Examensarbete

HÖGSKOLAN I SKÖVDE

ETISKA OCH SOCIAL KONSEKVENSER AV ML-MODELLER INOM CYBERBROTTSBEKÄMPNING

ETHICAL AND SOCIAL IMPLICATIONS OF ML MODELS IN CYBER CRIME FIGHTING

Examensarbete inom huvudområdet informationsteknologi

Grundnivå 30 Högskolepoäng
Vårtermin 2024

Asma Ibrahim Abdi

Handledare: Christian Lennerholt
Examinator: Mikael Berndtsson

Sammanfattning

Denna studie syftar till att undersöka de etiska och sociala konsekvenserna av att använda maskininlärning inom cyberbrottbekämpning. Maskininlärning är en växande teknik inom cybersäkerhetsvärlden, men det finns inte tillräckligt med forskning om de utmaningar som kan uppstå vid användning av maskininlärning för cyberbrottbekämpning. För att undersöka dessa utmaningar har en kvalitativ ansats använts, där intervjuer har genomförts för att förstå vad personer som arbetar med cybersäkerhet tycker om cybersäkerhetssystem som inkorporerar maskininlärningsmodeller.

Resultatet av intervjuerna visar att det finns många fördelar med att använda en effektiv teknik som maskininlärning. Men det finns också många utmaningar, såsom algoritmer som är partiska och förstärker fördomar. Maskininlärning kräver stora mängder träningsdata, vilket kan leda till förlust av människors integritet. Denna studie kommer att fördjupa sig i dessa och andra utmaningar med maskininlärning.

Nyckelord: Maskininlärning, ML, ML-modeller, cybersäkerhet, cyberbrottsbekämpning, brottsbekämpning, integritet, sekretess, datasäkerhet, partiskhet, maskininlärning konsekvenser

Innehållsförteckning

1 Inledning

Maskininlärning har nu sitt momentum och är det nya populära verktyget inom teknologivärlden. Det är en teknik som används för att göra förutsägelser. Många organisationer är mycket intresserade av maskininlärning och hur denna typ av verktyg kan bidra till deras verksamhet (Sharifani & Amini, 2023). Ett område där maskininlärning börjar användas är inom cybersäkerhet.

Cyberattacker ökar varje år, vilket har skapat ett behov av att använda maskininlärning för att motverka hot såsom malware, phishing, man in the middle attacker (MITM) och andra typer av cyberattacker (Gottsegen, 2023). ML-algoritmer kan identifiera mönster i data för att förutsäga attacker. Trots att maskininlärning har många fördelar, finns det också utmaningar med att använda tekniken för att bekämpa cyberbrott (Flower, 2023).

Diskussionen om etik och integritet har varit central när det gäller maskininlärning, särskilt nu när världen blir mer digitaliserad och mängden data ökar. Detta problem har redan påverkat många människor och kommer att fortsätta att påverka fler när samhället blir mer datadrivet.

Nu när maskininlärning börjar användas inom cybersäkerhetsområdet måste det finnas forskning om utmaningarna som kommer med att använda ML-modeller inom detta område. Elluri et al. (2023) påpekar att det borde finnas mer forskning om konsekvenserna av att använda maskininlärning inom cybersäkerhet. Syftet med denna studie är att undersöka detta problem. Frågeställningen som ligger till grund för undersökningen är: "*Vilka sociala och etiska konsekvenser uppstår vid implementeringen av ML-modeller inom cyberbrottbekämpning?*"

I denna studie kommer frågeställningen att besvaras genom en kvalitativ ansats. Deltagare med erfarenhet inom cybersäkerhet kommer att besvara frågor relaterade till de etiska och sociala konsekvenserna av maskininlärning inom cybersäkerhet. Den insamlade informationen från intervjuerna kommer sedan att analyseras för att komma fram till ett resultat.

Författaren kommer därefter att diskutera de vetenskapliga, sociala och etiska aspekterna av denna studie. Slutligen kommer förslag på framtida forskning inom detta område att presenteras.

2 Bakgrundskapitel

ML, vilket är en förkortning av ordet maskininlärning, är en gren av artificiell intelligens, även kallad AI. Maskininlärning hjälper datorer att lära sig från data för att göra förutsägelser utan explicit programmering. Denna typ av tekniska framsteg för att fatta bättre beslut gör vågor i många branscher som sjukvård, finans och marknadsföring, etc. (Sakharkar, 2024). Förutom att ML hjälper till med förutsägelser och att fatta bättre beslut samt att hjälpa organisationer att arbeta mer effektivt, hjälper det dessutom till med att spara pengar för företag. Det finns många fördelar med att inkorporera maskininlärning i ens organisation, vilket har hjälpt detta verktyg att bli mer populärt inom olika industrier (Flower, 2023).

Maskininlärning hjälper företag att utföra manuella uppgifter som är mycket svåra att genomföra i skala. Detta kan vara bearbetning av stor mängd data som genereras av digitala enheter. Många av de största företagen idag använder sig av maskininlärning, det är en central del av deras verksamhet. Google, Facebook och Uber är etablerade företag som drar nytta av fördelarna som kommer med maskininlärning (Tucci & Burns, 2023). Populariteten av maskininlärning blir mer uppenbar ur ett ekonomiskt perspektiv. Maskininlärning är en framväxande marknad, under år 2021 förde ML-marknaden in 15,44 miljarder amerikanska dollar. Det är en marknad som ständigt växer och förväntas dra in runt 210 miljarder amerikanska dollar år 2030, vilket innebär en ökning på 38,8 procent årligen (Flower, 2023).

AI med hjälp av ML kommer vara ett värdefullt verktyg i framtiden. Inom många industrier är mänsklig interaktion väsentlig när det kommer till säkerhet. Speciellt inom cybersäkerhet är det viktigt att integrera mänsklig input, trots detta finns det en utveckling i tekniken som gör det möjligt för datorer att utföra specifika uppgifter mycket bättre än människor. Med varje tekniskt framsteg kan samhället nå en framtid där mänskliga roller kan kompletteras mer effektivt (Kaspersky, 2023).

Ökningen av cyberattacker har skapat en nödvändighet för maskininlärning inom cybersäkerhet för många företag. Cyberattacker blir mer avancerade med nya tekniker och metoder för att få tillgång till företagets tillgångar (Gottsegen, 2023). Här är det där maskininlärning kommer in. Det är en ständigt utvecklande teknik som hjälper till att hantera de nya hoten, vilket gör den till en väsentlig del för att bekämpa cyberbrott. Maskininlärning är effektivt för att analysera mängder av data och mönster vilket gör det till ett kraftfullt verktyg för att upptäcka attacker i den tidiga fasen (Gottsegen, 2023).

Förutom att maskininlärning är ett bra verktyg för att stoppa malware-attacker kan det också vara effektivt mot nätfiske-e-postmeddelanden, vilket är ett stort problem. Maskininlärning används för att lättare upptäcka attacker genom att analysera

domännamn, länkar, bifogade dokument och knappar. Detta hjälper datorn att detektera hotet, blockera det och stärka brandväggen i protokollen (Abdiyeva-Aliyeva et al., 2022). Det finns flera sätt som maskininlärning kan hjälpa samhället att bekämpa cyberbrott. Det är ett av de mer lovande verktygen som kommer från teknikens framsteg och som har en positiv effekt på samhället.

Dessutom genereras en stor mängd data varje dag. Detta kommer naturligtvis att öka i framtiden, och det är sannolikt att detta kommer att skapa ännu större behov av maskininlärning i den moderna framtiden (Tucci & Burns, 2023)

2.1 Artificiell intelligens och kopplingen med maskininlärning

För att förstå hur maskininlärning används måste olika begrepp definieras som är kopplade till maskininlärning. Dessa begrepp inkluderar AI och maskininlärningsalgoritmer (Carta, 2022).

AI, är en gren inom datavetenskap. Det är en teknik som hjälper datorn att efterlikna mänskligt beteende. AI använder sig av smarta maskiner och olika applikationer för att kunna utföra komplexa uppgifter utan att behöva få hjälp från människor (Maheshwari, 2023) (Carta, 2022). Utan att komplicera det är AI en representation av en datoriserad maskin med intelligens på mänsklig nivå. För att kunna driva AI måste huvudverktyg som maskininlärning användas (Maheshwari, 2023).

AI använder sig av ML, och deras processer för att analysera mängder av träningsdata för att hitta mönster, som AI använder för att göra förutsägelser, är likadana. Ett exempel på detta är chatbots som tar in data från konversationer, vilket resulterar i att de kan identifiera skrivmönster från människor och sedan förstå hur de ska skriva som en människa (Laskowski & Tucci, 2023). AI omfattar användningen av olika tekniker för att möjliggöra för datorer att replikera mänskligt beteende. Detta är en effektiv metod för att lösa komplexa problem, men AI har också sina problemområden. Detta inkluderar kunskapsrepresentation, resonemang, lärande, planering, perception och mer. Tidigare forskning inom AI visar att utvecklare använde hårdkodade uttalanden på ett formellt sätt som sedan datorn använde för att tänka logiskt. Detta kallas kunskapsbaserad metod, men denna metod har sina begränsningar. Det är svårt att förklara komplexa tankar om människor inte kan uttrycka sig fritt utan att behöva tänka på hur de ska säga det på ett formellt sätt (Janiesch et al., 2021).

Här kan maskininlärning lösa problemet med begränsningarna som fanns med tidigare AI. Ett datorprogram som använder ML blir bara mer effektivt med erfarenhet. När detta upprepas blir programmet bättre på att lösa problem (Janiesch et al., 2021)

2.2 Maskininlärningsteknik

Det finns olika tekniker som ML-algoritmer använder, och den avgörande faktorn för vilken teknik som ska användas beror på situationen och målet som ska uppnås (Carta, 2022). Det finns tre inlärningstyper. Den första är "supervised learning", vilket på svenska heter övervakad inlärning. Den andra heter "unsupervised learning", även känd som oövervakad inlärning. Den tredje tekniken är förstärkningsinlärning (reinforcement learning) (Staff, 2023).

Övervakad inlärning (supervised learning, SL-teknik): Denna inlärningsmetod används för att träna modellen med träningsdata som redan har en etikett eller kategori kopplad till den. När träningen av modellen är avslutad används denna modell för att lära nya modeller med otränade data för att sedan kunna placera dem i rätt kategori. Detta lärande består av två huvudsteg. Första steget är att träna en klassificerare med kategoriserade exempel och det andra steget är att verifiera. Modellen testas sedan med det nya data för att undersöka om modellen ger det önskade resultatet (Verma et al., 2019). SL-tekniken har två typer av uppgifter som den kan utföra, vilka är regression och klassificering. Regression innebär att förutsäga ett numeriskt värde baserat på information om framtiden. Däremot handlar klassificering om att förutsäga vilken kategori något tillhör (El Mrabet et al., 2021).

Oövervakad inlärning (unsupervised learning, UL-teknik): När det kommer till oövervakad inlärning är detta motsatsen till övervakad inlärning. Målet med UL-tekniken är att identifiera grupper eller kluster, som de kallas, från datamängden utan att det finns förhandskännedom om tidigare klassetikett för varje indata i träningsfasen (El Mrabet et al., 2021). I denna metod ska algoritmen hitta de dolda mönstren. Denna metod har inget rätt eller fel svar, det är en teknik där datavetare låter systemet hitta mönster i data och ser vad resultatet blir efteråt utan att förvänta sig något specifikt (Carta, 2022).

Förstärkningsinlärning (reinforcement learning, RL-teknik): Jämfört med de tidigare metoderna har förstärkningsinlärning (RL) inte ett behov av historisk data om ett problem, eftersom den inte behöver information i förväg.
Förstärkningsinlärningsmetoden lärs över tid. Det är en metod som hjälper en agent att lära sig genom experiment. Genom att prova olika handlingar i en kontrollerad miljö kan förstärkningsinlärning maximera en total belöning (Castañón, 2019).

2.3 Träningsdata i maskininlärning

En maskininlärningsalgoritm är den centrala metoden som hjälper AI-system att utföra arbetsuppgifter, oftast genom att förutsäga vilka resultat som kommer att uppstå

baserat på den information som matas in. De olika sätten som dessa algoritmer kan göra detta på är genom två huvudprocesser som kallas klassificering och regression (Hashemi-Pour & Wigmore, 2023).

För att effektivt utföra sina uppgifter använder AI-system ML-algoritmer, som representerar en uppsättning matematiska processer eller tekniker. Det är datavetenskapsexperten som har ansvaret att tillhandahålla träningsdata till en ML-algoritm. Detta möjliggör att algoritmen lär sig från träningsdata och förbättrar sina beslutsfattande förmågor över tiden (Hashemi-Pour & Wigmore, 2023). När en datavetare har matat in data till en ML-algoritm styr forskaren algoritmen att undersöka specifika variabler inom dem för att identifiera mönster. Anledningen till detta är att forskare tänker att algoritmen ska lära sig över tid på egen hand. Om mer data matas in över tid blir analysen mycket lättare för ML-algoritmen utan att människor behöver bli involverade i processen (Hashemi-Pour & Wigmore, 2023).

Träningsdata för maskininlärning är väsentligt, det är nyckeln till att skapa en maskininlärningsmodell eller datadrivna system. Data kommer i olika former, dessa former är strukturerad, ostrukturerad, halvstrukturerad och metadata (Sarker, 2021).

Strukturerad: Denna typ av data har ordning och struktur och följer en datamodell enligt standardordning. Det är organiserat och lätt åtkomligt för olika enheter eller dataprogram som kan användas vid behov. Strukturerade data sparas vanligtvis i tabellformat och databaser. Några exempel på strukturerade data är namn, datum, adresser, kreditkortsnummer och mer (Sarker, 2021).

Ostrukturerad: Motsatsen till strukturerad data är ostrukturerad data. Denna typ av data saknar ordning och har inte ett förbestämt format, vilket gör det svårt att hantera och analysera. Detta innehåll kan vara text, multimedia, e-postmeddelanden, blogginlägg, wikis, affärsdokument och mer (Sarker, 2021).

Halvstrukturerad: Data som är halvstrukturerad lagras inte på samma sätt som strukturerad data och har inte lika strikta standarder för organisering. Även om det inte är lika strukturerat finns det fördelar med denna typ av data som kan användas av organisationer. Oftast kommer den i format som HTML, XML, JSON-dokument och NoSQL-databaser, vilka är exempel på semistrukturerad data (Sarker, 2021).

Metadata: Denna typ av data skiljer sig från de tidigare nämnda. En enkel beskrivning av metadata är att det är information om data. Till skillnad från vanliga data, som går att mäta och dokumentera för en organisations användning, beskriver metadata betydelsefull datainformation som ger större kontext för användaren. Ett exempel på metadata för ett dokument kan vara författare, filstorlek, skapelsedatum, nyckelord och andra beskrivande detaljer som definierar dokumentet (Sarker, 2021).

