INF FAKULTÄT FÜR INFORMATIK

Advanced Topics in Machine Learning Programming Task

Jannis Becke, Christian Lausberger and Maik Riestock

Advanced Topics in Machine Learning

June 22, 2015

Data Preparation

49 Attributes, 101.000 Instances

Reasons for attribute removal:

1. **Uniqueness**
2. **Missing Values**
3. **Unequal Distribution**
4. **Single Attribute**

No instance was removed

Choice of Algorithms

Algorithms:

- Naive Bayes
- SVM
- YATSI
- Decision Table

Implementation

Choice of libraries:

- weka
- libSVM
- collectiveClassifiers

Settings

Decision Table:

- Search: Best First
- Evaluation Measure: accuracy/RMSE

SVM:

- Type: C-SVM
- Kernel: radial basis function
- with normalizing, no weights

YATSI:

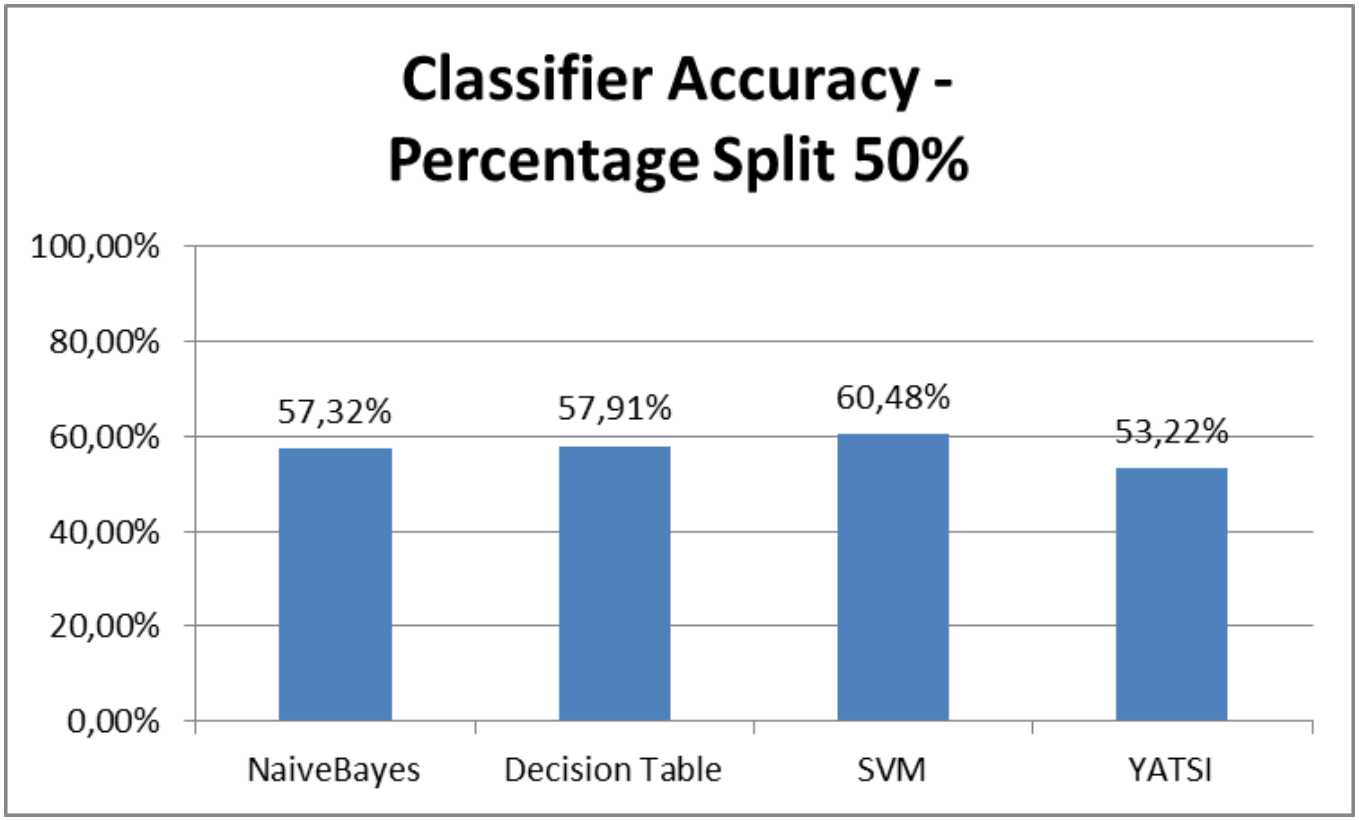
- KNNpe: 15

Experiments

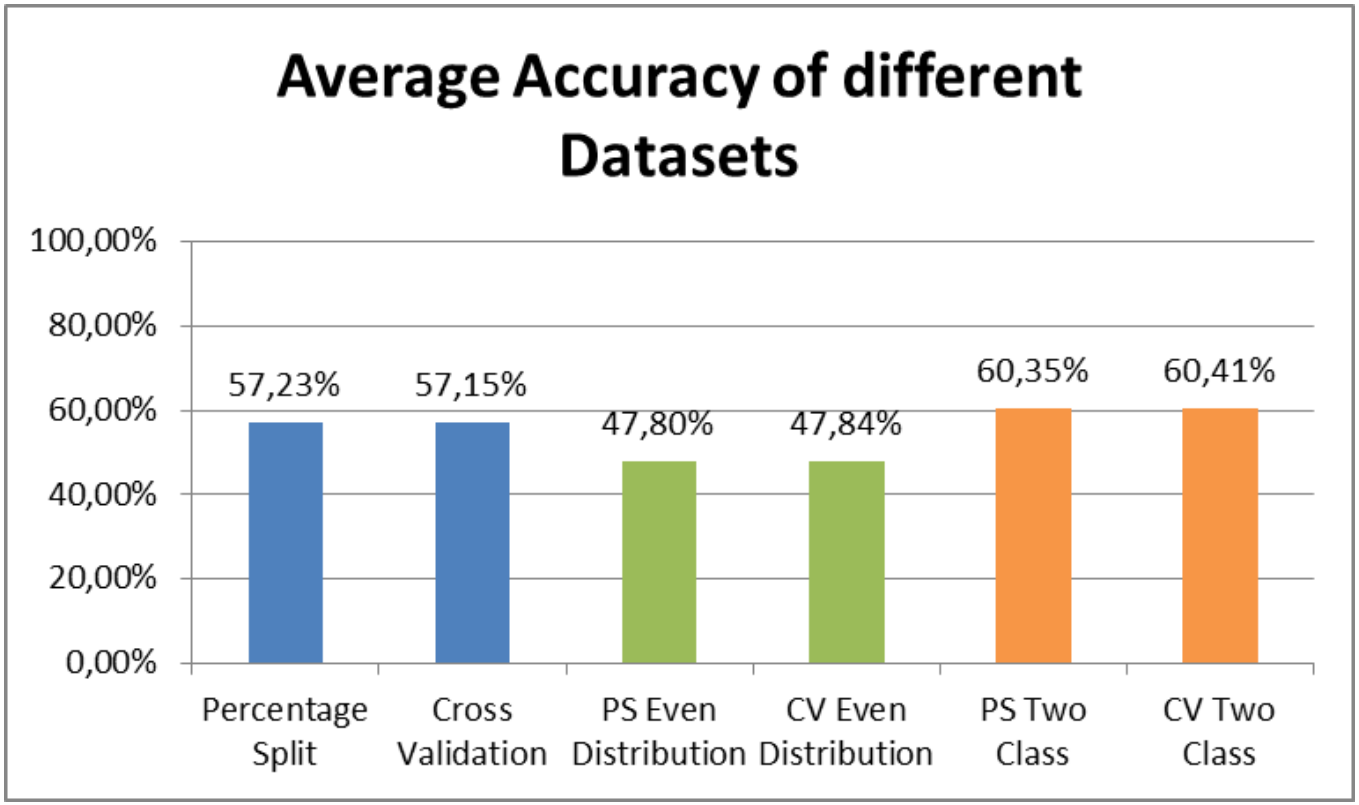
Settings:

- Percentage Split / Cross-Validation
- Accuracy of classification as quality measure
- two additional data sets:
  - even class dataset (equally distributed classification attribute)
  - two class dataset (join label > 30, < 30)

