ISSN: 2633-4828 Vol. 5 No.4, December, 2023

International Journal of Applied Engineering & Technology

IMPACT OF ATMANIRBHAR BHARAT ABHIYAN ON INDIAN BUYER'S MINDSET: A LEXICON SENTIMENT ANALYSIS, SUPERVISED MACHINE LEARNING AND HYBRID ENSEMBLE LEARNING APPROACH

Beena Kapadia[1] and Dr. Amita Jain[2]
[1]Research Scholar and [2]Research Associate, School of Computer Application and Technology, Career Point University, Kota, Rajasthan, India
[1]Beena.v.kapadia@gmail.com and [2]Amita.jain01@gmail.com
[1][0000-0002-5064-3480] and [2][0000-0002-1118-3389]

ABSTRACT
The Atmanirbhar Bharat Abhiyan initiative, launched to promote self-reliance and local manufacturing in India that has generated interest in understanding its impact on Indian consumers mindset towards Indian products. This research paper employs lexicon-based sentiment analysis, using the TextBlob, SpacyTextBlob and VADER sentiment analysis tools on the 'amazon_vfl_reviews.csv' dataset to investigate sentiments around products to compare sentiment scores before and after the implementation of the Atmanirbhar Bharat Abhiyan to identify changes in consumer perceptions. Also, this paper searched for topics such as 'Indian', made in India,' 'Indian Brand,' and 'Indian Product in the review text of the csv file taken from Kaggle. This study also employes the same analysis on another dataset created from respondents to find the popularity of Indian products due to "Atmanirbhar Bharat Abhiyan. Additionally, external factors such as media coverage, government policies, and market trends are considered to explore their correlation with shifts in product reviews. The analysis aims to uncover any causal relationship between the Atmanirbhar Bharat Abhiyan's announcements and changes in sentiment towards Indian products.

This paper applies various Machine Learning algorithms and the hybrid ensemble learning approach using the training and test ratio as 70:30 to find the accuracy of training the model with lexicon sentiment analysis using TextBlob, SpacyTextBlob and VADER tools. According to the related work using .csv file fetched from Kaggle, the popularity of Indian Brands increased after implementation of Atmanirbhar Bharat Abhiyan by 3.5 %. According to respondents' dataset, the popularity of Indian Brands increased after implementation of Atmanirbhar Bharat Abhiyan is 16.39 % on an average for all products into consideration taken from respondents for expensive, inexpensive and handcrafted products. The accuracy of Hybrid ensemble mode using TextBlob, SpacyTextBlob and VADER are 0.9026, 0.8959 and 0.9206 respectively.

*Index Terms**:** Atmanirbhar Bharat Abhiyan, self-reliance, media coverage, Indian consumers, government policies, Machine Learning algorithms, the hybrid ensemble learning, TextBlob, SpacyTextBlob, VADER*

1. INTRODUCTION
India and the entire world have suffered a lot due to the outbreak of pandemic and the resultant lockdown in 2020. People suffered on the health front as well as on economic front, which affected badly on the whole world economically resulting in fall of GDP of all major countries and a large-scale loss of human life in the world.

Our PM Narendra Modi came up with a call of 'Atmanirbhar Bharat' on 12th May 2020, which means Self-reliant India or self-sufficient India. The basic idea behind this concept is making India self-generating economy. The short-term goals of this program are economy, infrastructure, technology driven systems, vibrant demography, and demand.

Various slogans initiated under Atmanirbhar Bharat Abhiyan including 'Vocal for Local', 'Local for Global' and 'Make for World'. This paper focusses on finding out the popularity of Indian Brands and the mindset of Indians after Atmanirbhar Bharat Abhiyan (ABA) in the view of Vocal for Local. Self-reliant program would help in reviving the sectors of Indian economy in short term goal of this program and in long term it will build capacities and make the country strong enough to face unprecedented situations like COVID. The program emphasizes production locally.

In India, employment generation has always been a major concern for central and state government both, in public sector as well as in private sector. Self-reliant India can be a boon for young India. Besides encouraging entrepreneurship, the program will create jobs for all kinds of people – skilled or unskilled.

Cutting-edge technology like Artificial Intelligence (AI), Machine Learning (ML), Robotics, Deep Learning, Data Science, and Cloud Computing have become important elements of planning, production, and services.

Natural Language Processing is a branch of Artificial Intelligence that deals with the interaction between systems and humans using natural language. The objective of Natural Language Processing is to read, decode, understand, and make sense to perform desired task. The more data you collect, the more you can correct your algorithm's mistakes and reinforce its correct answers.

Sentiment analysis is the branch of NLP to find the opinion of the people as positive, negative or neutral from the given text. lexical sentiment analysis is a valuable technique for analyzing sentiment in text data, offering simplicity, efficiency, and language-independence. While it has limitations, such as difficulty handling context and sarcasm, lexical sentiment analysis remains a widely used approach in sentiment analysis tasks.

This study finds whether the Indian Brands are popular, and the popularity is increased or not amongst Indians after the implementation of AtmaNirbhar Bharat Abhiyan based on lexical sentiment analysis and various supervised Machine Learning Algorithm including hybrid ensembled machine learning approach.

2. LITERATURE REVIEW:

(goelyash.in, May 2020) The five pillars of self-reliant India are Economy, Infrastructure, System, Demography and Demand. (Bajaj, Bhumika, and Sudhir Narayan Singh, 2022) This study explores how the Atmanirbhar Bharat initiative aims to boost the economy by promoting self-reliance and local manufacturing. It emphasizes the revival of the economy during the pandemic and the impact on various sectors. (Upender Sethi, 2022) Atmanirbhar Bharat Abhiyan is a new version of 'Make in India' which was announced by the Hon'ble Prime Minister on 12 May 2020 with a new perspective. The government announced an economic recovery package of Rs 20 lakh crore (US$268.74 billion) and big-bang systemic reforms under Atma Nirbhar Bharat Abhiyan (independent India). The aim is to make the country and its citizens independent and self-reliant in all senses. This initiative has had significant implications for consumer sentiments and market dynamics across different industries in India.

Tubishat et al. conducted a systematic review of 45 research articles based on implicit extraction in sentiment analysis. Their work focuses on interdisciplinary knowledge acquisition and production.

