Frankenstein

A Machine Learning Project

Galia Appel

February 2021

https://github.com/Galia-A/frankenstein-ml

1. Objective

Was *"Frankenstein, or the Modern Prometheus"* written solely by Mary Shelley?

2. Background

1816, The year without summer. Five friends: Lord Byron, Percy Shelley, Mary Godwin, Dr. John Polidori and Claire Clairmont, have assembled at the Villa Diodati and were engaged in a friendly contest of writing horror stories.

In this setting, Mary (18 at the time) has written "Frankenstein, or the Modern Prometheus".

Why is there a question?

The first book was published anonymously, with Percy's preface and some of his poems. This created the assumption that Percy wrote it. Mary did publish the book twice more in her lifetime under her name, but the idea has stuck.
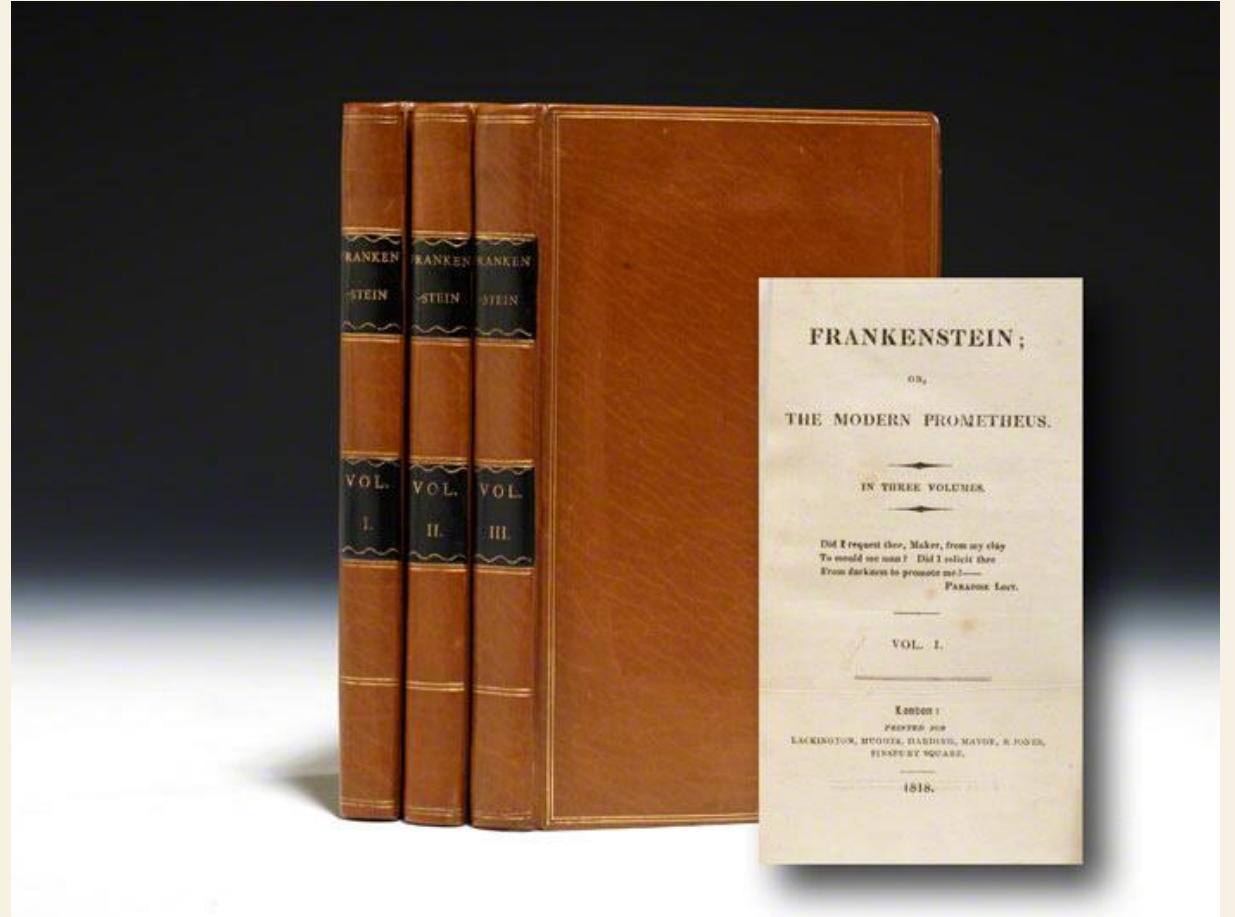
The Critics Vote for Percy

- The preface and poems were recognized as Percy Shelley's style.

