



Kandidatuppsats

IT-forensik & Informationssäkerhet 180hp

”Du är så mogen för din ålder...”

Identifiering av grooming med hjälp av en AI-språkmodell.

Digital Forensik 15hp

Halmstad, 2024-05-03

Jasmin Chroscielewski, Monia O’Neill



HÖGSKOLAN
I HALMSTAD

“Du är så mogen för din ålder...”

Identifiering av grooming med hjälp av en AI-språkmodell.

Kandidatuppsats

IT-forensik & Informationssäkerhet

Akademien för informationsteknologi

Högskolan i Halmstad

Maj 2024

Författare: Jasmin Chroscielewski, Monia O'Neill

Handledare: Eric Järpe, Peyman Mashhadi

Examinator: Urban Bilstrup

Sammanfattning

Genom litteratursökning och manuell datakompilering av sexualbrott mot barn, besvaras frågan “Vilka ord och fraser som förbrytare använder i konversationer är vanligt förekommande och kan användas som identifierande markörer av grooming?” och resulterade i en ordlista av könsord, sexuellt nedvärderande skällsord, och interjektioner som utrop, uppmaningar, och svordomar, som förekommer i högre utsträckning än i vardagliga konversationer. Denna lista användes för träning och test av en språkmodell som flaggar för skadlig data som kan indikera på grooming.

Med en semistrukturerad intervju, kompletterat med litteratursökningen av sexualbrottmål besvarades frågan “Vilka sociala plattformar används av förbrytare för att kontakta barn med syfte att utsätta dem för sexualbrott, och varför är dessa plattformar mer använda än andra?”. Dessa metoder påvisade att Snapchat hade en överväldigande majoritet och var den mest använda plattformen, följt av Instagram på en andraplats, samt Tiktok och Kik på en gemensam tredjeplats.

För att besvara den tredje frågeställningen “Kan identifiering av grooming underlättas genom Djupinlärning och Naturlig språkbehandling?” utfördes ett flertal experiment på den skapade detekteringsmodell med Naïve Bayes algoritmen som gav positiva utslag.

Motiveringen till användandet av AI var att underlätta för IT-forensiker och utredare i deras arbete genom att snabbt identifiera förekomsten av grooming. Eftersom mängden data som extraheras är väldigt omfattande och innehållsklassificering har stor potential för automatisering, kan AI-modeller avsevärt minska arbetsbördan och öka effektiviteten.

Nyckelord: Grooming; AI; Chattanalys; Ordinbäddning; Djupinlärning; Naturlig språkbehandling; Naïve Bayes; IT-Forensik; Text klassificering

Abstract

By investigating and analyzing court cases, the question of which are the most commonly used words and phrases during grooming attempts that could be used as grooming indicators. A list was compiled and utilized as “harmful” and “harmless” for a training- and test dataset for an AI-model. The list contained snippets of conversations where genital, sexually derogatory terms, commands, and swear words averaged higher than in daily conversation.

Through the methods of a semistructured interview and analyzing court cases, results of which social platforms perpetrators use to contact children could be compiled. This showed that Snapchat was by far the most prevalent platform used, followed by Instagram and in third place Tiktok and Kik Messaging. To answer this question, “harmless” data from the same platforms were used in the experiments.

The third and final question, pertaining to the possibility of using an AI in grooming detection, was answered through multiple experiments. In an effort to determine if the conversations contained grooming or not, similar in fashion to e-mail spam classification problems, a script with Naïve Bayes as the classifier produced positive results.

The goal of this study was to compile a list of words and phrases that, once used to train the model, could detect usage of these words and phrases. And notify the user if the current conversation has been flagged for suspected grooming attempts.

Keywords: Grooming; AI; Chat analysis; Word Embedding; Deep Learning; Natural Language Processing; Naïve Bayes; IT-Forensics; Text classification

Innehållsförteckning

Sammanfattning	2
Abstract	4
Innehållsförteckning	6
1. Introduktion	1
2. Problemformulering	3
2.1. Positionering av problemställningarna.....	3
2.2. Problematisering av frågeställning.....	4
3. Teori	5
3.1. Sociala plattformar.....	5
3.1.1. Instagram.....	5
3.1.2. Snapchat.....	5
3.2. Maskininläring och Tillämpningar.....	5
3.2.1. Naïve Bayes.....	5
3.2.2. PyTorch och SK-Learn.....	6
3.2.4. Hugging face transformers.....	6
3.3. Klassificeringsmått.....	6
4. Metod	9
4.1. Positionering av metod.....	9
4.2. Problematisering av metod.....	9
4.3. Intervju.....	10
4.3.1. Design.....	10
4.3.2. Deltagare.....	10
4.4. Relaterad litteratur.....	11
4.4.1. Avgränsningar.....	11
4.5. Detektionsexperiment med hjälp av maskininläring.....	11
4.5.1. Förbehandling av data.....	11
4.5.2. Val av modell.....	12
4.5.3. Val av maskininlärningsalgoritm.....	12
4.5.4. Träning av AI-modellen.....	12
4.5.5. Evaluering & test av klassificeringsmodellen.....	13
5. Resultat	15
5.1. Litteratursökning.....	15
5.2. Intervju.....	17
5.3. Experiment.....	18
5.3.1. Inhämtad data/ordlista.....	18
5.3.2. Träning.....	18
5.3.3. Evaluering.....	20
6. Diskussion	25

6.1. Fråga I.....	25
6.2. Fråga II.....	26
6.3. Fråga III.....	27
7. Etiska aspekter.....	31
8. Slutsats.....	33
9. Framtida forskningsinriktningar.....	35
Referenser.....	37
Appendix 1: Intervjufrågor.....	39
Appendix 2: Skadlig ordlista.....	42
Appendix 3: Oskadlig ordlista.....	54
Appendix 4: Koden för modellen.....	71
Appendix 5: Uträkningar.....	73

1. Introduktion

“Vad är det mest givande i din roll?”

“Varenda gång ett barn väljer att berätta om sin utsatthet...”

”Jag brukar säga det att, följer barnet med in i förhörssrummet så är allt därutöver en bonus.”

Allt fler barn, så unga som 7 år, använder internet i större utsträckning än tidigare. Covid-19 pandemin öppnade många dörrar ut till internet för låg- och mellanstadieelever när skolorna stängde ned. Med en ökad användning av internet, löper de även en högre risk för att exponeras av brottslighet.

Inom den svenska lagstiftningen refererar begreppet grooming till brotten “kontakt med barn för sexuellt syfte”, och “utnyttjande av barn för sexuell posering” i brottsbalkens sjätte kapitel. Med internets många funktioner för anonymitet och synonymitet har förövarens möjligheter ökat exponentiellt gentemot möjligheter i den fysiska världen. Förövare och offer har omedelbar tillgång till varandra genom chatt med text och bilder, röstsamtal, videosamtal och "livestreams".

Sexuellt utnyttjande av barn är ett växande problem, och med internets närvaro har det både blivit vanligare samt enklare för förövare att ta kontakt med barn i sexuellt syfte. Detta är ett stort problem, och trots att flera företag och plattformar försöker hindra grooming från att inträffa är detta inte tillräckligt. För barnen är fortfarande inte säkra på internet, och förövarna fortsätter att trakassera och begå övergrepp. (Ali et al., 2023; Lykousas & Patsakis, 2021).

Ett annat exponentiellt växande problem är hantering av den otroliga mängden data som produceras dagligen. För att hantera och kategorisera dessa datamängder är användningen av maskininlärning en användbar tillämpningsmetod. Liksom i fallet med skräppostfiltrering kan även groomingfraser filtreras genom att träna en språkmodell på stora datamängder som innehåller sådan kommunikation (Kumar et al., 2020). För närvarande är det utredare som manuellt går igenom stora mängder av detta material, vilket kan vara extremt psykiskt påfrestande. Genom att automatisera denna process med hjälp av AI blir arbetet inte bara mer tidseffektivt utan även den manuella arbetsbördan och den psykiska påfrestandningen minskar. Detta bidrar till en mer hållbar arbetsprocess och arbetsmiljö samt effektiviserar arbetet. Den tid som sparas kan istället läggas på andra viktiga delar av utredningsprocessen. (Riksrevisionen, 2021)

2. Problemformulering

Syftet med denna studie var att undersöka om man kan underlätta detekteringen av grooming vid brottsutredning med hjälp av en AI-modell, med fokus på att avlasta den fysiska och den psykologiska belastning som utredare hanterar. Detta uppnåddes genom att analysera språkmönster och ordval av gärningspersoner, samt deras val av sociala medier för att initiera kontakt med offer kunde två listor kompileras; en lista över sociala plattformar, och en lista för ord och fraser, som användes för att träna en AI-modell att detektera grooming.

Målet vid studiens slut var att skapa en modell som kunde detektera farliga ord och fraser i en chattkonversation. För att modellen sedan ska kunna avgöra om konversationen innehåller grooming eller inte. Genom att mata in chattkonversation till modellen ska den underlätta och effektivisera arbetet vid utredningen för dessa typer av brott.

Som underlag för studien och för att uppfylla syftet valdes tre frågeställningar:

- I. Vilka sociala plattformar används av förbrytare för att kontakta barn med syfte att utsätta dem för sexualbrott, och varför är dessa plattformar mer använda än andra?
- II. Vilka ord, fraser och tekniker är vanligt förekommande i konversationer med förövare och kan användas som identifierande markörer av grooming?
- III. Kan identifiering av grooming underlättas genom djupinlärning och naturlig språkbehandling?

2.1. Positionering av problemställningarna

En del av detta arbetes fokus var att identifiera förövarens val av plattformar, och undersöka hur och varför just dessa används. I likhet med andra arbeten har denna studie använt verktyg för att detektera grooming. Till skillnad från dessa, kommer denna studie att utveckla en egen AI-modell, en unik metod för att upptäcka grooming. Intervjuer kommer också att genomföras i kombination med en litteraturöversikt av rättsfall för att fastställa vilka plattformar som används mest. (Ashcroft et al., 2015; Borj et al., 2023)

Denna studie kommer likt Nikolovska (2020) att fokusera på en kombination av kvalitativa och kvantitativa metoder. Nikolovska gör en jämförelse av bredare teman och beteende som är cyberspecifika gentemot fall utanför internet. Nikolovskas datamängd representerade en förövare med flera offer; datamängden som denna studie använder utgör utdrag från 18 fall med en eller flera gärningspersoner mot ett eller flera offer. Brottmålen kommer analyseras med fokus, inte bara på varje specifik gärningpersons språkbruk, men även på identifiering av övergripande mönster och strukturer i konversationerna som är användbara för studiens syfte.

Experiment med simulerade barnkonton över olika plattformar för att kunna besvara problemformuleringen övervägdes också; ett exempel på detta var ett experiment utfört av Kamar et al., 2022, där tre "honeypot"-chattbotar skapades för att efterlikna unga flickor.

Studien undersökte hur föräldrars tillsyn kunde påverka hur dessa botar blev kontaktade. Denna typ av metod valdes dock bort då det är väldigt tidskrävande, och det finns ett behov av experter inom området samt juridiska myndigheter. (Kamar et al., 2022)

2.2. Problematisering av frågeställning

Det finns vissa svårigheter med att identifiera vilka plattformar som är mest använda för detta ändamål då det ständigt tillverkas nya applikationer och plattformar.

Det är viktigt att observera att bias kan oavsiktligt uppstå genom undersökning av frågan om vissa plattformar används i större utsträckning av förövare. Detta innebär dock inte nödvändigtvis att plattformen eller applikationen är indirekt farlig. Det gäller även applikationer som inte används i lika stor utsträckning, att dessa skulle vara mindre farliga handlar snarare om falsk säkerhet, gentemot förväntan av brott. Detta för att man ska vidta åtgärder på alla sociala medier oavsett statistisk utsatthet, men även att all statistik inte stämmer, då det existerar ett stort mörkertal.

Att identifiera ord/fraser som kan användas som identifierande markörer kan vara utmanande. Många ord som kan flaggas som misstänksamma i en konversation kan lika gärna vara oskyldiga i en annan. Det kan också gälla ironi, att orden/fraserna används på ett skämtsamt sätt, eller en konversation mellan två samtyckande vuxna. Detta är svårt att förstå utan en mänsklig bedömning. Så dessa markörer är inte en garanti på att ett brott har begåtts eller håller på att begås, men det är tänkt att vara ett verktyg som flaggar vid misstänkt grooming.

Denna kandidatuppsats har baserat forskningen på svenska lagar, svenskt rättsmaterial och andra svenska dokument. Arbetet fokuserade endast på Sverige och skapade en ordlista med svenska ord för att identifiera misstänkt grooming.

3. Teori

I detta teoriavsnitt presenteras relevant bakgrund för sociala plattformar, program och använda metoder.

3.1. Sociala plattformar

De plattformar som förekommer i relaterad litteratur, samt diskuteras i senare avsnitt, är främst Instagram och Snapchat då dessa förekom mest i den relaterade litteraturen. De som förekom färre tillfällen var Kik, Tiktok, messenger samt enstaka spelapplikationer.

3.1.1. Instagram

Instagram är en fotodelningsplattform med över 500 miljoner aktiva användare varje dag som tillåter användare att dela bilder med följare. Om kontot är offentligt kan det även ses av användare utanför sin följarkrets. Förutom att dela bilder erbjuder Instagram även möjlighet att skicka direktmeddelanden, vilket inkluderar text, foton och videor privat till andra användare. Användare har även alternativet att skicka bilder som raderas direkt efter visning, eller efter 24 timmar. Dessutom kan användare även radera meddelanden från chatten som de själva har skickat.

När det kommer till kommunikation med användare som man inte följer eller som inte följer en tillbaka, krävs det att den icke-följande parten godkänner en förfrågan för att delta i chatten. Denna förfrågan separeras från vanliga chattar och placeras i ett fält för förfrågningar tills den antingen godkänns eller tas bort. Användare som följer varandra kan däremot börja chatta direkt utan behov av någon förfrågan.

3.1.2. Snapchat

Snapchat är en populär multimedia-meddelandetjänst med cirka 422 miljoner aktiva användare varje dag. Plattformens huvudfokus är möjligheten att skicka och ta emot temporära meddelanden, bilder och videor. Denna funktion har gjort Snapchat till en populär plattform för delning av ögonblicksuppdateringar och spontana interaktioner mellan användare, då dessa ofta försvinner direkt efter visning.

En annan funktion på Snapchat är Snapkartan som tillåter användare att dela sin platsinformation i realtid med sina vänner. Denna information visas sedan på en interaktiv karta där användare även kan se när vänner senast var aktiva på plattformen.

Snapchat spotlight är en del av Snapchat som tillåter användare att skapa och dela korta videoklipp med andra användare. Här kan användare också interagera genom att dela, kommentera eller gilla dessa videoklipp.

3.2. Maskininlärning och Tillämpningar

3.2.1. Naïve Bayes

Naïve Bayes är en klassificeringsalgoritm som baseras på sannolikhetssteori och är en relativt enkel men användbar algoritm. Den är baserad på Bayes sats, som handlar om att

beräkna sannolikheten för en händelse baserat på tidigare kunskap om förhållanden som kan vara relaterade till händelsen. För klassificeringsproblem delas datamängden upp i olika klasser, Naive Bayes beräknar då sannolikheten vilken klass ett givet objekt tillhör.

Algoritmen använder träningsdata för att uppskatta sannolikheterna för olika klasser och attribut. Sedan jämförs dessa sannolikheter med attributen hos nya objekt för att klassificera dem i den mest sannolika klassen. Ett bra exempel på en tillämpning algoritmen är bra för är spamfiltrering då spam-mail innehåller specifika ord som är väldigt vanliga i dessa sammanhang. (Yang, 2018)

3.2.2. PyTorch och SK-Learn

PyTorch är ett bibliotek och ramverk för maskininlärning, skrivet i programmeringsspråket Python och används för att bygga modeller och applikationer. Vilket underlättar bland annat implementering och felsökning.

Likt PyTorch, är SK-Learn, eller SciKit-Learn också ett så kallat maskininlärningsbibliotek skrivet i python, som erbjuder olika verktyg, moduler och implementationer som automatiserar bland annat vektorisering, tokenisering, och normalisering av dataset.

Dessa bibliotek användes tillsammans för att skapa modellen som användes för experimenten.

3.2.3. Förbehandling av data

Preprocessing eller förbehandling är en viktig fas inom maskininlärning för bland annat normalisering och städning av datamängden. En viktig del av denna process är att ta bort ord som fyller ut en mening utan att tillföra betydande innehåll. Exempelvis "är", "den", "men", och andra vanliga ord som inte bidrar till att förstå meningens innebörd. (Al-jabery et al., 2020)

3.2.4. Hugging face transformers

Hugging face transformers är ett bibliotek som erbjuder förtränade transformer-modeller för olika ändamål. Exempelvis textklassificering, textgenerering eller objekt-detektering i fotografier. Det integreras också enkelt med olika ramverk som Tensorflow och PyTorch. En central del av dessa ändamål är naturlig språkbehandling (NLP) som handlar om att göra datorer kapabla att förstå, tolka och generera mänskligt språk.

En central del av NLP-processen är vektorisering, där mänskligt läsbar data omvandlas till numeriska vektorer som kan bearbetas av maskininlärningsalgoritmer. Hugging Faces Transformers-biblioteket tillhandahåller inte bara förtränade modeller för NLP-uppgifter utan även tokeniserare, som delar upp text i mindre enheter, eller "tokens", för att datorn effektivt kan bearbeta textdata.

