Multilabel Social Support Classification of Filipino-English Endocrinology Facebook Comments Using Machine Learning Classification Models

Romaine Dara Regala[1], Geoffrey Solano[1], and Iris Thiele Isip-Tan[2]

[1] Department of Physical Sciences and Mathematics, College of Arts and Sciences
dpsm-cas@upm.edu.ph,
[2] Medical Informatics Unit, College of Medicine, University of the Philippines Manila, Ermita, Manila, PH
rmregala@up.edu.ph, gasolano@up.edu.ph,icisiptan@up.edu.ph

Abstract. Social support refers to resources that are made available through social ties. Persons with diabetes often seek social support on social media. The Endocrine Witch Facebook page posts about diabetes mellitus and is moderated by an endocrinologist. This study introduces novel machine learning models tailored for multilabel social support classification of Filipino-English comments on this Facebook page. The objective is to effectively categorize comments according to content: spam and the types of social support: informational, emotional, appraisal and instrumental. Such a classifier will help the moderator better manage the Facebook page. The dataset underwent manual data cleaning and was subsequently divided into training and testing sets. Preprocessing techniques encompassing lowercasing, tokenization, and TF-IDF vectorization were employed on both sets. To address dataset imbalances, data augmentation techniques were implemented. Notably, the LP-SVM model emerged as the top performer and was seamlessly integrated into the developed application. These findings enhance our comprehension of social support dynamics and furnish practitioners with a user-friendly tool for social support text classification.

Keywords: natural language processing, social support, diabetes, machine learning, multilabel social support classification

1 INTRODUCTION

1.1 Background of the Study

Diabetes Mellitus is a chronic, metabolic disease from which patients are characterized by elevated levels of blood sugar. This disease affects about 422 million people and is one of the major causes of death worldwide as 1.5 million deaths are directly attributed to diabetes each year [28]. As the disease progress, 33% to 50% of people with diabetes would likely have diabetes distress as well as 20% of

J. Caro et al. (eds.), Proceedings of the Workshop on Computation: Theory and Practice (WCTP 2024) , Atlantis Highlights in Computer Sciences 23,
https://doi.org/10.2991/978-94-6463-684-0_19

patients with diabetes are more likely to experience anxiety than those without. These psychological problems affect and hinder their recovery [4]. Social support has been shown to reduce the psychological and physiological effects of stress and may increase immune function [3]. Social media platforms can offer social support for persons with diabetes through Facebook pages focused on diabetes, such as the Endocrine Witch Facebook page (www.facebook.com/ EndocrineWitch), moderated by one of the authors. On average, Filipinos spend four hours and six minutes on social media mainly on Facebook [1].

Types of social support that have previously been found in online diabetes communities [16] include informational support (from peer experiences or sharing of research that healthcare professionals may not be aware of nor have time to address at clinic visits), emotional support (from other members which enhances a sense of belonging in the community or as an outlet to vent frustrations), appraisal support (prompting self-reflection and empowering patients), and Instrumental support (either through sharing of community resources or financial aid).

Having a single moderator with limited time to review comments on the Endocrine Witch Facebook page, there is benefit in building a system that can effectively automate labeling or classifying comments into different types of social support.

1.2 Study Objectives

This study aims to classify comments on diabetes-related posts on the Endocrine Witch Facebook page into different types of social support using machine learning.

1.2.1 Research Objectives

Specifically, this study aims to develop a text classification model that can identify and/or categorize diabetes-related comments posted on a Facebook page into different types of social support by accomplishing the following :

1. Manual data cleaning for spelling corrections
2. Splitting the dataset into 80% training and 20% testing
3. Apply data augmentation using OpenAI in order to handle class imbalance
4. Benchmarking different multilabel text classification models, namely:
   (a) Logistic Regression
   (b) Support Vector Machine
   (c) Ensemble Classifier Chain based on Logistic Regression
   (d) Ensemble Classifier Chain based on Support Vector Machine
   (e) MultiLayer Perceptron

5. Evaluating the model using appropriate metrics
   (a) Precision
   (b) Recall
   (c) F1-score
   (d) Hamming Loss

1.2.2 System Objectives

The study also aims to develop a web application integrated with the model and allows the Facebook page-moderator to:

1. Classify a single text
2. Classify bulk of texts/comments through CSV format
3. View list of classified comments
4. Go to specific Facebook comment
5. Export data in CSV

This study carries significant implications for both research purposes and practical applications in the field of diabetes management. By effectively classifying comments on a Facebook page related to diabetes, the study provides valuable insights into the types of support that patients seek and offer in online communities. This understanding can greatly contribute to the development of tailored content and effective support programs that address the specific needs and preferences of diabetes patients.

Moreover, the classification system developed in this study can have practical benefits for professionals and page moderators. It enables them to efficiently analyze and categorize incoming comments, facilitating more targeted and personalized responses. By quickly identifying the type of support requested or offered, healthcare professionals can provide timely and appropriate information, guidance, and encouragement to patients. This not only enhances the quality of interactions on the Facebook page but also improves the overall management of the page by streamlining the moderation process.

Overall, the accurate classification of comments, whether for research purposes or to aid professionals in their responses, is crucial for optimizing the support and engagement experienced by diabetes patients on social media platforms. This study addresses this need, opening doors to more effective communication, improved content generation, and enhanced support programs for the management of diabetes.

1.3 Scope and Limitations

The scope of this study includes the development and evaluation of a text classification model that can accurately identify and categorize comments on

diabetes-related posts on an Endocrinologist-Moderated Facebook page into different types of social support. Consequently, the research will be subject to the following limitations:

1. The study is limited to comments written in English and Filipino.
2. The study is limited to analyzing only the comments posted on the Facebook page, and may not take into account other factors that could affect the nature of requested or offered support by patients such as demographic or disease stage.
3. The study is limited to comments posted on a specific Facebook page, and the results may not generalize to other social media platforms.
4. The dataset's spelling has been subjected to manual correction, introducing potential human errors. On the other hand, grammar is not checked or corrected.
5. The CSV file to be fed to the system represents the extracted Facebook comments.

