DIPLOMA: INFORMATION TECHNOLOGY

Subject	: Business Analysis 3 Module 2/ Business Analysis 3.2
Subject code	: AIBUY3A
Assessment	: Module ONE (Team Project)
Lecturer	: Mr. Matsela
Examiner	: Mr. Matsela
External Moderator	: Ms. S Matyila
Due date	:06 September 2023

Table of Contents

SOLUTION

As Codecrafters team we have decided to create an AI book recommendation system that will analyze a user's reading history, preferences, and behaviors. It then suggests personalized book recommendations based on these insights. By understanding individual tastes, this AI enhances the library experience, encouraging users to discover and engage with books that align with their interests. It promotes reading diversity, increases user satisfaction, and strengthens community ties through a tailored and engaging reading experience.

MAIN OBJECTIVES

Our main objective is to enhance the library experience and promote literacy within the chosen community by harnessing AI's power to provide personalized book recommendations. By implementing an AI-driven recommendation system, we plan to analyze users' reading preferences, history, and community-specific factors to suggest books and resources that are highly relevant to them.

Solving this problem using AI will benefit the local municipality and the community in several ways:

1. Increased Literacy: Personalized book recommendations will encourage more people to read, thereby promoting literacy and lifelong learning.

2. Efficient Resource Utilization: The AI system will optimize library resources by ensuring that books are borrowed and read more frequently, reducing resource waste.

3. Community Engagement: A more user-friendly and personalized library experience will attract and retain library users, fostering a sense of community engagement.

4. Inclusivity: The system will ensure that a wider range of books, including culturally relevant ones, are accessible to all members of the community, promoting inclusivity and diverse perspectives.

In summary, our AI solution for communities focuses on leveraging technology to enhance access to knowledge and promote literacy, benefiting both the local municipality and the community by optimizing library resources and improving the overall user experience.

Requirements, Constraints and Risks

Data Privacy: Strictly adhere to data privacy regulations to protect patron information.

Limited Data: Initial data may be limited, impacting the effectiveness of the recommendation system.

Resource Availability: Availability of skilled AI professionals and computational resources.

User Resistance: Some users may be hesitant about AI-driven recommendations, necessitating user education.

Algorithm Accuracy: Ensuring the accuracy and fairness of recommendations is crucial to user acceptance.

Initial assessment of tools and techniques

Collaborative filtering recommends items to users based on the preferences and behaviors of similar users or items.

Content-based filtering in a book recommendation system to suggest books to users based on the attributes and characteristics of the books themselves, such as genre, author, or keywords.

Data collection and preparation:

Used Python library called pandas for data manipulation and cleaning.

Model training and testing:

For splitting data into training and testing sets we use NumPy train-test split function.

Success Criteria

Accuracy: it accurately predicts what books the users will enjoy based on the data it has about them

Relevance: Recommends books that are highly relevant to the user's search, history, interests, and preferences.

Diversity: the recommendation is not limited to a narrow set of genres or authors, it introduces a diverse range of books and genres

PROBLEM DEFINITION

We aim to address the challenge of limited access to relevant books and resources, particularly in underserved communities. The problem we aim to solve is the lack of personalized book recommendations and the underutilization of library resources. Many library users, especially in underdeveloped areas, struggle to find books that align with their interests and needs, leading to reduced engagement with the library and missed educational opportunities. This problem aligns closely with the theme "an AI Solution for communities." as it focuses on utilizing AI to enhance community access to knowledge resources. Solving this problem through AI-based personalized book recommendations will significantly benefit the community and local municipality.

