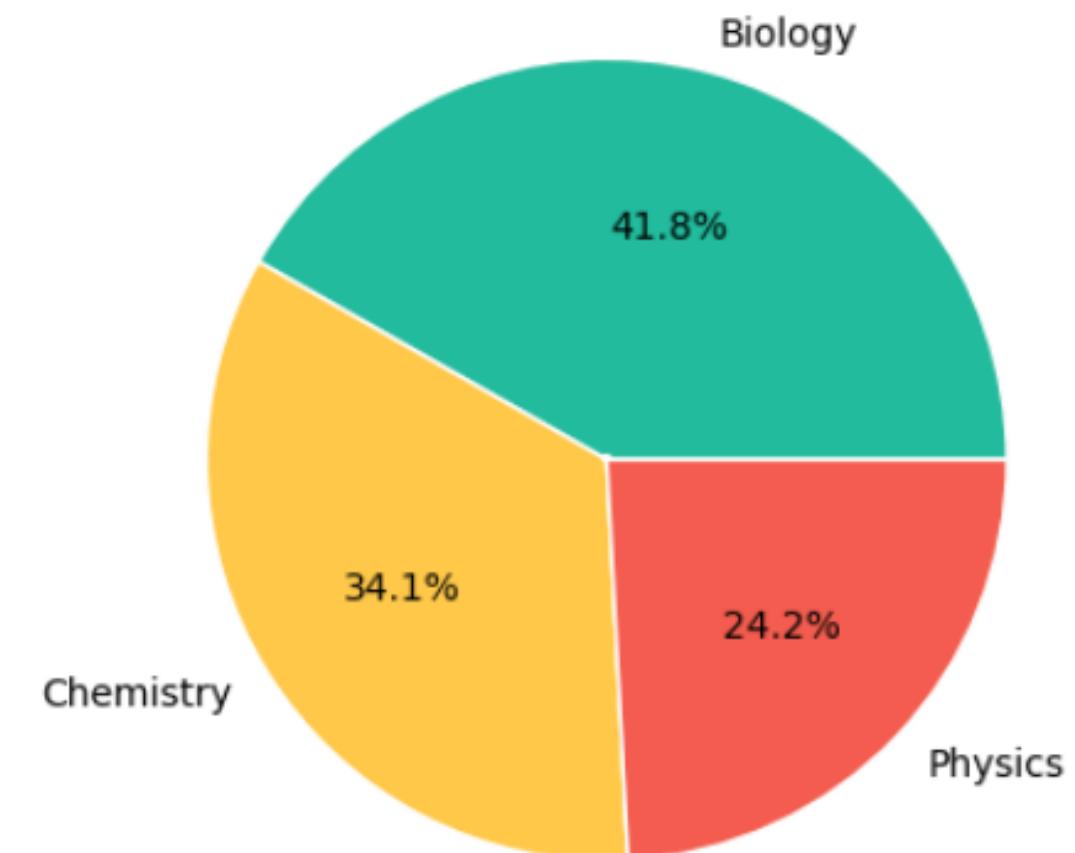
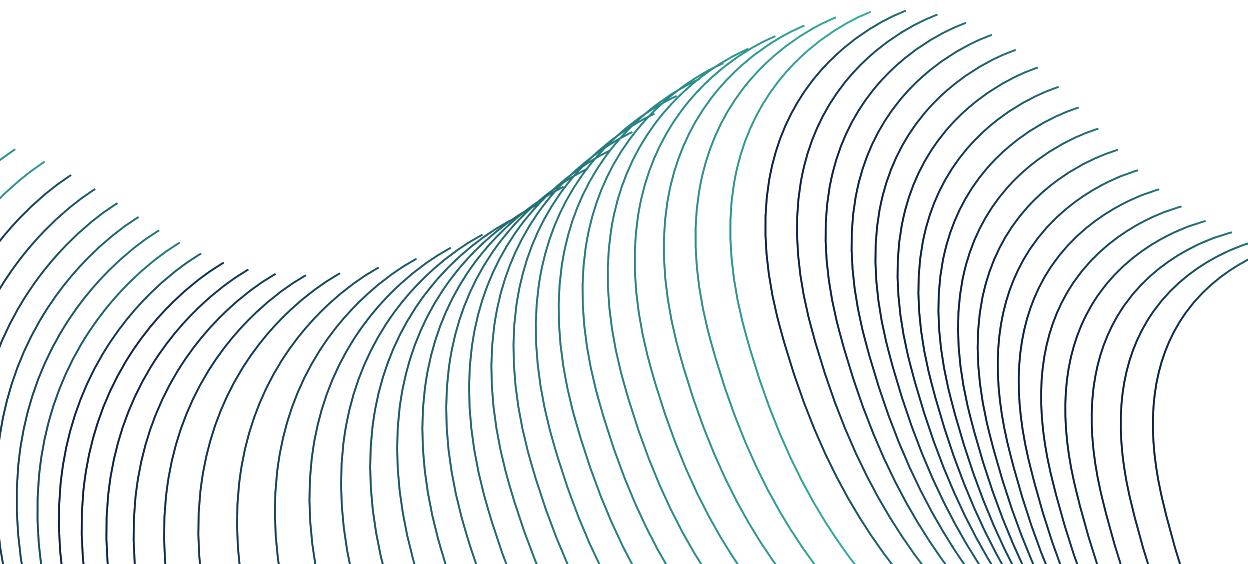


Physics vs Chemistry vs Biology

Sanja Nedeljković 166/19
Istraživanje podataka 1

Id	Comment	Topic
0 0x840	A few things. You might have negative- frequency dependent selection going on where the least common phenotype, reflected by genotype, is going to have an advantage in the environment. For instance, if a prey animal such as a vole were to have a light and a dark phenotype, a predator might recognize the more common phenotype as food. So if the light voles are more common, foxes may be keeping a closer eye out for light phenotypic voles, recognising them as good prey. This would reduce the light causing alleles due to increased predation and the dark genotypes would increase their proportion of the population until this scenario is reversed. This cycle continues perpetually. \n\nHowever, this is unlikely to be strictly yearly as it usually takes more time than a year for an entire populations allele frequencies to change enough to make a large enough difference to alter fitness. \n\nMore likely on a *year to year* basis, the population is experiencing fluctuating selection where al...	Biology
1 0xbf0	Is it so hard to believe that there exist particulars out that that we can't detect with anything we've invented so far.\n\nI mean look how long it took humans to find out a way to detect radiation.	Physics
2 0x1dfc		There are bees
3 0xc7e	I'm a medication technician. And that's alot of drugs on your liver. You probably won't die immediately you'll be fine. Take care of yourself tho that's definitely not good for your body	Biology
4 0xbba	Cesium is such a pretty metal.	Chemistry



**trening: 8695
test: 1586**

Dužine komentara

PHYSICS

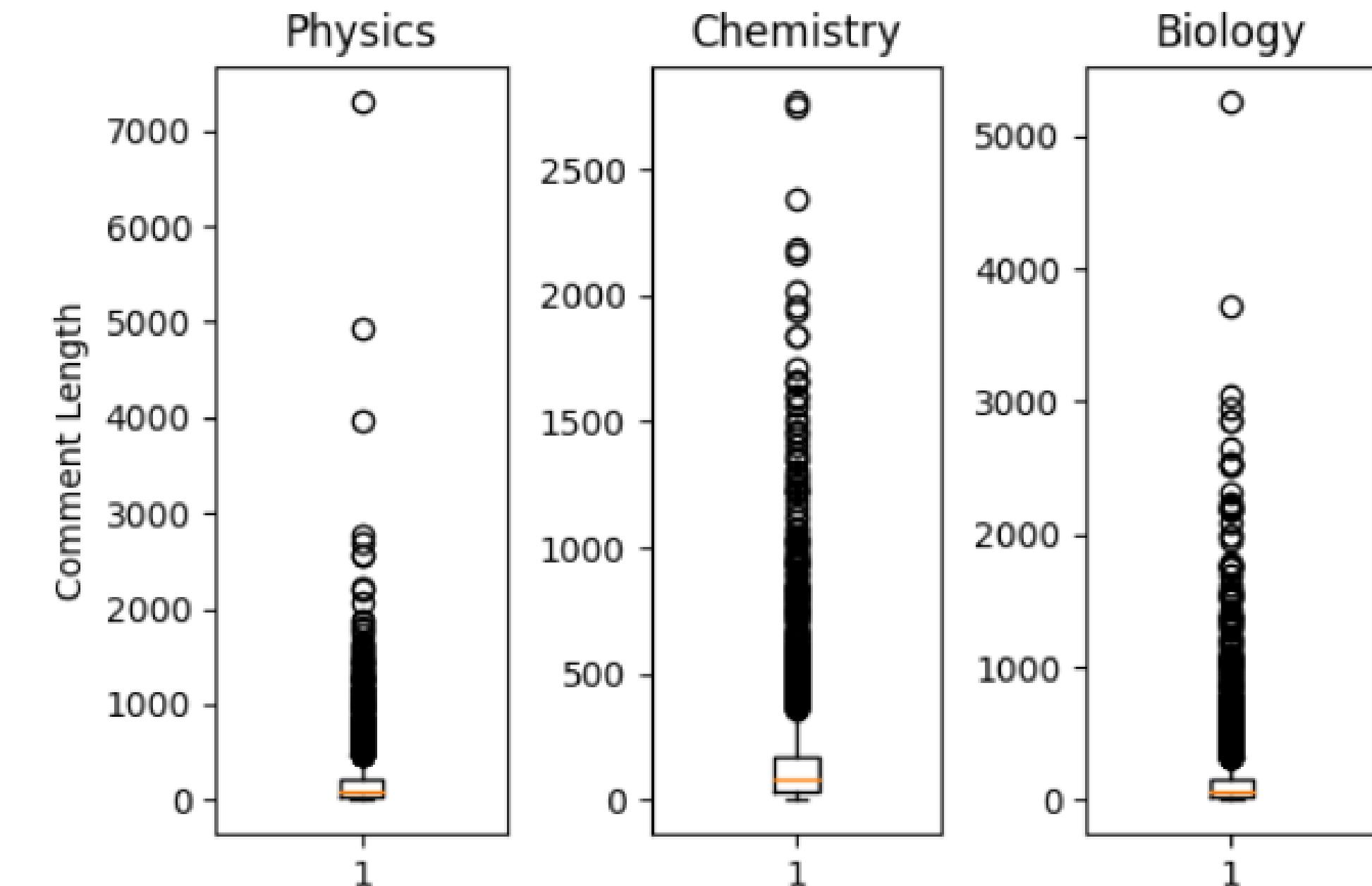
> 3000

CHEMISTRY

> 2000

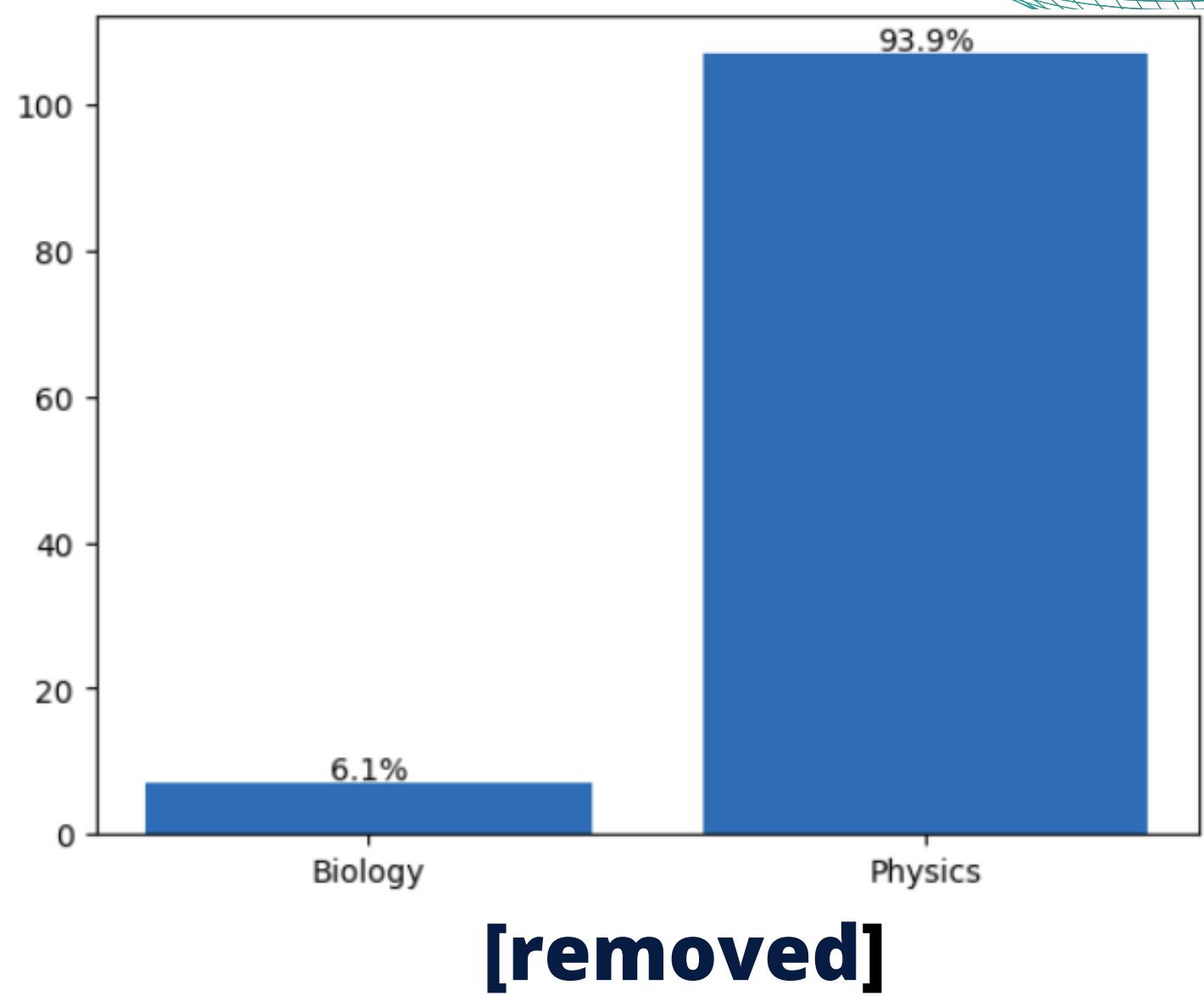
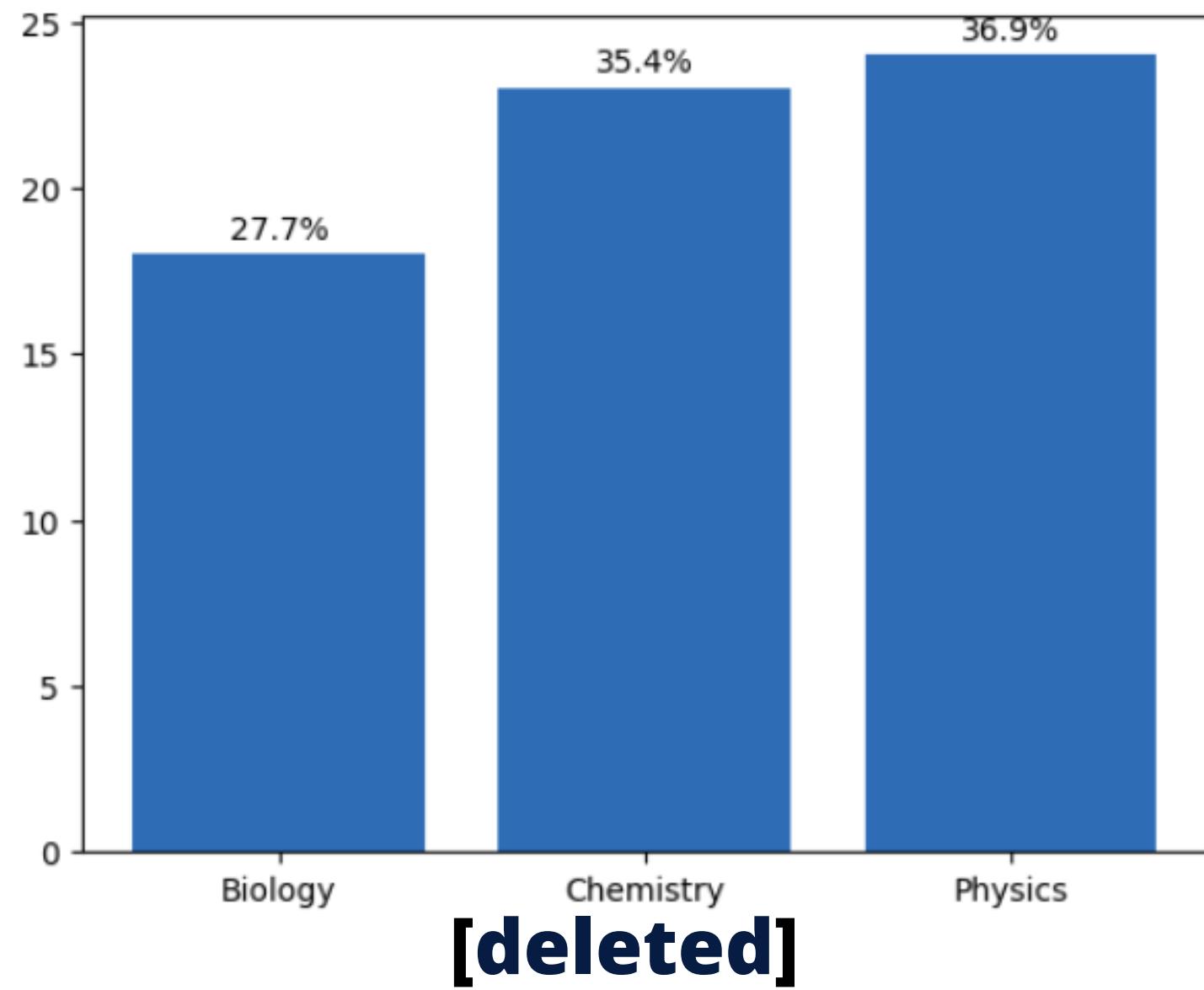
BIOLOGY

