

SURF, SORBETS ET STATISTIQUES

UNE INTRODUCTION AU MACHINE LEARNING
DANS LES AFFAIRES



CETTE BD, QUI SE VEUT UN OUTIL PÉDAGOGIQUE, VISE À DÉMYSTIFIER L'APPLICATION DU **MACHINE LEARNING** * DANS LE MONDE DES AFFAIRES POUR LES ÉTUDIANTS EN ÉCOLE DE COMMERCE.

NOUS TENONS À ADRESSER NOS PLUS SINCÈRES REMERCIEMENTS À **KEDGE BUSINESS SCHOOL** POUR SON SOUTIEN PRÉCIEUX, À TRAVERS LE FONDS D'INNOVATION PÉDAGOGIQUE, RENDANT POSSIBLE LA CRÉATION DE CETTE BANDES DESSINÉE.

UN GRAND MERCI ÉGALEMENT À **HEC LAUSANNE** ET À **KEDGE BUSINESS SCHOOL**, QUI NOUS ONT PERMIS DE CONCRÉTISER CE PROJET ENRICHISSANT.

AVERTISSEMENT : si vous trouvez que nos trois personnages principaux - un "SAVANT FOU" un peu âgé, un oncle italien fabricant de glaces et vendeur ambulant, et un neveu surfeur quelque peu désinvolte - ne brillent pas par leur diversité, c'est qu'ils sont inspirés de nous, les trois auteurs. Ce manque d'inclusivité n'est pas un choix délibéré mais le reflet de nos propres personnalités !

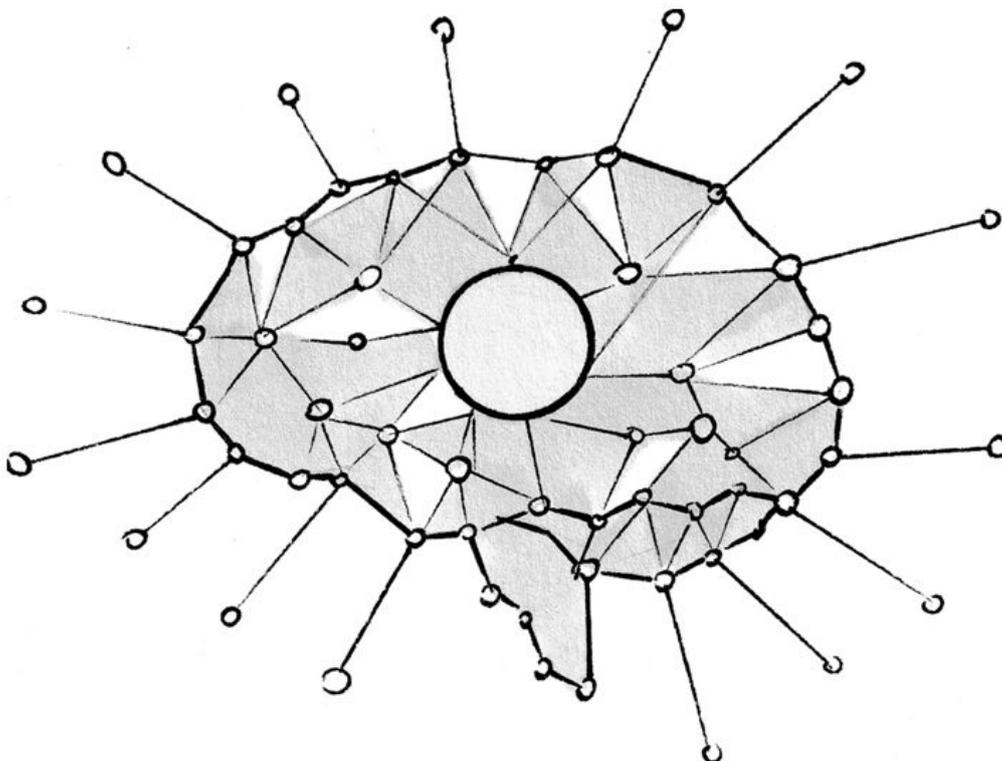
NOUS ESPÉRONS QUE CETTE BD VOUS OFFRIRA À LA FOIS DES CONNAISSANCES ET DES SOUÏRES.

LES AUTEURS :

OLIVIER DUPOÛET - KEDGE BUSINESS SCHOOL, 680 COURS DE LA LIBÉRATION, 33405 TALENCE, FRANCE.

GAUTIER STAUFFER - FACULTY OF BUSINESS AND ECONOMICS (HEC LAUSANNE), UNIVERSITY OF LAUSANNE, 1015 LAUSANNE, SWITZERLAND.

SÉBASTIEN TESSIER, DIT DAMOUR - AUTEUR DE BANDES DESSINÉES, BORDEAUX, FRANCE.



* CHAMP D'ÉTUDE DE L'INTELLIGENCE ARTIFICIELLE QUI SE FONDE SUR DES APPROCHES MATHÉMATIQUES ET STATISTIQUES POUR DONNER AUX ORDINATEURS LA CAPACITÉ À APPRENDRE À PARTIR DE DONNÉES.

BORDEAUX, GARE SAINT-JEAN.

DÉBUT DE L'ÉTÉ.

SALUT
TONKON!

SALUT
SIMON!



CONTENT
DE VENIR
PASSER CES
DEUX MOIS D'ÉTÉ
À BORDEAUX!

ET SUPER CONTENT
DE T'ACCOMPAGNER
SUR LES PLAGES DU
LITTORAL AVEC TON
CAMION DE GLACES!
REGARDE!!!

!!! J'AI
PRIS MA
FLANCHE!



TU VAS ÊTRE DÉÇU
SIMON MAIS JE NE VAIS
PLUS À L'OcéAN DEPUIS
QUELQUES ANNÉES.



JE "NAVIGUE" ENTRE LES DIFFÉ-
RENTES PLAGES DU CENTRE-VILLE,
TOUT L'ÉTÉ DÉSORMAIS.



AH MINCE!!! MAMAN
M'AVAIT DIT QU'ETU
PROFITAIS DE LA FRÉ-
QUENCE DES TOURISTES
L'ÉTÉ!!!

C'ÉTAIT
VRAI MAIS CES
DERNIÈRES
ANNÉES!!!

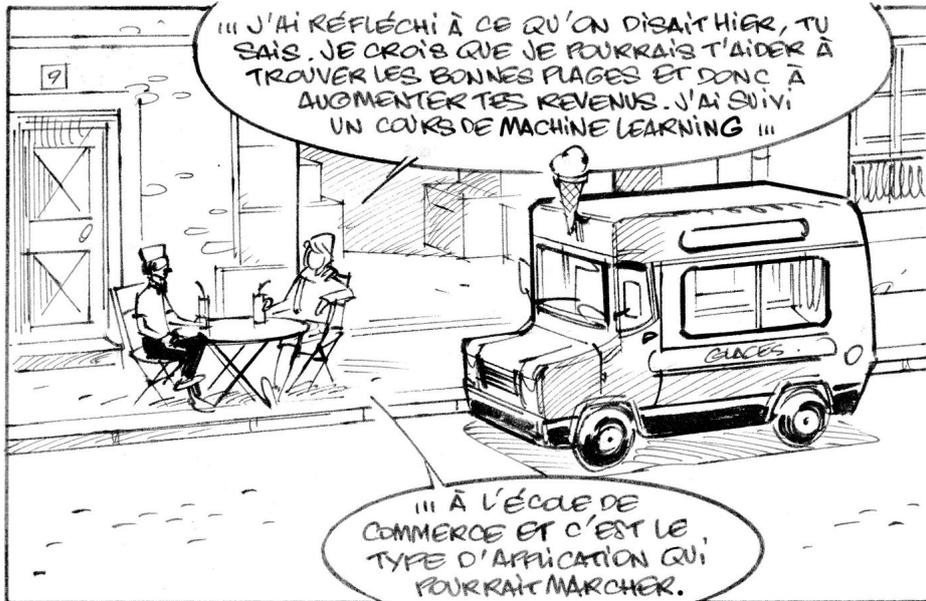
J'AI EU DES DIFFI-
CULTÉS À IDENTIFIER
LES BONNES PLAGES. TU
SAIS, UNE "BONNE PLACE"
DÉPEND PAS MAL DE LA
MÉTÉO ET JE CROIS
!!!

!!! QUE JEN'ÉTAIS PAS DOUÉ POUR
TROUVER LE BON SPOT. FINALEMENT,
JE ME SUIS RENDU COMPTE QUE J'AVAIS
DES REVENUS PLUS STABLES, ICI À
BORDEAUX. MAIS BON TU VERRAS,
BORDEAUX EST UNE SUPERBE
VILLE!



CHAPITRE 1 : ARBRES DE DÉCISION

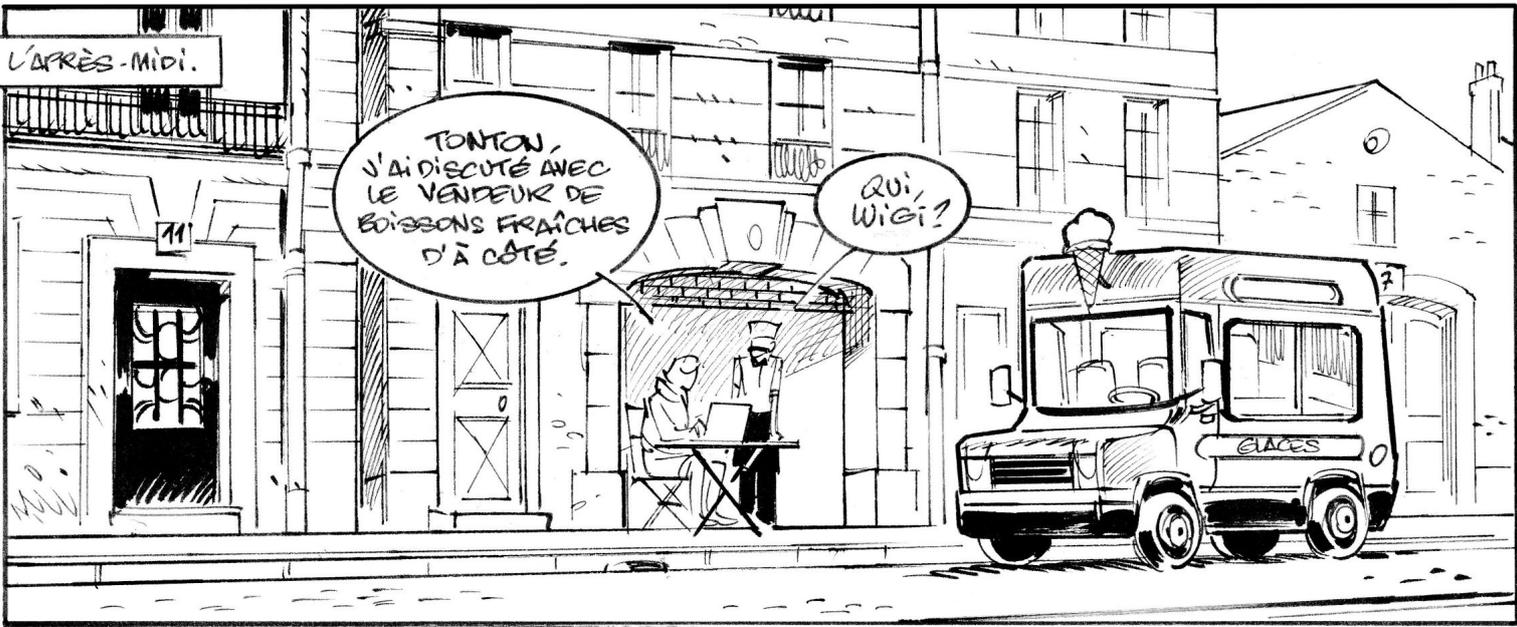
LE LENDEMAIN, AUX CHARTRONS, CHEZ "TONTON" GIOVANNI.



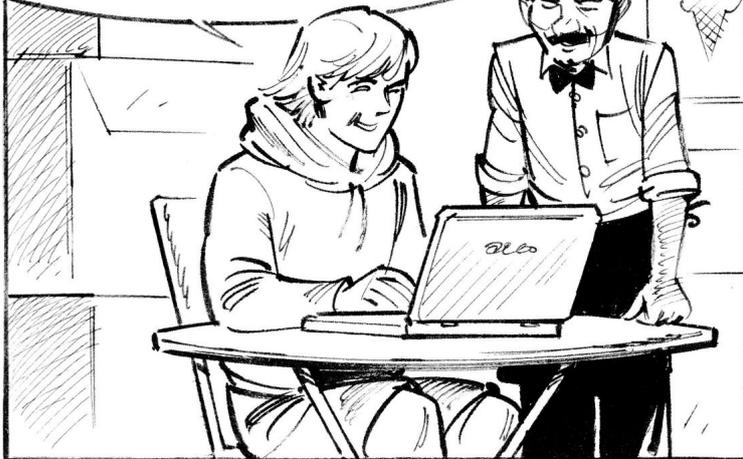
L'APRÈS-MIDI.

TONTON, J'AI DISCUTÉ AVEC LE VENDEUR DE BOISSONS FRAÎCHES D'À CÔTÉ.

Qui, Wigi?



QUI ET IL M'A DIT QU'IL AVAIT UNE ÉQUIPE DE TROIS EMPLOYÉS QUI PARCOURAIENT LE LITTORAL AVEC DES CAMIONS AMBULANTS. IL NE PEUT PAS PARTAGER SES CHIFFRES: "C'EST TROP CONFIDENTIELLE MON PITI!"



MAIS IL VEUT BIEN ME DONNER QUELQUES DONNÉES SUR SES PERFORMANCES, BONNES OU PAS. EN FONCTION PAR EXEMPLE, DE LA MÉTÉO OBSERVÉE (SOLEIL, NUAGE, PLUIE), DE LA TEMPÉRATURE, DU TYPE



DE PLAGE (URBAINE OU NATURE), DU JOUR DE LA SEMAINE (WEEK-END OU NON) ET DE LA PROXIMITÉ AVEC BORDAUX. TU CROIS QU'ON PEUT UTILISER ÇA COMME PREMIER INDICATEUR?

BEN ÉCOUTE, WIGI ET MOI, ON A SOUVENT LE SOURIRE EN MÊME TEMPS, ALORS J'IMAGINE

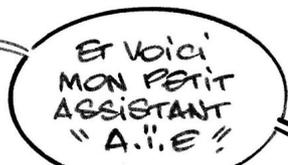


QUE NOS PERFORMANCES SONT ASSEZ EN PHASE!

SUPER, ALORS JE VAIS UTILISER SES DONNÉES POUR ESSAYER DE DÉTERMINER QUELLES SONT LES BONNES PLACES POUR TES GLACES! EN ESPÉRANT QU'ELLES SOIENT AUSSI BONNES POUR LE SURF!

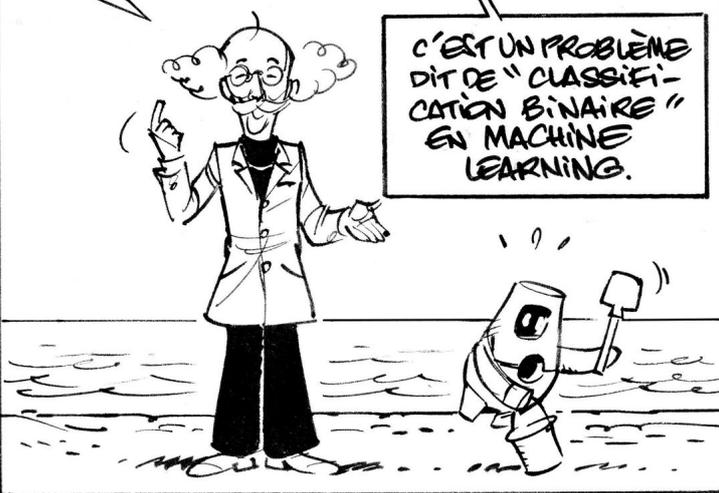


HA HA! FORMIDABLE!

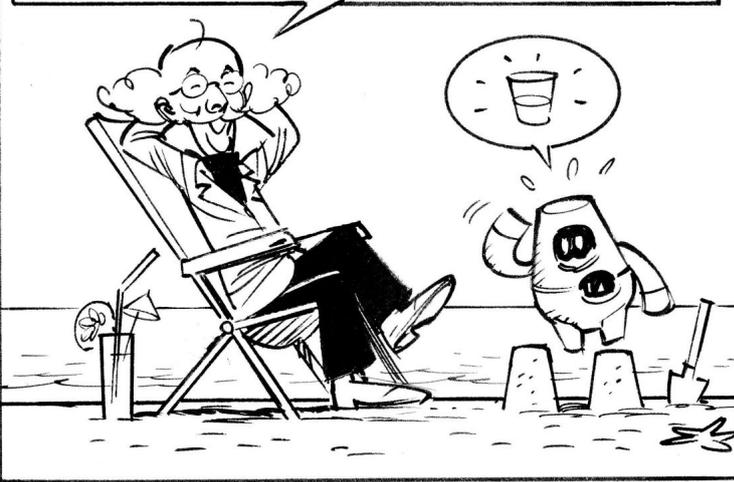


TEMP	METEO	WE	BEACH_CAT	BEACH_CAT_IND	METRO	BON_CA
26	SOLEIL	1	NATURE	0.4	PROCHE	TRUE
27	SOLEIL	0	NATURE	0.4	PROCHE	TRUE
25	SOLEIL	0	NATURE	0.5	LOIN	TRUE
25	NUAGE	1	URBAIN	0.9	LOIN	TRUE
26	NUAGE	1	NATURE	0.5	LOIN	TRUE
28	NUAGE	0	NATURE	0.5	PROCHE	TRUE
27	PLUIE	0	URBAIN	0.7	PROCHE	FALSE
28	PLUIE	0	NATURE	0.6	PROCHE	FALSE
21	SOLEIL	1	NATURE	0.2	PROCHE	TRUE
23	SOLEIL	0	URBAIN	1.0	PROCHE	FALSE
23	SOLEIL	0	NATURE	0.3	PROCHE	TRUE
21	NUAGE	1	NATURE	0.4	PROCHE	TRUE
22	PLUIE	1	URBAIN	0.8	LOIN	FALSE
23	PLUIE	1	NATURE	0.2	PROCHE	FALSE
21	PLUIE	0	NATURE	0.1	LOIN	FALSE
17	SOLEIL	1	URBAIN	0.8	LOIN	TRUE
18	GVN FI	1	NATURE	0.5	PROCHE	TRUE

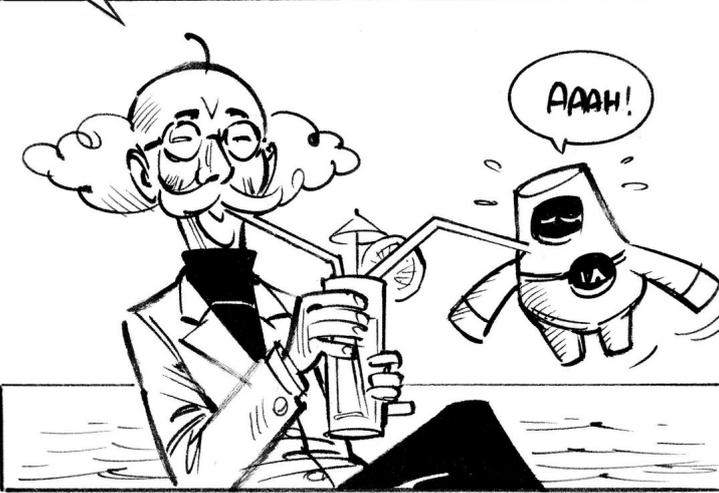
LE JEUNE SIMON ESPÈRE POUVOIR PRÉVOIR LA COLONNE BON-CA (BON CHIFFRE D'AFFAIRE) EN FONCTION DES AUTRES COLONNES.



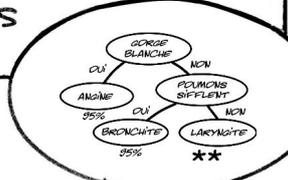
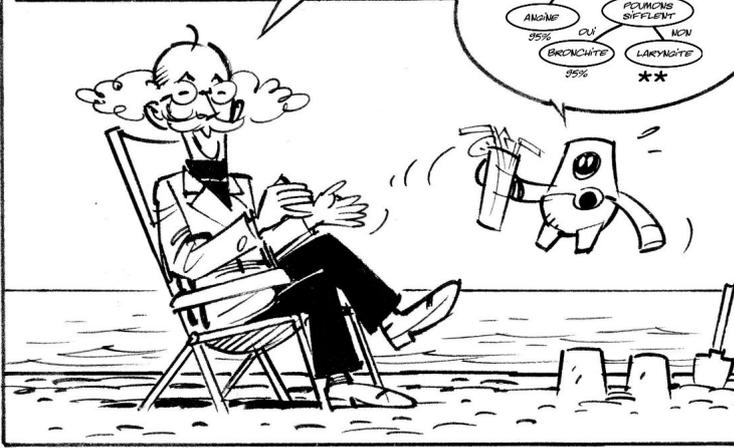
ON SOUHAITE PRÉDIRE UNE VARIABLE BINAIRE (VRAI/FAUX), IL EXISTE UN TYPE DE MODÈLE SIMPLE QUI PERMET CELA : LES ARBRES DE CLASSIFICATION.