POSTER

VANDERBILJPARK LIBRARY

BOOK RECOMMENDATiON SYSTEM

SOLUTION

Recommendation System: Implement an AI recommendation system that suggests books, articles, or other materials to library users based on their past preferences and reading habits.
Benefit: It promotes reading diversity, increases user satisfaction, and strengthens community ties through a tailored and engaging reading experience

BUSINESS OBJECTIVES

Enhance the library experience and boost literacy in our community with AI-driven personalized book recommendations. This AI system tailors suggestions by analyzing user preferences and community factors.
We aim to achieve the following:
1. Boosted Literacy: Encouraging reading promotes lifelong learning.
2. Resource Optimization: Efficient library resource use reduces waste.
3. Community Engagement: A personalized experience fosters a sense of belonging.
4. Inclusivity: Access to diverse, culturally relevant books enriches perspectives.

TOOLS & TECHNIQUES

PYTHON
VS CODE
COLABORATIVE & CONTENT-BASED FILTERING
CLUSTERING
NumPy
PANDAS

MACHINE LEARNING

- *USED UNSUPERVISED MACHINE LEARNING*
- *USED CLUSTERING TO GROUP READINGS BASED ON THEIR READING HISTORY.*
- *USED ASSOCIATION TO GATHER DATA ON LIBRARY'S BOOK COLLECTION AND USER BORROWING HISTORY.*

DEEP LEARNING

Empowers advanced feature extraction capabilities through CNN to learn complex patterns for more effective search and recommendation systems for personalized book suggestions.

VAAL UNIVERSITY OF TECHNOLOGY
APPLIED & COMPUTER SCIENCES

python

Theoretical Aspect

MACHINE LEARNING APPROACH

We used an unsupervised machine learning approach for our solution. Our hybrid model combines content-based filtering, collaborative filtering, and K-Means clustering, ensuring a well-planned and appropriate set of algorithms for the solution.

Data and collection preprocessing:

Gather data on the library's book collection, including book titles, authors, genres, publication dates, and user borrowing history. This step ensures that data relevant to the solution is clearly articulated and covers different forms of data.

User profiling:

Create user profiles that represent the reading history of each library patron. This can be done by encoding their borrowing history in a structured format, such as a user-item interaction matrix.

User Registration:

Allow users to create profiles, where they can specify their age and gender, enabling us to cluster them by reading preferences, favorite genres and authors using K-Means.

Behavior Tracking: Continuously update user profiles based on their reading and borrowing history, books they've rated or reviewed, and any explicit preferences they've set.

Recommendation Algorithms:

Collaborative Filtering: Implement collaborative filtering algorithms to recommend books based on the reading and borrowing history, including preferences of similar users. This can help users discover books that are popular among people with similar tastes.

Content-Based Filtering**:** Use content-based filtering to recommend books based on the content of the books users have previously read or borrowed and enjoyed. This involves analyzing book descriptions, author information, and genre tags.

Search functionality

- Allow users to search for books and authors directly. Book recommendations are integrated into the search results, promoting user discovery and improved user experience .

Clustering

- Apply clustering algorithms, such as k-means, DBSAN, or Hierarchical Clustering, to group similar users based on their reading histories. With each cluster representing a group of users with similar reading preferences.

User interface

- Integrate the recommendation system into the library's platform. Provide an interface where users can see their recommendations and explore additional books based on their cluster. This approach ensures that the model is user-centric and achievable, and that its performance is evaluated for accuracy.

DATA

When creating a dataset for a machine learning project that focuses on recommending books using an unsupervised approach, you'll need to structure the data in a way that allows your unsupervised learning algorithm to identify patterns and generate

recommendations based on user behavior and book attributes. Here's how you can prepare and structure such a dataset:

Data schema

- User data: Collect information about users, such as a unique user ID and any demographic information (age, gender, location) if available and relevant.

Data collection

- Gather data from various sources, such as library borrowing records, user-generated reviews and ratings, and book attributes from your library's catalog. Ensure that you have sufficient data to create meaningful recommendations.

Data preprocessing

- Data will be pre-processed and cleaned using pandas to handle missing data, outliers, and inconsistencies.

Data transformation

- For clustering, you can use the user-item interaction matrix directly. Each user's pattern of book interactions becomes a feature vector.

Model training and testing:

- For splitting data into training and testing sets we used NumPy train-test split function.

