

Metodbilaga för Gömda fel

Mörkertal för felaktiga utbetalningar från
arbetslöshetsförsäkringen



Innehåll

1	Syfte med bilagan och statistiken	4
1.1	Statistikens syfte och mottagare.....	4
2	Undersökningsdesign.....	4
2.1	Variabler och referenstider	4
2.1.1	Modellvariabler.....	4
2.1.2	Referenstider	7
2.2	Populationer och objekt.....	7
2.2.1	Mål- och rampopulationer	7
2.2.2	Avgränsningar.....	7
2.3	Urvalsmetod.....	9
2.4	Insamlingsmetod	9
2.5	Bearbetningar och kontroller	9
2.5.1	Bearbetningar	9
2.5.2	Granskning.....	10
2.6	Prediktiv modellering.....	10
2.6.1	Modellen består av tre delar i kombination via omröstning	10
2.6.2	Modellerna tränas på två kompletterande sätt	10
2.6.3	Modellerna identifierar avvikelser i träningsdata	11
2.6.4	Variabelreduktion minskar antalet dimensioner.....	11
2.6.5	Hantering av obalans sker via översampling	11
2.6.6	Förbehandlingsmetoder	12
2.6.7	Uppskattningen av mörkertalet baseras på förvirringsmatrisen	12
2.6.8	Prestandamått.....	13
2.7	Uppskattningar	13
2.7.1	Standardformel för uppskattning av mörkertalet.....	13
2.7.2	Anmärkningar för formeln för mörkertalet.....	14
2.7.3	Förbättringar av formeln för mörkertalet	14
2.7.4	Vår uppskattning av mörkertalet.....	15
2.7.5	Simulering av gömda återkrav ger underlag för korrigeringsfaktor	16
2.7.6	Sammanvägningar för alla arbetslöshetskassor.....	18
2.7.7	Omvandlingar från mörkertal för tidrapporter till antal dagar och belopp	18
3	Kvalitet	20
3.1	Tillförlitlighet totalt	20
3.2	Osäkerhetskällor	20
3.2.1	Urval.....	20
3.2.2	Ramtäckning	20
3.2.3	Mätning	20
3.2.4	Bortfall	20
3.2.5	Bearbetning.....	20
3.2.6	Modellantaganden	21



3.3	Jämförbarhet.....	21
4	Tillgång till och behandling av personuppgifter	22
4.1	Sekretess.....	22
4.2	Bevarande och gallring	22
4.3	Uppgiftsskyldighet.....	23



1 Syfte med bilagan och statistiken

Detta dokument är en bilaga till rapporten *Gömda fel – Mörkertal för felaktiga utbetalningar från arbetslöshetsförsäkringen*. Metodbilagan beskriver metoddesignen, genomförandet och den statistik relaterad till granskningen som presenteras i rapporten.

Ansvarig för statistiken för granskningen är Robert Granat. En extern konsult har faktagranskat delar av kapitel 2.6–2.7 i metodbilagan.

Kontakta statistik@iaf.se vid frågor.

1.1 Statistikens syfte och mottagare

Statistiken i rapporten beskriver uppskattningar av omfattningen av mörkertalet för de felaktiga utbetalningarna från arbetslöshetsförsäkringen, totalt och per arbetslöshetskassa.

Uppskattningarna beskriver andelen felaktiga tidrapporter som undgår upptäckt och motsvarande uppskattningar av det totala felbeloppet. Syftet är att ta reda på hur stor del av de felaktiga utbetalningarna från arbetslöshetsförsäkringen som inte hittas, och ange en tillförlitlig uppskattning av omfattningen av mörkertalet.

Huvudsakliga mottagare av statistiken är arbetslöshetskassorna och Arbetsmarknadsdepartementet.

2 Undersökningsdesign

I det här kapitlet beskriver vi vilka data som har samlats in i samband med granskningen och vilka statistiska mått vi har använt. Vi beskriver även hur data har samlats in och vilka skattningsmetoder vi har använt.

2.1 Variabler och referenstider

2.1.1 Modellvariabler

Tabell 1–2 redogör för de modellvariabler som ingått i träningen av de prediktiva modellerna i granskningen.



Tabell 1: Modellvariabler: Utbetalningsdata

Variabel	Exempel på innehåll
Ersättningsår – det år tidrapporten avser	2024
Gäller tidrapporten en alldeles ny ersättningsperiod?	Ja/Nej
Tidrapportvecka – det veckonummer tidrapporten avser	202212
Antal ersättningsdagar som utbetalats för den aktuella tidrapporten	5
Antal dagar kvar på ersättningsperioden	87
Utbetalt belopp som motsvarar tidrapporten (i kronor)	1872,50
Medeldagpenning för aktuell utbetalning som hör till tidrapporten (i kronor)	437
Antal deltidsdagar som sökanden deklarerat på tidrapporten	1.0
Antal förändringar i tidrapporten som sökanden gjort under ersättningsperioden	2.0
Antal dagar mellan första och sista förändringen i tidrapporten	5.0
Om någon av de tidrapporter som hör till utbetalningen har korrigerats eller kompletterats efter utbetalningsdatumet	Ja/Nej
Den typ av ersättning som var aktuell för tidrapportveckan	Inkomst/Grund
Har sökanden registrerat någon hindertid under tidrapportveckan?	Ja/Nej
Antal aktivitetsrapporter som sökanden lämnat in från ersättningsperiodens början fram till början av aktuell tidrapportvecka	6
Antal avvikelser och underrättelser som Arbetsförmedlingen registrerat och skickat för aktuell tidrapportvecka	2, 1
Antal avstängningsdagar under den utbetalning som hör till tidrapporten	1
Antal dagar mellan tidrapportens sista dag och dagen för motsvarande utbetalning	32
Ersättningsgrundande dagpenningen under tidrapportveckans början (i kronor)	749,0

All träningsdata genereras först på tidrapporter, sedan ackumuleras data på respektive utbetalning. Resultaten som presenteras i huvudrapporten utgår dock från träning på individuella tidrapporter för högsta möjliga detaljnivå.

