

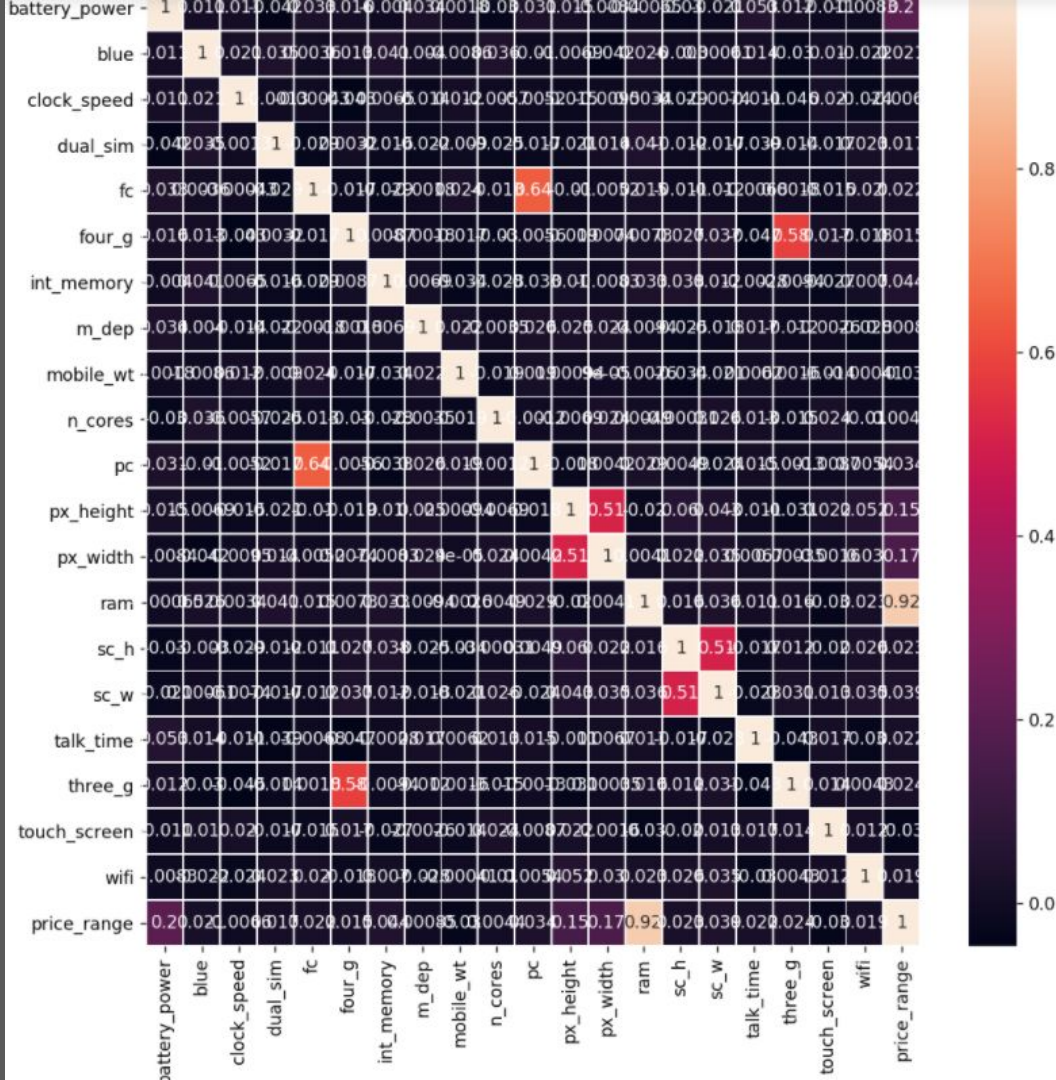
Pràctica 2 Classificació

Aprenentatge Computacional

Gerard Vinyes
Carlota Castro

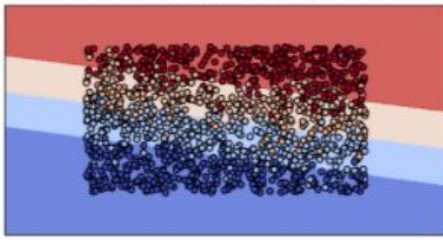
Introdució al dataset

- Classificació per preu de mòbils
- 2000 mostres, 21 variables
- Distribucions binàries minoritàries
- El nostre atribut objectiu és el rang de preu del mòbil
- Dataset dividit en dos arxius: train.csv i test.csv
- Correlació lineal baixa per a quasi totes les variables, moderada per al battery_power, px_height i px_width i alta per a la ram

