UNIVERSIDAD
POLITÉCNICA
DE MADRID

Escuela Técnica
Superior de
Ingenieros
Informáticos

Practical Application 1
Machine Learning

Jan Cerezo Pomykol
j.cerezo@alumnos.upm.es

Problem Description

Dry Bean Dataset:

- 13611 instances
- 16 variables
- 7 classes

https://archive.ics.uci.edu/ml/datasets/Dry+Bean+Dataset#

Area
Perimeter
Major axis length
Minor axis length
Aspect ratio
Eccentricity
Convex area
Equivalent diameter
Extent
Solidity
Roundness
Compactness
ShapeFactor1
ShapeFactor2
ShapeFactor3
ShapeFactor4

Seker
Barbunya
Bombay
Cali
Dermosan
Horoz
Sira

Methodology

- Software: **Weka**
- Classification algorithms:
  - k-Nearest Neighbour
  - Rule Induction – RIPPER
  - Support Vector Machine
  - Neural Network
  - Classification Tree – C4.5
- Feature Subset Selection:
  - No FSS
  - Univariant Filter
  - Multivariant Filter
  - Wrapper Approach

Algorithm	Weka Function
k-Nearest Neighbours	lazy.IBk
Rule Induction (RIPPER)	rules.JRip
Support Vector Machine	functions.SMO
Neural Network	functions.MultilayerPerceptron
Classification Tree (C4.5)	trees.J48

FSS algorithm	Weka Function
No FSS	-
Univariant Filter	attributeSelection.InfoGainAttributeEval
Multivariant Filter	attributeSelection.CfsSubsetEval
Wrapper Approach	attributeSelection.WrapperSubsetEval

Results

Selected attributes

Attribute	No FSS	Univariant	Multivariant	Wrapper (RIPPER)	Wrapper (kNN k=7)	Wrapper (SVM)	Wrapper (MLP)	Wrapper (C4.5)
Area	•	•						•
Perimeter	•	•	•	•	•	•	•	
MajorAxisLength	•	•	•					
MinorAxisLength	•	•	•				•	
AspectRatio	•		•				•	
Eccentricity	•				•		•	
ConvexArea	•	•	•			•		•
EquivDiameter	•	•		•		•		•
Extent	•		•			•	•	
Solidity	•			•	•	•	•	•
Roundness	•		•	•	•	•	•	•
Compactness	•		•	•		•	•	•
ShapeFactor1	•	•	•	•	•	•	•	
ShapeFactor2	•	•	•		•		•	•
ShapeFactor3	•			•				
ShapeFactor4	•		•	•	•	•	•	•
N attributes	16	8	11	8	7	9	11	8