> 2000



Prisustvo “[removed]” označava da su moderatori uklonili odgovarajući komentar. Sa druge strane, kada je prisutan termin “[deleted]”, to sugerije da je korisnik dobrovoljno odlučio ga ukloni.

Nedostajuće vrednosti



1. Mala slova:
'PhYSics' → 'physics'

2. Filtriranje reči:
'physics!' → 'physics'
'chemistry, bio,' → 'chemistry bio'

3. Uzastopno ponavljanje:
'physicsss' → "

4. Tokenizacija:
'about physics' → 'about', 'physics'

5. Stop words:
'there are bees' → 'bees'

6. Lematizacija:
'run' → 'run'
'running' → 'run'

7. Minimalna dužina:
'to' → "

Preprocesiranje

1 Is it so hard to believe that there exist particulars out that that we can't detect with anything we've invented so far.\n\nI mean look how long it took humans to find out a way to detect radiation.

exist particular detect invented took human detect radiation

2

There are bees

bee

3 I'm a medication technician. And that's alot of drugs on your liver. You probably won't die immediately you'll be fine. Take care of your self tho that's definitely not good for your body

medication technician alot drug liver die immediately fine care self tho

4

Cesium is such a pretty metal.

cesium metal

5

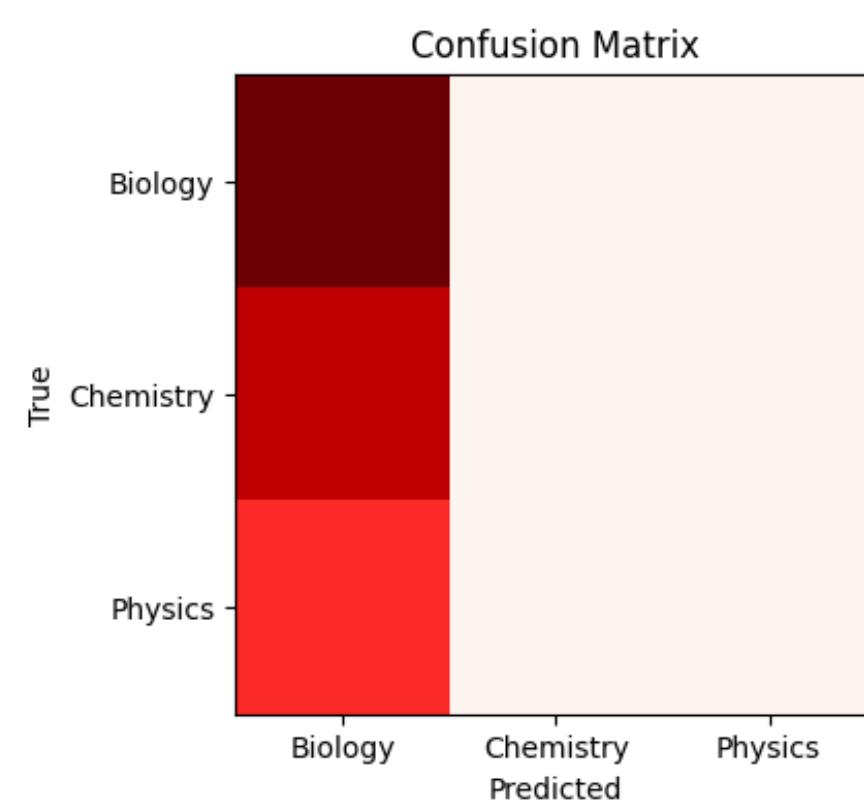
I meant that the question itself is unclear.

meant unclear

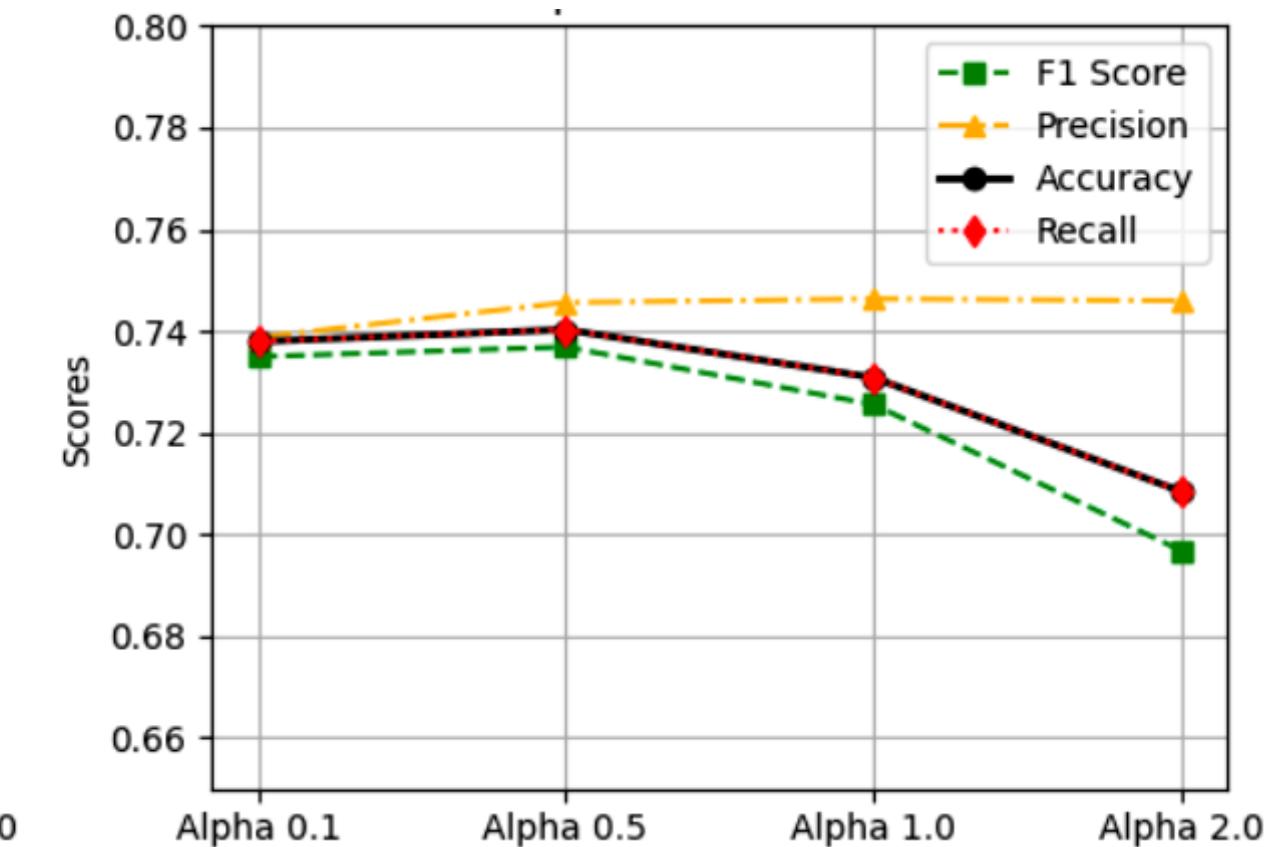
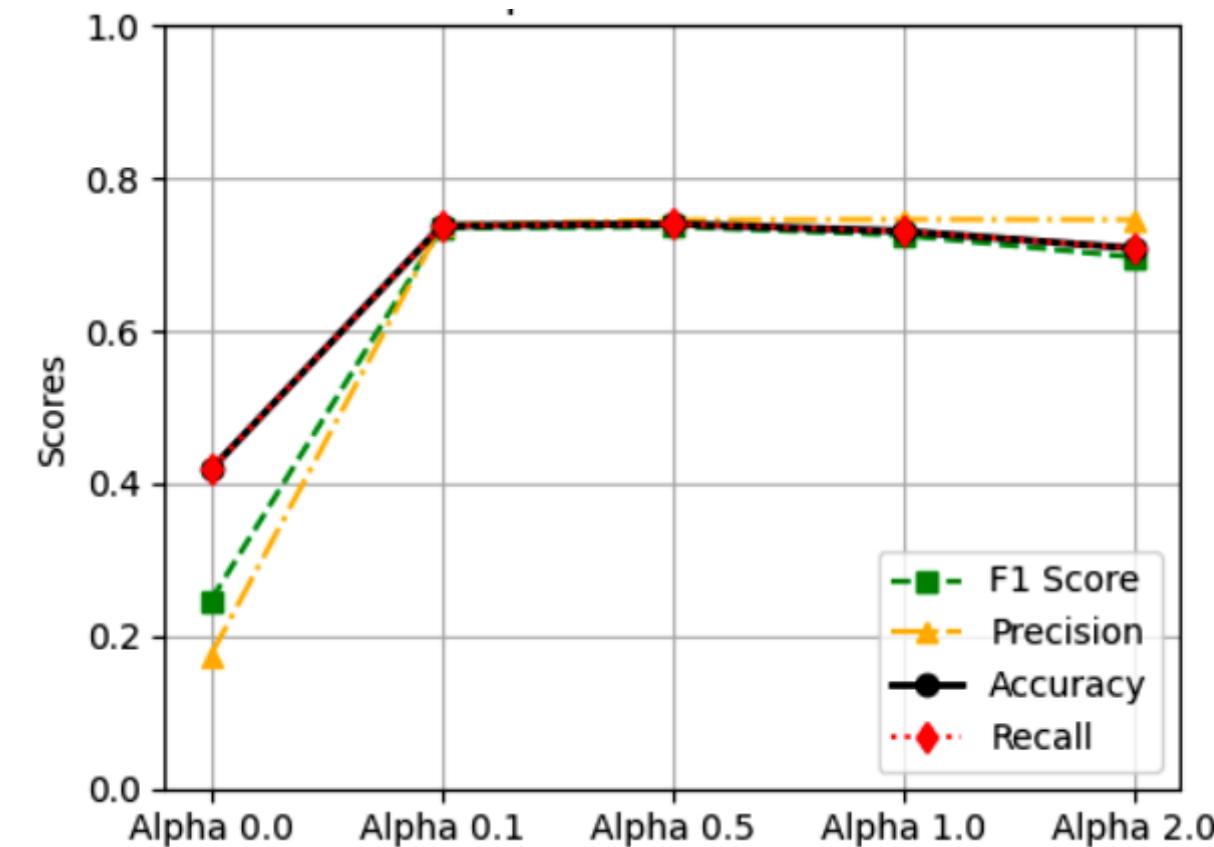
Transformacije i čuvanje podataka

Naivni Bajes

Ispitivanje performansi modela za razlicite vrednosti parametra alfa [0.0, 0.1, 0.5, 1.0, 2.0]



