Data Mining - Course Project

Aim: Internal positioning using WiFi signal strength

Team:

Prajual Pillai(193100069)
Sudip Walter Thomas(193100063)
Amitkumar Dubey(193100062)
Vivek Revi(193100076)

Data Information:

- 2000 rows, 8 columns
- 7 features (WiFi1,WiFi2,WiFi3,WiFi4,WiFi5,WiFi6,WiFi7)
- 1 output (Room No. 1/2/3/4)

	WiFi1	WiFi2	WiFi3	WiFi4	WiFi5	WiFi6	WiFi7	Room
0	-64	-56	-61	-66	-71	-82	-81	1
1	-68	-57	-61	-65	-71	-85	-85	1
2	-63	-60	-60	-67	-76	-85	-84	1
3	-61	-60	-68	-62	-77	-90	-80	1
4	-63	-65	-60	-63	-77	-81	-87	1

Knocking off 10% Data from the Data Set:

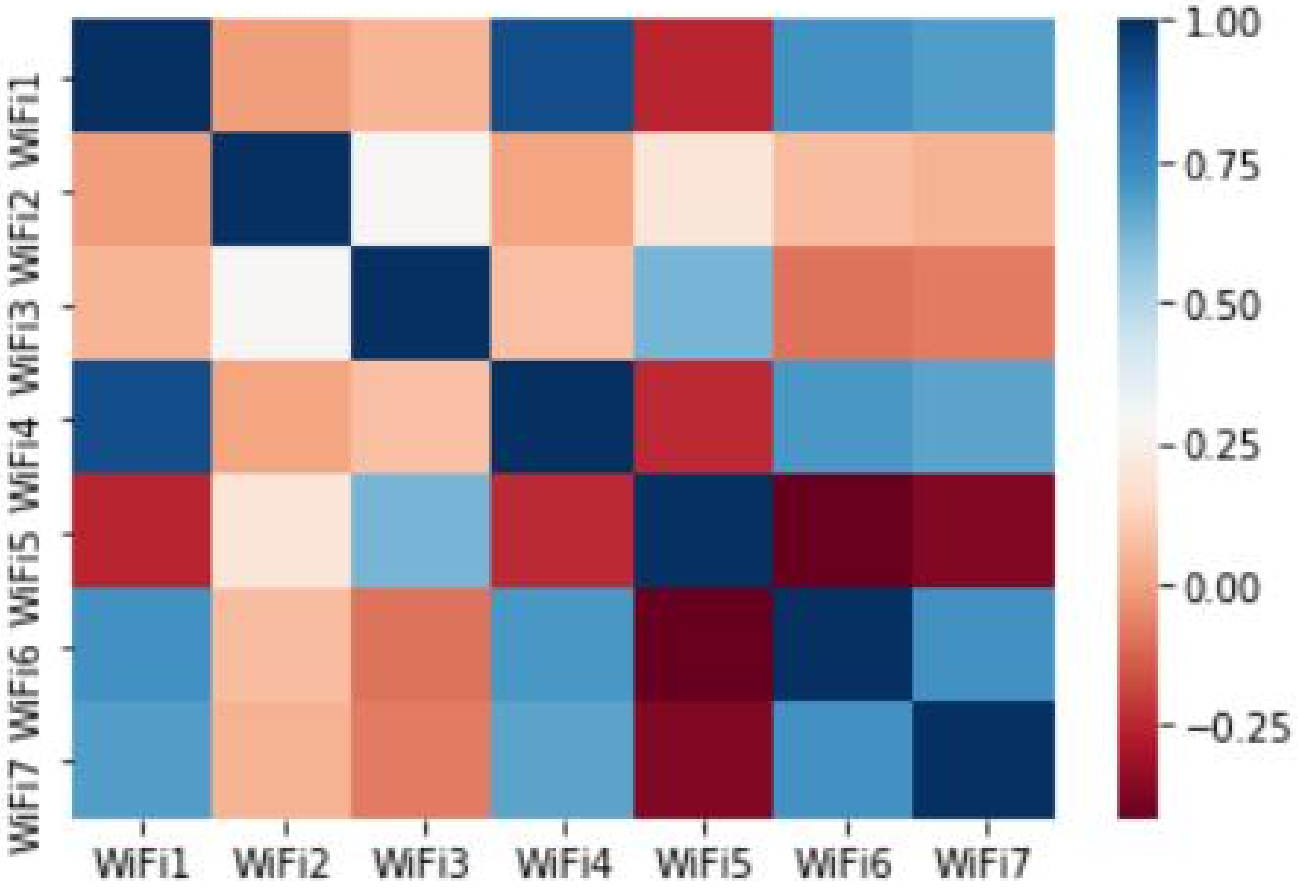
```
for i in range(1,8):
    X[f'WiFi{i}'] = X[f'WiFi{i}'].sample(frac=0.9)
```

