Odd set

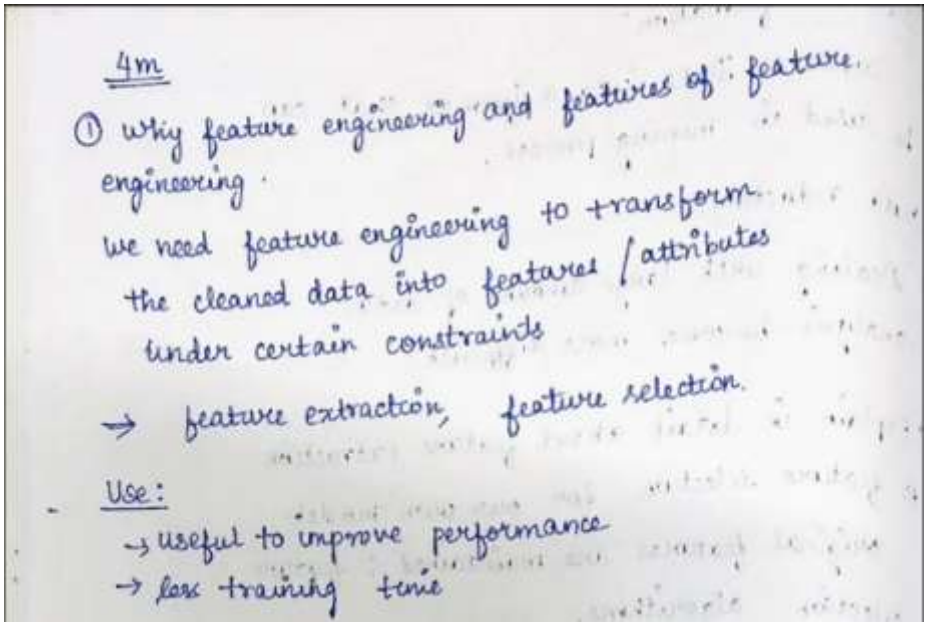
13341

12 m: Data data preprocessing techniques

Ans : https://www.kaggle.com/code/srivignesh/data-preprocessing-for-house-price-prediction

4m :1 q: feature eng and feature of feaure eng

Ans : https://www.javatpoint.com/feature-engineering-for-machine-learning



2 q: data cleaning and prepare ml data

Ans:

What is data cleaning?

Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset. When combining multiple data sources, there are many opportunities for data to be duplicated or mislabeled. If data is incorrect, outcomes and algorithms are unreliable, even though they may look correct. There is no one absolute way to prescribe the exact steps in the data cleaning process because the processes will vary from dataset to dataset. But it is crucial to establish a template for your data cleaning process so you know you are doing it the right way every time.

Problem formulation

Data collection and discovery

Data exploration

Data cleansing and validation

Data structuring

Feature engineering and selection

Even set

21343

12 m : feature of supervised models

Ans : https://www.simplilearn.com/tutorials/machine-learning-tutorial/feature-selection-in-machine-learning#:~:text=Supervised%20Models%3A%20Supervised%20feature%20selection,the%20efficiency%20of%20the%20model

4M 1q: data normalization:

Ans : https://coresignal.com/blog/data-normalization/#:~:text=The%20main%20objective%20of%20database,increase%20security%2C%20and%20lessen%20costs.

2Q: feature extraction vs selection

Ans : Extraction: Getting useful features from existing data.

Selection: Choosing a subset of the original pool of features.

We should apply feature selection, when there is a suspicion of redundancy or irrelevancy, since these affect the model accuracy or simply add noise at best. Sometimes, despite having relevant and non-redundant features, feature selection may be performed only to reduce the number of features, in order to favor interpretability and computing feasibility or to avoid the curse of dimensionality phenomena, i.e., too many features to describe not enough samples.