Tabell 2: Modellvariabler: Persondata

Variabel	Exempel på innehåll
Sökandens kön	Kvinna/Man
Vilken sökandekategori tillhör sökanden?	11
Sökandens senaste typ av kontakt med Arbetsförmedlingen	Telefon/Besök
Sökandens ålder	53
Sökandens åldersgrupp	25–34, 55–
Vilken kommungrupp enligt SKR:s kommungruppsindelning tillhör sökandens hemkommun?	A1
Sökandens hemkommun (kommunnummer)	180
Sökandens födelseregion	Sverige, Norden, Europa
Sökandens utbildningsnivå	Gymnasial
Sökandens yrkesområden registrerade som unika SSYK-koder i handlingsplanen	1120
Sökandens arbetslöshetskassa	68
Antal återkrav av arbetslöshetsersättning sökanden fått under den senaste ersättningsperioden	4
Antal återkrav av arbetslöshetsersättning sökanden fått under den undersökta tidsperioden	1
Antal inskrivningsperioder sökanden haft vid Arbetsförmedlingen under aktuell tidsperiod	10
Antal kända avstängningar från arbetslöshetsförsäkringen som sökanden fått under aktuell tidsperiod fram till innevarande vecka	2

Vi har övervägt om vi använt mer information än vad som är nödvändigt för granskningens ändamål. Till exempel kan kön, åldersgrupp och födelseregion typiskt sett vara uppgifter som bör hanteras med viss försiktighet. Mot bakgrund av granskningens syfte om att kunna göra en effektiv och ändamålsenlig uppföljning av arbetslöshetsförsäkringen och verka för en likformig och rättvis tillämpning av densamma är dock bedömningen att den mängd information som framgår av listan av variabler är nödvändig. Tidigare undersökningar har även visat att födelseregion och



kön påverkar risken att få återkrav¹. Det är IAF:s bedömning att informationen är nödvändig. Skälet är att vi då undviker konfundering, det vill säga att vissa variablers inverkan på ett resultat göms bakom andra genom dolda korrelationer.² Dessutom förväntar vi oss att de variabler som används i de prediktiva modellerna förbättrar kvaliteten på tillsynen och uppföljningen i väsentlig mån. I den här granskningen tar vi inte heller reda på hur variablerna viktats mot varandra inne i de prediktiva modellerna.

2.1.2 Referenstider

Statistiken avser tidrapporter som hör till utbetalningar från arbetslöshetsförsäkringen under perioden 1 januari 2021–31 december 2023.

2.2 Populationer och objekt

Detta avsnitt redogör för våra mål- och rampopulationer samt de avgränsningar som införts i behandlingen av materialet.

2.2.1 Mål- och rampopulationer

Målpopulationen är samtliga tidrapporter och tillhörande utbetalningar från arbetslöshetsförsäkringen.

Rampopulationen är samtliga tidrapporter som hör till de utbetalningar från den undersökta tidsperioden som finns tillgängliga i IAF:s statistik- och tillsynsdatas.

2.2.2 Avgränsningar

Det datauttag vi har använt i granskningens huvudrapport och i framställandet av den tillhörande statistiken genererades 16 februari 2024, och begränsades till tidrapporter som vid uttaget var minst 200 dagar gamla. Begränsningen infördes för att säkerställa att arbetslöshetskassornas ordinarie kontroller redan fångat upp så många felaktiga utbetalningar som möjligt, vilket i sin tur har stor betydelse för träningen av de prediktiva modellerna. I praktiken innebar det en avgränsning av materialet till utbetalningar från 1 januari 2021–23 juli 2023. Data avgränsades också till att gälla tidrapporter som inte innehöll karensdagar eller tillhörde utbetalningar som korrigerats i efterhand.

¹ IAF 2020:12 Arbetslöshetskassornas återkrav av arbetslöshetsersättning.

² Exempelvis kan variabeln Kön korrelera starkare med Deltid för kvinnor än för män.

Det totala antalet tidrapporter med eventuella återkrav och fördelningen över de olika arbetslöshetskassorna som ingår i granskningen framgår av tabell 3.

Tabell 3: Fördelning av tidrapporter och kända återkrav hos arbetslöshetskassorna.

