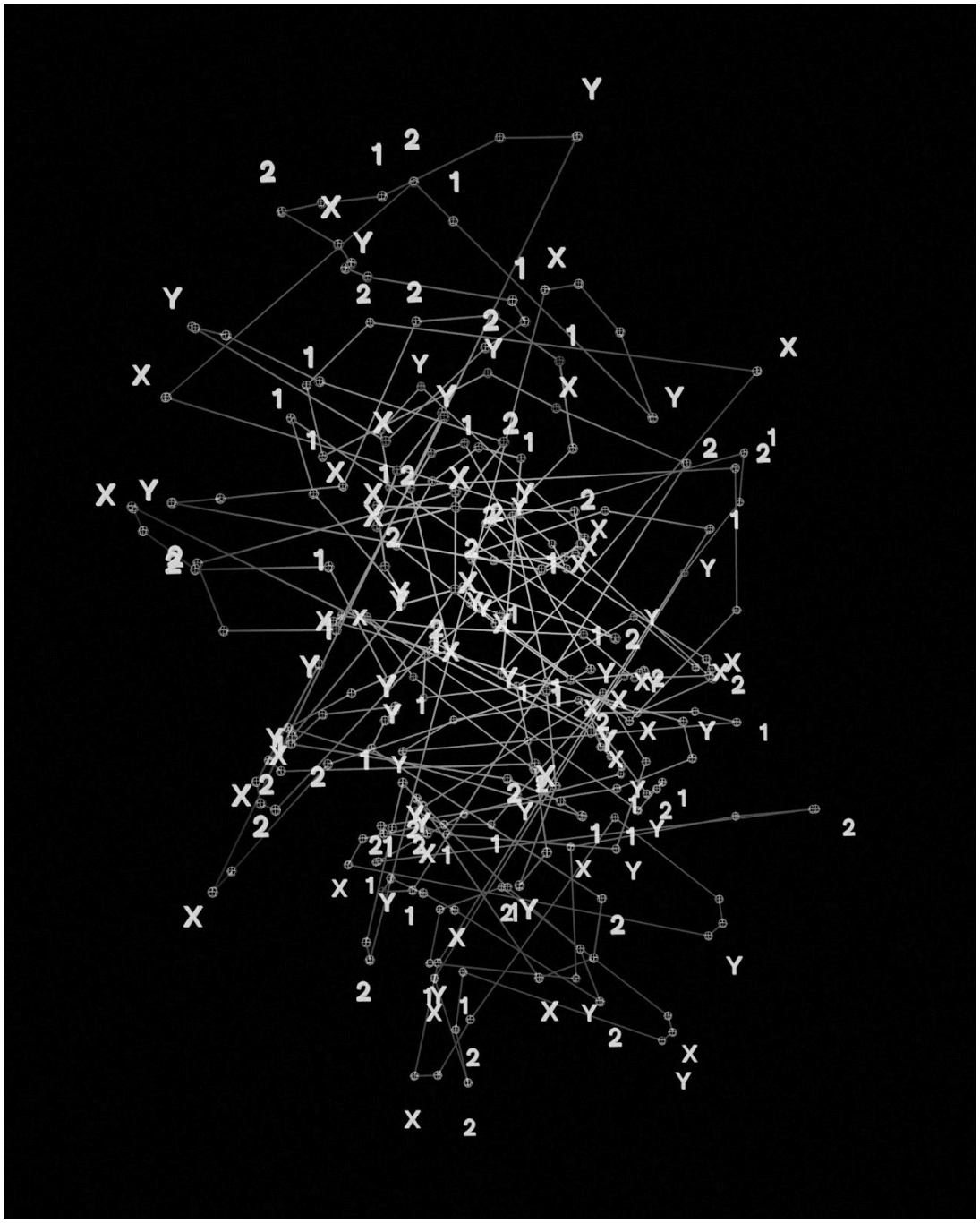
Kandidatspeciale: Machine Learning til brug i policyevaluering



Studerende: Mikkel Høj Bagger Larsson

Studienummer: 20203550

Vejleder: Jeppe Fjeldgaard Qvist

Antal anslag: 72128

Dato: 28.5.2025

Forord:

Indledende vil jeg gerne sige tak til Katie Henshaw og Rasmus Hove for at lægge friske øjne på specialet bidrage med gode input og rettelser af sammensatte ord. En tak skal også gå til Jeppe Fjeldgaard Qvist for god vejledning i øjenhøjde og CALDIS for at bidrage med data og vejledning når koden blev for rodet

Abstract

The purpose of this study is to examine and analyze the possibilities of advanced Machine Learning methods within political science. This study's subject field is the rise of Artificial Intelligence - also simply referred to as AI - tools in government and how the implementation of those fit within the framework of new governance models. When the Danish government introduced its plan for Artificial Intelligence, it introduced a round of financing for projects for municipalities to incorporate AI projects within their organisations. This sparked 140+ new Artificial Intelligence related projects throughout Denmark - around half of those projects retired after their trial periods ended. This study seeks to contribute to how these methods can be implemented in local as well as national governing tools. It utilises Twitter data from the Danish forum #dkpol - a forum used to platform discussions of Danish politics - from the years 2019 to 2021. The data has been embedded and classified in sentiment by a Danish BERT sentiment model. Afterwards a topic modelling was run with BERTopic, once again utilising a Danish fine tuned BERT model. The second part of the analysis runs three different Deep Learning models; LSTM, XGBoost and TextCNN. These models achieve overall high precision scores in their classifications, but the LSTM model achieves the highest. The third part of the analysis seeks to attribute these different steps in the analysis to the six steps in the policy cycle as described by Wayne Parsons and finds that the methods can be used throughout the model.

Indhold

1.0 Indledning: AI i det offentlige

Digitalisering af det offentlige er et emne, der fylder meget i den offentlige debat. Når denne digitalisering debatteres, er der med god grund et fokus på de store milliardprojekter, der har til formål at introducere nye store produkter, såsom MitID og E-boks, der omlægger, hvordan vi som samfund tilgår det offentlige. I den debat kan det være let at glemme de mindre projekter, der ikke nødvendigvis omvælter hele tilgangen i offentlig governance, men kan give et meget vigtigt løft i kvaliteten af service og beslutninger. Disse små ændringer er der dog et fokus på rundt om i administrationer i Danmark. I marts 2019 udgav Finansministeriet og Erhvervsministeriet den nationale strategi for Kunstig Intelligens. Denne strategi havde som målsætning at sætte Danmark på landkortet inden for Kunstig Intelligens samt sikre dansk udvikling inden for feltet. I deres vision for kunstig intelligens skriver regeringen:

> ”Algoritmerne skal sikre ligebehandling ved at være objektive, saglige og uafhængige af personlige forhold. […] Kunstig intelligens skal hjælpe os med at analysere, forstå og træffe bedre beslutninger. Men teknologien kan ikke – og skal ikke – erstatte mennesker eller træffe beslutninger for os. Det skal fx fortsat være lægen, der stiller den endelige diagnose for patienterne.” (Finansministeriet & Erhvervsministeriet 2019).

Med denne vision skriver regeringen sig ind i den etiske og praktiske debat for, hvilken rolle Kunstig Intelligens skal spille i fremtiden, og hvad den enkelte kommune og forvaltning bør gøre for at komme de teknologiske udviklinger i møde. Denne intention var særligt tydelig i den nationale strategi. Et af de centrale initiativer er en igangsættelse af signaturprojekter inden for Kunstig Intelligens i den offentlige sektor med fokus på at erfaringsudvikle på sundhedsområdet, social- og beskæftigelsesområdet og den tværgående sagsbehandling. Disse signaturprojekter har årligt fra 2020-2023 fået afsat 60-67 mio. kr., som kommunerne kunne søge for at igangsætte disse projekter (Digitaliseringsstyrelsen, U.Å, Signaturprojekter). Kommunernes Landsorganisation skriver, at 140 AI-projekter er blevet sat i gang siden 2020, hvoraf 78 af projekterne er i drift i dag – det er dog kun nogle af disse projekter, der er signaturprojekter. Følgende er eksempler trukket fra KL’s landekort over AI-projekter:

1. "Værktøjet Fleetoptimizer skaber klimasmart flådestyring ved hjælp af kunstig intelligens, der anvender moderne allokeringsalgoritmer til at optimere flådesammensætning og tildelingen af køretøjer til ture. I drift og udviklet som et tværgående kommunalt projekt med Aarhus som projektejer."

2. "Prøvehandling af Hugin Findr: Triangering af borgere på ældreområdet med henblik på at identificere borgere der er i fare for at blive (gen-)indlagt. Al observationsdata gennemgås og behandles af en algoritme baseret på Machine Learning."

3. "Smartmail 2.0 repræsenterer næste generation i udviklingen af intelligente digitale post-håndterings-systemer, videreudviklet fra fundamentet lagt af Signaturprojekt Fordeling og journalisering af post version 1.0. Formålet med Smartmail 2.0 har været at forbedre og udvide værktøjet til sortering af digital post, med særlig fokus på integration af avanceret kunstig intelligens og maskinlæringsteknologier" (KL, U.Å).

Ovenstående eksempler fra AI-projektdatabasen illustrerer, hvordan algoritmer kan benyttes til at forbedre drift og ressourceallokering i det offentlige. Som tidligere nævnt, er listen af projekter lang og spænder fra alt fra forskellige versioner af GPT (Generative Pre-trained Transformers)-modeller, som er trænet til at besvare spørgsmål om alt fra turisme til rotteområdet i forskellige kommuner, til 3D-rendering og risikovurderingen af træerne i en given kommune. Det er altså næsten alle felter inden for det offentlige, hvor disse projekter bliver oprettet. Det kan dog også være vigtigt at sikre meningsfuldheden, og hvilken værdi der bliver skabt, for eksempel ved at Sønderborg Kommune får trænet sin egen KommuneGPT.

I november 2024 blev der offentliggjort en evaluering af signaturprojekterne fra Deloitte. Hovedfundene fra den evaluering var, at signaturprojekterne, der blev sat i gang, skabte dialog mellem myndigheder om AIs rolle i det offentlige, og at de projekter, der er blevet gennemført, har skabt økonomiske gevinster for kommunerne. De centrale udfordringer for projekterne var dog, at det er vigtigt at få sat en ordentlig usecase op for projekterne, og at sikre, at man har et tilstrækkeligt og korrekt datagrundlag til den givne AI-model. Slutteligt var en central udfordring, at de juridiske rammer ikke var på plads. Dette betød, at succesfulde projekter ikke havde lovhjemmel til at drifte ud over den prøveperiode, projektet kørte i (Deloitte, 2024). Det er dermed tydeligt, at der er vilje til at gøre brug af de teknologiske udviklinger, men at potentielle projekter har brug for de rigtige rammer.

