

# UNDERSTANDING THE DATA

- This table represents the number of districts for employment type vs their pace of growth.
- e.g. The value 5 represents the Fast growing, Non-Agri districts.
- Here, for pace of growth we have used the 2019 prediction from ADI data.
- ADI Predicted values: 0-1 = Slow, 2 = Average, 3-4 = Fast

Number of Districts

	SLOW	AVG	FAST
UNEMP	124	75	29
AGRI	145	72	12
N-AGRI	99	32	5

This column represents the total number of unique articles for each collection.

Number of Articles for each subclass in each collection

COLLECTIONS	TOTAL NUMBER OF ARTICLES	UNEMP DISTRICTS			AGRI DISTRICTS			NON AGRI DISTRICTS		
		Slow (124)	Avg (75)	Fast (29)	Slow (145)	Avg (72)	Fast (12)	Slow (99)	Avg (32)	Fast (5)
AGRICULTURE	80221	7506	5155	4884	14513	5429	735	53840	8955	443
DEVELOPMENT	15984	1426	848	1034	2077	642	124	11936	1571	137
ENVIRONMENT	100038	8720	5704	4711	12141	3885	650	75165	9444	1261
INDUSTRIALIZATION	111291	6466	4017	3898	8379	3002	325	94602	9001	1126
LIFESTYLE	234158	18829	12904	15173	19616	6198	925	187286	19622	2017

These values are also represented in this table, as indicated by the arrow, for better understanding.

Note: Here the sum of articles in each subclass of a collection is more than the total number of articles, because an article can have multiple locations.

Here, 9001 means that there are 9001 number of articles for Average growing Non Agri Districts, for the industrialization collection.

# CREATING THE DATASETS

## Collections:

- Agriculture
- Development
- Environment
- Industrialization
- Lifestyle

## In these Collections, for each article we have:

- Article\_Id
- Article\_Title
- Article\_Text
- Article\_Date
- Location\_names

**Mapping of 2011 districts to 2001 districts**  
(640 to 593 districts)

**Mapping of Location Names to District Ids**

**Mapping of District Ids to their  
corresponding Employment Labels**

**Pace of Growth:**  
**2019 Change Predictions based on ADI**  
**0-1 : Slow, 2 : Average, 3-4 : Fast**

**Mapping of District Ids to their  
corresponding Industry Type**

## Dataset for each collection:

For each collection we form the following dataset and it contains the following fields:

1. **Article\_Id**
2. **Article\_Title**
3. **Article\_Text**
4. **Processed Text \***
5. **Article\_Date**
6. **District\_Id**
7. **Employment Type**
8. **Pace of growth**
9. **Industry Type**

Now in our dataset, Article\_Id and District\_Id together can form the primary key.

\* **Note:** To obtain the Processed Text, we have concatenated the article title and article text, and then applied the **stop-word removal, entity blinding, and stemming**.

# TAGGED DOCUMENTS

For each dataset corresponding to each collection we do the following:

## Dataset

## Tagged Documents

DATA	TAGS
Processed Text	Article_Id District_Id Employment Type Pace of growth

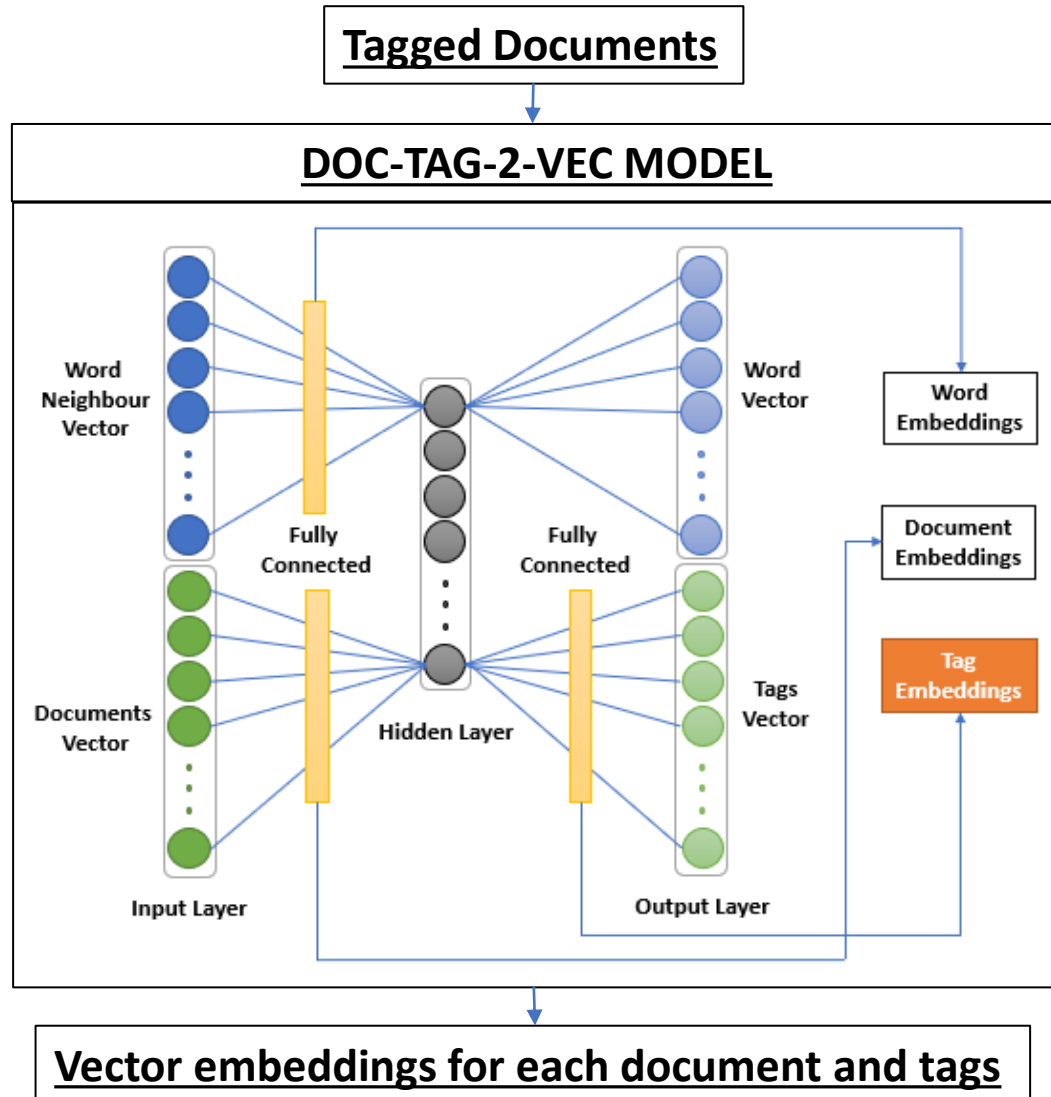
Here, for each article we provide the processed text as data, and Article\_Id, District\_Id, Employment Type, Pace of growth as tags.

## DOC-TAG-2-VEC MODEL

**Note:** At this point we have both datasets, as well as Tagged documents. Tagged Documents are used to train the DT2V Models, while datasets are used for further analysis.

# TRAINING DOC-TAG-2-VEC MODEL

For each tagged document corresponding to each collection we do the following:

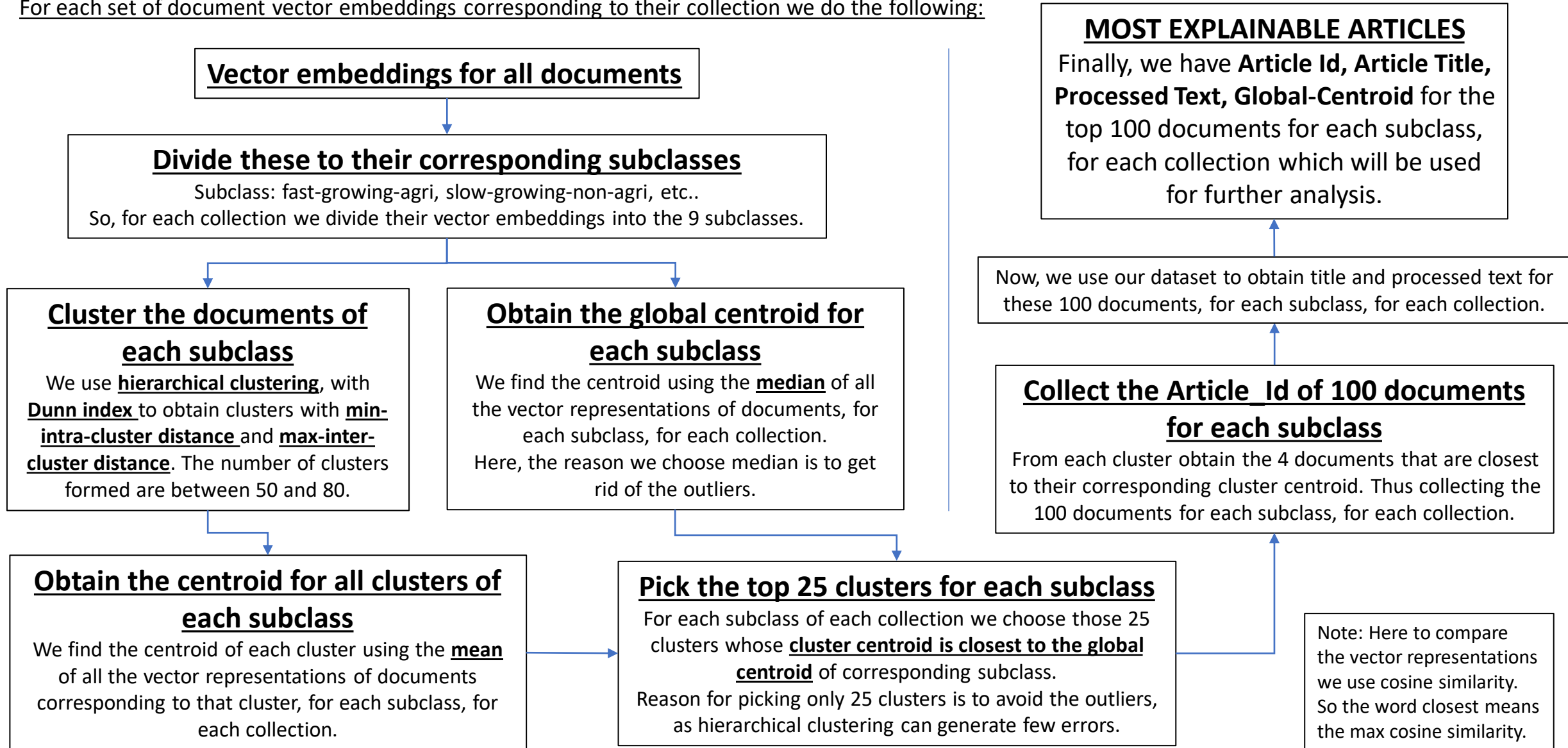


## Some points to note:

1. **Parameters** for training the model:
  - Epochs: 100
  - Learning rate: 0.1
  - Minimum learning rate: 0.001
  - Vector size: 50
  - Min occurrences for a word: 3
  - Distributed bag of words (D-BOW)
  - Window size: 3
2. In our model, the learning rate linearly falls down till the value of min-learning-rate.
3. Each article is considered as a separate document, along with appropriate tags.
4. Tags used: Article Id, District Id, Employment type, Pace of growth.
5. Here, we are getting the both the direct document embeddings as well as tag embeddings for each document. We are ignoring the direct document embeddings and are using the **tag embeddings** for each document.

# OBTAINING THE MOST EXPLAINABLE ARTICLES

For each set of document vector embeddings corresponding to their collection we do the following:



# ANALYSIS

For each collection we do the following:

## OBTAIN THE MOST EXPLAINABLE ARTICLES

Now, we have **Article Id, Article Title, Processed Text, Global-Centroid** for the top 100 documents for each subclass, of each collection.

## Similarity Matrix

We use cosine-similarity to obtain the similarity between the global centroids of all the subclasses, and form a 9x9 matrix. We do this for all the collections.

A global centroid of a subclass can be thought as a point in 50 dimensional vector space, which represents, that particular subclass.

## LIST OF PATTERNS

Using the tables shown on the first slide of this presentation, we obtain a list of patterns for which we would like to study or analyze the data.

## Keywords Extraction

We find out the **100 most explainable keywords** using **TF-IDF** on the **Processed Text**, and do this for all the subclasses of all the collections.

## Analysis using keywords

We use these keywords belonging to different subclasses to do the following:

1. Obtain those words that occur for all the subclasses of a collection.
2. Obtain those words that are unique to a particular subclass(i.e. the words that occur in a particular subclass but not in any other subclass).
3. For a pattern (subclass-A vs subclass-B) we find out those words that occur for both these subclasses (i.e. both A,B).
4. For a pattern (subclass-A vs subclass-B) we find out those words that occur only for subclass A, and only for subclass B.

We do this for all the collections.

## FINAL RESULTS

### ANALYSIS ACROSS PATTERNS

Finally we perform our analysis, using the similarity matrix, and keywords results, across all the patterns and for all the collections.

# ANALYSIS

## Initial List of Patterns:

1. Unemployment districts that are slow growing  
(vs) Unemployment districts that are fast growing.
2. Agriculture districts that are slow growing  
(vs) Agriculture districts that are fast growing.
3. Non Agri districts that are slow growing  
(vs) Non Agri districts that are fast growing.
4. Unemployment districts that are slow growing  
(vs) Agriculture districts that are fast growing.
5. Agriculture districts that are slow growing  
(vs) Non-Agri districts that are fast growing.
6. Unemployment districts that are slow growing  
(vs) Non-Agri districts that are fast growing.

## Some general analysis using keywords:

- For Agriculture collection, the following words: '**monsoon**', '**rain**', '**water**', '**farmer**', '**village**', '**land**', '**rainfall**', '**crop**' do come up on top **for all the subclasses**. On reading some of the articles manually, we see that there are a lot of talks regarding the weather, crops and farmers.
- For Development collection, the following words: '**project**', '**scheme**', '**develop**', '**central**', '**program**' do come up on top **for all the subclasses**. On reading some of the articles manually, we see that there are a lot of articles about the central government policies, schemes and projects.
- For Industrialization collection, the following words: '**project**', '**develop**', '**power**' do come up on top **for all the subclasses**. On reading some of the articles manually, we see that there are a quite some articles about the electricity, and ways of generating electricity from different sources.
- For Lifestyle collection, the following words: '**health**', '**hospital**', '**policies**' do come up on top **for all the subclasses**. On reading some of the articles manually, we see that there are many articles about the hospitals and human-health.

# SIMILARITY MATRICES

- We use cosine-similarity to obtain the similarity between the global centroids of all the subclasses, and form a 9x9 matrix. We do this for all the collections.
- A global centroid of a subclass can be thought as a point in 50 dimensional vector space, which represents, that particular subclass.
- Here, the value 0.63 represents the cosine similarity between the slow-growing Unemp districts and fast-growing Unemp districts.

## AGRICULTURE COLLECTION

	UNEMP_SLOW	UNEMP_AVG	UNEMP_FAST	AGRI_SLOW	AGRI_AVG	AGRI_FAST	NON_AGRI_SLOW	NON_AGRI_AVG	NON_AGRI_FAST
UNEMP_SLOW	1	0.83	0.63	0.4	0.29	0.42	-0.63	0.06	0.41
UNEMP_AVG	0.83	1	0.75	0.16	0.06	0.29	-0.49	-0.01	0.45
UNEMP_FAST	0.63	0.75	1	-0.2	-0.23	0.07	-0.13	-0.2	0.45
AGRI_SLOW	0.4	0.16	-0.2	1	0.56	0.35	-0.72	0.07	-0.34
AGRI_AVG	0.29	0.06	-0.23	0.56	1	0.85	-0.51	0.45	-0.27
AGRI_FAST	0.42	0.29	0.07	0.35	0.85	1	-0.47	0.42	0
NON_AGRI_SLOW	-0.63	-0.49	-0.13	-0.72	-0.51	-0.47	1	-0.34	0.05
NON_AGRI_AVG	0.06	-0.01	-0.2	0.07	0.45	0.42	-0.34	1	0.18
NON_AGRI_FAST	0.41	0.45	0.45	-0.34	-0.27	0	0.05	0.18	1

# SIMILARITY MATRICES

## DEVELOPMENT COLLECTION

	UNEMP_SLOW	UNEMP_AVG	UNEMP_FAST	AGRI_SLOW	AGRI_AVG	AGRI_FAST	NON_AGRI_SLOW	NON_AGRI_AVG	NON_AGRI_FAST
UNEMP_SLOW	1	0.83	0.65	0.63	0.59	0.45	-0.71	0.35	0.07
UNEMP_AVG	0.83	1	0.78	0.45	0.47	0.4	-0.62	0.28	-0.04
UNEMP_FAST	0.65	0.78	1	0.09	0.14	0.12	-0.35	0.08	-0.05
AGRI_SLOW	0.63	0.45	0.09	1	0.88	0.22	-0.83	0.38	-0.21
AGRI_AVG	0.59	0.47	0.14	0.88	1	0.4	-0.77	0.53	-0.2
AGRI_FAST	0.45	0.4	0.12	0.22	0.4	1	-0.18	0.49	0.37
NON_AGRI_SLOW	-0.71	-0.62	-0.35	-0.83	-0.77	-0.18	1	-0.47	0.14
NON_AGRI_AVG	0.35	0.28	0.08	0.38	0.53	0.49	-0.47	1	0.42
NON_AGRI_FAST	0.07	-0.04	-0.05	-0.21	-0.2	0.37	0.14	0.42	1

## ENVIRONMENT COLLECTION

	UNEMP_SLOW	UNEMP_AVG	UNEMP_FAST	AGRI_SLOW	AGRI_AVG	AGRI_FAST	NON_AGRI_SLOW	NON_AGRI_AVG	NON_AGRI_FAST
UNEMP_SLOW	1	0.69	0.41	0.65	0.54	0.47	-0.84	-0.03	-0.05
UNEMP_AVG	0.69	1	0.81	0.19	0.21	0.18	-0.58	0.14	0.09
UNEMP_FAST	0.41	0.81	1	-0.1	-0.06	0.07	-0.31	0.25	0.13
AGRI_SLOW	0.65	0.19	-0.1	1	0.77	0.51	-0.75	-0.01	-0.29
AGRI_AVG	0.54	0.21	-0.06	0.77	1	0.83	-0.62	0.15	-0.13
AGRI_FAST	0.47	0.18	0.07	0.51	0.83	1	-0.48	0.12	-0.09
NON_AGRI_SLOW	-0.84	-0.58	-0.31	-0.75	-0.62	-0.48	1	-0.16	0.14
NON_AGRI_AVG	-0.03	0.14	0.25	-0.01	0.15	0.12	-0.16	1	0.53
NON_AGRI_FAST	-0.05	0.09	0.13	-0.29	-0.13	-0.09	0.14	0.53	1



# SIMILARITY MATRICES

## INDUSTRIALIZATION COLLECTION

	UNEMP_SLOW	UNEMP_AVG	UNEMP_FAST	AGRI_SLOW	AGRI_AVG	AGRI_FAST	NON_AGRI_SLOW	NON_AGRI_AVG	NON_AGRI_FAST
UNEMP_SLOW	1	0.89	0.77	0.73	0.75	0.79	-0.83	0.51	0.19
UNEMP_AVG	0.89	1	0.86	0.52	0.63	0.73	-0.66	0.39	0.13
UNEMP_FAST	0.77	0.86	1	0.41	0.42	0.55	-0.56	0.52	0.19
AGRI_SLOW	0.73	0.52	0.41	1	0.82	0.57	-0.74	0.38	0.01
AGRI_AVG	0.75	0.63	0.42	0.82	1	0.77	-0.76	0.32	0.05
AGRI_FAST	0.79	0.73	0.55	0.57	0.77	1	-0.67	0.44	0.28
NON_AGRI_SLOW	-0.83	-0.66	-0.56	-0.74	-0.76	-0.67	1	-0.59	-0.09
NON_AGRI_AVG	0.51	0.39	0.52	0.38	0.32	0.44	-0.59	1	0.46
NON_AGRI_FAST	0.19	0.13	0.19	0.01	0.05	0.28	-0.09	0.46	1

