

# Afgrænsning af voksne i risiko for hjemløshed

Teknisk bilag til SØM

juni 2026

Publikationen er udgivet af  
Social- og Boligstyrelsen  
Lerchesgade 35, 5  
5000 Odense C  
Tlf: 72 42 37 00  
E-mail: [info@sbst.dk](mailto:info@sbst.dk)  
[www.sbst.dk](http://www.sbst.dk)

Udgivet juni 2026

Der kan frit citeres fra rapporten med angivelse af kilde.

# Indhold

Indledning .....	3
1. Datagrundlag og metodetilgang .....	4
1.1. Datagrundlag .....	4
1.1.1. Træningsdata .....	4
1.1.2. Testdata .....	5
1.2. Udvikling af risikomodeller og afgrænsning af målgrupper .....	6
1.3. Metoder bag risikomodeller .....	6
2. Afgrænsning af risikomålgrupper .....	7
2.1. Voksne i risiko for hjemløshed (18-29 år).....	7
2.2. Voksne i risiko for hjemløshed (30-64 år).....	8
Referencer.....	10
Bilag A: Parametre, der anvendes i risikomodellerne .....	11
Bilag B: Beskyttelses- og risikofaktorer .....	13
Bilag C: Indstillinger for Random Forest .....	15

# Indledning

Dette bilag beskriver datagrundlaget og den anvendte metode, der ligger til grund for afgrænsningen af målgrupperne af voksne i risiko for hjemløshed i Den Socialøkonomiske Investeringsmodel (SØM).

Afgrænsningen af målgrupperne af voksne i risiko for hjemløshed følger den generelle tilgang til afgrænsning af risikomålgrupper i SØM, der er beskrevet i dokumentationsrapporten for SØM version 3.4 (Social- og Boligstyrelsen, 2026)

# 1. Datagrundlag og metodetilgang

I dette kapitel beskrives datagrundlaget og den metodetilgang, der anvendes til at afgrænse målgrupperne af voksne i risiko for hjemløshed til SØM.

## 1.1. Datagrundlag

Risikomodellerne baseres på en population, der som udgangspunkt omfatter hele befolkningen i alderen 18-64 år (ultimo kalenderåret) fra 2008 og frem som for øvrige SØM-målgrupper.

Borgere, der har taget ophold på § 110-boformer, fx herberger og forsorgshjem, til og med 2008, fjernes dog fra populationen, da vi kun er interesseret i borgere, der oplever førstegangshjemløshed.

Indikatoren for om en borger oplever en hjemløshedssituation er i risikomodellerne opgjort som første gang, de indskrives på § 110-boformer. Det skyldes, at netop indskrivningsoplysninger på § 110-boformer pt. er de mest valide og kontinuerte individdata for borgere i hjemløshed. Populationen tilknyttes en lang række registerbaserede oplysninger fra SØMs datagrundlag, herunder karakteristika og forbrug af offentlige ydelser og aktiviteter. Derudover opgøres også øvrige karakteristika, der, på baggrund af en litteraturgennemgang, vurderes at være relevante i forhold til risiko for hjemløshed. En samlet liste over oplysninger, som anvendes i risikomodellen, findes i bilag A.

Alder for populationen forventes at påvirke, hvilke beskyttelses- og risikofaktorer, der har størst betydning for at opleve hjemløshed. Derfor opdeles populationen i henholdsvis 18-29-årige og 30-64-årige.

Derudover opdeles populationen i henholdsvis træningsdata (80 pct. af populationen) og testdata (20 pct. af populationen). Figur 1.1 illustrerer opdelingen. Det er træningsdata, der anvendes til at *træne* risikomodellen, mens testdata bruges til at estimere, hvor god risikomodellen er til at prædiktere risikoen for hjemløshed.

