16/01/2020

Project Report

Choosing the right tool to do upper airway surgery

Emma ARRIAU
Pierre LUSCAN

MALIS PROJECT

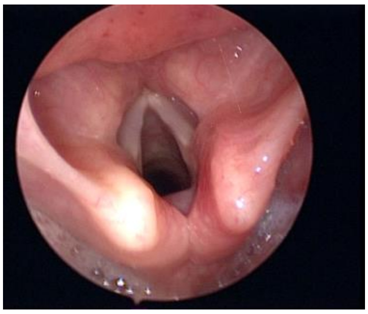
Context

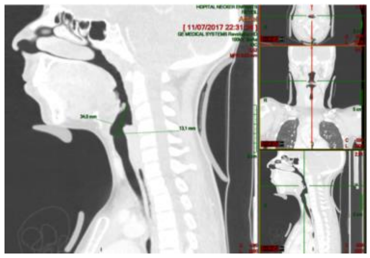
Romain LUSCAN and Briac THIERRY are pediatric ENT (otorhinolaryngologist) surgeons in “Necker enfants malades” Hospital (Paris). They are working on a brand-new project, which objective is to help surgeons and anesthetists involved in pediatric upper airway, by determining the adapted tool or probes for each child.

Decision-making in the upper airway area can have irreversible consequences for the child's life. Therefore, it is essential for any specialist involved in the child's airways to be able to estimate its dimensions as accurately as possible and to know its growth phases. The control of these data has many consequences for clinical practice, from the choice of the intubation tube by the anesthetist (1,000,000 pediatric anesthesia each year in France) to the decision making for the diagnosis and treatment of upper airway malformation (like laryngeal stenosis) by the ENT.

Today, to make decisions in the field of children's airways, surgeons have different bases, not always very scientific !

1. The clinician's experiences
2. Abacus inherited from various studies from the 1980s.
3. A simple method which compare the size of the little finger to define the size of one of the cartilages of the child's larynx (but not really precise!)





Problem definition

Our goal is to help surgeons choosing the right tool for each child. To do so, we want to analyse and predict some measures giving valuable knowledge on the kid’s upper airway (outputs), knowing the age, the sex and the weight of the patient (inputs). With these measures, the final aim is to predict the right tube to use among different sizes of tubes and constraints the tube must respect to fit the upper airway.

Data

The data at our disposal comes from morphometric studies on 192 children between 1 day to 14 years.
In addition to the age, the weight and the sex, 23 measures giving information regarding the upper airway (trachea, larynx...) are given. These measures were made from scanners. There are 6 measures of distance, 13 of diameters and 5 measures of surface (see annex next page for more details).

Methodology

1. Understand and analyze the problem: the objectives, the difficulties and the dataset
2. Detect dependency
   - Input/Output
   - Output/Output
3. Test different models of prediction : probably first the linear one and then trying others (depending on the results of the step two)

What about the size of our data set?

The (small) size of our dataset forces us to analyze the dependance between each input (age, sex and weight) and each output (each of the 23 morphologic data) but also between the outputs (finding a dependency between the outputs could allow us to have less than 23 measures to predict).

Final question

With constraints given by the doctors, how could we be able to predict the right tube without knowing the measures?

I. Prediction of the 23 measures regardless of the tube to use

- We tried to answer the first question the doctors asked us: is it possible to directly predict the 23 measures with only age, sex and weight in inputs?

Step 1 : Preparation of our Data Base

We asked R.L for the data. We had to turn it into an Excel file in order for us to be able to easily access it with our Jupyter Notebook.

29 columns/attributes
- 3 inputs : Age (in months), gender (0=M/1=F) and Weight (kg)
- 23 outputs : among them, measures of lengths, diameters and areas.
- 3 (Day of birth, day of the CT and age at the moment of the CT) that we do not plan to use.

193 lines (which is not much)

Step 2 : First analysis of our data

We first took a quick look into our Data Base. We plotted schemas of the distribution of different attributes.

