by

SUGANYA K(513421106042)
B.E(ECE)--III YEAR
University College of Engineering
Kanchipuram

AI_PROJECT PHASE-4

SENTIMENT ANALYSIS FOR MARKETING

INTRODUCTION TO SENTIMENT ANALYSIS

- A computational method called sentiment analysis, called opinion mining seeks to ascertain the sentiment or emotional tone expressed in a document. Sentiment analysis has become a crucial tool for organizations to understand client preferences and opinions as social media, online reviews, and customer feedback rise in importance. In this blog post, we'll look at how natural language processing (NLP) methods can be used to analyze the sentiment in customer reviews.

- Using algorithms and methodologies, sentiment analysis examines text data to determine the underlying sentiment. Businesses can better measure consumer satisfaction, pinpoint problem areas, and make educated decisions when they know whether the mood expressed is favorable, negative, or neutral. Sentiment analysis can examine various text data types, including social media posts, product reviews, survey replies, and correspondence with customer service representatives.

WHY NATURAL LANGUAGE PROCESSING (NLP) IS ESSENTIAL?

- In sentiment analysis, Natural Language Processing (NLP) is essential. NLP uses computational methods to interpret and comprehend human language. It includes several operations, including sentiment analysis, named entity recognition, part-of-speech tagging, and tokenization. NLP approaches allow computers to read, interpret, and comprehend language, enabling automated customer feedback analysis and accurate sentiment information extraction.
- NLP methods are employed in sentiment analysis to preprocess text input, extract pertinent features, and create predictive models to categorize sentiments. These methods include text cleaning and normalization, stopword removal, negation handling, and text representation utilizing numerical features like word embeddings, TF-IDF, or bag-of-words. Using machine learning algorithms, deep learning models, or hybrid strategies to categorize sentiments and offer insights into customer sentiment and preferences is also made possible by NLP.

NATURAL LANGUAGE PROCESSING (NLP) FUNDAMENTALS

- Artificial intelligence (AI) has a subfield called Natural Language Processing (NLP) that focuses on how computers and human language interact. It involves the creation of algorithms and methods that let computers meaningfully comprehend, decipher, and produce human language. Machine translation, sentiment analysis, information extraction, and question-answering systems are just a few of the many applications of NLP.
- NLP techniques include tokenization, part-of-speech tagging, named entity recognition, and word embeddings. Text is divided into tokens or individual words through the process of tokenization. It assists in word-level text analysis and processing, a crucial step in NLP activities. For machines to comprehend the syntactic structure of a sentence, part-of-speech tagging gives grammatical labels (such as nouns, verbs, and adjectives) to each word in a sentence. Many NLP activities, including parsing, language modeling, and text production, depend on this knowledge.

- Named Entity Recognition (NER) is the process of finding and categorizing named entities in text, such as names of individuals, groups, places, and dates. Information extraction, entity linking, and knowledge graph development depend heavily on NER. Word embeddings capture the semantic and contextual links between words and numerical representations of words. Word meanings are encoded via embeddings, allowing computers to recognize word relationships. Word2Vec, GloVe, and BERT have widely used word embedding methods.

PREPROCESSING TECHNIQUES FOR CUSTOMER FEEDBACK

- Data Cleaning and Normalization
- Stop word Removal
- Lemmatization and Stemming
- Handling Negation and Emojis
- Dealing with Noisy Text

FEATURE EXTRACTION FOR SENTIMENT ANALYSIS

- Bag-of-Words (BoW) Model:

The BoW model ignores word order and depicts text as a collection of singular words. It generates a frequency vector that shows whether certain words are present or absent in a document. The boW is straightforward and practical but disregards the semantic connections between words.

- Inverse Document Frequency Term Frequency (TF-IDF):

When determining a word's relevance, TF-IDF considers the word's frequency within the document and across the entire corpus. It gives terms used more frequently in one document but less frequently in another document heavier weights. The discriminative strength of words can be captured via TF-IDF.

- Word Embeddings:

By depicting words as dense vector representations, word embeddings can capture the semantic meaning of words. To learn word embeddings, techniques like Word2Vec and GloVe consider the context in which words are used. These embeddings enable algorithms to capture more complex sentiment data by capturing the links between words.

- Document Embeddings:

Word embeddings may now represent whole documents thanks to document embeddings like Doc2Vec. They record a document's contextual information, enabling models to comprehend the overall attitude portrayed. Instead of word-level analysis, document embeddings allow for sentiment analysis at the document level.

WHAT ARE THE SENTIMENT CLASSIFICATION TECHNIQUES?

Here are a few typical approaches:

Traditional Machine Learning Algorithms

Naive Bayes: Based on the occurrence of words or other features in the document, Naive Bayes determines the likelihood that a document belongs to a particular sentiment class. It is computationally effective and presupposes feature independence.

