

Prosjekt 3: Marked analysis with textembedding

Mina, Max og Hedda

1. oktober 2024

Sammendrag

Dette prosjektet utforsker bruken av tekstembedding i økonomisk analyse, med fokus på å følge trender over tid med tekster fra energiselskaper. Vi har anvendt unsupervised learning for å gruppere tekster etter temaer, og brukt KeyBERT for å trekke ut nøkkelord. Resultatene indikerer at tekstembedding kan være effektiv for å avdekke trendutvikling, men at nøyaktigheten kan påvirkes av små variasjoner i tekstinholdet.

1 Introduksjon

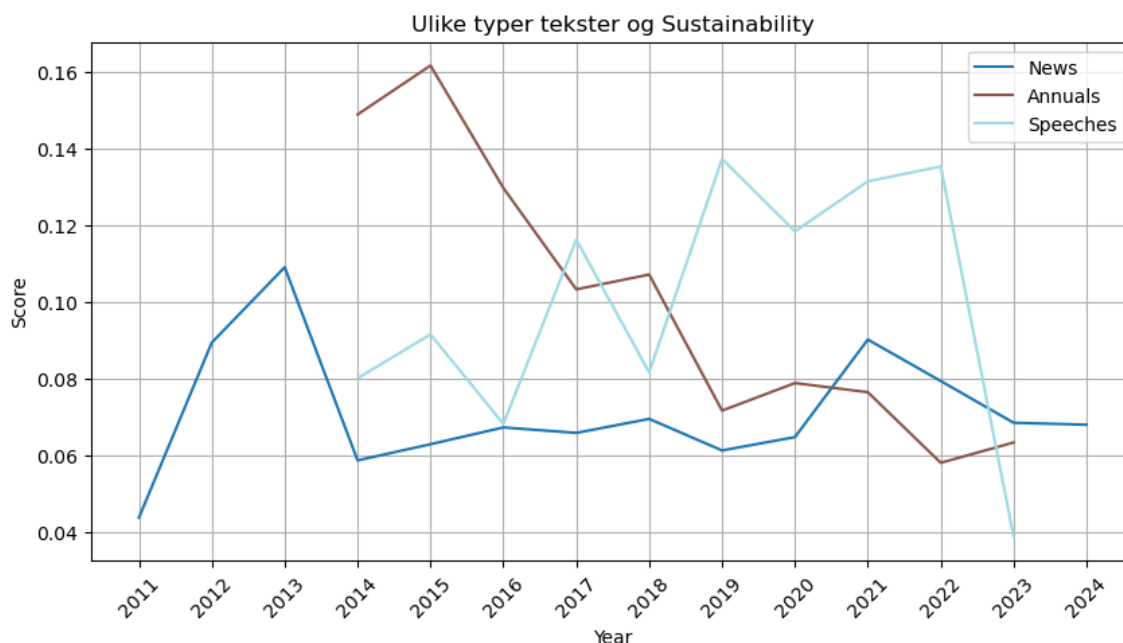
I dette prosjektet har vi utforsket bruken av tekstembedding for økonomisk analyse, med et spesielt fokus på energisektoren. Tilgjengelig tekstdata har vært essensielt for gjennomføringen av prosjektet. Gjennom nettskrapping har vi hentet ut nyhetsartikler, taler og årsrapporter fra nettsidene til flere ledende energiselskaper. De analyserte selskapene i prosjektet er Exxon, AkerBP, TotalEnergies, Equinor, Chevron og Shell, samt PetroChina, CNOOC og Reliance Industries.

Prosjektet fokuserer på tre hovedområder: identifisering av teksttyper som er best egnet for å avdekke og følge trender, metodene for å identifisere disse trendene og hvordan trendene utvikler seg over tid. Vi har undersøkt ulike teksttyper for å vurdere deres evne til å identifisere trender og om det er betydelige forskjeller mellom dem. For å identifisere trender har vi benyttet metoder som K-means og HDBSCAN for å gruppere tekstene, og deretter brukt KeyBERT til å trekke ut sentrale temaer. Til slutt har vi fulgt tre utvalgte trender over tid for forskjellige selskaper, der trendene er representert som centroids. Vi har beregnet tekstfordelingen mellom disse trendene over en gitt tidsperiode.

2 Hva slags tekstdata fungerer best?

Vi startet prosjektet med å undersøke om vi kunne trekke ut sammenlignbar informasjon fra ulike teksttyper. Shell ble valgt som case-selskap på grunn av tilgjengeligheten av forskjellige typer tekster på deres nettsider, inkludert taler, årsrapporter og egenpubliserte nyhetsartikler. Vår tilnærming innebar å beregne gjennomsnittet av embeddingene for hvert år og deretter måle avstanden mellom disse gjennomsnittene og ordet 'sustainability'. Målet var å vurdere om ulike teksttyper ga like resultater, og dermed indikere om de kunne brukes om hverandre i analyser.

Dataene ble hentet gjennom webskrapping ved hjelp av BeautifulSoup, en Python-pakke for å analysere HTML- og XML-dokumenter, og vi brukte deretter en transformer-modell for å embedde tekstene.



Figur 1: Avstanden til ordet 'Sustainability' for ulike teksttyper for Shell, snittet for hvert år

Basert på plottet ser vi at nyhetsartikler og taler viser en viss grad av likhet i enkelte år, men det er også betydelig variasjon, spesielt i de senere år. Dette antyder at disse teksttypene ikke alltid kan brukes om hverandre, og at både kontekst og årstall spiller en rolle i hvor nært de følger hverandre.

Vi observerte høy variasjon i embeddingene fra talene, noe som delvis kan forklares av det begrensede antallet tilgjengelige taler. Dette skapte utfordringer med å trekke konsistente konklusjoner basert på talene alene. Derfor valgte vi videre i prosjektet å fokusere på nyhetsartikler og årsrapporter, som ga et mer robust grunnlag for sammenligning og analyse.

Samtidig viser årsrapportene en mer stabil, men likevel avvikende score sammenlignet med både nyhetsartikler og taler. Dette kan tyde på at årsrapportene har en annen tilnærming til ordet 'sustainability', sannsynligvis på grunn av forskjeller i strukturen mellom årsrapporter og de andre teksttypene.

Oppsummert ser vi at valg av teksttype er avgjørende for analysens formål, hvor ulike teksttyper kan gi ulik score for samme tema (i vårt tilfellet ordet 'Sustainability'). Dette må tas i betraktning når man analyserer tekstdata for å sikre nøyaktige og sammenlignbare resultater.

