**DEVELOP A PREDICTIVE ALGORITHM THAT SUPPORTS PERSONAL FITNESS GOALS**

**Abstract**

This research proposal gives a description about the use of data science in the fitness sector. This technique will help in the accumulation of different types of data in the RAW and in pre-processing of the data forgetting the actual information which will be used for the conduction of the entire research. Machine learning and the internet of things are effective and modern technologies that can be incorporated into different sectors in recent times. These technologies can help them in achieving great results that are superior to human intelligence. Furthermore, computation and calculation of exercise, data can become easier for the athletes with the help of using different predictive algorithms in the internet of things. There are various predictive algorithms such as clustering model that can be used by organisations in order to create fitness tracking devices for athletes.

Table of Contents

[1. Title of the project and research questions 4](#_Toc95730814)

[2. Project Topic 4](#_Toc95730815)

[4. Motivation for the audience 5](#_Toc95730816)

[5. Primary Research Plan 5](#_Toc95730817)

[References 14](#_Toc95730818)

[Appendix 16](#_Toc95730819)

**List of Figures**

[Figure 1: Technologies of data science that can be incorporated in the fitness sector 5](#_Toc94877461)

[Figure 2: Advantages of executing data science in fitness 6](#_Toc94877462)

[Figure 3: Different predictive algorithms 8](#_Toc94877463)

[Figure 4: Gantt chart for scheduling the activities of project 10](#_Toc94877464)

# 1. Title of the project and research questions

The title of the project is ***“Develop a predictive algorithm that supports personal fitness goals''***

***Research objectives***

* To analyse different predictive algorithms that can be helpful in meeting personal fitness goals.
* To recommend a predictive algorithm among different algorithms that can be most useful in case of meeting the fitness goals of an individual.

***Research questions***

1. What are the different predictive algorithms that can be helpful in meeting personal fitness goals?
2. What is the most useful algorithm among different algorithms in case of meeting the fitness goals of an individual?

# 2. Project Topic

This project will be focusing on data science in the fitness sector with focus on how data can be used to help athletes achieve great fitness related results. Through analysing different predictive algorithm’s this research will evaluate and conclude which predictive algorithm is most effective in relation to fitness. Various predictive algorithms such as clustering models can be used by organisations in order to create fitness tracking devices for athletes.

**3. Expected outcomes**

This research work will enable fitness enthusiasts as well as organisations that are responsible for producing fitness products such as fitness trackers to gain information on different predictive algorithms. This will provide an insight to the sportsperson as well as the organisation on the utilisation of the internet of things in the fitness sector effectively. There will also be an algorithm created based on the analysed algorithm. The algorithm will tell the user how much progress they will make if they do a certain workout or eat a type of food daily. For example, it will say you will lose 300 calories if you run 15 minutes every day. And you will gain 200 calories if you eat croissant. The algorithm will specify the workouts based on the individual weight, height, and body type. Machine learning and the internet of things are effective and modern technology that can be incorporated into different sectors in recent times, thus this will be applied into the algorithm. These technologies can help them in achieving great results that are superior to human intelligence. Furthermore, computation and calculation of exercise, data can become easier for the athletes with the help of using different predictive algorithms on the internet of things.

# 4. Motivation for the audience

This project would be significant for athletes, it would enable them to get updates about the different algorithms that could be used for fitness tracking. In addition to this, this project is going to provide information about the use of the internet of things in the case of the fitness sector. Appropriate data collection can be done with the use of these technologies that can help athletes get information about their performance data, practices and exercise accordingly. Additionally, it can also be helpful for organisations that make health care products to provide better services to the athletes.

# 5. Primary Research Plan

**Data collection method**

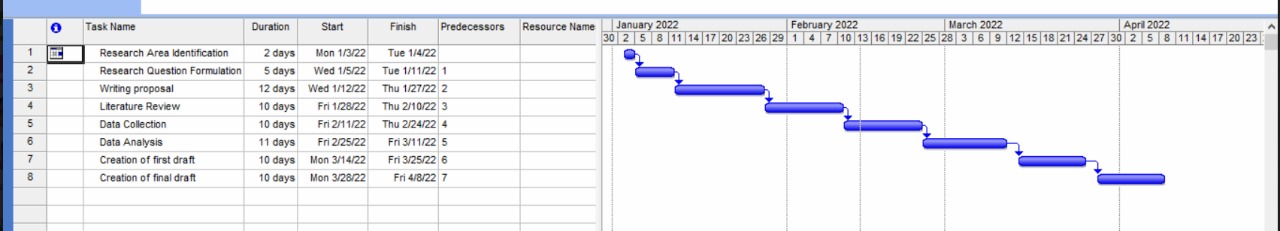
Primary methods of data of collection will be used throughout this research project. According to the statement of Tarofder*et al*. (2019), some of the key benefits of primary research is it allows for the research collected to be current and relevant to the topic. The interview method that has been selected for this research is face-to-face interviews. These interviews allow for a scrupulous data collection and in-depth comprehension.

