AIML FEATURE ENGINEERING.

The following are the answers to Assignment 1 in Feature Engineering course.

Q1. The purpose of aggregation when preprocessing raw data to obtain a dataset is to transform the data into a format, scale of unit that is more suitable for analysis. It also helps reduce variance in a dataset. It also reduces the number of attributes or objects.

Q2.

The purpose of attributes/objects into a single attribute/object is as follows

1. Data Reduction
2. Change of Scale
3. More stable data

Q3. In the use case as described in the question

It is best to use to Graph data.

Graphs are collection between nodes and edges which connect them.

Facebook students are represented by nodes.

Edges are connections between nodes.

With a graph built with nodes representing users and bi-directional edges to model equal connection between users.

Using this graph data paradigm, we can find the most influential person with more number of nodes and via their edges can broadcast the message.

Q4.

Decay in radioactive element is Continuous, Ratio

Pairwise distance between the cities is Continuous, Interval

Q5.

We will be using Stratified sampling. It is especially useful when the dataset is heterogeneous.

How will this approach reduce the training error?

It draws random samples from each partition and captures the entire population and reduces the training error.

In case of stratified sampling, the dataset is first separated into different partitions.

In our example classical and non-classical and thereafter samples are taken from each partition.

Therefore there is a good representation of both categories in training data and reduces the error in the training data.