CLIMATE CHANGE MODELING

CREATED BY: DIVYA KRISHNAKUMAR

TYPE OF PROJECT: INTERNSHIP PROJECT

TOOLS USED: PYTHON

SUBMITTED TO: UNIFIED MENTOR PVT LTD

DATE: 29/07/2025

EXECUTIVE SUMMARY

The project involves NLP(Natural Language Processing). The aim of this project is to gauge public opinion on climate change and NASA's communication strategies and to predict public opinion on unseen data. It also aims to identify shifts in public sentiments. The methodology followed consists of following steps: Data Collection, Data Wrangling, Data analysis and Visualisation, Model building, Evaluation.

The data is collected and loaded to a dataframe. The columns in the dataset were- date, likesCount, profileName, commentsCount, text. The dataset is checked for null values. The null values in the numerical columns are substituted with 0.0 and that of categorical columns are substituted with 'Unknown'. The date column is changed to datetime type. The text data is cleaned by removing emojis and wild card characters. The sentences are tokenized and tokens are created. Then the tokens are subjected to sentiment analysis. The distribution of public sentiments over NASA's communication strategies is visualised. The trends of shifts in public sentiments over a period is visualised.

During model building RNN(Recurrent Neural Network) model is built and results are predicted on unseen data. A model of Support Vector Machine is built and results predicted. Model is evaluated by displaying a classification report and a confusion matrix.

INTRODUCTION

The aim of the project is to analyse and predict public sentiments on climate change and NASA's Communication strategies. The project also involves analysis of trends in shifts in public opinions.

The key questions are :

- How is public sentiments on climate change and NASA's communication strategies are distributed?
- What are the trends of shifts in public sentiments over a period?

METHODOLOGY

The methodology followed in the project involves the following steps: Data Collection, Data Wrangling, Data analysis and Visualisation, Model building, Evaluation.

The data is cleaned, analysed, visualised and subjected to machine learning algorithms such as sequential RNN, Support Vector Machine.

Data collection: Data collected is loaded to a dataframe. The columns in the data set were: date, likesCount, profileName, commentsCount, text.

DATA WRANGLING

The null values are replaced with 0.0 in numerical columns and 'Unknown' in categorical columns. The data type of date column is changed to datetime. The emojis and wild card characters in the text column are removed. The sentences in the text column are tokenized and tokens are generated. Thus created tokens were used for sentiment analysis, trend analysis and machine learning. The dataframe after replacing null values is as follows:

	date	likesCount	profileName	commentsCount	text
0	2022-09-07 17:12:32+00:00	2	4dca617d86b3fdce80ba7e81fb16e048c9cd9798cdfd6d...	0.0	Neat comparison I have not heard it before.\n ...
1	2022-09-08 14:51:13+00:00	0	518ab97f2d115ba5b6f03b2fba2ef2b120540c9681288b...	0.0	An excellent way to visualise the invisible! T...
2	2022-09-07 17:19:41+00:00	1	d82e8e24eb633fd625b0aef9b3cb625cfb044ceb8483e1...	3.0	Does the CO2/ghg in the troposphere affect the...
3	2022-09-08 00:51:30+00:00	4	37a509fa0b5177a2233c7e2d0e2b2d6916695fa9fba3f2...	0.0	excellent post! I defo feel the difference - o...
4	2022-09-07 19:06:20+00:00	16	e54fbbd42a729af9d04d9a5cc1f9bbfe8081a31c219ecb...	26.0	Yes, and carbon dioxide does not harm the Eart...

