



LUNDS
UNIVERSITET

Institutionen för psykologi

**Improving Intelligent survey: Using Brute-Force method
to improve AI for diagnosing depression and anxiety**

**Förbättra Intelligent Survey: använda brute-force metod för
att förbättra diagnostisering av depression och ångest med
AI**

Sofia Jonasson & Jacob Lind & Joel Vikner

Kandidatuppsats VT22

Handledare: Sverker Sikström

Examinator: Joana Duarte

Abstract

The project Intelligent Survey (IS) was developed to predict depression and anxiety from natural language by translating individuals' qualitative answers from eleven open-ended questions to quantitative values, which are compared and predicted against the scales; Patient Health Questionnaire (PHQ-9), Generalized Anxiety Disorder scale (GAD -7), Harmony in Life scale (HILS) and Satisfaction With Life Scale (SWLS). In collaboration with the ongoing project, the aim of this study was to minimize the mean square error (MSE) in IS by modifying the question order, in order to obtain as much information as possible, with as few questions as possible. The method chosen was a brute-force method that tested all possible starting points to find the optimal first question. IS was trained with responses from 883 participants. The brute-force code resulted in 11x4 models with different starting points for IS. Differences were calculated to identify trends of change in MSE and to identify and implement the best starting point for IS. Two dependent t-tests were performed to be able to investigate with greater certainty whether, and to what extent, the new code reduced MSE in IS. The first t-test measured average MSE over the order of the questions and showed statistical significance with relatively high effect size for all scales except SWLS. However, a skepticism of the results was assumed as only PHQ-9 showed statistical significance in the second t-test that measured average MSE over participants. The brute-force method entailed a major reduction in MSE and proved suitable for examining different starting points for IS. Potentially, the brute-force method can be used to improve other similar AI systems in diagnostics.

Keywords: *depression, anxiety, artificial intelligence, mental health, brute-force method*

Sammanfattning

Projektet *Intelligent Survey* (IS) utvecklades för att predicera depression och ångest från naturligt språk genom att översätta individers kvalitativa svar från elva öppna frågor till kvantitativa värden som jämförs och prediceras mot skalorna; *Patient Health Questionnaire* (PHQ-9), *Generalized Anxiety Disorder scale* (GAD-7), *Harmony in Life scale* (HILS) och *Satisfaction With Life Scale* (SWLS). I samarbete med det pågående projektet var denna studies syfte att minimera *mean square error* (MSE) i IS genom att modifiera frågeordningen för att få så mycket information som möjligt med så få frågor som möjligt. Metoden som valdes var en brute-force metod som testade alla möjliga utgångspunkter för att hitta den optimala första frågan. IS tränades med svar från 883 deltagare. Brute-force koden resulterade i 11x4 modeller med olika utgångspunkter för IS. Mellanskillnader beräknades för att identifiera trender av förändring i MSE och för att urskilja och implementera den bästa utgångspunkten för IS. Två beroende t-test utfördes för att med större säkerhet kunna undersöka om, och med vilken effektstyrka den nya koden minskat MSE i IS. Första t-testet mätte genomsnittlig MSE över ordningen på frågorna och visade statistiskt signifikans med relativt höga effektstyrkor för alla skalor utom SWLS. Däremot antogs en skepticism av resultaten då enbart PHQ-9 visade statistisk signifikans i det andra t-testet som mätte genomsnittlig MSE över deltagare. Brute-force metoden innebar en huvudsaklig minskning av MSE och visade sig lämplig för att undersöka olika utgångspunkter för IS. Potentiellt kan brute-force metoden utnyttjas för att förbättra andra liknande AI-system inom diagnostisering.

Nyckelord: *depression, ångest, artificiell intelligens, mental hälsa, brute-Force metod*

Tack!

Vi skulle vilja tacka vår handledare Sverker Sikström för möjligheten att få vara en del av forskningsprojektet och ta del av den forskning han och hans team har gjort på ämnet. Tack för möjligheten att vi fick göra de framsteg i projektet som vi klarade av och hade förmågan till. Vi vill även tacka våra nära och kära för den hjälp vi har fått under arbetets gång och deras tålamod de haft med oss samt den hjälp vi fått med att förstå de program vi var tvungna att lära oss.

Förbättra Intelligent Survey: använda brute-force metod för att förbättra diagnostisering av depression och ångest med AI

Psykisk ohälsa är ett högaktuellt ämne idag och enligt World Health Organisation (u.å.) har det ökat med 13% på tio år när det mättes vid 2017. De klassar också depression som en vanligt förekommande sjukdom, estimerat hos 3,8% individer världen över (World Health Organisation, 2021). Datoriseringen av psykiatrin och bekämpningen av psykisk ohälsa har fått ett nytt kraftfullt verktyg, *Artificiell intelligens (AI)*, som kan behandla stora mängder data och bidra till en djupare förståelse och bättre beslut inom diagnos, prognos och behandling av människor som lider av psykisk ohälsa (Dwyer et al., 2018). Förmågan hos datorer att ha börjat förstå meningen av ord och naturligt språk är ett stort framsteg inom flera områden och speciellt viktig för tillämpningar inom mental hälsovård (Cambria, E. & White, B., 2014).

AI används alltmer i samhället idag och i takt med teknikens framfart breddas även dess användningsområden. Att använda AI inom psykisk hälsovård är enligt Graham et al. (2019) lovande, men kräver att expertis från olika discipliner samarbetar för att optimera praktiken och hälsan hos patienter, vilket inkluderar att de är aktiva i introduktionen av AI till kliniker. Utvecklingen av AI inom psykiatrin kan erbjuda bättre och mer objektiva definitioner av mental ohälsa för kliniker jämfört med DSM-5 och kan vara bidragande för att upptäcka tecken på psykopatologi i ett tidigare stadium, som kan göra interventioner mer effektiva (Graham et al., 2019; Bzdok & Meyer-Lindenberg, 2018; Dwyer et al., 2018).

Exempel på psykopatologier som redan idag undersöks ur ett maskininlärningsperspektiv är depression, schizofreni och Alzheimers (Shatte et al., 2019). Användandet av AI inom psykiatrin skapar möjligheter att utveckla undersökningsverktyg som tidigare och mer effektivt kan förutspå och identifiera diagnoser samt riskanalytiska modeller av individens anlag för att utveckla psykopatologi (Shatte et al., 2019). För att kunna implementera AI i psykisk hälsovård som ett långsiktigt mål, krävs bemästring av en metod som kan hantera stora mängder data, vilket går att uppnå med datorisering (Graham et al., 2019).

Det är också värt att notera hur användandet av AI-teknologin kan frigöra tid och utrymme för de arbetande i områden kring psykisk ohälsa och psykiatri. Implementeringen av AI algoritmer inom psykisk hälsovård och psykiatri kan bygga ut förståelsen av förekomsten

av psykisk ohälsa på populationsnivå, dra nya meningsfulla slutsatser utifrån stora dataset, göra nya fynd för biologiska mekanismer, risk- och förebyggande faktorer, erbjuda nya metoder för uppföljning av effektivitet i medicinering, erbjuda intelligenta skattningsmetoder för att urskilja allvarlighetsgrad av psykisk ohälsa och mycket mer. En viktig aspekt av dessa förbättringsområden är att appliceringen av AI-algoritmer för databehandling frigör tid och uppmärksamhet för de praktiserande att fokusera på de mänskliga aspekter som följer med behandling av psykisk ohälsa (Kim, J.W. et al., 2019; Topol, E., 2019).

Natural language processing (NLP) är en del av AI vars syfte är att förstå naturligt språk i text och tal samt sedan kunna operera på det genom algoritmer. Naturligt språk är inget som datorer kan förstå av sig själva utan det behöver översättas till programspråk för att datorer ska kunna tolka och förstå. NLP kan ses som länken mellan naturligt språk och programspråk. NLP kan användas inom många olika områden för att kommunicera med datorer och få datorer att förstå naturligt språk (IBM Cloud Education, 2020; Casey, K. 2019; Cambria, E & White, B. 2014).

Semantic excel är ett program utformat för att skapa möjligheten att utföra ett flertal olika typer av statistiska analyser på text av naturligt språk. Medföljande programmet är *semantiska representationer*, i form av nummer som beskriver semantiska likheter mellan texter och ord (Sikström et al., 2020). *Semantic excel* utför olika NLP analyser genom semantiska representationer som fungerar likt koordinater där varje ord får en position i ett *semantic space*, eller, semantiskt rum. Hur nära orden ligger varandra representerar hur lika varandra de är i innebörd (Kjell et al., 2019). Ett väldigt stort dataset krävs för att skapa bra semantiska representationer. Alternativt går det att undgå det kravet om ett semantiskt rum skapas, varav de semantiska representationerna sedan kan skapas från. *The Semantic Space* eller det semantiska rummet är en approach, likt Dumais et al., (1988) automatiska semantiska struktureringar kallat *Latent Semantic Analysis* (LSA), som skapar semantiska representationer från en stor mängd kvalitativ data och är en form av natural language processing (Kjell et al., 2019).

Semantic measures är ett samlingsbegrepp för ett flertal statistiska verktyg som skapats i svar på det komplicerade problemet att objektiva kvantifiera svar från öppna frågor. Metoden gör detta genom analys av NLP. *Semantic measures* skapades genom att semantiken i svar från öppna frågor i nio studier jämförts mot traditionella skattningsskalor (Kjell et al., 2019). De argumenterar att maskininlärning samt NLP kan lösa svårigheter med omvandlingen av ord till numerisk skala, något de anser bör ligga på den vetenskapliga metoden och inte på respondenten (Kjell et al., 2019). Traditionella skalor har visat mycket framgång, men Kjell

et al. (2019) argumenterar för att verbala svar från öppna frågor kan erbjuda en högre ekologisk och face validity vid mätning av psykologiska konstrukt. Således då individer inte behöver kognitivt omvandla semantik eller sitt naturliga språk som representerar deras mentala tillstånd till en endimensionell numerisk skala. Med detta anpassas svarsformatet utefter folket och det naturliga språket, något de anser är optimalt för framtida metodologi (Kjell et al., 2019). Följaktligen kan metoden generera stora framsteg i förståelsen av individens inre mentala stadier.

Den kan också innebära många praktiska användningsområden i både medicinsk och psykologisk forskning samt andra professionella sammanhang (Kjell et al., 2019). Det öppnar upp för möjligheten att individer med öppna ord kan berätta om deras känslor och mentala tillstånd och utefter deras svar potentiellt skatta och diagnostisera depression och ångest på effektivt sätt med hjälp av AI. Graham et al. (2019) förespråkar för att AI inom mental hälsovård verkar lovande men att det fortfarande är i ett tidigt skede, därmed att det krävs mer arbete för att det ska bli applicerbart i praktiken, varav en försiktighet bör antas så att tidiga resultat inte övertolkas.

Syfte & hypotes

Intelligent Survey (IS) är en AI i ett forskningsprojekt under utveckling som diagnostiserar mental hälsa genom att analysera deltagarnas deskriptiva ord från svar till öppna frågor. Orden analyseras av AI och NLP och får sedan värden baserat på semantiska representationer. Värdena används sedan för att predicera ångest och depression. Det semantiska rummet i studien skapades från en sammanslagning av en summering av engelska Google N-grams databas och det semantiska rummet från studien av Kjell et al. (2019). Likt studien användes samma metod, genom semantic excel, för att agera som en modell för de mindre dataset genererade i studien. Intelligent Survey innefattar en felvarians, *mean square error* (MSE), som kan effektivisera prediktionsmodellen om den minimeras. Vårt arbete pågick i samband med forskningsprojektet Intelligent Survey. Målet var att minimera MSE vid diagnostisering, genom att ändra ordningen på frågorna i regressionsmodellen för att så mycket information som möjligt genereras med så få frågor som möjligt. Således syftar vi till att finna en metod som kan implementeras i koden och därmed minimera MSE i prediktionen som resulterar i ett förbättrat mätredskap. Vår hypotes var att vi med hjälp av att undersöka alla möjliga frågor som utgångspunkt för IS, en så kallad brute-force metod, kunde minska MSE genom att ändra ordningen på första frågan.

Metod

Inledning

För att förstå arbetets metodologiska ansats är det viktigt att ha två saker i åtanke. Det första är att arbetet skett i ett större forskningsprojekt där Intelligent Survey redan innefattar en baskod. En kod som byggts upp till att kunna diagnostisera depression och ångest med hjälp av öppna frågor för att göra prediktioner på skalor för depression, ångest, harmoni i livet och tillfredsställelse i livet, samt en prediktion på MSE för varje deltagares svar och skala (Kroenke et al., 2001; Spitzer, et al., 2006; Diener et al., 1985; Kjell et al., 2016). Det andra är att arbetet var öppen för olika framsteg beroende på vilka försök som lyckades under metodens gång. Således innebar det att vår metod innefattade icke förutbestämda försök till att skapa förändringar i koden för att kunna minimera MSE genom ändringar i frågornas ordning. Utifrån denna metodologiska ansats har implementeringar lett till att en ny kod utvecklats, utifrån en hypotes om att en brute-force metod över ordningen för första frågan kan upptäcka modeller med lägre MSE och därmed högre effektivitet än baskoden.

