

Using the Twitter Data Classifier Web Application (TDCWA) to Identify Social Issues for the Philippines' Synod of Synodalities

Gian Lloyd B. Jacoba¹

ABSTRACT

Social Media has become one of the most reachable platforms for Filipinos to communicate with one another and share news and trending topics being discussed in the Philippines. Therefore, organizations can utilize a vast data collection and processing opportunity to help in their decision-making through Data Analytics and Machine Learning. One of the organizations that can benefit from Data Analytics and Machine Learning is the Roman Catholic Church. As the Church is currently holding its Synod of Synodalities, the Synod needs to be provided with unbiased societal issues that the Church must tackle as it moves towards an uncertain future. The paper offers a Data Analytics and Machine Learning solution using Twitter Data as the primary source of data to be processed. The study first scraped data from Twitter using a data-scraping application called Twint. Overall, 12,000 tweets were collected but had to be preprocessed. Descriptive analytics was utilized to determine the most frequently used words in the collected tweets. The social issues processed by the machine learning algorithm and discussed in the study can be used to augment and support the information already gathered by the Synod.²

¹**Gian Lloyd B. Jacoba** is currently pursuing his Masters in Information Technology at Saint Louis University, Baguio City, Philippines. He recently graduated with a Bachelor of Science in Information Technology in the same school. His research interests are Backend Development, Data Analytics and Cybersecurity.

² Acknowledgement: The study wishes to acknowledge the following persons that has helped clean the data gathered and given a hand in developing the Machine Learning models: Joshua Dullao, Klyde Rupert Formoso, Francis Nash Jasmin, Joemari Lopez, Trecia Mae Magsanoc, Athenie Andrei Marzan, Frederico João Trinidad, and Kristian Harrel Zuniga. The study would also like to thank Rico C. Jacoba and Beverly P. Ferrer for providing support in the writing of the paper.

Keywords: *data analytics, machine learning, Synod of Synodalities, Roman Catholic Church, social issues*

1. Introduction

1.1. The Church and Synodality

In October 2021, the Catholic Church began the long journey of one of the most significant consultative processes in human history aimed towards addressing various issues being faced by the Catholic Church (Sanem 2022). The process, called Synod on Synodality, allows the Catholic Church to listen intently to connect deeply to the Body of Christ (CBCP News 2021). A Synod usually only lasts about a month and involves different bishops around the universal Church. However, the Synod of Synodalities that Pope Francis has called is very special in that everyone baptized in Christ is invited to participate. The Synodality is a “journeying together,” meaning “one listens to one another.” Pope Francis writes in the Commemoration of the 50th anniversary of the institution of the Synod of Bishops in 2015 that the Church aims to be a Synodal Church, a listening church because we all have something to say and contribute. In today’s world, where technology and communication have become more accessible and widespread, the Church recognizes the importance of listening to the voice of the faithful who express their opinions and experiences on various platforms, such as social media.

1.2. Social Media and Data Analytics

Social media has become an integral part of modern society. Millions of people use platforms like Facebook, Twitter, and Instagram to share their thoughts, experiences, and opinions. As the popularity of social media has grown, so has the need for sophisticated tools and techniques to analyze the vast amount of data generated by these platforms. Other social media sites like Twitter are specialized in sharing short messages (Hudson 2020). Twitter is a microblogging platform that allows subscribers to send and receive quick notifications, known as tweets. Tweets can contain up to 280 characters and links to related websites and resources. It is considered the Internet’s pulse, providing a fast and straightforward way to discover

what's happening worldwide. Users can find the most recent news, events, and trends on Twitter (Arigo et al. 2018).

As social networking sites have expanded, data analytics techniques have been increasingly used to conduct research in fields such as politics, sociology, and finance by leveraging public data on these sites. Twitter data has become a particularly valuable source for social media analytics due to its accessibility and availability. Social media analytics involves analyzing data extracted from social media sites to generate useful information for various sectors of society, as Andryani, Negara, and Triadi (2019) point out. The business sector has primarily used social media analytics to understand the impact of their products and services on their customers. As the amount of data from social networking sites continues to grow, businesses have also realized the potential to improve their customer relations and marketing strategies by tapping into these resources (Carlson et al., 2018). With the increasing demand for social media analytics, the need for advanced data analytics techniques, such as machine learning, has risen, as noted by Alloghani et al. (2019).

According to Sandhu (2018), machine learning is a subset of artificial intelligence that employs automated methods to resolve issues based on historical data without needlessly changing the primary procedure. Creating algorithms and using other computing methods to make machines smarter is the essence of artificial intelligence. It includes algorithms that use techniques, usually beyond human grasp, to think, act, and carry out tasks. Machine learning aims to resolve issues based on historical or previous examples. This includes discovering hidden patterns in the data (data mining) and then using the practices to categorize or forecast a problem-related event. In other words, machine learning provides the information that intelligent machines need to maintain their functionality (Kadhim 2019).

Additionally, current research articles show that social media significantly impacts organizations (Bergström and Belfrage 2018). Organizations have identified data-driven methodologies as the perfect growth strategy (Salloum et al. 2017). After all, organizations such as the Roman Catholic Church can take advantage of Twitter's data to learn

people's reactions regarding social issues popping up without asking them individually. Twitter is ideal for knowing the consensus of a group of people. Organizations are allowed to see the public's sentiment (Anber, Salah, and El-Aziz 2016). The Roman Catholic Church, through Pope Francis, has maintained a presence in social media via his Twitter account, and the official Twitter account of the Pontiff has more than 19 million followers.

1.3. The Church and Social Media

On the part of the Body of Christ, where it refers to one of the images of the Church, social media has become ubiquitous and is beneficial in engaging them in evangelization, liturgical participation, and social responsibilities the Church promotes. They can express their opinions and views on how the Catholic Church addresses such social issues (Díaz 2021). Since technology is fast-changing, the Church should always pay attention to discussions taking place on social media platforms. The document entitled "CBCP Pastoral Guidelines on the Use of Social Media" released by the Catholic Bishops of the Philippines (CBCP) Episcopal Commission on the Laity exists for this reason. Social Media platforms, like Twitter, have become an integral part of modern communication and have a significant impact on public opinion and social trends.

Twitter and other social media platforms have revolutionized how people exchange information and interact. They've developed brand-new forums for discussion and debate that can sway public opinion and affect societal trends. Particularly on Twitter, people can post and debate current affairs as they happen, thanks to its quick-paced and real-time nature. The Church may better comprehend the issues and demands of its members and the larger society by participating in social media debates. It may also use these venues to spread its message and teachings, dispel rumors and untruths, and encourage harmony and understanding.

