IUPUI
SCHOOL OF INFORMATICS AND COMPUTING

Date: 4th December 2023

Election Pulse: Unveiling Voter Sentiments for 2024 US Election Forecasting

GROUP – 19
Hemanth Kumar Reddy Gunnam
Supraja Pericherla
Durga Pravallika Kuchipudi

INTRODUCTION

Aim and Objective:

This project aims to predict the outcome of the 2024 elections using vast amounts of data available online and advanced sentiment analysis techniques. By analyzing public opinion and sentiment from various online sources, we seek to gain insights into voter preferences and trends.

Understanding Public Opinion:

- In the modern political landscape, public opinion is often fragmented and rapidly changing. Our project seeks to capture this complexity through real-time sentiment analysis.
- The significance of this approach lies in its ability to provide a more accurate and nuanced understanding of voter sentiment, beyond what traditional polling methods can offer.

INTRODUCTION

The Role of Sentiment Analysis in Elections:

- Sentiment analysis allows us to process and interpret large-scale online discussions, opinions, and reactions related to political events and candidates.

- This approach is crucial for identifying emerging trends, voter concerns, and overall sentiment towards political entities, which are key indicators for election predictions.

Relevance in the Digital Age:

- In an era where social media and online platforms play a significant role in shaping public opinion, this project's approach is particularly relevant.

- It represents a step towards embracing more technologically advanced and data-driven methods in the field of political forecasting.

DATASET

Reddit Data Mining:

Subreddit Focused: Exploration centered on the 'politics' subreddit, a hub for political discourse.
Search Query: Used 'US elections 2024' to extract relevant discussions, opinions, and predictions from the subreddit.

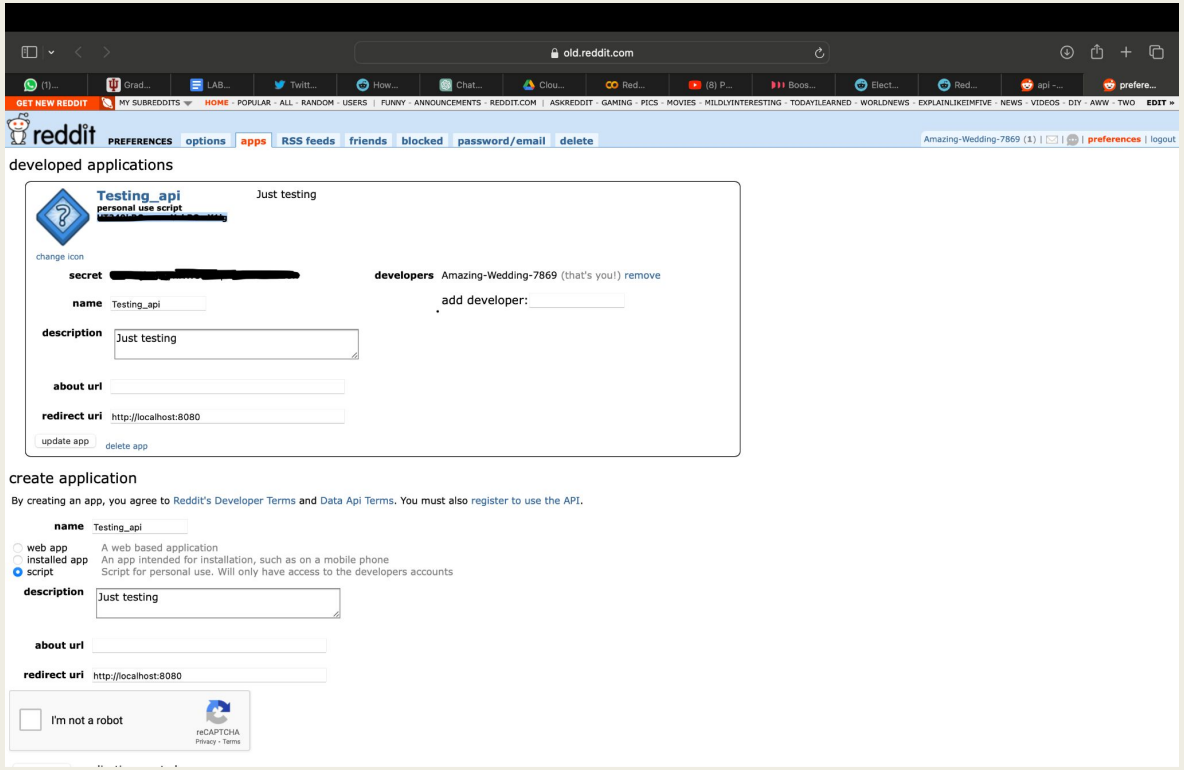


Fig 1: Reddit API

DATASET

Google API Extraction:

Broad Scope: Queries cast a wide net to capture diverse perspectives and information on the subject.
Specific Queries:

- 'US elections 2024' to gather general content related to the upcoming elections.
- '2024 US Presidential Election Candidates' to obtain information on the individuals participating in the race.
- 'US Election 2024 Predictions and Polls' to collect data on forecasts and public opinion polls.

