Gabrielle Beinars
Drug Reviews

Questions from the Audience

1. **How well do positive predictions compare with negative predictions?**
   A higher accuracy was observed for positive predictions than that for negative predictions across all models created for this project. This can be concluded through the results of the classification report.

2. **What makes a condition have more drug reviews than another?**
   There could be a couple explanations for this because the dataset is based on drug reviews, there could be more common conditions, or there could also be more drug options for a certain condition.

3. **Which model performed the best, and why?**
   The logistic regression model performed better than both of the Naïve Bayes models, which was interesting because there are more assumptions for logistic regression. This may just be because more assumptions were met.

4. **Where there any variables from the original data set that were not used, but could have been?**
   The short answer for this is yes – for example, the date could be used to look at reviews over time. With the increase in modern technology over the years, and this data being obtained from **online** reviews, we may hypothesize that there is an increase in online reviews over time. Additionally, the number of users that found each review useful could be used in modeling.

5. **Why remove stop words?**
   By removing stop words, the words that do not provide any value are removed form the text data, thus decreasing our word count and only include words that may have an impact on the results. For example, some stop words that were removed are “her”, “with” and “that”.

6. **What are some of the most common words used in the drug reviews?**

   The most common top ten words were the number, 10, “acne”, “ago”, “almost”, “also”, “anxiety”, “away”, “back”, “bad”, and “better”.

7. **What makes this project useful?**

   By evaluating drug reviews, a better understanding can be gained about what drugs may be preferred for treatment for a given condition.

8. **What types of modeling were explored?**

   The models explored for this project were logistic regression, and two commonly used Naïve Bayes classifiers for text analysis, Bernoulli and multinomial.

9. **What improvements could be made for this project?**

   Some different types of modeling that may also be used to gain some additional insight could by clustering, or additional types of classification, such as predicting a condition.

10. **What is a probabilistic classifier?**

    As mentioned, naïve bayes classifiers are considered probabilistic classifiers. This is because they involve predicting a probability distribution over a set of classes using a given input.