



Afidentificering af fritekstfelter – løsning udviklet af DEFACTUM

Automatiseret kommentar anonymisering

- Problem: Klinikere er glade for fritekstkommentarer, men patienterne skriver personoplysninger
- Lige nu løser vi det ved at begrænse adgangen
- Vi har kigget på om man kunne automatisere til vores anonymisering

Fremgangsmøde

- Alle kommentarer gennemses efter følgende:
 - Potentielle egenavne: simpel AI sprogmodel
 - Potentielle CPR-numre: Regelbaseret
 - Potentielle telefonnumre: Regelbaseret
 - Potentielle E-mailadresser: Regelbaseret
 - Potentielle datoer: Regelbaseret
 - Ord fra en prædefineret liste: Regelbaseret
- Desuden anonymiseres hele kommentaren hvis den er over 1000 tegn eller ikke på dansk

2025_kommentar_anonymisering - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

ntarer.R* x Untitled2* x kommentarer x fiktive_kommentarer x diagnoser x psyk_kommentarer.R x samlet x

```

1 #installer - køres lige efter nulpstilling, hvis spacy ikke kan loades
2 #source("installer.R")
3
4 #hent funktioner
5 source("kommentar_anonymiseringsfunktioner.R")
6
7 library(readxl)
8 fiktive_kommentarer <- read_excel("C:/Users/jehoeje/OneDrive - Region Midtjylland/fiktive_kommentarer.xlsx")
9
10 #ordlister
11 diagnoser <- read_delim("diagnoser.txt", delim = ";", escape_double = FALSE, trim_ws = TRUE)
12
13 #filtrer så det kun er overordnede diagnoser
14 diagnoseliste <- diagnoser %>%
15   filter(nchar(Kode)==4) %>%
16   sample_n(100)
17
18 diagnoseliste <- diagnoseliste$Kode
19
20 words_not_to_be_anonymized <- c("hat", "HAT", "HAT", "angst", "angsten")
21
22 kommentarer <- word_flag_personal_data(fiktive_kommentarer$`Fiktive kommentarer`, diagnoseliste)
23 resultater <- word_flagging_analyze_findings(kommentarer)
24
25
26 openxlsx::write.xlsx(kommentarer, file="demo.xlsx")
27

```

26:45 (Top Level) R Script

Console Terminal Background Jobs

R 4.5.1 N:/Begraenset/SUNKONKV_DataInd/kode_ad_hoc_analyser/2025_kommentar_anonymisering/

```

Antal Kommentarer i Alt      : 5      (100.0%)
Antal Med Egennavne         : 5      (100.0%)
Kommentarer Gennemsnitslængde : 190 tegn
Antal Kommentarer over 1000 Tegn : 0      (0.0%)
Antal Ikke Dansk            : 0      (0.0%)
Antal Med Cpr               : 1      (20.0%)
Antal Med Telefonnummer      : 1      (20.0%)
Antal Med Email             : 0      (0.0%)
Antal Med Dato              : 3      (60.0%)
Antal Med Ord Fra listen    : 0      (0.0%)

```

Environment History Connections Tutorial

R Global Environment

Data

- diagnoser 19867 obs. of 2 variables
- fiktive_kommentarer 5 obs. of 1 variable
- kommentarer 5 obs. of 14 variables
- resultater 1 obs. of 12 variables
- test3 List of 1

Values

- diagnoseliste chr [1:100] "DI21" "DM21" "DQ65" "DR56" "DJ92" "DK66" "D...
- python_environment__ TRUE
- python_environment__ "conda"
- words_not_to_be_ano... chr [1:5] "hat" "HAT" "HAT" "angst" "angsten"

Functions

- anonymiser_qualitat... function (df, kommentar_kolonne)

Files Plots Packages Help Viewer Presentation

New Folder New File Delete Rename More

N: > Begraenset > SUNKONKV_DataInd > kode_ad_hoc_analyser > 2025_kommentar_anonymisering

Name	Size	Modified
..		
2025_kommentar_anonymisering.Rproj	253 B	Nov 10, 2025, 9:18 AM
kommentar_anonymiseringsfunktioner.R	25.9 KB	Oct 10, 2025, 9:44 PM
psyk_kommentarer.R	5.2 KB	Oct 10, 2025, 4:47 PM
nlp.py	543 B	Oct 10, 2025, 4:47 PM
installer_miniconda.R	1.3 KB	Oct 10, 2025, 4:47 PM
installer_standard_python_311.R	2.7 KB	Oct 10, 2025, 4:47 PM
alle_kommentarer.R	1.5 KB	Oct 10, 2025, 4:47 PM
analyser_alle	840 B	Oct 10, 2025, 4:46 PM
.Rhistory	22.4 KB	Oct 10, 2025, 9:46 AM
.#gitignore	31 B	Oct 9, 2025, 8:13 PM
.gitignore	1.7 KB	Oct 9, 2025, 7:51 PM
psyk_kommentarer.xlsx	19.3 KB	Oct 7, 2025, 10:11 AM
kommentarer_almen_praksis.R	0 B	Sep 18, 2025, 10:07 AM
samlet_data_alle_kommentarer.xlsx	34.6 MB	Aug 8, 2025, 10:08 PM
alle_kommentarer.xlsx	23 MB	Aug 8, 2025, 5:50 PM

