

NPF i media

En semantisk analys



Lisa Pålsson

EC Utbildning

Projekt i Data Science

2025-10

Abstract

This project analyzes Swedish news articles related to neuropsychiatric disorders (NPF) using topic modeling and sentiment analysis. A dataset of articles was collected and processed to identify dominant themes and emotional tone. The BERTopic model revealed recurring topics such as education, healthcare, and social support. Sentiment analysis showed a predominantly neutral tone, with occasional shifts toward negative sentiment in discussions about policy and accessibility. The results highlight how NPF is framed in media and suggest areas for further investigation.

Innehållsförteckning

1	Inledning.....	1
1.1	Syfte och frågeställning.....	1
2	Teori.....	2
2.1	NPF som begrepp och i media	2
2.2	Semantisk analys och textanalys.....	2
3	Metod	3
3.1	Datainsamling	3
3.2	Datagranskning	3
3.3	Textbearbetning.....	4
3.4	Temaanalys med BERTopic	4
3.5	Sentimentanalys.....	4
3.6	Export och sammanställning.....	4
3.7	Agilt arbetssätt	5
3.8	AI i processen	5
4	Resultat och Diskussion	6
5	Slutsatser	8
6	Självutvärdering.....	9
	Appendix A	10
	Källförteckning.....	12

1 Inledning

Neuropsykiatriska funktionsnedsättningar (NPF) som ADHD, autism och Tourettes syndrom har under de senaste decennierna fått ökad uppmärksamhet i både forskning, samhällsdebatt och media. Mediernas sätt att rapportera om NPF påverkar inte bara allmänhetens förståelse utan kan också bidra till att forma attityder, policybeslut och tillgång till stödinsatser. Samtidigt är det svårt att få en samlad bild av hur NPF faktiskt framställs i nyhetsrapporteringen. Vilka teman lyfts fram, vilket tonläge används och vilka aktörer får komma till tals?

I takt med att AI-baserade textanalysverktyg utvecklas öppnas nya möjligheter att analysera stora mängder text på ett systematiskt sätt. Genom att kombinera topic modeling, sentimentanalys och ordfrekvensanalys kan man få en mer nyanserad bild av hur NPF representeras i media.

Denna rapport utgår från ett datadrivet angreppssätt för att undersöka hur NPF framställs i svensk nyhetsrapportering. Med hjälp av semantisk analys och maskininlärningstekniker har ett litet urval av artiklar samlats in, bearbetats och analyserats.

1.1 Syfte och frågeställning

Syftet med denna rapport är att undersöka hur neuropsykiatriska funktionsnedsättningar framställs i svensk media med fokus på teman, tonläge och språkliga mönster. För att uppfylla syftet kommer följande frågeställningar att besvaras:

1. Vilka återkommande teman kan identifieras i nyhetsrapporteringen om NPF?
2. Vilket tonläge används i artiklarna, och hur varierar det mellan olika teman?

2 Teori

2.1 NPF som begrepp och i media

Neuropsykiatriska funktionsnedsättningar (NPF) är ett samlingsnamn för flera olika diagnoser som i grunden handlar om att hjärnan arbetar och fungerar annorlunda.

De vanligaste diagnoserna är ADHD (och ADD) och Autism men även Touretts syndrom, dyslexi, dyskalkyli, språkstörning och OCD ingår. Det uppskattas att cirka 10 % av alla barn har någon form av NPF. De olika NPF-diagnoserna innebär olika svårigheter men kognitiva svårigheter som koncentrationssvårigheter, impulskontroll, minne, planering och sociala svårigheter är vanligt förekommande (Hjärnfonden).

I denna studie fokuserar analysen främst på ADHD och autism då dessa diagnoser förekommer mest frekvent i det insamlade nyhetsmaterialet.

Hur NPF beskrivs i media påverkar både allmänhetens uppfattning och den samhälleliga debatten om stöd och anpassningar.

2.2 Semantisk analys och textanalys

Textanalys handlar om att undersöka innehållet i texter för att hitta mönster, teman eller känslor. Det kan man göra manuellt men i detta projekt har automatiska metoder använts för att analysera ett större antal nyhetsartiklar. Med hjälp av datorbaserade verktyg kan man bearbeta stora textmängder och få fram statistik och strukturer som annars är svåra att upptäcka.

