TREBALL DE FI DE GRAU

Grau en Estadística

Títol: Una visió holística de la Intel·ligència Artificial: Anàlisis estadístic i enfocament ètic.

Autor: Sonia Julià Galindo

Director: Ernest Pons i Fanals

Departament: Departament d'Econometria, Estadística i Economia Aplicada

Convocatòria: Juny 2024

UNIVERSITAT DE BARCELONA

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Facultat de Matemàtiques i Estadística

Resum

Aquest treball és un espai de reflexió sobre la relació entre l'Estadística i la Intel·ligència Artificial, tenint en compte la vessant ètica. Pretén mostrar la necessitat d'una visió holística de la Intel·ligència Artificial, convertint-la en un camp multidisciplinari. L'escrit analitza els algorismes d'aprenentatge automàtic supervisat i els problemes que poden presentar des del punt de vista estadístics, i des d'un punt de vista moral. Finalment es presenta un plantejament ètic, manifestant la necessitat d'una formació ètica pels alumnes d'estadística.

Paraules clau: Intel·ligència Artificial, Aprenentatge automàtic supervisat, Algorisme, Biaix, Error, Moral, Filosofia.

Abstract

This work is a space for reflection on the relationship between Statistics and Artificial Intelligence, considering the ethical aspect. It aims to show the need for a holistic vision of Artificial Intelligence, turning it into a multidisciplinary field. The paper analyses supervised machine learning algorithms and the problems they can present from a statistical point of view, and from a moral point of view. Finally, an ethical approach is presented, showing the need for ethical training for statistics students.

Keywords: *Artificial Intelligence, Supervised Machine Learning, Algorithm, Bias, Error, Morality, Philosophy.*

Classificació AMS: 62-01 Instructional exposition
68T01 Artificial Intelligence, General
68T37 Reasoning under uncertainty
91D10 Models of societies, social and urban evolution
97R40 Artificial Intelligence

INDEX

I. INTRODUCCIÓ

Avui en dia, tothom ha escoltat parlar algun cop la Intel·ligència Artificial, ja sigui des d'una perspectiva de ciència ficció, on la IA ha dominat el món, ja sigui des del punt de vista actual, on per la majoria de la gent la IA es el Chat GPT. La Intel·ligència artificial és la simulació de la intel·ligència humana en màquines.

La Intel·ligència Artificial s'està aplicant ara mateix a molt àmbits, com per exemple la indústria manufacturera, per millorar processos i optimitzar el manteniment; també s'utilitza en banca, per prevenir activitats fraudulentes i atacs cibernètics; i fins i tot en l'àmbit de la sanitat analitzant dades clíniques dels pacients. Veiem doncs que, la IA s'aplica d'una manera transversal en el nostre dia a dia. L'estadística sembla tenir una gran relació amb la IA, ja que l'aprenentatge automàtic i l'aprenentatge automàtic profund son una rama de la Intel·ligència Artificial, i són processos bastats en models i dades, i per tant, pertinents a l'Estadística.

La participació dels estadístics en el desenvolupament de la Intel·ligència Artificial és clau en el desenvolupament d'aquesta tecnologia emergent. Com a tecnologia emergent és necessari analitzar i plantejar tots aquells problemes ètics que la IA pot tenir. Es pot fer des de diversos punts de vista, en aquest treball es farà des d'una perspectiva estadística. Cal analitzar la procedència de les dades, com s'han obtingut o si respecten la privacitat dels individus quan es creen perfils detallats; els biaixos que poden presentar els models entrenats amb dades que estaven esbiaixades, i per tant prenen decisions discriminatòries; els errors de tipus I i tipus II que poden presentar les decisions; tots aquests efectes poden tenir efectes negatius cap a la societat i poden potenciar discriminacions i desigualtats ja existents.

En tot això també caldria tenir en compte que les tècniques estadístiques estan ben aplicades, és a dir les hipòtesis, les distribucions de les dades, els estimadors, etc., tot i això aquest no és l'objectiu del treball. A causa de les limitacions del treball, únicament ens centrarem en analitzar de forma conceptual i general els algorismes d'aprenentatge automàtic supervisat, sense entrar en tots els detalls i possibilitats.