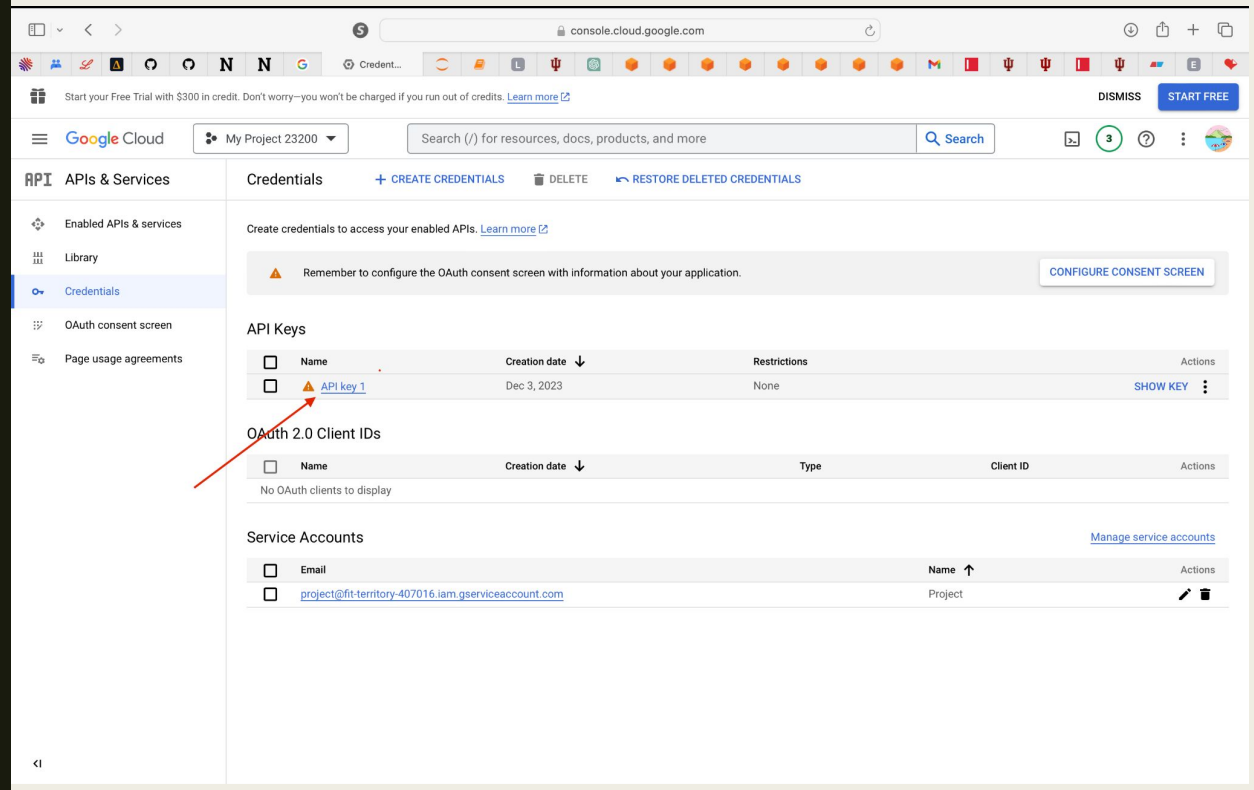


Fig 2: Google API

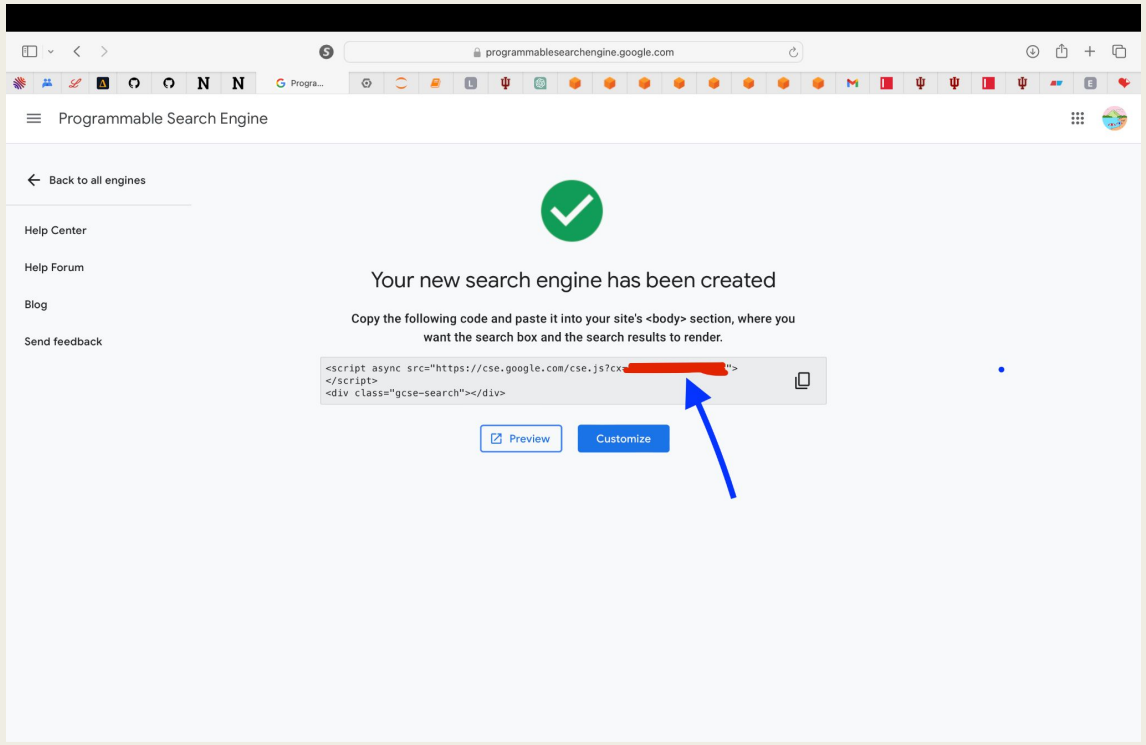


Fig 3: Google API search engine

Sources used for retrieving the data

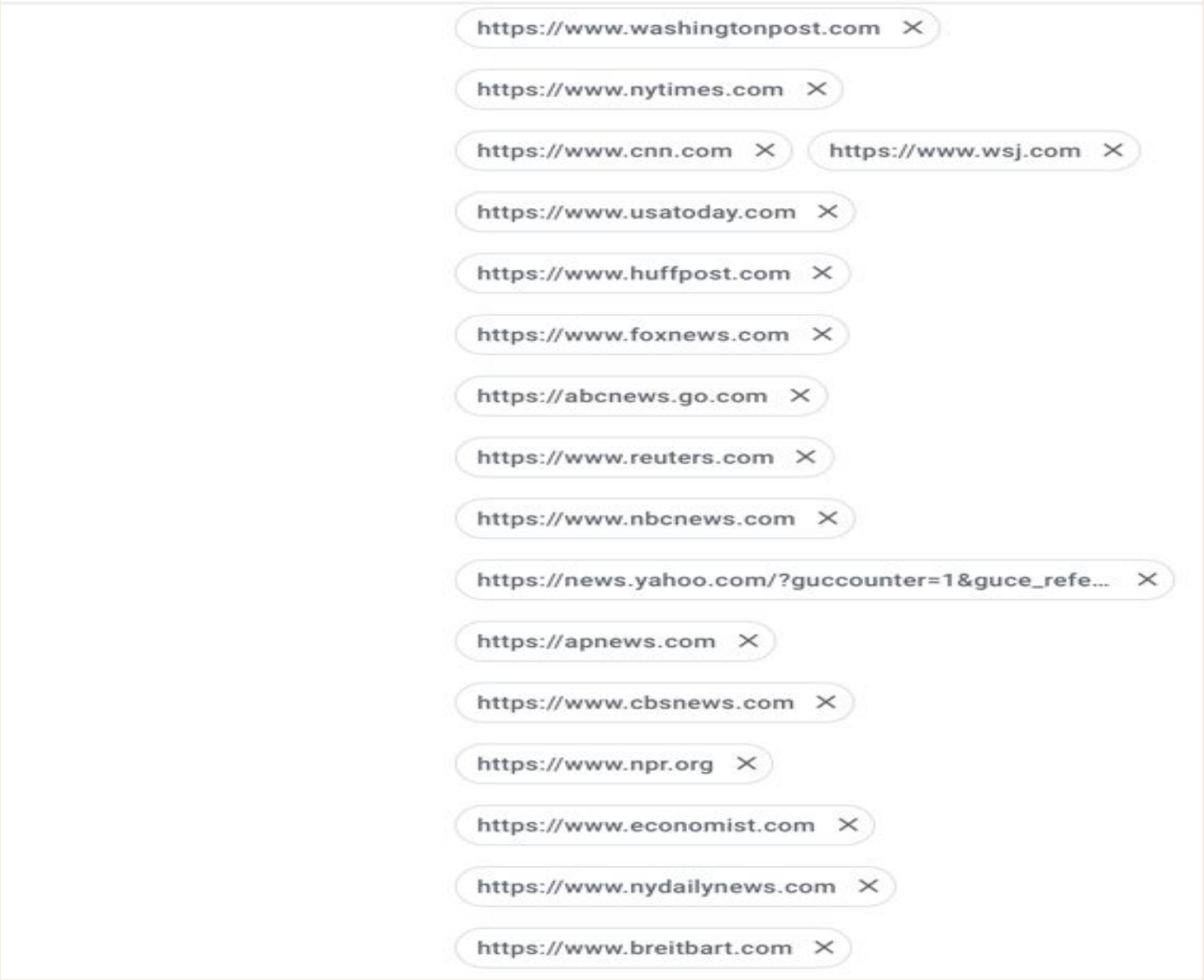


Fig 4: Source links

Comprehensive Methodology

