

Segmenteringsanalyse

Opgaven:

Ved hjælp af data fra filen "DataScientist-Case-Dataset.xlsx" og ved hjælp af visualisering og/eller statistik, skal du besvare spørgsmålet:

"Hvad er de vigtigste parametre for at forudse om en bruger er konverteret eller ej?"

Viden om hvad det betyder om en bruger er konverteret eller ej, er med vilje sløret for opgavens formål.

Felter:

<i>Felt</i>	<i>Beskrivelse</i>
<i>customer_id</i>	Numerisk ID af bruge
<i>converted</i>	Binær beskrivelse af om bruger er konverteret eller ej
<i>customer_segment</i>	Numerisk ID af hvilket segment brugeren tilhører
<i>gender</i>	Brugerens køn
<i>age</i>	Brugerens alder
<i>related_customers</i>	Antal personer som er relateret til brugeren
<i>family_size</i>	Antal af familiemedlemmer
<i>initial_fee_level</i>	Startbetaling for brugeren
<i>credit_account_id</i>	Hash for brugerens konto nr. None bliver vist som "9b2d5b4678781e53038e91ea5324530a03f27dc1d0e5f6c9bc9d493a23be9de0"
<i>branch</i>	Hvilken gren brugeren tilhører

Krav:

Vi fortrækker at du afleverer dit arbejde som et GIT repository i enten github eller bitbucket. Husk at dit repository skal være offentligt og send da link via mail.

Til nøds kan du sende din tilgangsmåde samt beskrivelse i en ZIP på mail.

Deadline er tirsdag d. 10/05 kl 23:59 og skal sendes til LWE011@PET.dk.

Anbefalet værktøjer/metoder:

- Brug statistik, visualisering eller/og andre analytiske metoder/tilgange til at understøtte dine argumenter. Vi lægger vægt din begrundelse og overvejelse bag dine valg og konklusioner og ikke det samlede antal timer/antal grafer i slutproduktet.
- Oplagte programmer er Python, Excel og Visualiseringsværktøjer som Power BI eller Tableau men der er ingen krav.
- Husk at inkludér din fulde besvarelse i PDF samt kode når du afsender.

Skriv til LWE011@PET.dk hvis du har spørgsmål omkring datasættet eller opgaven.