# Feature Selection Techniques:

Feature selection is an important process in any machine learning task to build a more accurate model with additional benefits like reducing the computation. Assume a scenario where our tabular dataset has 1000 columns/attributes, selecting all the 1000 attributes can lead to high computations and could result in inaccurate models.

Explaining these concepts with an example :- assume we have a dataset of students with various attributes/Features such as age, height, weight, hours of study, class rank, previous semester score, and exam score. The given task is to build a ML model to predict the exam score based on the other features or selecting features by feature selection technique.

**1. Feature Importance from Trees :-** Tree-based algorithms can provide a measure of feature importance by assigning importance scores to each feature based on how much they contribute to the overall performance of the model. These scores can be used to select the most important features for predictive modeling.

For example, if we have a dataset of students with various attributes such as age, height, weight, hours of study, class rank, previous semester score, and exam score, we could use a tree-based algorithm to predict the exam score based on the other features. The algorithm would assign importance scores to each feature, and we could then select the most important features for our model.

In our example, a tree-based model might reveal that hours of study, age, and weight are the most important features for predicting exam scores. This means that these features have the greatest impact on the model's ability to predict exam scores. We could then use these features to build a more accurate model.

Here is an example of how we could use feature importance to select features for predictive modeling:

1. We start by training a tree-based model on our dataset of student data.
2. The model assigns importance scores to each feature.
3. We select the top n features with the highest importance scores.
4. We build a new model using only the selected features.
5. We evaluate the performance of the new model.

If the new model performs better than the original model, then we can conclude that the selected features are important for predicting exam scores.

**2. Univariate selection :-** is a technique for selecting features based on their individual relationship with the dependent variable. This means that each feature is analyzed separately to determine its significance. Statistical tests can be used to identify the most significant features, such as chi-square for categorical features or correlation coefficients for numerical features.

For example, if we have a dataset of students with various attributes such as age, height, weight, hours of study, class rank, previous semester score, and exam score, we could use univariate selection to identify the features that are most correlated with exam scores. We could then use these features to build a more accurate model for predicting exam scores.

In our example, we might find that hours of study and height have a strong correlation with exam scores. This means that these features are likely to be important for predicting exam scores. We could then use these features to build a model that predicts exam scores based on hours of study and height.

Here is an example of how we could use univariate selection to select features for predictive modeling:

1. We start by calculating the correlation coefficient between each feature and the dependent variable (exam scores).
2. We select the features with the highest correlation coefficients.
3. We build a model using only the selected features.
4. We evaluate the performance of the model.

If the model performs better than a model that does not use feature selection, then we can conclude that the selected features are important for predicting exam scores.

**3. Recursive feature elimination (RFE) :-** is a technique for selecting features by recursively removing features and building a model iteratively to evaluate their impact on performance. In each iteration, the least important feature is eliminated until the optimal subset is obtained. This helps identify the subset of features that provide the best predictive power.

For example, if we have a dataset of students with various attributes such as age, height, weight, hours of study, class rank, previous semester score, and exam score, we could use RFE to identify the subset of features that best predicts exam scores. We would start by including all of the features in the model, and then iteratively eliminate the least important features until we find the best subset for predicting exam scores.

In our example, we might find that hours of study and height are the most important features for predicting exam scores. This means that these features have the greatest impact on the model's ability to predict exam scores. We could then use these features to build a more accurate model.

Here is an example of how we could use RFE to select features for predictive modeling:

1. We start by training a model on all of the features in our dataset.
2. We evaluate the performance of the model.
3. We identify the least important feature and remove it from the model.
4. We train a new model on the remaining features.
5. We evaluate the performance of the new model.
6. We repeat steps 3-5 until we find the optimal subset of features.

The optimal subset of features is the subset that produces the best model performance.