Les aplicacions de la IA també pot tenir greus conseqüències si no s'aplica d'una forma justa, és a dir, es poden crear algoritmes esbiaixats a propòsit, o ser aplicat amb fins il·legítims, per dur a terme estafes o decisions discriminatòries.

Aquest treball es centrarà en analitzar si l'estadística té una relació estreta amb el desenvolupament de la intel·ligència artificial i trobar quins problemes poden presentar els algorismes des d'una perspectiva estadística. També es vol fer un petit anàlisis sobre la importància de l'ètica a l'hora d'implementar els algorismes, i reflexionar sobre si és necessari que els estadístics, com a persones clau en el desenvolupament d'aquesta, requereixen d'una formació ètica. Aquest treball pretén ser un espai de reflexió entre la interacció de intel·ligència artificial i l'estadístic, tenint en compte també, alguns aspectes ètics.

En la primera part del treball s'introduiran alguns conceptes bàsics per tal de contextualitzar el tema. A continuació s'analitzaran els processos d'aprenentatge automàtic supervisat, ja que no és possible analitzar tots els processos d'aprenentatge automàtic i els de aprenentatge automàtic profund. Un cop els algorismes estiguin definits, es passarà a l'anàlisi dels problemes que aquests poden presentar. Finalment, s'aprofundiran els conceptes més ètics, per veure si és necessari que un estadístic en tingui una formació.

II. COS DEL TREBALL

1. DEFINCIÓ Y CONCEPTES BÀSICS

Per tant de contextualitzar el treball i que la seva lectura sigui més senzilla, caldria introduir alguns conceptes abans d'entrar en la discussió.

Agent: Quelcom que actua en un entorn; que fa alguna cosa. Els gossos, els avions, els humans, les formigues, el països i els termòstats, son agents, tant si tenen vida com si no, són agents. Els agents són jutjats per com actuen (Russell & Norvig, 2004, pp 37-38).

Una **agent computacional** és un agent del qual les seves decisions es poden explicar en càlculs, es poden desglossar en operacions primitives que es possible implementar en un dispositiu físic. Tant els homes com les màquines són agents computacionals, però fan el càlcul per prendre decisions de manera diferent, un humà realitza el càlcul en un forma de vida biològica, i un ordinador en "*hardware*" (Poole & Mackworth, 2023).

És difícil dir si tots els agents són intel·ligents, ja que alguns no son computacionals, com ara la pluja o el vent que erosionen un paisatge.

L'objectiu de la intel·ligència artificial és dissenyar agents que actuïn de forma intel·ligent. A més a més, la intel·ligència artificial pretén augmentar la intel·ligència i creativitat humana. Poden ajudar als metges a fer millors prediccions o fer traduccions automàtiques per ajudar a la gents a comunicar-se.

La intel·ligència artificial és un concepte complex de definir. Avui en dia, està en el centre de moltes tecnologies que utilitzem, com ara dispositius intel·ligents o assistents de veu, tot i això encara està en vies de desenvolupament. D'una forma senzilla podem dir que és el camp de desenvolupament de computadores i robots que siguin capaces de comportar-se de maneres que imiten i vagin més enllà de les capacitats humanes, com ara el raonament, l'aprenentatge, la creativitat o la capacitat de plantejar. La IA permet que els sistemes tecnològics percebin el seu entorn, s'hi relacionin i resolguin problemes i actuïn amb una finalitat específica. La comissió europea la defineix com sistemes de software (i possiblement també de hardware) dissenyats per humans que, davant d'un objectiu complex, actuen en la dimensió física o digital (Adán, 2023):

- Percebent el seu entorn, a traves de l'adquisició de dades estructurades o no estructurades
- Raonant sobre el coneixement, processant la informació derivada d'aquestes dades i decidint les millors accions per aconseguit l'objectiu.