3.3. Klassificeringsmått

Begreppen "True" och "False" indikerar om modellen tilldelade en konversation korrekt eller inte med den kategorin som konversationen faktiskt hade. Begreppet "positive" innebar att

modellen gav positivt utslag för grooming, medan “negative” innebar negativt utslag för grooming, oavsett korrekthet. (Kulkarni et al., 2020)

Ett True positive utslag innebar att modellen korrekt klassificerade grooming, de konversationer som tillhörde Harmful kategorin och identifierades av modellen som Harmful. Däremot, konversationer som identifierades som Harmful men tillhörde Harmless kategorin, benämns som ett False-positive utslag.

Konversationer som korrekt identifierades som icke-grooming, de konversationer som tillhörde Harmless kategorin och identifierades av modellen som Harmless benämns som True Negative. Däremot, konversationer som identifierades som Harmless men tillhörde Harmful kategorin, benämns som ett False negative utslag.

4. Metod

I den här studien användes intervju, litteraturöversikt och experiment som vetenskapliga metoder.

4.1. Positionering av metod

Möjligheten för AI att underlätta detektering av grooming har utretts tidigare, med liknande datamängd av Harmful och Harmless (Anderson et al., 2019). Genom en jämförelse av sex olika klassifiorare, däribland Gaussian Naïve Bayes, Random Forest, Logistisk Regression, och AdaBoost, kunde det konstateras att AI-tillämpningar var möjliga för groomingdetektering, men måste optimeras och förbättras innan dessa verktyg kan användas i aktuell verksamhet av utredare. Vikten av tidigare studier lades på jämförelsen mellan de olika klassificeringsalgoritmerna tillgängliga, samt deras effektivitet i förhållandet till förbehandlingsmetoder som appliceras på relevanta datamängder, medan denna studien fokuserade på effektiviteten av de kompillerade datamängder som används för att träna en klassificeringsmodell.

4.2. Problematisering av metod

Inhämtningen av den skadliga datamängden var från 18 olika domar. Det kan vara en felkälla, då det möjligtvis inte är så representativt för grooming i sin helhet. Det kan vara så att domarna som undersöktes kommunicerar på ett visst sätt som inte stämmer överens med hur det ser ut i andra fall. Trots detta är det viktigt att understryka att all inhämtad data härstammar från verkliga händelser, vilket inte alltid är fallet i andra studier.

Dessutom utvanns utdrag av konversationer och inte alltid en hel konversation, vilket minskar nyansen i ordlistan. Exempelvis fanns det inte alltid möjlighet att få den inledande konversationen, utan endast den skadliga delen där ett brott ofta redan begåtts, vilket inte är en fullständig representation av konversationen.

Dessa domar omfattas även av sekretess för att skydda målsäganden, vilket också kan ha bidragit till en datamängd med mindre kvalitet. Vissa domar hade väldigt mycket detaljer kring hur kommunikationen såg ut medan vissa knappt hade några exempel. Detta medförde väldigt mycket tid att gå igenom varje dom för att hitta relevant information.

Sedan användes endast domar som hade lett till en fällande dom vilket exkluderar fall där bevisningen inte varit tillräcklig. Detta innebär en selektionsbias där endast upptäckta och åtalade fall ingår, vilket kan påverka generaliserbarheten av resultaten.

Naïve Bayes algoritmen har även den en del begränsningar när det gäller att hantera komplex text, algoritmen antar att textdata är oberoende av varandra samt tar inte hänsyn till ordningsföljden av orden. Sedan har algoritmen svårare att fånga semantiska relationer och nyanser i språket. Den förstår inte meningsstrukturer, idiomatiska uttryck eller sarkasm, vilket kan leda till felaktiga klassificeringar. Felstavningar kan också påverka modellens prestation, därför kan det vara viktigt att förbehandla texten och rätta dessa. Då den arbetar med att matcha de exakta orden med varandra kan felstavade fraser gå obemärkta. I detta

arbete korrigerades inte felstavade ord, detta då det förekom en del felstavningar och för att få en så realistisk datamängd som möjligt behölls originaltexten. Felstavningar är en naturlig del av textbaserad kommunikation och vissa felstavningar kan ha betydelse i vissa sammanhang. Därför behölls felstavningarna för att öka modellens precision på verkliga scenarion. (Yang, 2018)

4.3. Intervju

För att besvara de två första frågeställningar; “Vilka sociala plattformar används av förbrytare för att kontakta barn med syfte att utsätta dem för sexualbrott, och varför är dessa plattformar mer använda än andra?” och “Vilka ord och fraser är vanligt förekommande i dokumenterade sexualbrott som förbrytare använder i konversationer med brottsoffer kan användas som identifierande markörer?” ska semi-strukturerade intervjuer hållas med brottsutredare inom ISÖB.

4.3.1. Design

En intervjuguide utvecklades baserat på problemformuleringarna. Intervjun inleddes med en kort introduktion av intervjuledarna. Därefter en kort bakgrund till studien, syftet, samt målet med intervjun, vilket innefattade vad som förväntades att bli besvarat och vad materialet skulle användas för.

Inledande frågor etablerade deltagarens relevanta bakgrund och erfarenhet till ämnet.

Därefter ställdes de mer utredande frågor som var av större relevans för studien.

Intervjun avslutades sedan med några avrundande frågor av mindre relevans, men som var utformad med avsikt att lämna deltagaren med en positiv känsla från intervjun, samt möjlighet för eventuell komplettering och feedback på intervjun för att utforska eventuella felkällor och bortfall av olika slag.

Efter avrundande frågor gavs information om återkoppling och publicering av arbetet den intervjuade vilket (Dalen, 2008) rekommenderar.

Vid utformningen av frågorna fanns flera saker i åtanke. Frågorna skulle vara enkelt formulerade och begripliga för alla parter, dock med vissa undantag av användning av ämnesspecifika begrepp som möter kandidatens kunskapsnivå. Däremot var frågorna specifikt utformade så att den intervjuade hade möjlighet att utveckla och förklara utförligt, för att undvika redundanta följdfrågor utöver fördjupningsfrågor, liksom metoden (Dalen, 2008) beskriver.

På grund av ämnets känslighet, fanns det etiska och moraliska aspekter att ha i åtanke. Frågorna gick inte in på djupet av individuella fall och var formulerade så objektivt som möjligt för att undvika influens eller påfrestning på kandidaten som skulle kunna leda till att denna avstod från att svara på frågan.

Intervjun var till för att sammanställa två tabeller som skulle diskuteras och redovisas i resultatet. På grund av att frågorna var fokuserade på data från brottsfall, hade intervjun statistiska inslag och undveks fall-specifika åsikter i majoriteten av frågorna.

4.3.2. Deltagare

För att ta reda på information för att utreda frågeställningarna söktes intervjupersoner efter yrkesrelevans och längd av erfarenhet inom ämnet; Dessvärre hade endast en respondent

möjlighet att medverka. Intervjupersonen har arbetat som cyberbrottsutredare med inriktning på Internetrelaterade Sexuella Övergrepp mot Barn, förkortat ISÖB. Deltagaren söktes genom LinkedIn för att underlätta kontakt och bredda den geografiska utsträckningen. Vidare hölls kontakten genom mejl och telefon.

4.4. Relaterad litteratur

Då kvaliteten på intervjuerna berodde på antal intervjuade, kompletterades frågeställning I och II med analys och datainsamling av relaterad litteratur i form av brottmål och lagstiftning. Den lagstiftning som studien förhöll sig till var sjätte kapitlet, 8 och 10a §§ i brottsbalken, *Utnyttjande av barn för sexuell posering* och *Kontakt för att träffa ett barn i sexuellt syfte*, samt de respektive försök till dessa brott. Brottmålen som kommer användas i denna studie återfinns på informationsportalen infotorg juridik.

För att utreda frågeställning två kommer brottmålen att granskas genom analys av skärmdumpar och annat bevismaterial som påvisar brottet. Utifrån detta kan en ordlista kompileras. Resultatet av denna metod kommer att användas som underlag för att besvara den tredje frågeställningen.

4.4.1. Avgränsningar

Vid val av vilka brottmål som ska användas har det valts att göra avgränsningar i vilka typ av brott det gäller, typ av bevismaterial och tidsram.

Vad gäller avgränsning av målen har endast de mål med brottsrubriceringen *Kontakt med barn i sexuellt syfte* och *Utnyttjande av barn för sexuell posering*, samt deras respektive försöksbrott valts för analys.

Då förövaren var drivande i brottet, samt att en vuxens slang-vokabulär inte förändras i lika stor utsträckning som hos ett barn, användes endast förövarens bidrag till konversationerna som dataset för att träna AI-modellen

Tidsramen på brottmålen som denna studien använde var så närtid som möjligt för att säkerställa aktuella och relevanta plattformar, därför valdes att begränsa studiens tidsintervall till de senaste fem åren.

4.5. Detektionsexperiment med hjälp av maskininlärning

För användning av maskininlärning utfördes ett flertal inledande steg. Pytorch valdes som ramverk för skapandet av klassificeringsmodellen. Användningen av ett ramverk var inte nödvändig, men underlättade skapandet av modellen med automatisering av vissa steg i processen, eller optimering av träning och behandling av data såsom att använda datorns grafikprocessor, om sådan finns, för att arbeta. Övriga fördelar med användning av PyTorch eller andra ramverk, som TensorFlow, är att de har en stor användarbas vilket underlättar i felsökningsprocessen för de som är nybörjare till maskininlärning.

4.5.1. Förbehandling av data

För att filtrera ut stoppord från datamängderna skapades ett python-script för att automatisera processen. En lista med vanliga svenska stoppord skapades där orden som användes var inhämtade från github. För att filtrera texten användes först `word.split()` för att

separera varje ord i frasen, sedan iterera över varje rad i datamängden och om orden inte är med i ordlistan, slås orden samman igen till en textsträng. Sedan sparades de nya textsträngarna i en ny lista för att sedan exportera den filtrerade datamängden till en ny csv-fil. (Al-jabery et al., 2020)

4.5.2. Val av modell

För denna studie användes en för-tränad modell från AI Sweden, modellen GPT-Sw3-356m som är en transformer-modell. Modellen är tränad på en datamängd som innehöll 320 miljarder tokens/ord på främst svenska men även danska, norska, isländska och engelska. Denna användes för att underlätta genereringen av representationsvektorer för det skapade datasetet som är nödvändigt för att modellen ska fungera. Modellen tokeniserar tränings- och testdatamängden och genererar sedan representationsvektorer för varje fras i datamängden. Genom att använda det yttersta lagret av modellen när vektorerna har extraherats, beräknas medelvärdet av dessa vektorer för att skapa sammanfattande representationsvektorer för varje fras.

4.5.3. Val av maskininlärningsalgoritm

Därefter togs steget att välja en lämplig maskininlärningsalgoritm för att bearbeta de genererade representationsvektorerna. Valet föll på en Naïve Bayes klassificerare, som ansågs vara lämplig för detta ändamål. De sammanfattade representationsvektorerna användes som inmatning i en Naïve Bayes klassificerare för att träna den på träningsdatamängden, och testa prestandan på testdatamängden. Först skapades klassificeraren med GaussianNB(), en variant av Naïve Bayes-distributionen som passar datamängder med normalfördelad varians och väntevärde, sedan tränades modellen med träningsdatamängden för respektive klass. Detta är grunden för att modellen ska kunna förutse om en konversation är skadlig eller inte. (Yang, 2018)

4.5.4. Träning av AI-modellen

För träningen av AI-modellen fördelades datamängden 80% för träning och 20% för validering och test av modellen. Den oskadliga datamängden inhämtas från Reddit-kommentarer, äldre personliga chattkonversationer, och Snapchat spotlight kommentarer, för att skapa en datamängd som är mångfaldig.

Först lästes träningsdatamängden in från en csv-fil som bestod av två kolumner, Fras och Värde där fras är "meddelandet" och värde är antingen harmful eller harmless. För att generera representationsvektorer för träningsdata, tokeniserades varje fras och passerades genom den förtränade GPT-modellen för att extrahera representationsvektorer från modellens sista lager. Alla vektorer sammanfogades därefter till en tvådimensionell tensor där varje rad representerar en fras.

En instans av GaussianNB() skapades och tränades med hjälp av dessa representationsvektorer. På samma sätt genererades representationsvektorer för valideringsdatamängden, vilka också sammanfogades till en tensor. Denna tensor användes för att utvärdera och finjustera modellens prestanda innan den slutliga testfasen av modellen.

Denna process av finjustering upprepades flera gånger för att hitta den optimala balansen och konfigurationen för modellen. Detta utfördes med ett experimentet för att undersöka hur varierande storlek på träningsdatamängden, borttagning av stoppord både från tränings- och valideringsdatamängden samt justering av klassfördelningen kunde påverka modellens prestanda. Efter varje justering inmatades den justerade valideringsdatamängden in i den tränade klassificeraren med predict-metoden för att undersöka hur förändringarna påverkade hur väl modellen kan förutse vilken klass en fras tillhör.

Efter att den optimala förbehandlingen av datamängden och den bästa konfigurationen för modellen har identifierats, var nästa steg att testa modellen för att utvärdera dess prestanda och tillförlitlighet mot data den tidigare inte har använt.

4.5.5. Evaluering & test av klassificeringsmodellen

En testdatamängd fördelat med lika många skadliga som oskadliga fraser användes för att evaluera modellens prestanda. Detta gjordes genom att applicera den tränade modellen på en separat datamängd som inte använts under träningsprocessen. Datamängden tokeniserades med samma teknik som tidigare, sedan användes tekniken predict igen. Utifrån modellens förutsägelser för testdatamängden applicerades även andra evalueringsmått för att utvärdera hur bra modellen förutsäger rätt klass.

Ytterligare ett test gjordes på en annan datamängd, denna innehöll uteslutande oskadliga fraser. Detta för att utvärdera hur bra modellen är på att klassificera fraser som endast innehåller ett typ av värde, för att visa att det inte är slumpmässigt och att modellen endast gissar sig fram då träningsdatamängden var fördelad jämlik. Denna metodik syftade till att verifiera att modellen inte lämnar förutsägelserna åt slumpen för att undvika att modellen överanpassar sig samt att den kan korrekt skilja mellan olika kategorier även när de är ensamma i datamängden.

Evaluering är en viktig del av processen vid skapandet av en modell. Det är ett viktigt steg för att ha en modell som kan klassificera korrekt. Utan evalueringsmetoder kan skadliga fraser missas vilket kan ha stora konsekvenser. Det finns flera val av metoder för evaluering av AI-modeller, bland annat träffsäkerhet. Alltså korrektheten av bedömningar vid positiva och negativa utslag. (Vujovic, 2021)

Accuracy eller träffsäkerhet mäter modellens förmåga att korrekt klassificera datamängden. Måttet skattar sannolikheten till true positive och true negative. Det beräknades genom att dividera det totala antalet korrekta förutsägelser, både negativa och positiva med hela datamängden.

$$Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + True\ Negative + False\ Negative + False\ Positive}$$

Ett annat sätt att utvärdera modellen är Precision. Där mätes andelen korrekta positiva förutsägelser, dividerat med det totala antalet positiva förutsägelser, både korrekta och falska. Detta mått skattar den betingade sannolikheten att en observation som klassats som positiv, verkligen är positiv, vilket kan uttryckas som $P(TP|Positive)$.

Denna metod är effektiv för att mäta hur ofta modellen klassar felaktiga positiva förutsägelser, detta är speciellt viktigt när konsekvenserna av dessa felaktiga klassningar är höga. Exempelvis vid bedömning av säkerhetshot.

$$Precision = \frac{True\ positive}{True\ Positive + False\ Positive}$$

Recall, ibland kallad känslighet, mäter modellens förmåga att klassificera alla faktiskt positiva exempel och undvika False negative.

Detta kan uttryckas som sannolikheten:

$$P(True\ Positive|Real\ Positive).$$

Recall beräknades som antalet True positive dividerat med summan av antalet True positive och False negative. Alltså dividerat på det riktiga totala antalet positiva fraser i hela datamängden.

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

Slutligen beräknades F1 som representerar både Precision och Recall för att ge en balanserad bedömning av modellens prestanda genom att beräkna det harmoniska medelvärdet för dessa två värden. Detta värde skattar sannolikheten för:

$$\frac{2 * P(TP|Real\ Positive)}{P(TP|Real\ Positive) + P(TP|Predicted\ Positive)}$$

och beräknas:

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall}$$

För dessa prestandamått beräknades standardfelen för vardera värde. Standardfelet anger standardavvikelsen för medelvärdet som i sin tur är en skattning av väntevärdet. Detta beräknades med följande formel.

$$StandardError_{95\%KI} = z * \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$Interval = \hat{p} \pm Standarderror$$

Där kan \hat{p} vara någon av de tidigare prestandamåtten Accuracy, Precision, Recall och F1-score. Beroende på vilket mått som väljs varierar n, baserat på nämnaren för varje uträkning.

