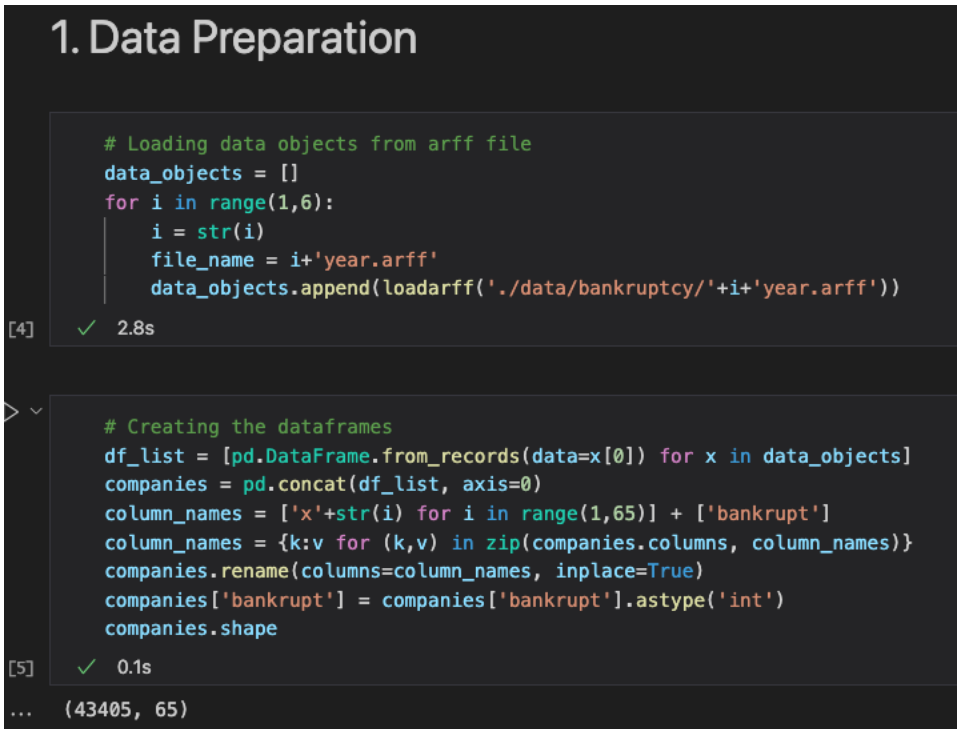
Föreläsning 8

Övning

- Predictiong Bankruptcy.
- Datan är på ett annorlunda format. Skriva av koden i bilden för att få det korrekt.
- Kom ihåh att hantera NaN värden. Rekommenderar att ta bara bort de kolumner som saknar mer än 2000 värden och använda fillna med median till övriga NaN punkter
- Är klassen är obalanserad? Ta eventuellt åtgärd med stratify
- Gör predikteringar med Logistic regression, Classification Tree och Naive Bayes
- Jämför prediktionerna av de olika modellerna med olika error metrics. Vad är viktigast att följa, precision eller recall?
- Få fram de tre viktigaste features från Classification tree genom att använda klassatributen feature_importances_



Frågor

1. Vad är Classification trees
    a. Målet
    b. Fördelar
    c. Nackdelar
2. Rita upp et möjligt flöde
3. Vad menas med Rule Inference
4. Förklara följande parametrar i scikit-learn (från DecisionTreeClassifier from the tree module):
    a. max_features
    b. max_dept
    c. min_sample_split
    d. min_sample_leaf

5. Vad intuitionen bakom Naive Bayes Models
6. Hur ser formeln ut
7. Vad är karaktäristiken för Naive Bayes Models
8. Vad är Gaussian ND Multinomial Naive Bayes
9. Vilken ”classifier” är lämplig till ”discrete features”
10. Vilken ”classifier” är lämplig till ”continuous features”

Länkar

- Decision Tree: build prune and visualize it using Python https://towardsdatascience.com/decision-tree-build-prune-and-visualize-it-using-python-12ceee9af752
- Post pruning Classification Trees https://scikit-learn.org/stable/auto_examples/tree/plot_cost_complexity_pruning.html
- Evaluation a Machine learning model https://www.jeremyjordan.me/evaluating-a-machine-learning-model/
- Oversampling and undersampling https://towardsdatascience.com/oversampling-and-undersampling-5e2bbaf56dcf
- Naive Bayes Classifier https://www.geeksforgeeks.org/naive-bayes-classifiers/

Data Camp

Supervised learning with scikit-learn https://app.datacamp.com/learn/courses/supervised-learning-with-scikit-learn