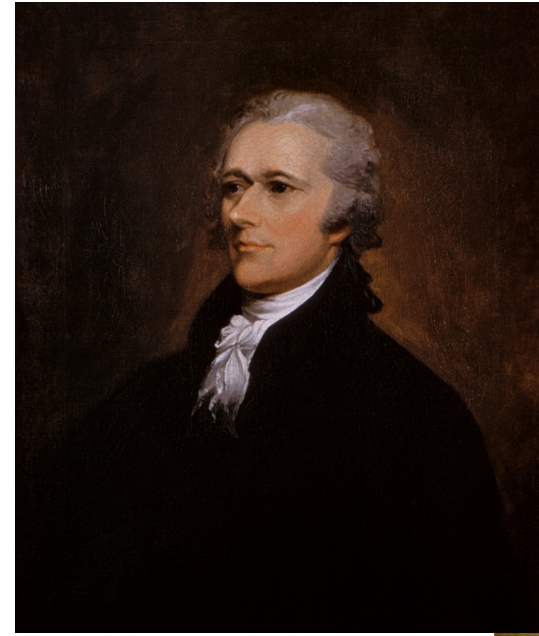


Alexander vs Hamilton: Final Project IST 736

By: Alison Reikher



Background

Alexander Hamilton

- Founding Father
- Promoted the US Constitution
- Established the national financial system
- Established the US Coast Guard
- Founded the New York Post newspaper
- Appears on the \$10 Bill

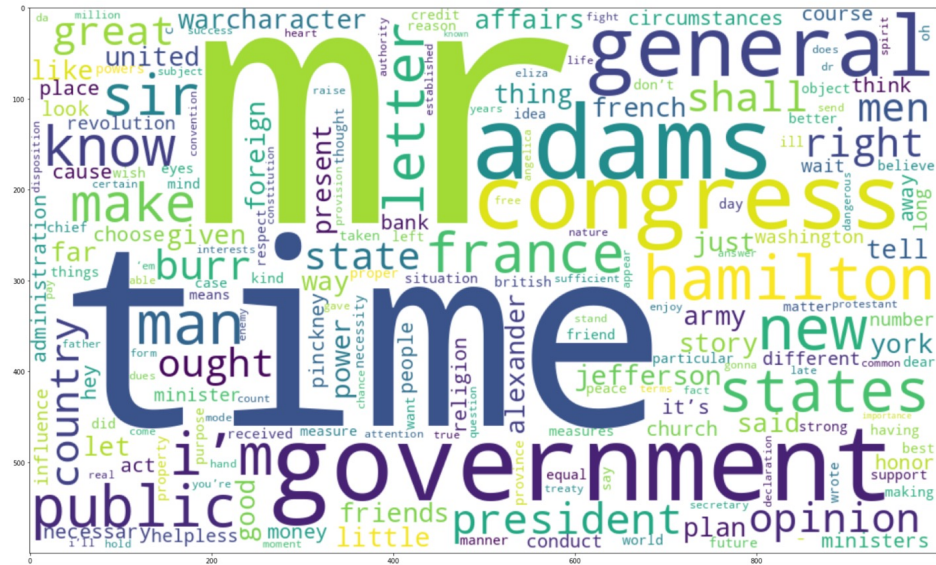
Hamilton: The Musical

- Musical written by Lin Manuel Miranda
- Premiered on Broadway in January 2015
- Most Tony nominated musical with 16 nominations
- Second most Tony wins with 11 wins including Best Musical and Best Original Score