## LIFESTYLE COLLECTION

	UNEMP_SLOW	UNEMP_AVG	UNEMP_FAST	AGRI_SLOW	AGRI_AVG	AGRI_FAST	NON_AGRI_SLOW	NON_AGRI_AVG	NON_AGRI_FAST
UNEMP_SLOW	1	0.88	0.57	0.46	0.53	0.56	-0.68	0.13	0.06
UNEMP_AVG	0.88	1	0.79	0.39	0.4	0.53	-0.69	0.19	0.11
UNEMP_FAST	0.57	0.79	1	0.13	0.12	0.39	-0.55	0.11	0.12
AGRI_SLOW	0.46	0.39	0.13	1	0.78	0.53	-0.62	0.31	-0.12
AGRI_AVG	0.53	0.4	0.12	0.78	1	0.83	-0.61	0.49	-0.1
AGRI_FAST	0.56	0.53	0.39	0.53	0.83	1	-0.51	0.47	0.04
NON_AGRI_SLOW	-0.68	-0.69	-0.55	-0.62	-0.61	-0.51	1	-0.42	-0.07
NON_AGRI_AVG	0.13	0.19	0.11	0.31	0.49	0.47	-0.42	1	0.48
NON_AGRI_FAST	0.06	0.11	0.12	-0.12	-0.1	0.04	-0.07	0.48	1

## **PATTERN-1: SLOW GROWING UNEMP DISTRICTS vs FAST GROWING UNEMP DISTRICTS**

<b>COLLECTION</b>	<b>GLOBAL CENTROID SIMILARITY</b>	<b>KEYWORDS THAT OCCUR IN BOTH SLOW-GROWING AND FAST-GROWING UNEMP DISTRICTS</b>	<b>KEYWORDS THAT OCCUR MUCH MORE OFTEN IN SLOW-GROWING UNEMP DISTRICTS</b>	<b>KEYWORDS THAT OCCUR MUCH MORE OFTEN IN FAST-GROWING UNEMP DISTRICTS</b>	<b>COMMENTS</b>
<b>AGRICULTURE</b>	0.63	loan, rainfall, suicide, well, agriculture, dam, rain, farmer, protest, monsoon, crop, hit, season, price, sow, collector, cultivation, water, drought, mm, bank, loss	kill, sugar, rabi, sugarcane, insurance, shower, waiver, factories, debt, tanker, storage	reservoir, tenant, field, tank, power, remain, river, canal, market, cattle, inspector, cotton, engine, kharif, onion	<ul style="list-style-type: none"> <li>• In both subclasses, there are a lot of talk about rainfall, climate changes, farmer protests, suicides and bank loans.</li> <li>• In slow-growing Unemp districts there are much more articles about cultivation of sugarcane, sugar, rabi crops, tanker, storage facility, and loan waivers.</li> <li>• In fast-growing Unemp districts there are more articles about canals, reservoirs, electricity, cotton, and kharif crops cultivation.</li> </ul>
<b>DEVELOPMENT</b>	0.65	worker, scheme, construction, agriculture, irrigation, drinking water, fodder, given, mgnrega, farmer, job, land, crop, panchayat, employ, policy, welfare, implement, farming, wage, water, drought, bank, labour	pond, fish, aadhaar, municipality, waste, relief, sabha, subsidy, forest, accord, account, current, set, poultry	nrega, school, death, tank, well, borewell, cattle, milk, fund, hospital, electricity, road, women, inspector, revenue, migration, target	<ul style="list-style-type: none"> <li>• In both subclasses, there are a lot of talk about schemes regarding farmers, workers, agriculture, irrigation, wages.</li> <li>• In slow-growing Unemp districts, articles are more about relief funds, subsidies, poultry farming and fishing.</li> <li>• In fast-growing Unemp districts, articles talk more about cattle, women, schools, migration, electricity, road, and hospitals.</li> </ul>

## **PATTERN-1: SLOW GROWING UNEMP DISTRICTS vs FAST GROWING UNEMP DISTRICTS**

<b>COLLECTION</b>	<b>GLOBAL CENTROID SIMILARITY</b>	<b>KEYWORDS THAT OCCUR IN BOTH SLOW-GROWING AND FAST-GROWING UNEMP DISTRICTS</b>	<b>KEYWORDS THAT OCCUR MUCH MORE OFTEN IN SLOW-GROWING UNEMP DISTRICTS</b>	<b>KEYWORDS THAT OCCUR MUCH MORE OFTEN IN FAST-GROWING UNEMP DISTRICTS</b>	<b>COMMENTS</b>
<b>ENVIRONMENT</b>	0.41	kill, death, poach, irrigation, reservoir, farmer, tribal, court, maoist, arrest, forest, policy, conserve, project, spot, wildlife, tiger, water, investigation, elephant, village	illegal, tree, cub, field, grant, export, power, well, trap, dam, wild, rain, chinkara, ore, honey, encroach, peacock, mine, bail, carcass, tourist, tourism, record, badger, sanctuary, skin	lake, herd, fish, skeleton, naxal, monkey, camp, waterfall, visitor, murder, tusker, women, attack, rape, smuggler, tribe, seize, gun, past, smuggle, adivasi	<ul style="list-style-type: none"> <li>• In both subclasses, there are a lot of talk about irrigation, tribal and maoist activities, and some illegal activities.</li> <li>• In slow-growing Unemp districts there are much more articles about mining, honey, death of wildlife animals, tourism in wildlife sanctuaries.</li> <li>• In fast-growing Unemp districts there are more articles about Adivasi, naxal activities, rape, and other crimes.</li> </ul>
<b>INDUSTRIALIZATION</b>	0.77	illegal, worker, power, scheme, river, construct, toilet, process, farmer, protest, land, panchayat, court, forest, mine, water, sand, miner, acre, coal, transport	handloom, die, cave, ore, mill, textile, weaver, arrest, women, dead, campaign, investigation	agriculture, reservoir, rain, fertile, sabha, urea, bank, health	<ul style="list-style-type: none"> <li>• In both subclasses, there are a lot of talk about mining and farmer protests.</li> <li>• In slow-growing Unemp districts there are much more articles about textile activities.</li> <li>• In fast-growing Unemp districts there are more articles about agricultural activities.</li> </ul>
<b>LIFESTYLE</b>	0.57	medic, death, river, die, treatment, farmer, hospital, doctor, protest, crime, children, women, patient, water, health	complaint, illegal, worker, dam, bull, rain, tribal, flu, crop, murder, court, jallikattu, dengue, forest, quarry, rape, mine, victim, farm, disease	school, student, college, culture, pollution, youth, idol, booth, voter, boat, life, festival, education	<ul style="list-style-type: none"> <li>• In both subclasses, there are a lot of talk about medical requirements, deaths, farmer protests.</li> <li>• In slow-growing Unemp districts there are much more articles about diseases, rape, mining and farming.</li> <li>• In fast-growing Unemp districts there are more articles about education, culture, voting, pollution and festival celebration.</li> </ul>

# PATTERN-1: SLOW GROWING UNEMP DISTRICTS vs FAST GROWING UNEMP DISTRICTS

## EXAMPLES OF TITLES OF ARTICLES FOR BOTH SUBCLASSES

COLLECTION	SLOW GROWING UNEMP	FAST GROWING UNEMP
AGRICULTURE	<ul style="list-style-type: none"> <li>• 3 more farmers commit suicide</li> <li>• New rice cultivation system boon for farmers</li> <li>• All 7 taluks get deficient rainfall in Mysuru</li> <li>• Shrinking dam spills water worries</li> <li>• Adopt a pond to save water</li> </ul>	<ul style="list-style-type: none"> <li>• Rail bridge gateway to fast, new life</li> <li>• Agriculture road map boosts output, minister says</li> <li>• Is CCE changing primary schools in Bihar?</li> <li>• First spell of rain exposes Patna's drainage system</li> <li>• Hot and humid conditions leave citizens in discomfort</li> <li>• EC recommends action against panchayat president for malpractice</li> </ul>
DEVELOPMENT	<ul style="list-style-type: none"> <li>• Jairam writes to Modi on NREGA complaints</li> <li>• Irregularity detected in rural employment scheme</li> <li>• 18 MGNREGS workers die of sunstroke in north Andhra this summer</li> <li>• Villagers fume over non-payment of wages in Eluru of Andhra Pradesh</li> <li>• Growth schemes crawl on ground</li> </ul>	<ul style="list-style-type: none"> <li>• Ensure proper execution of IAY scheme: CM</li> <li>• Mound spotted by CM yields rare artefacts</li> <li>• Floods threaten Tezpur University, air force station</li> <li>• Campus doors to open by April</li> </ul>
ENVIRONMENT	<ul style="list-style-type: none"> <li>• AP Police on toes as Maoists try to regain foothold in state</li> <li>• 2 top Pak terrorists killed in N Kashmir encounter</li> <li>• Elephant in musth kills 2</li> <li>• Bandipur tiger tranquilised and captured</li> <li>• 6 poachers held in joint raid</li> </ul>	<ul style="list-style-type: none"> <li>• 'Need to promote traditional varieties of pulses'</li> <li>• Pongal gift pack distribution begins</li> <li>• Govt sows green zeal seeds across state</li> <li>• Project push to cultivate rare paddy</li> <li>• Brand boosts student intake</li> </ul>
INDUSTRIALIZATION	<ul style="list-style-type: none"> <li>• 4 suspected Maoists killed in encounter</li> <li>• Reds blow up govt buildings, railway tracks</li> <li>• Rafiq murdered by rivals over land deal dispute?</li> <li>• Coal Sector, Banks Worst Affected</li> <li>• Tribal stir brings state to grinding halt</li> </ul>	<ul style="list-style-type: none"> <li>• Trade, industry hail state Budget</li> <li>• Industry demands funds boost</li> <li>• Irregularities in coal transportation for Sudha dairies</li> <li>• Sirhind canal water turns dark; DC warns against its use</li> <li>• Campus ready for convocation weekend</li> </ul>
LIFESTYLE	<ul style="list-style-type: none"> <li>• Mill attack</li> <li>• Two militants killed, one surrenders in Shopian encounter</li> <li>• Kashmir peaceful with restrictions in place</li> <li>• Strike paralyzes health services in districts</li> <li>• Will end Red menace in five years: Soren</li> </ul>	<ul style="list-style-type: none"> <li>• Patna police want gangster Bindu sent to Bhagalpur</li> <li>• Widow rape slur on prison guard</li> <li>• Fest to salute Maithili culture</li> <li>• Govt to prepare cultural road map, Nitish Kumar says</li> <li>• Hospital lacks anaesthetist</li> </ul>

