EXAMENSARBETE INOM TEKNIK,
GRUNDNIVÅ, 15 HP
STOCKHOLM, SVERIGE 2020

Machine Learning implementation for Stress-Detection

NICOLE MADJAR

FILIP LINDBLOM

KTH
SKOLAN FÖR ELEKTROTEKNIK OCH DATAVETENSKAP

Machine Learning implementation for Stress-Detection

N. Madjar, F. Lindblom

***Abstract*-This project is about trying to apply machine learning theories on a selection of data points in order to see if an improvement of current methodology within stress detection and measure selecting could be applicable for the company Linkura AB. Linkura AB is a medical technology company based in Linköping and handles among other things stress measuring for different companies employees, as well as health coaching for selecting measures. In this report we experiment with different methods and algorithms under the collective name of Unsupervised Learning, to identify visible patterns and behavior of data points and further on we analyze it with the quantity of data received. The methods that have been practiced on during the project are "K-means algorithm" and a dynamic hierarchical clustering algorithm. The correlation between the different data points parameters is analyzed to optimize the resource consumption, also experiments with different number of parameters are tested and discussed with an expert in stress coaching. The results stated that both algorithms can create clusters for the risk groups, however, the dynamic clustering method clearly demonstrate the optimal number of clusters that should be used. Having consulted with mentors and health coaches regarding the analysis of the produced clusters, a conclusion that the dynamic hierarchical cluster algorithm gives more accurate clusters to represent risk groups were done. The conclusion of this project is that the machine learning algorithms that have been used, can categorize data points with stress behavioral correlations, which is usable in measure testimonials. Further research should be done with a greater set of data for a more optimal result, where this project can form the basis for the implementations.**

***Sammanfattning*-Detta projekt handlar om att försöka applicera maskininlärningsmodeller på ett urval av datapunkter för att ta reda på huruvida en förbättring av nuvarande praxis inom stress-detektering och åtgärdshantering kan vara applicerbart för företaget Linkura AB. Linkura AB är ett medicintekniskt företag baserat i Linköping och hanterar bland annat stressmätning hos andra företags anställda, samt hälso-coachning för att ta fram åtgärdspunkter för förbättring. I denna rapport experimenterar vi med olika metoder under samlingsnamnet oövervakad maskininlärning för att identifiera synbara mönster och beteenden inom datapunkter, och vidare analyseras detta i förhållande till den mängden data vi fått tillgodosett. De modeller som har använts under projektets gang har varit "K-Means algoritm" samt en dynamisk hierarkisk klustermodell. Korrelationen mellan olika datapunktsparametrar analyseras för att optimera resurshantering, samt experimentering med olika antal parametrar inkluderade i datan testas och diskuteras med expertis inom hälso-coachning. Resultaten påvisade att båda algoritmerna kan generera kluster för riskgrupper, men där den dynamiska modellen tydligt påvisar antalet kluster som ska användas för optimalt resultat. Efter konsultering med mentorer samt expertis inom hälso-coachning så drogs en slutsats om att den dynamiska modellen levererar tydligare riskkkluster för att representera riskgrupper för stress. Slutsatsen för projektet blev att maskininlärningsmodeller kan kategorisera datapunkter med stressrelaterade korrelationer, vilket är användbart för åtgärdsbestämmelser. Framtida arbeten bör göras med ett större mängd data för mer optimerade resultat, där detta projekt kan ses som en grund för dessa implementeringar.**

***Index Terms*- BigQuery, Cloud Computing, Clustering, Dynamic clustering algorithm, Google Cloud Platform, Health coaching, HRV-values, K-Means algorithm, Machine learning, Stress, Unsupervised learning.**

I. INTRODUCTION

A. *Linkura*

We will develop a model for the company Linkura AB, a Swedish B2B-focused company situated in Linköping. The organization offers clients a service that calculate stress levels in a tangible format and suggest future measures for the client's employees based on their work situation. The type of service that Linkura is offering is categorized as "*Wellness as a service* " (WaaS), which is a growing niched type of service market as a result of an increased demand of healthier lifestyles.

The service offering is divided into two segments where the first action in the process is using ECG-sensors during a 3-day period in order to gather statistics regarding the user. These statistics include standard ECG-signals as well as sleeping habits and the amount of exercise the person is doing. In the second segment, the employees can see the results through the application/digital coach called Linus, combined with the availability of booking a certified health coach in order to create

Submitted for review 12th June 2020.

F. Lindblom is with the School of Industrial Engineering and Management, Royal Institute of Technology, Stockholm, Sweden (e-mail: filind@kth.se).

N. Madjar is with the School of Engineering Sciences in Chemistry, Biotechnology and Health, Royal Institute of Technology, Stockholm, Sweden (e-mail: madjar@kth.se).

a plan and put together a self-leadership based development[1].

The end goal for the customer having these measurements and health coach meetings, is to create sustainability and productivity by applying a more suitable healthy perspective on everyday routines, often instigated by employers.

B. *Stress*

Work-related stress is nowadays a known factor to depression and is closely connected to the ever-changing work situations worldwide, often caused by external forces like market changes and financial crises. Psychological disorders are impactful on an employees´ personal life and is something to be taken seriously[2]. From a social perspective, the society would gain a lot of benefits by increasing the amount of stress indicators available for employees working within hectic environments.

According to the European Neuropsychopharmacology, stress related illness in Sweden cost the Swedish government approximately 70 billion SEK every year in treatment and indirect by absence from production[3]. By doing this project we could build a stronger basis for detection of stress at an early stage with a better accuracy, which could eventually increase the mental health for individuals and society as a whole.

In order to further develop Linkura's methods, a digitalization and further optimization of the data analyzing could create a more effective process for the value offering from Linkura to its clients, hence why this project was created.

C. *Problem definition*

The project is about improving a stress bot called Linus, used by the company Linkura AB with support of the cloud-based consulting company Devoteam, by using machine learning applied to biodata and subjective data in order to support improved well-being.

The purpose of this project is to use Linkura's extensive database to increase the performance of stress bot Linus to the extent that it can interpret user data history with a higher precision, so it can give adequate measures and advices to the user. The scope for this project in order to increase the performance of the stress bot, is to create a platform for automatization in regards to classifying and analyzing users data to immediate spot patterns or behavior that is stress related and can indicate if or what measures are needed.

This approach would in the long term be a more sustainable and scalable method for Linkura. Since human health coaches often is quite expensive for the clients to hire, a digitalization of the services would increase the possibilities for more companies affording being a client. Further advantages for employers with an increasing healthy lifestyle for the employees are lower absence rate and higher productivity.

D. *Devoteam*

The "Big Data" for this project is processed with the cooperation of the cloud consulting company Devoteam based in Stockholm. Devoteam presents solutions for change management regarding digitalization and cloud-based solutions for other companies. This has made it possible for the data to be imported into GCP(Google Cloud Platform) and BigQuery which will be the main platforms used in the project. Provided by Devoteam is also mentoring, GCP related education and agile framework in order to plan the project in sprints and make sure we achieve our results in time.

We will be focusing on cloud-based queries regarding current status for the user and what suitable actions could be taken. The challenge lies primarily within choosing an adequate machine learning method with tangible evaluable results. Depending on the type of data, some correlations might have to be figured out before trying out different machine learning models. A lot of focus must be put on choosing what type of information should be relevant or necessary. When working with raw data, analysis can differ in results depending on the chosen method, especially within the area of machine learning.

