Paulo Cortez @ 2024/5

Project: Intelligent Decision Support System for *Favorita* Stores

OBJECTIVES:

- Apply Forecasting and Optimization methods in a complex real-world task, managing and documenting the project.
- Corresponds to 75% of the final Course Unit grade;

> Teams

- **Preferably teams / groups of 4 students**.
- **All team members** need to develop the project**.**
- Each team needs to elect a **leader** (responsible for managing the team and assuring a good development of the project through the semester).
- During classes there will be an **assessment** of the team behavior..

> Project Group Self-Assessment (A)

- At the end of the project execution, each team proposes a final project grade, from 0 (0%) to 20 “valores” (100%): **A**. The value of A should be justified. **Example** “The team believes that the project deserves 16 ”valores”, because we executed well these objectives..., but we were unable to achieve the objectives...”

- Each project is evaluated by the teacher, resulting in grade **P**, which can be equal or different than **A**.

Notes:

- Accomplishing all the objectives should result in a good evaluation, but not necessarily an excellent grade (e.g., 19 “valores”). The quality of the achieved results (programing code, analysis, report and presentation) is also relevant.
- It is not necessary to achieve all project objectives to be approved for the Curricular Unit.

> Project Individual Assessment

- Each group can propose an **individual self-differentiation**, clarifying the individual contributions for each member of the group.
- **Individual self-differentiation** follows these rules:
  - the individual sum of the grades cannot be greater than M x P, where M is the number of group members and P is the final grade of the project (assessed by the teacher).
  - a student cannot fail due to an increase of other grades (to do this, the group would have to "dismiss" the student in advance);
- Individual self-differentiation must be **justified in the report**, with a description of individual performance (see slide about report).
- The final decision made by the teacher (for example, via attendance and participation in practical classes, performance in presenting the project).

Example: for a P grade of 14 and 4 students *A, B, C, D*:
- *A* performed better than the average, {*B, C*} contributed with an average performance and D contributed less: A=15, B=14, C=14, D=13.

Ethical rules and concerns

Do **not falsify results**, **nor plagiarize projects from other groups** or **use Internet content without being properly identified and referenced** (who is the author, where it was published). The use of other sources cannot be exaggerated given the total work developed. Students are responsible for ensuring that all work submitted is original and references consulted are properly cited.

Artificial Intelligence (AI) tools, such as Grammarly, Chat-GPT, Gemini, and others, **can be used but with caution**. Any copied and pasted knowledge from AI tools should be **understandable** by the group members (e.g., to know what it does). Moreover, the **use of AI tools must be made explicit in the report and code** (e.g., which queries were used by which tools, which results were obtained). This information provides transparency and accountability for the use of AI in academic work and helps to avoid potential plagiarism issues.

> 

- ***Favorita* Stores** are popular and large Ecuadorian-based grocery retailer.
- ***Favorita*** created a database with daily sales from two stores (Id. **11** and **12**) of three important food family products: **d**airy, **e**ggs and **b**read/bakery, from January 2013 to August 2018 (total of 1684 days).
- For each day, the database includes the product **s**ales and the total amount of products (from the same family) that were being **p**romoted at the store.