Semantisk analys fokuserar på betydelsen i texten, alltså vad orden betyder i sitt sammanhang. Det handlar om att förstå hur olika begrepp hänger ihop och hur texter kan grupperas utifrån innehåll snarare än enskilda ord.

Ett exempel på semantisk analys är temamodellering, där en algoritm identifierar ämnen som återkommer i textmaterialet. I detta projekt har BERTopic använts för att hitta teman i nyhetsrapporteringen om NPF. BERTopic är en metod som kombinerar moderna språkmodeller med statistiska mått för att gruppera texter efter innehåll och lyfta fram viktiga nyckelord i varje tema. Metoden gör det möjligt att analysera stora textmängder och få fram tydliga ämneskluster. För mer information om hur BERTopic fungerar hänvisas till den officiella dokumentationen (BERTopic, 2024).

Sentimentanalys är en annan typ av textanalys som fokuserar på tonläget i texten, det vill säga om innehållet är positivt, negativt eller neutralt. I detta projekt har metoden använts för att undersöka hur olika teman kring NPF framställs i media.

3 Metod

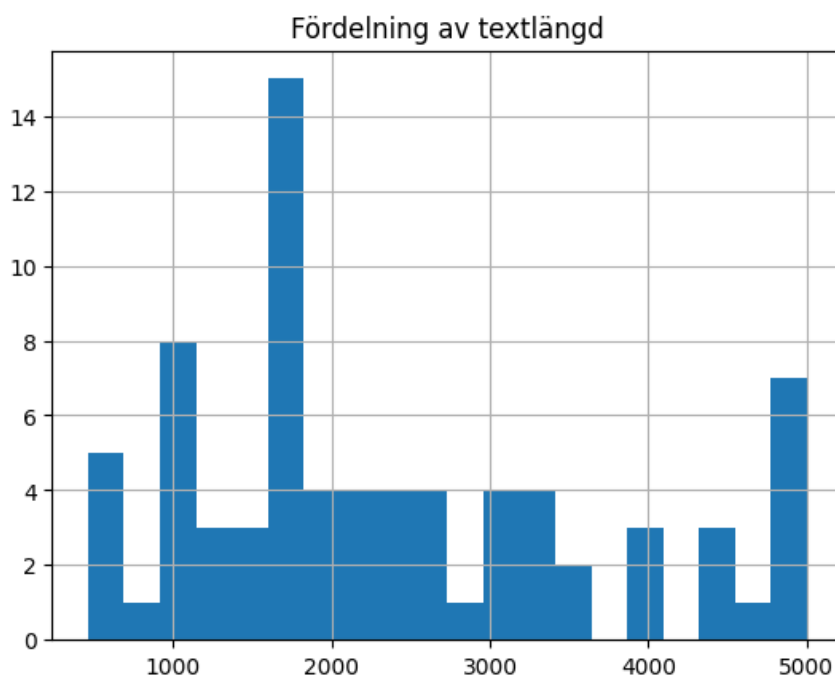
3.1 Datainsamling

Underlaget består av 76 nyhetsartiklar om neuropsykiatriska funktionsnedsättningar (NPF), insamlade via web scraping från svenska nyhetssajter. Inledningsvis gjordes flera försök att använda ett skript för att automatiskt hitta relevanta artiklar men detta fungerade inte. Istället valdes en manuell metod där URL:er samlades i en CSV-fil och därefter hämtades med hjälp av ett Python-script som extraherade rubrik och innehåll från respektive sida. Resultatet sparades i en strukturerad CSV-fil med kolumner för URL, titel, innehåll och status.

Datainsamlingen genomfördes manuellt genom Google-sökningar på begreppet "NPF" direkt på några av Sveriges största nyhetssajter såsom SVT, DN och Expressen. Relevanta artiklar valdes ut och URL:erna samlades i en Excel-fil, som sedan användes som grund för web scraping. Detta tillvägagångssätt gav bättre kontroll över urvalet och säkerställde att materialet var aktuellt och relevant.

