

Skulle AI-algoritmer kunna underlätta SSYK-användandet och spara samhället pengar?

### 0.0.1 Sammanfattning. Sammanfattning

Jobtech har utvecklat ett flertal algoritmer för etikettering av SSYK, varav den bästa är Bert Classifier. Innan föreliggande rapport antogs dess precision vara 86%, vilket uppmättes genom att använda människans etikettering som facit. Vidare förekom en ovisshet kring människans precision, men den antogs vara omkring 95%. Rapporten problematiserar båda dessa siffror och finner nya värden för de angivna precisionerna genom att utföra två stickprov. Den första av dem, som går att läsa om under experiment 1, kommer fram till att människans precision är omkring 78%. Bert Classifier undersöks sedan, i experiment 2, genom att djupgräva i de 14% där Bert anses ha etiketterat felaktigt. Resultatet påvisade att Bert egentligen etiketterade korrekt i majoriteten av dessa fall och att människans etikettering, som agerat facit, varit felaktig i en betydligt högre grad. Det faktiska värdet av Berts precision är enligt analysen cirka 97%. Därmed är Berts precision signifikant högre än människans, vilket beror på en rad faktorer.

### Slutsats och rekommendation

Rapportens rekommendation är att automatisera all SSYK-etikettering på JobTech, med vidare motivering under rubriken Business Case. Slutligen krävs ytterligare analyser för att bekräfta båda mätningarnas resultat, där exempelvis stickprovets storlek bör utökas eller metoden varierar.

# Contents

0.0.1	Sammanfattning. Sammanfattning . . . . .	1
0.1	Förord och huvudsyfte . . . . .	2
0.1.1	Tack . . . . .	3
0.1.2	Målgrupp och praktiskt . . . . .	3
0.2	Uppdragsbeskrivning . . . . .	3
<b>1</b>	<b>Teoretisk del</b> . . . . .	<b>4</b>
1.1	SSYK . . . . .	4
1.1.1	Bakgrund . . . . .	4
1.1.2	Användning för Jobtech . . . . .	5
1.1.3	Användning av SSYK i samhället i stort . . . . .	5
1.2	Platsbanken . . . . .	6
1.2.1	Hur Platsbanken fungerar . . . . .	6
1.2.2	Platsbankens koppling till SSYK . . . . .	6
<b>2</b>	<b>Undersökande del</b> . . . . .	<b>8</b>
2.1	Algoritmer på Jobtech . . . . .	8
2.1.1	Nuvarande Algoritm, text2ssyk . . . . .	8
2.1.2	Bert Classifier . . . . .	8
2.1.3	SSYKGPT . . . . .	9
2.1.4	Pricksäkerhet . . . . .	9
2.2	Analys 1 . . . . .	9
2.3	Människans pricksäkerhet . . . . .	9
2.3.1	Inledning . . . . .	9
2.3.2	Utförande . . . . .	10
2.3.3	Resultat . . . . .	10
2.3.4	Diskussion . . . . .	11
2.4	Experiment 2 . . . . .	13
2.5	Berts pricksäkerhet . . . . .	13
2.5.1	Inledning . . . . .	13
2.5.2	Resultat . . . . .	14
2.5.3	Diskussion . . . . .	17
2.6	Business Case:Skulle algoritmer kunna underlätta SSYK-användandet och spara samhället pengar? . . . . .	18
2.7	Vidare forskning . . . . .	19
2.8	Appendix . . . . .	19
2.8.1	Fler än två klasser . . . . .	22

## 0.1 Förord och huvudsyfte

Huvudsyftet med rapporten är att ge JobTech ett underlag för att bedöma om SSYK-etiketteringen av jobbannonser på Platsbanken bör automatiseras eller att fortsätta nuvarande arbetsuppgifter manuellt.

Rapporten expanderar JobTechs nuvarande kunskap genom

- Utförandet av ett stickprov i syfte att undersöka hur bra människan är på att sätta rätt SSYK på en annons; en siffra JobTech tidigare varit ovetande om.
- Sammanställer Job Techs nuvarande algoritmer för SSYK-sättning och anger precisionen för samtliga, vilket sedan jämförs med människan
- Ytterligare ett stickprov i syfte att undersöka hur korrekt den nuvarande angivna precisionen av den bästa algoritmen är.
- Går igenom vissa aspekter av psykologiska samt ekonomiska fördelar för samhället i stort genom att låta algoritmer sköta SSYK-etikettering istället för människan

### 0.1.1 Tack

Jag vill rikta ett extra stort tack till Jonas Södergren som initierade uppgiften, beskrev uppdragets behov tydligt, vägledde mig samt gav den informationsbas som faciliterade rapporten. David Norman som påvisade ett stort engagemang och hjälpsamhet när han exempelvis, utan tvekan, gick igenom 100 stycken annonser samt att bidra till tolkningen av resultatet. Jonas Östlund som möjliggjorde de båda stickproven, förklarade hur metoden för mätning av algoritmernas pricksäkerhet fungerar, samt var otroligt hjälpsam, exempelvis när han visade sin algoritm. Slutligen, Per Weinitz, som har väglett mig mycket, gett mig ett varmt välkomnande till JobTech och alltid varit behjälplig vid funderingar.

### 0.1.2 Målgrupp och praktiskt

Rapporten är tänkt att kunna läsas av individer som ej är insatta i SSYK och därför inleds rapporten av en *teoretisk del* innehållande flera förklarande moment. Är läsare insatt i vad SSYK är, vad det används till och hur platsbanken fungerar kan denne med fördel hoppa till rapportens *undersökande del*. Denna utgörs av rubrikerna om Jobtechs algoritmer, rapportens två experiment samt business-caset. Rapporten är tänkt att skapa en röd tråd genom att i sin undersökande del gå igenom hur bra människor är på att sätta korrekt SSYK. Därefter hur bra algoritmerna är och använda resultaten av de båda undersökningar som underlag för att undersöka vilken metod som är optimal.

## 0.2 Uppdragsbeskrivning

Utvärdera befintliga prototyper för att förbättra den svenska arbetsmarknaden genom att öka korrektheten av uppmärkning av SSYK-koder i jobbannonser

Uppgiften består i att hjälpa Jobtech utvärdera befintliga prototyper som har utvecklats för att förbättra den svenska arbetsmarknaden genom rätt uppmärkning av SSYK-koder i jobbannonser.

Syftet med uppdraget är att samla in data om vilken metod som är mest effektiv och mest användbar, och utvärdera vad effekten på arbetsmarknaden skulle kunna bli vid en sådan förändring.

Din uppgift kommer att vara att utvärdera prototyperna och dokumentera resultaten, såsom graden av korrekt SSYK-kodning, noggrannhet och praktisk användbarhet. Du kommer också att identifiera eventuella brister och föreslå förbättringar som kan göras för att öka prototypernas effektivitet.

För att klara av uppdraget ska du ha erfarenhet av databehandling, matematisk modellering och analys. Uppgiften underlättas av erfarenhet inom området. Vi erbjuder dig möjligheten att bidra till att förbättra och forma den svenska arbetsmarknaden. Du kommer att ha vårt fulla stöd och samarbete under hela uppdragstiden.

# Chapter 1

## Teoretisk del

### 1.1 SSYK

#### 1.1.1 Bakgrund

Standard för svensk yrkesklassificering (SSYK) är ett system som grupperar individens yrken eller arbetsuppgifter. Det är tänkt att användas inom arbetsmarknads- och individstatistiken och anses vara ett viktigt verktyg på arbetsmarknaden. Ett av de främsta användningsområdena för SSYK är att organisera information om lediga jobb, jobbsökande och yrkesgrupper. Företag och organisationer kan teoretiskt även använda det för att kategorisera sina anställdas kompetenser. Däremot är det vanligt att andra klassificeringsverktyg används intern på myndigheter och företag. När det kommer till den praktiska användningen av SSYK är informationen bristfällig, men verktyget kan enligt intervjuer betraktas som relativt oanvändbart och byråkratiskt.

Det finns två versioner av SSYK, kallade SSYK 96 och SSYK 2012. Sedan 2014 är det SSYK 2012 som används inom officiell statistik. Denna version är en moderniserad uppdatering av SSYK 96 och har ersatt denna. Anledningen till uppdateringen var att man behövde en yrkesindelning som bättre reflekterar den nuvarande yrkesstrukturen i Sverige. Däremot är alltså den senaste utgåvan elva år gammal, och många jobbtitlar har tillkommit sedan dess. Förändringarna var omfattande och det innebar ett tidsseriebrott, eftersom SSYK 96 och SSYK 2012 inte är helt kompatibla. SSYK 96 användes i statistiken mellan 2001-2013.

SSYK 2012 baseras på den internationella yrkesklassifikationen ISCO-08 (International Standard Classification of Occupation 2008), som är framtagen av den Internationella arbetsorganisationen (ILO). ISCO-08 är i sin tur en uppdatering av ISCO-88 och EU-varianten ISCO-88 (COM), som var grunden för SSYK 96. Målen med ISCO är många, samtidigt som det finns flera överlappningar med SSYKs syfte. Ett mål är att ISCO ska fungera som en mall för nationella och regionala yrkesklassifikationer. Dessutom är det tänkt vara användbar för länder som inte har sina egna nationella yrkesklassifikationer och ge en aktuell och relevant grund för internationell rapportering, jämförelser och utbyte av yrkesinformation. Detta kan skapa frågan varför Sverige inte valt att använda ISCO istället för SSYK till att börja med, då detta hade underlättat för internationella jämförelser av Sveriges statistik, samtidigt som sparat in nationella resurser vid uppdateringar av taxonomin.

