Exploring User Sentiments: ChatGPT’s Impact in the Education Sector

Team Members: Bhanu Prakash Reddy Gaddam, Nithin Shastry , Atharv Jangam

INDIANA UNIVERSITY BLOOMINGTON

INTRODUCTION

Purpose

- To identify sentiment of twitter users on how ChatGPT can impact in the field of education
- We obtained tweets between 2021 and 2022 to conduct the experiment

Why is this study important?

- User sentiment analysis is essential for evaluating the effectiveness of software/products.
- Use case study - Tay AI which was shut down within 16 hours and users termed it as corrupt AI
- User sentiments can further help to improve and update software/products

INTRODUCTION

Research Questions

- How twitter users are reacting on ChatGPT's impact in the field of education?
- How to build a good dataset which can serve the purpose of the experiment

Contributions of this study

- Creating exclusive dataset can help to study further about sentiments of ChatGPT and education in the future
- Outcome of this study may be compared with influence of other AI bots (BARD etc) in the field of education

Exploratory Data Analysis

- EDA was performed on around 100000 tweets that were scrapped related to the field of education and ChatGPT
- We used Snscrape to scrape tweets and to create dataset
- Creation of dataset was challenging as it took many trials to come up with quality dataset

Exploratory Data Analysis

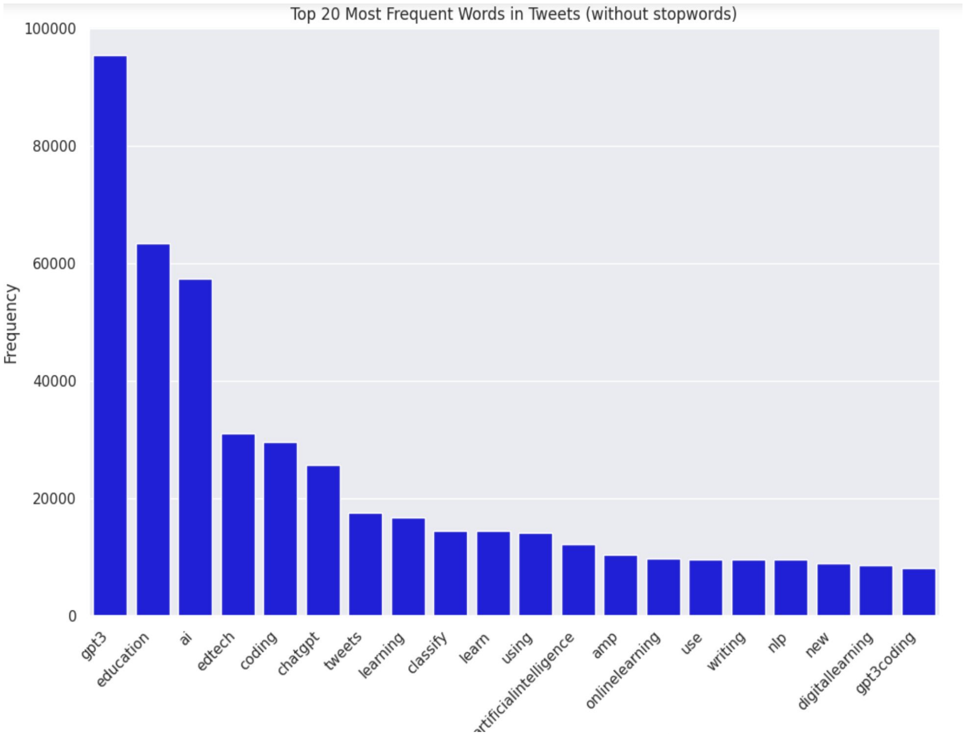


Fig 1 - Word Frequencies

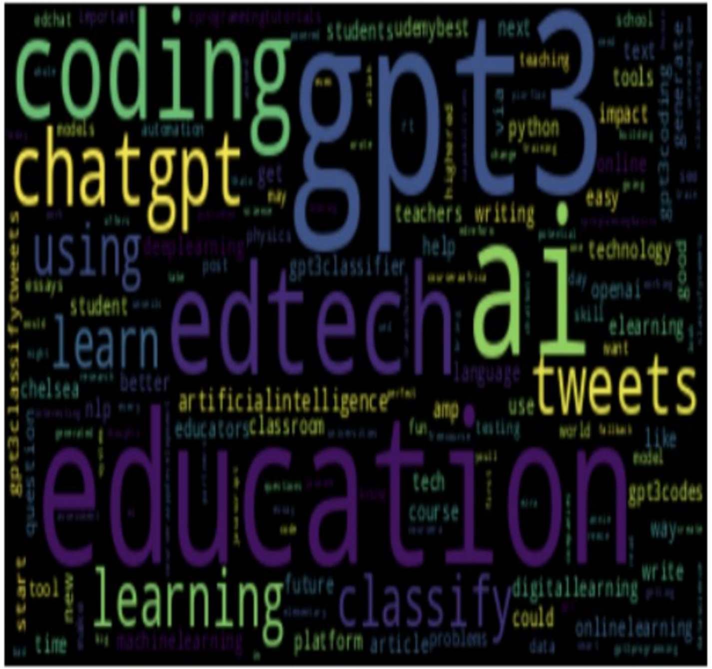


Fig 2 - Word cloud

VADER Model

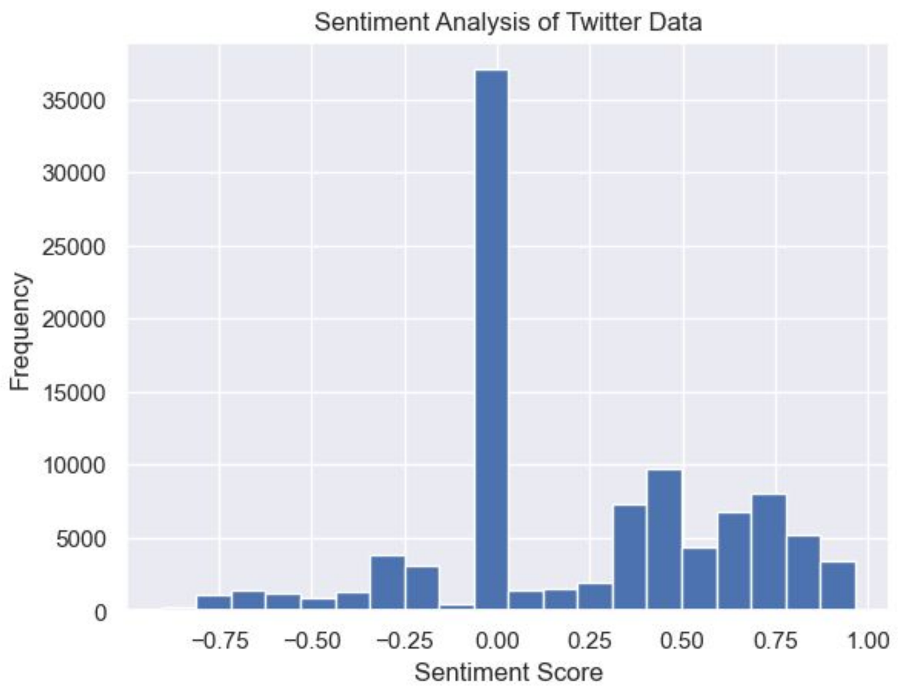
- VADER (Valence Aware Dictionary and Sentiment Reasoner) is a lexicon and rule-based sentiment analysis tool.
- It is designed to handle unique challenges of sentiment analysis in social media texts, where the use of non-standard language, sarcasm, and slang can make sentiment analysis difficult.
- It contains over 7500 lexical features that have been manually annotated with sentiment scores.
- Each feature has a positive and negative polarity score and an intensity score that ranges from 0 to 4.
- We used SentimentIntensityAnalyzer() class from nltk.sentiment module, which is used to calculate the sentiment of text.