Podríem classificar les definicions segons dos conceptes, que són intel·ligència i racionalitat. Definim intel·ligència com la capacitat d'entendre o comprendre i de resoldre problemes, de manera que la intel·ligència d'una màquina es mesura amb termes de fidelitat en la forma d'actuar dels humans. I definim racionalitat com el concepte ideal d'intel·ligència. A més a més, es té en compte si la definició fa referència a la conducta mental de la IA o la seva forma d'actuar. De manera que ens queda aquesta classificació (Russell & Norvig, 2004, p. 2):

1. Sistemes que pensen com humans.
2. Sistemes que pensen racionalment.
3. Sistemes que actuen com humans.
4. Sistemes que actuen racionalment.

Per tant, a més a més, de ensenyar a les màquines a pensar, aprendre i actuar, hi ha un altra dimensió que té a veure en com fem que les màquines facin, és a dir, com podem impartir més capacitat cognitives i sensorials a les màquines, per exemple, reconèixer patrons o analitzar imatges i vídeos. D'alguna manera la IA ha de reemplaçar alguns dels treballs que fan els humans, com per exemple revisió de publicacions a les xarxes socials, per veure quins continguts s'han d'eliminar. Aquests tipus de tasques són massa costoses de fer per humans, i per tant una màquina, no sols imita el comportament que tindria un humà, sinó que ha d'anar més enllà. Per tant, podem dir que la IA és una aplicació de la informàtica per resoldre problemes de forma intel·ligent mitjançant algoritmes, realitza tasques de manera automàtica sense l'ajuda dels éssers humans. La IA extreu conclusions de dades que no són obvies, i que par tant, aprèn patrons d'unes dades i els reprodueix en unes altres, de la mateixa manera que fem els humans, les màquines que tenen intel·ligència artificial aprenen de l'experiència.

Es tracta d'un tipus d'informàtica nou, fins ara els sistemes programables, estaven basats en principis matemàtics, i programats a partir de regles i lògica destinats a derivar respostes matemàtiques precises, seguint un rígid enfoc d'arbre de decisions. Actualment la gran quantitat de dades provoca que les decisions que s'han de prendre en funció d'aquestes siguin molt més

complexes, i per tant un enfocament rígid no aconsegueix mantenir-se al dia amb la informació. La computació cognitiva permet trobar patrons i respostes entre aquests volums tant grans de dades. Aquestes dades de les que aprèn poden ser estructurades, no estructurades o semiestructurades. Fem una definició bàsica per cada un d'ells (Education, I. S., 2020).

- Dades estructurades: són aquelles dades que es troben ordenades. Per exemple taules d'excel, fulles de càlcul, o bases de dades que es troben organitzades en files i columnes. Son les que se solen utilitzar en la major part de bases de dades relacionals, i es gestionen mitjançant SQL (Structured Query Language)

- Dades semiestructurades: tenen un nivell mitjà d'estructuració i rígides organitzativa. Tenen una mica d'estructura, jerarquia i organització, però no tenen esquema fix. Contenen metadades que s'utilitzen per agrupar i descriure com s'emmagatzemen. Tenen algunes propietats organitzatives que faciliten el seu anàlisis, i si es processen es poden arribar a emmagatzemar en bases de dades relacional i tenir files i columnes.

- Dades no estructurades: son el 80% de les dades existeixen en qualsevol organització i son molt complicades de tractar i analitzar. No es poden utilitzar en bases de dades tradicionals, ja que seria impossible ajustar-les en files i columnes estandarditzades. Poden ser imatges, vídeos, dades de xarxes, informes, etc.

Cada tipus de dada es tractada d'una forma diferent per tal d'extreure'n informació. Les dades estructurades donen una visió general, però les no estructurades poden donar una compressió molt més profunda i poden descobrir-se patrons. Les dades no estructurades s'analitzen mitjançant diferents tipus d'IA, que són l'aprenentatge automàtic o *machine learning* i l'aprenentatge automàtic profund o *deep learning*,. Més endavant definirem aquests conceptes. Aquests sistemes cognitius segueixen el mateix procediment que els humans per entendre i prendre decisions. En primer lloc observem els fenòmens i les proves visibles. En segon lloc, basant-nos en el que sabem i les experiències viscudes interpretem i generem hipòtesis sobre el significat. En tercer lloc, duem a terme una avaluació sobre si les hipòtesis són correctes o incorrectes. Finalment, escollim aquella opció que creiem millor en conseqüència de tot l'anàlisi. Les màquines poden dur aquest procés a gran escala i més ràpid, de manera que s'optimitza la feina. Aquestes dades no estructurades són ambigües, implícites i complexes d'analitzar. Un àudio d'una persona parlant, pot ser tot un repte d'analitzar en funció del dialecte, l'entonació o les expressions semàntiques, entre molts altres factors que poden alterar

el significat d'una frase. Analitzen això com una persona, i per tant, poden estar també subjectes al context, a la informació donada, etc. Per tant, les principals diferències entre els sistemes de computació cognitiu i els sistemes convencionals són:

1. Llegeixen i interpreten dades no estructurades, comprenent no només el significat de les paraules, sinó també la intenció i el context.

2. Raonen sobre els problemes com els humans i prenen decisions.

3. Aprenen amb el temps de les seves interaccions amb els humans i es van fent intel·ligents.

A més a més, cal afegir un concepte que serà necessari conèixer, és el concepte de biaix. En l'ésser humà es coneix com biaix cognitiu, i és l'alteració del processament d'informació, que provoca que les persones distorsioni la seva interpretació de la realitat. D'una forma més tècnica es pot definir com la tendència a preferir una hipòtesis sobre un altra. Els algorismes presenten algunes preferències inherents cap a certs valors o preferències, aquests normalment venen incorporats a causa de les dades que se li han introduït. D'aquí sorgeixen diferents qüestions com si es possible eliminar-los? Es poden especificar? Son transparents per als usuaris? (Poole & Mackworth, 2023)

2. HISTÒRIA DE LA INTEL·LIGÈNCIA ARTIFICIAL

Els orígens de la IA es poden rastrejar fins a les especulacions filosòfiques sobre la naturalesa de la ment i la possibilitat de crear màquines pensants. Alan Turing, un dels pares fundadors de la IA, va abordar aquestes qüestions en el seu article seminal "Computing Machinery and Intelligence" (1950), on proposava el famós test de Turing com una manera de determinar si una màquina pot exhibir comportament intel·ligent equivalent al d'un humà. Turing, i altres pioners com John McCarthy, Marvin Minsky, i Herbert Simon, van establir les bases de la IA com una disciplina acadèmica, imaginant màquines capaces de resoldre problemes, aprendre i fins i tot tenir una certa comprensió del món.

En la seva evolució, la IA ha adoptat diverses tècniques i models per millorar la seva capacitat de presa de decisions i predicció. Entre aquests, els models de Bayes i Markov han tingut un paper fonamental.

El teorema de Bayes, una fórmula matemàtica utilitzada per actualitzar les probabilitats a la llum de nova informació, és crucial en molts aspectes de la IA, especialment en el camp de l'aprenentatge automàtic (machine learning). Els models bayesians permeten la inferència probabilística, ajudant a les màquines a prendre decisions basades en l'evidència disponible. Aquests models són útils en aplicacions com la detecció de correu brossa, el reconeixement de veu, i la diagnosi mèdica, on les decisions s'han de prendre en contextos d'incertesa.

Els processos de Markov i les cadenes de Markov són eines fonamentals per modelar sistemes que evolucionen de manera probabilística al llarg del temps. Una cadena de Markov és un model matemàtic que descriu un sistema que es mou d'un estat a un altre, amb la propietat que la probabilitat de transició només depèn de l'estat actual, no de la seqüència d'estats anteriors. Aquest concepte és essencial en moltes aplicacions de la IA, incloent el processament de llenguatge natural, on les cadenes de Markov ocultes (HMM) s'utilitzen per tasques com l'etiquetatge de parts del discurs i la reconeixença de patrons en dades temporals.

En els darrers anys, el desenvolupament de l'aprenentatge automàtic i de l'aprenentatge automàtic profund ha transformat radicalment el camp de la IA. L' aprenentatge automàtic, una branca de la IA centrada en la construcció d'algoritmes que poden aprendre a partir de dades, ha permès avenços significatius en àrees com la visió per computador, el reconeixement de veu i la traducció automàtica. Dins del aprenentatge automàtic, l'aprenentatge automàtic profund, que utilitza xarxes neuronals profundes amb múltiples capes, ha permès la creació de models capaços de superar els humans en tasques complexes com el reconeixement d'imatges i la comprensió del llenguatge natural.

