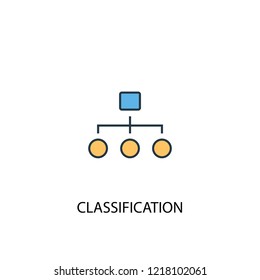
Practica 2.- Classificació



Grup: 304

Víctor Bosch Pueyo 1566583

Álvaro Caravaca Hernández 1566685

Juan Carlos Martínez Moreno 1566936

Índex

[Apartat B: Comparativa de models 3](#_Toc88211440)

[Introducció a la base de dades 3](#_Toc88211441)

[Comparació de models 3](#_Toc88211442)

[Apartat A: Classificació Numèrica 8](#_Toc88211443)

[EDA (exploratory data analysis) 10](#_Toc88211444)

[Preprocessing 10](#_Toc88211445)

[Model Selection 10](#_Toc88211446)

[Crossvalidation 10](#_Toc88211447)

[Metric Analysis 10](#_Toc88211448)

[Hyperparameter Search 10](#_Toc88211449)

# Apartat B: Comparativa de models

//bla bla bla, explicació apartat b

//alvaro

## Introducció a la base de dades

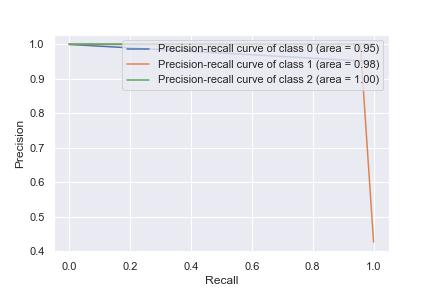
//alvaro

## Comparació de models

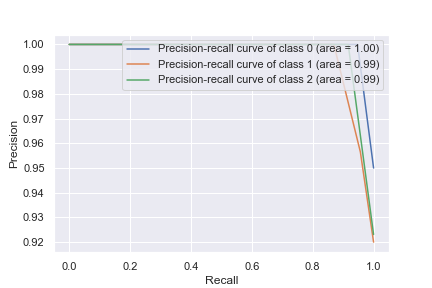
//alvaro i juankers

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Logistic | SVM | KNN | Decision Tree |
| 50% train | 0.9775 | 0.9887 | 0.9775 | 0.8988 |
| 80% train | 0.9722 | 0.9722 | 0.9166 | 0.8888 |
| 70% train | 0.9814 | 0.9629 | 0.9814 | 0.9629 |