Dessa typer av data används för att träna ML-algoritmer och all data har olika användningsområden. För att analysera och undersöka sådan data inom ett specifikt område, och för att hitta viktig information som hjälper människor att skapa användbara applikationer, måste olika maskininlärningstekniker utnyttjas (Sarker, 2021).

2.4 Användningen av maskininlärning inom cybersäkerhet

Samhället blir allt mer digitaliserat och det kommer ständigt nya teknologiska innovationer. Internet har förändrat hur människor lever, arbetar och interagerar med varandra. Människor är nu mer beroende av internet eftersom det utgör en stor del av deras vardag. Denna ökade digitalisering har resulterat i en ökning av säkerhetshot på många sätt.

Den ökade digitaliseringen har lett till att cyberbrottslighet sprider sig i stor omfattning. Forskare inom cybersäkerhet betonar vikten av att vara proaktiv med att förebygga cyberbrott för att uppnå en säkrare internetmiljö. Med detta menar de att det är viktigt att rapportera misstänkta cyberbrottsliga handlingar till myndigheterna för vidare utredning och åtgärder (Veena et al., 2022) (Halbouni et al., 2022).

Det är viktigt att människor är mer kunniga inom cybersäkerhet. Definitionen av cybersäkerhet är processen att implementera cyberskyddsåtgärder och policys för att skydda data, program, servrar och nätverksinfrastrukturer från obehörig åtkomst och potentiella ändringar av dessa tillgångar (Halbouni et al., 2022). De flesta datorsystem och nätverksinfrastrukturer är sammanlänkade tack vare internet. Som ett resultat av detta har cybersäkerhet vuxit fram som ryggraden för många företag, regeringar och även allmänheten för att säkerställa integritet (Halbouni et al., 2022).

Människor skickar och tar emot data över nätverksinfrastrukturer, såsom en router. Genom att hacka en router kan brottslingar få tillgång till människors data, som de sedan kan ändra och använda för sina egna syften. Användningen av internet har bidragit till en ökning av data som dessutom är mer komplex än tidigare, vilket har lett till uppkomsten av ”Big Data” (Halbouni et al., 2022). Kombinationen av att människor använder internet allt mer varje dag, vilket leder till omfattande data, har skapat ett behov av ett pålitligt system för intrångsdetektering (Halbouni et al., 2022).

Spridningen av cyberattacker har gjort att traditionella säkerhetslösningar inte längre är tillräckliga för att stoppa cyberbrott. Att dra nytta av nya teknologiska innovationer, såsom artificiell intelligens, och särskilt maskininlärningsteknik, är väsentligt för att utveckla och förbättra ett dynamiskt, automatiserat och uppdaterat säkerhetssystem genom att analysera säkerhetsdata (Sarker, 2022). Cyberbrottslingar är snabba att anpassa sig till nya miljöer, vilket är nödvändigt för dem. Om säkerheten förbättras kommer de att hitta nya lösningar för att undvika att bli upptäckta. Dessutom innehåller

nya malware-kit redan nya metoder för att undkomma antivirusprogram och andra system för hotdetektion (Sarker, 2022).

För att lösa problemet med cyberbrottslighet finns det inte förutbestämda data för att lösa cyberbrottsfall, därför har det uppstått ett behov av maskininlärningsmodeller. Genom maskininlärning kan data klassificeras genom analys och förutsägelser (Veena et al., 2022). Målet med att införa maskininlärningstekniker inom cybersäkerhet är att förbättra nätverkssäkerheten för att skydda det från angripare. Genom att använda klusterberäkningstekniker och utvärdera prestandan för olika upptäcktsmetoder för cyberbrott i realtid (Veena et al., 2022).

Maskininlärning, specifikt inom cybersäkerhet, kan i hög grad bidra till att dra insikter från data. Cybersäkerhetsdata kan vara strukturerad eller ostrukturerad och kan komma från olika källor. Genom att utvinna värdefulla insikter från dessa datakällor kan skapandet av olika applikationer för att förhindra olika hot bli möjligt. Det är viktigt att verkliga applikationer inom cybersäkerhet har tillgång till intelligenta dataanalysverktyg och metoder som kan utvinna värdefull kunskap från data (Sarker, 2022).

2.5 Hur kan maskininlärning hjälpa att stoppa cyberattacker?

Maskininlärning har förmågan att sortera igenom miljontals filer för att identifiera potentiella hot. Ett exempel på hur effektiv maskininlärning är på att stoppa attacker är när Microsoft Windows Defender under året 2018 hindrade en skadlig programvara från att infektera över 400 000 användare med en gruvarbetare för kryptovaluta under en 12-timmarsperiod. Detta system består av flera lager av maskininlärning för att identifiera och blockera upplevda hot (Gottsegen, 2023).

Under de senaste åren har maskininlärningstekniker använts för många olika problem relaterade till cybersäkerhet. Exempel på detta är intrångsdetektering, analys och upptäckt av skadlig programvara, skräppostfiltrering, avvikelse- och bedrägeriupptäckt, detektering av nolldagsattacker, upptäckt av nätmobbning, IoT-attacker och hotanalys (Sarker, 2022).

Data är en central del av maskininlärningsalgoritmer. Det är användningen av data som bygger dessa modeller, därför är det viktigt att ha en grundläggande förståelse om cybersäkerhetsdata. Det finns många datamängder tillgängliga för analys inom cybersäkerhet, dessa kan innehålla data från olika cyberattacker som intrångsdetektering, upptäckt av skadlig programvara och spamfiltrering (Sarker, 2022).

Den mest använda datamängden kallas KDD'99 Cup-datasetet. Den består av 41 attribut med klassificering av attacker, dessa attacker är indelade i kategorierna Denial of

Service (DoS), Remote-to-Local (R2L) intrång, User-to-Root (U2R) och PROB (Sarker, 2022). Det finns också en uppdaterad version av KDD'99 som kallas NSL-KDD, skillnaden är att onödiga poster inte ingår för att förhindra snedvridning av maskininlärningsmodeller mot frekventa händelser. För att kunna utvärdera dessa intrångsdetektionssystem har forskare och tekniker vid MIT Lincoln Laboratory gjort olika datamängder offentliga. Detta kommer att hjälpa dem att få återkoppling för att förbättra olika system som förhindrar cyberattacker (Sarker, 2022).

Det finns flera olika datamängder som är tillgängliga för allmänheten. Dessa data, tillsammans med deras olika attribut och cyberattacker, skulle kunna användas för att förstärka och utveckla pålitliga cyberapplikationer genom maskininlärningsbaserad analys. Genom att undersöka och bearbeta dessa säkerhetsdata blir det lättare att skapa maskininlärningsmodeller. Detta kan i sin tur leda till intelligenta cybersäkerhetstjänster (Sarker, 2022).

2.6 Vad för typer av cyberattacker finns det?

En cyberattack inträffar när en hackare använder datateknik för att få tillgång till offrets nätverk eller dator genom att kringgå skyddshinder (Dasgupta et al., 2020). Institute for Security Technology har beskrivit en cyberattack som en attack mot ett datorsystem som kompromitterar konfidentialiteten, integriteten eller tillgängligheten för informationen i systemet (Dasgupta et al., 2020). Cyberattacker kan kategoriseras i olika typer av attacker.

Malware-attacker: Malware står för skadlig programvara. Dessa attacker kan förekomma på alla typer av enheter och operativsystem (Dasgupta et al., 2020) (Sudar et al., 2020). Attacken sker när en brottsling utvecklar och installerar skadlig programvara som hjälper dem att få åtkomst till offrets dator och deras information utan att offret vet om det. Anledningen till att angripare använder denna typ av attack är för att få tag i offrets personliga information, som inloggningsuppgifter och konfidentiell information (Sudar et al., 2020). Det är en typ av attack som har funnits i flera år för att uppnå angriparens mål att stjäla information eller förstöra ett system. Det finns olika typer av malware-attacker beroende på vad angriparen vill uppnå med sin attack. Dessa typer av malware inkluderar spionprogram, virus, ransomware, trojaner, scareware, bots och rootkits (Dasgupta et al., 2020).

SQL-injektionsattacker: Denna typ av attack utvecklades för att angripa databasdrivna webbplatser. Angriparen använder SQL-frågor mot databasen för att injicera indata från klienten till servern. De försöker identifiera sårbarheter i databasen för att få tillgång till information de inte har behörighet till. De använder fördefinierade SQL-kommandon istället för förväntade data för en postbegäran. Om detta kommando exekveras kan de läsa känsliga data från databasen eller ändra eller ta bort konfidentiella data från

databasen, vilket ändrar applikationen som använder databasen för att lagra information eller operativsystemet (Dasgupta et al., 2020) (Sudar et al., 2020).

Kompromiss med användaråtkomst: Att få åtkomst till en användares personliga information, som deras lösenord, är den vanligaste taktiken för denna typ av attack. Hur detta blir möjligt är genom att övervaka nätverkstrafiken för att fånga upp okrypterade lösenord, använda social ingenjörskonst för att komma åt lösenordsdatabaser eller genom att använda brute-force-attacker eller ordboksattacker för att gissa lösenord (Dasgupta et al., 2020). En annan populär metod för att få åtkomst till användarinformation är genom nätfiskeattacker. Denna metod innebär att skicka e-postmeddelanden som ser trovärdiga ut för att locka offret att utföra vissa åtgärder, som att klicka på en länk som tar dig till en sida där du behöver skriva in dina inloggningsuppgifter, eller en bilaga som vill att offret ska ladda ner skadlig programvara (Dasgupta et al., 2020). Detta kallas "social engineering" och innebär att utnyttja människors förtroende för att få åtkomst till personlig information. Det finns också attacker som är riktade till specifika personer eller företag, vilket kallas "spear phishing". För att utföra denna attack måste hackaren forska på målen för att skapa ett trovärdigt meddelande. För att få ett meddelande att se trovärdigt ut används e-postspoofing, vilket innebär att skicka ett mejl till offret från en av deras ledningstjänstemän eller någon annan känd person som deras företagspartner. Denna taktik kräver mer arbete eftersom hackaren måste ta sig tid att skapa webbsidor som är kopior av trovärdiga webbplatser (Dasgupta et al., 2020).

DoS- och DDoS-attacker: Detta är förkortningar av "Denial of Service (DoS)" och "Distributed Denial of Service (DDoS)". Syftet med denna attack är att begränsa behöriga användares åtkomst till resurser genom att skapa fler falska förfrågningar än vad servern kan hantera. Attackens huvudsakliga syfte är att överanstränga och överväldiga systemet, vilket i sin tur saktar ner eller till och med avbryter processen för att hantera legitima förfrågningar. Anledningen till att denna attack används är för ekonomisk vinning mot stora företag; om denna attack utförs störs deras tjänster, vilket skadar hela företaget och påverkar deras ekonomi (Sudar et al., 2020).

Missbruk av resursattacker: Det finns anställda som inte är kunniga inom cybersäkerhet, vilket leder till att de är omedvetna om hot som kan uppstå. Dessa anställda är ofta överdrivet tillitsfulla, vilket kan leda till säkerhetsintrång och ge angripare åtkomst till organisationsdata. Den mest kända typen av missbruk av resursattacker kallas "Man in the Middle (MITM)". Denna attack inträffar när en hackare placerar sig själv mellan en pålitlig klient och dess server. Ett vanligt exempel på en MITM-taktik kallas sessionkapning. Angriparen använder denna attack för att infoga sig själv i en session mellan offret och servern. Sedan ersätter angriparen offrets internetprotokoll (IP) med sin egen IP-adress för att fortsätta sessionen med servern. Under denna process avbryter angriparen datoranslutningen från offrets dator och

manipulerar även offrets sekvensnummer och organisationsinformation (Dasgupta et al., 2020).

2.7 Utmaningar som kommer med maskininlärning

I filosofin finns det tre grenar av etik: metaetik, normativ etik och tillämpad etik. Tillämpad etik är en filosofisk inriktning som har blivit starkare eftersom den kombinerar olika tillvägagångssätt för moraliskt tänkande. Enligt Philip (1985) kan affärsetik definieras som "regler, standarder, koder eller principer som vägleder moraliskt korrekt beteende och ärlighet i specifika situationer". Detta koncept kan även appliceras på digital etik, vilken omfattar "värdesystem och moraliska principer för att navigera elektroniska interaktioner". Inom ramen för AI innefattar dessa elektroniska interaktioner både människa-maskin och maskin-maskin-interaktioner. Digital etik kan ses som socialt konstruerad av teknologiska och samhälleliga moraliska värderingar (Ashok et al., 2022).

Ansvaret att tillämpa etik inom tekniska problem framträder mer. Anledningen till att teknisk etik framträder inom teknologi och varför det är viktigt att tillämpa det är för att teknologi har gett människor mer makt än de tidigare kunde ha förväntat sig. Detta innebär att människor måste göra val som de inte tidigare förväntades göra, eftersom människors handlingar förr var mycket begränsade på grund av deras svagheter. Detta har förändrats; nu, med teknologins framväxt, måste människor lära sig hur de frivilligt begränsas av sitt omdöme och sin etik (Patrick Green, n.d.).

Ett exempel på en teknik som ger människor mycket makt är AI. Det är en teknik som hjälper människor att utforska världen bättre. Människor vill bli klokare genom teknologi som AI, vilket kan hjälpa dem att fatta bättre beslut. Eftersom ny teknologi hjälper människor och samhället att bli mer effektiva på att lösa problem, måste effektivitet inte förväxlas med moralitet. Något kan vara mycket effektivt, men det är viktigt att människor utforskar etiska frågor när det kommer till teknologins framväxt (Patrick Green, n.d.).

Inom maskininlärning har det uppkommit frågor kring etik också. Maskininlärningens snabba utveckling har skapat en nödvändighet att hantera etiska problem som uppstår när ML-modeller tillämpas inom olika system. Eftersom ML-algoritmer behöver träningsdata för att fungera och utföra uppgifter, måste de använda sig av stora mängder träningsdata, vilket kan skapa etiska utmaningar. Dessa modeller kan oavsiktligt lära sig fördomar som finns i träningsdata, vilket leder till diskriminerande resultat. Detta kan i sin tur bidra till samhälleliga ojämlikheter och förstärka befintliga fördomar. Problemet har väckt frågor kring samtycke, transparens och potentiellt missbruk (Ghosh, 2023).

2.7.1 Partiskhet

Maskininlärning har många fördelar och har hjälpt många företag och organisationer att förbättra deras säkerhet. Men ingenting kan vara helt perfekt, och detta gäller också för ML. Det finns brister inom maskininlärning och en av dem är partiskhet. Hela poängen med maskininlärningsalgoritmer är att använda data för att träna algoritmen att hitta mönster. Problemet är att en algoritm kan bara vara lika bra som den data den tränas på. Det innebär att om träningsdata är partisk kommer det att påverka resultatet av algoritmen. Partiskhet inom maskininlärning innebär att maskininlärning kommer att göra samma misstag konstant på grund av data som används. Att begränsa inlärningen av en algoritm kan få allvarliga konsekvenser och leda till fördomar, stereotyper och eskalering av social ojämlikhet (Shaibu, 2023). En rättvis algoritm måste ta hänsyn till demografiska faktorer som ras, kön, socioekonomisk bakgrund och politisk inställning (Gatha, n.d.).

