

Smart grid stability

Francesc Gallego
1467414

<https://github.com/cescjunior24/APC-KAGGLE>





Introducció

Aquest dataset correspon a una versió augmentada del "Conjunt de dades simulat d'estabilitat de la xarxa elèctrica", creat per Vadim Arzamasov i donat al Repositori d'aprenentatge automàtic de la Universitat de Califòrnia (UCI).

- Informació sobre una red elèctrica amb diferents atributs
- Conté un dataset pel train de 60000 i un pel test de 29998
- El nostre objectiu serà predir si la red elèctrica és estable o inestable



Anàlisi de les dades

Dataset amb 14 atributs:

0	tau1	60000	non-null	float64
1	tau2	60000	non-null	float64
2	tau3	60000	non-null	float64
3	tau4	60000	non-null	float64
4	p1	60000	non-null	float64
5	p2	60000	non-null	float64
6	p3	60000	non-null	float64
7	p4	60000	non-null	float64
8	g1	60000	non-null	float64
9	g2	60000	non-null	float64
10	g3	60000	non-null	float64
11	g4	60000	non-null	float64
12	stab	60000	non-null	float64
13	stabf	60000	non-null	object

Atribut objectiu: stabf

Tipus d'atributs

- Float
- Object

El tipus object el passarem a binari

I he mirat també que no hi ha valors nulls



Anàlisi de les dades



Correlació entre les dades

Per la variable objectiu stabf

Correlacions més altes:

- tau1, tau2, tau3, tau4
- g1, g2, g3, g4

Eliminarem p1, p2, p3, p4
perquè tenen poca correlació
amb els altres atributs



Mètode d'aprenentatge

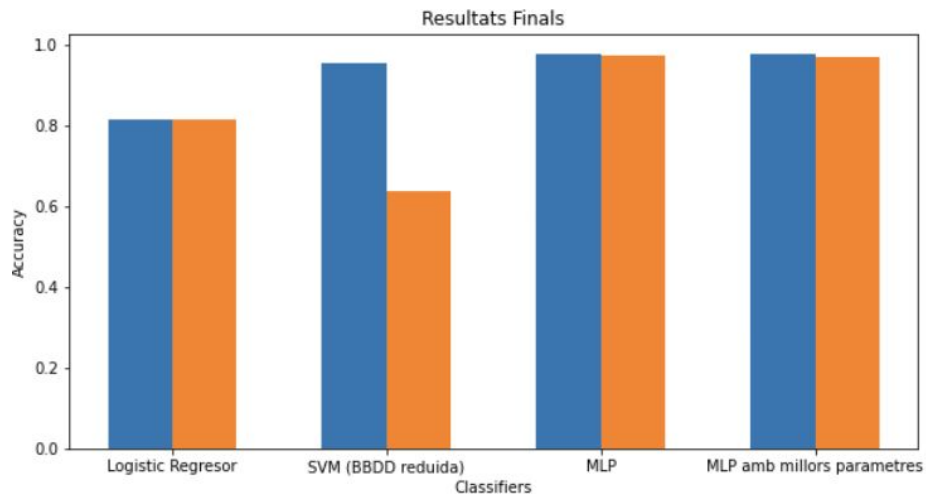
Al tractar-se de un target binari, estable o no estable, he implementat diferents classificadors:

- Regressor logístic
- SVM (Support Vector Machine)
- MLP (Multi-Layer Perception)

Mètrica escollida: Accuracy



Resultats



Millor model:

MLP amb paràmetres

Accuracy: 97,8%



Conclusions

He aconseguit una millora en el model MLP al trobar els millors paràmetres, el que he fet es baixar el learning rate a 0.001 i la alpha a 0.001

Degut a la dimensió de la BBDD no he pogut aplicar una SVM a tot el dataset, només a 10000 files, i això ha fet que el accuracy no fos tan alt, tot i així molt content amb el resultat que ha donat

El regressor lògistic té el accuracy més baix comparat amb els altres tres models

Els accuracy dels 4 models són bastant elevats, he estat veient resultats d'altre gent, i suposo que al tractar millor les dades han aconseguit resultats molts més elevats al voltant del 99.1% de accuracy.