Benson Yee

Professor Mounjed Moussalem

CIS 140

14 December 2023

# Project Proposal: Predicting Emotion From Text using Big Data and NLP

## TEAM

**Lead Researcher:** Benson Yee

## SUBJECT

**Natural Language Processing (NLP)** is a machine learning technique that allows computers to understand, interpret, and manipulate human language. It allows programs to accept large volumes of voice and text data, analyze the intent or sentiment, then respond in real time akin to human conversation.

The goal of this project is to (a) train a rudimentary NLP model to predict emotion from text using publicly available datasets, then (b) feed new text to reveal trends of both my specific model and human language as a whole. This allows us to challenge common conceptions and intuitions about inferring emotion, and how **artificial intelligence (AI)** interprets feelings versus humans.

# DATA

To train our model, we will be using the **"NLP Emotion Dataset"** from prathap kumar (@gymprathap) (https://data.world/gymprathap/nlp-emotion-dataset). The dataset contains 21,459 texts associated with one of five emotions: anger, fear, happy, love, sadness, and surprise. These have been manually inferred by human interpreters instead of directly observed as empirical evidence, as emotion (as we currently know it) is a subjective experience. Below are several examples from this dataset. The complete dataset can be found under **CREDITS**.

| Text | Emotion |
|---|---|
| i wake up feeling cranky and out of sorts | anger |
| im one of girl who feel insecure about herself always | fear |
| i can stay awake whole night feeling all energetic and stuff | happy |
| i feel cared for and accepted | love |
| i know its too late to crawl back to you but im feeling so alone | sadness |

**Table 1: NLP Emotion Dataset**

# FEATURES

There are two features present: text and emotion. All features are raw input from the "NLP Emotion Dataset". Emotion can be one of five feelings (anger, fear, happy, love, sadness, surprise), where text can be any string that conveys that sentiment.

These are appropriate features for several reasons. The goal of the project is predicting emotion from text. The input is text, and the output is emotion. Furthermore, emotion is a complicated, convoluted experience that can often only be described and not directly observed. People can feel several—even conflicting—emotions simultaneously. In addition, context can completely change a text's emotion. For example, if I say, "I spent all day at the ice rink", this could be classified as "happy" or "love" for someone who loves to skate. However, if I say, "My family dragged me to spend all day at the ice rink", this additional context of obligation could twist this scenario to feel angry. Someone terrified of skating could twist this scenario to feel fear.

By having only two features—text and emotion—and identifying only five common, but diverse emotions, we can narrow typically complicated experiences into simpler phenomena we can study.

## EXPERIMENTS

We begin our experiment by creating a Python script. This script performs the following functions in order:

1. Load "NLP Emotion Dataset" into a Pandas DataFrame (`load_data`)
2. Preprocess text data (`preprocess_text`)
3. Train and evaluate accuracy of model (`train_model`)
4. Accept text and predict emotion (`predict_emotion_probabilities`)
5. Appends text and emotion probabilities to CSV file (`append_to_csv`)
6. Repeat steps 4 - 5 until the user enters 'q'

The script uses **Term Frequency-Inverse Document Frequency (TF-IDF)**, a numerical statistic that reflects how important a word is to a document in a collection. TF-IDF is commonly used in NLP and information retrieval for text analysis.

**Term Frequency (TF)** measures how often a term (word) appears in a document. It calculates the ratio of occurrences of a term to the total terms in the document $TF(t, d) = \frac{Times\ term\ t\ appears\ in\ document\ d}{Total\ terms\ in\ document\ d}$.

**Inverse Document Frequency (IDF)** measures how important a term is across a collection of documents. It's calculated as the logarithm of the ratio of total documents to number of documents containing that term, plus one to avoid division by zero $IDF(t, D) = log(\frac{Total\ documents\ in\ corpus\ N}{1 + Documents\ containing\ term\ t})$

The TF-IDF score for a term in a document is the product of its TF and IDF $TF - IDF(t, d, D) = TF(t, d) \times IDF(t, D)$.

The script uses TF-IDF by transforming a collection of raw documents (texts) into a matrix of TF-IDF features. This matrix is used by a **Support Vector Machine (SVM)**, an algorithm that maximizes decision boundaries to nearest data points. This improves accuracy and performance. The SVM model is trained using the TF-IDF matrix before evaluating its performance upon the original dataset.

The complete script as well as the SVM model used in our experiments can be found under [CREDITS].

# RESULTS

After building and training our model, we first evaluate its performance against the original dataset. The script uses several metrics to measure the model's effectiveness.