Support Vector Machines (SVM): SVM is a supervised learning technique that divides data into various sentiment groups by locating an ideal hyperplane. SVM maximizes the margin between classes while considering the data's high-dimensional feature space.

Random Forests: Random Forests classify emotion using a collection of decision trees. The final prediction is based on the majority vote of all decision trees, and each tree is independently constructed. High-dimensional data may be handled, and complicated relationships can be captured with the help of random forests.

- Deep Learning Approaches
- Convolutional Neural Networks (CNN): Although their main application is image analysis, CNNs can also be utilized for text classification problems. Convolutional layers enable CNNs to automatically learn features by capturing local patterns and hierarchical representations in the text data.
- Recurrent neural networks (RNNs): RNNs can accurately represent temporal dependencies in text and are well suited for sequential data. They process text by preserving a hidden state that saves data from prior words, allowing them to gather contextual data.
- Transformer-Based Models: Modern performance in sentiment classification has been attained by transformer-based models, such as BERT (Bidirectional Encoder Representations from Transformers). These models capture global dependencies and context-aware representations via self-attention methods.
- Ensemble Techniques: These methods integrate several models to get a single prediction. This can be done by averaging the predictions of various models or training numerous models with various hyperparameters or topologies. Often, ensemble approaches increase sentiment classification's robustness and accuracy.

PRACTICAL IMPLEMENTATION AND CASE STUDIES

- ## Building a Sentiment Analysis Pipeline

Data collection, preprocessing, feature extraction, model training, and evaluation are all steps in the pipeline development process for sentiment analysis. It entails gathering data from multiple sources, cleaning and preparing it, choosing pertinent features, training and optimizing the sentiment analysis model, and assessing its performance using relevant metrics.

- ## Analysing Customer Feedback in E-commerce

In this case study, consumer feedback, reviews, and ratings for e-commerce platforms can be analyzed using sentiment analysis. The sentiment analysis pipeline can be used to measure overall customer happiness, highlight areas for improvement, and detect positive and negative feelings expressed by customers.

- ## Sentiment Analysis for Social Media Monitoring

User-generated information, such as posts, tweets, and comments, is abundant on social networking platforms. To track social media sentiment regarding a brand, item, or event, sentiment analysis can be used. The pipeline can be used to monitor trends in public opinion, find hot subjects, and gain insight into client preferences.

Ethical Considerations in Sentiment Analysis

- Sentiment analysis ethical considerations cover a number of significant areas, including:
- Bias and Fairness Issues: Sentiment analysis models may be biased, resulting in unjust treatment or discriminating outcomes. To achieve fair representation and equal treatment for all demographic groups, bias must be addressed.
- Privacy and Data Protection: Sentiment analysis frequently entails processing personal data, which raises privacy and data protection issues. Protecting user data, getting informed consent, and ensuring all applicable privacy laws are followed is crucial.
- Transparency and Accountability: Accountability and Transparency: In terms of sentiment analysis, transparency is making the procedures, formulas, and selection criteria transparent and understandable. It's crucial to take responsibility for the sentiment analysis results and offer justifications for the choices taken.

Modern AI-based sentiment analysis tools can:

- Process text feedback at massive scales
- Identify trends across the entire customer journey
- Provide recommendations with the most impact on CSAT and NPS

sentiment analysis solution is probably the way to go:

- You need to work with a large amount of data at scale: No matter how great your team is, no one can sort through thousands of surveys, reviews, chats, and support tickets manually. Sentiment analysis does this automatically in a matter of minutes.

- You need a consistent analysis throughout: Sentiment can be pretty subjective. If you ask each of your team members to individually read through the same set of customer comments, they'll likely come up with different results. Instead, AI-powered sentiment analysis uses a consistent system to handle all comments equally.

- You need to analyze information in real-time: Sentiment analysis solutions that integrate with your tech stack can import new data as it comes in. Otherwise, you're left to constantly export and upload data whenever you want to perform a new analysis!