## PATTERN-2: SLOW GROWING AGRI DISTRICTS vs FAST GROWING AGRI DISTRICTS

COLLECTION	GLOBAL CENTROID SIMILARITY	KEYWORDS THAT OCCUR IN BOTH SLOW-GROWING AND FAST-GROWING AGRI DISTRICTS	KEYWORDS THAT OCCUR MUCH MORE OFTEN IN SLOW-GROWING AGRI DISTRICTS	KEYWORDS THAT OCCUR MUCH MORE OFTEN IN FAST-GROWING AGRI DISTRICTS	COMMENTS
AGRICULTURE	0.35	loan, rainfall, agriculture, rain, farmer, protest, monsoon, water, loan, crop, farm, village, demand	system, variety, suicide, plant, mani, irrigation, grape, dam, wheat, hectare, hailstorm, season, flood, cultivate, paddy	forecast, field, problem, power, consumption, tariff, protest, waiver, electricity, growth, corrupt, sector, develop	<ul style="list-style-type: none"> <li>• In both subclasses, there are a lot of talk about rainfall, climate changes, farmer protests, bank loans.</li> <li>• In slow-growing Agri districts there are much more articles about cultivation of crops, floods, irrigation projects.</li> <li>• In fast-growing Agri districts there are more articles about loan and electricity waivers.</li> </ul>
DEVELOPMENT	0.22	air, scheme, connect, market, route, airport, train, skill, employ, invest, industry, project, flight, technology, develop	msme, assist, worker, manage, power, farmer, agriculture, irrigation, land, crop, cultivate, plant	policy, student, business, fund, service, job, company, investor, health, infrastructure, digital, data, research	<ul style="list-style-type: none"> <li>• In both subclasses, there are a lot of talk about schemes regarding airports and technology related developments.</li> <li>• In slow-growing Unemp districts, articles are more about agricultural activities and related issues.</li> <li>• In fast-growing Unemp districts, articles talk more about health, investments and digital development.</li> </ul>

## **PATTERN-2: SLOW GROWING AGRI DISTRICTS vs FAST GROWING AGRI DISTRICTS**

<b>COLLECTION</b>	<b>GLOBAL CENTROID SIMILARITY</b>	<b>KEYWORDS THAT OCCUR IN BOTH SLOW-GROWING AND FAST-GROWING AGRI DISTRICTS</b>	<b>KEYWORDS THAT OCCUR MUCH MORE OFTEN IN SLOW-GROWING AGRI DISTRICTS</b>	<b>KEYWORDS THAT OCCUR MUCH MORE OFTEN IN FAST-GROWING AGRI DISTRICTS</b>	<b>COMMENTS</b>
<b>ENVIRONMENT</b>	0.51	air, system, student, school, manage, river, waste, land, court, pollution, environment, industry, plant, pollution, develop	medic, tree, survey, agriculture, farmer, conserve, activist, water, village, drain, tourism, forest, proposal	illegal, consult, suicide, construct, poor, electricity, vehicle, clearance, public, education, transport	<ul style="list-style-type: none"> <li>• In both subclasses, there are a lot of talk about pollution, waste and other development activities.</li> <li>• In slow-growing Agri districts there are much more articles about agricultural activities and forest conservation.</li> <li>• In fast-growing Agri districts there are more articles about forest clearances and transportation.</li> </ul>
<b>INDUSTRIALIZATION</b>	0.57	manage, power, real, estate, market, business, opportunity, invest, industry, technology, develop	passenger, air, survey, river, hike, waste, land, grass, pollution, water, city, village, demand, steel, sanitisation	policy, term, fuel, airline, domestic, investor, money, spectrum, bank, growth, firm, global	<ul style="list-style-type: none"> <li>• In both subclasses, there are a lot of talk about business and industry development.</li> <li>• In slow-growing Agri districts there are much more articles about pollution and waste related problems.</li> <li>• In fast-growing Agri districts there are more articles about investment and monetary policies.</li> </ul>
<b>LIFESTYLE</b>	0.53	medic, air, college, service, pollution, tourist, water, environment, develop, health	worker, care, problem, student, traffic, agriculture, vehicle, organisation, business, doctor, festival	theatre, show, cbi, culture, market, travel, violence, protest, social, film, industry, invest, terror, crime	<ul style="list-style-type: none"> <li>• In both subclasses, there are a lot of talk about protests, tourism, hospital services.</li> <li>• In slow-growing Agri districts there are much more articles about festivals, facilities, policies.</li> <li>• In fast-growing Agri districts there are more articles about violence, protests, crimes.</li> </ul>

## PATTERN-2: SLOW GROWING AGRI DISTRICTS vs FAST GROWING AGRI DISTRICTS

### EXAMPLES OF TITLES OF ARTICLES FOR BOTH SUBCLASSES

COLLECTION	SLOW GROWING AGRI	FAST GROWING AGRI
AGRICULTURE	<ul style="list-style-type: none"> <li>• Farmers wait for subsidised farm equipment</li> <li>• Mallu promises relief to flood-hit farmers of Madhira</li> <li>• Raichur to get irrigation water after August 5</li> <li>• Plan alternative crops: Collector</li> <li>• Central team assesses drought in Arsikere</li> </ul>	<ul style="list-style-type: none"> <li>• Farmers with land along streams lost almost all the crop</li> <li>• Rain fails to push up dam levels</li> <li>• Agri staff stop work to protest corruption</li> <li>• High-level committee assesses crop damage in Ariyalur district</li> <li>• Monsoon ends with 12% shortfall</li> </ul>
DEVELOPMENT	<ul style="list-style-type: none"> <li>• 2 lakh labourers to get employment under NREGS</li> <li>• Drought-hit farmers pour out their woes</li> <li>• Agricultural implements donated</li> <li>• 'Tribal areas are well connected with NREGS'</li> <li>• 'Focus on development works in grama sabha meetings'</li> </ul>	<ul style="list-style-type: none"> <li>• UDAN's first flight: City connects with Porbander</li> <li>• Maharashtra nod for regional plan in eight districts</li> <li>• Forest Department launches afforestation drive in Ariyalur</li> <li>• Yogi govt plans airport terminal in Chitrakoot</li> <li>• Farmers pine hopes on north east monsoon</li> </ul>
ENVIRONMENT	<ul style="list-style-type: none"> <li>• Elephants kill two in Davangere district</li> <li>• Tribal people resume struggle for land</li> <li>• Farmers seek an end to monkey raids</li> <li>• Crop-raiding tusker captured in Dharmapuri</li> <li>• Naxal gunned down in Sukma weapons recovered</li> </ul>	<ul style="list-style-type: none"> <li>• Pranhita sanctuary only on paper</li> <li>• Tiger kills 1 in Gadchiroli</li> <li>• Forest department: 89% of 10-year-old plantations are unsuccessful</li> <li>• Mulak takes up cudgels for wildlife</li> <li>• Forest department evaluation confirms plantation scams</li> </ul>
INDUSTRIALIZATION	<ul style="list-style-type: none"> <li>• Ponnamm dares KCR to inspect Karimnagar town</li> <li>• Rural sanitation takes a beating</li> <li>• Scouts take out cycle rally</li> <li>• Power crisis to get worse as coal stocks plunge</li> <li>• SC notice to Rajasthan on illegal mining</li> </ul>	<ul style="list-style-type: none"> <li>• Chaos at environmental hearing for JSW project</li> <li>• Illegal sand mining continues unabated</li> <li>• No visible improvement in road infrastructure in Erode</li> <li>• Signals from Pak mobile companies reach bordering areas</li> <li>• Drought hits freight movement</li> </ul>
LIFESTYLE	<ul style="list-style-type: none"> <li>• Flamingo festival in tourism calendar</li> <li>• Bandh shuts down the city</li> <li>• Fervour marks Lord Ganesha immersion</li> <li>• Out of forests and into mines</li> <li>• Universal Health Coverage programme inaugurated</li> </ul>	<ul style="list-style-type: none"> <li>• Second phase of pulse polio campaign tomorrow</li> <li>• Rs 80 lakh public wealth destroyed in public violence</li> <li>• Govt zeros in on 6 Red-hit districts to tackle Maoists</li> <li>• Chela gets award along with guru</li> <li>• Clashes tensions during Holi relieves in Rajasthan</li> </ul>

