

UNIVERSITAT AUTÒNOMA DE BARCELONA  
FACULTAT DE CIÈNCIES



# MEMÒRIA DE PRÀCTIQUES

*Pràctiques Externes*



Eric Jiménez Barril

Tutora de l'assignatura: M<sup>a</sup> Dolors Herbera Espinal

Tutors de l'empresa: Marc Jordà Mascaró i Felipe Basurto Barrio

# Índex

<b>1</b>	<b>Carta de presentació</b>	<b>2</b>
<b>2</b>	<b>Curriculum Vitae Actualitzat</b>	<b>3</b>
<b>3</b>	<b>Dades generals de l'entitat i de l'estada de pràctiques</b>	<b>4</b>
3.1	Dates de l'estada a l'entitat i tutor assignat per l'entitat . . . . .	4
3.2	Dades de l'entitat . . . . .	4
<b>4</b>	<b>Procés per aconseguir una entitat per fer les pràctiques</b>	<b>5</b>
4.1	Procés de cerca de l'entitat . . . . .	5
4.2	Consells per afrontar una entrevista de feina . . . . .	6
<b>5</b>	<b>Tasques realitzades dins l'empresa</b>	<b>7</b>
5.1	Tasques generals i primera setmana: Onboarding . . . . .	7
5.2	Segona setmana: Documentació específica de <i>QA</i> i primera tasca . . . . .	7
5.3	Tercera setmana: Darrera tasca a <i>CBAI</i> i introducció a <i>Lab Site Health</i> . . . . .	8
5.4	Lab Site Health . . . . .	9
5.4.1	EDA of Scores . . . . .	9
5.4.2	Gain Charts . . . . .	10
5.4.3	<i>Non-Pyramidal KPIs</i> . . . . .	10
5.4.4	Validation Pipeline . . . . .	11
5.4.5	Real Example of Fuzzy Logic . . . . .	12
5.5	Ajudes en altres projectes a companys . . . . .	13
5.5.1	Audit Case . . . . .	13
5.5.2	Score Recommendation Rank . . . . .	13
5.6	Eines Utilitzades . . . . .	14
<b>6</b>	<b>Condicions de feina</b>	<b>14</b>
<b>7</b>	<b>Impresions personals</b>	<b>15</b>
<b>8</b>	<b>Annexos</b>	<b>16</b>
8.1	Exemple codi . . . . .	16
8.2	Exemple Matriu de Confusió Multiclasse . . . . .	17
8.3	Example of some graphics of EDA of Score . . . . .	17
8.4	Grafics realitzats per explicar la funció del <i>Non-Pyramidal KPI</i> . . . . .	18
8.5	Exemple d'un <i>Jump</i> i una <i>Insensitivity</i> . . . . .	18
8.6	Exemple del Validation Bot . . . . .	19
8.7	Exemple del Slack Alert Bot . . . . .	19
8.8	Fuzzy Logic Real Example . . . . .	20

# 1 Carta de presentació

Estimades professores de l'assignatura i estudiants lectors:

Em permeto dirigir-me a vostès per presentar la memòria final de l'assignatura de Pràctiques Externes, que he tingut l'honor de realitzar sota la supervisió de Marc Jordà i Felipe Basurto a Aily Labs, durant l'últim any del grau en Matemàtiques a la Universitat Autònoma de Barcelona.

El meu nom és Eric Jiménez Barril, i a través d'aquest document, desitjo compartir no només les experiències i coneixements adquirits durant el període de pràctiques, sinó també com aquestes han enriquit la meva formació acadèmica i professional.

A l'inici dels meus estudis en el grau de Matemàtiques, no sabia molt bé on volia enfocar el meu futur laboral. Vaig escollir aquesta carrera segurament per la facilitat que tenia per les matemàtiques durant els períodes escolars anteriors i aprofitant que, amb la nota obtinguda en el batxillerat, se'm va obrir un ventall de possibilitats que em van permetre escollir gaire bé qualsevol grau, sent el de Matemàtiques el que em proporcionava més sortides laborals variades i per al qual tenia més afinitat.

Sempre he sigut una persona molt motivada per millorar dia a dia i preocupada pel meu futur, cosa que em va portar des d'un primer moment a buscar l'excel·lència acadèmica, així com, alhora, a explorar quines possibilitats laborals podien motivar-me per enfocar la meva carrera.

Ràpidament em vaig adonar que el món relacionat amb la ciència de dades i la intel·ligència artificial és el que més em motiva per centrar la meva carrera. Encara que no sigui un apassionat per la programació, poder utilitzar aquesta com una eina per aplicar diferents processos i coneixements per automatitzar i optimitzar sistemes em sembla fascinant i molt útil dins del món empresarial.

És en aquest moment on les pràctiques a Aily Labs m'estan brindant una oportunitat excepcional per submergir-me en el món de la ciència i l'anàlisi de dades, sempre des d'una perspectiva innovadora i on es prioritza el producte. Em fascina el fet que a Aily Labs hagi tingut la oportunitat d'obtenir una formació àmplia i alhora m'han permès desenvolupar la meua creativitat dins d'un marc matemàtic, podent modelitzar solucions per diferents problemes des d'un punt de vista matemàtic. Això, juntament amb els coneixements adquirits dins del marc informàtic (com temes enfocats a bases de dades o la utilització de diferents softwares) ha fet que aquesta experiència hagi sigut excel·lent tenint en compte la seva curta duració.

Ha sigut sorprenent com dins d'un entorn col·laboratiu i àgil, on en un primer moment pensava que tot seria mecanitzat, hagi trobat a més un lloc on prioritzen la creativitat i sobretot la innovació en les solucions. M'ha semblat fascinant poder treballar colze a colze amb persones amb molta més experiència i altament qualificades on (un cop adequat a l'entorn) em van fer sentir un més i podia aplicar els meus coneixements per ajudar l'equip en tasques realment importants i útils per al projecte que s'estava desenvolupant.

Estic profundament agraït per l'oportunitat que em van brindar de portar a terme aquestes pràctiques, les quals considero fonamentals en el meu camí cap al desenvolupament professional. L'experiència ha estat summament enriquidora, permetent-me no solament aplicar en la pràctica els meus coneixements teòrics, sinó també adquirir noves habilitats i competències. És tanta la meua gratitud a l'empresa que ara estic en la indecisió de saber si continuar treballant en aquesta o començar els meus estudis en un Màster de Data Science o Artificial Intelligence a la UPC.<sup>1</sup> Estic convençut que ambdós camins seran bons pel meu futur però haig de valorar quin pot ser més decisiu per continuar mantenint l'excel·lència en la meua carrera.

Agraeixo la seva atenció i el suport brindat durant aquest procés. Quedo a la seva disposició per discutir qualsevol aspecte de la memòria o proporcionar informació addicional que considerin pertinent, així com per respondre qualsevol tipus de dubte que puguin tenir.

Atentament,

Eric Jiménez Barri a 30 d'abril de 2024,

Assignatura de Pràctiques Externes, Grau en Matemàtiques, Universitat Autònoma de Barcelona.

---

<sup>1</sup> Això ho vaig escriure justament abans d'acabar les pràctiques, a dia d'avui puc dir que finalment continuaré a la empresa com a *Junior Data Science 1*

## 2 Currículum Vitae Actualitzat



23/08/2002

### Contact Information



+34-691498916



ericjiba3@gmail.com



Carrer Tamarit 84,  
Ripollet, 08291, España



ericjimenezbarril



ericjimenezbarril

### Additional Information

- Driving license.
- Own vehicle.

### Skills

- Strong learning ability and perseverance.
- Effective problem-solving and analytical thinking.
- Active and effective communication.
- Leadership and motivation.
- Teamwork and collaboration.

### Programming Skills

- Python
- C
- R
- SQL
- Java
- SageMath

### Extracurricular Activities

Octubre 2017 - Mayo 2018

- Volunteer at Esplai Estel Ripollet.

# ERIC JIMÉNEZ BARRIL

## Mathematics student

Final year Mathematics student with a passion for specialising in Data Science and Artificial Intelligence. I aim to apply my theoretical knowledge in business environments to solve complex problems and add value.

### Work Experience

**AILY LABS:** February 2024 - Present

#### • Intern Data Science

- Responsibilities: designing, developing, and industrializing Machine Learning and Deep Learning models.
- Simulation of scenarios.
- Focused on Python OOP for optimising Data Preprocessing, Feature Engineering, and Modeling processes.

