Copenhagen Business School

MSc in Business Administration and Information Systems (E-Business)

Master's Thesis (CEBUO2000E) - Kontraktnr: 10774

Data Analysis of Public Health Parameters

Deriving Insights About Motivation for Physical Activities Through Supervised Machine Learning Techniques

Date: 15 May 2018
Page count: 80 STU-count: 133947
Supervisor: Raghava Rao Mukkamala

Student: Tsveta Toshkova Atanasova: 61538

Abstract

Health care organizations are increasingly arousing concerns about the lack of physical activities, which eventually leads to a range of sedentary lifestyle diseases. One of the most cited reasons for lack of physical activities is motivation. This study considered motivation as a public health parameter, which if studied and applied in practise can promote physical activity and with this a healthier lifestyle. Therefore, the aim of this study was to explore how predictors of motivation for physical activities can be derived from social media texts, and which insights this investigation would provide for the promotion of healthier lifestyle.
A systematic literature review of Self-Determination Theory and its application in the fields of physical activities and social media was conducted for the development of conceptual model of motivation. A total of 55259 social media texts from five Facebook pages and three online forums, discussing physical activities, collected for this investigation. Supervised machine learning were applied on the data by firstly manually coding a training set, and afterwards applying five text classification algorithms.
The findings point that supervised text classification is an efficient method for deriving predictors of motivation from social media texts. Additionally, the results show consistent support that intrinsic motivation is most often associated with physical activities. Intrinsic motivation was also most often related to intrinsic goals and the needs of autonomy and competence. Intrinsic people were also most often related to open, conscience and agreeable personalities. The text classification also extracted valuable reflections on topics from the social texts. Among others, physical injuries, bodybuilding, advice seeking, and encouragement tips were the most often discussed topics. Based on these insights it was concluded that being present when people need support, turning extrinsically motivated into into intrinsically, and engaging on social media are appropriate guidelines for the promotion of motivation for physical activities, which practitioners, health care organisations or sport centres should take into account in order to stimulate healthier lifestyle.

Key words: Self-Determination Theory, Motivation, Physical Activities, Text Classification, Supervised Machine Learning, social Data Analytics

1. Introduction 5

2.Background 7

2.1 Self-Determination Theory (SDT) 8

2.1.1 Understanding Motivation 9

2.1.1.1 Self-Determination Continuum 9

2.1.1.2 Intrinsic Motivation 11

2.1.1.4 Amotivation 13

2.1.2 Basic Needs 13

2.1.2.1 Autonomy 14

2.1.2.2 Competence 14

2.1.2.3 Relatedness 14

2.1.3 Goal Contents Theory 15

2.1.3.1 Intrinsic Goals 15

2.1.3.2 Extrinsic Goals 15

2.2 Application of SDT in Academic Research 16

2.2.1 Intrinsic and Extrinsic Motivation for Physical Activities 16

2.2.2 Motivational Goals 17

2.2.3 Exercise Variety 17

2.2.4 Social Media Engagement 18

2.3 Conceptual Model 19

2.4 Implications of Motivation for Physical Activities 21

2.5 Impact of Big Social Data 23

2.5.1 Big Social Data 23

2.5.1.1 Attributes of Big Social Data 24

2.5.2 Big Social Data Application in Academic Research 25

3. Methodology 26

3.1 Research Design 26

3.2 Research Strategy 27

3.3 Data Collection and Preparation 30

3.3.1 SODATO 30

3.3.2 Sample Description 31

3.3.3 Data Preparation 32

3.4 Data Modeling and Analysis 32

3.4.1 Manual Text Classification 33

3.4.2 Supervised Machine Learning Techniques 34

3.4.2.1 Naive Bayes Multinomial 36

3.4.2.2 Linear Support Vector Classifier 37

3.4.2.3 Support Vector Machine (Stochastic Gradient Descent) 38
3.4.2.5 Passive Aggressive Classifier 38
3.4.2.4 Logistic Regression Classifier 39
3.4.3 Performance of the Models 39

4. Results 41
4.1 Performance of the Overall Model 41
4.1.1 Overall Model Performance 41
4.1.2 Basic Needs Model Performance 42
4.1.3 Goals Model Performance 46
4.1.4 Personality Traits Model Performance 49
4.1.4 Types of Motivation Model Performance 53
4.2 Overall Distribution of Types of Motivation 57
4.2.1 Intrinsic Motivation 57
4.2.2 Extrinsic Motivation 60
4.2.3 Amotivation 61
4.2.4 Motivation and Social Media Activity 63
4.3 Datasets - Actors and Actions 63
4.3.1 Facebook 63
4.3.2 Forums 64

5. Discussion 65
5.1 Evaluation of the Model 66
5.1.1 Review of the Process 66
5.1.2 Ethical Considerations 67
5.1.3 Dataset Limitations 68
5.2 Turning Data into Information 69
5.2.1 Predictors of Motivation 69
5.2.2 Commonly Identified topics 71
5.2.2.1 Physical Injuries 72
5.2.2.2 Bodybuilding 72
5.2.2.3 Advice Seeking 73
5.2.2.4 Encouraging Tips 74
5.3 Turning Information into Actionable Knowledge 75
5.3.1 Being Present When People Need Support 75
5.3.2 Turning Extrinsic into Intrinsic Motivation 75
5.3.3 Social Media Engagement 76
5.4 Reflections, Limitations and Future Research 78

6. Conclusion 78

7. Acknowledgement 80

8. Reference list 81

9. Appendix 90

9.1 CRISP-DM - data life cycle (Chapman et al., 2000) 90
9.2 Data Source Description 90
9.3 MUTATO Architecture 93
9.4 Detailed Confusion Matrix 94
9.5 Total Basic Needs Model Performance 95
9.6 Total Goals Model Performance 96
9.7 Total Personality Traits Model Performance 97
9.8 Total Types of Motivation Model Performance 98
9.9 Motivation and its Predictors 99
9.10 Distribution of Categories Across Datasets 100

1. Introduction

Lack of regular physical activity is a critical public-health concern, which has in the last two decades gained more and more attention around the world (Kwan et al., 2011). According to practitioners, lack of physical activity is a risk factor not only for sedentary lifestyle but even more critically for common diseases like cardiovascular disease, obesity, diabetes, osteoporosis and some cancers (Fortier et al., 2011). Many government around the world are advising people to exercise, as there seem to be a tendency that people are not physically active as much as they are being recommended (Wilson and Rodgers, 2004).
Data from Eurobarometer survey on sport and physical activity conducted in 2014 shows that 41% of Europeans exercise or play sport once a week, while 59% never or seldom do so (TNS Opinion and Social, 2014). What is more frustrating is that the overall proportion of non-exercisers has increased from 39% to 42% between 2009 and 2014 respectively. Even though Northern Europe is more physically active than the South and East, the survey has proven that people in general are not following the health recommendation on exercise and physical activities. The survey also shows that local authorities in particular could do more to encourage citizens to be physically active (ibid).
Another survey conducted by The World Health Organization (WHO) has also shown that a larger proportion of individuals are not physically active at the recommended levels not only in Europe but all around the world (Fortier et al., 2011).
Previous research has tried to uncover the reasons why people do not exercise, and found that motivation is among the top five reported barriers to physical activity (Fortier et al., 2011). For inactive people, motivation is the greatest reported barrier to physical activities. Additionally, there can be many various reasons to why people are experiencing motivation as a barrier and understanding the mechanisms of motivation by which people adopt and maintain regular exercise is a priority (Kwan et al., 2011; Wilson and Rodgers, 2004).
Recently there has been a tendency that people are looking for health care information online (Li and Wang, 2018). Even though, this information has been considered as something private before, a study from 2013 shows that in USA 80% of internet users have looked for health information online within the past year (msnbc.com. 2018). The study also found that there is a tendency that people are much more interested in reading and creating content about exercise and health care online.

Because a lot of content is generated online, many researchers have started to use social data analytics to learn and understand diseases better and be better at diagnosing and preventing them. As seen from the literature, lack of physical activity is a reason for many diseases, therefore motivation should be studied as a prevention measure to this outcome. Even though a considerable volume of research has studied motivation, at the time of this research none has tried to investigate predictors of motivation for physical activities derived from social media texts. In this fashion, this study believes that motivation can be studied in a more innovative way, and therefore it intends to apply supervised machine learning to a sample of social media data collected from five Facebook pages and three forums, all related to physical exercise.

Social data in the form of user-generated content especially from social media platforms can possibly be a new innovative way to extract valuable insights and to understand motivation for physical activities and exercise. Therefore this study aims at answering these questions:

1. How can predictors of motivation for physical activities be identified from social media texts?
2. Which insights about motivation for physical activities can be derived from social media conversations?
3. How can these insights be used as guidelines for promoting motivation for physical activities to the benefit of practitioners, health care organisations or sport centres?

In order to understand motivation in the field of exercise and physical activities, this study followed the recommendations from previous research, and adopt Self-Determination Theory (hereinafter SDT) as the framework for exploring this topic (Kwan et al., 2011; Wilson and Rodgers, 2004). Even though SDT has been applied intensively, previous research has identified that there is a strong need for further research that identifies predictors of motivation for physical activities (Silva et al., 2008; Kwan et al., 2011; Wilson and Rodgers, 2004).

The motivation behind this study was to provide actionable insights for the benefit of those who are affected by lack of motivation for physical activities, namely patients, doctors, public healthcare organization, and pharmaceutical companies. Additionally, this study aimed, in regard to the evidence from EU and WHO, at providing guidelines for promoting physical activity, which is clearly an increasing public health priority (Edmunds, Ntoumanis and Duda, 2006).

This study unfolds firstly into a Background section which lays the foundation of this research. The section elaborates on the fundamental elements of SDT and provides a detailed explanation of the different types of motivation. The theory is further supported by systematic literature review in the field of physical activities and social media. The section sums up the arguments for why SDT is relevant and an effective tool for identifying motivation, and the arguments for why this study should apply social data analytics. Thereafter, the paper discusses people's tendency of reading and creating health care content online, and explanation of big social data.
Next section provides a justification for the methodology consideration undertaken in this study. This included research design, research strategy, data collection and preparation, and data modelling and analysis. In the results section, the findings are visualized, and further explained and disputed in regard to the research questions in the discussion. The paper is summarized in the conclusion.

2.Background

In order to promote physical activities, previous research highlights the need for a clearer understanding of motivation and how it facilitates physical activities (Silva et al., 2008). To gain understanding of motivation, the current study adopted Self-Determination Theory as the theoretical framework. The theoretical framework was developed on the basis of literature review on SDT (Saunders, Lewis and Thornhill, 2009).
Firstly, SDT developed by its founders and authors was reviewed. Texts written by Edward L. Deci and Richard Ryan were selected, discussed and utilized as the backbone for the understanding of SDT.
Secondly, a review of academic literature adopting SDT was conducted by searching for SDT in ScienceDirect, Scopus and Google Scholar databases. The focus of the study was narrowed down to literature investigating on one hand motivation for physical activities, and on the other hand research conducted in the settings of social media. This was done in order to discover whether previous research has seen opportunities in uncovering

motivation out of social media text. In order to find relevant studies, the databases were searched for “Self-Determination Theory”, “Motivation”, “Physical activities” in combination with “Social Media”, “Social data”, “Big Social data”, “Big data”. In total of 523 articles were found. Articles not written in English, with no full-article access, and not having motivation as the main focus were excluded. In total of 24 articles were included since these discussed motivation for physical activities, social media, and relation to motivation. These articles were believed to provide a valuable primary understanding and facts about social media in the realms of motivation for physical activities. None of the articles derived motivation out of social media data and that provided an argument for why this study should use data analytics for the investigation of motivation.

The following section unfolds SDT and underlines how we can understand motivation in general, and also in the context of physical activities and social media.

2.1 Self-Determination Theory (SDT)

The initial work leading to SDT dates back to the 1970s, though the first relatively comprehensive statement of SDT was presented in 1980s (Deci and Ryan, 1985). SDT has reached its flourishment period in the beginning of the 20s, where a particular amount of research applying SDT to fields as education, health care, relationships, psychotherapy, sports and exercise, goals and others were conducted (Deci and Ryan, 2008). Today, SDT has increasingly become a basic framework in the areas of health and physical activity promotion (Ryan and Patrick, 2009). A lot of research has contributed to explain motivation, especially motivation in physical activities (Hagger and Chatzisarantis, 2007). Following is an in-depth explanation of Self-Determination Theory.

SDT is organismic theory which stems from psychology (Deci and Ryan, 2015). It is a “motivational theory of personality, development, and social processes that examines how social contexts and individual differences facilitate different types of motivation”, and among others, it predicts physical activities (Deci and Ryan, 2015, p.1). SDT embraces both an organismic and a dialectical framework for the study of personality growth, development, wellness and goal pursuit (Ryan and Deci, 2002; Deci and Ryan, 2008; Lutz, Karoly and Okun, 2008, Brickell and Chatzisarantis, 2007). SDT is based on the assumption that human beings have “natural innate and constructive tendencies to develop an even more elaborated and unified sense of self” (Ryan and Deci, 2002, p.5).

SDT consists of among others, five mini theories which contribute to its complex and detailed nature. At first, SDT presents the Cognitive Evaluation Theory (CET) which distinguishes between intrinsic, extrinsic motivation and amotivation. Organismic Integration Theory (OIT) further represents four different subtypes of extrinsic motivation. According to Causality Orientation Theory (COT), there are autonomous, controlled and impersonal orientations. Based on the degree of orientation which goes from autonomous, controlled and at last impersonal, motivation is represented on a continuum starting from intrinsic, extrinsic and at last amotivation, respectively. The types of motivation are further dependent on the satisfaction of competence, autonomy and relatedness needs, which are outlined by Basic Psychological Needs Theory (BPNT). Finally, Goal Contents Theory (GCT) distinguishes between intrinsic and extrinsic aspiration which also influence the types of motivation. The following sections discussed these in turn.

2.1.1 Understanding Motivation

According to SDT, people are active, willing to integrate new material into their own sense of self (Deci and Ryan, 2015). In the same time SDT accepts that environment plays an important role as it either nourishes or disrupts this integrative process (Ryan and Deci, 2002). Thus, it is assumed that the active, growth-oriented individual and the social context are the central explanations of motivational behaviour and development of humans (Deci and Ryan, 2015). Ryan and Deci (2000) explained motivation as consisting of energy, direction and persistence. SDT looks at motivation in a way that tries to uncover what type of motivation human exhibit at any given time (Ryan and Deci, 2000). Therefore, SDT focuses on types rather than amount of motivation (Deci and Ryan, 2008).

2.1.1.1 Self-Determination Continuum

Deci and Ryan (1985) established the self-determination continuum, which postulates that the types of motivation lay on a continuum going from intrinsic to extrinsic and amotivation (see Figure 1). Thus, the types of motivation depends on the degree to which people's orientation is autonomous, controlled, or impersonal, respectively.

Autonomous orientation is also known as self-determined. It occurs when people act with a full sense of willingness and volition (Deci and Ryan, 2008). Their behaviour is motivated out of their personal interest and values, and sense of enjoyment. Deci and Ryan (2015) also explain that autonomous orientation is typically accompanied by the experience of positive affect, flexibility, and choice.

Controlled orientation occurs when people act out of coercion, seduction or obligation. Their actions are influenced by pressure and compulsion, rather than a result of their own choices (Deci and Ryan, 2015).

Impersonal orientation occurs when people tend to focus on indications from the social context, which would increase their sense of incompetence and inability to obtain certain outcomes (Deci and Ryan, 2015).

Previous research has provided that intrinsic motivation would correspond to more autonomous orientations, while extrinsic motivation would correspond to more controlled orientations. Lastly, amotivation would correspond to impersonal orientations (Silva et al., 2008). The authors state that despite of whether it is autonomous or controlled orientation, there is a presence of motivation. On the other hand, impersonality orientation lead to amotivation, or the absence of motivation (Deci and Ryan, 2015). The ability to distinguish between autonomous and controlled orientations indicates that positive motivation consequences, such as persistence and psychological well-being, are influenced by autonomous orientations rather than controlled (Wilson and Rodgers, 2004). Therefore, being able to access these different types of orientation can be used for predicting the quality and maintenance of particular types of motivation, which would promote physical activities (Deci and Ryan, 2015).

Each of these orientations exist to some degree within every person. However, people cannot be categorized as one or another, but rather as a combined measure of degree of each (Deci and Ryan, 2015). Therefore, this study would not consider these as direct predictors of types of motivation.

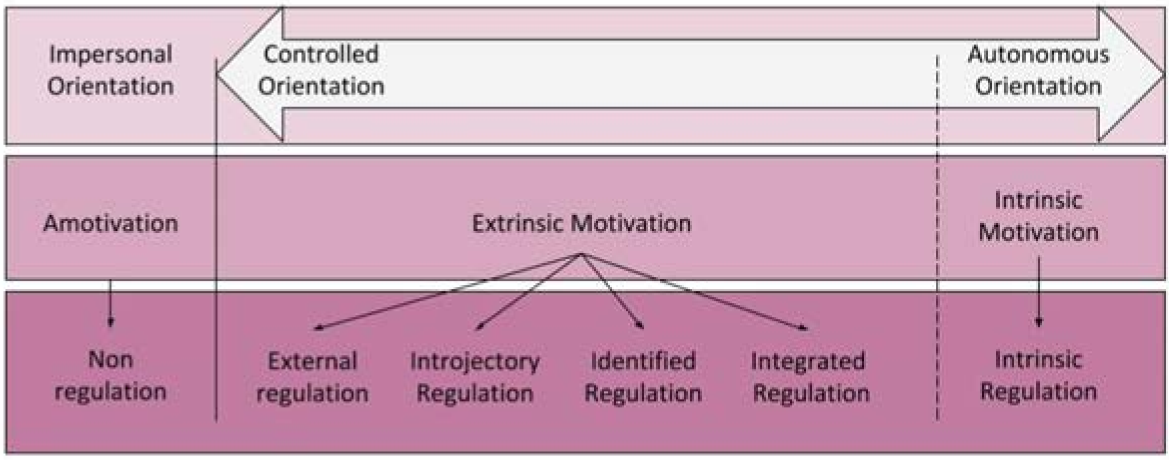


Figure 1. Self-determination Continuum - the arrow shows how types of motivation range from Amotivation to Intrinsic motivation based on the orientation type.

The distinction between the three types of motivation is enabled by the Cognitive Evaluation Theory (CET), and each of the types are explained in turns.

2.1.1.2 Intrinsic Motivation

Intrinsic motivation is a manifestation of the human tendency towards learning and creativity (Ryan and Deci, 2000). When people are intrinsically motivated they would engage in activities which they find interesting, enjoyable and fun (Deci and Ryan, 2015). Also, intrinsic motivation is exhibited in one's seeks of novelty, challenges, and one's seek to improve capacities, or explore and learn (Ryan and Deci, 2000; Brickell and Chatzisarantis, 2007). A well-known example of intrinsic motivation is kids and their engagement in playing games. In regard to physical activities, the main reason for people to engage in these activities is because of the inherent pleasures, enjoyment and satisfactions these provide (Ryan and Patrick, 2009).

It has been examined that receiving a positive feedback, or acknowledging one's feeling and perspectives, having a choice and opportunity of self-direction enhance the intrinsic motivation, because people have more autonomy (Ryan and Deci, 2000). On the other hand, punishment, threads, deadlines, surveillance can undermine intrinsic motivation (Deci and Ryan, 2015). Intrinsic motivation is associated with intrinsic regulation, which is believed to be the most self-determined regulation (Silva et al., 2008).

2.1.1.3 Extrinsic Motivation

When people are extrinsically motivated they would engage in activities because there is some kind of reinforcers. In the past people were extrinsically motivated to take an action because they would be able to get hold on water, food. Nowadays, extrinsic motivation is also exhibited in the pursuit of rewards such as money, prizes, or the avoidance of undesirable situation, or in the wish for social approval (Deci and Ryan, 2015). Many scholars explain this type of motivation with the example of the “carrot and stick” (Deci and Ryan, 2015). In regard to physical activities, extrinsic motivation is prevalent because of one’s gain of carrying the activity (Ryan and Patrick, 2009). Possible reasons for extrinsic motivation in sports are improvement of health, appearance, physical shape (Ryan and Patrick, 2009).

Giving these extrinsic rewards to intrinsically motivated people acting out of own interest, would result in diminishing their intrinsic motivation for the activity. Thus, extrinsic rewards would have negatively effect on intrinsic motivation (Deci and Ryan, 2015).

Organismic Integration Theory (OIT) is another mini-theory used in SDT (Ryan and Deci, 2000). OIT explores the degree to which an activity is considered as autonomous or internalized (Ryan and Deci, 2000). In this regard, IOT refers to internalization as the degree to which individuals “accept” values and regulations as their own, and integration as the act of transforming these regulation as one’s own, emanated from one’s sense of self (Ryan and Deci, 2000). Based on OIT, there are four regulations of extrinsic motivation, which are influenced by contextual factors to either promote or hinder the internalization and integration of physical activities (Ryan and Deci, 2000; Edmunds, Ntoumanis and Duda, 2006).

External regulation is explicit when people act because of external contingencies of reward or punishment that have not been internalized (Deci and Ryan, 2015). For example, a patient might join a gym because his physician recommends it (Fortier et al., 2011; Brickell and Chatzisarantis, 2007).

Introjected regulation occurs when people take in values and regulations, but only partially accept and internalize them as their own (Deci and Ryan, 2015; Ryan and Deci, 2000). For example, people who exercise to avoid feeling guilty if they do not act according to someone else’s standards (Brickell and Chatzisarantis, 2007; Fortier et al., 2011).

Identified regulation occurs when people identify their behaviour with personal value and importance, and accept the behaviour as it is their own (Deci and Ryan, 2015; Ryan and Patrick, 2009). For example, people would exercise because of the value, usefulness, or importance of exercise (Brickell and Chatzisarantis, 2007; Fortier et al., 2011).
Integrated regulation occurs when people act with a full sense of volition and choice. It also means that people have fully accepted their identified regulations as their own (Ryan and Deci, 2000). For example, people who exercise to maintain good shape, because being fit is part of who they are (Brickell and Chatzisarantis, 2007).

2.1.1.4 Amotivation

Another type of motivation is amotivation. It occurs when people have no motivation, or do not feel content to take an action, or do not value the outcome of the action, or do not feel competent to do it (Deci and Ryan, 2015, Ryan and Deci, 2000). Amotivation is complex because there can be many reasons to why people are not motivated to take a certain action. One explanation to amotivation can be the fact that one does not feel competent, knowledgeable and skilful in carrying the activity (Ryan and Patrick, 2009). Another explanation can be that people do not feel connection between the action and the desired outcomes (Ryan and Patrick, 2009). Lastly, amotivation can simply be caused by the fact that people do not want to act, do not see value or interest in the activity (Ryan and Patrick, 2009). Amotivation is associated to impersonal type of orientation, which is related to self-derogation, anxious, and ineffective and depressive symptoms. This can be caused by diminishing effects from the social environment and the inability to satisfy one's needs, or ill-being.

2.1.2 Basic Needs

Outlined in the Basic Psychological Needs Theory (BPNT) is that there are three basic, universal and cross-developmental psychological needs, namely the needs for competence, autonomy and relatedness (Ryan and Patrick, 2009). The basic needs are believed to be innate, not learned, and seen in humans across genders and cultures. Within the SDT terminology, the basic needs are referred to as "basic psychological needs" (Ryan and Deci, 2002).

2.1.2.1 Autonomy

The need of autonomy refers to “being the perceived origin or source of one’s own behaviour” (Ryan and Deci, 2002, p.8). Thus, autonomy is the act of going things from personal interest and integrated values. When people feel autonomous, they experience their behaviour as self-organized and endorsed, and fully engaged in the activity (Ryan and Patrick, 2009).

2.1.2.2 Competence

The need for competence refers to “feeling effective in one’s ongoing interactions with the social environment and experiencing opportunities to exercise and express one’s capacities” (Ryan and Deci, 2002, p.6). Competence make people to see challenges and to maintain and develop their skills and capacities. Thus, competence is a “felt sense of confidence and affectance in action” (Ryan and Deci, 2002, p.7). In regard to physical activities, it has been proven that positive feedback enhances the feeling of competence and leads to increase in motivation, and the other way around (Ryan and Patrick, 2009).

2.1.2.3 Relatedness

The need for relatedness refers to “feeling connected to others...having sense of belongingness and connectedness both with other individuals and with one’s community” (Ryan and Deci, 2002, p.7). Relatedness is represented in the tendency to connect with others and be integral part of the community. Relatedness can be endorsed by the warmth, care and involvement other individuals convey (Ryan and Patrick, 2009).

When the social environment satisfies these three basic needs, a vital human functioning is supported, whereas when these three basic needs are thwarted, it is suggested that the environment is antagonistic (Ryan and Deci, 2002). This was proven to be the case regardless of countries, cultures, traditional values and other individualist, egalitarian values (Deci and Ryan, 2008). In the same time, human beings are “living for” satisfying their needs, which sets the concept of basic needs as a criterium of what is important to life (Deci and Ryan, 2000). Additionally, satisfaction of the needs is an important nutrient for effective functioning and wellness, and it also promotes the optimal motivational traits and states of intrinsic motivation and intrinsic aspirations, which facilitate psychological health and effective engagement with the world (Deci and Ryan, 2015). As Deci and Ryan (2015, p.3) have explained, “the satisfaction versus thwarting of the basic psychological needs for autonomy, competence and relatedness explain the enhancement versus undermining of intrinsic motivation, the internalization of extrinsic motivation, and the development of general causality orientations”. In this regard, the need of autonomy and the need of competence are believed to have impact on intrinsic motivation. More specifically, if these two needs are satisfied, they can have a range of positive consequences. On the other hand, if they are thwarted then can have a range of negative consequences (Deci and Ryan, 2015). In the context of health, the social environment which facilitates the satisfaction of these needs will promote the internalization of protective and preventive health behaviours (Silva et al., 2008). This would lead to autonomous engagement and maintenance of health-related activities and satisfaction of these needs in the long term (Silva et al., 2008).

2.1.3 Goal Contents Theory

SDT also focuses on goals, which are defined as desirable outcomes that people value and expect to attain by engaging in particular behaviour (Deci and Ryan, 2015). The Goal Contents Theory (GCT) further divides these into two groups, namely intrinsic goals and extrinsic goals (Deci and Ryan, 2008).

2.1.3.1 Intrinsic Goals

The intrinsic goals are personal growth, relationships, social affiliation, and generativity, community contribution, social affiliation, health and fitness, and self-acceptance (Deci and Ryan, 2009; Deci and Ryan, 2015; Sebire, Standage and Vansteenkiste, 2009). In general, intrinsic goals are assumed to be inherently satisfying and fostering the fulfilment of the three basic needs (Ryan and Patrick, 2009).

2.1.3.2 Extrinsic Goals

Extrinsic goals are such as wealth, fame, image, attractiveness, financial success, appearance, popularity, power, and conformity (Deci and Ryan, 2009; Deci and Ryan, 2015; Sebire, Standage and Vansteenkiste, 2009). In general, extrinsic goals are assumed to not be inherently satisfying the basic needs, and consequently impede the motivation for physical activities (Ryan and Patrick, 2009).

Previous research has proven that if people value more extrinsic goals than intrinsic, then they tend to display poorer psychological health, and the other way around (Deci and Ryan, 2015). Additionally, the pursuit and attainment of intrinsic goals such as self-exploration, meaningful relationships and community contributions tent to provide satisfaction for the basic psychological needs, whereas the pursuit and attainment of the extrinsic goals of material possession, social recognition and attractive image are at best indirectly satisfying the basic needs and may even be antagonistic to them (Deci and Ryan, 2015). Additionally, satisfaction of intrinsic goals has proven to be associated with greater health, well-being and performance in contrast to extrinsic aspiration (Deci and Ryan, 2008). In regard to physical activities, intrinsic goals such as improved health or enhanced appearance are a common foundation for intrinsic motivation to exercise (Sebire, Standage and Vansteenkiste, 2009).

2.2 Application of SDT in Academic Research

This research has systematically reviewed academic literature applying SDT in the fields of physical activities and social media. This provided support for using SDT as a framework for understanding predictors of motivation for physical activity, which is also reflected below (Silva et al., 2010; Hagger and Chatzisarantis, 2008).