LES MÉDECINS UTILISENT CES ARBRES POUR POSER UN DIAGNOSTIC (VRAISEMBLABLE), POUR DÉTERMINER SI VOUS AVEZ UNE ANGINE, PAR EXEMPLE.



IL S'AGIT DONC DE POSER ITÉRATIVEMENT DES "QUESTIONS" BINAIRES (LA RÉPONSE EST OUI OU NON), POUR REGROUPEZ LES DONNÉES DANS DES GROUPE DE PLUS EN PLUS HOMOGÈNES.



* CECI N'EST QU'UN EXTRAIT DES DONNÉES. METEO = SOLEIL, PLUIE OU NUAGE - TEMP = TEMPÉRATURE EN °C - WE = 1 SI SAMEDI OU DIMANCHE, 0 SINON - METRO = PROXIMITÉ DE BORDEAUX (PROCHE OU LOIN) - BEACH-CAT = CATÉGORIE DE PLAGE (NATURE OU URBAINE) - BEACH-CAT-IND = INDICE D'URBANISME (0 = 100% NATURE 1 = 100% URBAINE)

** CET ARBRE N'EST QU'UNE ILLUSTRATION, L'ARBRE DE DÉCISION "RÉEL" EST BIEN PLUS COMPLEXE.

DANS CET EXEMPLE, ON CLASSIFIERAIT COMME BON CHIFFRE D'AFFAIRE TOUTE DONNÉE AVEC WEATHER-IND > 0,17 (ON SE TROMPERAIT ALORS UNE FOIS SUR QUATRE) ET DE MAUVAIS CA TOUTE DONNÉE AVEC WEATHER-IND < 0,17 (ET ON SE TROMPERAIT DANS 11% DES CAS).



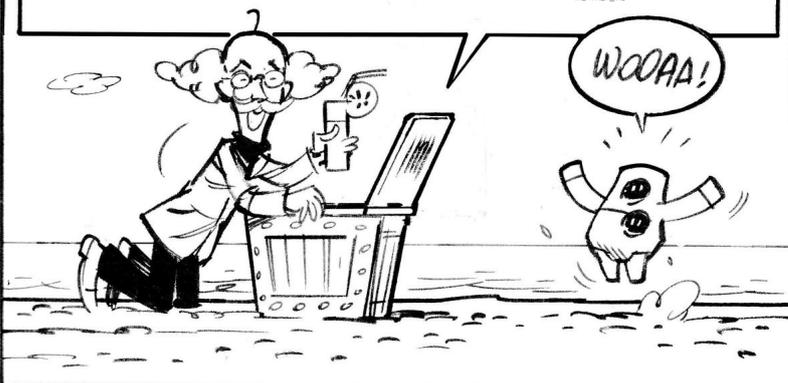
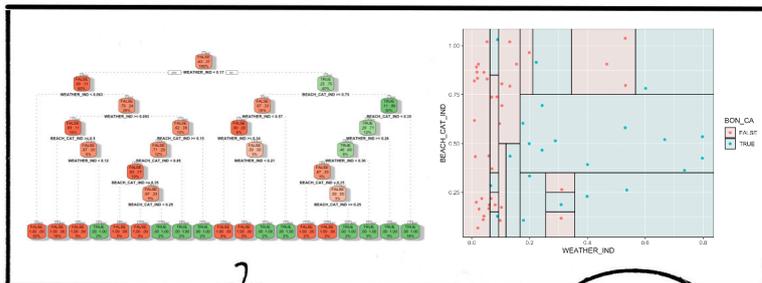
ON POURRAIT DÉVELOPPER DES ARBRES PLUS COMPLEXES EN POSANT D'AUTRES QUESTIONS BINAIRES.



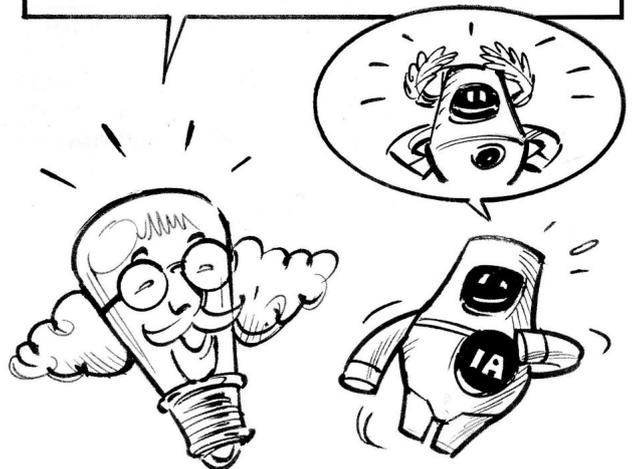
POUR SAVOIR SI ON PROGRESSE (ET DONC SI LES SOUS-GROUPES SONT PLUS HOMOGÈNES), IL FAUT INTRODUIRE UNE MESURE D'HOMOGÉNÉITÉ : ON PEUT UTILISER LA VARIANCE*.



SI ON ITÈRE CES QUESTIONS TANT QUE LES SOUS-GROUPES NE SONT PAS PARFAITEMENT HOMOGÈNES, VOILÀ CE QU'ON OBTIENT (DANS LA CASE 5)...



UNE CLASSIFICATION PARFAITE DES DONNÉES HISTORIQUES A.I.E. ! VOYONS CE QUE CELA DONNE EN PRATIQUE ET CE QUE SIMON EN FERA !!!



* [HTTPS://FR.WIKIPEDIA.ORG/WIKI/VARIANCE_\(MATHÉMATIQUES\)](https://fr.wikipedia.org/wiki/Variance_(mathématiques))

MUNI DE L'ARBRE PRÉCÉDENT, SIMON REVIENT VOIR SON ONCLE.

TONTON ! J'AI DÉVELOPPÉ UN OUTIL POUR DÉTERMINER SI UNE PLAGE EST BONNE POUR TES AFFAIRES !

ON PEUT L'ESSAYER SI TU VEUX !



AH OUI ! COMMENT ÇA MARCHE ?

C'EST SIMPLE : JE REGARDE L'INDICE MÉTÉO DU JOUR SUR "METEOINDICE.FR", JE CHERCHE L'INDICE D'URBANISATION DE LA PLAGE SUR "PLAGEINDICE.FR" ET JE CONSULTE MON ARBRE DE CLASSIFICATION.



DEMAIN J'AI UN INDICE MÉTÉO DE 0,8 ET UN INDICE D'URBANISATION DE 0,9 POUR LA PLAGE DU GRAND CROÛT, SI JE REGARDE MON ARBRE, JE VOIS QUE ÇA VA ÊTRE TOP FEUR TON CA !

ON Y VA ?

BANCO BAMBINO ! TU AS L'AIR DE SAVOIR CE QUE TU FAIS, ANDIAMO !



AU BORD DE L'OcéAN.



BON ALLEZ, ON
REMBALLE!



PAS FRANCHEMENT
CONQUANT CE PREMIER
ESSAI, SIMON!



DÉSOLÉ TONTON, JE NE COMPRENDS PAS,
L'ARBRE DE CLASSIFICATION AVAIT POUR-
TANT L'AIR TRÈS BON!



JE VAIS RECON-
SULTER MES NOTES DE
COURS POUR VOIR D'OÙ
VIENT LE PROBLÈME



LE LENDEMAIN AU PETIT DÉJEUNER.

TU N'AS PAS L'AIR D'AVOIR BEAUCOUP DORMI...

ÇA VA !!!



!!! J'AI COMPRIS D'OÙ VENAIT LE PROBLÈME. J'AI OUBLIÉ QU'EN MATIÈRE DE MACHINE LEARNING, UN DES GROS RISQUES EST LE SUR-APPRENTISSAGE.



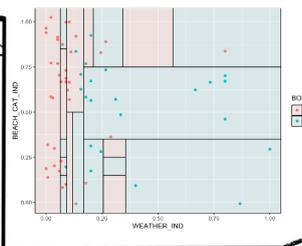
L'IDÉE EST SIMPLE : SI ON ESSAYE DE TROP BIEN REPRODUIRE DES DONNÉES HISTORIQUES* AVEC UN MODÈLE, ON RISQUE DE FAIRE BEAUCOUP D'ERREURS SUR DES DONNÉES JAMAIS VUES JUSQUE LÀ.



JE SUIS RETOURNÉ VOIR WIGI ET IL M'A FOURNI DES DONNÉES COMPLÉMENTAIRES COLLECTÉES LES ANNÉES PRÉCÉDENTES. LE RÉSULTAT EST FANTASTIQUE : SI JE REGARDE LES POINTS SUR MA CLASSIFICATION "IDÉALE", JE VOIS QUE JE ME TROMPE SOUVENT, NOTAMMENT DANS LES ZONES À GAUCHE. REGARDE !!!



AH OUI, EN EFFET! TU PRÉDIS QU'ELLES SERONT VERTES ET ELLES SONT SOUVENT ROUGES**



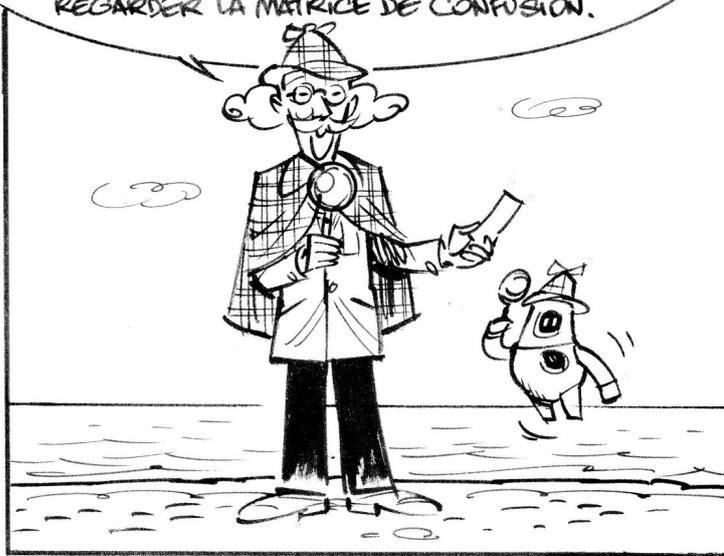
OUI EN BIEN, LÀ "SCRENTCH"!!! ON A AFFAIRE À UN EXEMPLE DE SUR-APPRENTISSAGE : MON MODÈLE A TENDANCE À REPRODUIRE PARFAITEMENT MES DONNÉES D'APPRENTISSAGE MAIS IL A DES PERFORMANCES TRÈS MÉDIOCRES SUR D'AUTRES DONNÉES, DITES DE TEST.



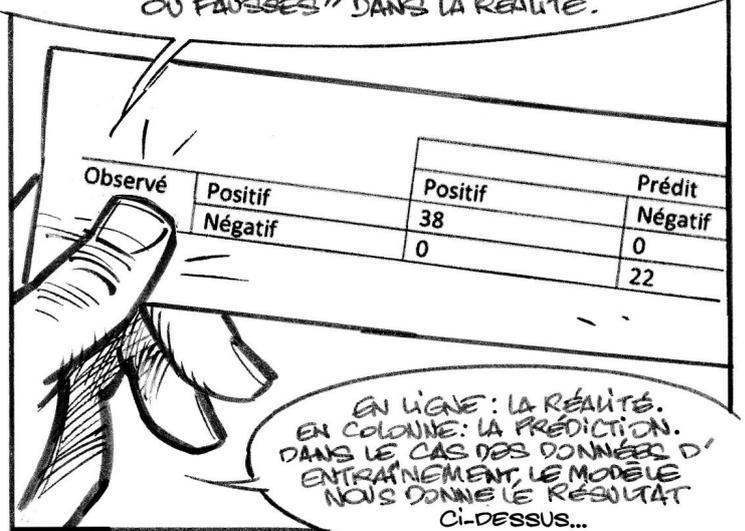
* ON PARLE DE DONNÉES D'APPRENTISSAGE OU D'ENTRAÎNEMENT.

** OU INVERSEMENT.

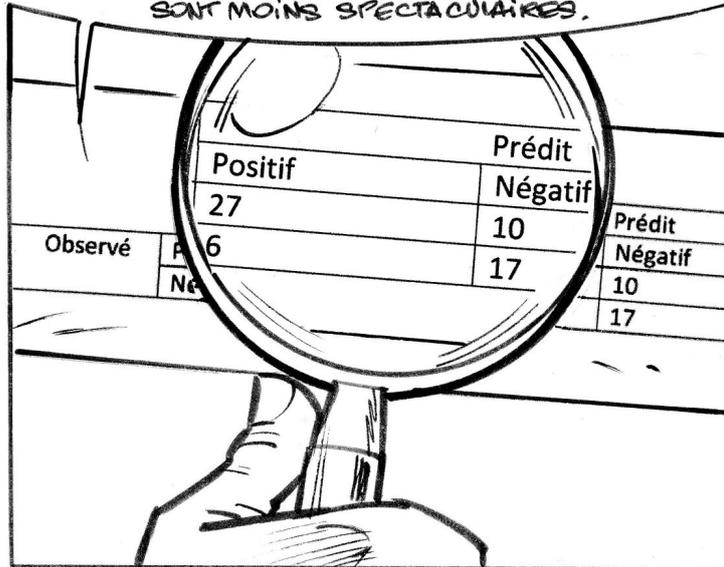
POUR CARACTÉRISER LA PERFORMANCE D'UN ALGORITHME DE CLASSIFICATION BINAIRE, ON PEUT REGARDER LA MATRICE DE CONFUSION.



CETTE TABLE 2X2 ILLUSTRÉ LES DONNÉES QUI SONT CLASSÉES "VRAIES OU FAUSSES" PAR LE MODÈLE. ON DIT ALORS POSITIF OU NÉGATIF ET "VRAIES OU FAUSSES" DANS LA RÉALITÉ.



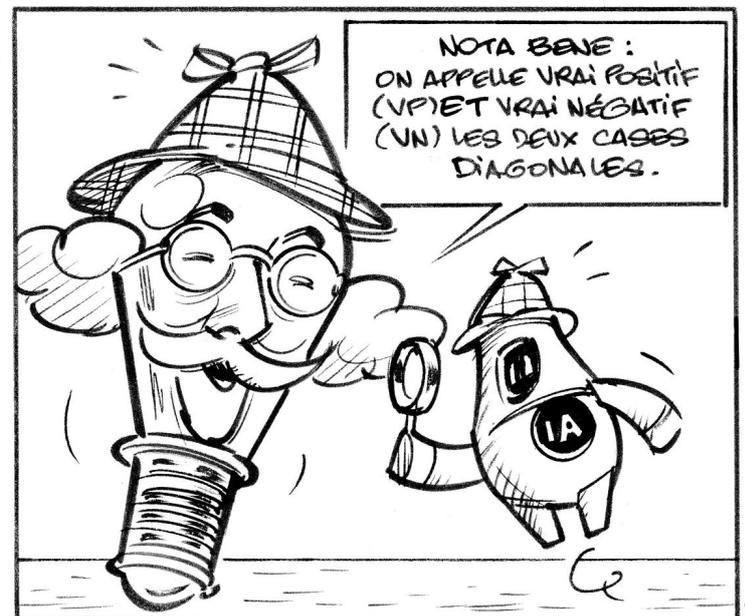
ON VOIT ICI QU'ON NE SE TROMPE JAMAIS. SUR LES DONNÉES TEST EN REVANCHE, LES PERFORMANCES SONT MOINS SPECTACULAIRES.



IL Y A DONC ICI 16 ERREURS DE PRÉVISION: 10 FAUX NÉGATIFS (FN), DES DONNÉES CORRESPONDANT À BON - CA = VRAI ET POUR LESQUELLES LE MODÈLE PRÉDIT FAUX.



ET 6 FAUX POSITIFS (FP), DES DONNÉES CORRESPONDANT À BON - CA = FAUX ET POUR LESQUELLES LE MODÈLE PRÉDIT VRAI.



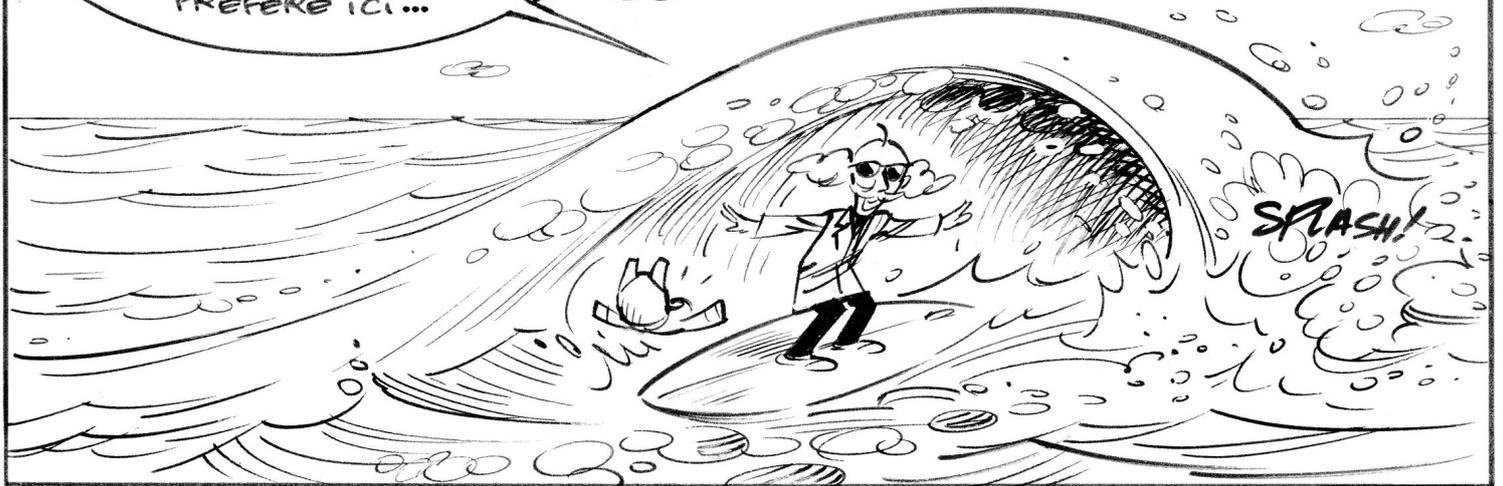
FAIRE DES ERREURS A UN COÛT QUI N'EST GÉNÉRALEMENT PAS SYMÉTRIQUE ENTRE LES ERREURS DE TYPE FAUX NÉGATIF ET FAUX POSITIF.

GIOVANNI, LE GLACIER, PRÉFÉRERA SÛREMENT AVOIR ICI DES FAUX NÉGATIFS (ET PENSER QUE LA PLAGE N'EST PAS RÉMUNÉRATRICE ALORS QU'ELLE L'EST, PUISQUE LES CONSÉQUENCES SONT FAIBLES) : IL CHOISIRA UNE PLAGE EN LAQUELLE IL A CONFIANCE



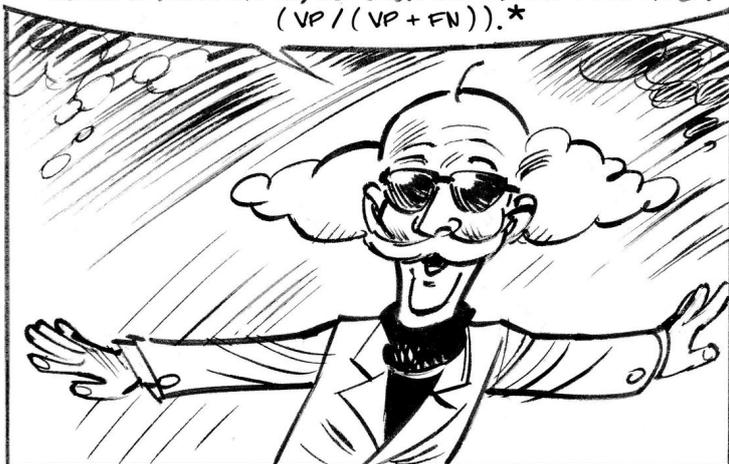
... PLUTÔT QUE D'AVOIR DES FAUX POSITIFS (PENSER QUE LA PLAGE EST RÉMUNÉRATRICE ALORS QU'ELLE NE L'EST PAS, PUISQUE L'IMPACT EST PLUS FÂCHEUX : IL SE DÉPLACERA POUR RIEN). UN MODÈLE AVEC UNE MATRICE DE CONFUSION, COMME EN HAUT À DROITE, SERA DONC SÛREMENT PRÉFÉRÉ ICI...

