

A dark blue vertical bar on the left side of the slide. A blue arrow points to the right from the bar, containing the date.

09/11/2022

A5 - Algoritmes de Classificació

Rapid Miner

Several thin, curved lines in dark blue and light grey originate from the bottom left and curve upwards and to the right.

Guillem Calaf Vidal

IES BERNAT EL FERRER
M05 – PROJECTES BIG DATA
CURS: IABD 2022-2023

Contingut

- 1. Taula d'observacions..... 2
- 2. Observacions 3
 - a. Decision Tree 3
 - b. Logistic Regression 3
 - c. Random Forest 3
 - d. Support Vector Machine 3
 - e. KNN..... 3
 - f. Naive Bayes 3

1. Taula d'observacions

	Parametre 1	Parametre 2	Parametre 3	Accuracy
Decision Tree	maximal_depth=20	Confidence=0.25	Minimal gain=0.1	78.69%
Decision Tree	maximal_depth=15	Confidence=0.25	Minimal gain=0.1	78.69%
Logistic Regression	C=5	Max-iter=100000	Penalty o Solver=N/A	ERROR
Random Forest	n_estimators=50	maximal_depth=20	Confidence=0.25	83.80%
Support Vector Machine	C=10	Kernel=radial	Gamma=5	ERROR
Support Vector Machine	C=10	Kernel diferent al anterior=anova	Gamma igual al anterior=5	ERROR
KNN	k_neightbors=3	weight=uniform	p=2	65.39%
KNN	k_neightbors=5	weight=uniform	p=2	65.62%
Naive Bayes	Laplace_correction=true	var_smoothing	priors	89.69%

2. Observacions

a. Decision Tree

Aquest algoritme s'ha executat pràcticament de forma immediata amb els paràmetres definits a la taula, mostra un nivell de precisió alt, no necessita d'operadors addicionals per a l'execució i mostra una interpretació bastant clara dels resultats, com a observació, canviant els paràmetres la precisió no canvia.

b. Logistic Regression

Aquest algoritme no s'ha pogut provar amb el Dataset utilitzat, per tant no es pot saber el seu nivell de precisió, necessita dos operadors addicionals ja que no pot processar dades binomials, la primera és un operador per a transformar les dades a nominals, i el segon és per a eliminar els camps vuits, com a observació, el camp de max-iter es troba ocult a mostrar paràmetres avançats, i el Penalty o Solver no existeix com a paràmetre dins l'operador.

c. Random Forest

El temps d'execució d'aquest algoritme varia molt segons el número d'arbres que s'hi declari, mostra un nivell de precisió més elevat que el Decision Tree i com aquest, tampoc necessita d'operadors addicionals, mostra una interpretació més caòtica dels resultats malgrat l'alta precisió.

d. Support Vector Machine

Aquest algoritme no s'ha pogut provar amb el Dataset utilitzat, per tant no es pot saber el seu nivell de precisió, necessita dos operadors addicionals ja que no pot processar dades binomials, la primera és un operador per a transformar les dades a nominals, i el segon és per a eliminar els camps vuits, com a observació, només els kernels radial i anova contenen el paràmetre gamma.

e. KNN

Aquest algoritme, com els demás, ha tingut un temps d'execució pràcticament immediat, té un nivell de precisió més baix que la resta d'algoritmes utilitzats en aquesta pràctica, no necessita d'operadors addicionals per a funcionar, mostra un nivell d'interpretació bastant senzill i caòtic, i la precisió canvia entre les dues proves amb aquest algoritme.

f. Naive Bayes

Aquest últim algoritme és el que obté un nivell de precisió més elevat, el temps d'execució no varí gaire en comparació als demás, només té un paràmetre per a poder utilitzar ja que s'obté un model més estandarditzat sacrificant la precisió de la predicció degut a la baixa varietat classificatòria.