(Ghosal, 2021) The Atmanirbhar Bharat Abhiyan sought to reduce dependence on imports and boost domestic production and consumption of goods and services. As a result, there has been a noticeable shift in consumer sentiments, with a growing preference for domestically manufactured products and services. Consumers are increasingly inclined towards supporting local brands and products, driven by patriotic sentiments and the desire to contribute to the nation's economic growth.

(The Economic Times, 2021) Furthermore, the Atmanirbhar Bharat Abhiyan has influenced market dynamics by fostering the growth of indigenous industries and startups. Various sectors, including manufacturing, agriculture, technology, and healthcare, have witnessed increased investments and innovation to meet the objectives of self-reliance. Domestic companies are leveraging government incentives and initiatives to enhance their competitiveness and expand their market presence.

(Business Standard, 2021) Additionally, the government's emphasis on promoting Atmanirbhar Bharat has led to changes in trade policies and regulations, encouraging domestic manufacturing and reducing barriers to entry for local businesses. This has resulted in a more conducive environment for entrepreneurship and business expansion, leading to a vibrant and dynamic marketplace.

In conclusion, the Atmanirbhar Bharat Abhiyan has had a profound impact on consumer sentiments and market dynamics in India. By fostering self-reliance and promoting domestic production, the initiative has instilled a sense of confidence among consumers and catalyzed growth and innovation across various sectors of the economy.

(Singh, G., Kumar, B., Gaur, L., & Tyagi, A., 2019) Multinomial Naïve Bayes classifier works on the concept of term frequency (TF). TF means the number of times the word occur in a document. Multinomial Naïve Bayes is used to find two facts that whether the word exists in a document or not as well as that words frequency in that document.

(Borg, A., & Boldt, M., 2020) Support Vectors are simply the coordinates of individual observation. It is a classifier and is a frontier which best segregates the classes. In SVM, it is required to select the hyper-plane which segregates the two classes better. If we can't find linear hyperplane between the two classes, then it is required to introduce additional feature.

(Singh, S. N., & Sarraf, T., 2020) Random Forest algorithm is a supervised machine learning algorithm used for classification. It works based on the concept of ensemble learning, in which number of decision trees get various subsets of the data set. Here, all decision trees predict the output for the new data and the final class of that new data is considered as most of the outcome predicted from all decision trees.

(Irawaty, I., Andreswari, R., & Pramesti, D., 2020) The K Nearest Neighbor algorithm takes all the given data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a category, which is most suitable. That is, this algorithm at the training phase just stores the dataset and when it gets new data, then it classifies that data into a category that is much like the new data.

This algorithm first selects the number K of the neighbors (to get the exact k-value we need to test the model for each expected k-value.). Then, it calculates the minkowski distance of K number of neighbors. After that, it takes the k Nearest Neighbors as per the calculated minkowski distance. Among the identified k neighbors k neighbors, it counts the number of the data points in each category. It then assigns the new data points to that category for which the number of the neighbor is maximum.

(Hutto, C.J., & Gilbert, E., 2014). VADER is abbreviation for "Valence Aware Dictionary and sEntiment Reasoner" and is available under the MIT License. The VADER tool was released in 2014. It uses a lexicon driven approach and additional heuristics for rating the input. It offers consistent ratings and requires no training data, as VADER is not a machine learning approach. It achieved some remarkable scores for multiple domains such as tweets, movie or product reviews.

VADER is a rule-based sentiment analysis tool to express the sentiments on given text. Yang, S. E. (n.d.). It is used to label the dataset into positive or negative sentiment score, based on whether its value exceeds 0.5 or not. (V. Bonta, 2019) The compound score can be calculated as the sum of all lexicon ratings which are normalized ones. VADER works better than TextBlob for the text, taken from either social media or any web sources.

Lexical sentiment analysis, also known as lexicon-based sentiment analysis, is a fundamental approach in natural language processing (NLP) that aims to analyze the sentiment or opinion expressed in text data. Unlike machine learning-based approaches that require labeled training data, lexicon-based methods rely on pre-defined sentiment lexicons or dictionaries containing words annotated with their associated sentiment polarity (positive, negative, or neutral).

In lexical sentiment analysis, each word in the text is assigned a sentiment score based on its presence in the sentiment lexicon. The sentiment score can be a continuous value representing the intensity of sentiment or a discrete value indicating the polarity (positive, negative, or neutral). By aggregating the sentiment scores of individual words, the overall sentiment of the text can be determined.

One of the key advantages of lexical sentiment analysis is its simplicity and efficiency. Since it does not require training on labeled data, lexicon-based methods can be quickly applied to analyze sentiment in large volumes of text data. Additionally, lexical sentiment analysis is language-independent and can be adapted to different languages by using language-specific sentiment lexicons.

However, lexical sentiment analysis also has its limitations. Sentiment lexicons may not capture context-dependent tones or sarcasm, leading to inaccuracies in sentiment analysis. Moreover, lexicon-based methods may struggle with out-of-vocabulary words or domain-specific terminology not covered by the sentiment lexicon.

Several popular sentiment lexicons have been developed for lexical sentiment analysis. (Hutto & Gilbert, 2014) One of the most widely used lexicons is the VADER (Valence Aware Dictionary and sEntiment Reasoner) lexicon, which is specifically designed for social media text and incorporates both unigram and bigram sentiment scores along with punctuation and capitalization rules. (Wilson, 2022; Tomažič et al., 2021) Other lexicons include TextBlob and SpaCyTextBlob, which are Python libraries that offer easy-to-use interfaces for sentiment analysis, providing accurate sentiment polarity scores for text data.

3. RESEARCH GAP AND RESEARCH OBJECTIVE:

Previous studies have extensively explored sentiment analysis in the context of government initiatives and its impact on consumer behavior. These studies have provided valuable insights into understanding how government policies influence public sentiment and consumer attitudes.

For instance, research by Singh et al. (2019) examined the sentiment of social media discussions related to government initiatives such as demonetization and Goods and Services Tax (GST) in India. The study found that sentiment varied across different segments of society and was influenced by factors such as economic implications, political affiliations, and media coverage.

