Tab 1

**Model Name**

Topic #: key words

* Notes, definitions, and context
* Suggested topic name
* Coherence Score 0-5

**NMF**

0,0,ha indic report mighti earth respons rapid

* Mighty earth rapid response report, always occurs together
* Usually about deforestation
* Ha, is used to talk about how many ha of land is affected
* Mighty Earth and subsidiary deforestation

1,1,respond complain alleg did

* General violations, often between workers but also about ignoring regulations
* Interpersonal relations and issues, use of the words respondent and “alleged complaints”
* General Violations

2,2,palm oil ran sustain report publish

* Palm oil, is being talked about in contexts of reporting products which are not certified
* Non RSPO Certified Products

3,3,eof pt plantat illeg investig forest estat supplier compani

* \*\*REPEATS

4,4,worker work casual union daili employ

* daily/casual workers
* Worker rights violations
* Union issues
* Workers Rights Violations, Human Abusers

5,5,alleg deforest bhd sdn link supplier

* Bhd and sdh are in company names
* The work link is being used between thing describe but also as a literally link to websites
* Some alleged deforestation?
* ???

6,6,park nation bukit ffb cpo eof

* Ffb sourcing illegally
* National parks are the main places
* Illegally Sourcing FFB from National Parks

7,7,right labour violat alleg issu

* Not following labor or human rights laws
* Labor Rights

8,8,clearanc biodivers forest impact leuser ecosystem

* Land designation and protection
* Clearance of land which was menat to be protected
* Land Clearing HCS/HCV

9,9,non compli palm product enter wilmar chain supplier suppli alleg

* Non RSPO Compliant Plam Products

10,10,mention greenpeac countdown final report

* “Mentioned in greenpeace final countdown report”
* Greenpeace final countdown report

11,11,rspo complaint compani plant land new member

* F
* RSPO Member rules/conflicts

12,12,forest clear hectar total pt report

* FOrest Claerence in protected areas, affecting Biodiveristy

13,13,pt gar jaya deforest chain report compani suppli

* Vague!!
* 0

14,14,communiti land villag local plasma compens desa

* Plasma land
* Shares of revenue
* Shares, Compensation, and Plasma for taken land
* 2

**LDA**

Topic 0: 0.012\*"worker" + 0.008\*"plantat" + 0.007\*"2" + 0.007\*"1" + 0.006\*"member" + 0.006\*"deforest" + 0.006\*"respons" + 0.005\*"forest" + 0.005\*"gar" + 0.005\*"conduct"

* Workers, union workers
* respondent/alleged, complainant issue
* Trash bin, related to workers
* Workers/trash bin
* 2

Topic 1: 0.008\*"plantat" + 0.007\*"plant" + 0.006\*"forest" + 0.005\*"claim" + 0.005\*"inform" + 0.005\*"certif" + 0.005\*"2" + 0.005\*"develop" + 0.005\*"use" + 0.004\*"commun"

* developers/development effects on community
* Failure to consult/inform community
* Misleading claims
* Poorly defined
* Companies and Developer being Misleading
* 2

Topic 2: 0.014\*"forest" + 0.013\*"plantat" + 0.008\*"respons" + 0.007\*"clear" + 0.007\*"deforest" + 0.006\*"rapid" + 0.006\*"concess" + 0.006\*"gar" + 0.006\*"right" + 0.006\*"ha"

* Poorly defined
* Bad
* Labor is kind of just showing up all over the place
* Saw labor, deforestation, payment, pollution etc.
* ??
* 0

Topic 3: 0.011\*"forest" + 0.010\*"deforest" + 0.009\*"earth" + 0.009\*"ha" + 0.009\*"mighti" + 0.008\*"clear" + 0.007\*"respons" + 0.007\*"plantat" + 0.007\*"concess" + 0.006\*"indic"

* Notes, definitions, and context
* Suggested topic name
* Coherence Score 0-5

Topic 4: 0.008\*"gar" + 0.007\*"plantat" + 0.007\*"local" + 0.006\*"deforest" + 0.005\*"alleg" + 0.005\*"violat" + 0.005\*"ffb" + 0.005\*"member" + 0.005\*"oper" + 0.004\*"agro"