2.2.1 Intrinsic and Extrinsic Motivation for Physical Activities

Previous research has commonly proven that intrinsic motivation has positive association with physical activities (Lonsdale et al., 2013; Kwan et al., 2011), while extrinsic motivation and amotivation were found to have weak, negative associations with physical activities (Lonsdale et al., 2013). When people are physically active as a result of inherent behaviour, they feel happiness and subjective vitality (Ryan and Patrick, 2009). Consequently, this behaviour enhances psychological needs that contribute to the overall sense of wellness (Ryan and Patrick, 2009). Overall, in the physical activity domain, more intrinsic motivation has been linked to more positive attitudes towards physical activity and greater physical activity levels (Fortier et al., 2011).

Another study suggested that intrinsic motivation may not be the most important predictor of engagement in exercise because maintenance of regular exercise behaviour is unlikely to be supported by fun and enjoyment (Edmunds, Ntoumanis and Duda, 2006). Another important finding is that in order to participate in intensive exercise behaviors, which necessitate considerable physical and mental exertion and stamina, individuals must place some value on the exercise and recognize its importance in terms of health and well-being (Edmunds, Ntoumanis and Duda, 2006). In general, active engagement in physical activities is seen as dependent on the significance of the activity and its values and benefits on the individual. Another classification is that new participants reported health benefits as their reason for exercise adoption, while long-term participants reported enjoyment as their principal reason for continuing (Edmunds, Ntoumanis and Duda, 2006).

Another study found that both intrinsic and extrinsic motivation are important to both physical activity adoption and maintenance, and can be decomposed to reveal a complex substructure of factors that are related to exercise stage of motivational readiness and behavior over time (Buckworth et al., 2007).

2.2.2 Motivational Goals

A common theme emerging from the problem of lack of exercise concerns the role of people's reasons for exercising (Silva et al., 2008). Thus, perceived as dwindling motivation are extrinsic motives such as weight control and doctors recommendation. While intrinsic motives such as enjoyment and interest are perceived as enhancing the motivation (Silva et al., 2008). Therefore, in regard to maintaining long lasting results, previous research has concluded that an individual's motivational focus needs to shift from extrinsic to intrinsic regulations (Silva et al., 2008; Edmunds, Ntoumanis and Duda, 2006).

Intrinsic motivation is also associated with sustained over time health promoting behaviour as physical activities and exercise (Lonsdale et al., 2013). A study examining regular exercise, found that intrinsically motivated people tend to report more regular physical activities, including positive outcomes in exercise participation (Rodgers et al., 2010). There is a tendency that regular exerciser would endorse intrinsic motivation more strongly than extrinsic motivation (Rodgers et al., 2010; Kwan et al., 2011). Wilson and Rodgers (2004) found that support from friends and social agents appears to endorse more intrinsic motivation, which is further associated with intentions to continue exercising.

2.2.3 Exercise Variety

Another study found that variety in environment, social context and personal perceptions has influence over motivation for physical activities (Sylvester et al., 2018). Authors concluded that intrinsic motivation is positively associated with exercise variety, especially when the needs for autonomy, competence, and relatedness are satisfied. The study further discussed that based on SDT people are growth oriented and have an innate propensity to satisfy their basic needs. Thus, in the seek of diverse ways to satisfy people’s needs, variety has a stronger relationship with exercise behavior via intrinsic motivation because a change in activity diverts interest and attention in the short term, and offers potential to have one's psychological needs satisfied (Sylvester et al., 2018). When focusing on the relationships between SDT with frequency, intensity, and duration of exercise, it has been suggested that some people may persist at sport and exercise despite being extrinsically motivated (Duncan et al., 2010).

2.2.4 Social Media Engagement

Another more recent stream of research has focused attention to social media in the investigation of motivation to physical activities (Jih-Hsuan, 2016). It has identified that people are increasingly seeking health information and discussing health related topics on social media (Li and Wang, 2018). Even more interestingly, “more than 40 percent of consumers say that information found via social media affects the way they deal with their health” (Li and Wang, 2018, p. 1). Even though many people would have some trust issues associated with the validity of the information found on social media, there are a big amount of people who would share ideas and seek support related to health care. Factors such as information support, emotional support, and the satisfaction of one’s autonomy and relatedness needs play an important role in people’s health-information-seeking intentions on social media (Li and Wang, 2018).

Another study found that sharing physical activity content on Facebook can provide social support through friends’ encouraging comments and can allow comparison of one’s own results with others (Klenk, Reifegerste and Renatus, 2017). Despite of gender, social media usage for posting of physical activities can reflect higher motivation for health-oriented behaviour (Klenk, Reifegerste and Renatus, 2017). Some people also tend to use social media for competition and challenge.

Another study found significant relationship between need satisfaction for relatedness and autonomy, and intention to like, comment and share exercise posts on social media. Their findings “highlight the importance of elements of exercise posts, which in turn satisfy viewers’ need for relatedness and autonomy” (Kim and Park, 2016, p.128). Their study has proven implications to practise by recommending practitioners to design exercise activity posts in ways that highlights the viewer’s role and their feelings of being appreciated and connected (Kim and Park, 2016). This was seen as a possible way to boost patients’ motivation for healthier lifestyle.

Another study examined whether being part or not being part of online community can have an effect on people’s motivation to physical activities. The findings point at the tendency that participating in social media groups has positive effects for physical activity levels, especially when the other members are highly physically active. Thus, relatedness with other physically active people was found to increase motivation. Another interesting insight from the study is that the density of the online community has negative effect on people’s behavior for physical activities, as it is assumed that people can experience too much pressure leading to reduction in motivation (Groenewegen et al., 2012).

2.3 Conceptual Model

Based on the theoretical framework of SDT and previous research conducted in the field of physical activities and social media, this study developed a conceptual model (Figure 2). The model illustrates that Basic Needs and Goals, which according to the theory have mutual influence over each other, and further have an effect on the motivation. According to previous research, intrinsic motivation would be a strong predictor of physical activities, while extrinsic and amotivation would be weak or negative predictors of physical activities. Additionally, this study found it beneficial to include personality identification to the investigation of motivation, by including the Personality Trait model, which is external for SDT (Digman, 1990). As Deci and Ryan (2015) put it “SDT is a motivational theory of personality… that examines how social contexts and individual differences facilitate different types of motivation”. Thus this study, in the seek to account for predictors of motivation, saw a benefit in using the Personality Traits, which provide a general classification of human personality and individual differences (Digman, 1990). Additionally, it is believed that the model would provide deeper insights into the human personality of the actors, who are involved in the collected data (Iacobelli and Culotta, 2013). Personality traits is also believed to be reflected in humans' interactions and communications, and therefore the social media texts was believed to contain those (Digman, 1990). Personality detection from text has also been seen as useful for extracting behavior characteristics of authors who has written the text (Agarwal, 2014). The personality traits model has five-factors, namely Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism. A description of each is provided in Table 3 (on page 35).

The conceptual model in Figure 2 illustrates that the three Basic needs of autonomy, competence and relatedness and the intrinsic and extrinsic goal can be used as predictors of the three types of motivation. Additionally, the five personality traits of openness, conscientiousness, agreeableness, extraversion, and neuroticism are also used as predictors of motivation. The conceptual model was not only used to provide a coherent representation of the predictors of motivation, but it was also applied in the data analysis as a classification model. This is further discussed in the methodology section.

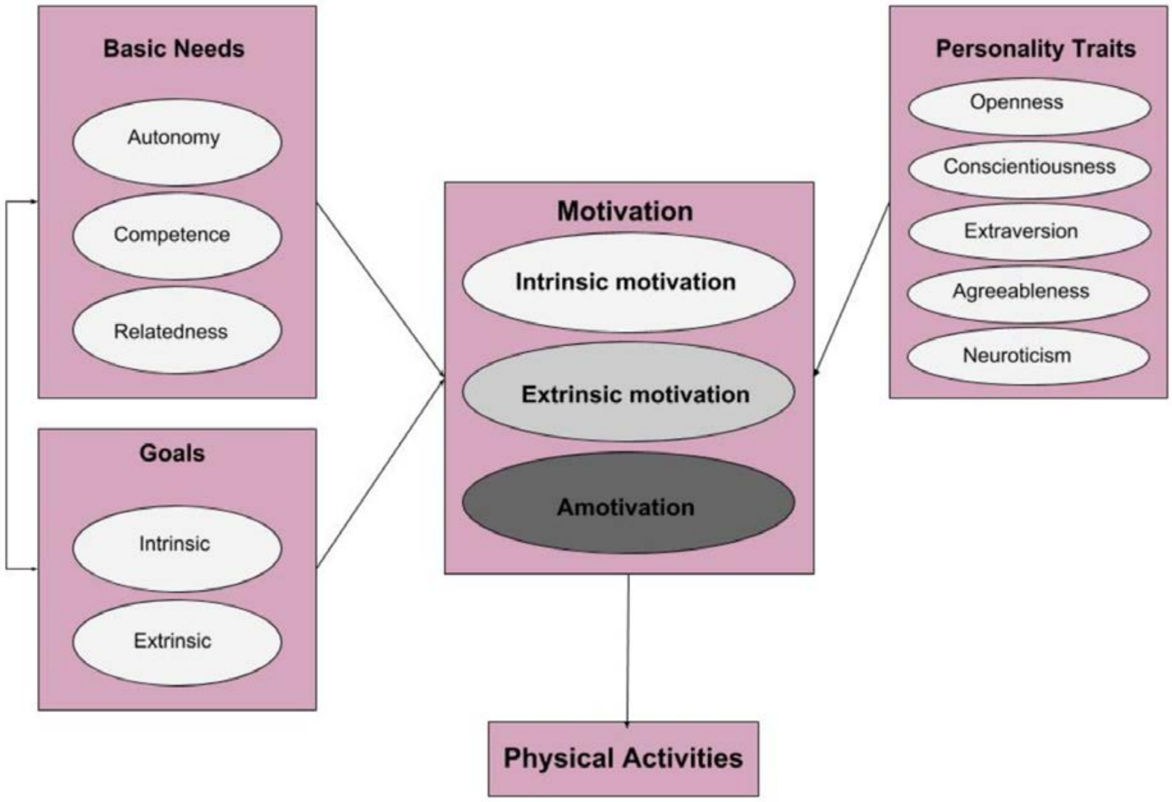


Figure 2. Conceptual Model

2.4 Implications of Motivation for Physical Activities

SDT has identified that the three Basic needs of autonomy, competence and relatedness and the intrinsic and extrinsic goal can be used as predictors of the three types of motivation. Additionally, think study included the five personality traits of openness, conscientiousness, agreeableness, extraversion, and neuroticism as predictors of motivation. Previous research examining predictors of motivation for physical activities has also applied other theories. Among those were Goal-centered Self-regulatory Theory used in the investigation of perceived goals (Lutz, Karoly and Okun, 2008), Self-Efficacy Theory used to examine quantity of motivation that people have for conducting tasks based on the degree to which they feel competent to do so (Silva et al., 2008). Theory of Planned Behaviour was another example. The theory aligns with the examination of quantity of motivation conceptualization and has been found to be consistently linked with physical activity behavior (Fortier et al., 2011). However, all these theories treat motivation as an unitary concept focused on the quantity of motivation people have for particular behaviour.

However, the current study is interested in examining quality of motivation as a measure of engagement for physical activities, and in this regard SDT is the theory that most precisely outlines that. Academic research also support this choice of theoretical foundation because they argue that SDT "is the only empirically-derived theory of motivation which posits that autonomy is essential for maintained behavior change and because there are validated psychometric instruments for each construct of the theory" (Silva et al., 2008). Also, SDT represents a viable framework of an array of initiation and adherence issues within exercise, and thus useful tool for investigating exercise motivation (Wilson, Mack and Grattan, 2008). Even though previous literature has investigated motivation for physical activities in various settings and contexts, it was highlighted that there is a need for further application of SDT, especially in the physical activities aspects (Silva et al., 2008). Therefore, this study employed SDT as the main theoretical framework for the investigation of motivation for physical activities.

Another important outcome from the conducted literature review points at a gap in regard of methodology. Despite of the numerous research examining motivation towards physical activities, none had actually used social data analytics as a possible methodology.
All reviewed studies used predefined questionnaires and surveys for their examination (Markland and Tobin, 2004; Godin and Shephard, 1985; Gillison et al., 2009; Sylvester et al., 2018; McDonough and Crocker, 2007). A prominent example was the Behavioral Regulation in Exercise Questionnaire (BREQ), which was used to measure motivation and assess the reasons why people exercise (Brickell and Chatzisarantis, 2007; Fortier et al., 2011; Rodgers et al., 2010; Sylvester et al., 2018; Sebire, Standage and Vansteenkiste, 2009; Edmunds,Ntoumanis and Duda, 2006; Duncan et al,. 2010). Even though SDT has been applied to many fields and has been useful and efficient in measuring motivation in the physical activities context, in their review, Hagger and Chatzisarantis (2008) mentioned that there is not a lot of variance in the use of measurement of SDT, explicitating that the Association Test (IAT) and BREQ are the mainly used ones. Therefore, in order to assess predictors of motivation, this study employed social data analytics. Before going to the methodological consideration of this research, the next section emphasis and justifies the need of applying social data analytics from both topic point of view and methodology point of view.

2.5 Impact of Big Social Data

Big social data refers to data produced through user-generated content often on social media platform. Social media is often referred to as community provider because it provides users with the possibilities to interact with people with similar interests, ideas and concerns and give them a sense of belonging (C., K., 2016). Social media is also attractive to users because they can broadcast their thoughts and opinions on the great variety of topics (C., K., 2016). Social media and its mechanism for social support, social engagement, social influence, social pressure and access to resources, are seen as an opportunity to promote physical activities (Groenewegen et al., 2012). On social media, the users are encouraged to generate and share content which includes text, images, videos and others.

Social media has also been defined as “scalable communications technologies that turn Internet-based communications into an interactive dialogue platform” (Vatrapu, 2013). Thus, in recent years, social media platforms have become popular not only for users, but also for worldwide academic literature (Jih-Hsuan, 2016).

2.5.1 Big Social Data

Social media provides large volumes of unstructured social text, which is also just termed “big data” (Vatrapu, 2013) or more specifically “big social data” (Vatrapu et al., 2016). Big data has no rigorous definition, even some scholars believe that “big data is … a poor term“ (Boyd and Crawford, 2012). Though our common understanding of it, big data is related to the volume of information that has grown so much that it has changed the traditional ways of handling data. As a result, one of its definition emphasis on the size, thus “big data can be defined based on large volumes of extensively varied data that are generated, captured, and processed at high velocity” (Günther et al., 2017). Other scholars, emphasis on the analytical part of it and define big data as “algorithm-based analyses of large-scale, disparate digital data for purposes of prediction, measurement and governance“ (Flyverbom and Madsen, 2015).

Big data is also commonly described by three dominant characteristics, which are known in the academic world as the Vs model (Lugmayr et al., 2017; Gandomi and Haider, 2015). Data velocity refers to the rate data is produced and the speed at which it is analyzed and utilized (Gandomi and Haider, 2015). Data volume refers to the size of the produced data. Data variety refers to the structural heterogeneity in a dataset, thus the use of structured, semi-structured and unstructured data (Gandomi and Haider, 2015).

From a more technical point of view, big data is enabled by the processes of datafication, which are taking “information about all things under the sun - including ones we never used to think of as information at all, such as a person's location, the vibrations of an engine, or the stress on a bridge - and transforming it into a data format to make it quantified” (Mayer-Schönberger and Cukier, 2013, p. 15). When quantified information is captured, then we can refer to it as data.

Thus, in the realm of big data there is an important distinction between data, information and knowledge. Data on its own is understood as the “raw material”, which is unprocessed. Information on its own is understood as organized data in a way that there is a logic. Knowledge is more than information, because it is understood as “data organized in a way that has an effect”, which makes us take an action (Flyverbom and Madsen, 2015). In this regard, data seen as raw material is only useful when it is stored, accessed and consequently processes and analysed into information, which is interpreted and utilized into actionable knowledge (Flyverbom and Madsen, 2015; Stubbs, 2014). The processing and analysing of data is called data mining or data analytics.

2.5.1.1 Attributes of Big Social Data

Data coming from social media platforms is often textual data and considered unstructured. With newer characteristics of the social media platforms, data is continuously becoming more heterogeneous, unstructured, sparse and high dimensional (Aggarwal, 2012). All these characteristics of social media data challenge the traditional data mining techniques (Stubbs, 2014). These have become inefficient and therefore text mining technique are preferred when working with social media data (Aggarwal, 2012).

As social media data is generated by user, the content reflects primarily on them, their communication and their interactions (Günther et al., 2017). In general, big social data is a result of electronic trace data, which is generated by two types of socio-technical interaction (Vatrapu et al., 2016).

The first one, social graph, represents the technology interactions in the form of what is being done by the users. These interaction are associated with the technology itself, and emerge from actions like using the Facebook's webpage or app in order to like, post, comment, share etc. These actions consists of users, their actions, their activities, and the artifacts created through these interactions (Mukkamala, Hussain and Vatrapu, 2014; Vatrapu et al., 2016; Hussain et al., 2014).

The second one, social text, represents the peers interactions in the form of what is being said through user's interactions with other users. This type consists of all communicative and linguistic aspects of social media conversations. It involves the topics, keywords, and sentiment communicated thought the content (Mukkamala, Hussain and Vatrapu, 2014; Vatrapu et al., 2016; Hussain et al., 2014). These two types of data can provide measures of motivation to physical activities through Facebook. This study is primarily interested in the social text, as these interactions involve the communicative part, where content regarding motivation can be communicated (Mukkamala, Hussain and Vatrapu, 2014).

2.5.2 Big Social Data Application in Academic Research

Big data is also defined as the "ability of society to harness information in novel ways to produce useful insights or goods and services of significant value" (Mayer-Schönberger and Cukier, 2013, p.3). The processing of user-generated content has helped many different industries including health care to extract valuable knowledge for prevention purposes (Mayer-Schönberger and Cukier, 2013, p.3). Big data used in research is believed to add social value because it can improve our social well-being in regard to education, health care, public services (Boyd and Crawford, 2012).

Therefore, the use of big data in academic research has been discussed widely in the last decade. The availability of large volumes of data, better warehousing, advanced computing power and software have made this concept widespread (Linoff and Berry, 2011). While big data and analytics have been rapidly accepted by the industry actors, this has left no time for the concept to mature in the academic world. Thus, there is a lack of coherent understanding and terminology of the concept, leading big data and analytics as a research method to be lacking behind in the academic studies (Gandomi and Haider, 2015).

In general, the use of big data and analytics in academic research can be seen as revolutionary, because it reforms the way we think about data and the way we use data

(Mayer-Schönberger and Cukier, 2013). Boyd and Crawford (2012) believe that by big data analytics we refer not only to a large datasets and the advanced techniques and methods used for analysing them, but also to a "computational turn in thought and research". Thus, the more advanced the tools and technique for analysing big data become, the better we will be at conducting research.

With this in mind, the current study will challenge our understanding of motivation by added a more quantitative measurement, which could better illustrate what is our understanding of predictors of motivation to physical activities. The next section explains in detail the methodological considerations undertaken for this study.

3. Methodology

This section reflects on the general methodological consideration, including research design, research strategy, data collection and preparation, and data modelling and analysis.

3.1 Research Design

The current study employed pragmatic philosophy and a cross-sectional design in order to examine motivation for physical activities leading to healthy lifestyle.

The purpose was to be able to discover how predictors of motivation for physical activities can be derived from social media text, and afterwards to explore these insights. Lastly, provide recommendations for the promotion of motivation for physical activities.

In order to find answers to the research questions, this study undertook an explorative design, which helped gaining understanding and insights, and consequently a comprehensive clarification of the topic of interest (Saunders, Lewis and Thornhill, 2009).

The research started at more abstract and broader level by conducting a systematic literature review of self-determination theory, which is believed to be the foundation of motivation. Afterwards, a more specific focus on physical activities and social media was adopted in order to set the scope.

CRISP-DM, which is a data analytics methodology, was adopted as a research strategy (Chapman et al., 2000). The unit of analysis was social media text generated by the users of social media platforms. The data sources for this research were five Facebook public pages and three online forums. The data was collected and prepared for analysis with the help of a tool called SODATO.

For data modeling and analysis, firstly a training set was randomly selected and manually coded. Afterwards it was passed to a tool called MUTATO, where five different supervised machine learning algorithms were used for text classification. A deductive quantitative approach, founded in the theory was utilized for this step. The outcomes were visualized with the help of Tableau and Excel, and represented in the results section.

The whole process is illustrated in Figure 3 and the next sections provide a detailed verification and justification of each of these steps, starting with the research strategy.

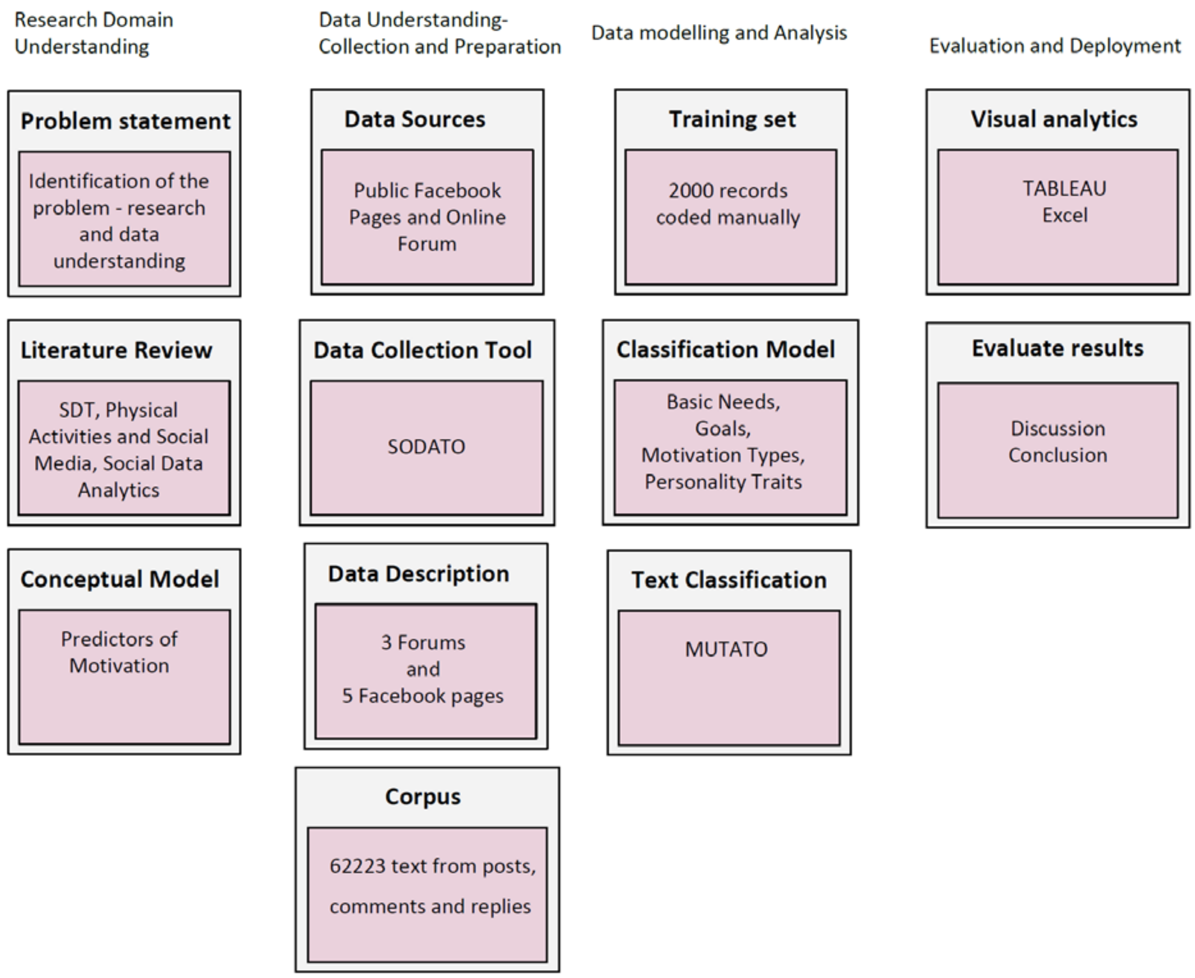


Figure 3. Research Design - reflecting CRISP-DM in this study's methodology

3.2 Research Strategy

Data analytics are based on mathematical algorithms. Thus, it is not a new term as it has been existing even before the invention of the computers (Linoff and Berry, 2011). Though with the development of technology, big data analytics has also advanced and new techniques has been founded. On its own the term data analytics refers to a "business process for exploring large amounts of data to discover meaningful patterns and rules" (Linoff and Berry, 2011, p.2).

Computational social science is one of the areas where big data analytics has its greatest impact because of the potential to replace traditional strategy as surveys, sampling studies and questionnaires (Mayer-Schönberger and Cukier, 2013). The need for new research strategies has also been explicitated in the literature review, which concluded that other strategies than questionnaire are highly needed for the study of motivation. Therefore, in order to justify the use of data analytics, this study followed an industry accepted data analytics methodology called CRISP-DM (Chapman et al., 2000).

CRISP-DM is usually applied by data mining experts, who use it in order to tackle industry specific problems. CRISP-DM stands for cross-industry standard process for data mining (Chapman et al., 2000). Fundamental for this strategy is the data mining life cycle illustrated in Table 1 as consisting of six steps. The use of these steps in this study is illustrated in Figure 3.

CRISP always starts with (1) identification of the problem, which involves both determining the research needs and the data analytics goals. In the case of this study, this initial step involved the formulation of problem statement, research questions and scope. From more technical point of view, this step also involved considerations about the data sources and tools which can be used further on in the analysis. This was necessary because the different data analytics techniques used for extracting valuable information from data depend on the data format. For example, this study used social media text which is most often analysed with text analytics techniques (Kannan et al., 2016). Text analytics employ different machine learning models, which are provided with special abilities for mining dynamic data containing poor and non-standard vocabulary (Kannan et al., 2016; Aggarwal, 2012). In this regard, text analytics were applied in order to discover knowledge about the predictor of motivation for physical activities from the collected data. Text analysis techniques are often text classification and topic modeling. Text classification is used for organising textual data into a set of categories or models (Mukkamala and Vatrapu, 2016).

Next step was (2) Data Collection, followed by (3) Data Preparation. Here, considerations about data understanding, selection, exploration and descriptions of the collected data took place before data was prepared for analysis. Next was (4) Data Modeling and Analysis, where manual text classification and supervised machine learning techniques were applied. Finally, the data mining cycle ends with (5) evaluation of the model and results from the data, and (6) considerations and application of the results in practise, which are reflected in the discussion and conclusion sections of this study. Data analytics is a iterate process, which involves going back and forth between the steps, which is also illustrated in the data life cycle, available in Appendix 9.1.

Big data analytics is believed to be a valid and reliable method for measuring motivation, because in order to understand what motivation is about, we need to be able to know what motivated people versus not-motivated people say or write. This information can be extracted unbiasedly from social media texts. Being able to recognize that language is central to the study of motivation is something this study aims at providing (Grimmer and Stewart, 2013).

