# Empowerment of Breast Cancer Patients Using Social Media: Analysis of Breast Cancer Facebook Pages Using Supervised Machine Learning Approaches

by Ewa Przybył

Master's thesis submitted for the degree of

MSc in Business Administration and Information Systems

cand.merc.(it.) E-business

Supervisor: Raghava Rao Mukkamala

Number of pages: 71

Number of characters: 130,547

Hand-in date: 15.03.2018

# Abstract

Breast cancer is the most frequently diagnosed type of cancer among women and has the highest survival rate among all cancer. Breast cancer treatment has a negative effect on upper extremity function, which adversely impacts patients' quality of life. Therefore, in order to fully recover, patients need to undergo rehabilitation therapy. Nevertheless, due to intense multimodality therapy, patients also suffer from cancer-related fatigue and diminished self-efficacy, which become the main obstacles to starting rehabilitation. In order to find a solution to this problem, scholars developed Rehabilitation Empowerment Model, which prioritizes the role of rehabilitation practitioner in enhancing self-efficacy through empowerment. Nevertheless, patients report being unaware of necessity of rehabilitation and existing opportunities, which should be communicated to them by healthcare practitioners. Since it turns out that breast cancer patients often seek support online before they undergo the treatment, this thesis investigates whether social media discussions can compliment the role of rehabilitation practitioner in enhancing self-efficacy through empowerment. In order to do it, the data in form of text corpus was fetched from four most popular Facebook pages devoted to breast cancer. Using supervised machine learning the text was labelled into number of constructs of domain-specific models for text classification, which were created based on relevant theories of self-efficacy and empowerment. The analysis was based on the results obtained from text classification. It turns out that both user comments and admin post do reflect constructs of Self-efficacy Development Model, Perceived Self-efficacy Model and Model of Psychological Empowerment. These and other insights serve as useful information for healthcare sector to explore the potential of using social media for informative purposes but also empowering patients to take control over post-treatment management of their illness.

# Table of Contents

# Introduction

## Context

Breast cancer is said to be the most frequently diagnosed type of cancer among women worldwide. Based on U.S. Breast Cancer Statistics (2018), “about 1 in 8 U.S. women (about 12.4%) will develop invasive breast cancer over the course of her lifetime”. Fortunately, breast cancer became a chronic disease rather than fatal one. Medical development and therapeutic advances, have contributed to the increase of survival rates in the last couple of decades. Just in the United States, the survival rates have increased from 63% in 1960s to 90% in the year of 2011. Although, the rates has been increasing dramatically, so has been the number of women who need to be aware of and undergo post-cancer treatments.

Breast cancer survivors suffer from side effects of cancer and its multimodality therapy, which cause long-term problems and require a range of rehabilitation methods. Significant part of the treatment of breast cancer constitutes the number of invasive surgeries. First of all, in order to assess the signs of disease, patients receive lymph node biopsy (SLNB) or lymph node dissection (ALND). Afterwards, to remove detected breast tumor, doctors need to mastectomy or lumpectomy (Velonesi et al. 2002). Unfortunately, these procedures cause pain and physical impairments. As listed by Yim Loh & Musa (2015) commonly reported effects of such surgeries include “upper body symptoms such as shoulder functions, breast/arm swelling (or lymphedema) with deformity, impairment of functionality, physical discomfort and numbness of the skin on upper arm and impaired arm” (p. 2). Consequently, these impairments limit women in performing everyday life activities and leisure tasks. These physical deficits can be both prevented and reduced if a patient ongoing breast cancer adhere to rehabilitation program given that she is aware of its necessity and opportunities.

As suggested by Loh (2015), “comprehensiveness and effectiveness of post-operative breast cancer rehabilitation should consider patients' self-management approaches towards lifestyle redesign, and incorporate health promotion aspects” (p. 1). Patients’ self-management skills

are however often buffered by diminished sense of self-efficacy caused by emotional distress which results from painful treatment, above-mentioned physical impairments, body-shaming or even concern of cancer recurrence. One of the most common side effects is *cancer-related fatigue*, which appears in form of tiredness, weakness, and lack of energy (Patarca-Montero, 2004). Therefore, due to physical and emotional difficulties, women tend to lack self-belief in adhering tasks that appear as slightly more challenging (such as physiotherapy). This requires new empowerment strategies, which will help them believe in their task accomplishment abilities (Fisher & Howell, 2010).

Before breast cancer patients are provided with guidelines by healthcare professionals, they often seek information regarding coping with the illness from experiences of other patients. Studies report that chronic disease sufferers take advantage of the access to the Internet and opportunity to find communities of people, who are in the same situation. The most prevalent source of mutual support are social media, which serve as a perfect environment for exchanging health information and suggestions regarding management of the disease (Meroli, Gray & Martin-Sanchez, 2013). Social media such as Facebook and Twitter as well as Internet forums play an important role in enabling people to communicate with each other, share knowledge and experiences. Facebook (www.Facebook.com) appears to be the most popular online community, which also makes it second most visited site right after Google (www.Google.com) (Bender, Jimenez-Marroquin & Jadad, 2011). The social network turned out to be an effective tool for rising awareness, fund rising as well as support seeking. Although Facebook appears to be the most popular among young adults, the fastest growing age group of users is women in the age above 55 years old, which is exactly the age group of women who are most likely to be diagnosed with breast cancer (Ibid.).

Phenomenon of social media and its infinite number of individual, user-centric and collaborative experiences shared among users, made it become an important source of information about social behavior (Olshannikova et al. 2017). More specifically, online communication networks such as Facebook make it possible to retrieve large amounts of public, human-generated data for research purposes. By using machine learning and data

processing techniques, large volumes of social data, which include details of people's daily activity on social media, can be conceptualized and transformed into valuable insights.

# Background

Mikkelsen et al. (2008) investigated the main psychosocial needs of cancer patients after they get discharged from the hospital. As mentioned by the scholars "the cancer survivor often has psychosocial rehabilitation needs which are met neither by the healthcare system nor by others". Based on the interviews, it has been reported that patients are not provided with any information about post-treatment side effects, rehabilitation needs or available opportunities. Moreover, the interviewees voiced the need of joining a support group in order to acquire relevant information on post-illness self-management. After the extensive analyses of the interviewees of the study, the researchers identified several needs of cancer patients (in general), which also included needs for continuous support and information about rehabilitation opportunities as well as needs for social support.

Nowadays, in order to receive information and social support, one only needs the access to the Internet. Undoubtedly, a great source of easily accessible information is Facebook. Using this social network one can follow a lot of breast cancer-related Facebook pages. Most of them aim to increase awareness of the disease, inform women about the risk factors, educate women self-examination steps as well as raise funds for further research or support patients in need of financial aid.Some of the most popular pages are (a) The Breast Cancer Site with over 6 million followers; (b) Breast Cancer Now with more than 646.000 followers; (c) Breast Cancer Support which number of followers exceeds 309.000 people (d) Breast Cancer Care followed by over 153.000 users[1]. On top of promoting ways of supporting breast cancer research or sharing relevant information with their followers, they also provide a lot of valuable insights about very specific aspects such as types and effects of treatment or even recommendations for special lingerie for women after mastectomy as well as beauty products

[1] Based on on-page statistics from March 5, 2018

for women suffering from physical side effects. Sadly, while scrolling through the pages' walls, one can find hardly any rehabilitation-related posts, which lead us to the assumption that rehabilitation seems to be neglected.

Although lack of post-treatment rehabilitation-related posts is clearly visible on social media, one can find a lot of information on how individuals cope with the illness. Nevertheless, as mentioned before many side-effects experienced by breast cancer survivors simply demotivate them from initiating rehabilitation process. These barriers do not only contain of the number of physical deficits mentioned in the previous section, but also various psychological impairments such as decreased self-esteem or loss of a sense of control. Subsequently, deficits become factors which diminish self-efficacy, which is a personal judgement on "how well one can execute courses of action required to deal with prospective situations" (Bandura, 1982, p. 37). Nonetheless, self-efficacy can be enhanced by empowerment, which is the result of support received by an individual (Fisher & Howell, 2010). Based on theories, which supported these assumptions, in order to put them into practice, Fisher & Howell (2010) developed a model that serves to define the role of a rehabilitation practitioner in enhancing self-efficacy through empowerment, which shall result in maximised participation of patients in rehabilitation activities. Yet given that after the hospital discharge, patients are not even likely to receive information on rehabilitation possibilities, the effectiveness of the model, might be limited, as in order for it to take effect, it requires the key component, which is *rehabilitation practitioner*. However, considering the amount of people following breast cancer awareness pages, there is no doubt that patients do seek support and information online. Thus, Facebook might be a great channel to develop complementary activities, which will empower breast cancer survivors to begin rehabilitation. However, in order to test the potential of the model, one shall look into existing discussions among breast cancer patients and find out how the main constructs of the Rehabilitation Empowerment Model, which are self-efficacy and empowerment, can be expressed and measured.

The first step to do it is the extraction of knowledge from data, which is social text. Since the aforementioned breast cancer awareness Facebook pages have a great amount of followers,

one can expect satisfying user engagement, which can sustain collection of significant sample of text for the analysis and valid results. One of the ways to fetch the text data from Facebook walls, such as page admin posts and comments, as well as user posts and comments is the use of social data analytics tool (SODATO) (Hussein et al. 2014). Once the data is fetched, the data shall be conceptualised and classified using machine learning. One of the tools, which is perfect for performing domain-specific text mining and multiclass classification is multi-dimensional tool for text classification (MUTATO). It enables researchers to use domain-specific models and trains the data themselves (See Appendix A).

In order to create valuable domain-specific models, they shall be based on the theories, that will facilitate conceptualisation of data and assign data texts into specific labels. As the main subjects of analysis are self-efficacy and empowerment, one shall look deeper into those concepts in order to find a way of measuring them. Bandura (1977), psychologist who is a pioneer in self-efficacy research, lists four ways to develop self-efficacy: (1) performance accomplishments, (2) vicarious experiences, (3) verbal persuasion, (4) physiological state, which are created based on information acquired from other individuals. These are the most effective in introducing of new behavior when shared by individuals in similar situation. The scholar also states that self-efficacy is developed in a cognitive process, which is based on four factors (1) conception of ability, (2) social comparison influences, (3) framing feedback and perceived controllability. These constructs are individual, however, one can track them once expressed in the written form. Regarding conceptualisation of empowerment, Perkins & Zimmerman's (1995) developed theory of psychological empowerment, which keys out three components (1) interactional, (2) intrapersonal and (3) behavioral. The authors emphasize that in order for a person to become empowered, those three component shall be stimulated. In this case, text analysis can let us discover, which of the components has the strongest volume and which one needs to be given more focus in order for the patients or survivors to feel empowered. Nevertheless, person's behavior is also moderated by personality traits. Therefore, in order to learn how patients can cope with different types of information, it is worth investigating what are the most dominant personality traits among breast cancer patients.

Once it is detected how self-efficacy and empowerment can be released in online discussions, one shall investigate what are the current biggest barriers for physical exercise. This will provide useful insights on the struggles of individuals ongoing breast cancer and give an idea of what information shall be provided in order to overcome those barriers and in this way help to empower and motivate the patients to start exercising. What is more, the time the information is directed to the patient might influence its effectiveness. Therefore, it would be valuable to acquire better understanding of the audience. This can be done by exploring different steps of breast cancer journey and investigating in which one of them, breast cancer patients are the most active on social media.

# Relevance

First and foremost, post-treatment rehabilitation play a crucial role in reducing fatigue and improving physical function, safety, well-being as well as enhancing patient's participation and societal reintegration. It is therefore crucial to make sure that women are well-informed about the importance of physical activity during as well as after the treatment. What is more, since individuals undergoing breast cancer have diminished feeling of self-efficacy, it is important to take actions to restore their self-belief in being able to participate and successfully complete rehabilitation process.

One can also list several reasons why breast cancer patient rehabilitation is a trending topic in medical and academic research Since post-treatment health-care institutions have limited resources to constantly follow up the patients, mainly due to the lack of funds or staffing, researchers try to discover the ways to empower patients to develop ability of self-management. Moreover, as proved by Mikkelsen et al. (2001) breast cancer patients do seek continuous support and information about rehabilitation opportunities. They also report the need of family support as well as psychological help addressing the fear for relapse. However, the big part in their recovery plays social support which can be easily found and provided online. Therefore, it is important to make sure that the communication on social media platforms is conducted in a way that can inform, empower and motivate women

towards rehabilitation. Finally, big social data make it possible to analyse user-generated content, meaning the information obtained from the source (breast cancer patients), which serves as a great fundament for relevant findings.

# Research Question

Considering the key components of Rehabilitation Empowerment Model, the main focus of this thesis is to explore how can the concepts of Self-efficacy and Empowerment be realised from the social media conversations of breast cancer patients. To answer this research question , I decided to formulate one main proposition and four sub-propositions which will be empirically analyzed.

Analysed Facebook pages aim to provide information and support for people undergoing breast cancer. While the Rehabilitation Empowerment Model lists a *rehabilitation practitioner* as the main catalyser of self-efficacy through empowerment. With easy access to these pages, patients and survivors have an opportunity to find relevant information before they actually get to meet a rehabilitation practitioner. Therefore, my main proposition is that online discussions can serve as a complement of the role of a rehabilitation practitioner in the Model for Rehabilitation Empowerment.

- Sub-proposition 1: Breast cancer patients seek support and find support by following and actively participating in discussions on Facebook walls.
- Sub-proposition 2: Content of online discussions contain information that can potentially influence self-belief in breast cancer patients.
- Sub-proposition 3: Online discussions reflect the topics which can trigger the increase of self-efficacy within cognitive process of breast cancer.
- Sub-proposition 4: Breast cancer survivors can facilitate the process of psychological empowerment of individuals undergoing the illness.

Testing these sub-propositions will not only direct me towards answering the research question, but also finding other useful insights on patient's behavior and personality traits.

All together, it will help to point out relevant factors in development of communication strategy. Furthermore, these insights might be of great help to those who are affected by breast cancer, from patients to doctors as well as public health organisations. First of all, the analysis will lead to discovering most important needs of breast cancer patients, which will lead to valuable recommendations for healthcare institutions on what actions shall be taken or improved.

What is more, I believe that discovering various aspects of breast cancer patient behavior, will also lead to valuable findings which will benefit healthcare in understanding the needs of breast cancer patients in relation to their state at the point of receiving information. The analysis of Facebook walls can be used for improving communication with cancer patient regarding rehabilitation and providing them with information which will both empower them and increase their levels of self-efficacy, which in the end might result in higher participation in rehabilitation practices.

## Motivation

The main motivation for this study is the opportunity to make a difference and be able to draw attention and find solutions to the problem, which affects people I personally know. I believe that I can find numerous behavioral patterns, which will enable me to provide useful recommendations for health practitioners, which might lead to great developments in the approach towards breast cancer patients. Considering that this is a master's thesis paper, I seeked to find a topic, which will prove that I mastered the major topics, which I studied during the program: MSc in Business Administration and Information Systems (E-business). Research and analytical skills which I gained throughout the course of my studies, turn out to be a great asset to conduct a research, which will be reliable, insightful and based on a vast amount of social data, that can help us generalise the findings and make them applicable.

## Delimitations

Facebook pages that the data was retrieved from are public. This means that any person can view, comment or post on the page. Considering new Facebook algorithms, once a person

comments under a post, this particular post might even appear in News Feeds of that particular person's friends. All together, it is not the most suitable setting for women to share their intimate feelings regarding their fears, concerns or perceived weaknesses. However, the Facebook pages have are a valuable asset towards reaching breast cancer patients with valuable information by creating the posts. Therefore, the findings can still lead us to important conclusions regarding the adjustment of the content of various Facebook pages dedicated to issues related with breast cancer.

# Theoretical Framework

Cancer diagnosis is always a tremendous psychological shock not only for patients, but also for their relatives. There is a widespread belief that cancer is always fatal. However, early detection of cancer often leads to a complete recovery (Katherine et al. 2016). Even when diagnosed at the advanced state of cancer, breast cancer patients can live long provided that the appropriate treatment and rehabilitation process have been completed. (Ibid.) Depending on different stages of breast cancer determined at the time of diagnosis, patients require different types of therapy. Breast cancer develops predominantly in the form of a tumor located in different parts of the breast. That is why, it's treatment requires removal of a tumor through a surgical intervention.

There are three types of surgeries that are the most common in breast cancer treatment (Murawa et al, 2016). The first type is called a breast conserving therapy (BCT). It is performed when cancer is lobular (formed from dairy cells) and it is detected in the pre-invasive stage. During this procedure, only a tumor with a margin of healthy tissue is cut out. The second type is a called a saving surgery. This type of surgery is required in the invasive type of cancer (if lymph nodes are already in use or distant metastases already exist). Cancerous lesions and regional lymph nodes are then removed. The last type of the surgery is called mastectomy, which is the amputation (removal) of the whole breast. Mastectomy is performed in patients with breast cancer with the tumor appearing in more than one quarter of the breast. If breast cancer has been diagnosed at an early stage, mastectomy is effective in

more than 90 percent of women. Additional treatment, for example hormone therapy, radiotherapy and chemotherapy, increases the chances of post-mastectomy patients to avoid relapse of the cancer.

During and following the treatment, breast cancer patients are therefore exposed to a lot of mental and physical challenges (Tkaczuk et al. 2016). Long oncological therapy leads to impairment of physical fitness through weight loss, reduction of systemic capacity and physical activity, including decreased muscle activity of the shoulder girdle and lymphatic edema of the limb on the operated side. Removal of breast creates unfavorable conditions for maintaining the correct torso statics by not being able to balance its symmetrical elements. It results in postural disorders such as shoulder elevation, withdrawal of the shoulder blades, lateral curvature of the spine. The efficiency of the limb is influenced by individual characteristics, especially the tendency to hypertrophic scar formation resulting in adhesions, contractures and limited lymphatic drainage through cavities and lymph nodes. Disturbances in the torso statics particularly affect women with large breasts, which causes abnormal posture in the form of an elevation or lowering of the shoulder, protrusion of the shoulder and curvature of the spine. The procedure consists of kinesitherapy with the help of properly selected movement exercises. In the event of intraoperative damage to the nerve, muscle electrostimulation is introduced. This is most often the case after the damage of the long thoracic nerve, which leads to the withdrawal of the shoulder blade. In such situation, the anterior toothed muscle is electro-stimulated. Therefore physiotherapy is an important process that can bring women back to performing regular tasks. It aims to strengthen the muscles of the back, stimulating the muscles located in the area of the shoulder.

Physiotherapy is inevitably linked to medical rehabilitation. There are several definitions of rehabilitation recognised in medicine. British Society of Rehabilitation defines the notion as "A process of active change by which the person who becomes disable acquires the knowledge and skills needed for optimal physical, psychological and social function" (Gottenbrunner, Ward & Chamberlain, 2007). WHO organisation defines rehabilitation as the "use of all means aimed at reducing the impact of disabling and handicapping conditions and

enabling disabled people to achieve optimal social integration" (Ibid.). From these two definitions, one can distinguish two factors (1) skills and knowledge and (2) high degree of motivation. Since, it has been reported that after being discharged from the hospital, breast cancer patients seldomly get any information about rehabilitation or rehabilitation opportunities, they are deprived from acquiring necessary skills, knowledge and motivation. Therefore, it is important how these factors can be extracted elsewhere.

## Cancer-related fatigue and barriers towards physical exercise

As mentioned in the introduction, breast cancer patients suffer from many physical side effects of both surgical and pharmacological treatment of their disease. Those include physical reactions such as nausea, weakness, lymphedema and decreased range of motion (Smith-Turchyn et al. 2016). However, a log list of physical deficits also lead to common psychological side effects such as depression, anxiety, stress, mood disturbances, decreased self-esteem, loss of a sense of control, poor body image and diminished femininity, which along the way become barriers for starting rehabilitation process (Tkaczuk et al. 2016). Furthermore, one of the common problems of almost all cancer patients is so called Cancer-Related Fatigue (CRF) (Borneman, 2012). It is defined as "a persistent, subjective sense of physical, emotional, or cognitive tiredness or exhaustion related to cancer and cancer treatment that is not proportionate to activity and interferes with normal functioning" (Ibid. p. 2). The fatigue combines physical, emotional and mental components and is mainly triggered by hormonal imbalances caused by the treatment itself and stress which is the effect of cancer diagnosis and the fear of cancer recurrence. It turns out to be induced by all kinds of treatment: chemotherapeutic treatment radiotherapy as well as surgical treatment of cancer all induce fatigue, however with different patterns. Barriers to exercise can be attributed to the side effects of treatment. Therefore it is extremely important to investigate how the barriers can be overcome and how participation can be maximized.

Blaney et al. (2010) conducted a research in which they investigated what are the barriers to and facilitators of exercise among patients with cancer-related fatigue. Some of the most common barriers that were derived based on the focus group interviews were unpredictability

of fatigue and changes of its intensity from day to day. Nevertheless, next to various cognitive side effects of treatment such as lack of confidence and self-image issues, the major problem in initiating as well as facilitating physical exercise was lack of motivation. On the other hand, participants of the study reported that group support is a great facilitator of the exercise at it creates the feeling of friendship, solidarity and belonging. What is more, the research proved, that it is possible to transform the nature of fatigue through exercise, which in the end can serve as a mean of escapism. Through physical activity, patients have an opportunity to achieve sense of normality and progression from illness, which can become a motivator and facilitator to the regular exercise. transform the nature of fatigue through exercise, which in the end can serve as a mean of escapism. By performing physical activity, patients have an opportunity to achieve sense of normality and progression from illness, which can become a motivator and facilitator to the regular exercise.