Similarly, a study by Chen et al. (2020) investigated consumer sentiment towards environmental policies implemented by the Chinese government. The research utilized sentiment analysis techniques to analyze social media data and found that positive sentiment towards environmental initiatives positively correlated with consumer trust in the government and willingness to support eco-friendly products.

Furthermore, research by Li et al. (2018) focused on sentiment analysis in the context of healthcare policies in the United States. The study examined public sentiment towards the Affordable Care Act (ACA) using text mining and sentiment analysis of online discussions. The findings revealed a complex interplay of positive and negative sentiments influenced by factors such as political ideology, personal experiences with healthcare, and media framing.

Overall, previous studies have demonstrated the utility of sentiment analysis in understanding public perception and consumer behavior in response to government initiatives. By leveraging sentiment analysis techniques, researchers have been able to uncover valuable insights that inform policymakers and stakeholders about the effectiveness and impact of various government policies.

But in our opinion and best of the studies and knowledge none of the studies have been observed on sentiment analysis for Atmanirbhar Bharat Abhiyan or Vocal for Local policy launched by Government.

i. To find out the change in the Indian buyers' mindset, because of the role of "Atmanirbhar Bharat Abhiyan" in promoting the local products.

ii. To find out the popularity of Indian products due to "Atmanirbhar Bharat Abhiyan".

4. BACKGROUND AND RELATED WORK:

Figure I shows a process approach overview to get the overall idea of this study.

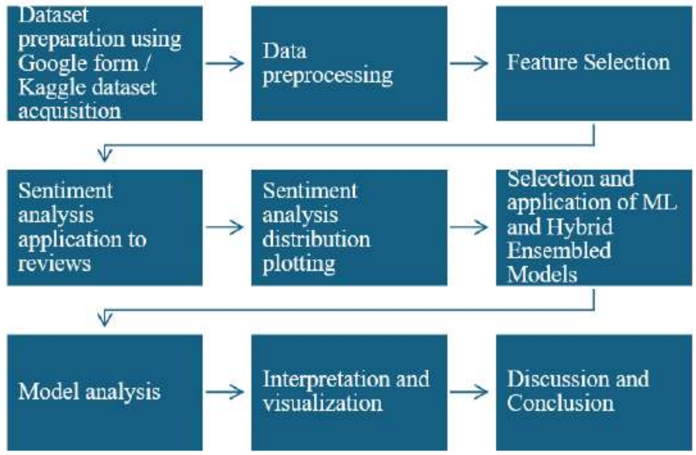

Figure I: A Process Approach Overview Using Amazon_Vfl_Reviews.Csv and Respondents' Data

A. Dataset Preparation Using Google form / Kaggle Dataset Acquisition:

There are two Datasets Taken in this Study.

The first dataset has been taken for the first research objective, which is 'amazon_vfl_reviews.csv' dataset from the website https://www.kaggle.com/code/abhishek1mohapatra/amazon-review-classification-using-nlp/comments having 2772 data in the dataset. It is observed that all brands in this dataset is not the Indian Brands. So, after vigorous research on all Brands given in the dataset, the data has been filtered only for the Indian Brands. We have made two separate files for two different studies – i) popularity of Indian Brands before the Atmanirbhar Bharat Abhiyan has been announced, to which the excel file prepared as amazon_vfl_reviews_Indian_Brands_before_12May_2020_Filtered.csv and ii) popularity of Indian Brands after the Atmanirbhar Bharat Abhiyan has been announced, to which the excel file prepared as amazon_vfl_reviews_Indian_Brands_after_12May_2020_Filtered.csv. The keyword Filtered at the end kept for both these files as some brands were common to both time zones – before 12th May 2020 and after 12th May 2020 but not all. So, whichever common brands are available, only that have considered. Here we have applied topic search for searching the exact sentiment on the topic 'Indian', 'made in India' or 'Indian products' kinds of topics so stemming, and lemmatization not applied for this study. TFIDF and removal of stop words are used at machine learning algorithms just before splitting the dataset into training and test data. Here, the training versus test ratio is taken as 70:30.

The other dataset is the primary dataset which is collected through a google form. This dataset consists of 696 respondents' data collected across 49 cities distributed across 8 states of India – Bihar, Delhi, Gujarat, Haryana, Karnataka, Maharashtra, Rajasthan and Tamil Nadu of India for various products. The objective is to study the difference in buying behavior due to Atmanirbhar Bharat Abhiyan. For that first we asked about whether the respondent is aware of Atmanirbhar Bharat Abhiyan (ABN) or not. We have considered only those respondents who are aware of ABA. A total of 656 data have been considered for those who were aware of ABN and 40 data we have discarded. We have considered five expensive Indian brands and five inexpensive Indian brands in our questionnaire to understand whether any rise in the popularity after ABA or not. The five expensive brands are taken into consideration like – watches, shoes, clothing, automobiles and Jewelry and five less expensive items what we use in our day-to-day life are considered as – pen/pencil, lipstick, tea-coffee, soap-handwash-facewash, and shampoo-hair cleanser. For all these products the question is asked to choose one of the options amongst Preferable even before ABA, not preferable before pandemic but started buying after ABA and never preferred this brand and not sure about future preference. And the last column is the review about Atmanirbhar Bharat

Abhiyan. Generally, the products we use in day-to-day life are considered for the study. In this study some expensive Indian brands are considered like Allen Solly, Provogue, Biba, GINI & JONY, Van Heusen for clothes, Titan, Fastrack, Sonata for watches, Bajaj, TVS, Tata, Mahindra for automobiles, Tanishq, Kalyan Jewellers, Waman Hari pethe, Nakshatra Jewellers for Jewellery and Campus, Liberty, Woodland, Lakhani, Sparx for shoes. There are some inexpensive Indian Brands also included in this study like Hamam, Santoor, Lifebuoy, khadi for soap, Khadi, Ayur, Patanjali, Indulekha, Dabur for shampoo / hair cleanser, Cello, Camlin, Flair, Natraj for pen / pencil, Nykaa, Biotique, Lakme, Elle18 for Lipstick and Organic India, Wagh Bakri Tea, Tata, Tulsi Tea for Tea or coffee.