Business Understanding	Data Understanding	Data Preparation	Modeling	Evaluation	Deployment
Determine Business Objectives *Background* *Business Objectives* *Business Success Criteria* **Assess Situation** *Inventory of Resources* *Requirements, Assumptions, and Constraints* *Risks and Contingencies* *Terminology* *Costs and Benefits* **Determine Data Mining Goals** *Data Mining Goals* *Data Mining Success Criteria* **Produce Project Plan** *Project Plan* *Initial Assessment of Tools and Techniques*	**Collect Initial Data** *Initial Data Collection Report* **Describe Data** *Data Description Report* **Explore Data** *Data Exploration Report* **Verify Data Quality** *Data Quality Report*	**Select Data** *Rationale for Inclusion/ Exclusion* **Clean Data** *Data Cleaning Report* **Construct Data** *Derived Attributes* *Generated Records* **Integrate Data** *Merged Data* **Format Data** *Reformatted Data* *Dataset* *Dataset Description*	**Select Modeling Techniques** *Modeling Technique* *Modeling Assumptions* **Generate Test Design** *Test Design* **Build Model** *Parameter Settings* *Models* *Model Descriptions* **Assess Model** *Model Assessment* *Revised Parameter Settings*	**Evaluate Results** *Assessment of Data Mining Results w.r.t. Business Success Criteria* *Approved Models* **Review Process** *Review of Process* **Determine Next Steps** *List of Possible Actions* *Decision*	**Plan Deployment** *Deployment Plan* **Plan Monitoring and Maintenance** *Monitoring and Maintenance Plan* **Produce Final Report** *Final Report* *Final Presentation* **Review Project** *Experience Documentation*

Table 1. Data Mining strategy CRISP-DM (Chapman et al., 2000)

3.3 Data Collection and Preparation

Data collection is a crucial step in the research process, where consideration about data sources are taken into account. Having access to the right data sources would ensure valid, reliable and more insightful results. The target for the data collection were forums and Facebook public pages where motivation for physical activities was widely discussed. Getting an overview and access to these sources was the starting point of the data collection process in this study. Collecting data from these data sources also provides another benefit, because the data is collected “passively while people do what they normally do anyway” and in this way, it is ensured that “the old biases associated with sampling and questionnaires disappear” (Mayer-Schönberger and Cukier, 2013).

In order to provide verification, the unit of analysis of this study was social text. All the text associated and collected for this investigation is called corpus (Grimmer and Stewart, 2013; Linoff and Berry, 2011). The corpus is comprised of .cvs files, called datasets. Each dataset contained fields and records, which are any string of characters, usually separated by semicolon (Linoff and Berry, 2011). The data was collected with the help of SODATO. The following parts provide a detailed explanation of the data collection tool, data sources and data preparation.

3.3.1 SODATO

SODATO is a Social Data Analytics Tools for data extraction, which was developed by the Computational Social Science Laboratory at CBS. The main purpose of the tool is to sustain fetching of data and preparing the data for analysis (Hussain and Vatrapu, 2014). SODATO is “designed, developed and evaluated to collect, store, analyze, and report big social data emanating from the social media” (Hussain and Vatrapu, 2014, p.4). Additional documentation about the tool is available in SODATO (2018).

With the help of SODATO data from three forums and five Facebook pages were fetched. SODATO also cleaned the data for missing values and duplicates. In order to ensure that the tools for modelling and classification can read the data, all data files were saved as UTF-8 with the help of the Sublime Text software, and were also manually cleaned for any additional spam texts.

3.3.2 Sample Description

In regard to what is feasible with time of this project and compliant with the general data privacy regulation (GDPR) in the EU, the study could get hold on data from three online forums and five public groups on Facebook. The eight in total data sources were all social media platforms. A full description of the data sources is available in Appendix 9.2, and a summary in Table 2. This research has only considered these online communication channels because data is produced at higher velocity and volume, and provides unbiased data.

Moreover, only Danish forums and pages, written in mainly Danish language, were included because the scope of this study was on local level of understanding.

The Facebook pages were selected on the basis on being public and users are allowed to post, comment or like. After that they were evaluated on the basis of popularity - how many followers there were and lastly whether they have been active on the long run.

The forums were selected with the help of Google search, questing for “motivation”, “sports”, “fitness”. Forums were also selected on the basis of being public and any one can read and write content.

The total corpus contains 55259 row of text in the form of posts, comments and replies. The corpus is represented in Table 2 below.

Data Description					
Data srouce: Facebook	**Description**	**Posts**	**Comments**	**Replies**	**Total:**
Abilica Online	Discusses weekly motivation and inspiration for training	441	32	7	480
Fitnessdk	Discussing fun and challenging facts about training	903	199	19	1121
MotionOnlinedk	Discussing healthy lifestyle, training, diet and sports.	172	28	0	200
Træning og kost	Discusssing training tips and good diet advice	302	21	0	323
Træning og Motivation	Discussing training, motivation, bodybuilding, fitness and sports	1145	283	6	1434
	Total:	2963	563	32	**3558**
Data srouce: Forums	**Description**		**Comments**	**Replies**	**Total:**
Dindebat.dk	Discusses lifestyle, health, training and motion		17621	3696	21317
Musclezone.dk	Discussing training, inspiration and diets		948	0	948
Motion-online.dk	Discussing healthy lifestyle, training, diet and sports.		23048	6388	29436
	Total:		41617	10084	**51701**
				Total number of observations across datasets:	**55259**

Table 2. Data description - including name of data source and description of the source

3.3.3 Data Preparation

A training set consisting of a total of 2000 randomly selected texts was comprise out of the corpus. This was used as a training set, which was manually coded, and used to support the supervised machine learning algorithm utilized for the data analysis. In order to provide consistency in the training set, an equal proportion of Facebook and Forum data was selected by assigning a random value to each record. Before collecting the training set, each dataset was given an unique identifier column called "ID", and all text fields were collected under one column called "Text". The data relationship diagram illustrates the process in Figure 4 below. The next section explains the actual coding of the training set.

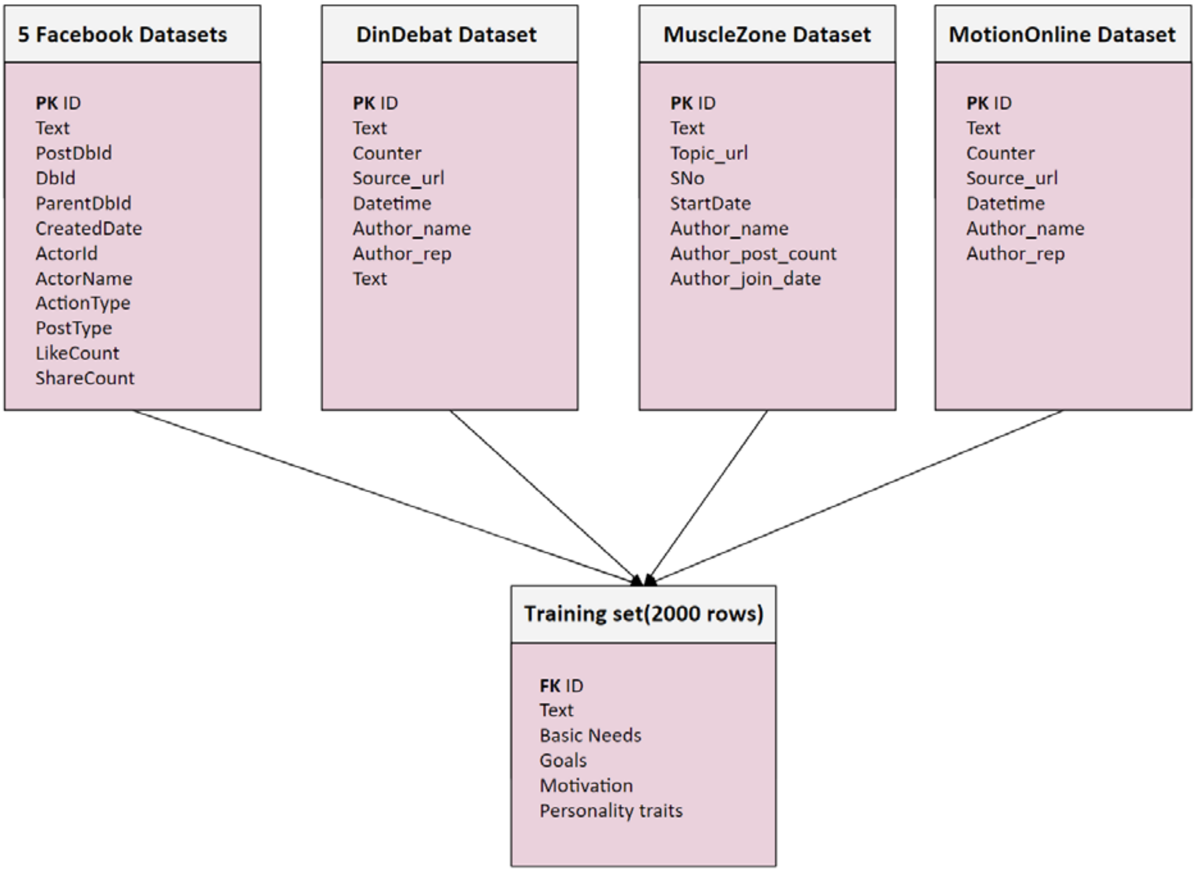


Figure 4. Data Relationship Diagram (PK - Private Key; FK - Foreign Key)

3.4 Data Modeling and Analysis

The data modelling and analysis of this study involved the incorporation of the conceptual model identified in the background section with the collected data. The process involved two steps, manual text classification and automated supervised machine learning classification (Mukkamala and Vatrapu, 2016).

3.4.1 Manual Text Classification

In the first step, the conceptual model was added to a list, where each individual model was assigned a set of categories and a comprehensive description for each category. This is shown on Table 3. The conceptual model is applied in the training set by manually classifying each record with a category from the models (see Figure 4). An excel sheet was created and the categories were added as lists in order to make sure that spelling and other errors are omitted (Mukkamala and Vatrapu, 2016).
The validity and reliability of the categorization process was ensured by developing a set of coding rules which were boned in the theory, and a comprehensive description of each of the categories (Morris, 1994).
While classifying the data, it became apparent that some texts are not corresponding to any of the category and therefore an “None” category was added to each of the models. In this case “None” was used for records which were among others, commercials, irrelevant in context, other discussions not concerning motivation to physical activities, blank, common facts, notifications, or simply insufficient to determined the model.

The manual categorization process was inspired from content analysis. Content analysis is a research technique applied to textual data for systematically inferences about intentions, attitudes, values, sentiment, understanding of individuals in a specific topic (Morris, 1994). The aim of content analysis is to systematically categorise the text into predefined categories. This allows for the systematic coding on large volumes of text corpora (Mukkamala and Vatrapu, 2016). During the process, memos and notes were written down, and used later in the discussion of topics. After the training set was categorized, supervised machine learning techniques were trained to classify the rest of the data.

Label	Definition
	Model 1: Basic needs
Autonomity	The need to act out of self-determination because of personal interest and integrated values.
Competence	The need to act out of feeling confidence, capability and affectance when undertaking the action.
Relatedness	The need to act because of having a sense of belongingness and connectedness both with other individuals and within the community.
	Model 2: Goals
Intrinsic	Personal growth, relationships, social affiliation, and generativity, community contribution, social affiliation, health and fitness, and self-acceptance.
Extrinsic	Wealth, fame, image, attractiveness, financial success, appearance, popularity, power, and conformity.
	Model 3: Types of motivation
Intrinsic motivation	People engage in activities which they find interesting, enjoyable and fun. People seek novelty, challenges, and wish to improve capacities, explore and learn. People to engage in physical activities because of the inherent pleasures, enjoyment and satisfactions they provide.
Extrinsic motivation	People engage in activities because of reinforcers like rewards, money, prizes, or because of the avoidance of undesirable situation, or in the wish for social approval.
Amotivation	People do not engage in activities because of lack of motivation, or lack of value and interest in the outcome. In physical activities, lack of competence, knowledgeable and skills are seen as increasing amotivation.
	Model 4: Personality trails
Openness	People with broad range of interests. People are adventurous, creative, open to challenges, think abstract.
Conscientiousness	People are thoughtful, goal-orientated, pay attential to details, structured.
Extraversion	People are extrovert and sociable, like to be center of attention
Agreeableness	People possess attributes as trust, kindness, affection, coorperativeness, care about others, enjoy contributing to the happiness of the others.
Neuroticism	People are sad, moody, and emotionally instable, worried about

Table 3. Classification model based on the conceptual model

3.4.2 Supervised Machine Learning Techniques

The second step of the data modelling and analysis included supervised machine learning techniques for building the predictive model. Multi-dimensional Text Analytics Tool (MUTATO) was used for performing the five different algorithms for text classification. The tools is developed by the Computational Social Science Laboratory at CBS. The overall architecture of MUTATO is shown in Appendix 9.3.

The general representation of how the supervised machine learning technique works is shown in Figure 5. The technique requires three inputs, namely a cleaned and prepared for analytics dataset, fixed set of categories, and a manually coded training set. The machine learning technique would train with the help of the training set and as an output, it will classify the dataset.

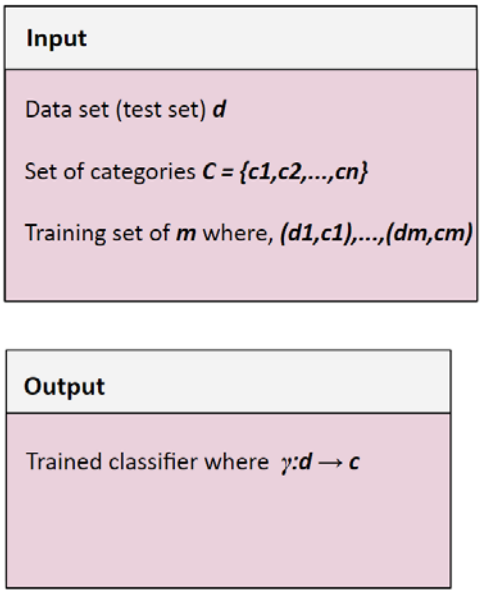


Figure 5. Supervised Machine Learning Method

As mentioned in the the research strategy, using data analytics for predictions is an iterative process. Thus, after running the algorithms for the first time, it became apparent that there is a need of merging some of the categories from the Basic Needs model, as illustrated in Table 5 below. This means that categories that contained a combination of needs were actually split. For example the Autonomy/Competence was split into two separate and the each of the observation containing this category was classified both as Autonomy and as Competence.

Additionally, the training set was over populated with categories of “None” for all four models. This left the prediction of the rest of the categories to be very low in accuracy. In order to improve the accuracy of the classifiers, there was a need for balancing the set. This involved oversampling (also known as stratified sampling). This was done in order to improve the overall performance of the classifiers. With oversampling a set which has an equal number of records from the different categories was generated. This reduces the size of the set but in the same time provides an equal representation of each category (Linoff and Berry, 2011).

Basic Needs
Autonomity
Competence
Relatedness
Autonomity/Competence
Autonomity/Relatedness
Competence/Relatedness
Autonomity/Competence/ Relatedness
None

Basic Needs
Autonomity
Competence
Relatedness
None

Table 5. Category amendment after first round of text classification

For this study, Naive Bayes Multinomial Classifier, Linear Support Vector Classifier, Logistic Regression Classifier, Passive Aggressive Classifier, and Support Vector Machine SGD Classifier were used. The classifiers were trained by 75% of the manually coded observation, and the rest 25% were used as a test set.

3.4.2.1 Naive Bayes Multinomial

Naive Bayes is a modelling technique used for predicting the probability of whether a data point is or is not classified as a particular category (Linoff and Berry, 2011). The technique is based on Bayes' Theorem of conditional probabilities, and work on the Bag of Words approach. Bag of Words (BOW) treats the dataset simply as a list of words by disregarding grammar and word order (Linoff and Berry, 2011). The list contains the tokens representing the single words and the frequency of how often the words appeared in the dataset. The equation below illustrates how the technique works, given the inputs shown in Figure 5. Here, P(d) is the probability of the dataset, and P(c) is the probabilities of the category, both independent of each other. P(c|d) is the conditional probability of the category *c* given that the dataset *d* occurred. Similarly, P(d|c) is the conditional probability of the dataset *d* given that the category *c* occurred (Linoff and Berry, 2011).

$$P(c \mid d) = \frac{P(d \mid c)P(c)}{P(d)}$$

The technique finds the conditional probability of each word given the category. This way it checks how often the category occurs given each word in the dataset. All this is done based on the training set (Aggarwal, 2012). This techniques was used because it is quite simple to apply, and in the same time it assumes that each category is of equal importance for the model.

3.4.2.2 Linear Support Vector Classifier

This classifier uses the characteristics of the text to identify which category it belongs to (Aggarwal, 2012). The Support Vector Machine (SVM) is a geometric method of separating two categories of the same model by finding the best hyperplane that puts one category above it and the other below (Linoff and Berry, 2011). A simple two-dimensional representation of SVM is shown in Figure 6. The hyperplane is a decision surface, or simple a line, that separated the two categories, and is illustrated with a dotted line in the middle. The green category are separated to one side of the line and go one way, while the red category are separated to the the other side, and go the other way. The boundary line are called supporting hyperplanes, and the data points on these boundaries are the support vectors. The support vectors have an effect on the support vector classifier, thus the same of the algorithm. The distance between one support hyperplane to other is called margin. The purpose of the SVM is to find the decision surface that maximizes the margin (Linoff and Berry, 2011). Though, this study, used textual data, which is characterized as very unstructured and heterogenous, leading the maximal margin hyperplane to be extremely sensitive to change in a single data point (James *et al.*, 2013). In order to get a greater robustness to individual observation and to classify better the most of the training data points, the classifiers used a Support Vector Classifier, which is also called soft margin classifier. The margin is defined as soft because it can be violated by some of the training data points (James *et al.*, 2013). Simply put, the classifier could miscategorise few training data points in order to better classify the remaining data points.

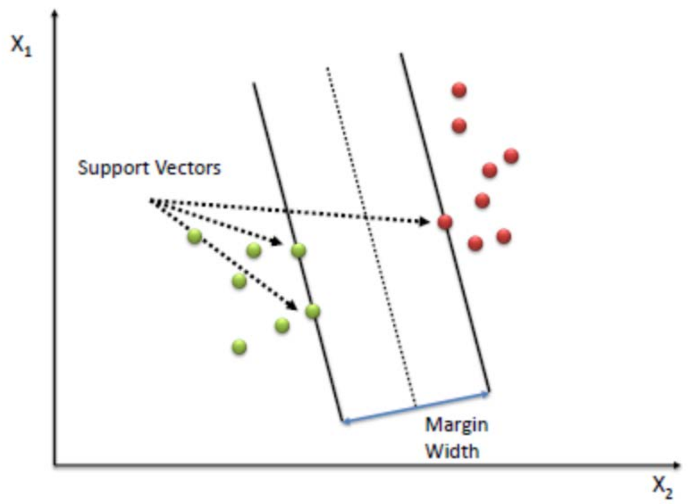


Figure 6. Support Vector Machine - a one-dimensional line separating points on a two-dimensional plane (Linoff and Berry, 2011).

3.4.2.3 Support Vector Machine (Stochastic Gradient Descent)

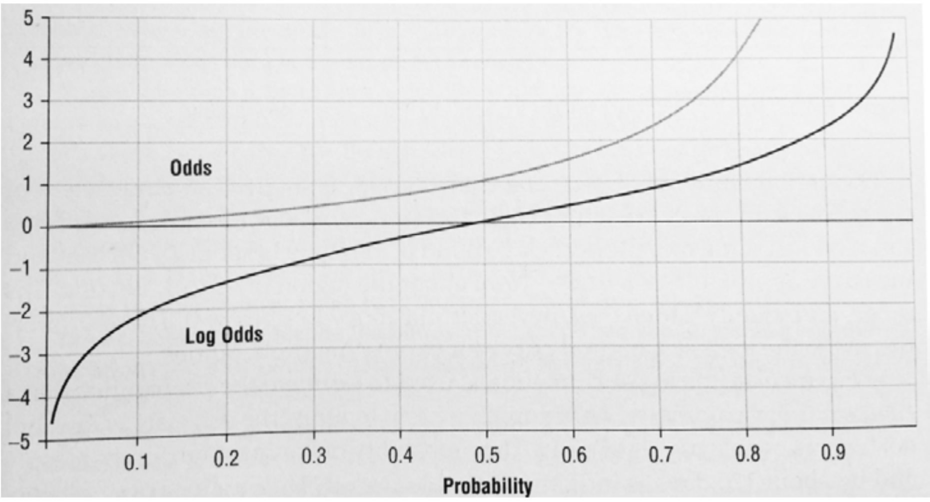
This classifiers has the same fundamentals as the Support Vector Machine introduced in the previous classifier. Though here, Stochastic Gradient Descent (SGD) was applied, in order to improve the efficiency, to better support the multi-class classification, and to provide more tuning opportunities (Scikit-learn 0.19.1 documentation. 2018). SGD combines the categories in a “one versus all” scheme. For each category, the classifier is trained to discriminate between that category and any other category.

3.4.2.5 Passive Aggressive Classifier

This classifier was introduced by Crammer at al.(2006) as a simple online algorithm. In a binary classification, the classifier predicts a yes/no (or +/-) outcome after each observation. The algorithm uses the same function as the Support Vector Machine above (Bonaccorso, G. 2018). Once it has made a prediction, it receives a feedback from the training set indicating the correct outcome. When the observation is classified on the correct side of the hyperplane, the algorithm is passive. But when a misclassification occurs, based on the feedback, the classifier may modify its prediction mechanism, in order to improve the accuracy of the prediction for the remaining observations (Crammer et al., 2006). This means that the algorithm becomes aggressive, because it is looking for a new margin which is as close as possible to the previous one. With time the algorithm slowly forgets the previous decision and learns from the new, that is why it improves its accuracy.

3.4.2.4 Logistic Regression Classifier

The Logistic Regression are build on the bases of linear regression algorithm. While the linear regression uses the equation for a straight line, which has no minima and maxima, the logistic regression bends the regression line into more appropriate shape. The logistic function takes the values between 0 and 1 (Linoff and Berry, 2011). This specific characteristic of logistic regression makes it very suited for assigning categories to records. Thus, it calculated the probability that a particular record belongs to one specific category. At first logistic regressions, transforms the probability *p* into odds, by taking the ratio of *p* over *p - 1* (Graph 1). Odds and probabilities are similar, though probabilities are restricted to the range from 0 to 1, while odds take values from 0 to infinity. By taking the *log* of the odds, the function goes from negative to positive infinity.



Graph 1. Logistic Function - comparison between odds and log odds. (Linoff and Berry, 2011)

3.4.3 Performance of the Models

Identifying the classifiers for the machine learning technique is an important milestone in this research because the choice is highly dependent on the type of data. It is also important to remember that the different documents would yell different results in regard to use of mining techniques (Linoff and Berry, 2011).

Analysing social data is still considered as a challenge because of it is often unstructured (Aggarwal, 2012). In order to ensure that the method is applied correctly, because of the complexity of natural language, a set of measures were used to assess the accuracy of each

of the classifiers (Grimmer and Stewart, 2013). In Figure 7, the confusion matrix represents the possible output from the classifier and the actual values from the training set.
Precision was used to measure the percentage of the observations which the classifier detected and are in fact true positives. In this case true positive refers to accuracy of detecting something that it fact is positive according to the training set.
Recall is used to measure the percentage of observations that are actually present in the input and were correctly identified by the classifier. In other words, recall is the number of true positive outcomes, divided by the number of positive results that should have been returned (Yedidia, 2016).
F1 score is a combined measure of precision and recall, which weights the harmonic means, or the weighted average of the precision and recall. The equation of F1 score is shown below. F1 score reaches its best value at 1 and its worst at 0 (Yedidia, 2016). A more detailed illustration of these measures is available in Appendix 9.4.

$$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

Accuracy is used to measure the percentage of observations that are actually present in the input and were correctly identified or correctly rejected by the classifier. An overall score of 0.80 or higher is considered of a high accuracy.

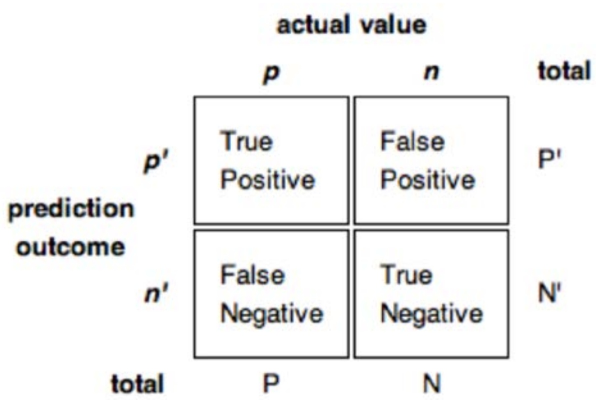


Figure 7. Confusion matrix

4. Results

The results from performed text classification were collected in log files, where the accuracy of each classifiers per model was reflected. Additionally, the log files contained the words with highest and lowest probability to identify the categories in the models. The information from the log files was collected in Excel tables and visually computed with the help of Excel and Tableau. The following section provides the results from the classifiers' performance based on models. Moreover, the results from the text classification are depicted with the help of different graphs, where meaningful insights are easy to discover.

After the manual coding of the data and the results from the classification, it was apparent that the None categories did not bring any meaningful and valuable insights to the investigation of motivation to physical activities. These categories were additionally assigned in order to cope with sorting out the text that was not representative for any of the meaningful categories. Thus, the result below do not take None into considerations.

4.1 Performance of the Overall Model

This section represents the results considering the performance of the overall model, followed by the results from each individual model and each individual category related to the model.

4.1.1 Overall Model Performance

In order to report on the performance of the classifiers, the represented in the Methodology (section 3.4.3) measures are used. For evaluating the models, it is important to look at the accuracy measure for each classifier. The accuracy of all the classifiers for every model ranged between 0.73 and 0.97. Based on the voted accuracy for each model, the Linear SVC and the Passive Aggressive classifiers performed the best with accuracy score higher than 0.83 for each model. The classifiers that got the voted accuracy for the model were used to classify the corpus. In Table 6, all classifiers for each model are presented with their evaluation measures. The classifiers performing best are highlighted with grey background color.

Text Classification Performance				
Model 1: Basic needs				
Classifier	Precision	Recall	F1-score	Accuracy
NaiveBayesMultinomialClassifier	0.80	0.78	0.79	0.78
LinearSVCClassifier	0.83	0.82	0.82	0.82
LogisticRegressionClassifier	0.76	0.75	0.74	0.75
PassiveAggressiveClassifier	0.84	0.83	0.83	0.83
SVMSGDClassifier	0.77	0.75	0.75	0.75
Model 2: Goals				
NaiveBayesMultinomialClassifier	0.94	0.94	0.94	0.94
LinearSVCClassifier	0.98	0.98	0.98	0.98
LogisticRegressionClassifier	0.86	0.85	0.84	0.85
PassiveAggressiveClassifier	0.98	0.98	0.98	0.98
SVMSGDClassifier	0.91	0.91	0.91	0.91
Model 3: Types of motivation				
NaiveBayesMultinomialClassifier	0.95	0.94	0.94	0.94
LinearSVCClassifier	0.97	0.97	0.97	0.97
LogisticRegressionClassifier	0.80	0.76	0.74	0.76
PassiveAggressiveClassifier	0.96	0.96	0.96	0.96
SVMSGDClassifier	0.88	0.88	0.87	0.88
Model 4: Personality traits				
NaiveBayesMultinomialClassifier	0.90	0.90	0.90	0.90
LinearSVCClassifier	0.95	0.95	0.95	0.95
LogisticRegressionClassifier	0.77	0.73	0.70	0.73
PassiveAggressiveClassifier	0.95	0.94	0.94	0.94
SVMSGDClassifier	0.85	0.85	0.84	0.85

Table 6. Total Model Performance (classifiers with highest accuracy score are in grey color).

4.1.2 Basic Needs Model Performance

For the Basic Needs model, the accuracy of the classifiers ranged between 0.75 and 0.83, where the Passive Aggressive Classifier has scored highest. When looking more detailed in the measurements, it becomes apparent that except from the None category, the rest were not that well classified (the table is available in Appendix 9.6). The F1- score was under 0.80 for Autonomy, Competence and Relatedness. This might imply that some classifiers were randomly assigning the three categories to the texts. Even the Passive Aggressive Classifier score ranges between 0.71 and 0.79. Even though the manual classification of the model was revised by merging the categories, which had in fact improved the overall performance of the classifiers, this was not enough to contribute to an higher acceptable accuracy score. This is also reflected in the overall representation of categories, where Autonomy, Competence and especially Relatedness were not representative in the population, representing only 15.91%, 10.50% and 6.11% from the total amount of observations from this model, respectively (see Table 7).

Additionally, when looking deeper into the work classification, it is apparent that there were many words that have a meaningful relationship to the category but there were also many words which added no value. Looking a level deeper in the classification by randomly selecting observations, which represents the category, are presented in Table 8. The table contain both examples from the texts categorized by the classifiers and examples from the manually categorized training sets. Already in the training set, it become apparent that the Basic Needs Model was harder to identify than the rest of the Models. This is mainly because the texts did not include direct reflection of the needs.