Arbetslöshetskassa	Antal tidrapporter	Andel tidrapporter i procent	Antal återkrav	Andel återkrav i procent
Akademikernas	1 030 784	8,36	12 603	1,22
Alfa	1 243 555	10,08	25 427	2,04
Byggnadsarbetarnas	470 725	3,82	10 420	2,21
Elektrikernas	38 421	0,31	1 005	2,62
Fastighets	273 829	2,22	7 761	2,83
Finans- och Försäkringsbranschens	67 480	0,55	1 375	2,04
GS	150 259	1,22	3 079	2,05
Handels	750 958	6,09	17 707	2,36
Hotell- och restauranganställdas	862 437	6,99	21 699	2,52
IF Metalls	721 362	5,85	15 675	2,17
Journalisternas	32 972	0,27	600	1,82
Kommunalarbetarnas	2 331 455	18,91	53 667	2,30
Ledarnas	145 953	1,18	2 705	1,85
Livsmedelsarbetarnas	120 293	0,98	2 043	1,70
Lärarnas	227 271	1,84	5 446	2,40
Pappersindustriarbetarnas	26 860	0,22	611	2,27
Sekos	247 561	2,01	5 628	2,27
Småföretagarnas	388 117	3,15	8 230	2,12
STs	116 955	0,95	2 411	2,06
Sveriges arbetares	45 983	0,37	1 711	3,72
Säljarnas	33 700	0,27	994	2,95
Transportarbetarnas	495 857	4,02	15 246	3,07
Unionens	2 231 645	18,10	42 048	1,88
Vision	277 526	2,25	6 671	2,40
Totalt	12 331 958	100,00	264 762	100,00

Källa: IAF:s statistik- och tillsynsdatabas.



2.3 Urvalsmetod

När vi uppskattat det verkliga mörkertalet har vi dragit ett obundet slumpmässigt urval (OSU) ur rampopulationen. Urvalsstorleken för att uppskatta en andel för ett OSU ges i vanliga fall av formeln:

$$n = \frac{n_0}{1 + \frac{n_0}{N}}$$

där $n_0 = z_{\alpha/2}^2 s^2 / e^2$, n är urvalsstorleken, N är populationstotalen, z är ett värde från den standardiserade normalfördelningen, $s^2 = p(1 - p)$ och e är den önskade felmarginalen. Sedan sätts e , p och z , exempelvis till 0,025, 0,5 och 1,96 (för att få ett 95-procentigt konfidensintervall).

I vår granskning gäller att uppskattningen av det verkliga mörkertalet bygger på en urvalsstorlek n och ett antal observationer p som redan är kända på förhand. Därför blir uppgiften att beräkna felmarginalen:

$$e = z_{\alpha/2} \cdot \sqrt{s^2 \frac{N - n}{n \cdot N}}$$

Konfidensintervallet ges då av $p \pm e$.

När det gäller de uppskattningar som beräknas av de prediktiva modellerna har vi inte dragit något urval från rampopulationen. En prediktiv modell tränas för varje arbetslöshetskassas samtliga tidrapporter och resultatet sammanvägs till ett viktat gemensamt genomsnitt. Den delen av granskningen utgör en totalundersökning.

2.4 Insamlingsmetod

Statistiken är uteslutande baserad på data från IAF:s statistik- och tillsynsdatabas.

2.5 Bearbetningar och kontroller

2.5.1 Bearbetningar

Förutom de aggregerade variabler som finns i tabell 1–2 bearbetas inte några data innan de används i de prediktiva modellerna och den tillhörande statistiken.



2.5.2 Granskning

Grunddata har granskats så långt det är möjligt för att inga orimliga värden ska användas i beräkningarna. Den framställda statistiken kontrolleras också så att till exempel olika andelar summeras till 100 procent.

2.6 Prediktiv modellering

I det här avsnittet redogör vi för hur vi arbetat med den prediktiva modellering som ligger till grund för framställningen av statistiken.

2.6.1 Modellen består av tre delar i kombination via omröstning

I vår granskning har vi kombinerat tre prediktiva modeller³ – random forest, adaboost⁴ och neurala nätverk – via ”soft voting”⁵. Den kombinationsmetoden innebär att alla tre modellerna tränas på samma data och sedan röstar om vilken kategori som är mest trolig för vissa data. Omröstningen resulterar även i ett gemensamt beräknat sannolikhetsvärde, utöver klasstillhörighet.

2.6.2 Modellerna tränas på två kompletterande sätt

Vi har tränat modellerna på två olika kompletterande sätt:

- Den *korsvaliderade* modellen innebär 10-faldig korsvalidering⁶ med träning på 80 procent av träningsdata som sedan utvärderas på 20 procent av träningsdata (testdata).
- Den *omtränade* modellen innebär träning på 100 procent av träningsdata och utvärdering på 100 procent av träningsdata.

³ *Introduction to Statistical Learning*, Springer Texts in Statistics, https://doi.org/10.1007/978-3-031-38747-0_1 [hämtat 2024-10-31]

⁴ Kortform för Adaptive Boosting

⁵ *Soft voting* i maskininlärning är en metod där varje modell i en grupp ger sannolikheter för olika klasser, i stället för bara en slutlig klassprediktion. Sannolikheterna vägs samman, ofta genom ett beräknat medelvärde, för att avgöra vilken klass som ska väljas som slutgiltigt resultat.

⁶ *k*-faldig korsvalidering är en metod för att utvärdera hur bra en prediktiv modell fungerar genom att dela upp datamängden i flera delmängder. I stället för att träna modellen på hela datasetet från början, delas datasetet in i *k*-delar (vanliga värden på *k* är 5 eller 10). Modellen tränas sedan på alla delar utom en, som används för att testa modellen. Processen upprepas sedan så att varje del utlämnas en gång och används som testdata. Resultatet ger en mer robust utvärdering, eftersom modellen testas på olika delar av data, vilket minskar risken för överanpassning och ger en bättre bild av modellens generella prestanda när den utsätts för data den inte sett förut.