2 REVIEW OF RELATED LITERATURE

It is evident that NLP is considered beneficial to every field. NLP helps in understanding and analyzing written texts and spoken for more objective and accurate analysis. This study aims to use NLP techniques with Hybrid deep learning in the classification of social support.

2.1 Diabetes Management

Persons with diabetes find it challenging to manage the disease on their own daily. The goal of diabetes management is to keep blood pressure, glucose and cholesterol levels on target [25] while adhering to the recommended diet. A study has shown effort in managing the diet of patients with Type 2 Diabetes by creating a program called Beyond Good Intention (BGI). It shows that the participants have greater improvement in dietary quality, especially those who have no nutrition goal baseline [26].

Social Support Social support can be beneficial to diabetes management as it can reduce psychological repercussions of the challenges of the progressing disease [3]. There is a study that focused on the relationship between social support and self-care behavior in patients with diabetes where they perform a cross-sectional study on diabetes patients who are over 30 years of age in Mashad, Iran. The result showed that 40.7% of the samples had good social support, 46.2% has average social support, and the rest had poor social support which shows that the mean score of social support was average, and the majority had moderate to poor self-care. They use Spearman's rank correlation

coefficient between the domains of diabetes self-care behaviors and social support (healthy eating behaviors=0.758 correlation coefficient, physical activity=0.893 correlation coefficient, blood glucose monitoring = 0.680 correlation coefficient, and proper medication use = 0.737) which indicates that social support and diabetes self-care behaviors are significantly and directly correlated [22].

Patient Engagement Patient engagement is defined as the patient's desire and capability to actively choose to participate in care, with cooperation with a healthcare provider or institution, in order to maximize the outcomes or improve experiences of care [10]. We cannot argue that patient engagement is necessary for diabetes management but according to the study of Iqbal(2020), there are still barriers that should be overcome in this aspect. There are multiple barriers like access to services, understanding of referral requirements, and further practitioner/patient education. Improvement in these areas will hopefully lead to a change in diabetes care to prevent diabetes complications [11].

Health Communication Health communication is an area of study that examines how the use of different communication strategies can keep people informed about their health and influence their behavior so they can live healthier lives. Public health experts also recognize health communication as important to different health programs [24]. A study focused on exploring the association between patients' perceptions of communication quality with their provider and a range of patients' outcomes in type 2 diabetes revealed that their higher perceived quality of provider-patient communication was associated with improved self-management, adherence to diabetes care, and greater well-being, perceived personal control, self-efficacy, and less diabetes distress. In addition to that it is also revealed that effective communication was more related to the provision of support [21].

2.2 Social Media

With the emergence of social media, understanding human behavior as well as mental health studies become a new interest. Platforms like Facebook, Instagram, Twitter, as well as Reddit, can be a new source of support for everybody, and the exchange of information can be easily accessed. A study on patient experience and cancer survivorship on social media reveals that Instagram can be a platform for patient experience and can be beneficial for head and neck surgeons and other oncology providers. They can obtain a better understanding of the daily experience of survivorship in order for them to provide support and comprehensive cancer care [7].

2.3 Multilabel Text Classification

With the significant growth of technology around the world there have been numerous efforts with the use of NLP to understand and analyze data. Text

classification is one of the fundamental tasks in natural language processing that is used in numerous applications like sentiment analysis, and spam detection [6].

Learning Models There are numerous studies utilizing text classification with various methods. A study on the applicability of machine learning methods to Multilabel medical text classification benchmarks different methods of handling multilabel text classification. Clinical documents written in Russia of more than 250 thousand patients were used for the research for which it includes four labels. The researchers performed some basic preprocessing like cleaning symbols and extra spaces, minimizing notes, correcting syntax, spelling correction, tokenization, and lastly, vectorizing both the training set and test sets using Bag of Words. The best-performing machine learning models are Logistic Regression and Support Vector Machine but the performance metrics become better with the Ensemble Classifier Chains [15] as shown in Figure 1.

Model	Precision		Recall		F-measure	
	Micro	Macro	Micro	Macro	Micro	Macro
Shallow classifiers						
MNB	0.781	0.764	0.864	0.873	0.864	0.852
LR	**0.866**	**0.850**	**0.920**	0.915	**0.920**	**0.910**
Linear SVM	**0.865**	**0.849**	**0.919**	0.916	**0.919**	**0.909**
k-NN	0.694	0.715	0.803	0.827	0.803	0.809
Ensembles of classifier chains						
ECCLR	**0.867**	**0.852**	**0.925**	**0.921**	**0.922**	**0.912**
ECCSVM	**0.872**	**0.855**	**0.927**	**0.922**	**0.924**	**0.914**

Fig. 1. Performance results of the classification models

Another study on multi-label Arabic text categorization also performed benchmark and baseline comparisons of multi-label learning algorithms. Multilabel text classification has two popular methods, which are the transformation-based methods (Binary Relevance, Classifier Chains, Calibrated Ranking by Pairwise Comparison, and Label Powerset) and the adaption-based methods (Multi-label k-Nearest Neighbors, Instance-Based Learning by Logistic Regression Multilabel, Binary Relevance kNN, and RFBoost), where the three well-known multiclass algorithms (Support Vector Machine, k-Nearest-Neighbors, and Random Forest) were used as the base learners for the transformation-based methods. The results showed that RFBoost significantly outperforms the other methods and the transformation-based methods perform well when the Support Vector Machine is used as a base classifier [2].

