TOI insights

Akash Gupta[1]

Abstract

18 opinion articles form the month of May 2020 were mined from Times of India archives relating to the migrant workers crisis. In this analysis the text has been broken down into vectors, tokens and ngrams so as to analyse the context of important phrases and words that link most intimately to the migrant issue.

Preparing the data

The data has first been mined from the internet and parsed from html format to clean text. Further, the text data is organised into a machine-readable dataframe and unrequired special characters like **punctuation** marks and **Upper case** letters have been filtered out.

```
head(article_collection)
```

```
## # A tibble: 6 x 4
##   title text                              num clean_text
##   <chr> <chr>                           <int> <chr>
## 1 TOI   The so-called symbol of ‘livin~     1 "the so called symbol of living m~
## 2 TOI   My heart sank as I saw the vir~     2 "my heart sank as i saw the viral~
## 3 TOI   Several commentators have writ~     3 "several commentators have writte~
## 4 TOI   It’s painful for any human to ~     4 "it s painful for any human to se~
## 5 TOI   As the Mahatma Gandhi National~     5 "as the mahatma gandhi national r~
## 6 TOI   The lockdown has shown up the ~     6 "the lockdown has shown up the se~
```

Creating wordclouds

The next step is to make a wordcloud of all the important **Nouns** used in all the **TOI** articles put together. We will do the same for the other publications since this will give us an overview comparison of - what are the different topics/subjects the publications are talking about within the framework of the migrant issue. Below is the wordcloud for TOI in which we can quickly notice some of the important subjects of discussion - **Bihar, Aadhaar, Adityanath, Dhan, Employment, Congress**.

Email address: pgdm19akashgupta@mse.ac.in (Akash Gupta)

Figure 1: Figure 1: wordcloud

N-gram analysis

Here we set **n=4** and essentially extract phrases of 4 neighboring words around a **Noun/Subject** so as to get an idea about the context in which the subject is being talked about.

In this first table, **word1** has been filtered to show the **sentiment lexicon** terms which essentially set the tone for the rest of the phrase. For examples notice the first row - ***"bitterly fractious ideological battle"*** : **bitterly** sets the negative tone in this regard. The second row - ***"affluent city dwellers deride"*** : has been said in the context of **Nrega** : We can hence infer from this statement that - **city folk tend to not support policies like NREGA**. In index number 18 - the phrase **"diverted payments"** also paints an intersting picture of how these schemes might be perceived.

```
## # A tibble: 21 x 5
##       X1 word1        word2          word3        word4
##    <dbl> <chr>        <chr>          <chr>        <chr>
##  1     5 bitterly     fractious      ideological  battle
##  2     6 affluent     city           dwellers     deride
##  3     6 expensive    gravy          train        money
##  4     7 righteous    corporate      india        blame
##  5    18 jam          trinity        jan          dhan
##  6    18 rejected     payments       diverted     payments
```

```
## 7      20 innovative  citizen       centric     social
## 8      25 leading     development   economist   pranab
## 9      27 conflicting income        support     schemes
## 10     30 strong      institutional social      audits
## # ... with 11 more rows
```

In yet another set of 4-grams we can see in index number 27 - ***"multiple conflicing income support"*** : tends to suggest that the various income support schemes for the poor tend to be viewed as chaotic and confusing since they conflict with each other. Index number 105 - ***"extra revolutionary arvind kejriwal"*** : suggests a negative, sarcastic view of the Delhi chief minister in this context.

```
## # A tibble: 12 x 5
##        X1 word1        word2         word3       word4
##     <dbl> <chr>        <chr>         <chr>       <chr>
## 1       5 bitterly     fractious     ideological battle
## 2      18 leveraging   jam           trinity     jan
## 3      27 multiple     conflicting   income      support
## 4      31 income       support       welfare     programmes
## 5     105 extra        revolutionary arvind      kejriwal
## 6     125 home         sick          ness        kindled
## 7     154 india        unequal       india       india
## 8     156 reverse      hard          fought      gains
## 9     189 pushed       disabled      gayoor      ahmed's
## 10    224 hospitals    trauma        centres     orphanages
## 11    246 workers      undermining   physical    distancing
## 12    274 tremendously rich          temples     religious
```