MODEL

To assess the accuracy of a Book Recommendation AI model, we employ various evaluation metrics and establish a strong training and validation pipeline. Here's how these components were implemented:

Metrics for evaluating performance:

In recommendation systems, precision measures the accuracy of suggested items by comparing correct recommendations to the overall suggestions. A high precision score indicates that an AI system excels at suggesting items tailored to user preferences. On the other hand, recall assesses if a system can successfully identify all relevant items by considering correctly recommended ones in relation to all pertinent options available. A higher recall value suggests that a system is proficient in avoiding omissions of important recommendations. To ensure accurate and comprehensive suggestions, it is vital for recommendation systems to strike a balance between precision and recall metrics.

The Pipeline for Training and Validation:

Data Splitting: The dataset is divided into three subsets - training, validation, and test sets. The purpose of the training set is to train the model while the validation set assists in tuning hyperparameters. Lastly, the test set serves for final evaluation purposes.

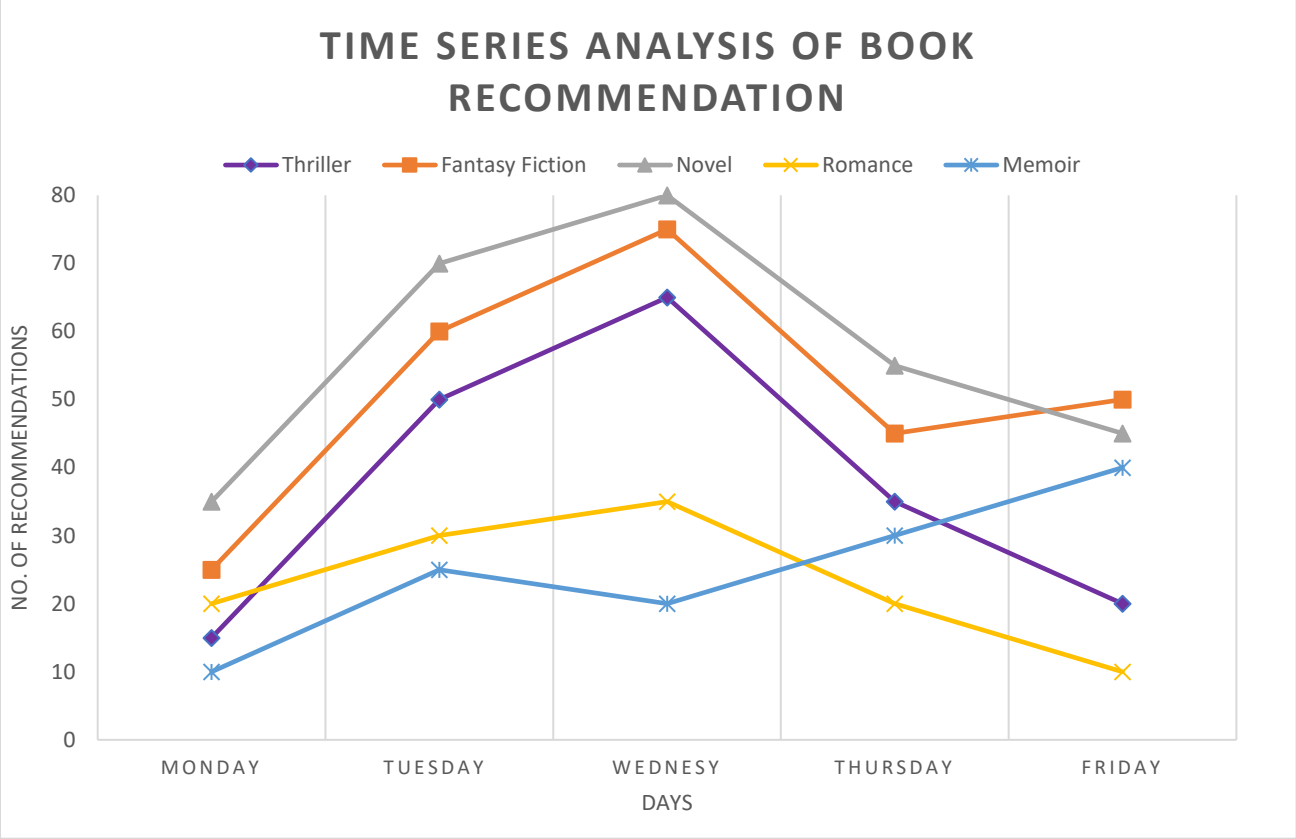
Training of the Model: Employ suitable algorithms (such as collaborative filtering, deep learning) to train the recommendation model with the training dataset. Consistently assess and track the performance of the model on a validation set.

