Comprehensive Sentiment Analysis of Religious Content Naive Bayes Algorithm Model

Hersatoto Listiyono1,*, Zuly Budiarsao1, Susi Susilowati2, Agus Perdana Windarto3

1Prodi Manajemen Informatika, Universitas Stikubank, Semarang, Indonesia
2Prodi Sistem Informasi Kampus Kota Bogor, Universitas Bina Sarana Informatika, Jakarta, Indonesia
3Prodi Sistem Informasi, STIKOM Tunas Bangsa, Pematangsiantar, Indonesia

Email: 1*hersatotolistiyono@edu.unisbank.ac.id, 2zulubudiarsao@edu.unisbank.ac.id, 3susi.sss@bsi.ac.id,
4agus.perdana@amiktununsbangsa.ac.id

Correspondence Author Email: hersatotolistiyono@edu.unisbank.ac.id

Abstract—This paper delves into sentiment analysis of online religious content utilizing the Naive Bayes algorithm to decipher the array of sentiments present in religious discussions. By tailoring this algorithm to the complexities of religious language, the study reveals hidden sentiments, offering valuable insights for researchers, policymakers, and communities. The findings demonstrate that the sentiment analysis model performs robustly, with a precision of 84.78%, a recall of 82.98%, and a balanced F1 Score of 83.87%, indicating high accuracy in sentiment identification and effectiveness in capturing a significant portion of actual sentiments. The overall accuracy of the model stands at 75.10%, affirming its successful adaptation to the intricacies of religious discourse. These results not only deepen our understanding of sentiment analysis in the realm of faith and spirituality but also have practical implications for enhancing interfaith dialogue, fostering mutual understanding, and guiding decision-making in religious and social organizations. This research makes a significant contribution to the growing field of sentiment analysis, providing a methodological framework for exploring the nuanced sentiment landscape within the domain of faith and spirituality.

Keywords: Sentiment Analysis; Naive Bayes Algorithm; Religious; Twitter; Interfaith Dialogue; Digital Humanities; Faith; Social Implications

1. INTRODUCTION

In the digital age, social media platforms and online forums have become hubs for expressing diverse opinions, beliefs, and sentiments [1], [2]. One prominent area of discussion on these platforms revolves around religious content, where individuals share their thoughts, beliefs, and emotions related to various religious topics. Religion is an important dimension in human life that includes beliefs, values, and spiritual practices that shape a person's identity and life orientation.

This is a very personal and in-depth aspect, guiding individuals in understanding the meaning of life, existential goals, and the moral ethics that govern their behavior. Religion is often manifested through involvement in worship practices, rituals, and the development of spirituality. Religious content refers to any form of information, material or expression related to religious beliefs and spirituality. This can include religious writings, religious speeches, religious services, scriptures, and more. Religious content has a significant role in spreading religious values, strengthening religious identity, and facilitating spiritual growth in communities[3], [4].

In society, the lives of religious communities can be reflected in various aspects. First of all, religion plays a key role in shaping the moral and ethical norms applied in society. Religious people may have ethical guidelines provided by their religious teachings to guide daily behavior. In addition, the lives of religious communities can be reflected in their involvement in social and humanitarian activities, with many religious communities active in providing assistance to those in need.

Apart from that, interactions between religious communities also play an important role in building tolerance and respect for diversity. Although there may be differences in beliefs, religious communities often work together to create an environment that is mutually supportive and encourages cooperation. Therefore, the diversity of religious life in society is an integral part of social balance and harmony between individuals with different religious backgrounds [5].

Understanding the sentiments expressed in religious content is crucial for researchers, policymakers, and communities alike, as it provides valuable insights into the public's attitudes, opinions, and concerns [6]. Despite the vast amount of religious content available online, comprehensively analyzing the sentiments expressed within it remains a challenging task. Traditional methods of sentiment analysis often struggle to handle the complexity and nuances present in religious discourse. This is where advanced techniques, particularly those from the realm of artificial intelligence, come into play.