Dataset

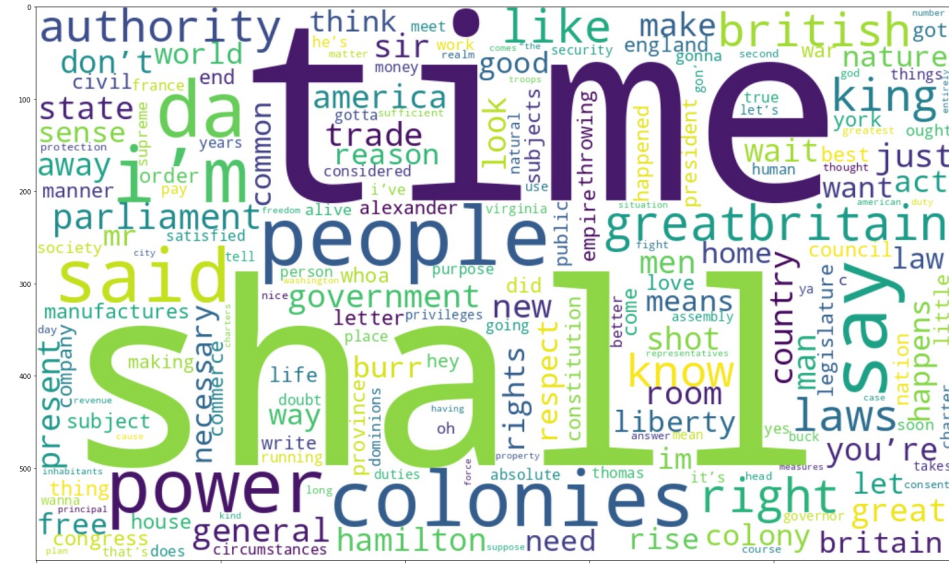
Letters

- 46 letters selected from, to, or about Alexander Hamilton



Lyrics

- 46 song lyrics



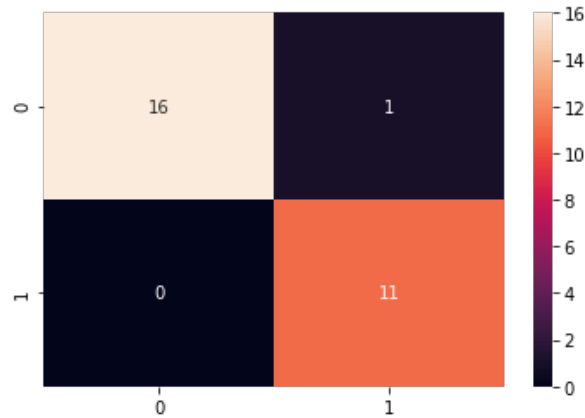
Models

Goal of analysis: determine which algorithm can classify the adapted vs original dataset best

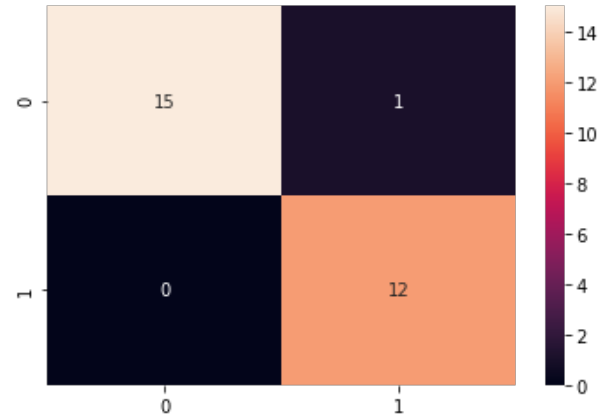
- 4 Multinomial Naïve Bayes
- 1 Bernoulli Naïve Bayes
- 5 Decision Trees
- 4 SVM models
- 1 K Means Clusters
- LDA Topic Modeling



