

MIUUL PROJECT

miuul

# LINKEDIN İŞ İLANI ANALİZİ

Sunan: İsmail Asilkan VAR

# Gündem

Problem Tanımı  
Kullanılan teknolojiler  
Proje Akış Şeması  
Data & NLTK  
Power BI

# Problem Tanımı

Günümüzde verinin bu kadar erişilebilir olduğu bir durumda özellikle veri bilimi iş ilanlarının sayısı giderek arttığı görülmektedir. Bu noktada bazen yüzlerce ilan üzerinden öncelikli edinilmesi gereken özellikleri belirlemek oldukça önem kazanmaktadır. Projemin amacı iş arayan bir veri bilimcinin LinkedIn üzerinden erişebildiği bir takım ilan üzerinde veri analizi yaparak en çok istenen özellikleri belirlemesi üzerindedir.



# KULLANILAN TEKNOLOJİLER



WEB SCRAPING

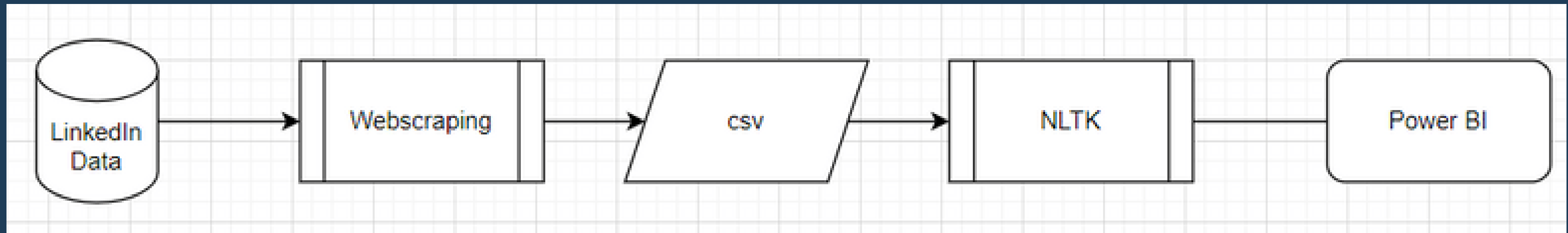


NLTK



POWER BI

# Akış Şeması





DATA

**job\_name**  
**company\_name**  
**location**  
**job\_details**

**Data Scientist** → job\_name  
Enerjisa Üretim → company  
Istanbul, Turkey (Hybrid) → location  
3 days ago - Over 200 applicants

Full-time · Associate  
501-1,000 employees · Electric Power Generation

See recent hiring trends for Enerjisa Üretim. [Try Premium for free](#)

Actively recruiting

[Easy Apply](#) [Save](#)

job\_details

We are looking for "Data Scientist" who can demonstrate their knowledge and skills in Enerjisa Üretim Data Analytics department with qualifications below:

- Bachelor's degree in quantitative field (Mathematics, Statistics, Industrial Engineering etc.)
- Hands on experience in scripting languages (Python, R)
- Interest in real-world optimization problems and ability to develop large MIP algorithms or exact heuristics to solve these problems (experience of coding in XPRESS, CPLEX or GUROBI)
- Ability to analyze big data and manage complex calculations
- SQL knowledge is an asset
- Machine Learning experience is an asset
- Knowledge of Turkish energy sector is an asset

**JOB DESCRIPTION**

- Developing quantitative solutions centered around energy generation, trading, and derivatives
- Developing statistical forecasting models, simulation tools, optimization algorithms and mathematical models for stochastic processes
- Performing fundamental data analysis
- Preparing and maintaining methodology documents
- Assisting IT teams for model implementation by regarding the secure software development life cycle.



NLTK

Natural Language toolkit olarak bilinen paket, NLP (doğal dil işleme) için en çok kullanılan paketlerden biridir. Standart bir akışta stopword, tokenizer, lemmazation / stemming kullanılsa da ben ingilizce dilini referans alarak sadece stopwords kullanmayı tercih ettim.

## stopwords

Birçok dil için tanımlanmış analiz için genellikle gereksiz kelimeleri, çalışmanızdan atmak için kullanılır.

```
1 import nltk
2 from nltk.corpus import stopwords
3 stop = stopwords.words('english')
4 print(stop)
```

```
['i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves', 'you', "you're", "you've", "you'll", "you'd", 'your', 'yours', 'yourselves', 'he', 'him', 'his', 'himself', 'she', "she's", 'her', 'hers', 'herself', 'it', "it's", 'its', 'itself', 'they', 'them', 'their', 'theirs', 'themselves', 'what', 'which', 'who', 'whom', 'this', 'that', "that'll", 'these', 'those', 'am', 'is', 'are', 'was', 'were', 'be', 'been', 'being', 'have', 'has', 'had', 'having', 'do', 'does', 'did', 'doing', 'a', 'an', 'the', 'and', 'but', 'if', 'or', 'because', 'as', 'until', 'while', 'of', 'at', 'by', 'for', 'with', 'about', 'against', 'between', 'into', 'through', 'during', 'before', 'after', 'above', 'below', 'to', 'from', 'up', 'down', 'in', 'out', 'on', 'off', 'over', 'under', 'again', 'further', 'then', 'once', 'here', 'there', 'when', 'where', 'why', 'how', 'all', 'any', 'both', 'each', 'few', 'more', 'most', 'other', 'some', 'such', 'no', 'nor', 'not', 'only', 'own', 'same', 'so', 'than', 'too', 'very', 's', 't', 'can', 'will', 'just', 'don', "don't", 'should', "should've", 'now', 'd', 'll', 'm', 'o', 're', 've', 'y', 'ain', 'aren', "aren't", 'couldn', "couldn't", 'didn', "didn't", 'doesn', "doesn't", 'hadn', "hadn't", 'hasn', "hasn't", 'haven', "haven't", 'isn', "isn't", 'ma', 'mightn', "mightn't", 'mustn', "mustn't", 'needn', "needn't", 'shan', "shan't", 'shouldn', "shouldn't", 'wasn', "wasn't", 'weren', "weren't", 'won', "won't", 'wouldn', "wouldn't"]
```



# POWER BI

Oluşturulan analizler bir dataframe üzerinden csv dönüştürüldükten sonra Power BI üzerinde sql tabloları üzerinde düzenlenerek, görselleştirme yapıldı.

Word_Count	words	prob_languages_total	prob_languages
62	python	0.837837838	0.373493976
45	sql	0.608108108	0.271084337
20	r	0.27027027	0.120481928
15	java	0.202702703	0.090361446
10	C++	0.135135135	0.060240964
7	Scala	0.094594595	0.042168675
5	C	0.067567568	0.030120482
2	julia	0.027027027	0.012048193



Oğuz, İbrahim, Blirind ve Hande

**TEŞEKKÜRLER**

