



# GÖTEBORGS UNIVERSITET

STUDENT

**0003-RSN**

TENTAMEN

**TIG113 Tentamen**

---

Kurskod	--
Bedömningsform	DT
Starttid	31.10.2023 08:00
Sluttid	31.10.2023 12:00
Bedömningsfrist	--
PDF skapad	17.09.2024 15:29
Skapad av	Lyudmila Vafaeva

---

## **i Information**

### TENTAMEN

#### **Big data och framtidens beslutsfattande, 15hp**

DAG: 31/10 -2023 TID: 08.00 – 12.00

Ansvarig: Faramarz Agahi

Förfrågningar: 031-786 28 22

Resultat: Anslås senast den 21/11 -2023

Betygsgränser: Godkänd 23p. Väl godkänd 33p.  
Maximal poängsumma är 40p.

Hjälpmedel: Inga

Skrivningsfrågor består av 8 frågor om sammanlagt 40 poäng. Varje fråga ger maximalt 5 poäng. Du får maximalt poäng på varje fråga om ditt svar får mig att tro att du kan allt om vad uppgiften handlar om. Detta innebär att du behöver svara relativt omfattande så att jag har chansen att göra denna typ av bedömning.

Lycka till!

# 1 Brus

Diskutera kopplingen mellan följande begrepp inom ramen för dataanalys och bedömning: Bias, Systembrus, Nivåbrus, Mönsterbrus och Situationsbrus. Analysera hur dessa begrepp är sammankopplade och påverkar varandra när du bedömer och utvärderar data. (5p)

**Skriv in ditt svar här**

Brus är variabiliteten i våra judgement och bias är systematiska fel i. Bias hade kunnat visas med exemplet, att en läkare systematiskt överdiagnostiserar en specifik sjukdom, medan ett exempel på brus hade varit att flera läkare kollar på samma sjukdom och de alla ger personen olika diagnoser.

Situationsbrus: Situationsbrus är något som man kan säga beror på yttre påverkan, det kan vara att man är trött, hungrig eller så är det en specifik tid på dygnet.

Nivåbrus: Nivåbrus kan beskrivas som nivåavvikelse, så till exempel om en chef ger mig högre lön vid ett lönesamtal än vad en annan chef hade gjort

Mönsterbrus: Mönsterbrus kan beskrivas som mönsteravvikelse där till exempel en kassapersonal alltid är trevligare mot äldre kunder.

Systembrus är en av de vanligaste typerna av brus och förekommer ofta inom företag osv.

Om vi använder exemplet om domare och straff, så kan vi diskutera många av dessa begrepp i samband med det, först och främst kan vi ha en domare som kanske tenderar att dömma en viss grupp människor strängare än en annan. Just på den dagen när han skulle dömma ett ärende, för den personen som tillhör den gruppen människor han tenderar att ge högre straff till, är han också hungrig. Detta leder då till att han är mer otålig och vill ha sin lunchrast så fort som möjligt och med tanke på hans tidigare val, så lägger han på något år på straffet (situationsbrus). Nu råkar den här domaren också inte tycka om personer som är i 30 års åldern och begår brott och även till de är han taskig (mönsterbrus). För ett brott där rekommenderad högsta straff är 5 år, dömmar denna domare systematiskt brottslingarna till 7 år istället, vilket då är en bias. Till sist har vi då nivåbrus där en annan domare i efterhand också fick titta på straffet som blev högt, och tyckte att det var alldeles för högt straff och skulle aldrig göra det. Alla dessa begrepp brus/bias gjorde att den dömda fick ett väldigt högt straff gentemot vad en annan kanske hade fått, vilket i sig visar på brus. Om vi istället vände på det och det blev positivt, att domaren istället tenderade att ge mildare straff, att han inte var hungrig, att han inte har något emot 30 åringar, så hade utfallet istället blivit tvärtom. Detta i sig visar på systembrus och det faktum att det kan få så många konsekvenser där de andra typerna av brus kan påverka mycket.

Ord: 411

Besvarad.