Den fyrsiffriga koden, eller "placeringen" av ett yrke grundar sig på arbetsuppgifter, arbetserfarenhet och nivån på kvalifikationer. Ett yrke identifieras huvudsakligen utifrån arbetsuppgifter och ansvarsområden som en individ, antingen som anställd eller som egen företagare, förväntas utföra. Kvalifikationer representerar de nödvändiga kunskaperna och färdigheterna som behövs för att utföra arbetet.

Nivån på kvalifikationer bedöms genom att man tittar på den formella utbildning som vanligtvis är relaterad till yrket, definierat enligt ISCED (International Standard Classification of Education). Men detta betyder inte att en formell utbildning alltid är nödvändig för att erhålla de färdigheter och kunskaper som krävs för ett specifikt jobb. Dessa kan också förvärfvas genom arbetserfarenhet och kompetens. Här syftas på de kvalifikationer som jobbet kräver, inte nödvändigtvis de kvalifikationer som den anställda

besitter.

Kvalifikationsnivån används huvudsakligen som ett kriterium på en grov nivå, medan specialisering ofta leder till mer detaljerade nivåer. ISCO-08 och SSYK 2012 har definierat fyra kvalifikationsnivåer enligt ISCED 1997.

För exempel på statistik baserad på SSYK, se appendix.

### 1.1.2 Användning för Jobtech

Idag fyller SSYK inte något väsentligt internt användningsområde för JobTech. Istället används koncept-ID i hög grad. SSYK saknar information som är väsentlig för en jobbannons inom JobTech, framförallt vilka förmågor en arbetssökande bör besitta för att kunna vara lämplig kandidat för arbetet. Man kan säga att koncept-id från Job Techs taxonomi är mer informationsrik och detaljerad än vad SSYK är. IDt är baserad på ett nätverk av noder, och kan innehålla information såsom utbildningsgrad och huruvida körkort krävs för jobbannonsen.

Varje lagrad post i JobTechs taxonomidatabas har ett unikt ID som kallas koncept-ID. Taxonomidatabasen innehåller ett antal tabeller, däribland tabellen över SSYK-koderna - varje SSYK-kod är en post i denna databas, så därför har varje SSYK-kod också ett koncept-ID.

### 1.1.3 Användning av SSYK i samhället i stort

- Standardisering och jämförbarhet: SSYK är tänkt att vara en standardisering av yrken, vilket i sin tur möjliggör jämförelser över tid och mellan olika sektorer eller geografiska områden. Detta underlättar i analysen av arbetsmarknadstrender och i utvärderingen av politikens effekter. Risker är dock att fel SSYK sätts på en annons på grund av slarvfel, eller att taxanomin inte är uppdaterad för dagens arbetsmarknad vilket leder till att annonsörerna tvingas sätta ett SSYK som inte passar. Detta kan leda till att det ser ut som att Sverige har fler eller färre yrken inom en viss kategori än vi egentligen har. I slutändan står denna statistik, om inte praktiskt så i alla fall teoretiskt, som beslutsunderlag för arbetsskapande organ.
- En konsekvens av uppdaterad taxonomi för SSYK eller felaktig etikettering kan alltså vara att exempelvis fler YH-utbildningar än nödvändigt skapas inom en viss yrkesgrupp.
- Utbildning och arbetsmarknadens behov: Genom att analysera data baserad på SSYK kan utbildningsinstitutioner och politiska beslutsfattare identifiera vilka yrken eller kompetenser som är mest efterfrågade på arbetsmarknaden. Detta kan i sin tur leda till utveckling av utbildningsprogram eller arbetsmarknadspolitik som är mer anpassade till arbetsmarknadens behov.
- Arbetsmarknadstjänster: SSYK används också inom Arbetsförmedlingen för att organisera information om lediga platser och arbetssökande. Detta hjälper till att matcha arbetslösa individer med lämpliga jobbomöjligheter.
- Företag och organisationer: Inom organisationer kan SSYK användas för att kategorisera de anställda, vilket kan hjälpa till i planeringen och hanteringen av personalresurser.
- SSYK försäkrar att annonser som språkligt kan utgöra två olika jobb istället blir ett jobb. Exempelvis får "mjukvaruutvecklare" och "JAVA-developer" samma SSYK, istället för att de betraktas som två olika yrken vilket underlättar kategorisering.
- Skulle SSYK försvinna behöver mycket av dagens infrastruktur som redan används göra om
- SSYK används internt för statistik inom AF, men ej av SCB
- Sverige hade istället kunnat använda ESCO, vilket hade underlättat jämförelser med övriga Europa samt att man hade kunnat ta del av ett större ramverk. ESCO fungerar snarlikt SSYK.
- En nackdel med SSYK är att alla jobb ej finns representerade eftersom taxanomin inte uppdaterats sedan 2012.

## Användning för SCB

Efter en intervju kom en bekräftelse på att SSYK systemet spelar en betydande roll i praktiken för SCB.

SSYK används bl.a. för att koda yrken i SCB:s interna undersökningar AKU (arbetskraftsundersökningarna) samt ULF (Undersökningar av levnadsförhållanden) men även olika undersökningar som SCB får in på uppdrag. I undersökningen Yrkesregistret, som är en enkät som skickas till företrädesvis mindre företag, ber SCB företagen fylla i de anställdas huvudsakliga arbetsuppgift med hjälp av SSYK-koder. SCB har även löneundersökningar där man frågar efter de anställdas huvudsakliga arbetsuppgift och där företagen, förutom lön, arbetade timmar etc, också ska fylla i personalens arbetsuppgifter med SSYK-koder.

Exempel: I AKU och ULF-undersökningarna, som är individundersökningar och genomförs via intervjuer, så gäller en av frågorna i formuläret vilket yrke/huvudsakliga arbetsuppgifter undersökningspersonen har. I de fall intervjuarna inte kan hitta rätt yrke i den yrkeslista som de har som stöd i formuläret så måste de istället skriva en klartext där arbetsuppgifterna beskrivs så detaljerat som möjligt. Dessa "poster" med klartexter samt övrig info om undersökningspersonens arbetsplats skickas sen till oss i kodningsgruppen för att SCB ska utreda och koda yrket, baserat på den information som intervjuaren lämnat om intervjupersonens arbetsuppgifter. Det innebär att SCB måste ha expertkunskap om koderna för att kunna sätta korrekt SSYK-kod, vilket SCB alltså gör mer "manuellt" i det system SCB jobbar i.

Ett exempel kan vara att UP (undersökningspersonen) är läkare. Det finns 4 olika koder för läkare beroende på om man är specialiserad, ST- eller AT-läkare samt underläkare. Intervjuaren ska inte vara expert på SSYK-koder, därför har de endast tillgång till en begränsad yrkeslista att söka efter koder i, detta då många yrkesbenämningar har flera alternativ på koder. Läkare är ett sådant exempel, alltså finns den inte i intervjuarnas lista. Här måste intervjuaren fråga UP om arbetsuppgifterna och skriva en klartext till oss. SCB som jobbar med koder på heltid har tillgång till ett komplett index plus att SCB ska kunna skillnaderna mellan de olika koderna. Om SCB kan se av intervjuarens info att personen är specialistläkare inom något område sätter SCBi SSYK-kod 2211. En ST-läkare har istället kod 2212, en AT-läkare 2213 och en underläkare 2219.

## 1.2 Platsbanken

### 1.2.1 Hur Platsbanken fungerar

Platsbanken är en tjänst som drivs av Arbetsförmedlingen i Sverige och det är en av landets största jobbsajter. Dess primära syfte är att koppla samman arbetssökande med arbetsgivare som har lediga tjänster.

Arbetssökande kan söka efter jobb baserat på olika kriterier som yrke, plats, anställningstyp, och mer. Arbetsgivare kan publicera sina lediga tjänster och bläddra bland CV:n för att hitta lämpliga kandidater. Tjänsten är gratis för både arbetssökande och arbetsgivare.

### 1.2.2 Platsbankens koppling till SSYK

Platsbanken har som krav att alla annonser etiketteras med SSYK av annonsörerna. Detta används sedan som grund för testdatan algoritmerna använder. Joblinks, som har annonser från utomstående sidor, har inte detta krav. I fallet Joblinks sätts istället annonsernas SSYK av algoritmen text2ssyk.

Teoretiskt sett ska Platsbankens koppling till SSYK medföra fördelar. Exempelvis ska det underlätta för arbetsgivare att specificera vilken typ av roll de behöver fylla samtidigt som det hjälper arbetssökande att söka efter jobb som matchar deras färdigheter och erfarenheter.

Genom att använda SSYK i Platsbanken kan olika aktörer ha en standardiserad metod för att kategorisera och söka efter jobb. Detta kan även resultera i att arbetsmarknadsdata blir mer jämförbara och analyserbara, vilket i sin tur kan påverka politiska beslut och forskning om arbetsmarknaden.

Platsbanken och Joblinks täcker 50% respektive 30% av totala antalet jobbannonser i Sverige. Joblinks sysslar med inhämtning av annonser från externa webbsidor såsom Monster och Blocket Jobb. Alltså täcks totalt omkring 80% av alla arbetsannonser i Sverige av Platsbanken. Observera att dessa siffror är grova estimat. Det exakta värdet av hur många jobbannonser som finns i Sverige idag är svårt att säga. Därmed uppstår svårigheter vid kvantifiering av platsbankens överlappning med denna helhet.



