Assignment 2 Report

Drake Storm Hackett, 000783796

# Data Description

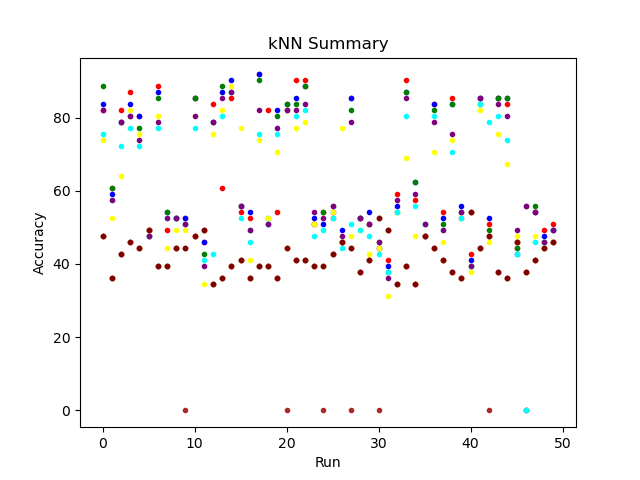
The Algerian Forest Fires Dataset, <https://archive.ics.uci.edu/ml/datasets/Algerian+Forest+Fires+Dataset++#>. This dataset has 12 features including the day and month (June – September of 2012), the temperature in degrees Celsius (22-42), relative humidity (RH) in % (21 - 90), wind speed (Ws) in km/h (6 - 29), rain total for the day in mm (0-16.8), fine fuel moisture code (FFMC) index from the FWI system (28.6-92.5), duff moisture code (DMC) index from the FWI system (1.1-65.9), drought code (DC) index from the FWI system (7-220.4), initial spread index (ISI) index from the FWI system (0-18.5), buildup index (BUI) index from the FWI system (1.1-68), fire weather index (FWI, 0-31.1), then it has the labels or classes of either fire (138 instances) or not fire (106 instances). The task is to use the features to identify fires.

# The Tests

## Part 1: k-Nearest Neighbour

For kNN I did 50 runs of 12 different versions. Changing values for K (3, 7, and 15), distance calculation (Euclidian or manhattan), and normalized data or non-normalized. The training/testing split was done using train\_test\_split(theData, labels, train\_size=0.75) for each run.