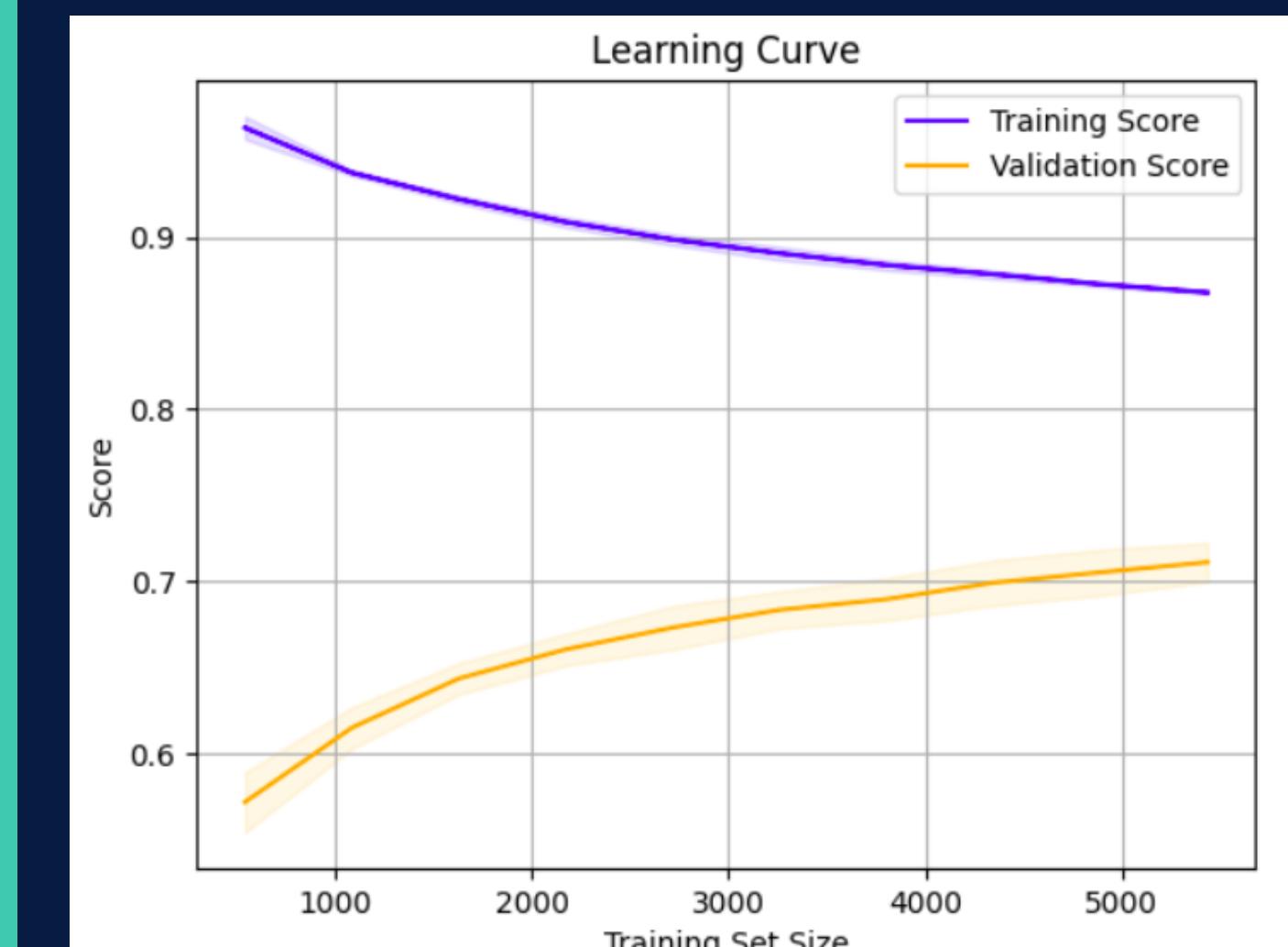
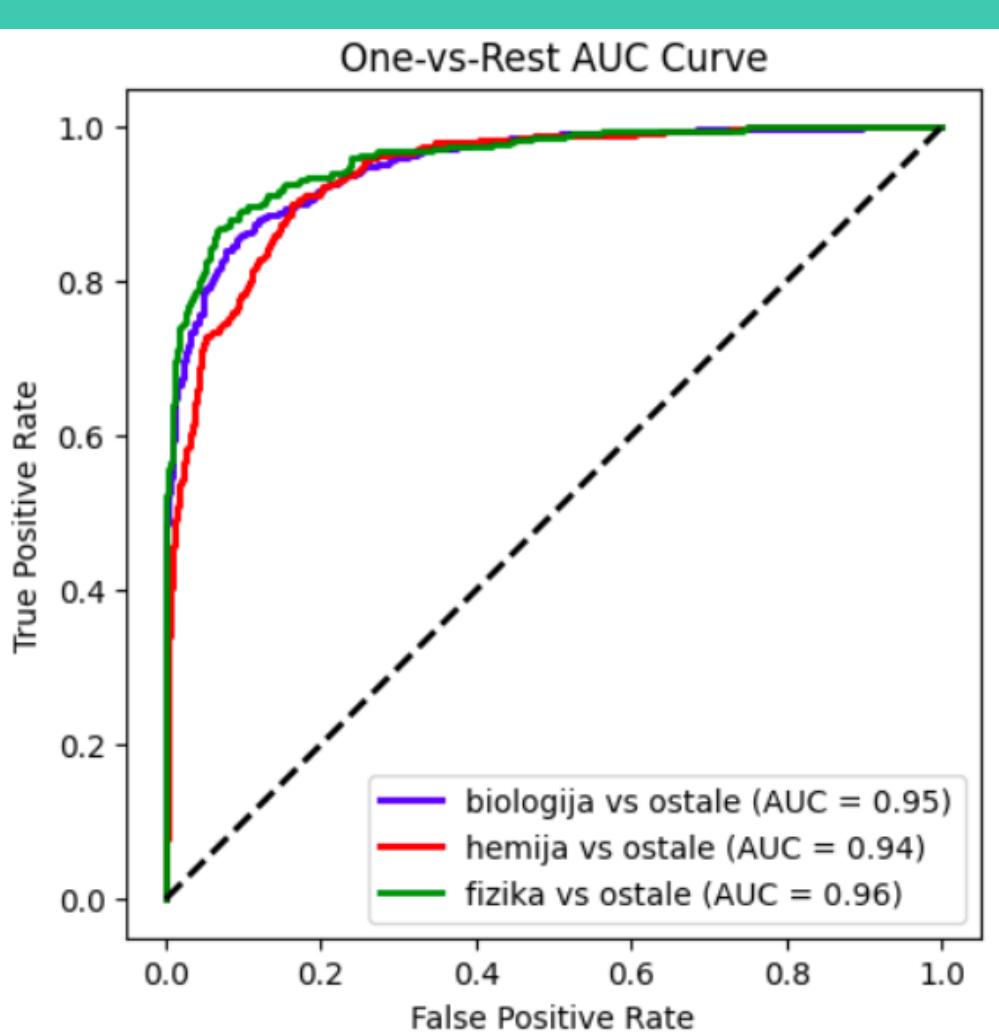
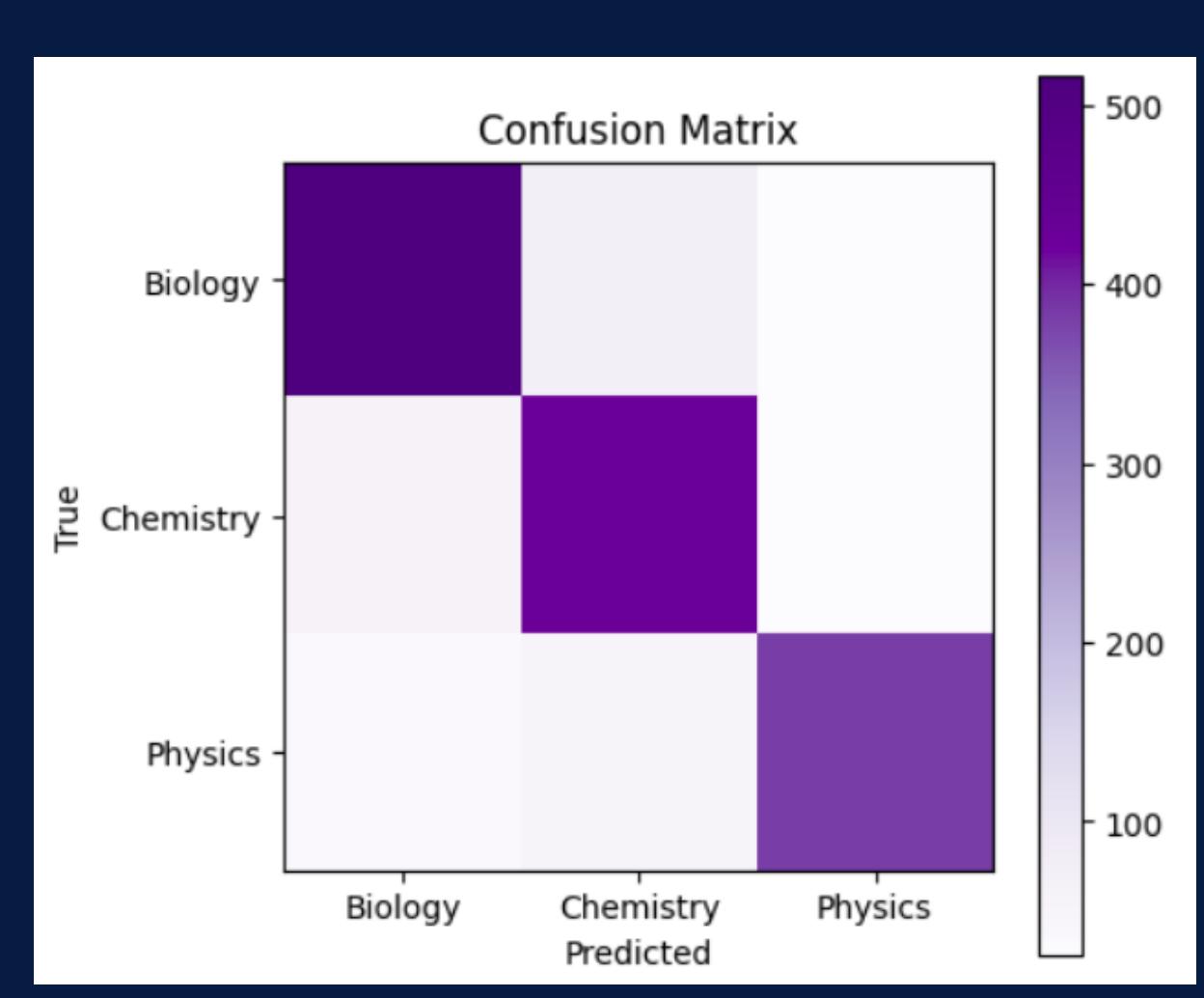
alpha = 0



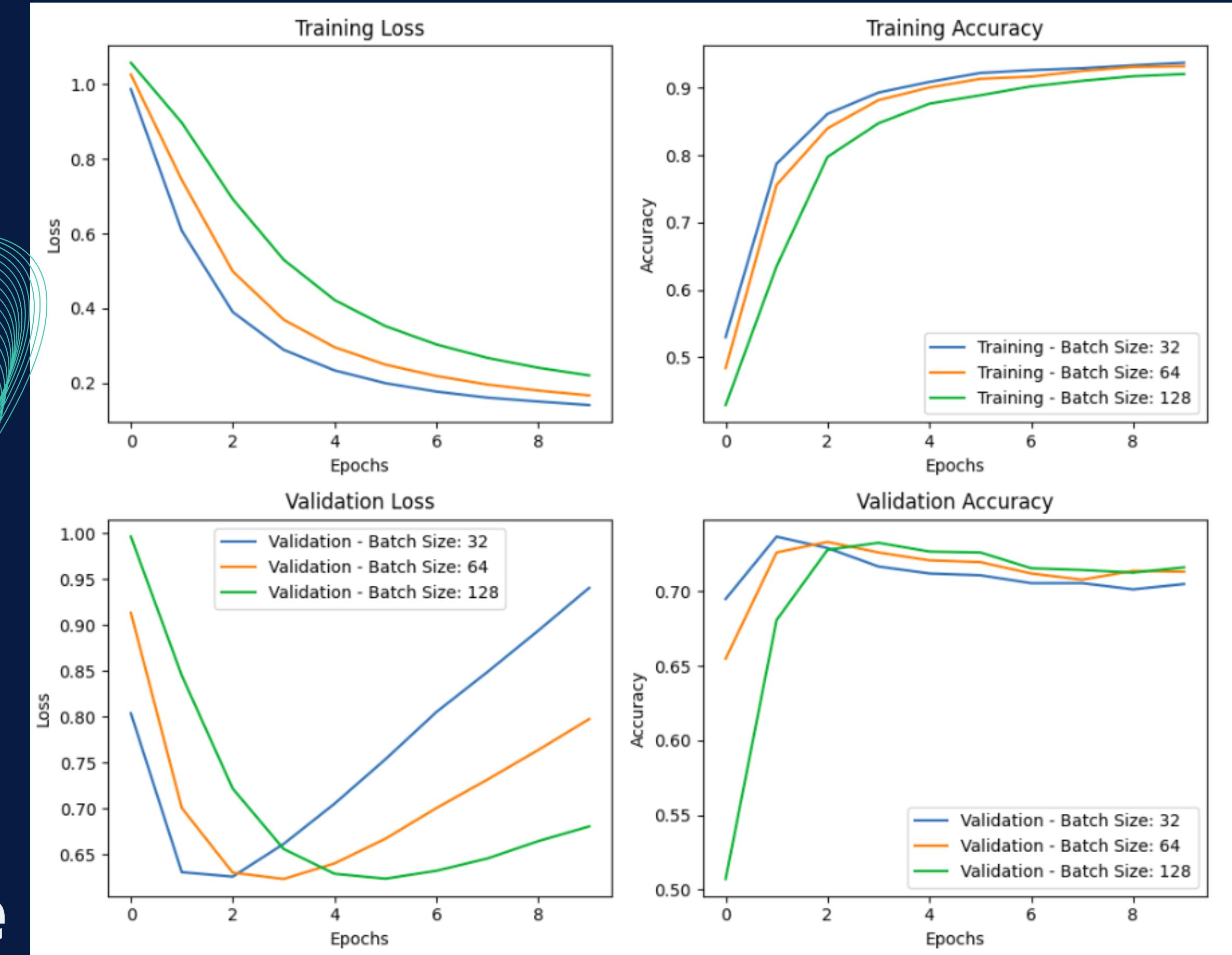
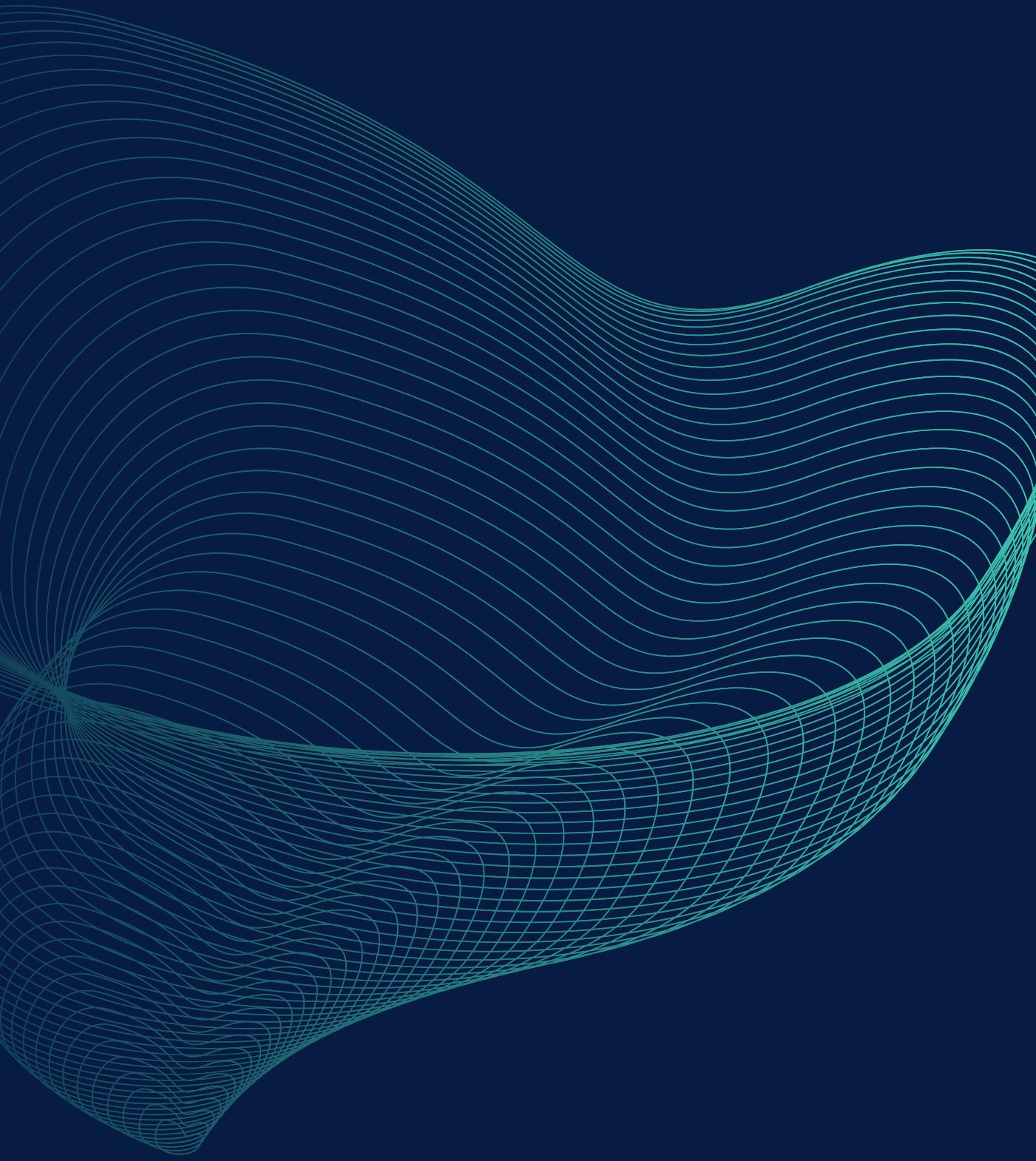
α	Preciznost	Odziv	F1-mera	Tačnost
0.1	0.74	0.74	0.73	0.74
0.5	0.75	0.74	0.74	0.74
1.0	0.75	0.73	0.73	0.73
2.0	0.75	0.71	0.70	0.71

