UNIFIED MENTOR
YOUR SKILL, SUCCESS & JOURNEY

PROJECT – 3

NAME - AAGASH.M
UNID - UMID11062542183

PROJECT TITLE – “Climate Change Modeling”

Abstract

This project conducts a comprehensive analysis of public discourse on climate change using Natural Language Processing (NLP) and machine learning techniques. The study utilizes a dataset of over 500 user comments from NASA's official Climate Change Facebook page, posted between 2020 and 2023. The primary objectives were to perform an exploratory data analysis (EDA) to understand data characteristics, conduct sentiment analysis to classify public opinion, and apply topic modeling to identify key themes in the conversation. Methodologies included text preprocessing, VADER for sentiment scoring, and Latent Dirichlet Allocation (LDA) for topic extraction. Key findings indicate a mixed sentiment distribution, with 43.2% neutral, 32.1% positive, and 24.7% negative comments, suggesting the discourse is less polarized than often assumed. Topic modeling identified five major themes: general climate science, the impact of warming on ice and sea levels, public skepticism, long-term effects, and environmental solutions. The results highlight a significant public interest in factual information and tangible environmental impacts, providing valuable insights for science communicators and policymakers.

Table of Contents

1. Introduction

1.1. Background

Climate change is one of the most pressing issues of our time, sparking widespread public discourse across various platforms. Social media has become a primary arena for these conversations, offering a rich source of data for understanding public perception, concerns, and misconceptions. NASA, as a leading scientific authority, plays a crucial role in disseminating information about climate change, and the public's engagement with their content provides a unique window into this discourse.

1.2. Problem Statement

Understanding the nuances of public discourse on climate change is essential for scientists, policymakers, and communicators to tailor their outreach and

educational strategies effectively. Raw text data from social media is unstructured and voluminous, making manual analysis impractical. This project addresses the need for an automated, data-driven approach to extract meaningful insights from public comments on NASA's climate change communications.

2. Objectives

By applying Natural Language Processing (NLP) and machine learning techniques, this project aims to analyze the provided dataset to achieve the following primary objectives:

- **Exploratory Data Analysis (EDA):** To perform an initial analysis to understand the dataset's underlying characteristics, such as comment frequency over time and user engagement patterns.
- **Sentiment Analysis:** To automatically gauge the overall public opinion within the comments and classify each comment as positive, negative, or neutral.
- **Trend Analysis:** To identify potential shifts or patterns in public sentiment and discussion topics over the specified period (2020-2023).
- **Topic Modeling:** To discover and interpret the most prevalent themes and subjects discussed by the public in relation to NASA's climate communications.

3. Methodology

3.1. Dataset Description

The dataset contains over 500 user comments posted on NASA's Climate Change Facebook page between 2020 and 2023. The data includes the comment text, date of posting, and engagement metrics.

Key Columns:

- date: Timestamp of the comment.
- likesCount: Number of likes on the comment.
- text: The raw text content of the comment.

3.2. Data Preprocessing

Before analysis, the text data required significant cleaning to ensure the accuracy of the NLP models. The following steps were taken:

1. **Handling Missing Values:** Removed rows with no comment text.
2. **Text Cleaning:** Converted all text to lowercase. Removed URLs, special characters, punctuation, and numbers.
3. **Tokenization:** Broke down sentences into individual words (tokens).
4. **Stop Word Removal:** Eliminated common English stop words (e.g., "the", "a", "is").
5. **Lemmatization:** Reduced words to their base or root form (e.g., "running" becomes "run") to standardize the vocabulary.

3.3. Exploratory Data Analysis (EDA)

EDA was conducted to identify initial patterns and trends. This involved visualizing the distribution of likesCount and plotting the frequency of comments over the years to perform a basic trend analysis.

3.4. Sentiment Analysis

We used the VADER (Valence Aware Dictionary and sEntiment Reasoner) sentiment analysis tool, which is specifically tuned for social media text. VADER calculates a compound score for each comment, which is a normalized score between -1 (most negative) and +1 (most positive). These scores were used to classify comments into three categories.

3.5. Topic Modeling

To identify the underlying themes in the comments, we employed Latent Dirichlet Allocation (LDA), a popular unsupervised learning algorithm for topic modeling. LDA groups words that frequently co-occur into distinct topics, allowing us to infer the main themes of the corpus.

4. Results and Discussion

4.1. Exploratory Data Analysis Findings

The EDA revealed that the distribution of likesCount is heavily skewed, with most comments receiving between 0 and 10 likes, indicating that virally popular comments are rare. The time-series analysis of comment frequency shows fluctuating but consistent engagement over the years, providing a steady stream of data for analysis.

4.2. Sentiment Analysis Results

The analysis revealed a mixed but predominantly neutral and positive sentiment among the commenters.

- **Positive Sentiment (32.1%):** Comments expressing appreciation for NASA's work, agreement with findings, or general positivity.
- **Negative Sentiment (24.7%):** Comments showing skepticism, criticism of climate science, or expressing concern and fear about the future.
- **Neutral Sentiment (43.2%):** Comments that were purely factual, asking questions, or did not express a strong emotion.

This distribution suggests the discourse is not entirely polarized. The large portion of neutral engagement presents an opportunity for public education and outreach.

4.3. Topic Modeling Results

The LDA model identified 5 key topics from the comment corpus.

Identified Topics:

- **Topic 1: General Climate Science & CO2:** (Keywords: climate, change, co2, science, data, nasa, earth, temperature) - General discussions about the science of climate change, CO2 levels, and NASA's role.
- **Topic 2: Ice, Sea Levels & Warming:** (Keywords: ice, water, sea, level, warm, melt, heat, arctic) - Focused on the consequences of warming, such as melting ice caps and rising sea levels.
- **Topic 3: Skepticism & Alternative Theories:** (Keywords: people, think, believe, know, like, money, make, real) - This topic captures skeptical viewpoints, questioning the consensus or attributing climate change to other factors.
- **Topic 4: Time, Years & Future Impact:** (Keywords: year, time, ago, long, cause, effect, human, world) - Discussions centered around the timeline of climate change, historical data, and its long-term impact.
- **Topic 5: Energy & Environmental Solutions:** (Keywords: energy, sun, plant, carbon, use, need, food, fuel) - Conversations about solutions, alternative energy sources, and the role of carbon in the environment.

5. Conclusion and Future Work

5.1. Conclusion

This analysis provides a valuable snapshot of public discourse on climate change within a specific social media context. The sentiment is mixed, with a substantial number of neutral comments indicating public curiosity and a desire for information. The topic modeling reveals that the conversation is multifaceted, covering a wide range of subjects from core climate science to skepticism and potential solutions.

Key Takeaways:

- Public sentiment is not as polarized as often assumed.
- There is a clear need for accessible information, as shown by the high volume of neutral, question-based comments.
- Discussions frequently revolve around tangible effects like melting ice and sea-level rise.

5.2. Future Work

- **Advanced Trend Analysis:** Analyze the evolution of topics and sentiment over time more granularly to identify specific trends.
- **Engagement Analysis:** Correlate comment sentiment and topics with the specific content of NASA's posts to understand what drives engagement.
- **Advanced NLP Models:** Use more advanced contextual models like BERT for a deeper, more nuanced understanding of the comments.

DATA SET - https://drive.google.com/file/d/1lBYJ_y26zjS0yEzG2pEZDFJbriXRwKsF/view?usp=sharing

DATA VISVALIZATION COLAB LINK

https://colab.research.google.com/drive/1lgeqkHF3mLGuXsBIC-_KXLB_PhfUfWSV?usp=sharing