

Feature Engineering-5

Assignment Questions



- Q1.** What is the difference between Ordinal Encoding and Label Encoding? Provide an example of when you might choose one over the other.
- Q2.** Explain how Target Guided Ordinal Encoding works and provide an example of when you might use it in a machine learning project.
- Q3.** Define covariance and explain why it is important in statistical analysis. How is covariance calculated?
- Q4.** For a dataset with the following categorical variables: Color (red, green, blue), Size (small, medium, large), and Material (wood, metal, plastic), perform label encoding using Python's scikit-learn library. Show your code and explain the output.
- Q5.** Calculate the covariance matrix for the following variables in a dataset: Age, Income, and Education level. Interpret the results.
- Q6.** You are working on a machine learning project with a dataset containing several categorical variables, including "Gender" (Male/Female), "Education Level" (High School/Bachelor's/Master's/PhD), and "Employment Status" (Unemployed/Part-Time/Full-Time). Which encoding method would you use for each variable, and why?
- Q7.** You are analyzing a dataset with two continuous variables, "Temperature" and "Humidity", and two categorical variables, "Weather Condition" (Sunny/Cloudy/Rainy) and "Wind Direction" (North/South/East/West). Calculate the covariance between each pair of variables and interpret the results.
- Note:** Create your assignment in Jupyter notebook and upload it to GitHub & share that github repository link through your dashboard. Make sure the repository is public.