DICTIONARY-BASED DIABETES DISTRESS DETECTION MECHANISM USING FACEBOOK REACTIONS

MARIAN CYNTHIA MARTIN

FACULTY OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY
UNIVERSITY OF MALAYA
KUALA LUMPUR

2020

DICTIONARY-BASED DIABETES DISTRESS DETECTION MECHANISM USING FACEBOOK REACTIONS

MARIAN CYNTHIA MARTIN

DISSERTATION SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF COMPUTER SCIENCE (APPLIED COMPUTING)

FACULTY OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY
UNIVERSITY OF MALAYA
KUALA LUMPUR

2020

UNIVERSITY OF MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: Marian Cynthia Martin

Matric No: WOA160034

Name of Degree: Degree of Master of Computer Science (Applied Computing) Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"): Dictionary-Based Diabetes Distress Detection Mechanism using Facebook Reactions Field of Study: Information Systems

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date:

Subscribed and solemnly declared before,

Witness's Signature Date:

Name:

Designation:

DICTIONARY-BASED DIABETES DISTRESS DETECTION MECHANISM USING FACEBOOK REACTIONS

ABSTRACT

Over the last decade, the internet has paved the way for Facebook to become the digital hub for social networking transforming the ways of sharing information carrying rich and valuable information of users' perspectives. Facebook support groups compromised of online diabetes communities facilitates users to connect to different people with similar conditions and share information wrapped in their own sentiments and emotions. Diabetes being a major life threatening health issue, results in diabetes distress among the online diabetes community. However, the detection of diabetes distress had been carried out manually using surveys and questionnaires, accentuating the lack of studies in automated diabetes distress detection. Hence, this research aims to leverage on information from public Facebook diabetes support pages to extract extended features such as reactions along with posts to build a diabetes distress detection mechanism. An evaluation illustrates that the developed detection mechanism results with 62% accuracy, indicating that the proposed mechanism provides a feasible solution to detect diabetes distress in the online diabetes community. Finally, a comparison was done with the baseline study and the results depict a significant improvement in the overall accuracy of the proposed mechanism.

Keywords: Diabetes distress, Dictionary-based, Facebook, Reactions

MEKANISMA DETEKSI 'DIABETES DISTRESS' BERASASKAN KAEDAH 'DICTIONARY' MENGGUNAKAN 'FACEBOOK REACTIONS'

ABSTRAK

Sepanjang dekad yang lalu, internet telah mencipta laluan bagi Facebook untuk menjadi hab digital untuk rangkaian sosial mengubah cara berkongsi maklumat dan membawa maklumat yang kaya dan berharga tentang perspektif pengguna. Kumpulan sokongan Facebook yang berkompromi dari komuniti diabetes atas talian memudahkan pengguna untuk berhubung dengan orang berbeza dengan keadaan yang sama dan berkongsi maklumat yang mengandungi sentimen dan emosi mereka sendiri. 'Diabetes' menjadi isu kesihatan kritikal yang menyebabkan 'diabetes distress' di kalangan komuniti diabetes atas talian. Walau bagaimanapun, pengesanan 'diabetes distress' telah dijalankan secara manual menggunakan kaji selidik dan soal selidik, menunjukkan kekurangan kajian dalam pengesanan 'diabetes distress' secara automatik. Oleh itu, penyelidikan ini bertujuan memanfaatkan maklumat dari halaman sokongan diabetes Facebook umum untuk mengekstrak ciri-ciri lanjutan seperti 'reaction' bersama dengan 'post' untuk membina mekanisma pengesanan 'diabetes distress'. Penilaian menunjukkan bahawa mekanisma pengesanan dibangunkan menghasilkan ketepatan 62% yang menunjukkan bahawa mekanisme yang dicadangkan menyediakan penyelesaian yang layak untuk mengesan 'diabetes distress' di kalangan komuniti diabetes atas talian. Akhirnya, perbandingan telah dilakukan dengan kajian dasar dan hasilnya menggambarkan peningkatan yang ketara dalam ketepatan keseluruhan mekanisma yang dicadangkan.

Kata Kunci: 'Diabetes distress', 'Dictionary-based', 'Facebook', 'Reactions'

ACKNOWLEDGEMENTS

First and foremost, I would like to express my profound gratitude and respect to several important individuals who have been involved directly and indirectly throughout the completion of this dissertation. I am pleased to take this opportunity to thank Dr. Vimala Balakrishnan for her invaluable guidance and supervision in making this research possible for me.

I would like to express my gratitude to Wandeep Kaur, who opened up the possibility to access the diabetes dataset, as well as guidance and sharing on the specifics of the dataset. Besides that, I would also like to thank the management staff of the University of Malaya (UM) Faculty of Computer Science and Information Technology for their for their cooperation with providing valuable information, assistance and guidance throughout my study.

Last but not least, my deepest thanks and gratitude to my parents, family and friends for their encouragement, constructive suggestions and full mental support in completing the dissertation.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

AUC	:	Area under the curve
DAIC	:	Distress Analysis Interview Corpus
DDDM	:	Diabetes Distress Detection Mechanism
DDDM-B	:	Diabetes Distress Detection Mechanism - Baseline
DDDM-FB_R	:	Diabetes Distress Detection Mechanism with Facebook Reactions
DIWC	:	DAIC Inquiry and Word Count
ED	:	Extreme Distress
LCM	:	Linguistic Category Model
LIWC	:	Linguistic Inquiry and Word Count
LD	:	Low Distress
MD	:	Moderate Distress
ND	:	No Distress
SVM	:	Support Vector Machine
PAID	:	Problem Areas in Diabetes
POS	:	Part of speech
SUDS	:	Subjective Unit of Distress Scale

CHAPTER 1: INTRODUCTION

1.1 Overview

The emergence of Web 2.0 with a more interactive and dynamic web experiences, social media platforms like Facebook, Twitter and Instagram provide a framework for people to share and discuss ideas, experiences and opinions on various topics (Balazs & Velásquez, 2016; Meire et al., 2016; Ravi & Ravi, 2015; Yang et al., 2016). These topics include political issues, food retail, sports, healthcare, etc. (Krebs et al., 2017; Ravi & Ravi, 2015; Yadollahi et al., 2017). Within the past decade, social media platforms have spread globally and witnessed a rapid growth in the number of users, reaching people from various demographic groups, ethnicities and occupations (Perrin, 2015).

Social media has become the digital hub in which users express themselves frequently and naturally (Ortigosa et al., 2014). According to Ortigosa et al. (2014), the number of users who interact with one another via social networks is increasing rapidly. Facebook is the most popular social networking site around the world with an average of over 829 million daily active users (Blachnio et al., 2015). Whilst other social networks focus mostly, as sources of information, Facebook acts as a platform for posting and sharing random messages and for users to speak their minds more naturally (Ortigosa et al., 2014). It provides the environment for knowledge sharing and peer support (Salas Zarate et al., 2017; Wu & Peng, 2015).

The explosion of user-generated contents from these social media platforms contain rich and valuable information of users' perspectives (Yang et al., 2016). They express opinions and perceptions towards a topic of interest, and thus producing a huge range of text-based data (Akter & Aziz, 2016; Balazs & Velásquez, 2016; Liu & Zhang, 2013; Salas Zarate et al., 2017; Yang et al., 2016). This eventually led to the birth of opinion mining, which refers to the statistical analysis of natural language expressions in the detection of specific textual emotions (Bandhakavi et al., 2017; Medhat et al., 2014).

1.2 Facebook Reactions

In a generic posting on Facebook, people appear to be more expressive using written text. However, once the post is shared, it usually gains attention in the form of "likes" rather than long descriptive comments (Pool & Nissim, 2016). Instead of having the ambiguous "like" as the only wordless response to a post, a new collection of more expressive reactions has been introduced, as seen in Figure 1 (i.e. love, haha, wow, sad and angry) (Pool & Nissim, 2016). These reactions are used as alternatives for emotion tags linked to posts which are to express an 'emotion' towards the posted content (Kaur et al., 2018; Krebs et al., 2017; Pool & Nissim, 2016).

In contrast to "likes," graphical features such as reactions enable a user to reflect different emotions in response to a post, such that a post could then be wordlessly labelled as saying "joy" or "surprise" rather than a cliched "like" (Kaur et al., 2018; Pool & Nissim, 2016). This representation of emotion via Facebook reactions will be beneficial to improvise text classification and opinion mining (Kaur et al., 2018). The following sub-sections discuss further on emotion analysis and diabetes distress.

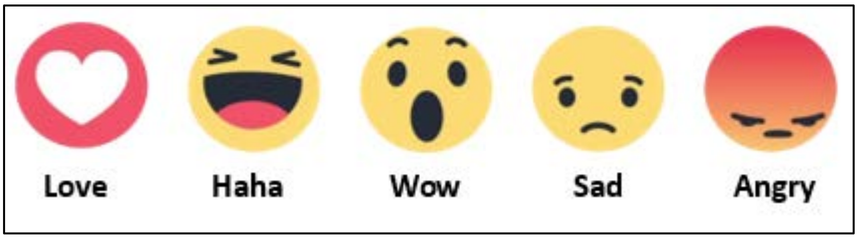


Figure 1.1 Facebook Reactions

1.3 Emotion Analysis

Opinion mining is the field of research that analyses people's views, feelings, judgments, and emotions towards various elements and their features (Liu & Zhang, 2013). According to Medhat et al. (2014), opinion mining is described as the computational study of the opinions, evaluations, behaviours and emotions of individuals towards organizations, individuals, challenges, activities, subjects and characteristics of individual people. There are many fields related to opinion mining including sentiment analysis, emotion analysis, etc. (Medhat et al., 2014).

In other words, analysis of sentiments involves the retrieval of opinions in order to extract and categorize opinions from certain documents (Abirami & Gayathri, 2017; Liu & Zhang, 2013; Salas Zarate et al., 2017). Typically the polarity of sentiment defined in terms of positive or negative opinion (Abirami & Gayathri, 2017; Medhat et al., 2014). In addition, sentiment polarity can be associated with emotion analysis like sad, anger, fearful, happy, etc. The strength of sentiment opinions are interconnected to the intensities of certain emotions, i.e., positive sentiment related to joy, trust, etc. and negative sentiment related to anger, sadness, fear, etc. (Liu & Zhang, 2013).

Emotion is any conscious experience in which personality, emotional state, temperament and determination are interconnected. It plays a vital role in influencing human behavior in which rationale, decision-making and interactions are affected (Bandhakavi et al., 2017). Emotion analysis is an opinion mining task that concerns the computational study of natural language expressions in recognizing various emotions from text (Bandhakavi et al., 2017; Medhat et al., 2014). Analysis of emotions aims to illustrate the text's content in different emotions like joy, anger, fear, surprise and sadness (Medhat et al., 2014). It can be administered on text, emoticons, reactions, facial expressions and audios, although the common ones are text and audio (Bhaskar et al., 2015). Emotion analysis has been performed on various domains such as political science (Ravi & Ravi, 2015; Yadollahi et al., 2017), retail businesses (Krebs et al., 2017), sports (Meire et al., 2016), education (Ortigosa et al., 2014) and healthcare (Salas Zarate et al., 2017; Wu & Peng, 2015; Yang et al., 2016), etc.

An example of a post with the sentiment and emotion is as below:

Post: "I wished I had. Now I have more issues. Never knew it would affect my teeth til my dentist told me almost ten years ago. Gotta get the rest of a tooth pulled. Been married almost six years and no kids yet. If I had only listened when I was younger."

Emotion= Sadness

Sentiment polarity= Negative

In healthcare domain, the rapid development of the Internet has resulted more users to share their medical stories and experiences or interact with other people in the online health communities (Yang et al., 2016). Generally, people suffering from chronic illnesses will have periodic contacts with healthcare professionals, but they also need to have the skills, attitude, and support for self-management of their condition (Berry et al., 2015; Fisher et al., 2015). Therefore, social networks such as Facebook are an excellent resource for the patients since it helps to build a bridge to connect different people who have similar conditions and experiences (Salas Zarate et al., 2017). The sharing of information by patients, enclosed in their own thoughts and feelings, is the driving factor in the analysis of sentiments and emotions (Agarwal et al., 2018; Salas Zarate et al., 2017). One medical condition with increased life-threatening health problems is diabetes which results in higher medical costs, reduced quality of life and increased mortality (Cho et al., 2018). The following sub-section discusses on Diabetes Mellitus, online diabetes community and the role of social media in this community.

1.4 Diabetes Distress

Diabetes Mellitus or simply known as diabetes relates to a group of metabolic disorders identified by high blood glucose levels over an extended span of time (Cho et al., 2018). There are several types of diabetes, some of which are more common than the others. The types of diabetes include Type I, Type II and Gestational diabetes (Cho et al., 2018; Fisher et al., 2015). According to Cho et al. (2018), globally 451 million people aged 18 to 99 are diagnosed with diabetes and this number is likely to rise more in the future. Diabetes is a complex chronic disorder that requires continuous medical care and patient self-management for control of abnormal glucose levels to prevent or minimize acute and long-term complications such as a kidney failure, heart attack and stroke (Cho et al., 2018; DeFronzo et al., 2015).

Diabetes distress is a coherent reaction to the menace of a life changing illness. Negative emotions such as fear, concerns and worries are normally experienced by diabetic patients due to the burden of potential complications in relation to Diabetes condition (Fisher et al., 2015). Distinct from depression, diabetes distress is typically entrenched with the stipulation of diabetes management and is a result of emotional reconciliation (Berry et al., 2015). It is the emotional strain often linked to the health issues and its complications and health impacts caused by poor lifestyle and health management (Berry et al., 2015; Dieter & Lauerer, 2018; Fenwick et al., 2018; Fisher et al., 2015; Lašaite et al., 2016; Schmitt et al., 2015; Sturt et al., 2015). According to Snoek et al. (2015), extreme diabetes distress exist in 10 – 30% diabetic patients, subjected to case mix and different countries.

Generally, people with diabetes have an increased risk of complications such as major life-threatening health problems in relation to higher medical care costs, reduced quality of life and increased mortality implicating presence of diabetes distress (Cho et al., 2018;

Fisher et al., 2015). Facebook support pages also composed of a community of online diabetes such as patients, caregivers support networks, and healthcare professionals. There is a high level of diabetes distress in this community implying that there is a need to address diabetes distress (Fisher et al., 2015). To aid this study, Facebook is used as the online media platform to extract relevant data from public diabetes support pages. Facebook was chosen due to its massive number of users where huge data are daily generated (Blachnio et al., 2015; Ortigosa et al., 2014; Wu & Peng, 2015).

The following sections presents research problems, research objectives, research questions, research scope and organization of thesis.

1.5 Research Problems

One of the problems faced is the lack of automated diabetes distress detection mechanism. Automated diabetes distress detection mechanism will be a tool to detect the level of distress present in text. An automated distress detection mechanism could aid in the prevention of future burden of diabetes distress, and this is necessary for allocating community and health resources to create strategies to counteract these rising trends of increasing health complication (Cho et al., 2018). The current tool for diabetes distress assessment and detection is in the form of surveys and questionnaires (Benjamin et al., 2010; Dieter & Lauerer, 2018; Fenwick et al., 2018; Sturt et al., 2015). These surveys are focused on clinical care and could be diagnosed depending on patients' acknowledgment and participation (Dieter & Lauerer, 2018; Fenwick et al., 2018). With the aid of internet, patients are more actively expressing themselves and sharing their opinions via social media (Fisher et al., 2015). This information could be analyzed for the presence of diabetes using automated distress detection mechanism. Surveys such as Problem Areas in Diabetes (PAID) and Diabetes Distress Scale (DDS) are more commonly used to collect relevant information for diabetes distress detection and analysis compared to automated detection (Benjamin et al., 2010; Dieter & Lauerer, 2018; Fenwick et al., 2018; Sturt et al., 2015).