# Rehabilitation Empowerment Model

As mentioned in the introduction, the problem with lack of participation of people with disabilities in rehabilitation processes has also been recognized by World Health Organisation (2001). That is why, in 2001, World Health Assembly approved the International Classification of Functioning, Disability and Health to standard to describe and measure health and disability (ICF, See Figure 1). More specifically, ICF provides a framework, which integrates various theories of disability and aims to set a common language for health practitioners in order for them to be able to use it in rehabilitation practices worldwide. Considering that disability of an individual occurs in a context, the model is based of four main principles: environmental influence, universality, neutrality as well as parity and aetiological neutrality. The main components of the model include body functions and structures. Body functions component stands for physiological and psychological functions of a body, while body structures are anatomical parts of the body such as organs. Another factor playing an important role in order for a patient to take part in an activity is participation, which is a representation of patient's involvement into certain action. Environmental factors include various circumstances of the condition such as social

attitudes, legal or social structures or even a climate change. The last significant component is wide range of personal factors which include social background, gender, age or education.

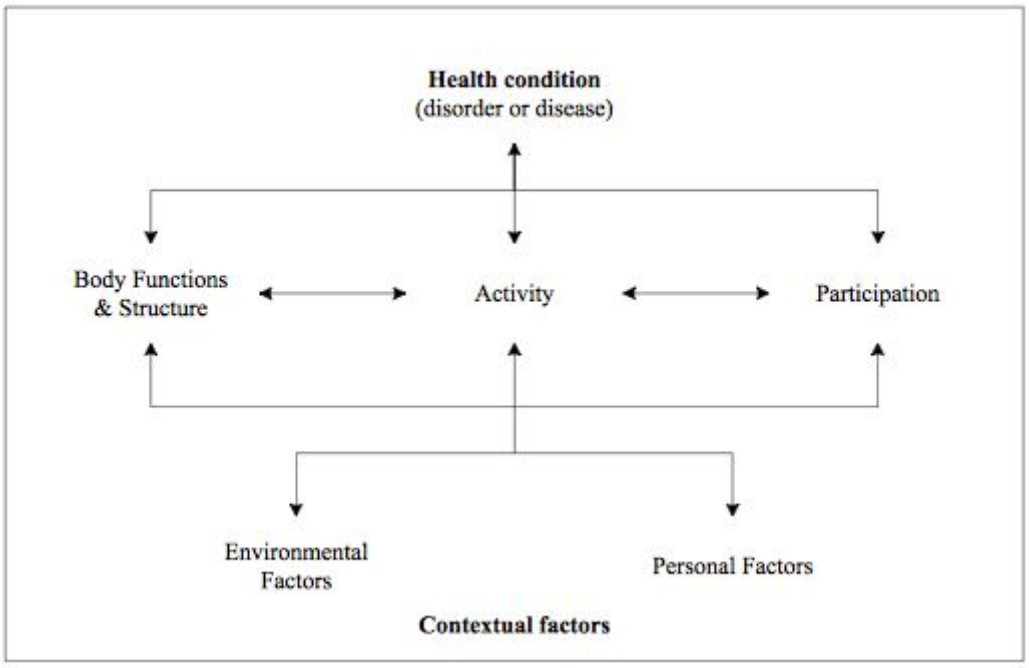



Figure 1. International Classification of Functioning, Disability and Health, World Health Organisation, 2001

ICF does not serve as a model for classification of people, but investigation of possible characteristics of certain condition of human health in the context of one's individual life situation considering the influence of the surrounding environment (McDoughal, Wright & Rosenbaum, 2010). The model, therefore, intends to improve rehabilitation practitioners' understanding not only of the physical needs of patients, but also enable them to recognize the context and state a patients is in, to be able to adopt the right approach and communication strategy.

Fisher and Howell (2010) studied the ICF model and applied breast cancer as a health condition, which further lead them to develop the Rehabilitation Empowerment Model (See Figure 2 aiming to improve self-efficacy and upper extremity function of survivors of breast cancer.

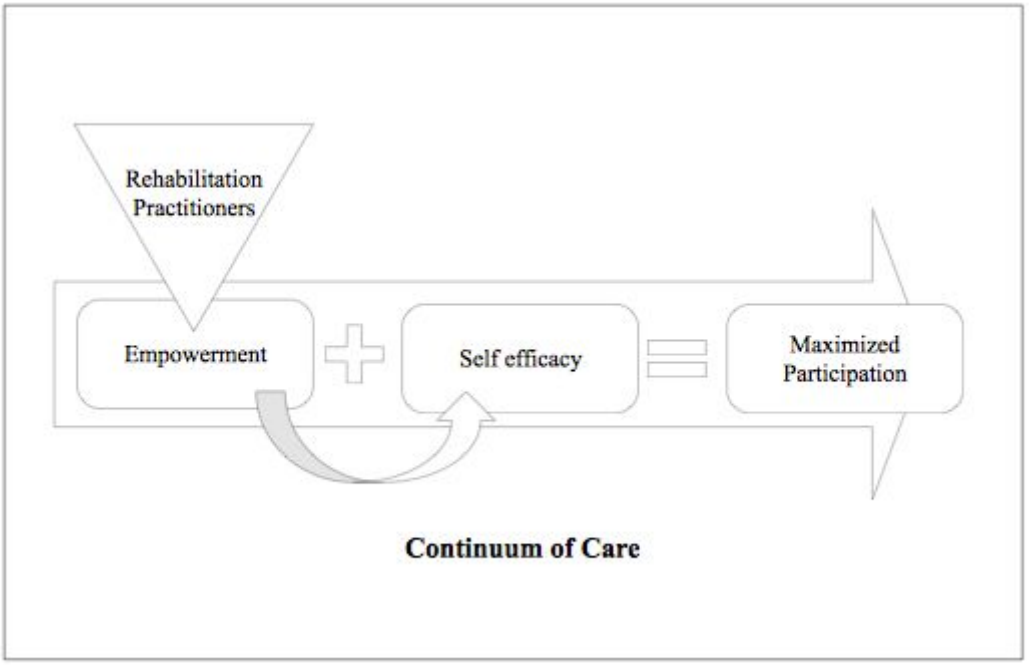



Figure 2. Rehabilitation Empowerment Model, Fisher & Howell, 2010

The above mentioned Model for Rehabilitation Empowerment is based on factors which affect ability of self-control of a patient and management of their illness, which in the light of limited healthcare resources are ideal features of breast cancer survivors. The first unique factor is *self-efficacy* which reflects the belief of ability to manage one's situation. It can heavily influence person's health, as once a person is not capable of performing a certain task, for instance physiotherapeutic exercise, he or she will lack motivation for initiating that or similar task in the future. However, based on the model, the level of *self-efficacy* can be enhanced by another factor which is *empowerment*. *Empowerment* can therefore increase patient's self confidence and positive self-belief, which might then result in performing the task.

In order to lead to the positive outcomes and desired behavior, which in this case is participation in rehabilitation, both components need to constantly interplay with each other (Fisher & Howell, 2010). Nevertheless, *empowerment* functions as the main driver that according to the model can be initiated by rehabilitation practitioners. Opposite to self-efficacy, which is processed very individually, empowerment is strictly dependent on

external factors such as social influence. Therefore, practitioners' empowerment skills and techniques are crucial for overall improvement of treatment outcomes.

The model aims to enhance the patient's ability to take control over the direction of their health instead of being totally dependent on medical care provider. Based on the model one can conclude that if a person does not believe in the possibility of reaching certain goals, he or she lacks motivation for adhering to certain behavior that would otherwise lead to reaching that goal.

Although the models are meant to serve as a framework for health practitioners, Fisher and Howell (2010) also stress the importance of social persuasion in women empowerment. As the example, they list support and awareness groups, which also carry some educational components that empower survivors by providing necessary information and therefore increase their level of self-efficacy. To test if one could transform the model into the social media framework, before the analysis of discussions conducted on Facebook walls, one shall take a closer look into the notions of self-efficacy and empowerment in order to fully understand how those can be expressed in the online discussions and conceptualized for text classification.

# Understanding Self-efficacy

Self-efficacy beliefs are a key aspect of human motivation and behavior as it influences actions that can affect one's life (Bandura, 1995). Self-efficacy does not refer to a personality characteristic; instead, individual's efficacy expectations will vary greatly depending on particular task and having given conditions or circumstances with which the efficacy judgement is associated. Bandura (1995) explains that self-efficacy "refers to beliefs in one's capabilities to organize and execute the courses of action required to manage prospective situations" (p. 2). To put it more simply, self-efficacy is what an individual believes he or she can accomplish using his or her skills under certain circumstances. The scholar stresses, that while it is a personal sense of control that facilitates a change of health behavior, cognition determines whether certain behavior will be initiated as well as how long will it last, given

possible obstacles and failures. It influences changes that people take on. Thus, people with high self-efficacy tend to set more challenging goals as they focus on opportunities and achievements, rather than obstacles and possible failures. Individuals with low self-efficacy about a particular task may ruminate about their personal deficiencies rather than think about accomplishing or attending to the task at hand; this, in turn, impedes successful performance of the task. Self-efficacy is also seen to be a task-specific version of self-esteem. The basic principle behind Self-efficacy Theory is that individuals are more likely to engage in activities for which they have high self-efficacy and less likely to engage in those they don't.

In order to explain how individuals acquire and maintain certain behavior, one shall investigate Social Cognitive Theory (SCT; Bandura, 1997), which is also relevant to health communication and health promotion. It defines how people acquire certain behavioral patterns that appear through the learning process and lists what factors influence this process. The primary assumption of the theory is however that individuals use self-reflective cognitive process to manage their behavior and adapt to necessary changes. Bandura indicated that observational learning occurs when people observe other people's behavior. The scholar argued that when people see other people rewarded for certain behavior, they are more keep to adopt the same behavior. What is more, they are more likely to imitate the behavior of people with whom they identify. More specifically, the behavior is influenced by dynamic and reciprocal interaction of three factors: environment, people and behavior.

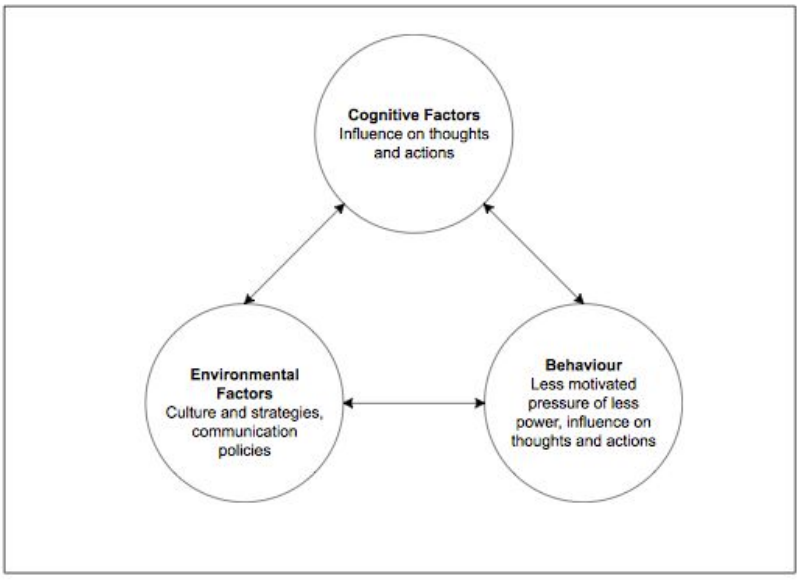



Figure 3. Social Cognitive Theory by Albert Bandura, 1989

Behavior is the physical action, more specifically, representation of responses that individual reflects based on environmental and cognitive factors which influence his decision. Cognition represents the influence of thoughts and reflection over previous similar experiences. In this construct self-efficacy towards capability of performing certain behavior plays the most important role. One should also take into consideration possible environmental aspects which can set the conditions both for decision-making as well as performing certain behavior.

Bandura (1989) emphasized that "to exercise some measure of control over one's developmental course requires, in addition to effective tools of personal agency, a great deal of social support" (p. 8). The model is basically a person who demonstrates behavior for someone else. Models can also come from media and social media as behavior does not only have to be reflected verbally but can also be described using text. Therefore, social media can create a great environment for sharing experiences between people with similar experiences and participating in discussions.

Self efficacy is also one of the core constructs in Social Cognitive Theory model of health behavior (Bandura, 1997). The model can help to determine whether the person will engage into certain behavior. Self-efficacy is said to have a direct effect on goal setting as well as can

result in indirect effects, being mediated by other constructs which are socio-structural factors and outcome expectations.

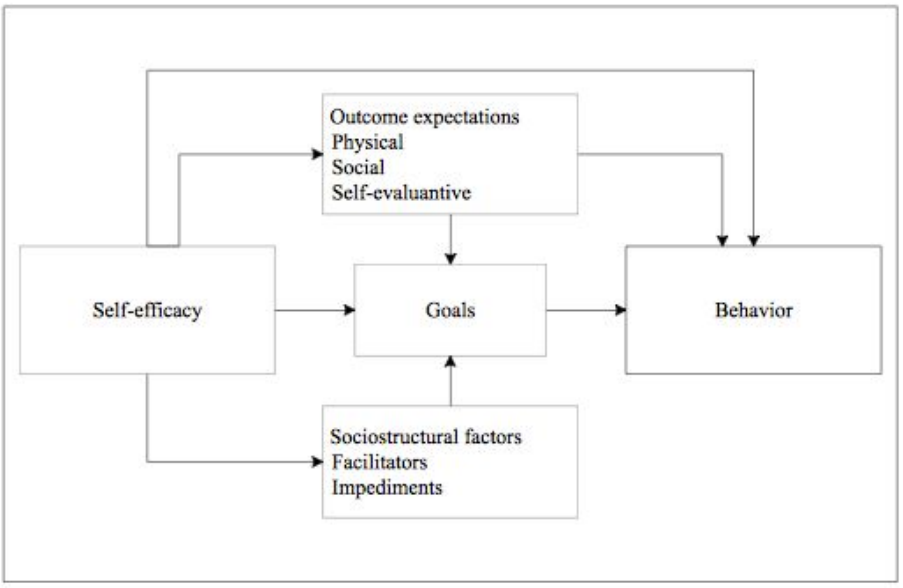



Figure 4. The Social Cognitive Theory Model of Health Behaviour, Young et al. 2014

Central assumption of the theory is one's belief that given behavior will lead to valuable and positive outcomes (Young et al. 2014). Bandura outlined three types of outcomes expectations, which are considered when deciding whether certain action should be taken. Physical outcomes include material and physical gains or losses. Social outcomes are perceived approval or disapproval of social environments. Self-evaluative outcomes are individual's judgement on how she will feel after performing a task. Self-efficacy is also mediated by sociostructural factors, which are various facilitarors and impedimets to behavior such as social support and perceived environment.

There are also ways to order a tasks by their level of difficulty. This can be determined through "self-efficacy magnitude" (Van der Bijl & Shortridge-Baggett, 2002). High-magnitude expectations lead the person to feeling capable to performing the hardest task. Self-efficacy magnitude measures the difficulty level (e.g. easy, moderate, and hard) an individual feels is required to perform a certain task (Ibid.). For instance, the patient might wonder how difficult will their exercises be or whether the exercises easy or hard? One can

also measure self-efficacy strength, which refers to the amount of conviction an individual has about performing successfully at diverse levels of difficulty (Ibid.). The patients might therefore wonder how confident are they that they can excel at their tasks and try to determine how sure are they that they can succeed in completing them. The last generality of self-efficacy refers to the "degree to which the expectation is generalized across situations" (Lunenburg, 2011). This then makes a patient question whether what she has learned will be possible to be applied in successful performance of the task.

Self-efficacy beliefs can have diverse effects on four main processes: cognitive, motivational, affective and selection processes (Bandura, 1993). Bandura (1993) stresses that "it is difficult to achieve much while fighting a self-doubt" (p. 118). As most of actions are actually in thought, so in our cognitive system, certain sense of efficacy can help people visualize the effects of their actions. If people dwell on many things that can go wrong and visualise negative actions of their behavior, they are less likely to take up a certain task, which leads them to the number of selection processes. Cognitive processes also have a function of predicting certain events through processing information, which helps to develop strategies in order to control the outcome. This can then influence motivational processes, as once outcomes are determined as positive and a person can visualise benefits of performing a certain task, it is easier for them to initiate behavior. Behavior can also be catalysed through the affective processes, which are the result of someone else taking an action.

Based on the above model, self-efficacy again turns out to be one of the most important factors in evaluation of learning process. Considering the structure SCT, one can conclude that people with high self efficacy set more challenging goals and they focus on opportunities and achievements, not on obstacles and possible failures. In terms of breast cancer rehabilitation such achievements can be an increase in the range of movements in the joints of the limbs, increase in the strength of the hand muscles, prevention or removal swelling within the limb, prevention of inflammation resulting from tissue edema (so-called lymphatic edema), prevention of postural defects and their possible correction, improvement of the body's efficiency (Gutman et al, 1990).

## Self-efficacy Development Model

Bandura (1977) also developed a model, which illustrates four main sources of information that influence individuals when judging their efficacy and help them to determine whether they are able to perform a specific task. Thus, based on these sources one can develop either high or low level of self-efficacy. These include performance accomplishments, various experiences, verbal persuasion and physiological state (or physiological feedback).

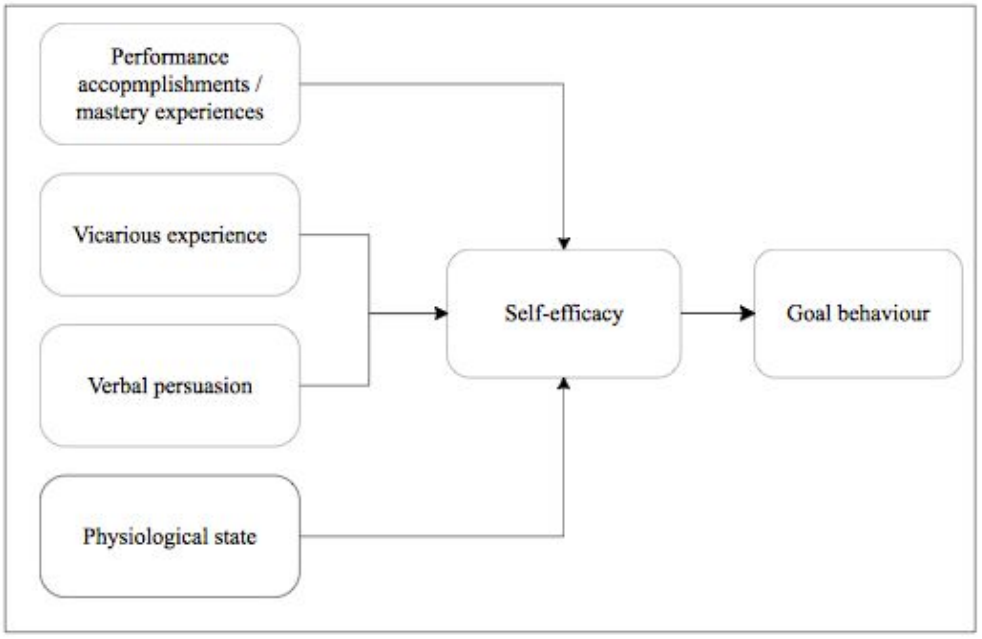



Figure 5. Self-efficacy development model, Bandura, 1977

According to the author, the most effective source is performance accomplishments also called as mastery experiences, which are the reflection of previous experiences of adherence. Bandura (1994) stresses that "a resilient sense of efficacy requires experience in overcoming obstacles through perseverance effort." (p. 2) In the discourse of success it is commonly recognized that effort is the most important determinant of success. Thus, patients who reassure themselves that they are capable of making an effort that lead them to success are less likely to adversity and quickly rebound from setbacks.

*Performance accomplishments*, which are the positive results of certain action, can serve the reflection of one's mastery influence and reinforce one's self of control. Therefore, if a person fails to deliver or complete a task, this experience can undermine her self-efficacy.

*Vicarious experiences* are the reflection of person's successful or unsuccessful attempts at certain adherence. Seeing people similar to oneself, being in a similar situation or dealing with similar problems and succeeding in performing a task can greatly increase one's level of self-efficacy. However, the greater the assumed similarity, the greater the impact and the more persuasive are the results of other people's experiences on self. Human beings are hardly ever able to objectively assess their ability to perform a task. Therefore, it is natural for them to compare themselves to peers. The people, who the patients compare their abilities to, have a big influence on judging their own abilities. That is why, it is important that the peer situation is as much alike as possible. *Social comparison* is also an important driver for people's satisfaction with their achievements, which then strengthen their self-efficacy. It however also help to develop self-regulatory mechanisms.

*Social persuasion* is yet another important source of information influencing other people's belief on whether they can or cannot succeed in performing a task. If patients are persuaded verbally that they have certain characteristics, which make them capable of performing the task, they are significantly more likely to do it. Bandura (1994) argues that "persuasive boosts in perceived self-efficacy lead people to try hard enough to succeed, they promote development of skills and a sense of personal efficacy" (p. 3). Positive efficacy builders do not only provide efficacy appraisal, but also serve to set a structure of the situations which assure the patients that a certain way of behavior will bring them expectable outcomes. It is a useful tool which helps to avoid unwanted failure and disappointing results of one's effort.