## 2 Algoritmer

Beskriva de farhågor som tas upp i kurslitteraturen angående användningen av algoritmer i beslutsfattande. Hur kan algoritmer potentiellt förstärka fördomar, och vad är implikationen för demokratin, som framhållits av Cathy O'Neil? Vidare, baserat på forskningsresultat som nämns i Brus-boken, under vilka omständigheter kan algoritmer överträffa mänskliga bedömare i beslutsfattande, och vilka faktorer bidrar till denna framgång? (5p)

**Skriv in ditt svar här**

Negative feedback loops, det faktum att en människa programmerat den. Mekaniska kliniska Det som bland annat är ett stort problem bland modeller och algoritmer är att man utgår från att de inte har någon bias och att de är helt rationella eftersom det inte är en människa som styr den, problemet blir ofta att man glömmer bort att modellen har skapats av en människa och trots att den personen troligtvis gjort sitt bästa i att undvika bias, brus och stereotyper, så kommer det alltid finnas där. Det innebär att inte heller modellen och sättet algoritmerna jobbar på, är helt objektiv utan bias, för den helt enkelt har programmerats på det sättet. Problemet är också att modellen måste byggas och utgå från något, vilket ofta är någon typ av tidigare information/historik. Dessa bedömningar har då också gjorts av människor med alla sina bias, vilket är ännu ett sätt bias tar plats i en modell. I och med detta så finns ju fördomar involverade, vilket leder till att även modellen blir fördomsfull som kan orsaka skada hos människor.

Det som också orsakar problem är också icke transperansen i själva modellerna och algoritmerna, där vi inte egentligen vet hur den beräknar saker och ting utan ser egentligen bara vad som går in och vad som kommer ut. Detta i sig gör att det finns en stor risk för bias och fördomar och väldigt svårt för en att kunna åtgärda det.

När det kommer till demokratin så drabbas ju människor av algoritmerna som är biased vilket gör att det för skada för människor och att de kan hamna i negative feedback looper som göra att det blir sämre och sämre för de som redan blivit utsatta. Detta leder till ett ganska unfair samhälle som gynnar de rika och de fattiga glöms bort, de rika får förmåner osv. De rika har en högre röst och är "viktigare" vilket då också leder till mer inflytande och då är det inte längre hela populationen som ska bestämma något utan endast de med högt inflytande. Algoritmerna kan sikta in sig på vissa grupper som oftast redan är drabbade, allt från fattiga till vissa minoriteter, ålder osv. Då kan det väldigt snabbt också bli farligt, eftersom det inte finns någon gräns, nu handlar det mest om pengar, men man kan mycket lätt i framtiden använda det till något mer farligt eftersom det numera finns så extremt mycket information om oss som individer. Det kan då användas för att förtrycka en och tysta folk och ja det är inte så demokratiskt.

När det gäller brus boken så pratar de om mekaniska och kliniska verktyg. I det exemplet räknas vi människor in i de kliniska och maskiner och t.ex linear regression ingår då i de mekaniska. Det som visades var att det mekaniska överträffade oss människors förmåga i princip alla gånger det gjordes en jämförelse och till och med när det gällde att skapa en modell av individen själv. Oavsett vad har människor oftast mycket mer bias och generella tankar, och det blir lätt att man kanske övertänker något eller så tolkar man informationen på fel sätt. Skillnaden blir att de mekaniska endast utgår från den informationen den har fått och det lämnas egentligen inte mycket utrymme för spekulering, så medan vi människor förvirrar oss själva så går maskinerna rakt på sak. Dessutom har det mekaniska också många gånger tillgång till historik och vi människor har inte den kapaciteten att komma ihåg allt, vilket också kan göra att maskiner har mer information för att komma fram till något.

Ord: 587

### 3 Minska bruset

"Om vårt syfte är att minska bruset står valet mellan två sätt att reglera människors beteenden: regler respektive normer." Diskutera detta påstående genom att definiera regler och normer, ge praktiska exempel på var och en och diskutera vilken strategi som skulle vara mer lämplig för olika typer av organisationer. (5p)

#### Skriv in ditt svar här

Regler är något som är konkret, där återstår inte plats för att tolka vad som menas, antingen följer du regeln eller inte. Ett exempel är att hastighetsbegränsningen är 50 km/tim, kör du bara en kilometer snabbare har du brutit mot regeln, det är något som är bestämt och ingenting som du som individ kan göra något åt. Detta reducerar också väldigt mycket brus just på grund av att det inte lämnas utrymme för att tolkas. Å andra sidan har vi då normer som just är tolkning där varje individ kan tolka det på sitt egna sätt, ett exempel relaterat till det första hade till exempel varit att man ska köra försiktigt. Försiktigt för en förare kan innebära 10 km/tim medan för en annan betyda 50 km/tim, och problemet blir att du inte riktigt kan säga att någon har fel, eftersom vi alla har rätt att tolka det på vårt egna sätt.