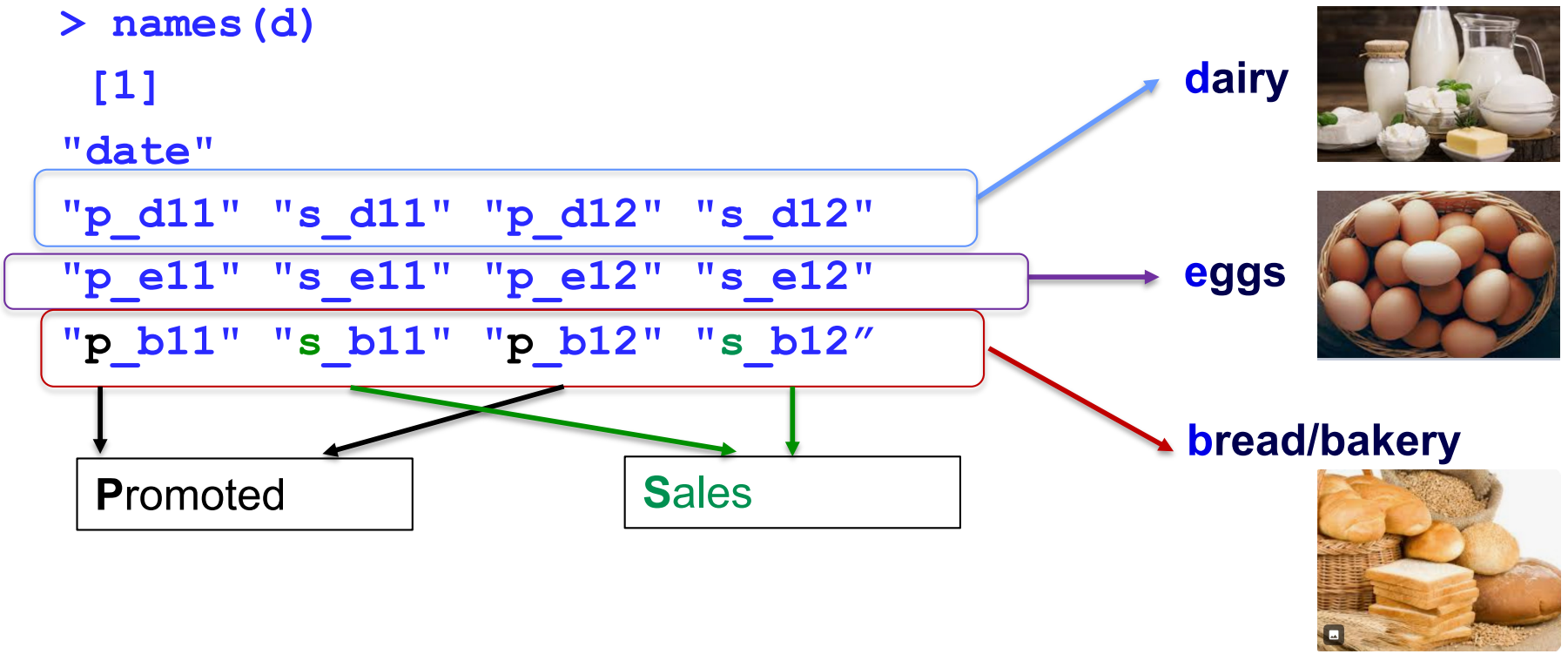
> Data

Available via a CSV file: **sales.csv**

In R, the data can be loaded using the command:

```
> d=read.table("sales.csv",header=TRUE,sep=";")

# Attributes (e.g., s_d11 are the dairy sales for store 11)
> names(d)
 [1]
"date"
"p_d11" "s_d11" "p_d12" "s_d12"
"p_e11" "s_e11" "p_e12" "s_e12"
"p_b11" "s_b11" "p_b12" "s_b12"
```



```
> summary(d)
     date               p_d11            s_d11            p_d12
 Length:1684        Min.   :  0.00   Min.   :   0.0   Min.   :  0.000
 Class :character   1st Qu.:  0.00   1st Qu.: 372.0   1st Qu.:  0.000
 Mode  :character   Median :  0.00   Median : 500.0   Median :  0.000
                    Mean   : 11.27   Mean   : 512.9   Mean   :  9.144
                    3rd Qu.:  8.00   3rd Qu.: 642.0   3rd Qu.:  5.000
                    Max.   :137.00   Max.   :1515.0   Max.   :109.000
     s_d12            p_e11            s_e11            p_e12
 Min.   :  0.0   Min.   : 0.000   Min.   :   0.0   Min.   : 0.00
 1st Qu.:251.0   1st Qu.: 0.000   1st Qu.:  80.0   1st Qu.: 0.00
 Median :348.0   Median : 0.000   Median : 130.0   Median : 0.00
 Mean   :335.4   Mean   : 1.992   Mean   : 157.5   Mean   : 1.26
 3rd Qu.:436.0   3rd Qu.: 1.000   3rd Qu.: 195.0   3rd Qu.: 0.00
 Max.   :879.0   Max.   :20.000   Max.   :1181.0   Max.   :15.00
     s_e12            p_b11            s_b11            p_b12
 Min.   :  0.00   Min.   : 0.000   Min.   :   0.0   Min.   : 0.000
 1st Qu.: 39.00   1st Qu.: 0.000   1st Qu.: 522.0   1st Qu.: 0.000
 Median : 60.00   Median : 0.000   Median : 585.0   Median : 0.000
 Mean   : 74.95   Mean   : 6.807   Mean   : 606.2   Mean   : 4.232
 3rd Qu.: 97.00   3rd Qu.: 5.000   3rd Qu.: 664.0   3rd Qu.: 3.000
 Max.   :370.00   Max.   :80.000   Max.   :2080.0   Max.   :50.000
     s_b12
 Min.   :   0.0
 1st Qu.: 148.8
 Median : 179.0
 Mean   : 186.5
 3rd Qu.: 217.0
 Max.   :1164.0
> d[1:3,]
        date p_d11 s_d11 p_d12 s_d12 p_e11 s_e11 p_e12 s_e12 p_b11 s_b11 p_b12 s_b12
1 2013-01-01     0     0     0     0     0     0     0     0     0     0     0     0
2 2013-01-02     0   296     0   151     0   140     0    75     0   763     0   337
3 2013-01-03     0   173     0   114     0    74     0    41     0   658     0   206
```

