

Sample of the dataset

"Classic left strategy. Blame the other guy "

"#वायुप्रदूषण का समाधान 'स्विच ऑन' और 'स्विच ऑफ' करने जैसा नहीं है। यह निरंतर प्रयास है जो प्रदूषण को कम करेगा। #वायुप्रदूषण से निपटने के लिए सभी राज्य एजेंसियों और लोगों को इस प्रयास में "अगर सिगरेट पीने से टार जम जाता है सीने में तो अगर कुछ दिन और यही मौसम रहा तो आदमी टार मूतने भी लगेगा।"

" And it's also not that #oddeven is the only action being taken by Delhi govt . They are taking many other steps as well ."

"As if naman gusto ng mga taga probinsya dito sa Maynila. The lack of opportunities and liveable wage force them to come here. So again, just like us in the ci

" now you are talking, so where lies the root cause ??? everything produce pollution, even you or me also produces pollution ... stop using vehicles, stop fa

"People of Punjab and Haryana are suffering equally. No steps taken by the govt of Punjab and Haryana. Why is all the pollution level talks Delhi centric when

"Study--"levels of fine particulate matter--inhalable specks of liquids & solids that make up air pollution--increased by 11.5% from 2016-2018--Vulnerable people d

"DSGMC takes a humble step towards environment protection by installing Waste Management System at Gurdwara Sri Bangla Sahib 🙏 "

" Yeah,Your are right let BJP rule the Delhi then you will see the same situation that happens with economy ,it will happen with you 😊."

"Delhi Dy Chief Minister Manish Sisodia: Centre has made 63,000 machines (to curb stubble burning) available in 2 yrs. How will it help 26 lakhs farmers? With

" All adjacent mountains can be seen having trees. Only this portion had no trees but bushes. Because someone had cut the trees. Point is same - make house in

"फिर से, मैं सभी एजेंसियों अपील करता हूँ कि वे #वायुप्रदूषण से निपटने के लिए एक साथ मिलकर काम करें और लोगों को राहत दें तथा सस्ती राजनीति में शामिल न हों। #AirPollution"

"If I hear one more time that we can't afford the #GreenNewDeal, I will demand they come to my neighborhood & breathe in the dirty air. The daily average of

"Class Viith Student of Govt Boys Lower Secondary School HyderShah , Hyderabad Sindh (Rural) TandoJam speaking about #pollution we should control together on

"आँखों में जलन, सीनों में दर्द सा क्यूँ है | साँसों की तलाश में इंसान भ्रमता सा नज़र आता है| दिल्ली शहर में हर शख्स परेशान सा नज़र आता है || #DelhiAirEmergency #DelhiPollution"

$\alpha=500$ $\beta=6$ no. of high-fq. terms=372 Accepted and cleaned=1056/1100

"#delhiairquality writes says delhis pollution puts nation bad light"

"delhi police lawyers clash #tishazari court police vehicles vandalised"

"cant imagine feels like dilliwalli hello bangalore"

"# #airpollution"

"dilliwallahs unwilling accept role vehicular pollution easier blame farmers alter lifestyles"

"hi government delhi take governments haryana punjab court failing prevent crop burning"

"needs see scifi horrors delhi little 8 pm tonight sent buddy ashish anand nightmare governments help us"

"air pollution level increased gzb occupying top slot india itsaqi 470 severe category conferring fear thatair pollution impac

"emissions prohibiting states setting higher standards forced use lower federal ones lowered obamas fuel economy strictions ta

"nitrogen oide share 2wheelers 20 4wheelers 17 carbon monoide co highest vehicle types 43 politics election year"

"white men produced great art literature whiteness maleness superior western civilization largely product white men moved towa

"hello unpadh andhe bhakt bcoz stubble burning entire north india choking bcoz delhi talk like educated person atleast"

"environment pollution prevention control authority national capital region urgent directions regarding grap air quality delhi

"good area matlab"

$\alpha=500$ $\beta=6$ no. of high-fq. terms=372 Rejected and cleaned=44/1100

"pourquoi donc faits observes"

"friend"

"londiya bheju"

"46"

"140 permanent damage 107"

"hand understand played part defeating hitler"

