlighting the positive effect of big data analytics on sustainable product development and its subsequent impact on organizational performance. In this discussion the confirmed mediation role of sustainable program development between big data analytics and organizational performance.

Yufan Wang and Haili Zhang et.Al. (2020) The impact of big data-embedded processes in new product development (NPD) on sustainable innovation performance. In this study they conducts a cross-national comparison using data collected from 1858 NPD projects in the USA, UK, and Australia.In this they influence of big data-embedded business analysis, product design, testing, and commercialization on sustainable innovation performance and various financial metrics in the three countries. They explores the existing literature that delves into the relationship between NPD and sustainable organizational performance. This might include studiesdiscussing this impact of innovation on long-term sustainability, market competitiveness, environmental impact,and social responsibility.This research focusing on the significance of data-driven approaches in NPD processes. This could cover studies that compare the efficacy of experience-driven decision-making with data-driven decision-making, emphasizing how the latter contributes to improved outcomes in product development.That specifically examines how big data-embedded processes,including business analysis, product design, testing,and commercialization, affect sustainable innovation performance and financial metrics. They highlighting the influence of these processes on sales growth, gross margin,and other financial indicators.

Yajun Wang , Shengming chen et.Al. (2021) In this studies the use of big data in agricultural practices, including aquaculture.Here they investigating how big data is collected, processed, and applied in improving various aspects of food production, such as crop growth, livestock management, and specifically, seafood production.

The challenges and solutions regarding distributed database systems, focusing on efficient storage, retrieval, and the limitations of traditional relational databases in handling big data and the challenges posed by the increasing volume, velocity, and variety of data.This data management of large-scale, geographically dispersed data.The optimization of storage and processing methods, such as parallel computing, to handle big data challenges efficiently.The trade-offs between real-time database systems and high-performance computing in managing large-scale datasets. Here how real-time systems impact data acquisition resolution, server load,and cost implications.

Chih-Hsuan Ashley Cheng et.Al.(2022) In this studies they discuss the broader impact of BDA on firms operations, emphasizing its potential in creating business value, enhancing decision-making, and transforming business processes.Here how BDA has been instrumental in various industries and sectors.In this paper the outline challenges faced by firms in adopting and implementing BDA.They understanding the difficulties in making sense of the data,adapting organizational processes, and harnessing the full potential of BDA. The role of BDA in driving firm performance, as seen in profitability,market share,customer retention, and financial outcomes.And they understanding the existing research landscape on the impact of BDA on various aspects of business operations and outcomes.They summarized how the BDA has been utilized to enhance product innovation performance.The transfer of technology-specific know-how to product innovation, and how BDA facilitates or hinders this process.The gaps in the literature that examine the role of BDA in fostering innovation.

Michael Riesenera , Günther Schuha , Christian Döllea et.Al. (2019) In this paper provides an overview of the Industrie 4.0 concepts,including data utilization and itsimpact on manufacturing processes.And here they explains the challenges of data utilization, including data overload,and strategies for transforming data into valuable insights.This papers covers various aspects of business intelligence,data integration, and analytics, which are essential for making informed decisions.

The integration of business rules and predictive analytics in decision-making systems,emphasizing the role of data in this process.Data Mining explores the importance of a data-driven culture in organizations and strategies to overcome challenges related to data utilization.They provides insights into data mining techniques and methodologies that could be useful in extracting meaningful information from a large volume of data.Here they focuses on the strategic use of enterprise architecture,which includes data modeling and integration for effective decision-making.And here they discusses tools and strategies for managing large volumes of data, particularly in the context of scientific research but applicable to broader data challenges.