E. *Scientific question*

With this background, a scientific question could be stated regarding the scope of the project. Can we apply machine learning models in order to improve the results of an stress related bot that grants an analysis of the users current status using large databases of medical information like ECG values?

What potential risks might this implementation present in the future? Since we are dealing with a project in the medical segment, risks will always be present and the main reason for lack of change.

When trying to implement this type of solution, what could be some important factors in order for the end users to accept a more digitalized solution instead of the current best practice?

Machine learning is a hot topic today within a large number of sectors of the tech industry and the implementation of different types of algorithms may be able to optimize current standards and methods with better speed and accuracy. Because of the focus on applying machine learning, the project is relevant to current scientific research. We can observe more and more works and studies about machine learning and the span of possibilities when applied to everyday situations and problems.

F. *Expected results*

The expected result of this project according to the scope is a model that provides a solution to a more optimized chatbot that has a more solid basis for partially or fully replacing the human health coaches. Since we are doing an unsupervised learning approach, analytic discussions regarding the resulting clusters will be of significance for us to see whether the model

[1] Så fungerar Linkura, [Online], Available: https://www.linkura.se/about/sa-fungerar-linkura

[2] Tennant, C. (2001). Work-related stress and depressive disorders. *Journal of psychosomatic research*, 51(5), 697-704.

[3] Gustavsson, A., …& Olesen, J. (2011). Cost of Disorders of the Brain in Europe 2010. *European neuropsychopharmacology*, *21*(10), 718-779

is applicable. Based on previous works and the theories of machine learning, we expect to achieve a model that can push current practices in the right direction. The end goal is to have a digitalized and optimized platform where the user could get results and suggested measures in an instant by the chatbot going through history and statistics regarding the user.

II. Theories

A. HRV

An important part of the data being analyzed in this project is HRV-data granted by sensors worn by the patients. In order to know what to look for when implementing machine learning models, an understanding of what these HRV-values actually mean is needed.

HRV stand for *Heart Rate Variability* and is the fluctuation of heart rate in time intervals between adjacent heartbeats. The heart rate is the amount of heartbeats per minute. The main reason that this metric is used is because of the human heart being complex and non-linear when we look at the oscillations. Even if a heart is healthy, we get huge variation, Shaffer and Ginsberg describe this phenomena as “Mathematical chaos”. Often, a non-linear system’s variability indicates the adaptability to quickly cope with uncertain and changing environments. Diseases within complex healthy biological systems sometimes involves either a loss or an increase of complexity. Higher HRV-values tends to be correlated with a healthier situation but not always since they can be initiated by for example pathological conditions that is strongly linked to increased mortality.

The HRV-values provided in this project follows the standard of *RMSSD*, also known as the root mean square of successive differences between normal heartbeats. This is done by calculating each successive time difference in between heartbeats, measured in milliseconds. This is followed up by squaring all of the values and then averaging the result before taking the square root of the total[4].

B. Machine Learning

In order to apply a machine learning model in the project, a clear definition of what is expected from the end product is necessary for planning the project structure. Thrun and Pratt defines what it means for an algorithm to be able to learn how to learn, with a number of given conditions. First of we have a family of tasks, as well as training experience of these tasks and at last a family of performance measures. They state that “an algorithm is said to learn to learn if its performances at each task improves with experience and with the number of tasks”[5].

Machine learning can often be seen as a solution to real world scenarios where multiple learning problems occur parallel to each other. Thrun and Pratt mentions examples of service robots with an initial purpose of fetching objects also could be trained to do perceptual tasks like recognizing objects, landmarks or even people[5]. This could be seen as an argument of the increasing interest of Machine Learning within a lot of markets and sectors due to multi purposing and optimization of systems and products.

C. Unsupervised/Supervised Machine Learning

The most common practice of machine learning is through Supervised learning. That is when you have an input variable, an output variable and you use an algorithm to map out the correlating function in between. In other words, you have approximated the mapping function in such a great degree that you can predict the output variable just by using an input variable. Often, we talk about having data points with a certain number of features or parameters, combined with labeling features that decides if a data points belong to a certain classification or not.

The data is distributed into training, testing and validation data in order to compare the results with already present and labeled data. This ensures that comparing ratios like accuracy, precision, recall etc. can be used as evaluation metrics. The opposite to this is called unsupervised machine learning, the difference within the data is that the data points are not labeled with a correct classification. This means that the user does not know at first sight where a data point should belong.

Unsupervised learning focuses on spotting patterns and similar behavior between groups of data points. This is done by different sets of algorithms that spots similarities, for example clustering. *Clustering* is when you want to gather the constitutional groups of a dataset that shares different similarities[6].

D. Google Cloud Platform

Within our project, we will analyze data of users being tracked by sensors. We will have access to over 2000 days of data divided into 3-days segments for each data point and tracking. This means that we must consider iterating over large amounts of data, making it require structure and fast computing.

In order to cope with these large amounts of data(Big Data), a larger data ecosystem is often a good approach when doing projects including machine learning since it often requires a hefty amount of data for the programs to learn the right way. Google Cloud Platform is such a data ecosystem developed by Google, that includes several types of offerings such as processing power, storage and analytics. First off, the cloud processing presents a computing engine to the user in form of virtual engines. These virtual engines can be used to run software like testing machine learning algorithms on data. Within the cloud storage, BigQuery is a storage API that enables the user to store data in structured rows and columns of data[7].

Cloud analytics is a department where machine learning models can be trained in order for them to be applied to the user data. BigQuery can also be used within the analytics department by running SQL statements over the user data. For example,

[4] Shaffer, F., & Ginsberg, J.P. (2017). An Overview of Heart Rate Variability Metric and Norms. *Frontiers in public health*, *5*, 258.

[5] Thrun, S., Pratt, L. (2012) Learning to Learn. *Springer Science & Business Media*, 4-12.

[6] Brownlee, J. Supervised and unsupervised machine learning algorithms, [Online], *Available:* https://machinelearningmastery.com/supervised-and-unsuper vised-machine-learning-algorithms/

[7] Tigani, A., Naidu, S. (2014) *BigQuery Fundamentals Analytics*. John Wiley & Sons Inc. Indianapolis, Indiana, 8-14.

large JSON-files can be uploaded into a BigQuery table in order to structure it into a user-friendly viewable format. These tables also support running code in the terminal provided, allowing for example SQL-statements to be run in order to analyze and perform computations on the data within the table.

There is also built in machine learning techniques, ready to be applied onto the data. By running SQL-statements, machine learning models can be created and saved in the BigQuery storage for future recreatable tests. When using BigQuery, SQL queries is being run on a lot of disks parallel which grants the ability of granting answers to complex questions with a quick response time[7].

BigQuery utilizes the ability to run queries on a *computing cluster*, while pipelining the queries in order to minimize the amount of time spent on the computing compared to for example using a single core locally.

E. Tensorflow

Since Google Cloud Platform is a suitable solution for heavy computing on larger databases and the abilities and the efficiency of storing “Big Data” within Google's BigQuery system, a link between the user software to the GCP would be needed for continuous testing and development. Tensorflow is a service used by several Google services in order to allow users to perform larger machine learning tasks with for example parallel computing on different disks or clusters. Tensorflow is applicable to several different programming languages and enables the user to run their codes efficient on larger data within the GCP environment like BigQuery[8].

