KSA Ministry of Justice

NLP project and cosine similarity

SDAIA
الهيئة السعودية للبيانات والذكاء الاصطناعي
Saudi Data & AI Authority

Problem statement

The data contains judicial information, including final commercial judgments issued by commercial courts, this is where our project will help find the most similar judgments.

contents

1 **collect data**

Web scrapping
and tools

2 **Preprocessing**

- Data cleaning
- Text preprocessing

3 **Convert to vector**

TF –IDF
SVD

4 **Distances matrix**

Cosine similarity

5

Tools

- Python and Jupiter Notebook
- NumPy and Pandas for data manipulation
- Matplotlib and Seaborn for plotting visualization
- SkLearn

1

Collect data

Web scrapping

بيانات الحكم

المحكمة: المحكمة العامة
المدينه: بريدة
رقم القضية - القرار: ٣٣٢ تاريخها: ٢٦/١٠/١٤٤٢
محكمة الاستئناف: المحكمة التجارية
المدينة: منطقة الرياض
رقم القرار: ٥٥٢٨ تاريخه: ٢٤/١٠/١٤٤٣

التصنيف

عنوان الحكم

نص الحكم

الاستئناف

Cleaning data

1. Rename columns
2. Drop duplicate values
3. Fill null values

2

Preprocessing

NLP on Arabic data

Preprocessing

1. Remove punctuations
2. remove \n
3. remove numbers

Preprocessing

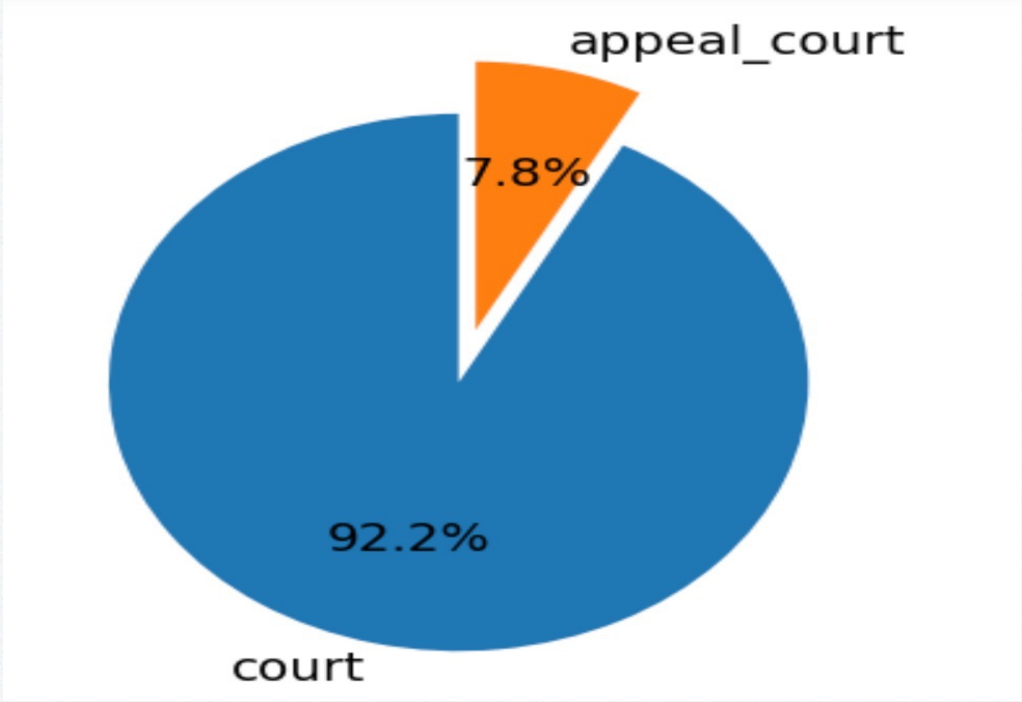
1. Remove stop words
2. Correct words
3. remove English words