- The drafts revealed Percy Shelley made many corrections.

- The novel and its themes drew from Percy Shelley's life.

> One hardly knows whether to regard him as editor or minor collaborator.
>
> — James Rieger, 1974

Against The Critics

- The preface and poems were indeed written by Percy Shelley. They both were correctly credited in a later edition.

- The handwriting is hard to distinguish between the two authors' and most of the critics didn't get a firsthand examination of the notebooks.

- The themes and ideas which appear in the novel appeared in Mary's father, William Godwin's works. Both Percy and Mary admired him and learned them, each on his own.

Yet, the argument continues…

3. Data

All the texts were taken from the Project Gutenberg site

Mary Shelley

- Frankenstein; Or, The Modern Prometheus
- Falkner: A Novel
- Mathilda
- Tales and Stories
- The Last Man
- Valperga Volume 1
- Valperga Volume 2
- Valperga Volume 3

Percy Shelley

- A Vindication of Natural Diet
- A Defence of Poetry and Other Essays
- The Complete Poetical Work

4. Pre-processing

1. Manually :

   1. Remove Gutenberg's header and footer, keep only the relevant book.

   2. Remove references and publisher's notes.

   3. Remove prefaces by other authors.

   4. Remove italic mark (underscore).

   5. Remove spelling and grammar fixes.

4. Pre-processing

2. NLTK (Natural Language Toolkit) - tokenize sentences.

```python
import nltk

def extract_sentences(book_collection):
    sents = []
    for book in book_collection:
        with open(book, encoding="utf-8") as f:
            text = f.read().replace("\n", " ").replace("***", "")
            text = re.sub(r"\n+", "", text)
            sents.extend(nltk.sent_tokenize(text))
    return sents

mary_sents = extract_sentences(mary)
percy_sents = extract_sentences(percy)
```

4. Pre-processing

3. As we can see from the left plot below, the dataset is imbalanced, with Mary having almost twice as many sentences as Percy. I randomly discard some of Mary's sentences to match the number of sentences from Percy.
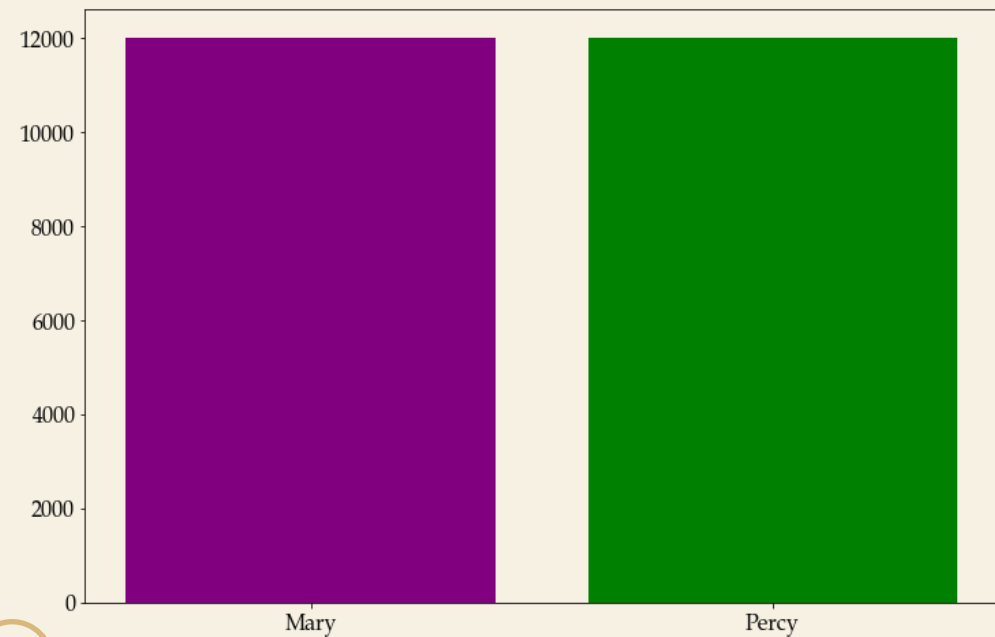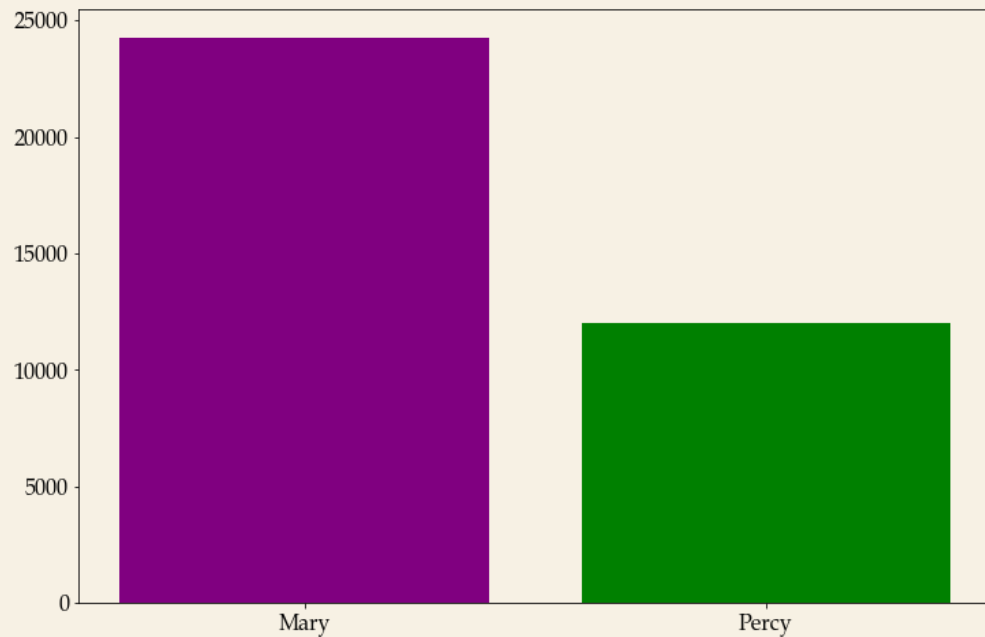
5. Machine Learning Pipeline