### Academic Background

**Bachelor's Degree in Mathematics, Universitat Autònoma de Barcelona** September 2020 - June 2024

- Grade: 9.07 (210/240 cdt)
- Credits with honour: 90/210
- **Award for Outstanding First Cycle Student in Mathematics.**
- Relevant subjects: Modelling, Numerical Methods and Integration, Statistics, Advanced Programming, Time Series, Machine Learning, Financial Management.
- Significant projects:
  - 1st year. *Initial Coin Offering - Blockchain & Cryptocurrencies.*
  - 2nd year. *An automatic parking system for autonomous vehicles.*
  - 4th year. *4 in a row with C.*
  - 4th year. *Ghost 'n Goblins with Python.*
  - 4th year. *Unsupervised Learning analysis in PISA (End of degree project).*

**Scientific-Technological Baccalaureate, INS CAN MAS, Ripollet**  
Septiembre 2018 - Julio 2020

- Grade : 9.44 (Selectividad: 12.989)
- Significant project:
  - Power activo bicycle (Awarded with Honours)

### Additional Training

**IBM Data Analyst Professional Certificate** - Coursera

1. *Python for Data Science, AI & Development*
2. *Data Analysis & Visualization with Python*

### Intro to SQL

- Kaggle

### Advanced SQL

- Kaggle

### Languages

**Spanish and Catalan:** Native.

**English:** Upper - Intermediate (Currently studying level B2.2).

### Other Work Experience

**Private Maths Tutoring:** October 2020 - May 2022

**Accept@-Servicios Integrales:** July 2023 - August 2023

- Telephone Sales Assistant.
- Course and Training Program for Insurance and Reinsurance Distributors (Level II - 163 hours)

**Emergen Disc SL:** Mayo 2022 - Noviembre 2022

- Waiter assistance and access control at a nightlife company

## 3 Dades generals de l'entitat i de l'estada de pràctiques

### 3.1 Dates de l'estada a l'entitat i tutor assignat per l'entitat

Durant la meua estada de pràctiques curriculars (del dilluns 5 de febrer al 30 d'abril) he treballat generalment de 9:00 a 13:00 i de 14:00 a 15:00, de dilluns a divendres (no festius). En la secció 6 de Condicions de Treball detallaré més sobre les condicions. Durant el període, els meus tutors assignats per l'entitat de pràctiques han estat, en primer lloc, Marc Jordà Mascaró, llicenciat en Física i Matemàtiques per la UB i especialitzat en Enginyeria i Ciència de Dades, i Felipe Basurto Barrio, llicenciat en Enginyeria Informàtica per la Universitat de Burgos i especialitzat en *Business Analytics and Big Data* a l'IE School.

Com explicaré més endavant, l'empresa treballa en petits grups; dins del meu grup, en Marc era el líder, actualment com a *Senior Data Scientist*; per tant, les seves tasques com a tutor han estat orientar-me sobre quines tasques havia de realitzar, explicar-me el funcionament de l'empresa i corregir i revisar les mateixes.

D'altra banda, en Felipe era un altre membre del grup, actualment *Junior Data Scientist I*, va ser assignat el primer dia com el meu *buddy* i el seu paper ha estat ajudar-me i orientar-me en les tasques diàries, així com en el funcionament dels softwares utilitzats i els dubtes sobre els projectes en què treballàvem.

### 3.2 Dades de l'entitat

*Aily Labs* és una empresa de tecnologia amb la missió de democratitzar la *AI*, portant els productes d'*Aily* a les empreses més grans del món. L'organització està dissenyada de manera que aquesta missió es pugui acomplir de la manera més eficaç possible.

L'organització està dividida en 3 àrees principals:

1. Demanda i Mercat (portant negocis): liderada pel COO i el *Lead of Business Impact*, encarregats de gestionar, fer créixer i adquirir nous comptes de clients.
2. Producte (dissenyant i decidint sobre el pla de desenvolupament del producte): liderada pel Director d'Experiència (CXO) i el Vicepresident de Producte. Són els encarregats de decidir què es construeix per al mercat i com es veurà.
3. Subministrament (construcció del producte): liderada pel Cap d'IA i Tecnologia. Aquesta àrea és responsable de construir el millor producte possible, amb els estàndards més alts i satisfent les necessitats del mercat (i on he treballat jo).

La empresa va ser formada al juny del 2020 per 3 treballadors de la farmacèutica *Novartis* que van decidir emprendre amb aquesta empresa liderada per Bianca Diana Anghelina (CEO), Christoph Schmitt (COO) i Sara Bisbe (C-AI-T-O).

La empresa té dues oficines a Espanya (Barcelona i Madrid) i una a Alemanya (Munic). Únicament a les dues oficines espanyoles hi ha representació del departament d'*AI*, representats principalment (a part de la Sara Bisbe) per Pau Ferrer (*VP of AI*) a Barcelona i Jaime Zamora (*Head of Madrid Office*) a Madrid. A més, el Jaime Zamora és *Head of AI Correlator + M&S (Supply, CAPA, QA, OCR, M&S)*, per tant com veureu més endavant ha sigut un dels meus *Principal DS*.

Actualment, té al voltant de 300 treballadors, una xifra que s'ha incrementat en més d'un 100% al 2023 i tenen millors objectius per al 2024. Al departament d'*AI* som prop de 150 treballadors i és un dels departaments que més ha crescut.

## 4 Procés per aconseguir una entitat per fer les pràctiques

### 4.1 Procés de cerca de l'entitat

Per la meua part, vaig començar la recerca de pràctiques potser massa aviat. Per un tema econòmic personal, vaig decidir començar la recerca de pràctiques a partir del segon semestre del tercer curs del grau.

Els primers passos que vaig fer van ser actualitzar el meu currículum vitae així com explorar diferents sectors que actualment poguessin ajustar-se al que volia, i explorar plataformes com InfoJobs i aplicacions com LinkedIn, on sobretot en la segona, les oportunitats disponibles al camp de la ciència de dades són molt àmplies.

A més, vaig contactar amb la meua actual tutora de pràctiques, Dolors Herbera, amb qui vaig tenir un parell de reunions inicials per saber com podia començar la cerca i vaig ser afegit al grup de correu on arribaven diferents ofertes.

Un cop tot estava llest, vaig enviar diversos currículums i cartes de presentació a entitats que s'ajustaven al meu gust o que al menys, oferien llocs de treball que cridaven la meua atenció. Però totes les respostes van ser negatives. No vaig arribar ni a fer cap entrevista. Al ser del grau de matemàtiques, al tercer curs encara no tenim una formació específica en cap àmbit de la ciència de dades (*Machine Learning (ML)*, *Deep Learning (DL)*, coneixements avançats en programació en *Python*), a més, gran part de les empreses busquen estudiants d'últim any per després oferir un contracte a la empresa. Per tant vaig decidir deixar les pràctiques curriculars per al quart curs i buscar un treball d'estiu en un altre sector.

Al començar el quart curs, va sorgir la possibilitat d'obtenir una beca d'*Aily Labs*, que donava l'opunitat de fer pràctiques remunerades amb ells a més d'una beca escolar. Des del dia que es van fer les jornades ocupacionals m'havia cridat l'atenció l'empresa, i vaig endinsar-me en adaptar el currículum a les preferències de la empresa, informant-me de què feien i quin perfil de treballadors tenien (via LinkedIn). No obstant, no vaig rebre resposta d'aquesta sol·licitud durant més de dos mesos, en els quals vaig estar centrat en superar el curs de la millor manera possible.

La última setmana de novembre vaig rebre resposta que havien seleccionat el meu perfil per començar el procés de selecció, el qual constava en principi de quatre parts:

1. Un formulari de contractació. Aquest simplement era un formulari amb les teves dades personals i preguntes personals.
2. Un examen *online* amb límit de temps de ciència de dades, que barrejava preguntes bàsiques de matemàtiques i estadística amb altres de *ML* i *SQL* més centrades en *Software*.
3. Una entrevista amb un membre de recursos humans.
4. Una entrevista amb un membre de l'equip de Data Science.

Dos dies (laborables) després vaig realitzar la primera i finalment única entrevista. Vaig realitzar aquesta entrevista amb Clara Zaldivar (*Data Scientist*) i Miguel Toquero (actual *Senior Data Scientist*). La entrevista va ser 100% tècnica. Encara que no va ser 100% en anglès com pensava en un primer moment, va tenir preguntes tant en castellà com en anglès. Totes les preguntes (llevat de què estudiava i la meua assignatura preferida) van estar relacionades amb *ML* i *Software* en general.

El meu error va ser pensar que la entrevista seria com les típiques on et pregunten sobre tu, i no havia preparat gaire en profunditat aquests temes, per tant algunes preguntes no vaig saber respondre, però en tot moment vaig ser jo mateix, em vaig expressar tal com soc i amb honestitat i naturalitat. Crec que aquest punt juntament amb el bon resultat de l'examen va derivar en la bona impressió que van tenir de mi.

Després d'un parell de setmanes, vaig rebre un correu que confirmava que havia superat el procés de selecció d'*Aily Labs* i em van donar total llibertat per escollir quan volia començar les meves pràctiques. Va ser un moment d'il·lusió i satisfacció de saber que al primer i únic lloc fins al moment on havia començat el procés de selecció, vaig obtenir resultat satisfactori. No obstant, encara que vaig contestar aquell correu indicant quines dates (aproximades) s'ajustaven millor al meu calendari escolar, no vaig rebre resposta d'aquell missatge (ni a altres 3 missatges enviats durant desembre i principis de gener sense obtenir resposta).

En aquest transcurs de temps i durant el període de Nadal, vaig retomar la cerca de pràctiques curriculars, sempre a l'espera d'una resposta d'*Aily Labs*. Com que sabia que havia superat aquest procés de selecció de forma satisfactòria, em vaig posar a treballar els aspectes que pitjor havia tingut. Vaig realitzar fins a 4 cursos de *Python for Data Science* a *Coursera* (que són gratuïts a través de la UAB) i 2 cursos de *SQL* a *Kaggle*, una web (gratuïta) que em va recomanar l'entrevistador Miguel, on hi ha tant cursos de formació de diferents àmbits de Data Science (*ML*, *DL*, *SQL*, *Neural Networks* ).

Em van trucar del Banc Sabadell (sense haver aplicat a cap programa) ja que estaven interessats pel meu perfil (ja que la Anna Ratera, professora de problemes de Sèries Temporals, em va animar a fer les pràctiques al seu grup de treball i va entregar el meu CV), però la trucada va ser només per parlar amb mi i motivar-me a fer el procés de selecció.