* Notes, definitions, and context
* Suggested topic name
* Coherence Score 0-5

Topic 5: 0.017\*"plantat" + 0.011\*"eof" + 0.010\*"investig" + 0.009\*"gar" + 0.008\*"illeg" + 0.008\*"mill" + 0.007\*"supplier" + 0.007\*"riau" + 0.007\*"right" + 0.007\*"forest"

* Notes, definitions, and context
* Suggested topic name
* Coherence Score 0-5

Topic 6: 0.012\*"forest" + 0.008\*"clear" + 0.007\*"ffb" + 0.007\*"worker" + 0.007\*"gar" + 0.006\*"nation" + 0.006\*"alleg" + 0.006\*"concess" + 0.005\*"respons" + 0.005\*"clearanc"

* Notes, definitions, and context
* Suggested topic name
* Coherence Score 0-5

**LSA**

Topic 0: 0.330\*"land" + 0.228\*"forest" + 0.210\*"area" + 0.193\*"plantat" + 0.162\*"right" + 0.161\*"principl"

Topic 1: 0.241\*"grower" + 0.200\*"us" + -0.196\*"forest" + 0.179\*"small" + 0.163\*"river" + 0.159\*"holder"

Topic 2: 0.337\*"plantat" + 0.247\*"eof" + -0.227\*"land" + 0.207\*"mill" + 0.188\*"illeg" + 0.173\*"investig"

Topic 3: 0.260\*"worker" + 0.221\*"ioi" + 0.219\*"land" + -0.187\*"forest" + 0.170\*"complaint" + 0.169\*"rspo"

Topic 4: 0.674\*"worker" + 0.240\*"employ" + 0.207\*"casual" + 0.205\*"daili" + 0.187\*"provid" + -0.171\*"land"

Topic 5: 0.285\*"deforest" + 0.272\*"respons" + 0.260\*"earth" + 0.257\*"mighti" + 0.245\*"rapid" + -0.204\*"land"

Topic 6: -0.467\*"land" + 0.231\*"ioi" + -0.190\*"gar" + 0.182\*"rspo" + 0.181\*"area" + 0.174\*"forest"

**LSA TFIDF**

* 0 and 2 hold the most documents but we didnt like them, they seem like the “trash bins” of the models
* We suspect it could be the word land
* Or area

Topic 0: -0.227\*"eof" + -0.181\*"plantat" + -0.168\*"forest" + -0.155\*"illeg" + -0.147\*"gar"

* Eof = eyes on the forest
* “Might erath rapid”
* random!!
* ???
* 0

Topic 1: 0.256\*"eof" + -0.216\*"mighti" + -0.215\*"rapid" + -0.214\*"earth" + -0.197\*"ha"

* “Mighty earth rapid”
* Deforestation
* Ha
* Deforestation, mostly MIghty Earth
* 5

Topic 2: 0.214\*"land" + 0.181\*"right" + 0.178\*"worker" + -0.161\*"rapid" + -0.159\*"mighti"

* Complainant and respondent repeated
* 0

Topic 3: -0.358\*"park" + -0.333\*"cpo" + -0.316\*"ffb" + -0.267\*"nation" + -0.266\*"bukit"

* National parks
* Sourcing FFB illegally and from national parks
* Contaminated FFb
* Illegal Sourcing of FFB from National Parks/Contaminated FFB
* 5

Topic 4: 0.409\*"biodivers" + 0.380\*"leuser" + 0.340\*"clearanc" + 0.321\*"ecosystem" + 0.300\*"impact"

* leuser ecosystem is a protected rainforest
* Envirmentment Impact
* Environment and Biodiveristy impact around leuser ecosystem
* 4

Topic 5: -0.218\*"link" + -0.194\*"jaya" + -0.189\*"suppli" + -0.187\*"chain" + -0.166\*"greenpeac"

* Deforestation but also greenpeace
* Suplliers, supply chain
* Duplicates : 337, 338
* Deforestation mostly Greenpeace
* 3