Dette eksempel tydeliggør, at AI har gjort sit indtræden på scenen for offentlige myndigheder. Det kan være svært at følge med den massive udvikling, der sker inden for data science, når det kommer til udbredelsen af AI-baserede værktøjer, også med specifikt øje på at forbedre governance. Her ser man næsten ugentlig udvikling inden for algoritmer og værktøjer, som kan bruges alsidigt. Her spiller

Open Source en vigtig rolle i at udbrede algoritmer, da denne tilgang til projekter gør det muligt – for private såvel som offentlige aktører – at bruge dem som udgangspunkt til at udvikle nye værktøjer, der kan forbedre deres drift. Man kan dermed se, at Open Source kan hjælpe med at skubbe startlinjen for Closed Source-projekter. Det er en relativt ny udvikling inden for offentlig policyudvikling at benytte sig af data science og data scientists til at skabe værktøjer og opkvalificere beslutningsprocesser. Policyudvikling er i sagens natur et tværfagligt felt, og det kan derfor virke naturligt at inddrage andre felter i processen. Der er i dag flere anerkendte data science-værktøjer til rådighed for en datadreven offentlig administration. Her kan data science-værktøjer være med til at informere beslutningerne, der leder til præcisionspolicy (Chen et al, 2021). De mest kendte nærliggende metoder kan være at lave eksplorative analyser på datasæt for at kvalificere beslutninger inden for nye emner. Mulighederne for brugen af disse værktøjer kan spænde langt bredere. Man kan også bruge regressionsanalyser til at forudse fremtidige udgifter, eller man kan bruge Machine Learning-algoritmer med henblik på at finde sammenhænge mellem policyimplementeringer og deres konsekvenser (ibid.). Dette åbner op for spørgsmålet om, hvilke værktøjer der skal bruges til fremtidens beslutningsprocesser, og hvordan man i dag kan bruge værktøjer, der allerede er i brug i andre felter til at styrke de politiske processer i offentligt regi og dermed styrke demokratiet.

I et samfund, der bliver mere og mere individualiseret, bliver det sværere at drive et direkte og meningsfuldt demokrati. Det kan blandt andet ses på den faldende opbakning til politiske partier (Folketingets Oplysning, 2025) og en stigende frivillig interesse for enkelte mærkesager, eksempelvis klimademonstrationer eller studenteroprøret (VIVE, 2024). Med den manglende input direkte fra medborgere – ud over på særlige enkelte mærkesager – kan policyudviklere blive nødt til i større grad at læne sig op af tal og statistik for at udvikle meningsfuld og effektiv policy. Det offentlige rum er i samme ombæring i større grad blevet lagt om fra virkeligheden til det digitale, og her bliver sociale medier taget op som et rum til debat. For at sikre en kvalitet i policy feedback skal der derfor nye metoder til, der tager højde for, at det offentlige forum er blevet digitaliseret.

Dette studie har derfor til formål at undersøge, hvordan policyevalueringsprocessen kan styrkes gennem brug af Machine Learning-metoder. Der vil blive taget udgangspunkt i teorien om policy cyklussen, herunder policy feedback. Empirien består af Twitter-data (platformen er nu kendt som "X") fra årene 2019-2021 som et udsnit af offentlig debat på digitale platforme. Specifikt omfatter datasættet

tweets, der brugte hashtagget #dkpol mellem 2019 og 2021. Da datasættet dækker folketingsvalget i 2019 og Coronaepidemien, kommer de temaer naturligvis til at være fremtrædende. Dette er en god case, da Coronaepidemien var en tid med behov for hastig policyudvikling og -implementering. Her var offentligheden essentiel at have med på udviklingen for at sikre national sundhed. Der vil derfor særligt blive fokuseret på sentiment, som i dette speciales sammenhæng skal forstås som en scoring af ord, der bruges til at vurdere den emotionelle værdi af tekst. Der vil blive trænet et neuralt netværk til at teste brugbarheden af prædiktiv sentimentscoring. Datasættet vil blive delt op i månedlige tidsserier med henblik på at udvinde månedlige topics, som den offentlige debat omkredser, hvorefter sentiment til disse emner vil blive udvundet ved hjælp af sentimentscoring på tweets, der falder inden for disse topics. Der vil derefter blive trænet en Long Short-Term Memory-model (LSTM), en Convolutional Neural Network-model og en XGBoosting-model, som testes mod hinanden. For at teste performance på disse modeller vil der blive brugt præcision og loss som evalueringsmetrikker. Der vil i metodeafsnittet blive gået i dybden med, hvad der karakteriserer disse modeller. Der vil blive inddraget public governance-teorier for at kunne diskutere applicerbarheden af disse metoder i offentlige bureaukratier.

1.1 Problemformulering

På baggrund af ovenstående indledning vil dette speciale undersøge:

Hvordan kan neurale netværk anvendes til at forudsige offentlighedens holdning til policy på Twitter, med henblik på at understøtte feedback-elementet i policy cyklussen?

Her udvælges Neurale Netværk som værktøj, da dette er en kategori inden for maskinlæring, der ser stor udvikling, men også stor mulighed inden for det at fremskrive (forecasting) og klassificere. Twitter bliver i dette speciale brugt som udsnit af den offentlige holdning. Det bliver det ud fra perspektivet, som Twitters ejer italesætter i et tweet d. 22 marts 2022:

> "*Given that Twitter serves as the de facto public town square, failing to adhere to free speech principles fundamentally undermines democracy.*
>
> *What should be done?*" (Musk, 2022)

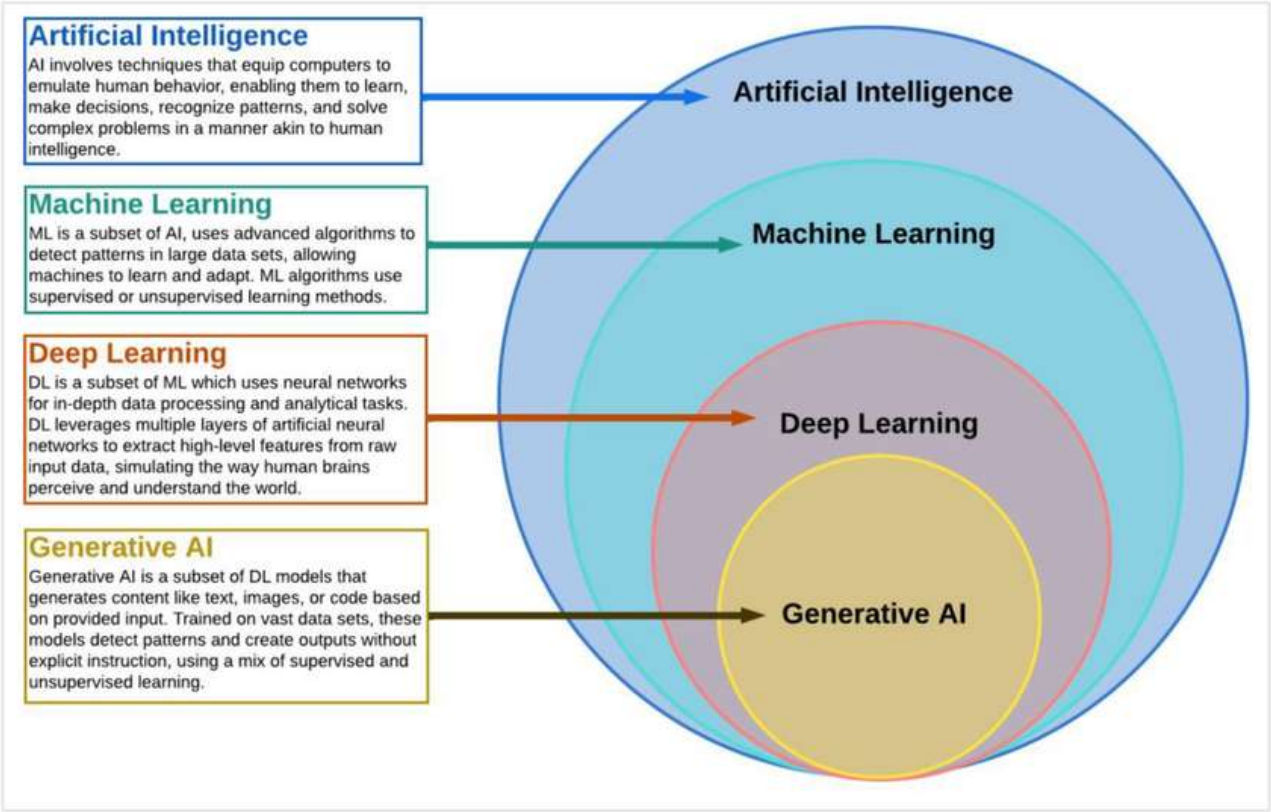
Dette speciale deler overbevisningen, at sociale medier, herunder Twitter, er blevet til en form for "public square", altså et led af det offentlige rum, og derfor kan data derfra bruges som et udsnit af

det offentliges holdning. Datasættet har også ud fra et effektivitetsperspektiv en styrke i, at man har mulighed for at benytte allerede tilgængelige data til at blive klogere på policy feedback i stedet for at indsamle data på mere.

1.2 Begrebsafklaring

AI, Machine Learning og data science er brede begreber, der ofte anses som værende "buzz words". Det kan afkonkretisere et emnefelt, hvis ikke man er præcis med definitionerne på disse ord. Det er derfor nærliggende for dette projekt at afgrænse og kvalificere brugen af en række ord, der kommer til at fremgå i projektet:

AI, Machine Learning og Deep Learning: Dette er tre brede begreber, der bruges omskifteligt, men i virkeligheden er det tre begreber, der inden for samme genstandsfelt er på hvert sit niveau. Dette kan forstås, som at AI er et paraplybegreb, der indfanger mange metoder, hvoraf Machine Learning er en underkategori, der beskriver et sæt af metoder inden for AI. Inden for Machine Learning er der en underkategori, der hedder Deep learning, som beskriver mange af de mest fremtrædende metoder, vi bruger i dag. Alle de Machine Learning-metoder, der vil blive brugt i dette projekt, falder inden for Deep Learning-kategorien. Nedenfor er en populær graf, der beskriver relationerne mellem disse begreber.



(Zhuhadar, Lily & Lytras, Miltiadis, 2023)

Big Data er et emne, der fylder meget, når man snakker om AI. Dette skyldes, at de store modeller trænes på gigantiske datasæt for at være mest muligt retvisende. Dette kommer dog på bekostning af, at 'Big Data' flyder ind i almensproget og kan blive en smule ukonkret. Derfor bruger dette projekt 'Big Data' som en beskrivelse af et datasæt, der er så stort, at det ikke kan åbnes på en normal computer. Det vil sige, at trods dette projekt arbejder med et stort datasæt, så vil det ikke klassificeres som Big Data, omend det er i projektets overbevisning, at metoderne er skalerbare til et større datasæt. Projektet bevæger sig dog i retningen af Big Data, når der bliver genereret embeddings senere i projektet, da disse gør dataen mere kompleks (Google Cloud, U.Å.).

1.3 State of the art

Forskning inden for AI og offentlig administration er et voksende felt. I takt med at Machine Learning og AI-løsninger bliver mere udbredte, så følger offentlige administrationer også med. Denne udvikling kan spores i regeringens førnævnte digitaliseringsstrategi og tilhørende projekter, som også har resulteret i øget fokus i denne udvikling. Selvom der er mange danske eksempler, er de præget af den

nationale digitaliseringsstrategi og dets projektudbydelse, samt hvilke metoder der kan tilkøbes af administrationer. Det kan derfor være gavnligt at se udenlands for at se på eksempler på projekter, der gavner uden for digitaliseringsstrategiens faste projekter. Her er den citerede kilde "Data Science for Public Policy" af C.J. Chen en god kilde til projekter, hvor man benytter sig af data science-metoder. Inden for forskning er feltet policyanalyse traditionsrigt, men anvendelse af disse metoder i praksis er en relativt ny tendens. Der kan derfor også være plads til udvikling, i forhold til hvordan disse metoder skal benyttes. Nedenfor er et par studier, som dette studie er inspireret af, og som benytter sig af data science til at blive klogere på policy feedback.