Another 4-gram set brings out more interesting observations. Index number 7 - ***"corporate india blame nrega"*** : another phrase suggesting negative view towards nrega among the urban population of India. An important phrase at index number 61 - ***"largest humanitarian crisis dwarfing"*** : suggests that the migrant crisis is by no means to be taken lightly since in this context it is viewed as an extremely serious issue that requires immediate attention.

```
## # A tibble: 13 x 5
##        X1 word1       word2        word3     word4
##     <dbl> <chr>       <chr>        <chr>     <chr>
## 1       7 corporate   india        blame     nrega
## 2      10 dhan        yojna        smart     city
## 3      27 basic       income       support   combined
## 4      27 conflicting income       support   schemes
## 5      29 communities land         reforms   beneficiaries
## 6      61 largest     humanitarian crisis    dwarfing
## 7      70 easing      direct       benefit   transfers
## 8     154 growing     india        unequal   india
```

```
## 9    187 fellow      migrants     hard        scrabbling
## 10   189 anirudh     pushed       disabled    gayoor
## 11   215 india       reopening    accessible  testing
## 12   246 migrant     workers      undermining physical
## 13   261 strongly    worded       passionate  observations
```

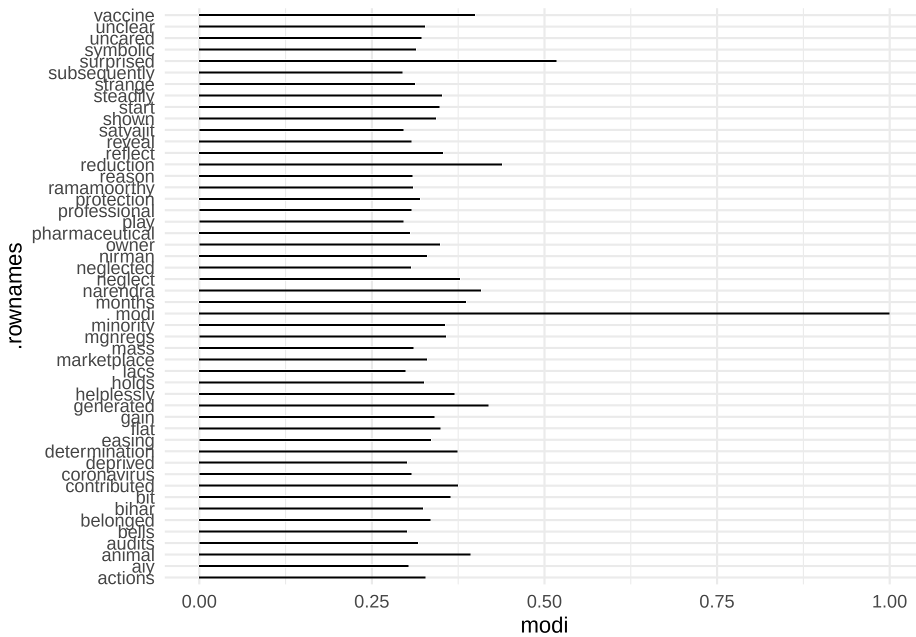
Word similarities

Now we shall essentially convert these words into vectors and see how similar they are on the basis of - How frequently do they occur in the same context or within a sentence of each other.

We will now check words that are similar to the word **Modi** and bring out the top 50 closely correlated words. We can notice that words that are often mentioned in the context of **modi** appear to be - **surprised, outlined, reassure, providing, mismanagement, aadhaar, offerred, opportunities**. Some terms like **surprised** seem to be in line with anecdotal evidence of how the prime minister directs policy at times - in surprising moves. Words like **providing, reassure** suggest that the prime minister has taken action regarding this crisis.

```
##           modi      surprised      reduction      generated       narendra
##      1.0000000      0.5175068      0.4382753      0.4187787      0.4078502
##        vaccine         animal         months        neglect    contributed
##      0.3996068      0.3930762      0.3860814      0.3773198      0.3749539
##  determination     helplessly            bit        mgnregs       minority
##      0.3740319      0.3698159      0.3639314      0.3569366      0.3559855
##        reflect       steadily           flat          owner          start
##      0.3527601      0.3513485      0.3494147      0.3483307      0.3478893
##          shown           gain         easing       belonged         nirman
##      0.3429381      0.3404632      0.3351734      0.3347612      0.3300565
##    marketplace        actions        unclear          holds          bihar
##      0.3296125      0.3273781      0.3271464      0.3256649      0.3241468
##        uncared     protection         audits       symbolic        strange
##      0.3217929      0.3196669      0.3164055      0.3137493      0.3120267
##           mass    ramamoorthy         reason         reveal   professional
##      0.3102876      0.3095673      0.3086962      0.3072613      0.3070423
##    coronavirus      neglected pharmaceutical            aiy          bells
##      0.3069518      0.3069158      0.3053290      0.3031864      0.3009449
##       deprived           lacs       satyajit           play   subsequently
##      0.3007853      0.2988081      0.2953505      0.2953356      0.2941496
```

```
## Warning: 'tidy.matrix' is deprecated.
## See help("Deprecated")
```