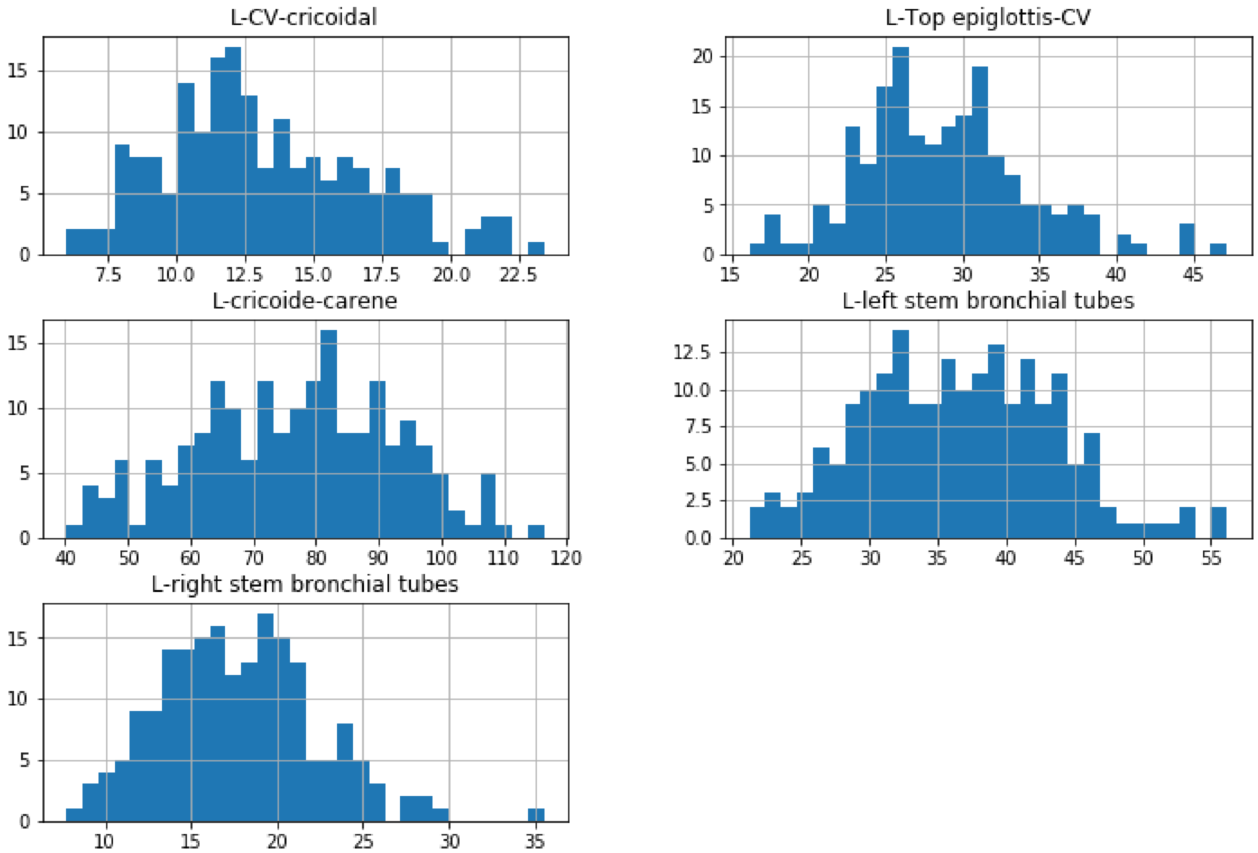


Fig 1 : Distribution of the 5 length attributes from our dataBase

We did the same for diameter and area attributes.

Conclusions: We can infer that all 23 outputs are more or less correlated. Indeed, it seems pretty logic that a child who has a long trachea has also a large trachea a big area of distal trachea for example.

Step 3 : Machine Learning

Idea: We selected one particular output ('L-CV-cricoidal') and tried to predict it.
For both methods, we first split the data into two different sets: **training set** and **testing set.**

1. Linear Prediction

Related files : Notebook 1 – part 1

Inputs : age (in months), sex and weight
Outputs : 'L-CV-cricoidal'

TRAINING :

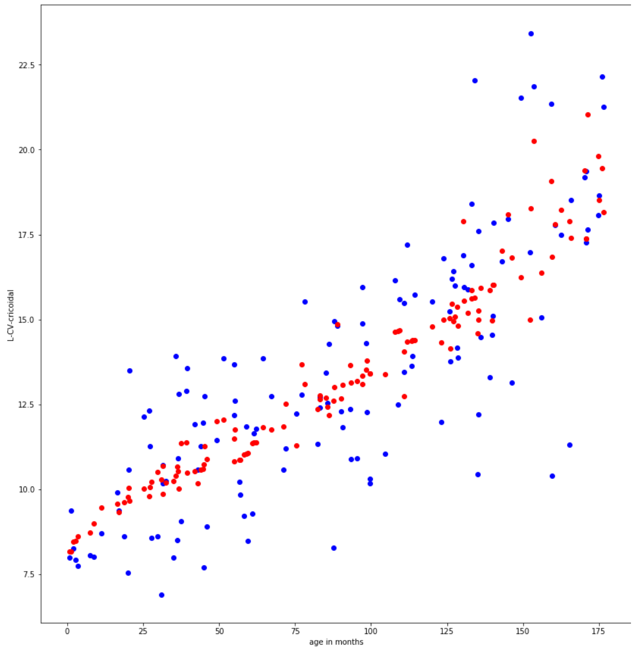


Fig 2 : Length training values according to one of the 2 inputs : here, according to the age (in months).
—> In blue : the data base values
—> In red : the prediction values
MSE = 4.38

TESTING :

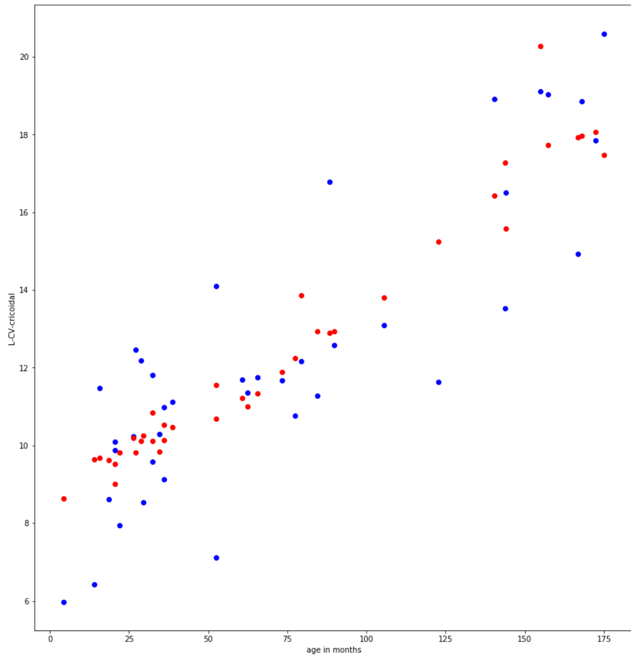


Fig 3 : Length testing values according to one of the 2 inputs : here, according to the age (in months).
—> In blue : the data base values
—> In red : the prediction values
MSE = 4.55

Conclusion : Using Linear Prediction seems, here, quite efficient as the MSE value is quite low.

We also wanted to check whether using linear dependency with regards to the age and weight was the best model. We therefore tried with different polynomial models to see which is the best. That's what we obtained (cf. following figure).