Urval rekrytering och deltagare

För att träna AI:n används 883 deltagare vars data tidigare samlats in under det pågående projektet. Alla deltagare talade engelska som modersmål. Majoriteten av deltagarna var från USA (83%) medans resterande var från Indien (10%), Storbritannien (3%) och andra länder (4%). Medelåldern på deltagarna var 35,5, där deras ålder varierade mellan 18 till 70 år (standardavvikelse 11,8). 539 av deltagarna var kvinnor och 334 var män, resterande 10 identifierade sig som något annat. För att kategorisera variationen av de diagnoser som ämnades att mätas blev deltagarna undersökta med SDAS, ett mätinstrument som används vid diagnostisering av major depression disorder (MDD) och generaliserat ångestsyndrom (GAD). Antalet deltagare med MDD var 61, 138 deltagare hade GAD och 259 deltagare hade båda diagnoserna. Resterande 425 deltagare led varken av MDD eller GAD vilket utökade variationen i urvalet från vilket AI:n får information att träna på. Deltagarna gavs fyra dollar för sitt deltagande och studien tog 20 minuter att genomföra.

Design

Syftet med IS är att den ska kunna på ett precist och effektivt sätt skatta och diagnostisera depression och ångest med hjälp av svar med 2 -5 deskriptiva ord från öppna frågor (se appendix A). Ju färre frågor som krävs för en bra prediktion med låg MSE, ju

bättre. Målet med IS är därför att få ut så mycket information som möjligt genom att använda så få frågor som möjligt.

Genom att fråga efter öppna svar som “*Describe your mental health with five descriptive words*” istället för att ställa stängda frågor som “*On a scale from 1-7, how are you?*”, genereras svar som är mer i linje med hur den subjektiva upplevelsen uttrycks i naturligt språk. Exempelvis svar som “*Fine and Happy*”, istället för svar som “7”. Svaren omvandlas till kvantitativa värden beroende på likheten till semantiska representationer som bygger på ett semantic space (Kjell et al., 2019). Svaren från frågorna jämförs sedan med deltagares självskattningar på skalorna *Patient Health Questionnaire* (PHQ-9), *Generalized Anxiety Disorder scale* (GAD-7), *Harmony in Life scale* (HILS) och *Satisfaction With Life Scale* (SWLS).

PHQ-9 är ett verktyg som utvecklats för att diagnostisera depression baserat på nio kriterier från DSM-IV. Individer under självskattning får besvara på ett antal frågor, varav en lyder huruvida de upplevt sig nedstämda, deprimerade eller en hopplöshet under de senaste två veckorna. Frågorna graderas från 0 (inte alls) till 3 (nästan varje dag). PHQ-9 kan även med hög reliabilitet och validitet mäta allvarlighetsgraden av depressionen, från 5 (måttlig) till 20 (allvarlig depression) (Kroenke, et al., 2001).

GAD-7 är ett validerat och effektivt mätinstrument, med 7 items, som diagnostiserar generalized anxiety disorder (GAD) och allvarlighetsgraden. Dessa skattas sedan från en 0a (inte alls) till en 3a (nästan dagligen) och innehåller uttryck som “*feeling nervous, anxious or on edge*”. Reliabiliteten av GAD-7 i återtester har visat goda resultat, med intraklasskorrelation på 0,83. Den interna konsistensen har en cronbach alpha på 0,92. Således är GAD-7 ett bra mätinstrument för både praktiken och i research för diagnostisering av GAD och dess allvarlighetsnivå (Spitzer, et al., 2006).

SWLS mäter global livstillfredsställelse. SWLS har hög intern konsistensitet och hög temporal reliabilitet. Korrelationen mellan SWLS och andra mätinstrument för objektiva välmående är måttlig till hög. SWLS är lämplig för att användas till olika åldersgrupper. SWLS har designats för att övergripande mäta någons bedömning av dennes liv. Detta möjliggör till mätningen av konceptet livstillfredsställelse (Diener, et. al., 1985).

HILS är ett mätinstrument som tar fram psykologisk balans och flexibilitet i livet. HILS korrelerar med välmående-relaterade skalor. HILS kompletterar SWLS genom att det formar ett mer holistiskt förstående av subjektivt välbefinnande. HILS visar på validitet, reliabilitet och unikheter (Kjell, et. al., 2016).

I baskoden benämns de fyra ovannämnda skalorna som “trainVariables” och de deskriptiva ord som deltagarna besvarade med i de öppna frågorna som skalorna jämförs med benämns som “variableToCreateSemanticRepresentationFrom”. Svaren som deltagarna gav till de öppna frågorna gjordes vid samma tillfälle som självskattningarna på skalorna. När frågorna ställdes till deltagarna var det i slumpmässig ordning för samtliga deltagare. Därför går det att anta att den ursprungliga slumpmässiga ordningen som deltagarna besvarade frågorna på inte påverkat vilka deskriptiva ord de ger som svar på varje enskild fråga, randomiseringen motverkar confounders. Analyserna och beräkningarna av IS görs alltså på redan befintliga svar från de öppna frågorna. Ändringar av frågeföljden innebär i praktiken ändringar på ordningen av svaren och deras semantiska värden.

Baskoden

Inledningsvis räknar AI:n ut värden av semantiska representationer från svaren i frågorna “variableToCreateSemanticRepresentationFrom” genom att jämföra dessa med de semantiska representationerna som finns i det semantiska rummet. Sedan med “pred{Y, i}.d=getX(s, pred{Y, i}.index);” tränas koden att utföra prediktioner mellan frågorna, “variableToCreateSemanticRepresentationFrom”, och de fyra skalorna, “trainVariables”. Därefter med “MSE{Y}(i, :)=(pred{Y, i}.pred-pred{Y, i}.data).^2;” tränas AI:n predicera MSE för frågorna och “trainVariables”.

Efter att AI:n tränats för att predicera skalorna och MSE för prediktionen över skalorna kör koden igenom alla deltagares svar med tre olika modeller, som på olika sätt avgör i vilken ordning frågorna placeras i följd. De tre modellerna är Random, Fixed och Intelligent Survey. Frågornas predicerade MSE plottas över respektive modells frågeordning fram till 10e frågan, med varsin graf för varje skala av “trainVariables”. Plottningen gör det lättöverskådligt för jämförelse av de olika modellernas påverkan på MSE.

I modellen Random placeras frågorna i slumpmässig följd för varje deltagare. I modellen Fixed har samtliga deltagare samma ordning på frågorna. Ordningen för frågeföljden i Fixed utgår från vilka frågor som prediceras minska feltermen mest utifrån en generalisering av värdena som svaren från deltagarna gett på frågorna och är olika för de olika skalorna.

Modellen Intelligent survey (IS) genererar en ordning på frågorna som utgår från vad varje person har svarat på föregående fråga. IS använder en multipel linjär regression genom en prediktionsmodell som tar de genererade värdena från varje fråga och sammanställer dem i

en vektor där de aggregeras och ger ett nytt värde för varje frågas predicerade prediktionsvärde, genom koden

“`[s, ~, pred{Y,i}]=scriptTrain(s,par,func)`”. IS tolkar med vilken styrka en fråga kan bidra till minskandet av MSE genom att analysera värdena som produceras av varje fråga individuellt. IS lägger sen till den fråga som predicerar störst minskning i MSE utifrån denna prediktionsmodell efter varje fråga på varje deltagare. Eftersom ordningen på frågorna i IS utgår ifrån svaren på den föregående frågan är den första frågan samma för alla deltagare, för att agera som utgångspunkt. För att bestämma första frågan åt IS över alla skalorna användes första frågan från mallen för modellen fixed, dvs. nummer 1 på X-axeln i “`index1`” vilket var fråga 3; frågan om ens generella mentala hälsa (se appendix A). Anledningen för att utgå från “`index1`” är att den redan utgick från den predicerat bästa kollektiva ordningen, specificerat för de olika skalorna. I koden är det här en 4x11 matris, kallad “`index1`”, där de fyra raderna på Y-axeln är skalorna i “`trainVariables`”. De 11 kolumnerna på X-axeln är de olika frågorna av “`variableToCreateSemanticRepresentationFrom`” i den bästa ordningen för korresponderande skala, från 1-11.

När koden har loopat alla deltagare över alla modeller plottas en figur med linjära regressionsmodeller som visar hur de olika modellernas MSE varierar över de tio frågorna i test order i den skala de loopats över (se appendix B, figur B1). De olika modellerna varierar eftersom de använder olika metoder för att bestämma ordningen på frågorna och således sjunker MSE olika snabbt över de elva frågorna. Därefter börjar koden på nästa skala och upprepar proceduren för att plotta regressions-figurer med samtliga modeller för varje skala. Slutligen, efter att alla skalor har gått igenom, plottas även histogram-figurer som visar hur variablerna (frågorna) har fördelats i ordning efter antal deltagare (se appendix B, figur B2).

Procedur

Genom ett antal möjliga metoder kan regressionsmodeller förbättras, därmed kan modifikationer av Intelligent Survey innebära att dess *Mean Square Error* (MSE) blir lägre. I projektet har en metod valts ut för att begränsa arbetet och hålla det inom den tidsram som finns tillgänglig. Den valda metoden syftar till att se hur ordningen som frågorna läggs in i prediktionen i IS kan öka eller minska MSE, med målet att få en så låg MSE som möjligt så tidigt som möjligt. En så kallad *brute-force* metod tillämpades för att systematiskt testa alla möjliga frågor som första fråga, därmed testa alla möjliga utgångspunkter för AI:n och identifiera de som bäst predicerar utkomsten och därmed hitta möjligheter till minskad MSE.

Genom att skapa en ny for-loop mellan loopen över skalorna och loopen över modellerna, med en ny variabel, “a”, som är lika med “1 till längden av frågor”, och sedan välja variabeln “a” som första fråga över alla skalorna i stället för nummer 1 från `index1`, testar AI:n sin prediktionsmodell med alla möjliga olika frågor som utgångspunkt innan den går vidare till nästa skala. I originalkoden utgår IS från första frågan i `index1`, och implementerar sin prediktionsmodell utifrån den utgångspunkten. I brute-force koden gör IS samma sak, fast för alla möjliga frågor i `index1`. Förändringen innebär i praktiken att alla möjliga utgångspunkter för prediktionsmodellen att utgå ifrån körs för att kunna analysera hur IS prediktionsmodell reagerar på dem och för att se om någon av dem innebär lägre MSE än originalkodens utgångspunkt. Den här förändringen refereras hädanefter i arbetet som brute-force koden. Som nämnt ovan anpassas frågeföljden i IS utefter deltagarnas svar på frågorna. För att bibehålla AI:ns funktion i en brute-force applicering implementerades förändringen i koden vid initieringsmomentet för AI:ns utgångspunkt för varje deltagare.

I Baskoden innan förändringen tog initieringsmomentet sig uttryck genom:

```
Y=1:length(trainVariables)
indexStart=indexText
indexText=index1(Y,1)
```

I Brute-force koden efter förändringen tog initieringsmomentet sig uttryck genom:

```
Y=1:length(trainVariables)
a=1:length(variablesToCreateSemanticRepresentationsFrom)
indexStart=indexText
indexText=index1(Y,a)
```

“`indexStart`” initierar IS prediktionsmodell och behöver ett värde att utgå ifrån. “`indexText`” förklarar vad prediktionsmodellen ska utgå ifrån och fungerar som vägen mellan initieringen och värdet som den utgår ifrån. “`index1`” är en matris med 4x11 celler som fungerar som mall från vilken värdet för utgångspunkten i IS hämtas från. De fyra raderna på Y-axeln i matrisen är de fyra skalorna, och är samma som “`trainVariables`”; PHQ-9, GAD-7, HILS och SWLS. De 11 kolumnerna på x-axeln är samma som “`variablesToCreateSemanticRepresentationsFrom`”, och är de 11 frågor som ligger till grund för deltagarnas svar och de semantiska representationerna.

Den här ändringen i koden gjorde så att IS inte längre utgick ifrån den fråga som predicerade lägst MSE enligt de generaliserade värden som första fråga, nummer 1 i `index1`. Istället testade den alla möjliga frågor som första fråga och bildade en fullständig

regressionsmodell med varje. Det innebar att modellen med frågan om ens generella mentala hälsa, “*describe your mental health with five descriptive words*” som första fråga, som predicerats bidra till lägst MSE, kunde jämföras med hur modellen hade sett ut med alla andra möjliga frågor i `index1` som utgångspunkt för AI:n.

Efter att alla möjliga alternativ till första frågan testats var det tydligt genom figurerna som plottats efter varje loop över samtliga modeller vilken fråga som var den optimala för att få en utgångspunkt med så låg MSE som möjligt i IS prediktionsmodell. En sammanställning av MSE över test order, samt genomsnittlig MSE för modellen som helhet gjordes på samtliga modeller (se appendix C). Efter att alla variabler testats som utgångspunkt för AI:n, selekterades den fråga som bidrog till lägst MSE, vid utgångspunkten samt över modellen som helhet, manuellt. Den nyligen tillagda for-loopen och ändringen i koden med variabel “a” togs bort och ersattes med det nummer som den optimala utgångspunkten hade på X-axeln i `index1`, i det här fallet nummer 3 “`index1(Y,3)`”, för att implementera selektionen till koden. AI:n fungerar efter ändringarna likadant som innan dem, fast med en annan utgångspunkt med lägre MSE.

I koden efter brute-force metodens resultat analyserats och implementerats för att optimera utgångspunkten tog initieringsmomentet sig uttryck genom:

```
Y=1:length(trainVariables)
indexStart=indexText
indexText=index1(Y,3)
```

Den här förändringen i koden refereras hädanefter i arbetet som “*efter brute-force metoden*” eller “*den nya modellen*”.