Also, study has been conducted on other Indian products like Diya, Colours, Lights, handicrafts and clothes made by Hathkargha Udyog (handloom). Further study is carried out to find any change in respondent's buying behavior during and after ABA- like buying Indian masks only or tried to buy as much Indian products as possible, which respondent had not thought about it ever before pandemic.

B. Data Preprocessing:

There are two datasets we have used – the first one is amazon_vfl_reviews.csv which is available on Kaggle, and another one is created from the respondents' data. The amazon_vfl_reviews.csv file of Kaggle is having total 2782 rows but having some non-Indian brands reviews also like - Coca-Cola (American brand), Mazza (American brand), Maggi (Swiss company Nestlé), Dettol (United Kingdom), and Savlon (British brand). After filtration of these non-Indian Brands, we got total 2256 rows.

Further, for respondents' data, some respondents were unaware about Atmanirbhar Bharat Abhiyan. Hence, those data are filtered from that file and worked on 656 data out of 696.

This study drops any row containing the missing text in the 'review' column of csv file. Also, it converts any other datatype to the string data type for applying the sentiment analysis. Various data preprocessing techniques applied to clean the text like replaced non alphabets to blank space, converted each sentence (review) split into different words (list), removed stop words from the word list, joined the words to frame sentence again and squeezed number of spaces into a single space. We have also used Bag of words and Term Frequency-Inverse Document Frequency technique to convert text into numbers. Both are used for converting text to numbers or vectors from text as any classification algorithms like multinomial Naïve Bayes algorithm, support vector machine algorithm, random forest algorithm etc. takes numbers only as input for classification.

C. Feature Selection

As mentioned earlier, we have made two files from amazon_vfl_reviews.csv as amazon_vfl_reviews_Indian_Brands_before_12May_2020_Filtered.csv and amazon_vfl_reviews_Indian_Brands_after_12May_2020_Filtered.csv, in which only Indian Brands are considered and also, we have checked the common brands in both the files. We have further filtered any uncommon brands in any of the files.

From both the files only the review column is used out of all four rows- asin(product code), name, date and review.

Also 10 different files have been generated from the single respondents' file – one for each product. The five expensive brands are taken into consideration like – watches, shoes, clothing, automobiles and Jewelry and five less expensive items what we use in our day-to-day life are considered as – pen/pencil, lipstick, tea-coffee, soap-handwash-facewash, and shampoo-hair cleanser. The names of the files are kept as product_name.csv as per all ten products. Each product can have 4 to 5 brands of varieties of Indian Brands to choose from in the questionnaire and for each variety, 3 radio buttons have been kept choosing from as – 'Preferable even before Atmanirbhar Bharat Abhiyan', 'Not preferable before pandemic but started buying after Atmanirbhar Bharat Abhiyan', and 'never preferred this brand and not sure about future preference'.

D. Sentiment Analysis Application and Plotting:

SentimentIntensityAnalyzer() is used on amazon_vfl_reviews.cvs file by importing SentimentIntensityAnalyzer from nltk.sentiment.vader to apply the sentiment analysis on ‘review before and after the launch of ABA. Figure II shows the DataFrame with Sentiment Scores before implementation of 'Atmanirbhar Bharat Abhiyan'. The TextBlob is imported from textblob to create the object of textblob using its constructor with a parameter as review text. On that object sentiment.polarity is called to find its polarity row wise. For SpacyTextBlob working, the SpacyTextBlob class is imported from spacytextblob.spacytextblob to call spacy.load("en_core_web_sm"). add_pipe("spacytextblob") to create the object named nlp. Another doc object has been created using nlp(review). Finally, doc._.polarity has been called to generate the sentiment score using SpacyTextBlob.

	asin	name	date	rating	review	sentiment_score
0	B07QSQCF9R	Mamaearth-Moisturizing-Baby-Bathing-Oatmeal	6/19/2019	5	Mamaearth is always best. Every product of th...	0.9864
1	B07QSQCF9R	Mamaearth-Moisturizing-Baby-Bathing-Oatmeal	7/23/2019	1	Update: horrible bathing that I have used so f...	-0.9052
2	B07QSQCF9R	Mamaearth-Moisturizing-Baby-Bathing-Oatmeal	6/23/2019	3	Although, it makes a very rich and creamy lath...	0.9851
3	B07QSQCF9R	Mamaearth-Moisturizing-Baby-Bathing-Oatmeal	6/22/2019	1	Very bad product worth ZERO star. The entire s...	0.9389
4	B07QSQCF9R	Mamaearth-Moisturizing-Baby-Bathing-Oatmeal	4/29/2019	5	I really have started appreciating this brand ...	0.9892

Figure II: Dataframe with Sentiment Scores before Implementation of 'Atmanirbhar Bharat Abhiyan' Applied on Amazon_Vfl_Reviews.Csv File

A new column ‘sentiment_scores’ is created to store the compound sentiment of each review. Finally, a new csv file is created with the name amazon_sentiments.csv to have the fifth column – ‘sentiment_score’, which can be any fraction number from -1 to 1. The sentiment score has been converted to the whole number -1, 0 or 1 to work further for Machine Learning and hybrid ensemble model, where x is the review and y is the sentiment_score, which is -1, 0 or 1. -1 indicates the negative sentiment, 0 indicates the neutral sentiment and 1 indicates the positive sentiment. These ‘sentiment_score’ is created separately with TextBlob, SpacyTextBlob and VADER sentiment Analysis tool based on lexical sentiment analysis. Further a pie chart has been generated to show negative, neutral and positive sentiments.

The study has been made also to check whether any of the strings from - 'made in India', 'Indian Brand', 'Indian Product', 'Local Product', 'Vocal for Local', 'Atmanirbhar Bharat Abhiyan' exists in the review or not for the data after the launch of ABA as shown in figure III. We could find that there are the two terms - 'made in India', 'Indian Brand' are available two times in this review.

```
Occurrences of 'indian': 12

Occurrences of 'made in india': 8

Occurrences of 'indian brand': 8

Occurrences of 'indian product': 0

Occurrences of 'local product': 0

Occurrences of 'vocal for local': 0

Occurrences of 'atmanirbhar bharat abhiyan': 0
```

Figure III: Outcome of the Study to Check Whether Any of The Strings from - 'Made in India', 'Indian Brand', 'Indian Product', 'Local Product', 'Vocal for Local', 'Atmanirbhar Bharat Abhiyan' Exists in the Review or Not, Which is Applied on Amazon_Vfl_Reviews.Csv File

The Plotting of sentiment analysis for the popularity of Indian brands before and after is shown in the figure VIII a and b respectively.