Första träningsmetoden minskar generellt riskerna för överträning⁷, men kan överskatta antalet gömda fel i träningsdatan. Den andra träningsmetoden ökar risken för överträning, men kan i stället underskatta antalet gömda fel i träningsdatan, särskilt om modellerna inte är bra på att upptäcka avvikelser.

I den här granskningen har vi använt båda modellvarianterna, för att ta fram ett intervall för det uppskattade mörkertalet. Se avsnitt 2.7.4.

2.6.3 Modellerna identifierar avvikelser i träningsdata

All träningsdata som används av prediktiva modeller innehåller avvikelser eller gömda fel. Så är också fallet i vår granskning, eftersom samtliga felaktiga utbetalningar inte hittas. Vi tränar därför våra prediktiva modeller för att hitta avvikelser i själva träningsdatan⁸.

Efter träningen appliceras modellvarianterna på testdata respektive träningsdata, och de avvikelser som identifieras blir sedan underlag för uppskattningar av mörkertalet och manuell kontroll. De identifierade avvikelserna kan sorteras på risknivå inför den manuella kontrollen, men det gör vi inte i den här granskningen.

2.6.4 Variabelreduktion minskar antalet dimensioner

Innan de prediktiva modellerna tränas använder vi oss av variabelreduktion genom principalkomponentanalys (PCA). Det är en statistisk metod som minskar antalet dimensioner i data med bibehållen önskad nivå på den statistiska variansen – i vårt fall 99 procent – och gör de kvarvarande dimensionerna (principalkomponenterna) oberoende av varandra (ortogonala). Syftet med variabelreduktionen är att rensa bort brus ur träningsdatan och göra det lättare för de prediktiva modellerna att separera de olika klasserna av data.

2.6.5 Hantering av obalans sker via översampling

Data som modelleringen bygger på är obalanserat, det vill säga att proportionerna mellan de ingående klasserna är ungefär 2:100. Det innebär att det går två tidrapportveckor med återkrav på varje hundratal utan återkrav. Sådana obalanser kan skapa problem när man ska träna modellerna och påverka prestandan negativt.

⁷ Överträning (*eng.* overfitting) är ett fenomen där den prediktiva modellen presterar mycket bra på träningsdata, men betydligt sämre på data den inte sett tidigare. Förklaringen är att modellen enbart "memorerat" träningsdata, men inte lärt sig att göra några direkta förutsägelser.

⁸ *Eng.* outlier detection.



Vi har använt oss av en teknik för s.k. översampling⁹ som kallas SMOTE under träningen av de prediktiva modellerna¹⁰. SMOTE skapar fiktiva datapunkter längs hypotetiska ”räta linjer” mellan datapunkter i minoritetsklassen, så att de ser ut att tillhöra samma klass (i vårt fall tidrapporter med återkrav). Syftet är att ge modellerna ett mer balanserat material att arbeta med.

2.6.6 Förbehandlingsmetoder

Första steget i den prediktiva modellen är att data förbehandlas, så att de omvandlas till numeriska värden och skalas ner till intervallet $[0, 1]$.

2.6.7 Uppskattningen av mörkertalet baseras på förvirringsmatrisen

Våra uppskattningar av mörkertalet av felaktiga utbetalningar baseras på antalet hittade avvikelser. En förvirringsmatris för de prediktiva modellerna hjälper oss att se hur ofta och på vilket sätt de prediktiva modellerna gör fel:

$$F = \begin{bmatrix} TP & FN \\ FP & TN \end{bmatrix}.$$

Förvirringsmatrisens diagonalelement TP, TN visar antalet tidrapporter som de prediktiva modellerna kategoriserat rätt under träningen, medan FP, FN visar antalet tidrapporter som modellerna kategoriserat fel.

TP, FP, FN, TN är standardnotation¹¹ för förvirringsmatrisen för binär klassificering och utläses ”true positives”, ”false positives”, ”false negatives” och ”true negatives”. ”Positive” och ”negative” motsvarar träff eller inte träff för vad vi letar efter, i vårt fall tidrapporter med känt återkrav respektive utan återkrav.

Vår uppskattning av mörkertalet nedan kommer att utgå från följande antaganden:

- De tidrapporter som hamnar i FN avspeglar främst brister i den prediktiva modellens förmåga att korrekt identifiera existerande återkrav i träningsdata.
- De tidrapporter som hamnar i FP avspeglar främst misstänkta felaktiga utbetalningar som (ännu) inte gett upphov till återkrav.

⁹ Eng. översampling.

¹⁰ Prediktiva modeller som tränats med SMOTE brukar fungera klart sämre för förutsägelser på okända data.

¹¹ Det finns olika konventioner när det gäller betydelsen av förvirringsmatrisens rader respektive kolumner, men vi följer den som är standard i programbiblioteket [scikit-learn](https://scikit-learn.org/): rader betecknar verklig klass och kolumnerna förutsagd klass.