		PRÉDIT	
		POSITIF	NÉGATIF
OBSERVÉ	POSITIF	23	14
	NÉGATIF	2	21

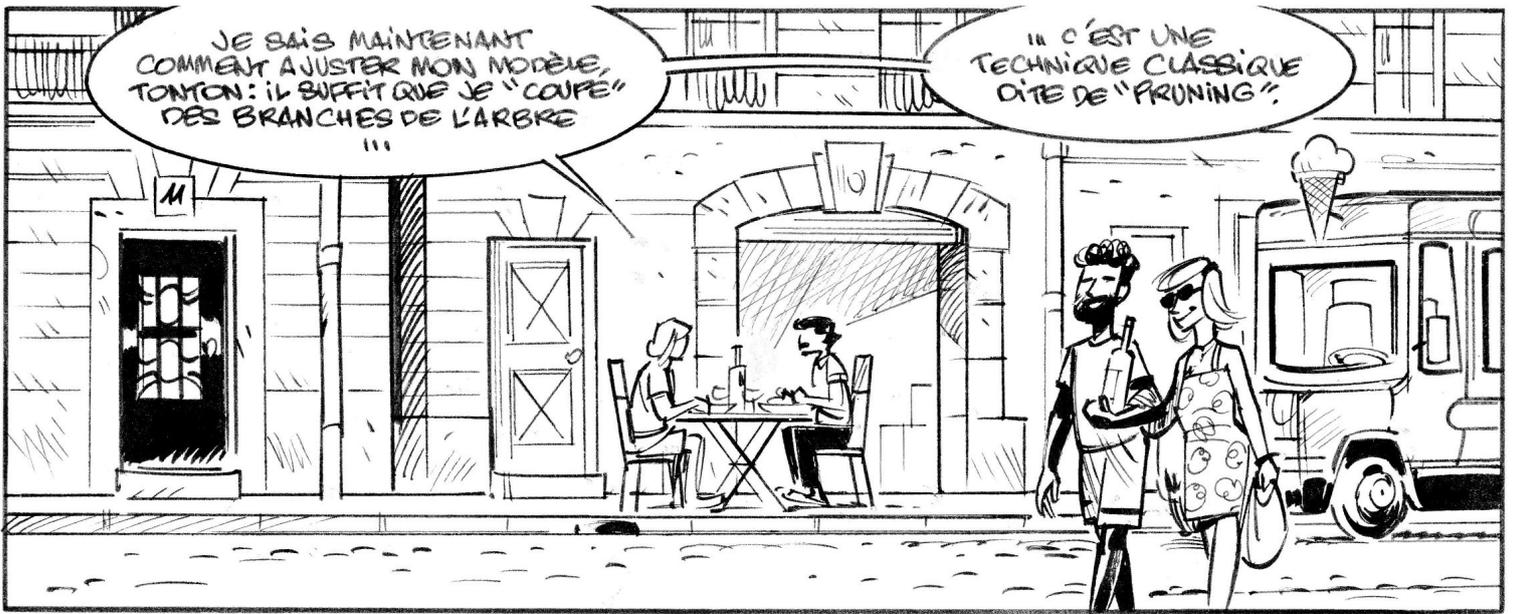


POUR COMPARER DEUX MATRICES DE CONFUSION, ON REGARDE SOUVENT LA PRÉCISION : $(VP + VN) / (VP + VN + FP + FN)$, LE TAUX DE FAUX POSITIFS $FP / (FP + VN)$ ET LE TAUX DE FAUX NÉGATIFS $FN / (VP + FN)$ OU SON COMPLÉMENTAIRE, LE TAUX DE VRAIS POSITIFS : $(VP / (VP + FN))$.*

UN MODÈLE EST EXCELLENT SI LE TAUX DE FAUX POSITIFS EST PROCHE DE 0 ET, SIMULTANÉMENT, QUE LE TAUX DE VRAIS POSITIFS EST PROCHE DE 1.



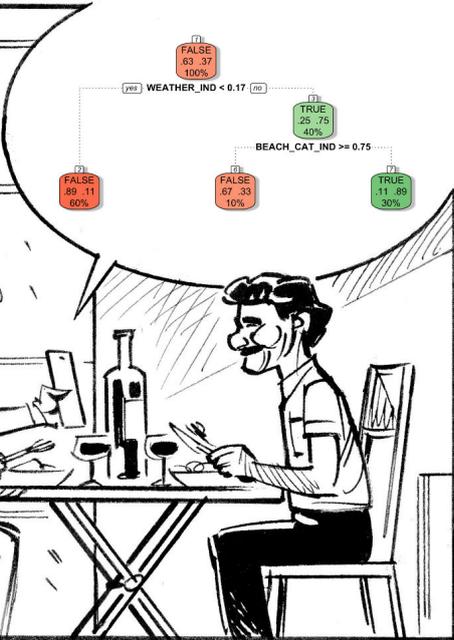
* PAR LA SUITE, ON NOTE TPR (TRUE POSITIVE RATE) ET FPR (FALSE POSITIVE RATE). LE TAUX DE VRAIS POSITIFS ET DE FAUX POSITIFS, RESPECTIVEMENT.



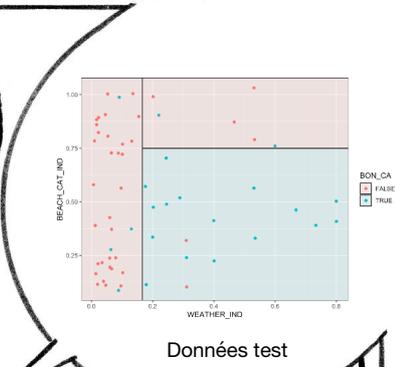
JE SAIS MAINTENANT COMMENT AJUSTER MON MODÈLE, TONTON: IL SUFFIT QUE JE "COUPE" DES BRANCHES DE L'ARBRE !!!

!!! C'EST UNE TECHNIQUE CLASSIQUE DITE DE "PRUNING".

VOICI CE QUE ÇA POURRAIT DONNER PAR EXEMPLE, EN IMPOSANT QUE JE GARDE SUFFISAMMENT DE REPRÉSENTANTS DANS CHAQUE FEUILLE.



TU VOIS C'EST MOINS BON SUR LES DONNÉES D'ENTRAÎNEMENT MAIS EN REVANCHE, C'EST BIEN MEILLEUR SUR LES DONNÉES DE TEST!



AH OUI, MAINTENANT J'AI PLUS ENVIE DE FAIRE CONFIANCE À TON MODÈLE! MAIS IL Y A QUAND MÊME DES CHANCES DE SE TROMPER, N'EST-CE PAS?



ON NE PEUT PAS GAGNER À TOUS LES COUPS MAIS STATISTIQUEMENT TU SERAS PLUTÔT GAGNANT EN PRINCIPLE!

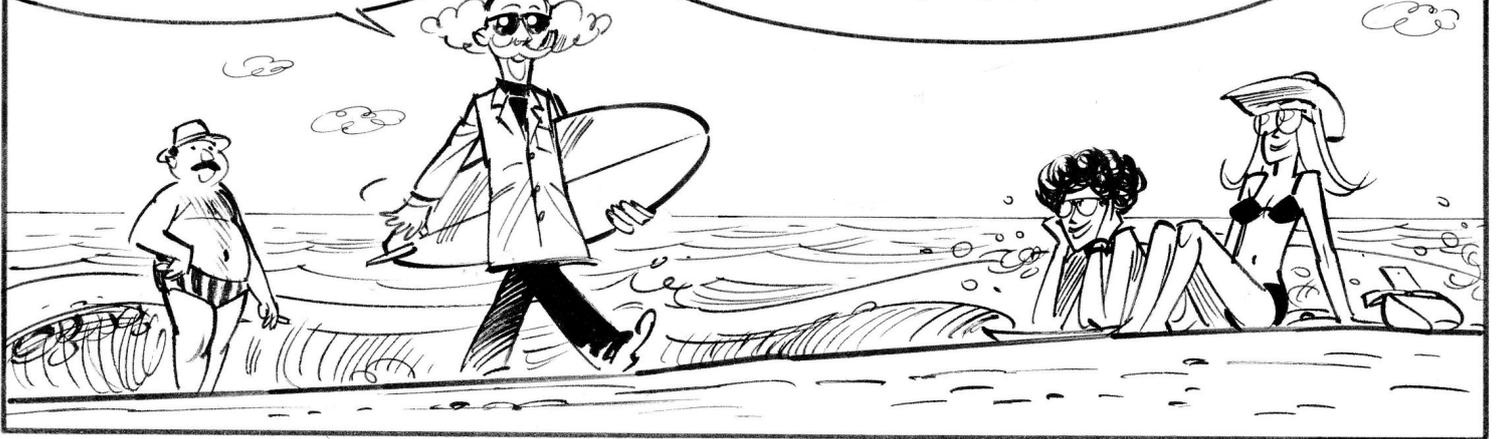
EN FAIT, ÇA DÉPEND UN PEU DE CE QUE TU PÉRDS SI ON SE TROMPE MAIS ON POURRAIT L'ÉVALUER EXACTEMENT SI TU ME DONNAIS LES CHIFFRES!



HÉ! HÉ!

QUAND ON NE SAIT PAS EXACTEMENT QUELS SONT LES COÛTS DES ERREURS, ON UTILISE SOUVENT LA FAMEUSE COURBE ROC.*

ON FAIT VARIER UN PARAMÈTRE IMPORTANT EN CLASSIFICATION : LE SEUIL À PARTIR DUQUEL ON PRÉDIT UN POSITIF EN CLASSIFICATION BINAIRE.

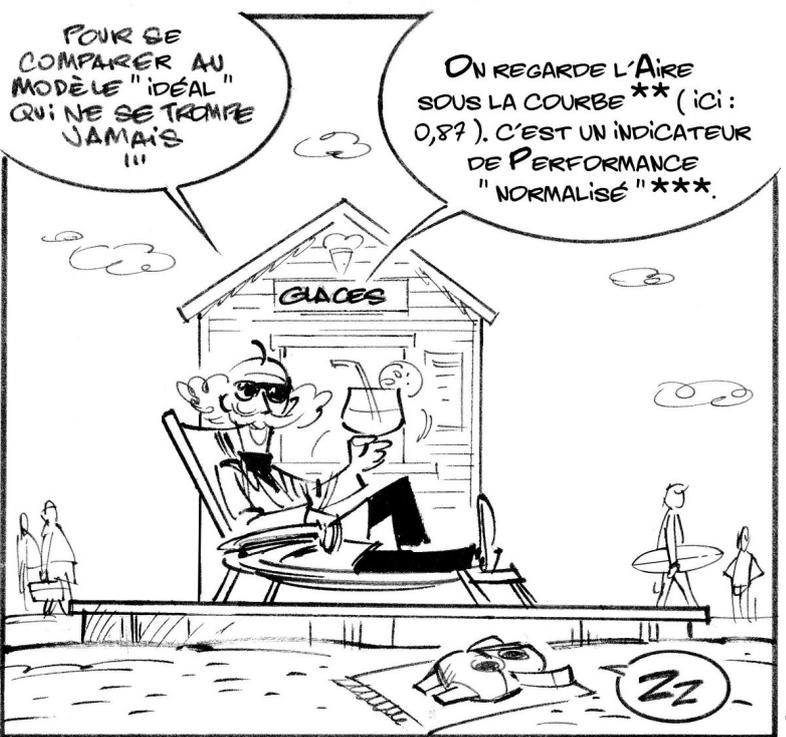
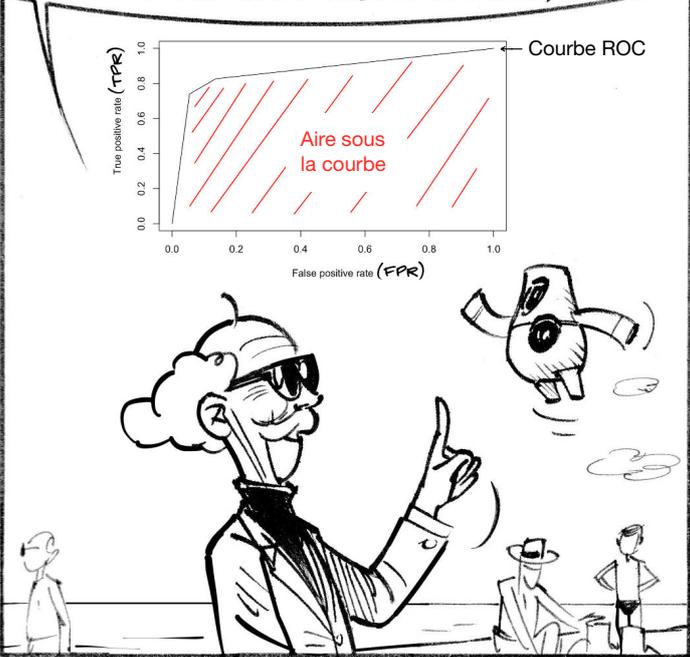


ON A SUPPOSÉ JUSQU'ICI QUE C'ÉTAIT AU-DESSUS DE 50% DE CHANCES MAIS ON POURRAIT ÊTRE PLUS CONSERVATEUR (e.g. 95%) OU PRENDRE PLUS DE RISQUES (e.g. 20%).

LA COURBE ROC REPRÉSENTE LES PERFORMANCES (TAUX DE VRAIS POSITIFS ET DE FAUX NÉGATIFS) D'UN MODÈLE POUR TOUTS LES SEUILS POSSIBLES DE 0% À 100%.



VOICI CE QUE ÇA DONNE POUR LE JEU DE TEST (LE RÉSULTAT EST SENSIBLEMENT IDENTIQUE SUR LE JEU D'ENTRAÎNEMENT).



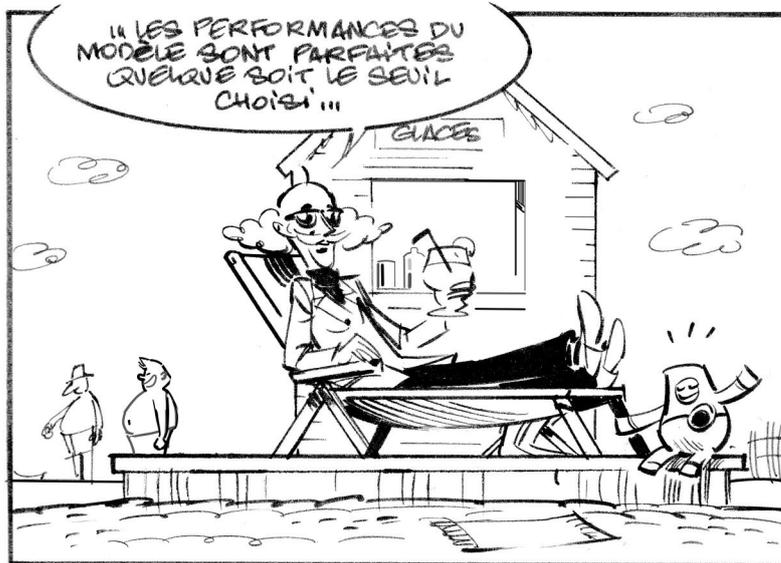
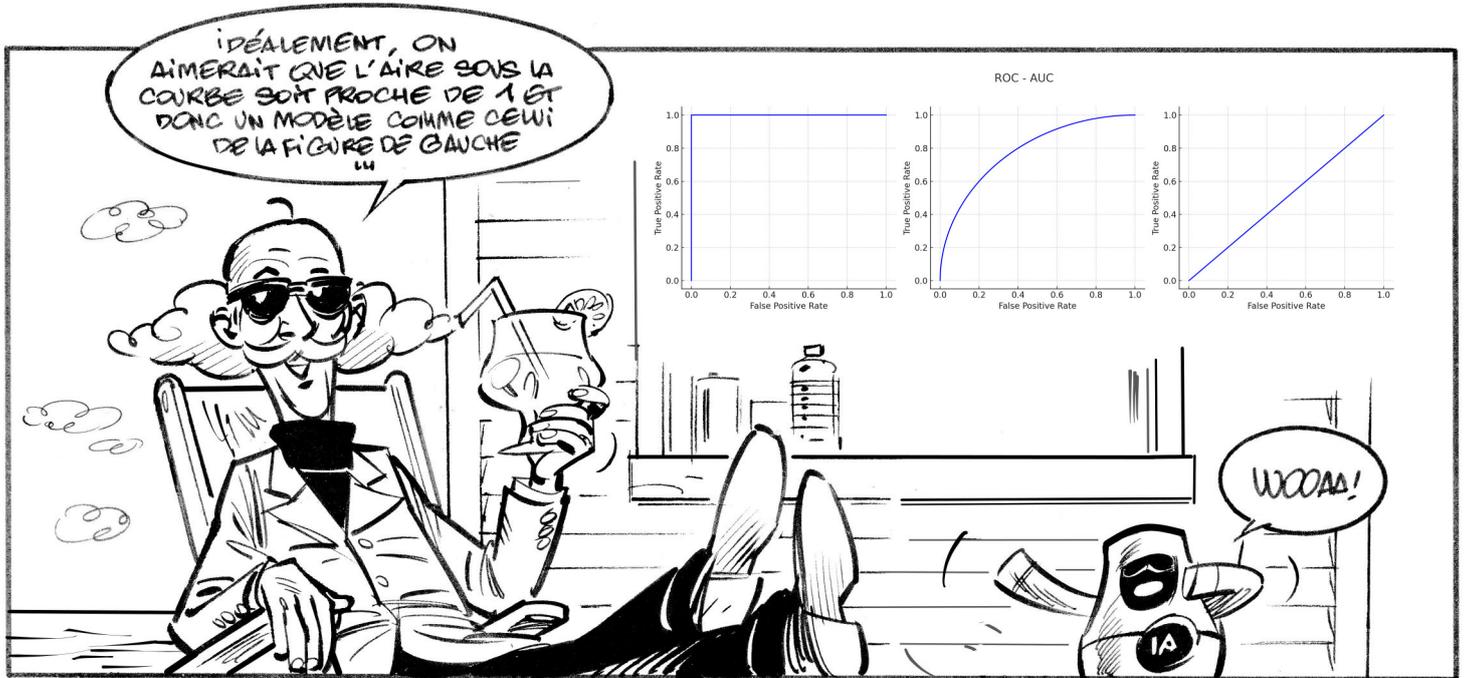
POUR SE COMPARER AU MODÈLE "IDÉAL" QUI NE SE TROMPE JAMAIS !!!

ON REGARDE L'AIRE SOUS LA COURBE** (ici: 0,87). C'EST UN INDICATEUR DE PERFORMANCE "NORMALISÉ"***.

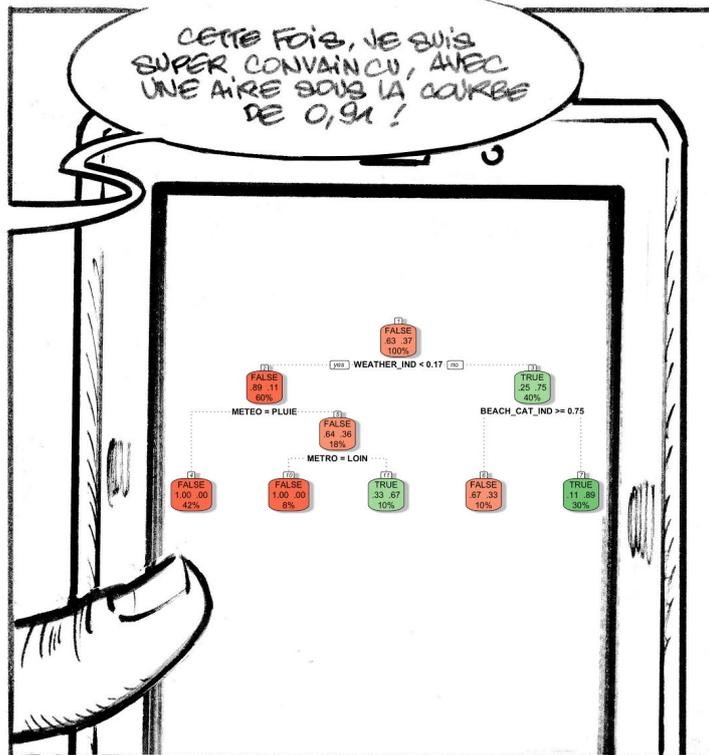
* https://fr.wikipedia.org/wiki/Courbe_ROC

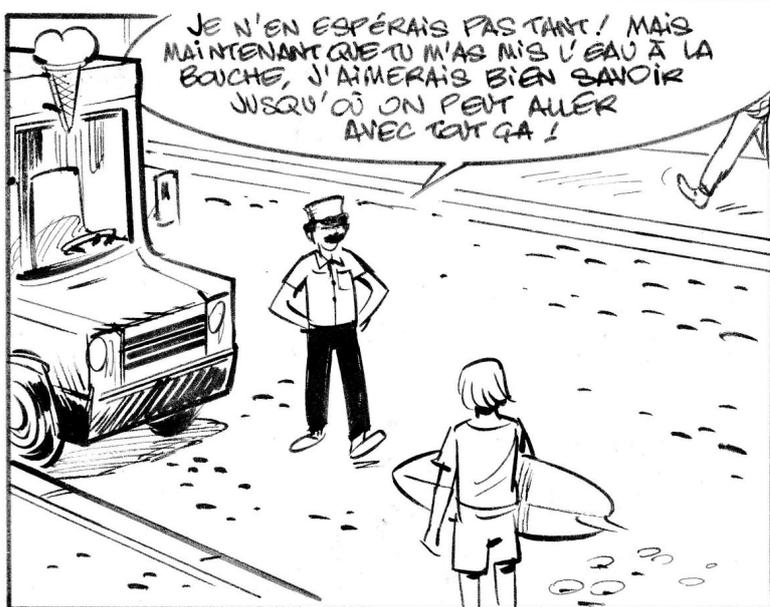
** LA SURFACE HACHURÉE SUR LA FIGURE : AUC EN ANGLAIS.