Rezultati na test skupu

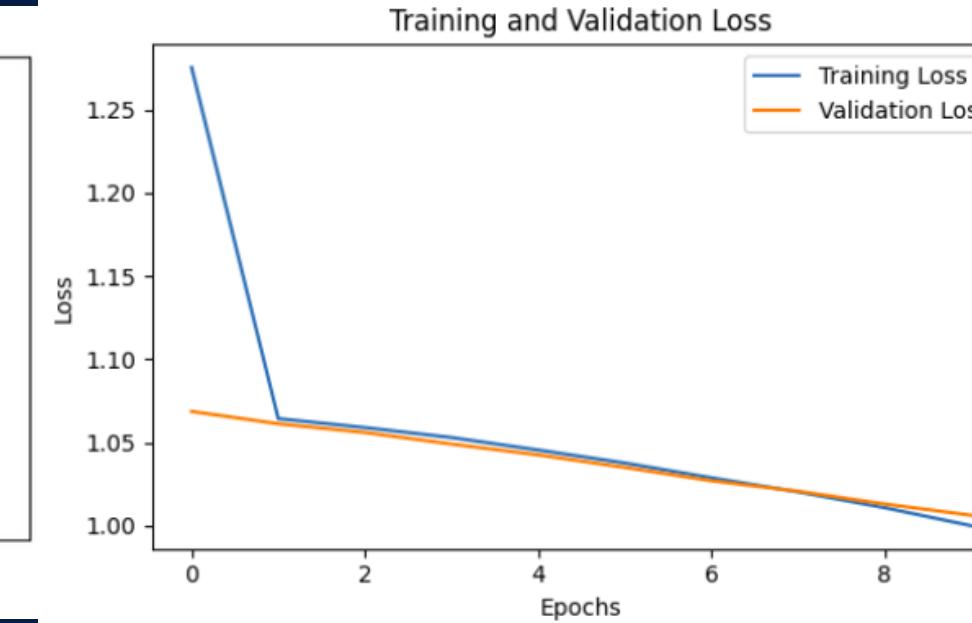
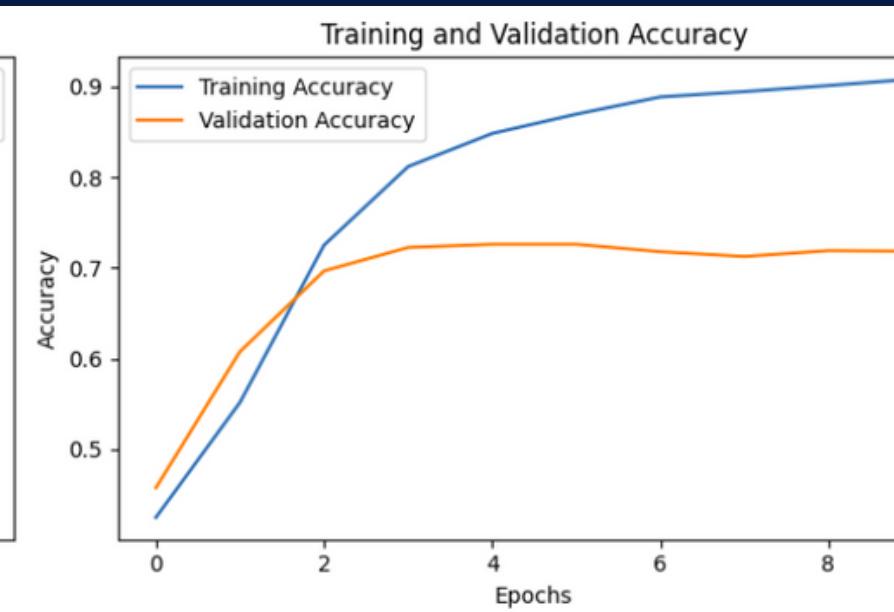
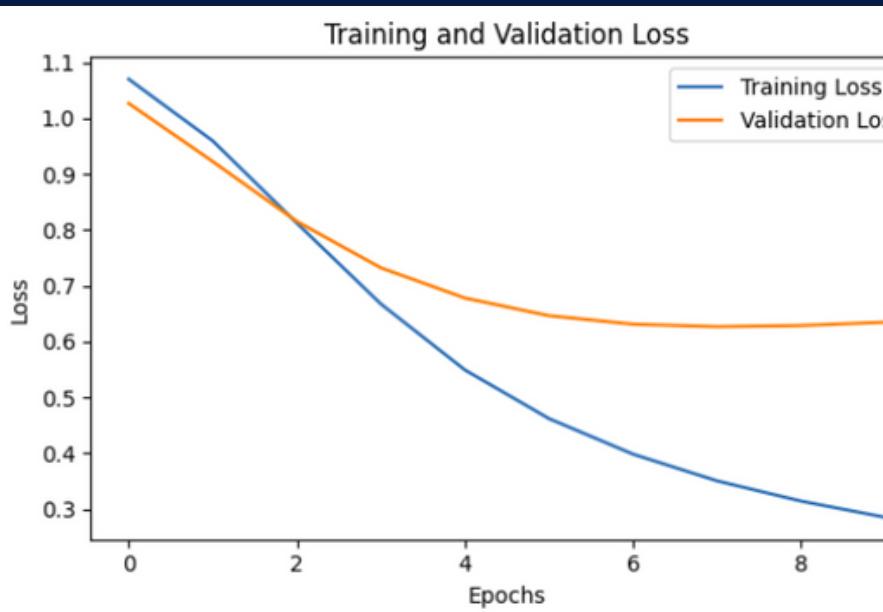
	Preciznost	Odziv	F1-mera
0	0.85	0.84	0.85
1	0.77	0.84	0.80
2	0.88	0.82	0.85
Tačnost	-	-	0.83
Makro sredina	0.84	0.83	0.83
Težinska sredina	0.84	0.83	0.83



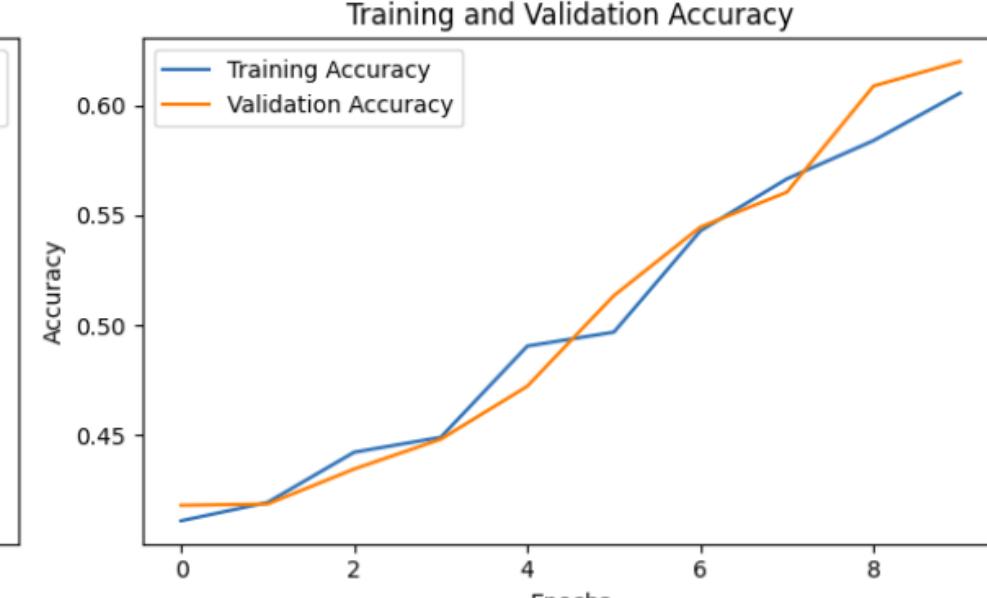
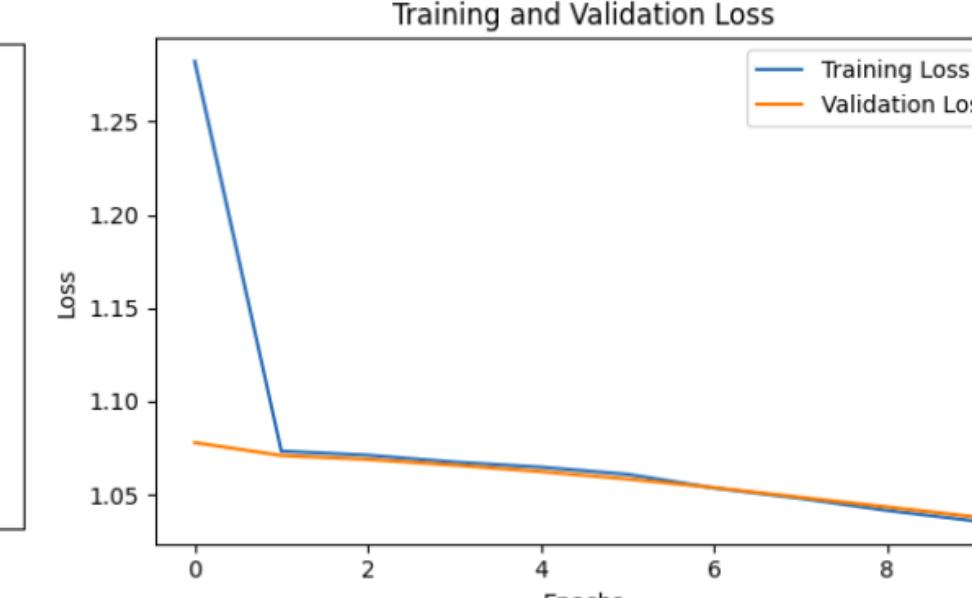
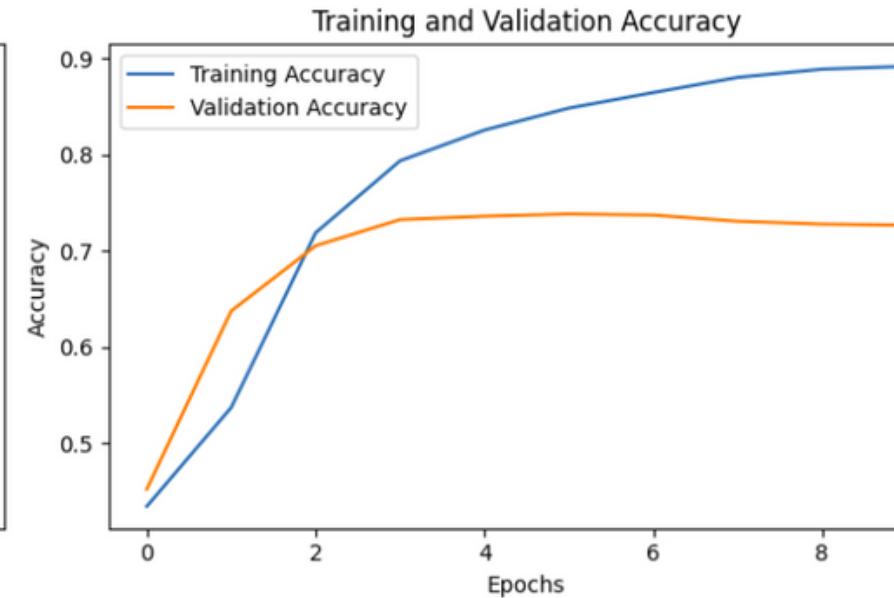
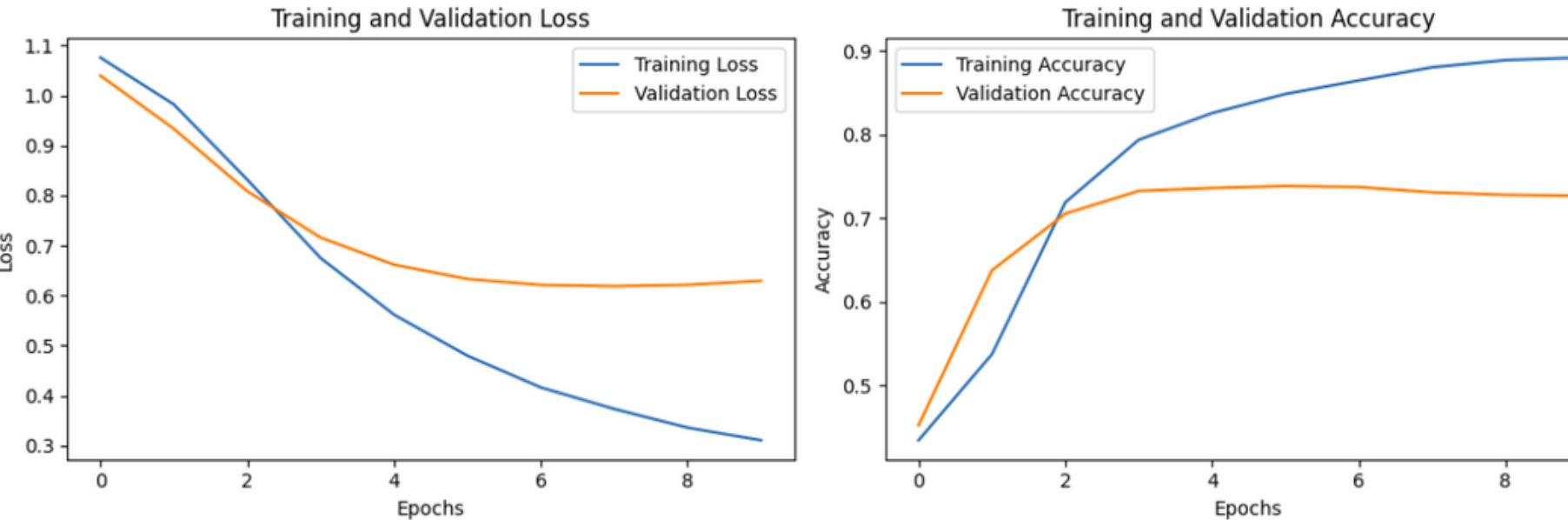
Neuronske mreže