Motivat per aquesta trucada el vaig realitzar (van ser en aquest cas 6 examens, 4 típics de coeficient intel·lectual i destresa i 2 d'anglès) i superar (el procés de selecció) i vaig fer una primera entrevista amb una empresa externa de recursos humans en la qual em van seleccionar pel lloc de treball. Curiosament aquell mateix dia vaig enviar un altre correu a una altra persona d'*Aily Labs* de recursos humans com a última esperança ja que, sent honest, no m'agradava gens la feina que m'havien descrit per al Banc Sabadell.

Finalment aquell mateix dia em van contestar el correu d'*Aily Labs* i va arribar al seu fi el procés de selecció. Em van indicar el dia i l'hora per començar el procés d'*onboarding* i vam parlar de com fer el conveni.

Vaig recomanar a l'Anna al meu company Arnau Perich per fer les pràctiques al seu grup i finalment va ser ell qui va ocupar la plaça al Banc Sabadell.

## 4.2 Consells per afrontar una entrevista de feina

Com a consell per a tots aquells lectors que Si estàs pensant en buscar pràctiques o ja estàs en pràctiques, recomanaria els següents punts, deixant de banda els típics consells com investigar l'empresa, ressaltar les habilitats tècniques i personals que encaixin al lloc de treball, i mantenir una actitud positiva:

1. Demuestra les teves habilitats analítiques i de resolució de problemes. Explica com has resolt problemes tècnics en els teus estudis o projectes personals, incloent exemples dels anàlisis que has realitzat. Això demostrarà la teva capacitat per aplicar els coneixements en situacions pràctiques.
2. Parla sobre la teva experiència amb eines d'anàlisi de dades, especialment *Python* i *SQL*, i descriu en quins projectes les has utilitzat.
3. Expressa la teva passió per l'aprenentatge continu. Destaca les teves ganes d'aprendre més sobre el camp en el qual desitges fer les pràctiques.

Per arribar a una entrevista, és essencial trobar un lloc que et motive realment. Aquí tens alguns consells per destacar en la recerca de pràctiques:

1. Presenta't de manera professional amb un bon currículum i perfil de LinkedIn ben cuidat.
2. Incrementa les interaccions en LinkedIn, començant per connectar-te amb companys de classe i gradualment expandint la teva xarxa professional.
3. No et limitis a la formació acadèmica; explora cursos online en plataformes com *Coursera*, *DataCamp*, *Udemy*, i *Kaggle*, que ofereixen coneixement tècnic complementari.
4. Assistir a xerrades i fires relacionades amb els teus interessos, com els esdeveniments organitzats per la UAB o altres entitats professionals, pot ser molt enriquidor i útil per ampliar els teus contactes.

## 5 Tasques realitzades dins l'empresa

### 5.1 Tasques generals i primera setmana: Onboarding

Durant el meu període de pràctiques dins l'empresa, he pogut viure l'experiència de treballar en un equip de treball com si fos un *Junior Data Scientist* més.

El primer dia d'*onboarding* es va basar principalment en la instal·lació i configuració de l'ordinador de feina, així com en l'assignació de l'equip de treball al qual anava a pertànyer. Aquest equip va ser *QA* (*Quality Assurance*), dins del grup de *M&S* (*Manufacturing and Supply*) del departament de *AI*. Durant aquest dia, en Marc Jordà (tutor dins l'entitat) es va reunir amb mi amb la finalitat de conèixer-nos a nivell personal i explicar-me com treballava l'empresa, la seva experiència tant a nivell professional com acadèmica prèvia, com funcionava l'equip de treball i alhora, transmetre'm tota la seva confiança. També em van assignar al meu *buddy*, la persona que m'acompanyaria durant tota l'estada de pràctiques, en Felipe Basurto, *Junior Data Scientist* de l'equip.

A diari, a primera hora es realitzava una *DSU* (*Daily Stand-Up*). En aquestes reunions de 15-30 minuts, tots els membres de l'equip de *QA* exposaven el treball realitzat el dia anterior tant als seus companys com al Senior de l'equip (en aquest cas, en Marc Jordà). Explicaven quins eren els seus objectius de treball i reunions per aquell dia (excepte dilluns, on s'exposaven els objectius setmanals i es parlava també de com havia anat el cap de setmana). A més, un encarregat (escollit diàriament de forma aleatòria per un *bot* programat) escrivia el *report* de la sessió.

Aquest *report* s'exposava en una reunió posterior (*Daily M&S*). En aquest *meeting*, en el qual es reunien els 5 equips dins del grup de *M&S* (*QA*, *Supply*, *Clinical Supply*, *CAPA* i *M&S*), cadascun dels encarregats de fer el report del seu equip l'exposava als encarregats de l'equip, a la Felícia Reinders (*Principal Data Science*), qui va començar el mateix dia que jo a l'empresa i a Jaime Zornoza.

Durant la primera setmana, les pràctiques es van basar principalment en realitzar l'*onboarding*. Vaig haver de fer totes les configuracions, instal·lacions de *software* i demanar diferents permisos per poder accedir a les diferents eines de treball.

Per últim, cada setmana se celebren dues reunions, una obligatòria (*AI-Weekly Catch Up*) i una opcional (*KSS Session*). En la primera, es fa un resum de tots els equips dins de *AI*, exposant els resultats més rellevants (tant a nivell tecnològic com de negoci) i cada setmana un dels equips presenta les últimes actualitzacions del seu projecte.

D'altra banda, la *KSS Session* és una reunió opcional on els assistents exposen temes relacionats amb *AI* dels quals estiguin interessats però que no es treballin a *Aily Labs* o coses que poden ser útils.

### 5.2 Segona setmana: Documentació específica de *QA* i primera tasca

La segona setmana es va dividir en dues parts:

1. **Documentació de *QA*:** Com ve he mencionat, s'em va assignar el grup de *QA* (*Quality Assurance*). Durant aquesta primera setmana vaig haver de llegir-me tota la documentació sobre el projecte principal del equip: ***Complaints by AI* (CBAI)**. Haig de dir que no va ser molt productiu ja que, vaig estar llegint tota la setmana documentació sobre aquest projecte per fer dos tasques molt petites i que no em van aportar gran benefici (però és el que s'em va manar). El projecte, del qual estava encarregat en Felipe és basava en un laboratori en el qual el propòsit era predir diferents característiques relacionades amb les queixes dels productes de *Sanofi* (principal client de *Aily Labs*). Per això, s'havien realitzat 4 models de *ML* basants en *NLP* (Natural Language Processing):

- (a) **Complaints priority:** (Classificació Multiclasse) Donava una prioritat del 1 al 4.
- (b) **Complaint defects:** (Classificació Multietiqueta) Cadascuna de les queixes podia estar associada a diferents defectes de productes (hi havia  $\sim 50$ ).
- (c) **Complaint resolution:** (Classificació Multiclasse) A cada queixa se li assignava una causa de resolució (de 17 possibles).
- (d) **Complaint finding and cause:** (Classificació multietiqueta) A cada queixa se li assignava la causa originària d'aquesta, hi havia més de 500 possibles.



2. **Primera tasca: Defect higher granularity - Error Analysis.** En el model (b), el Felipe havia afegit molts més possibles defectes (184) proposats i discutits entre els *BILs* (**B**usiness **I**mpact **L**ead) de *Sanofi* i *Aily Labs*. Eren defectes més específics i particulars. Aleshores la meua feina va estar fer un analysis dels resultats i les mètriques obtingudes pel model per veure si realment tenia un impacte o no afegir aquests models i veure quins d'aquest realment eren rellevants. Aquí haig de fer una menció a la meua companya d'assignatura Judit Panisello, la qual va fer la mateixa tasca però pel primer model i per tant vaig poder aprofitar el seu codi de *Python* per fer l'anàlisi (obviament vaig haver de fer canvis perquè el model havia canviat, però vaig poder aprofitar la base, el qual em va ajudar per acostumar-me a la forma de programar d'*Aily Labs*). A continuació, destacaré algunes de les coses relacionades amb el que vaig veure o vaig aprendre amb aquesta tasca.

- **Programació Avançada (PA) /Eines Informàtiques per les Matemàtiques (EIM)**

- **Docstrings a les funcions:** Aquest procés em va servir per veure per primer cop una cosa que mai havia vist i que en cap de les dues assignatures de programació de la carrera s'havia ensenyat, i és que les funcions a qualsevol llenguatge de programació i a qualsevol *script* han de tenir un *docstring* (explicació de la funció al inici).
- Indicar de quin tipus esperes que sigui cada variable i el *output* quan defineixes una funció.
- Respectar sempre les mateixes regles en quant a espais, salts de línia i estructura.

Per ser no sembla tant important, però després d'aquesta època a la empresa m'he adonat que he entès super bé i ràpid funcions molt més complexes que les que havia treballat, i al crear el meu *GitHub* personal que vaig revisar programes meus antics en els que havia obtingut un 10 (inclús un parell d'ells amb un missatge de felicitacions per la bona estructuració), eren un desastre en comparació. Podeu trobar un exemple representatiu al Annexes 8.1

- **Estadística Matemàtica / Investigació Operativa (Aprentatge Automàtic)**

- Durant tots els cursos d'aquestes assignatures, els models que havíem vist classificatius a la pràctica eres binaris, però com s'interpreta una matriu de confusió quan té dimensions més grans? Com bé he destacat, en aquest *error analysis* vaig haver d'interpretar matrius de confusió amb 184 possibles sortides. Per tant vaig haver d'investigar sobre el tema (en Felipe em va ajudar amb això) i aprendre a interpretar-les. L'objectiu era analitzar quins eren realment els defectes més importants i quins eres els que millor predien. Etudiant també els que sempre es predien malament, podem concloure que aquests no eres bons.
- El mateix va succeir amb les mètriques que havia estudiat a les assignatures (*Recall*, *Precision*, *F1-score* o *Accuracy*). Com s'interpreten aquestes mètriques? Ara ja no tenia només 4 opcions (FP, TP, FN, TN), per tant vaig haver d'aprendre a interpretar-les, el que m'ha aportat un coneixement útil pel futur.
- Finalment vaig aprendre que tots aquests anàlisis són una generalització de l'anàlisi binari en el qual s'estudien els percentatges de pertanyer a cada classe (al anàlisi binari no fem més que establir un *threshold* a 50% per cada classe, però podria ser diferent) i estudiar el top 10, top 3 o top 1 per cada classe (segons ens interessi).