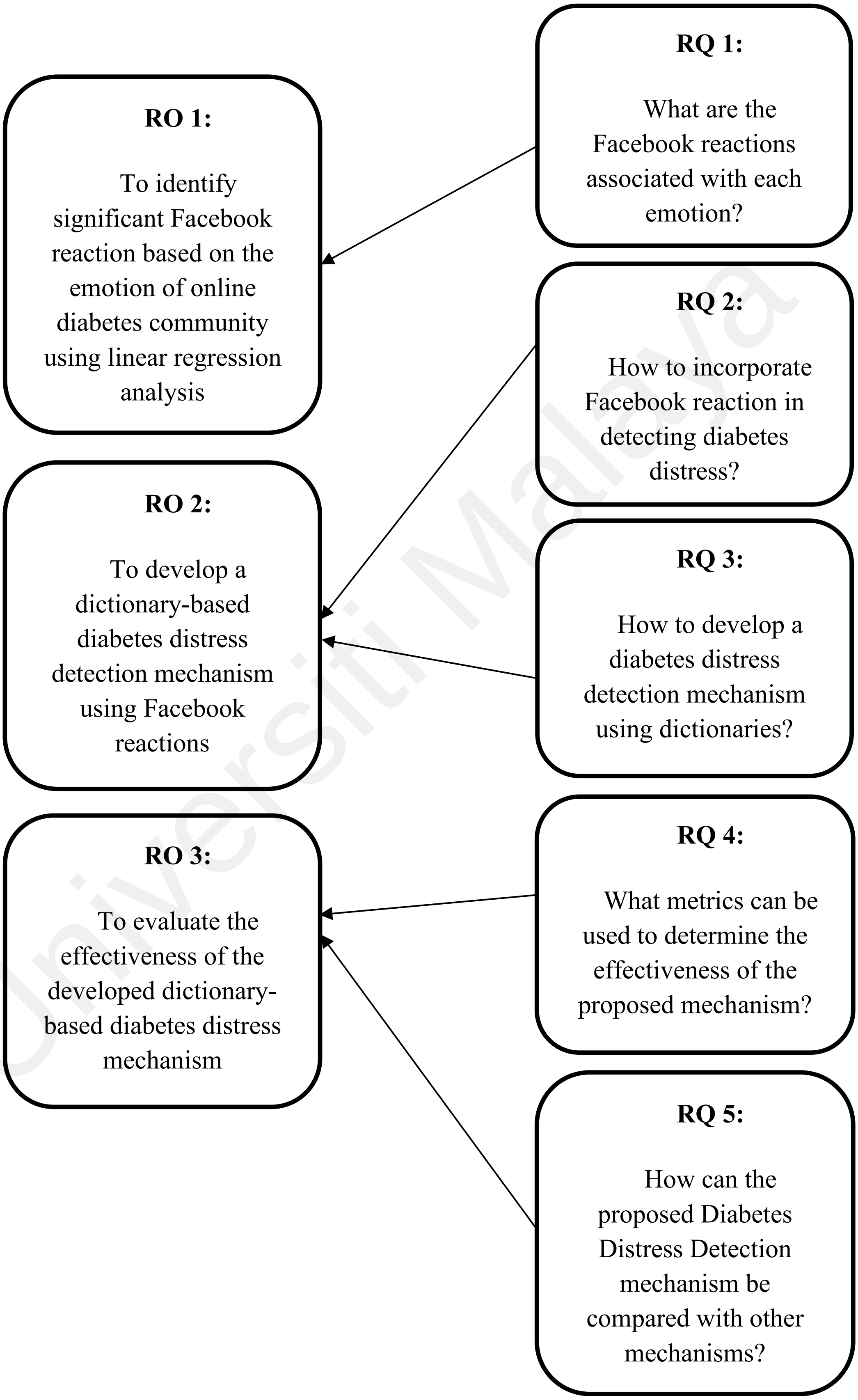
Furthermore, the dictionary-based approach is widely used in emotion analysis (Mandal & Gupta, 2016; Seih et al., 2016; Settanni & Marengo, 2015) but has not been applied in distress detection mechanism in recent studies. The current studies, such as (Dieter & Lauerer, 2018; Fenwick et al., 2018; Gahlan, Rajput, Gehlawat, & Gupta, 2018; Schinckus et al., 2017) have all been based on questionnaire surveys for which PAID and DDS were often the prevalent survey approach. This created a space to study the effectiveness of a dictionary-based approach by integrating Facebook reactions as predictors to diabetic distress detection, which have not been done to the best of knowledge.

Besides, emotion detection is conducted mainly on the main Facebook posts, where extra features such as ' likes, ' ' reactions ' and ' comments ' are not commonly used as they are considered as noise (Meire et al., 2016). According to Meire et al. (2016), using these features gives an improvised predictive performance. In another study, a framework for predicting the Facebook post reaction distribution based on a customer service dataset from several supermarket Facebook posts was developed to enhance customer experience analytics with results revealing that Facebook features improves emotion classification (Krebs et al., 2017).

Therefore, the study aims to investigate and develop an automated technique using dictionary-based approach and Facebook reactions in the context of building an automated diabetes distress detection mechanism. This research, unlike the existing ones, takes advantage of the common dictionary-based approach by using regressed Facebook reactions as independent variables for the distress detection mechanism.

1.6 Research Objectives and Research Questions

This section presents the research objectives and research questions.



The first objective is to identify significant Facebook reaction based on the emotion of online diabetes community. Studies have shown that including these features in the analysis improves the overall prediction (Kaur et al., 2018; Krebs et al., 2017; Meire et al., 2016). The graphical representation using Facebook reactions will be advantageous to improvise distress level classification by incorporating with text analysis. This objective was achieved using Linear Regressions to identify the significant Facebook reaction.

The second research objective is to develop a dictionary-based diabetes distress detection mechanism. Limited studies had been carried out to automate the process of distress detection in an online diabetes community (Benjamin et al., 2010; Dieter & Lauerer, 2018; Fenwick et al., 2018; Sturt et al., 2015). Dictionary Inquiry and Word Count technique was used to develop an automated diabetes distress detection mechanism which could accelerate the process of distress detection.

The third objective is to evaluate the effectiveness of the proposed diabetes distress mechanism against human annotated results to test the reliability and effectiveness of the proposed approach. This step is crucial to evaluate and compare the baseline mechanism with the improvised mechanism to see if the accuracy of the mechanism has improved (Kaur et al., 2018).

1.7 Research Scope

This study focused on the development of diabetes distress detection mechanism using dictionary-based approach and how the approach can be improved by integrating significant Facebook reactions.

The dataset was Facebook posts collected from six diabetes-related groups which have been active with full user participation, with an average of 42 posts per day since 2014 (Kaur et al., 2018). This dataset was taken from an earlier study in which public posts related to diabetes were extracted from Facebook using Graph API3 over six months, from July 2016 to January 2017 (Kaur et al., 2018). Besides posts and comments, additional features such as the number of likes, reactions, comments and shares were collected (Kaur et al., 2018).

The scope of this study is limited to only Facebook post tagged with reaction looking at one of the objectives of the study; which are to identify significant Facebook reaction based on the emotion of online diabetes community using linear regression analysis.

1.8 Organization of Thesis

The thesis is structured and organized into 5 chapters.

Chapter 1 discussed about the background of the study followed by the problem statement, research objectives and research questions. The chapter ends with the organization of the thesis.

This chapter describes the different methodologies and approaches that can be applied for emotion classification and diabetes distress level detection. The literature review includes discussion on machine learning and lexicon-based approaches. Definition and challenges of emotion analysis and distress detection are introduced and commonly used algorithms for classification are discussed.

This chapter provides the information about the data used for training and testing the classifiers, also importance of data pre-processing is explained. Moreover, applied algorithms for emotion analysis and diabetes distress detection mechanism is explained.

Chapter 4 presents the experimentations and results achieved in this research work based on the obtained results, analysis is performed and compare the performance of the algorithm used for Diabetes Distress Level Mechanism.

The final chapter provides the conclusion, limitation and future research perspectives. Factors affecting the performance, evaluations and challenges are also provided in this chapter.

CHAPTER 2: LITERATURE REVIEW

2.1 Overview

This chapter covers all aspects of the various approaches that can be used to evaluate emotions as well as brief descriptions of algorithms used by researchers. Diabetes Mellitus is introduced in Section 2.2, followed by Diabetes Distress in Section 2.3. The role of social media is explained in Section 2.4 along with social media and diabetes in Section 2.5. Following that, Section 2.6 and 2.7 discuss emotion analysis method and diabetes distress detection techniques respectively. Section 2.8 explains Dictionary-based distress analysis and emotion analysis techniques. Finally, Section 2.9 is the overall summary for Chapter 2.

2.2 Diabetes Mellitus

Diabetes Mellitus, commonly known as diabetes is a metabolic disorder that results in high blood glucose level (Alotaibi et al., 2017; Cho et al., 2018) . Diabetes is one of the largest global health concerns with almost 451 million people (age 18-99 years) diagnosed with diabetes worldwide (Cho et al., 2018). According to Cho et al (2018), the finding reveals this number to increase to 693 million by 2045. About 2.5 million individuals in Malaysia are diagnosed with diabetes (Al-Naggar et al., 2017).

Long term diabetes lead to an increased risk of developing various life-threatening health problems including retinopathy, nephropathy, neuropathy and macrovascular complications such as heart attack and stroke (Al-Naggar et al., 2017; Cho et al., 2018). These chronic complications generally require continuous medical care and patient self-management for the control of high glucose level (Al-Naggar et al., 2017).

2.2.1 Types of Diabetes

The common types of diabetes are type 1, type 2 and gestational diabetes. Type 1 diabetes comprises 5% to 10% of diabetes population and is caused by auto-immune condition where the pancreas is attacked by one's own antibodies (Kharroubi & Darwish, 2015). Type 1 is profoundly diagnosed among children and young adults with severe dependence on insulin treatment to sustain (DeFronzo et al., 2015; Kharroubi & Darwish, 2015).

Type 2 diabetes is the most common among the three major types of diabetes which occurs in middle-ages and older population (DeFronzo et al., 2015; Kharroubi & Darwish, 2015). Type 2 is due to a low insulin production caused by an impaired pancreas (DeFronzo et al., 2015). The root of the type 2 is a combination of genetic and lifestyle factors (Kharroubi & Darwish, 2015).

Gestational diabetes is the result of high blood glucose during pregnancy with an increased risk of developing type 2 diabetes later in life (Kharroubi & Darwish, 2015). Risk of gestational diabetes occurs among pregnant women with an overweight issue, genetic history of diabetes, history of polycystic ovarian syndrome, age etc. (Kharroubi & Darwish, 2015).

2.3 Diabetes Distress

Diabetes distress is the emotional retaliation to the burden of life-threating illness (Berry et al., 2015; Lašaite et al., 2016). Diabetes distress is a normal response associated with agitation and anxiety about developing other complication in relation to diabetes (Fisher et al., 2015). This response is often associated with poor glycemic control, increased complications, decreased life expectancy, interpersonal issues, eating distress, powerlessness, negative social perceptions, relationships with caregivers and healthcare professionals (Berry et al., 2015; Dieter & Lauerer, 2018; Fisher et al., 2015; Lašaite et al., 2016). Concurrent diabetes condition and distress could negatively affect life quality with augmented complications, and shriveled lifespan (Fisher et al., 2015).

In a study to explore gender and age difference in diabetes distress by Lašaite et al. (2016), it was found that type 1 diabetes patients have higher implication of diabetes distress in adulthood compared to adolescence. Similar findings were discovered in Fisher et al. (2015), where seven major sources of diabetes distress were identified with high prevalence of diabetes distress in women. These findings advocate the need to address diabetes distress in clinical care for diabetes distress management (Fisher et al., 2015; Lašaite et al., 2016).

Another study on the necessity for enhanced acknowledgement on diabetes distress, found 50% of patients were undiagnosed and untreated despite the availability of screening tools (Dieter & Lauerer, 2018). Consistent mental health screening is also vital for proper management of diabetes distress to improve the quality of life in patients (Dieter & Lauerer, 2018).

The next sub-sections will introduce and discuss the conventional techniques used for the assessment of diabetes distress detection. Following this, the role of social media in the online diabetes community and the features of the selected social media platform are explained.

2.4 Diabetes Distress Detection

To the best of our knowledge, studies assessing diabetes distress to date were all based on questionnaire surveys (Beiter et al., 2015; Berry et al., 2015; Gahlan et al., 2018; Schinckus et al., 2018; Sidhu & Tang, 2017). Screening or assessment tools used in diabetes distress detection includes Problem Areas in Diabetes (PAID) questionnaire (Berry et al., 2015; Dieter & Lauerer, 2018; Fenwick et al., 2018; Sturt et al., 2015), Diabetes Distress Scale (Gahlan et al., 2018; Sidhu & Tang, 2017), Diabetes Self-Management Questionnaire (Schinckus et al., 2018) and World Health Organization Five Item Well Being Index (WHO-5) (Dieter & Lauerer, 2018).

However, two commonly used questionnaires for diabetes distress assessment are Problem Areas in Diabetes (PAID) and the Diabetes Distress Scale (Fenwick et al., 2018; Sturt et al., 2015). The PAID questionnaire is a self-report tool designed to measure and identify emotional difficulties of diabetes distress in relation to living with diabetes (Dieter & Lauerer, 2018). Previous studies showed PAID questionnaire to have a consistent reliability and validity with a strong correlation with emotional distress, depressive symptoms, disordered eating, fear of hypoglycemia, etc. (Dieter & Lauerer, 2018; Fenwick et al., 2018). PAID is a complete 20-item of six-point scale questionnaire focusing on emotional concerns, diet and diabetes complications (Berry et al., 2015; Fenwick et al., 2018).

Diabetes Distress Scale (DDS) is a revision of PAID to address discern limitations of PAID such as lack of addressing respondents' reaction on items such as healthcare professionals (Fenwick et al., 2018) and the care provided to them (Berry et al., 2015). Typically, diabetes distress has been identified and diagnosed using the survey and questionnaires such as PAID and DDS scales more extensively, thus no further studies on automated diabetes distress detection mechanism has been made (Berry et al., 2015; Dieter & Lauerer, 2018; Fenwick et al., 2018; Sturt et al., 2015).

2.5 Social Media

The endless growth and usage of Internet leads to a more interactive and dynamic web experiences with social media platforms like Facebook, Twitter and Instagram, which provides a framework for people to share and discuss ideas, experiences and opinions on various topics (Perrin, 2015; Pool & Nissim, 2016; Ravi & Ravi, 2015; Salas Zarate et al., 2017; Yang et al., 2016). These topics include political issues, food retail, sports, healthcare, etc. (Krebs et al., 2017; Ravi & Ravi, 2015; Yadollahi et al., 2017).

Within the past decade, social media platforms have spread globally and witnessed a rapid growth in the number of users, reaching people from various demographic groups, ethnicities and occupations (Perrin, 2015). Social media has become the digital hub in which the users express themselves frequently and naturally (Ortigosa et al., 2014). According to Ortigosa et al. (2014), the number of users interacting with others through social networks is growing exponentially.

The explosion of user-generated contents from these social media platforms contain rich and valuable information of users' perspectives (Yang et al., 2016). They express opinions and perceptions towards a topic of interest, producing massive volume of data in the form of text, thus resulting opinion mining to become an active domain for knowledge extraction (Balazs & Velásquez, 2016; Blachnio et al., 2015; Salas Zarate et al., 2017; Yang et al., 2016).

The following sub-sections address the role and involvement of Facebook posts and additional features such as reactions in opinion mining activities.

2.5.1 Facebook

Facebook is the more popular social networking site around the world with an average of over 829 million daily active users (Blachnio et al., 2015). Whilst other social networks focus mostly as sources of information, Facebook serves as a platform to share and communicate where messages in Facebook are spontaneous and users express their emotions more naturally, and thus providing the environment and the tools for knowledge sharing and peer support (Ortigosa et al., 2014; Salas Zarate et al., 2017; Wu & Peng, 2015). Facebook delivers a distinct advantage for this research in addition to its popularity, which is the volatile content dense with emotions and sentiments for an opinion mining activity (Ortigosa et al., 2014).

In recent times, a growing number of studies have examined the relationship between Facebook user behaviors and behavioral constructs, with variations in personality being one of the most active research areas (Settanni & Marengo, 2015). Significant connections between Facebook activities and psychological well-being factors were also identified, hence reflecting the possible use of Facebook-related data to identify individuals with associated risk profiles, such as those Facebook users at risk of depression (Ortigosa et al., 2014; Settanni & Marengo, 2015).

One of several works that studied opinion mining using Facebook data, was the one presented in Settanni and Marengo (2015) which evaluated the connection between emotion-related linguistic factors and initiatives of self-report on the emotional well-being of Facebook users. The findings showed the existence of strong correlations between emotional categories and the emotional well-being of users (Settanni & Marengo, 2015). Negative emotional expression was found to be positively correlated with symptoms of anxiety, depression and stress. A strong correlation between the expression of sadness and all the aspects of mental distress was also observed in this study.

Another opinion mining study using data from Foodbank, a well-known group on Facebook for food market basket analysis (Akter & Aziz, 2016). This study aimed to determine a particular product or event's recent market trend or market value based on the interest of people. Two methods were used, namely Naïve Bayes and lexicons. The results showed Naïve Bayes to perform poorly as the Facebook posts were extremely noisy. The lexicon-based method in predicting sentiment were shown to be effective compared to the machine learning method (Akter & Aziz, 2016).

2.5.2 Facebook Reactions

In addition to the written text post, Facebook incorporated five additional options (i.e. love, haha, wow, sad and angry) to the ' Like ' button (Kaur et al., 2018). In this study, the authors used Facebook features (like, share, comment and reaction) to improve sentiment classification accuracy by including these features as weights in addition to the frequently used emojis to express an 'emotion' towards the posted content (Kaur et al., 2018).

According to Pool and Nissim (2016), reaction features in Facebook allow users to react to the initial post to express how they feel rather than using the 'Like" button which expresses minimal emotion (Pool & Nissim, 2016). This representation of emotion via Facebook reaction feature will be beneficial to aid the present study on the effect of reactions towards opinion mining.

In a study by Meire et al. (2016), the added value of available information before (i.e. leading information) and after (i.e. lagging information) the development period of the post in Facebook sentiment analysis were analyzed to evaluate most relevant predictors and to investigate the relationship between key predictors and sentiment. Leading information are the details available even before the main post is shared (e.g. user profiles, previous posts) whereas lagging information are generated after the main content has been posted (e.g. interactions such as likes, comments or reactions). The findings showed that Model 1 (i.e. primary content of the post) obtained an average AUC of 0.751, whereas Model 2 (i.e. primary content of the post and leading information) scored an average AUC of 0.775. Meanwhile, Model 3 (i.e. primary content of the post, leading information and lagging information) scored the highest average AUC of 0.812 (Meire et al., 2016). The results clearly demonstrate that leading and lagging information add predictive value to the models of sentiment analysis that have been developed.

In the next section, the relationship between social media platforms and the mode of interaction among healthcare community along with its impact towards diabetes distress is explained.

2.6 Social Media and Diabetes Distress

The emergence of social media has resulted in more users to share their medical stories and experiences or interact with other people in the online health communities (Yang et al., 2016). People with chronic conditions will usually have regular contact with their health care professionals, but they also need the skills, behaviors and support to control their condition on their own (Berry et al., 2015; Fisher et al., 2015). Therefore, social networks like Facebook have therefore become an incredible platform for patients, as it helps build a bridge to connect different people of similar problems and similar experiences. (Salas Zarate et al., 2017).

