



Modelowanie statystyczne w zarządzaniu wierzytelnościami masowymi – wprowadzenie

Laboratorium 1 Temat: Kurs, dane, R

Sebastian Kuzara

KRUK S.A.

Statistical Methods Development Area

Wrocław, 2024



Kurs



1. Przedmiot: Modelowanie statystyczne w zarządzaniu wierzytelnościami masowymi
2. czwartki, godz. 10:15 - 11:45
3. Kurs organizowany we współpracy UWr i firmy Kruk S.A.

Oferty pracy: [oferty pracy](#)

Źródło danych: repozytorium GitHub: <https://github.com/KrukUwr/Lab2024>

Ankieta “zapoznawcza”:

<https://shorturl.at/fUNUR>

Wymagania techniczne



1. R- rekomendowana wersja 4.4.1

<https://cran.r-project.org/>

2. RStudio, również zalecana najnowsza wersja

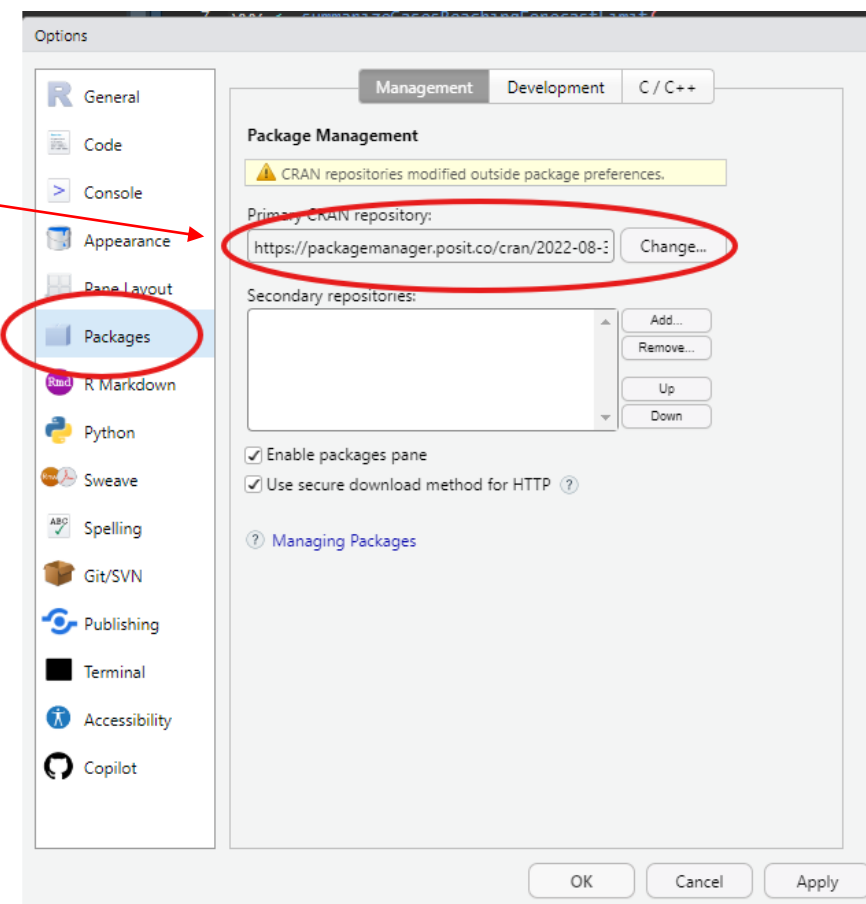
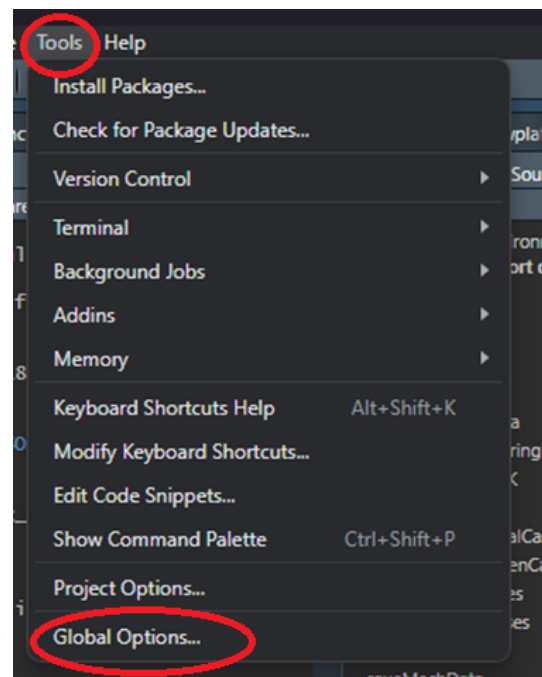
<https://posit.co/download/rstudio-desktop/>