The process of conducting interviews and surveys allows understanding the facial expression of the participant of research. Visual techniques can be used for the conduction of interviews (Researchgate, 2022). However, interviews can be very much time consuming and can be considered as an expensive technique for accumulating information of search. Biased responses can be delivered by the participants of the research. The ability of the participants of the research needs to be clearly examined before the process of interview. On the other hand, survey methods have some advantages and disadvantages. Survey methods are less expensive, they consume lesser amount of time Tarofder*et al*. (2019). It is also extremely beneficial in the current climate with the outbreak of Covid-19 it can be conducted remotely by the use of online mode. Furthermore, Geographical dependence can be decreased by considering this method of data accumulation. However, There are certain limitations of using survey methods also such as often the participants do not give honest and accurate answers. They can also give wrong information due to their limited knowledge of the topic at hand, this research will try to counter this by making the surveys anonymous and also selecting individuals who have knowledge of fitness. In this research, both ***interview method*** and ***survey method*** will be used for data accumulation.

The following research will also follow secondary data collection process whereby data shall be collected through doing research and collecting information from journals. However other secondary sources that will be followed in this research are online articles, books and previous research scholar data. There are certain processes that will be followed in this ongoing research that are stated accordingly. According to the statement of Nayak and Narayan (2019), it helps in finding the competitive advantages of the research. Interviews will target fitness trainers, after their consent and the face-to-face interviews will be conducted.

**Project timeline and organization**

The author of this project will be Suhayb Mumin. Since the project is large, I will break it down in smaller tasks and set duration to complete each task. This will enable me to organize the project and follow the steps to complete it. Therefore, I have created a Gannt chart as it is very effective for time management.



**Figure 4: Detailed Gantt chart**

(Source: Self-developed)

A Gantt chart has been constructed for the scheduling of all the activities that are required for the execution of the entire research. It is essential for the completion of all the tasks of the research within the deadline.

**Step 1:** Collecting data process

**Step 2:** Analysing the required elements based on algorithms on fitness

**Step 3:** Implementing the data gathered from sources

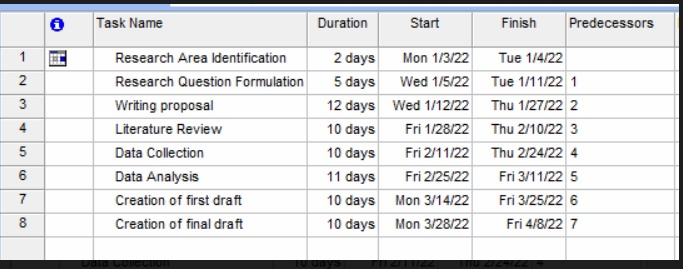
**Step 4:** Making a proper detailed overview about the information gathered

**Step 5:** Making required thematic analysis on the gathered data

**Step 6:** Making final draft progress

**Step 7:** Finalising and proofreading the draft

**Step 8:** Submission



#### Figure 5: Gantt chart for scheduling the activities of project

(Source: Created by Learner)

From the Gantt chart it is seen that the project will be requiring 13 weeks for the completion. Hence, this research will be helpful for understanding the various types of applications of ***predictive algorithms in the fitness field.***

**Research limitation**

In ongoing research there are phases that hinder the research process. This research also experiences some limitation that hinder the research progress. For example, the lack of time available, the little information that is available on this specific topic are key limitations of this research. Another key limitation is focusing on an algorithm task can be quite problematic and hence results in high failures. Lastly, low budget is also another limitation in the research hampering research progress negatively.

**2.1. Concept of predictive algorithms (PA)**

Predictive algorithms for predictive modelling are statistical techniques that are broadly used in machine learning as well as in data mining. Wiens (2019) explains that this technique is very helpful in prediction for the forecast of future outcomes with the help of previous existing collected and available data.