VADER Model

	date	tweet_text	processed_text	sentiment_score
0	2022-01-27 14:42:36+00:00	RT GPT-3-like models with extended training co...	[rt, gpt, like, models, extended, training, fu...	0.0000
1	2022-01-08 08:46:33+00:00	RT Testing GPT-3 on Elementary Physics Unveils...	[rt, testing, gpt, elementary, physics, unveil...	-0.2263
2	2022-01-08 08:37:45+00:00	Testing GPT-3 on Elementary Physics Unveils So...	[testing, gpt, elementary, physics, unveils, i...	-0.2263
3	2021-11-26 16:49:24+00:00	Now we know how @BorisJohnson wrote his speech...	[know, borisjohnson, wrote, speech, gpt, piorf...	0.0000
4	2021-11-26 15:07:12+00:00	@sharplm shares how GPT-3 generates believable...	[sharplm, shares, gpt, generates, believable, ...	-0.3382
5	2021-11-05 09:39:02+00:00	#ArtificialIntelligence is getting better at #...	[getting, better, writing, amp, universities, ...	0.0000
6	2022-01-30 09:45:02+00:00	Classifying tweets with GPT-3 is easy! You can...	[classifying, tweets, gpt, easy, use, tool, ca...	0.4926
7	2022-01-27 14:42:36+00:00	RT GPT-3-like models with extended training co...	[rt, gpt, like, models, extended, training, fu...	0.0000
8	2022-01-08 08:46:33+00:00	RT Testing GPT-3 on Elementary Physics Unveils...	[rt, testing, gpt, elementary, physics, unveil...	-0.2263
9	2022-01-08 08:37:45+00:00	Testing GPT-3 on Elementary Physics Unveils So...	[testing, gpt, elementary, physics, unveils, i...	-0.2263

VADER Model

- Total Positive Tweets: 48,242
- Total Neutral Tweets: 38,300
- Total Negative Tweets:13484



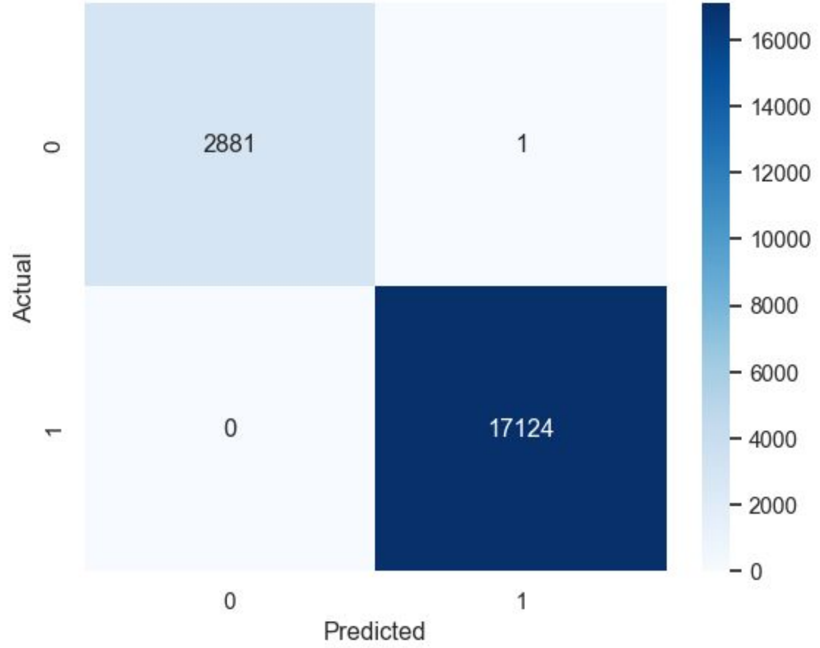
Logistic Regression Model

- We got sentiment scores from VADER model and we used it to perform supervised learning on that data.
- We used Logistic Regression model, which is a popular choice for sentiment analysis, as it can handle binary and multi-class classification problems and is relatively simple and fast to train.
- We used 80% of the data to train the model and 20% to test the model.
- We applied TF-IDF Vectorizer to build a sparse matrix and remove English stop words, and converted the sentiment scores to binary labels.
- Then we trained the Logistic Regression model and did prediction on test data.

