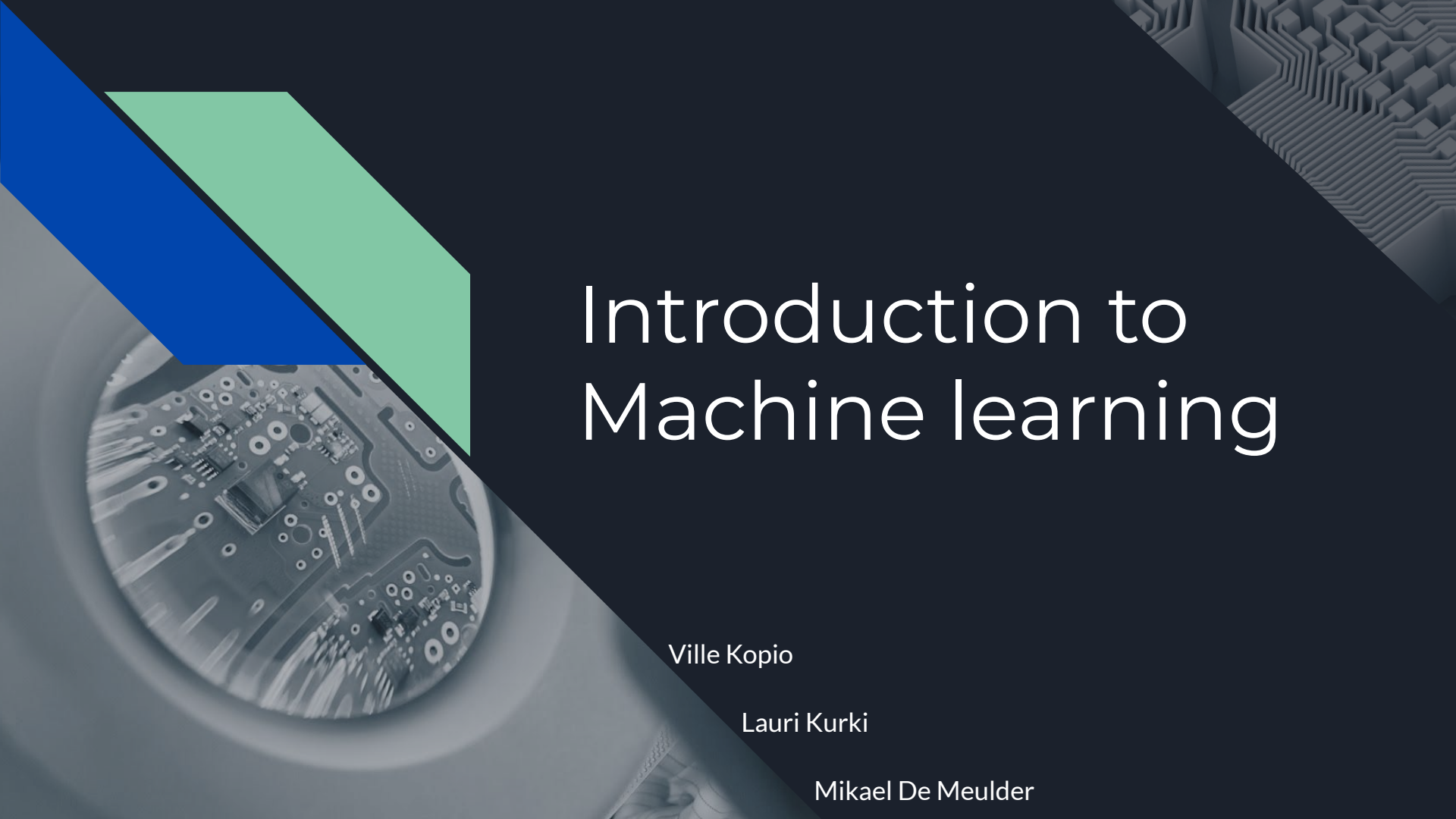


Introduction to Machine learning

Ville Kopio

Lauri Kurki

Mikael De Meulder

TOC

Binary Classifier

Target

- Classify the particle formations to Events and Non-events

Multiclass Classifier

Target

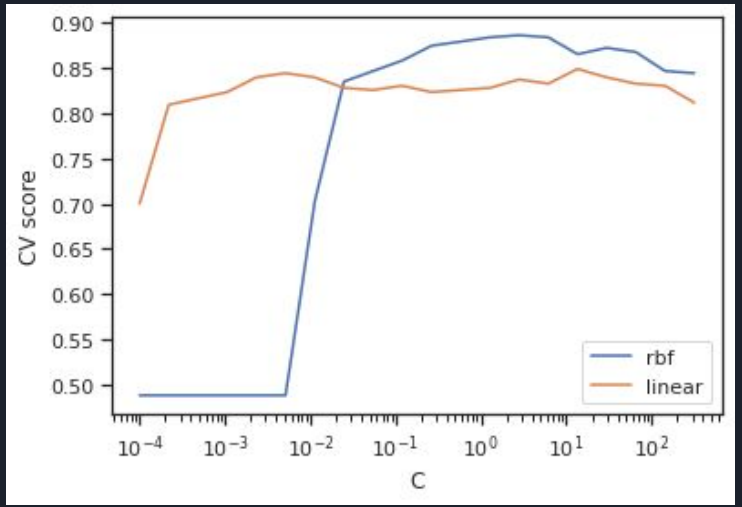
- Classify the different event to Non-events and the different particle formation events
(Ia, Ib and II)

Binary Classifier

- Model selection:
  - Logistic Regression (LR)
  - Support Vector Classifier (SVC)
  - Random Forest Classifier (RFC)
  - Naive Bayes (NB)
- Comparison done with 10-fold cross-validation

Model	CV loss
LR	0.819
SVC	**0.884**
RFC	0.879
NB	0.807

- Optimization of model parameters:
  - Kernel
  - Regularization parameter C
- Kernel = rbf, C = 2.2



Multiclass Classifier

- Prosessing the std's out and doing the normalization
  - Unsupervised learning methods
    - K-means
    - Hierarchical clustering
  - Supervised learning methods
    - Multiple different
  - In both learning types affect of feature selection was done by sklearns selectKBest function
  - Also the effect of different scaling methods was tested

Multiclass Classifier

Unsupervised

K-Means	Hierarchical	Scaling
0.61	0.57	Normal
2.57	0.59	MinMax
0.59	0.57	Standard
0.58	0.58	No scaling

K-mean with optimized features: 0.59

Unsupervised

75/25 test/train split

LR	**0.70**
Decision tree	0.64
K-NN	0.67
LDA	0.64
NB	0.55
SVC	0.70

Decided to use Logistic Regression