**Documentatie Proiect ML (Smartphone User Classification)**

SUBMISIA 2

**Cuprins:**

1. **Descrierea submisiei**
2. **Descrierea modelului ales** *(detaliat in submisia precedenta)*
3. **Media preciziei obtinute in urma folosirii unui 3-fold cross-validation pe setul de date de antrenare. Matricile de confuzie pt fiecare combinatie**
4. **Alte metode incercate** *(detaliate in submisia precedenta)*
5. **Descrierea submisiei**

Clasificator ales: SVM.SVC()

Parametri:

* C = 9
* Kernel = "rbf"

Precizie:

* locala: 0.9165
* Kaggle:
  + clasament public: 0.853
  + clasament privat: 0.84825

Citirea si memorarea datelor:

* Acestea sunt retinute intr-o matrice de dimensiune "variabila"

Extragerea unor features + Modelarea datelor de antrenare si de testare:

* Calcularea fft-urilor semnalelor accelerometrului (Fast Fourier Transform = un algoritm care converteste un semnal prin impartirea acestuia in componente periodice intr-o reprezentare a frecventelor acestor componente repetitive) cu ajutorl bibliotecii SciPy
* Aplicarea unor functii statistice pe coloane(atat pt x, y, z, cat si pt fft\_x, fft\_y si fft\_z) implementate in biblioteca Statistics
* Pentru a mai obtine informatii din numarul restrans de date, am cuplat pe coloane mai multe linii asupra carora am aplicat ulterior aceleasi functii de statistica
* In final sunt 1119 features

Normalizarea datelor:

* Am folosit scalarea standard: fiecare dimensiune este scalată folosind deviația standard corespunzătoare, transformă vectorii caracteristici astfel incat fiecare sa aiba media 0 si deviatia standard 1

Deoarece alegerea parametrilor C si gamma este critica in performanta SVM-ului de kernel "rbf", am gasit cea mai buna combinatie a acestora pentru clasifcatorul ales si datele modelate prin parcurgerea unor vectori de valori distincte.

1. **Media preciziei obtinuta in urma folosirii unui 3-fold cross-validation pe setul de date de antrenare. Matricile de confuzie pt fiecare combinatie**



