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

Table of Contents

[Chapter 1 Introduction 3](#_Toc97646767)

[1.1 Introduction 3](#_Toc97646768)

[1.2 Background 3](#_Toc97646769)

[1.3 Rationale 4](#_Toc97646770)

[1.4 Research aim 5](#_Toc97646771)

[1.6 Research Questions 6](#_Toc97646772)

[1.7 Research Significance 6](#_Toc97646773)

[1.8 Structure of the dissertation 6](#_Toc97646774)

[Chapter 2: Literature Review 8](#_Toc97646775)

[2.1 Introduction 8](#_Toc97646776)

[2.2 Concept and significance of predictive algorithm 8](#_Toc97646777)

[2.2 Features of clustering algorithm 8](#_Toc97646778)

[2.3 Steps of the k-means clustering algorithm in context of fitness world 10](#_Toc97646779)

[2.4 Impact of the mentioned algorithm on fitness sector 12](#_Toc97646780)

[2.5 Challenges of k means clustering algorithm in fitness sector 13](#_Toc97646781)

[2.6 Strategies for the implementation of the clustering algorithm 14](#_Toc97646782)

[2.7 Theoretical Framework 15](#_Toc97646783)

[2.8 Conceptual framework 18](#_Toc97646784)

[2.9 Literature Gap 19](#_Toc97646785)

[2.10 Summary 19](#_Toc97646786)

[References 20](#_Toc97646787)

**List of figures**

[Figure 1.1: Market size of fitness sector of UK 4](#_Toc97568367)

[Figure 1.2: Revenue that has been generated by the fitness industry of UK from 2015 to 2019 5](#_Toc97568368)

[Figure 1.3: Structure of Dissertation 7](#_Toc97568369)

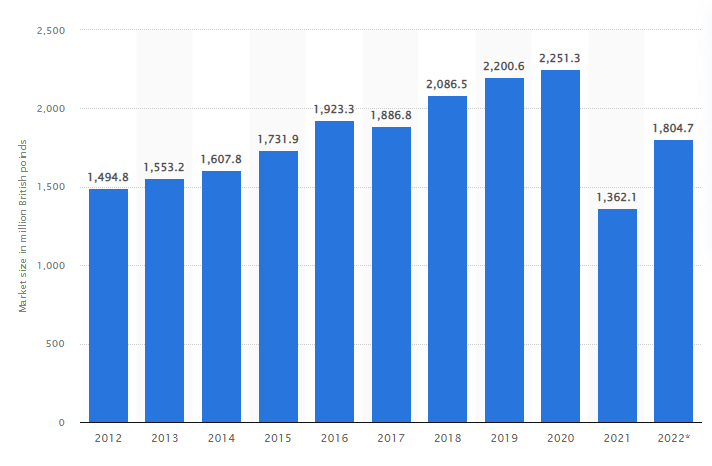
# Chapter 1 Introduction

## 1.1 Introduction

The Clustering model algorithm can be selected as a predictive algorithm for supporting the goals of fitness in a fitness organization. This algorithm can be described as a method in which the input is not in the form of a labeled one. The solving of problems using this algorithm can be based on experience which has been gained by this algorithm for solving similar types of problems as a schedule of training. It also gives descriptions about the aims and objectives of the research and the significance of this research.

## 1.2 Background

K means clustering type of algorithm is based on centroid. This algorithm can be described as the simplest and supervised type of algorithm. Minimization of the variance related with the data points are done within the presence of a cluster and hence it is known as clustering algorithm (Free code camp.org, 2020). This algorithm can be applied in the fitness sector for the improvement of the performances by constantly monitoring the activities that are being performed by the members of the mentioned sector.



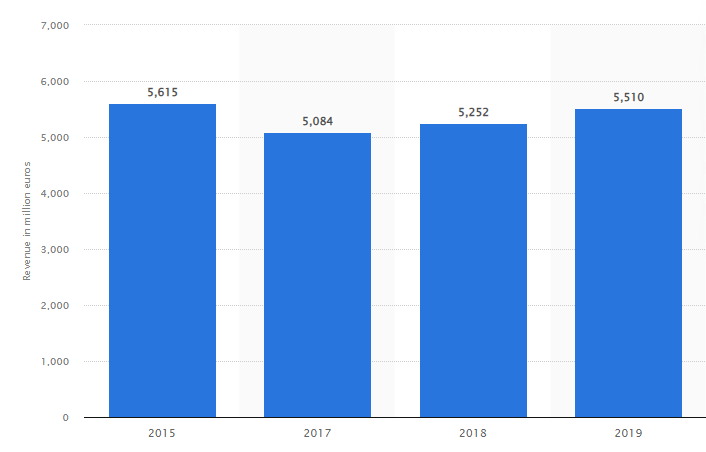
#### Figure 1.1: Market size of fitness sector of UK

