Rezultati za sve značajke bez kombinacije značajki koji jako koreliraju zbog fizikalne sličnosti

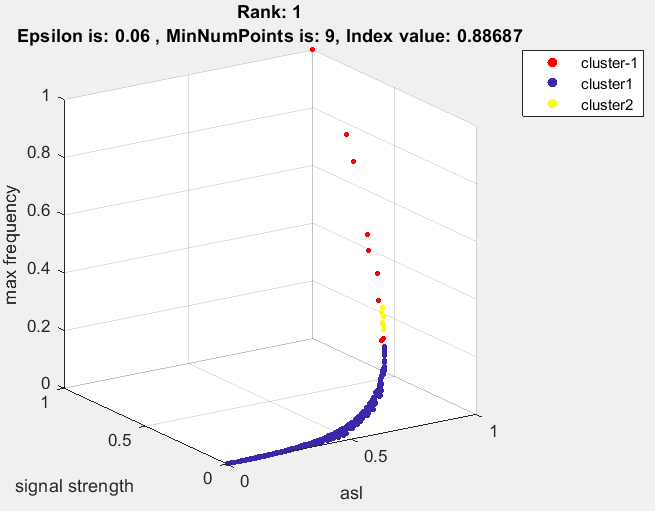
Ispod je prikazan koreleogram svih značajki za korišteni skup podataka.

Prva grupa značajki koje jako koreliraju su PEAK AMP, RMS, ABSOULUTE ENERGY i SIGNAL STRENGTH.

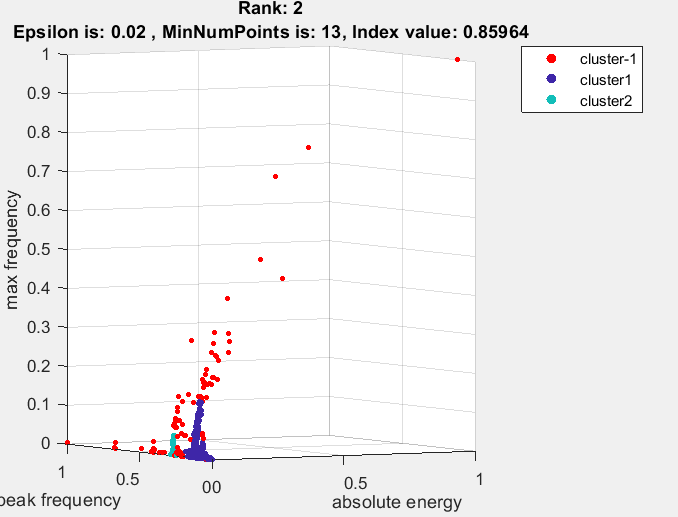
Druga grupa značajki koje jako koreliraju su WEIGHTED PEAK FREQUENCY i PEAK FREQUENCY.

Selekcija značajki provedena je kroz vrijednosti:

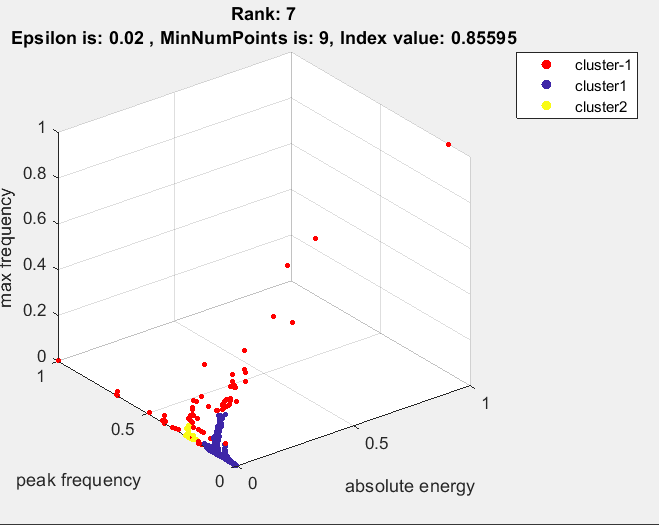
1. Epsilon – 0.02 do 0.06 s korakom 0.01
2. NMinPoints – 9 do 20 s korakom 1



