**Introduction**

The Federalist Papers are a series of 85 essays written by Alexander Hamilton, John Jay, and James Madison between October 1787 and May 1788. The essays were published anonymously, under the pen name "Publius," in various New York state newspapers of the time. The Federalist Papers were written to urge New Yorkers to approve the proposed United States Constitution, which was drafted in Philadelphia in the summer of 1787. In petitioning for adoption of the Constitution over the Articles of Confederation, the essays explain provisions of the Constitution in detail. James Madison's Federalist no. 10 is said to be one of the most important and enduring statements of American political theory. Its reasoned statement explains what an expanding nation might do if it accepted the basic premise of majority rule, a balanced government of three separate branches, and a commitment to balance all the diverse interests through a system of checks and balances. The Federalist Papers are often used to help interpret the intentions of those drafting the Constitution.

The Federalist Papers were published primarily in two New York state newspapers: The New York Packet and The Independent Journal. They were reprinted in other newspapers in New York state and in several cities in other states. A bound edition, with revisions and corrections by Hamilton, was published in 1788 by printers J. and A. McLean. An edition published by printer Jacob Gideon in 1818, with revisions and corrections by Madison, was the first to identify each essay by its author's name. Because of its publishing history, the assignment of authorship, numbering, and exact wording may vary with different editions of The Federalist. 11 essays are authored by “Hamilton or Madison”. These are the famous essays with disputed authorship. Hamilton wrote to claim the authorship before he was killed in a duel. Later Madison also claimed authorship. Historians were trying to find out which one was the real author.

**About the Data**

The text files of the essays are brought in as a corpus. In the author column, there are 74 essays with identified authors: 51 essays written by Hamilton, 15 by Madison, 3 by Hamilton and Madison, 5 by Jay. The remaining 11 essays are authored by “Hamilton or Madison”.

The corpus needs to be turned into a Document Term Matrix. Punctuation, numbers, and separators need to be removed. The text needs to be lowercase. The model was tested with and without stopwords, the results were more accurate with the stopwords removed.

The Document Term Matrix needs to be broken in to different data frames. The papers by Jay are removed. The papers written by Hamilton Madison are removed to not confuse the training model. 5 data frames are created. A Madison Test, a Madison Train, a Hamilton Test, a Hamilton Train, and a Disputed Author data frame. Columns are added to the data frames to indicate whether it is a Madison (0) or Hamilton (1) paper.

The next steps are to combine the test data frames in to a single test data frame and combine the training data frames in to a single train data frame. Then convert the Author columns to factors.

The CSV file has a slightly different approach. The disputed articles are removed from the main data frame and put into a predicting data frame. The Jay and HM articles are removed from the main data frame. Random sampling is done to split the data frame in to a train set with 2/3 of the, the remaining 3rd is made into a testing frame.

**Analysis**

The first J48 model was created with no specifications and produced very poor results.

=== Summary ===

Correctly Classified Instances 0 0 %

Incorrectly Classified Instances 38 77.551 %

Kappa statistic -0.0509

Mean absolute error 0.5

Root mean squared error 0.6996

Relative absolute error 278.6481 %

Root relative squared error 258.9269 %

UnClassified Instances 11 22.449 %

Total Number of Instances 49

=== Confusion Matrix ===

a b c d <-- classified as

0 31 0 1 | a = Hamilton

0 0 2 0 | b = HM

0 0 0 4 | c = Jay

0 0 0 0 | d = Madison

The Weka Classifier using a 10 Fold Cross Validation produced excellent results.

=== 10 Fold Cross Validation ===

=== Summary ===

Correctly Classified Instances 64 96.9697 %

Incorrectly Classified Instances 2 3.0303 %

Kappa statistic 0.9137

Mean absolute error 0.0429

Root mean squared error 0.1766

Relative absolute error 12.0347 %

Root relative squared error 42.0218 %

Total Number of Instances 66

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.980 0.067 0.980 0.980 0.980 0.914 0.949 0.976 Hamilton

0.933 0.020 0.933 0.933 0.933 0.914 0.949 0.831 Madison

Weighted Avg. 0.970 0.056 0.970 0.970 0.970 0.914 0.949 0.943

=== Confusion Matrix ===

a b <-- classified as

50 1 | a = Hamilton

1 14 | b = Madison

The second J48 model also produced poor results, predicting every paper was written by Madison.

=== Summary ===

Correctly Classified Instances 12 27.2727 %

Incorrectly Classified Instances 32 72.7273 %

Kappa statistic 0

Mean absolute error 0.8427

Root mean squared error 0.9122

Relative absolute error 210.0417 %

Root relative squared error 204.7791 %

Total Number of Instances 44

=== Confusion Matrix ===

a b <-- classified as

0 32 | a = Hamilton

0 12 | b = Madison

The Weka Classifier using a 30 Fold Cross Validation produced excellent results.

