Current state and future challenges in Optional Weaving

Constanze Michaelis

Student Conference on Software Engineering and Database Systems

June 27, 2009

FAKULTÄT FÜR
INFORMATIK

INF

OTTO VON GUERICKE
UNIVERSITÄT
MAGDEBURG

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
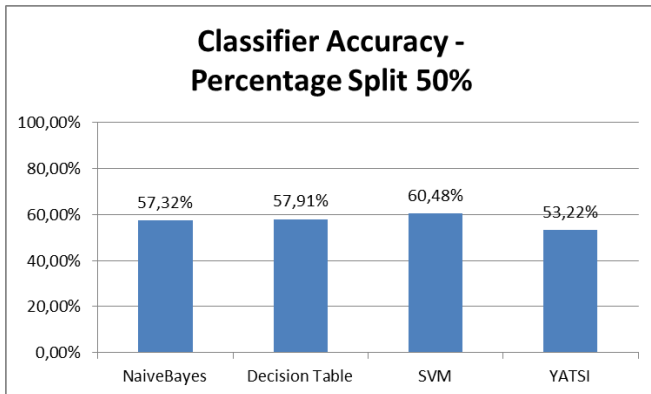
Implementation

Choice of libraries:

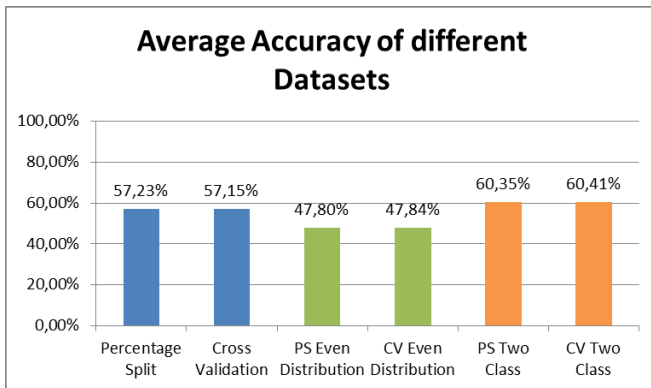- weka
- libSVM
- collectiveClassifiers

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%

Classifier	Accuracy
NaiveBayes	57,32%
Decision Table	57,91%
SVM	60,48%
YATSI	53,22%

Results

Average Accuracy of different Datasets

Dataset	Accuracy
Percentage Split	57,23%
Cross Validation	57,15%
PS Even Distribution	47,80%
CV Even Distribution	47,84%
PS Two Class	60,35%
CV Two Class	60,41%

Thanks for Your attention!

OTTO VON GUERICKE
UNIVERSITÄT
MAGDEBURG

INF

FAKULTÄT FÜR
INFORMATIK