1. Create dataset and targets:

```python
from sklearn.model_selection import train_test_split

X = mary_sents + percy_sents
y = ["Mary"] * len(mary_sents) + ["Percy"] * len(percy_sents)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

5. Machine Learning Pipeline

2. Fit the Bag-of-Words to the Training Set

    Scikit-Learn - Using `CountVectorizer` to turn a sentence into a feature vector.

    A Bag-of-Words representation of text is a vector of length equal to the number of unique words in the training set.

    For example, given the two sentences: "I want to go get some ice-cream", "I wish to go get some coffee", we have 9 unique words, so each Bag-of-Words vector is of length 9, and the resulting vectors would look like this:

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 1 \end{bmatrix}$$

5. Machine Learning Pipeline

2. Fit the Bag-of-Words to the Training Set

```python
from sklearn.feature_extraction.text import CountVectorizer

bow = CountVectorizer() # init
bow.fit(X_train) # learning all the words

x_train_bow = bow.transform(X_train) # creates vectors
x_test_bow = bow.transform(X_test)
```

6. Trying Different Models

As we've seen in class, the models I am trying are:

- Decision Tree
- Random Forest
- Support Vector Machine
- Multilayer Perceptron

7. Evaluation

I tested the accuracy of each model on the held-out test set.

- Decision Tree: 75.3%

- Random Forest: 83%

- SVM: 85.7%

- MLP: 87.7%

7. Evaluation

The Multilayer Perceptron got the best score out of the four models, so I'll use it for my analysis.

The following confusion matrix shows how many times the model was "confused": it classified Mary when it should have classified Percy and vice versa.
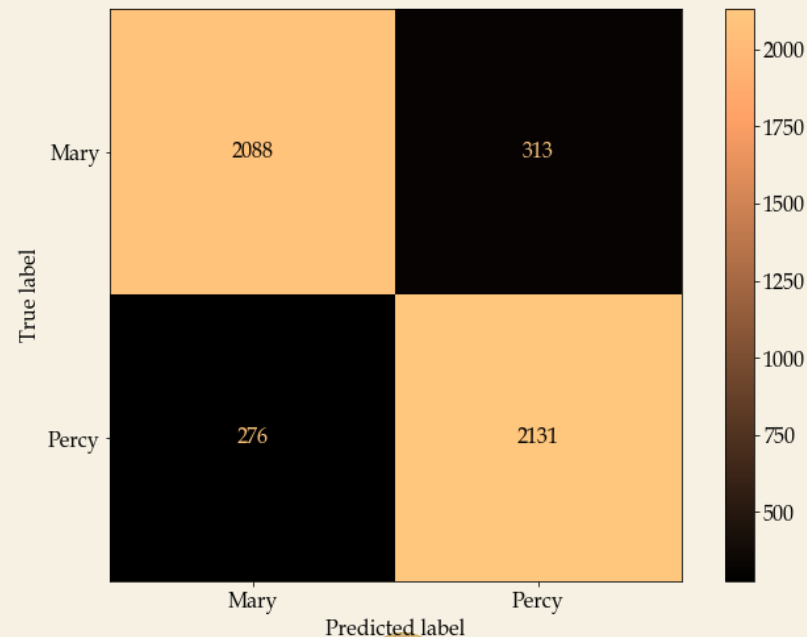
8. Frankenstein Analysis

Transform the sentences into a Bag-of-Words using the same Bag-of-Words that was fit on the original training set.

Now, I predict using the MLP model which author wrote which sentence in Frankenstein.
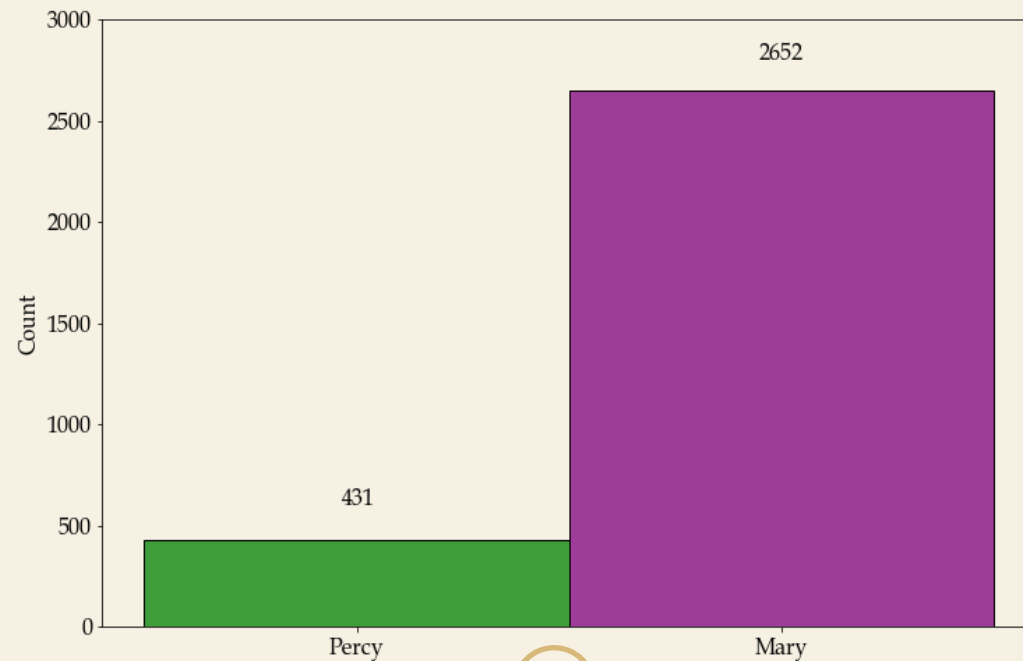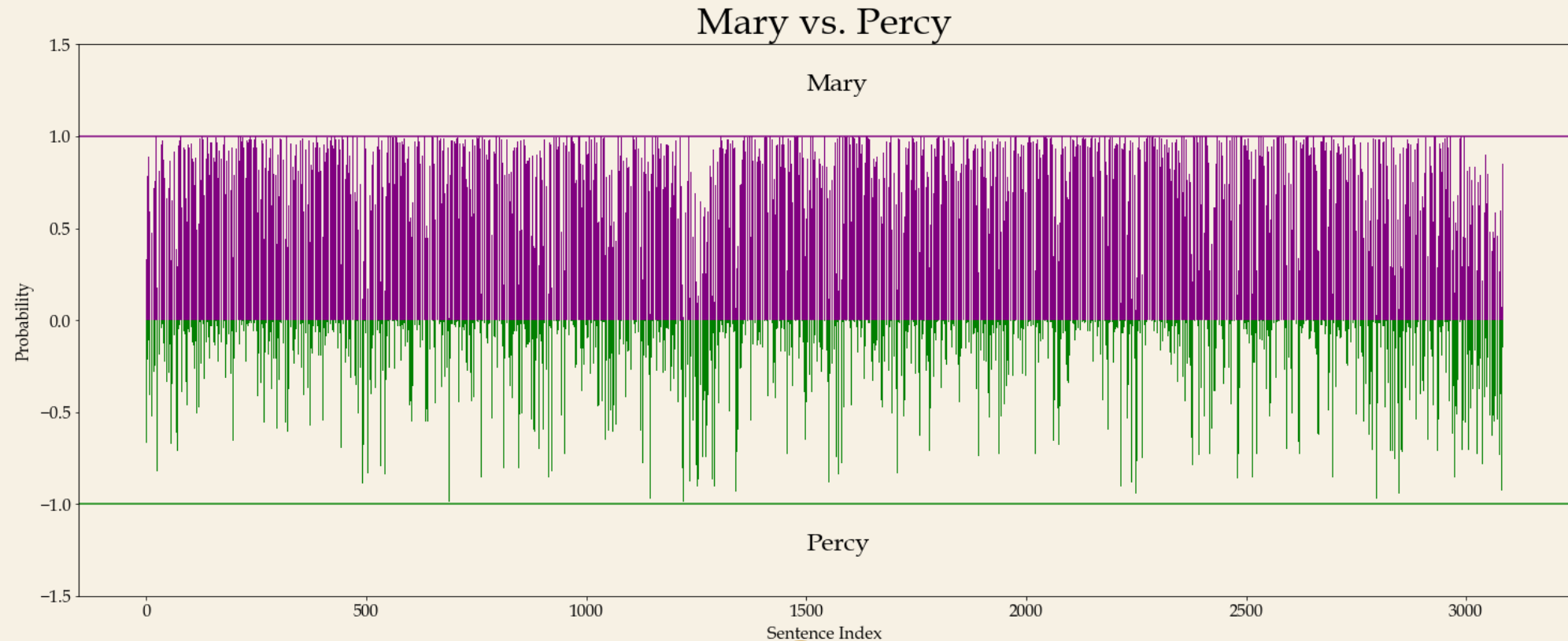
8. Frankenstein Analysis

In my research, one of the hypotheses was that Percy Shelley wrote between 4,000 and 5,000 words in "Frankenstein". Let's see what the model thinks:

```
>>> sum([len(nltk.word_tokenize(frank_sents[idx])) for idx in percy_sent_idxs])
8280
```

8. Frankenstein Analysis

I also checked the probability for each sentence to be written by either of them:

Mary vs. Percy

9. Conclusion

The model trained on both Mary and Percy Shelley's works does an acceptable job classifying a sentence to its corresponding author (less than 600 misclassifications out of ~4K sentences in the test set), and as such, can give us a very rough estimate of an answer to Frankenstein's authorship question: **according to the model, the majority of Frankenstein was written by Mary Shelley, with Percy Shelley's "touches".**
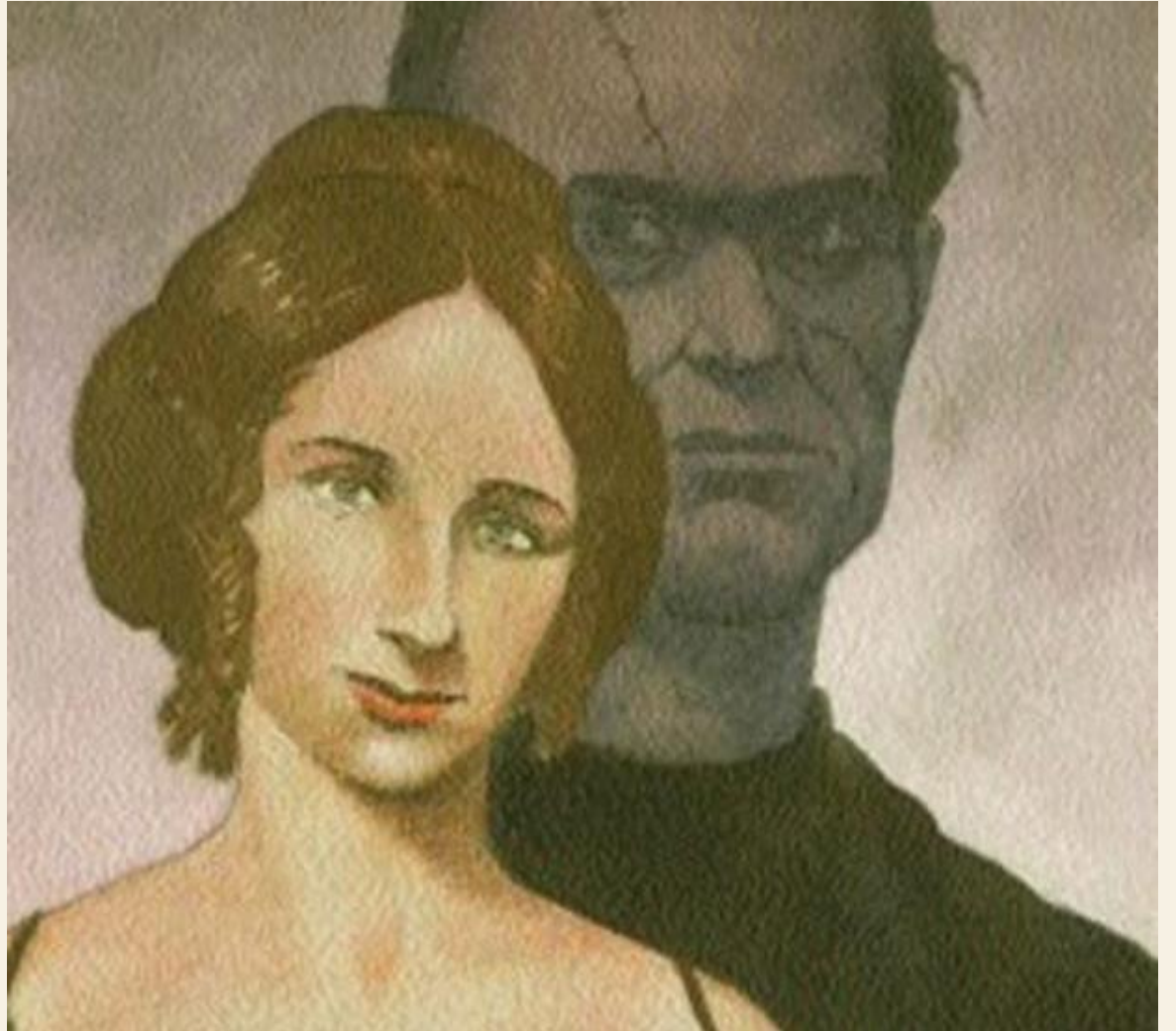
10. Discussion

My results are very similar to those of Charles E. Robinson's full survey (1996). Robinson used modern tools to create photofacsimiles , with them he created a new transcript. His conclusion was:

> As this edition will make evident, MWS is the creative genius by which this novel was conceived and developed, we can call PBS an able midwife who helped his wife bring her monster to life.
>
> — Charles E. Robinson, 1996

> PBS's contributions to Frankenstein were no more than what most publishers' editors have provided new (or old) authors or, in fact, what colleagues have provided to each other after reading each other's works in progress.
>
> — Charles E. Robinson, 1996

10. Discussion

It should be noted that even though the classifier achieved 87% accuracy, this result should be taken with a grain of salt:

- The training corpus was relatively small, with only about 24K sentences in total (12K for each author).

- 87% still means that more than one sentence in ten could be misclassified, and with Frankenstein's 3,083 tokenized sentences, this can add up to more than 300 misclassified sentences.

While Robinson estimated Percy Shelley have contributed around 4,000-5,000 words, my results show his contribution was as high as 8,280 words. This might indicate that further tuning is in order.

Moreover, we must take into consideration the fact that many hands are involved in every book publication (until this very day), therefore it cannot be said that any of the texts used in this project were written solely by either or both the Shelleys.

11. A Personal Note

As I read about Mary Shelley, I found out that although she never went to school, her father (William Godwin, a political philosopher) provided her a rich and informal education. His library contained not only his works, but also her late mother's (Mary Wollstonecraft Godwin) which had a reputation of her own as a feminist philosopher and writer.