One medical condition with increased life-threatening health problems is diabetes which results in higher medical care costs, reduced quality of life and increased mortality (Cho et al., 2018). Facebook support pages are often comprised of online diabetes community such as patients, caregivers, support groups and healthcare professionals. The presence of diabetes distress is noticed to be high in this community, arguing for a need to address diabetes distress (Fisher et al., 2015). A study among diabetes related Facebook pages to identify features on user engagement revealed messages with images had higher likes and shares in Facebook (Rus & Cameron, 2016). This finding signified the communication pattern in healthcare related social media usage.

The next sub-sections will introduce as well as discuss different opinion mining and text analysis techniques in depth.

2.7 Emotion Analysis

Sentiment analysis can be considered as an opinion mining task where it focuses in specifying positive or negative opinions, but emotion analysis is concerned with detecting various emotions from text (Medhat et al., 2014). Opinion mining has been studied since the 90s (Abirami & Gayathri, 2017; Liu & Zhang, 2013), however, with the explosive growth of social media the opinionated data has pushed research in this area to a new stage (Liu & Zhang, 2013).

According to Liu & Zhang (2012),

"Sentiment analysis or opinion mining is the computational study of people's opinions, appraisals, attitudes, and emotions toward entities, individuals, issues, events, topics and their attributes."

In other words, opinion mining is the process of computationally identifying and categorizing opinions expressed in a piece of text, especially in order to determine the degree of emotion or polarity towards a particular entity. The polarity of sentiment is usually expressed in terms of positive or negative opinion (Liu & Zhang, 2012; Medhat et al., 2014). However, it can also be a multi-class classification (Medhat et al., 2014), hence sentiment may have a neutral label or even broadened variation of labels like very positive, positive, neutral, negative, very negative, also labels can be associated with emotions like sad, anger, fearful, happy, etc.

Sentiment analysis aims at recognizing positive, neutral or negative feelings from text, while an emotion analysis aims to identify and distinguish types of feelings through the transmission of texts such as anger, disgust, fear, joy, sadness and surprise. (Medhat et al., 2014). Emotion analysis may benefit in situations such as determining how happy people are pertaining different factors such as environmental, health, economic or social factors.

The primary emotions are love, joy, surprise, anger, sadness, and fear where each emotion can also have different intensities (Liu & Zhang, 2013; Medhat et al., 2014; Rout et al., 2017). Generally, sentiment reflects feeling or emotion while emotion reflects attitude (Medhat et al., 2014). As an opinion mining task, emotion analysis can be implemented using Machine Learning approach or Lexicon-based approach, but Lexicon-based approach is more frequently used (Medhat et al., 2014).

Emotion analysis plays a significant role in our daily decision-making process. These decisions may range from purchasing a product such as mobile phone to reviewing a movie to making investments. Nowadays, people are seeking feedbacks on the products before buying any product or service in the market. Emotion analysis is a developing area that increases the interest of humans and especially organizations because it can be used for decision making (Liu & Zhang, 2013). Individuals are no longer restricted to having to ask friends ' views about a specific good or service; such information can be openly found online. Furthermore, organizations may save time and money by avoiding surveys; instead they can concentrate on processing the opinions that can be obtained from the Web freely. Nevertheless, it should be noted that sources containing opinions are usually noisy, so it is crucial to obtain the essential meaning for further application and analysis from this information (Abirami & Gayathri, 2017).

Bandhakavi et al. (2017) developed a domain specific emotion lexicon-based technique on a generational unigram mixture model. This technique derived labeled texts from blogs, news headlines and incident reports and weakly labeled emotional text from tweets and developed a technique to extract appropriate emotional classification features (Bandhakavi et al., 2017). The findings confirm that the features derived from the proposed lexicon outperform those from state-of-the-art lexicons (Bandhakavi et al., 2017).

In another study, SVM was used by Bhaskar et al. (2015) where text and audio features were combined to form a single function vector and then fed to the classifier as input features. Lexicon approaches were used to derive the emotions from the text while multiclass SVM was used for emotion classification (Bhaskar et al., 2015). The results showed that the reliability and precision of the hybrid approach were significantly higher with an accuracy score of 90% compared to 57.1% and 76% for the individual approach (Bhaskar et al., 2015).

2.8 Dictionary-based Distress Analysis

The basic steps in this approach is the collection of a small set of opinion words with known orientations (Liu & Zhang, 2013; Medhat et al., 2014). This set is then bootstrapped using online dictionaries such as WordNet or thesaurus to search for their synonyms and antonyms. The newly found words will be added to a seed list, and this process stops when new words are exhausted.

Mandal and Gupta (2016) used the dictionary-based text classification approach to analyze and predict the polarity of users' sentiments from online reviews. In addition, the authors have used negation words to convey how the program's accuracy can be enhanced. The finding shows that the dataset without negation scored an accuracy of 86.5% whereas the dataset with negations obtained an accuracy of 97.1% (Mandal & Gupta, 2016). The model achieved a better accuracy when the algorithm used the pool of words of negation to identify text compared to the model that disregarded negations.

Example of similar approach includes the work of Seih et al. (2016) who developed a computerized text analysis method for the Linguistic Category Model (LCM) using three different types of verb categories to describe emotional or mental states, to characterize more general behaviour and to describe a specific and observable action. The authors had contrasted two writing tasks, namely writing in viewpoints of first person and third person (Seih et al., 2016). Text written from a third-person's perspective had higher LCM scores than text written in a first-person's perspective. These results indicated that typically while writing in a third person's viewpoint, more abstract or concrete words were used (Seih et al., 2016).

On the other hand, Settanni and Marengo (2015) proposed a Linguistic Inquiry and Word Count (LIWC) technique to predict emotion related textual indicators via data collected from an online questionnaire in Facebook. Following this, the authors performed a correlation analysis to identify individuals with higher levels of depression and anxiety level. Findings revealed strong associations between the emotion-related categories of LIWC and the psychological well-being of users. As a whole, negative emotional expressions correlated positively with symptoms of anxiety, depression and stress (Settanni & Marengo, 2015).

LIWC is an analysis tool used to provide insights and calculate percentage of a text-corpora on a word-by-word basis that falls within various domains such as emotion, cognitive, social etc. (Dehghani et al., 2017; Firmin et al., 2016; Settanni & Marengo, 2015). One prominent advantage of LIWC include the feature that allows users to develop their own dictionaries for analyzing any given dimension of language use (Seih et al., 2016).

In this research, the advantage of LIWC technique will be leveraged to develop the Diabetes Distress Detection Mechanism based on the Distress Analysis Interview Corpus-Wizard of Oz (DAIC-WOZ) as dictionary and Subjective Units of Distress Scale (SUDS) (Benjamin et al., 2010; Kiyimba & O'Reilly, 2017; Ringeval et al., 2017). The technique is adapted and developed based on the Linguistic Inquiry and Word Count (LIWC) technique, which is one of the most used technique by data scientists for text analysis (Dehghani et al., 2017; Firmin et al., 2016; Settanni & Marengo, 2015). The adapted technique is named DAIC Inquiry and Word Count (DIWC) with reference to the dictionary to be used to develop the diabetes distress detection mechanism.

The Distress Analysis Interview Corpus (DAIC) (Parekh & Patil, 2017; Rana et al., 2019; Ringeval et al., 2017) includes a database built using semi-structured clinical interviews of participants to enable the diagnosis of psychological distress conditions such as depression, anxiety and post-traumatic stress disorder. A sample is shown in Figure 2.1 below.

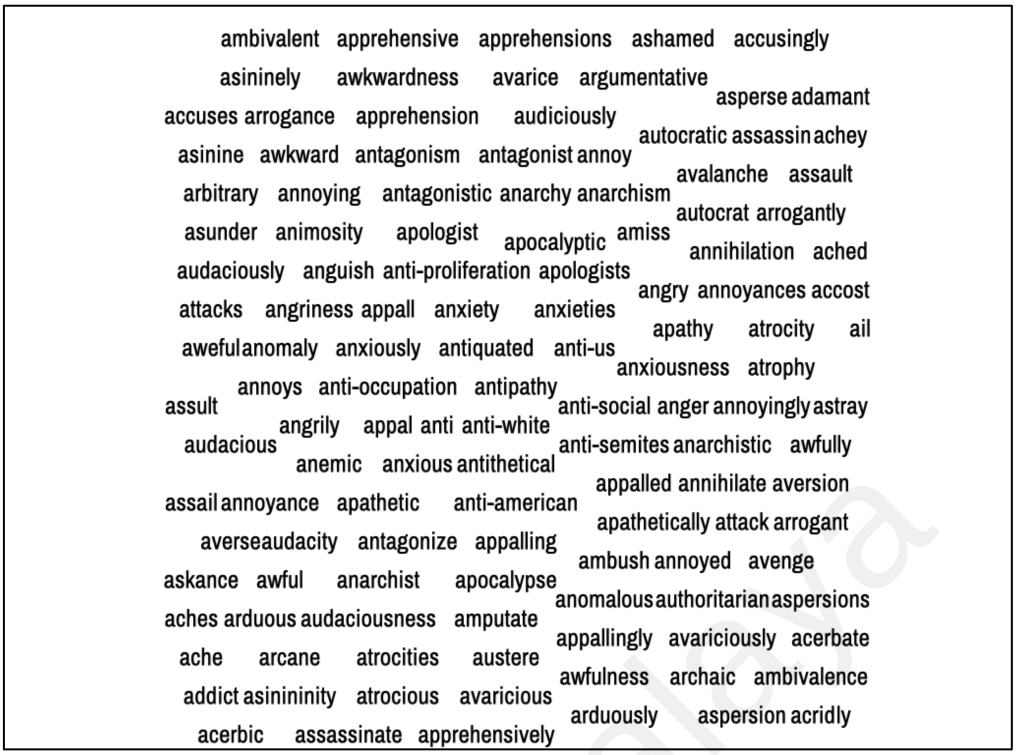


Figure 2.1 Sample words in Distress Analysis Interview Corpus

In this current research, the DAIC database will be used as the dictionary in DAIC Inquiry and Word Count (DIWC) technique to aid the development of automated diabetes distress detection mechanism and Subjective Unit of Distress Scale (SUDS) will be used to classify the detection into 4 levels of 'No Distress', 'Low Distress', 'Moderate Distress' and 'Extreme Distress' (Benjamin et al., 2010; Kiyimba & O'Reilly, 2017). SUDS will be used to determine the subjective intensity of disturbance or distress currently experienced by an individual (Benjamin et al., 2010; Kiyimba & O'Reilly, 2017). A sample of SUDS distress level breakdown is shown in Figure 2.2 below.

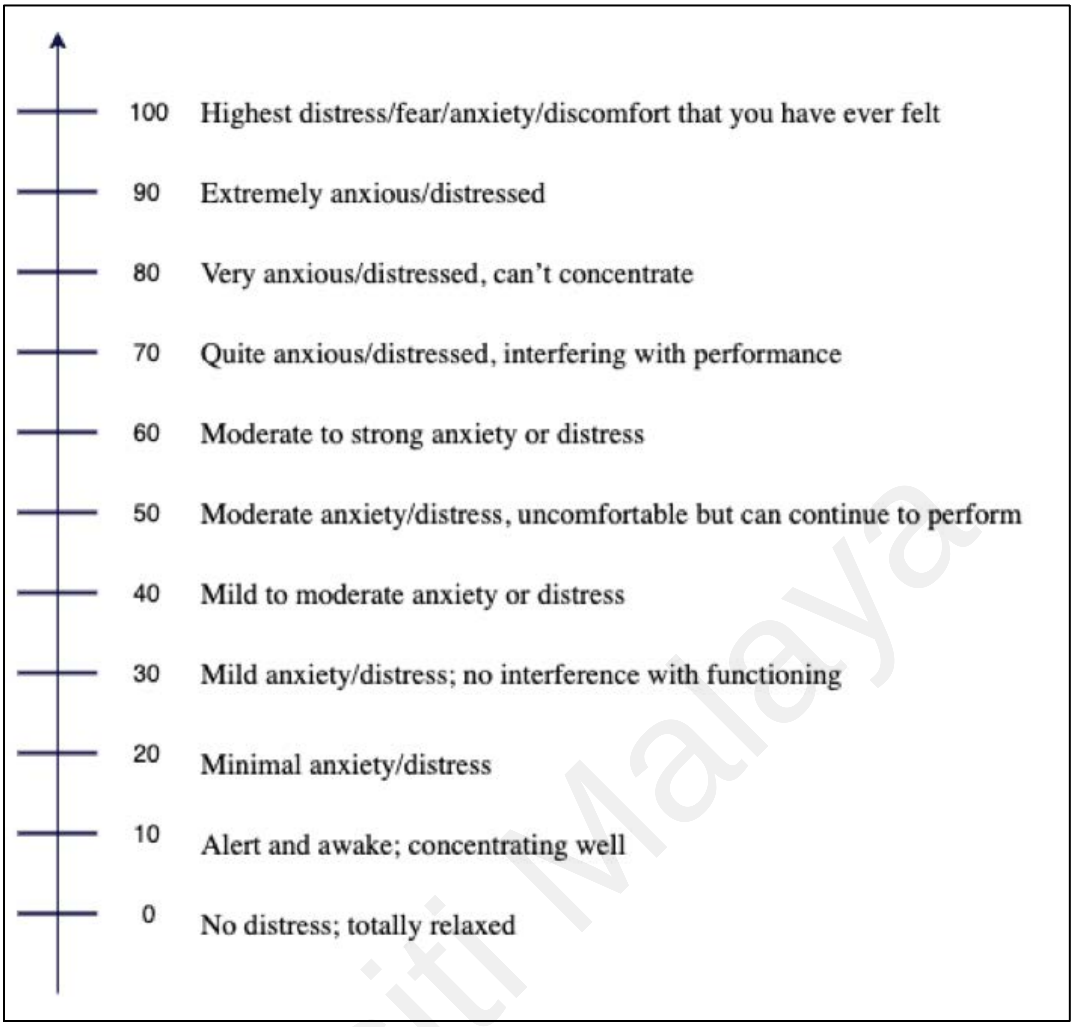


Figure 2.2 Subjective Units of Distress Scale (Benjamin et al., 2010)

2.9 Summary of research gap

Table 2.1 below depicts the comparison of techniques and approaches used in the previous study against the proposed technique in this study. Predominantly, diabetes distress was detected and diagnosed more widely using surveys and questionnaires such as PAID and DDS (Berry et al., 2015; Dieter & Lauerer, 2018). In Section 2.7, survey and questionnaire are shown to be the most common technique used for diabetes distress detection whereas studies on the usage of Facebook reaction to improvise text analysis are relatively low (Beiter et al., 2015; Berry et al., 2015; Gahlan et al., 2018; Schinckus et al., 2018; Sidhu & Tang, 2017) which is outlined in Table 2.1 below. The table also depicts the studies conducted using additional feature apart from text such as Facebook reaction in enhancing emotion or sentiment analysis. Studies showed that incorporation of additional features such as like, share, comment and reaction improves the classification accuracy (Ortigosa et al., 2014; Salas Zarate et al., 2017).

As depicted in Table 2.1, the dictionary-based approach is a popular technique that is extensively used in emotion analysis but has not been applied to the detection of distress in recent studies (Fenwick et al., 2018; Sturt et al., 2015). A dictionary-based approach is a computational approach that uses lexicon-based word mapping to evaluate the emotion that the text conveys to the reader (Dehghani et al., 2017; Firmin et al., 2016; Settanni & Marengo, 2015). Mostly in simplest form, sentiment has a classification model: positive or negative, and can be expanded to various dimensions, such as fear, sadness, anger and happiness. A popular dictionary-based approach is the Linguistic Inquiry and Word Count (LIWC) technique as detailed in section 2.8 (Dehghani et al., 2017; Firmin et al., 2016; Settanni & Marengo, 2015). One of the key benefits of LIWC is the feature that enables users to develop their own dictionaries to analyze any given language usage aspect (Seih et al., 2016). In this research, the advantage of LIWC technique will be leveraged to develop the Diabetes Distress Detection Mechanism based on the Distress Analysis

Interview Corpus-Wizard of Oz (DAIC-WOZ) as dictionary and Subjective Units of Distress Scale (SUDS).

This study incorporates all the techniques presented in the table except survey and questionnaire, which are Linguistic Inquiry and Word Count (LIWC), Distress Analysis Interview Corpus (DAIC), Subjective Unit of Distress Scale (SUDS) and Facebook reactions (Kaur et al., 2018; Meire et al., 2016; Pool & Nissim, 2016). This study emphasizes the development of diabetes distress detection mechanism using a dictionary-based approach and how the approach can be improved by integrating significant Facebook reactions. This integration is evaluated against a baseline mechanism which is executed without integrating the Facebook reaction. Analysis and evaluation of the results are based on the experimental results obtained.

Table 2.1 Summary of Existing Approaches vs Proposed

		Existing			
Techniques	FBR	SQ	LIWC	DAIC	SUDS
Author(s)					
Pool & Nissim, 2016)	/	/			
Meire et al., 2016		/			
Kaur et al., 2018	/	/			
Berry et al., 2015		/			
Fenwick et al., 2018		/			
Gahlan et al., 2018		/			
Dieter & Lauerer, 2018		/			
Mandal & Gupta, 2016			/		
Seih et al., 2016			/		
Settanni & Marengo, 2015			/		
Dehghani et al., 2017			/		
Firmin et al., 2016				/	
Benjamin et al., 2010					/
Kiyimba & O'Reilly, 2017;					/
Ringeval et al., 2017					/
		Proposed			
Current Research	/		/	/	/

FBR: Facebook Reactions, SQ: Survey and questionnaires, LIWC: Linguistic Inquiry and Word Count, DAIC: Distress Analysis Interview Corpus, SUDS: Subjective Unit of Distress Scale

2.10 Conclusion

This chapter provided brief explanations of the domain of study, techniques and approaches of previous work that were leveraged in this study for the development of diabetes distress detection mechanism. The purpose of this review was to contemplate the trends within the past years on how automated distress detection tool has been developed. It is clear from the research reviewed that automation in distress detection among online diabetes community is not evident. Most of the research found was based on surveys. More research and studies are required for the development of an effective automated diabetes distress detection mechanism.

