

NOTAT

Beskrivelse af profilafklaringsværktøjet til dagpengemodtagere

16. januar 2020

J.nr.

VOA

CSNI

Baggrund

Det statistiske profilafklaringsværktøj udgør en frivillig mulighed for nyledige dagpengemodtagere. Profilafklaringsværktøjet skal ses i forlængelse af andre værktøjer, som stilles til rådighed for borgeren, den kommunale administration og a-kassernes indsats for at hjælpe den ledige tilbage på arbejdsmarkedet.

Profilafklaringsværktøjet blev implementeret som led i aftalen om beskæftigelsesreformen fra 2014, og har været i drift siden slutningen af 2015. I efteråret 2017 blev der foretaget en række forbedringer af værktøjet, som bestod af hhv. et forbedrende spørgeskema til borgeren samt en videreudvikling af den bagvedliggende statistisk model. Den 13. december 2019 er den seneste opdatering blevet idriftsat. Opdateringen består i en gen-kalibreringen af modellen på nyere data.

Formålet med profilafklaringsværktøjet er, at hjælpe den ledige til at forberede sig bedst muligt på et aktivt og jobrettet kontaktførelse, og at etablere et fælles udgangspunkt for forløbet mellem borgeren og sagsbehandleren ved de efterfølgende jobsamtaler.

Derudover understøtter profilafklaringsværktøjet sagsbehandlerens tilrettelæggelse af en indsats, der er tilpasset den enkelte lediges konkrete behov. Værktøjet understøtter således sagsbehandlerens arbejde, når de vurderer den enkelte borgers muligheder for at komme i job og tilrettelægger indsatsen ud fra den enkeltes forudsætninger. Dog således, at valg af tilbud eller jobstøtteaktiviteter altid baserer sig på sagsbehandlerens samlede faglige og individuelle vurdering.

Værktøjet understøtter sagsbehandleren ved systematisk at indsamle, behandle og præsentere en række udvalgte informationer, som har betydning for den enkeltes jobmuligheder allerede før den første samtale i jobcenteret og a-kassen. Sagsbehandleren har hermed mulighed for at bruge informationen både i sin forberedelse til den første samtale. Og som dialogværktøj i selve samtalen.

Hvad er profilafklaringsværktøjet?

Profilafklaringsværktøjet, som nyledig dagpengemodtagere møder på Jobnet.dk i begyndelsen af deres ledighedsforløb, består af to dele:

1. Et spørgeskema (forberedelsesskemaet) hvor den ledige skal besvare 10 spørgsmål om sig selv, sin jobsøgning og uddannelsesbaggrund mv. og
2. En statistisk baseret vurdering af borgerens risiko for at være ledig i mere end 6 måneder (herefter ”langtidsledig”).

Når en nyledig har udfyldt forberedelsesskemaet modtager borgeren en kvitterings-tekst, hvoraf det fremgår, om der er risiko for, at borgeren kan opleve en lidt længere vej til nyt job, og dermed tilhører risikogruppen. Hertil kommer værktøjet med et forslag til hvad borgeren kan gøre allerede før vedkommende kommer til den første samtale.

Den nyledige har mulighed for at printe svaret og kan desuden genfinde resultatet på 'Min side' på Jobnet.dk. Forberedelsesskemaet skal ses som en optakt til de første samtaler, hvorfor muligheden for at udfylde spørgeskemaet forsvinder, hvis den ledige ikke har udfyldt det inden for de første tre måneder. Sagsbehandlere i jobcentre og a-kasser kan se resultatet af profilafklaringen og borgerens svar på forberedelsesskemaet.

Kun borgere, der udfylder det frivillige forberedelsesskema bliver profileret.

Spørgsmål i forberedelsesskemaet

Når en nyledig dagpengemodtager melder sig ledig på Jobnet.dk, får vedkommende mulighed for at udfylde forberedelsesskemaet. Spørgsmålene i forberedelsesskemaet er udvalgt og formuleret på baggrund af viden fra både forskning, STARs egne analyser og konkrete erfaringer fra kommuner og a-kasser. Spørgsmålene er udviklet i samarbejde med kommuner og a-kasser. (Se tabel 1.)