Moreover, social media sites like Twitter have a global audience. They may assist the Church in connecting with people everywhere, even those who might not have access to conventional means of communication. Pope Francis, for instance, emphasized the value of social media and digital communications in his message for World Communications Day 2021

to encourage intercultural conversation, understanding, and encounters (Pope Francis 2021). The Pontifical Council for Social Communications' document "The Church and Internet" highlights the Church's responsibility to evangelize through new media and use the Internet and social media to spread the Gospel and promote unity and understanding (Pontifical Council for Social Communications 2002). Pope Benedict XVI's message for World Communications Day 2009 emphasizes the potential of new technologies and social media to promote dialogue, understanding, and friendship among people and urges Catholics to use these tools to spread the message of the Gospel (Pope Benedict XVI 2009).

The Roman Catholic Church has not fully adapted to technologies such as Data Analytics and Machine Learning. Still, it relies on traditional media forms to obtain information and gain awareness of current social issues (Díaz 2021). But the Church has called on its members to use science and technology fully and constructively, provided it is for the common good and the inner purpose of creation (John Paul II 2002). Pope St. John Paul II also stated that using these technologies to spread the Christian Message and the Church's authentic teaching is not enough. It is essential to integrate the Church's message into the 'new culture' created by these modern communications (John Paul II 1990). The introduction of Data Analytics and Machine Learning toward knowing the consensus of the Body of Christ can significantly help the Church identify sectors that need attention (Pope Francis 2014).

1.4. Aim of the Study

While many research studies examine the use of social media data, such as Wprostkiewicz, Sosnowska, and Wójciszyn-Wasil (2022) article discussing "The Catholic influencer as a challenge for spiritual leadership in the age of social media" and Brazal's (2023) "Synodality and the New Media," there are limited discussions that specifically tackle the use of machine learning for the use of the Synod on Synodality of the Roman Catholic Church. Machine learning has increasingly become an essential tool for analyzing large amounts of data and making predictions. Still, its use in religious institutions like the Roman Catholic Church remains relatively unexplored. As the Church seeks to engage with its followers and address pressing social issues through the Synodality, there is an opportunity to

explore how machine learning can be used to understand better the needs and perspectives of the Church's members and how it can help inform decision-making and policy development.

The current study aims to provide new insights into how the Catholic Church can enhance its provision of services to its followers in the contemporary era through the integration of data analytics and machine learning methods, utilizing Twitter data as the primary source of information. Twitter was selected as the primary data source due to its unique characteristics and relevance to the research objectives. Specifically, Twitter is a real-time social media platform where users generate and interact with content, making it an advantageous source of data for studies interested in investigating how people respond to unfolding events or issues. Additionally, Twitter users discuss a diverse range of topics, rendering it a valuable source of data for research endeavors that seek to explore a wide array of issues. Moreover, the majority of content on Twitter is publicly available, facilitating the collection and analysis of substantial data volumes. Finally, Twitter provides researchers with access to a wealth of user demographic data, which can aid in comprehending the sample and controlling for certain variables in the analysis.

Utilizing social media data analysis, the Church can gain a deeper understanding of the needs and concerns of its members, which can inform its decision-making processes and pastoral efforts. This approach can be a valuable tool to support the Synod on Synodality's emphasis on collaboration, dialogue, and a synodal approach to decision-making. The data collection period spanned the second quarter of 2021 and focused on the Twitter ecosystem in the Philippines. A significant proportion of Filipinos utilize Twitter as a platform to express a diverse range of concerns and opinions relating to the Church's social issues.

To achieve the goal of the study, the following goals were determined:

- a. To describe the pressing social issues faced by the Catholic Church in the Philippines to provide supplemental information to the Synodal process;
- b. To develop a machine-learned model for the classification of Filipino social issues;

- c. To develop a web application that integrates the developed machine-learned model.

The rest of the paper is organized as follows: Section 2 describes the methodology used in the paper. Section 3 will discuss the results. Section 4 will discuss the conclusion, and section 5 introduces the limitations and future work for the topic.

2. Methodology

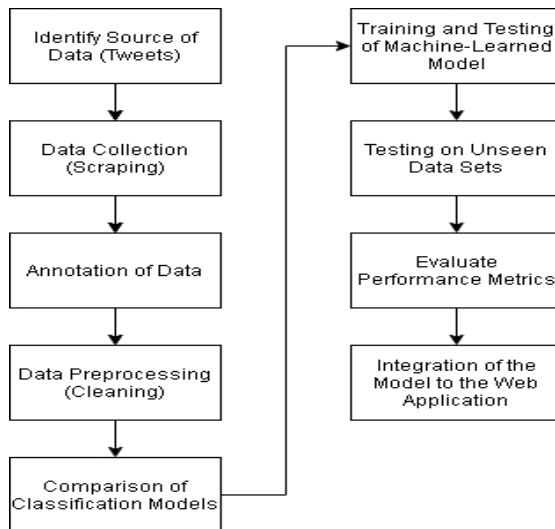


Figure 1: Framework of the Web Application

To achieve the aforementioned research goals, the study will need to start with identifying the tweets that will be used as the basis of the machine learning to identify and classify new Tweets into different topics. Therefore, the study has identified three criteria for the selection of Tweets for the training dataset of the machine:

1. Tweets that are written by reputable sources
2. Tweets that are relevant to the Philippines
3. Tweets are readily available

The study has identified two reputable sources of Tweets that will be fed into our machine-learning algorithms: GMA News and CNN Philippines. The Tweets that come from these sources will be utilized mainly because of their credibility and experience in writing trustworthy news and Tweets. GMA News and CNN Philippines have different journalists and fact-checkers available to verify the credibility of the news that they are about to Tweet, thus eliminating the doubt that the Tweets might not be credible, which can severely affect how the machine learning algorithms classify the Tweets that we will feed it later on (Li, Zhan, and Li 2018).

The research primarily selected secular news sources to guarantee the credibility and impartiality of the data. The study's objective is to leverage machine-learning algorithms for tweet classification, thus necessitating the use of unbiased and dependable information sources. Although Catholic news sources may offer valuable insights and viewpoints, they could also be influenced by religious biases and interests. Conversely, secular news sources are known for keeping their best at dedicating and providing factual and unbiased reporting. However, it is impossible to entirely eliminate bias as it is inherent in human interpretation and presentation of information. Therefore, this study aimed to employ credible and dependable news sources in the Philippines recognized for their commitment to unbiased and factual reporting. GMA News and CNN Philippines are generally considered to have a lower degree of bias in the Philippines (PUBLiCUS 2022).