I studiet "*Can Artificial Intelligence Technology Help Achieving Good Governance: A Public Policy Evaluation Method Based on Artificial Neural Network"* (Xu et al., 2025) ser forskerne Zhinhan Xu, Zijun Liu og Hang Luo på, hvordan man kan bruge neurale netværk til at analysere social medie feedback på policy til at opnå bedre governance. Dette studie har været en stor inspiration for dette speciales forskningsdesign, da dette studie tester hvordan forskellige neurale netværk kan bruges til at sortere, og dermed vurdere "public sentiment". Public sentiment skal forstås som det offentliges holdning til politik fordelt på en positiv-negativ skala. Studiet tager udgangspunkt i populære datasæt inden for politisk analyse og udvikler et Convolutional Neural Network (CNN) model, og tester det mod andre populære modeller for at vurdere deres models brugbarhed. De finder, at deres Convolutional Neural Network model er 93.4% præcis, når man tester mod baseline modeller. Her bruger de også Text Graphs til at optimere deres models præcision (ibid.). Dette studie er inddraget ud fra dets design og metode. Der bliver fokuseret på de regionale governance niveauer, og hvordan de kan gøre brug af disse AI-metoder, men de tester også hvilke metoder, der kan fungere bedst til den rolle.

I studiet "*Machine Learning and Public Health Policy Evaluation: Research Dynamics and Prospects for Challenges"* undersøger forskergruppen Zhengyin Li, Hui Zhou, Zhen Xu og Qingyang Ma, hvordan sundhedspolicyevaluering kan bliver styrket og optimeret med AI- og Big Data-gennembruddet, der er sket de senere år. I artiklen undersøger de, hvordan – og til hvilken grad – Machine Learning-metoder kan bruges til at opnå disse mål. De analyserer brugbarheden og principperne bag traditionelle sundhedspolicyevalueringsmetoder som Difference in Difference, Regression Discontinuity Design og Synthetic Control og analyserer hvilke udfordringer, de får med udviklingen inden for AI og Big Data. Som led i deres analyse ser de på, hvordan Machine Learning kan bruges til at håndtere

store og ustrukturerede datasæt med henblik på at lindre begrænsningerne, der ses i de større modeller. De rejser også kritikpunkter inden for de specifikke udfordringer, som de nye metoder kan skabe – specifikt Black Box-modeller, der kan skabe udfordringer, når man skal forstå resultaterne af analyser, da det ikke er lige til at forstå beslutningerne, modellen har taget (Zhengyi et al., 2025). Der kan være datasikkerhedsudfordringer, og der er også udfordringer inden for bias i data, der kan styrke diskriminerende strukturer. Studiet foreslår en række strategier, der kan bruges til at udvikle Machine Learning-strategier for sundhedspolicyevaluering. De argumenterer for, at en data- og teoridrevet tilgang til modelfortolkning af resultater kan modarbejde Black Box-modeller. At arbejde med flere niveauer af data kan adressere data bias-udfordringer, og det er vigtigt at sætte tekniske, juridiske og sociale sikkerhedsmekanikker på plads for at sikre god og etisk brug af disse metoder. De mener, at praktisk integration af Machine Learning-metoder til sundhedspolicyevaluering kræver hybridmodeller, der trækker på både de tekniske styrker men også en teoretisk bundlinje, der kan skabe balance og forståelse for genstandsfeltet (ibid.).

Dette studie er inddraget, da det er et godt udsnit af den metodologiske debat, der er i gang inden for policyevaluering. Den fokuserer specifikt på det sundhedspolitiske, men kommer med argumenter og udfordringer, der er lette at drage paralleller til. Konkret kan dette bidrag gavne en debat om specialets resultater.

Disse to ovenstående studier er blevet valgt, da den ene bidrager metodisk og det andet bidrager teoretisk. Der er dog mange studier, som kommer med relevante bidrag til området, og som kunne være relevante at inddrage, men disse er blevet fravalgt til fordel for en mere uddybende beskrivelse af de ovenstående. Et eksempel på en fravalgt tekst er "*Sentiment Analysis Of Government Policy On Corona Case Using Naive Bayes Algorithm*" af Isnain, Marga og Alita (2021). Dette studie bruger en Naive Bayes-algoritme til at lave en sentimentanalyse på tweets om Corona-policy i Indonesien. Studiet opnår en præcisionsscore på 84%. Trods genstandsfeltet er sammenligneligt og den metodologiske diskussion er relevant, inddrages studiet ikke. Den beslutning er taget på baggrund af studiets opnåede præcisionscore, da Zhu et al. (2025) opnår en højere præcision med mere relevante metoder for dette speciale.

2.0 Teori

2.1 Public Governance-teorier: Fra New Public Management til New Public Governance og Digital Era Governance

Dette afsnit har til formål at belyse en af de centrale teorier for at forstå incitamentet for og debatten om organisering i den offentlige sektor i det 21. århundrede. Der er bred enighed om, at man er gået væk fra eller har udviklet sig fra New Public Management (NPM). Dette afsnit har til formål at afdække denne overgang og beskrive de nye teoretiske forståelser af udviklingen inden for governance (Torfing & Traintafillou, 2017, s. 25).

NPM tog i 1980'erne sit indtog i den offentlige sektor som et svar på en offentlig sektor, som på mange måder var udfordret (Torfing & Traintafillou, 2017, s. 19). NPM har spillet en central rolle i vestlige offentlige bureaukratier siden før årtusindskiftet. Det har været et governance-regime, som bredt kategoriseret har handlet om en opdeling af bureaukratier til mindre enheder, en individualisering af arbejdskraften og en konkurrenceliggørelse af arbejdsopgaver. Lederens rolle i NPM er særlig, da en central del af styringsparadigmet handler om lederens strategiske ledelse. Ledelsen bør understøttes af styrings- og ledelsesteknologier, man ser fra den private sektor. Der bør under NPM være et armslængdeprincip for ledelseslag til konkret sagsbehandling, og ledelsen bør understøtte et virksomhedssprog, hvor man i højere grad ser borgere som kunder, og fagprofessionelle som servicemedarbejdere (ibid:18-19).

NPM, som styringsparadigme for det offentlige, har mange komponenter, der opgør den brede definition. De behøver ikke nødvendigvis at optræde universelt i alle NPM-regimer, men er blot generelle regler, der sikrer en bredere definition. Eksempler på dets komponenter kan være kvasi-markedsliggørelse af offentlige ydelser (f.eks. private botilbud), Individualiseret Performance-fokus i afdelinger (KPI'er) og en opdeling af bredere bureaukratiske afdelinger til mindre enheder eller organisationer. En af grundtankerne bag NPM er en målorienteret virksomhedslogik med optimering og effektivisering i fokus (Dunleavy et al., 2006). NPM har ikke været uden sin kritik, og flere forskere skriver om, at offentlige administrationer i større grad bevæger sig væk fra NPM og mod nye styringsparadigmer, der bryder med NPM-logikkerne. Selvom NPM bredt er blevet erklæret stagneret eller død, så er der uenighed i litteraturen om hvilket regime, der i dag gør sig gældende, og hvor langt det er fra NPM

(Torfing & Traintafillou, 2017, s. 25). Dette er en kompleks debat, da offentlige administrationer er komplekse systemer, og da de ikke ledes ens, er det vigtigt at konkretisere sin forståelse af dem, da der med stor sandsynlighed kommer til at være forskel på offentlige administrationer på tværs af lande. Den måske mest fremtrædende teori inden for styringsparadigmer er New Public Governance (NPG). Dette styringsparadigme adskiller sig fra NPM på en række områder, herunder effektiviseringenslogikken, som var drivende for NPM, og har mere fokus på samskabelse på tværs af sektorer, og en ny organisationsstruktur, der fokuserer på netværker (ibid:25-26).

En anden fremtrædende post-NPM teori, der er særligt relevant for dette speciale, er Digital Era Governance (DEG). DEG er en teori, der bygger på IT-systemers voksende rolle i offentlige administrationer. Disse systemer understøttes af ændringer i administrationer med fokus på informationshåndtering. Denne øgede digitale arbejdsform kan bane vejen for yderligere effektivisering og nye organiseringsformer i det offentlige. Denne DEG-udvikling er ofte en beskrivelse af bevægelsen væk fra det analoge bureaukrati over til det digitale bureaukrati med mails, intranet og automatiseringer af arbejde (Dunleavy et al., 2006, s. 14-15). De tre primære temaer, som Dunleavey og Margetts (2006) argumenterer for kommer til at fylde i DEG, er:

1. Reintegration af noget af den opdeling, som NPM logikken drog med sig,
2. En agil tilgang til governance med en holistisk tilgang,
3. Digitaliseringsændringer mod en fuldkommen digital administration (ibid).

Disse 3 temaer opgør de primære karakteristika for governance-modellen, men forskerne bag denne model har opdateret, hvordan de mener, modellen har udviklet sig i takt med teknologien og brugen heraf har udviklet sig. De beskriver tre bølger inden for DEG: den første bølge som værende den, der først modarbejdede NPM's strukturer, den anden bølge som værende den, hvor teknologien og brugen af metoderne modnede, og til sidst en tredje bølge, som er en videre modning, men også en udvikling af potentialet af data science og AI i governance (Dunleavey & Margetts, 2023, s. 40). Hvilken governance teori, der gør sig gældende, er der stadig debat om i dag. De to modeller, der er blevet præsenteret ovenstående, er et eksempel på, hvordan teorierne også har hver deres fokus, og det kan være med til at bidrage til uafklarethed i debatten.

2.2 Policyforskning og Policy Cyklussen

Policyforskning er ofte tværvidenskabelig, da den policy, som analysen beskæftiger sig med, næsten altid berør et tværsnit af felter. Et smalt og konkret eksempel kan være implementeringen af inklusionsreformen, hvor pædagoger, lærere og psykologers faglighed spiller ind. Et mere bredt eksempel kan være den grønne trepart, der har fået oprettet sit eget ministerium for at sikre eksekveringen af reformen på tværs af flere danske sektorer.

Forskningen skelner altså mellem policyanalyse på ene side og implementering og evaluering på den anden side, hvoraf det første foregår præ-implementering og de andre post-implementering af et givent stykke policy (Albæk, 2009, s. 1035-1036). En af de mere fremtrædende teorier, som prøver at indfange hele policyprocessen, er policy cyklussen. Policy cyklussen beskriver policyprocessen som en 6-trinscyklus, der starter med et problem, som policyen skal adressere. Parsons beskriver det i sit værk "*Public Policy: An Introduction to the Theory and Practice of Policy Analysis*" (1995) som en distancering mellem et "issue" og et "problem", hvor et "issue" er en bredere problemstilling, som vi ofte alle kan blive enige om, er der. Et "problem" er derimod en fastsættelse af, hvad denne bredere problemstilling består af. Et eksempel på et issue kan være stigende temperaturer på globalt plan. Her kan problemet være CO^2-udledninger, hvorefter man kan foreslå et stykke policy for at gøre noget ved dette problem –for eksempel et loft på CO^2-udledninger. Det første trin i denne cyklus er en politisk fortolkning af et issue og dermed en samfundsmæssig anerkendelse af den problemstilling. Dernæst er der en problemdefinition, som skal forstås som, at man oversætter problemet til at være et policy-problem. Således går det fra en bredere samfundstendens til noget, der kan handles på – her er agendasetting og -framing relevante begreber, da det er her, man framer et problem og agendasetter et policyområde (Parsons, 1995, s. 87-89). Næste punkt i cyklussen er "identifying alternative responses or solutions", og derefter bliver disse muligheder vægtet og evalueret, og der bliver valgt en policy option ud af de muligheder, man har fundet frem til. Dette trin kan også anses som en form for legitimering af den udvalgte policy. Dernæst kommer implementeringen af den valgte policy og efter det en evaluering. Parsons lægger i denne teori vægt på, at alle disse punkter i policy cyklussen har valg i sig – hvilke alternativer man finder frem til, og hvilke metoder man bruger til at evaluere med, er vigtige spørgsmål og led i cyklussen (ibid:245).