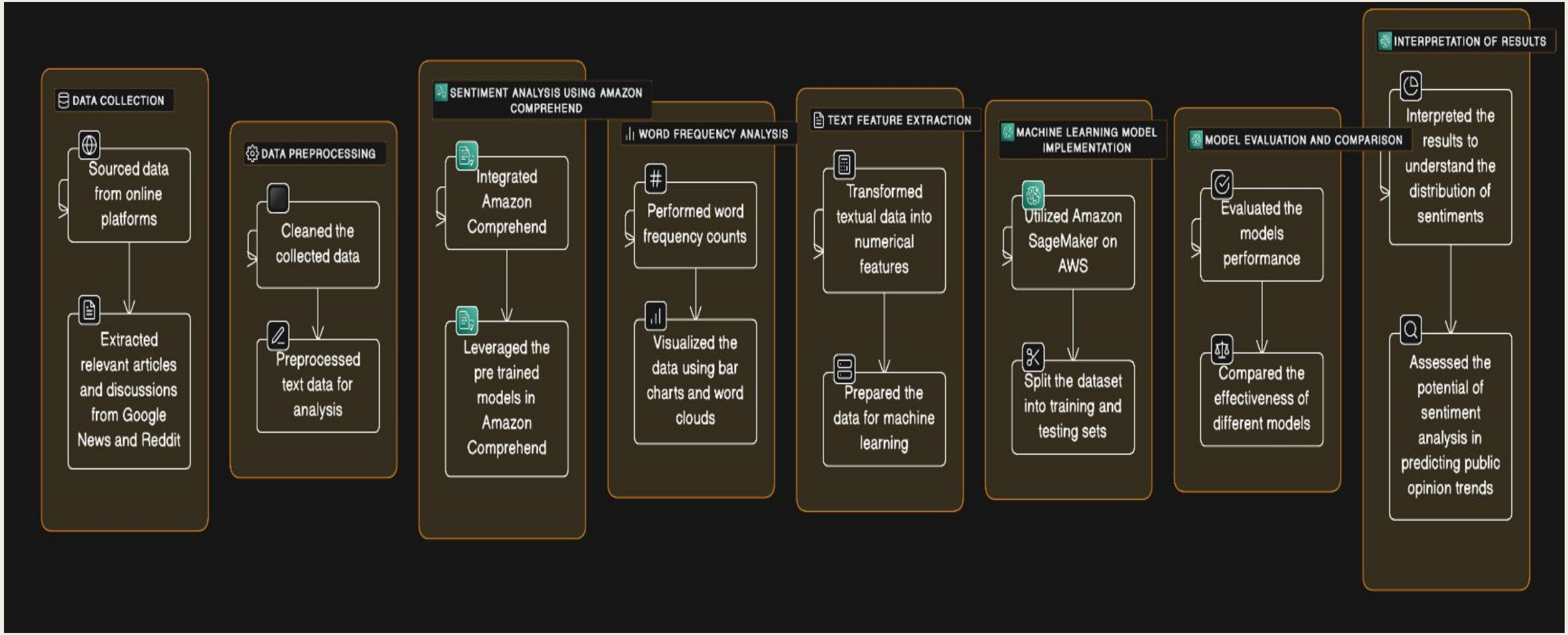


Fig 5: Methodology

Cloud Architecture

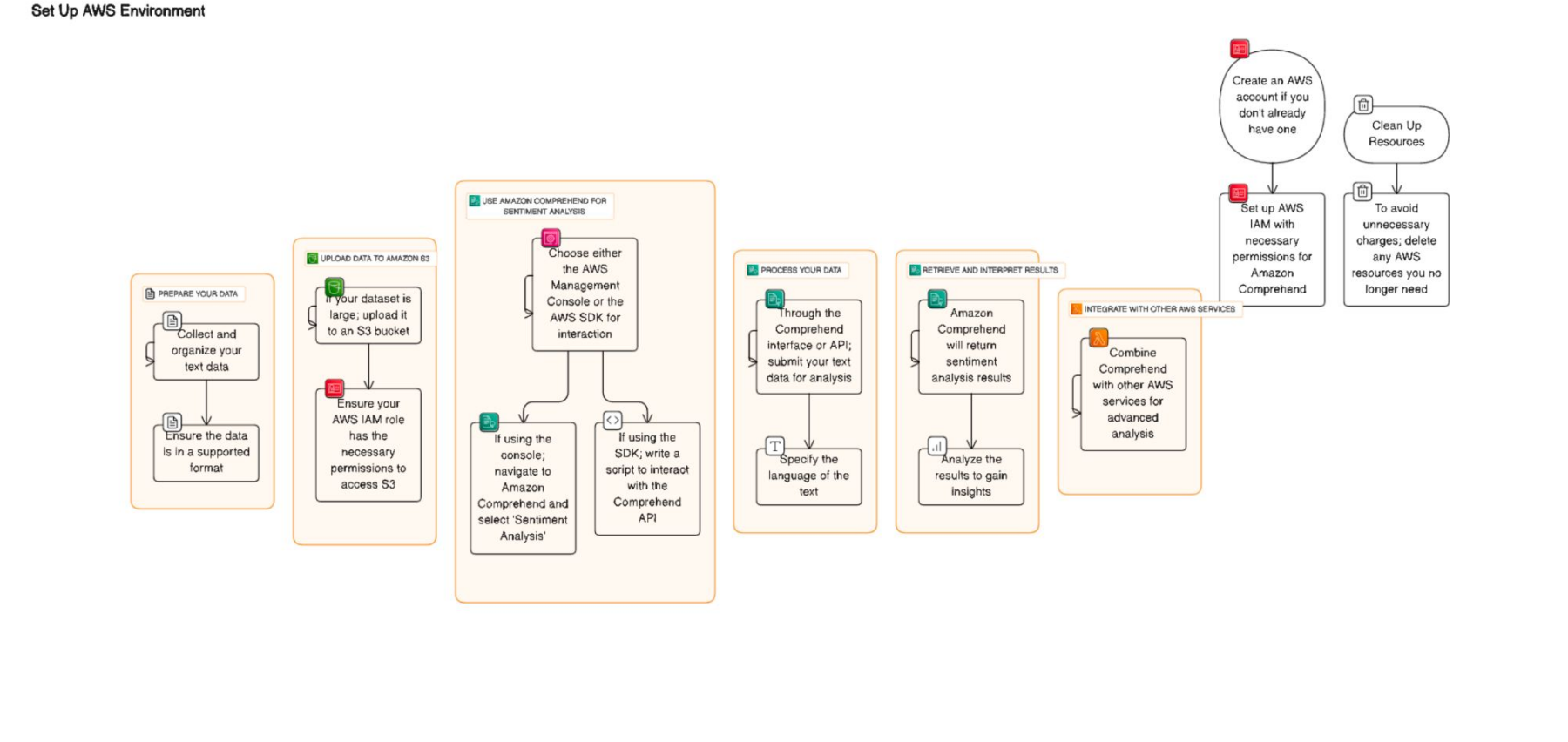
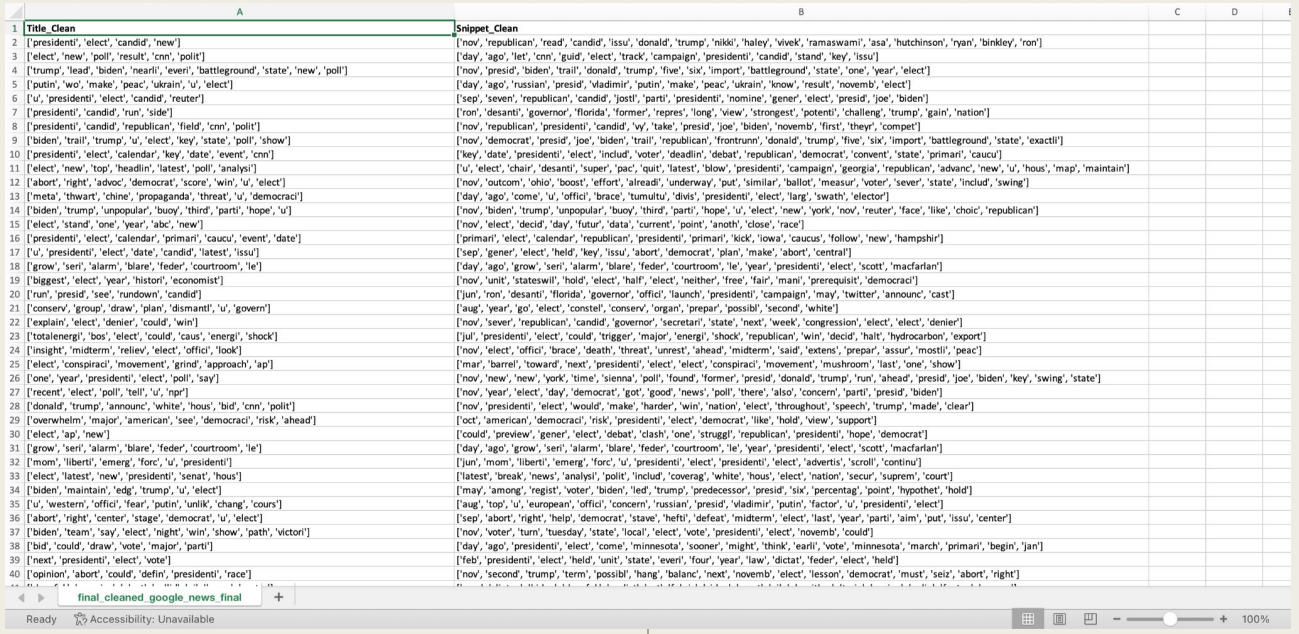


Fig 6: Cloud architecture

Results
For google

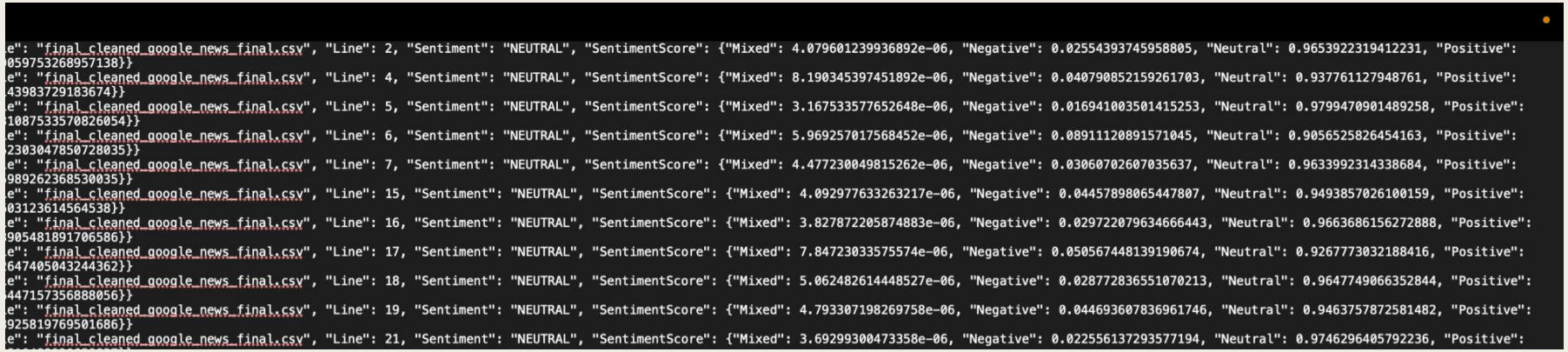
IUPUI
SCHOOL OF INFORMATICS AND COMPUTING