(Source: Statista, 2022 a)

It can be seen from the graph that there has been a slight decrease in the market size of the mentioned sector in the present times in the UK. It can be because of the presence of the pandemic situation. Therefore the mentioned algorithm is needed to be implemented in the fitness sector for the improvement of the experience of the members. In this manner the market size of the fitness sector can be increased.

## 1.3 Rationale

The issues that are currently bothering the fitness centers include lack of monitoring of the exercises that are performed by the members of this sector. It also includes lack of motivation of the members and also limitation of motivation of the trainer for giving training about a particular exercise. There are also the technologies that can be used in the mentioned sector for measuring the heart rate ,blood pressure, sugar level in blood of the patient, sugar level in blood of the patients which are required in the fitness sector for the scheduling of the exercises. The total amount of annual turnover from the fitness sectors which are present in the UK is about ***100 to 250 pounds*** (Statista, 2022 b). There has been a decrease of the annual turnover of the mentioned sector in the UK in recent times because of the presence of the pandemic situation.



#### Figure 1.2: Revenue that has been generated by the fitness industry of UK from 2015 to 2019

(Source: Statista, 2019)

It has been estimated that the total number of health sectors which are present in the mentioned country will increase from ***9 million*** to about ***14 million*** (Statista, 2022 c). Implementation of the k-means clustering algorithm which is an algorithm of machine learning in the fitness sector can help in the constant monitoring of the performance of the members. In this manner the experience of the members can be increased by constant rectification of the mistakes by the trainers of the organization. They will also get improvement in their result after the performance of exercises. Implementation of the k-means clustering algorithm which is an algorithm of machine learning in the fitness sector can help in the constant monitoring of the performance of the members. In this manner the experience of the members can be increased by constant rectification of the mistakes by the trainers of the organization. They will also get improvement in their result after the performance of exercises.

## 1.4 Research aim

The aim of the research is development of An appropriate predictive algorithm for supporting personal goals of fitness.

**1.5 Research Objectives**

* To identify the issues that are present in the fitness organizations
* To analyze the predictive type of algorithm which can be used for fulfilling the goals of fitness sector
* To evaluate the results that can be obtained in the fitness sector after the implementation of the predictive algorithm
* To recommend a particular predictive algorithm which will be useful for the improvement of the performance in the fitness sector

## 1.6 Research Questions

1. What are the issues that can be present in fitness organizations?

2. What is the analysis of the predictive algorithm that can be used for fulfilling the goals of the fitness sector?

3. What are the results that can be obtained in the fitness sector after the implementation of the predictive algorithm?

4. What are the particular predictive algorithms that can be recommended for the improvement of the performance in the fitness sector?

## 1.7 Research Significance

This dissertation gives a description about the k-means clustering algorithm which can be used in the fitness sector for the improvement of the experience of the members of the mentioned sector. Technologies of machine learning are required to be implemented in the k-means clustering algorithm. It can be said that this dissertation can help in solving the issues that are currently faced by the fitness sectors and thereby help in improvement of these sectors. The mentioned algorithm can help in monitoring the performance of the members of the fitness sector and can regularly take the heart rate, blood pressure, blood sugar of the patients. These records of the patients are needed to be taken into consideration during the scheduling of the exercises by the trainers so that appropriate exercises can be given to the patients for the improvement of their health.

## 1.8 Structure of the dissertation

The structure of the dissertation is given in the following figure.

#### Figure 1.3: Structure of Dissertation

(Source: Created by Learner)

**1.9 Summary**

K-means clustering algorithm can be used in the fitness sector for the improvement of the performance. This chapter gives the description of the clustering algorithm which can be used in the fitness sector for the improvement of the performance. Therefore it can be said that implementation of clustering algorithm using technologies of machine learning is very much significant in the fitness sector for the improvement of the experience of the members.

# Chapter 2: Literature Review

## 2.1 Introduction

Technologies needed to be implemented in the fitness sector for the tracking of activities of the members. Incorporation of clustering algorithms gives the advantage of instant discovery of natural type of grouping in the data. This chapter gives description about the concepts of clustering algorithm, implementation of clustering algorithm in the effective sector and the theories related to it.