Logistic Regression Model

	precision	recall	f1-score	support
0	1.00	1.00	1.00	2882
1	1.00	1.00	1.00	17124
accuracy			1.00	20006
macro avg	1.00	1.00	1.00	20006
weighted avg	1.00	1.00	1.00	20006

- We got accuracy of 0.99995 to be precise.
- This is because we don't have any manually annotated data. We used the data generated by VADER model.

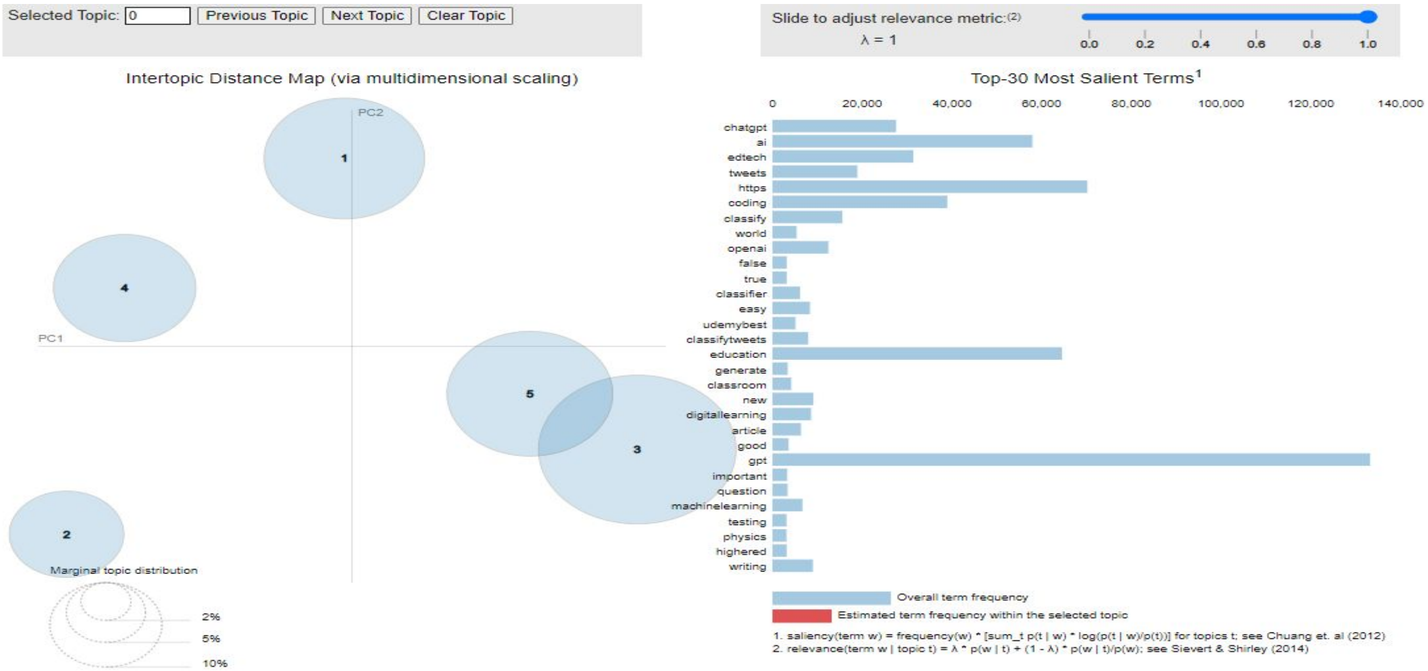


CONFUSION MATRIX

LDA(Latent Dirichlet Allocation)

- It is a topic modeling technique that allows discovering hidden topics or themes in a large collection of texts, such as a corpus of documents, by analyzing their word frequency patterns.
- The LDA model will represent each article as a mixture of topics, where each topic is a probability distribution over words
- Using these representations, we can interpret the topics and label them based on the most frequent words and their meaning