CHAPTER 3: METHODOLOGY

3.1 Introduction

This chapter discusses in detail the research methodology that was adopted in this study beginning from data collection to data pre-processing, emotion classification, followed by the implementing dictionary-based technique and linear regression analysis using significant Facebook reactions.

3.2 Methodology Design and Development

In the context of an automated diabetes distress detection system, the development methodology is to build and implement a mechanism based on dictionary-based and regression techniques. The latter involved the process of data analysis, which includes the systematic application of logical techniques for data collection and analysis, development of distress detection mechanism, and eventually analyzing and evaluating results. Figure 3.1 depicts the three phases of data analysis, namely the data collection phase, classification phase and evaluation phase.

The data collection phase involves data extraction, data cleaning and data prepping for the next phase. The data set used in this analysis was from a previous study (Kaur et al., 2018) and in Section 3.3 the data cleaning and pre-processing involved in explained thoroughly. The second phase, which involved development of distress detection mechanism comprises 3 stages, which are emotion analysis, identifying significant Facebook reactions and development of diabetes distress detection mechanism. Finally, the third phase is the evaluation process of the outcomes from phase 2 in two different stages. Detailed description of the entire design and development of the methodology are presented in the subsequent sub-sections.

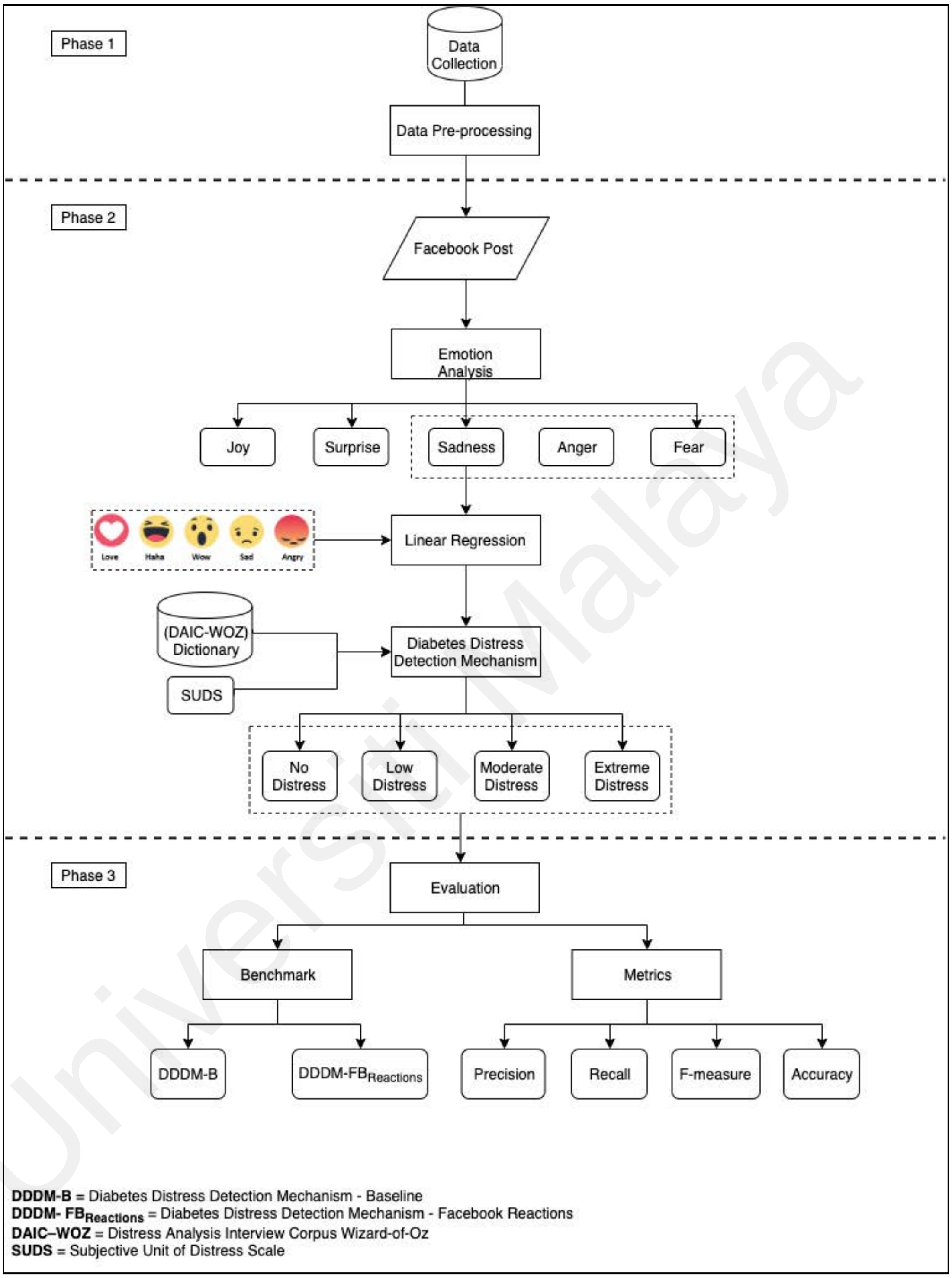


Figure 3.1 Overview of Diabetes Distress Detection Mechanism

3.3 Data Collection

The next sub-section discusses the data collection process, which involves the collection and evaluation of information which could be used to address the research questions and evaluate the outcomes.

3.3.1 Facebook Dataset

The first phase of an analysis process is the empirical phase which involves the collection of data and the preparation of data. The dataset used for this study were from a prior study whereby diabetes related public posts were extracted from Facebook using Graph API3 over 6 months, from July 2016 to January 2017 (Kaur et al., 2018). The collected Facebook posts were from six diabetes related groups (not specified due to confidentiality reasons). These groups have been active with a full participation of users with an average of 42 posts per day since 2014 (Kaur et al., 2018). Besides posts and comments, the researcher have also collected the number of likes, comments and shares (Kaur et al., 2018). This also comprises the number of different reactions such as 'love', 'haha', 'wow', 'sad' and 'angry' derived for each post.

The following sub-section explains the process of data cleaning and pre-processing involved in this study.

3.3.2 Data Cleaning and Pre-processing

Figure 3.2 illustrates the processes of data collection, pre-processing, annotation and sample selection. The total number of collected post was 105,540, where 23,420 posts and comments were removed (i.e. 5,216 posts with emojis only, 7,119 spams and 11,085 posts with only the usernames tagged), leaving 82,120 posts for analysis. This is a simple and crucial step since the removal of any non-textual posts will ensure a smooth classification of the emotions of the text (Kaur et al., 2018).

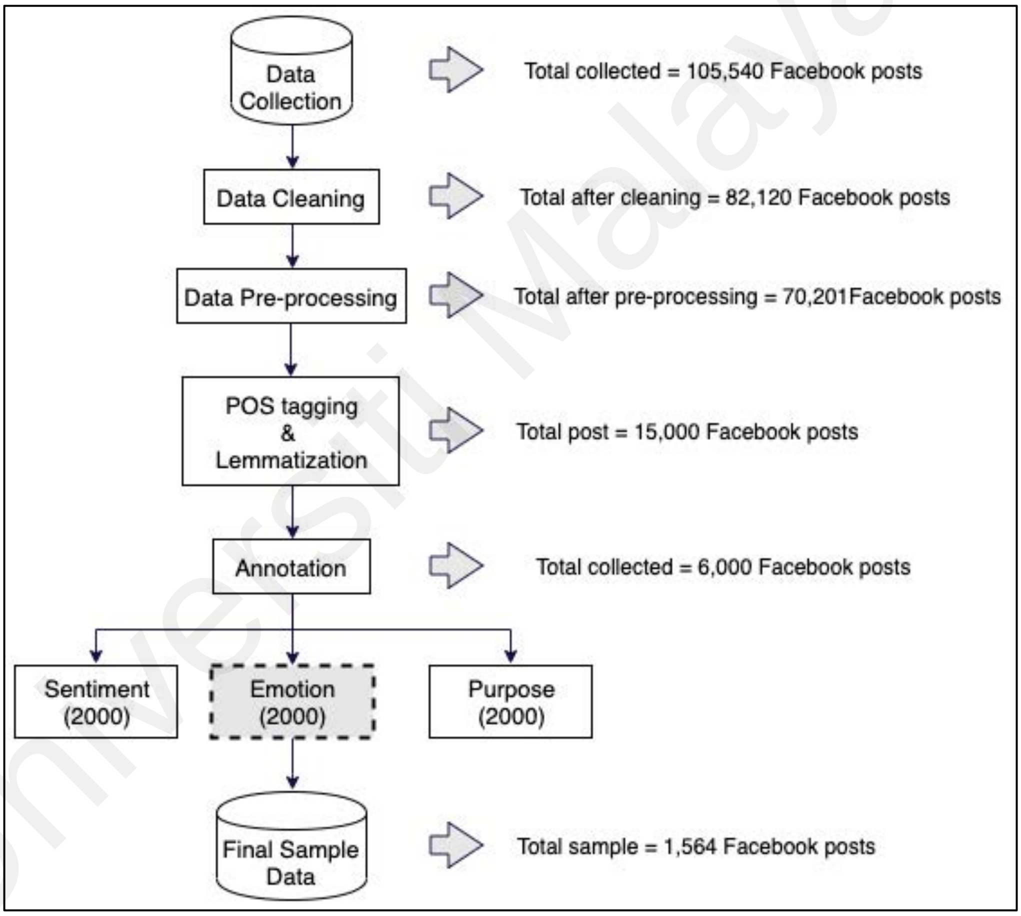


Figure 3.2 Data Cleaning and Pre-processing (Kaur et al., 2018)

Generally, social media data carries a lot of noise (i.e. insignificant data), therefore pre-processing is important to remove noise so that the classification process is not disrupted (Effrosynidis et al., 2017; Kaur et al., 2018; Ravi & Ravi, 2015). The excluded elements are as follows (Kaur et al., 2018):

a) Hashtags and URL links

b) Emoticons and emojis

   i. Emoticons: Set of punctuation marks, letters and numbers arranged to resemble a human face. E.g. :-D means laughing or a big grin, :-O is for surprise.

   ii. Emoji: Pictogram, a small picture that can show anything from a smiling face, a fruit or an animal.

c) Non-textual posts and comments (photo, video, GIF file etc.)

d) Posts and comments that were fewer than 3 words long

e) Posts and comments that were written in language other than English

f) Posts and comments that had five or more misspelled words

g) Special characters (@, #, $, etc.)

Misspelled words were classified as words which have been incorrectly spelt often because of human error, such as 'sometimes' spelled as 'sumtimes', or 'you' as 'u', etc. During the pre-processing stage, any text that had five or more incorrect words was deleted. The remaining posts after these pre-processing steps were then checked against the Wordnik API dictionary where misspellings were corrected, else the word was discarded (Kaur et al., 2018). The process of data-preprocessing resulted in a final dataset of 70,201 posts and approximately 20% (i.e. 15, 000) of these were chosen randomly for POS tagging and lemmatization process (Kaur et al., 2018).

Following the POS tagging and lemmatization process, only 6000 Facebook post were then selected for annotation process. Seven human annotators who are linguistic and medical experts annotated the posts for sentiment, emotion and purpose equally (i.e. 2000 posts). An inter-rater reliability (IRR) analysis was carried out and the resulting Krippendorff's alpha was found to be 89% in agreement (Kaur et al., 2018).

The annotated posts categorized for emotion were then extracted for this study for diabetes distress detection mechanism. Of the 2000 annotated posts for emotion, only 1564 posts were tagged with Facebook reactions, and these were selected as a sample for the emotion detection and distress level analysis. Additional annotation were again administered by three other annotators comprising of medical experts and linguists in order to categorize the posts into different distress levels, namely ‘No Distress’, ‘Low Distress’, ‘Moderate Distress’ and ‘Extreme Distress’ over 2 months period, from January 2019 to March 2019. This annotation were then analysed for inter-rater reliability (IRR) was performed and the Krippendorff alpha was observed to be 84% in agreement.

A few examples of the final sample are as follows:

Table 3.1 Examples of Facebook Posts and Reactions

No of Facebook Reactions / Facebook Post	Love	Wow	Haha	Sad	Angry
"I know just how you feel. I've been told to go and inject my insulin in the bathroom, even in my own home. I'm sick of people trying to make one feel ashamed. Did we ask for Diabetes 1?"	2	9	0	23	59
"I was really pissed i had diabetes till i had a heart attack, got diagnosed with coronary artery disease, open angle glaucoma, hypothyroid, pustular psoriasis and depression. Now, I WISH i only had diabetes."	5	23	0	21	2
"Yes, it's terrifying, I can't see and my whole body shakes, I purposely raise my blood sugar before bed because I'm so scared, I won't wake up"	9	2	0	14	3

The following section describes the different data analysis techniques, including emotion analysis, regression analysis and the distress detection mechanism.

3.4 Diabetes Distress Detection Mechanism

In the second phase of the mechanism, the sample of 1564 Facebook posts were analyzed and classified in three different stages. The first stage was the emotion classification based on the Facebook text posts whereas in the second stage linear regressions were performed on posts that were classified as Sadness, Anger and Fear. This posts were selected based on emotions that were identified as the negative emotion which has been higlighted used dotted closure in Figure 3.3. The third stage was the development of the Diabetes Distress Detection Mechanism (DDDM) using the dictionary-based technique. The following sub-sections describe each stage thoroughly.

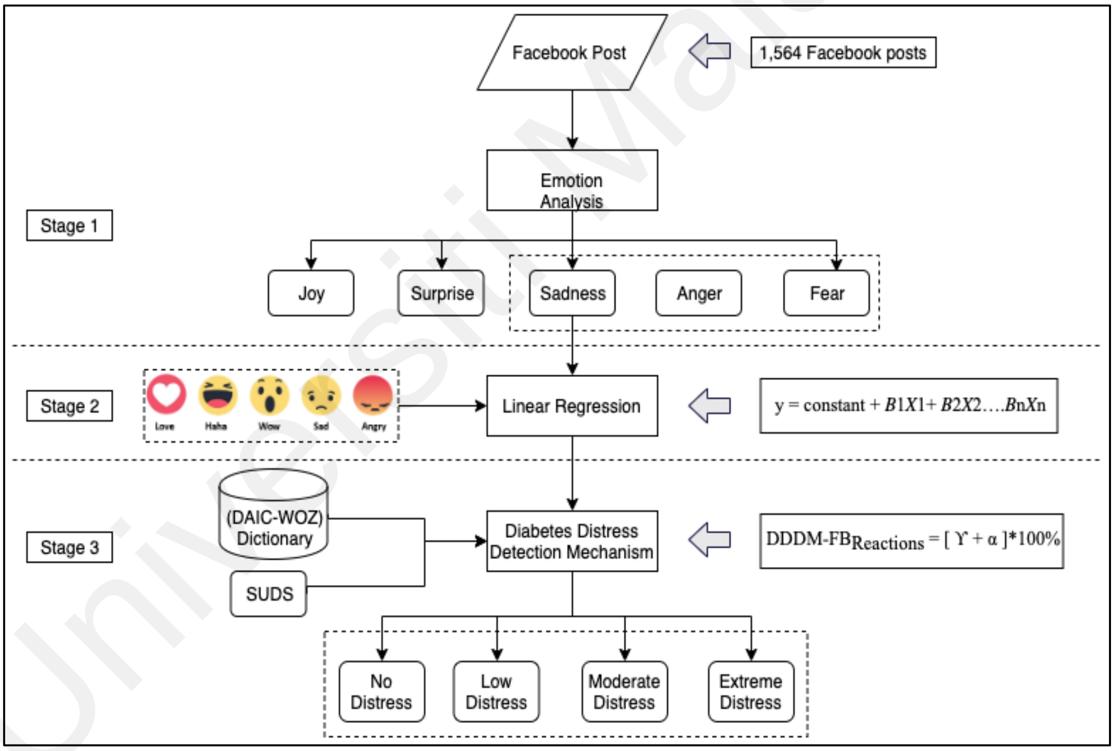


Figure 3.3 Diabetes Distress Detection Mechanism

3.4.1 Emotion Analysis

Figure 3.4 illustrates on how Indico API was used to perform emotion classifications. Indico API's text analysis approach allows a machine to acquire a range of information from plain text data using machine learning method (Agarwal et al., 2018). The emotion model in Indico API allows a machine to predict a variety of emotion expressed in a plain text input using its in-built functions (Agarwal et al., 2018). This function returns a dictionary that maps 5 emotions (i.e. anger, fear, joy, sadness, surprise) to the probability that the text is expressing the respective emotion (Agarwal et al., 2018; Ahmad et al., 2017).

Indico API was executed using PhP and MySQL scripting which gives an output of a weighted average score for all the 5 emotions, with the value scale of 0 to 1. The highest value among the 5 emotions (i.e. anger, fear, joy, sadness, surprise) indicates a very strong emotion while lower values indicate a moderate to weak emotion. The post is then classified based on its highest weight score, where Table 3.2 illustrates an example of this.

