

**Document de specificare a cerințelor de sistem
ale aplicației software
(Software requirements specification)**

**Proiectarea și implementarea unui model
k-means pentru un set de date**

Cuprins:

Cuprins:	2
Scopul documentului:	3
Conținutul documentului:	3
Descriere generală:	3
Situția inițială:	3
Scopul produsului:	3
Contextul produsului:	3
Rezultate urmărite:	4
Descriere funcțională:	4
Actori:	4
Implementare:	4
Cerințe nefuncționale:	5
Interfața cu utilizatorul:	5
Disponibilitate și valabilitate:	5

Scopul documentului:

Acest document este realizat pentru a descrie cu acuratețe detalii despre structura și cerințele proiectului, oferind o privire de ansamblu asupra modului de funcționare cât și al actorilor prezenți. Documentul este folosit pentru construirea soluției de către echipa de dezvoltare a proiectului.

Conținutul documentului:

Documentul acoperă următoarele secțiuni:

- O descriere generală a problemei și a produsului ce oferă context cu privire la starea inițială, abordarea problemei și a rezultatului obținut.
- O descriere funcțională ce cuprinde cerințele produsului software, astfel vom descrie sistemul, actorii și cazurile de utilizare.
- Ultima secțiune cuprinde cerințe nefuncționale precum interfața cu utilizatorul și performanța.

Descriere generală:

Situația inițială:

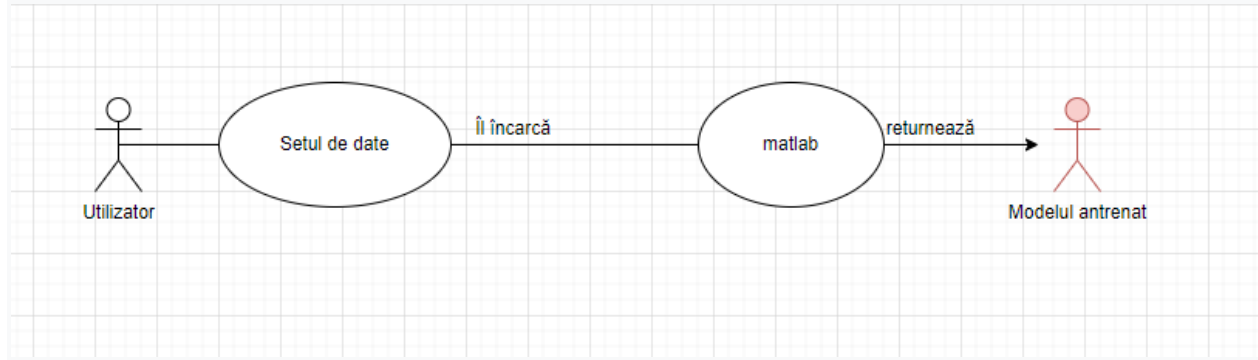
Se presupune problema clasificării muzicii în funcție de originea geografică. Astfel pornim de la un set de date care a fost construit dintr-o colecție de 1059 de piese care acoperă 33 de țări/zone. Muzica folosită este tradițională, etnică sau numai „lumească”, așa cum este clasificată de editorii produsului pe care apare. Orice muzică occidentală nu este inclusă deoarece influența sa este globală - ceea ce căutăm sunt aspectele muzicii care influențează cel mai mult locația. Astfel, posibilitatea de a specifica o locație cu influență puternică asupra muzicii este centrală.

Scopul produsului:

Rezolvarea problemei presupune construirea unui model k-means împărțind setul de date în 2 secțiuni: una de antrenare și una țintă. În urma construirii acestui model se va evalua acuratețea, precizia, sensibilitatea și scorul F1 a modelului în urma clasificării a melodiilor din setul de testare. Un model cu acuratețe și precizie cât mai mare va putea clasifica corect melodii introduse de utilizatori, din afara setului de date.

Contextul produsului:

Implementarea va fi realizată în matlab și rulată local, setul de date va fi încărcat de către utilizator. Tot utilizatorul alege cum se împarte setul de date pentru antrenare și testare.



Rezultate urmărite:

Scopul produsului este analiza rezultatelor finale și a performanței modelului obținut. Se urmărește o acuratețe și precizie cât mai mare a clasificărilor.

Descriere funcțională:

Actori:

Actorii sunt utilizatorii produsului ce reprezintă echipa de dezvoltare a produsului care va analiza ulterior rezultatele obținute prin rulare.

Implementare:

Implementarea presupune dezvoltarea unui model k-means utilizând matlab.

K-means este o metodă de clasificare nesupervizată ce presupune gruparea obiectelor dintr-un set de date în mai multe cluster, fiecare cluster conținând un set de obiecte dintr-o anumită categorie.

Presupunem că se dorește gruparea obiectelor în K cluster. algoritmul implică următorii pași:

1. Se stabilesc cele K centre în spațiul obiectelor ce trebuie clasificate. Pozițiile centrelor pot fi, de exemplu, generate aleator sau stabilite manual de către utilizator. În mod ideal, aceste centre vor fi cât mai îndepărtate unul de altul.

2. Pentru fiecare obiect, se determină clusterul căruia îi aparține ca fiind clusterul cu centrul cel mai apropiat de poziția obiectului.

3. Odată grupate toate obiectele, se determină baricentrul fiecărui cluster (de exemplu, în cazul unor puncte din plan se poate face media aritmetică a coordonatelor) și se consideră noul centru al clusterului ca fiind acest baricentru.

4. Se repetă pașii 2, 3 până când pozițiile centrelor clusterelor nu se mai modifică (în practică, se stabilește o valoare prag sub care trebuie să se situeze diferența dintre pozițiile centrelor clusterelor determinate în cadrul iterației curente și cele determinate în iterația anterioară).

Cerințe nefuncționale:

Interfața cu utilizatorul:

Interfața cu utilizatorul va fi cea din matlab, implementarea constă într-un script care va fi rulat direct din program.

Disponibilitate și valabilitate:

Produsul este rulat local, deci va fi disponibil oricând iar rezultatele vor fi stocate tot local și păstrate pentru a putea fi vizualizate și analizate ulterior.