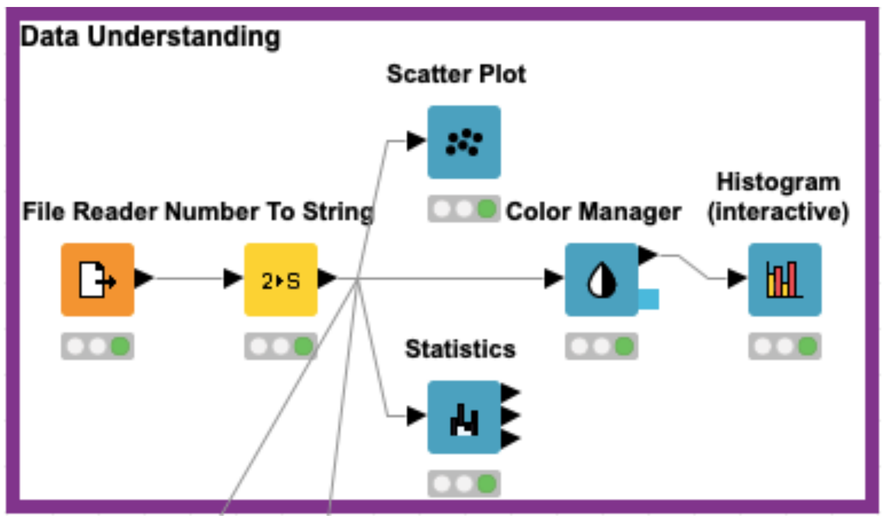
Group 1: Phone Prices

Vincent Kretzschmar, Suhas Shantharam, Asma Mohammed, Munasif Hassan, Showmik Das

Data Understanding

- No missing values
- However: violations of semantic accuracy
- Not many features seem to have a large impact on the price range except for **ram**