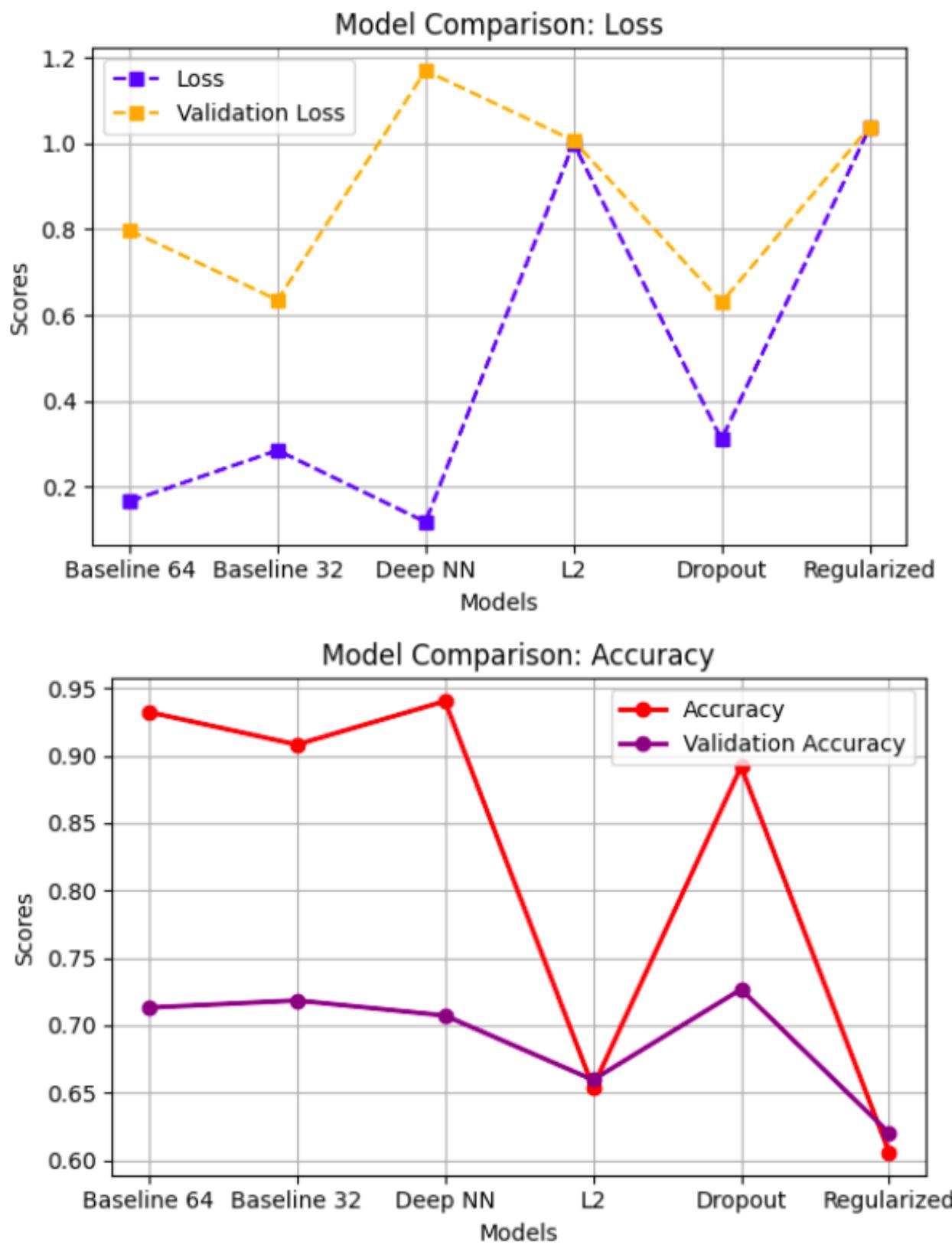
Smanjenje broja neurona



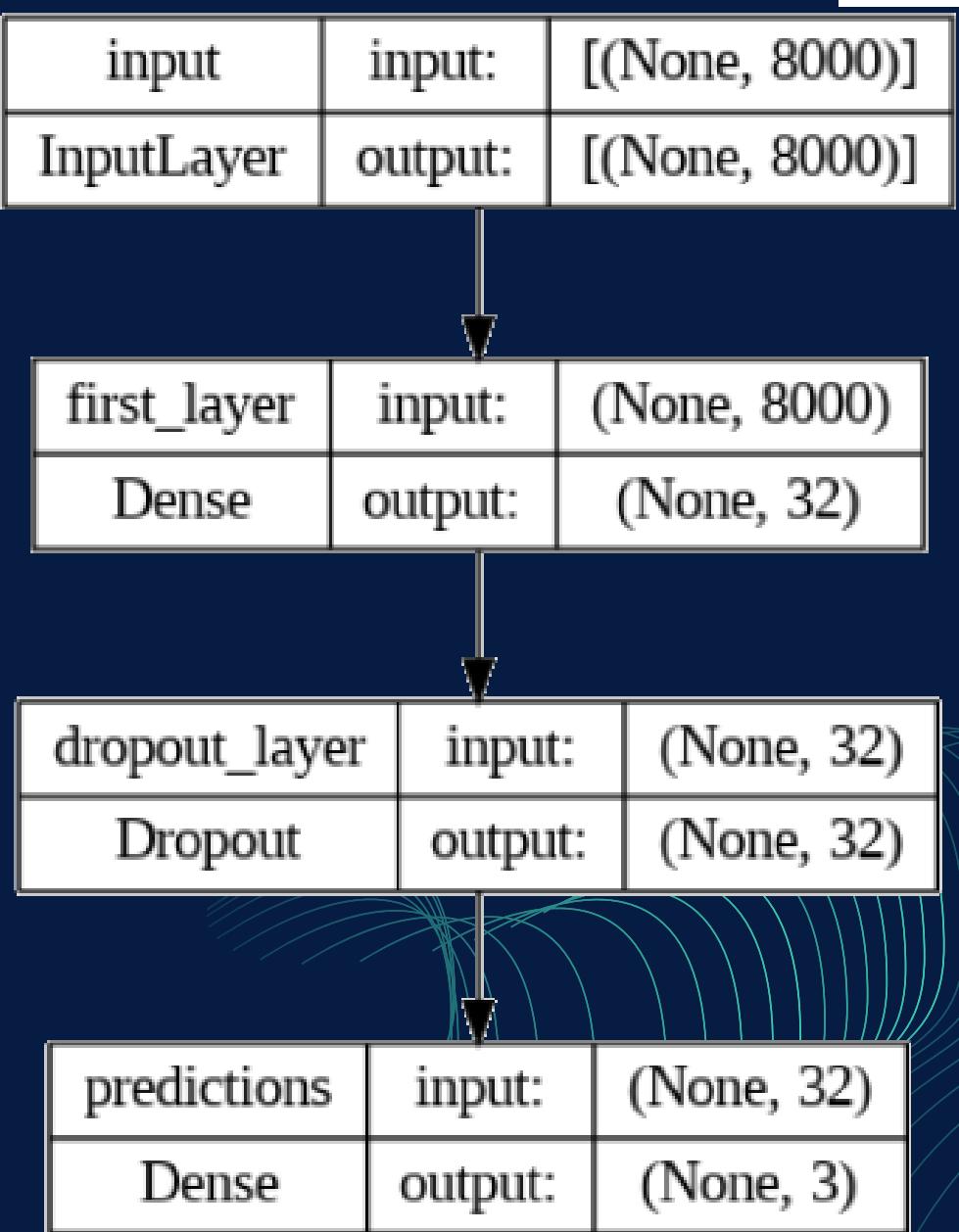
Dropout



L2 + dropout



```
50/50 - 0s - loss: 0.4742 - accuracy: 0.8380
Test loss: 0.474152535200119
Test accuracy: 0.8379571437835693
```



K sredina – 3 klastera

Cluster 0

phd magic youtube degree cream ice mushroom chemistry physic look

Cluster 1

effect research year sense solution science electron matter make thing

Cluster 2

sidebar hell physicsmemes mod username question explain stop check please

Biology

Chemistry

Physics

Biology

Physics

38.7%

30.0%

31.4%

Chemistry

Biology

Physics

42.2%

24.5%

33.3%

Chemistry

Biology

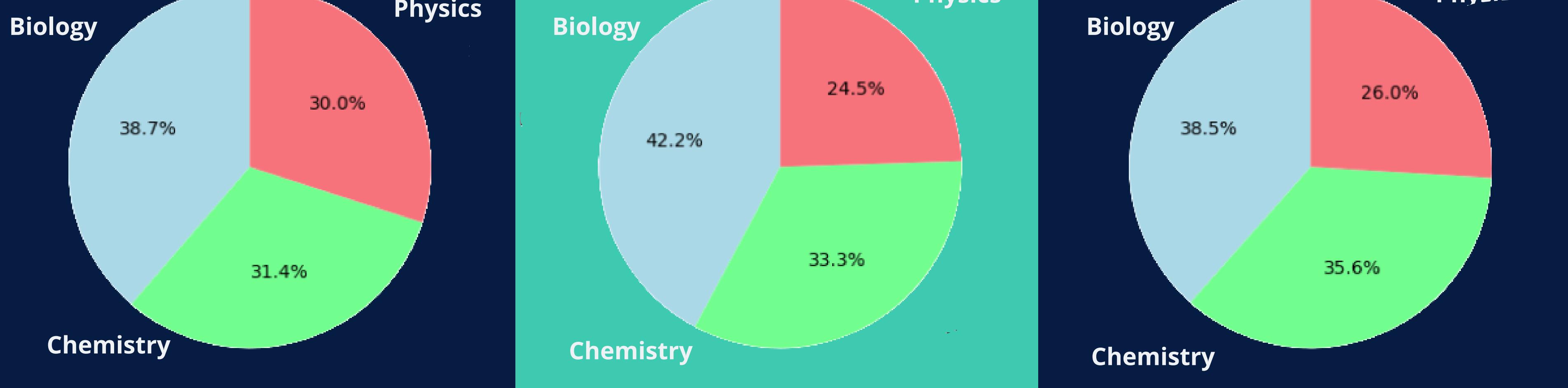
38.5%

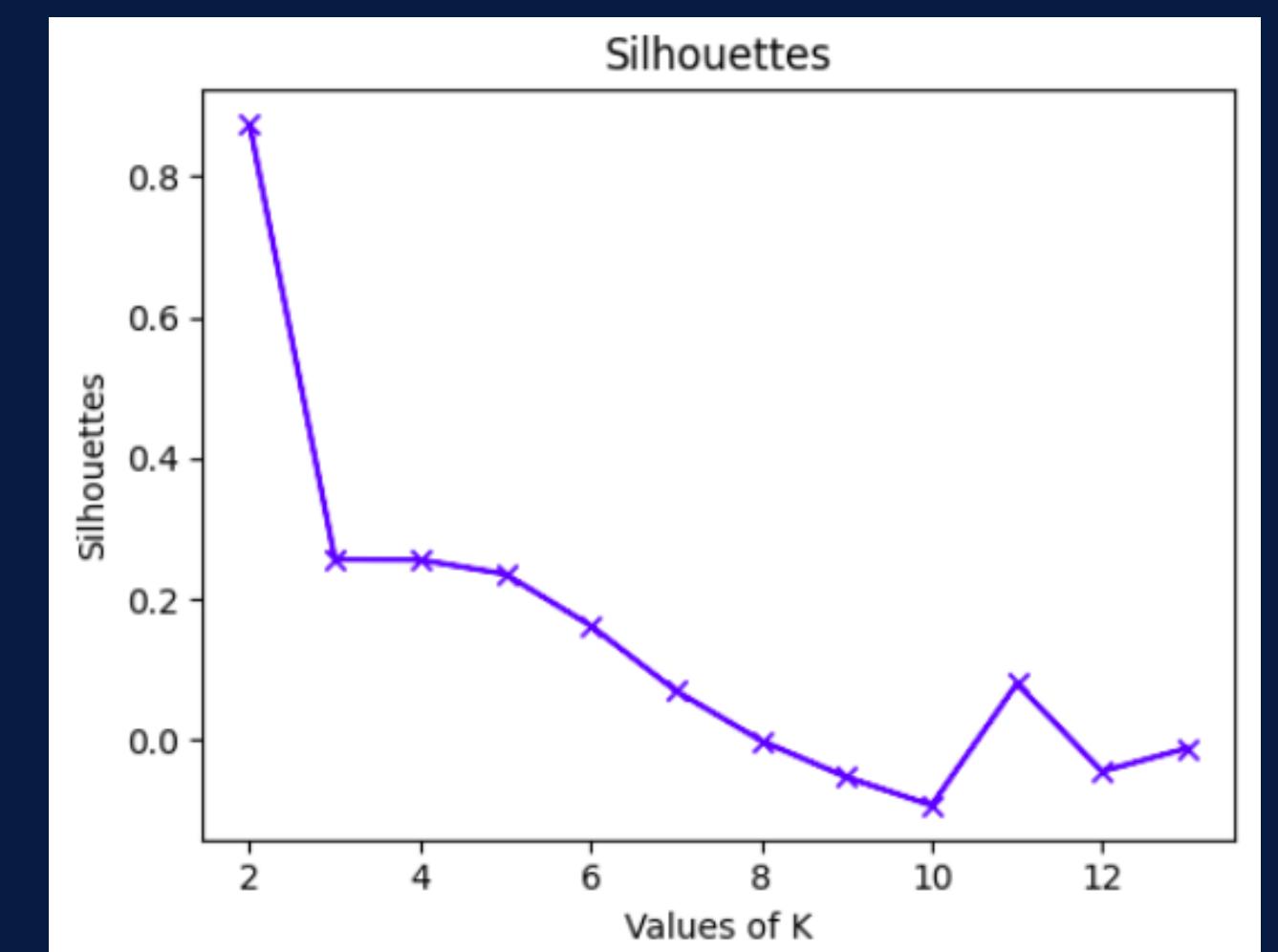
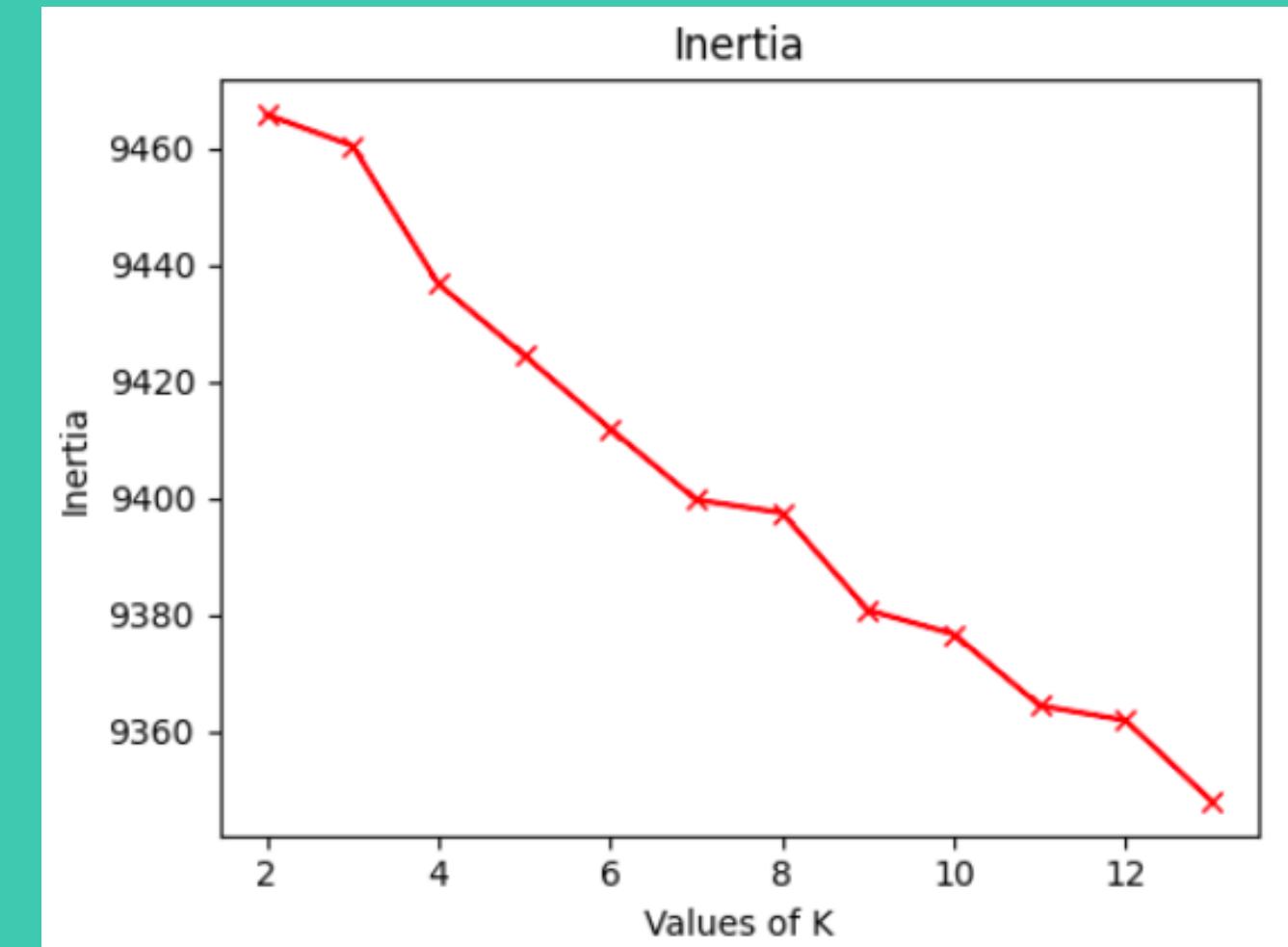
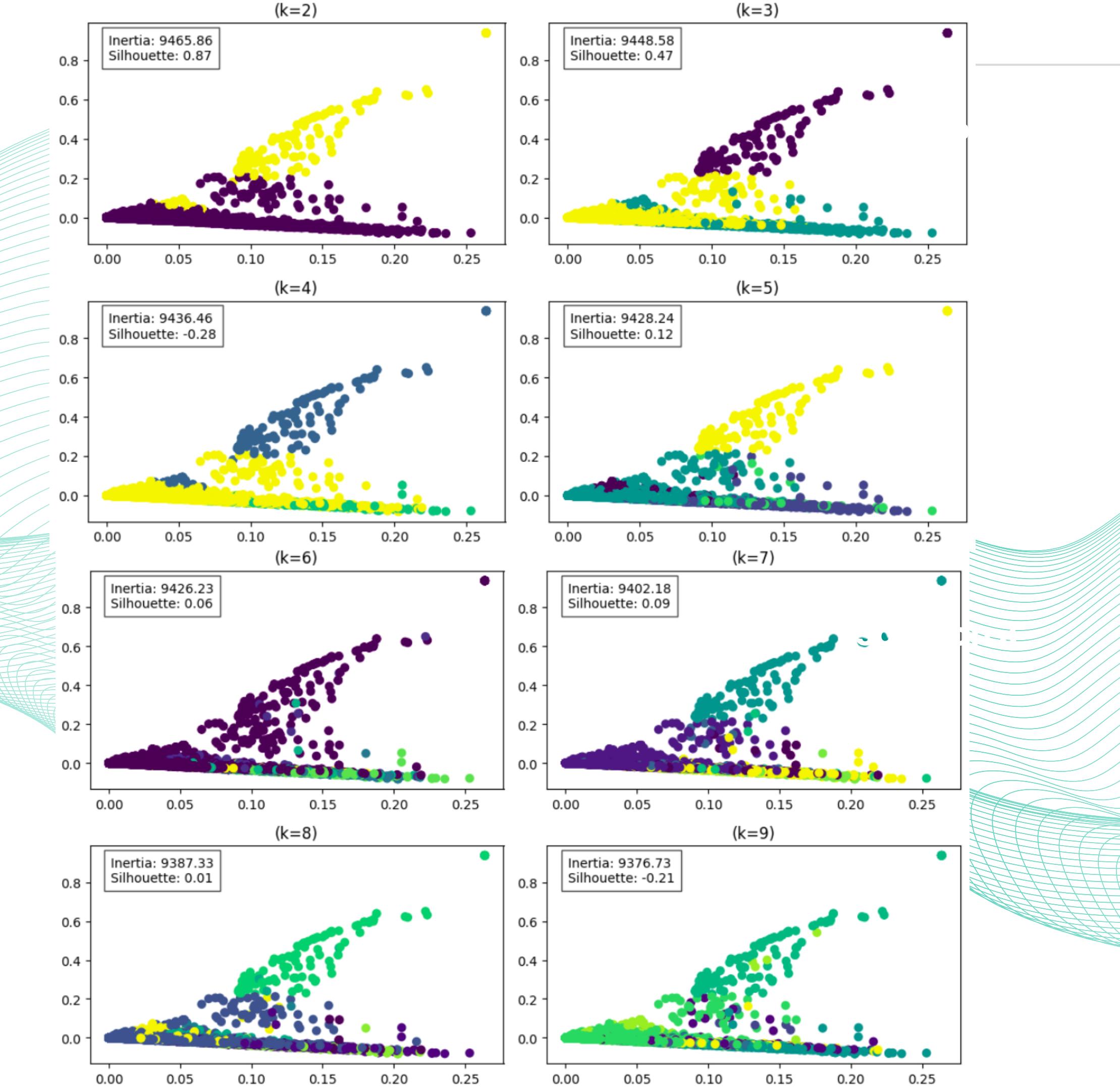
26.0%