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| anger | 0.95 | 0.93 | 0.94 | 2993 |
| fear | 0.91 | 0.94 | 0.93 | 2652 |
| happy | 0.94 | 0.97 | 0.95 | 7029 |
| love | 0.94 | 0.82 | 0.88 | 1641 |
| sadness | 0.96 | 0.96 | 0.96 | 6265 |
| surprise | 0.94 | 0.84 | 0.89 | 879 |
| | | | | |
| accuracy | | | 0.94 | 21459 |
| macro avg | 0.94 | 0.91 | 0.92 | 21459 |
| weighted avg | 0.94 | 0.94 | 0.94 | 21459 |

**Table 2: Emotion Classification Evaluation**

Here are the key terms and their definitions in context of the script:

- **Precision:** The ratio of correctly predicted positive observations (true positives) to the total predicted positives (true positives + false positives)
- **Recall:** The ratio of correctly predicted positive observations (true positives) to the total actual positives (true positives + false negatives)
- **F1-Score:** The harmonic mean of precision and recall, providing a balanced measure that considers both false positives and false negatives

- **Support:** The number of actual occurrences of each emotion class in the dataset
- **Accuracy:** The ratio of correctly predicted observations (true positives and true negatives) to total observations
- **Macro Avg (Macro-Averaging):** The average of performance metrics (precision, recall, F1-score) across all classes
- **Weighted Avg (Weighted-Averaging):** The average of performance metrics, weighted by each class' contribution by its support (number of instances)

Across all metrics, we see our model is extremely precise, with a total weighted accuracy of 94%. This gives us a robust framework to begin our investigation.

We begin by feeding texts that contain a singular emotion. This is to test our model's **zero-shot learning (ZSL)**, or performing a task without explicit training upon it.

| Text | anger | fear | happy | love | sadness | surprise |
|---|---|---|---|---|---|---|
| i feel angry | 1.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| i feel fearful | 0.0000 | 0.9999 | 0.0000 | 0.0000 | 0.0000 | 0.0001 |
| i feel happy | 0.0000 | 0.0000 | 0.9994 | 0.0006 | 0.0000 | 0.0000 |
| i feel loved | 0.0000 | 0.0000 | 0.0000 | 1.0000 | 0.0000 | 0.0000 |
| i feel sad | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 1.0000 | 0.0000 |
| i feel surprised | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 1.0000 |

**Table 3: Singular Emotion Prediction**

We see our model correctly identifies texts with a singular emotion with extremely high confidence. This opens the door for more complex texts.

| Text | anger | fear | happy | love | sadness | surprise |
|---|---|---|---|---|---|---|
| i feel angry and fearful | 0.0807 | 0.9187 | 0.0002 | 0.0000 | 0.0000 | 0.0003 |
| i feel angry and happy | 0.9999 | 0.0000 | 0.0001 | 0.0000 | 0.0000 | 0.0000 |
| i feel angry and loved | 0.0218 | 0.0000 | 0.0017 | 0.9765 | 0.0000 | 0.0000 |
| i feel angry and sad | 0.9956 | 0.0000 | 0.0000 | 0.0000 | 0.0043 | 0.0000 |
| i feel angry and surprised | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 1.0000 |
| i feel loved and angry | 0.0218 | 0.0000 | 0.0017 | 0.9765 | 0.0000 | 0.0000 |

**Table 4: Dual-Emotion Prediction**

Here, we see the model correctly identify one of the two emotions present with very high confidence. The other emotion falls between 0.00 - 2.18%. Notice when given two emotions, the model *always* chooses the emotion with less support.

- "Fear" (2652 occurrences) is chosen over "Anger" (2993 occurrences)
- "Anger" (2993 occurrences) is chosen over "Happy" (7029 occurrences)
- "Love" (1641 occurrences) is chosen over "Anger" (2993 occurrences)
- "Anger" (2993 occurrences) is chosen over "Sadness" (6265 occurrences)
- "Surprise" (879 occurrences) is chosen over "Anger" (2993 occurrences)

Also notice the order the emotions are stated does not change the predictions. "i feel angry and loved" has the same probabilities as "i feel loved and angry".

| Text | anger | fear | happy | love | sadness | surprise |
|---|---|---|---|---|---|---|
| i feel equally angry and loved | 0.3969 | 0.0002 | 0.0117 | 0.5903 | 0.0007 | 0.0001 |
| i feel equally loved and angry | 0.3969 | 0.0002 | 0.0117 | 0.5903 | 0.0007 | 0.0001 |
| i feel more angry than loved | 0.1130 | 0.0001 | 0.0101 | 0.8765 | 0.0002 | 0.0001 |
| i feel more loved than angry | 0.1130 | 0.0001 | 0.0101 | 0.8765 | 0.0002 | 0.0001 |
| i feel so angry, i love you | 0.9999 | 0.0000 | 0.0000 | 0.0001 | 0.0000 | 0.0000 |
| i feel so loved, i'm angry | 0.0826 | 0.0001 | 0.0029 | 0.9142 | 0.0001 | 0.0000 |

**Table 5: Weighted Dual-Emotion Prediction**

Here, we modified our dual-emotion texts to add weight (bias) toward one or both of the emotions.