Classifier	Macro-Precision		Macro-Recall		Macro-F1		Micro-Precision		Micro-Recall		Micro-F1	
	score	f. size	score	f. size	score	f. size	score	f. size	score	f. size	score	f. size
Problem transformation approaches												
BR-*k*NN	**83.34**	2000	47.60	200	56.22	200	**85.66**	4000	55.26	200	64.77	200
BR-RF	82.80	4000	57.43	200	62.23	200	84.81	3,000	61.99	200	67.56	200
BR-SVM	76.89	500	61.24	4000	65.88	4000	82.39	200	64.20	4000	69.89	500
CLR-*k*NN	82.83	2000	47.43	200	55.57	200	85.03	4000	55.33	200	64.21	200
CLR-RF	82.14	3000	59.31	200	63.50	200	82.41	3000	63.98	200	68.38	200
CLR-SVM	76.05	500	64.59	4000	67.84	4000	81.56	200	67.30	4000	70.72	4000
CC-*k*NN	81.89	4000	50.92	200	57.16	200	80.55	4000	57.10	200	63.52	200
CC-RF	82.91	4000	57.91	200	61.78	200	84.80	4000	62.44	200	69.01	500
CC-SVM	73.12	500	62.02	2000	66.58	4000	76.67	500	65.08	4000	69.57	500
LP-*k*NN	74.37	500	52.16	200	57.99	200	71.27	200	58.08	200	64.00	200
LP-RF	68.42	500	52.80	200	57.80	200	72.19	500	59.68	200	65.28	500
LP-SVM	76.70	4000	**64.86**	4000	**69.79**	4000	79.72	4000	**67.39**	4000	**73.04**	4000
Algorithm adaptation approaches												
RFBoost	77.11	4000	63.74	500	**68.98**	4000	77.44	3000	**68.57**	2000	**72.57**	3000
ML*k*NN	79.21	3000	48.37	1000	56.39	1000	83.92	4000	53.52	1000	64.97	1000
BR*k*NN	**85.57**	1000	43.40	200	52.68	200	**90.10**	4000	51.41	200	62.83	200
IBLRML	79.01	500	47.50	200	56.44	200	84.43	2000	51.88	500	64.08	500

Fig. 2. The best-obtained scores (%) of all compared methods.

A master's thesis [9] focuses on a multi-label text classification task using a small dataset of bilingual short texts (emails) in both English and Swedish. The dataset consists of 5,800 emails that are categorized into 107 classes, with the majority of the emails containing both languages. Various models, including Support Vector Machines (SVM), Gated Recurrent Units (GRU), Convolutional Neural Network (CNN), Quasi Recurrent Neural Network (QRNN), and Transformers, were employed to handle this task.

The experimental results reveal that the SVM model outperforms the other models in terms of the weighted average F1 score, achieving a score of 0.96. The CNN model follows with a score of 0.89, and the QRNN model obtains a score of 0.80. These findings highlight the effectiveness of the SVM model for the given multi-label text classification task on the bilingual email dataset [14].

	SVM	GRU	Very deep CNN	QRNN	Transformers
Weighted avg F1	0.96	0.67	0.89	0.80	0.53
Hamming loss	0.0077	0.0447	0.0235	0.0375	0.0461

Fig. 3. Metrics comparison among the models

In summary, there are different methods one can utilize to effectively perform text classification to understand a specific topic. The lack of research about social support types being requested and offered online is evidence that this can be a gap available for future studies.

3 ESSENTIAL CONCEPTS

Natural Language Processing (NLP) has existed for more than 50 years and has a variety of real-world applications in different fields, including medical research, search engines, and even business intelligence [17]. NLP is a subfield of Artificial Intelligence that aims to process, analyze, and natural language data (spoken and written). The goal of NLP is to enable machines to interact with humans in a way that resembles natural human communication [27].

3.1 Text classification

Text classification refers to a technique in Natural Language Processing that focuses on categorizing text documents into two or more categories. The most common form of text classification is binary classification, which assigns a document to one of two categories. The process of text classification involves extracting a set of features that describe the document and then applying an algorithm to use these features to classify the document into its appropriate category [18].

3.2 Label Powerset

The label powerset method takes a multi-label dataset and changes it to a single multi-class dataset by treating each label combination as a separate class. It achieves multi-label categorization by associating an instance with a class composed of a set of labels. A class is assigned to each unique label set, after which a multi-class classifier is trained to assign an instance to one of the classes [19].

3.3 Classifier Chain

This method is comparable to binary relevance. However, it considers label correlation. This method employs a chain of classifiers, with each classifier taking as input the predictions of all previous classifiers. The number of classifiers is the same as the number of classes [8].

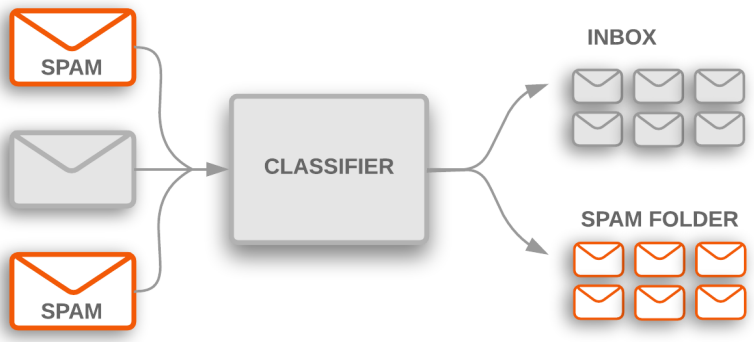

Fig. 4. Example of a Text Classification Application

3.4 Support Vector Machine

The Support Vector Machine (SVM) is a powerful supervised machine learning algorithm that excels in classification tasks. It utilizes labeled input data, divided into two distinct classes, and aims to predict the classification of a query sample. By transforming the data into a higher-dimensional space, SVM employs a support vector classifier to determine a threshold, known as a hyperplane, which effectively separates the two classes with minimal error. This technique enables SVM to make accurate predictions and effectively classify new samples based on their characteristics [13].