	WiFi1	WiFi2	WiFi3	WiFi4	WiFi5	WiFi6	WiFi7
0	-64.0	-56.0	-61.0	-66.0	-71.0	-82.0	-81.0
1	NaN	-57.0	-61.0	-65.0	-71.0	-85.0	-85.0
2	-63.0	-60.0	NaN	-67.0	-76.0	-85.0	-84.0
3	NaN	-60.0	-68.0	-62.0	-77.0	-90.0	-80.0
4	NaN	-65.0	-60.0	-63.0	-77.0	-81.0	-87.0

The obtained file is exported as a csv file for further use

Plotting the Correlation Heatmap between features:

- Some features are correlated
- No. of features are less
- We keep all the features



Imputation Methods:

- Mean
- Median
- KNN regression
- Decision Tree

Classification Models:

- LDA
- QDA
- SVM
- Decision tree
- KNN

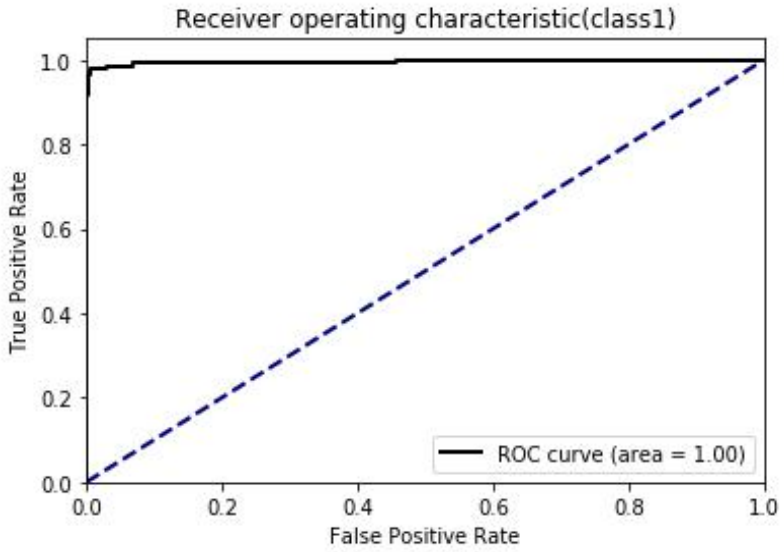
Approach:

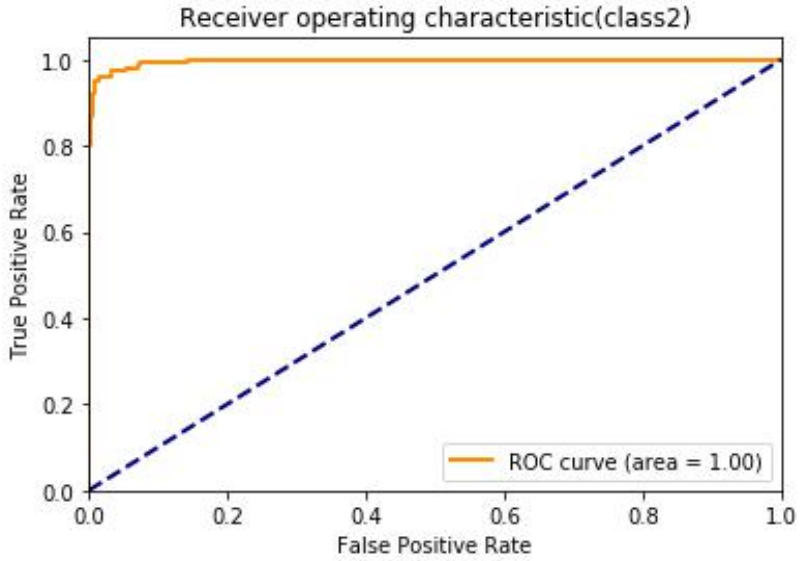
- All the above mentioned Classification Models are created for each Imputation techniques.
- They are compared to find the most accurate model.

