

Project Title:

Fine-tuning and Assessing NLP Models on Pushcart and Non-Pushcart Poetry

Team Members and Contributions

- **Mann Nada:** Led POS tagging, lexical statistical analysis, topic modeling, and sentiment scoring. Built Knowledge Graph pipelines and ran all notebook experiments end-to-end. Integrated transformer-based poem generation and prompt engineering.
-

Abstract

This project investigates stylistic and semantic differences between Pushcart-nominated and non-nominated poetry using Natural Language Processing (NLP) methods. We used parts-of-speech analysis, topic modeling, sentiment scoring, and knowledge graphs to identify linguistic trends that distinguish critically acclaimed work. Using GPT-style prompting, we generated poetry in selected styles and evaluated its alignment with Pushcart standards. Our findings indicate that lexical richness, abstraction, and conceptual layering are key distinguishing traits of Pushcart poetry. These insights can guide future AI-generated literary evaluation.

Introduction

This research explores what makes a poem "Pushcart-worthy" by comparing two categories: Pushcart-nominated poets like Edmund Spenser and lesser-known, non-nominated poets like Will Rout. By combining traditional NLP techniques and transformer-based generative tools, we attempted to simulate and distinguish poetic quality computationally.

Our objectives were:

- Quantify syntactic, semantic, and stylistic features of both poem sets
 - Use topic modeling and sentiment to extract common themes and tonal shifts
 - Generate poems using GPT-based prompting
 - Build and compare Knowledge Graphs derived from poetic texts
-

Literature Review

Prior research on AI and poetry includes stylometry, author mimicry using fine-tuned GPT-2/3 models, and emotion-driven text generation. However, limited studies assess award-worthy literary features or conduct direct comparisons across poet categories. Existing tools like BERTopic and LDA provide useful thematic modeling but are rarely applied to poetry. Our work integrates these along with semantic embeddings and graph-based techniques to build a comprehensive poetic analysis pipeline.

Methodology

We adopted a modular pipeline across four core components:

1. POS and Lexical Analysis:

- Tools: spaCy, pandas
- Metrics: Lexical density, noun/verb ratio, POS frequency
- Input: 10 poems each from Pushcart-nominated and Non-Pushcart categories

2. Topic Modeling:

- Tools: Gensim (LDA), BERTopic, pyLDAvis
- Methods: Tokenization, lemmatization, removal of stopwords, topic extraction, and visualization
- Input: Combined corpus grouped by poet type

3. Sentiment Analysis and Generation:

- Tools: VADER, DistilBERT, OpenAI GPT-3.5, Claude
- Methods: Sentence-level sentiment scoring; Prompt-driven poem generation with stylistic constraints

4. Knowledge Graph Construction:

- Tools: NetworkX, sentence-transformers, FAISS
- Methods: Entity extraction → KG formation → semantic embedding → article retrieval

Experimentation and Results

Below is a detailed breakdown of each homework-based experiment, including objectives, inputs and outputs, comparative strategies (what worked and what failed), and conclusions drawn.

Experiment 1: NLP POS Substitutions and Tones for Poets

Colab:

<https://colab.research.google.com/drive/1nBo5VyhlbMu1QGSpQklcQ74mshcLrIti?usp=sharing>

Objective: Analyze the stylistic impact of grammatical elements in poetry by computing POS statistics and testing the effect of POS substitutions on tone.

Inputs:

- 10 Pushcart-nominated poems (scraped from Iowa Pushcart site)
- 10 Non-Pushcart poems (collected from Poets.org)

Process:

- Used spaCy to perform POS tagging.
- Computed lexical density, noun/verb ratios, and POS proportions.
- Replaced adjectives and verbs using synonym/antonym dictionaries and observed tonal shifts in the modified poems.

Outputs:

- CSV table with all lexical statistics.
- Substituted versions of poems for manual inspection.

Comparative Results:

- Pushcart poems maintained emotional depth and coherence post-substitution.
- Non-Pushcart poems often lost coherence and showed abrupt tonal changes.

Failures & Learnings:

- Sentiment models like VADER struggled with archaic Pushcart diction.
- Adjective substitution led to tone flattening in modern, simpler poems.

Conclusion:

Higher abstraction and structure in Pushcart poetry made tone resilient to lexical

replacement. Non-Pushcart poetry became inconsistent or fragmented.**
Higher abstraction and structure in Pushcart poetry made tone resilient to lexical replacement. Non-Pushcart poetry became inconsistent or fragmented.

Experiment 2: Homework Week 8b NLP Part 2 – Topic Modeling

Colab:

<https://colab.research.google.com/drive/1V1DDxJkU746uZrWlQzy6JQkcZvm6QRbL?usp=s>
haring

Objective: Extract dominant themes from each poet group and evaluate how topic abstraction differs between Pushcart and non-Pushcart poems.

Inputs:

- Combined corpus of 20 poems, grouped into two categories.
- Preprocessing pipeline with stopwords removal, lemmatization, and frequency thresholding.