The study does not assume that these sources are entirely unbiased but recognizes that they are known for their impartial and factual reporting, which will significantly reduce bias impact and enhance the accuracy of the algorithms. To mitigate bias, the study employed two data sources instead of one, although it acknowledges the need for more data sources that represent diverse perspectives on various categories. It is crucial to note that the selection of news sources was not influenced by any ideological leanings but rather their reputation for accurate and factual reporting.

Before being fed into a machine for text categorization, tweets should be reliable and credible since the accuracy of the machine learning

algorithm depends on the quality of the data it is trained on. The algorithm may learn to make erroneous predictions or classifications based on erroneous information if the tweets being fed into it are untrustworthy, meaning they contain fake information, rumors, or misleading claims (Dreisbach et al. 2019). This is crucial when actions based on incorrect forecasts might have serious repercussions, including news analysis.

After which, data collection will follow, followed by annotation and data preprocessing. After the data is cleaned, it will be fed into several classification models to identify the best model to use. After placing the model used, the data will undergo training and testing of the Machine-Learned model. Rigorous testing on data sets not present during the previous iterations will be undertaken, then finally, the evaluation of the performance of the said model will be done. Lastly, this information will then be passed onto the web application to show the product of the process. The study will utilize data mining techniques and text classification using Natural Language Processing (NLP).

2.1. Tools and Technologies

Before delving into the program's technicalities, it is crucial to discuss the tools and technologies used. The data used in the program was collected from Twitter using Twint, an advanced Twitter scraping tool in Python that enables scraping tweets without using Twitter's API. Twint was preferred over other data scraping tools due to its unlimited scraping capability, which is advantageous for cost-effective data collection in large-scale projects. Twint uses Twitter's search operators to scrape tweets from specific users or those related to particular topics, hashtags, and trends. The program's ability to extract sensitive information such as email and phone numbers from tweets makes it an excellent data processing tool.

Python Programming Language was used to create the machine learning models as Python is one of the best languages for developing a machine learning model (Raschka 2015). The model was then integrated and turned into a Web Application prototype, which can be seen at the end of the paper. Other open-source tools and technologies for Python

that was also utilized can be seen in the table below.

Table 1: Tools and Technologies

Technology	Description
Scikit-learn	An open-source Python library for machine learning containing features such as regression, clustering and classification. Scikit-learn has split the data set into training and testing sets, logistic regression, count vectorizer, and TF-IDF.
NLTK library	A natural language processing tool for Python. This library was used for the tokenizer of words, and removing of stop words.

2.2. Data Collection

The study collected 12,000 tweets from chosen trustworthy news sites, GMA News and CNN Philippines, on the 13th of April, 2021. These collected tweets were only scraped via credible news outlets to guarantee that the data collected is trustworthy and is not tainted with noisy data. The tweets collected serve as the primary foundation for the machine-learning process and guide the device in predicting and categorizing the tweets for our web application. However, there are inherent risks when collecting tweets based on hashtags to gain insights into a specific topic or event.

One potential risk is that certain hashtags may be associated with inaccurate or misleading information, especially if promoted by individuals or organizations with particular agendas. Such hashtags could disseminate false information, conspiracy theories, or hate speech, leading to the inadvertent acquisition and use of erroneous or damaging information.

Another risk is that hashtags may be taken over or controlled by bots or trolls who use them to disseminate false information or highlight specific points of view. Because of this, it can be challenging to categorize tweets appropriately based on their content, resulting in a misleading image of the debate surrounding a particular issue. Although news organizations are often held to higher standards of accuracy and trustworthiness, collecting tweets from news organizations can offer a more trustworthy source of information.

2.3. Annotation

The data analysis process involved the individual examination of each dataset with the help of some volunteers, who then added tags (i.e., categories) to each item to facilitate classification. To ensure the accuracy and reliability of the annotations, the volunteers also compared the annotations made by volunteers and deliberated any discrepancies to determine the appropriate category assignment for a consistent outcome. The resulting annotated dataset was subsequently utilized for both model training and testing.

2.4. Preprocessing

Extracting the features from the documents was essential in training the machine learning models. Each unique word in the dataset is considered a feature when using the “bag of words” approach in data preprocessing. In this way, the number of columns within the data frame has the exact dimensions as the number of unique words in the data set. Count vectorization is applied to the processed data wherein the number of occurrences of the word in a text is assigned to its respective feature. As a result, count vectorization cannot represent the specific ordering or combination of words.

The term frequency-inverse document frequency (TF-IDF) approach is used in text mining, where each word is quantified and a weight that determines the importance of a comment in a text is computed (Scott 2019). It combines two metrics: Term Frequency and Inverse Document Frequency. Term frequency is computed by dividing the number of occurrences of that specific word over the total number of words within a text or document. On the other hand, inverse document frequency helps in identifying the most important and relevant terms within a document or corpus of documents. Words that are used frequently across many documents are considered to have a lower relevance than words that occur rarely but are specific to a particular document. The resulting TF-IDF score for each term can be used to rank and retrieve documents based on their relevance to a given search query.

For the first part of the preprocessing, a standard cleaning of data was performed, such as removing duplicate data, noisy data, and anomalous data that were unintentionally scraped. Noisy data are tweets that are

meaningless. An example of noisy data is the tweet “Read more: <https://t.co/VNqHsw7FQ8>”. Anomalous data are outliers in the dataset. An example of anomalous data from the tweets is “*Nagtira pa ng 8.97 pesos.* <https://t.co/HL2tKljIiW>” (They left 8.97 pesos). The tweets were also cleaned by removing several stop words, special characters, and punctuation marks. Stop words are the commonly used words in a language that does not affect the meaning of a sentence even if it is removed such as, “is,” “are,” “the,” “were,” and “at.” (Teja 2020). Additionally, the tweets will also be made into lowercase and lemmatization of the words will be implemented. Lemmatization is a process of extracting a root word by considering the vocabulary. For example, “good,” “better”, or “best” is lemmatized (changed) into “good.” The stop words are specified by a library called NLTK, which was discussed earlier in the tools and technologies used.

To illustrate, a sample tweet extracted from the dataset reads, “A group of health professionals is demanding urgent changes to the government’s COVID-19 response to prevent a ‘vicious cycle’ of surge in infections.” After subjecting this text to data cleaning techniques, the resulting dataset would contain the following cleaned data: “group health professionals demanding urgent changes government covid response prevent vicious cycle surge infections.” Notably, stop words like “of” and “the” were removed, punctuations were omitted, and capitalized words such as “COVID” were converted to lowercase.

2.5. Descriptive Analytics

The descriptive analytics used in the data collected by the group included the most frequently used words for each category. With this, text classification was utilized. Text classification is an iterative process or perhaps a linear process. The statistical patterns and trends identified during the predictive model training phases should be applied to data preprocessing and, conversely, the predictive model to determine optimal parameters and anticipate high accuracy.

