



RAPPORT ASSIGNMENT 3

Etik & Könsanalys av datamängd

RICKARD SÖRLIN

Kurs: Artificiell intelligens 1
Kurskod: DVA264
Högskolepoäng: [7.5 hp]
Program: Kandidatprogrammet tillämpad AI

Handledare: Christina Björkman
Examinator: Christina Björkman
Datum: 2023-05-23
E-post: rsn21005@student.mdu.se

INNEHÅLL

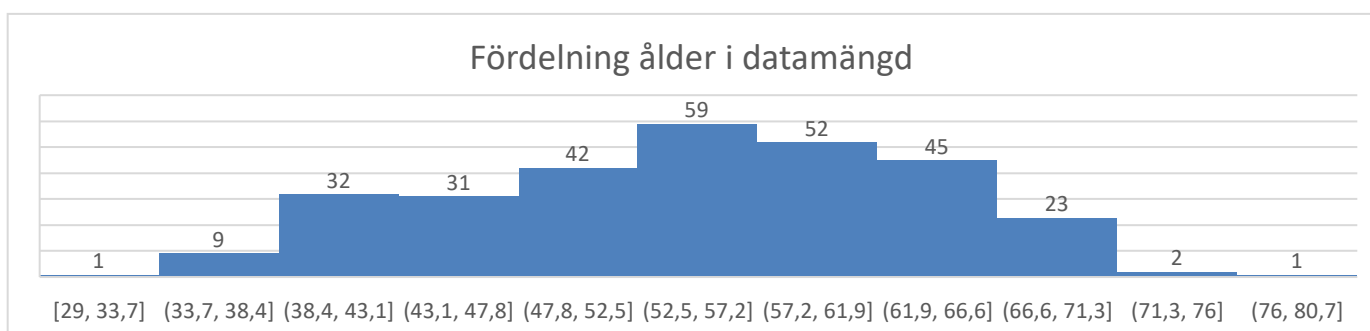
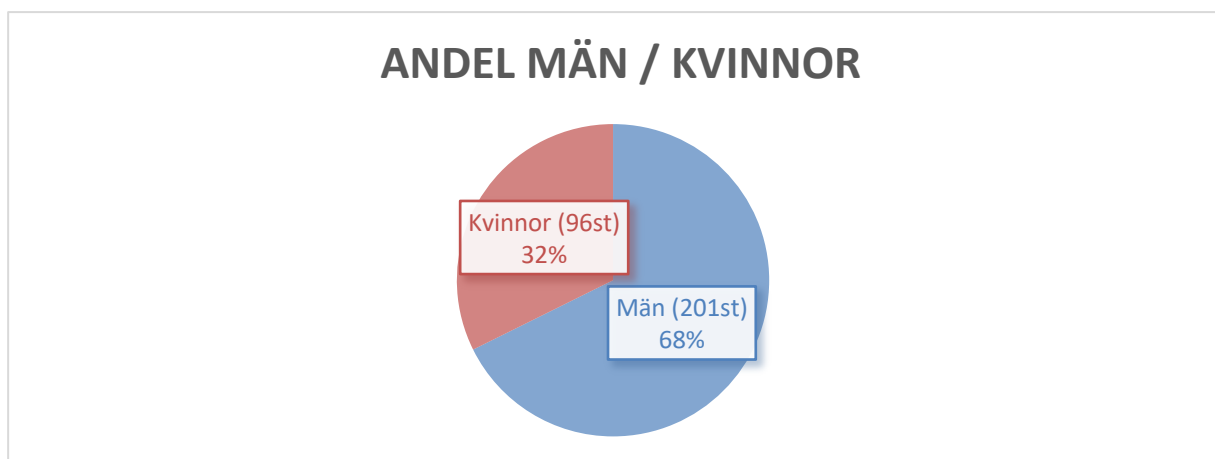
1	DISKUSSION & ANALYS.....	2
1.1	Bias i datamängd	3
1.2	Kön i anslutning till datamängden	4
1.3	Relevans av data för män och kvinnor	4
2	SLUTSATSER	5
	REFERENSER	6

1 DISKUSSION & ANALYS

1.1 Bias i datamängd

Aggregations bias [2, s.115:5] kan tänkas finnas i datasetet bland antalet ifrån datan om antalet förträngningar i större artärer ”ca” som tolkas att inte vara helt relevant för kvinnor enligt kommande analys samt ifrån chest pain type ”CP” men är svårt att tolka utan domänkunskap.