För att beräkna konfidensintervallet med 95% konfidensnivå, multiplicerades standardfelen med 1.96, vilket motsvarar percentilvärdet som används till ett dubbelsidigt 95% konfidensintervall. Anledningen till att normalpercentiler används är att medelvärdet, enligt centrala gränsvärdessatsen, blir asymptotiskt normalfördelat oberoende av variabelns fördelning. Det innebär att om urvalsprocessen upprepas många gånger, förväntas 95% av konfidensintervallen innehålla det verkliga värdet av populationens parameter. (Bobbitt, 2021)(Wooldridge, 2023)

För att undersöka vilka sorts fraser som klassificerades fel skapades en kod som går igenom alla förutsägelser och jämför varje förutsägelse med det verkliga värdet. Om förutsägelsen inte stämmer överens med det verkliga värdet, läggs den felaktiga förutsägelsen till i en lista som innehåller frasen, den sanna etiketten och den förutsagda etiketten.

5. Resultat

I detta avsnitt redovisades resultatet från litteratursökningen, intervjun med Erica Engberg, och experiment av språkmodellen.

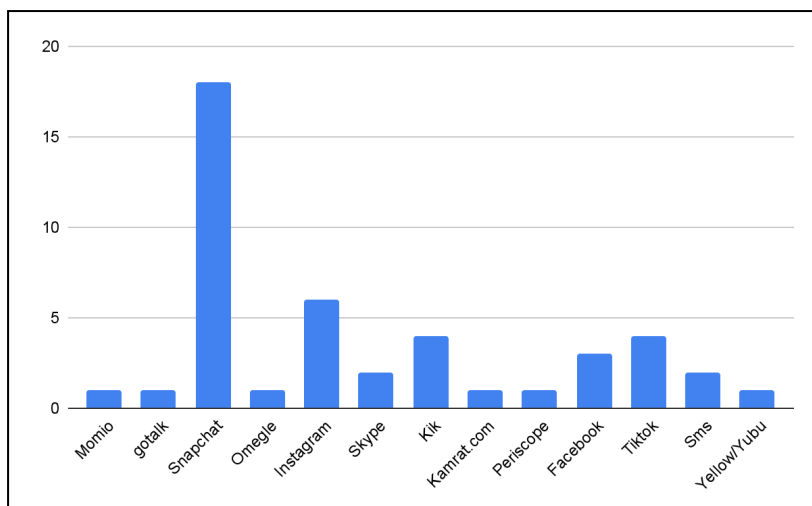
5.1. Litteratursökning

För påbörjandet av litteratursökningen valdes endast brottmål där relevant brottsrubricering tillämpades och gärningspersonen dömdes, övriga brottsrubriceringar togs inte i beaktande, inte heller de brottmål där brottsrubriceringarna tillämpades men den tilltalade friades från domen.

Via *Infotorg Juridik* söktes bland annat termerna “Utnyttjande av barn för sexuell posering” och “Kontakt för att träffa barn i sexuellt syfte.” Sökresultatet filtrerades efter datum och valdes i tur ordning. Över 130 sexualbrottmål analyserades och totalt 18 mål redovisade utdrag ur chatt mellan förövare och offer som bevismaterial. Det insamlade materialet i sin helhet återfinns i Appendix 2: Ordlista från litteratursökning.

De ord och fraser som var främst förekommande var varierande men följde liknande mönster.

Könsord som “snopp/kuk”, “snippa/fitta” förekom flest gånger i olika sammanhang som uppmaningar, eller önskemål på bilder med 11 respektive 25 gånger; ord på kroppsdelar som i olika sammanhang sexualiserades var “rumpa/röv” och “bröst/boobs/tuttar”, som förekom 18 gånger; kläder som nämndes främst, särskilt i sammanhang av avklädning, var trosa, bh, topp, magtröja, strumpor, bikini på 40 gånger i olika former; 36 gånger krävde förövaren skickande av “bild(er)” eller “video(r)” och i sammanhängande eller liknande kontext uppmanades vederbörande 34 gånger att “visa” något; ord som uttryckligen implicerade sexuella handlingar eller tillstånd, och dess olika böjelser förekom: “smeka” 12 gånger, “licka” 9 gånger, “trycka” 4 gånger, “pulla” 11 gånger, “ knull” 11 gånger, “spruta” 18 gånger, “våt” 11 gånger, “orgasm” 5 gånger, “kåt” 22 gånger, “oskuld/virgin” 6 gånger; samt förekomsten av orden “sex” och “sexig” i olika format och sammanhang förekom 16 gånger sammanlagt.



Figur 1: Förekomsten av kontaktmedel mellan offer och förövare: Momio,1; GoTalk,1; Snapchat,18; Omegle,1; Instagram,6; Skype,2; Kik,4; Kamrat.com,1; Periscope,1; Facebook,3; Tiktok,4; SMS,2; Yellow/Yubo,1;

Vid undersökningen av vilka kontaktmedel som förekom flest gånger i brottmålen visualiseras i Figur 1. Det visade att Snapchat var allra vanligast eftersom det förekom i alla använda brottmål, 18 gånger. Efter det kom Instagram och på en delad tredjeplats, Kik och Tiktok. Viktigt att notera är att dessa inte är för varje incident, utan endast räknat som en gång beroende på om förövaren använt plattformen eller inte.

På Snapchat förekom det ofta att förövarna gick med i gruppchatter för att hitta sina offer, eller via en funktion på Snapchat som kallas "lägg till snabbt" som tillåter användare att lägga till nya gemensamma vänner. Om en förövare lyckats få tag i ett offers Snapchat-konto fanns det större möjlighet att hitta nya offer.

Förövarna låtsades vid flera instanser att de var unga flickor och skickade bilder av andra brottsoffer, detta för att övertala det nya brottsoffret att våga skicka bilder för att förövaren har vågat skicka bilder "på sig själv". På detta sätt lyckats manipulera brottsoffren att det inte är något konstigt att utbyta bilder, för de är ju bara två tjejer som skickar till varandra.

Det observerades även att förövarna utpressade offren genom hot om att skada dem eller någon i deras närhet. Vanliga hot innefattade att sprida bilderna till pedofilsidor, föräldrar, skolan eller vänner samt att skapa fejkprofiler med deras bilder. Hot som exempelvis, "alla ska få se vad du har gjort," eller "Du borde berätta för din mamma vad du gjort för videos och hur många killar o tjejer du visat din snippa för."

Förövarna använde också andra tekniker för att manipulera offren, såsom att erbjuda pengar eller andra former av belöningar för att utföra vissa handlingar. Dessa belöningar kunde inkludera överföringar, presentkort, berusningsmedel eller tobak.

En funktion på plattformen är Snapkartan, som unga även använder. När dessa har Snapkartan aktiverad, har det funnits tillfällen där förövaren utnyttjar detta till hota barnen för egen vinning.

Även på Instagram som förekom i 6 olika brottmål uttrycktes komplimanger och skickande av bilder mellan offer och förövare där förövaren ofta låtsades vara en jämnårig tjej eller kille. I ett av brottsfallen där förövaren uppgett sig vara en jämgammal tjej som

offret, övertygat offret att följa förövarens konto tillbaka, och sedan kommunicera med varandra, uppgav offret att det till en början hade känts kul att ha kontakt med en jämnårig kompis att dela sina tankar med. När det sedan övergick till att skicka bilder blev det oerhört jobbigt. När målsägande återfann bilder av hennes "nya kompis" på Google, förstod hon snabbt att det var någon annan hon pratade med.

Ett annat vanligt scenario var att förövaren låtsades vara en tonårspojke och använde romantiska fraser för att manipulera offren. Trots att det kunde börja som en oskyldig konversation, eskalerade det ofta till krav på sexuella bilder och andra övergrepp.

Dessutom förekom det ofta att brottsoffren raderade och blockerade förövarnas konton. Att förövarna då istället gjorde nya konton och fortsatte trakassera samma personer. Målsäganden i ett av fallen hade totalt blockerat över 20 olika konton som tillhörde samma person.

I ett fall där målsägande blivit kontaktad via Snapchat av samma gärningsman flera gånger, uppgav hon att ett flertal försök till att blockera honom hade gjorts, men att denne, fem minuter senare, skapat ett nytt konto och kontaktat offret igen. Enligt brottmålet hade denne förövare skapat över 100 konton med olika namn, inklusive offrets vänners namn. Han hade gjort nya konton och kontaktat henne varje dag i en månad, möjligtvis två, och hade gjort liknande försök på ett flertal av hennes vänner, samt hotat om att kidnappa målsägande i skolan.

5.2. Intervju

En intervju utfördes med Erica Engberg som arbetar som utredare på polisens avdelning Internetrelaterade Sexuella Övergrepp mot Barn. Engberg är i grunden utbildad socionom sedan 2008, och har arbetat med brottsutsatta barn nästan hela sin karriär, och arbetat på ISÖB sedan 2019. Där arbetar Engberg som barnförhørsledare och granskare. Under intervjun framkom det att förbrytare inte använder sig av några specifika fraser som direkt går att identifiera att det är grooming, och att begreppet grooming förekommer mer sällan nuförtiden. Förövare är i dagsläget väldigt snabba, och att den initiala kontakten går väldigt snabbt över till en förfrågan om bilder och liknande. Det förekommer också fall där det är längre kontakt än så, vilket kan gälla en längre period där offret exempelvis blir förälskad i sin förövare. Engberg nämner också att brottet "kontakt med barn i sexuellt syfte" oftast övergår till utnyttjande av sexuell posering, då man ofta förmår barnet att skicka bilder.

Plattformarna Snapchat och Instagram upplever Engberg förekomma mest i sitt arbete, vilket beror på att plattformarna är lättillgängliga och populära, särskilt bland yngre. Varför förövarna använder dessa plattformar är främst för att bilder och liknande försvinner på en gång och att plattformen upplevs som anonym av förövare. Den främsta anledningen dock, är att förövarna alltid kommer att finnas där potentiella offer finns. Vid enstaka tillfällen förekommer även Tik Tok, Yubo och Roblox som initial kontakt, men oftast förflyttas kontakten till en annan plattform, då kommunikation på dessa plattformar övervakas delvis av moderatorer och inte har samma möjligheter att gömma chattar.

I de flesta fall som Engberg har arbetat med ljuger förövaren om sin ålder och låtsas vara yngre; det är endast få fall där förövaren är jämnårig. Vuxna förövare försöker använda åldersadekvata ord för att verka yngre än vad de är, men använder inte förkortningar på samma sätt som barn gör. Detta kan i sin tur kan göra att barnen misstänker att det är en vuxen de pratar med.

Det är oftast det manipulativa beteendet i förbrytares språk som utgör varningssignaler. Tjatande och övertygande om att få bilder, och smickrande komplimanger, till exempel “han sa att jag var snygg”, “han sa att jag var söt” och liknande meddelanden återberättar offer bland annat i förhör. Den största utmaningen när det gäller dessa brott är för utredare, att låta offer komma förbi dess skuld och våga berätta sin upplevelse.

Kunskap saknas mest när det gäller dessa typer av brott, och Engberg förklarar att vårdnadshavare måste sluta förfasas över att barnen skickar bilder eller att de har kontakter på internet. Vårdnadshavare måste finnas där deras barn finns och måste förstå vad barnen gör på internet. Alla vuxna som möter barn måste ha bättre koll på hur de sociala sfärerna runt barnen ser ut, och kunskap om risker om dessa brott borde inkluderas redan i läroplanen.

En av de viktigaste sakerna som Engberg ansåg behövs för att förhindra barns utsatthet, är att vuxna och barn behöver ha en bättre öppen dialog. Det är alltid den vuxnes ansvar att hålla koll på vad barnen gör på internet, utöka kunskapen om hur teknologin eller applikationer fungerar samt veta hur klimatet ser ut på internet. Bemötande med ilska, skam och skuldbeläggning främjar inte god dialog med barnet, utan gör det svårare för brottsoffer att vara öppna om de skulle utsättas flera gånger.

Att hålla koll och ha kunskap är viktigt, men det är även en fråga om integritet, och behöver inte nödvändigtvis leda till att kolla igenom sitt barns konton. Med stigande ålder måste hänsyn tas till den personliga integriteten och även den sexuella integriteten. Det är bättre att fråga vad barnet gör på nätet utan att behöva inkräkta på barnets integritet och förminska dess tilltro till den vuxne.

När frågan ställs om det används något liknande verktyg inom polisen som det detta arbete har planerat att skapa, nämner Engberg att verktygen som finns inte används för att söka på specifika fraser. De används främst för spårning och bevissäkring, vilket utredarna hanterar manuellt i programmen.

5.3. Experiment

Experimenten bestod av organisering samt uppdelning av den insamlade datamängden följt av byggande, träning, validering och test av modellen.

5.3.1. Inhämtad data/ordlista

416 skadliga fraser och 480 oskadliga fraser inhämtades. Den oskadliga datamängden bestod av chattkonversationer, innehållande vardagligt prat. Det inkluderade separata utdrag från chattkonversationer mellan enbart barn, mellan barn och vuxna, och mellan enbart vuxna för att skapa en mångsidig och bred datamängd. Cirka 250 fraser reserverades för validering och test av modellen, och resterande användes för att träna modellen.

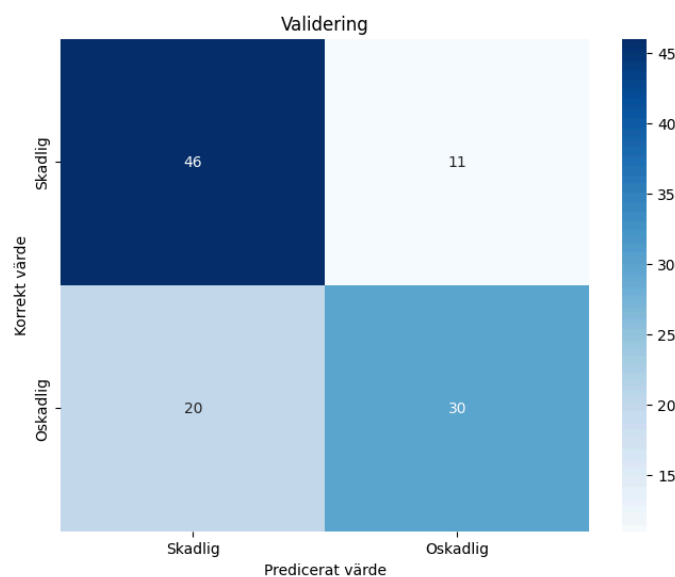
5.3.2. Träning

Vid träning av klassificeraren användes “predict” för att undersöka hur bra modellen kunde förutse klasserna för testdatamängden. I Tabell 1 presenteras hur olika justeringar påverkar träffsäkerheten. Vid första mätningen var 58.33% av modellens förutsägelser korrekta. På det andra försöket, tränades modellen om med fler rader oskadlig data och fick då ett högre resultat. Det visade en ökning på träffsäkerheten med 7.4% från föregående

mätning. Det undersöktes även om stoppord har någon betydelse för träffsäkerheten i detta fall. Det visade sig att utan stopporden blev träffsäkerheten sämre och sjönk med ca 3.7%. När den oskadliga datan övergick 60% av den totala mängden började träffsäkerheten sjunka. Det bästa resultatet var på en fördelning av 55% oskadlig och 45% skadlig data, en datamängd på 690 rader och när stopporden inte hade filtrerats bort.

Fördelning % Harmless/Harmful	50/50	55/45	55/45	55/45	65/35	61/39
Träffsäkerhet	0.5833	0.6574	0.5278	0.7102	0.5093	0.5556
Radantal	615	690	690	690	883	794
Filtrerad?	Nej	Nej	Endast träningsdatamängden	Ja	Nej	Nej

Tabell 1. I tabellen visualiseras hur träffsäkerheten påverkas av olika faktorer såsom större datamängd, ojämlig fördelning och filtrering av stoppord. Den första raden beskriver fördelningen av klasserna i procent. Andra raden är Accuracy ($\frac{TP+TN}{Total\ antal}$). Tredje raden visar totalt antal rader, och sista raden om texten var filtrerad från stoppord.



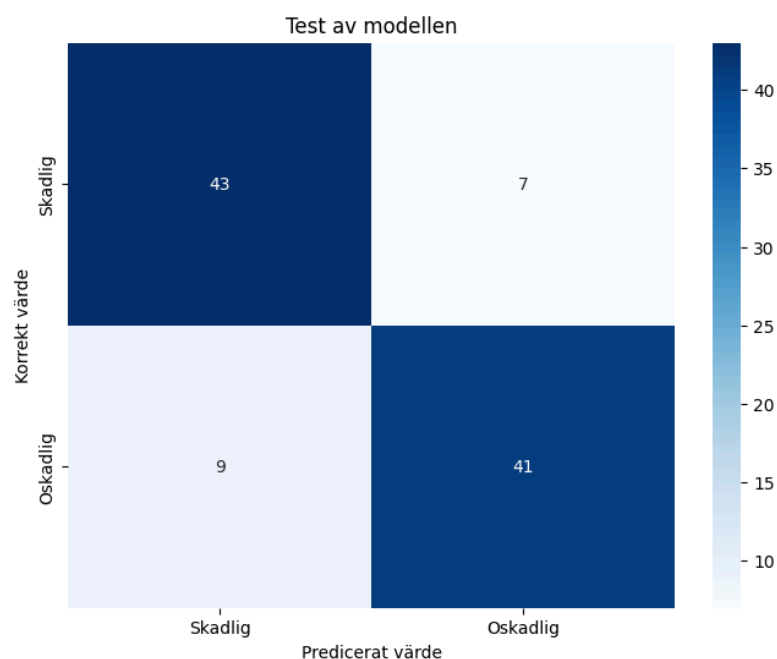
Figur 2. X-axeln visar det predicerade värdet, Y-axeln representerar det korrekta värdet. Matrisen visar att modellen förutsåg 46 skadliga fraser rätt och elva stycken klassades fel. För den oskadliga datamängden kunde modellen endast förutse 30 stycken fraser rätt, och 20 stycken blev klassade som skadliga istället.