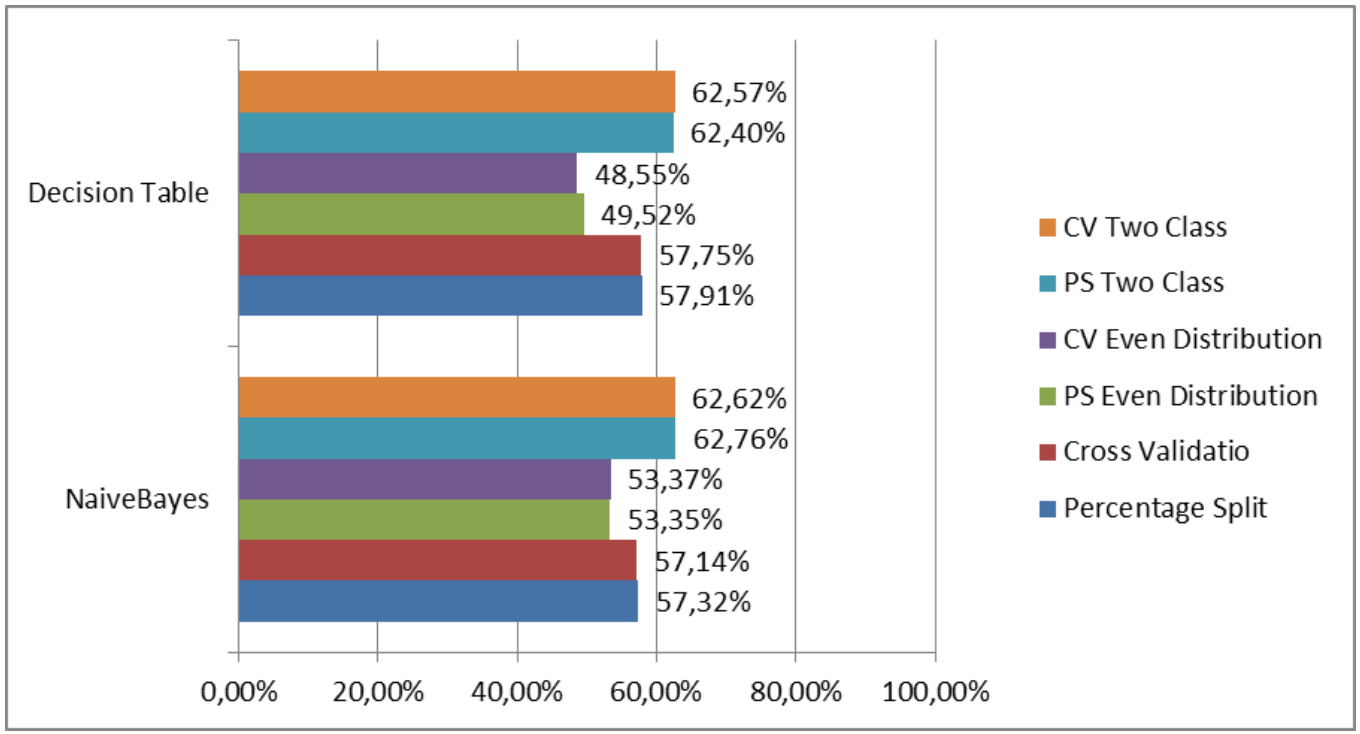
Results

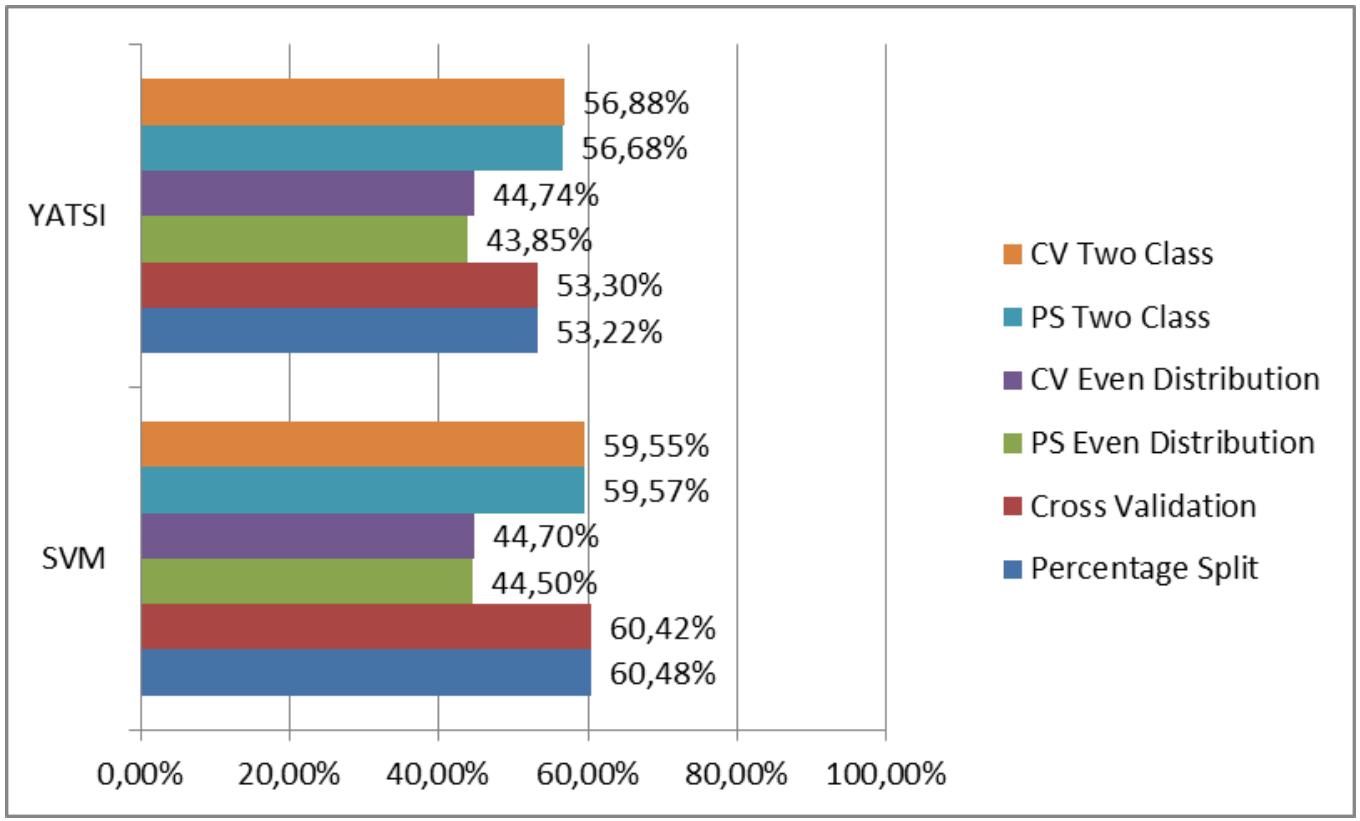

Classifier Accuracy - Percentage Split 50%
100,00%
80,00%
60,00%
40,00%
20,00%
0,00%
57,32%
57,91%
60,48%
53,22%
NaiveBayes
Decision Table
SVM
YATSI

Results


Average Accuracy of different Datasets
100,00%
80,00%
60,00%
40,00%
20,00%
0,00%
57,23%
57,15%
47,80%
47,84%
60,35%
60,41%
Percentage Split
Cross Validation
PS Even Distribution
CV Even Distribution
PS Two Class
CV Two Class

Thanks for Your attention!


Decision Table
NaiveBayes
62,57%
62,40%
48,55%
49,52%
57,75%
57,91%
62,62%
62,76%
53,37%
53,35%
57,14%
57,32%
0,00%
20,00%
40,00%
60,00%
80,00%
100,00%
CV Two Class
PS Two Class
CV Even Distribution
PS Even Distribution
Cross Validatio
Percentage Split


YATSI
SVM
56,88%
56,68%
44,74%
43,85%
53,30%
53,22%
59,55%
59,57%
44,70%
44,50%
60,42%
60,48%
0,00%
20,00%
40,00%
60,00%
80,00%
100,00%
CV Two Class
PS Two Class
CV Even Distribution
PS Even Distribution
Cross Validation
Percentage Split