**2.2. The way the predictive algorithms can be used in the fitness sector**

Athletes all over the world use fitness trackers and devices that help them by predicting their physical fitness and in this case, predictive algorithms can be used effectively. Predictive algorithms are basically part of the internet of things (IoT). According to Matsui (2019), IoT along with machine learning can be considered as to most useful technologies that are being used by organisations and sectors all over the world. ***In the case of the fitness sector, the machine learning algorithm assesses from the entire athlete’s information, which determinants are crucial to develop their running speed (Ander, 2021).*** Thus, through thus algorithm the athletes can have a sequential list of training optimizations that they can incorporate in their routines to improve.

#### Figure 2: Advantages of executing predictive algorithm in fitness

(Source: Created by Learner)

**2.3. Analysis of different predictive algorithms**

There are various predictive algorithms that can be used by organisations in order to create fitness tracking devices for athletes.

***Clustering model or algorithm***

The clustering algorithm can be considered as one of the most effective PA that can be useful in predicting the physical fitness of athletes. This algorithm divides and classifies multiple data into different categories that are called clusters. According to Liang *et al.* (2019), it helps in achieving useful information easily and systematically. This algorithm is able to research the data to understand the difference between dense and sparse regions. Therefore, it becomes possible for them to correlate different data attributes, thus it is an easy and effective clustering algorithm (Sinaga and Yang, 2020). Similar workout sessions can provide fitness freaks a better understanding of the overall performance and that way they can gain new insights and improve themselves. The use of a clustering algorithm enables them to cluster their data into two different classes which provide them with better analysis and understanding. It involves automatic discovery of grouping of data and information. It can help in monitoring all the information whether available regarding the performance of fitness exercises by the patients.

***Classification model***

Classification model or algorithm can be considered as the simplest algorithm among different predictive algorithms. This algorithm is mainly used in order to identify different categories of new observations of collected data. ***According to Garifi et al. (2018), the classification algorithm programs can learn from the data set that they are provided with. In addition to that, the observations can be made from the different number of classes or groups.***This can be used for the determination of new forms of observations that can be achieved according to the training data that has been used in the research.  The calculation can be done using KNN algorithm and decision tree method. One of the main goals that the classification algorithm has used is to identify the category in which a given data set comes into (Uy*et al.* 2019). Therefore, the exercise data includes the weight and the amount of exercise an individual does in accordance with the days that can be put into this algorithm. This algorithm can be applied in the fitness sector for the achievement of appropriate results.

***Forecast model***

The forecast model is another predictive algorithm that can be applied at a time when historical statistical data is available. ***As argued by Iglesias et al. (2018), there are multiple input parameters in this algorithm as well because along with the historical data. In addition, to this it also takes into consideration different factors that can impact those data.*** Therefore, in the case of fitness tracking the amount of exercise and the weight can be calculated with the help of this algorithm. Calculation can be done using regression analysis technique and factor analysis technique using this algorithm. Accuracy information can be obtained using these techniques.

That also depends on different factors of athletes. Different factors such as exhaustion levels, the amount and quality of food an individual intake along with that their physical condition as well. It can be said that this algorithm can also be used by athletes for their effective fitness tracking.

***Outlier’s model***

This model is related to anomalous data and entries within the data set which effectively identify different anomalous factors within data sets. ***Zou et al. (2021) argued that different organisations in multiple sectors can use this algorithm which can help them in their development of their business effectively.*** In the fitness sector as well, it can be incorporated with the fitness data of athletes which can provide them on the inside if sometimes data can be incorrect. It is used for the analysis of abnormal type of observations that is present in dataset. It can be done using if mathematical formula known as outlier’s formula. Accuracy can be determined by the comparison of definitions of the datasets in the output.

On the other hand, it does not have much effect on predicting the results that are helpful for fitness tracking. It can be said that this predictive algorithm is not as effective as others in the case of the fitness sector and for athletes. Therefore, it cannot be recommended to the organisations as well as athletes to use this algorithm in their fitness tracking devices.

#### Figure 3: Different predictive algorithms

(Source: Created by learner)

***Time series model***

This model is based on the evaluation of sequences as well as data points in relevance with the time. As an example, data of last year is used in this predictive algorithm in order to develop a numerical Matrix that can provide prediction on next year or next month data (Parmezan *et al.* 2019). In case of fitness sector predictive algorithm can be used. In that case it will take the fitness record of previous year of an athlete and can provide them with the amount of exercise that they need to perform to improve this year. The accuracy can be measured by the fitting of points with the historical type of data and then comparing them.