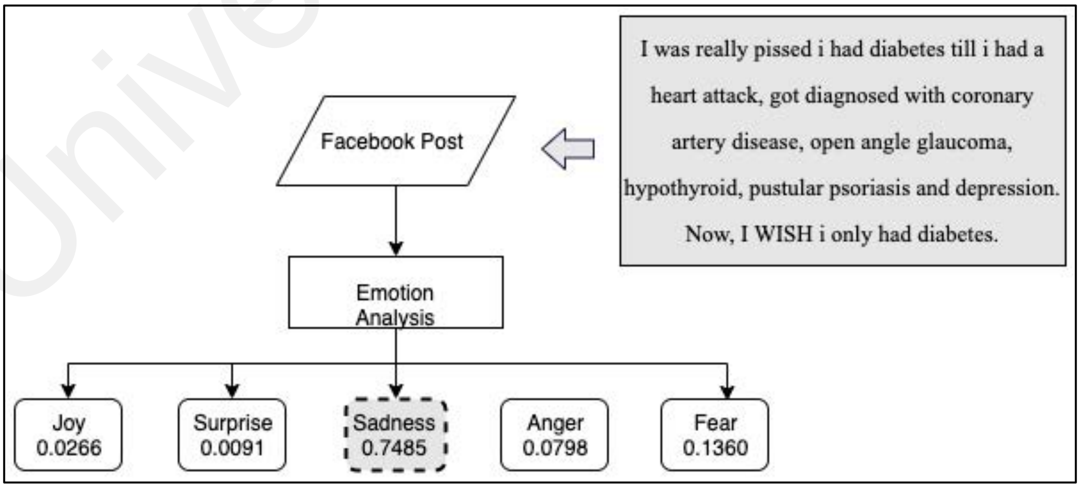


Figure 3.4 Emotion Analysis Stage

Table 3.2 Sample Post with Emotion Score

Facebook Post	I was really pissed i had diabetes till i had a heart attack, got diagnosed with coronary artery disease, open angle glaucoma, hypothyroid, pustular psoriasis and depression. Now, I WISH i only had diabetes.
anger_score	0.079823852
joy_score	0.026580459
fear_score	0.135965601
sadness_score	0.748547256
surprise_score	0.009082856

Based on the emotion scores for the post above, the strongest emotion is Sadness with a score of 0.748547256, indicating a negative emotion.

The next sub-section discusses the next stage which is identifying the significant Facebook reactions for each post.

3.4.2 Linear Regression on Facebook Reactions

Following the emotion classifications, a linear regression analysis was then administered to identify significant Facebook reactions as depicted in Figure 3.5. This was done to determine which Facebook reactions affect the specific emotions (Effrosynidis et al., 2017). Linear regression is a reliable method of identifying which variables have impact on a topic of interest (Rezende et al., 2016; Yi & Yu, 2018). (Rezende et al., 2016; Yi & Yu, 2018). The dependent variable is the main factor to be predicted and Independent Variables are the factors that hypothesize to have an impact on the dependent variable (Effrosynidis et al., 2017).

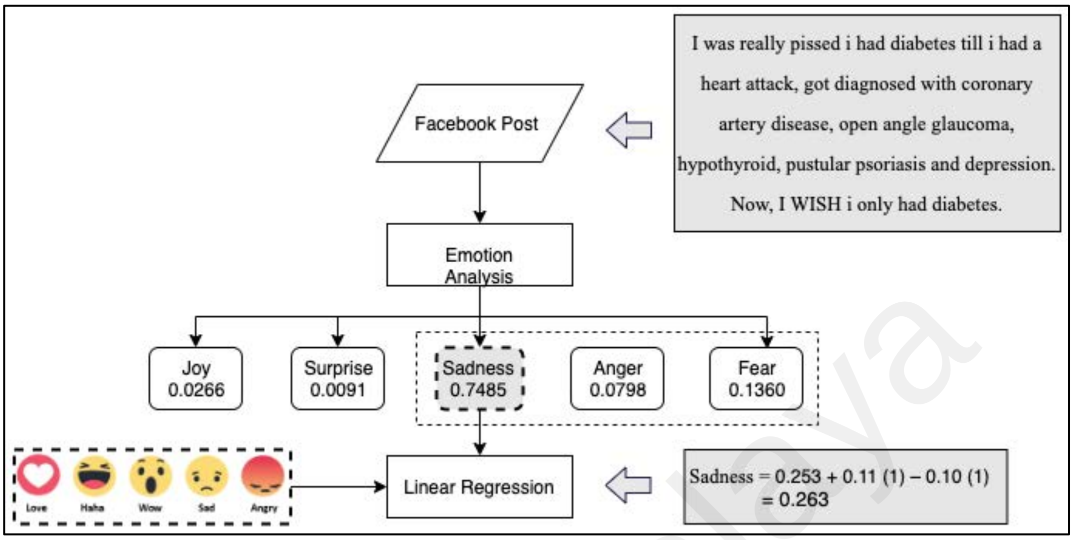


Figure 3.5 Significant Facebook Reactions

IBM SPSS Statistics 24 was used to perform the linear regression analysis. The dependent variables were the 5 emotions classified using Indico API (i.e. Anger, Fear, Sadness, Joy and Surprise) and the independent variables were the Facebook reactions (Love, Wow, Haha, Sad and Angry). The general linear regression equation is as stated in Equation 1 below:

$$y = \text{constant} + \beta 1\chi 1 + \beta 2\chi 2 + \ldots . \beta n\chi n; \qquad (1)$$

where y = dependent variable, β= unstandardized β value from regression results,

χ= weight of significant variable

The weight of χ was determined by multiple experiments using different weights of reactions (i.e. number of reactions on a post) to determine the final weight to be applied to each reaction attribute (i.e. Love, Wow, Haha, Sad and Angry) (Effrosynidis et al., 2017; Kaur et al., 2018; Yi & Yu, 2018). Due to huge variance in the number of reactions towards a particular post, the weight assigned for every reaction was normalized between

0 to 1 (Kaur et al., 2018). For instance, the number of reactions within the range of 1 to 5 was assigned a weight of 0.2, whereas the number of reactions above 18 is assigned a weight of 1. A similar concept was applied to other range of reactions, as shown in Table 3.3.

Table 3.3 Normalized Facebook Reaction Frequency

Reactions	Frequency Range		Normalized Value
Love, Wow, Haha, Sad and Angry	0	0	0
	1	5	0.2
	6	10	0.4
	11	12	0.6
	13	17	0.8
	≥ 18		1

Each significant reaction is weighted using the regression formula above and is later factored in the Diabetes Distress Detection Mechanism to determine the significance of Facebook reactions towards the distress level. A sample calculation is as follows:

Sample text: *I was really pissed i had diabetes till i had a heart attack, got diagnosed with coronary artery disease, open angle glaucoma, hypothyroid, pustular psoriasis and depression. Now, I WISH i only had diabetes.*

Assume the sample post above has been categorized as Sadness. The regression analysis assigns the above post with a 0.253 constant score. Assume the number of significant reactions is as shown in Table 3.4.

Table 3.4 Frequency of Facebook Reaction

Love Frequency	Wow Frequency	Haha Frequency	Sad Frequency	Angry Frequency
5	23	0	21	2

Therefore, the final score determined by factoring in the significant reactions is:

Sadness = 0.253 + 0.11 (Sad) – 0.010 (Wow)
Sadness = 0.253 + 0.11 (1) – 0.10 (1)
Sadness = 0.263

Based on linear regression results, the significant reactions for 'Sadness' emotion is 'Sad' and 'Wow'. The calculation above assigned the normalized value for both the reactions based on the frequency as shown in Table 3.3.

3.4.3 Diabetes Distress Detection Mechanism

Stage 3 was to develop a diabetes distress detection mechanism using the DAIC Inquiry and Word Count (DIWC) technique (Bravo-Marquez et al., 2016; Effrosynidis et al., 2017). This technique calculates the percentage of various categories of depressed words used in a text and determines the level of distress in the text. The DIWC technique was adapted and developed based on the Linguistic Inquiry and Word Count (LIWC) technique which is one of the most used technique by data scientists for text analysis (Dehghani et al., 2017; Firmin et al., 2016; Settanni & Marengo, 2015). LIWC is an analysis technique used to provide insights and calculate percentage of a text-corpora on a word-by-word basis that falls within various domains such as emotion, cognitive, social and etc. (Dehghani et al., 2017; Firmin et al., 2016; Settanni & Marengo, 2015).

One prominent advantage of LIWC includes the feature that allows users to develop their own dictionaries for analyzing any given dimension of the language used (Seih et al., 2016). The developed diabetes distress detection mechanism in this research leveraged on the advantage mentioned above to incorporate the Distress Analysis Interview Corpus – Wizard of Oz (DAIC- WOZ) as dictionary to develop a customized technique, which was DAIC Inquiry and Word Count (DIWC) technique.

The dictionary used in this research is composed of 36885 distressed words and word stems. This dictionary is part of a larger corpus, the Distress Analysis Interview Corpus (DAIC) (Gratch et al., 2014), which includes a database built using semi-structured clinical interviews of participants to enable the diagnosis of psychological distress conditions such as depression, anxiety and post-traumatic stress disorder. This DAIC database was used as the dictionary in DIWC to aid the development of automated diabetes distress detection mechanism whereas the Subjective Unit of Distress Scale (SUDS) was used to classify the detection into 4 levels of 'No Distress', 'Low Distress', 'Moderate Distress' and 'Extreme Distress' (Benjamin et al., 2010; Kiyimba & O'Reilly, 2017). SUDS is commonly used to determine the subjective intensity of disturbance or distress currently experienced by an individual (Benjamin et al., 2010; Kiyimba & O'Reilly, 2017). Figure 3.6 and 3.7 shows the sample words in distress analysis interview corpus and the subjective unit of distress scale respectively.

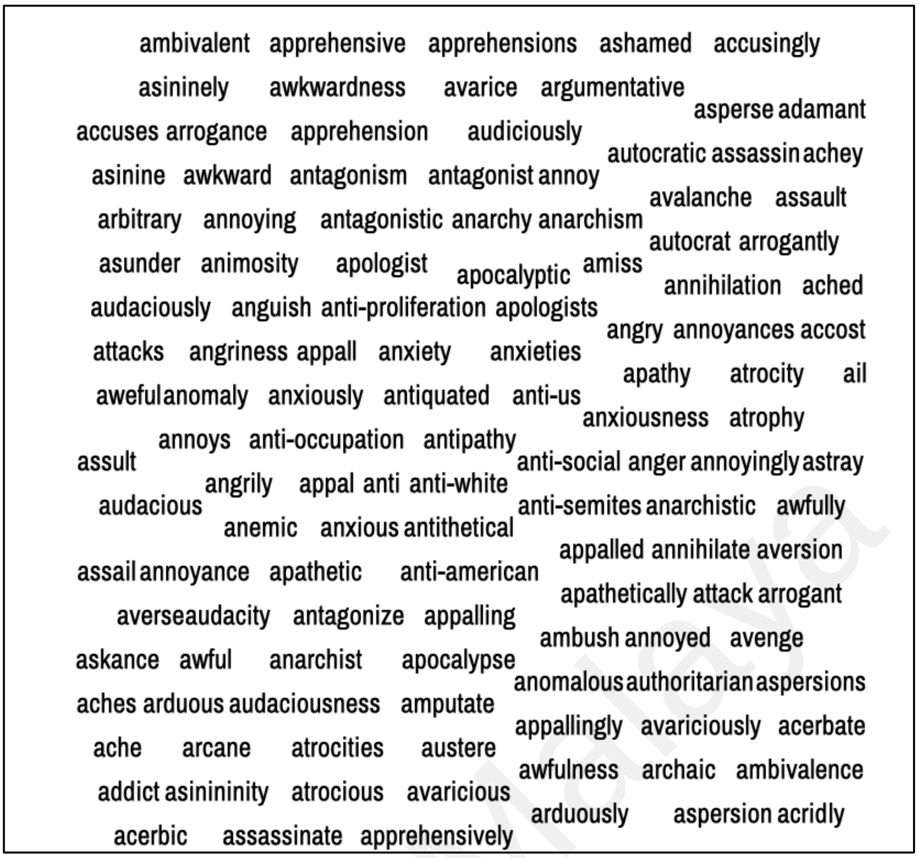


Figure 3.6 Sample words in Distress Analysis Interview Corpus

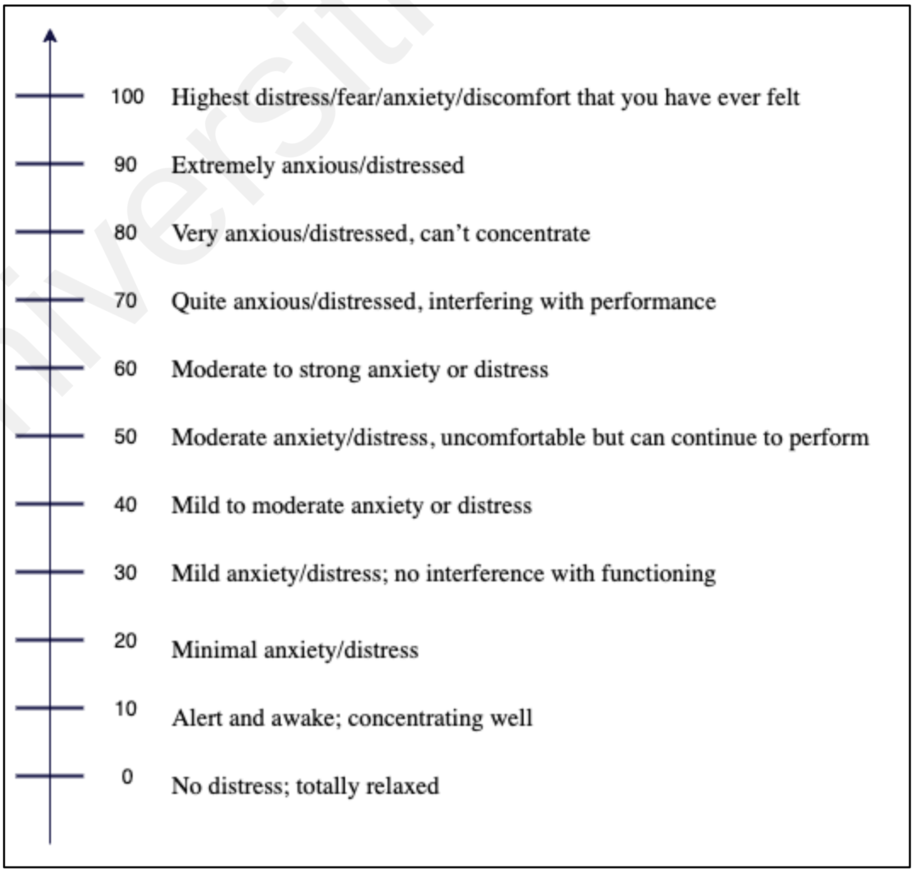


Figure 3.7 Subjective Units of Distress Scale (SUDS)

From the initial Emotion Detection results, a total of 1,176 posts that were classified as ‘Anger’, ‘Sadness’ and ‘Fear’ were extracted to run the analysis to determine the level of distress present in the post. Visual Basic for Applications Macros was used to determine the percentage of distress level present in the post using the DIWC technique. The program is automated to check each word in a post against the words in the dictionary to determine the number of matched words. DIWC contains a linguistic dictionary of 36,885 distressed words that performs a word-by-word assessment of emotional and cognitive components in the text post. The DIWC program for text analysis searches for distress-related words based on these 36,885 words. This analysis is first performed with a baseline mechanism and then improvised mechanism.

3.4.3.1 Experimental Set Up

For benchmarking purposes, the baseline mechanism (DDDM-B) in Figure 3.8 was set without including the significant Facebook reaction scores, contrary to the improvised mechanism (DDDM-$FB_{Reactions}$) in Figure 3.9.