*Physiological state* is a personal interpretation of whether or not, a person is capable of adhering to certain practice. Every person is capable of judging their personal indicators for poor performance and vulnerability. In activities involving physical performance, people automatically judge the reaction of their body such as fatigue, aches and pains. This is the

sign of their physical disability which enables them to judge their personal efficacy towards performing the task as a whole. Therefore, Bandura suggests making an effort to reduce people's stress reactions, because they lead to the wrong interpretation of their physical state.

One can conclude, that overall Model for Self-efficacy Development consists of two components (social persuasion and vicarious experiences) that require other people perspective, which control sense of belief and capability of performing the task. These factors further support Bandura's argument on the importance of social support in developing one's ability to exercise self-control.

Self-efficacy theory imposes that overall efficacy expectations are the central determinant of the behavior. Christina Lee (1984) argued that outcome expectations, which consist of beliefs about whether a given behavior will lead to given outcomes, can be even more important factor. She conducted a study in which she analyzed performance in a snake-handling task with a population of 33 non-phobic man. Though again, it turned out that efficacy expectations are a better factor in predicting the performance than outcome expectations. Thus, yet this is another study that stresses the importance of one's level of self-efficacy.

## Model for Perceived Efficacy within Cognitive Process

According to Bandura (1993) perceived efficacy appears through four major process: (1) motivational, (2) affective, (3) selection and (4) cognitive. Those processes consist of conditions which influence development of self-efficacy. Patients undergoing breast cancer face many ambiguities and uncertainties regarding their life and their future. In order to construct and integrate predictive factors, they need to raise and test their own judgements towards the immediate results of prospective actions. Giving social repercussions and meeting situational demands, such as getting physically strong and getting back to the normal life, requires a strong sense of self-efficacy. Strong sense of self-efficacy can be developed through a cognitive processing of information and thought. Cognitive process consists of several stages: *conception of ability*, *social comparison influences*, *framing feedback* and *perceived controllability* (See Figure 6).

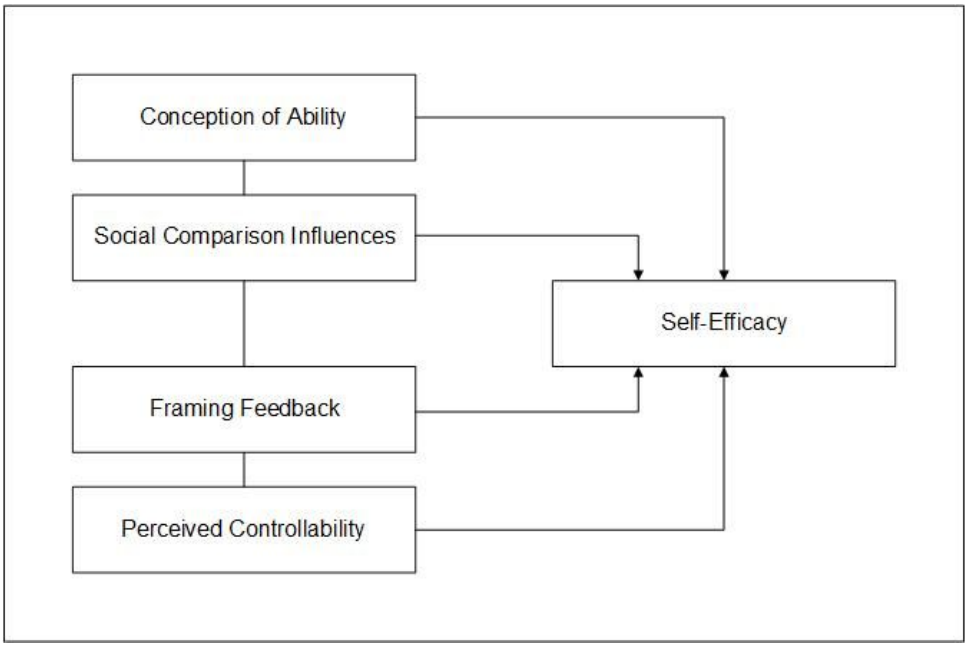



Figure 6. Model for Perceived Self-Efficacy within Cognitive Process

Human beings are hardly ever able to objectively assess their ability to perform a task (Bandura, 1993). Therefore, it is natural for them to compare themselves to peers. *Conception of ability* is developed through comparing one's abilities with other people's abilities. Thus, in order for it to result in increase of self-efficacy, it is important that the peer situation is as much alike as possible. *Social comparison* is also an important driver for people's satisfaction with their achievements as it helps to develop self-regulatory mechanisms. Receiving evaluation on one's progress is also a significant factor influencing one's perceived self-efficacy. In order for a *framing feedback* to underscore person's capabilities, it should focus on achieved progress. The last construct which helps to examine one's level of self-efficacy is *perception of controllability*. Exercise of control is based on two aspects. One of them is a level of strength of one's belief to implement and sustain certain changes, another is the modifiability of the given environment. The other aspect is possessing necessary skills to achieve desired outcome. *Perceived controllability* is therefore defined as one's judgement on ability of controlling specific situational outcomes and ability to readjust the strategy in case the primarily adopted strategy turns out to fail in bringing desirable outcomes.

## The Big-Five Personality Theory: Impact on Self-efficacy

Both human cognition and personality have an influential role in human behavior. Self-efficacy is not a personality characteristic. Lawrence (1993) defines personality as pattern of psychological behavior, such as thoughts or feelings. It determines how one will act or react to a given situation. However, it does not determine behavior itself. Behavior, on the other hand, is a result of a context and depends on the requirements of a given situation. According to Bandura (1982) self-efficacy is considered as situation-specific construct and it leads to assessment of ability. Personality, however, can influence this assessment of ability and therefore increase or decrease one's self-efficacy. For instance, a student can be good at performing all the tasks and preparing for the oral exam, however, due to a personality trait such as shyness, he might not be able to perform well at the oral exam. Similar example can also be found in the context of rehabilitation. For instance, a patient might be aware of the importance of physiotherapy and hard work that it requires to put, however, due to their pessimism (which is a personality trait), their perception focuses on the lack of positive results. That is why, it is worth analyzing what personality traits are dominant among breast cancer survivors or people undergoing breast cancer as they also have an influence on behavioral outcome. When studying personality-influenced topic, scholars often refer to Big Five personality model, also known as the five factor model (FFM) (See Figure 7).

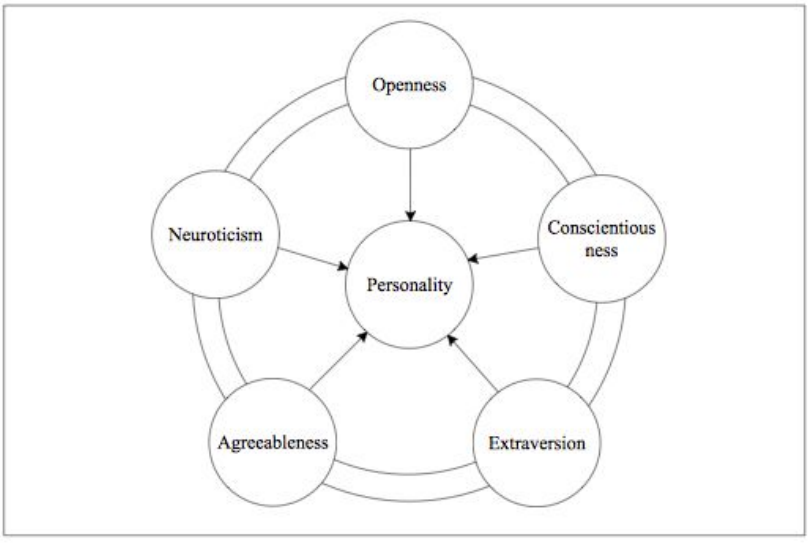



Figure 7. The Big Five Personality Theory: The five-factor model

Consciousness is a feature that reflects being open to new learnings and experiences as well as being curious, imaginative, adventurous and focused on tackling new challenges. Extraversion, which is opposite to introversion, is reflected in behaviour such as having a desire and enjoying meeting new people, making new friends, being center of attention. An extrovert is outgoing, social and talkative. Openness is defined as being intellectually curious and open to emotion and experience as well as interested in learning. Person holding this trait is open to new learnings and experiences as well as being curious, imaginative, adventurous and focused on tackling new challenges. Agreeableness includes attributes such as conforming to social norms, trust, altruism, kindness, affection, generosity, optimism and trustworthiness. Finally, neuroticism is linked to sadness, moodiness, and emotional instability. People with this trait are prone to experiencing mood swings, anxiety and irritability.

The Big Five traits are also seen as a great influence on self-efficacy in a way that they can either increase or reduce self-efficacy. Conscientiousness is a great facilitator of task engagement and effort, while agreeableness facilitates entry into new activities. Extraversion, let a person to focus on positive reactions from others, which also helps to facilitate the increase in self-efficacy. Sanchez-Cardona et al. (2012) argues that openness leads to shifting the perception of demand into a challenge. Opposite to all these traits, neuroticism can have a negative effect on self-efficacy as it increases anxiety. Furthermore, past studies of patient adherence reported the correlation of personality and adherence to prescribed medical regimen (Christensen & Smith, 1995). Costa and McCrae (1992) labeled Conscientiousness as the most accurate trait for adherence behaviour. People of this feature are highly -purposeful and self-disciplined . They also have strong will to achieve and self-control, which are ideal features, which can stimulate patient's ability of post treatment self-management and coping to medical advice such as rehabilitation.

# Understanding Empowerment

Empowerment is the term that refers to more than one phenomenon. Therefore, before one can conduct text analysis, it is important to define and clarify specific aspects of the concept and investigate how can empowerment be expressed in a textual way. For the past few decades the term was used in different discourses and in the end became established in the social-political-professional discourse all around the world. According to Sadan (1993) "empowerment integrates well into the discourse on contemporary social ideologies and values, contributes to this discourse, and provides it with an important moral criterion. (p. 31) Empowerment however can be defined as a "construct that links individual strengths and competencies, natural helping systems, and proactive behaviors to social policy and social change" (Perkins & Zimmerman, 1995, p. 569). Empowerment-oriented actions have positive influence on person's wellness through solving problems and providing opportunities to develop skills.

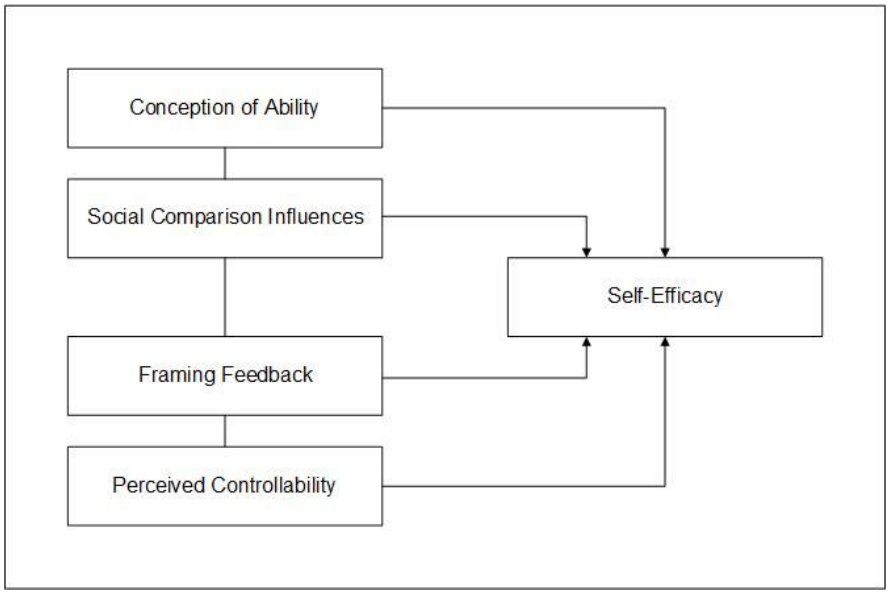



Figure 8. Model for Perceived Self-efficacy within Cognitive Process

One can distinguish between individual empowerment and group empowerment. According to Sadan (1993) "individual empowerment process occurs in a man's or a woman's life, they begin to believe that they are capable of having better control over their lives' they

understand their situation. And begin to act to improve their lives and their environment." (p. 193). The author discusses the importance of the group in the process of empowerment of the individual. Namely, while belonging to a group, participants feel equal, therefore they are braver to express their feelings. The group also enables them to develop solutions and ways to act in their own favour and towards being able to solve their problems. Moreover, as author lists, participation in the groups make people committed to a cause and "a person whose circumstances and condition led her to participate in an empowerment-encouraging group has a better chance of becoming empowered than someone who has not participated in such a group." (p. 194) The benefits of empowerment turn out to be enhanced social well-being, increased self-efficacy, becoming better informed, improved confidence in treatment, improved acceptance of their illness, and feeling more competent and in control (Coulson & Shaw, 2013).

## Psychological Empowerment Model

As mentioned by Perkins & Zimmerman (1995) it is important to keep in mind that empowerment theories cover both empowerment processes and empowerment outcomes. Empowerment processes are operationalizations of empowerment that allow us to study the results (outcomes) of empowering processes. The researchers argue that it is actions and suggestions which serve to empower, while the results are analyzed to illustrate the level to which a person has been empowered. On the individual level, empowering process might be the result of participation in a certain community. However when studying individual empowerment outcomes, one needs to keep in mind that they are affected by perceived control and resource mobilization skills. Empowerment approach goes by positive aspects of a given situation.

In order to analyze data, we will use Perkins & Zimmerman's (1995) theory of psychological empowerment, which has already been used by the number of researchers in conceptualising empowerment. The theory outlines that both empowerment outcomes and empowerment process depend on the situation of an individual patient, her personal understanding, self-management and self-control. Therefore, the outcomes are seen as individual and

context-specific. However, Zimmermann (1995) created a model of empowerment which consists of three main components which are needed in order for a person to become empowered.

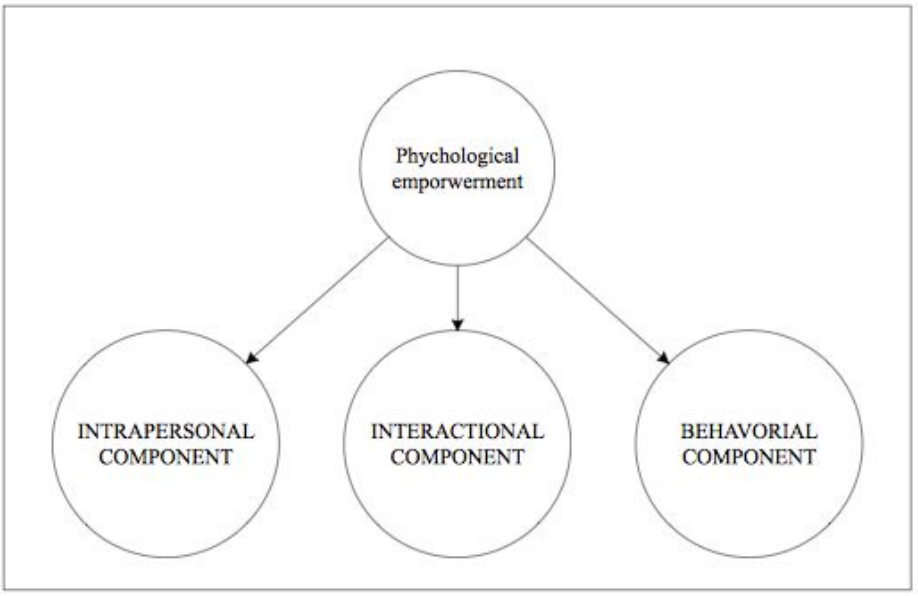



Figure 9. Psychological Empowerment Model, Perkins & Zimmermann, 1995

The first one of them, the intrapersonal component reflects individual perception of possibility for and motivation for having a mastery over their actions. Interactional component shows individual's skills and knowledge to perform certain action. Behavioral component represents actions taken by an individual to achieve or influence specific outcomes.

The notion of empowerment has always been commonly used in relation to feminism, society and politics. While studying history or recent news, one can encounter a lot of information on empowerment of certain communities, minorities or people of certain gender and sexual preferences in order to give them certain rights which can enable them to take control over their actions, which is an important factor leading to social and political harmony (Al-Haj et al., 2007). Empowerment appears also on more personal levels of human constructs, which

can relate to their adherence to the certain group such as for instance employees of a company or medical patients (Ibid.).

Patients involvement started being recognized as an important factor in health-care from several perspectives (Johnsen 2016). First of all, empowerment is seen as a social right and everybody should have opportunities to gain mastery of their lives. Secondly, empowerment and self-control can add to patient's quality of life and welfare. Lastly, limited resources of the healthcare system require more patient involvement in the successful treatment. Therefore patient empowerment became an important aspect of medical research (Britten, 2009). First it appeared in the literature as an "emergent state", followed by being described as "processes", which leaded to the "emergent states" (Jorgensen et al. 2017). In the end patient empowerment has been seen as as "behaviors" proving that patients are able to self-control themselves. "Behavior" is therefore seen as a proof of patient's skills, knowledge and control of exercising certain behavior, which reflect that they are empowered. Following this conception, "empowered patient" may or may not choose to perform a behavior that would have an impact on their health.

In 2012, World Health Organization hosted the first European Conference on Patient Empowerment. In the conference report one can find a definition of the process of empowering a patient. It is defined as "providing them with the opportunities and the environment to develop the skills, confidence and knowledge to move from being a passive recipient of care to an active partner in their health care" (p. 5).The report also includes several specific characteristics of an empowered patient such as understanding one's health condition and its effects on one's body, feeling the ability of participating in health-related decision making, feeling able to make choices about treatment, understanding the needs of making certain changes to lifestyle in order to be able to manage a health condition as well as seeking, evaluating and making use of information.

According to Chiauzzi (2016) as patient's empowerment is a crucial element triggering successful self-management of patients with chronic disease, "providers and healthcare

organizations have begun recognizing the importance of patient empowerment as a driver of patient-centered care" (p. 6). The researcher seeked to understand the concept of empowerment in relation to health information. He analysed important factors in patient empowerment and how those factors are associated with patient's characteristics. He conducted a survey which creation required identification of concepts and domains of empowerment, which are positive patient-provider interaction and knowledge and personal control. Positive patient-provider reaction can lead to providing precious knowledge and confidence in decision making. The effect of personal control is positive attitude and sense of control in performing a task. Patient empowerment also results from factors such as patient satisfaction, comprehension, and active involvement in treatment.

Within the cancer discourse, empowerment is defined as "individual's feelings of being able to manage the challenges of the cancer experience and of having a sense of control over one's life" (Maunsell, 2014)**.** Since many cancer survivors experience the fatigue and pain as well as require rehabilitation and constant control after the treatment, there are also some studies on empowerment in relation to cancer follow-up (Jorgensen et al. 2017). Those studies reviewed web-based interventions of empowering cancer survivors, which triggered Jorgensen et al. (2017) to conduct the study which aimed to find both facilitators and barriers towards patient empowerment in cancer research". One of the key facilitators of patients empowerment turned out to be knowledge which is seen as the main tool for patient's self-control and the key trigger of self-management. This condition is most commonly met after the patients are provided with information, which they find relatable and helpful. It turns out that many of them seek information online through following educational intervention groups and joining educational programs. Having an active role and being capable of advising others also turns out to be an element that accelerate patient's confidence and makes them feel empowered. This however is usually possible when patients can interact with people in significantly similar situation and when themselves they are able to get support from participating in the online group. The researchers also found out that religion and spirituality played an important role in enhancing patient empowerment.

# Breast Cancer Patient's Journey

In order to understand lived experiences of women who have been diagnosed, treated and became cancer-free as survivors, it is crucial to investigate their care pathway, which is called a patient journey. Patient journey also helps to keep the patients engaged in throughout the care continuum. Understanding different stages, can also help to determine when is the most appropriate time to initiate exercise, so therefore, in what moment of breast cancer patient journey, social media support might be the most effective.

Faustine Williams (2015) identified different stages of breast cancer journey. By conducting interviews with women ongoing breast cancer, she depicted 8 unique themes, which represent step-by-step process of dealing with this illness.