3

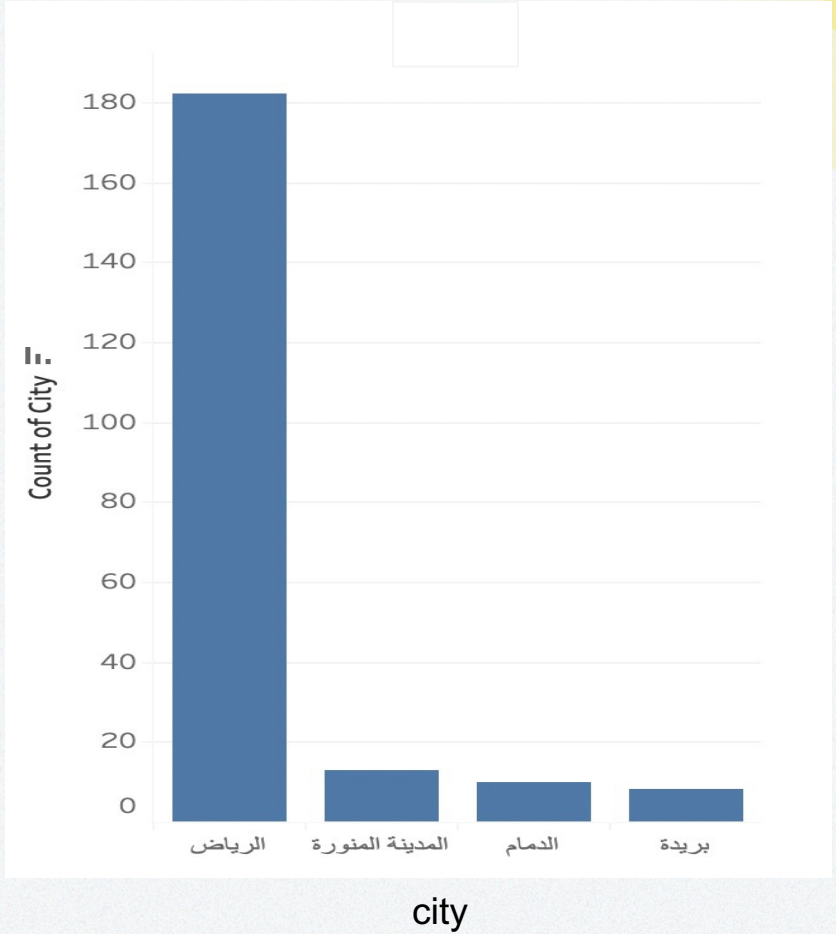
EDA

Shape data : (2043, 8)

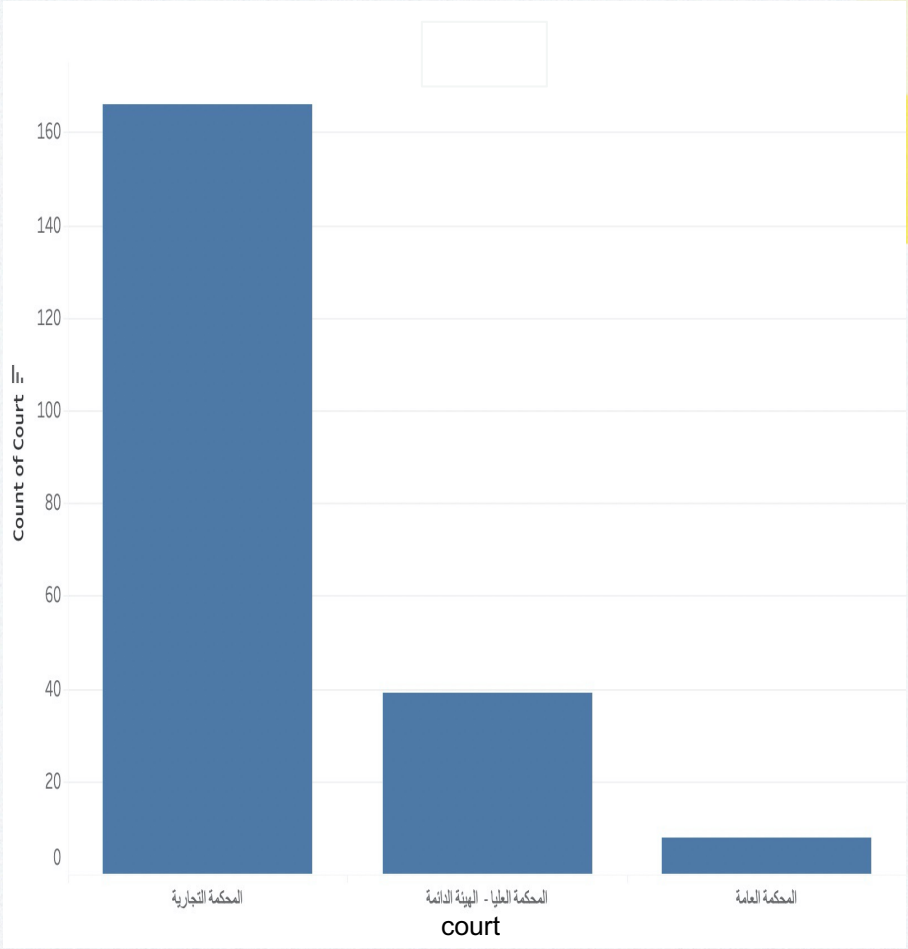
Appeal court percentage according to Court percentage