Analysmetod

Mätredskapet blir effektivare om IS får ut så mycket information som möjligt från så få frågor som möjligt, vilket ter sig i de plottade figurerna genom att modellen genererar en så låg MSE som möjligt så tidigt i test order som möjligt. Därmed jämfördes MSE för varje fråga mellan baskoden och efter implementeringen av brute-force metoden “`indexText=index1(Y,3)`”. Medelvärdeskillnaden för genomsnittlig MSE över både test order och deltagare räknades ut, för att skapa en överblick på potentiella trender av förändring i MSE. Mellanskillnaderna av genomsnittliga MSE i test order räknades ut genom att för varje skala “`trainVariables`” summera varje frågeordnings (1-10) MSE, delat på antalet frågor (N=10) i den gamla modellen “`indexText=index1(Y,1)`”. Därefter subtraherades varje frågeordnings genomsnittliga MSE med motsvarande från den nya modellen “`indexText=index1(Y,3)`”. Proceduren upprepades därefter för samtliga skalor för

att få mellanskillnaden av den genomsnittliga MSE:n över test order i alla skalor “trainVariables”. Beräkningen av mellanskillnaden för deltagarnas genomsnittliga MSE räknades ut genom summering av varje deltagares MSE på varje skala “trainVariables” i den gamla modellen “indexText=index1(Y,1)”. Sedan delades värdet på andelen deltagare utan att använda saknade värden. Därefter subtraherades genomsnittliga MSE:n med motsvarande från den nya modellen “indexText=index1(Y,3)”.

Två beroende t-test utfördes för att med större säkerhet kunna undersöka om, och med vilken effektstyrka den nya koden efter brute-force metoden förbättrat modellen i helhet. Om den nya utgångspunkten “indexText=index1(Y,3)” för test order i IS innebär en statistiskt signifikant minskning av MSE över både test order (t9) men också av MSE över alla deltagare (t739), som i sin tur genererats från ett test med mer power, styrker det säkerheten av resultaten. Det första t-testet ämnade till att hypotetiskt testa ifall den nya utgångspunkten för IS i test order “indexText=index1(Y,3)” innebar en statistiskt signifikant mindre genomsnittlig MSE jämfört med MSE i baskoden med “indexText=index1(Y,1)” som utgångspunkt. Därmed användes genomsnittlig MSE från varje fråga i test order på varje skala “trainVariables” som datapunkt. Det andra t-testet ämnade till att undersöka om den nya utgångspunkten “indexText=index1(Y,3)” för test order i IS, innebar en statistiskt signifikant mindre genomsnittlig MSE för deltagarna. Således bestod datapunkterna av varje deltagares genomsnittliga MSE på varje skala “trainVariables”. Det ena t-testet testar IS genomsnittliga MSE över test order, det andra t-testet testar deltagares genomsnittliga MSE i IS.

Etik

Det pågående arbetet pågick i ett skede då ingen ytterligare data från deltagare samlades in, utöver redan insamlad data tidigare under projektet. Sveriges nationella etikprövningsmyndighet bedömde forskningsprojektets metodik och undersökning som ej behövande av en etisk prövning enligt svensk lag (etikdeklaration 2020-00730).

Vid tidigare tillfälle för datainsamling från deltagare gav samtliga informerat samtycke och deltog frivilligt med möjlighet att konsekvenslöst avsluta sitt deltagande när som helst. Samtliga deltagare tilldelades id-nummer och förblev därmed helt anonyma. Informationen som användes, (inklusive demografiskt information) kan inte spåras tillbaka till dessa individer, varken av oss eller av andra utomstående.

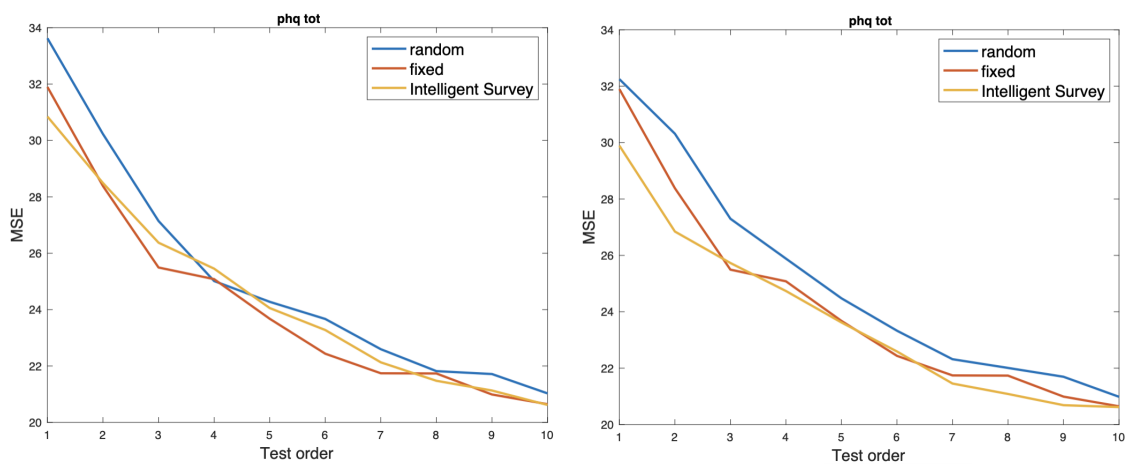
Resultat

PHQ-9

MSE för första frågan i PHQ-9 var innan brute-force metoden “ $\text{indexText}=\text{index1}(Y,1)$ ” 30,85 och 29,89 efter brute-force metoden “ $\text{indexText}=\text{index1}(Y,3)$ ”, en minskning på 0,96. I genomsnitt är MSE 0,659 lägre vid varje fråga i test order från den gamla modellen till den nya (se figur 1; tabell 1 och 2).

Figur 1

MSE för PHQ-9 över Test Order Innan och Efter Brute-Force Metoden.



Kommentar: Figuren till vänster, A, visar hur MSE i PHQ-9 ser ut innan brute-force

“ $\text{indexText}=\text{index1}(Y,1)$ ” metoden tillämpats, med frågan om ens generella mentala hälsa som utgångspunkt för IS regressionsmodell. Figuren till höger, B, visar hur MSE i PHQ-9 ser ut efter brute-force metoden tillämpats “ $\text{indexText}=\text{index1}(Y,3)$ ”, med frågan om depression som utgångspunkt för IS regressionsmodell. I den högra figuren visar IS en regressionsmodell med i genomsnitt 0,659 lägre MSE vid varje fråga i test order än den vänstra.

Tabell 1

MSE i Intelligent Survey över Test Order före och efter Brute-Force Metoden.

| Test order | PHQ | Post-BF | GAD | Post-BF | HILS | Post-BF | SWLS | Post-BF |
|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| | <i>MSE</i> | <i>MSE</i> | <i>MSE</i> | <i>MSE</i> | <i>MSE</i> | <i>MSE</i> | <i>MSE</i> | <i>MSE</i> |
| 1 | 30.85 | 29.89 | 22.1 | 21.86 | 27.52 | 27.02 | 34.74 | 34.77 |
| 2 | 28.48 | 26.85 | 20.57 | 19.74 | 26.15 | 24.43 | 33.91 | 33.54 |
| 3 | 26.37 | 25.73 | 19.53 | 18.33 | 25.25 | 24.04 | 33.9 | 33.33 |
| 4 | 25.45 | 24.73 | 18.79 | 18.19 | 24.71 | 23.59 | 32.84 | 32.62 |
| 5 | 24.06 | 23.62 | 18.27 | 18.3 | 23.98 | 22.95 | 32.56 | 32.88 |
| 6 | 23.27 | 22.59 | 17.99 | 17.8 | 23.79 | 22.88 | 31.68 | 32.47 |
| 7 | 22.13 | 21.45 | 17.82 | 17.55 | 23.23 | 22.72 | 31.89 | 32.25 |
| 8 | 21.48 | 21.08 | 17.42 | 17.19 | 22.86 | 22.7 | 31.95 | 32.46 |
| 9 | 21.13 | 20.69 | 16.86 | 16.69 | 22.87 | 22.89 | 32.24 | 32.33 |
| 10 | 20.62 | 20.62 | 16.36 | 16.36 | 23.41 | 23.41 | 32.69 | 32.7 |

Kommentar. Figuren visar MSE i Intelligent Survey över test order i PHQ-9, GAD-7, HILS och SWLS före och efter brute-force metoden implementerats. "Post-BF" innebär "efter brute-force" och syftar till resultaten från efter implementeringen "indexText=index1(Y, 3)" av brute-force metoden.

Tabell 2

Mellanskillnaden i MSE i Intelligent Survey över Test Order från Innan och Efter Brute-Force Metoden.

| Test order | PHQ <i>Mellanskillnad av MSE</i> | GAD <i>Mellanskillnad av MSE</i> | HILS <i>Mellanskillnad av MSE</i> | SWLS <i>Mellanskillnad av MSE</i> |
|------------|---|---|--|--|
| 1 | 0.96 | 0.24 | 0.5 | -0.03 |
| 2 | 1.63 | 0.83 | 1.72 | 0.37 |
| 3 | 0.64 | 1.2 | 1.21 | 0.57 |
| 4 | 0.72 | 0.6 | 1.12 | 0.22 |
| 5 | 0.44 | -0.03 | 1.03 | -0.32 |
| 6 | 0.68 | 0.19 | 0.91 | -0.79 |
| 7 | 0.68 | 0.27 | 0.51 | -0.36 |
| 8 | 0.4 | 0.23 | 0.16 | -0.51 |
| 9 | 0.44 | 0.17 | -0.02 | -0.09 |
| 10 | 0 | 0 | 0 | -0.01 |

Kommentar. Tabellen visar minskningen av MSE i IS prediktionsmodell över samtliga skalor, efter den nya utgångspunkten “indexText=index1 (Y, 3)“ implementerats, över test order. Negativa värden innebär en ökning av MSE i den nya modellen.

Medelvärdesskillnader efter brute-force metoden “indexText=index1 (Y, 3)“ visar att MSE på PHQ-9 överlag har sjunkit över hela modellen. Den genomsnittliga minskningen av MSE efter brute-force metoden för varje fråga i test order jämfört med den gamla modellen är 0,659 (se tabell 3) och för alla deltagare 0,894 (se tabell 5).

Det första t-testet $t_9 = 4,888$ ($p < 0,001$) visade att genomsnittlig MSE i baskoden var statistiskt signifikant med en uppskattad hög effektstyrka på Cohen's $d = 1,546$ (se tabell 3 och 4). Det andra t-testet $t_{739} = 1,7$ ($p = 0,045$) visade att baskoden var statistiskt signifikant större i baskoden och med en hög uppskattad effekstyrka på Cohen's $d = 0,0625$ (se tabell 5 och 6).

Tabell 3

Paired Samples T-test mellan Medelvärden för MSE över Test Order i Skalorna Innan och Efter Brute-Force Metoden.

| Paired Samples T-Test för test order | | T-statistics | df | p | Mean difference | SE difference | Effect Size | | |
|---|------------------|--------------|--------|------|-----------------|---------------|-------------|-----------|--------|
| PHQ MSE | Post-BF PHQ MSE | Student's t | 4,888 | 9,00 | < 0,001 | 0,6590 | 0,135 | Cohen's d | 1,546 |
| GAD MSE | Post-BF GAD MSE | Student's t | 3,000 | 9,00 | 0,007 | 0,3700 | 0,123 | Cohen's d | 0,949 |
| HILS MSE | Post-BF HILS MSE | Student's t | 3,908 | 9,00 | 0,002 | 0,7140 | 0,183 | Cohen's d | 1,236 |
| SWLS MSE | Post-BF SWLS MSE | Student's t | -0,725 | 9,00 | 0,757 | -0,0950 | 0,131 | Cohen's d | -0,229 |

Kommentar: $H_a \mu \text{ Measure 1} - \text{Measure 2} > 0$. Post-BF står för “efter brute-force” och syftar till de värden genererade av IS prediktionsmodell efter att brute-force metoden “indexText=index1 (Y, 3)” implementerats.

Tabell 4

Deskriptiva Data för MSE över Test Order I Samtliga Skolor Innan och Efter Brute-Force Metoden.

| Descriptives | | | | | |
|------------------|----|------|--------|-------|-------|
| | N | Mean | Median | SD | SE |
| PHQ MSE | 10 | 24,4 | 23,7 | 3,388 | 1,071 |
| Post-BF PHQ MSE | 10 | 23,7 | 23,1 | 3,075 | 0,973 |
| GAD MSE | 10 | 18,6 | 18,1 | 1,750 | 0,553 |
| Post-BF GAD MSE | 10 | 18,2 | 18,0 | 1,601 | 0,506 |
| HILS MSE | 10 | 24,4 | 23,9 | 1,536 | 0,486 |
| Post-BF HILS MSE | 10 | 23,7 | 23,2 | 1,317 | 0,416 |
| SWLS MSE | 10 | 32,8 | 32,6 | 1,021 | 0,323 |
| Post-BF SWLS MSE | 10 | 32,9 | 32,7 | 0,769 | 0,243 |

Kommentar: Post-BF syftar till “efter brute-force” och innefattar resultaten av IS prediktionsmodell efter brute-force metoden “indexText=index1 (Y, 3)” tillämpats.

Tabell 5

Paired Samples T-test mellan Medelvärden för MSE över Deltagare i Samtliga Skalor Innan och Efter Brute-Force Metoden.

| Paired Samples T-Test för deltagare | | T-statistics | df | p | Mean difference | SE difference | Effect Size | | |
|--|------------------|--------------|--------|-----|-----------------|---------------|-------------|-----------|--------|
| PHQ MSE | Post-BF PHQ MSE | Student's t | 1,7 | 739 | 0,045 | 0,894 | 0,526 | Cohen's d | 0,062 |
| GAD MSE | Post-BF GAD MSE | Student's t | 1,109 | 739 | 0,134 | 0,405 | 0,366 | Cohen's d | 0,041 |
| HILS MSE | Post-BF HILS MSE | Student's t | 1,401 | 739 | 0,081 | 0,640 | 0,457 | Cohen's d | 0,051 |
| SWLS MSE | Post-BF SWLS MSE | Student's t | -0,253 | 739 | 0,600 | -0,110 | 0,436 | Cohen's d | -0,009 |

Kommentar: $H_a \mu \text{ Measure 1} - \text{Measure 2} > 0$. Post-BF syftar till “efter brute-force” och innefattar resultaten av IS prediktionsmodell efter brute-force metoden “indexText=index1 (Y, 3)” tillämpats.