Datamängden låg på totalt 107 fraser, där skadlig data var 57 stycken och oskadlig 50 stycken fraser. I figur 2, identifierades 46 skadliga fraser korrekt som skadliga, vilket motsvarar True Positive. Däremot klassificerades endast 30 oskadliga fraser korrekt som oskadliga, vilket representerar True Negative. Antalet False Negative var 11, vilket innebär att 11 skadliga fraser felaktigt klassificerades som oskadliga. Sedan klassificerades 20

oskadliga fraser felaktigt som skadliga, vilket är False Positive. Efter att ha justerat modellen med hjälp av valideringsdatamängden uppnåddes en träffsäkerhet på 71%. Detta indikerar att 71% av observationerna klassificerades korrekt.

5.3.3. Evaluering

För evaluering och test av AI-modellens prestanda användes den datamängd som sparats för test av modellen, som låg på 150 fraser.



Figur 3. X-axeln visar det predicerade värdet, Y-axeln representerar det korrekta värdet. Matrisen ovan visar att modellen klassificerade 43 skadliga fraser av 50 totalt, rätt och sju fel. Av de oskadliga fraserna klassades 41 stycken rätt och nio stycken fel av 50.

I figur 3, identifierades 43 skadliga fraser korrekt som skadliga, vilket motsvarar TP. Sedan klassificerades 41 oskadliga fraser korrekt som oskadliga, vilket representerar TN. 7 skadliga fraser klassificerades felaktigt som oskadliga. Sedan klassificerades 9 oskadliga fraser felaktigt som skadliga, vilket är False Positive.

I Tabell 2 återfinns de fraser som modellen felaktigt klassificerade som Harmless eller Harmful från testdatamängden. Där fördelningen av klasserna var lika för datamängden.

Felaktiga förutsägelser	Predicerat	Verkligt
jätte kåt	Harmless	Harmful
Jag skriver vad du ska skicka så ska du skicka det till mig oavsett vad det är	Harmless	Harmful
vill ha rövbilder och tungbilder/videos	Harmless	Harmful
skicka mig bilder annars skickar jag vidare	Harmless	Harmful
Tryck in pennan i fittan	Harmless	Harmful
din kropp bby	Harmless	Harmful
du får 500 kr nu, Ta av alla dina kläder, Så sprutar jag fortare	Harmless	Harmful
Ingen aning, men jag är redo att bli skrämnd haha! Vill du joina?	Harmful	Harmless
Tack! Jag kämpar mig igenom det, men det är alltid bra med lite extra övning.	Harmful	Harmless
är du med på festen ikv	Harmful	Harmless
färgar mitt hår nu	Harmful	Harmless
Ses då!	Harmful	Harmless
Fett! Det låter som du har koll på vad du gillar.	Harmful	Harmless
Äh, jag tror jag drar åt naturvetenskapligt håll.	Harmful	Harmless
Kom ihåg att ta hand om dig själv, det är viktigt att ta det lugnt ibland.	Harmful	Harmless
Absolut, om du behöver hjälp är jag här.	Harmful	Harmless

Tabell 2. Hur modellens felaktiga förutsägelser såg ut för den normalfördelade testdatamängden, där korrekta värdet samt det predicerade värdet presenterats.

```
Klassificering av text: vill ha rövbilder och tungbilder/videos
Resultat av klassificeringen: ['Harmless']
Klassificering av text: vill ha rövbilder
Resultat av klassificeringen: ['Harmless']
Klassificering av text: vill ha tungbilder
Resultat av klassificeringen: ['Harmfull']
Klassificering av text: vill ha tungbilder/videos
Resultat av klassificeringen: ['Harmless']
Klassificering av text: vill ha bilder
Resultat av klassificeringen: ['Harmfull']
```

Figur 4: En närmare granskning av hur modellen klassar delar av den felklassade frasen “vill ha rövbilder och tung bilder/videos”.

I figur 4 klassade modellen “vill ha rövbilder” och “vill ha rumbilder/videos” som harmless, däremot klassade den “vill ha bilder” och “vill ha tungbilder” som harmful.

```
Skriv en rad text eller 'n' för att avsluta:
> Tryck in pennan i fittan
> tryck in pennan i fittan
> n
Klassificering av text: Tryck in pennan i fittan
Resultat av klassificeringen: ['Harmless']
Klassificering av text: tryck in pennan i fittan
Resultat av klassificeringen: ['Harmfull']
```

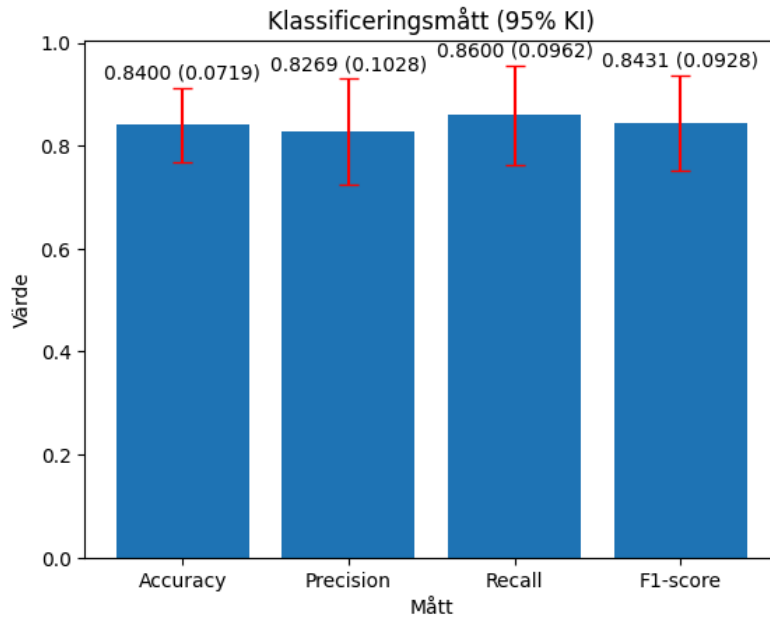
Figur 5: Hur utdatan såg ut vid närmare granskning av false negative frasen “Tryck in pennan i fittan” som återfinns i tabell 2.

Figur 5 visar hur modellen klassar en false negative fras från tabell 2 om den ändrats till en liten inledande bokstav.

```
Klassificering av text: du får 500 kr nu, Ta av alla dina kläder, Så sprutar jag fortare
Resultat av klassificeringen: ['Harmless']
Klassificering av text: du får 500 kr nu
Resultat av klassificeringen: ['Harmless']
Klassificering av text: Ta av alla dina kläder
Resultat av klassificeringen: ['Harmfull']
Klassificering av text: Så sprutar jag fortare
Resultat av klassificeringen: ['Harmfull']
```

Figur 6: En granskning av false negative frasen uppdelad separat. Modellen klassar två delar av frasen som harmful.

I figur 6 presenteras en längre fras som delades upp vid kommatecknet i frasen, det visade att varje fras var för sig klassades som harmful.



Figur 7: Diagrammet visar prestandamått för modellen med 95% konfidensintervall. De olika prestandamåtten inkluderar Accuracy med värdet 0.8400, intervall: 76.81%-91.19%. Precision med värdet 0.8269, intervall: 72.41%- 92.97%. Recall med värdet 0.8600, intervall: 76.38%-95.62%. F1-score med värdet 0.8431, intervall: 75.03%-93.59%.

I Figur 4 presenterades prestandamåtten för modellen med 95% signifikansnivå. Den ger en visuell representation av modellens klassificeringsförmåga.

Accuracy hade ett värde på 0.8400 vid testdatamängden. Med 95% signifikans kan det fastställas att noggrannheten för modellen ligger inom intervallet mellan 76.81% och 91.19%.

Precision fick ett värde på 0.8269 vid testdatamängden. Med 95% signifikans kan det fastställas att noggrannheten för modellen ligger inom intervallet mellan 72.41% och 92.97%.

Recall fick ett värde på 0.8600 vid testdatamängden. Med 95% signifikans kan det fastställas att noggrannheten för modellen ligger inom intervallet mellan 76.38% och 95.62%.

F1-Score fick ett värde på 0.8431 vid användning av testdatamängden. Med 95% signifikans kan det fastställas att noggrannheten för modellen ligger inom intervallet mellan 75.03% och 93.59%.

Vid testet av datamängden som endast innehöll oskadliga fraser uppnådde modellen ett värde på 0.7143 i Precision. De fraserna som klassades felaktigt som skadliga har presenterats i tabell 3.

False Positive

Chillar bara, du?	Din story var så cringe igår, haha.
Tja, ska vi dra o kaka nåt?	Dags att flexa våra moves
Sjukt najs väder idag, va?	Några chillplans för ikväll?
Dina snaps är alltid så kul!	Den där memen lol
Asså seriöst, så trött på skolan.	Några dope tracks att rekommendera?
Fett trött idag, behöver en powernap.	Ska vi chilla snart?
Ska vi linka sen?	Kan du fixa några snacks?

Tabell 3: Oskadliga fraser som modellen har klassificerat felaktigt som skadliga vid det andra testet med endast oskadlig data.

```
Skriv en rad text eller 'n' för att avsluta:  
> Chillar bara, du?  
> Chillar bara  
> Bara, du?  
> n  
Klassificering av text: Chillar bara, du?  
Resultat av klassificeringen: ['Harmfull']  
Klassificering av text: Chillar bara  
Resultat av klassificeringen: ['Harmless']  
Klassificering av text: Bara, du?  
Resultat av klassificeringen: ['Harmfull']
```

Figur 8: En skärmdump som visar hur modellen klassificerar delar av en mening som var false positive under föregående test.

Figur 8 illustrerar hur modellen hanterar olika delar av frasen “Chillar bara, du?” som klassades som skadlig av modellen. I figuren klassar modellen “chillar bara” som harmless men “Bara, du?” som harmful.

```
Klassificering av text: Ska vi chilla snart; Klassificering av text: ska vi chilla  
Resultat av klassificeringen: ['Harmfull'] Resultat av klassificeringen: ['Harmfull']  
Klassificering av text: ska Klassificering av text: ska vi snart  
Resultat av klassificeringen: ['Harmless'] Resultat av klassificeringen: ['Harmfull']  
Klassificering av text: vi Klassificering av text: ska snart  
Resultat av klassificeringen: ['Harmless'] Resultat av klassificeringen: ['Harmfull']  
Klassificering av text: chilla Klassificering av text: ska chilla  
Resultat av klassificeringen: ['Harmless'] Resultat av klassificeringen: ['Harmless']  
Klassificering av text: snart Klassificering av text: ska vi  
Resultat av klassificeringen: ['Harmless'] Resultat av klassificeringen: ['Harmless']
```

Figur 9 & 10: Figurerna visar hur modellen klassificerade olika delar av meningen “Ska vi chilla snart?”

I Figur 9 & 10 visualiseras hur modellen klassificerar olika kombinationer av meningen “Ska vi chilla snart?” som återfinns i tabell 3. Modellen klassade alla orden separat som harmless men “ska vi chilla”, “ska vi snart” och “ska snart” klassades alla som harmful.

6. Diskussion

Syftet med denna studie var att undersöka hur detekteringen av grooming kunde automatiseras vid en brottsutredning med hjälp av en AI-modell. Syftet uppfylldes genom att utreda studiens tre frågeställningar:

- I. Vilka sociala plattformar används av förbrytare för att kontakta barn med syfte att utsätta dem för sexualbrott, och varför är dessa plattformar mer använda än andra?
- II. Vilka ord och fraser som förbrytare använder i konversationer är vanligt förekommande och kan användas som identifierande markörer av grooming?
- III. Kan identifiering av grooming underlättas genom djupinlärning och naturlig språkbehandling?

Dessa frågeställningar besvarades i resultatdelen och diskuteras ytterligare i detta avsnitt.

6.1. Fråga I

Varför Snapchat förekom flest gånger kan bero på att det erbjuder möjligheten att radera meddelanden, vilket innebär att det försvinner för båda parter. Även chattmeddelanden försvinner efter 24 timmar om de inte sparas av någon av parterna, vilket också gäller videor och bilder. Detta kan skapa en falsk trygghet hos förövarna, som tror att de är anonyma, och därmed bidrar till valet av denna plattform.

Instagram som förekom flest gånger efter Snapchat har liknande funktioner. Möjligheten att automatisk eller manuellt radera chatinnehåll, som på Snapchat och Instagram, kan försvåra en utredning eller dölja misstanken att ett brott har begåtts från första början. Detta kan leda till ökat missbruk av plattformarna och underlätta för förövare att begå övergrepp. Dessa plattformar är också väldigt populära bland barn och ungdomar, trots befintliga åldersgränser för användare.

Mycket kan diskuteras kring varför förövare väljer dessa plattformar. Att Instagram och Snapchat förekom främst i denna undersökningen kan bero på exempelvis vilka brottmål som valdes för analys. För att få så relevant information som möjligt, valdes mål i nära tidsspann till varandra samt närhet till studiens utformning. Om denna studie skulle återskapas, kan resulterande plattformar förändras på grund av bland annat popularitet, användarbas, och tid. Plattformar kan läggas ner, likt Omegle som var en populär mötesplattform som förekom i många sexuallbrottmål mot barn, under dess tid av drift fram till november 2023.

Förutom vuxna i förtroendeställning till barn, bör även domän- och plattformsägare hållas ansvariga, särskilt för de beteenden som deras plattformar främjar. Exempelvis kan autentiseringskrav ställas mot användaren med tredje part certifikat för att exempelvis förhindra att en användare skapar flertal konton under olika pseudonymer, eller sparar nya konton när de tidigare spärrats och blockerats. Ökad kunskap om hur förövare utnyttjar funktioner på en plattform för att begå brott är det första och viktigaste steget för att förhindra dessa typer av internetbrott. Det är av synnerlig vikt på de domäner och plattformar vars riktade målgrupp är barn och ungdomar.

Förövare är starkt motiverade individer, och trots att de utgör en minoritet av användarna på diverse plattformar, måste deras förekomst tas i åtanke för att kunna begränsa

missbruk av något som är till för majoriteten. Begränsningar kan göras utan att för stor frihet för vanliga användare tas bort, men kan behöva göras i olika stora omfattningar beroende på vilken funktion som ska begränsas.

Förutom tidigare nämnda, är Snapkartan en annan funktion på Snapchat som också kan missbrukas. Detta komplicerar situationen ytterligare för barn på plattformen då förövare kan utnyttja funktionen för sina egna syften. Möjligheten för en förövare att få tillgång till Snapkartan och därmed kunna lokalisera sina offer, utgör en betydande fara, särskilt med tanke på hot och andra riskfyllda beteenden.

Sammanfattningsvis kan det konstateras att den plattform förövare använder mest är främst Snapchat, som förekom i alla brottmål samt var den plattform som nämndes först under intervjun. Sedan var Instagram väldigt populärt. Detta kan bero på egenskaper hos plattformarna som utgör en fara för integriteten och svårigheten att betydelsefull data lätt kan försvinna eller tas bort.

6.2. Fråga II

Främst är det smicker som utgör en stor del av konversationerna, med kommentarer som *“jävlar vad du är snygg”*, *“Söt tjej”* eller *“omg hur kan du ens ha så fin kropp om du e 10?”*. Dessa kommentarer används ofta för att vinna offrets förtroende och ge en känsla av bekräftelse till offret. Detta följs tätt av upprepade efterfrågningar om bilder och videor, med uppmaningar som *“kan du snälla skicka bild”*, *“visa dig med topp en gång ok”* eller *“skicka naket”*. Detta beteende kan beskrivas som tjatigt och psykiskt påfrestande, där förövaren försöker övertala offret att utföra handlingar de annars inte skulle göra.

Varför smicker förekommer, och blir så framgångsrikt i alla dessa fall kan bero på nyfikenheten hos barn och ungdomar, speciellt vid särskilt unga offer. Många av de offer som utsätts är under pubertetsutvecklingen, en period av ungdomens kroppsliga och psykiska utveckling mot vuxenstadiet. Ofta är det en tid av nyfikenhet och förvirring som är en naturlig process, men som utnyttjas av dessa förövare. Genom överflödigt och omedelbar bekräftelse, falska löften om förhållande eller utlåtande om kärlek och omsorg manipuleras offer att ge med sig, även om det går emot magkänsla att något är fel. Exempelvis *“Tänker på dig”*, *“jag har saknat dig baby, jag saknar bilder du skickade”* eller *“Saknar se dig”*.

Förövarna använder sig ofta av andra manipulationstekniker också. Bland annat uppgav de falska identiteter och ljög om sin ålder för att vinna förtroende hos brottsoffret vilket oftast ingick i moment av den initiala kontakten med offret. Vid flera instanser skickade förövaren egna önskade bilder till offret för att försöka motivera offret att skicka bilder på sig själva i gengäld. Motivation i form av monetär kompensation eller annan form som exempelvis berusningsmedel, förekom också, ofta i fall där offret var något äldre, eller hade nekat förövarns tidigare försök av smicker och tjat. Exempel på detta kan vara fraser som *“Om du camar ikväll får du 1500 när vi ses”*, *“fyller på din mob om jag får se lite”* eller *“400 avrunkning, 1000 avsugning, 2000 ligg”*.

