



The fourth project of the Data Science Bootcamp T5

Unsupervised Learning and Natural Learning Processing
(NLP)

Job Postings

Modhi Alhbrdi

Modhi.alhbrdi@outlook.com

Rahaf Alyousef

rhof5111@gmail.com

November 2021



Abstract

The fourth project of the Data Science Bootcamp T5 is called job posting. Since finding and hiring the right people – fast – is essential to your business's success. Smart employers know their job posting plays an important role in their ability to attract top talent. If your job posting doesn't compel job seekers to apply, your time to hire will deteriorate, and so will your results.



Design

Using unsupervised learning and Natural Learning Processing (NLP), throughout the project we will classify jobs based on their job description.

That will give the requirement employee better performance to the advertiser for an available position, and the right position for the job seeker also, he can find similar jobs in his interest.



Data

This dataset contains job postings available on the jobs site. Found the dataset on the Kaggle website: [Here](#).

The dataset consists of 2,946 posts And includes many features it consists of Job ID, Job Category, Job Description, Preferred Skills, etc.



Algorithms

The project process started from NLP of determined columns: job description, business title and job type. Also apply machine learning topic modeling which is process of discovering “topics” that occur in collection of documents also a form of dimensionality reduction. Here the use of topic modeling is to reduce the dimensions of the data records we determine to reduce dimensions of the data to 4 main topics: Engineering & Planning, Technology, Environment, and Legal Affairs with 10 leading words.

By using all topic modeling models we got in **LDA** a probabilistic model, fit with TF-IDF vectorizer a clear leading words for each topic, but not the right result in modeling results are attached in communication.

LSA a matrix factorization machine learning approach that can also be used for topic modeling, fit with CountVectorizer method, the model performance giving a confusing leading words to the topics, so it's been hard to determine which is which, that's why we need to apply all the models to find the closer result to reality.

The last model **NMF** also a matrix factorization topic model fit with TF-IDF vectorizer, give the clearest and the leading words making sense and easy to determine the topics, and after training the model it gives the right results in test set, and by using cosine_similarity method of the vectors is another good approval, T- SNE clustering delivered by NMF model to visualize are results, results are attached in communication.

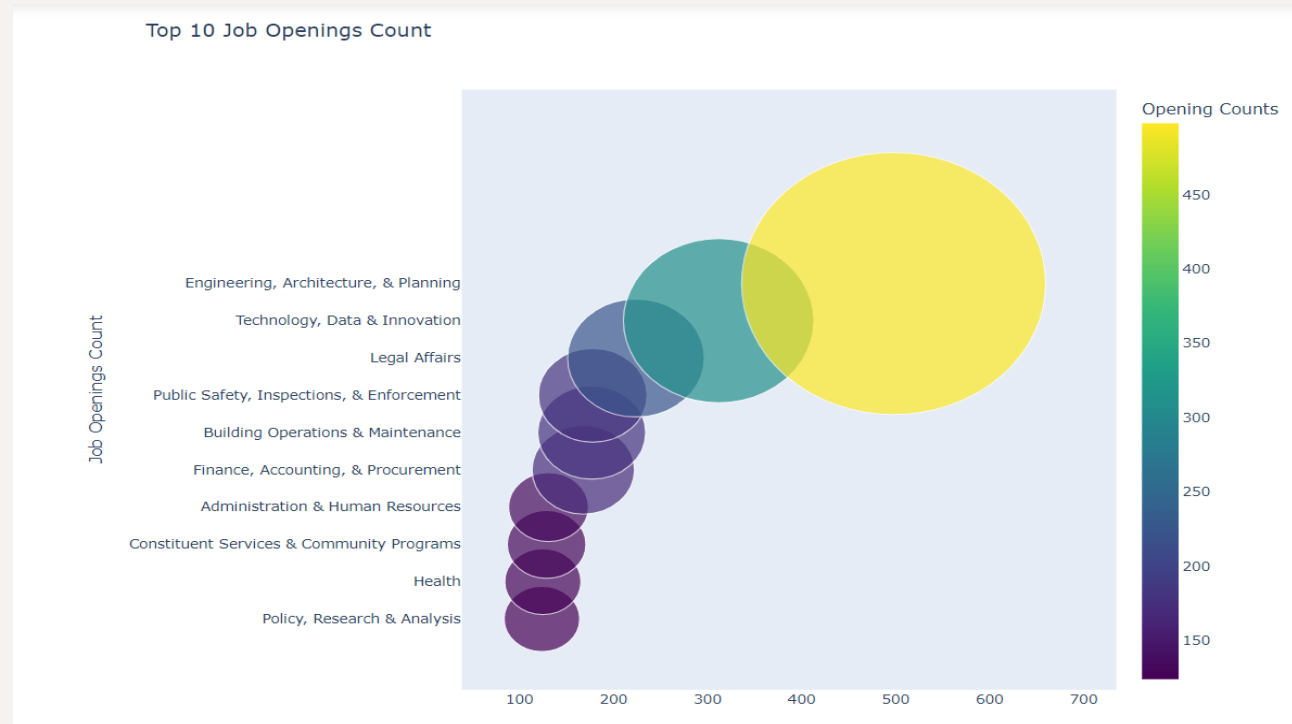


Tools

- **Technologies:** Python - Jupyter Notebook.
- **Libraries:** Pandas – NumPy – Matplotlib – Seaborn – WordCloud - plotly – bokeh - Pandas_profiling - nltk – Sklearn - TSNE