Category	Percentage	Observations
Model 1: Basic needs		
Autonomy	15.91%	8793
Competence	10.50%	5804
Relatedness	6.11%	3378
None	67.47%	37284
Total Observations	**100.00%**	**55259**

Table 7 Basic Needs Categories in Percentage

In Figure 8, the words with highest probability to identify the model are presented on the left side, followed by the words with lowest probability to identify the model, presented on the right side. The bubbles were created based on most descriptive words mainly from the best performing classifier but also as commonly identified by the rest of the classifiers in order to evaluate whether other words were better representatives. Words like “she”, “he”, “here”, “now”, “there”, “this”, “is”, etc. were excluded from the representation. While some of the words did not reflect any meaningful insights, which is because of the low F1 score, many of the words were valuable indicators of the categories. Size of bubbles reflect the probability score. The bigger the bubble the higher the probability that the word can classify the category.

Autonomy was best predicted with word like burning, training experience, dadbod, bulk, today, beginning, and others. On the other side of the spectrum lied January, bigger, muscles, course of action, fastest, restart, remember.

For competence the words best predicting the category were program, injury, leg press, breasts, chest, steroids, exercise, and others. The words with least probability were fun, day, preferably, should, and others that do not provide any special implication.

Relatedness was best predicted by stretching, teams plan, atmosphere, holiday, chest, cool, closer, the blog. The words with least probability for relatedness were Monday, excellent, fit, training, motion, New Years resolution, problem, shoes.

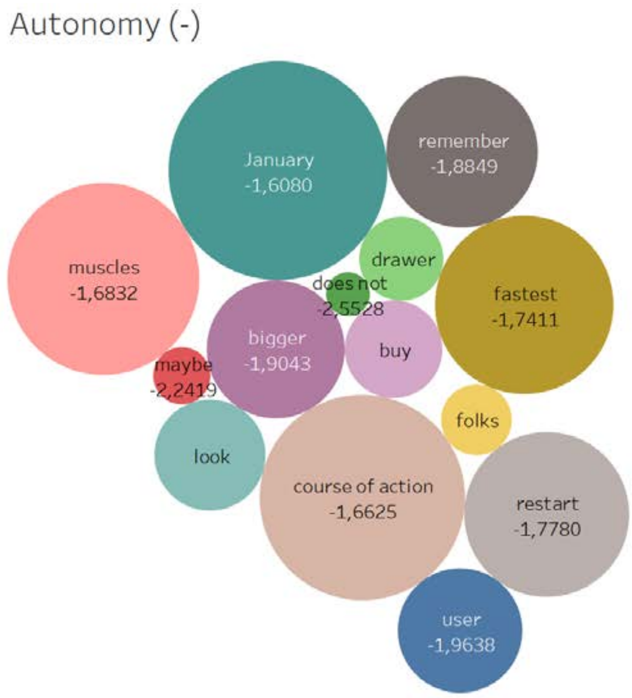


Figure 8. Basic Needs Model sorted by most probable vs. least probable features for each category.

Category	Example from classifier	English translation
	Model 1: Basic needs	
Category	**Example from classifier**	**English translation**
Autonomy	Ugens tip: Løb en kort tur men gør det ofte. Vi er nok mange der synes at hvis vi skal ud så skal det være mindst 7 km og hurtigt (og så bliver det ikke i dag) men 3 km roligt er meget bedre end 0 km hjemme i sofaen. (MotionOnline - FB)	Weeks tip: Run a short tour but do it often. We are sure many who think that if we go out, so it should be at least 7 km and fast (and this becomes not today) but 3 km calm is much better than 0 km home in the sofa.
Competence	Jeg har prøvet Stram op, hvilket jeg synes var hårdt...Jeg har også været til StepFit intro et på gange. Det er lidt svært i starten, da man også skal huske noget koreografi, men jeg synes faktisk det var sjovt nok og godt alternativ til løbetræningen en gang imellem. (Din Debat)	I have tried Stram op, which I think was hard...I have also been to StepFit intro a few times. It is a bit hard in the beggining as one should also remember some choreophraphy, but I think actually it was fun enough and a good alternative to the running training once in a while.
Relatedness	Det er aldrig for sent og altid rart med en partner! (Træning og Motivation)	It is never too late and always nice with a partner!
Category	**Example from training set**	**English translation**
Autonomy	Det der driver mig, er nok mest at jeg sætter mig nogle mål. Fx i denne uge vil jeg nå op på denne vægtklasse osv.	What drives me is mainly to put some goals for myself. For example, this week I will maage on to this weight class, etc.
Competence	Og det kan anbefales! Super træning - hårdt, hårdt og tjah.. hmm.. hårdt Det er rigtig fed træning.	And it can be recommended! Super training - hard, hard, and tjah...hmm..hard...This is really cool training.
Relatedness	Kom og vær med til en festlig og svedig weekend med de nye koreografier og den nye musik.	Can and participate in a festive and sweaty weekend with the new chorography and the new music.

Table 8. Random Examples of Basic Needs from both Training set and Corpus

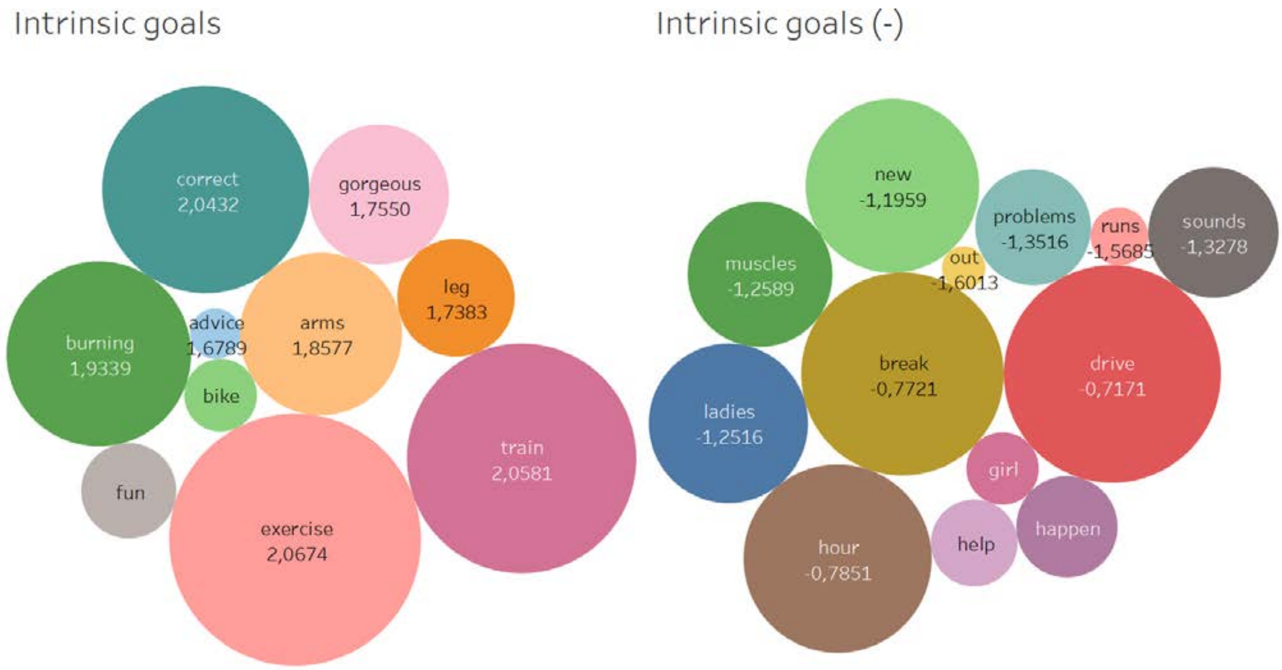
4.1.3 Goals Model Performance

For the Goals model, the accuracy of the classifiers ranged between 0.85 and 0.98, where the Passive Aggressive Classifier and the Linear SVC Classifier have scored the same highest score. All classifiers have performed good in detecting all three categories except from the Extrinsic category detected by Logistic Regression classifier (Appendix 9.6). The F1- score was over 0.81 for all other classifiers and categories. The None category was again the most dominant in the model, leaving Intrinsic and Extrinsic goals with 10.25% and only 5.27% representation respectively (Table 9).

Model 2: Goals		
Intrinsic	10.25%	5665
Extrinsic	5.27%	2912
None	84.48%	46682
Total Observations	**100.00%**	**55259**

Table 9. Goals Categories in Percentages

Figure 9 represents the probability words for this model, which were chosen based on all classifiers, but the main focus was on the classifiers with highest accuracy.
For the intrinsic goals, the words that best predict the category were exercise, correct, train, burning, again, fun, arms, gorgeous, advice, and others. The words with least probability were break, hours, drive, muscles, sounds, ladies, new, problems, etc. Extrinsic goals were best predicted by the words running injuries, training, weight loss, training program, goal, hiit (high-intensity interval training), kg, and other. On the opposite side were the words which lowest reflection on the category. These included recommend, high, km, knee, space, run, easily and others. In order to reflect these observations, examples of intrinsic and extrinsic goals were randomly selected from both the datasets and the training set and shown in Table 10.



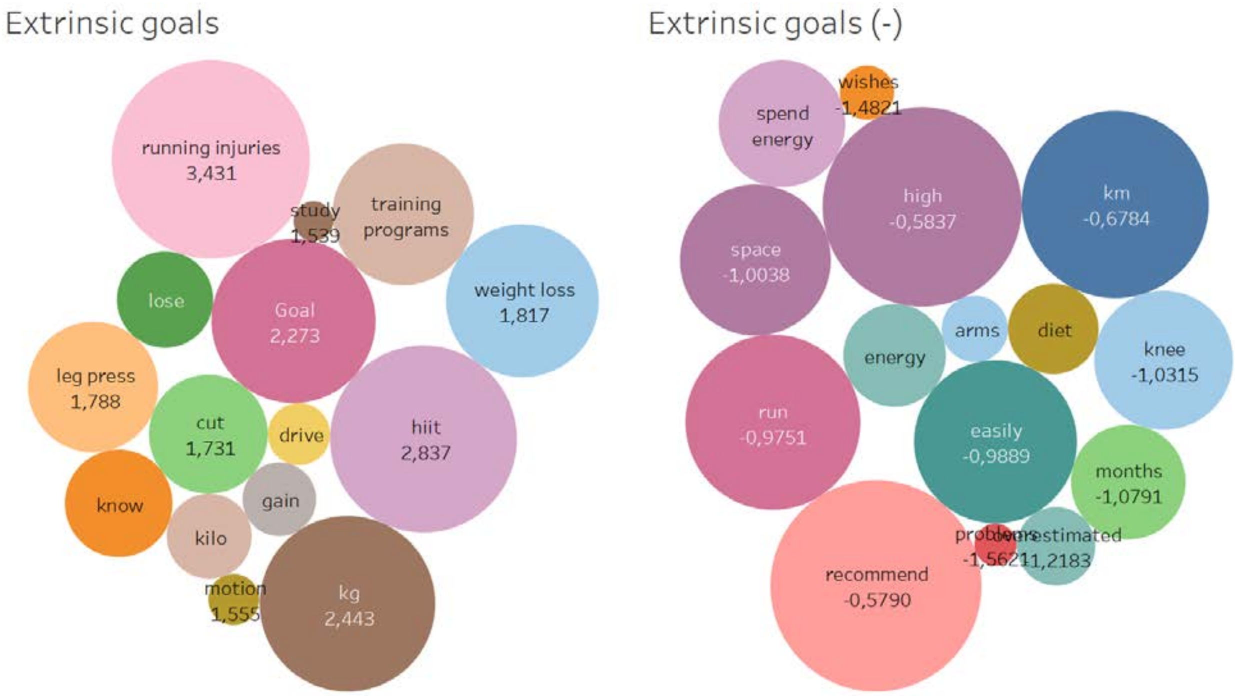


Figure 9. Goals Model sorted by most probable vs. least probable features for each category

Model 2: Goals		
Category	**Example from classifier**	**English translation**
Intrinsic	og siden jeg begyndte at bygge min form op med en realistisk progression og tålmodighed, så har jeg ikke haft skader i læggene.Det tager ca 7 år og lave en løbekrop og byg basis op før end du begynder at løbe i et tempo der kan udfordre understellets fibre for kraftigt. (Motion Online)	and since jeg began to build my form up with a realistic progression and patience, so I haven't been injured. It takes about 7 years to make a racer body and build base before you start to be running at a pace that can challenge undercarriage fibers too strong.
Extrinsic	Er sikker på du ville tabe dig drastisk ...Alle de der forskellige kemikalier er ikke nødvendige...Jeg har også engang brugt clean og diverse... Men prøv og hør... Der bare ikke noget der virker lige så godt som en diet...Det bedst sætte sig ned med sin bærbar på skødet og så ellers give en helt ny omlægning af sin diet en chance. (Muscle Zone)	You sure you want to lose weight drastically...All these different chemicals are not necessary...Jeg have once tried clean and diverse...But try to listen...It is not something that works as good as a diet. The best is to sit down with you laptop and so otherwise give a completely new conversion of their diet a chance ..
Category	**Example from training set**	**English translation**
Intrinsic	Ingen billeder før formen er blevet lidt bedre, syntes der er lang vej stadig Men de skal nok komme	No pictures before the form has become a bit better, I think it is a long way. But it will come.
Extrinsic	er begyndt at motionere og har hørt at cardiomaskiner skulle være suer gode når man vil tabe sig og strammes op (det typiske)- men nu spørger jeg nok lidt dumt, hvad er en cardiomaskine???er de virkelig så effektive når man vil tabe sig??	Has started to motionate and have heard that cardio machines should be super good when you want to lose weight and tighten up (the typical)- but now i might ask little stupid, what is a cardio machine? Is it really so effective when you want to lose weight?

Table 10. Random Examples of Goals from both Training set and Corpus

4.1.4 Personality Traits Model Performance

For the Personality Traits model, the accuracy of the classifiers ranged between 0.73 and 0.94, where again the Passive Aggressive Classifier and the Linear SVC Classifier have scored the almost the same highest score, 0.96 and 0.97 respectively. All classifiers have performed good, except from the Logistic Regression Classifier, where the accuracy was the lowest. In regard to predicting the categories, only Passive Aggressive Classifier and the Linear SVC Classifier were able to predict all categories. Openness has scored very low in regard to the other categories for the Naive Bayes, Logistic Regression, and SVMSGD classifiers. This model was also considered hard to predict because of the multiple categories. All classifiers and categories are presented in Appendix 9.7.

In regard to the distribution of the categories, shown in Table 11, it is visible that None is dominating also this model, with 77.96%. Openness and Agreeableness were coming after in the rank, but still taking a small part in the model, reflecting on 8.33% and 9.94%, respectively. Conscientiousness represented the model with 3.30%. While Extraversion and Neuroticism were representing less than 1% each, with total number of observation of 119 and 137, respectively. Examples from the classified dataset and the training set are presented in Table 12.

Model 4: Personality Traits		
Openness	8.33%	4603
Conscientiousness	3.30%	1825
Extraversion	0.22%	119
Agreeableness	9.94%	5494
Neuroticism	0.25%	137
None	77.96%	43081
Total Observations	**100.00%**	**55259**

Table 11. Personality Traits Categories in Percentage

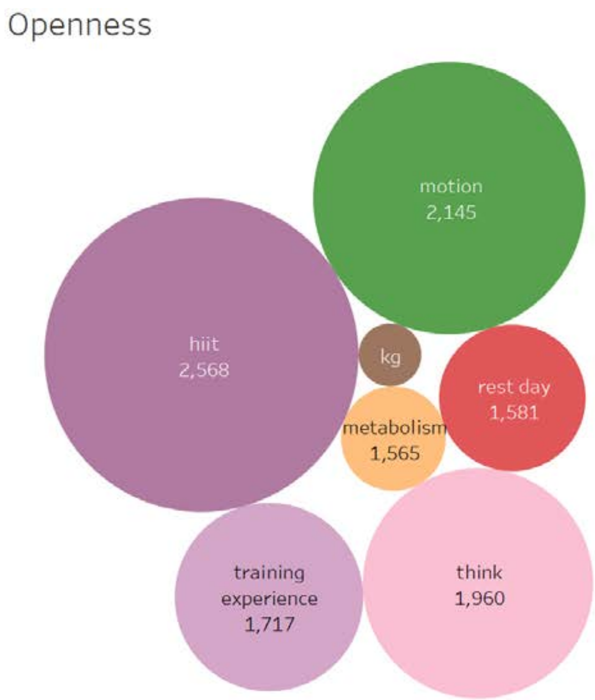
In Figure 10 it is represented that Openness was best predicted with words like hiit (high-intensity interval training), motion, think, training experience, rest day and others. On the other end, Openness was least predicted with nearest/almost, ride a bike, needle, extends, good, headache, folks, and others.

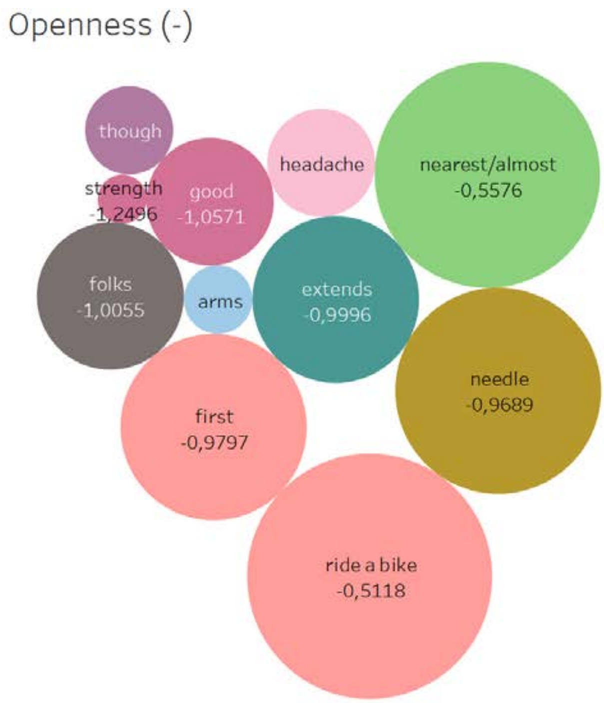
Conscientiousness was predicted with words like running injuries, train, dadbod, will, the lust, and others. While the words with least probability were pace, lose, shoe, ladies, little, weeks, test, gyno, muscles, and others.

Extraversion was best predicted with the words exercise, ladies, iceman, leg press, stretching, know, steroids, transformation, humor and others. On the other end were the words remember, large, better, beast, can, thanks and others.

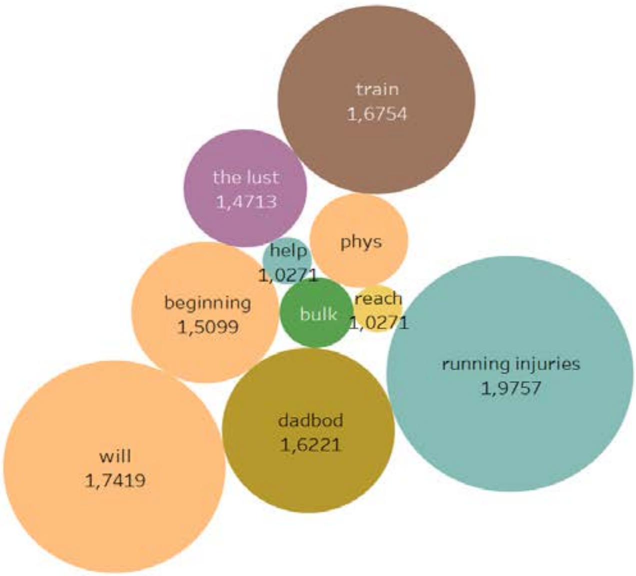
Agreeableness was best predicted with words like personally, advice, drive, bad, fun, willpower, beginners, and other. On the other end, the words with least probability to predict the category were times, searching eat, map/short, the rest, strongly, cool, healthy and others.

Neuroticism was mostly predicted by willpower, comes, possible, problems,bad, conscience, after, operation, share and others. On the other end were words like wait, week, beter, knee, hard, obese/cool, up, sleep and others.

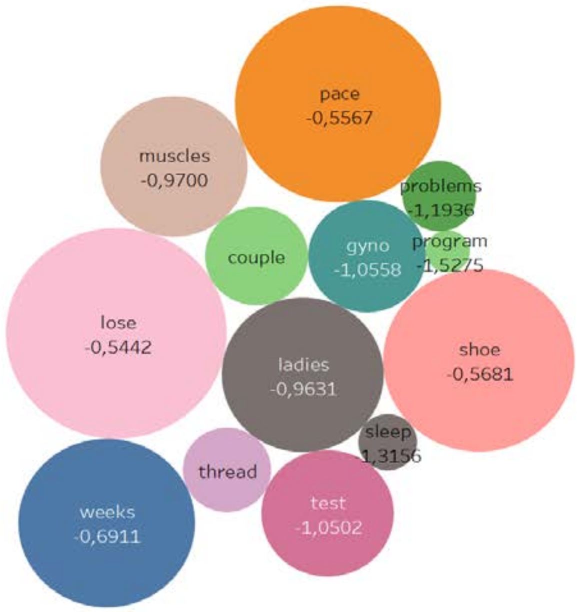




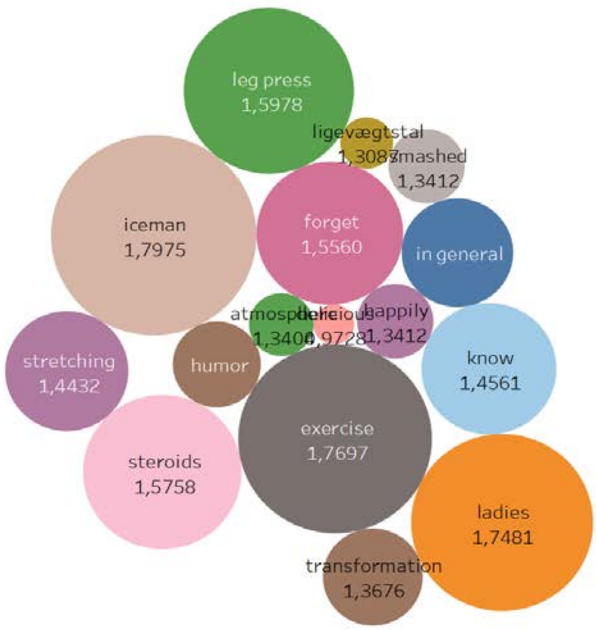
Conscientiousness



Conscientiousness (-)



Extraversion



Extraversion (-)

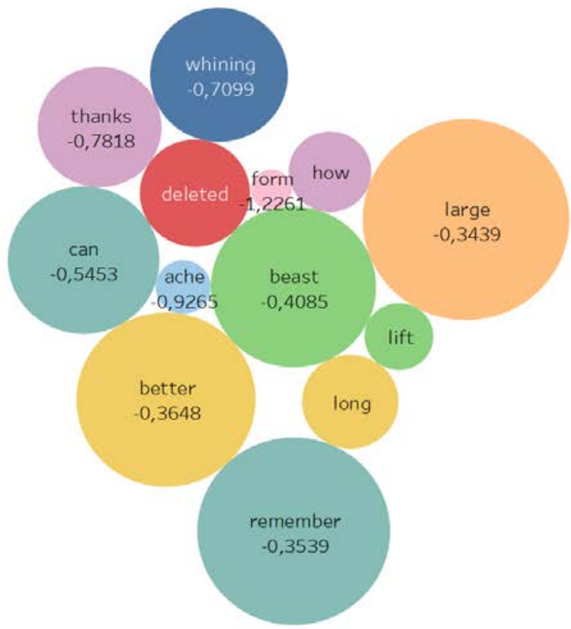


Figure 10. Goals Model sorted by most probable vs. least probable features for each category

Model 4: Personality Traits		
Category	**Example from classifier**	**English translation**
Openness	Jamen vis andre kan så kan jeg squ da også...! De første 27 kg er da ihvertfald røget... (Træning og Motion)	Well, if others can, so can I then too ...! The first 27 kg is at least smoked ...
Conscientiousness	Kan du huske Louise? Hendes mål var da hun startede med 100 dages træning sidste sommer at hun ville stå skarpt i brudekjolen på sin bryllupsdag. I lørdags blev hun så gift og hun har i dén grad holdt formen. (Fitness DK)	Do you remember Louise? Her goal was, when she started with 100 days of training last summer, that she would be sharp in the wedding dress on her wedding day. On Saturday she became married and she has kept the shape like that :)
Extraversion	lige for en lille update - så var jeg til træning idag , og nej hvor er jeg smadret.. Havde lykkeligt glemt den følelse. Hehe (Muscle Zone)	Just for a small update - so I was at training today, and oh no how I am exhausted...Had happily forgotten the feeling. Hehe
Agreeableness	Mange stiller så høje krav til sig selv og når de ikke kan overholde dem går de fluks i den anden grøft "når jeg alligevel har spist både konfekt og flæskesvær kan jeg jo ligeså godt fortsætte. Giv dig selv lidt line! Hvis du har spist mere end planlagt – tilgiv dig selv og tag udgangspunkt i hvor du er lige nu. Lige nu og her er er et fint tidspunkt at starte på en frisk. (Abilica)	Many people put so high demands on themselves and when they can not meet them, they fly in the other ditch "when I have eaten both chocolates and pork fat, I can as well continue". Give yourself a little line! If you have eaten more than planned - forgive yourself and start from where you are right now. Right now and here is a great time to start a fresh one.
Neuroticism	Jeg har anskaffet mig en mini stepmaskine, men når jeg bruger den får jeg ondt i mit ene knæ og derefter brænder det i knæet længe efter...Ødelægger stepmaskinen mit knæ eller er det noget der skal trænes væk?Har tidligere haft problemer med begge mine knæ... (Motion Online)	I have bought a mini-step machine, but when I use it, I feel pain in my one knee and then it burn in the knee long after ... Does the step machine destroy my knee or is it something to be trained? Previously had problems with both my knees ...
Category	**Example from training set**	**English translation**
Openness	Prøv med et højere split, og de øvelser du altid skal kæmpe dig igennem med en sur mine, skift dem ud. Det hjælper!	Try with a higer split, and the exercises you always fight youselft through with an angry face, change them. It helps!
Conscientiousness	Jeg træner altid kun ved høj feber lader jeg værre, jeg trapper godt nok ned, men det minimerer altid formtabet, har jeg erfaring for både i styrkeløft og ultraløb.	I always only train at high fever til I get worse. But this always minimize losing the shape, I have experience for both racing and ultrasound.
Extraversion	I skal lige alle sammen huske at der er ingen der har sagt dette var den perfekte slankekur . Nanna er ny på siden og har delt sit vægttab. people really would like to know how this could be done so quickly	You should all remember that nobody said that this is the perfect diet. Nanna is new on the page and had shared her weight lost. If you want to know more about how
Agreeableness	du kan sagtens forbrænde meget mere end 200 kcal. bare hold dig til en motionsform som ikke er alt for belastende (fx motionscykel). Du skal endelig ikke gå ud og løbe 5 km uden at træne dig langsomt op.	You can definitely burn much more than 200kcal. Jusl hold yourself to a form of motion that is not too stressful (like motion biking). You should eventually not go out and run 5 km without training yourself slowly up.
Neuroticism	...tænker da det bliver frustrerende at skulle skifte mellem centrene :pHar været en tur rundt ved dørene i de forskellige centre og tjekke efter skiltet, og har desværre ikke fundet nogen	...This might be frustrating to change between centres :p Have been on a tour around the doors in the different centers and checked for signs, and have unfortunately not found any

Table 12. Random Examples of Personality Traits from both Training set and Corpus

4.1.4 Types of Motivation Model Performance

For the Types of Motivation model, the accuracy of the classifiers ranged between 0.76 and 0.97, where again the Passive Aggressive Classifier and the Linear SVC Classifier have scored the almost the same highest score of 0.94 and 0.95 respectively. All classifiers had performed good, except from the Logistic Regression Classifier, where the accuracy was only 0.76. In regard to predicting the categories, Amotivation has the highest F1-score of 1.00. The F1- score was also high was Intrinsic and Extrinsic motivation, 0.92 and 0.98 respectively (Appendix 9.8). The distribution of the categories in the types of motivation model (see Table 13), showed that the None category was again the most dominant, representing 78.51%. Intrinsic motivation was categorized as 18.39% from the total observation. While Extrinsic motivation and Amotivation were the minority, representing 3.03% and 0.07%, which corresponds to 1675 and only 38 observation in total, respectively. Example from these categories from both the dataset and the training set are presented in Table 14.