2.6. Data Analysis

The tweets were analyzed using a data classification technique

which will group the data into several categories. These categories were various topics that are commonly talked about on social media platforms. The project team used several classification algorithms, such as Stochastic Gradient Classifier, Support Vector Machine (SVM), and Logistic Regression.

In the initial phase, the first dataset was partitioned into two subsets: a training and testing datasets, with 80% and 20% of the data, respectively. This partitioning aimed to determine the most suitable algorithm for training the machine. Subsequently, iterative machine learning was carried out in the study, where new and previously unseen tweets were obtained and inputted into three algorithms for classification. The algorithms' performance metrics, including accuracy and confusion matrix, were evaluated through multiple iterations to enhance their performance. Following this, the project team compared the accuracies and the outcomes of the confusion matrices of the algorithms to identify the optimal algorithm for classification in the web application.

2.7. Web Application

After the data is analyzed and categorized, it is displayed in a web application, primarily built on Django, a Python framework. The plan in creating the web application was to see the trending category, and the user will be able to see the tweets in the category that will be selected. The user can also input a number of how many tweets the web application will scrape.

3. Results and Discussion

3.1. Categories

The data collected from Twitter were analyzed and manually classified into several categories. The group did the categorization manually, as it would be used to train the model for organizing unseen data. The categories determined by the group were based on the news categories that were identified by Jahara, Sharif, and Hoque (2022), and HuffPost (2014), an online news aggregator website. These categories can be seen below:

Table 2: Categories

Categories
Crimes, Accidents, and Human Affairs
Health and Wellness
Politics
Business, Finance, Economy & Tourism
Travel and Transportation
Religion
Education and Literacy
Environment
Others

The categories used in this study were selected based on common problems encountered and discussed in the country. Originally, there were nineteen categories, but the research team narrowed them down to nine main categories considered the most relevant for ranking purposes. Unfortunately, the distribution of tweets across the different categories was highly uneven, with the most populated categories being politics and health. In contrast, the travel and transportation category had the least number of tweets. These findings were based on data gathered during the second quarter of 2021.

Moreover, the Religion category was one of the least discussed topics in Twitter based on the tweets that were collected. This may be attributed to the fact that societal, political, and health issues dominated the public discourse during data collection, while discussions about the Church and religion were relatively scarce. This is consistent with previous researches (Woodward and Kimmons 2018; Hosseinzadeh 2011; Cheong 2016), which found that individuals are often reluctant to discuss religious topics on social media due to concerns about causing tension or arguments (Barna Group 2016). Furthermore, some people may not feel knowledgeable enough to discuss religious topics or may not be interested in such discussions.

The reason behind this is that according to “Why People Are Reluctant to Discuss Faith” by the Barna Group (2016), people using Twitter may be afraid of talking about their faith in Twitter for mainly two reasons: avoidance and ambivalence. Avoidance because majority of the respondents in the said study said that “Religious conversations always seem to create tension or arguments”. Ambivalence because they either “do not know enough to talk about religion” or “I am not religious, and don’t care about these kinds of topics”. In the tweets that the study reviewed during the data cleaning phase, it has been found that tweets that discuss religion, particularly the Catholic Church, are some of the places where people usually argue.

3.2. Training and Testing

In order to prepare for the next stage of the process, the first set of tweets was utilized to train a machine learning model. The tweets were sourced from various verified accounts, manually categorized, and limited to English only. Any Tagalog tweets were translated to ensure consistency. To improve the accuracy of the model and reduce noise, several preprocessing steps were taken on the tweets. This involved removing punctuation, converting text to lowercase, and eliminating stop words using the Natural Language Toolkit (NLTK). The study also employed lemmatization to extract root words based on vocabulary, after which the preprocessed tweets were vectorized using the TF-IDF (Term Frequency-Inverse Document Frequency) technique.

During the dataset annotation process, the study initially relied on manual annotation, but it was discovered that some items were not correctly annotated, which led to reduced accuracy of the results. To address this issue, the study reduced the number of categories in the dataset. Categories such as “fashion and beauty” and “entertainment” were deemed irrelevant and removed. Some categories were merged because of their similarities, such as merging “travel” with “transportation,” “human affairs” with “crimes and accidents,” and “tourism” with “business, finance, and economy.”

Lastly, the text classification framework was trained into several classifiers with several features created previously. The study’s approach

was to explore traditional machine learning models first such as SVM, Stochastic Gradient Classifier, and the Logistic Regression. The training and testing had been split into 80% training and 20% testing.

3.3. Model Metrics

A preprocessed data set was used to train the models in the first iteration of machine learning. The models were then tested and results of its performance metrics (accuracy and confusion matrix) were matched with the expected output. If the output did not meet the desired values, then the models were then fed new, unseen data to further improve and refine the results of its performance metrics.

These models went through a word embedding layer. Word embedding was described as a form of representing words and documents through dense vector representation. These were the different tools that are needed to build the embedding:

1. Use tokenizer methods from the TF-IDF vectorizer,
2. Make a vocabulary of words,
3. Make the text into a sequence to convert the words to numbers,
4. Make fixed-length sequences.

	model	accuracy
0	SGD	68.56
1	SVC	66.24
2	Logistic Regression	86.44

Figure 2: Accuracy


```

Confusion matrix :
[[458 39 29 31 37 31 28 26 13]
 [ 28 846 62 68 57 54 49 64 26]
 [ 32 44 834 76 45 53 54 62 26]
 [ 33 59 65 868 68 57 47 64 27]
 [ 29 56 57 60 851 46 57 64 22]
 [ 31 55 52 47 49 844 43 58 15]
 [ 31 58 53 50 56 42 820 50 24]
 [ 35 70 52 66 58 52 53 902 23]
 [ 15 27 22 22 21 15 33 21 433]]
Outcome values :
458 0 0 0
Classification report :
              precision    recall  f1-score   support

     1           0.66       0.66       0.66         692
     0           0.00       0.00       0.00           0

   micro avg       0.66       0.66       0.66         692
   macro avg       0.33       0.33       0.33         692
weighted avg       0.66       0.66       0.66         692

```

Figure 3: Confusion Matrix for SGD

```

Confusion matrix :
[[[417 49 27 31 38 41 32 46 11]
 [ 49 770 45 83 70 57 59 91 30]
 [ 36 64 749 75 57 67 66 81 31]
 [ 37 74 89 778 81 62 65 72 30]
 [ 32 69 73 77 777 61 59 63 31]
 [ 35 59 74 55 62 763 60 63 23]
 [ 28 53 68 70 63 64 739 62 37]
 [ 47 86 72 75 67 60 68 811 25]
 [ 11 30 29 44 27 19 36 22 391]]]
Outcome values :
417 0 0 0
Classification report :
              precision    recall  f1-score   support

     1           0.60       0.60       0.60         692
     0           0.00       0.00       0.00           0

   micro avg       0.60       0.60       0.60         692
   macro avg       0.30       0.30       0.30         692
weighted avg       0.60       0.60       0.60         692

