NavaJyoti, International Journal of Multi-Disciplinary Research
Volume 7, Issue 2, June 2025

SENTIMENT ANALYSIS OF PFIZER VACCINE ON X DATA

[1]Maria John, **[2]Sanjana Sharma M**
[1] Assistant Professor, Jyoti Nivas College Autonomous, Bengaluru, India
[2] Student, Jyoti Nivas College Autonomous, Bengaluru, India

Abstract

The growing role of social media in shaping public opinion has made sentiment analysis a crucial tool for understanding vaccine perceptions. This study focuses on sentiment analysis of Pfizer vaccine-related tweets using machine learning techniques to classify public sentiment into positive, negative, or neutral categories.The dataset, sourced from X data, undergoes text preprocessing involving text cleaning, removal of URLs, punctuation, and stopwords. Sentiment scores are extracted using VADER (Valence Aware Dictionary and Sentiment Reasoner), while TextBlob is used to compute polarity and subjectivity scores. Feature extraction is performed using TF-IDF vectorization, combined with additional numerical features for improved classification.To classify sentiments, four machine learning models are implemented: Logistic Regression, Support Vector Machine (SVM), Random Forest, and Naïve Bayes. The models are trained and evaluated using accuracy scores and classification reports. The study further visualizes model performance, sentiment distribution, and comparisons of sentiment predictions using bar charts and histograms.The findings provide insights into how people perceive the Pfizer vaccine, with potential applications for policymakers, healthcare professionals, and researchers aiming to address vaccine hesitancy and misinformation. By leveraging sentiment analysis, this study contributes to understanding the influence of social media on public health discussions.

Keywords:

Sentiment Analysis, Pfizer Vaccine, Social Media, Machine Learning, Public Perception

1. Introduction

The rapid spread of information on social media platforms has significantly influenced public perceptions of vaccines. With the Pfizer COVID-19 vaccine being a crucial component of global immunization efforts, it is essential to understand how people perceive it. Sentiment analysis, a field of natural language processing (NLP), helps in analyzing and classifying public opinion into positive, negative, or neutral sentiments.

This study aims to analyze public sentiment regarding the Pfizer vaccine by leveraging machine learning techniques. The research focuses on collecting X data data, preprocessing it, extracting sentiment scores, and classifying sentiments using various machine learning models.

2. Methodology

2.1 Data Collection

Tweets related to the Pfizer vaccine were extracted using the X data API. Data collection included filtering tweets based on relevant hashtags and keywords such as "#PfizerVaccine" and "#COVIDVaccine".