De båda reglerar människors beteenden fast till olika grader, regler är något kommer nog påverka mer än vad normer kommer.

Regler är något som nog hade varit lämpligt i många organisationer, det både reducerar brus och man har något konkret att förhålla sig till, men då blir det också direkt ganska strikt och kan göra organisationen stel. Med normer är ju största problemet just att det finns utrymme att tolka som man vill, vilket också leder till mycket brus. I de flesta styrdokument använde de sig mer att normer med lösa definitioner. Om vi pratar inom rättsväsendet där det gäller mer allvarliga saker, så är regler betydligt bättre, för då finns det kanske konsekvenser om du inte följer det. Jag tror att de allra flesta organisationer fungerar bäst med en kombination av båda, där det finns regler för allvarliga delar som mobbning, men också normer som att det hade varit bra att komma till jobbet en kvart innan man börjar. Detta just för att undvika en alldeles för strikt miljö.

Ord: 321

Besvarad.

## 4 Modeller

Farliga modeller, särskilt Weapons of Math Destruction, har tre saker gemensamt. Vilka är dessa? Förklara och ge exempel på ditt svar. (5p).

### Skriv in ditt svar här

Skala: Detta är skalan modellen implementeras på. Vilka modellen påverkar

Opacitet: Kan förklaras som en svart låda, där vi ser vad som kommer in och ut, men vi vet inte riktigt vad som händer i mitten.

Skada: Det faktum att modellen kan gynna en del människor, men också skada en del.

Dessa tre saker gör själva modellen farlig, om vi använder skolrankingen till exempel där en skola blir lägre rankad än vanligt, vilket då leder till att top elever väljer en annan skola och att några kanske väljer att byta, så blir de nästa år ännu sämre rankade vilket gör att ännu färre söker osv. Alltså en negativ feedback loop, som också kan göra en hel del skada, eftersom att det kan vara ganska svårt att ta sig ur den.

Ett exempel är problemet med lärarmodellen som togs upp, som bara mätte hur bra eleverna presterade på provet, vilket gjorde att många omtycka och bra lärare sparkades, eftersom de fick låga poäng. Problemet med modeller blev att även när lärarna bad om att få veta varför den fick så låga poäng, så kunde de inte det eftersom att ingen riktigt vet alla parametrar den använder. Här blir direkt opaciteten/icke transparensen ett problem, för även om lärarna vill förbättra sig, så har de inget konkret att utgå från, vilket då kan vara svårt att jobba utifrån. Detta skapar skada dels för lärarna som blev sparkade, men också till eleverna som blev berövade en bra lärare. Om detta helt plötsligt börjades implementeras på alla skolor, så hade det gett extremt skeva resultat, där dåliga lärare hade kunnat jobba kvar medan bra hade blivit sparkade, det hade kunnat påverka väldigt många. Det som tillsist var poängen är att lärande inte är någonting du kan mäta. Problemet är att modellen inte tar hänsyn till självaste klassen, ibland finns klasser med riktigt smarta och begåvade elever, ibland finns det elever som behöver lite mer hjälp. Och om en lärare ena året hade de smartaste eleverna där nästan alla fick fullpott och sedan året därefter fick elever som hade det svårare och då presterade sämre och om man då jämförde resultaten så hade modellen troligtvis klassat läraren som dålig, trots att lärarens roll inte var särskilt stor där.

Ord: 372

Besvarad.

