**Introduction:**

**General:**

The Federalist Papers were written and circulated in order to persuade New Yorkers to ratify the proposed United States Constitution, which had been drafted in Philadelphia during the summer of 1787. The essays discuss specific aspects of the Constitution in depth in order to advocate for its adoption over the current Articles of Confederation. Originally, intended to consist of 25 documents, The Federalist Papers eventually totaled to 85. Between October 1787 and August 1788, seventy-seven of the essays were published in a series in The Independent Journal, The New York Packet, and The Daily Advertiser[[1]](#footnote-0). Afterwards, the final eight were added. The "Federalist Papers" were initially referred to as "The Federalist.". Because of this, and due to Hamilton, Jay and Madison, the founding fathers of the Constitutional Convention, the Federalists are often consulted today to interpret the purposes of those who drafted the Constitution.

The essays were written by Alexander Hamilton, James Madison, and John Jay the three men collaborated under the name Publius. The authorship of the papers was a closely guarded secret at the time of publication. A list naming Hamilton as one of the writers was not made public until after his death. Hamilton received almost two-thirds of the essays. Madison would later refute many of these claims, claiming that he had written a number of the papers credited to Hamilton. The matters of authorship of the individual essays remained a matter of controversy. That is, not until two mathematicians named Frederick Mosteller and David Wallace decided to spend 3 years combing through Madison's and Hamilton's written works for subtle patterns in word choice—linguistic fingerprints that could be used to pinpoint the contested papers' roots back to their original author.

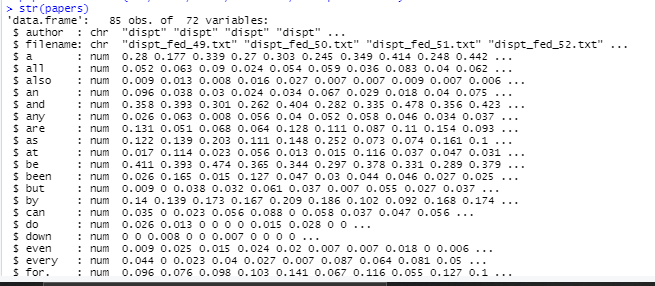
Mosteller and Wallace, statisticians, studied the frequency distributions of different function terms in the Federalist Papers. Was Hamilton, the sole author of these essays? Was there another author who was the sole author? Or were the essays the product of the three men working together? When Mosteller and Wallace published their findings, they not only solved a long-standing mystery, but they also developed a new mathematical method for analyzing written text. Suddenly, statisticians and computer scientists can be called upon to act as textual detectives. Authorship attribution is now a common problem in data mining.

**Analysis:**

**About the data**

The data was read in from the csv file provided and examined using the structure function. Initially, the data looked like Figure 1 below.

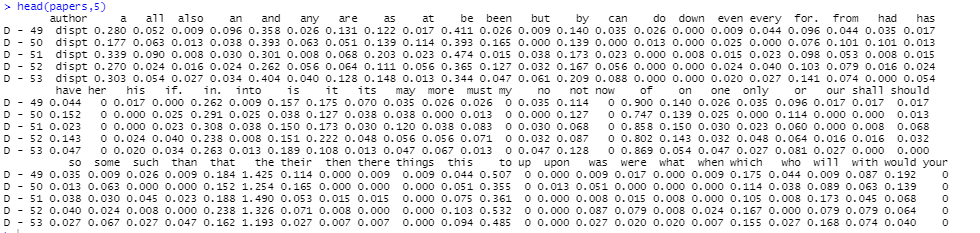
Figure 1: Original Structure of the dataframe



Data Manipulation

Once the Federalist dataset is loaded some pre-processing steps are used to clean the data. To start, a new column was created with an abbreviation of the author name. Secondly, the file name and file number was split. A new column was thus created to combine the author name along with the file number. Dropped unused levels, unwanted columns and only concerned on the authorship of the disputed articles and only among Hamilton and Madison. Lastly, since there were a few changes to the dataset, reviewed the first top 5 rows as shown in Figure 2.

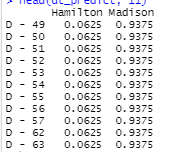
Figure 2: First 5 rows post manipulation



Data Preparation

The aim of using a Decision Tree is to build a training model that can be used to predict the class or value of target variables by learning decision rules from prior data (training data). Based on the author's name, the data is divided into training and testing. The unused levels were dropped and the decision tree analysis was run on the l (train) dataset. As a result, the contested papers are 93.75 percent likely to belong to Madison as shown in Figure 3.