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
id	user_name	user_location	user_description	user_created	user_followers	user_friends	user_favourites	user_verifi	date	text	hashtags	source	retweets	favorites	is_retweet
1.34054E+18	Rachel Roh	La Crescenta-Montr	Aggregator of Asian /	08-04-2009 17:52	405	1692	3247	FALSE	########	Same folks	['PfizerBio	Twitter for	0	0	FALSE
1.33816E+18	Albert Fong	San Francisco, CA	Marketing dude, tech	21-09-2009 15:27	834	666	178	FALSE	########	While the world has b		Twitter We	1	1	FALSE
1.33786E+18	eliðŸ‡±ðŸ‡¹ðŸ‡ªðŸ‡ºŒ	Your Bed	heil, hydra ðŸ–â˜°	25-06-2020 23:30	10	88	155	FALSE	########	#coronavii	['coronavi	Twitter for	0	0	FALSE
1.33786E+18	Charles Adler	Vancouver, BC - Car	Hosting "CharlesAdle	10-09-2008 11:28	49165	3933	21853	TRUE	########	Facts are immutable,		Twitter We	446	2129	FALSE
1.33785E+18	Citizen News Channel		Citizen News Channe	23-04-2020 17:58	152	580	1473	FALSE	########	Explain to	['whereare	Twitter for	0	0	FALSE
1.33785E+18	Dee	Birmingham, Englan	Gastroenterology tra	26-01-2020 21:43	105	108	106	FALSE	########	Does anyone have an		Twitter for	0	0	FALSE
1.33785E+18	Gunther Fehlinger	Austria, Ukraine anc	End North Stream 2 r	10-06-2013 17:49	2731	5001	69344	FALSE	########	it is a bit sa	['vaccinati	Twitter We	0	4	FALSE
1.33785E+18	Dr.Krutika Kuppalli		ID, Global Health, VH	25-03-2019 04:14	21924	593	7815	TRUE	########	There	['BidenHar	Twitter for	2	22	FALSE
1.33785E+18	Erin Despas		Designing&selling on	30-10-2009 17:53	887	1515	9639	FALSE	########	Covid	['CovidVac	Twitter We	2	1	FALSE
1.33784E+18	Ch.Amjad Ali	Islamabad	#ProudPakistani	12-11-2012 04:18	671	2368	20469	FALSE	########	#CovidVa	['CovidVac	Twitter We	0	0	FALSE
1.33784E+18	Tamer Yazar	Turkey-Israel	Im Market Analyst,	17-09-2009 16:45	1302	78	339	FALSE	########	while	['PfizerBio	Twitter We	0	0	FALSE
1.33784E+18	VoiceM		campaigner & optimi	31-08-2020 10:38	2	25	20	FALSE	########	@cnnbrk #	['COVID19	Twitter We	0	0	FALSE
1.33782E+18	WION	India	#WION: World Is On	21-03-2016 03:44	292510	91	7531	TRUE	########	The agency also relea		TweetDeck	0	18	FALSE
1.33781E+18	Dr.Krutika Kuppalli		ID, Global Health, VH	25-03-2019 04:14	21924	593	7815	TRUE	########	For all the	['PfizerBio	Twitter for	48	82	FALSE
1.33781E+18	Opoyi		High-quality trusted c	13-01-2019 18:33	10332	49	16	FALSE	########	"Expect 145 sites acr		TweetDeck	0	0	FALSE
1.33779E+18	City A.M.	London, England	London's business ne	09-06-2009 13:53	66224	603	771	TRUE	########	Trump	['vaccine']	Twitter for	0	1	FALSE
1.33779E+18	STOPCOMMONPASS.ORG	Global	'Trust' is not carte-bl	25-10-2020 20:33	406	176	479	FALSE	########	UPDATED	['YellowFe	Twitter We	2	2	FALSE
1.33778E+18	ILKHA	TÃ¼rkiye	Official Twitter acco	22-05-2015 08:31	4056	6	3	TRUE	########	Coronavir	['Iran', 'cor	TweetDeck	3	5	FALSE
1.33778E+18	Braderz73ðŸŒ¹#GTTO ðŸ‡¹ðŸ‡¹	Bristol, UK	One of those lefty	24-07-2012 08:18	6430	6292	45007	FALSE	########	.@Pfizer w	['CovidVac	Twitter for	3	3	FALSE
1.33778E+18	Alex Vie	Los Angeles, CA	Marine vet. Yogi. Kra	24-01-2010 04:43	125	442	5401	FALSE	########	The trump	['COVIDIO	Twitter for	0	0	FALSE
1.33777E+18	Mani			10-10-2019 13:41	26	33	2515	FALSE	########	How much	['fda', 'vac	Twitter We	0	0	FALSE
1.33777E+18	Richard Dunne, MD	Rochester, NY	Husband, Girl Dad, GI	23-04-2012 12:18	1982	608	9110	FALSE	########	Anyone wa	['PfizerBio	Twitter for	0	2	FALSE
1.33777E+18	City A.M.	London, England	London's business ne	09-06-2009 13:53	66224	603	771	TRUE	########	Trump	['vaccine']	Buffer	1	0	FALSE

2.2 Preprocessing

The collected tweets were cleaned by:

- Removing URLs, special characters, and punctuation
- Eliminating stopwords to reduce noise
- Converting text to lowercase for uniformity

2.3 Sentiment Scoring

- **VADER** was used for lexicon-based sentiment analysis
- **TextBlob** provided polarity and subjectivity scores

2.4 Feature Extraction

TF-IDF vectorization was applied to convert text into numerical features for model training.

