Gender Classification in 19th Century Literature

Alex Kimbrell & Tori Pirtle

Carli Arnett, Jeffrey Mendez,
Amelia Green & Matthew Carpenter

Our Problem

Using an assortment of novels from the 19th century, compare and contrast works written by male and female authors to determine if there is any systemic difference to distinguish them.

Our Theory

Female Authors:

- Longer, more descriptive sentences
- More dialogue
- More adjective usage

Male Authors:

- Shorter sentences
- Less dialogue
- Less adjective usage, less descriptive

Things to Calculate:

- Average Sentence Length
- Average Length of Quotations
- Number of Quotations Used
- Adjective Ratio

Our Data:

- 215 novels (rows)
  - 27 novels written by Female Authors
  - 188 novels written by Male Authors
- 6 Attributes for Each Novel
  - Title
  - Male/Female
  - **Number of Quotes**
  - **Average Quote Length**
  - **Adjective Ratio**
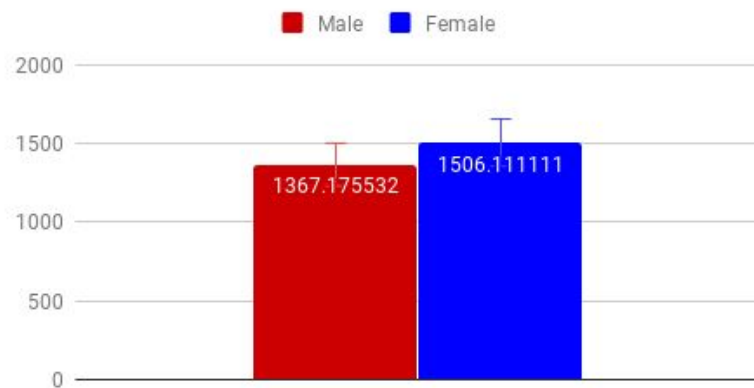 - **Average Sentence Length**

Our Data:

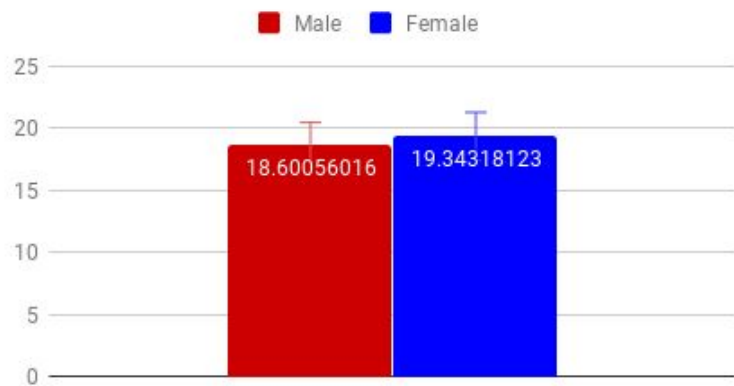|	Novel	Male/Female	Number Of Quotes	Average Quote Length	Adjective Ratio	Average Sentence Length
0	A Christmas Story	Male	49	5.755102041	0.05386552506	14.15539066
1	ecticut Yankee in King Arthu	Male	1251	25.54036771	0.0534505004	18.72300542
2	uet With an Occasional Cho	Male	83	11.36144578	0.06328709254	11.56114398
3	ckoning_A tale of Bush Life	Male	2130	21.6	0.03833040603	16.47139224
4	A Jacobite Exile	Male	1678	34.74016687	0.04017649292	19.99491008
5	White Cross_A tale of the S	Male	1543	48.88788075	0.04326394216	24.03685423
6	A Lady of Quality	Female	1499	16.50233489	0.05106634268	21.31205324
7	Laodicean_A Story of Toda	Male	47	15.40425532	0.05429400717	16.73111032
8	A Pair of Blue Eyes	Male	153	17.15686275	0.05714337594	14.22726302
9	A Search for a Secret Vol. 1	Male	434	82.34101382	0.05441615856	23.51088963
10	A Search for a Secret Vol. 2	Male	670	43.03731343	0.0461739105	21.73150582

Average Adjective Ratio

■ Male ■ Female

Male: 0.05175889239
Female: 0.0560031917

Average Number of Quotations

■ Male ■ Female

Male: 1367.175532
Female: 1506.111111

Average Quote Length

■ Male ■ Female

Male: 36.22595665
Female: 33.26677962

Average Sentence Length

■ Male ■ Female

Male: 18.60056016
Female: 19.34318123
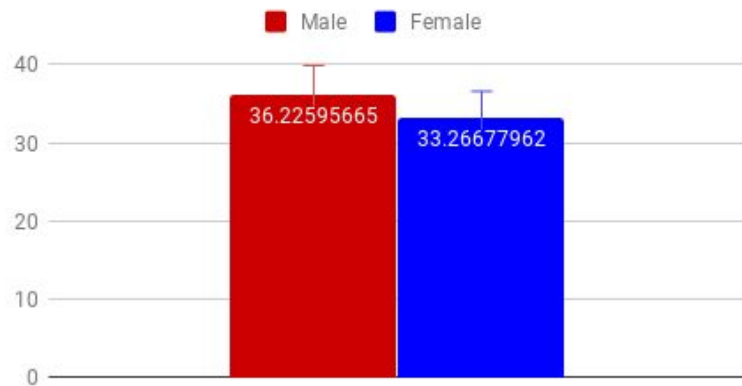
Our Models

