International Journal on Future Revolution in Computer Science & Communication Engineering
Volume: 4 Issue: 3
ISSN: 2454-4248
632 – 636

Diabetes and Hard Disease Detection Algorithm

Gaurav Divtelwar[1], Prof. Roshni Talmale[2], Prof. Rajesh Babu[3]
[1]*divtelwargaurav@gmail.com*, [1]+91-7276248217,
[2]roshnitalmale.cse@tgpcet.com, [3]rajeshbabu.cse@tgpcet.com

Abstract: diebetise detection we are using a detection technique using an algorithm based on k map technique this technique is helpful for the detection of disease in effective manner and smooth way. The k map method has two comparative way which has to be compared and get result in the form of graph or table .this algorithm is to be achieved in R programming**.**

*****

I. Introduction

In recent years, diebetise is the most comman disease find among the world. After the cross aperticular age the hormonal inbalance of body tends to more disease in human body.the factor called insulin is mostly control the sugar body level in human body. But body of some people get disturbed its function due to some factors.to diagnose these factors some parameters are taken into a consideration ,and for its better and proper consideration we require some test methods and processs through which we can come to proper conclusion .These test is done by proper step by step tests and hence I performed this test using k means algorithm.

Similarly, result of algorithms remains to be adjusted and improved to meet searchable encryption in cloud k means. Therefore, how to design a searchable encryption scheme with support of both personalized ranking and test extension is the problem that we try to tackle in this paper. We study and solve the problem of modified multi-keyword ranked search over encrypted input.while preserve privacy in the result analysis. with the help of k means and research interest model for individual user is built by analyzing the user's diagnosis. And we adopt a scoring mechanism to express user interest smartly by calculating the similarity score between different types of related words and the keyword.

II. Literature Review

In principle, there are three steps for diagnosing any diseaseusing machine learning: (1) Data collection, (2) Preprocessing(3) Diagnosing disease using an appropriate classification model. In this work, we concentrate on both pre-processing and classification part as a proof-of-concept methodology for a diabetes diagnosis. Therefore, to classify the diabetic or nondiabetic subjects,Matlab Classification Learner Toolbox isused that allows easy experimentation with different architecture.

This section presents the process of diagnosing diabetesMellitus. A. Data Collection

In this investigation, experiments are performed using the UCI machine learning respiratory diabetes database which is taken from a large data set supported by the National Institutes

of Diabetes and Digestive and Kidney Diseases. All subjects in this diabetes database are women from Pima Indian heritage having age of at least 21 years old. This data set consists of 768 samples which is divided into two classes 0 or 1, represent negative and positive test respectively. The class distribution is -

