



Krannert School of Management

CRAIGSLIST RESUME MATCHING

MGMT 59000
AUD Boilers

December 8, 2021



About the Client

Craigslist is an American classified advertisements website with sections devoted to jobs, housing, for sale, items wanted, services, community, gigs, resumes, and discussion forums. It is one of the largest peer-to-peer user-generated advertisements website serving 570 cities in 70 countries.

The website is advertiser oriented. Advertisers can easily publish advertisements in a flexible and unstructured format, typically shown as text description and/or uploaded images. This unstructured format leads to the following problems –

1. High cost incurred for ad-viewers to extract information
2. High cost for platform managers to maintain and organize the website

Segments Targeted

We targeted the 'Resumes' and 'Jobs' page on Craigslist for our project. Our reason for selecting these subsections was two-fold:

1. The Jobs section on Craigslist generates revenue through a handful of revenue streams. Job postings is one of the top-most revenue generating sections of the website. Craigslist charges a \$25 fee to post a job listing in six major U.S. cities. For example, it charges \$75 for a job listing in the San Francisco area.
2. The AIM Group projected 2020 revenue of \$565 million for Craigslist. That's down from more than \$1 billion two years ago. While some of the drop in 2020 may have been caused by Covid-19, it began well before the pandemic started. Revenue declined in 2020 in almost every ad category except "gigs." **Job ads took the biggest hit, down 32%; auto ads were down 22%**



Problem Identified

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. Recruitment agencies must match hundreds of candidates with hundreds of job listings a month. It is a very time-consuming process to manually sift through such a pile of documents.

Business Analysis

Current Scenario:

- Aside from the sophisticated job portals, employers and employees use Craigslist as a medium to find each other.
- The jobs category on the website is properly filtered and sub-categorized while the resume category requires improvement.

Our proposed improvement is a three-part process:

1. To provide parallel systems in case of the Craigslist 'Jobs' and 'Resume' section, we created a classification model that will identify which job category best fits each resume. The model will be trained on job ads with the existing job categories as the target variable.
2. To provide employers with the top 3 resumes that fit their vacancy. This will be accomplished by using classic similarity measures to compare a specific job description with resumes under the same job category.
3. To provide job seekers with the top 3 job openings that fit their resume. This will also be accomplished using the similarity measures described above.

Data Analysis

Our Data Analysis and Pre-processing process consisted of the following six steps:

Step 1: Web Scraping

There exists a range of data that has been scraped from Craigslist in the past, but none of them are related to resume and job postings on the site. Due to the lack of data in this category, we needed to manually scrape the web to gather this data.

This method was performed by using two Python libraries: urllib and BeautifulSoup. Urllib was used specifically for gathering the static HTML from the site via HTTP requests and responses. After the HTML had been received, BeautifulSoup was implemented to parse through the documents and gather header and body data from each of the posts.

Step 2: Dimension Reduction

The raw data scraped from Craigslist needed to be groomed so it would work better with classification models. In order to accomplish this goal, the raw text was passed through a dimension reduction process: tokenization, lemmatization, and removal of stop words and punctuation.

- Tokenization: dividing data up into smaller chunks (i.e. groups of words)

- Lemmatization: reducing a token down to its base meaning (i.e. running --> run, happier --> happy)
- Removal of stop words: stop words are words that are so common, they appear in nearly every document (i.e. where, and, this, but, me, etc.)

By performing lemmatization and removing stop words, the data become much less cluttered and can be processed more accurately.

Step 3: Words frequency Normalisation using TF-IDF Vectorization

Term Frequency (TF) - It is a measure of the frequency of a word (w) in a document (d). TF is defined as the ratio of a word's occurrence in a document to the total number of words in a document.

$$TF(w, d) = \frac{\text{occurrences of } w \text{ in document } d}{\text{total number of words in document } d}$$

Inverse Document Frequency (IDF)

It is the measure of the importance of a word. Term frequency (TF) does not consider the importance of words. Some words such as 'of', 'and', etc. can be most frequently present but are of little significance. IDF provides weightage to each word based on its frequency in the corpus D.

$$IDF(w, D) = \ln\left(\frac{\text{Total number of documents } (N) \text{ in corpus } D}{\text{number of documents containing } w}\right)$$

Term Frequency — Inverse Document Frequency (TFIDF)

It is the product of TF and IDF. TFIDF gives more weightage to the word that is rare in the corpus (all the documents). TFIDF provides more importance to the word that is more frequent in the document.

$$TFIDF(w, d, D) = TF(w, d) * IDF(w, D)$$

Step 4: Classification of Resumes into Categories using different Models

We examined five different models to solve the classification problem of categorizing the resumes into the existing job classes. The following are the models we explored along with its accuracy rates:

1. **Naïve Bayes** with accuracy score of 57.98%
2. **Support Vector Machine (SVM)** – Highest accuracy of 78.93%
3. **Decision Tree** with accuracy score of 72.5%
4. **Random Forest** with accuracy score of 71.9%
5. **Artificial Neural Network (ANN)** with accuracy score of 64.76%

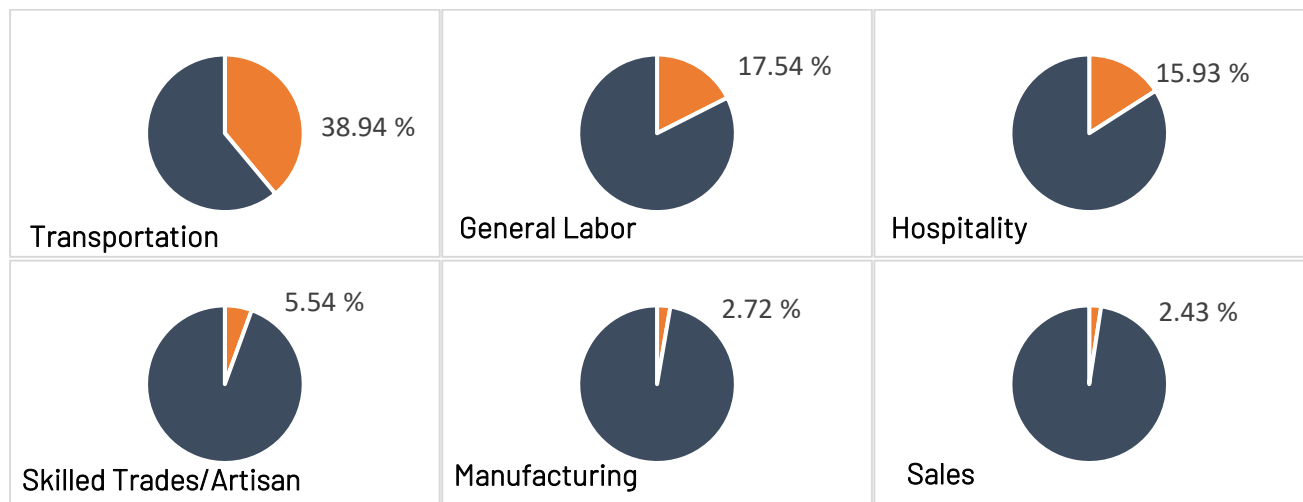
SVM (Support Vector Machine) is a supervised machine learning algorithm which can be used for classification or regression problems. It uses a technique called the kernel trick to transform your data and then based on these transformations it finds an optimal boundary between the possible outputs.

Benefits of SVM

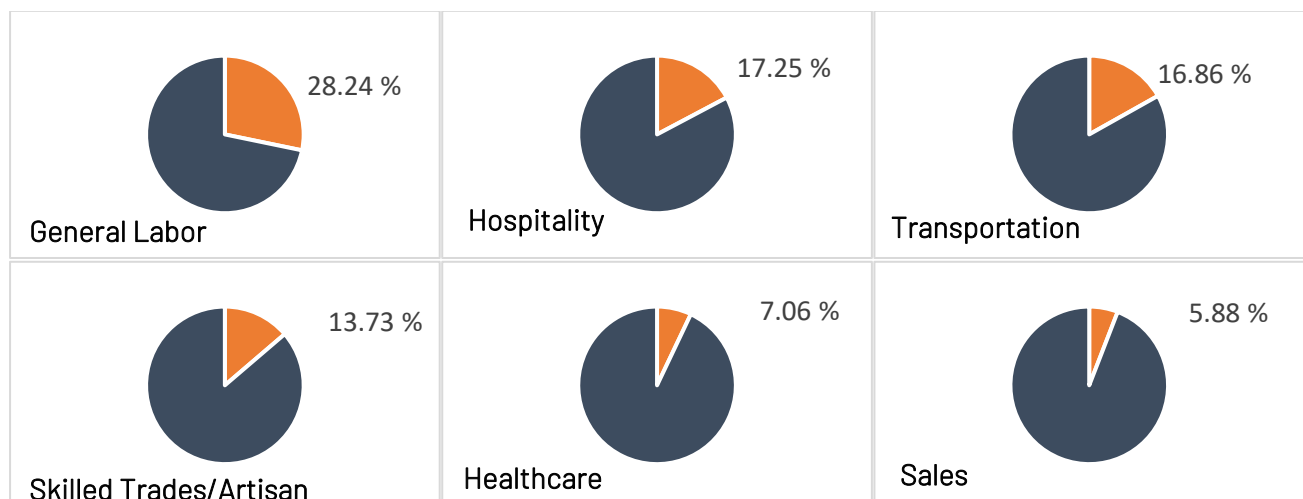
- It works well with a clear margin of separation
- It is effective in high dimensional spaces.
- It is effective in cases where the number of dimensions is greater than the number of samples.
- It uses a subset of training points in the decision function (called support vectors), so it is also memory efficient.