```

Figure 4: Confusion Matrix for SVN

```

Confusion matrix :
[[[ 598 13 12 11 11 11 8 23 5]
 [ 13 1081 15 39 24 27 21 25 9]
 [ 12 15 1069 32 23 19 23 20 13]
 [ 11 39 32 1099 20 21 25 31 10]
 [ 11 24 23 20 1074 28 22 32 8]
 [ 11 27 19 21 28 1033 21 24 10]
 [ 8 21 23 25 22 21 1030 22 12]
 [ 23 25 20 31 32 24 22 1126 8]
 [ 5 9 13 10 8 10 12 8 534]]]
Outcome values :
598 0 0 0
Classification report :
              precision    recall  f1-score   support

     1           0.86       0.86       0.86         692
     0           0.00       0.00       0.00           0

   micro avg       0.86       0.86       0.86         692
   macro avg       0.43       0.43       0.43         692
weighted avg       0.86       0.86       0.86         692

```

Figure 5: Confusion Matrix for Logistic Regression

Based on the figures above, the study had decided that Logistic Regression will be used for the project. After which, the study had then proceeded to produce the data, starting with the top five most frequent words used per category. The top three categories are presented below with their respective top 5 words based on their frequency being present in tweets.

The study initially had 19 categories for classification but due to the limited number of datasets, the accuracy level was low. To improve the accuracy, the study decided to reduce the number of categories to 9 and increase the number of datasets for training. The volunteers manually classified 12,000 tweets, but after data cleaning, only 10,000 tweets were left to train the model.

Prior to training the model, the collected tweets underwent a rigorous process of data cleaning to remove any irrelevant or noisy data that could adversely affect the model's accuracy. The data cleaning process involved several steps, including the elimination of duplicate tweets and tweets written in languages other than English. Additionally, tweets that did not fall within any of the predefined categories were removed, and any unnecessary words or punctuations were eliminated while spelling errors were corrected. These measures ensured that only relevant and high-quality data were utilized in the model training phase, thereby contributing to more precise outcomes.

Subsequently, the study developed a prototype web application capable of categorizing and classifying tweets according to the designated categories. The web application demonstrated commendable capabilities in terms of classification and categorization. However, the paper did not present the results of preliminary experiments that were conducted.

3.4. Top Three Topics

Table 3: Top Five Terms for Crimes, Accidents, and Human Affairs

Terms	Frequency
kill	124.576891
police	110.473147
city	112.339649
suspect	104.761438
dead	100.013144

Table 4: Top Five Terms for Health and Wellness

Terms	Frequency
covid	265.567243
vaccine	143.095385
health	141.673592
Philippines	134.100935
DOH	132.803924

Table 5: Top Five Terms for Politics

Terms	Frequency
Duterte	225.870495
sea	123.889451
China	113.204521
Philippine	111.421589
roque	111.067942

Based on the results obtained by the study last 2021, the issues with the most frequent tweets were about Health and Wellness, Crimes, Accidents, and Human Affairs, and Politics. In the category of Health and Wellness, the terms that came up frequently were “COVID” and “vaccination.” Given the current situation regarding the global pandemic crisis, it was evident that these terms were frequently posted on Twitter by the news platforms the study collected the data from. 2021 was a year when vaccines were starting to roll out in the Philippines and initially, there were fears among the faithful of the side-effects that it brings because of the apparent “fast” and “experimental” the development of these vaccines underwent (Quibranza III 2021). There was also opposition from other religious communities because getting vaccinated “shows a lack of faith,” which, unfortunately, hid their vaccine hesitancy behind religious reasons (Quibranza III 2021).

3.5. Health and Wellness

The Catholic Bishops’ Conference of the Philippines (CBCP) has expressed their willingness to be vaccinated against COVID-19 to encourage people to get inoculated (Patinio 2021). The CBCP has also offered church facilities as venues for vaccination activities and partnered with local government units to allay people’s fears of vaccines. The government has commented that the CBCP’s sphere of influence could help in the vaccine roll-out (Parrocha 2021). However,

the CBCP's Bioethics chairman Ricardo Baccay has also stated in a pastoral letter that some COVID vaccines were manufactured using the remains of an aborted fetus, and people should be free to decide to be vaccinated or not according to their conscience (Aquino 2021).

In nations such as the Philippines, the utilization of vaccines derived from aborted fetuses has become an inevitable circumstance. In light of this, the Vatican has issued a statement acknowledging that in certain countries, access to ethically uncontroversial Covid-19 vaccines may prove challenging due to factors such as limited availability, inadequate storage and transportation, or a lack of options for citizens to select their preferred vaccine. Given these circumstances, the Catholic Church recognizes the moral acceptability of receiving Covid-19 vaccines that employed cell lines from aborted fetuses during their research and production stages, provided that no other viable options exist. This decision was grounded on the principle of the common good, which places a premium on safeguarding oneself and others from the spread of the virus (Aquino 2021).

The Pontifical Academy of Life has already been advocating the use of vaccines even before the COVID Pandemic started, stating in their document entitled "Clarifications on the Medical and Scientific Nature of Vaccination" dated July 31, 2017, that vaccines do not come from freshly voluntary aborted fetuses nowadays because the vaccines that are being developed are far from the cell lines that were originally aborted back in the 1960s. Therefore, it no longer implies a "bond of moral cooperation indispensable for an ethically negative evaluation of their use" (Pontifical Academy for Life - National Office for Health Pastoral Care (CEI) - Association of Italian Catholic Doctors 2017). They went on with their argument saying that vaccination is a very urgent concern to safeguard the safety of others especially toward immunodeficient and pregnant women.

3.6. Crimes, Accidents and Human Affairs

The category of Crimes, Accidents, and Human Affairs has been the topic of discussion in the twitter verse in 2021 with the following keywords of "kill", "police", "suspect", "dead", and "city". Ever since

Duterte stepped into office last 2016, the amount of people being killed was growing and being labeled as extrajudicial killings. Last 2021, the issue of police killings was once again brought up after a drunk police officer killed a 52-year-old mother, which led to several reforms in the Philippine National Police because of the swift public outcry (Aspinwall 2021).

