Warzer Mohammed / 01-22-00117

Special Topics in Software Engineering

University of KURDISTAN Hewlêr

Book Genre Predictor

1. Project Goal

My project is a program that predicts what genre a book belongs to. You can type in details like the book's title, author, and some keywords, and it will try to guess the genre (like Fantasy, Mystery, Romance, etc.).

2. What It Does

This project has three main parts that work together:

- **Preparing the Data (preprocess.py):** This script is like cleaning up ingredients before cooking. It takes raw book information (titles, authors, tags) and makes it neat. It also figures out the main genre for each book from its tags, because the original tags are too many. It creates new useful numbers about books (like how many ratings they have).
- **Teaching the Computer (train_model.py):** Once the data is clean, this script "teaches" a computer model (called a Random Forest Classifier) how to identify patterns in the book data that lead to different genres. It uses both the cleaned text and the new numbers we created. After teaching, it saves the trained model so we can use it later.
- **The App You Use (gui_app.py):** This is the part with a window where you type in book details. It loads the "taught" model and uses it to make a prediction. Then, it shows you the predicted genre and how confident the model is about its guess.

3. Key Files

Here are the important files in my project:

- archive/: This folder holds the original book data.
- preprocess.py: The script that cleans and prepares the book data.
- train_model.py: The script that trains the genre prediction model.
- gui_app.py: The application with the user interface.
- cleaned_books_genre.csv: This file stores the prepared book data after preprocess.py runs.

- book_genre_model.joblib: This is the saved "brain" of our model after train_model.py finishes.
- tfidf_vectorizer.joblib: This is a tool saved by train_model.py that helps the app understand new text you type in.
- requirements.txt: A list of all the special tools (Python libraries) needed to run this project.
- README.md: A document explaining how to set up and run the project.

4. Tools I Used

I built this project using:

- **Python:** The programming language.
- **Pandas and NumPy:** For handling and working with the data.
- **Scikit-learn:** This is the main library for the machine learning part (training the model).
- **Tkinter:** For building the graphical app.
- **NLTK:** For cleaning up text better.

5. Conclusion

This project successfully builds a simple machine learning system. It takes raw book data, processes it, trains a model to predict genres, and then lets users get predictions through a simple interface.

Warzer Mohammed / 01-22-00117

Special Topics in Software Engineering

University of KURDISTAN Hewlêr

Some Screenshots of the GUI with examples