CS 6375 Project - Results

Table of Results

Key:

- Green section - Logistic regression experiments; the learning rate and iterations were changed for each experiment
- Yellow section - KNN experiments; the value of K was changed for each experiment
- Blue section - Decision trees experiments; the value of min number of samples and max tree depth were changed for each experiment
- "**" next to the experiment number indicates the experiment that had the best results

Experiment #	Algorithm Used	Parameter(s)	Results
1	Logistic Regression	Learning rate = 0.001 Iterations = 1000	Train/Test Split: 80:20 Size of dataset: 10,000 Accuracy: 0.546 Precision: 0.6335211226942232 Recall: 0.5424124329072448 F1 Score: 0.5120471063217218
2	Logistic Regression	Learning rate = 0.01 Iterations = 1000	Train/Test Split: 80:20 Size of dataset: 10,000 Accuracy: 0.386 Precision: 0.7979594957489058 Recall: 0.386491794405048 F1 Score: 0.27596970424722783
3	Logistic Regression	Learning rate = 0.1 Iterations = 1000	Train/Test Split: 80:20 Size of dataset: 10,000 Accuracy: 0.329 Precision: 0.7667231933386718 Recall: 0.32995289426551316 F1 Score: 0.22928512488906427
4	Logistic Regression	Learning rate = 0.001 Iterations = 500	Train/Test Split: 80:20 Size of dataset: 10,000 Accuracy: 0.507 Precision: 0.742332407689565 Recall: 0.508280403289354 F1 Score: 0.4293330452842256
5 **	Logistic Regression	Learning rate = 0.001 Iterations = 2000	Train/Test Split: 80:20 Size of dataset: 10,000

CS 6375 Project - Results

			Accuracy: 0.5465 Precision: 0.6227017606815628 Recall: 0.5455614503077837 F1 Score: 0.49790568850624095
1	KNN	K = 5	Train/Test Split: 80:20 Size of dataset: 10,000 Accuracy: 0.608 Precision: 0.6027969816019223 Recall: 0.6034633518408842 F1 Score: 0.6000491684695035
2	KNN	K = 10	Train/Test Split: 80:20 Size of dataset: 10,000 Accuracy: 0.624 Precision: 0.6172530739873586 Recall: 0.6192791730145735 F1 Score: 0.6139032082286967
3	KNN	K = 15	Train/Test Split: 80:20 Size of dataset: 10,000 Accuracy: 0.63 Precision: 0.6231969963679339 Recall: 0.6249314712016996 F1 Score: 0.6191021493661564
4	KNN	K = 20	Train/Test Split: 80:20 Size of dataset: 10,000 Accuracy: 0.63 Precision: 0.6226206953386234 Recall: 0.6253107559590814 F1 Score: 0.6184034275335252
5 **	KNN	K = 19	Train/Test Split: 80:20 Size of dataset: 10,000 Accuracy: 0.6335 Precision: 0.6275920912764241 Recall: 0.6288817473992449 F1 Score: 0.621962746662528