The baseline calculates the word count of matched words against total word count of a post using the following formula:

$$\text{DDDM-B} = \left(\frac{\text{No. of word counts matched}}{\text{Total No. of word count in a post}} \right) * 100\% \tag{2}$$

To improvise the results, the second analysis was factored in with the regression weight of significant Facebook reactions using the following formula:

$$\text{DDDM-FB}_{Reactions} = [\ \Upsilon + \alpha\ ] * 100\% \ ; \tag{3}$$

where Υ = regression weight of significant Facebook reactions, α = Distress level value from DDDM-B

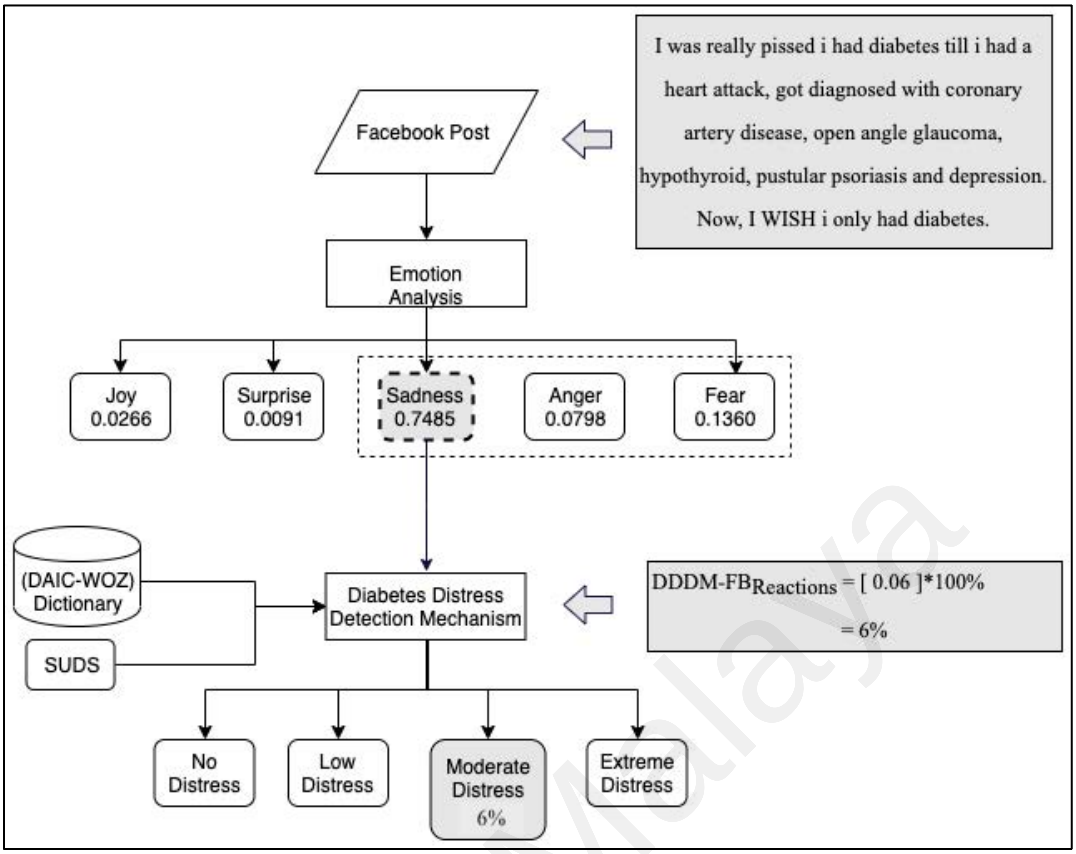


Figure 3.8 Diabetes Distress Detection Mechanism – Baseline

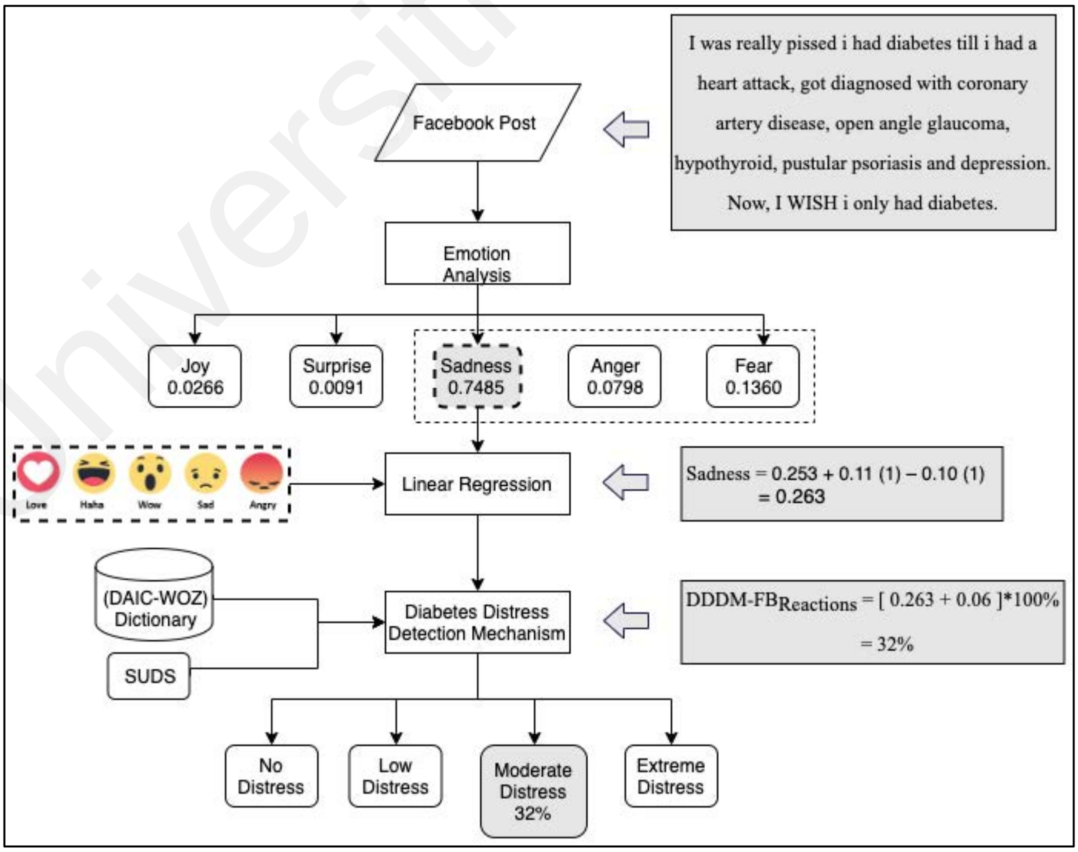


Figure 3.9 Diabetes Distress Detection Mechanism with Facebook Reactions

The DDDM-FB$_{Reactions}$ mechanism determines the significance and impact of Facebook reactions towards the intensity of distress level present in a post. The comparison for baseline and improvised model for each emotion is shown in the table below:

Table 3.5 Comparison of Baseline and Improvised Mechanism

Facebook Post	Emotion	DDDM-B	DDDM-FB$_{Reactions}$
"Frightening to wake up soaked to the skin, disorientated and confused. Change of nightie & sometimes bedlinen in the early hours, then freezing cold & wide awake .... not funny."	Anger	17 % (Low Distress)	44% (Moderate Distress)
"Scared of losing my eyesight. I have mottling of the macula. Couldn't get new glasses. Can't drive anymore. Scared to death. Being tired all of the time. Just want to sit. "	Fear	16 % (Low Distress)	37% (Moderate Distress)
"Had a lot of stress in my life here lately went to the doctor today and my sugar was 4 37 they put me on Invokana and keep raising my insulin. I feel absolutely terrible."	Sadness	6 % (Low Distress)	32% (Moderate Distress)

The results of these two analyses are then tabulated to evaluate any significant improvement in the distress level results. This is performed by comparing the overall accuracy of the two mechanism in detecting distress level in the posts.

3.5 Evaluation

The final phase is to evaluate the outcome of the results which was conducted in two stages. The first stage investigates the performance of DDDM with and without the regressed Facebook reactions (i.e. DDDM-B versus DDDM-FBReactions). This is followed by the findings of the two mechanisms against human annotated results to see if the mechanism's accuracy is enhanced when Facebook reactions is included. The performance of the mechanism was evaluated using the standard metrics such as accuracy, precision, recall and F-measure to test the reliability and effectiveness of the proposed mechanism.

The total number of posts that were used for the experiment was 1,564 posts. Of these, 1,176 posts classified as ' Anger, ' ' Sadness ' and ' Fear ' are extracted for human annotation to determine the level of distress in the post. Then the human experts read each post manually and categorized the latter as no distress, low distress, moderate distress or extreme distress.

$$\text{Precision} = \frac{\sum \text{post that match human classified count}}{\text{Total post produced by tool}} \tag{4}$$

$$\text{Recall} = \frac{\text{Total post that match human results}}{\text{Total human classified post}} \tag{5}$$

$$\text{F-measure} = \frac{2 \text{ X precision X recall}}{\text{precision + recall}} \tag{6}$$

$$\text{Accuracy} = \frac{\text{Total correctly classified post}}{\text{Total post}} \tag{7}$$

To evaluate the precision-recall curve, F-measurement metric was applied which has the range varying from 0 (the worst results) to 1 (the best results). A higher F-measure is an indicator of a good classification performance (Kaur et al., 2018; Lee, Hu, & Lu, 2018).

3.6 Summary

This chapter describes the research design employed in this study. Data collection, pre-processing and sample selection are discussed prior the introduction of methodology used to conduct this research. The following chapter discusses the outcome of the mechanism and the results.

CHAPTER 4: RESULTS AND DISCUSSION

4.1 Introduction

This chapter discusses the results that were obtained after conducting the experiments to build a Diabetes Distress Detection Mechanism. Two mechanisms are built, i.e. a baseline mechanism with DIWC technique and an improvised mechanism with DIWC technique and Facebook reactions. To evaluate the quality of the built mechanisms, human annotation three main metrics are used, namely precision, recall, and accuracy. Overall, the results of the built mechanisms are compared and discussed. Based on the obtained information, the conclusion about the most efficient model is given.

4.2 Experimental Setup and Results

This section presents the findings of this study that has been divided into a few parts. The first section evaluates the emotion analysis model's output of different emotions. The following section introduces significant Facebook reactions towards the post. The finding of the earlier two parts is preceded with the experimentation of distress level before (baseline) and after the Facebook reaction has been implemented. This is followed by the evaluation of the findings against human annotation classification using standard metrics. The final section summarizes the research findings of individual experimentations (Facebook posts and Facebook reactions) and combined experimentations (Facebook posts with Facebook reactions).

4.2.1 Emotion Analysis

The first phase of the experiment was emotion analysis on the 1,564 data sample where Indico API is used to perform emotion classification. The Indico API's feature returns a dictionary that maps the probability of the text reflecting the associated emotion. The Indico API's emotional model predicts a variety of emotions from a plain text input into emotions such as anger, fear, joy, sadness and surprise.

Indico API generates a weighted average score output for all 5 emotions, ranging from 0 to 1. This is subsequently classified into a single, predominant emotion based on the emotions ' highest weight score. The following table and chart show the frequency of the different emotions.

Table 4.1 Emotion Analysis Results

Emotion	Emotion Frequency
Anger	422
Fear	261
Sadness	493
Surprise	85
Joy	303
Total	**1564**

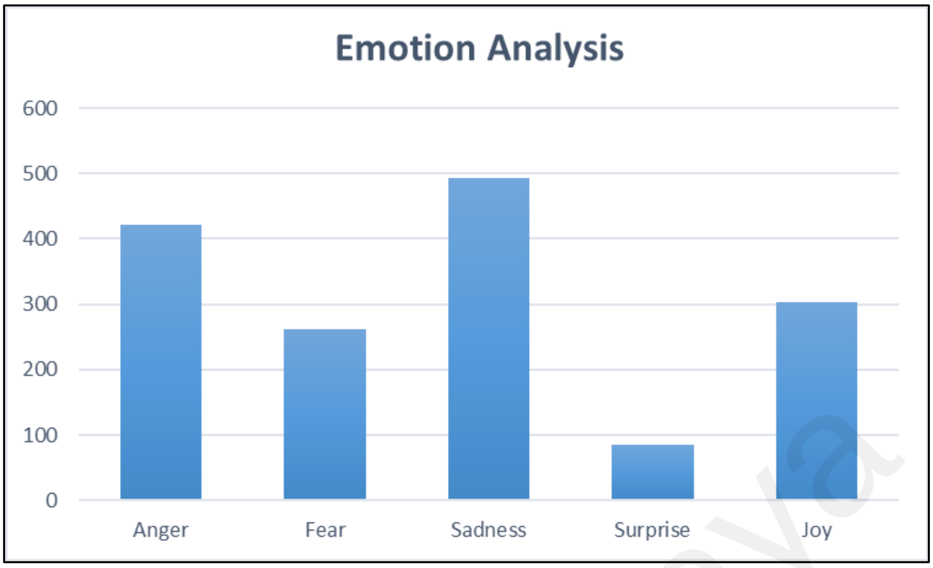


Figure 4.1 Emotion Analysis Results

Table 4.1 and Figure 4.1 illustrates the breakdown of the classification of the five emotions based on the Indico API emotional analysis. In terms of emotions, sadness ranked the highest (493), followed by anger (422) and joy (303). Generally, it can be observed that sadness and anger emerged as the prominent negative emotion whereas joy ranked top for the positive emotions. Nevertheless, the emergence of all three negative emotions can be associated with diabetes-related emotions which comprise a range of negative emotions of fragility, sadness, and fear, which can trigger depression, anxiety, panic, social alienation, and philosophical crisis (Fisher et al., 2015; Snoek et al., 2015).

4.2.2 Significant Facebook Reactions

Following the emotion analysis on the 1564 post, regression analysis was applied to identify significant the Facebook reactions towards each post. Literature has shown how Facebook's reaction feature allows users to react to the original post to express their thoughts instead of using the ' Like ' button that communicates minimal emotion (Pool & Nissim, 2016).

The dependent variables in this analysis are the five emotions classified earlier (i.e. Anger, Fear, Sadness, Joy and Surprise) and the independent variables are the Facebook reactions (Love, Wow, Haha, Sad and Angry) tagged to each post in the dataset. In order to identify which independent variable contributes most towards the dependent variable, linear regression analysis was executed. Table 4.2 below is derived from the IBM SPSS Statistics platform's regression analysis model summary with parameters such as R-square, B coefficients and p-values. The required parameters for regression analysis are the R-square, B coefficients and p-values. The table below depicts the significant Facebook reactions towards each emotion which was derived from the linear regression analysis. The B coefficients represent the sum of the increase in the emotion scores that would be predicted by an increase of 1 unit of the independent variables, whereas, the p-values illustrate that the B coefficient is statistically significant if its p-value is less than 0.05. A significant relationship was commonly observed between ‘angry’, ‘love’, ‘wow’ and ‘sad’ reactions towards the emotions. It is also observed that ‘haha’ reaction does not show any significant relationship towards any of the emotions. The findings of this regression study provided valuable insight into the variability of reactions in a Facebook post. For instance, ' angry ' reaction has a significant effect on a post's overall anger emotion; high ' angry ' reactions to a post are indicates stronger anger emotion. Anger is among the most prevalent negative emotions that could be conveyed is based on the actions of other users in the online diabetes community. Thus, whenever an angry article is posted or shared, the online community is generally responding appropriately to show that they consent, and that is when an angry reaction is expressed.

Table 4.2 Linear Regression Results

Emotion	Reactions	R^2	B	F *(p-value)*	t	*p*-value
Anger		0.17	0.16	3.074 *(0.001)*		
	Angry				1.92	0.044
Fear		0.06		2.133 *(0.024)*		
	Love				-3.248	0.001
			-0.10			
	Wow				3.07	0.002
Sadness		0.26		4.163 *(0.000)*		
	Wow		-0.10		-3.514	0.000
	Sad		0.11		3.754	0.000

Note: Only significant results are shown; Anger = 0.267 +0.16(Angry); Fear = 0.213 – 0.10(Love) + 0.10(Wow); Sadness = 0.253 – 0.10(Wow) + 0.11(Sad)

Either wow and love have also been found to significantly predict two different emotions, that is, joy and fear. As shown in Table 4.2, wow is viewed as "ambiguous" which signifies that it is challenging to affiliate this reaction with a specific emotion (i.e. Fear or Sadness). This partly explains the existence of wow as a significant reaction for both sadness and fear emotions. Fear can be correlated with patients feeling anxious or scared. Online diabetes community could convey their surprise at users' fear as a sign of comforting them down and supporting them to resist their emotions. Sad reactions are often used as a supportive reaction to convey that they perceive the emotions of the Facebook user. As such an ambiguous reaction, members responded with a wow to sadness in order to acknowledge the emotion they expressed. The user could even look forward to a positive opportunity to shift easily to the future from their otherwise frustrating or distressing situation.

4.2.3 Diabetes Distress Detection Mechanism

The next phase of the experiment involved the development of the Diabetes Distress Detection Mechanism (DDDM) which was built using a Dictionary Inquiry and Word Count technique. This technique analyzes the percentage of matched words of distress in a post with the dictionary of DAIC-WOZ and determines the level of distress in that post. The DAIC-WOZ dictionary used in this research is part of a larger corpus, the Distress Analysis Interview Corpus (Gratch et al.,2014), composed of 36885 distressed words and word stems. The level of diabetes distress is classified into 4 levels of 'No Distress', 'Low Distress', 'Moderate Distress' and 'Extreme Distress' (Benjamin et al., 2010; Kiyimba & O'Reilly, 2017).

Across the dataset of 1564 Facebook post used for distress detection in this study, approximately only 75% (i.e. 1176 posts) of those data that were classified as 'Anger', 'Sadness' and 'Fear' are extracted to run the analysis to determine the level of distress present in the post. This study was first performed with a baseline mechanism focusing primarily on text post and the improvised mechanism included significant Facebook reactions.

4.2.3.1 Diabetes Distress Detection Mechanism - Baseline

The baseline mechanism (DDDM-B) was developed without including significant Facebook result scores, in contrast to the improvised mechanism (DDDM-FBReactions) for evaluation purposes.

The table and pie chart below depicts the frequencies of detected distress levels with different emotions. Table 4.3 is derived from the output of DAIC Inquiry and Word Count (DIWC) technique which calculated the percentage of distinct types of depressing words used within the text and evaluated the extent of distress in the text.