midtjylland.dk

Fiktive eksempler på anonymiserede kommentarer

- Der var en super professionel behandling på afdelingen i den tid jeg var indlagt. En stor tak til personalet. Med venlig hilsen Jens Jensen, Nordre Ringegade 4a, 32451678
- Fund: Jens, Jensen, Nordre, Ringegade, 32451678
- Der var en super professionel behandling på afdelingen i den tid jeg var indlagt. En stor tak til personalet. Med venlig hilsen [...] [...]en, [...] [...] 4a, [...]

Fiktive eksempler på anonymiserede kommentarer

- Alt for lang ventetid, over et år på operation. Hvorfor indkalder man 5 personer på samme tid. Der kunne vel indkaldes med 30 min mellemrum, så man undgår den lange ventetid. Ellers var alt meget fint. Mvh Jens Peter Jensen, 000000-0000
- Fund: Jens, Peter, Jensen, 000000-0000, 000000
- Alt for lang ventetid, over et år på operation. Hvorfor indkalder man 5 personer på samme tid. Der kunne vel indkaldes med 30 min mellemrum, så man undgår den lange ventetid. Ellers var alt meget fint. Mvh [...] [...] [...]en, [...]

Fiktive eksempler på anonymiserede kommentarer

- Da jeg ringede ind den 2. februar for at få en tid til at snakke med lægen fik jeg af vide at det var svært at få tid men at han ville ringe mig op dagen efter, hvilket ikke skete. På anden dagen blev jeg ringet op af sygeplejersken. Sagde jeg ikke skulle snakke med hende men med Lægen Ole Hansen. Ole ringede mig op en uge efter.
- Fund: Ole, Hansen, 2. februar
- Da jeg ringede ind den [...] for at få en tid til at snakke med lægen fik jeg af vide at det var svært at få tid men at han ville ringe mig op dagen efter, hvilket ikke skete. På anden dagen blev jeg ringet op af sygeplejersken. Sagde jeg ikke skulle snakke med hende men med Lægen [...] [...]. [...] ringede mig op en uge efter.

Fiktive eksempler på anonymiserede kommentarer

- Lægen Hanne Hansen skældte ud over jeg var der. Det var meget utrygt og jeg frygter, at jeg ikke har fået den rette behandling.
- Fund: Hanne, Hansen
- Lægen [...] [...] skældte ud over jeg var der. Det var meget utrygt og jeg frygter, at jeg ikke har fået den rette behandling.

Fiktive eksempler på anonymiserede kommentarer

- Jeg har kun ros tilovers for jer. Jeg er meget, meget tryk hos jer. Mvh Kanya Saengthong
- Fund: Kanya, Saengthong
- Jeg har kun ros tilovers for jer. Jeg er meget, meget tryk hos jer. Mvh [...] [...]

Resultat alle vores kommentarer

■	📄 Resultater		
■	=====		
■	Antal Kommentarer i Alt	: 226288	(100.0%)
■	Antal Kommentarer over 1000 Tegn	: 2774	(1.2%)
■	Antal Ikke Dansk	: 4704	(2.1%)
■	Kommentarer Gennemsnitslængde	: 169 tegn	
■	Antal Med Egnavne	: 22501	(9.9%)
■	Antal Med Cpr	: 69	(0.0%)
■	Antal Med Telefonnummer	: 111	(0.0%)
■	Antal Med Email	: 25	(0.0%)
■	Antal Med Dato	: 1118	(0.5%)
■	Antal Med Ord Fra Listen	: 0	(0.0%)
■	Antal Whitelistede Ord Flagget	: 8	(0.0%)
■	Antal Flaggede Kommentarer	: 23175	(10.2%)
■	📊 Andel Flaggede Kommentarer	: 10.2%	(23175 ud af 226288)

Svagheder

- Ikke perfekt anonymisering
 - Alt er kontekstuel (eks. "Sygeplejersken")
 - Detaljerige kommentarer er svære
 - Fjerner nogle gange for meget
- Virker ikke når kommentarene ikke på dansk

Styrker

- Tidsbesparende, selvom det ikke kan stå alene.
- Alt kører lokalt

Disclaimer

- Vi kan ikke garantere perfekt anonymisering og I er ansvarlige for hvordan i udstiller data