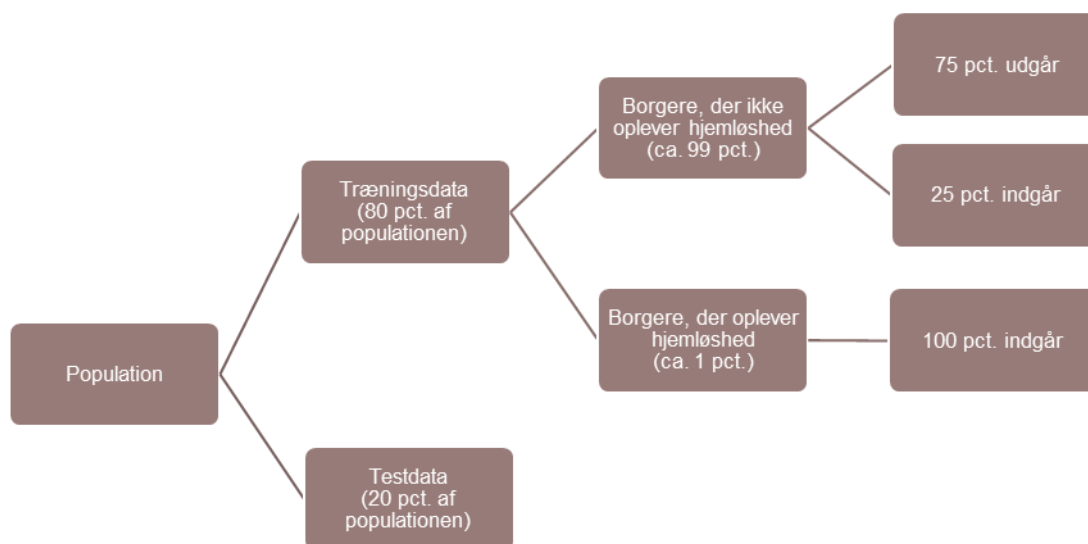
### 1.1.1. Træningsdata

I træningsdatasættene optræder hvert individ én gang. Borgere, der på et tidspunkt i løbet af analyseperioden oplever hjemløshed, inkluderes i kalenderåret inden de første gang oplever hjemløshed, mens borgere, der ikke oplever hjemløshed, inkluderes i et tilfældigt år.

Fordelingen mellem dem, der hhv. oplever og ikke oplever hjemløshed, er meget skæv i træningsdatasættene. Derfor udvælges tilfældigt 25 pct. af de borgere, der ikke har oplevet hjemløshed, for at sikre en mindre skæv fordeling mellem de to grupperinger. Figur 1.1 illustrerer opdelingen. Fordelingen af borgere i hjemløshed i de endelige træningsdatasæt er ca. 2,38 pct. for 18-29-årige og ca. 2,36 pct. for 30-64-årige.

Figur 1.1

Illustration af opdeling af population i test- og træningsdata



### 1.1.2. Testdata

Testdata anvendes til at beregne risiko for hjemløshed på baggrund af risikomodellen. Det er således testdata, som risikomålgrupperne baserer sig på.

I testdatasættene optræder et individ flere gange. Borgere, der på et tidspunkt i løbet af analyseperioden oplever hjemløshed, inkluderes til og med kalenderåret inden de første gang oplever hjemløshed. Borgere, der ikke oplever hjemløshed, inkluderes i alle kalenderår i analyseperioden.

I tabel 1.1 nedenfor fremgår et overblik over antallet af observationer i de fire datasæt.

Tabel 1.1

#### Antal observationer i trænings- og testdatasæt

	18-29-årige	30-64-årige
Træningsdata	475.096	831.217
Testdata	2.965.238	8.788.937

Anm.: Årsagen til det markant højere antal observationer i testdatasættene er, at enkelte individer kan fremgå i flere år, mens de i træningsdatasættet kun optræder i ét år.

Kilde: Egne beregninger på Danmarks Statistiks registerdata.

## 1.2. Udvikling af risikomodeller og afgrænsning af målgrupper

Først udvikles en risikomodel, der estimerer risikoen for hjemløshed. Modellen testes derefter ved, at hvert individ i testdatasættet får beregnet sin risiko for hjemløshed. Afsnit 1.3 beskriver de metoder, der anvendes i udviklingen af risikomodellen samt hvilken metode, der performer bedst.

Når modellen er færdigudviklet, identificeres de væsentligste beskyttelses- og risikofaktorer, og til sidst afgrænses risikomålgruppen ud fra en tærskelværdi, som fastlægger den nedre grænse for risikoniveauet i målgruppen. De væsentligste beskyttelses- og risikofaktorer fremgår af bilag B, mens afgrænsningen af risikomålgruppen beskrives i afsnit 2.

## 1.3. Metoder bag risikomodeller

I forbindelse med udviklingen af risikomodellerne afprøves fire statistiske og maskinlæringsbaserede metoder: Logistisk regression, LASSO, Random Forest og Extreme Gradient Boosting.