Exempel 1. Följande numeriska förvirringsmatris för totalt 100 datapunkter

$$\begin{bmatrix} 3 & 1 \\ 2 & 94 \end{bmatrix}$$

betyder att modellerna identifierat 3 kända positiva och 94 negativa korrekt, men kategoriserat 1 positiv och 2 negativa inkorrekt.

2.6.8 Prestandamått

I tabell 4 visar vi tre prestandamått för binär klassificering som bygger på förvirringsmatrisen ovan. För alla prestandamåtten gäller att de ligger mellan 0 och 1 och ju högre värde, desto bättre är klassificeringen.

Tabell 4: Prestandamått för klassificering.

Mått	Formel	Förklaring
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$	Mäter den övergripande prestandan i att klassificera data rätt oavsett klass.
Sensitivity (Recall) "True Positive rate"	$TPr = \frac{TP}{TP + FN} = 1 - \frac{FN}{TP + FN}$	Mäter hur väl modellen kan komma ihåg den ursprungliga klassen för positiva data.
Specificity "True Negative rate"	$TNr = \frac{TN}{TN + FP} = 1 - \frac{FP}{TN + FP}$	Mäter hur väl modellen kan komma ihåg den ursprungliga klassen för negativa data.

Exempel 1 (forts.) Den numeriska förvirringsmatrisen ger värden på måtten i tabell 4 som 97, 75 respektive 98 procent.

Notera att ett utfall där förvirringsmatrisens diagonalelement inte är dominerande antyder att den prediktiva modellen inte presterar bra. Då blir eventuella följdberäkningar mindre tillförlitliga.

2.7 Uppskattningar

2.7.1 Standardformel för uppskattning av mörkertalet

Definition. Vår uppskattning av mörkertalet M för den positiva klassen definierar vi med utgångspunkt i förvirringsmatrisens fyra element som:

$$M = \frac{FP}{TN + FP} \times \left(1 + \frac{FN}{TP + FN} \right)$$

Första faktorn från vänster återspeglar andelen negativa data som de prediktiva modellerna identifierat som positiva. Den andra faktorn kompenserar för att de prediktiva modellerna kan brista i identifieringen av en del av den ursprungliga positiva klassen. Det senare återspeglas i att FN blir större än noll och kan också innebära att motsvarande del av FP inte heller identifieras korrekt.

Prestandamåtten i tabell 4 ger följande samband för de två ingående faktorerna:

$$\frac{FP}{TN + FP} = \frac{FP + TN - TN}{TN + FP} = 1 - \frac{TN}{TN + FP} = 1 - TNr$$

$$1 + \frac{FN}{TP + FN} = 1 + 1 - \frac{TP}{TP + FN} = 2 - TPr$$

vilket medför att vi kan skriva om formeln som:

$$M = (1 - TNr) \times (2 - TPr).$$

Uppskattningen kan således beräknas som en produkt av två faktorer som båda beror på hur väl modellen lyckats separera den positiva klassen från den negativa och vice versa. Högre prestanda i klassificeringen ger lägre mörkertal, lägre prestanda ger högre mörkertal. Tillförlitligheten i uppskattningen kan dock bero på vilken typ av modell som används (se nedan).

Exempel 1 (forts.) För den numeriska förvirringsmatrisen ovan blir första faktorn 0,02 och den andra blir 1,25. Det ger en uppskattning av mörkertalet på 2,6 procent.

2.7.2 Anmärkningar för formeln för mörkertalet

Följande anmärkningar för formeln för mörkertalet är viktiga:

- Formeln kan generaliseras direkt till multiklass-fallet genom att vi använder motsvarande definitioner av *Specificitet* och *Sensitivitet*.
- Formeln är applikations- och modelloberoende i bemärkelsen att den kan användas för andra tillämpningar där ett mörkertal ska beräknas, under förutsättning att den prediktiva modell som används är effektiv för detektering av avvikelser.

Den sistnämnda anmärkningen medför att anledningen till ett eventuellt dåligt utfall vid ett riskbaserat urval främst bör sökas hos den prediktiva modell som används, och inte främst hos den här beräkningsmetoden.

2.7.3 Förbättringar av formeln för mörkertalet

Å ena sidan kompenserar den andra faktorn i standardformeln för brister i identifieringen av den positiva klassen, men formeln innehåller ingen motsvarande kompensation för eventuella brister i hur modellen presterar i att identifiera dolda positiva dataelement. Eftersom det dolda antalet positiva av uppenbara skäl är okänt



på förhand, kan standardformeln behöva skalas upp med en faktor som tas fram empiriskt.

Å andra sidan tar formeln inte hänsyn till att fördelningen av träningsdata i förvirringsmatrisen kan ske med olika sannolikhet för varje dataelement. De dataelement som hamnar i exempelvis *FP* och *FN* hamnar inte där med samma sannolikhet, och mörkertalet kan därför behöva skalas ner med en faktor som avspeglar sannolikheterna i fördelningen.

En förbättring som går i linje med det resonemanget är att lägga till en gemensam korrektionsfaktor α :

$$M = \alpha \times (1 - TNr) \times (2 - TPr).$$

som tas fram empiriskt, till exempel:

- *direkt* genom att beräkna ett genomsnitt av sannolikheten för prediktionen för varje dataelement
- *indirekt* genom någon sorts kalibreringsinformation från den prediktiva modellen som visar hur överoptimistisk eller underpessimistisk den är
- *kombinerade direkta och indirekta metoder* där hänsyn till exempel tas både till sannolikheten för varje prediktion och annan kalibreringsinformation.