Most city where judgments issued



What is the most redundunt court



Word cloud

Word cloud judgment_text

3

TF-IDF

- In this step after cleaning and preprocessing the TF-IDF is used to give weight to word that has less redundancy in the whole document

4

Cosine similarity

Output distance Cosine matrix

```
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]])
```

Exsample:

```
df_train.judgment_text.iloc[25]
```

'الحمد لله والصلاة والسلام على رسول لله أما بعدفلدى الدائرة التجارية العاشرة وبناء على القضية رقم لعام هالمقامة من عبد لله بن عبدالرحمن بن محمد الرشيد سجل تجاري ضد مصنع الغدير للمنتجات الورقية ورق غير ذلك القاضي سلمان بن عبدالرحمن الو لبعي رئيسا الوقائعتتحصل وقائع هذه القضية بالقدر اللازم لإصدار هذا الحكم بتقدم وكيل المدعي إلى المحكمة التجارية بالدم ام بلائحة ادعاء يختصم فيها المدعى عليها تتلخص في طلبه إلزام المدعى عليها بدفع مبلغ قدره سبعة وستون ألفا واثنان وثلاث ون ريالا والتي تمثل قيمة زخام وحجر قام موكله بتوريدها للمدعى عليها ولم تقم بسداد ثمنها وإلزامها بدفع مبلغ قدره عشرو ن ألف ﷼ كأتعاب للمحاماة قيدت القضية بالرقم المشار إليه أعلاه وأحيلت إلى هذه الدائرة وفي سبيل نظرها حددت لها جلسة هذا اليوم وفيها حضر محمد عبدالله بن محمد الحليبي هوية وطنية رقم بصفته وكيل اعن المدعي بالوكالة رقم وتاريخ فيما تبين عدم حضور من يمثل المدعى عليها رغم تبلغها بموعد هذه الجلسة وبسؤال المدعي أصالة عن دعواه أحال على لائحة الدعوى وبسؤاله عن بينته على الدعوى أحال على مطابقة المدعى عليها على الرصيد بتاريخ والمنبلة بتوقيع وختم المدعى عليها عليه قررت رفع ال جلسة لإصدار الحكم الأسبابوبعد سماع الدعوى والاطلاع على أوراق القضية ومستنداتها وحيث أن دعوى وكيل المدعي تنحصر في طلبه إلزام المدعى عليها بدفع مبلغ قدره سبعة وستون ألفا واثنان وثلاثون ريالا والتي تمثل قيمة زخام وحجر قام موكله بتوريدها للمدعى عليها ولم تقم بسداد ثمنها وإلزامها بدفع مبلغ قدره عشرون