La filosofia no només ha influït en els fonaments teòrics de la IA, sinó que també continua sent vital en l'exploració de les seves implicacions ètiques i epistemològiques. Els filòsofs han plantejat qüestions importants sobre la naturalesa de la consciència i la identitat personal en un món on les màquines poden imitar comportaments humans de manera cada cop més precisa. Això porta a debats sobre si les màquines poden realment "pensar" o si simplement simulen la intel·ligència humana.

L'augment de l'ús de la IA en una àmplia gamma d'aplicacions ha plantejat importants qüestions ètiques. Els sistemes d'IA sovint prenen decisions que tenen impactes significatius en les persones i en la societat, des de sistemes de recomanació de continguts fins a vehicles autònoms.

Cal abordar qüestions sobre la responsabilitat, la transparència i la justícia en el disseny i ús de la IA per garantir que els seus beneficis es distribueixin de manera justa i equitativa.

Malgrat els reptes i les incerteses, el futur de la IA promet innovació i impacte continuats en tots els aspectes de la nostra vida. L'aplicació de la IA en àmbits com la medicina, la indústria, la mobilitat i l'entreteniment continuarà transformant la manera com vivim i treballem. No obstant això, és crucial abordar els reptes ètics, legals i socials associats amb el desenvolupament i desplegament de la IA per garantir un futur sostenible i inclusiu.

La història de la IA és una narrativa fascinant que reflecteix la complexitat de la ment humana, el poder de la tecnologia i els reptes de la societat. Des dels seus inicis filosòfics fins als avanços tecnològics actuals, la IA ha captivat la nostra imaginació i ha generat esperança, expectativa i preocupació. Mentre ens endinsem en un futur cada cop més dominat per la IA, és crucial mantenir un diàleg ric i inclusiu sobre els seus impactes i implicacions, assegurant que la intel·ligència artificial serveixi els interessos humans i promogui el benestar col·lectiu.

La filosofia continua sent fonamental en aquest diàleg, oferint una llum crítica sobre les qüestions ètiques, epistemològiques i ontològiques que planteja la IA. En aquesta intersecció entre la filosofia i la tecnologia, podem trobar una comprensió més profunda de la nostra pròpia naturalesa, així com una visió més clara del futur que estem creant.

3. TIPUS DE INTEL·LIGÈNCIA ARTIFCIAL

3.1. IA Forta vs IA Dèbil

Segons la força, l'amplitud i l'aplicació, la Intel·ligència artificial es pot classificar com IA dèbil i IA forta. La IA dèbil s'aplica a un domini específic, per exemple els traductors d'idiomes, els assistents virtuals, els vehicles autònoms o els filtres intel·ligents de *spam*. Aquest tipus de IA pot realitzar tasques específiques, prendre decisions basades en algoritmes programats i dades d'entrenament, però no pot aprendre taques noves. Aquest tipus d'IA ajuda al ésser humà en algunes tasques, milloren la productivitat i efectivitat del home, però no el superen.

La IA forta o generalitzada pot interactuar i executar una gran quantitat de tasques que son independents i no relacionades entre elles. El caràcter d'aquest tipus d'IA és com a mínim el d'un humà. Pot aprendre noves tasques per resoldre nous problemes i, n'aprèn ensenyant-se a ella mateixa noves estratègies. Podria fer judicis, comunicar mitjançant llenguatge natural, etc. La IA generalitzada és la combinació de moltes estratègies de IA que aprenen de la experiència

i poden funcionar a nivell d'intel·ligència humana. Alguns exemples de IA forta son l'aprenentatge autonòmic (*machine learning i deep learning*), la robòtica autònoma i els sistemes experts.

La capacitat de la intel·ligència artificial es mesura mitjançant el **Test de Turing**, que va ser dissenyat el 1950 per Turing però que encara s'utilitza avui en dia. El test consisteix en un joc d'imitació, on l'interrogador pregunta a un sistema via Interface, si l'interrogador no és capaç de distingir entre les respostes d'una persona i les del sistema, significa que el sistema és intel·ligent (Poole & Mackworth, 2023).

