1. Features are the basic building blocks of datasets. The quality of the features in your dataset has a major impact on the quality of the insights you will gain when you use that dataset for machine learning.

Additionally, different business problems within the same industry do not necessarily require the same features, which is why it is important to have a strong understanding of the business goals of your data science project.

2. The features in your data will directly influence the predictive models you use and the results you can achieve. Your results are dependent on many inter-dependent properties. You need great features that describe the structures inherent in your data. Better features means flexibility.

3. Nominal data is made of discrete values with no numerical relationship between the different categories — mean and median are meaningless. Animal species is one example. For example, pig is not higher than bird and lower than fish.

4. Converting categorical features into numeric features using domain knowledge. For example, we are given a list of countries and say we know the distance to these countries from India then we can replace it with distance from India. So, every country can be represented as its distance from India.

5. Wrapper methods measure the “usefulness” of features based on the classifier performance. In contrast, the filter methods pick up the intrinsic properties of the features (i.e., the “relevance” of the features) measured via univariate statistics instead of cross-validation performance.

The wrapper classification algorithms with joint dimensionality reduction and classification can also be used but these methods have high computation cost, lower discriminative power. Moreover, these methods depend on the efficient selection of classifiers for obtaining high accuracy.

6. Features are considered relevant if they are either strongly or weakly relevant, and are considered irrelevant otherwise.

Irrelevant features can never contribute to prediction accuracy, by definition. Also to quantify it we need to first check the list of features. There are three types of feature selection:

* **Wrapper methods** (forward, backward, and stepwise selection)
* **Filter methods** (ANOVA, Pearson correlation, variance thresholding)
* **Embedded methods** (Lasso, Ridge, Decision Tree).

7. If two features {X1, X2} are highly correlated, then the two features become redundant features since they have same information in terms of correlation measure. In other words, the correlation measure provides statistical association between any given a pair of features.

Minimum redundancy feature selection is an algorithm frequently used in a method to accurately identify characteristics of genes and phenotypes.

8. Three of the most commonly used distance measures in machine learning are as follows:

* Hamming Distance.
* Euclidean Distance
* Manhattan Distance.

9. Euclidean & Hamming distances are used to measure similarity or dissimilarity between two sequences. Euclidean distance is extensively applied in analysis of convolutional codes and Trellis codes.

Hamming distance is frequently encountered in the analysis of block codes.

10. Feature selection is for filtering irrelevant or redundant features from your dataset. The key difference between feature selection and extraction is that feature selection keeps a subset of the original features while feature extraction creates brand new ones.

11. (i) **The width of the silhouette**: The silhouette width, *s*(*i*), is defined as:

urn:x-wiley:20457758:media:ece35774:ece35774-math-0002

*s*(*i*) ranges between −1 and 1. Values near 1 indicate that object *i* is much closer to the other objects in the same cluster than to objects of the closest other cluster, implying a correct classification. If *s*(*i*) is near 0, the correct classification of the focal object is doubtful, thus suggesting intermediate position between two clusters. *s*(*i*) near −1 indicates obvious misclassification. Accordingly, averaging silhouette widths over a cluster gives an assessment of the “goodness” of that cluster, or a sample-wise average can be used as an index of the validity of the entire classification.

(ii) **Receiver operating characteristic curve**: The receiver operating characteristic (ROC) curve, which is defined as a plot of test sensitivity as the y coordinate versus its 1-specificity or false positive rate (FPR) as the x coordinate, is an effective method of evaluating the quality or performance of diagnostic tests, and is widely used in radiology to evaluate the performance of many radiological tests.