Sobia Wassana , Xi Chenb et.Al.(2021) In this paper they Opinion Mining and Sentiment Analysis"by Bo Pang and Lillian Lee provides an in-depth view of sentiment analysis and opinion mining. The paper discusses methods for identifying sentiment in text data, which is particularly useful in understanding consumer opinions and behavior. It also emphasizes the extraction of meaningful information (positive or negative sentiments)from large datasets like those found in Amazon consumer

reviews."The Impact of Social Media on Consumer Buying Intentions" by Rana Tassabehji and Christine Moor describes the influence of social web sites on consumer behavior and decision-making. It discusses how these platforms,through user-generated content like reviews, recommendations, and ratings, impact purchasing decisions and influence the perception of product quality."Social Network Analysis:Methods and Applications" by Stanley Wasserman and Katherine Faust offers insights into social network analysis methods and their application in understanding network structures. It discusses the analysis of large networks and graph structures, which could be relevant in understanding relationships and interactions betweenconsumers and products in online reviews and social networks.

Monika Mishra, Jaydeep Chopde et.Al.(2019) In this paper they explores the analysis of click stream data to understand and predict consumer behavior in e-commerce.This study delves into the importance of click stream data in identifying patterns and preferences among online shoppers.The Opportunities and Challenges of E-Commerce in the Digital Age" by F. Javier Miranda, Julia Pollak, and John Haltiwanger addresses the rapid growth of e-commerce and its impact on businesses and consumers.In this study they discusses the challenges and opportunities presented by the digital market place,emphasizing the need for efficient analysis of customer sentiments and shopping patterns. The authors examine the utilization of big data analytics in e-commerce.The paper highlights the significance of managing and analyzing large-scale data in e-commerce using frameworks like Hadoop. Hive – A Petabyte Scale Data Warehouse Using Hadoop" by Ashish Thusoo and JoydeepSen Sarma discusses the application of Hive as a data warehouse solution built on top of Hadoop. The study emphasizes the scalability and efficiency of Hive in handling large-scale data for analysis in the context of e-commerce.A Survey on Collaborative Filtering-based Recommender Systems"by X. Amatriain, Alejandro Jaimes,N. Oliver, and Josep M. Pujol presents an overview of collaborative filtering techniques used in recommendation systems. The paper explores how these systems learn user preferences and behavior to provide personalized recommendations, which hold significant value for businesses in e-commerce.

Bickey Kumar Shah, Amar Kumar et.Al.(2021) In this paper they discusses the prevalence of online product usage among the majority of people and the impact of these reviews on purchase decisions facilitated by major e-commerce platforms like Flipkart, Myntra, and Amazon.And they suggests the utilization of machine learning methods to conduct sentiment analysis, categorizing product reviews into positive, neutral, and negative sentiments. They mentions the use of various algorithms such as Logistic Regression,Naive Bayes, and Random Forest for classifying feedback and determining the most effective method for precision, with the Random Forest algorithm identified as the most accurate among those tested.In this studies that explore sentiment analysis techniques applied to consumer reviews and their influence on purchasing decisions. This could include research on the impact of online reviews on consumer behavior and decision-making processes.The significance of using specific datasets for sentiment analysis, such as the Kaggle dataset of Amazon product reviews mentioned in their abstract. Understanding the importance of diverse, high-quality data sources for training and testing sentiment analysis models.In the case studies or real-world applications where sentiment analysis has been used in e-commerce settings to understand consumer sentiments and predict product success or failure.

METHODOLOGY:

1) Data Collection and visualization:

Amazon product reviews has a record of consumer reviews for Amazon products. From the dataset important features are extracted and, the four important features that are relevant for further processing from the dataset that is (ID, Review text, Review rating). The data are then visualized and find out that the data is heavily biased towards positive reviews that's why some neutral and negative reviews are mixed to make dataset balance.

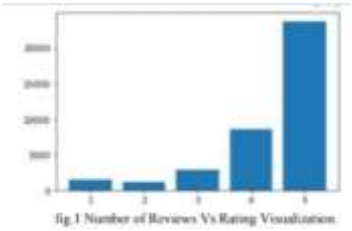


fig.1 Number of Reviews Vs Rating Visualization

B) Preprocessing:

Tokenization is designated as the technique by which a succession of words is separated into individual words, such as names, keywords, expressions, and tokens. Tokens could be phrases, even full sentences of single or individual words. Certain characters are discard, like punctuation marks, vowels and many more in the tokenization process.For the distinct course of action such as parsing and text mining, the tokens act as input values.

