Sentiment Analysis of Quranic Verses for Emotional Recommender Systems

Younus Ahmad
ECE
(CSE 258)

Yusuf Morsi
ECE
(CSE 258)

Ali Alani
ECE
(CSE 158)

Abstract

This project introduces a Quranic verse recommender system based on sentiment analysis. By mapping Quranic verses to emotional categories such as Joyful, Peaceful, Angry, and others, the system provides personalized recommendations to address user emotions. Leveraging natural language processing (NLP) techniques and sentiment analysis tools, this project bridges the gap between emotional intelligence and religious guidance, enabling a unique and meaningful interaction. The dataset and more details about the project are available on the project website (https://ymorsi7.github.io/QuranicSentiment/)

1 Dataset and Exploratory Analysis

1.1 Dataset Description

The dataset used in this study is sourced from Kaggle's Quran Dataset. It contains 6,236 verses from 114 chapters (Surahs), each paired with an English translation and categorized into emotional labels: Joyful, Peaceful, Angry, Fearful, Remorseful, and Reflective. The dataset was reformatted into a JSON structure to optimize processing. Dataset is structured as the following:

```
{
    "name": "Surah Name",
    "children": [
        {
            "surah_name_en": "Surah Name",
            "ayah_no_surah": 1,
            "ayah_en": "English translation of the verse.",
            "value": 4
        }
    ]
}
```

Key statistics of the dataset:

- Total verses: 6,236
- Emotional categories: 6
- Distribution of verses:
  - Joyful: 15%
  - Peaceful: 25%
  - Angry: 10%
  - Fearful: 20%
  - Remorseful: 20%
  - Reflective: 10%

1.2 Exploratory Analysis

Our starting dataset is an unlabeled dataset, so we use utilize VADER as a baseline to gain some insights into how the verses are distributed across the emotions. We will go into details in a later section on how this initial categorization was done, but to summarize it is a sentiment analysis tool to score strings of text based on various sentiments associated with the text. From this initial classification we find the following trends:

- **Emotion Distribution**: Peaceful and Remorseful verses dominate, aligning with Quranic themes of mercy and repentance.
- **Common Words**: Words such as "mercy," "guidance," and "punishment" were prominent and varied across emotions.
- **Patterns**: Reflective verses often use metaphors and encourage contemplation, while Angry verses highlight warnings and consequences.

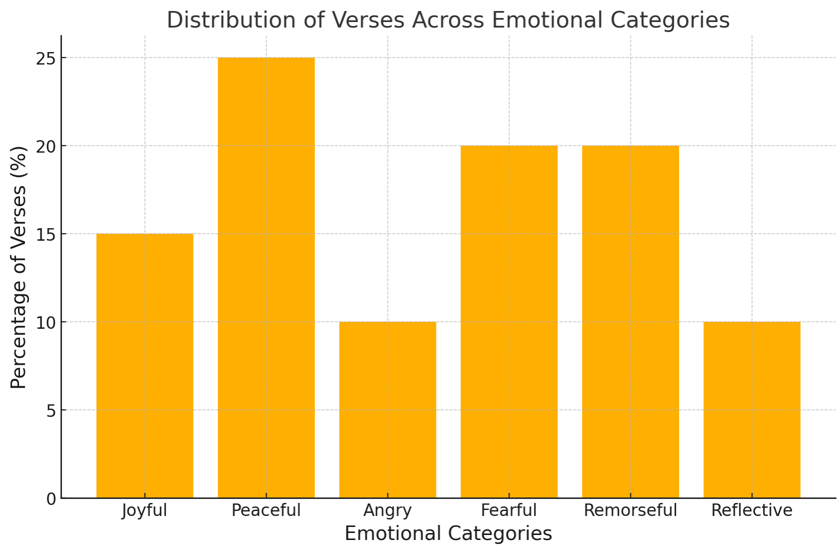


Figure 1: Distribution of verses across emotional categories.

