



# ONE BOAT

## SOLUTIONS



### Project

## BOOKIFY.COM PROJECT PROPOSAL

### Project Outline

Bookify.com is a growing online bookstore that aims to enhance customer experience by providing personalized book recommendations and gaining insights from user reviews. With a large collection of books and an increasing number of users interacting with the platform, Bookify.com seeks to implement an AI-powered recommendation system and a sentiment analysis tool to improve book discovery and customer satisfaction.



# Project Proposal: Book Recommendation & Sentiment Analysis for Bookify.com

## Project Background

**Bookify.com** is a growing online bookstore that aims to enhance customer experience by providing personalized book recommendations and gaining insights from user reviews. With a large collection of books and an increasing number of users interacting with the platform, Bookify.com seeks to implement an **AI-powered recommendation system** and a **sentiment analysis tool** to improve book discovery and customer satisfaction.

**Note:** Download the Dataset from here ([Click Here](#))

Currently, the platform has three structured datasets:

- **books.csv** – Contains book details such as **ISBN, title, author, publisher, and publication year**.
- **ratings.csv** – Stores user ratings for books, providing insights into book popularity.
- **users.csv** – Includes user demographic information such as **location and age**.

Additionally, Bookify.com has collected customer reviews from various sources, but the data is inconsistent and unstructured across multiple files:

- **book\_review.csv** – Contains **ratings and text reviews**.
- **customer\_review.csv** – Includes detailed review information such as **review title, reviewer name, reviewer rating, description, and verification status**.
- **all\_review.csv** – Stores various user-generated **reviews, summaries, timestamps, and reviewer details**.

To leverage this data effectively, Bookify.com requires a system that can:

1. Recommend books to users based on their interests, past interactions, and ratings.
2. Analyze customer reviews to determine sentiment—positive, neutral, or negative.
3. Build a sentiment prediction model that can classify new customer reviews automatically.
4. Provide actionable insights for business decisions and customer engagement.

## Project Objectives & Requirements

### 1. Book Recommendation System

- Develop a **collaborative filtering** and **content-based filtering** recommendation engine.
- Utilize **user ratings and book attributes** to generate personalized recommendations.
- Implement a **hybrid approach** to improve recommendation accuracy.

### 2. Sentiment Analysis on Customer Reviews

- Preprocess and clean raw review data from multiple sources.
- Apply **Natural Language Processing (NLP)** techniques to classify customer reviews as **positive, neutral, or negative**.
- Provide insights on overall customer sentiment for different books.

### 3. Sentiment Prediction Model

- Train a machine learning model (e.g., Logistic Regression, Random Forest, SVM) or deep learning model (e.g., LSTM, BERT) for sentiment prediction.
- Use labeled review data to build a supervised classification model.
- Evaluate model performance using appropriate metrics (accuracy, F1-score, confusion matrix).
- Deploy the model to predict the sentiment of new customer reviews in real-time.

### 4. Data Cleaning & Preprocessing

- Standardize and merge **inconsistent review datasets**.
- Handle **missing values, duplicate data, and text normalization** for accurate analysis.

## 5. Interactive Dashboard for Business Insights

- Develop a **visualization dashboard** displaying:
  - **Top-rated books** based on user ratings.
  - **Sentiment distribution of customer reviews.**
  - **Predicted sentiment for new reviews.**
  - **Most recommended books** for different user segments.
- Implement filters to analyze data based on **genres, publication years, and user demographics.**

## Expected Outcomes

- A **recommendation system** that improves book discovery for users.
- An **NLP-powered sentiment analysis tool** providing insights into customer reviews.
- A **sentiment prediction model** that classifies new reviews automatically.
- A **clean, structured dataset** for advanced analytics and decision-making.
- A **dashboard with key insights**, enabling Bookify.com to enhance customer engagement.

## Step-by-Step Process for Developing Book Recommendation & Sentiment Analysis System

This project will be completed in a following structured **Software Development Life Cycle (SDLC)** approach. The project is divided into two major components:

- **Book Recommendation System**
- **Sentiment Analysis & Prediction Model**

### Week 1: Planning & Requirement Gathering (SDLC Phase 1 - Requirement Analysis & System Design)

Before beginning development, it is essential to follow the **Software Development Life Cycle (SDLC)** process:

#### Step 1: Understanding the Project Scope

- Read and analyze the problem statement.
- Identify the key objectives:

- Build a **book recommendation system** based on user ratings and book metadata.
- Develop a **sentiment analysis model** to classify customer reviews.

## Step 2: Gathering & Understanding Requirements

- List down all datasets available and their attributes:
  - **Structured Data:** Books, Users, Ratings
  - **Unstructured Data:** Customer Reviews
- Identify key functionalities:
  - **User Inputs:** Ratings, book preferences, past interactions
  - **Outputs:** Recommended books, sentiment scores, analytics dashboard
- Identify **use cases** and system requirements:
  - What kind of recommendations should the system generate?
  - How will sentiment analysis be used for business insights?

