Capstone Project Proposal

<Sung Jun Eun>

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?	In music streaming application 'Tidal', there's many different kinds of playlists. However, I always wanted playlists of typical mood. For example, playlist of songs about 'loneliness' or songs about 'passionate love'. So, in this project proposal, I would think myself as an AI product manager of company 'Tidal'. To generate the accurate mood playlist, I would use AI technology. AI technology will train based on the lyrics of songs and it will label each song to certain moods.
Business Case 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.	Diverse and accurate playlist are very important factor when users choose their music streaming platform. Specifically, all of us will have an experience that certain days, we are caught up in a certain mood and just feel it with right music. My mood playlists would be perfect playlists for these days. In conclusion, my playlists would increase the customer happiness and also attract new users, too. Then revenue and market share will naturally increase.
Application of ML/AI What precise task will you use ML/AI to accomplish? What business outcome or objective	I want to get the representative semantic mean of each song based on NLP analysis of the song's lyrics. For example, after training the model, 'Won't go home without you' by Maroon5 and 'If I can't have you' Shawn Mendes could be classified into one semantic mean 'foolishness'. This classification based on semantic

will you achieve?	mean will not matter on genre, age of song, bpm but solely on the semantic mean of the song based on the lytics.

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.	Success metrics would be user time playing this mood playlist. Another long term success metrics would be overall number of users difference after implementing this new playlist.

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	In website called 'lyrics.com' (https://www.lyrics.com/about.php), I could get the hundred of thousand lyrics with the track information. I guess it isn't sensitive data. I guess data should be update throughout future, because new music comes out every day and we should be able to add these songs to playlists.

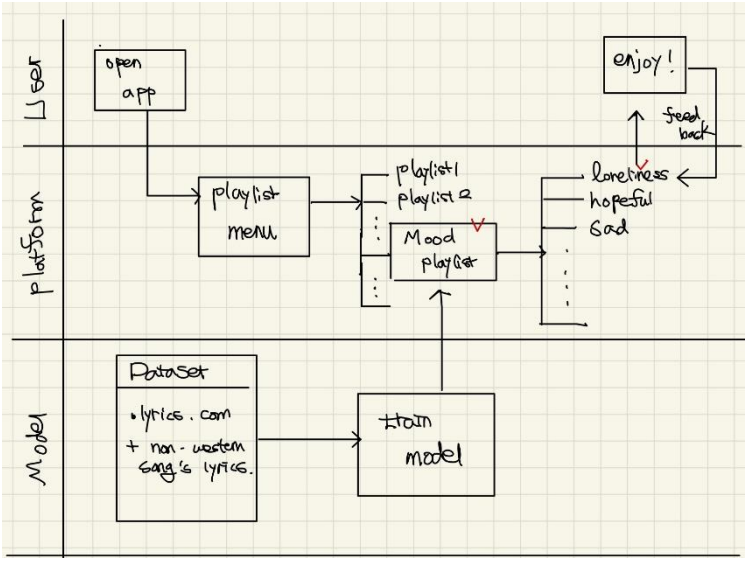
available on an ongoing basis, or will you acquire a large batch of data that will need to be refreshed?	
Data Source Consider the size and source of your data; what biases are built into the data and how might the data be improved?	As most of the data will be lyrics of western songs, playlists would be trained based on western songs. Then If we label not-western songs, model could make some errors. As, we are living in a globalized world, I should get lyric data from not-western songs and add to dataset.
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?	Label would be kind of emotions, 'loneliness', 'sad', 'love', 'passionate love', 'hopeful', etc. I chose these labels based on my experiences, so for real implementation, I will do some quick customer survey about what emotions do they hear about top 100 songs. Then I could choose top 16 or 32 emotions as a label.

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?	I would make model using in-house team. As NLP open sources are well-made and available, I think I don't need to outsource the model training. We could just modify the opensource model and implement it.
Evaluating Results Which model performance metrics are appropriate to measure the success of your model? What level of	Because I am making a recommendation base product, it's important to make precise recommendation. So, precision would be appropriate to measure the success.

performance is required?	

Minimum Viable Product (MVP)

Design What does your minimum viable product look like? Include sketches of your product.	As I will implement my product inside an existing music streaming platform, I will add the functional flow chart of my product. 
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?	I'm designing for music listeners who want playlists for typical mood. The major case would be; you got scold from your boss and you got in a traffic jam. After you arrived at your home, you can release your stress with the playlist of songs about anger. My product would be access at the playlist menu of the application 'Tidal'.
Roll-out How will this be adopted? What does the go-to-market plan look like?	As it will be implemented to existing music application, we don't need any go-to-market plan.

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?	As listeners add or delete tracks in the mood playlists, it would be act as a feedback to the model. Also, I would increase the number of labels which could cover more emotions. Lyrics of music songs don't have any different in real-word and training data. New data would be added to labels and increase the overall accuracy. After adding new label, I would use 80% of traffic to the current model and use 20% to updated model and check the performance metrics.
Monitor Bias How do you plan to monitor or mitigate unwanted bias in your model?	User's feedback would be a key to monitor and eliminate the unwanted bias in model.