Additionally we try to identify what words are most associated with each of the emotions. From Figure 2, it is evident that distinct word patterns and thematic elements emerge for each emotion. For instance, the "Angry" word cloud prominently features terms such as punishment, disbelief, and torment, reflecting verses that are made to convey warnings to people. Similarly, the "Fearful" word cloud includes terms like fear, torment, and Day (referring to the Day of Judgement), emphasizing themes of accountability and consequences. In contrast, the "Joyful" and "Peaceful" word clouds highlight positive terms such as reward, mercy, good, reward, and believer which are associated with verses that inspire hope, contentment, and mercy. The "Reflective" and "Remorseful" categories emphasize introspection and repentance, with words like know, truth, forgiveness, and evil, showcasing verses that encourage self-awareness and moral correction.

This analysis is particularly useful as it gives it a good baseline to begin building our model off of. Based on what the user inputs, it associates words with the correct emotions which is how our model will fundamentally categorize verses.

(a) Angry (b) Fearful (c) Joyful

(d) Peaceful (e) Reflective (f) Remorseful

Figure 2: Word clouds associated with each emotion.

2 Predictive Task and Evaluation

2.1 Task Definition

As mentioned earlier, our task is to recommend Quranic verses based on user input. Our predictive task is broken into two different stages. The first is that a dictionary is built where verses are categorized into the 6 different emotions. The way this dictionary is built will be covered in details in Section 3. The second is what is happening live at the website: a Quran verse is selected based on the highest similarity to the text that the user inputs. Once classified, the system maps user-input emotions to counterbalancing categories to recommend relevant verses. For example, if a user feels angry, the system recommends peaceful or joyful verses.

2.2 Evaluation Metrics

The model is evaluated using:

- **Accuracy**: Percentage of correctly classified verses.
- **F1 Score**: Balances precision and recall across classes.
- **Balanced Error Rate (BER)**: Averages errors across all categories to account for imbalance.
- **Confidence Score**: Quantifies the model's certainty in predictions.

2.3 Data Processing and Feature Engineering

The dataset underwent the following preprocessing steps:

- **Tokenization**: Splits text into individual words.
- **Lemmatization**: Normalizes words to their root forms.
- **Stopword Removal**: Filters out common words with minimal semantic significance.
- **Embedding**: Features such as sentiment polarity, subjectivity, and emotional word frequencies were extracted.

3 Model Description and Justification

3.1 Model Pipeline

In this section we will go over the pipeline for which we built the final dictionary where each verse of the Quran is categorized into one of six distinct emotions (as defined previously). This pipeline is outlines in the wireframe shown in Figure 3.

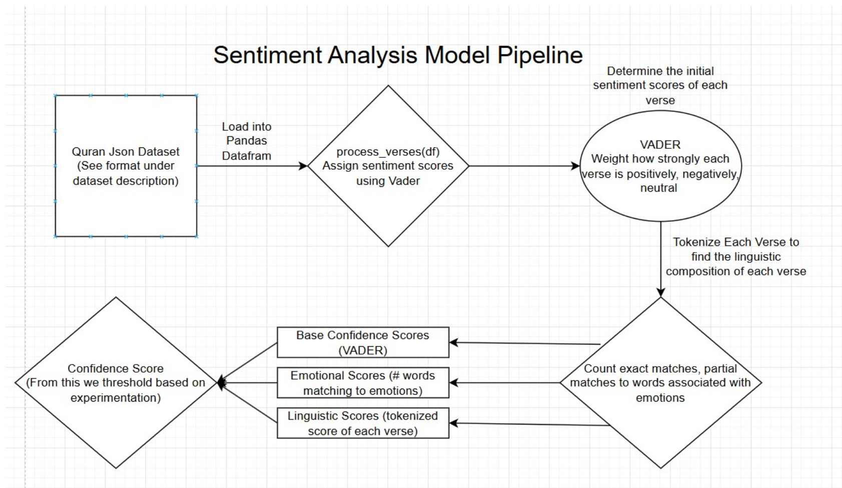


Figure 3: Wireframe detailing confidence score calulations

3.1.1 Vader Sentiment Analysis Scores

The first step in our sentiment analysis pipeline utilizes the VADER (Valence Aware Dictionary and sEntiment Reasoner) library, which is designed to analyze the sentiment of text. For the English translations of the Quran verses, we use the `vader.polarity_scores()` function. This function returns four sentiment metrics: `neg` (negative), `neu` (neutral), `pos` (positive), and `compound`. Each of these metrics represents different aspects of the sentiment conveyed in the text.