LDA plot



Selected Topic: 0
Previous Topic
Next Topic
Clear Topic
Slide to adjust relevance metric:(2)
λ = 1
0.0
0.2
0.4
0.6
0.8
1.0
Intertopic Distance Map (via multidimensional scaling)
PC2
PC1
1
2
3
4
5
Marginal topic distribution
2%
5%
10%
Top-30 Most Salient Terms[1]
0
20,000
40,000
60,000
80,000
100,000
120,000
140,000
chatgpt
ai
edtech
tweets
https
coding
classify
world
openai
false
true
classifier
easy
udemybest
classifytweets
education
generate
classroom
new
digitallearning
article
good
gpt
important
question
machinelearning
testing
physics
highered
writing
Overall term frequency
Estimated term frequency within the selected topic
1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

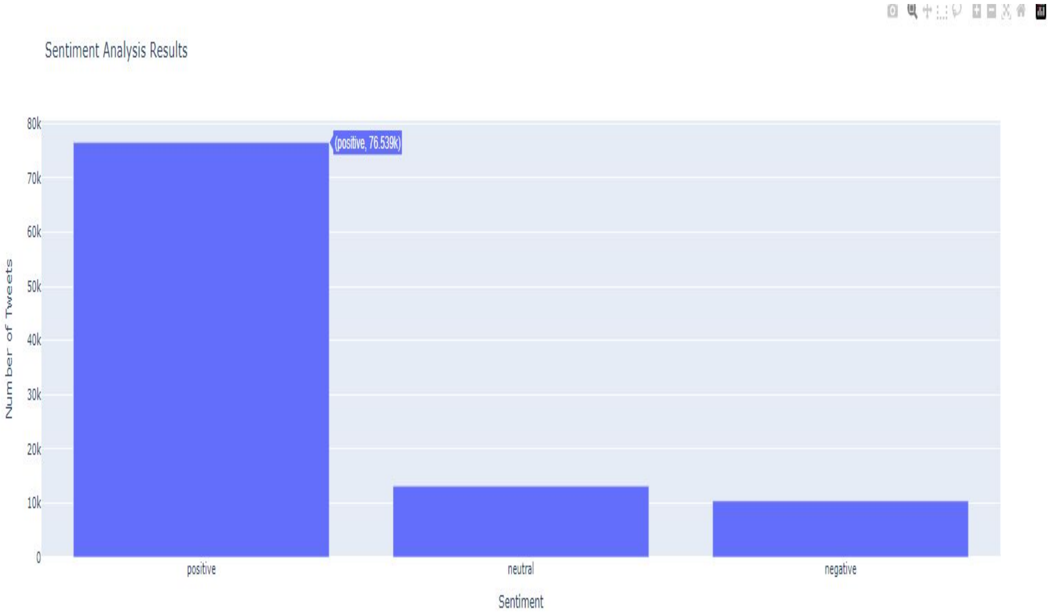
SentiWordNet

- SentiWordNet is a lexical resource for sentiment analysis that assigns to each synset (a set of synonyms that share a common meaning) in WordNet three sentiment scores: positivity, negativity, and objectivity.

- These scores indicate the degree of positivity, negativity, and objectivity associated with the synset. SentiWordNet is built using the WordNet lexical database, which provides a large set of synsets organized into a network of semantic relations.

- The overall sentiment polarity of the text can be determined by aggregating the sentiment scores of the synsets.

Results



There are total of 76k positive tweets, 13k neutral and 10k negative tweets.

This gives us a very clear picture of how people think about chatgpt when it comes to education.

Conclusion

- As discussed, data collection was one of the biggest challenge for sentiment analysis.
- LDA helped us in knowing the top 30 words in the corpus.
- VADER model may produce incorrect sentiment scores when dealing with complex and nuanced language use.
- SentiWordNet helped us in analyzing the sentiment of tweets in positive, negative and neural way.
- Both VADER and SentiWordNet was able to predict accurately and performed well on our dataset.
- Results from different models indicate that people are positive about ChatGPT in education, but some people are also concerned about integrity and plagiarism.
- Future scope could be, determining characteristics of person who made the tweet to analyze the behaviour patterns, and how people are changing their opinion with time and updation of ChatGPT.

Thank you!

Questions?