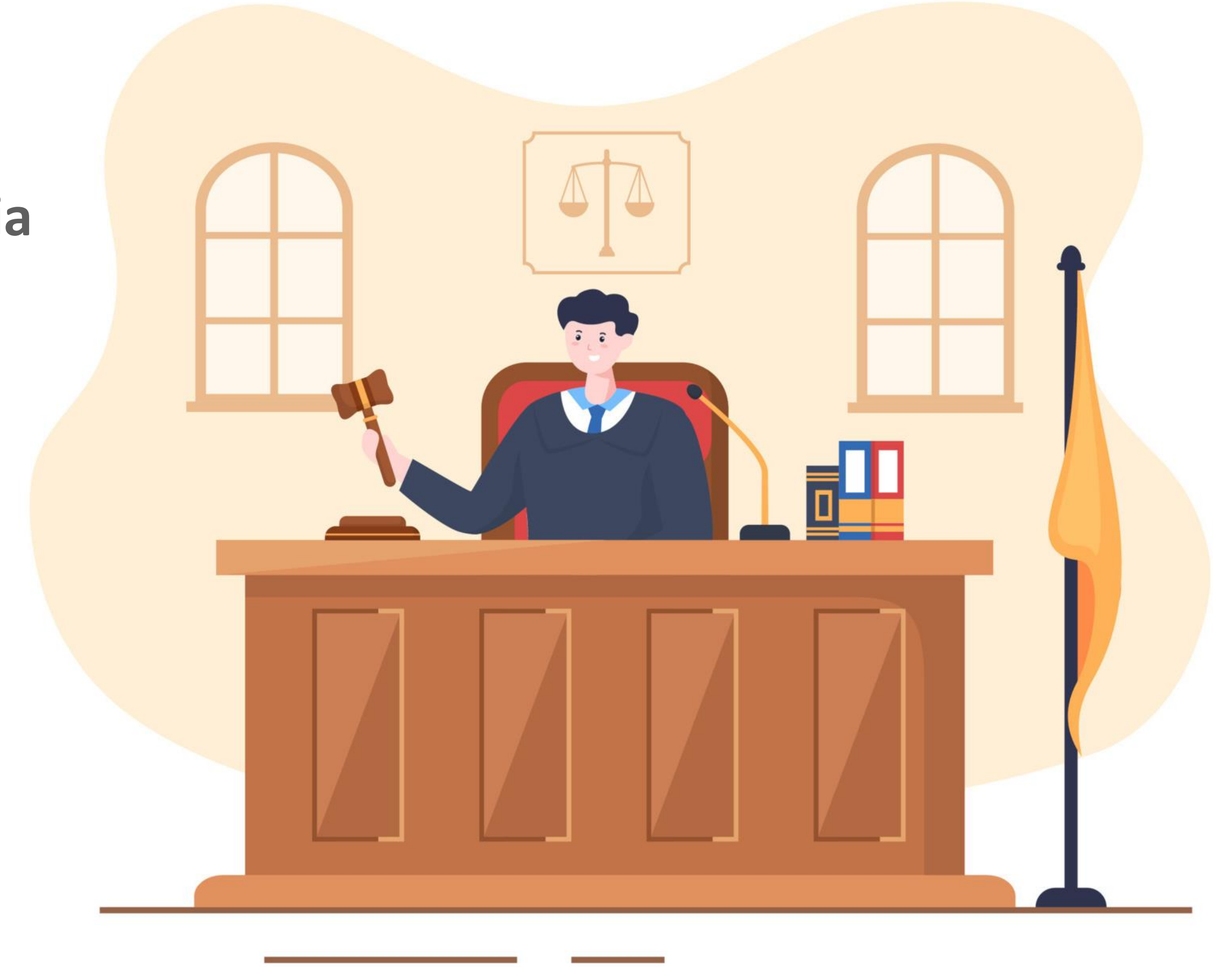
LAWYER CASE RECOMMENDATION SYSTEM

Recommendation **S**ystems

Prof. **R**aghuram **B**hardwaj

INTRODUCTION

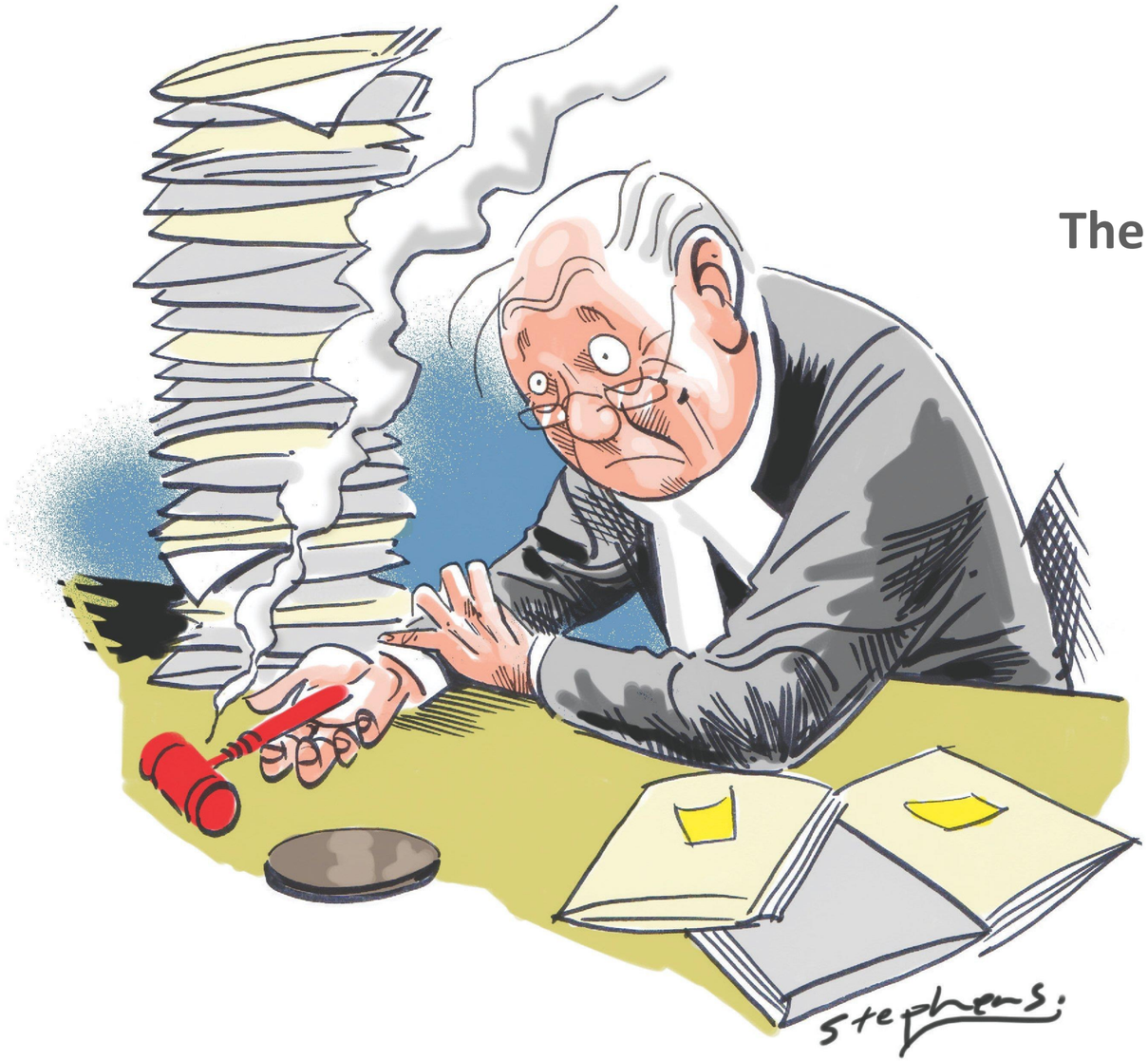
Judicial System of India



- The judicial system in India is one of the most laborious institutions in the world thanks in part to the large population and the attention to detail every case requires.
- One of the ways to support the system is the ability to streamline the references of a case to previous cases.
- Case law is the act of invoking a previous case's details and ruling to support or reject statements of the current
- Thus having a previous case ruling greatly helps any judge to drive their point across with much more weight,
- With over 7 decades of independence there is more than likely chance that there exists a case similar to any new one.

Case law, also used interchangeably with common law, is ***law that is based on precedents, that is the judicial decisions from previous cases, rather than law based on constitutions, statutes, or regulations****. Case law uses the detailed facts of a legal case that have been resolved by courts or similar tribunals.*

Wikipedia



CURRENT SCENARIO

The problem we want to address

- However, with our access we have **at least 47,000 cases in the supreme court alone** since independence. And there are **25 high courts and 672 district courts** which will have much higher case frequency as well.