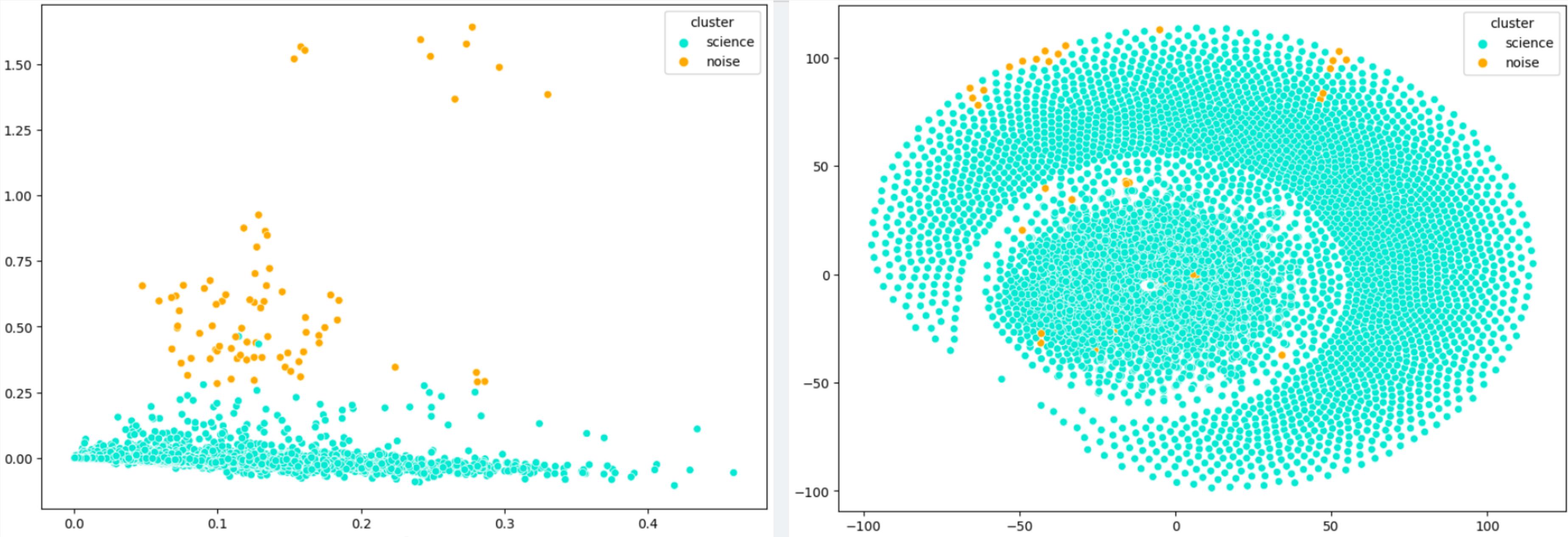
35.6%

Chemistry

Physics







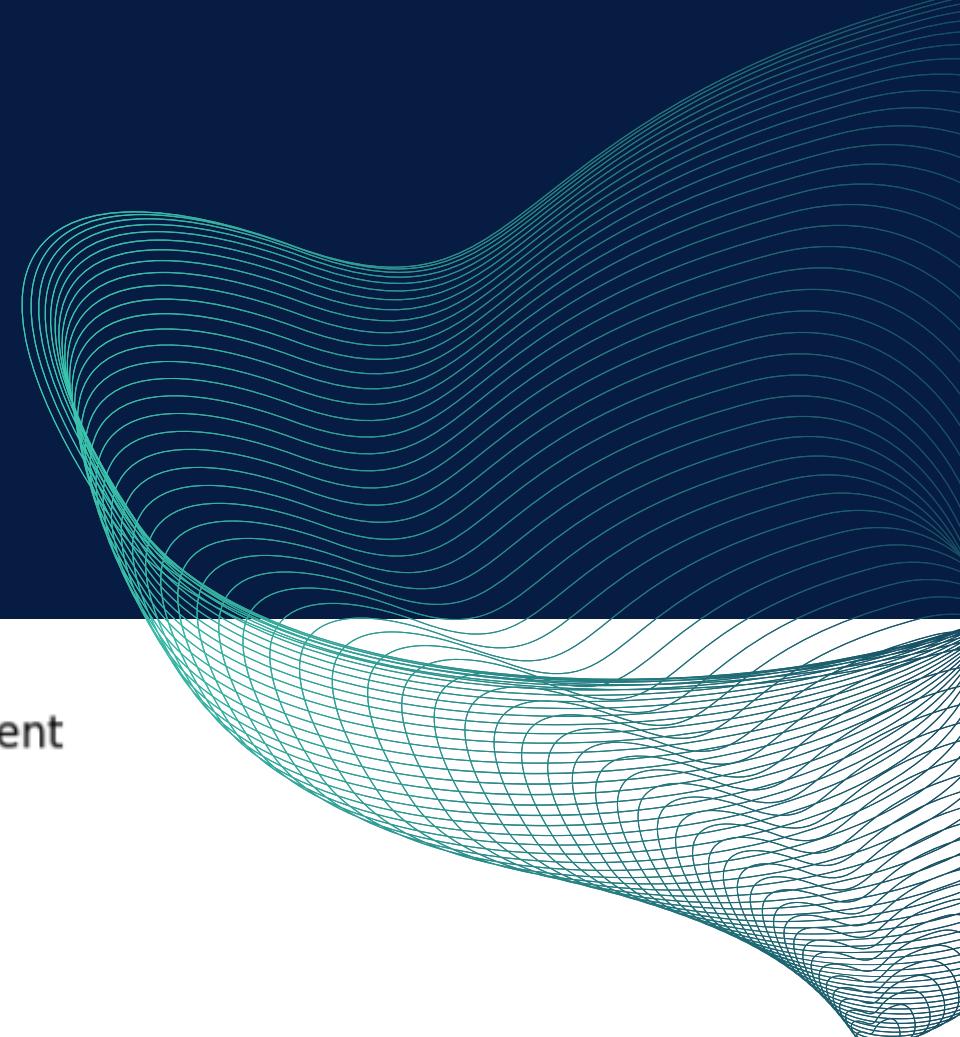
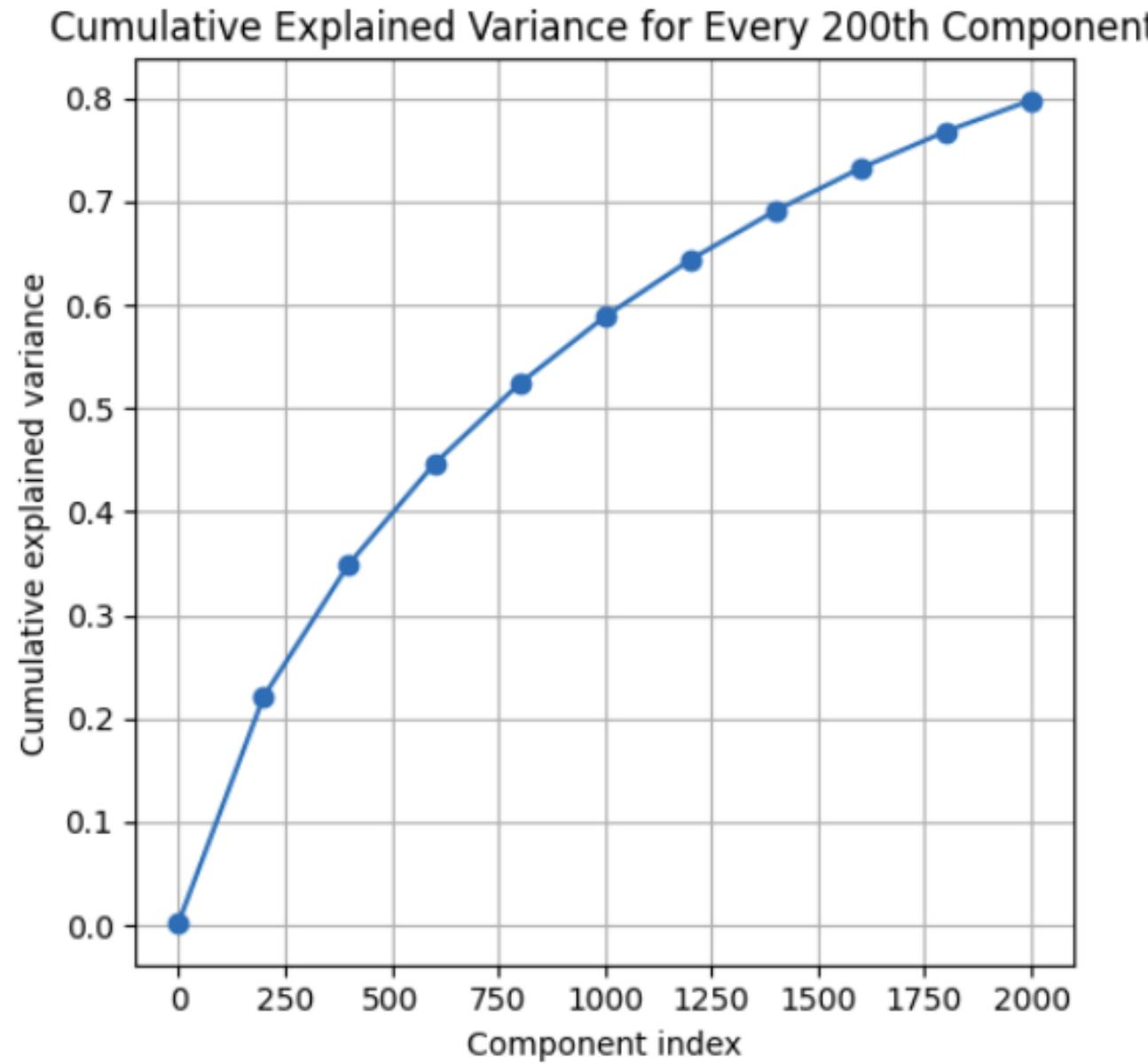
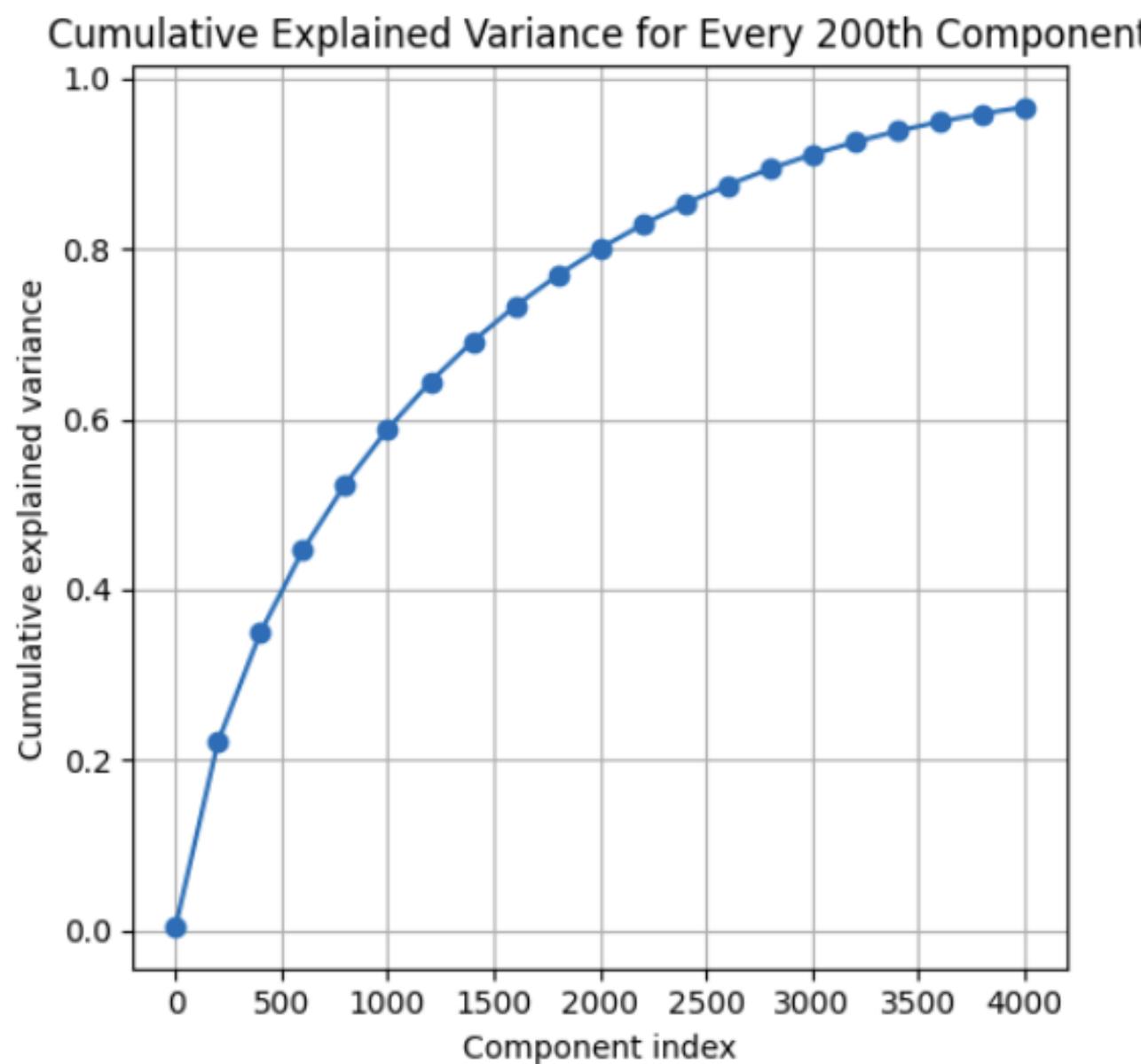
Normalizovani podaci