F. K-means algorithm

K-means algorithm is a clustering method to identify *k* numbers of centroids, designate it to a cluster nearby and keeping the centroids small. The centroids represent the imaginary center of a cluster in a vector space. This model initiates *k* -number of centroids in the vector space, and then by calculating distances between points and continuously iterating until requirements or limits are met. The goal is to find underlying arrangements in these clusters of data points in order to spot patterns and behavior that could be used in data driven decisions[9].

G. Hierarchical Clustering Algorithm

Hierarchical clustering is a dynamic clustering method used for classifying data points. By calling it a dynamic model, we say that we are not setting a fixed value for the number of clusters, this is calculated for us within the algorithm. The algorithms work by starting off by treating every single data point as a single separate cluster. Then it repeatedly identifies and merges the two clusters that are closest to each other. This iterative process continues until all the clusters has been merged into one large cluster. How the distances between the different clusters are calculated depends on what type of metric is being used. Examples of different metric could be *Euclidean distance, Manhattan distance* or *Cosine Similarity*.

There are also different options when it comes to selecting a linkage method when calculating distance between data points within clusters. Some examples are *Single-Linkage* where the two most similar points in the clusters are selected to compute the distance, *Complete-Linkage* where the two least similar points are used or *Mean-Linkage* where the center of the centroid are used for calculations. The most common default linkage selection is *Ward's method* that reduces the sum of squared distances of each observation from the mean observation in clusters[10].

Hierarchical clustering can be graphically presented with an output in form of a *Dendrogram*. A *Dendrogram* is a diagram that shows the hierarchical relationships between data points or clusters. The advantage of using a *Dendrogram* is that it makes it easier to find the best way of allocating data points to clusters and the number of clusters to be used. By looking at the diagram, the height of links shows how similar or different two points are to each other. The higher the link, the more different the two data points are.

By looking at the heights of the links and locating the longest vertical lines, the number of clusters can be accurately selected. By drawing a horizontal line through the longest link, you can count the number of vertical lines it crosses and that can indicate how many clusters should be used. This is not a definitive answer of the optimal amount of clusters but could be a pointer in the right direction[11].

III. Previous work

Using machine learning in order to improve medical diagnostic results is a method that has been used in previous projects. Supervised learning with a Convolutional Neural Network has been used in order to achieve a better detection rate for irregular heart rhythms called *arrhythmias*[12]. What the authors did was to firstly build a data set that was significantly larger than previously studied corpora that was to be used by their trained neural network.

The reason for this previous work to be done was to increase the performance of early detection for heart diseases compared to at the time current valuation methods. By using a 34-layer convolutional neural network, the authors were able to classify and segment input into twelve different arrhythmia types. This project used data consisting of ECG signals and corresponding data which is what we will use as data points as well in order to lay a foundation in order to increase the performance of chat bot Linus, hence why this type of project is highly relevant to our research.

In their method, ECG-signals is taken as input in order to

[8] Abadi, M., …& Kudlur, M. (2016). Tensorflow: A system for large Scale Machine Learning. *In 12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI}16)* (pp. 265-283).

[9] M. Garbadi, “*Understanding K-means Clustering in Machine Learning*”, [Online], Available: https://towardsdatascience.com/understanding-k-means-cluster ing-in-machine-learning-6a6e67336aa1

[10] T.Bock, “*What is Hierarchical Clustering?*”, [Online], Available: https://www.displayr.com/what-is-hierarchical-clustering/

[11] T.Bock, “*What is a Dendrogram?*”, [Online], Available: https://www.displayr.com/what-is-dendrogram/

[12] Hannun, A. Y., …& Ng, A. Y. (2019). Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature medicine, 25*(1), 65.

create an output set of labels that can take on one of the different rhythm classes. Later, this output would be used with a cross-entropy objective function in order to calculate probability of assigning the network to these rhythm classes. One main difference between the project is that their data points consists of solely ECG-signals over 30 seconds while our data points is spread over the period of 3-days including data on sleeping habits and exercise.

The end result is for the authors to classify the rhythms in order to recognize patterns for arrhythmia while we combine all the data points to see whether a user has a stress condition where measures are needed at if so, what type of measure?

The authors performed supervised learning, the training data was changed to test data in order to compare the results to human performed labeling and diagnostics. The results for the model clearly outperformed the labeling from actual cardiologists when it came to most of the different types of rhythms, which can be seen in table 1 in the report[12].

However our project will be using unsupervised learning due to the lack of labels in our data, nonetheless the results of the study shows that there is potential for our project to create an increased performance in classification of stress levels and choosing correct measures for the personal user.

When trying to implement a digitalization in order to change a current practice that can be seen as more manual work, studies often show the importance of the acceptance of the solution within the client. Paul j. Hu et al, discuss this matter in their work regarding the *Technology Acceptance Model*(TAM)[13]. In their study, they examine the acceptance of the implementation of Telemedicine among physicians working at hospitals in Hong Kong. The purpose of the study is to show in an empirical way, an understanding for the acceptance among health-care professionals who are becoming increasingly dependent on information technology.

In the study, telemedicine is presented as a supportive and enhancing tool for the current practices within the hospitals, which is comparable to the purpose of our study. The connection to the national and global economies and the health-care sector´s impact in it is the purpose for their study and they also highlight the unique characteristics within the sector.

The empirical study was done with questionnaires being sent to the physicians at the different clinics to be answered, resulting in 421 completed answers. Based on these collected answers, the study presents the acceptance values and goodness-of-fit indexes, which resulted in a conclusion that TAM could show applicable to this type of measurement and implementation. The result showed that perceived usefulness had a strong influence on whether the physician had an intention to use the technology. It also had a high impact on the physicians' attitude towards the implementation. Since the attitude also had an effect on the intent to use, perceived usefulness was therefore a main factor identified for the intention to use the technology. Perceived usefulness combined with attitude towards telemedicine accounted for 44 percent of the variances of the physicians´ intent of using the technology.

On the contrary, perception of easy usability had close to none effect on attitude and perceived usefulness. This is discussed in the study as a possible result of a demographic group that is highly educated and being able to learn new technology quickly.

IV. METHOD

The dataset that have been given to us do not have labels, which means in comparison to the earlier study we only have the input variable and not the output variable compared to supervised learning. Therefore, our method of choice was unsupervised machine learning with the approach of clustering our data in order to spot patterns and behavior.

Combining the theories discussed with the information and instructions gained by meeting and with companies Linkura AB and Devoteam our methodology was structured into different subtasks such as following:

First step was to upload the data provided by Linkura into Google storage within GCP. A virtual machine was created, and a Jupyter Notebook server was started through Google AI Platform, also within GCP. By using the Jupyter Notebook server, we could create a Python program that easily could access and write to data saved within the Google storage and BigQuery.

Before we could start analyzing the data and start creating clusters and models, we had to preprocess the raw data and to choose the relevant features. The format of the file we had access to, was a JSON-file with several characteristics that did not really fit into the requirements of BigQuery. Therefor we had to convert some of the data types, replace illegal tokens and fill in Null or None values for BigQuery to understand it. The file also contained many nested parameters that had to be unraveled before running SQL statements on it for simplicity and consistency. When the preliminary preprocessing of the data was done, the JSON-file could be exported into a table format in BigQuery.