Imputation Method → Mean

Model	Accuracy Score	Confusion Matrix
Decision Tree	0.950	Confusion Matrix: [[134 0 1 7] [0 148 9 0] [2 7 133 4] [0 0 0 155]]
SVM	0.701	Confusion Matrix: [[86 56 0 0] [0 155 2 0] [0 59 87 0] [1 61 0 93]]
LDA	0.933	Confusion Matrix: [[137 0 4 1] [1 136 20 0] [1 1 141 3] [2 0 7 146]]
QDA	0.968	Confusion Matrix: [[140 0 1 1] [0 151 6 0] [3 2 136 5] [0 0 1 154]]
KNN	0.966	Confusion Matrix: [[139 0 1 2] [0 150 7 0] [1 1 141 3] [2 0 3 150]]

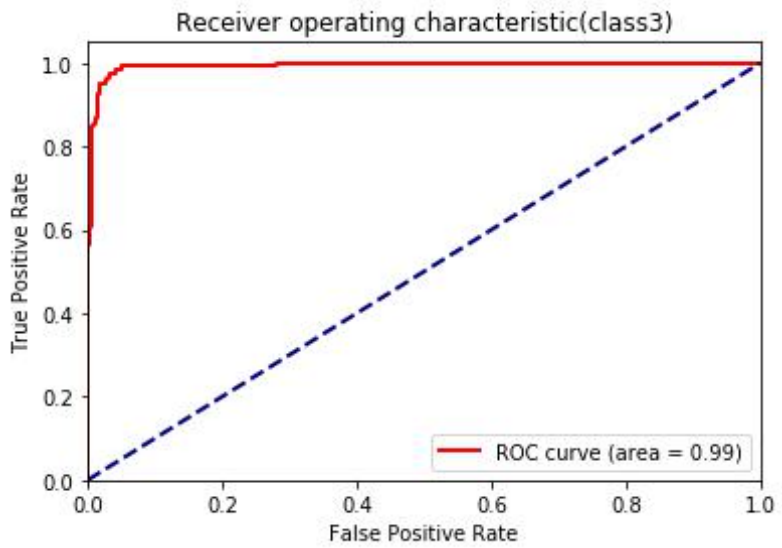
Best Model → QDA

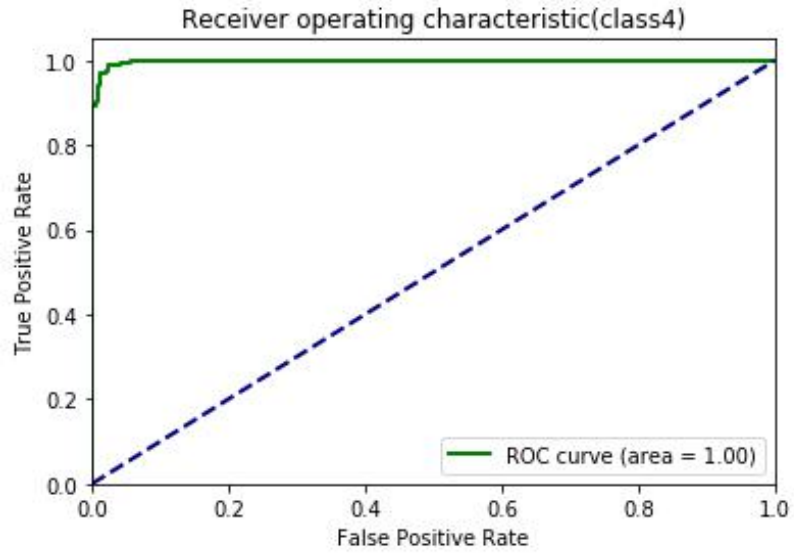





```
Confusion Matrix:
[[140   0   1   1]
 [  0 151   6   0]
 [  3   2 136   5]
 [  0   0   1 154]]
```

