**Zadatak 1.**

a) Postoji jasna povezanost između koncentracije CO₂ i zauzetosti prostorije.

Temperatura također igra ulogu, ali čini se da sama po sebi nije dovoljan pokazatelj.

Postoji zona preklapanja između klasa, što može otežati klasifikaciju.

b) Broj primjera u skupu podataka: 10129

c) Radzioba:

* 0 - 8228
* 1 – 1901

**Zadatak 2.**

e) Što se događa s rezultatima ako se koristi veći odnosno manji broj susjeda?

* Manji broj susjeda (K=1, K=3)

Model postaje osjetljiviji na šum u podacima

Točnost može biti visoka na trening skupu, ali loša na testnom skupu

Model donosi odluku na temelju jednog (ili nekoliko) najbližih primjera, što ga čini nestabilnim

* Veći broj susjeda (K=10, K=15, K=20)

Model postaje glatkiji, manje osjetljiv na pojedinačne outliere

Može se dogoditi underfitting – model previše generalizira, što smanjuje preciznost

Točnost obično raste do optimalnog K, a zatim se smanjuje

f) Što se događa s rezultatima ako ne koristite skaliranje ulaznih veličina?

Ako ne koristimo skaliranje, KNN neće moći pravilno računati udaljenosti jer atributi imaju različite jedinice i raspon vrijednosti.

* Problem:

S5\_CO2 ima vrijednosti 400 - 2000 ppm, dok S3\_Temp ima vrijednosti 24 - 26°C

KNN izračunava udaljenosti, pa će atributi s većim vrijednostima (CO₂) dominirati nad atributima s manjim vrijednostima (temperatura)

Zadatak 3.