Representation bias [2, s.115:5] och **sampling bias** [2, s.115:6] kan tänkas förekomma då datasetet är obalanserad gällande kön och ålder. Då en större andel representeras av den övre medelåldern och resultera i att modellen generalisera felaktig för den lägre medelåldern för båda könen och bli mer partisk i mot manliga könet och den övre medelåldern.



Omitted variable bias [2, s.115:5] kan tänkas förekomma om till exempel var så att ärftliga faktorer kan tänkas öka risken för dolda hjärtproblem om det finns inom släkt och därmed tänkas resultera i att den tränade modellen generalisera sämre då dess data saknas.

1.2 Kön i anslutning till datamängden

Vid featureselektion var kön en av dom viktiga features för den tränade modellen vid ett av resultaten medan i andra resultatet så var inte kön relevant.

Kön är ett relevant attribut men är obalanserad i datamängden där ungefär 2/3 av datamängden representeras av manliga könet och kan tänkas orsaka att modellen generalisera felaktigt och vara mer partisk i mot manliga könet och visar på representations bias.

Dateringen av datasetet är satt till 1988-07-01 enligt [3] och med hänsyn till det Perez lyfter fram i boken [1, s.254] och återberättar om att den främsta orsaken till dödsfall bland kvinnor sedan 1989 i USA har orsakats av hjärt-kärlsjukdomar då symptom har missats då dom skiljer sig åt mellan män och kvinnor och delvis skulle kunna bero på modeller har tränats på obalanserade data gällande kön.

1.3 Relevans av data för män och kvinnor

Data setet innehåller data om antalet förträngningar i större artärer "**ca**" och tolkas inte vara helt relevant för kvinnor då dom oftast inte får det problemet och inte syns på *angiografen* och därmed möjligtvis orsaka att modellen kan komma att generalisera felaktigt för kvinnliga könet Perez lyfter fram detta i boken [1, s. 256].

Vid "feature selection" lyftes "**ca**" fram att var en av dom viktiga och visar på aggregations biaset som kan tänkas påverka att modellen kan komma att generalisera fel för kvinnliga könet.

Data setet innehåller också data om "**chest pain type (cp)**" där Perez lyfter fram i boken [1, s. 254] vikten av att särskilja på olika typer av bröstsmärtor då speciellt unga kvinnor oftast inte upplever en "Hollywoodinfarkt" som män & fel diagnostiseras.

Men data setet ser ut att särskilja på 4 st olika kategorier av typer av bröstsmärtor bland annat "**typical angina**" och "**atypical**". Om data setet inte skulle särskilja på olika typer av bröstsmärtor så kan det tänkas att modellen skulle generalisera fel och får svårt att särskilja samband mellan könen.

Resultatet av min "feature selection" påvisar inte att typen av bröstsmärta(**cp**) skulle vara en relevant feature i mitt resultat men borde vara det och visar på möjlig aggregations bias.

2 SLUTSATSER

Med hänsyn till min ovanstående analys av datamängden, relevans, bias och mitt resultat vid "feature selection" som också troligt visst saknar flertalet av dom övriga viktiga attributeten för att modellen skall kunna prognostisera korrekt.

Samt det Perez lyfter fram i boken [1] gör mig kritiskt till att dataset kan använda till att träna en och samma modell för att prognostisera hjärt-kärlsjukdomar för både kvinnor och män och generalisera korrekt.

REFERENSER

- [1] C. Criado Perez, *Osynliga kvinnor: Hur brist på data bygger en värld för män, Stockholm, Sweden: ROMANIUS & Selling, 2020.*
- [2] N.Mehrabi, F.Morstatter, N.Saxena, K.Lerman, A.Galstyan, "A Survey on Bias and Fairness in Machine Learning, *ACM Computing Surveys*, vol.54, nr. 6, Article 115, July. 2021
- [3] Hungarian Institute of Cardiology. Budapest: A.Janosi,University Hospital, Zurich, Switzerland: W.Steinbrunn,University ,Hospital, Basel, Switzerland: M.Pfisterer,V.A. Medical Center, Long Beach and Cleveland Clinic Foundation:R.Detrano, " UCI Machine Learning Reposotory," 2019. [Online]. Tillgänglig: <https://archive.ics.uci.edu/ml/datasets/Heart+Disease> (hämtad: 2023-05-23).