Table 4.3 Diabetes Distress Detection Mechanism – Baseline Results

DDDM - B	ND 0%	LD 1% - 29%	MD 30% - 69%	ED 70% - 100%	Total
Anger	113	309	0	0	422
Sadness	154	339	0	0	493
Fear	86	175	0	0	261
Total	**353**	**823**	**0**	**0**	**1176**

* DDDM-B = DDDM-Baseline, ND = No Distress, LD = Low Distress, MD = Moderate Distress, ED = Extreme Distress

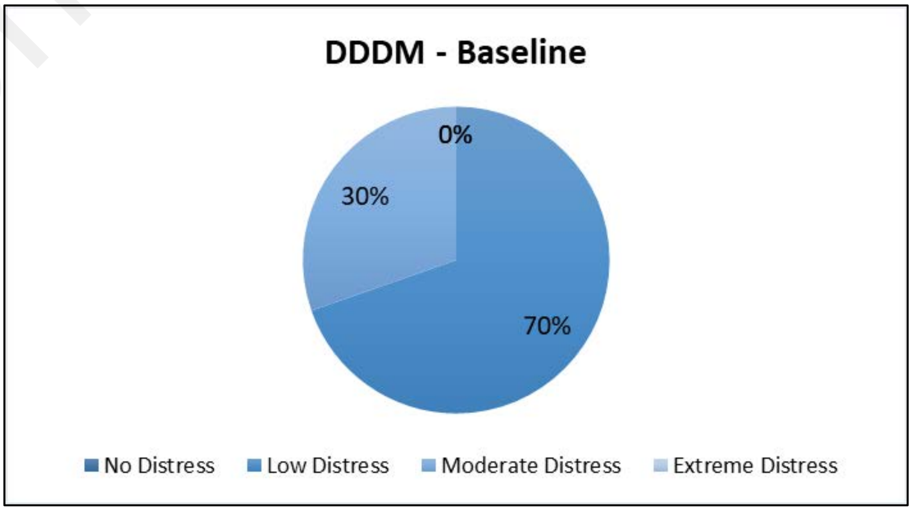


Figure 4.2 Diabetes Distress Detection Mechanism – Baseline Results

The results illustrate that the majority of the post falls under the category 'Low Distress' with a percentage of 70%. The remaining posts are classified as ‘No Distress’ which is 30%. It can also be observed that none of the posts are classified under 'Moderate Distress' or 'High Distress' category.

4.2.3.2 Diabetes Distress Detection Mechanism with Facebook Reactions

The next set of experiment involved factoring significant Facebook reactions to the developed mechanism to improvise the results of the baseline model. This improvised mechanism evaluates the relevance and impact of Facebook reactions towards the magnitude of distress level present in a post.

Table 4.4 exhibits the results of improvised mechanism where it can be observed that ‘Low Distress’ increased to 70% whereas ‘No Distress’ has dropped to 0%. ‘Moderate Distress’ was observed to be at 30% compared to baseline at 0%. Similar to the baseline mechanism, none of the posts are classified under ‘Extreme Distress’. It can also be observed that the majority of data falls under the level of ‘Low Distress’ in both mechanisms, indicating a substantial low distress presence in the online diabetes community.

Table 4.4 Diabetes Distress Detection Mechanism – Facebook Reaction Results

DDDM-FB$_R$	ND 0%	LD 1% - 29%	MD 30% - 69%	ED 70% - 100%	Total
Anger	0	240	182	0	422
Sadness	0	378	115	0	493
Fear	0	201	60	0	261
Total	**0**	**819**	**357**	**0**	**1176**

* DDDM-FB$_R$ = DDDM-Facebook Reactions, ND = No Distress, LD = Low Distress, MD = Moderate Distress, ED = Extreme Distress

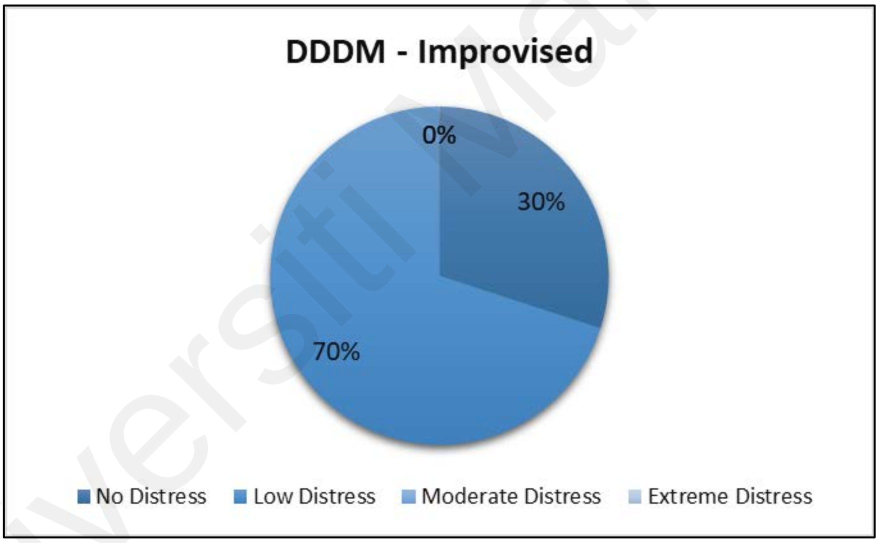


Figure 4.3 Diabetes Distress Detection Mechanism – Facebook Reaction Results

4.3 Evaluation Metrics of Mechanism Performance

The evaluation results of the different detection mechanisms are presented in Table 4.5 and Table 4.6. The effectiveness of the developed Diabetes Distress Detection Mechanism was evaluated using the standard metric evaluation i.e. precision, recall, F-measure and accuracy. Table 4.5 depicts the evaluation of the baseline and improvised mechanism with emotions (anger, fear and sadness) against the distress levels (No Distress, Low Distress, Moderate Distress and Extreme Distress). Table 4.6 illustrates the summary of overall distress level detected against a specific emotion.

The results of DDDM-Baseline were compared with the results of DDDM-$FB_{Reactions}$ using human annotation. From the results obtained, it can be seen that DDDM-$FB_{Reactions}$ could produce higher accuracy, thereby concluding that mechanism with the significant Facebook reactions enhanced the results compared to the baseline mechanism. It is proved essential to incorporate to the significant reactions to the classification of distress as this has ensured that the accuracy of the classification is enhanced.

Prior studies have indeed documented the impact of user engagement behaviors on Facebook, particularly in the use of like, share, comment and reaction (Kaur et al., 2018). Users are generally likely to use Facebook reactions to react when they feel connected or agreeable to a post, which is an acknowledgement of the post's emotion. Therefore, the accuracy of distress classification can be improved by adding these aspects to the distress classification as shown by the findings in the tables below.

Table 4.5 Comparison of Different Mechanism

Mechanisms			ND	LD	MD	ED
DDDM-B	Anger	Precision	0.69	0.50	0	0
		Recall	0.77	0.83	0	0
		F-measure	0.73	0.62	0	0
	Fear	Precision	0.43	0.54	0	0
		Recall	0.67	0.67	0	0
		F-measure	0.52	0.59	0	0
	Sadness	Precision	0.52	0.25	0	0
		Recall	0.57	0.59	0	0
		F-measure	0.54	0.34	0	0
DDDM-FB_R	Anger	Precision	0	0.57	0.58	0
		Recall	0	0.74	0.88	0
		F-measure	0	0.64	0.70	0
	Fear	Precision	0	0.63	0.72	0
		Recall	0	0.90	0.65	0
		F-measure	0	0.74	0.68	0
	Sadness	Precision	0	0.38	0.83	0
		Recall	0	0.97	0.47	0
		F-measure	0	0.55	0.60	0

** DDDM-B = DDDM-Baseline, DDDM-FB_R = DDDM-Facebook Reactions, ND = No Distress, LD = Low Distress, MD = Moderate Distress, ED = Extreme Distress

Table 4.6 Results Summary

Mechanisms	DDDM-B			DDDM-FB$_R$		
Metrics	Precision	Recall	F-measure	Precision	Recall	F-measure
DL-Anger	0.59	0.80	0.68	0.57	0.81	0.67
DL-Fear	0.48	0.67	0.56	0.67	0.78	0.71
DL-Sadness	0.38	0.57	0.44	0.61	0.72	0.58
Average	0.49	0.68	0.56	0.62	0.77	0.65
Accuracy	**49%**			**62%**		
Improvement	**DDDM-FB$_{Reactions}$ - DDDM-B= 13%**					

* DDDM-B = DDDM-Baseline, DDDM-FB$_R$ = DDDM-Facebook Reactions, DL = Distress Level

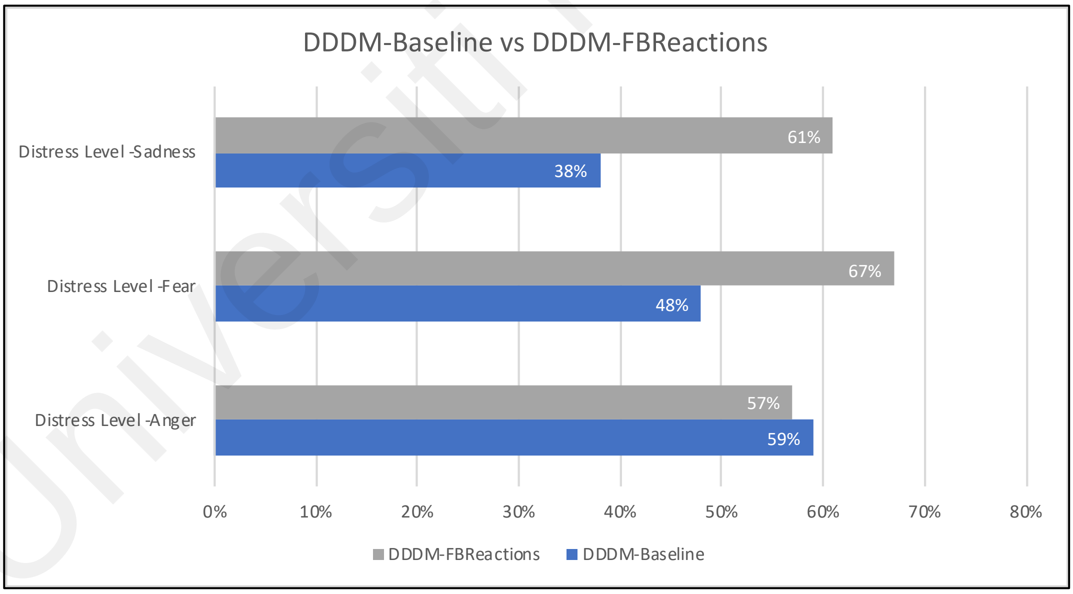


Figure 4.4 Accuracy Comparison with Distress Level and Emotion

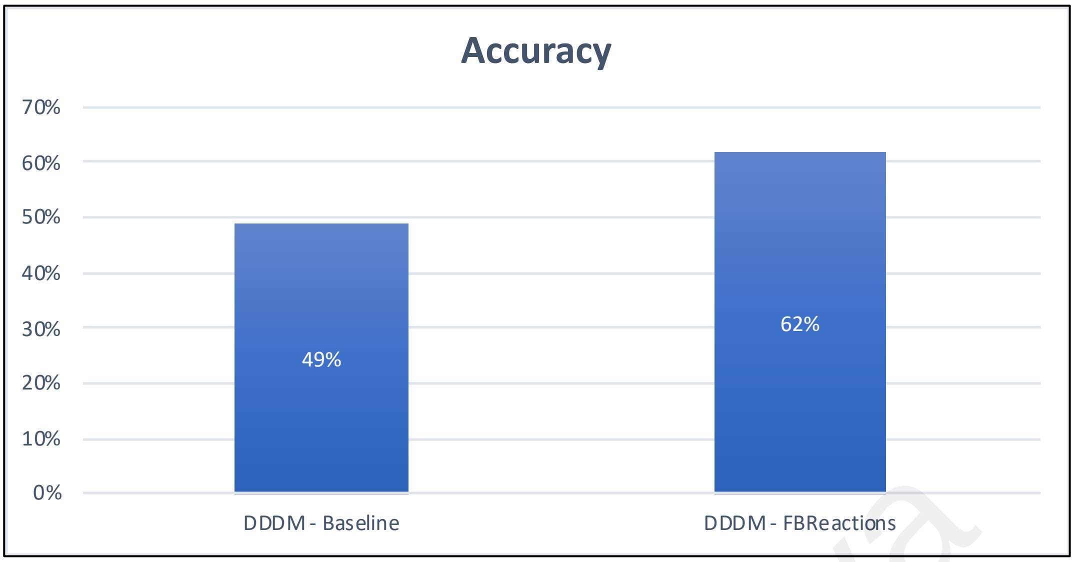


Figure 4.5 Accuracy Comparison of Diabetes Distress Detection Mechanisms

Based on the precision, recall, F-measure and accuracy results above, the DDDM-$FB_{Reactions}$ using DIWC technique with the significant Facebook reactions outperformed DDDM-B which is built with DIWC alone.

Contrary to the overall finding, individual emotion showed different results where sadness and fear had scored higher accuracy with DDDM-FBReactions, but anger showed declined accuracy. One factor which has affected the accuracy is the number of significant Facebook reaction for each emotion. In this study, both sadness and fear have 2 significant reactions, whereas anger has only one significant reaction. The regression analysis was performed to identify the strength of the effect of the Facebook reactions (independent variables) on the emotions (dependent variable) presented in a Facebook post. Thus, the number of independent variables (i.e. Facebook reactions) towards a post affects the intensity of the emotion in the post, which could affect the accuracy of a particular emotion.

Another factor which could have impacted the accuracy of anger is the dataset size. Specific to the research problem and the targeted statistical mechanism over which the test is focused, accuracy has been often determined by the size of the sample. Therefore, with these other factors taken into consideration as the sample size increases, the sampling deviation decreases or will become more accurate. Alternately, as the sample size increases, so does the accuracy of the predicted value.

Generally, findings indicate that including Facebook reactions improvised the accuracy of the detection mechanism with an average improvement of 13%. In conclusion, with a research and medical aspect, diabetes-disease needs a periodic assessment as advocated for kids and adults with diabetes. To this point, there have been no validated screening tools available for automated detection of diabetes distress that are feasible in routine practise (Dieter & Lauerer, 2018; Snoek et al., 2015). The advancement of automated techniques for early detection of emotional stress in the online diabetes community is essential as it provides a platform for long-term treatment and care. The latter allows people to plan accordingly while still being able to acquire practical advice and guidance as they face new challenges (Dieter & Lauerer, 2018). Prevention, early detection and proper management of distress are vitally crucial to improving the quality of life of patients as they manage diabetes.

4.4 Conclusion

This chapter presents the results of the conducted research on building a Diabetes Distress Detection Mechanism. The results of using Facebook reactions along with the DIWC technique in predicting distress levels of a post are proven to exhibit better results than the baseline mechanism with the DIWC technique alone. An analysis of 1564 Facebook posts revealed wow being closely associated with anger, fear and sadness within the online diabetes community. As far as positive reactions are considered, love appeared as the highest-ranking reaction to these posts, which was observed for fear. These findings portray that online health communities tend to support one another in times of crisis with their positive reactions.

The baseline mechanism demonstrates that the average of the posts fall under the 'low distress' category, with a percentage of 70%. The remaining posts are identified as 'No Distress,' which is 30%. Table 4.4 shows the results of the improvised mechanism, where it can be seen that 'Low Distress' increased to 70% while 'No Distress' decreased to 0%. 'Moderate stress' was found to be 30% compared to 0% of baseline. Similar to the baseline mechanism, none of the posts are classified under 'Extreme Distress.'

From the results obtained, it can be observed that using Facebook reaction predictors improved the accuracy of the model by 13% compared to the baseline model. Evolving automated techniques for early detection of emotional stress in the online diabetes community is vital because it provides a system for long-term treatment and care. Prevention, early detection and proper management of distress are essential to enhancing the quality of care of patients as they treat diabetes. This study could also be advantageous for healthcare facilities or non-governmental organisations to recognise and address patient mental health and concerns constructively.

CHAPTER 5: CONCLUSION

5.1 Overview

The main objective of this research is to develop diabetes distress detection mechanism to aid in classifying distress levels in the online diabetes community. The mechanism of classifying distress level was developed successfully using a dictionary-based technique and regressed Facebook reactions. This chapter summarizes the findings made with respect to the outcomes in the domain of diabetes and some suggestions for extended work.

5.2 A Brief Overview of the Research

The research problem addressed in this research was the lack of studies in an automated diabetes distress detection mechanism. The current method for assessing and detecting diabetes distress are in the form of surveys and questionnaires which are focused on clinical care and could only be diagnosed according to the acceptance and participation of patients. The active online diabetes community has generated rich and valuable user experience data which aided this study of developing an automated diabetes distress detection mechanism.

This study was aimed not only to develop a distress detection mechanism but also to improve and intensify the original distress score of the post using the reactions extracted from the Facebook diabetes community. Throughout the hours spent on Facebook, 44% of users spend reading posts and comments and clicking on either like, comment, share and reaction (Kaur et al., 2018; Pelletier & Horky, 2015). When users connect with the post content, they tend to click either of the buttons like, comment, share, or reaction. To acknowledge such context, it would only make sense to include these Facebook feature traits when it comes to classifying Facebook posts.

Several experimentation using different mechanism building, i.e. with and without using Facebook reactions are conducted to investigate the reaction capability in detecting and classifying the level of distress. Section 4.2 addressed the outcomes of the experiments where the results showed that the Facebook reaction incorporation improved the accuracy by 13%.

In general, the findings indicate the distress detection to improve when Facebook reaction was incorporated into the baseline model, implying that the addition of such features is capable of detecting distress more precisely. Presently no established diagnostic techniques have been viable for the automated detection of diabetes distress that are adequate. Deploying automated techniques for early detection of mental stress in the online diabetes community is critical in providing a framework for quick and reliable diabetes distress assessment and detection. This facilitates users to prepare prudently while still being able to reach out for help and emotional support as they face new challenges. Prevention, early detection and proper management of distress are crucial with improving the quality of life of individuals with diabetes.

5.3 Research Contribution

The online diabetes community has reported an increased presence of diabetes distress, implying that there is a need to address diabetes distress issue (Fisher et al., 2015). As discussed in the earlier chapters, traditional distress detection techniques and lack of studies in automated detection mechanism lead to the main contribution which is the development of automated diabetes distress detection mechanism. Ideally, this study hopes to start as a step forward to initiate the development of automated distress detection mechanism and produce further research involving the online health community. This development is supported by the formulation of an equation for diabetes distress detection using the dictionary-based technique. This formula is further enhanced by including significant Facebook reaction using linear regression technique, thus contributing to the second contribution.