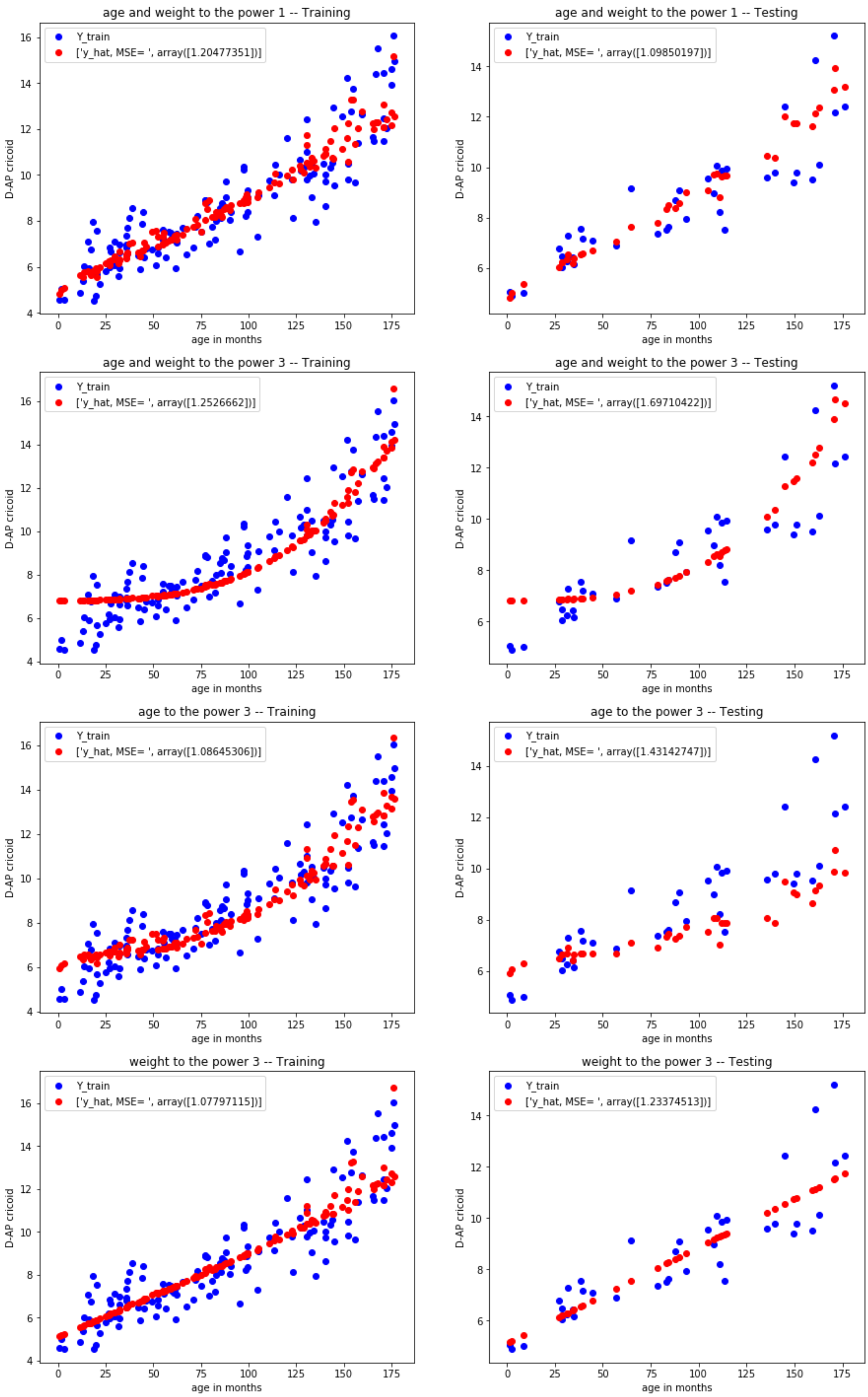


Fig 4: Results of the regression with different power for weight and/or age

Conclusion: We obtained the minimum MSE using linear dependency with respects to the weight and age. The influence of sex didn't change anything so we chose not to show it.

2. Linear Classification

Related files : Notebook 2

We also tried linear classification (as the final objective is to choose a tube among different tubes, it's interesting to study a classification directly one of the measures).
Given the distribution of the 'L-CV-cricoidal' outputs (cf fig1-first plot), we decided to distinguish 4 categories:

- cat1 : LTE in [0,25[
- cat2 : LTE in [25,30[
- cat3 : LTE in [30,35[
- cat4 : LTE in [35,…[

Inputs : age (in months), sex and weight
Output : Categorie (1,2,3 or 4)

The repartition of the training set data in each category can be seen in this graph :