- It is therefore highly unlikely that the lawyer can go through such unfathomable number of cases, let alone summarize and find which case is relevant to the one he is dealing with.
- Hence it is very valuable to a lawyer to have a quick and ready reference that can **suggest previous cases which are similar** to the one that he is currently working on.

54,000 pending cases in Supreme Court, 43,00,000 pending cases in High courts and 2,76,000,000 pending cases in subordinate courts

Statistics updated on 2021
Insights Editor

PROBLEM STATEMENT

Our goal is to *build a recommendation system that can suggest past cases that are similar to a prompt that is given or based on another case*

SELF SCRAPPED DATASET

We have scrapped the dataset ourselves and processed it to suit our needs

Extremely vast and unordered data available through public information outlet

We have streamlined the way to successfully and efficiently scrape the data from the web

MUTIPLE SUMMARISING ALGORITHMS

We have tried multiple summarizing algorithms to compare between generalized and specific summarizing models

Lots of NLP Models to summarize data, some even fine tune for legal domain

Compare and contrasting every algorithm to see how layman as well as lawyers can use

CLUSTERING APPROACHES

We have tried multiple clustering approaches to analyse how different clusters are formed

How different algorithms handle the encoded tokens of the summaries

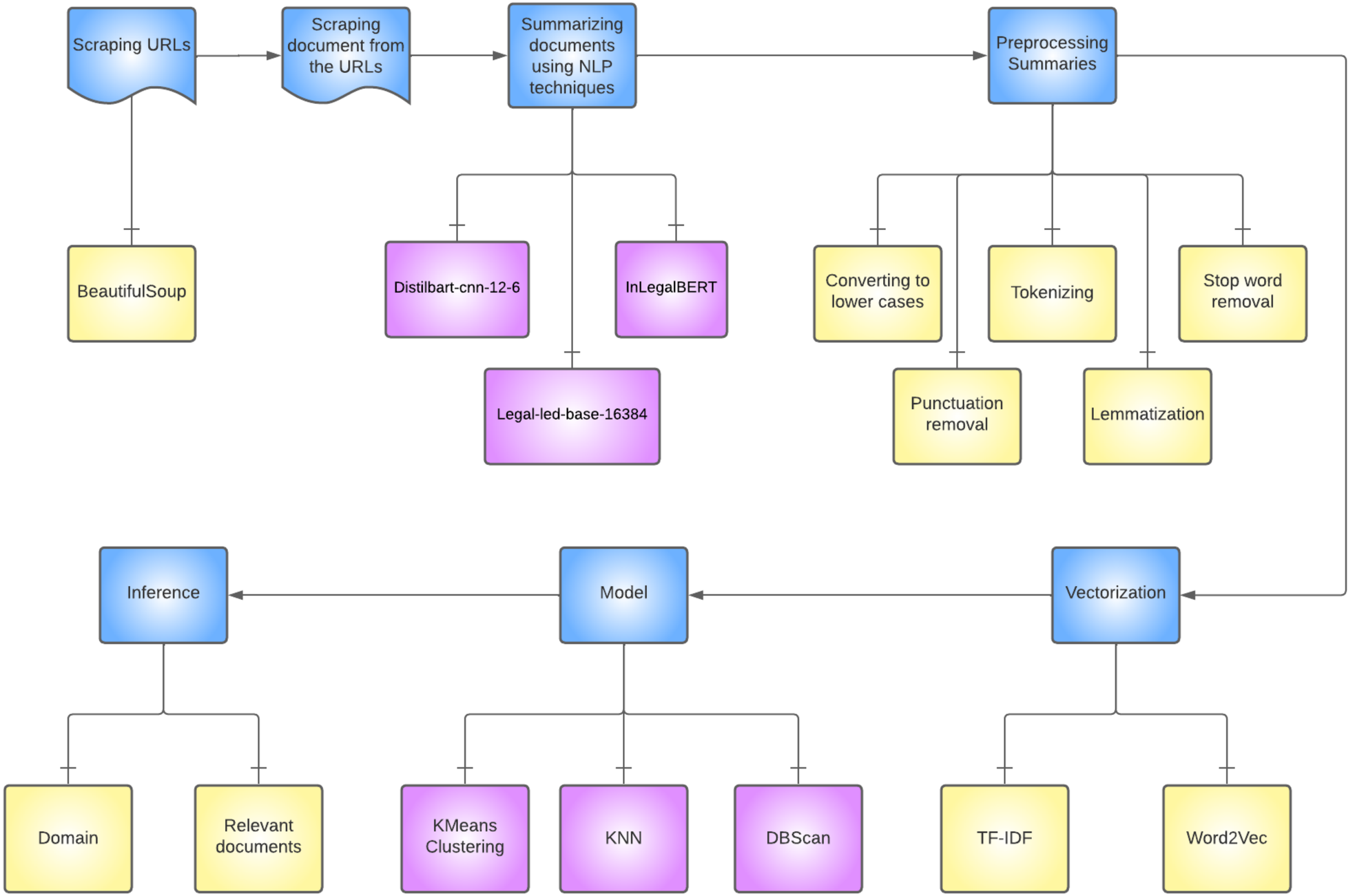
Comparing the clustering of various algorithms to see how the vectors are similar

