International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 02 February 2024
Page No: 169 - 177
https://irjaeh.com

A Comparative Study on Machine Learning Approaches for Sentiment Analysis

Mr. V. S. Raj kumar[1], Dr. T. Kumaresan[2]*

[1]Assistant Professor, Artificial Intelligence and Data Science, Bannari Amman Institute of Technology, Sathyamangalam, Erode, Tamilnadu, India.

[2]Associate Professor, Computer Science and Engineering, Bannari Amman Institute of Technology, Sathyamangalam, Erode, Tamilnadu, India.

Email id: *rajkumar.rbd@gmail.com[1], kumaresant@bitsathy.ac.in[2]*

****Corresponding Author Orcid ID:*** *0009-0005-8418-7451*

Abstract

Sentiment analysis plays a pivotal role in the operations of online product companies. User reviews are taken into account by others when they search for products, forming the cornerstone for delivering the right product based on user sentiments through sentiment analysis. Sentiment analysis involves the process of collecting, analyzing, and recommending reviews, which are often extensive and contain multiple paragraphs of content. This paper presents a comparative analysis of various machine learning models used to conduct sentiment analysis on customer reviews of Amazon products within the Electronics category. The initial models under scrutiny for our analysis include Logistic Regression, Decision Tree, Naive Bayes Classifier, Random Forest, Support Vector Machines, and BERT Model. The experimental result show that BERT classifier achieves higher accuracy when compare with other machine learning models.

Keywords: *Sentiment Analysis, Natural Language Processing (NLP), Product Reviews, Machine Learning.*

1. Introduction

Social media has seamlessly integrated into the daily routines of nearly everyone, wielding a profound impact on our interactions with the world. It plays a crucial role by offering a medium for individuals to voice their thoughts, feelings, and viewpoints, particularly in the context of products available on e-commerce websites. These expressions and evaluations function as a gauge for interpreting customer sentiment, which can range from enthusiastic endorsements to pointed criticisms regarding a specific product. In a dynamic and constantly evolving market where new products flood the scene regularly, online product companies increasingly depend on the wealth of customer feedback available on e-commerce platforms to navigate their selling decisions. These assessments are pivotal in shaping the decision-making process, directing potential buyers toward products that align with their needs and preferences. Consequently, they directly influence a product's success in the market. These customer evaluations cover a broad spectrum of product attributes and characteristics. They scrutinize the pricing, determining whether it offers good value for money. They consider the reputation and trustworthiness of the brand, a crucial factor in building consumer confidence. They also delve into the product's features and specifications, exploring what distinguishes it from the other competitors in the market. In essence, these reviews provide a comprehensive perspective on the product, considering its place within the market landscape. Our main objective is to extract valuable key performance indicators from this extensive reservoir of review data. By doing so, our aim is to identify and emphasize the reviews that are especially insightful and beneficial for online product companies. These reviews, brimming with valuable information and insights, serve as guiding lights for monitor consumers navigating the intricate realm of online shopping, aiding them in making well-informed decisions that align the consumers with their specific requirements and preferences.

2. Literature Review

Arwa S. M. AlQahtani [1] examined sentiment classification using several machine learning techniques and offers an analysis of the Amazon Reviews dataset. First, the reviews were converted into vector representation using a variety of techniques, including glove, bag-of-words, and Tf-Idf. Next, they trained a variety of machine learning algorithms, including Bert, Random Forest, Naïve Bayes, Bidirectional Long-Short Term Memory, and Logistic Regression. Then they used the Cross-Entropy Loss Function, Precision, Accuracy, F1-Score, and Recall to evaluate the models. Following that, the examined the sentiment classification of the best performing model by analyzing it. After conducting an experiment on multiclass classifications, is retrained the most effective model on the binary categorization. Najma Sultana, et.al [2] done a research on identifying the text as "Positive," "Negative," or "Neutral" by conducting sentiment analysis on customer reviews. In addition to discussing a theoretical approach to sentimental analysis, the paper analyses several algorithms for the same with comparable accuracies. It also provides a synopsis of earlier sentimental analysis methodologies. The paper's solution entails building an ML model in three main stages: data filtering, training, and testing. Data filtration entails utilizing only pertinent textual content for the model and pre-processing the text to exclude undesirable things. All of the characteristic words such as verbs, adverbs, and adjectivesare taken out and categorized throughout training. Using a dataset to train classification algorithms such as SVM, Linear Model, and Naive Bayes. Tanjim UI Haque & Nudrat Nawal Saber, et.al [3] conducted a study that uses the Amazon Review dataset for analysis, thereby rendering it as a perfect starting point. For reviews analysing process, they have mostly employed SVM and Multinominal Naive Bayesian classifiers. Considering the data that the model learns via active learning helps prevent bottleneck scenarios with unlabeled data. Because the authors compared the accuracy, precision, recall, and F1 score of several techniques (SVM, MNB, Stochastic Gradient Descent, Random Forest, etc.) for the supplied dataset, we found this work to be quite informative.