With this background, I was somewhat confused by some of the widely accepted opinions which surrounded the Shelley's relationship. The assumption that Percy Shelley was the writer started when the novel was published anonymously, and this assumption was only strengthened when the draft found at 1974 had many notes which were automatically assumed to be Percy Shelley's.

11. A Personal Note

Some of the critics' conclusions were:

In short, by means of condescension and egotism, he suppressed his wife's distinctive vision. Worse still, Mary Shelley 'shared Percy's opinion of her inferior literary abilities' and 'substituted Percy's style for her own'. It proved their marriage was one 'in which the husband played the dominant role'

— Anne K. Mellor, 1988

Mary Shelley felt unable 'to put [her]self forward unless led, cherished & supported,' and she perceived this need for support as feminine, 'the woman's love of looking up and being guided'. It might be, then, that this ideology of dependent femininity rendered her unable to write her own text without her husband's help. Moreover, collaboration forced by a more dominant writer on a less powerful and perhaps unwilling 'partner' is a kind of rape; if Frankenstein is the product of such a union, then it evinces a debilitating femininity.

— Johanna M. Smith, 1992

11. A Personal Note

After Robinson has published his work, the critics' voices has change, and I am happy to find out that after all these years not only Mary receives the honor of being recognized for her work, but also Percy is looked upon as an encouraging husband who takes interest in his wife's work.

> As the years have passed, the idea that husband and wife worked alongside each other in ways that did not entail bullying, harassment, or physical violence has begun to find acceptance.
>
> — Robinson, 1996

11. A Personal Note

Those who knew Shelley testify to 'quite different qualities: directness, precision, irony, command, qualities frequently found in the early journal entries'. Indeed, in Mary's own self-assessments, Leader finds evidence of her 'belief in herself as an author'. It is more likely that, when writing Frankenstein, she 'consciously, willingly welcomed Percy Shelley's contributions'.

— Zachary Leader, 1996

11. A Personal Note

