1. **Cross Validation:** A technique used to evaluate the performance of a machine learning model by partitioning data into training and testing sets multiple times. This helps assess how well the model generalizes to new, unseen data and avoids overfitting.
2. **Overfitting:** Occurs when a model learns the training data too well, including noise and irrelevant patterns, leading to poor performance on new data.
3. **Underfitting:** Occurs when a model is too simple to capture the underlying patterns in the data, resulting in poor performance on both training and test data.
4. **Out-of-Sample Testing:** Similar to cross-validation, it involves testing a model on data not used during training to estimate its real-world performance.
5. **K-fold Cross-Validation:** A specific cross-validation technique where data is divided into k equal-sized folds. The model is trained on k-1 folds and tested on the remaining fold. This process is repeated k times, with each fold used as the test set once.
6. **Repeated Random Subsampling:** A cross-validation method where data is randomly split into training and testing sets multiple times. It's less common than k-fold cross-validation because some data points might never be included in the test set, and others might be included more than once.
7. **Decision Tree Classifier:** A machine learning model that uses a tree-like structure to make decisions based on a series of rules inferred from the data. It's used as an example in the document to demonstrate cross-validation.
8. **Supervised Learning:** A type of machine learning where models learn from labeled data (input-output pairs) to predict outcomes for new, unseen data.
9. **Unsupervised Learning:** A type of machine learning where models learn from unlabeled data to discover patterns, structures, or relationships in the data.