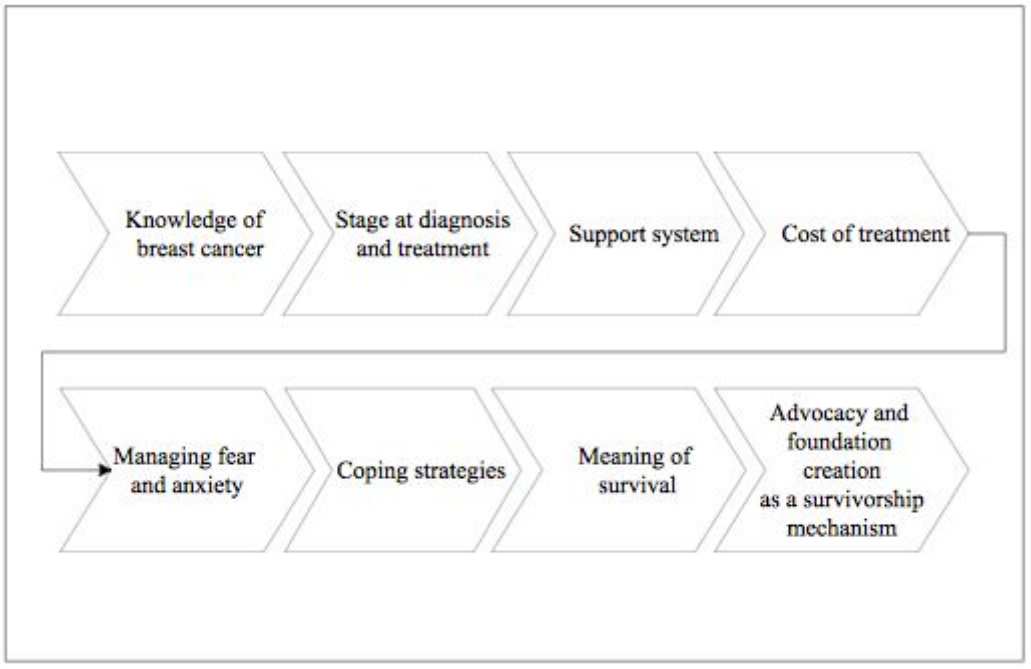



Figure 10. Breast Cancer Patient's Journey, Williams, 2015

Knowledge of breast cancer is a stage in which women with a prior knowledge of the symptoms of breast cancer, discover the problem themselve. They usually observe the changes in their breast themselve noticing a lump or tumor, which triggers the need of further

examination. According to participants of the study, becoming familiar with breast anatomy and physiology can help women detect breast cancer symptoms earlier. Consequently, earlier diagnose lead to a higher chance of total recovery from the illness. The next stage of Breast Cancer Patient Journey is diagnosis and treatment. Many women report being “shocked” and “scared” when they find out about their diagnoses. The stage of diagnoses also determines the type of treatment. The same applies for the tumor size, which then determines options available to the patient. In this stage women are also afraid of side effects such as hair loss, mastectomy scars or weight gain. Breast cancer patients also reported that support system was a crucial and provided them with strength, encouragement and hope. The most mentioned was the support by family, friends and co-workers, which the participants of the study saw as the “key” element in making the treatment bearable. What is more, women who participated in the social support groups reported a new bond and friendship created among the member who were able to share similar experiences. Cost of treatment is the next stage that patients need to face financial means of the treatment, which depend on the type of health insurance a patient have. Afterwards, women need to learn how to manage fear and anxiety regarding recurrence of the illness. In order to effectively deal with physical symptoms and side effects of the treatment patients also learn to develop coping strategies. However, many of the participants of the study made sure to get involved in various activities, which provided them with comfort and helped in maintaining positive attitude. Nevertheless, it turns out that meaning of survival differs from one person to another. After the treatment every woman adopts her personal meaning of survival. Many of them realize how precious their life is and express their gratitude to life. As the last stage, Williams listed advocacy and foundation creation as a survivorship mechanism. Some (but not all) participants of the study, felt the need of educating other women about breast cancer as well as creating awareness and rising funds. Few of them even shared their experiences in establishing foundations.

# Theory and Techniques for Social Media Analytics

Social media enables us to obtain digital information about individuals, organizations and institutions, which is now labeled as big data. Analysis of social media requires collection,

storage, analysis and reporting of the data that has been collected. This study aims to analyze discussions of breast cancer patients of survivors on specific Facebook pages. Based on social data sets, utilizing various methods and techniques of computational social science, I hope to provide meaningful insights regarded aforementioned concepts which go beyond traditional social science investigation.

## Theory of Social Data

Therefore, in order to collect and analyze big social data, one shall support conceptual and mathematical modeling with the theory of big social data. To be able to describe how individual data items come into being, I decided to use a theory of socio-technical interactions by Vartrapu (2016), as this theory will enable us to conceptualize perception of interaction between the users who generate analyzed content. According to the theory, social media platforms such as Facebook involve two types of user interactions: (1) Perception and appropriation of socio-technical affordances, (2) Structures and functions of technological intersubjectivity. Those practically reflect a user (a) interacting with technology (such as using Facebook on a user's tablet) and (b) interacting with others socially using technology (such as commenting on a picture posted on Facebook using a tablet). Those result in electronic trace data, which

Theory of socio-technical interactions leaded Vatrapu et al. (2016) to development of conceptual model of social data, which consists of two types of data (1) Interactions and (2) Conversations.

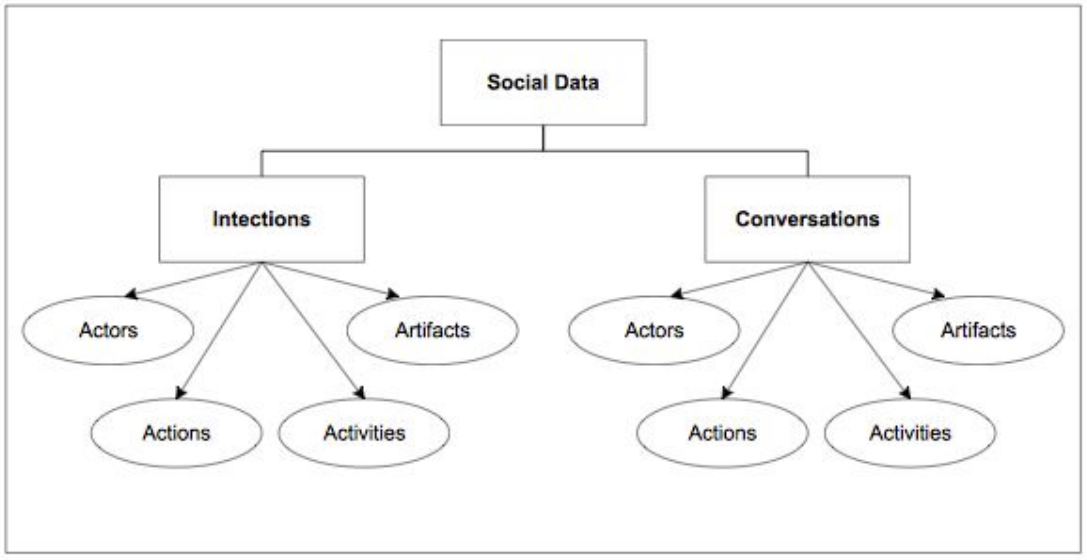



Figure 11. Model of Social Data, Vatrapu et al. 2016

`

Interactions reflect the relationship of different social media possibilities such as posting, linking, tagging, sharing, liking etc. Converdationst consists of the factors which make it possible for users to communicate, such as topics, keywords, pronouns and sentiments. Therefore, using social data in this study provides means to understand both interactional and conversational processes, which happen to be key catalyzers of the processes listed in the above mentioned theories (Self-efficacy Theory or Psychological Empowerment Model). Interactional processes help to examine in depth the engagement of the users. While, conversational processes serve as a source of data on the content and emotions expressed by the users.

## Supervised Machine Learning for Content Analysis

Big data gives the opportunity to test social theories on a large number of data. Having large amount of content, however, requires the use of computational tools. Machine learning is an area of computing, which is closely related to data mining and statistics (Dale, Moisl & Somers, 2000). There are three types of machine learning: (1) supervised machine learning, (2) unsupervised machine learning, and (3) reinforcement learning. Unsupervised machine learning is based on uncovering hidden patterns from unlabeled data, without knowing in

advance what categories are, which is known as clustering. Opposite to that, supervised machine learning is based on a training data, which provides examples of situations and for each of the examples it assigns an outcome (assigns a label). The machine then uses this training data to build a model, which will predict the outcome of the new data based on the past examples. Reinforcement learning in which machines are software agents interacting with environment which learn how to optimize their behavior given a system of rewards and punishments. To test and improve model specification and predictions, the most suitable is application of supervised machine learning. Since, this research is based solely on sociological and psychological theories explained in the previous subsections, it gives a perfect opportunity to create models for text classification and application of supervised machine learning.

In order to prepare do it, one shall follow the 5-step process for content analysis suggested by Rebecca Morris (1994) (See Figure 11) . The first step requires categorisation of a unit of text. The second step involves development of classification models, which will serve as a guide to train the coders. In the third step, the coders are being trained on text classification according to previously set models. In this step, they key activity is to make sure that the coders understand all the coding rules and concepts which will enable them to label the data points. In the next step, the coders will classify as set of randomly selected data, which will facilitate reliability coding, to test, whether there are any misunderstandings between the coders regarding the categories for classification. If the results will be reliable, then one can apply the code to the rest of the data set. If they are not, then the coding rules should be reviewed and the coders should repeat the classification.

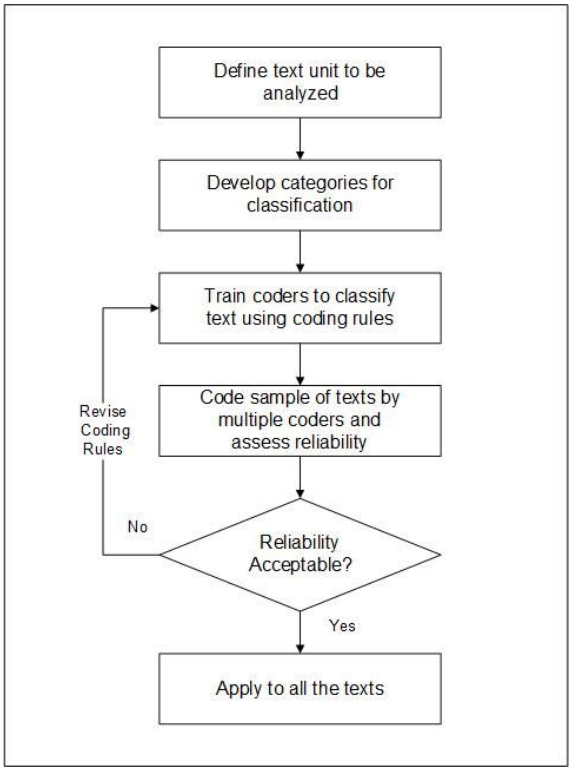



Figure 12. Five-step Data Analysis Process, Morris, 1994

# Related Work

## Empowerment and Self-efficacy in Breast Cancer Research

Both empowerment and self-efficacy have gotten a lot of attention among scholars for the past several years both in application to breast cancer as well as rehabilitation research. First of all, the need of improving low frequency of participation in rehabilitation leaded to the development of aforementioned Rehabilitation  Empowerment Model by Fisher and Howell. Other scholars also looked more specifically on the role of self-efficacy and empowerment while undergoing breast cancer.

As mentioned in one of the previous sections, Badura (1993) listed four aspects that highly influence self-efficacy are social comparison influences, framing feedback. Wood et. al (1985) studied the role of social comparison specifically in adjustment to breast cancer. Two

most important motives for social comparison turned out to be self-evaluation and self-enhancement. It is said that breast cancer patients value the opportunity of being able to compare their emotional reactions towards their illness to other patients who suffer from the same. Furthermore, undoubtedly breast cancer can pose great threats to patients' self-esteem. This can be influenced by changes in body image, social relationships or occupation. Therefore, the patients seek social comparison in order to meet their self-enhancement goals. Comparison is however based on four main perspectives. The similarity perspective explains that in order for a person to be able evaluate herself, she needs to compare herself with someone similar. Attributes that breast cancer patients usually see as similar are age, type of surgery and prognosis. According to the upward comparison perspective, patients tend to compare themselves to people who are better off, because it either helps them to measure their adjustment in relation to the person who is better off and / or it helps them feel better by identifying themselves to that person. Opposite to that, some breast cancer patients avoid comparing themselves with superior others, in order not to threaten their self-esteem. Instead, by looking at others who are in worse situation they can minimize their own sense of victimization, which is the reflection of the downward comparison perspective. The last perspective is called the supecoper perspective, which is drawn upon the source of comparison, which might be available in the media, which at the time of the research (the year of 1985) related only to the images depicted on TV, such as celebrities, who has easily came back to their lives. Nowadays, cancer patients have the opportunity to browse the Internet, which offers a wide range of stories, information and support, which enable them to choose, who they want to compare themselves with.

Importance of patient empowerment has also been widely discussed in the breast cancer literature. Teleghani et al. (2014) conducted a qualitative study in which they looked into empowerment needs of women with breast cancer. In their study the researchers interviewed 19 women with breast cancer in attempt of finding out what are their empowerment needs. Based on interviews, the researchers distinguished three main categories of empowerment needs from the participants' perspective, which were information, beliefs and skills. The patients seeked comprehensive as well as continuity of information about their condition and

treatment. They also expressed the need of relying on other people's trust and beliefs in their choices and actions. Finally, the often mentioned the need of communication and expression of their needs, receiving answers and use of the internet.

## Breast Cancer, Empowerment and Self-efficacy in Social Media Environment

Nowadays, social networking sites generate convenient ways to connect breast cancer patients with their peers, which can provide a wide array of practical and supportive information. (Kidd, 2008*)* Nazil et al. in their 2017 study on topics on breast cancer in social media compared different uses of forums and social networks. It turned out that support from a patient's family and friends, body care and body image during cancer, and sexuality were mostly discussed on Facebook. On the other hand, the Internet forums were more often used to discuss topics such as chemotherapy and its secondary effects, breast reconstruction, and breast cancer as daily battle.

There are many support groups, which people diagnosed with breast cancer can join in order to share their experiences. Those however are usually private. Facebook also offers the possibility of creating pages, which gather that users of special interest can follow in order to receive certain information that is exposed in their News Feeds. As mentioned by Bandura, individual self-efficacy can be strongly influenced by the social feedback. Comparison influences and framing feedback were one of his forefront constructs of cognitive development of self-efficacy. Therefore social support serves as an important aspect that enhances self-efficacy.

In one of the earlier sections, it has been stated that physical efficacy is one of the most relevant triggers towards rehabilitation process. Since the results of the study showed that patients tend to analyze their physical attributes over Facebook, it gives us confidence to use the Facebook data to analyze the topic of self-efficacy.

Social media serves as a great tool in sharing knowledge and experiences, which in the end help to empower peers, who struggle with the same problem. Facebook also turned out to be a great tool in patient empowerment towards medicine use.

## Big Social Data in Breast Cancer Research

Nazil et al. (2017) attempted to analyze the quality of life of patients with breast cancer by conducting topic analysis for social media on breast cancer. The biggest frequency of discussion had "hair loss, work life during cancer, support from patient's family and friends, treatment period, diagnosis, and family members with breast cancer". Koval, Li & Lyst (2017) also visualized social media trends about cancer. The researchers used SAS software and managed to identify trends from the twitter data regarding different types of cancer. Breast cancer turned out to have the highest number of tweets. Moreover, together with ovarian cancer, they were one of the topics that shifted around fundraising and campaign tweets. The scholars also depicted high celebrity influence as the top mentions included Taylor Swift's mother, Ellen Show's mother, Kylie Minogue, Joan Lunden, Carolina Herrera, Christina Applegate, Robin Roberts, and Oprah. This is a great representation of previously mentioned supecoper perspective of patient empowerment, which is more diverse due to the access to the Internet.

Bender, Jimenez-Maroquin & Jadad (2011) analyzed the content of breast cancer groups on Facebook. Topics of discussions on Facebook groups revolve around shared experiences of the illness. Opposite to forums, where patients share information, the breast cancer support groups on Facebook turned out to be personal and socially stigmatizing topics. Through the ability of reading other people's personal stories and receiving advice on the issues that can be posted in the group, the groups help patients to reduce the feeling of isolation, buffer anxiety, cope with painful treatments and be prepared for interaction with health system. The groups are therefore seen as "bridging social capital" due to the fact that they provide access to the information as well as "bonding social capital" as people can find emotional support there.

Opposite to Facebook groups, Facebook pages are public, thus interactions that occur there are visible to all Facebook users. Therefore, this feature appears to make them more suitable for figures or organisations, which want to attract number of followers in order to serve some purpose. They mostly serve to raise fundings, increase social awareness and promote-a-site groups in particular. Abramson, Keefe & Chou (2015) conducted a qualitative analysis of one of the breast cancer awareness pages on Facebook. It turns out that they serve as an effective tool for open space for self-expression. As using public Facebook pages, users can also initiate discussions by making a post. The researchers found that the topics are very diverse, varying from products and services to sharing personal stories and opinions.

## Research Gap

Although researchers already proved that self-efficacy and empowerment are important factors in medical adherence behavior, as well as they looked into online discussions of breast cancer patients and mapped some of the most important trends and topics discussed on social media, no research has been done to investigate whether online discussions can actually release the feelings of self-efficacy and empowerment in breast cancer patients.

# Methodology

As defined by Anderson (2010) ”Qualitative research involves the collection, analysis, and interpretation of data that are not easily reduced to numbers. These data relate to the social world and the concepts and behaviors of people within it” (p. 1). This chapter aims to provide the insights on specific processes of the research as well as justification of their choice, characteristic elements and the appropriateness of those elements for the large purpose of the study. The following diagram illustrates the research and analysis process.

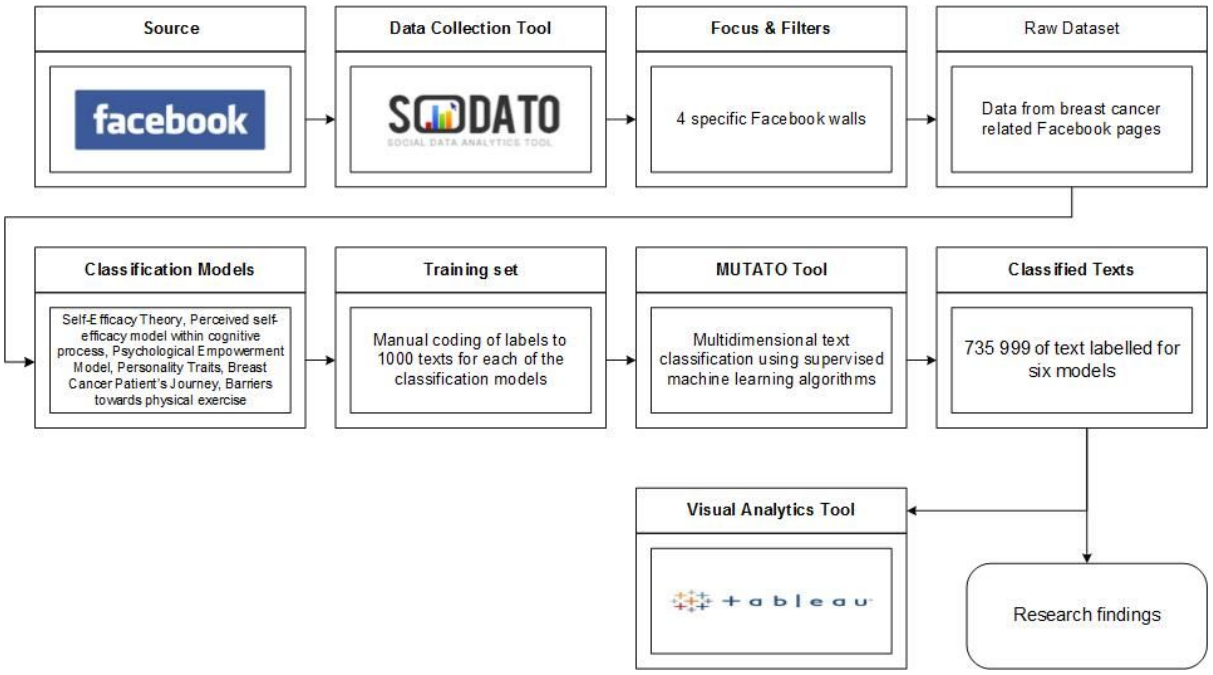



Figure 13. Analysis Process Diagram

# Research Approach

Data collection and analysis are highly dependent on methodological approach used in a study. (Bryman, 2012) In order to illustrate effective progression through which this research methodology was designed, I followed the stages of research onion approach developed by Saunders et al. (2007). The stages cover the identification nature of social reality the nature of social reality (known as ontology) ways of knowing (known as epistemology), and ethics and value systems (known as axiology ), ways of knowing (known as epistemology) as well as well as time horizons and data collection methods (See Appendix A).

## Philosophy of Science

Since the most outer layer of research onion deals with philosophies, one shall specify the philosophical paradigm adopted in order to inform the research design. Paradigm is a set of assumptions, theories or theories that serve as the foundation of concepts like identity or institutions, which in the end provides explanation on how research was undertaken (Flick, 2011). The main ontological frameworks used in social sciences are: positivism and

interpretivism. Positivists see the world as external (Carson et al., 2001). They construct a clear research topics, supporting it with appropriate set of hypothesis. What is more, they keep a clear distinction between science and personal experience. That is why, positivist researchers create distance between themselves and participants to remain emotionally neutral and creare clear distinction between science and personal experience. On the other hand, interpretivism sees reality as constructed through the meanings created by individuals. It is complex and unpredictable (Carson. et al 2001). Thus, it is characterised with flexible research structure capturing meanings in human interaction, rather than fixed research design.Therefore, typical research methods used by interpretivists are unstructured interviews or participants observations.

The aim of this research is to gain a deep understanding of why the patients act in a way they do and how can this behavior (facebook comments) influence their level of self-efficacy and feeling of empowerment. the more suitable approach is interpretivism as it allows me as a researcher to understand motives, meanings and experiences, which drive certain feelings among breast cancer patients. What is more, receiving a huge amount of text data and applying different theoretical models enables me to gain in depth insights into the online discussions.

## Research Approach

Another stage of the research onion suggests to specify a research approach. There are two types of research approaches: inductive and deductive. The most crucial difference between both approaches is the start of the research. Deductive approach starts with a theory which through testing is to be confirmed or rejected, while inductive approach starts with an observation and aims to develop a theory through inductive reasoning.

Therefore, following this approach to research, I began with a research question, which guided me in collection of empirical data, which was used to generate testical or tentitive hypothesis, which I confirmed with additional data. Those then became a basis for forming a social theory, set of ideas that I used to explain the topic. Analysis was based on constant

comparison of the content of facebook comments to their counterparts included in different domain specific models for text classification. In the end, I was able to develop a new theory and give further recommendations.