Roshan Pramod Samineedi Joseph [4] used reinforcement learning and a pre-trained BERT model, to categorize Amazon reviews into binary classes and specific multi-school categories. The researcher employed a product-based data gathering methods from Amazon's website. The goal of the research is to classify data as positive, negative, or neutral and to use algorithms such as BERT and LSM to anticipate and measure the sentiment behind the analysis. To get around the issue of long-term reliance, LSTM may be a particular kind of RNN. Because it propagates forward, it processes data that transmits information. Wanliang Tan, et.al [11] investigated the relationship between consumer ratings and product reviews on Amazon's website. They utilize both deep neural networks, such as Recurrent Neural Network (RNN), and standard machine learning techniques, such as Naive Bayes analysis, Support Vector Machines, and the KNearestNeighbor approach. We may be able to learn more about these algorithms by contrasting these outcomes. They could also support other techniques for detecting fraud scores. Ali Hasan, et.al [5] suggested a machine learning based sentiment analysis using mathematical and computational applications for twitter accounts. Customers have the opportunity to rate and review items on e-commerce platforms. This holds substantial sway on future buyers who wish to purchase the same item. As a result, it gives businesses the ability to examine opinion mining in sentiment analysis of reviews and ratings in order to track market pricing and product sales. Sentiment analysis was [6] done during the sentencing stage, and to help determine the polarity of sentiment, a sophisticated lexicon with preset positive and negative terms was employed.

3. Dataset

Amazon provides a space for small businesses and those with limited resources to expand their reach. Due to its widespread popularity, individuals invest time in crafting detailed reviews for both the brand and its products. Analyzing this wealth of information can offer valuable insights to companies, guiding them on product improvements. However, the sheer volume of data makes manual analysis impractical. Now, enter the realm of machine learning, specifically Natural Language [7]8 Processing (NLP), to tackle the challenge posed by massive datasets. The objective at hand is to forecast whether a given review leans towards the positive or negative spectrum. Considering the real dataset, which could comprise millions of reviews after scraping the website, we've undertaken the crucial task of preprocessing the data to facilitate the analysis.

For this investigation, we've employed the Amazon dataset for [10] analysis. The dataset encompasses key components such as customer details (User ID, Profile Name), product information (Product ID), and the reviews themselves (Score, Summary, Review Text, and Review Score). Notably, the Review Score serves as a labeled feature, categorizing reviews as positive and negative with values 1 and 0, respectively.

4. Data- Preprocessing

Text pre-processing plays a pivotal role in refining the quality of textual data in NLP, as illustrated in Figure 1 for the Amazon dataset in this study. The reviews underwent several pre-processing steps, starting with the conversion of all letters to lowercase to ensure uniformity (e.g., "Excellent" and "uSaBle" becoming "excellent" and "usable"). Punctuation and common stop words—those with minimal impact on meaning, such as "-, /, :, ?, the, a"—were subsequently removed. Following this, the reviews underwent tokenization, a process of breaking down sentences into sequences of words, or "tokens," separated by space characters. This tokenizing process relies on spaces for word separations [8] [9]. Finally, a lemmatization process was applied to revert all tokens to their base or dictionary form.

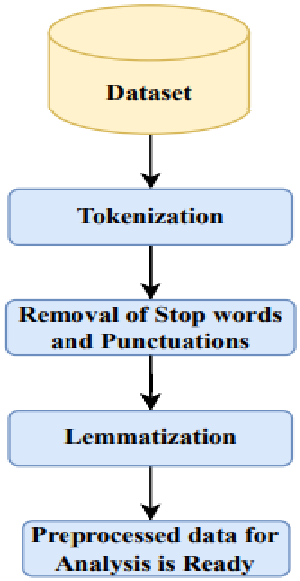

Figure 1 Pre-Processing Steps

After that, the dataset was partitioned into a 75% training set and a 25% validation set. This split aids in training the model on a substantial portion of the data while retaining a separate subset for validation purposes.

