Domain

I will collect data from kaggle's dataset Mobile App Statistics (Apple iOS app store). I would like to build up a model to predict users' rating on new version of an app based on features including its size, price, number of ratings, etc. This model could help developer identify whether users will like this new version or not and what improvement could be made. Because there are many level of ratings, so I decide to stratified it into three levels: excellent(3.5 – 5), fair(2 – 3.5), poor(0-2).

Data

Variable	Type	Description
size_millionbytes	float	Storage the app occupies
price	float	The amount of charges if the user wants to download the app
Rating_count_tot	int	Total number of ratings the app has.
Rating_count_ver	int	Total number of ratings the current version of app has.
User_rating	float	Average user rating value(for all version)
User_rating_ver	float	Average user rating value for current version
Sup_devices.num	int	Number of supporting devices

ipadSc_urls.num	int	Number of screenshots showed for display
Lang.num	int	Number of supported languages
Prime_genre	str	Primary Genre

Known Unknowns
- I might change my predicted variable when I figure out something more reasonable.
- I might change my dataset when I figure out something more reasonable.