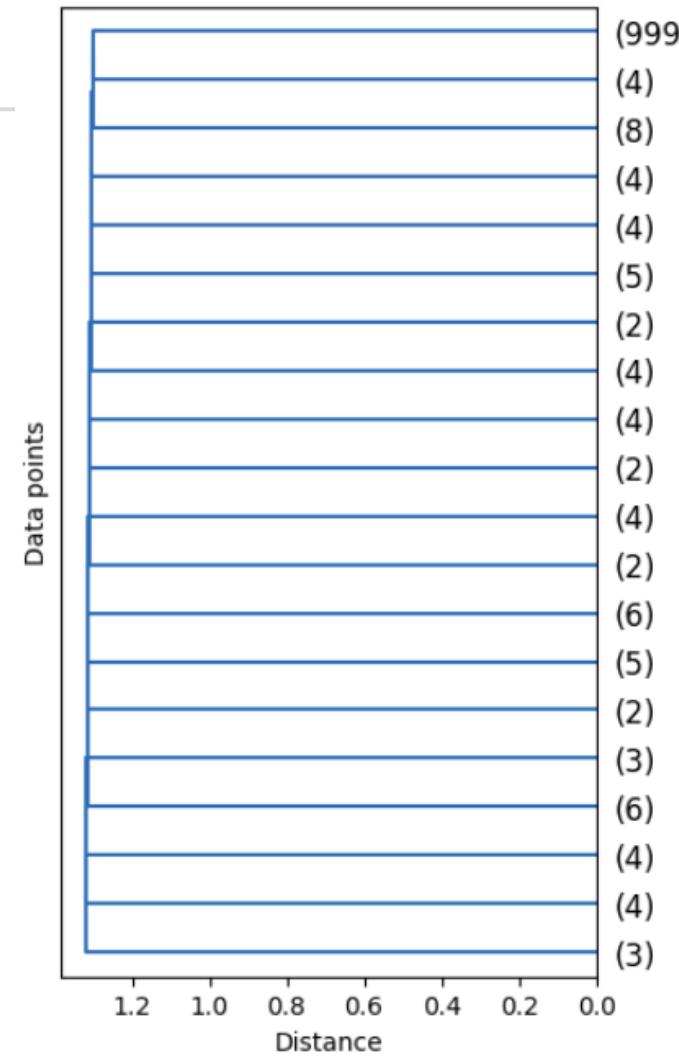
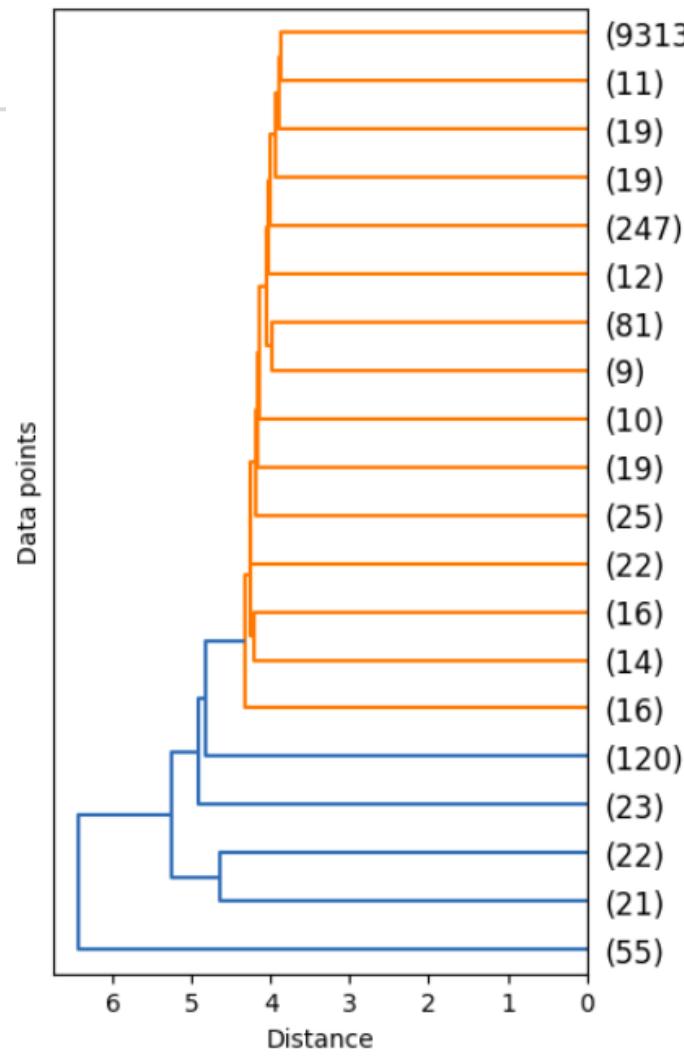
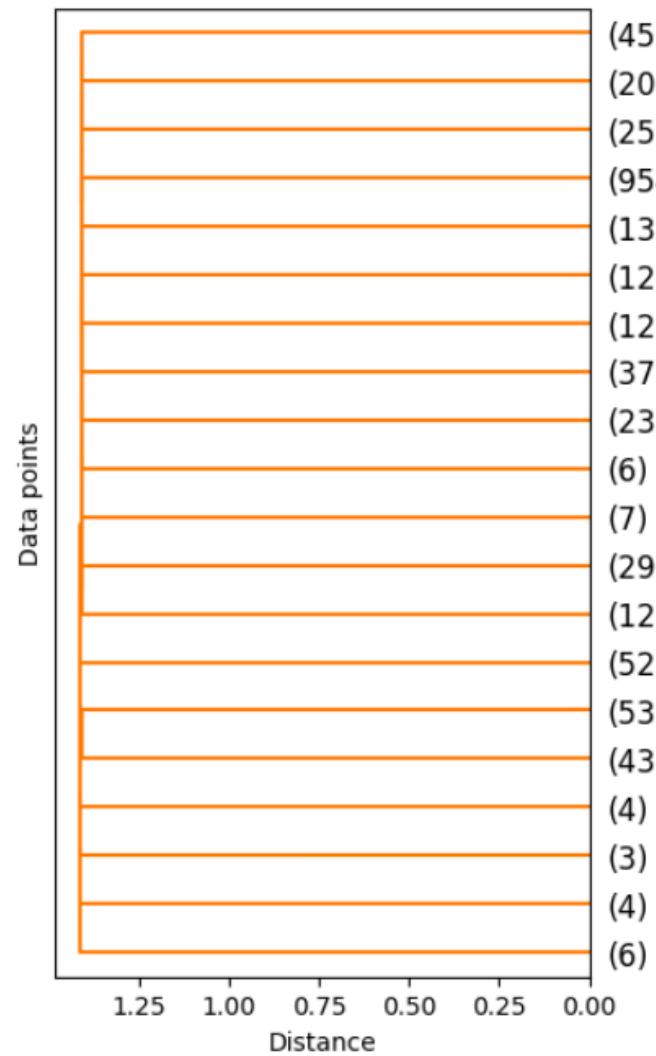
Cluster 0
message github compose sujal wikisummarizer wikisummarizerbot optout opt wikipedia wiki

Cluster 1
state reaction level particle mass chemistry thing make physic look

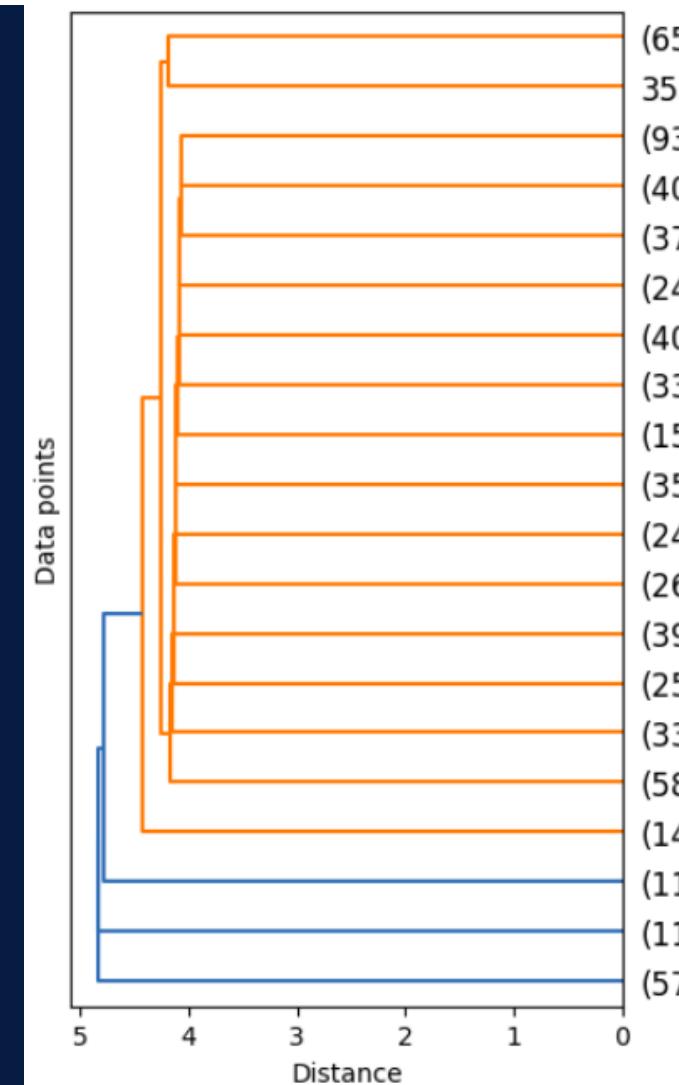
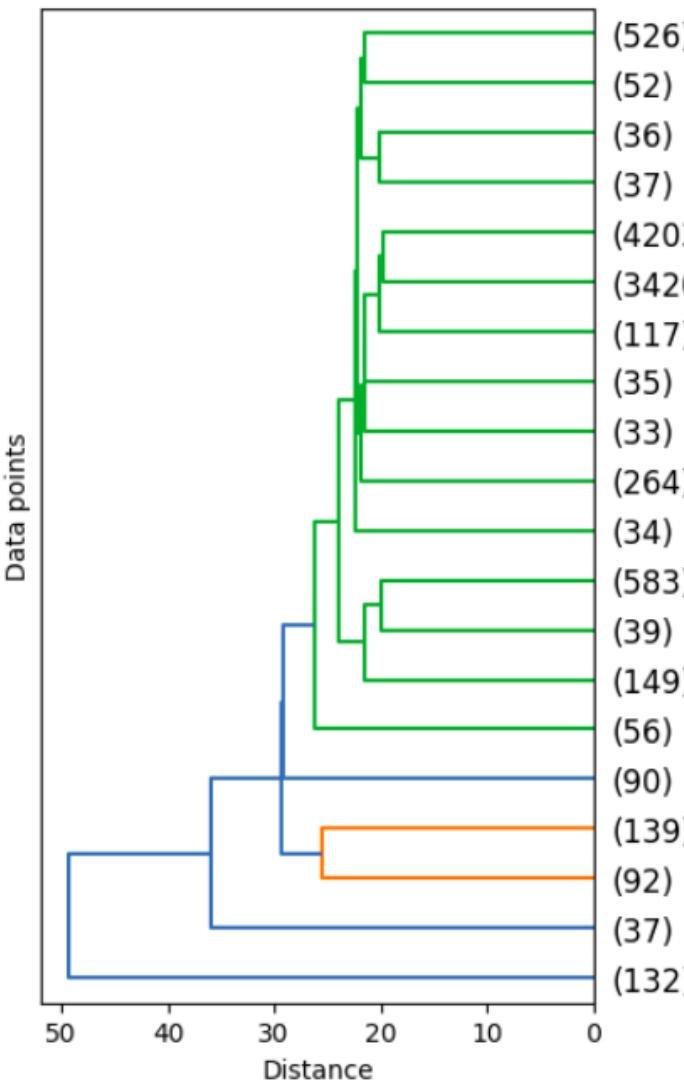
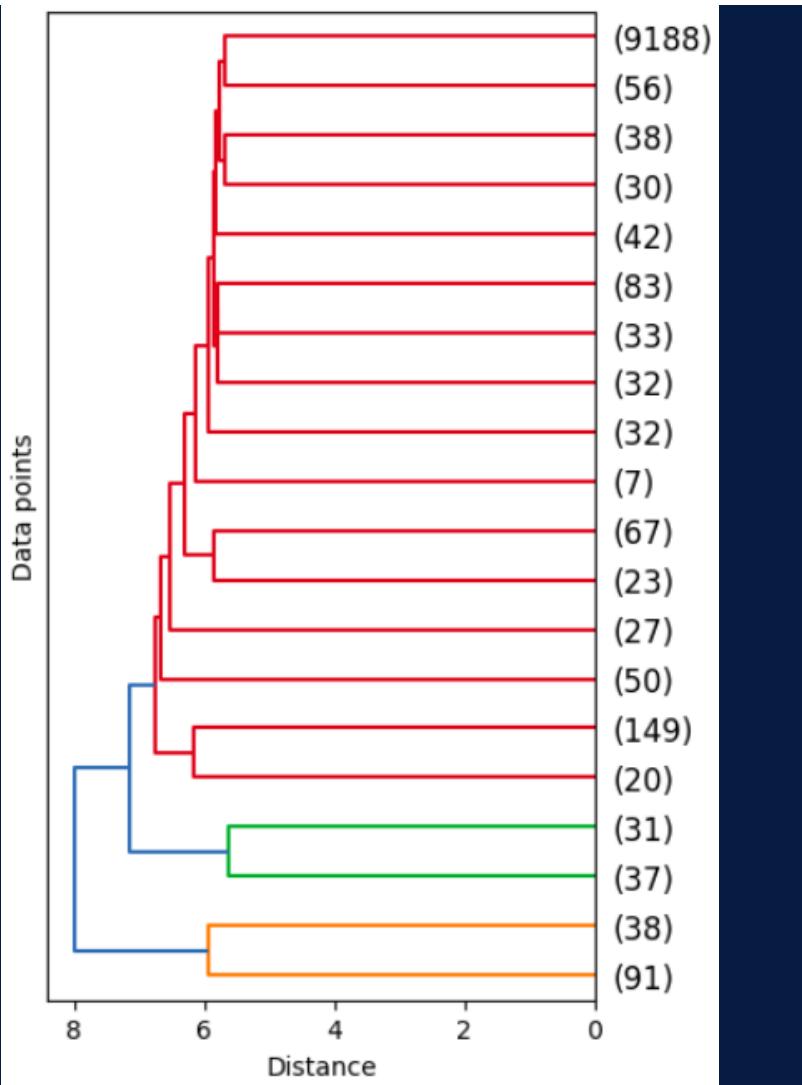
Cluster 2
thy snewd ebooks beep boop shakespeare fordo ncommands shakespeareinsult swapp