# Chapter 2

## Undersökande del

### 2.1 Algoritmer på Jobtech

På JobTech återfinns idag minst tre algoritmer för etikettering av SSYK. Dessa är Bert Classifier, SSYKGPT och text2ssyk, varav den sistnämnda har flera olika modeller av sig.

Följande tabell visar ett utdrag av ett mer omfattande utvärderingsresultat, vilket återfinns i appendix.

Bert Classifier	SSYKGPT	text2ssyk (i bruk)	text2ssyk (bästa)
86%	53%	59%	78%

#### 2.1.1 Nuvarande Algoritm, text2ssyk

I dagsläget används text2ssyk för att etikettera jobbannonserna på platsbanken som kommer från externa webbplatser, och därmed inte har en ursprunglig etikett av arbetsgivaren. Versionen som är i bruk, alltså 0.1.0, har en pricksäkerhet på 59%, vilket kan tyckas vara lågt. Däremot finns ett flertal versioner av den, med pricksäkerhet som varierar inom intervallet 49-78%, vilket redovisas i nedanstående tabell.

Version:	0.1.0	config0	config1	config2	config6	sbert-cov
Precision:	59%	66%	73%	74%	78%	49%

Högst precision av dessa har "config6". Den skapades genom att dels justera diverse parametrar för algoritmen, och dels genom att öka dess träningsdata. Version 0.1.0 var tränad på tre hundra tusen annonser. Motsvarande siffra för config 6 är två miljoner. Detta förklaras av att träningsdata av hög kvalitet är väsentligt AI och maskininlärningssammanhang. Alltså existerar korrelationen mellan ökad mängd träningsdata och skarpere precision hos text2ssyk eftersom algoritmen är baserad på testdata från mänskliga klassifikationer, som kan vara felaktiga. Ju större mängd träningsdata, desto mindre blir effekten av den andel annonser där ssyk-etiketteringen är felaktig. Vikten av att undersöka algoritmens beteende vid olika parametrar belyses även av ovanstående tabell. I merparten av de olika versionerna var detta den enda varierande faktorn, inte träningsdatan. Nämnvärt är även att config6 kräver betydligt mer högre RAM-kapacitet, närmare tiofaldigt. Under business case diskuteras om en sådan investering är lönsam, för att öka precisionen.

#### 2.1.2 Bert Classifier

Den av Jonas Östlund utvecklade algoritmen benämnd Bert Classifier utgår från ett neuralt nätverk och transformers. Bert(Bidirectional Encoder Representations from Transformers) har hittills nått en

imponerande pricksäkerhet eller "accuracy" på 86%, vilket är det högsta uppmätta värdet hittills av JobTechs algoritmer. Det hör även till saken att detta är på en begränsad mängd övningsdata, ju mer Bert tränar, desto bättre blir precision. Dessutom är den endast testad med en inställning av parametrar. Skulle den fortsätta att experimenteras med flera olika kombinationer av parametrarnas inställningar, samt låta Bert träna på mer data kan precisionen komma att höjas avsevärt. Ett flertal test av olika parametrar krävdes för denna avsevärda förbättring.

### 2.1.3 SSYKGPT

SSYKGPTs precision testades i samband med denna rapport. Således har algoritmen, likt Bert, inte erhållit testning och blivit anpassad för att höja precisionen; den var snarare ett experiment för att utreda den tekniska möjligheten till användningen av en ChatGPT baserad algoritm. Hade ChatGPT använts i dagsläget, när begränsningarna på inmatningen expanderats, hade algoritmen programmerats på annat sätt. Hade den gjorts för att maximera precision, hade den likt Bert krävt mer testdata och justeringar av algoritmens parametrar. SSYKGPT har en precision på 53% i dagsläget, vilket är nära den version av text2ssyk som brukas idag med en precision på 59%. Därmed är det inte orimligt att justering av parametrarna bör leda till en åtminstone med text2ssyk jämförbar uppgång. Således på närmare 20 procent. Vidare info om SSYKGPT återfinns i appendix.

Författaren har med stor framgång, vilket tas upp i experiment 2, använt ChatGPTs fjärde version åt etikettering. Utifrån stickprovet att bedöma, etiketterade den likt författaren i 15 av 16 fall. Detta kan påvisa att en ny algoritm baserad på den senaste versionen av ChatGPT, som dels är smartare samt tillåter en större inmatning, kan vara intressant att undersöka. En annan aspekt är att Bert har en så hög precision rate, vilket framkommer senare i experiment 2, att författaren inte rekommenderar att fortsätta lägga ned tid och resurser på SSYKGPT. Istället borde fokus läggas på den bästa, som även har stor potential till förbättring.

### 2.1.4 Pricksäkerhet

Tidigare data om precision har utgått från jämförelser av algoritmens svar gentemot människans svar, som även har agerat som facit. Ett antagande har varit att om algoritmen erhåller ett svar och människan ett annat, ska algoritmens svar vara felaktigt. Efter rapporten vet vi hur ofta människan har fel och huruvida denna metod är optimal eller ej, eftersom det kan förekomma fall då algoritmen har rätt och människan har fel. För att erhålla ett mer exakt värde har ett stickprov utförts när vi kontrollerar fall där algoritmen fått "fel" och djupdyker i om det faktiskt är fel, eller om det är människan som gjort fel. En annan viktig aspekt är att man enbart kan sätta en SSYK per jobbannons; men att vissa annonser söker efter flera olika yrkesgrupper. Detta kan leda till att algoritmen ser ut att sätta ett felaktigt SSYK, som egentligen är rätt men enbart annorlunda än människans. Vidare är taxonomin inte uppdaterad sedan 2012, vilket gör det väldigt luddigt vad som faktiskt är rätt SSYK och som tidigare så kan flera SSYK:s mer eller mindre korrekt sättas på en annons. Därmed är den nuvarande, binära metodiken bristfällig i sin angivelse om algoritmers precision. I appendix går det att läsa ett utdrag av Jonas Östlund som går djupare in i hur pricksäkerheten beräknas.

## 2.2 Analys 1

## 2.3 Människans pricksäkerhet

### 2.3.1 Inledning

Tidigare har JobTech varit ovetande om människans precision gällande etikettering av SSYK, men antagit att denna varit omkring 95%. Att kunna jämföra algoritmens pricksäkerhet med människans är centralt för att argumentera för huruvida etiketteringen ska ske manuellt eller per automatik.

Därför initierades experiment 1; för att få ett värde på hur pricksäker människan är vid etikettering av arbetsannonser. Därpå kommer detta värde jämföras med motsvarande värde för algoritmerna under rubriken business case.

Information: "Arbetsgivaren" och "människan" växlas fritt i rapporten.

### 2.3.2 Utförande

Experimentet utfördes genom att författaren, som kan klassas som en nybörjare när det kommer till SSYK-klassificering, etiketterade 100 stycken av Jonas Östlund slumpmässigt valda jobbannonser. David Norman utförde samma uppgift parallellt. När författarens och Davids svar angivits sammanställdes de i en tabell.

Därefter korrigerade David, sina misstag med hjälp av sina egna, författarens och arbetsgivarens angivna etiketter. De bästa valdes, vilket producerade ett förhoppningsvis felfritt facit.

Information: David är expert inom SSYK medan Jacob är en nybörjare.

### 2.3.3 Resultat

Excelarket innehåller flera olika intressanta mängder av rader, varifrån olika siffror på människans pricksäkerhet kan framställas.

- En delmängd av en större mängd illustreras av nedanstående bild. Den utmärks av att alla etiketterare fått samma SSYK för varje annons. Det finns en jobbannons per rad. I kolumnerna har, i ordning från vänster håll, arbetsgivaren, David och författaren satt sina respektive SSYKs.
- Denna mängd, där alla tre parter etiketterar lika, ger oss en pessimistisk skattning av människans precision. Detta skedde i 48 av 100 fall och ger en precision rate på 48%. Delmängden kallas för "garanterat korrekt".

96	IT Test Mar	Does the cl	2514	2514	2514
97	Utveckland	I tjänsten ir	5343	5343	5343
98	Arbetstera	Just nu sök	2273	2273	2273
99	Arbetsleda	Vi söker dig	3121	3121	3121
100	Personlig a	GIL står för	5343	5343	5343

- Arbetsgivaren anses ha "garanterat fel" när både författaren och David har fått en kod, men arbetsgivaren en annan. Åtta sådana fall förekom, vilket ger en precision rate på 92%. Detta värde utgör en övre gräns av precisionen. Delmängden illustreras av nedanstående bild, där arbetsgivaren återigen utgör kolumnen längst åt vänster, David den i mitten och författaren längst åt höger.

15	En unik mö	Om Doktor	2212	2219	2219
16	Administra	Om Bravur	4113	4222	4222
17	Biomedicin	Lunds univ	2313	3212	3212
18	IT Support	Guided by	2511	3512	3512

- Människans etikettering jämfört med Davids, där Davids värde agerar facit. Metoden utgår från att alla gånger arbetsgivaren satt något som inte överensstämmer med Davids facit, är det felaktigt. Detta gav en precision rate för människan på 78%.

Ytterligare intressanta delmängder förekom, men dessa anses inte kunna användas för att mäta precision.

- 14 annonser återfanns där alla tre parter genererat olika SSYK-etiketter
- När alla parter etiketterade SSYK blint återfanns 24 fall där David Norman och arbetsgivaren fått ett SSYK och författaren ett annat. Motsvarande siffra för David, alltså att författaren och arbetsgivaren fått ett SSYK och David ett annat, var 7.
- David Normans facit innehåller endast tre SSYK-etiketteringar där författaren och arbetsgivaren satt samma etikett och David varit själv.