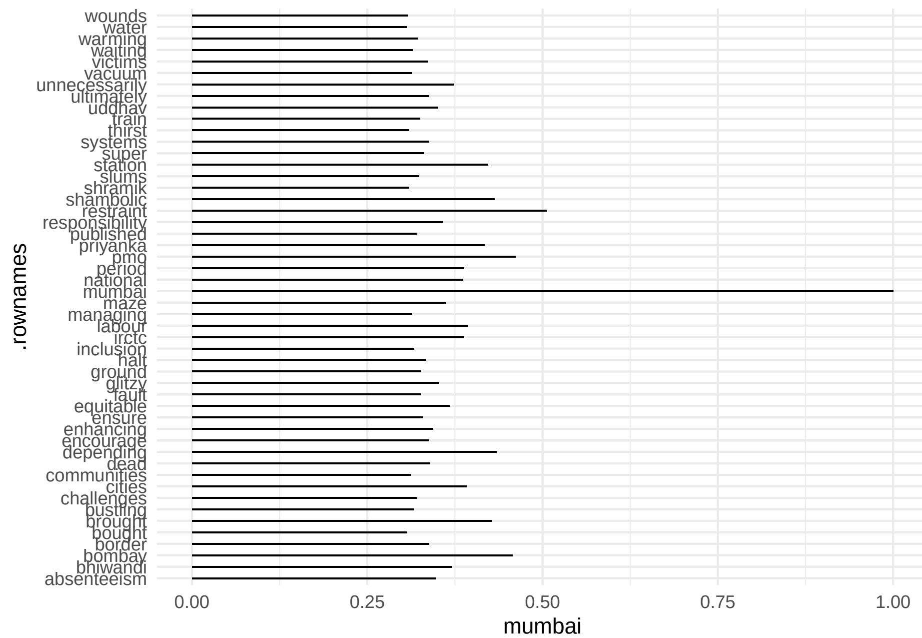


Now we see the top correlated terms for **mumbai** to get an idea about the context in which this city was mentioned as the migrant crisis unfolded. Words like **arrest, jail** suggest that the police was probably handling this issue by hard enforcement of lockdown norms. Words like **critisized** suggest that this city has been critisized for the way it handled this crisis. Note that we can make these judgements because analysis suggests a close correlation between these words to the main subject. So in essence, since the word **critisized** has been present within a few words of the word **mumbai** we can say it was probably referring to **mumbai** and not something else.

```
##         mumbai      restraint            pmo         bombay      depending
##      1.0000000      0.5063127      0.4614587      0.4573265      0.4343722
##      shambolic        brought        station       priyanka         labour
##      0.4316517      0.4274483      0.4222558      0.4176109      0.3932278
##         cities          irctc         period       national unnecessarily
##      0.3923932      0.3879903      0.3878940      0.3864918      0.3729191
##       bhiwandi      equitable           maze responsibility         glitzy
##      0.3700400      0.3679636      0.3626228      0.3584705      0.3518065
##         uddhav    absenteeism      enhancing           dead         border
##      0.3503430      0.3473395      0.3441686      0.3391745      0.3381475
##      encourage     ultimately        systems        victims           halt
##      0.3379318      0.3377582      0.3371529      0.3360302      0.3334402
##          super         ensure         ground          fault          train
##      0.3308369      0.3298977      0.3260425      0.3259468      0.3254716
##          slums        warming      published     challenges      inclusion
##      0.3236417      0.3223772      0.3210351      0.3210092      0.3164950
```

```
##      bustling        waiting       managing         vacuum    communities
##     0.3158299      0.3145329      0.3137902      0.3132575      0.3126908
##        thirst        shramik         wounds          water         bought
##     0.3099424      0.3097115      0.3075812      0.3061962      0.3061707
```

```
## Warning: 'tidy.matrix' is deprecated.
## See help("Deprecated")
```