2.3 Data science til offentlige administrationer

Som tidligere beskrevet er det en relativt ny tendens for offentlig policy-udvikling at benytte sig af data science og data scientists. Trods dette er der i dag mange kendte data science-værktøjer til rådighed for en datadreven offentlig administration. Her kan data science-værktøjer være med til at informere beslutningerne, der leder til præcisionspolicy (Chen et al., 2021). I en krisetid hvor udviklingen ikke altid er til at forudsige, er det en styrke at være omstillingsparat som administration, og det er vigtigt at kunne rette kurs ud fra feedback. Den klassiske udfordring kan være, at feedback-loopet for policy kan være langsom og tage flere år. Det er her, data science-metoder kan hjælpe beslutningstagere med at blive klogere på effekten af policy (ibid.).

Der er flere metoder, der kan være relevante for en administration at benytte sig af. I det følgende stykke kommer der en række eksempler:

- Man kan lave eksplorative analyser for at forberede ministersvar inden for nyere emner
- Man kan visualisere data med henblik på at belyse en beslutningsproces
- Man kan lave regressionsanalyser eller algoritmer med henblik på at fremskrive fremtidige omkostninger eller forbrug

- Man kan lave kausale Machine Learning-algoritmer med henblik på at finde sammenhænge mellem policyimplementeringer og deres konsekvenser.

Der er altså flere metoder i dag, der er relevante for offentlige administrationer, men forskningen inden for feltet kan i dag ofte arbejde i helikopterperspektiv uden at undersøge, hvordan moderne metoder kan benyttes i praksis.

2.4 Teoretisk sammenfatning

Dette afsnit har til formål at sammenfatte specialets teoriafsnit med henblik på at klargøre specialets metodeafsnit og analyse. Den akademiske debat om governance-teorier er relevant for dette speciale, da det lægger grunden for, hvordan nye metoder også bliver introduceret og arbejdet med. Hvor NPG fokuserer på netværksdriften af organisationen og samarbejdet på tværs af sektorer, fokuserer DEG mere på den instrumentale og metodiske udvikling inden for governance. De to teoretiske forståelser har forskellige fokus, men de overlapper også, da de begge er reaktioner på NPM. Det er for eksempel begge teorier, der argumenterer for, at der er sket en decentralisering i kølvandet på NPM. Teorierne er inddraget, da de er baggrunden for at forstå, hvordan administration udvikler sig, og de kommer derfor til at have en forklarende rolle i diskussionen. Policyevaluering og -cyklussen spiller en vigtig rolle i analysen, da det er teorien, som analysens fund kan bidrage ind i. Slutteligt bliver data science for offentlige administrationer brugt til at rammesætte de metodiske valg. Det er altså en teori, der skal bidrage til, hvordan studier som dette kan medvirke til policyevalueringer.

3.0 Metode

Det følgende afsnit har til formål at gå i dybden med de designovervejelser, metoder, styrker og begrænsninger, der ligger til grund for besvarelsen af problemformuleringen.

3.1 Data og afgrænsning

Datagrundlaget for dette speciale er Twitter-data trukket ved hjælp af Twitters API i årene 2019-2021. Dataene er afgrænset til tweets, der benytter sig af det populære #dkpol, som er et hashtag, der bliver brugt til at skabe et forum for debat af dansk politik. Der er dog vigtige overvejelser ved brugen af denne data, som er følgende:

1) Dataene er udelukkende trukket fra Twitter, som er en af de mindre danske social medieplatforme. Det er en platform, hvor debatten også i stor grad er præget af, at deltagerne selv er en aktiv del af politik (Thielke, 2021, KL).

2) Dette resulterer i en form for cirkulær feedback – hvis man som beslutningstager selv bidrager en overvægt til datasættet, så bliver feedbacken også farvet af det input, og man skal derfor også have det for øje, når man laver analyser.

3) Det er et datasæt, som ikke kan opdateres på Twitter uden en investering i Twitters nye API-omkostninger.

4) Der er også et problem med bots i dataene. Hvis målsætningen er at være så tæt på at få borgernes egen evaluering ud af denne analyse som muligt, er det vigtigt at frasortere bots, der automatisk kan forstørre holdninger. Dette er en udfordring, som flere sociale medier selv arbejder med ved hjælp af deres bot moderation. Med henblik på at holde fokus på metodeudnyttelse vil dette speciale ikke gå ind i bot detection og moderation i datasæt.

Der er ligeledes nogle etiske overvejelser at have in mente, hvilket vil blive uddybet i et senere afsnit.

Datasættet er som sagt trukket fra #dkpol i årene 2019-2021. Denne periode er, som tidligere nævnt, særligt præget af Coronaepidemien, som var en tid, hvor bølgerne gik højt i den offentlige debat,

men det var også en tid med ekstreme politiske udviklinger. Dette kan ses i ændringer i tilslutninger; for eksempel blev der argumenteret for, at man i flere lande oplevede en "rally around the flag effect" (TV 2, 2020). Der var også store ændringer i Corona-restriktioner, og man lærte at bruge nye mobilapplikationer til at tracke sin bevægelse med henblik på at opspore smitte, og der blev implementeret en regulering af størrelsen på forsamlinger. Der var altså mildest talt meget, som danskerne – og resten af verden – skulle tage stilling til. Derfor er dette datasæt også et udsnit i en politisk debat, hvor vi ved, at der har været meget diskussion.

3.1.1 Etiske overvejelser

For at sikre en etisk behandling af offentlighedens data, er det vigtigt at være tydelig omkring de overvejelser, der går ind i at designe et studie, der behandler så offentlige data. Der kan være uklare grænser når dataene er offentligt tilgængelige, men brugerne med stor sandsynlighed ikke har gjort sig overvejelser om, at det, de skriver, kan inddrages i studier som disse. Det ville være enormt omfattende og en urealistisk målsætning at indhente samtykke fra samtlige brugere, der optræder i dette studie. Dette vurderes dog ikke at være nødvendigt i dette tilfælde, da der, som tidligere nævnt, er tale om brugere, som alle skriver offentligt om et emne. Der er dog truffet en beslutning om at anonymisere brugerne i datasættet ved ikke at inkludere deres brugernavne. Derfor er der heller ikke blevet inddraget kontostatistikker, som ellers kunne være relevante for at undersøge tendenserne i, hvem der fylder mest på platformen. Specialet har valgt stadig at forholde sig til de konkrete, men forarbejdede tweets. Dette kan i sig selv have sine udfordringer, da tweetene hverken står uredigeret, men flere er heller ikke uigenkendelige til fra det originale tweet og der er derfor risiko for, at man kan søge sig frem til de originale tweets, hvis de stadig er at finde på Twitter og dermed belyse en borger, der ikke har samtykket. Yderligere så kan clustering metoderne på sådanne datasæt skabe uetiske situationer, hvis analytiker ikke finder måder at anonymisere borgere på. Et eksempel herpå er at man kan ville kunne oprette cluster alle brugerne der udtaler sig kritisk om politikere. Det er med disse overvejelser specialets design er opstillet.

3.2 Forskningsdesign: Tidsserieanalyse

Dette speciales design er inspireret af flere studier, der passer i samme felt, herunder Xhu et al. (2025), der tester Machine Learning-modeller på populære datasæt inden for politisk analyse. De laver et tværsnitsdesign på de enkelte datasæt og modeller for at kunne teste, hvor præcise deres modeller er til at vurdere sentiment. Sentiment scoring skal, som tidligere nævnt, forstås som en vurdering af indholdet af tekst. Denne vurdering kan designes på mange måder, men for simplicitetens skyld bruger dette speciale en skala, der hedder positiv = 1, negativ = -1, neutral = 0. Datasættene, som bruges i studierne henvist til tidligere, er blandt andre IMDB-anmeldelser og BBC-artikler – altså store datasæt, som rækker fra alt fra politik til film og medier. I dette projekt bruges data fra den danske politiske diskurs på Twitter, og det kan derfor fungere som en illustration af den politiske holdning og en test på Machine Learning-metoderne i en politisk kontekst. Projektet benytter sig desuden af forskningsdesignet Time Series Cross-Sectional til at afdække både det temporale, men også en dybere indholdsside af datasættet. Dette har til formål at gøre det muligt at teste forskellige Deep Learning-modeller på deres evne til at forudsige politiske emners udvikling. En vigtig overvejelser der er blevet gjort er de compitutionelle ressourcer der er til rådighed for projektet, der vil derfor bliver brugt de første 4 måneder af hele datasættet for at sikre at der er ressourcer nok til at koden der ligger til grund for specialet effektivt.

3.3 Machine Learning-modeller

De to typer Neurale Netværk, der vil blive benyttet i dette speciale, er Transformer og Convolutional Neural Network. En transformer-model er en model, der benytter sig af opmærksomhedsmekanismer i dets arkitektur til at udnytte kontekst for præcist at forecaste (Gerrón, 2023, s. 609-611), hvorimod Convolutional Neurale Netværk baserer sig på convolution, herunder convolutional layers, til at give kontekst og struktur på data (ibid:479). Det er vigtigt at nævne, at pre-træning spiller en rolle, når det kommer til transformer-modellerne; det betyder, at de er trænet på data forinden for at finde de vægte, der bliver brugt til at analysere data.

I første trin af analysen gøres der brug af to transformer-modeller udviklet ud fra BERT-arkitekturen til at klassificere datasættet, således at det kan bruges til mere meningsfuld analyse. BERT (Bidirectional Encoder Representations from Transformers) er en Deep Learning-sprogmodel, der ved

at bruge kontekst ud fra tidligere tekster i datasættene og efter en given tekst kan danne en dyb forståelse for indholdet. Dette er en proces, der sker i alle lag af det Neurale Netværk og derfor har kontekst med i hele processen.

De konkrete Deep Learning-modeller der bliver brugt til den prædiktive del af analysen, hedder henholdsvis LSTM, TextCNN og XGBoost. Disse tre modeller er bygget på forskellige tilgange til Deep Learning, og de har hver deres styrker. Long-Short Term-modeller er en af de mere anvendte Machine Learning-modeller, som fungerer ved at gøre brug af "forget gates". Det betyder, at når man bruger sekventielt data, så træner modellen både med de tidligere måneder i hukommelsen, men har også vægte, der ikke har tidligere data med (Gerrón, 2023, s. 568-69). TextCNN er, som tidligere nævnt, et Convolutional Neuralt Netværk, der bruger flere lag af convolution til at forstå sekventielt data. CNN bruges typisk i billedbehandling, men kan også bruges til Natural Language Processing (NLP), som dette speciale beskæftiger sig med (ibid:479). Slutteligt er der XGBoost (Extreme Gradient Boosting), som adskiller sig fra de andre modeller ved, at det ikke er et Neuralt Netværk, men i stedet gør brug af Gradient-Boosted Decision Trees, som er en algoritme, der minder om random forest. Måden, metoden fungerer på, er, at den bygger iterativt parallelle decision trees, hvor den gradient booster i den forstand, at den tager gennemsnittet af de træer, den bygger, og arbejder sig mod færre fejl (NVIDIA, U.Å., XGBoosting).

