**Assignment**

Siddhartha Prithvi Chebiyyam

University of the Cumberlands

Artificial Intelligence for Human-Computer Interaction

Prof. Ying Lin

August 10, 2025

**Content-Based Book Recommendation System using TF-IDF and Cosine Similarity in Python for Enhanced Human-Computer Interaction**

**Abstract**

Smart recommendation systems play a central role in defining modern experiences of human-computer interaction. All of this is made possible in these systems as they allow personalised communication, informational overload, and enhance the efficiency of decisions. The project was aimed at developing a Python-based book recommendation system based on content-based filtering. The system could recommend analogous book titles based on a user's choice by doing text vectorisation with Term Frequency and Inverse Document Frequency to use in calculating content similarity, as well as using cosine similarity with it. The implementation encouraged the possibility of easily handling it using a terminal interface, thus playing a real recommendation assistant. The fact that it was successful in producing recommendations confirmed the model design. Additional comments outlined shortcomings of the system, areas to be improved upon, and the provision of the changing role of intelligent systems to offer proactive user engagement.

**Introduction**

Digital platform expansion has escalated the necessity to have smart systems to give customizable user experiences. Such intelligence can be encountered in the form of recommendation systems found on media streaming, shopping, education, and digital libraries across platforms. A recommendation system is mainly designed to identify the needs or interests of the user and provide him/her with content that goes in line with their preferences. By integrating such systems into the common interfaces, it is possible to drastically improve the human-computer interaction. This project presented a book recommendation system that ran on the similarity of the contents, as opposed to user behaviour. Natural language processing and machine learning were employed in the methodology aimed at developing an intuitive but efficient recommendation system. The system presented a somewhat interactive approach in which users would get real-time suggestions with regard to a chosen book title. The project has shown that the least amount of input by the user can lead to a meaningful result, highlighting the capabilities of intelligent systems to make the digital interaction as simple as possible.

**Technical Background and Recommendation Techniques**

The recommendation systems usually fall into three categories. Collaborative filtering is based on the actions of users in the form of ratings, clicks, or preferences. This technique was used in this project and is called content-based filtering. It is based on the use of item attributes to identify the similarity among items. Mixing types are hybrid versions of both techniques. Content-based filtering works well in a scenario where the specific user information is not provided. In this model, the book title and the name of the author were the main features that were considered regarding similarity estimation.

Vectorisation of the text characteristics was performed with the assistance of Term Frequency-Inverse Document Frequency. In this approach, each word was given a weight whereby a larger such as value meant more significance in a document and less usage in the whole. TF-IDF allowed a precise mathematical comparison by converting the book metadata into a numerical form. Cosine similarity was then used to estimate the similarity of books by computing the angle between two vectors. This meant that a smaller angle was more similar. This mathematical method provided the book recommendation scheme based on books that were very related to the chosen one in terms of content. TF-IDF in conjunction with the cosine similarity gave it the framework of a lightweight and scalable system with applications spread across various areas that make use of text.

**Methodology and Implementation Workflow**

The first step of the implementation was reading a dataset of books with the pandas library. This data set included several fields, yet only the titles of books and authors were maintained to analyse. To achieve data quality, cleaning operations were used to remove missing entries or duplicate entries. The dataset had to be narrowed down to a feasible number of 3000 records in order to improve system performance. By combining book titles and authors, a new content field was produced. This mixed text was used as a feed into the vectorisation.