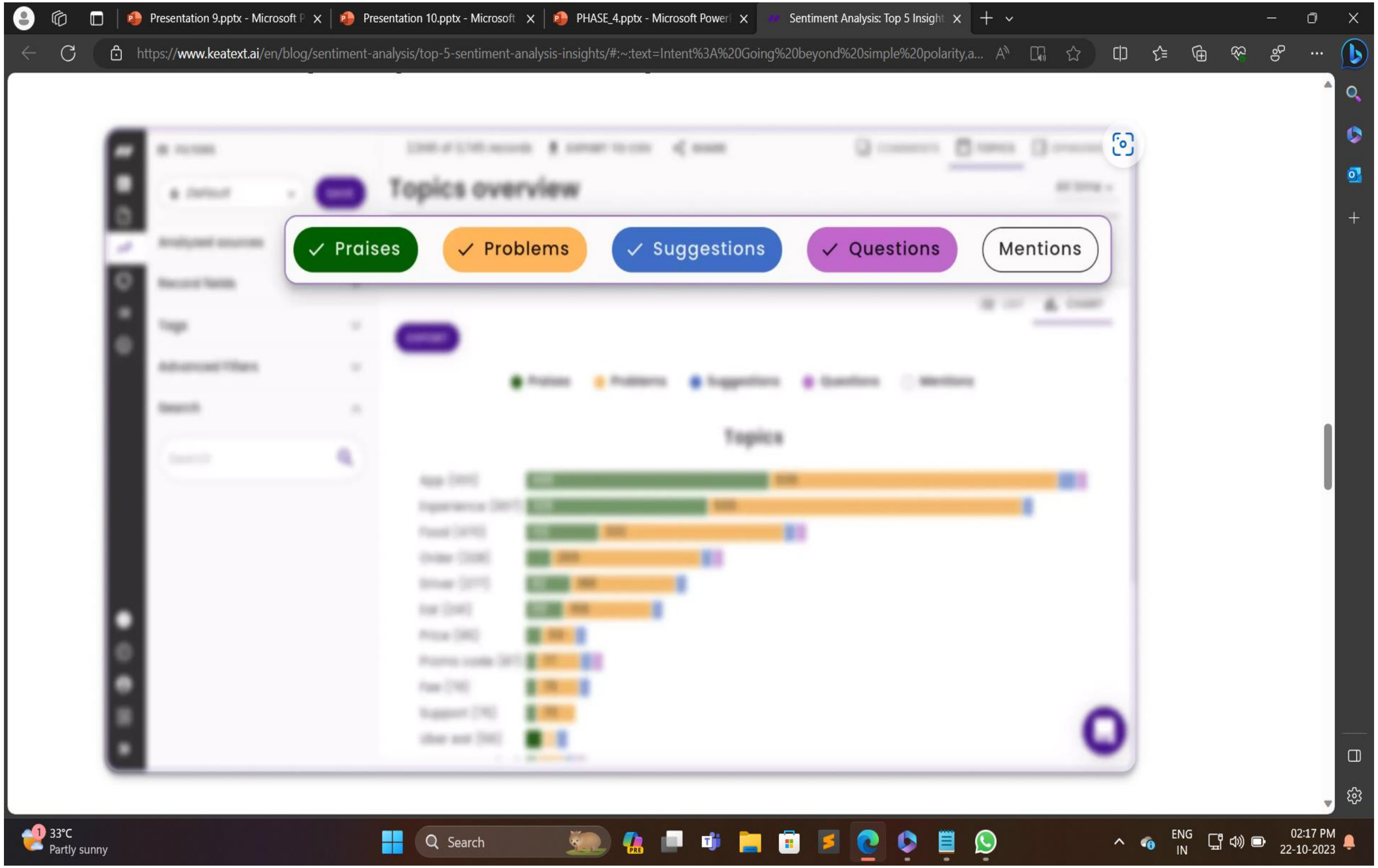
How you can improve their experience?

- Topics: Sentiment analysis can identify the main themes that are coming up in your data – like product quality, shipping, customer service, pricing, and so on. You can get pretty granular by analyzing sentiment for each topic to prioritize where you should focus your efforts. Often times, a single customer comment will cover many topics at once.
- Intent: Going beyond simple polarity (positive or negative), sentiment analysis can uncover where customers are asking questions or even giving suggestions. Looking at intent can provide more qualitative context behind customer comments.
- Emotion: Of course, this is the bread and butter of sentiment analysis. What are your customers feeling in their interactions with your brand? Delight, frustration, excitement, anger? NLP does the heavy lifting here to identify and categorize customer sentiment.

- Root Cause: When customers are having issues, sentiment analysis can get to the root cause, a powerful outcome for any company. Details like at which stage of the customer journey a problem occurs enable you to take action to resolve it quickly and efficiently.
- Sentiment Score: This is a pretty common metric designed to benchmark the general sentiment for analyzed text data. Drawing correlations between trends in the sentiment score and business decisions, new product offerings, or other milestones can indicate how customers felt about a specific product, service, or experience.