## 2.3.4 Diskussion

### Bästa siffran

Garanterat rätt och garanterat fel ger oss ett intervall mellan 48-92% avseende precision av etikettering. En tredje siffra på 78% för arbetsgivarens precision fås då David agerar facit. Denna metod är den mest trovärdiga att använda eftersom David är expert inom SSYK. Till sin hjälp fick han författarens och arbetsgivarens värden, vilka levererades blint. Således hade alla tre parter tagit fram värden oberoende av varandra. David hade minst delmängd där han givit ett svar och de andra två fått samma. Denna siffra var 7 för David, 8 för arbetsgivaren och 24 för författaren. En tolkning av detta är att David är bättre än både arbetsgivaren och författaren på etikettering.

Vidare fungerade den blinda datamängden som stöd vid bildandet av facit. Det gav förslag och problematiserade Davids initiala värden. David kunde därför jämföras med författarens, sin egna och arbetsgivarens etikett för att ta fram den bästa. I de fall som det förekom när arbetsgivaren och författaren fick en SSYK och David ett annat var den blinda datamängden till stor hjälp. Att med hjälp av detta kunna rätta sina fel, i kombination med att etiketterna redan satts av en expert, leder till att ett nästan felfritt facit uppstår. Något som styrker kvaliteten av Davids facit är att det endast förekom tre fall där människan och Jacob båda höll med varandra och David var ensam med sitt svar, vilket utgjorde den minsta undergruppen av etiketteringarna.

### Felkällor

Idag etiketterar arbetsgivaren via DXA system, vilket står för direktöverförda annonser. Detta fungerar på en fundamental nivå likt JobTech Atlas, vilket var verktyget som användes i experimentet, men är inte snarlikt. Däremot anses systemen fungera så pass lika varandra att experimentets etikettering inte bör avvika från människan i övrigt. Utöver detta svarade både författaren och David i vissa av dem antingen blankt eller N/A. Detta kan påverka resultatet.

Bakgrundsinformation är att många arbetsannonser söker en bredare målgrupp än just en specifik yrkesgrupp. Exempelvis kan annonsen vara riktad mot sjuksköterskor för att få många sökande, men arbetsgivaren är enbart ute efter psykiatrisjuksköterskor. Vid en sådan annons kan två SSYK-etiketter passa lika bra. Därför måste det inte, trots att alla tre inte fått samma värden, nödvändigtvis betyda att en har fått fel i strikt bemärkelse. Däremot utgår experimentet ifrån att det alltid finns ett SSYK som passar arbetsbeskrivning bättre, eftersom detta är sant i de allra flesta fall.

### Svårigheter med SSYK

14% av annonserna hade etiketter där alla tre parter etiketter olika. Detta är en betydande och problematisk andel. Vissa mönster återfanns dock mellan dessa annonser, vilka i kombination kan förklara delmängden.

- Vissa arbetsbeskrivningar lägger fokus och tyngd på erfarenhet inom en bransch snarare än yrkestitel. Erfarenhet fångas inte upp av SSYK, utan endast utbildningsgrad. Ett exempel på problemet kan vara "Vi söker dig med 5 års erfarenhet inom rör- och anläggning..."
- Chefsyrken är svåra att sätta SSYK på, eftersom dessa ofta är branschöverskridande. Exempelvis förekom ett fall där arbetsannonsen beskrev en chef inom bygg och miljö, medan SSYK endast klassar för chefer inom bygg eller chefer för miljö. Här måste avgränsning göras som är snarlik med att singla slant. Att då påstå att en etikettering är korrekt blir felaktigt. Att hitta rätt SSYK i sådana fall är tidskrävande.
- Etiketteringen blir även svår när annonsen lägger stort fokus på kompetens inom någon nisch. Exempelvis när en "expert inom instrument för ventilationssystem" sökes. Att då välja exempelvis "servicetekniker" istället för "civilingenjör" motiveras då servicetekniker sysslar med sådana instrument. Detsamma gäller dock ingenjörer, vilka dessutom har en djupare teoretisk bakgrund avseende teknik och därmed kan vara mer påtagliga att kallas "experter".
- Till hjälp kan då arbetsgivare använda verktyget "skills", vilket återfinns i JobTech atlas och DAX-system. Där matchas olika kunskaper eller förmågor, som återfinns i arbetsbeskrivningen, mot jobbtitlar som alla har en lista med skills under sig. I vissa fall blir etiketteringen tydlig. Däremot kan detta även försvåra processen om arbetsbeskrivningen innehåller skills från en mängd olika yrken.

### Stickprovets storlek

Medan ett stickprov på 100 testade exempel kan ge betydande insikter, är det viktigt att notera att pricksäkerheten ofta förbättras med en ökning av provstorleken. Mindre provstorlekar kan påverkas av slumpmässig variation mer jämfört med större provstorlekar. Detta kan leda till felaktiga slutsatser då man utgår från en bristfällig storlek på datamängden.

- Om urvalet är slumpmässigt och alla punkter i datamängden har samma chans att bli valda ökar pricksäkerheten, vilket är fallet i detta experiment
- Generellt när det kommer till stickprov är det viktigt att ta den större datamängdens storlek i beaktning. Om datamängden är mycket stor, vilket är fallet, kan ett stickprov på 100 vara för litet för att ge tillförlitliga resultat.
- Variabiliteten inom datamängden spelar också en stor roll för pricksäkerheten. Om det finns stor variation inom datapunkterna kan ett stickprov på 100 exempel vara otillräckligt för att representera datamängden tillförlitligt. I detta fall anses de olika arbetsannonserna vara väldigt liknande.

### Herzbergs tvåfaktorsmodell kopplat till SSYK

Enligt Herzbergs tvåfaktorsmodell finns faktorer vilka försämrar arbetsklimatet på arbetsplatser, vilket leder till missnöjdhet bland anställda och därav lägre produktivitet<sup>1</sup>. För att analysera ifall SSYK-etiketteringen kan utgöra en hygienfaktor för anställda med uppgiften att etikettera undersöktes fallen där människan gjort fel.

Ett vanligt fel är att arbetsgivaren är väldigt nära att etikettera korrekt SSYK, men att den sista siffran av koden blir felaktig. Exempelvis förekom många fall där koden 2511 etiketterats istället för 2512. Detta motsvarar att "Systemanalytiker och IT-arkitekter m.fl." etiketterats istället för "Mjukvaru- och systemutvecklare m.fl." Detta kan antyda att kompetensen för de som sätter SSYK i vissa fall inte är tillräckligt branschspecifik för att etikettera korrekt. Även för de som jobbar inom branschen kan det vara svårt att bedöma skillnaden mellan dessa två etiketter. Att förvänta sig korrekt etikettering av någon som aldrig jobbat med sådant kan vara ett stressmoment.

<sup>1</sup><https://www.sciencedirect.com/science/article/abs/pii/0030507376900167?via%3Dihub>

förekommer en hygienfaktor att arbeta med monotona arbetsuppgifter, vilka anställda inte förstår syftet med, och därför väljer slumpbaserat.

Att vara oförstående om varför en syssla utförs eller att uppleva sin kompetens som otillräcklig för en syssla är alla faktorer vilka kan bidra till minskad trivsel på arbetsplatser; hygienfaktorer enligt Herzberg. Antagandet att en viss andel av de som sysslar med manuell SSYK-hantering känner detta är rimligt då detta indikeras av datan.

### Slutsats

- Människan har en precision rate någonstans mellan 48-92%. Den bästa estimeringen är att människan är på en nivå av 78%.
- SSYK-användandet kan innebära hygienfaktorer på arbetsplatsen

## 2.4 Experiment 2

## 2.5 Berts pricksäkerhet

### 2.5.1 Inledning

Därför initierades experiment 2; Anledningen till att just Bert algoritmen valdes var att den, som beskrivits nedan, är den överlägset bästa algoritmen avseende pricksäkerhet.

#### Syfte och bakgrund

Från experiment 1 blev det tydligt att människans precision var lägre än förväntat. Att då använda dessa SSYK-etiketter som facit då en algoritms precision mäts är problematiskt. Exempelvis kan en algoritm få rätt SSYK-etikett, samtidigt som det markeras som ett "fel" eftersom algoritmens SSYK-etikett överensstämmer med facit. I själva verket har människan i vissa av dessa fall satt fel SSYK-kod. Resultatet blir att det i precisionsmätningar ser ut som att algoritmen får fel, när i själva verket facit är felaktigt.

Följaktligen kan den reella precisionen hos Bert vara betydligt högre än 86%. Syftet med experiment 2 är att undersöka hur rimligt detta är i fallet Bert Classifier; kan man säkerställa en algoritms precision via denna metod? Vad är den egentliga precisionen och hur förhåller den sig till människans precision? Anledningen till att just Bert valdes var att den, som ovanstående beskrivits, är den överlägset bästa algoritmen.

Ett annat sätt att se saken är att Bert-classifier enligt tidigare tester antas sätta fel SSYK-kod i 14% av fallen. Om det istället är människan som satt fel i majoriteten av dessa fall, minskar andelen felaktigt satta SSYK-etiketter avsevärt.