3.5 Hamming Loss

In multiclass classification, the evaluation of model performance is often measured using the Hamming loss, which quantifies the similarity between the true labels (y_true) and the predicted labels (y_pred). This loss can be viewed as the Hamming distance between the two label sets, or equivalently, as a subset zero-one loss when the normalized parameter is set to True [23].

In contrast, multilabel classification introduces a different interpretation of the Hamming loss compared to the subset zero-one loss. The zero-one loss considers an entire set of labels for a specific sample as incorrect if it does not perfectly match the true set of labels. On the other hand, the Hamming loss takes a more lenient approach and penalizes only the individual labels that are incorrectly predicted [23].

4 DESIGN AND IMPLEMENTATION

4.1 Dataset

The dataset utilized for modeling consists of manually annotated/labeled comments extracted from Dr. Isip-Tan's Facebook page. The dataset includes a total of 9,767 comments that were posted on 196 diabetes-related posts from 1 January 2018 to 31 March 2021, spanning a period of several years. The dataset used for modeling is the manually annotated/labeled comments from Dr. Isip-Tan's Facebook page.

In identifying the themes, a six-phase reflexive thematic analysis was conducted following the procedure of Braun and Clarke [5] : data familiarization, generating initial codes, constructing themes, revising themes, defining themes, and then writing the report. Comments on the lone Facebook live video were excluded from the analysis. Using the above-mentioned steps, Dr. Isip-Tan performed thematic analysis [5] which aimed to explore the types of support provided, received, and offered to Facebook users on the page. Five themes were identified in the dataset:

- informational : sharing of peer experiences or research
- emotional : enhancing a sense of belonging or as an outlet to vent frustrations
- appraisal : prompting self-reflection
- instrumental : sharing of resources
- spam : anything not related to diabetes

Dr. Isip-Tan and her team then labeled the comments accordingly.

4.2 Use Cases

The use-case diagram (Figure 5) presents the system to be used by the Facebook moderator. The Facebook moderator should be able to classify a single text or classify bulk comments inside a CSV file. The Facebook moderator should also see the list of classified comments. Lastly, the Facebook moderator should be able to export the data to a CSV file.

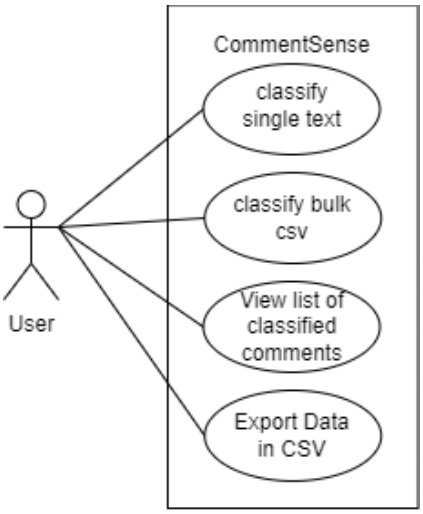

Fig. 5. Use-Case Diagram of System

4.3 System Design

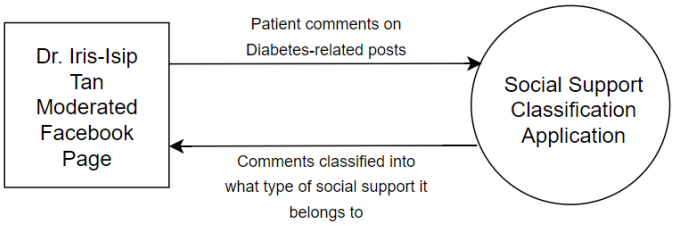

Fig. 6. Context Diagram

4.3.1 Context Diagram

Figure 6 represents how the system works. Patient comments on Diabetes-related posts on Dr. Isip-Tan's Facebook page are the data to be processed by the system. The system then outputs the classified comments based on what type of social support it belongs to.

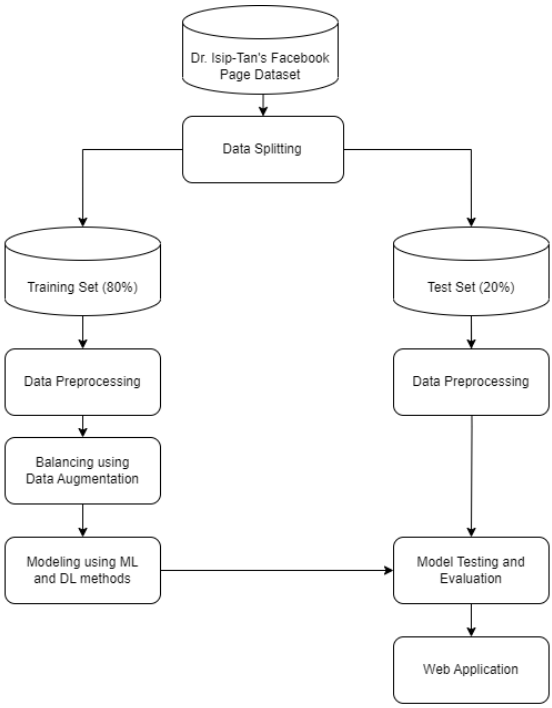

Fig. 7. Data Flow Diagram

4.3.2 Data Flow Diagram

Figure 7 shows the data flow from the dataset to the web application development. The dataset is from Dr. Iris Isip-Tan's labeled comments extracted from her moderated Facebook page. We will perform data splitting into training and testing sets where both of them will be subject to data preprocessing. After data preprocessing, the training set will be balanced using data augmentation utilizing OpenAI. After rebalancing,it will now proceed to modeling using LP-Logistic Regression, LP-Support Vector Machine, ECCLR, ECCSVM, and MLP. After training the model, accuracy, precision, recall, F1-score, and Hamming loss will be used to evaluate the model. Lastly, the best-trained model is used for the development of the web application.

4.4 System Architecture

The system makes use of Python as its main programming language. For the development of web application, the system also makes use of SQLite for its database server while using Django as its framework. For modeling, python machine learning libraries like scikit-multilearn and scikit-learn will be utilized as well as libraries like pandas, numpy, and joblib. These are the detailed list of the important libraries.

