UNIVERSITÉ Grenoble Alpes

App Shiny machine learning

Komi – Azat – Lucas – Mamadou - Soheil

Didier Morel

BD

UNIVERSITÉ Grenoble Alpes

Table of contents

- Context presentation
- Expected rendering
- Organization
- First step

Komi Agblodoe - Azat Aleksanyan - Lucas Chabeau – Mamadou Diamban – Soheil Salmani
Didier Morel

BD

Project proposed by

Didier MOREL, Data Scientist @ BD

- Becton, Dickinson and Company is an American medical technology company
- Offers innovative solutions

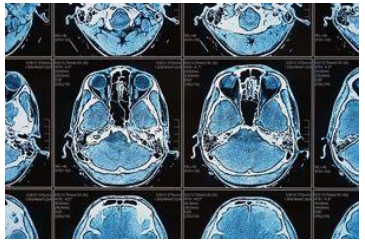
BD's workflow and our integration

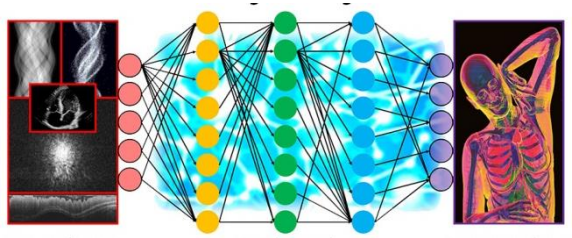
Komi Agblodoe - Azat Aleksanyan - Lucas Chabeau – Mamadou Diamban – Soheil Salmani
Didier Morel

BD

ML and healthcare industry

" Machine learning is swiftly infiltrating many areas within the healthcare industry, from diagnosis and prognosis to drug development and epidemiology, with significant potential to transform the medical landscape." © nature.com





Presentation of the application

Develop a Shiny application to test and compare ML algorithms over any kind of dataset

Issues

- Extract the data from its source (.txt, .csv, Excel, etc.)
  - Different code for each format
- Identify the type of each variable
  - Can be tedious to do that manually for a large dataset with lots of variables
- Manage missing values
  - Several methods exist to deal with missing values

Issues

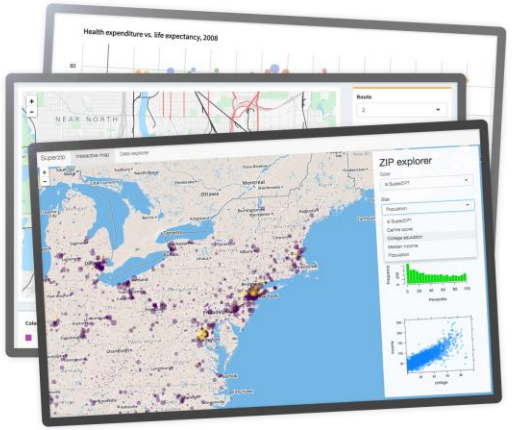
- Select ML algorithm and his settings (called *hyperparameters*)
  - Many differents types of ML methods exist
  - Often, we don't know which algorithm to choose until we test some of them
- Launch the training phase
  - Can be long and it takes time if we test different ML methods with different hyperparameters
- Analyze the quality of the model

The need

- A graphical interface : the user doesn't need to code
- Flexible way to :
  - Import data and to do the data preprocessing
  - Select the method to deal with missing values
  - Identify each type of variable automatically
  - Choose ML algorithms and hyperparameters we want to test and compare
- Generate a quick analytical report of the results (with tables and graphics)

Graphical interface with Shiny

Shiny is an R package that makes it easy to build interactive web apps straight from R.



Features

- The application must be compatible with any type of dataset (CSV, Excel, etc.)
- The application can detect automatically each type of variable for a dataset
- The user can select easily the set of variables to consider
- The user can choose easily a list of ML algorithms and hyperparameters to use for a dataset
- After the process, the application can generate a quick analytical report of the results and export it