ألف ﷼ كأتعاب للمحاماة مستندا في دعواه على مطابقة ال مدعى عليها على الرصيد بتاريخ والمنبلة بتوقيع وختم المدعى عليها وقد نصت المادة من نظام المحكمة التجارية على تعد ال ورقة العادية صادرة ممن نسبت إليه مالم ينكر صراحة ما هو منسوب إليه فيها من خط أو إمضاء أو ختم أو بصمة يؤيد ذلك تخلف المدعى عليه عن الحضور رغم تبلغها عليه فإن الدائرة تذهب إلى إلزام المدعى عليها بما ترتب في ذمتها وأما طلب وكيل المدع ي بإلزام المدعى عليه بأتعاب المحاماة فإن الحكم باستحقاق صاحب الحق للنفقات التي غرمها لأجل الدعوى إنما يكون حال ثبوت مماطلة من تعلق الحق بذمته وامتنع عن الوفاء به وشرط إلزام المماطل في أداء الحق بهذه النفقات أن يكون غرمها على وجه مع تاد كما نص على ذلك شيخ الإسلام ابن تيمية رحمه لله في الاختيارات إذ قال ومن مطل صاحب الحق حقه حتى أحوجه إلى الشكاية فما غ رمه بسبب ذلك فهو على الظالم المبطل إذا كان غرمه على الوجه المعتاد وقال الإمام المرداوي رحمه لله في الإنصاف لو مطل غريمه حتى أحوجه إلى الشكاية فما غرمه بسبب ذلك يلزم المماطل وحيث إن المدعى عليها قد امتنعت عن السداد رغم إقرارها بالمبلغ م ما اضطر المدعي إلى اللجوء إلى المحكمة لانتزاع حقه من المدعى عليها فإن المدعي مستحق لهذا الطلب ولكن الدائرة ترى بأن م بلغ المطالبة كأتعاب للمحاماة مبالغ فيه حيث أن القضية يسيرة كما أنه لم يبذل فيها مزيد جهد وفيها مصادقة وفصل فيها بجل سة واحدة فقط وبما للدائرة من سلطة تقديرية فهي تنتهي إلى استحقاق المدعية لمبلغ قدره ستة آلاف سبعمائة وثلاثة ريالات فقط كأتعاب للمحاماة والتقاضي وهو ما يمثل ما نسبته تقريبا من المبلغ المحكوم فيه منطوق الحكمحكمت الدائرة بإلزام المدعى ع ليه مصنع الغدير للمنتجات الورقية ورق سجل تجاري رقم بأن تدفع للمدعي عبدالله بن عبدالرحمن بن محمد الرشيد هوية وطنية رقم مبلغ قدره ثلاثة وسبعون ألفا وسبعمائة وخمسة وثلاثون ريالا ولله الموفق وصلى لله وسلم على نبينا محمد وعلى آله وصحبه أجمعينرئيس الدائرةسلمان بن عبدالرحمن الوليعي'

```
df_train.judgment_text.iloc[1108]
```

