



ADACTA



# Client Behaviour Prediction: A Machine Learning Challenge

## THE DESCRIPTION OF THE TASK

---

Bank clients often behave differently than was agreed upon, among other things – they often withdraw their deposits before the maturity, or they pay off their loans earlier. Such changes, should they happen often, affect the everyday decision making in the bank. Subsequently, banks consider the actual client behaviour as they are developing their strategies and make business decision.

The bank has also noticed the dependency of clients' behaviour on different factors, internal and external – such as demography (client's age, number of children, etc.), macroeconomic data (economy strength, employment rate) and, naturally, interest rates in the market.

The bank is attempting to proactively govern the risks originating in such behaviour and predict possible scenarios through a certain time period. The goal is to adjust the offer according to newly predicted conditions and in that way, diminish the negative effects on their financial results.

## WHAT EXACTLY IS THE TASK?

It is necessary to try and predict the clients who will potentially change their contractual relationship with the bank be it by paying off their loan or extending the maturity of their deposit through historical and external data (key measures of economic performance). Optionally, it is possible to visualize the influence of the external factors on the client's behaviour. Through the extracted data, it is necessary to prepare personalized offers to clients.

## EDUCATIONAL COMPONENT

The competitors will have an opportunity to learn how the new methods of predictive analytics are used in the banking sector to enhance the service, but also decrease the variability of the business and decrease the overall risk.

The competitors will also have an opportunity to learn and gain experience in machine learning on real-world data.



**ADACTA**



**Raiffeisen  
BANK**

## POTENTIAL APPLICATIONS

Both examples described in the intro can have strong impact on the financial results of the bank and the business segment to which the defined clients belong.

## POTENTIAL PROBLEMS WHICH MIGHT OCCUR WHILE SOLVING THE TASK

Adacta shall keep a FAQ document updated with answer regarding any technical issues should they arise in the course of the competition. The given data is already used by RBA in simulations predicting the loan closure so learning from the data using machine learning is possible.

**Data Specificity:** The bank tracks every contract through the calendar year even if the contract was closed since the cumulative year results account for the closed contracts as well. As such, the same party might show up several times during a given year marked as closed early. In principle, it shows as such until the year ends.

**Possible issues in pre-processing:** The problems might occur in the pre-processing stage due to a large amount of information and missing values. It can be solved through data structure analysis and with different pre-processing methods.

## RECOMMENDED PROGRAMMING LANGUAGES AND/OR TOOLS TO SOLVE THE TASK

---

Qlik and Python are the recommended choices, but the final choice belongs to the competitors.

### USEFUL TOOLS:

- Jupyter Notebook (Python)
- Scikit-learn (Python)
- Bokeh (Visualization)
- Orange (Python)
- Weka (Java)

### THE EXAMPLE OF THE DATASET THE COMPETITORS WILL GET AND AMOUNTS PERTAINING TO IT

A list of transactional account of individuals or legal persons and/or the property loans of natural persons through several time series. The data shall be stored in csv files.

- The amount of data shall not exceed 2GB.
- There are 17 columns (16 parameters and 1 class variable)



ADACTA



## THE DESCRIPTION OF THE INPUT VARIABLES GIVEN BY COMPETITORS

The data about the parties is structures as follows:

Columns name	Data Type	Description
DATUM_IZVJESTAVANJA	datetime	Date of the report
KLIJENT_ID	numeric	ID
OZNAKA_PARTIJE	numeric	The loan
DATUM_OTVARANJA	datetime	Account opening date
PLANIRANI_DATUM_ZATVARANJA	datetime	Planned closure date
DATUM_ZATVARANJA	datetime	Real closure date
UGOVORENI_IZNOS	numeric	Original amount
STANJE_NA_KRAJU_PRETH_KVARTALA	numeric	Remaining loan amount at the end of the previous quarter
STANJE_NA_KRAJU_KVARTALA	numeric	Remaining loan amount at the end of the quarter
VALUTA	numeric	Currency
VRSTA_KLIJENTA	numeric	Client segment
PROIZVOD	categorical	Product
VRSTA_PROIZVODA	categorical	The product type
VISINA_KAMATE	numeric	Interest rate (percentage)
TIP_KAMATE	categorical	Interest rate type
AGE	numeric	Client Age
PRIJEVREMENI_RASKID	categorical	Was the account closed early (Y/N)