En algoritm får inte heller ge förmånsbehandling eller diskriminerande behandling till en specifik grupp av människor. Dessa problem började uppstå när människor insåg att det fanns obalanserade datamängder, vilket ledde till att algoritmen blev partisk, särskilt inom kritiska områden som sjukvård och rättsliga organisationer (Gatha, n.d.).

Detta är ett av de främsta etiska problemen inom maskininlärning. Det är mycket möjligt att fördomar kan införas under inkörningsperioden av en ML-algoritm (Shaibu, 2023).

2.7.2 Sekretess/Integritet

Inom datavetenskap är en stor mängd data väsentlig. Detta innebär att data kan innehålla personuppgifter, vilket har skapat en diskussion om integritet kring maskininlärning och AI. När en person har rätt till integritet innebär det att de har förmågan att hålla sin information osynlig för andra. Människor har rätt till att leva ett privatliv. När det kommer till sekretess innebär det att vara medveten om vilken information som samlas in och hur den används av olika företag. Problemet med maskininlärning är att det har introducerat nya integritetshot (Kästner, 2022).

Under de senaste åren har efterfrågan på skydd av individuell information ökat. Detta har lett till införandet av regler och förordningar som GDPR, HIPAA och CCPA. Dessa regler ger människor en förståelse för hur data måste samlas in från användare och hanteras av tjänsteleverantörer med hänsyn till dess syfte. Dessa regler betonar också hur viktigt det är med informerat samtycke. Om en användare tillåter insamling av sina data måste de informeras om hur dessa data kommer att användas, lagras och eventuellt delas vidare. Facebook-Cambridge Analytica-skandalen är ett exempel som väckte oro bland människor. Personuppgifter hade använts för att rikta sig till specifika demografier och driva rekommendationer för politiska och ekonomiska vinster. Detta är ett exempel på en oetisk användning av personuppgifter (Gatha, n.d.).

Företag måste ta ansvar när det kommer till att diskutera hur data används inom deras verksamhet och hur de kan skydda kunders personliga information (De Harder, 2023).

3 Problemområde

Under senare år har maskininlärning (ML) och djupinlärning (DL) fått ökad uppmärksamhet och användning inom en rad olika branscher, inklusive hälsovård, finans och detaljhandel (Sharifani & Amini, 2023). Dessa tekniker utnyttjar algoritmer för att lära sig från historiska data och har visat sig vara värdefulla för att förutse och hantera olika typer av problem. Trots dessa fördelar har det dock uppstått diskussioner kring de etiska aspekterna av ML, särskilt när det gäller frågor som datasekretess, transparens och partiskhet (Sharifani & Amini, 2023).

Utmaningar som partiskhet och rättvisa är redan välkända inom ML och det finns en generell förståelse för konsekvenserna av ML. Detta förhindrar inte att dessa konsekvenser och utmaningar utvecklas i takt med framsteg inom teknik och applikationer. Barbierato och Gatti (2024) skriver att det är viktigt med en kontinuerlig diskussion om de utmaningar som följer med ML inom ML-samhället. Författarna betonar att bara för att dessa frågor och utmaningar redan är kända, minskar det inte deras betydelse. En öppen och pågående dialog om dessa problem kan förbättra praxis, policyer och utbildningar inom området (Barbierato & Gatti, 2024).

Diskussionen om konsekvenserna och utmaningarna med ML borde spridas vidare, vilket kommer att leda till en fördjupad förståelse för hur ML påverkar olika sektorer. Denna diskussion börjar nu uppkomma inom cybersäkerhetsområdet. Med den ökande användningen av ML-modeller inom cybersäkerhet för att upptäcka och motverka skadlig kod, nätfiske, skräppost och bedrägeri (Bharadiya, 2023), uppstår ett behov av att förstå de sociala och etiska konsekvenserna av dessa teknologiers implementering inom cybersäkerhetssystem. Cyberbrott utgör inte bara en betydande global risk med långvariga effekter på samhället, utan de påverkar också miljontals människors liv och kan resultera i omfattande ekonomiska förluster och förtroendeförlust för organisationer (Monteith et al., 2021; Elluri et al., 2023).

Därför är det viktigt att inte bara fokusera på att utveckla mer avancerade ML-modeller för att bekämpa cyberbrott, utan också att undersöka hur dessa tekniker kan implementeras på ett etiskt och socialt ansvarsfullt sätt (Elluri et al., 2023).

Elluri et al., 2023, påpekar att det behövs framtida forskning för att adressera utmaningar såsom potentiell fördom i utbildningsdata och integritetsproblem. Elluri et al., 2023, betonar vikten av etiska överväganden när dessa maskininlärningsmodeller används inom cybersäkerhet. Elluri et al., 2023, presenterar flera forskningsområden som är värda att utforska inom forskningsvärlden, en av dem är konsekvenserna som följer av att använda maskininlärningsmodeller för förutsägelser av cyberbrott.

Genom att noggrant undersöka och förstå de sociala och etiska implikationerna av ML-modeller inom cyberbrottsbekämpning går det att säkerställa att dessa verktyg används på ett ansvarfullt och etiskt sätt.

3.1 Problem/fråga

Syftet med denna studie är att förstå hur samhället kan bli påverkad av ML-modeller specifikt inom cybersäkerhet. Det finns inte mycket forskning inom det etiska och social konsekvenserna av ML inom cyberbrottbekämpning. Med detta i grund har detta examensarbetet arbetat mot följande frågeställning:

- Vilka sociala och etiska konsekvenser uppstår vid implementeringen av ML-modeller inom cyberbrottbekämpning?

3.2 Avgränsningar

Det finns olika metoder för att förutsäga cyberbrott, fokuset i den här rapporten kommer att ligga på maskininlärning. I diskussioner om ML tas AI upp väldigt ofta, i denna studie finns det förklaring på hur dessa två korrelerar med varandra, dock kommer det inte finnas en fördjupning inom AI och dess konsekvenser.

3.3 Förväntat resultat

Fokuset på studien kommer att vara på att intervjua människor med kunskap inom området cybersäkerhet. Litteraturen som har läst har inte avgränsat från varandra väldigt mycket, samma problemområde tas upp. Det som förväntas är att respondenter kommer hålla med litteraturen och ge ungefär samma svar som diskuteras i litteraturen. Studien kommer utgå från att vara öppensinnad för alla typer av svar.

4 Metod

4.1 Forskningsdesign

Den mest lämpliga forskningsmetoden för att besvara frågan ”*Vilka sociala och etiska konsekvenser uppstår vid implementeringen av ML-modeller inom cyberbrottbekämpning?*” är en kvalitativ ansats. Denna metod används för att få insikt i människors erfarenheter och perspektiv gällande olika ämnen. Fokus ligger ofta på specifika individer, händelser och sammanhang. Det är en ideografisk analysstil (Hammarberg et al., 2016; Gerring, 2017). Data som samlas in med denna metod är inte något som ska mätas kvantitativt. Att göra en kvalitativ forskning innebär att utföra smågrupps diskussioner eller individuella intervjuer för att besvara forskningsfrågan (Hammarberg et al., 2016). Detta ger information som är mycket djupare och mer grundad i verkliga problem.

Anledningen till att kvalitativ forskning ger bra resultat är att det är möjligt att ställa öppna frågor. Svaren som samlas in från öppna frågor går inte att kvantifiera. På grund av detta är en kvalitativ forskningsdesign inte lika linjär som kvantitativ forskning, som är mer baserad på räkning och mätbarhet (Tenny, 2022). Kvalitativ forskning är mycket effektiv eftersom den förklarar processer och mönster i människors beteenden som inte går att kvantifiera. Saker som människors erfarenheter, attityder och beteenden är svåra att fånga genom en kvantitativ undersökning. Genom en kvalitativ ansats kan deltagarna själva förklara hur, varför och vad de tänker under intervjun (Tenny, 2022).

I denna studie kommer känsliga frågor att tacklas, såsom fördomar, utsatta grupper, partiskhet och mer. Det är komplicerat att få nyanserad data genom en kvantitativ ansats. För att få nyanserade svar var det mest lämpligt att genomföra intervjuer, där det var möjligt att ställa följdfrågor om det behövdes för att få djupare svar från deltagarna.

4.2 Fältstudie

För att undersöka frågan som ligger till grund för denna studie kommer en fältstudie att genomföras. En fältstudie är en forskningsstrategi där forskare genomför observationer och samlar in data i en naturlig miljö. Det finns olika metoder för att samla in data som är tillgängliga när en fältstudie utförs. Dessa metoder är intervjuer, observationer och att interagera med människor i deras miljö (Dovetail Editorial Team, 2023).

För denna studie valdes fältstudie som den mest lämpliga forskningsstrategin. Att samla information från olika organisationer som arbetar inom cybersäkerhet var viktigt för att få mer nyanserade svar. Det var också betydelsefullt att hitta personer med olika åsikter för att se om det finns nya aspekter att undersöka och analysera. Att intervjua enbart ett

företag kan leda till grupptänkande inom organisationen. Genom att intervjua människor från olika områden och företag kunde ett mer intressant och varierat resultat uppnås.

4.3 Datainsamling

Att utföra kvalitativ forskning innebär en viss flexibilitet, öppenhet och lyhördhet för att samla in och analysera data jämfört med kvantitativ forskning. Denna ansats tillåter forskare att arbeta mer iterativt istället för att följa strikta steg. Detta innebär att forskare kan gå fram och tillbaka mellan att samla in data och analysera den (Busetto et al., 2020). Nya idéer kan uppstå som kan leda till ändringar i planen. Arbetet slutförs när forskarna anser att de inte kan få mer information eller när de är nöjda och känner att de har relevant data för att besvara sina frågor (Busetto et al., 2020).

Det finns flera olika metoder för att samla in data inom kvalitativ forskning. En forskare kan samla in data genom litteraturstudier, observationer, intervjuer och fokusgrupper (Busetto et al., 2020). I denna studie har två metoder använts: litteraturstudier och intervjuer.

Litteraturstudie: En litteraturstudie innebär att en forskare granskar tidigare litteratur. Detta kan vara personliga dokument, icke-personliga dokument som arkiv, årsredovisningar, riktlinjer, policydokument, dagböcker eller brev (Busetto et al., 2020). I denna studie var det nödvändigt att granska tidigare material om de utmaningar som förekommer med användningen av maskininlärning. Det var viktigt att förstå vad den generella diskussionen om maskininlärning var för att skapa en bra grund för att formulera frågor som relaterar till materialet. För att hitta relevanta artiklar användes sökmotorn Google för populära artiklar, medan Google Scholar och Högskolan i Skövdes fulltextdatabas användes för vetenskapliga artiklar. Högskolan i Skövdes fulltextdatabas hade inte tillräckligt med relevanta artiklar för studien. Därför var Google Scholar den mest lämpliga databasen för att hitta referenser, eftersom den erbjuder ett större utbud av artiklar om det aktuella ämnet. Både populärvetenskapliga och vetenskapliga artiklar behövde vara skrivna under de senaste åren, mellan 2019 och 2024, eftersom teknologin utvecklas snabbt och ny information om olika aspekter av teknologi ständigt publiceras. För att hitta dessa artiklar användes söktermer som:

- ethical and social challenges of machine learning
- machine learning cyber security
- machine learning cybersecurity ethics

Dessa är några exempel på söktermer som användes för att hitta referenserna som användes i denna studie.

Intervju: Intervjuer är en bra datainsamlingsmetod eftersom de ger forskare en djupare insikt i en persons upplevelser, åsikter och motivationer. Det finns tre olika sätt att genomföra en intervju: semistrukturerad, öppen eller strukturerad. I en strukturerad intervju används frågeformulär, medan en öppen intervju innebär att prata fritt, som vid en självbiografisk intervju. I denna studie kommer semistrukturerade intervjuer att användas. Denna typ av intervju kännetecknas av öppna frågor och användning av en intervjuguide där frågorna är definierade (Busetto et al., 2020).

Fördelarna med en kvalitativ intervju är flexibiliteten; ibland kan vissa frågor hoppas över om de inte är nödvändiga eller om tiden rinner ut. Ett problem som kan uppstå är att intervjupersonen inte vill svara på vissa frågor av personliga skäl (Busetto et al., 2020). Under intervjun ställdes åtta frågor, för att inte dra ut på tiden och för att ställa frågor som är rakt på sak. Frågorna behandlade ämnen som etiska och sociala utmaningar med ML inom cybersäkerhet, integritet, dataskydd samt riktlinjer och åtgärder. Inspirationen för frågorna till intervjun kom från tidigare litteratur. Detta är ett exempel på en av frågorna som respondenterna svarade på:

- Vilka sociala utmaningar kan uppstå när ML-modeller implementeras i cyberbrottsbekämpning? Hur påverkar det samhället?

En annan fördel med kvalitativa intervjuer är att fokus ligger på att ha ett interaktivt samtal med respondenten. Därför sker intervjuer sällan i skriftlig form, utan intervjuaren är aktivt engagerad i samtalet med respondenten. Intervjuer kan vara ljud- eller videoinspelade; i vissa situationer har intervjuaren dock inte tillåtelse att spela in något, och då får de nöja sig med att anteckna svaren (Busetto et al., 2020). För denna studie var det inte möjligt att intervjua människor fysiskt, därför skedde samtalen digitalt. Detta var positivt eftersom det var möjligt att spela in samtalen, vilket underlättar transkriberingen av konversationerna.

4.4 Urval och deltagare

För denna studie var det viktigt att få tag på personer med god förståelse för ämnet. Det var nödvändigt att kontakta personer som har bra kunskap om cybersäkerhet. Många olika företag kontaktades; specifikt kontaktades 8 företag som arbetar med cybersäkerhet. Det var lämpligt att ha företag som faktiskt arbetar i denna typ av miljö eftersom det skulle gynna studien. En annan metod för att hitta deltagare var att läsa artiklar som behandlade samma eller liknande ämnen. Två författare till dessa artiklar kontaktades för en intervju, en av dem var svensk och den andra var en internationell författare. Detta var urvalsprocessen för studien. Dock uppstod problem med hur många som var villiga att delta. Detta leder till att reliabiliteten och validiteten av studien kan

ifrågasättas eftersom, av alla som kontaktades, var det endast två personer från två olika företag som var villiga att ställa upp på en intervju. Litteraturen kommer att vara ryggraden i denna studie och deltagarnas synpunkter kommer att jämföras med det som finns i litteraturen.

4.5 Analysmetod

Att arbeta kvalitativt innebär att samla in data som inte är sifferbaserat. Kvalitativa data tar ofta formen av intervjuanteckningar, dokument och öppna enkätsvar. Detta är dock inte dess enda former; det kan även innefatta tolkning av bilder och videor (Warren, 2023).

Vid analys av kvalitativa data används inte statistik som vid kvantitativa metoder. Istället är fokus riktat mot ord, beskrivningar, begrepp eller idéer, vilket skiljer sig från den kvantitativa ansatsen där siffror och statistik står i centrum (Warren, 2023).