***Comparison among all the algorithms***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Clustering algorithm | Classification algorithm | Forecast algorithm | Outlier's algorithm | Time series algorithm |
| * It deals with interpretation of output data. * It involves finding natural groups in space. * Effective in solving problems (Cai *et al*. 2020). * Can be used in fitness sector for output interpretation. | * Homogeneity and stability are the major characteristics. * Linear classification algorithm uses logistic regression technique. * KNN algorithm and decision tree algorithm required in it (Chen *et al.* 2020). | * It is based on AI estimation techniques. * Used for making predictions. * Can be used in fitness sector for estimating the outcomes. | * Helps in the determination of anomalous factors   (Zou *et al.* 2021).   * Accuracy can be obtained by comparing datasets. * It can be applied for tracking and monitoring purposes in the fitness sector. | * It deals with evaluation of different sequences. * It involves evaluation of data points with time. * Can be applied in fitness sector for the evaluation purpose. |

**Table 1: Comparison among all the algorithms**

(Source: Created by Learner)

***Clustering algorithm*** will be suitable for application in the fitness sector. It is widely used because of its effectiveness in solving problems. It can be used for the interpretation of output data. Accuracy can be calculated using a confusion matrix and hence it can be applied for updating calculation of all the data that can be applied in the fitness sector. Therefore, it can be said that it is a relevant algorithm for application in the fitness sector.

# References

**Journals**

Ander. R.V. (2020).*Running Smart with Machine Learning and Strava.* Available at:https://towardsdatascience.com/running-smart-with-machine-learning-and-strava-9ba186decde0 [Accessed 15 January 2022]

Cai, J., Wei, H., Yang, H. and Zhao, X., 2020. A novel clustering algorithm based on DPC and PSO. *IEEE Access*, *8*, pp.88200-88214.

Chen, Z., Zhou, L.J., Da Li, X., Zhang, J.N. and Huo, W.J., 2020. The Lao text classification method based on KNN. *Procedia Computer Science*, *166*, pp.523-528.

Garifi, K., Baker, K., Touri, B. and Christensen, D., 2018, August. Stochastic model predictive control for demand response in a home energy management system. In *2018 IEEE Power & Energy Society General Meeting (PESGM)* (pp. 1-5). IEEE.

Iglesias, R., Rossi, F., Wang, K., Hallac, D., Leskovec, J. and Pavone, M., 2018, May. Data-driven model predictive control of autonomous mobility-on-demand systems. In *2018 IEEE international conference on robotics and automation (ICRA)* (pp. 6019-6025). IEEE.

Liang, W., Li, K.C., Long, J., Kui, X. and Zomaya, A.Y., 2019. An industrial network intrusion detection algorithm based on multifeature data clustering optimization model. *IEEE Transactions on Industrial Informatics*, *16*(3), pp.2063-2071.

Matsui, K., Yamagata, Y. and Kawakubo, S., 2019. Real-time sensing in residential area using IoT technology for finding usage patterns to suggest action plan to conserve energy. *Energy Procedia*, *158*, pp.6438-6445.

Nayak, M.S.D.P. and Narayan, K.A., 2019. Strengths and weaknesses of online surveys. *technology*, *6*, p.7.

Parmezan, A.R.S., Souza, V.M. and Batista, G.E., 2019. Evaluation of statistical and machine learning models for time series prediction: Identifying the state-of-the-art and the best conditions for the use of each model. *Information sciences*, *484*, pp.302-337.

Sinaga, K.P. and Yang, M.S., 2020. Unsupervised K-means clustering algorithm. *IEEE access*, *8*, pp.80716-80727.

Tarofder, A.K., Jawabri, A., Haque, A., Azam, S.M. and Sherief, S.R., 2019. Competitive advantages through it-enabled supply chain management (SCM) context. *Polish Journal of Management Studies*, *19*.

Uy, M.A., Pham, Q.H., Hua, B.S., Nguyen, T. and Yeung, S.K., 2019. Revisiting point cloud classification: A new benchmark dataset and classification model on real-world data. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 1588-1597).

Wiens, T., 2019. Engine speed reduction for hydraulic machinery using predictive algorithms. *International Journal of Hydromechatronics*, *2*(1), pp.16-31.

Zou, L., Wang, Z., Geng, H. and Liu, X., 2021. Set-membership filtering subject to impulsive measurement outliers: A recursive algorithm. *IEEE/CAA Journal of AutomaticaSinica*, *8*(2), pp.377-388.

**Websites**

Researchgate. (2022). *Beyond Projection. Using Collaborative Visualizations to Conduct Qualitative Interviews.* Available at: https://www.researchgate.net/publication/264003144\_Beyond\_Projection\_Using\_Collaborative\_Visualizations\_to\_Conduct\_Qualitative\_Interviews [Accessed on: 14 February 2022]