- Django 4.2.3
- emoji 2.6.0
- joblib 1.3.1
- nltk 3.8.1
- numpy 1.25.0
- pandas 2.0.3
- requests 2.31.0
- scikit-learn 1.3.0
- scikit-multilearn 0.2.0
- scikit-optimize 0.9.0
- scipy 1.11.1

4.5 Technical Architecture

Initially, the system should be accessible via a link that works on any browser and can be viewed either from a computer or a smartphone but during the course of development and time restrictions it is only deployed using localhost through Django.

The system requirement would be a computer unit with the following minimum specifications :

- Browser (Chrome, Edge, Safari, etc)
- Windows 7 or higher
- Stable internet connection

5 RESULTS

5.1 Exploratory Data Analysis

The dataset consists of comments from diabetes-related posts on the Endocrine Witch Facebook page from 1st of January 2018 to 31st of March 2021. These comments were copied into Word documents per post. These comments were labeled by one of the authors according to the types of social support, and spam on nVivo. Names of Facebook users were coded to hide their identities.

To ensure the accuracy of the data, the comments underwent a manual review to identify and correct any spelling errors. After this meticulous data-cleaning process, the resulting dataset contained 2941 rows. The figure below presents an overview of the dataset's information.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2942 entries, 0 to 2941
Data columns (total 3 columns):
 #   Column   Non-Null Count  Dtype
---  ------   --------------  -----
 0   id       2942 non-null   object
 1   label    2942 non-null   object
 2   comment  2942 non-null   object
dtypes: object(3)
memory usage: 69.1+ KB
```

Fig. 8. Information of the dataset

Figure 8 shows that the dataset has 3 columns mainly: id with type object, label with type object, and comment with type object. The comment column will be the feature variable while the label column is the target variable.

The data is then partitioned for training and testing.

5.2 Handling Class Imbalance using OpenAI Data Augmentation

Upon analyzing the dataset, a class imbalance is observed among the labels. Such class imbalance in a multilabel text classification cannot be handled by the Imbalanced-learn Python library for which novel class balancing methods is provided (SMOTE, Random Undersampling, etc). In order to handle class imbalance, OpenAI's API is utilized to create augmented sentences as needed by the training set.

In Figure 9, the imbalance is visible among the labels as the Emotional count is relatively high compared to others and the Spam label count is relatively low among the others. There is at most 484 difference in the count of the emotional labels and spam labels.

The duplicates or noise in the Emotional label are manually removed leaving a count of 734 emotional labeled comments. This becomes the basis of how much augmented data should be added per label in the training set.

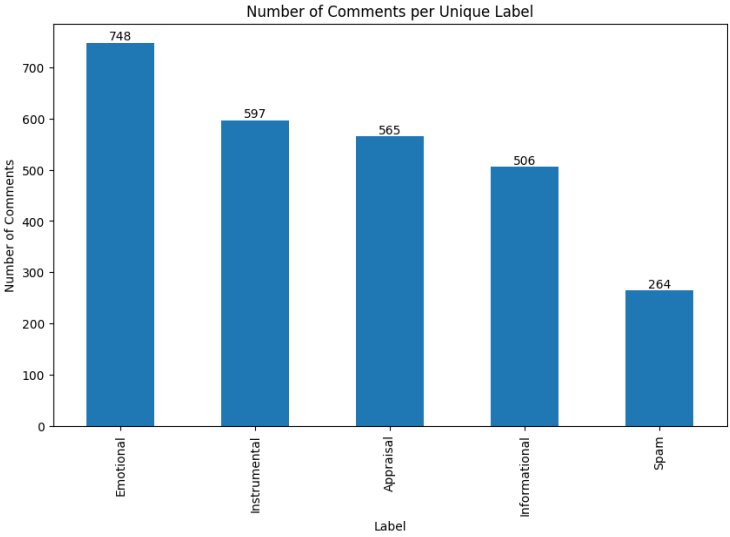

Fig. 9. Trainset imbalance

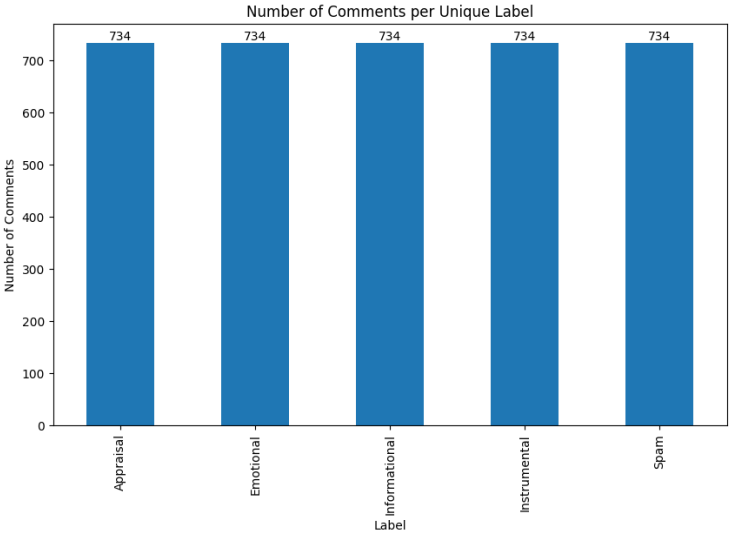

Fig. 10. Rebalanced trainset using Data Augmentation

5.3 Data Preprocessing

Both the imbalanced dataset and the balanced dataset were processed in a similar manner, employing a set of preprocessing methods. For the training set, several steps were applied, including lowercasing letters, removing numbers, eliminating single characters, removing multiple spaces, tokenization, removing stop words, and TF-IDF vectorization. These steps aimed to standardize the textual data and extract meaningful features for training purposes. Similarly, for the test set, preprocessing steps such as removing multiple spaces and converting the data to binary using the MultiLabelBinarizer were performed. These preprocessing techniques were essential in ensuring data consistency and preparing the dataset for subsequent analysis and modeling tasks.

5.4 Model Evaluation

The model training process is divided into two distinct batches: training with imbalanced data and training with a balanced training set. For evaluating the performance of each model, both the Classification R eport a nd t he hamming loss were utilized.

	Imbalanced			Balanced		
Labels	Precision	Recall	F1-Score	Precision	Recall	F1-Score
Appraisal	0.68	0.63	**0.66**	**0.69**	**0.64**	**0.66**
Emotional	0.79	**0.77**	0.78	**0.83**	0.76	**0.79**
Informational	0.56	0.44	0.49	**0.55**	**0.45**	**0.50**
Instrumental	0.69	**0.61**	**0.65**	**0.70**	0.56	0.63
Spam	**0.86**	0.48	0.62	0.78	**0.62**	**0.70**
Micro AVG	0.70	**0.61**	0.65	**0.71**	**0.61**	**0.66**
Macro AVG	**0.72**	0.59	0.64	0.71	**0.61**	**0.65**
Weighted AVG	0.70	**0.61**	0.65	**0.71**	**0.61**	**0.66**
Samples AVG	0.70	**0.65**	**0.67**	0.71	**0.65**	**0.67**

Table 1. Classification Report For Imbalanced and Balanced datasets - Label Powerset Logistic Regression

Table 1, shows that most of the label in LP-Logistic Regression performs a little better in augmented data. It is notable that the most constant labels are Emotional and Spam. On the other hand, Instrumental Label seems to decrease in metrics when augmented data is applied. Overall, this model performs much better in imbalanced data.