Det finns sex metoder för dataanalys inom forskning: innehållsanalys, narrativ analys, diskursanalys, tematisk analys, grundad teori (GT) och tolkningsfenomenologisk analys (IPA) (Warren, 2023). I denna studie kommer en tematisk analys att utföras. Att tillämpa tematisk analys för dataanalys innebär att, efter att data samlats in, söka och identifiera mönster och olika teman som återkommer i datamängden. Det är ett sätt att beskriva data där tolkning ingår i processen för att välja koder och skapa teman (Kiger & Varpio, 2020).

Tematisk analys utmärker sig som en kraftfull metod genom dess förmåga att bistå forskare i att förstå olika upplevelser, tankar eller beteenden som framträder inom en datamängd. Det centrala i tematisk analys är att söka efter likheter i upplevelser och gemensamma betydelser. Denna analysmetod lämpar sig mindre väl för att undersöka unika upplevelser från enskilda dataobjekt (Kiger & Varpio, 2020).

4.6 Forskningsetiska överväganden

Forskning kan involvera många olika parter i undersökningen, därför är det viktigt att tänka på de etiska aspekterna av forskning. I Sverige är det enligt lagen (2019:504) krav på att forskare måste vara ansvarsfulla och följa god forskningssed i sitt arbete. Om detta inte sker och forskare ignorerar reglerna kan det leda till felaktiga resultat och skada människor, djur och naturen. Detta kan dessutom påverka förtroendet mellan allmänheten och forskarna negativt (Vetenskapsrådet, 2023).

Forskning är en viktig del av samhället och hjälper människor och världen att utvecklas. Däremot är det nödvändigt att följa riktlinjerna för att uppnå forskning som är pålitlig och av hög kvalitet. Dessa regler betecknas som forskningskravet, vilket innebär att kunskap utvecklas och förbättras över tid samtidigt som forskare förfinar sina

forskningsmetoder. Det är viktigt att ta hänsyn till människors privatliv och att deras personliga uppgifter skyddas. Forskare måste ansvara för att inte orsaka psykisk eller fysisk skada, förnedring eller kränkning för de personer som deltar i deras studie. Detta kallas för individskyddskravet. Individskyddskravet kan brytas ner till fyra olika krav (Vetenskapsrådet, 2002).

Informationskravet: *"Forskaren skall informera uppgiftslämnare och undersökningsdeltagare om deras uppgift i projektet och vilka villkor som gäller för deras deltagande. De skall därvid upplysas om att deltagandet är frivilligt och om att de har rätt att avbryta sin medverkan. Informationen skall omfatta alla de inslag i den aktuella undersökningen som rimligen kan tänkas påverka deras villighet att delta." – Vetenskapsrådet*

Studien uppfyllde detta krav genom att informera deltagare via e-post om vad som skulle diskuteras. I e-posten framhölls vad ämnet för uppsatsen var och vilken fråga som skulle besvaras. När respondenter accepterade att delta i en intervju skickades intervjufrågorna därefter. Detta var viktigt för att deltagarna skulle vara förberedda på vad som skulle frågas och känna sig bekväma med vad som skulle diskuteras i förväg. Innan intervjun startade, betonas det igen vad studien handlar om.

Samtyckeskravet: *"Forskaren skall inhämta uppgiftslämnares och undersökningsdeltagares samtycke. I vissa fall bör samtycke dessutom inhämtas från förälder/vårdnadshavare (t.ex. om de undersökta är under 15 år och undersökningen är av etiskt känslig karaktär)." - Vetenskapsrådet*

Respondenterna informerades om att de hade frihet att delta och att de hade rätt att säga ifrån om de inte kände sig bekväma med något. Om något i intervjun skulle kännas olämpligt efteråt, skulle det vara forskarens ansvar att ta bort det.

Konfidentialitetskravet: *"All personal i forskningsprojekt som omfattar användning av etiskt känsliga uppgifter om enskilda, identifierbara personer bör underteckna en förbindelse om tystnadsplikt beträffande sådana uppgifter." - Vetenskapsrådet*

Deltagarnas personliga information kommer inte att ingå i denna studie. I början av intervjun togs anonymitet upp, och det diskuterades att alla som deltar i denna studie kommer att vara anonyma. Deltagarna representerar sig själva och inte företagen de arbetar på. För denna studie var det inte viktigt att inkludera information om deltagarnas namn, ras, ålder, kön och adress. Det som var viktigt var deras åsikter kring olika frågor.

Nyttjandekravet: *"Uppgifter om enskilda, insamlade för forskningsändamål, får inte användas eller utlånas för kommersiellt bruk eller andra icke-vetenskapliga syften." – Vetenskapsrådet*

All information som har samlats in från intervjuerna kommer endast att användas för denna studie. Efter studiens avslutning kommer allt material att raderas eftersom det har uppfyllt sitt syfte och inte längre är nödvändigt att behålla.

5 Analys

I detta kapitel ska materialet som har samlats in från intervjuerna granskas. Varje fråga kommer att behandlas i ett eget avsnitt, där data från intervjuerna och relevant litteratur jämförs för att se om respondenternas svar stämmer överens med vad som framkommit i tidigare forskning.

5.1 Materialpresentation

Data som kommer att användas för analyskapitlet har samlats in genom kvalitativa intervjuer. Deltagarna som var villiga att intervjuas arbetar inom cybersäkerhetsområdet. Båda deltagarna har ledarpositioner i sina företag, vilket tyder på att deras åsikter har hög validitet.

Data som har samlats in genom intervjuer är från två individer. Intervjuerna genomfördes under samma månad och varje intervju varade inte längre än 30 minuter. Samtalen spelades in via Zoom. Transkriberingen av båda intervjuerna omfattade totalt 14 sidor.

Respondent	Yrkestitel	Kunskapsnivå inom ämnet	Intervjulängd	Utförandet
R1	Ledarposition	fördjupad	29 minuter	Zoom
R2	Ledarposition	fördjupad	29 minuter	Zoom

Genom att analysera den insamlade data är det möjligt att identifiera mönster, vilket gör det möjligt att kategorisera data i olika frågor som sedan kommer att jämföras med litteraturen.

5.2 Fördelarna av att använda ML-modeller för cybersäkerhet

Den första frågan syftar till att ge en förståelse för respondenternas åsikter om maskininlärning inom cybersäkerhet och om de anser att det finns fördelar med att använda tekniker som ML i detta område. Frågan ska ge läsaren en insikt i den allmänna åsikten kring användningen av ML i cybersäkerhet från personer som arbetar inom detta fält.

Litteraturen visar att det finns positiva effekter av att kombinera ML med cybersäkerhet. Maskininlärning är ett verktyg som uppskattas inom olika branscher på grund av dess förmåga att svara på verkliga frågor, vilket är en anledning till dess popularitet. Trenden

att använda maskininlärning har även påverkat cybersäkerhetsområdet, där detektionssystem uppgraderas med ML-komponenter (Apruzzese et al., 2018).

I nuläget är AI och ML verktyg som är nödvändiga inom cybersäkerhetsområdet på grund av den pågående kampen mot cyberhot. Detta har lett till en stor förändring i hur människor försvarar sig mot hot jämfört med de traditionella metoder som inte var lika adaptiva och proaktiva. Förändringen att inkorporera ML och AI för att bekämpa cyberbrott kom från att traditionella cybersäkerhetsåtgärder var beroende av signaturbaserade detektionssystem och statiska regeluppsättningar (Shah, 2022). Problemet som uppstod var att teknologin utvecklas och blir mer avancerad med tiden, vilket leder till att traditionella cybersäkerhetsåtgärder inte är tillräckliga för att hantera moderna hot. Att integrera AI och ML i cybersäkerhet har lett till en mer lovande framtid där cyberförsvar är mer effektivt och erbjuder smidighet, skalbarhet och förutsägbarhet, vilket är nödvändigt för att hantera och motverka cyberattacker i dagens samhälle (Shah, 2022).

Maskininlärningsalgoritmer använder stora mängder märkt och omärkt data, och algoritmerna kan upptäcka komplicerade mönster och ovanliga händelser som kan indikera skadliga aktiviteter inom komplexa datasamlingar (Shah, 2022). R1 hade liknande tankar kring fördelarna med ML inom cybersäkerhet.

> *Jag tycker att det finns många fördelar med att använda ML-modeller. Det första är att man kan upptäcka anomalier, alltså man kan upptäcka ovanliga aktiviteter. Det kan man göra i nätverkstrafik och man kan upptäcka det i användarbeteende eller i olika loggar och då kan man få hjälp och få indikationer om potentiella säkerhetshot. Det kan ju vara intrång eller skadlig aktivitet...sen det andra är ju också att genom att analysera historiska data så kan ju ML -modeller hitta mönster som är förknippade med kända hot och tidigare attacker och då kan man ju också upptäcka liknande attacker i framtiden… så jag tänker att ML-modeller för att bekämpa cyberbrott det hjälper till med både noggrannhet och effektivitet. Det gör ju också att det går snabbare så jag tänker att det är ett jättebra komplement till människan.* - (R1)

R1 betonar effektiviteten i att hitta ovanliga aktiviteter med hjälp av ML-modeller, vilket kan hjälpa datorer att agera snabbt mot olika typer av hot. ML-modeller inom cybersäkerhet kan använda mängder av historiska data för att granska och hitta mönster som är förknippade med tidigare attacker. Detta, som R1 säger, hjälper till att upptäcka framtida attacker. R2 har samma åsikter, vilket återspeglas i deras synpunkter på fördelarna med ML-modeller inom cybersäkerhet.

> *Som jag nämnde, jag är inte specialist på alla typer av ML-modeller, men ett tillämpningsområde som jag själv har jobbat med och som är en fördel handlar om anomalidetektion alltså att upptäcka avvikelser och det är väldigt viktigt när det handlar om till exempel att analysera trafik inom nätverk och inom system där bara de*

betrodda användarna ska ha tillgång. Traditionellt har man byggt säkerhet genom att försöka ha höga murar genom traditionella brandväggar och antivirus och så vidare. Men ett problem med det är att om någon tar sig förbi muren eller kanske är en insider så att man har tillgång men man avser att missbruka tillgången till en viss digital miljö så är man väldigt dålig på att upptäcka det. Men med hjälp av maskinlärande metoder till exempel så kan man göra avvikelsedetektion kopplat till beteenden eller signaler på nätverket eller på ett digitalt system som hänger ihop med hot eller risk så att om man har en modell som dels kan mäta hur vanligt eller avvikande ett beteende är för en viss användare eller en viss enhet och kan kombinera den mätningen med annan kunskap till exempel om hur angripare brukar bete sig på nätverk eller i digitala system som de har tagit sig in i och tagit sig förbi gränsskyddet så får man en mycket bättre detektionsförmåga och en bättre träffsäkerhet i detektionsförmågan. - (R2)

Både R1 och R2 tar upp anomalidetektion som ett exempel på varför ML-modeller är en effektiv metod för att bekämpa cyberbrott. Respondenterna har överlag en positiv inställning till att använda ML-modeller inom cybersäkerhet. Även om de kan se brister inom detta område, vilket de tar upp i de kommande frågorna, anser de ändå att ML har många positiva egenskaper. Detta stämmer överens med litteraturen; ML har hjälpt många olika branscher och ökat effektiviteten, och användningen av ML i cybersäkerhet är nödvändig för att bemöta dagens moderna hot.

5.3 Sociala utmaningar av att använda ML-modeller i cyberbrottsbekämpning

Användningen av ML inom cybersäkerhet kommer med många fördelar, men det medför också problem och utmaningar. Ett stort problem som uppstår med ML-modeller är att de ofta använder stora mängder känslig data, vilket har lett till oro och diskussioner kring personlig integritet (ShardSecure, 2024). I dagens samhälle är data centralt för nästan alla företag och organisationer. Detta beror på att företagen har insett värdet av att samla in, omvandla och använda både interna och externa data för att fatta beslut, vilket i sin tur ger dem en konkurrensfördel. Tidigare var detta inte möjligt eftersom resurser som hårdvara och mjukvara var för dyra eller ännu inte existerade, men nu finns det ett överflöd av dessa resurser. Som en följd av detta finns det idag mer varierande data som existerar i större volymer än tidigare (Monteiro et al., 2021).

Många länder har ännu inte implementerat dataskyddslagar som är specifika för AI och ML, vilket kan vara problematiskt, särskilt med tanke på att teknologin blir alltmer avancerad. För att kunna anpassa sig till samhällets utveckling krävs att data används, vilket kan leda till att känslig information om en persons geografiska plats och identitet kompromitteras (ShardSecure, 2024). AI och ML är bra verktyg för att stödja samhället, men det kommer inte utan risker såsom integritet, rättvisa och jämlikhet (ShardSecure, 2024). R1 anser att en av de sociala utmaningarna med att använda ML-modeller för cybersäkerhet är människors personliga information: *”Det finns ju jättemycket viktiga*

sociala krav som man måste ta hänsyn till och det handlar ju mycket om sekretess och dataskydd tänker jag. Det är ju jätteviktigt att den personliga integriteten och vår personliga information respekteras."

En stor anledning till att integritet är i människors medvetande är på grund av Facebooks dataintegritetsskandal som inträffade under 2018. Denna skandal ledde till en ökad uppmärksamhet kring integritetsutmaningar. Dessa samtal återupptas till följd av smart teknologi, och Big Data-revolutionen är en stor orsak till användningen av dessa nya tekniska verktyg såsom maskininlärningstekniker (Liu et al., 2021). Säkerhet är en viktig del av ett system som fungerar optimalt. Det är utvecklarnas ansvar att säkerställa att sekretessen för data eller känslig information om användare är en integrerad del av säkerhetslösningen (Himthani et al., 2020).

Även om ML-tekniker har problem med integritet är det möjligt att använda dessa tekniker för att skapa modeller som hjälper till att säkerställa integritet. Himthani et al., 2020, föreslår att det finns olika maskininlärningstekniker som kan användas för att utveckla modeller som kan hantera Big Data, såsom artificiella neuronnätverk (ANN), SVM, fuzzy logic, Bayesiska klassificerare, rough set theory, regression och andra. De skriver att dessa tekniker kan tillämpas på Big Data både för analys och för att bevara integritet. Problemet som uppstår med detta förslag är att det kan fungera om det används begränsade datamängder, men utan detta kan det uppstå komplikationer. Traditionella ML-modeller är utformade för att kräva stora mängder data för att kunna fungera, vilket gör förslaget att använda begränsade data komplicerat att genomföra och bedöma dess effektivitet.

Ett annat problem som uppstår med Big Data är dess höga kostnader för hantering, lagring, bearbetning, frisläppande och analys av data i exabyte-storlek. Detta leder till att företag måste investera i välkonfigurerad infrastruktur och mjukvaruverktyg (Himthani et al., 2020).