## 2.2 Concept and significance of predictive algorithm

Implementation of technology in the fitness sector can also help in evaluation and assessment of the effectiveness of the members in the fitness regime. Therefore, it can be said that implementation of technology is very much significant in the fitness sector for the identification of the effectiveness of the members and suggesting them several means for the improvement of their efficiencies. ***Predictive algorithms*** utilize historical data for the prediction of the events of the future (Mathworks, 2022). In machine learning, this algorithm can be described as a statistical method for the prediction of outcomes of the future with the help of existing and historical type of data using the technologies of data mining.According to the statement of Wiens (2019), the predictive algorithms are very much significant for the forecasting of the outcomes with available data that has been collected for the purpose of the research and existing type of data. There can be classification algorithms and clustering algorithms which can be used as predictive algorithms for the prediction of the outcomes. According to the statement of Dias and Cordeiro (2020), the classification algorithms can provide the highest amount of accuracy of outcomes. However, there are several disadvantages of the classification algorithms such as these classification algorithms cannot be appropriate for non linear types of problems. These algorithms cannot be applied to the case of a large number of features and also complex types of features (Geeks for geeks, 2020). ***K means clustering algorithm*** has been used in this research and it can be found that this algorithm is most advantageous as it is very fast and convenient.

## 2.2 Features of clustering algorithm

Clustering algorithm can be defined as an unsupervised method of machine learning. It is used for interpretation of input data finding the presence of natural types of clusters which can be present in the feature space. According to the statement of Yuan and Yang (2019), among all the clustering algorithms ***k-means clustering algorithm*** can be said as the best algorithm because it can be described as a fast, convergent and a simple type of algorithm. The classification can be done based on the behavior of the user and development of the principles of recommendations based on the interests of the user. The advantages of the mentioned algorithm include this algorithm involves simple structure of mathematics; this algorithm is fast and convergent. Implementation of the mentioned algorithm is also very easy.

**Figure 2.1: Advantages of k means clustering algorithm**

(Source: Created by Learner)

***Overview of k means clustering algorithm***

In the mentioned algorithms there is presence of a given set of data. This set of data has a specific set of features and for each feature a feature vector can be calculated. This algorithm is responsible for categorizing the items into ‘k’ number of groups according to similarities. According to the statement of Sinaga and Yang (2020), the total number of clusters can be calculated using this algorithm without any type of utilization and selection of parameters. For the calculation of similarity, Euclidean distance is needed to be measured. Euclidean distance can be defined as the length of the segment of line which is present in between the two points.



**Figure 2.2: Mathematical formula for calculating “Euclidean distance”**

(Source: Cut-the-knot.org, 2018)

For the implementation of the k-means clustering algorithm in the fitness sector, the formula of euclidean distance is needed to be calculated. The steps of the algorithm are needed to be known along with the code and mathematical formulae for its implementation and application in the mentioned sector.

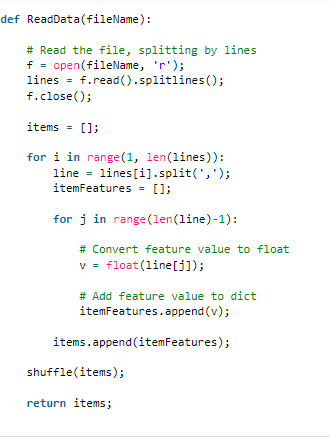
## 2.3 Steps of the k-means clustering algorithm in context of fitness world

The steps are given in the following points

Step 1: In the first step random initialization of ‘k’ points is done.

Step 2: In the next day categorization of each of the items is done up to the closest mean. Updating of the coordinates of the mean is done in the next step. In the mean, averages of the categorized items have been kept.

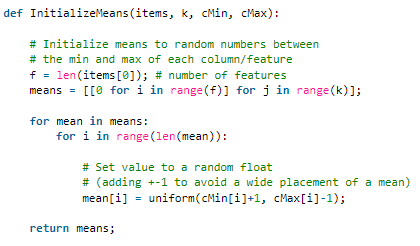
Step 3: The steps are repeated up to the end of the iterative procedures and at the end of the iteration the clusters can be obtained.