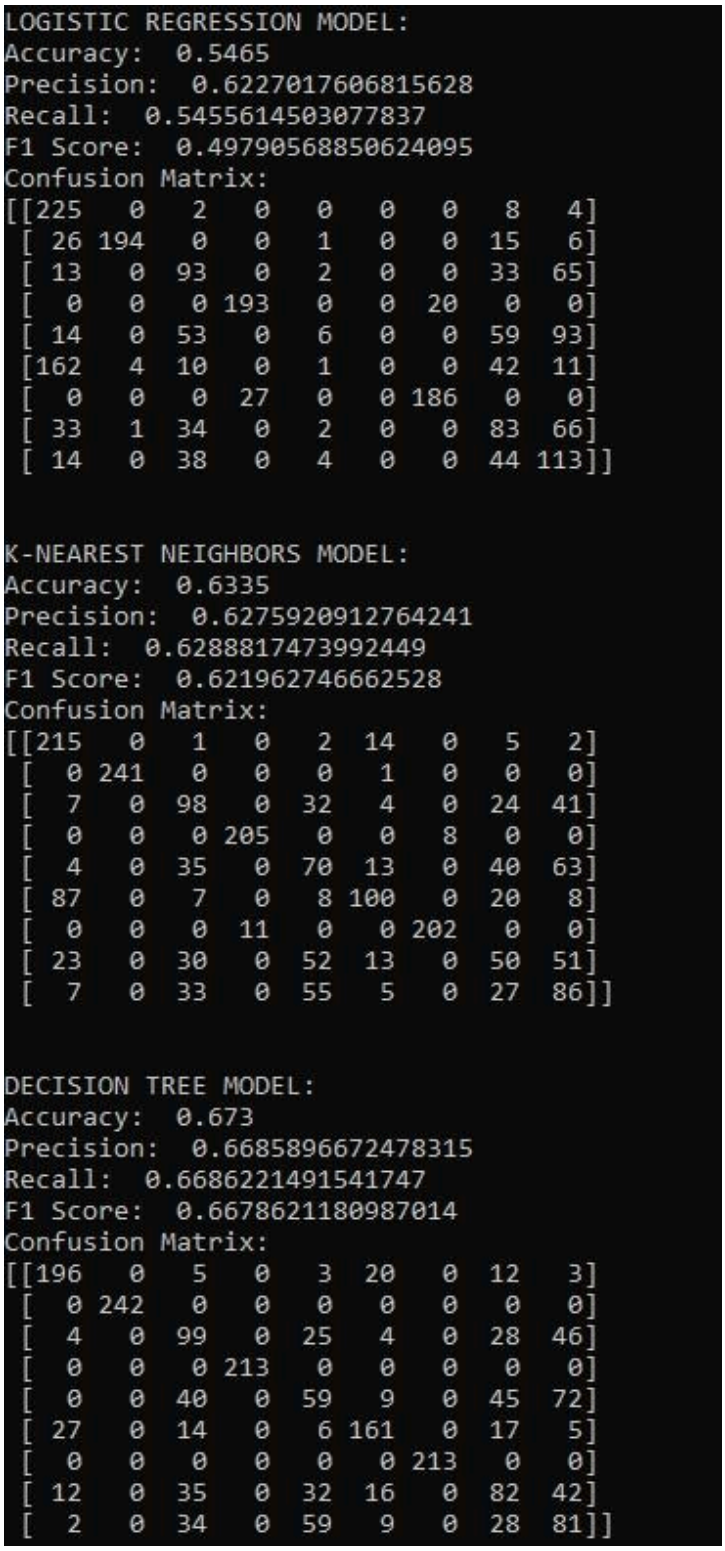
CS 6375 Project - Results

---

1 **	Decision Trees	Min # of samples = 3 Max tree depth = 100	Train/Test Split: 80:20 Size of dataset: 10,000 Accuracy: 0.677 Precision: 0.6713338748380095 Recall: 0.6724877813055057 F1 Score: 0.6712170265704344
2	Decision Trees	Min # of samples = 5 Max tree depth = 100	Train/Test Split: 80:20 Size of dataset: 10,000 Accuracy: 0.676 Precision: 0.6730410785657202 Recall: 0.6720159059008721 F1 Score: 0.671540879369559
3	Decision Trees	Min # of samples = 1 Max tree depth = 100	Train/Test Split: 80:20 Size of dataset: 10,000 Accuracy: 0.674 Precision: 0.6704544060547161 Recall: 0.6698148690011201 F1 Score: 0.6692339024956581
4	Decision Trees	Min # of samples = 3 Max tree depth = 50	Train/Test Split: 80:20 Size of dataset: 10,000 Accuracy: 0.6715 Precision: 0.6681703494787898 Recall: 0.6671840945376403 F1 Score: 0.6671166742671333
5	Decision Trees	Min # of samples = 3 Max tree depth = 200	Train/Test Split: 80:20 Size of dataset: 10,000 Accuracy: 0.674 Precision: 0.6713918547284184 Recall: 0.6696541177953734 F1 Score: 0.6697396076678773

CS 6375 Project - Results

Screenshot of Program Output



```
LOGISTIC REGRESSION MODEL:
Accuracy:  0.5465
Precision:  0.6227017606815628
Recall:  0.5455614503077837
F1 Score:  0.49790568850624095
Confusion Matrix:
[[225   0   2   0   0   0   0   8   4]
 [ 26 194   0   0   1   0   0  15   6]
 [ 13   0  93   0   2   0   0  33  65]
 [  0   0   0 193   0   0  20   0   0]
 [ 14   0  53   0   6   0   0  59  93]
 [162   4  10   0   1   0   0  42  11]
 [  0   0   0  27   0   0 186   0   0]
 [ 33   1  34   0   2   0   0  83  66]
 [ 14   0  38   0   4   0   0  44 113]]


K-NEAREST NEIGHBORS MODEL:
Accuracy:  0.6335
Precision:  0.6275920912764241
Recall:  0.6288817473992449
F1 Score:  0.621962746662528
Confusion Matrix:
[[215   0   1   0   2  14   0   5   2]
 [  0 241   0   0   0   1   0   0   0]
 [  7   0  98   0  32   4   0  24  41]
 [  0   0   0 205   0   0   8   0   0]
 [  4   0  35   0  70  13   0  40  63]
 [ 87   0   7   0   8 100   0  20   8]
 [  0   0   0  11   0   0 202   0   0]
 [ 23   0  30   0  52  13   0  50  51]
 [  7   0  33   0  55   5   0  27  86]]


DECISION TREE MODEL:
Accuracy:  0.673
Precision:  0.6685896672478315
Recall:  0.6686221491541747
F1 Score:  0.6678621180987014
Confusion Matrix:
[[196   0   5   0   3  20   0  12   3]
 [  0 242   0   0   0   0   0   0   0]
 [  4   0  99   0  25   4   0  28  46]
 [  0   0   0 213   0   0   0   0   0]
 [  0   0  40   0  59   9   0  45  72]
 [ 27   0  14   0   6 161   0  17   5]
 [  0   0   0   0   0   0 213   0   0]
 [ 12   0  35   0  32  16   0  82  42]
 [  2   0  34   0  59   9   0  28  81]]
```