Lower case: Word to lower case (TREE -> tree) conversion.All of the words are converted into a lowercase letter as to bring uniformity and a direction in one way.Depending on the country and their culture and languageetc., there are various kinds of stop words in different formats. There are some stop words in the English format,so it has to be deleted.

Stemming: It is the method of normalizing the word into its root form.

Assigning sentiment scores: Sentiment scores are assigned for different ratings. A score of 0 to reviews having ratings 1 and 2, score 1 to reviews having ratings 3, and score 2 to reviews having ratings 4 and 5 respectively are assigned.

Score 0: negative sentiment

Score 1: neutral sentiment

Score 2: positive sentiment

C) Feature Extraction TF-IDF:

It is an analytical metric that evaluates how foremost a term is to a document or record. This is computed by two metrics: how many times a word appears and inverse document frequency of word across the document. Every word or expression has a TF and IDF score of its own. It can be optimized accordingly that the unlikely the word, and the other way round, the maximum is the TF*IDF measure .

The TF of a term or an expression is the number of times in a text that a term or expression appears. The IDF of a term is the course of action of the value in the corpus of that name. It will always be encompassed in the head search results when terms have a large TF*IDF measure in expression/name/term/figure, so everyone can:

1.Avoid thinking about stop-words being used

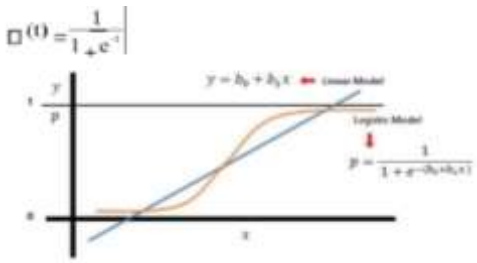
2.Find terms that are searched for the maximum number of times and lower competition successfully. It is extracted function that is found from preprocessing from the cleaned text.

D)Categorization:

Categorization is the process of analyzing opinions based on their emotions into three divisions: Positive, Neutral, and Negative. 80% of data is used for priming the model and 20% for experiment in.

1) Classifiers:

We used several classifiers Logistic regression, Naive Bayes, Random forest classifier. Logistic regression: The probability of a categorical subordinate variable is anticipated by Calculated Relapse. The subordinate encompasses a parallel variable coded as yes or no. Calculated relapse works superior to the large test estimate. The calculated work may be a sigmoid work, which takes any genuine input x and yields esteem between zero and one.



Naïve Bayes Method:

The probabilistic classifier technique is Naïve Bayes . It depends on the Bayes theorem There are two different naïve Bayes variants for text. Multi-nomial naïve Bayes and Bernoulli naïve Bayes. In the Multinomial naïve Bayes method data simply follows a multinomial distribution and here each feature value is count. Bernoulli naïve Bayes data follows a multivariate distribution and each feature is binary. Since proof Y is calculated by the Bayes rule by the Finding emotion, the conditional likelihood of event X occurs.

P(X/Y) = [P(X) P(Y/X)] / P(Y)

Random forest:

Having supremacy on the top of a single decision tree concerning reliability and efficiency, the random forest classifier was chosen . It is a method of ensemble which is based on bulging.

Sentiment score	precision	recall	F1 score
0	0.79	0.63	0.70
1	0.62	0.35	0.45
2	0.93	0.98	0.96

Table 2 Classification report of Logistic regression

Sentiment score	precision	recall	F1 score
0	0.81	0.25	0.38
1	0.54	0.01	0.02
2	0.87	1.00	0.93

Table 3 Classification report of Multinomial Naive Bayes

Sentiment score	precision	recall	F1 score
0	0.61	0.48	0.54
1	0.37	0.23	0.28
2	0.91	0.95	0.93

Table.4 Classification report of Bernoulli Naive Bayes

Sentiment score	precision	recall	F1 score
0	0.89	0.63	0.77
1	0.90	0.46	0.60
2	0.94	1.00	0.96