## **PATTERN-3: SLOW GROWING NON-AGRI DISTRICTS vs FAST GROWING NON-AGRI DISTRICTS**

<b>COLLECTION</b>	<b>GLOBAL CENTROID SIMILARITY</b>	<b>KEYWORDS THAT OCCUR IN BOTH SLOW-GROWING AND FAST-GROWING NON-AGRI DISTRICTS</b>	<b>KEYWORDS THAT OCCUR MUCH MORE OFTEN IN SLOW-GROWING NON- AGRI DISTRICTS</b>	<b>KEYWORDS THAT OCCUR MUCH MORE OFTEN IN FAST-GROWING NON- AGRI DISTRICTS</b>	<b>COMMENTS</b>
<b>AGRICULTURE</b>	0.05	loan, march, rainfall, kill, field, normal, death, suicide, agriculture, compensation, rain, farmer, flood, harvest, monsoon, protest, paddy, village, forest	Fire, school, student, die, condition, scheme, seed, hailstorm, season, cultivate, education	sanction, power, construct, irrigation, dam, wheat, delay, rape, expressway, weather, life, develop	<ul style="list-style-type: none"> <li>• In both subclasses, there are a lot of talk about rainfall, floods and farmer suicides.</li> <li>• In slow-growing Non-Agri districts there are much more articles about hailstorms and crop cultivation.</li> <li>• In fast-growing Non-Agri districts there are more articles about irrigation, dams and sowing crops.</li> </ul>
<b>DEVELOPMENT</b>	0.14	Complaint, worker, person, scheme, connect, water, wage, mnrega, health, develop, programme, farmer, job, connect	School, problem, sanction, airport, bank, household, claim, inquiry, law, labour	Fire, dam, grant, poor, budget, scam, protest, elect, corrupt, arrest, dalits, infrastructure	<ul style="list-style-type: none"> <li>• In both subclasses, there are a lot of talk about schemes regarding agriculture and technology related developments.</li> <li>• In slow-growing Non-Agri districts, articles are more about sanction of funds and law related claims.</li> <li>• In fast-growing Non-Agri districts, articles talk more about scams, protests and corruption.</li> </ul>



## **PATTERN-3: SLOW GROWING NON-AGRI DISTRICTS vs FAST GROWING NON-AGRI DISTRICTS**

<b>COLLECTION</b>	<b>GLOBAL CENTROID SIMILARITY</b>	<b>KEYWORDS THAT OCCUR IN BOTH SLOW-GROWING AND FAST-GROWING NON-AGRI DISTRICTS</b>	<b>KEYWORDS THAT OCCUR MUCH MORE OFTEN IN SLOW-GROWING NON- AGRI DISTRICTS</b>	<b>KEYWORDS THAT OCCUR MUCH MORE OFTEN IN FAST-GROWING NON- AGRI DISTRICTS</b>	<b>COMMENTS</b>
<b>ENVIRONMENT</b>	0.14	fire, kill, death, herd, river, die, poacher, wildlife, wild, rain, rescue, tourism, forest, attack, animal, develop	Poach, crocodile, reserve, farmer, land, crop, blackbuck, tourist, protest, bauxit, mine	Ganga, march, student, suicide, leopard, drive, pollution, clean, troop, zoo, protect, city	<ul style="list-style-type: none"> <li>• In both subclasses, there are a lot of talk about poaching and rescue of wildlife.</li> <li>• In slow-growing Non-Agri districts there are much more articles about agriculture and mining activities and tourism.</li> <li>• In fast-growing Non-Agri districts there are more articles about pollution and protection of wildlife.</li> </ul>
<b>INDUSTRIALIZATION</b>	-0.09	Illegal, worker, school, kill, student, power, river, bandh, recover, train, rescue, area, plant, coal, develop	Fire, estate, vehicle, support, toilet, rain, tribal, protest, flood, maoist, weaver, attack, rape, victim, protest	Yatra, pond, death, farmer, criminal, maffia, miner, labour, heath, city, health, industry	<ul style="list-style-type: none"> <li>• In both subclasses, there are a lot of talk about worker issues and mining activities.</li> <li>• In slow-growing Non-Agri districts there are much more articles about pollution and protests.</li> <li>• In fast-growing Non-Agri districts there are more articles about criminal activities and healthcare.</li> </ul>
<b>LIFESTYLE</b>	-0.07	Medic, fire, night, kill, death, power, violence, hospital, doctor, protest, murder, crime, flood, rape, attack	Die, vehicle, military, rain, highway, curfew, strike, youth, dispute	Worker, politician, loot, relief, criminal, terror, cop, women, rally, demand, team	<ul style="list-style-type: none"> <li>• In both subclasses, there are a lot of talk about rapes, attacks, murders and other crimes.</li> <li>• In slow-growing Non-Agri districts there are much more articles about disputes and defence activities.</li> <li>• In fast-growing Non-Agri districts there are more articles about violence, crime and terrorism.</li> </ul>

## **PATTERN-3: SLOW GROWING NON-AGRI DISTRICTS vs FAST GROWING NON-AGRI DISTRICTS**

### **EXAMPLES OF TITLES OF ARTICLES FOR BOTH SUBCLASSES**

<b>COLLECTION</b>	<b>SLOW GROWING NON-AGRI</b>	<b>FAST GROWING NON-AGRI</b>
<b>AGRICULTURE</b>	<ul style="list-style-type: none"> <li>• Bankers fume at loan waiver muddle</li> <li>• Attempt to divide farmers: Activists</li> <li>• Anti-green corridor farmers stage protest in Salem</li> <li>• Shoddy work causes road under Metro to flood</li> <li>• Plea to stop field trials of GM crops</li> </ul>	<ul style="list-style-type: none"> <li>• MCG plans to revive seven water bodies before monsoon</li> <li>• NGT seeks reply on plea against disinfectant fumigation</li> <li>• Ozone pollution highest in Delhi among NCR cities in last 3 years</li> <li>• No witch-hunt in land scams probe: Haryana</li> <li>• Rain in Delhi hits traffic on DND</li> </ul>
<b>DEVELOPMENT</b>	<ul style="list-style-type: none"> <li>• Make in India: Will the programme do well to evolve beyond the 'take-make-use-dispose' model?</li> <li>• Modi 2.0 budget skips some important economic issues</li> <li>• Rahul fears job violence</li> <li>• Malnutrition high among urban poor children</li> <li>• Two major institutions set to boost city's health care status</li> </ul>	<ul style="list-style-type: none"> <li>• IT professionals bring lesser known geo-spatial technology to Goa</li> <li>• IT sector needs serious re-inventing</li> <li>• Life insurance companies struggling to retain employees</li> <li>• Centre advisory on cyber attacks</li> <li>• From verdant city to vertical slum</li> </ul>
<b>ENVIRONMENT</b>	<ul style="list-style-type: none"> <li>• Air pollution problem 'very serious'</li> <li>• No relief from toxic mix of pollutants until Sunday</li> <li>• No waste mgmt: City turns garbage pit</li> <li>• PM Modi to launch Eastern Peripheral Expressway meant to cut traffic</li> <li>• City bridges underutilized</li> </ul>	<ul style="list-style-type: none"> <li>• Gurgaon identifies 16 dumping sites</li> <li>• NCR cities need air clean-up plan</li> <li>• SC talks tough on illegal Aravali mines</li> <li>• Main reasons behind pollution in Gurugram: Diesel autos</li> <li>• Dengue fear grips city slum dwellers</li> </ul>
<b>INDUSTRIALIZATION</b>	<ul style="list-style-type: none"> <li>• No rate hike worries for cash-rich cos</li> <li>• Bharti Airtel set to sell stake in landline</li> <li>• 2G scam: Arun Shourie advises Raja to turn approver and expose all</li> <li>• Coal supply pact with power producers soon</li> <li>• Mumbai: Decline in slum TDR sharpest in 20 years</li> </ul>	<ul style="list-style-type: none"> <li>• Dip in illegal mining in the Aravalis over last three fiscal years</li> <li>• Waste dumped inside protected biodiversity park after MCG's 'do whatever' remark</li> <li>• Conclave next week for IT sector pushBus rape spooks working women</li> </ul>
<b>LIFESTYLE</b>	<ul style="list-style-type: none"> <li>• Using hate to challenge modernism</li> <li>• Infosys fastest to seize industry shift towards high-value enterprise services</li> <li>• Railways to get modest hike of Rs 2000 crore for next fiscal</li> <li>• 'Brace for worst waste crisis'</li> <li>• Crime against women</li> </ul>	<ul style="list-style-type: none"> <li>• B2B e-marketplace on investors' radar</li> <li>• CIFA develops new technology for murrel breeding</li> <li>• Water: Luxury vs scarcity</li> <li>• Toxic tentacles spread</li> <li>• More children fall prey to cyber crime as web users get younger</li> </ul>

## **PATTERN-4: SLOW GROWING UNEMP DISTRICTS vs FAST GROWING AGRI DISTRICTS**

<b>COLLECTION</b>	<b>GLOBAL CENTROID SIMILARITY</b>	<b>KEYWORDS THAT OCCUR IN BOTH SLOW-GROWING UNEMP AND FAST- GROWING AGRI DISTRICTS</b>	<b>KEYWORDS THAT OCCUR MUCH MORE OFTEN IN SLOW-GROWING UNEMP DISTRICTS</b>	<b>KEYWORDS THAT OCCUR MUCH MORE OFTEN IN FAST-GROWING-AGRI DISTRICTS</b>	<b>COMMENTS</b>
<b>AGRICULTURE</b>	0.42	loan, rainfall, supply, well, agriculture, rain, farmer, protest, monsoon, land, crop, waiver, center, house, project, farm, water, bank, committee, village, demand	sugar, suicide, meter, rabi, sugarcane, drink, fodder, dam, shower, relief, factory, season, debt, sow, tanker, damage, drought, loss, amount, storage	forecast, home, power, consume, tariff, elect, company, secure, court, draft, food, metro, invest, charge, law, park, growth, corrupt, opposite, finance, increase	<ul style="list-style-type: none"> <li>• In both subclasses, there are a lot of talk about rainfall, waivers and agriculture.</li> <li>• In slow-growing Unemp districts there are much more articles about farmer suicides and crop cultivation.</li> <li>• In fast-growing Agri districts there are more articles about tariff, metro and laws.</li> </ul>
<b>DEVELOPMENT</b>	0.45	scheme, aadhaar, job, employ, children, center, house, project, wage, product, bank, program, develop	sanction, fish, mgnrega, farmer, relief, land, crop, season, forest, village, worker, power, agriculture, poultry, construct, waste, sabha, subsidy, accord, welfare, drought,, panchayat, water, labour	policy, air, ticket, support, budget, skill, elect, invest, investor, mission, health, airport, research, manufacture, smart, technology, infrastructure, connect, market, fund, company, railway, industry, flight, school, student, train, court	<ul style="list-style-type: none"> <li>• In both subclasses, there are a lot of talk about schemes regarding employment, aadhaar and wages.</li> <li>• In slow-growing Unemp districts, articles are more about agriculture and poultry related activities.</li> <li>• In fast-growing Agri districts, articles talk more about technology, infrastructure and transportation.</li> </ul>