Model 3: Types of motivation		
Intrinsic	18.39%	10160
Extrinsic	3.03%	1675
Amotivation	0.07%	38
None	78.51%	43386
Total Observations	**100.00%**	**55259**

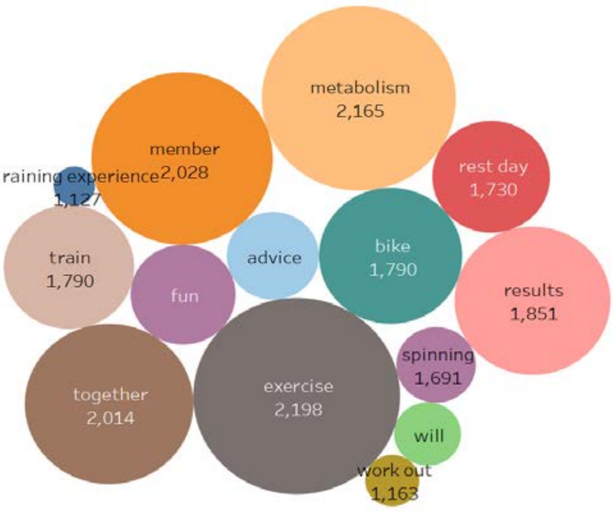
Table 13. Types of Motivation Categories in Percentage

Taking a step further in the category classification, represented in Figure 11 are the words for highest and lowest probability to predict the category. In regard to intrinsic motivation, the words were metabolism, exercise, member, together, results, rest day, train, fun, work out, advice, will, bike, training experience, and others. On the other end, were the word like weeks, find, bigger, muscles, share, around, remember, which were least probable in predicting the category.

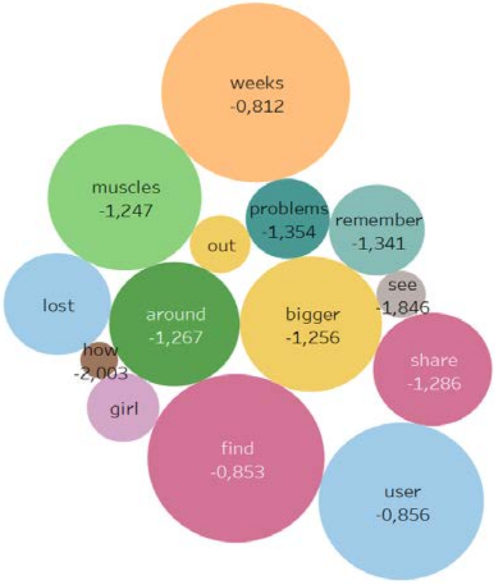
Extrinsic motivation was best predicted by words like training programs, kg, running injuries, breasts, hiit, motion, study, for free, leg press and others. The words predicting least the category were km, rest day, right, direct, correct, diet, knee, thanks, and others.

Amotivation, was characterised with words like will power, heart, weeks, problems, beaten, points, drop, bad, conscience, the doctor, operation and others. The words least predicting the category were hard, self, today, correct, day, great, gyno, better, out, back and others.

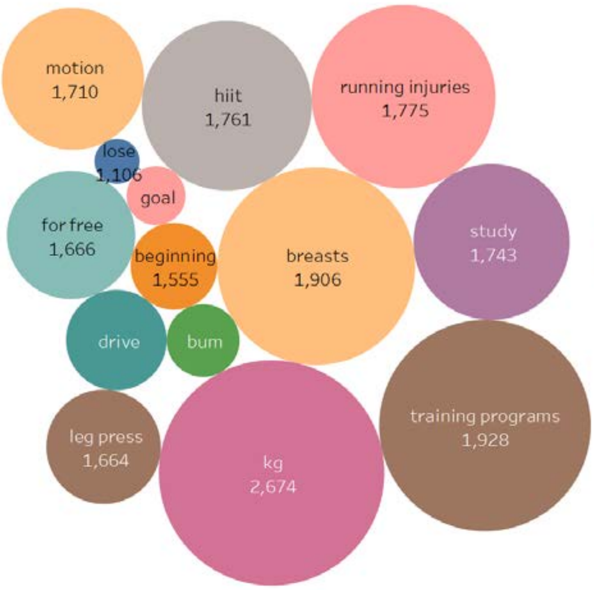
Intrinsic Motivation



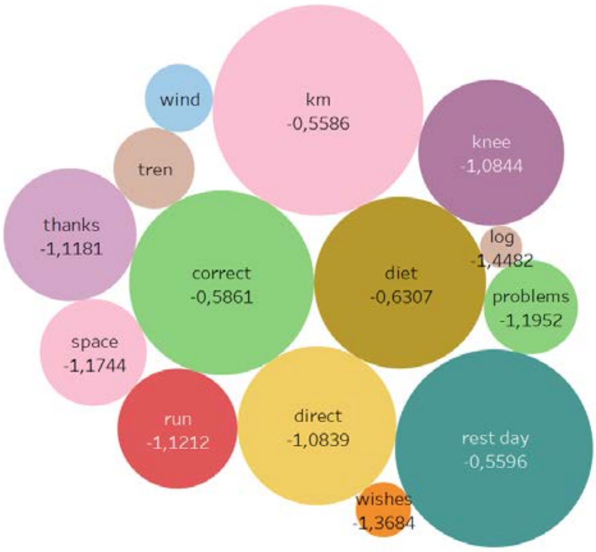
Intrinsic Motivation (-)



Extrinsic Motivation



Extrinsic Motivation (-)



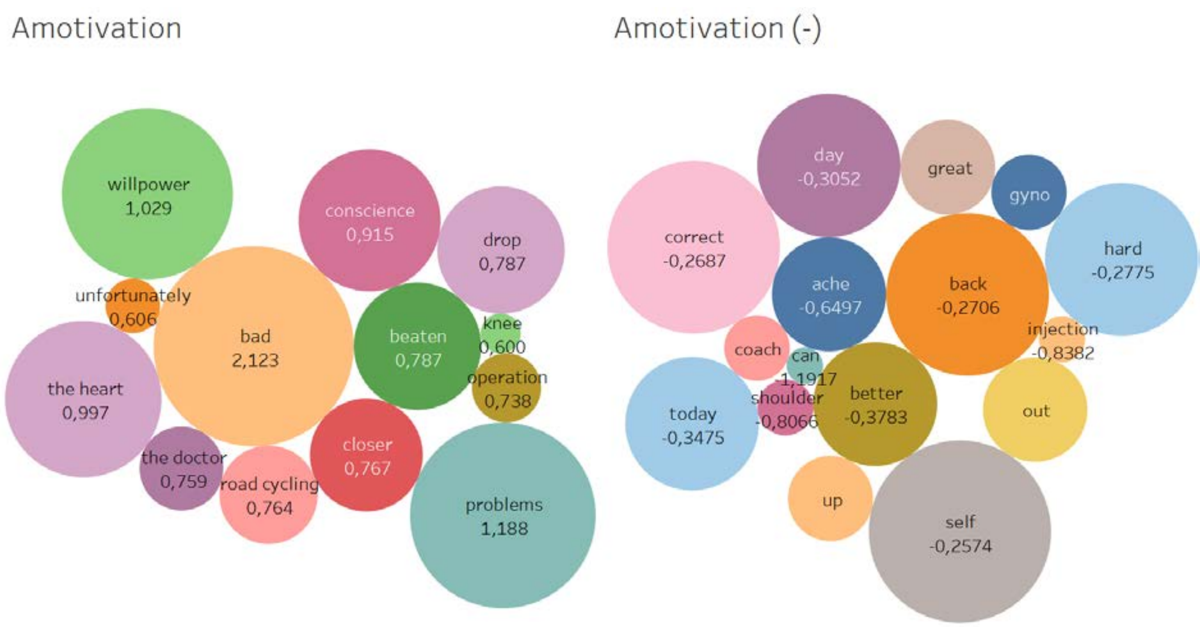


Figure 11. Goals Model sorted by most probable vs. least probable features for each category

Model 3: Types of motivation		
Category	**Example from classifier**	**English translation**
Intrinsic	Selvom man er skadet kan man næsten altid lave en form for cardio eller træne en anden kropsdel så man ikke er helt udlukket fra træning. (Træning og Kost)	Even thoug you are injured, you can almost always do some form of cardio or train another body part. So you are not completely excluded from exercise
Extrinsic	Hvis du med en vægt på 77 kg skal kunne nå det mål, vil du have en maximal iltoptagelse på 5.37L/Min. Det er 0.1L/Min mere end Haile Gebrsellasie, så for at istemme med Rune: Du når det ikke medmindre du taber dig.	If you with weight of 77kg should manage the goal, you need to have a maximal oxigen intake at 5.37L/Min. It is 0.1L/Min more than Haile Gebrsellasie, so to settle with Rune: You can not do that unless you lose weight.
Amotivation	Blandet med en smule dårlig samvittighed, jeg kan simpelthen ikke forestille mig at træne 5-6 dage om ugen. (Din Debat)	Mixed with a bit of bad conscience, I simply can not imagine myself training 5-6 days a week
Category	**Example from training set**	**English translation**
Intrinsic	Efter min mening skal træning være krævende, hårdt og frem for alt så motiverende og eller sjovt at man har LYSTEN til det. Prøv nye ting, og fokuser på dit/dine mål, ikke kun vejen frem til dem.	In my opinion, training must be demanding, hard and above all so motivating and or fun that you have the joy for it. Try new things and focus on your goals, not just the way forward to them.
Extrinsic	jeg er begyndet at træne for ca. 2 mdr. siden og har været der ned 21 gange dvs. ca. 3 gange om ugen.jeg går der for at tabe mig.	I began to train ca. 2 months ago and have been down 21 times. This means ca. 3 times a week. I do it to lose weight.
Amotivation	Mit problem er bare, at jeg er SÅ dårlig til at komme der reglmæssigt - dvs jeg ville gerne træne et par gange om ugen, men kommer der måske max et par gange om måneden...Jeg vil SÅ gerne have resultaterne (lækker krop og i god form) derfra, men er jo alt for dårlig til at tage mig sammen og komme derhen	My problem is just, I am SO bad at going regularly. This means that I would like to train few times a week, but train max few times a month. I would SO much want to have results (hot body and good form) from that, but i am just too bad at put myself together and go there.

Table 14. Random Examples of Types of Motivation from both Training set and Corpus

4.2 Overall Distribution of Types of Motivation

In order to illustrate how types of motivation are related to the rest of the models, the datasets were filtered from None categories. This was believe to provide a more relative representation of how the prediction of motivation were connected. Therefore, the resulting dataset contained a total of 5603 observations out of the 55259 observations from the corpus. Intrinsic motivation category represented 4811 observations, extrinsic motivation - 788 observations, and amotivation only 4 observations. Intrinsic motivation category is six time more represented in the sample than extrinsic motivation. Amotivation is rather considered a minority across the whole sample. The overall distribution of Types of motivation is presented in Appendix 9.9.

4.2.1 Intrinsic Motivation

In Figure 13 below, the Intrinsic type of motivation is described based on the categories from the other models. In regard to basic needs, autonomy is dominating the observation, followed by competence and relatedness at last. Intrinsic goals are also more recently connected to intrinsic motivation, than extrinsic goals are. Personality traits like openness, conscientiousness and agreeableness were the most common ones. While extraversion and especially neuroticism are not well presented in the corpus. It is also visible that the combination of autonomy needs and having intrinsic goals is most often related to intrinsic motivation. Other dominating relationships are between autonomy, intrinsic goals and conscientiousness, and also between the need of competence, intrinsic goals and agreeableness.

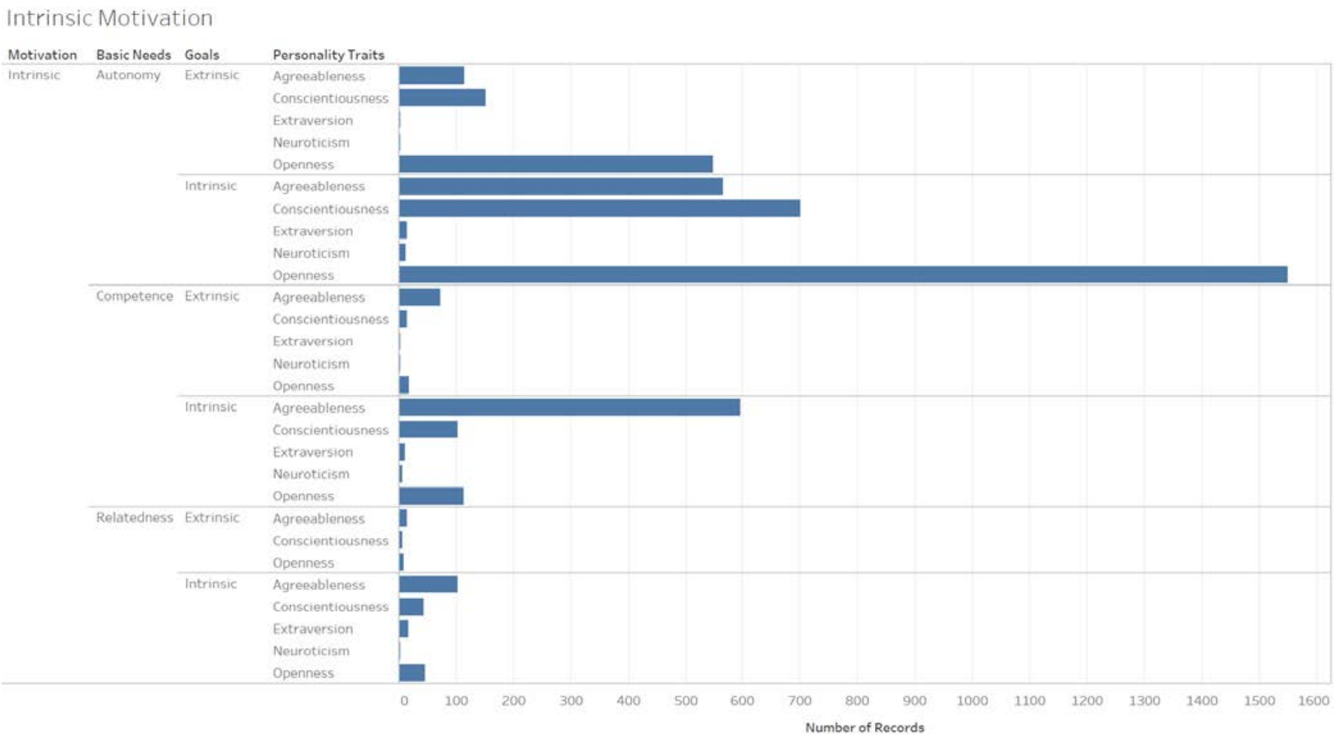


Figure 13. Distribution of Intrinsic motivation

When looking deeper into what these combination of categories had in common based on the most probable words to predict the categories, it became apparent that there were many common words (Figure 14). Intrinsically motivated discussions included texts about exercise being fun. Many texts also discussed physical exercises like running, biking, spinning, workout, and gym training, cut and bulk training, burning fat, exercises. Additionally, people were conversing about training programs, training experience, and results from the training, or seeking advice for specific matters. In these categories the words will, fun, and lust were mentioned multiple times. On the other hand, the words also implied discussions about injuries, physical problems and advices for overcoming these.

Figure 14. Common words between Intrinsic motivation and the rest of the categories

4.2.2 Extrinsic Motivation

Extrinsic type of motivation is described based on the categories from the other models in Figure 15 below. In regard to basic needs, it seemed like need of autonomy is also dominating extrinsic motivation, followed by competence and relatedness at last. In regard to Goals, extrinsic goals are more often present than intrinsic goals. Openness and agreeableness are dominating the sample, followed by conscientiousness, extraversion and neuroticism. From the sample below it is visible that extrinsic motivation is best predicted by the combination of needs of autonomy, extrinsic goals and openness. Another common combination was competence, extrinsic goals and agreeableness. This was followed by autonomy, extrinsic goals and conscientiousness.

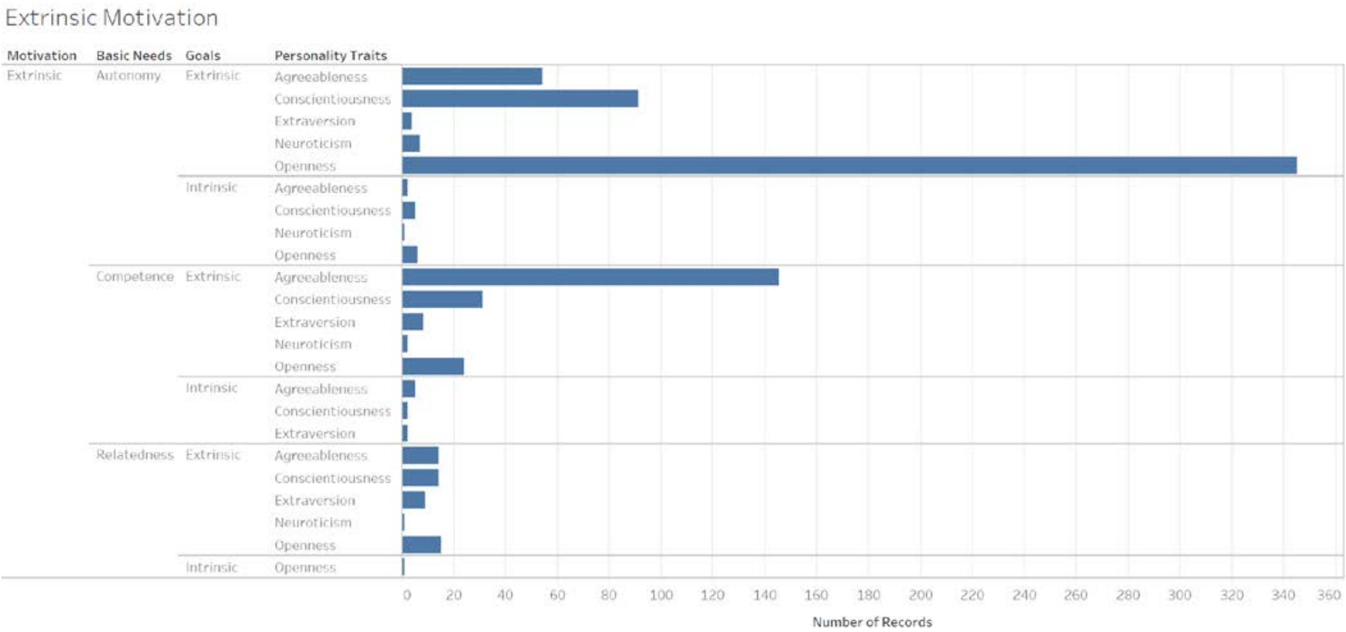


Figure 15. Distribution of Extrinsic motivation

These combinations for Extrinsic motivation led to the considerations of the probability words, shown in Figure 16. The combinations for extrinsic motivation shared many words with intrinsic motivation, mainly because the Basic needs and Personality traits were the same. This pointed at the consideration that there are many common topics and discussions which both extrinsically and intrinsically motivated people engage in. Specific for extrinsic motivation was the focus on “kg” or “kilo”, and physical appearance, looking good, losing weights, having a good bum, going something that is free of charge. These were also reflected in the extrinsic goals. It seems that many of the topics were including new beginners.

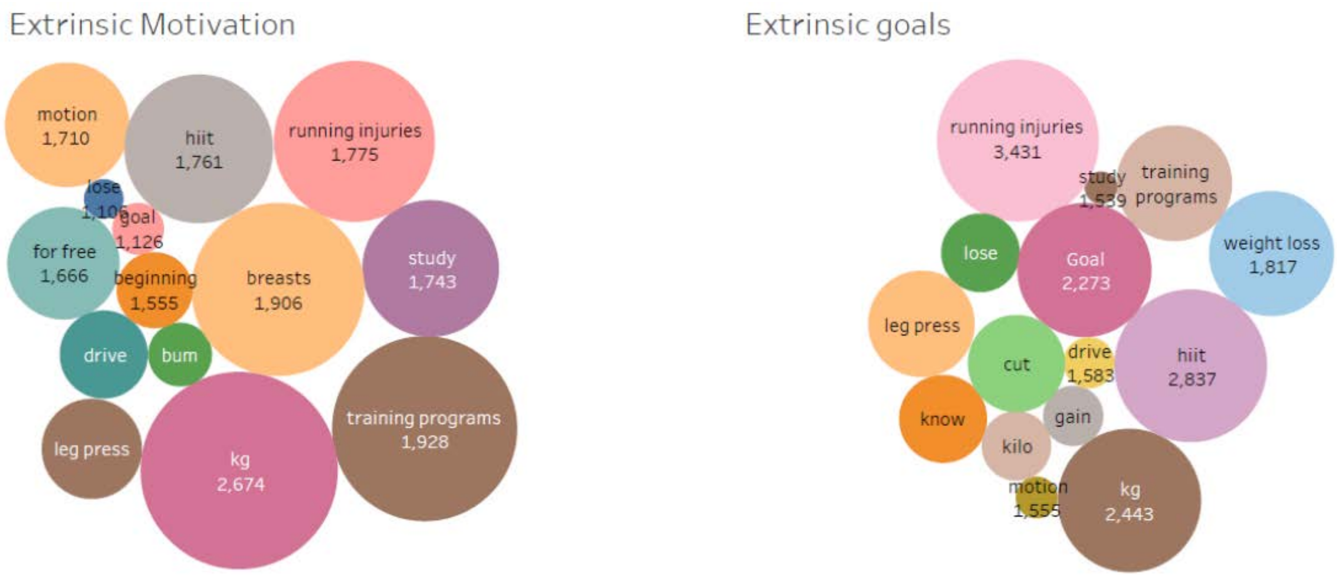


Figure 16. Common words between Extrinsic motivation and Extrinsic goals. The rest of the categories are the same as Figure 14.

4.2.3 Amotivation

Amotivation is the least representative category in the whole sample. After removing all None categories, this led to only 4 observation having this category (Figure 17). As it is a very small sample it is hard to generalize. Therefore, the Figure 18 provides a representation of amotivation, while excluding only from None type of motivation, and including the other None categories for Basic needs, Goals and Personality traits. This was done in order to compare the findings and check whether there is a meaningful explanation of Amotivation. In Figure 18, it is visible that the None Basic need category in combination with None goals are most presentable. In both figures Neuroticism is most often associated with Amotivation.

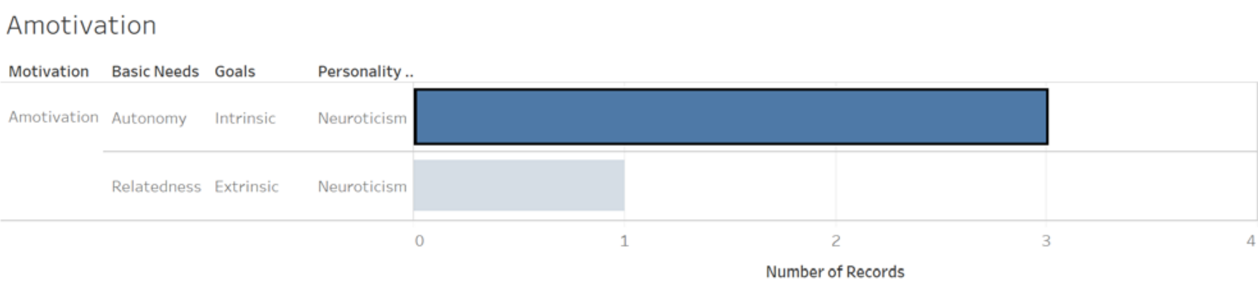


Figure 17. Distribution of Amotivation, when None is completely excluded

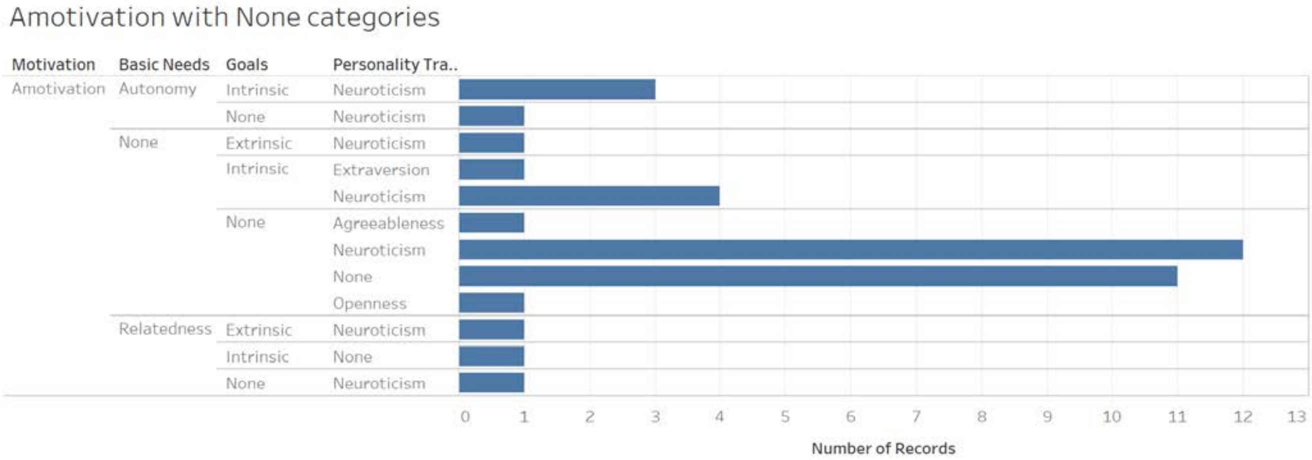


Figure 18. Distribution of Amotivation, when None is included for Basic needs, Goals and Personality Traits.

Even though the small number of observation, it was possible to depict Amotivation through the common words and topics discussed in the data (Figure 19). One part of the text discussed the inability of doing physical exercises because of “injuries”and “problems” or being hard to work out “after operations”. Many people were dropping out of the trainings because there lack of “willpower”, which eventually led to “bad conscience”.

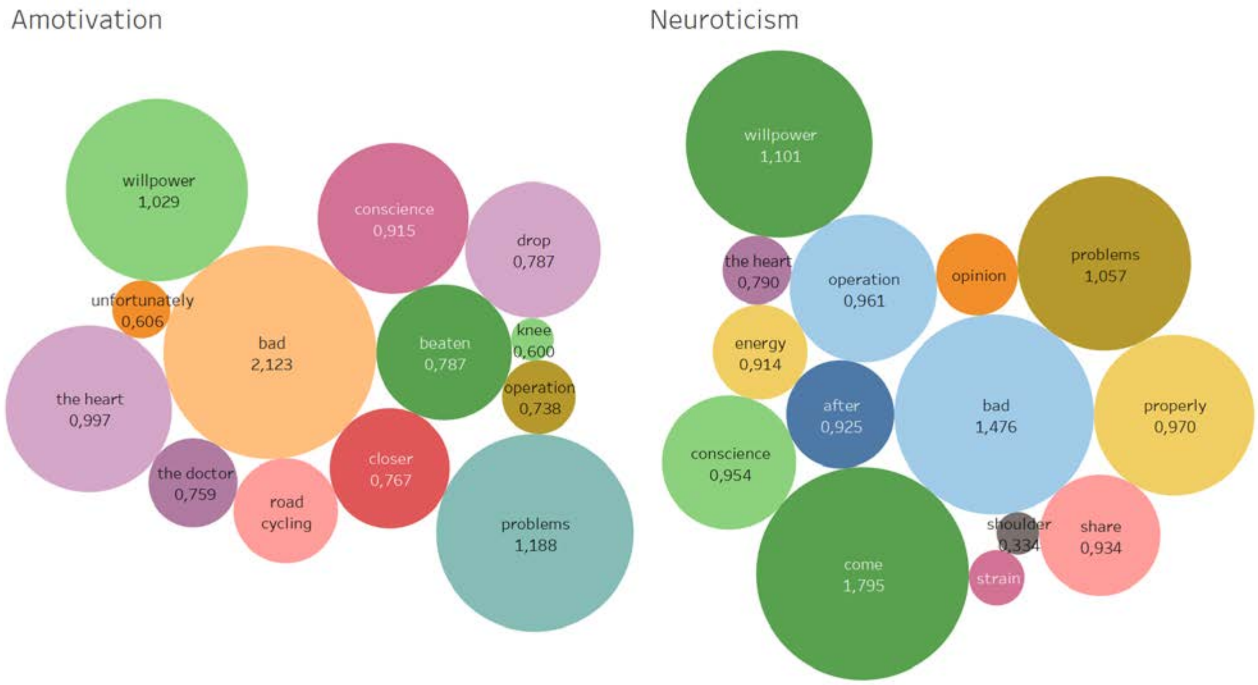


Figure 19. Common words between Intrinsic motivation and the rest of the categories

4.2.4 Motivation and Social Media Activity

The Facebook datasets provided an extra field measuring the type of activity and type of post. This enabled a representation of the distribution of intrinsic and extrinsic motivation in regard to activities. The most common activity for both intrinsic and extrinsic motivation was a photo post, followed by a comment and link post.