- EDA



- LDA

	Word 0	Word 1	Word 2	Word 3	Word 4	Word 5	Word 6	Word 7	Word 8	Word 9
Technology	develop	manag	program	agenc	support	data	work	plan	servic	provid
Environment	water	wastewat	manag	oper	protect	bureau	engin	project	environment	work
Engineering & Planning	project	manag	contract	program	agenc	review	report	construct	work	servic
Legal Affairs	case	assist	investig	perform	offic	agenc	work	staff	provid	servic

	Technology	Environment	Engineering & Planning	Legal Affairs	dominant_topic
Doc0	0.41	0.00	0.59	0.0	Engineering & Planning

	Processed_Desc	Business Title	Division/Work Unit	Job Type
108	supervision latitude independent action exercise independent judgment infrastructure project manager responsible scoping planning execution infrastructure project project focus delivery new capability application service infrastructure candidate specific duty include work project stakeholder clearly define manage expectation objective assumption measurement schedule milestone scope requirement risk issue escalation change metric success criterion communication change control successful deployment actively communicate project stakeholder ensure alignment visibility transparency aspect assigned project manage vendor engagement foster accountability deliverable timeline quality project delivery develop maintain project charter project planschedules implementation plan status report risk mitigation contingency plan qa plan project budget cost capital funding request rfp strategic plan business continuity plan account resource capacity organizational change competing initiative project constraint utilize project management process methodology tool ensure project delivered time within budget adhere high quality standard meet stakeholder expectation build lead mentor diverse project team throughout project lifecycles collaborate unit throughout project lifecycles successful implementation project deliverable	infrastructure project manager	mi adminnetwork sysms dvlpmt	information technology telecommunication

■ LSA (SVD)

Topic 0	manag, project, develop, program, work, includ, agenc, water, respons, plan	???
Topic 1	water, wastewat, engin, bureau, protect, construct, project, sewer, environment, suppli	Environment ??
Topic 2	project, manag, energi, construct, design, capit, program, schedul, scope, effici	???
Topic 3	develop, program, build, plan, water, energi, hous, track, polici, sewer	Engineering & Planning ??

■ NMF

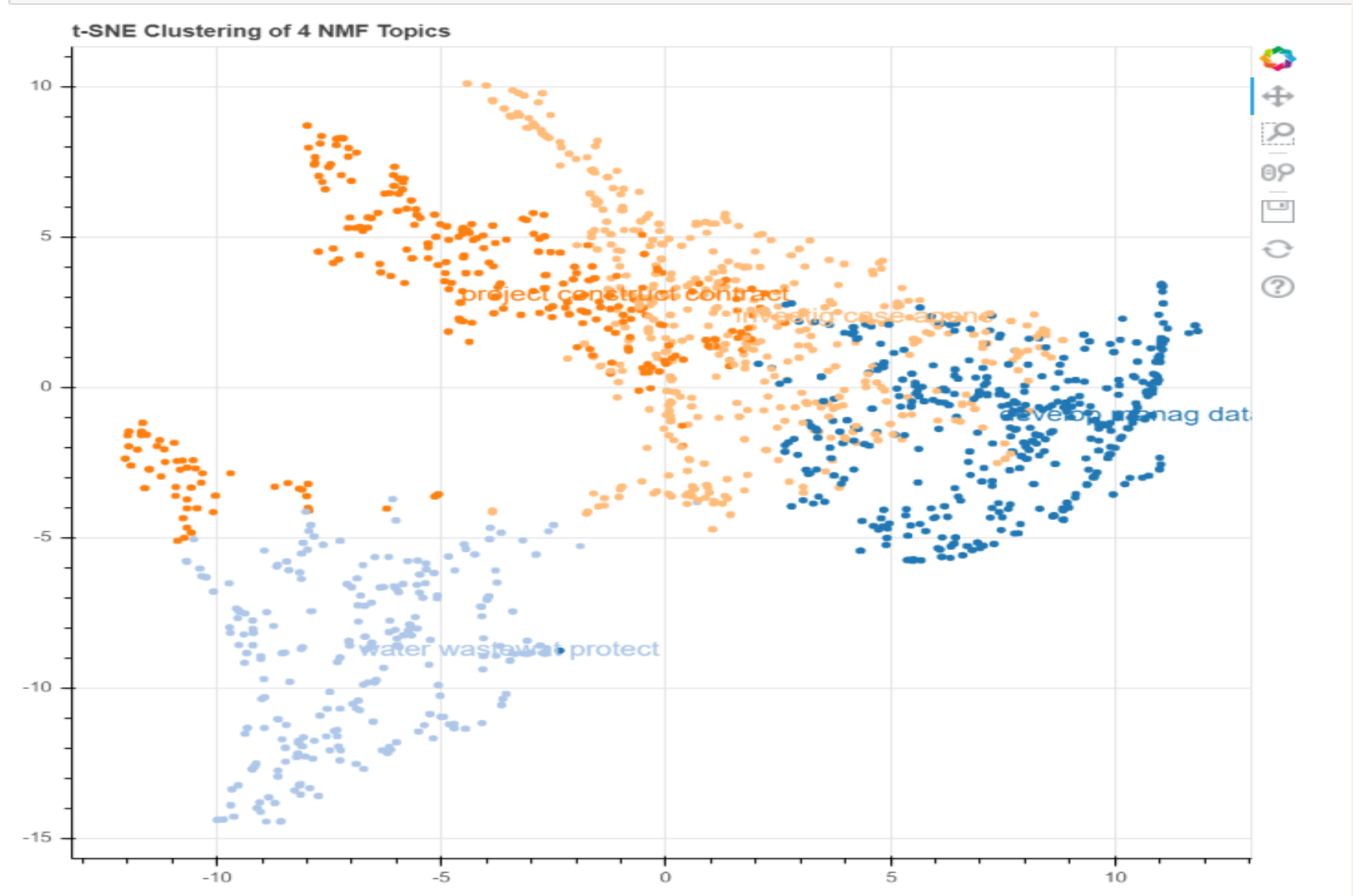
	Word 0	Word 1	Word 2	Word 3	Word 4	Word 5	Word 6	Word 7	Word 8	Word 9
Technology	develop	data	manag	support	technolog	busi	servic	agenc	team	plan
Environment	water	wastewat	protect	sewer	bureau	environment	oper	drink	drink water	gallon
Engineering & Planning	project	construct	contract	design	manag	engin	capit	budget	design construct	project manag
Legal Affairs	investig	case	attorney	assist	enforc	legal	nycha	perform	court	agenc