The initial problem with the resulting table was the sheer number of parameters and how it was structured. The next step was to analyze and understand what the parameters meant and contributed to the actual data points. One of the reasons that the number of parameters were so high, was based on the decision not to have nested arrays as values for parameters. Since we had 3-day data points, we divided these types of parameters into three separate ones for more manageable calculations and organization. Some of the parameters could be removed because of them being obviously irrelevant or doublets, for example being in both text and numerical form.

The table still consisted of a huge amount of parameters and naturally the next step turned out to be doing numerical analysis of the remaining parameters with the purpose of spotting instances that was not necessary for generating an accurate clustering model. Since the project were about creating a future improvement on a system dealing with large amount of data, a lot of parameters could mean longer times needed for calculations when applied on larger data sets. The numerical analysis we applied on the data was creating a correlation

[13] Hu, P. J., Chau, P. Y., Sheng, O. R. L., & Tam, K. Y. (1999). Examining the Technology Acceptance Model Using Physician Acceptance of Telemedicine Technology. *Journal of management information systems*, *16*(2), 91-112.

matrix in order to spot dependencies between the parameters in the table. This correlation matrix would have a set threshold value in order to single out parameters that were highly dependent on others, and therefore could be removed without damaging the cluster results in the end.

When the total amount of relevant parameters had been selected, the next step was to start training unsupervised machine learning models on the data in order to analyze and spot patterns within the data set. A selection of two different type of clustering models were selected supported by our concurrent discussions with our mentors at Devoteam. The decision was to use one Dynamic clustering model and one model where the number of clusters were to be set by the user which would be the *K-Means* algorithm.

We started off with applying the *K-Means* algorithm on the data set, with four different variations. The variations consisted of the number of clusters we created for the data set. The numbers tried were two as minimum and 5 as maximum. The *K-Means* algorithm is a built-in function within BigQuery, hence why we started with it, since it would give us an immediate overview of how the cluster would look like. The models were generated by using the uploaded data in BigQuery and then running SQL-statements defined for the API with number of clusters and method as parameters.

Then we built a Python program within the Jupyter Notebook server created in GCP with the purpose of creating the clusters with the dynamic clustering algorithm. First off since we were using a dynamic algorithm, we had to apply the theories of Dendrograms in order to calculate the optimal number of clusters for the algorithm. With the results from the Dendrogram, we could then use the *scikit* library in python to build a program that created clusters based on a hierarchical agglomerative approach of dynamic clustering.

When both types of models had been created, ongoing discussions with Devoteam and Linkura AB were being held in order to analyze the resulting clusters. The next step consisted of experimenting with different number of parameters, where 3-day data would be merged into single parameters and survey results would also be merged into one definitive parameter. The experiments would consist of running the *K-Means* algorithm as well as the dynamic algorithm on either all or 15/18 parameters.

In order to analyze these two different methods and the different numbers of clusters and parameters tried in the experiments, concurrent meetings with representatives of Linkura and Devoteam were held in order to rate and evaluate the results. The expertise from Linkura consisted of for example input from experienced health coaches and representatives with medical engineering background. The different methods were weighed against each other in order to select a method that is most suitable for the cause of predicting whether or not a person should receive future health coaching or measures due to their stress situation.

V. Results

This chapter is structured in the chronological order of our method, so that we can present our way of thinking along the way and to get an understanding of our end result.

A. *Preprocessing*

After the data was processed enough to be usable for the different algorithms, the selection of parameters would be significant, both for resource management and for analytic purpose. Hence why the correlation matrix was created for the remaining parameters. A higher value of a correlation between two parameters would then show dependency of variables, meaning that they would behave the same way subjected to change.

By doing this matrix, a few more parameters could be removed because of high dependency, but the table still consisted of 46 parameters. The threshold set for the correlation matrix was that higher than 0.4 would be considered a dependent correlation.

The main reason that the number of parameters could not be decreased a lot was that most of the cases where dependency would be spotted between parameters, were connected to parameters spread of the 3-day interval. This meant that a person´s HRV-values were correlated to the same value the next day which to us made a whole lot of sense. Since the analysis is largely based on these values, the correlations were not enough to remove all of them but one, since the result would differ a lot. The risk of removing all but one or using mean values would be that we could lose large variations or fluctuations from day to day that could indicate stress factors.

B. *K-Means*

Through GCP, by using the built in API “BigQuery”, several models would be trained with a differing amount of parameters and internal *K-Means* related variables like the *K-Means++* algorithm that is a different approach of finding the initial values of the cluster. These models were then presented in the graphical interface in BigQuery so that discussions could be made with expertise on the stress evaluation area, see examples in Table 1. These tables were created for 3 iterations, either with all 46 parameters, or 15/18. Regarding the iterations with 15 or 18, parameters with 3-day data were joined into single mean value parameters as well as survey parameters were merged into single ones. These iterations were also divided into four sub iterations with either 2, 3, 4 or 5 clusters. That meant that the number of options to be discussed granted by the *K-Means* algorithm was 12.

Presented to the stress expertise within Linkura, the *K-Means* algorithm generated several options, both regarding on how many parameters should be used, and the number of clusters to apply on the model. When including all 46 parameters, some of the cluster iterations did not show obvious “risk-clusters”, and the iterations were narrowed down to either three or four clusters. The resulting tables for the experiments with lesser parameters were all showing clear “risk-clusters”, hence the tables for 15 or 18 parameters were kept as options for selection of optimal methodology.

C. *Dynamic Hierarchical Clustering*

In order to test a different method than the *K-Means*, a dynamical model was chosen for comparison. The model chosen were a hierarchical agglomerative one, meaning taking on a bottom-up approach on finding the optimal number of clusters. First off, the number of clusters was determined by analyzing a *Dendrogram* created by the data set. This gave us an idea of how the clusters were structured.

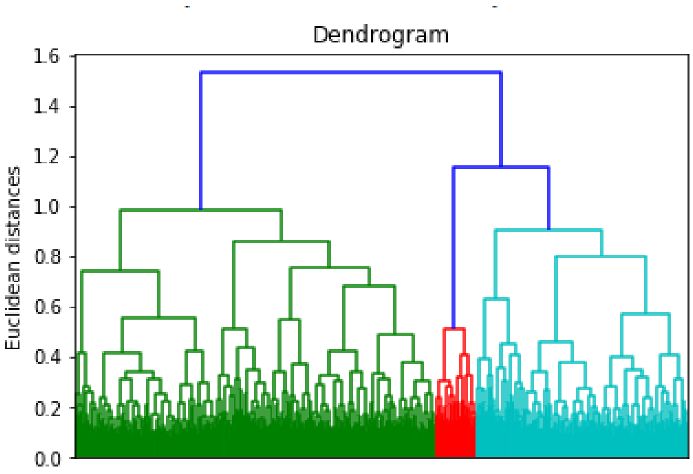

Fig. 1. Dendrogram showing the optimal number of clusters for a set of certain stress related data points. The three colors represent the three different clusters generated. The graph represents how originally the data points each belong to a single cluster, later to be merged into a lesser and optimal amount, in this case we have three.

As seen in Fig.1, the number of clusters selected as optimal for the dynamic clustering model was set to three. During the project, different number of clusters were tested in order to get more extensive knowledge regarding the data behavior. This Dendrogram was calculated by including all the 46 parameters relevant for the data points.