Tabell 6

Deskriptiva Data för MSE över Deltagare i Samtliga Skalor Innan och Efter Brute-Force Metoden.

| Descriptives | | | | | |
|------------------|-----|------|--------|------|------|
| | N | Mean | Median | SD | SE |
| PHQ MSE | 740 | 24,3 | 11,1 | 34,9 | 1,28 |
| Post-BF PHQ MSE | 740 | 23,4 | 10,3 | 33,7 | 1,24 |
| GAD MSE | 740 | 18,4 | 8,26 | 27,4 | 1,01 |
| Post-BF GAD MSE | 740 | 18,0 | 7,75 | 27,8 | 1,02 |
| HILS MSE | 740 | 24,6 | 11,76 | 36,2 | 1,33 |
| Post-BF HILS MSE | 740 | 23,9 | 11,33 | 35,2 | 1,29 |
| SWLS MSE | 740 | 32,9 | 16,16 | 45,7 | 1,68 |
| Post-BF SWLS MSE | 740 | 33,0 | 16,10 | 46,2 | 1,70 |

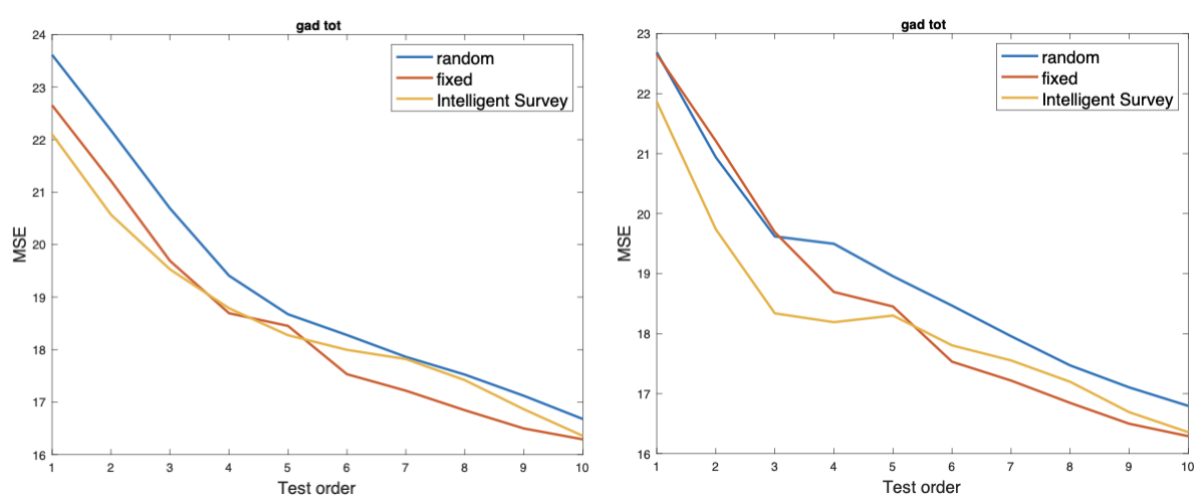
Kommentar: Post-BF syftar till “efter brute-force” och innefattar resultaten av IS prediktionsmodell efter att brute-force metoden “indexText=index1 (Y, 3)” tillämpats.

GAD-7

MSE för IS i GAD-7 var innan brute-force metoden 22,10 på första frågan och 21,86 efter brute-force metoden, vilket är en minskning på 0,24. I genomsnitt är MSE 0,370 lägre vid varje fråga i test order i den nya modellen “indexText=index1(Y, 3)” jämfört med den gamla “indexText=index1(Y, 1)”. (se figur 2; tabell 1 och 2).

Figur 2

MSE för GAD-7 över Test Order Innan och Efter Brute-Force Metoden.



Kommentar: Figuren till vänster visar MSE i GAD-7 innan brute-force metoden tillämpats

“indexText=index1(Y, 1)”, med frågan om ens generella mentala hälsa som utgångspunkt för IS. Figuren till höger visar MSE i GAD-7 efter brute-force metoden tillämpats “indexText=index1(Y, 3)”, med frågan om depression som utgångspunkt för IS. I den högra figuren visar IS en regressionsmodell med i genomsnitt 0,370 lägre MSE vid varje fråga i test order än den vänstra.

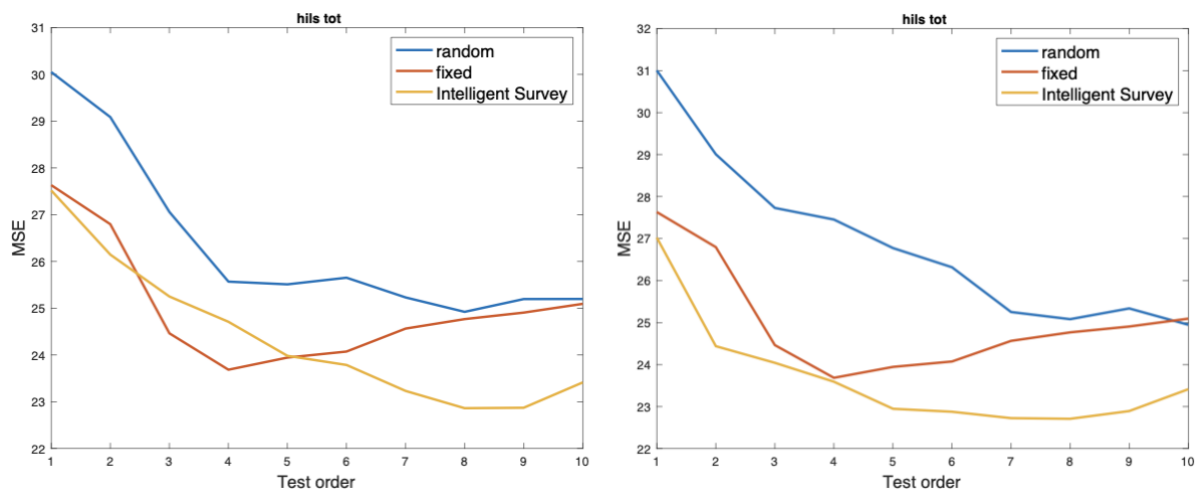
Resultaten efter brute-force metoden “indexText=index1(Y, 3)” visade att MSE på GAD-7 överlag har sjunkit över hela modellen med en mean difference på 0,370 i test order och 0,405 för deltagarna (se tabell 3 och 5). Första t-testet $t_9 = 3,0$ ($p = 0,007$) visade en statistiskt signifikant högre MSE i baskoden “indexText=index1(Y, 1)” och en uppskattat hög effektstyrka på Cohen’s $d = 0,949$ (se tabell 3 och 4). Det andra t-testet $t_{739} = 1,109$ ($p = 0,134$) var MSE i baskoden inte statistiskt signifikant större än efter brute-force metoden (se tabell 5 och 6).

HILS

På första frågan låg MSE i HILS innan brute-force metoden “ $\text{indexText}=\text{index1}(Y, 3)$ ” på 27,52 MSE och efter metoden “ $\text{indexText}=\text{index1}(Y, 3)$ ” på 27,02, en minskning på 0,50 vid interceptet. Den genomsnittliga minskningen av MSE för varje fråga i test order från den gamla modellen till den nya var 0,714. (se figur 3; tabell 1 och 2).

Figur 3

MSE för HILS över Test Order Innan och Efter Brute-Force Metoden.



Kommentar: Figuren till vänster visar MSE över test order i HILS innan brute-force metoden tillämpats “ $\text{indexText}=\text{index1}(Y, 1)$ “, med frågan om tillfredsställelse i livet som utgångspunkt för IS. Figuren till höger visar MSE över test order i HILS efter brute-force metoden tillämpats “ $\text{indexText}=\text{index1}(Y, 3)$ “, med frågan om ens självbild som utgångspunkt för IS. I den högra figuren visar IS en regressionsmodell med i genomsnitt 0,714 lägre MSE vid varje fråga i test order än den vänstra.

I HILS minskade MSE överlag efter brute-force metoden

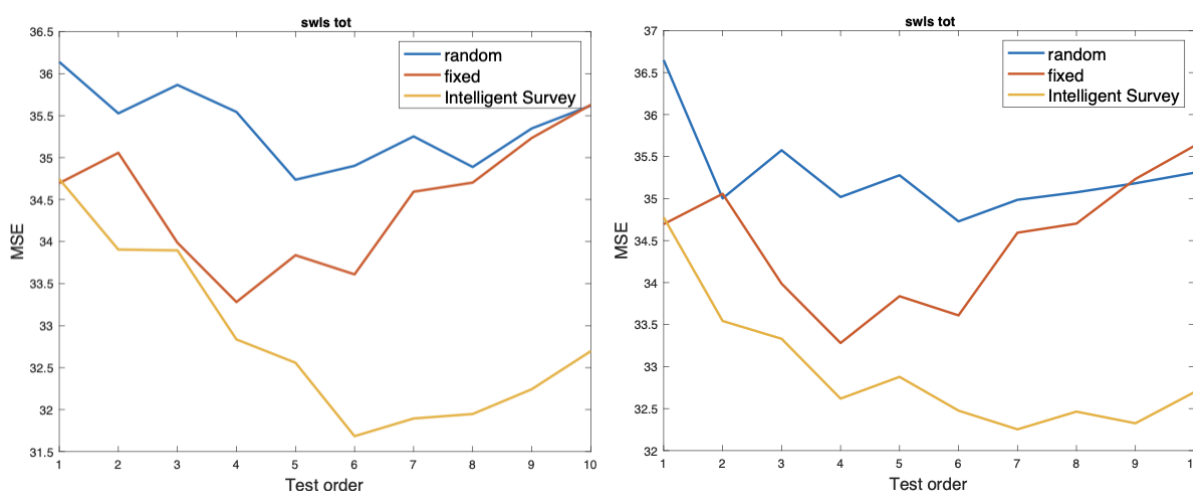
“ $\text{indexText}=\text{index1}(Y, 3)$ “, både över test order och över deltagares genomsnittliga MSE. Mellanvärdeskillnaderna för test orders genomsnittliga MSE låg på 0,714 och för deltagarnas genomsnittliga MSE 0,640 (se tabell 3 och 5). Det första t-testet $t_9 = 3,908$ ($p = 0,002$) visade att MSE i baskoden var signifikant högre och med en uppskattad hög effektstyrka på Cohen’s $d = 1,236$ (se tabell 3 och 4). I det andra t-testet $t_{739} = 1,401$ ($p = 0,081$) var MSE inte signifikant högre i baskoden (se tabell 5 och 6).

SWLS

Vid första frågan blev MSE i SWLS innan brute-force metoden “indexText=index1(Y,1)” 34,74 och 34,77 efter “indexText=index1(Y,3)”, en ökning på 0,03. Den genomsnittliga ökningen av MSE vid varje fråga i test order över hela modellen, i den nya modellen jämfört med den gamla var 0,095 (se figur 4; tabell 1 och 2).

Figur 4

MSE för SWLS över Test Order Innan och Efter Brute-Force Metoden.



Kommentar: Figuren till vänster visar MSE över test order i SWLS innan brute-force metoden tillämpats

“indexText=index1(Y,3)“, med frågan om ens tillfredsställelse i livet som utgångspunkt för IS. Figuren till höger visar MSE över test order i SWLS efter brute-force metoden tillämpats

“indexText=index1(Y,3)“, med frågan om ens självbild som utgångspunkt för IS. I den högra figuren visar IS en regressionsmodell med i genomsnitt 0,095 högre MSE vid varje fråga i test order jämfört med den vänstra.

Genomsnittligen ökade MSE på SWLS, både för test order och för deltagare, då mean difference i test order blev -0,095 MSE och -0,110 för deltagare efter brute-force metoden “indexText=index1(Y,3)” (se tabell 3 och 5).

Det första t-testet $t_9 = -0,725$ ($p = 0,757$) var MSE i baskoden inte signifikant högre än den efter brute-force metoden (se tabell 3 och 4). Det andra t-testet $t_{739} = -0,253$ ($p = 0,6$) gav heller inget signifikant resultat (se tabell 5 och 6).

Diskussion

Efter brute-force metoden med nya utgångspunkten “indexText=index1(Y, 3)” i stället för “indexText=index1(Y, 1)” visar PHQ-9 en minskning på 0,659 MSE för genomsnittet över test order och 0,894 MSE för genomsnittet över deltagare. Båda t-testen visade på statistisk signifikans med hög effektstyrka för PHQ-9. GAD-7 visade en minskning av genomsnittlig MSE över test order på 0,37 och 0,405 över deltagarna. För GAD-7 är det enbart t-testet för test order som är statistiskt signifikant, med en hög effektstyrka på 0,949. HILS visade minskning med 0,714 i test orders genomsnittliga MSE och 0,64 över deltagarnas genomsnittliga MSE. Av t-testen är enbart den för test order statistiskt signifikant för HILS, med en hög effektstyrka på 1,236. I SWLS ökar genomsnittlig MSE med -0,095 för test order och -0,110 för deltagarna. Inget av t-testen visar någon signifikans för SWLS.