Cross-Validation: Implement cross-validation techniques like k-fold cross-validation to assess model stability and generalization.

We consistently observe the performance of the model throughout its training and validation processes, ensuring to record significant metrics for future reference.

After being satisfied with the model's performance on the validation set, it is assessed on a distinct test set to obtain an impartial estimation of its accuracy.

TIME SERIES ANALYSIS ON DATA



In the above time series analysis, we are analyzing the user's engagements per book genre, we observe that the number of recommendations weekly shared a similar patten. It has demonstrated a consistent growth in user engagement and recommendations generated over a month, it appears to be effective in maintaining a healthy CTR, indicating that users are interested in the recommended Books.

SOLUTION TECHNIQUES

Creating an effective book recommendation system involves using various solution techniques. Here are some commonly used approaches:

- Collaborative Filtering:
  User-Based: Recommends books based on the preferences of similar users.
  Item-Based: Suggests books like those a user has previously liked or borrowed.
- Content-Based Filtering:
  Analyses the content of books and recommends similar ones based on attributes like genre, author, keywords, and themes.
- Deep Learning:
  Utilizes neural networks to extract complex patterns from user and book data for personalized recommendations.
- Natural Language Processing (NLP):
  Analyses book descriptions, reviews, or user-generated content to understand book content and make recommendations.
- Cold Start Solutions:
  Addresses the "cold start" problem by using techniques like popularity-based recommendations for new users or items.
- Association Rule Mining:
  Identifies patterns and associations in user behaviour to suggest books often borrowed together.

NATURAL LANGUAGE PROCESSING, SPEECH RECOGNITION OR SPEECH SYNTHESIS

Natural Language Processing (NLP) pertains to the interaction between computers and human language. It holds a critical significance in our Book recommendation systems, fundamentally transforming how libraries engage with patrons and efficiently handle their extensive collections. Our System leverages speech recognition by means of voice commands as well as text-based instructions for book recommendations.

The system utilizes a technique to transmute the audio signal into a computer-readable format, thus enabling processing. This method incorporates phonemic identification which refers to identifying the smallest sound units within language. Subsequently, it proceeds with matching the spoken information against textual commands for further comprehension and execution.

Through the integration of Natural Language Processing (NLP), speech recognition, or speech synthesis technologies, the Book Recommendation AI achieves enhanced versatility and user-friendliness. This advanced system accommodates a broader range of individuals with varying preferences and accessibility requirements, thus aligning well with the objective of facilitating library communities' improvement while fostering inclusivity in their services.

DEEP LEARNING

For our project of automating the library book recommendation system, we implemented a Convolutional Neural Network, CNN model.

We chose CNN for its ability to detect complex features in data.

Our training data for the CNN model is collected from the library’s historical user interactions and preferences, including book descriptions.

The model will learn to extract meaningful features from the textual descriptions and make personalized recommendations based on user input and preferences.

The model is trained using the TensorFlow backend.

Here is an overview of the CNN layers and the training process:

Input Layer

- Text data is received and represented as a sequence of words or embeddings.

Embedding Layer

- Input text is converted into dense vector representations, work embeddings that capture semantic meaning.