Förövare använder ofta hot som en manipulativ taktik för att utpressa offret och tvinga dem att fortsätta med de handlingar som begärs. Dessa hot kan vara direkta, som att hota offret med våld om de inte följer förövarns instruktioner. Exempelvis kan hoten innefatta fraser som *“skicka bilder innan du är död”* eller *“visa dina bröst för mig innan jag skjuter dina föräldrar.”* Det förekom även hot om att sprida materialet som redan hade

tillhandahållits av förövaren vidare till andra i offrets närhet eller sociala kretsar. Vilket är ett utnyttjande av offrets naturliga rädsla för ensamhet och skada inom familjära sociala kretsar.

Vidare diskussioner om risk- och psykologiska faktorer som påverkar offers handlingar och utsatthet kommer dock inte att lyftas i denna studie, då detta faller utanför studiens omfattning. Hittills har tekniska aspekter diskuterat, såsom valet av plattformar, och endast kort berört förövarens motivationer, som språkbruk, då dessa kan variera kraftigt beroende på många faktorer. Kultur, språk, nationalitet och ålder, både hos förövare och offer, kan alla påverka språkbruket och temat i konversationerna.

6.3. Fråga III

Det är viktigt att förstå att endast enstaka ord och fraser inte är tillräckliga för att korrekt identifiera grooming; det är kontexten och sammanhanget där orden används som får betydelse. Att använda AI och naturlig språkbehandling för att klassificera konversationer som grooming är ett steg i rätt riktning. Likväl, precis som med allt som rör artificiell intelligens, tar det tid att uppnå hög noggrannhet. Språket förändras kontinuerligt, och de korpusar som används för att träna NLP-modeller behöver ständigt utökas och förbättras. I fall som grooming, där språkbruket varierar starkt från situation till situation och från språk till språk, är det alltid nödvändigt med större och mer aktuella datamängder. Under genomförandet av denna studie uppkom utmaningar relaterade till tillgången på moderna svenska korpusar. Detta belyser behovet av ytterligare forskning och tillgång till resurser inom detta område för att främja fortsatt utveckling och förbättringar. Dessutom hade skapandet av en egen modell varit ett betydligt mer tidskrävande moment om inga svenska förtränade modeller kunnat hittas.

Mängden data spelade även stor roll; idealt hade en större datamängd kompilerats för att träna Naïve Bayes modellen. Ett alternativ som övervägdes för att utöka datamängden var att använda befintliga datamängder från engelskspråkiga studier och sedan översätta dem till svenska. Den metoden valdes dock bort på grund av egen orimlighet: bland annat ytterligare tid för egen översättning, alternativt korrekturläsning och redigering vid automatisk översättning då vissa semantiska begrepp inte alltid har en exakt och likvärdig ekvivalent i andra språk.

Trots en mindre datamängd än som hade varit önskvärt, visade resultaten i föregående kapitel att det går att automatisera detektion av grooming genom användning av djupinlärning och naturlig språkbehandling, med relativt hög träffsäkerhet på cirka 84%. Denna siffra är lovande, men kan inte garantera sådan träffsäkerhet varje gång. Den kan variera beroende på datamängden som används. Exempelvis kan datamängden som användes för test av modellen vara lik den datamängden som modellen tränades på, detta påverkar självklart träffsäkerheten.

Genom att beräkna standardfelet för varje prestandamått och sedan tillämpa en signifikansnivå på 95% för att fastställa konfidensintervallet, kunde träffsäkerheten på modellen förväntas att ligga inom ett visst intervall 95% av fallen. Om intervallet är brett kan det betyda att prestandamåttet är osäkert, i denna studie låg bredden mellan 15%-20% som kan klassas som relativt brett. För att undvika detta och minska bredden på intervallet kan man använda större dataset vid test av modellen för att undvika osäkerheten.

Trots att 84% är ett förhållandevis högt värde, finns det rum för förbättring. Målet är att få ett värde så nära 100% som möjligt, men detta faller tillbaka i bristande tillgång till en

större datamängd. Det kan också bero på datamängden modellen tränades på, om det är felaktigheter i de fraserna påverkas träffsäkerheten. Dessutom användes fraser som inte i sig är skadliga, exempelvis *“vill du bli ihop”*. Sådana fraser kunde modellen reagera på som skadlig även när det gäller i sammanhang där meningen är oskadlig.

Detta eftersom Naïve Bayes använder sannolikhet för att klassificera data, vilket medför att vissa ord som inte är skadliga klassas som skadliga och vice versa. Modellen beräknar hur många gånger varje ord i ett “meddelande” förekommer i de två klasserna harmless och harmful, och gör sedan en samlad bedömning av sannolikheten att frasen tillhör en viss klass baserat på hur orden förekommit i träningsdatamängden.

Detta medförde en ökad risk för felklassificeringar, särskilt i fall där ord eller fraser, såsom "träffa mig", kan tolkas på olika sätt och därmed vara både skadliga och potentiellt oskadliga. Trots att fullständig tillgång till konversationer skulle ha ökat modellens känslighet och därmed offrat precisionen på oskadlig data var vår högsta prioritet att identifiera skadligt innehåll med hög Precision. Därför är det bättre att modellen är känslig och klassar “vad snygg du är” som skadlig, hellre än oskadlig. Därför skulle det ha underlättat om den inledande fasen av dessa konversationer hade varit tillgängliga.

I tabell 2 och tabell 3 redovisades fraser från testdatamängden som modellen felaktigt klassificerade. Bland de oskadliga fraserna som klassificerades felaktigt återfanns uttrycket "Kan du X?" flera gånger. Detta kan härledas till att frasen "Kan du" återfanns upprepade gånger i den datamängd som modellen tränades på. Som resultat reagerar modellen på dylika fraser och misstolkar dem som tecken på grooming. Liknande mönster observerades även för fraser som innehöll "Ska vi X?" och "Vill du X?". Exempelvis förekom orden “vill” 41 gånger i den skadliga datamängden medan den oskadliga datamängden endast innehöll “vill” 10 gånger. “Du” förekom 102 gånger i den skadliga träningsdatamängden och hälften så många gånger i den oskadliga datamängden.

I tabell 2 fanns det utdrag av den skadliga testdatamängden som klassades som ofarliga, trots att det för blotta ögat var uppenbart att de innehöll grooming. Dessa felklassificeringarna kan väcka frågor om tillförlitligheten att modellen kan klassa rätt. Och även om modellen har lyckats klassificera de flesta fraserna korrekt är det oroväckande att så uppenbara fraser som återfinns i tabellen undgår filtret. Trots att en fullständig förklaring till varför dessa fraser inte klassades som farliga kan ges, valdes dessa fraser för att göra en närmare undersökning. Detta för att undersöka hur och varför detta kunde ske, undersöktes alternativa meddelanden, likt de felaktiga fraserna fast med viss variation, för att utreda om vissa grammatiska variationer påverkar klassifikationen. Exempel på dessa variationer presenterades i figurerna 4-6 och hämtades ursprungligen från tabell 2.

I figur 4 presenterades hur frasen *“vill ha rövbilder och tung bilder/videos”* när den delades upp, påverkade klassningen. *“Vill ha rövbilder”* blev klassad som ofarlig, detsamma gällde för *“vill ha tung bilder/videos”*. Däremot klassificerades *“vill ha tungbilder”* och *“vill ha bilder”* som skadliga. Fraserna *“vill ha”* och *“bilder”* är ord som modellen nästan överpresterar på i vanliga fall då de orden förekommer så oerhört mycket i den skadliga datamängden. Däremot hanterar algoritmen varje ord för sig, vilket betyder att *“tung bilder/videos”* klassas som två ord, *“tung”* och *“bilder/videos”*. Och när det gällde orden *“rövbild”*, *“tung”* och *“bilder/videos”* förekom inte alls i någon av datamängderna.

Även den väldigt tydliga frasen *“Tryck in pennan i fittan”* i figur 5 från tabell 2 klassades som ofarlig, men när den stora bokstaven togs bort klassificerades frasen korrekt. Detta kan bero på att *“Tryck”* inte förekom i någon av datamängderna men *“tryck”* med liten bokstav förekom däremot i den skadliga mängden. Dessutom fanns inte ordet *“pennan”* i någon av datamängderna. Liknande mönster observerades för frasen i figur 6, när skiljetecken togs bort och fraserna klassificerades separat var alla klassificerade som skadliga. Denna trend kan sannolikt härledas till att förövare ofta använt ett språkbruk som är felstavat, ofärdigt och inte följer korrekt meningsbyggnad. Dessa avvikelser från konventionell språkanvändning utmanar modellens förmåga att korrekt klassificera skadligt innehåll, och understryker vikten av att ytterligare utforska och anpassa vår modells förmåga att hantera sådana språkliga variationer.

Varför dessa uppenbara avvikelser har klassificerats inkorrekt kan ha berott på brist i variation av den skadliga datamängden, som grammatisk variation och avvikande språkbruk i testdatamängden. En potentiell förändring för att öka Accuracy är att träna om modellen på större datamängder, för både *“Harmless”* och synnerligen *“Harmful”*, och därmed hantera problemet att modellen missar uppenbara skadliga fraser endast för att de var med skiljetecken, stor bokstav eller sårskrivna.

Detta problem kan också hanteras genom att texten förbehandlas så att alla ord konverteras till gemener och alla skiljetecken tas bort. En sådan förbehandling skulle kunna minska antalet felklassificeringar, men det medför även att en stor del av identifieringsförmågan försvinner eftersom förövarnas typiska språkbruk, såsom felstavningar, små bokstäver och avsaknad av skiljetecken, försvinner. Om all text normaliseras och felskrivningar rättas, riskerar modellen att förlora sin förmåga att korrekt klassificera verkliga fall.

Detta utgör en betydande begränsning för den skapade modellen, särskilt när enklare algoritmer används. En möjlig lösning på detta problem skulle kunna vara att implementera regler där en text klassificeras som skadlig om den innehåller specifika ord, exempelvis *“naken”*, oavsett kontext. Transformer-modeller har exempelvis större potential eftersom de kan förstå samband och kontext bättre än enklare algoritmer. Modellen skulle dessutom förbättras med en större datamängd, vilket skulle öka sannolikheten för att ovanligare ord förekommer i träningsdatamängden och därmed förbättra modellens noggrannhet och prestanda.

Trots brist på tillgång till hela konversationer mellan förövare och offer, kunde modellen identifiera skadliga fraser relativt bra. Detta begränsade dock omfånget i variation av skadlig data, då harmful data även innehåller inslag av oskadligt innehåll. Sällan inleds konversationer omedelbart med konversation som tyder på grooming, eller endast består av sådant material.

Valet av Naïve Bayes motiverades främst på grund av dess enkelhet att implementera, särskilt med tanke på en begränsad kunskap inom maskininlärning. Dock medför denna algoritmen vissa begränsningar som kan vara en felkälla. Exempelvis behandlar Naïve Bayes varje fras som en separat funktion, så om något ord är felstavat kan det påverka klassningen. Detta är ett problem då förövare kan kringgå detta genom att medvetet stava fel på skadliga ord.

Modellen utvecklades med målet att underlätta för brottsutredare att identifiera grooming efter misstänkt utfört brott som del av brottsutredningen. På grund av viss lägre känslighet, vilket tillåter flest True positive men även många falskt positiva utslag för grooming, är fokuset vid redan utförda brott att mer bevismaterial samlas och icke relevant material bortses. Vid ett alternativt syfte som realtidsbearbetning skulle en hög känslighet inte vara lika lämplig då flest False positive tillåts; i realtidsscenario kan falskt positiva utslag göra motsatt nytta, då konsekvensen är att programmet deaktiveras på grund av för många falska varningar.

7. Etiska aspekter

Känsligt material rörande barns integritet kan vara svårt att utreda i vetenskapliga studier, då frågor om sekretess måste lyftas. Materialet som analyserats under resultatavsnittet, var offentliga domslut som är tillgängliga för allmänheten via tingsrätternas målen prövats på. I dessa domslut är alla individer med sekretessbelagda uppgifter redan anonymiserade av domstolen så inga fler åtgärder har tagits av författarna på sekretess aspekten.

Andra etiska aspekter att ha i åtanke är de potentiellt negativa effekter som kan medföras vid felaktig användning av klassifieringmodellen. Om modellen publicerats open-source, kan förövare använda modellen som ett filter för sina konversationer för att se om modellen ger avslag på grooming, och sedan avsiktligt felstava eller formulera om innehållet för att försöka undgå framtida detektering.

Förhoppningen är att då denna studien endast producerat en prototyp av ett detekteringsverktyg och inte är optimerad nog för att arbeta utanför simuleringsmiljö, kommer missbruk av denna iterationen inte försvåra eller förhindra en utredning. Om studien däremot skrivits, exempelvis i samarbete med polisen eller annan utredande myndighet, hade kravet på sekretess varit högre då forensiska verktyg används regelbundet och är avgörande för bevisföring av en utredning. Om ett potentiellt verktyg som detekterar grooming som är ämnat åt brottsutredande verksamhet blir känt för förövare, och kan kringgås, blir verktyget förlegat och meningslöst.

Författarna lägger stor vikt på att osäkerhet råder kring de potentiella användningsområdena då liknande verktyg i bruk av brottsutredande verksamheter är få eller okända.

En av de främsta motivationerna till att automatisera detektion och analys av känsligt material som grooming eller dokumenterat övergreppsmaterial, är för att bespara det psykiska måendet hos utredare, samt effektivisera arbetet av den överväldigande mängden

material. Detta material är inte alltid förutsägbart och som utredare kan man bli utsatt för väldigt påfrestande material, så genom att processen automatiseras kan materialet som klassas som skadligt av modellen reserveras för specifika utredare. Grooming förekommer vid utredningen av andra typer av brott och därmed kan “vanliga” utredare utsättas för detta av misstag. Polismyndigheten erkänner själva att tiden för att granska dessa typer av brott inte alltid finns till, samt avsätts inte tillräckligt mycket IT-stöd som behövs för utredningsarbetet inom ISÖB samt att det skulle underlätta med ett verktyg som automatiserar denna process. (Riksrevisionen, 2021)

8. Slutsats

Denna studie visar att modellen har möjlighet att agera som ett första-steps filtreringsverktyg som kan detektera grooming. Algoritmen som användes i detta arbete var Naïve Bayes tillsammans med en förtränad transformer-modell, GPT-sw3-356m från AI Sweden för att underlätta ordinbäddning. Sedan tränades, validerades och testades Naïve Bayes modellen på en datamängd, kategoriserat efter fraser klassade som skadliga eller oskadliga.

Med en signifikansnivå på 95% fick detektionsmodellen följande intervall: Accuracy 76.81%-91.19%, Precision:72.41%- 92.97%, Recall: 76.38%-95.62% samt F1-score: 75.03%-93.59%. Dessa värden pekade på relativt hög trovärdighet och förmåga att detektera när grooming förekommer i konversationer. Det är även konstaterat att denna modell skapades med syftet att underlätta i polisutredningar under forensisk analys, inte i brottsförebyggande syfte.

Med syftet i åtanke, fungerar modellen väl, trots en mindre datamängd; dock bör en större datamängd alltid eftersträvas vid maskininlärning och naturlig språkbehandling för att öka träffsäkerheten.

Litteratursökning och intervju var de vetenskapliga metoder som valdes för att kompilera datamängden som användes för träning, validering och test av modellen, samt besvarande av frågeställning II. Denna studie och liknande arbeten hade gynnats stort av ett samarbete med utredande myndigheter som polisen för att få tillgång till resurser som bevismaterial, i form av skadliga fraser, vilket la grunden för studiens experiment.

De plattformar som förövare använder sig av mest, som besvarande frågeställning I, utreddes genom litteratursökningar och intervju. Snapchat var den populäraste plattformen och förekom i samtliga dokumenterade fall. Instagram kom på andra plats, med en markant marginal till Snapchat. Därefter följde TikTok och Kik, som förekom lika ofta och delade tredjeplatsen.

De fraser som besvarande frågeställning II, tillhörande "Harmful" datamängden, innehöll bland annat smickrande kommentarer, oftast i förhållande till offrets kropp eller ansikte. Tillsägelser och förfrågningar om bilder eller videor på offrets kropp eller kläder; mutor om pengar eller annat av värde om offret nekat initiala förfrågningar. Detta ofta följt av fler förfrågningar om nya bilder med särskilda instruktioner; det förekom många gånger hot om våld eller utpressning genom spridning av det delade materialet. Språket uppvisade mönster av upprepning och beskrevs som tjtande samt psykiskt påfrestande.

Trots de tre metodernas sammanlagda relevans och validitet, kan resultatet komma att förändras vid framtida upprepade försök, då äldre brottmål förlorar relevans mot nyare mål, och innehållet av dessa vad gäller frågeställning I. Potentiella förändringar tyder inte nödvändigtvis på motbevisning av resultat, snarare en vidareutveckling.

9. Framtida forskningsinriktningar

Eftersom detta arbete är inriktat på att detektera grooming i förhandsutvunnet material, särskilt i icke-strömmande data, skulle det vara intressant att utforska och utveckla realtidsdetektion i framtiden, kanske i form av ett lokalt webbläsartillägg. Ett sådant projekt skulle kunna utvecklas med liknande metoder genom att sammanställa en ordlista och sedan integrera den med tredjepartsprogram för att fungera i bakgrunden av chattar i realtid. En utmaning vid utvecklingen av ett sådant projekt är dock frågan om kompatibilitet och tillgänglighet av data med olika valda plattformar, vilket har identifierats under loppet av denna studie.

