

Hausarbeit im Modul „Data Science und Machine Learning“ WS22/23 - Teil 1

Churn Prediction

You have been hired by a young FinTech startup that serves European customers with different financial products (e.g., banking, credit cards, ...). The startup is currently facing some headwinds because many of the customers are leaving the FinTech. As a data scientist, your job is to build a machine learning model that predicts which customers are about to leave. With this, the FinTech startup can actively engage with the identified customers and try to prevent them from leaving.

The data set contains the following attributes:

- CCreditScore: The customer's credit score
- CGeography: The customer's country of residence
- CGender: The customer's biological gender
- CAge: The customer's age
- CTenure: The number of years the customer has been a registered customer
- CBalance: The amount the customer has in the account
- CNumOfProducts: The products they own
- CHasCrCard: The status of whether the customer has a credit card
- CIsActiveMember: The status of whether the customer is an active member
- CEstimatedSalary: The estimated salary of the customer
- Churn: The status of whether the customer churned

1. Apply three different simple machine learning models, which we discussed in lecture 02. Describe your approach, interpret the performance, and justify (based on the results) which machine learning model you would choose.

2. Select the machine learning model that you have chosen in task 1. Now try optimizing your results by applying the advanced concepts (lecture 04). Use cross-validation (lecture 04) when evaluating your results. Describe your approach, interpret the performance, and justify which machine learning model you would choose.

3. Apply one ensemble learning technique (lecture 05) to see whether those techniques lead to a better performance than in task 1 and task 2. You should also apply the advanced concepts (lecture 04) to improve the model. Describe your approach, interpret the performance, and justify the final machine learning selection.