In recent years, machine learning algorithms have proven to be effective in deciphering complex patterns within textual data, making them invaluable tools for sentiment analysis [7]. The Naive Bayes algorithm, a probabilistic machine learning model, has garnered significant attention in the field of natural language processing. Its simplicity and efficiency make it an appealing choice for sentiment analysis tasks[8], [9]. However, applying the Naive Bayes algorithm to religious content brings about its own set of challenges. Religious discourse often involves figurative language, metaphors, and context-specific meanings, making it distinct from other forms of communication. Consequently, adapting the Naive Bayes algorithm to effectively analyze sentiments within religious content requires a nuanced approach [10]. In previous research conducted by [11] with the title...
"Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler: Twitter" it was obtained from the results of this research that the positive sentiment polarity value for the Jokowi-Ma'ruf Amin pair was 45.45% and the negative value was 45.45%. 54.55%, while the Prabowo-Sandiaga pair received a positive sentiment score of 44.32% and negative. 55.68%. Then the combined data was tested from the training data used by each presidential candidate and obtained an accuracy of 80.90% = 80.1%.

In this study, a comparison was carried out using the Naïve Bayes, SVM and K-Nearest Neighbor (K-NN) methods which were tested using RapidMiner, producing a Naïve Bayes accuracy value of 75.58%, an SVM accuracy value of 63.99% and a K-NN accuracy value of 73. 34%. Meanwhile, the next previous research was conducted by [8] with the title "Twitter Sentiment Analysis as an Evaluation and Service Base on Python Textblob", this research classified tweets with the keywords indihome, myindihome, useetv, and wifi.di. Next, several data preprocessing techniques, sentiment analysis, and visualization were carried out in the form of histograms, pie charts, and word clouds. Of the 3324 tweets analyzed, 34.4% of the tweets were positive, 16.1% of the tweets were negative, and 49.6% of the tweets were neutral. In this article, we delve into the realm of sentiment analysis in religious content, aiming for a comprehensive understanding of the various sentiments expressed by individuals across different religious discussions [12]. We focus specifically on harnessing the power of the Naïve Bayes algorithm to navigate the intricacies of religious language and uncover underlying sentiments[13], [14]. By developing a tailored Naïve Bayes model, we address the unique challenges posed by religious discourse and pave the way for a nuanced analysis of sentiments in this domain [15], [16].

This study not only contributes to the academic understanding of sentiment analysis but also holds practical implications for diverse fields, including sociology, religious studies, and digital humanities. By gaining insights into the sentiments prevalent in religious content, we can foster better interfaith dialogue, promote understanding among diverse communities, and facilitate more informed decision-making processes for religious and social organizations[17]. In the following sections, we will explore the methodology employed in adapting the Naïve Bayes algorithm for religious sentiment analysis, present our findings, and discuss the implications of this research. Through this comprehensive analysis, we aim to shed light on the intricate tapestry of sentiments within religious content, offering a valuable resource for researchers and practitioners seeking a deeper understanding of the human experience in the realm of faith and spirituality.

2. RESEARCH METHODOLOGY

2.1 Related Work

a) Sentiment Analysis in Textual Content an integral aspect of natural language processing (NLP), has witnessed significant advancements in recent years. Researchers have explored various techniques, including rule-based approaches, machine learning, and statistical algorithms, to discern emotional tones and opinions embedded in textual content. The evolution of sentiment analysis methodologies provides a foundational backdrop for understanding the current landscape of techniques applicable to diverse domains[18].

b) Naïve Bayes Algorithm in Sentiment Analysis Among the diverse algorithms employed in sentiment analysis, the Naïve Bayes algorithm has proven to be particularly effective. Leveraging probabilistic principles, Naïve Bayes classifiers have demonstrated notable success in text classification tasks, making them a popular choice for sentiment analysis applications. Previous studies showcase instances where Naïve Bayes has been adeptly applied to analyze sentiment in various contexts, highlighting its versatility and reliability in sentiment classification tasks [19], [20].

c) Religious Content Analysis introduces unique challenges due to the distinct nature of religious language, cultural nuances, and diverse interpretations. Researchers have delved into the intricacies of sentiment within religious texts, considering the impact of religious vocabulary and the complexity of conveying emotions in a context steeped in cultural and spiritual significance. Understanding the nuances of sentiment analysis within religious content is pivotal for the development of effective models tailored to this specific domain [21].

d) Hybrid Approaches in Sentiment Analysis Combining different methodologies such as statistical techniques and machine learning models, have emerged as promising solutions to enhance the accuracy and robustness of sentiment analysis systems. Researchers have explored the integration of diverse techniques, seeking to capitalize on the strengths of each approach. These hybrid models present an opportunity to address the challenges posed by the intricate nature of sentiment analysis, offering improved performance and adaptability across varied content types, including religious discourse [11], [22]–[24].

