RAVELRY PROJECT RECOMMENDER

BUSINESS UNDERSTANDING

I enjoy knitting and crocheting and, along with millions of other users, utilize a website called Ravelry, a database-based website, to find patterns and track progress on my projects

Crafting, like Netflix or Spotify, constitutes a form of entertainment and merits a recommender system so that users who have generated input by purchasing or downloading, and completing certain patterns, can benefit from feedback, suggesting which patterns best suit their making habits.

The website has a pattern recommender, but it seems to be based more on clicks than actual projects completed. I think this could be improved upon and better tailored to users specific making interests.

DATA UNDERSTANDING

The website does not have an app and so makes its data available through an API so that apps may be developed by third parties.

I will use this API collect data on the users and their projects.

DATA PREPARATION

The data will be stored in a CSV. All variable types will strings. I imagine there will be some preprocessing to eliminate users who do not track their projects. Similarly, the discrepancies between projects based on a published pattern, more than one published pattern, or no published pattern will have to be addressed.

Ravelry has 9 million users, 1 million of whom are monthly users. The average number of projects per user being around 30, I think there could be 30 million rows of data at a minimum.

MODELLING

I plan to develop a recommender system, recommending published patterns to users based on other projects they have made.

EVALUATION

Precision@k and recall@k are common evaluation metrics for recommendation systems. The minimum viable product is a recommender system based on published patterns. The level up product would be to expand the recommendations to other users projects, not only published patterns.

DEPLOYMENT

I have no plans for deployment at this point. This recommendation system could be incorporated into an app built on the API.

TOOLS / METHODOLOGIES

I plan to use PySpark to build a recommender system. Data will be stored and analysis performed on my machine.