Als Annexos 8.2 podeu veure un exemple de matrius de confusió i gràfiques de *Recall* i *Precision* que vaig interpretar. No adjunto els quadres amb les dades per confidencialitat amb *Sanofi*.

### 5.3 Tercera setmana: Darrera tasca a *CBAI* i introducció a *Lab Site Health*

La tercera setmana es va dividir en 4 parts:

- 1.
2. Vaig continuar (durant tota la setmana en estones que estava bloquejat o esperant alguna resposta) continuant llegint-me documentació de **CBAI**.
3. Vaig acabar de pulir els detalls de la primera tasca i va ser el primer cop que vaig realitzar una de les que pot ser ha sigut una de les tasques que més m'ha costat acostumar-me però de les més útils pel futur: **Aprendre què és GitHub i com s'utilitza**. A la assignatura de **Programació Avançada** vam fer una classe *teòrica* de què és *GitHub*, la qual a la pràctica no em va servir per res.

Amb ajuda vaig crear la meua primera branca i vaig realitzar el primer *Pull-Request*, havent-ho de modificar-lo amb les propostes de millors i finalment fent *merge* amb la branca *develop* del projecte. Aquesta aprenentatge va ser molt útil i he hagut de realitzar-ho com a mínim 2-3 cops en les següents setmanes. A més m'ha servit per poder-me crear un GitHub personal com he mencionat prèviament.

4. **Model Cards - QA:** Tots els *pipelines* s'havien de documentar adequadament, afegint descripcions, clients, *developers*, *BILs*, versions, models utilitzats, etc. Dels 4 models que hi havia a *QA*, només hi havia una realitzada, i vaig realitzar les 3 restants. Sincerament ha sigut la tasca que menys m'ha agradat ja que únicament havia d'entendre el codi i buscar tota la informació i reunir-la en unes taules predeterminades. Per sort, em van recomanar que copiés la que hi havia feta i s'havien de canviar detalls a cadascuna d'ella. En un parell de dies ho vaig poder acabar. Va ser molt útil per conèixer tots els Models de *CBAI* i el projecte, però va ser la última tasca que vaig realitzar d'aquest projecte per tant no va ser gaire útil.
5. Introducció a **Lab Site Health:** Aquest ha sigut l'altre projecte en el qual he estat involucrat directament en la resta de les meves pràctiques. De igual manera que amb **CBAI** vaig estar llegint documentació sobre el projecte i em van introduir a la primera tasca dins d'aquest projecte.

El que resta de la tercera setmana i la resta de pràctiques, es van centrar en el projecte de **Lab Site Health**, per tant dedicaré la següent secció a parlar sobre això.

## 5.4 Lab Site Health

**Lab Site Health** es un projecte que té com a objectiu donar a cadascuna de les fàbriques (*sites*) de *Sanofi* una puntuació única entre 0 y 100, indicant la situació ("Salut") d'aquesta. Per ella, *Aily Labs* disposa d'uns *KPIs* (*key performance indicator*) amb els quals s'ha creat una sèrie de àrees per agrupar-les i amb aquestes Àrees, s'ha creat un *Health Pulse Score (HP)*, que és la mesura que les empreses veuen a la seva app (obviament poden veure totes les marques que obtenen a cadascun dels *KPIs* i àrees també). Donaré una mica de context sobre el projecte (sense entrar en molt detall) per poder explicar les meves tasques en ell. El *HP* està calculat a través de les següents àrees (de moment):

- |              |                |  |
|--------------|----------------|--|
| 1. Costos    | 3. Entregues   | 5. Qualitat: La qual té 4 sub-categories |
| 2. Seguretat | 4. Intervenció |  |

Cadascun d'aquests té diferents *KPIs* (entre 1 i 5) i aquests *KPIs* poden ser de diferents tipus. Tots tenen 3 valors claus: *best value* (millor valor), *target* (valor a partir del qual no es considera bo) o *threshold* (valor a partir del qual es considera dolent). Durant el projecte, tots els valors millors que el *best value* (*bv*) o pitjors que *threshold* (*trsh*), es consideraven *bv* o *trsh* respectivament. A més, tenen un objectiu: Estar lo més aprop del *best value*. Per tant depenent com fos la seva natura es buscava o bé: maximitzar (tots els valors estaven per sota del *best value*, per exemple: percentatge de demandes a entregades a temps) o minimitzar (tots els valors eren per sobre o iguals del *best value*, per exemple: accidents mortals) o eren piramidals, aquestes podien ser de dos pitus, piramidals simètriques (la distància al *best value* era el que realment importava), asimètriques (es donava diferent importància si estaves per sota o per sobre del *best value*). A més, a cadascun dels *sites* pertanyia a un arquetip diferent (hi havia 4 diferents). La última cosa necessària per tenir tot el context és que tant *Sanofi* (*external*) com *Aily* (*internal*) tenia uns *sites* ficticis simulats per tenir referències amb les dades (*Benchmarks*).

Per fer el càlcul del valor dels Scores de les àrees a través del valor dels *KPIs* i a de àrees a *HP Score*, es fa amb *Fuzzy Logic*. No explicaré detalladament què és la *Fuzzy Logic* però em sembla algo **molt últim** que permet implementar *NLR* (*Natural Language Rules*) i obtenir un valor numèric. És una tècnica matemàtica que m'ha semblat molt atractiva i útil per aplicar les matemàtiques en el món real i considero que aquests és un dels principals motius pels quals vaig decidir estudiar Matemàtiques.

### 5.4.1 EDA of Scores

La primera tasca que vaig realitzar al projecte va ser un *EDA of Scores* (*Exploratory Data Analysis*), per estudiar segons arquetip (dividint els *sites* per arquetip) els *HP Score* i els *Scores* de les àrees dels *Sites*. D'aquesta manera vam poder treure conclusions dels millors *sites*, pitjors i veure la diferència entre

els internals i externals benchmarks (no hi havia) i va ser molt útil per que l'equip tingués les seves conclusions. A mi em va servir per tenir una primera aproximació a la realitat de programar a *Python* un codi llarg i important i amb dades reals, el que va resultar molt satisfactori per a mi. Al Annex 8.3 podeu trobar exemples d'algunes gràfiques que vaig realitzar. Va estar també relacionat amb les Assignatures de **Aprenentatge Automàtic i Estadística** ja que s'havien d'interpretar el *top decile*, *top quartile*, *mediana* entre d'altres mètriques, tant el gràfic de barres com **box-plot**.

### 5.4.2 Gain Charts

Em van demanar investigar sobre les Gain Charts, implementades amb l'índex de Gini; són eines utilitzades per a l'anàlisi de models predictius. Permeten avaluar l'efectivitat d'un model comparant la distribució acumulativa d'esdeveniments (com respostes positives) entre una mostra ordenada pel model i una distribució aleatòria i ajuda a identificar què tan bé el model pot distingir entre classes o predir un esdeveniment d'interès, proporcionant una visualització clara de la capacitat predictiva del model respecte a l'atzar. L'objectiu era veure si amb aquestes eines podiem validar la *Fuzzy Logic AI Framework* del *Lab Site Health* però vaig concloure que no eren útils ja que aquestes utilitzen la probabilitat de pertanyer a una variable objectiu per validar el problema i la idea que es proposava era utilitzar l'*Score* per fer aquesta validació, però no teniem cap variable objectiu. Per tant finalment vaig concloure que no es podien utilitzar després d'una petita investigació i fer algunes conclusions.

Encara que la tasca no va ser efectiva, va ser molt valorada pel meu equip (sobre tot pel Marc) ja que van valorar la meua capacitat de ser crític i saber dir que no (que ho vaig haver de fer 3 cops, ja que els 2 primers em van replantejar el problema de forma diferent).

### 5.4.3 Non-Pyramidal KPIs

Aquesta va ser la primera tasca on vaig aplicar els meus coneixements matemàtics de la carrera (no relacionats amb optatives de quart). Pot ser no relacionats amb cap assignatura però si amb la capacitat d'extracció. Un problema que hi tenia l'equip era que no sabia com tractar els KPIs que no eren simètrics. Com havíem de tenir tots els *Scores* entre 0 i 100, si un KPI era simètric i tenia un valor negatiu, simplement es feia:

$$\text{Mirror Value} = 2 \times \text{Best Value} - \text{Value}$$

que és un canvi de referència simple, on pases d'un costat del triangle a l'altre (pensat en un triangle equilàter on el punt més alt és el best value). El problema amb els KPIs no simètrics era que això no es podia fer, ja que el pes a cada costat del *best value* era diferent. Per tant vaig desenvolupar la següent funció (que sincerament em va costar poc temps):

$$f: X \rightarrow X^*$$

$$x \mapsto \begin{cases} x_0 + f_1(x) & x_1 \leq x \leq x_0 \\ x_0 + f_1(x_1) + f_2(x) & x_2 \leq x \leq x_1 \\ x_0 + f_1(x_1) + f_2(x_2) + f_3(x) & x \leq x_2 \end{cases}$$

$$\begin{array}{lll} f_1: [x_1, x_0] \rightarrow [x_0, x_1^*] & f_2: [x_2, x_1] \rightarrow [0, x_2^* - x_1^*] & f_3: (-\infty, x_2] \rightarrow [0, \infty) \\ x \mapsto \frac{x_1^* - x_0}{x_0 - x_1}(x_0 - x) & x \mapsto \frac{x_2^* - x_1^*}{x_1 - x_2}(x_1 - x) & x \mapsto (x_2 - x) \end{array}$$

on  $X = (-\infty, x_0)$  and  $X^* = (x_0, \infty)$ ,  $x_0$ : Best value,  $x_1$ : Target (under the best value),  $x_2$ : Threshold (under the best value),  $x_1^*$ : Target (over the best value),  $x_2^*$ : Threshold (over the best value),  $x$ : Value under the best value for which we want to find its "symmetric".

