

Capstone Project Proposal



Nourah Alhassan

Business Goals

Project Overview and Goal

What is the industry problem you are trying to solve? Why use ML/AI in solving this task? Be as specific as you can when describing how ML/AI can provide value. For example, if you're labeling images, how will this help the business?

The publishing world is thriving with hundreds and thousands of books and literary material being published every year and since it's more accessible than ever, Improving the user experience for readers by personalizing and optimizing the recommendation model will help book sites sell more books and increase the rate of return customers and customers satisfaction. form my personal experience using multiple book-related sites (Goodreads, audible, Scribd..Etc.) Often when I create an account I'm asked to select 3-5 favorite genres, and then I'm provided an initial list based on that. I have had some of my accounts for almost 10 years and my taste naturally has changed a lot during that time and yet you don't see the recommendation system adapting to the change and offering books that might interest me, instead recommending books that are "similar"; meaning share one or two genera but nothing deeper or more meaningful than that.

Using AI/ML model is going to personalize the experience of looking for the next read by analyzing a large amount of data, the history of past reads combined with the user rating and tagging system. This would switch the current recommendation system that is based on "what other readers also enjoyed" and "More like *book name*" to a more helpful optimized recommendation model, which will increase customer satisfaction and improve repeat customer rates.

<p>Business Case</p> <p>Why is this an important problem to solve? Make a case for building this product in terms of its impact on recurring revenue, market share, customer happiness and/or other drivers of business success.</p>	<p>Having a system that will be able to look past the bestseller list and the viral books and have the ability to look deeply for books worth recommending to a specific reader based on personal taste, and historical data won't only improve the user experience but will also increase customer satisfaction and raise the return customer rates.</p>
<p>Application of ML/AI</p> <p>What precise task will you use ML/AI to accomplish? What business outcome or objective will you achieve?</p>	<p>The model will be able to suggest relevant books, using both collaborative filtering and content-based methods are suitable for this kind of recommendation system, it considers past interaction to make prediction based on:</p> <ul style="list-style-type: none"> • Rating is given by a user to a book • Time spent interacting with the content page of a book • Books clicked when suggested or not • Books tagged and shelved by the user • Information of the item, characteristics such as (Tropes, Genera, Targeted audience, Story elements,...) <p>And when more interaction happens with books on the site the more accurate and successful the prediction will be.</p>

Success Metrics

Success Metrics

What business metrics will you apply to determine the success of your product? Good metrics are clearly defined and easily measurable. Specify how you will establish a baseline value to provide a point of comparison.

- Increase customer satisfaction and customer return rates –Feedback surveys and ratings, the number of times of visiting the website during a set period of time.
- Increase the number of new and active users
- Increase the accuracy and efficiency of the model by 100% within six months

Data

Data Acquisition Where will you source your data from? What is the cost to acquire these data? Are there any personally identifying information (PII) or data sensitivity issues you will need to overcome? Will data become available on an ongoing basis, or will you acquire a large batch of data that will need to be refreshed?	<p>The data will be connected constantly through the life of the model, the data will be acquired from current active users of the site or from open sources.</p> <p>The cost will be determined by the compute power needed to run the model.</p> <p>The model will not be acquiring any personal or identifying information of the users so there will be no issues in that regards.</p>
Data Source Consider the size and source of your data; what biases are built into the data and how might the data be improved?	<p>The source of the data will come in-house from the old recommendation model and additional information from the website content and users. The data will improve by time and training.</p>
Choice of Data Labels What labels did you decide to add to your data? And why did you decide on these labels versus any other option?	<p>Initially the model will have simple labeling system then will expand to more complicated system with training and data availability, the types of label:</p> <ul style="list-style-type: none">-Rating-Key words in review-Time of interaction-Books Tag-Genres <p>The purpose of this model it to suggest possible interesting next reads for the user using the given data.</p>

Model

Model Building How will you resource building the model that you need? Will you outsource model training and/or hosting to an external platform, or will you build the model using an in-house team, and why?	Inhouse training is more expensive and slower and with the lack of experience building AI model its more fit to outsource. Using tools like Google AutoML will save time and help experiment and build experimental models to help the team learn and gain experience.
Evaluating Results Which model performance metrics are appropriate to measure the success of your model? What level of performance is required?	Precision and recall are the metrics that need to be evaluated and monitored regularly for as long as the model is running and more data is trained.

Design

What does your minimum viable product look like? Include sketches of your product.

GoodBooks

Welcome back
Looking for your next read?

Based on your last reads



High fantasy with *dragons* and *quests*



Use Cases

What persona are you designing for? Can you describe the major epic-level use cases your product addresses? How will users access this product?

We target readers who use online book sites to search for relevant books. to ensure customer satisfaction the model must provide relevant book suggestions to the user to increase the site sales and return rates.

Roll-out

How will this be adopted? What does the go-to-market plan look like?

We will target readers so the product will be adopted by having different organizations try out the new model which will be made available for different sites e.g Book Review sites, Online libraries, Online Book stores.

Feedback will be encouraged from all users trying out the product.

Post-MVP-Deployment

Designing for Longevity How might you improve your product in the long-term? How might real-world data be different from the training data? How will your product learn from new data? How might you employ A/B testing to improve your product?	<p>To ensure the model stays relevant to its users and provides an excellent user experience we must plan for longevity. Improving the model in the long-term includes:</p> <ul style="list-style-type: none">• Collecting Feedback/new information from users and utilizing the feedback to release new versions of the model• Constant repeat of prototyping testing, and iterating process with new information• Updating the model constantly with new data in order to prevent the model from being outdated when the predictive power of the ML model decreases over time as strengths or user state changes
Monitor Bias How do you plan to monitor or mitigate unwanted bias in your model?	<p>Model bias is expected, and it could be caused by:</p> <ul style="list-style-type: none">• Unbalanced dataset• Dirty data• Special case data• Irrelevant data <p>All could lead to biased or low-performing models.</p> <p>To manage and avoid major faults in our model we must do testing and ask for feedback from users to help us improve the model as its output is relative and varies from user to user.</p>