	Imbalanced			Balanced		
Labels	Precision	Recall	F1-Score	Precision	Recall	F1-Score
Appraisal	0.66	0.64	0.65	**0.68**	**0.66**	**0.67**
Emotional	0.86	**0.75**	0.80	**0.89**	**0.75**	**0.81**
Informational	**0.56**	**0.47**	**0.51**	0.54	0.46	0.50
Instrumental	0.69	**0.64**	0.66	**0.75**	**0.64**	**0.69**
Spam	**0.96**	0.41	0.57	0.90	**0.56**	**0.69**
Micro AVG	0.71	0.61	0.66	**0.74**	**0.63**	**0.68**
Macro AVG	**0.75**	0.58	0.64	0.74	**0.63**	**0.68**
Weighted AVG	0.73	0.61	0.66	**0.75**	**0.63**	**0.68**
Samples AVG	0.71	0.65	0.67	**0.74**	**0.67**	**0.69**

Table 2. Classification Report For Imbalanced and Balanced datasets - Label Powerset Support Vector Machine

Table 2, shows that most of the label in the LP-Support Vector Machine model performs a little bit better in augmented data, particularly Appraisal, Emotional, and Instrumental. Just like in LP-LR model, it is notable that the most constant label is Emotional. On the other hand, Informational Label seems to decrease in metrics when augmented data is applied. Overall, this model performs much better in imbalanced data.

	Imbalanced			Balanced		
Labels	Precision	Recall	F1-Score	Precision	Recall	F1-Score
Appraisal	0.70	**0.48**	**0.57**	**0.72**	0.47	**0.57**
Emotional	0.86	**0.74**	**0.80**	**0.88**	0.72	0.79
Informational	**0.59**	0.43	0.50	0.58	**0.45**	**0.51**
Instrumental	**0.56**	**0.72**	**0.63**	**0.56**	0.65	0.61
Spam	**0.74**	0.48	0.58	0.58	**0.66**	**0.62**
Micro AVG	**0.67**	**0.59**	**0.63**	**0.67**	**0.59**	0.62
Macro AVG	**0.69**	0.57	**0.62**	0.67	**0.59**	**0.62**
Weighted AVG	**0.69**	**0.59**	**0.63**	0.68	**0.59**	**0.63**
Samples AVG	**0.68**	**0.62**	**0.64**	**0.67**	0.62	**0.63**

Table 3. Classification Report For Imbalanced and Balanced datasets - Ensemble Classifier Chain Logistic Regression

Table 3, shows that most of the label in Ensemble Classifier Chain Logistic Regression does not perform well in augmented data than the two earlier models, particularly Instrumental. Overall, this model performs slightly better in imbalanced data.

	Imbalanced			Balanced		
Labels	Precision	Recall	F1-Score	Precision	Recall	F1-Score
Appraisal	0.75	**0.43**	**0.55**	**0.76**	0.42	0.54
Emotional	0.84	**0.76**	**0.80**	**0.85**	**0.76**	**0.80**
Informational	**0.52**	0.48	0.50	**0.52**	**0.50**	**0.51**
Instrumental	0.60	0.70	**0.65**	0.62	**0.71**	**0.66**
Spam	**0.85**	0.52	**0.64**	0.77	**0.53**	0.63
Micro AVG	**0.69**	**0.60**	**0.64**	**0.69**	**0.60**	**0.64**
Macro AVG	**0.71**	0.58	**0.63**	**0.71**	0.58	**0.63**
Weighted AVG	0.70	**0.60**	0.63	**0.71**	**0.60**	**0.64**
Samples AVG	**0.69**	**0.63**	**0.65**	**0.69**	**0.63**	**0.65**

Table 4. Classification Report For Imbalanced and Balanced datasets - Ensemble Classifier Chain Support Vector Machine

Table 4, shows that most of the label in the Ensemble Classifier Chain Support Vector Machine model performs a little bit better in augmented data, particularly Emotional, and Informational. Just like in other models, it is notable that the most constant label is Emotional. On the other hand, Appraisal and Informational Label seems to decrease in metrics when augmented data is applied. Overall, this model performs much better in balanced data.

	Imbalanced			Balanced		
Labels	Precision	Recall	F1-Score	Precision	Recall	F1-Score
Appraisal	**0.66**	**0.51**	**0.58**	0.64	**0.51**	0.57
Emotional	0.76	**0.76**	0.76	**0.82**	0.72	**0.77**
Informational	**0.59**	**0.41**	**0.49**	0.52	0.35	0.42
Instrumental	**0.62**	0.49	**0.55**	0.59	**0.50**	0.54
Spam	**0.72**	0.44	0.54	0.63	**0.56**	**0.60**
Micro AVG	**0.67**	**0.54**	**0.60**	0.65	0.53	0.59
Macro AVG	**0.67**	0.52	**0.58**	0.64	**0.53**	**0.58**
Weighted AVG	**0.67**	**0.54**	**0.59**	0.65	0.53	0.58
Samples AVG	**0.58**	**0.58**	**0.57**	0.57	**0.58**	0.56

Table 5. Classification Report For Imbalanced and Balanced datasets - MultiLayer Perceptron

Table 5, shows that most of the label in the MultiLayer Perceptron model does not perform well in augmented data, particularly Appraisal, Instrumental and Informational. Just like in other models, it is notable that the most constant label is Emotional. Overall, this model performs much better in imbalanced data.

Metrics	LP-LR	LP-SVM	ECC-LR	ECC-SVM	MLP
Weighted AVG(Precision)	0.70	**0.73**	0.69	0.69	0.67
Weighted AVG(Recall)	**0.61**	**0.61**	0.59	0.60	0.54
Weighted AVG(F1-Score)	**0.71**	0.66	0.63	0.63	0.59
Hamming Loss	0.152	**0.149**	0.163	0.159	0.170

Table 6. Imbalanced models comparison

Metrics	LP-LR	LP-SVM	ECC-LR	ECC-SVM	MLP
Weighted AVG(Precision)	0.71	**0.75**	0.68	0.71	0.65
Weighted AVG(Recall)	0.61	**0.63**	0.59	0.60	0.53
Weighted AVG(F1-Score)	0.66	**0.68**	0.63	0.64	0.58
Hamming Loss	0.150	**0.140**	0.165	0.157	0.176

Table 7. Balanced models comparison

Upon analyzing tables 6 and 7, even though with below average metrics, Label Powerset Support Vector Machine performs better than the other models.

6 DISCUSSIONS