Table.5 Classification report of Random forest

Approach:

Approach	Methodologies used
Machine Learning	1)Support Vector Machine (SVM) 2)Decision Trees 3)Neural Networks 4)Naïve Bayes 5)Logistic Regression
Wordbook Approach (lexicon)	1)Dictionary-based approach 2)Manual opinion approach 3) Corpus-based approach

Table.1 Methodologies used

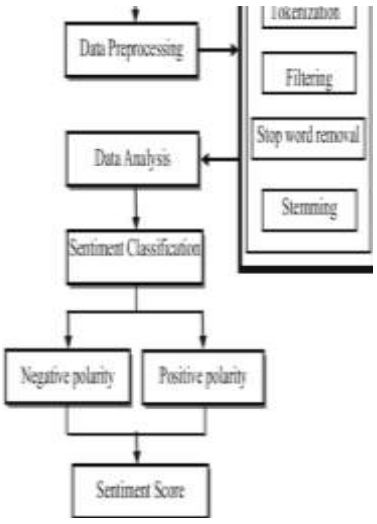
Here they use Machine Learning and word book approach.

Accuracy:

Method	Accuracy
Logistic regression	90.88
Multinomial Naive Bayes	87.09
Bernoulli Naive Bayes	86.46
Random Forest	93.17

Table.6 measured Accuracy

Architecture:



Case Study:

1)

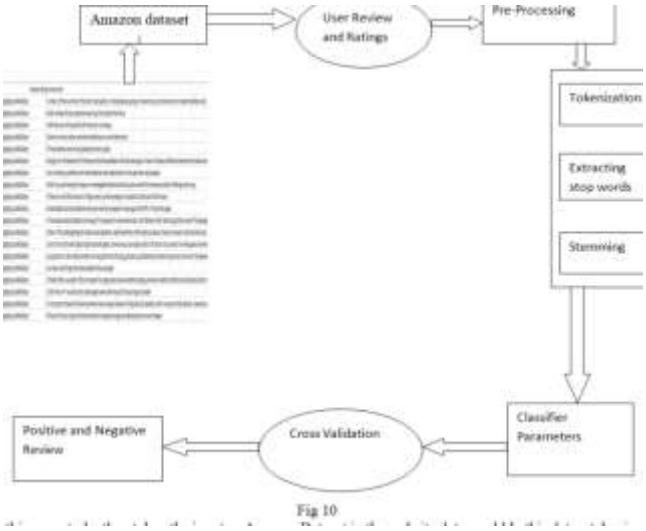


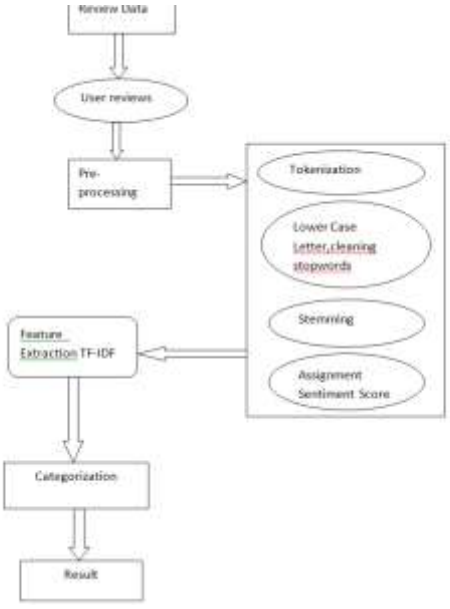
Fig 10

In this case study ,they taken the input as Amazon Dataset in the website data.world. In this dataset having Id's, Reviews and Ratings. Dataset having nearly 28,000 product review and ratings which consists of 60 different products. In this dataset they determining that ratings which consists of 1,2,3,4,5 numbers which representing as 1 and 2 is negative ,3 is neutral,4 and 5 is positive. And here they mentioned that polarity and subjectivity which means emotions which containing positive and negative.

