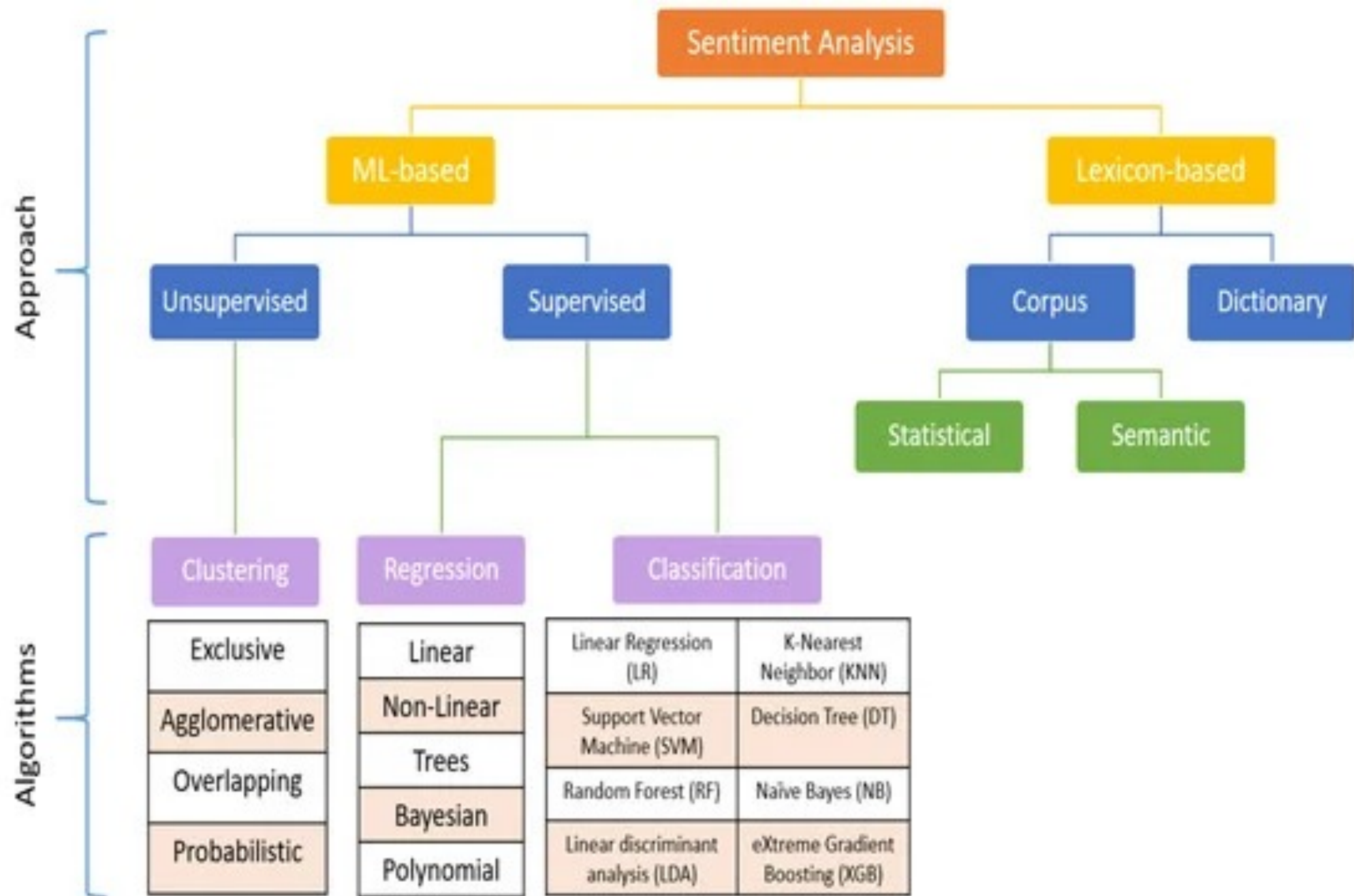


Sentiment Analysis

Approaches and Algorithms

Mustafa Coşkun



Classification Algorithm

1. Train the Classification Model:

- Provide a labeled dataset with input features (X) and corresponding target labels (Y).
- Initialize or choose a specific classification algorithm (e.g., Decision Tree, Random Forest, Logistic Regression).
- Train the model using the labeled dataset:
 - Decision Tree Example:
 - Create a tree node that represents the entire dataset.
 - Choose the best feature to split the data based on some criterion (e.g., Gini impurity, information gain).
 - Split the data into child nodes.
 - Recursively repeat the splitting process for child nodes until a stopping condition is met (e.g., maximum depth, minimum samples per leaf).

2. Make Predictions:

- Given new, unlabeled data (X_new):
 - Input the data into the trained classification model.
- Use the model to make predictions:
 - For a decision tree, traverse the tree from the root node to a leaf node based on feature values.
 - The leaf node's label or class is the prediction.

3. Evaluate Model Performance (Optional):

- If you have a separate dataset for testing, you can evaluate the model's performance by comparing the predicted labels to the true labels.