In texts using the word, 'equally', we see both emotions predicted with roughly equal confidence, biased around 9.3% toward the emotion with less support.

In texts where one emotion is stated more than the other, we see roughly an 11.3% increase in one emotion, compared to a near 100% confidence in the other emotion. The order, once again, does not affect these probabilities.

In texts where we state we feel one emotion so strongly, we feel the other, we see a high confidence in the strong emotion, and a low confidence in the other emotion. This time, however, the order *does* affect these predictions, biased toward the strong emotion.

| Long Text 1 — Author A | | | | | |
|---|---|---|---|---|---|
| Willow felt out of place. Exposed. Out in the open. No cover or concealment within sight. Without safety, and without weapon, the only instinct that remained was run. Run. Run. The gentle warmth of a hand. It wraps around Willow's left shoulder. She tenses, like a twig snapped. Like gunshot cracked. Doe eyes grow wide. Like a beast trapped in headlights. Willow gave Reni the opportunity of solace. To end social interaction and be by herself, if she wanted it. That meant if she felt uncomfortable, she could have let Willow leave. There were only a few possible reasons Reni might not want Willow to leave. She could be looking to exploit her. To take the doegirl's manna, information, or resources. She could want to get close to her. To get her to drop her guard. Lower her defenses. Say or share something vulnerable about themselves. Blackmail, connections, scams. Or maybe. Just maybe. She liked her. And wanted to spend a little more time with her. Willow takes a deep breath in. She's quiet. Drawn in. Reserved. For a long, long time. Lavender eyes flick. Bottom lip bit. Gears turning in her head. Before she sighs. \"Well, uhm. If we're... going to keep talking... the least I could do is buy you something to drink. Coffee? Tea? I think I want a hot c-chocolate myself.\" | | | | | |
| **anger** | **fear** | **happy** | **love** | **sadness** | **surprise** |
| 0.1165 | 0.0840 | 0.1699 | 0.5435 | 0.0734 | 0.0126 |

| Long Text 2 — Author B | | | | | |
|---|---|---|---|---|---|
| Rose flinches as well as he does, looking quite nervous as he quickly looks over at them. Worried they did something wrong, they wait with baited breath until the word comes, at which point they let it out, face melting into relief and a smile. \"No need to thank me for just being cute.\" | | | | | |
| **anger** | **fear** | **happy** | **love** | **sadness** | **surprise** |
| 0.0033 | 0.3573 | 0.6239 | 0.0018 | 0.0055 | 0.0081 |

| Long Text 3 — Author B | | | | | |
|---|---|---|---|---|---|
| Rose smiles, wrapping an arm around his shoulder. \"So, I saw your eyes flash down that red lit street. Anything catch you eye~?\" | | | | | |
| **anger** | **fear** | **happy** | **love** | **sadness** | **surprise** |
| 0.0513 | 0.0944 | 0.5543 | 0.0151 | 0.2700 | 0.0149 |

| Long Text 4 — Author C |
|---|
| Oriana frowns as she stares up into the clouds \"Hmph. I shall admit, your \'everstorm\' is quite grand. Why, the sheer size alone would make even she who bestowed the curse upon my lineage envious of its tenacity.\" \"But!\" She snaps her fingers, creating a boom like a burst of thunder \"A true Scion of Zetylyn cannot settle! My ancestors would surely shame me if I simple kneeled down to one who created a larger storm than I. Doubly so, if it were a bunch of foppish wizards and their nonsensical instruments.\" \"Nay. The only honorable path that lies before me is to exceed it. Thus, I must humbly decline.\" She concludes, dipping her head respectfully to Aella \"For as your storm is mighty, I intend on creating one that surpasses it in every way.\" |

| anger | fear | happy | love | sadness | surprise |
|---|---|---|---|---|---|
| 0.5767 | 0.0711 | 0.0814 | 0.0137 | 0.2375 | 0.0195 |

