

Entregable WEKA

Laura Rodríguez Navas

Marzo 2020

Preparación de Datos

Quitar valores missing
aleatorizar la base de datos
crear conjunto de entrenamiento

Uso de los clasificadores

cross-validation (5cv)

NaiveBayes

col	Accuracy	F-measure
NaiveBayes	cell5	cell6
C4.5	cell8	cell9

J48 (C4.5)

Técnicas de preprocesamiento

Selección de variables
AttributeSelectedClassifier
FilteredClassifier

Visualitization

Confusion matrix ROC

1. inicialmente lanzaremos los clasificadores sobre la base de datos inicial, obteniendo así una medida inicial que intentaremos mejorar.
2. mejorar con la selección de variables
de variables es no lanzar una selección de tipo multivariada (filter o wrapper) sobre todas las variables, puesto que podría llevar mucho tiempo. Se recomienda primero hacer una selección univariada tipo ranker y quedarse con los mejores 200-500 atributos.