Patient's Feedback Platform for Quality of Services via "Free Text Analysis" in Healthcare Industry

Muhammad Awais [1], Shazia Batool [1] , Amir Mehmood Mirza [1], *Ahthasham Sajid [1], Amir Shahzad Khokhar [1], Afia Zafar [2]

1.Department of Computer Science, Faculty of ICT, BUITMS
Quetta, Pakistan
2. Department of Computer Science, NUTECH University
Islamabad, Pakistan
Email: m.awais890@gmail.com ,sbatool566@gmail.com ,mirza.amir@buitms.edu.pk ,
ahthasham.sajid@buitms.edu.pk , amir.shahzad@buitms.edu.pk , afia.zafar9@gmail.com
*Corresponding Author: ahthasham.sajid@buitms.edu.pk

Abstract

Data analysis of social media posting continues to offer a huge variety of information about the health situation faced by an individual. Social networking or social media websites provide us a wealth of information generated by users in a variety of domains, that generated information are unstructured and unlabeled and are not captured in an exceedingly systematic manner, as info generated is not humanly possible to process due to its size. One traditional way of collecting patient's experience is by conducting surveys and questionnaires, as these methods ask fixed questions and are expensive to administer. In this paper, a patient feedback platform (PFP) using free text sentiment analysis is developed to computationally identify and categorize the polarity expressed in a piece of text. Results achieved have shown 88 % accuracy in predicting whether or not a patient will recommend the service. Based on which it is recommended that developed (PFP) patient feedback platform could be used to improve E-health care services indeed.

Keywords: Sentiment Analysis, Polarity, Healthcare.

1   INTRODUCTION

With the advancement in technology, various fields of computing have emerged such as cloud computing, Internet of Things (IoT), big data etc. This technological advancement has resulted in the generation of vast information on the world wide web, especially on social networking platforms. These platforms provide a convenient way to share user experience in the form of tweets, updates, reviews, etc. Such tweets, comments, blogs, etc. have also been referred to as free text and are unstructured and are freely accessible for people to read and learn from. According to a survey conducted in the

United States, 85% of the adults use the internet and 25% have read someone else feedback on healthcare whereas only 11% of those consulted based on that feedback [1]. To deliver efficient care, health service providers need to have a holistic view of the day-to-day operation of businesses. The wellbeing of the patient and the effectiveness of the practitioner are two of the most significant factors in the area of e-health. Besides, any medical practice must continually generate value for its business by providing real and authentic customer service. Thus, without increasing the number of employees, these enterprises cannot effectively collect enough data to improve the overall service; an effort that can be both cumbersome and uneconomic at the same time.

Twitter data has been utilized for various tasks, for instance, predicting stock and financial market [2], monitoring occurrences of major events such as disease outbreak, natural disasters such as tsunami or earthquake and the reactions of the users to events [3][4]. SA is one of the frequently used methods to analyze the opinions of users expressed over the internet. SA is also widely used for product reviews such as kitchen appliances, beauty products, etc.

In this study, we have developed a platform to collect patients' feedback in the form of free text. Furthermore, we have analyzed this free text by employing machine learning to classify the underlying sentiment to formulate predicted system accuracy; further details could be found in section 4. The main objective of this study is to analyze free text to assess whether a patient will recommend a healthcare service or vice versa so that the healthcare service provider improves the services.