3 Kan vi finne trender med unsupervised learning?

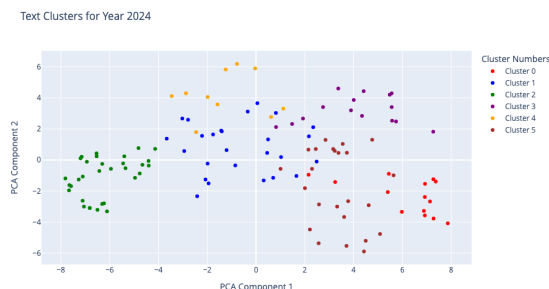
3.1 Clustering

Unsupervised learning handler i hovedsak om at en modell selv skal finne mønstre og sammenhenger i en mengde umerket data. Målet med dette i vårt tilfelle er å gruppere nyhetsartikler og årsrapporter som tilsynelatende handler om det samme, sammen. Dette kan brukes til å få et bedre overordnet bilde av hva et selskap fokuserer på i sine tekster om seg selv gjennom årene.

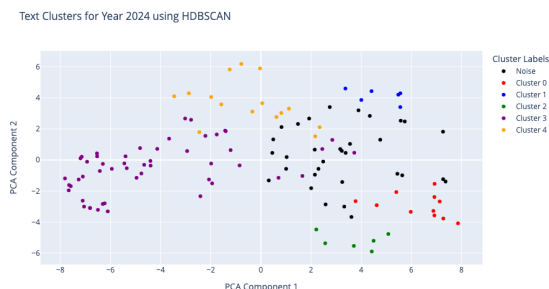
Før vi utfører clustering, benytter vi oss av en teknikk for dimensjonsreduksjon, Principal Component Analysis (PCA). PCA hjelper oss med å redusere kompleksiteten i dataene ved å komprimere informasjonen ned til et mindre antall dimensjoner, samtidig som de mest relevante komponentene beholdes. Dette er spesielt viktig når vi arbeider med høy-dimensional data, som tekstembeddinger, da høy dimensionalitet kan gjøre det utfordrende å utføre clustering effektivt.

Når dataene er redusert i dimensjoner ved hjelp av PCA, går vi videre til clustering. Vi har fokusert på to forskjellige clustering-algoritmer: K-means og HDBSCAN. K-means lar oss spesifisere et bestemt antall grupper (k clustere), der hver cluster er representert av en centroeide som ligger nærmest gjennomsnittet av de tilhørende punktene. HDBSCAN, derimot, grupperer basert på tettheten av punkter, og kan også identifisere uteliggere (støy) som ikke passer naturlig inn i noen cluster.

Vi valgte Equinors nyhetsartikler fra 1998 til 2024 som forskningsobjekt i denne delen av oppgaven. Under kan man se en clustering utført på de 79 artiklene som så langt er publisert i 2024, ved bruk av både K-means og HDBSCAN etter at dataene er blitt redusert i dimensjon med PCA. Selv om plottene ser relativt like ut, er gruppene ulike på grunn av forskjellene i hvordan de to algoritmene fungerer. Videre i analysen ser vi nærmere på nøkkelord som er knyttet til de ulike clusterne.



Figur 2: Clustering med K-means (K=6).



Figur 3: Clustering med HDBSCAN.

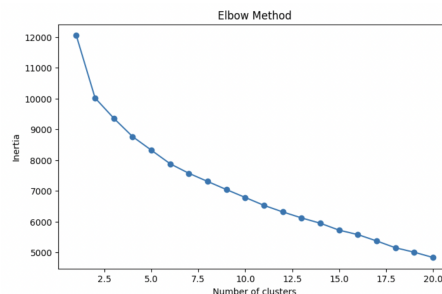
3.2 KeyBERT

Vi har brukt python-pakken 'keyBERT', som henter ut nøkkelord fra en gitt tekst. På omtrent samme måte som i andre deler av oppgaven vår, trekker keyBERT ut den semantiske meningen ved hvert ord, og ved hele teksten, ved hjelp av embeddings. Deretter blir cosine-avstanden målt mellom ordene og teksten. Ordene som ligger nærmest betydningen til avsnittet blir plukket ut som nøkkelord.

Etter clustering brukte vi altså keyBERT til å trekke ut viktige ord fra hver cluster. Målet med dette var å ha en oversiktlig måte å hente ut temaer fra clustrene uten å måtte lese alle tekstene. Om disse ordene faktisk er representative for alle tekstene i hver cluster, er varierende.

Språkmodellen vi har brukt til keyBERT kalles 'paraphrase-MiniLM-L6-v2'. Denne språkmodellen er rask og er skalerbar, slik at den fungerer godt på store datasett, slik som tekstene vi bruker.

Ved hjelp av 'Elbow-metoden', kan man finne et tilnærmet optimalt antall clustre gitt våre embeddings. Man vil altså identifisere et punkt der tettheten til punktene rundt hver centroeide ('inertia') begynner å avta saktere. På figuren til høyre har vi identifisert 6 clustre som 'optimalt' for året 2024 basert på formen til grafen. Dette varierer for hvert år. Likevel har vi valgt å holde oss til 6 clustre når det kommer til K-means, alle år, for enkelhetens skyld.



Når det kommer til nøkkelord, har vi også hatt muligheten til å velge antall selv og prøve oss frem. La oss velge 2024 som et eksempel igjen.

Med maks ett nøkkelord til hver av de 6 clustrene får vi:

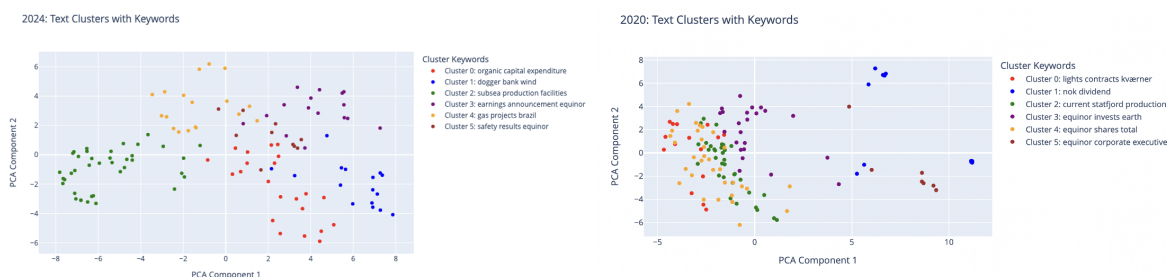
- Cluster 0: poland, Score: 0.5225
- Cluster 1: windfarms, Score: 0.4456
- Cluster 2: equinor, Score: 0.528
- Cluster 3: equinor, Score: 0.5439
- Cluster 4: turbines, Score: 0.3861
- Cluster 5: rig, Score: 0.4475