*** UN PEU COMME LE R^2 EN RÉGRESSION LINÉAIRE (VOIR UN PEU PLUS LOIN).



* POUR UN SEUIL, DISONS DE 33,3 %, ON LANCERA UN DÉ ET ON DIRA VRAI À CHAQUE FOIS QUE L'ON TOMBERA SUR LES CHIFFRES 1 OU 2.





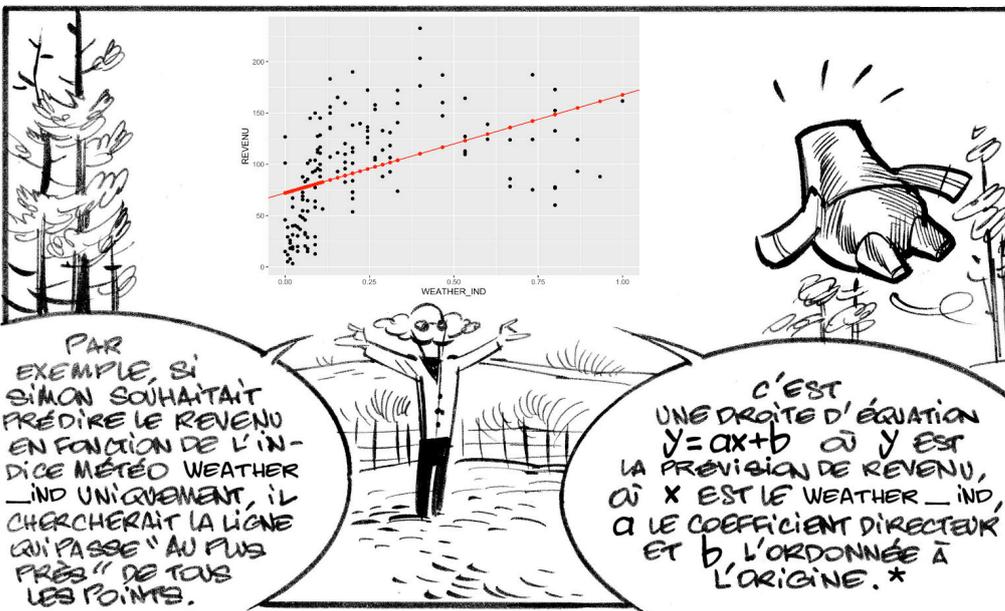
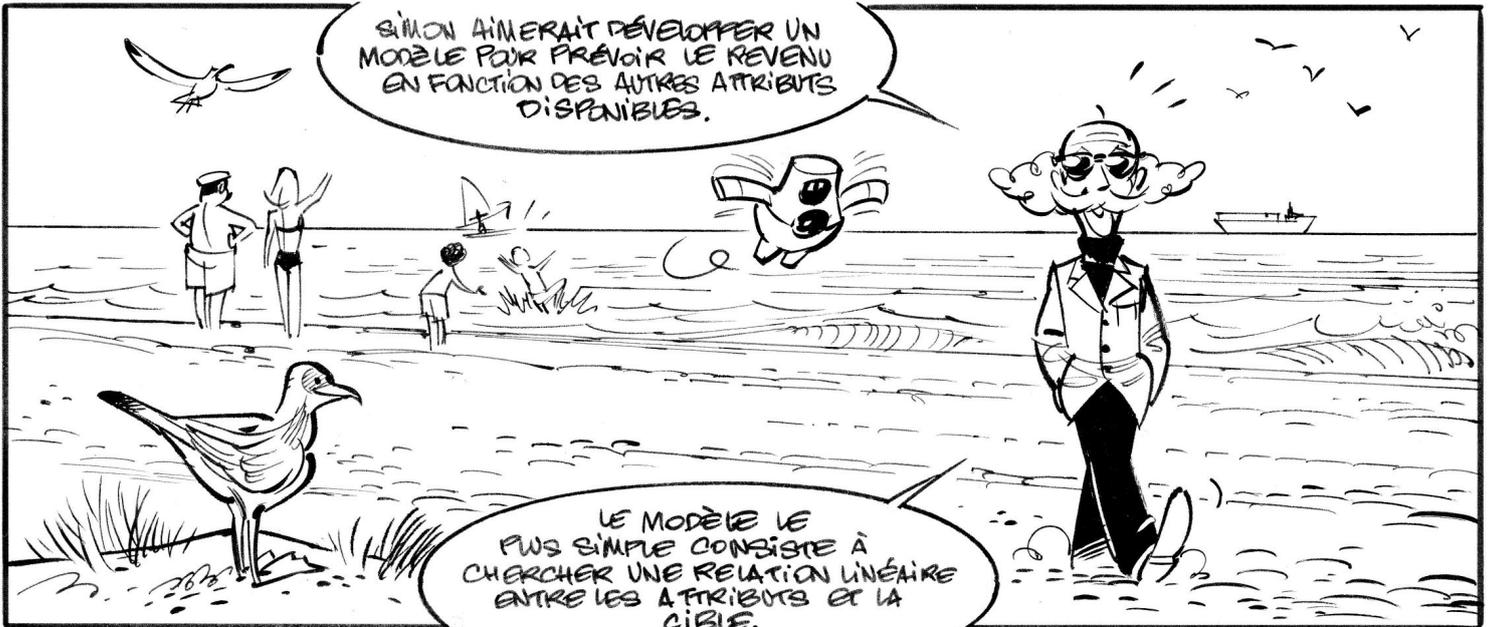


QUAND ON CHERCHE À PRÉVOIR DES DONNÉES CONTINUES COMME LE REVENU, ON PARLE DE RÉGRESSION. ET LES MODÈLES LES PLUS SIMPLES SONT LES MODÈLES DE RÉGRESSION LINÉAIRE. TU PEUX ME DONNER LES RÉSULTATS DES DERNIÈRES SEMAINES ?

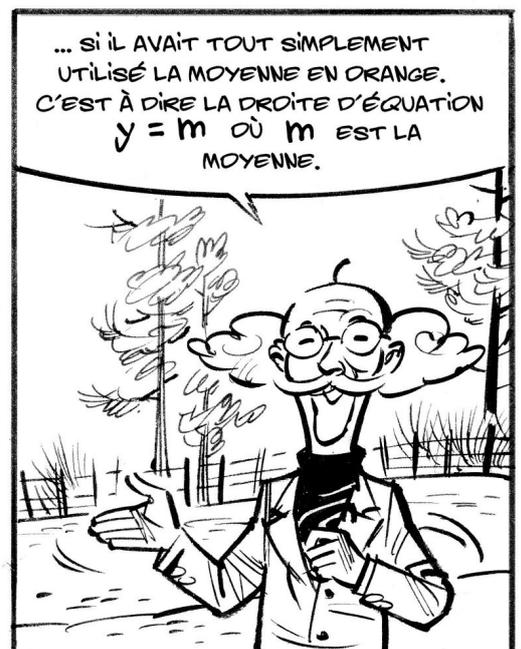
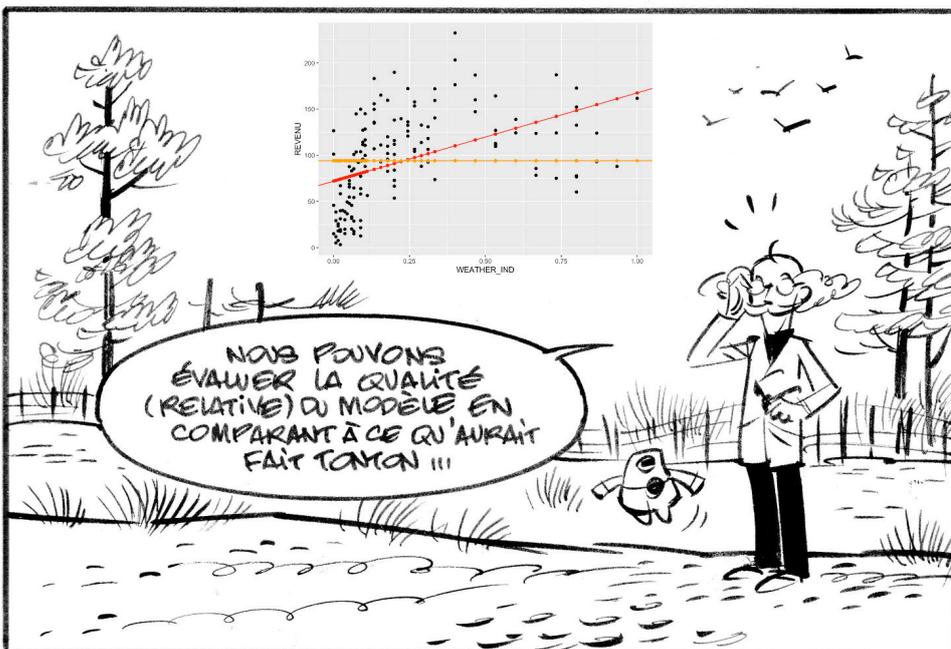
TEMP_CAT	TEMP	METEO	WE	BEACH_CAT	BEACH_CAT_IND	METRO	REVENU	WEATHER_IND	
FROID	15	PLUIE	1	URBAIN		0.7	PROCHE	46	0.00000000
NORMAL	24	PLUIE	0	NATURE		0.4	PROCHE	29	0.06666667
FROID	18	PLUIE	1	URBAIN		0.9	LOIN	40	0.02222222
NORMAL	24	PLUIE	0	URBAIN		0.7	PROCHE	54	0.06666667
FROID	19	NUAGE	0	URBAIN		0.7	LOIN	58	0.08888889
CANICULE	25	SOLEIL	1	URBAIN		0.6	LOIN	86	0.66666667



* LA MOYENNE EST LE PRÉDICTEUR LE PLUS NAÏF MAIS LE MEILLEUR, AU SENS DES MOINDRES CARRÉS, SI ON N'UTILISE AUCUN AUTRE ATTRIBUT DE PRÉVISION.



ON PEUT DÉFINIR "AU PLUS PRÈS" AU SENS DES MOINDRES CARRÉS**, C'EST À DIRE QUE L'ON SOMME LES ERREURS AU CARRÉ ENTRE LA PRÉVISION QUE L'ON FERAIT SI ON UTILISAIT LE POINT SUR LA LIGNE (POUR UNE VALEUR DE WEATHER_IND DONNÉE) ET LA RÉALITÉ.



* TROUVER LA MEILLEURE LIGNE REVIENT À DÉTERMINER LES MEILLEURS a ET b .

** ON UTILISE LES MOINDRES CARRÉS ESSENTIELLEMENT POUR RÉSOUDRE LE PROBLÈME FACILEMENT MAIS, DANS L'ABSOLU, ON POURRAIT UTILISER D'AUTRES MESURES DE L'ERREUR.



L'INDICATEUR DE PERFORMANCE DE RÉFÉRENCE POUR LES MODÈLES LINÉAIRES MESURE D'AILLEURS LA QUALITÉ RELATIVE D'UN MODÈLE LINÉAIRE EN CONSIDÉRANT LE RATIO ENTRE LA SOMME DES ERREURS ROUGE AU CARRÉ (E) ET LA SOMME DES ERREURS ORANGE AU CARRÉ (E'). C'EST CE QU'ON APPELLE LE R CARRÉ OÙ $R^2 = 1 - E/E'$.



ON PEUT GÉNÉRALISER À DES MODÈLES QUI INTÈGENT PLUSIEURS ATTRIBUTS D'ENTRÉE, C'EST À DIRE QU'ON CONSIDÈRE UNE APPROXIMATION DE LA FORME :

$$"y = a_1x_1 + a_2x_2 + \dots + a_nx_n + b"$$

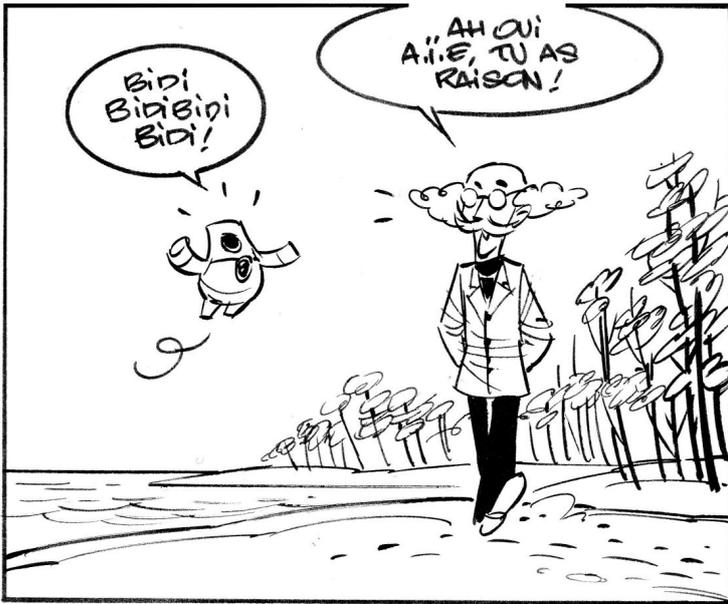
(OÙ y EST LA PRÉVISION, x_1, \dots, x_n SONT LES ATTRIBUTS ET a_1, \dots, a_n, b SONT DES PARAMÈTRES À DÉTERMINER**).

POUR UTILISER LE MODÈLE AFIN DE FAIRE DE LA PRÉVISION ON VEILLERA À UTILISER, COMME POUR LES ARBRES DE DÉCISION, UN JEU D'ENTRAÎNEMENT ET UN JEU DE TEST. ON POURRAIT ALLER PLUS LOIN ET FAIRE CE QU'ON APPELLE DE LA VALIDATION CROISÉE.

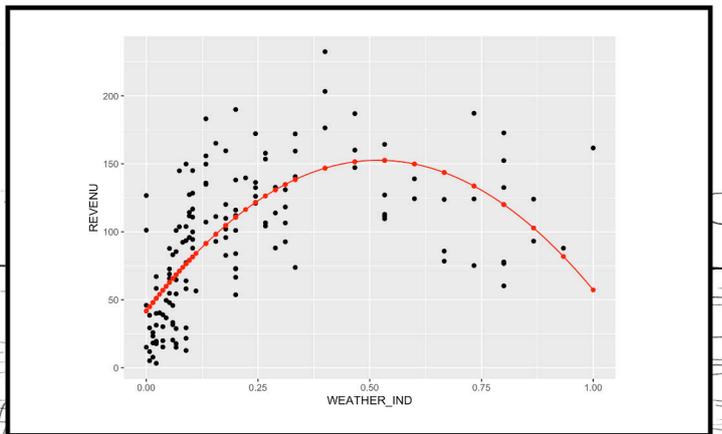
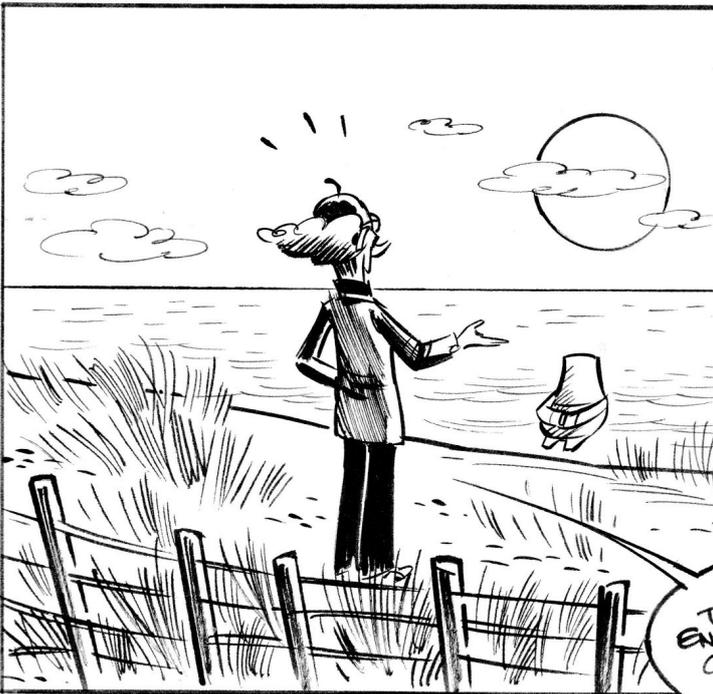
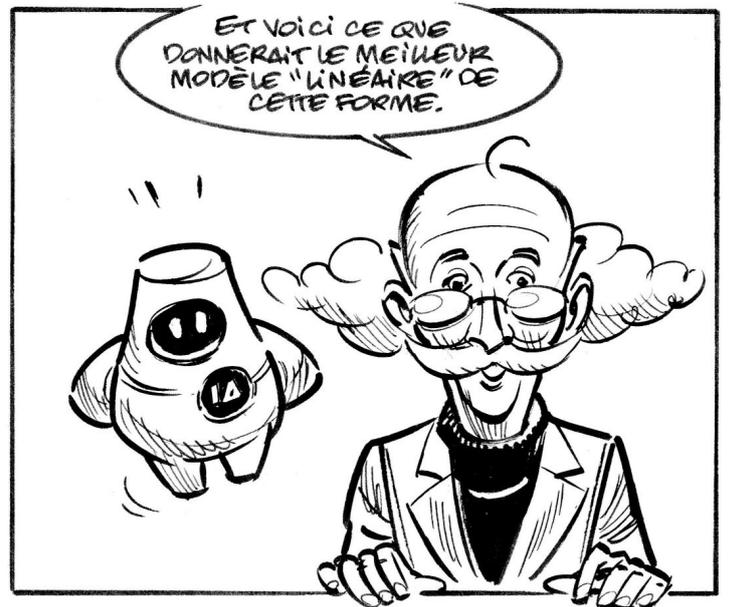
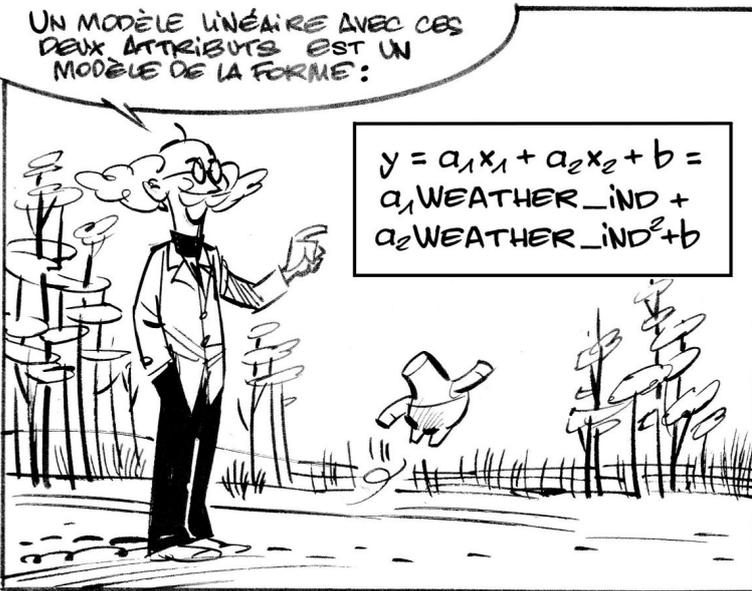


* SUR LES DONNÉES D'ENTRAÎNEMENT.

** ON VA ENCORE CHERCHER ICI LES MEILLEURS a_1, \dots, a_n, b AU SENS DES MOINDRES CARRÉS.



J'AI OUBLIÉ DE MENTIONNER QU'ON PEUT INTRODUIRE DES NON-LINÉARITÉS DANS DES MODÈLES LINÉAIRES EN AJOUTANT DES ATTRIBUTS FICTIFS. PAR EXEMPLE, ON POURRAIT AJOUTER UNE COLONNE AVEC LE CARRÉ DE L'INDICE MÉTÉO DANS LES DONNÉES.



TRouver LES BONs ATTRIBUTs ET LES BONNES FONCTIONS DE TRANSFORMATION S'APPELE DU "FEATURE ENGINEERING" ET DEMANDE SOUVENT UNE BONNE CONNAISSANCE MÉTIER, DE L'ANALYSE FINE ET BEAUCOUP D'ESSAIS !!



TEMP_CAT FROID	TEMP_CAT NORMAL	TEMP	METEO PLUIE	METEO SOLEIL	WE1	BEACH_CATU RBAIN	BEACH_CAT_IND	METRO PROCHE	WEATHER_IND	b
-87.13	-24.09	1.44	-47.85	66.78	22.5	6.66	13.25	10.66	-123.16	93.62



* ON OBTIENT LES COEFFICIENTS OPTIMAUX (AU SENS DES MOINDRES CARRÉS) FACILEMENT, EN RÉSOUVANT UN SYSTÈME D'ÉQUATION LINÉAIRE. C'EST AUSSI L'INTÉRÊT DE REGARDER LE MEILLEUR MODÈLE EN CE SENS.

SIMON, C'EST UN PEU DU CHARABIA POUR MOI, TU PEUX ME PRÉCISER COMMENT JE PEUX UTILISER TON MODÈLE ?



EH BIEN...

!!! IL SUFFIT DE REGARDER DANS QUEL CAS DE FIGURE TU TE TROUVES. PAR EXEMPLE : DISONS QUE TU SOUHAITES ÉVALUER LE REVENU CE WEEK-END SUR LA PLAGE DE LACANAU : PLAGE D'INDICE URBAIN 0,9 ET PROCHE DE BORDEAUX !!!