2   RELATED WORKS

The determination of public views on social issues is essential to governance and decision-making. Madhukar et al. [5] has addressed this issue and proposed a hybrid method for sentiment analysis of social issues i.e. a girl child in India. There are big quantities of unstructured free, text healthcare data accessible on the internet in blogs, social networks, and websites that are not systematically captured. Greaves at el. [1] employed machine learning to comprehend the unstructured remarks of patients about their care. Compared to the patient's quantitative rating of their care, the proposed solution automatically predicts whether a patient would recommend a hospital-based on two parameters (clean and dignity). Wallace et al. [6] define a probabilistic generative model that captures latent sentiment through aspects of care (e.g., interpersonal way). The research addresses particular aspects by using a small set of manually annotated comments. The authors conducted a regression analysis to determine whether model performance enhances correlation with state-level healthcare measures in the USA and correlates with state-level quality healthcare measures, including the probability that patients will visit their primary care physician within 14 days of discharge. Ivanciuc et al. [7] assessed whether the free-text answers system could automatically predict the opinions of

patients on several subjects. A classification method for machine learning was selected in which an algorithm learns to classify remarks from a specified set of examples into classifications using an open source data mining software Weka. There are elements of patient remarks that are very difficult to process for feeling assessment. Irony, sarcasm, and humor, frequently adopted by English speakers when talking about their care, cannot be easily detected using this process and the same is true for any other language [8].

A semantic analysis approach has been in Saudi telecom companies to determine customer perception [9]. They have used semantic analysis over collecting real time twitter data for the enhancement of customer relationship management (CRM). Authors have used NLP tools to filter tweets and further stored it in the relevant database containing semantic oriented words. POS tagging is further applied to words in the tweet and given sentiment score, at the end average score is calculated overall. One important but sensitive aspect of our wellbeing is Healthcare. We undoubtedly have interacted with healthcare in some way. While common ailments will be treated in a very prescribed manner with uniform processes, healthcare is an extremely personal service. obsessed with the condition and therefore the resources offered to treat it, the outcome of a patient's expertise with a health service will greatly vary. Treatment had gone extremely well and a high commonplace of care was received or treatment went terribly and the standard of care was below from what was expected [10].

Based on how the healthcare industry is currently structured around the world, health providers and organizations can be categorized into small, medium-sized and large-scale companies. For small and medium-sized companies, it is relatively easy to facilitate their patients and clients. But for large-scale health providers, continuously measuring the quality and improving the quality of the services could be an expensive job.

With the emergence of social networking platforms such as twitter etc., more and more practices are turning to the online world of reviews, marketing and organic reach-growth. A good review seems to be a good way to make any good practice authentic. While many industries are evolving by collecting feedback from their customers, medical practices can also gain traction by collecting, correlating and tracking patient's experience through their online feedback. Yet synchronizing such reviews from different types of platforms to one equivalent & specialized one which is uniquely designed for medical practices, isn't readily available for the market. A Gap that's truly setting the industry back both in the online world and in customer reach & acquisition.

3   ORIGINALITY

To address afformentioned issues, we propose a solution to analyzing free text for assessing patient's experience to improve the quality of services in the healthcare industry. Ideally, a high standard of care would be preferred

but there are many factors involved that can affect a patient's experience, good or bad patient feedback can be collected about healthcare. This patient feedback is processed using sentiment analysis and using tools like Valence Aware Dictionary and Sentiment Reasoner (VADER) [11] dataset for SA could be created [12] to find the polarity of free text and classifying the text into one of three categories i.e. Positive, Negative and Neutral. Thus, this feedback is input into the proposed solution to identify and categorize the opinion expressed and output is the polarity of the sentence either positive, negative or neutral.

Our proposed solution will help both the patient and the health care organizations, as the patients' review is being analyzed and listen to by the healthcare organizations in finding out the quality of healthcare services they are providing.

The novelty of our proposed solution is that rather than using Twitter or any other social media platform we have developed a web application where patients can provide feedback about the services provided. Unlike Greaves et al. [1] where user opinion restricted to only two parameters i.e. cleanliness and dignity that could be rated from five given options, we have not restricted the user's opinion and processed free text to make our solution more generic.