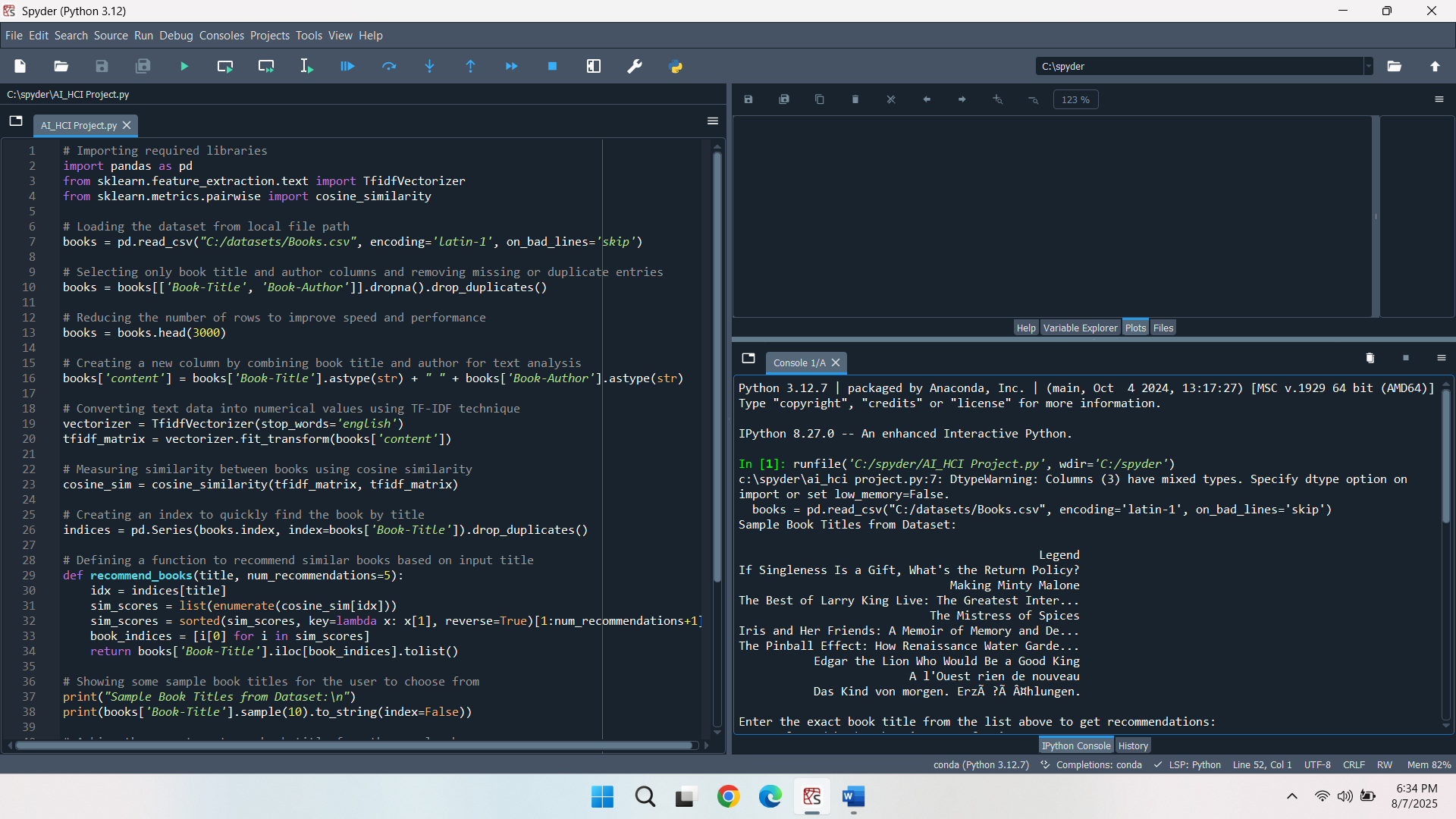
The scikit-learn library contains TfidfVectorizer, which was applied to convert the textual information to the weighted term frequency matrix. The output matrix mapped each book to a high-dimensional vector (Chibb et al., 2024). This matrix was then used in cosine similarity to get the similarity score of each book with each of the other books in the dataset. A look-up index was also constructed so a quick access by any book could quickly access its respective row within the matrix. The recommendation functionality was also exercised to find the most comparable book titles using a selected book.

