

Predikcija cijena telefona



Uvod

Glavni cilj projekta je analizirati koji od testiranih algoritama ima najbolje rezultate pri određivanju opsega cijene telefona na osnovu njegovih karakteristika. Odлучili smo se za anaziranje *logističke regresije*, *naivnog Bayes-ovog klasifikatora* i *KNN-a (K- nearest neighbours)*, pri čemu je najbolje rezultate postigla logistička regresija.

Skup podataka

Skup sadrži informacije o:

- **1000 telefona** opisanih sa
- **20 obilježja**

Zbog dominantnosti određenih podataka prije analize izvršićemo skaliranje vrijednosti **standardizacijom**.

Nakon toga, podatke dijelimo na:

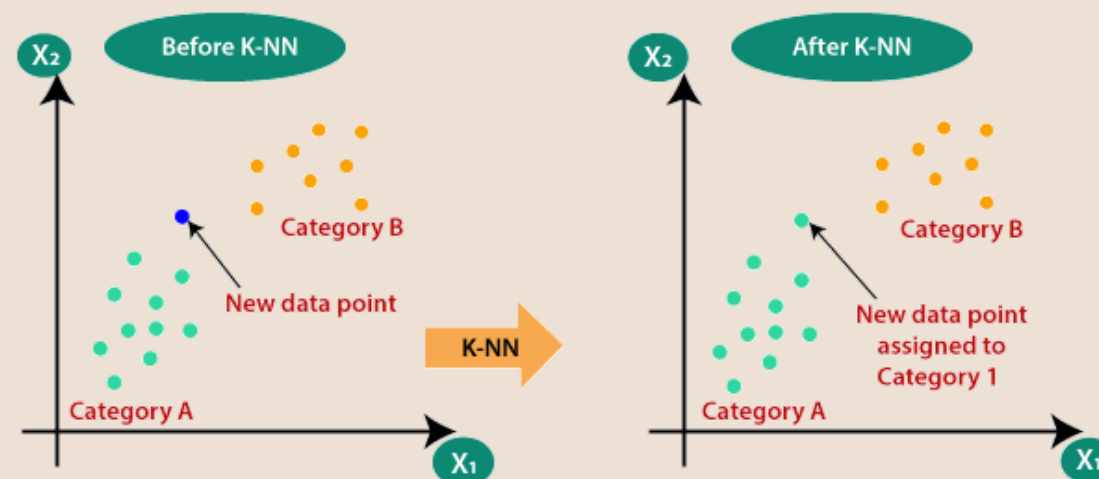
- trenirajući skup - 90%
- testirajući skup - 10%



Metodologije

- KNN

Podatak smjestamo u onu klasu kojoj priprada najviše od izabranih k susjeda.



- Logistička regresija

Koristićemo multiklasnu regresiju koja upotrebom *sigmoidne* funkcije računa vrijednosti klasifikatora.

- Naivni Bayes-ov kasifikator

Zasniva se na Bayes-ovoj teoremi sa pretpostavkom o nezavisnosti podataka.

Krajnje rješenje je ona vrijednost za koju klasifikator ima maksimalnu vrijednost.

Metrike

Za evaluaciju rješenja koristićemo:

- o **tačnost**
- o **preciznost**
- o **osjetljivost**
- o **F1 score**

Rezultati

