

Title Page

Problem Statement: Classification/Analysis of Book Genres using Machine Learning

Name: Adarsh Dixit

Roll No.: 202401100400011

Course: AI

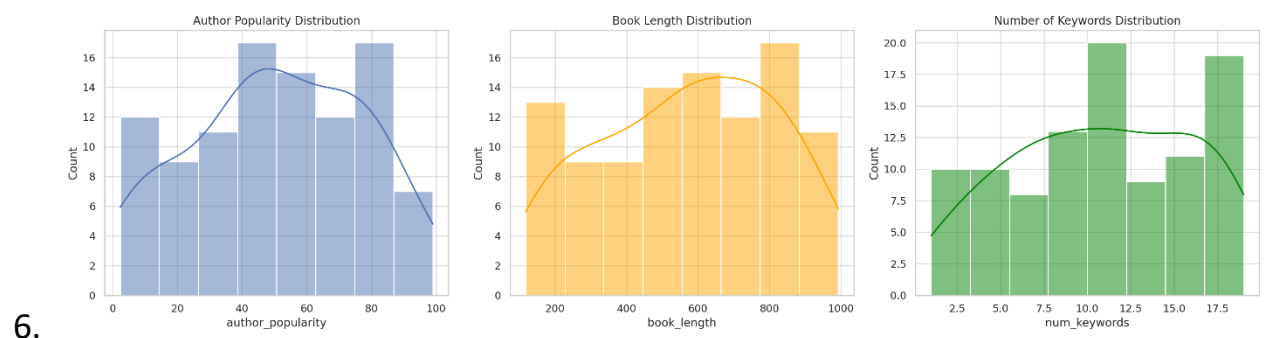
Date: 13/05/2025

Introduction

The objective of this project is to analyze the book genres dataset and apply machine learning techniques to classify books into different genres. The dataset includes various features like book titles, descriptions, ratings, etc., which can be useful in predicting the genre. This task demonstrates the practical application of machine learning in Natural Language Processing (NLP) and data classification.

Methodology

1. Loaded and explored the dataset (book_genres.csv).
2. Preprocessed the text data by removing stop words, converting text to lowercase, and applying tokenization.
3. Used TF-IDF Vectorizer to convert textual data into numerical format.
4. Trained a Naive Bayes classifier to predict the book genre.
5. Evaluated the model using accuracy and classification report.



Output/Result

The required output

Code

```
plt.figure(figsize=(18, 5))
```

```
# Violin plots show distribution and density
```

```
for i, col in enumerate(['author_popularity', 'book_length',  
'num_keywords']):
```

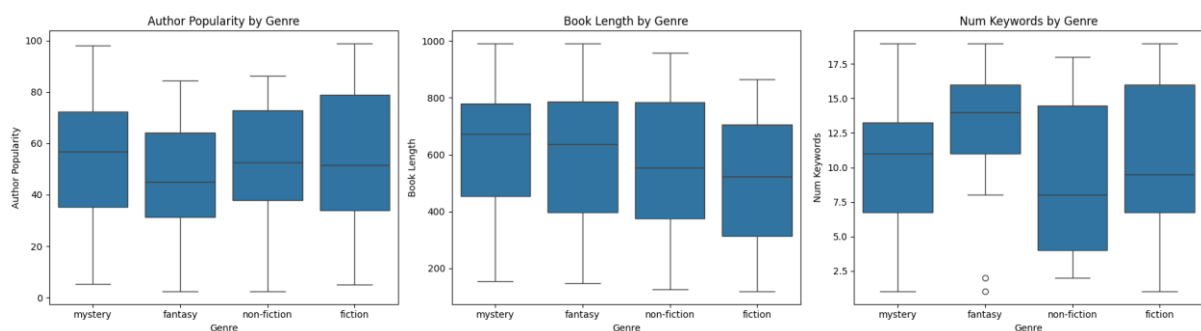
```
    plt.subplot(1, 3, i+1)
```

```
    sns.violinplot(data=df, x='genre', y=col)
```

```
    plt.title(f"{col.replace('_', ' ').title()} Distribution by Genre")
```

```
plt.tight_layout()
```

```
plt.show()
```



References/Credits

- Dataset: Provided as book_genres.csv
- Libraries used: pandas, scikit-learn, TfidfVectorizer