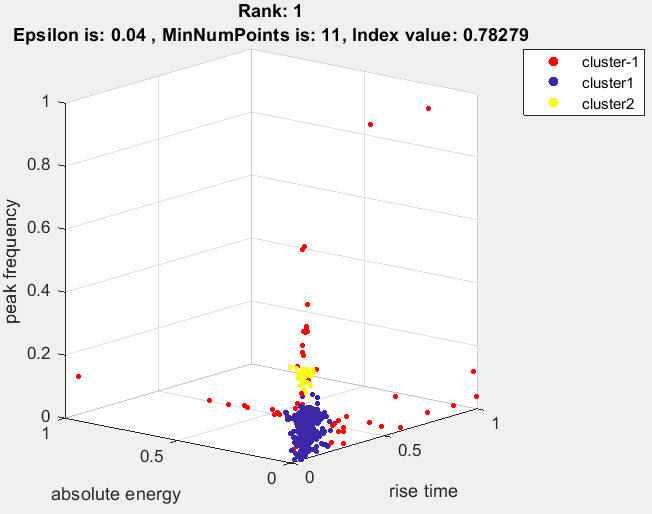
Zaključak da je potrebno koristiti veći broj NMinPoints jer kod male razlike u broju outliera indeks više preferira manji broj zgusnutih točaka.



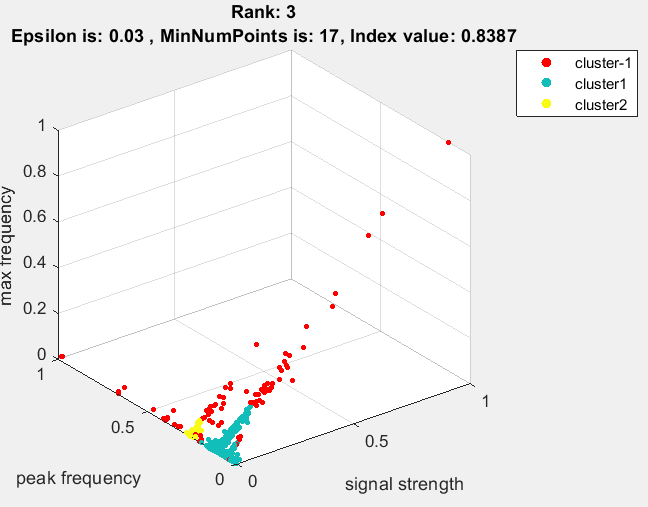
Ovdje se barem izabire veći broj točaka u svakom clusteru.



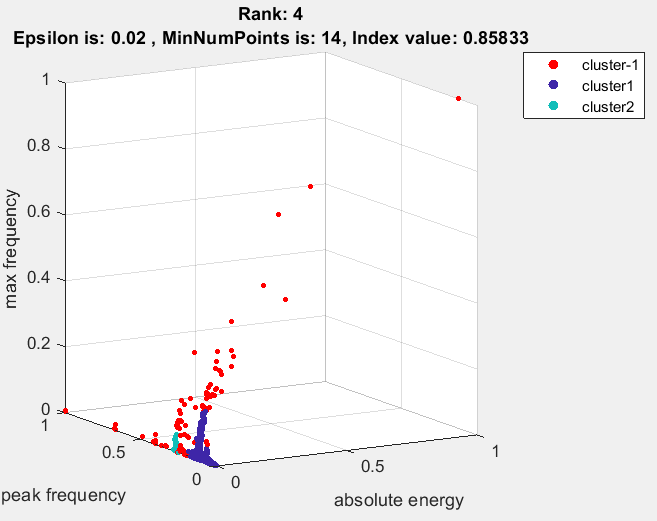
Za slihouette dobivamo dosta dobre rezultate.