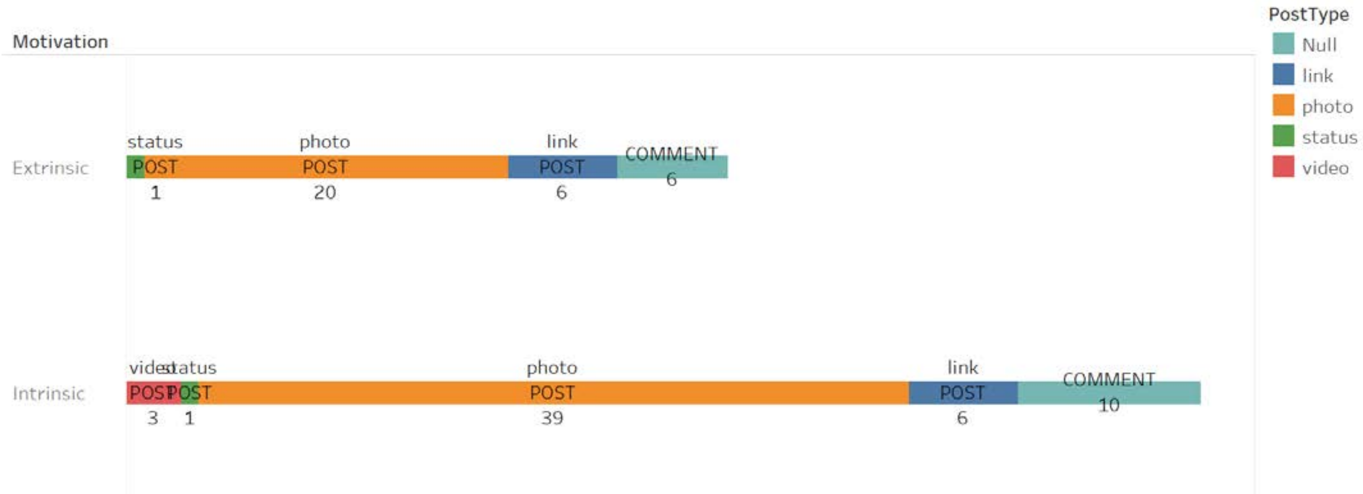


Figure 20. Social media activity and motivation types across the sample

4.3 Datasets - Actors and Actions

This study found it interesting and intriguing to discover how the categories were distributed in regard to the data sources. This was believed to show whether there were some topics that were more specific to the particular data source. In order to provide a more comprehensive representation, the findings were grouped in two categories. First, the five Facebook pages because of their similarities in regard to fields. Second, the forums were grouped together, in order to be able to reflect on their similarities but also possible differences.

4.3.1 Facebook

The Facebook data did not contain any records from Abilica, which were not categorized by None. This also reflected the overall purpose of the page, which was to promote sport equipment, even though its description said that they are posting inspirational tips for exercise. Similarly, the Fitness DK dataset contained only 13 records. The other 3 pages contained more records. Figure 21 represents the same overall tendency of social media activity as represented in Figure 20, when controlling for type of motivation. Photo posts are the most common action, which is followed by link posting and also commenting.

Figure 21 Social media activity on Facebook

4.3.2 Forums

For some of the forum data, there were also Topic field. Therefore, after the classification these topics could be linked to the classified texts. It was possible to depict the top 10 most commented topics, and consequently link them to types of motivation (Figure 22). Amotivation was not among these topics, and there were few example of extrinsic motivation. The topics classified as intrinsic dealt mainly with tips and sharing of experience like “Running program”, “Spinning”, “Running/Cycling/Swimming”, and similar. Many of the topics focused on putting one’s motivation up by “when does it become fun?”, “remembering to complimenting yourself”, “what to do to keep motivated”, “finding the joy for running”, “goal with your training”. Among the topics there were also some addressing the need for support with the training like “eating”, “program”, “help”, “new beginners” and generally tips for the different types of sports like running, spinning, cycling, bodybuilding. Another category of topics arousing attention were “Stop smoking and cardio”, “Obesity”, “anorexia”, “pain”, “restitution”, “sore muscles”, “knee problems”, which pointed at the fact that many people are looking for support or help from others with similar experiences.

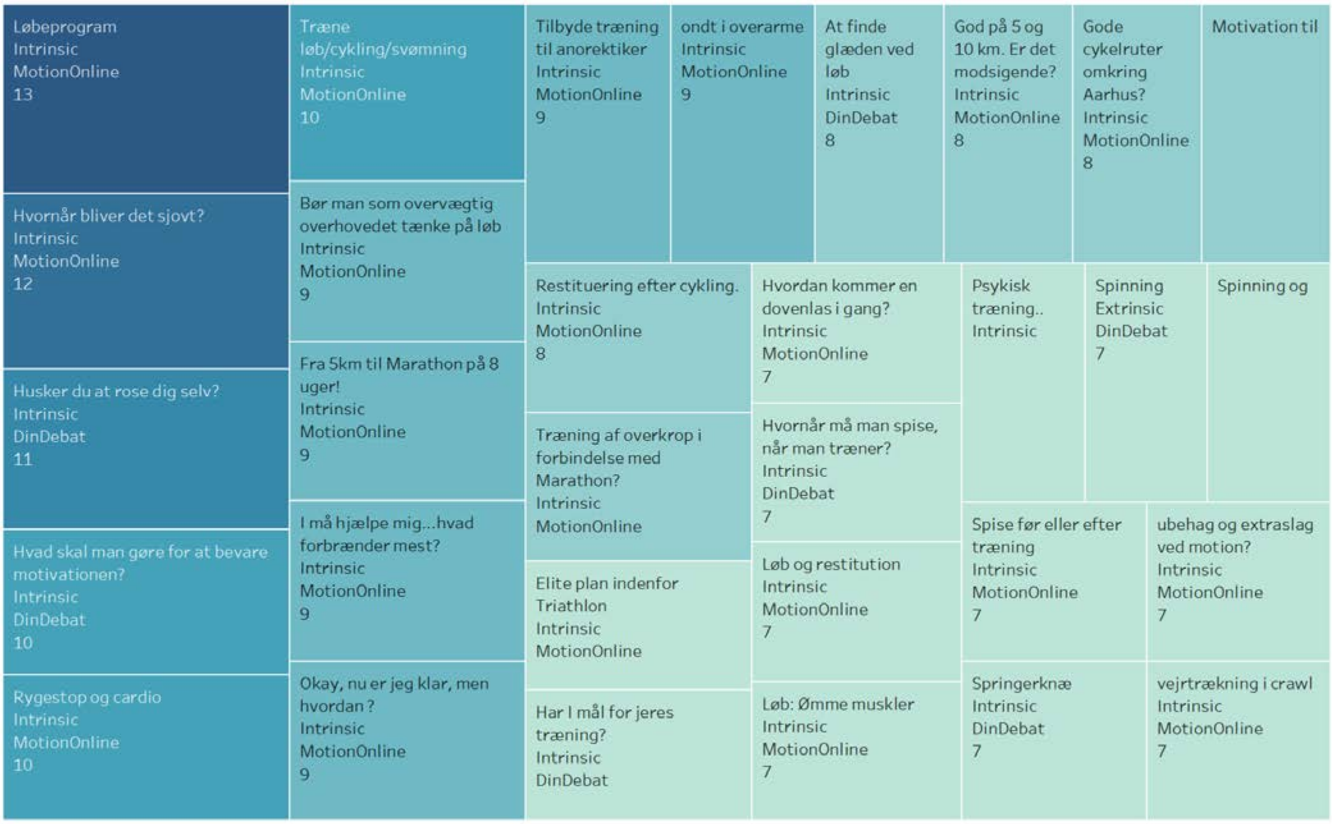


Figure 22. Distribution of most popular topics divided by type of motivation

5. Discussion

The process of data analytics involves the collection, modelling and analysis of large volumes of data, and deployment of findings, and this study went through these steps. At the data collection step, this study had gained access to data. As Flyverbom and Madsen (2015) stressed, data is like “raw material”, which is only useful when turned into information and further on turned into actionable knowledge. The visualization of the processed data was the step where the data started taking the shape of organized information. But it is first in the discussion where the information is turned into a knowledge which can be acted upon. Thus, this section aims at evaluating the data analytics process and afterwards depicting the valuable insights uncovered from the data analytics process. At last providing actionable guidance for the promotion of motivation to physical activities.

5.1 Evaluation of the Model

This research has identified that majority of previous studies of motivation tend to apply one and the same questionnaire (Behavioral Regulation in Exercise Questionnaire) for the investigation of motivation in the physical activity settings (Hagger and Chatzisarantis, 2008). On one hand, this has yelled many valuable and confirmed conclusions about this phenomenon, which were also taken into considerations in this study. On the other hand one should also consider the limitations of applying the same questionnaire, in regard to biases, limited number of observations and participants, ethical considerations, participants' errors, and the ability to uncover newer and richer knowledge. Therefore, in the seek to uncover newer insights and actionable knowledge, which would have a more spread application for the academic world and the practise, this study undertook a data analytics strategy (Vatrapu et al., 2016). Even though data analytics is continuously fighting its way in into the academic literature, there is still lack of predefined research designs and approaches. In order to provide a reliable and valid results, CRISP-DM was utilized as a research strategy, while having traditional quantitative research as a referential framework. The study followed the outlined six steps and data analytics life cycle from CRISP-DM.

5.1.1 Review of the Process

This study had as an initial purpose to explore how predictors of motivation for physical activities can be identified from social media texts. In order to answer this question, both the research and the methodological consideration were used as starting point. Gaining an understanding about the predictors of motivation was done with the help of literature review of Self-Determination Theory. The theory has been applied to many different fields among which physical activity and exercise was one. The theory postulated that there are three different types of motivation, which are distinguished upon a continuum going from autonomous, controlled to impersonal orientations. Types of motivation were further differentiated between basen on three categories of basic needs and two categories of goals. Motivation was explained to be affected by the social context on one hand, and the individual's differences on the other hand. This has led to consider that the three basic needs of autonomy, competence and relatedness were directly dependent and influenced from the social context, and the intrinsic and extrinsic goals were more personal (Deci and

Ryan, 2015). In order to reflect on the individual differences of the persons, the personality traits, which are external for the SDT, were considered as adding a supplementary value for the investigation and thus included in the overall conceptual model of motivation. With the help of SDT and already existing research in the field of physical activities, intrinsic motivation was seen as the type of motivation most commonly associated with physical activities.

The conceptual model was tested against the collected data, firstly by manually categorizing the training set and secondly by applying text classification with the help of five supervised machine learning techniques. The classified models were quality evaluated based on precision, recall, F1 score and accuracy. The overall performance of the Passive Aggressive and Linear SVC Classifiers was satisfactory in regard to accuracy, scoring over 0.83. The classifiers also identified a set of 20 words which had highest probability to predict a category, and 20 words which had the lowest probability to predict a category.

The current study was able to apply data analytics and obtain highly acceptable accuracy score, and this is how motivation was derived from the texts. Moreover, this strategy was also seen as an advantage for the study because it avoided participants bias by collecting data from relevant pages, where participants did not know that they might be subjects of investigation. This was beneficial because they would discuss the topics of interest as they will discuss it anyways and their use of language was as what they will normally use in order to describe their opinions, feeling, experiences, etc.

5.1.2 Ethical Considerations

This study was also reflective in regard to ethical considerations involved with personal and sensitive data that might be disclosed in the texts, or tracked down to the persons behind the texts. This study also took into account that the used examples from the texts can be Googled and can eventually lead to the username behind the text. Anyhow, it is also worth noting that usernames are publicly available and the texts can be Googled despite of this study. Thus, this study has no way of harming or undermining the privacy of the users. Moreover, with the coming into force on 25th May 2018 General Data Privacy Regulation (GDPR) many of the previously available data about the data subjects became unavailable and the study respected this. Unfortunately, the scandal with Facebook and Cambridge Analytica forced Facebook to straighten its privacy policy even more. This led to anonymising the data subjects' names by replacement them with the name of the page they were posting in, and consequently made it impossible to report on number of unique or returning users. Only the forum datasets had the usernames of the users which were registered on the page, and those who were visiting were blank or called "guest".

5.1.3 Dataset Limitations

Apart from not being able to link the texts with the usernames, there were a limited number of other fields available in the datasets. Those provided information about date of the activity and type of activity. During data collection, all texts on the pages were selected, without applying any filters. This allowed for unbiased selection of data subjects, meaning that these pages were all public and anyone would be able to read, like, post and comment. Furthermore, the training set comprised of 2000 texts was also randomly selected, in order to provide an unbiased sample for the training.

The five Facebook dataset did not include that many records. Some of them only included 200 records in total. Also, the five Facebook pages range differently in regard to the purpose of the page. For example Abilica and Fitness DK were more of business promoting pages. Even though in their description they mentioned that they are posting motivational and inspirational tips for physical activities, there was mainly marketing content. These pages also allowed users to post, comment and like but these texts were rather a minority.

The other three Facebook pages were more of user created public page where user-generated content was prevalent. This allowed for the exploration of wider scope of topics discussed between the users.

The forums provided with even wider ranges on topics, but they were more centered around the overall discussion of motivation to physical activities. These dataset characteristics made it harder to generalize, but in the same time this provided a wider overview, which was believed to provide wider range of insights.

The following discussion takes a step deeper into the text classification process, and provides reflection on the valuable insights and actionable knowledge extracted from the text classification.

5.2 Turning Data into Information

The conducted classification of the social text has also enabled a more overall presentation of how the predictors of motivation are related to each other. This help the analysis with identifying the common connections between the categories and the common word that predicted them

5.2.1 Predictors of Motivation

This study benefited from a theory based conceptual model of motivation, which had guided the data modelling and analysis. The conceptual model was also backed by previous literature in the field of physical activity, which has proven multiple times the relationship between the categories of Basic Needs and Goals, and the types of motivation (Lonsdale et al., 2013; Kwan et al., 2011). The finding from these study were also an illustration of the widespread establishment of SDT in the field of physical activities because it was used as a tool to measure motivation (Hagger and Chatzisarantis, 2007). This study was able to find support for the strong relationship between intrinsic motivation and physical activities. More than 85% of the total observations were classifier as intrinsic, leading to the assumption that intrinsically motivated people are more likely to be conducting physical activities than extrinsically motivated people, which represented only 14% on the sample. This finding support the postulation that more intrinsic motivation has been linked to more representation of physical activity (Fortier et al., 2011). Also, the findings support the tendency that intrinsically motivated people are more interested in reporting their activities, which was reflected in the content (Rodgers et al., 2010). This was also illustrated in Figure 20, where intrinsic people are more active than extrinsic. Amotivation has also been proven to have weakest and even negative association of physical activities, which is also reflected in the sample where less than 1% of the texts were classified as amotivation (Lonsdale et al., 2013).

When computing the relationship between the three types of motivation and the other models (section 4.2), it was shown that the most common predictors of extrinsic and intrinsic motivation in regard to basic needs were autonomy and competence. Even though the accuracy of classifiers for this model were under the desirable level, it was possible to spot the relationship between the words that were common for basic needs and types of motivation. As the theory postulated, the three basic needs are highly dependent on the social environment, which acts as either satisfying or thwarting the normal functioning of the human. It is in our human nature to satisfy these needs, and that is how motivation to act in regard to physical activities is boosted. Thus autonomy and competence are predictors for both extrinsic and intrinsic motivation (Deci and Ryan, 2000). The need of relatedness was also a predictor of intrinsic and extrinsic motivation, even though it was not well presented in the observations. The texts classified as relatedness, involved sports like biking, running and working-out in fitness, which are not necessary team sports. This might be a reason why relatedness was not often reflected in the results. People who are looking for relatedness would probably engage in more team oriented physical activities like team sports, where the social setting would enhance their need of relatedness (Silva et al., 2008). Other words predicting relatedness were "team plan" or "blog", "closer", which reflect on the desire to be surrounded with others, follow each others development and feel connected with them (Ryan and Patrick, 2009).

Also common predictors of intrinsic and extrinsic motivation were the personality traits for openness, agreeableness and conscientiousness. Openness was the most common predictor for both types of motivation, which reflect that people have a broad range of interests and are willing to trying something new in regard to training. This also implies that they are willing and open to new ideas which would provide various training opportunities. Conscientiousness was more common for intrinsic than extrinsic motivation, which also reflects that intrinsically motivated people would be more structured and more determined to achieve their goals connected to physical activities. A lot of people were looking for support and advice, and there were also many that were there to help and cooperate. This was reflected by the agreeableness trait, classified as a predictor for intrinsic and extrinsic motivation.

It seemed that the only difference between these two types of motivation in regards to predictors were the goals. Intrinsic motivation was best predicted by intrinsic goals, while extrinsic motivation was best predicted by extrinsic goals (Deci and Ryan, 2008; 2015). The findings also reflected that intrinsic goals would aim at the satisfaction of the basic needs and also the prediction of intrinsic motivation. The intrinsic goals reflected discussions about exercise that is fun, about desire for personal progress and self-acceptance through exercise. Extrinsic goals, on the other hand are were descripted by topics of weight lost and obesity, which are not necessary satisfying the basic needs and can eventually lead to lack of internationalization of the inspirations. These goals can be triggered in the desire to be socially accepted, which might not sustain long-term physical activity.

In the few observations representing the relationship between amotivation and neuroticism, a set of common words were identified. In support with the literature, the findings also pointed out that the discussions involved bad conscience, which had a diminishing effect on ones motivation. People were mentioning that it was hard to get out and go the gym. Another stream of topics reflected on the physical inabilities to satisfy the basic needs which also led to lack of motivation (Ryan and Patrick, 2009). Ryan and Patrick (2009) had also concluded that amotivation is complex in its nature because there are multiple reasons for why people would experience lack of motivation for a specific action. The findings support that bad consciousness of dropping or skipping a training was seen as one of the reasons. Health issues, physical inabilities, and recent operations were also seen as obstacles for the motivation, and were rather a fact that can't be done so much about. People were seeking support and tips from experienced and other in their situation, in order to find a solution that would fix or improve their situation. One of the user simply put it “...can’t imagine myself to train”. If this is a common understanding, then amotivated people will not be participating in online discussions about physical activities. They will not show up on these page, mainly because of lack on interest to physical activities. This understanding was also supported by the finding where a max of 38 records were classified as Amotivation.

5.2.2 Commonly Identified topics

Language was central for this investigation because of the use of many different words and expressions. These offered a representation of wide range of opinions, discussions and concerns motivated and not that motivated people shared with each other. Language is seen as the medium of information because information is produced, distributed and communicated with the use of written and spoken language. This is considered beneficial for this study because language can provide a wide range of insights especially with the availability of big social data (Grimmer and Stewart, 2013). This part of the discussion focused on extracting information from the data by reflection on the topics that emerged from the analysis. The topics are discussed in regard to the words identified by the classifiers, topics identified through the manual coding and previous literature.

5.2.2.1 Physical Injuries

Physical activities are seen as advantage for our health but in some cases, they can also have side effects. In cases where people are not experienced and competence in for example running, cycling and doing body building, there seem to be a common symptoms of knee injuries, back and arm pains. Even for intrinsically motivated people that would be seen as an obstacle and possibly reduce the enjoyment associated with the activity. It seemed like many were frustrated from their physical inabilities and previous injuries that were arousing from their current training. For both intrinsically and extrinsically, these seemed to be common topics.

The fact that people are actively and autonomously looking for advice and tips implies that they are encouraged and goal-orientated about their training, which was also reflected in the personality traits. Additionally, it seemed that many would ask for advice in the social community instead of consulting with practitioner. This might be because of the trust that there are other people that had similar injuries and experiences, and they can relate to each other. In a lot of the texts in the training set there were discussions about knee problems and that running shoes is one of the most common reason for that. This led to discussions about products, brands and stores where people can get more personalized products that would solve their problems.

5.2.2.2 Bodybuilding

Mainly from the Muscle Zone forum, the main conversations included the use of supplements, steroids, drugs and others that are used for muscle gain, quick fat loss and improvement of the body. These topics would include discussions about specific low fat and high protein diets that are combined with specific type of training. It seemed that many of these conversations were marked with extrinsic motives, aiming at being attractive, improving one’s appearance, being socially accepted, and being ego-centered. As suggested

from the model, these were also people open to try and experience with new things. Therefore there were many that used online sources to gain information and purchase supplements that were not necessary legal to use. In the same time, these people were very structured and goal-oriented, which was reflected in their strict training programs and their will to train six times a week. In the seek to achieve their goals, it seemed like there is a high idealization and look-up into professional bodybuilders, involving following their blogs, videos and guidance. Also, people were supporting each other with feedback on training programs and diets, and with advice of how to do specific types of exercises. This type of behaviour was mainly reflected a more agreeableness.
Another stream of conversations were related to the side effects of the intake of the supplements and steroids/drugs. Leading many people to experience headache, lack of sleep, and even undertaking surgeries for the adjustment of Gynecomastia (enlarged male breasts due to imbalance of hormones), for example. Again, there were few cases where the people seeked support from the doctors, but still many were seeking advice from the community before going for a professional check-up.

5.2.2.3 Advice Seeking

Acting possibly out of the need of relatedness there were a lot of discussions addressing the seek of advice. Among the topics about injuries and help for training programs and diets, there also were inquiries for feedback about where to train, which gym were better than other, and what classes to choose from. In regard to weather conditions and very cold temperatures in the winter, many seeks advice about appropriated outdoor outfit.
New beginners were also common participants in the conversations. They mainly seeked support from those with training experience about how to reach their goals. Especially, when people were motivated to lose weight and to get in shape they were seeking advice, ideas and suggestions for suitable training and diet that would let me achieve their goals as fast as possible. These topics might have also been inspired by the seek of positive feedback and acknowledgement of one’s efforts, which would eventually enhance their motivation (Ryan and Deci, 2000). Sustained physical activities can also be enhanced by the seek of social support which is underlined by the needs of relatedness (Wilson and Rodgers, 2004).

5.2.2.4 Encouraging Tips

Figure 22 shows the most common topics from the forums. Many people were interested in tips about how to keep one's motivation up. Some of the advices included the use of making plans one day at the time and trying to keep to them as much as possible. Also, the setting up of feasible goals was better than big goals, which might bring bad conscience and have the reverse effect.

As SDT postulated, having intrinsic motivation into particular physical activity, would require that one have an interest. Finding a training that brings positive emotions or is seen as enjoyable would increase the sense of intrinsic motivation. The corpus reflected the overall Danish culture and the understanding of "hygge". The topics reflected that by implying that training needs to be "hyggeligt", meaning that you have to feel pleasure, comfort and coziness by the activity. Having fun and having lust for the training is another factor predicting intrinsic motivation, and is especially applicable in the danish context.

On the other hand, more extrinsically motivated people would mean that exercise is only meaningful when one gives all that it takes, "No pain, no gain" was mentioned at least four time in the dataset.

Seem more of a shared goal was the desire to live healthier, which was reflected in texts about how to get in shape, how to get in good condition, both of which would eventually lead to good appearance. Some texts were expressions of motivational experiences from people that have previously been obese but had managed to get into the desired shape. Fighting obesity seemed to be a common goals for many, especially for mothers that have just given birth. Some of these texts included photos illustrating the progress one made, which was the main social media activity in the corpus. In this regard, people sharing experience were also encouraged others to be patient because it takes time to see the results and one needs to remember that.

Another approach to keeping motivation up was by setting up a milestones. Conversations about how to achieve the necessary skills for running a marathon, or participating in a competitions, where one need to prepare for many months. These would involve discussions about active days where one would exercise, and passive days, where one would rest and restiturate the body.

Discussions also included the need for variety in the exercise in order to keep the thrill and joy of physical activities. For the body builders these included exercising different body part, while for others it involved trying new classes at the gyms, or trying new sports (Sylvester et al., 2018).

5.3 Turning Information into Actionable Knowledge

This part of the discussion reflects upon the identified topics and thus provides guidelines, which practitioners, health care organisations or sport centres could employ in the promotion of motivation for physical activities.

5.3.1 Being Present When People Need Support

The finding pointed at the assumption that many people would find health care information regarding injurings and health problems from online sources or by asking others for advice. This tendency was rather more popular than the tendency of seeing the doctor or talking with a practitioner. However, one of the reasons for amotivation was actually the fact that people could not be physically active because of injuries or physical inabilities. In order to provide timely support and professional advice that would benefit the individuals in their motivation to be physically active, it is important to be there, where the people are. Timely support can actually prevent the occurrence of injuries caused by the lack of exercise competence.

People seeking advice online, should be able to get hold on professionals also online. Thus the promotion of motivation would require the availability of online community of specialists, where people can get professional guidance in order to avoid injuries and get advice of what exercise and training technique are beneficial for their specific body, goal and lifestyle. A page on Facebook or a forum sustained by professional staff might promote a enhancement of motivation for physical activities by preventing side effects of exercise (Li and Wang, 2018).

5.3.2 Turning Extrinsic into Intrinsic Motivation

This research was able to identify the differences between extrinsic and intrinsic types of motivation, when looking at the predictor of each. The results have proven that both types of motivation are related to physical activities, and that it is important to reassure the sustainability of intrinsic motivation (Buckworth et al., 2007). Previous research has found that maintained and long-lasting physical engagement is highly dependent on turning extrinsically motivated people into more intrinsically motivated (Silva et al., 2008; Edmunds,Ntoumanis and Duda, 2006). Therefore learning from what intrinsically motivated people do to satisfy their needs and achieve their goals is central for the promotion of physical activities (Lonsdale et al., 2013).

Another point to consider is that intrinsically motivated people might not necessary be doing physical activities because of the perceived enjoyment from the exercise, but because of them being very goal-oriented and persistent. This assumption is supported by the relationship between intrinsic motivation and conscientious personality trait, which is one of the dominating in the sample. The data reflected on this point with topics that contained tips for how to keep yourself motivated by setting up small goals, to be proud of the small achievements and to be patient. These factors are important to consider for new-beginners, whose goals are more extrinsic than intrinsic (Edmunds, Ntoumanis and Duda, 2006). In the Danish context, it seemed that enjoyment and fun related to the exercise needed to be sustained and promoted in order to boost one's motivation and engagement with physical activities (Edmunds, Ntoumanis and Duda, 2006).

5.3.3 Social Media Engagement

Another important fact in the promotion of motivation is that the different types of motivation would be exhibited differently in regard to online activity. The findings point at the tendency that highly motivated people would be also interested in being active in sharing their experiences online. In the seek of social support reflected from the satisfaction of relatedness, social media can be seen as the medium to satisfy one's needs of autonomy, and competence. Social media is often related to the overall understanding of online social community, where people benefit from social support, social interactions and social sharing. The study has proven that participating on social media has a positive effects on the physical activities, especially when others are also physically active (Groenewegen et al., 2012). Sharing everyday experiences and seeking of advice from others with the same interests can eventually contribute to the need to be part of a community, give a sense of belonging and increase one's connectedness to the activity itself. In the same time a social support and feedback from the online community can be seen as boosting one's motivation and encouraging their intentions as there is a sense of competition and challenge, and sense of idealizing what the others are doing.

