**Project Title:** Guess Who: The Office Edition

**Team:** Morris LaGrand

**Objectives:**

* To leverage the skills, techniques and libraries learned throughout the semester to create an NLP classifier that will predict which character from The Office said the given monologues.
* Create a simple UI that allows users to interface with the data and train a series of classifiers using various methods

**Usefulness:**

This application is a fun, novel project for individuals who are interested in data science but have little knowledge of the domain. Users will have the ability to train multiple machine learning models and compare the performance using only a graphical interface. To my knowledge, there aren’t any websites or services that perform the actions that my application will perform. It’s a great way to increase interest in data science while working with a fun dataset.

**Dataset:** The Office Quotes Dataset

* Description: The Office Quotes Dataset is a Kaggle dataset that features two separate csv files. One csv contains monologues and the characters that said them. The other csv contains quotes from a character responding to another character. I will leverage the monologues csv for this project.
* Data Summary:
  + Records: 1728
  + Labels: Yes
  + Text field only
  + Cleaning: Yes, will do standard pre-processing (stop removal, punctuation, etc.) and refine cleaning method as needed

**Functions:**

* Pipeline to pre-process data and train an NLP classifier
* Train and validate multiple classifiers via UI
* Explore and filter data via UI
* “Guess Who” game where the user can compete against the trained model to correctly guess who said the quote (time permitting)

**Communication:**

N/A: I will complete this independently.

**Personal Contribution Statement:**I took Applied Database Technologies under Dr. Scrivner and tackled a project with a similar scope of effort. For that project, I felt it was easiest to create Milestones to guide me through the project efforts. Below are the tentative milestones to keep me on track.

**Milestones:**

* Week 7: Project Description
* Week 9: Data Cleaning
* Week 11: Web App Mock and Analytic Development
* Week 13: UI Development
* Week 14: Video Presentation
* Week 15: Full Demo