Feature Engineering-4

Assignment Questions

PW SKILLS

Q1. What is data encoding? How is it useful in data science?

Q2. What is nominal encoding? Provide an example of how you would use it in a real-world scenario.

Q3. In what situations is nominal encoding preferred over one-hot encoding? Provide a practical example.

Q4. Suppose you have a dataset containing categorical data with 5 unique values. Which encoding technique would you use to transform this data into a format suitable for machine learning algorithms? Explain why you made this choice.

Q5. In a machine learning project, you have a dataset with 1000 rows and 5 columns. Two of the columns are categorical, and the remaining three columns are numerical. If you were to use nominal encoding to transform the categorical data, how many new columns would be created? Show your calculations.

Q6. You are working with a dataset containing information about different types of animals, including their species, habitat, and diet. Which encoding technique would you use to transform the categorical data into a format suitable for machine learning algorithms? Justify your answer.

Q7.You are working on a project that involves predicting customer churn for a telecommunications company. You have a dataset with 5 features, including the customer's gender, age, contract type, monthly charges, and tenure. Which encoding technique(s) would you use to transform the categorical data into numerical data? Provide a step-by-step explanation of how you would implement the encoding.

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.*