4.1. Feature Extraction

Natural Language Processing (NLP) requires computers to make sense of human language. The first step involves transforming textual data into a numerical format that can seamlessly integrate with machine learning models. This conversion facilitates the computational understanding of the nuances present in human language. The Figure 1 provided below illustrates the comparison of data distribution across different categories (Positive (1) and Negative (0)) of review sentiment. As indicated in Figure 2, it's evident that there are more reviews expressing positive sentiment compared to those with a negative tone.

To gain a clearer insight into the significance of words, let's generate Wordclouds for two categories: sentiment = 1 (positive) and sentiment = 0 (negative). This visual representation will highlight the prominent words in each sentiment category.

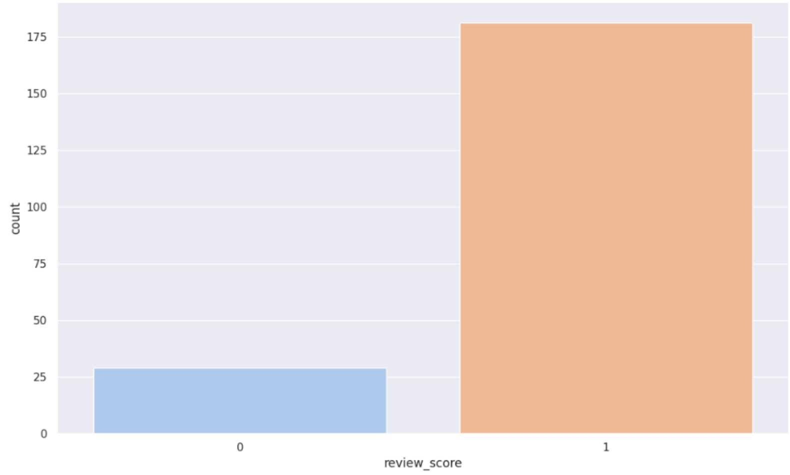

Figure 2 **Positive and Negative Review Comparison**

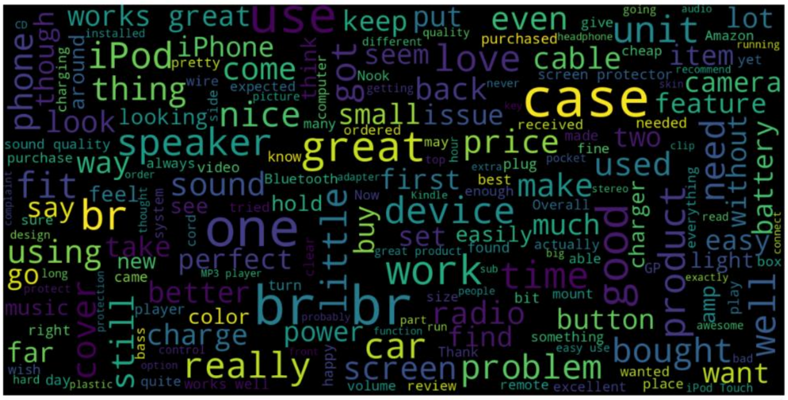

Figure 3 **Word Cloud for Positive Reviews**

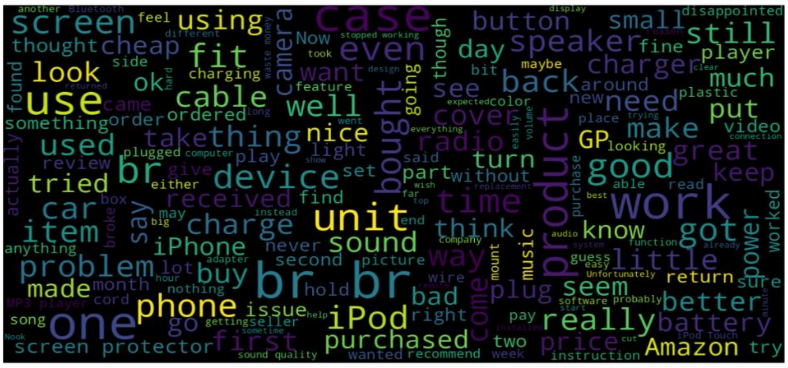

Figure 4 **Word Cloud for Negative Reviews**

5. Experimental Results

In each of the models implemented below, we have incorporated the selection of the top k highest-scoring features from the data for our model. Additionally, we proceeded to fine-tune the parameters of the previously executed models and evaluated their performance using both CountVectorizer and TF-IDF Tokenizer. [12]