4. Output:

- The model's predictions for the input data (X_new).

Clustering Algorithm

1. Initialize:

- Choose the number of clusters (K).
- Initialize K cluster centroids randomly (or using some other method).
- Assign each data point to the nearest centroid.

2. Repeat until convergence:

- Update centroids:

for k in range(1 to K):

➡ Calculate the mean of all data points assigned to cluster k.

- Update the centroid of cluster k to the calculated mean.

- Reassign data points:

for each data point:

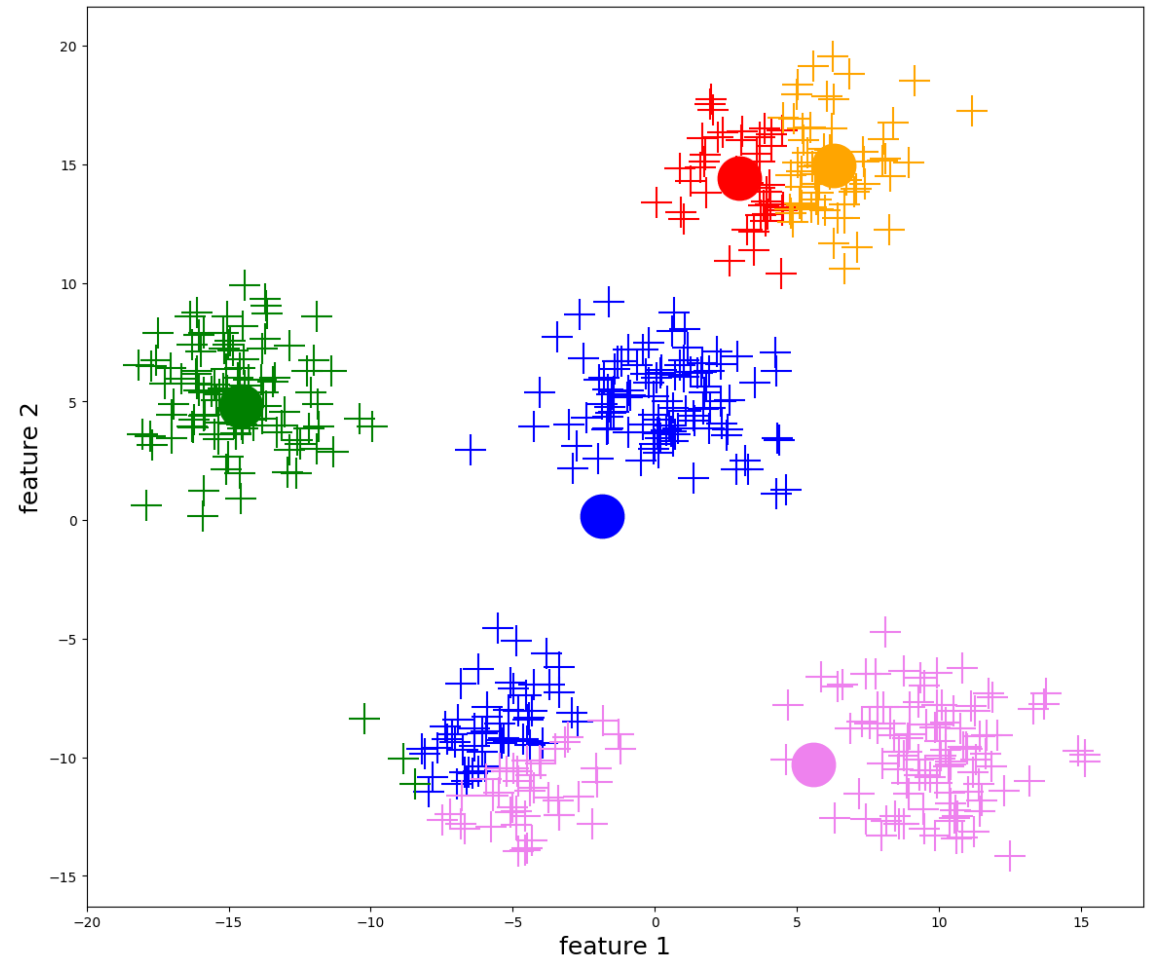
Assign it to the nearest centroid.

3. Convergence criteria:

- Check for convergence. If the centroids do not change significantly or a maximum number of iterations is reached, exit the loop.

4. Output:

- The final cluster assignments for each data point.



4, 2, 10, 11, 20, 3, 12, 30, 25

2 3 4 10 11 | 12 20
25
30

1) Centroids: $\frac{1}{4}$
4, 2, 3

2) $\frac{1}{3}$
4, 2, 3, 10

$\frac{2}{12}$
10, 11, 20, 12, 30, 25

$\frac{2}{18}$
11, 20, 12, 30, 25

3) $\frac{1}{4.75}$
4, 2, 10, 11, 3
12 | $\frac{2}{19.6}$
20, 30, 25

4) $\frac{1}{7}$
4 2 10, 11, 12, 3 | $\frac{2}{25}$
20, 30, 25

5) $\frac{1}{7}$ | $\frac{2}{25}$