Hijerarhijsko klasterovanje



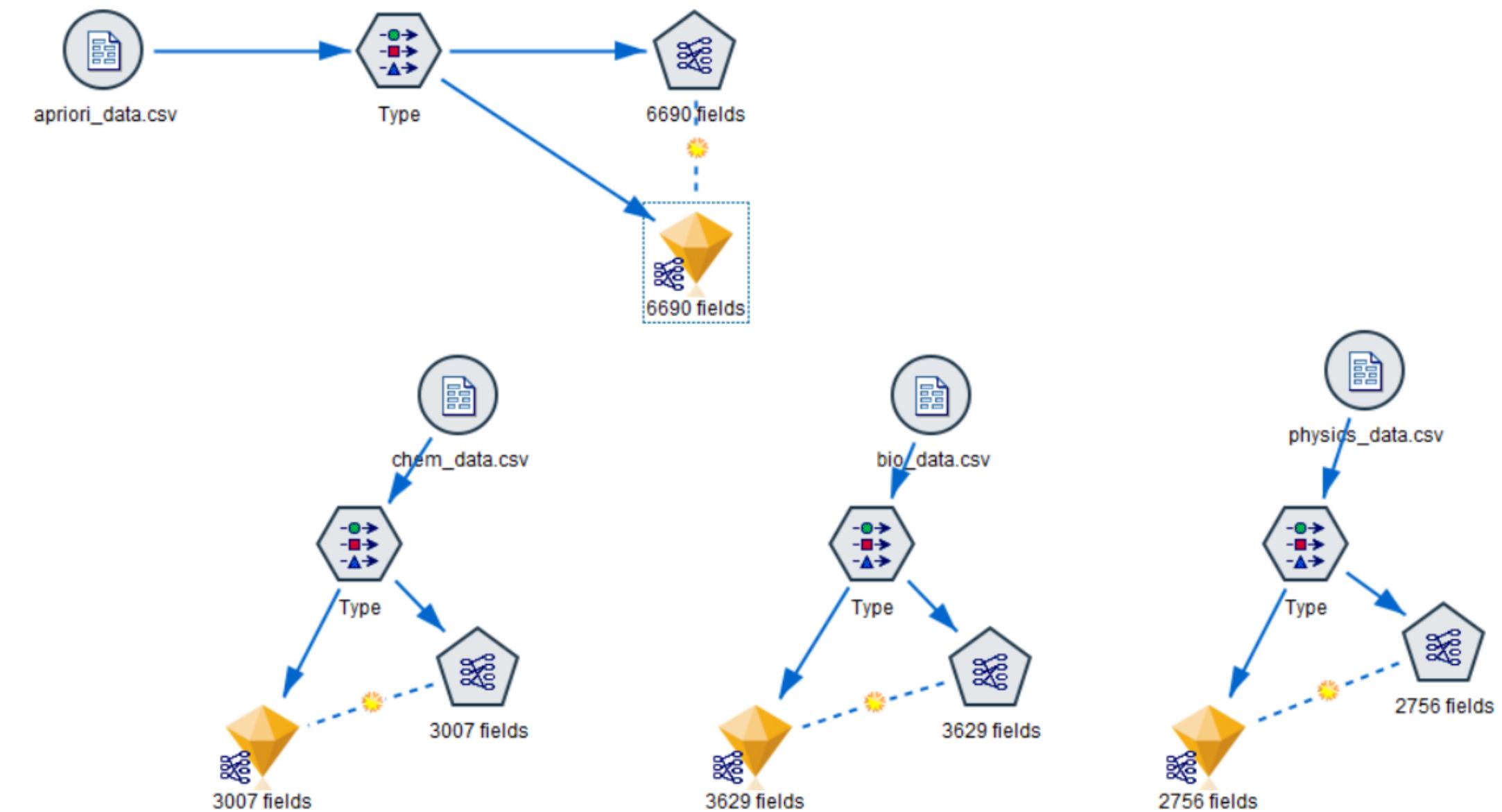
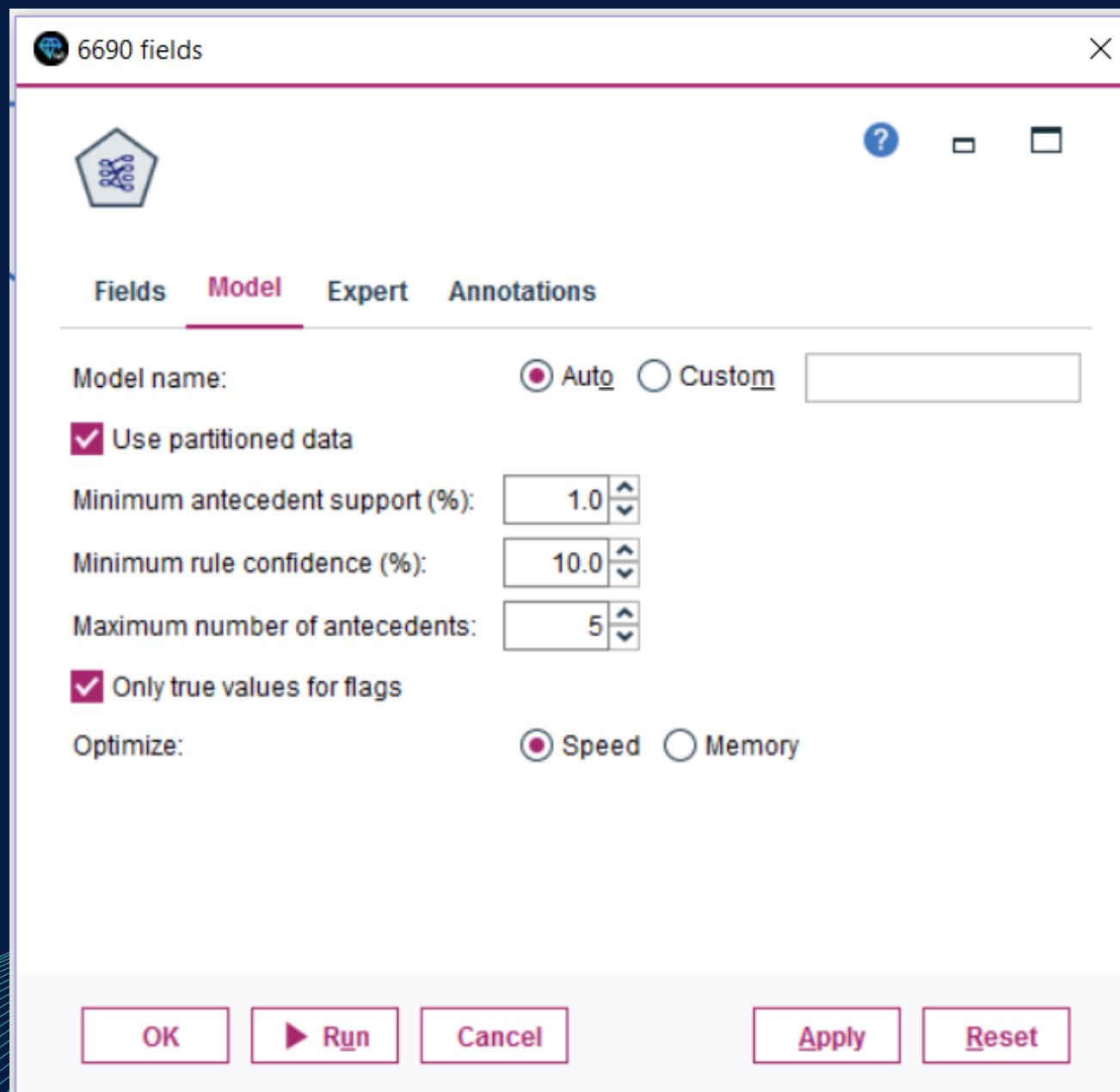


euklidsko rastojanje



kosinusno rastojanje

Pravila pridruživanja



Sort by: Lift

356 of 356

Consequent	Antecedent	Support %	Confidence %	Lift
compose	reddit	1.032	11.538	96.865
utm	reddit	1.032	10.577	96.865
vaccination	covid	1.042	10.476	81.182
mrna	vaccine	1.072	32.407	64.014
comment	reddit	1.032	37.5	58.119
mechanic	quantum	1.082	25.688	57.507
salary	phd	1.092	10.0	55.967
classical	quantum	1.082	11.009	48.22
immunity	vaccine	1.072	14.815	46.639
message	reddit	1.032	14.423	45.406
vaccinated	covid	1.042	12.381	43.009
relativity	general	1.032	10.577	42.621
travel	speed	1.082	24.771	39.609
journal	data	1.142	12.174	38.325
oxidation	state	1.112	12.5	38.159
vaccine	covid	1.042	40.0	37.311
covid	vaccine	1.072	38.889	37.311
cat	quantum	1.082	11.009	36.969
heart	covid	1.042	16.19	36.245
bullet	speed	1.082	18.349	36.244

Sort by: Lift

1022 of 1022

Consequent	Antecedent	Support %	Confidence %	Lift
pitched	higher	1.081	11.111	92.467
teneral	adult	1.081	13.333	92.467
ncommands	word	1.106	15.217	90.457
fordo	word	1.106	15.217	90.457
swapp	word	1.106	15.217	90.457
shakespeare	word	1.106	15.217	90.457
shakespeareinsult	word	1.106	15.217	90.457
thy	word	1.106	15.217	90.457
cream	ice	1.057	88.636	89.955
creationevolution	reddit	1.154	12.5	86.688
compose	reddit	1.154	12.5	86.688
compromised	immune	1.178	10.204	84.918
meal	food	1.202	10.0	83.22
opt	reddit	1.154	12.5	74.304
reindeer	drink	1.033	13.953	72.576
model	nit	1.033	13.953	72.576
bull	shit	1.226	11.765	69.933
funding	scientific	1.033	11.628	69.12
beep	reddit	1.154	14.583	67.424
boop	reddit	1.154	14.583	67.424
handled	bug	1.178	12.245	63.689
integrated	immune	1.178	12.245	63.689
web	spider	1.009	16.667	63.045
integrated	mrna	1.226	11.765	61.191
vaccination	vaccine	1.009	19.048	60.967
covid				
tea	drink	1.033	11.628	60.48
strawberry	ice	1.057	43.182	59.893

Sort by: Lift ▾

1571 of 1571

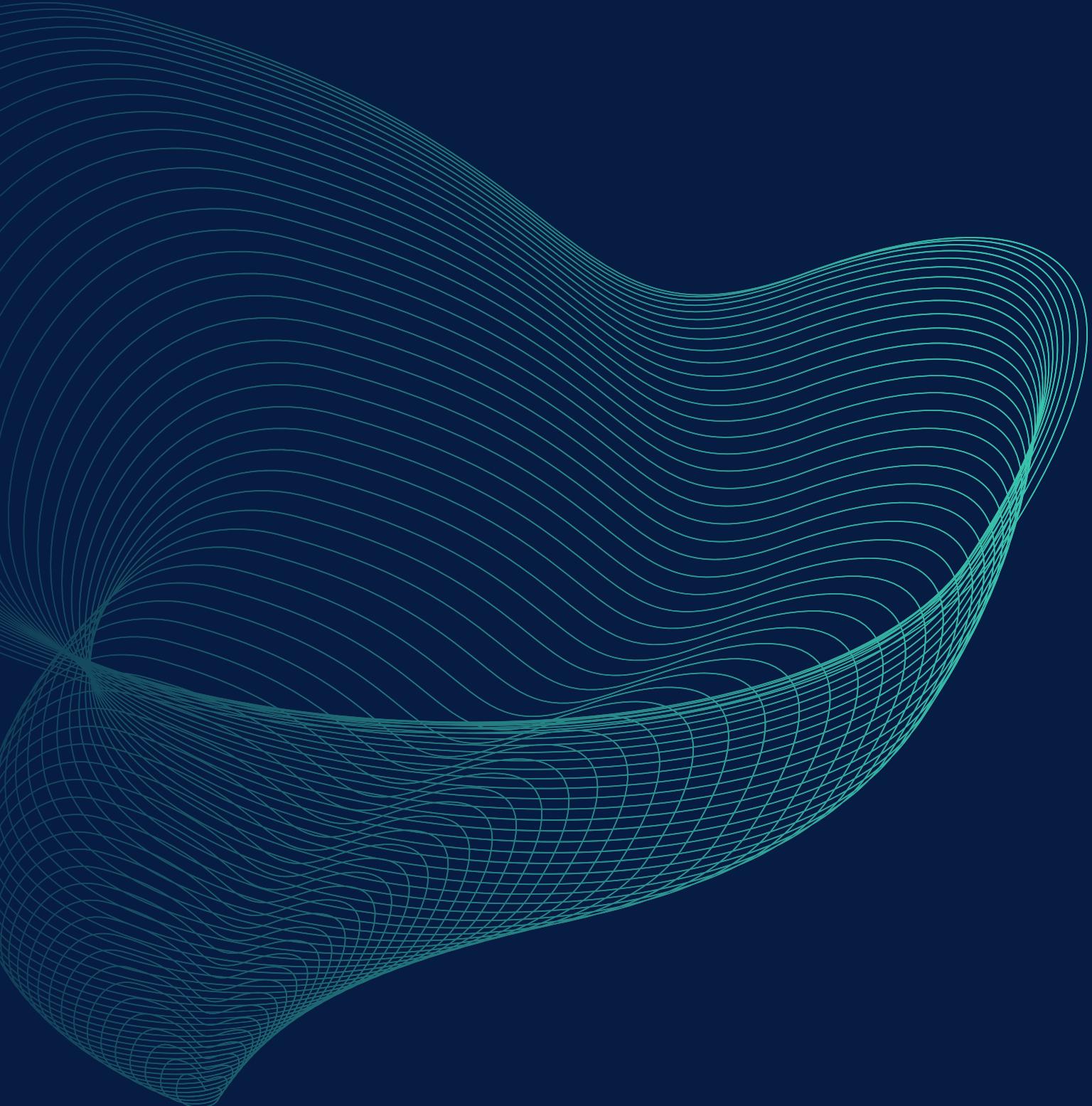
Consequent	Antecedent	Support %	Confidence %	Lift
spectrometry	mass	1.148	10.256	87.103
office	hour	1.148	20.513	87.103
rubbing	alcohol	1.089	13.514	76.509
tailor	job	1.178	12.5	70.771
helium	oxygen	1.266	16.279	69.125
wick	flame	1.178	10.0	67.94
manager	job	1.178	10.0	67.94
culture	phd chemistry	1.178	10.0	67.94
isopropyl	alcohol	1.089	29.73	67.328
hemoglobin	oxygen	1.266	11.628	65.833
octet	rule	1.119	18.421	62.576
peak	show	1.207	14.634	62.14
lattice	crystal	1.384	12.766	61.951
sanitize	alcohol	1.089	10.811	61.207
upon	come	1.089	10.811	61.207
valence	electron	1.649	21.429	60.661
reacts	chlorine	1.001	20.588	58.282
ocl	mass	1.148	10.256	58.068
dioxide	carbon	1.472	32.0	57.213
hydrochloric	chlorine	1.001	11.765	57.092

Sort by: Lift ▾

5960 of 5960

Consequent	Antecedent	Support %	Confidence %	Lift
alternatively	calculation	1.033	15.385	96.769
redirected	calculation	1.033	19.231	96.769
forum	calculation	1.033	19.231	96.769
assisting	calculation	1.033	19.231	96.769
predicting	value	1.033	11.538	96.769
distinct	open	1.033	11.538	96.769
distinct	towards	1.033	11.538	96.769
witness	travel speed	1.033	11.538	96.769
alternatively	asking	1.073	14.815	93.185
hgzfx	name	1.073	11.111	93.185
high efficiency	name	1.073	14.815	93.185
redirected	asking	1.073	18.519	93.185
forum	asking	1.073	18.519	93.185
assisting	asking	1.073	18.519	93.185
iossmf	name	1.073	18.519	93.185
app	name	1.073	18.519	93.185
inflaton	field	1.073	11.111	93.185
worthwhile	done	1.073	11.111	93.185
captured	capture	1.073	14.815	93.185
flue	capture	1.073	14.815	93.185
adsorbent	capture	1.073	18.519	93.185
thermodynamic	capture	1.073	18.519	93.185
renewables	capture	1.073	29.63	93.185
div	equal	1.153	10.345	86.759
symmetric	equal	1.153	10.345	86.759
counterintuitive	observation	1.153	10.345	86.759
cart	kinetic	1.153	20.69	86.759

Zaključak



Hvala na
pažnji!