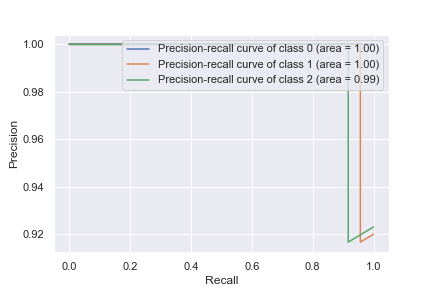
PR DECISION TREE



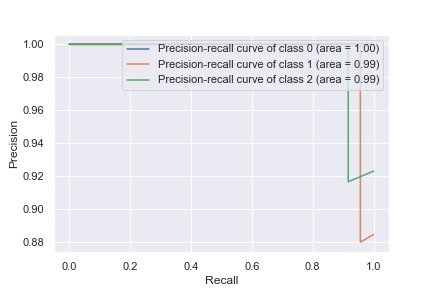
PR KNN



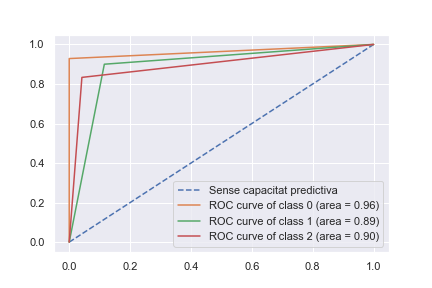
PR LOGISTIC REGRESSION



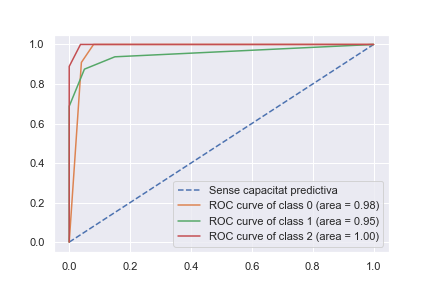
PR SVM



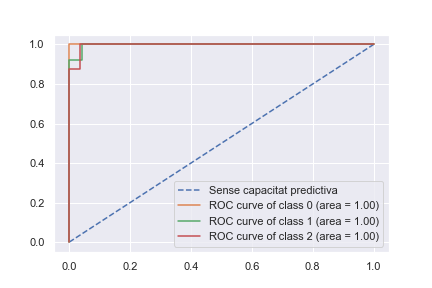
ROC DECISION TREE



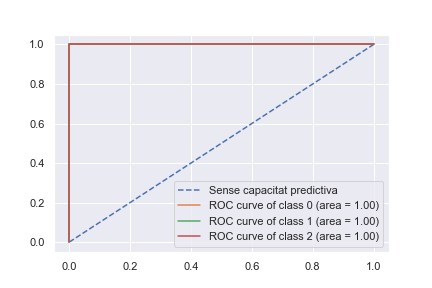
ROC KNN



ROC LOGISTIC REGRESSION



ROC SVM



Un dibujo de un perro

Descripción generada automáticamente con confianza baja

# Apartat A: Classificació Numèrica

//bla bla bla, descripcio de l’apartat A, taules obtigudes sobre les que analitzarem els resultats son aquestes:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| PERCEPTRÓ | Precisió | alpha | fit\_intercept | penalty | Tol |
| Bàsica |  |  |  |  |  |
| 50% train, 50% test | 0.706 | 0.0001 | True | None | 0.001 |
| 80% train, 20% test | 0.7175 | 0.0001 | True | None | 0.001 |
| 70% train, 30% test | 0.765 | 0.0001 | True | None | 0.001 |
|  |  |  |  |  |  |
| K-Fold |  |  |  |  |  |
| K = 2 | 0.713 | 0.0001 | True | None | 0.001 |
| K = 3 | 0.72101 | 0.0001 | True | None | 0.001 |
| K = 4 | 0.74 | 0.0001 | True | None | 0.001 |
| K = 5 | 0.702 | 0.0001 | True | None | 0.001 |
| K = 6 | 0.726 | 0.0001 | True | None | 0.001 |
|  |  |  |  |  |  |
| LOOCV | 0.727 | 0.0001 | True | None | 0.001 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| KNN | Precisió | leaf\_size | n\_neighbors | metric | weights |
| Bàsica |  |  |  |  |  |
| 50% train, 50% test | 0.506 | 30 | 5 | minkowski | uniform |
| 80% train, 20% test | 0.52 | 30 | 5 | minkowski | uniform |
| 70% train, 30% test | 0.51167 | 30 | 5 | minkowski | uniform |
|  |  |  |  |  |  |
| K-Fold |  |  |  |  |  |
| K = 2 | 0.4985 | 30 | 5 | minkowski | uniform |
| K = 3 | 0.5085 | 30 | 5 | minkowski | uniform |
| K = 4 | 0.5 | 30 | 5 | minkowski | uniform |
| K = 5 | 0.503 | 30 | 5 | minkowski | uniform |
| K = 6 | 0.5015 | 30 | 5 | minkowski | uniform |
|  |  |  |  |  |  |
| LOOCV | 0.5145 | 30 | 5 | minkowski | uniform |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Decision Tree | Precisió | criterion | splitter | max\_leaf\_nodes | max\_depth |
| Bàsica |  |  |  |  |  |
| 50% train, 50% test | 0.809 | gini | best | None | None |
| 80% train, 20% test | 0.81 | gini | best | None | None |
| 70% train, 30% test | 0.8 | gini | best | None | None |
|  |  |  |  |  |  |