Training Data (Job ads) Distribution:

The below graph shows the top six categories that were present in the training data. Since most of the records are within these six categories, the trained model will predict unseen data that belong to these classes more accurately.



Resumes Data Distribution:



Step 5: Validation of our classification models

We have divided the jobs records into 70/30 training and test data sets to assess the model accuracy. The same model was then applied on resumes that were scraped from Craigslist. Since the resumes section did not have any categorization, we used manual check to validate the results of the model on the resume section. Below are few examples:

Categorization of Resumes:

Transportation

Driver Job Wanted (Chicago) CDL-Class-B Looking for Small Truck Delivery work Prefer Days Thanks Jeff show contact info

Foods Beverages

Private Chef/Executive Chef (Lakeview) Private chef available to cook in your home.

Manufacturing

quality tech control (west burbs) Manufacturing with metal and plastics. CMM, GD&T.

Step 6: Text Similarity Analysis to match jobs to resumes and vice-versa

We used cosine similarity for our text similarity analysis. Cosine similarity is a popular NLP method for approximating how similar two word/sentence vectors are. The intuition behind cosine similarity is to determine the cosine of the angle between the two vectors in order to quantify how similar two documents are. This ranges from 0 to 1, with 0 being the lowest (the least similar) and 1 being the highest (the most similar).

Based on this cosine similarity score we populated the top 3 resumes that best fit a particular job description on Craigslist and similarly we populated top 3 job descriptions that best match a particular resume on Craigslist. (Appendix A)

Business Impact

1. Job Seekers do not need to filter out job descriptions based on their preferred
 - Pay
 - Work schedule
 - Location
 - Qualifications
 - Duties
 - Benefits, etc.
2. Employers do not need to filter out resumes based on applicants'
 - Qualifications
 - Availability
 - Location
 - Past experiences
 - Remuneration, etc.

Also, the approach would aid the recruiter in expediting profile shortlisting while also ensuring the shortlisting process's authenticity, since they would be able to examine a large number of resumes in a short period of time, also with the proper fit, which a human would not be able to perform in near real time. This would help to make the process of recruiting individuals more efficient and successful in terms of selecting the best candidates.

3. Employers can easily filter resumes in the 'Resumes' section of Craigslist based on the categories we assigned, instead of using just the search engine to shortlist resumes

Future Scope

1. Gathering more data would help us to develop a deep learning model and we will be able to better predict the resume category.
2. Manually Tagging the resume so that results can be measured on similarity scores.
3. Use word-embedding and compare accuracy scores with TF-IDF vectorization

Conclusion

1. The testing accuracy on the job categories dataset turned out to be 78.93% based on which the categories for resumes were furnished. (Assumption being that the model predicts the categories for resumes on a $\pm 5\%$ of the above accuracy).
2. First similarity analysis produced three potential resumes for every job listing.
3. Second similarity analysis produced three potential job listings for every resume
4. This was done to reduce the hesitancy of people looking to find jobs on Craigslist and turn the website into a more job-friendly environment.

Appendix – A

1. Matching Job openings to resumes Example

Resume:

Dog sitter (Lake forest) Have worked in rescue with Fetching Tails and Almost home foundation. Worked with all breeds and sizes. My Great Dane just passed away and still have a lab. I'm a single mom abs looking to making some extra income by watching your dog at my home. I have ample space and would also be able to exercise and walk your pet. I have crates if needed. I'm flexible with price as I have a true passion for animals. I live in lake forest near Everett and Waukegan

Matched Job Openings:

Job Opening 1

Dog Daycare Handler/Overnight - Full time weekdays & weekends (Ravenswood)
POSITION: Dog handler/overnight attendant (experience preferred but not required)
HOURS: M-F 3p-11p Saturday/Sunday 3p-11p PAY: dependent on experience Bow Wow Lounge is looking for fellow dog lovers to join our pack! As a dog handler and overnight attendant at Bow Wow Lounge you will be responsible for safely monitoring dogs during off leash play during business hours. After 7pm the overnight attendant completes deep cleaning as well as attends to all overnight dogs until leaving at 11pm. This position is great for someone who is self-motivated to work alone for long periods! POSITION RESPONSIBILITIES Monitor evaluate and react as needed for dog body language pack dynamics and play styles. Manage non-aggressive and non-barking playgroups. Monitor safe and appropriate play correcting inappropriate behaviors in a consistent way. Feed monitor and care for overnight dogs after business hours Be a compassionate communicator Knowledge of harnesses/leashes Communicate observations about dog actions/behaviors to co-workers and Manager Clean and sanitize entire facility daily Keep supplies organized. Report supply stock levels to management for ordering before it runs out. SKILLS DEMANDS & WORKPLACE Dog daycare or boarding background preferred. Some dog handling skills and comfortable walking through 40+ dogs off leash Knowledge of general dog behavioral cues indicators and health issues to watch for. Be able to lift 40+ pounds and restrain dogs that weigh over 100 pounds. Be willing and able to stand for 8 hours (with a 30 min break in between) Must be able to use highly concentrated cleaning supplies. Must not be allergic to or afraid of dogs Must have a valid driver's license or state ID. Must be willing to get dirty and be comfortable around canine urine feces and vomit. JOB PERKS Optional employee healthcare (after 6 months) Bring your dog to work! TO APPLY Submit your application & resume via our website <http://www.bowwowlounge.com/contact/employment-opportunities/,7405437236>,

Job Opening 2

Dog Daycare/Boarding Pack Leader (Chicago) JOIN OUR SQUAD! Dog Lover Wanted- Part and Full time positions available open availability some weekends. Able to work some holidays will be given first consideration. Responsibilities: Supervise and ensure the safety of dogs in a pack setting encourage play and good behavior. Perform dog hotel duties including customer service and general tasks such as cleaning feedings baths taking social media pics. This position requires multitasking and attention to detail. Punctuality and Reliability: We have dogs dog owners and employees that rely on us and must always know they can count on us to be there for them. Requirements: Previous dog related work experience preferred (handling dog walking volunteer or other) but will train the right candidate. Ambition is valued just like experience. Must be confident in handling different size dogs. Compensation: TBD based on experience. Opportunity for

more hours and advancement an evaluation will be given after 90 days. We are loyal to our pack and promote from within. *Perks: Free daycare and boarding services for your dog.,

Opening 3

Dog bather (Northfield) Loving Care a grooming facility is looking for a full-time Dog bather. The hours are Monday through Friday 8 am. \xe2\x80\x93 4 pm. Saturday at 8:00AM \xe2\x80\x93 2 pm. Here one can expect to learn how to properly bathe and dry dogs and build relationships with not only coworkers but also the dogs! We also will ask to be flexible with taking the dogs to potty area throughout the day. There are no prerequisites that pertain to this job as you can learn while you work. The ideal employee for this position is one that loves animals just as much as we do. We pride ourselves on showing love and care to each dog we take in and making sure we meet all customers\xe2\x80\x99 wants/needs. Pay starts at \$13 per hour but there is plenty of room to grow. As you grow within the company so will your pay!

2. Matching resumes to job openings

Job Opening:

Find a job now in Chicago - Earn up to Up to \$16.80/hour (Chicago) Hello Amazon is hiring in Chicago. Earn up to Up to \$16.80/hour/hr as a Amazon Delivery Station Warehouse Associate. Health and safety are a top priority with all of our roles and sites. Learn more about the precautions Amazon is taking in our buildings to keep people healthy. Questions? Visit our FAQ page for more information. Amazon is always looking for ways to wow their customers and sometimes that means we\xe2\x80\x99re opening a new facility or one of our current locations is in high demand. This creates a unique opportunity that benefits new hires: bonus incentives based on the role or the location but only for a limited-time. We may be hiring near you so it literally pays to start now! Join the Amazon team and work at a warehouse close to home! No experience? Amazon warehouse jobs are a great way to start earning money with or without experience in a distribution center factory or warehouse. Amazon warehouse employees come from a variety of industries and experience levels. From Uber to Trader Joe\xe2\x80\x99s Amazon warehouses are home to associates with backgrounds ranging from transportation retail Walmart warehouses Target distribution centers FedEx warehouse package handlers UPS warehouses packing jobs shipping & receiving customer service creative industries general labor and more! Our employees have worked at Lowes Home Depot HD Supply OnTrac DHL and more\xe2\x80\x93 Ready to join our versatile team? Apply for a full-time part-time or seasonal job today!! Amazon is an Equal Opportunity Employer,

Matched Resumes:

Resume 1

Union laborer -demolition (Cicero) Hello I\xe2\x80\x99m a apprentice union laborer laid off looking for a new company to work for

Resume 2

General labor work (Chicago) My name is Mike I am looking for job as a Laborer I have experience in moving landscaping demolition and just general labor work. I am a hard worker have my own transportation

Resume 3

Looking for anything to do I\'m looking for something that\'s cleaning out apartments or yard work. I can do it after 2pm.

Appendix - B

Complete code can be found in the following GitHub repository:

https://github.com/ryan-egbert/aud_final_proj