Sammanfattningsvis innebär resultaten att prediktionsmodellen för IS efter förändringen med “indexText=index1(Y, 3)”, visar trender av förbättring överlag i majoriteten av skalorna, förutom SWLS där den i stället blivit sämre då MSE stigit. Dessutom har alla förutom just SWLS fått statistisk signifikans på åtminstone ett av t-testen.

Analys av resultat

Argumentet av Graham et al. (2019), att en försiktighet bör antas så att resultat inte övertolkas för AI inom hälsovård, var något som vi hade i åtanke och motiverade våra analyser av resultaten. Valet med utförandet av två t-test gjordes för att vi med större säkerhet skulle kunna undersöka om, och med vilken effektstyrka den nya koden från brute-force metoden förbättrat modellen i helhet. Något vi anser ha mer belegg för att göra om den nya utgångspunkten “indexText=index1(Y, 3)” för frågeordningen i IS innebär en statistiskt signifikant minskning av MSE för både test order (t_0) men också i MSE för alla deltagare (t_{739}) som i sin tur genererats från ett test med mer power.

Resultaten efter brute-force metoden med nya utgångspunkten “indexText=index1(Y, 3)” visar att användningen av frågan om depression, “*During the last two weeks, describe with five descriptive words whether you felt depression or not.*” i stället för “indexText=index1(Y, 1)” med frågan om ens generella mentala hälsa, “*Describe your mental health with five descriptive words.*”, överlag visar trender av lägre MSE både över test order och över deltagare i IS prediktionsmodell för PHQ-9. Mellanskillnaderna i PHQ-9 visar den högsta trendande minskningen av MSE. Dessutom är PHQ-9 den enda som fick en statistiskt signifikant mindre MSE i båda t-testen, men med en högre effektstyrka för test order (Cohen’s $d = 1,546$) än för deltagarna (Cohen’s $d = 0,0625$). Appliceringen av den

nya koden “`indexText=index1 (Y, 3)`” från brute-force metodens resultat har påverkat IS prestation över PHQ-9, som predicerar depression, genom en lägre MSE. Den signifikanta minskningen av MSE visar en modell som blivit mer effektiv i en sådan prediktion. Dessutom indikerar det andra t-testet på att förändringen signifikant förbättrat prediktionen av depression över deltagare generellt.

I GAD-7 med samma utgångspunkt “`indexText=index1 (Y, 3)`” med frågan om depression syns också trender av minskad MSE, men t-testen visade enbart en statistisk signifikant minskning i t-testet av test-order med en hög effektstyrka (Cohen’s $d = 0,949$). Vi tolkar resultaten som att den nya modellen skapat en optimal utgångspunkt för test order i IS och att den har till viss del förbättrat modellens prediktion av generaliserande ångestsyndrom. Däremot då det inte är statistiskt signifikant bättre i det andra t-testet är vi försiktiga med att dra slutsatser för hur stor påverkan, denna förbättring har för en sådan predicering för deltagare i helhet.

Resultaten av HILS med “`indexText=index1 (Y, 3)`” och frågan om ens självbild, “*Describe your self, with two descriptive words*” som utgångspunkt i test order för IS, visar likt som tidigare, trender av minskad MSE och därmed trender mot förbättring av modellen som helhet. Men precis som i GAD-7 visar enbart t-testet för test order på statistiskt signifikant mindre MSE, däremot med en högre effektstyrka än GAD-7 (Cohen’s $d = 1,236$). Därmed tolkar vi också resultaten i HILS likt som GAD-7, dvs. att vi tror den nya modellen signifikant förbättrat modellens prediktion av harmoni i livet, men att vi är försiktiga med hur pass stor påverkan förbättringen har för en sådan prediktion hos deltagare i helhet.

SWLS med sin nya utgångspunkt “`indexText=index1 (Y, 3)`” med frågan om ens självbild är den enda som visade trender på att modellen försämrats i mellanskillnaderna och var den enda skalan där ingen av t-testen är statistiskt signifikanta. Vi tolkar det som att den nya modellen i stället gjort att IS genom en ökad MSE, blivit lite sämre på att predicera tillfredsställelse med livet både över test order och över deltagare.

Resultaten av brute-force implementeringen var huvudsakligen positiva. Dock är forskningsområdet emellertid fortfarande inom ett tidigt skede, vilket kan innebära att metoder som används idag inom sin tid kan vara föråldrade eller att resultat kan övertolkas. Likt Graham et al. (2019) ståndpunkt om ett antagande om försiktighet inom övertolkning av resultat reserveras därför en varsamhet kring tolkningen av resultaten. Trots att en statistiskt signifikant förbättring går att finna i PHQ-9, både över test order och deltagare, finns det inte tillräckligt med gjorda studier på området av experter från olika discipliner för att med

säkerhet säga att resultaten från brute-force metoden talar för något fundamentalt inom förbättring av AI-system för prediktion av psykopatologier.

Analys frågeordningen

Konceptuellt är det inte otänkbart att ställa hypotesen att frågan om depression bör ha starkare prediktionseffekt på PHQ-9 jämfört med frågan om ens generella mentala hälsa, då den specifikt riktar sig mot uttryck för depression. Förbättringen av prediktionsmodellen för skalorna HILS och GAD-7 är mer intressant. Resultaten visar att uttryck för depression i större grad än uttryck för ens generella mentala hälsa minskar MSE vid prediktion för GAD-7. Detta är eventuellt inte helt oväntat att frågan som är bäst på att predicera depression och MDD även är bäst för att predicera GAD, eftersom depressions- och ångestsjukdomar är bland de vanligaste psykopatologierna samt att det finns en hög samsjuklighet för dessa diagnoser (Kalin, 2020). Dessutom är det skrivet i *Diagnostic and Statistical Manual of Mental Disorders* att personer med GAD mer ofta lider eller har lidit av någon form av unipolär depression (American Psychiatric Association. DSM-5 Task Force, 2013). Användandet av frågan om depression som utgångspunkt förbättrar prediktionsmodellen kraftigt under de första tre frågorna i ordningen relativt till de senare jämfört med användandet av frågan om ens generella mentala hälsa (se figur 2). Det tyder på att den nya modellen ger IS mer information på färre frågor.

Resultaten visar också att uttryck för ens självbild i större grad än uttryck för ens tillfredsställelse i livet minskar MSE vid prediktion för harmoni i livet, HILS. I studien av Kjell et al. (2016) hade HILS starkast korrelation till självacceptans och personlig utveckling samt mening med livet. HILS var dessutom mer korrelerat än SWLS till hur individer definierar sig själva. Således är det inte otänkbart att dessa korrelationer ligger till grund för att frågan *“Describe your self, with two descriptive words”* minimerar MSE mer än *“Overall in your life, describe with three descriptive words, whether you are satisfied or not?”*.

IS efter brute-force metoden i SWLS visar en förbättring av prediktionsmodellen för de första fyra frågorna i test order, men visar sedan försämring vid femte frågan och framåt. Hur resultaten för den nya modellen i SWLS ska tolkas beror på vilka prioriteringar i snabbhet och precision som finns. Om målet är att få en så bra prediktion som möjligt efter 1-4 frågor kan man tolka resultaten som positiva men små. Är målet å andra sidan att få en så precis prediktion som möjligt efter 5-10 frågor är resultatet negativt för SWLS. Överlag har förändringen i MSE över den totala modellen varit negativ då en summering av MSE-förändringen över samtliga frågor visar på en ökning i MSE på SWLS, till skillnad från

de andra skalorna. Den fråga som visar på störst förbättring av modellen som helhet jämfört med den gamla modellen är frågan om sömn, “*Describe your sleep with two descriptive words*”.

Metoddiskussion

Brute-force metoder är å ena sidan inte den mest effektiva form av metod, då de ställer höga krav på tid och datorkapacitet genom att beräkna samtliga möjligheter. Å andra sidan kan de vara väldigt informationsrika och bidra till en ökad överblick över de modeller som är möjliga under givna premisser. Att använda en brute-force metod för att identifiera hur IS modell ser ut med alla möjliga utgångspunkter gav en inblick i hur olika frågor ökar eller minskar MSE, vid interceptet såväl som för modellen som helhet, givet att de fungerar som utgångspunkt för AI:n.

Tillvägagångssättet med *brute-force* metoden fungerar genom att testa alla möjliga scenarion, vilket i sin tur tar lång tid för koden att köra igenom och ställer höga krav på tid och dator. Eftersom det endast var den första frågan som redan var samma för alla deltagare i AI:n kändes det passande att endast applicera brute-force metoden över den parametern, att testa alla möjliga utgångspunkter för AI:n att köra sin prediktionsmodell genom. Det var också mest praktiskt sett till tid och datorkraft då det tog 8 timmar för koden att köra igenom och testa de olika möjliga utgångspunkterna för AI:n. Att använda brute-force metoden på samtliga variabler, för att få fram alla möjliga kombinationer av frågeordningar över hela modellen, hade tagit cirka 968 timmar. Utöver det hade en fullständig brute-force applicering över samtliga variabler ersatt IS och blivit en fixed modell i ny ordning där deltagares frågeföljd inte baserats på svaren givna på föregående frågan, vilket hade förstört syftet med forskningsprojektet.

Metodens största fördel är att den provar alla möjliga alternativ. Den kan finna samtliga möjliga olika modeller som skiljer sig från den ursprungliga, inom de parametrar som satts, och kan hitta den som innebär minst MSE. Däremot är metoden, som nämnt ovan, mycket resurs- och tidskrävande.

Implementeringen av den optimala förstafrågan gjordes över samtliga skalor kollektivt, eftersom ändringen $\text{indexText}=\text{index1}(Y, 3)$ innebär att tredje frågan i index1 används som första över samtliga skalor. Det fungerade relativt bra då den tredje frågan i index1 var den optimala utgångspunkten i tre av fyra skalor. Dock, kan en tänka sig en tillämpning av de optimala utgångspunkterna individuellt. I skalorna PHQ-9, GAD-7 och HILS hade det fortfarande varit $\text{indexText}=\text{index1}(Y, 3)$, men i SWLS hade det varit

`indexText=index1(Y,4)`. Kunskapen för att göra en sådan individuell implementering av optimal utgångspunkt i koden var otillräcklig under detta arbete och gjordes därför inte, vilket lämnar utrymme för framtida forskning.

En annan möjlig metod för att förbättra IS prediktionsmodell som också hade inneburit en förändring av utgångspunkten hade varit att tillämpa en backwards elimination i IS regressionsmodell. Med "`indexText=index1(Y,:)`" hade IS utgått från en modell med alla frågor inlagda som utgångspunkt. Sedan genom att ändra i sorterings- och selektionsmekanismen i koden hade IS kunnat skala av den som är statistiskt minst signifikant i förklaringsmodellen iterativt tills det inte går att skala av mer utan att få en statistisk signifikant förlust i förklaringsmodellen. Resterande frågor inkluderas då i rangordning efter statistisk signifikans. Det är sannolikt att en backwards elimination hade inneburit andra resultat, men det är svårt att förutspå hur de resultaten hade sett ut. Det är också svårt att veta hur tillämpningen hade påverkat IS funktion, då det skulle kunna vara olika antal avskalade frågor över deltagarna. Det är inte heller uppenbart vad som hade hänt med de avskalade frågorna. Om de exkluderas helt ur beräkningen är datan svår att jämföra mellan deltagare då de har olika antal variabler, men om de inkluderas kan deras signifikansvärden förändras efter inkluderingen och någon premis för i vilken ordning de inkluderas måste ställas.

Inverkan på området och vidare forskning

Med minskande MSE och ökande effektivitet av AI:n går det att argumentera för att IS innebär en större konkurrens mot de mer traditionella skalorna som mäter psykologiska konstrukt, liksom resultaten av Kjell et al. (2019), med hjälp av det naturliga språket från öppna frågor. Då datorers förståelse av naturliga språket enligt Cambria, E. och White, B. (2014) är en speciellt viktig tillämpning inom mental hälsovård, är verktyg som IS tongivande i utvecklingen. IS som kliniskt verktyg för prediktioner av depression och ångest har genom brute-force metoden förbättrats ytterligare, vilket kan argumenteras för genom resultaten i studien. Det går även att argumentera för att IS är tidsmässigt konkurrerande mot traditionell diagnostisering, både för patienter och praktiserande inom mental hälsovård. Eftersom tidsåtgången för diagnostisering med IS till största del går till att respondenter svarar på de öppna frågorna kan det innebära mer tid åt de praktiserande såväl som patienterna, om IS tillåts utföra diagnostiseringen istället för en långvarig diagnostisering av en psykolog. IS går ihop mer adekvat till idén om att AI och algoritmer för databehandling kan utgöra en viktig

del i behandling av psykisk ohälsa, eftersom det kan frigöra mycket tid och uppmärksamhet för de praktiserande (Kim, J.W. et al., 2019; Topol, E., 2019).

I linje med vad som uttrycks om datoriseringens potential inom diagnostisering och skattning av psykisk ohälsa i Dwyer et al. (2018) och Kjell et al. (2019) visar resultaten av IS och brute-force metoden på exempel hur den psykiska hälsovården potentiellt kan utvecklas genom datoriseringen. Att tolka människors känslor utifrån vad de säger eller skriver har än så länge varit människors uppgift, inte datorers. Det är för att människor, inte minst psykologer och terapeuter, förstår den semantiska innebörden av ord. Människor har dock i jämförelse med datorer en begränsad förmåga att behandla stora mängder data, vilket skapar ett incitament för möjligheten att låta datorer behandla den typen av data. Genom att lära datorer att förstå naturligt språk kan de även analysera det och göra insikter om mönster eller prediktorer på minuter eller timmar som hade tagit de flesta människor dagar eller år. Den här typen av analys av ord för att skatta och diagnostisera depression och ångest som IS gör hade inte varit möjlig utan beräknings-kraften som datorer besitter.