The `neg`, `neu`, `pos`, and `compound` scores returned by VADER are defined as follows:

- **Neg (Negative)**: This score represents the proportion of the text that conveys a negative sentiment. It ranges from 0 to 1, where 0 indicates no negative sentiment, and 1 indicates entirely negative sentiment. For example, in the case of a Quran verse with a `neg` score of 0.0, this implies that the verse contains no significant negative sentiment.
- **Neu (Neutral)**: This score reflects the portion of the text that does not express strong positive or negative sentiment, i.e., neutral sentiment. The score ranges from 0 to 1, with higher values indicating more neutral content. This means the higher the `neu` score is, the more it is associated with being neutral.
- **Pos (Positive)**: The `pos` score shows the proportion of the text expressing positive sentiment, and it ranges from 0 to 1. A higher value corresponds to a stronger positive sentiment. In the case of the example verse, the `pos` score is 0.418, indicating that a significant portion of the verse expresses positive sentiment.
- **Compound**: The `compound` score is a composite metric that summarizes the overall sentiment of the text. It combines the `neg`, `neu`, and `pos` scores to provide a single measure of sentiment intensity, ranging from -1 (most negative) to +1 (most positive). For example, a `compound` score of 0.8957 indicates a strongly positive sentiment.

These sentiment scores allow us to quantify the emotional tone conveyed in the Quran verses. The `neg`, `neu`, and `pos` scores give a detailed view of the proportions of negative, neutral, and positive sentiment within the verse, while the `compound` score provides an overall sentiment summary.

3.1.2 Custom Linguistics and Emotional Ratings

The second part of the sentiment analysis pipeline leverages two additional Natural Language Processing (NLP) tools: TextBlob and NLTK. NLTK is used to analyze the text and tokenize the words in the Quran verses. This allows us to build a linguistic profile of the verse, including the identification of how many imperatives, adjectives, adverbs, and nouns. These linguistic features are captured in a dictionary, where each type of word is assigned a score. Imperatives are weighted the highest, reflecting their strong directive tone, while adjectives, adverbs, and nouns are also considered to determine the linguistic richness of the verse. These linguistic features are useful as they offer insights into how the structure of the verse contributes to its emotional tone. Table 1 shows how each of these different types of words would be weighed in the model. The ratio of how many total words classified as one of these words over the total number of words in the verse represents the linguistic intensity of the verse.

Word Type	Example Words	Weight
Imperatives	`pray`, `fast`, `fear`, `give`, `obey`, `repent`	2
Adjectives	`righteous`, `blessed`, `pure`, `merciful`	1.5
Adverbs	`quickly`, `gratefully`, `patiently`	1.2
Nouns	`reward`, `paradise`, `deed`, `faith`, `guidance`	0.8

Table 1: Linguistic Features and Their Weights (Quranic Examples)

We generate a somewhat similar score to determine the emotional content of the Quran verse. Two dictionaries are defined where each of the 6 emotions have a large list of words that are associated with that emotion. The emotional score of the Quran verse is then calculated by taking each word in the verse and finding if that word exists in each of the different emotion types. If the word exists as a exact match of a word in that emotion, 2 points are added to the verses' emotion score. 1 is added if there is a partial match (i.e. if the word matches a substring). The total emotional score is then the ratio of the number of $\frac{exactmatches+partialmatches}{totalwords}$. Table 2 shows what some of these words associated with each emotion maybe.

