

Feature Engineering-1

Assignment Questions



Q1: What are missing values in a dataset? Why is it essential to handle missing values? Name some algorithms that are not affected by missing values.

Q2: List down techniques used to handle missing data. Give an example of each with python code.

Q3: Explain the imbalanced data. What will happen if imbalanced data is not handled?

Q4: What are Up-sampling and Down-sampling? Explain with an example when up-sampling and down-sampling are required.

Q5: What is data Augmentation? Explain SMOTE.

Q6: What are outliers in a dataset? Why is it essential to handle outliers?

Q7: You are working on a project that requires analyzing customer data. However, you notice that some of the data is missing. What are some techniques you can use to handle the missing data in your analysis?

Q8: You are working with a large dataset and find that a small percentage of the data is missing. What are some strategies you can use to determine if the missing data is missing at random or if there is a pattern to the missing data?

Q9: Suppose you are working on a medical diagnosis project and find that the majority of patients in the dataset do not have the condition of interest, while a small percentage do. What are some strategies you can use to evaluate the performance of your machine learning model on this imbalanced dataset?

Q10: When attempting to estimate customer satisfaction for a project, you discover that the dataset is unbalanced, with the bulk of customers reporting being satisfied. What methods can you employ to balance the dataset and down-sample the majority class?

Q11: You discover that the dataset is unbalanced with a low percentage of occurrences while working on a project that requires you to estimate the occurrence of a rare event. What methods can you employ to balance the dataset and up-sample the minority class?

Note: *Create your assignment in Jupyter notebook and upload it in GitHub & share that github repository link through your dashboard. Make sure the repository is public.*