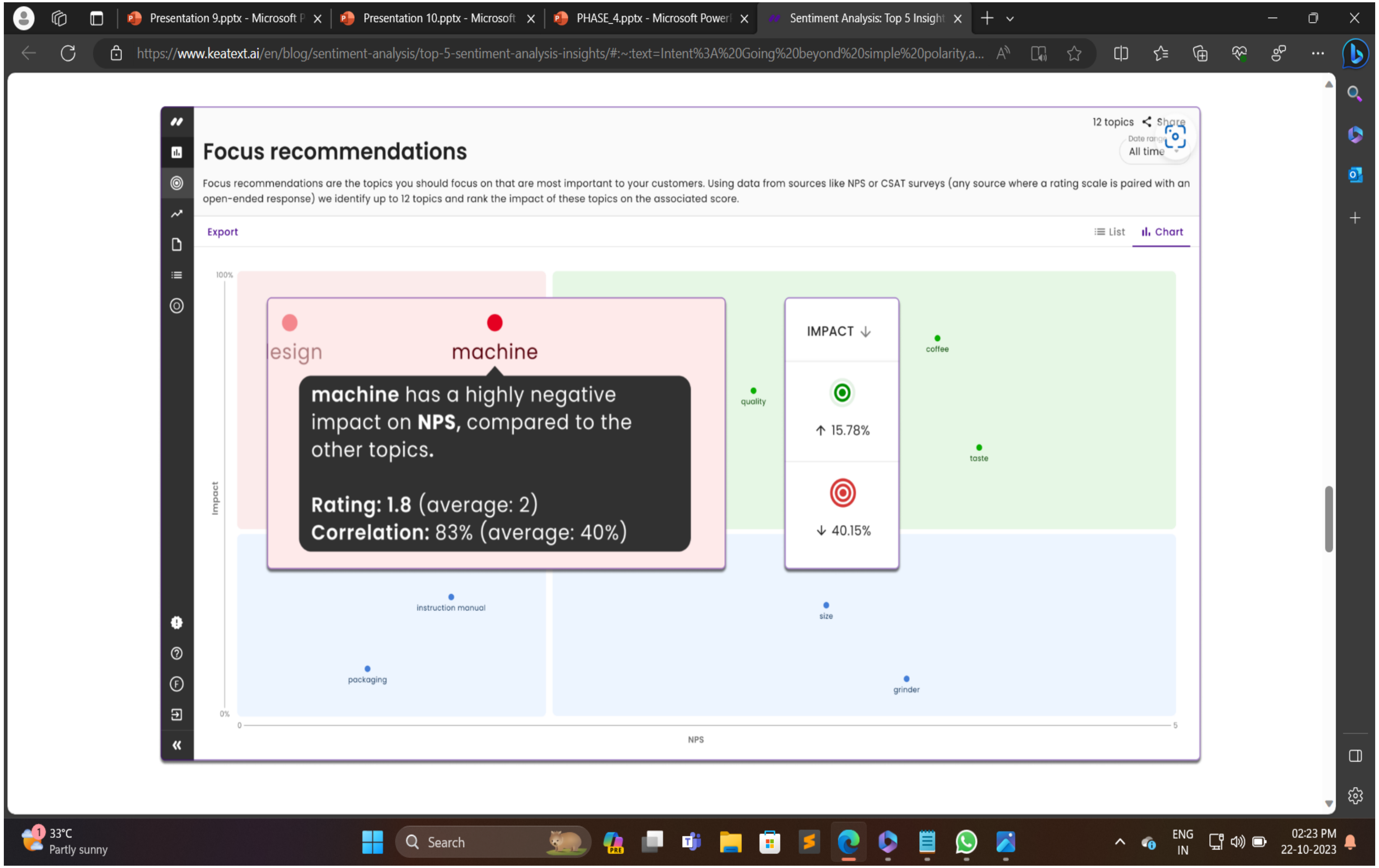
Advanced intent analysis

Keatext's powerful intent analysis identifies praises, problems, questions, and suggestions from customers, so you can really see what customers are feeling. While polarity insights (positive and negative feedback) are certainly useful, what are customers asking for, or even telling you to do in order to improve a product, service, or experience?


Presentation 9.pptx - Microsoft P
Presentation 10.pptx - Microsoft
PHASE_4.pptx - Microsoft Power
Sentiment Analysis: Top 5 Insight
https://www.keatext.ai/en/blog/sentiment-analysis/top-5-sentiment-analysis-insights/#:~:text=Intent%3A%20Going%20beyond%20simple%20polarity,a...
Topics overview
Praises
Problems
Suggestions
Questions
Mentions
Topics
33°C
Partly sunny
Search
ENG
IN
02:17 PM
22-10-2023