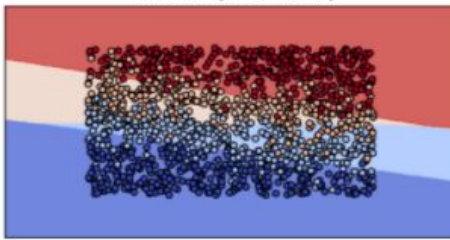


Anàlisi de les classificacions

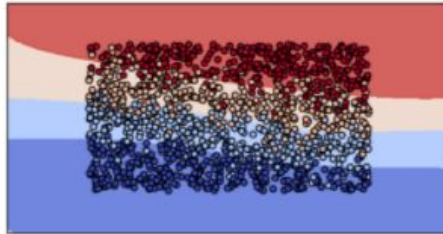
SVC with linear kernel



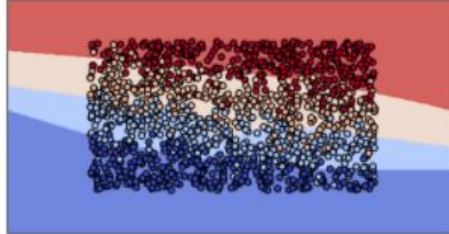
LinearSVC (linear kernel)



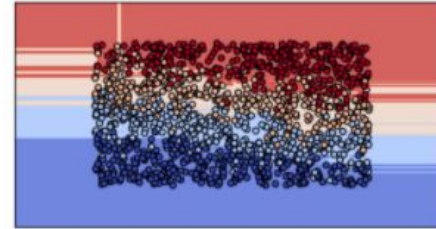
SVC with RBF kernel



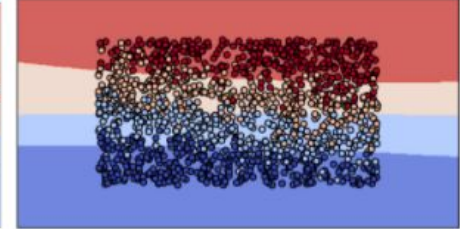
SVC with polynomial (degree 3) kernel



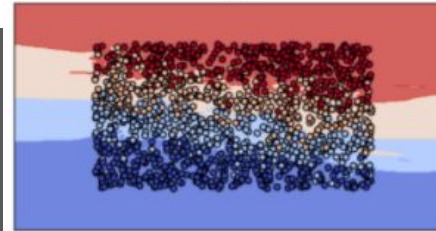
Random Forest



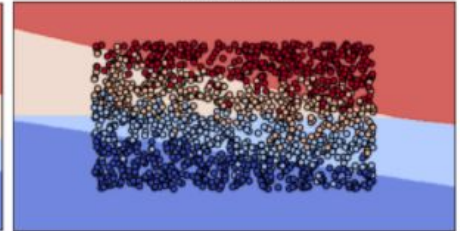
Gaussian Naive Bayes



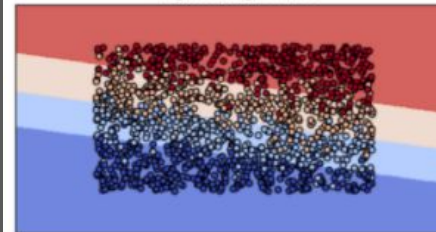
KNN



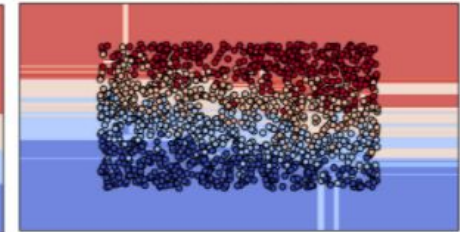
KN Centroids



Logistic Regression



Decision Tree



Acuracy of SVC with linear kernel : 0.8257142857142857
Acuracy of LinearSVC (linear kernel) : 0.7607142857142857
Acuracy of SVC with RBF kernel : 0.8214285714285714
Acuracy of SVC with polynomial (degree 3) kernel : 0.7864285714285715
Acuracy of Random Forest : 1.0
Acuracy of Gaussian Naive Bayes : 0.7664285714285715
Acuracy of KNN : 0.8721428571428571
Acuracy of KN Centroids : 0.7907142857142857
Acuracy of Logistic Regression : 0.8242857142857143
Acuracy of Decision Tree : 1.0