> Forecasting Goal:

Forecast the daily sales for each store and family product, using a maximum horizon event of one week.

- Both univariate and multivariate time series models can be adopted, as well as mathematical and machine learning algorithms.
- Models should be trained on older data and tested on newer data.
- The time horizon must involve a prediction from 1 to 7 days ahead (one week, up to H=7 multi-step ahead predictions).

Forecasting requirements:

- At least, each group member should model (code, experiment, test) one different forecasting method.
- The quality performance measures and evaluation procedures should be defined by the group.

Better assessment if different and more interesting forecasting methods are explored.

> Retail distribution

Retail distribution is the process of getting products to consumers.

It covers:

- Transportation
- Warehousing
- Inventory management
- The actual sale

> Retail distribution

Favorita uses 2 main types of distribution resources:

a) **Warehousing**:

Includes employees, distribution boxes and other materials required for transportation.

b) **Transportation**:

Three types of distribution vehicles.

> Warehousing Resources

- The cost of 1 resource warehouse unit (includes warehouse workers and packing materials) for a particular day **d** ($\mathbf{W_d}$=1) is:
  - 10 EUR for a **d** working day.
  - 15 EUR for a **d** weekend day (Saturday or Sunday).
- For a particular **d** day, $\mathbf{W_d}$ resources (integer value) can be allocated (e.g., by using temporary workers).
- One resource unit allows to pack and upload up to 220 products per day (of any family type).
- Only packed and vehicle uploaded products can be sent to stores and then sold.

> Distribution vehicles

- Rented on a daily basis, $\mathbf{V_d}$ vehicles for **d** day.
- There are three main types of vehicles (cost includes the renting and driver):
  - **V1**: fixed cost of 40 EUR, can distribute up to 200 products of any family (e.g., dairy) per day.
  - **V2**: fixed cost of 50 EUR, can distribute up to 300 products of any family (e.g., dairy) per day.
  - **V3**: fixed cost of 56 EUR, can distribute up to 390 products of any family (e.g., dairy) per day.
  - If **d** is a weekend (Saturday or Sunday), the cost is increased by 10 EUR per day for each vehicle.

> Store details

Each product sold on the same **d** day (one unit) has this revenue for a store:

- 1 dairy product: 1.0 EUR.
- 1 eggs product: 0.5 EUR
- 1 bread/bakery product: 0.20 EUR

Any product not sold on the same day has depreciation and storage costs:

- 1 dairy product: 0.10 EUR per day.
- 1 egg and bread/bakery products: 0.05 EUR per day

At the end of a week, any remainder unit product produces an average retail cost of 0.10 EUR (dairy) or 0.05 EUR (other products) per unit.

> Store details

At the beginning of a distribution week planning, the stores are empty (thus the stock $\mathbf{S_d=0}$ for all family products).