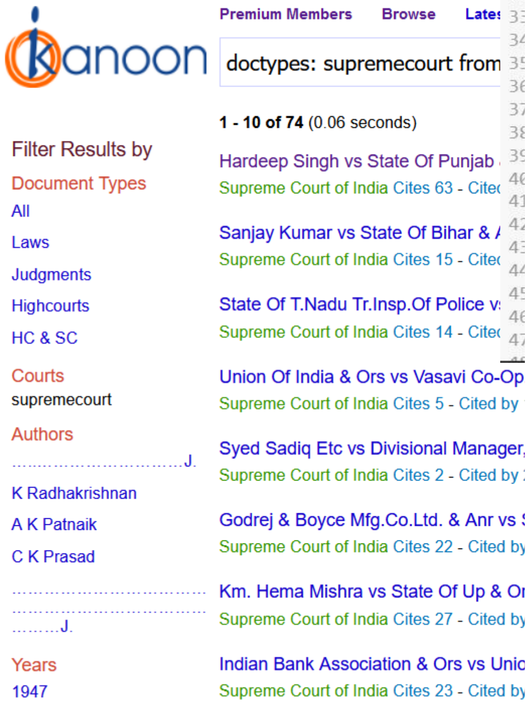
OUR MOTIVATION

- One lawyer cannot be up to date on all court cases in India.
- The number of cases will only continue to increase exponentially as the population increases
- Rather than a search on just keywords found within a case, we want to summarize and extract the content and find similar content to it.
- By trying multiple abstractive summarizing algorithms of various tune, we can try to make a layman specific or law specific pipelines as well.

OUR PIPELINE

A sneak peak of our entire projects flow






```python
import requests
from bs4 import BeautifulSoup
from transformers import pipeline
import pandas as pd

def summarize_legal_document(url):
    # Retrieve the HTML content of the legal document
    response = requests.get(url)
    html_content = response.content
    soup = BeautifulSoup(html_content, 'html.parser')
    # Extract the main text content of the legal document
    main_content = soup.find("div", {"class": "judgments"})
    if main_content is None:
        return "Error: Could not find main content"
    paragraphs = main_content.find_all('p')
    text_content = ' '.join([p.text for p in paragraphs])
    if not text_content:
        return "Error: Could not extract text content"
    document_length = len(text_content)
    if document_length < 1000:
        max_length = 100
    elif document_length < 2000:
        max_length = 150
    elif document_length < 3000:
```

Supreme Court of India

Hardeep Singh vs State Of Punjab & Ors on 10 January, 1947

Author: . B Chauhan

Bench: P Sathasivam, B.S. Chauhan, Ranjana Prakash Desai, Ranjan Gogoi, S.A. Bobde

REPORTABLE

IN THE SUPREME COURT OF INDIA

CRIMINAL APPELLATE JURISDICTION

CRIMINAL APPEAL NO. 1750 OF 2008

Hardeep Singh
…Appellant

Versus

State of Punjab & Ors.
…Respondents

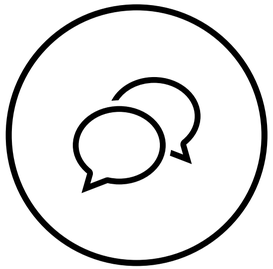
With

DATASET SCRAPING

We have scrapped the data from indiankanoon.org using Beautiful Soup

DATASET CREATION

Methodology behind the dataset that we scraped



INDIANKANOON API

With some formal communication and request, we can get access to their API to make scraping the data much more faster

SUBSET

Due to the time constraints of the summarizing algorithm we have decided to scrape only 8819 cases compared to the total 46536 cases available

RESEARCH

We ensured that the spread of cases where balanced initially but then we purposefully added imbalanced data in the form of 4000 tenant key word cases to see how the clustering will handle it

We also ensured that the source was not DDoSed by our large requests by sensibly running the scraping algorithm for a few hours at a time. This was apparent in the beginning when the scrapped content had copyright restrictions if we scarped too many at once

SUMMARIZING ALGORITHMS

Comparing between 3 summarizing models

DISTILBART CNN-12-6

Regular Summarizing algorithm

LEGAL LED BASE 16384

US SEC Trained BART

inLEGALBART

LEGAL LED Tuned for Indian Law

CHOOSE YOUR ALGORITHM

DISTILBART CNN-12-6

A simple yet efficient model to summarize a given input of text

```
text_content = '''Supreme Court of India
Municipal Council, Ratlam vs Vardichan And Ors. on 29 July, 1980
Equivalent citations: AIR 1980 SC 1622, 1980 CriLJ 1075, (1980) 4 SCC 162, 1981 1 SCR 97
Author: K Iyer
Bench: O C Reddy, V K Iyer

ORDER Krishna Iyer, J.

1. 'It is procedural rules', as this appeal proves, 'which infuse life into substantive rights, which activate them to make them effective'. Here, before us, is what

2. The circumstances of the case are typical and overflow the particular municipality and the solutions to the key questions emerging from the matrix of facts are ca

123. Duties of Council.-(1) In addition to the duties imposed upon it by or under this Act or any other enactment for the time being in force, it shall be the duty o

X X X

(b) cleansing public streets, places and sewers, and all places, not being private property, which are open to the enjoyment of the public whether such places are ve

(c) disposing of night-soil and rubbish and preparation of compost manure from night-soil and rubbish.

And yet the municipality was obvious to this obligation towards human well-being and was directly guilty of breach of duty and public nuisance and active neglect. Th

3. The Magistrate, whose activist application of Section 133 Cr.P.C., for the larger purpose of making the Ratlam municipal body to do its duty and abate the nuisanc

New Road, Ratlam, is a very important road and so many prosperous and educated persons are living on this Road. On the southern side of this Road some houses are sit
```

LAYMAN SUMMARY

The summary that it proves is very simple English and understandable even by those not well versed with law

FAST GENERATION

Fastest summarizing model within the 3 that we have selected

TRAINING

Trained on CNN Daily Mail and XSUM(BBC) datasets, which are written by journalists for summarization task.

```
 The Ratlam municipal town, like many Indian urban centers, is populous with human and sub-human species . The rich have bungal
ows and toilets, the poor live on pavements and litter the street with human excreta because they use roadsides as latrines . T
he city fathers being too busy with other issues to bother about the human condition, cesspools and stinks, dirtied the place b
eyond endurance which made the well-to-do citizens protest, but the crying demand for basic sanitation and public drains fell o
n deaf ears .
```

