

DATA 607 - Final Project Proposal

Author:

Kory Martin

Topic:

Data Science Job Posting Classifier System

Overview:

For my final project, I will be extending the work from Project 3 to build a supervised learning based classifier model that will take a job posting and classify it as being a Data Science job posting or not. Furthermore, I will be refactoring my code related to how I generated the corpus of text to take advantage of some of the text mining tools learned in Week 10.

Motivation:

I have a personal interest in the commercial applications of machine learning and predictive analytics, with a particular focus on how machine learning can be used to improve the personalization features of digital products and services in order to increase business goals (e.g. drive product adoption, increase marketing ROI, increase sales, increase customer retention, etc.). Classification and text mining are very important components of this process, so I want to expand my familiarity and competence in both of these areas.

Primary Data Sources:

The two primary types of data that I will be using are:

- Relational - I will utilize the corpus of Data Science keywords that was created in Project 3, which is currently held in a PostgreSQL database
- Web Scraping - Additionally I will be using web scraping to pull additional job postings that will be used to expand my training and test datasets