Accuracy Score: 0.968





K-fold Cross Validation

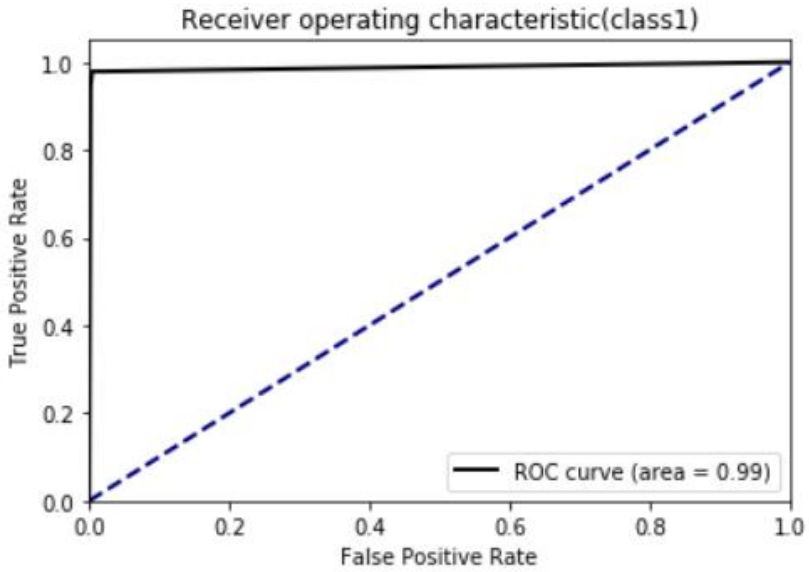
Max Accuracy: 0.969

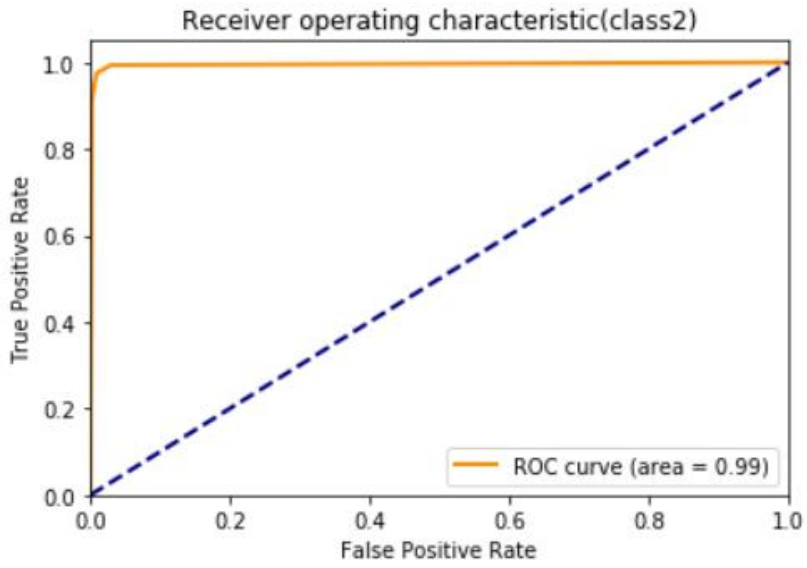
No. Folds: 340

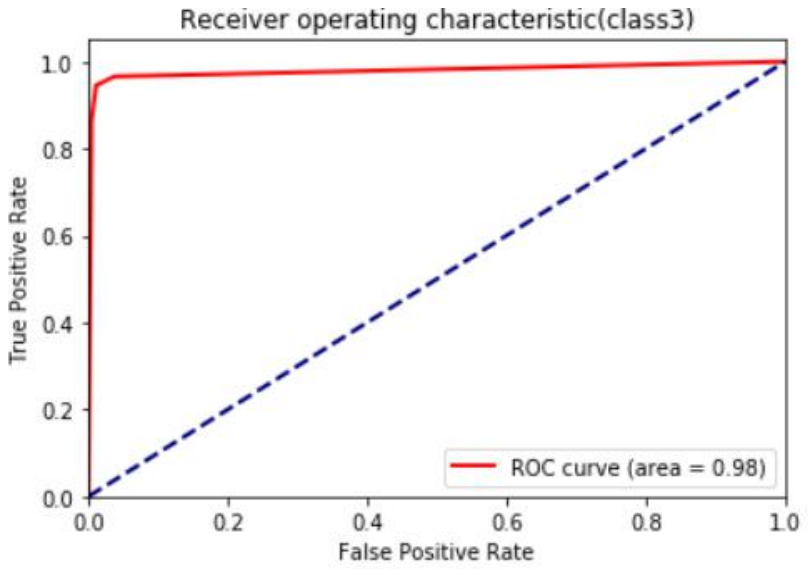
Imputation Method → Decision Tree

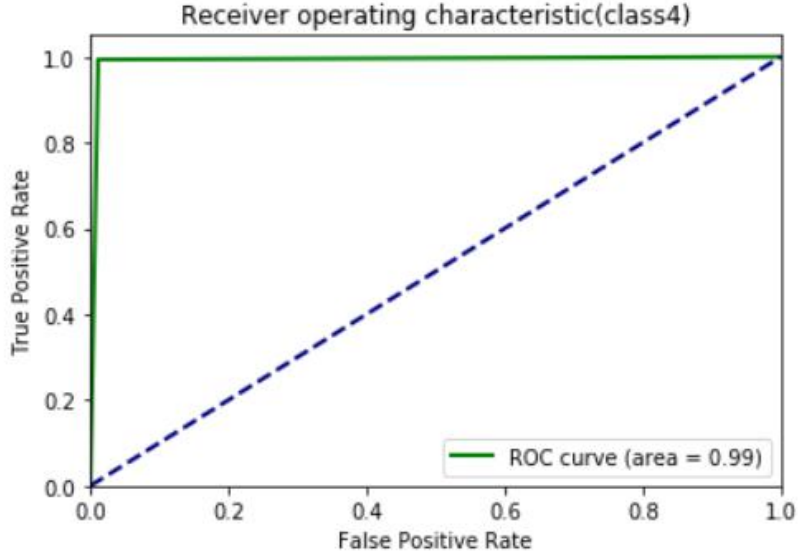
Model	Accuracy Score	Confusion Matrix
Decision Tree	0.953	Confusion Matrix: [[137 0 1 4] [0 150 7 0] [4 7 131 4] [0 0 1 154]]
SVM	0.770	Confusion Matrix: [[99 42 0 1] [0 156 1 0] [0 53 91 2] [0 39 0 116]]
LDA	0.958	Confusion Matrix: [[139 0 2 1] [0 141 16 0] [2 1 141 2] [0 0 1 154]]
QDA	0.968	Confusion Matrix: [[140 0 1 1] [0 153 4 0] [2 5 134 5] [0 0 1 154]]
KNN	0.973	Confusion Matrix: [[139 1 0 2] [0 153 4 0] [2 3 138 3] [0 0 1 154]]

Best Model → KNN










```
Confusion Matrix:
[[139   1   0   2]
 [  0 153   4   0]
 [  2   3 138   3]
 [  0   0   1 154]]
```