#### Utförande

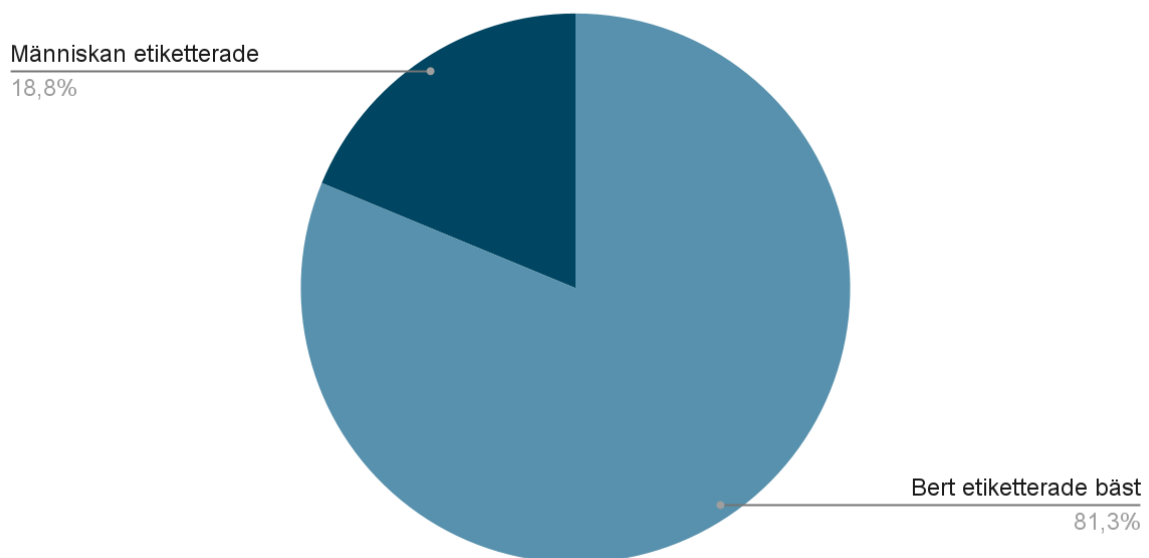
För att utreda detta skapades ett Excel-dokument utgjort av alla felen från Bert Classifiers träningsdata. Filen innehöll jobbtitel, beskrivning av jobbet, människans SSYK och Berts SSYK för tiotusentals annonser i slumpmässig ordning. En djupdykning gjordes i 25 av dessa där varje annons utreddes grundligt för att analysera om antingen människans eller Berts etikettering ansågs mest korrekt. Det skulle visa sig vara en betydligt mer tidskrävande process att gå igenom dessa annonser än motsvarande mängd genomsnittliga annonser från experiment 1.

Vid undersökningens start återfanns endast arbetsannonser kopplade till två olika SSYK-etiketter; människans och Berts. Därefter gjordes en djup analys kring varje annons för att bestämma vilken av de två etiketterna som var lämpligast.

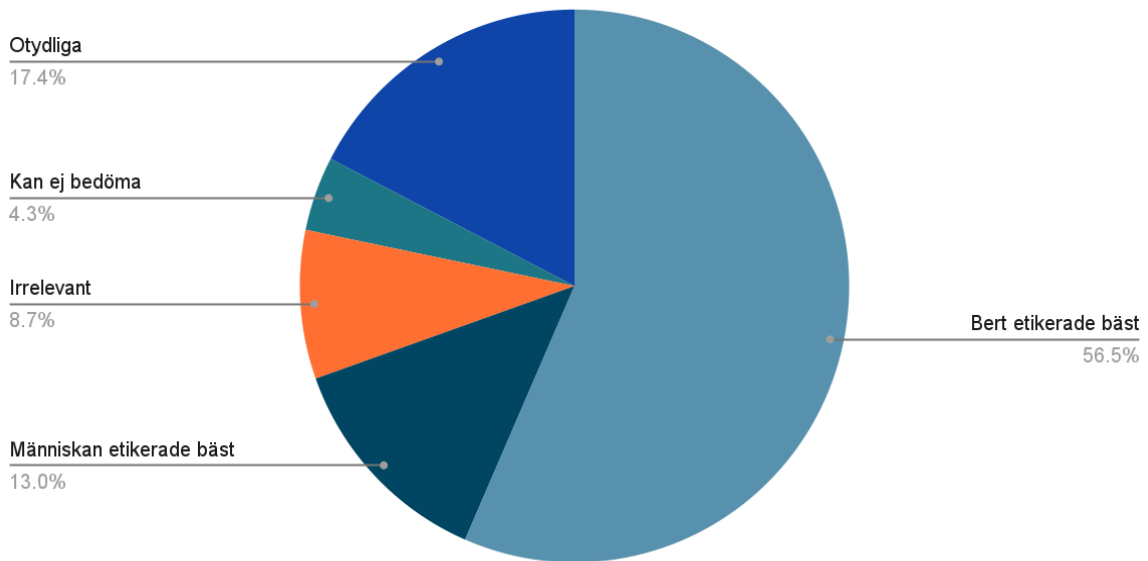
För att få en andra åsikt nyttjades den fjärde versionen av ChatGPT, betalversionen. Utfördes med att kopiera arbetsbeskrivningen in till ChatGPT, följt av frågan “vilken jobbtitel passar arbetsbeskrivningen bäst, (arbetstitel till SSYK av människan) eller (arbetstitel till SSYK av Bert). Anledningen var att någon form av extra kontroll ansågs vara fördelaktig och intressant.

## 2.5.2 Resultat

En jämförelse av storleken på andelarna där Bert etiketterar bäst jämfört med motsvarande siffra för människan



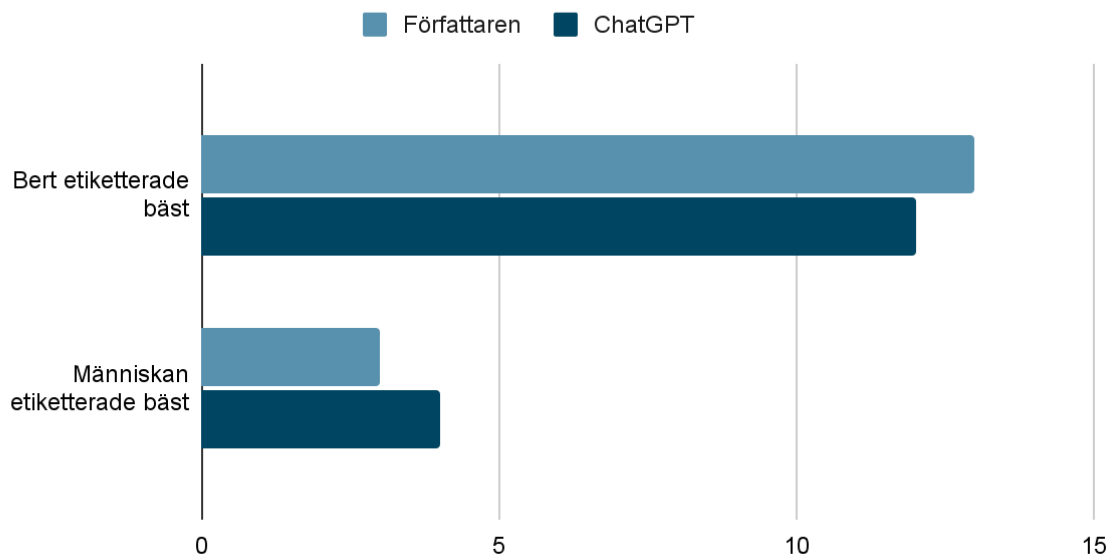
## Hur författaren kategorisererat 25 annonser där Bert etiketterat "felaktigt"



Notera

att nedanstående tabell, olikt cirkeldiagrammet, innehåller ChatGPTs svar.

## Författarens etikettering i jämförelse med fjärde versionen av ChatGPT



- I vissa fall har "lika bra" använts. Här placeras annonser där flera SSK:s passar eftersom de är breda och det inte finns ett föredraget alternativ mellan människans och Berts etikettering.



Exempelvis fick Bert fram "Civilingenjörsyrken inom elektroteknik" och människan "Ingenjörer och tekniker inom elektroteknik". Annonser tillät högskoleingenjörer men ansåg att en titel som civilingenjör vore meriterande. Många liknande exempel, där annonserna riktar sig till en bred grupp av yrken, samt fall med betydande överlappning och det ej går att finna ett föredraget alternativ hamnar här.

- Två annonser innehöll ingen jobbtitel eller arbetsbeskrivning. I båda dessa fält stod det istället "tillsatt". Beslutet att ej exkludera dessa fall ur studien baseras på syftet; att avgöra hur rimligt det är att ange "precision rate" som ett mått på hur duktig Bert är. Ifall det förekommer annonser där den enda informationen som förekommer är "tillsatt" och Bert tvingas sätta ett SSYK på detta, säger det inget om Berts förmåga att sätta rätt SSYK. Därav blir resultatet något missvisande.
- Författaren och ChatGPT fick samma svar i alla undersökta fall förutom ett. Detta talar för hur duktiga algoritmer av detta slag är på just språkhantering.
- Den mänskliga faktorn, som inkluderar uppenbara slarvfel där etiketteringen absolut inte hör till annonsen, görs inte av Bert. Ett sådant fall hittades.
- Ett av de 25 undersökta fallen markerades som "kan ej bedöma". Djurskötare, som människan satt, passade inte eftersom arbetssysslorna inte hade något med djurskötsel att göra. De hade enbart telefonsamtal och inbokningar av djur till djurskötare att göra. Däremot krävdes kompetens inom djurskötsel, och ett års arbete på djurklinik. Detta fångades inte upp av Berts etikettering som kundtjänstpersonal. Den matchar arbetssysslorna bättre men missar det väsentliga kravet om erfarenhet från djurklinik. Både författaren och ChatGPT ansåg att båda SSYK:s etiketter passade dåligt. Efter utredning hittades inget SSYK som var uppenbart lämpligare än ett annat. Bilden nedan är inkluderad för att påvisa ChatGPTs förmåga att resonera där dens svar på denna fråga används som exempel:



Den jobbbannons du beskriver inkluderar både kliniska och administrativa uppgifter, vilket gör det svårt att placera den i en strikt kategori. Eftersom rollen inkluderar att hantera remisser, koordinera bokningar, erbjuda kundservice och samtidigt kräver klinisk erfarenhet inom djursjukvård, verkar det som om det behövs en unik kombination av färdigheter.