3.2 Datagranskning

Efter insamlingen genomfördes en kvalitetskontroll av materialet. Samtliga artiklar innehöll både titel och brödtext och den genomsnittliga textlängden uppgick till cirka 2 400 tecken. En histogramvisualisering (Figur 1) av textlängderna visade en varierad fördelning vilket tyder på att materialet är tillräckligt varierat för att möjliggöra semantisk analys. Det fanns även en variation i textlängden vilket indikerar att det är en blandning mellan både korta nyhetsnotiser och längre reportage.



Figur 1. Fördelning av textlängder

3.3 Textbearbetning

Texten förbehandlades genom att fylla i saknade värden, rensa bort vanliga svenska stopwords och extrahera relevanta ord. Detta låg till grund för senare ordfrekvensanalys och visualiseringar.

Ordet *nyheter* förekommer ofta i materialet, vilket är naturligt eftersom källorna är nyhetsartiklar. I denna analys har jag valt att behålla det, även om det i vissa fall kan betraktas som ett domänspecifikt stoppord. Det påverkar inte helhetsbilden men kan filtreras bort i framtida iterationer.

Ordmolnet ger en visuell överblick över de mest frekventa orden i artiklarna. Det fungerar som ett pedagogiskt komplement till den kvantitativa analysen och hjälper till att identifiera dominerande ord samt eventuella brusord som kan behöva filtreras bort.

3.4 Temaanalys med BERTopic

För att hitta återkommande teman i artiklarna användes BERTopic, en avancerad språkmodell som grupperar texter med liknande innehåll. Modellen anpassades för svenska texter och skapade automatiskt teman baserat på artiklarnas innehåll. Varje tema beskrivs med nyckelord som sammanfattar vad artiklarna i gruppen handlar om.

Modellen skapade också en visualisering som gjorde det lättare att se hur temana fördelade sig. Genom temaanalysen gick det att identifiera vilka ämnen som dominerar i rapporteringen om NPF.

3.5 Sentimentanalys

För att bedöma tonläget i varje artikel användes en enkel regelbaserad metod för sentimentanalys. Texten jämfördes mot fördefinierade listor med positiva och negativa ord och artiklarna klassificerades som positiva, negativa eller neutrala beroende på ordens förekomst. Ordlistan togs fram med hjälp av AI-stöd och bygger på allmänt vedertagna känslord samt begrepp som är särskilt relevanta i NPF-kontexten såsom *stöd*, *diagnos* och *utanförskap*.

För att fördjupa analysen något kombinerades temamodellen med sentimentklassificeringen vilket gjorde det möjligt att jämföra tonläge mellan olika teman. Denna kombination gav en tydligare bild av vilka ämnen som hade en positiv respektive negativ rapportering och bidrog till en mer nyanserad tolkning av mediebildens kring NPF.

3.6 Export och sammanställning

Samtliga resultat exporterades till CSV- och bildfiler. En sammanfattning genererades såsom antal artiklar, genomsnittlig textlängd, antal identifierade teman och sentimentfördelning.

3.7 Agilt arbetssätt

Projektet genomfördes som ett individuellt arbete, vilket innebar att vissa delar av det agila arbetssättet (som samarbete och gemensam planering) inte tillämpades i sin helhet. Däremot har jag arbetat iterativt, testat olika lösningar och reflekterat över metodval i linje med agil utveckling.