Accuracy Score: 0.973

K-fold Cross Validation

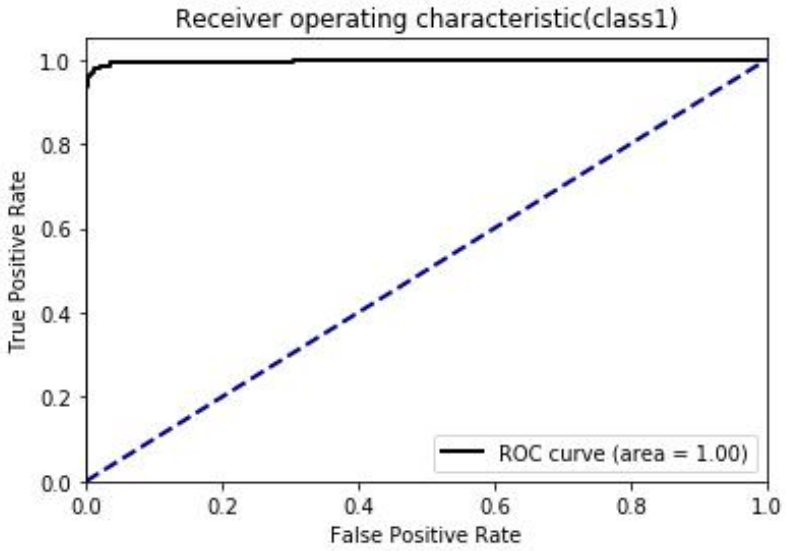
Max Accuracy: 0.975

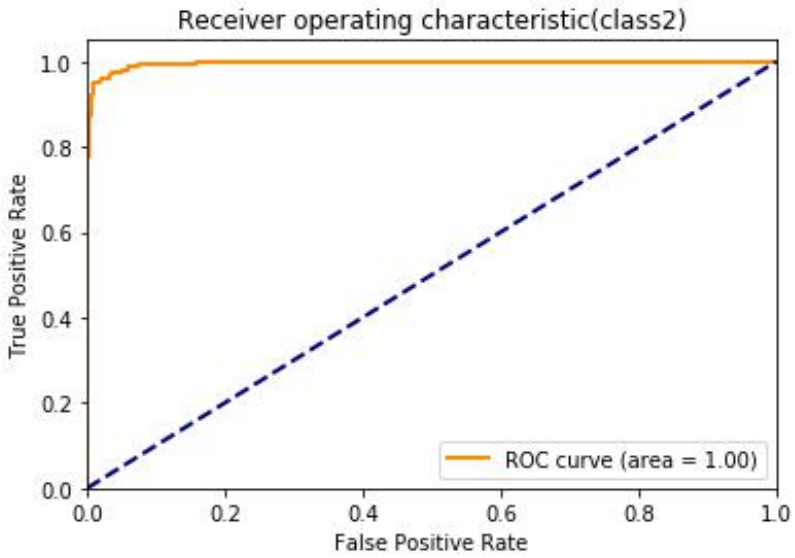
No. Folds: 340

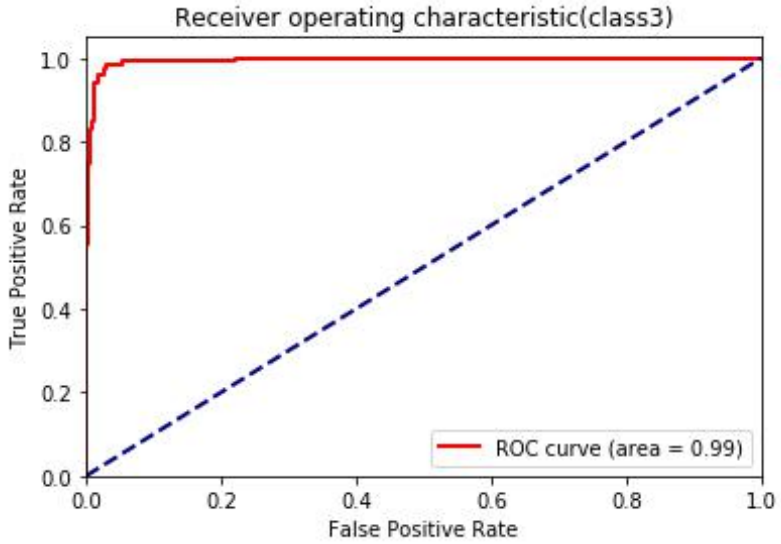
Imputation Method → Median

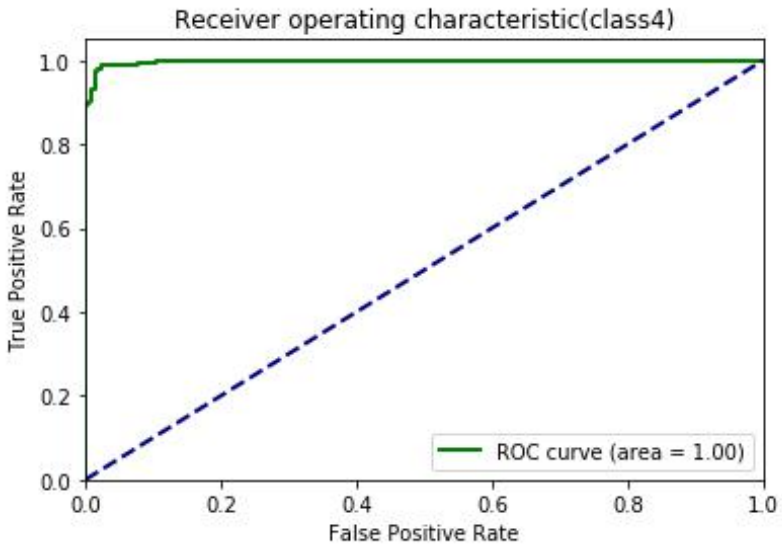
Model	Accuracy Score	Confusion Matrix
Decision Tree	0.948	Confusion Matrix: [[137 0 1 4] [0 147 10 0] [1 7 133 5] [1 0 2 152]]
SVM	0.708	Confusion Matrix: [[90 52 0 0] [0 156 1 0] [0 62 83 1] [1 58 0 96]]
LDA	0.928	Confusion Matrix: [[138 0 4 0] [2 133 22 0] [1 1 141 3] [2 0 8 145]]
QDA	0.970	Confusion Matrix: [[140 0 1 1] [0 151 6 0] [2 2 137 5] [0 0 1 154]]
KNN	0.948	Confusion Matrix: [[141 0 1 0] [1 144 12 0] [2 0 142 2] [1 0 1 153]]

Best Model → QDA










```
Confusion Matrix:
[[140   0   1   1]
 [  0 151   6   0]
 [  2   2 137   5]
 [  0   0   1 154]]
```