'الحمد لله والصلاة والسلام على رسول لله أما بعدفلدى الدائرة التجارية العاشرة وبناء على القضية رقم لعام هالمقامة من شرك ة مصنع فابكو للبلاستيك المحدودة سجل تجاري ضد سعيد عوضان القحطاني هوية وطنية القاضي سلمان بن عبدالرحمن الوليعي رئ يسا الوقائعتتحصل وقائع هذه القضية بالقدر اللازم لإصدار هذا الحكم بتقدم وكيلة المدعية إلى المحكمة التجارية بالدمام بلائ حة ادعاء تختصم فيها المدعى عليها تتلخص في طلبها إلزام المدعى عليه بدفع مبلغ قدره مائتان وخمسة الاف وتسعمائة واثنان واربعون ﷼ والتي تمثل قيمة توريد مواسير بلاستيكية قامت موكلتها بتوريدها للمدعى عليه ولم يقم المدعى عليه بدفع ثمنها وإ لزامه بدفع مبلغ قدره خمسة وعشرون ألف ﷼ كأتعاب للمحاماة قيدت القضية بالرقم المشار إليه أعلاه وأحيلت إلى هذه الدائرة وفي سبيل نظرها حددت لها جلسة هذا اليوم وفيها حضرت بشائر احمد عبدالوهاب القناص هوية وطنية رقم بصفتها وكيلة المدعية وكالة رقم بتاريخ فيما تبين عدم حضور من يمثل المدعى عليه رغم تبلغه بموعد هذه الجلسة وبسؤال المدعية وكالة عن دعوى م وكلتها أحالت على لائحة الدعوى وبسؤالها عن بينتها على الدعوى أحالت على مطابقة المدعى عليه على الرصيد بتاريخ والمنبل بتوقيع المدعى عليه وختم مؤسسته عليه قررت رفع الجلسة لإصدار الحكم الأسباب وبعد سماع الدعوى والاطلاع على أوراق القضية ومس تنداتها وحيث أن دعوى وكيلة المدعية تنحصر في طلبها إلزام المدعى عليه بدفع مبلغ قدره مائتان وخمسة الاف وتسعمائة واثنا ن واربعون ﷼ والتي تمثل قيمة توريد مواسير بلاستيكية قامت موكلتها بتوريدها للمدعى عليه ولم يقم المدعى عليه بدفع ثمنها وإلزامه بدفع مبلغ قدره خمسة وعشرون ألف ﷼ كأتعاب للمحاماة مستندة في دعواها على مطابقة المدعى عليه على الرصيد بتاريخ والمنبل بتوقيع المدعى عليه وختم مؤسسته وقد نصت المادة من نظام المحكمة التجارية على تعد الورقة العادية صادرة ممن نس بت إليه مالم ينكر صراحة ما هو منسوب إليه فيها من خط أو إمضاء أو ختم أو بصمة يؤيد ذلك تخلف المدعى عليه عن الحضور رغم تبلغه عليه فإن الدائرة تذهب إلى إلزام المدعى عليه بما ترتب في ذمتها وأما طلب وكيلة المدعية بإلزام المدعى عليه بأتعا ب المحاماة فإن الحكم باستحقاق صاحب الحق للنفقات التي غرمها لأجل الدعوى إنما يكون حال ثبوت مماطلة من تعلق الحق بذمته وامتنع عن الوفاء به وشرط إلزام المماطل في أداء الحق بهذه النفقات أن يكون غرمها على وجه معتاد كما نص على ذلك شيخ الإس لام ابن تيمية رحمه لله في الاختيارات إذ قال ومن مطل صاحب الحق حقه حتى أحوجه إلى الشكاية فما غرمه بسبب ذلك فهو على الظال م المبطل إذا كان غرمه على الوجه المعتاد وقال الإمام المرداوي رحمه لله في الإنصاف لو مطل غريمه حتى أحوجه إلى الشكاية فما غرمه بسبب ذلك يلزم المماطل وحيث إن المدعى عليه قد امتنع عن سداد ما ترتب في ذمته مما اضطر المدعية إلى اللجوء إلى الم حكمة لانتزاع حقها من المدعى عليه فإن المدعية مستحق لهذا الطلب ولكن الدائرة ترى بأن مبلغ المطالبة كأتعاب للمحاماة مبا لغ فيه حيث أن القضية يسيرة وليست معقدة كما أنه لم يبذل فيها مزيد جهد وفيها مصادقة وفصل فيها بجلسة واحدة فقط وبما لل دائرة من سلطة تقديرية فهي تنتهي إلى استحقاق المدعية لمبلغ قدره عشرون ألفا وخمسمائة وأربعة وتسعون ريالا فقط كأتعاب ل لمحاماة والتقاضي وهو ما يمثل ما نسبته تقريبا من المبلغ المحكوم فيه منطوق الحكمحكمت الدائرة بإلزام المدعى عليه سعيد عوضان القحطاني هوية وطنية رقم بأن يدفع للمدعية شركة مصنع فابكو للبلاستيك المحدودة سجل تجاري رقم مبلغا قدره مائتان وستة وعشرون ألفا وخمسمائة وستة وثلاثون ريالا ولله الموفق وصلى لله وسلم على نبينا محمد وعلى آله وصحبه أجمعينرئيس الدائرة سلمان بن عبدالرحمن الوليعي'

Thanks!

Do you have any questions?