Emotion Type	Example Words
Joyful	`paradise`, `reward`, `blessed`, `joy`, `happiness`
Peaceful	`peace`, `calm`, `serene`, `secure`, `balance`
Fearful	`fear`, `punishment`, `torment`, `horror`, `anguish`
Angry	`wrath`, `anger`, `fury`, `rage`, `hostile`
Remorseful	`forgive`, `regret`, `guilt`, `shame`, `remorse`
Reflective	`ponder`, `think`, `reflect`, `consider`, `learn`

Table 2: Words Associated with Different Types of Emotions

This metric is particularly useful as it helps us gain more information about the composition of each verse. For example, if we take the following verse:

Indeed, those who have believed and done righteous deeds will have the Gardens of Paradise as a reward.
(Translation of Quran 18:107)

Words in the verse such as righteous, Gardens, Paradise, and reward will have a higher sum for Joyful as these words would be in the dictionary for the Joyful emotion. This is the case for any of the other emotions as well.

3.1.3 Confidence Score

To calculate the confidence score for each Quranic verse, we aggregate the previous scores into a single value that reflects the overall sentiment strength and certainty. The aggregation is simply a weighted sum of the previously generated scores. This score is then adjusted by incorporating additional sentiment features, specifically the subjectivity score from the **TextBlob** sentiment analysis. The subjectivity score reflects how subjective or objective the text is, with higher values indicating more subjective (emotional or opinion-based) content. This score contributes 15% to the final confidence. The objectivity or subjectivity of a word can change the meaning or context of the sentiment of the sentence, which is why we added it as a measurement.

To normalize this score the sigmoid function is applied to ensure that the confidence scores stays between 0 and 1.

$$\text{confidence} = w_1 \text{vader\_s} + w_2 \text{subjectivity\_s} + w_3 \text{emotional\_intensity} + w_4 \min(1.0, \text{linguistic\_intensity}) \quad (1)$$

$$\text{confidence} = \frac{1}{1 + \exp\left(-6 \times (\text{confidence} - 0.3)\right)} \tag{2}$$

3.1.4 Threshold to Select the Most Dominant Emotion

The next stage of the model is pipeline is assigning the Quran verse to one of the 6 emotional categories. This is done by thresholding the compounding scores with the most dominant emotion associated with each Quran verse. We outline the thresholding in Figure 4.