Several processes are performed in this word processor, firstly we collect data, in this study we using data tweets are collected from Twitter social media by using a crawler. Furthermore, we parse the tweets are get by describing it verbatim. Hereinafter, we do the tokenization process that is cleaning the tweet and selecting the meaningful words. Then, we do text mining using naïve bayes method.

2.2 Method

This research uses a flowchart methodology as shown in Figure 1. Figure 1 is a flowchart from a sentiment analysis program as a whole. The detailed steps in the research is explained as follows:
In Figure 1, the research workflow is carried out, which involves collecting data in one day, which is then collected and compiled into one. Once the data is collected, the classification process begins. Using the textblob library, the goal of this step is to differentiate positive, negative, and neutral data. After that, the pre-processing step starts by cleaning the tweet data. Text pre-processing includes symbol removal (URLs, special characters), negation handling (word standardization), stemming (removal of affixes and suffixes), tokenization (separating a series of words in a sentence), and case folding (converting to lowercase). Then the final step is data visualization that shows these three types. There are histograms, pie charts, and word clouds.

1. Data Collection:
The first step involved the collection of a diverse dataset comprising religious content from various online platforms, including social media networks, forums, blogs, and religious websites. The dataset encompassed discussions from different faith traditions to ensure a comprehensive analysis of sentiments across religions.

<table>
<thead>
<tr>
<th>id</th>
<th>label</th>
<th>tweet</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>Cricket is a religion not just in India but in whole subcontinent. DEADLINE: NOVEMBER 30 2023 â€œ PrimeProgressng seeks applications for its Religion for Change Fellowship aimed at training and supporting Nigerian/African journalists. 10 successful applicants wi... @Ade__Ayomide @JakesOlasupo @GazetteNGR You’re a big â€œidiot you disrespect these ministers of God practicing their religion peaceful with accusations of hate and vengeance against non Christians. U ... @OleOestlid @enur72 Value and purpose is subjective. You can empirically prove it by not finding two individuals with exactly the same choices in life. This has nothing to do with religion or beli... @_NavyBrat There you have it folks. the religion of peace. The enemy is inside the gates. @PoppetSaab @ZakHu18111965 @tathagata2 @TheSyedHaq I don't condone disrespecting others religion. its â€œid is a religion of peaceâ€œ@QamareAlam21 @TheSyedHaq And there has been many new religions along the way and most of them is like sub-part of Sanatan Dharm Jain. Buddha. Sikh As all these religions celebrate the existence ...</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>Islam has no place in the west. We are being forced to call Nazism a religion. Itâ€™s time to stop the insanity. There are a ton of rich Muslim countries. the only reason Muslims are in the west i... @metpoliceuk The final stage is the establishment of a totalitarian Islamic theocracy. eradicating all other religions and freedoms. @NdabeniMzukisi @eNCA Itâ€™s patheticâ€œthe ANC having nothing to do with any sporting, financial or any other...</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>17</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>18</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>19</td>
</tr>
</tbody>
</table>
but if your going to mention flying donkey or whatever which isn't even accu...

@Bellius_27 Religion in Europe. caused more wars than anything. for example crusades. and it is responsible for more deaths than III Reich and USSR combined. People are getting smarter and they tr...

Any religion that has nothing but palangis at the top is a major red flag.

Presidents going all the way to the beginning of the cult all y? δŶ$©

[[&gt Sixth Commandment &lt ]] You shall not murder///God values humans highly as he wants us to values cq&amp;gt choose life as well Add Commandments 3.8.9.10.11.12.13.14.15 to this [Orthodoxy Re...

@jacksonhinklle What about jews religion?

For the record, on the label the number 0 is for negative content, 1 is for positive content.

2. Data Preprocessing:
The collected textual data underwent preprocessing techniques to clean and prepare it for analysis. This stage included tasks such as text normalization, tokenization, and removal of stop words, special characters, and irrelevant symbols. Additionally, stemming and lemmatization were applied to standardize the words, ensuring consistency in the dataset.