For at kunne udvælge den foretrukne risikomodel er der behov for at kunne vurdere modellernes evne til at forudsige risikoen for hjemløshed. Til det formål anvendes to udbredte performancemål: AUC-ROC og AUC-PR. Der henvises til dokumentationsrapporten for en nærmere beskrivelse (Social- og Boligstyrelsen, 2026).

I tabel 1.2 fremgår AUC-ROC og AUC-PR-scoringer for de fire metoder. En højere score indikerer en bedre præcision i forhold til at forudsige, hvorvidt en borger i fremtiden vil opleve hjemløshed. På baggrund af performancemålene fremstår Random Forest som den mest effektive metode for begge delmålgrupper. De anvendte indstillinger til Random Forest beskrives i bilag C.

Tabel 1.2

### Performancemål for risikomodellerne

		Logistisk Regression	LASSO	Random Forest	Extreme Gradient Boosting
18-29-årige	AUC-ROC	0,9202	0,9233	0,9802	0,9547
	AUC-PR	0,0316	0,000	0,3476	0,0640
30-64-årige	AUC-ROC	0,9258	0,9263	0,9849	0,9562
	AUC-PR	0,0259	0,000	0,2120	0,0544

Kilde: Egne beregninger på Danmarks Statistiks registerdata.

## 2. Afgrensning af risikomålgrupper

Risikomodellen har, som tidligere nævnt, tildelt en risiko for at opleve hjemløshed til alle observationer i testdatasættet. På baggrund af denne risiko skal hver observation klassificeres som enten 'i risiko for hjemløshed' eller 'ikke i risiko for hjemløshed'. Denne opdeling sker ved fastsættelse af en tærskelværdi.

Formålet med tærskelværdien er at definere en nedre grænse for størrelsen, den beregnede risiko bør have, for at en borger ifølge modellen er i højere risiko for at opleve hjemløshed. Det er således alene tærskelværdien, som bestemmer, hvem der bliver en del af risikomålgruppen, da hvert individs risikoværdi ikke ændrer sig afhængigt af tærskelværdien.

Til beregningen af tærskelværdien anvendes Youdens J-indeks, som forsøger at balancere gruppestørrelse og præcision (for nærmere beskrivelse, se (Social- og Boligstyrelsen, 2026)).

I de efterfølgende afsnit beskrives den fastsatte tærskelværdi for de to aldersgrupper, herunder betydningen for størrelsen af risikomålgrupperne.

### 2.1. Voksne i risiko for hjemløshed (18-29 år)

Der er fastsat en tærskelværdi på 27 pct. for risikoen for at opleve hjemløshed blandt borgere i alderen 18-29 år. Det betyder, at alle borgere med en beregnet risiko over 27 pct. indgår i risikomålgruppen. Tabel 2.1 viser en confusion matrice, der illustrerer, hvor mange borgere modellen forudsiger vil opleve hjemløshed (prædikeret), sammenholdt med hvor mange der faktisk gør det (reference).

Tabel 2.1

#### Confusion matrice over hele testdatasættet (18-29-årige)

	Borgere, der oplever hjemløshed	Borgere, der ikke oplever hjemløshed	I alt
Hjemløshed (prædikeret)	3.123	47.698	50.821
Ikke hjemløshed (prædikeret)	933	2.913.484	2.914.417
I alt	4.056	2.961.182	2.965.238

Anm.: Confusion matricen er udarbejdet på baggrund af beregnede risikoer fra risikomodellen, der er beskrevet nærmere i afsnit 2. Borgere, som ifølge modellen har en forhøjet risiko for hjemløshed, vil fremgå under 'Hjemløshed (prædikeret)'.

Kilde: Egne beregninger på registerdata fra Danmarks Statistik, Sundhedsdatastyrelsen og Styrelsen for Arbejdsmarked og Rekruttering.

Da tabel 2.1 viser hele testdatasættet, vil individer indgå i flere år, og derfor kan vi ikke bedømme størrelsen på risikomålgruppen ud fra tabellen.

En måde at gøre det på, er i stedet at udtage hvert individ én gang fra testdatasættet. Helt specifikt udtages hvert individ i målgruppen i det første år, de har en beregnet risiko, der er højere end tærskelværdien.

Tabel 2.2 viser størrelsen på den endelige risikomålgruppe. Summen af den første række er størrelsen på risikomålgruppen, når den består af unikke borgere.