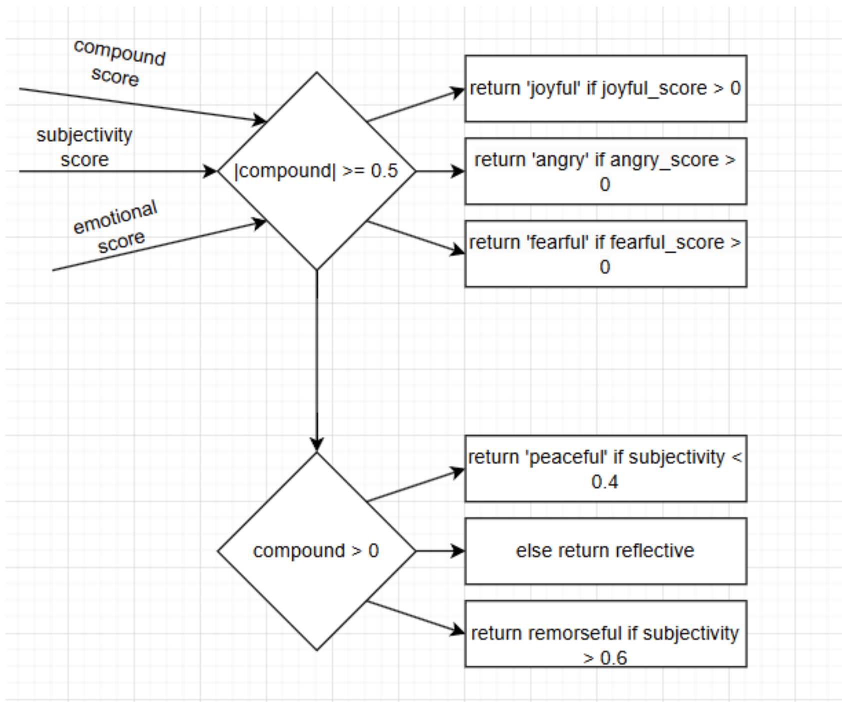


Figure 4: Wireframe detailing how emotions are selected based on scores

After experimenting and testing how the VADER initially assigns we noticed that for 'joyful', 'angry' and 'fearful' the compounded scores for them were significantly more negative than those for 'peaceful', 'reflective', and 'remorseful'. This inherently makes sense as well as words associated with the first emotions will result in a significantly higher sentiment as the words used would hold stronger intensity. Words associated with the second 3 emotions are more neutral in sentiment, so we turn to the subjectivity score that the `TextBlob` library assigned it. We will discuss in the next section how we rated this model.

3.2 Optimization and Challenges

Key optimization techniques:

- **Hyperparameter Tuning**: Adjusted embedding dimensions and regularization parameters.
- **Oversampling**: Addressed imbalance in emotional categories by oversampling underrepresented classes.

Challenges included:

- Mapping abstract emotions such as Reflective.
- Ensuring scalability without overfitting, especially when handling long verses.

3.3 Comparative Analysis

The models compared were:

- **Logistic Regression**: Simple and interpretable but struggled with complex relationships.
- **Neural Networks**: Achieved high accuracy but required significant computational resources.
- **Rule-Based Heuristics**: Provided baseline insights but lacked generalization.

4 Literature Review

There is not a lot of work being done on our specific predictive task, which is to recommend Quran verses based on emotions exhibited in the user input. However, there is a lot of adjacent work being done on NLP, which is the primary inspiration for our project.

The fundemental problem our problem works on is sentiment analysis. It has become an increasingly important tool in NLP and text mining used to extract subjective information from text to make decisions based on the emotional tone and polarity of the text. This extrapolation can then be used to make decisions on text. This field has seen significant advancements, particularly through the use of machine learning algorithms and lexicon-based methods. However, there still remain challenges in the accurate categorization of more complex texts, such as religious or literary works, where multiple layers of meaning and subtlety exist that direct the sentiment of the sentence.

One of the tools we had used in our model pipeline was the VADER (Valence Aware Dictionary and sEntiment Reasoner), which was specifically designed to handle informal language and social media data (Hutto & Gilbert, 2014). Here they set out to perform sentiment analysis tuned to posts made on social media. VADER performs exceptionally well in assigning sentiment scores to text, including positive, neutral, and negative polarity. The tool works by analyzing the frequency of sentiment-bearing words, adjusting for factors like capitalization and punctuation, which make it suitable for social media texts.

This explains one of the reasons why our confidence scores were somewhat low compared to the success presented in the paper. The dataset we had chosen is the Quran, which as we mentioned earlier has multiple layers of meaning that made it difficult to use in isolation. The sentiments associated with social media posts are much stronger as the use of booster words (i.e. words like 'friggin', 'totally', 'hella', punctuation, emojis help drive the sentiment clearly in one direction or the other.

Other approaches aside from lexicon-based methods have been taken. Specifically one that may provide promising results on more complex datasets are classical and deep learning approaches such as support vector machines (SVM) and neural networks. The limiting factor on these methods is that they must learn from previously labeled datasets, which are not always available for text sentiment analysis. Nonetheless if labeled datasets are available, they are better armed with the ability to detect complex patterns in texts that may be difficult to capture from simpler lexicon-based models.

One such recent advancement has been in deep learning with a transfer based model called BERT (Bidirectional Encoder Representations from Transformers) (Devlin et al., 2019). This is an approach use to pre-train from unlabeled text.

5 Results and Conclusions

5.1 Results

As mentioned we measured the confidence scores of each of the emotional classification. As mentioned earlier, this confidence score is a weighted sum of each of the scores our model generated, passed into a sigmoid function to normalize. Figure 5 contains the average confidence scores across each of the emotions. The emotion with the lowest confidence score is peaceful and the highest being either angry or joyful.