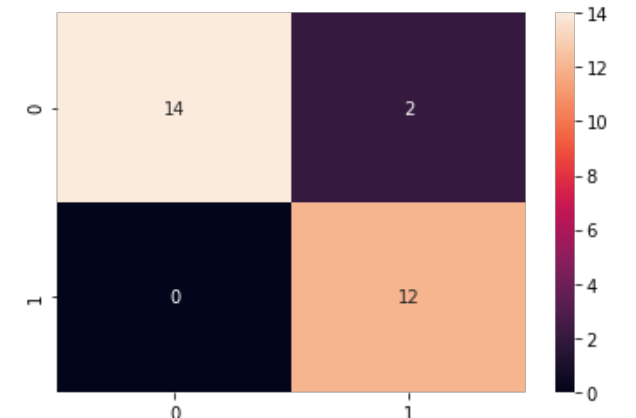
Naïve Bayes Results



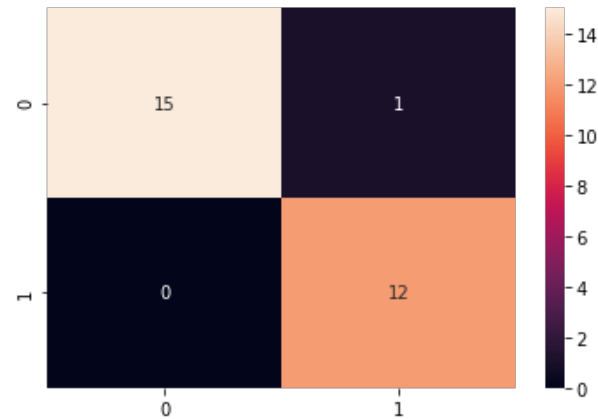
CountVectorizer Stemmed Matix



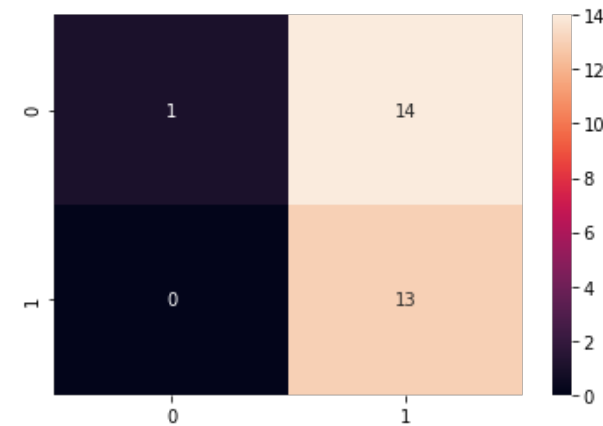
TFIDF No Stemming Matrix



CountVectorizer Punctuation Matrix



TFIDF Stemmed Matrix



Bernoulli Matrix

Naïve Bayes Cross-Validation Results

Score 1 = CountVectorizer Stemmed Matrix

Score 2 = TFIDF No Stemming Matrix

Score 3 = TFIDF Stemmed Matrix

Score 4 = Bernoulli Matrix

Score 5 = CountVectorizer Punctuation Matrix

Score 1: 0.95 accuracy with a standard deviation of 0.04

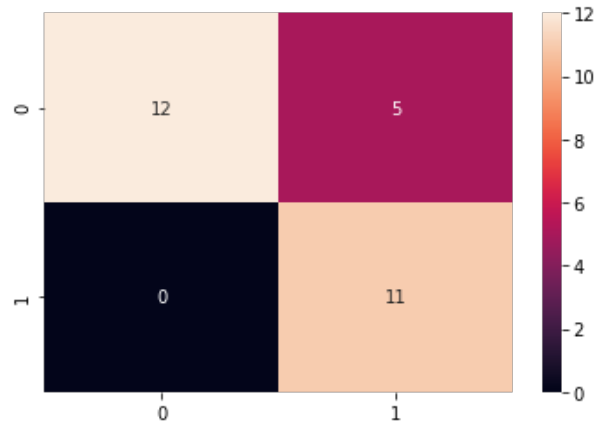
Score 2: 0.87 accuracy with a standard deviation of 0.08

Score 3: 0.94 accuracy with a standard deviation of 0.09

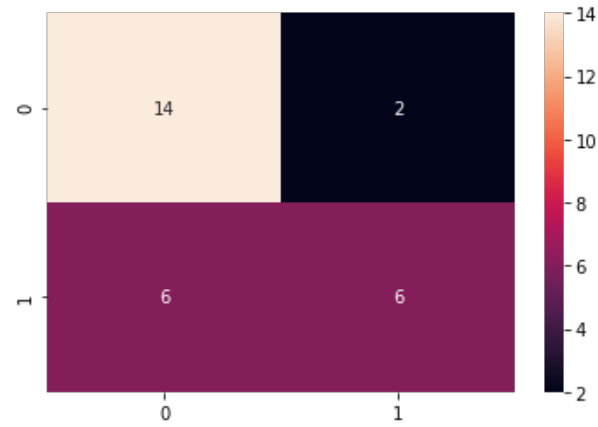
Score 4: 0.59 accuracy with a standard deviation of 0.10