Table 2. Tweets Data after Pre-processing

<table>
<thead>
<tr>
<th>id</th>
<th>label</th>
<th>tweet</th>
<th>id</th>
<th>label</th>
<th>tidy_tweet</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>0</td>
<td>a sin to do that. but if your going to mention flying donkey or whatever which isn't even accu... gains created by good honest South Africans of all colour. religion or creed. Voetseki...</td>
<td>21</td>
<td>1</td>
<td>Any religion that has nothing but palangis at the top is a major red flag. Presidents going all the way to the beginning of the cult all y? δŶ$© And he has reverted to Islam. May Allaah reward Tasleem for his efforts &amp; make you steadfast upon the religion. Allaah is the one who guides whoever He wills &amp; He has willed to grant you ...</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>Any religion that has nothing but palangis at the top is a major red flag...</td>
<td>26</td>
<td>1</td>
<td>@9ja_gistme @drzakiranaik Essence of all religions is exactly same yet people like Zakir Naik get royal treatment as if there are no great saints in Islam. What a shame. There are very highly en...</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>you're a big !diot you disrespect these ministers of God practicing their religion peaceful with accusations of hate and vengeance against non Christians.U ...</td>
<td>31</td>
<td>1</td>
<td>@DougAMacgregor disagree.Erdogan uses this Muslim brotherhood tactic where he barks some rebellious. courageous speeches/threats just to keep his angry base under control.Gaza isn't his fight neith</td>
</tr>
</tbody>
</table>

3. Comprehensive Sentiment Analysis:
The trained Naive Bayes model was applied to the entire dataset, conducting a comprehensive sentiment analysis of the religious content. The analysis involved categorizing the sentiments expressed into positive, negative, neutral, and nuanced emotional states. The results were analyzed and interpreted to identify patterns, trends, and insights regarding the sentiments prevalent in different religious discussions.
4. Ethical Considerations:
Ethical guidelines and considerations were adhered to throughout the research process. Anonymity and privacy of the users contributing to the dataset were preserved, and the research aimed to respect the diverse beliefs and opinions expressed within the religious content.

5. Model Training and Testing:
The adapted Naive Bayes algorithm was trained on a portion of the preprocessed dataset and testing on another distinct subset. Testing techniques were employed to assess the model's performance, ensuring its accuracy, precision, recall, and F1-score across various sentiment categories. Rigorous testing was conducted to validate the model's effectiveness in capturing the nuances of sentiments in religious content.

6. Feature Extraction:
To facilitate the Naive Bayes algorithm, feature extraction was employed to convert the textual data into numerical vectors. Term Frequency-Inverse Document Frequency (TF-IDF) and word embeddings techniques were utilized to capture the semantic meanings and importance of words within the religious content.

7. Naive Bayes Algorithm Adaptation:
The Naive Bayes algorithm was adapted to the unique characteristics of religious discourse. Special attention was given to handling figurative language, idiomatic expressions, and context-specific meanings commonly found in religious texts. Training the model involved using a labeled dataset, carefully annotated to reflect the diverse range of sentiments expressed in religious discussions.

8. Conclusion and Implications:
The research methodology's findings were analyzed, leading to conclusions about the sentiments expressed in online religious content. Implications for interfaith dialogue, community understanding, and decision-making processes within religious and social organizations were discussed, highlighting the practical significance of the study's outcomes.

2.3 Sentiment Scoring
The polarity of words has the basic number of features based on the selection process. English language words assign a score by referring dictionary. This scoring module determines the score of sentiments during the analysis of data. Naïve Bayes classifiers are used to classify the sentiment. Naïve Bayes algorithm is used to classify the sentiment and this sentiment orientation performs well with more accuracy.

2.4 Naïve Bayes Classifier
Naïve Bayes is a simple probabilistic classifier that calculates a set of probabilities by summing the frequencies and combinations of values from a given dataset. Naïve Bayes put forward by the English scientist Thomas Bayes, namely predicting future opportunities based on previous experience [25], [26]. The advantage of using Naïve Bayes is that this method only requires a small amount of training data to determine the parameter estimates needed in the classification process. To solve the Naive Bayes method, it can be done with the following equations [27]:

\[ P(H\mid X) = \frac{P(H)\cdot P(X\mid H)}{P(X)} \]  

Where:
- \( X \): Data with unknown class
- \( H \): Hypothesis data is a specific class
- \( P(H\mid X) \): The probability of the hypothesis \( H \) under condition \( X \) (probability posteriori)
- \( P(H) \): Probability of the hypothesis \( H \) (probability prior)
- \( P(X\mid H) \): The probability of \( X \) based on the conditions in hypothesis \( H \)
- \( P(X) \): Probability of \( X \)

Further elaboration of the Naive Bayes formula is carried out by explaining in detail (C|X1,...,Xn) using the multiplication rule as follows.