Convolutional Layer

- Several convolutional layers are applied to detect patterns and features within the word embeddings.
- Each convolutional layer uses a set of filters or kernels to scan the text and capture different features at various scales.
- These filters slide over the embedding, performing element-wise multiplications and aggregating information.

Max Pooling Layer

- After each convolutional layer, a max pooling layer is applied to down sample the feature maps.
- Max pooling extracts the most important information from the feature maps, reducing dimensionality.

Flatten Layer

- The flattened feature maps are then converted into a one-dimensional vector, ready for further processing.

Fully Connected Layer

- Takes flattened vector as input and performs traditional network operations.
- It learns complex patterns and associations in the feature representations.

Output Layer

- Produces predictions for book recommendations.
- We will use a soft max activation function for output.

REFERENCES

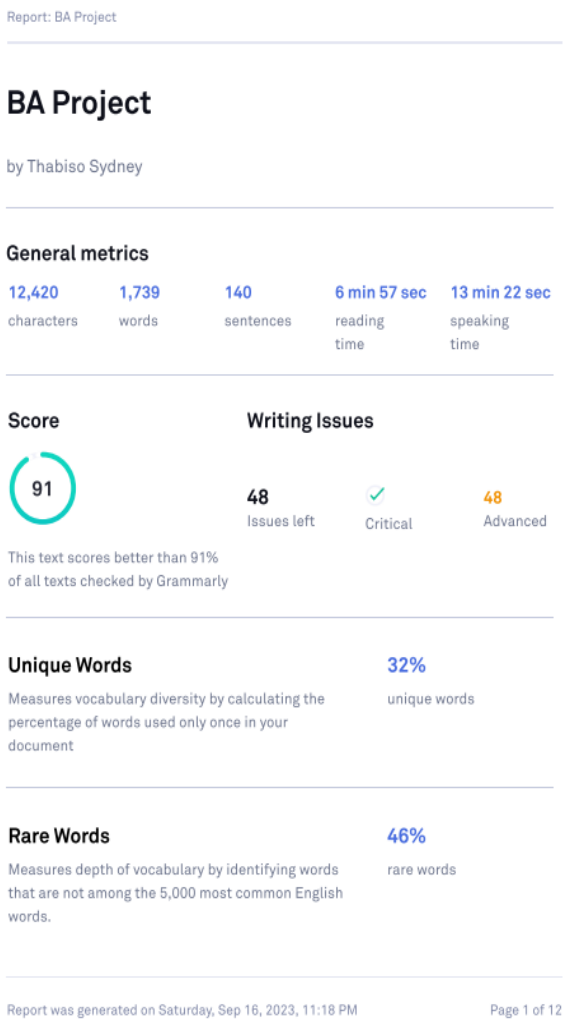
Zaher Talab (2021) 'Machine Learning vs Deep Learning', at https://www.developer.com/guides/machine-learning-vs-deep-learning/ (Accessed:03 August 2023)

(2023) 'Natural Language Processing' at https://www.deeplearning.ai/resources/natural-language-processing/ (Accessed:14 Sep 2023)

'Artificial intelligence, data' at https://en.wikipedia.org/wiki/Artificial_intelligence (Accessed:14 September 2023)

Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies by John D. Kelleher, Brian Mac Namee, and Aoife D'Arcy

GRAMMARLY REPORT



Report: BA Project

BA Project

by Thabiso Sydney

General metrics

12,420	1,739	140	6 min 57 sec	13 min 22 sec
characters	words	sentences	reading time	speaking time

Score

91

This text scores better than 91% of all texts checked by Grammarly

Writing Issues

48 Issues left

Critical

48 Advanced

Unique Words

Measures vocabulary diversity by calculating the percentage of words used only once in your document