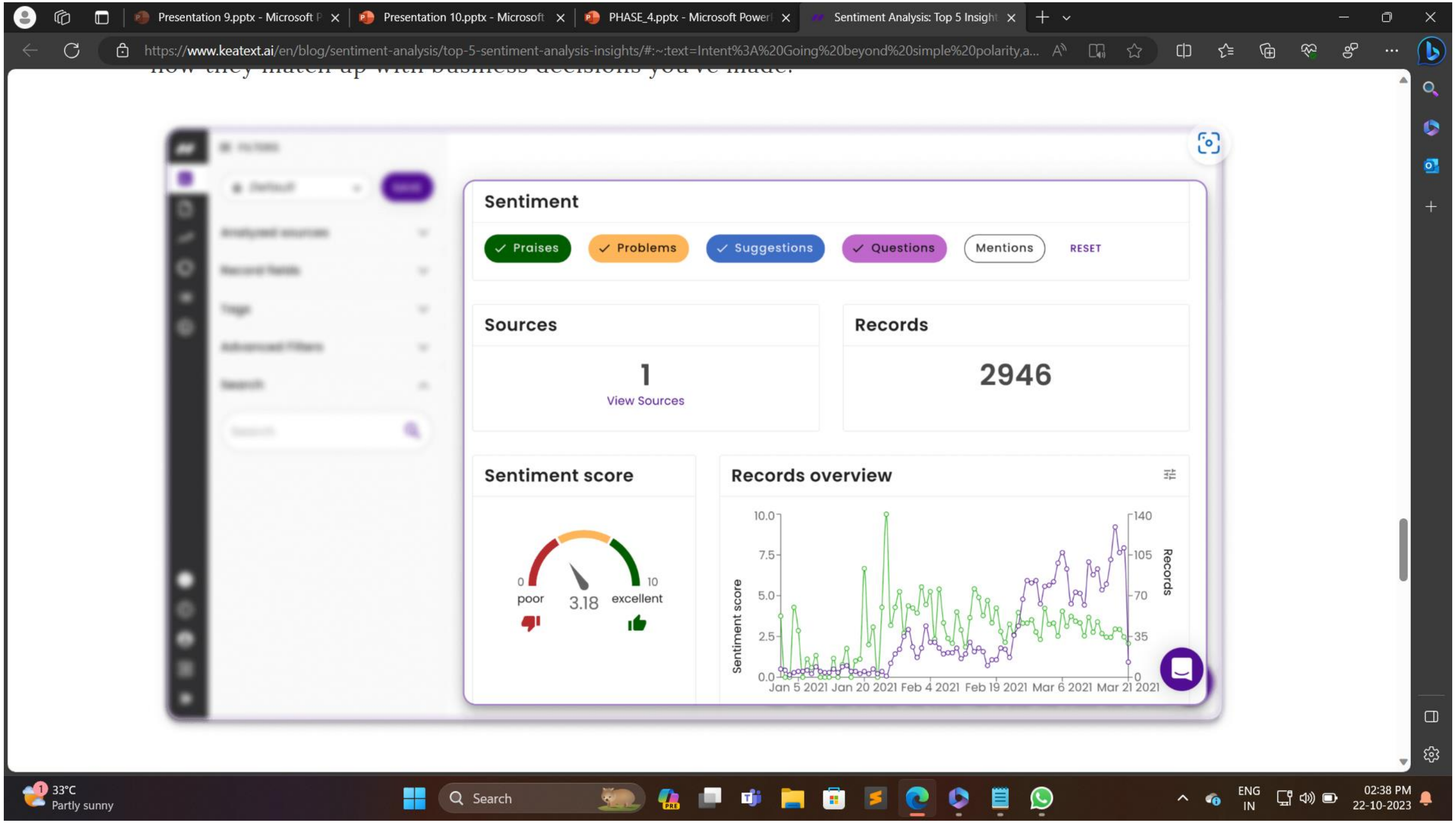
Recommendations for top action items

You can also get powerful recommendations from your data with our platform. Keatext uses predictive analytics to bring together your qualitative customer feedback with the associated CSAT or NPS score to identify areas that have the most impact on customer satisfaction. Leveraging these focus recommendations enables your company to take the most impactful action to improve customer experience.


Presentation 9.pptx - Microsoft P
Presentation 10.pptx - Microsoft
PHASE_4.pptx - Microsoft Power
Sentiment Analysis: Top 5 Insight
https://www.keatext.ai/en/blog/sentiment-analysis/top-5-sentiment-analysis-insights/#:~:text=Intent%3A%20Going%20beyond%20simple%20polarity,a...
12 topics
Share
All time
Focus recommendations
Focus recommendations are the topics you should focus on that are most important to your customers. Using data from sources like NPS or CSAT surveys (any source where a rating scale is paired with an open-ended response) we identify up to 12 topics and rank the impact of these topics on the associated score.
Export
List
Chart
100%
Impact
0%
esign
machine
machine has a highly negative impact on NPS, compared to the other topics.
Rating: 1.8 (average: 2)
Correlation: 83% (average: 40%)
IMPACT
15.78%
40.15%
coffee
quality
taste
instruction manual
size
packaging
grinder
0
5
NPS
33°C
Partly sunny
Search
ENG
IN
02:23 PM
22-10-2023

Sentiment score

- On your Keatext dashboard, it's easy to add a sentiment score widget to see the overall sentiment in your data. The benefit of this is that with our advanced dashboard filtering, you can see the score from a 360 perspective or get as granular as you want for a specific customer segment, topic, etc. Plus, you can monitor changes in the score over time to spot trends and how they match up with business decisions you've made.
- We hope you learned something useful or found your "aha!" moment from our article. If you want to learn more about this topic, check out our Learning Hub for more articles and guides to support your sentiment analysis journey!


