**MGMT59000 Analyzing Unstructured Data**

**AUD Boilers**

**Project Milestone 1**

*Gagan Pahuja, Mrugshitr Rajiv Bansal, Puja Gupta, Ryan Egbert, Zainab Aljaroudi*

**Business Scenario**

We are trying to improve website traction for the ‘Resume’ page on Craigslist which will lead to higher visibility and more business for the website.

**Problem Statement**

Currently, the ‘Resumes’ page on Craigslist does not offer any filtering service for employees wishing to target a specific job category and employers looking to fill posted positions. However, the ‘Jobs’ page does offer this filtering mechanism. This results in a much lower number of resumes relative to the number of jobs posted in a particular area.

Text, application

Description automatically generated

**Proposed Improvement**

Our proposed improvement is a two-part process:

1- In order to provide parallel systems, we will create a classification model that will identify which job category best fits each resume. The model will be trained on job descriptions with the existing job categories as the target variable.

2- We will also provide employers with the top 3 resumes that fit their position. This will be accomplished by using classic similarity measures to compare a specific job description with resumes under the same job category.

**Data Requirement**

The necessary data includes job and resume descriptions along with the job categories. The data has been scraped directly from Craigslist using Python using BeautifulSoup.

Attached is a sample of the scraped data from Craigslist.

Column 1: Category (resumes / jobs wanted is the current category for resumes, the rest are specific job categories)

Column 2: Post description

Column 3: Post id