!! S'ILS ANNONCENT QU'IL VA FAIRE UNE TEMPÉRATURE DE 24 DEGRÉS (DONC, NORMAL POUR LA SAISON), QU'IL Y AURA DE LA PLUIE ET QUE LE WEATHER_IND EST DE 0.03333 (IL EST CALCULÉ SELON UNE FORMULE SAVANTE À PARTIR DE LA TEMPÉRATURE ET DE LA MÉTÉO) !!!



!!! LE REVENU SERA ESTIMÉ À :

$$93.62 - 24.09 \times 1 + 1.44 \times 24 - 47.85 \times 1 + 22.5 \times 1 + 6.66 \times 1 + 13.25 \times 0.9 + 10.66 \times 1 - 123.16 \times 0.03333 = 103.88$$



EUH !!! POURQUOI "FOIS 1" POUR -24.09, -47.85, 22.5, 6.66 ET 10.66 ?

AH PARDON ! CE SONT DES VARIABLES BINAIRES QUI INDICENT DANS QUELLE CLASSE ON SE TROUVE ET DONC ÇA REVIENT À PRENDRE SEULEMENT LES COEFFICIENTS DES CATÉGORIES QUI NOUS CONCERNENT.



AH D'ACCORD ! DONC SI, PAR EXEMPLE, ÇA N'ÉTAIT PAS LE WEEK-END, JE SUPPRIMERAIS LE COEFFICIENT 22.5 ?

EXACTEMENT TANTON !! TU COMPRENDS VITE !

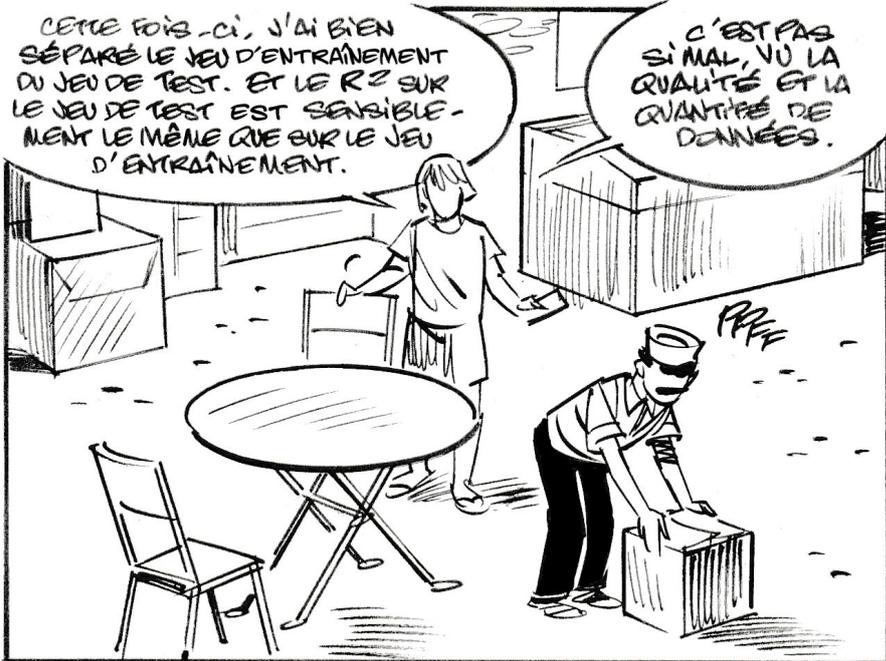




AH TU SAIS MOI, QUAND
ÇA CONCERNE MA RECETTE
JE ME CONCENTRE VITE !!
DONC JE COMPRENDS QUE
JE PEUX AINSI COMPARER
FACILEMENT DIFFÉ-
RENTS PLATS

!! EN APPLIQUANT LA
FORMULE POUR CHACUNE
D'EUX. GÉNIAL!

ET COMMENT SAVOIR SI JE
PEUX AVOIR CONFIANCE EN TON
MODÈLE ? ON A EU DES SURPRISES
AVEC TON PREMIER ARBRE
DE DÉCISION !!



CETTE FOIS-CI, J'AI BIEN
SÉPARÉ LE JEU D'ENTRAÎNEMENT
DU JEU DE TEST. ET LE R² SUR
LE JEU DE TEST EST SENSIBLE-
MENT LE MÊME QUE SUR LE JEU
D'ENTRAÎNEMENT.

C'EST PAS
SI MAL, VU LA
QUALITÉ ET LA
QUANTITÉ DE
DONNÉES.

ÇA VEUT DIRE QUE TU RÉDUIS L'ERREUR,
DE 72 % (AU SENS DES MOINDRES CARRÉS,
PLUTÔT QU'UTILISER NAÏVEMENT LA
MOYENNE COMME PRÉDICTION).



OUCK!



AH OUI, ÇA SEMBLE SIGNIFICATIF. MAIS
J'AI UNE QUESTION BÊTE SIMON : JE NE
CONNAIS PAS LA MÉTÉO DE CE WEEK-END !!!
COMMENT JE FAIS ? JE PRENDS SIM-
PLEMENT LES PRÉVISIONS MÉTÉO
?

TA QUESTION N'EST PAS SI NAÏVE
TONTON ! ON TOUCHE ICI À UN POINT IMPORTANT :
TU AS RAISON, TU NE CONNAIS PAS LA MÉTÉO
AVEC PRÉCISION LORSQUE TU DOIS DÉCIDER.
DONC, SI TU DÉVELOPPES UN MODÈLE DE
PRÉVISION, IL DOIT INTÉGRER TA CONNAISSANCE
DE LA MÉTÉO AU MOMENT OÙ TU
UTILISES LE MODÈLE.



ET TU NE PEUX PAS UTILISER DE DONNÉES D'ENTRAÎNEMENT QUI UTILISERAIENT, PAR EXEMPLE, L'INDICE MÉTÉO (WEATHER-IND) EFFECTIVEMENT RÉALISÉ. CAR AU MOMENT OÙ TU DOIS DÉCIDER, TU NE LES CONNAIS PAS



ON PARLE DE FUGITE TEMPORELLE SI TU UTILISES DE LA DONNÉE POUR T'ENTRAÎNER, QUI NE SERA PAS DISPONIBLE AU MOMENT D'UTILISER LE MODÈLE. C'EST UNE ERREUR TRÈS CLASSIQUE EN MACHINE LEARNING, ET PARFOIS, SUFFISAMMENT À DÉTECTER



DONC ICI, J'AI UTILISÉ LES PRÉVISIONS MÉTÉO UNE SEMAINE À L'AVANCE POUR ENTRAÎNER LE MODÈLE, PUISQUE TU DOIS DÉCIDER LE LUNDI DE TA PRODUCTION POUR LA SEMAINE.

MAINTENANT, ON POURRAIT AUSSI DÉVELOPPER UN MODÈLE COMPLÉMENTAIRE POUR SAVOIR OÙ ALLER CHAQUE MATIN EN FONCTION DE CONDITIONS MÉTÉO PUS FINES POUR MAXIMISER LES VENTES SUR LA JOURNÉE. MAIS ÇA SERAIT UN MODÈLE UN PEU DIFFÉRENT, SÛREMENT.



AH D'ACCORD! DANS CE CAS, JE POURRAIS UTILISER LE MODÈLE PUS FIN POUR ESSAYER DE TERMINER MA SEMAINE PUS TÔT!



C'EST GÉNIAL TON TRUC, ÇA ME PERMETTRAIT PEUT-ÊTRE DE DÉGAGER PARFOIS DU TEMPS POUR FRANCISCA. TU SAIS, TU TANTE ME RE-PROCHE BEAUCOUP DE NE PAS ÊTRE TRÈS DISPONIBLE L'ÉTÉ !!



JE TE DÉVELOPPE ÇA TANTON!

LE WEEK-END SUIVANT.

IL EST
SUPER CE
PETIT!

C'EST LA PREMIÈRE
FOIS QU'ON A DU TEMPS
POUR NOUS PENDANT
L'ÉTÉ!



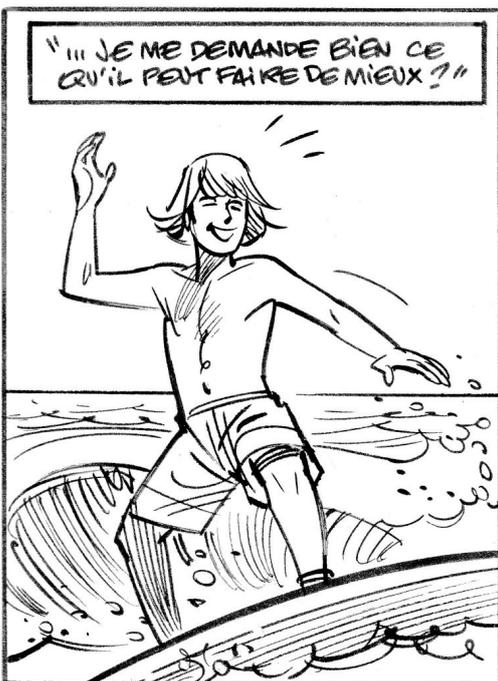
C'EST VRAI ET
ATTENDS, ÇA N'EST
PAS TOUT.



IL M'A DIT QU'IL
AVAIT UNE AUTRE
SUPER IDÉE POUR
NOUS!



"... JE ME DEMANDE BIEN CE
QU'IL PEUT FAIRE DE MEUX ?"



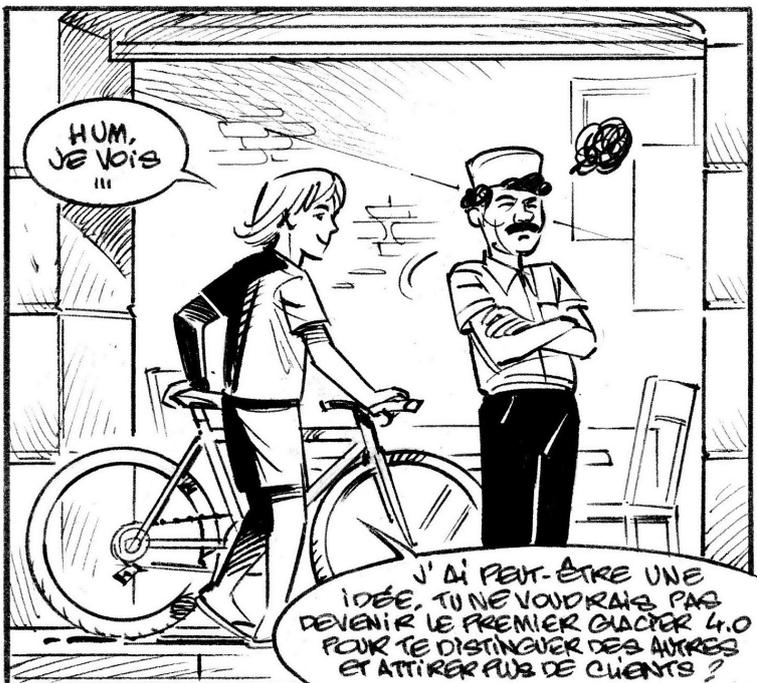
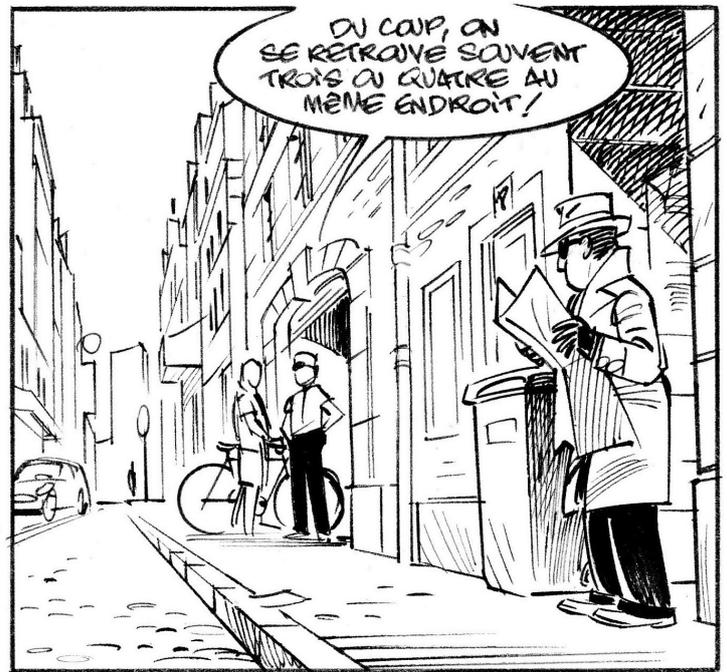
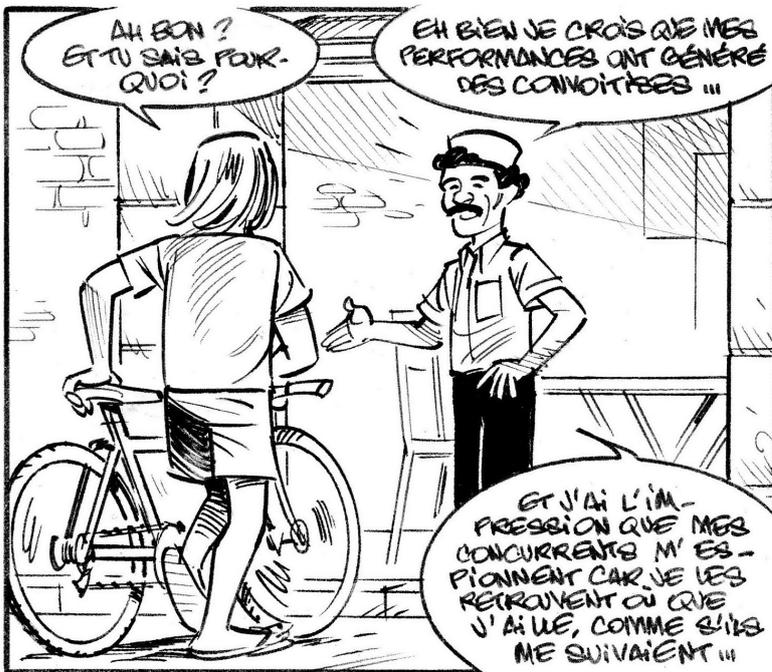
PEUT-ÊTRE
UN ROBOT POUR QUE TU
PUISSES TE LA COULER
D'UNE TOUT L'ÉTÉ?

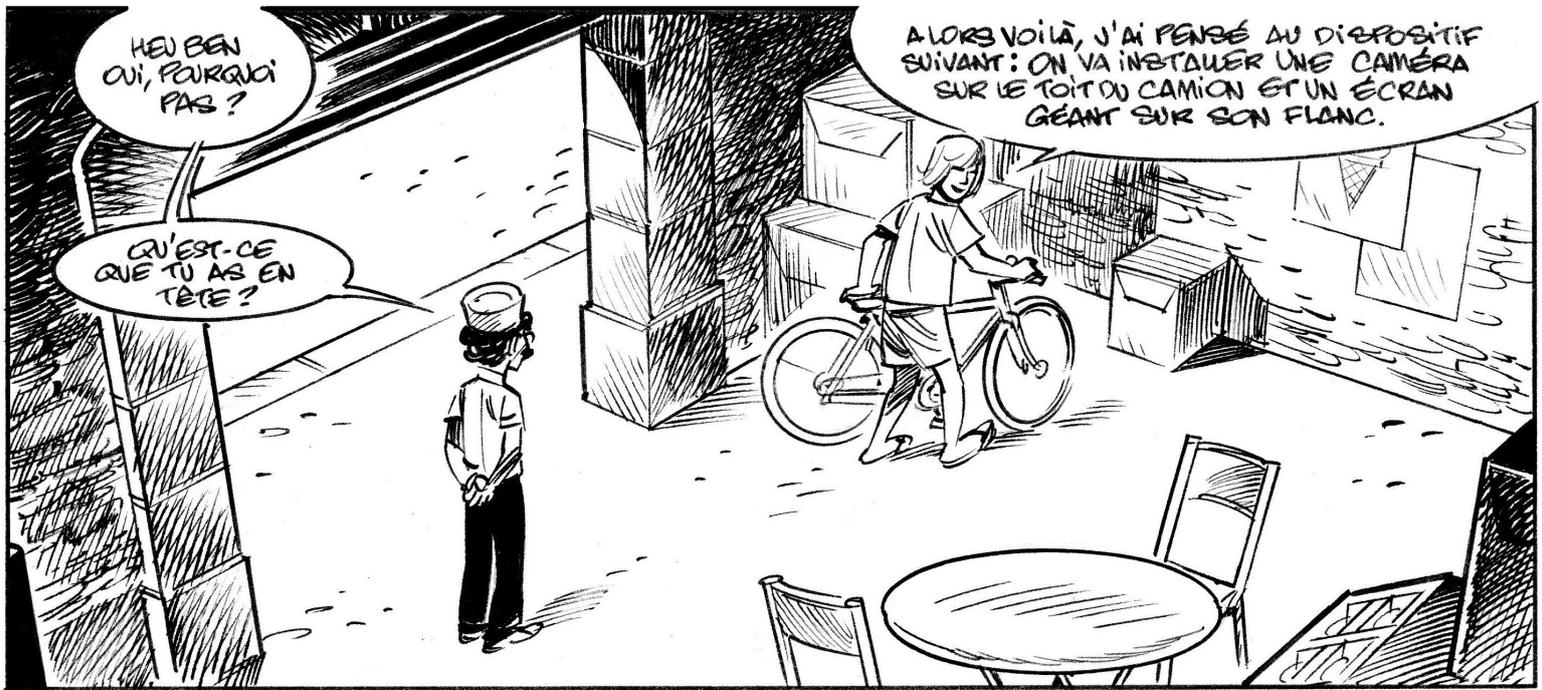


HA!
HA! HA!



CHAPITRE 2 : RÉSEAUX DE NEURONES





HEU BEN QUI, POURQUOI PAS ?

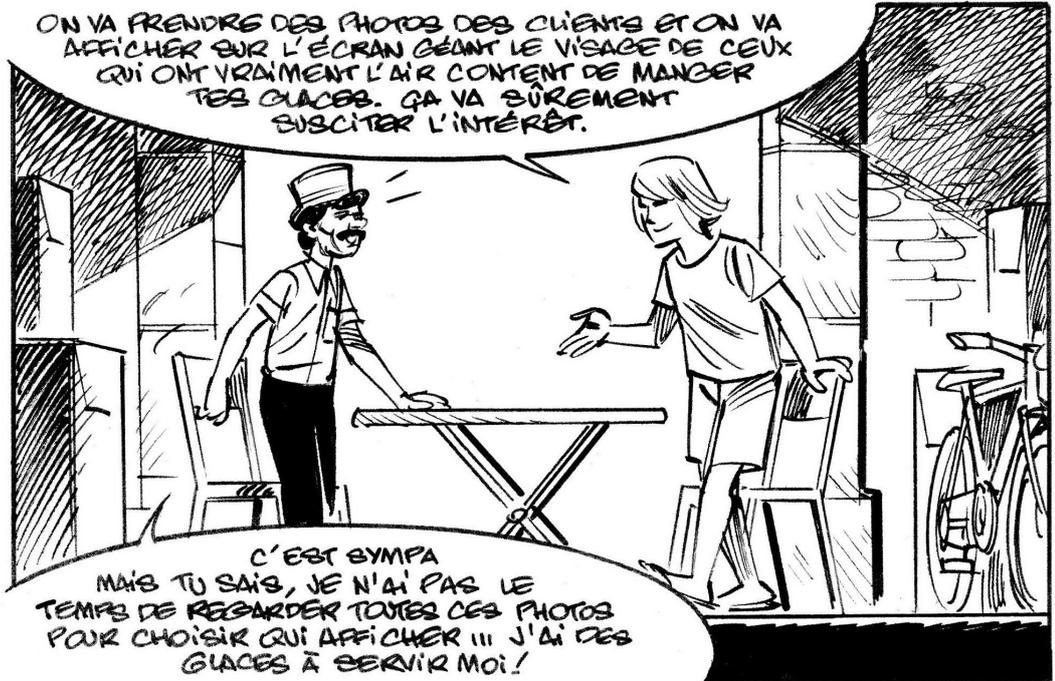
QU'EST-CE QUE TU AS EN TÊTE ?

ALORS VOILÀ, J'AI PENSÉ AU DISPOSITIF SUIVANT : ON VA INSTALLER UNE CAMÉRA SUR LE TOIT DU CAMION ET UN ÉCRAN GÉANT SUR SON FLANC.



HHEUUU... D'AAACCORDO!

GRAT GRAT



ON VA PRENDRE DES PHOTOS DES CLIENTS ET ON VA AFFICHER SUR L'ÉCRAN GÉANT LE VISAGE DE CEUX QUI ONT VRAIMENT L'AIR CONTENT DE MANGER TES GLACES. ÇA VA SÛREMENT SUSCITER L'INTÉRÊT.

C'EST SYMPA MAIS TU SAIS, JE N'AI PAS LE TEMPS DE REGARDER TOUTES CES PHOTOS POUR CHOISIR QUI AFFICHER !!! J'AI DES GLACES À SERVIR MOI !