Sentiment
Praises
Problems
Suggestions
Questions
Mentions
RESET
Sources
1
View Sources
Records
2946
Sentiment score
0
poor
3.18
10
excellent
Records overview
Sentiment score
Records
Jan 5 2021 Jan 20 2021 Feb 4 2021 Feb 19 2021 Mar 6 2021 Mar 21 2021

GIVEN DATA

•

1	tweet_id	airline_se	airline_se	negativer	negativer	airline	airline_se	name	negativer	retweet_c	text	tweet_co	tweet_cre	tweet_loc	user_timezone
2	6E+17	neutral	1			Virgin America		cairdin		0	@VirginAmerica Wh		######		Eastern Time (US
3	6E+17	positive	0.3486		0	Virgin America		jnardino		0	@VirginAmerica plus		######		Pacific Time (US &
4	6E+17	neutral	0.6837			Virgin America		yvonnalynn		0	@VirginAmerica I dic		######	Lets Play	Central Time (US
5	6E+17	negative	1	Bad Flight	0.7033	Virgin America		jnardino		0	@VirginAmerica it's		######		Pacific Time (US &
6	6E+17	negative	1	Can't Tell	1	Virgin America		jnardino		0	@VirginAmerica and		######		Pacific Time (US &
7	6E+17	negative	1	Can't Tell	0.6842	Virgin America		jnardino		0	@Virgin America seriousl y would pay $30 a flight for seats that didn't have this playing. it's really		######		Pacific Time (US &
8	6E+17	positive	0.6745		0	Virgin America		cjmcginnis		0	@VirginAmerica yes,		######	San Franc	Pacific Time (US &
9	6E+17	neutral	0.634			Virgin America		pilot		0	@VirginAmerica Rea		######	Los Angel	Pacific Time (US &
10	6E+17	positive	0.6559			Virgin America		dhepburn		0	@virginamerica Wel		######	San Diego	Pacific Time (US &
11	6E+17	positive	1			Virgin America		YupitsTate		0	@VirginAmerica it w		######	Los Angel	Eastern Time (US
12	6E+17	neutral	0.6769		0	Virgin America		idk_but_youtube		0	@VirginAmerica did		######	1/1 loner	Eastern Time (US
13	6E+17	positive	1			Virgin America		HyperCamiLax		0	@VirginAmerica I &I		######	NYC	America/New_Yo
14	6E+17	positive	1			Virgin America		HyperCamiLax		0	@VirginAmerica This		######	NYC	America/New_Yo
15	6E+17	positive	0.6451			Virgin America		mollanderson		0	@VirginAmerica @vi		######		Eastern Time (US
16	6E+17	positive	1			Virgin America		sjespers		0	@VirginAmerica Tha		######	San Franc	Pacific Time (US &
17	6E+17	negative	0.6842	Late Fligh	0.3684	Virgin America		smartwatermelon		0	@VirginAmerica SFC		######	palo alto,	Pacific Time (US &
18	6E+17	positive	1			Virgin America		ItzBrianHunty		0	@VirginAmerica So c		######	west covi	Pacific Time (US &
19	6E+17	negative	1	Bad Flight	1	Virgin America		heatherovieda		0	@VirginAmerica I fl		######	this place	Eastern Time (US
20	6E+17	positive	1			Virgin America		thebrandiray		0	I ãū_ flying @VirginA		######	Somewhe	Atlantic Time (Ca
21	6E+17	positive	1			Virgin America		JNLpierce		0	@VirginAmerica you		######	Boston | '	Quito
22	6E+17	negative	0.6705	Can't Tell	0.3614	Virgin America		MISSGJ		0	@VirginAmerica why		######		
23	6E+17	positive	1			Virgin America		DT_Les		0	@VirginA [40.74804		######		
24	6E+17	positive	1			Virgin America		ElvinaBeck		0	@VirginAmerica I lov		######	Los Angel	Pacific Time (US &
25	6E+17	neutral	1			Virgin America		rjlynch21086		0	@VirginAmerica will		######	Boston, M	Eastern Time (US
26	6E+17	negative	1	Customer	0.3557	Virgin America		ayeevickiee		0	@VirginAmerica you		######	714	Mountain Time (
27	6E+17	negative	1	Customer	1	Virgin America		Leora13		0	@VirginAmerica stat		######		
