Group 30: Animal Shelter

Abhinav Srivastava, Albert Mäkinen, Harsh Rajeshbhai Solanki, René Degenkolbe

Challenge

Challenge was to try to predict the outcome of cats and dogs in an animal shelter.

Dataset

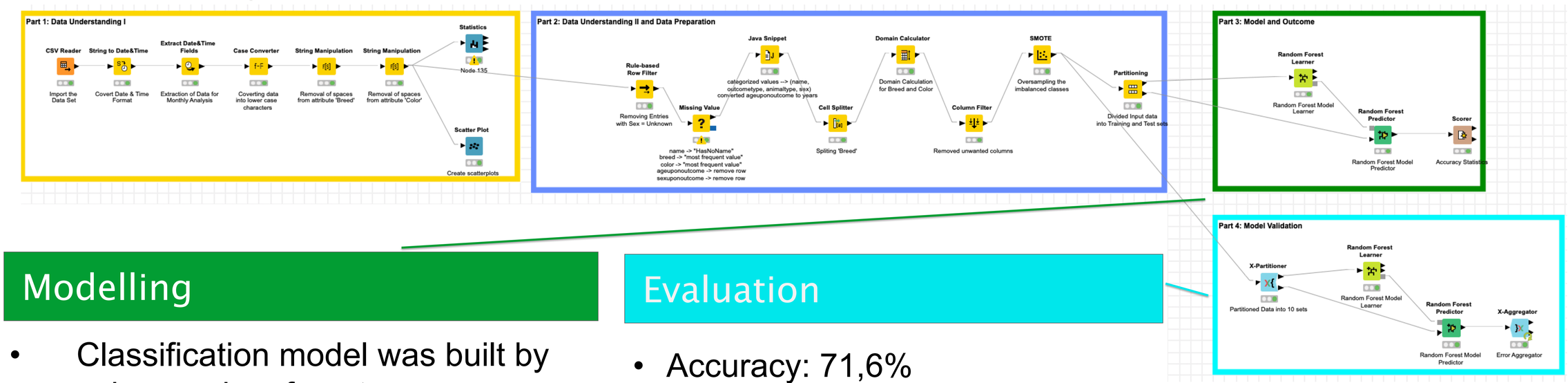
- Dataset consisted of 20000 rows and 9 columns.
- Dogs and cats, their breed, color, sex, age, name and outcome type
- Outcomes: Adopted, returned to owner, Euthanasia, Transfer, Died

Data Understanding & Preparation

- Converting date & time, removing spaces breed and color
- Extracting date & time for monthly analysis
- Overall basic data preparation and string manipulation
- There was trends in adoption between months

Data Understanding & Preparation

- Further manipulation that is essential to make the model work
- New variables like ”Hasnoname” and sex = ”unknown” were created
- Combining values in breed
- SMOTE; oversampling the imbalanced classes
- Partitioning



Modelling

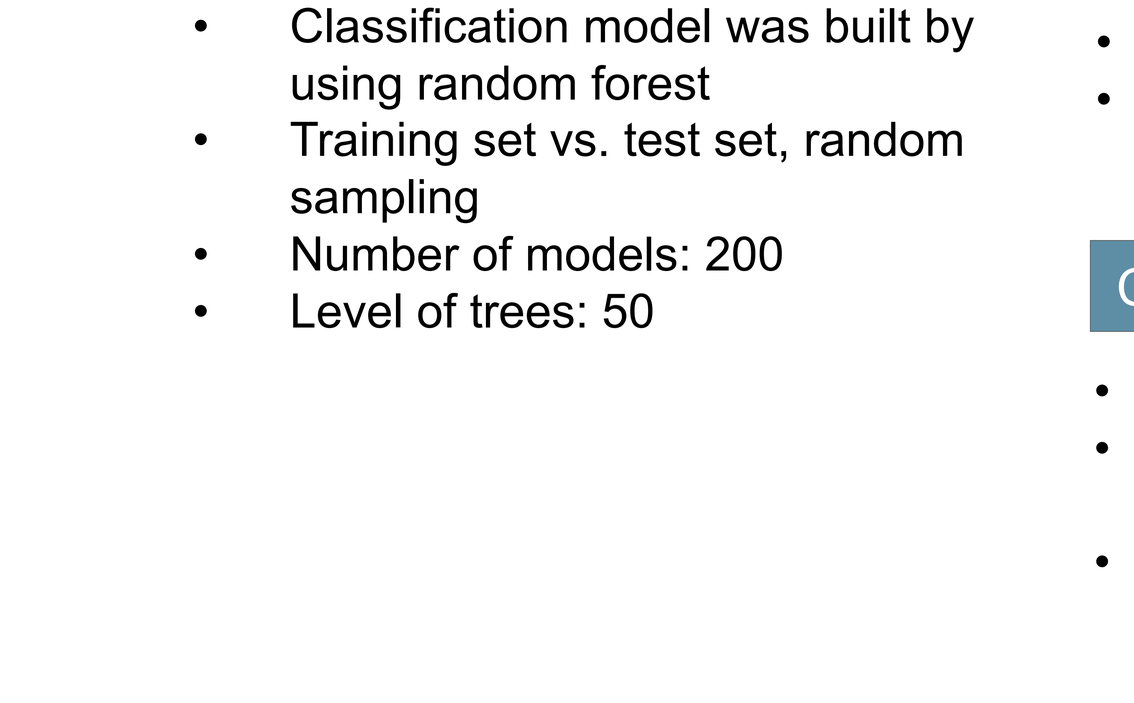
- Classification model was built by using random forest
- Training set vs. test set, random sampling
- Number of models: 200
- Level of trees: 50