| K-Fold |  |  |  |  |  |
| K = 2 | 0.8295 | gini | best | None | None |
| K = 3 | 0.8245 | gini | best | None | None |
| K = 4 | 0.8295 | gini | best | None | None |
| K = 5 | 0.821 | gini | best | None | None |
| K = 6 | 0.8265 | gini | best | None | None |
|  |  |  |  |  |  |
| LOOCV | 0.839 | gini | best | None | None |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Random Forest | Precisió | criterion | bootstrap | max\_leaf\_nodes | max\_depth |
| Bàsica |  |  |  |  |  |
| 50% train, 50% test | 0.864 | gini | True | None | None |
| 80% train, 20% test | 0.88 | gini | True | None | None |
| 70% train, 30% test | 0.88 | gini | True | None | None |
|  |  |  |  |  |  |
| K-Fold |  |  |  |  |  |
| K = 2 | 0.8675 | gini | True | None | None |
| K = 3 | 0.873 | gini | True | None | None |
| K = 4 | 0.879 | gini | True | None | None |
| K = 5 | 0.877 | gini | True | None | None |
| K = 6 | 0.8865 | gini | True | None | None |
|  |  |  |  |  |  |
| LOOCV | 0.881 | gini | True | None | None |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Regressió logística | Precisió | C | fit\_intercept | penalty | tol |
| Bàsica |  |  |  |  |  |
| 50% train, 50% test | 0.948 | 1.0 | True | None | 0.0001 |
| 80% train, 20% test | 0.965 | 1.0 | True | None | 0.0001 |
| 70% train, 30% test | 0.96 | 1.0 | True | None | 0.0001 |
|  |  |  |  |  |  |
| K-Fold |  |  |  |  |  |
| K = 2 | 0.954 | 1.0 | True | None | 0.0001 |
| K = 3 | 0.956 | 1.0 | True | None | 0.0001 |
| K = 4 | 0.9605 | 1.0 | True | None | 0.0001 |
| K = 5 | 0.9625 | 1.0 | True | None | 0.0001 |
| K = 6 | 0.9635 | 1.0 | True | None | 0.0001 |
|  |  |  |  |  |  |
| LOOCV | 0.966 | 1.0 | True | None | 0.0001 |

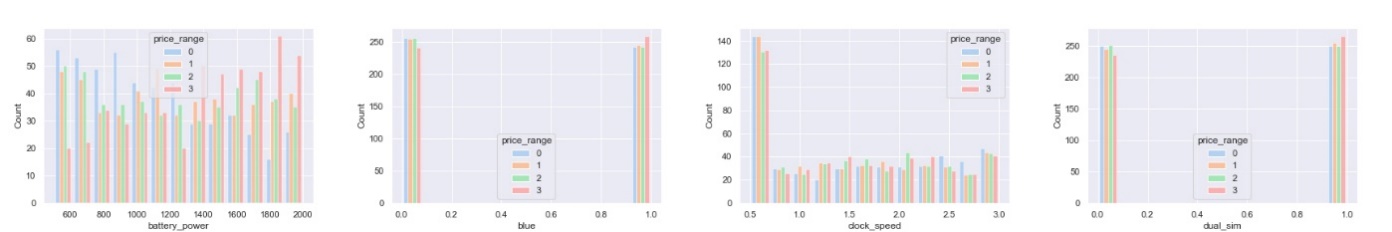
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| SVM | Precisió | C | kernel | decision\_function\_shape | tol |
| Bàsica |  |  |  |  |  |
| 50% train, 50% test | 0.854 | 1.0 | rbf | ovr | 0.001 |
| 80% train, 20% test | 0.855 | 1.0 | rbf | ovr | 0.001 |
| 70% train, 30% test | 0.85167 | 1.0 | rbf | ovr | 0.001 |
|  |  |  |  |  |  |
| K-Fold |  |  |  |  |  |
| K = 2 | 0.8635 | 1.0 | rbf | ovr | 0.001 |
| K = 3 | 0.868 | 1.0 | rbf | ovr | 0.001 |
| K = 4 | 0.886 | 1.0 | rbf | ovr | 0.001 |
| K = 5 | 0.885 | 1.0 | rbf | ovr | 0.001 |
| K = 6 | 0.885 | 1.0 | rbf | ovr | 0.001 |
|  |  |  |  |  |  |
| LOOCV | 0.888 | 1.0 | rbf | ovr | 0.001 |