28	6E+17	negative	1	Can't Tell	0.6614	Virgin America		meredithjlynn		0	@VirginAmerica Wh		######		
29	6E+17	neutral	0.6854			Virgin America		AdamSinger		0	@VirginAmerica do y		######	San Franc	Central Time (US
30	6E+17	negative	1	Bad Flight	1	Virgin America		blackjackpro911		0	@VirginA [42.36101		######	San Mateo, CA & Las Vegas,	
31	6E+17	neutral	0.615		0	Virgin America		TenantsUpstairs		0	@VirginA [33.94540		######	Brooklyn	Atlantic Time (Ca
32	6E+17	negative	1	Flight Boo	1	Virgin America		jordanpichler		0	@VirginAmerica hi! I		######		Vienna
33	6E+17	neutral	1			Virgin America		JCervantezzz		0	@VirginAmerica Are		######	California	Pacific Time (US &
34	6E+17	negative	1	Customer	1	Virgin America		Cuschoolie1		0	@VirginA [33.94209		######	Washingt	Quito
35	6E+17	negative	1	Customer	1	Virgin America		amanduhmccarty		0	@VirginAmerica awa		######		Pacific Time (US &
36	6E+17	positive	1			Virgin America		NorthTxHomeTeam		0	@VirginA [33.21450		######	Texas	Central Time (US
37	6E+17	neutral	0.6207			Virgin America		miaerolinea		0	Nice RT @VirginAme		######	Worldwid	Caracas
38	6E+17	positive	1			Virgin America		Nicsplace		0	@VirginAmerica Mo		######	Central Texas	
39	6E+17	positive	1			Virgin America		Nicsplace		0	@VirginAmerica @fr		######	Central Texas	
40	6E+17	neutral	0.6791		0	Virgin America		elisha_malulani		0	@VirginAmerica whe		######	i'm creati	Pacific Time (US &
41	6E+17	negative	1	Customer	1	Virgin America		DannyDouglass		0	@VirginAmerica You		######	San Franc	Pacific Time (US &
42	6E+17	positive	0.6639			Virgin America		jamesferrandini		0	@VirginAmerica Vie		######		
43	6E+17	negative	0.6688	Flight Boo	0.6688	Virgin America		will_lenzenjr		0	@VirginAmerica Hey		######	Iowa City	Central Time (US
44	6E+17	neutral	1			Virgin America		GottAmanda		0	@VirginA [34.02198		######	Los Angeles	
45	6E+17	neutral	0.6578		0	Virgin America		KGervaise		0	@VirginAmerica I ha		######	Georgia	Pacific Time (US &
46	6E+17	neutral	1			Virgin America		papamurat		0	@VirginAmerica are		######		
47	6E+17	positive	1			Virgin America		arieldaie		0	@VirginAmerica I'm		######	Los Angeles	
48	6E+17	neutral	0.6799			Virgin America		vacations7		0	@VirginAmerica DR		######	Turks and caicos	
49	6E+17	positive	1			Virgin America		ChelseaPoe666		0	@VirginAmerica wo		######	Oakland v	Atlantic Time (Ca
50	6E+17	neutral	1			Virgin America		BobGlavinVO		0	@VirginAmerica @la		######	New York	Eastern Time (US
51	6E+17	neutral	0.6436			Virgin America		lisaaiko		0	@VirginAmerica @la		######		
52	6E+17	neutral	0.6764		0	Virgin America		grantbrowne		0	@VirginAmerica Is fl		######	Worldwid	Central Time (US
53	6E+17	positive	0.657			Virgin America		joyabsalon		0	@VirginAmerica @la		######	Northern	Eastern Time (US
54	6E+17	neutral	1			Virgin America		2v		0	@VirginAmerica wisl		######	Los Angel	Eastern Time (US
55	6E+17	neutral	0.7118		0	Virgin America		KSmithFoundHere		0	@VirginAmerica @la		######		Atlantic Time (Ca
56	6E+17	neutral	1			Virgin America		papamurat		0	@VirginAmerica Will		######		
57	6E+17	negative	0.6939	Flight Boo	0.6939	Virgin America		murphicus		0	@VirginAmerica hi! i		######	new york,	Eastern Time (US
58	6E+17	positive	1			Virgin America		VinnieFerra		0	@VirginAmerica you		######	brooklyn,	Pacific Time (US &
59	6E+17	positive	0.635			Virgin America		KevinDemsi		0	@VirginAmerica @la		######	Bali, Repu	Kuala Lumpur