This study focuses on applying various learning models (LP-LR, LP-SVM, ECC-LR, ECC-SVM, and MLP) to classify texts or comments from a Facebook page moderated by an endocrinologist, specifically related to diabetes. The goal is to categorize these comments into different types of social support, including informational (sharing of peer experiences or research), emotional (enhancing a sense of belonging or as an outlet to vent frustrations), appraisal (prompting self-reflection), instrumental (sharing of resources), and spam (anything not related to diabetes). The evaluation process revealed that the Label Powerset Support Vector Machine model performed the best and was integrated into a user-friendly application called CommentSense.

The study timeline was significantly influenced by the dataset, which required manual cleaning, revision, and augmentation. The initial dataset had a poor structure, posing challenges for the models to achieve optimal performance. However, data augmentation was applied to improve the initial metrics.

To evaluate the performance of each model, a classification report was utilized, providing a detailed analysis of the metrics. Examining the performance of each model across different labels, the Emotional label consistently achieved the highest metrics, followed by Appraisal or Spam. The Instrumental label had relatively lower metrics, and the Informational label performed last.

Overall, CommentSense provides an intuitive interface for users to effectively classify texts into various social support categories and identify spam. While Emotional support stands out as the most consistent performer, further analysis

of each model's performance sheds light on the strengths and weaknesses across different support categories.

7 CONCLUSIONS

This study presents innovative machine learning models for multilabel social support classification of Filipino-English Endocrinology Facebook Comments. The data, collected and labeled by Dr. Iris Isip-Tan, underwent manual preprocessing, including inputting the data into Excel, performing data cleaning tasks such as spelling corrections, removing whitespace, eliminating Dr. Iris Isip-Tan's replies, and removing commentors' aliases. The dataset was then split into training and testing sets for further analysis.

Both the training set and test set underwent preprocessing techniques. For the training set, a series of steps were applied, including lowercase conversion, number removal, elimination of single characters, removal of multiple spaces, tokenization, stop word removal, and TF-IDF vectorization. Similarly, the testing set underwent preprocessing steps such as removing multiple spaces and converting the data to binary using the MultiLabelBinarizer.

The researcher observed class imbalance in the dataset and addressed it through data augmentation using OpenAI. The model training process consisted of two batches: one for the imbalanced dataset and one for the balanced dataset. The models were trained and evaluated using the classification report, which provided insights into the performance per label and overall weighted averages, as well as the hamming loss. The results indicated that the LP-SVM model outperformed the other models, followed by the LP-LR model. However, it was noted that the metrics of the LP-SVM model were still lower than expected. As a result, the LP-SVM model with a threshold of 0.3 was integrated into the CommentSense application, allowing users to classify texts and comments efficiently which can facilitate a better understanding of how social support is requested or offered.

8 RECOMMENDATIONS

In order to improve the performance of the models used in this study, it is recommended to have a clean and well-structured dataset that can clearly represent the labels. The comments should be tagged individually and not be thread or bulk in order for the models to understand the patterns clearly. In addition to that, it is recommended to create an automation that can automate spelling checks to minimize human error.

Exploring ensemble methods and incorporating contextual word representations offer promising avenues for enhancing the study's outcomes thus it is recommended. Ensemble methods, which involve combining multiple classifiers or utilizing ensemble learning techniques, have the potential to significantly improve

the overall performance of the models. By leveraging the diversity and complementary strengths of different classifiers, ensemble methods can enhance accuracy and robustness. Furthermore, incorporating contextual word embeddings into the machine learning models enables a deeper understanding of the comments and texts, capturing rich semantic and contextual information. Contextual word representations have demonstrated their effectiveness in improving classification accuracy and robustness across various domains. By exploring ensemble methods and contextual word representations in a study like this, further improvements can be achieved, advancing the understanding and categorization of social support in the domain of Endocrinology and other domains.