Med maks tre nøkkelord til hver av clustrene får vi:

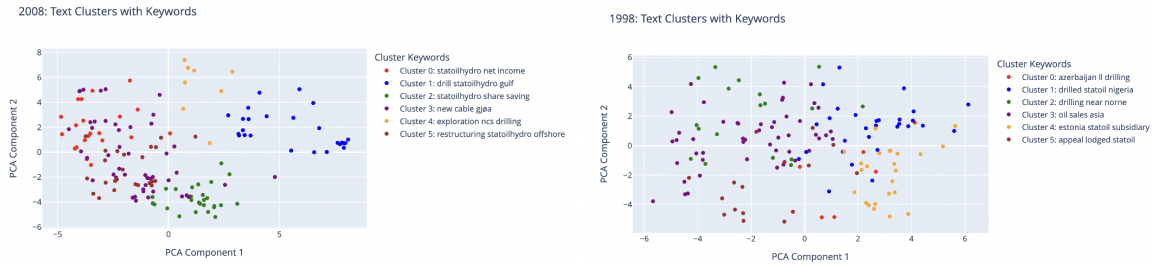
- Cluster 0: organic capital expenditure, Score: 0.6387
- Cluster 1: dogger bank wind, Score: 0.6571
- Cluster 2: subsea production facilities, Score: 0.5797
- Cluster 3: earnings announcement equinor, Score: 0.7278
- Cluster 4: gas projects brazil, Score: 0.5733
- Cluster 5: safety results equinor, Score: 0.7467

Som man kan se er det en betydelig forskjell i mengden informasjon man får angående clusteren, basert på antall nøkkelord. Et hovedtrekk vi fant etter å ha gjort det samme med mange ulike selskaper, er at om man bare lar modellen velge ett nøkkelord, er det sannsynlig at navnet til selskapet dukker opp for å representere flere av clusterene. Selv om det gir mening at selskapet snakker endel om seg selv, sier ikke dette oss så mye om de spesifikke temaene i tekstene. Med flere ord, eller 'n-grams', får man et mer nyansert innblikk i hva tekstene handler om. Allikevel vil det alltid kunne være utelligere blant tekstene som ikke blir representert like godt av nøkkelordene.

I eksemplene over ser man en 'score'. Dette er noe keyBERT beregner hver gang den finner et nøkkelord i en tekst, og viser hvor likt ordets semantiske mening er sammenliknet med hele teksten. Som man kan se, gir flere nøkkelord ofte høyere score. Vi velger derfor å fokusere på 3 nøkkelord til hver cluster fremover.



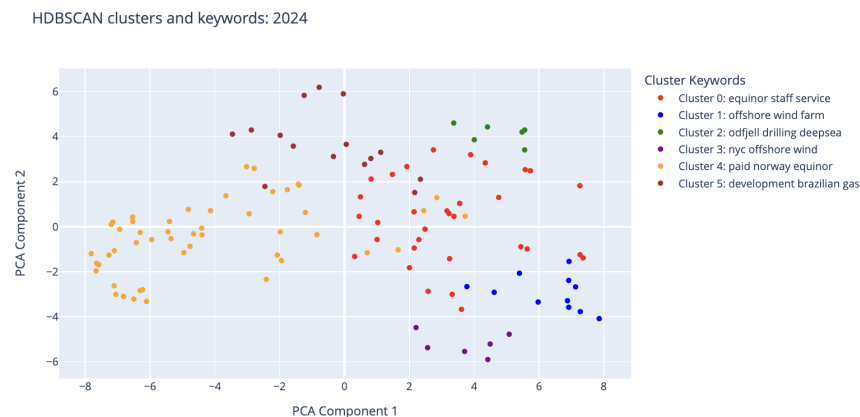
Figur 4: Tre nøkkelord, 2024 og 2020



Figur 5: Tre nøkkelord, 2008 og 1998

Figur 4 og 5 viser nøkkelord og clusters fra fire utvalgte år, 2024, 2020, 2008, og 1998. Fra 2024 har vi nøkkelordene: 'organic capital expenditure', 'dogger bank wind', 'subsea production facilities', 'earnings announcement equinor', 'gas projects brazil', 'safety results equinor'. Fra 2020 har vi: 'norges bank fixing', 'management statfjord øst', 'equinor share saving', 'collaboration equinor supply', 'ulrica equinor leadership', 'dogger bank eni'. Fra 2008 har vi: 'statoilhydro net income', 'drill statoilhydro gulf', 'statoilhydro share saving', 'new cable gjae', 'exploration ncs drilling', 'restructuring statoilhydro offshore'. Fra 1998 har vi: 'azerbaijan ll drilling', 'drilled statoil nigeria', 'drilling near norne', 'oil sales asia', 'estonia statoil subsidiary', 'appeal lodged statoil'.

Her er det mulig å se hvordan tematikken i nyhetsartiklene til Equinor har endret seg over tid. I 1998 var fokuset i stor grad på drilling etter olje, salg og akkseverd. I årene etterpå har tematikken i større grad dreid seg om sikkerhet og andre former for energi, blant annet vindkraft. Dette står i stil med økende fokus på fornybare løsninger i dagens samfunn.



Figur 6: Clustering og nøkkelord med HDBSCAN

Deretter testet vi det samme med den andre clustering-algoritmen HDBSCAN for året 2024. Som man kan se, får man ut litt ulike nøkkelord sammenliknet med K-means algoritmen. Grupperingen er gjort annerledes selv om plottet ser likt ut. Nøkkelord som 'development brazilian gas' fra HDBSCAN plottet og 'gas projects brazil' fra K-means, matcher ganske godt, mens nøkkelordene til de store clustrene varierer med de ulike algoritmene. Allikevel gir disse like nøkkelordene oss en indikasjon på at keyBERT faktisk klarer å hente ut hovedessensen fra clustrene.