LEGAL LED BASE 16384

A BART based system that was fine tuned for US SEC data

HALLUCINATION

Since the model was fine tuned for US Law dataset and rulings, the model comes up with nonsense

MORE LAWYER LINGO

The summaries generated are much more in line with the language that lawyers use

GRAMATICALLY SENSIBLE

Out of all 3 models, this model is consistently more grammatically correct

TRAINING

Trained on litigations released by the US securities and exchange commission (SEC)

```
text_content = '''Supreme Court of India
Municipal Council, Ratlam vs Vardichan And Ors. on 29 July, 1980
Equivalent citations: AIR 1980 SC 1622, 1980 CriLJ 1075, (1980) 4 SCC 162, 1981 1 SCR 97
Author: K Iyer
Bench: O C Reddy, V K Iyer

ORDER Krishna Iyer, J.

1. 'It is procedural rules', as this appeal proves, 'which infuse life into substantive rights, which activate them to make them effective'. Here, before us, is what

2. The circumstances of the case are typical and overflow the particular municipality and the solutions to the key questions emerging from the matrix of facts are ca

123. Duties of Council.-(1) In addition to the duties imposed upon it by or under this Act or any other enactment for the time being in force, it shall be the duty o

X X X

(b) cleansing public streets, places and sewers, and all places, not being private property, which are open to the enjoyment of the public whether such places are ve
```

The U.S. District Court for the Western District of New York yesterday entered a final judgment against Ratlam Municipality for violating the antifraud provisions of Section 133 of the Municipal Code. According to the court's order, filed in the United States District Court in New York, Ratlam Municipal Corporation ("Municipality") has agreed to the entry of an order requiring it to construct drains and sewers at a time when it has no drainage system. The court found that the M.P. Municipality had failed to adequately deal with the issue of public nuisance. As alleged in the court order, the municipal council and its executive officers spent half a million dollars on cleaning up the street and constructing the drains by rousing the residents of the area and laying out the city's limited financial resources. Without admitting or denying the allegations of the court, the Municipal Council and the executive officers have agreed to settle the matter by consenting to a judgment that permanently enjoins them from violating Section 133 and orders them to pay disgorgement of ill-gotten gains plus prejudgment interest, a civil penalty, and an officer-and-director bar. In a parallel action, the U.K. Attorney's Office today announced criminal charges against the municipal authorities for their role in this matter. The SEC's complaint charges Ratlam with violating Sections 133 and (1) and (2) of the Securities Act of 1933 ("Securities Act") and Section 10(b) and Rule 10b-5 thereunder, and seeks permanent injunctions, civil penalties, and injunctive relief. It also seeks to bar Ratlam from serving as an officer or director of a public company, from participating in the issuance, purchase, offer, or sale of any municipal securities, and from participating as a municipal agent in any matter of public law. For further information, see Litigation Release No. 13-20-2017 (January 17, 2018). The case is being handled by Krishna Iyer, J. Shankar, and B. Garth, of the New York Regional Office, under the supervision of C.J. Kerstetter. Â The SEC appreciates the assistance of the United Kingdom Attorney General's Office, the Federal Bureau of Investigation, and the Financial Industry Regulatory Authority (FINRA).

inLEGALBART

Taking LEGAL LED Base fine tuned on US Law to fine tune on Indian Law

```
text_content = '''Supreme Court of India
Municipal Council, Ratlam vs Vardichan And Ors. on 29 July, 1980
Equivalent citations: AIR 1980 SC 1622, 1980 CriLJ 1075, (1980) 4 SCC 162, 1981 1 SCR 97
Author: K Iyer
Bench: O C Reddy, V K Iyer

ORDER Krishna Iyer, J.

1. 'It is procedural rules', as this appeal proves, 'which infuse life into substantive rights, which activate them to make them effective'. Here, before us, is what

2. The circumstances of the case are typical and overflow the particular municipality and the solutions to the key questions emerging from the matrix of facts are ca

123. Duties of Council.-(1) In addition to the duties imposed upon it by or under this Act or any other enactment for the time being in force, it shall be the duty o

X X X

(b) cleansing public streets, places and sewers, and all places, not being private property, which are open to the enjoyment of the public whether such places are ve

(c) disposing of night-soil and rubbish and preparation of compost manure from night-soil and rubbish.

And yet the municipality was obvious to this obligation towards human well-being and was directly guilty of breach of duty and public nuisance and active neglect. Th

3. The Magistrate, whose activist application of Section 133 Cr.P.C., for the larger purpose of making the Ratlam municipal body to do its duty and abate the nuisanc

New Road, Ratlam, is a very important road and so many prosperous and educated persons are living on this Road. On the southern side of this Road some houses are sit
```

The Ratlam Municipal Council, Ratlam vs Vardichan And Ors., ;, challenged the sense and soundness of the High Court's affirmation of the trial court's order directing the construction of drainage facilities and the like which had spiralled up to this Court. It was contended on behalf of the Municipality that : (1) The circumstances of the case were typical and overflow the particular municipality and the solutions to the key questions emerging from the matrix of facts are capable of universal application, especially in the Third World humanscape of silent subjection of groups of people to squalor and of callous public bodies habituated to deleterious inaction; (2) Section 133 Cr.P.C. casts a mandate: "Duties of Council" In addition to the duties imposed upon it by or under this Act or any other enactment for the time being in force, it shall be the duty of the Council to make reasonable and adequate matters within the limits of XXXXXXX(X)(b), namely, (i) cleansing public streets, places and sewers, and (ii) removing all obstructions and nuisances which are not open to public enjoyment; (iii) abating the pollution caused by the use of roads and

LAWYER LINGO

Just as Legal LED it is able to generate more formal summaries

SLOW GENERATION

Slowest model. Took multiple minutes even with CUDA

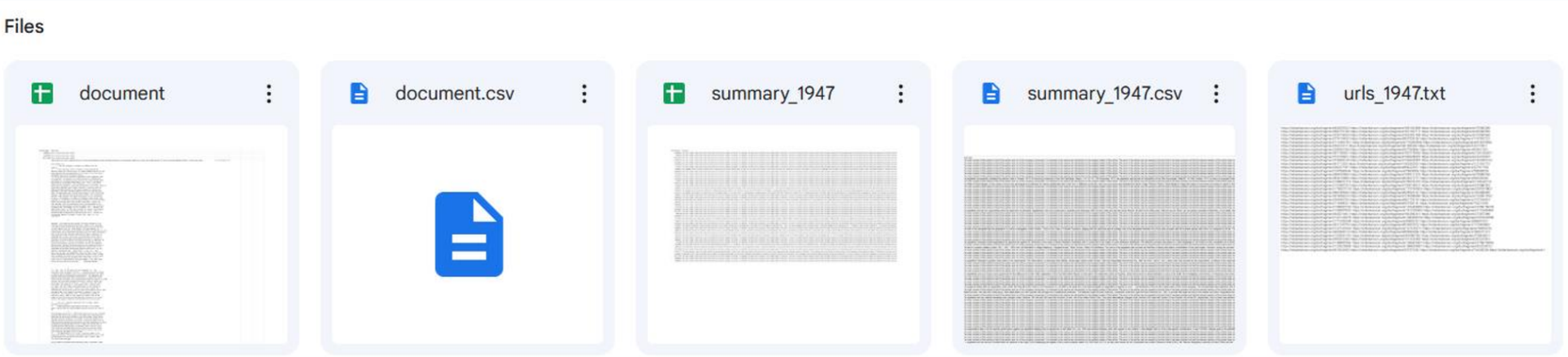
NO MORE NONSENSE

The model was very well fine tuned for Indian Law and case filings that the there are no hallucinations