Regrettably, the Duterte Government has been criticizing the Philippine Roman Catholic Church because of its stance against drug users and pushers since these both organizations have contradicting ideas on how drug abuse can be stopped in the Philippines (Willis 2019). The Catholic Church is one of the most prominent voices as time and time again, the CBCP has been voicing out the extrajudicial killings that were happening ever since former President Duterte has been voted into office. A pastoral letter by the archbishops of Nueva Segovia, Lingayen-Dagupan, and Tuguegarao in northern Luzon has urged its faithful to resist the “culture of murder and plunder” (CNA 2021). The pastoral letter has condemned the killings of journalists, political opposition members, lawyers, and priests. The Church has been attempting to propose a non-violent solution to the helpless killings by initiating peaceful assemblies wherein the issues of the Filipino people, via sober discussion that the Gospel guides must be the path that must be chosen.

The topics identified by the study, namely human affairs, crime, accidents, politics, and health and wellness, are intricately linked and exert a significant influence on individuals and communities. These matters are particularly relevant to the Catholic Church’s pursuit of a synodal mode of being. Synodality seeks to address these challenges in a holistic and comprehensive manner by fostering participation, dialogue, and collaboration, recognizing their interdependence and integration into a broader system. This approach is of critical importance in the Philippines, where these issues are particularly acute and exert a significant impact on the lives of millions of individuals.

The spirit of synodality upholds the inherent dignity and freedom of each individual, acknowledging their intrinsic worth regardless of their social standing, background, or situation. The ongoing COVID-19 pandemic underscores the need for the Catholic Church to deliberate

on its stance concerning vaccination and public health matters, a topic that can be fruitfully examined in the Synod of Synodality, facilitating a more sophisticated and informed approach. In the healthcare domain, a synodal process prioritizes patients' well-being and dignity above business or political considerations, enabling them to participate actively in treatment decisions, securing access to high-quality healthcare services, and promoting preventative health measures conducive to a meaningful existence.

Embracing synodality as a guiding principle can likewise enhance the Catholic Church's response to human concerns such as crime, accidents, and other issues. By engaging diverse perspectives and expertise, including those of laypersons and those directly impacted by these challenges, the Church can formulate more efficacious and empathetic solutions. A synodal approach places a premium on the common good over partisan interests and accords equal weight to the distinct contributions and viewpoints of all segments of society. This paradigm may entail advocating for legislation that fosters social equity and parity, collaborating constructively with politicians from divergent ideological persuasions, and empowering marginalized groups to participate actively in decision-making procedures.

3.7. Politics

The Politics category revealed that the terms "duterte," "government," and "sea" frequently appeared in the collected tweets, with most issues relating to the government's handling of the West Philippine Sea dispute and the previous administration's governance. It's worth noting that the Philippines won its case against Mainland China in the Permanent Court of Arbitration in 2016. This territorial dispute was also significant to the Philippine Catholic Church. The Catholic Church's advocacy for peace in territorial disputes is rooted in its concern for the welfare of people who depend on the sea for their livelihood, and because of how important the sea is to us, Sea Sunday is celebrated every second of July by many Catholic Churches as an "International Day of Remembrance, prayer and celebration, and an opportunity to thank seafarers and fishermen who work tirelessly throughout the year" (Gorecho 2021).

Bishop Broderick Pabillo, the CBCP's President, has been critical

of the government's treatment of Filipino fishermen harassed in the West Philippine Sea. In line with this, Pope Francis also acknowledged the vital role of seafarers and fishermen, stating that "without the people of the sea, many in the world would starve." The Church advocates for peaceful dialogue between the Philippines and China to resolve this long-standing conflict (Torres and Saludes 2016). The Church may actively contribute to advancing social justice and the common good by encouraging communication and cooperation among various political actors and interaction with the larger society.

Based on the presented findings and discussions, it is recommended that the Catholic Church accord priority to engaging with its congregation through social media and actively solicit their feedback and insights on pertinent issues. Social media platforms can serve as a valuable tool for gathering data, fostering collaboration, and augmenting communication among the laity. It is important to note, however, that the Synod on Synodality did not explicitly address the role of social media in information gathering and analysis. To realize a synodal decision-making process that values cooperative and inclusive deliberations among all members of the Church, it is proposed that the Church ensure the transparency, responsiveness, and inclusivity of its decision-making procedures. By embracing this approach, the Church can prioritize the active involvement and engagement of all members of its community in the decision-making process.

4. Conclusion

The innovative use of data analytics and machine learning in this study provides valuable insights into the social issues being discussed on Twitter, which can supplement the Catholic Church's traditional means of data collection. By leveraging technology and data analysis tools, the Church can better understand the needs and concerns of the people it serves and prioritize its efforts accordingly. This study highlights the importance of the Catholic Church addressing social issues such as health, politics, and human affairs. The Church's mission is not limited to faith and worship but extends to the broader social and political sphere. By prioritizing these issues, the Church can better serve the needs of the

people and work towards the common good. In addition to serving as a supplement to traditional data collection methods, the findings of this study can also inform the Church's approach to synodality. Synodality is a matter of ecclesial governance and a way of living out the Church's mission in the world. A synodal approach that prioritizes the dignity and freedom of every human person can help the Church to fulfill its mission of eliminating evil and achieving the Kingdom of God here on Earth.

Furthermore, the Synod of Synodality provides a platform for the Church to engage with the wider community and address these issues holistically and integrated manner. Through the synodal process, the Church can listen to the voices of the people and work with them towards common goals. This can lead to greater unity and collaboration and a more effective and efficient use of resources. In conclusion, the use of data analytics and machine learning in this study is a promising development that can help the Church better serve the people's needs. By prioritizing social issues, embracing a synodal approach, and working towards the common good, the Church can continue to fulfill its mission of bringing the Kingdom of God to Earth.

While the Twitter Data Classifier and similar tools can provide valuable insights into the opinions and concerns of Church members on social media, they should be viewed as a supplemental tool to the Synod on Synodality. The Synod emphasizes the importance of fostering greater collaboration and dialogue among all members of the Church, including the laity, clergy, and hierarchy, in order to ensure that decision-making processes are transparent, participatory, and reflective of the needs and concerns of all members.

5. Limitations and Future Work

The primary focus of the study on social issues as presented on Twitter may constrain its ability to fully capture the "sensus fidelium" of the Roman Catholic Church. The term "sensus fidelium" refers to the collective voice of the faithful, which conveys the Church's beliefs

and teachings. While social issues hold significance, the Church is also concerned with other internal matters, including but not limited to women's ordination, faith, church leadership, and other pertinent topics. Hence, it is recommended that future studies broaden their data sources and subject matter to provide a more comprehensive comprehension of the "sensus fidelium."