Höga kostnader av välkonfigurerad infrastruktur och mjukvaruverktyg leder till en annan social utmaning av ML-modeller, organisationer som inte har råd att använda dessa avancerade tekniska innovationer som ML. *"Men sen tror jag också så här att om man tänker att det kommer säkert finnas företag och organisationer i framtiden som inte har möjlighet att använda machine learning i samma utsträckning som andra. Man kanske har bristande resurser och sämre investeringsförmåga och sådana här saker och då finns det ju en potential att den typen av organisationer och företag också skulle kunna bli sårbara på grund av att man inte har möjlighet att använda de här modellerna…det finns ju flera aspekter tänker jag på konsekvensen av det där då just ur ett socialt och etiskt perspektiv*."-R1.

Den senaste tekniska boomtiden för AI har hjälpt människor att förstå hur kostsam det är att både utveckla och underhålla programvara. Kostnaderna kan vara höga för de

företag som utvecklar dessa teknologier och för företag som använder AI för att driva sin egen programvara. Den höga kostnaden för ML har lett till att riskkapitalister vänder sig mot företag som redan är värda biljoner, såsom Microsoft, Google och Meta. Dessa företag har möjlighet att använda ML för att utveckla ett försprång inom tekniken. Detta leder till att mindre företag inte kan hålla jämna steg med de stora företagen (Vanian & Leswing, 2023). Även om små företag har råd att använda ny teknologi, har de inte samma starka ramverk för cybersäkerhet som större företag. Mycket av dagens teknologi är beroende av att samla in och bearbeta mängder av känsliga data, vilket kan skapa nya datasäkerhetsproblem (TechCurators, 2023).

ML-modeller har fler utmaningar än bara integritet. Den ökande användningen av AI och ML har skapat diskussioner om opacitet och partiskhet som följer med dessa teknologier. Opacitet i detta sammanhang innebär svårigheten för en användare att förstå hur en algoritm har kommit fram till ett visst beslut baserat på indata. Partiskhet innebär fördomar och favorisering av individer eller grupper baserat på vissa egenskaper (Franzen et al., 2021). Att skapa en ML-modell innebär att människor behöver mata algoritmen med data. Människor själva är defekta och har fördomar som de kanske inte är medvetna om, vilket kan leda till skapandet av en modell som också är partisk. Resultatet blir en modell som gör förutsägelser som är partiska (Milaninia, 2021). Processen att samla in data kan leda till att data överrepresenterar en viss grupp eller händelse medan andra grupper och händelser är underrepresenterade. Ett exempel på överrepresentation är bildigenkännings ML-modeller. En datavetare vid University of Virginia skapade en bildigenkänningsmodell som associerade bilder av kök med kvinnor. Datavetaren tränade algoritmen med bilder som visade matlagning och städning, där kvinnor ofta syntes. Detta ledde till att modellen associerade dessa specifika aktiviteter med kvinnor, vilket förstärkte könsstereotyper. Sådana här modeller och andra Big Data-analysverktyg kan vara mycket felaktiga och leda till ökad fördomsfullhet (Milaninia, 2021).

De två problemen, partiskhet och opacitet, har lett till ett behov av transparens i hur dessa algoritmer utvecklas och vilken data som används under träningsstadiet för att förhindra eller mildra negativa konsekvenser (Franzen et al., 2021). Dessa konsekvenser påverkar alla områden där ML-algoritmer används, vilket gör behovet av transparens ännu större. Att övervaka datakvaliteten blir allt viktigare vid användning av ML-modeller. När det gäller opacitet i ML kan detta problem uppstå i olika former och nyligen har problemet med black box-effekten framträtt som en utmaning (Franzen et al., 2021). Ett black box-system innebär att människor kan observera indata och utdata men inte den interna processen. Anledningen till denna utmaning är att ML-algoritmer ofta är mycket komplexa, vilket gör det svårt för människor att förstå deras interna beslutsprocesser. Dessa algoritmer är skapade för att uppnå bästa möjliga prestanda, vilket gör att fördelarna ofta överväger nackdelarna (Franzen et al., 2021). Black box-effekten kan dock bidra till ökade fördomar och orättvisor som kan påverka människors liv. Franzen et al. (2021) skriver att det är rekommenderat att undvika användningen av

dessa algoritmer vid beslut med höga insatser inom områden som rättvisa, sjukvård och arbetsmarknaden.

Akademisk forskning visar att algoritmer som skapas med ofullständiga eller partiska data kan upprepa eller förstärka samma fördomar när de genererar resultat. Callahan (2023) genomförde en intervju med Ngozi Okidegbe, professor vid Juridiska institutionen och fakulteten för datavetenskap. Okidegbe förklarar att användningen av ML-algoritmer inom brottsbekämpning kan få allvarliga konsekvenser, särskilt för minoriteter (Callahan, 2023). Under intervjun ger Okidegbe ett exempel där en domare får en algoritmiskt genererad återfallsriskpoäng som en del av en rapport om en dömd brottsling. Poängen används för att informera domaren om hur sannolikt det är för en brottsling att begå ett annat brott i framtiden. Om domaren observerar att poängen är hög kan detta leda till fördomar om att personen är en återfallsförbrytare. I sådana situationer kan domaren ta hänsyn till poängen, vilket kan leda till längre fängelsestraff för en person med hög återfallsrisk (Callahan, 2023).

Cabrera et al. (2020) påpekar att en av de stora utmaningarna inom maskininlärning är frågan om rättvisa. Eftersom olika populationer har varierande basnivåer av vad rättvisa innebär, blir det matematiskt omöjligt att uppfylla alla definitioner av rättvisa. Detta problem formaliserades genom Arrows omöjlighetsteorem, vilket innebär att det är omöjligt att tillgodose olika rättviseaspekter samtidigt. Cabrera et al. (2020) hänvisar till två forskningsartiklar som visar på problemet när olika grupper använder olika fördelningar av sina etiketter. Detta blir problematiskt eftersom det är svårt att garantera de tre grundläggande aspekterna av rättvisa: balans i positiva klasser, balans i negativa klasser och en korrekt justerad modell. Datavetare måste därför fatta beslut om vilka mätningar av rättvisa som är mest lämpliga när de konstruerar en modell (Cabrera et al., 2020).

Dessa konsekvenser blev tydligare i verktyget COMPAS, som är ett system som används för att förutsäga sannolikheten för återfall hos personer som överväger att släppas mot borgen. Organisationen ProPublica publicerade en artikel om COMPAS-systemet som visade att systemet var partiskt mot svarta individer. Systemet tenderade att ge svarta individer en högre riskbedömning jämfört med vita individer, även när de hade samma grundläggande risknivåer. COMPAS är ett exempel på ett system som inte kan uppfylla alla definitioner av rättvisa samtidigt (Cabrera et al., 2020).

Respondenten delade också denna oro när det gäller de sociala utmaningarna som följer med ML-modeller: ”*om vi håller oss till cybersäkerhet och säkerhetsdimensionen så finns det kanske risker för att om man överlåter en del av beslutsfattandet eller en del av produktionen av beslutsunderlag för beslutsfattande när det gäller till exempel rättsutövning och brottsbekämpning och om dessa modeller innehåller felaktigheter kan det medföra risker för rättssäkerheten hos dem som då blir misstänkta för brott.*” - R2

5.4 Etiska dilemman av att använda ML-modeller i cyberbrottsbekämpning

Ett av de etiska dilemman som respondenten tog upp var problemet med transparens och hur viktigt det är för företag som skapar dessa algoritmer att vara transparenta. R1 sa: *"men sen tror jag också att det är jätteviktigt att när man tränar de här algoritmerna så är det viktigt att det är tydligt vem som har ansvar och att det är transparent också så att det finns en tydlighet för vad man vill uppnå med de här algoritmerna…"*

Transparens innebär i vilken grad något är tydligt. För de som utvecklar informations- och kommunikationsteknologi innebär det tydlighet kring hur information skapas, bearbetas, hanteras, överförs, lagras och delas (Rouse, 2023). Om företag som skapar dessa system är transparenta kan det hjälpa både företag och intressenter. Problem som kan uppstå identifieras och företag kan komma på förbättringar. Denna process av att vara transparent under hela utvecklingsprocessen säkerställer ansvarsskyldighet. Det hjälper intressenter att gå tillbaka till grundorsaken till ett problem och identifiera individer som var ansvariga för att åtgärda det (Rouse, 2023). R1 sa: *"… [det är] väldigt lätt även inom cybervärlden att det kan leda till diskriminering eller orättvis behandling baserat på en mängd olika faktorer och det är ju också jätteviktigt att datan som används är korrekt så att man inte tränar de här algoritmerna på data som i grunden kanske är snedvridet på olika sätt. Så jag tror också att det här med transparens och tydligt ansvar är jätteviktigt för att det här ska kunna hanteras på rätt sätt ur ett samhällsperspektiv."*

En stor fördel med att vara transparent är att det motverkar korruption, slöseri och maktmissbruk. Det är betydligt svårare för individer och organisationer att använda sig av korrupta metoder om det finns system med kontroller och balanser. Detta kan leda till att företag tänker mer etiskt och arbetar mer ansvarsfullt (Rouse, 2023).

Inom cybersäkerhetsvärlden är bristen på transparens mellan leverantörer och användare en faktor som har hindrat utbyggnaden av AI/ML inom cybersäkerhet. Algoritmer blir mer komplicerade, vilket leder till att användare spekulerar kring hur algoritmerna kommer fram till vissa resultat (Chamberlain, 2023). Denna brist på transparens kan resultera i system som inte fungerar korrekt och som har inbyggda fördomar (Bharadiya, 2023). Leverantörer är ofta inte öppna med att ge detaljer om sina produkters funktioner eftersom de är rädda för att avslöja konfidentiell information och skydda sina immateriella rättigheter. Detta leder till att användare av dessa system inte litar på leverantörerna och istället förlitar sig på äldre tekniker som de är bekanta med (Chamberlain, 2023).

En annan etisk fråga som respondenterna tog upp var dataskydd och sekretess. R1 ” […] *men det är ju precis det här som jag sa att de här etiska kraven är jätteviktiga kring och just dataskydd. Jag tänker lite grann på samma tema som GDPR att varje person måste ju äga sin egen information. Det är jätteviktigt för att det ska bli rättvist och rättssäkert också och att också säkerställa att de algoritmer och metoder som används verkligen är*

rättvisa. Det är också viktigt att använda data som har samlats in på ett lagligt sätt och etiskt sätt när man tränar de här algoritmerna så att man inte använder data som man kanske har samlat in utan att någon har gett sitt samtycke eller liknande. Det är väl det här som känns som att det inte finns tydliga regler kring än."

GDPR, General Data Protection Regulation, är regler som gäller för EU-medborgare oavsett var de befinner sig i världen. Integritet genom design är en central princip, vilket innebär att integritet ska vara standard och att användarens integritet ska respekteras som en grundläggande princip. Om ett företag samlar in data som innehåller personlig information, måste företaget försöka minimera mängden personuppgifter som samlas in, vara tydliga med syftet med uppgifterna och begränsa hur länge dessa uppgifter kommer att lagras. GDPR-reglerna förklarar också att de som samlar in data måste få ett tydligt och informerat samtycke från användarna för att ha rätt att samla in och använda deras personuppgifter. Detta innebär att användarna måste uttrycka sitt samtycke för att ge tillstånd till att deras uppgifter används för specifika ändamål (Thaine, 2021).

Introduktionen av GDPR markerade ett viktigt ögonblick inom dataskyddsområdet. Reglerna hjälper inte bara till att skydda användarinformation utan har även motiverat organisationer att förbättra sin cybersäkerhet. GDPR-regler kräver att organisationer ligger steget före genom att ha översyn och genomföra uppgraderingar av sina cybersäkerhetspraxis, vilket i sin tur leder till minskade risker för cyberattacker (Bašić, 2023). När GDPR introducerades år 2018 blev alla påverkade av det och ML-team från hela världen behövde anpassa sig till reglerna om hur de kan använda EU-medborgares data.

GDPR introducerade många regler som tog hänsyn till användarnas intressen. GDPR betonar hur viktigt det är för ML-team att förhindra diskriminering. Detta beror på att många ML-system tränas med partisk data vilket kan resultera i diskriminerande resultat. Om ML-team vill använda EU-medborgares personuppgifter måste de arbeta för att skapa system som följer GDPR strikta regler. Att inte följa GDPR medför konsekvenser; företag kan drabbas av stora böter om de inte följer reglerna (Klushin, 2022).

5.5 Grupper eller samhällen som kan vara sårbara för det etiska och social konsekvenserna av ML-modeller i cyberbrottsbekämpning

AI och ML skapades ursprungligen utan att ta etiska frågor i åtanke. ML kan ses som en teknik som försöker efterlikna mänsklig intelligens. Mänsklig intelligens kan leda till både etiska och oetiska beteenden. Därför blir etik relevant i sociala sammanhang, särskilt när mänsklig intelligens integreras med maskinens intelligens. Nu när samhället har utvecklats och AI och ML integreras djupt i dess struktur, har detta lett till en offentlig diskussion om deras påverkan på samhället. Diskussionen om etik i samband

med AI och ML har blivit större i takt med att användningen av dessa tekniker ökar (Bacciu et al., 2019).

Rörande vilka grupper som kan vara sårbara för ML inom cyberbrottsbekämpning har R2 nämnt människor som är digitalt omogna. *"[Dem]…löper väl lite större risk att drabbas på något sätt eftersom man kanske har dålig förståelse och insikt i hur det här fungerar...det kan ju ibland handla om övervakning eller anpassningar av information man presenteras för eller tar del av och man har inte så att säga insikt eller förståelse för [informationen] och därmed också en viss kritiskt tänkande kring att det kan vara fel eller det kan bli fel…*"

Ruoslahti et al. (2021) skriver att cybersäkerhet är ett allvarligt problem i det moderna samhället. Människor som har tillgång till IKT-enheter är i riskzonen. De flesta har inte kunskapen inom cybersäkerhet eller misslyckas med att praktisera korrekta cybersäkerhetsåtgärder. När det gäller vilka grupper som är mer sannolika att bli utsatta för cyberattacker är det yngre personer. Den yngre åldersgruppen följer inte säkerhetsrutiner, förlitar sig på säkerheten hos sina personliga enheter, har inte den kunskap som behövs för att förstå ny teknik inom sociala medier, eller är ofta engagerade i näthandel (Ruoslahti et al., 2021).

Vuxna har också problem med att undvika risker när de använder IKT-enheter. De är mer sannolika att använda offentliga Wi-Fi-nätverk, klicka på okända länkar eller använda samma lösenord för flera onlinekonton. Seniorer brukar inte vara lika cyberkunniga som de yngre generationerna, vilket gör dem mer pålitliga, något som kan utnyttjas av hackare och därmed gör dem till en sårbar grupp. (Ruoslahti et al., 2021)

Den vardagliga människan utan expertkunskaper om integritet och cybersäkerhet kan lätt bli förvirrad när de använder ny teknologi. Denna brist på kunskap och förståelse för integritetshot, dold information och påträngande verktyg för att skydda personuppgifter i digitala miljöer leder till att användaren inte har en bra förståelse för integritetsskydd (Barth et al., 2022). Om en användare inte har den kunskap de behöver för att förstå vad som pågår, kan detta leda till att de delar för mycket information och att det uppstår en obalans i makt mellan användare och leverantörer av onlinetjänster. Om en användare i denna situation får mer stöd och information, kan det hjälpa dem att fatta bättre beslut om sitt integritetsskydd (Barth et al., 2022).