3.4 BERTopic

Topic-modelling bliver i dette projekt gjort med BERTopic. Dette er en topic modelling-løsning, der gør brug af BERT-arkitekturen. En af dens største styrker er dens modularitet. BERTopic har 6 trin:

1) Embeddings – her bliver input-dokumenter omdannet til numeriske repræsentationer ud fra transformer-modeller, der hedder sentence transformers. Standardmodellen er "all-MiniLM-L6-v2". Dette projekt bruger dog "Maltehb/danish-bert-botxo", da den er trænet specifikt til dansk. Den er trænet på dansk Wikipedia, så der er heller ikke risiko for, at præ-træningsdataene overlapper med dette speciales trænings-, test- og valideringsdata (huggingface.co)

2) Næste skridt er dimensionsreduktion. Dette er vigtigt, da embeddings-trinnet skaber store matricer, der gør det svært at skabe clusters grundet "the curse of dimensionality". Udgangspunktet for BERTopic er Umap, men der er flere forskellige dimensioneringsværktøjer, der virker med BERTopic. Dette speciale benytter Umap, hvor man kan skrue på fire metrikker for at ændre, hvordan dimensioneringen fungerer.

3) Det næste trin er den reelle clustering. Her er ligesom de andre trin i BERTopic mulighed for at vælge forskellige metoder. Det er i dette trin, hvor de dimensionaliserede embeddings bliver brugt til at skabe clusterne, der skal fungere som specialets topics. HDBSCAN er standardalgoritmen for BERTopic, men der kan være mange gode usecases for de andre algoritmer. Eksempelvis kan man bruge k-Means, som også er en effektiv metode, der er mere stringent i forhold til hvordan den skaber sine clusters, hvor man på mere semi-superviseret vis kan bestemme mængden af clusters. Derudover så bliver alle topics sorteret i clusters til forskel for den mere flydende HDBscan, der bliver brugt i denne analyse (Grootendorst, U.Å.). Grunden til, at dette speciale holder sig til HDBscan, er, at det anvendte datasæt er så bredt, at der uden tvivl kommer til at være tekster, der falder uden for de forskellige emner, og der kommer til at være outliers. Det giver HDBscan plads til.

4) Vektorisering er det næste trin for BERTopic og skaber topic-repræsentationen. På dette trin kan man også bruge mange forskellige vektor-modeller. Her er der ikke en standardløsning - dette speciale gør brug af CountVectorizer, da det er en ligetil tilgang, som har alle de parametre, der er behov for, for at finetune topic-repræsentationen. Det er her stopwords kan bruges. Stopwords - eller stopord - er en liste af ord, som ikke giver værdi i repræsentationen, og man kan derfor på baggrund af dette lave en liste af ord eller tegn, der gør, at de ikke holder vægt i analysen. Grunden til, at de ikke fjernes tidligere – ligesom der bliver gjort med links og andre Twitter-specifikke ting – er, som tidligere nævnt, at BERT-modeller fungerer ud fra kontekst, så ved at fjerne dem helt vil man forværre topic clusteringen. Stopordslisten, der er blevet brugt, er dels brugt ud fra en liste fra Github (Torp, 2020), dels med feedback fra Claude AI og dels ud fra en iterativ proces, hvor ord der fylder meget i topic repræsentationen, men ikke skaber værdi i forbindelse med forståelsen af topics bliver skrevet på listen.

5) Næste trin i BERTopic-modellen er C-TF-IDF. Dette er en ny version af den klassiske TF-IDF-ligning, hvor C repræsenterer clustrene, som bliver omdannet til samlede dokumenter, hvor de mest hyppigste ord for hvert cluster bliver udvundet. Det er disse ord, der repræsenterer hvert topic. Der er også her mulighed for parameter tuning, men specialet bruger her standardvægtene, der ligger i C-TF-IDF-ligningen.

6) Slutteligt er der mulighed for ekstra fine tuning af topics ved hjælp af ekstra modeller, Large Language Models (LLM) eller generativ AI. Der er mange måder at gøre dette på, og man behøver ikke at vælge en enkelt. Man kan for eksempel bruge en ny repræsentationsmodel til at forbedre topic-repræsentationen og så bagefter gøre brug af LLM eller generativ AI til at videre arbejde topics ud fra prompter (Grootendorst U.Å.). Dette trin gør dette speciale ikke brug af, da det, der er centralt for den videre analyse, er, at tweets bliver sorteret i topics, og ikke så meget hvor god beskrivelsen af disse topics er.

3.5 Kvalitetskriterier

For at sikre specialets kvalitet og gennemsigtighed vil der i det kommende stykke blive diskuteret kvalitetskriterier.

I forbindelse med specialets validitet kan man argumentere for, at validiteten både internt og eksternt er stærk ud fra, at det, som specialet sætter sig for at undersøge, er, hvordan Machine Learning-metoder kan bruges til at undersøge Twitters politiske diskurs, og #dkpol er et fremtrædende forum for politisk diskussion på Twitter. Dette kan også bakkes op af mængden af tweets, som der er at finde i datasættet. Der er 25 måneders tweets på dette forum, som svarer til ca. 650.000 tweets. Hertil er en andel af det tal retweets, bot-opslag og en del opslag, der bliver sorteret ud, når vi fjerner stopord og derefter tomme tweets. Der er altså en chance for, at vi ikke måler korrekt på alt i datasættet, og der er også anden politisk diskussion på andre hashtags som for eksempel #dkgreen. Grunden til, at dette projekt dog stadig udelukkende bruger #dkpol-datasættet, er, at det er et rendyrket debatudsnit, som rammer flere forskellige emner, og det giver mulighed for at have et bredere sigte. Administrationer kan overveje, at hvis deres sigte for eksempel kun dækker det grønne område eller det sociale område, så kan der findes og benyttes sociale medie-sider, der specifikt omhandler deres genstandsfelt med henblik på at få så rammende et datasæt på den virkelighed, man prøver at analysere.

En generel udfordring med mange af disse præ-trænede Machine Learning-modeller er den manglende gennemsigtighed på træningsdata. Dette kan være en udfordring for validiteten, både fra et gennemsigtighedsperspektiv, men også ud fra, at den data, man træner på, potentielt kan være med til at træne disse præ-trænede modeller. Man kan derfor i værste tilfælde lande i en "garbage in garbage out"-situation, som betyder, at hvis man leverer et dårligt datagrundlag, så kan man også forvente en dårligt rammende analyse. Denne overvejelse er dog taget med i modeludvalget, og der er taget hensyn til denne udfordring i udvalget af Machine Learning-modeller og forarbejdelsen af data ved at undersøge datagrundlaget for modellerne i brug. Det er desværre ikke alle, der er gennemsigtige med dette, men på trods af denne udfordring, så er specialet stadig let at replikere, da både data og koden, der ligger til grund for analysen, er offentlig tilgængelig.

3.6 Analysestrategi

Første delanalyse har til formål at illustrere, hvordan moderne Deep Learning-metoder kan benyttes til at belyse policyfeedback. Transformers-modeller kommer til at blive brugt til sentimentscoring og topic modelling med henblik på at kvantificere det offentliges udviklende holdning.

Først indsamles data og opdeles i tidsserier, og derefter bliver der forarbejdet data for at sikre kvalitet i de kommende processer. Forarbejdelsen består af fjernelse af Twitter-specifikt indhold såsom RT (en indikator for at dette var et genopslag/retweet), hashtags og stopord (ofte brugte ord i dansk sprogbrug som beskrevet i et tidligere afsnit), der ville kunne mudre en topic clustering. Derefter er der foretaget en sentimentanalyse på de enkelte tweets og embeddings (den sproglige placering), som bliver genbrugt, således at sentimentanalysen og topic-modelleringen bliver udført på de samme vilkår. Topics bliver genereret ud fra den måned, de er i, for at udvikle denne førnævnte tidsserie.

En relevant overvejelse har været, hvorvidt det har givet mening at køre topic-modellering på hele datasættet eller at opdele det på månedlig basis. Hvis man kører det på tværs af hele datasættet, får man et billede på de generelle tendenser i ens datasæt, hvorimod det at køre en topic modellering for hver enkelte måned giver et bedre billede på udviklingen på tværs, og hvad der bliver diskuteret og hvornår i bedre detaljer. Det er også her, hvor det at bruge embeddings på tværs er en styrkelse af analysen, da det beholder teksternes placering, og man kan derfor spore topics over tid. Da analysen bruger et Time Series Cross Sectional Design, er det nærliggende at benytte sig af månedlige

tilgang, som beskrevet her. En anden ting, som er vigtig at notere, er, at tweetene bliver tokeniseret – det vil sige, at de enkelte tekster opdeles til tokens – da dette er en del af topic-modelleringsprocessen.

Efter disse indledende trin er der blevet trænet de tre tidligere nævnte modeller med det formål at teste, hvilke af de tre modeller der bedst kan prædiktere ud fra en feature engineering, der baserer sig på topics på en månedlig basis. Det, at sentiment, tekster og embeddings er med, gør, at man kan samle et sekventielt design, der passer til specialets forskningsdesign. Først køres en LSTM-model, derefter en TextCNN og til sidst en XGboost. Slutteligt vil metrikkerne blive sammenlignet på tværs af modellerne.

3.7 Bekendtgørelse om generativ AI

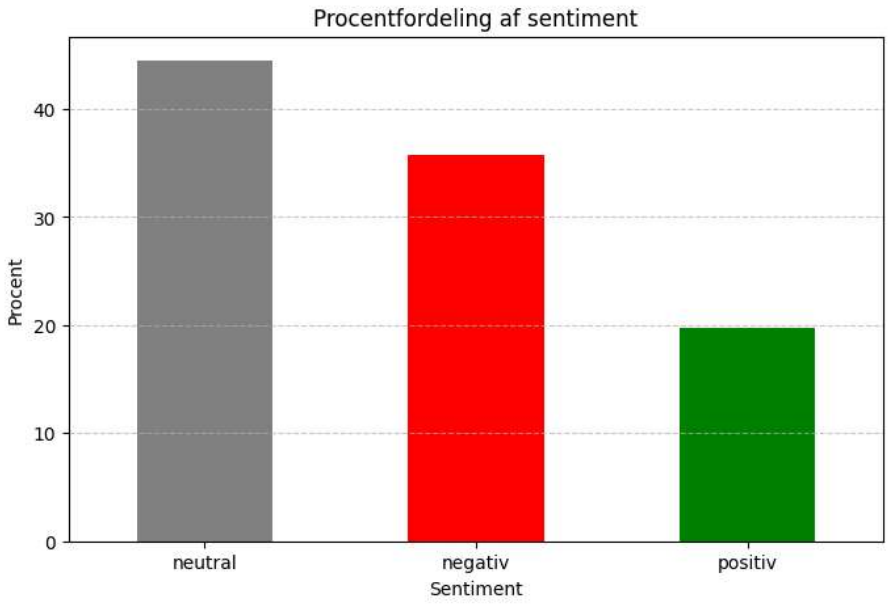
Generativ AI er i dette speciale blevet brugt i forbindelse med debugging og udvikling af kode, der ligger til grund for grafikker og regex statements. Det er også blevet benyttet til generel brainstorming og udvikling af stopords liste.