**Figure 2.3: Code that can be used for the reading of data**

(Source: Geeks for geeks.org, 2021)

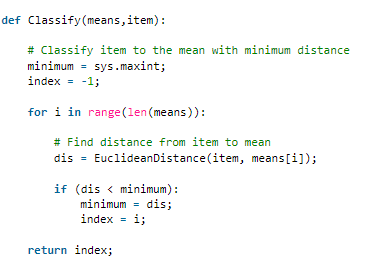
After receiving the data as a text file, each line can be represented as an item containing numerical values. Each numerical value can be used for representing one new feature.



**Figure 2.4 Python code for the calculation of mean**

(Source: Geeks for geeks.org, 2021)

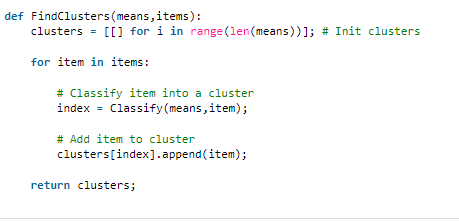
Mean is required to be calculated for finding the euclidean distance which is required for the successful execution of the algorithm.



**Figure 2.5 Python code for the classification of items**

(Source: Geeks for geeks.org, 2021)

Classification of each of the items is needed to be done into one cluster. Similarity of each of the items with the mean is needed to be calculated and classification of the items is needed to be done.



**Figure 2.6: Code that can be used for calculating clusters**

(Source: Geeks for geeks.org, 2021)

The mentioned algorithm can be used for the calculation of the clusters and it is required to be done for its implementation in the fitness sector. The clusters can be obtained at the end of the iterative procedure and it is required to be calculated for finding the application of the mentioned algorithm in the fitness field.

## 2.4 Impact of the mentioned algorithm on fitness sector

Impact of the mentioned algorithm includes convergence of the data that can be obtained from this algorithm. It can be used for the generalization of the clusters having different shapes and sizes. The efficiency of this specified algorithm depends on the selection of the centroids in a random manner. According to the statement of Fard *et al*. (2020), discrete steps of optimization have been used in the mentioned algorithm. The selection of the initial type of centroids in case of the mentioned algorithm can be done randomly. In this manner, different types of clusters can be obtained for different executions of the same data that have been taken as input.

It can be used in the fitness sector, as there are different types of data related to the health of the patients. This data can include scheduling of the exercises of different patients, names of the patients, their blood pressure, blood sugar, heart rate. The chosen algorithm supports the presence of data of different kinds. The quality of the results of the final clustering of this algorithm mainly depends on the selection of the centroids. Therefore it can be said that selection of the data of the fitness sector needs to be done appropriately for obtaining the most appropriate form of clusters. The mentioned algorithm can be used for the segmentation of the patients in the fitness sector, for analyzing the data of the patients and also for keeping records of them.

## 2.5 Challenges of k means clustering algorithm in fitness sector

The challenges of the mentioned algorithm are given in the following table.

| ***Challenges of k means clustering algorithm*** | ***Explanation*** |
| --- | --- |
| Difficulty in the prediction of k value | It can be said that one of the major challenges of the mentioned algorithm is that assumptions about the ‘k’ value are very much difficult. |
| It has been observed that when the initial partitions are different, different clusters can be obtained as the output (Yuan and Yang, 2019). | Often relevant outcomes cannot be obtained in case of the presence of different initial partitions |
| Specification of the total clusters present is needed to be done at the initial stage. Noisy type of data cannot be handled by the mentioned algorithm. | It can be said that as the mentioned algorithm cannot support noisy type of data therefore application has been restricted of the algorithm for those fields which contain noisy data. Identification of the clusters having non convex types of shapes is also not possible by the chosen algorithm. |
| There can be degradation of the performance of the mentioned algorithm for the presence of globular type of clusters (Ray, 2019). | There can be degradation of the performance of the mentioned algorithm when the density and the size of the input data which is present for the execution of the algorithm. In this case appropriate and relevant outcomes from the mentioned algorithm cannot be available. The results of this algorithm can be sensitive to local optimal points and initial points. |

**Table 2.1 Challenges of k means clustering algorithm**

(Source: Created by Learner)

Apart from the presence of these challenges there can also be the presence of cyber security challenges as the confidential information from the patients have been collected regarding the health of the patients. All these data are kept in the databases and they are used for the successful implementation of the mentioned algorithm. Therefore care needs to be taken for maintaining privacy of the data. Cyber security procedures are needed to be adopted for the prevention of any sort of data breaching of the data related to patient health. According to the statement of Linyao and Jianguo (2018), assumption of the K value is very much difficult in case of the mentioned algorithm. Moreover, the mentioned algorithm can be sensitive to measure type of data and relevant outcomes cannot be available by the application of the mentioned data as input. Therefore, the noisy type of data in the fitness sector is required to be avoided to be given as the input of the algorithm for the implementation of the algorithm in the fitness sector. It is required to be done for obtaining the optimal results in the fitness field by the incorporation of the mentioned algorithm.