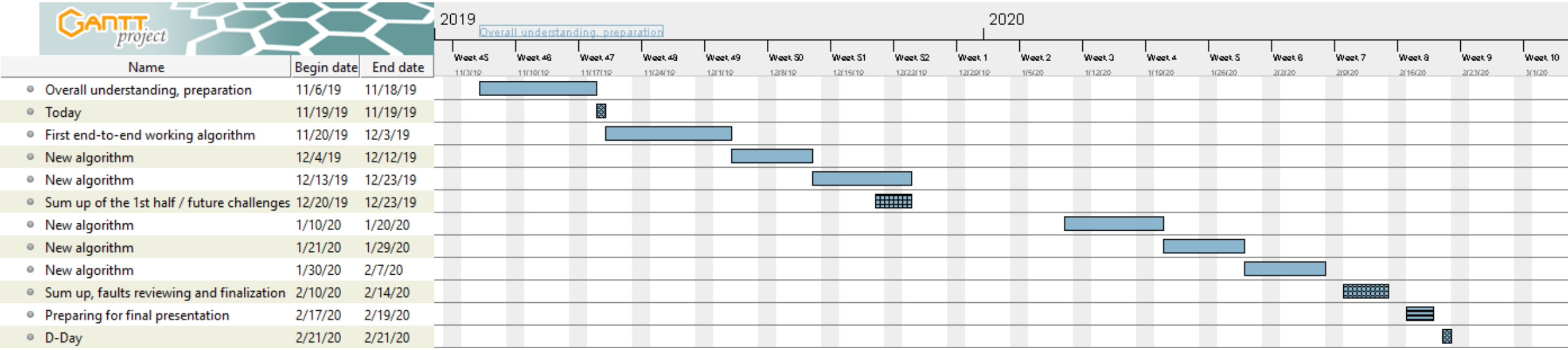
Algorithms used

- Machine Learning
- Selection of important variables
- Different algorithms
- Library caret

Planning

- Analysis phase (functional) or design
- Implementation or programing phase
- Delivery phase

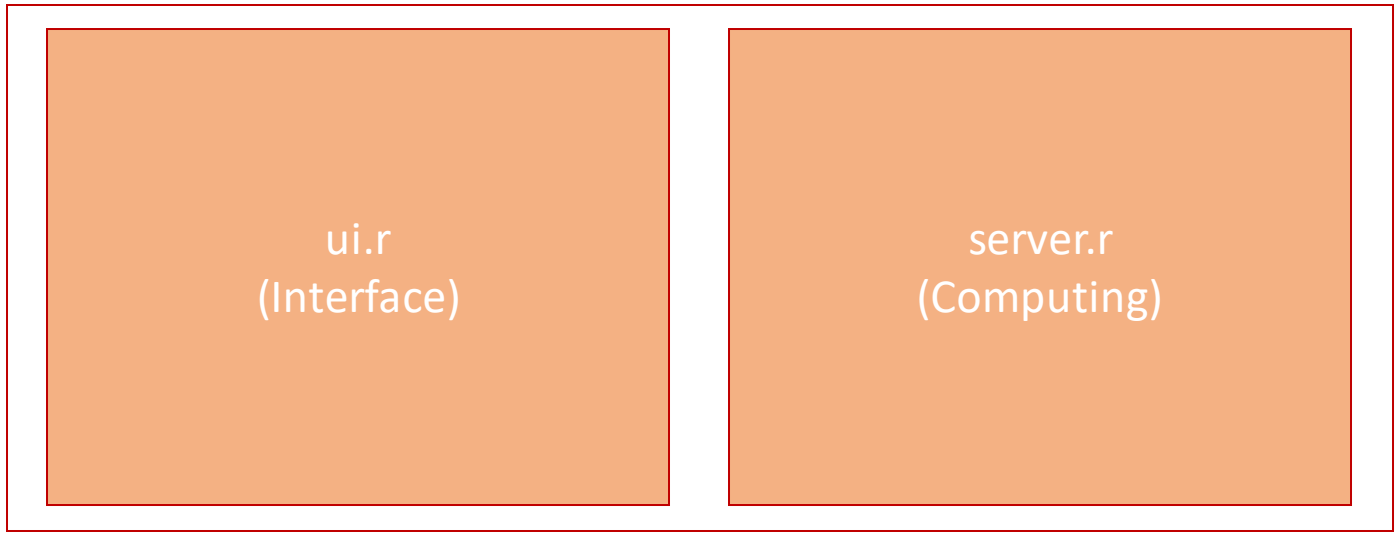
Gantt



Name	Begin date	End date
Overall understanding, preparation	11/6/19	11/18/19
Today	11/19/19	11/19/19
First end-to-end working algorithm	11/20/19	12/3/19
New algorithm	12/4/19	12/12/19
New algorithm	12/13/19	12/23/19
Sum up of the 1st half / future challenges	12/20/19	12/23/19
New algorithm	1/10/20	1/20/20
New algorithm	1/21/20	1/29/20
New algorithm	1/30/20	2/7/20
Sum up, faults reviewing and finalization	2/10/20	2/14/20
Preparing for final presentation	2/17/20	2/19/20
D-Day	2/21/20	2/21/20

UNIVERSITÉ Grenoble Alpes

Structure of app



UNIVERSITÉ Grenoble Alpes

Data uploading

- csv / tsv
- flexible
- First limit with big datasets

SampleID	sex	age	ag
1	NA	NA	
2	NA	NA	
3	NA	NA	
4	NA	NA	
5	NA	NA	
6	NA	NA	
7	NA	NA	
8	NA	NA	
9	NA	NA	
10	NA	NA	
11	NA	NA	
12	NA	NA	
13	NA	NA	
14	NA	NA	
15	NA	NA	
16	NA	NA	
17	NA	NA	
18	NA	NA	

Komi Agblodoe - Azat Aleksanyan - Lucas Chabeau – Mamadou Diamban – Soheil Salmani
Didier Morel

BD

UNIVERSITÉ Grenoble Alpes

Thanks for your attention.

Have you got any questions?

Komi Agblodoe - Azat Aleksanyan - Lucas Chabeau – Mamadou Diamban – Soheil Salmani
Didier Morel

BD