CITATION

Model courtesy of IIT Kharagpur who have made it publicly available

DATASET PRE PROCESSING



LOWER CASE

Converting the flow of text all into lower case

REMOVING STOP WORDS

Removing words like I, me, etc which don't add any extra meaning to the sentence

PUNCTUATION REMOVAL

Removing comma, full stop, apostrophe, etc.

TOKENISING

Converting each word in a sentence into an entry in a list

LEMMATIZING

Grouping together multiple form of a word as a single word

VECTORIZATION

Bye Bye English and Hello vectors

```
In [75]:  1  km1.vectorized_summaries

Out[75]: <8819x17032 sparse matrix of type '<class 'numpy.float64'>'
                with 201954 stored elements in Compressed Sparse Row format>
```

```
In [62]:  1  # km1.getSummaryOfCluster(4)
```

```
self.top_words_per_cluster = []
#Vectorisation on Preprocessed data:
preprocessed_summaries = self.complete_data['PreProcessedSummary'].tolist()
self.vectorizer = TfidfVectorizer()
self.vectorized_summaries = self.vectorizer.fit_transform(preprocessed_summaries)

#Fitting Model
cluster_labels = self.kmeans.fit_predict(self.vectorized_summaries)


#Updating data
self.complete_data['cluster']=cluster_labels
```

```
28      def predict_cluster(self,new_document):
29
30          self.target_doc=new_document
31
32          new_doc_preprosum = preprocess_summary(new_document)
33
34          self.target_doc_vector = self.vectorizer.transform([new_doc_preprosum])
35          self.target_cluster = self.kmeans.predict( self.target_doc_vector)
36          print(" TARGET CLUSTER : ",self.target_cluster)
37
```

WHY VECTORS

We went from large case file to summary since the entire content of case file was too large. Now, why are we going from summary to vectors ?

VECTOR SPACE & SIMILARITY

Once we vectorize the summary we are able to plot the summary in our vector space. This then allows us to mathematically compare two vectors to identify how similar they are.

We have use 2 approaches:

WORD2VEC

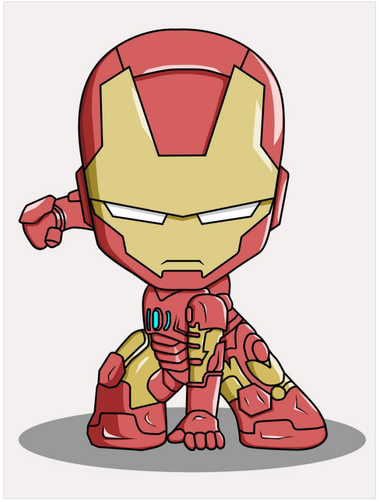
Additionally when faced with subpar performance in DBScan we tried Word 2 Vec as well

TF-IDF

We have tried it out for all our clustering algorithms

RECOMMENDATION APPROACHES

Comparing between 3 clustering and similarity algorithms

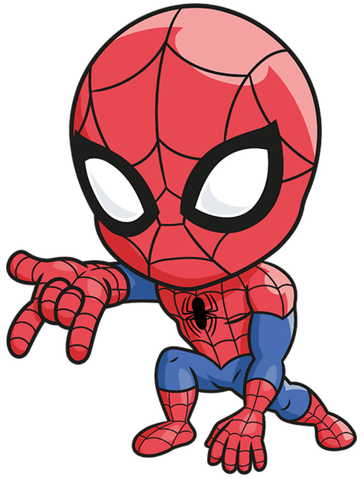


K Means Clustering

Clustering similar documents

K Nearest Neighbors

Closest similar documents



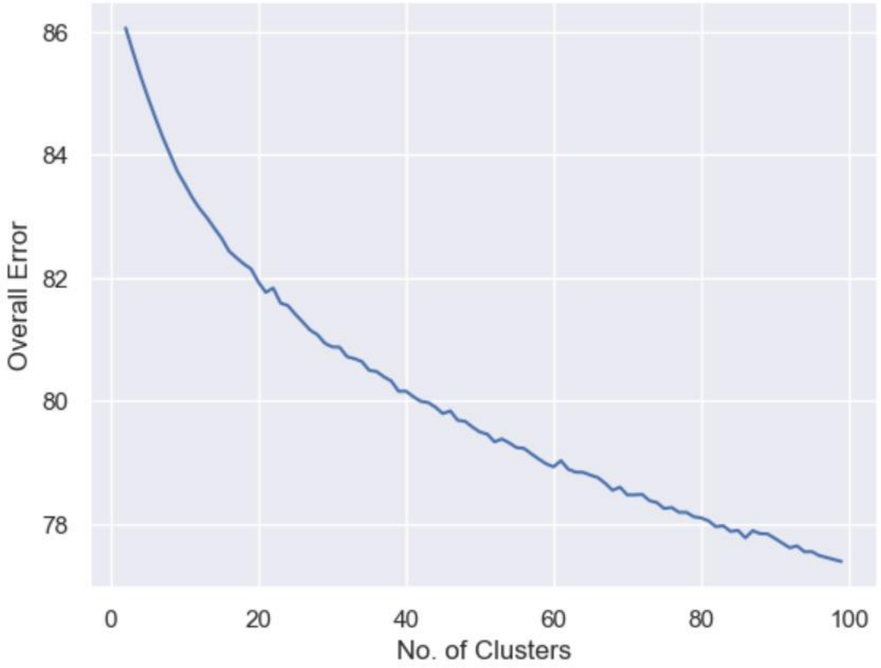
DBScan

Clustering based on density

CHOOSE YOUR ALGORITHM

K MEANS CLUSTERING

Clustering algorithm that tries to minimize intra class and maximize inter class distance




```
class KNN_Law:
    def __init__(self,data,context="Summary"):
        '''To vectorise the data.
        Context : 1) "Summary" (defualt) - Algorithm uses the summaries
                  2) "PreProcessed" - Algorithn uses Pre preocessed summaries'''
        self.summaries = data[context].tolist()
        self.complete_data = data
        self.vectorizer = TfidfVectorizer(stop_words='english')
        self.X = self.vectorizer.fit_transform(self.summaries)

    def findNN(self,no_of_neighbours, new_document,context="Summary" , what="URL"):
        '''Prints the URL of top K nearest neighbours.
        -------------------------
        Paramaters:
        Context : 1) "Summary" (defualt) - Algorithm uses the summaries
                  2) "PreProcessed" - Algorithn uses Pre preocessed summaries

        What    : 1)"URL" : (default) Prints the URL for the document
                  2)"Summary" : Prints the Summary/Pre Processed Summary based on the context
        '''
```