Va agradar molt a tot l'equip i van quedar molt contents ja que solucionava el problema i era molt senzilla d'implementar a *Python* (molt més que escribia a nivell matemàtic). No obstant per explicar-la a clients i jefes vaig haver de realitzar algun gràfic per que fos més entendible (que adjunto al Annex 8.4. Posteriorment el vaig implementar a la *pipeline* principal on es feien tots els càlculs dels *KPIs* .

#### 5.4.4 Validation Pipeline

Aquest ha sigut el projecte que ha sigut més important a les meves pràctiques. Vaig començar-lo la quarta setmana i s'ha allargat fins el final de les meves pràctiques. La idea darrera d'aquesta tasca era realitzar una *pipeline* que s'executés a diari de forma automàtica i validés els *Scores* obtinguts a cada area i el *HP Score* de cada *site* i alhora, enviés un missatge a *Slack* (plataforma per on ens comuniquem i on tenim tots els grups per parlar) fent un resum de les validacions trobades. Aquest procés s'ha dividit en diferents passos que indico a continuació:

1. **Definició de les validacions: Inconsistències.** En total, havia de fer 6 tipus de validacions que ja estaven definides. El meu treball va ser pensar com podem detectar-les matemàticament i implementar-les a *Python*. Gran part del temps el vaig invertir en posar en consens amb l'equip les coses que pensava i en solucionar problemes que feia programant, més que pensant els problemes i programant-los. Un resum de les validacions i l'*approach* matemàtic és el següent:

- Primerament, hi havia les validacions que únicament s'havien de programar (sense raonament matemàtic): Comprovar que els *HP Score* no eren ni menors ni major que totes les seves areas (o subcategories), veure que les referències dels *benchmarks* (*top-decile*, *top-quartile*, *median*) estaven ben ordenades (en aquest ordre), veure que tots els *Scores* estaven entre 0 i 100, i les *trend inconsistencies*, que bàsicament es basaven en veure que si totes les areas o subcategories tenien la mateixa tendència (no creixent o no decreixent), aleshores el *HP Score* tampoc era creixent o decreixent respecte els anteriors valors (el mateix amb les areas i els respectius *KPIs*). Aquestes validacions van tenir complicacions de programació alhora de tractar amb bases de dades i grans quantitats de dades i diferents casos i errors.
- Validacions amb *approach* matemàtic:
  - (a) *Jumps and Insensitivities* Va ser la principal validació que interessava al equip, ja que les altres estalviaven temps però es podien detectar mirant les taules. Degut a la definició de la *Fuzzy Logic*, hi havia dos esdeveniments anòmals. Eren de dos tipus, *jumps* (quan hi havia un alt gran a un area en un canvi petit de *KPIs*) o *insensitivities* (hi havia un gran canvi en els *KPIs* però no en el *Score* de l'area. Com que s'havia de detectar de forma automàtica, després de moltes proves i moltes idees la que millor va funcionar va ser pendre una regressió lineal que tenia com a *target* el *Score* de l'area i com a variables independents cada *Score* de cada *KPI*. D'aquesta manera podem fer una regressió amb els *Scores* de tots els sites i estudiar els errors que eren majors a 1.5 *standard deviation* dels errors. Encara que resumit sembla algo senzill, va ser difícil ja que van haver diferents idees previes i s'havia de convencer a tot l'equip de que això funcionava. A més no és una cosa molt intuïtiva si no tens context matemàtic. Finalment va ser satisfactori i va detectar a la perfecció tant els *jumps* com *insensitivities*. Al Annex 8.5 podeu trobar un exemple de *Jump* i de *Insensitivity*
  - (b) *Imperfect areas with Perfect KPIs*: L'objectiu d'aquesta validació era veure que si tots els *KPIs* d'un area eren perfectes, aleshores l'area havia de ser també perfecte. Però la dificultat era: Què és ser perfecte? Aquestes parts en la validació són les que han suposat més dificultats per mi, ja que s'han de tomar decisions que no estan definides, no existeix res que et digui: a partir de tal nombre ho podem considerar perfecte, quin  $\epsilon$  hagafem? Finalment, vaig decidir que un *KPI* era perfecte si el seu *Score* estava en el 10% dels millors valors entre el *best value* i el *target*, com podem veure a la foto a continuació:

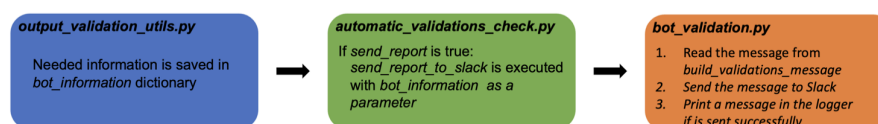


Amb això, l'únic que faltava per definir era un valor *target* pels areas i categories (ja que el *best value* era 100 però no tenien target. Això va ser l'equip que treballa amb les dades

i va decidir posar-l'ho en 90.

- (c) *Closeness benchmark references* L'objectiu era definir si dos *benchmark references* eren massa prop. Per això calia definir (de nou) que volia dir ser massa a prop. Finalment vaig definir que dos referències eren massa properes si la distància entre aquestes eren menor o igual a un mig de la *standard deviation* que hi havia entre les 3.

2. **Desenvolupament de la pipeline a Python** Ha sigut un procés molt productiu en el qual he pogut aprendre a treballar en una *pipeline* (un conjunt de *scripts*) i com poder unir-los a Python i ahora amb *GitHub*. Adjunto un esquema que pot ser explicatiu per la *pipeline* de validacions:



3. **Creació del *Health Pulse Validation Bot* per Slack** Pot ser ha estat la part més entretinguda de les pràctiques. Per crear aquest *bot* per *Slack* vaig crear un canal nou on només vaig incloure al Felipe i a mi al principi i posteriorment a tot l'equip on fer proves amb el missatge que volíem enviar. *Slack* està sincronitzat amb *GitHub*, *PyCharm* (on utilitzem *Python*), *AWS* (on emmagatzemem tot). D'aquesta manera, vam generar un *secret name* amb el nostre canal de *Slack*, que vaig emmagatzemar a *AWS* i d'aquesta manera podem sincronitzar *PyCharm* i el canal d'*Slack*. Després vaig programar (s'ha de fer d'una forma molt específica però vaig tenir un exemple d'un altre bot) el missatge i vaig anar fent diferents proves. El resultat final es pot veure al Annex 8.6.
4. **Creació del *Site Pulse Error Alert* per Slack** Aquesta va ser la meua darrera tasca a la empresa. Va ser semblant al bot anterior però implementant sistemes *try/except* o *if* a totes les *pipelines* de *Lab Site Health* per tal de captar tots els errors i produir un *raise* amb l'objectiu de no correr la *pipeline* amb errors, o taules o data frames buits, o inclús quan algun *KPI* no estigués disponible ja que no tenien gaire sentit i podia enviar taules a *S3-AWS* (on almacenem totes les taules) amb dades sense sentit. Això em va portar tots la última setmana juntament amb el Exemple 5.4.5. El resultat ho podeu veure al Annex 8.7
5. **Documentació de tot el procés de validació** Simplement va tractar de fer pàgines documentat tot el procés de validacions, dividit en 5 parts: *Goal of the pipeline*, *Validations Details*, *Structure of the Code*, *Structure of the output* i *Slack Messages*, les quals explicaven de forma més detallada i concisa els punts 1-4 que he explicat aquí.

#### 5.4.5 Real Example of Fuzzy Logic

Tot el que fem al meu equip de treball va destinat a la empresa de *Sanofi*. Per aquest motiu, tots els models que desenvolupem han de ser explicables, per tal que com a mínim els *BILs* de *Sanofi* siguin capaços d'entendre com treballem. Per aquest motiu em van demanar desenvolupar una presentació per enviar a l'equip de *Sanofi* sobre com implementàvem la *Fuzzy Logic* a *Lab Site Health* i que fos amb un exemple REAL i de forma clara i entendible per tots els públics. En primer lloc vam voler fer-ho per un site complet i explicar el procés de calcular el *HP Score*, però amb la nova actualització de *KPIs* tot va complicar-se i finalment vam decidir fer-lo únicament pel càlcul d'un Score. Al Annex 8.8 podeu veure alguns exemples de diapositives realitzades. Va ser una feina **molt necessària** i probablement on vaig tenir més comentaris de millora, ja que és molt difícil explicar una cosa tan complexa i que a mi mateix em costava entendre al principi, sense utilitzar conceptes matemàtics. La *Fuzzy Logic* que utilitzàvem era amb *Inferència de Mamdani* i per això ens quedem amb màxim valor de Score quan hi ha un "or" (s'ha d'escollir una situació) i amb el mínim quan hi ha un "and" (s'han de donar totes les situacions). Als *BILs* de l'empresa i *Principal DS* els hi va agradar molt la presentació, no se la opinió de *Sanofi* ja que va ser enviada la meua darrera setmana a l'empresa i no se he rebut el *feedback* sobre la mateixa.