## 2.6 Strategies for the implementation of the clustering algorithm

Technologies are required to be implemented in the fitness sector for the creation of the insights that are related to fitness. The technologies are required to be implemented for the tracking of activities of the members and also for the scheduling of the relevant and appropriate exercises by the training manager for obtaining the best results. According to the statement of Sakr *et al*. (2018), algorithms of machine learning are required to be implemented in the fitness sector for the evaluation of the performance of the members of the mentioned sector. Several strategies are needed to be implemented for the incorporation of the k-means clustering algorithm in the fitness sector for the monitoring of the performance of the members. These strategies are given in the following points.

***Implementation of k means clustering algorithm without noisy data***

The mentioned algorithm does not give reliable and appropriate output when noisy types of data are taken into consideration (Ray, 2019). Therefore noisy types of data are needed to be avoided for obtaining optimal results at the end of iteration and execution of this algorithm.

***Proper assumption of K value***

Assumption of the K value is needed to be appropriate for obtaining optimal results at the end of the iterative procedure of the algorithm.

***Proper data protection against security breaches***

Strategies are needed to be adopted for the proper protection of the data against any sort of data breaches. According to the statement of Plachkinova and Maurer (2018), confidential and sensitive data is needed to be protected from any cert of databases. This is required to be done as collection of the confidential information of the patients is done and these data are needed to be kept safely.

## 2.7 Theoretical Framework

**2.7.1 Theory of machine learning**

This theory aims at understanding the basic principles related to the computational process. It can be used for combining the tools of statistics and computer science. It deals with the designing of the programs and creation of the rules that can help in the improvement of the performance of the system (cs.cmu.edu, 2022). It can be used for the creation of mathematical models that can capture the features of machine learning algorithms. It can be used for the analysis of different problems of learning. The concept of machine learning theory can be implemented in the k-means clustering algorithm.

**Figure 2.7: Features of machine learning theory**

(Source: Created by Learner)

Clustering can be defined as a technique of machine learning which is used for grouping unlabelled sets of data. This concept can be applied in the fitness sector for the creation of the rules and for the improvement of the experience of the athletes by the development of the new rules. it can also be used for the tracking of the activities of the athletes and for keeping the records the all the data related to their health

**2.7.2 Diffusion theory**

This theory is concerned with the spreading of innovation. This theory has the relative benefits of managing computability and complexity (Sphweb, 2022). It can be used for implementation of an idea of a program in a particular sector. Therefore this theory can be implemented in the fitness sector for the incorporation of new innovative approaches which can be available from the k-means clustering algorithm.

**Figure 2.8: Features of diffusion theory**

(Source: Created by Learner)

It can be used for the implementation of technology in the fitness world for monitoring the activities of the athletes and for the improvement of the performance.

## 2.8 Conceptual framework

***Computability***

***Diffusion Theory***

***Develop a predictive algorithm that supports personal fitness goals***

***Implementation in fitness sector***

***Relative Advantages***

***Machine Learning Theory***

***Analysis of learning problems***

***Combination of computer technology and statistics***

***Development of mathematical models***

***Gap of Literature***

**Figure 2.9 Conceptual Framework**

(Source: Created by Learner)

## 2.9 Literature Gap

In the previously published journals, codes and flowcharts have been given regarding the application of the K means clustering algorithm. These journals give descriptions about the advantages and disadvantages of the mentioned algorithms. They also give descriptions about the fields of application of the chosen algorithm. However, the previously published journals fail to give any information regarding the practical examples where the mentioned algorithm can be applied. Proper means of the assumption of 'K’ values and requirement of taking appropriate user inputs for the delivery of correct outcomes are not properly mentioned in the journals. It has been mentioned in the journal that noisy data result in the failure of appropriate outcomes by the execution of this algorithm. However, the reason for failures using this data has not been mentioned.