Previous research has also concluded that intrinsically motivated people are more often online, which also reflected the idea that social media participation can boost one's motivation. On the other hand, the findings suggested that amotivated people would not be participating in discussion about physical activities, simply because they do not have interest for these. Therefore, amotivation was hardly presented in the corpus. Though, it suggestions that amotivated people are not part of these specific pages, but they are still on Facebook. Thus, attracting their attention and gaining their trust would possibly enable them to reconsider their intentions to physical activities. Promoting motivation to physical activities would then require to get amotivated people to participate in conversation about physical activities in order to get inspired and satisfy their needs. Attracting the attention of amotivated people can possibly be done by sharing inspirational photos, videos, content of others who overcome their amotivation and reached their goals. The findings also reflected that many photo post were connected to intrinsic motivation. Learning from motivated people would also lead the way to helping amotivated into physical activities. Social media is already bombarded with target marketing, then maybe motivational advertisement content would promote the physical activities (Li and Wang, 2018).
In regard to relatedness, social media can also be seen as the medium for receiving a positive feedback. The facebook data indicated that there is a majority of photo and link posts done by intrinsically motivated people. This implied that sharing experiences and getting "likes" on social media can be beneficial for one's motivation. Therefore, the findings support the believe that positive feedback enhanced the feeling of competence and led to increase in motivation (Ryan and Patrick, 2009). Previous research has also reflected that sharing physical activity content provides social support through friends' encouraging comments, which for open and conscientious people would mean new challenges and stimulation for achieving goals (Klenk, Reifegerste and Renatus, 2017). The sense of competition and challenge can be used for promoting motivation for physical activities by creating engaging posts on social media (Kim and Park, 2016).

5.4 Reflections, Limitations and Future Research

As any other study, this one also had its limitations. The use of social data analytics in academic research suggest new innovative ways for the search of ideas, concepts, and processes previously undiscovered. This also implies considerations about data subjects' right to privacy and ownership. For publicly available data, it is important to take into considerations about data quality, and why this data is available.

In regard to identifying predictors of types of motivations, this study was not able to analyse deeper into the distinction of the four subcategories of extrinsic motivation. This would have provided a further discussion on what possible recommendation that can be in order to turn extrinsically motivated people into intrinsically motivated.

The use of both Facebook and forum data was on one hand beneficial for getting hold on many different topics regarding motivation, but it also made it harder to generalize. Abilica and Fitness DK Facebook pagers were more of a marketing tools rather than motivation promoting content, despite the fact that their description said so. Involving them into the investigation did not benefit the research because there were very few observations where motivation could be predicted. The other three Facebook pages, were anonymized, while the forum data did not contain fields about actions, and that made it impossible to add more coherent representation of socio-technical interactions into the paper.

This study only focus on the Danish context, which according to the WHO has one of the highest numbers of physically active people. However in order to fight the overall problem of promoting motivation, it would be beneficial if future research compare these findings against eastern counties, where according to WHO people's lifestyles is under the recommended level for physical activity. Additionally, it would be beneficial to involve demographic data into the research. Gender, age, location etc. can be used to compute possible differences in people's motivation and physical activity behaviour. Another beneficial source of information would be lifestyle factors as job, income, education and similar, where the lifestyle and personality of the people can be better extracted.

6. Conclusion

This study considered motivation for physical activities as a public health care parameter for the prevention of sedentary lifestyle diseases. Through the framework of Self-Determination

Theory and previous research in the fields of physical activities and social media, a conceptual model was developed. Based on it, motivation for physical activities can be predicted through the effects from the social environment emphasised by the basic needs for autonomy, competence and relatedness, on one hand. On the other hand, motivation is also predicted by more personal influences such as intrinsic and extrinsic goals. SDT is a motivational theory of personality and development, therefore the conceptual model also employed personality traits of openness, conscientiousness, agreeableness, extraversion and neuroticism for the exploration of personal differences. The unit of analysis of the study was social text, which was collected from Facebook and online forums, discussing physical activities. A random sample of 2000 texts were manually coded with the help of content analysis. Thereafter, the training set was passed to Supervised machine learning techniques, among which Linear Support Vector Classifier and Passive Aggressive Classifier were used for automated text classification of the 55259 texts.

The findings point that supervised text classification is an efficient method for deriving predictors of motivation from social texts, where the classifiers scored higher than 0.83 in accuracy. With the help of the words identified by the classifiers, content analysis of the training set and previous literature, the study found consistent support that intrinsic motivation is most often associated with physical activities. Intrinsic motivation was also most often related to intrinsic goals and the needs of autonomy and competence. Intrinsic people were also most often related to open, conscience and agreeable personalities. Furthermore, the classifications of the models led to a deeper exploration of the relationships between the categories and consequently to the extraction of common topics among the texts. Among others discussions about physical injuries, bodybuilding, advice seeking, and encouragement tips were the most often topics expressed by intrinsically motivated people. Based on these insights it was concluded that practitioners, health care organisations or sport centres can promote motivation for physical activities by being present when people need support. Additionally, an effort into turning extrinsically motivated into intrinsically motivated people would ensure a regular and sustained motivation for physical activities. Finally, engaging on social media is another guideline for catching the interest of the people and helping them into the pursue of interest toward physical activities.

7. Acknowledgement

I would like to thank my supervisor Raghava Rao Mukkamala for the hours and efforts spent on supporting this study. Furthermore, I would like to thank him and his team for helping with the data collection and the machine learning classification, which were only possible with his help.

I would also like to acknowledge the people who have proofread his paper.

8. Reference list

- Agarwal, B. Personality Detection from Text: A Review. p.4.
- Agarwal, Basant.(2014). “Personality Detection from Text: A Review,” n.d., 4. International Journal of Computer System
- Aggarwal, C. C., & Zhai, C. (2012). Mining text data: Springer Science & Business Media.
- B. Owen, K., Smith, J., Lubans, D. R., Ng, J. Y. Y. and Lonsdale, C. (2014). Self-determined motivation and physical activity in children and adolescents: A systematic review and meta-analysis. Preventive Medicine, 67, pp.270–279. [Online]. Available at: doi:10.1016/j.ypmed.2014.07.033.
- Boldt, L. C., Vinayagamoorthy, V., Winder, F., Schnittger, M., Ekran, M., Mukkamala, R. R., Lassen, N. B., Flesch, B., Hussain, A. and Vatrapu, R. (2016). Forecasting Nike’s sales using Facebook data. In: Big Data (Big Data), 2016 IEEE International Conference on. 2016. IEEE. pp.2447–2456.
- Bonaccorso, G. 2018. ML Algorithms addendum: Passive Aggressive Algorithms - Giuseppe Bonaccorso. [ONLINE] Available at: https://www.bonaccorso.eu/2017/10/06/ml-algorithms-addendum-passive-aggressive-algorithms/. [Accessed 05 May 2018].
- Boyd, Danah, and Kate Crawford. “CRITICAL QUESTIONS FOR BIG DATA: Provocations for a Cultural, Technological, and Scholarly Phenomenon.” Information, Communication & Society 15, no. 5 (June 2012): 662–79. https://doi.org/10.1080/1369118X.2012.678878.
- Brickell, T. A. and Chatzisarantis, N. L. D. (2007). Using self-determination theory to examine the motivational correlates and predictive utility of spontaneous exercise implementation intentions. Psychology of Sport and Exercise, 8 (5), pp.758–770. [Online]. Available at: doi:10.1016/j.psychsport.2006.11.001.
- Buckworth, J., Lee, R. E., Regan, G., Schneider, L. K. and DiClemente, C. C. (2007). Decomposing intrinsic and extrinsic motivation for exercise: Application to stages of motivational readiness. Psychology of Sport and Exercise, 8 (4), pp.441–461. [Online]. Available at: doi:10.1016/j.psychsport.2006.06.007.
- C., K., 2016. E-Commerce 2016: Business, Technology, Society (12th Edition). Pearson.
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. & Wirth, R. (2000). CRISP-DM 1.0: Step-by-step data mining guide. Chicago, IL: SPSS.
- Chen, H., Chiang, R. H. L. and Storey, V. C. (2012). BUSINESS INTELLIGENCE AND ANALYTICS: FROM BIG DATA TO BIG IMPACT. 36 (4), p.24.
- Crammer, Koby, Ofer Dekel, Joseph Keshet, Shai Shalev-Shwartz, and Yoram Singer. (2006) “Online Passive-Aggressive Algorithms,” n.d., 35.
- Deci, E. L. and Ryan, R. M. (2000). The" what" and" why" of goal pursuits: Human needs and the self-determination of behavior. Psychological inquiry, 11 (4), pp.227–268.
- Deci, E. L. and Ryan, R. M. (2008). Self-determination theory: A macrotheory of human motivation, development, and health. Canadian Psychology/Psychologie canadienne, 49 (3), pp.182–185. [Online]. Available at: doi:10.1037/a0012801.
- Deci, E. L. and Ryan, R. M. (2015). Self-Determination Theory. In: Wright, J. D. (Ed). International Encyclopedia of the Social & Behavioral Sciences (Second Edition). Oxford :

Elsevier. pp.486–491. [Online]. Available at: doi:10.1016/B978-0-08-097086-8.26036-4 [Accessed 9 February 2018].

❖ Digman, J. M. (1990). Personality Structure: Emergence of the Five-Factor Model. Annual Review of Psychology, 41 (1), pp.417–440. [Online]. Available at: doi:10.1146/annurev.ps.41.020190.002221.

❖ Duncan, L. R., Hall, C. R., Wilson, P. M. and Jenny, O. (2010). Exercise motivation: a cross-sectional analysis examining its relationships with frequency, intensity, and duration of exercise. International Journal of Behavioral Nutrition and Physical Activity, 7, p.7. [Online]. Available at: doi:10.1186/1479-5868-7-7.

❖ Edmunds, J., Ntoumanis, N. and Duda, J. L. (2006). A test of self-determination theory in the exercise domain. Journal of Applied Social Psychology, 36 (9), pp.2240–2265.

❖ f_score.pdf. [Online]. Available at: https://adamyedidia.files.wordpress.com/2014/11/f_score.pdf [Accessed 5 May 2018a].

❖ Flyverbom, M. and Madsen, A. K. (2015). Sorting data out Unpacking big data value chains and algorithmic knowledge production. In: Süssenguth, F. (Ed). Die Gesellschaft der Daten. Bielefeld : transcript Verlag. [Online]. Available at: doi:10.14361/9783839427644-006 [Accessed 25 March 2018].

❖ Fortier, M. S., Sweet, S. N., O'Sullivan, T. L. and Williams, G. C. (2007). A self-determination process model of physical activity adoption in the context of a randomized controlled trial. Psychology of Sport and Exercise, 8 (5), pp.741–757. [Online]. Available at: doi:10.1016/j.psychsport.2006.10.006.

❖ Fortier, M. S., Wiseman, E., Sweet, S. N., O'Sullivan, T. L., Blanchard, C. M., Sigal, R. J. and Hogg, W. (2011). A moderated mediation of motivation on physical activity in the context of the Physical Activity Counseling randomized control trial. Psychology of Sport and Exercise, 12 (2), pp.71–78. [Online]. Available at: doi:10.1016/j.psychsport.2010.08.001.

❖ Fox, K. (2012). Intervention design using self-determination theory to promote activity in depressed patients, older adults, and slimmers. Journal of Science and Medicine in Sport, 15, p.S4. [Online]. Available at: doi:10.1016/j.jsams.2012.11.011.

❖ Gandomi, A. and Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management, 35 (2), pp.137–144. [Online]. Available at: doi:10.1016/j.ijinfomgt.2014.10.007.

❖ Gao, J., Koronios, A. and Selle, S. (2015). Towards A Process View on Critical Success Factors in Big Data Analytics Projects. AMCIS 2015 Proceedings. [Online]. Available at: https://aisel.aisnet.org/amcis2015/BizAnalytics/GeneralPresentations/16.

❖ Gibson, R., Cantijoch, M. and Ward, S. (2014). Analysing social media data and web networks. [Online]. Available at: http://public.eblib.com/choice/publicfullrecord.aspx?p=1880226 [Accessed 10 April 2018].

❖ Gillison, F., Osborn, M., Standage, M. and Skevington, S. (2009). Exploring the experience of introjected regulation for exercise across gender in adolescence. Psychology of Sport and Exercise, 10 (3), pp.309–319. [Online]. Available at: doi:10.1016/j.psychsport.2008.10.004.

❖ Godin, G. and Shephard, R. (1985). A Simple Method to Assess Exercise Behavior in the Community. Canadian journal of applied sport sciences. Journal canadien des sciences appliquées au sport, 10, pp.141–146.

- González-Cutre, D., Sicilia, Á., Sierra, A. C., Ferriz, R. and Hagger, M. S. (2016). Understanding the need for novelty from the perspective of self-determination theory. Personality and Individual Differences, 102, pp.159–169. [Online]. Available at: doi:10.1016/j.paid.2016.06.036.
- Gourlan, M., Trouilloud, D. and Sarrazin, P. (2013). Motivational characteristics of obese adolescents toward physical activity: Contribution of self-determination theory. Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology, 63 (4), pp.209–218. [Online]. Available at: doi:10.1016/j.erap.2013.02.001.
- Griffiths, F., Lindenmeyer, A., Powell, J., Lowe, P. and Thorogood, M. (2006). Why Are Health Care Interventions Delivered Over the Internet? A Systematic Review of the Published Literature. Journal of Medical Internet Research, 8 (2). [Online]. Available at: doi:10.2196/jmir.8.2.e10 [Accessed 27 February 2018].
- Grimmer, J. and Stewart, B. M. (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. Political Analysis, 21 (03), pp.267–297. [Online]. Available at: doi:10.1093/pan/mps028.
- Groenewegen, Stoyanov, Deichmann, Helteren. (2012) “Connecting with Active People Matters: The Influence of an Online Community on Physical Activity Behaviour,” n.d.
- Günther, W. A., Rezazade Mehrizi, M. H., Huysman, M. and Feldberg, F. (2017). Debating big data: A literature review on realizing value from big data. The Journal of Strategic Information Systems, 26 (3), pp.191–209. [Online]. Available at: doi:10.1016/j.jsis.2017.07.003.
- Hagger, M. and Chatzisarantis, N. (2008). Self-determination Theory and the psychology of exercise. International Review of Sport and Exercise Psychology, 1 (1), pp.79–103. [Online]. Available at: doi:10.1080/17509840701827437.
- Hagger, M. S. and Chatzisarantis, N. L. D. (2007). Advances in self-determination theory research in sport and exercise. Psychology of Sport and Exercise, 8 (5), pp.597–599. [Online]. Available at: doi:10.1016/j.psychsport.2007.06.003.
- Hagger, M. S., Chatzisarantis, N. L. D. and Harris, J. (2006). From Psychological Need Satisfaction to Intentional Behavior: Testing a Motivational Sequence in Two Behavioral Contexts. Personality and Social Psychology Bulletin, 32 (2), pp.131–148. [Online]. Available at: doi:10.1177/0146167205279905.
- Howard, J., Gagné, M., Morin, A. J. S. and Van den Broeck, A. (2016). Motivation profiles at work: A self-determination theory approach. Journal of Vocational Behavior, 95–96, pp.74–89. [Online]. Available at: doi:10.1016/j.jvb.2016.07.004.
- Hussain, A. and Vatrapu, R. (2014). Social Data Analytics Tool (SODATO). In: Tremblay, M. C., VanderMeer, D., Rothenberger, M., Gupta, A. and Yoon, V. (Eds). Advancing the Impact of Design Science: Moving from Theory to Practice. 8463. Cham : Springer International Publishing. pp.368–372. [Online]. Available at: doi:10.1007/978-3-319-06701-8_27 [Accessed 20 March 2018].
- Hussain, A., Vatrapu, R., Hardt, D. and Jaffari, Z. A. (2014). Social Data Analytics Tool: A Demonstrative Case Study of Methodology and Software. p.20.
- Iacobelli, Francisco. (2013) “Too Neurotic, Not Too Friendly: Structured Personality Classification on Textual Data,”

- ❖ Ingledew, D. K. and Markland, D. (2008). The role of motives in exercise participation†. Psychology & Health, 23 (7), pp.807–828. [Online]. Available at: doi:10.1080/08870440701405704.
- ❖ Ingledew, D. K., Markland, D. and Ferguson, E. (2009). Three Levels of Exercise Motivation. Applied Psychology: Health and Well-Being, 1 (3), pp.336–355. [Online]. Available at: doi:10.1111/j.1758-0854.2009.01015.x.
- ❖ James, G., Witten, D., Hastie, T. and Tibshirani, R. (Eds). (2013). An introduction to statistical learning: with applications in R, Springer texts in statistics 103. New York : Springer.
- ❖ Jih-Hsuan Lin (2016). Need for Relatedness: A Self-Determination Approach to Examining Attachment Styles, Facebook Use, and Psychological Well-Being: Asian Journal of Communication: Vol 26, No 2.
- ❖ Kannan, S., Karuppusamy, S., Nedunchezhian, A., Venkateshan, P., Wang, P., Bojja, N. and Kejariwal, A. (2016). Big Data Analytics for Social Media. In: Big Data. Elsevier. pp.63–94. [Online]. Available at: doi:10.1016/B978-0-12-805394-2.00003-9 [Accessed 9 March 2018].
- ❖ Kim, W. and Park, T. (2016). Factors Leading to Viral Intention on Exercise Posts. In: Social Computing and Social Media. Lecture Notes in Computer Science. 17 July 2016. Springer, Cham. pp.123–129. [Online]. Available at: doi:10.1007/978-3-319-39910-2_12 [Accessed 27 February 2018].
- ❖ Klenk, S., Reifegerste, D. and Renatus, R. (2017). Gender differences in gratifications from fitness app use and implications for health interventions. Mobile Media & Communication, 5 (2), pp.178–193. [Online]. Available at: doi:10.1177/2050157917691557.
- ❖ Kwan, B. M., Hooper, A. E. C., Magnan, R. E. and Bryan, A. D. (2011). A longitudinal diary study of the effects of causalityorientations on exercise-related affect. Self and Identity, 10 (3), pp.363–374. [Online]. Available at: doi:10.1080/15298868.2010.534238.
- ❖ L1 CRISP-DM 1.0 Step-by-step data mining guide.pdf.
- ❖ Li, Y. and Wang, X. (2018a). Seeking health information on social media: A perspective of trust, self-determination, and social support. Journal of Organizational and End User Computing, 30 (1), pp.1–22. Scopus [Online]. Available at: doi:10.4018/JOEUC.2018010101.
- ❖ Lindwall, M., Weman-Josefsson, K., Sebire, S. J. and Standage, M. (2016). Viewing exercise goal content through a person-oriented lens: A self-determination perspective. Psychology of Sport and Exercise, 27, pp.85–92. [Online]. Available at: doi:10.1016/j.psychsport.2016.06.011.
- ❖ Linoff, G., Berry, M., 2011. Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management. Wiley, John & Sons, Incorporated.
- ❖ Lonsdale, C., Owen, K., Smith, J. and Lubans, D. (2013). Motivation and physical activity in children and adolescents: A systematic review and meta-analysis of evidence from studies framed by self-determination theory. Journal of Science and Medicine in Sport, 16, p.e80. [Online]. Available at: doi:10.1016/j.jsams.2013.10.192.
- ❖ Lugmayr, A., Stockleben, B., Scheib, C. and Mailaparampil, M. A. (2017). Cognitive Big Data: survey and review on Big Data research and its implications. What is really “new” in Big Data? Journal of Knowledge Management, 21 (1), pp.197–212. [Online]. Available at: doi:10.1108/JKM-07-2016-0307.

- Lutz, R. S., Karoly, P. and Okun, M. A. (2008). The why and the how of goal pursuit: Self-determination, goal process cognition, and participation in physical exercise. Psychology of Sport and Exercise, 9 (5), pp.559–575. [Online]. Available at: doi:10.1016/j.psychsport.2007.07.006.
- Mallett, C., Kawabata, M. and Newcombe, P. (2007). Progressing measurement in sport motivation with the SMS-6: A response to Pelletier, Vallerand, and Sarrazin. Psychology of Sport and Exercise, 8 (5), pp.622–631. [Online]. Available at: doi:10.1016/j.psychsport.2007.05.001.
- Mallett, C., Kawabata, M., Newcombe, P., Otero-Forero, A. and Jackson, S. (2007). Sport motivation scale-6 (SMS-6): A revised six-factor sport motivation scale. Psychology of Sport and Exercise, 8 (5), pp.600–614. [Online]. Available at: doi:10.1016/j.psychsport.2006.12.005.
- Manzoor, A., Mollee, J. S., Aráujo, E. F. M., Halteren, A. T. V. and Klein, M. C. A. (2016). Online Sharing of Physical Activity: Does It Accelerate the Impact of a Health Promotion Program? In: 2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom) (BDCloud-SocialCom-SustainCom). October 2016. pp.201–208. [Online]. Available at: doi:10.1109/BDCloud-SocialCom-SustainCom.2016.40.
- Mayer-Schönberger, V. (2009). Delete: the virtue of forgetting in the digital age. Princeton : Princeton University Press.
- Mayer-Schönberger, V. and Cukier, K. (2013). Big data: a revolution that will transform how we live, work, and think. Houghton Mifflin Harcourt.
- Miquelon, P. and Vallerand, R. J. (2008). Goal motives, well-being, and physical health: An integrative model. Canadian Psychology/Psychologie canadienne, 49 (3), pp.241–249. [Online]. Available at: doi:10.1037/a0012759.
- Moreno, J. A., González-Cutre, D., Sicilia, Á. and Spray, C. M. (2010). Motivation in the exercise setting: Integrating constructs from the approach–avoidance achievement goal framework and self-determination theory. Psychology of Sport and Exercise, 11 (6), pp.542–550. [Online]. Available at: doi:10.1016/j.psychsport.2010.06.003.
- Morris, R. (1994). Computerized content analysis in management research: A demonstration of advantages & limitations. Journal of Management, 20 (4), pp.903–931.
- msnbc.com. 2018. More people search for health online | NBC News. [ONLINE] Available at: http://www.nbcnews.com/id/3077086/t/more-people-search-health-online/#.WqunsujwaM8. [Accessed 25 March 2018].
- Mukkamala and Vatrapu, (2016) “Multidimensional-Text-Classification-TR.Pdf,”
- Mukkamala, R. R., Hussain, A. and Vatrapu, R. (2014). “Towards a Set Theoretical Approach to Big Data Analytics”in Big Data (BigData Congress), 2014 IEEE International Congress. Copenhagen, Denmark
- Mukkamala, R. R., Hussain, A. and Vatrapu, R. (2014). Fuzzy-Set Based Sentiment Analysis of Big Social Data. In: September 2014. IEEE. pp.71–80. [Online]. Available at: doi:10.1109/EDOC.2014.19 [Accessed 20 March 2018].
- Mukkamala, R. R., Hussain, A. and Vatrapu, R. (2014b). Towards a Set Theoretical Approach to Big Data Analytics. In: June 2014. IEEE. pp.629–636. [Online]. Available at: doi:10.1109/BigData.Congress.2014.96 [Accessed 20 March 2018].

- Jih-Hsuan Lin, (2015) Need for relatedness: a self-determination approach to examining attachment styles, Facebook use, and psychological well-being: Asian Journal of Communication: Vol 26, No 2. [Online]. Available at: https://www-tandfonline-com.esc-web.lib.cbs.dk:8443/doi/full/10.1080/01292986.2015.1126749?scroll=top&needAccess=true [Accessed 27 February 2018d].
- O'Connor, B., Balasubramanyan, R., Routledge, B. R. and Smith, N. A. From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series. p.8.
- Pang, B. and Lee, L. Opinion mining and sentiment analysis. p.94.
- Pelletier, L. G., Vallerand, R. J. and Sarrazin, P. (2007). The revised six-factor Sport Motivation Scale (Mallett, Kawabata, Newcombe, Otero-Forero, & Jackson, 2007): Something old, something new, and something borrowed. Psychology of Sport and Exercise, 8 (5), pp.615–621. [Online]. Available at: doi:10.1016/j.psychsport.2007.03.006.
- Reichert, J.-R., Kristensen, K. L., Mukkamala, R. R. and Vatrapu, R. (2017). A supervised machine learning study of online discussion forums about type-2 diabetes. In: October 2017. IEEE. pp.1–7. [Online]. Available at: doi:10.1109/HealthCom.2017.8210815 [Accessed 10 April 2018].
- Rodgers, W. M., Hall, C. R., Duncan, L. R., Pearson, E. and Milne, M. I. (2010). Becoming a regular exerciser: Examining change in behavioural regulations among exercise initiates. Psychology of Sport and Exercise, 11 (5), pp.378–386. [Online]. Available at: doi:10.1016/j.psychsport.2010.04.007.
- Ryan, R. M. and Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American Psychologist, 55 (1), pp.68–78. [Online]. Available at: doi:10.1037/0003-066X.55.1.68.
- Ryan, R. M. and Deci, E. L. (2002). Overview of Self-Determination Theory: An Organismic Dialectical Perspective. EL Deci, RM Ryan. Handbook of Self-determination Research.
- Ryan, R. M. and Patrick, H. (2009). Self-determination theory: the dynamics of motivation in development and wellness. Hellenic journal of psychology, 6, pp.107–124.
- Saunders, M. N. K., Lewis, P. and Thornhill, A. (2009). Research methods for business students. 5th ed. New York : Prentice Hall.
- Saxena, R. N. and Srinivasan, A. (2013). Business analytics: a practitioner's guide, International series in operations research & management science volume 186. New York : Springer-Verlag New York.
- Schalock, R. L. (2004). Quality of Life from a Motivational Perspective. In: International Review of Research in Mental Retardation. Personality and Motivational Systems in Mental Retardation. 28. Academic Press. pp.303–319. [Online]. Available at: doi:10.1016/S0074-7750(04)28009-0 [Accessed 9 February 2018].
- Schneider, M. L. and Kwan, B. M. (2013). Psychological need satisfaction, intrinsic motivation and affective response to exercise in adolescents. Psychology of Sport and Exercise, 14 (5), pp.776–785. [Online]. Available at: doi:10.1016/j.psychsport.2013.04.005.
- scikit-learn 0.19.1 documentation. 2018. 1.5. Stochastic Gradient Descent — scikit-learn 0.19.1 documentation. [ONLINE] Available at: http://scikit-learn.org/stable/modules/sgd.html. [Accessed 04 May 2018].

❖ Sebire, S. J., Standage, M. and Vansteenkiste, M. (2009). Examining intrinsic versus extrinsic exercise goals: Cognitive, affective, and behavioral outcomes. Journal of Sport and exercise Psychology, 31 (2), pp.189–210.

❖ Silva, M. N., Markland, D., Minderico, C. S., Vieira, P. N., Castro, M. M., Coutinho, S. R., Santos, T. C., Matos, M. G., Sardinha, L. B. and Teixeira, P. J. (2008). A randomized controlled trial to evaluate self-determination theory for exercise adherence and weight control: rationale and intervention description. BMC Public Health, 8, p.234. [Online]. Available at: doi:10.1186/1471-2458-8-234.

❖ Silva, M. N., Markland, D., Vieira, P. N., Coutinho, S. R., Carraça, E. V., Palmeira, A. L., Minderico, C. S., Matos, M. G., Sardinha, L. B. and Teixeira, P. J. (2010a). Helping overweight women become more active: Need support and motivational regulations for different forms of physical activity. Psychology of Sport and Exercise, 11 (6), pp.591–601. [Online]. Available at: doi:10.1016/j.psychsport.2010.06.011.

❖ Silva, M. N., Vieira, P. N., Coutinho, S. R., Minderico, C. S., Matos, M. G., Sardinha, L. B. and Teixeira, P. J. (2010b). Using self-determination theory to promote physical activity and weight control: a randomized controlled trial in women. Journal of Behavioral Medicine, 33 (2), pp.110–122. [Online]. Available at: doi:10.1007/s10865-009-9239-y.