## Research Strategy

This brings me to narrowing down the approach to anxiological level of the research, which explain the process of decision making during the process of this research study. One of the types of inductive approach is Grounded Theory approach. This theory is seen as a perfect approach for analysis of large quantities of unstructured or semi-structured qualitative data (such as big social data used in this study).

In order to position a new theory derived from this study, I examined several existing theories, which led me to constructing middle-level theories, which served as sequential guidelines for conducting qualitative research.

## Choice of Methods

Since the research is strictly focused on gathering one type of information, which is text from the Facebook walls, one can conclude that the mono method has been applied.

## Time Horizons

The study aims to analyse the content of existing Facebook pages. Neither it does consider any changes happening through any period of time, nor requires conducting of the study through the long period of time. Therefore it can be classified as cross sectional study as it enables me to to solve a research question at any given time once the data is gathered.

# Procedure

The overall process consists of three important steps:  Data collection, data processing, quality assurance and data analysis.

## Data collection

In order to fetch the social data, I used A Social Data Analytics Tool (SODATO), which utilizes APIs provided by social network vendors and provides a server to store and pre-process the data. In order to find useful findings, the data was fetched from 4 most popular Facebook Pages (See Table II). Those pages include: (1) Breast Cancer Now, Breast Cancer Site, (2) Breast Cancer Care, (3) Breast Cancer Support UK, (4) Breast Cancer Now. They have been chosen due to their significant amount of followers, which reached up to hundreds of thousands of followers, that gave me confidence of collecting a large enough amount of data to serve as a valid sample. What is more, the content of the aforementioned Facebook pages fit the purpose of the study. The pages has clearly been created to support breast cancer patients, increase breast cancer awareness as well as raise funds for persons ongoing breast cancer and survivors. The content of the walls ranges from tailored information for anyone affected by breast cancer to sharing of experiences of fellow survivors or persons who are ongoing the illness. That is why, I believe, I can find valuable insights to be able to answer the research question.

| | **Breast Cancer Care** | **The Breast Cancer Site** | **Breast Cancer Support UK** | **Breast Cancer Now** | **Total** |
|---|---|---|---|---|---|
| **Posts** | 5500 | 6622 | 504 | 6712 | **17338** |
| **Comments** | 42788 | 88441 | 3662 | 44137 | **178728** |
| **Comment replies** | 9644 | 17743 | 1380 | 6352 | **35199** |
| **Likes** | 1122285 | 2583472 | 118242 | 1173016 | **4997015** |

Table I
DATASET DESCRIPTION

Most texts were in English and most originates from UK & the US. The texts in different languages were excluded during the dataset training. The texts were labeled according to the model constructs they promote and reflect.

## Models for Text Classification

Since, this research is based on grounded-theory approach, the logical step is to go forward with supervised machine learning technique. Various models and theories listed in the theoretical framework of this thesis, serve as a solid ground for developing domain-specific text classification models. The models include: (1) Self-efficacy Development Model (Refer to p. 19 ), (2) Model for Perceived Self efficacy within Cognitive Process (Refer to p. 22), (3) Psychological Empowerment Model (Refer to p. 23), (4) Personality Traits: The Five-factor model (Refer to p. 26), (5) Breast Cancer Patient's Journey (Refer to p. 29), (6) Barriers towards physical exercise (Refer to p. 11). As these models and theories consist of numerous constructs, they serve as a perfect tool to label texts into specific categories. The models were used for dataset training, which afterwards enabled the software classify all the data collected from Facebook walls.

| LABEL | DEFINITION |
|---|---|
| **Model 1: Self-efficacy Development Model** | |
| Performance Accomplishments | Previous experiences of adherence |
| Vicarious Experiences | Successful or unsuccessful attempts at adherence |
| Verbal Persuasion | Support or encouragement by significant others and peers |
| Physiological State | Patient's interpretation of whether or not she / he is capable of adhering at treatment |
| **Model 2: Model for Perceived Self-efficacy within Cognitive Process** | |
| Conception of self-ability | Readiness or possibility to show self-ability |
| Social comparison | Relating and comparing to others |
| Framing feedback | Talking about or evaluating behaviour of other people. |
| Perception of controllability | Exhibiting self-control ability / ability of management of process |
| **Model 3: Psychological Empowerment Model** | |
| Interpersonal | Am I motivated enough to do it? |
| Interactional | Show that one has skills knowledge to go for next step |
| Behavioral | Represents actions taken by an individual to achieve or influence specific outcomes |
| **Model 4: Personality Traits - the Five-Factor Model** | |

| | |
|---|---|
| Openness | Being open to new learnings and experiences as well as being curious, imaginative, adventurous and focused on tackling new challenges. |
| Conscientiousness | An individual with this trait characterizes with goal-directed behavior, enjoys having a set schedule and pays attention to details. |
| Extraversion | Opposite to introversion; extrovert likes meeting new people, making new friends, being center of attention. Such an individual is outgoing, social and talkative. |
| Agreeableness | This trait includes attributes such as conforming to social norms, trust, altruism, kindness, affection, generosity, optimism and trustworthiness. |
| Neuroticism | This characteristic is linked to sadness, moodiness, and emotional instability. People with this trait are prone to experiencing mood swings, anxiety and irritability. |
| **Model 5: Breast Cancer Patient's Journey** | |
| Knowledge of breast cancer | Prior knowledge and understanding about breast health |
| Stage at diagnosis and treatment | Emotions tapped at the time of breast cancer diagnosis |
| Support system | Social support as a critical ingredient to physical and mental health when facing cancer care |
| Cost of treatment | Financial situation and health insurance in relation to treatment costs |
| Coping strategies | Fear and anxiety regarding recurrence |
| Meaning of survival | Appreciate and realize how "precious life is"and the importance of living "one step at a time |
| Advocacy and foundation creation as a survivorship mechanism | Educating other women and creating foundations to provide education, generate awareness and raise funds for breast cancer research |
| **Model 6: Most Common Barriers towards Physical Exercise** | |
| Physical barriers | Lack of energy or vigor, lethargy, feeling tired, decreased strength, and trouble concentrating |
| Emotional barriers | Anxiety, depression, fear and stress |
| Cognitive barriers | Lack of motivation, lack of perceived control, lack of perceived ability |

Table II
DESCRIPTION DOMAIN-SPECIFIC MODEL FOR TEXT CLASSIFICATION

From all the comments collected, a total of 1000 text was chosen and were manually coded for 6 domain specific models. Coding of texts was done by two coders: 1000 texts were coded by coder 1: mysel. Out out of these 1000 texts, 200 texts were randomly allocated to the coder 2: a fellow student, who was familiarized with the theories and text classification model together with the examples (See Appendix C) before the start of classification procedure. This procedure was conducted in order to assure consistency of the data and

accuracy of the agreement between the coders, which has been practically tested by calculating Cohen's kappa coefficient.

## Machine Learning and Text Classification Performance

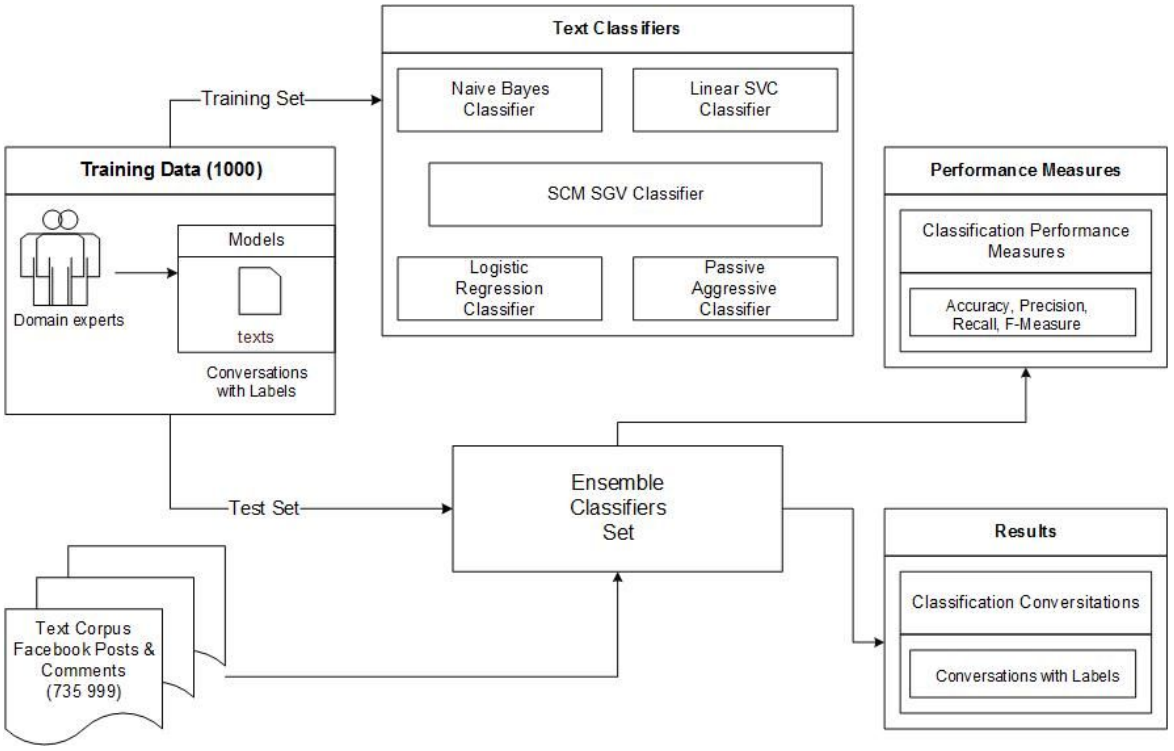



Figure 14. Text Classification Architecture

After manual coding, the next step is model training. In order to do so, I used supervised machine learning technique, text classification with several algorithms (Fig. 14) to classify the texts using custom-build tool Multi-dimensional Text Analytics Tool (MUTATO )developed by the Computational Social Science Lab at Copenhagen Business School. MUTATO is a tool that can perform several tasks ranging from text mining, text classification as well as topic modelling. It is based on keyword analysis, word frequency analysis and collocation analysis. In order to perform text classification, an ensemble learning method has been used. To assign a given textual comment to a specific label, the program is based on 5 supervised text classification algorithms. These classifiers has been built using

scikit learn machine learning library in Python. As shown in Figure 11, there are 5 different algorithms for text classification: (1) Multinomial Na ¨ıve Bayes classifier (Multinomial NB) (2) Linear Support Vector Classifier (Linear SVC) (3) Logistic Regression Max Entropy classifier (Logistic Regression) (4) Passive-Aggressive classifier (Passive Aggressive) and (5) Support Vector Machines with stochastic gradient descent (SVM SGD). Furthermore, to see the effectiveness of performance of classifiers, the program applies several performance measures, which are built over specific statistical variables and serve to indicate correct and incorrect values of predictions done by a certain algorithm. The first performance measure is *precision*, which indicates the ratio of positive values to predicted positive values obtained during data training process. The second one, *recall* is a ratio of correctly predicted positive values over all positive values. Once values or *precision* and *recall* are set, *F1-score* is used to test harmonic mean of precision and recall. In the end *accuracy* provides information about overall performance of all classifiers. Once training of the classifiers is completed, the rest of the remaining text data was classified.

| **Model 1: Model for Development of Self-efficacy** | | | | |
|---|---|---|---|---|
| **Classifiers** | **Precision** | **Recall** | **F1-score** | **Accuracy** |
| SVM SGD | 0.53 | 0.61 | 0.53 | 0.614 |
| Voted Accuracy | - | - | - | 0.596 |
| **Model 2: Perceived Self-Efficacy Model within cognitive process** | | | | |
| **Classifiers** | **Precision** | **Recall** | **F1-score** | **Accuracy** |
| Linear SVC | 0.71 | 0.61 | 0.59 | 0.606 |
| Voted Accuracy | - | - | - | 0.606 |
| **Model 3: Psychological Empowerment Model** | | | | |
| **Classifiers** | **Precision** | **Recall** | **F1-score** | **Accuracy** |
| Multinomial NB | 0.52 | 0.46 | 0.43 | 0.463 |
| Voted Accuracy | - | - | - | 0.463 |
| **Model 4: Personality Traits - the five-factor model** | | | | |
| **Classifiers** | **Precision** | **Recall** | **F1-score** | **Accuracy** |
| SVM SGD | 0.48 | 0.43 | 0.39 | 0.428 |
| Voted Accuracy | - | - | - | 0.392 |
| **Model 5: Breast Cancer Patient's Journey** | | | | |

| Classifiers | Precision | Recall | F1-score | Accuracy |
|---|---|---|---|---|
| Linear SVC | 0.62 | 0.64 | 0.61 | 0.639 |
| Voted Accuracy | - | - | - | 0.628 |
| **Model 6: Most common barriers towards physical exercise** | | | | |
| **Classifiers** | **Precision** | **Recall** | **F1-score** | **Accuracy** |
| Linear SVC | 0.75 | 0.67 | 0.65 | 0.666 |
| Voted Accuracy | - | - | - | 0.666 |

TABLES III
PERFORMANCE MEASURES OF THE CLASSIFIERS

All the performance measures of each of the classifiers has been computed and displayed in a table, which can be found in the Appendix D. In most of the Models (1, 2, 5 & 6) voted accuracies varied in between 0,596 and 0,666, which shows that the prediction of classifiers was highly accurate. Among all classifiers the best performing was Linerar SVS and SVM SGD.

# Data Processing and Visualisation

In order to understand and visualise the data results obtained from text classification using MUTATO, I used the data visualisation software called Tableau. It is a smart software which helps to visualise unstructured data and make them informative through generating very appealing graphs.

As the aim of the data processing and visualisation is to enable me to discover main trends in social media discussion among breast cancer patients, considering the constructs of models used in data classification and therefore answer the research question (*how the indicators of Rehabilitation Empowerment Model can be found by analysing Facebook walls interactions between people undergoing breast cancer*).

The research question will be answered after investigating adequacy of the main proposition (*Online discussions can serve as a complement of the rehabilitation practitioner construct in the Model for Rehabilitation Empowerment*) and subpropositions, following several steps.

1. Visualisation of topics related to certain steps of breast cancer patient journey model, will enable me to investigate Sub-proposition 1 (*Breast cancer patients seek and find support by following and actively participating in discussions on Facebook walls*).
2. Visualisation of constructs of Theory of Self-efficacy will enable me to determine whether discussions reflect 4 sources of information (theory constructs) which can increase one's self efficacy. Therefore, it will be possible to investigate Sub-proposition 2 (*Topics of online discussions contain information that can positively influence self-belief in breast cancer patients*)
3. Visualisation of the amount of comments representing certain concepts of the Model for Perceived Self-efficacy within Cognitive Process will enable me to apply findings into Sub-proposition 3 (*Online discussions reflect content which can trigger the increase of self-efficacy within cognitive process of breast cancer patients*)
4. Visualisation of data results of text classification in Psychological Empowerment Model will facilitate a closer look on Sub-proposition 4 (*Breast cancer survivors can facilitate the process of psychological empowerment of individuals undergoing the illness*)
5. Visualisation of personality traits used for text classification will let me find main personality traits of BC patients and find the link between them and self-efficacy as well as empowerment level (which will be demonstrated in the previous steps)
6. Conducting the analysis of common barriers toward physical exercise will be a great asset for recommendations on what shall be improved, or what the focus should be put on when providing patients with different kind of information, so in the end it has the potential to empower them and increase their self-belief towards exercising.

# Quality Assurance

The first one to mention shall be the aforementioned process of analysing performance measure of text classifiers, which provided the researcher and readers with the information on accuracy, f-measure, precision and recall of overall performance of the classifiers and the results turned out to be viable. Quality of qualitative research can also be assessed by its

validity, credibility, reliability, and generalizability (Leung, 2015). Validity of the research can be set in the early stage by selecting appropriate research approach, as well as methods, tools, process and data necessary for the research to be conducted in an appropriate way. This research design has been based on academic sources and articles, in which researchers used similar methods in order to explore similar phenomenon. What is more, reliability was constantly tested by making the form and context of the data comprehensive and inclusive with reference to existing theories. Since the study aims to research a phenomenon in a certain social group, focused in a particular context, generalizability of research findings might not be its expected attribute. However the findings can be generalized and methods can be used for similar theoretical model or social context. This research is clearly dependable on various academic theories, which has been proven by clearly presenting step-by-step data collection and analysis process. In order to make data "credible", I made sure that it reflects participants experience.

## Inter-rater Reliability and Cohen's Kappa

Cohen's Kappa is the reliability coefficient of two measurements of the same variable, which are a (1) nominal variable: proportion of units in which coders agreed and (2) dependent variable: proportion of units in which agreement happened randomly. The value can be calculated using the following equation:

$$k = \frac{p_o - p_e}{1 - p_e}$$

Measurement of a variable can be made by one competent judge (twice, in a time interval) or two different competent judges at the same time. Kappa Cohen takes values from -1 to 1. The closer to the value of 1, the more the two judges (or one judge in a time interval) are more consistent. In turn, the closer to 0, the more divergent the grades. Values from -1 to 0 are practically unused, because the negative Kappa Cohen's value would mean that the ratings would be less consistent than if they were completely random variables.

As an example Cohen's Kappa was calculated for Patient's Rehabilitation Journey Model, which resulted in 0.65 score. Although, it still shows some level of uncertainty, considering that the second coder was a person who volunteered to train the data, although she familiarized herself with the theories and definitions, having equally competent partner who would classify the same amount of data would undoubtedly results in a higher score.

# Analysis and Results

In this section I want to present the final results of text classification. Using graphs generated in Tableau to visualise the results, through application of the theories mentioned in theoretical framework, I will try to not only report, but also organize and explain the findings. I will do it by distinguishing between different types of phenomena which can be interpreted from the results.

## Results of text classification

### Openness to Sharing Information and Experiences

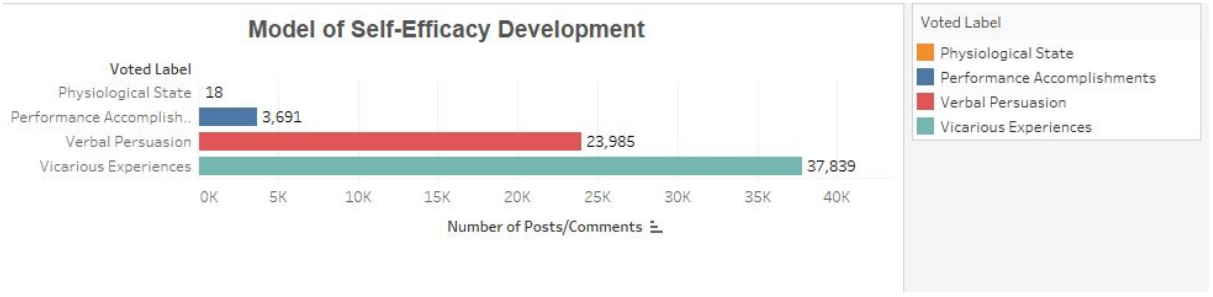



Figure 15. Factor Distribution of Self Efficacy Development Model

It turns out that a lot of patients share their experiences no matter if they were positive, negative or resulted in successful or unsuccessful outcomes. This has been clearly represented by a high number of *vicarious experiences* reflected in the comments (Fig. 15). Great amount of data was also classified as *performance accomplishments,* which shows that patients and survivors are willing to share their achievements such as treatment

accomplishment. As mentioned in the theoretical framework, *vicarious experiences* and *performance accomplishments* have a high impact on increasing one's self-efficacy as according to Bandura, relating to stories of people in similar situation, especially those that have positive outcomes, significantly help to restore one's self-belief in implementing a task successfully.

## Low Degree of Self-ability and Control

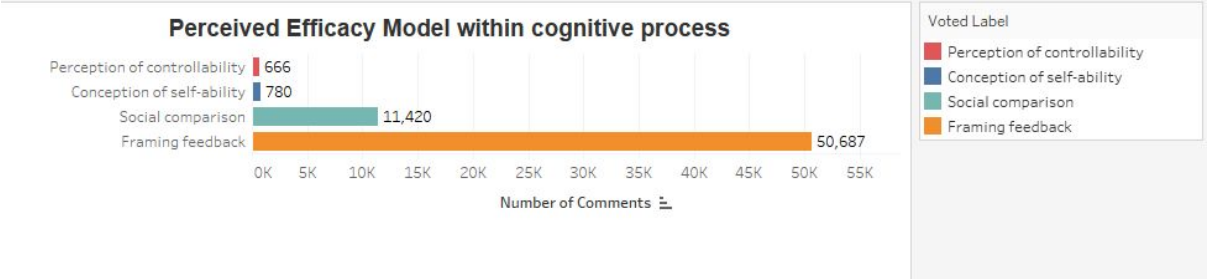



Figure 16. Factor Distribution of Model of Self-efficacy within Cognitive Process

One could suspect that the nature of the medium, meaning the fact that the Facebook pages are public and everybody can view individual's comments, limits the amount of personal information being shared by the patients. *Conception of self-ability* and *perception of controllability* are represent the the comments that include statements showing either once vulnerabilities or strengths. Nevertheless, it is more common and natural for people to share positive stories rather than failures in the public eye. Relatively low number of comments labeled as *conception of self-ability* and *perception of controllability* displayed in Fig. 16, could therefore indicate that patients do struggle with not having a sense of ability or control, as otherwise they would have shared it and the the above mentioned constructs would score higher. What is more, *vicarious experiences* construct distribution score is the proof of general openness to sharing stories. However, since *conception of self-ability* and *perception of controllability* are scored so low, one can assume that vicarious experiences are the comments, which include neutral or negative perception of self-control and self-ability, rather then positive one.