R1 säger att människor utan teknisk kunskap kommer att vara en av de grupper som kan bli sårbara. Båda respondenterna påpekar att ML kan bidra till att förstärka fördomar mot samhällsgrupper som redan är utsatta. R2- *"Det här med att prata om intersektionalitet och det råkar vara så att ML-modellen då som används och på något sätt kanske har något inbyggt fel eller jävighet eller producerar dåligt beslutsunderlag agerar i ett sammanhang där också sådana grupper är implicerade eller inblandade på något sätt så kan ju det ge liksom en kaka på kaka effekt lite grann "*.

Respondenterna tog upp tre olika grupper som kan drabbas hårt av ML inom cybersäkerhet. För det första, människor som redan är utsatta i samhället, där ML kan förvärra deras situation eftersom tekniken ofta inte utvecklas med etik i åtanke. För det andra, människor som saknar teknisk kunskap, vilket kan leda till att de hamnar i problematiska situationer. För det tredje, företag som inte har råd att tillämpa ML för sina säkerhetssystem. Detta kan vara problematiskt eftersom dessa företag löper stor risk att bli attackerade.

5.6 Hur bör frågor om integritet och dataskydd hanteras om ML används för att bekämpa cyberbrott?

Respondenterna hade liknande tankar kring denna fråga. R2 betonade vikten av att definiera riskerna och vara transparent om hur de har kommit fram till dem samt hur riskerna kan åtgärdas. Respondenten pratade om ansvaret att undersöka vilka konsekvenser som kan uppstå av att tillämpa algoritmer som medför oacceptabla risker och påverkar samhället negativt, både ur ett socialt och etiskt perspektiv. R2 lyfte även fram att det borde finnas människor på plats som reglerar vem som kan använda denna typ av teknologi.

R1 framhöll vikten av GDPR och att varje person har rätt att äga sin information. Det finns regler inom olika områden av teknologi, och dessa regler måste tillämpas även för ML inom cybersäkerhet. Respondenten betonade också vikten av transparens, det är viktigt att användare vet hur ett företag samlar in och använder deras data. R1 ”*Man ska endast använda den här informationen för det den är godkänd för att användas till så det är ju viktigt tänker jag att man är tydlig om man samlar in data för att till exempel träna en ML-modell så är det ju viktigt att man informerar om det.*”

R1 säger att företag ofta samlar in mycket data för att skapa en algoritm som blir bättre och bättre. Mer data kan hjälpa ML-algoritmer att bli mer avancerade, men det kan också påverka människors integritet. Respondenten föreslår att minska mängden träningsdata som används för dessa algoritmer för att försäkra att människors integritet inte påverkas. R1-”*Men jag tror att vi kan bli bättre på att minska mängden data och verkligen bara samla in den data som behövs för ändamålet och då skulle man också kunna minska risken lite grann för att man kränker den personliga integriteten... men sen också som jag sa, jag tror att väldigt mycket beteendedata och saker vi gör, den fungerar lika bra om den är anonym så det behöver inte finnas en koppling till en person.*”

Båda respondenterna tog upp transparens. Ansvaret ligger i företagsledarnas händer att förmedla hur användarnas personliga uppgifter används. Respondent 2 anser att problemområden måste identifieras när ett system skapas, då detta kan påverka samhället negativt.

5.7 Säkerställandet av att användningen av ML-modeller i cyberbrottsbekämpning sker på ett ansvarsfullt och etiskt sätt

Respondent 1 säger att den bästa metoden för att säkerställa att företag använder ML-modeller i cyberbrottsbekämpning är genom diskussion. Om företag tar sig tid att diskutera med allmänheten om ML-modeller och hur de tänker använda dessa tekniker kan det skapa förtroende mellan utvecklare och intressenter. R1- *" […] det är också det lite grann att jag ändå tror att är man transparent och har en viss öppenhet och insyn så kan man också bygga förtroende och visa. Men det behöver ju debatteras det här och då menar jag liksom det behöver ju debatteras från alla håll*."

Respondenten fortsätter med att säga att alla behöver vara involverade i diskussionen, medborgare, experter, myndigheter och företag som är inblandade. Det är viktigt att ägna sig åt att involvera samhället i sådana diskussioner. *" […] för att kunskapen hos oss vanliga medborgare om man säger så, den är ju alldeles för låg. Även om liksom många förstår kanske […] hade vi sagt det [till] någon för tre år sedan vad en large language model var så var det ju ingen som skulle veta det. Nu finns det ju ändå ganska många som vet vad det är så att utvecklingen går ju jättefort. Men sen hänger ju inte våra regelverk och lagar med och det är jätteviktigt."*

R1 betonar vikten av att föra debatter om dessa ämnen och tar upp exemplet när människor insåg hur mycket data som Google och Microsoft använde för sina system. Detta ledde till många diskussioner och debatter, vilket i sin tur resulterade i att lagar och riktlinjer skapades för att begränsa användningen av människors personuppgifter. R2 håller med och säger att diskussioner hjälper samhället att förbättras och hitta lösningar.

> *"Jag tror det mest direkta man kan göra som samhällsmedborgare är ju att ställa de här frågorna längs de linjerna jag nämnde nyss i de sammanhang man är engagerad så i skolan, på jobbet, i föreningarna, i idrottslivet, i politiken. För att det väcker medvetenhet och frågan sätter ju också igång processer där man behöver hitta svar på de här frågorna och behöver man hitta svar då behöver man metodik och behöver man metodik då behöver man kompetens och så vidare så att fråga är ju bland det mest kraftfulla vi kan göra i olika roller och kommer en fråga upp tillräckligt ofta så börjar saker och ting röra på sig typiskt."-R2*

Att upprätthålla cybersäkerhet i dagens värld kommer med många utmaningar. Produktutvecklare anstränger sig inte för att kommunicera viktig information om cybersäkerheten för sina produkter (Megas et al., 2023). Om produktutvecklare vill att deras kunder ska ha förtroende för dem måste det finnas kommunikation mellan alla involverade parter, vilket minskar riskerna. Kommunikation innebär att ge den information som varje målgrupp behöver (Megas et al., 2023).

5.8 Riktlinjer eller åtgärder som organisationer bör vidta för att minimera de negativa sociala och etiska konsekvenserna av ML i cyberbrottsbekämpning

R1 tar upp att det ska finnas tydliga riktlinjer och lagar på plats för att hantera de olika utmaningar som kan uppstå med ML-modeller inom cyberbrottsbekämpning. Återigen betonar R1 hur viktigt det är med transparens; organisationer måste vara ärliga med hur intressenters data används. De går vidare och pratar om riskerna; idag går teknikutvecklingen väldigt snabbt, vilket gör att utvecklare inte har tid att göra riskbedömningar. Företag är ivriga att gå framåt utan att ta hänsyn till framtiden och hur deras produkter kan påverka samhället. R1 fortsätter att prata om vikten av GDPR. ” […] *men också tänker jag så här att lite grann också som GDPR att där säger ju regelverket att du får spara information så länge du behöver den och kan motivera att du behöver den och det ska också vara så lite information som möjligt så du får inte bara ta en massa information som du har tillgänglig utan det ska vara den information som du verkligen behöver och att kunna få regler som är lite grann på det sättet kring det här det tror jag är väldigt brådskande och viktigt att få på plats.”*

R2 säger samma sak, att samhället har gjort mycket bra arbete med att hantera frågor om ML. Det finns lagar som tar itu med dessa problem. Samhällets och myndigheternas engagemang i att föra debatt om dessa frågor leder till att nya standarder utvecklas

> *”Man arbetar för att få fram en AI-strategi brett, men man arbetar också för att få fram en ny cybersäkerhetsstrategi brett och de här behöver ju vara nära förknippade med varandra så att man från samhällets sida, den cybersäkerhetsstrategi man formar och formulerar säkerställer att man [har] tillräckligt [med] hänsyn och uppmärksamhet på möjliga etiska och sociala konsekvenser. Balansera det mot andra uppenbara risker i förhållande till antagonistiska hot eller aktörsdrivna hot, alltså hot från andra stater till exempel så vi får väldigt mycket uppmärksamhet nu, vilket ju kanske är rimligt om man tittar på konsekvenserna. Men man måste vara uppmärksam på de här potentiella krypande långsamma konsekvenserna som man kanske inte upptäcker så jättetydligt på kort sikt men som får enorma konsekvenser på lite längre sikt.”-R2*

Manoharan och Sarker (2022) skriver att även om ML och AI är kraftfulla verktyg att använda inom cybersäkerhet, är de inte en perfekt lösning. För att hantera den komplexa naturen av cybersäkerhet behövs en nyanserad strategi som integrerar nya tekniker med traditionella metoder och involverar experter inom området som beslutsfattare, juridiska rådgivare och tekniska experter. Om människor som har möjlighet att skapa förändring inom AI- och ML-samhället kommer samman och samarbetar, är det möjligt att säkerställa effektiviteten och den etiska och ansvarsfulla användningen av dessa teknologier (Manoharan & Sarker, 2022).

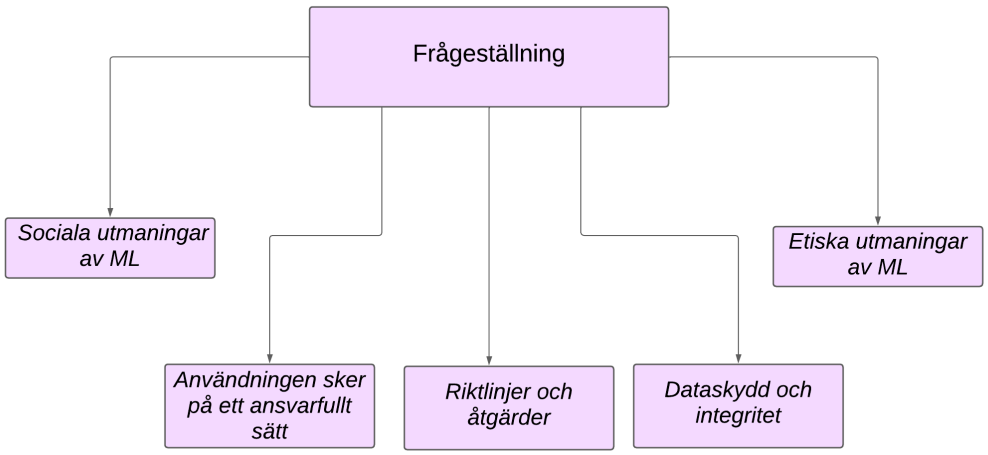
6 Resultat

Resultatkapitlet är avsett att presentera syftet med studien och redogöra för vad som har framkommit genom analysen av litteraturen och respondenternas åsikter om ämnet.

Studiens frågeställning har varit: *Vilka sociala och etiska konsekvenser uppstår vid implementeringen av ML-modeller inom cyberbrottsbekämpning?*

Studien bidrar till den pågående diskussionen om ML och dess utmaningar vid användning av en sådan teknik inom cybersäkerhet.

Figur 1 är sammanställningen av resultatet som kom fram i analysen. Det hittades fem faktorer som lyftes fram under intervjuerna och litteraturen.



Figur 1 – Resultatet av frågeställningen

6.1 Sociala utmaningar av ML

En av utmaningarna som kommer med användningen av ML är dess påverkan på samhället, de sociala utmaningarna. I denna studie var det viktigt att inkludera de sociala konsekvenserna av ML, särskilt när det gäller brottsbekämpning och hur det kan påverka människor i framtiden. Tidigare litteratur visar att AI och ML har gynnat samhället, men dessa verktyg kommer inte utan risker (Shah, 2021). En av respondenterna tog upp att det är viktigt att beakta sociala krav och människors personliga information när dessa ML-modeller utvecklas. Många ML-modeller kräver stora mängder data för att kunna producera bra resultat, och här uppstår utmaningarna

med ML. Att använda stora datamängder innebär en kompromiss av integriteten. Himthani et al., 2020, skriver att metoder för att minska mängden träningsdata inom ML är begränsade, även om det finns sätt att reducera datamängden medför det komplikationer vid tillämpningen av dessa metoder. Kostnaden för att hantera data är mycket hög, vilket leder till en annan social utmaning som samma respondent tog upp.

Att tillämpa avancerade tekniker som ML för att förbättra ens företag kan leda till mycket bra resultat, men inte alla har råd att använda ML-tekniker (Himthani et al., 2020). Respondenten nämnde att många organisationer inte kommer att kunna använda ML i samma utsträckning som andra. Företag som redan är värda biljoner påverkas inte av dessa kostnader (Vanian & Leswing, 2023). Litteraturen visar att även om ett litet företag har råd att använda ML medför det risker; dessa företag har inte samma starka ramverk för cybersäkerhet som större företag har. Även om de vill expandera medför det stora risker som sätter dem i fara. Det kan leda till att de kanske inte agerar för att använda ML, vilket i sin tur gynnar de stora företagen eftersom de har mindre konkurrens (TechCurators, 2023; Vanian & Leswing, 2023)

En av de största utmaningarna med ML är partiskhet. För att skapa dessa algoritmer som ska användas för att skapa modeller krävs träningsdata och någon som är ansvarig för att mata in denna data. Om det finns människor som hjälper till att skapa dessa algoritmer kan det hända att algoritmen blir partisk. Människor är inte perfekta varelser; de har fördomar som de är medvetna om eller inte medvetna om. Om en person med fördomar hjälper till att skapa ML-modeller kommer modellen att göra förutsägelser som är partiska (Milaninia, 2021).

Det finns ett behov av att förstå hur dessa ML-algoritmer utvecklas. ML-algoritmer är väldigt komplexa, vilket leder till att människor inte förstår hur dessa algoritmer kommer fram till olika beslut. Det är här black-box-effekten inträder; människor kan se indata och utdata men inte den interna processen. Även om dessa black-box-system är komplexa anser många att det är värt att använda dem. Detta leder till att fördelarna överväger nackdelarna. Detta är problematiskt eftersom black-box-system kan bidra till att fördomar förstärks (Franzen et al., 2021). En respondent tog upp att detta kan vara ett problem; om beslutsfattandet överlåts till ML-modeller när det gäller brottsbekämpning öppnar det upp stora risker för att något kan gå fel. Det finns redan studier om hur algoritmer kan påverka minoriteter som har begått brott, där algoritmen är partisk mot dem vilket leder till att de får längre strafftider.

6.2 Etiska utmaningar av ML

Respondenterna betonade vikten av transparens och att företag måste vara ärliga om hur dessa algoritmer skapas. Om företag är transparenta under hela utvecklingsprocessen kan de lättare arbeta mot att lösa problem som uppstår (Rouse, 2023). Leverantörer som skapar dessa algoritmer inom cybersäkerhet är ofta inte ärliga

om hur en algoritm kommer fram till ett specifikt resultat, vilket leder till att användare av dessa algoritmer inte har förtroende för leverantören (Chamberlain, 2023). Detta kan leda till problem i framtiden, eftersom bristen på transparens från leverantörerna kan resultera i ett partiskt system.