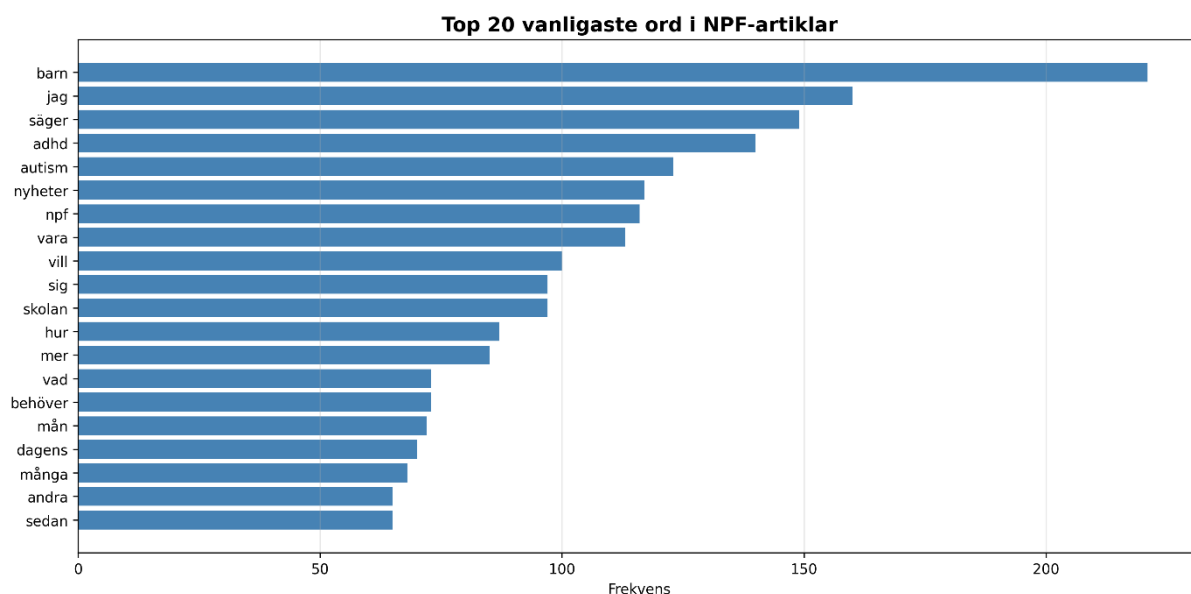
3.8 AI i processen

Generativ AI har använts som ett stöd i arbetsprocessen, i enlighet med EC:s riktlinjer. AI har fungerat som en iterativ partner i både tekniska och analytiska moment, exempelvis vid kodstrukturering, felsökning, metodbeskrivning och resultatreflektion. Genom att använda AI på detta sätt har arbetet gått mer effektivt samtidigt som kontrollen över innehållet behållits, för att säkerställa kvalitet.

4 Resultat och Diskussion

Analysen av 76 nyhetsartiklar om neuropsykiatriska funktionsnedsättningar (NPF) visar att rapporteringen innehåller flera återkommande teman. Skola, diagnoser, stödinsatser och samhällsdebatt är mest framträdande.

Ordfrekvensanalysen bekräftar detta (Figur 2). Ord som *barn*, *adhd*, *autism*, *säger* och *skolan* förekommer ofta, vilket tyder på att fokus ligger på unga personer och deras situation i utbildningssystemet. Ordmolnet och frekvensdiagrammet fungerade som visuella stöd för att identifiera centrala begrepp och eventuella brusord.