\[ P(C|X1,...,Xn) = P(C) P(x1,..., xn|C) \]
\[ = P(C) P(X1|C) P(X2,...,Xn|C,X1) \]
\[ = P(C) P(X1|C) P(X2|C,X1) P(X3,...Xn|C,X1,X2|C) P(X1|C) P(X2|C,X1) \]
\[ = P(X1|C) P(X2|C,X1) P(X3|C,X1,X2) P(Xn|C,X1,X2,...,Xn-1) \]  

It can be seen that the more complex factors that affect the probability value, the more impossible it is to calculate these values one by one. As a result, the calculation process will be increasingly difficult to do, so this is where the assumption of very high independence is used, that each attribute can be independent of one another. With these assumptions, equation is needed:

\[ P(Xi|Xj) = \frac{P(Xi) P(Xj)}{P(Xi)} = P(Xi) \]
For $i\neq j$, so that 
\[ P(X_i|C, X_j) = P(X_i|C) \]

From equation (3), it can be concluded that the assumption of independence makes the calculation requirements simpler. Furthermore, the description of $P(C|X_1, \ldots, X_n)$ can be simplified into From equation:

\[ P(X_2|C)P(X_3|C) \ldots P(C|X_1, \ldots, X_n) = P(X_1|C) = \prod_{i=1}^{n} P(X_i|C) \]  

Information:
\[ \prod_{i=1}^{n} P(X_i|C) = \text{branch multiplication between attributes.} \]

The above equation is Bayes' theorem which will then be used to perform classification calculations. For classification with continuous data or numerical data, use the Gaussian distribution formula with 2 parameters: mean $\mu$ and variance $\sigma$:

\[ p(X_i|C = c_j) = \sqrt{\frac{2\pi\sigma_{ij}}{2\sigma^2}} \exp \left( -\frac{(x_i - \mu_{ij})^2}{2\sigma^2} \right) \]  

Q: Opportunity
Xi : Attribute to i
Xj : Attribute value to i
C : The class you are looking for
Ci : Subclass Y to be searched for
$\mu$ : Stating the average of all attributes
$\sigma$ : Standard deviation, expresses the variance of all attributes.

In the naive Bayes method, training data and test data are needed to be classified, in naive Bayes, the more training data that is involved, the better the predicted results are given. Calculate $P(C_i)$ which is the prior probability for each subclass $C$ that will generated using the equation:

\[ P(c_i) = \frac{S_i}{S} \]  

Where $S_i$ is the amount of training data from the $C_i$ category, and $S$ is the total amount of training data. Calculating $P(X_i|C_i)$ which is the posterior probability of $X_i$ with condition $C$ using equation 4.

The advantages and disadvantages of the Naive Bayes method namely:

a) Strong against interference isolation on data.
b) If there is a missing value case when the computation process is in progress, then the object will be ignored.  
c) Can be used for irrelevant data.
Some of the deficiencies found in the Naive Bayes method, namely:

a) Must assume that between features are not related (independent) in reality, the relationship exists.
b) This relationship cannot be modeled by the Naive Bayes Classifier.

It does not require complicated iterative parameter estimation schemes, which means it can be applied to large data sets.

3. RESULT AND DISCUSSION

3.1 Result Testing Program

Python is a programming language with many paradigms, which supports object-oriented, procedural and functional programming. The program was created using Google Colab as a text editor and Python programming compiler which the author used to test this research. To test whether the program is able to meet the demands that have been set, the researcher carried out several tests as follows in table 4. The results of program testing are shown in several sections below, including:

![Figure 2. Distribution of datasets for training and testing](image)

Figure 2 shows the distribution of the dataset for training and testing, where for training data 70% of the data used for training is used, while for testing 30% of the data is used. From the graph above, you can see the level of sensitivity between training and testing data.
Table 3. Example of data that will be tokenized

<table>
<thead>
<tr>
<th>No</th>
<th>Sentimen</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Negatif</td>
<td>[Cricket, religion, just, India, whole, subcontinent]</td>
</tr>
<tr>
<td>1</td>
<td>Positif</td>
<td>[DEDLINE, NOVEMBER, seeks, applications, Religion, Change, Fellowship, aimed, training, supporting, Nigerian, African, journalists, successful, applicants, will, trained, after, which, pitches, w...]</td>
</tr>
<tr>
<td>2</td>
<td>Negatif</td>
<td>[hate, vengeance, against, Christians, failed, using, tribal, sentiments, polarise, country, want, rel...]</td>
</tr>
<tr>
<td>3</td>
<td>Positif</td>
<td>[Value, purpose, subjective, empirically, prove, finding, individuals, with, exactly, same, choices, life, This, nothing, with, religion, beliefs]</td>
</tr>
<tr>
<td>4</td>
<td>Negatif</td>
<td>[There, have, folks, religion, peace, enemy, inside, gates]</td>
</tr>
</tbody>
</table>

In table 3, the process of processing sentences into several words has been separated by character and words that have value have been taken. Table 3 is the result of the text parsing process and tokenization of sample data in Table 2. After the data is tokenized, the text processing process continues and displays the results of all the text in the tweet which will enter the text management process, to be classified by Naive Bayes. In Figure 3 below, the Wordcloud visualization is displayed as a whole.