5.1. Logistic Regression

Logistic Regression is a commonly utilized machine learning algorithm designed for binary classification tasks involving a two-class outcome variable. Despite its name, it is used for classification rather than regression. The algorithm employs the sigmoid function to map a linear combination of input features to probabilities within the range of 0 to 1. During training, the model learns coefficients, and predictions are made by applying a threshold, usually set at 0.5, to the predicted probabilities. Logistic Regression is appreciated for its simplicity, interpretability, and effectiveness in applications such as spam detection and medical diagnosis. The Table 1 given shows the experimental result for logistic regression model for both training set and test set. [13]

Table 1 Logistic Regression Result

		0	1	Accuracy	macro avg	weighted avg
Train Result	precision	0.853	0.854	0.854	0.854	0.854
	recall	0.912	0.763	0.854	0.839	0.854
	f1-score	0.881	0.810	0.854	0.845	0.852
	support	11282	7467	0.854	18749	18749
Test Result	precision	0.820	0.813	0.820	0.820	0.816
	recall	0.890	0.707	0.820	0.804	0.824
	f1-score	0.852	0.757	0.820	0.801	0.813
	support	3718	2532	0.820	6250	6250

5.2. Decision Tree Classifier

The Decision Tree Classifier is a versatile algorithm for classification and regression. It constructs a tree where nodes represent decisions based on features, and leaves indicate class labels. Known for interpretability, it uses metrics like Gini impurity for recursive dataset splits. To [14] Prevent over fitting, methods like Random Forests or Gradient Boosting are often used. Table 2 presents the experimental results for the Decision Tree Classifier model on both the training and test sets. Word Cloud for positive reviews in Figure 3.

Table 2 Decision Tree Classifier Result

		0	1	Accuracy	macro avg	weighted avg
Train Result	precision	0.998	0.993	0.995	0.994	0.995
	recall	0.995	0.996	0.995	0.996	0.995
	f1-score	0.997	0.991	0.995	0.991	0.995
	support	11282	7467	0.995	18749	18749
Test Result	precision	0.734	0.600	0.679	0.668	0.681
	recall	0.722	0.626	0.679	0.667	0.678
	f1-score	0.726	0.615	0.679	0.667	0.679
	support	3718	2532	0.679	6250	6250

5.3. Naive Bayes Classifier

Gaussian Naive Bayes (Gaussian NB) is a classification algorithm for tasks with continuous features assumed to have a Gaussian distribution. It calculates probabilities based on feature independence given the class label. Quick and simple, it's used in applications like spam filtering, But performance may be impacted if assumptions aren't met. Table 3 displays the experimental outcomes for the Gaussian Naive Bayes (Gaussian NB) model on both the training and test sets. Word cloud for Negative review in Figure 4.

Table 3 Gaussian Naive Bayes (Gaussian NB) Result

		0	1	Accuracy	macro avg	weighted avg
Train Result	precision	0.857	0.677	0.768	0.765	0.782
	recall	0.742	0.811	0.768	0.776	0.768
	f1-score	0.791	0.735	0.768	0.766	0.772
	support	11282	7467	0.768	18749	18749
Test Result	precision	0.790	0.627	0.714	0.711	0.728
	recall	0.702	0.730	0.714	0.725	0.725
	f1-score	0.746	0.679	0.714	0.711	0.728
	support	3718	2532	0.714	6250	6250

5.4. Random Forest Classifier

Random Forest Classifier is an ensemble algorithm for classification and regression, constructing multiple trees during training and combining their predictions for improved performance. It features random selection of features, bagging, and versatility with various data Types, and provides insights into feature importance. Table 4 presents the experimental results, showcasing the performance of the Random Forest Classifier model across both the training and test datasets. [15]