32% unique words

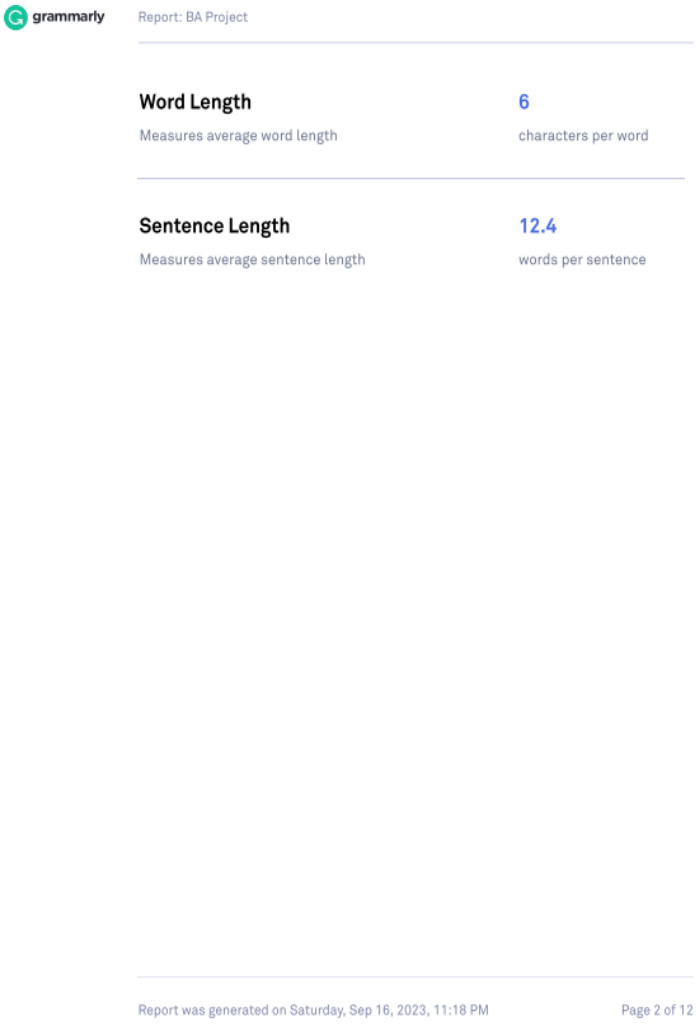
Rare Words

Measures depth of vocabulary by identifying words that are not among the 5,000 most common English words.

46% rare words

Report was generated on Saturday, Sep 16, 2023, 11:18 PM

Page 1 of 12



grammarly

Report: BA Project

Word Length

Measures average word length

6 characters per word

Sentence Length

Measures average sentence length

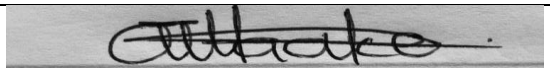
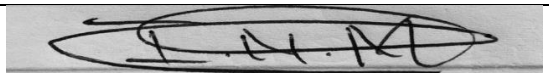
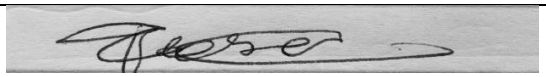
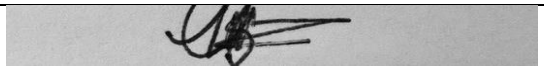
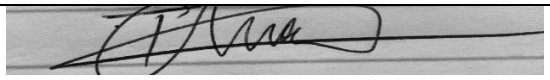
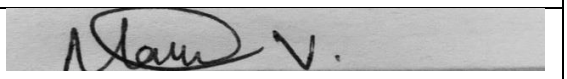
12.4 words per sentence

Report was generated on Saturday, Sep 16, 2023, 11:18 PM

Page 2 of 12

DECLARATION

We declare that this assignment is an original work submitted by the following group members who have all actively contributed. Any other work of a similar nature has been appropriately referenced in this assignment.

Surname and Initials	Student Number	Signature
Maake TS	218590822	
Monyeseala I	221768963	
Lebeko T	221686274	
Mofokeng T	221025138	
Sereme TH	221442375	
Tshavhani NG	221347801	
Mafanywa V	221138846	
Marunga TJ	221766960