AI Exam Report: Book Genre Classification

Objective

To classify books into their correct genre using metadata features like Author Popularity, Book Length, and Number of Keywords.

Dataset Overview

The dataset includes the following columns:

- author_popularity
- book_length
- num_keywords
- genre (target variable)

Genres: Fantasy, Fiction, Mystery, Non-Fiction.

Model and Method

Algorithm: Random Forest Classifier

Data split: 80% training, 20% testing

Evaluation Metrics: Accuracy, Precision (macro), Recall (macro), F1-score, Confusion Matrix

Results

Performance Metrics:

- Accuracy: 0.50
- Precision (macro avg): 0.64
- Recall (macro avg): 0.60

AI Exam Report: Book Genre Classification

Classification Report (per genre):

Fantasy - Precision: 0.25, Recall: 0.50, F1-Score: 0.33, Support: 4

Fiction - Precision: 1.00, Recall: 1.00, F1-Score: 1.00, Support: 1

Mystery - Precision: 0.63, Recall: 0.50, F1-Score: 0.56, Support: 10

Non-Fiction - Precision: 0.67, Recall: 0.40, F1-Score: 0.50, Support: 5

Analysis

The model achieves an accuracy of 50%. It performs best on Fiction but shows weaker results for Fantasy. Macro-averaged metrics indicate fair overall performance.

Improvements

Suggestions:

- Use more data
- Engineer better features (e.g., NLP on keywords)
- Try different classifiers
- Handle class imbalance

Conclusion

Random Forest provides a starting point for genre classification using metadata. There is potential for significant improvement through feature and model enhancement.

Confusion Matrix

AI Exam Report: Book Genre Classification