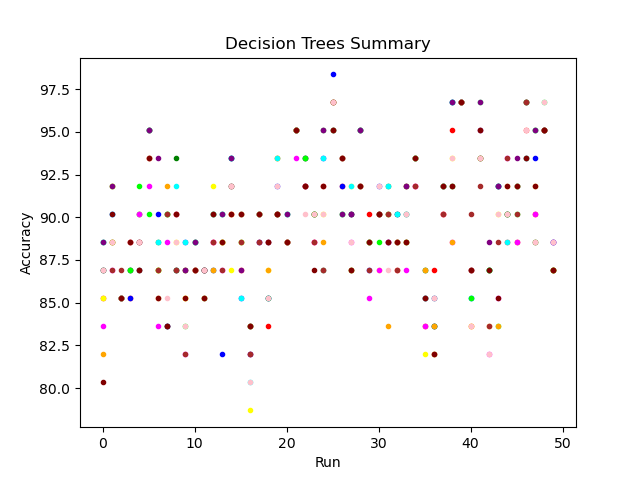
|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | K=3  Euclidean | K=7  Euclidean | K=15  Euclidean | K=3  Manhattan | K=7  Manhattan | K=15  Manhattan | K=3  Euclidean  normalized | K=7  Euclidean  normalized | K=15  Euclidean  normalized | K=3  Manhattan  normalized | K=7  Manhattan  normalized | K=15  Manhattan  normalized |
| Avg | 63.87 | 64.2 | 64.69 | 58.62 | 59.21 | 63.21 | 41.97 | 41.97 | 41.97 | 36.52 | 41.97 | 41.97 |  |
| 1 | 81.97 | 83.61 | 88.52 | 73.77 | 75.41 | 81.97 | 47.54 | 47.54 | 47.54 | 47.54 | 47.54 | 47.54 |  |
| 2 | 60.66 | 59.02 | 60.66 | 52.46 | 57.38 | 57.38 | 36.07 | 36.07 | 36.07 | 36.07 | 36.07 | 36.07 |  |
| 3 | 81.97 | 78.69 | 78.69 | 63.93 | 72.13 | 78.69 | 42.62 | 42.62 | 42.62 | 42.62 | 42.62 | 42.62 |  |
| 4 | 86.89 | 83.61 | 80.33 | 81.97 | 77.05 | 80.33 | 45.9 | 45.9 | 45.9 | 45.9 | 45.9 | 45.9 |  |
| 5 | 80.33 | 80.33 | 77.05 | 75.41 | 72.13 | 73.77 | 44.26 | 44.26 | 44.26 | 44.26 | 44.26 | 44.26 |  |
| 6 | 49.18 | 47.54 | 47.54 | 47.54 | 47.54 | 47.54 | 49.18 | 49.18 | 49.18 | 49.18 | 49.18 | 49.18 |  |
| 7 | 88.52 | 86.89 | 85.25 | 80.33 | 77.05 | 78.69 | 39.34 | 39.34 | 39.34 | 39.34 | 39.34 | 39.34 |  |
| 8 | 49.18 | 54.1 | 54.1 | 44.26 | 52.46 | 52.46 | 39.34 | 39.34 | 39.34 | 39.34 | 39.34 | 39.34 |  |
| 9 | 52.46 | 52.46 | 52.46 | 49.18 | 52.46 | 52.46 | 44.26 | 44.26 | 44.26 | 44.26 | 44.26 | 44.26 |  |
| 10 | 52.46 | 52.46 | 50.82 | 49.18 | 50.82 | 50.82 | 44.26 | 44.26 | 44.26 | 0. | 44.26 | 44.26 |  |
| 11 | 85.25 | 85.25 | 85.25 | 80.33 | 77.05 | 80.33 | 47.54 | 47.54 | 47.54 | 47.54 | 47.54 | 47.54 |  |
| 12 | 45.9 | 45.9 | 42.62 | 34.43 | 40.98 | 39.34 | 49.18 | 49.18 | 49.18 | 49.18 | 49.18 | 49.18 |  |
| 13 | 83.61 | 78.69 | 78.69 | 75.41 | 42.62 | 78.69 | 34.43 | 34.43 | 34.43 | 34.43 | 34.43 | 34.43 |  |
| 14 | 60.66 | 86.89 | 88.52 | 81.97 | 80.33 | 85.25 | 36.07 | 36.07 | 36.07 | 36.07 | 36.07 | 36.07 |  |
| 15 | 85.25 | 90.16 | 86.89 | 88.52 | 86.89 | 86.89 | 39.34 | 39.34 | 39.34 | 39.34 | 39.34 | 39.34 |  |
| 16 | 54.1 | 55.74 | 55.74 | 77.05 | 52.46 | 55.74 | 40.98 | 40.98 | 40.98 | 40.98 | 40.98 | 40.98 |  |
| 17 | 52.46 | 54.1 | 49.18 | 40.98 | 45.9 | 49.18 | 36.07 | 36.07 | 36.07 | 36.07 | 36.07 | 36.07 |  |
| 18 | 91.8 | 91.8 | 90.16 | 73.77 | 75.41 | 81.97 | 39.34 | 39.34 | 39.34 | 39.34 | 39.34 | 39.34 |  |
| 19 | 81.97 | 50.82 | 52.46 | 52.46 | 50.82 | 50.82 | 39.34 | 39.34 | 39.34 | 39.34 | 39.34 | 39.34 |  |
| 20 | 54.1 | 81.97 | 80.33 | 70.49 | 75.41 | 77.05 | 36.07 | 36.07 | 36.07 | 36.07 | 36.07 | 36.07 |  |
| 21 | 81.97 | 83.61 | 83.61 | 81.97 | 81.97 | 81.97 | 44.26 | 44.26 | 44.26 | 0. | 44.26 | 44.26 |  |
| 22 | 90.16 | 85.25 | 83.61 | 77.05 | 80.33 | 81.97 | 40.98 | 40.98 | 40.98 | 40.98 | 40.98 | 40.98 |  |