### MACROECONOMIC MEASURES FOR LEGAL PERSONS:

GDP, average import rate, Foreign Direct Investment – Impact on the transactional account of the company



ADACTA



Raiffeisen  
BANK

## MACROECONOMIC MEASURES FOR NATURAL:

Unemployment rate, average salary, inflation – impact on the retain and property loans

## DOES THE CONTRACTOR HAVE EXPERIENCE SOLVING THE SAME OR SIMILAR TASK?

---

RBA continuously runs similar simulations and ADACTA has an extensive experience in data analysis and machine learning. They encounter similar issues which require client analysis and profiling, the analysis of client-company interactions, training ML models to predict future interactions or client responses. Such models are usually combined with integrated solutions in the form of recommendation systems or decision support systems.

## ADDITIONAL INFORMATION AND REFERENCES

---

- <https://www.hnb.hr/statistika>
- <https://www.dzs.hr/>
- <https://www.hnb.hr/statistika/sdds>
- <https://www.hnb.hr/statistika/statisticki-podaci>
- <https://www.hnb.hr/statistika/glavni-makroekonomski-indikatori>
- <https://www.hnb.hr/statistika/informacije-za-korisnike-statistickih-podataka>
- Elements of Statistical Learning - Jerome H. Friedman, Robert Tibshirani, and Trevor Hastie
- Python: <https://www.dataquest.io/blog/sci-kit-learn-tutorial/>
- Java: <https://www.cs.waikato.ac.nz/ml/weka/book.html>
- Coursera - Machine learning by Andrew Ng
- Machine Learning - Mitchell
- Pattern Recognition, 2nd ed - Duda, Hart & Stork
- Macroeconomics, 3rd ed - O. Blanchard
- Macroeconomics-Study Guide - D.W. Findlay
- Modern econometrics - R. L. Thomas
- Time Series Analysis - J. D. Hamilton
- Reducing Customer Attrition with Predictive Analytics for Financial Institutions, Nate Derby, Mark Keintz, MWSUG 2016, Paper BI04
- Maximizing Cross-Sell Opportunities with Predictive Analytics for Financial Institutions, Nate Derby
- Using SAS® GTL to Visualize Your Data When There is Too Much of It to Visualize, Perry Watts, Nate Derby



ADACTA



Raiffeisen  
BANK

- <http://stakana.com/>

## REQUIRED ITEMS AND SOLUTION FORMATS

---

The user and technical documentation (a prepared template), presentation, the report of achieved results daily (at least once), the report of achieved validation results (final accuracy mark), codes.

## THE SCORING CRITERIA

---

### PRESENTATION

The presentation should have the following structure, which will be discussed during the workshop

- Motivation
- Problem definition and dataset analysis
- Chosen method
- Results
- Conclusion

### DOCUMENTATION – PDF DOCUMENT

The documentation should follow the given template.

### SOFTWARE – SW

Software should be structured, clear and well documented. The documentation and the software should be understandable to technical audience.

### THE ACCURACY OF THE TRAINED MODELS (EVALUATION)

#### Daily Rankings:

The competitors will be able to evaluate their results through a web service where a ranking list of competitors will be updated daily.



**ADACTA**



**Raiffeisen  
BANK**

## THE ACCURACY OF THE TRAINED MODELS (VALIDATION)

On-the-spot Rankings:

In the end of the competition, the finalists will have to classify a completely new dataset on-the-spot. This will give the final accuracy of the solution.

## THE SOLUTION QUALITY

The quality and viability of the solution will be assessed.

## THE SOLUTION INNOVATION

In case interesting insights into data which can be used to enhance the solution are presented, it shall be considered as an innovation. The usage of innovative algorithms shall also be treated as an innovation.

Criteria	Points	Criteria Weight
<b>Presentation</b>	0-10	10%
<b>Documentation</b>	0-10	10%
<b>SW – solution quality</b>	0-35	35%
<b>Solution Innovation</b>	0-10	10%
<b>The accuracy of the trained models (evaluation)</b>	0-15	15%
<b>The accuracy of the trained models (validation)</b>	0-20	20%

## ADDITIONAL BENEFITS

---

Possibility of an internship in RBA and Adacta.