Process:

- Used Gensim LDA and BERTopic for separate topic extraction.
- Visualized results using pyLDAvis.

Outputs:

- Top 3 topics per group with 6-word keyword summaries.
- Coherence score comparisons between groups.

Comparative Results:

- Pushcart topics leaned abstract (soul, myth, virtue), with consistent emotional tone.
- Non-Pushcart topics were grounded in physical setting and urban life.

Failures & Learnings:

- BERTopic struggled with short, highly figurative poems (topic drift).
- LDA required manual tuning of topic count (k=3 performed best).

Conclusion:

Pushcart poems conveyed metaphorical and mythical elements; Non-Pushcart poems

leaned on everyday objects and events. Topic models showed higher coherence and deeper abstraction in Pushcart sets.**

Pushcart poems conveyed metaphorical and mythical elements; Non-Pushcart poems leaned on everyday objects and events. Topic models showed higher coherence and deeper abstraction in Pushcart sets.

Experiment 3: Week 9; Part 5 – From Poems to News Articles (KG's)

Colab:

<https://colab.research.google.com/drive/1FujWTVdabjj38lsjIXmHZjt7oqm4tAYD?usp=sharing>

Objective: Link poetic concepts to real-world contexts by building knowledge graphs and embedding them for semantic comparison.

Inputs:

- Pre-tagged poems from both poet groups.
- External article corpus to compare.

Process:

- Extracted named entities and noun phrases to form KGs.
- Used sentence-transformers to embed graph paths.
- Queried vector DBs (FAISS) for related real-world texts.

Outputs:

- Graphs visualized in NetworkX.
- Matched articles ranked by cosine similarity.

Comparative Results:

- Pushcart KGs produced 2x more cross-linked nodes.
- Their article matches leaned academic/philosophical.
- Non-Pushcart queries returned simple lifestyle pieces.

Failures & Learnings:

- Named Entity Recognition failed on metaphor-heavy lines.

- Some embeddings failed to cluster meaningfully without dense text.

Conclusion:

Knowledge graphs amplify conceptual understanding and indicate stronger thematic grounding in Pushcart texts.**

Knowledge graphs amplify conceptual understanding and indicate stronger thematic grounding in Pushcart texts.

Experiment 4: Homework 9 NLP – Analyzing Gold Standard Poems**Colab:**

https://colab.research.google.com/drive/1Ua7wQCrOkSu3pOilyPhKBx7v_UfMwYC?usp=sharing

Objective: Identify measurable linguistic patterns in Pushcart-nominated poems and build a classifier to evaluate poetic quality.

Inputs:

- 2022 Pushcart poetry samples scraped from official archive.
- Control set of 10 non-nominated poems.

Process:

- Used spaCy to extract POS and compute metrics.
- Constructed thresholds for a rule-based Pushcart classifier.

Outputs:

- Lexical and POS stats for all 20 poems.
- Classification label (Likely/Unlikely Pushcart).

Comparative Results:

- Classifier correctly predicted 80% of true Pushcart poems.
- Only 3/10 non-Pushcart poems crossed all thresholds.
- Top indicators: High adj/noun ratio and abstract word usage.

Failures & Learnings:

- Overfit on small sample set.

- Did not account for line structure or prosody.

Conclusion:

The POS profile of a poem is predictive of its critical acclaim in small-sample evaluations. Pushcart poems are linguistically denser and more compositionally balanced.**

The POS profile of a poem is predictive of its critical acclaim in small-sample evaluations. Pushcart poems are linguistically denser and more compositionally balanced.

Discussion

Linguistic metrics like noun/verb ratios and lexical density correlate well with literary quality. Topic modeling and sentiment scoring highlight structural depth and emotional variability in Pushcart poems. Knowledge graphs provide a novel interpretive layer to detect conceptual patterns and real-world alignments.

While GPT models can approximate tone, they struggle with generating layered abstraction without guided prompts. This limitation reinforces the human element of poetic excellence, but also illustrates where AI can aid creative writing and literary education.

Conclusion

This project demonstrated that a combined NLP+ML pipeline can partially emulate, differentiate, and evaluate poetic quality. POS metrics, LDA themes, and knowledge graphs are valuable tools in analyzing literary works.

Future work could involve:

- Full fine-tuning of a small GPT model on Pushcart data
 - Extending KG analysis to multi-lingual or historical corpora
 - Designing a generative poetry scoring framework
-

References

- spaCy: <https://spacy.io/>
- Gensim & LDA: <https://radimrehurek.com/gensim/>
- BERTopic: <https://maartengr.github.io/BERTopic/>

- Hugging Face: <https://huggingface.co/>
 - OpenAI GPT: <https://platform.openai.com/>
 - Iowa Pushcart Prize Nominations: <https://iowapoetry.com/pushcart2022.htm>
-

Appendices

- Appendix A: Delta Stats CSV
- Appendix B: Poem samples pre- and post-POS substitution
- Appendix C: LDA + BERTopic Visualizations
- Appendix D: GPT Prompt Templates
- Appendix E: KG Graph Screenshots