## Empowering Feature of Discussions

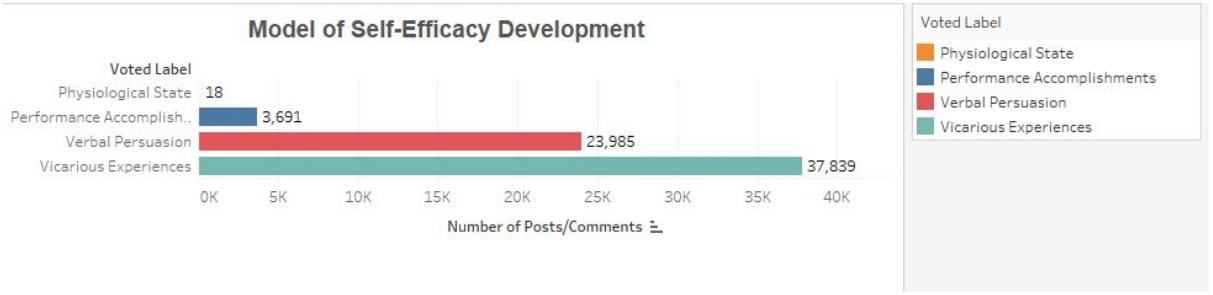



Figure 17. Factor distribution - Psychological Empowerment Model

Based on the theory of psychological empowerment by Perkins and Zimmerman, the above graph shows three main factors of the theory: *intrapersonal*, which expresses the level of motivation of a person to adhere to certain behavior, *behavioral*, which indicates whether a person has skills or knowledge to go the next step, and *interactional* which reflects actions taken by an individual to achieve or influence specific outcomes. Once these three factors are met, a person is expected to feel empowered. The highest amount of comments were labelled as *intrapersonal* aspect of empowerment. Respectively, Teleghani et al. (2014) proved that "information, beliefs and skills" are the most important components that foster one's degree of empowerment. As most of user comments were appear to express interactional aspects of empowerment, inevitably they must contain information about user's skills and knowledge. Therefore, it giver opportunity to other patients to gain information and inspiration from other people's experiences.

## Facebook as the Source of Social Support

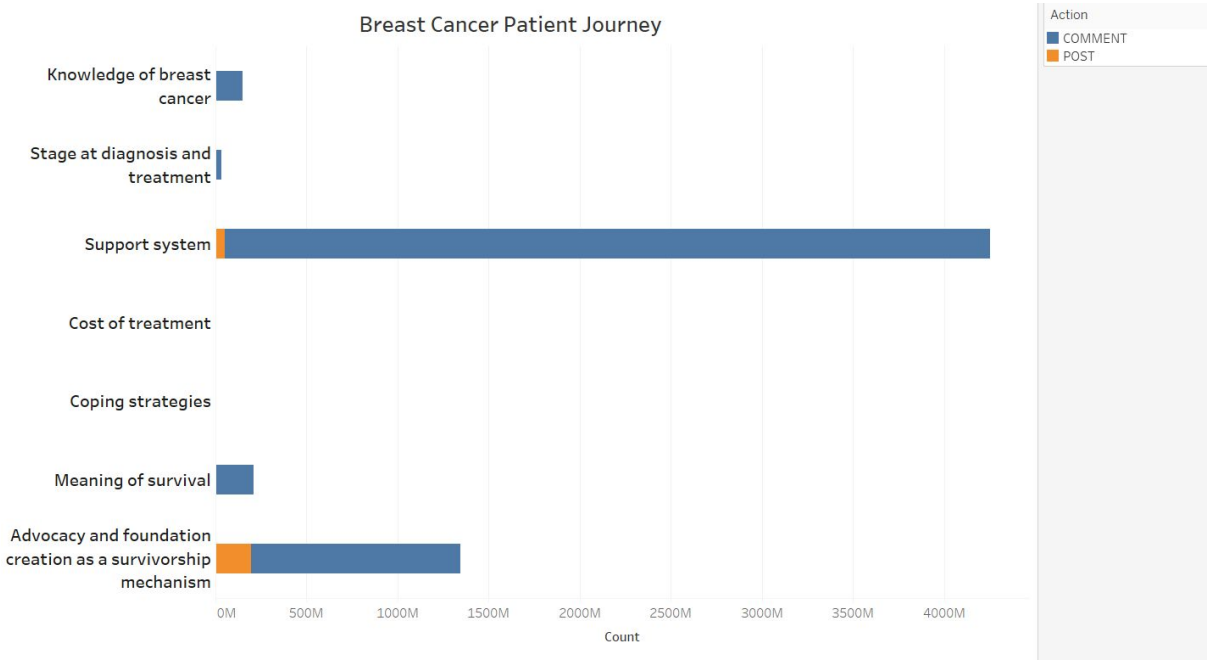



Figure 18. Distribution of Stages of Breast Cancer Patient Journey

Figure 18. Is visualisation of the results of text classification and distribution of different stages of Breast Cancer Patients Journey. Considering the purpose of Facebook pages devoted to breast cancer, which is awareness and support, it came as no surprise that the highest number of activity on Facebook is reported during *support system* stage of the journey.

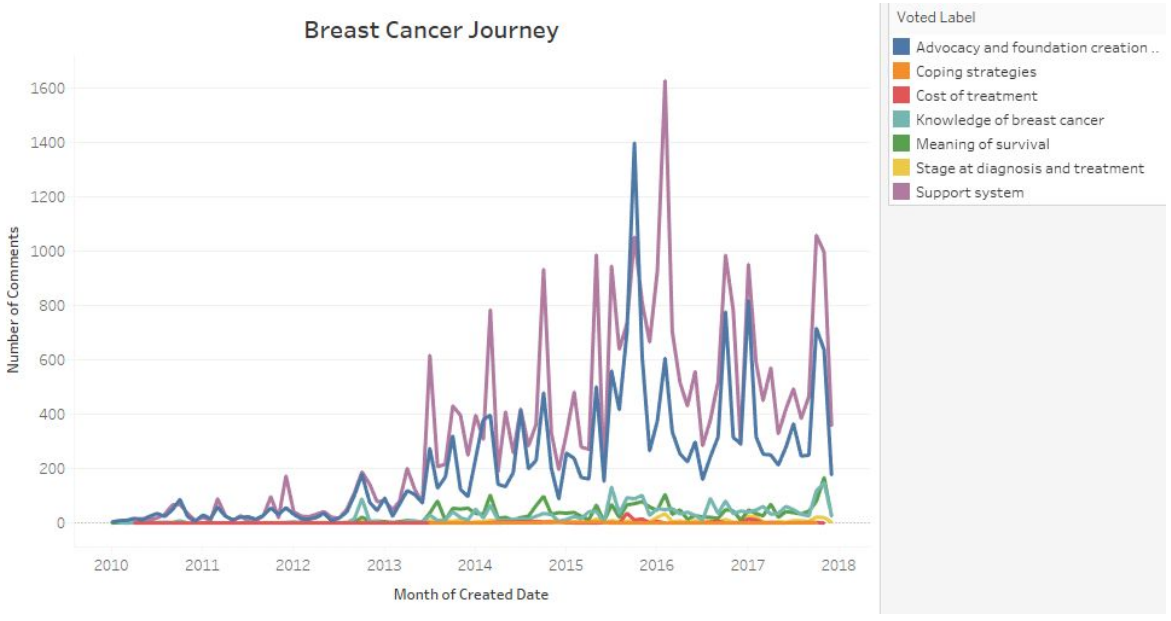



Figure 19. Distribution of Comments Classified to the Stages of Breast Cancer Journey over Time

Since the pages are devoted to providing support for people undergoing breast cancer, it self-explanatory that patients can easily come across supportive posts. To see whether they actually express the need of support or provide support to fellow patients, I created the graph that separates user posts and comments (See Fig. 18). The amount of user comments labelled as *support system* is beyond expected and proved that most of the patients become extremely active on Facebook walls in the stage of *support seeking*. Fig. 19 presents the distribution of comments in different stages of breast cancer journey over time. Once can easily notice, that the amount of comments of users in the *support system* stage varies dramatically and must be dependent on content displayed by page admins.In the interviews conducted by Teleghani et al. (2014) patients expressed the need of relying on other people's trust and beliefs in their choices and actions. Therefore, since Fig. 16 shows a great amount of *framing feedback* in the user discussions, as well as verbal persuasion in Fig. 15 , patients can definitely rely on Facebook walls in getting remarks and comments on their choices of actions from people who have been in similar situation.

Vast amount of data has also been classified to a *social comparison* label. This means that a lot of users comments do indicate comparison to other patients, which might potentially lead to self-evaluation and self-enhancement. Considering four comparison perspectives listed by Bandura (1) upward, (2) downward, (3) similarity, (4) supercoper, considering diversity of the comments, patients with different comparison needs can surely come across comments, which either motivate them (upward perspective), decrease their sense of victimisation (downward perspective), compare themselves to people same age with the same prognosis (similarity perspective) and potentially find role models (supercoper perspective). Fig. 16 also proves that a big amount of comments was labeled as *verbal persuasion* is a great indication of social supports as it is classified to the comments which content expresses encouragement of others to take certain action.

## Physical Barriers as the Biggest Obstacle

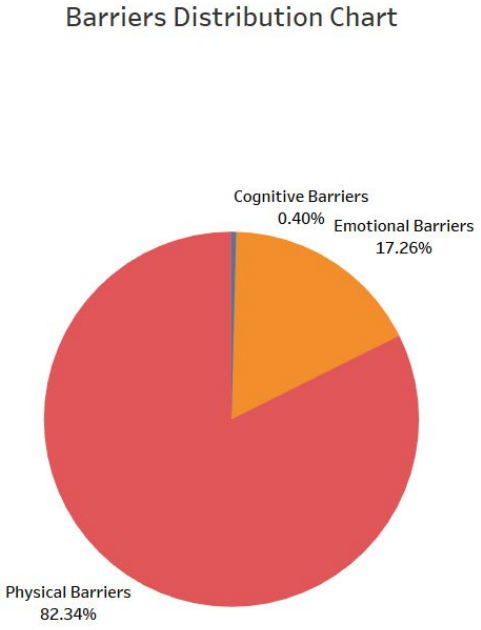



Figure 20. Most Common Barriers towards Physical Exercise

People undergoing breast cancer seem to be open for sharing of their struggles and frustrations. Although breast-cancer fatigue fosters emotional barriers towards physical exercise, the most reported turned out to be physical barriers. Over 82% of comments were classified as *physical barriers*, because they indicate lack of energy or vigor, lethargy, tiredness, decreased strength, and trouble concentrating. These are also the symptoms of *cancer-related fatigue*, what proves to cause wreak havoc among breast cancer patients. This is an indicator that more initiatives shall be taken in order to help patients to overcome the fatigue, which will then increase the chances that the patients will be more willing to initiate rehabilitation process. A little bit over 17% of comments included some of emotional barriers such as anxiety, depression, fear and stress. As these are usually caused by possibility of recurrence of the illness or many life-changing consequences of the treatment, by a great amount of support that Facebook walls proved to provide, those feelings can possibly be buffered if the admins adopt the right approach.

## Involvement in the Community

It turns out that breast cancer patients not only actively support each other on social media, but also try to actively contribute to research, spreading awareness and raising funds. By looking at Figure 18 *Advocacy and foundation creation as a survivorship* has a high score in text classification. Foundation creation seems to be a very popular activity among breast cancer survivors. Advocacy, which is defined as public support for or recommendation of a particular cause is valuable asset both for increasing self-efficacy and empowerment. As mentioned before, "information, beliefs and skills" are the fundamental factors of empowerment. Thus, high degree of advocacy as well as support system can foster information and beliefs, which can motivate a patient to leave comfort zone and therefore gain some skills.

## Agreeableness as dominant trait

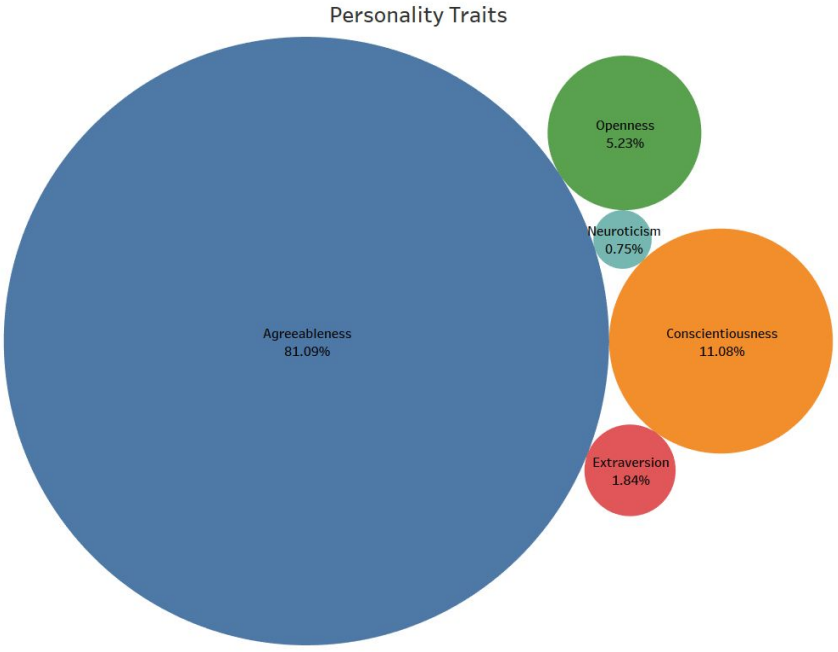



Figure 21. Big Five Personality Traits Distribution Among Users

In order to analyze personality traits of breast cancer patients, I filtered only user comments, which were classified into 6 labels and generated the graph showing percentages of comments that reflected these personality traits (See Fig.21). Although, *consciousness* was labeled by many researchers as an ideal feature for behavior agherence, the dominant personality trait turned out to be *agreeableness*, which was assigned to 81.09% of the comments. *Agreeableness*, however, also reflects the traits which are in favor of behavior adherence such as conforming to social norms, trust, altruism, kindness, affection, generosity, optimism and trustworthiness. Over 5% of the patients appear to share the trait of openness, which indicates that they are intellectually curious and open to emotion and experience as well as interested in learning. The only trait imposing some of the negative characteristics, *neuroticism,* turned out to be classified to less than 1% of the comments, which is a good sign. This trait causes anxiety, which firmly lowers the level of self-efficacy. However, again, the fact that Facebook pages are public might discourage users from sharing sadness, moodiness, and emotional instability.

Since, personality traits determine the likelihood of a person adhering to clinic-based aspects of rehabilitation program, having in mind that majority of patients holds a personality trait of agreeableness, communication strategies should be developed in a way that resonates with patient's personality.

# Discussion

Based on the analysis and findings reported and explained in the previous chapter, in this chapter I am to examine main-proposition and subpropositons, which will lead to finding the answer to the research question. What is more, I will list some of the limitations of the study and provide recommendations on what could be improved in the future work. I will also reflect on my personal learnings and academic performance. Before, the thesis is concluded I will also provide recommendations for healthcare sector and rehabilitation practitioners on how they can use the findings of this thesis.

# Factor Dynamics and Emerging Trends

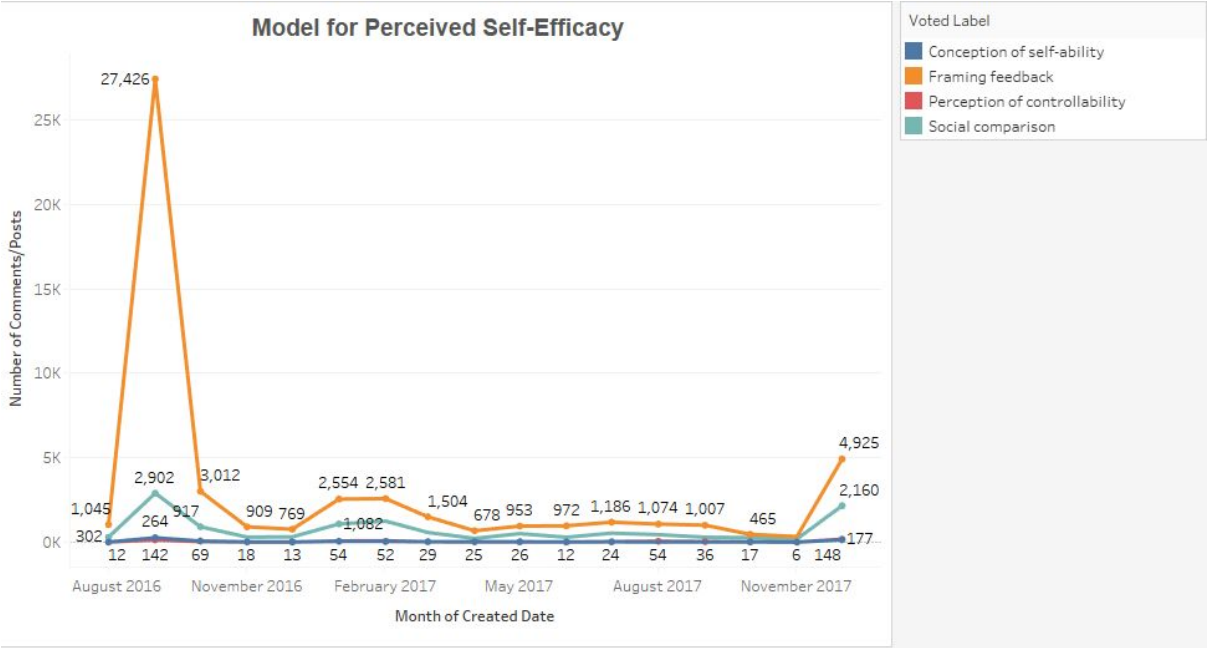



Figure 22. Distribution of Perceived Self-Efficacy Model Constructs Over Time

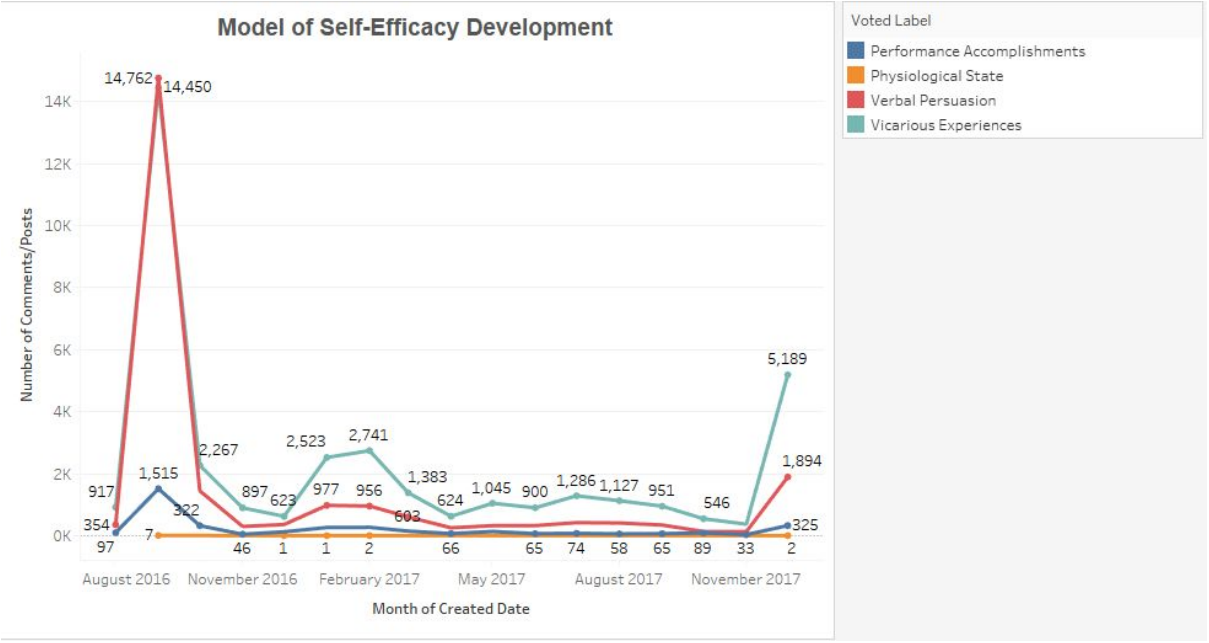



Figure 23. Distribution of Self-Efficacy Development Model Constructs Over Time

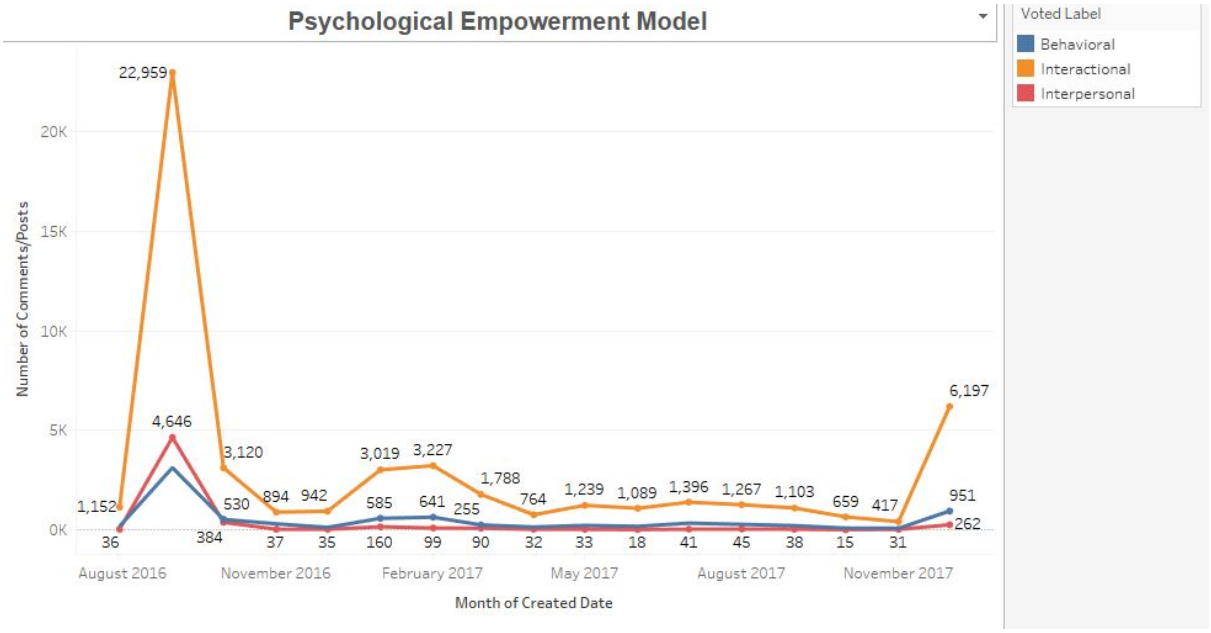



Figure 24. Distribution of Factors of Psychological Empowerment Model Over Time

By looking at Figures 22, 23 & 24, one can clearly observe that there is no exponential growth of any of the factors, as one could assume that considering number of Facebook followers has been increasing over time. One can however notice a similar pattern in redistribution of comments reflecting factors such as *verbal persuasion*, *framing feedback* and *interactional empowerment.* Based on the above results, one can however take on the argument that potentially many comments were labelled to all three factors. This shows the tendency of providing constructive feedback between the patients, as considering classification, the comments must have included the feedback, as well as one's abilities, and persuasive content.