Tabel 1. Forberedelsesskemaet på jobnet.dk

	<i>Spørgsmål</i>	<i>Svarmuligheder</i>
1	Hvad er din højeste gennemførte uddannelse?	<ul style="list-style-type: none"> • 18 uddannelseskategorier
1.a	Inden for hvilken fagretning er du uddannet? (tag udgangspunkt i din højst gennemførte uddannelse)	<ul style="list-style-type: none"> • Pædagogiske uddannelser • Humanistiske, religiøse og æstetiske uddannelser • Samfunds-, kontor- og handelsuddannelser • Naturvidenskabelige uddannelser • Industri- og håndværkstekniske uddannelser • Jordbrugs, fiskeri-, og levnedsmiddeluddannelser • Transport- og kommunikationstekniske uddannelser • Sundhedsuddannelser • Uddannelser vedrørende offentlig sikkerhed mv.
1.b	Inden for hvilken fagretning er du uddannet? (tag udgangspunkt i din højst gennemførte uddannelse)	<ul style="list-style-type: none"> • Ikke erhvervsrettede uddannelser (fx HF, STX) • Samfunds-, kontor- og handelsuddannelser (fx HHX) • Industri- og håndværkstekniske uddannelser (fx HTX)
2	Hvor hurtigt tror du, at du får et arbejde (sæt kun ét kryds):	<ul style="list-style-type: none"> • Jeg har fået nyt arbejde, men er ikke startet endnu • Inden for 1 måned • Inden for 3 måneder • Inden for 6 måneder • Der går mere end 6 måneder • Jeg forventer snart at gå på barsel • Jeg forventer snart at gå på efterløn eller pension • Ved ikke
3	Er du blevet opsagt, eller har du selv sagt op, fra dit tidligere arbejde?	<ul style="list-style-type: none"> • Jeg er blevet opsagt • Jeg har selv sagt op • Jeg har indtil nu været selvstændigt erhvervsdrivende • Jeg har ikke haft et job (Fx hvis du har modtaget syge- eller barseldagpenge mv., eller har afbrudt din uddannelse)

		<ul style="list-style-type: none"> • Jeg er dimittend • Jeg er midlertidigt hjemsendt • Andet
4	Hvor længe har du vidst, at du ville blive ledig?	<ul style="list-style-type: none"> • Jeg har vidst det i mindre end en uge • Jeg har vidst det den seneste måned • Jeg har vidst det de seneste tre måneder • Jeg har vidst det de seneste seks måneder • Jeg har vidst det i over seks måneder
5	Hvor mange job har du søgt, fra du vidste, at du blev ledig til nu?	<ul style="list-style-type: none"> • Jeg har ikke søgt job • Jeg har søgt 1 job • Jeg har søgt 2-5 job • Jeg har søgt over 5 job
6	Hvordan vurderer du dine jobmuligheder inden for de områder, hvor du søger arbejde?	<ul style="list-style-type: none"> • Mine jobmuligheder er gode • Mine jobmuligheder er nogenlunde gode • Mine jobmuligheder er ikke så gode • Mine jobmuligheder er dårlige • Ved ikke
7	Ved du hvilket arbejde, du gerne vil have?	<ul style="list-style-type: none"> • Jeg ved lige hvilket arbejde, jeg gerne vil have • Jeg har en idé om hvilke typer arbejde, jeg gerne vil have • Jeg har nogle få, men uklare idéer om hvilke typer arbejde, jeg gerne vil have • Jeg ved ikke hvilke typer arbejde, jeg gerne vil have •
8	Hvordan søger du/planlægger du at søge arbejde? (Sæt gerne flere kryds)	<ul style="list-style-type: none"> • Svarer på annoncer i aviser, blade, internet, mv. • Sender uopfordrede ansøgninger • Henvender mig personligt eller telefonisk til arbejdsgiver • Bruger sociale medier (Facebook, LinkedIn mv.) • Bruger mit netværk (kontaktet venner, familie, tidligere studie- eller arbejdskolleger mv.) • Vikarbureau
9	Hvilke af følgende overvejelser gør du dig for at øge dine jobmuligheder? (Sæt gerne flere kryds)	<ul style="list-style-type: none"> • Jeg overvejer at skifte branche • Jeg overvejer at pendle langt • Jeg overvejer at flytte • Jeg overvejer at tage en anden uddannelse • Jeg overvejer at gå ned i løn • Jeg overvejer ingen af de nævnte
10	Er der noget, der gør det svært for dig at få et arbejde?	<ul style="list-style-type: none"> • Ja • Nej • Ved ikke
10a	Hvis "ja" i spørgsmål 11: Hvad gør det svært for dig at få et arbejde? (Sæt gerne flere kryds)	<ul style="list-style-type: none"> • Mit fysiske helbred • Mit psykiske helbred • Personlige- eller sociale forhold • Økonomiske forhold • Alkohol- eller stofmisbrug • Manglende læse- stave eller regnefærdigheder • Manglende IT færdigheder • Sprogbarrierer • Andet, angiv gerne _____

