

CASE STUDY-TEXT MINING ONLINE JOB POSTINGS

Copyright:



SCO 394, Sector-29

Next to Iffco Metro Station, Gurgaon – 122011

Website: www.analytixlabs.co.in

Email: info@analytixlabsl.co.in

Disclaimer: This material is protected under copyright act AnalytixLabs©, 2011-2018. Unauthorized use and/ or duplication of this material or any part of this material including data, in any form without explicit and written permission from AnalytixLabs is strictly prohibited. Any violation of this copyright will attract legal actions.

Business Context:

The project seeks to understand the overall demand for labour in the Armenian online job market from the 19,000 job postings from 2004 to 2015 posted on CareerCenter, an Armenian human resource portal. Through text mining on this data, we will be able to understand the nature of the ever-changing job market, as well as the overall demand for labour in the Armenia economy. The data was originally scraped from a Yahoo! Mailing group.

Business Objectives:

Our main business objectives are to understand the dynamics of the labour market of Armenia using the online job portal post as a proxy. A secondary objective is to implement advanced text analytics as a proof of concept to create additional features such as enhanced search function that can add additional value to the users of the job portal.

Business questions answering to our business objectives are defined as follows:

Job Nature and Company Profiles: What are the types of jobs that are in demand in Armenia? How are the job natures changing over time?

Desired Characteristics and Skill-Sets:

What are the desired characteristics and skill-set of the candidates based on the job description dataset? How these are desired characteristics changing over time?

IT Job Classification: Build a classifier that can tell us from the job description and company description whether a job is IT or not, so that this column can be automatically populated for new job postings. After doing so, understand what important factors are which drives this classification.

Similarity of Jobs: Given a job title, find the 5 top jobs that are of a similar nature, based on the job post.

Text Mining Goals:

The text mining goals is a set of sub-goals to answer our business questions:

For the IT Job classification business question, you should aim to create supervised learning classification models that are able to classify based on the job text data accurately, is it an IT job.

On the business question of Job Nature and Company Profiles. Unsupervised learning techniques, such as topic modelling and other techniques such as term frequency counting will be applied to the data, including time period segmented dataset. Qualitative assessment will be done on the results to help us understand the job postings.

To understand the desired characteristics and skill-sets demanded by employers in the job ads, unsupervised learning methods such as K-means clustering will be used after appropriate dimension reduction.

For Job Queries business question, we propose exploring the usage of Latent Semantic Model and Matrix Similarity methods for information retrieval. The results will be assessed qualitatively. To return the top 5 most similar job posting, the job text data are vectorised using different models such as word2vec, and doc2vec and similarity scores are obtained using cosine similarity scores, ranked and returned as the answer which is then evaluated individually for relevance.

Data Understanding:

The data was obtained from Kaggle competition. Each row represents a job post. The dataset representation is tabular, but many of the columns are textual/unstructured in nature. Most notably, the columns jobDescription, JobRequirement, RequiredQual, ApplicationP and AboutC are textual. The column jobpost is an amalgamation of these various textual columns.

Key Data Columns Description

Col Name	Description
Jobpost	The full text for the job post
Title	Title of the job
Announcement Code	Announcement code, which is some internal code and is usually missing.
Eligibility	Eligibility of the candidates.
StartDate	Start date of work.
Location	Employment location.
JobRequirement	Job requirements.
Salary	Job salary.
OpeningDate	Opening date of the job announcement.
Notes	Additional notes.
Attach	Attachments.
Month	The month of the announcement (derived from the field date).

Col Name	Description
Date	Date that the job was posted
Company	Company for the job
Term	Announcement code, which is some internal code and is usually missing.
Audience	Who can apply?
Duration	Duration of the employment.
JobDescription	Job Description.
RequiredQual	Required qualifications.
ApplicationP	Application procedure.
Deadline	Deadline for the job announcement.
AboutC	About the company.
Year	Year of the announcement (derived from the field date).
IT	TRUE if the job is an IT job.

Also provided sample job posting (attached with data set)

TITLE: Database Developer

LOCATION: Yerevan, Armenia

JOB DESCRIPTION: IUNetworks LLC is looking for a qualified Database Developer to design stable databases, according to the Company's needs. The incumbent will be responsible for developing, testing, improving and maintaining new and existing databases, ensuring the database systems run effectively and securely on a daily basis. He/ she will work closely with developers to ensure system consistency. He/ she will also collaborate with administrators and clients to provide technical support and identify new requirements.

JOB RESPONSIBILITIES:

- Design stable, reliable and effective databases;
- Optimize and maintain legacy systems;
- Modify databases according to requests and perform tests;
- Solve database usage issues and malfunctions;
- Liaise with developers to improve applications and establish best practices;
- Gather user requirements and identify new features;
- Develop technical and training manuals;
- Provide data management support to users;
- Ensure all database programs meet Company and performance requirements;
- Research and suggest new database products, services and protocols.

REQUIRED QUALIFICATIONS:

- BS In Computer Science or a relevant field;
- Proven work experience as a Database Developer;
- In-depth understanding of data management (e.g. permissions, recovery, security and monitoring);
- Knowledge of software development;
- Hands-on experience with SQL;
- Experience in programming with Oracle 11g Server, MS SQL Server, MySQL and PostgreSQL;
- Knowledge of Oracle Exadata is a plus;
- Experience in NoSQL (Couchbase/ MongoDB) is a plus;
- Excellent analytical and organization skills;
- Ability to understand users' requirements; a problem-solving attitude;
- Excellent verbal and written communication skills.

REMUNERATION/ SALARY: Competitive salary, based on skills and experience, plus a medical insurance and biannual company events.

APPLICATION PROCEDURES: Please apply to this job by sending your resumes to: job@iunetworks.am. Please mention the name of the position you are applying for in the subject line of the letter.

Please clearly mention in your application letter that you learned of this job opportunity through Career Center and mention the URL of its website - www.careercenter.am. Thanks.

OPENING DATE: 12 April 2017

APPLICATION DEADLINE: 08 May 2017

ABOUT COMPANY: IU Networks LLC is an information technology company that provides integrated solutions of hardware supply and software development. The Company was founded in March 2008.