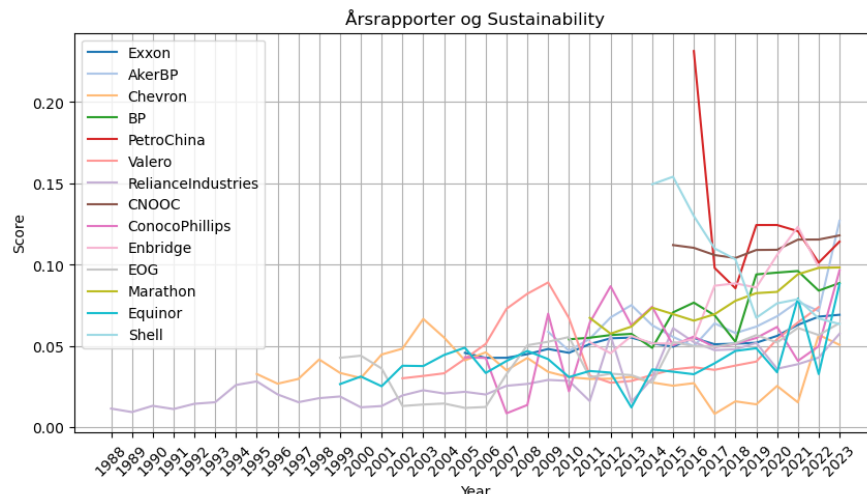
Denne samme idéen kan brukes på andre selskaper for å forsøke å dra ut informasjon fra nyhetsartikler, taler, årsrapporter og annen tekstdata. Ved bruk av K-means med flere clustre vil man kunne fange opp mer nyanserte temaer innad i tekstene.

4 Trender over tid

4.1 Sustainability

I denne analysen har vi regnet ut avstanden til ordet 'sustainability' i årsrapportene til flere store energiselskaper over tid. Grunnen til at vi fokuserer på dette ordet er at det har blitt et sentralt tema innenfor både bedriftsstrategi og i den offentlige samtalen, der det er blitt et økt fokus på miljø og samfunnsansvar. Grafen nedenfor (Figur 6) viser likhets-scoren mellom ordet 'sustainability' og de gjennomsnittlige tekstembeddingene for alle årsrapportene fra hvert selskap.

Selv om grafen kan virke noe rotete, kan vi likevel identifisere en generell økning i fokus på 'sustainability' i mange selskaper, spesielt etter 2015. Dette kan skyldes økende global bevissthet og reguleringer rundt klimaendringer.



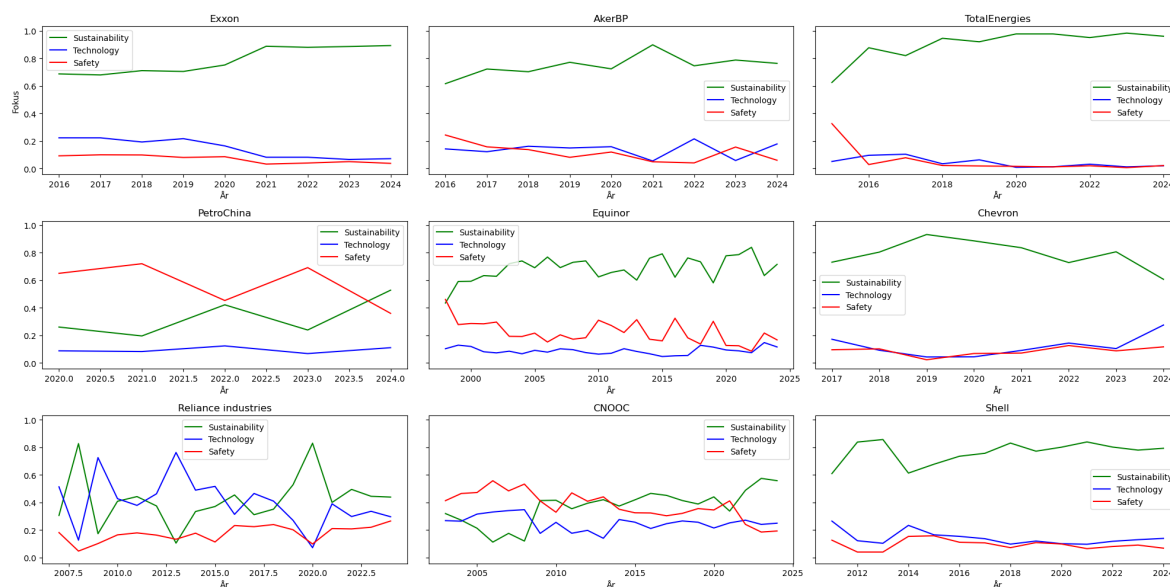
Figur 7: Likhets-score mellom ordet 'sustainability' og gjennomsnittlige embeddings for årsrapportene fra hvert selskap.

4.2 Egenvalgte trender

Vi har valgt ut tre sentrale temaer – 'sustainability', 'technology' og 'safety' – for å se hvordan fokuset på disse temaene har utviklet seg over tid. Scorene er vektinger som til sammen summerer opp til én, og representerer således det relative fokuset på hver av disse trendene. Analysen er basert på nyhetsartikler som selskapene selv har publisert på sine nettsider, da dette gir et innblikk i hvordan de ønsker å fremstille seg selv overfor offentligheten.

Figuren nedenfor viser tydelig at sustainability er den dominerende trenden for selskaper som Exxon, AkerBP, TotalEnergies, Equinor, Chevron og Shell. Dette reflekterer deres økte fokus på bærekraft og miljøansvar i tråd med globale krav og forventninger. På den annen side ser vi at PetroChina, CNOOC og Reliance Industries har lavere score på sustainability, noe som samsvarer med deres historiske fokusområder og mindre engasjement på miljøfronten.

Disse resultatene gir innsikt i hvordan ulike selskaper prioriterer mellom bærekraft, teknologi og sikkerhet i deres eksterne kommunikasjon, noe som kan være en refleksjon av deres strategiske fokus og tilpasning til eksterne krav.



Figur 8: Relativ score for trendene 'sustainability', 'technology' og 'safety' basert på selskapenes publiserte nyhetsartikler.

4.3 Hvordan skriver selskapene om bærekraft

Det er kjent at oljeselskaper har et større fokus på klima i dag, enn for 30 år siden. Om de ikke står for det selv, er det et stadig økt press fra konsumentene om at selskapene skal være mer bærekraftige. En måte å fremstå som mer bærekraftig er å uttale seg mer om klima. Vi ønsket å sjekke om det er mulig å bruke textembedding til å se på *hvordan* selskapene skriver om klima, og om dette har endret seg over tid. Vi har lest to artikler som handler om forskjellige måter å vinkle seg inn på temaet, som vist i variant 1 og variant 2.