In this dataset they collecting data from data. world which contains review and ratings ,after this they doing pre-processing techinique .In this preprocessing they doing tokenization which means dividing the tokens, removing of stop words and stemming. In this pre-processing they importing some libraries like NLTK(Natural Language Tool Kit),textblob. After the preprocessing they doing classifier parameters which contains classification algorithms(like random forest etc).After this parameters they doing cross validation for model evaluation, model selection .After the cross validation the occurring result is positive and negative review.

In this study the final output is polarity and subjectivity which means positive and negative reviews.

2)



In this case study they taken the input as Amazon Dataset which consists of review & ratings. And the dataset contains ratings from 1 to 5. The data are then visualized and find out that the data is heavily biased towards positive reviews that's why some neutral and negative reviews are mixed to make dataset balance. Exploratory data analysis is carried out to find out the most common words in the text of the reviews .

After the dataset taken as input they doing preprocessing .In the preprocessing first they doing tokenization means the technique by which a succession of words is separated into individual words, such as names, keywords, expressions, and tokens. Certain characters are discard, like punctuation marks, vowels and many more in the tokenization process. For the distinct course of action such as parsing and text mining, the tokens act as input values.

Lower case: Word to lower case (TREE -> tree) conversion. All of the words are converted into a lowercase letter as to bring uniformity and a direction in one way.

Cleaning Stop Words: Stop words are terms in an expression that are not required or needed in text analytics in any field.

Stemming: It is the method of normalizing the word into its root form

Assigning sentiment scores: sentiment scores are assigned for different ratings. A score of 0 to reviews having ratings 1 and 2, score 1 to reviews having ratings 3, and score 2 to reviews having ratings 4 and 5 respectively are assigned.

Score 0: negative sentiment

Score 1: neutral sentiment

Score 2: positive sentiment

After the pre-processing they doing that Feature Extraction

TF-IDF: It is an analytical metric that evaluates how foremost a term is to a document or record.

After this they doing categorization using classification algorithms like Logistic Regression, Naïve Bayes, Random Forest.

They using this classification algorithms and getting best accuracy .

After this they doing Evaluation Metrices in this they doing precision, Recall and F1 score .

Precision::P=TP / TP+FP

Recall::R = TP / TP+FN

F1 Score::F = 2P*R / P+R

And finally they getting accuracy using this classification algorithms based on precision, recall and f1 score.

This is the accuracy they getting

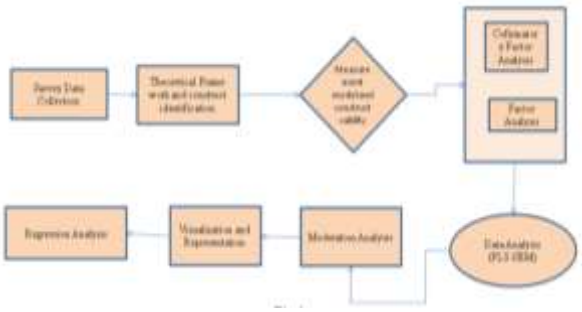
Logistic regression 90.88

Multinomial Naive Bayes 87.09

Bernoulli Naive Bayes 86.46

Random Forest 93.17

3)



1) Survey Data Collection: No specific algorithms, but typically involves survey design, questionnaire creation, and data gathering techniques such as online surveys, questionnaires, or interviews.

2) Theoretical Framework and Construct Identification: No specific algorithms but involves theoretical frameworks like the Dynamic Capabilities View (DCV) that guide the conceptualization of constructs and relationships.

3) Measurement Model and Construct Validity:

Factor Analysis (FA): Used for assessing the validity and reliability of the constructs derived from survey data.

Confirmatory Factor Analysis (CFA): To confirm the measurement model's fit with the data collected.

4) Data Analysis (PLS-SEM):

Partial Least Squares Structural Equation Modeling (PLS-SEM): Employed for hypothesis testing and examining relationships among constructs. PLS-SEM is a multivariate statistical technique used for analyzing structural models. It doesn't rely on assumptions of multivariate normality and is often preferred when dealing with complex models.