For a particular **d** day, if the total expected (forecasted) sales is X and what is available in a store is Y, then:

if X>Y, then only **Y** is sold that day. The stock does not increase $\mathbf{S_{d+1}=S_d}$

Else then **X** is sold and the the stock of that product increases $\mathbf{S_{d+1}=S_d+(Y-X)}$

> Optimization Goal:

At the end of day d, the aim is to define a week resource use plan (storage and distribution) for the next 7 days: between days d+1 and d+7. The resource plan needs to include the following daily integer numbers: distributed products (per family type) for each store (11 or 12), warehousing resources and types of vehicles used.

There are 2 objectives:

O1 - Maximize profit = sales revenue – costs for the planned week.

O2 – Maximize O1 and Minimize the total number of vehicles and warehousing resources of the weekly distribution plan.

distribution plan example

```
[1] "> Forecasted sales:"
     sd1 sd2 se1 se2  sb1 sb2
[1,] 626 366 177  38  472 181
[2,] 575 341 340  92  868 175
[3,] 608 387 163  65  634 145
[4,] 527 294 170  43  582 139
[5,] 569 381 338  45  582 184
[6,] 603 411 275 117  583 275
[7,] 611 373 438  15 2080 176
> Distributed products per day:
     wd1 wd2 we1 we2 wb1 wb2
[1,] 200 150 175  50 990 500
[2,] 100  64 300  20 155 101
[3,]   0   0   0   0   0   0
[4,] 700 300 100  35 200  23
[5,] 200 234 140  78 100  75
[6,] 100  50   8 230 300  23
[7,]   0   0   0   0   0   0
> Warehousing resources: 10 4 0 7 4 4 0
> total Warehousing resources cost= 345
check if Warehousing resources can distribute all products:
[1] TRUE TRUE TRUE TRUE TRUE TRUE TRUE
>Vehicle resources per day:
     v1 v2 v3
[1,] 11  0  0
[2,]  0  1  2
[3,]  0  0  0
[4,]  4  2  0
[5,]  0  3  0
[6,]  0  2  1
[7,]  0  0  0
> vehicle total cost= 1258
check if vehicles can distribute all products:
[1] TRUE TRUE TRUE TRUE TRUE TRUE TRUE
```

Weekend: "2017-08-12" "2017-08-13"

Elements of the plan

distribution plan example

```
[1] "True sales:"
     [,1] [,2] [,3] [,4] [,5] [,6]
[1,]  200  150  175   38  472  181
[2,]  100   64  300   32  673  175
[3,]    0    0    0    0    0  145
[4,]  527  294  100   35  200  123
[5,]  373  240  140   45  100   75
[6,]  100   50    8  117  300   23
[7,]    0    0    0   15    0    0
> total sales revenue= 3093.9
[1] "Stock:"
     [,1] [,2] [,3] [,4] [,5] [,6]
[1,]    0    0    0   12  518  319
[2,]    0    0    0    0    0  245
[3,]    0    0    0    0    0  100
[4,]  173    6    0    0    0    0
[5,]    0    0    0   33    0    0
[6,]    0    0    0  146    0    0
[7,]    0    0    0  131    0    0
> total stock cost= 93.1
[1] "Total costs:"
> total costs= 345 + 1258 + 93.1 = 1696.1
>> Week profit= 3093.9 - 1696.1 = 1397.8 EUR
>> Week resources= 29 + 26 = 55
```

Optimization requirements:

- For the optimization, assume:
  - forecasted values for the next 7 days (best) OR
  - use actual values;
- Assume 2 optimization possibilities:
  - Maximize the store's profit for a given week (O1).
  - Multi-objective: maximize the store's profit (O1) and minimize the total number of warehouse and distribution resources (O2).
- At least one optimization model must be used by each group member. Each member must be responsible for implementing (code), configuring (setting parameters) and experimenting with this optimization model.

What is evaluated?

- Quality of the forecasting and optimization study (**report**).
- System developed, including a console or graphical (better) mode interface (e.g., via shiny.rstudio.com) to demonstrate the complete Intelligent Data Analysis system (prediction and optimization) at work.
- Better if:
  - several forecasting / optimization methods are compared;
  - several R (or Python or …) packages are explored;
  - forecasts are used by the optimization methods;
  - a more robust evaluation is adopted;
  - more realism is introduced (e.g. other optimization scenarios);
  - study the effect of changing some of the forecasting / optimization parameters.