## **PATTERN-4: SLOW GROWING UNEMP DISTRICTS vs FAST GROWING AGRI DISTRICTS**

COLLECTION	GLOBAL CENTROID SIMILARITY	KEYWORDS THAT OCCUR IN BOTH SLOW-GROWING UNEMP AND FAST-GROWING AGRI DISTRICTS	KEYWORDS THAT OCCUR MUCH MORE OFTEN IN SLOW-GROWING UNEMP DISTRICTS	KEYWORDS THAT OCCUR MUCH MORE OFTEN IN FAST-GROWING-AGRI DISTRICTS	COMMENTS
ENVIRONMENT	0.47	illegal, power, project, land, company, reside, court, well	kill, death, farmer, maoist, arrest, encroach, forest, conserve, tourist, dam, rain, protect, bail, sanctuary, demand, cbi, export, leopard, carcass, tourism, badger, ore, tribal, honey, mine, wildlife	air, river, electricity, pollute, metro, road, oil, <b>ngt</b> , children, invest, quality, ban, education, system, consult, problem, traffic, global, construct, poor, vehicle, municipal, women, industry, city, school, student	<ul style="list-style-type: none"> <li>• In both subclasses, there are a lot of talk illegal activities.</li> <li>• In slow-growing Unemp districts there are much more articles about farmers, tourism, conservation of wildlife.</li> <li>• In fast-growing Agri districts there are more articles about pollution, metro, education and transportation.</li> </ul>
INDUSTRIALIZATION	0.79	cbi, fund, high, power, project, company, house, industry, court, coal	river, temple, farmer, electricity, mill, textile, forest, miner, activist, driver, worker, protest, construct, weaver, women, iron, transport, illegal, handloom, bench, process, ore, mine, water	sale, invest, quality, data, airline, estate, call, institute, global, manufacture, sell, technology, fuel, market, kyc, telecom, gst, bank, student, import, wto, trade, spectrum, service, tax, foreign, point, drug, dollar, aircraft	<ul style="list-style-type: none"> <li>• In both subclasses, there are a lot of talk about electricity, coal and funding activities.</li> <li>• In slow-growing Unemp districts there are much more articles about workers, protests, transportation.</li> <li>• In fast-growing Agri districts there are more articles about real estate, telecommunication, manufacturing, banking and finance activities.</li> </ul>
LIFESTYLE	0.56	medic, hospital, protest, elect, court, crime, accuse, arrest, women, rape, team, police, project, water, tourism, health, demand	worker, river, swine, dam, bull, farmer, rain, flu, treatment, tribal, doctor, murder, children, jallikattu, dengue, forest, quarry, mine, disease, woman, screen, program	theatre, air, college, political, culture, poll, market, travel, violence, pollute, gender, leader, agency, life, interest, tourist, brand, industry, terror, drug, law, movie, film	<ul style="list-style-type: none"> <li>• In both subclasses, there are a lot of talk about rapes, women, medical facilities and also tourism.</li> <li>• In slow-growing Unemp districts there are much more articles about mining and tribal activities.</li> <li>• In fast-growing Agri districts there are more articles about theatre, travel, violence, and tourism.</li> </ul>

## PATTERN-4: SLOW GROWING UNEMP DISTRICTS vs FAST GROWING AGRI DISTRICTS

### EXAMPLES OF TITLES OF ARTICLES FOR BOTH SUBCLASSES

COLLECTION	SLOW GROWING UNEMP	FAST GROWING AGRI
AGRICULTURE	<ul style="list-style-type: none"> <li>• Baghpat on edge after sugarcane farmer dies during dharna</li> <li>• Punjab Finance Minister Badal favours all-India farm debt waiver scheme</li> <li>• Season of showers arrives in state</li> <li>• New rice cultivation system boon for farmers</li> <li>• 50% of crops lost in Dharwad district due to rain deficit</li> </ul>	<ul style="list-style-type: none"> <li>• Farmers with land along streams lost almost all the crop</li> <li>• Rain fails to push up dam levels</li> <li>• Agri staff stop work to protest corruption</li> <li>• High-level committee assesses crop damage in Ariyalur district</li> <li>• Monsoon ends with 12% shortfall</li> </ul>
DEVELOPMENT	<ul style="list-style-type: none"> <li>• FIR against District Fisheries Officer, Junior Engineer for Rs 18.06 lakh fraud</li> <li>• Beti Bachao, Beti Padhao: PM Modi to visit Jhunjhunu on March 8</li> <li>• As the rain gods fail, Chawadal villagers give up hope</li> <li>• Rice mill launch</li> <li>• Teachers to boost Centre's Beti Bachao Beti Padhao scheme in Rajasthan</li> </ul>	<ul style="list-style-type: none"> <li>• UDAN's first flight: City connects with Porbander</li> <li>• Maharashtra nod for regional plan in eight districts</li> <li>• Forest Department launches afforestation drive in Ariyalur</li> <li>• Yogi govt plans airport terminal in Chitrakoot</li> <li>• Farmers pine hopes on north east monsoon</li> </ul>
ENVIRONMENT	<ul style="list-style-type: none"> <li>• Illegal mining of laterite continues unabated</li> <li>• UP farmers catch leopard alive, hang it upside down</li> <li>• Forest officials rescue elephant</li> <li>• Poachers kill another Kaziranga rhino</li> <li>• Cyclone Phailin causes blackbuck migration</li> </ul>	<ul style="list-style-type: none"> <li>• Pranhita sanctuary only on paper</li> <li>• Tiger kills 1 in Gadchiroli</li> <li>• Forest department: 89% of 10-year-old plantations are unsuccessful</li> <li>• Mulak takes up cudgels for wildlife</li> <li>• Forest department evaluation confirms plantation scams</li> </ul>
INDUSTRIALIZATION	<ul style="list-style-type: none"> <li>• More help on way for drought-hit farmers</li> <li>• Three workers killed after gas leak in Chhattisgarh coal mine</li> <li>• Bride gets Rs 2L award for demanding toilet facilities</li> <li>• Water crisis deepens in Odisha with rise in temperature</li> <li>• Woman gang-raped, killed; juvenile nabbed</li> </ul>	<ul style="list-style-type: none"> <li>• Chaos at environmental hearing for JSW project</li> <li>• Illegal sand mining continues unabated</li> <li>• No visible improvement in road infrastructure in Erode</li> <li>• Signals from Pak mobile companies reach bordering areas</li> <li>• Drought hits freight movement</li> </ul>
LIFESTYLE	<ul style="list-style-type: none"> <li>• Welfare schemes: Guntur makes a mark</li> <li>• Deal with crimes against women promptly, CM to police</li> <li>• Sundargarh gets new health centres</li> <li>• Two 'kidnappers' posing as cops arrested, sent to 14-day custody</li> <li>• Piped water hope for 57 villages</li> </ul>	<ul style="list-style-type: none"> <li>• Second phase of pulse polio campaign tomorrow</li> <li>• Rs 80 lakh public wealth destroyed in public violence</li> <li>• Govt zeros in on 6 Red-hit districts to tackle Maoists</li> <li>• Chela gets award along with guru</li> <li>• Clashes tensions during Holi relieves in Rajasthan</li> </ul>

## **PATTERN-5: SLOW GROWING AGRI DISTRICTS vs FAST GROWING NON-AGRI DISTRICTS**

<b>COLLECTION</b>	<b>GLOBAL CENTROID SIMILARITY</b>	<b>KEYWORDS THAT OCCUR IN BOTH SLOW-GROWING AGRI AND FAST-GROWING NON-AGRI DISTRICTS</b>	<b>KEYWORDS THAT OCCUR MUCH MORE OFTEN IN SLOW-GROWING AGRI DISTRICTS</b>	<b>KEYWORDS THAT OCCUR MUCH MORE OFTEN IN FAST-GROWING NON- AGRI DISTRICTS</b>	<b>COMMENTS</b>
<b>AGRICULTURE</b>	-0.34	loan, rainfall, suicide, agriculture, dam, rain, farmer, monsoon, crop, wheat, price, flood, forest, farm, water, bank, loss, weather, harvest, village, demand	supply, variety, pond, scheme, market, stock, grape, shower, hailstorm, season, strike, winery, sow, horticulture, tanker, temperature, onion,	kill, field, sanction, death, power, college, river, gang, construct, relief, hospital, protest, court, attack, rape, expressway, life, police, tiger, product, claim, cabinet, develop	<ul style="list-style-type: none"> <li>• In both subclasses, there are a lot of talk about rainfall, floods and farmer suicides.</li> <li>• In slow-growing Agri districts there are much more articles about different cultivation practices, schemes to resolve farmers issues, strikes of farmers, situation of market, effect of temperature.</li> <li>• In fast-growing non-agri districts there are much more articles about sanction of funds, power, construction, hospital, field of crops.</li> </ul>
<b>DEVELOPMENT</b>	-0.21	worker, scheme, construct, connect, airport, farmer, tribal, highway, train, land, employ, women, project, implement, water, program, village, develop	air, power, college, estate, agriculture, prison, market, subsidy, skill, crop, enterprise, textile, bird, institute, youth, invest, horticulture, industry, flight, park, smart, entrepreneurship, storage, city, plant, technology, entrepreneur	complaint, school, sanction, student, die, girl, fund, yojana, job, court, road, bridge, mnrega, maoist, child, commission, police, beti, wage, pmgsy, campaign, engine, household, law, bank, monitor, labour, health, demand	<ul style="list-style-type: none"> <li>• In both subclasses, there are a lot of talk about schemes regarding village development schemes and technology related developments.</li> <li>• In slow-growing Agri districts, articles are more about horticulture, production of crops, implementation of technology, skills of farmers, facilities to farmers.</li> <li>• In fast-growing non-agri districts, articles talk more about students , IT jobs, air pollution, bank and health facilities,</li> </ul>