Tabel 2.2

**Confusion matrice over unikke borgere (18-29-årige)**

	Borgere, der oplever hjemløshed	Borgere, der ikke oplever hjemløshed	I alt
Hjemløshed (prædikteret)	2.760	16.672	19.432
Ikke hjemløshed (prædikteret)	661	447.734	448.395
I alt	3.421	464.406	467.827

Anm.: Confusion matricen er udarbejdet på baggrund af beregnede risikoer fra risikomodellen, der er beskrevet nærmere i afsnit 2. Borgere, som ifølge modellen har en forhøjet risiko for hjemløshed, vil fremgå under 'Hjemløshed (prædikteret)'.

Kilde: Egne beregninger på registerdata fra Danmarks Statistik, Sundhedsdatastyrelsen og Styrelsen for Arbejdsmarked og Rekruttering.

Testdatasættet omfatter 467.827 unikke borgere, hvoraf 3.421 (<1 pct.) oplever hjemløshed. Med en tærskelværdi på 27 pct. udgør risikomålgruppen 19.432 borgere, og blandt disse ender 2.760 (14 pct.) i hjemløshed. Samtidig er der 661 borgere (19 pct. ud af 3.421), som faktisk oplever hjemløshed, men som ikke indgår i risikomålgruppen, fordi deres beregnede risiko ligger under tærsklen.

## 2.2. Voksne i risiko for hjemløshed (30-64 år)

Der er fastsat en tærskelværdi på 25 pct. for risikoen for at opleve hjemløshed blandt borgere i alderen 30-64 år. Det betyder, at alle borgere i testdatasættet med en beregnet risiko over 25 pct. indgår i risikomålgruppen. I tabel 2.3 fremgår en confusion matrice over hvor mange borgere, modellen forudsiger, vil opleve hjemløshed (prædikteret), sammenholdt med hvor mange, der faktisk gør det (reference).

Tabel 2.3

**Confusion matrice over hele testdatasættet (30-64-årige)**

	Borgere, der oplever hjemløshed	Borgere, der ikke oplever hjemløshed	I alt
Hjemløshed (prædikteret)	6.132	143.946	150.078
Ikke hjemløshed (prædikteret)	1.046	8.637.813	8.638.859
I alt	7.178	8.781.759	8.788.937

Anm.: Confusion matricen er udarbejdet på baggrund af beregnede risikoer fra risikomodellen, der er beskrevet nærmere i afsnit 2. Borgere, som ifølge modellen har en forhøjet risiko for hjemløshed, vil fremgå under 'Hjemløshed (prædikteret)'.

Kilde: Egne beregninger på registerdata fra Danmarks Statistik, Sundhedsdatastyrelsen og Styrelsen for Arbejdsmarked og Rekruttering.

Da tabel 2.3 viser hele testdatasættet, vil individer indgå i flere år, og derfor kan vi ikke bedømme størrelsen på målgruppen ud fra tabellen. Derfor udtages hvert individ i målgruppen i det første år, de har en risiko, der er højere end tærskelværdien.

Tabel 2.4

**Confusion matrice over unikke borgere (30-64-årige)**

	Borgere, der oplever hjemløshed	Borgere, der ikke oplever hjemløshed	I alt
Hjemløshed (prædikeret)	5.092	42.410	47.502
Ikke hjemløshed (prædikeret)	676	770.551	771.227
I alt	5.768	812.961	818.729

Anm.: Confusion matricen er udarbejdet på baggrund af beregnede risici fra risikomodellen, der er beskrevet nærmere i afsnit 2. Borgere, som ifølge modellen har en forhøjet risiko for hjemløshed, vil fremgå under 'Hjemløshed (prædikeret)'.

Kilde: Egne beregninger på registerdata fra Danmarks Statistik, Sundhedsdatastyrelsen og Styrelsen for Arbejdsmarked og Rekruttering.

Tabel 2.4 viser størrelsen på den endelige risikomålgruppe for 30-64-årige. Testdatasættet omfatter 818.729 unikke borgere, hvoraf 5.768 borgere (<1 pct.) oplever hjemløshed. Med en tærskelværdi på 25 pct. består risikomålgruppen af 47.502 borgere, hvoraf 5.092 (11 pct.) oplever hjemløshed. Samtidig er der 676 borgere (12 pct. ud af de 5.768), som faktisk oplever hjemløshed, men som ikke indgår i risikomålgruppen.