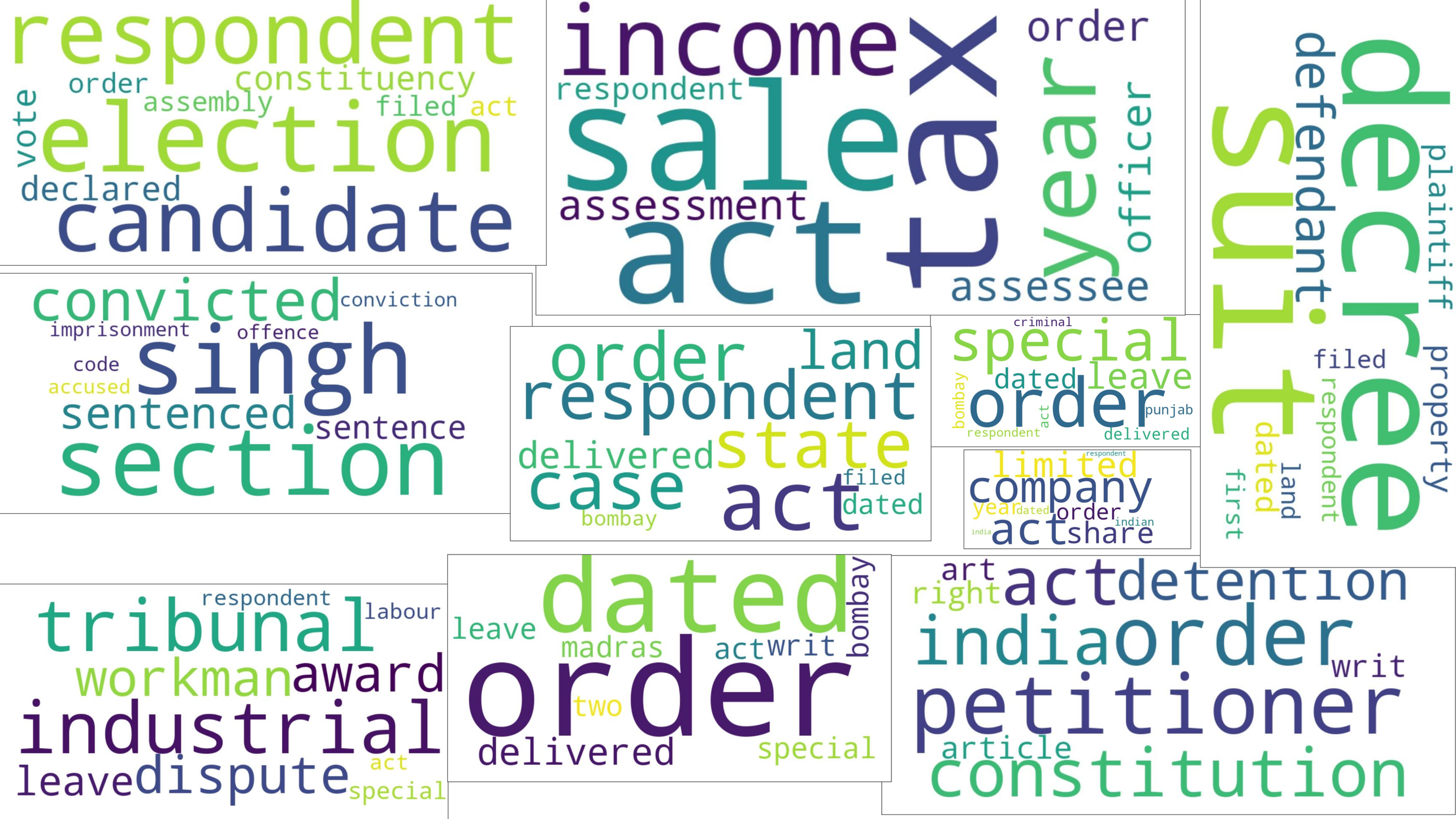
APPROACH

TF-IDF vectorization itself was giving good performance with clusters formed of similar words and cases.

We used Euclidean measure and cosine similarity to determine the closeness between 2 vectors.

We settled on using cosine similarity as it was performing better reason being pointing in a similar direction in vector space is a better representation than the distance between the vectors.

```
: 1 km1.predict_cluster(new_doc)

  TARGET CLUSTER :  [6]

: <8819x17032 sparse matrix of type '<class 'numpy.float64'>'
        with 201954 stored elements in Compressed Sparse Row format>

: 1 km1.getTopXDoc(no_of_doc=5, distance_measure='cos', what="URL")

Top 5 nearest documents in cluster 6:
https://indiankanoon.org/doc/622466
https://indiankanoon.org/doc/77493
https://indiankanoon.org/doc/134461
https://indiankanoon.org/doc/983410
https://indiankanoon.org/doc/151821
https://indiankanoon.org/doc/50761490

: 1 km1.getTopXDoc(no_of_doc=5, distance_measure='euc', what="URL")

Top 5 nearest documents in cluster 6:
https://indiankanoon.org/doc/445974
https://indiankanoon.org/doc/1027347
https://indiankanoon.org/doc/225066
https://indiankanoon.org/doc/1595917
https://indiankanoon.org/doc/241705
https://indiankanoon.org/doc/626233

: 1 km1.getTopXDoc(no_of_doc=6, distance_measure='cos', what="Summary")

Top 6 nearest documents in cluster 6:
--------------------
 Petition No. 77 of 1958 was filed under Art. 32 of the Constitution of India . Petitione
pect of their business under the provisions of the Bengal Finance (Sales Tax)   Act, as i
--------------------
 The appellant was a registered dealer under the Bihar Sales Tax Act . He was assessed to
ods . The appellant paid the principal amounts of the certificates . The Certificate Offi
e on the certificates. The appellant filed objections disputing its liability to pay inte
```


respondent
election
order
constituency
assembly
filed
act
vote
declared
candidate
income
tax
sale
act
year
order
respondent
assessment
officer
assessee
decree
suit
defendant
plaintiff
property
filed
respondent
dated
land
first
convicted
conviction
imprisonment
offence
singh
code
accused
sentenced
sentence
section
order
land
respondent
delivered
state
case
act
filed
dated
bombay
special
criminal
dated
leave
order
bombay
act
punjab
respondent
delivered
limited
respondent
company
year
dated
order
indian
india
act
share
tribunal
respondent
labour
workman
award
industrial
leave
dispute
act
special
dated
order
leave
madras
act
writ
bombay
two
delivered
special
art
act
detention
right
india
order
writ
petitioner
article
constitution