## PATTERN-5: SLOW GROWING AGRI DISTRICTS vs FAST GROWING NON-AGRI DISTRICTS

COLLECTION	GLOBAL CENTROID SIMILARITY	KEYWORDS THAT OCCUR IN BOTH SLOW-GROWING AGRI AND FAST- GROWING NON-AGRI DISTRICTS	KEYWORDS THAT OCCUR MUCH MORE OFTEN IN SLOW- GROWING AGRI DISTRICTS	KEYWORDS THAT OCCUR MUCH MORE OFTEN IN FAST-GROWING NON- AGRI DISTRICTS	COMMENTS
ENVIRONMENT	-0.29	river, farmer, land, court, bird, area, forest, police, project, water, tourism, reside, village, develop	medic, air, plastic, hospital, conserve, quality, activist, mayor, clean, noise, plant, green, municipal, waste, commission, industry, drain, school, student, NGO, treatment, puja, dispose, cover, swachh	fire, kill, death, poach, electricity, blackbuck, relief, crop, maoist, arrest, guard, tourist, rain, camp, recover, protest, flood, attack, tiger, sanctuary, college, crocodile, poacher, wild, park, herd, tribal, wildlife	<ul style="list-style-type: none"> <li>In both subclasses, there are a lot of talk about forest clearance wildlifes and tourism.</li> <li>In slow-growing Agri districts there are much more articles about programmes for wildlifes, Forest clearance, plantation, industries,</li> <li>In fast-growing non-agri districts there are more articles about relief factors, wildlife, mining, tourism, agriculture fire.</li> </ul>
INDUSTRIALIZATION	0.01	school, student, power, river, estate, construct, well, toilet, train, land, company, sanity, mine, water, reside, village, plant, develop	air, survey, connect, hike, shop, market, airport, waste, pollute, tax, textile, garbage, institute, children, commission, invest, industry, product, engine, manufacture, wine, technology, demand, steel	illegal, fire, worker, kill, vehicle, rain, tribal, protest, murder, rescue, raid, flood, weaver, arrest, forest, dengue, attack, rape, agency, police, victim, rebel, coal	<ul style="list-style-type: none"> <li>In both subclasses, there are a lot of talk about illegal mining and vllagers issues</li> <li>In slow-growing Agri districts there are much more articles about industry, textile, studies, pollution, investment.</li> <li>In fast-growing Non-Agri districts there are more articles about forest, illegal mining , rape, cop duties, temperature.</li> </ul>
LIFESTYLE	-0.12	medic, vehicle, service, hospital, doctor, flood, police, patient, water, village, health, develop	air, kumbh, farmer, pollute, children, tourist, worker, problem, traffic, agriculture, dam, institute, research, disease, product, college, toilet, municipal, mela, women, tourism, screen, program, school, student, drive, fever, treatment, train, irctc	fire, kill, death, die, violence, land, elect, strike, arrest, power, poll, rain, highway, protest, spot, law, force, reside, murder, rape, vote, dispute, secure, curfew, court, crime, youth, victim, army, life	<ul style="list-style-type: none"> <li>In both subclasses, there are a lot of talk about health awareness and crimes like rape, violence.</li> <li>In slow-growing Agri districts there are much more articles about pollution, tourism, woman and children, treatment, health program, dam.</li> <li>In fast-growing Non-Agri districts there are more articles about strikes, minorities success, murder, sexual harassmt, raids.</li> </ul>

**PATTERN-5: SLOW GROWING AGRI DISTRICTS vs FAST GROWING NON-AGRI DISTRICTS**

**EXAMPLES OF TITLES OF ARTICLES FOR BOTH SUBCLASSES**

COLLECTION	SLOW GROWING AGRI	FAST GROWING NON-AGRI
AGRICULTURE	<ul style="list-style-type: none"><li>• ‘Assess crop loss, provide compensation to farmers’</li><li>• Plan alternative crops: Collector</li><li>• BJP Kisan Morcha stages dharna in front of Collectorate</li><li>• Paddy purchase going on smoothly</li><li>• Dry spell in Andhra forces tenant farmers to take a summer break</li></ul>	<ul style="list-style-type: none"><li>• Load rejig in peak hours may ease power cuts</li><li>• Haryana sanctions DIF of over Rs. 1.93 crore</li><li>• Haryana clears Rs. 255 crore road-widening project in NCR</li><li>• Health department fighting dengue with dud gun</li><li>• 10L acre wheat damaged in Haryana</li></ul>
DEVELOPMENT	<ul style="list-style-type: none"><li>• State increases target for horticulture crops</li><li>• 1, 500 ha of jowar to be produced for Anna Bhagya scheme</li><li>• Biometric machine installation at govt offices at snail’s pace</li><li>• Farmers grow green fodder to tackle shortage</li><li>• Five lakh ‘Agathi’ seedlings to be distributed to farmers</li></ul>	<ul style="list-style-type: none"><li>• New challenges in IT job prospects</li><li>• Students baffled by CBSE results delay</li><li>• Nod to new border road agency</li><li>• World Health Organisation to South-East Asian countries: Accelerate efforts to address air pollution</li><li>• World Bank scouts for innovative social projects</li></ul>
ENVIRONMENT	<ul style="list-style-type: none"><li>• AForest clearance for Palamuru-RR LI</li><li>• Non-teak species to get greater priority in growing plantations from next year</li><li>• Govt acts tough with tendu contractors</li><li>• Uma allays ryots’ fear over Pattiseema project</li><li>• Social forestry wing to help farmers take commercial route</li></ul>	<ul style="list-style-type: none"><li>• Punjab, Haryana told to check agricultural fires</li><li>• DMRC launches e-rickshaw service in Ghaziabad</li><li>• Manesar’s wild side: A peek at a leopard family in Aravali forest</li><li>• SC talks tough on illegal Aravali mines</li><li>• Tourist hotspot on the anvil</li></ul>
INDUSTRIALIZATION	<ul style="list-style-type: none"><li>• Industry prepared to pay more for power</li><li>• WB studies biometric, Aadhaar-enabled services</li><li>• Textile parks will be established in all taluks, says Anjaneya</li><li>• Civic workers' strike raises a stink</li><li>• Chief Minister invites U.S. investment in aerospace sector</li></ul>	<ul style="list-style-type: none"><li>• With summer on, demand for gensets goes up in city</li><li>• Unitech promoters sent to five-day police custody for FD scheme probe</li><li>• Bus rape spooks working women</li><li>• Illegal mining ruined Aravalis in Haryana, Rajasthan'</li><li>• Aravalis a forest? Survey to decide</li></ul>
LIFESTYLE	<ul style="list-style-type: none"><li>• Telangana to get tourism boost with heliports</li><li>• Woman, infant die after nurse botches up delivery</li><li>• State plans two tertiary centres for cancer treatment</li><li>• Universal Health Coverage programme inaugurated</li><li>• Bandh shuts down City</li></ul>	<ul style="list-style-type: none"><li>• Trade unions' strike paralyzes region</li><li>• Free coaching for minorities and women at Jamia</li><li>• Raids on to nab 3 key members of Kaushal gang</li><li>• 7 more arrested for murder of Congress leader Vikas Chaudhary</li><li>• Two girls questioned: ‘Knew about Crazy Sumit video, didn’t know he’ll upload it’</li></ul>



## PATTERN-6: SLOW GROWING UNEMP DISTRICTS vs FAST GROWING NON-AGRI DISTRICTS

COLLECTION	GLOBAL CENTROID SIMILARITY	KEYWORDS THAT OCCUR IN BOTH SLOW-GROWING UNEMP AND FAST-GROWING NON-AGRI DISTRICTS	KEYWORDS THAT OCCUR MUCH MORE OFTEN IN SLOW-GROWING UNEMP DISTRICTS	KEYWORDS THAT OCCUR MUCH MORE OFTEN IN FAST-GROWING NON-AGRI DISTRICTS	COMMENTS
AGRICULTURE	0.41	loan, rainfall, kill, suicide, agriculture, dam, rain, farmer, relief, protest, monsoon, crop, policy, farm, water, mm, loss, part, village, demand	sugar, rabi, sugarcane, fodder, shower, waiver, factory, season, debt, sow, collector, cultivation, taluka, drought, storage, committee	sanction, power, river, gang, college, construct, hospital, wheat, flood, arrest, forest, express-way, rape, product, cabinet, weather, harvest, develop	<ul style="list-style-type: none"> <li>• In both subclasses, there are a lot of talk about monsoon, projects, farmers issues and facilities</li> <li>• In slow-growing unemp districts there are much more articles about schemes, sugarcane, season, cultivation, drought.</li> <li>• In fast-growing non-agri districts there are much more articles about sanction of funds, paddy crop, weather, colleges.</li> </ul>
DEVELOPMENT	0.07	worker, sanction, scheme, construct, mgnrega, farmer, job, panchayat, children, area, police, implement, water, wage, bank, committee, labour, program, sever, demand, develop	benefit, pond, power, fish, agriculture, fodder, aadhaar, municipal, waste, relief, subsidy, crop, season, rate, zilla, forest, welfare, farmer, product, drought, poultry, progress, yield, distribute	school, student, connect, airport, fund, yojana, highway, train, court, road, bridge, commission, payment, women, beti, pmgsy, campaign, engine, household, law, health	<ul style="list-style-type: none"> <li>• In both subclasses, there are a lot of talk about technology development, bank facilities, schemes and programs for farmers and workers.</li> <li>• In slow-growing unemp districts, articles are more about Fisheries, welfare, drought, production of crop, studies.</li> <li>• In fast-growing non-agri districts, articles talk more about schools ranking, airport facilities, bank projects, internet connectivity.</li> </ul>