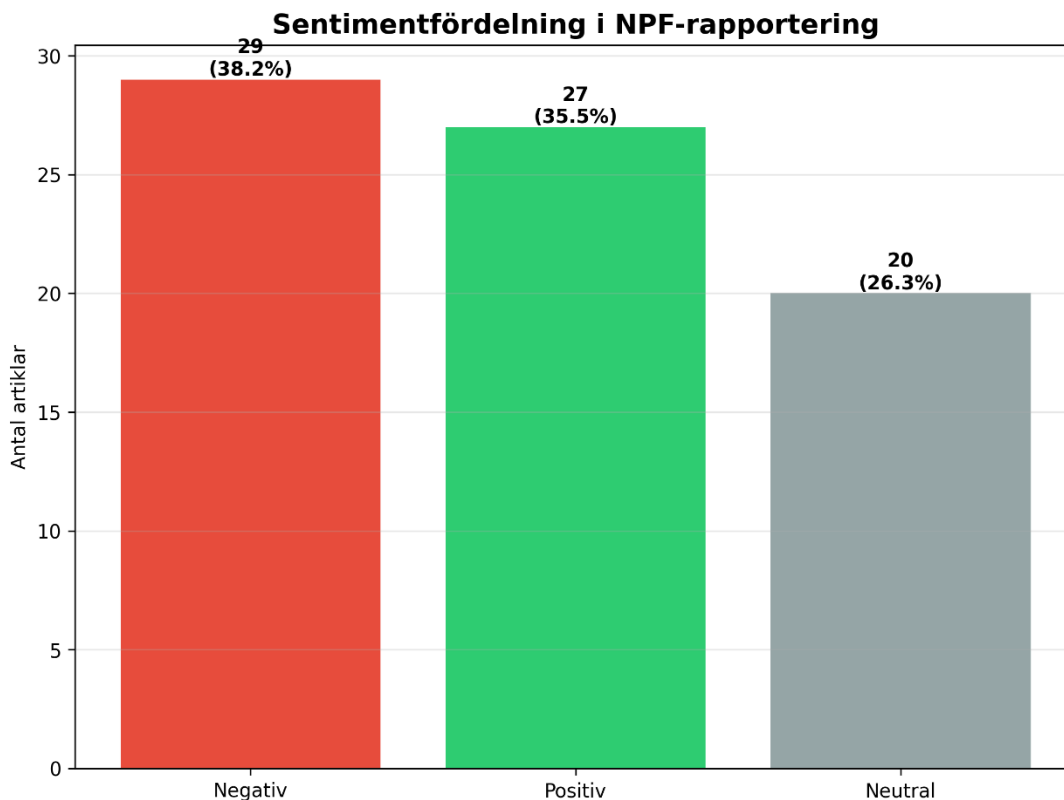


Figur 2. Ordfrekvensanalys som visar på de vanligaste orden i artiklarna.

Temaanalysen med BERTopic gav sex olika teman med olika antal artiklar i varje. Det största temat handlade om autism och barns rätt till stöd. Andra teman tog upp diagnosprocesser, vuxnas erfarenheter och politiska frågor. Modellen valdes eftersom den kan identifiera teman automatiskt. Den är särskilt användbar för att tolka språkets betydelse, vilket passar nyhetsartiklar med varierat innehåll.

Sentimentanalysen visade att tonläget i artiklarna är ganska blandat men något mer negativt än positivt. 38 % av artiklarna var negativa, 36 % positiva och 26 % neutrala. När jag kopplade sentimentet till temana såg jag att vissa teman, särskilt de som handlar om skola och brist på stöd, ofta har ett negativt tonläge. Andra teman, som handlar om framsteg eller samhällsinitiativ, är mer positiva eller neutrala.

Jag valde en enkel regelbaserad metod för sentimentanalysen. Den bygger på att jämföra texten med listor över positiva och negativa ord. Metoden är lätt att förstå och ändra, vilket passar bra i ett skolprojekt. Metoden har dock begränsningar när det gäller att tolka ironi och språkliga nyanser.