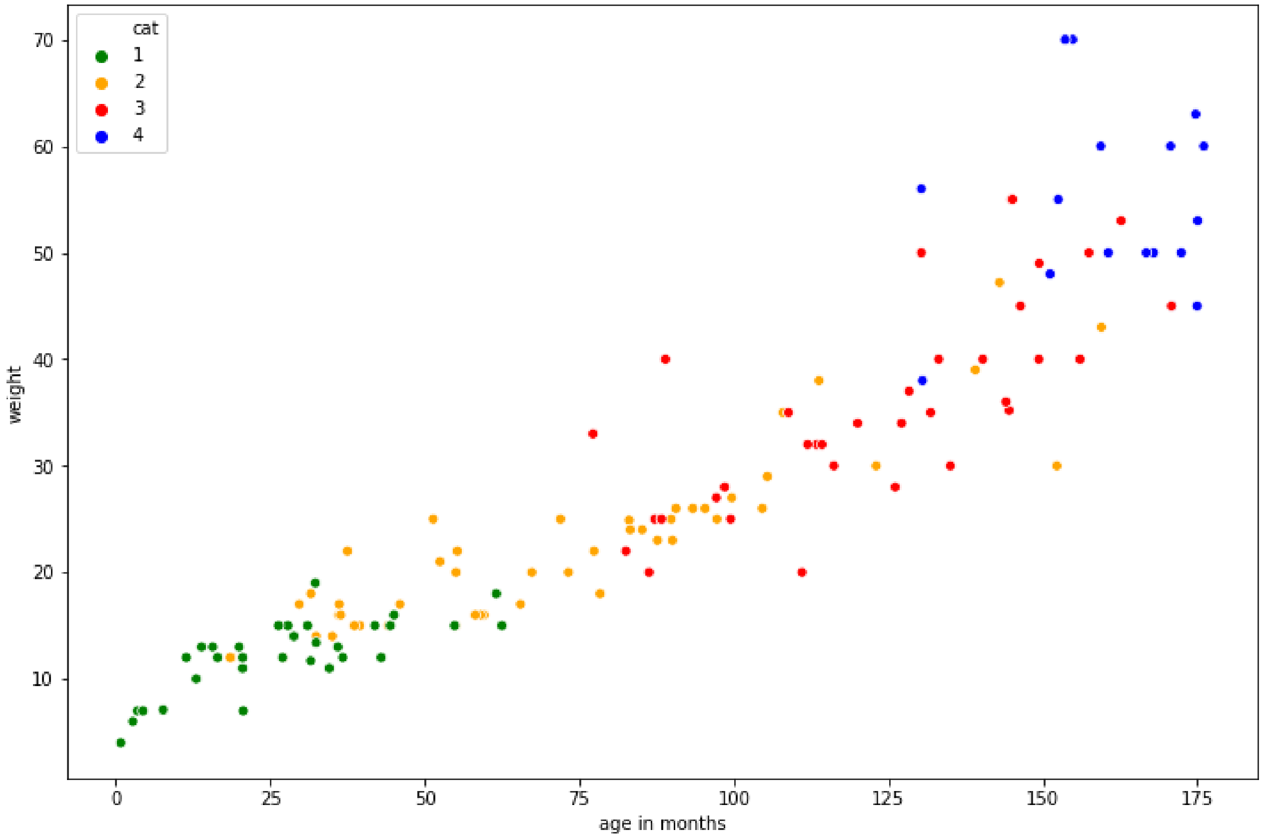


Fig 5: Repartition of the training data set in each category.
Abscisse : Age (in month)
Ordonnées : Weight

TRAINING:

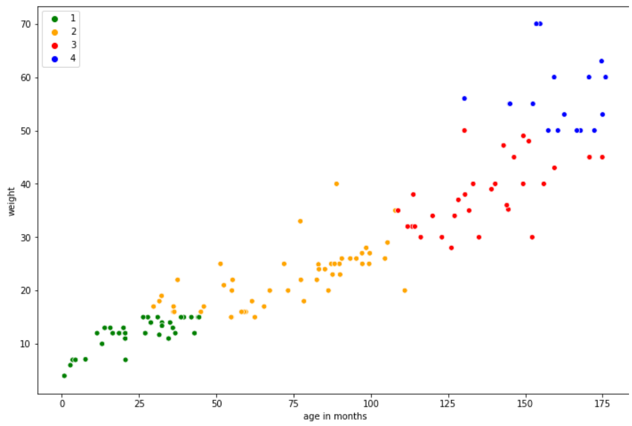


Fig 6: Repartition of the training results in each category.
Abscisse : Age (in month)
Ordonnées : Weight
Accuracy : 72,4%

TESTING:

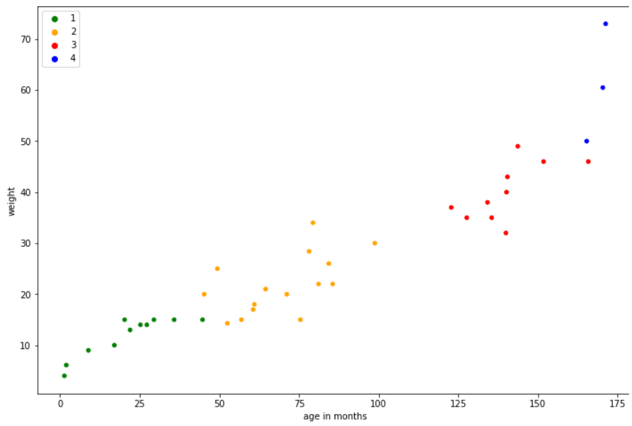
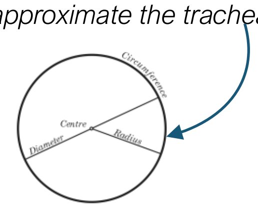


Fig 7: Repartition of the training results in each category.
Abscisse : Age (in month)
Ordonnées : Weight
Accuracy : 64,1%

Conclusion : Using Linear Classification does not seem, here, quite efficient. It is logic as we can see in fig5 that the data is far from being linearly separable.

- We realized that it couldn't be a solution to go over the same processes 23 times for each outputs. Moreover, it didn't make sense to us: we don't even know what these values refer to.
  - Idea 1 : As we discover from the distribution of each attribute, it exists a strong correlation between them. When looking at what these attributes refer to, maybe we can try to use our own models?
    Example : Let us approximate the trachea with a cylinder. Then, knowing its length L, we also know its area (πL).



  - Idea 2 : As there are **few tools** that a doctor can use to pursue surgery on a child, maybe we can try to predict the right tool according to the age, weight and sex of the child instead of wanting to know the exact 23 outputs.

We decided to go along with the second idea and contacted R.L to have a clear idea about what these tools were.

II. Prediction of the tube to use

- We saw that it was not relevant to predict all 23 measures directly from age, weight and sex. Indeed, having that few inputs was making the predication inaccurate. But by classifying, we could allow us to be less specific. So, we specified the objective: we asked for the **exact constraints** that allow the doctor to choose the adapted tool for a child !