Percy's attitude towards his wife's writing was positive. He encouraged her to translate Alfieri's Mirra, 'supplied lyrics for her mythological dramas Proserpine and Midas (1820)', and welcomed her contributions to his translation of Spinoza's Tractatus Theologico-Politicus.

— Nora Crook, 2013

11. A Personal Note

I personally prefer to give both of them the benefit of the doubt and conjecture that (1) PBS suggested and made alterations to the text of Frankenstein for the purpose of improving an already excellent narrative (in [?February 1818] he wrote a review that judged the published novel "one of the most original and complete productions of the day") and that (2) MWS accepted the suggestions and alterations that she agreed with.

— Robinson, 1996

Image Sources

- By Robertgrassi - Own work, Public Domain, https://commons.wikimedia.org/w/index.php?curid=4463463

- By Richard Rothwell - Public Domain, https://commons.wikimedia.org/w/index.php?curid=59420814

- By After Amelia Curran - Public Domain, https://commons.wikimedia.org/w/index.php?curid=6370522

- Illustration of the Villa Diodati, British Library - Public Domain, https://www.bl.uk/collection-items/illustration-of-the-villa-diodati

- Frankenstein original cover 1818, https://todayinbritishhistory.com/2014/01/first-edition-of-frankenstein-mary-shelley-published/

- Mary Shelley and Frankenstein, couldn't find the source.

- Main draft of Frankenstein, the Bodleian Libraries at the University of Oxford, https://www.mentalfloss.com/article/80812/flip-through-mary-shelleys-draft-notebooks-frankenstein

External Links and Sources

- Duncan Wu (2015). *30 Great Myths about the Romantics*. WIley-Blackwell. p. 214. ISBN 978-1118843260.

- Shelley-Godwin Archive website http://shelleygodwinarchive.org/contents/frankenstein/the-frankenstein-notebooks-introduction

- Works by Mary Wollstonecraft Shelley at Project Gutenberg http://www.gutenberg.org/ebooks/author/61

- Works by Percy Bysshe Shelley at Project Gutenberg https://www.gutenberg.org/ebooks/author/1529

- The Frankenstein authorship question at Wikipedia https://en.wikipedia.org/wiki/Frankenstein_authorship_question

- Villa Diodati at Wikipedia https://en.wikipedia.org/wiki/Villa_Diodati

- מנהיים, נועה. *הרשת התרבותית. הוצאת גרף*, 2019, פרק 38.

- טריאסט איילת, מנהיים נועה. 27.6.2019, פרנקנשטיין (פרק 1) [פודקאסט] ב-*האחיות גרים*, גלי צה"ל.

https://glz.co.il/08-07-2020-2031האחיות-גרים/האחיות-גרים/תוכניות/גלצ

Thank You!

Galia Appel

https://github.com/Galia-A/frankenstein-ml