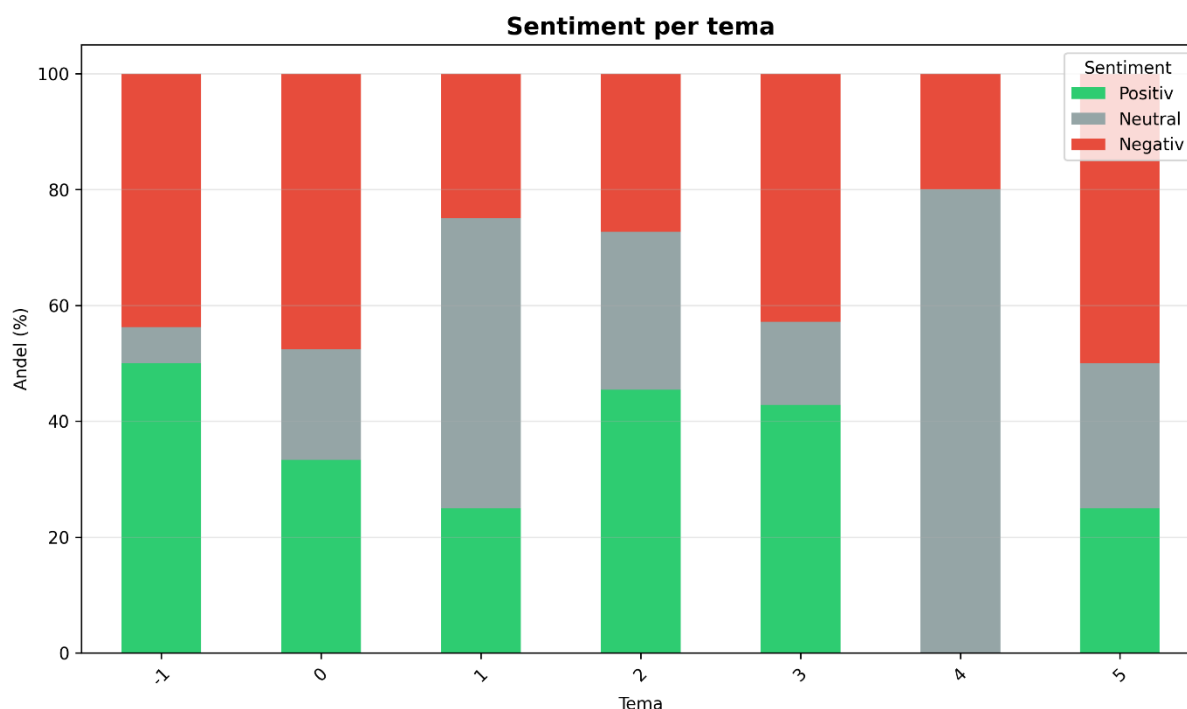


Figur 3. Fördelning av sentimentanalys

När jag kombinerade temaanalysen med sentimentet (Figur 4) blev det tydligt att vissa ämnen är mer känsloladdade än andra:

- **Tema 0 – Diagnos och barnperspektiv:** Mest negativt. Artiklarna handlar ofta om bristande stöd och frustration.
- **Tema 1 – Bruskluster från DN:** Verkar mest bestå av teknisk text från låsta DN-artiklar, som prenumerationserbjudanden och metadata.
- **Tema 2 – Skola och personliga berättelser:** Blandat tonläge. Frustration men också hopp.
- **Tema 3 – Statistik och utbildning:** Mest neutralt. Fakta och siffror dominerar.
- **Tema 4 – Flickor och diagnos:** Blandat. Kritik mot bristande förståelse, men också positiva exempel.
- **Tema 5 – Diagnosprocess och privata aktörer:** Tydligt negativt. Artiklarna handlar om ojämlik tillgång och kritik mot privata utredningar.

Att Tema 1 blev ett slags brus-tema visar en svaghet i modellen. När texten innehåller mycket teknisk information eller delar som inte är själva artikeln kan modellen skapa kluster som inte har något tydligt innehåll. Det visar att det är viktigt att rensa texten noga innan analysen och att man alltid behöver tolka resultaten med sunt förnuft.



Figur 4. Sentimentfördelning per tema (inklusive bruskcluster (-1))

5 Slutsatser

Syftet med denna rapport var att undersöka hur neuropsykiatriska funktionsnedsättningar framställs i svensk media, med fokus på teman, tonläge och språkliga mönster.

NPF framställs i svensk media som ett engagerande och emotionellt laddat ämne, med stort fokus på barn, skola och diagnoser. Tonläget varierar beroende på tema, och språket präglas av personliga berättelser och samhällsdebatt. Projektet visar hur textanalys kan ge insikt i både innehåll och attityder – och öppnar för vidare undersökningar kring representation, inkludering och påverkan.

Studien har gett en första inblick i hur NPF framställs i svensk nyhetsrapportering. För framtida utveckling skulle man kunna fördjupa analysen genom att jämföra fler olika mediekällor, analysera förändringar över tid eller använda mer avancerade modeller för sentimentanalys exempelvis KB-BERT.

Analysen skulle även kunna koppla till verkliga konsekvenser till exempel hur mediebilderna påverkar attityder i skolan eller vården. Ett nästa steg vore att bygga ett interaktivt gränssnitt där användaren kan filtrera artiklar efter tema eller tonläge vilket skulle göra analysen mer tillgänglig för exempelvis pedagoger eller beslutsfattare.

Jag har lärt mig mycket om hur man kan använda textanalys för att undersöka samhällsfrågor och hur viktigt det är att tolka resultaten kritiskt, särskilt när man använder automatiska metoder. Jag har också lärt mig vikten av att granska data noggrant innan analysen, exempelvis missen med låsta DN-artiklar. Projektet har också visat hur dataanalys kan användas för att förstå komplexa samhällsfrågor på ett strukturerat sätt vilket jag gärna hade arbetat vidare med.