- The tube must fit as much as possible the child's trachea in order for the surgery to be successful. The most important constraints are that the tube's outer diameter (OD) must be smaller than both trachea and cricoid diameters.
- In conclusion, knowing the Antero-posterior trachea diameter, the Antero-posterior cricoid diameter and the transversal trachea diameter of the child, the tube that best fits is to one who have the bigger diameter Di that verify the following constraints :

Di ≤ Antero-posterior trachea diameter

D i ≤ Antero-posterior cricoid diameter

D i ≤ transversal trachea diameter

Step 1: Preparation of our Data Base and analysis of the data

- The doctors gave us the references of 17 different tubes' sizes (cf. annex for more details on tubes). Thus, we added a column in our dataset and determined the matching tube for each input of our dataset (with respect to the above constraints). We plotted the results to see how the repartition seem to be like.

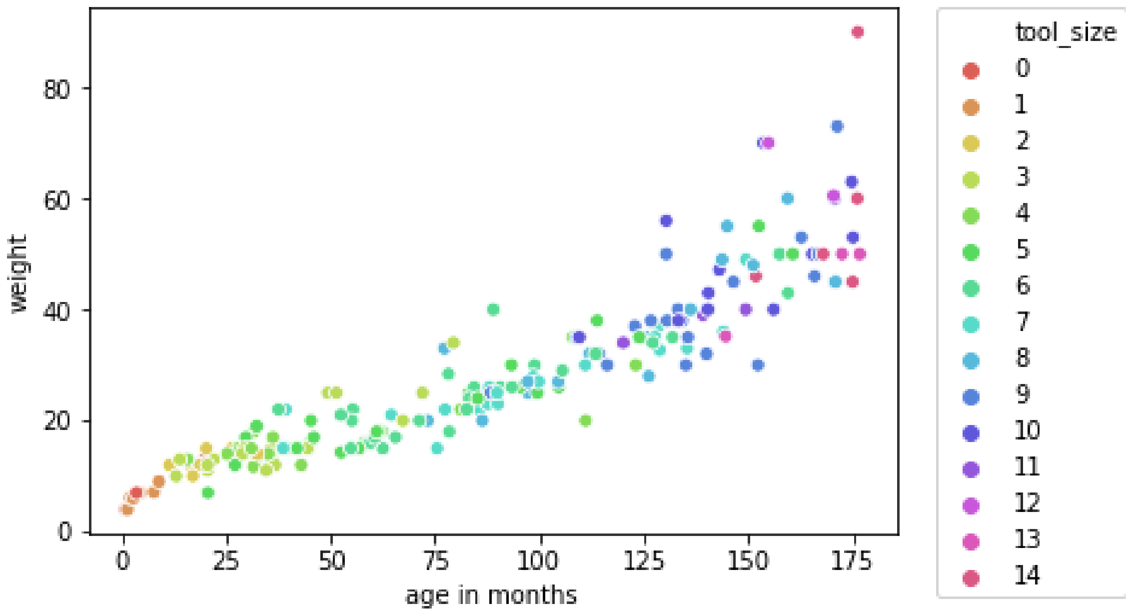


Fig 8 : Repartition of the children of data we have among the different tube sizes

- We immediately saw that it will be difficult to have a great accuracy seeing the amount of data we have and the number of classes but we needed to go deeper to see what is possible or not and what are the effects of the different constraints.

Step 2: Machine Learning

a. Classification using 3 linear independent models for the three measures

Related files : Notebook 1 – part 2

- First, we wanted to check that each measure influencing the choice of the tube (each of the three involved in the above constraints) had more or less a linear dependency with respects to the age, weight and sex.