Ett annat utmärkt framtida forskningsområde skulle kunna vara ett liknande arbete fast i samarbete med polisen. Genom att samarbeta med polisen skulle forskare få tillgång till ett bredare och mer varierat material, vilket skulle kunna förbättra och förfina detekteraren. I samband med detta kan möjligheten att göra en liknande detekterare gällande bildgranskning undersökas. Genomgången av barnpornografimaterial är både tidskrävande och mentalt påfrestande för utredare. Att automatisera den processen av bildigenkänning kan utredares arbetsbörda lättas rejält då automatiserade system snabbt och effektivt kan utföra den första filtreringen.

Slutligen, även om detta arbete inte kunde ge en fullständig förklaring till varför vissa uppenbara fall undgick modellens filter, var det tydligt att språkliga variationer och subtila nyanser spelade en viktig roll. För att handskas med detta kan framtida forskning fokusera på att förbättra modellens förmåga att hantera sådana variationer, möjligtvis genom att använda mer avancerade transformer-modeller som kan förstå samband och kontext bättre än enklare algoritmer.

Referenser

- Adekolu, A. M., & Chen, L. (2019). *A Simple Approach to Develop a Parental Control via a Monitoring Home Network*. <https://urn.kb.se/resolve?urn=urn:nbn:se:hh:diva-38994>
- Ali, S., Haykal, H. A., & Youssef, E. Y. M. (2023). Child Sexual Abuse and the Internet—A Systematic Review. *Human Arenas*, 6(2), 404–421.
<https://doi.org/10.1007/s42087-021-00228-9>
- Al-jabery, K. K., Obafemi-Ajayi, T., Olbricht, G. R., & Wunsch II, D. C. (2020, January 1). 2 - *Data preprocessing* (K. K. Al-jabery, T. Obafemi-Ajayi, G. R. Olbricht, & D. C. Wunsch II, Eds.). ScienceDirect; Academic Press.
<https://www.sciencedirect.com/science/article/abs/pii/B9780128144824000024>
- Anderson, P., Zuo, Z., Yang, L., & Qu, Y. (2019). An Intelligent Online Grooming Detection System Using AI Technologies. *2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 1–6. <https://doi.org/10.1109/FUZZ-IEEE.2019.8858973>
- Bobbitt, Z. (2021, April 14). *Standard Error of the Proportion: Formula & Example*. Statology. <https://www.statology.org/standard-error-of-proportion/>
- Dalen, M. (2008). *Intervju som metod*. Gleerups utbildning. ISBN: 978-91-40-65247-8
- Kamar, E., Maimon, D., Weisburd, D., & Shabat, D. (2022). Parental guardianship and online sexual grooming of teenagers: A honeypot experiment. *Computers in Human Behavior*, 137, 107386. <https://doi.org/10.1016/j.chb.2022.107386>
- Kulkarni, A., Chong, D., & Batarseh, F. A. (2020). 5—*Foundations of data imbalance and solutions for a data democracy in Data Democracy*. In F. A. Batarseh & R. Yang (Eds.), (pp. 83–106). Academic Press.
<https://doi.org/10.1016/B978-0-12-818366-3.00005-8>
- Kumar, N., Sonowal, S., & Nishant. (2020). Email Spam Detection Using Machine Learning

- Algorithms. *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*, 108–113.
<https://doi.org/10.1109/ICIRCA48905.2020.9183098>
- Lykousas, N., & Patsakis, C. (2021). Large-scale analysis of grooming in modern social networks. *Expert Systems with Applications*, *176*, 114808.
<https://doi.org/10.1016/j.eswa.2021.114808>
- M. Ashcroft, L. Kaati and M. Meyer, "A Step Towards Detecting Online Grooming -- Identifying Adults Pretending to be Children," 2015 European Intelligence and Security Informatics Conference, Manchester, UK, 2015, pp. 98-104, doi: 10.1109/EISIC.2015.41.
- Nikolovska, M. (2020). The Internet as a creator of a criminal mind and child vulnerabilities in the cyber grooming of children. *JYU Dissertations*.
<https://jyx.jyu.fi/handle/123456789/67584>
- Riksrevisionen. (2021). *Internetrelaterade sexuella övergrepp mot barn – stora utmaningar för polis och åklagare*. Riksdagens Interntryckeri.
<https://www.riksrevisionen.se/rapporter/granskningsrapporter/2021/internetrelaterade-sexuella-overgrepp-mot-barn---stora-utmaningar-for-polis-och-aklagare.html>
- Vujovic, Ž. Đ. (2021). *Classification Model Evaluation Metrics*. International Journal of Advanced Computer Science and Applications, *12*(6).
<https://doi.org/10.14569/ijacsa.2021.0120670>
- Wohlwend, B. (2023, July 16). *Machine Learning — Evaluating classification models*. Medium.
<https://medium.com/@brandon93.w/machine-learning-evaluating-classification-models-18713af3d764>
- Wooldridge, J. M. (2023). *What is a standard error? (And how should we compute it?)*.

Journal of Econometrics, 237(2), 105517–105517. ScienceDirect.

<https://doi.org/10.1016/j.jeconom.2023.105517>

Yang, F.-J. (2018, December 1). *An Implementation of Naïve Bayes Classifier*. IEEE Xplore.

<https://doi.org/10.1109/CSCI46756.2018.00065>

Appendix 1: Intervjufrågor

1. Kan du börja med att berätta lite om dig själv?
 - a. Arbetsroll; Längd av erfarenhet inom din kompetens; Några exempel på arbetsuppgifter.
 - b. (Utbildning; Tidigare relevant erfarenhet.)
2. Brb Kap 6 §10a definierar “grooming” i lagboken under “kontakt med barn för sexuellt syfte”. Anser du att dagens beskrivning helt omfattar hur verkligheten av dessa brott kan se ut? Finns det ett behov att omdefiniera/uppdatera groominglagen?
3. Hur stor del av de fallen du arbetar med förekommer det grooming? (antingen endast, eller i samband med andra brott)
4. Vilka typer av kommunikationsmedium förekommer främst i fall av grooming
 - a. Chattmeddelande, röstmeddelande/samtal, videosamtal, live-video/chatt
 - b. Varför används dessa mest?
5. Vilka plattformar/sociala medier förekommer oftast där chatt meddelande är det huvudsakliga kommunikationsmedium?
6. Hur beter sig förövare vid första kontakt med ett offer?
 - a. Påstår de sig vara offrets ålder?
 - b. Hur brukar det inledas?
7. Vilka ord eller fraser används ofta? Finns det några ord/fraser som indikerar att detta kan vara en vuxen?
8. Finns det några teman eller samtalsämnen som förövare använder för att skapa bekvämlighet med offret?
 - a. t.ex. Skola, kläder, hobbyn.
 - b. Rollspel
9. Finns det särskilda ord eller fraser som kan tyda på att förövaren har ett sexuellt intresse av offret?
10. Finns det särskilda ord eller fraser som förekommer som tydligt markerar att grooming sker?
11. Vad tror du är den största utmaningen gällande grooming?
12. Använder ni, i verksamheten, verktyg för att detektera grooming?
 - a. Vilket typ av verktyg?

- b. Ex. extraheringsprogram, osint verktyg, andra metoder som hemliga tvångsmedel.
13. Finns det något som du anser fattas/som hindrar bekämpningen av grooming?
 14. Vad kan individer själva göra för att minska risken att utsattas för grooming.
 15. Vad kan vårdnadshavare göra för att skydda sina barn på nätet?
 16. Vad är det mest givande i din roll? Vad är det som gör ditt arbete värt det?
 17. Finns det något mer som du skulle vilja tillägga eller påpeka som du känner att vi missade att fråga?

Appendix 2: Skadlig ordlista

Dataset training

Gör bara vad du känner dig bekväm med och Vad har du för Swish nr?
visa din
kan du stoppa in något
typ tandborste eller sminkborste
pulla
kör snabbt å hårt
sen sug på ditt finger
pulla
kan du de?
Alltså videochatta och visa ansiktet och pulla dig medan tills du kommer i orgasm
pulla
Ditt först
Och fitta
pulla
visa lite då
mmm ta av
naket
skicka naket
nice runkbild
rövbild
vilken annan go bild får jag då
hej är du också 13
vill du se min snippa
visar om jag får se snopp
runka
runka mer
om jag får 5 önskingar får du 10
kan du stoppa in ett finger i rumpan
kan du stoppa in en tandborste
pinne eller något
gör mig kåt
runka
naken
vill du se min snippa

visa snopp så får du
spruta för i mig
du gör mig så våt
spruta för i mig
spruta så får du min snap
bli ihop
få se en gång hur du ser ut då. så man vet att de e du
så att man vet att du är tjej
ta en likadan och gör tumme upp me handen
ta en bit ifrån så att hela du syns
visa en gång hur smal du e
visa frammifrån
svårt o se me den tröjan. magtröja då?
gör tumme upp me handen så man ser att de va nu
är du lång? smal?
få se en bit ifrån
svårt o se så. kan du inte ha en magtröja ist då
visa dig med topp en gång ok
fast både frammifrån och från sidan då
men gör en kul saak också då – stå på alla fyra som en hund i sängen HAHA de ser roligt ut
gör en rolig grej då – ha den vita bikinin och stå på alla fyra
svårt att se i den där tröjan, kan du inte ha magtröja istället
gå in på den där videosidan så man kan se dig igen när man chattar
få se hur du ser ut en gång bara

trycka in en sminkborste
Mm smek lite fitta
blir stolt över dig
Nästa gång ska du trycka in o knulla dig jätte fort
Hade vart nice se dig
sexig video till mig privat
blir kåt när ja ser snygg tjej med glasögon
få se en gång hur du ser ut då. så man vet att de e du
ta en likadan och gör tumme upp me handen
ta en bit ifrån så att hela du syns
visa en gång hur smal du e
visa frammifrån

Vill se dig sexiga kora shorts
fan saknat dig
svårt o se me den tröjan. magtröja då?
Blev jätte kåt när ja såg din video
Saknar se dig
Tänker på dig
testa dig fram hur man hånglar med en vägg
rulla ut med tungan, låsas de min fitta
Ta en sexig bild på dig
gö video vil se om man ser genom
gör tumme upp me handen så man ser att de va nu
är du lång?smal?
få se en bit ifrån
svårt o se så. kan du inte ha en magtröja ist då
Ta trosan vid sidan
vill slicka din tajta fitta
kan kolla på den nu
också gjort några videos så får ni en se några knullfilmer sen
Okej visa en här
Skicka videon
visa dig med topp en gång ok
fast både frammifrån och från sidan då
så ska ni få se knullfilmer
Filma med vanliga kameran då och sen spelar du den och skickar chatten här
Jättebra den blev superbra va duktiga ni är
Nu tycker ju vi att det är din tur att göra en sån video oxå
testa att smeka klitoris också
kanske också sära på fittläpparna så ser vi om du är lika våt i snippan
Tror du är minst lika bra
Okej vi visar första delen av en knullfilm
Kolla nu och njut samtidigt som ni smeker era snippor
du är helt plaskvåt i fittan det nästan rinner ur den wow bra jobbat
Nu blir det mera knull gillar ni hans stora snopp?
Måste vi smeka våra fittor?
Vill ni se mer knullfilm?
Ja det är det som är själva grejen när man ska lära sig det här med onani men vad skulle ni vilja göra då?
Först börjar man med själva smekandet och speciellt på och runt och vid själva klitoris för det är så man blir riktigt våt innan man kan börja att onanera på riktigt

Nej det var det inte men ni kan få se en video kort video på mig när jag kör med tandborsten?
Först körde jag i rumpan sen i fittan
Nej det är superskönt
testa med ett finger i snippan så långt in som det går
Sen rör du fingret in och ut så snabbt du kan
sen kan du smeka klitoris samtidigt
För då kanske du får en orgasm och sprutar ut fittsaft
Glöm inte att suga på fingret så det blir fuktigt och blött innan du stoppar in det i snippan
vågar slicka varandras fittor
Jag var våt oxå
Längst ner på snippan precis innan rumphålet
På andra sidan av snippan i botten inte toppen där klitoris är
Ju längre video ni gör desto våtare blir hon och då är det skönare och bättre
Är ni våta hahahahshshhaha
testa att stoppa in ett finger så långt in du kan och sen köra in och ut med det då blir du supervåt och kan få en orgasm
är jävligt kåt
Så att filma nu är inga problem alls och du gjorde så bra från början också
Så du ska vara stolt för at du lär dig så snabbt
Vi gör inte det för att vara taskiga tänk på att vi själva började med att onanera när vi var 6 och 8 år så vi vet hur det är
ska jag filma eller?
Jättebra ni är ju hur duktiga som helst
försök att nu göra varsin längre videon där ni kör så snabbt ni kan ut och in med fingret och blöt ner det innan ni stoppar in det
var inte så fäg...
Kom på en idé men kan du lova mig att radera den här chatten helt med mig här
jo, kom igen...
vill se alla tre
Vill du va vän eller inte? Blockar dig nu för ska ändå göra ett nytt konto så du vet. Du borde berätta för din mamma vad du gjort för videos och hur många killar o tjejer du visat din snippa för och när du pullat o sånt innan nån du känner i din familj får se dom filmerna du gjort och bilderna med
Det ja tänkte var ifall ni kunde göra typ bara tiktoks tillsammans typ i trosor eller så för om du säger till henne att du gör såna eller visar henne nån du gjort till mig så hon fattar att du gör såna till en kille kanske hon med vill va med om du säger att jag gärna skulle vilja det
stå upp närmre
alla visar bröst på snap
filma lite nu
skicka dina nudes

För ja tror lätt hon skulle vilja va med o göra såna med om du bara frågade och visade henne hur du brukar göra
Okej men vad är det som gör att du inte vill fråga för hon var ju med naken i en video ju och det kändes ju som hon gillade det
jag blir kåt när jag tar kontroll, när jag får en bitch att göra ngt hon it vill
hur fan kan du ha såna bröst näru e 09a
Du ska ta av allt, Även strumpor Ska av ALLT SKA av
smeka fittan
bli ordentligt kåt
omg so sexiiiiig
meeeeer
För ja tror hon kan dansa lika sexigt som dig typ
men du måste fixa en utan bh juuuuu
orättvist annars ju hahah
Gör som du vill men kan du göra fler dansvideos idag tror du?
Skulle gärna vilja se när du gör i alla olika trosor du har så kan ja säga vilka som är sexigast sen
kläm på tuttar
kom ska jag värma dig
kom hit istället
Eller om du kan göra i bara strumpbyxor utan trosor fan de hade varit så najs att se för älskar såna
Gör när du kan innan det är dags att sova
din body är asnice, bröst utan bh eller helkroppsbild
Gör dom så porriga du kan och så nära kameran du kan typ som du modellerar och ska visa upp det sexigaste mer Varje trosa eller strumpbyxa kommer blir hur bra som helst
Två snyggingar på väg till träning
Naaaw va gulliga ni är
hehe niice, lite närmare!
på boobsen asså hehe
vill se alla tre
Kör hårt på träningen och var rädda om rumporna nu
din body är fan goals
nu är det din tur att skicka på din
visa en privat del på dig själv eller någon annan
smek dig hårdare
tryck och gör det skönt för dig
smek dig utanpå
stå upp närmre
ta in handen