5) Moderation Analysis:

Interaction Effects Testing: Methods within PLS-SEM or other techniques to test the moderating role of Information Visibility (IV) on the relationships between AMC, BDAC, and NPD.

6) Visualization and Representation:

Path Diagrams or Structural Equation Models: Graphical representations of the relationships between constructs derived from PLS-SEM.

7) Regression Analysis:

Regression techniques within PLS-SEM used to estimate the relationships and the extent of influence between constructs.

Results and Discussions:

Utilizing advanced analytics for NPD, resulting in more targeted product development, greater innovation novelty, improved market responsiveness, and increased customer satisfaction. Creation of a systematic 10-step schema for conceptual Product

-Service System (PSS) design, addressing the need for structured guidance in PSS design processes.

Mapping potential areas (Steps 1, 2, 3, 4, 6, 7, and 9)in the PSS design schema that can be enhanced by Big Data Analytics(BDA), providing a foundation for future research and practical application. Measure the financial gains or cost savings derived from the adoption of big data analytics and sustainable product development in the automobile sector, assessing profitability, cost efficiency, and revenue growth. Measure the speed and efficiency of innovation within the NPD process, examining the time taken to introduce new products, services, or processes by integrating big data analytics. Practical testing has shown that the proposed storage optimization and data retrieval methods are supportive in constructing a big data resource management platform for seafood products. The multidata source parallel retrieval execution time is significantly faster than the standard Hadoop scheme(32% of the time), and the multichannel data fusion feature extraction algorithm is also notably faster (35% of the time of the standard Hadoop scheme).we show that BDA leads to positive product innovation performance. As such, this study contributes to the research fields of big data in following aspects. First, it contributes to big data literature by extending the growing research on BDA to the context of product innovation. The predominant view in prior research is that BDA plays an important role in determining firm performance . However, researchers have devoted limited efforts to examining whether firms can use BDA to enhance product innovation performance. This study thus augments extant big data literature through confirming that firms can utilize their BDA to enhance product innovation performance. Second, the deployment of BDA presents critical challenges for firms.

The key challenges and solutions related to data utilization in the context of "Industrie 4.0" or the fourth industrial revolution, which emphasizes the use of data and intelligent analysis methods. the need for a structured approach to manage and derive valuable insights from the overwhelming volume of available data, specifically in the context of industrial processes and decision-making within the framework of Industrie 4.0. It doesn’t provide specific performance measures regarding the efficiency or success of the proposed methodology. Instead, it emphasizes the challenges and solutions associated with data utilization in a dynamic business environment. They provided the growth of e-commerce, its advantages, challenges, the role of reviews and ratings in online shopping, the emergence of Big Data due to the vast amount of data, and the significance of recommendation systems. However, it does not explicitly mention performance measures related to these concepts. Performance measures can be variedand may relate to different aspects of e-commerce, data analysis, and recommendation systems. Amazon has become the most popular hub for online purchasing.The useability of each user can be detected by using user comments. The user comments can show the behavior of the buyers satisfactions. The main implication of this research is that we can develop a system that can give us a certain report for any product sold from the Amazon website. We can develop this system by using different machine learning techniques. We have trained our system on different user reviews and check the polarity of different user feedbacks. After We can then test the reliability of the product and increase the product's productivity. Here they mentioned a word best precicion this word considered as a main performance measure. Here they using some classification algorithms for best accuracy. And they using confusion matrix to refers the sample of proportios.

CONCLUSION:

The integration of product data analytics within the framework of big data in manufacturing is transformative. It has revolutionized the industry by providing unparalleled insights into product lifecycles, manufacturing processes, and customer interactions. Through the analysis of vast and varied data sets, manufacturers gain invaluable advantages that significantly impact decision-making, operational efficiency, and customer satisfaction. Product data analytics ,when powered by big data, equips manufacturers with detailed insights into consumer behavior, market trends, and product performance. This information empowers data-driven decision-making and fuels innovation by identifying emerging needs and trends, thereby enhancing product development and design. By harnessing big data analytics on product data, manufacturers streamline operations, minimize inefficiencies, and optimize resource allocation. Insights derived from these analyses help in refining production processes, reducing waste, and improving overall operational efficiency. The utilization of big data analytics on product data enables predictive maintenance, identifying potential issues in products before they occur.