In addition to the limitations of the study's focus on social issues presented on Twitter, there is a potential oversight of important topics that are not as popular or frequently discussed on social media. The study's focus on "trending" topics on Twitter may exclude critical issues that are not receiving as much attention, such as ecological concerns, which may not garner as much attention on Twitter compared to other social issues. It is essential to note that the Roman Catholic Church has consistently expressed its concern about ecological issues, such as climate change and environmental degradation, through official statements and encyclicals. Thus, future studies that explore the "sensus fidelium" of the Church should consider a broader range of issues, including those that may not be as popular or trending on social media. To address this limitation, future studies could explore ways to identify and incorporate a more diverse set of topics relevant to the Church's teachings and beliefs. For example, they could use a variety of data sources beyond social media, such as official Church documents, surveys, and interviews with the faithful. By including a more diverse set of issues and data sources, studies can provide a more comprehensive understanding of the "sensus fidelium" of the Roman Catholic Church.

Furthermore, while Twitter may be an important source of information and preoccupation for the Church, relying solely on tweets may not provide a complete picture of the issues that matter most to the Church's members. The use of social media as a data source has limitations, such as the exclusion of groups that are not active on these platforms. Therefore, future research could explore how social media can be utilized alongside other sources of information and perspectives to create a more synodal Church. For instance, combining data from official Church documents, surveys, interviews, and social media could provide a more nuanced understanding of the collective voice of the faithful. Moreover, understanding the limitations and strengths of

different data sources can help guide the Church's decision-making. For example, social media can provide real-time insights into the concerns and opinions of the faithful, allowing the Church to address emerging issues promptly. However, it is also essential to consider the broader context, such as the historical and theological perspectives, when making decisions that impact the Church and its members.

Lastly, as the web app is still in its development and enhancement stage, future work could focus on improving its functionality and usability, as well as exploring ways to integrate it with other tools and platforms that the Church uses to engage with its members and the wider community. Doing so will enable the Church to better serve its members and engage with them on important issues.

REFERENCES

- Alloghani, Mohamed, Dhiya Al-Jumeily, Jamila Mustafina, Abir Hussain, and Ahmed J. Aljaaf. "A Systematic Review on Supervised and Unsupervised Machine Learning Algorithms for Data Science." *Unsupervised and Semi-Supervised Learning*, September 5, 2019, 3–21. https://doi.org/10.1007/978-3-030-22475-2_1.
- Anber, Hana, Akram Salah, and A. A. Abd El-Aziz. "A Literature Review on Twitter Data Analysis." *International Journal of Computer and Electrical Engineering* 8, no. 3 (2016): 241–49. <https://doi.org/10.17706/ijcee.2016.8.3.241-249>.
- Andryani, Ria, Edi Surya Negara, and Dendi Triadi. "Social Media Analytics: Data Utilization of Social Media for Research." *Journal of Information Systems and Informatics* 1, no. 2 (September 2, 2019): 193–205. <https://doi.org/10.33557/journalisi.v1i2.23>.
- Aquino, Leslie Ann. "CBCP Urges Filipinos to Get Immunized against COVID-19." *Manila Bulletin*, January 14, 2021. <https://mb.com.ph/2021/01/14/cbc-urges-filipinos-to-get-immunized-against-covid-19/>.
- Arigo, Danielle, Sherry Pagoto, Lisa Carter-Harris, Sarah E Lillie, and Camille Nebeker. "Using Social Media for Health Research: Methodological and Ethical Considerations for Recruitment and Intervention Delivery." *Digital Health* 4 (January 2018): 205520761877175. <https://doi.org/10.1177/2055207618771757>.
- Aspinwall, Nick. "Police Killing Leads to Calls for Reform in the Philippines." *The Diplomat*, June 7, 2021. <https://thediplomat.com/2021/06/police-killing-leads-to-calls-for-reform-in-the-philippines/>.
- Banerjee, Amarabha. "Iterative Machine Learning: A Step towards Model Accuracy." *Packt Hub*, December 1, 2017. <https://hub.packtpub.com/iterative-machine-learning-step-towards-model-accuracy>.
- Barna Group. "Why People Are Reluctant to Discuss Faith." *Barna Group*, 2016. <https://www.barna.com/research/reasons-for-reluctance/>.
- Bergström, Annika, and Maria Jervelycke Belfrage. "News in Social Media." *Digital Journalism* 6, no. 5 (January 12, 2018): 583–98. <https://doi.org/10.1080/21670811.2018.1423625>.
- Brazal, Agnes M. "Synodality and the New Media." *Theological Studies* 84, no. 1 (March 2023): 95–109. <https://doi.org/10.1177/00405639221150888>.
- Carlson, Jamie, Mohammad Rahman, Ranjit Voola, and Natalie De Vries. "Customer Engagement Behaviours in Social Media: Capturing Innovation Opportunities." *Journal of Services Marketing* 32, no. 1 (February 12, 2018): 83–94. <https://doi.org/10.1108/jsm-02-2017-0059>.
- CBCP News. "CBCP: Synod Must Discern 'Signs of Our Times.'" *CBCP News*, October 15, 2021. <https://cbcnews.net/cbcnews/cbc-synod-must-discern-signs-of-our-times/>.
- Cheong, Pauline Hope. "The Vitality of New Media and Religion: Communicative Perspectives, Practices, and Changing Authority in Spiritual Organization."