❖ Simoneau, H. and Bergeron, J. (2003). Factors affecting motivation during the first six weeks of treatment. Addictive Behaviors, 28 (7), pp.1219–1241. [Online]. Available at: doi:10.1016/S0306-4603(02)00257-5.

❖ SODATO . 2018. SODATO . [ONLINE] Available at: http://cssl.cbs.dk/software/sodato/. [Accessed 02 May 2018].

❖ Special Eurobarometer 334: Sport and Physical Activity - ecodp.common.ckan.site_title. [Online]. Available at: https://data.europa.eu/euodp/data/dataset/S776_72_3_EBS334 [Accessed 27 February 2018e].

❖ Special Eurobarometer 412: Sport and physical activity - ecodp.common.ckan.site_title. [Online]. Available at: https://data.europa.eu/euodp/data/dataset/S1116_80_2_412 [Accessed 27 February 2018f].

❖ Statista. 2018. Number of internet users worldwide 2005-2017 | Statista. [ONLINE] Available at: https://www.statista.com/statistics/273018/number-of-internet-users-worldwide/. [Accessed 21 March 2018].

❖ Stubbs, E 2014, Big Data, Big Innovation : Enabling Competitive Differentiation through Business Analytics, Wiley, Somerset. Available from: ProQuest Ebook Central.

❖ Sun, M. (2017). Study on the Construction and Positioning of Sports Goals for College Students based on Big Data Analysis. Boletín Técnico, ISSN: 0376-723X, 55 (18).

❖ Sylvester, B. D., Curran, T., Standage, M., Sabiston, C. M. and Beauchamp, M. R. (2018). Predicting exercise motivation and exercise behavior: A moderated mediation model testing the interaction between perceived exercise variety and basic psychological needs satisfaction. Psychology of Sport and Exercise, 36, pp.50–56. [Online]. Available at: doi:10.1016/j.psychsport.2018.01.004.

❖ Saavedra, J. M., García-Hermoso, A., Escalante, Y. and Domínguez, A. M. (2014). Self-determined motivation, physical exercise and diet in obese children: A three-year follow-up study. International Journal of Clinical and Health Psychology, 14 (3), pp.195–201. [Online]. Available at: doi:10.1016/j.ijchp.2013.12.001.

❖ Taboada, M., Brooke, J., Tofiloski, M., Voll, K. and Stede, M. (2011). Lexicon-Based Methods for Sentiment Analysis. Computational Linguistics, 37 (2), pp.267–307. [Online]. Available at: doi:10.1162/COLI_a_00049.

❖ Teixeira, P. J., Carraça, E. V., Markland, D., Silva, M. N. and Ryan, R. M. (2012). Exercise, physical activity, and self-determination theory: a systematic review. International Journal of Behavioral Nutrition and Physical Activity, 9 (1), p.78.

❖ Thomas, M., Pang, B. and Lee, L. (2006). Get out the vote: determining support or opposition from congressional floor-debate transcripts. In: 2006. Association for Computational Linguistics. p.327. [Online]. Available at: doi:10.3115/1610075.1610122 [Accessed 30 March 2018].

❖ TNS Opinion and Social. (2014). Sport and physical activity: special Euro barometer 412. Brussels : European Commission.

❖ Tsai, T.-H., Chang, H.-T., Lin, Y.-W., Hu, Y.-H., Chen, C.-W., Chen, Y.-C. and Wu, W.-H. (2016). Running on the gatherun cloud platform: Using self-determination theory to increase motivation to participate in sporting events. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 9737, pp.223–230. Scopus [Online]. Available at: doi:10.1007/978-3-319-40250-5_22.

❖ Vansteenkiste, M., Matos, L., Lens, W. and Soenens, B. (2007). Understanding the impact of intrinsic versus extrinsic goal framing on exercise performance: The conflicting role of task and ego involvement. Psychology of Sport and Exercise, 8 (5), pp.771–794. [Online]. Available at: doi:10.1016/j.psychsport.2006.04.006.

❖ Vatrapu, R. (2013). Understanding Social Business. In: Akhilesh, K. B. (Ed). Emerging Dimensions of Technology Management. India : Springer India. pp.147–158. [Online]. Available at: doi:10.1007/978-81-322-0792-4_11 [Accessed 20 March 2018].

❖ Vatrapu, R., Mukkamala, R. R., Hussain, A. and Flesch, B. (2016). Social Set Analysis: A Set Theoretical Approach to Big Data Analytics. IEEE Access, 4, pp.2542–2571. [Online]. Available at: doi:10.1109/ACCESS.2016.2559584.

❖ Vatrapu, Ravi. “Understanding Social Business.” In Emerging Dimensions of Technology Management, edited by K.B. Akhilesh, 147–58. India: Springer India, 2013. https://doi.org/10.1007/978-81-322-0792-4_11.

❖ Vlachopoulos, S. P., Kaperoni, M. and Moustaka, F. C. (2011). The relationship of self-determination theory variables to exercise identity. Psychology of Sport and Exercise, 12 (3), pp.265–272. [Online]. Available at: doi:10.1016/j.psychsport.2010.11.006.

❖ Wang, T., Yeh, R. K.-J., Chen, C. and Tsydypov, Z. (2016). What drives electronic word-of-mouth on social networking sites? Perspectives of social capital and self-determination. Telematics and Informatics, 33 (4), pp.1034–1047. [Online]. Available at: doi:10.1016/j.tele.2016.03.005.

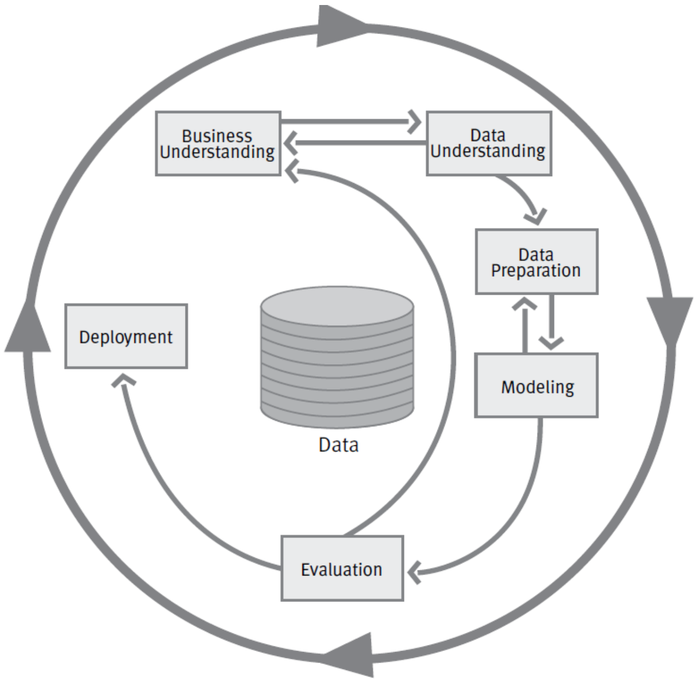
❖ Wang, X. and Li, Y. (2014). Trust, psychological need, and motivation to produce user-generated content: A self-determination perspective. Journal of Electronic Commerce Research, 15 (3), pp.241–253. Scopus.

❖ Wang, Y. and Hajli, N. (2017). Exploring the path to big data analytics success in healthcare. Journal of Business Research, 70, pp.287–299. [Online]. Available at: doi:10.1016/j.jbusres.2016.08.002.

- Wilson, P. M. and Rodgers, W. M. (2004). The relationship between perceived autonomy support, exercise regulations and behavioral intentions in women. Psychology of Sport and Exercise, 5 (3), pp.229–242. [Online]. Available at: doi:10.1016/S1469-0292(03)00003-7.
- Wilson, P. M., Mack, D. E. and Grattan, K. P. (2008). Understanding motivation for exercise: A self-determination theory perspective. Canadian Psychology,49.
- Wilson, P. M., Rogers, W. T., Rodgers, W. M. and Wild, T. C. (2006a). The psychological need satisfaction in exercise scale. Journal of Sport and Exercise Psychology, 28 (3), pp.231–251.
- Wilson, P., Rodgers, W., M. Blanchard, C. and Gessell, J. (2006b). The Relationship Between Psychological Needs, Self-Determined Motivation, Exercise Attitudes, and Physical Fitness1. Journal of Applied Social Psychology, 33, pp.2373–2392. [Online]. Available at: doi:10.1111/j.1559-1816.2003.tb01890.x.
- Wininger, S. R. (2007). Self-Determination Theory and Exercise Behavior: An Examination of the Psychometric Properties of the Exercise Motivation Scale. Journal of Applied Sport Psychology, 19 (4), pp.471–486. [Online]. Available at: doi:10.1080/10413200701601466.
- World population - Wikipedia. 2018. World population - Wikipedia. [ONLINE] Available at: https://en.wikipedia.org/wiki/World_population. [Accessed 21 March 2018].
- Yedidia, Adam (2016) "Against the F-score"

9. Appendix

9.1 CRISP-DM - data life cycle (Chapman et al., 2000)



9.2 Data Source Description

Dindebat.dk - forum

It is a platform where users can read news, have email account, play games, chat and debate. The forum consists of a wide range of topic, separated into different categories. The category used in this study was “Lifestyle/ Health/ Training and motion”. Under it there were more than 3200 different sub-topics at the time of fetching the data. The platform does not have a dedicated Facebook public page for more discussions.

The fetched dataset contained a total of records 24517 records, which were divided into 3200 unique topic categories, and written by 2780 unique users. Data was created in the period between June 2004 and November 2017.

Musclezone.dk - forum

It is a forum for discussing training, exciting articles, inspiration and help with diet and training. The forum is specializing in fitness and it is helping with tips for getting in a better shape by trying harder. The forum provided a small statistics insights, and on the time of collection there were a total of 103791 posts, a total 8722 topics, and a total of 9193 registered members.

The fetched dataset contained a total of 1908 records, which were divided into 960 unique topic categories and 356 unique users. Data was created in the period between May 2012 and February 2018.

Motion-online.dk - forum

It is an online forum created only for discussion of healthy lifestyle, training, diet and sports. There is unfortunately no information and history about the forum. The topics are divided into 11 different categories, which had a total of 819522 comments at the time of data collection. As described on its Facebook page, the forum is "Denmark's biggest forum for discussion of training and health".

The fetched dataset contained a total of records 32240 records, which were divided into 2804 unique topic categories and created by 3145 unique users. Data represents the period between July 2002 and April 2018.

MotionOnlinedk - Facebook page

It is the social media page of the above mentioned forum. In its description, it says that it is mainly for discussing how to become stronger, healthier, faster, lose weight or just become more informed about these topics. At the time of the collection, the page had a total of 4519 Likes and total of 4389 follows. No information about when the page was founded.

The fetched dataset contained a total of records 200 records, and represents the period between January 2001 and April 2018.

Træning og Motivation - Facebook page

It is a Facebook public page for the purpose of discussing training and motivation, bodybuilding, fitness and sports. At the time of the collection, the page had a total of 9916 Likes and total of 9744 follows. The page was founded in January 2012. The fetched dataset contained a total of 1434 records, created in the period between January 2013 Marts 2018.

Træning og kost - Facebook page

It is a Facebook public page, which discusses training and diet as main topics. It als also a place where people would find information about training tips and good diet advice, which can inspire and help others to achieve the best possible results. At the time of the collection, the page had a total of 3614 Likes and a total of 3580 follows. The page was founded in 2013.

The fetched dataset contained a total of records 323 records, representing the period between February 2013 and January 2016.

Abilica Online - Træning med mening - Facebook page

It is a Facebook public page which discusses weekly motivation and inspiration for “training for a purpose”. It is focused on training inspiration in the busy everyday. At the time of the collection, the page had a total of 5863 Likes and a total of 5806 follows. The page was founded in 2007. It has additionally a online webshop which sells sports equipment.

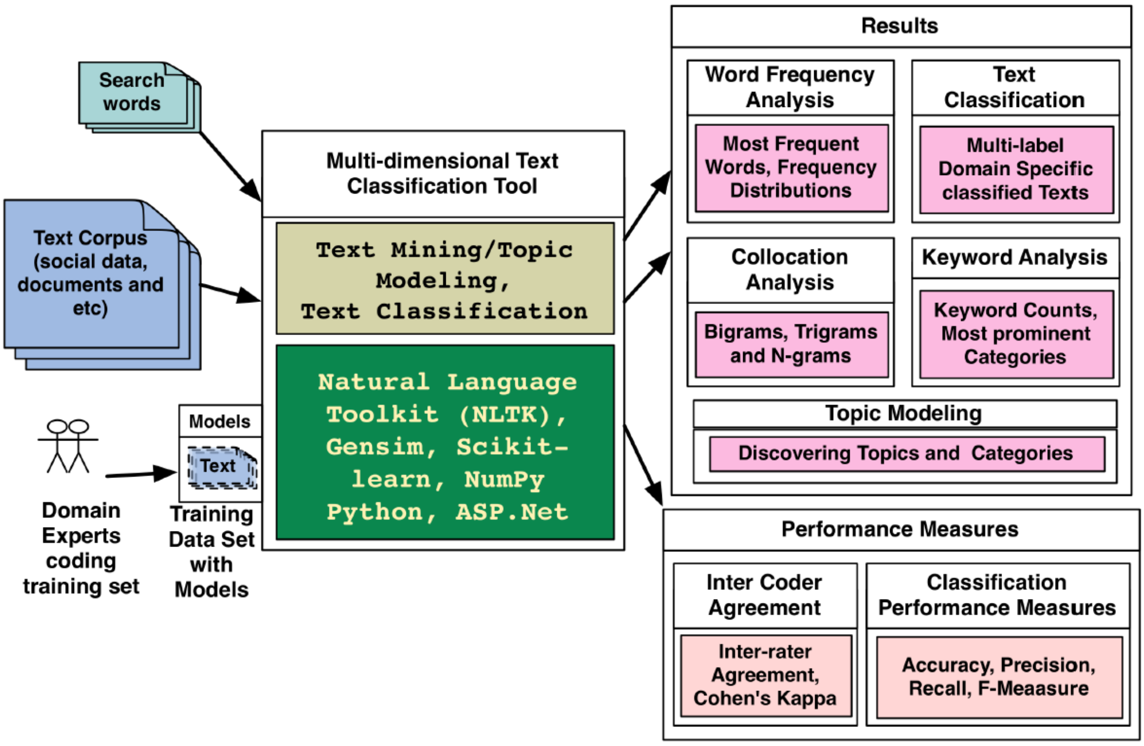
The fetched dataset contained a total of 480 records, which were created in the period between October 2008 until February 2018.

Fitnessdk - Facebook page

Fitnessdk is one of the competing fitness centers around the whole of Denmark. This is their social media Facebook public page. It is aimed at providing training which is fun and challenging, and giving the users the best experience. At the time of the collection, the page had a total of 47332 Likes and a total of 45346 follows. The page was founded in 2001. The fetched dataset contained a total of 1121 records, created the period between October 2008 and February 2018.

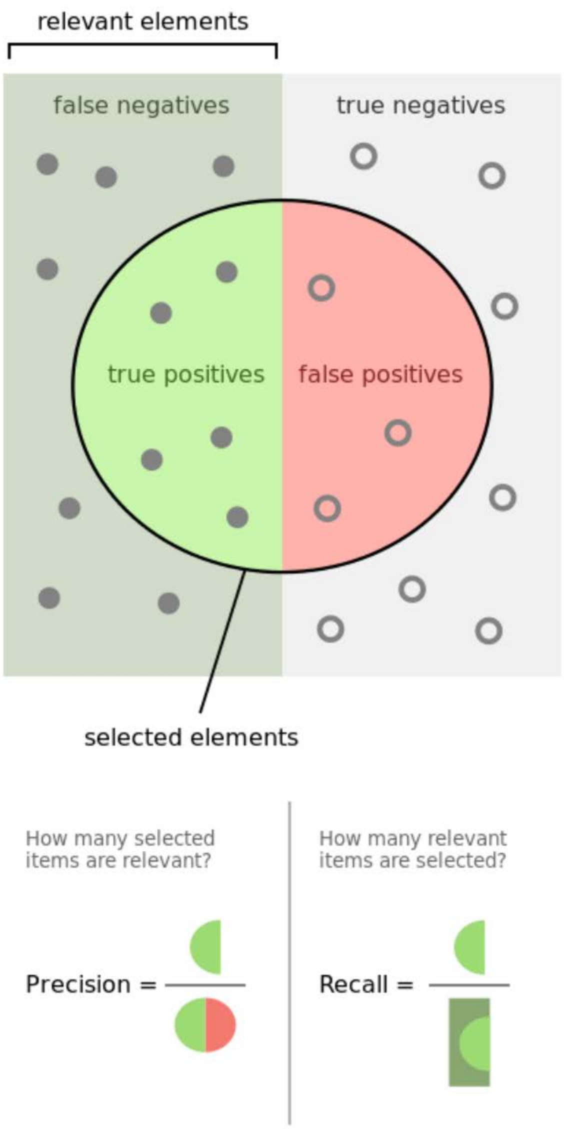
9.3 MUTATO Architecture

(Mukkamala and Vatrapu, 2016)



9.4 Detailed Confusion Matrix

F1 - score, Precision and Recall (Yedidia, 2016).



9.5 Total Basic Needs Model Performance

Model 1: Basic needs			
NaiveBayesMultinomialClassifier			
Category	Precision	Recall	F1-score
Autonomy	0.63	0.73	0.68
Competence	0.62	0.74	0.76
Relatedness	0.80	0.74	0.76
None	0.97	0.82	0.89
LinearSVCClassifier			
Category	Precision	Recall	F1-score
Autonomy	0.71	0.73	0.72
Competence	0.66	0.85	0.74
Relatedness	0.83	0.70	0.76
None	0.98	0.91	0.94
LogisticRegressionClassifier			
Category	Precision	Recall	F1-score
Autonomy	0.64	0.62	0.63
Competence	0.62	0.81	0.70
Relatedness	0.84	0.48	0.61
None	0.84	0.90	0.87
PassiveAggressiveClassifier			
Category	Precision	Recall	F1-score
Autonomy	0.68	0.76	0.71
Competence	0.69	0.81	0.75
Relatedness	0.88	0.71	0.79
None	0.98	0.94	0.96
SVMSGDClassifier			
Category	Precision	Recall	F1-score
Autonomy	0.59	0.66	0.62
Competence	0.65	0.79	0.71
Relatedness	0.86	0.50	0.63
None	0.86	0.89	0.88

9.6 Total Goals Model Performance

Model 2: Goals			
NaiveBayesMultinomialClassifier			
Category	Precision	Recall	F1-score
Intrinsic	0.88	0.94	0.91
Extrinsic	0.91	0.95	0.93
None	0.96	0.93	0.95
LinearSVCClassifier			
Category	Precision	Recall	F1-score
Intrinsic	0.94	1.00	0.97
Extrinsic	0.96	1.00	0.98
None	1.00	0.96	0.98
LogisticRegressionClassifier			
Category	Precision	Recall	F1-score
Intrinsic	0.90	0.74	0.81
Extrinsic	0.97	0.58	0.73
None	0.81	0.97	0.88
PassiveAggressiveClassifier			
Category	Precision	Recall	F1-score
Intrinsic	0.94	1.00	0.97
Extrinsic	0.95	1.00	0.97
None	1.00	0.96	0.98
SVMSGDClassifier			
Category	Precision	Recall	F1-score
Intrinsic	0.90	0.87	0.89
Extrinsic	0.95	0.81	0.87
None	0.90	0.95	0.93

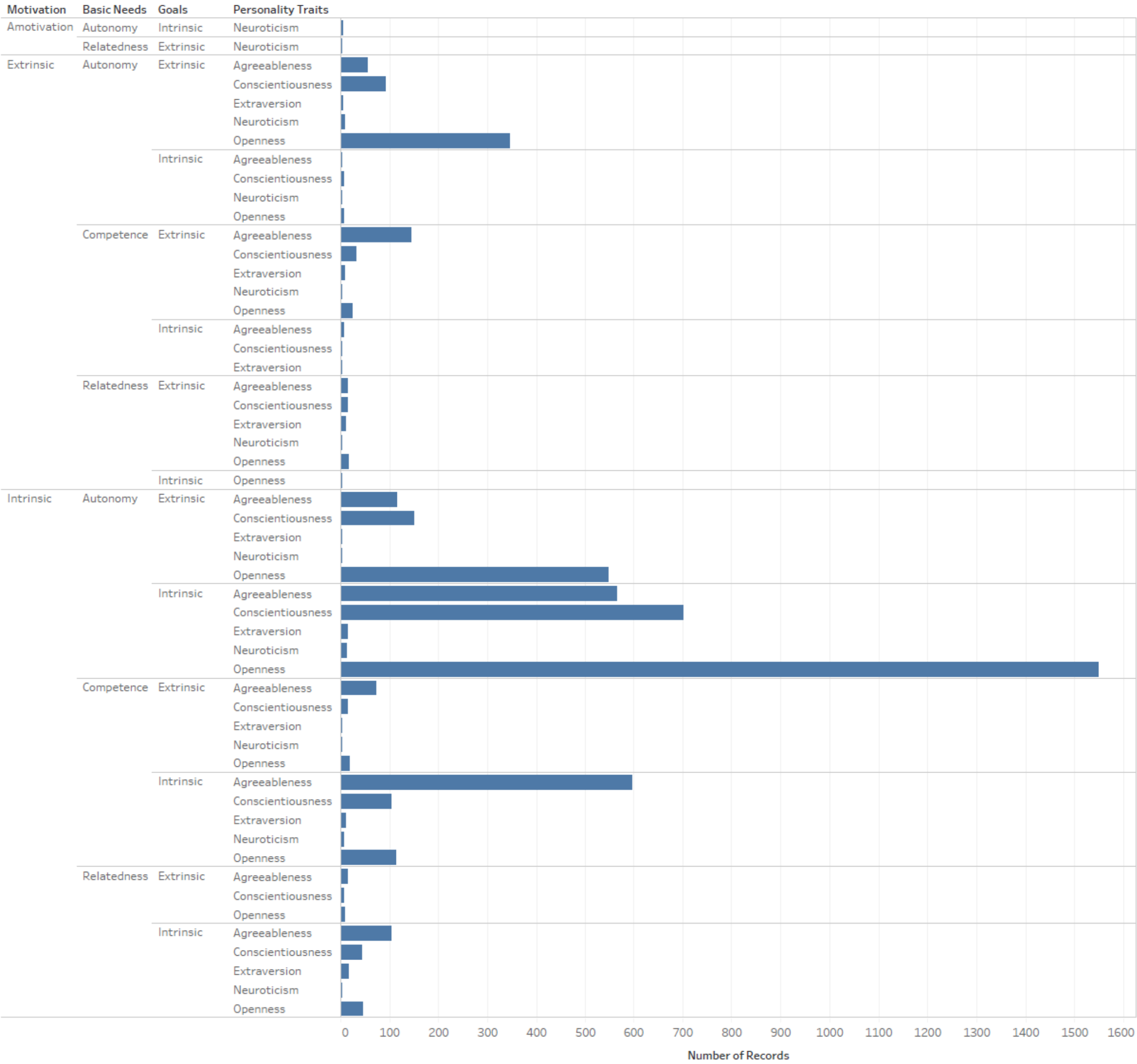
9.7 Total Personality Traits Model Performance

Model 4: Personality Traits			
NaiveBayesMultinomialClassifier			
Category	Precision	Recall	F1-score
Openness	0.76	0.68	0.72
Conscientiousness	0.95	0.81	0.88
Extraversion	0.95	1.00	0,98
Agreeableness	0.88	0.88	0.88
Neuroticism	1.00	1.00	1.00
None	0.90	0.94	0.92
LinearSVCClassifier			
Category	Precision	Recall	F1-score
Openness	0.85	0.83	0.84
Conscientiousness	0.97	0.88	0.92
Extraversion	1.00	1.00	1.00
Agreeableness	0.87	1.00	0.93
Neuroticism	1.00	1.00	1.00
None	0.96	0.96	0.96
LogisticRegressionClassifier			
Category	Precision	Recall	F1-score
Openness	0.56	0.19	0.29
Conscientiousness	1.00	0.29	0.45
Extraversion	1.00	0.45	0.62
Agreeableness	0.72	0,70	0,71
Neuroticism	1.00	0.66	0.79
None	0.70	0.97	0.81
PassiveAggressiveClassifier			
Category	Precision	Recall	F1-score
Openness	0.80	0.83	0.81
Conscientiousness	0.99	0.88	0.93
Extraversion	0.98	1.00	0.99
Agreeableness	0.88	1.00	0.94
Neuroticism	1.00	1.00	1.00
None	0.96	0.95	0.96
SVMSGDClassifier			
Category	Precision	Recall	F1-score
Openness	0.69	0.43	0.53
Conscientiousness	0.93	0.53	0.68
Extraversion	0.97	0.93	0.95
Agreeableness	0.86	0.94	0.90
Neuroticism	1.00	1.00	1.00
None	0.83	0.95	0.88

9.8 Total Types of Motivation Model Performance

Model 3 Types of Motivation			
NaiveBayesMultinomialClassifier			
Category	Precision	Recall	F1-score
Intrinsic	0.85	0.97	0.91
Extrinsic	0.94	0.98	0.96
Amotivation	1.00	0.84	0.91
None	0.98	0.93	0.95
LinearSVCClassifier			
Category	Precision	Recall	F1-score
Intrinsic	0.87	1.00	0.93
Extrinsic	0.98	1.00	0.99
Amotivation	1.00	1.00	1.00
None	1.00	0.94	0.97
LogisticRegressionClassifier			
Category	Precision	Recall	F1-score
Intrinsic	0.76	0.62	0.68
Extrinsic	1.00	0.44	0.61
Amotivation	1.00	0.26	0.42
None	0.73	0.96	0.83
PassiveAggressiveClassifier			
Category	Precision	Recall	F1-score
Intrinsic	0.85	1.00	0.92
Extrinsic	0.96	1.00	0.98
Amotivation	1.00	1.00	1.00
None	1.00	0.93	0.96
SVMSGDClassifier			
Category	Precision	Recall	F1-score
Intrinsic	0.82	0.71	0.76
Extrinsic	0.82	0.71	0.76
Amotivation	1.00	0.84	0.91
None	0.86	0.95	0.90

Motivation



Sum of Number of Records for each Personality Traits broken down by Motivation, Basic Needs and Goals.

9.10 Distribution of Categories Across Datasets

Category	Percentage	Din Debat	Muscle Zone	Motion Online - forum	Motion Online - FB	Abilica	Fitness DK	Træning og Motion	Træning og Kost
Model 1: Basic needs									
Autonomy	15.91%	4461	139	4114	5	3	18	45	8
Competence	10.50%	2151	132	3418	31	12	29	25	6
Relatedness	6.11%	483	38	502	162	445	134	1308	306
None	67.47%	14222	639	21402	2	20	940	56	3
Total Observations	**55259**	**21317**	**948**	**29436**	**200**	**480**	**1121**	**1434**	**323**
Model 2: Goals									
Intrinsic	10.25%	2675	172	2708	13	1	23	61	12
Extrinsic	5.27%	1701	116	985	18	3	30	55	4
None	84.48%	16941	660	25743	169	476	1068	1318	307
Total Observations	**55259**	**21317**	**948**	**29436**	**200**	**480**	**1121**	**1434**	**323**
Model 3: Types of motivation									
Intrinsic	18.39%	4692	216	5141	9	3	26	63	10
Extrinsic	3.03%	1025	86	483	18	0	9	49	5
Amotivation	0.07%	18	11	8	0	0	1	0	0
None	78.51%	15582	635	23804	173	477	1085	1322	308
Total Observations	**55259**	**21317**	**948**	**29436**	**200**	**480**	**1121**	**1434**	**323**
Model 4: Personality Traits									
Openness	8.33%	2367	91	2132	1	0	1	10	1
Conscientiousness	3.30%	976	67	715	7	0	16	40	4
Extraversion	0.22%	31	18	39	4	0	3	20	4
Agreeableness	9.94%	2126	100	3256	5	1	2	3	1
Neuroticism	0.25%	69	13	55	0	0	0	0	0
None	77.96%	15748	659	23239	183	479	1099	1361	313
Total Observations	**55259**	**21317**	**948**	**29436**	**200**	**480**	**1121**	**1434**	**323**