α skulle också kunna tas fram via expertutlåtanden, till exempel genom manuell granskning av en representativ del av träningsdatan.

2.7.4 Vår uppskattning av mörkertalet

För vår granskning har vi uppskattat mörkertalet genom att använda standardformeln med en α -korrektionsfaktor, se avsnitt 2.7.3. Korrektionsfaktorn i den här granskningen bestäms genom att

- simulera gömda fel bland de ingående tidrapporterna, vilket ger en mer pessimistisk uppskattning
- multiplicera den beräknade korrektionsfaktorn med ett viktat genomsnitt av sannolikheterna i fördelningen av felklassificeringar mellan *FP* och *FN*, vilket ger en mer optimistisk uppskattning.

Vårt slutliga intervall för uppskattningen av mörkertalet i den här granskningen utgörs av

- övre gränsen för den pessimistiska uppskattningen baserat på den korsvaliderade prediktiva modellen applicerat på 20 procent av träningsdatan
- nedre gränsen för den optimistiska uppskattningen baserat både på den korsvaliderade och den omtränade modellen applicerat på 100 procent av testdatan.

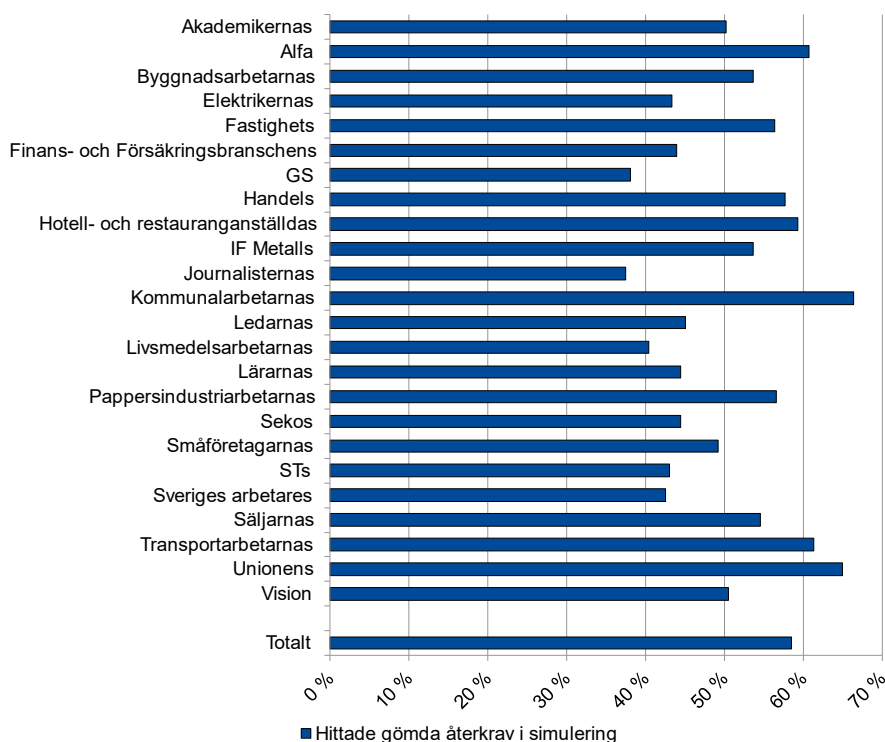
För den nedre gränsen utgör förvirringsmatrisens icke-diagonal-element de data som minst en av modellerna klassificerar fel¹². Det verkliga mörkertalet bör ligga inom det intervall vi beräknat med metoden.

2.7.5 Simulering av gömda återkrav ger underlag för korrigeringsfaktorn

Simuleringen genomförde vi genom att gömma undan 20 procent av de kända felen (återkraven) i träningsdatan för varje arbetslöshetskassa. Det innebar en försämring av träningsunderlaget för de prediktiva modellerna som motsvarar testdelen avsedd för den korsvaliderade prediktiva modellen (se avsnitt 2.6.2), men kan mildras i viss mån av översamplingstekniken vi använder (se avsnitt 2.6.5).

Efter försämringen av dataunderlaget tränade vi båda de prediktiva modellvarianterna. Resultatet visade att modellerna tillsammans återfann i genomsnitt 58 procent av de undångömda felen i simuleringen. Skillnaderna mellan de enskilda arbetslöshetskassorna var dock stor, se diagram 1 och tabell 5.

Diagram 1: Andel av 20 procent av simulering av gömda fel från 2021–2023 som upphittats med prediktiv modellering.



Källa: Prediktiv modellering med data från IAF:s statistik- och tillsynsdatas.