PAS DE PROBLÈME, ON VA AUTOMATISER LA SÉLECTION DES PHOTOS À AFFICHER.

LAISSE-MOI DEVINER, IL YA UN MODÈLE DE MACHING LEARNING POUR ÇA ?



BINGO ! C'EST LE DEEP LEARNING !

AH !

ÇA S'INSPIRE DU FONCTIONNEMENT DU CERVEAU QUI FONCTIONNE PAR TRAITEMENT ET TRANSMISSION D'INFORMATIONS ENTRE LES NEURONES POUR APPRENDRE À GÉNÉRALISER À PARTIR D'EXEMPLES.



HEU TU SAIS, MES COURS DE BIOLOGIE REMONTENT À LOIN MAINTENANT... ET PUIS JE N'ÉTAIS PAS VRAIMENT UNE FLÈCHE EN LA MATIÈRE... LE CERVEAU ÇA M'ÉVOQUE PLUTÔT UN FILM FRANCO-ITALIEN DES ANNÉES 60 AVEC JEAN-PAUL BELMONDO...



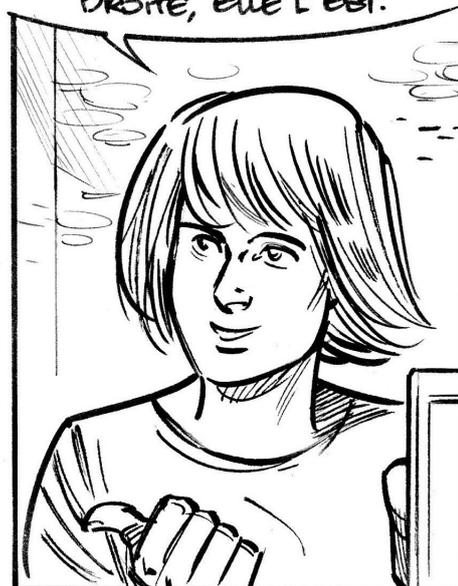
PAS DE PANIQUE TONTON ! EN FAIT, C'EST ASSEZ SIMPLE CONCEPTUELLEMENT, ON APPELLE ÇA UN RÉSEAU DE NEURONES ARTIFICIELS. JE VAIS T'EXPLIQUER.



ON VA DONNER DES IMAGES* DE CE TYPE AU MODÈLE.



EN LUI PRÉCISANT QUE SUR LA PHOTO DE GAUCHE, LA DAME N'EST PAS SATISFAITE, TANDIS QUE SUR LA PHOTO DE DROITE, ELLE L'EST.



* DISONS DE 32 PIXELS PAR 32 POUR ILLUSTRER, SOIT 1024 PIXELS EN TOUT.

TU VOIS, ON PEUT REPRÉSENTER CES INFORMATIONS DANS UNE TABLE EN NUMÉROTANT LES PIXELS DE 1 À 1024. ON INDIQUE POUR CHAQUE LIGNE/IMAGE LE NIVEAU DE GRIS (DE 0: NOIR À 1: BLANC) DE CHAQUE PIXEL ET ON AJOUTE UNE COLONNE SATISFAIT AVEC UNE VALEUR 0 OU 1 QUI INDIQUE LE RÉSULTAT SUR LA PHOTO EN QUESTION.

Pixel 1	...	Pixel 1024	SATISFAIT
0.24	...	0.76	1
0.36		0.27	0

COMME LES ARBRES (DE DÉCISION), Ton DEEP LEARNING VA APPRENDRE DE CES EXEMPLES, POUR LESQUELS IL CONNAÎT LA RÉPONSE, POUR TENTER ENSEMBLE DE CLASSIFIER DE NOUVELLES IMAGES, ENTRE SATISFAIT ET INSATISFAIT ?

VOILÀ, C'EST ÇA !

LE DEEP LEARNING FONCTIONNE PARTICULIÈREMENT BIEN AVEC DES IMAGES.

EN FAIT, L'ANALYSE DES IMAGES NÉCESSITE DE NOMBREUX PRÉTRAITEMENTS. ON NE PARLE ICI QUE DU COEUR DE L'ALGORITHME, LE RÉSEAU DE NEURONES.

MAIS AVANT D'ÊTRE ENTRÉES DANS LE RÉSEAU, LES IMAGES DOIVENT ÊTRE PRÉTRAITÉES !!!

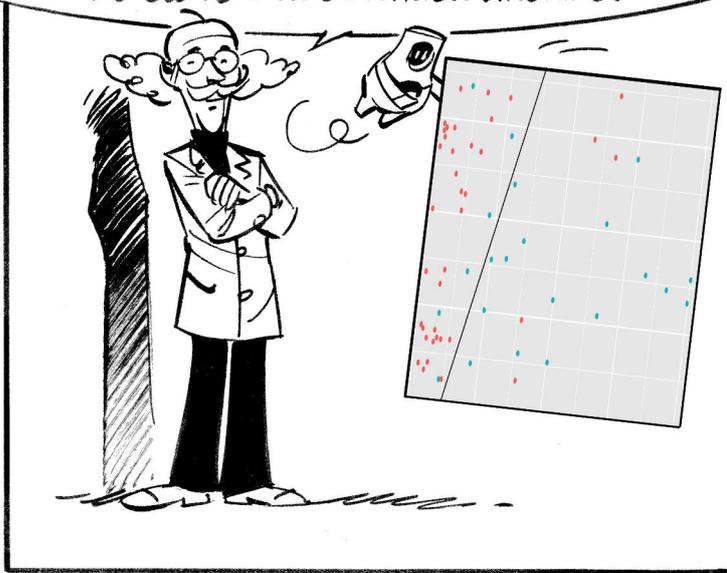
TAP! TAP!

Pixel 1	...
0.24	...

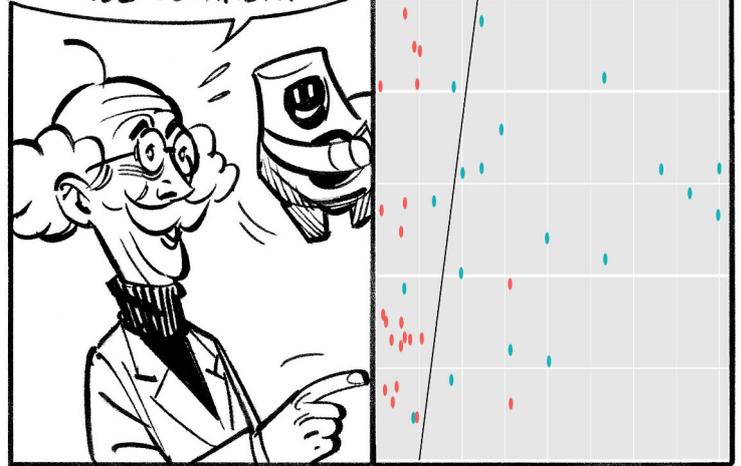
!!! POUR NE GARDER QUE LES ATTRIBUTS-CLÉS : LES YEUX, LA BOUCHE VOIRE LES OREILLES, COMME ICI. UNE SORTE DE PASSAGE À LA MOULINETTE!

IMPRESSIONNANT SI ON PEUT VRAIMENT FAIRE ÇA ! LA TÊTE DE LUIGI, MON COPAIN VENDEUR DE BOISSONS, QUAND IL VERRA ÇA ! MAIS ÇA MARCHE COMMENT ALORS, UN RÉSEAU DE NEURONES ARTIFICIELS ?

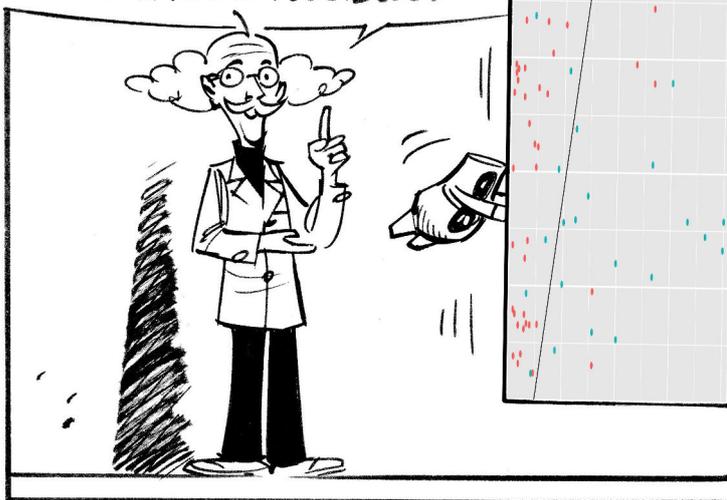
AVANT DE PARLER D'UN RÉSEAU DE NEURONES, COMMENÇONS PAR VOIR CE QU'EST UN NEURONE ARTIFICIEL. UN NEURONE EST ESSENTIELLEMENT UN MODÈLE DE CLASSIFICATION LINÉAIRE.



DANS LE CAS DE LA CLASSIFICATION BINAIRE, ON ESSAIE DE DISTINGUER LES BONS POINTS (EN VERT) DES MAUVAIS (EN ROUGE), EN IDENTIFIANT UNE LIGNE QUI POURRAIT "BIEN" LES SÉPARER.



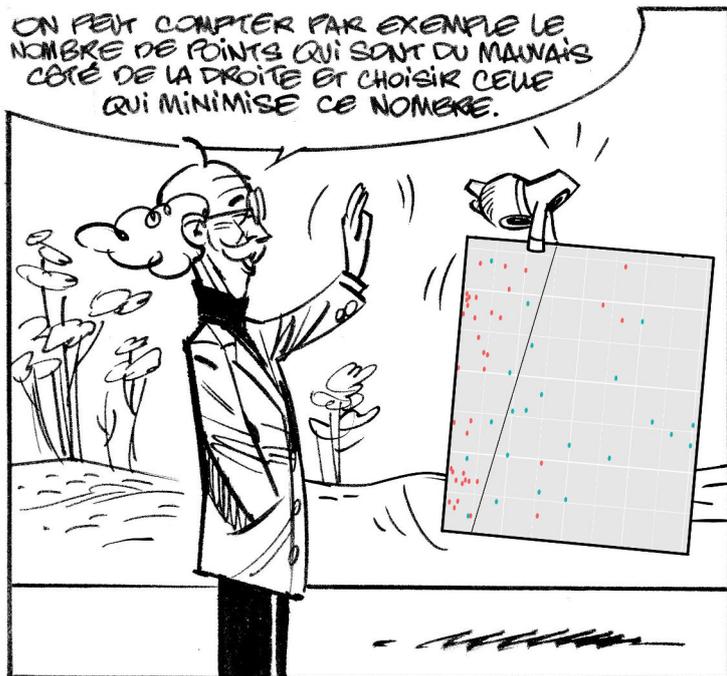
POUR ILLUSTRER SUR NOTRE PROBLÈME DE CLASSIFICATION DU CA*, ON CHERCHE UNE DROITE QUI SÉPARE NOS POINTS AVEC LE MOINS D'ERREURS POSSIBLES.



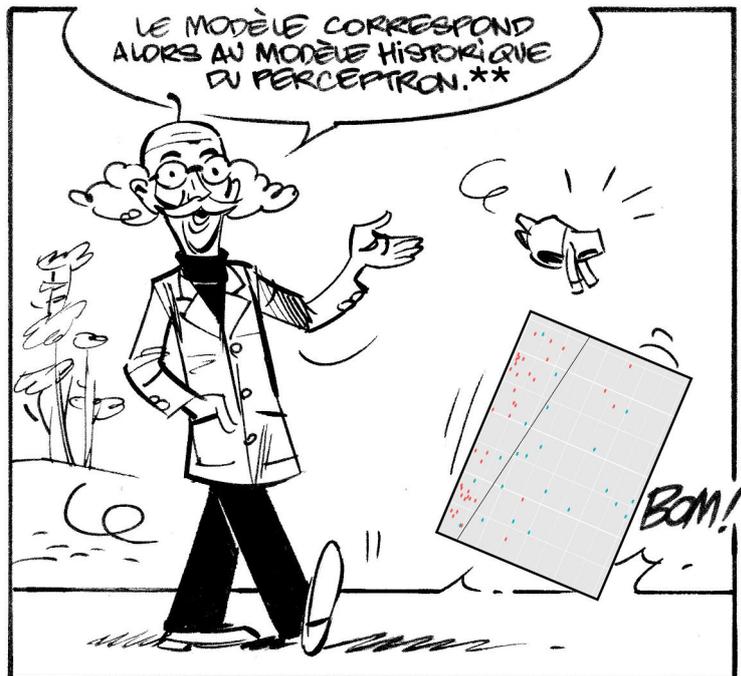
IL Y A DIFFÉRENTES MANIÈRES DE COMPTER LES ERREURS QUI DONNENT LIEU À DIFFÉRENTES VARIANTES.



ON FEIT COMPTER PAR EXEMPLE LE NOMBRE DE POINTS QUI SONT DU MAUVAIS CÔTÉ DE LA DROITE ET CHOISIR CELLE QUI MINIMISE CE NOMBRE.



LE MODÈLE CORRESPOND ALORS AU MODÈLE HISTORIQUE DU PERCEPTRON.**



* CHIFFRE D'AFFAIRE : VOIR CHAPITRE 1.

** [HTTPS://FR.WIKIPEDIA.ORG/WIKI/PERCEPTRON](https://fr.wikipedia.org/wiki/Perceptron)

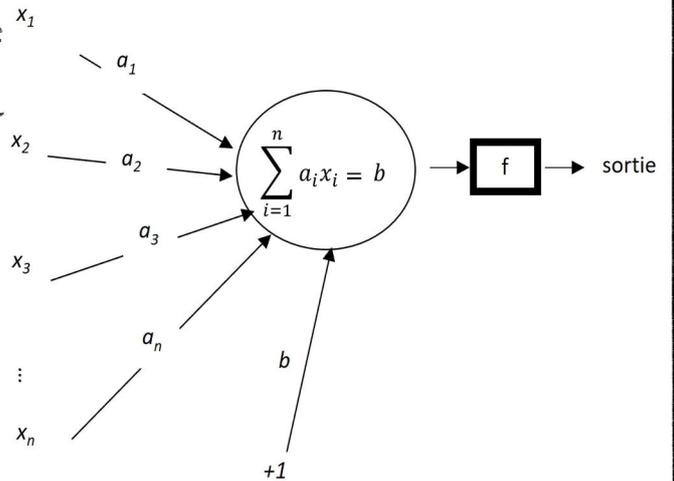
EN FAIT, PLUS GÉNÉRALEMENT, UN NEURONE PEUT ÊTRE VU COMME UN PETIT MOTEUR DE CALCUL QUI VA PRENDRE EN ENTRÉE LES DONNÉES : DANS LE CAS D'UNE IMAGE, LE NIVEAU DE GRIS DES PIXELS QUI COMPOSENT LA PHOTO (CF LA TABLE PAGE 29). PUIS LE NEURONE VA RETOURNER UNE FONCTION f DE L'ÉCART À LA DROITE* OU PLUTÔT À L'HYPER PLAN S'IL YA PLUS DE DEUX ATTRIBUTS.

DISONS x_1, x_2, \dots, x_n : LE PLAN DE SÉPARATION EST ALORS DE LA FORME :

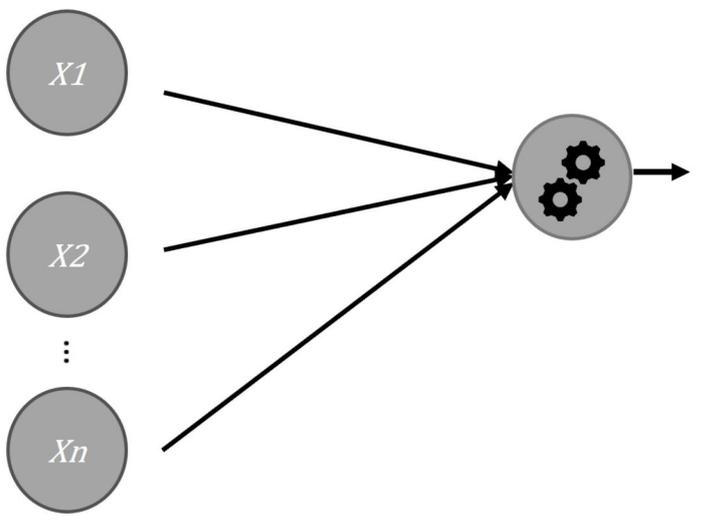
$$\sum_{i=1}^n a_i x_i = b.$$



ON PEUT INTERPRÉTER LE COEFFICIENT a_i COMME L'IMPORTANCE QU'IL ATTRIBUE À L'ENTRÉE CORRESPONDANTE DANS LE CADRE DE LA TÂCHE QU'ON LUI A ASSIGNÉE, ON PEUT DONC REPRÉSENTER SYNTHÉTIQUEMENT UN NEURONE PAR UN PETIT GRAPHE COMME CELUI-CI :



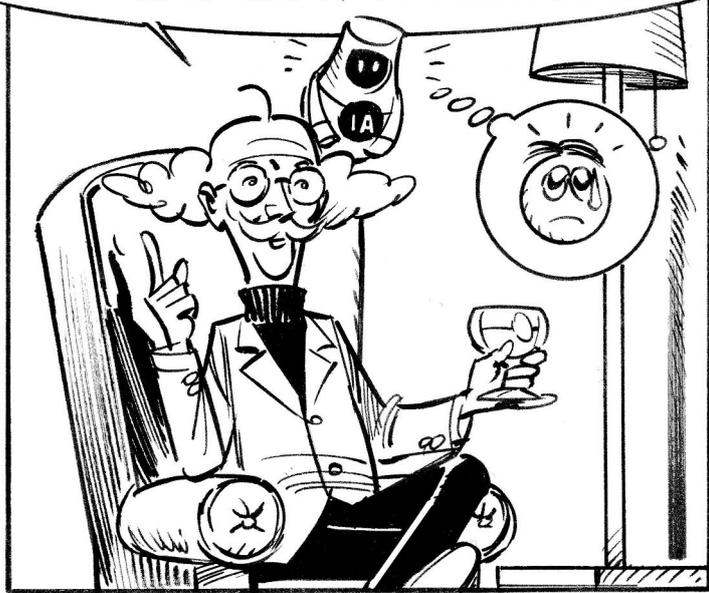
f EST APPELÉE FONCTION D'ACTIVATION.



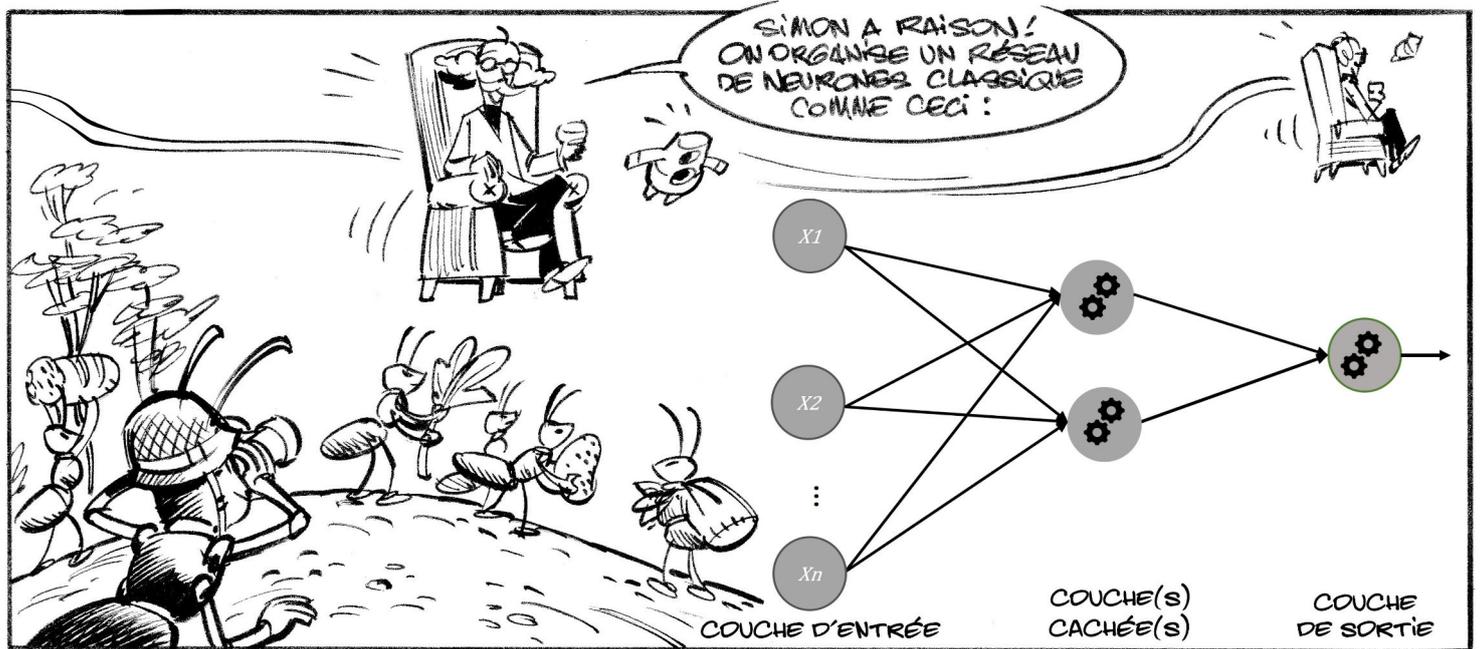
* DANS LE CAS DU PERCEPTRON, LA FONCTION RETOURNE 1 SI ON EST D'UN CÔTÉ DE LA DROITE, 0 SI ON EST DE L'AUTRE.