Den statistiske profilafklaring

Den statistiske profilafklaring, der følger når forberedelsesskemaet er blevet udfyldt, anvender maskinlæring (machine learning) til at analysere tidligere ledighedsforløb. I analysen bliver data, bestående af registerdata koblet med besvarelser

af forberedelsesskemaet, brugt til at få viden om de faktorer, der har betydende statistisk sammenhæng med nyledige dagpengemodtageres risiko for langtidsledighed.

Modellen

Princippet bag den statistiske profilafklaring er, at man ved at se på tidligere ledighedsforløb kan få en ide om, hvem i en gruppe af nyledige dagpengemodtagere, der er i risiko for at blive langtidsledig. Den statistiske vurdering i profilafklaringsværktøjet forsøger således at forudsige om den enkelte borger vil være ledig i mere end 6 måneder. Dette gøres ved at sammenligne borgeren med andre nyledige dagpengemodtagere, der historisk set er blevet langtidsledige. Maskinlæringsalgoritmen danner på denne baggrund kombinationer af faktorer der, *i gennemsnit*, bedst forudser om borgeren kan *forventes* at blive langtidsledig.

Mere præcist består modellen af et statistisk genereret beslutningstræ. Beslutningstræsalgoritmen kategoriserer i én af to grupper: risiko og ikke-risiko for langtidsledighed. Algoritmen er dermed trænet, på data, til at identificere de faktorer, der er de vigtigste for at forudsige langtidsledighed. Kategoriseringen starter med den faktor der minimerer usikkerheden ved forudsigelsen mest. I kombination med denne faktor vælges så den næste faktor, på en måde hvor de to tilsammen minimerer usikkerheden mest, og så fremdeles. Faktorer og kombinationer af faktorer, som ikke væsentligt bidrager til at mindske usikkerheden, medtages ikke. Algoritmen fører til ”grene” på beslutningstræet.

Data

Den statistiske profilafklaring er baseret på en analyse af påbegyndte dagpengeforløb i perioden maj 2016 til juni 2018, hvor borgeren samtidig har udfyldt forberedelsesskemaet i forbindelse med forløbets start (ca. 152.000 forløb). Data består af udvalgte registerdata og besvarelserne på forberedelsesskemaet. Registerdata og spørgsmål i forberedelsesskema er udvalgt med udgangspunkt i den eksisterende viden om de mekanismer, der har betydning for ledighed.

Resultatet

Ved at analysere data findes de kombinationer af faktorer, som bedst forudser om en nyledig dagpengemodtager tilhører risikogruppen. Den endelige statistiske profilafklaring består af 4 grene på beslutningstræet. Ud af de ca. 152.000 analyseret forløb bliver 20,4 pct. profileret til at være i risiko for langtidsledighed. (Se også Tabel 2.)

Den faktor, der har størst statistisk betydning for, om en et forløb er i risiko for langtidsledighed, er borgerens egen vurdering af, hvor hurtigt han/hun finder et nyt arbejde.

Borgere, der vurderer, at han/hun kommer hurtigt i arbejde eller ikke ved det, vil dog, *i gennemsnit*, tilhøre risikogruppen, hvis han/hun er indvandrer fra et vestligt land, eller efterkommer fra et vestligt eller ikke-vestligt land. Begrebet ”herkomst”, som defineret af Danmarks Statistik, er dermed betydende for, om borgeren, af den statistiske profilering, vurderes i risiko for at blive langtidsledig. Det samme gør sig gældende, hvis borgeren er over 56 år, hvilket således er en tredje faktor, som bidrager til om man tilhører risikogruppen, osv. Tabel 2 indeholder de faktorer og

kombinationer heraf, som modellen udskiller som udslagsgivende for at være i risikogruppen.