4.3.1 Variant 1: fire trender

Vi har hentet inspirasjon fra forskningsartikkelen: Better poison is the cure? Critically examining fossil fuel companies, climate change framing, and corporate sustainability reports (Megura, 2022), som skriver om forskjellige måter å omtale bærekraft. Vi har plukket ut 4 måter å fremme interesse for bærekraft, som vi synes var relevante. Vi hentet ut deres eksempletekster, men brukte chatGPT for å gjøre tekstene mer generiske. Vi har brukt metoden for å finne avstand mellom flere centroids og sett på vektingen mellom dem, per år, per selskap.

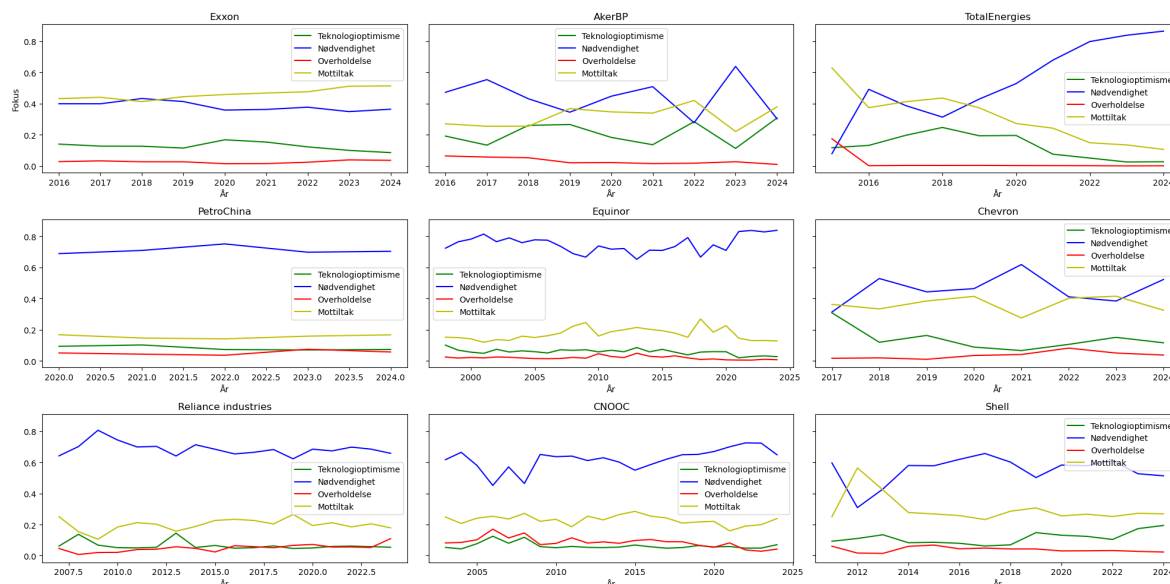
Trend nr. 1 er teknologioptimisme, der selskapene uttrykker at de har tro på at ny teknologi vil bidra til å løse klimakrisen. Trend nr. 2 er nødvendighet, der selskapet skriver om viktigheten av deres bærekraftige energikilde. Trend nr. 3 er overholdelse, der selskapet skriver om hvordan de gjør ting i henhold til regler og forskrifter. Til slutt har vi trend nr. 4 som tar for seg mottiltak. Her skriver selskapet om hvordan de investerer i andre grønne løsninger slik at de bedrer miljøet på en indirekte måte. Vi har laget fire centroids med tekster som representerer trendene;

1. "We believe that technological innovations will play a crucial role in resolving the climate crisis. Through sustainable solutions and advanced technologies, we are committed to reducing our emissions and protecting the environment for future generations."

2. "We are proud to be a vital supplier of renewable energy for society. Our commitment to harnessing sustainable energy sources ensures a greener future and supports communities in their transition to a low-carbon economy."

3. "By strictly adhering to regulatory standards, we ensure that our sustainable energy initiatives are executed responsibly and ethically, promoting both environmental stewardship and legal integrity."

4. “Our investment in cutting-edge technologies that reduce emissions underscores our commitment to sustainability. By supporting innovative solutions, we contribute indirectly to a greener future and promote responsible energy practices.”



I følge grafen er det vanligst å nevne nødvendigheten av selskapets fornybare kilde. Dersom vi endret litt på setningene ga det store endringer i grafene. Dette tyder på at det ligger mye usikkerhet i svarene.

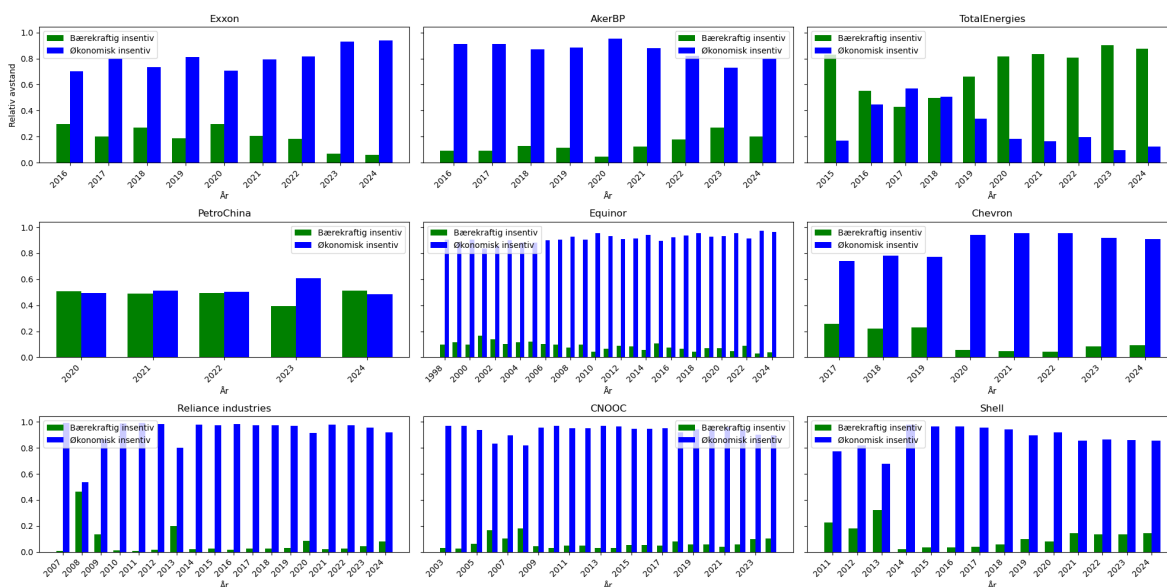
4.3.2 Variant 2: bærekraftig vs økonomisk insentiv