One can observe a similar pattern between vicarious experiences and behavioral empowerment, which could be an indication that once patients share their vicarious experiences, the same comments also express actions taken to achieve or influence specific outcomes. By providing explanation on how achieved desirable outcomes, the reader who can relate to this situation, can feel motivated to undertake the same steps in adhering to certain behavior. Therefore, all posts and comments combining text that has been labelled to several constructs serve as a powerful foundation to increasing one's self-efficacy.

One of the interesting patterns of the findings is the difference between the frequency of sharing physical aspects, such as barriers and experiences versus more personal, emotional aspects such as feelings, or *conception of self-ability* and *perception of control*. The patients and survivors are also keen on giving each other feedback and motivating each other to adhere to certain behavior, which is reflected in the high number of posts and comments expressing factors such as *verbal persuasion* and *framing feedback*. Analysis of distribution of these factors over time shows dependency of both factors, therefore, one could assume that usually comments both: provide feedback on one's performance, but also motivate to take certain action.

# Conceptualisation of Findings

After analysis of the results of text classification and findings, as well as mapping out certain trends arising from the results, the next natural step is to conceptualise these findings and apply them into sub-propositions and the main proposition.

a. Sub-proposition 1 - *Breast cancer patients seek support and find support by following and actively participating in discussion on Facebook walls.*

Figure 17. shows a huge demand of social support in online discussions. Patients do only seek seek support but also provide it. They are very active in sharing information through writing about *vicarious experiences*, and providing *framing feedback*. Patients and survivors often express that they relate to someone else, or compare their situation to the situation of another person. Thus, *social comparison* factor is also relatively high. Since the frequency of involvement in the discussions is the highest in the third stage of breast cancer journey, which is the stage in which patients seeks support right after diagnosis, the content of online discussions have a great potential to buffer anxiety and increase the feelings of self-efficacy and empowerment.

*b.* Sub-proposition 2 - *Content of online discussions contain information that can potentially influence self-belief in breast cancer patients.*

Sources of information which influence person's self-efficacy has been listed by Albert Bandura in his Self-efficacy Development Model, which are *performance accomplishments, physiological state, vicarious experiences, verbal persuasion* (See Fig. 15). Analysis of the walls proved, that patients do share their experiences, as *vicarious experiences* score was high. They also try to motivate each other which was exhibited by a high number of comments classified to the *verbal persuasion* factor. However, patients do not share much information on their physiological state, which again, might be the result of privacy issue of Facebook walls. Concerning factor might be however a small amount of *performance accomplishments*, which are the most influential triggers of self-efficacy as they convey positive experiences of behavior adherence, which can make others believe that they can also succeed. Nevertheless, overall, users seem to share enough information, which can positive influence one's self-belief.

c. Sub-proposition 3 - *Online discussions reflect the topics which can trigger the increase of self-efficacy within cognitive process of breast cancer.*

Albert Bandura developed a Model of Self-efficacy within cognitive process, which contains factors, which trigger the process of developing self-efficacy. These include *conception of self-ability, framing feedback, perception of controllability, social comparison.* Considering the low degree of sharing information which would reflect one's conception of self-ability or perception of controllability, one could assume that the self-efficacy levels are rather low among breast cancer patients. However, the need for social comparison and framing feedback was overruling the other two factors, which can serve as a proof that, when given feedback or opportunity to compare oneself to a person in similar situation, one can develop self-efficacy. Thus, social media do serve as a good "food for thought" for factors of Model of Self-efficacy within cognitive process, and therefore can lead to development of one positive judgement about adherence behavior.

*d.* Sub-proposition 4 - *Breast cancer survivors can facilitate the process of psychological empowerment of individuals undergoing the illness.*

Although, according to the theory all three constructs of Psychological Empowerment Model need to be fulfilled in order for a person to be considered as empowered, the use of the model for text classification aimed to investigate whether these constructs can even be reflected in patients' comments. The reason for conducting this procedure was to test whether the comments reflect any signs of patients empowerment, as they do have a potential to empower other patients.

The above arguments serve as a proof of the main proposition suggesting that online discussions can serve as a complement of the role of a rehabilitation practitioner in the Model for Rehabilitation Empowerment (See Figure 2). In the model, rehabilitation practitioner constitutes the main trigger for enhancing self-efficacy through empowerment. The findings of the analysis has proven that machine learning can help to realise how the concepts of self-efficacy and empowerment can be realised from social media conversations. By creation of domain-specific models for text classification based on existing theories, which explain how people develop self-efficacy and become empowered, it was possible to train the data and measure the amount of comments expressing the constructs of those theories. The results of text classification visualised in the graphs illustrate that many constructs of theories of self-efficacy and empowerment are manifested both in the posts and the comments. A lot of experience and information sharing, as well as support given through feedback and persuasion, serve as a great source of enhancement self-efficacy and empowerment and can make patients not only aware of necessity of rehabilitation, but also motivated to overcome barriers and initiate the process.

# Implication for Research

This thesis provides a new angle on the topic that has recently gotten a lot of attention from researchers. It explores potential solutions of the rehabilitation problem and ways of enhancing patient empowerment. It also outlines a solid theoretical ground for researchers to

apply similar methodology into the analysis of comparable phenomenon. Theoretical framework could also be applied for the analysis of other online discussion sites, such as Internet forums. It also serves as a proof that machine learning provides effective ways of analysing big amount of data and makes it possible to conceptualise the findings.

# Recommendations for Healthcare Sector

Facebook walls again proved that they contain a lot of information on how patients are coping with specific aspects of their illness. Therefore, they serve as valuable source of information for the healthcare institutions that enables them to monitor the needs of patients and even observe some defects of the healthcare system that can be improved.

## Advantages of Big Social Data and Machine Learning

Just how big companies use big social data for analyzing their audience and taking strategic decisions, public health institutions should also take an advantage from availability of information coming straight from the patients. It also enables public health institutions to observe how patients interact with health information, which might improve health information management. Although big companies use big social data and create business intelligence departments to manage their strategic processes, healthcare institutions have an advantage of being able to cooperate with the researcher institutions or universities, which could conduct research on topics, that still lack support of social data.

## Strategy and Communication

Since the physical barriers turn out to be the most commonly reported, more effort should be put into developing solutions which will help the patients to overcome these barriers. Research proved that there is a link between a healthcare team member's communication skills and a patient's ability to follow instructions and recommendations, which increase their capability to self-manage a chronic medical condition (Wenzer, 2004). Renzi et al. (2017) lists several factors which increase patient's satisfaction. Among many of those factors, one can find communication, meaning that patients receive information, which is explained clearly and based on viable options. What is more, continuity of care also turns out to be one

of the crucial factors increasing satisfaction level of a patient. They do appreciate the follow up information and continuity of care from the same healthcare provider. As this research proved that support is clearly needed and provided by breast cancer patients mutually, healthcare institutions could either develop the ways for online communication or enter into cooperation with existing Facebook awareness page admins to provide even more medical information.

### Recommendation for Rehabilitation Practitioners

By approving International Classification of Functioning, Disability and Health (ICF, See Figure 1) World Health Assembly provided grounds for common language for health practitioners in order for them to be able to use it in patients rehabilitation worldwide,

Based on theoretical framework and the research conducted by Blaney et al. one shall recommend the practitioners to adopt a new approach towards the patients. Namely, they should try to transform the nature of fatigue through exercise, which in the end can serve as a mean of escapism. Through physical activity, patients have an opportunity to achieve sense of normality and progression from illness, which can become a motivator and facilitator to the regular exercise.

## Limitations of the Study

Although, the methods and theories has been carefully selected to match the purpose of the study and bring the answer to the research question, upon completion of the research process, I can point out some aspects that have limited the findings and should be considered for the future work.

Firstly, the data used for this research is fetched from public Facebook walls. Visibility of posts and comments might have negatively influenced patient's comfort of sharing personal information.

Although it seemed like domain-specific model for text classification consisted of specific constructs, it could be useful to develop even more specific constructs out of those already create. For instance, *vicarious experiences* are the comments in which the patients share both of the positive and negative experiences. Since positive experiences have greater potential to positively influence self-efficacy than those negative ones, it would be useful to find out which type of experiences is more common. If there are more negative experiences than positive ones, it would be beneficial to find out what are the specific problems the patients are dealing with, as they might also be an obstacle to adhering to rehabilitation practices.

Subsequently, division of the labels could be established in accordance to an author of a comment. For instance, during data training procedure, I noticed many comments coming from family members, who started raising money for a great cause:

> *"Hiya. My dad has recently been diagnosed with an aggressive cancer. I have started a diary page as a way of self help and i would like to build it so that it may offer support to others who are affected."*

Although, during the data training process I excluded these comments, comments coming from the survivors have very similar content. Therefore, there is a risk that MUTATO might classified such comments as *Advocacy and foundation creation as a survivorship mechanism* stage of Breast Cancer Patient's Journey, which might have created bias.

Furthermore, during text classification, I have noticed that many comments include corporate advertisements, such as the comment below:

> *"Please share your story to help inspire others... I promise this is NOT spam....just a wonderful contest that my company is holding to reward all of your efforts for helping others fight Breast Cancer survive Breast Cancer or just sharing your infinite knowledge of Breast Cancer Awareness. Please take a look... CLICK TO VIEW CONTEST"*

Although excluded during data training, again, they might have been labelled as *advocacy and foundation creation as a survivorship mechanism*. That is why, for future research, I would recommend creating sub-labels to identify who is the author of the comment and be able to distinguish the comments written by relevant subjects.

# Learning Reflections

In this thesis, I aimed to demonstrate different analytical skills acquired throughout the course of my studies. As an Information Systems student I wanted to present that I am able to find solutions into existing social problems by using technological advances, such as in this case machine learning. Although, I do not have any background in public health, I wanted to find a topic which would enable me to have an impact and provide help where it is needed. Thus, in the end, this thesis enabled me to present my academic skills, but also gain the new ones. This wouldn't be possible without support and academic supervision of Professor Raghava Mao Mukkamala, who guided me throughout the whole process.

# Future Work

Something that future research should look into is perhaps getting the rights to fetch the data from private forums for breast cancer patients, where they need to register. First of all, it gives a guarantee that majority of comments will be written by patients, which will help to avoid classifying irrelevant comments such as those including advertisements. Secondly, privacy might help patients to open up about their feelings, which might influence the results of constructs reflecting more personal information such as emotional barriers towards *physical exercise* or *physiological state*. What is more, since Internet forums are even richer in discussions, one could hope that patients discuss topics related to rehabilitation and physiotherapy, which make it possible to investigate whether increased self-efficacy through empowerment has any direct effect on participation in rehabilitation.

# Conclusion

The findings of the analysis demonstrated that social media discussions contain content that can be classified as motivating, empowering and therefore having potential to increase patient's self-efficacy. Feelings of self-efficacy and empowerment have been possible to release through application of relevant sociological and psychological theories and conceptualising them using machine learning. Both admin posts and user comments reflected the constructs of Model for Efficacy Development, which indicates that information that is provided in the posts and comments can trigger self-efficacy. What is more, patients do seem to evince the characteristic, which are aligned with the constructs of Perceived Self-efficacy Model within Cognitive Process. Application of text into the constructs of this model helped to determine what are the current levels of perceived self-efficacy of breast cancer patients. This leaded to the conclusion, that patients have low belief in their abilities and negative perception of control. Nevertheless, they actively compare themselves to peers as social media creates a perfect environment to obtain information on other people's state and experiences. The content of discussions also turned out to be informative. As information and support are one of the key elements for development of empowerment, discussions again proved to have a potential to increase one's self-efficacy. Empowerment can also be generated by supportive messages and constructive feedback that patients provide each other.

The use of machine learning approach turned out to be an effective way of analysing large amount of data in terms of relevant theories. These theories served to develop domain-specific model for text classification, which helped to measure the amount of data reflecting certain constructs. Following Five-Step Content Analysis Process provided research reliability assurance.Performance measurement scores and inter-reliability agreement scores could perhaps be slightly. However, it was the result of being the single student writing this thesis. Having a partner who equally understand the theories and their constructs and who would have equal amount of time for training of the same amount of text would have undoubtedly lead for highers scores.

Using the same methodology, this research study could be developed further by analysing Internet forums devoted to breast cancer, as opposite to public Facebook walls, they provide much more privacy, which can increase patient's comfort to discuss more personal matters. What is more, since forums contain more data, one could track whether patients bring up rehabilitation-related topics. This way it would possible to test whether provided information that help to develop self-efficacy, perceived self-efficacy and empowerment realised from online discussions have direct result on rehabilitation performance. In the end, healthcare institutions should make use of the findings of this research and apply them in development of communication strategies between them and the patients.

# Bibliography

Abramson, K., Keefe, B., & Chou, W. Y. S. (2015). Communicating about cancer through Facebook: a qualitative analysis of a breast cancer awareness page. *Journal of Health Communication*, *20*(2), 237-243. http://dx.doi.org/10.1080/10810730.2014.927034

Amatya, B., Khan, F., & Galea, M. P. (2017). Optimizing post-acute care in breast cancer survivors: a rehabilitation perspective. *Journal of multidisciplinary healthcare*, *10*, 347. http://dx.doi.org/10.2147/JMDH.S117362

Anderson, C. (2010). Presenting and evaluating qualitative research. *American journal of pharmaceutical education*, *74*(8), 141.

Bandura, A. (1977). Self-efficacy: Toward unifying theory of behavioral change. Stanford University, Psychological review, Vol. 84, No. 2, 191-215

Bandura, A. (1993a). Perceived self-efficacy in cognitive development and functioning. Educational psychologist, 28(2), 117-148.

Bandura, A. (1993b). Perceived self-efficacy in cognitive development and functioning. Educational psychologist, 28(2), 117-148.

Bender, J. L., Jimenez-Marroquin, M. C., & Jadad, A. R. (2011). Seeking support on facebook: a content analysis of breast cancer groups. Journal of medical Internet research, 13(1). http://dx.doi.org/10.2196/jmir.1560

Blaney, J., Lowe-Strong, A., Rankin, J., Campbell, A., Allen, J., & Gracey, J. (2016). The cancer rehabilitation journey: barriers to and facilitators of exercise among patients with cancer-related fatigue. Physical therapy, 90(8), 1135-1147.

Breast Cancer Care (n.d.) In Facebook [Public Page], Retrieved from: https://www.facebook.com/breastcancercare/

Breast Cancer Now (n.d.) In Facebook [Public Page], Retrieved from: https://www.facebook.com/breastcancernow/

Breast Cancer Site (n.d.) In Facebook [Public Page], Retrieved from: https://www.facebook.com/TheBreastCancerSite/

Breast Cancer Support (n.d.) In Facebook [Public Page], Retrieved from: https://www.facebook.com/breastcancersupportuk/

Breast Cancer Facts & Figures 2017-2018. (2017). Atlanta: American Cancer Society. Retrieved from https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/breast-cancer-facts-and-figures/breast-cancer-facts-and-figures-2017-2018.pdf

Bryman, A., & Bell, E. (2011). Business Research Methods (3rd ed.) Oxford: Oxford University Press

Chiauzzi, E., DasMahapatra, P., Cochin, E., Bunce, M., Khoury, R., & Dave, P. (2016). Factors in Patient Empowerment: A Survey of an Online Patient Research Network. The Patient - Patient-Centered Outcomes Research, 9(6), 511-523. http://dx.doi.org/10.1007/s40271-016-0171-2

Christensen, A. J., Smith, T. W., Turner, T. W., Holman, J. M., and Gregory, M. C. (1990). Type of hemodialysis and preference for behavioral involvement: Interactive effects on adherence in end-stage renal disease.*Health Psychol.* 9: 225–236.

Coulson, N. S., & Shaw, R. L. (2013). Nurturing health-related online support groups: Exploring the experiences of patient moderators. Computers in Human Behavior, 29(4), 1695-1701. http://dx.doi.org/10.1016/j.chb.2013.02.003

Fisher, B., Anderson, S., Bryant, J., Margolese, R. G., Deutsch, M., Fisher, E. R., & Wolmark, N. (2002). Twenty-year follow-up of a randomized trial comparing total mastectomy, lumpectomy, and lumpectomy plus irradiation for the treatment of invasive breast cancer. New England Journal of Medicine, 347(16), 1233-1241.

Fisher, M. I., MSPT, O., & OTD, O. (2010). The power of empowerment: An ICF-based model to improve self-efficacy and upper extremity function of survivors of breast cancer. Rehabilitation Oncology, 28(3), 19.

Flick, U. (2015). Introducing research methodology: A beginner's guide to doing a research project. Sage.

Glanz, K., Rimer, B. K., & Viswanath, K. (Eds.). (2008). Health behavior and health education: theory, research, and practice. John Wiley & Sons.

Gutenbrunner C, Ward AB, Chamberlain MA, Bardot A, Barat M, Bensoussan L, et al. White Book on Physical and Rehabilitation Medicine in Europe. J Rehabil Med 2007; 39 suppl 45: 1–48

Dale, R., Moisl, H. and Somers H. (2000), *Handbook of Natural Language Processing*. CRC Press.

Hayes, S. C., Johansson, K., Stout, N. L., Prosnitz, R., Armer, J. M., Gabram, S., & Schmitz, K. H. (2012). Upper‐body morbidity after breast cancer. Cancer, 118(S8), 2237-2249.

Hoffman, A. J., von Eye, A., Gift, A. G., Given, B. A., Given, C. W., & Rothert, M. (2009). Testing a theoretical model of perceived self-efficacy for cancer-related fatigue self-management and optimal physical functional status. Nursing research, 58(1), 32. http://dx.doi.org/0.1097/NNR.0b013e3181903d7b

Hussain, A., Vatrapu, R., Hardt, D., & Jaffari, Z. A. (2014). Social data analytics tool: A demonstrative case study of methodology and software. In *Analyzing Social Media Data and Web Networks* (pp. 99-118). Palgrave Macmillan, London

Johnsen, A. T., Eskildsen, N. B., Thomsen, T. G., Grønvold, M., Ross, L., & Jørgensen, C. R. (2017). Conceptualizing patient empowerment in cancer follow-up by combining theory and qualitative data. Acta Oncologica, 56(2), 232-238. http://dx.doi.org/10.1080/0284186X.2016.1267403

Kidd MR. (2008) Personal electronic health records: MySpace or HealthSpace? (7652):1029-30)

Kim, H. S., & Sundar, S. S. (2014). Can online buddies and bandwagon cues enhance user participation in online health communities?. Computers in Human Behavior, 37, 319-333.

Lau, D. H. (2002). Patient empowerment--a patient-centred approach to improve care. Hong Kong medical journal= Xianggang yi xue za zhi, 8(5), 372.

Loane, S. S., & D'Alessandro, S. (2014). Empowered and knowledgeable health consumers: the impact of online support groups on the doctor–patient relationship. Australasian Marketing Journal (AMJ), 22(3), 238-245.

Lyst, M., Koval, S., & Li, Y. (2017). A Big Data Challenge: Visualizing Social Media Trends about Cancer using SAS® Text Miner. Midwest SAS Users Group.

Maunsell, E., Lauzier, S., Brunet, J., Pelletier, S., Osborne, R. H., & Campbell, H. S. (2014). Health‐related empowerment in cancer: Validity of scales from the Health Education Impact Questionnaire. Cancer, 120(20), 3228-3236. http://dx.doi.org/10.1002/cncr.28847

Mayer, R. (2010). Motivation Based on Self-Efficacy. Retrieved from https://www.education.com/reference/article/motivation-based-self-efficacy/

McCobin, A. A Theory of Empowerment. Students for Liberty.

McKinley, C. J., & Wright, P. J. (2014). Informational social support and online health information seeking: Examining the association between factors contributing to healthy eating behavior. Computers in Human Behavior, 37, 107-116.