Title_Clean	Snippet_Clean
['presidenti', 'elect', 'candid', 'new']	['nov', 'republican', 'read', 'candid', 'issu', 'donald', 'trump', 'nikki', 'haley', 'vivek', 'ramaswami', 'asa', 'hutchinson', 'ryan', 'binkley', 'ron']
['elect', 'new', 'poll', 'result', 'cnn', 'polit']	['day', 'ago', 'let', 'cnn', 'guid', 'elect', 'track', 'campaign', 'presidenti', 'candid', 'stand', 'key', 'issu']
['trump', 'lead', 'biden', 'nearli', 'everi', 'battleground', 'state', 'new', 'poll']	['nov', 'presid', 'biden', 'trail', 'donald', 'trump', 'five', 'six', 'import', 'battleground', 'state', 'one', 'year', 'elect']
['putin', 'wo', 'make', 'peac', 'ukrain', 'u', 'elect']	['day', 'ago', 'russian', 'presid', 'vladimir', 'putin', 'make', 'peac', 'ukrain', 'know', 'result', 'novemb', 'elect']
['u', 'presidenti', 'elect', 'candid', 'reuter']	['sep', 'seven', 'republican', 'candid', 'jostl', 'parti', 'presidenti', 'nomine', 'gener', 'elect', 'presid', 'joe', 'biden']
['presidenti', 'candid', 'run', 'side']	['ron', 'desanti', 'governor', 'florida', 'former', 'repres', 'long', 'view', 'strongest', 'potenti', 'challeng', 'trump', 'gain', 'nation']
['presidenti', 'candid', 'republican', 'field', 'cnn', 'polit']	['nov', 'republican', 'presidenti', 'candid', 'vy', 'take', 'presid', 'joe', 'biden', 'novemb', 'first', 'theyr', 'compet']
['biden', 'trail', 'trump', 'u', 'elect', 'key', 'state', 'poll', 'show']	['nov', 'democrat', 'presid', 'joe', 'biden', 'trail', 'republican', 'frontrunn', 'donald', 'trump', 'five', 'six', 'import', 'battleground', 'state', 'exactli']
['presidenti', 'elect', 'calendar', 'key', 'date', 'event', 'cnn']	['key', 'date', 'presidenti', 'elect', 'includ', 'voter', 'deadlin', 'debat', 'republican', 'democrat', 'convent', 'state', 'primari', 'caucu']
['elect', 'new', 'top', 'headlin', 'latest', 'poll', 'analysi']	['u', 'elect', 'chair', 'desanti', 'super', 'pac', 'quit', 'latest', 'blow', 'presidenti', 'campaign', 'georgia', 'republican', 'advanc', 'new', 'u', 'hous', 'map', 'maintain']
['abort', 'right', 'advoc', 'democrat', 'score', 'win', 'u', 'elect']	['nov', 'outcom', 'ohio', 'boost', 'effort', 'alreadi', 'underway', 'put', 'similar', 'ballot', 'measur', 'voter', 'sever', 'state', 'includ', 'swing']
['meta', 'thwart', 'chine', 'propaganda', 'threat', 'u', 'democraci']	['day', 'ago', 'come', 'u', 'offici', 'brace', 'tumultu', 'divis', 'presidenti', 'elect', 'larg', 'swath', 'elector']
['biden', 'trump', 'unpopular', 'buoy', 'third', 'parti', 'hope', 'u']	['nov', 'biden', 'trump', 'unpopular', 'buoy', 'third', 'parti', 'hope', 'u', 'elect', 'new', 'york', 'nov', 'reuter', 'face', 'like', 'choic', 'republican']
['elect', 'stand', 'one', 'year', 'abc', 'new']	['nov', 'elect', 'decid', 'day', 'futur', 'data', 'current', 'point', 'anoth', 'close', 'race']
['presidenti', 'elect', 'calendar', 'primari', 'caucu', 'event', 'date']	['primari', 'elect', 'calendar', 'republican', 'presidenti', 'primari', 'kick', 'iowa', 'caucus', 'follow', 'new', 'hampshir']
['u', 'presidenti', 'elect', 'date', 'candid', 'latest', 'issu']	['sep', 'gener', 'elect', 'held', 'key', 'issu', 'abort', 'democrat', 'plan', 'make', 'abort', 'central']
['grow', 'seri', 'alarm', 'blare', 'feder', 'courtroom', 'le']	['day', 'ago', 'grow', 'seri', 'alarm', 'blare', 'feder', 'courtroom', 'le', 'year', 'presidenti', 'elect', 'scott', 'macfarlan']
['biggest', 'elect', 'year', 'histori', 'economist']	['nov', 'unit', 'stateswil', 'hold', 'elect', 'half', 'elect', 'neither', 'free', 'fair', 'mani', 'prerequisit', 'democraci']
['run', 'presid', 'see', 'rundown', 'candid']	['jun', 'ron', 'desanti', 'florida', 'governor', 'offici', 'launch', 'presidenti', 'campaign', 'may', 'twitter', 'announc', 'cast']
['conserv', 'group', 'draw', 'plan', 'dismantl', 'u', 'govern']	['aug', 'year', 'go', 'elect', 'constel', 'conserv', 'organ', 'prepar', 'possibl', 'second', 'white']
['explain', 'elect', 'denier', 'could', 'win']	['nov', 'sever', 'republican', 'candid', 'governor', 'secretari', 'state', 'next', 'week', 'congression', 'elect', 'elect', 'denier']
['totalenergi', 'bos', 'elect', 'could', 'caus', 'energi', 'shock']	['jul', 'presidenti', 'elect', 'could', 'trigger', 'major', 'energi', 'shock', 'republican', 'win', 'decid', 'halt', 'hydrocarbon', 'export']
['insight', 'midterm', 'reliev', 'elect', 'offici', 'look']	['nov', 'elect', 'offici', 'brace', 'death', 'threat', 'unrest', 'ahead', 'midterm', 'said', 'extens', 'prepar', 'assur', 'mostli', 'peac']
['elect', 'conspiraci', 'movement', 'grind', 'approach', 'ap']	['mar', 'barrel', 'toward', 'next', 'presidenti', 'elect', 'elect', 'conspiraci', 'movement', 'mushroom', 'last', 'one', 'show']
['one', 'year', 'presidenti', 'elect', 'poll', 'say']	['nov', 'new', 'new', 'york', 'time', 'sienna', 'poll', 'found', 'former', 'presid', 'donald', 'trump', 'run', 'ahead', 'presid', 'joe', 'biden', 'key', 'swing', 'state']
['recent', 'elect', 'poll', 'tell', 'u', 'npr']	['nov', 'year', 'elect', 'day', 'democrat', 'got', 'good', 'news', 'poll', 'there', 'also', 'concern', 'parti', 'presid', 'biden']
['donald', 'trump', 'announc', 'white', 'hous', 'bid', 'cnn', 'polit']	['nov', 'presidenti', 'elect', 'would', 'make', 'harder', 'win', 'nation', 'elect', 'throughout', 'speech', 'trump', 'made', 'clear']
['overwhelm', 'major', 'american', 'see', 'democraci', 'risk', 'ahead']	['oct', 'american', 'democraci', 'risk', 'presidenti', 'elect', 'democrat', 'like', 'hold', 'view', 'support']
['elect', 'ap', 'new']	['could', 'preview', 'gener', 'elect', 'debat', 'clash', 'one', 'struggl', 'republican', 'presidenti', 'hope', 'democrat']
['grow', 'seri', 'alarm', 'blare', 'feder', 'courtroom', 'le']	['day', 'ago', 'grow', 'seri', 'alarm', 'blare', 'feder', 'courtroom', 'le', 'year', 'presidenti', 'elect', 'scott', 'macfarlan']
['mom', 'liberti', 'emerg', 'forc', 'u', 'presidenti']	['jun', 'mom', 'liberti', 'emerg', 'forc', 'u', 'presidenti', 'elect', 'presidenti', 'elect', 'advertis', 'scroll', 'continu']
['elect', 'latest', 'new', 'presidenti', 'senat', 'hous']	['latest', 'break', 'news', 'analysi', 'polit', 'includ', 'coverag', 'white', 'hous', 'elect', 'nation', 'secur', 'suprem', 'court']
['biden', 'maintain', 'edg', 'trump', 'u', 'elect']	['may', 'among', 'regist', 'voter', 'biden', 'led', 'trump', 'predecessor', 'presid', 'six', 'percentag', 'point', 'hypothet', 'hold']
['u', 'western', 'offici', 'fear', 'putin', 'unlik', 'chang', 'cours']	['aug', 'top', 'u', 'european', 'offici', 'concern', 'russian', 'presid', 'vladimir', 'putin', 'factor', 'u', 'presidenti', 'elect']
['abort', 'right', 'center', 'stage', 'democrat', 'u', 'elect']	['sep', 'abort', 'right', 'help', 'democrat', 'stave', 'hefti', 'defeat', 'midterm', 'elect', 'last', 'year', 'parti', 'aim', 'put', 'issu', 'center']
['biden', 'team', 'say', 'elect', 'night', 'win', 'show', 'path', 'victori']	['nov', 'voter', 'turn', 'tuesday', 'state', 'local', 'elect', 'vote', 'presidenti', 'elect', 'novemb', 'could']
['bid', 'could', 'draw', 'vote', 'major', 'parti']	['day', 'ago', 'presidenti', 'elect', 'come', 'minnesota', 'sooner', 'might', 'think', 'earli', 'vote', 'minnesota', 'march', 'primari', 'begin', 'jan']
['next', 'presidenti', 'elect', 'vote']	['feb', 'presidenti', 'elect', 'held', 'unit', 'state', 'everi', 'four', 'year', 'law', 'dictat', 'feder', 'elect', 'held']
['opinion', 'abort', 'could', 'defin', 'presidenti', 'race']	['nov', 'second', 'trump', 'term', 'possibl', 'hang', 'balanc', 'next', 'novemb', 'elect', 'lesson', 'democrat', 'must', 'seiz', 'abort', 'right']

final_cleaned_google_news_final

Ready Accessibility: Unavailable 100%




```
e": "final_cleaned_google_news_final.csv", "Line": 2, "Sentiment": "NEUTRAL", "SentimentScore": {"Mixed": 4.079601239936892e-06, "Negative": 0.02554393745958805, "Neutral": 0.9653922319412231, "Positive":
059753268957138}}
e": "final_cleaned_google_news_final.csv", "Line": 4, "Sentiment": "NEUTRAL", "SentimentScore": {"Mixed": 8.190345397451892e-06, "Negative": 0.040790852159261703, "Neutral": 0.937761127948761, "Positive":
43983729183674}}
e": "final_cleaned_google_news_final.csv", "Line": 5, "Sentiment": "NEUTRAL", "SentimentScore": {"Mixed": 3.167533577652648e-06, "Negative": 0.016941003501415253, "Neutral": 0.9799470901489258, "Positive":
1087533570826054}}
e": "final_cleaned_google_news_final.csv", "Line": 6, "Sentiment": "NEUTRAL", "SentimentScore": {"Mixed": 5.969257017568452e-06, "Negative": 0.08911120891571045, "Neutral": 0.9056525826454163, "Positive":
2303047850728035}}
e": "final_cleaned_google_news_final.csv", "Line": 7, "Sentiment": "NEUTRAL", "SentimentScore": {"Mixed": 4.477230049815262e-06, "Negative": 0.03060702607035637, "Neutral": 0.9633992314338684, "Positive":
98926236853003 5}}
e": "final_cleaned_google_news_final.csv", "Line": 15, "Sentiment": "NEUTRAL", "SentimentScore": {"Mixed": 4.092977633263217e-06, "Negative": 0.04457898065447807, "Neutral": 0.9493857026100159, "Positive":
03123614564538}}
e": "final_cleaned_google_news_final.csv", "Line": 16, "Sentiment": "NEUTRAL", "SentimentScore": {"Mixed": 3.827872205874883e-06, "Negative": 0.029722079634666443, "Neutral": 0.9663686156272888, "Positive":
905481891706586}}
e": "final_cleaned_google_news_final.csv", "Line": 17, "Sentiment": "NEUTRAL", "SentimentScore": {"Mixed": 7.84723033575574e-06, "Negative": 0.050567448139190674, "Neutral": 0.9267773032188416, "Positive":
6474705043244362}}
e": "final_cleaned_google_news_final.csv", "Line": 18, "Sentiment": "NEUTRAL", "SentimentScore": {"Mixed": 5.062482614448527e-06, "Negative": 0.028772836551070213, "Neutral": 0.9647749066352844, "Positive":
447157356888056}}
e": "final_cleaned_google_news_final.csv", "Line": 19, "Sentiment": "NEUTRAL", "SentimentScore": {"Mixed": 4.793307198269758e-06, "Negative": 0.044693607836961746, "Neutral": 0.9463757872581482, "Positive":
925819769501686}}
e": "final_cleaned_google_news_final.csv", "Line": 21, "Sentiment": "NEUTRAL", "SentimentScore": {"Mixed": 3.692993004733 58e-06, "Negative": 0.022556137293577194, "Neutral": 0.9746296405792236, "Positive":
```