| Long Text 5 — Author C | | | | | |
|---|---|---|---|---|---|
| There's a squeak from the bed. A soft grunt as Elena works up the strength to drag herself over to the side. Then a warm hand, settling atop Katias shoulder, giving it a reassuring squeeze. A hug would've been ideal, but this was all she could manage right now \"Not a war of conquest, but a war of subjugation all the same.\" She murmurs, staring down at the bed \"I know how it feels to look a person in the eye after you've murdered their husband. Their sons. Then to justify how that -somehow- makes them safer. I've felt the hate and disgust they have towards people like me, how all they see is the personification of the TSC's wrath.\" \"It hurts. It hurts so much being forced to do wrong. Even more when you're punished for something you never had a choice in. Or worse yet -rewarded.\" Her voice begins to weaken a bit, her grip tightening. \"The alternative was ignoring it. But that's like a knife chipping away a little bit of you every time you do it. Over and over until you're not -you- anymore. Just a machine, a tool, nothing more.\" \"I understand why you do it. I did it myself for a while. But I couldn't anymore. No matter how much it hurts, I can't lose myself again.\" She sniffles and tries to steady herself with a quick breath, but ends up breaking into a fit of wheezing. Realizing she let her emotions get the best of her, she clears her throat, then looks back up towards Katia \"Sorry. I... probably said too much. I'm not going to ask you to do the same, but I want you to think about whether a lie is the life you want to live.\" | | | | | |
| anger | fear | happy | love | sadness | surprise |
| 0.2310 | 0.0061 | 0.0374 | 0.0051 | 0.7129 | 0.0075 |

| Long Text 6 — Author D | | | | | |
|---|---|---|---|---|---|
| \"Could be tempted with that, I suppose.\" He shuffles in place again, getting comfortable before looking upwards at the station's \'sky.\' | | | | | |
| anger | fear | happy | love | sadness | surprise |
| 0.1386 | 0.0051 | 0.8320 | 0.0035 | 0.0113 | 0.0096 |

**Table 6: Complex Emotion Prediction**

Here are several fiction roleplay posts written by four different people, analyzed with consent from their respective authors. They contain context, dialogue, introspection, nuance, and subtext. These are examples of situations that may appear in the real world. In all these examples, not only does the model correctly identify the primary emotion, but also secondary and sometimes even tertiary emotions. This satisfies the goal we had for our project.

## INTERPRETATIONS AND DISCUSSIONS

From our **RESULTS**, we conclude our NLP model is fairly accurate, sans the following faults:

(a) When given two emotions, the model is biased toward the emotion with the LEAST support (i.e the emotion with the least training is the emotion selected)

(b) The order emotions are stated typically does not affect predictions. This causes problems in texts where order is important (i.e "i feel more angry than loved" yields the same results as "i feel more loved than angry")

For (a), this yields an interesting insight about knowledge and interpretation. When faced with ambiguity, the model predicts the category it knows less about. Even when weight is given (such as "i feel equally angry and loved"), the model biases in-favor of the category it knows less about. A metaphor would be a student who is both an expert programmer but an average statistician. In projects that require both programming and statistics, the student would be more likely to classify ambiguous problems as statistics—the field they know less about—even if the problem requires equal amounts of both. This **multifield expert bias**, or the bias of classifying ambiguous problems as the result of inadequacy in fields you aren't an expert in, may be a phenomena that humans experience too. This flaw could be fixed rudimentarily by having equal amounts of testing data for each category. Another possible fix would be training the model on texts with multiple emotions. Equal weight could be given per emotion (i.e a text with “anger” and “happy” would be treated as 0.5 anger and 0.5 happy). Or, specific weights could be given (i.e 0.7 anger, 0.3 happy).

For (b), this is likely due to how our NLP model works. TF-IDF does not care about order. It simply cares about which terms are associated with each category, and how many of those terms are in the text. The term “upset”, for example, may be associated with “anger” and “sadness”. It doesn't matter if the sentence is, “i am feeling upset” or “i am not feeling upset anymore”. Patching this flaw would likely involve putting emphasis on the order where terms appear in-addition to TF-IDF.

## CONCLUSION

Given more time, there would be several avenues we would investigate. The first would be normalizing our training data so that every emotion has the same amount of support. This would be to test the multifield expert bias to see if it is a result of differing support. Another would be obtaining or building new training data that could accept multiple emotions, making our model more complex and multifaceted. Finally, we would compare our model to **large language models** (LLM) such as ChatGPT and Google Gemini to see how they compare to ours.

# CREDITS

- GitHub - https://github.com/ToothlessTheNightFury/bigDataFinalProject
- Special thanks to Author A, Author B, Author C, and Author D for allowing their texts to be analyzed by our model