En respondent tog upp problemet med dataskydd och sekretess som en etisk konsekvens. Respondenten säger att det är viktigt att människor äger sin egen information. Företag måste få samtycke från användare för att använda deras data. GDPR är till för att skydda EU-medborgare och deras integritet. Det finns regler på plats som betonar vikten av samtycke och om det inte finns samtycke mellan företag och användare kan detta påverka företaget negativt (Thaine, 2021).

I denna studie var det inte bara viktigt att veta vilka de sociala och etiska utmaningarna med ML inom cyberbrottsbekämpning var, utan också vilka grupper som kommer att påverkas av dessa konsekvenser. Både respondenterna ansåg att människor som inte har mycket kunskap inom teknologi kan bli påverkade. De löper större risk att drabbas på grund av sin bristande kunskap. Respondenterna betonade även att ML-modeller kan påverka människor som redan är sårbara i samhället. Användningen av dessa verktyg kan öka deras sårbarhet.

6.3 Dataskydd och integritet

Båda respondenterna delade samma åsikter om denna fråga. De anser att det är viktigt att definiera riskerna och vara transparenta om hur dessa risker identifieras och hanteras. De anser att lösningen ligger hos dem som skapar dessa algoritmer och att de måste vara medvetna om de problem som kan uppstå när en ML-modell skapas för användning inom cybersäkerhet. Respondenterna betonar återigen vikten av transparens; företag måste både få användarnas samtycke och informera dem om hur deras data kommer att användas.

En av respondenterna föreslog att det inte är nödvändigt att använda stora datamängder som ML-modeller använder idag. Dessa verktyg bör bara använda den data som är absolut nödvändig, och i många fall behöver inte ens personlig information inkluderas i träningsdata, särskilt inte inom området cybersäkerhet.

6.4 Användningen sker på ett ansvarsfullt sätt

Respondenterna anser att det är viktigt med diskussioner mellan experter och allmänheten för att möta de utmaningar som kan uppstå med ML inom cyberbrottsbekämpning. Det är nödvändigt att alla är involverade i dessa diskussioner eftersom alla kan påverkas av dem. Experter, myndigheter och företag måste engagera sig i att främja debatter, vilket kan öka allmänhetens förtroende för dem.

En respondent tog upp ett exempel på när människor insåg hur mycket data både Google och Microsoft använde, vilket ledde till diskussioner och debatter. Detta resulterade i att lagar och riktlinjer infördes för att begränsa stora företags möjlighet att samla in vilken typ av data de ville ha.

6.5 Riktlinjer och åtgärder

I dagens samhälle är det nästan viktigare att vara först med att skapa något nytt, vilket leder till att utvecklare inte genomför riskbedömningar eftersom de är ivriga med att skapa något. Det tas inte hänsyn till vilken påverkan ens produkter kan ha på samhället. Respondenten tar upp redan befintliga riktlinjer och förklarar att det är olämpligt att spara information som inte behövs. Det bör finnas motiveringar till varför viss information samlas in och sparas. En annan respondent betonar att samhället gör det bra redan nu med att hantera alla problem. Det finns regler på plats för att förhindra företag från att missbruka data.

7 Diskussion

Detta kapitel är avsett för reflektion kring den genomförda studien. Det ger författaren möjlighet att diskutera studiens process ur ett samhälleligt, vetenskapligt och etiskt perspektiv.

7.1 Syfte och frågeställning

Syftet med denna studie var att utforska konsekvenserna av att använda ML-modeller inom cyberbrottsbekämpning ur etiskt och socialt perspektiv, samt hur detta skulle påverka dagens samhälle. Författarna Elluri et al. (2023) pekade på forskningsgapet inom detta område, vilket tydde på att det var nödvändigt att starta diskussioner och forskning kring ämnet. Om teknologier som ML-modeller ska användas måste människor förstå hur det påverkar dem och vilka grupper som kan bli mer berörda av att implementera nya tekniker som maskininlärning inom cybersäkerhet.

I början av studien antogs det att respondenterna skulle framföra liknande åsikter som de som presenteras i litteraturen om maskininlärning inom cybersäkerhet. Både respondenterna visade sig ha en god förståelse för cybersäkerhet, och det förväntades att deras tankar skulle likna de hos många andra individer som arbetar inom området. Både respondenterna och litteraturen som studerades kom till liknande slutsatser. Maskininlärning är en mycket effektiv teknik för att automatisera processer och underlätta för företag att snabbare reagera på hot från hackare. Det har förenklat processer som annars skulle vara mycket mer tidskrävande för människor att utföra. Det finns dock nackdelar med att förlita sig på maskininlärning i känsliga situationer som brottsbekämpning, och det är viktigt att överväga vilka som kan drabbas negativt av att använda algoritmer i sådana sammanhang. Det är avgörande att utvecklare och leverantörer är transparenta när det gäller hur deras lösningar har skapats. Personer som är involverade i utvecklingen av ML-modeller måste vara ärliga mot allmänheten om vilken träningsdata som används och vilka åtgärder som vidtas för att undvika att skapa modeller som förstärker fördomar eller kan användas för att skada människor.

Genom diskussioner mellan alla berörda parter kan problematiska aspekter av maskininlärning identifieras och lösas, vilket kan leda till införandet av lagar och riktlinjer för att förhindra utvecklingen av teknik som kan vara skadlig för människor. GDPR är ett exempel på detta, där det har skapats tydliga regler för hantering av EU-medborgares information efter diskussioner och debatter mellan människor angående deras personliga integritet.

7.2 Metodval

Kvalitativ ansats användes för att nå studiens resultat. Denna ansats ger forskare möjlighet att utforska komplexa sociala fenomen. Studien fokuserade på att utforska ML-modellers användning inom cybersäkerhet och vilka konsekvenser detta hade på samhället, vilket gjorde den kvalitativa ansatsen till den mest lämpliga metoden för att undersöka studiens frågeställning. Intervjuer användes som datainsamlingsmetod, vilket hjälper till att få djupare och mer nyanserade insikter från deltagare jämfört med kvantitativa metoder.

Även om det teoretiskt sett hade varit möjligt att använda en kvantitativ ansats, skulle det vara svårt att samla in insikter som behandlar ämnen som diskriminering, fördomar och partiskhet. Respondenterna gav under intervjuerna mycket komplexa och långa svar, vilket inte hade varit möjligt att få genom en enkät eftersom den inte ger en djup förståelse för människors åsikter på samma sätt som en intervju eller observation. Majoriteten av intervjuerna resulterade i nyanserade svar från respondenterna, och följdfrågor ställdes ibland för att få en tydligare förståelse av deras synpunkter, vilket möjliggjorde fördjupade diskussioner. Detta hade inte varit möjligt med en kvantitativ ansats.

Överlag var intervjuer den bästa metoden för att samla data, eftersom de ger människor utrymme att förklara vad de känner på ett mer djupgående sätt. En kvalitativ ansats öppnar upp möjligheten för forskare att observera människor för att samla in data. I denna studie handlade det dock inte om att observera hur människor arbetar, utan om deras åsikter, vilket gjorde observation som datainsamlingsteknik olämplig. Om det hade funnits möjlighet att observera utvecklare som skapar ML-modeller och se hur utvecklingsprocessen ser ut, skulle detta ha varit en mycket bra metod. Det skulle ha öppnat möjligheten att utforska och diskutera de aspekter och utmaningar utvecklare möter när de skapar ML-modeller och hur de arbetar för att förhindra att skapa modeller som påverkar samhället negativt.

Att hitta lämpliga personer att intervjua var en komplicerad uppgift. Många individer var inte villiga att delta i intervjuer, antingen på grund av brist på intresse eller tidsbegränsningar. Även om intervjuer är den mest lämpliga metoden för att samla in data, kommer de inte utan brister. Att schemalägga intervjuer, som kan ligga veckor ifrån varandra, är problematiskt, särskilt när det finns en tidsplan att följa. Det är viktigt att vara tillmötesgående mot deltagarna och anpassa sig till deras schema, vilket kan leda till lång väntetid mellan att de svarar på mailet och att intervjun genomförs.

För denna studie var det endast möjligt att hitta två deltagare. Detta kan innebära att studien betraktas som mindre tillförlitlig eftersom den endast bygger på åsikter från två personer. Detta kan ses som en svaghet och göra studien mindre trovärdig för läsarna. För att stärka studiens trovärdighet byggdes den på både litteratur och respondenternas

åsikter. Det var viktigt att ha deltagare som kunde bekräfta och komplettera de resultat som framkom i litteraturen.

Kvalitativ ansats har många fördelar eftersom det ger forskare djupare insikter om olika fenomen. Dock kommer det med begränsningar; i detta fall går det inte att generalisera respondenternas resultat eftersom det inte fanns tillräckligt med deltagare för studien. Det finns alltid en chans att deltagares åsikter och upplevelser är unika för deras verklighet, vilket är en nackdel med intervjuer. Respondenterna som intervjuades för denna studie hade erfarenhet av att arbeta med cybersäkerhet, och deras åsikter kunde jämföras med litteraturen, där det fanns många likheter mellan vad de sa och vad forskningen visade.

Om det hade funnits möjlighet att börja om och förhindra problem som uppstod vid användning av intervjuer som datainsamlingsteknik skulle det finnas ändringar. För att denna studie skulle ha varit bättre behövdes fler deltagare för att få en bredare förståelse för hur människor inom cybersäkerhetsfältet tänker kring användningen av ML-modeller. Omkring tio personer skulle ha varit ett bra antal, vilket skulle ha hjälpt till med att generalisera åsikterna inom detta område. Att vänta för länge på svar från potentiella deltagare var ett misstag som gjordes under studien, då det är mer sannolikt att människor inte har tid för intervjuer. Därför är det viktigt att kontakta flera personer för att öka chansen att få fler intervjuer och därmed mer data.

För att transkribera intervjun användes Klang.ai, som är ett AI-drivet transkriptionsverktyg. Att använda AI för att transkribera intervjun underlättade analysprocessen. Även om det var två personer som intervjuades skulle det vara tidskrävande att sitta ner och skriva ner allt som respondenterna sa. Anledningen till att Klang.ai användes var att en av respondenterna påpekade att det fanns transkriptionsverktyg som kunde användas.

Analysmetoden som användes för denna studie var tematisk analys. Det finns flera sätt att analysera data vid en kvalitativ analys, såsom tematisk analys, innehållsanalys, diskursanalys och grundad teori. Anledningen till att tematisk analys valdes var att det redan fanns en viss förståelse för hur denna typ av analys ska genomföras. Metoden användes för att nå studiens resultat och visade sig vara effektiv eftersom den går ut på att hitta och kategorisera teman. Detta underlättar strukturen och skrivandet av analysen. Respondenterna hade samma åsikter som togs upp i litteraturen, vilket underlättade analysen. Det var lätt att jämföra litteraturen med det respondenterna sa. Tematisk analys är också bra när forskare vill förstå individers eller gruppers erfarenheter och perspektiv. Denna analysmetod är mycket populär inom psykologiska, sociologiska, antropologiska och pedagogiska discipliner (Zaveri, 2023), vilket gör den lämplig för denna studie.

För att hitta källor för studien användes Google, där olika sökord skrevs in för att hitta källor relaterade till ämnet. Populärvetenskapliga artiklar användes i studien, och dessa källor hittades via Google-sökningar. För vetenskapliga artiklar användes Google Scholar som databas. Det fanns inte tillräckligt många artiklar om de etiska och sociala konsekvenserna av att använda ML-modeller inom cyberbrottsbekämpning. Det finns visserligen artiklar om hur ML används inom cybersäkerhet, men forskning om dess effekter på samhället är bristfällig. Detta kan sänka kvaliteten på studien eftersom det saknas litteratur som kan jämföras och som behandlar samma ämnen, vilket kan försvåra jämförelser av resultaten. På grund av svårigheten att hitta referenser inom detta ämne slösades mycket tid bort på att söka källor, vilket förlängde skrivprocessen.

7.3 Etiska aspekter

Alla forskningsetiska principer följdes under denna studie. Det var viktigt att informera deltagarna om syftet med studien. All nödvändig information för att bedöma om de ville delta gavs i det första mailet, och när de svarade tillbaka fick de intervjuguiden. Det var viktigt att deltagarna förstod vilka frågor som skulle ställas under intervjun, och om det fanns några frågor de inte ville svara på kunde de meddela det innan intervjun. Deltagarna blev också informerade om att de skulle vara helt anonyma.
Allt inspelat material kommer att raderas när studien är helt avslutad. Det finns ingen anledning att behålla materialet längre efter att studien är slutförd.

7.4 Sociala aspekter

Forskning som tacklar de sociala utmaningarna inom ett ämne kommer alltid att vara nödvändig då det hjälper samhället att bli bättre. Det är på detta sätt människor lär sig om problem som de kanske inte hade tänkt på tidigare. Litteraturen visar att maskininlärning kan ha allvarliga konsekvenser när det kommer till fördomar. Sådana konsekvenser kan ha en stor påverkan på samhället och kan drabba redan utsatta grupper. Att undersöka de sociala konsekvenserna av ML-modeller inom cyberbrottbekämpning gynnar människor och skapar ett mer rättvist samhälle.

7.5 Vetenskapliga aspekter

Att hitta bra referenser som behandlar maskininlärning inom cybersäkerhet och dess utmaningar var inte lätt. Det finns artiklar där författare skriver om maskininlärning inom cybersäkerhet fast de utforskar andra frågor. Även i dessa artiklar betonas det att det inte finns tillräckligt med forskning om användningen av maskininlärning i cybersäkerhet. Maskininlärning är inget nytt fenomen; det har blivit betydligt mer populärt de senaste åren. Det bör finnas tillräckligt med forskning om olika användningsområden för maskininlärning och hur det har påverkat samhället. Forskning behövs även om hur utmaningar som uppstår med maskininlärning kan

hanteras, såsom användningen av stora datamängder och hur ML-modeller kan förhindras från att använda människors personliga information.

7.6 Framtida forskning

Framtida forskning inom detta område bör fördjupa förståelsen för hur utvecklare reflekterar kring etiska frågor och hur de arbetar för att skapa system som inte negativt påverkar samhället. Att undersöka tankesättet hos dem som skapar dessa modeller skulle ge värdefull insikt och ge allmänheten förståelsen för varför vissa beslut fattas i processen att utveckla dem.

Att intervjua personer som är involverade i utformningen av lagar och riktlinjer för att förhindra missbruk av personlig information kan vara fördelaktigt för forskningen. Dessa personer kan bidra med perspektiv om hur maskininlärning kan påverka samhället i framtiden och hur människor kan hantera nya utmaningar som kan uppstå med dess användning inom cybersäkerhet. Genom att intervjua representanter från olika sektorer kan en mer nyanserad diskussion om ämnet uppnås.

