**Introduction**

The Federalist Papers were a series of eighty-five essays urging the citizens of New York to ratify the new United States Constitution. Written by Alexander Hamilton, James Madison, and John Jay, the essays originally appeared anonymously in New York newspapers in 1787 and 1788 under the pen name "Publius." A bound edition of the essays was first published in 1788, but it was not until the 1818 edition published by the printer Jacob Gideon that the authors of each essay were identified by name. The Federalist Papers are considered one of the most important sources for interpreting and understanding the original intent of the Constitution.

As one could imagine, maintaining accurate authorship of historical writings is extremely important. It establishes credibility and also gives credit where credit is due. In order to maintain an accurate account of history, we must accurately account history to those who wrote it.

**Analysis**

**About the Data**

There are a total of 85 essays, 74 of which have 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”, or “dispute”. The features are a set of “function words”, for example, “upon”. The feature value is the percentage of the word occurrence in an essay. For example, for the essay “Hamilton\_fed\_31.txt”, if the function word “upon” appeared 3 times, and the total number of words in this essay is 1000, the feature value is 3/1000=0.3%.

For observing the data, note that “dispute” records are records numbers 1 through 11, “Hamilton” are numbers 12 through 62, and “Madison” are numbers 71 through 85.

**Preprocessing**

Since the dispute is only between Hamilton and Madison, records for John Jay were removed from the Analysis. The “filename” field was also removed since it should not be a factor in the analysis. All papers with known authors were used to train our model, and all papers with disputed authors were tested against it.

**Results**

|  |  |
| --- | --- |
|  | x |
| 1 | Madison |
| 2 | Madison |
| 3 | Madison |
| 4 | Madison |
| 5 | Madison |
| 6 | Madison |
| 7 | Madison |
| 8 | Madison |
| 9 | Madison |
| 10 | Madison |
| 11 | Madison |

Using the J48 model for Decision Tree analysis trained with data from the papers with known authors, the disputed papers were tested against it were all classified as Madison being the author.

Model evaluation rated the model as being 94.2% correct with a kappa of .85. This means our model was not only highly accurate but also highly repeatable, making it highly reliable with this particular set of data.

**Conclusions**

Evidence suggests that the writing style of all 11 disputed papers matches that of James Madison.

As it turns out, it seems that Hamilton was trying claim more credit than he was due. Every paper Madison challenged authorship on does not match Hamilton’s typical writing style. Whether intentionally or unintentionally, there is every reason to believe that Hamilton was overstepping his credibility.