> Project Execution

- From today until the date of the project presentation.
- Teacher Support is performed during AID classes.
- During the classes, a preliminary evaluation (non-formal, with a reduced weight) will be carried out for each group and project (all members of the group should be present): this contributes to the **PA*** project assessment component.

* PA - Active Participation assessment.

> Week Summary:

From 27/03/2025, all groups must show the teacher a “weekly project work summary report” (preferably via Googledocs, sharing a link with the option “anyone with the link can view”).

The summary should include 2 parts:

- Work team summary: what was accomplished, how many group meetings were held and their duration (in hours), examples of outputs (e.g., code, graphs, tables) and list of doubts.
- For each element of the project: what you did individually and how many hours you dedicated to the project during the week.

> Week Summary example:

Semana 21/03 a 28/03

Trabalho de Grupo	
Resumo do trabalho realizado: Criação de dois ficheiros diferentes. Um growing window para o forecast e outro growing window para o rminer. Distribuição dos modelos pelos elementos de grupo. Cada elemento ficou responsável por elaborar/adaptar o código do Growing Window para 2 modelos distintos.	
Número de Reuniões Realizadas: 3	
Reunião Nº1	**Dia: 25/03/2024**
	Duração: 2h30
	Objetivos: Estudo do growing window. Criação de growing window para os modelos de rminer e forecast. Compreensão e testes.
Reunião Nº2	**Dia: 26/03/2024**
	Duração: 2h
	Objetivos: Continuação da reunião do dia anterior. Criação de dois ficheiros diferentes. Um growing window para o forecast e outro growing window para o rminer para termos o código mais organizado.

> Week Summary example:

Dúvidas:

- ☑ ~~**Ao fazer o nrow no D, não tamos a perder dados? no CaseSeries**~~

Próximos passos temos:

→ Usar métodos de divisão entre dados teste e dados de treino, sem ser à "mão" - holdout
→ Testar com multivariados
→ Usar algoritmos de rolling e growing window

Trabalho Individual		
Elemento	**Tarefas Realizadas**	**Horas dedicadas ao projeto**
John Doe PG11111	Elaboração do ficheiro "Predict_ ' (xgboost + lm) Registo das métricas do modelo no excel, para os vários cenários.	**5h**
Jane Doe PG2222	Elaboração do ficheiro "Predict_ (random forest + mlpe) Registo das métricas do modelo no excel, para os vários cenários.	**5h**

> Project upload and presentation

- **Dates:** 2025-05-29
- **Group presentation:** in-class with all elements, includes project presentation and question answering. Only one group will be present at the classroom for each presentation. A group schedule will be provided at the elearning system.
- **Project upload:** Report (in pdf) – mandatory; presentation slides (in pdf format) – optional; R code attached to the zip file - mandatory => single .zip file to be uploaded in the elearning system.

Demo Video

- A computer video must be recorded (maximum of 5 minutes, ideally with voice narration) where the developed system is demonstrated (in R and/or other tools) and uploaded to **youtube**.
- **Include in the report** the link of the youtube video.

To search for free screen video capturing software do a google search:

screen capture video

> Report

- **Cover**

1. **Introduction:** brief description of the group, project goals and how the report is structured.
2. **Project execution**: project planning, how the group worked, self-assessment (**A** grade proposed for the project), proposal of individual self-assessment, etc. Use half a page per group element, describing what she/he did in terms of the project execution (e.g., effort, tasks, code, tests).
3. **Prediction task**: what was done and obtained results.
4. **Optimization task**: what was done and obtained results.
5. **System demonstration**: of the developed intelligent data analysis system, use 1 sentence with the youtube link.
6. **Conclusions**: brief appreciation and discussion of what has been accomplished

- **Bibliography**: optional
- **Attachments**: optional

Try to have a maximum of around 20 to 40 pages of the body document (without cover, indexes, bibliography and attachments). Use a direct and concise writing.