→ Seems to be the most important attribute →



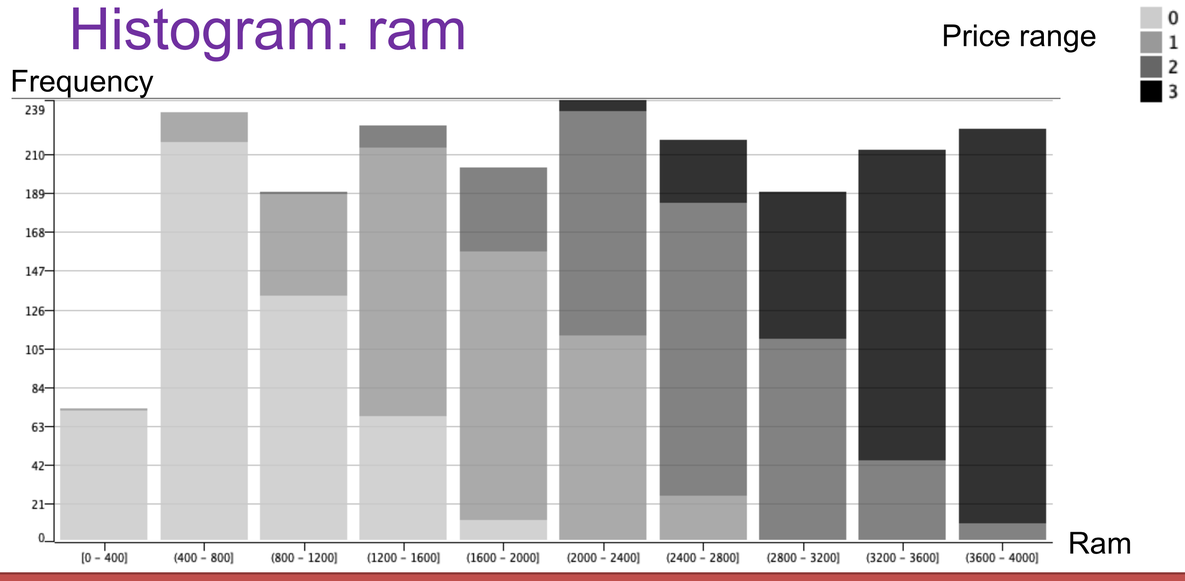
Challenge

- Predicting the **price category** of phones

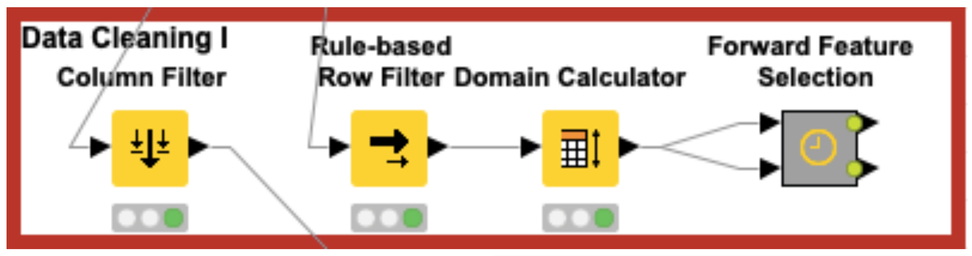
→ 0 to 3

Dataset

- Describes various attributes of mobile phones
- 2000 records with 21 different features

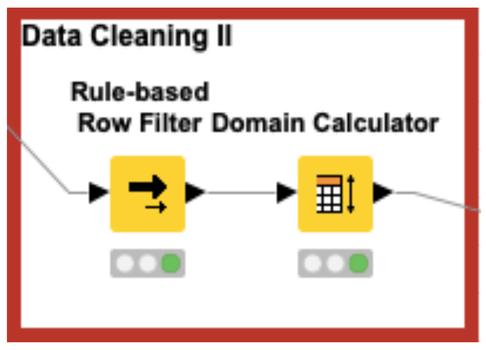


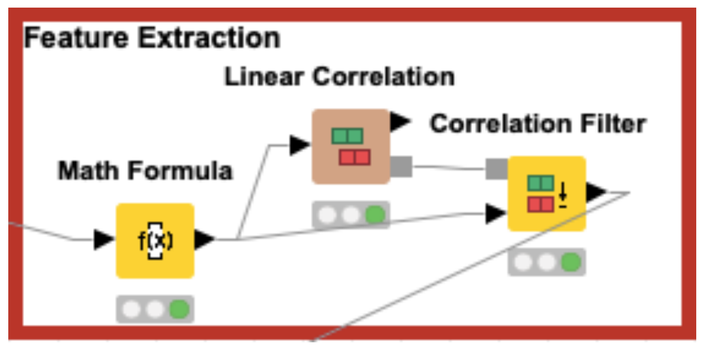
Data Preparation



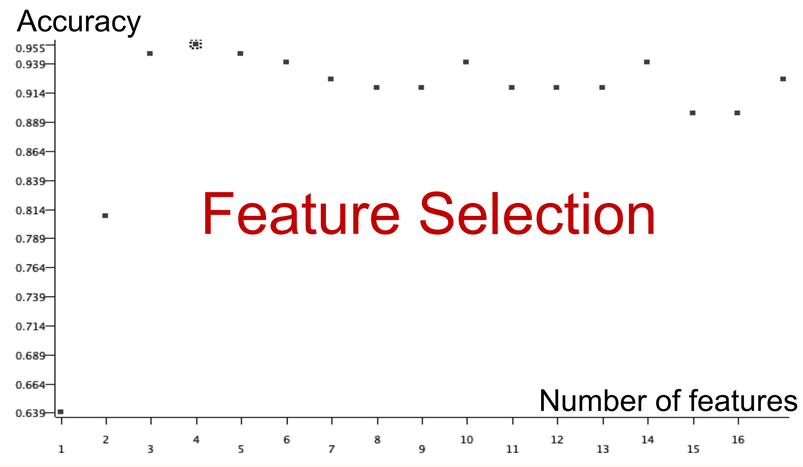
- Thought of applying the Rule-based Row Filter to overcome potential problems concerning semantic accuracy (screen width and screen height) → as too many rows would be eliminated, the Forward Selection node is applied to check if the accuracy is high enough without those features & the columns were eliminated instead