Om vi måste välja mellan "Djursjukskötare" och "Kundtjänstpersonal" skulle "Djursjukskötare" vara närmare eftersom den innebär en viss mängd klinisk erfarenhet, vilket är ett av kraven för jobbet. Men det bör noteras att jobbet också innebär mycket kundserviceuppgifter.

En bättre titel kanske vore "Remisskoordinator med klinisk erfarenhet" eller "Djursjukskötare med kundserviceansvar". Dessa titlar återspeglar både de kliniska och kundservicebaserade aspekterna av positionen.

Beräkning av den nya precisionen:

- Vi exkluderar de som ansågs lika bra, irrelevanta och inte kunde bedömas. Då blir andelen korrekta bland felen  $13/16$ . Detta betyder att bland de felen, som utgjorde 14%, så var egentligen  $13/16$  korrekt.
- $13/16 * 0,14 = 0,11375$ . Detta läggs ihop med 0,86% som var bekräftat korrekt.  $0,86 + 0,11375 = 0,97375$
- Alltså är den nya precisionen, enligt författarens mått ungefär 97,4%!
- Efter samma beräkning med ChatGPTs värden, som utgör det undre intervallet, får vi en precision på 96.5%
- Observera att "precision rate" på 86%, vilket användes i beräkningen, är ett avrundat tal.

- Båda dessa resultat är högre än vad människan har för precision enligt experiment 1.

### 2.5.3 Diskussion

Att endast 25 annonser undersöktes var en brist, men gjordes i mån av tid. För ett mer pricksäkert resultat rekommenderas att urvalsgruppen ska vara större. Däremot pekar resultatet i en tydlig riktning; att Bert är exaktare än människan. Detta kommer troligtvis inte ändras om urvalsgruppen ökar i storlek dels eftersom skillnaden är så stor, och dels eftersom annonserna var helt slumpmässigt utvalda.

ChatGPT visade sig ha en förvånansvärt djup kunskap om alla olika branscher och tillhörande jobbtitlar, något som kan vara svårt för en människa att uppnå. Vidare påvisade ChatGPT en förmåga att resonera kring och motivera tydligt och relevant kring vilken jobbtitel som hör ihop bäst med en jobbannons.

Att arbeta med dessa annonser, där Bert anses ha gjort fel, var betydligt mer tidskrävande än att sätta SSYK-etiketter på slumpmässiga annonser. Detta på grund av att i fallen där Bert satt "fel", är det vanligt att andra människor sätter fel SSYK. Anledningarna till detta kan vara flera. Till att börja med innehöll vissa av dessa arbetsannonser mycket information och ofta luddig sådan. Att Bert får rätt oftare än människan i dessa fall kan förklaras av att algoritmen är bättre på att tolka stora textfiler. Detta ser vi i flera fall där människan verkar ha glömt bort viktiga detaljer, som Bert inte missat i textmassan, återfanns.

Vidare finns ett vanligt förekommande mönster hos människan bland dessa annonser; att etikettera arbeten som chefsjobb när de egentligen inte är det och vice versa. exempelvis kan "chefsförsäljare" väljas istället för "företagsförsäljare", när annonsen inte innehåller någon information alls om gäller en chefsposition. Författarens hypotes är att detta beror på att människan sökt i databasen och försökt finna rätt SSYK men tyvärr misslyckats. Vidare medvetet valde en SSYK som innehöll felaktig information eftersom man inte haft tid, korrekt information eller energi att etikettera korrekt. Bert är inte präglad av dessa svagheter utan istället snabbt och effektivt kan söka sig igenom dels stora och svårhanterliga arbetsbeskrivningar, men framförallt genom hela databasen för SSYK.

Vidare krävs en stor mängd branschspecifik kunskap för att etikettera korrekt, exempelvis har två annonser förekommit där man förutsätts veta skillnaden mellan "Mjukvaru- och systemutvecklare" och "Systemanalytiker och IT-arkitekter". Människan har i samtliga sådana fall etiketterat felaktigt. Bert har vid dessa fall etiketterat korrekt. Fakta är att det finns en betydande överlappning mellan de två arbetstitlarna. En ytterligare anledning till varför processen för dessa annonser är tidskrävande och mödosam är att annonser där taxonomin för SSYK, som kom 2012, ej är uppdaterad för dessa nya yrkestitlar. Scenariot kan liknas med att tvingas välja mellan det ena eller det andra men inte veta vilket som är korrekt, men ändå finns det något som passar något bättre än det resterande. Här visar resultatet att Bert har en edge över människan när Bert tolkar fler korrekt.

Slutligen utgjordes vissa av dessa annonser, där etiketteringen skiljde sig, av människor som uppenbart inte är experter på SSYK. Ett exempel var en sushirestaurang som arbetsannonsen, som var dåligt skriven och full av stavfel, sökte en kock samtidigt som människans SSYK-etikettering var för servitör. I dessa fall lyckas Bert när människan är okunnig och misslyckas med rätt yrke.

### Experiment 2 Slutsats

Berts precisionrate är inte 86%, utan den befinner sig i intervallet 96,5-97,4%. Medelvärdet av dessa är 96,95 alltså ungefär 97%.

## 2.6 Business Case:Skulle algoritmer kunna underlätta SSYK-användandet och spara samhället pengar?

### Inledning

Business Case har som syfte att argumentera för för- och nackdelar av automatiseringen av SSYK-etiketteringen. Slutmålet är att kunna ge en tydlig recommendation för Jobtech genom att kvantifiera för- och nackdelarna och väga dessa mot varandra.

### Business case

.Som vi ser från experiment 1 och 2 har Bert en högre precision än människan. Bert med 97% jämfört med människans 78%. Således har Bert en högre precision än människan med 19%. Att Bert är bättre än människan anses vara utom rimligt tvivel och är i linje med vad nya rön kring denna typ av AI påvisar<sup>2</sup>. Därmed är det lämpligt att byta till en automatiserad process enbart för att ge möjlighet till att generera mer exakt statistik åt Platsbanken. Detta kan i sin tur komma att ge beslutsfattare i offentliga organ bättre data, vilket bör öka precisionen. Ett exempel är när nya platser och yrken på YH-utbildningar ska matcha arbetsmarknaden bättre.

En ytterligare fördel är att företag härmed kan spara in både tid och resurser. Varje dag publiceras ungefär 2000 annonser på platsbanken(enligt `grep VIA_PLATSBANKEN_DXA 2022.jsonl | wc -l | awk '{printf "%.2f\n", $1/365}' => 2090.87`). Det är rimligt att påstå att genomsnittstiden är 2 minuter per annons. Siffran baseras på personlig erfarenhet; vissa tar längre tid än andra. En uppskattning är att det läggs ungefär 67 timmar per dag av diverse anställda på arbetet med SSYK etikettering. En överslagsberäkning; Om vi antar att en HR-anställd får 150 kr i timmen, sparar samhället in ungefär 10.000 kronor varje dag. Detta blir en summa på över tre och en halv miljon kronor per år. Summan överskrider definitivt JobTechs driftskostnader, och införsäkring av eventuellt tillskott i form av RAM- och GPU kraft.

Något som är svårare att kvantifiera men värt att nämna är att förändringar från manuell till automatiserad etikettering hade inneburit avlägsnandet av arbetsmiljöförsämrande faktorer från bolag riket över. När anställda slipper de monotona uppgiften bör de istället prestera bättre i andra avseenden. En fördel kan alltså vara att HR-personal slipper känna oro för ifall de satt korrekt SSYK-etikett, eftersom Bert, som har högre precision och sköter den processen och arbetsuppgiften.

Om vi istället tar det orimligt pessimistiska värdet för Bert, som är 86%, och jämför det med det ännu mindre rimliga men optimistiska värdet för människan om 92%, ligger dessa värden nära varandra. Detta kan fungera som sanity check; det sämsta värdet för Bert och det bästa värdet för människan skiljer endast 6% enheter ifrån varandra. Även om detta innebär en högre precision för människan, hade denna data för sig, öppnat för frågan om det är lämpligare att låta Bert sköta etiketteringen. Dels för att beslutet hade sparat på tid och därmed pengar, dels på grund av hygienfaktorer på arbetsplatsen.

### Slutsats

- Användandet av Bert kan leda till minskade kostnader på 3.5 miljoner kronor per år
- Dessutom blir statistiken, som utgår från SSYK, exaktare, vilket mynnar ut i många svårberäknliga men samtidigt markanta, fördelar
- Bert användandet kan även medföra psykologiska fördelar
- Kombinationen av fördelarna överväger eventuellt ökade kostnader i form av exempelvis RAM

---

<sup>2</sup><https://www.refuel.ai/blog-posts/llm-labeling-technical-report>

## 2.7 Vidare forskning

Rapporten kan ge en indikation på hur JobTech bör agera. För att säkerställas om att resultatet är korrekt krävs däremot vidare och bredare undersökningar.