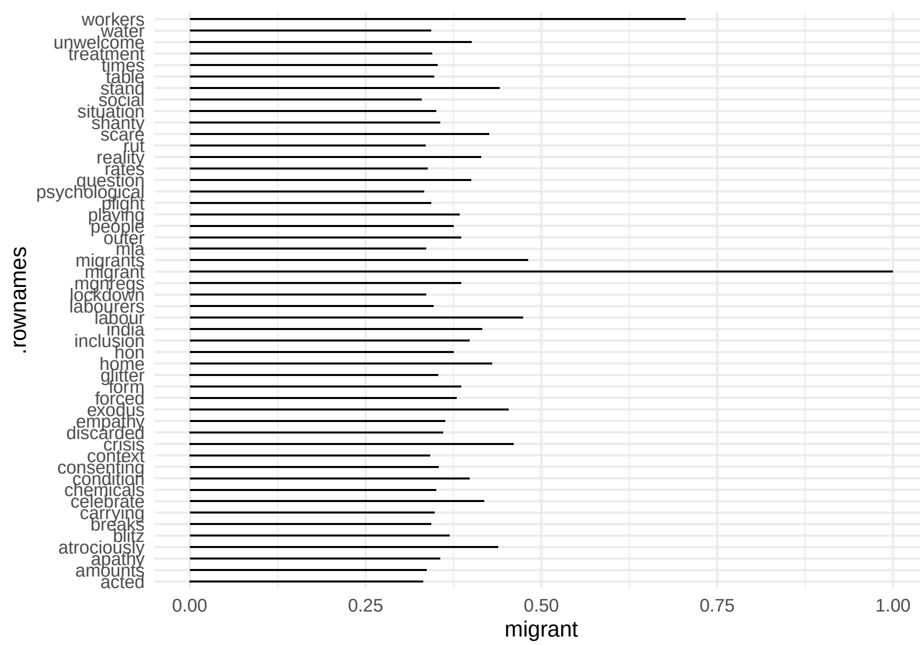


We will now see top correlated terms with the word **migrant** to see the relevant context. We can see that most descriptions of the migrant workers had words like - **tragedy, plight, aggravating, crisis, exodus, unemployment, lockdown** - which give us a view that the conditions of these people are quite horrible and this resulted mainly from the effect of the lockdown. Words like - **governance, delays, semiskilled, passivity** - imply bottlenecks in governance and administration regarding handling this issue.

```
##       migrant      workers     migrants       labour       crisis
##     1.0000000    0.7048118    0.4807049    0.4735181    0.4605119
##        exodus        stand  atrociously         home        scare
##     0.4534140    0.4402268    0.4382190    0.4295203    0.4254909
##     celebrate        india      reality    unwelcome     question
##     0.4184058    0.4156643    0.4138634    0.4009450    0.3997030
##     inclusion    condition        outer      mgnregs         form
##     0.3977806    0.3974949    0.3860645    0.3860130    0.3856370
##       playing       forced       people          hon        blitz
##     0.3833354    0.3793967    0.3750546    0.3748726    0.3691406
##       empathy    discarded       shanty       apathy   consenting
```

```
##     0.3626706     0.3603040     0.3561335     0.3557080     0.3534620
##       glitter         times     situation     chemicals      carrying
##     0.3533026     0.3523568     0.3505019     0.3499113     0.3480529
##         table     labourers     treatment        breaks         water
##     0.3471636     0.3467378     0.3445367     0.3430608     0.3429685
##        plight       context         rates       amounts           mla
##     0.3429000     0.3418471     0.3381601     0.3364791     0.3362329
##      lockdown           rut psychological        acted        social
##     0.3356980     0.3350790     0.3329768     0.3320390     0.3297252
```

```
## Warning: 'tidy.matrix' is deprecated.
## See help("Deprecated")
```