Vi ble inspirert av rapporten: Sustainability or Profitability? How Communicated Motives for Environmental Policy Affect Public Perceptions of Corporate Greenwashing (Terwel, 2013). Denne rapporten skriver om hvordan oljeselskapene mister troverdighet når de fremmer bærekraft for miljøets skyld, da de fort blir anklaget for greenwashing. I følge rapporten fremstod selskapene derimot mer troverdige dersom de heller skrev at de investerte i fornybar energi fordi det er økonomisk lønnsomt å investere grønt. Vi ønsket derfor å teste om tekstene deres har vridt seg mer mot et økonomisk insentiv de siste årene. Vi skrev to tekster som snakket positivt om miljø, der den ene fremmet et fokus på bærekraftig mens den andre argumenterte for økonomisk gevinst.

Vi valgte de to centroidene:

‘Our commitment to a sustainable future drives our investments in renewable energy sources. We recognize the crucial role the energy sector plays in addressing climate change, and we are dedicated to making a positive impact on our environment. By investing in solar, wind, and other renewable technologies, we aim to reduce our carbon footprint and pave the way for a cleaner, greener planet. Our efforts not only support global environmental goals but also ensure that future generations inherit a world where energy is both reliable and sustainable. Join us on our journey towards a more responsible energy future.’

‘We are always on the lookout for strategic investments that drive future profitability. Our latest move into renewable energy is part of this forward-thinking approach. As the world shifts towards more sustainable energy solutions, we see immense potential for growth and profitability in this sector. By diversifying our energy portfolio to include solar, wind, and other renewable technologies, we are positioning ourselves to capitalise on emerging market trends. This strategic investment not only enhances our financial performance but also aligns with global energy transitions. It’s a win-win for both our business and our shareholders.’



Vi forventet å få resultater som vred seg mer mot økonomisk gevinst de siste årene, men ser at den blå søylen i de fleste tilfellene er høyest. PetroChina er det eneste selskapet der fordelingen ble veldig jevn. Dette er lite troverdig da selskapet ikke er kjent for å ha fokus på fornybar energi, spesielt ikke for klimaets skyld.

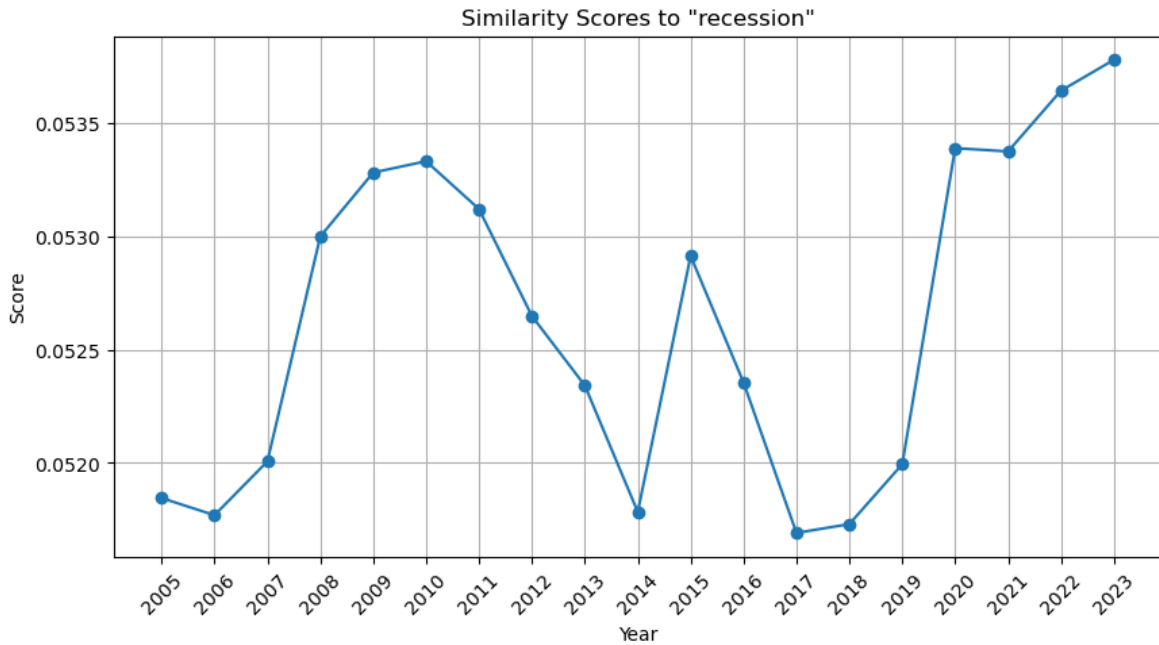
4.4 Hvordan fungerte metodene?

Vi oppdaget på både variant 1 og variant 2 at svarene endret seg ganske mye dersom vi endret litt på setningene. Små endringer i centroidene kunne gi ganske store utslag på grafene. Eksempelvis for variant 2 startet vi med to centroider som brukte forkortelsen CCS. Vi prøvde å endre CCS til "carbon capture and storage". Dette ga endringer i tallene, selv om det ikke burde ha gjort det. Derfor endte vi med å bruke ordet renewable energy sources. For variant 2 ble heller ikke svarene som vi forventet. Dette kan skyldes at det er noen fremtredende ord i centroiden som gjør at den nesten alltid scorer høyest. Av erfaring har vi funnet ut at et par ord kan være veldig dominerende, selv i en lenger tekst.

Igjen peker det på at denne metoden ikke fungerer så bra på mer kompliserte problemstillinger. Forsøket på å vinkle ett tema på forskjellige måter ga oss lite informerende svar. Text-embedding fungerte bedre da vi heller fokuserer på ett tema, for eksempel å sammenligne med ordet sustainability.

