

# KSA Ministry of Justice NLP

## **Abstract:**

The goal of this project is to be able to reach the nearest five cases similar to the same entered case In terms of the type of trial and judgment

## **Design:**

Distance Cosine matrix (Cosine similarity)

## **Data:**

<https://sjp.moj.gov.sa/Filter?isFilterButtonClicked=True>

The data was extracted from the Saudi Ministry of Justice website via(web scraping )  
It consists of more than 2000 rows and 13 columns.

## **Algorithms:**

The data was extracted from the site

Data cleaning

Preprocessing (stop word, punctuation, spelling correction, delete numbers, English word)

Steamer

TF-IDF

Cosine similarity

## **Tools:**

Python and Jupiter Notebook

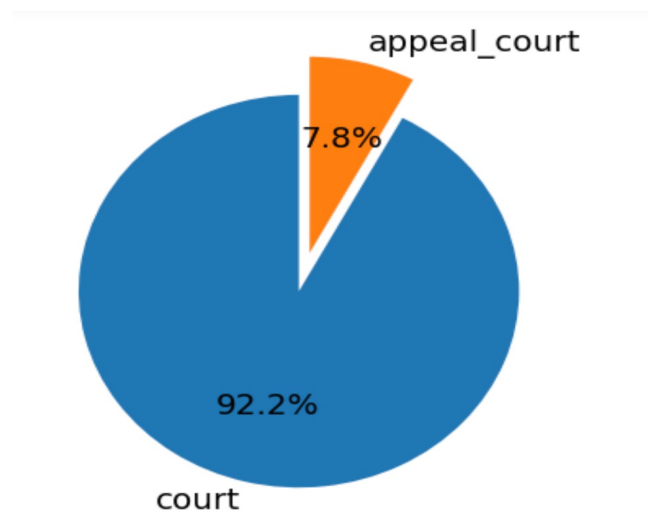
NumPy and Pandas for data manipulation

Matplotlib and Seaborn for plotting visualization

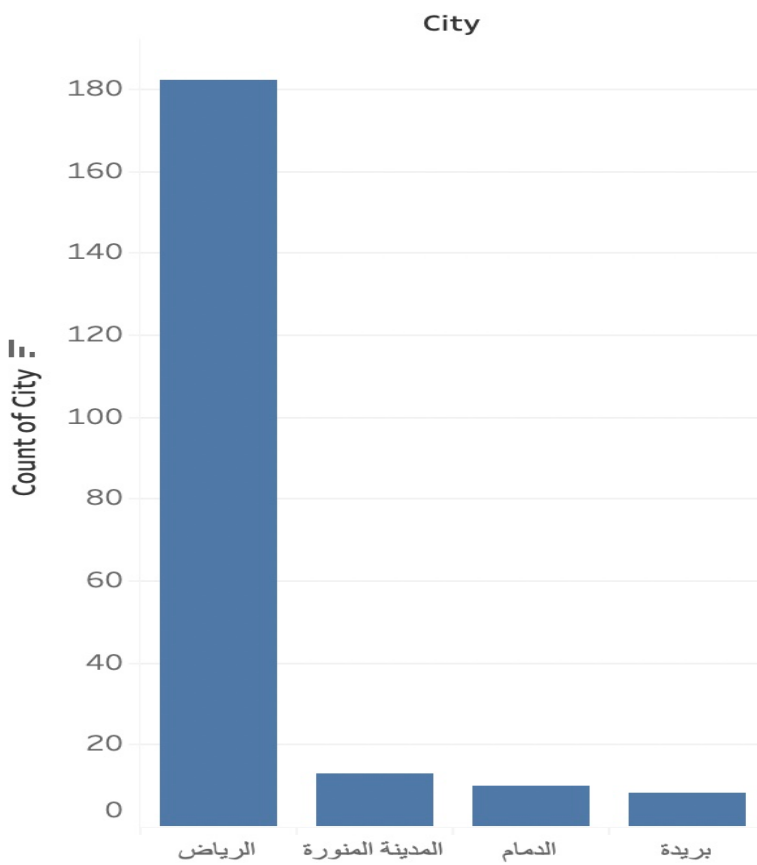
SkLearn

### Communication:

#### Appeal court percentage according to Court percentage



#### Most city where judgments issued



Court	Count of Court
المحكمة التجارية	165
المحكمة العليا - البيئة العامة	40
المحكمة العامة	5

Word cloud judgment\_text

### Cosine similarity

```
cosine_similarity(dt_tfidf.loc[25:25,:],dt_tfidf).argsort()[0][-6:]  
array([ 235,  982,  735,   87, 1108,   25])
```

```
cosine_similarity(dt_tfidf.loc[25:25,:],dt_tfidf.loc[1108:1108,:])  
array([[0.70741606]])
```

**By:**

**Maria Felemban**

**Ghaidaa Khalid**

**Reem Alghamdi**