4   SYSTEM DESIGN

Our proposed solution named Patient Feedback Platform (PFP) is comprised of five: data gathering, data processing, model engineering, model tuning and last but not least the experimentation and results phase. Our proposed solution is depicted in Figure 1.
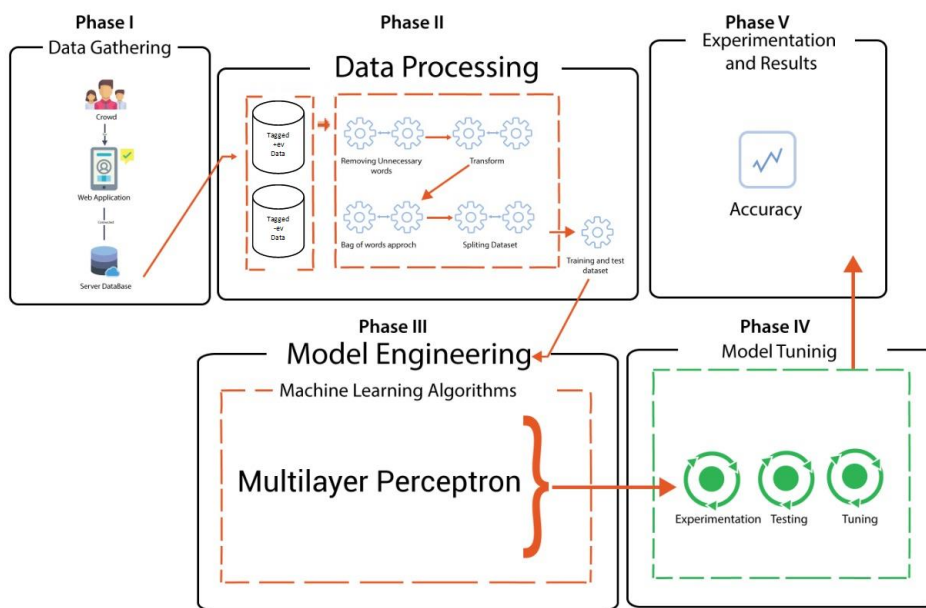
Figure 1: System Design of Patient Feedback Platform (PFP)

In this research study, we have developed a platform to gather feedback from patients. The primary aim of developing the platform is to collect post-interaction details from patients regarding his/her experience with a healthcare provider. These details could range from physicians' efficiency to healthcare service providing the organization's throughput. Furthermore, the platform is also needed to be used as a trustworthy place for any patient to look upon the physician/healthcare providing organization beforehand, based on their reputation observed/reported by their ex-patients. Moreover, the developed platform need to display the physician's information based on patients health concerns, gather feedback for prospective patients, develop insight into post-care for the patient, help mitigate commonly occurring mistakes by the physician and most importantly develop insight from the feedback for better management and training of the staff of the organization. With the unique set of solutions, our proposed solution has to provide, it also came with its challenges, for instance, understanding huge volume data to interpolate better information for the organization was a major part of it.

This analysis of data is required to extract the information from free text to offering and maintaining a high quality of service without increasing staff to decrypt the data. To address this problem, Authors in this research study have used machine learning to analyze the data to develop insight into the physicians' performance, gave acute instructions about the language tone as well as understand what the patient said in the statement. These elements later needed to be collectively analyzed and interpolated both for the physicians' performance enhancement and the organization's training protocols for new and currently employed recruits. Therefore, in essence, the platform to be developed would uniquely satisfy the needs of both the patient and the physician, while furthermore helping the healthcare enterprise use the information for quality assurance and training purposes.

4.1   Phase I: Data Gathering

In phase I of the proposed platform, the data gathering process carried out. For this purpose, in this research work authors have developed a web-based application using an open source content management system called WordPress. While MySQL database management system is used to store the patient's review. The crowd used the application to provide feedback about the physician and healthcare services providing organization. Using this application, we have collected 8000 reviews which are then used for validating the trained model accuracy.