| 23 | 90.16 | 88.52 | 88.52 | 78.69 | 81.97 | 83.61 | 40.98 | 40.98 | 40.98 | 40.98 | 40.98 | 40.98 |  |
| 24 | 50.82 | 52.46 | 50.82 | 50.82 | 47.54 | 54.1 | 39.34 | 39.34 | 39.34 | 39.34 | 39.34 | 39.34 |  |
| 25 | 54.1 | 50.82 | 54.1 | 49.18 | 49.18 | 52.46 | 39.34 | 39.34 | 39.34 | 0. | 39.34 | 39.34 |  |
| 26 | 52.46 | 54.1 | 55.74 | 54.1 | 52.46 | 55.74 | 42.62 | 42.62 | 42.62 | 42.62 | 42.62 | 42.62 |  |
| 27 | 45.9 | 49.18 | 45.9 | 77.05 | 44.26 | 47.54 | 45.9 | 45.9 | 45.9 | 45.9 | 45.9 | 45.9 |  |
| 28 | 85.25 | 85.25 | 81.97 | 47.54 | 50.82 | 78.69 | 44.26 | 44.26 | 44.26 | 0. | 44.26 | 44.26 |  |
| 29 | 52.46 | 52.46 | 52.46 | 49.18 | 49.18 | 52.46 | 37.7 | 37.7 | 37.7 | 37.7 | 37.7 | 37.7 |  |
| 30 | 50.82 | 54.1 | 50.82 | 42.62 | 47.54 | 50.82 | 40.98 | 40.98 | 40.98 | 40.98 | 40.98 | 40.98 |  |
| 31 | 44.26 | 44.26 | 45.9 | 44.26 | 42.62 | 45.9 | 52.46 | 52.46 | 52.46 | 0. | 52.46 | 52.46 |  |
| 32 | 40.98 | 39.34 | 37.7 | 31.15 | 37.7 | 36.07 | 49.18 | 49.18 | 49.18 | 49.18 | 49.18 | 49.18 |  |
| 33 | 59.02 | 55.74 | 54.1 | 54.1 | 54.1 | 57.38 | 34.43 | 34.43 | 34.43 | 34.43 | 34.43 | 34.43 |  |
| 34 | 90.16 | 86.89 | 86.89 | 68.85 | 80.33 | 85.25 | 39.34 | 39.34 | 39.34 | 39.34 | 39.34 | 39.34 |  |
| 35 | 57.38 | 62.3 | 62.3 | 47.54 | 55.74 | 59.02 | 34.43 | 34.43 | 34.43 | 34.43 | 34.43 | 34.43 |  |
| 36 | 47.54 | 47.54 | 50.82 | 47.54 | 50.82 | 50.82 | 47.54 | 47.54 | 47.54 | 47.54 | 47.54 | 47.54 |  |
| 37 | 83.61 | 83.61 | 81.97 | 70.49 | 80.33 | 78.69 | 44.26 | 44.26 | 44.26 | 44.26 | 44.26 | 44.26 |  |
| 38 | 54.1 | 52.46 | 50.82 | 45.9 | 49.18 | 49.18 | 40.98 | 40.98 | 40.98 | 40.98 | 40.98 | 40.98 |  |
| 39 | 85.25 | 83.61 | 83.61 | 73.77 | 70.49 | 75.41 | 37.7 | 37.7 | 37.7 | 37.7 | 37.7 | 37.7 |  |
| 40 | 52.46 | 55.74 | 54.1 | 52.46 | 52.46 | 54.1 | 36.07 | 36.07 | 36.07 | 36.07 | 36.07 | 36.07 |  |
| 41 | 42.62 | 40.98 | 39.34 | 37.7 | 39.34 | 39.34 | 54.1 | 54.1 | 54.1 | 54.1 | 54.1 | 54.1 |  |
| 42 | 85.25 | 85.25 | 83.61 | 81.97 | 83.61 | 85.25 | 44.26 | 44.26 | 44.26 | 44.26 | 44.26 | 44.26 |  |
| 43 | 50.82 | 52.46 | 49.18 | 45.9 | 78.69 | 47.54 | 47.54 | 47.54 | 47.54 | 0. | 47.54 | 47.54 |  |
| 44 | 85.25 | 85.25 | 85.25 | 75.41 | 80.33 | 83.61 | 37.7 | 37.7 | 37.7 | 37.7 | 37.7 | 37.7 |  |
| 45 | 83.61 | 85.25 | 85.25 | 67.21 | 73.77 | 80.33 | 36.07 | 36.07 | 36.07 | 36.07 | 36.07 | 36.07 |  |
| 46 | 44.26 | 42.62 | 44.26 | 47.54 | 42.62 | 49.18 | 45.9 | 45.9 | 45.9 | 45.9 | 45.9 | 45.9 |  |
| 47 | 0. | 0. | 55.74 | 0. | 0. | 55.74 | 37.7 | 37.7 | 37.7 | 37.7 | 37.7 | 37.7 |  |
| 48 | 54.1 | 54.1 | 55.74 | 47.54 | 45.9 | 54.1 | 40.98 | 40.98 | 40.98 | 40.98 | 40.98 | 40.98 |  |
| 49 | 49.18 | 47.54 | 45.9 | 44.26 | 45.9 | 45.9 | 44.26 | 44.26 | 44.26 | 44.26 | 44.26 | 44.26 |  |
| 50 | 50.82 | 49.18 | 49.18 | 45.90 | 49.18 | 49.18 | 45.9 | 45.9 | 45.9 | 45.9 | 45.9 | 45.9 |  |