Fig 7: Sentiment scores document

Results

For google

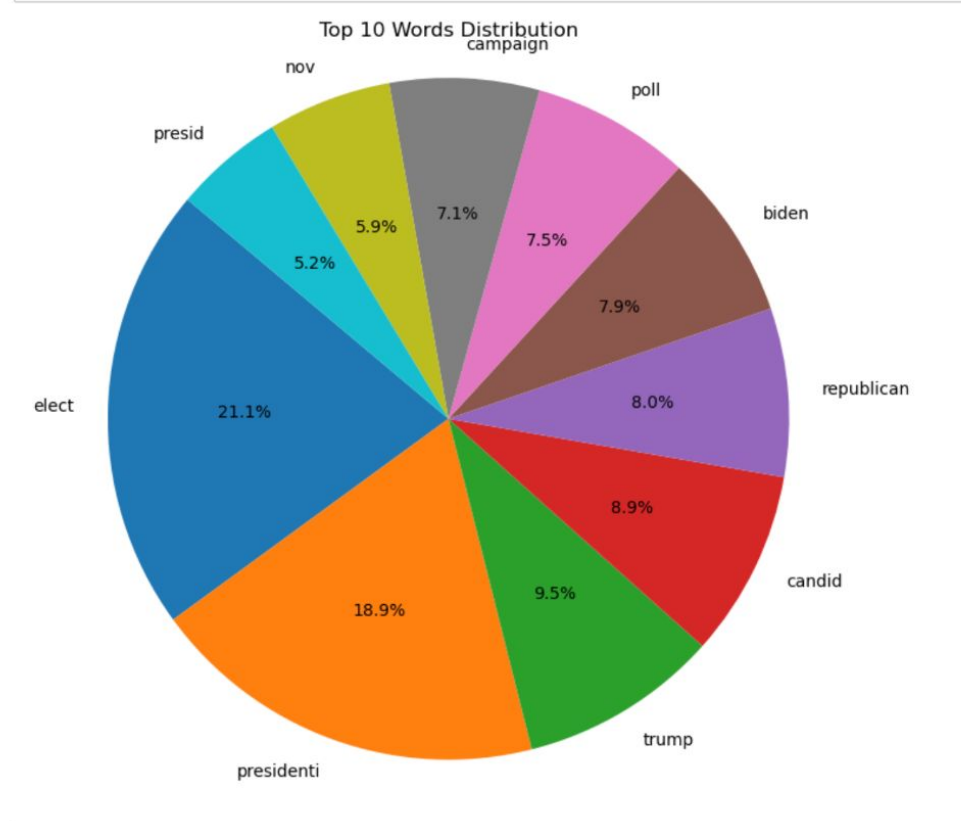


Fig 8: Pie chart of top 10 words

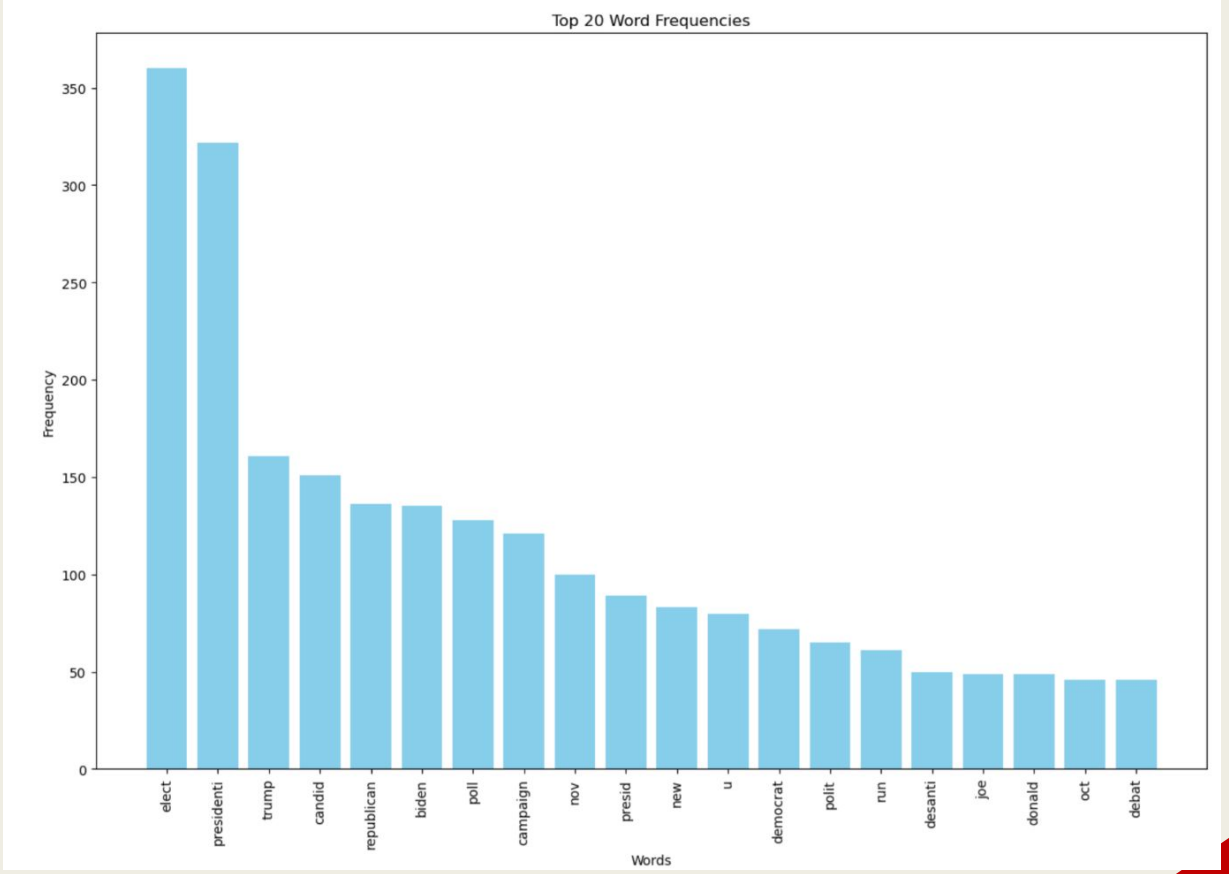


Fig 9: Bar graph showing top 20 word frequencies

Results

IUPUI
SCHOOL OF INFORMATICS AND COMPUTING

Reddit

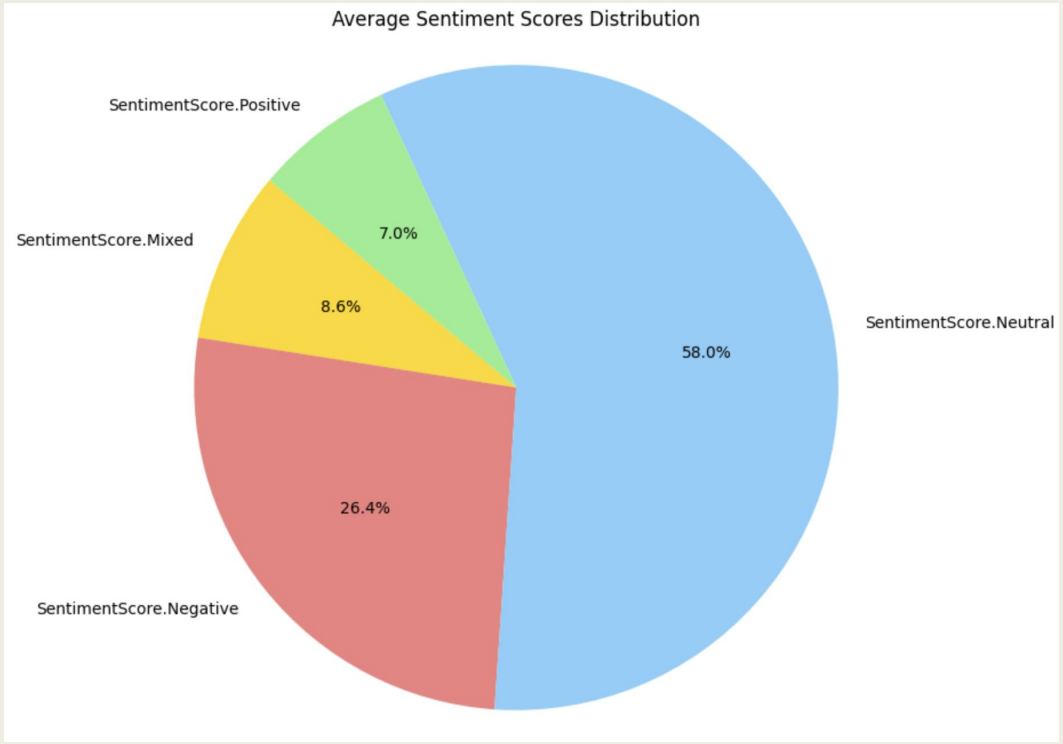


Fig 10: Average sentiment scores distribution from Reddit

Google

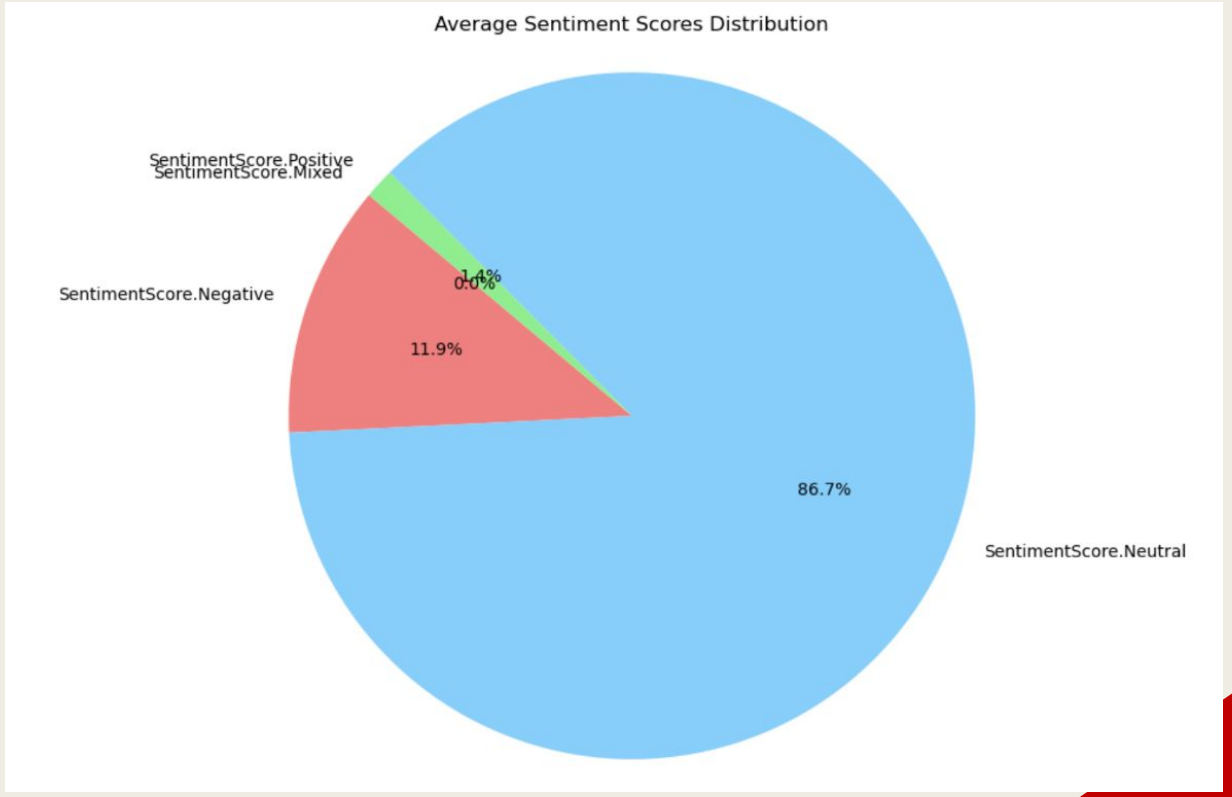