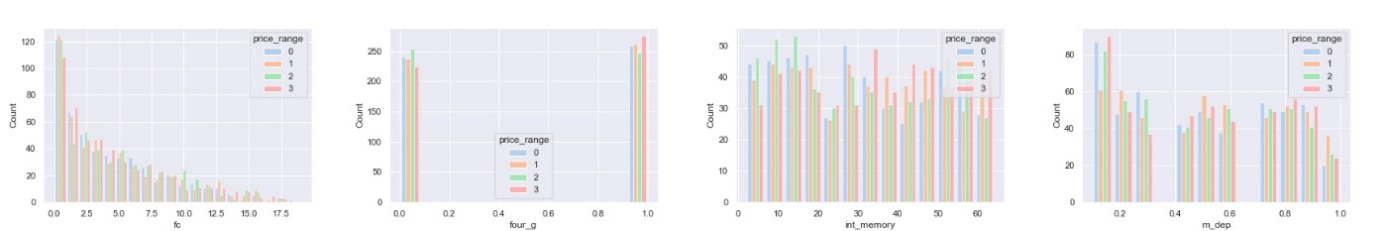
## Introducció a la base de dades

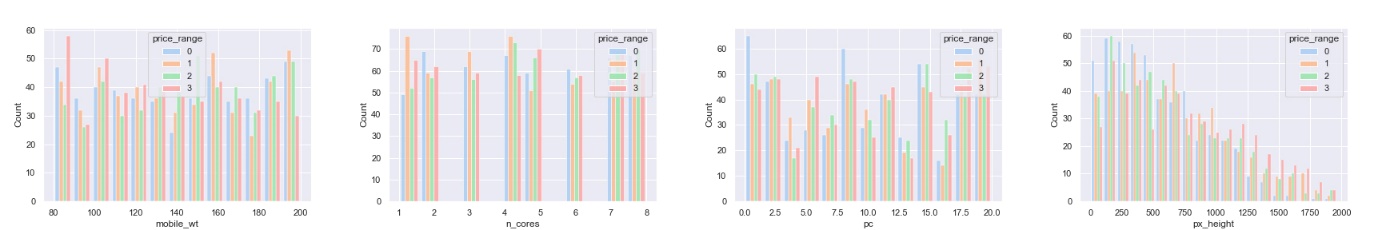
//la base de dades es la de mòbils

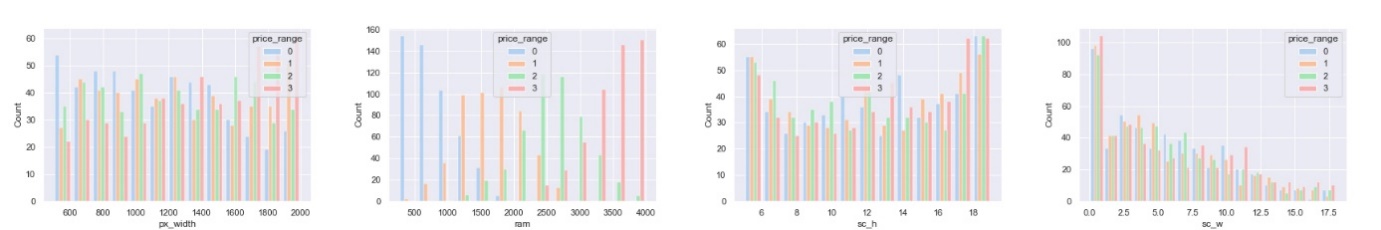
Per a poder fer un estudi adient de la nostra base de dades, fem una visualització dels histogrames i les gràfiques de dispersió de cada atribut. D’aquesta manera entendrem millor amb quines dades estem treballant. Les gràfiques són les següents:

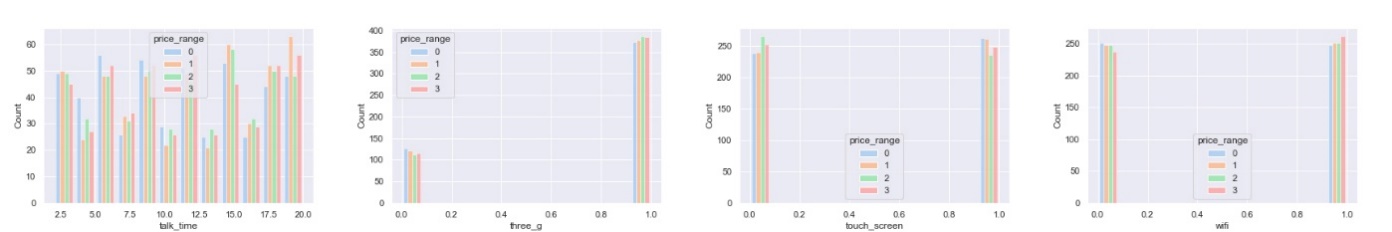
### HISTOGRAMES DELS ATRIBUTS

Histogrames dels atributs 1-4

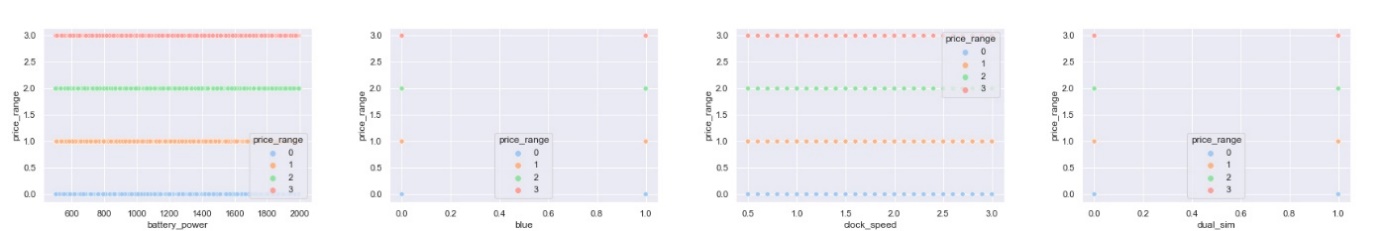
Histogrames dels atributs 5-8

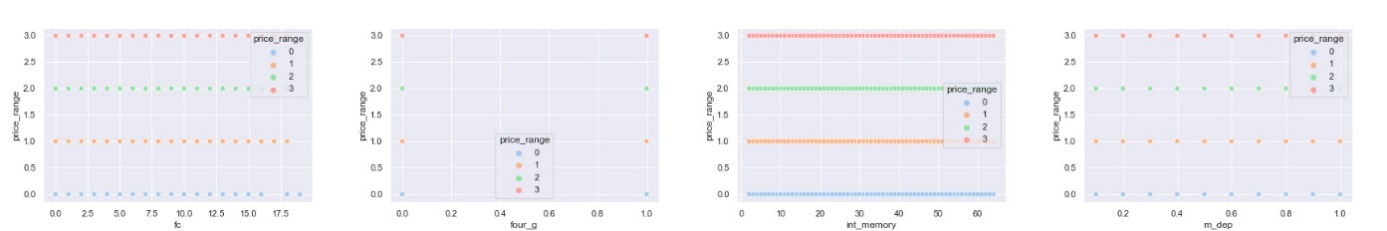
Histogrames dels atributs 9-12

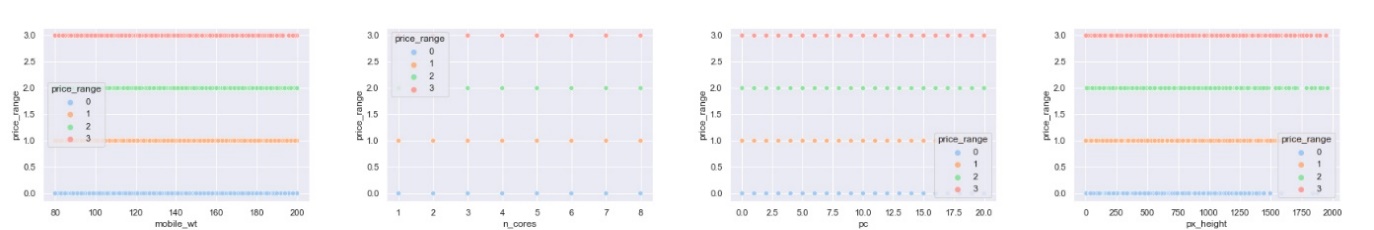
Histogrames dels atributs 13-16

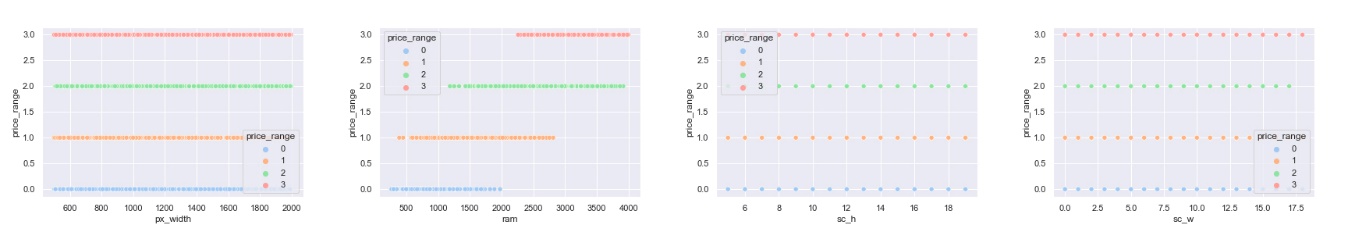
Histogrames dels atributs 17-20