4.0 Analyse

4.1 Første delanalyse: Eksplorativ og klassificerende analyse

Første del af analysen vil belyse baggrunden for de kommende Machine Learning-modellers virke ved at fokusere på funktionaliteten af de to transformer-modellers sentiment og topic-analyse. Analysen vil benytte den historiske Twitter-data til at illustrere, hvordan administrationer kan bearbejde data med henblik på at forbedre forståelsen for udvikling i holdning, og hvad der fylder i debatten.

Indledende bliver dataene forarbejdet for at sikre, at transformer-modellerne kan bearbejde data ordentligt. Dette inkluderer at fjerne links og Twitter-specifikke specialtegn og håndtag/brugernavne. Med henblik på at forbedre de kommende analyser bliver der også fjernet populære danske ord fra datasættet ved hjælp af en stopordsliste. Det inkluderer eksempelvis "for", "egen", "din" og "og". Dette gør, at tekster, der ellers ville ligge tæt sammen på grund af hyppigt brugte ord i topic-modellingen, nu ligger det rigtige sted på baggrund af substansord. Det samme princip gælder for sentimentanalysen, så brugen af denne stopordsliste gør, at tekster ikke bliver trukket ned af for mange neutrale ord, og dermed sorterer tweets, der ellers havde haft positivt eller negativt sentiment. Det er derfor vigtigt at sikre, at sentimentanalysen er retvisende ved at se på procentdelen af tweets, der bliver sorteret som neutrale, da modellen sorterer tekster som neutrale, når den ikke er sikker. Det er derfor vigtigt at sikre, at tweetene rent faktisk bliver sorteret. Fordelingen af tweets kan ses nedenfor og falder således, at der er en overvægt af neutrale tweets. Det er dog forventeligt, da preprocessing kommer til at gøre nogle tweets utydelige ved at fjerne tweets, hvor specialtegn, links og så videre fylder størstedelen af indholdet. Det giver dog også god mening, fordi en god del af tweetene ikke kommer til at have noget sentiment. Det kan tilskrives retweeting bots og deling af artikler, hvoraf der ikke er noget emotionelt indhold i selve teksten. Derefter ser vi en mindre overvægt af negativitet - dette kan måske virke som en bekræftelse af en intuitiv forståelse af, at den politiske debat på internettet ofte er drevet af negativt sentiment. Vi ser dog heller ikke en voldsom lav andel af positivitet, og det kan måske forklares med, at sociale medier – her særligt #dkpol – kan være en platform eller et sted, hvor policy tales op. Vi ser dog ikke en ligelig fordeling mellem positiv og negativ, og den overvejende vægt mod det negative kan give et billede af, at størstedelen af opslag med tydelige politiske holdninger er negative i denne analyse.



Topic-modelling er det andet led af den eksplorative analyse. Her er der flere metrikker, der kan sikre kvaliteten af analysen. Den første og mest intuitive, man kan se på, er outlier-procenten, det vil sige tweets, der ikke passer ind i den clustering, som er sat op.

Da denne analyse er nødsaget til kun at benytte fire måneders data for at mindske de computationelle krav, er der en risiko for, at disse måneder fungerer særligt godt på den udvalgte BERTopic med disse parametre. Parametrene der bliver brugt i HDBscan-clusteringen og dermed kan justeres er den minimum topic-størrelse, hvordan afstand måles og hvilken metode clusterne bliver samlet på.

Der ses på tværs af de fire måneder 44.808 outliers (bilag 1). Dette er en relativt høj procent, da der på tværs af alle måneder i alt er 140.000 tweets. Det vil sige, at det er under en tredjedel af tweets, der ikke bliver sorteret. Det er dog vigtigt at have for øje, at en begrænsning, der er blevet sat, er at topics skal indeholde mindst 50 tekster for ikke at blive sorteret fra. Det kan være en begrænsning, da mindre og lokale emner ikke nødvendigvis indeholder så store mængder af enkelte tweets om et givent emne på en måned.

Et andet element der er relevant at have for øje, er, hvor stor variation der er mellem de topics, som er at finde baseret på månederne. Et eksempel herpå er, at den første måned i datasættet, som er 2019-04 (april 2019), har 36 topics. De ord, som er blevet brugt til at repræsentere dette topic, er "`0_danmark_paludan_df_politikere`". Dette tolkes som, at dette topic har indfanget noget nationalistisk politik, hvor DF og Rasmus Paludan optræder som gennemgående tendenser. For at sikre kvaliteten og forbedre forståelsen for, hvordan topics hænger sammen bliver der også udvalgt et eksempel ud fra teksterne i clusteringen. Her er uddraget for topic 0 i topic listen 2019-04: "*Rasmus Paludans Stram Kurs får stor opmærksomhed…*". Vi ser altså her, hvordan topics kan bruges til at forstå det generelle billede af en debat, og hvordan tekster udvælges inden for cluster til at repræsenterer det. Det gør det muligt at læse de konkrete tekster for kvalitativt at danne en bedre forståelse for emnerne.

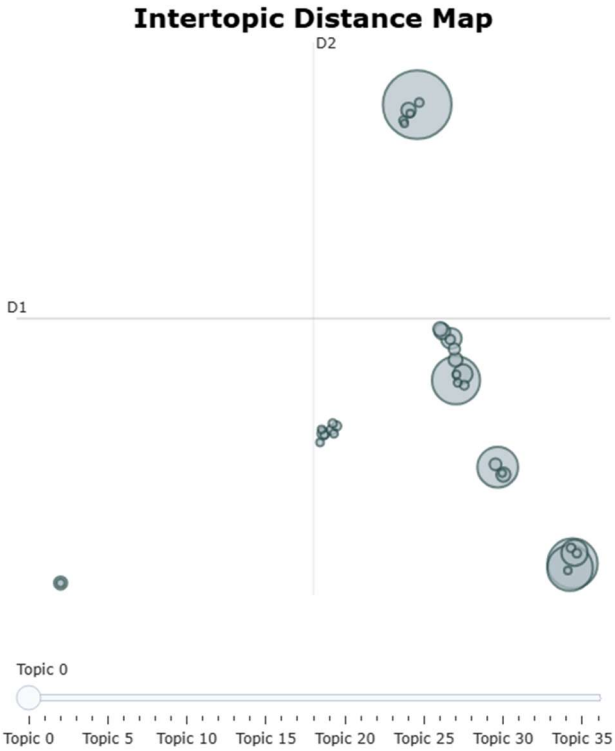
Der er dog også en stor variation i mængden af topics, der er på tværs af måneder, da der også er stor variation i, hvor mange tweets der er på tværs af månederne. Når man for eksempel sammenligner de 35 topics fra april med mængden af topics fra maj måned, er der en stigning på 85 nye topics. Denne variation kan tilskrives flere variable. Man kunne forestille sig en intensivering i politisk aktivitet på sociale medier generelt op til folketingsvalget i juni 2019, og der er også en stigning i topics frem mod valget.

En anden tendens, der er at finde i datasættet, er, at politikerne fylder en del ved navn. Dette kan både være en styrke og en svaghed, alt efter hvad målet er med ens analyse. For eksempel kan partier spore offentlighedens holdning på sociale mediers holdningen til sine kandidater ud fra disse clusterings. Et eksempel herpå er topic 34 i "`34_saxohvalp_gal_alex_vanopslagh`", hvor teksten er "*alex vanopslagh er et af en gal saxohvalp…*". Hvis man dog ikke er særligt interresseret i politikernes rolle og har et fokus på den konkrete politik, kan dette være en forstyrrende faktor. En måde at undgå dette er ved at lave em politiker stopords liste.

Eksemplet ovenfor viser tydeligt, at clusteret handler specifikt om Alex Vanopslagh. Når man ser på sentimentfordeling på dette topic, ser man 100% negativitet. Det er måske intuitivt forstået, at når man ser på ordene "saxohvalp" og "gal". Dette kan dog være et eksempel på, at nogle af clustrene potentielt kun indfanger de negative eller positive dele af et emne. Dette er ikke nødvendigvis hensigtsmæssigt og derfor noget, som kan arbejdes med, når man sammensætter sine parametre til fremtidige analyser. Der er dog også andre ting, der kan bruges til at evaluere, hvor præcis en given

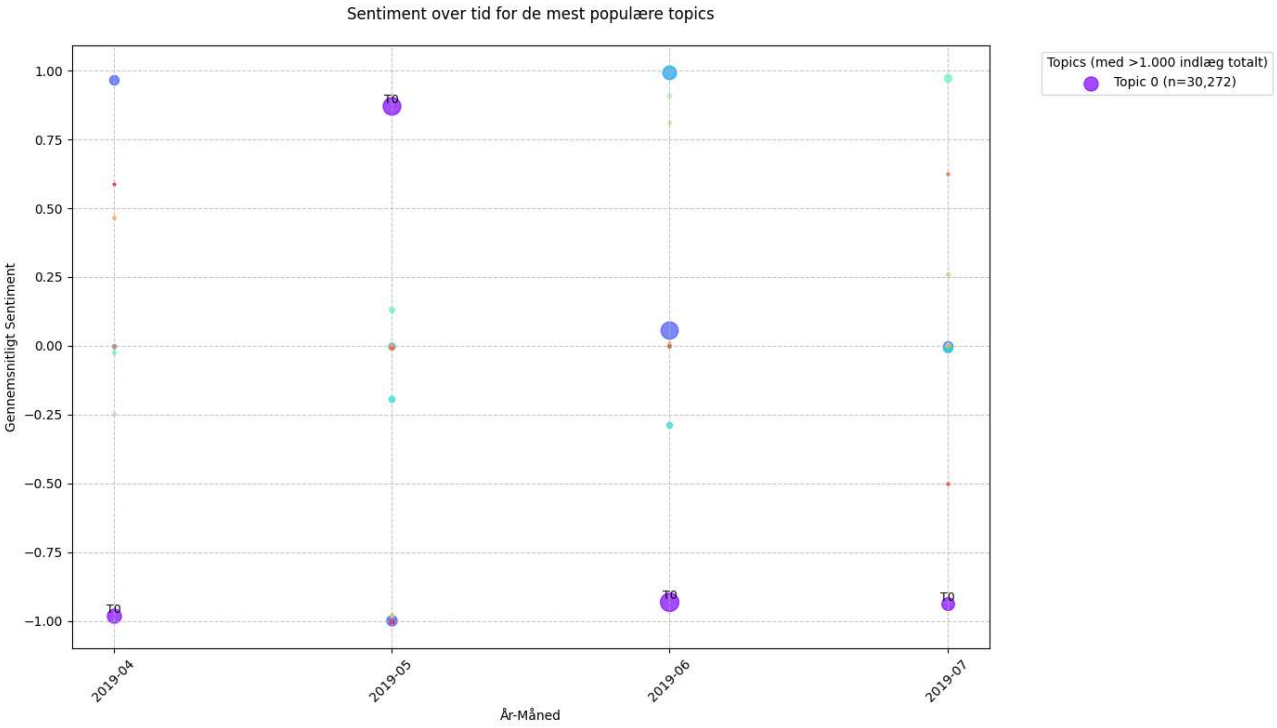
topic clustering er. Fordelingen af topics er vigtig, da det muliggør, at man kan se hvor store clusterne er, i forhold til hvor langt hvert emne ligger fra hinanden og dermed kan man sikre at de ikke overlapper, men vi kan også bruge Umap til at evaluere modellen. Modellen køres, som tidligere nævnt, ud fra batches baseret på den måned, tweetet er blevet postet i. Dette skaber derfor et behov for at sikre, at disse topics stadig er at spore på tværs af månederne for at sikre en kontinuitet og dermed projektets time-series.

ved at se på, hvordan topic-clusteringen ligger i en 2D-illustration for juli 2019, ses der fire primære clusters, hvoraf der er flere mindre clusters, som ligger omkring, og enkelte clusters der adskiller sig helt. Det er vigtigt at sige, at denne type illustration har sine begrænsninger, da 2D-repræsentationen ikke kan illustrere de mange dimensioner, der ligger i embeddings. Det, illustrationen kan give os, er et billede af overlap mellem clusterings - hvis der er store topics, hvoraf der ligger mindre topics indeni, så er der risiko for, at clustering ikke er optimalt afgrænset. I sådan en situation kan der være behov for yderligere parameter tuning.