Furthermore, extending the application of CommentSense to other social media platforms commonly used by individuals seeking diabetes-related support is recommended. By expanding the application's reach, a more comprehensive and holistic approach to social support classification and spam detection can be achieved, catering to a broader user base.

References

1. Amurthalingam, S.: Social Media Statistics in the Philippines [2022]. Meltwater (2022). https://www.meltwater.com/en/blog/social-media-statistics-philippines
2. Al-Salemi, B., Ayob, M., Kendall, G., Noah, S.A.M.: Multi-label Arabic text categorization: A benchmark and baseline comparison of multi-label learning algorithms. *Information Processing & Management* 56(1), 212–227 (2019). Elsevier.
3. Bakken, E.E.: Social Support. University of Minnesota (2022). https://www.takingcharge.csh.umn.edu/social-support
4. Centers for Disease Control and Prevention: Diabetes and Mental Health. (2022). https://www.cdc.gov/diabetes/living-with/mental-health.html
5. Clarke, V. & Braun, V.: Successful Qualitative Research: A Practical Guide for Beginners.(2013).
6. Deutschman, Z.: Multi-Label Text Classification. (2019). https://towardsdatascience.com/multi-label-text-classification-5c505fdedca8
7. Gao, R.W., Smith, J.D., Malloy, K.M.: Head and neck cancer and social media: the patient experience and cancer survivorship. The Laryngoscope 131(4), E1214–E1219 (2021). Wiley Online Library.
8. George, J.A.: An introduction to multi-label text classification. Medium. https://medium.com/analytics-vidhya/an-introduction-to-multi-label-text-classification-b1bcb7c7364c (2020).
9. Harsha Kadam, S., Paniskaki, K.: Text analysis for email multi label classification.(2020).
10. Higgins, T., Larson, E., Schnall, R.: Unraveling the meaning of patient engagement: a concept analysis. *Patient Education and Counseling* 100(1), 30–36 (2017). Elsevier.
11. Iqbal, M.: What are the Barriers to Patient Engagement in Managing Type 2 Diabetes? How Can We Overcome Them?: A Review Article. Asian Journal of Research and Reports in Endocrinology, 1–5 (2020).
12. Isip-Tan, I.T.: Endocrine Witch. https://www.facebook.com/EndocrineWitch (2024).

13. Jain, K.: What is support vector machine? Medium. https://towardsdatascience.com/what-is-support-vector-machine-870a0171e690 (2021).
14. Kadam, S.H., Paniskaki, K.: Text analysis for email multi label classification. (2020).
15. Lenivtceva, I., Slasten, E., Kashina, M., Kopanitsa, G.: Applicability of machine learning methods to multi-label medical text classification. In: Computational Science–ICCS 2020: 20th International Conference, Amsterdam, The Netherlands, June 3–5, 2020, Proceedings, Part IV 20, pp. 509–522. Springer (2020).
16. Litchman, M. L., Walker, H. R., Ng, A. H., Wawrzynski, S. E., Oser, S. M., Greenwood, D. A., Gee, P. M., Lackey, M., Oser, T. K. (2019). State of the Science: A Scoping Review and Gap Analysis of Diabetes Online Communities. Journal of diabetes science and technology, 13(3), 466–492. https://doi.org/10.1177/1932296819831042
17. Lutkevich, B.: Natural Language Processing (NLP). (2021). https://www.techtarget.com/searchenterpriseai/definition/natural-language-processing-NLP
18. Miner, G., Delen, D., Elder, J., Fast, A., Hill, T., Nisbet, R.A.: Chapter 7 - Text Classification and Categorization. In: Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications, pp. 881–892. Academic Press, Boston (2012). https://doi.org/10.1016/B978-0-12-386979-1.00035-9
19. Motro, Y.: Overview on multi-label classification. Tasq.ai. https://www.tasq.ai/blog/multi-label-classification (2022).
20. NVIVO : a software program used for qualitative and mixed-methods research. https://lumivero.com/products/nvivo/ (2024).
21. Peimani, M., Nasli-Esfahani, E., Sadeghi, R.: Patients' perceptions of patient–provider communication and diabetes care: A systematic review of quantitative and qualitative studies. Chronic Illness 16(1), 3–22 (2020). SAGE Publications Sage UK: London, England.
22. Sarpooshi, D., Mahdizadeh, M., Jaferi, A., Robatsarpooshi, H., Haddadi, M., Peyman, N.: The relationship between social support and self-care behavior in patients with diabetes mellitus. Family Medicine & Primary Care Review 23(2), 227–231 (2021). Termedia.
23. Tsoumakas, G., Katakis, I.: Multi-label classification. International Journal of Data Warehousing and Mining 3(3), 1–13 (2007). doi:10.4018/jdwm.2007070101
24. Tulane University: 10 Strategies for Effective Health Communication. (2020). https://publichealth.tulane.edu/blog/health-communication-effective-strategies
25. UCSF Health: Diabetes Mellitus Treatments. https://www.ucsfhealth.org/conditions/diabetes-mellitus/treatment
26. Van Der Velde, L.A., Kiefte-de Jong, J.C., Rutten, G.E., Vos, R.C.: Effectiveness of the Beyond Good Intentions Program on Improving Dietary Quality Among People With Type 2 Diabetes Mellitus: A Randomized Controlled Trial. Frontiers in Nutrition 8, 583125 (2021). Frontiers Media SA.
27. Vasiliev, Y.: Natural Language Processing with Python and SpaCy: A Practical Introduction. No Starch Press (2020).
28. WHO: Diabetes. World Health Organization (n.d.). https://www.who.int/news-room/fact-sheets/detail/diabetes