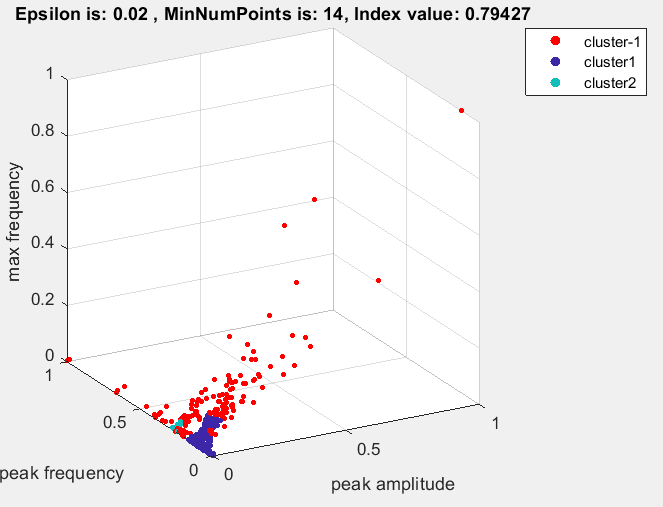
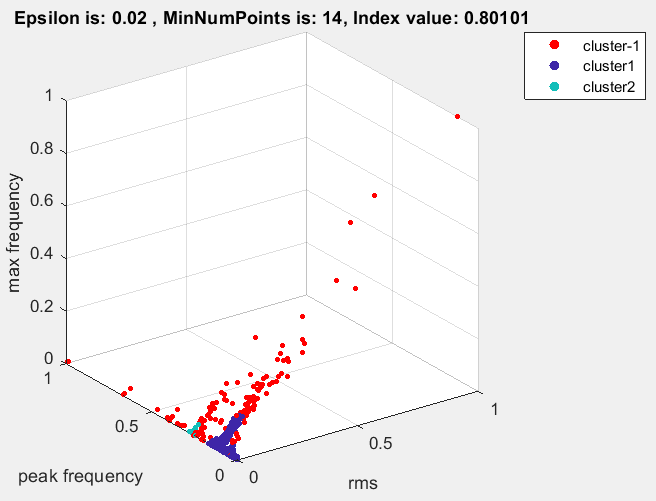
Promjenom algoritma selekcije značajki da za svaki podskup značajki prvo izračuna optimalni cluster rezultat pa onda prikazuje 10 najboljih subsetova.

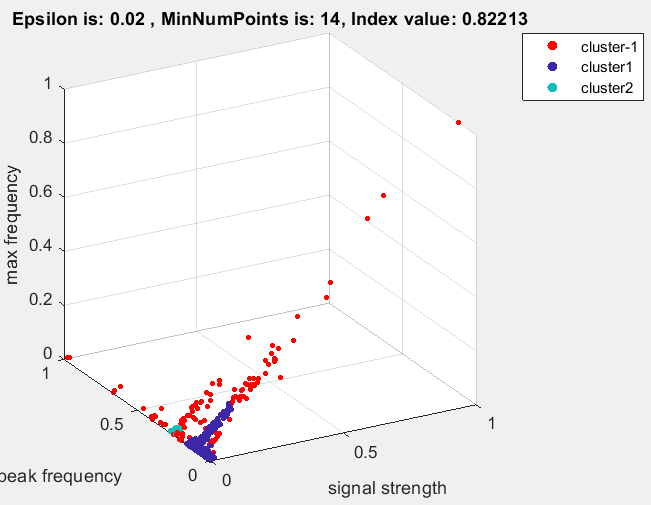
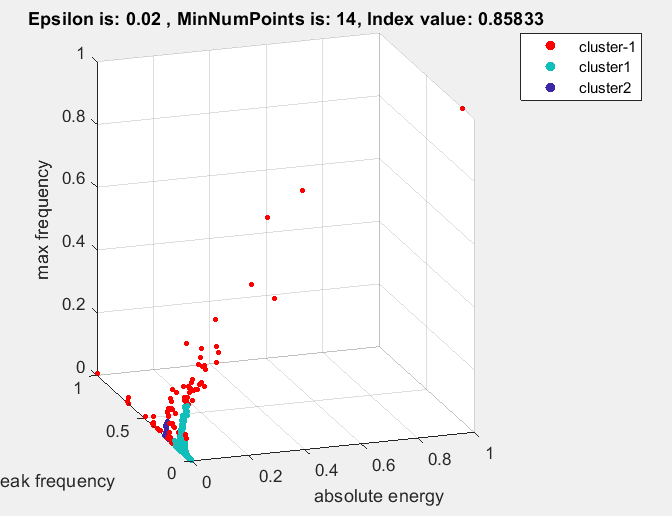


Vidimo kako abs energy zamjenjuje signal strength jer su dosta korelirani, možemo za smanjiti feature selection samo raditi s jednim od koreliranih značajki.