III. Methodology

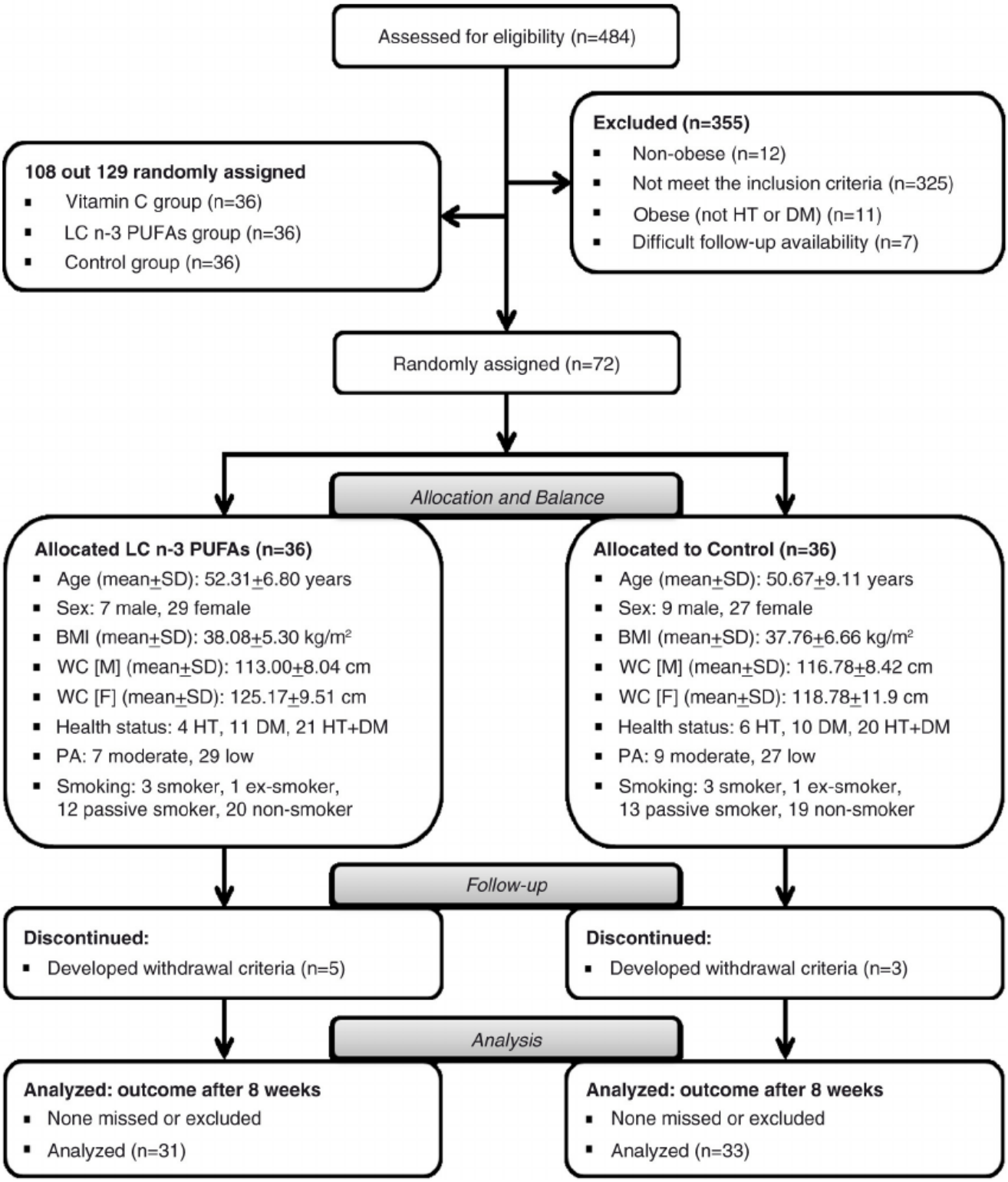
This data gives the percent of constructive responses for each department. In the summary output we can see that for the variable **privileges** in the midst of all 30 departments the minimum percent of favourable response was 30 and the most was 83. In other words, one section had only 30% of responses favourable when it came to assessing 'privileges' and one department had 83% of favourable responses when it came to assessing 'privileges', and a lot of other favourable response levels in between.

When performing clustering, some important concepts must be tackled. One of them is how to deal with data that contains multiple (or more than 2) variables. First option would be to carry out Principal Component Analysis (PCA) and then plot the first two vectors and maybe in addition apply K-Means. Check to be made the data should be uniform, whether the number of clusters obtain are truly representing the original pattern found in the data, could be other clustering algorithms or parameters to be taken, etc. It is recommended to perform clustering algorithms with different approach and slightly test the clustering results with independent datasets. Particularly, it is very important to be careful with the way the results are reported and used.