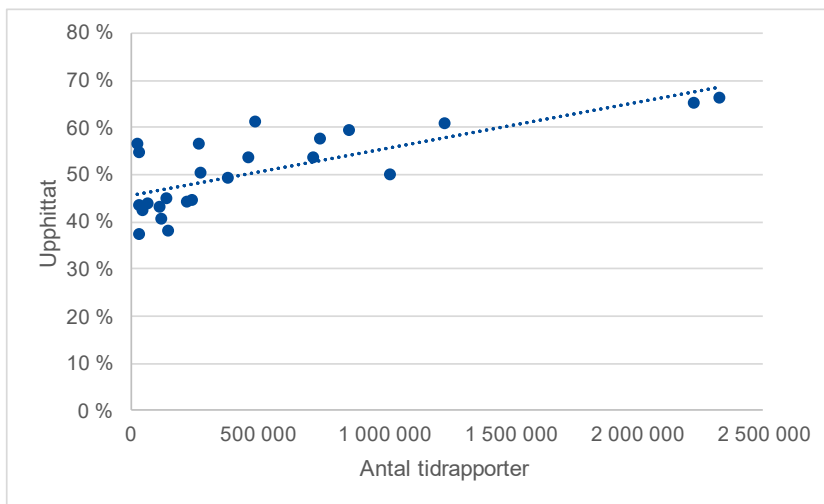
En uppföljande regressionsanalys visar att det fanns ett tydligt samband mellan antalet tidrapporter och mängden återfunna fel i simuleringen, se diagram 2. Det

¹² Den nedre gränsen kan göras mer optimistisk genom att man enbart tar hänsyn till den omtränade modellen vid beräkningen av förvirringsmatrisen för den optimistiska uppskattningen.



visar att träffsäkerheten hos vår metod för att leta felaktiga tidrapporter ökar med ökande datamängd.

Diagram 2: Regressionsanalys av resultat av simulering av 20 procent gömda fel från 2021–23 som upphittats med prediktiv modellering.



Källa: Prediktiv modellering med data från IAF:s statistik- och tillsynsdatabas.

Tabell 5: Återfunna återkrav i simuleringen med 20 procent gömda fel och resulterande korrektionsfaktor.

Arbetslöshetskassa	Andel upphittade återkrav i procent	Korrektionsfaktor
Akademikernas	50,2	1,99
Alfa	60,7	1,65
Byggnadsarbetarnas	53,7	1,86
Elektrikernas	43,3	2,31
Fastighets	56,4	1,77
Finans- och Försäkringsbranschens	43,9	2,28
GS	38,1	2,63
Handels	57,7	1,73
Hotell- och restauranganställdas	59,3	1,69
IF Metalls	53,7	1,86
Journalisternas	37,5	2,67
Kommunalarbetarnas	66,4	1,51
Ledarnas	45,0	2,22
Livsmedelsarbetarnas	40,5	2,47

Arbetslöshetskassa	Andel upphittade återkrav i procent	Korrektionsfaktor
Lärarnas	44,4	2,25
Pappersindustriarbetarnas	56,6	1,77
Sekos	44,4	2,25
Småföretagarnas	49,2	2,03
STs	43,0	2,32
Sveriges arbetares	42,6	2,35
Säljarnas	54,5	1,83
Transportarbetarnas	61,3	1,63
Unionens	65,0	1,54
Vision	50,5	1,98
Viktat genomsnitt	58,5	1,71

Källa: IAF:s statistik- och tillsynsdatas.

2.7.6 Sammanvägningar för alla arbetslöshetskassor

Den gemensamma uppskattningen av mörkertalet för alla arbetslöshetskassorna beräknas som ett viktat genomsnitt av uppskattningarna av mörkertalen för samtliga arbetslöshetskassor, viktade med respektive arbetslöshetskassas andel av de tidrapporter som ingår i rampopulationen.

2.7.7 Omvandlingar från mörkertal för tidrapporter till antal dagar och belopp

För att uppskatta mörkertalet för antalet felaktiga ersättningsdagar och andelen felaktigt utbetald arbetslöshetsersättning utifrån uppskattningarna av andelen tidrapporter med gömda fel har vi använt oss av uppgifter för genomsnittlig dagpenning och genomsnittligt antal dagar som återkrävs för 2023. Vi redovisar de uppgifterna i tabell 6.

Tabell 6: Beräknade värden för uppräknig av mörkertalet.

Genomsnittlig dagpenning	Genomsnittlig andel av en tidrapportvecka som återkrävs
903 kr	71,6 %

Källa: IAF:s statistik- och tillsynsdatas



2.7.7.1 Omvandling av uppskattningarna baserat på de prediktiva modellerna

Den här uppskattningen utförde vi genom att iterativt skapa slumpmässiga utvalda andelar av träningsdata för modellerna vars storlekar låg inom det uppskattade mörkertalsintervallet. Andelarna summerades, och uppskattningarna av belopp och dagar beräknades sedan med hjälp av värdena i tabell 6.

2.7.7.2 Omvandling av uppskattningarna baserat på de manuella granskningarna

Antalet felaktiga ersättningsdagar uppskattades genom antalet dagar för tidrapporter med bekräftade fel samt en tredjedel av tidrapporterna med misstänkta fel. Det multiplicerades med genomsnittlig andel återkrävd tidrapportvecka och genomsnittlig dagpenning, som framgår av tabell 6.



3 Kvalitet

3.1 Tillförlitlighet totalt

Den totala tillförlitligheten påverkas av ett antal osäkerhetskällor. Den främsta osäkerhetskällan är kvaliteten på det underliggande datasetet. Träningsdata för prediktiva modeller behöver generellt uppdateras kontinuerligt för att de ska skyddas från självförstärkande skevheter, exempelvis genom slumpvisa stickprovskontroller av målpopulationen.

Därefter är det den prediktiva modellen som är den största osäkerhetskällan.