K NEAREST NEIGBOURS

Given a vector, we find the top K closest neighbors

APPRAOCH

We tried K values from 1 to 100 to try and find a suitable elbow.
However there was no definite elbow but we noticed the graph lose smoothness beyond K = 20
With some trail and error we settled for K = 10 as our optimum K

PARALLELS WITH K MEANS

In general we are getting similar recommendations compared to K Means clustering.
We also tried two approaches with K Means:

1. Vectorised matrix of summary data without preprocessing
2. Vectorised matrix of preprocessed summary data

While in K Means Clustering clearly showed better result on preprocessed data, the difference was negligible in K Nearest Neighbours with non preprocessed data giving slightly better result than preprocessed in some cases
We attribute this to the preprocessing removing certain elements of the sentence like auxiliary verbs which reduces the meaning

```python
class KNN_Law:
    def __init__(self,data,context="Summary"):
        '''To vectorise the data.
        Context : 1) "Summary" (defualt) - Algorithm uses the summaries
                  2) "PreProcessed" - Algorithn uses Pre preocessed summaries'''
        self.summaries = data[context].tolist()
        self.complete_data = data
        self.vectorizer = TfidfVectorizer(stop_words='english')
        self.X = self.vectorizer.fit_transform(self.summaries)

    def findNN(self,no_of_neighbours, new_document,context="Summary" , what="URL"):
        '''Prints the URL of top K nearest neighbours.
        -------------------------
        Paramaters:
        Context : 1) "Summary" (defualt) - Algorithm uses the summaries
                  2) "PreProcessed" - Algorithn uses Pre preocessed summaries

        What    : 1)"URL" : (default) Prints the URL for the document
                  2)"Summary" : Prints the Summary/Pre Processed Summary based on the context
        '''
```

```python
knnObj_Summary.findNN(no_of_neighbours=5,new_document=new_doc,context="Summary",what="URL")
```

```
The 5 nearest summaries to 'He did not pay tax' are:
https://indiankanoon.org/doc/35670
https://indiankanoon.org/doc/1378734
https://indiankanoon.org/doc/151821
https://indiankanoon.org/doc/1219983
https://indiankanoon.org/doc/1731367
```

```python
knnObj_PreProSum.findNN(no_of_neighbours=5,new_document=new_doc,context="PreProcessedSummary",what="PreProcessedSummary")
```

```
LOG : Summary PreProcessed
The 5 nearest summaries to 'pay tax' are:
-------------------------
registered dealer bihar sale tax act assessed pay sale tax four different period paid principal amount certificate certificate officer claimed payment interest due certificate filed objection disputing liability pay interest
-------------------------
filed art constitution india petitioner liable pay sale tax respect business provision bengal finance sale tax act force delhi
-------------------------
vaidyanatha aiyer found guilty paying bribe underestimating income assessed along along officer coimbatore since notice issued march read act failed pay advance tax year discovered pay advance tax found paid bribe r income tax officer
-------------------------
small cause ahmedabad ordered appellant pay rent municipal tax appellant also agreed pay municipal tax electricity charge respondent obtained order issue distress warrant recovery amount due municipal tax distress levied file order confirmed
-------------------------
mysore writ brought certificate behalf income tax officer mangalore tax imposed income tax act called act bearing respectively
```

DBSCAN

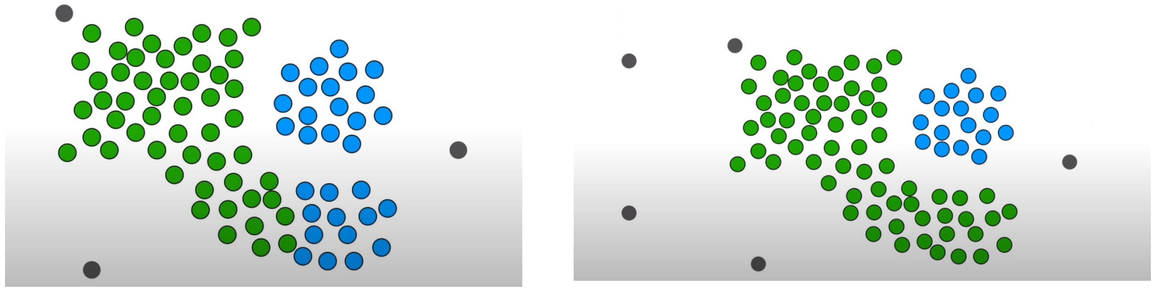
Clustering algorithm that tries to create clusters based on the density

In [136]:

```
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.cluster import DBSCAN
from sklearn.neighbors import NearestNeighbors


summaries = new_data['PreProcessedSummary'].tolist()
# create a TfidfVectorizer object to convert summaries into feature vectors
vectorizer = TfidfVectorizer(stop_words='english')
X = vectorizer.fit_transform(summaries)

dbscan = DBSCAN(eps=0.5, min_samples=2)
dbscan.fit(X)

new_doc_summ = vectorizer.transform([new_doc])
new_doc_summ

```

Out[136]:

```
<1x16864 sparse matrix of type '<class 'numpy.float64'>'
        with 0 stored elements in Compressed Sparse Row format>
```



APPRAOCH

We tried eps values from 0.01 to 2 to try and find a suitable clusters

The eps varied based on what vectorisation algorithm that we are using. Word 2 Vec requires very small eps (0.00001) while TF-IDF requires larger eps values (0.7)

However, even with multiple attempts at tuning the model we get sub par clusters. Almost all the data get clustered in one or two clusters

The expected reasoning is that there are common words within the summary that is confusing the density algorithm requiring much more fine tuned eps values

The possible ways to improve are:

- Preprocessing your data
- Changing the feature representation
- Tuning other hyper parameters
- Trying a different clustering algorithm