- New Media & Society* 19, no. 1 (July 9, 2016): 25–33. <https://doi.org/10.1177/1461444816649913>.
- CNA. “Catholic Bishops Condemn Country’s ‘Murderous and Corrupt Public Order.’” Catholic News Agency, September 2021. <https://www.catholicnewsagency.com/news/248965/filipino-catholic-bishops-condemn-countrys-murderous-and-corrupt-public-order>.
- Díaz, Israel. “Considering the Efficacy of Digital Technology as a Means of Evangelization in Christian Religious Education.” *Religious Education* 116, no. 1 (January 1, 2021): 3–15. <https://doi.org/10.1080/00344087.2021.1872001>.
- Dreisbach, Caitlin, Theresa A. Koleck, Philip E. Bourne, and Suzanne Bakken. “A Systematic Review of Natural Language Processing and Text Mining of Symptoms from Electronic Patient-Authored Text Data.” *International Journal of Medical Informatics* 125 (2019): 37–46. <https://doi.org/10.1016/j.ijmedinf.2019.02.008>.
- Gorecho, Dennis. “‘Sea Sunday,’ Fishermen, and the West Philippine Sea.” Catholic News Philippines, July 13, 2021. <https://philippines.licas.news/2021/07/13/sea-sunday-fishermen-and-the-west-philippine-sea/>.
- Hosseinzadeh, Afsaneh. “The Role of Social Media in Religion: Dialogues or Conversations.” MA in communications sciences, 2011. http://www.academia.edu/14382725/The_Role_of_Social_Media_in_Religion_Dialogues_or_conversations.
- Hudson, Matthew. “What Is Social Media?” The Balance Small Business, 2020. <https://www.thebalancesmb.com/what-is-social-media-2890301>.
- Huffpost.com. “HuffPost - Breaking News, U.S. and World News,” 2014. <https://www.huffpost.com/>.
- Jahara, Fatima, Omar Sharif, and Mohammed Moshiul Hoque. “Automatic Categorization of News Articles and Headlines Using Multi-Layer Perceptron.” *Intelligent Computing & Optimization*, 2022, 155–66. https://doi.org/10.1007/978-3-030-93247-3_16.
- John Paul II. “Redemptoris Missio.” Vatican Website, December 7, 1990. https://www.vatican.va/content/john-paul-ii/en/encyclicals/documents/hf_jp-ii_enc_07121990_redemptoris-missio.html.
- Kadhim, Ammar Ismael. “Survey on Supervised Machine Learning Techniques for Automatic Text Classification.” *Artificial Intelligence Review* (January 19, 2019). <https://doi.org/10.1007/s10462-018-09677-1>.
- Parrocha, Azer. “CBCP ‘Sphere of Influence’ to Help Allay Vaccine Fears: Panelo.” Philippine News Agency, January 31, 2021. <https://www.pna.gov.ph/articles/1129115>.
- Patinio, Ferdinand. “CBCP Leaders Willing to Be Vaccinated vs. Covid-19.” Philippine News Agency, January 28, 2021. <https://www.pna.gov.ph/articles/1128886>.
- Pontifical Academy for Life - National Office for Health Pastoral Care (CEI) - Association of Italian Catholic Doctors. “Note on Italian Vaccine Issue.” Vatican Academy for Life, July 31, 2017. <https://www.academyforlife.va/content/pav/en/the-academy/activity-academy/note-vaccini.html>.

- Pontifical Council for Social Communications. "The Church and Internet." Vatican Website, 2002. http://www.vatican.va/roman_curia/pontifical_councils/pccs/documents/rc_pc_pccs_doc_20020228_church-internet_en.html.
- Pope Benedict XVI. "43rd World Communications Day, 2009 - New Technologies, New Relationships. Promoting a Culture of Respect, Dialogue and Friendship." Vatican Website, 2009. https://www.vatican.va/content/benedict-xvi/en/messages/communications/documents/hf_ben-xvi_mes_20090124_43rd-world-communications-day.html.
- Pope Francis. "Ceremony Commemorating the 50th Anniversary of the Institution of the Synod of Bishops." Vatican Website, October 17, 2015. https://www.vatican.va/content/francesco/en/speeches/2015/october/documents/papa-francesco_20151017_50-anniversario-sinodo.html.
- . "LV World Communications Day, 2021 - 'Come and See' (Jn 1:46). Communicating by Encountering People Where and as They Are." Vatican Website, January 23, 2021. https://www.vatican.va/content/francesco/en/messages/communications/documents/papa-francesco_20210123_messaggio-comunicazioni-sociali.html.
- . "XLVIII World Communications Day, 2014 - Communications at the Service of an Authentic Culture of Encounter." Vatican Website, June 1, 2014. https://www.vatican.va/content/francesco/en/messages/communications/documents/papa-francesco_20140124_messaggio-comunicazioni-sociali.html.
- Pope John Paul II. "Common Declaration on Environmental Ethics Signed by the Holy Father and the Ecumenical Patriarch His Holiness Bartholomew." Vatican Website, June 10, 2002. https://www.vatican.va/content/john-paul-ii/en/speeches/2002/june/documents/hf_jp-ii_spe_20020610_venice-declaration.html.
- PUBLiCUS. 2022. "Executive Summary 2022 Pahayag Second Quarter Survey." PUBLiCUS Asia, Inc. June 28, 2022. <https://www.publicusasia.com/executive-summary-2022-pahayag-q2/>.
- Quibranza III, Luis A. "Why Are Some People Afraid to Get Vaccinated?" Sunstar, August 30, 2021. <https://www.sunstar.com.ph/article/1905672/cebu/lifestyle/why-are-some-people-afraid-to-get-vaccinated>.
- Raschka, Sebastian, and Randal S. Olson. *Python Machine Learning: Unlock Deeper Insights into Machine Learning with This Vital Guide to Cutting-Edge Predictive Analytics*. Birmingham, UK: Packt Publishing, 2015.
- Salloum, Said A., Mostafa Al-Emran, Azza Abdel Monem, and Khaled Shaalan. "A Survey of Text Mining in Social Media: Facebook and Twitter Perspectives." *Advances in Science, Technology and Engineering Systems Journal* 2, no. 1 (January 2017): 127–33. <https://doi.org/10.25046/aj020115>.
- Sandhu, Tejenderkaur H. "Machine Learning and Natural Language Processing – A Review." *International Journal of Advanced Research in Computer Science* 9, no. 2 (February 20, 2018): 582–84. <https://doi.org/10.26483/ijarcs.v9i2.5799>.
- Sanem, Michael J. *Your Church Wants to Hear from You: What Is the Synod on*

Synodality? Collegeville, Minnesota: Liturgical Press, 2022.

- Scott, William. "TF-IDF for Document Ranking from Scratch in Python on Real World Dataset." Medium, May 21, 2019. <https://towardsdatascience.com/tf-idf-for-document-ranking-from-scratch-in-python-on-real-world-dataset-796d339a4089>.
- Teja, Sai. "Stop Words in NLP." Medium, June 12, 2020. <https://medium.com/@saitejaponugoti/stop-words-in-nlp-5b248dadad47>.
- Torres, Joe, and Mark Saludes. "South China Sea Ruling 'No Cause for Alarm.'" UCA News July 13, 2016. <https://www.ucanews.com/news/south-china-sea-ruling-no-cause-for-alarm/76564>.
- Willis, Adam. "The Catholic Rebels Resisting the Philippines' Deadly War on Drugs." The Guardian, July 22, 2019. <https://www.theguardian.com/world/2019/jul/18/philippines-rodrigo-duterte-war-on-drugs-catholic-church>.
- Woodward, Scott, and Royce Kimmons. "Religious Implications of Social Media in Education." *Religion & Education* 46, no. 2 (September 21, 2018): 1–23. <https://doi.org/10.1080/15507394.2018.1510249>.
- Wyrostkiewicz, Michał, Joanna Sosnowska, and Aneta Wójciszyn-Wasil. "The Catholic Influencer as a Challenge for Spiritual Leadership in the Age of Social Media." *Repozytorium*, 2022. <https://repozytorium.kul.pl/handle/20.500.12153/3876>.