**Lab Week 4**

**Feature Selection**

**Part A: Wrapper Method**

Topic 1: Forward Selection

1. Install feature selection library. “pip install mlxtend”
2. Import libraries:
   1. pandas
   2. from mlxtend.feature\_selection import SequentialFeatureSelector
3. Read “wine.csv” file
4. View top five row of the data frame.
5. Inspect the number of rows and columns in the data frame.
6. Check the number of null values in each column.
7. Set X value as independent variable.
8. Set y value as dependent variable (‘target’).
9. View top five records of X.
10. View y.
11. Import sklearn libraries:
    1. From sklearn.model\_selection import train\_test\_split
    2. From sklearn.ensemble import RandomForestClassifier
12. Split X and y using train\_test\_split with 80% for training and 20% for testing
13. View X\_train.
14. View X\_test.
15. View y\_train.
16. View y\_test.
17. Use SequentialFeatureSelector model with parameters:
    1. RandomForestClassifier(n\_jobs = -1)
    2. k\_features = 6
    3. forward = True
    4. floating = False
    5. verbose = 2
    6. scoring = “accuracy”
    7. cv = 5
18. Fit the feature selection model with X\_train and y\_train.
19. Identify index with important feature: from the model.k\_feature\_idx
20. Identify name with important feature: from the model.k\_feature\_names
21. Check the model score: model.k\_score
22. View the score in data frame: pd.DataFrame.from\_dict(model.get\_metric\_dict()).T
23. Repeat step 17 to 22 by considering all 13 features (1-13).

Topic 2: Backward Elimination

1. Use SequentialFeatureSelector model with parameters:
   1. RandomForestClassifier(n\_jobs = -1)
   2. k\_features = 8
   3. forward = False
   4. floating = False
   5. verbose = 2
   6. scoring = “accuracy”
   7. cv = 5
2. Fit the feature selection model with X\_train and y\_train.
3. Identify index with important feature: from the model.k\_feature\_idx
4. Identify name with important feature: from the model.k\_feature\_names
5. Check the model score: model.k\_score
6. Repeat step 1 to 5 by considering all 13 features (1-13).

Topic 3: Exhaustive Search

1. Import libraries
   1. pandas
2. Load the dataset from ‘sampledata.csv’.
3. View top five records from data frame.
4. Inspect the number of rows and columns in the data frame.
5. Check the number of null values in each column.
6. Set X value as independent variable.
7. Set y value as dependent variable (‘target’).
8. View top five records of X.
9. View y.
10. Import sklearn libraries:
    1. From sklearn.model\_selection import train\_test\_split
    2. From sklearn.ensemble import RandomForestClassifier
11. Split X and y using train\_test\_split with 80% for training and 20% for testing
12. View X\_train.
13. View X\_test.
14. View y\_train.
15. View y\_test.
16. Import library from mlxtend.feature\_selection import ExhaustiveFeatureSelector
17. Use ExhaustiveFeatureSelector model with parameters:
    1. RandomForestClassifier(n\_jobs = -1)
    2. min\_features = 1
    3. max\_features = 4
    4. scoring = “accuracy”
    5. cv = 2
    6. n\_jobs = -1
18. Fit the feature selection model with X\_train and y\_train.
19. Identify feature name that are important: from the model.best\_feature\_names
20. Check the model score: model.best\_score\_
21. View the score in data frame: pd.DataFrame.from\_dict(model.get\_metric\_dict()).T