Score 5: 0.90 accuracy with a standard deviation of 0.13

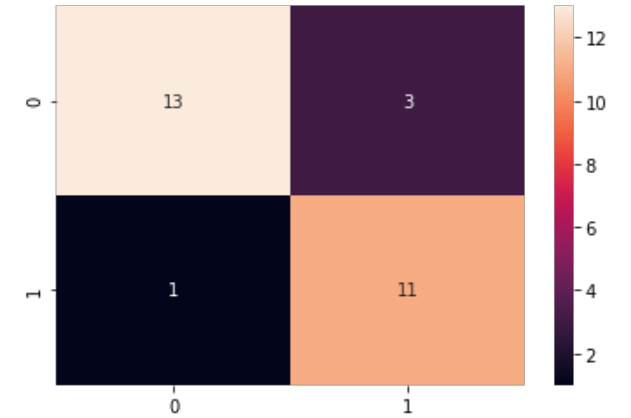
Decision Tree Results



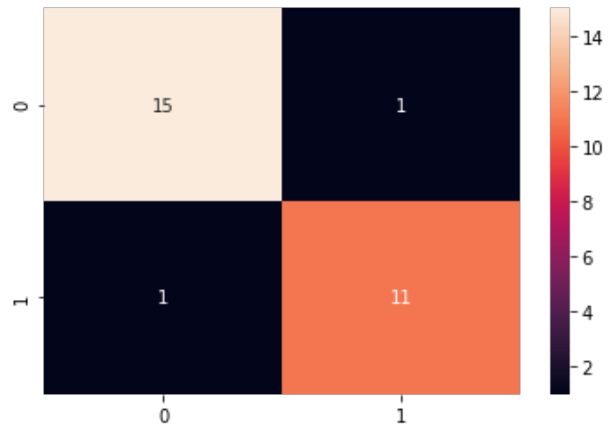
CountVectorizer Stemmed Matix



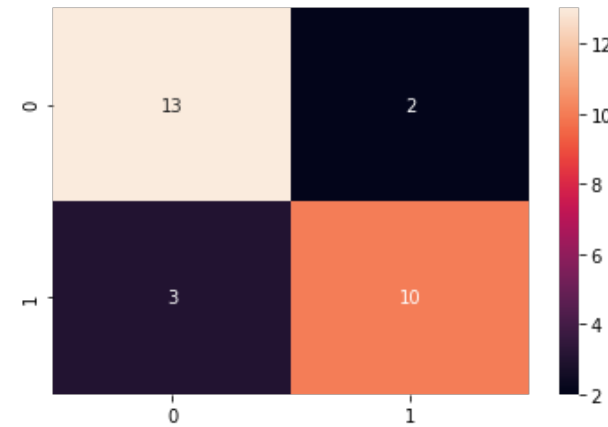
TFIDF No Stemming Matrix



CountVectorizer Punctuation Matrix



TFIDF Stemmed Matrix



Bernoulli Matrix

Decision Tree Cross-Validation Results

Score 6 = CountVectorizer Stemmed Matrix

Score 7 = TFIDF No Stemming Matrix

Score 8 = TFIDF Stemmed Matrix

Score 9 = Bernoulli Matrix

Score 10 = CountVectorizer Punctuation Matrix

Score 6: 0.83 accuracy with a standard deviation of 0.06

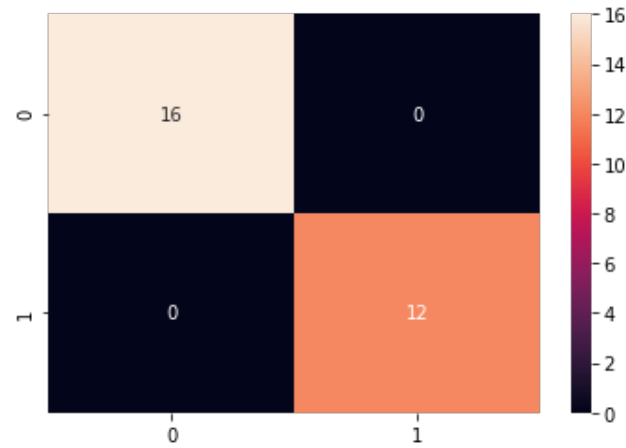
Score 7: 0.80 accuracy with a standard deviation of 0.08