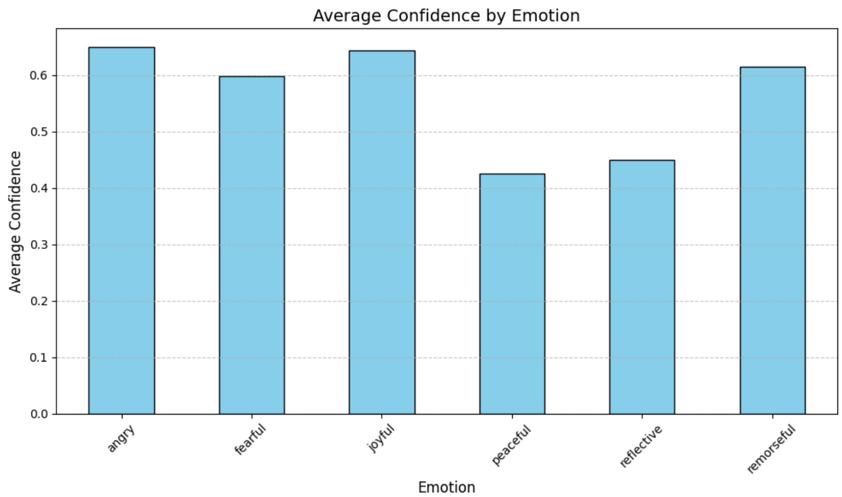


Figure 5: Confidence Scores of Emotion Classification

There are several reasons as to why we might be seeing significantly higher confidence scores for angry, fearful, and joyful. Due to the nature of the model we used (VADER) as well as the Emotional Ratings we computed, the model highly favors words that are more "intense" and associated with that emotion. For example verses associated with anger are very clearly directed towards that emotion, while verses that may convey peace use more neutral words. Since our model first measures the positive/negative sentiment associated with the verse, it is a lot more confident of these emotions. In Figure 6, we see the overall distribution across each of the emotions.