Results

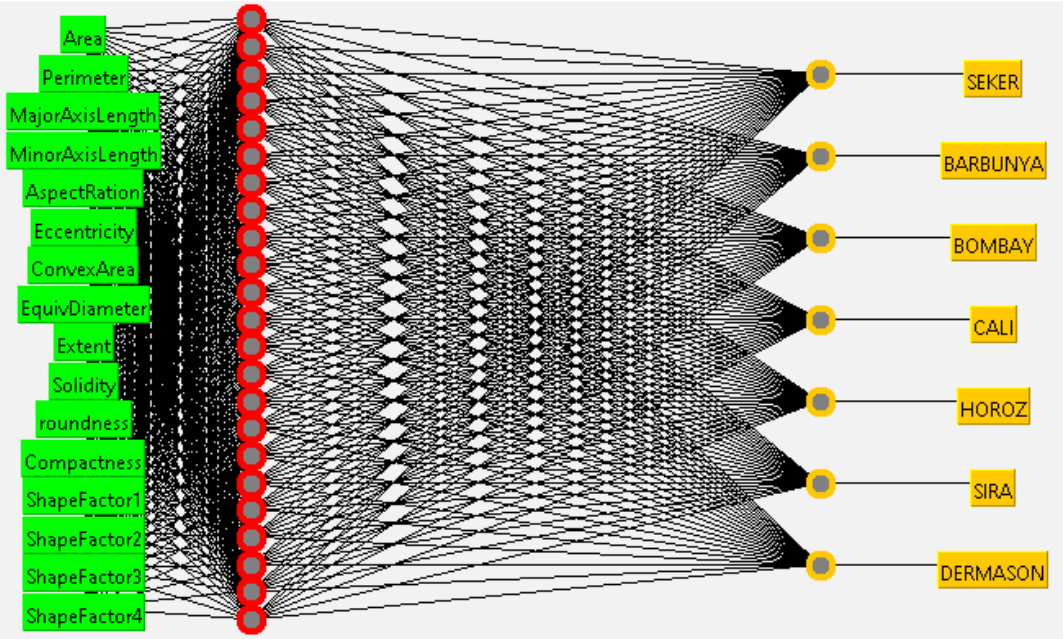
Classifier scores

Dataset	kNN	RIPPER	SVM	MLP	C4.5
Original	91.9477	91.4775	92.5648	92.7852	91.4995
Univariate Filter	90.8897	90.2432	89.9052	91.2718	90.3681
Multivariate Filter	92.1901	91.1395	92.4987	92.5428	91.0073
Wrapper (kNN)	92.4840	91.4114	92.462	92.5428	91.3967
Wrapper (RIPPER)	92.5428	91.4261	92.4106	92.4693	91.4628
Wrapper (SVM)	92.3077	91.4187	92.5281	92.7044	91.3452
Wrapper (MLP)	92.0065	91.4334	92.5648	92.6971	91.3085
Wrapper (C4.5)	92.4326	91.4628	92.3297	92.4546	91.4040

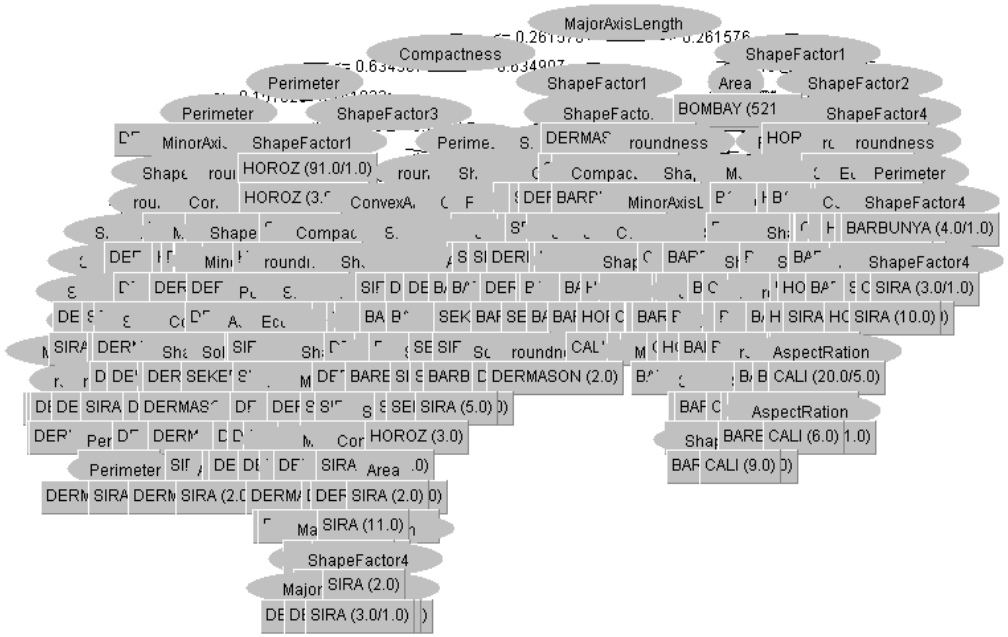
Training time

Dataset	kNN	RIPPER	SVM	MLP	C4.5
Original	<0.1	2.38	7.98	33.37	0.3
Univariate Filter	<0.1	1.89	12.11	14.87	0.1
Multivariate Filter	<0.1	2.43	6.16	20.67	0.18
Wrapper (kNN)	<0.1	1.78	5.58	13.25	0.11
Wrapper (RIPPER)	<0.1	1.67	6.82	14.89	0.12
Wrapper (SVM)	<0.1	1.74	6.13	16.64	0.14
Wrapper (MLP)	<0.1	1.91	6.71	20.56	0.17
Wrapper (C4.5)	<0.1	1.99	6.33	15.19	0.14

Results



Multilayer Perceptron Network



C4.5 Tree

UNIVERSIDAD
POLITÉCNICA
DE MADRID

Escuela Técnica
Superior de
Ingenieros
Informáticos

Practical Application 1
Machine Learning

Jan Cerezo Pomykol
j.cerezo@alumnos.upm.es