Considering now the respondents’ data, the output of applying VADER sentiment algorithm, without customized VADER sentiment analysis is shown in figure IV. Where in the 0th row, the soap brand Hamam, Santoor and Lifebuoy are preferable by respondent, hence, its sentiment should be positive; even though in Anyother column '-

' is written. Here, we have applied the condition, which states that if respondent uses any of the four given Indian Brands, then '-' or blank or any punctuation should be replaced with yes; and if all four not selected then the punctuation or blank space should be replaced with 'no'. In rows 2 and 4, the other Indian Brand name is mentioned – 'Hamam', 'Santoor', 'Lifebuoy' and 'Khadi' respectively, so the respondent uses Indian brand only, but the other Indian brand. So, ideally sentiment should be positive. But as the brand name doesn't show any sentiment in general, the sentiment score using VADER sentiment analysis is neutral. Also, in row 3, respondent has conveyed 'Any other local brand that compete any global brands', also has positive sentiment towards buying Indian soap brand. But the sentiment score for the same is 0.0, that is neutral.

	Hamam	Santoor	Lifebuoy	khadi	Anyother	MAnyother	sentiment_score
0	Preferable even before pandemic	Preferable even before pandemic	Preferable even before pandemic	never preferred this brand and not sure about ...	'-'	yes	1.0
1	Preferable even before pandemic	Preferable even before pandemic	Not preferable before pandemic but started buy...	Not preferable before pandemic but started buy...	'Yes'	yes	1.0
2	never preferred this brand and not sure about ...	never preferred this brand and not sure about ...	Preferable even before pandemic	never preferred this brand and not sure about ...	'Wild Stone'	wild stone	0.0
3	Preferable even before pandemic	Preferable even before pandemic	Preferable even before pandemic	Preferable even before pandemic	'Any local brand that compete any global brands'	local brand compet global brand	0.0
4	never preferred this brand and not sure about ...	Preferable even before pandemic	Preferable even before pandemic	never preferred this brand and not sure about ...	'Dettol'	dettol	0.0

Figure IV: Applied Vader Sentiment Algorithm to Generate Sentiment Score Which is applied on Respondents' Data

To overcome that problem and to get the correct sentiment score of the respondent, we used the concept of 'Named Entity Recognition'[7]. We can get the correct sentiment based on the data by using customized VADER Sentiment Analysis. There were two situations, where VADER sentiment algorithm didn't work properly. First is when some noun is given. The noun is nothing but the Indian product name. By writing Indian Product / Brand name, the respondent supports Indian brand only, but it doesn't have happy / sad or yes/no kind of answer. By applying the logic of 'if content is a noun, then change the sentiment score to 1' to where it was zero sentiment by VADER sentiment algorithm, ultimately correct emotions can be retrieved. Secondly, if respondent has already selected any Indian Brand from the given list, then he / she may write '-' or 'no' for one more brand in the text. So, we considered previous choices also, if sentiment score is resulted to zero by VADER sentiment algorithm. Thus, we developed a customized VADER algorithm for our study as shown in figure V. The logic of customized VADER algorithm is shown in figure VI.

	Hamam	Santoor	Lifebuoy	khadi	Anyother	MAnyother	sentiment_score
0	Preferable even before pandemic	Preferable even before pandemic	Preferable even before pandemic	never preferred this brand and not sure about ...	'-'	yes	1.0
1	Preferable even before pandemic	Preferable even before pandemic	Not preferable before pandemic but started buy...	Not preferable before pandemic but started buy...	'Yes'	yes	1.0
2	never preferred this brand and not sure about ...	never preferred this brand and not sure about ...	Preferable even before pandemic	never preferred this brand and not sure about ...	'Wild Stone'	wild stone	1.0
3	Preferable even before pandemic	Preferable even before pandemic	Preferable even before pandemic	Preferable even before pandemic	'Any local brand that compete any global brands'	local brand compet global brand	1.0
4	never preferred this brand and not sure about ...	Preferable even before pandemic	Preferable even before pandemic	never preferred this brand and not sure about ...	'Dettol'	dettol	1.0

Figure V: Applied Customized Vader Sentiment Algorithm to Generate Sentiment Score, Which is Applied on Respondents' Data

As per the respondents' data, the popularity of Indian brands as shown in figure IX a, b and c.

F. Selection and application of ML and Hybrid Ensembled Models:

Kapadia, B., Jain, A. (2021) the four best machine learning model for sentiment analysis on e-commerce data are – support vector machine (SVM), Naïve Bayes classification, Random Forest Algorithm and K Nearest Neighbor. All these algorithms are selected for this study.

For developing Hybrid Ensemble classification Model, we defined each of the four machine learning models – SVM, MNB, RFC and KNN - 5 times that results in a combination of a total of 20 weak learners as shown in figure VII. Finally, we used Max Voting Classifier method with 10-fold cross-validation for all 5 categories of learning, where