Table 4 Random Forest Classifier Result

		0	1	Accuracy	macro avg	weighted avg
Train Result	precision	0.997	0.992	0.994	0.994	0.991
	recall	0.996	0.992	0.994	0.996	0.991
	f1-score	0.997	0.995	0.996	0.996	0.991
	support	11282	7467	0.996	18749	18749
Test Result	precision	0.795	0.781	0.791	0.783	0.788
	recall	0.872	0.669	0.791	0.771	0.794
	f1-score	0.832	0.721	0.791	0.775	0.787
	support	3718	2532	0.791	6250	6250

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 02 February 2024
Page No: 169 - 177
https://irjaeh.com

5.5. BERT (Bidirectional Encoder Representations from Transformers) Algorithm

BERT, designed by researchers at Google AI Language, achieves state-of-the-art results across various NLP tasks. It utilizes the Transformer architecture to learn contextual relationships between words or sub-words in a text. Unlike [16] Directional models that read text sequentially, BERT processes the entire word simultaneously with the help of Transformers. This approach enhances its ability to capture contextual information and improves performance in a wide range of natural language processing tasks. The Table 5 given shows the experimental result for BERT model for both training set and test set.

Table 5 BERT Model Result

		0	1	Accuracy	macro avg	weighted avg
Train Result	precision	0.991	0.994	0.991	0.992	0.997
	recall	0.992	0.993	0.993	0.991	0.995
	f1-score	0.991	0.994	0.995	0.992	0.993
	support	11282	7467	0.995	18749	18749
Test Result	precision	0.774	0.647	0.687	0.661	0.684
	recall	0.744	0.638	0.684	0.663	0.683
	f1-score	0.743	0.627	0.684	0.665	0.684
	support	3718	2532	0.683	6250	6250

6. Result and Discussion

After identifying the optimal hyper parameters for each model, we conducted a comparison to assess their relative performance. This analysis helps us determine which model is worth further [17]. Exploration in our subsequent studies. The comparative Analysis of Sentiment Analysis across distinct Machine Learning Models is presented in Table 6. The following Figure (Figure 5) illustrates the conclusive outcomes of all models employed in this research.

Table 6 Comparative Analysis on different ML Models for Sentiment Analysis

Model	Accuracy
Logistic Regression	81.60%
Decision Tree Classifier	85.87%
Gaussian Naive Bayes (Gaussian NB)	73.78%
Random Forest Classifier	87.12%
BERT Model	92.45%

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 02 February 2024
Page No: 169 - 177
https://irjaeh.com

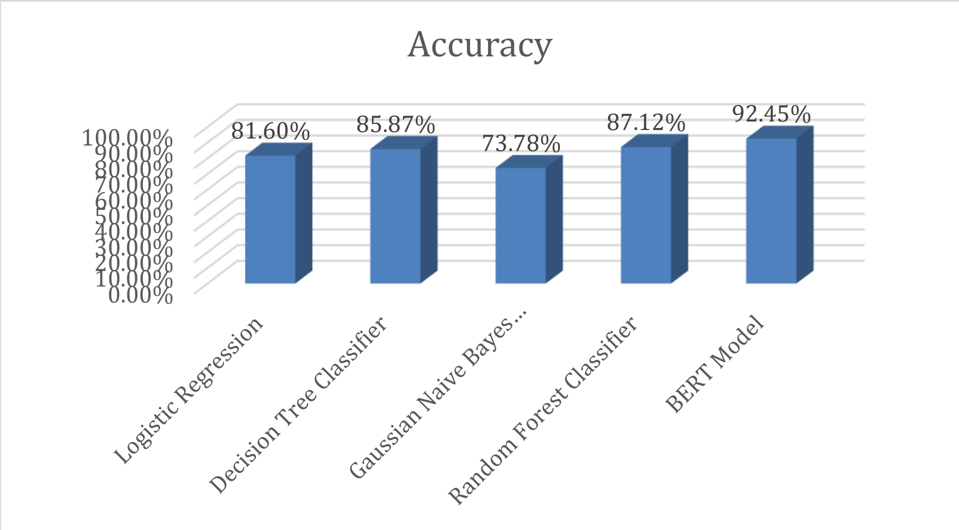

Figure 5 Graphical Representation of Final Result

Conclusion

Through multiple model iterations and rigorous testing, the BERT classifier emerged as the most effective in estimating sentiment, boasting an impressive accuracy of above 90%. While our testing and analysis were conducted at a foundational level, we foresee significant utility across various domains, particularly in product and user relationship analysis. One compelling application lies in recommendation systems. Leveraging BERT's sentiment analysis, users can be efficiently clustered based on their similar reviews on platforms like Amazon. This not only enhances the precision of recommendations but also contributes to a more personalized and user-centric experience in the realm of online product evaluations. This project utilizes various machine learning models for sentiment analysis, including fine-tuning the BERT model on Amazon customer reviews. Future plans involve integrating word2vec, exploring alternative classifiers (e.g., SVM, GRU), and extending the analysis to broader Amazon customer reviews.