- In contrast to screen width & height, pixel resolution helps predicting the price range → application of the Rule-based Row Filter to eliminate rows where pixel resolution height < 360

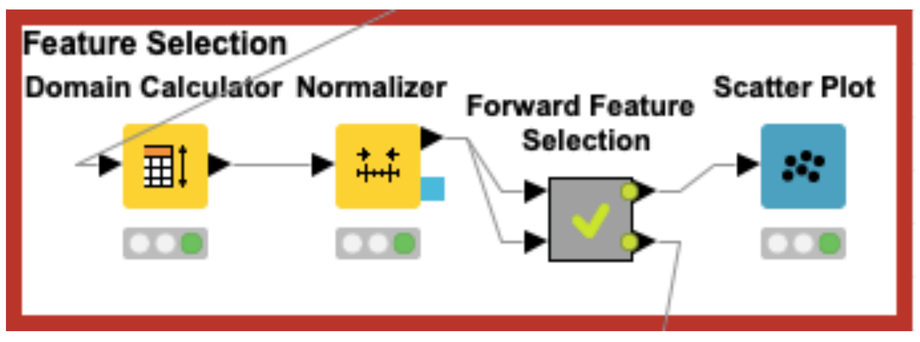




- Further reduction of features:
  - ➢ New feature: total pixel resolution = pixel resolution height * pixel resolution width



Feature Selection



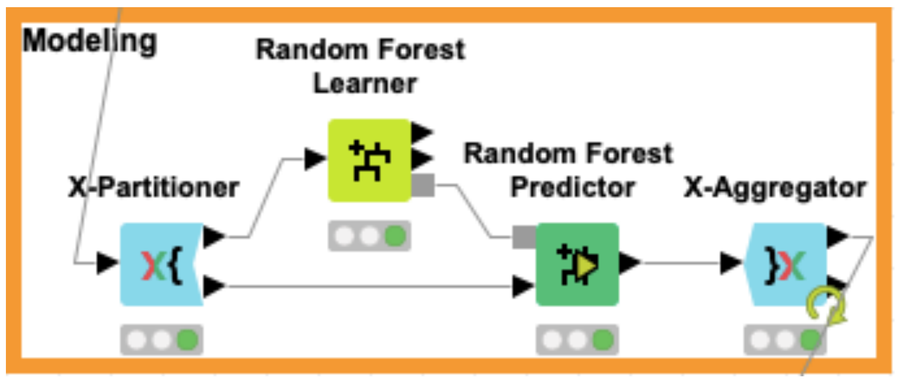
Accuracy	Nr. of features
0.956	4

- Feature selection: since not all features seemed helpful, the most important features needed to be found
  → **Forward Feature Selection:**
- **4 features** with an accuracy of 95.6 %
  → selected features: **ram, total pixel resolution, battery power, Bluetooth**

Modeling

Model	Accuracy	F-measure
Random Forest Classifier	**0.908**	**0.903**
Naïve Bayes Classifier	0.796	0.797
Decision Tree Model	0.849	0.849

- Use of **cross-validation** to protect against overfitting and to maximize the use of the data
- different classification models were built for comparison
- Most important quality measure: Accuracy
- Final model: **Random Forrest Classifier** (Ensemble Classifier) →provides the advantage of a higher efficiency



Conclusion

- 21 features in the beginning → 4 features are now used to predict the price category with an accuracy of 90.8%
- If a phone has a high ram value → very likely to fall into the highest price category

Results

- Accuracy: **90.768%**
- Cross-Validation: **10 folds**