# Referencer

Social- og Boligstyrelsen. (2026). *Den Socialøkonomiske Investeringsmodel (SØM) version 3.4*. Odense: Social- og Boligstyrelsen.

# Bilag A: Parametre, der anvendes i risikomodelle

Nedenfor fremgår en liste over alle de parametre, som anvendes i risikomodelle.

Tabel A.1

## Liste over parametre anvendt som datagrundlag til risikomodelle

Alder	Ambulant kontakt i psykiatrien
År	Kontakt til almen praksis
Køn	Kontakt til psykiater
Antal børn i E-Familien	Kontakt til psykolog
Etnicitet	Kontakt til fysioterapeut
Region	Kontakt til andre behandlere
Oprindelse	Psykofarmaka recepter
Gift/Samlevende	Øvrige recepter
Højeste uddannelsesniveau gennemført	Boligstøtte
Socioøkonomisk klassifikation	Kontanthjælp
Log(Disponibel indkomst)	Førtidspension
Ophold på midlertidige botilbud	Øvrige midlertidige overførsler
Ophold på længerevarende botilbud	SU
Støtte på botilbud	Fleksløn
Støtte i hjemmet	Løntilskud
Tilskud til hjælp	Øvrige beskæftigelsesindsatser
Handicaphjælp	Vejledning og opkvalificering
Ledsagerordning og kontaktperson	Mentorstøtte
Beskyttet beskæftigelse eller aktivitet	Beskæftigelse
Individuel/gruppebaseret støtteforløb	Elev på ungdomsuddannelse
Ophold på kvindekrisecenter	Elev på videregående uddannelse
Merudgifter til voksenområdet	Domme vedrørende vold og sædelighedsforbrydelser
Hjælp til enkeltudgifter	Domme vedrørende indbrud, tyveri og hærværk
Øvrige kontante ydelser	Domme vedrørende færdselsloven og andre særlove
Tabt arbejdsfortjeneste	Frihedsberøvelse
Efterværn (Trin 3)	Sigtelser
Efterværn (Trin 4)	Fysisk funktionsnedsættelse
Efterværn (Trin 5)	Kognitiv eller intellektuel funktionsnedsættelse
Efterværn (Trin 6)	Moderat psykisk lidelse
Alkoholmisbrugsbehandling	Svær psykisk lidelse
Stofmisbrugsbehandling	Anden psykisk lidelse
Indlæggelse i somatikken	Diagnoser vedrørende selvskade
Akutte ophold i somatikken	Misbrugsdiagnoser

(tabellen fortsættes på næste side)

Tabel A.2 (fortsat)

**Liste over parametre anvendt som datagrundlag til risikomodelle**

Ambulant kontakt i somatikken	Diagnoser vedrørende tics
Indlæggelse i psykiatrien	Anbragt i barndommen
Akutte ophold i psykiatrien	Forebyggende eller støttende foranstaltning i barndommen

Anm.: For definition af de enkelte parametre se Social- og Boligstyrelsen (2026). *Den Socialøkonomiske Investeringsmodel (SØM) version 3.4. Dokumentationsrapport.*

Kilde: Egne beregninger på registerdata fra Danmarks Statistik, Styrelsen for Arbejdsmarked og Rekruttering og Sundhedsdatastyrelsen.

# Bilag B: Beskyttelses- og risikofaktorer

De væsentligste beskyttelses- og risikofaktorer er fundet via 'Impurity'-metoden. Denne metode ser på, hvilke faktorer der enkeltvist er bedst til at opdele data i hjemløsheds- og ikke-hjemløshedstilfælde. 'Impurity'-metoden rangerer faktorerne ift. væsentlighed. Det er dog ikke muligt at sige noget om, hvor væsentlige beskyttelses- og risikofaktorerne er relativt til hinanden. Ej heller hvorvidt der er tale om, at den enkelte faktor øger eller reducerer sandsynligheden for hjemløshed.