Tabel 2. Faktorer og kombinationer af faktorer, for risikogruppen.

Variabel 1	Variabel 2	Variabel 3	Variabel 4	Forløb	Risiko
SPM2: Hvor hurtigt tror du, at du får et arbejde?				5.158 (3,4 pct.)	83,1 pct.
Svar:					
- Der går mere end 6 mdr.					
- Jeg forventer snart at gå på barsel					
- Jeg forventer snart at gå på efterløn					
SPM2: Hvor hurtigt tror du, at du får et arbejde?	Herkomst			13.976 (9,2 pct.)	67,7 pct.
Svar:	Indvandrere fra vestlige lande, og efterkommer fra vestlige eller ikke-vestlige lande				
- Inden for 6 mdr.					
- Ved ikke					
SPM2: Hvor hurtigt tror du, at du får et arbejde?	Herkomst	Alder		6.941 (4,6 pct.)	64,3 pct.
Svar:	Dansk eller indvandrere fra ikke-vestlige lande (inkl. uoplyst)	Over 56 år			
- Inden for 6 mdr.					
- Ved ikke					
SPM2: Hvor hurtigt tror du, at du får et arbejde?	Herkomst	Alder	Beskæftigelsesgrad de seneste 36 mdr.	4.896 (3,2 pct.)	65,1 pct.
Svar:	Dansk eller indvandrere fra ikke-vestlige lande (inkl. uoplyst)	Under 56 år (inkl. uoplyst)	Mindre end 0,08		
- Inden for 6 mdr.					
- Ved ikke					
I alt				30.971 (20,4 pct.)	69,9 pct.

Anm.: Hver række i tabellen angiver en konkret kombination af faktorer, eller "sti gennem beslutningstræet", der fører til udskillelse i risikogruppen. I kolonnerne til højre fremgår det, hvor mange og hvor stor en andel af analysepopulationen, der har de pågældende karakteristika. Yderst til højre vises den gennemsnitlige sandsynlighed (risiko) for langtidsledighed for hver enkelt undergruppe

Performance

Det er forbundet med usikkerhed at forsøge at forudse, om et påbegyndt dagpengeforløb vil ende i langtidsledighed. Fordelen ved at bruge statistisk profilering er, at denne usikkerhed kan kvantificeres.

STAR kvantificerer performance ved brug af to gængse mål: præcision og dækningsgrad (også kendt som "recall"). Præcisionen beskriver andelen af forløb der bliver identificeret af værktøjet (risikogruppen), som der rent faktisk ender i langtidsledighed. Dækningsgraden derimod beskriver andelen af alle forløb, som endte i langtidsledighed, som værktøjet også havde identificeret som værende i risiko.

Der eksisterer en afvejning af præcision versus dækningsgrad. Man kan fx afgrænse risikogruppen til dem man statistisk set er mest sikre på, hvis man ønsker en højere præcision. Men det betyder samtidigt, at flere forløb der rent faktisk ender i langtidsledighed, ikke bliver identificeret.

Andelen af alle påbegyndte dagpengeforløb der rent faktisk ender i langtidsledighed er 45 pct.

Ud af de forløb som det statistiske profilafklaringsværktøj identificerer, som værende i risiko, ender 70 pct. i langtidsledighed. Dette er værktøjets *præcision*.

Samtidig er det dog kun 32 pct. af de forløb der rent faktisk ender i langtidsledighed, som værktøjet har været i stand til - ud fra data - at identificere som værende i