Score 8: 0.92 accuracy with a standard deviation of 0.05

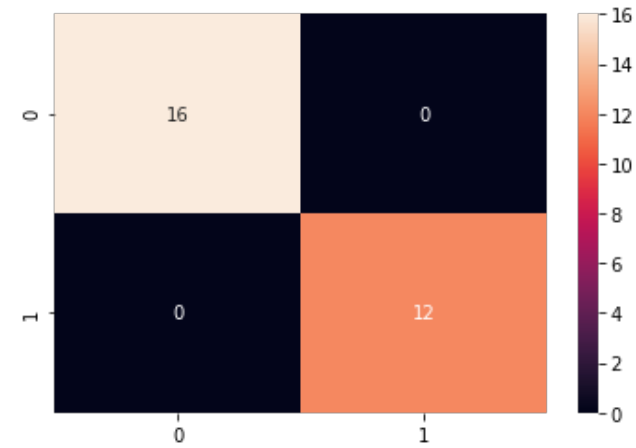
Score 9: 0.73 accuracy with a standard deviation of 0.04

Score 10: 0.76 accuracy with a standard deviation of 0.06

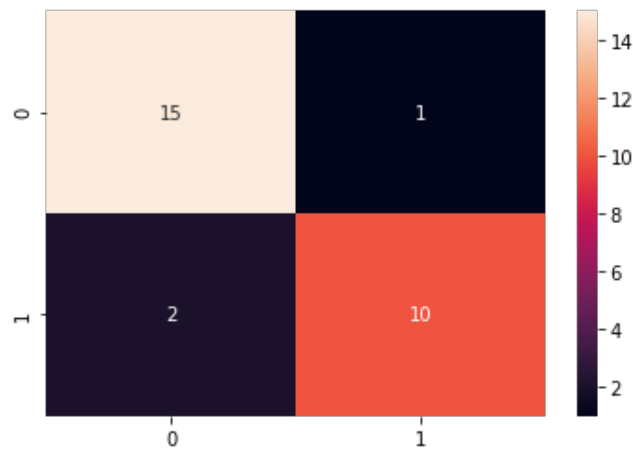
SVM Results



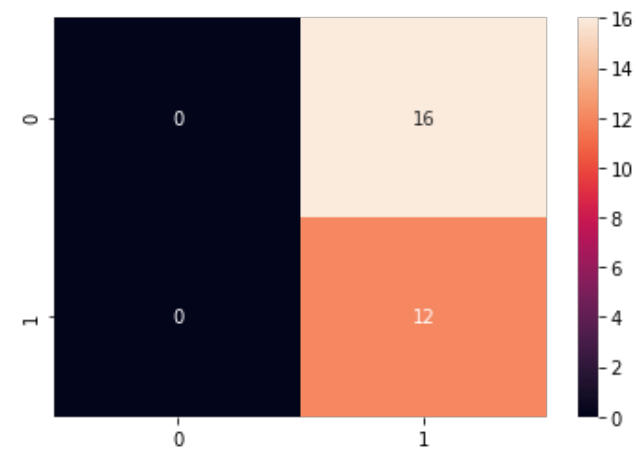