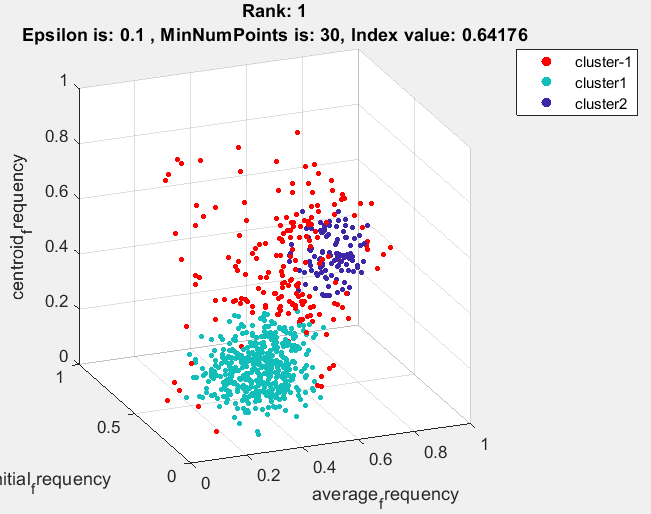
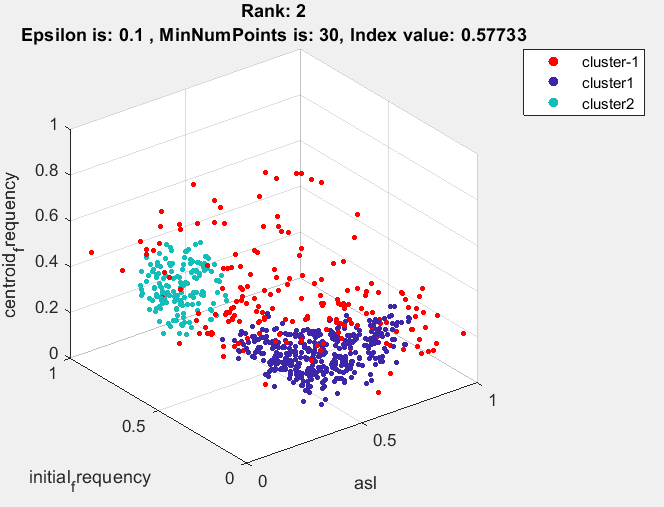
Za 4 korelirane značajke PEAK AMP, RMS, ABSOULUTE ENERGY i SIGNAL STRENGTH vidimo kako zamjenom jedne korelirane značajke s drugom u podskupu značajki čini slične skupove podataka te rezultira sličnim clusteringom.

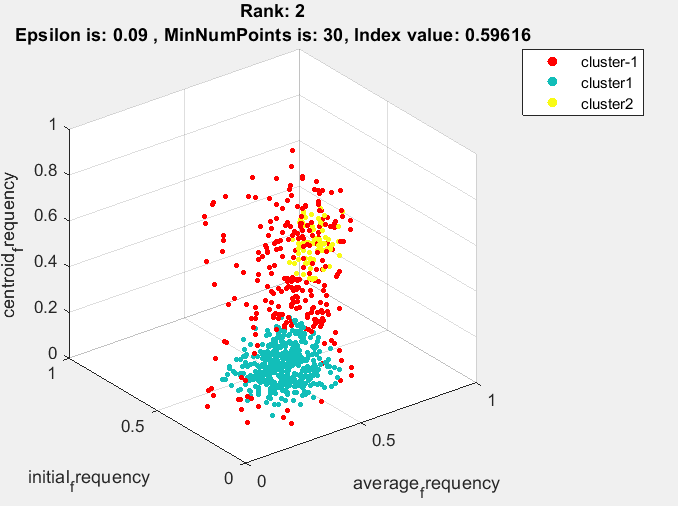
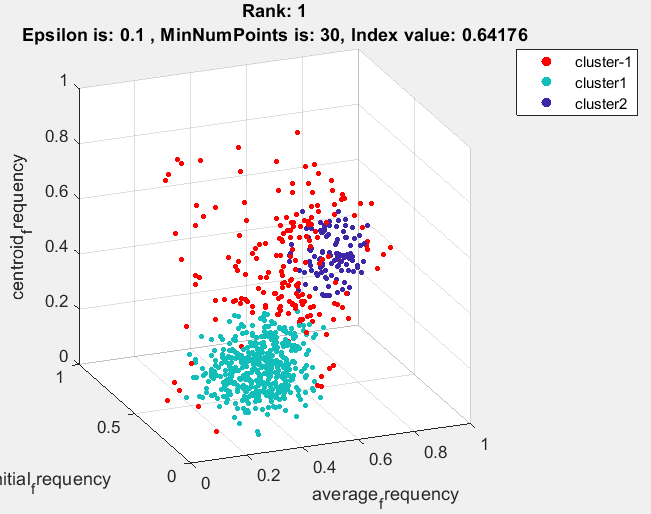
 

