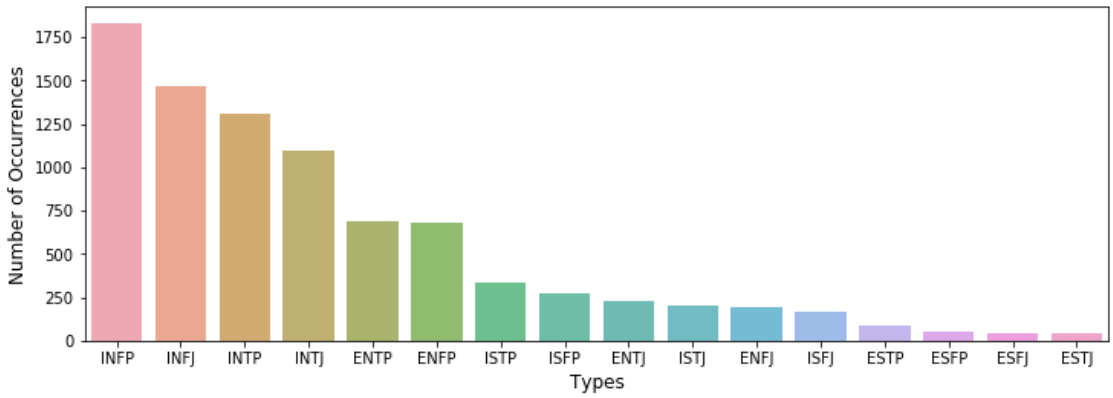
After training the classes, we take a look at the dataset, which looks very unbalanced.



The performance for the Extra Trees Classifier with SVD

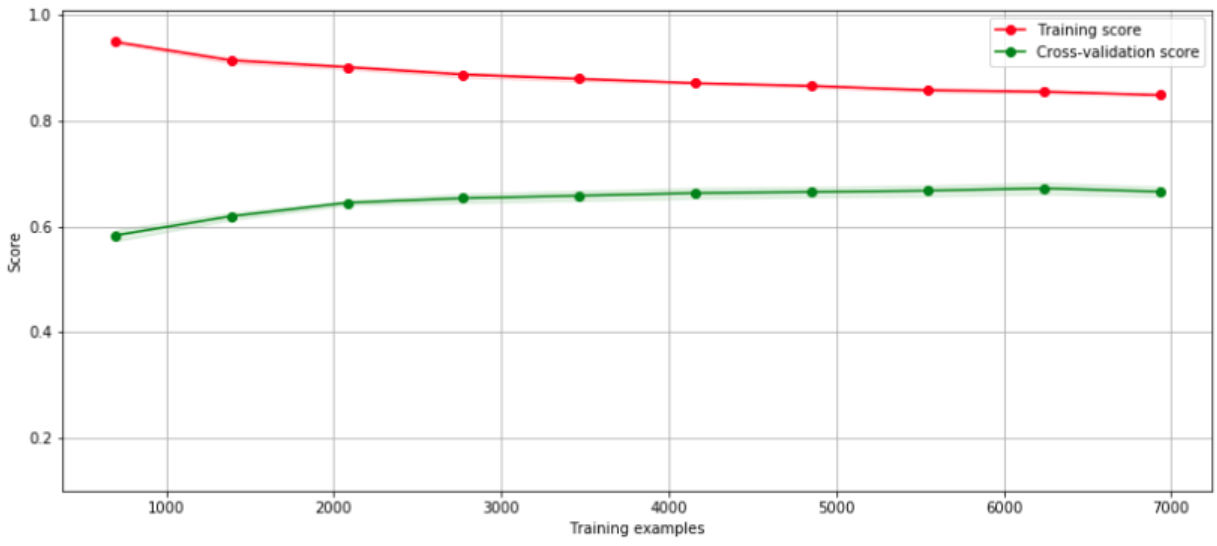
CV Accuracy: 0.2930 (+/- 0.0148)
CV F1: 0.2930 (+/- 0.0148)
CV Logloss: 2.1430 (+/- 0.0247)

Naive Bayes:

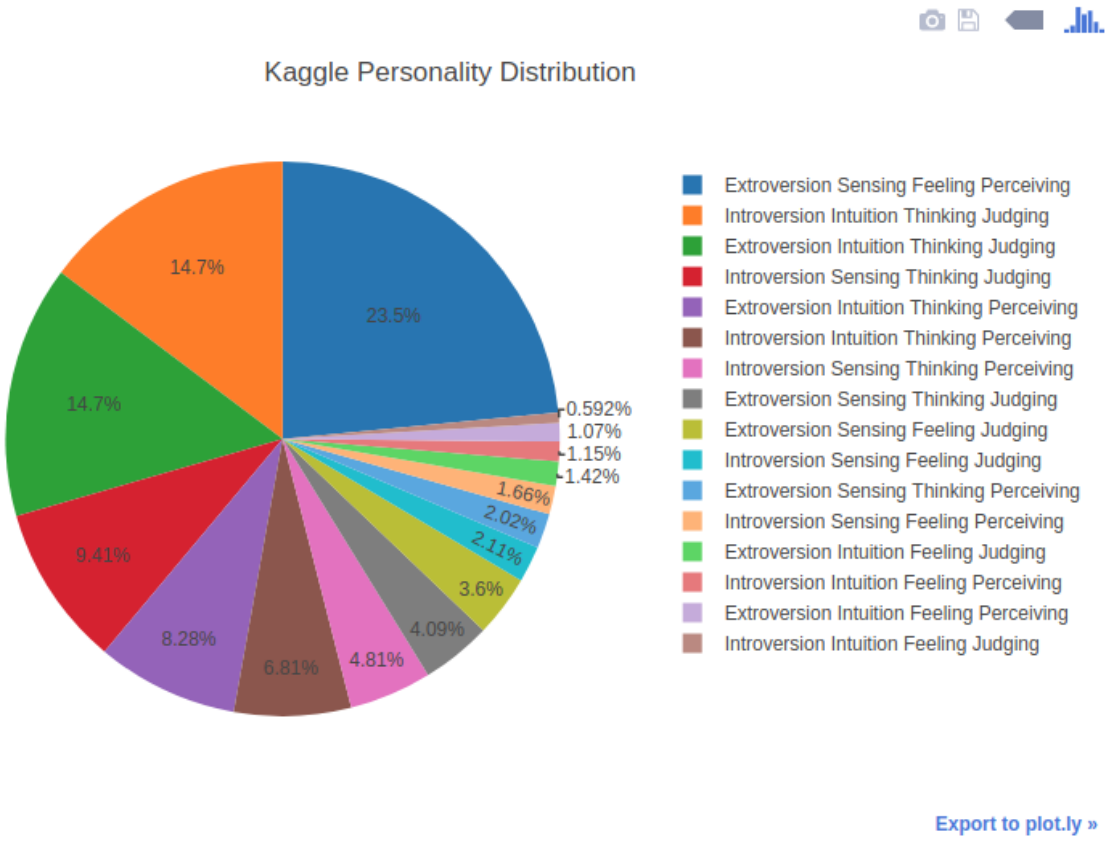
CV Accuracy: 0.5589 (+/- 0.0097)
CV F1: 0.5589 (+/- 0.0097)
CV Logloss: 6.2748 (+/- 0.2918)
Logistic Regression:

CV Accuracy: 0.6659 (+/- 0.0090)
CV F1: 0.6659 (+/- 0.0090)
CV Logloss: 1.2501 (+/- 0.0238)

It looks like that with more data, the model gets better and it’s not overflowing.



Applying our obersations to the Kaggle users comments, we can see what the most common type of users on kaggle are.



Conclusion:

We can see from the piechart that the most common users personality is ESFP (Extroversion Sensing Feeling Perceiving), but we are getting this onclusion based on users comments: It is reasonable to think that users who participate in more comment writing are more extroverted. My sample data mainly comes from Kaggle users who write comments, so the conclusion can’t be applied to all Kaggle users, but only to those who write comments.
More accuracy models could be obtained with more data.