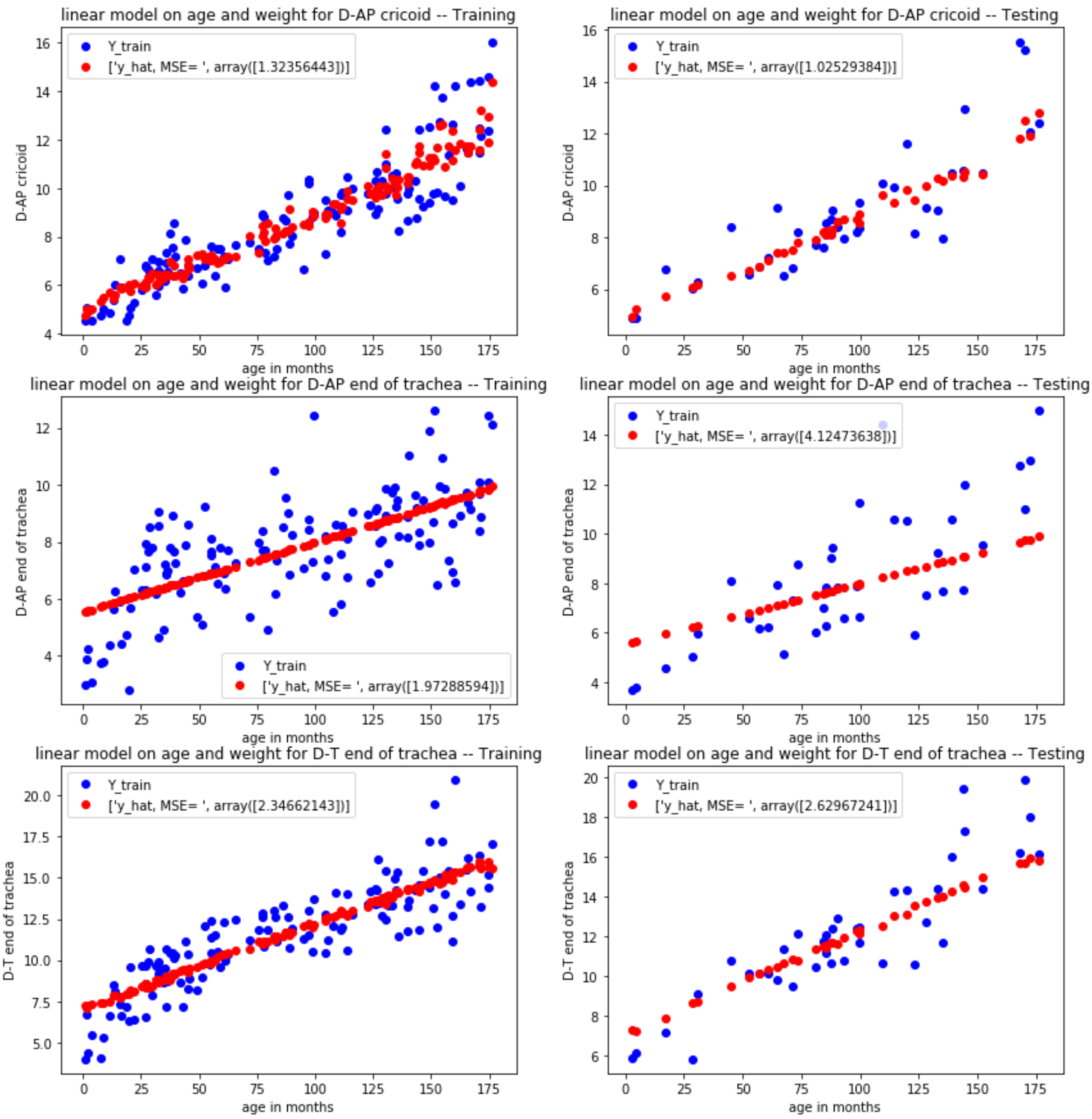
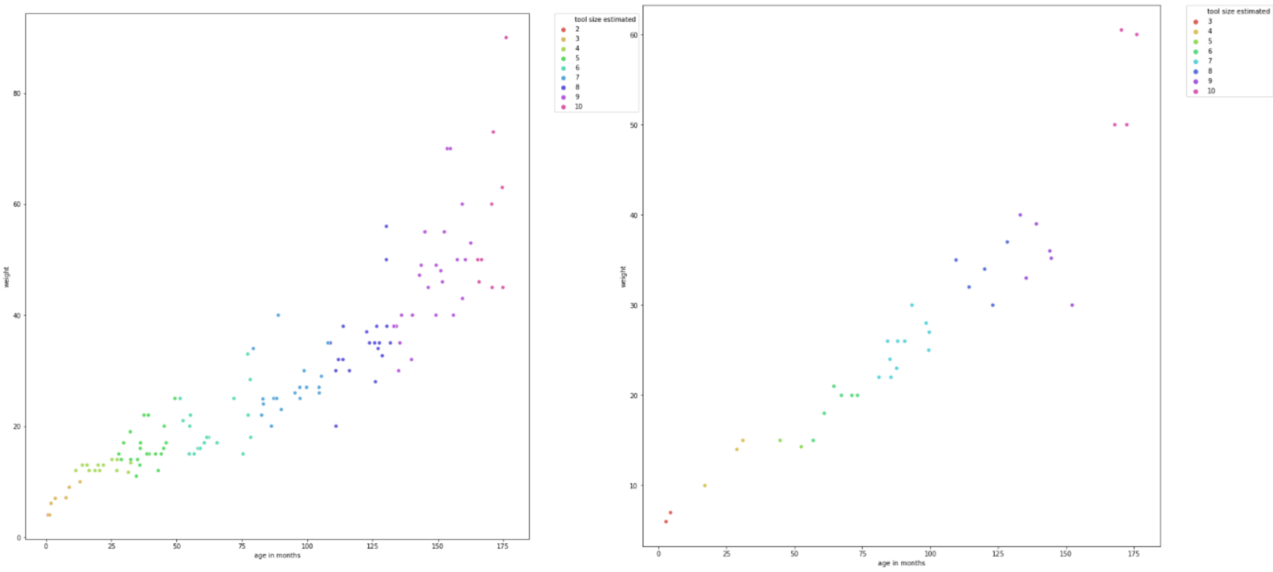


Fig 9 : Results of linear regression with MSE value for each of the 3 measures under constraint

- Remark: Here, we only plotted the data according to the age in month. We did the same according to the weight, and, it appeared that the tendencies of the plots were quite the same.
- After running **independent** linear machine learning algorithms to predict all of these 3 measures, we added these values in the data frame and applied the constraints to find the right tube for each child. We obtained these graphs:



TRAINING TESTING

(age and weight in axis)

Fig 10 : Results of the classification after 3 independent linear models on the 3 measures under constraint

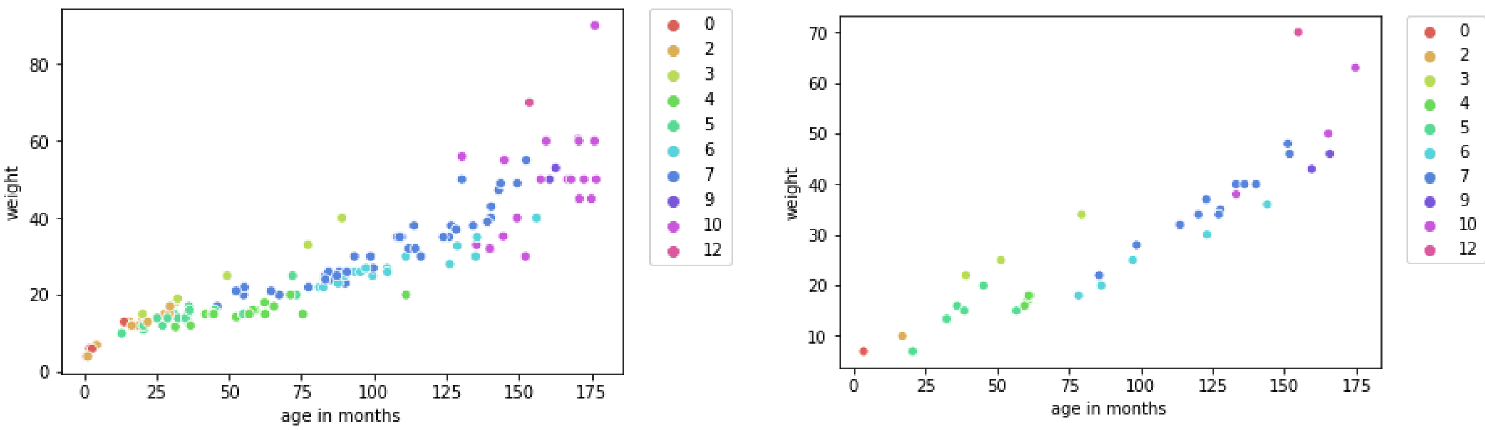
TRAINING ACCURACY = 29% TESTING ACCURACY = 23%