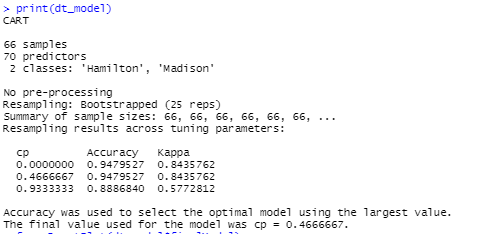
Figure 3: 93.75% probability the disputed articles belonging to Madison

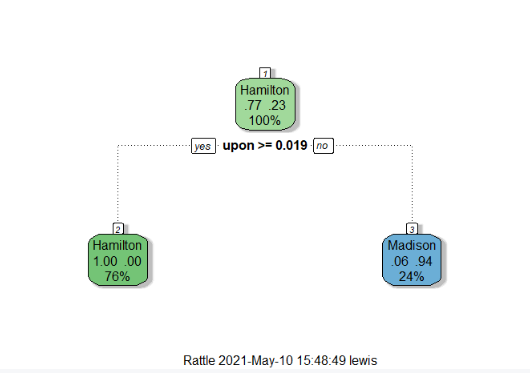


**Model**

Printing and Plotting the final model shown in Figure 4.

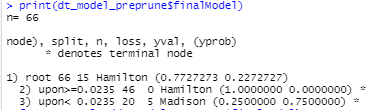
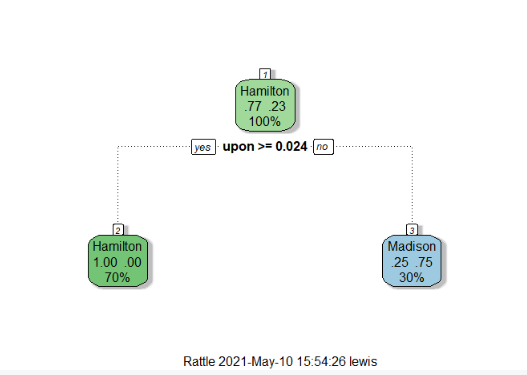
Figure 4: Print and Plot of the training final model





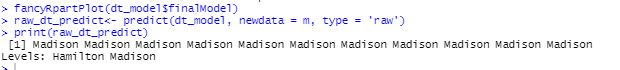
Next step was building and tuning the decision tree model (Figure 5). Increased the required frequency from 0.019 to 0.024.

Figure 5: Tuning model



**Prediction:**

Figure 6: Predicting Model which reconfirms that the disputed articles are authored by Madison



**Results:**

One can see that the term 'upon' plays a major role in both of the models above. Surprisingly, the frequency of this word appears to dictate the authorship of the entire file. The needed frequency has also increased from 0.019 to 0.024 as a result of the tuning and pruning. If it is greater than the specified value, the file is Hamilton's; otherwise, it was written by Madison. Using the Decision Tree Algorithm, the probability the disputed articles belong to Madison is 93.75%. The DT Model reaches the same conclusion as the clustering algorithm.

**Conclusion:**

The essays were co-written by Alexander Hamilton, James Madison, and John Jay, who went by the alias Publius. At the time of print, the authorship of the papers was a tightly guarded secret. The matters of authorship of the individual essays remained a matter of controversy. That is, not until two mathematicians named Frederick Mosteller and David Wallace decided to join in on the controversy and try to solve the mystery of who wrote the disputed papers. The frequency distributions of different function words in the Federalist Papers were analyzed by statisticians Mosteller and Wallace. Was Hamilton the only person who wrote these essays? Is there any person who wrote the essays by themselves? Or were the essays the product of collaboration between the three men?

The evidence indicates that Madison was the author of the contested papers after reviewing the findings gathered. Madison’s total essays authored are 29. Hamilton, in turn, wrote 51 essays, with Jay providing the remaining 5. Despite the fact that Hamilton is the most prolific author of the three, this assertion significantly extends Madison's contribution to the papers, making this a significant historical moment. This isn't the first time The Federalist Papers have been studied in order to solve this puzzle. The essays have played a significant role in American history, both in the past and present. Identifying the writers of each of these seminal works will reveal essential truths about the founding fathers' values and achievements.

Some people were concerned that the Constitution would grant the federal government too much control. Not only did people think that the government would have too much power but also restrict individual liberty. To help ease these concerns, Hamilton, Jay, and Madison studied the Constitution in depth. The authors explained the built-in checks and balances intended to balance power. The Federalist Papers are a valuable compilation today. The papers are valuable because they provide insight into the intentions of key individuals who discussed the elements of the Constitution. Thank you Mosteller and Wallace for developing a new mathematical method for analyzing written text.

1. Dan T. Coenen, “Fifteen Curious Facts About the Federalist Papers.” [University of Georgia School of Law](https://digitalcommons.law.uga.edu/cgi/viewcontent.cgi?article=1001&context=fac_pm), April 1, 2007. [↑](#footnote-ref-0)