## Part 2: Decision Trees

For the decision tree I did 50 runs of 12 different versions. Changing values for criterion (gini, entropy), max depth (None, 5), and min sample split (2, 4, 6). The training/testing split was done using train\_test\_split(theData, labels, train\_size=0.75) for each run.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | gini  None  2 | gini  None  4 | gini  nONE  6 | gini  5  2 | gini  5  4 | gini  5  6 | Entropy  None  2 | Entropy  None  4 | Entropy  None  6 | Entropy  5  2 | Entropy  5  4 | Entropy  5  6 |
| Avg | 89.61 | 89.87 | 90.56 | 89.84 | 90.49 | 90.49 | 88.89 | 89.34 | 89.15 | 89.21 | 89.44 | 89.28 |  |
| 1 | 85.25 | 86.89 | 88.52 | 85.25 | 88.52 | 88.52 | 83.61 | 86.89 | 81.97 | 86.89 | 86.89 | 80.33 |  |
| 2 | 88.52 | 90.16 | 91.80 | 91.80 | 90.16 | 91.80 | 86.89 | 88.52 | 88.52 | 86.89 | 88.52 | 90.16 |  |
| 3 | 85.25 | 85.25 | 85.25 | 85.25 | 85.25 | 85.25 | 85.25 | 85.25 | 85.25 | 86.89 | 85.25 | 85.25 |  |
| 4 | 85.25 | 85.25 | 86.89 | 86.89 | 86.89 | 86.89 | 88.52 | 86.89 | 88.52 | 88.52 | 88.52 | 88.52 |  |
| 5 | 90.16 | 88.52 | 86.89 | 90.16 | 90.16 | 86.89 | 90.16 | 91.80 | 88.52 | 88.52 | 88.52 | 86.89 |  |
| 6 | 95.08 | 95.08 | 95.08 | 95.08 | 95.08 | 95.08 | 91.80 | 90.16 | 93.44 | 93.44 | 93.44 | 93.44 |  |
| 7 | 86.89 | 90.16 | 88.52 | 86.89 | 88.52 | 93.44 | 83.61 | 86.89 | 85.25 | 86.89 | 85.25 | 85.25 |  |
| 8 | 90.16 | 83.61 | 83.61 | 90.16 | 90.16 | 83.61 | 88.52 | 90.16 | 91.80 | 90.16 | 85.25 | 83.61 |  |
| 9 | 86.89 | 86.89 | 93.44 | 88.52 | 91.80 | 90.16 | 90.16 | 86.89 | 90.16 | 86.89 | 88.52 | 90.16 |  |
| 10 | 86.89 | 88.52 | 86.89 | 85.25 | 88.52 | 86.89 | 81.97 | 83.61 | 85.25 | 81.97 | 83.61 | 85.25 |  |
| 11 | 86.89 | 88.52 | 88.52 | 88.52 | 88.52 | 88.52 | 86.89 | 86.89 | 86.89 | 86.89 | 86.89 | 86.89 |  |
| 12 | 86.89 | 86.89 | 86.89 | 86.89 | 86.89 | 86.89 | 85.25 | 85.25 | 85.25 | 86.89 | 86.89 | 85.25 |  |
| 13 | 88.52 | 90.16 | 90.16 | 91.80 | 86.89 | 86.89 | 88.52 | 90.16 | 86.89 | 88.52 | 90.16 | 90.16 |  |
| 14 | 88.52 | 81.97 | 90.16 | 88.52 | 90.16 | 90.16 | 86.89 | 88.52 | 88.52 | 86.89 | 88.52 | 88.52 |  |
| 15 | 90.16 | 93.44 | 93.44 | 86.89 | 93.44 | 93.44 | 91.80 | 91.80 | 90.16 | 91.80 | 91.80 | 90.16 |  |
| 16 | 86.89 | 85.25 | 88.52 | 88.52 | 85.25 | 86.89 | 90.16 | 90.16 | 90.16 | 88.52 | 90.16 | 90.16 |  |
| 17 | 81.97 | 81.97 | 83.61 | 78.69 | 80.33 | 83.61 | 83.61 | 83.61 | 83.61 | 81.97 | 80.33 | 83.61 |  |
| 18 | 88.52 | 88.52 | 90.16 | 90.16 | 90.16 | 90.16 | 90.16 | 90.16 | 90.16 | 88.52 | 90.16 | 90.16 |  |
| 19 | 83.61 | 85.25 | 88.52 | 86.89 | 85.25 | 85.25 | 88.52 | 85.25 | 86.89 | 88.52 | 85.25 | 88.52 |  |
| 20 | 93.44 | 91.80 | 90.16 | 93.44 | 93.44 | 90.16 | 90.16 | 90.16 | 90.16 | 90.16 | 91.80 | 90.16 |  |
| 21 | 88.52 | 88.52 | 90.16 | 88.52 | 90.16 | 90.16 | 88.52 | 88.52 | 88.52 | 88.52 | 88.52 | 88.52 |  |
| 22 | 95.08 | 95.08 | 95.08 | 95.08 | 95.08 | 95.08 | 93.44 | 95.08 | 95.08 | 95.08 | 95.08 | 95.08 |  |
