I pledge on my honor that I have not given or received any unauthorized assistance on this

assignment/examination. I further pledge that I have not copied any material from a book, article,

the Internet or any other source except where I have expressly cited the source.

Signature: Kanika Yadav

Date: 10/27/2022

Topic name -

Name - Kanika Yadav

Date: October 27, 2022

W&A Chapter - 8

Question No – Q15 & Q 8

Page no – 460 & 461

**Management Overview**

**Problem Statement:**

W&A **Chapter 8**

**1. Classification (Cluster analysis)**

4th Edition Ch 8.  Q15 **or** 3rd Edition Q15

The file P08\_15.xlsx contains the following information about the top 25 MBA programs (according to the 1997 Business Week Guide): percentage of applicants accepted, percentage of accepted applicants who enroll, mean GMAT score of enrollees, mean under- graduate GPA of enrollees, annual cost of school (for state schools, this is the cost for out-of-state students), percentage of students who are minorities, percentage of students who are non-U.S. residents, and mean starting salary of graduates (in thousands of dollars). Use these data to divide the top 25 schools into four clusters. Then interpret your clusters.

**Data Sources:** Use Data [P08\_15.xlsx](https://sit.instructure.com/courses/60732/files/10002692?wrap=1)[Download P08\_15.xlsx](https://sit.instructure.com/courses/60732/files/10002692/download?download_frd=1)

**Model Approach:** Genetic algorithm with Evolutionary Solving method.

**Solution Analysis:**

**Graphical user interface, application, table, Excel

Description automatically generated**

The solution in above Figure, which uses Stanford, Columbia, NYU, Indiana, is the best we found using the solver for Evolutionary Genetic model solution. The Stanford cluster consists of top MBA school like Harvard based on best mean GPA and Mean GMAT scores. The mean starting salary too is the highest for these schools.   
The Columbia and NYU followed by Indiana center has the best mean starting salary with percent of non Us high whereas less for nYU i.e 0.98. The total cost is lowest in case of Indiana cluster centre. The total Sum disrance counts upto minimum of 23.465.

The optimal number of cluster for typical approach is to stop adding cluster when sum of distances fails to decrease by substantial amount.

**Conclusion:** With further individual city calculation for the distance from the School center we deduce the minimal distance from the given other center. This helps to identify and classify the clusters accordingly. **Table

Description automatically generated**

Finally out of all the calculation s made the Top MBA schools can be formulated into 4 clusters for Stanford, Columbia, NYU and Indiana as -   
Diagram, table

Description automatically generated

**Discriminant Analysis**

**Problem Statement:**

**There are two parts to this question:**

8.

a.  Solve W&A Example 8.8 using the data in the textbook for yourself. You should come up with the answer in the textbook.

b.   You computed the optimal accuracy of the discriminant analysis in part a. Suppose WSJ wants to use discriminant analysis to maximize its expected revenue.   Reformulate the GA model for this objective.  Is accuracy the best measure for this? Solve this new problem with the following data:

|  |  |  |  |
| --- | --- | --- | --- |
|  | ADVERTIZING REVENUE | |  |
|  | Predict Pos | Predict Neg |  |
| Actual Pos | 20 | 0 |  |
| Actual Neg | -12 | 0 |  |
|  |  |  |  |

**Data Sources:** Use Data [WSJ subscribers data.xlsx](https://sit.instructure.com/courses/60732/files/10002691?wrap=1)[Download WSJ subscribers data.xlsx](https://sit.instructure.com/courses/60732/files/10002691/download?download_frd=1) under Modules.

**Model Approach:** We use Genetic Algorithm model for this objective to solve with the help of Evolutionary solver technique.

**Solution Analysis:** Since we used the Genetic algorithm which is generally used to solve complex problems like this one, it uses the evolutionary method in solving the weigths and cut off values for optimal solution which is 91.67% if classification rate. Also it is valid to note that only six of the 84 people are misclassified—four subscribers are misclassified as nonsubscribers and two nonsubscribers are misclassified as subscribers. Also, you can see from the weights that the classification is based primarily on the investment amount, with very little weight placed on income.

Because of the positive weight on the investment amount, people with large investment amounts tend to be classified as subscribers. Therefore, a subscriber such as person 3 is misclassified because his investment amount is abnormally small relative to other subscribers. On the other hand, a nonsubscriber such as person 8 is misclassified because his investment amount is abnormally large relative to other nonsubscribers.

Along with part b where we provide sumproduct of classification matrix along with theAdvertising revenue estimation.

**Table

Description automatically generated**

**Conclusion:**

The solution shown in Figure 8.31 is certainly not unique. Many other sets of weights and cutoff values can obtain a 92.86% correct classification rate, and you will probably obtain a different solution from ours. Note that only six of the 84 people are misclassified—four subscribers are misclassified as nonsubscribers and two nonsubscribers are misclassified as subscribers. Also, you can see from the weights that the classification is based primarily on the investment amount, with very little weight placed on income. Because of the posi- tive weight on the investment amount, people with large investment amounts tend to be classified as subscribers. Therefore, a subscriber such as person 3 is misclassified because his investment amount is abnormally small relative to other subscribers. On the other hand, a nonsubscriber such as person 8 is misclassified because his investment amount is abnor- mally large relative to other nonsubscribers.

In a real application, you would use this analysis for people other than the 84 in the training sample. That is, you would calculate a discriminant score for each such person and then classify each as a nonsubscriber if her discriminant score is less than the cutoff value. However, the percentage correctly classified would typically be less—maybe considerably less—than the 92.86% rate achieved in the training sample. The reason is that the opti- mization procedure takes advantage of all the data for these particular 84 people to derive the weights and the cutoff score. Unfortunately, there is no reason to believe that these will work as well for another group of people. ■