## **PATTERN-6: SLOW GROWING UNEMP DISTRICTS vs FAST GROWING NON-AGRI DISTRICTS**

<b>COLLECTION</b>	<b>GLOBAL CENTROID SIMILARITY</b>	<b>KEYWORDS THAT OCCUR IN BOTH SLOW-GROWING UNEMP AND FAST-GROWING NON-AGRI DISTRICTS</b>	<b>KEYWORDS THAT OCCUR MUCH MORE OFTEN IN SLOW-GROWING UNEMP DISTRICTS</b>	<b>KEYWORDS THAT OCCUR MUCH MORE OFTEN IN FAST-GROWING NON- AGRI DISTRICTS</b>	<b>COMMENTS</b>
<b>ENVIRONMENT</b>	-0.05	kill, death, wild, rain, farmer, tribal, court, maoist, arrest, forest, mine, police, spot, wildlife, tiger, tourist, water, tourism, sanctuary, village, demand	illegal, cub, field, grant, export, power, leopard, trap, dam, chinkara, company, cost, honey, encroach, peacock, protect, conserve, carcass, badger, skin, leas	fire, college, river, seat, die, herd, poacher, blackbuck, camp, relief, protest, crop, murder, secure, train, rescue, flood, bird, morn, attack,, guard, park, encounter	<ul style="list-style-type: none"> <li>• In both subclasses, there are a lot of talk about illegal mining, wildlife conservation.</li> <li>• In slow-growing unemp districts there are much more articles about Illegal activities, conservation of wildlife, carcass, relocation of animals.</li> <li>• In fast-growing non-agri districts there are more articles about murder security, crop, rescue, relief.</li> </ul>
<b>INDUSTRIALIZATION</b>	0.19	illegal, worker, power, river, construct, toilet, protest, elect, land, company, panchayat, court, arrest, area, forest, attack, mine, police, project, water, sand, coal, develop	justice, handloom, die, temple, trap, fund, mine, ore, farmer, mill, textile, women, ban, dead, iron, campaign, miner, park, activist, file, driver, transport, leas	fire, school, kill, student, estate, vehicle, rain, tribal, train, rescue, murder, raid, flood, maoist, dengue, agency, rape, victim, stone, record, seize, temperature, plant, rebel	<ul style="list-style-type: none"> <li>• In both subclasses, there are a lot of talk about illegal mining, coal sector, worker's issues.</li> <li>• In slow-growing unemp districts there are much more articles about farmers help, demand of facilities, environment, mining, woman issues</li> <li>• In fast-growing Non-Agri districts there are more articles about real estate, raids, defence, residents facilities.</li> </ul>
<b>LIFESTYLE</b>	0.06	medic, fire, death, die, communication, rain, service, hospital, doctor, protest, land, elect, murder, court, crime, station, arrest, rape, police, patient, victim, water, village, sever, health, develop	illegal, worker, scheme, river, swine, dam, bull, farmer, flu, treatment, high, tribal, crop, children, jallikattu, dengue, forest, quarri, mine, farm, ban, disease, tourism, record, woman, screen, program	kill, power, gang, girl, vehicle, military, minor, violence, highway, curfew, strike, flood, maoist, youth, life, army, restrict, vote, law, force, dispute, surrender	<ul style="list-style-type: none"> <li>• In both subclasses, there are a lot of talk about health issues, crime, rape.</li> <li>• In slow-growing unemp districts there are much more articles about welfare scheme, health and care, woman, illegal activities, facilities.</li> <li>• In fast-growing Non-Agri districts there are more articles about youth success, strike, army, safety and security, disputes, law.</li> </ul>

## PATTERN-6: SLOW GROWING UNEMP DISTRICTS vs FAST GROWING NON-AGRI DISTRICTS

### EXAMPLES OF TITLES OF ARTICLES FOR BOTH SUBCLASSES

COLLECTION	SLOW GROWING UNEMP	FAST GROWING NON-AGRI
AGRICULTURE	<ul style="list-style-type: none"> <li>• Baghpat on edge after sugarcane farmer dies during dharna</li> <li>• Punjab Finance Minister Badal favours all-India farm debt waiver scheme</li> <li>• Season of showers arrives in state</li> <li>• New rice cultivation system boon for farmers</li> <li>• 50% of crops lost in Dharwad district due to rain deficit</li> </ul>	<ul style="list-style-type: none"> <li>• Haryana sanctions DIF of over Rs. 1.93 crore</li> <li>• 20 peacocks found dead in Khurda forest</li> <li>• Rains bring down temperatures in Delhi-NCR, but cause traffic jams</li> <li>• Cold damages paddy seedlings</li> <li>• BHU entrance exams conducted at 40 centres</li> </ul>
DEVELOPMENT	<ul style="list-style-type: none"> <li>• FIR against District Fisheries Officer, Junior Engineer for Rs 18.06 lakh fraud</li> <li>• Beti Bachao, Beti Padhao: PM Modi to visit Jhunjhunu on March 8</li> <li>• As the rain gods fail, Chawadal villagers give up hope</li> <li>• Rice mill launch</li> <li>• Teachers to boost Centre's Beti Bachao Beti Padhao scheme in Rajasthan</li> </ul>	<ul style="list-style-type: none"> <li>• 3 Bangalore schools among India's best</li> <li>• VSS airport records high traffic in 1 month</li> <li>• World Bank scouts for innovative social projects</li> <li>• Tiruchi Junction a Wi-Fi hotspot now</li> <li>• Doubts on fortified midday-meal salt</li> </ul>
ENVIRONMENT	<ul style="list-style-type: none"> <li>• Illegal mining of laterite continues unabated</li> <li>• UP farmers catch leopard alive, hang it upside down</li> <li>• Forest officials rescue elephant</li> <li>• Poachers kill another Kaziranga rhino</li> <li>• Cyclone Phailin causes blackbuck migration</li> </ul>	<ul style="list-style-type: none"> <li>• Gurgaon: Forest department moots 500 home guards for Aravalli forest</li> <li>• GMDA ropes in private agency for water security study</li> <li>• Punjab, Haryana told to check agricultural fires</li> <li>• Leopard, two cubs spotted in Manesar</li> <li>• Tender floated for 250 more CNG buses</li> </ul>
INDUSTRIALIZATION	<ul style="list-style-type: none"> <li>• More help on way for drought-hit farmers</li> <li>• Three workers killed after gas leak in Chhattisgarh coal mine</li> <li>• Bride gets Rs 2L award for demanding toilet facilities</li> <li>• Water crisis deepens in Odisha with rise in temperature</li> <li>• Woman gang-raped, killed; juvenile nabbed</li> </ul>	<ul style="list-style-type: none"> <li>• Leg-up for real estate, traffic to improve as minister inaugurates Kundli-Manesar-Palwal Expressway</li> <li>• CBI raids Hooda, officials in Rs 1,500 cr land grab case</li> <li>• Official who cleared Vadra deal booked Why residential property prices never fell in cities</li> </ul>
LIFESTYLE	<ul style="list-style-type: none"> <li>• Welfare schemes: Guntur makes a mark</li> <li>• Deal with crimes against women promptly, CM to police</li> <li>• Sundargarh gets new health centres</li> <li>• Two 'kidnappers' posing as cops arrested, sent to 14-day custody</li> <li>• Piped water hope for 57 villages</li> </ul>	<ul style="list-style-type: none"> <li>• Free coaching for minorities and women at Jamia</li> <li>• No Mumbai rerun, cab strike has little effect in Delhi-NCR</li> <li>• Jat stir: Army conducts flag march in Rohtak, Bhiwani</li> <li>• Budding designers check in</li> <li>• 100 seats for MBBS aspirants at AIIMS in Raipu</li> </ul>

# DISTRICT LEVEL ANALYSIS USING TAGS EMBEDDINGS

## OBTAIN THE DISTRICT VECTORS

First we obtain the district vectors from tag embeddings of the models that we had trained. We do this for all the collections.



## 250-dimensional representation for each district

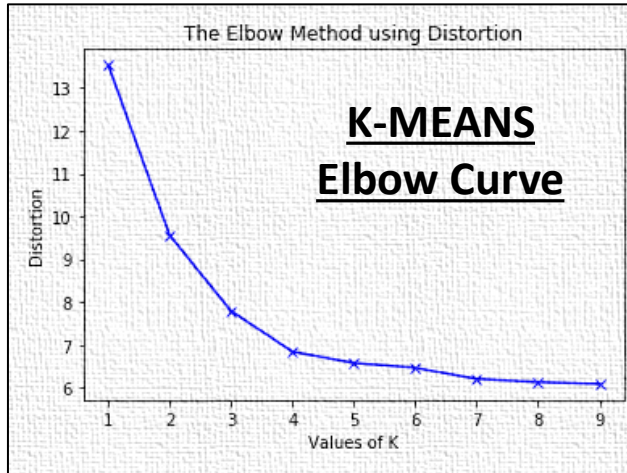
Each district, for a collection has a 50 dimensional tag embedding. Since we have 5 different collections, so by concatenating all these vectors we form a 250 dimensional vector.

## Data Visualization using t-SNE

We then use t-SNE to convert this 250 dimensional vector to a 2-dimensional vector, for data visualization.

## Clustering using K-Means

We use K-Means to cluster these district vectors. First, we start by trying different values for k, and plotting the elbow curve.



Using this elbow curve we choose the value for k as 4.

The clusters formed have the following number of districts in them:

Clusters-1 : 62  
Clusters-2 : 82  
Clusters-3 : 216  
Clusters-4 : 176

## T-SNE Plot

This plot shows us a 2-dim representation of districts