| 23 | 91.80 | 91.80 | 93.44 | 93.44 | 93.44 | 93.44 | 91.80 | 93.44 | 91.80 | 91.80 | 90.16 | 91.80 |  |
| 24 | 88.52 | 90.16 | 90.16 | 90.16 | 90.16 | 90.16 | 88.52 | 90.16 | 90.16 | 88.52 | 90.16 | 86.89 |  |
| 25 | 91.80 | 93.44 | 95.08 | 90.16 | 93.44 | 95.08 | 86.89 | 86.89 | 88.52 | 86.89 | 90.16 | 91.80 |  |
| 26 | 96.72 | 98.36 | 95.08 | 96.72 | 95.08 | 95.08 | 96.72 | 96.72 | 95.08 | 96.72 | 96.72 | 95.08 |  |
| 27 | 91.80 | 91.80 | 90.16 | 90.16 | 90.16 | 90.16 | 93.44 | 93.44 | 93.44 | 93.44 | 93.44 | 93.44 |  |
| 28 | 90.16 | 90.16 | 90.16 | 91.80 | 91.80 | 90.16 | 88.52 | 86.89 | 86.89 | 86.89 | 88.52 | 86.89 |  |
| 29 | 95.08 | 95.08 | 95.08 | 95.08 | 95.08 | 95.08 | 91.80 | 91.80 | 91.80 | 91.80 | 91.80 | 91.80 |  |
| 30 | 90.16 | 86.89 | 88.52 | 88.52 | 88.52 | 88.52 | 85.25 | 88.52 | 88.52 | 86.89 | 88.52 | 88.52 |  |
| 31 | 90.16 | 91.80 | 90.16 | 86.89 | 90.16 | 90.16 | 86.89 | 88.52 | 90.16 | 90.16 | 91.80 | 90.16 |  |
| 32 | 90.16 | 88.52 | 91.80 | 88.52 | 91.80 | 88.52 | 88.52 | 90.16 | 83.61 | 90.16 | 86.89 | 88.52 |  |
| 33 | 86.89 | 90.16 | 90.16 | 88.52 | 90.16 | 88.52 | 88.52 | 88.52 | 88.52 | 86.89 | 88.52 | 88.52 |  |
| 34 | 91.80 | 90.16 | 91.80 | 90.16 | 90.16 | 91.80 | 86.89 | 88.52 | 88.52 | 88.52 | 90.16 | 88.52 |  |
| 35 | 91.80 | 93.44 | 93.44 | 93.44 | 93.44 | 93.44 | 91.80 | 93.44 | 93.44 | 91.80 | 93.44 | 93.44 |  |
| 36 | 83.61 | 85.25 | 85.25 | 81.97 | 85.25 | 86.89 | 83.61 | 86.89 | 86.89 | 85.25 | 85.25 | 85.25 |  |
| 37 | 86.89 | 83.61 | 83.61 | 81.97 | 85.25 | 85.25 | 83.61 | 83.61 | 83.61 | 81.97 | 85.25 | 81.97 |  |
| 38 | 90.16 | 91.80 | 91.80 | 90.16 | 91.80 | 91.80 | 90.16 | 91.80 | 91.80 | 90.16 | 91.80 | 91.80 |  |
| 39 | 95.08 | 96.72 | 96.72 | 93.44 | 96.72 | 96.72 | 88.52 | 91.80 | 88.52 | 91.80 | 93.44 | 91.80 |  |
| 40 | 96.72 | 96.72 | 96.72 | 96.72 | 96.72 | 96.72 | 96.72 | 96.72 | 96.72 | 96.72 | 96.72 | 96.72 |  |
| 41 | 83.61 | 85.25 | 86.89 | 83.61 | 85.25 | 86.89 | 86.89 | 85.25 | 86.89 | 90.16 | 83.61 | 86.89 |  |
| 42 | 93.44 | 93.44 | 95.08 | 96.72 | 96.72 | 96.72 | 95.08 | 93.44 | 93.44 | 91.80 | 93.44 | 95.08 |  |
| 43 | 86.89 | 86.89 | 86.89 | 86.89 | 86.89 | 88.52 | 81.97 | 86.89 | 83.61 | 83.61 | 81.97 | 86.89 |  |
| 44 | 91.80 | 91.80 | 91.80 | 90.16 | 91.80 | 91.80 | 85.25 | 83.61 | 83.61 | 88.52 | 90.16 | 85.25 |  |
| 45 | 88.52 | 90.16 | 90.16 | 90.16 | 88.52 | 90.16 | 93.44 | 90.16 | 91.80 | 93.44 | 90.16 | 91.80 |  |
| 46 | 88.52 | 91.80 | 91.80 | 91.80 | 91.80 | 93.44 | 88.52 | 90.16 | 91.80 | 90.16 | 91.80 | 91.80 |  |
| 47 | 95.08 | 93.44 | 96.72 | 96.72 | 93.44 | 96.72 | 96.72 | 96.72 | 93.44 | 96.72 | 95.08 | 93.44 |  |
| 48 | 90.16 | 93.44 | 95.08 | 95.08 | 95.08 | 95.08 | 90.16 | 88.52 | 91.80 | 88.52 | 88.52 | 91.80 |  |
| 49 | 95.08 | 95.08 | 95.08 | 96.72 | 96.72 | 95.08 | 95.08 | 95.08 | 95.08 | 95.08 | 96.72 | 95.08 |  |
| 50 | 88.52 | 88.52 | 86.89 | 86.89 | 88.52 | 86.89 | 88.52 | 86.89 | 86.89 | 86.89 | 88.52 | 86.89 |  |