4.2   Phase II: Data Processing

In phase II data processing is performed on the feedback collected in phase 1. When the data is collected from different sources, it is in a raw format that cannot be analyzed using machine learning algorithms. Thus, the collected feedback is also in the form of free text and it needs to be transformed into a more suitable format for the machine to process and train against. In data preprocessing the free text information is smoothen and converted in a more suitable format that the algorithm can fathom. This method is composed of two steps i.e. removing the unnecessary words and

developing a bag of words (BOW). Before proceeding towards the BOW approach, the free text is first refined by removing unwanted words and symbols sometimes also called stop words.   These stop words such as prepositions, articles and conjunctions which occurs frequently in a language have no meaning when standalone thus have no significance for sentiment classification.

4.2.1   Bag of Words Model

Bag-of-words (BOW) model addresses the issues about the occurrence of words inside the document. The rationale for employing the BOW is that it is one of the simplest and easiest methods of extracting choices from the text. Using the bag of words approach, we converted text into the matrix of the incidence of words within a document. To elaborate on how the BOW model is employed in this study, let's consider an example. Suppose that we have three documents with some text as shown in Table 1.

TABLE 1: An example of Beg of Words

Document ID	Text in Document
Doc 1.	The doctor was punctuation.
Doc 2.	The doctor had me waiting.
Doc 3.	The treatment received was exceptionally good.

The next step is developing the Document-Term Matrix (DTM) [13]. DTM is a matrix of documents and words by numeration the incidence of words among the given document. For documents listed in Table 1, DTM is developed as illustrated in Table 2.

TABLE 1: Document term matrix

Document ID		Doc1	Doc 2	Doc 3
Words	The	1	1	1
	Doctor	1	1	0
	Was	1	0	1
	Punctual	1	0	0
	Had	0	1	0
	Me	0	1	0
	Waiting	0	1	0

4.2.2   Generation of Features with Term Frequency – Inverse Document Frequency

Term Frequency (TF) is the output of the BOW model and in TF, the number of words that occurred in each document is counted. The main problem with TF is that it will give more weight to longer records. While the Inverse Document Frequency (IDF) measures the amount of information a given word provides across the document. Equation 1 shows the formulae for measuring the amount of information a given word provides [13].

$$\text{IDF (Words)} = \log \frac{\neq(documents)}{\neq(documents\ containing\ word\ W)} \qquad (1)$$

4.2.3   Splitting the Dataset and Training the Model

To employ machine learning we first have to train the algorithm. For training the algorithm, a dataset is required and this dataset is called a

training dataset. It is worth mentioning here that we have selected multilayer perceptron algorithm for training our model. Once the algorithm is trained, it needs to be tuned before it could be used to perform the classification of the free text on an unseen (by algorithm) dataset. For the training and tuning of the model classifier, we have used UCI Drug Review dataset with 21,5063 uncleaned records [12]. This training dataset consists of 21, 5063 records (70%) and the remaining 30% of the dataset i.e. 21,5063 records are used for tuning the model. The UCI Drug Review dataset contains 5 columns out of which only 3 were relevant. For this particular dataset's pre-processing, we removed the unwanted columns and then tuned the dataset. The result of this stage was a dataset consisting of lowercase words, free from unwanted symbols and numbers.

For the actual classification of free-text reviews, we have used the dataset which we obtained through our web application. This dataset was comprised of 8000 records. Pre-processing discussed in section 3.2 is performed on our collected data.

4.3   Phase III: Model Engineering

The ML model/classifier is trained with 70% of the pre-processed dataset using MLP and tested with the remaining 30%. The evaluation resulted in 88% of the accuracy rate. The patient feedback in the form of user-entered words was transformed into machine understandable language and fed to the classifier as input. The classifier, now trained, gives the prediction of whether the words carry a positive, negative or neutral sentiment.

4.4   Phase IV: Model Tuning

Once the data is prepared and algorithms are trained, the next stage is model tuning, this step is a cycle of experimentation, testing and tuning to optimize the performance of the algorithm and refine the results. As when a model is first trained with the training dataset upon completion then it comes to an iterative process of checking its performance and reevaluating until desired results are achieved.