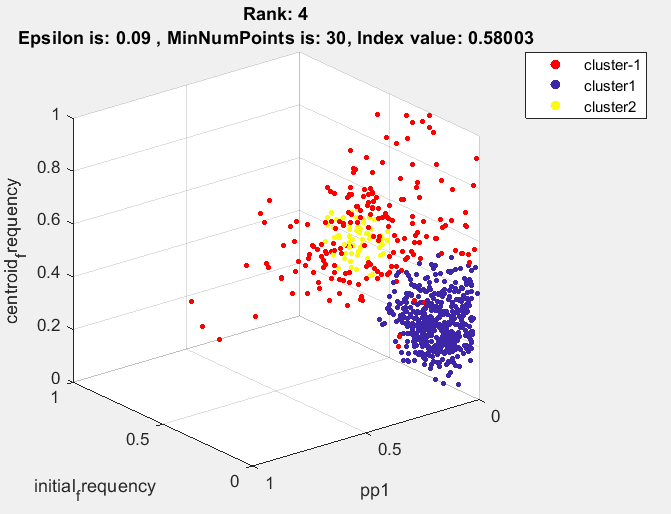
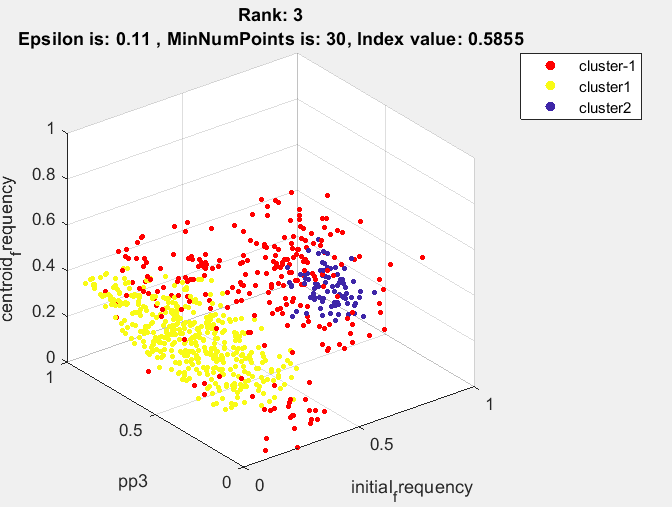
Želimo odrediti optimalne granice epsilona. Znamo da što je manji epsilon biti će potrebno smanjiti NminPoints za kvalitetan rezultat.

Za epsilon od 0.1, 0.15 i 0.2 uz NminPoints = 30 algoritam uvijek izabire epsilon = 0.1 te sam clustering nije optimalan gdje je najveća vrijednost indeksa 0.64176

Za NminPoints = 30 i epsilon 0.09 , 0.1, 0.11 dobivamo podvojene rezultate za najbolji epsilon gdje je najbolji rezultat za epsilon 0.1.





Još jedan koristan tip vizualizacije:

Obzirom da vrijeme ne uzimamo kao zasebnu značajku kod svrstavanja AE značajki u klastere, ne znamo:

* u koje vrijem se formirao pojedini klaster
* koliko je točaka u nekom vremenskom intervalu detektirano da pripada pojedinom klasteru

Ideja bi bila, **nakon** klasteriranja, vizualizirati na x-y grafovima u ovisnosti o vremenu:

* kumulativni broj AE emisija pridijeljen svakom klasteru (+ odvojeno neklasterirane točke)
* broj AE emisija pridijeljenih pojedinom klasteru po jednom satu

To pomaže npr. da se vidi npr. u kojoj fazi suše se javlja koji klaster i koliko je značajan.

To bi nam bilo korisno i za ugradbenu implementaciju.

Dobivamo zanimljivi rezultat za nevedeni skup značajki : PEAK FREQ, ABS ENERGY, REVB FREQ