This proactive approach minimizes downtime and elevates after-sales service, improving customer satisfaction and loyalty. This integration isn't just a technological advancement; it's a cornerstone for competitiveness and success in the ever-evolving landscape of the manufacturing industry.

REFERENCES:

[1] Dubey, R., Bryde, D. J., Graham, G., Foropon, C., Kumari, S., & Gupta, O. (2021). The role of alliance management, big data analytics and information visibility on new-product development capability. Annals of Operations Research, 1-25.

[2] Riesener, M., Schuh, G., Dölle, C., & Tönnes, C. (2019). The digital shadow as enabler for data analytics in product life cycle management. Procedia CIRP, 80, 729-734.

[3] Sakao, T., & Neramballi, A. (2020). A product/service system design schema: Application to big data analytics. Sustainability, 12(8), 3484.

Reference3: Ali, S., Poulova, P., Yasmin, F., Danish, M., Akhtar, W., & Javed, H. M. U. (2020). How big data analytics boosts organizational performance: The mediating role of the sustainable product development. Journal of Open Innovation: Technology, Market, and Complexity, 6(4), 190.

[4] Wang, Y., & Zhang, H. (2020). Achieving sustainable new product development by implementing big data-embedded new product development process. Sustainability, 12(11), 4681.

[5] Micus, C., Smeets, J., Böhm, M., & Krcmar, H. (2022). Customer integration in product development using big data: An example of using fleet data to determine parking behavior.

[6] Wang, Y., Cheng, S., Zhang, X., Leng, J., & Liu, J. (2021). Block storage optimization and parallel data processing and analysis of product big data based on the hadoop platform. Mathematical Problems in Engineering, 2021, 1-14.

[7] Cheng, C. H. A., Hong, M. Y., & Cheng, C. C. (2022). A Study on The Effect of Big Data Analytics on Product Innovation Performance. Journal of Student Research, 11(4).

[8] Riesener, M., Schuh, G., Dölle, C., & Tönnes, C. (2019). The digital shadow as enabler for data analytics in product life cycle management. Procedia CIRP, 80, 729-734.

[9] Mishra, M., Chopde, J., Shah, M., Parikh, P., Babu, R. C., & Woo, J. (2019). Big data predictive analysis of Amazon product review. In KSII the 14th Asia Pacific international conference on information science and technology (APIC-IST) (pp. 2093-0542).

[10] Shah, B. K., Jaiswal, A. K., Shroff, A., Dixit, A. K., Kushwaha, O. N., & Shah, N. K. (2021, January). Sentiments detection for amazon product review. In 2021 International conference on computer communication and informatics (ICCCI) (pp. 1-6). IEEE.

[11] Ren, S., Zhang, Y., Liu, Y., Sakao, T., Huisingh, D., & Almeida, C. M. (2019). A comprehensive review of big data analytics throughout product lifecycle to support sustainable smart manufacturing: A framework, challenges and future research directions. Journal of cleaner production, 210, 1343-1365.

[12] Wilberg, J., Schäfer, F., Kandlbinder, P., Hollauer, C., Omer, M., & Lindemann, U. (2017, December). Data analytics in product development: Implications from expert interviews. In 2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM) (pp. 818-822). IEEE.

[13] Wassan, S., Chen, X., Shen, T., Waqar, M., & Jhanjhi, N. Z. (2021). Amazon product sentiment analysis using machine learning techniques. Revista Argentina de Clínica Psicológica, 30(1), 695

[14] Jagtap, S., & Duong, L. N. K. (2019). Improving the new product development using big data: A case study of a food company. British Food Journal, 121(11), 2835-2848.

[15] Wilberg, J., Schäfer, F., Kandlbinder, P., Hollauer, C., Omer, M., & Lindemann, U. (2017, December). Data analytics in product development: Implications from expert interviews. In 2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM) (pp. 818-822). IEEE.