4.5   Phase V: Experimentation and Results

To evaluate the correctness of our proposed platform as presented in figure 1, we have conducted experiments and analyzed the results. The experiemtns and obtaioned results are discussed in detatail in Section 5.

5   EXPERIEMENT AND ANALYSIS

We employed perceptron multilayer (MLP) machine learning algorithm for classifying the sentiment into positive, negative and neutral categories. A perceptron multilayer (MLP) is a class of feedforward artificial neural network and in this type of neural network, a connection between the nodes does not form a circle. An MLP is made up of at least three node layers: an input layer, a hidden layer, and an output layer. Except for the input node, each node is a neuron using a nonlinear activation function.

we have classified free text in three categorized: positive, negative or neutral polarity. These polarities are represented as +,- and * in Table 2.

Table 2: Free text classification based on polarity

TEXT	Classification
The Doctor was punctual and had a great experience consulting with him.	+
The area and the environment of the clinic was clean	+
I'm sorry but I had a very bad experience and I would recommend consulting with him	-
Why was that doctor so rude	*
Why on earth do you say bad things about the patient's personality	+
Doctor was professional but didn't had sympathy about the patient health condition	-
Had panic attacks social anxiety starting from adolescence. Powered postpartum depression second child started severe mood swings, irritability panic attacks (dizzy spells, heart racing, fainting, chest pain) even I wasn't feeling stress. Started Effexor 37.5 mg almost instant achievement (within week).	-
Accuracy in %	**88**

We have used confusion matrix along with tagged data for calculating the accuracy of our model in this study. The confusion matrix is a table used to describe the performance of a classification model. Besides, the confusion matrix also helps to visualize the true values of the algorithm and the performance of an algorithm. We have calculated the accuracy of our model based on the following formulae depicted as Equation 2.

$$\text{Accuracy} = \frac{\text{rue positive} + \text{true negative}}{\text{true positive} + \text{true negative} + \text{false positive} + \text{false negative}} \quad (2)$$

A comparative analysis of selected contemporary studies as presented in Table 3.

Table 3: Comparison with selected contemporary studies

Selected Study	Our proposed Solution
In [1], User opinion restricted to only two parameters i.e. cleanliness and dignity that could be rated from five given options	We have not restricted the user's opinion and processed free text to make our solution more generic.
In [6], data from RateMDs, a website for rating physician has been used. Rating is only possible based on four categories. Correlates state-level healthcare services in the USA.	We have developed our web application for the collection of patients' feedback. We have not restricted the user's opinion and processed free text to make our solution more generic.

6   CONCLUSION

The main objective of this study was to develop a framework that would process the free-text available online in form of comments given by patients about their experience with healthcare providers. These comments are processed using machine learning algorithm to categorized it into positive,

negative or neutral sentiment. These sentiments, in the context of performance and quality, then represent whether a healthcare provider have done a satisfactory job from patients' perspective. Such performance evaluations are used by many people in the pursuit for quality medical treatment, which helps in encouraging them to select a better choice. Using our proposed Patient's Feedback Platform, it could be predicted with 88% accuracy if a patent will recommend the services or vise versa. This insight could help mitigating challenges about assessing weaknesses and areas requiring further attention and improvement thus the quality of service provided by healthcare providers could be improved.

7   FUTURE WORK

The accuracy of the model can be improved by increasing the volume of the training dataset to obtain more realistic insight into the quality of services provided by a physician/organization. Moreover, a solution could be further improved by further classification of sentiments into categories such as strong positive, positive, neutral, negative, strong negative rather than simply having positive or negative sentiment. Additionally, understanding the statement in terms of what emotions the message is carrying such as anger, disappointment, pleasure, etc. can help in determining their brand loyalty and likely hood of returning. A depression detection algorithm can also be applied to social networking sites, to create a psyche profile of the user, capable of understanding a wide range of psychiatric problems (e.g. detecting a trend to suicidal tendencies or a bipolar disorder, etc.). Such mechanics would help and equip the physician to understand the patients' overall health and have access to data that would've previously taken years to develop through manual labor.