## 5.5 Ajudes en altres projectes a companys

### 5.5.1 Audit Case

Un company estava realitzant un projecte relacionat amb les auditories que havia de fer cada site, en el qual l'objectiu era, donats 2 factors: la data màxima en la qual s'havia de fer una autotòria, el *Score* dels KPIs de risc (diferents als de *Lab Site Health*), que indiquen del 0 al 100 quan de malament està l'empresa (sent 0 res i 100 molt malament). Necessitaven obtenir un *Score* global amb aquests dos factors i que satisfés un munt de condicions:

- Que la data només afegís risc si era menor a un any.
- Que el risc fos màxim si la data era menor a 3 mesos.
- Que si la data era menor a un any, l'*Score* de risc dels KPIs només influís si era major a 50.
- Que l'*Score* final mínim si el temps era menys d'un any i més de 3 mesos fos major o igual a 80 i menor que 100.
- Que com més a prop de 3 mesos fos la *deadline*, més risc hi hagués (podia no ser lineal).

Amb tantes restriccions, vaig veure fàcilment que s'havia de pensar en una funció a trossos i que a més fos exponencial en el temps. Vaig desenvolupar la següent fórmula:

$$SC(TR) = 100 \cdot \left( e^{-\beta(\max(\alpha, TR) - \alpha)} \right) \cdot \mathbb{1}_{(TR \leq \lambda)}$$

$$Score(TR, SKPI) = SC(TR) + SKPI \cdot \mathbb{1}_{(SC(TR)=0)} + (100 - SC(TR)) \cdot \left( \frac{(SKPI - \gamma)}{100} \right) \cdot \mathbb{1}_{(SC(TR)>0, SKPI \geq \gamma)}$$

where

- |   |  |
|---|--|
| $\lambda = 1$   | in order to have $SC(TR) = 0$ for $TR$ bigger than 1 and satisfying the 1st and 4th constraint |
| $\gamma = 50$   | in order to consider $SKPI$ if is bigger than 50 and satisfying the 2nd and 3rd constraint     |
| $\beta = \log \left( \left( \frac{4}{5} \right)^{-\frac{4}{3}} \right)$ | in order to have $SC(1) = 80$ and satisfying the 5th constraint                                |
| $\alpha = 0.25$   | in order to have $SC(TR) = 100$ for $TR$ less than 0.25 and satisfying the 6th constraint      |

Com podeu observar depèn de  $\alpha, \beta, \gamma, \lambda$ , això va ser una cosa que els hi va encantar ja que permetia generalitzar l'*approach* a futurs problemes. No adjunto més detall del Audit Case, però va ser gratificant que tot el grup que treballava en ell van estar super agraïts amb el meu treball i impresionats (ja que em van avisar un dia a les 11 si ho podia fer per aquell mateix dia a les 14, i no tenia ni context del tema, però ho vaig poder fer). Ells van poder presentar el projecte i durant els següents dies vaig escriure documentació sobre el tema amb exemples, gràfiques i casos reals per tal que futurs treballadors no matemàtics puguin entendre com es crea la funció, per què es així i com funciona. En aquest cas també la vaig implementar a *Python* però una versió "cutre" ja que no tenia context del *pipeline* del cas, i va ser amb l'objectiu de facilitar la feina a qui treballava en el cas.

### 5.5.2 Score Recommendation Rank

Semblant al cas anterior, un altre grup de treball necessitaven ajuda amb un problema que era: **Recomenació per nous Scores**, és a dir, com podien canviar els Scores en funció del valor actual i de lo lluny o prop que estaven del *Best Value*. La idea era crear una funció que donats:

- |                         |                                 |                      |
|-------------------------|---------------------------------|----------------------|
| 1. <i>Current Value</i> | 2. <i>New recommended value</i> | 3. <i>Best value</i> |
|-------------------------|---------------------------------|----------------------|

et donés un Score entre 0 y 100 per dir quina probabilitat del nou score. Això s'havia de basar en la distància que tenia del *score* actual (contra més lluny, pitjor) i de la distància al *Best Value* (contra més aprop, pitjor, ja que és "més difícil" arribar). Amb les indicacions vaig desenvolupar la següent funció:

$$f(current, new, best) = e^{-\alpha abs(new - current)} e^{-\beta abs(best - new)}$$

where  $\alpha, \beta > 0$ .

On vaig deixar-ho en funció d' $\alpha$  i  $\beta$  per que ells mateixos, un cop ho vaig implementar a *Python* ho poguessin experimentar i escollir de la millor manera possible.

Com a conclusió general diria que totes les tasques han estat molt valorades, sobre tot la *Validation Pipeline* i el *Bot*, ja que és on més he treballat (ha representat 2/3 de les meves pràctiques encara que pot ser no es reflexa a l'informe ja que hi ha hagut **molt** treball de programació, configuració de funcions, moltes PR (*Pull Requests*) però finalment, just un parell de dies abans d'acabar les pràctiques, la vaig poder finalitzar i estic molt orgullós.

## 5.6 Eines Utilitzades

Encara que la majoria les he anat mencionant durant l'explicació de les tasques, m'agradaria fer un resum de les eines que he utilitzat durant la meua estada de pràctiques, ja que han sigut una cosa important en aquesta (aprendre a utilitzar-les i adaptar-me).

1. *PyCharm*; tota la programació en *Python*.
2. *Confluence* (*Altassian*); On es troba tota la documentació i on he afegit tota la meua nova documentació.
3. *Jira Board* (*Altassian*); Diferents taules on es troben tant les tasques a realitzar amb totes les descripcions necessàries com taules per donar queixes o suggerències. També per demanar coses a altres equips, així com permisos o suport.
4. *DBaver*; Per fer totes les *query* de *SQL*.
5. *GitHub*; Per almacenar tots els codis de *Python* i *SQL*.
6. *Slack*; Comuniació entre equips i companys.
7. *Power Point*; per fer presentacions a clients i per les *meetings* entre equips.
8. *Teams* & *Zoom*; Per fer les *meetings*.
9. *AWS* (S3); Almacenatge de tot tipus d'informació confidencial, taules i arxius.

## 6 Condicions de feina

Les condicions de feina han sigut excel·lents durant aquesta estada de pràctiques curriculars.

L'horari és totalment flexible, permetent a cadascú establir el seu propi horari, amb l'únic requisit d'assistir a les reunions diàries en línia. Personalment, vaig establir un horari de 9 a 13 i de 14 a 15 de dilluns a divendres, però molts dies començava a les 8:30 i marxava a les 13:30.

Aquesta flexibilitat m'ha permès combinar perfectament la vida laboral amb la universitària.

A nivell de remuneració, rebo una compensació de 14€ per hora (bruts), que són 1400€ mensuals, amb un descompte del 2% d'IRPF. A més, s'han compromès a cobrir el 100% de la matrícula d'aquest quart curs de carrera (470€)

Quant a materials i instal·lacions, el primer dia em van donar un MacBook Pro, a més de tots els complements necessaris per a la compatibilitat amb altres dispositius i una samarreta de l'empresa. La oficina disposa de llocs amb pantalles grans per a cada persona, tres sales de reunions, i una sala de menjar amb cafè, te, vaixel·la, microones i dues neveres amb aigua i altres begudes per a tots. Quan venien clients, compraven diferents tipus de fruita i pastisseria que erem lliures d'agafar.

A nivell interpersonal, s'fomenta un ambient inclusiu i agradable. El lema de l'empresa és "*Fun, Agility, Innovation*", motivant la col·laboració i igualtat entre tots els membres dels equips, des d'*Interns* fins a *Seniors*. També s'organitzen activitats regulars com dinars en grup, sortides després del treball i festes *post office* amb jocs de taula, consoles i karaoke, així com classes en línia de gimnàstica.

La oficina també té zones amb sofàs, TV, altaveus i pufs on es pot treballar o descansar durant el dia, i una sala tranquil·la per a reunions individuals o estudi.

En resum, l'experiència a nivell de condicions ha sigut immillorable, en un ambient de respecte, multicultural (amb personal de diversos països i amb anglès com a llengua principal) i de suport mutu entre companys.

## 7 Impresions personals

Com a impressió general i com a resum de la meva estada de pràctiques a *Aily Labs*, diria que ha sigut una experiència molt bona per a mi, molt enriquidora i en la qual he pogut introduir-me al món laboral en una empresa jove, amb treballadors joves. Això m'ha donat l'oportunitat de sentir-me valorat, de sentir que el meu treball contribuïa a un projecte real i que era útil i necessari. Això segurament no hauria sigut possible en moltes altres empreses on ja està tot fet i l'únic objectiu és realitzar tasques repetitives i monòtones.

D'altra banda, l'oportunitat de treballar amb un equip de *Data Science* en un grup multidisciplinar format per un físicomatemàtic, un enginyer de *Software*, un economista i un químic, m'ha fet veure dues coses molt útils per a la meva visió del món laboral: en primer lloc, la importància de tenir diferents punts de vista quan s'ha d'afrontar un projecte, i la realitat que, encara que molts matemàtics pensin que sí, no tot és matemàtic i, encara que tots els problemes es poden pensar de forma matemàtica, al món real no tothom ho entendreà i s'ha de tenir la capacitat de transmetre els problemes que poden tenir una solució matemàtica de manera que sigui comprensible per a tothom. D'altra banda, m'ha fet veure que, havent estudiat matemàtiques, puc arribar a tenir un bon lloc de treball i fins i tot arribar a ser Senior o *Head of AI* (qui també és matemàtic i de fet va estudiar a la UAB, encara que va fer tres carreres). Tot això m'ha aportat uns coneixements personals impossibles de valorar.