References

[1] Arwa S. M. AlQahtani, "Product sentiment analysis for Amazon reviews," International Journal of Computer Science & Information Technology (IJCSIT) Vol 13, No 3, June 2021.

[2] Najma Sultana, Pintu Kumar, Monika Rani Patra, Sourabh Chandra and S.K. Safikul Alam," Sentiment analysis for product review," ICTACT journal on soft computing, Vol 09, Issue 03, April 2019.

[3] Haque, Tanjim & Saber, Nudrat & Shah, Faisal," Sentiment analysis on large scale Amazon product reviews," IEEE international conference on Innovative Research and development (ICIRD), 10.1109/ICIRD.2018.8376299, June 2018.

[4] Dublin, Griffith & Joseph, Roshan, "Amazon Reviews Sentiment Analysis: A Reinforcement Learning Approach," . 10.13140/RG.2.2.31842.35523, 2020.

[5] Ali Hasan,Sana Moin,Ahmad Karim, Shahaboddin Shamshirband,"Machine Learning-Based Sentiment Analysis for Twitter Accounts. Mathematical and Computational Applications," Vol 23, Issue 1, 2018.

[6] S. A. a. A. N. S. Aljuhani, "A Comparison of Sentiment Analysis Methods on Amazon Reviews of Mobile Phones," International Journal of Advanced Computer Science and Applications, vol. 10, 2019.

[7] X. C. T. S. M. W. N. J. Sobia Wassan, “Amazon Product Sentiment Analysis using Machine,” Revista Argentina de Clínica Psicológica, pp. 695-703, 2021.
[8] Bahrawi, “Sentiment Analysis Using Random Forest Algorithm-Online Social Media Based,” Journal of information technology and its utilization, vol. 2, pp. 29-33, 2019.
[9] M. a. S. R. Fikri, “A Comparative Study of Sentiment Analysis using SVM and SentiWordNet,” Indonesian Journal of Electrical Engineering and Computer Science, vol. 13, pp. 902-909, 2019.
[10]Y. a. K. V. Saito, “Classifying User Reviews at Sentence and Review Levels Utilizing Naïve Bayes,” 21st International Conference on Advanced Communication Technology (ICACT), PyeongChang Kwangwoon_Do, Korea (South), 2019.
[11]https://www.semanticscholar.org/paper/Analysis-for-Amazon-Reviews-Tan-wang/06babbc8d4b5dce8d6154f74b75c83cbc1f16de7?utm_source=direct_link http://jmcauley.ucsd.edu/data/amazon/
[12]Islam, Ariful, et al. "Latex reinforced waste buffing dust-jeans cotton composites and its characterization." Journal of Polymer Research 28.8 (2021): 322.
[13]Jaganathan, Saravana Kumar, et al. "Enriched physicochemical and blood-compatible properties of nanofibrous polyurethane patch engrafted with juniper oil and titanium dioxide for cardiac tissue engineering." International Journal of Polymer Analysis and Characterization 24.8 (2019): 696-708.
[14]Moganapriya, C., et al. "Sustainable hard machining of AISI 304 stainless steel through TiAlN, AlTiN, and TiAlSiN coating and multi-criteria decision making using grey fuzzy coupled taguchi method." Journal of Materials Engineering and Performance 31.9 (2022): 7302-7314.
[15]Kaliyannan, Gobinath Velu, et al. "Mechanical and tribological behavior of SiC and fly ash reinforced Al 7075 composites compared to SAE 65 bronze." Materials Testing 60.12 (2018): 1225-1231.
[16]Moganapriya, Chinnasamy, et al. "Tribomechanical behavior of TiCN/TiAlN/WC-C multilayer film on cutting tool inserts for machining." Materials Testing 59.7-8 (2017): 703-707.