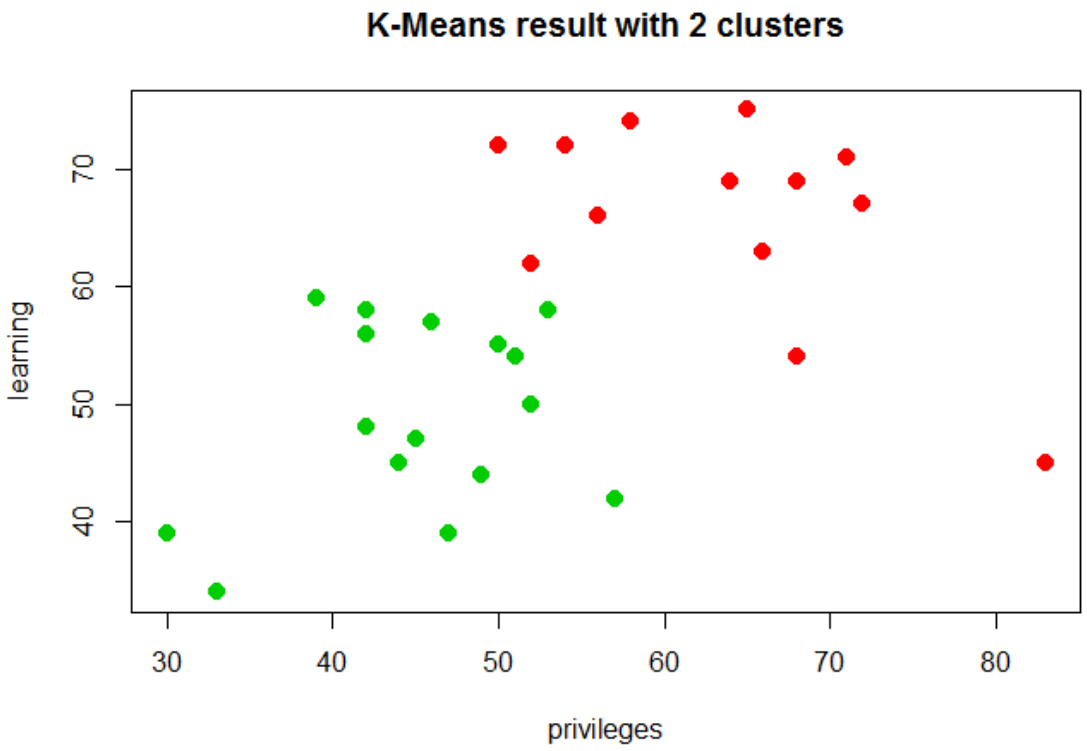
 ISSN: 2454-4248
632 – 636

We're not going to begin most of this concern in this example but they should always be part of a more strong work.

In the example, we will take a separation of the attitude dataset and consider only two variables in our K-Means clustering implement. I would like to cluster the stance dataset with the response from all 30 departments when it comes to 'privileges' and 'learning' and we would like to recognize whether there are commonalities among positive departments when it comes to these two variables



 ISSN: 2454-4248
632 – 636



IV. Calculations

With the data division and the plot above we can see how each department's score act across Privilege and knowledge compare to each other. In the most naive sense, we can apply K-Means clustering to this data set and try to assign each department to a specific number of clusters that are "similar".

Let's use the **kmeans** function from R base stats package:

```
# Perform K-Means with 2 clusters
set.seed(7)
km1=kmeans(dat, 2, nstart=100)

# Plot results
plot(dat, col=(km1$cluster+1) , main="K-Means result with 2 clusters", pch=20, cex=2)
```

The decisions to be made while performing K-Means clustering is to decide on the numbers of clusters to use. In practice, there is no easy answer and it's important to try different ways and numbers of clusters to decide which options is the most useful, applicable or interpretable solution.

We randomly chose the number of clusters to be 2 for design purposes only.