Du e ju ba 14
bokar isåfall direkt
kan du gapa stort när du är naken älskling
Jag ska komma och ta dig och våldta dig
Behöver ju veta så vi får hotell
Vill du träffa mig kanske
Ska vi träffas
ring om du behöver när som god natt finaste
Kåt på dig
Mår skit och funderar på att försöka en annan sak igen så alla slipper mig
Hur lång är du
är du kåt
Skicka bild på hela din kropp
ligga för 10 000 kr?
Vilka typer av killar gillar du?
Har du stora bröst?
posa
Hur många har du legat med?
se kuk
Vill du se något stort
Ja är kåt på dig
skitsnygg och en jävla massa
sexig
snygga bröst
fina bröst
fina grejer
ta av dig upptill då
fyller på din mob om jag får se lite
är så kåt
får jag se ditt ansikte
är du rakad
blir mera tänd då
opps så snygg
gillar kramas djupa kyssar, smeka bröstet slicka mm
älskar rumpor och bröst
om man får se hihi
nice 10 poängare

fan vad du är snygg
på med cam och testa lite kläder
jävlar vad du är snygg
är du oskuld
är du nyfiken på att förlora oskulden
tagit en oskuld för 6 år sen
älskar att slicka en tjej på ett speciellt sätt
det är tråkigt att runka, roligare att få hjälp
pulla
tänder lite på det
älskar att slicka också
Ligger i sängen med världens stånd. Vill du se?
fick jag bild på din trånga fiffin
nice urringning
tänder på att ge cash för sex
aja får köpa mig en brud tills jag flyttar hem, säkert någon som vill ha ett bra betalt sommar jobb en timma i månaden
behöver spruta
är du min slampa på onsdagkväll snygging
då får jag runka istället
vågar du testa min gubbkuk
sex och hångel menar jag
har just runkat ligger i sängen
vad gillar du för hångel
så busig så du kanske suger av mig
har en fetisch och tänder på att strula med tjejer och ge cash, vet du någon tjej som vill ses och testa?
Dessutom är det lättförtjänta cash. 3000 på en timma
Tänder mer på bilsex, cash och sånt
Jag älskar att slicka min specialare. Är bäst på det
400 avrunkning, 1000 avsugning, 2000 ligg
Har du haft sex i bilen ngn?
du får välja att jag ser på din bild och gör det, eller att du ser mig på Skype när jag gör det
avsugen av en 03a
slicka först jag blir sugen då
sex eller bj

fyllekåt
vill spruta i cam
ska runka
ville bara spruta
är ni sugna
blir myssjuk av cider
är du 06a
vill lära dig något busigt
vill göra en grej
vad kostar det att knulla dig
se mig runka nu det är en order
hej vill ni se mig spruta
jag vill spruta för dig hehe
se mig dra lite skinn
är du virgin
har du sett en kille spruta i cam någon gång
med händer, rep eller bojor
får man se din rumpa
har du sett killar leka i cam
vågar du se en äldre spruta
älskar att slicka
vill se när du pullar
kan få pengar om du ser mig leka lite också
kååt
mm nice
du skall göra mig kåt så jag kan runka för dig sen efter fortsätter du att strippa
stoppa in något och låt det gå för dig
smek dig till orgasm
så kåt nu kan inte tänka
fan behöver spruta
Fan du gör mig kåt ju
sprutar sjukt mycket det är ett problem
en kåt kille
vill ni se mig runka
strippa då

sprängfylld
vill lätta på trycket
vill spruta för dig innan du dör
full och kåt
se mig runka så vill du leva
inte min grej, vill ha ett vanligt jävla hårt knull och slicka fittan till orgasm
fegis
lögnare
svara för fan då
har kollat på dina bilder och inte svårt att få 18 cm hårt
max vill se underkläder
ska jag svara ärligt
ångrar att jag har tigha byxor
runkar och kollar på dina bilder
jag ser hela din sexiga kropp
Lust o hjälpa mig?
Kommer du ihåg när du visade o fick se min?
Du kan få nått för det om du visar
Aldrig sett kuk innan?
Aldrig smekt dig själv heller?
Blev kåt när jag såg dig
Vill du inte visa alls på snap?
Vill se dig i bh
Jag fyller 16. Gillar du äldre?
Men du är ju bara 13
Vill gärna prova med dig
Vill du ha nått för det eller så
Veta om du skulle klara min
Räcker att du visar underkläder när du ligger ner
Ju mer du visar desto mer saker eller för mer pengar
Visa vad du har på dig
Om jag visar min kuk. Gillar du den visar du oxå o hjälper mig spruta?
Men vill se mer av dig typ magen eller bh och trosor
Söt tjej
Men du vet nog att jag runkar

Kan visa min kuk för dig och du skulle kunna få nått
vad säger du om att visa för varandra ikväll
tänkte ner dug i underkläder
vill inte se ner än det typ
runkar och sprutar för dig
Vill du se när jag runkar samtidigt som du visar dig i bh o trosor
trångt och gör ont efter jag såg dig
18cm i redan tighta jeans
kåt?
en snäll tjej som är oskyldig eller en lydig snuskis
Så kanske du kan hjälpa mig spruta?
Om jag visart vad som finns innanför? Gillar du det så visar du lite av dig och hjälper mig
låt mig få slita av dina trosor och knulla dig
I helt naken isf så man ser bröst och fitta
Snälla visa magen
Bh?
Tänkte om du ville se min...o jag får se ditt
Hej sötis vrf blockade du kik på snap
Men jah har 18 cm som gillar din kropp
Vill runka till dig som jag gjorde för din syster
Vill du tjäna enkla pengar idag?
Genom o runka och kanske suga av mig?
Om du camar ikväll får du 1500 när vi ses
Vill du tjäna 2000kr
vill du se mig runka samtidigt?
vad tycker du om min kuk?
Kåt? Fortsätt nu då
Pulla för mig, med fingrar eller borste
Kåt flicka?
Har du nån borste du kan använda också
Pulla med 2-3 fingrar djupt och viska att du är en snäll flicka
Säkert ännu mer när du är 14
Bra Oh 14?
In på skoltoan o ta bild på outfit och även trosorna fram o bak
Brösten o fittan

Ok för tänkte att om du visade lite o jag får spruta att du skulle få presentkort på typ hm eller Gina för 1000kr

18 cm rakt upp

Ok ska göra en fejkprofil ned dina bilder o skriva du är en hora hejdå

Är du busig?

På ett sexuellt sätt?

Jag älskar ha sex med yngre

Nice o se en 15-åring med min mellan läpparna Du hade sett bra ut med det oxå

Prova? Den är grov

Låt mig komma då? Vila kan du göra ikväll och med mig? Dessutom får vi väl det överstökad eftersom du inte vill fler gånger sen

Tar kontonummer när vi träffas så är 10.000kr inne senast tisdag

Kan vi inte ses direkt då? Helst om vi kan vara hos dig o jag för knulla dig annars utomhus men jag vet inga ställen

få se helbild så man ser hur lång du e

är du smal?

visa från sidan en gång

du e snygg

svårt o se me den tröjan, magtröja då?

Ok men visa trosorna en gång så lovar jag att ta bort då. Ok?

Jag skriver vad du ska skicka så ska du skicka det till mig oavsett vad det är

klä av dig

skicka bilder på din kropp

skicka bilder med borsten i

jävligt sexig i tajts

omg hur kan du ens ha så fin kropp om du e 10?

filma din rumpa igen för du hade asnajs kropp

vill du visa igen me ba trosor och dra in trosorna lite i rumpan innan jag går?

vad tar du för en avsugning, eller förresten jag ska ha en

din jävla slyna kommer fan våldta dig lyssna på mig

visa din fitta nu innan jag krossar din ruta på din bil

visa dina bröst för mig innan jag skjuter dina föräldrar

skicka bilder innan du är död

Så jävla kul när man har dig som uppskattar mer än missionären

anser att vi har ett bra sexliv nu
längtar tills vi har sex
kan du skicka bild
kan du skicka bild på dina bröst
kan du snälla skicka bild
nej men snälla skicka en till bild den var jättefin
får jag en bild
nej kan du inte bara skicka bild snälla
ta mobilen som om du ska ta en bild men istället filma när du sätter in dina fingrar i snippan och sen tar ut det och in och ut om igen
gör det med en sminkborste då
Tryck in pennan i fittan
vill spruta min sperma i eran mun
ta på din fitta igenom
vill du att jag ska spruta min pedofil sperma i din lilla 12 åriga fitta bby?
sen tänder jag på att se en tjejs röv, bröst, fitta
vill runka till din sexiga kropp
din kropp bby
vi kommer knulla som fan
du får 500 kr nu, Ta av alla dina kläder, Så sprutar jag fortare
rätt go kropp asså
skit sexig kropp asså
hade hellre sett utan kläder dock
jätte kåt
vill ha rövbilder och tungbilder/videos
vill knulla sönder din lilla 13 åriga kropp
pulla
pulla dig själv
vad vill du ha för ligg eller avsugning
Men du kan suga av mig?
Vill att ni suger av mig
jag såg din insta du är svinsöt vill du skicka några bilder till mig
jag har saknat dig baby, jag saknar bilder du skickade
skicka mig bilder annars skickar jag vidare
vill känna dina läppar mot mina

får jag fråga hur mycket du runkar
snygging

Appendix 3: Oskadlig ordlista

sluta försök vara cool
drömutsende
varför får hon ha samma kläder som du har
förlåt det blev så för vi är bffs
men aja vi pratar sen
men det bara hände det var inte meningen
oompa kan inte svenska
tydligen it du hellee
jo han kan men inte du
tvärt om
men asså seriöst det finns något som heter missclick
men asså seriöst asså
jävla svene
sluta va så taskig
ingen bryr sig
bryr mig inte eftersom jag inte kan hjälpa dig
men snälla wtf
jävla skill issue
det kommer att bli bra
stackars dig
näe gud stakarn
wtf pick me
jag håller med dig
har också haft en sån kompis
vilket datum tar du studenten
5 juni eller 9e minns ej
oki
dd
ja haha
ska du komma
venne vgn

jag äter chips i vardagsrummet
synd
mamma är ledsen för du ej kom
mannen jag kunde inte vafan ska jag göra
kom
vrf e du lack
ramlade i trappen också
på estelle
o nu har min axel gått ur led elr når
sen sov vi i någon soffå
o drog ner massa tavlor
sen kommer vakter mot oss o vi gömmer dom under soffan
beklagar sorgen
men kanske bara jag
visa respekt för i helvete
adolf hitler
ditt barn mst hetta hitler asså
din respektlösa skitunge
hoppas hon får det bra däruppe med din morfar

Mums roadkill med bröd
Nä fyfan. Jag var mätt redan, men nu vill jag ha mera mat.
Man blir ju sugen på att köpa matchande till hela familjen och dra och tälta
Vore oironiskt skitkul att komma till plugget en dag mes Lidl merch
Den tröjan är ju dope as fuck
Jag ska fan cop:a en Lidl jacka, skiten är hård
Lidls sneakers gick på 2lax i andraperson sist jag hörde om dem haha
Ny läanguage just dropped.
Han behöver ingen bra dator, han spelar roblox.
Jag älskar användning av tidningar för att göra memes
Har aldrig käkat Coops så jag röstar på ICAs, dom smiskar
låter som något en Peddo Janne vid coop hade kunat säga
Är det inte därför man äter bröd? För att få njuta av alla smörfyllda hål?
Jahaja.. vissa har pengar att flexa med ser jag.
Du är en sjuk jävel, vet du det?

Mentalsjukhuset. Ditt ska du...
Way to personal info att dela
Nä jag golar inte
Haha aa de haru
Brorsan har aldrig åkt nedförsbacke i en kundvagn
Fett tragiskt att vagnen trillat och slagit sig, hoppas den klarar sig
Man får inte ta med dem hem, då lägger man den där ist
Tror det är veganska alternativet till att välta kor.
Ibland när jag går hem från krogen tar jag med mig en kundvagn
Ser ut som hon borde chilla lite kanske
Kalla det vad du vill men det är fortfarande knark!
Inte ok att ta knark på jobbet
Legit, äter bananer genom att ta en av en bit med andra handen.
Bitchboy vem äter en macka på det viset??
Subway är fan underskattat
Hej edvin
hej
vill du bli ihop
Jaa
vill du pussas
såklart
Cha bror ska vi spela forre eller kom bror
slay kolla min nya maskara
Vad ska ni göra idag jag ska rida på spirit
Tjejer kan inte spela fotboll
Ey bror ska vi spela forre efter skolan eller
omg du är min bff slay omg
ej bror ska vi spela grabbar lite forre är jag sugen på bre jalla vi börjar
Vill någon bli ihop, e 11 om någon vill veta! Om ni gillar mig så
Du vet att det finns snuskgubbar
E du 2012 elr 2013
Adda mig
Ja järna ada mig på snap

Jag vill men jag är ful och jag är äldre så nej tack
När ens situation är så fucked att den inte ens finns på Reddit eller Quora
Varför är ni så eelaka
för vi är boys
det är inte nån flex att vara elak
jag kan mobba er mer än ni kan
För att di ä så ful
Boys bete er ni måste fan vara hjältemannar
Om vi retats de byteder att man gillar henne om han typ slår henne då han gillar tjejen ni måste fatta lite
Vi är boysen ju
Vi är boys och är coola
Vne asså
För vi är sigma
holll käften nu
Varför är ni pick me girl då??
Ronaldo är bäst i hela världen
du är bäst ingen ska reta dig
ok vem bryr sig
Asså tänk att det finns folk med epelsi
vad fan är epelsi
jag har det
menar du epilepsi
o massor av andra sjukdomar
womp womp lil nga
du är rasist
man stativer epilepsi
bror ingen bryr sig om dina problem
min brorsa har det
Ni är så söta
jag minns
Jag får inte ha okända personer på snap tyvärr
Snälla adda inte nån okänd för aa ni fattar om ni någonsin gör det men snälla inte

Jag addade du, du kan adda tillbaka lm di vill
Vän med mig snälla har knappt några vänner
Sådär komer du inte kunna bli ihop
Han menar it att han vill gifta sig med ngn
dum i huvet
hon är 6 år
wtf är det här du är typ 7
ha säämmmsst
visst mannen du är typ 6 wallah
kan du hålla käften
jage också 10
swiftie imagien heta der 8 åring
Nä, du ska inte stoppa snoppen i smöret hörnu!
Ska ni verkligen jogga i detta väder?
Får se, vikingarna korsade ju Atlanten, d day tror du de soldaterna brydde sig om regnet
Yeah bruh. Date night
Hej, vad händer i helgen
Behöver inte vara en förnedring, kan ju vara en rolig svenssexan, som gjorde till min bror
Du blir omkörd mycket men bra kört ändå
Okej, bra, får ha invigning sen, så kommer jag med champagne
Åh så söta ni är!!
Ja men absolut!!! Den va super fin annars
aa visst
okej
det sög ju
Hade de varit samtidigt skulle det vart perg
Juste är det dyrt med tågbiljetter?
Ja
Men ugh
Ehm
rädda mig nu vad ska jag göra
igga bara dp
då

Okej känner mig elak
Låtsas att du inte använder facebook?
Man ser inte om man iggar en förfrågan
okeej
Kollar du aldrig på film eller
kollar aldrig på filmer jag inte sett
Vill se mkt film
Spanar in mkt film
Men ser aldrig
Men lol
kollar du aldrig serier heller?
Du missar något stort
Jo
alltid
Jag råkade tömma luften i din madrass oj oj
Ja det närmar sig snabbare och snabbare för varje år tycker jag
Det är modigt av dig tycker jag
Det är riktigt skönt att bara mysa lite ibland också
Det känns så nervöst att flytta ensam till en helt ny stad bara!
Ska jag lägga in de eller vill du?
Man ba gubben lämna mej ifred
hahahahhahahah relaterar
Gå in
Va
Håll in den blåa
Sa bara att dem inte ska vara i tc typ 12 på natten utan hellre får komma hit
Va vilka snackar du om
Vill du göra något på lovet
Gärna!
Vilken dag kan du?
Kan alla vardagar
Kan du på måndag
aa

Ska vi leta ngn fest annars mm dag som vi kan gå på
På måndag? tror inte så många har då
Menar under veckan
Aha ja det kan vu
Det finns säkert ngn i veckan då alla har lov
Men kan vi äta eller nåt på måndag
Aa låter bra
Yes
När är du framme
14:00
E där om 1 min
Okej är framme om 5 min
Vi sed där uppe vid ica
Har fest i sollentuna
När?
Ikväll?
Vet du var han bor
Hur gammal e han
Okek nice ska kolla buss
Vrf har jag skickat denna video till massa människor??
gahaha
vilka
Robin
sen massa andra människor som jag ej typ känner
like wtf
asså dom måste ju tro jag är galen
Vilken tid kunde du smyga ut
Kanska tolv
Det kanske e för sent?? Vet ej
Tänk om jag kommee sen slutar det
Är det en stor fest?
Ye, vill inte utnyttja kompispris
För tält borde du det xD kanoter kan jag gå med på att vi betalar för

Btw så kostar tält fick jag reda på, men 100kr per tält
Men ska jag säga 11 imorgon?
Jag har blivit nykterist
Nudist kanske
Vi skiter i sova, det kan man göra i graven
Sant, man kan ju alltid sova under öppen himmel!
Jag ska elda hela natten på fredag
Nej fuck miljön, det är ju östkusten
Var ska vi träffas, på folketshus eller fika först?
så bois vad blir det? har vi något allmänt konsensus eller?
Om vi bara är hemma hos mig blir det ju betydligt billigare
Okej snyggaste killarna skrev
kolla tinder snabbt
AA vet men robin fixa snygga killar som jag var tvungen att godkänna
Han ville göra samma med tjejerna
Okej stelt har redan sagt till dom hehehhe
vad gör jag om han säger nej
haha säg bara att jag råka bjuda för många
börjar typ svettas nu
haha
skriv när han gett besked
hej vad gör ni
inget dd
inte nåt heller
ha fråga inte henne hon e dum
va vad har jag gjort
fråga din mamma
det är it jag
vart är ni
Wait no, jag är inte klar med mina julklappar fuck
Inga problem broder, jag kör typ 11.40 hemifrån
Ses imorgon
Blev faktiskt några ganska bra bilder haha ä, speciellt med såpbubblorna