6 Självutvärdering

1. Vad tycker du har varit roligast i kunskapskontrollen?
Att bygga denna typ av analysverktyg.
2. Hur har du hanterat utmaningar? Vilka lärdomar tar du med dig till framtida kurser?
Att hålla mig på en rimlig nivå, började för högt och fick backa då jag inte fick ihop det.
3. Vilket betyg anser du att du ska ha och varför?
Bör få ett G för att ha genomfört ett projekt. Dock har jag inte alls arbetat på det sätt jag brukar. Hade gärna gjort ett gränssnitt osv.
4. Något du vill lyfta till Antonio?

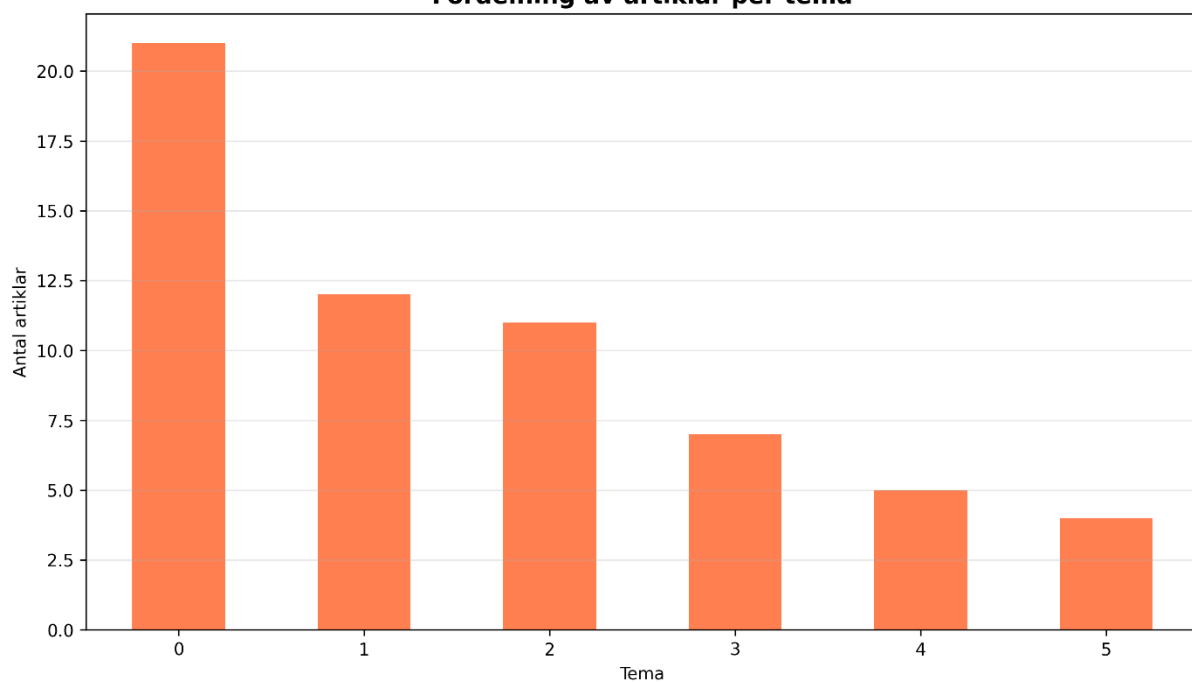
Appendix A

Ordmoln: NPF-relaterad nyhetsrapportering

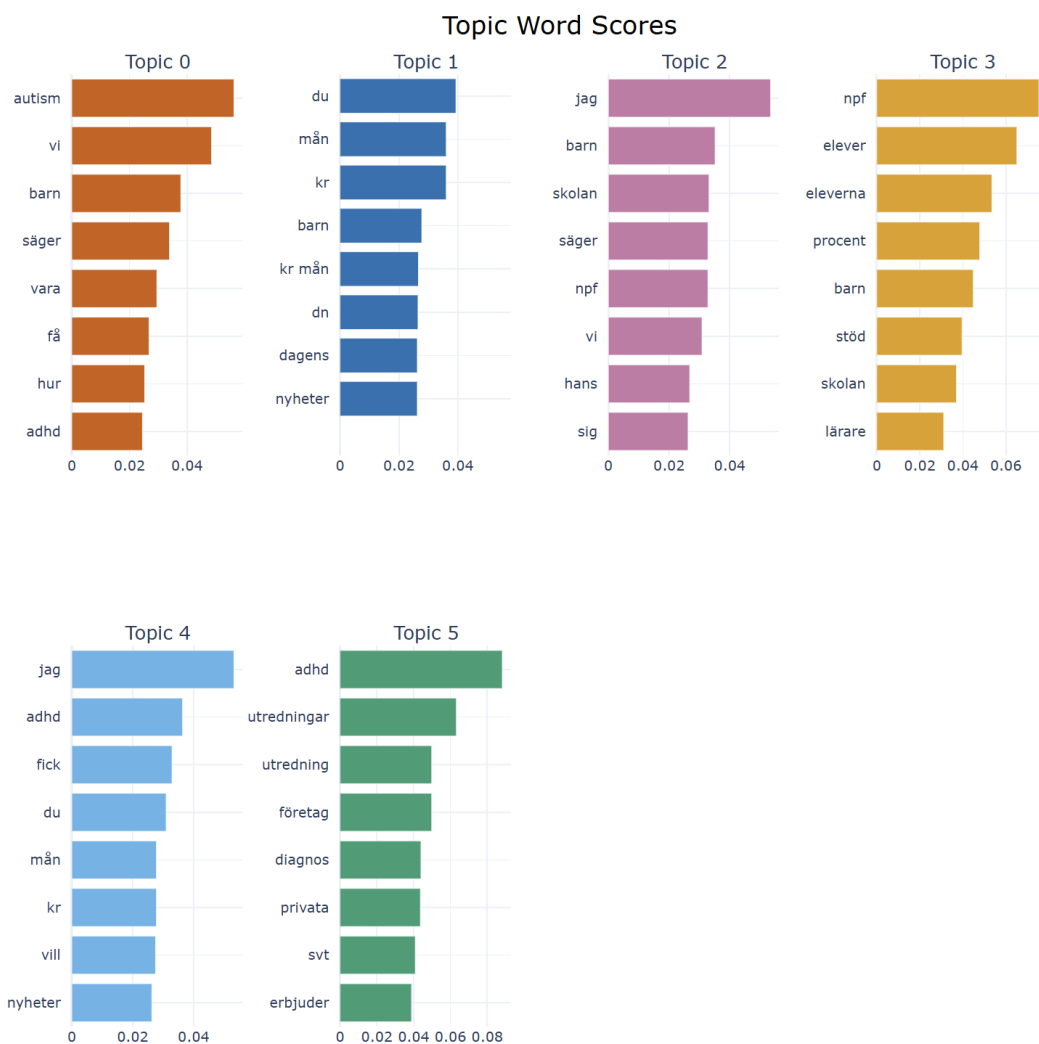


Figur 5 visar ett ordmoln baserat på de mest frekventa orden i nyhetsartiklarna om NPF. Visualiseringen ger en översiktlig bild av vilka begrepp som dominerar i materialet.

Fördelning av artiklar per tema



Figur 6. Stapeldiagrammet visar hur de 76 nyhetsartiklarna fördelar sig över de sex teman som identifierades med hjälp av BERTopic. Tema 0 är störst med 21 artiklar och handlar främst om autism och barns rätt till stöd.



Figur 7. Figuren visar de tio mest relevanta orden för varje tema som identifierades med BERTopic. Staplarna representerar hur viktiga orden är inom respektive tema, baserat på deras poäng i modellen.

Källförteckning

Hjärnfonden (u.å) *Fokusområde Neuropsykiatriska funktionsnedsättningar, Fakta om NPF.*

Tillgänglig på:

<https://www.hjarnfonden.se/om-hjarnan/fokusomraden/fokus-neuropsykiatriska-funktionsnedsattningar-npf/fakta-om-npf/> (Hämtad: 2025-10-31)

BERTopic (2024) *Home.* Tillgänglig på:

<https://maartengr.github.io/BERTopic/index.html> (Hämtad: 2025-10-31)