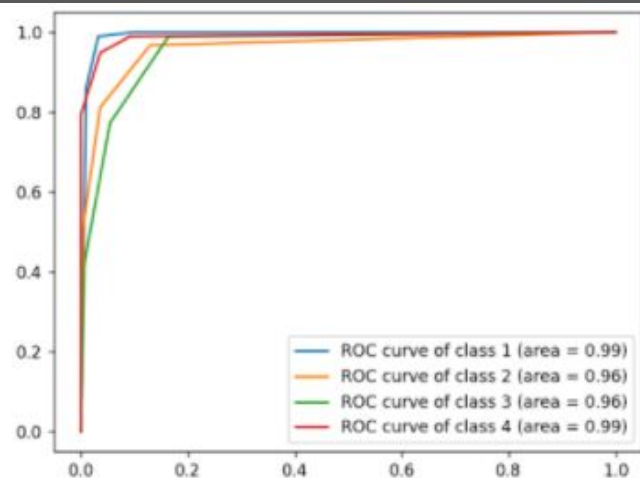
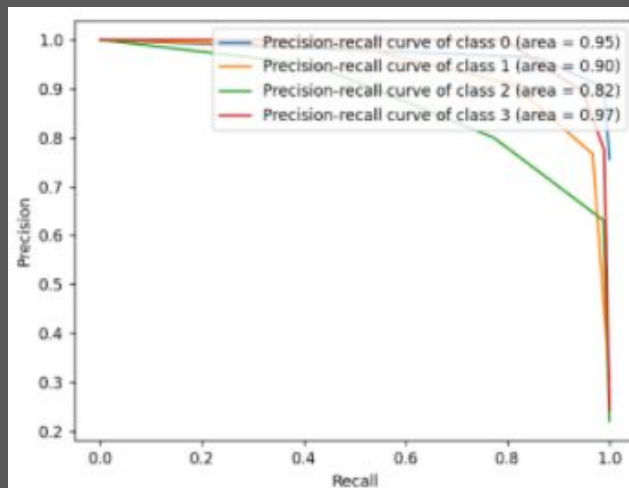
Crossvalidation

```
2 : 0.785625
3 : 0.7862428062482872
4 : 0.7943749999999999
5 : 0.791875
6 : 0.7912125820168399
7 : 0.7868388219456939
8 : 0.7887500000000001
9 : 0.7843444282217849
10 : 0.7849999999999999
11 : 0.7868037961094174
12 : 0.7849427673661767
13 : 0.7862272791462406
14 : 0.7874359812574916
15 : 0.7848939046611416
16 : 0.78625
17 : 0.7850009880772016
18 : 0.7861976506639426
19 : 0.7855521155830754
20 : 0.7825
```

- Avaluació dels models
- Utilitzem KFold com a estratègia
- $K = 4$
- LeaveOneOut massa costós en temps

Metric Analytics

	precision	recall	f1-score	support
0	0.91	1.00	0.95	93
1	0.92	0.86	0.89	122
2	0.85	0.82	0.83	88
3	0.93	0.95	0.94	97
accuracy			0.91	400
macro avg	0.90	0.91	0.90	400
weighted avg	0.90	0.91	0.90	400



