

UNSUPERVISED LEARNING

TOTAL MARKS:70

Dataset Description

Students' feedback was collected in the form of questionnaire from different classes for which three instructors were teaching. The university wants to segment the students based on the feedback given as positive or negative. You can also explore neutral feedback. This will enable the university to understand the concerns of the students who have given negative feedback to improve the rating of the instructors and hence the university. There is a total of 28 course-specific questions and additional 5 attributes.

Note: Consider all values as continuous.

Attribute Information

S.NO	Column Names	Description
1.	instr:	Instructor's identifier; values taken from {1,2,3}
2.	class:	Course code (descriptor); values taken from {1-13}
3.	repeat	Number of times the student is taking this course; values are taken from {0,1,2,3,...}
4.	attendance	Code of the level of attendance; values from {0, 1, 2, 3, 4}
5.	difficulty	Level of difficulty of the course as perceived by the student; values taken from {1,2,3,4,5}
6.	Q1	The semester course content, teaching method, and evaluation system were provided at the start.
7.	Q2	The course aims and objectives were clearly stated at the beginning of the period.
8.	Q3	The course was worth the amount of credit assigned to it.
9.	Q4	The course was taught according to the syllabus announced on the first day of class.
10.	Q5	The class discussions, homework assignments, applications and studies were satisfactory.
11.	Q6	The textbook and other courses resources were sufficient and up to date.
12.	Q7	The course allowed field work, applications, laboratory, discussion and other studies.
13.	Q8	The quizzes, assignments, projects and exams contributed to helping the learning.

UNSUPERVISED LEARNING

14.	Q9	I greatly enjoyed the class and was eager to actively participate during the lectures.
15.	Q10	My initial expectations about the course were met at the end of the period or year.
16.	Q11	The course was relevant and beneficial to my professional development.
17.	Q12	The course helped me look at life and the world with a new perspective.
18.	Q13	The Instructor's knowledge was relevant and up to date.
19.	Q14	The Instructor came prepared for classes.
20.	Q15	The Instructor taught in accordance with the announced lesson plan.
21.	Q16	The Instructor was committed to the course and was understandable.
22.	Q17	The Instructor arrived on time for classes.
23.	Q18	The Instructor has a smooth and easy to follow delivery/speech.
24.	Q19	The Instructor made effective use of class hours.
25.	Q20	The Instructor explained the course and was eager to be helpful to students.
26.	Q21	The Instructor demonstrated a positive approach to students.
27.	Q22	The Instructor was open and respectful of the views of students about the course.
28.	Q23	The Instructor encouraged participation in the course.
29.	Q24	The Instructor gave relevant homework assignments/projects, and helped/guided students.
30.	Q25	The Instructor responded to questions about the course inside and outside of the course.
31.	Q26	The Instructor's evaluation system (midterm and final questions, projects, assignments, etc.) effectively measured the course objectives.
32.	Q27	The Instructor provided solutions to exams and discussed them with students.
33.	Q28	The Instructor treated all students in a right and objective manner.
34.	Q1-Q28	are all Likert-type, meaning that the values are taken from {1,2,3,4,5}

1. Read the dataset (tab, csv, xls, txt, inbuilt dataset)

2. Summarize important observations from the data set (5 Marks)

Some pointers which would help you, but don't be limited by these

a. Find out number of rows; no. & types of variables (continuous, categorical etc.)

- b. Calculate five-point summary for numerical variables*
 - c. Summarize observations for categorical variables – no. of categories, % observations in each category*
 - d. Generate the covariance and correlation tables for the data*
- 3. Perform Exploratory Data Analysis (5 Marks)
Some pointers which would help you, but don't be limited by these
 - a. Create Visualization plots to find the pattern amongst the variables.*
 - b. Check which clustering is better suited for the data.*
- 4. In case of K Means Clustering, Find out the optimal cluster size (K Value) for the data. (5 Marks)
Some pointers which would help you, but don't be limited by these
 - a. Find the optimal K Value in case of K Means Clustering*
 - b. Check if the optimal cluster size is correct*
- 5. Apply Clustering and Generate Visualization (10 marks)
 - a. Check if the Data has been scaled before applying clustering.*
 - b. Apply Clustering and find out if the data points have been clustered correctly.**Check what all Clustering Metrics will be used to support the answer.*
- 6. Split dataset into train and test (70:30) (5 marks)
 - a. Are both train and test representative of the overall data? How would you ascertain this statistically?*

7. Fit the Clustering models and explain the reason of selecting that model. Please write your key observations (15 marks)

- a. What is the overall Accuracy? Please comment on whether it is good or not.*
- b. What is Precision, Recall and WCSS, Silhouette - Score that can justify the deployment of the model.*
- c. What is Cohen's Kappa Value and what inference do you make from the model*
- d. Which other key model output parameters do you want to look at?*

8. In case of a Supervised Machine Learning Problem, how will you decide when to Apply PCA & How do you improve the accuracy of the model? Write clearly the changes that you will make before re-fitting the model. Fit the final model. (20 marks)

Please feel free to have any number of iterations to get to the final answer. Marks are awarded based on the quality of final model you are able to achieve.

9. Summarize as follows (10 marks)

- 1. Summarize the overall fit of the model and list down the measures to prove that it is a good model*
- 2. Write down a business interpretation/explanation of the model – which variables are affecting the target the most and explain the relationship. Feel free to use charts or graphs to explain.*
- 3. What changes from the base model had the most effect on model performance?*
- 4. What are the key risks to your results and interpretation?*