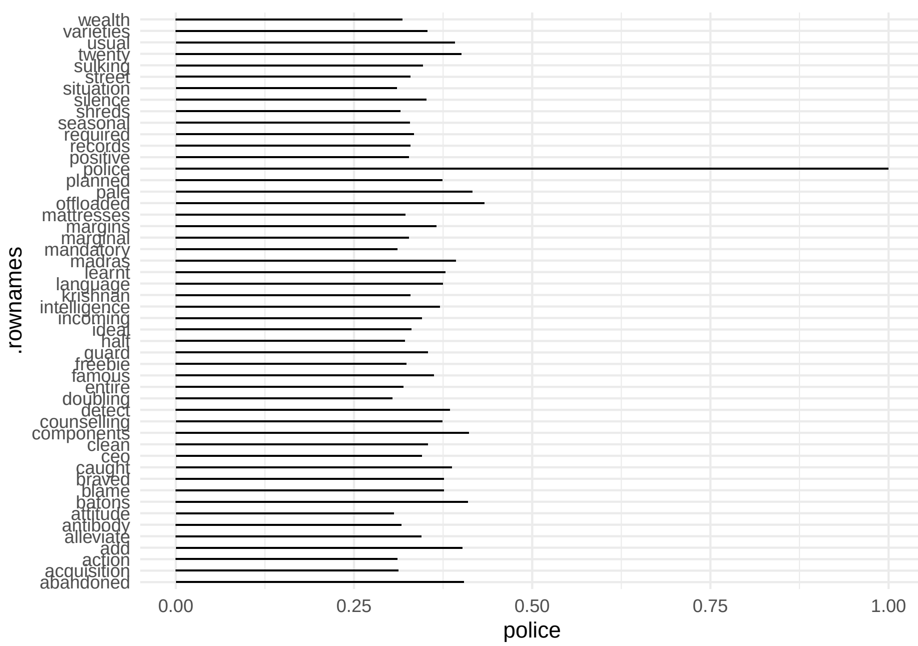


Now we look at top correlations with the word **police** to get an idea about how the police tackled these problems. Words that immediately stand out are - **batons, hawkers, violence, thousands, kilometeres** - which suggests that as hawkers and migrant workers were walking towards their cities, the police sparked violence by hitting them with batons (a probable judgement - not to be taken as fact).

```
##        police     offloaded          pale    components        batons     abandoned
##     1.0000000     0.4328430     0.4159509     0.4112612     0.4097914     0.4039869
##           add        twenty        madras         usual        caught        detect
##     0.4018132     0.4007953     0.3932753     0.3913439     0.3872070     0.3848601
##        learnt         blame        braved      language       planned   counselling
##     0.3781997     0.3763572     0.3761518     0.3746807     0.3740583     0.3737859
```

```
## intelligence      margins       famous        clean        guard    varieties
##    0.3705544    0.3657804    0.3623137    0.3537709    0.3537261    0.3532699
##      silence      sulking     incoming          ceo    alleviate     required
##    0.3514557    0.3465073    0.3453491    0.3450783    0.3448615    0.3337751
##        ideal      records       street     krishnan     seasonal     marginal
##    0.3307246    0.3292142    0.3291955    0.3291886    0.3280849    0.3273618
##     positive      freebie   mattresses         half       entire       wealth
##    0.3268171    0.3235474    0.3223522    0.3215703    0.3193382    0.3180540
##     antibody       shreds  acquisition       action    mandatory    situation
##    0.3165451    0.3148779    0.3122983    0.3111028    0.3106232    0.3100186
##     attitude     doubling
##    0.3056700    0.3040659
```

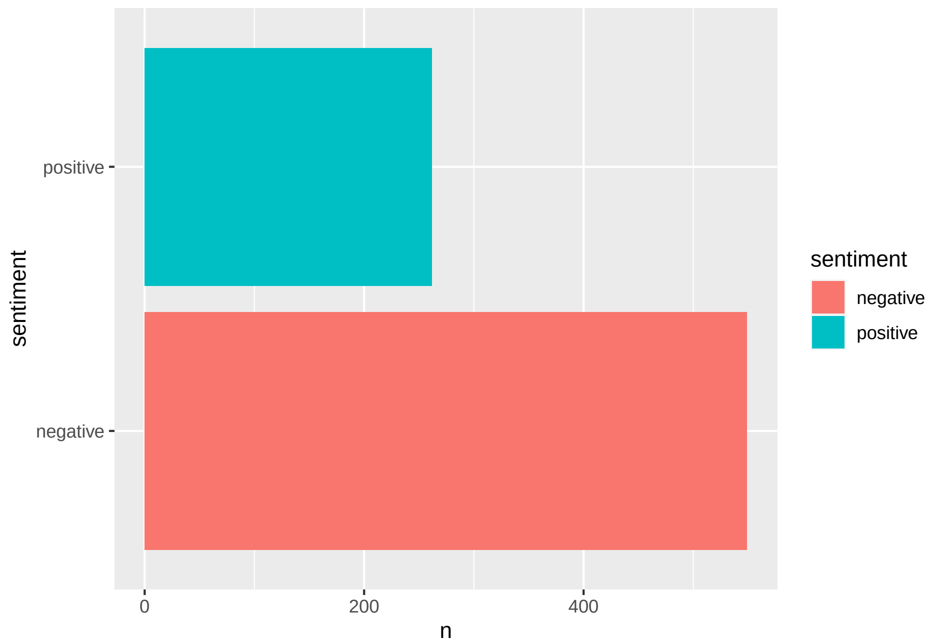
```
## Warning: 'tidy.matrix' is deprecated.
## See help("Deprecated")
```



