



# Gruppenarbeit: Vorhersage einer «buy, hold, sell» - Strategie

## Gruppe 1

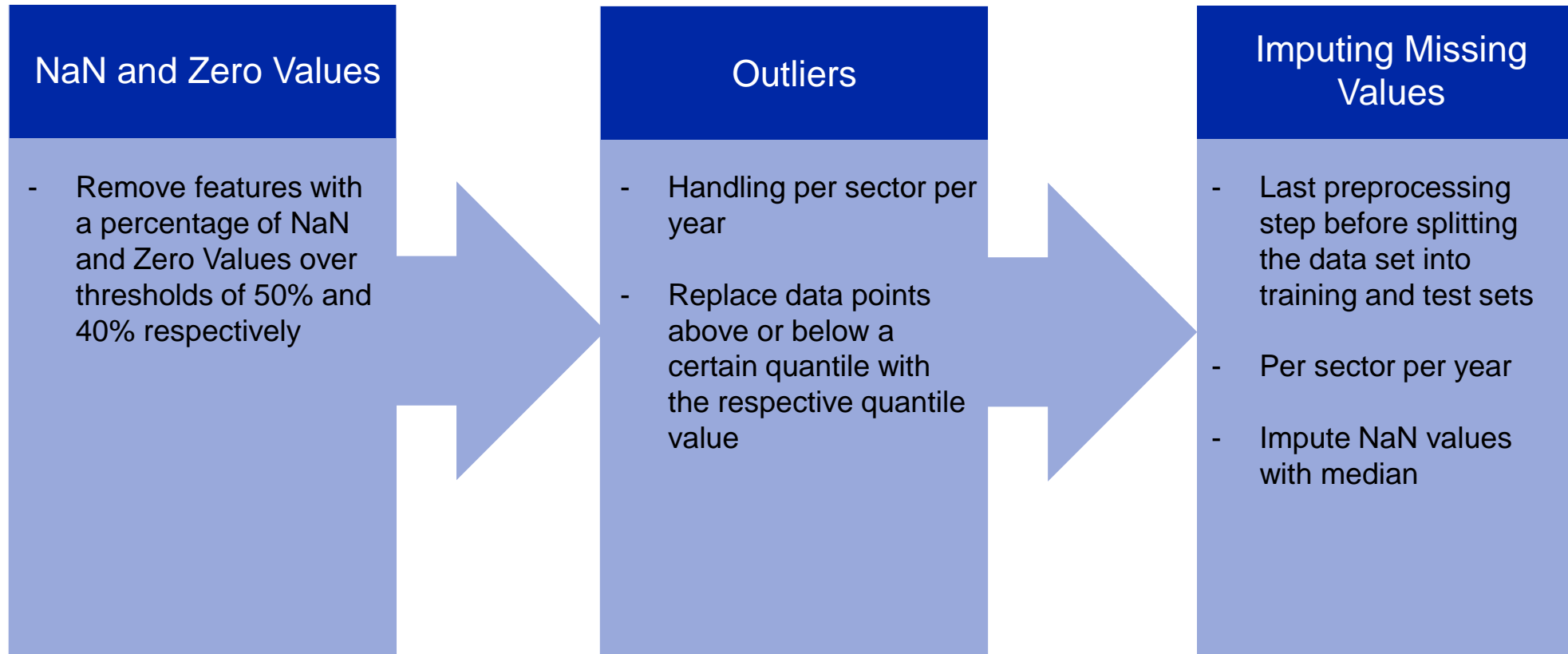
Tim Ehrensperger, Marco Heiniger, Pascal Huser, Marc Tschudi