Relacionat amb el primer punt, m'ha donat l'oportunitat de poder aplicar coneixements matemàtics a problemes reals i veure que, encara que haguessis d'adaptar la resposta per la visió empresarial, podies aplicar un objecte matemàtic a un problema útil, i això crec que és molt important després de 4 anys on, llevat d'una assignatura (taller de modelització) no havíem vist problemes de la vida laboral.

Per destacar una impressió negativa (puntual), diria que m'hauria agradat fer més coses relacionades amb *AI* o *ML*. He treballat en entorns on hi havia models de *ML* o *deep learning*, però no he pogut treballar directament amb ells. No obstant, soc conscient que això és probablement per un tema de temps, ja que els models han de sortir a producció i en 3 mesos l'empresa no pot arriscar-se a que algú que no coneix bé el producte desenvolupi un model que surti a producció.

A més, per acabar, encara que potser fa 3 mesos no opinava el mateix, crec que per primer cop valoro el fet d'haver estudiat matemàtiques. Tot i que, com bé he dit, és un món on tothom té lloc i no cal estudiar matemàtiques per dedicar-te a ell, som molt valorats.

És per tot això que he decidit que a partir de juliol començaré la meva segona etapa com a *Junior DS 1* a *Aily Labs* a jornada completa i en el mateix equip de treball, per continuar vivint aquesta experiència amb els meus companys actuals i continuar creixent com a professional.



## 8 Annexos

### 8.1 Exemple codi

A continuació adjunto un exemple d'un codi realitzat a l'empresa. És un exemple molt senzill i amb estructures molt sencilles però amb el qual es poden veure els 3 punts comentats a la Secció 5.2.2.

```
1 def kpi_improve_or_get_worse(value: float, previous_value: float, kpi: str, kpi_improved:
  int,
2                                     kpi_got_worse: int) -> tuple[int, int]:
3     """
4     Return the values of kpi_improved and kpi_got_worse uploaded. If the kpi improve,
5     then kpi_improved improve in 1
6     point. On the other hand, if the kpi wet worse, then kpi_got_worse improve in 1 point
7     .
8     Parameters:
9     - value (float): current value of the KPI
10    - previous_value (float): previous value of the KPI
11    - kpi (str): kpi we want to rescale
12    - kpi_improved (int): count of kpi improved so far on this site.
13    - kpi_got_worse (int): count of kpi
14
15    Returns:
16    - tuple[int, int]:
17      int: uploaded count of kpi improved on this site.
18      int: uploaded count of kpi got worse on this site.
19    """
20    rename_kpi = next(key for key, value in kpi_map_rename.items() if value == kpi)
21    info_kpi_values = kpi_info.get(rename_kpi)
22    kpi_objective = info_kpi_values.get("objective")
23    best_value = info_kpi_values.get("best")
24    threshold_value = info_kpi_values.get("threshold")
25    # Firstly, in the asymmetrical kpi(s), we mirror all of them to right side
26    asymmetrical = info_kpi_values.get("asymmetrical")
27
28
29    if kpi_objective == "pyramidal" and asymmetrical:
30        threshold_value = mirror_value(threshold_value, rename_kpi, asymmetrical)
31        if (value < best_value and previous_value > best_value):
32            value = mirror_value(value, rename_kpi, True)
33        elif (value > best_value and previous_value < best_value):
34            previous_value = mirror_value(previous_value, rename_kpi, True)
35
36    value_moved = move_value_to_variable_universe(value, best_value, threshold_value,
37    kpi_objective)
38    previous_value_moved = move_value_to_variable_universe(previous_value, best_value,
39    threshold_value, kpi_objective)
40
41    if kpi_objective == "max":
42        kpi_improved += (value_moved > previous_value_moved)
43        kpi_got_worse += (value_moved < previous_value_moved)
44    elif kpi_objective == "min":
45        kpi_improved += (value_moved < previous_value_moved)
46        kpi_got_worse += (value_moved > previous_value_moved)
47    elif kpi_objective == "pyramidal":
48        dist_value_to_best_value = abs(value_moved - best_value)
49        dist_previous_value_to_best_value = abs(previous_value_moved - best_value)
50        kpi_improved += (dist_value_to_best_value < dist_previous_value_to_best_value)
51        kpi_got_worse += (dist_value_to_best_value > dist_previous_value_to_best_value)
52    else:
53        logger.error(f"Invalid KPI objective: '{kpi_objective}'. This objective has not
54        been studied")
55
56    return kpi_improved, kpi_got_worse
```

## 8.2 Exemple Matriu de Confusió Multiclasse



Figura 1: Exemple de la matriu de confusió

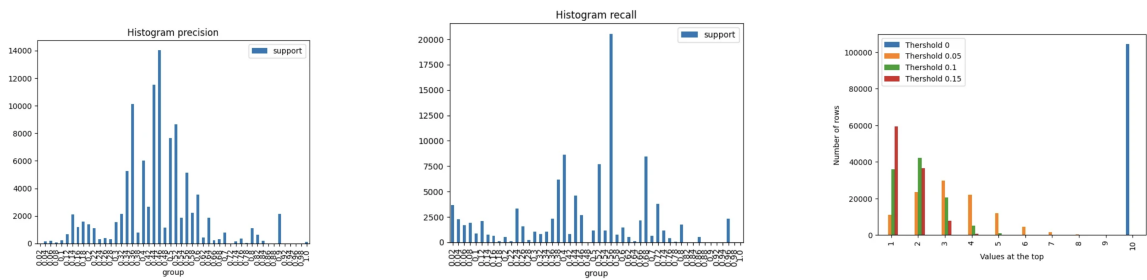


Figura 2: Exemple dels 3 histogrames de les mètriques: A l'esquerra de *precision*, al centre de *recall* i a la dreta la freqüència amb el *thresholds*

## 8.3 Example of some graphics of EDA of Score

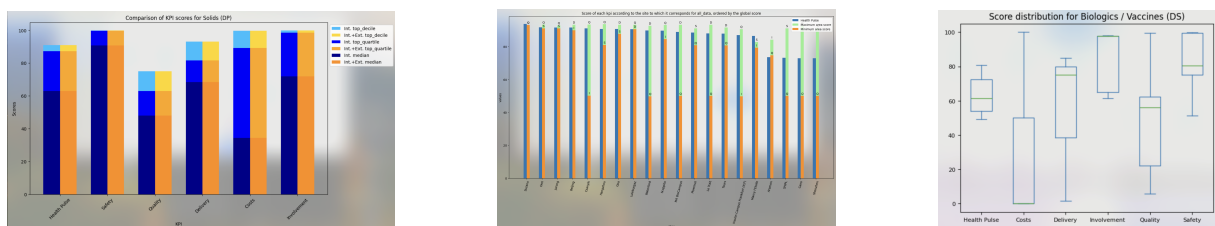


Figura 3: Exemple dels 3 gràfics dels Score a cada site: A l'esquerra la comparació dels *internal* vs *external* benchmarks pels sites d'arquetip *Solids*, al centre la gràfica del *Scores* per totes les dades i a la dreta el *box-plot* per cada area del arquetip de *Biologic / Vaccines*.

## 8.4 Gràfics realitzats per explicar la funció del *Non-Pyramidal KPI*

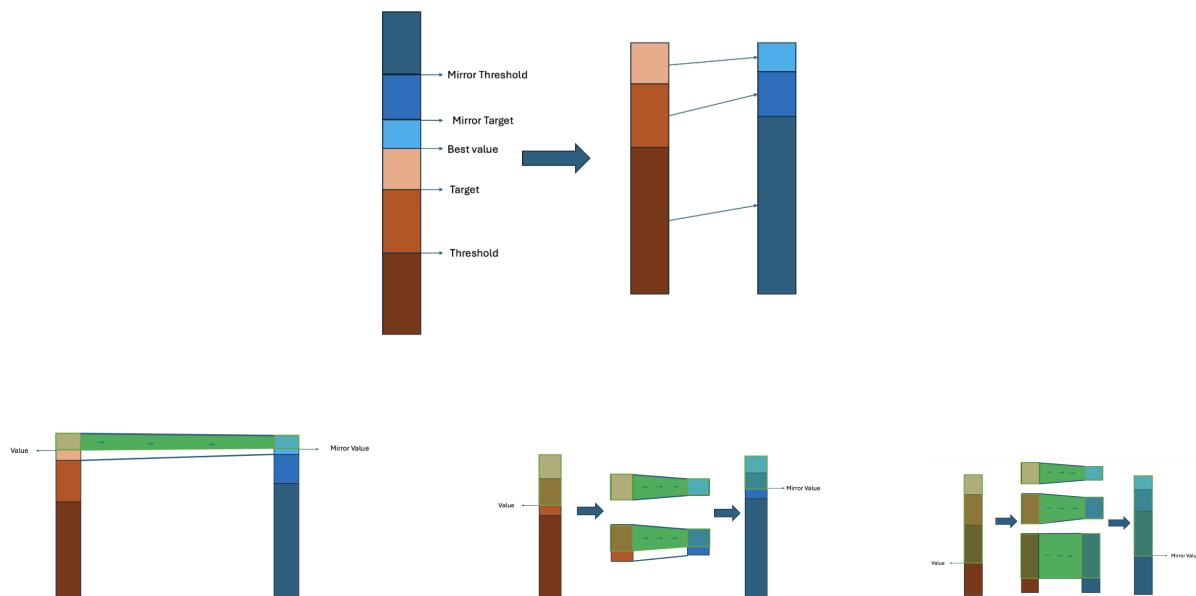


Figura 4: Gràfics explicatius de la funció que passa els Scores d'un costat a un altre del *best value* quan el KPI es piramidal i asimètric

