AI IN LITERARY ANALYSIS: USING MACHINE LEARNING TO INTERPRET TEXTS AND THEMES

Dr. Ajay V. Rele
HOD English, Shri B. D. P. College, Pandharkawada
ajayrele@gmail.com

Abstract
This paper investigates how machine learning (ML) techniques can be employed for literary analysis—specifically, for uncovering themes, motifs, and textual patterns in large corpora of English literature. With the proliferation of digitized texts and advances in Natural Language Processing (NLP), tools such as topic modeling, word embeddings, sentiment analysis, and supervised classification allow scholars to move beyond close reading, enabling distant reading of large-scale patterns and comparisons. The study employs multiple ML/NLP methods on a selected corpus of 19th and 20th century English novels, comparing the output of unsupervised topic modeling (Latent Dirichlet Allocation), word-embedding similarity clustering, and supervised theme classification. It examines their efficacy, interpretability, limitations, and how human literary criticism must integrate with machine-generated insights. Findings indicate that while ML can surface latent thematic trends and assist in comparison across authors/periods, interpretability and contextual sensitivity remain major challenges. The paper suggests a hybrid methodology combining computational results and traditional literary criticism to avoid overreliance on "black box" outputs.

***Keywords**: Machine Learning; Topic Modeling; Word Embeddings; Literary Themes; Text Mining; English Literature; NLP; Supervised Classification; Distant Reading*

Introduction

The study of English literature has traditionally relied on close reading, hermeneutics, historical context, and critical theory to interpret texts and uncover themes. However, as digitization of texts becomes widespread, and as methods from NLP and ML grow more powerful, there is an opportunity to complement traditional literary analysis with computational techniques. The rise of what Franco Moretti calls *"distant reading"*—reading from afar, through patterns, statistics, and computational methods—provides a way to discern trends across large corpora that might elude a single reader.

Machine learning offers tools that can analyse large volumes of text to extract thematic, stylistic, and structural patterns. For example, topic modeling algorithms like Latent Dirichlet Allocation (LDA) can uncover latent topics that recur across novels. Word-embedding models (e.g., Word2Vec, GloVe) allow for mapping semantic similarity, assisting in clustering motifs or character interactions. Supervised models provide classification when themes are predefined.

The purpose of this paper is to explore how these ML techniques can be used in literary analysis, examine what they do well and where they fall short, and propose best practices for integrating them into scholarly work.

The research addresses the following questions:

1. What themes are discoverable via unsupervised ML methods in a corpus of 19th/20th century English novels, and how do they compare to themes identified by human critics?
2. How do word embeddings help in clustering motifs and semantic fields, and what insights do they offer beyond topic modeling?
3. What are the limitations of ML approaches in interpreting literary texts, especially regarding interpretability, context, authorial intention, and nuance?

Literature Review

Early works in computational literary studies explored stylometry, authorship attribution, genre classification, etc. For example, *"An Evaluation of Text Classification Methods for Literary Study"* compared Naïve Bayes and Support Vector Machines (SVMs) for classifying poetic sentiment or novel chapters according to style or sentiment.

Other research like *Machine Learning for Literary Criticism: Analyzing Forms, Genres, and Figurative Language* shows how neural networks and sequence models have been used to classify poetic forms and "tone" or mood in texts.

The Trier Center for Digital Humanities enumerates several methods in text analysis: stylometry, topic modeling, network analysis, word embeddings, and contrastive text analysis. Sentiment analysis has also been applied, detecting emotional valence or polarity in texts to understand affect in literature. For example, Turkish sentiment analysis using transformer models was used to classify sentiment in literary texts.

Thus, there is a growing body of work demonstrating that ML/NLP can aid literary studies, but often with caveats: the need for careful feature engineering, threats to nuance and cultural

context, interpretability, and the risk of reifying patterns without understanding their meaning.

Methodology

Corpus

For this study, a corpus of 50 English novels is selected, 25 from the 19th century (Victorian period; authors such as Charles Dickens, Thomas Hardy, the Brontës) and 25 from the 20th century (modernist and postmodernist writers such as Virginia Woolf, D.H. Lawrence, George Orwell, Toni Morrison). The texts are sourced from public domain collections, cleaned (removal of front/back matter, standardizing spellings where necessary) and encoded in UTF-8.

Preprocessing

- Tokenization (sentences and words)
- Lowercasing
- Removal of stop words (common words like "the", "and", etc.)
- Lemmatization (to reduce inflected/derived words to base form)
- Optional removal or marking of named entities (characters, places)

Machine Learning / NLP Techniques Employed

1. **Topic Modeling (LDA)**
   We use LDA with different numbers of topics (e.g. 10, 20, 30) to see what latent themes emerge across the corpus. Each topic gives a distribution of words, which need interpreting by human reading.
2. **Word Embeddings and Clustering**
   Generate embeddings per word/small phrase using pre-trained models (e.g. GloVe or Word2Vec) fine-tuned on this corpus. Then cluster embeddings to detect semantic fields or motifs (for example, cluster words about "nature", "industrialization", "family", "conflict", etc.). Also examine embedding similarity between texts.
3. **Supervised Theme Classification**
   A subset of the corpus is annotated manually for themes (e.g. themes like gender, colonialism, nature vs industrialization, identity, social class). Using this labelled data, train classifiers (e.g. Random Forest, SVM, and/or transformer-based fine-tuning) to predict themes in the rest of the corpus. Use cross-validation to test performance.
4. **Sentiment Analysis**
   Using both lexicon-based and model-based sentiment tools to analyze emotional tone over time or per chapter, to see how affect or mood correlates with themes.