Trots de lovande stora framgångarna med AI och datorisering som förstår det naturliga språket, har processen i projektet visat att stor vikt ligger på att det även finns en förståelse för funktionen och applicerbarheten av AI och datoriseringen hos de praktiserande. Likt som Graham et al. (2019) menade att experter från olika discipliner behöver vara delaktiga i introduktionen och optimeringen av AI inom vården, är det även viktigt att lyfta att expertisen kommer vara nödvändig för att tekniken ska kunna användas fortlöpande. Därmed att optimeringen och hanteringen av AI:n eventuellt blir en ny del av arbetsuppgifterna hos de som praktiserar inom mental hälsovård och därmed förändrar professionen fundamentalt. Med andra ord att den fundamentala förändringen göra att den ovannämnda frigjorda tiden och uppmärksamheten för att fokusera på mänskliga aspekter, eventuellt skiftar till att exempelvis psykologer kommer spendera mer tid till datoriserade aspekter och AI. En förändring som eventuellt inte kommer uppskattas hos praktiserande som exempelvis valt sitt yrke på grund av den höga graden av mänsklig interaktion. Således väcker det viktiga och intressanta framtida forskningsfrågor, om hur attityder ser ut hos de praktiserande för en sådan potentiell förändring.

Användandet av AI för att predicera depression genom text-data och naturligt språk är ett växande område. Resultaten av IS går i linje med många andra studiers resultat och visar att det är en passande metod med hög potential för att göra bra prediktioner (Al Hanai, T. et al., 2018; Uddin, M.Z. et al., 2022). Vidare kan Intelligent Survey som verktyg potentiellt användas inom andra studier, projekt eller liknande verktyg för att inkorporera dess

prediktionsförmåga. Brute-force metoden ger dessutom en fingervisning för exempel på hur man kan utveckla de AI:n som redan idag opererar eller ämnar att göra så inom prediktioner för psykisk ohälsa. Exempelvis menade Graham et al (2019) att det är viktigt för framtida utveckling för datorprogram i mental hälsa också behöver kunna diagnostisera gravheten av sjukdomar något som saknats inom området. IS använder skalor som kan mäta severity (PHQ-9 & GAD-7) och inte bara diagnostisera, trots det är AI:n enbart tränad för att bara predicera och diagnostisera. Därmed hade det varit bra för framtida forskning att utföra tester med detta i fokus i stället för enbart binär diagnostisering.

Det hade varit intressant för vidare utveckling av projektet att kolla hur väl det fungerar att ha den nya första frågan som första fråga i formuläret när de besvaras av deltagare. Därefter generera nästkommande frågor på samma sätt som IS beräknas med för att se hur det ter sig i verkligheten och om de svaren skiljer sig mot de beräkningar som finns just nu. I den insamlade datan från IS fick alla deltagare svara på frågorna i en slumpmässig ordning, vilket säkerställer att confounders i form av frågornas ordning randomiseras ut och därmed inte påverkar den totala datan som används i beräkningarna. Däremot för varje individ går det inte att utsluta att detta sker, vilket är en nackdel då vi inte kan kontrollera ifall frågorna har någon påverkan på hur individerna svarar på de andra frågorna.

Slutsats

Denna studie visar att brute-force metoden är lämplig att använda för att utforska AI:n IS olika utgångspunkter och därmed ett bra verktyg för att undersöka hur ordningen på frågorna påverkar MSE. Följaktligen ledde metoden till att en statistiskt signifikant minskning av MSE över test order medföljde den nya modellen i PHQ-9, GAD-7 och HILS men inte SWLS som visade en liten ökning av MSE. I t-testen över deltagarnas genomsnittliga MSE är det enbart i PHQ-9 som en statistiskt signifikant minskning av MSE uppvisas. Således indikerar t-testen att mätredskapet verkar ha blivit bättre på att predicera depression, GAD och harmoni i livet. Däremot att en skepticism bör vidtas för förbättringens faktiska effekt på prediktionerna för deltagare. Således verkar brute-force metoden visat prov på tillämpbar användbarhet med vissa signifikanta effekter på IS. Men med tanke på metodens resurskrav och oförmågenhet att tillämpas över flera variabler utan att förhindra AI:ns funktion, samt forskningsområdets infantila stadie, är det inte uteslutet att andra metoder hade haft större effekt på minskningen av MSE.

Referenser

- Al Hanai, T., Ghassemi, M. M., & Glass, J. R. (2018). Detecting Depression with Audio/Text Sequence Modeling of Interviews. *Interspeech 2018*, 1716-1720.
<http://dx.doi.org/10.21437/Interspeech.2018-2522>
- American Psychiatric Association. DSM-5 Task Force. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5*. (5th ed.)
- Bzdok, D., & Meyer-Lindenberg, A. (2018). Machine Learning for Precision Psychiatry: Opportunities and Challenges. *Biological psychiatry. Cognitive neuroscience and neuroimaging*, 3(3), 223–230. <https://doi.org/10.1016/j.bpsc.2017.11.007>
- Cambria, E. & White, B. (2014). Jumping NLP curves: A Review of Natural Language Processing Research [Review Article]. *IEEE Computational Intelligence Magazine*. 9(2), 48-57. <https://doi.org/10.1109/MCI.2014.2307227>
- Casey, K. (2019, 17 September). *How to explain natural language processing (NLP) in plain English*. The enterprisers project. Hämtad 4 maj, 2022, från <https://enterpriseproject.com/article/2019/9/natural-language-processing-nlp-explained-plain-english>
- Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction With Life Scale. *Journal of personality assessment*, 49(1), 71–75.
https://doi.org/10.1207/s15327752jpa4901_13
- Dumais, S. T., Furnas, G. W., Landauer, T. K., Deerwester, S., & Harshman, R. (1988). Using latent semantic analysis to improve access to textual information. *Proceedings of the SIGCHI Conference: Human Factors in Computing Systems (9780201142372)*, 281–285. <https://doi.org/10.1145/57167.57214>
- Dwyer, D. B., Falkai, P., & Koutsouleris, N. (2018). Machine Learning Approaches for Clinical Psychology and Psychiatry. *Annual review of clinical psychology*, 14(1), 91–118. <https://doi.org/10.1146/annurev-clinpsy-032816-045037>

- Graham, S., Depp, C., Lee, E., Nebeker, C., Tu, X., Kim, H., & Jeste, D. (2019). Artificial Intelligence for Mental Health and Mental Illnesses: an Overview. *Current psychiatry reports*, 21(11), 116. <http://dx.doi.org/10.1007/s11920-019-1094-0>
- IBM Cloud Education. (2020, 2 Juli). *Natural language processing (NLP)*. IBM.com. Hämtad den 4 maj, 2022, från <https://www.ibm.com/cloud/learn/natural-language-processing>
- Kalin, N. H. (2020). The Critical Relationship Between Anxiety and Depression. *American Journal of Psychiatry*, 177(5), 365–367. <https://doi.org/10.1176/appi.ajp.2020.20030305>
- Kim, J. W., Jones, K. L., & D'Angelo, E. (2019). How to Prepare Prospective Psychiatrists in the Era of Artificial Intelligence. *Academic psychiatry : the journal of the American Association of Directors of Psychiatric Residency Training and the Association for Academic Psychiatry*, 43(3), 337–339. <https://doi.org/10.1007/s40596-019-01025-x>
- Kjell, O., Daukantaitė, D., Hefferon, K., & Sikström, S. (2016). The Harmony in Life Scale Complements the Satisfaction with Life Scale: Expanding the Conceptualization of the Cognitive Component of Subjective Well-Being. *Social Indicators Research*, 126, 893–919. <https://doi.org/10.1007/s11205-015-0903-z>
- Kjell, O., Kjell, K., Garcia, D., & Sikström, S. (2019). Semantic measures: Using natural language processing to measure, differentiate, and describe psychological constructs. *Psychological methods*, 24(1), 92–115. <https://doi.org/10.1037/met0000191>
- Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of general internal medicine*, 16(9), 606–613. <https://doi.org/10.1046/j.1525-1497.2001.016009606.x>
- Shatte, A., Hutchinson, D. M., & Teague, S. J. (2019). Machine learning in mental health: a scoping review of methods and applications. *Psychological medicine*, 49(9), 1426–1448. <https://doi.org/10.1017/S0033291719000151>

- Sikström, S., Kjell, O.N.E., Kjell, K. (2020). SemanticExcel.com: An Online Software for Statistical Analyses of Text Data Based on Natural Language Processing. In: Sikström, S., Garcia, D. (Red.) *Statistical Semantics* (ss.87-103). Springer, Cham.
https://doi.org/10.1007/978-3-030-37250-7_6
- Spitzer, R. L., Kroenke, K., Williams, J. B., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of internal medicine*, 166(10), 1092–1097. <https://doi.org/10.1001/archinte.166.10.1092>
- Topol, E. J. (2019). *Deep medicine: How artificial intelligence can make healthcare human again*. (First edition). Basic Books.
- Uddin, M. Z., Dysthe, K. K., Følstad, A. & Brandtzaeg, P. A. (2022). Deep learning for prediction of depressive symptoms in a large textual dataset. *Neural Computing & Applications* 34, 721–744. <https://doi.org/10.1007/s00521-021-06426-4>
- World Health Organisation. (u.å.). *Mental Health*. World Health Organisation. Hämtad 28 april 2022 från https://www.who.int/health-topics/mental-health#tab=tab_2
- World Health Organisation. (2021). *Depression*. World Health Organisation. Hämtad 28 april 2022 från <https://www.who.int/news-room/fact-sheets/detail/depression>

Appendix A

Öppna svarsfrågor

Fråga 1, Dep_all: for *depression* “During the last two weeks, describe with five descriptive words whether you felt depression or not”,

Fråga 2, Wor_all: *anxiety* was “During the last two weeks, describe with five descriptive words whether you have you felt anxious or not”,

Fråga 3, Gen_all: The question for *general* mental health was “Describe your mental health with five descriptive words”

Fråga 4, Sleep: “Describe your sleep with two descriptive words”.

Fråga 5, Concentration: “Describe your concentration with two descriptive words”.

Fråga 6, Appetite: “Describe your appetite with two descriptive words”.

Fråga 7, Energy: “Describe your sleep energy with two descriptive words”.

Fråga 8, Self: “Describe your self, with two descriptive words”.

Fråga 9, movement: “Describe your movement with two descriptive words”.

Fråga 10, Har_all: “Overall in your life, describe with three descriptive whether you are in harmony or not?”.

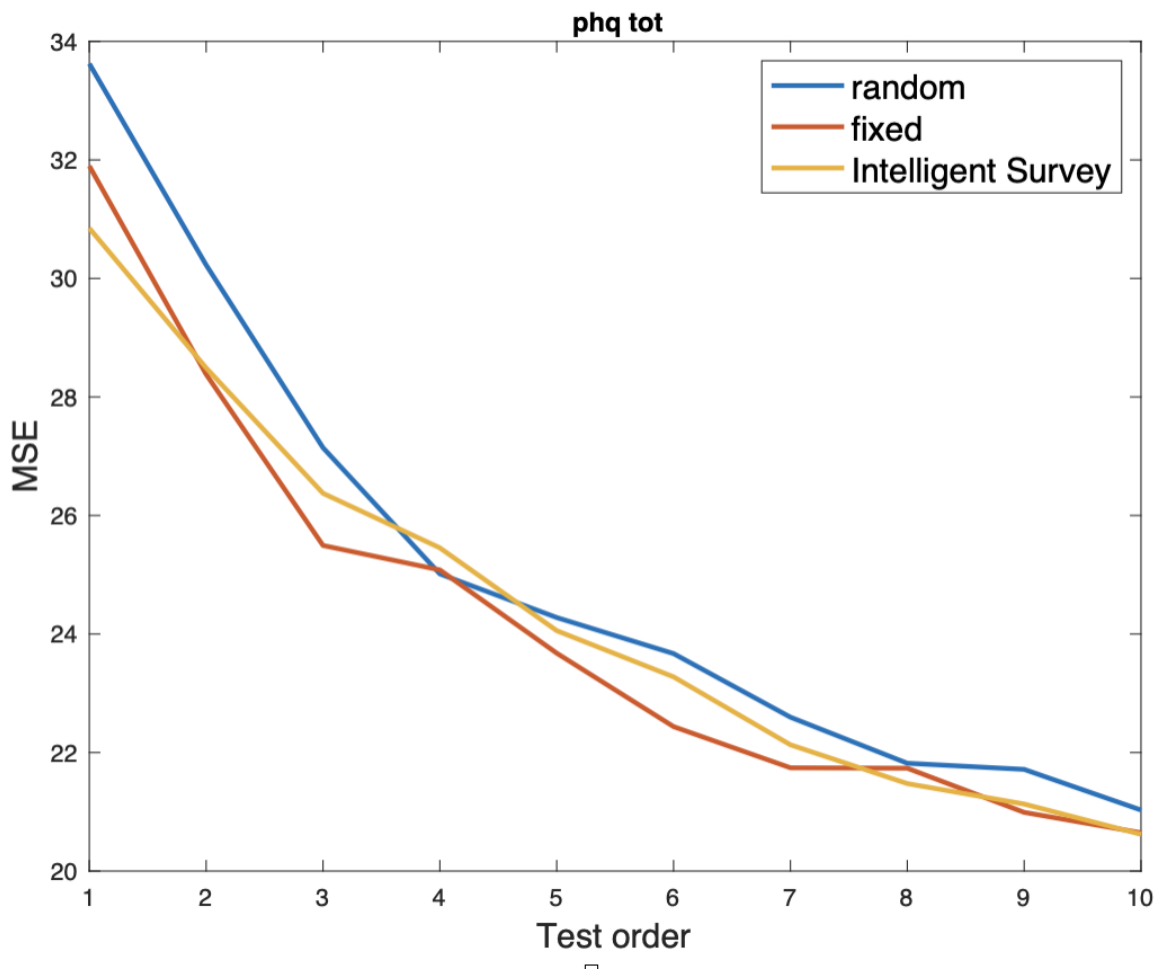
Fråga 11, Sat_all: “Overall in your life, describe with three descriptive words, whether you are satisfied or not?”