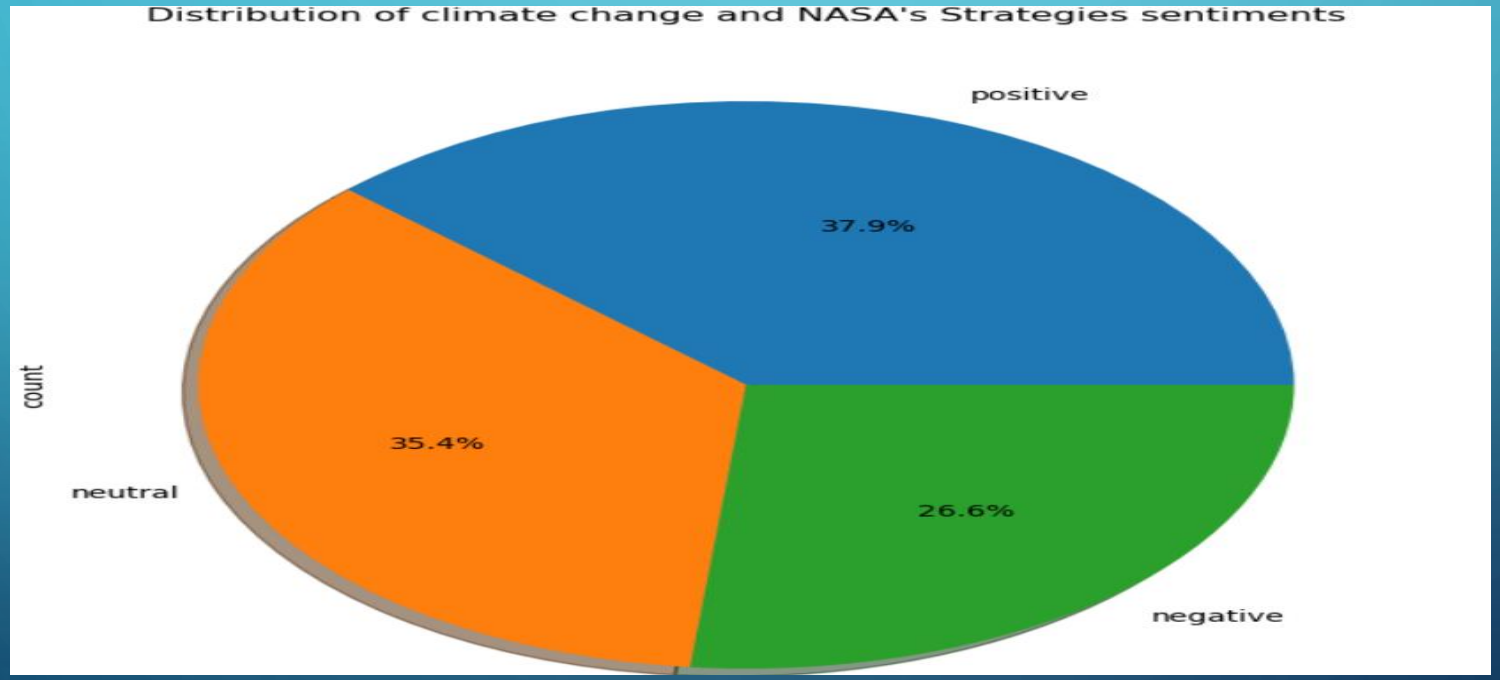
DATA WRANGLING

The text column after removal of emojis and wild card characters

Neat comparison I have not heard it before.
I would say that CO2 is like a disco ball so energy and collide with countless molecules as it bounces around in our atmosph re.
An excellent way to visualise the invisible Thanks
Does the CO2 ghg in the troposphere affect the stratosphere Being warmer the troposphere must be expanding
excellent post I defo feel the difference one i can drink and the other makes me feel sick.
Yes and carbon dioxide does not harm the Earth like you people are making it out to be. Fake climate change rhetoric as u ual.
Excellent example very relatable to the Gen Pop. Complex topics explained in understandable ways.
I am so proud of whoever created this content. Please feel very very good about yourself today
Has the Earth ever had a concentration of over 1000 ppm
Yes and not even half a cup later that poor soul is gonna know it too.
Historical co2 data would make it an interesting post
why the northern hemisphere warm more than the southern hemisphere
So what are we gonna do about it
Yes. I fall asleep on decaf and stay awake with real coffee. You can keep your brown sadness water thanks.
Plus you can literally measure the heat retention effect of CO2 in a glass container in a lab
The two most abundant gases nitrogen and oxygen comprise 99 of the atmosphere. Neither traps heat. By definition all g eenhouse gases are trace gases because they live in the remaining 1 . It s too bad the CO2 is naturally so sparse. If CO2 we

