BookGenics

Course name: Software Engineering (CSN – 302)

Supervisor: Prof. Mayank Gupta

Team Members (Group ID – 15):

- Shivam Raina (18103015, shivamraina.be18cse@pec.edu.in)
- Akshit Monga (18103074, akshitmonga.be18cse@pec.edu.in)
- Harshit Mittal (18103113, harshitmittal.be18cse@pec.edu.in)
- Neetish Kumar (18103009, neetishkumar.be18cse@pec.edu.in)

Project Introduction and Relevance:

The goal of the project is constructing a website/webapp which will have a user-friendly interface capable of predicting genres of short and long stories from their pdf/docx files. It is difficult for humans to classify the book based off its title or cover and quite tedious to manually read the whole book. So, this project is an attempt to provide the solution of the above problem. We didn't find any particular organized work regarding the same, so, we decided to inculcate this in our project.

Working/ Technology:

We plan to use NLP to achieve this. We will use a supervised classification model based on the concepts of topic modelling and topic classification. The concepts of multi label text classification will be used to classify the text. The technology stack that will be used is MERN stack, we are aimed at making the website using the cutting-edge technologies like React.js, Node.js, Express.js, Scikit Learn, etc. The languages used will most probably be JavaScript and Python. The database used will be mongo DB as we have to store json formats also.

Features:

It will contain the following features:

- Every user must have an account to use the features of the website.
- During Registration, a user must select his favorite genres.
- On the home page, user can see the list of reduced versions of books based off his preferred genres and can expand to view the summary of each book.
- User can add any book to his favorites.
- User can sort the books in any order based off the date on which it was uploaded and also apply multi-filtering based off Genres, Title, Author or Uploader of the book. Searching based off Title, Author or Uploader uses

Regex searching.

- User will be able to give the summary of book pdf/doc as an input and the backend will predict its genre(s).
- User can also add the book the database. The book will be cached in the database for future references.
- A User can edit/delete only those books which he has uploaded whereas Admin will have the power to edit/delete any book present in the database.
- A User can also change his account settings. (general details, password, genres preferred, etc.)

Scaling:

- A payments system can be setup in the future for premium usage.
- Also, the input type can be extended to json/bson format files of the book.
- User can give book as input and will get the summary of that book.
- Based on the searches of the user, he will also be recommended books

Tech-Stack Used:

- **Backend:** Node.js, Express.js
- **Frontend:** React.js
- **Database:** MongoDB
- **ML-Model:** Scikit-Learn (Machine Learning Library), Matplotlib and Seaborn (for visualizations), NLTK (Natural Language Processing Library) and some other Python Libraries.