

Automated 'Customer email classification' {

[Data Science @UL-FRI]

< Ekipa Orca - Miha Likar, Neža Marija
Slosar >

< Mentroja - doc. dr. Slavko Žitnik, Andrej
Miščič >

}

Uvod - pregled podatkov; {

'60.000+ mailov kategoriziranih v 15 kategorij iz obdobja januar - junij 2021'

<p Asistenca, Asistenca brez primera, Banke, CORIS, Drajev, GDPR, Help desk - interna javnost, Marketinške akcije, Odjava B2B, PDPZ, Premoženje, Zavarovanja - zunanji partnerji, Zdravje, drugo, Življenje + ne kategorizirani>

</p>

}

Metodologija {

Korak 01 Standardni postopek
predprocesiranja

Korak 02 Dodatno CLASSLA predprocesiranje

Korak 03 Klastiranje in re-kategorizacija

Korak 04 Poganjanje različnih modelov

}

Pogostost besed –

‘Predstavitev’

Asistenca brez primera

Beseda	Pojavnost
xd	54736
pt	53943
span	48087
style	44889
font	37862
div	32196
nbsp	24365
msonormal	23947
class	23396
si	21680
color	20444
family	19630
border	17861
br	16365
width	16119
cm	15234
triglav	15205
size	14266
td	14043
serif	13645

Drajv

Beseda	Pojavnost
si	7150
triglav	6756
div	5398
drajv	5152
https	5017
com	4988
outlook	3299
protection	3230
eur	3208
safelinks	3170
http	3159
url	3154
data	3127
odrajv	2995
ceefdcbbfdee	2949
img	2927
cunknown	2894
sdata	2790
reserved	2710
xd	2425

Metodologija {

Korak 01 Standardni postopek
predprocesiranja

Korak 02 Dodatno CLASSLA predprocesiranje

Korak 03 Klastiranje in re-kategorizacija

Korak 04 Poganjanje različnih modelov

}

Pogostost besed – ‘CLASSLA’

Asistenca brez primera

Beseda	Pojavnost
vozilo	8374
cenitev	6872
black	6022
naloga	5111
stranka	4147
asistenčen	3794
kalkulacija	3764
primer	3434
storitev	3259
nadomesten	3059
avtovleka	3024
izvedba	2923
ura	2890
potnislog	2857
datum	2850
cesta	2671
izdaja	2610
pošiljati	2469
amzs	2368
lokacija	2287

Drajv

Beseda	Pojavnost
drajv	4676
vožnja	3785
aplikacija	3523
cdrajv	2948
eglasnik	1747
popust	1510
koda	1272
imeti	1196
prejeti	1062
uporaba	1034
mesec	878
nov	864
telefon	814
varen	672
prípis	630
podatek	628
pogoj	599
uporabnik	582
obvestiti	563
pravilno	549

Metodologija {

Korak 01 Standardni postopek
predprocesiranja

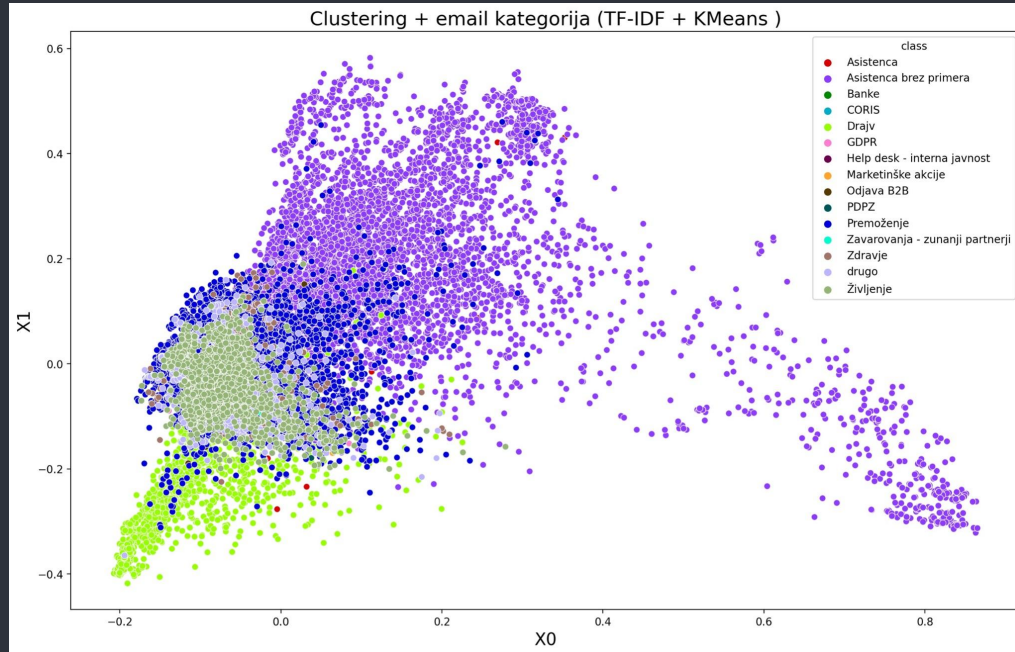
Korak 02 Dodatno CLASSLA predprocesiranje

Korak 03 Klastriiranje in re-kategorizacija

Korak 04 Poganjanje različnih modelov

}

Klastriranje 'TF-IDF + KMeans'



Clustiranje (TF-IDF + KMeans)