LES IMAGES SONT GÉNÉRALEMENT TROP COMPLIQUÉES POUR QU'UN SEUL PETIT NEURONE (DONC UN MODÈLE DE CLASSIFICATION LINÉAIRE) PUISSE FONCTIONNER.

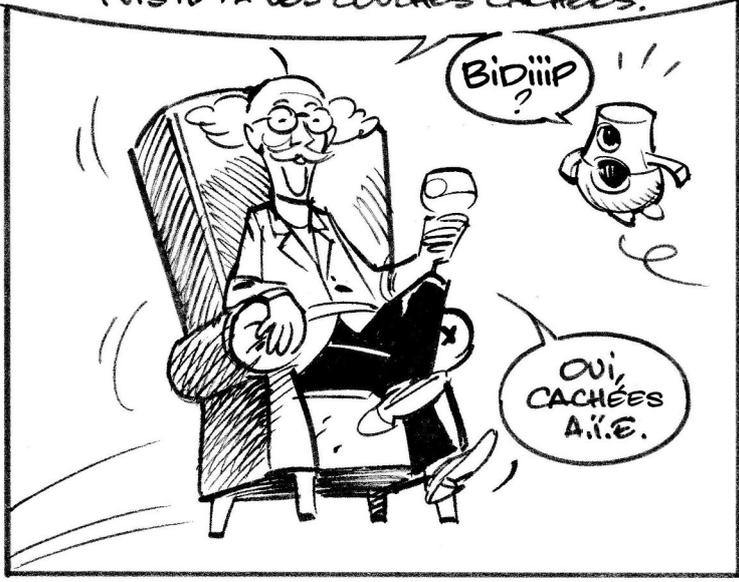
NE T'INQUIÈTE PAS, AÏE! C'EST LÀ QU'ENTRE EN JEU LA NOTION DE RÉSEAU. SI UN SEUL NEURONE A DES CAPACITÉS LIMITÉES, UN TRÈS GRAND NOMBRE DE NEURONES CONNECTÉS LES UNS AUX AUTRES A DES CAPACITÉS ILLIMITÉES. EN THÉORIE HEÏN!



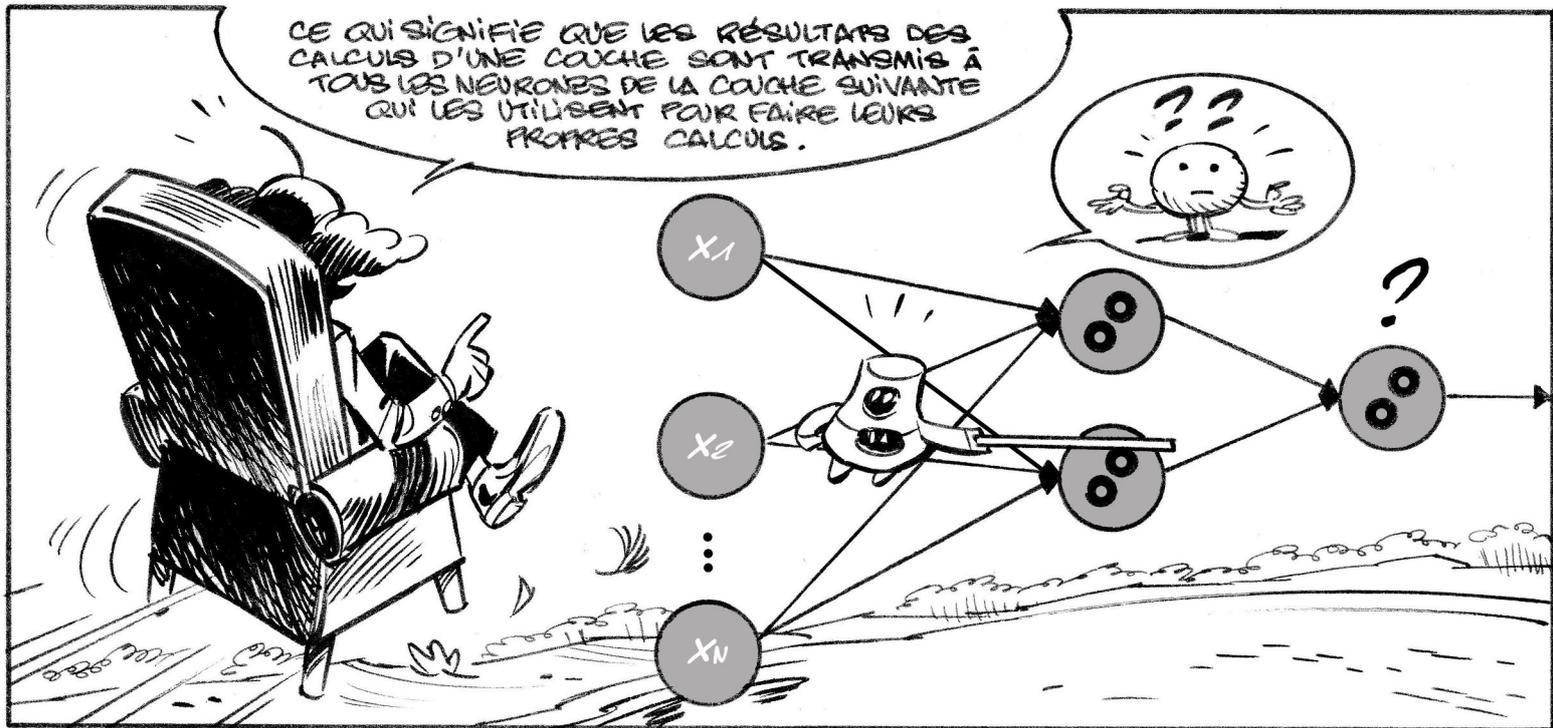
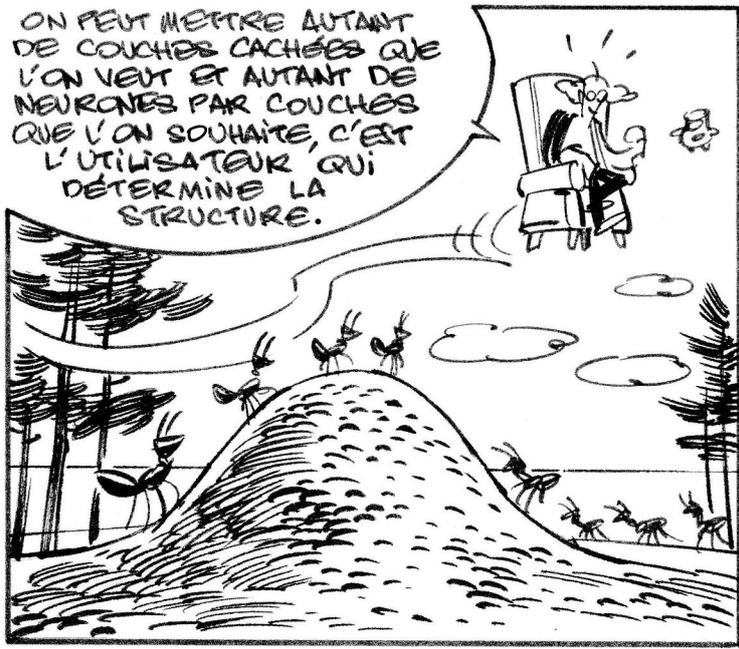
NON TONEN. IMAGINE : ON PEUT DIRE QU'UN NEURONE EST UNE FOURMI CAPABLE, SEULE, DE CHOSSES TRÈS BASIQUES !!!



COMME TU VOIS A.I.I.E, LES NEURONES SONT ORGANISÉS EN COUCHES. IL YA LA COUCHE D'ENTRÉE QUI VA ACCUEILLIR LES DONNÉES D'ENTRÉE, ICI LES VALEURS DE NOS PIXELS. PUIS IL YA LES COUCHES CACHÉES.

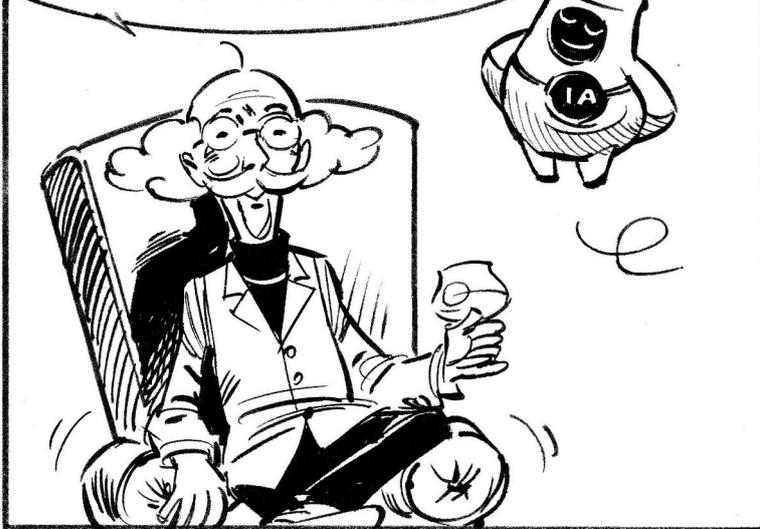


ON LES APPELLE CACHÉES CAR ON NE SAIT PAS EXACTEMENT CE QUI SE PASSE DEDANS.

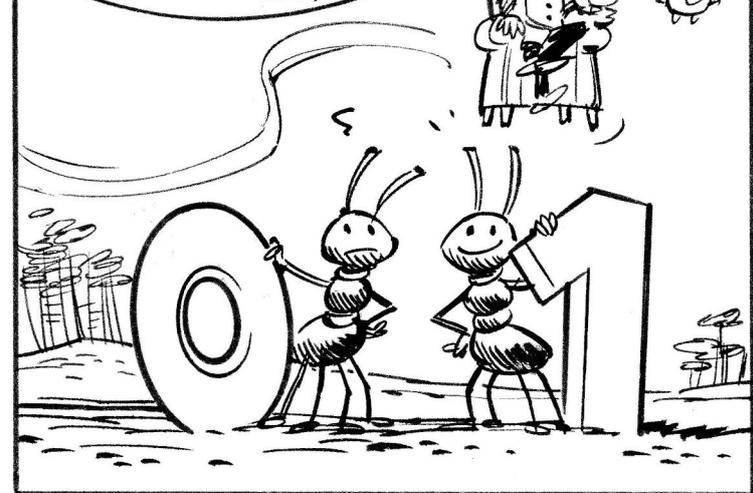


* DANS LES RÉSEAUX LES PLUS SIMPLES.

BONNE QUESTION A.I.E! COMME ON S'ATTACHE ICI A UN PROBLEME DE CLASSIFICATION BINAIRE, ON MET UN SEUL NEURONE SUR LA DERNIERE COUCHE * QUI RENVERRA L'INDICE DE SATISFACTION DE L'IMAGE!!!

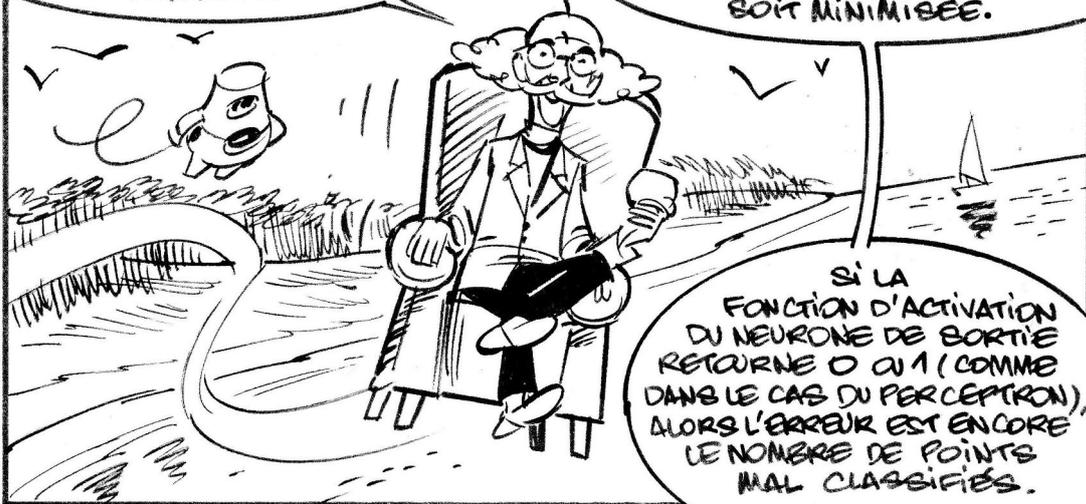


!!! O SI LE MODELE EST CONVAINCU QUE LA PHOTO REPRESENTE UN CHIEN INSATISFAIT ET 1, A L'INVERSE (AVEC TOUTES LES NUANCES POSSIBLES).



UNE FOIS LA STRUCTURE CHOISIE, L'UTILISATEUR DETERMINE QUELLES FONCTIONS D'ACTIVATION UTILISER.

LES COEFFICIENTS VONT ETRE AJUSTES AUTOMATICQUEMENT PAR UN ALGORITHME D'OPTIMISATION POUR QUE L'ERREUR DE PREVISION SOIT MINIMISEE.



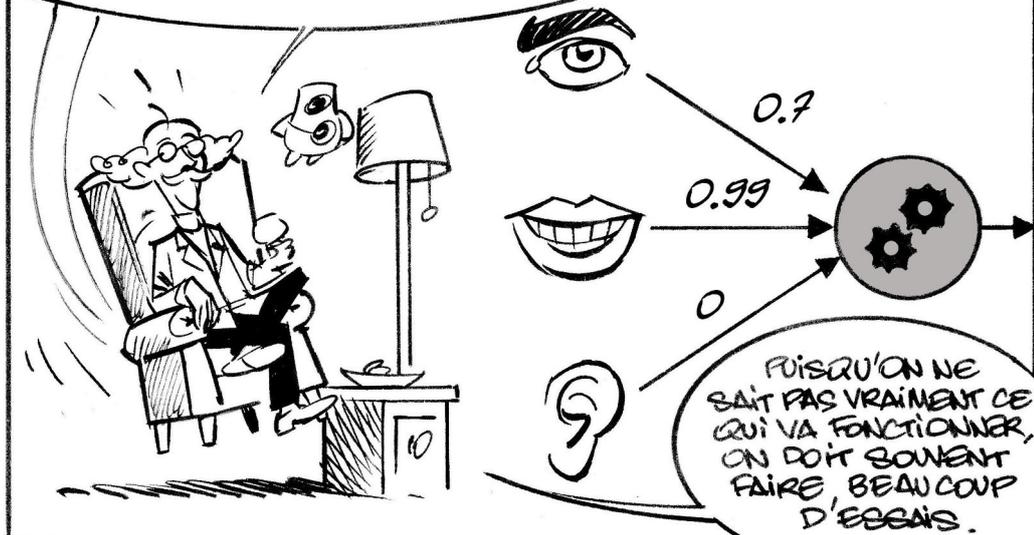
SI LA FONCTION D'ACTIVATION DU NEURONE DE SORTIE RETOURNE 0 OU 1 (COMME DANS LE CAS DU PERCEPTRON), ALORS L'ERREUR EST ENCORE LE NOMBRE DE POINTS MAL CLASSIFIES.



SURP!
MHH

LORSQUE LE RESEAU EST GRAND, ON NE SAIT PAS BIEN INTERPRETER LES TRANSFORMATIONS EN GENERAL. ON PARLE DE MODELE "BOITE NOIRE".

MAIS ON S'ATTEND A CE QUE LE RESEAU DONNE PLUS DE POIDS ET COMBINE LES PIXELS ASSOCIES A UN OEIL RIEUR ET A UNE BOUCHE SOURIANTE POUR DETECTER LES PERSONNES SATISFAITES.



PUISQU'ON NE SAIT PAS VRAIMENT CE QUI VA FONCTIONNER, ON DOIT SOUVENT FAIRE BEAUCOUP D'ESSAIS.

* si ON AVAIT PLUS DE CAS A CLASSER, ON METTRAIT AUTANT DE NEURONES QUE DE CLASSES!



LE LENDEMAIN, À L'OCEAN.

OK TONTON, J'AI MIS LA GOMME, AVEC UN MAXIMUM DE COUCHES ET UN MAXIMUM DE NEURONES PAR COUCHES.

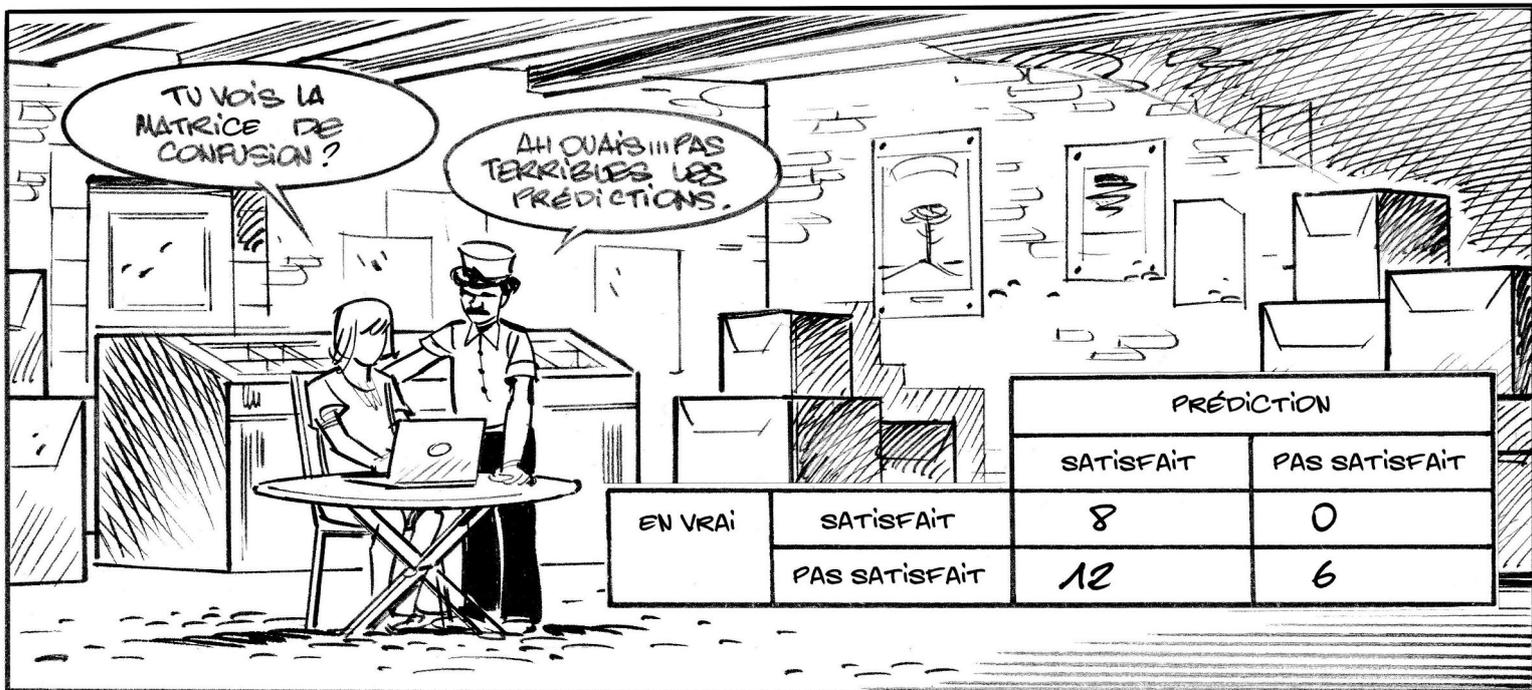
J'AI DIVISÉ LE STOCK D'IMAGES QU'ON VIENT DE PRENDRE EN TROIS : UN JEU D'APPRENTISSAGE, UN JEU DE VALIDATION* ET UN JEU DE TEST.



AH BEN MAHEUREUSEMENT, LÀ C'EST PAS TERRIBLE!!!



