Level 4 Project Presentation

Health Misinformation, Covid-19, and Search

Shahzeb Zafar, 2371704

Motivation

- The project has two major objectives:
  1. Understand how misinformation changes overtime when the quantity and quality of data increases.
  2. Identify the main characteristics of misinformation to make it easier to identify and filter out fake news.

Methodology

- The project was split into 5 distinct tasks:
  1. Gathering Twitter Data
  2. Binary Classification
  3. Data Clustering
  4. Sentiment & Emotion Classification
  5. Analysis of Data

Gathering Twitter Data

- Collected around 1M tweets from a large scientific dataset for Covid-19 tweets for 8 time periods.
- The time periods are: January, April, July, October 2020/21

Binary Classification

- Classified the tweets as either true or false.
- Used a pre-trained BERT model (Covid-Twitter-Bert-v2)that was trained using over 1B tweets.
- Finetuned the model using a training dataset of around 24,000 tweets.

Data Clustering

- Grouped the tweets into 15 topics.
- Used Jaccard Similarity as a substitute for clustering algorithms.

Sentiment & Emotion Classification

- Classified the tweets as either positive, negative, or neutral.
- For the emotions we have joy, sadness, anger, surprise, disgust, and fear.
- Both were done using the pysentimiento package.

Analysis of Data

- Calculated the Pearson Correlation Coefficient for sentiments and emotions for each topic.
- Visualisation done using tables and line graphs.

Results

- Negative sentiment, anger, and disgust have a moderate positive correlation with misinformation (coefficient >= 0.4).

Sentiment	Pearson Correlation Coefficient
Positive	-0.608
Negative	0.499
Neutral	0.067

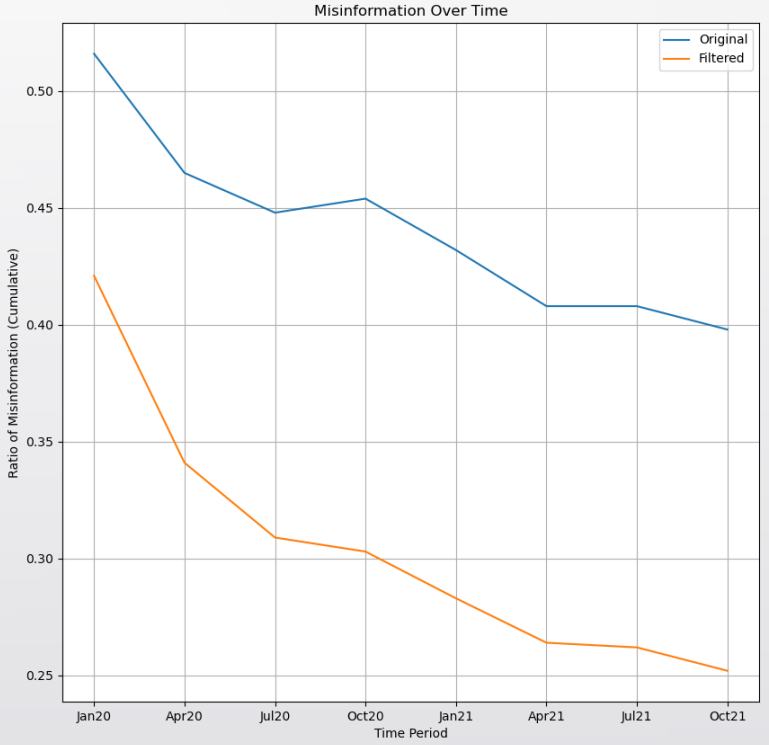
Table 5.18: *Average Pearson Correlation Coefficients for Sentiment*

Emotion	Pearson Correlation Coefficient
Joy	-0.264
Sadness	-0.190
Anger	0.428
Surprise	0.188
Disgust	0.497
Fear	-0.256

Table 5.19: *Average Pearson Correlation Coefficients for Emotions*

Results

- Blue line represents original dataset without any alterations. We see misinformation naturally decrease as time passes.
- Orange line represents filtered dataset after removing tweets with negative sentiment + anger/disgust. There's a drastic decrease in misinformation.



Conclusion

- Successfully identified key characteristics of misinformation.
- Can use this to develop advanced deep learning models to detect information with such traits and remove/ignore it.

Acknowledgements

- Dr Jesus Rodriguez Perez for being a great supervisor and helping me throughout the course of the project.
- Dr Graham McDonald and Una Marie Darragh for helping me get an extension for medical reasons.

Thank You!