The y-axis shows the Euclidean distance between clusters while the x-axis represents the individual clusters which is not a metric itself. From the bottom we have one cluster for every data point in the set, and higher up in the graph we can see that its merging into a fewer number of clusters. Since we are using the dynamic algorithm with the dendrogram, as mentioned before, at the starting point all the individual data points are considered as separate clusters that will be merged throughout the process.

In the same way that was done with parameter selection for the *K-Means* algorithm, tests were run with 15 and 18 parameters selected as well on the dynamic model in order for having different options and cluster to compare for the stress expertise within Linkura AB. The selection of the 15 or 18 parameters, were done as previous experiments, by merging parameters representing measurements from 3-day periods and surveys, as well as analyzing scatter plots on certain parameters behavior.

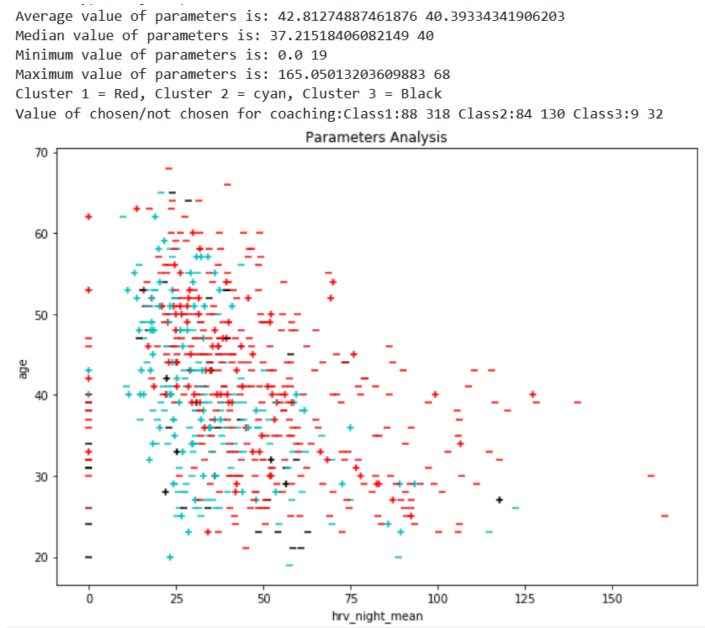

Fig. 2. Scatter plot representing three different clusters showing in different colors, as well as different shapes for a “Semi-label” based on a lot of soft values like surveys and health coaches´ judgements.

By using the likes of scatter plots seen in Fig.2, characteristics on certain parameters would be analyzed. For example, if the centroids of clusters differ a lot, or if certain clusters are dominantly connected to the “semi-label”, then this parameter would be of significant value for the experiments. In this figure we can see a present relation between a higher age and a lower HRV-value during the night since we have no data points in the top right corner. We can also see from the statistics of the different shapes in this graph that the cyan cluster contains a higher percentage of people that were chosen for further coaching as well as it is shifted to the left compared to the red cluster. The “semi-label” that was used were a Boolean parameter that said whether a patient had been selected or not for health coaching.

The reason that we could not use this parameter as an adequate label and use supervised learning were that the decision on giving health coaching as an option were based on a lot of soft values as well, hence why it would not be suitable as a label for this project. This type of scatter plot also showed us the dependencies between two variables which also gave us information regarding what parameters to keep in the experiments. The scatter plot analysis was combined with analyzing *Histograms* on singular parameters for further understanding of the data.

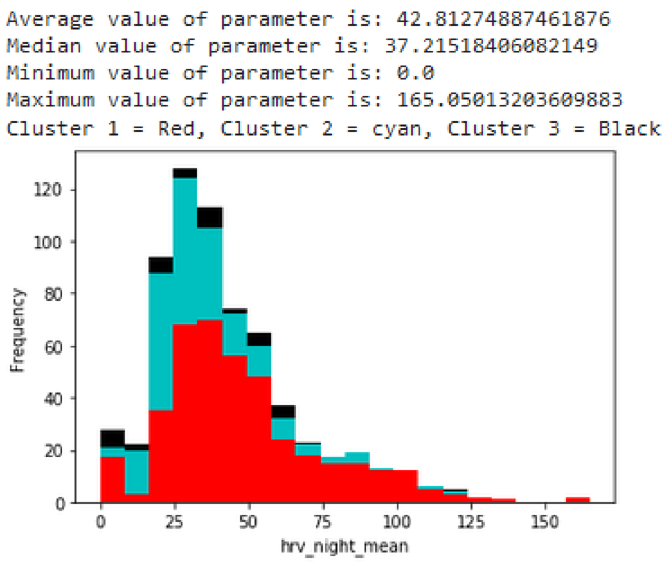

Fig. 3. Example of a Histogram showing the spread of a single parameter and how the clusters were allocated. The different colors represent the different clusters. Machine learning method used for the analysis is the Dynamic hierarchical clustering method.

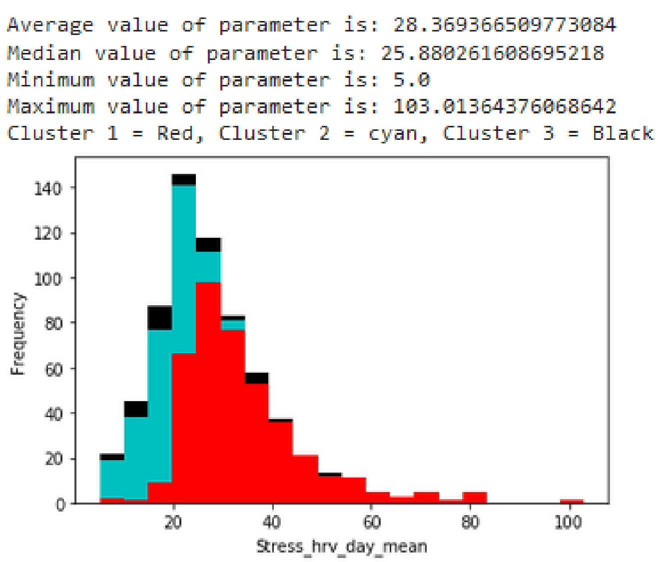

Fig. 4. Histogram showing the spread of the average HRV-value for the data points, divided into separate colors based on clusters.

Combining the analysis of the *Histograms* and the *Scatter plots* and discussing with mentors at Devoteam and representatives within Linkura we could progress further when selecting parameters for extensive testing.

These graphical methods gave some important input on which parameters to keep in order to progress with the experiments. In general, there were a few parameters that stood out with the spread of the clusters. Mainly we could see large differences in HRV-values (Fig.4) measured by Linkura´s sensors as well as sleep, sedentary time and physical activity. In both Fig.3 and Fig.4, we can see the shift of representation of the red cluster percentagewise the higher value we get on the x-axis. This means that these parameters have a clear differential effect on the clusters, hence why they should be further analyzed. When these shifts did not occur for a specific parameter, then that parameter would be an option to be removed for further experimentation. By spotting these important parameters we moved into the next stage of the project. Summarized, these end results of this graphical analysis of the parameters led to the two options of either keeping 15 or 18 parameters for additional testing instead of the original 46. These 15 or 18 parameters were tested with both *K-Means* algorithm as well as the dynamic hierarchical clustering model.