Referenser

Abdiyeva-Aliyeva, G., Aliyev, J., & Sadigov, U. (2022). Application of classification algorithms of Machine learning in cybersecurity. *Procedia Computer Science*, *215*, 909–919. https://doi.org/10.1016/j.procs.2022.12.093

Ashok, M., Madan, R., Joha, A., & Sivarajah, U. (2022). Ethical framework for Artificial Intelligence and Digital technologies. *International Journal of Information Management*, *62*, 102433. https://doi.org/10.1016/j.ijinfomgt.2021.102433

Apruzzese, G., Colajanni, M., Ferretti, L., Guido, A., & Marchetti, M. (2018). On the effectiveness of machine and deep learning for cyber security. In IEEE Conference Publication | IEEE Xplore (pp. 371–390). Institute of Electrical and Electronics Engineers. https://doi.org/10.23919/CYCON.2018.8405026

Bacciu, D., Biggio, B., Lisboa, P. J. G., Martín, J. D., Oneto, L., & Vellido, A. (2019). Societal issues in machine learning: when learning from data is not enough. In ESANN 2019 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (p. 455). European Symposium on Artificial Neural Networks (ESANN). https://www.esann.org/sites/default/files/proceedings/legacy/es2019-6.pdf

Barbierato, E., & Gatti, A. (2024). The Challenges of Machine Learning: A Critical review. Electronics, 13(2), 416. https://doi.org/10.3390/electronics13020416

Barth, S., De Jong, M. D., & Junger, M. (2022). Lost in privacy? Online privacy from a cybersecurity expert perspective. Telematics and Informatics, 68, 101782. https://doi.org/10.1016/j.tele.2022.101782

Bašić, B. (2023, September 15). The impact of GDPR on cybersecurity and data protection. https://www.linkedin.com/pulse/impact-gdpr-cybersecurity-data-protection-bo%C5%BEa-ba%C5%A1i%C4%87-1f/

Bharadiya, J. P. (2023). AI-Driven Security: How Machine learning will shape the future of Cybersecurity and Web 3. American Journal of Neural Networks and Applications. https://doi.org/10.11648/j.ajnna.20230901.11

Bharadiya, J. P. (2023). Machine learning in cybersecurity: Techniques and challenges. *European Journal of Technology*, *7*(2), 1–14. https://doi.org/10.47672/ejt.1486

Busetto, L., Wick, W., & Gumbinger, C. (2020). How to use and assess qualitative research methods. Neurological Research and Practice, 2(1). https://doi.org/10.1186/s42466-020-00059-z

Cabrera, Á. A., Epperson, W., Hohman, F., Kahng, M., Morgenstern, J., & Horng Chau, D. (2020). FAIRVIS: Visual analytics for discovering intersectional bias in machine learning. In IEEE Conference Publication. Institute of Electrical and Electronics Engineers. https://doi.org/10.1109/VAST47406.2019.8986948

Callahan, M. (2023, February 24). Algorithms were supposed to reduce bias in criminal Justice—Do they? Boston University. https://www.bu.edu/articles/2023/do-algorithms-reduce-bias-in-criminal-justice/

Carta, S. (Ed.). (2022). ML Between routines and wonder. In *Machine Learning and the City: Applications in Architecture and Urban Design*. John Wiley & Sons Ltd. https://doi.org/10.1002/9781119815075

Castañón, J. (2019, May 1). 10 Machine Learning Methods that Every Data Scientist Should Know. *Medium*. https://towardsdatascience.com/10-machine-learning-methods-that-every-data-scientist-should-know-3cc96e0eeee9

Chamberlain, C. (2023, December 8). Looking beyond the hype cycle of AI/ML in cybersecurity. https://www.darkreading.com/vulnerabilities-threats/looking-beyond-hype-cycle-ai-ml-cybersecurity

Dasgupta, D., Akhtar, Z., & Sen, S. (2020). Machine learning in cybersecurity: a comprehensive survey. *The Journal of Defense Modeling and Simulation*, *19*(1), 57–106. https://doi.org/10.1177/1548512920951275

De Harder, H. (2023, July 13). Ethical considerations in machine learning projects. *Medium*. https://towardsdatascience.com/ethical-considerations-in-machine-learning-projects-e17cb283e072

Dovetail Editorial Team. (2023, April 18). Field Study Guide: Definition, steps & Examples. https://dovetail.com/research/field-study/

Elluri, L., Mandalapu, V., Vyas, P., & Roy, N. (2023). Recent advancements in machine learning for cybercrime prediction. *Journal of Computer Information Systems*, 1–15. https://doi.org/10.1080/08874417.2023.2270457

El Mrabet, M. A., El Makkaoui, K., & Faize, A. (Eds.). (2021). Supervised machine learning: A survey. In IEEE Conference Publication. Institute of Electrical and Electronics Engineers. https://doi.org/10.1109/CommNet52204.2021.9641998

Flower, D. (2023, July 25). The Power of Machine Learning: The Business Impact on Real-Time Data. *Forbes*. https://www.forbes.com/sites/forbestechcouncil/2023/07/25/the-power-of-machine-learning-the-business-impact-on-real-time-data/

Franzen, M., Kloetzer, L., Ponti, M., Trojan, J., & Vicens, J. (2021). Machine learning in citizen Science: promises and implications. In Springer eBooks (pp. 183–198). https://doi.org/10.1007/978-3-030-58278-4_10

Gatha. (n.d.). *The importance of ethics in machine learning*. https://censius.ai/blogs/the-importance-of-ethics-in-machine-learning

Gerring, J. (2017). Qualitative methods. *Annual Review of Political Science*, *20*(1), 15–36. https://doi.org/10.1146/annurev-polisci-092415-024158

Ghosh, P. (2023, October 12). *Top Ethical Issues with AI and Machine Learning - DATAVERSITY*. DATAVERSITY. https://www.dataversity.net/top-ethical-issues-with-ai-and-machine-learning/

Gottsegen, G. (2023, August 7). *Machine learning in cybersecurity: How it works and companies to know*. Built In. https://builtin.com/artificial-intelligence/machine-learning-cybersecurity

Halbouni, A., Surya Gunawan, T., Hadi Habaebi, M., Halbouni, M., Kartiwi, M., & Ahmad, R. (2022). Machine Learning and Deep Learning Approaches for CyberSecurity: A

review. IEEE Journals & Magazine | IEEE Xplore, 10, 19572–19585. https://doi.org/10.1109/ACCESS.2022.3151248

Hammarberg, K., Kirkman, M., & De Lacey, S. (2016). Qualitative research methods: when to use them and how to judge them. *Human Reproduction*, *31*(3), 498–501. https://doi.org/10.1093/humrep/dev334

Hashemi-Pour, C., & Wigmore, I. (2023, October 16). *machine learning algorithm*. WhatIs. https://www.techtarget.com/whatis/definition/machine-learning-algorithm

Himthani, P., Prasad Dubey, G., Mohan Sharma, B., & Taneja, A. (Eds.). (2020). Big data privacy and challenges for machine learning. IEEE Conference Publication | IEEE Xplore. Institute of Electrical and Electronics Engineers. https://doi.org/10.1109/I-SMAC49090.2020.9243527

Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets*, *31*(3), 685–695. https://doi.org/10.1007/s12525-021-00475-2

kaspersky. (2023, April 19). *AI and Machine Learning in Cybersecurity — How They Will Shape the Future*. www.kaspersky.com. https://www.kaspersky.com/resource-center/definitions/ai-cybersecurity

Kiger, M. E., & Varpio, L. (2020). Thematic analysis of qualitative data: AMEE Guide No. 131. Medical Teacher, 42(8), 846–854. https://doi.org/10.1080/0142159x.2020.1755030

Klushin, P. (2022, July 13). GDPR considerations when training machine learning models | Qwak. https://www.qwak.com/post/gdpr-considerations-when-training-machine-learning-models

Kästner, C. (2022, December 20). Security and Privacy in ML-Enabled Systems - Christian Kästner - Medium. *Medium*. https://ckaestne.medium.com/security-and-privacy-in-ml-enabled-systems-1855f561b894

Laskowski, N., & Tucci, L. (2023, November 13). *artificial intelligence (AI)*. Enterprise AI. https://www.techtarget.com/searchenterpriseai/definition/AI-Artificial-Intelligence

Liu, B., Ding, M., Shaham, S., Rahayu, W., Farokhi, F., & Lin, Z. (2021). When machine learning meets privacy. ACM Computing Surveys, 54(2), 1–36. https://doi.org/10.1145/3436755

Maheshwari, R. (2023, April 3). What is artificial intelligence (AI) and how does it work? *Forbes Advisor INDIA*. https://www.forbes.com/advisor/in/business/software/what-is-ai/

Manoharan, A., & Sarker, M. (2022). Revolutionizing Cybersecurity: Unleashing the power of artificial intelligence and machine learning for Next-Generation threat detection. International Research Journal of Modernization in Engineering Technology and Science, 04(12). https://doi.org/10.56726/irjmets32644

Megas, K., Smith, A., Newton, E., Elazari, A., & Cuthill, B. (2023, April 3). The importance of transparency – fueling trust and security through communication. NIST. https://www.nist.gov/blogs/cybersecurity-insights/importance-transparency-fueling-trust-and-security-through

Milaninia, N. (2021). Biases in machine learning models and big data analytics: The international criminal and humanitarian law implications. International Review of the Red Cross, 102(913), 199–234. https://doi.org/10.1017/s1816383121000096

Monteiro, J. P., Ramos, D., Carneiro, D., Duarte, F., Fernandes, J. M., & Novais, P. (2021). Meta-learning and the new challenges of machine learning. International Journal of Intelligent Systems, 36(11), 6240–6272. https://doi.org/10.1002/int.22549

Monteith, S., Bauer, M., Alda, M., Geddes, J., Whybrow, P. C., & Glenn, T. (2021). Increasing cybercrime since the pandemic: Concerns for psychiatry. Current Psychiatry Reports, 23(4). https://doi.org/10.1007/s11920-021-01228-w

Patrick Green, B. (n.d.). *Technology Ethics - Markkula Center for Applied Ethics*. https://www.scu.edu/ethics/focus-areas/technology-ethics/

Rouse, M. (2023, December 15). Transparency. Techopedia. https://www.techopedia.com/definition/30368/transparency-data

Ruoslahti, H., Coburn, J., Trent, A., & Tikanmäki, I. (2021). Cyber Skills Gaps – A Systematic Review of the Academic literature. Connections: The Quarterly Journal, 20(2), 33–45. https://doi.org/10.11610/Connections.20.2.04

Shah, V. (2022). Machine Learning Algorithms for Cybersecurity: Detecting and Preventing Threats. Spanish Journal of Scientific Documentation (REDC), 15, 4. https://doi.org/10.5281/zenodo.10779509

Sakharkar, A. (2024, February 16). New machine-learning technique promises 40% speed boost in real-world datasets. *Tech Explorist*. https://www.techexplorist.com/new-machine-learning-technique-promises-40-speed-boost-real-world-datasets/81190/

Sarker, I. H. (2021). Machine learning: algorithms, Real-World applications and research directions. *SN Computer Science*, *2*(3). https://doi.org/10.1007/s42979-021-00592-x

Sarker, I. H. (2022). Machine learning for intelligent data analysis and automation in cybersecurity: Current and future Prospects. *Annals of Data Science*, *10*(6), 1473–1498. https://doi.org/10.1007/s40745-022-00444-2

Shaibu, S. (2023, April 15). The Ethics of Machine Learning: What you need to know. *Medium*. https://medium.com/@Samietex/the-ethics-of-machine-learning-what-you-need-to-know-8f827568c554

ShardSecure. (2024, January 22). Machine Learning and Cybersecurity: A Double-Edged Sword. https://shardsecure.com/. https://shardsecure.com/blog/machine-learning-cybersecurity

Sharifani, K., & Amini, M. (2023). Machine Learning and Deep Learning: A review of Methods and applications. World Information Technology and Engineering Journal, 10(07), 3897–3904. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4458723#paper-citations-widget

Staff, C. (2023, November 29). *3 types of machine learning you should know*. Coursera. https://www.coursera.org/articles/types-of-machine-learning

Sudar, K. M., Deepalakshmi, P., Nagaraj, P., & Muneeswaran, V. (2020). Analysis of cyberattacks and its detection mechanisms. In 2020 Fifth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN). IEEE. https://doi.org/10.1109/ICRCICN50933.2020.9296178

TechCurators, a TC Group Company. (2023, November 30). Overcoming the Labyrinth: The Difficulties of AI integration in small Enterprises - Possibilities and barriers. https://www.linkedin.com/pulse/overcoming-labyrinth-difficulties-ai-integration-small-enterprises-dapwc/

Tenny, S. (2022, September 18). *Qualitative study*. StatPearls - NCBI Bookshelf. https://www.ncbi.nlm.nih.gov/books/NBK470395/

Thaine, P. (2021, November 17). 5 things ML teams should know about privacy and the GDPR. https://www.darkreading.com/cyber-risk/5-things-ml-teams-should-know-about-privacy-and-the-gdpr

Tucci, L., & Burns, E. (2023, September 15). *What is machine learning and how does it work? In-depth guide*. Enterprise AI. https://www.techtarget.com/searchenterpriseai/definition/machine-learning-ML

Vanian, J., & Leswing, K. (2023, April 17). ChatGPT and generative AI are booming, but the costs can be extraordinary. CNBC. https://www.cnbc.com/2023/03/13/chatgpt-and-generative-ai-are-booming-but-at-a-very-expensive-price.html

Veena, K., Meena, K., Kuppusamy, R., Teekaraman, Y., Angadi, R. V., & Thelkar, A. R. (2022). Cybercrime: Identification and prediction using Machine learning techniques. *Computational Intelligence and Neuroscience*, *2022*, 1–10. https://doi.org/10.1155/2022/8237421

Verma, K. K., Singh, B. M., & Dixit, A. R. (2019). A review of supervised and unsupervised machine learning techniques for suspicious behavior recognition in intelligent surveillance system. *International Journal of Information Technology*, *14*(1), 397–410. https://doi.org/10.1007/s41870-019-00364-0

Vetenskapsrådet. (2002). Forskningsetiska principer inom humanistisk-samhällsvetenskaplig forskning. In https://www.vr.se/. https://www.vr.se/download/18.68c009f71769c7698a41df/1610103120390/Forskningsetiska_principer_VR_2002.pdf

Vetenskapsrådet. (2023, December). Etik i forskningen och god forskningssed. https://www.vr.se/. https://www.vr.se/uppdrag/etik/etik-i-forskningen.html

Warren, K. (2023, April). Qualitative Data Analysis Methods 101: The Big 6 Methods (Including examples). Grad Coach. https://gradcoach.com/qualitative-data-analysis-methods/

Zaveri, A. (2023a, July 9). Innehållsanalys vs tematisk analys: En närmare titt. Mind the Graph Blogg. https://mindthegraph.com/blog/sv_se/innehallsanalys-vs-tematisk-analys/