

Avinash Bunga

Information Systems and Business Analytics, Park University

CIS607HOF1P2024 Applied Business Forecasting

Professor: Dr. Abdelmonaem Jornaz

September 12, 2024

Unit 5: Discussion

Classification vs. Clustering

Introduction

After reviewing the lecture notes and relevant articles, I understood the differences between clustering and classification. Clustering is an unsupervised learning method that groups data based on similarities without predefined labels, while classification is a supervised learning method that assigns data to known groups based on learned patterns. To demonstrate these differences, I have created a hypothetical dataset related to car manufacturing to illustrate how each method works within the same context (Simplilearn, 2024).

Hypothetical Dataset for Car Manufacturing

Car Model	Number of Airbags	ABS	ADAS Level	Crash Test Score	Structural Safety Features	Safety Rating
Toyota Corolla	4	Yes	Level 1	4	Reinforced Steel Frame	4
Honda Accord	6	Yes	Level 2	5	Advanced Crumple Zones	5
Ford Fiesta	2	No	Level 0	3	Basic Structure	3
Hyundai Sonata	4	Yes	Level 1	4	Reinforced Steel Frame	4
Volvo XC90	8	Yes	Level 3	5	Advanced Crumple Zones	5

Example 1: Clustering in Car Manufacturing – Segmenting Car Models Based on Safety Features

Problem: The car manufacturer wants to understand how to position its car models in the market based on their safety and structural features. By using clustering techniques, specifically the k-means algorithm, the cars were grouped based on attributes like the number of airbags, ABS, ADAS level, crash test scores, and structural safety features.

Clustering Output: Grouped by Safety and Structural Features

Cluster	Car Models
Basic Safety	Ford Fiesta
Standard Safety	Toyota Corolla, Hyundai Sonata
High Safety	Honda Accord, Volvo XC90

Method and Reason for Selection: Clustering was selected because the manufacturer did not have predefined categories and wanted to explore how their cars compare in terms of safety features. The k-means algorithm was chosen because it efficiently handles the multiple safety-related features and helps identify natural groupings within the car models.

Outcome and Conclusion: The clustering analysis provided the manufacturer with clear segments that highlight how each model compares on safety standards, allowing them to better communicate the safety benefits of their models to customers and adjust their designs to meet market needs (Frost, n.d.).

Example 2: Classification in Car Manufacturing – Predicting Safety Ratings

Problem: To enhance the safety design process, the manufacturer needed to predict the safety rating of new car models based on their features, including advanced safety components. I used classification, specifically a decision tree classifier, to assign a safety rating (1 to 5) based on factors directly related to safety.

Classification Input Features for New Car Model

Feature	Value
Number of Airbags	5
ABS	Yes
ADAS Level	2
Crash Test Score	4
Structural Safety Features	Reinforced Frame

Predicted Safety Rating: 4

Method and Reason for Selection: Classification was appropriate here because the safety ratings were predefined categories, and the goal was to predict which category a new model would belong to. The decision tree was chosen for its effectiveness in handling complex relationships between safety features and its easy-to-interpret decision rules.

Outcome and Conclusion: This classification model provides valuable insights during the design phase, allowing the manufacturer to anticipate safety ratings early and make adjustments to meet safety standards before production (Vitello & Ameduri, 2023).

Comparison and General Conclusion:

The major difference between clustering and classification is their approach to handling data: clustering is used when the goal is to explore and discover new groupings without predefined labels, as shown in the segmentation of car models based on safety features. Classification, on the other hand, assigns data to known categories based on existing patterns, as demonstrated by safety rating predictions.

In car manufacturing, clustering can reveal hidden patterns in safety features, while classification helps make informed decisions aligned with predefined safety standards. Deciding between these methods depends on whether the task involves discovering new patterns (clustering) or sorting data into known groups (classification).

References

Frost, J. (n.d.). *What is K Means Clustering? With an Example*. Statisticsbyjim.

Retrieved September 11, 2024, from

<https://statisticsbyjim.com/basics/k-means-clustering/>

Simplilearn (2024, July 1). *Classification vs. Clustering: Key Differences Explained*.

<https://www.simplilearn.com/tutorials/data-analytics-tutorial/classification-vs-clustering#:~:text=Classification%20involves%20assigning%20data%20into,on%20similarities%20without%20predefined%20labels.>

Vitello, A., & Ameduri, A. (2023, January 31). *Learn About Classification Systems*

With Snacks. Sciencefriday.

<https://www.sciencefriday.com/educational-resources/classification-systems-lesson/>