As mentioned, and as could be seen in Fig.1, the optimal selection of number of clusters were three, but in the experiments, tables were created with two and four clusters as well for comparison. This meant that the *Dynamic Clustering Algorithm* led to nine options, either 15, 18 or 46 parameters and either 2, 3 or 4 clusters to be discussed for definitive selection of method.

D. *Selection of method and parameters*

With nine available options for the *Dynamic clustering* algorithm and twelve options for the *K-Means* algorithm, a selection or ranking of these methods was done in order to lay a foundation for future work. With the medical technology expertise from Linkura, the options were narrowed down to either using the *Dynamic* algorithm with all the parameters or the *K-Means* with all 15 or 18 parameters with three or four clusters.

In the end the conclusion was drawn that the clusters for the *Dynamic* algorithm was showing clearer risk groups and with a more consistent result when changing the number of clusters used. When changing the number of clusters between 2, 3 and 4 we could see that the optimal choice of clusters was three, just as we could interpret from the *Dendrogram*. With two clusters, one were a clear risk group, with three we could see that the risk group were divided into two sub clusters of risk groups, and lastly with four clusters the no-risk cluster was divided into two which is irrelevant. Hence why it unanimously seemed most suitable and reliable method for this project.

VI. Discussion

A. *Current practices*

There is a lot of interesting parts to discuss regarding these results, but we must start with going in-depth regarding the current practices to understand whether or not our result is usable and applicable. The data that we have used in this project is a combination of strict metric values as HRV-measurements from people wearing sensors and softer value parameters like surveys that the test subject fills in themself.

With the current practices, these measurements and survey answers are combined with the test subject having interviews with a health coach regarding their situation. The health coach looks at the measurements and discuss with the people that are examined in order to collect a lot of “soft values” in order for the coach to make a subjective evaluation. This subjective evaluation is what lays the foundation for whether the person needs some type of stress management or measure in the future. Since there is a lot of soft values included in the current practice there is always the risk of missing important aspects when using strict numerical and machine learning methods. There are some more factors that needs to be looked at as well that could have negative consequences on the result.

B. *Unsupervised Learning*

One advantage of doing supervised learning instead of unsupervised learning is that you from the beginning have definitive answer to predictions. You start off with data points that has some type of label or several that shows where a data point belong and in what direction. In this case we have had to use unsupervised learning in order to try to spot behavior and patterns between data points and parameters. The main difference is that it makes the result harder to analyze in form of prediction accuracy. In this project we have had expertise from a medical engineering-oriented company in order to examine and analyze the resulting clusters that we have created from our two methods. This grants us professional information regarding if our methodology is following a correct trajectory in terms of correctly classifying groups of people. However, the analysis is based on average values of clusters parameters since it would be too time consuming to manually discuss every data point and person that belongs to the data set.

What this means is that these clusters have outliers that probably differs a lot from data points situated in the center of the centroid, which means most definitely that some of the data points would be wrongly classified. Within a medical business, false positives and false negatives can have a huge impact a person's life, hence why a metric performance evaluation always is a good way to see how the method is delivering accurate answers. Since this is not possible, for future work and development, labeled data would be a good step forward in order to determine whether or not this approach is definitely a suitable solution to improving stress prediction.

C. *Amount of Data*

When discussing the results of this project, it must be mentioned that machine learning methods perform different based on the amount of data that is granted. In most cases, the more data, the better performance you get because of the resulting effect of single extreme fluctuations on smaller data sets. Originally the plan was that we would have access to over 10000 days of data, but some logistical problems hampered that plan and made us do the experiments with around 2000 days' worth of data instead. This is definitely something that could have had an impact on the performance of the clustering and the accuracy.

D. *Response from expertise*

When we have mentioned the potential risks with the results, we can now discuss the positive feedback granted from presenting the different clustering methods. First of all, the results showed that three different approaches brought clear results in the form of accurate risk clusters based on the parameters mean values. Both the K-*Means* algorithm with either all or 15/18 parameters and the dynamic algorithm performed well and could define risk clusters. However, from a numerical point of view, the *K-Means* algorithm was not as consistent as the dynamic one in terms of result when experimenting with the number of clusters. For example, we came across cluster combinations without any example of clusters that stood out with their respective HRV-values.

This was one of the key parameters that we could immediately recognize as important due to almost all the previous cluster experiments. The HRV-values for both daytime and nighttime is currently one of the parameters primarily looked and when analyzing the patients and must have importance in the study. These inconsistencies in correlation between HRV-data makes prediction of patterns and behavior within the data much more difficult using the *K-Means* algorithm.

When using the *K-Means* algorithm, the risk groups were divided more inconsistent and less clearly defined. For example when increasing the amount of clusters, the new clusters consisted of combination of data points from both regular clusters and "risk-groups", making them difficult to label as risk group or not. Of course, there is possible that a combination of data points of these clusters are more accurate when working with a larger data set, but given the scope of the project and the conclusions drawn by discussions with Linkura and Devoteam's expertise, the result seemed more inconsistent compared to the dynamic algorithm.

With the dynamic algorithm we could clearly see that when trying different number of clusters, the result was clearly pointing towards and optimal choice of three. When trying out different number of clusters with the dynamic algorithm, the clusters divided into clear sub clusters of either risk groups or healthy groups. Regardless of the number of parameters used in the experiment, or the number of clusters, the dynamic algorithm performed clear clusters for risk groups and non-risk groups. This consistency combined with the simplicity of analyzing the dynamic clusters, and also the positive indication that the key parameters of the HRV-values were constantly distinguishable between the clusters, made this algorithm seem like a viable option.

Compared to the previous works granted by Hannun, we could definitely see some positive trends regarding spotting behavior and patterns within the clustering results[12]. However, there is the obvious difference in the actual type of result since the models of machine learning differ quite a bit. In our project we use unsupervised learning which means that the evaluation of the results is in this case based on discussions and soft value parameters combined with the actual data in form of clusters. This means that the result in the end is affected by human factors and the interpretation of the resulting clusters and their reliability. In the project by Hannun, supervised learning was done with actual metric evaluation, with accuracy and precision as tools. This could be done thanks to the data being labeled with actual outcomes of previous existing cases. We could work with some semi-labels but since the measure selection for patients currently is based on a lot of soft and subjective parameters, we had to go for the unsupervised learning approach. This of course makes the end result mean different things even if they both point to a positive trend regarding using machine learning as a solution for improving current practices in real life scenarios, especially in a medical market where lives could be saved on a daily basis.

There are some other differences between the projects as well. We had some logistic problems with getting the amount of data points that we initially would prefer to have, while the other discussed project had a larger data set than any previous work in that area. This is something we think influenced the result and would be interesting to investigate further into with a new larger data set.

While the differences in methods and evaluation between the projects are quite significant, the outcome of them both is that a positive potential of implementing a machine learning approach into medical science and “Wellness as a Service” exists and is shown by the results.

E. Economic perspective

The automatization of this service delivery can also provide a platform for Linkura to gain a larger and wider customer segment. Partly because of the faster results leading to a different resource allocation situation, but also because of the pricing point of the service being driven down. In the current situation Linkura is focusing on B2B solutions and a large part of the current clients are larger businesses that can afford this treatment subjected to their employees. This is for the moment a heavy cost investment for the companies with hope of achieving a healthier workplace and increased productivity.
The automatization of the value chain could not only achieve results in a faster manner, but also make the value offering more appealing to smaller businesses and private investors due to the availability of offering less expensive alternatives. If the machine learning models is to be implemented, the need and necessity of the involvement of expensive health coaches would decrease or with optimization even fully be replaced.