R - pakiety



W celu zapewnienia spójności wersji rekomenduję, aby wskazać repozytorium CRAN-owe „zamrożone” na dzień 31.08.2024.

<https://packagemanager.posit.co/cran/2024-08-31>



Zasady zaliczenia laboratorium



1. Uzyskanie co najmniej 2 plusów na zajęciach
2. Prezentacja oraz „obrona” projektu zaliczeniowego
3. Dopuszczalne 2 nieobecności nieusprawiedliwione

Harmonogram



Harmonogram znajduje się w repozytorium na githubie:

<https://github.com/KrukUwr/Lab2024>

Tematy zajęć



1. Wprowadzenie do R i danych
2. EDA
3. Modele uczone bez nadzoru (*unsupervised learning*)
4. Idea podziału zbioru na uczący i testowy
5. Omówienie tematów projektów
6. Modele oparte na drzewach
7. Uzupełnianie braków danych
8. PCA
9. Regresja liniowa
10. Wybór zmiennych do modelu.
Wprowadzenie do ANN (h2o)
11. Regresja logistyczna
12. Regresja GAM
13. Metody porównywania klasyfikatorów
14. Konsultacje projektów
15. Prezentacje

Prowadzący: Sebastian, Piotr

Organizacja laboratorium

(dotyczy mojej części zajęć)



1. Zaprezentowanie wyników Waszej pracy z poprzednich zajęć i dokończonych w domu (dla chętnych) - możliwe zdobycie +
2. Krótka prezentacja wprowadzająca do tematu zajęć
3. Wasza praca nad wybranymi problemami biznesowymi. Możliwa praca indywidualnie lub grupowo
4. Prezentacja wyników Waszej pracy na zajęciach (dla chętnych) - możliwe zdobycie +
5. (opcjonalnie) quizy w aplikacji kahoot – do zdobycie tylko +, brak ocen negatywnych