## 2.10 Summary

It can be summarized that a means clustering algorithm is the most suitable for the implementation in the fitness sector because of its advantage of having convergence. Theory of machine learning and theory of diffusion can be applied in the fitness sector for the implementation of technology and for the improvement of the performance of the athletes.

# References

**Journals**

Yuan, C. and Yang, H., 2019. Research on K-value selection method of K-means clustering algorithm. *J*, *2*(2), pp.226-235.

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

Fard, M.M., Thonet, T. and Gaussier, E., 2020. Deep k-means: Jointly clustering with k-means and learning representations. *Pattern Recognition Letters*, *138*, pp.185-192.

Ray, S., 2019, February. A quick review of machine learning algorithms. In *2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon)* (pp. 35-39). IEEE.

Linyao, X. and Jianguo, W., 2018, April. Improved k-means algorithm based on optimizing initial cluster centers and its application. In *2018 Second International Conference of Sensor Network and Computer Engineering (ICSNCE 2018)* (pp. 5-10). Atlantis Press.

Sakr, S., Elshawi, R., Ahmed, A., Qureshi, W.T., Brawner, C., Keteyian, S., Blaha, M.J. and Al-Mallah, M.H., 2018. Using machine learning on cardiorespiratory fitness data for predicting hypertension: The Henry Ford ExercIse Testing (FIT) Project. *PLoS One*, *13*(4), p.e0195344.

Dias Canedo, E. and Cordeiro Mendes, B., 2020. Software requirements classification using machine learning algorithms. *Entropy*, *22*(9), p.1057.

Plachkinova, M. and Maurer, C., 2018. Security breach at target. *Journal of Information Systems Education*, *29*(1), pp.11-20.

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

**Websites**

Statista. (2022 a). *Market size of the gym, health & fitness clubs industry in the United Kingdom (UK) from 2012 to 2022.* Available at: https://www.statista.com/statistics/1194831/fitness-health-club-market-size-uk/ [Accessed on: 7 March 2022]

Free code camp.org. (2020). *8 Clustering Algorithms in Machine Learning that All Data Scientists Should Know.* Available at: https://www.freecodecamp.org/news/8-clustering-algorithms-in-machine-learning-that-all-data-scientists-should-know/ [Accessed on: 7 March 2022]

Statista. (2022 b). *Number of users in the eHealth fitness segment in the United Kingdom (UK) from 2016 to 2022.* Available at: https://www.statista.com/statistics/515788/ehealth-fitness-users-digital-market-outlook-uk/ [Accessed on: 7 March 2022]

Statista. (2022 c). *Revenue of the eHealth fitness segment in the United Kingdom (UK) from 2016 to 2022.* Available at: https://www.statista.com/statistics/515744/ehealth-fitness-revenue-digital-market-outlook-uk/ [Accessed on: 7 March 2022]

Statista.(2019). *Revenue of the fitness market in the United Kingdom (UK) from 2015 to 2019.* Available at: https://www.statista.com/statistics/515788/ehealth-fitness-users-digital-market-outlook-uk/ [Accessed on: 7 March 2022]

Mathworks. (2022). *Predictive Analytics.* Available at: https://in.mathworks.com/discovery/predictive-analytics.html [Accessed on: 8 March 2022]

Cut the knot.org. (2018). *The Distance Formula.* Available at: https://www.cut-the-knot.org/pythagoras/DistanceFormula.shtml [Accessed on: 8 March 2022]

Geeks for geeks. (2020). *Advantages and Disadvantages of different Classification Models.* Available at: https://www.geeksforgeeks.org/advantages-and-disadvantages-of-different-classification-models/#:~:text=Not%20appropriate%20for%20non%2Dlinear,for%20large%20number%20of%20features.&text=High%20performance%20on%20non%20%E2%80%93%20linear,number%20of%20features%2C%20more%20complex. [Accessed on: 8 March 2022]

Geeks for geeks. (2021). *K means Clustering – Introduction.* Available at: https://www.geeksforgeeks.org/k-means-clustering-introduction/ [Accessed on: 8 March 2022]

Cs.emu.edu. (2022). *Machine Learning Theory.* Available at: https://www.cs.cmu.edu/~avrim/Talks/mlt.pdf [Accessed on: 8 March 2022]

Sphweb. (2022). *Diffusion of Innovation Theory.* Available at: https://sphweb.bumc.bu.edu/otlt/mph-modules/sb/behavioralchangetheories/behavioralchangetheories4.html [Accessed on: 8 March 2022]