One solution used to identifiy the best number of clusters is called the **Elbow** method and it involves observe a set of possible information of clusters relative to how they minimise the within-cluster sum of squares. The Elbow method examines the within-cluster difference as a function of the number of clusters. Below is a visual representation of the method:

```
# Check for the best number of cluster given the data

mydata<-dat
wss<-(nrow(mydata)-1)*sum(apply(mydata,2,var))
for(iin2:15)wss[i]<-sum(kmeans(mydata,
centers=i)$withinss)
plot(1:15, wss, type="b", xlab="Number of Clusters",
ylab="Within groups sum of squares",
main="Assessing the Optimal Number of Clusters with the Elbow Method",
pch=20, cex=2)
```

With the Elbow method, the solution criterion value (within groups sum of squares) will tend to decrease substantially with each successive increase in the number of clusters. Simplistically, an optimal number of clusters is identified once a "kink" in the line plot is observed. As you can grasp, identifying the point in which a "kink" exists is not a very objective approach and is very prone to heuristic processes.

But from the example above, we can say that after 6 clusters the observed difference in the within-cluster dissimilarity is not substantial. Consequently, we can say with some reasonable confidence that the optimal number of clusters to be used is 6.

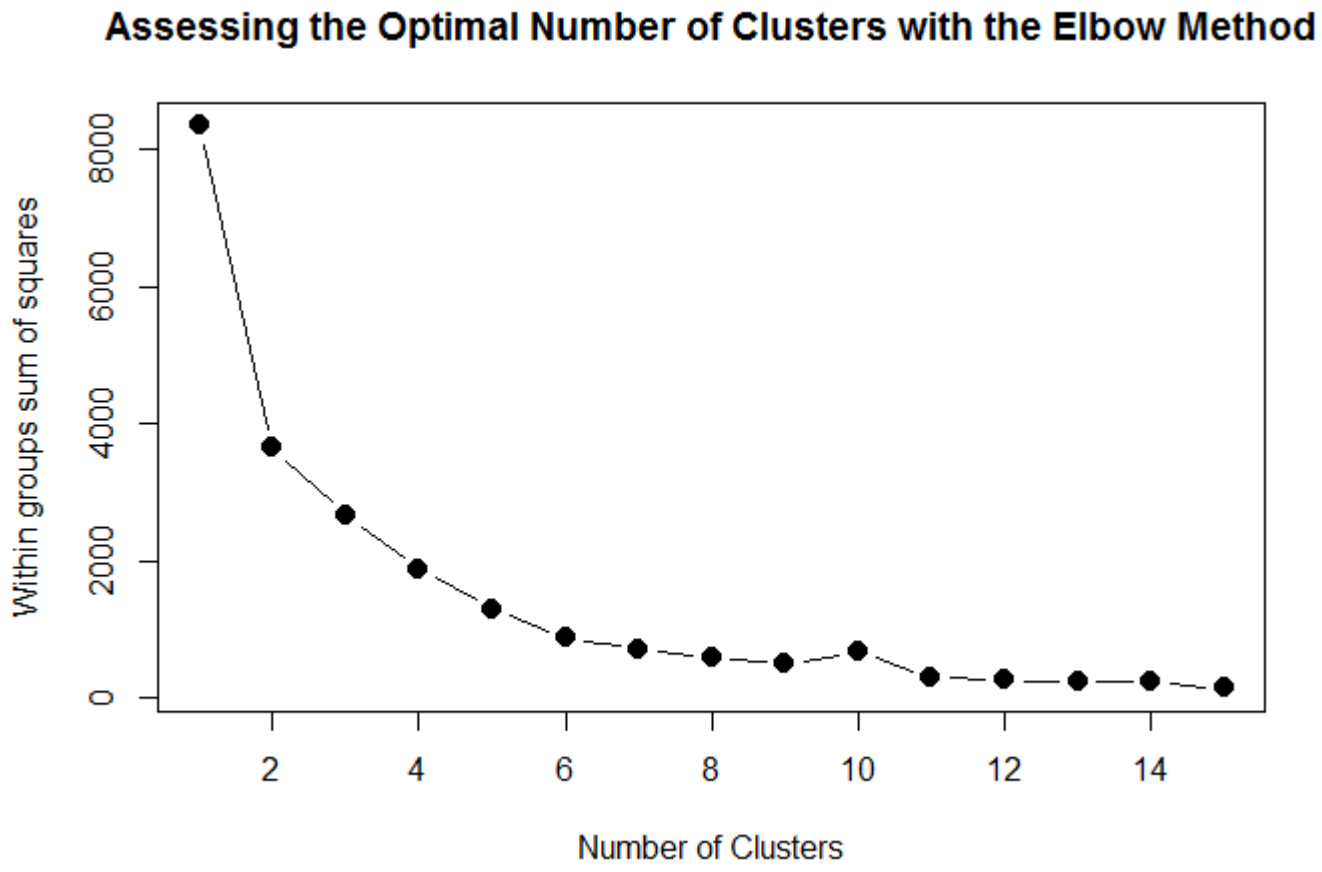
Assuming this assertion is valid, we can go on and apply the identified number of clusters onto the K-Means algorithm and plot the results:

 ISSN: 2454-4248
632 – 636

```
# Perform K-Means with the optimal number of clusters identified from the Elbow method
set.seed(7)
km2=kmeans(dat, 6, nstart=100)

# Examine the result of the clustering algorithm
km2
## K-means clustering with 6 clusters of sizes 4, 2, 8, 6, 8, 2
##
## Cluster means:
##   privileges learning
## 1   54.50000   71.000
## 2   75.50000   49.500
## 3   47.62500   45.250
## 4   67.66667   69.000
## 5   46.87500   57.375
## 6   31.50000   36.500
##
## Clustering vector:
##  [1] 6 5 4 3 1 3 5 5 4 3 5 3 3 2 1 1 4 4 5 2 6 5 3 5 3 4 1 3 4 5
##
## Within cluster sum of squares by cluster:
##  [1]  71.0000 153.0000 255.3750 133.3333 244.7500  17.0000
##  (between_SS / total_SS =  89.5 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"
## [5] "tot.withinss" "betweenss"    "size"         "iter"
## [9] "ifault"
# Plot results
plot(dat, col=(km2$cluster+1) , main="K-Means result with 6 clusters", pch=20, cex=2)
```

From the results above we can see that there is a relatively well defined set of groups of departments that are relatively distinct when it comes to answering favourably around Privileges and Learning in the survey. It is only natural to think the next steps from this sort of output. One could start to devise strategies to understand why certain departments rate these two different measures the way they do and what to do about it. But we will leave this to another exercise.



 ISSN: 2454-4248
632 – 636

V. Conclusions

It has been observed that the application gives an accuracy of 65-68%. This has been concluded after testing this application with many key-words. Study has shown the large number of tweet in a dataset is not for all time an indication for a large language size. This application is also user-friendly and has an instinctive interface which was the basic requirement of the project. This application is also capable of showing data in bar-chart. This helps the user to visualize the data efficiently.

References

[1]. FerouiAmeli "Improvement of the hard exudates detection method used for computer – aided diagnosis of diabetic retinopathy", I.J.Image, Graphics and Signal Processing, DOI: 10.5815/ijigsp.2012.04.03.

[2]. H. Li and O. Chutatiape, "Automated feature extraction in color retinal images by a model based approach," IEEE Trans. on Medical Engineering, vol. 51, pp. 246-254, 2004

[3]. C. I. Sanchez, M. Garcia, A. Mayo, M. Lopez and R. Horniero, "Retinal image analysis based on mixture models to detect hard exudates," Medical Image Analysis, vol. 13, pp. 650-658, 2009. 129

[4]. M. Garcia, C. I. Sanchez, M. I. Lopez, R.Horneiro and D.Abasolo, "Neural network based detection of hard exudates in retinal images," Computer Methods and Programs in Biomedicine, vol. 93, pp. 9-19, 2009