## 5 Kostnader

Kostnader och fördelar med brusreducering: Förklara vikten av att bedöma om brusdämpande åtgärder är värda sitt pris i detta sammanhang. Ge exempel på kostnader och överväganden som påverkar denna utvärdering. (5p)

**Skriv in ditt svar här**

Det första man nog borde göra är en brusinventering och ta reda på var problemet sitter och hur stort det är. Vi har det exemplet med försäkringsföretaget där de anställda alla fick 3 fall var för att bedöma med tillgång till samma information. Cheferna trodde att man skulle hitta en variation på ca 10%, men brusinventeringen visade på en skillnad på över 50%, vilket var chockande för cheferna som då förstod att de förlorade miljontals kronor på grund av detta. Detta var ju en stor förlust för företaget och i deras fall hade de nog tyckt att det hade varit mycket värt att göra ändringar för att förebygga brus, kanske sätta riktlinjer med intervall för olika typer av fall. Hade det däremot varit ett annat företag som kanske inte tjänar på att göra några förändringar, kanske på grund av att förändringarna hade inneburit en ännu större kostnad, så kanske det hade varit värt att ha kvar det så. Kostnaderna beror mycket på vilket problem som skapar brus, ibland kan det krävas att man får lägga till någon regel och det är nog värt kostnaden, men ibland kanske det kräver en beteendeförändring vilket å andra sidan hade påverkat alla anställda på företaget och hade krävt tid och pengar på att få alla att förstå ett nytt system till exempel, då kanske det inte hade varit värt kostnaden ifall det hade skapat större problem istället.

Kostnader kan vara från saker som förändringar på websidan (kanske lägga till någon ny funktion), ändra till ett nytt system (det gamla kanske är utdaterat och svårförstått) eller så kanske alla lever i enighetensillusionen (man hade kanske behövt diskutera mer och se till att alla är på samma sida). Det finns massa olika kostnader och alla kräver mer eller mindre ändringar och det är upp till företaget själv att bedöma om de är villiga att ta det.

För mina exempel ovan så kanske de har gjort en studie och sett att personer som använder deras websida har svårt att förstå var man ska klicka för att kunna boka ett samtal med dom. Då när företaget hör detta så kanske de tänker att det är värt det, för om man gör det tydligare så kommer de få fler möten och då kan de tjäna mer pengar till exempel. För det nya systemet kanske man är lite mer osäker, man vet att det är en stor kostnad, att det hade tagit mycket tid och tänker att det för tillfället kanske inte är värt det, eftersom de flesta klarar ändå av det för tillfället. För det sista exemplet med att alla tror att de andra tänker lika som de kanske medför stora förluster (t.ex försäkringsbolaget), detta leder till att införa samtal och kanske fler riktlinjer inte alls hade varit en stor kostnad för företaget, relativt lätt att införa och hade troligtvis fort fått en positiv effekt.

Det finns ju också optimalt brus som går ut på att det krävs att hitta en balans, det kanske inte alltid är bra att få bort brus om det istället leder till att företaget inte kan anpassa sig till nya situationer, så det gäller att hitta en balans.

Ord: 524

Besvarad.

## 6 Datalager

Definiera vad ett datalager (Data Warehouse) är och markera de viktigaste skillnaderna som skiljer det från en vanlig databas. Diskutera egenskaperna hos datalager och förklara varför de är avgörande för specifika analytiska uppgifter. (5p)

**Skriv in ditt svar här**

Datalager:

- Består av databaser
- Företag samlar all sin information i datalagret
- Flerdimensionell
- Använder OLAP

Databas:

- Används mer för specifika grejer t.ex transaktioner
- Tvådimensionell
- Använder OLTP

De största skillnaderna blir främst att datalagret är flerdimensionellt och inte databaser, datalager har mer data om själva företaget, medan databasen oftast är fokuserat på ett specifikt område.

Med tanke på att datalager har all information om ett företag så är det ett bra verktyg om man vill analysera data och sortera den. Det kan vara ett praktiskt sätt att ta fram specifik och detaljerad information som vilket vara som köpts mest, eller hur mycket vinst man har gjort det senaste halvåret osv. Där har du tillgång till all nödvändig information som då kan hjälpa dig i många aspekter, så som vilken vara du borde annonsera, vilken vara man borde ta bort från utbudet. Vad är det som drar mest pengar och hur ska vi åtgärda det, alltså inom väldigt många områden som kan hjälpa ett företag förstå sig själva bättre, men också deras kunder. Det handlar främst om att du kan få en bra överblick av själva företaget som är ganska lättöverskådligt. Du kan använda det för att ta reda på det mesta. Man har möjlighet att gå tillbaka och se historiken, inom vilka perioder säljs en specifik vara mest.