REFERENCES

[1] F. Greaves, D. Ramirez-Cano, C. Millett, A. Darzi, and L. Donaldson, **Use of sentiment analysis for capturing patient experience from free-text comments posted online**, *J. Med. Internet Res.*, vol. 15, no. 11, Nov. 2013, doi: 10.2196/jmir.2721.

[2] J. Bollen, H. Mao, and X.-J. Zeng, **Twitter mood predicts the stock market**, *J. Comput. Sci.*, vol. 2, no. 1, pp. 1–8, 2011, doi: 10.1016/j.jocs.2010.12.007.

[3] M. Walther and M. Kaisser, **Geo-spatial event detection in the Twitter stream**, in *Advances in Information Retrieval*, 2013, pp. 356–367, doi: 10.1007/978-3-642-36973-5_30.

[4] D. Corney, C. Martin, and A. Göker, **Spot the ball: Detecting sports events on twitter**, in *European Conference on Information Retrieval*, 2014, vol. 8416 LNCS, pp. 449–454, doi: 10.1007/978-3-319-06028-6_40.

[5] M. Madhukar and S. Verma, **Hybrid Semantic Analysis of Tweets: A Case Study of Tweets on Girl-Child in India**, *Eng. Technol. Appl. Sci. Res.*, vol. 7, no. 5, pp. 2014–2016, 2017.

[6] B. C. Wallace, M. J. Paul, U. Sarkar, T. A. Trikalinos, and M. Dredze, **A large-scale quantitative analysis of latent factors and sentiment in online doctor reviews**, *J. Am. Med. Inform. Assoc.*, vol. 21, no. 6, pp. 1098–1103, 2014, doi: 10.1136/amiajnl-2014-002711.

[7] O. Ivanciuc, **Weka machine learning for predicting the phospholipidosis inducing potential**, *Curr. Top. Med. Chem.*, vol. 8, no. 18, pp. 1691–709, 2008, doi: 10.2174/156802608786786589.

[8] B. Liu, **Sentiment analysis and opinion mining**, *Synth. Lect. Hum. Lang. Technol.*, vol. 5, no. 1, pp. 1–184, 2012, doi: 10.2200/S00416ED1V01Y201204HLT016.

[9] S. Khatoon, **Real-time Twitter Data Analysis of Saudi Telecom Companies for Enhanced Customer Relationship Management**, *Int. J. Comput. Sci. Netw. Secur.*, vol. 17, no. 2, pp. 141–147, 2017.

[10] P. Smith, **Sentiment Analysis Of Patient Feedback**, Doctoral dissertation, University of Birmingham, 2017.

[11] Parul Pandey, **Simplifying Sentiment Analysis using VADER in Python (on Social Media Text)**. https://www.datacamp.com/community/news/simplifying-sentiment-analysis-using-vader-in-python-on-social-media-zoewd3c40d (accessed Feb. 02, 2020).

[12] **Opinion-Mining-using-the-UCI-Drug-Review-Dataset/drug**, review processed.csv.gz at master · WuraolaOyewusi/Opinion-Mining-using-the-UCI-Drug-Review-Dataset · GitHub. https://github.com/WuraolaOyewusi/Opinion-Mining-using-the-UCI-Drug-Review-Dataset/blob/master/drug review processed.csv.gz (accessed Feb. 02, 2020).

[13] A. Bin Raies, H. Mansour, R. Incitti, and V. B. Bajic, **Combining Position Weight Matrices and Document-Term Matrix for Efficient Extraction of Associations of Methylated Genes and Diseases from Free Text**, *PLoS ONE*, vol. 8, no. 10, p. e77848, Oct. 2013, doi: 10.1371/journal.pone.0077848.