# Discussion

Part 1: The version that provided the best accuracy was when k = 15, distance calculation = Euclidian, and the data was not normalized. Between the k values I tested it seemed the higher the value the more accurate it would be. Each Euclidian distance accuracy was higher than any calculated with Manhattan distance. When the data was normalized it greatly decreased the accuracy and none of the other parameters seemed to matter, the accuracies didn’t change but they were the poorest overall. As the measurement features in the dataset are very specific while other features such as the day or month are just referencing the date, this would cause changing the values to not be the best idea which may be why normalizing data did far worse. Therefore, the best accuracy and my recommendation for this kNN algorithm is k: 15, Distance: Euclidian, Normalized: No.

Part 2: The versions that provided the best accuracy was when the criterion = gini, I provided a max depth, and min sample split = 4 or 6. Each version when criterion = entropy did worse than the ones when criterion = gini. Each time I provided a max depth rather than setting it to none the accuracy improved. When changing the min sample split there wasn’t a lot of change in accuracy, especially when changing from 4 to 6, however the better accuracies still came from when it was equal to 4 or 6, mostly 4. The Decision Trees all performed greatly better than the kNN algorithm. This may be due to the fact that the decision tree versions were not including features deemed unimportant for classification, such as the day/month. Overall, the version with the best accuracy for classifying this dataset was the Decision Tree when criterion: gini, max\_depth: 5, min\_sample\_split: 4 and this is what I would recommend out of all these versions.

# Future Work

If I had more time, I would try a few extra versions. Some would be a higher k value to test when the accuracy would decline, and which was most accurate. I would also try some different values for the max-depth in the Decision Trees, here I only did None or 5 as I had 12 versions running 50 times, which showed the having a max depth provided a higher accuracy however I would like to see the effect of different values for max depth.