2.5 Model Selection and Training

Four machine learning models were used:

- Logistic Regression
- Support Vector Machine (SVM)
- Random Forest
- Naïve Bayes

2.6 Methods

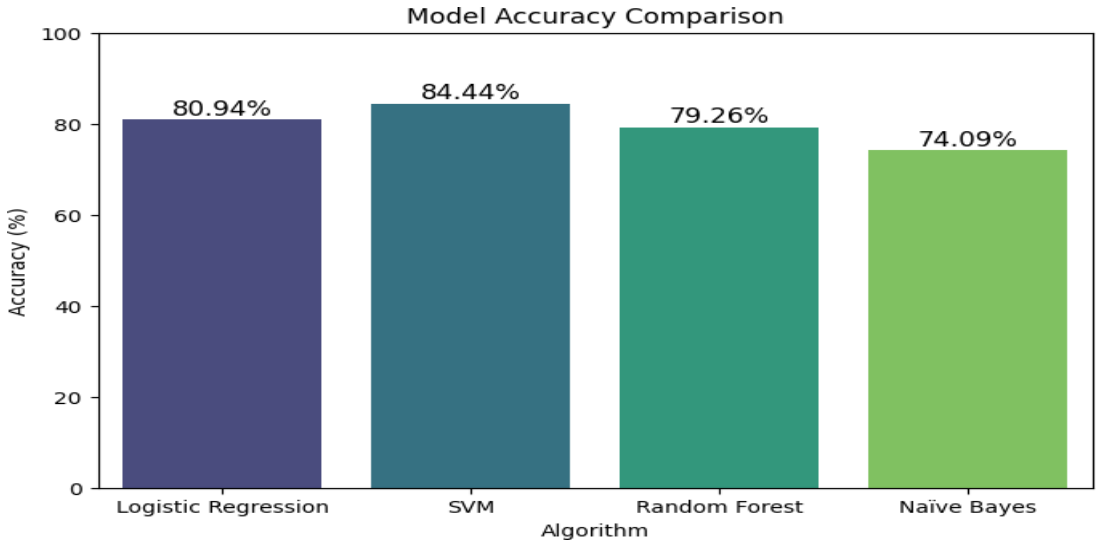
This study employs a structured approach to analyze public sentiment regarding the Pfizer vaccine using machine learning techniques. The dataset was collected from X (formerly Twitter) using relevant keywords and hashtags, ensuring that only English-language tweets were considered while filtering out retweets and irrelevant posts. The preprocessing stage involved cleaning the text by removing URLs, punctuation, special characters, and stopwords, followed by text normalization techniques such as tokenization and lemmatization. Sentiment scores were generated using VADER and TextBlob, which helped classify tweets as positive, negative, or neutral based on sentiment polarity and subjectivity. For feature extraction, TF-IDF vectorization was applied to convert textual data into numerical representations, along with additional features such as tweet length and word count to enhance classification accuracy. The processed data was then used to train four machine learning models—Logistic Regression, Support Vector Machine (SVM), Random Forest, and Naïve Bayes—which were evaluated using accuracy, precision, recall, and F1-score. Visualizations such as bar charts and histograms were used to compare model performance, providing insights into public perception of the Pfizer vaccine.

Each model was trained and evaluated using classification metrics such as accuracy, precision, recall, and F1-score.

3. Results and Discussion

- **Sentiment Distribution:** The majority of tweets were classified as positive, followed by neutral and negative sentiments.
- **Model Performance:** Among the tested models, SVM and Logistic Regression achieved the highest accuracy.
- **Insights:** The study highlighted key public concerns and positive reactions towards the Pfizer vaccine.

	Accuracy	Precision	
Logistic Regression Classifier	80.94	Negative Positive Neutral	0.78 0.84 0.79
Support Vector Machine	84.44	Negative Positive Neutral	0.79 0.88 0.83
Random Forest Classifier	79.26	Negative Positive Neutral	0.80 0.81 0.78
Naïve Bayes Classifier	74.09	Negative Positive Neutral	0.89 0.70 0.78



4. Conclusion

This study demonstrates the effectiveness of sentiment analysis in understanding vaccine perceptions. The findings can help policymakers, healthcare officials, and researchers address misinformation and enhance public trust in vaccination efforts.

5. References

Mandloi, Lokesh, and Ruchi Patel. "Twitter sentiments analysis using machine learninig methods." *2020 international conference for emerging technology (INCET)*. IEEE, 2020.

Bhalerao, Ashish A., et al. "Sentiment Analysis on Covid-19 Vaccination Using Machine Learning Techniques." *International Conference on Applications of Machine Intelligence and Data Analytics (ICAMIDA 2022)*. Atlantis Press, 2023.