- These accuracies are the one of choosing the best adapted tool (which fit as much as possible to the trachea's dimensions).

b. <u>Classification using directly LDA on weight, age and sex</u>

Related files : Notebook 3 – part 1

- We then did the linear discriminant algorithm to predict directly the most adapted tool given the age, the weight and the sex of the childs. That is where the previous preparation of our data set was important.



TRAINING TESTING

Accuracy = 18% Accuracy = 52 %

Fig 11: Results of LDA algorithm on classification of our training and testing set

- Thus, even with this method, these accuracies are still not really high! It seemed pretty logic that, using one method or the other would lead to approximatively the same results, as the 3 measures are directly linked with the choice of the tools.

c. Classification using KNN

Related files : Notebook 3 – part 2

- Even if we don't have lot of data (that is a problem for kNN), we decided to try kNN algorithm with different values of k. In fact, it seems that there are lots of groups of children classified with the same tool at some space (cf. p.7).

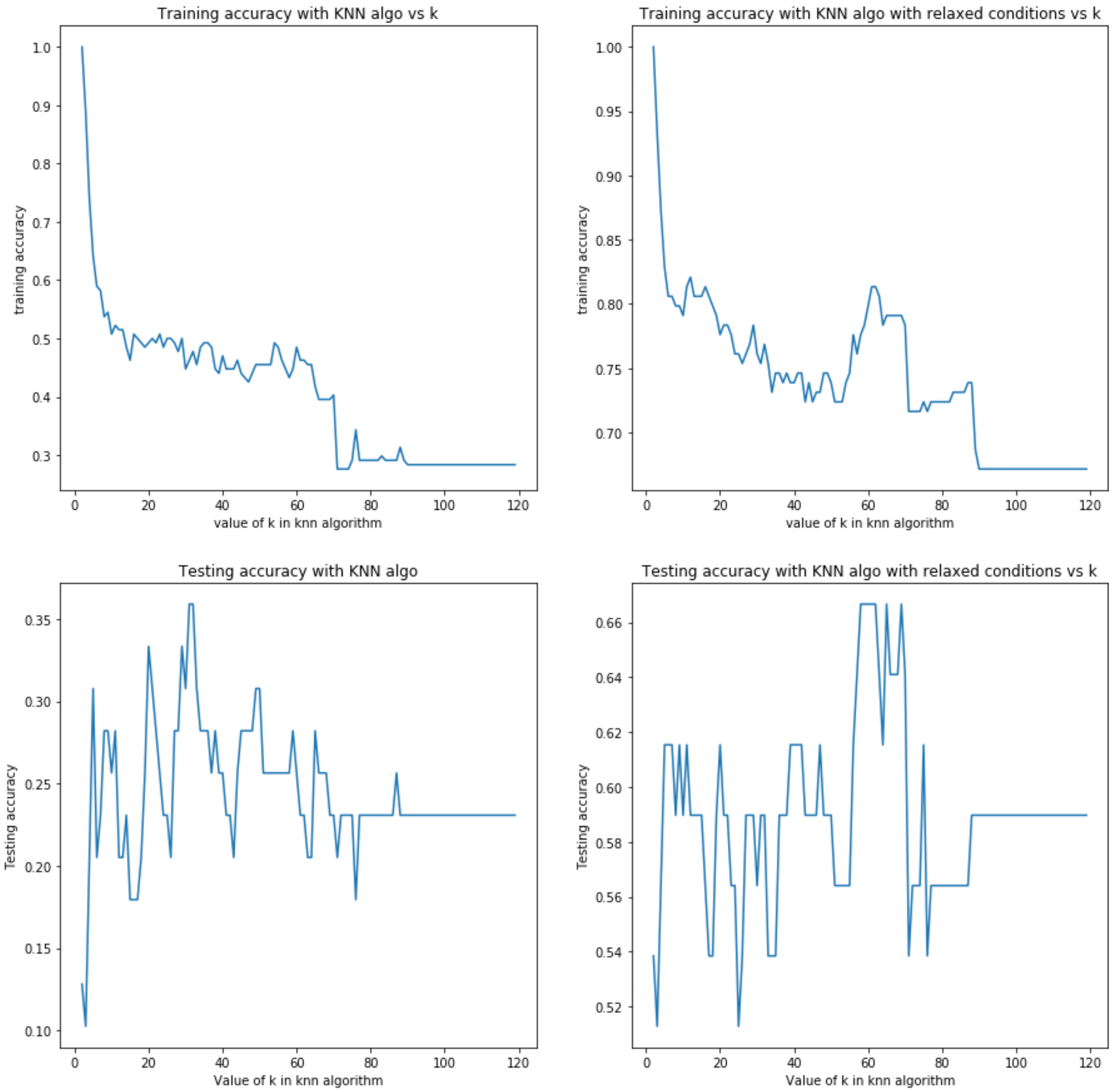


Fig 12: Repartition of the children of data we have among the different tube sizes (only left side, the right side is explained in the next section)

Conclusion: The accuracy is still really low. There is no real good trade-off between overfitting with a small k and good testing accuracy. One of the best values is for k = 60, we obtain a training accuracy of 50% and a testing accuracy of 26%.