Linear SVM $c=10$



Linear SVM $c=50$



RBF SVM



Poly SVM

SVM Cross-Validation Results

Score 11 = Linear SVM $c=10$

Score 12 = Linear SVM $c=50$

Score 13 = RBF SVM

Score 14 = Poly SVM

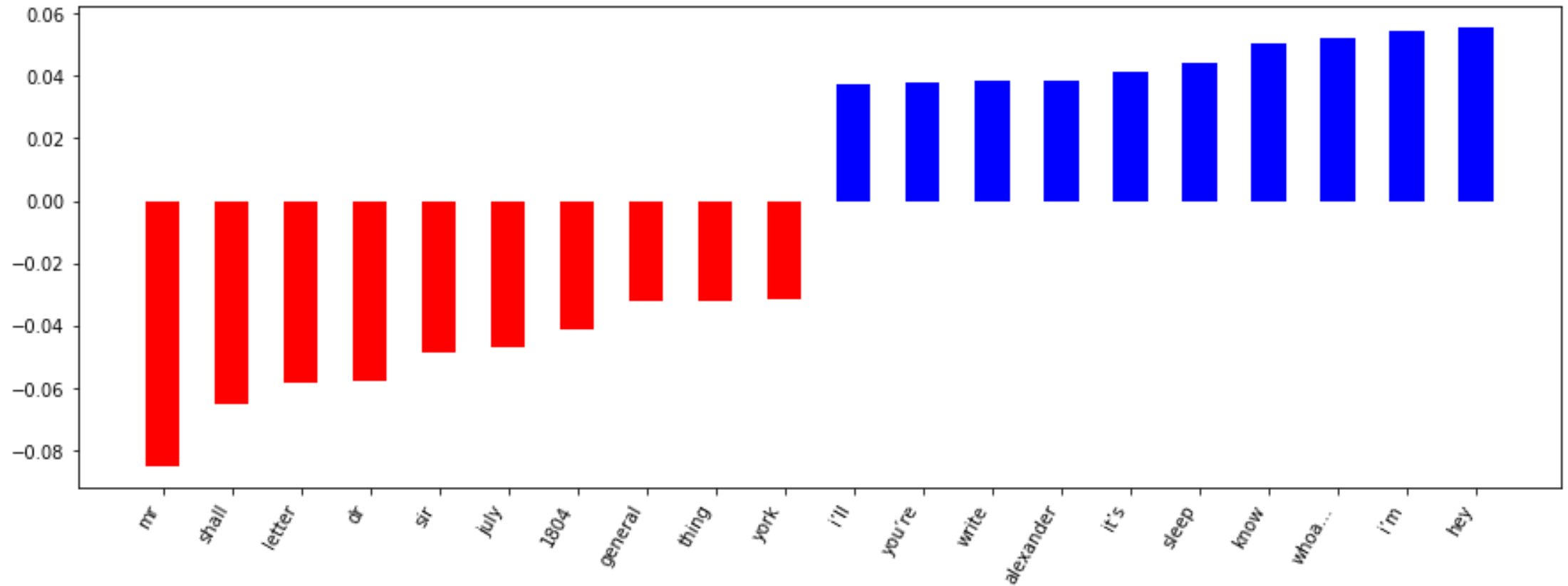
Score 11: 0.98 accuracy with a standard deviation of 0.03