- Installing NLTK and downloading the data, Tokenizing the data, Normalizing the data

```
pip install nltk==3.3
python3
import nltk
nltk.download('twitter_samples')
nano nlp_test.py
from nltk.corpus import twitter_samples
from nltk.corpus import twitter_samples

positive_tweets = twitter_samples.strings('positive_tweets.json')
negative_tweets = twitter_samples.strings('negative_tweets.json')
text = twitter_samples.strings('tweets.20150430-223406.json')
python3
import nltk
nltk.download('punkt')
from nltk.corpus import twitter_samples
```

```
positive_tweets = twitter_samples.strings('positive_tweets.json')
negative_tweets = twitter_samples.strings('negative_tweets.json')
text = twitter_samples.strings('tweets.20150430-223406.json')
tweet_tokens = twitter_samples.tokenized('positive_tweets.json')
from nltk.corpus import twitter_samples
positive_tweets = twitter_samples.strings('positive_tweets.json')
negative_tweets = twitter_samples.strings('negative_tweets.json')
text = twitter_samples.strings('tweets.20150430-223406.json')
tweet_tokens = twitter_samples.tokenized('positive_tweets.json')[0]
print(tweet_tokens[0])
python3 nlp_test.py
```

OUTPUT

```
['#FollowFriday',
'@France_Inte',
'@PKuchly57',
```

```
'@Milipol_Paris',
'for',
'being',
'top',
'engaged',
'members',
'in',
'my',
'community',
'this',
'week',
':)']
---------------------------------------------------------------------------------------------------
from nltk.corpus import twitter_samples

positive_tweets = twitter_samples.strings('positive_tweets.json')
negative_tweets = twitter_samples.strings('negative_tweets.json')
```

```
text = twitter_samples.strings('tweets.20150430-223406.json')
tw8eet_tokens = twitter_samples.tokenized('positive_tweets.json')[0]
#print(tweet_tokens[0])
python3
import nltk
nltk.download('wordnet')
nltk.download('averaged_perceptron_tagger')
from nltk.tag import pos_tag
from nltk.corpus import twitter_samples
tweet_tokens = twitter_samples.tokenized('positive_tweets.json')
print(pos_tag(tweet_tokens[0]))
```

OUTPUT

```
[('#FollowFriday', 'JJ'),
('@France_Inte', 'NNP'),
```

```
('@PKuchly57', 'NNP'),
('@Milipol_Paris', 'NNP'),
('for', 'IN'),
('being', 'VBG'),
('top', 'JJ'),
('engaged', 'VBN'),
('members', 'NNS'),
('in', 'IN'),
('my', 'PRP$'),
('community', 'NN'),
('this', 'DT'),
('week', 'NN'),
(':)', 'NN')]
------------------------------------------------------------------------------------------------------------
```

```
from nltk.tag import pos_tag
from nltk.stem.wordnet  import WordNetLemmatizer

def lemmatize_sentence(tokens):
    lemmatizer  = WordNetLemmatizer()
    lemmatized_sentence  = []
    for word, tag in pos_tag(tokens):
        if tag.startswith('NN'):
            pos = 'n'
        elif tag.startswith('VB'):
            pos = 'v'
        else:
            pos = 'a'
        lemmatized_sentence.append(lemmatizer.lemmatize(word, pos))
    return lemmatized_sentence
```

```
print(lemmatize_sentence(tweet_tokens[0]))
python3 nlp_test.py
```

OUTPUT

```
['#FollowFriday',
 '@France_Inte',
 '@PKuchly57',
 '@Milipol_Paris',
 'for',
 'be',
 'top',
 'engage',
 'member',
 'in',
 'my',
```

```
'community',
'this',
'week',
':)']
--------------------------------------------------------------------------------
import re, string
def remove_noise(tweet_tokens, stop_words = ()):
    cleaned_tokens = []
for token, tag in pos_tag(tweet_tokens):
token = re.sub('http[s]?://(?:[a-zA-Z]|[0-9]|[$-_@.&+#]|[!*\(\),]|'\
                   '(?:%[0-9a-fA-F][0-9a-fA-F]))+','', token)
        token = re.sub("(@[A-Za-z0-9_]+)","",token)
```

```
if tag.startswith("NN"):
            pos = 'n'
        elif tag.startswith('VB'):
            pos = 'v'
        else:
            pos = 'a'
        lemmatizer = WordNetLemmatizer()
        token = lemmatizer.lemmatize(token, pos)
        if len(token) > 0 and token not in string.punctuation and token.lower() not in stop_words:
            cleaned_tokens.append(token.lower())
    return cleaned_tokens
nltk.download('stopwords')
from nltk.corpus import stopwords
```

```
stop_words = stopwords.words('english')
print(remove_noise(tweet_tokens[0], stop_words))
```

OUTPUT

```
['#followfriday', 'top', 'engage', 'member', 'community', 'week', ':)']
```

CONCLUSION

NLP uses computational methods to interpret and comprehend human language. National language processing is the field of computer science and linguistic that deals with the interaction between computers and human languages