Topic 6: 0.439\*"worker" + 0.208\*"labour" + 0.195\*"union" + -0.190\*"land" + 0.175\*"employ"

* Workers rights
* Unions
* Land, plasma, local community
* Plasma seems to be an outlier
* Workers Rights/Unions and Violations
* 4

**Vocab**

**High Conservation Values (HCVs)**

**LSA Uni-Bi-Gram**

Topic 0:eof, illegal, forest, plantation, gar, plantations, grievance\_handling, investigations\_engaged, pla, sarimas\_bukit

* EOF: Eyes on the Forest (report)
* Repeat grievances 252, 253 and 308-316
* Sarimas Bukit is a mill
* PLA has to do with the magistrate decree
* “Embezzing funds and fraud” agreement violations
* Animal abuse (orangutans) and illegal logging did
* Clearing forest in protected ecosystem, land clearing, road damage
* Illegal plantation on peatlands in protected ecosystem
* Illegal palm oil plantations and sell of illegal fruit”
* Possibly take out the word “plantation”, “RAN”, “GAR”, “EOF”
* Plantations on Protected Nature Reserves,
* \*SUBTOPIC: Illegal Palm Oil Plantations and trade of Illegal Fruit
* 2

Topic 1: mighty\_earth, rapid\_response, ha, indications, deforestation, concession, total, eof, illegal, sep

* Sep is just sepember, take out dates
* Concession: granting a company the right to operate
* Lots of “Tree cover loss” and “land clearing”
* Also using area in a wildlife reserve, taking land from farmers, expanding without permission
* Compensation for land
* RSPO code of conduct?
* Could be two topics maybe
* Deforestation and Stolen Land
* 4

Topic 2: national\_park, cpo, ffb, contaminated, bukit\_tigapuluh, tesso\_nilo, tiger\_corridor, bukit\_betabuh, alleging, number

* Tesso Nilo and Bukit Tigapuluh are National Park
* Duplicate document 195-196 and 226-227
* Bukit Betabuh and Tiger Corridor are a nature habitat
* “Restrictive competition practices”
* “Unsustainable sources”
* Bought FFB illegally and sold to RSPO members, contaminating CPO
* Conflicts with RSPO standards
* Negative press for palm oil? Then still uses palm oil in production?
* Planting trees outside HBU, dust affecting village health, air pollution, liquid waste, river pollution
* Doc 223 feels like it should be in worker topic
* Illegal FFB and Planting outside HBU
* SUBTOPIC unsustainable sources and RSPO standards conflicts
* 3

Topic 3:biodiversity, leuser\_ecosystem, workers, clearance, allegations, impact, rapid\_response, mighty\_earth, indications, forest

* Leuser Ecosystem is a vast rainforest in Indonesia
* “Mighty Earth Rapid Response”
* Worker rights, extra hours, compensation
* Labor laws, worker violations
* "Human Rights Defenders”, Restricting access to water and seizure of land
* Relationship with community, violating agreement with community
* Another off topic doc 4: Biodiversity and forest conservation (2)
* Doc 19 off topic: land clearing/deforestation (4)
* Doc 30 seems like topic 2 planting outside HBU
* Doc 16: seems off-topic air and noise pollution
* Worker rights, property rights, and community rights
* SUBTOPIC first clearing/HBU rules broken
* 3

Topic 4: biodiversity, clearance, leuser\_ecosystem, impact, forest, linked, potential, rayatama\_jaya, krida\_dharma, kahuripan

* Duplicates 173, 224 and 225
* Allegations of forest clearance, deforestation
* Forest clearance in high eco-diversity area
* Withdrawn from RSPO
* Doc 356 and 357 off-topic non-compliant palm products
* Doc 207 seems like topic 3 worker rights
* Doc 381 off-topic makes no sense
* Deforestation in High Biodiversity Areas
* 4.5

Topic 5:biodiversity, clearance, linked, leuser\_ecosystem, workers, rayatama\_jaya, kahuripan, krida\_dharma, potential, supply\_chain