Fig 11: Average sentiment scores distribution from google

Results

IUPUI
SCHOOL OF INFORMATICS AND COMPUTING

Word Rank and Word frequencies from google and reddit

Reddit:

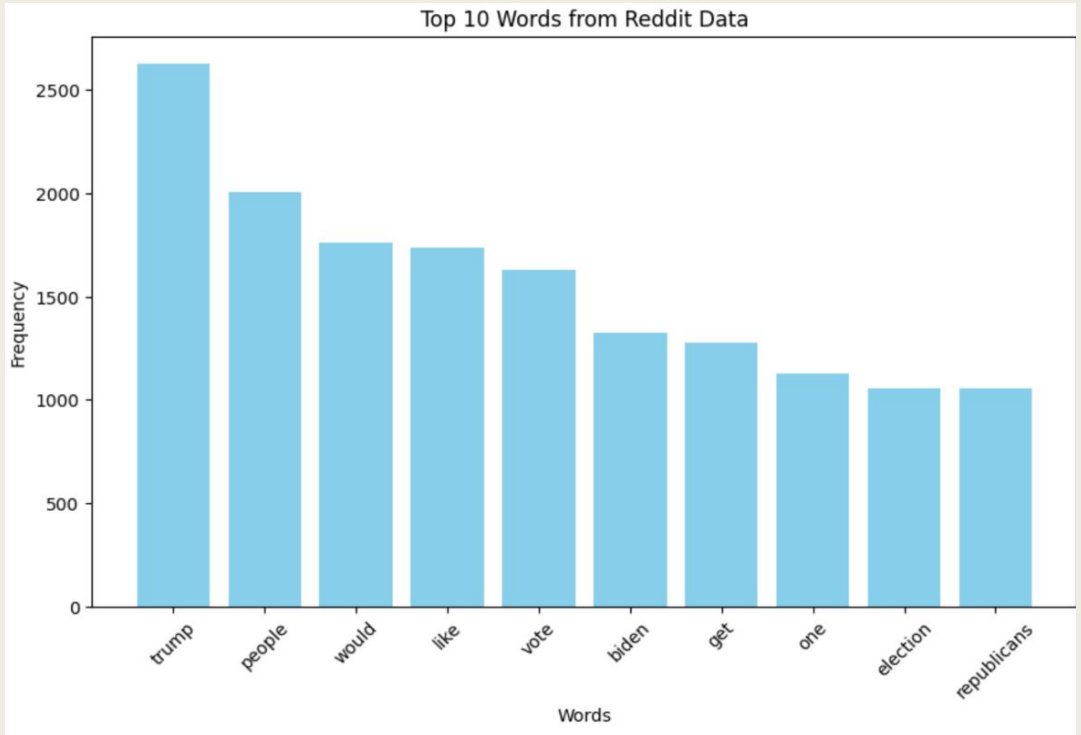
	Word	Frequency	Rank
0	donald	178	237.0
1	trump	2748	1.0
2	iowa	21	1869.0
3	accuses	3	6586.0
4	biden	1357	6.0

Fig 12: Word rank of Reddit

Google news:

	Word	Frequency	Rank
0	elect	360	1.0
1	presidenti	322	2.0
2	trump	161	3.0
3	candid	151	4.0
4	republican	136	5.0