Appendix B

Resultat från baskoden

Figur B1

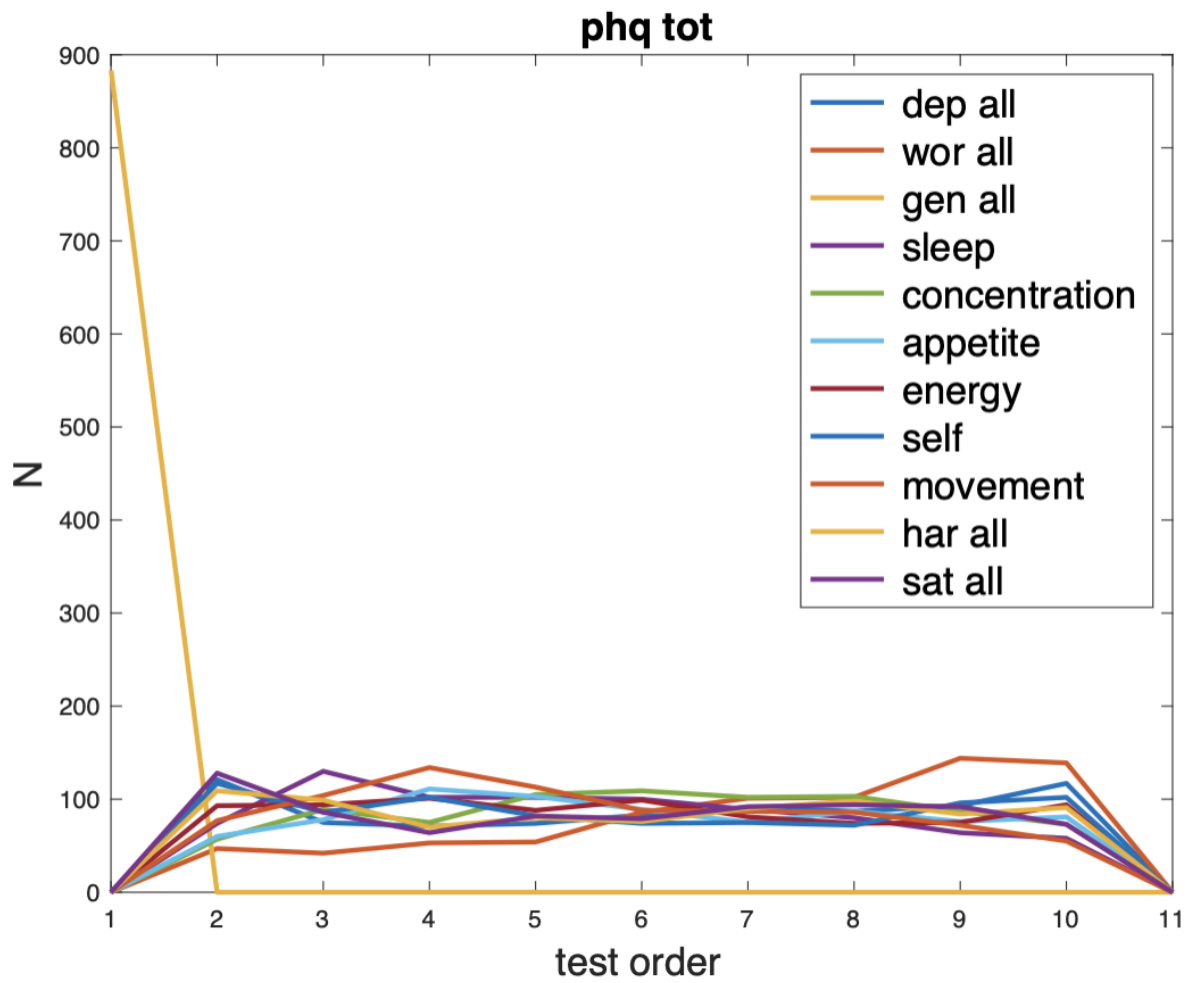
MSE för PHQ-9 över Test Order.



Kommentar: Figuren visar MSE för de olika modellerna Random, Fixed och Intelligent Survey över test order i PHQ-9. Figuren visar resultaten innan brute-force metoden.

Figur B2

Antal Deltagare för Varje Fråga över Test Order i PHQ-9.



Kommentar: Figuren visar hur många deltagare som i IS prediktionsmodell har fått respektive fråga i respektive ordning över test order i PHQ-9. Figuren visar resultaten innan brute-force metoden.

Tabell B1

Tabell för MSE i Intelligent Survey över Test Order i PHQ-9, GAD-7, HILS och SWLS.

| Test order | PHQ | GAD | HILS | SWLS |
|------------|-------|-------|-------|-------|
| 1 | 30.85 | 22.1 | 27.52 | 34.74 |
| 2 | 28.48 | 20.57 | 26.15 | 33.91 |
| 3 | 26.37 | 19.53 | 25.25 | 33.9 |
| 4 | 25.45 | 18.79 | 24.71 | 32.84 |
| 5 | 24.06 | 18.27 | 23.98 | 32.56 |
| 6 | 23.27 | 17.99 | 23.79 | 31.68 |
| 7 | 22.13 | 17.82 | 23.23 | 31.89 |
| 8 | 21.48 | 17.42 | 22.86 | 31.95 |
| 9 | 21.13 | 16.86 | 22.87 | 32.24 |
| 10 | 20.62 | 16.36 | 23.41 | 32.69 |

Kommentar: Tabellen visar MSE i Intelligent Survey över test order i ska PHQ-9, GAD-7, HILS och SWLS innan brute-force metoden implementerats, med baskoden som underlag för prediktionsmodellen.

Appendix C

Samtliga modeller från brute-force metoden med MSE för varje test-order samt genomsnittet för respektive modell.

Tabell C1

MSE över frågeordningen med alla möjliga alternativ som förstafrågan

| PHQ | | GAD | | HILS | | SWLS | |
|------------------------|-------------------------------|------------------------|------------------------------|------------------------|--------------------------------|------------------------|--------------------------------|
| indexText=index1 (Y,1) | | indexText=index1 (Y,1) | | indexText=index1 (Y,1) | | indexText=index1 (Y,1) | |
| Test Order | MSE - PHQ fråga 3 som första | Test Order | MSE - GAD fråga 3 som första | Test Order | MSE - HILS fråga 11 som första | Test Order | MSE - SWLS fråga 11 som första |
| 1 | 30.848 | 1 | 22.1 | 1 | 27.52 | 1 | 34.74 |
| 2 | 28.489 | 2 | 20.57 | 2 | 26.15 | 2 | 33.91 |
| 3 | 26.374 | 3 | 19.53 | 3 | 25.25 | 3 | 33.9 |
| 4 | 25.453 | 4 | 18.79 | 4 | 24.71 | 4 | 32.84 |
| 5 | 24.056 | 5 | 18.27 | 5 | 23.98 | 5 | 32.56 |
| 6 | 23.276 | 6 | 17.99 | 6 | 23.79 | 6 | 31.68 |
| 7 | 22.13 | 7 | 17.82 | 7 | 23.23 | 7 | 31.89 |
| 8 | 21.476 | 8 | 17.42 | 8 | 22.86 | 8 | 31.95 |
| 9 | 21.131 | 9 | 16.86 | 9 | 22.87 | 9 | 32.24 |
| 10 | 20.62 | 10 | 16.36 | 10 | 23.41 | 10 | 32.7 |
| Mean | 24.3853 | Mean | 18.571 | Mean | 24.377 | Mean | 32.841 |
| indexText=index1 (Y,2) | | indexText=index1 (Y,2) | | indexText=index1 (Y,2) | | indexText=index1 (Y,2) | |
| Test Order | MSE - PHQ fråga 8 som första | Test Order | MSE - GAD fråga 8 som första | Test Order | MSE - HILS fråga 10 som första | Test Order | MSE - SWLS fråga 10 som första |
| 1 | 30.208 | 1 | 22.33 | 1 | 27.51 | 1 | 34.74 |
| 2 | 28.54 | 2 | 21.07 | 2 | 25.33 | 2 | 34.06 |
| 3 | 27.638 | 3 | 20.32 | 3 | 24.88 | 3 | 33.35 |
| 4 | 25.925 | 4 | 19.61 | 4 | 23.77 | 4 | 32.15 |
| 5 | 24.452 | 5 | 19.05 | 5 | 23.02 | 5 | 31.63 |
| 6 | 24.018 | 6 | 18.31 | 6 | 22.45 | 6 | 31.91 |
| 7 | 22.807 | 7 | 17.9 | 7 | 22.96 | 7 | 31.67 |
| 8 | 22.397 | 8 | 17.32 | 8 | 22.86 | 8 | 32.23 |
| 9 | 21.552 | 9 | 16.82 | 9 | 22.97 | 9 | 32.9 |
| 10 | 20.62 | 10 | 16.36 | 10 | 23.41 | 10 | 32.7 |
| Mean | 24.8157 | Mean | 18.909 | Mean | 23.916 | Mean | 32.734 |
| indexText=index1 (Y,3) | | indexText=index1 (Y,3) | | indexText=index1 (Y,3) | | indexText=index1 (Y,3) | |
| Test Order | MSE - PHQ-fråga 1 som första | Test Order | MSE - GAD fråga 1 som första | Test Order | MSE - HILS fråga 8 som första | Test Order | MSE - SWLS fråga 8 som första |
| 1 | 28.892 | 1 | 21.86 | 1 | 27.02 | 1 | 34.77 |
| 2 | 26.846 | 2 | 19.74 | 2 | 24.43 | 2 | 33.54 |
| 3 | 25.73 | 3 | 18.33 | 3 | 24.04 | 3 | 33.33 |
| 4 | 24.73 | 4 | 18.19 | 4 | 23.59 | 4 | 32.62 |
| 5 | 23.62 | 5 | 18.3 | 5 | 22.95 | 5 | 32.88 |
| 6 | 22.59 | 6 | 17.8 | 6 | 22.88 | 6 | 32.47 |
| 7 | 21.45 | 7 | 17.55 | 7 | 22.72 | 7 | 32.25 |
| 8 | 21.08 | 8 | 17.19 | 8 | 22.7 | 8 | 32.46 |
| 9 | 20.69 | 9 | 16.69 | 9 | 22.89 | 9 | 32.33 |
| 10 | 20.62 | 10 | 16.36 | 10 | 23.41 | 10 | 32.7 |
| Mean | 23.6248 | Mean | 18.201 | Mean | 23.663 | Mean | 32.935 |
| indexText=index1 (Y,4) | | indexText=index1 (Y,4) | | indexText=index1 (Y,4) | | indexText=index1 (Y,4) | |
| Test Order | MSE - PHQ-fråga 11 som första | Test Order | MSE - GAD fråga 9 som första | Test Order | MSE - HILS fråga 3 som första | Test Order | MSE - SWLS fråga 3 som första |
| 1 | 35.511 | 1 | 22.58 | 1 | 28.4 | 1 | 33.72 |
| 2 | 30.456 | 2 | 20.76 | 2 | 24.84 | 2 | 32.76 |
| 3 | 28.342 | 3 | 19.96 | 3 | 24.59 | 3 | 32.63 |
| 4 | 26.241 | 4 | 19.01 | 4 | 23.77 | 4 | 32.3 |
| 5 | 25.068 | 5 | 18.63 | 5 | 23.4 | 5 | 31.52 |
| 6 | 23.777 | 6 | 18.34 | 6 | 23.17 | 6 | 31.28 |
| 7 | 22.891 | 7 | 18.02 | 7 | 23.13 | 7 | 31.41 |
| 8 | 21.832 | 8 | 17.5 | 8 | 23.2 | 8 | 32.27 |
| 9 | 21.41 | 9 | 16.83 | 9 | 22.94 | 9 | 32.62 |
| 10 | 20.62 | 10 | 16.36 | 10 | 23.41 | 10 | 32.7 |
| Mean | 25.6148 | Mean | 18.799 | Mean | 24.085 | Mean | 32.321 |

| indexText=index1 (Y,5) | | indexText=index1 (Y,5) | | indexText=index1 (Y,5) | | indexText=index1 (Y,5) | |
|------------------------|-------------------------------|------------------------|-------------------------------|------------------------|-------------------------------|------------------------|-------------------------------|
| Test Order | MSE - PHQ fråga 10 som första | Test Order | MSE - GAD fråga 11 som första | Test Order | MSE - HILS fråga 1 som första | Test Order | MSE - SWLS fråga 1 som första |
| 1 | 35.51 | 1 | 25.43 | 1 | 29.28 | 1 | 35.18 |
| 2 | 30.223 | 2 | 21.17 | 2 | 25.69 | 2 | 33.06 |
| 3 | 26.803 | 3 | 19.89 | 3 | 25.07 | 3 | 32.28 |
| 4 | 25.052 | 4 | 19.3 | 4 | 24.66 | 4 | 32.05 |
| 5 | 23.772 | 5 | 18.2 | 5 | 23.66 | 5 | 31.99 |
| 6 | 22.801 | 6 | 17.91 | 6 | 22.95 | 6 | 31.2 |
| 7 | 22.209 | 7 | 17.73 | 7 | 23.17 | 7 | 31.46 |
| 8 | 21.793 | 8 | 17.43 | 8 | 22.85 | 8 | 32.03 |
| 9 | 20.981 | 9 | 16.84 | 9 | 22.66 | 9 | 32.49 |
| 10 | 20.62 | 10 | 16.36 | 10 | 23.41 | 10 | 32.7 |
| Mean | 24.9764 | Mean | 19.026 | Mean | 24.34 | Mean | 32.444 |

