

# FINAL-Evaluation Instructions

**Deadline: 11:59 PM, 3<sup>rd</sup> February, 2022**

For your final evaluation, you will be creating an entire NLP project utilizing everything you have learned during the span of WoC.

## Datasets

You can choose any one of the following 3 datasets for your final project:

1. [Tweets about Products](#)
2. [Play Store app reviews](#)
3. [Social Tweets](#)

## Instructions

1. **Gain Insight from the data:** Download and explore the chosen dataset using various methods learned to gain insight on the data and the problem.
2. **Clean the data:** Fill missing values, use label encoding, optimize memory if necessary.
3. **Visualize the data:** Use Matplotlib/Seaborn to visualize the data (word clouds, histograms etc.) and gain more insight on what methods you will use to get the best results.
4. **Clean the text:** Clean the text to remove hyperlinks, email addresses, multiple spaces, unnecessary symbols and numbers etc if necessary.
5. **Preprocess the text:** Use tokenization, stemming/lemmatization, remove stopwords to prepare the text.
6. **Feature engineering:** Use Bag or Words/n-grams, TF-IDF, feature hashing or any other technique to convert the text to features for passing them as input to the learning algorithm.
7. **Choosing the model:** Choose the model you would want to use to train the data on.
8. **Train and Test the model:** Train the model using prepared features to get predictions.
9. **Evaluate the model:** Evaluate the accuracy of the model, tune hyperparameters, try other models if you'd like to get the best results.

**Note:** You may add/remove some particulars on the above given instructions based on your data set, you may also change the order of instructions if you want to, For example, if you want to split data before preprocessing or visualize the data after preprocessing, you are free to do so.

## **Submission instructions**

1. Save your notebook as “Final\_Project” and commit to your GitHub repository.
2. Fill up the submission [form](#).