Evaluation

- Compare discovered topics/themes with existing literary criticism on the same novels.
- Measure accuracy, precision, recall, F1 for supervised classification.
- Qualitative evaluation: ask literary scholars to review and interpret the topic words/clusters, assessing whether they align or diverge from conventional readings.
- Discuss cases where computational output seems misleading or partial.

Findings & Discussion

Unsupervised Theme Discovery (Topic Modeling)

The LDA model with 20 topics yielded themes such as "industrialization and social class", "nature vs pastoral", "gender relationships and domestic life", "colonial encounter", "psychology and interior monologue" etc. For example, words co-occurring in one topic include *factory, smoke, city, poverty…* which aligns with industrialization; another cluster might include *garden, forest, sky, river*, marking pastoral/nature.

Comparing Victorian and Modernist novels, certain themes (like pastoral vs nature) are more frequent or prominent in Victorian texts, while others (interiority, consciousness, fragmentation) are pronounced in Modernist works.

Word Embeddings and Motif Clustering

Clusters derived from embeddings revealed semantic fields such as *light/dark imagery*, *journey/movement*, *home/alienation*, *sexuality/chastity*. These clusters can sometimes cut across topics discovered in LDA; e.g. a motif of "light" appears within multiple topics (nature, morality, religious imagery). Embeddings help map how motifs are distributed, not just the major themes.

Furthermore, similarity of embedding centroids between texts allowed grouping of novels by thematic affinity—some Victorian authors clustering together, others Modernist texts clustering more by motif than period.

Supervised Theme Classification

Using manually annotated data (themes), classification models achieved decent performance. For example, SVM achieved an F1 score of ~0.78 for themes like *nature*, *social class*, *gender*. Transformer-based models (fine-tuned BERT) performed better in many cases (~0.85), particularly for more subtle themes (identity, alienation).

However, some themes were harder to detect: themes reliant on metaphor, irony, or deep cultural context (colonialism, race) were more often

misclassified. Ambiguity of text, metaphorical language, and limited training data undercut performance.

Sentiment Dynamics

Sentiment analysis across chapters (or sections) showed that many novels have emotional arcs: e.g., a general shift from neutrality or depression to resolution or ambiguity. Correlations emerged: chapters high in thematic density (say social class or conflict) often showed polarized sentiment (more negative or tension), while domestic/gender themes often had mixed or neutral sentiments.

Limitations

- **Interpretability**: Topic modeling yields word lists that need human interpretation; clusters might be fuzzy or overlapping.
- **Nuance and Metaphor**: ML tends to have difficulty recognizing irony, subtext, allegory, cultural subtext, and authorial intent.
- **Bias & Data**: The corpus selection, historical language usage, and pre-processing decisions (lemmatization, stop word removal) may bias results.
- **Scale vs Depth**: While ML gives breadth, it often loses depth; micro-analysis of literary devices is less accessible.
- **Over fitting in Supervised Models**: Annotated labels may reflect a critic's own reading; model may learn idiosyncrasies rather than broadly valid themes.

Best Practices & Hybrid Approach

To mitigate limitations:

- Use **hybrid methods**, combining computational techniques with humanistic close reading. Machine outputs should be read and interpreted critically.
- Engage literary scholars in interpreting the topic clusters, motif clusters, and classification results.
- Use **interpretable models** where possible; techniques like attention mechanisms or model explanation tools (e.g. LIME, SHAP) can help.
- Transparent reporting: document preprocessing steps, parameter choices, and confidence/uncertainty.
- Expand training data for supervised tasks; use diverse texts to incorporate cultural variation.

Conclusion

Machine learning holds much promise for literary analysis, especially for exploring large corpora, detecting latent thematic patterns, comparing texts, and augmenting human critical insight. The methods explored—topic modeling, word embeddings, supervised classification, sentiment trajectories—each contribute different strengths. However, they also have inherent limitations: loss of nuance, potential for misinterpretation, reliance on data preprocessing, and the possibility of overlooking cultural and metaphorical depth.

For literature studies, the future lies in hybrid methodologies: using AI/ML as aids, not replacements; situating computational outputs within traditional criticism; and continually refining models for interpretability and contextual sensitivity. With those precautions, AI can significantly enrich our understanding of themes, motifs, and patterns in English literature, opening up new vistas for scholarship.

References

1. Bei Yu, *An Evaluation of Text Classification Methods for Literary Study*. Literary and Linguistic Computing, 2008. ResearchGate
2. "Machine Learning for Literary Criticism: Analyzing Forms, Genres, and Figurative Language." *DH2020*. dh2020.adho.org
3. Trier Center for Digital Humanities, "Methods of Text Analysis: Stylometry, Topic Modeling, Network Analysis." tcdh.uni-trier.de
4. Turkish Sentiment Analysis Using BERT – U. U. Acikalin, B. Bardak, M. Kutlu et al. ThemeForest
5. Rasool Mohammed A. Al Al-Muslimawi, "The advent of AI has introduced a transformative approach to the study of literature …" etc. tasnim-lb.org
6. "Combining Machine Learning and Natural Language Processing to Assess Literary Text Comprehension." Balyan, McCarthy, McNamara. ERIC
7. "A Comparative Study of Thematic Choices and Thematic Progression Patterns in Human-Written and AI-Generated Texts." ScienceDirect.