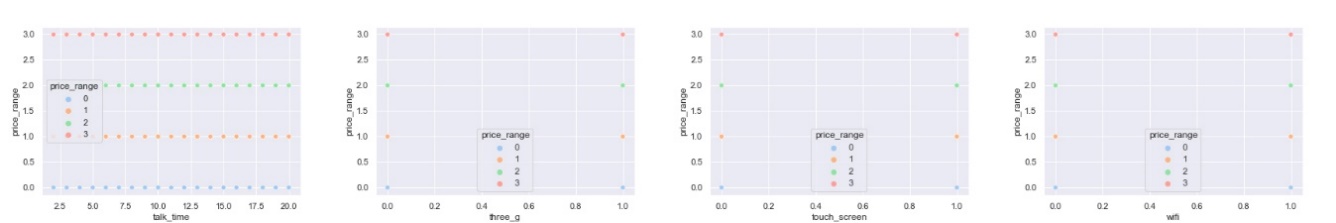
### GRÀFIQUES DE DISPERSIÓ

Dispersió dels atributs 1-4

Dispersió dels atributs 5-8

Dispersió dels atributs 9-12

Dispersió dels atributs 13-16

Dispersió dels atributs 17-20

## Preprocessing

//victor

## Model Selection

//victor

## Crossvalidation

//victor

## Metric Analysis

//juankers

## Hyperparameter Search

//juankers