Cluster 0	Cluster 1	Cluster 2 (asistenca)	Cluster 3	Cluster 4 (drajv)
podpis	tzavarovalnica	prevoz	prostiti	preveriti
deloven	obvestilo	gm	server	ometje
nalog	oseba	nadomesten	pošta	pravilno
underlina	elektronski	pošljati	leto	težava
račun	podpis	cenitev	račun	danes
škoden	matičen	dan	zavarovanje	uporaba
dokumentacija	okrožen	datum	sklenjen	napaka
potrditev	osnoven	poslati	zavarovalen	mese
obrazec	pravica	priloga	polica	telefon
zahtev	kapital	strošek	dan	koda
prijava	sodišče	najem	poslati	popust
škoda	ljubljana	avtovleka	stranka	cdrajv
priponka	pridržan	popravek	odgovor	apokazija
priloga	dokumentacija	vozilo	števila	vožnja
pošljati	zavarovalen	kalkulacija	zanimati	drajv
Cluster 5 (zdravje)	Cluster 6	Cluster 7 (asistenca)	Cluster 8 (asistenca)	Cluster 9
izjava	varstvo	pomotoma	kalkulacija	stranka
možen	vsebovati	priloga	škoda	odgovoriti
kratek	contents	pošljati	car	avto
January	zaupen	servis	prevzem	ind
zdravstven	prespevden	cenitev	stranka	odsoten
strokov	strictly	prevoz	vozilo	madam
podpora	razkritje	podaljšanje	naloga	target
služba	privileg	stranka	asistenčen	sir
stranka	kopiranje	asistenčen	izdan	vozilo
posredovati	namenjen	avtovleka	tip	out
oseben	obvestiti	izvedba	znamka	avtovleka
prijava	naslovnik	storitev	datum	dear
sprememba	use	naloga	lokacija	office
oznaka	vsebina	nadomesten	ura	best
obrazec	pošiljatelj	vozilo	izdaja	cenitev

Metodologija {

Korak 01 Standardni postopek
predprocesiranja

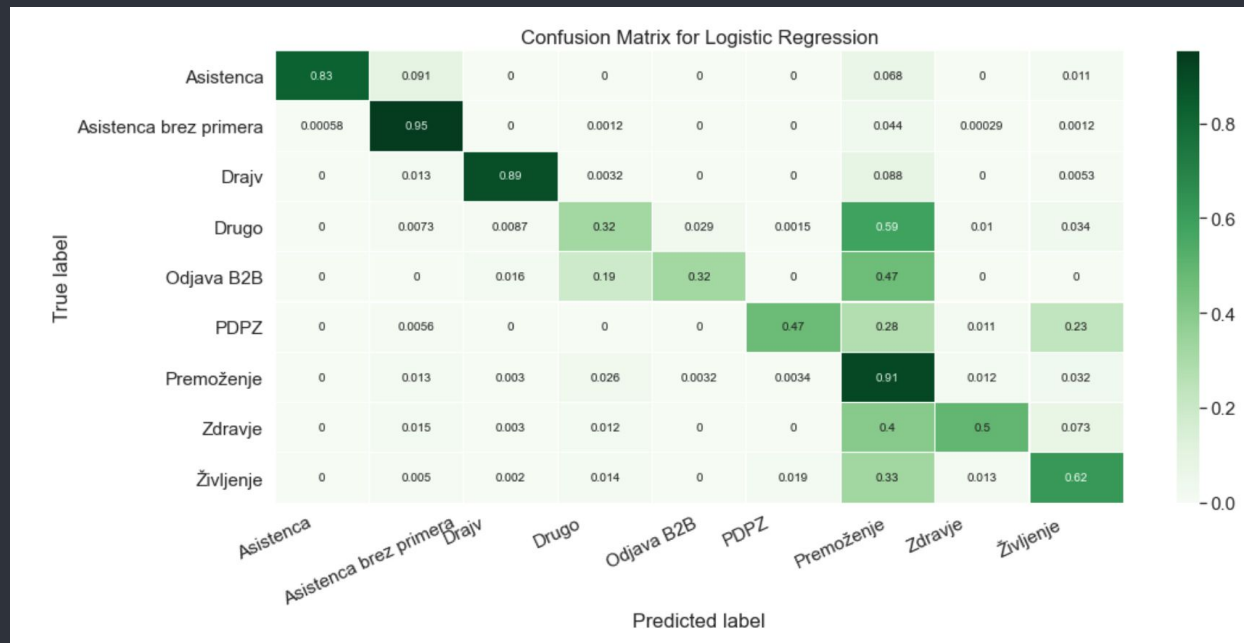
Korak 02 Dodatno CLASSLA predprocesiranje

Korak 03 Klastriiranje in re-kategorizacija

Korak 04 Poganjanje različnih modelov

}

Rezultati modela 'LogisticRegression'



Diskusija

	precision	recall	f1-score	support
Asistenca	0.97	0.83	0.90	88
Asistenca brez primera	0.97	0.95	0.96	3438
Drajv	0.97	0.89	0.93	937
Drugo	0.55	0.32	0.41	686
Odjava B2B	0.53	0.32	0.40	124
PDPZ	0.69	0.47	0.56	179
Premoženje	0.79	0.91	0.84	4997
Zdravje	0.67	0.50	0.57	331
Življenje	0.71	0.62	0.66	1001
accuracy			0.84	11781
macro avg	0.76	0.65	0.69	11781
weighted avg	0.83	0.84	0.83	11781

1
2
3
4
5
6
7
8
9
10
11
12
13
14

Hvala {
za pozornost;
|
}