- De båda stickproverna måste utvidgas i storlek och görasom. Exempelvis genom DXA-system istället för JobTech atlas.
- Det vore intressant med entabell som redovisar körtiderna för de olika modellerna, inteminst eftersom uppdragsbeskrivningen nämner att praktisk användbarhet skall undersökas. Om den bästa modellen har en orimligt lång körtid så kanske den i praktiken inte kan användas till att processa den nattliga annonslasten.
- Att komplettera siffrorna från experiment 1 och två med konfidensintervall hade givit betydande insikter.
- Vidare forskning inom Mattias algoritm.
- Vidare undersöka varför taxanomin på arbetsförmedlingen skiljer sig från den på SCB avseende SSYK. Omkring 30 rader är inte samma. Påverkar detta resultatet på något sätt?
- Enanalytisk sektion i rapporten som går igenom proceduren för träning och utvärdering - var träningsdatan kommer ifrån, hur den partitionerats, hur stort utvärderingsdatasettet är etc, samt en deklaration att träningsdatan inte kontaminerats med data från utvärderingsdatasettet. Idet andra experimentet är det viktigt för läsaren att veta mer om experimentkörningen och att annonserna här inte kan ha ingått i Berts träningsdata.
- Det är intressant att Bert-Classifer enligt experimentet fick högre precision än vad människan fick enligt experiment 1. 86% är högre än 78%. Detta kan indikera på att människans pricksäkerhet i själva verket är närmare 86%. Alltså, om Berts reella precision är väldigt nära 100%, och den uppmätt har en precision på 86%, kan detta tolkas som att människans precision är 86%.
- Kan algoritmen appliceras på andra områden än just platsbanken? Exempelvis inom SCBs undersökningar, såsom Yrkesregistret, där företag ska skicka in alla sina anställdas befattningar, med hjälp av SSYK.

Vidare hade forskning om huruvida concept-id kan användas istället för SSYK varit intressant. SSYK är en jämförbart simpel taxanomi som ej återspeglar komplexiteten av marknaden. Ett argument till varför vi har SSYK, är att det är enkelt för människor att etiketera i jämförelse med exempelvis concept-id. Om etiketteringen ändå ska ske automatiskt, varför inte använda något bredare än SSYK? Detta borde ge beslutfattare betydligt bättre underlag för exempelvis nya YH-utbildningar.

## 2.8 Appendix

Ett exempel på SCB:s användning av SSYK:

Tabellen listar de trettio vanligaste yrkena i Sverige, utifrån SSYK.

SSYK	Yrke	Totalt	Kvinnor	Män	Kvinnor %	Män %
5321	Undersköterskor, hemtjänst, hemsjukvård och äldreboende	129 350	114 820	14 530	89	11
2341	Grundskollärare	112 500	83 160	29 330	74	26
5223	Butikssäljare, fackhandel	106 810	64 620	42 190	60	40
4322	Lager- och terminalpersonal	97 370	23 650	73 710	24	76
2512	Mjukvaru- och systemutvecklare m.fl.	90 830	17 320	73 510	19	81
5311	Barnskötare	89 720	81 140	8 580	90	10
5222	Butikssäljare, dagligvaror	88 950	57 850	31 100	65	35
3322	Företagssäljare	86 510	23 500	63 010	27	73
5330	Vårdbiträden	79 030	56 290	22 740	71	29
9111	Städare	75 600	54 680	20 930	72	28
5342	Vårdare, boendestödare	72 320	52 380	19 940	72	28
4119	Övriga kontorsassistenter och sekreterare	67 410	53 230	14 180	79	21
5343	Personliga assistenter	66 050	47 150	18 900	71	29
2343	Förskollärare	65 580	62 860	2 720	96	4
9412	Restaurang- och köksbiträden m.fl.	65 240	38 360	26 880	59	41
2422	Planerare och utredare m.fl.	63 660	41 930	21 720	66	34
8332	Lastbilsförare m.fl.	55 130	4 530	50 610	8	92
2221	Grundutbildade sjuksköterskor	51 370	45 450	5 920	88	12
7223	Maskinställare och maskinoperatörer, metallarbete	50 910	7 840	43 070	15	85
7111	Träarbetare, snickare m.fl.	49 030	670	48 360	1	99
5152	Fastighetsskötare	41 720	7 330	34 390	18	82
5312	Eleassistenter m.fl.	40 160	25 250	14 910	63	37
2421	Lednings- och organisationsutvecklare	39 830	19 510	20 310	49	51
5120	Kockar och källskänkor	39 690	19 950	19 740	50	50
5323	Undersköterskor, vård- och specialavdelning	38 810	34 850	3 960	90	10
4111	Ekonomiassistenter m.fl.	36 250	31 280	4 970	86	14
7411	Installations- och serviceelektriker	34 660	870	33 790	3	97
4222	Kundtjänstpersonal	34 600	20 320	14 270	59	41
7231	Motorfordonsmekaniker och fordonsreparatörer	32 350	920	31 430	3	97
3113	Ingenjörer och tekniker inom elektroteknik	32 070	4 710	27 360	15	85

Antalet klasser för varje kvalifikationsnivå visas i följande tabell:

Yrkesområde		Huvud-grupper	Yrkes-grupper	Under-grupper	Kvalifi-kations-nivåer
1	Chefsyrken	7	30	57	3:e 4:e
2	Yrken med krav på fördjupad högskolekompetens	6	29 <sup>1</sup>	113	4:e
3	Yrken med krav på högskolekompetens eller motsvarande	5	20	68	3:e
4	Yrken inom administration och kundtjänst	4	7	22	2:a
5	Service-, omsorgs- och försäljningsarbete	4	15	44	2:a
6	Yrken inom lantbruk, trädgård, skogsbruk och fiske	2	5	10	2:a
7	Yrken inom byggverksamhet och tillverkning	6	13	50	2:a
8	Yrken inom maskinell tillverkning och transport med mera	3	15	46	2:a
9	Yrken med krav på kortare utbildning eller introduktion	6	10	16	1:a
0	Militärt arbete	3	3	3	2:a 3:e 4:e
Totalt		46	147	429	--

Djupgående text om hur klassificering sker, skriven av Jonast Östlund:

#### “Binär klassificering: Två klasser

Jag ska också försöka beskriva mer i detalj hur jag har räknat ut *precision*, *recall* och *f1*. I ett binärt klassificeringsproblem finns det två klasser, t.ex. *Ja* och *Nej*. Vi antar att vi för varje exempel har *den sanna klassen* (Ja eller Nej) och den av algoritmen förutsagda klassen, *prediktionen* som också är Ja eller Nej. Det betyder att det finns fyra olika kategorier för varje exempel: True Positive, True Negative, False Positive och False Negative.

- True Positive: Sann klass=Ja, Predicerad klass=Ja
- True Negative: Sann klass=Nej, Predicerad klass=Nej
- False Positive: Sann klass=Nej, Predicerad klass=Ja

- False Negative: Sann klass=Ja, Predicerad klass=Nej

Namngivningen för dessa kategorier följer alltså principen att False betyder att prediktionen var fel och True att prediktionen var rätt. Positive och Negative syftar på vad prediktionen var.

När man evaluerar en binär klassificeringsalgoritm (d.v.s. en algoritm med två olika klasser) på flera exempel räknar man antalet exempel som faller under var och en av dessa fyra grupper och refererar till dessa antal som TP (True Positive), TN (True Negative), FP (False Positive) och FN (False Negative). Utifrån dessa värden kan man sedan härleda måtten *precision* och *recall*, här är wikipedia-artikeln: [https://en.wikipedia.org/wiki/Precision\\_and\\_recall](https://en.wikipedia.org/wiki/Precision_and_recall)

### 2.8.1 Fler än två klasser

I vårt fall har vi emellertid fler än två olika klasser. I själva verket har vi 400 olika klasser som motsvaras av de olika SSK4-koderna i våra data. Dessa klasser listas i det kalkylark som jag skickade med resultat. Där framgår till exempel att vi har 338 exempel där den *sanna etiketten* är "Grundutbildade sjuksköterskor (2221)" och det totala antalet exempel är 6779. För varje klass redovisas värdena *precision*, *recall* och *f1* i raden för varje klass och för varje algoritm. Dessa är ju binära klassificeringsmått. Det vi gör för att räkna ut dessa värden är att vi för var och en av de 400 klasserna betraktar ett binärt klassificeringsproblem för just den klassen, där vi ställer den binära frågan: Tillhör exemplet den här klassen? Så om vi till exempel bara betraktar klassen sjuksköterskor, då frågar vi för vart och ett av de 6779 exemplen: "Ska den här platsannonsern klassificeras som Grundutbildade sjuksköterskor (2221)?" och tittar på om den *sanna* klassen är svarade Ja eller Nej, och huruvida den *predicerade* klassen svarade Ja eller Nej. Sedan kategoriserar vi varje exempel som True Positive, True Negative, False Positive och False Negative.

För att göra det mer konkret, I tabellen som jag skickade har vi 338 exempel som är märkta som klass 2221 (Grundutbildade sjuksköterskor). Antag att av dessa 338 exempel har algoritmen som vi provkör också korrekt svarat att 310 exempel är klass 2221. Då är TP=310. Då återstår 338-310 = 28 exempel där rätt svar är klass 2221 men där algoritmen har svarat någon av de andra 399 klasserna (t.e.x mjukvaruutvecklare). Då blir värdet FN=28. Nu betraktar vi alla återstående 6779-338=6441 exempel där den sanna etiketten är vad som helst utom just klass 2221. Antag att av dessa svarade algoritmen med någonting annat än just klass 2221 för 6300 exempel. Då blir TN=6300. Nu återstår 6441 - 6300 = 141 exempel där algoritmen felaktigt svarade att exemplen tillhör klass 2221. Då blir värdet FP=141. Sedan kan vi med hjälp av TP, TN, FP och FN beräkna precision och recall för sjuksköterskor. Och vi gör på samma sätt för de andra 399 klasserna.

