

Brugermanual

Indlæsning af CSV-fil

Efter programmets opstart præsenteres indlæsning af CSV-fil. Her er følgende valgmuligheder:

- Valg af fil
- Valg af separationskarakter mellem hvert felt
- Valg af separationskarakter mellem hver værdi
- Valg om fjernelse af mellemrum

Når der trykkes "Browse" bliver et stifinder vindue åbnet og CSV-filen skal lokaliseres og vælges. Herefter vælges separationskarakterer i filen, og der angives i checkboksen hvorvidt mellemrum skal fjernes.

Standardindstillinger

Som standard er filen "UserData.csv" valgt, hvilket er dette projekts database-fil. Der ud over er separationkaraktererne sat til csv standard, og mellemrum fjernes som udgangspunkt også. Når de ønskede indlæsningsindstillinger er valgt, kan der trykkes på "Import", og kerneprogrammet opstartes.

Kerneprogrammet

Så snart kerneprogrammet opstartes, vises en venteskærm, hvor der går et par min. inden programmet er klar. Når venteskærmen er væk kan der ses indstillinger til grafen i toppen af vinduet med profilerne plottet ind. Profildetaljer kan ses til højre i vinduet, men der er dog ikke vist detaljer indtil en profil er valgt. Nederst i vinduet kan der findes en liste over de profiler der mest sandsynligt er outliers.

Efter venteskærmen er der mulighed for ændre på indstillingerne for den viste graf, ved hjælp af indstillingerne i toppen af vinduet. Samt der er mulighed for at ændre på indstillingerne for outlier detektering, ved at gøre brug af indstillingerne umiddelbart under grafen til højre. Ønskes profildetaljer vist, kan der trykkes på et punkt i grafen. Derudover kan der trykkes

på feltet for det pågældende punkt/profil i datagridet, hvorefter punktet bliver fremhævet i grafen ud over at detaljerne bliver vist.

Idét eksportering af data ikke er implementeret, er det ikke muligt at gemme de ændringer der foretages, derfor trykkes der på knappen "Close" i stedet for "Commit".

Grafen

Til grafen er følgende valg mulige:

- Valg af distance metric
- Valg af clusermetode
- Antal af clusters

Dropdown-vinduet ved "Distance Metric" indeholder de forskellige distancemetoder, og ved "Clustering Method" indeholder dropdown-vinduet de forskellige clusteringmetoder. Herefter vælges ønskede antal af clusters, og der trykkes på knappen "Generate", som opdaterer grafen.

Standardindstillinger

Som standard er "Distance Metric" sat til "Euclidian distance", og "Clustering Method" er valgt til "K-means", samt antallet af clusters er som udgangspunkt 5.

Outlier Detektering

Følgende valgmuligheder er tilgængelige til outlier detektering:

- Valg af metode
- Antal profiler vist
- Antal nabopunkter/profiler det pågældende punkt/profil skal sammenlignes med

Dropdown-vinduet ved "Method" indikerer outlier detekterings metoden, "Limiter" er antallet af profiler der vises og "Neighbours to Compare" er antallet af profiler hvert punkt

sammenlignes med.

I kolonnen "False/fake" er der mulighed for at vælge om den pågældende profil er sand eller forkert eller om sandhedsstatussen ikke kan bedømmes. I kolonnen sættes sandhedsstatussen ved at klikke på det pågældende felt x antal gange. En gang for sand, to gange for forkert, tre gange for ukendt eller fire gange for at vende tilbage til standarden "Not Set".

Standardindstillinger

Som standard er den valgte outlier detekterings metode "Local Outlier Factor", "Limiter" sat til 50 og "Neighbours to Compare" er sat til 10. Herefter trykkes på den tilhørende "Generate"-knap, som opdaterer datagridet. Sandhedsstatussen til hver profil bliver bibeholdt selvom der ændres på indstillingerne.

Detaljer

I detaljer er det muligt at se profiloplysninger, samt den procentvise forskel mellem profilen og centroiden i det pågældende cluster. Profiloplysningerne er sorteret efter stigende procentmæssige lighed, og er markeret med farve der indikerer hvilken information som udskiller sig mest fra centroiden.