"random ass wave sadness"

"1950"

"already making noise mate"

"addressing community programme bangkok thailand watch"

"shame"

$\alpha=1100$ $\beta=6$ no. of high-fq. terms=828 Accepted and cleaned=1079/1100

increasing α (more generalising -less strictness)

"politics polluted air seems sad poor innocent lives suffer ill motives #airpollution #delhiaemergency"

"climate protesters fires damage last decade done biggest industrial complexes country dont care much pollution doe

"bhai itne pollution achanak ye prakash kaise dikhne laga"

"thoda sa fog hai buss kaha h pollution baarish hojayegi sab sahi hojayega 100 oxygen bhi thick nahi hoti dhool mitt

"yes order inder devta today rain tom would one full page today new ad flash every metro stations reduce pollution

"manila quezon city cesspool pollution traffic noise congestion compared provinces"

"yes vehicle produce pollution thats pollution reaches high valueodd even scheme arises"

"lives equally important everywhere problem bjp bhakts aap power delhi thats reason pollution talks delhi centric"

"air pollution increase since 2016 caused almost 10k premature deaths across usmore 40 caalbuterol use went uphous

"air suction machine absorb pollution better idea must adopted state borders politics avoided 555yrs janam din gur

"dont think delhi good condition right indias economy delhis pollution anyway delhi like shithole days bad air n s

"stop drama show people done capacity stop pollution"

"dont teach things himachalis make sure city pollution free"

$\alpha=1100$ $\beta=6$ no. of high-fq. terms=828 Rejected and cleaned=21/1100

increasing α (more generalising-less strictness)

"pourquoi donc faits observes"

"friend"

"londiya bheju"

"46"

"140 permanent damage 107"

"random ass wave sadness"

"1950"

"quite"

"hahhaaha"

"madness enhancement e"

"poke tubers showing leaks videos"

$\alpha=500$ $\beta=10$ no. of high-fq. terms=96 Accepted and cleaned=1002/1100
increasing β (less generalising-more strictness)

"2 legs always better vehicle 3 wheeler gives pollution nature #tvseurogrip"

"could except concentrate creating pollution"

"sir fighting air pollution taking life"

"delhi capital india delhi capital world pollution hhhhhhhhh"

"#nodapl #waterislife #environment #pollution #fossilfuels #climatechange"

"people forget planet formed violent eruptions poisonous gasses filled atmosphere tectonic movements force never imagine yet somehow pollutio

"leaders responsive aware citizens also responsibility questioning govt ones chosen represent ask made suffer like action taken combating pol

"instead taking steps reduce pollution modi spotted vacation #thailand #delhiemergency #delhibachao"

"biggest cause ncr pollution toxic vehicular smoke stays air winters dun nothing till date surprisingly ministers pm also breathe toxic air unl

"traffic pollution unbearable old city new old absolutely shabby"

"electrical tech endless wonders world especially world dont want pollution noise depletion vital resources"

"sirjee faridabad pollution ka kuch karo yaha saas lena mushkil hota ja raha hai"

"conrad two people say pollution 76 different ways favour oil costhey open mind wide openall others say pollution bad"

"come notice huge pollution coming industrial area sampla close delhi stubble burning plastic waste burning delhi people care delhi states"

$\alpha=500$ $\beta=10$ no. of high-fq. terms=96 Rejected and cleaned=98/1100

increasing β (less generalising-more strictness)

"#theprintopinion"

"rytt ethe ta pataake hje v chl rhe ne"

"lots different opinions heres official hse pages states 107db"

"frog gay energy"

"labour end hospital car parking fees"

"healthcare isnt right"

"despite democrats impeachment sham presidents agenda working"

"ta ultrarich"

"weeks column banning cars major manhattan road hasnt led worse traffic elsewhere tell managing overlong todo list"

"ohhh asi knsi bt jo dm mn e oni bnnda asy b kr sakta koi imp bt"

"california cant save allow policies ruin rest america california save"

Comparison of values

Sl no	α	β	High fq terms	Accepted comments	Rejected comments
1	500	6	372	1056	44
2	1100	6	828	1079	21
3	500	10	96	1002	98