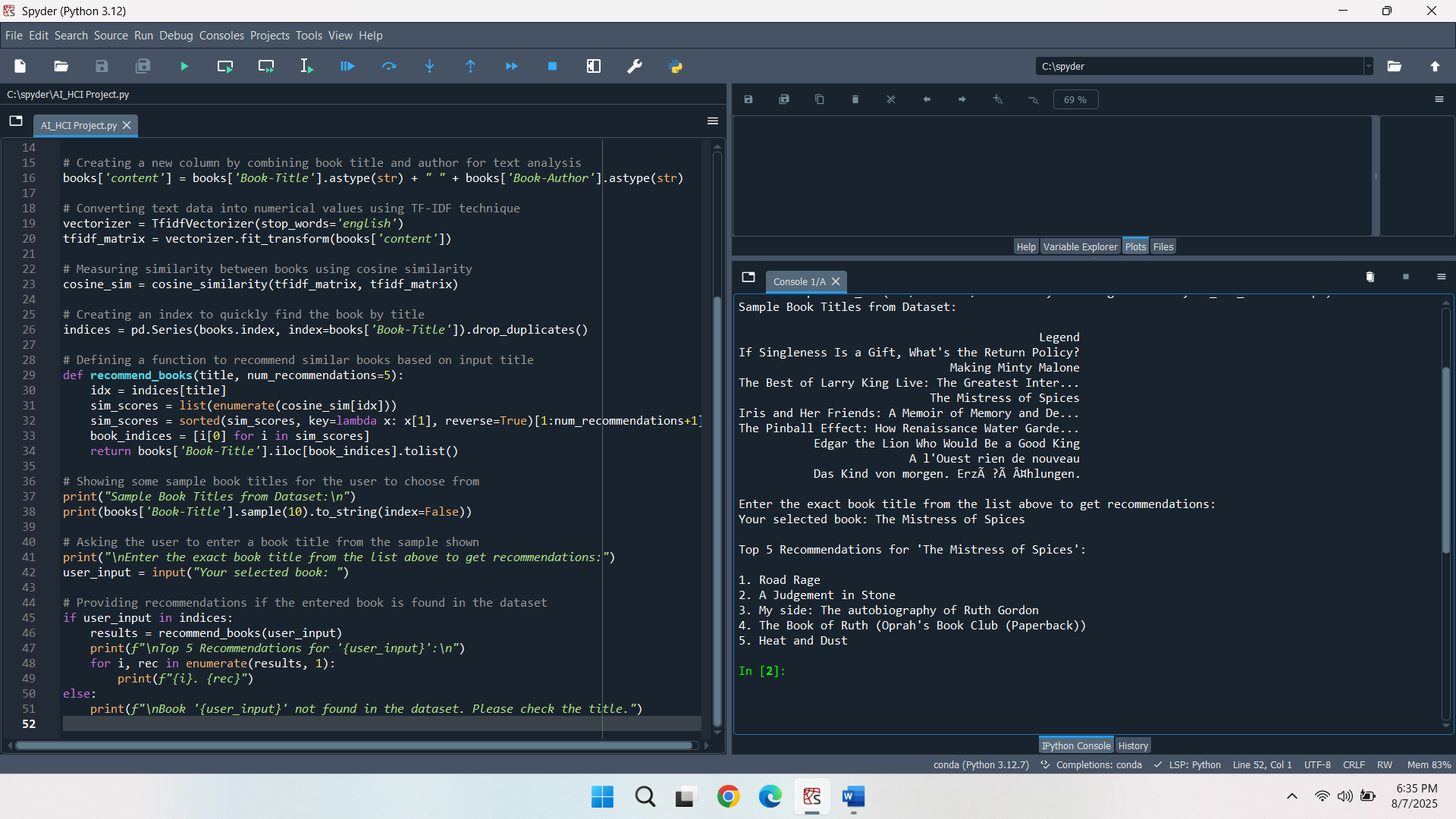
Ten titles of books randomly chosen from data dataset were shown on the screen to prompt the user. After feeding the text, the system compared the book with the index and responded with the five most similar ones. Its product was directly printed on the terminal interface.

**Results and Model Performance**

The model made appropriate recommendations related to content similarity. An example is when the user selected the book titled The Mistress of Spices, the system came up with five suggested books such as Road Rage, A Judgement in Stone, The Book of Ruth, and Heat and Dust. Such outputs revealed the fact that the TF-IDF and cosine similarity methods effectively detected the contextual patterns in the text. Books that comprised the sample had stylistic, thematic, or lexical similarities with the title input.