4.5 Andre interessante trender

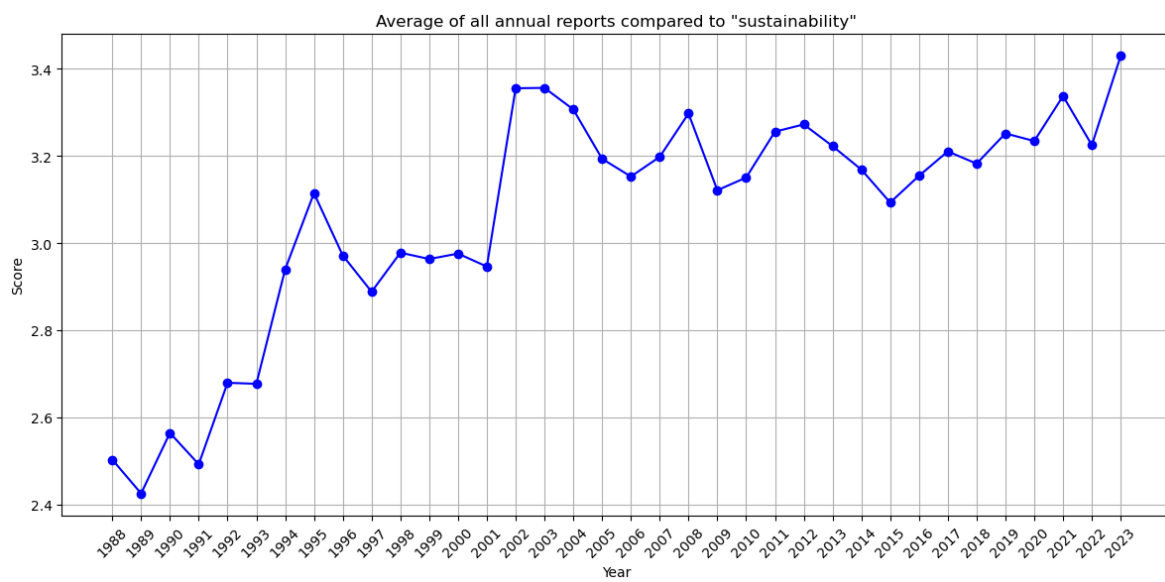
Vi har også undersøkt andre trender ved å bruke tekstembedding-metoder for å måle likheten mellom nøkkelord og tekstinholdet i årsrapportene. Vi vil her vise ulike bruksområder for tekstembedding, som kan bygges videre på i andre analyser. Figur 9 viser hvordan likhetsscoren mellom ordet 'recession' og ExxonMobil's årsrapporter har endret seg fra 2005 til 2023. Vi ser tydelig at embedding-scorene stiger i perioder med globale økonomiske utfordringer, som finanskrisen i 2008-2009, oljeprisfallet i 2015 og den økonomiske usikkerheten som fulgte pandemien i 2020-2023. Altså samsvarer embedding-analysen med den faktiske situasjonen vi så i verden.



Figur 9: Likhets-score mellom Exxon's årsrapporter og ordet 'recession' over tid.

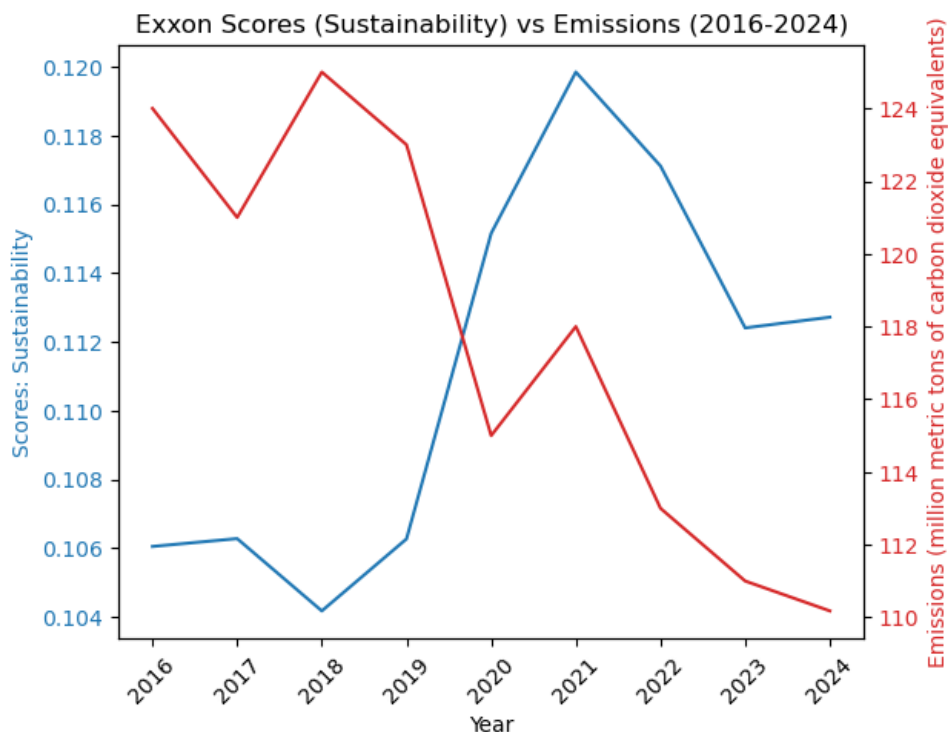
Figur 10 viser den gjennomsnittlige likhetsscoren mellom ordet 'sustainability' og alle årsrapporter samlet over flere tiår (gjennomsnittet av alle selskaper analysert). Her ser vi en jevn økning i embedding-scoren, spesielt etter år 2000, noe som indikerer et større fokus på bærekraft i årsrapportene for alle selskaper.

Disse analysene ved hjelp av tekstembedding gir verdifull innsikt i hvordan store selskaper har tilpasset sitt språk og endret fokus over tid. Økningen i embedding-scorer for ord som 'recession' og 'sustainability' viser at selskaper ikke bare reagerer på eksterne økonomiske og miljømessige press, men også aktivt inkorporerer disse temaene i deres offisielle kommunikasjon for å vise tilpasningsevne og ansvarlighet.



Figur 10: Gjennomsnittlig likhets-score mellom årsrapporter og ordet 'sustainability' over tid.