När vi har beräknat dessa mått kan vi beräkna *macro precision* och *macro recall* som i ekvationerna (14) och (15) i <https://arxiv.org/pdf/2008.05756.pdf>.

Samtliga versioner av text2ssyk och deras respektive precision:

59%	66%	73%	74%	78%	77%	49%
31%	38%	50%	54%	62%	58%	29%
67%	70%	75%	76%	80%	80%	56%
35%	42%	55%	57%	66%	62%	32%
text2ssyk	text2ssyk	text2ssyk	text2ssyk	text2ssyk	text2ssyk	text2ssyk-sbert-cov
0.1.0	tse-config0	tse-config1	tse-config2	tse-config6	tse-config7	0.1.0

SSYKGPT:

<https://gitlab.com/arbetsformedlingen/devops/documentation/report-ssykgpt/-/tags/v0.1-DRAFT>



58	Systemtestare och testledare (2514)	33.85%	82%	849.46%	80%	599.30%	100%	479.12%	80%	219.36%	92%	529.45%	79%	589.64%	85%	709.52%	85%	649.30%	57%	449	
59	Övriga administrations- och serviceschefer (1290)	32.91%	97%	849.79%	58%	679.53%	74%	629.60%	66%	669.75%	86%	809.66%	88%	759.78%	83%	819.72%	92%	819.72%	50%	599	
60	Förvaltnings- och planeringschefer (1236)	31.71%	54%	619.35%	14%	209.19%	43%	279.23%	33%	279.52%	44%	489.32%	53%	529.58%	51%	509.48%	50%	499.23%	23%	239	
61	Bemanningscentraler (222)	30.97%	100%	899.81%	81%	889.87%	84%	899.87%	90%	899.82%	100%	919.82%	90%	909.80%	100%	909.80%	90%	939.82%	81%	829	
62	Backofficepersonal m. fl. (4113)	29.66%	66%	669.26%	7%	119.21%	55%	309.34%	100%	519.52%	100%	689.55%	84%	679.59%	74%	699.55%	84%	679.24%	50%	339	
63	Civiltjänstgöringen inom bygg och anläggning (2142)	28.64%	86%	739.33%	86%	489.46%	25%	6%	7%	50%	129.30%	71%	489.21%	86%	349.30%	77%	489.32%	82%	489.14%	50%	229
64	Tandläkare (555)	28.96%	100%	899.25%	100%	409.22%	83%	879.89%	89%	879.89%	79%	849.89%	85%	879.86%	86%	819.100%	87%	929.58%	56%	579	
65	Intensivvårdsassistent (228)	25.88%	96%	929.65%	76%	709.12%	27%	179.16%	67%	209.64%	94%	769.60%	88%	719.68%	100%	729.12%	30%	179			
66	Pizzabagare m. fl. (8411)	24.100%	89%	849.75%	100%	869.50%	75%	609.39%	75%	509.50%	67%	579.54%	70%	639.58%	61%	609.58%	89%	709.54%	50%	529	
67	Inköps-, logistik- och transportchefer (1320)	23.91%	100%	959.87%	100%	939.74%	100%	899.87%	91%	899.91%	95%	939.87%	91%	899.87%	95%	919.87%	95%	919.79%	86%	829	
68	Taxibuss m. fl. (8321)	22.82%	86%	849.88%	41%	569.59%	81%	689.77%	89%	839.82%	75%	789.77%	81%	789.77%	81%	799.81%	74%	829.27%	40%	329	
69	Grovarbete inom bygg och anläggning (8310)	21.100%	100%	100.31%	83%	459.81%	100%	899.90%	100%	959.100%	95%	889.100%	100%	100.100%	95%	989.90%	95%	989.90%	95%	939	
70	Ekonomi- och finanschefer (1210)	20.90%	90%	809.100%	55%	719.40%	89%	569.45%	90%	609.55%	73%	639.55%	81%	729.70%	82%	709.65%	100%	799.15%	75%	259	
71	Ordinariearbete m. fl. (3324)	19.84%	75%	809.83%	65%	739.68%	62%	659.44%	50%	609.74%	47%	579.84%	67%	749.68%	62%	609.79%	71%	759.68%	27%	399	
72	Barnsköterskor (222)	18.94%	85%	899.67%	40%	509.11%	67%	199.17%	50%	259.44%	80%	579.61%	79%	699.50%	82%	629.56%	71%	629.11%	33%	179	
73	Säljande butikschef och avdelningschef i butik (5221)	17.59%	91%	719.0%	0%	0%	0%	100%	119.41%	88%	569.53%	82%	649.59%	91%	719.53%	90%	679.53%	79%	629.24%	44%	319
74	Övriga civiltjänstgöringen (2149)	16.15%	30%	239.0%	0%	0%	0%	2%	109.12%	40%	199.15%	42%	269.12%	50%	209.38%	76%	509.19%	60%	239.0%	0%	7%
75	Banströmar (5132)	15.73%	92%	819.17%	40%	349.20%	50%	299.13%	33%	199.27%	57%	369.27%	80%	409.40%	60%	489.27%	67%	389.20%	27%	239	
76	Bemanningspersonal (5225)	14.86%	100%	929.0%	0%	0%	36%	100%	539.21%	100%	399.64%	100%	789.50%	88%	649.57%	89%	709.89%	92%	899.7%	50%	129
77	Fisker (5141)	13.54%	88%	679.0%	0%	0%	31%	57%	439.49%	100%	639.54%	78%	649.54%	70%	619.52%	80%	704.54%	79%	619.15%	33%	219
78	Förskollära och förskollärautbildare (3321)	12.75%	64%	699.43%	75%	559.8%	100%	159.39%	57%	429.42%	62%	509.33%	57%	429.75%	66%	709.50%	67%	579.0%	0%	0%	
79	Laboratorieingenjörer (3215)	11.55%	67%	609.50%	60%	559.9%	50%	159.39%	80%	509.45%	83%	599.45%	56%	509.55%	67%	609.64%	88%	749.64%	28%	399	
80	Fästighetsförvaltare (3335)	10.100%	100%	100.100%	100%	100.70%	100%	829.30%	100%	489.60%	100%	759.90%	100%	569.80%	90%	909.70%	100%	829.50%	83%	629	
81	Tjänstemän (1121)	9.100%	100%	100.2%	12%	179.0%	0%	5%	100%	719.3%	75%	489.7%	100%	889.7%	100%	889.17%	70%	719.2%	50%	31%	
82	Maskinoperatörer, plastindustri (8142)	8.100%	100%	100.50%	50%	509.12%	100%	229.38%	79%	509.75%	100%	889.75%	100%	889.88%	100%	939.88%	100%	939.25%	67%	369	
83	Glastekniker (7124)	7.100%	88%	839.100%	100%	100.0%	0%	4%	100%	609.86%	100%	929.100%	100%	100.71%	100%	839.86%	100%	929.43%	100%	609	
84	Skadereglerare och värdare (3314)	6.63%	45%	599.0%	0%	0%	0%	0%	50%	75%	609.33%	100%	509.33%	67%	449.17%	33%	229.50%	75%	609.17%	50%	259
85	Utläskare inom spel och digitala media (2513)	5.80%	100%	899.75%	33%	469.80%	100%	899.80%	80%	809.100%	100%	100.80%	100%	899.80%	80%	809.60%	80%	809.60%	100%	759	
86	Optikerassistenter (524)	4.75%	60%	679.0%	0%	0%	0%	0%	50%	100%	679.25%	100%	409.25%	100%	409.25%	50%	339.25%	100%	409.0%	0%	0%
87	Arbetslösa och specialister inom hantverks m. m. (832)	3.32%	100%	509.0%	0%	0%	33%	50%	409.0%	0%	33%	100%	509.33%	100%	509.100%	100%	100.33%	100%	509.0%	0%	0%
88	Maskinoperatörer, påfyllning, packning och märkning (8181)	2.100%	67%	809.100%	20%	339.100%	87%	809.100%	67%	809.100%	100%	100.100%	87%	809.100%	67%	809.100%	67%	809.100%	67%	809.0%	0%
89	Prepress tekniker (7321)	1.100%	100%	100.0%	0%	0%	0%	0%	100%	100%	100.100%	100%	100.100%	100%	100.100%	100%	679.100%	100%	100.0%	0%	0%
90	@trained	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
91	Count	8883	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
92	Accuracy	86%		53%		59%		66%		73%		74%		78%		77%		63%		61%	
93	Recall	76%		39%		31%		39%		50%		54%		62%		58%		29%		29%	
94	Precision	81%		42%		47%		57%		75%		79%		80%		80%		56%		56%	
95	F1	75%		32%		35%		42%		55%		57%		66%		62%		32%		32%	
96	Algoritm	BertClassifier	SyzyGut		text2syzy	text2syzy		text2syzy		text2syzy		text2syzy		text2syzy		text2syzy		text2syzy		text2syzy	bert-cov
97	Version	1-cdms-sep	0		0.1.0			tse-conf0		tse-conf1		tse-conf2		tse-conf3		tse-conf7					0.1.0
98	@trained																				
99	@trained-sbert-name																				

Råa datan till experiment 2

Håller med:	Bert	Människan	Lika bra: 4	Irrelevant: 2	Kan ej bedöma: 1
ChatGPT		12	4		
Jacob		13	3		
Otydliga:		4			