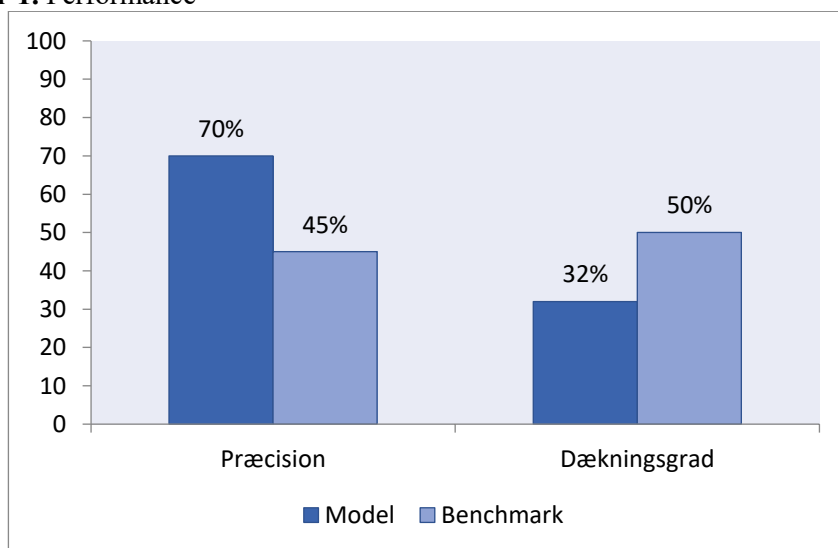
risiko. Dette er værktøjets *dækningsgrad*. (Se figur 1.) Er dette så en god eller dårlig performance?

Hvis man forestiller sig, at man slår "plat eller krone" om et påbegyndt forløb *forventes* at ende i langtidsledighed, så vil man blot ved tilfældighed, i gennemsnit, *korrekt* identificere hver andet forløb der *rent faktisk* ender i langtidsledighed. Dette svarer til en dækningsgrad på 50 pct. Så værktøjet identificerer i udgangspunktet faktisk færre af de langtidsledige, end man i teorien kan opnå med en mere 'naiv' model.

Ulempen ved en naiv model er, at man også *fejlagtigt* kommer til, at identificere hver andet forløb der *rent faktisk ikke* ender i langtidsledighed, som værende i risiko for langtidsledighed. Denne ulempe afspejles i den naive models præcision, som kun er 45 pct., da andelen af forløb i en tilfældigt udvalgt risikogruppe, der rent faktisk ender i langtidsledige, vil være lig andelen af alle påbegyndte dagpengeforløb der rent faktisk ender i langtidsledighed.

Da profilafklaringsværktøjet bruges til, at hjælpe nyledige dagpengemodtagere til at forberede sig bedst muligt på et aktivt og jobrettet kontaktføreløb, er det vigtigt at værktøjet ikke laver for mange fejl. Derfor vægtes værktøjets præcision højt. Baseret på profilafklaringsværktøjets formål, de krav der stilles til fx dataminimering og algoritmens forståelighed, og kendskab til lignede værktøjer i andre lande, vurderer STAR, at værktøjets performance samlet set er tilfredsstillende.

Figur 1. Performance



Anm.: Andelen af dem der bliver identificeret, som der rent faktisk ender i langtidsledighed ("precision"), og andelen af alle langtidsledige dagpengemodtagere der bliver identificeret (dækningsbidrag), for hhv. modellen og et naivt benchmark.

Modellen udskiller altså en delgruppe af de langtidsledige, som ud fra data kan identificeres med en høj præcision, hvilket sikrer, at omfanget af dem som modellen udskiller som værende i risiko, men som *ikke* bliver langtidsledige, er så lille som det lader sig gøre.

Men modellen udskiller langt fra alle langtidsledige. Der er således en stor gruppe, som bliver langtidsledige, men for hvem det er noget vanskeligere at forudsige ud fra data. Denne gruppe fanger modellen ikke.

STAR monitorerer løbende performance for at sikre, at præcisionen og dækningsgraden ikke løbende forringes. Dette gælder for alle nyledige, såvel som for undergrupper, fordelt på køn, alder og herkomst. Det betyder fx, at STAR løbende følger, om profileringen bliver mindre præcis, og dermed mere unfair, for fx kvinder end for mænd.

STAR vurderer også med jævne mellemrum, om nye og mere komplekse statistiske metoder kan øge værktøjets præcision. Én af ulemperne ved mere komplekse modeller kan dog være, at resultatet kan blive sværere at fortolke og forklare for den enkelte borger og sagsbehandler. Det er således en afvejning af begge forhold, som er bestemmende for, hvilken statistisk metode der anvendes i profilafklaringsværktøjet.

Spørgsmål og kontakt

Spørgsmål angående profilafklaringsværktøjet kan sendes til Specialkonsulent Carsten Søren Nielsen på csni@star.dk eller Fuldmægtig Rasmus Hoff på rho@star.dk.