Fig 13: Word rank of google



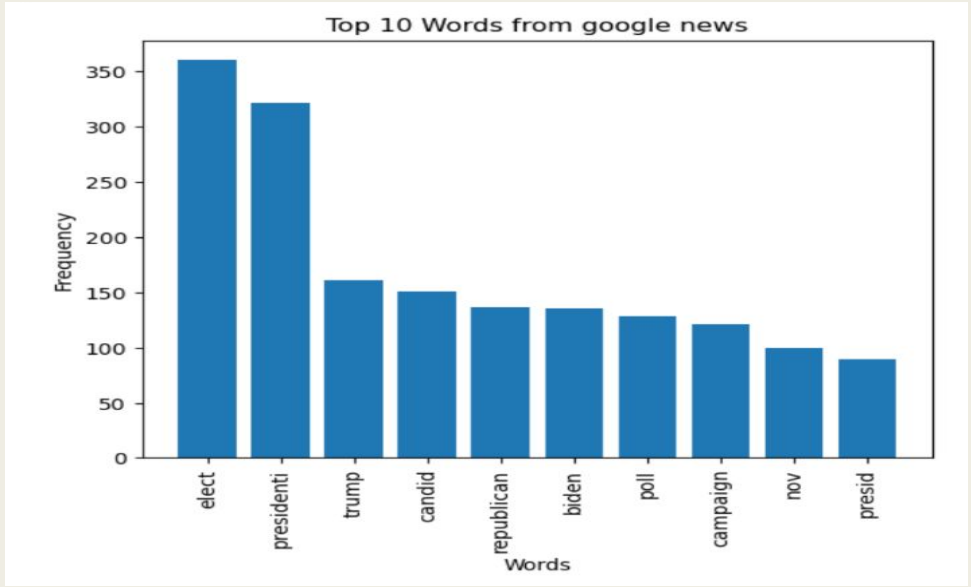


Fig 14: Top 10 words from google

Results - Word Cloud

IUPUI
SCHOOL OF INFORMATICS AND COMPUTING

Google

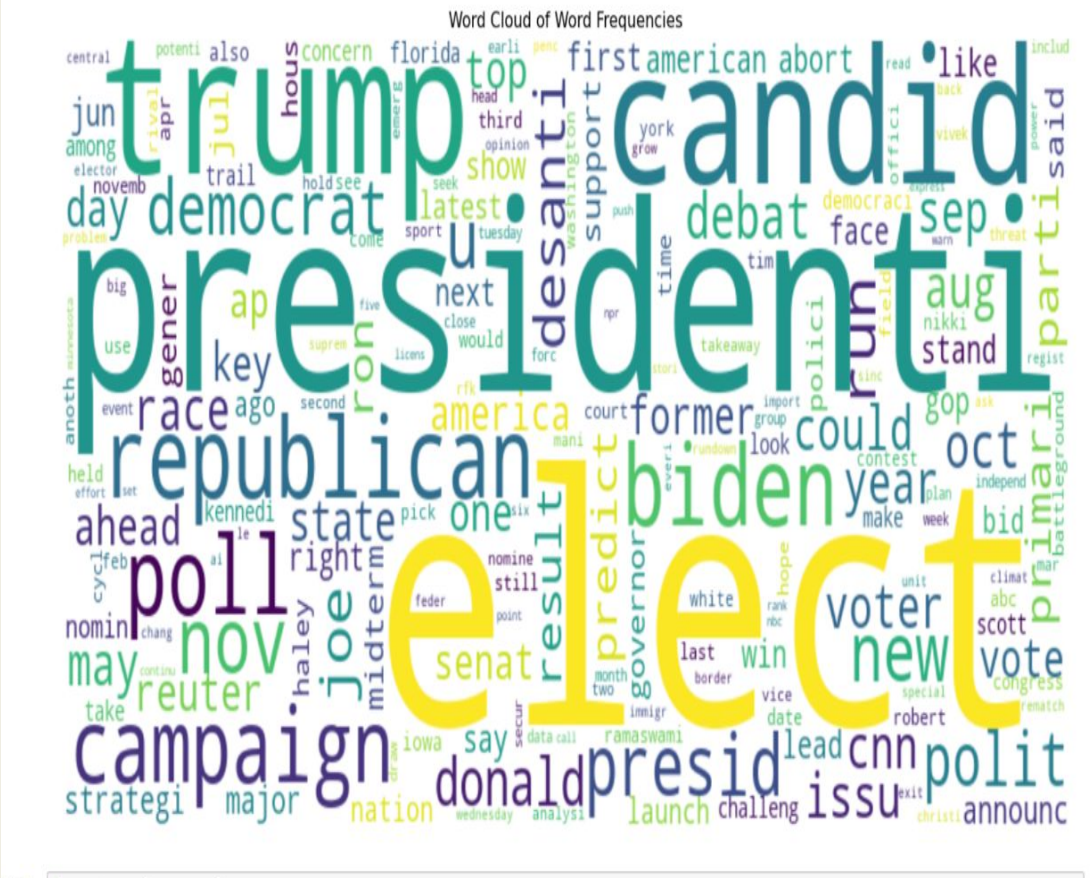


Fig 16: Word cloud from google

Reddit

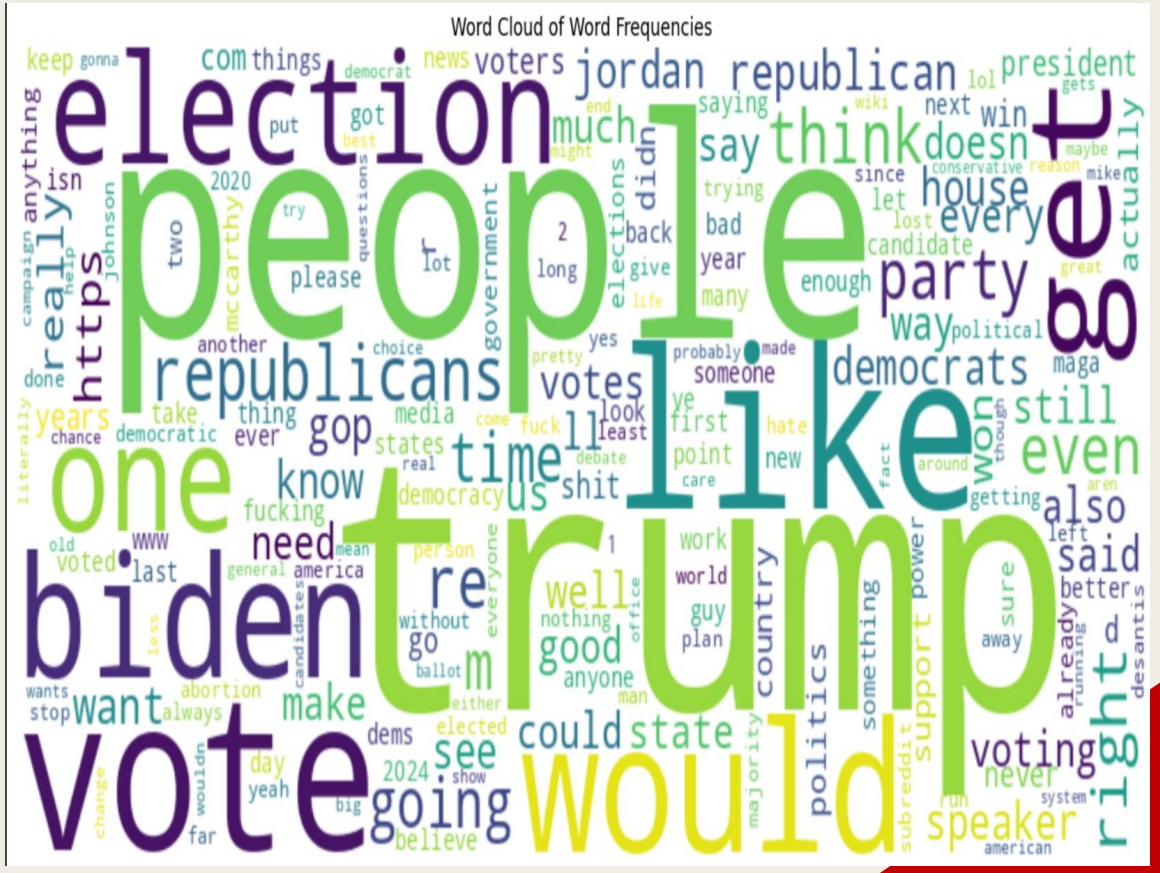


Fig 17: Word cloud from reddit

Results

For google

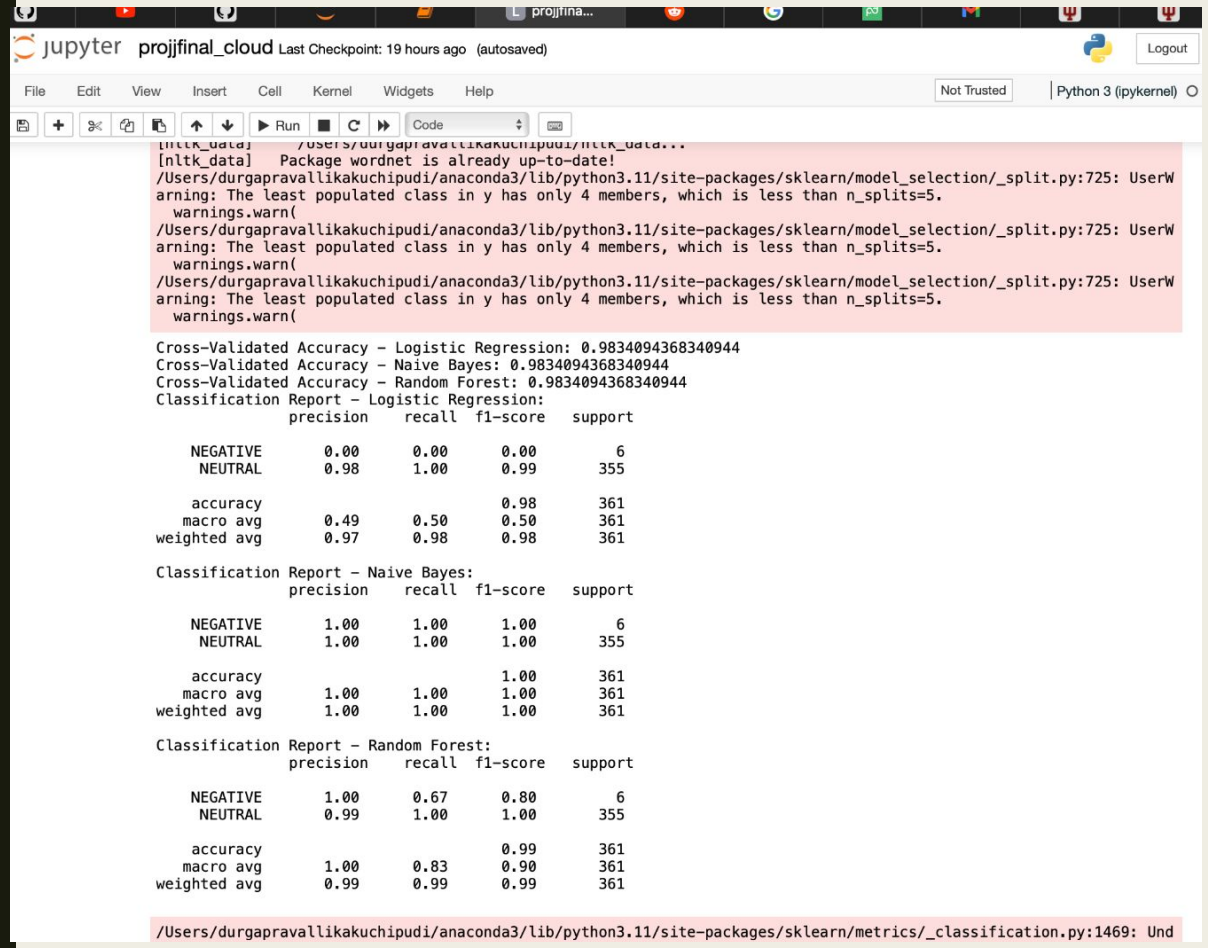


Fig 18: Cross-validated accuracy scores and classification reports of each model

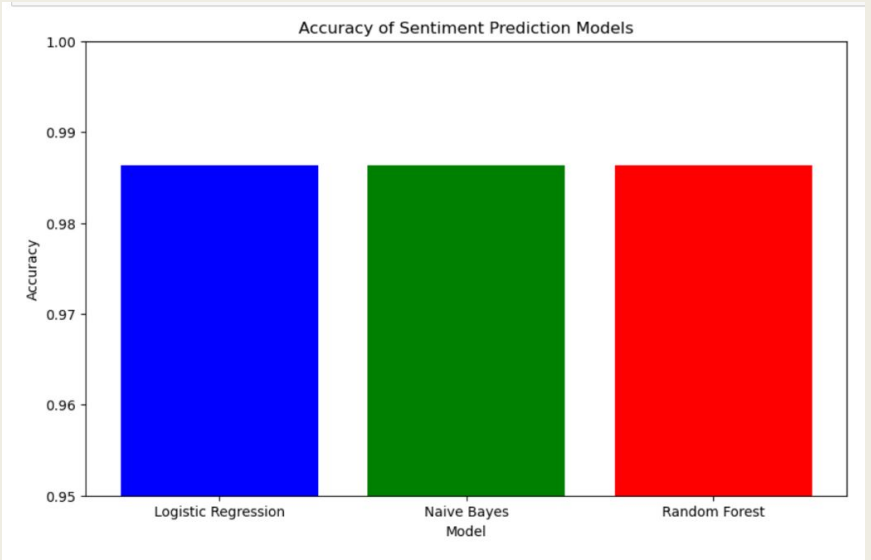


Fig 19: Accuracy score comparisons between different models

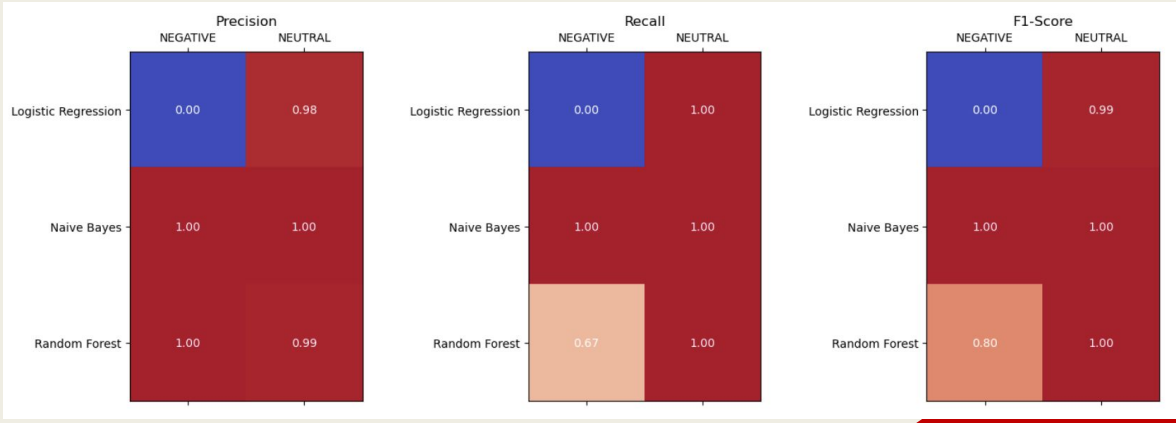


Fig 20: Classification Metrics for Different Models

Results

Reddit:

```
[nltk_data] Downloading package wordnet to /root/nltk_data...
Cross-Validated Accuracy - Logistic Regression: 0.6120182576333739
Cross-Validated Accuracy - Naive Bayes: 0.5945450810324535
Cross-Validated Accuracy - Random Forest: 0.6105921134716603
Classification Report - Logistic Regression:
              precision    recall  f1-score   support

       MIXED       1.00      0.14      0.25       770
    NEGATIVE       1.00      0.09      0.16      3152
     NEUTRAL       0.62      1.00      0.77      6720
    POSITIVE       1.00      0.09      0.17       575

    accuracy                           0.64     11217
   macro avg       0.90      0.33      0.34     11217
weighted avg       0.77      0.64      0.53     11217

/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_classification.py:1344: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to
  _warn_prf(average, modifier, msg_start, len(result))
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_classification.py:1344: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to
  _warn_prf(average, modifier, msg_start, len(result))
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_classification.py:1344: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to
  _warn_prf(average, modifier, msg_start, len(result))
Classification Report - Naive Bayes:
              precision    recall  f1-score   support

       MIXED       1.00      0.02      0.04       770
    NEGATIVE       0.72      0.04      0.08      3152
     NEUTRAL       0.61      1.00      0.76      6720
    POSITIVE       0.00      0.00      0.00       575

    accuracy                           0.61     11217
   macro avg       0.58      0.27      0.22     11217
weighted avg       0.64      0.61      0.48     11217

Classification Report - Random Forest:
              precision    recall  f1-score   support

       MIXED       0.99      0.14      0.25       770
    NEGATIVE       0.99      0.09      0.16      3152
     NEUTRAL       0.62      1.00      0.77      6720
    POSITIVE       1.00      0.09      0.17       575