Man kan med fordel bruge disse tre primære variable, Sentiment, Topics og måned, til at lave en graf, der viser de største clusters, og hvor de ligger i sentimentscoring for på den måde at vise den historiske udvikling, hvor de største emner i politik bliver vurderet på sentiment. Nedenfor vises en graf, der beskriver denne udvikling, hvor 1 repræsenterer 100% positivitet om et emne, og -1 repræsenterer 100% negativitet. Her ses det også, at der er flere store emner i juni måned, som også er der, hvor der skete valghandling i Danmark. Det er iøjnefaldende, at de største emner fordeler sig i ekstremerne på grafen, når der tidligere i analysen er blevet vist fordelingen af sentiment på tværs af hele datasættet. Dette kan potentielt forklares ved, at emner, der sorteres neutralt i sentiment, kan være så neutrale, at de bliver sværere at binde sammen med andre tekster semantisk, og derfor bliver de enten placeret i mindre clusters eller helt sorteret fra som outliers. Trods dette viser grafen også en masse mindre emner,

der fordeler sig bredere på skalaen af sentiment. Dette kan tolkes, som om at der stadig er en divers af de mindre emner på tværs af sentiment spektret.



4.1.2 Delkonklusion for første delanalyse

Ovenstående analyse beskriver processen, der rengør data til sentiment samt BERTopic-analyse. Der blev fundet en overvægt af negativt sentiment sammenlignet med det positive sentiment. Det blev derudover fundet, at topics, der blev genereret, havde en mindre andel outliers, men at de clusters, der blev dannet, lå tæt på hinanden. Det åbner dermed op for yderligere parameter-tuning i andre analyser. Slutteligt blev der sammenlignet udviklingen af topics på tværs af måneder, hvor det blev fundet, at de største emner fra hvert måned fordelte sig i ekstremerne, altså enten helt neutralt, positivt eller negativt. Der blev også fundet en tilvækst i tweets tæt på valget i 2019.

4.2 Anden delanalyse: Prædiktiv klassificering af sentiment

Den anden del af analysen har til formål at benytte de tre Deep Learning-modeller – LSTM, TextCNN og XGBoost – for at teste applicerbarheden i forbindelse med policyevaluering samt give et indblik i, hvordan policy kan komme til at udvikle sig i de kommende måneder.

Indledende bliver dataene bearbejdet til at kunne køre effektivt i disse tre modeller; det inkluderer sortering i kronologisk orden, adskillelse i test og træningsdata, at sikre at alt dataen er i det korrekte format og udvælgelse af features. Features er de input, som modellen skal benytte sig af til at træne sig selv på. Her bliver de primære variable valgt igen: sentiment på både topic og enkelt tweet-niveau bliver brugt, samt embeddings for tweets og topics. Disse er blevet udvalgt ud fra en forståelse af, at placeringen og dens sentimentværdi er et godt grundlag for at fremskrive fremtidige placeringer af topics og dokumenter, og dermed hvordan diskussion og sentiment udvikler sig.

Der vil som udgangspunkt blive gået i dybden med LSTM-modellen, og derefter sammenlignes med metrikker på tværs af modellerne. Som man kan se, kører LSTM-modellen 20 epoker som er udtrykket for den iterative træningsproces. Det er iøjnefaldende, at der er en meget lav loss rate på tværs af epoker. Dette betyder, at den ikke mister meget i den iterative proces, og det, at den falder løbende, viser også, at modellen lærer. Man kan se, at modellen er ekstraordinært præcis med en weighted average precision på 99%. Dette er så høj en præcision, at der må være en form for overfitting eller et alt for lille test datasæt. Det er muligt at se mængden af testdata, og hvordan det fordeler sig, ved at se på Support-sektionen i nedenstående tabel. Her ses, at over halvdelen af dataene er blevet sorteret i neutral sentiment, og at de samlede testdata er lige under 22.000. Hvis man ser på størrelsen af træningsdatasættet, så ligger det omkring de 110.00 linjer, og det passer dermed inden for den klassiske træningsopdeling, der ligger mellem 70-80% træning og 20-30% testdata.

```
--- Training LSTM Model ---

Using device: cpu

Epoch [5/20], Loss: 0.0178

Epoch [10/20], Loss: 0.0108

Epoch [15/20], Loss: 0.0076

Epoch [20/20], Loss: 0.0060


LSTM Results:

              precision    recall  f1-score   support


     negativ       0.99      0.98      0.98      6078
```

neutral	0.99	0.99	0.99	12702
positiv	0.98	0.97	0.98	3105
accuracy			0.99	21885
macro avg	0.99	0.98	0.98	21885
weighted avg	0.99	0.99	0.99	21885

Accuracy: 0.9864

Den første model der bliver kørt er LSTM modellen. Dens evaluerings metrikker bliver vist i ovensteående tabel. Den næste model er er XGBoost-modellen, og den opnår en ekstraordinær høj præcision på tværs af alle tre kategorier: 96.9-99.2%. De to andre modeller har ligeledes en høj præcision. Modellerne er på tværs af alle sentimenter meget præcise. Dette viser, at de features, vi har udvalgt, fungerer godt med modellerne. Det er vigtigt at have for øje, at det – på baggrund af det lidt smalle tidsserie, der er blevet nedsat grundet computerkraftbegrænsninger – derfor ikke er muligt at sætte et mere retvisende billede op af politiske svingninger. Dette har gjort, at dette datasæt, der er blevet sat op til den prædiktive analyse, kun har træningsdata fra dagene op til valget og testdata på dagene efter, og ud fra dette design kan det være oplagt at se en høj præcision. Dette er gavnligt af flere årsager. Man kan ud fra kortere politiske vinduer spore og effektivt danne et billede af den politiske udvikling på en kortere basis; dette fungerer i situationer så som når partier har brug for at danne et billede af, hvordan valget udvikler sig inden en valghandling, således at det er muligt at se, hvordan sentimenten udvikler sig. I nedenstående tabel er det tydeligt, at de forskellige modeller har forskellige styrker inden for sentiment, trods at de alle er stærkest inden for den neutrale klasse, hvilket giver mening, da denne klasse også er den med suverænt mest træningsdata tilknyttet. Det er dog stadig en tendens mod høj præcission på tværs af modellerne. Generelt så klarer TextCNN sig værre på tværs af modellerne og kommer helt ned på 78% effektivitet, hvilket isoleret set er relativt højt, men med plads til forbedring.

Accuracy by Sentiment Class:

Negativ class:

```
XGBoost:  0.9722

LSTM:     0.9780

TextCNN:  0.8583

Neutral class:

XGBoost:  0.9925

LSTM:     0.9935

TextCNN:  0.9587


Positiv class:

XGBoost:  0.9691

LSTM:     0.9739

TextCNN:  0.7833
```

Der er dog generelt en lavere tendens ved de positive tekster, som også er dem, der er i mindretal. Der vil derfor kunne spekuleres i, at et større datasæt med flere udsving vil resultere i en generelt lavere præcisionsrate, omend der er chance for, at positivmetrikken ville kunne forbedres relativt med mere data. Disse tal viser dog et stort potentiale inden for muligheden for at klassificere data, og derefter forudsige data. Med begrænsningerne i det udvalgte mindre datasæt giver det dog ikke god mening i at igangsætte en regressionsanalyse ud fra tre tidsserier. Ud fra den anvendte data har LSTM-modellen den bedste præcision ud af de tre modeller. Der er flere årsager hertil, der også kan være modelspecifikke, såsom udformningen af features. Her kunne man også have inddraget tekst-kolonnen for at give modellerne konkrete tekster at vægte. Dette kunne potentielt øge forvirringen i præcisionen, men ville kunne åbne op for flere muligheder – for eksempel kunne man gøre en træ-net model generativ, således man som beslutningstager kan se fiktive, men statistisk relevante feed-back-tekster inden for de forskellige sentimenter. Dette kan være en værdifuld måde at udvinde ar-gumenterne, der ligger i de forskellige politiske overbevisninger inden for et givent emne.

4.2.1 Delkonklusion for anden delanalyse

I ovenstående analyse er der blevet trænet tre modeller til at klassificere data inden for de tre sentimentkategorier ud fra embedding-placeringer. Der blev fundet en meget høj præcision på tværs af alle modellernes prediction, men LSTM-modellen var den, der havde den højeste præcision af alle modellerne med 98.64% på tværs af de tre sentimentklassifikationer.

4.3 Tredje delanalyse: Metoderne i praksis

Dette afsnit har til formål at binde de tidligere analyseafsnit tættere sammen med det teoretiske grundlag for analysen for bedre at besvare specialets problemformulering.

Som beskrevet i afsnittet om policycyklus, så anser Parson (1995) hvert et trin i cyklussen som værende et sted, hvor der tages beslutninger og dermed evalueres. Man kan med disse metoder præsenteret dermed holde en løbende evalueringsproces med direkte input fra vælgerne. Et eksempel herpå i issue-definition og agenda-setting-stadiet af cyklussen er, at man kan benytte topic modelling, som brugt i starten af analysen, til at blive klogere på problemer i samfundet, som ligger latent og bliver diskuteret, men endnu ikke er opfanget til et policy issue. Man kan altså med denne tilgang skabe muligheder for at opdage diskurser, der er under udvikling og følge dem tidligere end før.

Særligt sentimentanalysen er også vigtig til problemdefinition, da dette er et værktøj, der kan bruges til at danne en forståelse for sentiment i et bredt datasæt af tekster. Det behøver ikke nødvendigvis at være positiv-negativ-skalaen, det kan også være en klassificering i følelser. BERT-modellen "NikolajMunch/danish-emotion-classification" er en model, der kan fordele tekster i seks forskellige følelser; afsky, frygt, glæde, tristhed, overraskelse og vrede. Man kan med disse metoder udvinde, hvordan borgerne, der deltager i debatten, forholder sig til det givne emne på flere forskellige dimensioner.

Når det kommer til identificering af alternativer, kan man også gøre brug af disse metoder. Ved brug af topic modelling, så var en af de udfordringer, der blev identificeret i den tidligere analyse, at der var nogle topics, som handlede om det samme, men bare havde et andet sentiment og blev sorteret i et separat cluster på baggrund heraf. Dette er dog brugbart, når man gerne vil blive klogere på undergrupper i en bredere diskurs. Her kan man med udgangspunkt i et større emne og en Inter-

Topic-visualisering vise, hvilke clusters der ligger op ad hinanden. Dette kan bruges som en måde at finde alternativer, der bliver diskuteret i den offentlige debat.