<table>
<thead>
<tr>
<th>Figure 3. Distribution of datasets for training and testing</th>
</tr>
</thead>
</table>

Python is a programming language with many paradigms, which supports object-oriented, procedural and functional programming. From the research that has been carried out, it can be seen that the researcher's IDE is Google Colab and the researcher uses Python (3.8.1) as a programming language. Aiming to make the sentiment analysis application easy to use, researchers created a program based on visual text and images.

In Figure 3, a histogram visualization of racist and non-racist test data is displayed. Figure 4 shows a visualization of word cloud data with positive sentiment classification and neutral sentiment classification.

<table>
<thead>
<tr>
<th>Figure 4. histogram visualization of racist and non-racist test data is displayed.</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Figure 5. word cloud data visualization with positive sentiment classification and neutral sentiment classification</th>
</tr>
</thead>
</table>

3.2 Discussion

Word clouds are used to find out the words that appear most often. The more often a word appears, the larger its size, and conversely, the fewer times it appears, the smaller its size. This makes it easier to understand and find
information about each problem topic for each category. For the overall working cloud regarding sentiment towards religion, including positive, neutral or negative categories, it is shown in Figure 6 below.

![Figure 6. Overall Word Cloud for the Positive category](image)

Following are the results of the naive Bayes classification which are displayed in the form of a Confusion Matrix containing information that compares the results of the classification carried out by a system with the results that should be. Confusion Matrix is a method used to measure the performance of a classification method. There are four terms to represent the results of the classification process. There are True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). True Positive or TP value, is the number of positive data that is classified correctly by the system. True Negative Value or TN, is the amount of negative data that is classified correctly by the system. False Positive or FP values, are the number of positive data that are incorrectly classified by the system. False Negative or FN values, are the number of negative data that are incorrectly classified by the system.

![Figure 7. Confusion Matrix](image)

From Figure 9 below, the results of the Naive Bayes classification are displayed in the form of a Confusion Matrix containing information that compares the results of the classification carried out by a system with the results in the form of precision, recall, f-measure and accuracy values.

<table>
<thead>
<tr>
<th></th>
<th>Precision</th>
<th>Recall</th>
<th>F1 Score</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.8478</td>
<td>0.8298</td>
<td>0.8387</td>
<td>75.10%</td>
</tr>
</tbody>
</table>

The sentiment analysis model exhibits robust performance with a precision of 84.78%, indicating accurate identification of sentiments in religious content, while a recall of 82.98% underscores its effectiveness in capturing a significant proportion of actual sentiments. The balanced F1 Score of 83.87% reflects a comprehensive approach, ensuring accurate predictions without overlooking key sentiments. The model's overall accuracy stands at 75.10%, attesting to its successful adaptation to the complexities of religious discourse. These results collectively affirm the Naive Bayes algorithm's proficiency in providing nuanced insights into sentiments expressed across diverse religious discussions, offering valuable contributions to the understanding of emotional expressions within the realm of faith and spirituality.

4. CONCLUSION

In conclusion, our research on "Comprehensive Sentiment Analysis of Religious Content: Naive Bayes Algorithm Model" has yielded compelling insights into the intricate landscape of sentiments within diverse religious discussions. The Naive Bayes algorithm, tailored to the challenges of religious discourse, demonstrated commendable precision (84.78%), recall (82.98%), and a balanced F1 Score (83.87%), contributing to a nuanced understanding of sentiments. The model's overall accuracy at 75.10% underscores its successful adaptation to the complexities of religious language. These findings not only validate the efficacy of the applied algorithm in deciphering sentiments within online religious content but also provide a valuable tool for researchers, policymakers, and communities seeking deeper insights into the emotional expressions across various faith
traditions. This research contributes to the burgeoning field of sentiment analysis, offering a methodological framework for navigating the nuanced nuances of sentiments within the realm of faith and spirituality.

REFERENCES