| indexText=index1 (Y,6) | | indexText=index1 (Y,6) | | indexText=index1 (Y,6) | | indexText=index1 (Y,6) | |
|------------------------|------------------------------|------------------------|------------------------------|------------------------|-------------------------------|------------------------|-------------------------------|
| Test Order | MSE - PHQ fråga 9 som första | Test Order | MSE - GAD fråga 2 som första | Test Order | MSE - HILS fråga 7 som första | Test Order | MSE - SWLS fråga 7 som första |
| 1 | 32.51 | 1 | 22.76 | 1 | 30.65 | 1 | 34.51 |
| 2 | 28.73 | 2 | 19.83 | 2 | 26.8 | 2 | 32.86 |
| 3 | 27.15 | 3 | 19.17 | 3 | 25.44 | 3 | 32.18 |
| 4 | 25.38 | 4 | 18.2 | 4 | 24.59 | 4 | 32.05 |
| 5 | 24.19 | 5 | 17.89 | 5 | 23.99 | 5 | 32.44 |
| 6 | 23.45 | 6 | 17.56 | 6 | 23.58 | 6 | 32.25 |
| 7 | 22.1 | 7 | 17.28 | 7 | 23.13 | 7 | 32.04 |
| 8 | 21.88 | 8 | 17.03 | 8 | 23 | 8 | 32.33 |
| 9 | 21.2 | 9 | 16.63 | 9 | 22.87 | 9 | 32.45 |
| 10 | 20.62 | 10 | 16.36 | 10 | 23.41 | 10 | 32.7 |
| Mean | 24.721 | Mean | 18.271 | Mean | 24.746 | Mean | 32.581 |

| indexText=index1 (Y,7) | | indexText=index1 (Y,7) | | indexText=index1 (Y,7) | | indexText=index1 (Y,7) | |
|------------------------|------------------------------|------------------------|-------------------------------|------------------------|-------------------------------|------------------------|-------------------------------|
| Test Order | MSE - PHQ fråga 7 som första | Test Order | MSE - GAD fråga 10 som första | Test Order | MSE - HILS fråga 2 som första | Test Order | MSE - SWLS fråga 2 som första |
| 1 | 32.38 | 1 | 25.43 | 1 | 31.43 | 1 | 37.17 |
| 2 | 28.28 | 2 | 21.36 | 2 | 27.13 | 2 | 34.93 |
| 3 | 26.4 | 3 | 19.53 | 3 | 25.65 | 3 | 35 |
| 4 | 24.62 | 4 | 19.59 | 4 | 25.09 | 4 | 33.95 |
| 5 | 24 | 5 | 18.91 | 5 | 23.87 | 5 | 32.91 |
| 6 | 23 | 6 | 18.18 | 6 | 23.46 | 6 | 32.44 |
| 7 | 21.92 | 7 | 17.92 | 7 | 23.23 | 7 | 32.2 |
| 8 | 21.84 | 8 | 17.52 | 8 | 23.33 | 8 | 32.13 |
| 9 | 21.29 | 9 | 16.65 | 9 | 23.07 | 9 | 32.28 |
| 10 | 20.62 | 10 | 16.36 | 10 | 23.41 | 10 | 32.7 |
| Mean | 24.435 | Mean | 19.145 | Mean | 24.967 | Mean | 33.571 |

| indexText=index1 (Y,8) | | indexText=index1 (Y,8) | | indexText=index1 (Y,8) | | indexText=index1 (Y,8) | |
|------------------------|------------------------------|------------------------|------------------------------|------------------------|-------------------------------|------------------------|-------------------------------|
| Test Order | MSE - PHQ fråga 2 som första | Test Order | MSE - GAD fråga 5 som första | Test Order | MSE - HILS fråga 4 som första | Test Order | MSE - SWLS fråga 5 som första |
| 1 | 34.93 | 1 | 22.85 | 1 | 31.73 | 1 | 36.03 |
| 2 | 30.92 | 2 | 21 | 2 | 28.11 | 2 | 34.39 |
| 3 | 28.86 | 3 | 19.71 | 3 | 26.44 | 3 | 33.94 |
| 4 | 26.84 | 4 | 18.49 | 4 | 25.41 | 4 | 33.83 |
| 5 | 25.54 | 5 | 17.96 | 5 | 24.17 | 5 | 33.41 |
| 6 | 23.94 | 6 | 17.3 | 6 | 23.31 | 6 | 32.9 |
| 7 | 22.81 | 7 | 17.34 | 7 | 23.32 | 7 | 32.92 |
| 8 | 21.93 | 8 | 16.72 | 8 | 23.4 | 8 | 32.65 |
| 9 | 21.25 | 9 | 16.75 | 9 | 23.06 | 9 | 32.77 |
| 10 | 20.62 | 10 | 16.36 | 10 | 23.41 | 10 | 32.7 |
| Mean | 25.764 | Mean | 18.448 | Mean | 25.236 | Mean | 33.554 |

| indexText=index1 (Y,9) | | indexText=index1 (Y,9) | | indexText=index1 (Y,9) | | indexText=index1 (Y,9) | |
|------------------------|------------------------------|------------------------|------------------------------|------------------------|-------------------------------|------------------------|-------------------------------|
| Test Order | MSE - PHQ fråga 4 som första | Test Order | MSE - GAD fråga 7 som första | Test Order | MSE - HILS fråga 9 som första | Test Order | MSE - SWLS fråga 4 som första |
| 1 | 34.11 | 1 | 23.81 | 1 | 31.6 | 1 | 37.33 |
| 2 | 27.6 | 2 | 21.22 | 2 | 27.22 | 2 | 35.48 |
| 3 | 26.95 | 3 | 20.3 | 3 | 24.65 | 3 | 33.77 |
| 4 | 24.66 | 4 | 19.05 | 4 | 24.31 | 4 | 34.35 |
| 5 | 23.7 | 5 | 18.29 | 5 | 22.96 | 5 | 33.59 |
| 6 | 22.21 | 6 | 18.22 | 6 | 22.49 | 6 | 32.33 |
| 7 | 21.68 | 7 | 17.56 | 7 | 22.51 | 7 | 32.27 |
| 8 | 21.58 | 8 | 17.23 | 8 | 22.62 | 8 | 32.45 |
| 9 | 21.18 | 9 | 16.73 | 9 | 22.9 | 9 | 32.56 |
| 10 | 20.62 | 10 | 16.36 | 10 | 23.41 | 10 | 32.7 |
| Mean | 24.429 | Mean | 18.877 | Mean | 24.467 | Mean | 33.683 |

| indexText=index1 (Y,10) | | indexText=index1 (Y,10) | | indexText=index1 (Y,10) | | indexText=index1 (Y,10) | |
|-------------------------|------------------------------|-------------------------|------------------------------|-------------------------|-------------------------------|-------------------------|-------------------------------|
| Test Order | MSE - PHQ fråga 5 som första | Test Order | MSE - GAD fråga 4 som första | Test Order | MSE - HILS fråga 5 som första | Test Order | MSE - SWLS fråga 9 som första |
| 1 | 33.57 | 1 | 24.39 | 1 | 30.84 | 1 | 37.43 |
| 2 | 28.67 | 2 | 22.27 | 2 | 27.6 | 2 | 33.96 |
| 3 | 28.39 | 3 | 20.68 | 3 | 26.1 | 3 | 33.06 |
| 4 | 26.96 | 4 | 19.41 | 4 | 25.74 | 4 | 33.17 |
| 5 | 25.78 | 5 | 18.8 | 5 | 24.5 | 5 | 33.54 |
| 6 | 24.09 | 6 | 18.36 | 6 | 23.69 | 6 | 32.31 |
| 7 | 22.71 | 7 | 17.67 | 7 | 23.54 | 7 | 32.09 |
| 8 | 21.87 | 8 | 17.3 | 8 | 23.96 | 8 | 32.17 |
| 9 | 21.34 | 9 | 16.86 | 9 | 23.39 | 9 | 32.96 |
| 10 | 20.62 | 10 | 16.36 | 10 | 23.41 | 10 | 32.7 |
| Mean | 25.4 | Mean | 19.21 | Mean | 25.277 | Mean | 33.339 |

| indexText=index1 (Y,11) | | indexText=index1 (Y,11) | | indexText=index1 (Y,11) | | indexText=index1 (Y,11) | |
|-------------------------|------------------------------|-------------------------|------------------------------|-------------------------|-------------------------------|-------------------------|-------------------------------|
| Test Order | MSE - PHQ fråga 6 som första | Test Order | MSE - GAD fråga 6 som första | Test Order | MSE - HILS fråga 6 som första | Test Order | MSE - SWLS fråga 6 som första |
| 1 | 34.84 | 1 | 25.37 | 1 | 32.54 | 1 | 37.96 |
| 2 | 29.04 | 2 | 22.68 | 2 | 28.18 | 2 | 35.42 |
| 3 | 27.13 | 3 | 20.34 | 3 | 26.09 | 3 | 33.97 |
| 4 | 25.78 | 4 | 19.35 | 4 | 25.6 | 4 | 32.93 |
| 5 | 24.55 | 5 | 18.55 | 5 | 24.82 | 5 | 33.03 |
| 6 | 23.05 | 6 | 17.54 | 6 | 24.12 | 6 | 32.2 |
| 7 | 21.64 | 7 | 17.2 | 7 | 23.6 | 7 | 32.24 |
| 8 | 21.26 | 8 | 16.69 | 8 | 23.11 | 8 | 31.96 |
| 9 | 20.98 | 9 | 16.55 | 9 | 23.26 | 9 | 32.23 |
| 10 | 20.62 | 10 | 16.36 | 10 | 23.41 | 10 | 32.7 |
| Mean | 24.889 | Mean | 19.063 | Mean | 25.473 | Mean | 33.464 |

Kommentar: Tabellen visar resultaten i Intelligens Survey efter brute-force koden, där alla möjliga frågor som första frågan samt medelvärdet för MSE över respektive ordning. MSE över alla fyra skalorna finns med.

Tabell C2

Sammanställning av medelvärdena av MSE för alla möjliga första frågor för Intelligent survey

| Första Fråga | indexText= | Mean MSE: PHQ | Första Fråga | indexText= | Mean MSE: GAD |
|--------------|----------------|---------------|--------------|----------------|---------------|
| 3 | index1 (Y, 1) | 24.385 | 3 | index1 (Y, 1) | 18.571 |
| 8 | index1 (Y, 2) | 24.815 | 8 | index1 (Y, 2) | 18.909 |
| 1 | index1 (Y, 3) | 23.624 | 1 | index1 (Y, 3) | 18.201 |
| 11 | index1 (Y, 4) | 25.614 | 9 | index1 (Y, 4) | 18.799 |
| 10 | index1 (Y, 5) | 24.976 | 11 | index1 (Y, 5) | 19.026 |
| 9 | index1 (Y, 6) | 24.721 | 2 | index1 (Y, 6) | 18.271 |
| 7 | index1 (Y, 7) | 24.435 | 10 | index1 (Y, 7) | 19.145 |
| 2 | index1 (Y, 8) | 25.764 | 5 | index1 (Y, 8) | 18.448 |
| 4 | index1 (Y, 9) | 24.429 | 7 | index1 (Y, 9) | 18.877 |
| 5 | index1 (Y, 10) | 25.4 | 4 | index1 (Y, 10) | 19.21 |
| 6 | index1 (Y, 11) | 24.889 | 6 | index1 (Y, 11) | 19.063 |

| Första Fråga | indexText= | Mean MSE: HILS | Första Fråg | indexText= | Mean MSE: SWLS |
|--------------|----------------|----------------|-------------|----------------|----------------|
| 11 | index1 (Y, 1) | 24.377 | 11 | index1 (Y, 1) | 32.841 |
| 10 | index1 (Y, 2) | 23.916 | 10 | index1 (Y, 2) | 32.734 |
| 8 | index1 (Y, 3) | 23.663 | 8 | index1 (Y, 3) | 32.935 |
| 3 | index1 (Y, 4) | 24.085 | 3 | index1 (Y, 4) | 32.321 |
| 1 | index1 (Y, 5) | 24.34 | 1 | index1 (Y, 5) | 32.444 |
| 7 | index1 (Y, 6) | 24.746 | 7 | index1 (Y, 6) | 32.581 |
| 2 | index1 (Y, 7) | 24.967 | 2 | index1 (Y, 7) | 33.571 |
| 4 | index1 (Y, 8) | 25.236 | 5 | index1 (Y, 8) | 33.554 |
| 9 | index1 (Y, 9) | 24.467 | 4 | index1 (Y, 9) | 33.683 |
| 5 | index1 (Y, 10) | 25.277 | 9 | index1 (Y, 10) | 33.339 |
| 6 | index1 (Y, 11) | 25.473 | 6 | index1 (Y, 11) | 33.464 |

Kommentar: Tabellen visar genomsnittliga MSE för alla skalorna samt alla möjliga första frågor