Less expensive alternatives or replacement of current best practices would open doors for smaller companies and business to consumer markets where anyone could make an order of the service.

The benefits of diving deep into this type of medical science business and project spills over on different departments within society just as the project by Hannun[12]. The financial benefits for companies encouraged to engage within this niche market, as well as the productivity and well-being for the consumers summarize as a general benefit for society as a whole. Society achieves a healthier population as well as not having to spend the same amount of money on stress related healthcare. You could also argue for the double benefits of the population having a higher occupancy which is achieved by employees not having to stay at home due to mental illness. Instead of creating the financial need of both treatment of the patient, as well as the absence of that employee in the workplace that would provide a financial income for the government based on taxes and increased productivity. This general health effect is present regardless on whether the end product recognizes heart arrhythmia or in our case a stress related lifestyle where measures could be crucial. But in order to achieve these types of positive effects on society and the company, the implementation of the new technology has to be done in an adequate way first. An automatization and a possible effectivization is not going to happen if Linkura is not able to set this new methodology into practice.

When this type of digital implementation would or if to be implemented in the future, acceptance of this new technology could be crucial in order to actually make the digital transition. As we could see in the project by Paul j. Hu et al, using the *Technology Acceptance Model* could be an approach in order to see whether the clients and the market is ready for such an implementation of machine learning[13]. In their study, the result showed us the importance of perceived usability regarding the product or technology to be implemented. Due to the scope of this project, and external events like Covid-19, the empirical study of gathering information with questionnaires to relevant actors within Linkura and their clients has not been possible. Therefore, we discuss the empirical study of Paul J. Hu, and apply it on our project in order to get a wider picture of the implementation. In their study, the questionnaires were sent to physicians at hospitals in Hong Kong in order to see their acceptance of the new technology of telemedicine. In this case, our implementation would be used by the employees at Linkura with a medical technology background, which falls in line with the same demography with a more than average educational level.

Having the same type of demography, in order to apply this new type of approach with machine learning models instead of using health coaches, which would be the end goal, could grant the same characteristics in the results. This means that it is possible that the same demographic could result in focus having to lie withing trying to make the implementation focus on achieving a high value on perception of usefulness. Less focus would be on perceived ease of use since the subjects in this case has a high adaptability to new technology and advanced solutions. This could as with the physicians increase the positive attitude within the employees that are going to use these new methods when presenting solutions to their clients.

The comparativeness of the study on telemedicine and machine learning implementation to stress detection is also high when it comes to the purpose of the technology. Telemedicine is supposed to replace the need for physicians or practitioners to be on site for all types of operations, and instead apply a technology making it possible for these physicians to run their practices from a distance. With the machine learning implementation, the goal is to allow Linkura in the future to use a machine learning model for the data analytics part instead of or in support of having a health coach either present on site or via other connections. The studies are also similar in the fact that they are presented within the medical technology market, which is described by Paul J. Hu as a sensitive one when it comes to change management. A reason for this could be the large impact on people´s life that could occur with bad performing systems or implementations. It is often also important to avoid False Negatives within this market, so implemented methods often need strict numerical performance measures on forehand. These similarities are the reasons that we use that project for the comparison with our study.

However, it is not necessary only the employees at Linkura and their expertise that needs to have a genuine intent of using this new technology. We also have the company´s clients, in this case the other companies that hire Linkura for their stress detection methods. These companies probably have a HR-department that also must be willing to use this type of implementation. These departments could be used to different approaches, for example a high integration of human health coaches and the humane factors they bring. Since they could have different backgrounds than the employees and expertise within Linkura and the physicians in Paul j. Hu´s project, a different focus on acceptance factors could be needed. In order to see the willingness and intent of usage, a TAM could be used with that demographic group in focus. Paul J. Hu et al, also discusses the fact that most of the previous works regarding TAM has been used in an unprofessional environment, most common within student groups. In his work, the physicians represent a professional group of people within their market, this goes along well with the subjects at Linkura that would face the implementation as well due to their medical technology background. The result of that study as presented in the previous work segment, showed that the main factor for implementing this new technology within this certain type of demography was the perceived usefulness of the product. This could be interpreted as a reasonable approach as well for Linkura when eventually implementing a machine learning model to increase efficiency and productivity, since the respective subjects are closely comparable.

As seen by previous cases documented by Linkura, stress detection and measure recommendation have provided a solution for an increasement in productivity and *Return on Investment*(ROI) as well as a declining number in absence on their clients´ workplace[14]. For this to continue with the same positive results, a steppingstone for Linkura´s change management and implementation of an automated system could be to use the likes of TAM. Because of the visible importance of making the new technology implementation have a perceived usefulness as seen in the study regarding telemedicine, technology acceptance could play a large role in eventually successfully adopting this new methodology. Since these both scenarios include a demographic of a similar nature with a possible future implementation on the horizon, both in the medical technology market, using the TAM could be a suitable model where its likely to show that focus should lie on making the new technology have a clearly perceived usefulness. This will then spill over onto the employees' attitude towards the change and therefor increasing the possible intent to use the technology in question.

VII. CONCLUSION

A. *What has been done?*

In order to analyze the project as a whole and what or what have not been achieved, we will start by looking at the practical parts. By using Google Cloud Platform, we have managed through its vast amount of API's to create several machine learning models that can analyze and detect stress related behavior with the data points provided by Linkura´s HRV-sensors and patient surveys. Recreatable and reliable results have been created by the two different methods of the *K-Means* algorithm and hierarchical agglomerative dynamic clustering where the latter one outperformed the first one based on the discussions with mentors at Devoteam and Linkura.

When looking at the actual performance of the models and the clear patterns of the clusters provided by them, the answer to whether or not we can improve the performance of Linkura's stress bot is highly dependent on future work and optimization. Our results provide classification and grouping of data points that seemingly correctly analyses patient data which is of high importance when doing an automatization of the value chain for Linkura. Since this project is providing an unsupervised machine learning solution, testing over a longer period of time is needed for performance evaluation of the clustering model. This is based on the fact that in general, the more data, the more accurate the model is, and since we at the moment cannot use supervised learning and metric evaluation, it is harder to see the actual performance of the model.

Summarized, the output of this project produces increased possibilities for a better performance for the stress-bot in terms of being able to analyze customer data faster, and hopefully in the end, also more accurate.

These models and the results they are presenting has at least created a solid ground for developing and increasing the amount of automatization in the company's value chain when providing stress related analytics. This project is an indicator to the possible positive automatization methods that could be applied to the stress bot framework. For future work, we think that a necessity for improving the accuracy of the stress bot is to collect larger amounts of data and making sure to have a process that concurrently labels the results of customer experiments. In order to analyze the metric performance of the models, labeled data is required in order to implement a performance evaluation on the cluster models or future supervised machine learning models that could be applied to the data.

When comparing to the previous work within this area, we drew the conclusion that the project could produce a result that indicates a positive potential for increasing the performance of current practices within the niche market of "Wellness as a Service". This conclusion is based on the discussion with expertise within Devoteam and Linkura AB regarding the resulting clusters of the two machine learning models that we have tried to implement on the data points provided. The discussions mainly focused on whether the clusters were accurate in terms of dividing and highlighting key parameters connected to a stressful lifestyle, combined with granting clear patterns of other features like sleeping habits and physical activity.

