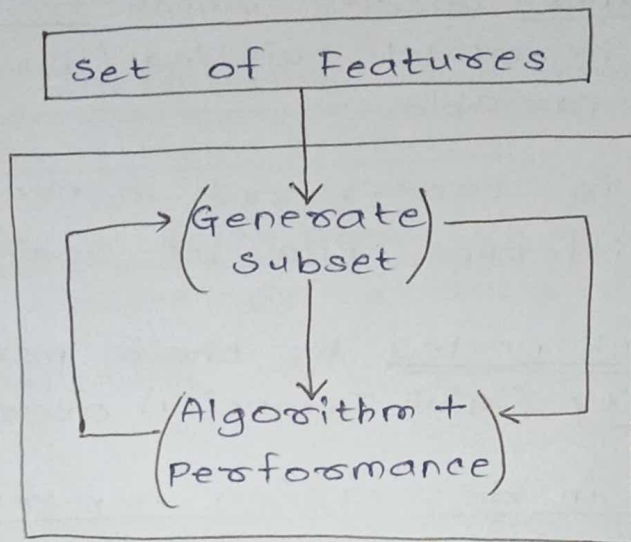


3) Embedded Methods:-

↳ Combine advantage of both filter & wrapper methods by considering iteration of features along with low computational cost.

↳ Similar to Filter methods, but fast & accurate.



• Steps in Embedded Methods:

① Choose ML Algo that support embedded feature selection (ex. Linear Regression with L1 Regularization, Decision Tree, Ensemble methods).

② Train model using entire feature set.

③ As part of model training process, Algo automatically evaluate importance of each feature.

④ Features selected / ranked on importance score.

⑤ Model is built using selected subset of features.

• Techniques in Embedded Methods:

① Random Forest Importance —

↳ Different tree-based methods help with features importance to provide way to select features.

↳ Feature importance: specify which feature has more importance in model building / has great impact on target variable.

↳ Random Forest is tree-based method, bagging Algo aggregates different num. of Decision trees.

↳ Automatically rank nodes by their performance / decrease in impurity (Gini Impurity) over all trees.

↳ Nodes arrange in decreasing impurity values, can prun trees below specific nodes.

↳ Remaining nodes create most imp features subset.

② Regularization —

↳ Add penalty term to different model params to avoid overfitting.

↳ Penalty Term added to coefficients, hence some coefficients shrink to 0.

↳ Can remove 0 coefficient features.

↳ Lasso (L1 Regularization): Add penalty term to loss funⁿ during model training, reduce coefficients of less important features to 0. Non-zero coeff features are selected.

↳ Elastic Net: combines L1 (Lasso) & L2 (Ridge) regularization, allow both feature selection & feature grouping.

• Advantages of Embedded Methods:

① Integration with Model Training — seamless integrate feature selection into model training \Rightarrow efficient & model-specific feature selection.

② Feature Interaction consideration — Capture interactions within features.

③ Automatic & Self-Adaptive — Automatic feature selection based on model's needs & reduce overfitting risk.

• Limitations of Embedded Features:

① Algorithm Dependency — Tied to specific ML Algo. Different Algo \Rightarrow different feature importance score.

② Limited Feature Engineering — Not suitable for extensive feature engineering / domain-specific feature creation required.

③ Difficult Interpret — Even these methods imp, but may not provide direct insights of data.