Evaluation

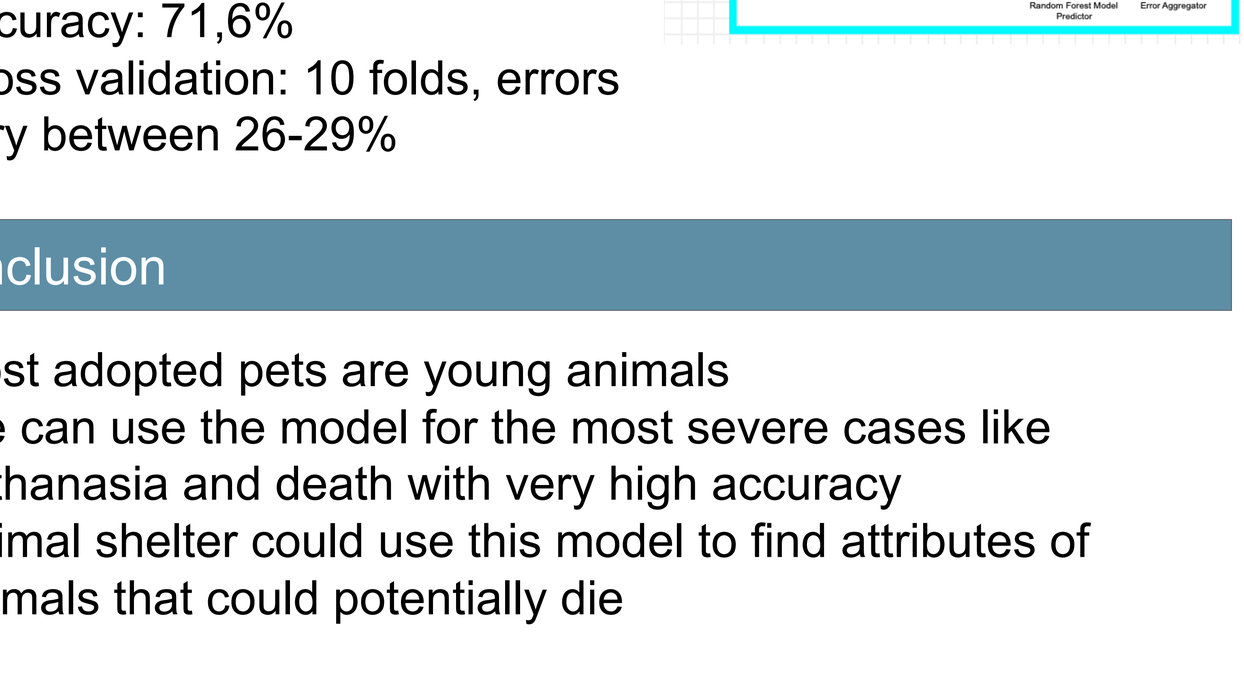
- Accuracy: 71,6%
- Cross validation: 10 folds, errors vary between 26-29%

Conclusion

- Most adopted pets are young animals
- We can use the model for the most severe cases like euthanasia and death with very high accuracy
- Animal shelter could use this model to find attributes of animals that could potentially die



Adoptions by age:
Adult, mature, very young, young, young adult



Adoptions by month