Kan vi ses en liten stund idag, och snacka om helgen
När passar det dig att komma?
Ring när du e redo
Ait, lemme gå till datorn
Hmu när du e fri
vi kan hålla om varandra och storböla tillsammans
får jag följa med dig efter skolan idag
asså jag ska städa
men jag kan hjälpa till
jag älskar prime
ps älskar er som gillar det
men för dyrt
true story
prime är fet gott
håller med gått men dyrt
sant
nä inte dyrt för oss kostar de 25 kr
nej as äckligt
det är ju bara vanlig saft
vrf vill man skriva hat till någon
det borde man inte iaf
för att vara dum
för att du ska ha nåt att fråga om
sry men pick me
så fkn sant
varför är typ alla i kommentarerna rasister
för att dem inte är smarta
du då eller
håll käft
som du
du ska inte säga nontin
ingen har respekt
för dem är efterblivna

käft
hål käften din mamma knullar uff
min mamma var 2 år gammal när min morfar dog av ett hål i hjärtat
går det änns
det går det är jätte sorgligt
man har svårt för saker
gud hur fick han hål i hjärtat
stackars dig
uuuh är det här barnvänligt
visa din pojkvän då
tbh det där ser ut som en tjejs hand
nejjj du ljuger
Jag har världens fucked up berättelse för dig
fan ska jag hoppa på en annan utbildning
Denn var najs
o shit visste inte att han var chef wtf
Ja privat är nog ganska svårt att komma in i som nyexaminerad.
Sen antar jag att det är just polisen som söker flest och är lättast att komma in på
Verkligen, asså fredag eftermiddag kommer va så jävla skön hahah
aight jag måste gå lägga mig nu
kommer du fredag
ja exakt
oki
vi måste hitta på nåt kul
klura lite på det
hej vad har du haft för dig på sistone
inte så mycket faktiskt
själv då
jag har börjat lära mig spela gitarr
bre han kommer inte in på sitt konto
bro hur addar man folk
han kom in på den
tänker också om det skulle hända något med personen som ansvarar för biljetterna så blir vi

lite fucked
åh det låter kul
sen kommer horan säkert sälja den
sen kmr han inte kunna göra ngt när vi står i hans trappa
då kommer vi sälja honom
ja
värre än saw
han ska inte ha ett alternativ att överleva
sen
vi snor hans mobil
och spottar på honom
vi krossar hans mobil
sen får han smörja in sig med glaset
äta upp batteriet
sen stoppa upp resten i hans urinrör
efter det vi ba "its just a prank bro"
prank gone sexual in the hood
social experiment
mycket pengar för en person att ligga ute med dock
har du haft nationella
skulle vi träffas idag
jag behöver räddning
kmr gråta adså
behöver du en riddare i nöden?
japp
ok
vart/när
vet ej när
Bra och snabb service.
Senaste beställningen var en bad blöja, förra året ett antal UV kläder. Super bra kvalitet. Jättenöjd!
Beställde 2 par simbyxor i storlek small 5-9 kg. Bra pris och snabb leverans. Det enda jag är missnöjd med är att simbyxorna är olika stora trots samma tillverkare och storlek. Ser ut att

skilja en hel storlek.
Blev kontaktad av företaget och de skickade en ny simbyxa som ska vara samma storlek. Väldigt fint bemötande! Kan verkligen rekommendera att beställa från dom då de bryr sig om kundnöjdheten.
Mycket bra service och riktigt skönt bemötande samt kunnig personal!
Snabb och bra service.
Väldigt snabb leverans och en trevlig lite lapp om vem som hade packat paketet. Bra länkar om en behövde hjälp med montering och dylikt.
Bra bemötande, samt hjälp med köp av reservdelar. Snabba svar.
Fungerade som förväntat, smärtfritt
Efter en vecka fick jag mail att de blivit fel och varan finns inte alls i lager som lovat, utan min soffa kommer om tre månader.
Sämsta leveransupplevelsen någonsin.
Så är du inte en superstark hantverkare så rekommenderar jag inte att köpa den
När jag köper varor oavsett storlek, små eller stora, och i samma kvitto, betalar jag för frakten som ska levereras till huset, vilket innebär att den kommer att levereras till huset och inte någon annan plats nära huset, t.ex. en butik.
allt gick som smort och paketet kom fort och produkterna funkade jättebra, nemas problemas
Att sedan leksaken var njutbar för oss båda gjorde inte saken sämre!
Diskret förpackning för den som är orolig för de
Biltemas motoroljor är ju väldigt bra pris på om man jämför med märkes oljorna, Men frågan är väl om det ändå inte är så att man får vad man betalar för till stor del
På mitt konto kan jag se att de har dragit summan.
Står ingenstans hur man tömmer vattenbehållaren när man strukit om det blir vatten kvar i järnet.
Kvittot kom och en dag senare fick jag pengarna på Swish. Kunde inte varit snabbare och smidigare!
Snabbt, smidigt och totalt sett helnöjd. Pengarna på kontot inom 5 minuter efter godkännande av värdering
Allt gick väldigt smidigt och kändes säkert. Inga konstigheter, tydlig information och man får det som står på hemsidan.
Väldigt snabb leverans och en trevlig lite lapp om vem som hade packat paketet. Bra länkar om en behövde hjälp med montering och dylikt.
Bra betalt och snabbt utbetalat, vilket naturligtvis uppskattas mycket.
Tydlig information i brevet som man beställer hem, snabb värdering och blixtnabb utbetalning med swish

Fick superbra pris för mitt guld och swishen den kom också superfört efter jag godkänt beloppet. Rekommenderas verkligen!
Det var bra att företaget skötte allt från början till slut, men jag hade uppskattat att få veta alla reparationer som behövdes på en gång i stället för i omgångar.
Tog lång tid innan bilen blev såld kanske för att jag inte fick info om däckbyte till snyggare fälgar och att bilen inte tvättades utvändigt förrän jag sa upp kontraktet.
Så roligt att höra att du är nöjd. Det kommer säkerligen att bli bra för din kanin att ligga på de mjuka fällarna.
Var och köpte en TV som Elgiganten, Bäckebo packat upp och kalibrerat. När vi kommer hem saknas delar.
Går jag på spa vill jag kunna ta det lugnt, stressa ner, och inte dela pooler med så många samtidigt.
Värsta lekland med badet full av ungar som tilläts göra "bomben" ned i vattnet och förstöra för andra i största allmänhet.
Va fan är det här, ta riktigt körkort
Avskaffa EPA helt och hållet, och låt 17-åringar ta körkort
Eller avskaffa epa och köp en moppe tills man är 18 istället
Hur många hjärnceller har en EPA-förare
En burk Ben&Jerry's innehåller cirka ett halvt dagsintag i kalorier. 2 per dag så har man sin energi för dagen. Borde vara mellan 8-10 i varje låda från leverantör Det är inte helt omöjligt ändå. Fast man har x antal år av föda i ren glass
Åldersgränsen ska också sänkas, från 18 år till 16 år med vårdnadshavares medgivande.
Finland ni är crazy
Och alla jävla miniförpackningar ... är det verkligen billigare än att bara ha vanliga morötter..
Helt irrelevant, fransmän är inte ens människor.
Jo men chokladmjölken är fat free så det är mer eller mindre hälsokost!
Vafan, dom får ju chokladmjolk! Det gör den ju automatiskt bättre
Ta med oboypulver och blanda med skolans mjölk
Har snart klarat första gymnasieåret, tacka fan att du sa detta. Ska göra det från och med terminsstart i tvåan!
Jag saknar mac n cheese
Skulle gissa att det där är typ 600kcal vilket är ganska rimligt för en lunch
8 cl bensin är också typ 600 kcal, men inte skulle jag dricka det för det.
2kr delicatoboll eller 5kr korv

BILTEMA-RUSHEN VID HÅLTIMME
Nämen blå!!
Jag när jag ser left-over pizza eller thaimat på morgonen när jag är bakfull
Usch! Röd mjölk ska det vara. Vem fan dricker ens mellanmjölk
Danskarna har ju fan konstant rea i Sverige också. Många jobbar ju kvar i Danmark
Bättre än så blir det inte, tjänar pengar i Danmark o spenderar sagda pengar här
Att flytta till ett u-land för att leva billigt är väl inget nytt?
Inte ens danskar gillar Danmark
Synd bara att danska är ännu mer obegripligt att förstå. Har nog svårt att höra vad skåningar snackar om
Det knasiga är att skåningarna och danskarna förstår varandra väldigt dåligt.
It's a trap guys! Det är deras plan för att ta tillbaka Skåne
lol... snacka om att de kommer ångra sig
Jag tror Gustav just nu ligger och vänder sig i sin grav så snabbt att hans kvarlevor genererar nog med ström för att förse landet med el tills det att vi tagit över Norge som en hämndaktion.
Jag laddar mynningsladdaren och ställer mig vid Svinesund. Vi ses i gryningen.
HELVETEEEE!
Håll din stora käft din jävla skitstövel
Är du dum?
Relativt till dig så nej
Vad I fan är en dansk dollar?
Du förstår nog inte ironi
Först vinner dom Vasaloppet, men vad kommer härnäst? Smyg kolonisering av Jämtland? Ockupation av Härjedalen? Preventivt anfall nu!
Känner samma inför mellofinalen
Exakt samma med mello
Eller dansk, fuck danmark
Går att applicera på melodifestivalen i år också!
Alltså fan.. jag älskar det här så innerligt! Talar du direkt till min själ?!
Hej! Vad gör du?
Tjena! Bara chillar hemma. Du då?
Samma här. Funderar på att dra ut en sväng senare, några förslag?

Kanske dra till bion? Det går en ny skräckfilm som verkar grym.
Låter spännande! Är den så läskig som alla säger?
Ingen aning, men jag är redo att bli skrämmd haha! Vill du joina?
Absolut! Låter som en plan. Ska vi säga kl. 19 vid biografen?
Perfekt! Ses där då!
Har ett stort matteprov nästa vecka.
Åh, matte är inte min grej.
Lycka till med det! Behöver du hjälp med något?
Tack! Jag kämpar mig igenom det, men det är alltid bra med lite extra övning.
Absolut, om du behöver hjälp är jag här.
Vill du komma över och spela lite efter provet?
Låter som en plan! Jag behöver definitivt en paus från allt plugg.
Ses då!
stackars dig
färgar mitt hår nu
är du med på festen ikv
har du sett nya netflix serien
Ska vi beställa pizza ikväll?
Har du hört talas om den nya restaurangen som öppnade?
Jag behöver verkligen en kaffe just nu.
Ska vi dra och ta en fika?
Vet du när nästa buss går?
Jag tror att jag glömde min mobil hemma.
Jag träffade Sara idag, vi gick runt i stan och pratade.
Kom ihåg att ta hand om dig själv, det är viktigt att ta det lugnt ibland.
Jag älskar att gå på promenader i parken och bara njuta av naturen.
Tänk att det redan gått så många år sedan vi tog studenten, tiden går verkligen fort.
Jag testade det där nya caféet runt hörnet, de hade grym kaffe.
Så jag var på gymmet idag och tränade, och nu känns mina armar som spaghetti.
Jag blev helt fast i den nya serien på HBO, binge-watched hela säsongen.
Haha, du måste se vad jag hittade på TikTok igår!
Tänkte dra till parken sen.

Ska vi hänga med?
Klart! Låter som en plan.
Vad vill du göra sen?
Kanske dra och köpa glass?
Låter som en bra idé!
Vilken smak föredrar du?
Jordgubb, alltid!
Vad tror du, naturvetenskapligt eller samhällsvetenskapligt?
Äh, jag tror jag drar åt naturvetenskapligt håll.
Låter som en bra plan! Vilka ämnen lockar dig där?
Matte och biologi är nog mina favoriter.
Ah, coolt! Jag lutar mer åt samhällsvetenskapliga, gillar historia och samhällskunskap.
Fett! Det låter som du har koll på vad du gillar.
Spännande! Låter som vi har olika intressen.
Absolut

Chillar bara, du?
Kom hit, vi drar till stan.
Asså, du missade en sjuk fest igår!
Lol, såg du vad han la upp på Insta?
Kan vi hänga imorrn? Behöver snacka.
Vart e du? Jag är vid torget.
Orka plugga, vill bara sova.
Haha, du är så rolig ibland!
Tja, ska vi dra o käka nåt?
Sjukt najs väder idag, va?
Sorry, glömde svara, vad sa du?
Är det fest hos dig ikväll?
Hörde att det var drama i skolan idag?
Dina snaps är alltid så kul!
Lyssna på den här låten, såå bra!
Blev klar med serien, ge mig tips på nåt nytt!

Orkar inte med allt drama just nu.
Asså seriöst, så trött på skolan.
Vi ses snart, tagga!
Wazzup? Läget?
Fett trött idag, behöver en powernap.
Ska vi linka sen?
Din story var så cringe igår, haha.
Let's grab some grub, jag är hungrig af.
Har du hört den nya låten?
Dags att flexa våra moves
Bror, du måste se den nya serien
Några chillplans för ikväll?
Läget? Allt good?
Den där memen lol
Ska vi kolla in den nya restaurangen?
Hörru, kan du sno mig en favvo?
Tjena! Vad gör du just nu?
Har du varit med den nya tjejen?
Dra t gymmet?
Ska vi köra några rundor på Fortnite ikväll?
Några dope tracks att rekommendera?
Bror, har du hört vad som hände på festen?
Skönt att bara chilla ibland, eller hur?
Några snackisar i skolan idag?
Ska vi chilla snart?
Blir det fest ikväll?
Kan du fixa några snacks?
Har du hört om det senaste?
Ska vi ses på lunchen?
Någon ny serie att rekommendera?
Hur har din dag varit?
Ska vi hänga med mina kompisar ikväll?
Ja gärna

Appendix 4: Kod för modellen

```
train_df = pd.read_csv('training1.csv', names=["Phrase", "Value"])
X_train = train_df["Phrase"]
Y_train = train_df["Value"]

val_df = pd.read_csv('validering.csv', names=["Phrase", "Value"])
X_val = val_df["Phrase"]
Y_val = val_df["Value"]
# Generera representationsvektorer
train_rep = []
for phrase in X_train:
    tokens = tokenizer(phrase, return_tensors="pt", padding=True, truncation=True)
    with torch.no_grad():
        outputs = model(**tokens)
    last_hidden_states = outputs[0]
    average_token_representation = torch.mean(last_hidden_states, dim=1)
    train_rep.append(average_token_representation)
X_train_tensor = torch.cat(train_rep, dim=0)
```

```
val_rep = []
for phrase in X_val:
    tokens = tokenizer(phrase, return_tensors="pt", padding=True, truncation=True)
    with torch.no_grad():
        outputs = model(**tokens)
    last_hidden_states = outputs[0]
    average_token_representation = torch.mean(last_hidden_states, dim=1)
    val_rep.append(average_token_representation)
X_val_tensor = torch.cat(val_rep, dim=0)

naive_bayes = GaussianNB()
naive_bayes.fit(X_train_tensor.numpy(), Y_train)
predictions = naive_bayes.predict(X_val_tensor.numpy())
accuracy = accuracy_score(Y_val, predictions)
print("Noggrannhet:", accuracy)
```

```

from transformers import AutoTokenizer, AutoModelForSequenceClassification
test_df = pd.read_csv('testdata.csv', names=["Phrase", "Value"])
X_testa = test_df["Phrase"]
Y_testa = test_df["Value"]
test_representations = []
for phrase in X_testa:
    tokens = tokenizer(phrase, return_tensors="pt", padding=True, truncation=True)
    with torch.no_grad():
        outputs = model(**tokens)
        last_hidden_states = outputs[0]
        average_token_representation = torch.mean(last_hidden_states, dim=1)
        test_representations.append(average_token_representation)
X_test_tensor = torch.cat(test_representations, dim=0)
predictions = naive_bayes_model.predict(X_test_tensor.numpy())

```

```

incorrect_predictions = []
for i in range(len(predictions)):
    if predictions[i] != Y_testa[i]:
        incorrect_predictions.append((X_testa[i], Y_testa[i], predictions[i]))
print("Felaktiga förutsägelser:")
for phrase, true_label, predicted_label in incorrect_predictions:
    print(f" {phrase}")

```

```

print("Skriv en rad text eller 'n' för att avsluta:")
texts = []
while True:
    user_input = input("> ")
    if user_input.lower() == 'n':
        break
    texts.append(user_input)
for text in texts:
    tokens = tokenizer(text, return_tensors="pt", padding=True, truncation=True)
    tokens = tokens.to(device)
    with torch.no_grad():
        outputs = model(**tokens)
    hidden_states = outputs[0]
    average_hidden_states = torch.mean(hidden_states, dim=1)
    prediction = naive_bayes_model.predict(average_hidden_states.cpu().numpy())
    print(f"Klassificering av text: {text}")
    print("Resultat av klassificeringen:", prediction)

```

Appendix 5: Uträkningar

$$\textit{Accuracy} = \frac{43+41}{43+7+9+41} = 0.8400$$

$$\textit{StandardError}_{\textit{Accuracy}} = 1.96 * \sqrt{\frac{0.8400(1-0.8400)}{100}} = 0.0719$$

$$\textit{Intervall}_{\textit{Accuracy}} = 0.8400 \pm 0.0719$$

$$\textit{Precision} = \frac{43}{43+9} = 0.8269$$

$$\textit{StandardError}_{\textit{Precision}} = 1.96 * \sqrt{\frac{0.8269(1-0.8269)}{43+9}} = 0.1028$$

$$\textit{Intervall}_{\textit{Precision}} = 0.8269 \pm 0.1028$$

$$\textit{Recall} = \frac{43}{43+7} = 0.8600$$

$$\textit{StandardError}_{\textit{Recall}} = 1.96 * \sqrt{\frac{0.8600(1-0.8600)}{43+7}} = 0.0962$$

$$\textit{Intervall}_{\textit{Recall}} = 0.8600 \pm 0.0962$$

$$F1 = \frac{2*0.8269*0.8600}{0.8269+0.8600} = 0.8431$$

$$\textit{StandardError}_{F1} = 1.96 * \sqrt{\frac{0.8431(1-0.8431)}{43+7+9}} = 0.0928$$

$$\textit{Intervall}_{F1} = 0.8431 \pm 0.0928$$