III. Need to relax the constraints

- As a result of these really low accuracies, we decided to relax the constraints. We discussed again with the doctor explaining the problem and the possibilities. We decided to say that the tube is ok if the diameter of the tube predicted is smaller than the “optimal perfect tube” satisfying theoretically all the constraints.
- We did again all the algorithms of the section 2. That are the results:

Methods/Accuracy	With all constraints		Relaxing constraints	
	Training	Testing	Training	Testing
Classification on 3 linear independent linear regression	29%	23%	53%	56%
LDA	20%	59%	65%	87%
kNN (for the optimal value of k: best trade-off) (cf. p.10 to see the result with relaxed constraints right graph)	50%	35%	82%	67%

Fig 13: Summary of the different accuracies we obtained in function of the method with all constraints or with relaxed constraints

IV. Analysis of the results and conclusion

- Of course, as expected from the beginning, the huge challenge of this project is the amount of data versus the number of classification classes (tube sizes).
- What we can say, for sure, to the doctor is that :
  - Knowing only the age and the weight of the child, he has in best cases around 30% of chance to choose the exact right tool.
  - He also has around 70% of chance that he will choose an adapted one (he might just need to take a bigger tube and to reintubate the child).

  It's even really better than what’s done currently. According to doctors, in around only 10% of the cases, there is no need to reintubate the child!

V. What he have learned about the importance of the data set size

- With this project, we really understood the complexity of the amount of data problem in medical field. It seems that, in this field, having a lot a data is a hard to achieve while it is particularly important :
  - Indeed, every child is different : even if we could find a general pattern, there will always be special cases that need to be detected by the algorithms (KNN was, for that a good approach as we saw above).
  - Remark : Moreover, the younger a child is, the more difficult it is to find a common pattern, even in general cases.
- You would have understood that we need more data!! But, as they are data very complicated to have, it’s important to define an amount which can be deterministic. Using statistical power estimation could be a way of doing it.

ANNEX

Location of the measurement	Type of measurement	Unit
Epiglottis - Glottic plane	Distance between the top of the epiglottis and the glottic plane	mm
Glottic plane	Antero-posterior diameter	mm
	Surface	mm^2
Glottic plane - lower cricoidal plane	Distance between the glottic plane and the lower edge of the cricoid	mm
Lower edge of the cricoidal cartilage	Antero-posterior diameter	mm
	Transverse diameter	mm
	Surface	mm^2
Trachea	Distance between the lower edge of the cricoid and the tracheal carina = tracheal length	mm
	Distance between the glottic plane and the tracheal carina	mm
Distal trachea	Antero-posterior diameter	mm
	Transverse diameter	mm
	Surface	mm^2
Right and left stem bronchial tubes	Length of the stem bronchial tubes	mm
Right and left stem bronchial tubes (proximal measurements)	Antero-posterior diameter	mm
	Transverse diameter	mm
	Surface	mm^2
Right and left stem bronchial tubes (distal measurements)	Antero-posterior diameter	mm
	Transverse diameter	mm

Fig 14: Different measures we have

ORDERING INFORMATION:

REF	Description	I.D. (mm)	O.D. (mm)	Length (mm)	Cuff Ø (mm)	
cuffless						
9320E	2.0mm Endotracheal Tube UNCUF	2.0	2.9	130	-	PAEDIATRIC
9325E	2.5mm Endotracheal Tube UNCUF	2.5	3.6	140	-	
9336E	3.0mm Endotracheal Tube UNCUF	3.0	4.2	160	-	
9335E	3.5mm Endotracheal Tube UNCUF	3.5	4.9	180	-	
9342E	4.0mm Endotracheal Tube UNCUF	4.0	5.5	200	-	
9345E	4.5mm Endotracheal Tube UNCUF	4.5	6.2	220	-	
9350E	5.0mm Endotracheal Tube UNCUF	5.0	6.8	240	-	
9360E	5.5mm Endotracheal Tube UNCUF	5.5	7.5	270	-	
9366E	6.0mm Endotracheal Tube UNCUF	6.0	8.2	280	-	
9365E	6.5mm Endotracheal Tube UNCUF	6.5	8.8	290	-	
9370E	7.0mm Endotracheal Tube UNCUF	7.0	9.6	310	-	
cuffed						
9430E	3.0mm Endotracheal Tube CUFD	3.0	4.2	160	8	PAEDIATRIC
9440E	4.0mm Endotracheal Tube CUFD	4.0	5.5	200	11	
9450E	5.0mm Endotracheal Tube CUFD	5.0	6.8	240	16	
9555E	5.5mm Endotracheal Tube CUFD	5.5	7.5	270	17	
9460E	6.0mm Endotracheal Tube CUFD	6.0	8.2	280	22	
9465E	6.5mm Endotracheal Tube CUFD	6.5	8.8	290	22	
9570E	7.0mm Endotracheal Tube CUFD	7.0	9.6	310	25	
9475E	7.5mm Endotracheal Tube CUFD	7.5	10.2	320	25	
9480E	8.0mm Endotracheal Tube CUFD	8.0	10.9	320	27	
9485E	8.5mm Endotracheal Tube CUFD	8.5	11.5	320	27	
9590E	9.0mm Endotracheal Tube CUFD	9.0	12.1	320	29	
9495E	9.5mm Endotracheal Tube CUFD	9.5	12.8	320	29	
9500E	10.0mm Endotracheal Tube CUFD	10.0	13.5	320	32	

Fig 15: Malinkrot Tubes (reference given by R.L.)