

# Raport Proiect SVM Dermatologie

## - Delcea Andrei Iacob -

Scopul programului este de a determina tipul de Boală Eritemato-Scuamoasă cu ajutorul unei baze de date numerice.

Baza de date constă în 34 de attribute, 12 clinice și 22 histopatologice, distribuite pe coloane, 8 dintre valorile ale atributului varstă lipsesc și au fost trecute cu simbolul '?', iar pe a 35 a coloană se află etichetele. În total există 366 de instanțe ale pacienților în baza de date, 70% dintre ele vor fi distribuite ca și date de antrenament, iar 30% ca și date de test.

Librăriile pe care le-am folosit sunt „Pandas” pentru extragerea și pregătirea datelor din baza de date și „SciKit Learn” pentru implementarea algoritmului Support Vector Machine.

Pentru măsurarea performanței am folosit metrica Accuracy Score din biblioteca SciKit Learn, care compară etichetele de test cu etichetele prezise și oferă un coeficient de precizie.

Variind costul între  $2^{-5}$  și  $2^7$  și folosind un nucleu linear am obținut rezultatele:

Cost	Precizie
$2^{-5}$	0.954545454545454
$2^{-3}$	0.972727272727272
$2^{-1}$	0.972727272727272
$2^1$	0.972727272727272
$2^3$	0.981818181818181
$2^5$	0.981818181818181
$2^7$	0.981818181818181

Se observă cum odata cu creșterea costului crește și precizia, în intervalul  $2^{-5}$  și  $2^7$ .