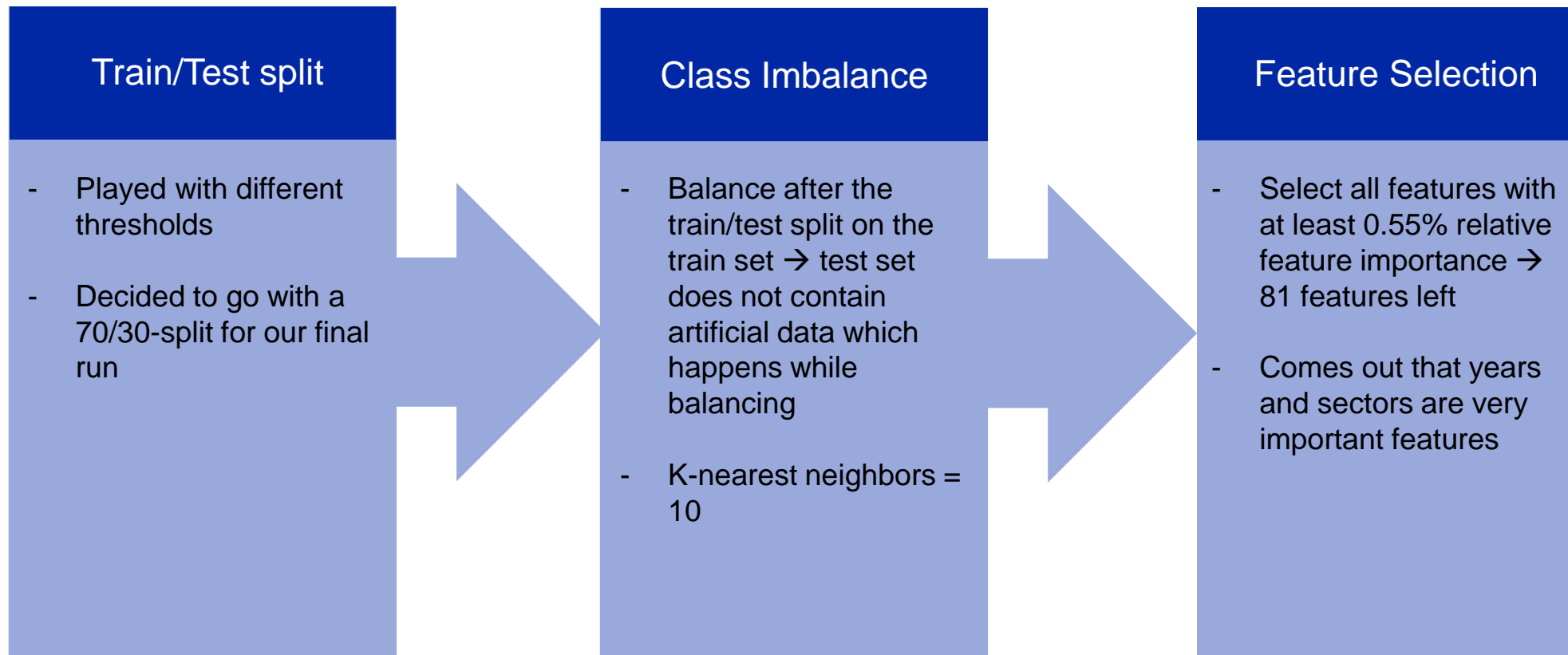
## Inhaltsverzeichnis

- **Data Preprocessing**
- **Algorithms**
- **Results**
- **Conclusion**

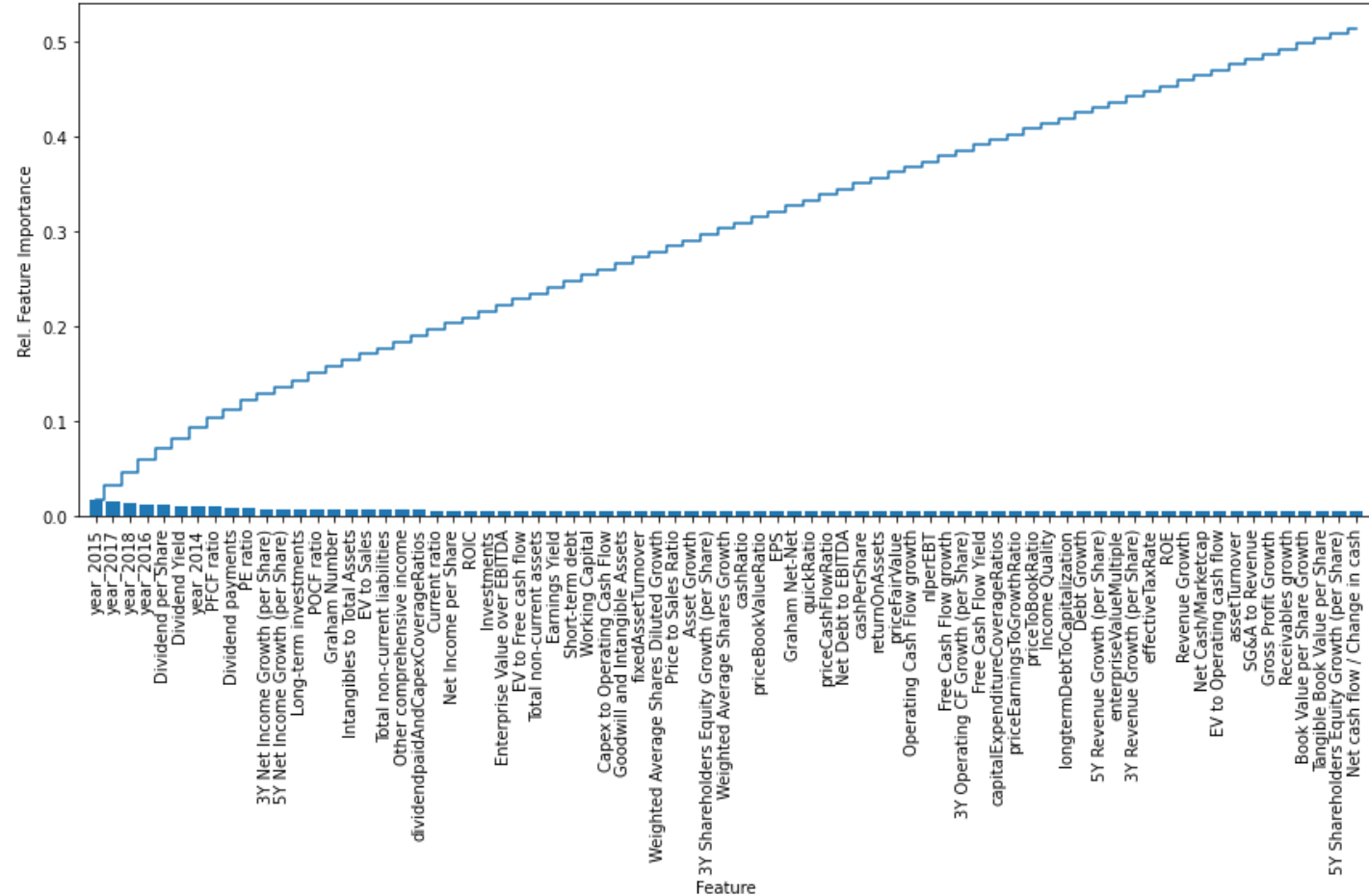
## Data Preprocessing I



## Data Preprocessing II



## Feature Selection



## Algorithms

Random Forest / Decision Tree	SVM	LDA / QDA	Keras Sequential Model
<ul style="list-style-type: none"><li>- Intuitive classification process</li><li>- Flexible</li><li>- Two of the most used algorithms for classification problems</li></ul>	<ul style="list-style-type: none"><li>- Rather disappointing score</li><li>- Runtimes went into hours → not 100 percent sure if we have found the best parameters</li><li>- Unlikely that the score would improve onto the level of Random Forest</li></ul>	<ul style="list-style-type: none"><li>- Dimensionality reduction technique</li><li>- Reduces high-dimensional data sets onto a lower-dimensional space</li><li>- Less computational costs because of lower dimension</li><li>- LDA assumption of a common covariance matrix</li></ul>	<ul style="list-style-type: none"><li>- Deep learning framework</li><li>- Most used framework among top-5 winning teams on Kaggle</li><li>- Layer-by-Layer model creation</li><li>- Three layers resulted in the best score</li></ul>



## Algorithms – Results

Method	Best Score
Random Forest	61.44%
Decision Tree	60.39%
LDA and QDA	56.60%
Keras Neural Network	56.58%
Support Vector Machine	41.09%



## Conclusion

- At first, we asked ourselves what accuracy score we would be aiming for → 65-75%
- Just predicting «sell» would have resulted in a score of 53% → we managed to push the score up by predicting «buy» and «hold» in an acceptable manner
- We are convinced that if we would get real world data we would achieve similar scores → we didn't artificially improve our test score by removing too many outliers
- More computational power would be useful for applying the GridSearch algorithm and trying out different threshold levels for the algorithms → threshold sensitivity as room for improvement
- Data preprocessing is a very important step and with a looser time restriction different ways of handling a task can be deeper investigated
- Feature selection as an important component