Since this project is based on medical and stress related data, there is obvious risks when trying to classify and organize data points or customer characteristics into either risk groups or not. Within the business, false negatives are the primarily the one result to avoid at all cost. In most of the medical market, wrongly classifying a patient as healthy when not can have tremendous consequences. Since we are dealing with a market where these types of classifications can have really negative impacts on individuals, these models have to be thoroughly optimized over a longer period of time and on larger data sets in order to actually implement them as a general go-to process.

[14] *En hållbar vardag med kommuninvest*, [Online], Available: https://www.linkura.se/about/referenser/referenscase-kommuninvest

B. Why has this been done?

As discussed in the introduction of the project, mental illness exists everywhere in society and constitutes a heavy economic burden on the Swedish government. *Wellness as a service* is a niche market that can lighten this burden and also provide economic growth, both for the providers but also the consumers. By increasing the general mental well-being, companies can achieve a work environment where employees not only have a sound and comforting feeling regarding their situation, they also could achieve an increase in performance and productivity.

The study have been compared to the study by Paul J. Hu, here we could see that the characteristics of the projects are aligned and that an implementation of machine learning models could need extensive technology acceptance analytics in order to align the company´s mind set with their clients in order for their value offering to be attractive. In order for the internal employees to accept the change of methods, a similar approach as described in the physician study would be recommended, where focus lies on the perceived usefulness of the technology. This could make the implementation more solid and effective for the company and in the end provide a successful change in practice.

Summarized we can conclude that we have successfully created a platform and a solid ground for further future works within the area. The patterns of the resulting clusters granted from implementing machine learning models looks promising and shows clear examples of data point classification. By discussing the results with medical technology expertise at Linkura and machine learning feedback from Devoteam, a conclusion can be drawn that this project lays a solid foundation for improving current practices. In order to implement the models, more testing on larger data sets combined with optimization and possibly trying to label the data in order to perform metric evaluation of the models is needed. Mainly this is needed because of the business being medical and stress oriented, meaning that wrongly classifying patients could have large impact on that person's health and well-being.

Acknowledgment

Thanks to examiners Joakim Gustafsson, Jonas Beskow and supervisor Mattias Wiggberg for their devoted guidance throughout the project, as well as feedback and support. Also, thanks to Linkura AB and Anders Tjernvik for providing the required data and vital medical technical expertise needed for analyzing the project results. Thanks as well to Devoteam for letting us take part in this project and giving us the tools needed throughout the whole process. In particular, special thanks to the mentorships of Matyas Manninger and Lu Zhao who consistently helped us with education and expertise regarding Google Cloud Platform as well as machine learning implementation. Also, thanks to Internship Manager Simone French for setting up the logistics for the project and our Agile Manager Dorota Bieniek for keeping the project on track.

References

[1] *Så fungerar Linkura*, Accessed On 2020-02-27 [Online], Available: https://www.linkura.se/about/sa-fungerar-linkura

[2] Tennant, C. (2001). Work-related stress and depressive disorders. *Journal of psychosomatic research*, 51(5), 697-704.

[3] Gustavsson, A., Svensson, M., Jacobi, F., Allgulander, C., Alonso, J., Beghi, E., …& Olesen, J. (2011). Cost of Disorders of the Brain in Europe 2010. *European neuropsychopharmacology*, *21*(10), 718-779

[4] Shaffer, F., & Ginsberg, J.P. (2017). An Overview of Heart Rate Variability Metric and Norms. *Frontiers in public health*, *5*, 258.

[5] Thrun, S., Pratt, L. (2012) Learning to Learn. *Springer Science & Business Media*, 4-12.

[6] Brownlee, J. *Supervised and unsupervised machine learning algorithms*, Accessed On 2020-03-27 [Online], *Available:* https://machinelearningmastery.com/supervised-and-unsuper vised-machine-learning-algorithms/

[7] Tigani, A., Naidu, S. (2014) *BigQuery Fundamentals Analytics*. John Wiley & Sons Inc. Indianapolis, Indiana, 8-14.

[8] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemewat, S., Irving, G., Isard, M., & Kudlur, M. (2016). Tensorflow: A system for large Scale Machine Learning. *In 12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI}16)* (pp. 265-283).

[9] Garbadi, M. *Understanding K-means Clustering in Machine Learning*, Accessed On 2020-03-27 [Online], Available: https://towardsdatascience.com/understanding-k-means-cluster ing-in-machine-learning-6a6e67336aa1

[10] Bock, T. *What is Hierarchical Clustering?*. Accessed On 2020-05-07[Online], Available: https://www.displayr.com/what-is-hierarchical-clustering/

[11] Bock, T. *What is a Dendrogram?*. Accessed On 2020-05-07[Online], Available: https://www.displayr.com/what-is-dendrogram/

[12] Hannun, A. Y., Rajpurkar, P., Haghpanahi, M., Tison, G. H., Bourn, C., Turakhia, M. P., & Ng, A. Y. (2019). Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature medicine, 25*(1), 65.

[13] Hu, P. J., Chau, P. Y., Sheng, O. R. L., & Tam, K. Y. (1999). Examining the Technology Acceptance Model Using Physician Acceptance of Telemedicine Technology. *Journal of management information systems, 16*(2), 91-112.

[14] *En hållbar vardag med kommuninvest*, Accessed On 2020-06-11 [Online], Available: https://www.linkura.se/about/referenser/referenscase-kommuninvest

VIII. Appendix

Table 1:
Example of cluster table in BigQuery for experiments with 2/3/4/5 clusters with K-Means algorithm.

Centroid Id	Count	age	Stress_hrv_day_mean	Stillasittande_sedentary	activity_1	hrv_night_mean
1	506	41.4427	32.0176	402.5692	53.8651	43.6063
2	155	36.9677	16.4775	340.9097	91.0194	40.2220

Centroid Id	Count	age	Stress_hrv_day_mean	Stillasittande_sedentary	activity_1	hrv_night_mean
1	265	41.2038	23.4551	450.6075	51.1914	30.4297
2	80	35.8125	18.7097	339.0375	118.3625	57.8181
3	316	40.8734	34.9447	348.1234	58.0032	49.3984

Centroid Id	Count	age	Stress_hrv_day_mean	Stillasittande_sedentary	activity_1	hrv_night_mean
1	164	42.2012	30.3874	442.6037	54.7317	41.1812
2	75	34.5200	16.9591	323.7067	116.5467	54.4850
3	245	40.6449	37.1794	325.2367	59.7429	53.0337
4	177	40.8588	19.1553	451.9379	50.9024	25.2309

Centroid Id	Count	age	Stress_hrv_day_mean	Stillasittande_sedentary	activity_1	hrv_night_mean
1	135	41.5111	30.9613	376.3037	55.5630	43.6709
2	46	35.3913	19.7134	414.1957	130.3261	68.3687
3	214	40.6776	37.9137	301.8271	60.7991	53.4606
4	112	43.1429	25.4710	653.1518	35.6175	29.1567
5	154	38.5130	17.5456	317.8117	70.5687	29.5621

TRITA EECS-EX-2020:442