Zbiory danych



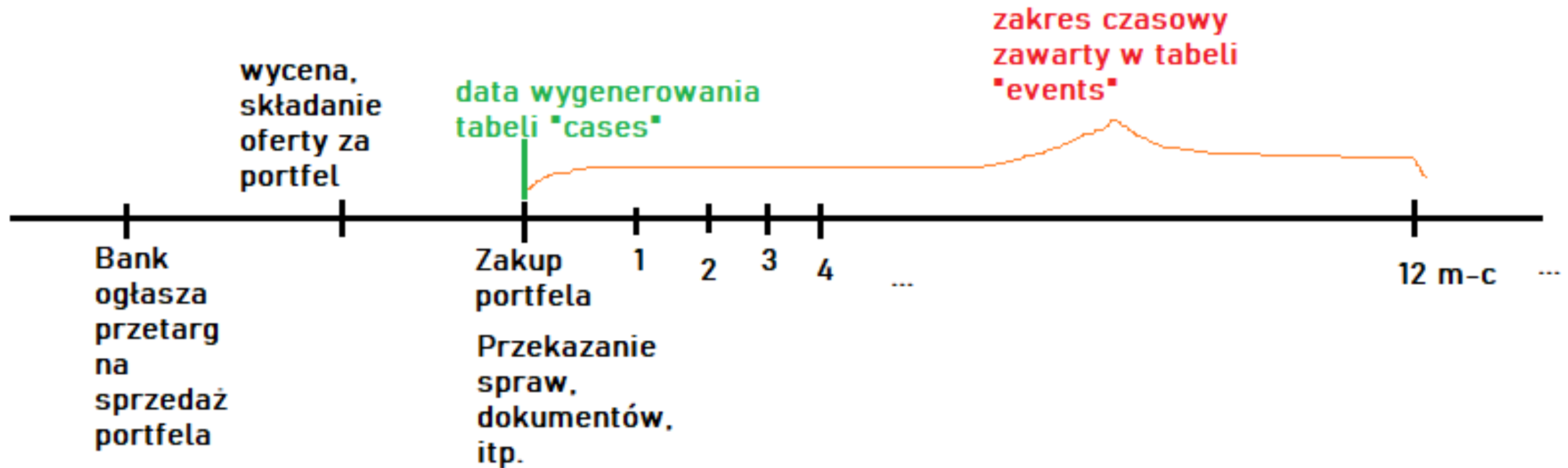
Plik: KrukUWr2024.RData - znajduje się w [repozytorium laboratorium na githubie](#)

Plik zawiera dwa obiekty data.table:

- **cases** - zbiór spraw wraz z ich atrybutami (cechy aplikacyjne) na moment zakupu*; identyfikator: *Caseld*
- **events** - zbiór zdarzeń po zakupie przypisanych do spraw; identyfikator: *Caseld* oraz *Month* (2 kolumny, „sprawomiesiąc”); NA oznacza brak zdarzenia; zawiera cechy behawioralne

* możecie się też spotkać z określeniami: data wygenerowania danych, data danych, data importu. W ramach naszych zajęć będziemy traktować je jak synonimy, choć jest to spore uproszczenie

Na osi czasu



Kolumny w tabeli *cases* (1)



- *Caseld* - identyfikator; numer sprawy
- *LoanAmount* - pierwotna kwota pożyczki (udzielonej przez bank lub inną instytucję udzielającą pożyczki). Kwota ta jest równoważna kapitałowi początkowemu. Kwota na moment udzielenia pożyczki przez bank.
- *TOA* - total outstanding amount - zadłużenie sprawy na moment jej zakupu (importu) przez Kruka
- *Principal* - kapitał pozostały do spłaty w momencie zakupu (importu) sprawy przez Kruka
- *Interest* - odsetki pozostałe do spłaty w momencie zakupu (importu) sprawy przez Kruka
- *Other* - inne składowe zadłużenia pozostałe do spłaty w momencie zakupu (importu) sprawy przez Kruka
- *Product* - nazwa produktu
- *D_ContractDateToImportDate* - liczba dni pomiędzy podpisaniem umowy pożyczki a zakupem (importem) sprawy przez Kruka
- *DPD* - days past due - dni od najstarszej zaległej płatności w momencie zakupu (importu) sprawy przez Kruka
- *ExternalAgency* - czy przed zakupem przez Kruka sprawa była obsługiwana przez agencję zewnętrzną, która specjalizuje się w odzyskiwaniu należności z wierzytelności.