- Run models on both balanced data and unbalanced data
    - Balanced: 27 Female Novels & 27 Male Novels
    - Unbalanced: 27 Female Novels & 188 Male Novels
- Models Run:
    - Logistic Regression
    - Naive Bayes
   - Decision Tree
    - SVM

Our Results: All Data

Train/Test	Logistic Regression	Naive Bayes	Decision Tree	SVM
30/70	86.0927	85.4305	77.4834	86.0927
50/50	90.7407	75.0000	84.2593	89.8148
70/30	92.3077	80.000	75.3846	92.3077

Our Results: Balanced Data

Train/Test	Logistic Regression	Naive Bayes	Decision Tree	SVM
30/70	50.0000	63.1575	65.7895	60.5263
50/50	51.8519	59.2593	62.9630	40.7407
70/30	47.0588	70.5882	70.5882	35.2941

Gender Classification
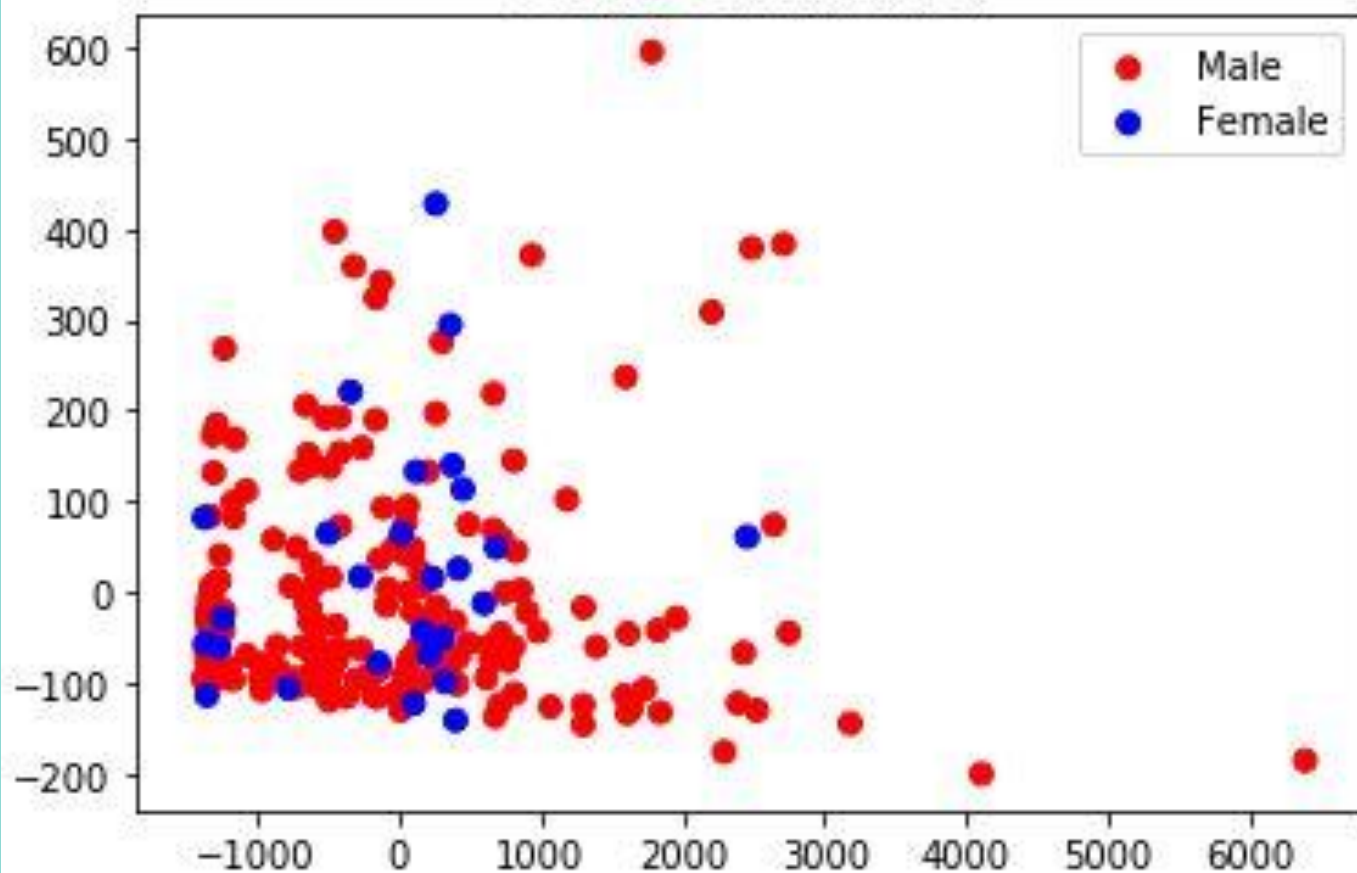
male	female
chief	pillows
rear	lace
civil	curls
bigger	dress
absolutely	china
enemy	skirt
fellows	curtains
king	cups
public	sheets
contact	shrugged

Our Results: With Masculine/Feminine Words

Train/Test	Logistic Regression	Naive Bayes	Decision Tree	SVM
30/70	65.7895	68.4211	63.1579	76.3158
50/50	70.3704	70.3704	62.9630	44.4444
70/30	82.3529	76.4706	70.5882	47.0588

Questions?

References

- https://qz.com/934268/words-that-give-away-a-writers-gender-from-ben-blatts-book-nabokovs-favorite-word-is-mauve/
- https://scikit-learn.org/stable/
- https://tenor.com/
- https://www.youtube.com/watch?v=dQw4w9WgXcQ