=== 30 Fold Cross Validation ===

=== Summary ===

Correctly Classified Instances 42 95.4545 %

Incorrectly Classified Instances 2 4.5455 %

Kappa statistic 0.8854

Mean absolute error 0.0644

Root mean squared error 0.2169

Relative absolute error 15.914 %

Root relative squared error 48.1552 %

Total Number of Instances 44

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.969 0.083 0.969 0.969 0.969 0.885 0.930 0.961 Hamilton

0.917 0.031 0.917 0.917 0.917 0.885 0.930 0.828 Madison

Weighted Avg. 0.955 0.069 0.955 0.955 0.955 0.885 0.930 0.925

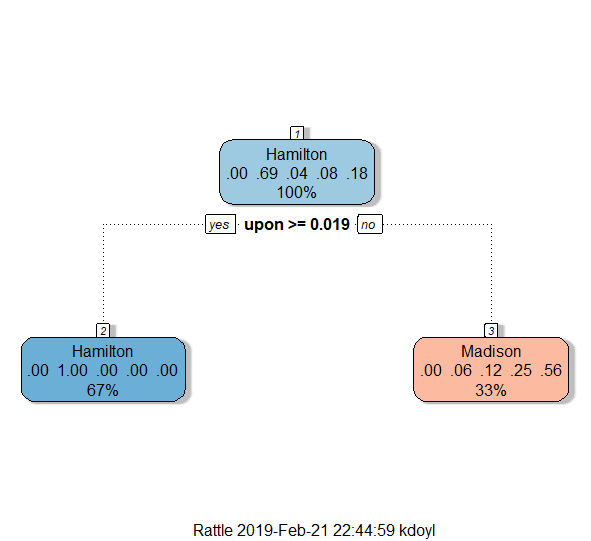
=== Confusion Matrix ===

a b <-- classified as

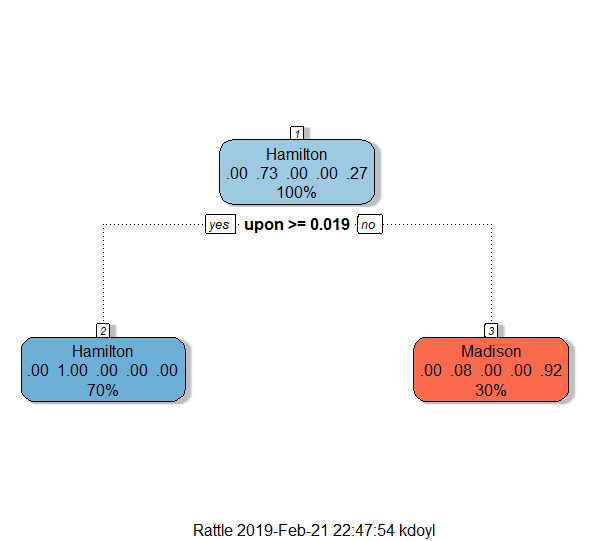
31 1 | a = Hamilton

1 11 | b = Madison

The rpart model with the Corpus produced disappointing results. This is a result of overfitting.



The rpart model with the CSV file produced the same tree as the corpus.



The prediction using the 2nd J48 model resulted in mostly Madison predictions.

|  |  |
| --- | --- |
| paper\_name | predict |
| dispt\_fed\_49.txt | Madison |
| dispt\_fed\_50.txt | Hamilton |
| dispt\_fed\_51.txt | Madison |
| dispt\_fed\_52.txt | Madison |
| dispt\_fed\_53.txt | Madison |
| dispt\_fed\_54.txt | Madison |
| dispt\_fed\_55.txt | Madison |
| dispt\_fed\_56.txt | Madison |
| dispt\_fed\_57.txt | Madison |
| dispt\_fed\_62.txt | Madison |
| dispt\_fed\_63.txt | Madison |

**Results**

The hierarchal model using Euclidean distance produces mostly Hamilton results.

* 50: Hamilton
* 49: Hamilton
* 63: Hamilton
* 53: Hamilton
* 52: Hamilton
* 54: Hamilton
* 55: Hamilton
* 57: Hamilton
* 58: Madison
* 51: Madison
* 62: Madison

The prediction using the 2nd J48 model resulted in mostly Madison predictions.

|  |  |
| --- | --- |
| paper\_name | predict |
| dispt\_fed\_49.txt | Madison |
| dispt\_fed\_50.txt | Hamilton |
| dispt\_fed\_51.txt | Madison |
| dispt\_fed\_52.txt | Madison |
| dispt\_fed\_53.txt | Madison |
| dispt\_fed\_54.txt | Madison |
| dispt\_fed\_55.txt | Madison |
| dispt\_fed\_56.txt | Madison |
| dispt\_fed\_57.txt | Madison |
| dispt\_fed\_62.txt | Madison |
| dispt\_fed\_63.txt | Madison |

**Conclusion**

By the time New York voted, ten states had already ratified the Constitution and had already passed because only nine states had to ratify it for the new government to be established. However, federal judges frequently use the Federalist papers to interpret the Constitution. There is a possibility if it were not for the big debate regarding the authors of the articles, would the papers be as famous as they are.

There were several different experiments performed on the data. The results are inconclusive. Different models produce very different results. Most studies show predominantly Madison results for the disputed articles. Hamilton and Madison both claimed to write the disputed articles, it is possible that it was a collaborative effort which would explain the questionable results. Due to the lack of confidence, models should be created with more articles by each author, articles with a variety of topics. The Federalist Papers are all addressed to New Yorkers urging them to approve the Constitution. By gathering articles of different topics would result in a better analysis of authors’ writing styles.