## Step 3: Defining System Architecture & Approach

- Choose the **recommendation model**:
  - **Collaborative Filtering** (user-based, item-based)
  - **Content-Based Filtering** (book metadata)
  - **Hybrid Approach** (combination of both)
- Define the **sentiment analysis approach**:
  - Preprocessing raw review data
  - Machine learning vs. deep learning sentiment classification
- Design a **data flow architecture**:
  - How datasets will be processed
  - How user inputs will be used for recommendations

## Deliverables for Week 1:

Project documentation (Problem statement, requirements, use cases, architecture) Initial dataset exploration (understanding missing values, data inconsistencies)

## Week 2: Data Collection, Cleaning & Preparation (SDLC Phase 2 - Data Engineering & Preprocessing)

### Step 4: Data Cleaning & Preprocessing

Before building any model, the dataset must be cleaned:

- **Books Dataset (books.csv)**
  - Remove duplicates and inconsistent data.
  - Ensure ISBN values are unique identifiers for books.
  - Standardize the Year-Of-Publication column.
- **Ratings Dataset (ratings.csv)**
  - Handle missing values in Book-Rating (if any).
  - Normalize ratings (e.g., scale between 1-5 if necessary).
  - Identify users with **insufficient rating history** and filter them.
- **Users Dataset (users.csv)**
  - Remove inconsistent or null values in Location and Age.
  - Normalize Age values (e.g., handle incorrect entries like 0 or 200).
- **Review Datasets (book\_review.csv, customer\_review.csv, all\_reviews.csv)** o Extract relevant fields (ratings, review text, reviewer details).
  - Merge datasets into a single standardized format.
  - Convert text to lowercase, remove special characters, and tokenize for NLP tasks.
- **Things to Learn & Research:**
- Data preprocessing techniques (handling missing values, normalization, deduplication) Data merging strategies (handling multiple sources of customer reviews)

### Deliverables for Week 2:

Cleaned and merged datasets for books, ratings, users, and reviews

## **Week 3: Building the Book Recommendation System (SDLC Phase 3 - Model Development)**

### **Step 5: Develop Recommendation System**

#### **1. Exploratory Data Analysis (EDA)**

- Analyze book popularity based on ratings.
- Identify users with similar reading patterns.

#### **2. Build Recommendation Models**

- **Collaborative Filtering:**
  - Identify similar users (User-User Filtering).
  - Identify similar books (Item-Item Filtering).
- **Content-Based Filtering:**
  - Extract features from book metadata (title, author, year, publisher).
- **Hybrid Approach:**
  - Combine collaborative filtering and content-based filtering for better recommendations.

#### **3. Evaluation & Optimization**

- Use metrics like **RMSE, MAE, and precision-recall** to evaluate the recommendation models.

#### **Things to Learn & Research:**

Machine Learning algorithms for recommendation (SVD, KNN, Matrix Factorization) Evaluation techniques for recommendation systems.

#### **Deliverables for Week 3:**

Trained recommendation models (collaborative, content-based, hybrid)

## Week 4: Sentiment Analysis - Data Processing & Text Cleaning

### Step 6: Preparing Data for Sentiment Analysis

- **Text Preprocessing**
  - Tokenization, stopwords removal, stemming, and lemmatization.
  - Handle emojis, special characters, and misspellings.
- **Sentiment Labeling**
  - Use existing **ratings** to label text reviews:
    - Ratings **4-5** → Positive
    - Ratings **3** → Neutral
    - Ratings **1-2** → Negative
- Manually review some samples to validate correctness.

### Things to Learn & Research:

- NLP preprocessing techniques
- Data augmentation for text classification

### Deliverables for Week 4:

Cleaned and labeled text review dataset



## Week 5: Building & Training the Sentiment Prediction Model

### Step 7: Model Development

- **Choose ML/DL Models for Sentiment Analysis**
  - Traditional ML models: **Logistic Regression, Random Forest, SVM**
  - Deep Learning models: **LSTM, BERT**
- **Train & Evaluate the Model**
  - Split data into **training (80%)** and **testing (20%)**.
  - Use performance metrics like **accuracy, F1-score, confusion matrix**.

### Things to Learn & Research:

- Supervised learning for text classification
- Transfer learning using pre-trained models (BERT, GPT)

### Deliverables for Week 5:

Trained sentiment analysis model

## Week 6: Deploying the Models & Developing Dashboard

### Step 8: Deployment & API Development

- **Deploy Recommendation System**
  - Create an API endpoint to return book recommendations.
- **Deploy Sentiment Analysis Model**
  - Develop an API to predict sentiment for new reviews.
- **Develop an Interactive Dashboard**
  - Visualize top-rated books, customer sentiment trends, and recommendations.

### Things to Learn & Research:

- API development (Flask, FastAPI, or Django)
- Data visualization techniques (Matplotlib, Seaborn, Plotly)

### **Deliverables for Week 6:**

Deployed models & interactive dashboard

### **Week 7: Testing, Finalizing & Documentation**

#### **Step 9: System Testing & Optimization**

- Conduct **unit testing & integration testing** to ensure models work correctly.
- Optimize recommendation & sentiment models for performance.

#### **Step 10: Project Documentation & Final Report**

- Document all processes, findings, and results.
- Prepare a presentation for project demonstration.

### **Deliverables for Week 7:**

- Fully functional recommendation & sentiment analysis system
- Complete project documentation & final presentation

### **Final Outcome**

By following this structured 7-week plan, you will develop a **real-world AI-powered recommendation and sentiment analysis system**, gaining hands-on experience in **machine learning, NLP, data engineering, and model deployment**.