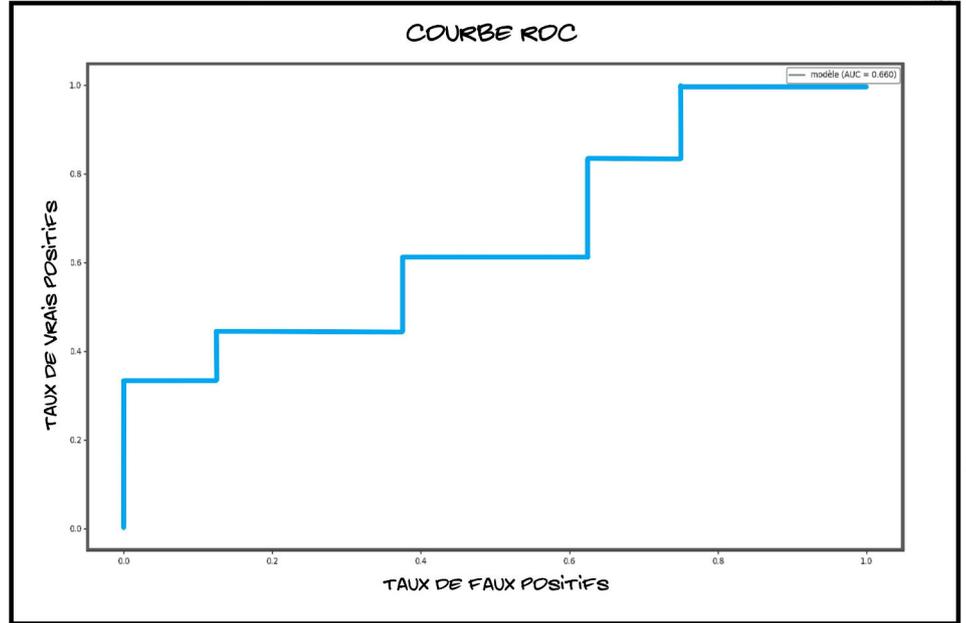
* LE JEU DE VALIDATION SERT À CALIBRER LES PARAMÈTRES DU MODÈLE (NOMBRE DE COUCHES, NOMBRE DE NEURONES PAR COUCHE, FONCTION D'ACTIVATION) AVANT DE TESTER. C'EST UN PEU UN JEU DE "PRÉ-TEST".



ET TU VOIS, LA COURBE ROC ET L'AIRE SOUS LA COURBE AUC, SUR LES DONNÉES DE VALIDATION, CONFIRMENT QUE LE MODÈLE POURRAIT FAIRE MEUX.

POUR DE LA CLASSIFICATION BINAIRE SIMPLE COMME ICI, ET SUR DES IMAGES "TRÈS PROPRES", ON S'ATTEND À UNE AIRE SOUS LA COURBE TRÈS PROCHE DE 1.

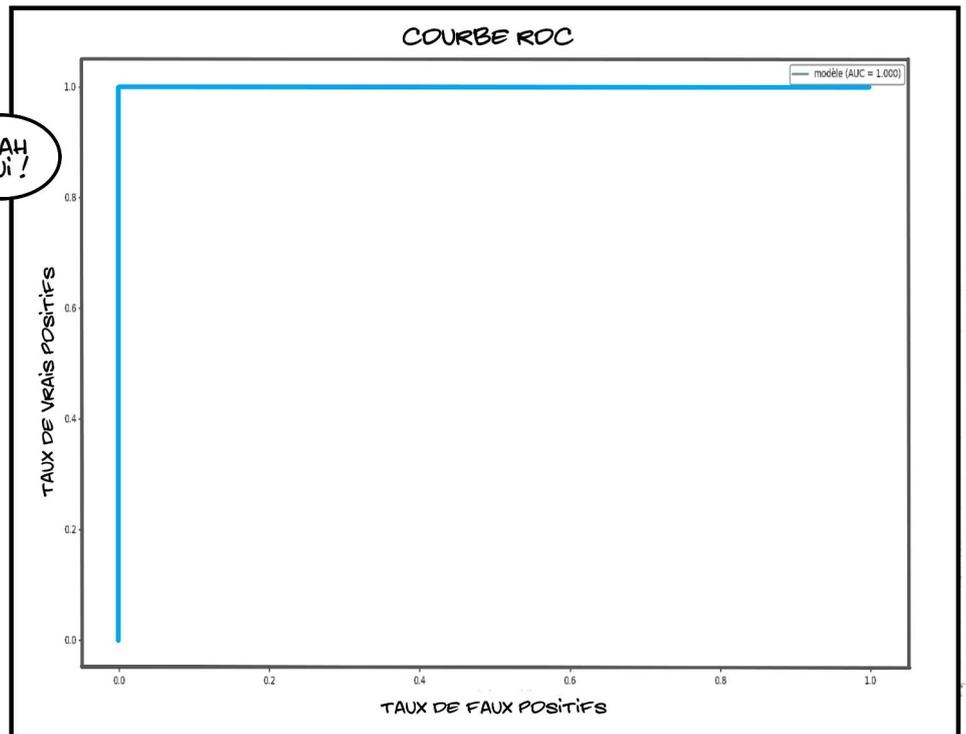
MHH!!!



LA COURBE CI-DESSUS EST ENCORE ASSEZ LOIN DU COIN SUPÉRIEUR GAUCHE, ÇA CONFIRME QUE LE MODÈLE POURRAIT ÊTRE AMÉLIORÉ.

OUI!!! REGARDE SI ON COMPARE À LA COURBE ROC SUR LE JEU D'APPRENTISSAGE.

AAH OUI!

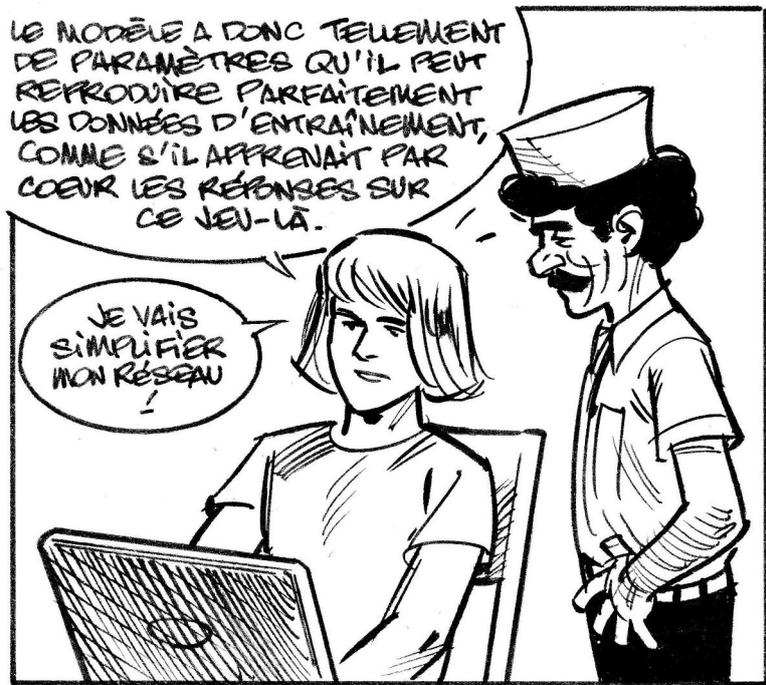


!!! IL A VRAIMENT PARFAITEMENT APPRIS DE CES EXEMPLES.



OUI, PROBABLEMENT PARCE QUE MON RÉSEAU EST TROP "FUISSANT"!! IL A TROP DE NEURONES ET DE COUCHES PAR RAPPORT À LA DIFFICULTÉ DE LA TÂCHE.

LE MODÈLE A DONC TELLEMENT DE PARAMÈTRES QU'IL PEUT REPRODUIRE PARFAITEMENT LES DONNÉES D'ENTRAÎNEMENT, COMME S'IL APPRENAIT PAR CŒUR LES RÉPONSES SUR CE JEU-LÀ.



JE VAIS SIMPLIFIER MON RÉSEAU!

QUELQUES HEURES PLUS TARD...



BON, J'AI BIEN SIMPLIFIÉ MON RÉSEAU !!!

ALORS ÇA DONNE QUOI?

!!! MAIS MAINTENANT, J'OBTIENS UNE PRÉCISION DE 0.58 SUR LE JEU D'ENTRAÎNEMENT ET DE 0.52 SUR LE JEU DE VALIDATION.

AH LÀ, ÇA A L'AIR ENCORE MOINS BON!!! JE CROIS QUE JE SAIS / MAINTENANT IL SOUS-APPREND CAR LES PERFORMANCES SONT ABSEZ PROCHES MAIS MAUVAISES SUR LE JEU D'ENTRAÎNEMENT ET SUR LE JEU DE VALIDATION.



OUI, C'EST ÉNERVANT HEIN? C'EST PARCE QUE MAINTENANT, MON RÉSEAU NE CONTIENT NI ASSEZ DE NEURONES NI ASSEZ DE COUCHES J'AI L'IMPRESSION !!!



AUEZ, AUEZ!

TAP TAP

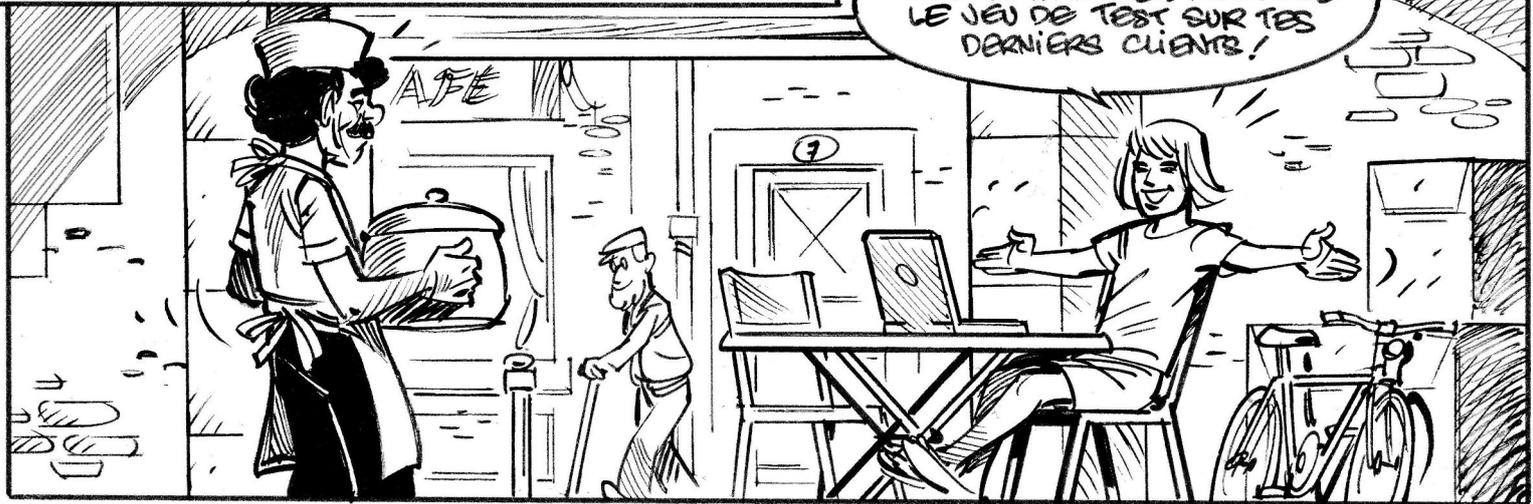
BON, JE VAIS PROGRESSIVEMENT AUGMENTER LE NOMBRE DE COUCHES ET DE NEURONES PAR COUCHES POUR VOIR SI ÇA AMÉLIORE LES CHOSSES SUR MON JEU DE VALIDATION.



ET MOI, JE VAIS TE CONCOCTER MON OSSO BUCO MAISON, ÇA VA TE REQUINQUER!

APRÈS ENCORE PLUSIEURS HEURES D'ESSAIS ET D'ERREURS ...

ÇA Y EST TONTON, CETTE FOIS ÇA MARCHE ! REGARDE LE JEU DE TEST SUR TES DERNIERS CLIENTS !



prédiction = pas satisfait



prédiction = satisfait



prédiction = pas satisfait



prédiction = satisfait



prédiction = satisfait



prédiction = pas satisfait



prédiction = satisfait



prédiction = pas satisfait



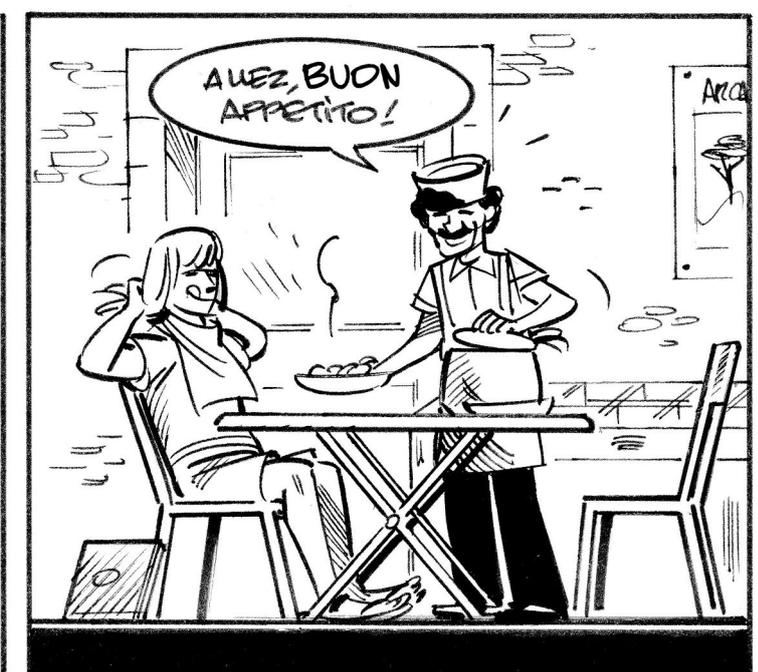
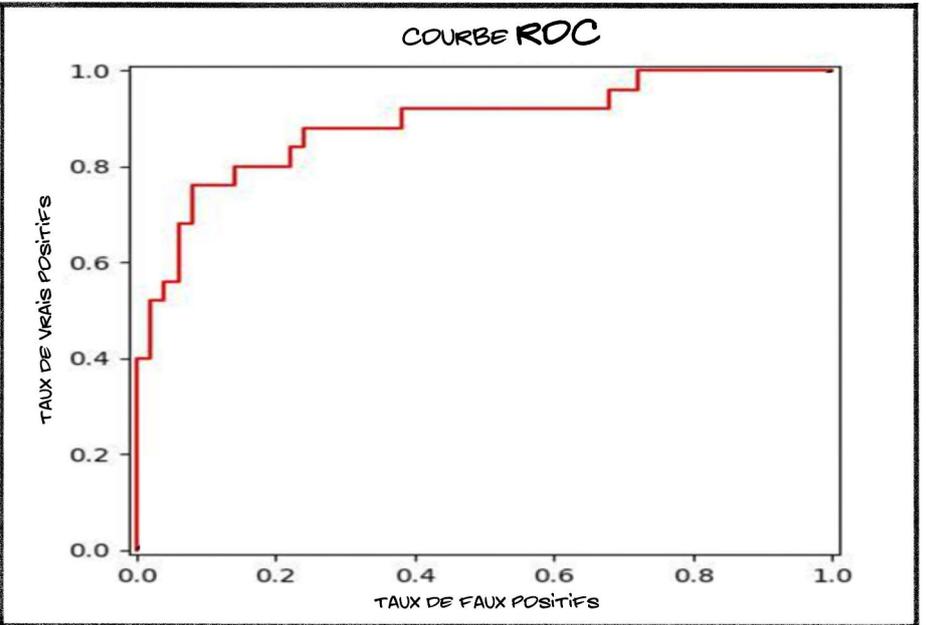
prédiction = satisfait

FORMIDABLE ! QUI JE ME RAPPELE L'AVANT-DERNIÈRE CLIENTE. FAUT DIRE QU'ELLE AVAIT PRIS CHAMALLOW - CAFÉ !



AUEZ, À TABLE ! TU AS VRAIMENT MÉRITÉ DE GOÛTER À MA SPÉCIALITÉ !







LÀ TU VOIS, JE TESTE DE NOUVELLES ASSOCIATIONS DE PARFUMS QUE LES GENS ONT DE FORTES CHANCES D'AIMER. EN PLUS DE LA TÊTE DE MES CLIENTS SATISFAITS OU PAS, JE GARDE DES CHOIX DE PARFUMS QU'ILS ONT FAITS. ÇA ME PERMET D'ESTIMER LE SUCCÈS PROBABLE D'UNE COMBINAISON DE SAVEURS.



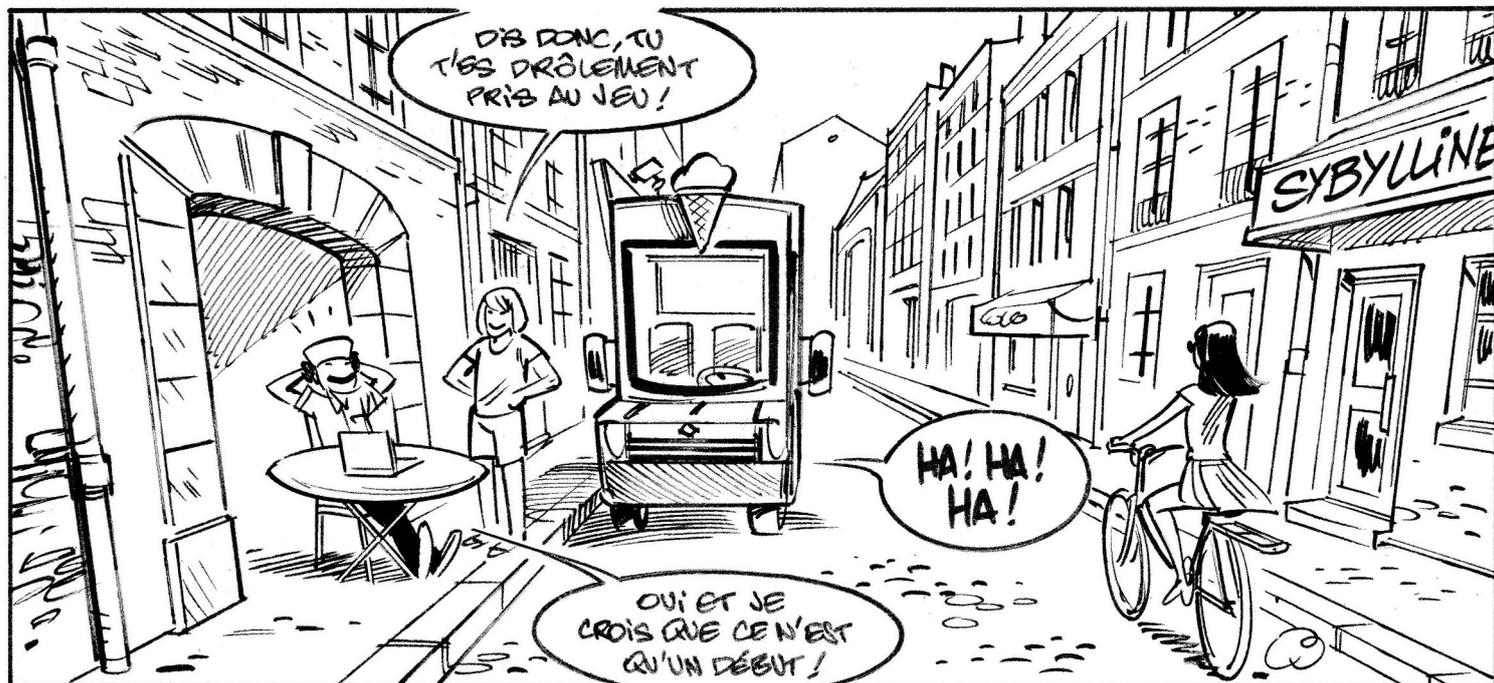
OUI, TU COUP J'EN AI AUSSI PROFITÉ POUR METTRE AU POINT UN PETIT SYSTÈME DE RECOMMANDATION QUI AFFICHE DES SUGGESTIONS SOUS LA CAISSE, EN FONCTION DES CHOIX PASSÉS, HÉ HÉ HÉ !



ET PUIS LÀ, J'AI UN PEU AMÉLIORÉ TES MODÈLES POUR ESTIMER LA QUANTITÉ DE CHAQUE PARFUM QUE JE DOIS COMMANDER À MON FOURNISSEUR EN FONCTION DES VENTES QUE J'ESPÈRE FAIRE.



ANEC TOUT ÇA, MA PETITE AFFAIRE MARCHE BEAUCOUP MIEUX QUE CEUX DE MES CONCURRENTS!



OUI ET JE CROIS QUE CE N'EST QU'UN DÉBUT!

PLONGEZ DANS " SURF, SORBETS ET STATISTIQUES ", UNE BD
RAFRAÎCHISSANTE OÙ LA TRADITION RENCONTRE LA TECHNOLOGIE !

SUIVEZ GIOVANI, VENDEUR DE GLACES AU COEUR GÉNÉREUX, ET SON NEVEU
SIMON, SURFEUR PASSIONNÉ, ALORS QU'ILS INTÈGRENT LES FONDAMENTAUX DE
L'INTELLIGENCE ARTIFICIELLE POUR MODERNISER L'ENTREPRISE DE
GLACES FAMILIALES. AVEC L'AIDE DU PROFESSEUR NIMBUS ET DU PETIT ROBOT
AÏE, ILS SE LANCENT DANS UNE AVENTURE AUSSI AMUSANTE D'INSTRUCTIVE.

DÉCOUVREZ UNE HISTOIRE OÙ CHAQUE RECETTE DE GLACE EST OPTIMISÉE,
CHAQUE VENTE EST PRÉVUE, OÙ LA SATISFACTION DE CHAQUE CLIENT EST
CAPTURÉE, GRÂCE AUX DONNÉES ET AU MACHINE LEARNING.

" SURF, SORBETS ET STATISTIQUES " EST UN COCKTAIL
DE SAVEURS ET DE TECHNOLOGIES À SAVOURER SANS MODÉRATION !