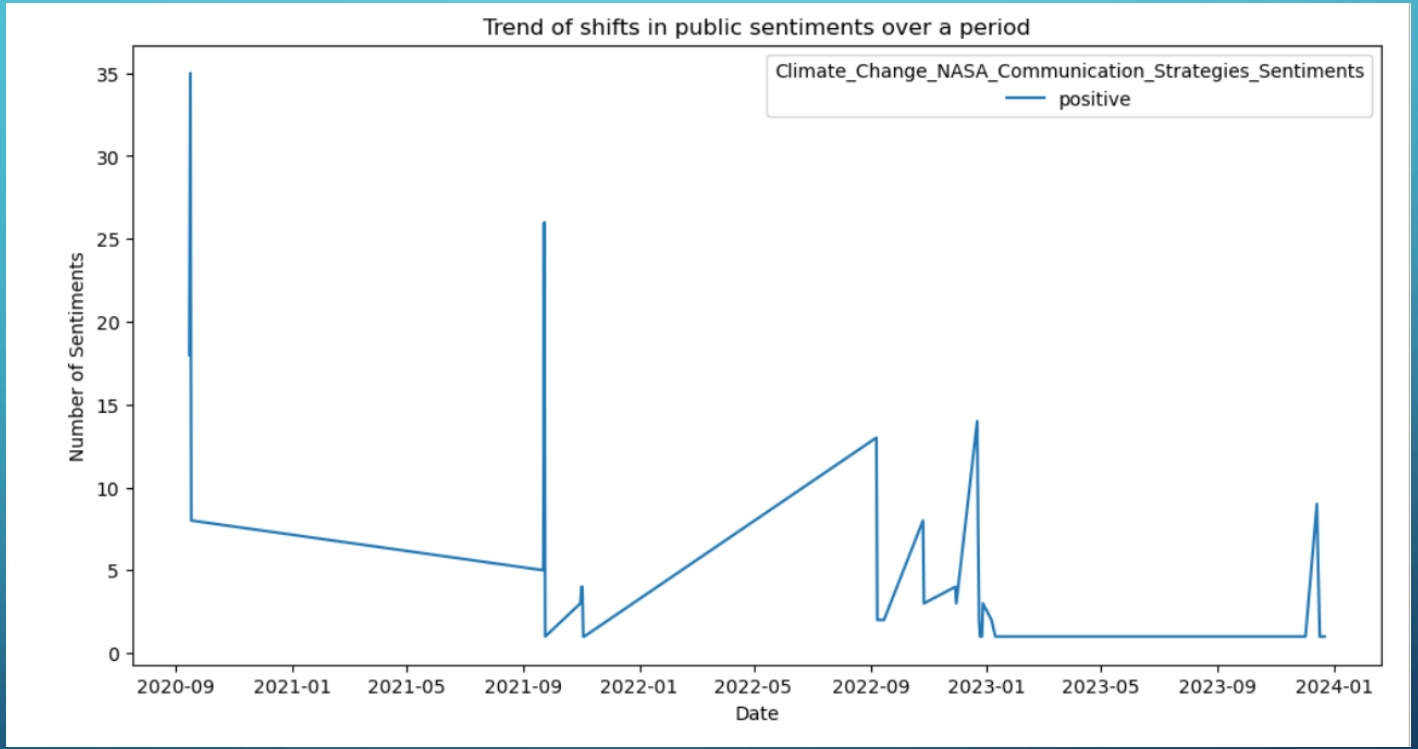
SENTIMENT ANALYSIS AND VISUALISATION

The generated tokens are subjected to generate sentiments. A pie chart is displayed to show the distribution of various sentiments.