KEY OBSERVATIONS

What we took away from the project

PRE PROCESSING

K Nearest Neighbors working better on non preprocessed data when dealing with aux verbs

```
In [183]: 1 knnObj_PreProSum.findNN(no_of_neighbours=5,new_document=new_doc,context="PreProcessedSummary",what="PreProcessedSummary")

LOG : Summary PreProcessed
The 5 nearest summaries to 'pay tax' are:
-----------------------
registered dealer bihar sale tax act assessed pay sale tax four different period paid principal amount certificate certificate
officer claimed payment interest due certificate filed objection disputing liability pay interest
-----------------------
filed art constitution india petitioner liable pay sale tax respect business provision bengal finance sale tax act force delhi
-----------------------
vaidyanatha aiyer found guilty paying bribe underestimating income assessed along along officer coimbatore since notice issued
march read act failed pay advance tax year discovered pay advance tax found paid bribe r income tax officer
-----------------------
small cause ahmedabad ordered appellant pay rent municipal tax appellant also agreed pay municipal tax electricity charge respo
ndent obtained order issue distress warrant recovery amount due municipal tax distress levied file order confirmed
-----------------------
mysore writ brought certificate behalf income tax officer mangalore tax imposed income tax act called act bearing respectively
total amount r made tax r

In [184]: 1 knnObj_Summary.findNN(no_of_neighbours=5,new_document=new_doc,context="Summary",what="Summary")

The 5 nearest summaries to 'He did not pay tax' are:
-----------------------
 The High Court of Gujarat dismissed summarily the applicant's application for revision of the judgment of the Principal Judge
of the City Civil Court, Ahmedabad . The appellant did not pay rent from June 1, 1956 for a period of over six months, in conse
quence of which the respondent gave a notice to him on February 20, 1957 terminating his tenancy . As the appellant did neither
vacate the premises, nor pay the arrears due from him, the respondent instituted a suit .
-----------------------
 Ex-Zamindar was assessed to agricultural income-tax in the assessment year 1360 F. corresponding to 1952-53 . He did not pay t
he assessed tax and was further assessed to a penalty . The High Court held that orders of the Agricultural Income-tax Assessin
g Officer and the Collector were wrong as the ground for refusing to accept the bonds .
-----------------------
 Vaidyanatha Aiyer was found guilty of paying a bribe for underestimating his income . He had been assessed to income-tax all a
long along with the Income-Tax Officer of Coimbatore since 1942 . A notice was issued to him on March 24, 1951 under s. 28 read
with s. 18-A (2) of the  Income-tax Act . He failed to pay advance tax for the year 1950-51 it was discovered that he didn't pa
y advance taxes . He was found to have paid a bribe of Rs. 1,000 to the Income Tax Officer .
-----------------------
 The Punjab High Court acquitted the respondent in the case . The High Court held that the credit sales were not sales not ille
gal . The respondent purchased goods worth Rs.2,876.20 on credit from cloth merchants as well as tailors . It was urged before
the High Court that when a person obtained goods on credit he did not obtain them without consideration and assumed that be did
not really intend to pay, even when he promised to pay .
```

KEY OBSERVATIONS

What we took away from the project

```
[31]:
1 new_doc = "The tenant fought with him" #Better than KNN | KNN mapped the fight keyword but KMeans mapped it to the tenant cl
2 km1.getTopXDoc(no_of_doc=2, distance_measure='cos', what="Summary")
```

```
Top 2 nearest documents in cluster 8:
----------
 Civil Appeal No. 389 of 1966 was heard in an appeal by a tenant in Delhi . The tenant had been evicted under the Delhi and Ajm
er Rent Control Act of 1952 . The landlord was ordered to pay the tenant Rs. 145 as arrears of rent . The High Court reversed i
ts earlier order and ordered eviction of the tenant .
----------
 The Bombay High Court granted an appeal against eviction of a tenant in a house in Sholapur, Maharashtra . The tenant had fail
ed to pay the rent on the 20th of each of the years  1951-52 and 1953-54 . The landlords had filed a .suit for recovery of the
rent and the tenant had . paid the tenant after his appeal against the decree passed against him was disposed of on June 8, 195
6 . The landlord received the rent in April 1952-53 . The appeal was dismissed on the ground that the tenant paid up the rent d
ue by him and there being no arrears at the .time of the application the appellants were,
----------
 An appeal by special leave against the judgment of the Bombay High Court in Special Application No. 2258 of 1955 . The appella
nt is the landlord and the respondent a protected tenant . The tenant objected to the right of the landlord to terminate the te
nancy of the tenant .
```

```
----------
directed dismissal allahabad election filed fought election jan sangh party ticket returned candidate fought congress party tic
ket election filed three ground publication two pamphlet ex contained false statement relating personal character conduct withi
n meaning representation people act
----------
heard tenant delhi tenant evicted delhi ajmer rent control act landlord ordered pay tenant r arrears rent reversed earlier orde
r ordered eviction tenant
```

```
:
1 knnObj_Summary.findNN(no_of_neighbours=2,new_document=new_doc,context="Summary",what="Summary")
```

```
The 2 nearest summaries to 'The tenant fought with him' are:
----------
 Appeal was directed against the dismissal by the High Court of Allahabad of the election petition filed by the appellant . The
appellant fought the election on the Jan Sangh party ticket, while the returned candidate fought it on the Congress party ticke
t . The election petition was filed on three grounds : (1) the publication of two pamphlets (Ex P-8 and P-9) contained false st
atements relating to his personal character and conduct within the meaning of the Representation of the People Act, 1951 .
----------
 Civil Appeal No. 389 of 1966 was heard in an appeal by a tenant in Delhi . The tenant had been evicted under the Delhi and Ajm
er Rent Control Act of 1952 . The landlord was ordered to pay the tenant Rs. 145 as arrears of rent . The High Court reversed i
ts earlier order and ordered eviction of the tenant .
```

```
Top 10 words for Cluster 8:
land (38.45)
act (25.69)
high (23.99)
tenant (20.83)
respondent (19.15)
rent (15.10)
landlord (13.56)
bombay (12.85)
acquisition (12.78)
order (12.70)
```

K MEANS BETTER THAN KNN for some examples

Working within a cluster means it gets within the cluster than getting nearest

KEY OBSERVATIONS

What we took away from the project

```
Top 10 words for Cluster 15:
wife (1.07)
husband (0.55)
high (0.51)
married (0.47)
decree (0.47)
filed (0.43)
ground (0.40)
separation (0.40)
judicial (0.40)
first (0.38)
```

DBScan IDENTIFIES OUTLIER CLUSTERS

Since this is density based it is able to group sparsely dense vectors much better than other algorithms

FUTURE OPTIMIZATIONS

What we could do differently to make things better

GIVEN ACCESS TO API

We get quicker and better format of words from indiankannon.org

USING INLEGAL BERT

With access to a CUDA Nvidia Desktop GPU for at least 100 Hrs

USING TRANSFORMERS

For vectorisation of the summary

WITH UNION OF RECOMMENDATION OF K MEANS & KNN

Sorting them as first from both, then second from both and so on

WE SHOULD GET AN ACCEPTABLE LEVEL OF PERFORMNACE

THANK YOU

MEET MANDHANE
MT2022061

NARASIMHAN N
MT2022062

JACOB MATHEW
MT2022150