The recommendation request was handled by the system in real time, which gave close to instantaneous results. The reduced user interface was also a one-input mimicking the model, which imitated low-effort interaction. This design was compatible with fundamental concepts of HCI, which promote intuitiveness in system behaviour. Lack of user preference or history was not a hitch in performance, as it proved a sound piece of the content-based model. The project ensured that it was possible to produce useful recommendations based on textual metadata alone. The consistency of results was also witnessed when tested under various book titles, which served to prove the reliability of the model against varying inputs.





**Discussion and Role in Human-Computer Interaction**

The recommendation engine demonstrated the possibilities of how artificial intelligence and machine learning may assist human-computer interaction, making the process of content discovery simpler. The system minimised the effort of a user, requiring only one book title and giving them individual suggestions. This model was also the origin of initial versions of adaptive systems that learn about user preferences using only some input. These systems make users more satisfied as they deliver interaction design matching user intent.

As successful as it was, the current model had limitations. The system was prone to suggesting similar wording literature, but the subject is not related. This is a typical disadvantage of content-based filtering called overspecialization. Lack of contextual metadata was a limitation to the recommendations (Alkatheiri, 2022). Moreover, the model did not have a way to consider user feedback that could be used to increase accuracy with time. These restraints are points of potential future improvement.

The programming of this system is in line with the current changes in human-computer interaction. Smart systems will no longer be limited to being servants instead, they will be servitors. Next-generation systems are going to interpret user behaviour, context and preferences without having to be told anything. Communication can no longer be based on typed messages, but could involve voice, gestures, and the environment. All these developments will enhance the system to predict and initiate a conversation, generating a smooth and customised user experience.

**Future Enhancements and Practical Applications**

There are various ways the system can be improved to be more functional and friendly. More metadata would make recommendations more pertinent. Collaborative filtering may be used to inject user feedback and behavioural patterns. This would enable the system to diversify and prevent redundancy. The terminal-based interface may be replaced by a graphical interface with the help of Flask, Streamlit or Dash. Such interfaces would render the system more user-friendly to coding amateurs (Kraus, 2021).

It would allow voice interaction with integration into speech recognition tools. This would be particularly useful within mobile or voice-assistant conditions. APIs would facilitate connections to live data sources and would allow the system to update in real time as well as to cover a greater breadth of content. Lastly, implementing the model in a cloud computing environment would facilitate scalability that would permit its application in large-scale systems like online bookstores, online libraries or e-learning systems.

**Conclusion**

The project was able to show how a content-based book recommendation engine can be implemented through artificial intelligence tools. This model used the vectorisation of TF-IDF and calculated content closeness through the cosine similarity. The results confirmed the functionality of the model to come up with valid and contextually relevant recommendations. The project underlined the significance of intelligent systems to enhance the user experience and the notion of human-computer interaction. Although the model included some limitations, it created a basis on which more elaborate applications can be derived, which combine user data, multi-modal interaction and real-time responsiveness. The smarter systems of the future will be the ones that anticipate, respond, and interact in manners that fall within human wants and social patterns. This project helped to achieve that vision by providing a solution based on practicality, scalability, and the intelligence of book recommendations.

**References**

Alkatheiri, M. S. (2022). Artificial intelligence assisted improved human-computer interactions for computer systems. *Computers and Electrical Engineering*, *101*, 107950.

Chibb, M., Vashisht, P., Katti, A., & Nrang, A. (2024, November). Enhancing Movie Recommendations: A Content Based Approach Using TF-IDF Weighted Word2Vec and Cosine Similarity. In *2024 4th International Conference on Technological Advancements in Computational Sciences (ICTACS)* (pp. 1449-1454). IEEE.

Kraus, M., Wagner, N., & Minker, W. (2021, October). Modelling and predicting trust for developing proactive dialogue strategies in mixed-initiative interaction. In *Proceedings of the 2021 International Conference on Multimodal Interaction* (pp. 131-140).