The evolution of automated techniques for early detection of mental stress in the online diabetes community is crucial as it paves the way for long term treatment and care. The latter helps people to plan ahead while still being able to obtain practical advice and support as they face new challenges (Dieter & Lauerer, 2018). In order to improve the quality of life of patients as they manage diabetes, prevention, early detection and proper management of distress are crucial. This study could also be beneficial for healthcare organizations or NGOs to recognize and address patient concerns proactively.

5.4 Limitations and Future work

There are few limitations present in this research work which creates a path for future enhancement and improvisations. The dataset used in this research was from a prior study, therefore resulting in a small sample size for experimentation purposes. This is because most of the cleaning, pre-processing and human annotation is done beforehand, leaving very limited sample selection choice. Furthermore, human annotation for evaluation purposes requires time and incurs a high cost which caused only a selected number of posts to be annotated for further experiments. Second, this study has used the Indico API to perform emotion analysis, with only five emotions detected. Future work can broaden the scope to use other wheels of emotions such as Ekman's six basic emotions (i.e. Anger, Disgust, Fear, Happiness, Sadness, Surprise) (Ekman & Friesen, 1986) or Plutchik 's eight emotions (joy, trust, fear, surprise, sadness, anticipation, anger, and disgust) (Plutchik, 2003). Another limitation of the improvised mechanism was the partial use of Facebook features, where other features such as likes, comments and shares could also be incorporated in order to compare the accuracy and implications of the mechanism developed. The final limitation is the confined use of other techniques or algorithms such as machine learning techniques while developing the mechanism for detecting distress for further comparison and enhancement.

Future enhancements that can be brought into this research are by modelling the impact of the other Facebook features such as likes, comments and share, which are the prognostic factors on the distress detection. The effect of these features on the outcome could be observed by comparing each factor individually on the mechanism and examining the results obtained. This approach of using multiple Facebook features has been done by Kaur et al. (2018) and Meire et al. (2016). Further investigation on distress detection using various machine learning techniques and algorithms are also in the pipeline where the best technique can be identified. More analysis is also intended to help us identify other disguised features within the content that could result in increased classification accuracy and it would be of significance to better understand the emotion of humans in the era of social media communication.

REFERENCES

Abirami, A. M., & Gayathri, V. (2017). A survey on sentiment analysis methods and approach. In *2016 Eighth International Conference on Advanced Computing (ICoAC)* (pp. 72–76). https://doi.org/10.1109/ICoAC.2017.7951748

Agarwal, S., Mittal, N., & Sureka, A. (2018). *Potholes and bad road conditions: mining Twitter to extract information on killer roads*. https://doi.org/10.1145/3152494.3152517

Ahmad, M. I., Mubin, O., & Orlando, J. (2017). Adaptive Social Robot for Sustaining Social Engagement during Long-Term Children–Robot Interaction. *International Journal of Human–Computer Interaction*, *33*(12), 943–962. https://doi.org/10.1080/10447318.2017.1300750

Akter, S., & Aziz, M. T. (2016). Sentiment analysis on facebook group using lexicon based approach. In *2016 3rd International Conference on Electrical Engineering and Information Communication Technology (ICEEICT)* (pp. 1–4). https://doi.org/10.1109/CEEICT.2016.7873080

Al-Naggar, R., Osman, M., Ismail, N., Ismail, Z., Noor, N., Ibrahim, N., … Selamat, M. I. (2017). Diabetes Mellitus among Selected Malaysian Population: A Cross-Sectional Study. *International Journal of Medical Research & Health Sciences*, *6*, 1–11.

Alotaibi, A., Perry, L., Gholizadeh, L., & Al-Ganmi, A. (2017). Incidence and prevalence rates of diabetes mellitus in Saudi Arabia: An overview. *Journal of Epidemiology and Global Health*, *7*(4), 211–218. https://doi.org/https://doi.org/10.1016/j.jegh.2017.10.001

Balazs, J. A., & Velásquez, J. D. (2016). Opinion Mining and Information Fusion: A survey. *Information Fusion*, *27*, 95–110. https://doi.org/https://doi.org/10.1016/j.inffus.2015.06.002

Bandhakavi, A., Wiratunga, N., Padmanabhan, D., & Massie, S. (2017). Lexicon based feature extraction for emotion text classification. *Pattern Recognition Letters*, *93*, 133–142. https://doi.org/https://doi.org/10.1016/j.patrec.2016.12.009

Beiter, R., Nash, R., McCrady, M., Rhoades, D., Linscomb, M., Clarahan, M., & Sammut, S. (2015). The prevalence and correlates of depression, anxiety, and stress in a sample of college students. *Journal of Affective Disorders*, *173*, 90–96. https://doi.org/10.1016/j.jad.2014.10.054

Benjamin, C. L., O'Neil, K. A., Crawley, S. A., Beidas, R. S., Coles, M., & Kendall, P. C. (2010). Patterns and predictors of subjective units of distress in anxious youth. *Behavioural and Cognitive Psychotherapy*, *38*(4), 497–504. https://doi.org/10.1017/S1352465810000287

Berry, E., Lockhart, S., Davies, M., Lindsay, J. R., & Dempster, M. (2015). Diabetes distress: understanding the hidden struggles of living with diabetes and exploring intervention strategies. *Postgraduate Medical Journal*, *91*(1075), 278–283. https://doi.org/10.1136/postgradmedj-2014-133017

Bhaskar, J., Sruthi, K., & Nedungadi, P. (2015). Hybrid Approach for Emotion Classification of Audio Conversation Based on Text and Speech Mining. *Procedia Computer Science*, *46*, 635–643. https://doi.org/https://doi.org/10.1016/j.procs.2015.02.112

Blachnio, A., Przepiorka, A., & Pantic, I. (2015). Internet use, Facebook intrusion, and depression: Results of a cross-sectional study. *European Psychiatry : The Journal of the Association of European Psychiatrists*, *30*(6), 681–684. https://doi.org/10.1016/j.eurpsy.2015.04.002

Bravo-Marquez, F., Frank, E., & Pfahringer, B. (2016). Building a Twitter opinion lexicon from automatically-annotated tweets. *Knowledge-Based Systems*, *108*, 65–78. https://doi.org/https://doi.org/10.1016/j.knosys.2016.05.018

Cho, N. H., Shaw, J., Karuranga, S., Huang, Y., da Rocha Fernandes, J. D., Ohlrogge, A., & Malanda, B. (2018). IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045. *Diabetes Research and Clinical Practice*, *138*. https://doi.org/10.1016/j.diabres.2018.02.023

DeFronzo, R. A., Ferrannini, E., Groop, L., Henry, R. R., Herman, W. H., Holst, J. J., … Weiss, R. (2015). Type 2 diabetes mellitus. *Nature Reviews. Disease Primers*, *1*, 15019. https://doi.org/10.1038/nrdp.2015.19

Dehghani, M., Johnson, K. M., Garten, J., Boghrati, R., Hoover, J., Balasubramanian, V., … Parmar, N. J. (2017). TACIT: An open-source text analysis, crawling, and interpretation tool. *Behavior Research Methods*, *49*(2), 538–547. https://doi.org/10.3758/s13428-016-0722-4

Dieter, T., & Lauerer, J. (2018). Depression or Diabetes Distress? *Perspectives in Psychiatric Care*, *54*(1), 84–87. https://doi.org/10.1111/ppc.12203

Effrosynidis, D., Symeonidis, S., & Arampatzis, A. (2017). *A Comparison of Pre-processing Techniques for Twitter Sentiment Analysis*. https://doi.org/10.13140/RG.2.2.18230.60484

Ekman, P., & Friesen, W. V. (1986). A new pan-cultural facial expression of emotion.

Motivation and Emotion, *10*(2), 159–168. https://doi.org/10.1007/BF00992253

Fenwick, E. K., Rees, G., Holmes-Truscott, E., Browne, J. L., Pouwer, F., & Speight, J. (2018). What is the best measure for assessing diabetes distress? A comparison of the Problem Areas in Diabetes and Diabetes Distress Scale: results from Diabetes MILES-Australia. *Journal of Health Psychology*, *23*(5), 667–680. https://doi.org/10.1177/1359105316642006

Firmin, R., Bonfils, K., Luther, L., Minor, K., & Salyers, M. (2016). Using Text-Analysis Computer Software and Thematic Analysis on the Same Qualitative Data: A Case Example. *Qualitative Psychology*, *4*. https://doi.org/10.1037/qup0000050

Fisher, L., Polonsky, W., Hessler, D., Masharani, U., Blumer, I., Peters, A., … Bowyer, V. (2015). Understanding the sources of diabetes distress in adults with type 1 diabetes. *Journal of Diabetes and Its Complications*, *29*. https://doi.org/10.1016/j.jdiacomp.2015.01.012

Gahlan, D., Rajput, R., Gehlawat, P., & Gupta, R. (2018). Prevalence and determinants of diabetes distress in patients of diabetes mellitus in a tertiary care centre. *Diabetes & Metabolic Syndrome*, *12*(3), 333–336. https://doi.org/10.1016/j.dsx.2017.12.024

Gratch, J., Arstein, R., Lucas, G., Stratou, G., Scherer, S., Nazarian, A., … Morency, L. (2014). *The Distress Analysis Interview Corpus of human and computer interviews.*

Kaur, W., Balakrishnan, V., Rana, O., & Sinniah, A. (2018). Liking, Sharing, Commenting and Reacting on Facebook: User behaviors' impact on Sentiment Intensity. *Telematics and Informatics*. https://doi.org/10.1016/j.tele.2018.12.005

Kharroubi, A. T., & Darwish, H. M. (2015). Diabetes mellitus: The epidemic of the century. *World Journal of Diabetes*, *6*(6), 850–867.

https://doi.org/10.4239/wjd.v6.i6.850

Kiyimba, N., & O'Reilly, M. (2017). The clinical use of Subjective Units of Distress scales (SUDs) in child mental health assessments: a thematic evaluation. *Journal of Mental Health (Abingdon, England)*, 1–6. https://doi.org/10.1080/09638237.2017.1340616

Krebs, F., Lubascher, B., Moers, T., Schaap, P., & Spanakis, G. (2017). *Social Emotion Mining Techniques for Facebook Posts Reaction Prediction*. https://doi.org/10.5220/0006656002110220

Lašaite, L., Dobrovolskienė, R., Danytė, E., Stankutė, I., Ražanskaitė-Virbickienė, D., Schwitzgebel, V., … Verkauskienė, R. (2016). Diabetes distress in males and females with type 1 diabetes in adolescence and emerging adulthood. *Journal of Diabetes and Its Complications*, *30*. https://doi.org/10.1016/j.jdiacomp.2016.08.013

Lee, P.-J., Hu, Y.-H., & Lu, K.-T. (2018). Assessing the helpfulness of online hotel reviews: A classification-based approach. *Telematics and Informatics*, *35*(2), 436–445. https://doi.org/https://doi.org/10.1016/j.tele.2018.01.001

Liu, B., & Zhang, L. (2013). A survey of opinion mining and sentiment analysis. *Mining Text Data*, 415–463. https://doi.org/10.1007/978-1-4614-3223-4_13

Mandal, S., & Gupta, S. (2016). *A Lexicon-based text classification model to analyse and predict sentiments from online reviews*. https://doi.org/10.1109/ICCECE.2016.8009549

Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, *5*(4), 1093–1113. https://doi.org/https://doi.org/10.1016/j.asej.2014.04.011

Meire, M., Ballings, M., & Van den Poel, D. (2016). The added value of auxiliary data in sentiment analysis of Facebook posts. *Decision Support Systems*, *89*, 98–112. https://doi.org/https://doi.org/10.1016/j.dss.2016.06.013

Ortigosa, A., Martin, J., & Carro, R. (2014). Sentiment analysis in Facebook and its application to e-learning. *Computers in Human Behavior*, *31*, 527–541. https://doi.org/10.1016/j.chb.2013.05.024

Parekh, P. R., & Patil, M. M. (2017). Clinical depression detection for adolescent by speech features. In *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)* (pp. 3453–3457). https://doi.org/10.1109/ICECDS.2017.8390102

Pelletier, M., & Horky, A. (2015). Exploring the Facebook Like: a product and service perspective. *Journal of Research in Interactive Marketing*, *9*. https://doi.org/10.1108/JRIM-09-2014-0059

Perrin, A. (2015). *Social Media Usage: 2005-2015: 65% of Adults Now Use Social Networking Sites--a Nearly Tenfold Jump in the Past Decade*. Pew Research Trust. Retrieved from https://books.google.com.my/books?id=OupAnQAACAAJ

Plutchik, R. (2003). *Emotions and life: Perspectives from psychology, biology, and evolution. Emotions and life: Perspectives from psychology, biology, and evolution.* Washington, DC, US: American Psychological Association.

Pool, C., & Nissim, M. (2016). Distant supervision for emotion detection using Facebook reactions.

Rana, R., Latif, S., Gururajan, R., Gray, A., Mackenzie, G., Humphris, G., & Dunn, J. (2019). *Automated Screening for Distress: A Perspective for the Future*.

Ravi, K., & Ravi, V. (2015). A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. *Knowledge-Based Systems*, *89*(Supplement C), 14–46. https://doi.org/https://doi.org/10.1016/j.knosys.2015.06.015

Rezende, M., Loguercio, A. D., Kossatz, S., & Reis, A. (2016). Predictive factors on the efficacy and risk/intensity of tooth sensitivity of dental bleaching: A multi regression and logistic analysis. *Journal of Dentistry*, *45*, 1–6. https://doi.org/10.1016/j.jdent.2015.11.003

Ringeval, F., Pantic, M., Schuller, B., Valstar, M., Gratch, J., Cowie, R., … Schmitt, M. (2017). *AVEC 2017: Real-life Depression, and Affect Recognition Workshop and Challenge*. https://doi.org/10.1145/3133944.3133953

Rout, J., Choo, K.-K., Dash, A., Bakshi, S., Jena, S., & Williams, K. (2017). A model for sentiment and emotion analysis of unstructured social media text. *Electronic Commerce Research*, *18*. https://doi.org/10.1007/s10660-017-9257-8

Rus, H. M., & Cameron, L. D. (2016). Health Communication in Social Media: Message Features Predicting User Engagement on Diabetes-Related Facebook Pages. *Annals of Behavioral Medicine*, *50*(5), 678–689. https://doi.org/10.1007/s12160-016-9793-9

Salas Zarate, M., Medina, J., Lagos-Ortiz, K., Luna Aveiga, H., Rodríguez-García, M., & Valencia-García, R. (2017). Sentiment Analysis on Tweets about Diabetes: An Aspect-Level Approach. *Computational and Mathematical Methods in Medicine*, *2017*, 1–9. https://doi.org/10.1155/2017/5140631

Schinckus, L., Dangoisse, F., Van den Broucke, S., & Mikolajczak, M. (2017). When knowing is not enough: Emotional distress and depression reduce the positive effects of health literacy on diabetes self-management. *Patient Education and Counseling*, *101*. https://doi.org/10.1016/j.pec.2017.08.006

Schinckus, L., Dangoisse, F., Van den Broucke, S., & Mikolajczak, M. (2018). When knowing is not enough: Emotional distress and depression reduce the positive effects of health literacy on diabetes self-management. *Patient Education and Counseling*, *101*(2), 324–330. https://doi.org/10.1016/j.pec.2017.08.006

Schmitt, A., Reimer, A., Kulzer, B., Haak, T., Gahr, A., & Hermanns, N. (2015). Negative association between depression and diabetes control only when accompanied by diabetes-specific distress. *Journal of Behavioral Medicine*, *38*(3), 556–564. https://doi.org/10.1007/s10865-014-9604-3

Seih, Y.-T., Beier, S., & Pennebaker, J. (2016). Development and Examination of the Linguistic Category Model in a Computerized Text Analysis Method. *Journal of Language and Social Psychology*, *36*. https://doi.org/10.1177/0261927X16657855

Settanni, M., & Marengo, D. (2015). Sharing Feelings Online: Studying Emotional Well-Being via Automated Text Analysis of Facebook Posts. *Frontiers in Psychology*, *6*, 1045. https://doi.org/10.3389/fpsyg.2015.01045

Sidhu, R., & Tang, T. S. (2017). Diabetes Distress and Depression in South Asian Canadians with Type 2 Diabetes. *Canadian Journal of Diabetes*, *41*(1), 69–72. https://doi.org/10.1016/j.jcjd.2016.07.008

Snoek, F. J., Bremmer, M. A., & Hermanns, N. (2015). Constructs of depression and distress in diabetes: time for an appraisal. *The Lancet. Diabetes & Endocrinology*, *3*(6), 450–460. https://doi.org/10.1016/S2213-8587(15)00135-7

Sturt, J., Dennick, K., Due-Christensen, M., & McCarthy, K. (2015). The detection and management of diabetes distress in people with type 1 diabetes. *Current Diabetes Reports*, *15*(11), 101. https://doi.org/10.1007/s11892-015-0660-z

Wu, B., & Peng, Y. (2015). *Sentiment Analysis in the Online Health Community*. https://doi.org/10.2991/ic3me-15.2015.254

Yadollahi, A., Shahraki, A. G., & Zaiane, O. R. (2017). Current State of Text Sentiment Analysis from Opinion to Emotion Mining. *ACM Comput. Surv.*, *50*(2), 25:1--25:33. https://doi.org/10.1145/3057270

Yang, F.-C., Lee, A. J. T., & Kuo, S.-C. (2016). Mining Health Social Media with Sentiment Analysis. *Journal of Medical Systems*, *40*(11), 236. https://doi.org/10.1007/s10916-016-0604-4

Yi, K., & Yu, Y. (2018). *Analysis about the Text Contents Effect on the Usefulness of Online Comments based on Search Type Products*. https://doi.org/10.2991/icmess-18.2018.340