* Rayatama Jaya is a plantation co
* Kahuripan is a place
* Krida Dharma: fulfilling ones duties and obligations
* Plasma allocation not received
* Deforestation (9)
* River contamination (1)
* Doc 146 kinda fits but also about villager protest
* Doc 355 seems like duplicate from topic 4
* Doc 200 environmental destruction/deforestation :(
* Plasma Allocation and Deforestation
* 3

Topic 6: released, 2018, public, september, case, workers, daily\_casual, studies, employment, union

* Duplicates 298-299 human rights abuses?
* Duplicated 373,374,375: (Public Report on this case was released on 19 September 2018.)
* Land clearing before assessment, clearing forest peatlands
* Illegally occupying property, land issue
* Forest destruction
* Lots of greenpeace
* Doc 335 should be in topic 2
* Doc 144 and 163 literally says nothing useful
* Land and Property Issues/Deforestation
* 3

Topic 7: workers, 2018, released, public, september, case, daily\_casual, indofood, committing, rayatama\_jaya

* Rayatama Jaya is a plantation co (maybe remove)
* Daily Casual workers are temp workers
* Worker rights/termination
* Land clearing outside Plantation Business license boundary
* Labor related issues, wages
* Embezzlement of retired workers payments
* Doc 57: ??? invasion forcible entry?
* Workers Rights
* 5

**LLM/Bert Fuzzy C-Means Uni-Bi-Gram**

**Groups are correct and doing good at topicing, but keywords seem like the wrong ones**

Average Score: 4.0 (T H)

Topic 0: mightyearth, ha, rapidresponse, concession, deforestation, cleared, forest, total, clearance, reportedindications

* Duplicates: 337-338
* Deforestation
* Orangutan habitat destroyed, land clearing
* Destroying rainforest
* Peatland (important for biodiversity) destruction and development
* Deforestation
* 5

Topic 1:complainantalleges, complainant, land, labour, rights, allegations, communities, members, complaint, bhd

* Duplicates: 173, 224 & 225, 252-253
* “BHD” Remove, just company suffix like co
* Disagreements over land use
* Expanding plantation into peatland
* Developing on peat area, lowland and rainforest, High carbon stock areas
* Doc 276 might be better in Topic 0, elephant habitat deforestation
* Overall topic 1 feels similar to topic 0, could be combined
* Peatland Violations
* 4

Topic 2:area, plasma, indonesia, forest, village, hcv, local, waste, permit, operations

* Duplicates:
* HCV: critical ecological areas, high biodiversity
* Unsustainably sourced FFB
* “Local violation standards”
* Withholding documents
* Comply with national law
* Stakeholders and small shareholders
* Lots of “allegations”, “appeals”, “consent”, “audits”, “transparency”, “regulations”, “dispute” etc
* Commercial violations
* “Member violated principles”
* Mostly government related formal complaints
* Dc 37 outlier: has every topic possible in its summary
* Doc 143 could be in a better topic
* Administrative Violations? Its giving law? Administration? Corporate?
* Compliance and Permit Conflicts
* 3

Topic 3: land, ffb, sustainable, rspos, newplanting, cpo, certification, communities, products, 2013

* “2013” should go fr
* Workers rights violations, wages, unjustifiably dismissed
* Union, labour issues
* Worker deaths
* Workers Rights
* 5

Topic 4:gar, plantation, eof, forest, mill, supplier, plantations, greenpeace, illegal, fruit

* Duplicates: 215 and 216, 308, 309, and 310
* Landowner complaints
* Stolen fruit
* greenpeace , land clearing
* GAR and Greenpeace audits
* Buying from destructive third party supplier
* Remove “GAR” and “Greenpeace” and “Mighty Earth”
* 'mightyearth''greenpeace'
* Greenpeace, GAR, and Mighty Earth
* 5

Topic 5:deforestation, gar, detected, internalmonitoring, ran, 2022, sawit, supplier, indications, raya

* Duplicates:
* Sawit: mushrooms that grow on empty FFB
* Raya: a company
* Work rights/labour related issues
* Doc 0: looks like topic 3
* Most look similar to topic 3, could potentially be combined
* Labour and Union Rights
* 5