Merolli, M., Gray, K., & Martin-Sanchez, F. (2013). Health outcomes and related effects of using social media in chronic disease management: a literature review and analysis of affordances. Journal of biomedical informatics, 46(6), 957-969.

Mielke, R., & Al-Haj, M. (Eds.). (2007). Cultural Diversity and the Empowerment of Minorities: Perspectives from Israel and Germany. Berghahn Books.

Mikkelsen, T. H., Søndergaard, J., Jensen, A. B., & Olesen, F. (2008). Cancer rehabilitation: psychosocial rehabilitation needs after discharge from hospital? A qualitative interview study. Scandinavian journal of primary health care, 26(4), 216-221. http://dx.doi.org/ 10.1080/02813430802295610

Morris, R., (1994) Computerized content analysis in management research: A demonstration of advantages & limitations, Journal of Management, vol. 20, no. 4, pp. 903–931, 1994

Mukkamala RR. Vatrapu R. Multi-dimensional Text Classification Tool, Unpublished manuscript, Copenhagen Business School

Murawa P., Murawa D., Adamczyk B., Polom K. (2014) Breast Cancer Trends: Actual methods of treatment and future trends, 19(3): 165–172.

Mustian, K. M., Katula, J. A., & Gill, D. L. (2002). Exercise: Complementary therapy for breast cancer rehabilitation. Women & Therapy, 25(2), 105-118. http://dx.doi.org/10.1300/J015v25n02_08

Olshannikova, E., Olsson, T., Huhtamäki, J. and Kärkkäinen, H. (2018). *Conceptualizing Big Social Data*. Journal of Big Data 2017. https://journalofbigdata.springeropen.com/articles/10.1186/s40537-017-0063-x

Patarca-Montero, R. (2004). Handbook of Cancer-Related Fatigue: What Does the Research Say?. CRC Press.

Perkins, D. D., & Zimmerman, M. A. (1995). Empowerment theory, research, and application. American journal of community psychology, 23(5), 569-579.

Reichert, J. R., Kristensen, K. L., Mukkamala, R. R., & Vatrapu, R. (2017). A Supervised Machine Learning Study of Online Discussion Forums about Type-2 Diabetes. In

IEEE Healthcom 17International Conference on E-health Networking, Application & Services (pp. 1-7). IEEE.

Renzi, C., Abeni, D., Picardi, A., Agostini, E., Melchi, C. F., Pasquini, P., Prudu, P., & Braga, M. (2001). Factors associated with patient satisfaction with care among dermatological outpatients. *British Journal of Dermatology, 145*, 617-623.

Sadan, E. (2004). Empowerment and community planning (pp. 193-218).

Scharkow, Michael. Online Content Analysis Using Super*vised Machine Learning*: An Empirical Evaluation, Paper presented at the annual meeting of the International Communication Association, TBA, Boston, MA, May 25, *2011*

Schwarzer, R., & Luszczynska, A. (2007). Self-efficacy. Health Behavior Constructs: Theory, Measurement, and Research. National Cancer Institute Website: http://cancercontrol. Cancer. Gov/constructs.

Shearer, N. B. C. (2009). Health empowerment theory as a guide for practice. Geriatric Nursing, 30(2), 4-10. http://dx.doi.org/10.1016/j.gerinurse.2009.02.003

Sherer, M., Maddux, J. E., Mercandante, B., Prentice-Dunn, S., Jacobs, B., & Rogers, R. W. (1982). The self-efficacy scale: Construction and validation. Psychological reports, 51(2), 663-671.

Siegel, R. L., Miller, K. D., & Jemal, A. (2013). Cancer statistics, 2013. CA: a cancer journal for clinicians, 68(1), 7-30. http://dx.doi.org/10.3322/caac.21166

Sirur, R., Richardson, J., Wishart, L., & Hanna, S. (2009). The role of theory in increasing adherence to prescribed practice. Physiotherapy Canada, 61(2), 68-77.

Strecher, V. J., McEvoy DeVellis, B., Becker, M. H., & Rosenstock, I. M. (1986). The role of self-efficacy in achieving health behavior change. Health education quarterly, 13(1), 73-92.

Tkaczuk, K. H., Kesmodel, S. B., & Feigenberg, S. J. (Eds.). (2016). Handbook of breast cancer and related breast disease. Springer Publishing Company.

Wanzer, M. B., Booth-Butterfield, M. & Gruber, K. (2004). Perceptions of health care providers' communication: Relationships between patient-centered communication and satisfaction. *Health Care Communication, 16*(3), 363-384

World Health Organization (2001) International Classification of Functioning, Disability and Health, WHO Press, http://psychiatr.ru/download/1313?view=name=CF_18.pdf

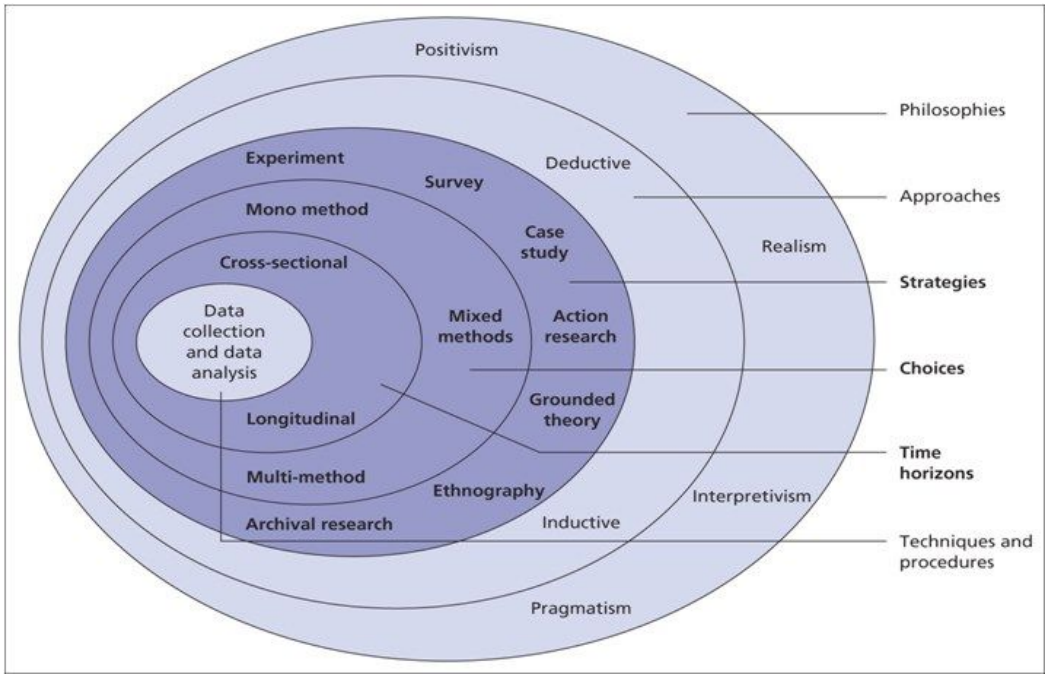
Van Berkel, J. J., Lambooij, M. S., & Hegger, I. (2015). Empowerment of patients in online discussions about medicine use. BMC medical informatics and decision making, 15(1), 24. http://dx.doi.org/10.1186/s12911-015-0146-6

Van't Spijker, A., Trijsburg, R. W., & Duivenvoorden, H. J. (1997). Psychological sequelae of cancer diagnosis: a meta-analytical review of 58 studies after 1980. Psychosomatic medicine, 59(3), 280-293.

# Appendices

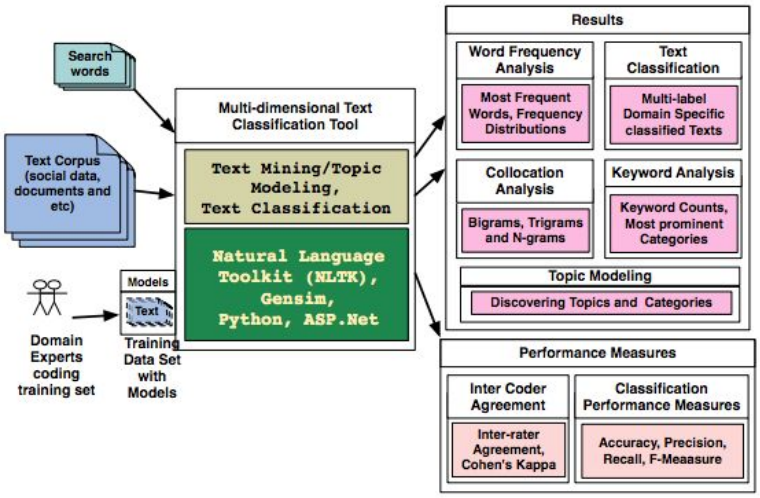
## Appendix A

Research Onion




## Appendix B

Mutato Architecture




# Appendix C

Domain Specific Model for Text Classification with the examples

| LABEL | DEFINITION | EXAMPLE<br>FACEBOOK COMMENT |
|---|---|---|
| **Self-Efficacy Theory: 4 sources of information** | | |
| Performance Accomplishments | Previous experiences of adherence | Grateful for being done with 34 rounds of radiation my family and friends' support and the wonderful people at the Radiation/Oncology Center and God's grace.<br><br>I dyed my hair pink ! This was my last day of radiotherapy. Now officially cancer free |
| Vicarious Experiences | Successful or unsuccessful attempts at adherence | My mum had breast cancer two years ago in November last year and had to have her breast off she has done fantastic and I am so proud of her hope |
| Verbal Persuasion | Support or encouragement by significant others and peers | My family have been very supportive but I need to socialise more and live a life and do more things and be encouraged to exercise more<br><br>Amen! And ladies be an advocate for your own care listen to your bodies if you feel something is wrong it probably is. This is your life! Get a second opinion if necessary. Thank you for sharing. I am a 3 yr survivor!<br><br>Go for the mastectomy Joan it's a walk in the park compared to chemotherapy and radiotherapy don't let these precancerous cells win |
| Physiological State | Patient's interpretation of whether or not she / | Sadly if mine returns after all this aggressive |

| | | |
|---|---|---|
| | he is capable of adhering at treatment | treatment it will return as stage 4....so I do have a fight ahead of me getting through this. But my will to live and beat this is stronger.... |
| **Perceived self efficacy model within cognitive process** | | |
| Conception of self-ability | Readiness or possibility to show self-ability | This is me I can't do wigs as they itch my head so I a bald and beautiful and am in the process of kicking cancers butt<br><br>currently fighitng breast cancer - diagnosed just after my 41st birthday with grade 2 - 2 lymph nodes also affected so just had my second lot of chemo yesterday. Its going to take a bit more than this to hold me back! |
| Social comparison | Relating and comparing to others | it was nice to chat to people on here that are going through the same situation as u are and understand how u feel<br><br>The initial shock is the worst I thought... it gets easier everyday. I'm in the battle with you since end of Feb. Stay strong and positive! You've got this!<br><br>Wow ive had some bad experiences but nothin that compare s to what youve been through. Wishin you every happiness for the future. .God Bless x |
| Framing feedback | Talking about or evaluating behaviour of other people. | Congrats on completion of your chemo. Stay strong and continue to smile. I can see in your face that you have everything you need to beat this.<br><br>I admire your strength and courage to continue with the journey of life I wish you the best that life has to offer.<br><br>You are beautiful and so strong!! So many people are gonna admire you for being so strong and staying positive!! |
| Perception of controllability | Exhibiting self-control ability / ability of management of process | Today is my last day of chemo. I will get 4 weeks off and then start radiation. One step closer! I am a very busy Mom with 6 kids so I won't let this bring me down |
| **Psychological Empowerment Model** | | |
| Interpersonal | Am I motivated enough to do it? | Sadly if mine returns after all this aggressive treatment it will return as stage 4....so I do have a fight ahead of me getting through this. But my will to live and beat this is stronger.... |
| Interactional | Show that one has skills knowledge to go for next step | Good luck Hun I'm going to have my left Brest operated On the 28th October this month but mine is 2 centre meters so glad I got it erly |
| Behavioral | Represents actions taken by an individual to achieve or influence specific outcomes | They were taken out Thursday. Feb 18 put in a drain. Went to the doctor this morning. Checked my drain healing fine. Doctors office just called. No sign of cancer in my nodes. Wheee. Still go to the hematologist Tuesday. Make sure the little |

| | | |
|---|---|---|
| | | creep didn't morph elsewhere. I'll do what I have to do.<br><br>I was diagnosed in Nov 2014 stage 1 ids with no lymph nodes effected. I had a lumpectomy with 36 treatments of radiation. I now have to take a pill for 5 years due to estrogen positive. To God be the glory. I will fight with all I have.. We are survivors! |
| **Personality Traits** | | |
| Openness | Being open to new learnings and experiences as well as being curious, imaginative, adventurous and focused on tackling new challenges. | How do I get a free mammogram I am a breast cancer survivor who doesn't have health insurance anymore. Thanks in advance.<br><br>This is me I can't do wigs as they itch my head so I a bald and beautiful and am in the process of kicking cancers butt |
| Conscientiousness | An individual with this trait characterizes with goal-directed behavior, enjoys having a set schedule and pays attention to details. | Last day of chemo! Now onto radiation. This journey has been a roller coaster ride. Being positive had made the difference. |
| Extraversion | Opposite to introversion; extrovert likes meeting new people, making new friends, being center of attention. Such an individual is outgoing, social and talkative. | Boy do I ever love to laugh and I also like to make other people laugh or make them smile that makes my day and also theirs :) |
| Agreeableness | This trait includes attributes such as conforming to social norms, trust, altruism, kindness, affection, generosity, optimism and trustworthiness. | God Bless every breast cancer patient and their personal fight with treatment!<br><br>I have been there i no what you are talking about we are truly bless and also beautiful let us all be strong in our faith. |
| Neuroticism | This characteristic is linked to sadness, moodiness, and emotional instability. People with this trait are prone to experiencing mood swings, anxiety and irritability. | Why then did I get Breast Cancer as I didn't drink or smoke ? Answer me this Why do Children get Cancer?<br><br>Protocel ... I'll never do Chemo or Radiation again. If my Cancer comes back I'm doing Protocel again and checking out Holistic medicine. I recommend that to any cancer patient. It worked for me. Chemo kills .. Good luck.<br><br>Thanks for the reminder that i am still not safe from this horrible disease nor will i ever be!!! Just what i need |
| **Breast Cancer Patient's Journey** | | |
| Knowledge of breast cancer | Prior knowledge and understanding about breast health | When my mum died I asked to go for a mammogram (I was in my very early thirties at the time) and my GP agreed that I should have regular screening.<br><br>I think i have breast cancer and im so scared to go I had gotten a lump overnight but i waited a month and it wont go away and i have an enlarged lymph node under armpit<br><br>I've just found one I'm really worried it's not like my other lumps I'm at the hospital on Wednesday I'm stressing so bad my nanna past |

| | | |
|---|---|---|
| | | away with breast cancer |
| Stage at diagnosis and treatment | Emotions tapped at the time of breast cancer diagnosis | I have been just diagnosed with invasive breast ductal carcinoma invloving the axillary lymph nodes. I am feeling lost and shocked. |
| Support system | Social support as a critical ingredient to physical and mental<br>health when facing cancer care | Thank you very much for all your kind comments. Joined the page suggested to me and I really hope for the best.<br><br>Thank you. I have contacted breast cancer care a few times since my diagnosis. You've been amazing and your information booklets have been a huge help to me |
| Cost of treatment | Financial situation and health insurance in relation to treatment costs | In the midst of battling cancer I wasn't able to work and that caused me to enter into financial debt. I have reached out to different churches organizations and contacted the CEO's of the companies. All of my bills are currently going to collections I am about to lose my house and I barely have food to feed my daughter. |
| Coping strategies | Fear and anxiety regarding recurrence | **I** will never win because the threat is always there.<br><br>9 years clear. still take meds and still wonder if it will return. |
| Meaning of survival | Appreciate and realize how “precious life is”and the importance of living “one step at a time | I had a wide local excision and axillary node clearance chemotherapy and radiotherapy. I now appreciate each and every day and take pleasure in the small things in life I've learned to 'stop and smell the roses'!. I'm happier and more confident.<br><br>Grateful for being done with 34 rounds of radiation my family and friends' support and the wonderful people at the Radiation/Oncology Center and God's grace. |
| Advocacy and foundation creation as a survivorship mechanism | Educating other women and creating foundations to provide education, generate awareness and raise funds for breast cancer research | A team of 17 people called simply the breast completed tough missed last sat including myself to celebrate my 5 years all clear from breast cancer and raised over £3k. It was very painful but so worth it. The bruises now need to heal!<br><br>Have always supported the pink ribbon campaign as it has always been a cause close to my heart. It was brought closer three years ago when I was diagnosed with triple negative breast cancer. Following surgeries and chemo have been in remission since June 2015. Please support the pink ribbon research into breast cancer saves lives information is available (when you're trying to make sense of it all) and support is available if you need to just talk. Keep up the excellent work. |
| **Most common barriers towards physical exercise** | | |
| Physical barriers | Lack of energy or vigor, lethargy, feeling tired, decreased strength, and trouble concentrating | Since she finished chemo over a year ago she has been aching so much with her back knees and ankles. |

| | | |
|---|---|---|
| | | I feel so sick all of the time. I sick for at least 2 weeks after Chemo. Does everyone feel this bad.<br><br>I am on letrozole with body aches especially my legs. I do stand up all day at work. My dr told me it is permanent side effect from chemo |
| Emotional barriers | Anxiety, depression, fear and stress | I find that reading light hearted stuff has helped me immensely with depression |
| Cognitive barriers | Lack of motivation, lack of perceived control, lack of perceived ability | Bless you. I have no kids at home and the chemo puts me down. I'm useless on chemo. |

# Appendix D.

Performance Measures of Text Classifiers

| **Model 1: Self-Efficacy Theory: 4 sources of information** | | | | |
|---|---|---|---|---|
| **Classifiers** | **Precision** | **Recall** | **F1-score** | **Accuracy** |
| Multinomial NB | 0.57 | 0.58 | 0.57 | 0.577 |
| Linear SVC | 0.56 | 0.57 | 0.55 | 0.568 |
| Logistic Regression | 0.54 | 0.63 | 0.56 | 0.633 |
| Passive Aggressive | 0.57 | 0.58 | 0.56 | 0.577 |
| SVM SGD | 0.53 | 0.61 | 0.53 | 0.614 |
| Voted Accuracy | - | - | - | 0.596 |
| **Model 2: Perceived Self-Efficacy Model within cognitive process** | | | | |
| **Classifiers** | **Precision** | **Recall** | **F1-score** | **Accuracy** |
| Multinomial NB | 0.55 | 0.54 | 0.53 | 0.534 |
| Linear SVC | 0.71 | 0.61 | 0.59 | 0.606 |

| | | | | |
|---|---|---|---|---|
| Logistic Regression | 0.57 | 0.58 | 0.53 | 0.575 |
| Passive Aggressive | 0.66 | 0.61 | 0.59 | 0.606 |
| SVM SGD | 0.67 | 0.61 | 0.58 | 0.606 |
| Voted Accuracy | - | - | - | 0.606 |
| **Model 3: Psychological Empowerment Model** | | | | |
| **Classifiers** | **Precision** | **Recall** | **F1-score** | **Accuracy** |
| Multinomial NB | 0.52 | 0.46 | 0.43 | 0.463 |
| Linear SVC | 0.23 | 0.39 | 0.29 | 0.390 |
| Logistic Regression | 0.19 | 0.44 | 0.27 | 0.439 |
| Passive Aggressive | 0.30 | 0.44 | 0.34 | 0.439 |
| SVM SGD | 0.20 | 0.44 | 0.27 | 0.439 |
| Voted Accuracy | - | - | - | 0.463 |
| **Model 4: Personality Traits - the five-factor model** | | | | |
| **Classifiers** | **Precision** | **Recall** | **F1-score** | **Accuracy** |
| Multinomial NB | 0.45 | 0.37 | 0.35 | 0.369 |
| Linear SVC | 0.45 | 0.38 | 0.34 | 0.380 |
| Logistic Regression | 0.48 | 0.39 | 0.34 | 0.392 |
| Passive Aggressive | 0.37 | 0.39 | 0.36 | 0.392 |
| SVM SGD | 0.48 | 0.43 | 0.39 | 0.428 |
| Voted Accuracy | - | - | - | 0.392 |
| **Model 5: Breast Cancer Patient's Journey** | | | | |
| **Classifiers** | **Precision** | **Recall** | **F1-score** | **Accuracy** |
| Multinomial NB | 0.56 | 0.57 | 0.56 | 0.573 |
| Linear SVC | 0.62 | 0.64 | 0.61 | 0.639 |
| Logistic Regression | 0.64 | 0.62 | 0.56 | 0.617 |
| Passive Aggressive | 0.60 | 0.63 | 0.61 | 0.633 |
| SVM SGD | 0.60 | 0.62 | 0.59 | 0.617 |
| Voted Accuracy | - | - | - | 0.628 |
| **Model 6: Most common barriers towards physical exercise** | | | | |
| **Classifiers** | **Precision** | **Recall** | **F1-score** | **Accuracy** |
| Multinomial NB | 0.79 | 0.69 | 0.64 | 0.611 |
| Linear SVC | 0.75 | 0.67 | 0.65 | 0.666 |
| Logistic Regression | 0.47 | 0.67 | 0.55 | 0.666 |

| | | | | |
|---|---|---|---|---|
| Passive Aggressive | 0.75 | 0.67 | 0.65 | 0.666 |
| SVM SGD | 0.75 | 0.67 | 0.65 | 0.666 |
| Voted Accuracy | - | - | - | 0.666 |