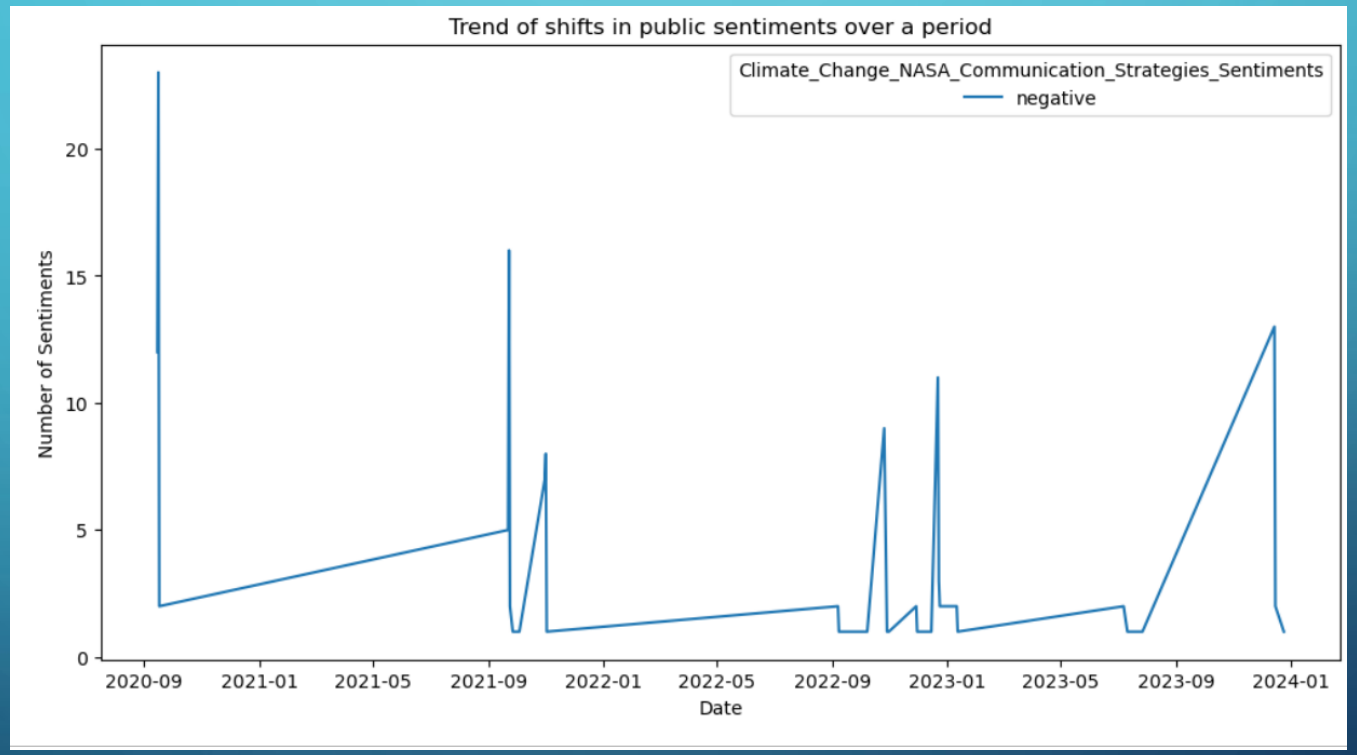
TREND ANALYSIS AND VISUALISATION

Trends of positive sentiments



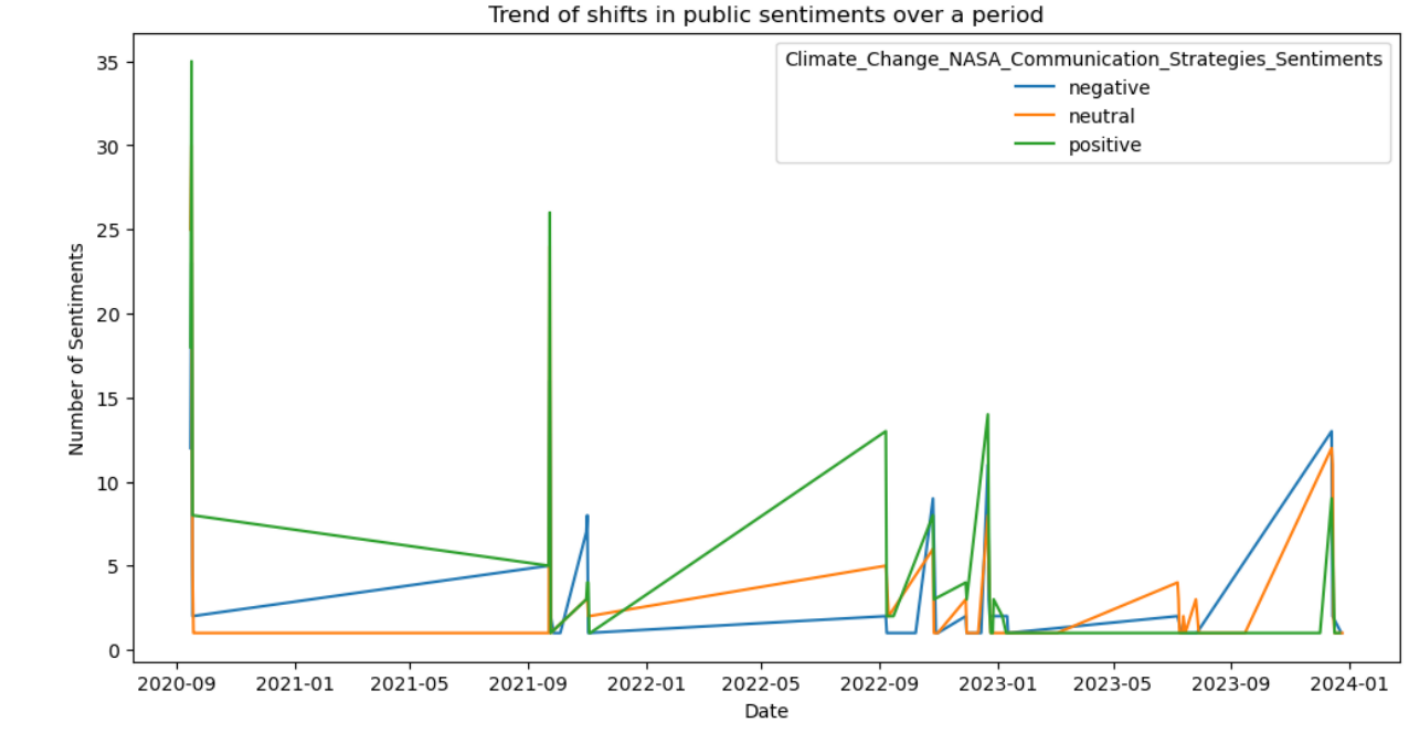
TREND ANALYSIS AND VISUALISATION

Trend of negative sentiments



TREND ANALYSIS AND VISUALISATION

Trend of sentiments over a period



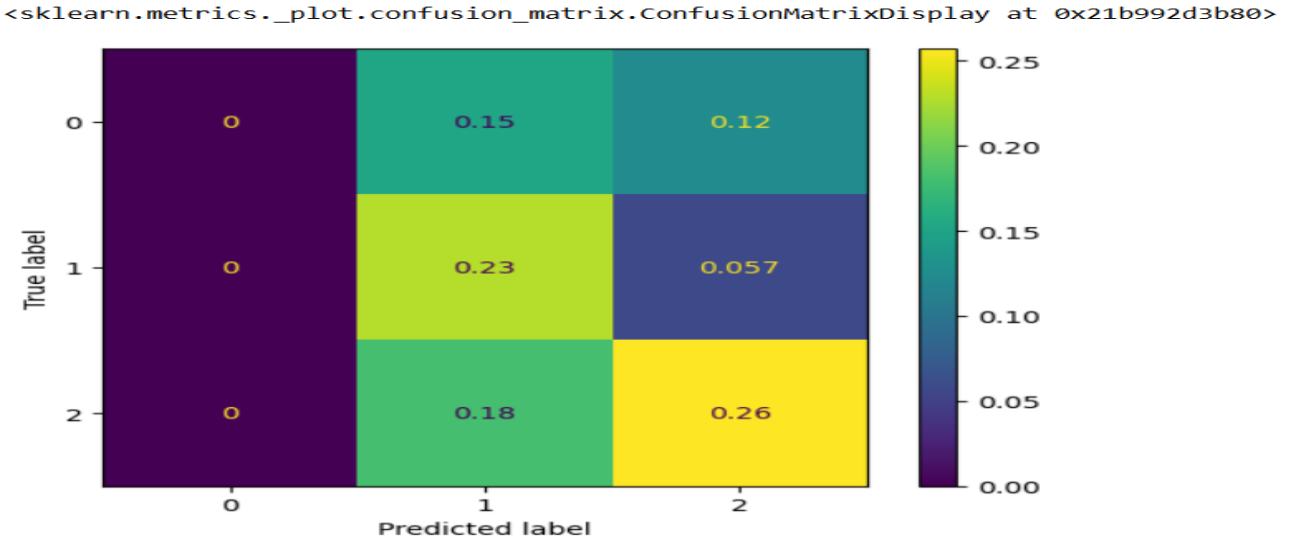
MODEL BUILDING AND EVALUATION

- A model is developed with sequential RNN(Recurrent Neural Network) and predicted the sentiment results The accuracy obtained were 40%.
- A model of Support Vector Machine is developed and fitted. Then the model predicted results on unseen data. The model's accuracy was displayed. The model was evaluated by generating a classification report and confusion matrix display. The classification report is as follows:

	precision	recall	f1-score	support
negative	0.00	0.00	0.00	29
neutral	0.41	0.80	0.54	30
positive	0.59	0.59	0.59	46
accuracy			0.49	105
macro avg	0.33	0.46	0.38	105
weighted avg	0.37	0.49	0.41	105

MODEL BUILDING AND EVALUATION

A normalised confusion matrix is displayed.



The accuracy score of svm model is 48.57%.

FINDINGS AND IMPLICATIONS

- The percentage of positive sentiments is 37.9%
- The percentage of negative sentiments is 35.4%
- The percentage of neutral sentiments is 35.4%.
- Initially positive sentiments were on rise, negative and neutral sentiments were less. Then total sentiments has a decreasing trend and all the three sentiments became almost equal.
- From the predictive results positives are 27 and neutral are 24 out of a total of 51 correct classification.

CONCLUSION

The percentage of positive and neutral sentiments were higher than negative sentiments. The trend of positive, negative and neutral sentiments were decreasing over period of time. Initially positive sentiments were higher in number and over a period of time positive, negative and neutral sentiments decreased and became almost he same.

From predictive analysis it can be concluded that positive sentiments are highest in number regarding NASA's climate change and communication strategies.

Future implications and suggestions:

- Advise or suggestion or educative information or a great thought could increase positive sentiments.