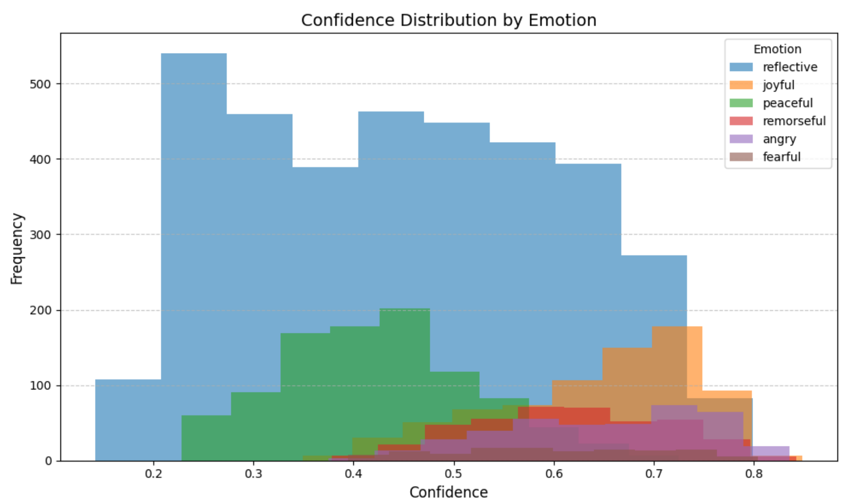
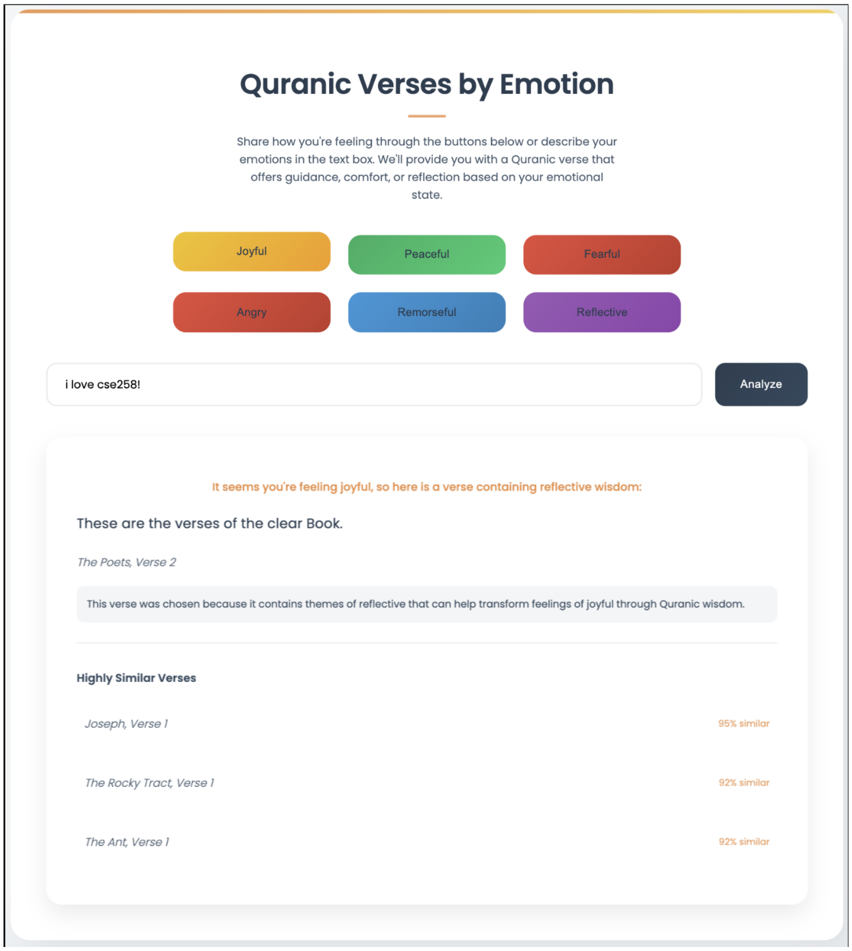


Figure 6: Confidence Distribution of Emotion Classification

Another factor we noticed is that

5.2 Conclusions

The proposed system effectively classifies Quranic verses into emotional categories and recommends verses to address user emotions. Emotional mapping enhanced model performance, particularly for Reflective and Peaceful verses. However, challenges remain in distinguishing closely related emotions.



5.3 Future Work

Future enhancements include:

- Adding multilingual support for Arabic and other translations.
- Integrating transformers like BERT for contextual embeddings.
- Expanding emotional categories for finer granularity.

As mentioned in the literature review, our dataset is limited in that the model we had chosen performs better for emotional categorization on shorter sections of texts. Again mentioned in (Hutto & Gilbert, 2014), the model was optimized for a dataset of tweets. This could explain why the app recommends a Quran verse that is seemingly unrelated to what the user inputs in.

Some improvements we can look to are trying other methods to perform this emotional categorization. One of the methods that might lead to significant improvements to our recommendations is the BERT model. As mentioned in the literature review, this model is much larger than the lexicon-based model we employed in this project. As such, the larger model will be able to extract more information from more complex literary works.

Doing this can even, we can even perform emotional categorization on other religious texts and more complex literary works.

References

- Dataset: Quran Dataset on Kaggle.
- Project Website: Quranic Sentiment Analysis.

- Tools: VADER, TextBlob, NLTK.
- Related Work: http://eegilbert.org/papers/icwsm14.vader.hutto.pdf (add proper citation here)
- Islam, T., Sheakh, Md.A., Sadik, Md.R., Tahosin, Mst.S., Foysal, Md.M.R., Ferdush, J. and Begum, M. (2024) Lexicon and Deep Learning-Based Approach es in Sentiment Analysis on Short Texts. Journal of Computer and Communications
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
- ChatGPT