3.2 Osäkerhetskällor

3.2.1 Urval

För de prediktiva modellerna är undersökningen en totalundersökning för den avgränsade tidsperioden 2021-01-01 till och med 2023-07-23. I övrigt har inget urval tillämpats.

För de kontroller arbetslöshetskassorna utfört inom ramen för granskningen har vi tillämpat ett obundet slumpmässigt urval.

3.2.2 Ramtäckning

Inga brister i ramtäckningen är kända för det underliggande datasetet.

3.2.3 Mätning

Inga mätningar görs för det underliggande datasetet.

3.2.4 Bortfall

Datasetet som använts i undersökningen har inget känt bortfall.

3.2.5 Bearbetning

Datasetet som använts i undersökningen har bearbetats i SQL samt i de förberedande delar av programmen som använder sig av de prediktiva modellerna. Inga data har förändrats medvetet, men ibland överförs till en annan typ, till exempel har olika konverteringar mellan numeriska och kategoriska variabler genomförts vid behov.



3.2.6 Modellantaganden

Kvaliteten på den prediktiva modellen som används är en stor osäkerhetsskälla. Den prediktiva modellen ska både ha förmågan att tränas så korrekt som möjligt på träningsdata av bristande kvalitet med ett visst etikettbrus och samtidigt passa för detektering av avvikelser. Vi har prövat flera olika prediktiva modeller och använde till slut den variant som gav det bästa resultatet på träningsdata. Vi ansåg att den gav rimliga uppskattningar av antalet avvikelser.

Prediktiva modeller kan bara tränas på det data som är tillgängligt. I vårt fall innebär det att vi antar att arbetslöshetskassorna hittat de flesta typer av fel som finns. Om det skulle finnas andra typer av fel som inte är kända, kan de inte heller hittas av de prediktiva modellerna, eftersom det inte finns några träningsdata som är tillgängliga för de felen.

Träffsäkerheten kommer också att bero på mängden tillgängligt träningsdata. Modellerna kan prestera sämre för arbetslöshetskassor med relativt färre ärenden, men prestera bättre för arbetslöshetskassor med relativt många ärenden, där det finns fler återkrav eller där återkraven följer tydliga mönster.

3.3 Jämförbarhet

Den producerade statistiken kan jämföras med tidigare undersökningar av mörkertalet för felaktiga utbetalningar som IAF gjort med andra metoder¹³. Det ger en god grund för uttalanden om ifall den producerade statistiken är rimlig eller inte.

¹³ IAF 2021:13, Felaktiga utbetalningar i arbetslöshetsförsäkringen 2021, IAF 2024:3, Upptäckta felaktiga utbetalningar från arbetslöshetskassorna 2021–2023 och IAF 2023:3, Upptäckta felaktiga utbetalningar från arbetslöshetskassorna 2020–2022.



4 Tillgång till och behandling av personuppgifter

IAF är en statlig myndighet med uppdrag att ha tillsyn över och följa upp arbetslöshetsförsäkringen. För att kunna utföra uppdraget måste IAF behandla en mängd olika personuppgifter.

IAF:s behandling av personuppgifter sker med stöd av EU:s dataskyddsförordning (EU 2016/679), dataskyddslagen (2018:218), lagen (2006:469) om behandling av personuppgifter vid Inspektionen för arbetslöshetsförsäkringen samt förordningen (2008:936) om behandling av personuppgifter vid Inspektionen för arbetslöshetsförsäkringen.

För bland annat tillsyn, uppföljning och framtagande av statistik har IAF en tillsyns- och statistikdatabas där personuppgifter får behandlas. I den här granskningen behandlas uppgifter från den databasen. IAF är personuppgiftsansvarig för de behandlingar som utförs vid myndigheten.

4.1 Sekretess

Uppgifter som IAF samlar in för sin tillsyn omfattas av samma sekretess som uppgifterna omfattades av hos uppgiftslämnaren, när de kommer in till IAF. Det följer av 11 kap 1 § offentlighets- och sekretesslagen (2009:400) (OSL).

I 28 kap OSL finns bestämmelser om sekretess hos Arbetsförmedlingen (11 §) och hos arbetslöshetskassorna (13 §).

4.2 Bevarande och gallring

Reglerna för gallring av uppgifter som IAF behandlar för sin tillsyn framgår av 15 § lagen (2006:469) om behandling av personuppgifter vid Inspektionen för arbetslöshetsförsäkringen.

Personuppgifter som behandlas för tillsyn ska gallras två år efter utgången av det år när behandlingen av uppgifterna har avslutats i ett ärende hos den handläggande myndigheten. Om personuppgifter enligt första stycket behövs för statistiska eller vetenskapliga ändamål ska de gallras tio år efter det att uppgifterna avskildes för de ändamålen.



4.3 Uppgiftsskyldighet

Arbetsförmedlingen är enligt 48 g § i lagen (1997:238) om arbetslöshetsförsäkring skyldig att lämna de uppgifter som har betydelse för IAF:s tillsyn över ärenden om arbetslöshetsersättning.

En arbetslöshetskassa ska enligt 90 § lagen (1997:239) om arbetslöshetskassor lämna uppgifter till IAF för tillsyn och uppföljning av arbetslöshetsförsäkringen. Vilka uppgifter som omfattas av den skyldigheten specificeras ytterligare i 5–7 §§ i förordningen (1997:836) om arbetslöshetskassor.