Tabel B.1

## Top-20 væsentligste beskyttelses- og risikofaktorer for hjemløshed

	Blandt 18-29-årige	Blandt 30-64-årige
1	Kontanthjælp	Disponibel indkomst
2	Stofmisbrugsbehandling	Kontanthjælp
3	Disponibel indkomst	Alder
4	Sigtelser	Misbrugsdiagnose
5	Vejledning og opkvalificering	Kontakter til almen praksis
6	Beskæftigelse	Alkoholmisbrugsbehandling
7	Kontakter til almen praksis	Akutte ophold i psykiatrien
8	Anbringelse i barndommen	Beskæftigelse
9	Alder	Indlæggelse i psykiatrien
10	Akutte ophold i psykiatrien	Stofmisbrugsbehandling
11	Misbrugsdiagnose	Sigtelser
12	Domme for indbrud, tyveri og hærværk	Region
13	Mentorstøtte	Vejledning og opkvalificering
14	Akutte ophold i somatikken	Kontakt til øvrige sygesikringskontakter (ekskl. almen praktiserende læge, psykiater, psykolog og fysioterapeut)
15	Indlæggelse i psykiatrien	Indlæste recepter (ekskl. psykofarmaka)
16	Region	Ambulante, somatiske sygehuskontakter
17	Domme for færdselsloven og andre særlove	Øvrige midlertidige overførsler (ekskl. kontanthjælp)
18	SU	Akutte ophold i somatikken
19	Ungdomsuddannelse	Boligstøtte
20	Hjælp til enkeltudgifter	Gift/samlevende

Anm.: Det er ikke muligt at sige noget om, hvor væsentlige beskyttelses- og risikofaktorerne er relativt til hinanden. Ej heller hvorvidt der er tale om, at den enkelte faktor øger eller reducerer sandsynligheden for hjemløshed.

Kilde: Egne beregninger på registerdata fra Danmarks Statistik, Sundhedsdatastyrelsen og Styrelsen for Arbejdsmarked og Rekruttering.

For begge aldersgrupper er forsørgelses- og indtægtsgrundlag blandt af de mest væsentlige beskyttelses- og risikofaktorer, særligt kontanthjælp (tidl. uddannelseshjælp for de 18-29-årige) og disponibel indkomst.

For de 18-29-årige er andre væsentlige beskyttelses- og risikofaktorer stofmisbrugsbehandling, sigtelser samt vejledning og opkvalificering. For de 30-64-årige er det derimod alder, brug af

sundhedsyndelser, særligt psykiatrien, samt misbrug (herunder -behandling og -diagnose), der udgør væsentlige beskyttelses- og risikofaktorer.

# Bilag C: Indstillinger for Random Forest

Random Forest har følgende indstillinger, som kan ændres, således at træerne bygges på en måde, der er hensigtsmæssig i forhold til det datagrundlag, som anvendes.

Tabel C.1

## Indstillinger i Random Forest

Indstilling	Mulige værdier for indstillingen
Det minimale antal observationer, der kræves for at foretage en ny opdeling i data	5, 10, 20, 50
Antallet af parametre, der overvejes ved hver opdeling af træet	5, 10, 20, 40
Den maksimale dybde af hvert træ	10, 30, 50
Reglen, der anvendes til at vælge hver opdeling i data	Gini Impurity, Extratrees
Antallet af træer	100, 200, 500

Anm.: Udvælgelse af hvilke værdier, der er mulige for indstillingerne at antage, er sket på baggrund af viden omkring Random Forest og datagrundlaget.

De anvendte indstillinger vælges pba. en hyper parameter tuning, hvor alle mulige kombinationer af indstillinger testes. Indstillingerne sikrer en afvejning af at minimere støj i datagrundlaget samt sikre mest mulig robusthed i modellen.

For målgruppen 18-29-årige er der valgt følgende indstillinger:

- Det minimale antal observationer, der kræves for at foretage en ny opdeling i data = 5
- Antallet af parametre, der overvejes ved hver opdeling af træet = 5
- Den maksimale dybde af hvert træ = 50
- Reglen, der anvendes til at vælge hver opdeling i data = Gini Impurity
- Antallet af træer = 500

For målgruppen 30-64-årige er der valgt følgende indstillinger:

- Det minimale antal observationer, der kræves for at foretage en ny opdeling i data = 10
- Antallet af parametre, der overvejes ved hver opdeling af træet = 10
- Den maksimale dybde af hvert træ = 50
- Reglen, der anvendes til at vælge hver opdeling i data = Gini Impurity
- Antallet af træer = 500



Social- og Boligstyrelsen  
Lerchesgade 35, 5  
5000 Odense C  
Tlf.: 72 42 37 00

[www.sbst.dk](http://www.sbst.dk)