Hyperparameter Search

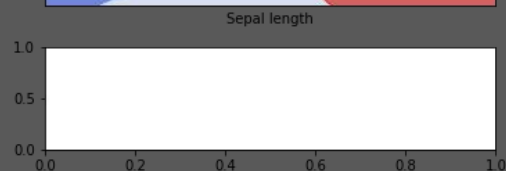
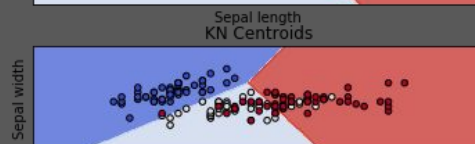
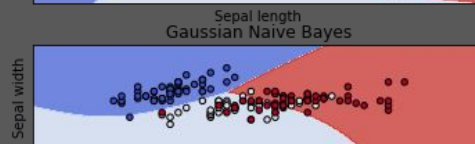
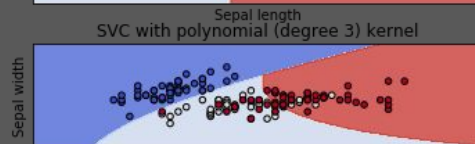
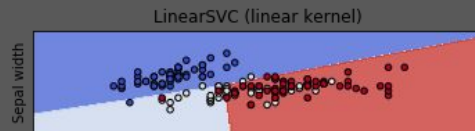
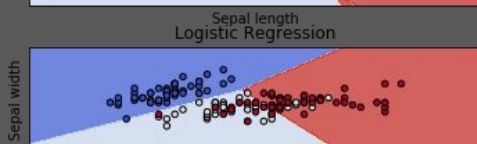
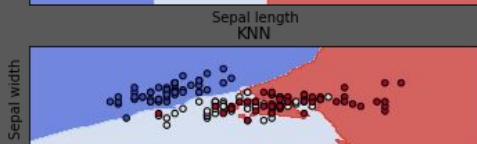
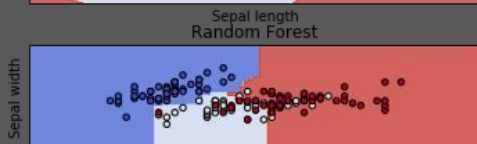
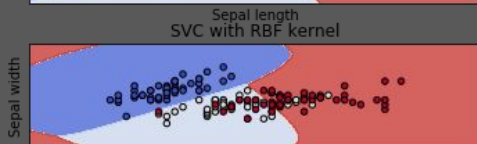
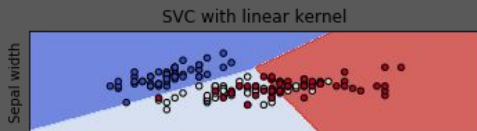
Randomized Parameter Optimization vs Exhaustive Grid Search

Després d'aquest procés, hem trobat que els millor valors eren els següents:

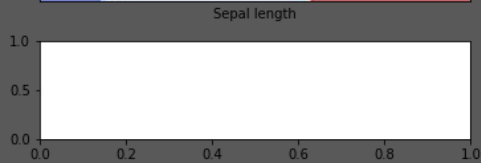
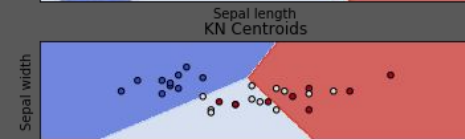
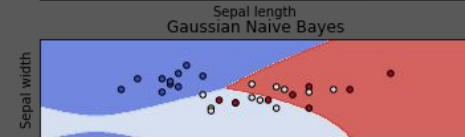
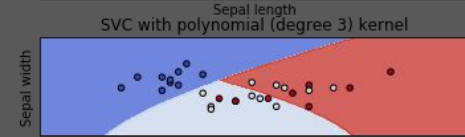
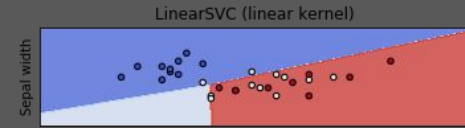
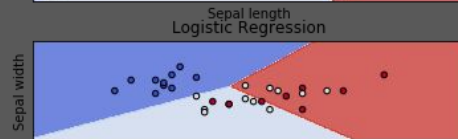
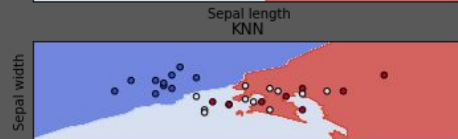
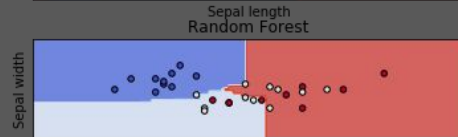
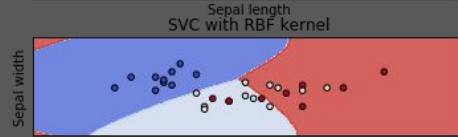
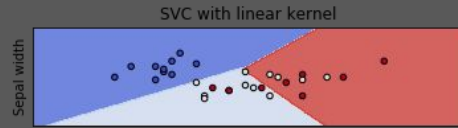
- Gaussian Naive Bayes: {'var_smoothing': 0.43287612810830584}
- KNN: {'leaf_size': 1, 'n_neighbors': 38, 'p': 1}
- Logistic Regression: {'C': 0.0001, 'penalty': 'none'}
- Decision Tree: {'criterion': 'entropy', 'max_depth': 20, 'min_samples_leaf': 5}

Apartat B

Classificació amb tot el conjunt de dades



Classificació amb el conjunt de dades d'entrenament



Problemes + Conclusions

- No em vist gaire dificultat en el que se'ns demanava
- Trobar el millor valor per els hiperparàmetres
- Triar la millor mètrica de classificació