Topic 6: plasma, indonesia, land, hcv, papua, yet, westkalimantan, village, districts, digoel

* Duplicates: 195, 196, and 234, 226-227
* Papua and West Kalimantan are places
* Digul is a river in indonesia
* Unsustainable FFB sources
* non -compliant products
* Mb shortening from MOI Foods
* Misleading
* Lots of “sustainable”
* Criticizing palm oil and claiming sustainability
* More GAR?
* Unsustainable FFB
* 4

Topic 7: labour, rights, violations, transnational, members, tpols, humanrights, reports, issues, violation

* Duplicates:
* TPOLS: Labour Union
* Human Rights Defenders
* Land clearing
* Restrictive Competition practices
* Private Property Infingement
* Unsustainable operations
* Air and noise pollution
* KKPA is a credit scheme to help small palm oil farmers
* Local government affairs
* Invaded farmland, small farmers, harassment, local police
* Land grabbing
* Local Law Enforcement
* 3

Topic 8:area, fire, forest, permit, sungai, decree, forestry, protected, indonesia, hgu

* Duplicates:
* HGU: land title to cultivate
* Human rights violations, safety, wages
* Workers rights, bribes
* Membership violations
* Exploited workers
* Child forced labour
* Human Rights Violations
* 4

Topic 9:products, ffb, cpo, nationalpark, contaminated, sustainable, eof, bukittigapuluh, tigercorridor, tessonilo

* Duplicates:
* Tssonoilo is a park in Indonesia
* Incests, pesticides
* Coconut plantation damage
* Cutting trees
* New planting on protected area
* Land clearing, land expansion
* Liquid waste, wastewater
* Toxic hazardous waste in natural reserve
* Air pollution, polluting river
* Trash Topic, Maybe: Pollution
* 2

**LLM/Bert Fuzzy C-Means**

Average Score = 3.3 (U, N)

Topic 0 Representation: ['land', 'communities', 'allegations', 'local', 'complainants', 'complaint', 'compensation', 'lands', 'process', 'dispute']

* Peatland and forest clearing
* Biodiversity impact
* Peatland Clearing and Biodiversity Impact
* 5

Topic 1 Representation: ['earth', 'mighty', 'ha', 'rapid', 'response', 'concession', 'deforestation', 'indications', 'total', 'hectares']

* Plasma disputes
* No transparency
* Not informing locals
* Plasma and other Compensations
* 4

Topic 2 Representation: ['river', 'ffb', 'national', 'sustainable', 'cpo', 'principle', 'waste', 'use', 'park', 'operations']

* Workers rights and unions
* Health issues
* Labor rights
* Payments
* Worker Abuses/Violations
* 5

Topic 3 Representation: ['workers', 'casual', 'union', 'daily', 'employment', 'bpjs', 'working', 'labour', 'termination', 'work']

* MIghty earth
* Deforestation
* MIghty Earth/Deforestation
* 5

Topic 4 Representation: ['deforestation', 'gar', 'chain', 'linked', 'mighty', 'earth', 'rapid', 'supply', 'response', 'jaya']

* Admin and other miscellaneous
* Admin/miscellaneous
* 3

Topic 5 Representation: ['eof', 'mill', 'forest', 'illegal', 'plantations', 'plantation', 'grievance', 'gar', 'investigations', 'riau']

* Land issues clearing
* Liabilities
* consessions
* Land Claims or wrongful acquisitions/mixed with random
* 3

Topic 6 Representation: ['land', 'hcv', 'rspos', 'plasma', 'npp', 'complaint', 'new', 'assessment', 'ioi', 'communities']

* Deforestation
* Graanpeace
* Gar
* Mighty earth
* Deforestation (Mighty Earth/GAR centered)
* 4

Topic 7 Representation: ['greenpeace', 'fire', 'jaya', 'indofood', 'suppliers', 'exploitation', 'ran', 'chain', 'supplier', 'media']

* Duplicates:
* Notes, definitions, and context
* Labour Rights
* 4

Topic 8 Representation: ['forests', 'leuser', 'clearance', 'forest', 'ecosystem', 'peatlands', 'acres', 'peat', 'ran', 'cleared']

