Student Number: [redacted] Name: Bryan Hoang

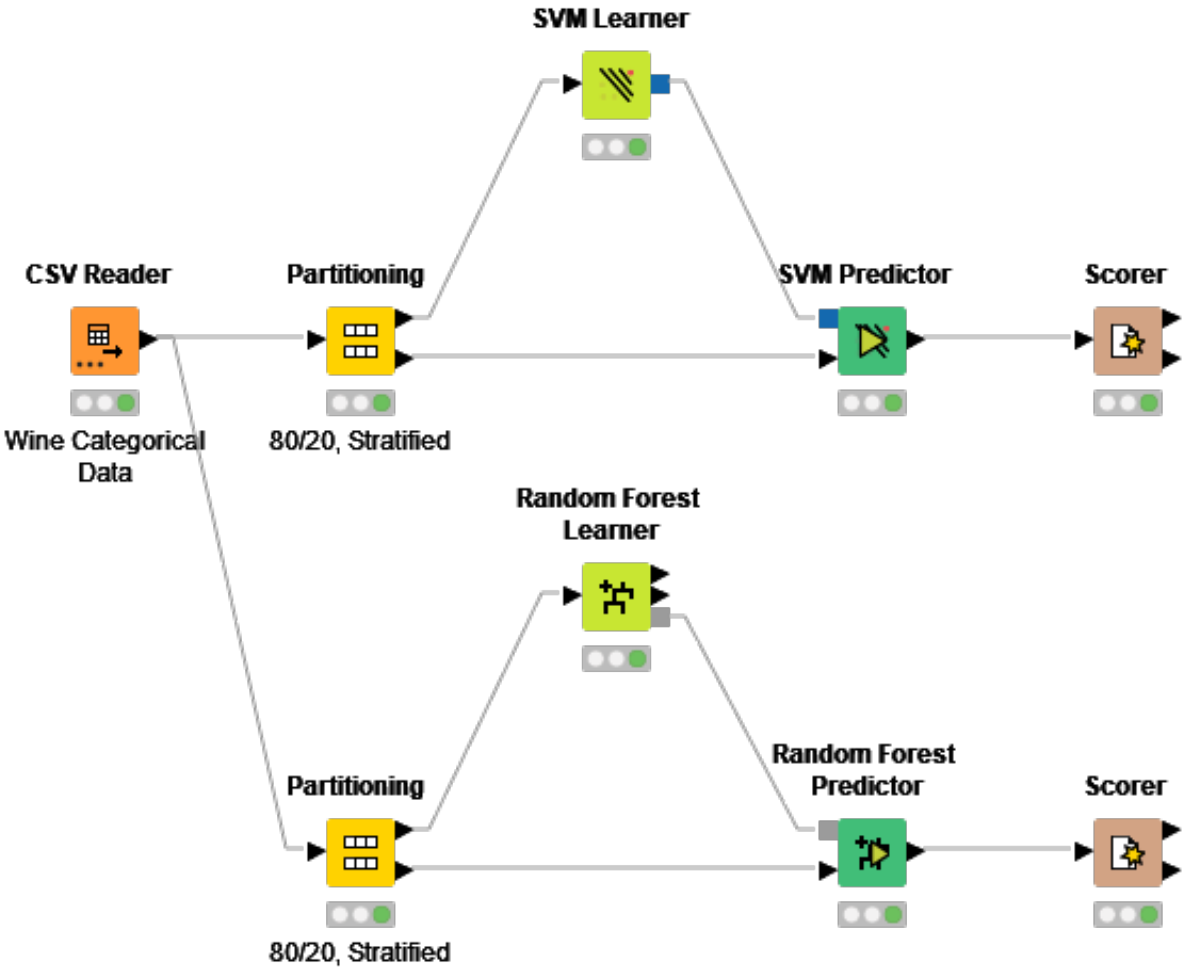

Figure 1: KNIME workflow used for SVM and RF analysis

1. (2 marks) Try using a Support Vector Machine to predict the wine type. Experiment with both the C parameter (penalty for points inside the block) and the choice of kernel. Write a brief comment on which configuration performs best.

   Answer:

   I tested all three kernel types, each with values of $C \in \{0.1, 1, 10\}$.

Kernel Type: Polynomial

I used degree 1 and degree 2 polynomial kernels, as seen in Figure 2 and Figure 3 respectively.