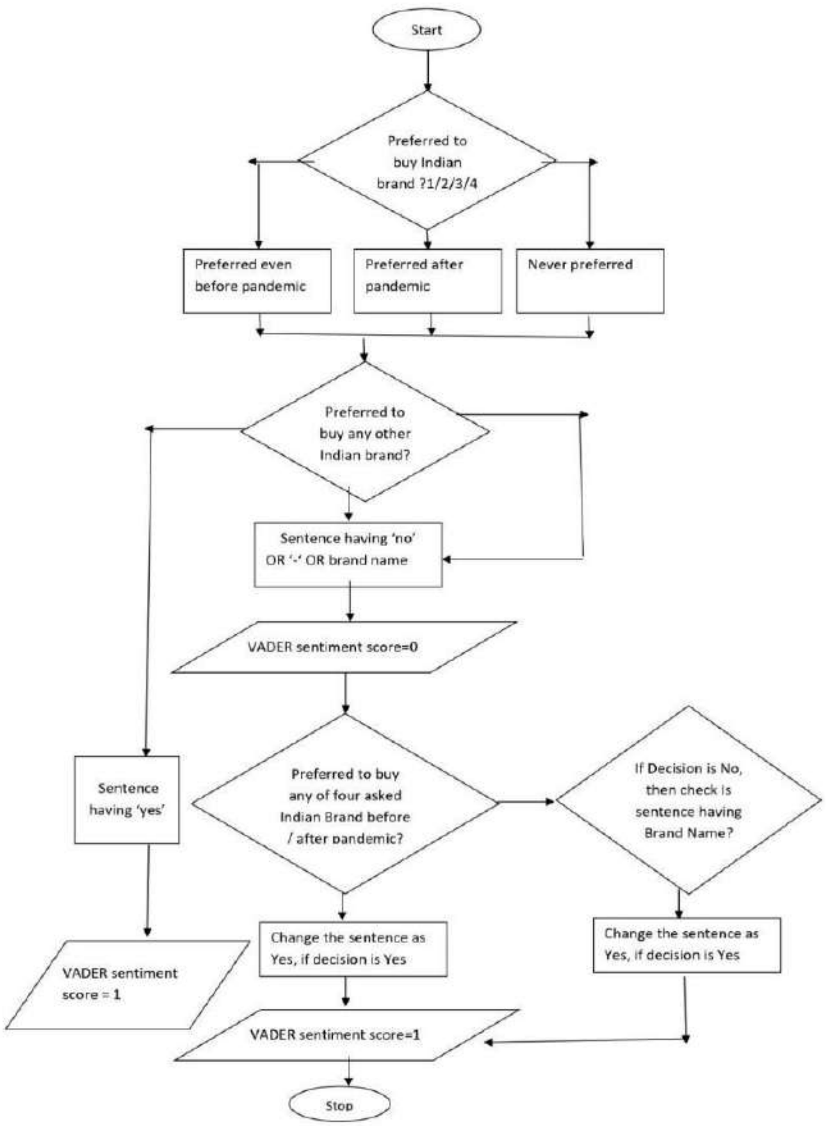

Figure VI: Flow Chart of Customized Vader Sentiment Analysis

the class which has been predicted mostly by the weak learners will be the final class prediction of the ensemble model. Dr. Vaibhav Kumar, (2021). The hybrid ensemble classifier is shown in figure VII.

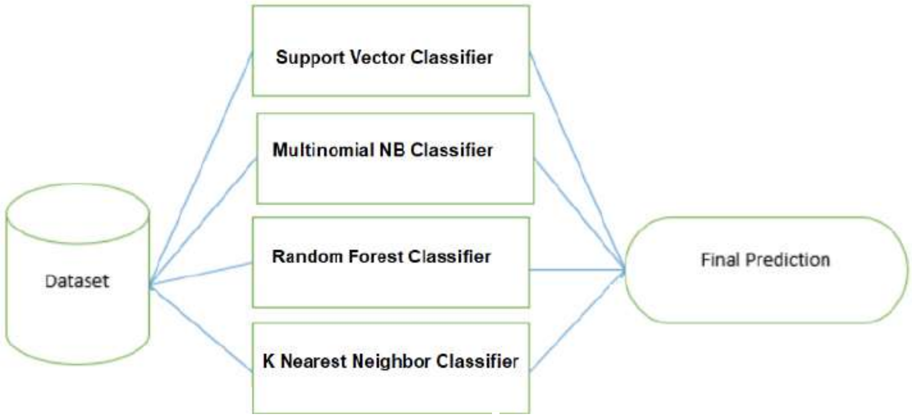

Figure VII: Hybrid Ensemble Classification Model Applied on Amazon_Vfl_Reviews.Csv File

For machine learning classification, data is divided as x-train, y-train as per the review and sentiment_score columns and x-test and y-test columns. All 4 models are defined as model1, model2, model3 and model4 respectively. Prediction for all models are made based on the x-test data. And the results have been generated for all models using 10-fold cross validation to get the robust outcome.

10-fold cross-validation is a common technique used in machine learning for evaluating the performance of a predictive model. The dataset is randomly partitioned into 10 equal-sized subsets, or "folds." The model is trained and evaluated 10 times, each time using a different fold as the test set and the remaining 9 folds as the training set. For each iteration, the model's performance metrics (such as accuracy, precision, recall, or F1-score) are computed on the test set. The final performance metric is typically computed as the average of the performance metrics from the 10 iterations.

Model Analysis and Visualization:

TABLE I: ANALYSIS OF LEXICAL TEXTBLOB, SPACYTEXTBLOB AND VADER SENTIMENT USING ML ALGORITHMS APPLIED ON AMAZON_VFL_REVIES.CSV.

Model Name	Model accuracy		
	Lexical-TextBlob	Lexical- SpacyTextBlob	Lexical- VADER
Support Vector Machine	0.8542	0.8628	0.8717
Multinomial Naïve Bayes	0.7832	0.7985	0.8010
Random Forest Classifier	0.8807	0.9292	0.9203
K Nearest Neighbor	0.7897	0.7154	0.6613
Hybrid Ensemble Model	0.9026	0.8959	0.9206

TABLE II: POPULARITY OF EXPENSIVE INDIAN BRANDS IN INDIA APPLIED ON RESPONDENTS' DATA

Model Name	Model accuracy for various Indian Brands				
	Clothing	watches	Automobiles	Jewellery	shoes
Multinomial Naïve Bayes	0.90	0.99	0.95	0.90	0.94
SVM	0.90	0.99	0.95	0.90	0.94
RF	0.90	0.99	0.95	0.90	0.94
KNN	0.90	0.99	0.95	0.90	0.94
Overall Popularity of Indian Brands	90.2%	98.8%	95.4 %	90.2%	94.2 %

Analyzing the model for the second objective, which is to find out the popularity of Indian products due to "Atmanirbhar Bharat Abhiyan". The popularity of Indian Brands for Clothing, Watches, Automobiles, Jewellery and shoes are respectively 90.2, 98.8, 95.4, 90.2 and 94.2 respectively which is shown in the following table II and table III for expensive products and inexpensive products respectively.

TABLE III: POPULARITY OF INEXPENSIVE INDIAN BRANDS IN INDIA APPLIED ON RESPONDENTS' DATA.