* Payment, funds, bribes
* duplicates
* TRASH
* 0

Topic 9 Representation: ['labour', 'rights', 'violations', 'transnational', 'tpols', 'human', 'reports', 'members', 'limited', 'right']

* Fires, river
* Ffb sourcing
* ASS CHEEKS
* -1

**If land appears with clearing, then combine the world if not remove land from corpus.**

pLSA\_new

**pLSA**

**Topic 0: 0.924\*"chain" + 0.904\*"forests" + 0.891\*"clearance" + 0.855\*"sawit" + 0.751\*"supply" + 0.684\*"supply chain" + 0.662\*"west" + 0.651\*"products" + 0.582\*"leuser" + 0.548\*"west kalimantan"**

Notes, definitions, and context

Severe air and noise pollution | **waste water disposed, poluted river**

Environmetal, **deforestation**, farm destrpyed (**affests lifes of the community**)

Community not notified, threat to food security

Outlier: villager was shot during the protests

animals

* Deforestation, affect on communities
* Coherence Score 3

**Topic 1: 5.390\*"concession" + 5.266\*"earth" + 4.632\*"rapid" + 3.459\*"earth rapid" + 2.544\*"total" + 2.110\*"development" + 1.568\*"sdn" + 1.407\*"hectares" + 1.397\*"may" + 1.122\*"according"**

Notes, definitions, and context

**Land clearing, unlawful land activity, mighty earth (deforestation)**

**Company violate regulations**

Outlier: worker rights (3)

Animals (2), community development, incident between community and company → burned a company excavator

* Land clearing, unlawful land activity (Company Violations, Ecosystem Disruption, and Social Tensions)
* Coherence Score 3.5

**Topic 2: 1.047\*"complainants" + 0.977\*"planting" + 0.968\*"rights" + 0.922\*"informed" + 0.918\*"consent" + 0.834\*"criteria" + 0.808\*"npp" + 0.804\*"hgu" + 0.790\*"operations" + 0.772\*"conservation"**

Notes, definitions, and context

**Company violate regulations**, damage to coconut plantation, irrigation system, **deforestation**, **clearing land affecting flora and fauna, development of roads, blocked river flow,** again water, **overtaking farms**

Fresh water,

* Suggested topic name: Environmental and Regulatory Impacts of Corporate Development
* Coherence Score 2.5

**Topic 3: 2.228\*"mill" + 1.079\*"fruit" + 0.913\*"ptpn" + 0.913\*"peputra" + 0.892\*"tunggal" + 0.839\*"stakeholders" + 0.830\*"located" + 0.802\*"alam" + 0.802\*"investigation" + 0.781\*"case"**

Notes, definitions, and context:

* 10 grievances overall
* 277 document is incomplete
* Trade of illegal fruit (308, 309, 310, 311, 312… 316) → same summaries but different companies for each (looks weird)
* Coherence Score 0-5

**Topic 4: 3.580\*"rights" + 3.468\*"agro" + 2.777\*"labour" + 1.718\*"violations" + 1.622\*"final" + 1.587\*"reports" + 1.481\*"lestari" + 1.452\*"human" + 1.442\*"greenpeace" + 1.336\*"labour rights"**

Notes, definitions, and context

Human rights violations (Bribes)

Deforestation

Damaged road

* Suggested topic name: Worker Rights Violations
* Coherence Score: 4.3

**Topic 5: 4.187\*"workers" + 1.186\*"agreement" + 0.989\*"plasma" + 0.919\*"casual" + 0.919\*"casual workers" + 0.868\*"working" + 0.810\*"work" + 0.802\*"ioi" + 0.788\*"water" + 0.770\*"use"**

Notes, definitions, and context

**Worker rights, union violations, minimum wages**

Embezzling, fraud

Plasma allocation/plantation, waste

* Suggested topic name
* Coherence Score 0-5

Workers rights

Deforestation

Fauna and Flora (coconut plantation, animals) →

Clearing land/illegal development → farms destroyed

water/river pollution + air/noise pollution