Score 12: 0.98 accuracy with a standard deviation of 0.03

Score 13: 0.83 accuracy with a standard deviation of 0.13

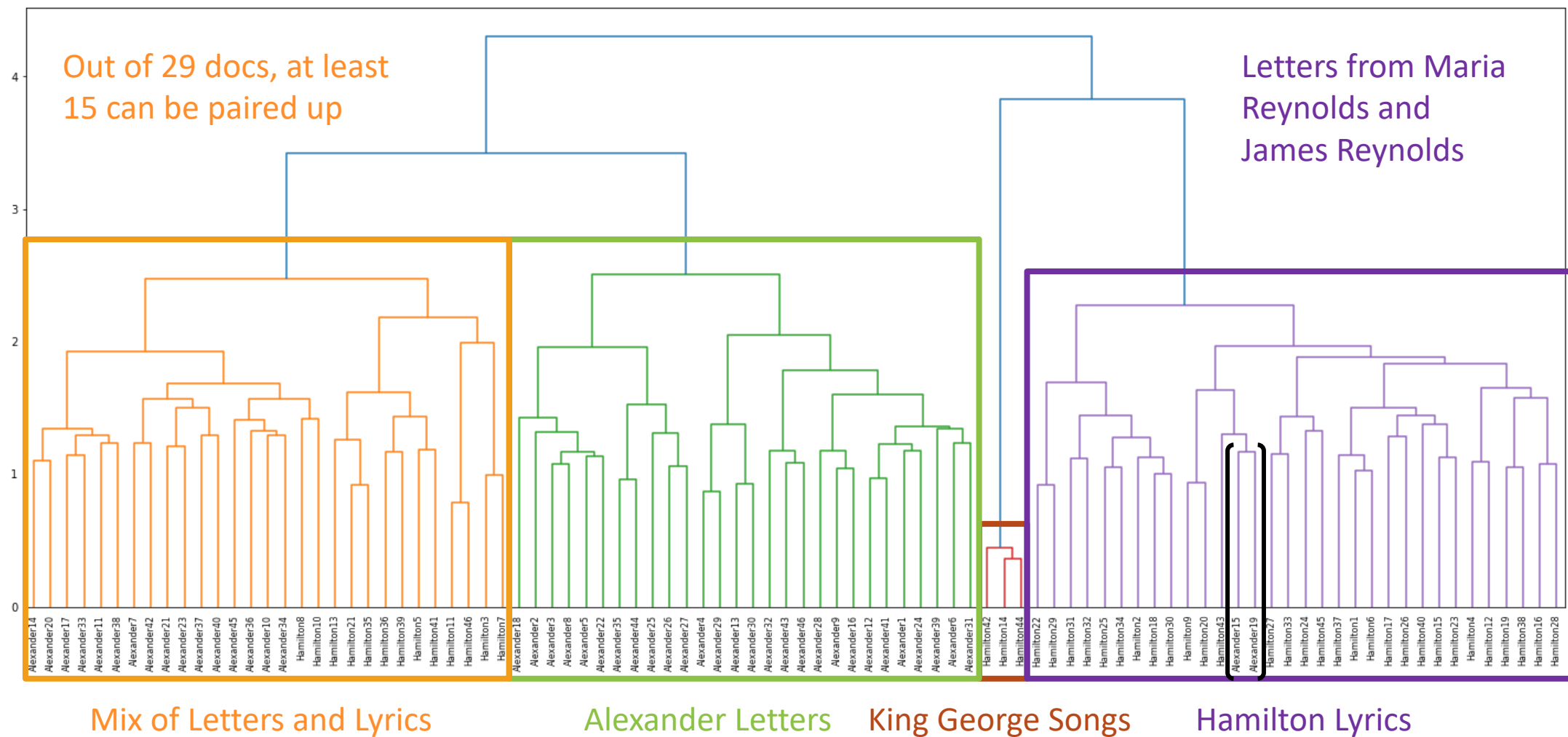
Score 14: 0.55 accuracy with a standard deviation of 0.04

Important Words from SVM model



Red = letters, Blue = lyrics

K Means Clustering



LDA Topic Modeling

| Topic #0 | Topic #1 | Topic #2 | Topic #3 | Topic #4 | Topic #5 | Topic #6 | Topic #7 | Topic #8 | Topic #9 |
|---|---|---|--|--|---|--|--|---|---|
| room i'm shot say time rise happens like away whoa you're know sir throwing happened gonna got don't burr buck | shall people time power colonies authority greatbritain said general laws government king congress british state make trade parliament right great | mr adams president government france man friends general public time country character ministers pinckney jefferson minister opinion letter states ought | know look i'm hey helpless angelica just new alive alexander satisfied eliza stay tell im sister tonight eyes york away | story time tells tell number lives dies orphanage — it's eliza madison negotiate seven alexander place face duel york tried | burr choose he's nice i'm know gon' oh it's hamilton wrote new don't you're sir mr wait like look letter | it col capt corps commanded redoubt right major detachment batalion van wounded honor received enemy left troops work advanced entering | love write said burn letters heart world amiable im wrote letter hey bed father ladies read forfeit watching memories built | thy thou oh mr united states june general million let treasury look livres god 10th revolution following millions farmers ye | da wait hamilton i'm ya don't im home sir you're takes world yo need thomas work fight gotta new man |



= Hamilton lyrics



= Alexander letters

Final Conclusion

- Best Models:
 - Linear SVM: Accuracy = 0.98
 - Stemmed Multinomial Naïve Bayes: Accuracy = 0.95
 - Normalized and Stemmed Decision Tree: Accuracy = 0.92
 - K Means Clustering
- Worst Models:
 - Bernoulli Naïve Bayes: Accuracy = 0.59
 - Poly SVM: Accuracy = 0.55