Model Name	**Model accuracy**				
	pen / pencil	Lipstick	Tea / coffee	soap / hand / face wash	shampoo / hair cleanser
Multinomial Naïve Bayes	0.98	0.83	0.95	0.96	0.84
SVM	0.98	0.84	0.95	0.96	0.85
RF	0.98	0.85	0.95	0.96	0.85
KNN	0.98	0.84	0.95	0.96	0.74
Overall Popularity of Indian Brands	98.8%	83.8%	94.8 %	96.4 %	85.7 %

This data is taken between Feb 2022 to May 2022, which is after ABA and a call from Hon'ble Prime Minister of India Shri Narendra Modiji.

Further, working for the second objective to find the popularity of Indian Brands due to launch of ABA, is shown in the following table IV and Table V for increase in the popularity of Indian Brands in terms of Expensive items and inexpensive items respectively. There is a rise in the popularity of buying Indian products after the Atmanirbhar Bharat Abhiyan announced.

TABLE IV: INCREASED POPULARITY PERCENTAGE OF EXPENSIVE INDIAN BRANDS AFTER ABA APPLIED ON RESPONDENTS' DATA.

Expensive Indian Brands Not preferred before but started preferring after ABA	**Clothing**	**watches**	**Automobiles**	**Jewellery**	**shoes**
Popularity percentage increased after ABA	21.6 %	15.2 %	14.3 %	20.1 %	16.2 %

TABLE V: INCREASED POPULARITY PERCENTAGE OF INEXPENSIVE INDIAN BRANDS AFTER ABA APPLIED ON RESPONDENTS' DATA

Inexpensive Indian Brands Not preferred before but started preferring after ABA	**pen / pencil**	**Lipstick**	**Tea / coffee**	**soap or hand wash or face wash**	**shampoo / hair cleanser**
Popularity percentage increased after ABA	14.0 %	11.3 %	15.5 %	18.0 %	17.7 %

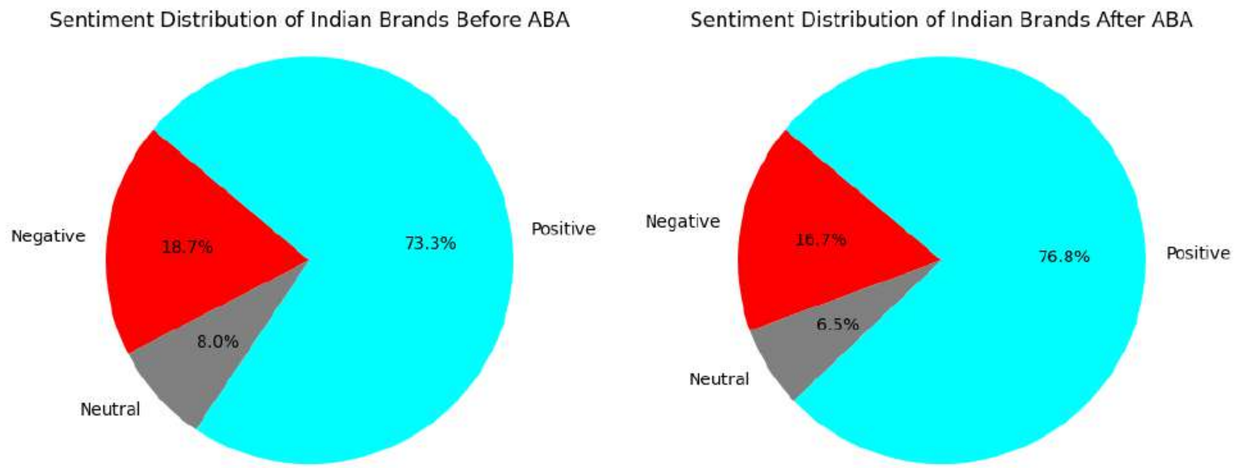

FIGURE VIII: A) Popularity of Indian Brands before Aba B) Popularity of Indian Brands after Aba Applied on Amazon_Vfl_Reviews.Csv File.

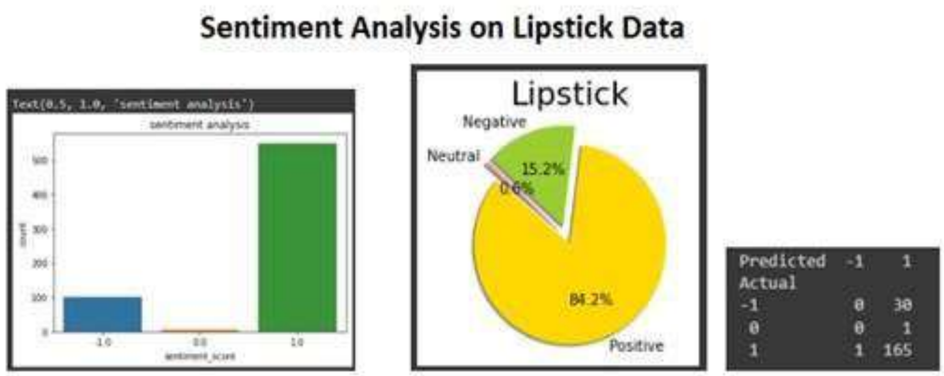

FIGURE IX: (A) Bar Chart, (B) Pie Chart And (C) Confusion Matrix Applied On Respondents' Data

DISCUSSION AND CONCLUSION:

Data in Amazon_vfl_reviews.csv data file is available only from 14th May 2020 to Sep 2020, whereas the respondent's dataset contains data up to Mar 2022.

According to amazon_vfl_reviews.csv dataset, figure VIII shows the rise in the popularity in after ABA which is 76.8, compared to popularity of Indian Brands before ABA which is 73.3, which shows that there is an increase in the popularity of Indian Brands by 3.5%. The accuracy of Hybrid ensemble mode using TextBlob, SpacyTextBlob and VADER are 0.9026, 0.8959 and 0.9206 respectively as shown in table I. Which proves that there is an increase in popularity after the launch of ABA, which is the first objective of this paper.