Accuracy Score: 0.970

K-fold Cross Validation

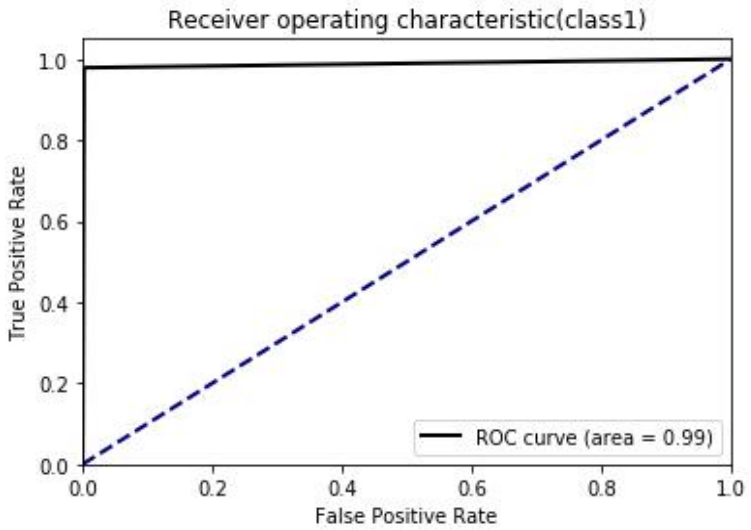
Max Accuracy: 0.972

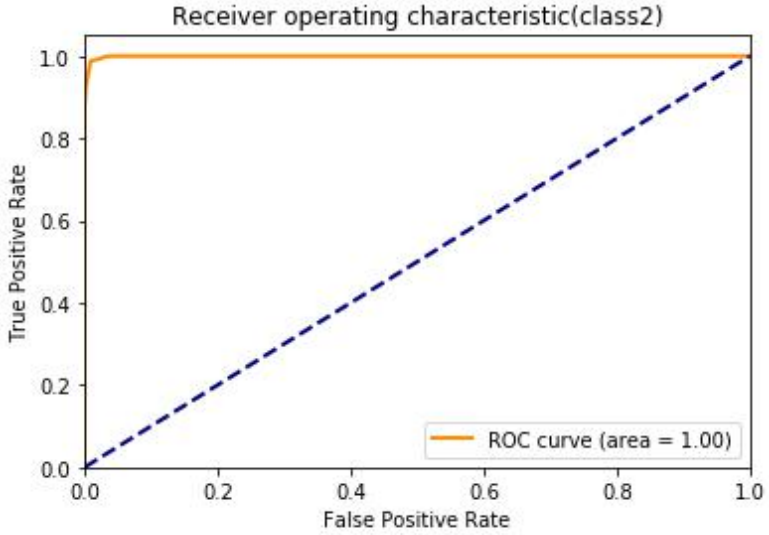
No. Folds: 340

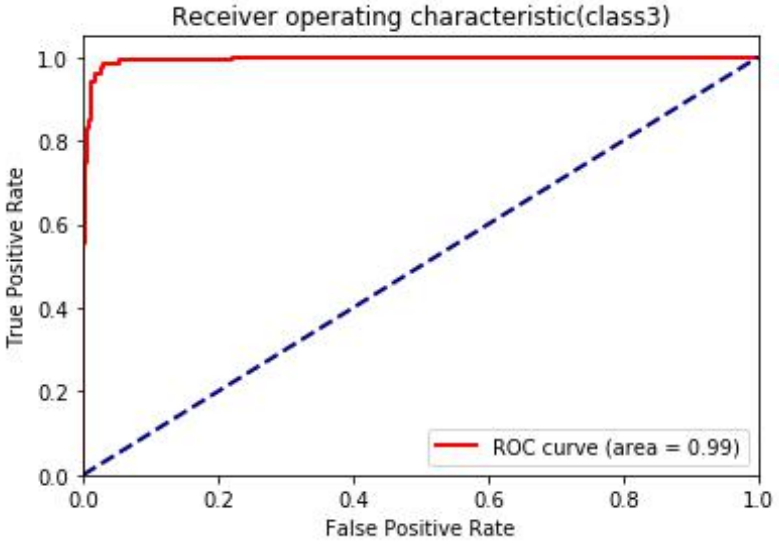
Imputation Method → KNN

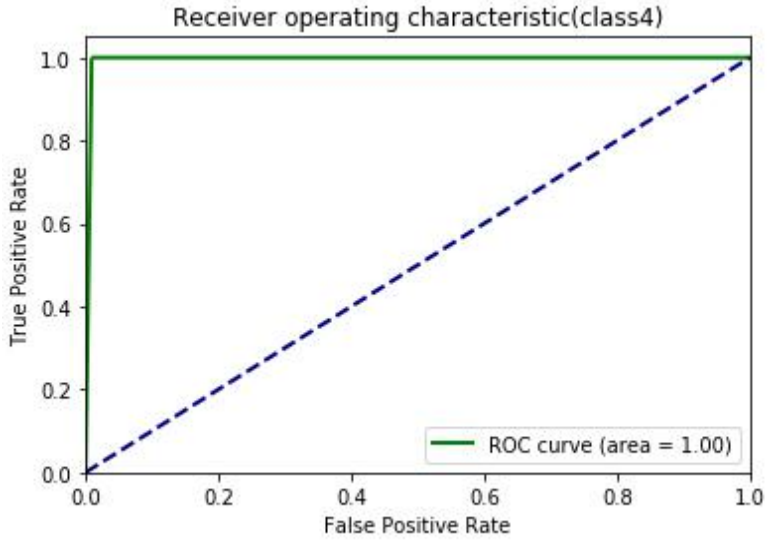
Model	Accuracy Score	Confusion Matrix
Decision Tree	0.953	Confusion Matrix: [[135 0 1 6] [0 149 8 0] [2 7 134 3] [0 0 1 154]]
SVM	0.791	Confusion Matrix: [[106 34 0 2] [0 157 0 0] [0 47 98 1] [0 40 1 114]]
LDA	0.960	Confusion Matrix: [[138 0 2 2] [0 141 16 0] [1 0 143 2] [0 0 1 154]]
QDA	0.973	Confusion Matrix: [[139 0 1 2] [0 152 5 0] [1 2 139 4] [0 0 1 154]]
KNN	0.976	Confusion Matrix: [[139 0 1 2] [0 155 2 0] [2 4 138 2] [0 0 1 154]]

Best Model → KNN










```
Confusion Matrix:
[[139   0   1   2]
 [  0 155   2   0]
 [  2   4 138   2]
 [  0   0   1 154]]
```

Accuracy Score: 0.976

K-fold Cross Validation

Max Accuracy: 0.977

No. Folds: 346

Imputation Method	Mean	Median	Decision Tree	KNN
Best Model	QDA	QDA	KNN	KNN
Accuracy	0.968	0.970	0.973	0.976
K-Fold Accuracy	0.969	0.972	0.975	0.977