Overall sentiment

Finally we can judge the overall sentiment of **TOI** coverage on this issue. Prsence of words like - plight, violence, struggling, suffer, unemployment - give this topic a more negative tilt as compared to usage of positive words like - relief, support, great initiative, praised.

```
## Joining, by = "word"
```



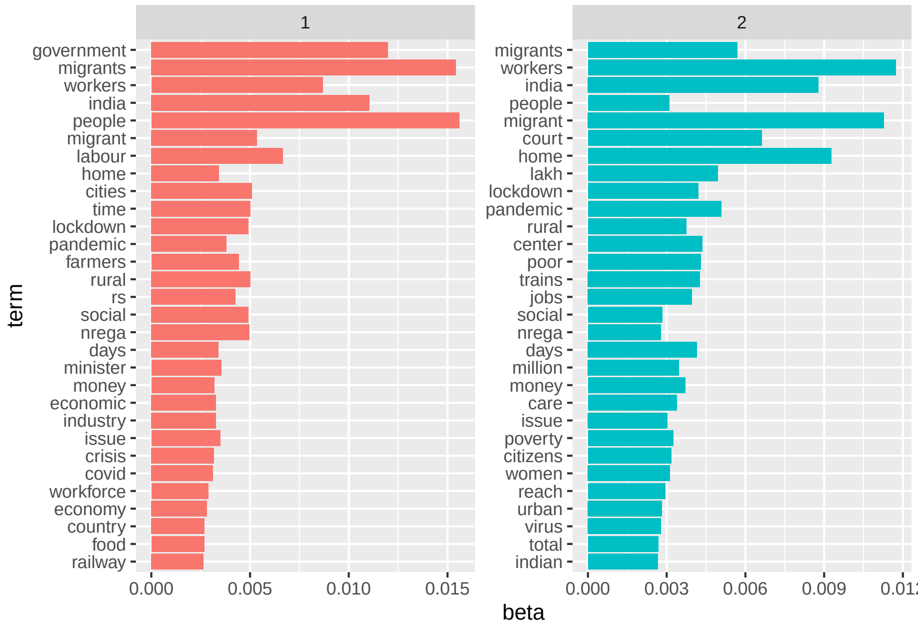
Topic analysis

Finally, we look at the broad topics on which the articles were largely focussed on within the **migrant workers** framework.

```
## Joining, by = "word"

## Warning: `tbl_df()` is deprecated as of dplyr 1.0.0.
## Please use `tibble::as_tibble()` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_warnings()` to see where this warning was generated.
```

We can see that broadly two topics have been in focus. The first topic has characteristic words like - **economy, economic, wages, jobs, crore, workforce** - suggesting a finance/economy focussed approach to the migrant issue. The second topic has characteristic words like - **trains, villages, welfare, political** - which suggests a more political and sympthetic view of the subject.


1
2
term
beta
government
migrants
workers
india
people
migrant
labour
home
cities
time
lockdown
pandemic
farmers
rural
rs
social
nrega
days
minister
money
economic
industry
issue
crisis
covid
workforce
economy
country
food
railway
0.000
0.005
0.010
0.015
migrants
workers
india
people
migrant
court
home
lakh
lockdown
pandemic
rural
center
poor
trains
jobs
social
nrega
days
million
money
care
issue
poverty
citizens
women
reach
urban
virus
total
indian
0.000
0.003
0.006
0.009
0.012