Kolumny w tabeli *cases* (2)



- *Bailiff* - czy sprawa była obsługiwana przez komornika przed zakupem (importem) przez Kruka
- *ClosedExecution* - czy egzekucja komornicza została zamknięta przed zakupem (importem) przez Kruka
- *Land* - Id jednostki terytorialnej, w której mieszka klient (dłużnik)
- *PopulationInCity* - przeskalowana liczba mieszkańców w miejscowości zamieszkania klienta (dłużnika)
- *Age* - wiek klienta (dłużnika)
- *Gender* - płeć klienta (dłużnika)
- *LastPaymentAmount* - kwota ostatniej wpłaty na sprawie przed zakupem (importem) przez Kruka
- *M_LastPaymentToImportDate* - liczba miesięcy od ostatniej wpłaty na sprawie przed zakupem (importem) przez Kruka
- *GDPPerCapita* - PKB per capita w Landzie
- *MeanSalary* - średnie wynagrodzenie w Landzie

Kolumny w tabeli *events* (2)



Caseld - identyfikator; numer sprawy

Month - numer miesiąca obsługi sprawy po stronie Kruka (po zakupie/importcie)

NumberOfCalls - liczba wykonanych telefonów

NumberOfLettersSent - liczba wysłanych listów

NumberOfVisits - liczba wizyt bezpośrednich

NumberOfCallsWithClient - liczba rozmów telefonicznych z klientem

NumberOfVisitsWithClient - liczba wizyt bezpośrednich kiedy skontaktowano się z klientem

NumberOfLettersReceived - liczba otrzymanych listów

NumberOfAgreementConcluded - liczba zawartych ugód

NumberOfAgreementSigned - liczba podpisanych ugód

TransferToLegalProcess - czy sprawa została przekazana do sądu w danym miesiącu

NumberOfPayment - liczba wpłat w miesiącu

PaymentAmount - suma wpłat w miesiącu

LoanAmount

TOA

Principal*

Interest*

Other*

wybrane pojęcia (1)



1. Skuteczność(*SuccessRate*, *SR*) = wpłaty / zadłużenie_początkowe
2. Składowe zadłużenia - Kapitał, Odsetki, Opłaty, Prowizje itp.

LoanAmount

TOA

Principal*

Interest*

Other*

* Proporcje przykładowe, mogą się różnić między sprawami

Wybrane pojęcia (2)



Chronologia dat:

Os czasu	Nazwa zdarzenia	Uwagi
1	Data podpisania umowy kredytu (<i>contract date</i>)	a
2	Data ostatniej wpłaty	b, c, d
3	Data due (od tej daty liczone jest DPD)	a, c
4	Data przekazania sprawy do obsługi w agencji zewnętrznej (<i>external agency</i>)	b, c
5	Data przekazania sprawy do komornika (<i>bailiff</i>)	b, c
6	Data zamknięcia egzekucji, np. egzekucja nieskuteczna (<i>closed execution</i>)	b, c
7	Sprzedaż sprawy do Kruka	

a - brak daty dla kart kredytowych

b - zdarzenie może nie występować dla danej sprawy

c - zdarzenie może wystąpić w każdym momencie pomiędzy datą podpisania umowy a sprzedażą do Kruka

d - data ostatniej wpłaty może wystąpić po dacie due, jeśli wpłata ta nie pokryje całości rat

R



Instalacja i wczytanie pakietu na przykładzie *data.table*:

```
install.packages("data.table")  
library(data.table)
```

Wczytanie danych z pliku binarnego *.RData*:

```
load("KrukDane2023.RData")
```

Sprawdzenie klasy obiektu:

```
class(cases)
```

Sprawdzenie struktury obiektu:

```
str(cases)
```

Utworzenie wektora np.:

```
wektor1 <- c(1,5,6)  
wektor2 <- 1:10
```


R - przydatne linki



Operacje na obiektach klasy data.table:

https://s3.amazonaws.com/assets.datacamp.com/blog_assets/datatable_Cheat_Sheet_R.pdf

JOINowanie obiektów data.table:

<https://medium.com/analytics-vidhya/r-data-table-joins-48f00b46ce29>

Winietka-tutorialem z wprowadzeniem do pakietu data.table:

<https://cran.r-project.org/web/packages/data.table/vignettes/datatable-intro.html>