2787	nyc department environmental protection dep enriches environment protects public health new yorkers providing billion gallon high quality drinking water managing wastewater stormwater reducing air noise hazardous material pollution dep largest combined municipal water wastewater utility country nearly employee deps water supply system comprised reservoir controlled lake throughout system square mile watershed extends mile north west city dep office energy resource recovery program dep energy office housed within commissioner office committed implementing energy management best practice across deps portfolio building operating facility help city meet ambitious greenhouse gas emission energy reduction goal dep energy office work track overall energy usage identify energysaving opportunity building implement energyefficient operation maintenance practice deliver energy efficiency retrofit project provide energy resource recovery performance tracking program assessment scientific research emergency planning internal auditing service dep energy office seeking hire energy manager deps bureau wastewater treatment work direction director dep energy office duty include limited identify implement energy resource recovery research opportunity project apply external grant program enrollment including limited energy funding reliability program report project progress support advise bureau throughout procurement design policyssetting order influence decisionmaking process toward energy resource efficiency share energy resource recovery best practice identify employee candidate energy resource recovery training identify energy conservation opportunity potential energy efficiency project associated wastewater conveyance treatment process building system equipment contribute project scope work cost estimate support measurement verification effort manage bureau heat light power budget including monthly billing review annual budgetsetting process investigate correct potential billing error determine report root cause change energy usage demand coordinate energy utility meter account service request inspection replacement interruption bureau provide project management assistance including developing maintaining project plan document managing coordinating project element reviewing milestone ensuring efficient progress implementation consistent monitoring progress initiative commitment assist senior staff project management responsibility including adhering internal external timeline ensuring completion highquality deliverable managing budget may also responsible one discrete project expected manage project completion track provide energy resource recovery data needed reporting compliance purpose prepare edit proofread present highquality report presentation article meeting correspondence document report energy resource recovery communication material datasets participate ad hoc project arise provide overall support dep energy office mission	bureau energy manager	wtenergy	engineering architecture planning policy research analysis
------	---	-----------------------------	----------	---

	Technology	Environment	Engineering & Planning	Legal Affairs	dominant_topic
Doc0	0.050	0.000	0.162	0.000	Engineering & Planning
Doc1	0.138	0.005	0.014	0.000	Technology
Doc2	0.142	0.000	0.000	0.021	Technology
Doc3	0.031	0.051	0.080	0.000	Engineering & Planning
Doc4	0.019	0.000	0.084	0.059	Engineering & Planning

```
vec1 = [0.138,0.005,0.014,0.000]
vec2 = [0.142,0.000,0.000,0.021] #same
print(cosine_similarity([vec1, vec2]).round())

[[1. 1.]
 [1. 1.]]
```



t-SNE Clustering of 4 Topics of NMF model