Student Number: [redacted] Name: Bryan Hoang

Kind of wine \ Prediction (Kind of wine)	Type1	Type2	Type3
Type1	11	1	0
Type2	0	13	1
Type3	0	1	9

Correct classified: 33 Wrong classified: 3

Accuracy: 91.667 % Error: 8.333 %

Cohen's kappa (κ) 0.874

(a) $C = 0.1$

Kind of wine \ Prediction (Kind of wine)	Type1	Type2	Type3
Type1	12	0	0
Type2	1	13	0
Type3	0	1	9

Correct classified: 34 Wrong classified: 2

Accuracy: 94.444 % Error: 5.556 %

Cohen's kappa (κ) 0.916

(b) $C = 1$

Kind of wine \ Prediction (Kind of wine)	Type1	Type2	Type3
Type1	12	0	0
Type2	0	13	1
Type3	0	0	10

Correct classified: 35 Wrong classified: 1

Accuracy: 97.222 % Error: 2.778 %

Cohen's kappa (κ) 0.958

(c) $C = 10$

Figure 2: Confusion matrices of the SVM with degree 1 polynomial kernels

Kind of wine \ Prediction (Kind of wine)	Type1	Type2	Type3
Type1	12	0	0
Type2	0	14	0
Type3	0	2	8

Correct classified: 34 Wrong classified: 2

Accuracy: 94.444 % Error: 5.556 %

Cohen's kappa (κ) 0.915

(a) $C = 0.1$

Kind of wine \ Prediction (Kind of wine)	Type1	Type2	Type3
Type1	11	1	0
Type2	0	14	0
Type3	0	0	10

Correct classified: 35 Wrong classified: 1

Accuracy: 97.222 % Error: 2.778 %

Cohen's kappa (κ) 0.958

(b) $C = 1$

Kind of wine \ Prediction (Kind of wine)	Type1	Type2	Type3
Type1	12	0	0
Type2	0	13	1
Type3	0	0	10

Correct classified: 35 Wrong classified: 1

Accuracy: 97.222 % Error: 2.778 %

Cohen's kappa (κ) 0.958

(c) $C = 10$

Figure 3: Confusion matrices of the SVM with degree 2 polynomial kernels

Degree 3 and above polynomial kernels could not find support vectors.

Kernel Type: Hyper Tangent

A Hyper Tangent kernel could not find support vectors.

Student Number: [redacted] Name: Bryan Hoang

Kernel Type: Radial basis function (RBF)

For all values of $C \in \{0.1, 1, 10\}$, the RBF kernel performed eactly the same, as seen in Figure 4.

Kind of wine \ Prediction (Kind of wine)	Type2	Type1	Type3
Type2	14	0	0
Type1	12	0	0
Type3	10	0	0

Correct classified: 14 Wrong classified: 22

Accuracy: 38.889 % Error: 61.111 %

Cohen's kappa (κ) 0

Figure 4: Confusion matrix of the SVM with a RBF kernel

Analysis

For the wine dataset, the SVM performed best with a polynomial kernel. From Figure 2 and Figure 3, we can deduce that **an SVM with a degree 2 polynomial kernel and a C value of 10 performs the best among the tested configurations**.

Student Number: [redacted] Name: Bryan Hoang

2. (2 marks) Try the same prediction using a random forest. Don't be afraid to try a larger number of trees than the default.

Answer:

I used a random forest (RF) with $t \in \{50, 100, 150, 300, 500\}$ trees.

Kind of wine \ Prediction (Kind of wine)	Type1	Type2	Type3
Type1	11	1	0
Type2	0	13	1
Type3	0	0	10

Correct classified: 34 Wrong classified: 2

Accuracy: 94.444 % Error: 5.556 %

Cohen's kappa (κ) 0.916

(a) $t = 50$

Kind of wine \ Prediction (Kind of wine)	Type1	Type2	Type3
Type1	10	2	0
Type2	0	14	0
Type3	0	0	10

Correct classified: 34 Wrong classified: 2

Accuracy: 94.444 % Error: 5.556 %

Cohen's kappa (κ) 0.915

(b) $t = 100$

Kind of wine \ Prediction (Kind of wine)	Type1	Type2	Type3
Type1	12	0	0
Type2	0	14	0
Type3	0	0	10

Correct classified: 36 Wrong classified: 0

Accuracy: 100 % Error: 0 %

Cohen's kappa (κ) 1

(c) $t = 150$

Kind of wine \ Prediction (Kind of wine)	Type1	Type2	Type3
Type1	12	0	0
Type2	0	14	0
Type3	0	0	10

Correct classified: 36 Wrong classified: 0

Accuracy: 100 % Error: 0 %

Cohen's kappa (κ) 1

(d) $t = 300$

Kind of wine \ Prediction (Kind of wine)	Type1	Type2	Type3
Type1	12	0	0
Type2	0	14	0
Type3	0	0	10

Correct classified: 36 Wrong classified: 0

Accuracy: 100 % Error: 0 %

Cohen's kappa (κ) 1

(e) $t = 500$

Figure 5: Confusion matrices of the RF with varying number of trees

From Figure 5, we can see that after hitting $t = 150$ trees, the RF performed perfectly on the dataset.