NLP Project - Prototyping

Time to build! By next week you should have a (minimally) functioning system that will perform basic tasks using the model you will have generated from your data. This week we are not overly concerned about how well it works, just that it functions and you are able to get some initial baseline measurements. As you will likely not have a huge amount of data, cross-validation is the best option to generate an estimate of how your system is performing.

I will again reiterate you are free to use existing libraries to build your model, but the model should be built using only your collected data and perhaps a lexicon to augment your vocabulary.

Each group will have a single submission:
> A basic implementation of your system. The prototype should have been run through at least 5-fold cross-validation (10-fold would be better, if possible) to test the resulting model before submission. The prototype will consist of:
> - The system, compiled as a self-contained executable.
>   - Any dependencies, libraries, etc, must be included in the executable
>   - The user (me) should not have to build, download, or configure anything.
>   - You may assume I have the necessary language SDKs and interpreters installed for interpreted languages (Java 18, Python 3, etc),
>   - The model generated by your training data, saved as a separate binary file that can be loaded by your system
>   - In other words, to use your system, I won't have to wait for the model to be built.
> - The set of training data that was used to build the model.
> - A set of test data
> - Clear instructions on how to use the system. This will include:
>   - How the user will feed the test data into the system
>   - How the output will be reported to the user
>   - How to interpret the output
> - A report detailing the output of the cross-validation tests, including the Precision, Recall, and F-measure of all folds and any actions taken as a result of these tests. Was any overfitting identified? Were any features under-represented? How close is the current performance to your vision statement's desired performance?