I implementeringsfasen såvel som evalueringsfasen er disse metoder også tydeligt brugbare. Her kan der med fokus på specifikke emner måles løbende holdninger til det emne og udvikle forecasts til, hvordan potentielle udsving i holdninger kommer til at se ud. Dette er relevant som et ekstra ben, som de løbende implementeringsevalueringer kan stå på. Det har dog også sin særlige plads ved evalueringsfasen, hvor det er muligt kvantitativt at analysere store mængder feedback og omdanne det til konkrete pointer, som beslutningstagere kan tage stilling til.

Dette projekt er designet med det formål at illustrere, hvordan nyere Machine Learning-metoder kan benyttes i offentlige administrationer til at forbedre governance. Dataen til at undersøge disse metoder har været den generelle debat på Twitter. Der er dog intet, der ligger til hinder for en afgrænsning af projektet til kun at fokusere på klima eller arbejdsmarkedspolitik. En konkret og simpel måde at implementere et fokus på arbejdsmarkedspolitik i denne er at ændre minimum clusters. Ligesom der er implementeret en stopordsliste, så kan en data scientist implementere startlister. Så hvis ikke der bliver nævnt "jobcentre", "arbejdsløshed", "dagpengesats" eller "aktiveringskrav", og så videre så skal data i datasættet sorteres fra eller vægtes mindre. På den måde kan man tilpasse sit data til det genstandsfelt, man beskæftiger sig med. Dette gør sig selvfølgelig gældende ud fra de data, man har tilgængelige, men når man benytter så bredt et datasæt som dette, er det en mulighed.

4.3.1 Delkonklusion for tredje delanalyse

Denne delanalyse har fokuseret på, hvordan de foregående analyser har passet ind i både policy cyklussen og kunne bidrage til Parsons løbende evaluerings- og beslutningsproces. Slutteligt blev der taget stilling til, hvordan disse tilgange konkret ville kunne blive brugt afgrænset til en administration, som kun beskæftiger sig med et afgrænset politisk arbejdsområde.

5.0 Diskussion

5.1 Tekniske begrænsninger

En central begrænsning for dette speciale har været de computationelle kræfter, der har været til rådighed for at bearbejde store datasæt og køre effektive Machine Learning-algoritmer. Måden, specialet har håndteret denne begrænsning på, har været at mindske datasættets størrelse og mængden af features på modellerne. Dette har ændret fokus til en mere metode- og usecase-definerende analyse, hvor fokus har været at analysere, hvordan metoderne kan bruges i praksis frem for at levere en analyse, der beskriver udviklingen i holdning de 25 måneder, som hele datasættet dækker over. Det har betydet, at et led i analysen er blevet droppet, hvor topics blev sporet på tværs af de måneder ved hjælp af deres embeddings til fordel for en mere deskriptiv tilgang. En anden spændende mulighed, der kunne forelægge med flere tekniske ressourcer til rådighed, kunne være er at udvide datasættene. Lige nu består datasættene af de tweets, der eksplicit gør brug af #dkpol, men ikke diskursen og diskussionen der opstår i kommentarsporene til disse tweets diskussionen der foregår i tweets der ikke nævner #dkpol eksplicit. Dette ville man kunne opnå med et bredere datasæt og flere ressourcer til at behandle dataen. Disse begrænsninger er ikke entydigt negative, da den begrænsede mængde ressourcer, også skabte et behov for prioritering og streamlining af analysen, så målet med analysen ikke blev undergravet af de ekstra features, man har kunnet have med og de ekstra aspekter, man har kunnet undersøge.

5.2 Konklusion

Specialet har taget udgangspunkt i debatten om digitaliseringen og AI i det offentlige. Der har været igangsat mange projekter, men der har været flere udfordringer med implementering. Dette speciale har taget udgangspunkt i denne problemstilling ved at beskæftige sig med, hvordan nye Machine Learning-metoder kan benyttes til at informere policyevaluering. Denne tilgang har haft det formål at undersøge hvordan man kan udnytte de nye metoder til værdiskabende analyser for offentlige forvaltninger. Derfor har dette speciale arbejdet med følgende problemformulering.

Hvordan kan neurale netværk anvendes til at forudsige offentlighedens holdning på Twitter til policy, med henblik på at understøtte feedback-elementet i policy cyklussen?

Til denne besvarelse er der blevet gjort brug af prævalente governance teorier til at danne baggrunden for analysen. Policyevaluering og herunder policy cyklussen, samt data science for public administration, bliver brugt som målestok for hvordan, de konkrete metoder kan indgå. Specialet har gjort brug af transformer modeller inden for sentimentanalyse og topic modelling med henblik på at klassificere og clustre politiske emner til yderligere analyse. I denne del af analysen blev der fundet, at de største emner fra hver måned fordelte sig i ekstremerne, altså enten helt neutralt, positivt eller negativt. Der blev også fundet en tilvækst i tweets tæt på valget i 2019. I forhold til clusteringen, så var der kun en mindre andel af teksterne, som blev frasorteret som outliers. Der blev lavet både en visualisering af overlap af clusters, samt en visualisering af hvordan de mest populære topics fordeler sig. Der blev også fundet en tendens til flere tweets i opløbet til folketingsvalget 2019, hvilket indikerer, at datasættet er relevant for aktuelle politiske udviklinger. Anden del af analysen fokuserede på Machine Learning-modellerne LSTM, XGBoost og TextCNN, hvor de blev trænet til at klassificere tweets ud fra features: Embeddings og sentiment, både på topic og på det enkelte tweet. Her opnåede modellerne på tværs meget høje præcisionstal. På tværs af de tre sentimenter havde modellerne sværest ved at klassificere positive topics og lettest ved at klassificere neutrale. Den mest succesfulde model var LSTM, som er særligt stærk til sekventielt data. Den sidste delanalyse undersøgte, hvordan metoderne kunne forenes i policyprocessen, og hvordan man ved at specificere emnefeltet for sine analyser kan skabe værdifulde værktøjer, selv for administrationer med et betydeligt smallere fokus end de tidligere analyser.

5.3 Bidrag

Dette speciale har bidraget til den offentlige adaption af ny-udviklede metoder indenfor Machine Learning, ved at tage udgangspunkt i organisering af offentlig administration og policyevaluering som eksempler på kerneanalytiske områder, der kan styrkes ved brugen af disse metoder. Ved at tage udgangspunkt i disse tilgange har specialet belyst, hvordan man kan gøre brug af disse metoder til at lave komplekse analyser og monitoreringsværktøjer til løbende evaluering af policy. Specialet tager udgangspunkt i prominente modeller, som har fyldt meget i data science-udviklingen, hvilket har været relevant at undersøge, eftersom dette stadig er et nyt område, som først nu begynder at blive implementeret i politologisk forskning.

5.4 Videre undersøgelser

Dette speciale viser, at der er potentiale for videre studie i fagspecifikke emner for at udvikle frameworks og værktøjer, der kan løfte administrationers evalueringsprocesser. Der er dog også plads og mulighed for undersøgelser, der har for øje, hvordan administrationer skal implementere data science baserede værktøjer i deres specifikke organisation. Specialet har desuden givet et generelt billede af hvordan Machine Learning-metoder kan bruges i politologiske undersøgelser, med henblik på at løfte niveauet i kvantitativ politologisk forskning.

Litteraturliste

Albæk, E. (2009). Policy Analyse. I L. B. Kaspersen, & J. Loftager (red.), *Klassisk og moderne politisk teori*. Hans Reitzels Forlag.

Chen, Jeffrey C., et al. *Data Science for Public Policy*. 1st Edition 2021, Springer International Publishing AG, 2021, https://doi.org/10.1007/978-3-030-71352-2.

Deloitte, 2024, "EVALUERING AF AI-SIGNATURPROJEKTER"

Dunleavy, P, et al. "New Public Management Is Dead—Long Live Digital-Era Governance." *Journal of Public Administration Research and Theory*, vol. 16, no. 3, 2006, pp. 467–94, https://doi.org/10.1093/jop-art/mui057.

Dunleavy, P., & Margetts, H. (2023). Data science, artificial intelligence and the third wave of digital era governance. *Public Policy and Administration*, *40*(2), 185-214. https://doi.org/10.1177/09520767231198737 (Original work published 2025)

Espersen H.H., Fridberg, T & Bastholm S, 2024 "hovedfundende: Frivillighedsundersøgelsen 2024 – En repræsentativ befolkningsundersøgelse af udviklingen i danskernes frivillige arbejde", Vive

Finansministeriet & Erhvervsministeriet, "National strategi for kunstig intelligens", 2019
Digitaliseringsstyrelsen, UÅ, Signaturprojekter; https://digst.dk/kunstig-intelligens/signaturprojekter/

Folketingets Oplysning, 2025, https://www.ft.dk/da/ofte-stillede-spoergsmaal/parti_hvor-mange-medlemmer-har-de-politiske-partier

Géron, A. (2022). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow : concepts, tools, and techniques to build intelligent systems* (3. edition.). O'Reilly.

Google Cloud. (U.Å.). *What is big data?* Google. https://cloud.google.com/learn/what-is-big-data

Grootendorst, M. (n.d.). Clustering in BERTopic with HDBSCAN. Maarten Grootendorst. [https://maartengr.github.io/BERTopic/getting_started/clustering/clustering.html]

Isnain, Marga og Alita (2021), Sentiment Analysis Of Government Policy On Corona Case Using Naive Bayes Algorithm"

Katrine Emme Thielke, 2021, "Overblik over sociale medier", KL

Kommunernes Landsforening (KL), U.Å., "Kommunernes AI-landkort", www.kl.dk

Li, Z., Zhou, H., Xu, Z., & Ma, Q. (2025). Machine learning and public health policy evaluation: research dynamics and prospects for challenges. _Frontiers in Public Health_, _13_. https://www.frontiersin.org/journals/public-health/articles/10.3389/fpubh.2025.1502599/full

Maltehb/danish-bert-botxo · Hugging Face, https://huggingface.co/Maltehb/danish-bert-botxo

Musk, E. [@elonmusk]. (2022, marts 26). [Tweet]. X, https://x.com/elonmusk/status/1507777261654605828

NVIDIA, U.Å., XGBoosting

Parsons, "Public policy: an introduction to the theory and practice of policy analysis" (1995)
J. C. Chen et al., 2021, Data Science for Public Policy, Springer Series in the Data Sciences, https://doi.org/10.1007/978-3-030-71352-2_1

Torfing, J., & Triantafillou, P. (2017). *New public governance på dansk.* (1. udgave. 1. oplag.). Akademisk Forlag.

Torp, B. (2020, March 6). Dansk stopords Liste / Danish stopwords. Github. Retrieved January 11, 2022, from (https://gist.github.com/berteltorp/0cf8a0c7afea7f25ed754f24cfc2467b)

TV 2. (2020, 29. marts). *Krisen har fået danskerne til at samle sig bag statsministeren – men herfra bliver det sværere*.

Zhuhadar, Lily & Lytras, Miltiadis. (2023). The Application of AutoML Techniques in Diabetes Diagnosis: Current Approaches, Performance, and Future Directions. Sustainability. 15. 13484. 10.3390/su151813484.

Xu, Z., Liu, Z., & Luo, H. (2025). Can Artificial Intelligence Technology Help Achieving Good Governance: A Public Policy Evaluation Method Based on Artificial Neural Network. *SAGE Open*, *15*(1). https://doi.org/10.1177/21582440251317833 (Original work published 2025)

Koden der er udviklet i forbindelse med specialet kan tilgås på https://github.com/munin-hbl/Speciale/.