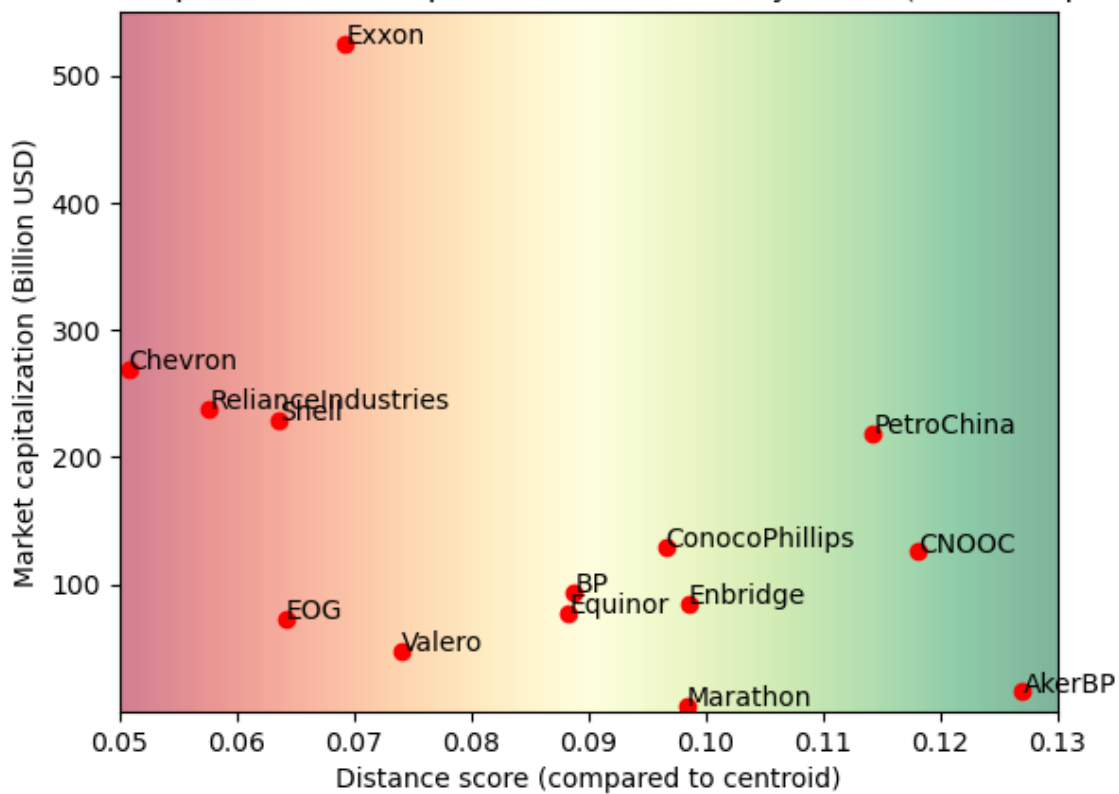
Figur 11 sammenligner utviklingen i score mellom 'sustainability' i ExxonMobil's årsrapporter og selskapets årlige utslipp fra 2016 til 2024. Som vi kan se, er det en tydelig motstridende trend mellom bærekraftsscoren og utslippene. Mens bærekraftsscoren har hatt en jevn økning, har utslippene hatt en gradvis nedgang. Denne observasjonen antyder at ExxonMobil har blitt mer bærekraftige i de senere årene, samtidig som dette også vises via embeddingene av årsrapportene.



Figur 11: Likhetsscore mellom Exxon's årsrapporter og sustainability, sammenliknet med selskapets årlige utslipp

Figur 12 viser sammenhengen mellom markedsverdi og bærekraftsscore basert på årsrapporter for de ulike selskapene vi har analysert i rapporten. Vi ser at ExxonMobil, som har den høyeste markedsverdien blant de undersøkte selskapene, plasserer seg på venstre side med en lavere bærekraftsscore. På den andre enden av skalaen finner vi selskaper som AkerBP, som har en lavere markedsverdi, men en høyere bærekraftsscore. Fargegradienten i bakgrunnen representerer en økende bærekraftsscore, fra rød (lav score) til grønn (høy score). Figuren viser hvordan større selskaper har lavere bærekraftspoenng sammenliknet med mindre aktører i bransjen, som kan reflektere utfordringene med å tilpasse bærekraftsinitiativer i store, etablerte organisasjoner.

Market capitalization compared to 'sustainability' score (annual reports)



Figur 12: Markedsverdi sammenlignet med bærekraftscore (Årsrapporter fra 2024)

5 Oppsummering

Vi har utforsket bruken av tekstembedding for økonomisk analyse, med særlig fokus på energisektoren. Vårt arbeid inkluderte analyse av ulike typer tekster, som årsrapporter og nyhetsartikler, ved bruk av usupervisert læringsteknikker som K-means og HDBSCAN for å identifisere trender og sentrale temaer.

Vi undersøkte flere energiselskaper, inkludert Exxon, Shell, Equinor og TotalEnergies, for å studere hvordan deres fokus på bærekraft, teknologi og sikkerhet har utviklet seg over tid. Resultatene viser at bærekraft har blitt en dominerende trend for mange av disse selskapene, spesielt etter 2015, noe som reflekterer det økende globale fokuset på miljøansvar.

Videre analyserte vi hvordan selskapene skriver om bærekraft og identifiserte fire hovedretorikker: teknologioptimisme, nødvendighet, overholdelse og mottiltak. Våre funn viser at omtale av nødvendigheten av selskapenes fornybare energikilder er mest utbredt. Vi forsøkte også å analysere om selskapenes kommunikasjon har dreid seg mot økonomiske insentiver snarere enn bærekraft, men resultatene her var usikre og viste stor variasjon basert på små endringer i tekstene.

Til slutt undersøkte vi andre potensielle bruksområder for tekstembedding i økonomisk analyse, som fokus på økonomiske resesjoner og bærekraft. Vi fant her blant at selskapenes omtale av økonomiske resesjoner korrelerte med globale økonomiske utfordringer, og at fokuset på bærekraft har økt jevnt etter år 2000.

Oppsummert viser vår analyse at tekstembedding kan være nyttig for å analysere økonomiske trender, men metoden kan være sensitiv for små tekstendringer og dermed mindre pålitelig enn mer komplekse analyser.

Referanser

Bart W. Terwel, Naomi Ellemers, Dancker D. L. Daamen (2013, 04.07). Sustainability or Profitability? How Communicated Motives for Environmental Policy Affect Public Perceptions of Corporate Greenwashing. Wiley online library. <https://onlinelibrary.wiley.com/doi/10.1002/csr.1327>

Matthew Megura, Ryan Gunderson (2022, 03). Better poison is the cure? Critically examining fossil fuel companies, climate change framing, and corporate sustainability reports. ScienceDirect. <https://www.sciencedirect.com/science/article/pii/S221462962100476X?via%3Dihub>