## 8.5 Exemple d'un *Jump* i una *Insensitivity*

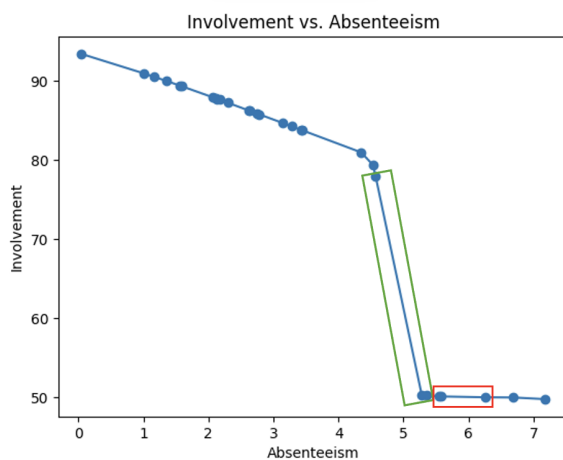
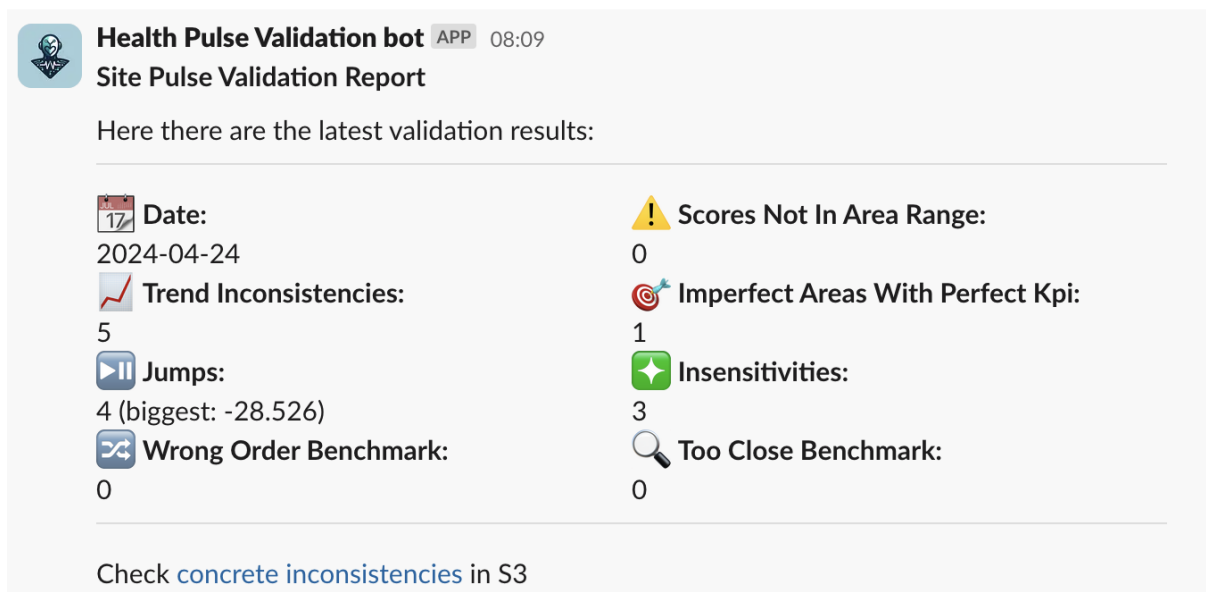


Figura 5: En verd, un *jump*, en vermell, una *insensitivity*









## 8.6 Exemple del Validation Bot



**Health Pulse Validation bot** APP 08:09

**Site Pulse Validation Report**

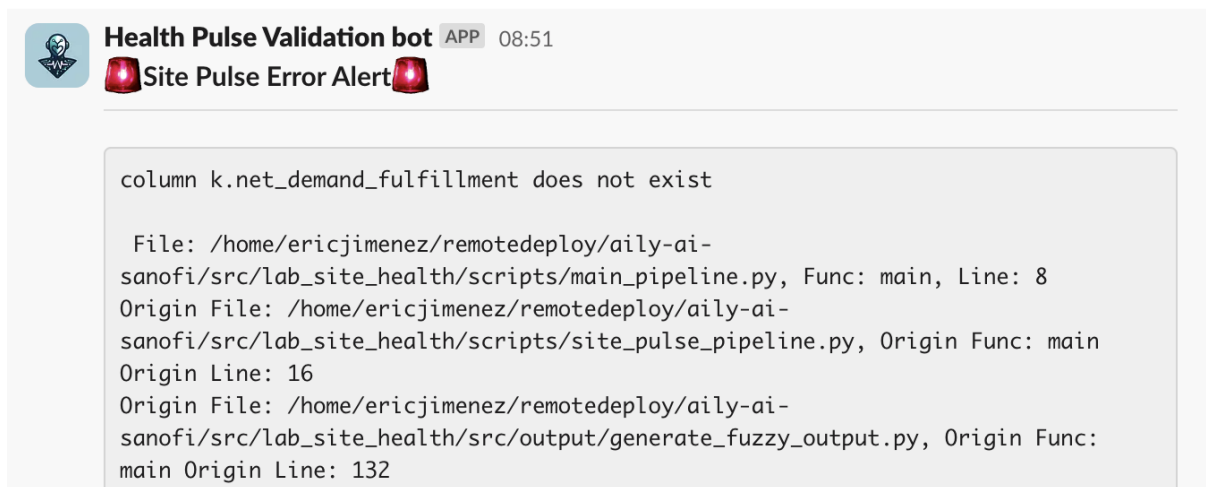
Here there are the latest validation results:

 <b>Date:</b> 2024-04-24	 <b>Scores Not In Area Range:</b> 0
 <b>Trend Inconsistencies:</b> 5	 <b>Imperfect Areas With Perfect Kpi:</b> 1
 <b>Jumps:</b> 4 (biggest: -28.526)	 <b>Insensitivities:</b> 3
 <b>Wrong Order Benchmark:</b> 0	 <b>Too Close Benchmark:</b> 0



Check [concrete inconsistencies](#) in S3

Figura 6: Exemple del Missatge rebut a diari sobre les Validacions (sempre que no falli)

## 8.7 Exemple del Slack Alert Bot



**Health Pulse Validation bot** APP 08:51

 **Site Pulse Error Alert** 

```
column k.net_demand_fulfillment does not exist

File: /home/ericjimenez/remotedeploy/aily-ai-
sanofi/src/lab_site_health/scripts/main_pipeline.py, Func: main, Line: 8
Origin File: /home/ericjimenez/remotedeploy/aily-ai-
sanofi/src/lab_site_health/scripts/site_pulse_pipeline.py, Origin Func: main
Origin Line: 16
Origin File: /home/ericjimenez/remotedeploy/aily-ai-
sanofi/src/lab_site_health/src/output/generate_fuzzy_output.py, Origin Func:
main Origin Line: 132
```

Figura 7: Exemple de la Alerta d'Error rebut a diari sobre les la *Lab Site Health Pipeline* quan no falli.

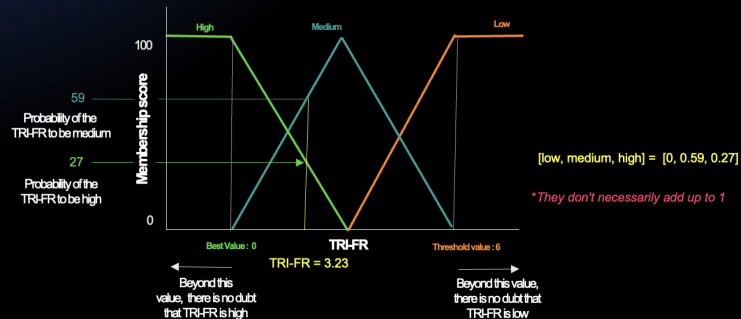
## 8.8 Fuzzy Logic Real Example

### Fuzzification: creation of membership functions

Assessment of the TRI-FR

#### Create Memberships

Define the associated thresholds for each KPI and create the membership functions



### Probability for each scenario

	TRI-FR	PSI	SIF		
(3 H):	27% H	+ 31% H	+ 100% H	We need the 3 things to happen minimum %	27% Superhigh
(2 H + 1 M):	27% H	+ 50% M	+ 100% H		27% High
(2 H + 1 M):	59% M	+ 31% H	+ 100% H		31% High
(1 H + 2 M):	59% M	+ 50% M	+ 100% H		50% High
				We need at least one scenario to happen. maximum %	50%

We have only 2 different final scenarios (**Superhigh** and **High**)

Following the Natural Language rules, we do not obtain any **Middle-High** (or worst) scenario.

### Leveraging Fuzzy Logic, domain knowledge and external benchmarks

Select the centroid in the selected area

#### Apply the Fuzzy Logic

Transform the initial KPIs into their respective scores using fuzzy logic

\*Let's assume that the blue area has weight, and that it weighs at each vertical line on the 'Score value' axis as much as the value itself (from 60 to 100 in this case). The **centroid** is the point where, if we placed this area on the tip of a pencil, it would balance without moving.\*