Ord: 216

---

Besvarad.



## 7 stjärnschema

Beskriv syftet med ett stjärnschema och förklara varför det används i samband med datalagring. Definiera vad en dimensionstabell är och dess roll i ett stjärnschema. Hur förbättrar denna schemastruktur effektiviteten i dataanalys? (5p)

**Skriv in ditt svar här**

Stjärnschema använder sig av RDMS. Innehåller faktatabeller med den viktigaste informationen, men också nycklar till de olika dimensionerna som till exempel tid. Dimensionerna innehåller också data om själva företaget och datan uppdateras ständigt. Du använder faktatabellens nycklar för att ta dig till dimensionerna, faktatabellens information är också "oföränderlig" till skillnad från dimensionerna.

Det används främst för att kunna få en överblick och kunna se vilka delar som ingår, också i dimensionerna vi kan lägga till ny information och nya punkter vi är intresserade utav.

Datalagret, stjärnschemat och dimensionerna är sammankopplade. Eftersom stjärnschemat är sammankopplat med dimensionerna och de uppdateras så pass ofta så blir det därigenom datalagret också får sin information och lägesrapporter. Stjärnschemat är den som håller ihop dimensionerna och gör det lättöverskådligt att se vilka områden/dimensioner som finns. Effektiviteten förbättras eftersom allt redan är strukturerat och gör det lätt att analysera områdena. Också lätt att lägga till ny information och även nycklar. Nycklarna finns för att se vad vi har valt att fokusera på i vårt datalager och ge oss tillgång till dimensionerna där den nya informationen finns.

Ord: 180

---

Besvarad.

## 8 Relation och samband

Kursen diskuterade tre olika kategorier av dataanalysverktyg, nämligen Deskriptiva, Prediktiva och Preskriptiva analysverktyg, samt två grundläggande typer av relation och samband: Kausalitet och Korrelation. Beskriv och ge exempel på vilka dataanalysverktyg som mest sannolikt genererar kausala relationer och/eller som är mer lämpade för att identifiera korrelationssamband. (5p)

**Skriv in ditt svar här**

**Deskriptiva:** Använder historisk data och svarar på frågan vad som har hänt. Använder data mining och data aggregate

**Prediktiva:** Använder också historisk data och svarar på frågan vad kan komma att ske. Använder prognostisering och statistiska modeller. För att förutspå vad som kan komma att hända

**Preskriptiva:** Använder både de deskriptiva verktygen och prediktiva verktygen och svara på frågan, vad bör vi göra. Använder optimerings och simulerings algoritmer för att visa vad för väg som fungerar bäst för vårt mål.

Kausala: Orsakssamband

Korrelation: Om x ändras så ändras även y.

Prediktiva använder sig av den historiska datan + statistiska modeller + prognostisering för att förutspå vad som kan komma att ske i framtiden. Jag skulle nog säga att den känns mer som något som skulle generera mer kausala relationer eftersom den bara försöker visa på vad som kan komma att hända. Den har kanske kollat på den historiska datan och identifierat att det verkar finnas något koppling och sen utgår från bland annat det för att få fram en prognos, här tror jag det mer utgår från det tidigare för att ge ett förslag på något som kanske hade kunnat hjälpa.

Preskriptiva använder sig av både de deskriptiva & prediktiva verktygen + optimerings och simuleringsalgoritmer. Det känns som att den har en bättre bild av vad som påverkar vad och jag tror därför den har en större chans att generera en korrelation. Den har tillgång till allt som de båda ovan har plus en simuleringsalgoritm som gör att den kan simulera vad som kan hända och vad händer om du väljer att göra det ena, och vad händer om du väljer att göra det andra. Då har den nog identifierat korrelationer för att ta fram det mest optimala lösningen.

Det deskriptiva verktyget använder historiska datan + data mining + data aggregate. Jag tänker att den kanske har en chans att eventuellt kunna hitta ett orsakssamband genom att analysera det som har hänt likt det med de prediktiva verktygen. Den hade då kunnat analysera datan och hitta saker som kanske kan orsaka något.

Ord: 342

Besvarad.