According to respondents' dataset, the popularity of Indian Brands increased after implementation of Atmanirbhar Bharat Abhiyan is 16.39 % on an average of all 10 products, which proves the popularity of Indian products due to "Atmanirbhar Bharat Abhiyan". By applying VADER Sentiment algorithm instead of TextBlob, the average accuracy was 88.5 and the popularity increased by 9.8 %. But with customized VADER Sentiment Algorithm, which used the Named Entity Recognition to identify the Indian Brand name in the review text, the average accuracy could be achieved as 92.1 and the average popularity is 92.83 %. 67.5 percent of respondents believe that Vocal for Local is a new form of globalization and not a rejection of globalization. 63.5% of respondents believe that during COVID lockdown period Local Brands served better than Global Brands. 89.1 % respondents agreed about the fact that they bought Indian products like Diya, Colours, Lights etc. to celebrate Diwali last year.

69.3 % respondents feel that the Indian products and handicrafts taking over the market during festivals. 62.9 % of respondents use clothes made by Hathkargha Udyog (handloom), out of which 8.3 % respondents were not preferring to use before pandemic but started buying after pandemic. More data can be collected for a better outcome. Also, popularity of Indian Brands is more inclined towards men than women and though too much of popularity for Indian brands is not there amongst youngsters but at least some of them started liking Indian brands.

We achieved 89% accuracy using SVM, 87% accuracy using Multinomial Naïve Bays, 90% accuracy with RF, 89% accuracy with KNN on all 11 products data set of total 7216 data with rule-based popularity as 88.3% positive response, 8.2 % negative response and 3.5% neutral response on an average including diya and other articles as one item totaling to 11 products of 656 respondents.

REFERENCES

[1] Bajaj, Bhumika and Narayan Singh, Sudhir, The Role of Atmanirbhar Bharat Abhiyan in Indian Economy Transformation through Innovation (January 4, 2022). Proceedings of the International Conference on Advances in Management Practices (ICAMP 2021), Available at SSRN: https://ssrn.com/abstract=4000746 or http://dx.doi.org/10.2139/ssrn.4000746

[2] Atmanirbhar Bharat Abhiyan: A Campaign for Self Reliant India - Upender Sethi - IJFMR Volume 4, Issue 4, July-August 2022.

[3] https://www.goelyash.in/post/atmanirbhar-bharat-abhiyan as on 27th Mar 2024.

[4] Press Information Bureau. (2020). Cabinet approves 'Atmanirbhar Bharat Abhiyan' package for allocation of foodgrains to the migrants. Government of India. Retrieved from https://pib.gov.in/PressReleasePage.aspx?PRID=1623638

[5] Ghosal, S. (2021). How 'Atmanirbhar Bharat' is reshaping consumer behavior. Business Today. Retrieved from https://www.businesstoday.in/opinion/columns/how-atmanirbhar-bharat-is-reshaping-consumer-behaviour/story/439047.html

[6] The Economic Times. (2021). How Atmanirbhar Bharat is set to change the Indian economy. Retrieved from https://economictimes.indiatimes.com/news/economy/policy/how-atmanirbhar-bharat-is-set-to-change-the-indian-economy/articleshow/82841367.cms

[7] Business Standard. (2021). Explained: What is the Atmanirbhar Bharat Abhiyan? Business Standard. Retrieved from https://www.business-standard.com/about/what-is-atmanirbhar-bharat-abhiyan#collapse

[8] Singh, S., Srivastava, S., & Joshi, A. (2019). Sentiment Analysis of Public Opinion on Demonetization and GST Using Twitter Data. In 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS) (pp. 143-148). IEEE.

[9] Chen, Y., Liu, Y., & Wu, Y. (2020). Analyzing the public sentiment and its effect on eco-friendly consumption behavior under the environmental policy perspective. Journal of Cleaner Production, 275, 123205.

[10] Li, Y., Zheng, K., Mao, J., Chen, Y., & Zhu, H. (2018). Analyzing the sentiment distribution of the Affordable Care Act using Twitter data. In Proceedings of the 2018 International Conference on Smart Computing and Communication (pp. 243-248). ACM.

[11] Kapadia, B., Jain, A. (2021). Analysis of Papers Based on Sentiment Analysis Applications on E-Commerce Data. In: Abraham, A., Sasaki, H., Rios, R., Gandhi, N., Singh, U., Ma, K. (eds) Innovations in Bio-Inspired Computing and Applications. IBICA 2020. Advances in Intelligent Systems and Computing, vol 1372. Springer, Cham. https://doi.org/10.1007/978-3-030-73603-3_30

[12] Dr. Vaibhav Kumar (2021). A Hands-on Guide To Hybrid Ensemble Learning Models, With Python Code, AI Mysteries.

[13] Singh, G., Kumar, B., Gaur, L., & Tyagi, A. (2019). Comparison between Multinomial and Bernoulli Naïve Bayes for Text Classification. 2019 International Conference on Automation, Computational and Technology Management (ICACTM).

[14] Borg, A., & Boldt, M. (2020). Using VADER Sentiment and SVM for Predicting Customer Response Sentiment. Expert Systems with Applications, 113746.

[15] Singh, S. N., & Sarraf, T. (2020). Sentiment Analysis of a Product based on User Reviews using Random Forests Algorithm. 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence).

[16] Irawaty, I., Andreswari, R., & Pramesti, D. (2020). Development of Youtube Sentiment Analysis Application using K-Nearest Neighbors (Nokia Case Study). 2020 3rd International Conference on Information and Communications Technology (ICOIACT).

[17] Hutto, C.J., & Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of social media Text. ICWSM.

[13] Yang, S. E. (n.d.). Twitter Sentiment Analysis Using Natural Language Toolkit and VADER Sentiment. In M. 1.-1. IMECS 2019 (Ed.). ISBN: 978-988-14048-5-5

[18] V. Bonta, N. K. N. J. for Comprehensive Study on Lexicon Based Approaches for Sentiment Analysis, Asian Journal of Computer Science and Technology vol. 8 pp. 1 -6, 2019.

[19] Wilson, T. (2022). TextBlob: Simplified Text Processing. Retrieved from https://textblob.readthedocs.io/en/dev/

[20] Tomažič, Š., Pollak, S., & Košir, A. (2021). SpaCyTextBlob: A simple and efficient NLP library for sentiment analysis. Journal of Open Source Software, 6(65), 3441. https://doi.org/10.21105/joss.03441