✓ Connected to Python 3 Google Compute Engine backend
```

Fig 21: Cross-validated accuracy scores and classification reports of each model of Reddit

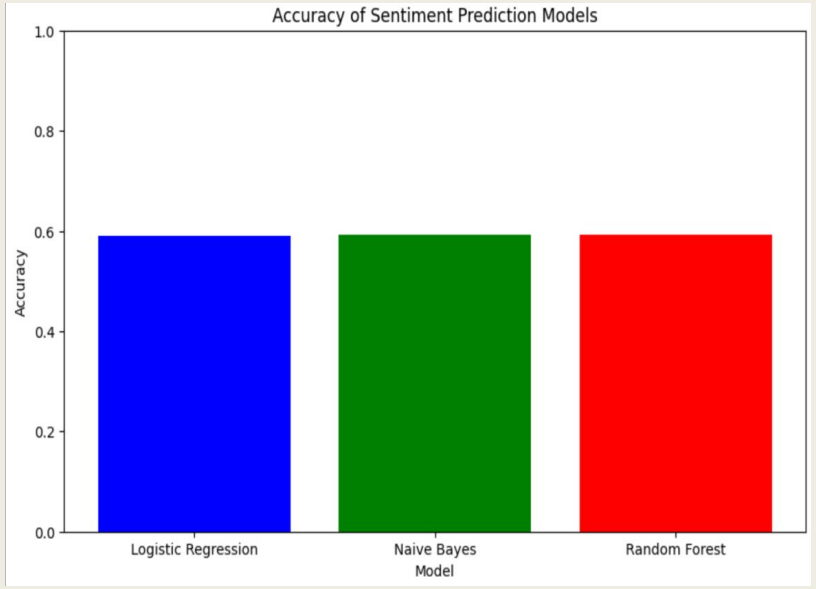


Fig 22: Accuracy score comparisons between different models

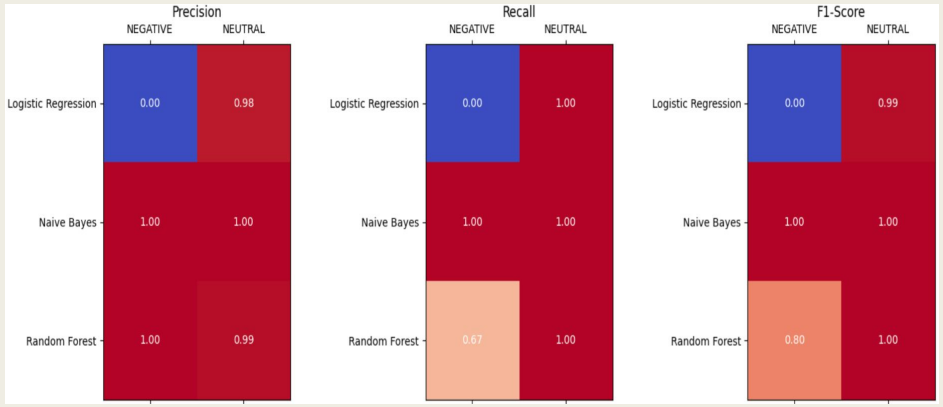


Fig 23: Classification Metrics for Different Models

Limitations

1. **Class Imbalance**: Potential issue in accurately identifying negative sentiments due to overshadowing by neutral sentiments.
2. **Model Overfitting**: Perfect scores of Naive Bayes raise concerns about memorizing training data instead of generalizing.
3. **Data Diversity**: Lack of diverse negative sentiment examples may lead to inadequate model training.
4. **Model Suitability**:Not all models equally suitable for sentiment classes, indicating a need for tailored approaches or ensemble methods.
5. **Evaluation Metrics**:Reliance on accuracy alone is insufficient; consider AUC-ROC, precision-recall curves, or confusion matrices for deeper insights.

Future Work

1. Enhanced Data Collection: Incorporate a wider range of real-time data from diverse sources and languages to capture a more global and comprehensive sentiment landscape.

2. Advanced Analytics: Employ cutting-edge machine learning techniques and natural language processing algorithms for deeper sentiment contextualization and more accurate prediction models.

3.Temporal and Demographic Insights: Implement time-series analysis to monitor sentiment trends and conduct demographic studies to understand sentiment variations across different voter segments.

4. Interactive Tools: Develop an interactive dashboard for real-time sentiment tracking and analysis, providing a dynamic tool for various stakeholders.

5. Ethical Framework: Strengthen ethical data handling practices, focusing on privacy, consent, and the responsible use of sentiment analysis in political contexts.

Conclusion

- **Accuracy Discrepancy**: High accuracy across Logistic Regression, Naive Bayes, and Random Forest, but performance varies across sentiment categories.
- **Precision, Recall, F1-score Differences**:Metrics reveal significant variations in correctly identifying negative sentiments among the models.
- **Naive Bayes Perfect Scores**: Raises concerns about potential overfitting, as achieving perfect scores may indicate memorization of training data.
- **Logistic Regression Failure**: Fails to identify any negative sentiments, highlighting limitations in its performance for this sentiment class.
- **Random Forest Partial Success**: Shows partial success in identifying negative sentiments, suggesting a potential for improvement or optimization.

Project link

https://github.com/Supraja-p/Cloud_computing