En els següents apartats es farà una definició més tècnica de la Intel·ligència artificial.

3.2. Sistemes experts

Un sistema expert és un tipus de programa d'ordinador que utilitza coneixement i raonament per a resoldre problemes complexos en un domini específic, de manera similar a com ho faria un expert humà. Aquests sistemes són una branca important de la intel·ligència artificial (IA) i es basen en una àmplia col·lecció de regles derivades del coneixement d'experts humans (InnovaciónDigital360, 2024) .

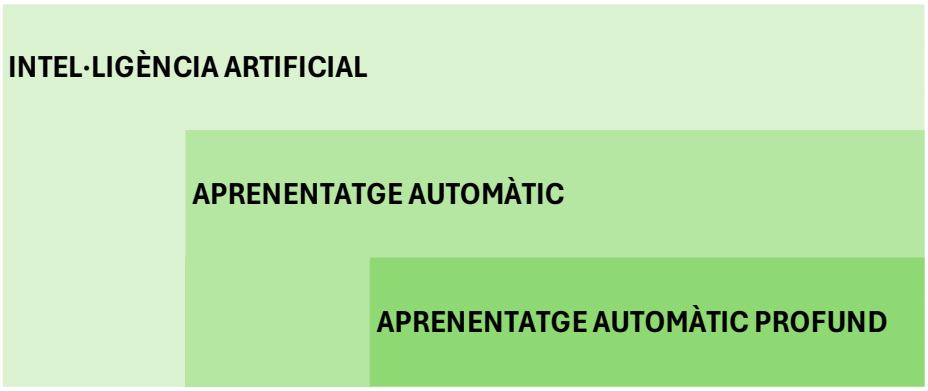
El component principal d'un sistema expert és la seva base de coneixement, que conté fets i regles relacionades amb el domini d'aplicació. Aquestes regles solen ser de la forma "si-a, llavors-b", on "a" representa una condició i "b" una acció o conclusió. Per exemple, en un sistema expert mèdic, una regla podria ser: "Si el pacient té febre alta i tos persistent, llavors pot tenir pneumònia". Aquestes regles permeten al sistema deduir noves informacions a partir dels fets coneguts.

Un altre component clau és el motor d'inferència, que és responsable de processar les regles i fets de la base de coneixement per a arribar a conclusions o recomanacions. El motor d'inferència utilitza tècniques de raonament com ara el raonament endavant (*forward chaining*) i el raonament endarrere (*backward chaining*) per a explorar les possibles solucions a un problema.

Els sistemes experts tenen diverses aplicacions en àrees com la medicina, l'enginyeria, les finances i l'assessorament legal. A més d'oferir recomanacions precises, poden millorar la consistència en la presa de decisions i augmentar l'eficiència en la resolució de problemes.

No obstant això, també tenen limitacions. La seva efectivitat depèn en gran mesura de la qualitat i l'abast del coneixement que han adquirit. A més, els sistemes experts sovint troben dificultats en situacions que requereixen coneixements fora del seu domini específic o que necessiten raonaments més creatius (Universidades, S., 2024).

Els sistemes experts són la forma més senzilla d'Intel·ligència artificial. Dintre d'aquesta hi ha l'aprenentatge automàtic, com un camp més específic, i un pas encara més enllà s'hi troba l'aprenentatge automàtic profund. El següent esquema és una visualització senzilla per fer-ho més entenedor.



Il·lustració 1. Tipus d'Intel·ligència Artificial

A continuació es procedeix a la explicació de l'aprenentatge automàtic, i de l'aprenentatge automàtic profund.

3.3. Aprenentatge automàtic

Una gran quantitat d'avenços de la IA han estat tingut lloc en el subcamp de l'aprenentatge automàtic. L'aprenentatge és la capacitat d'un agent de millorar el seu comportament basant-se en les seves experiències i coneixements. L'aprenentatge no és un fi en si mateix, en el cas de l'aprenentatge automàtic, l'objectiu és que l'agent, a partir de dades, comprengui i analitzi el món, trobi patrons o grups socials, per tal de poder decisions més encertades. Es pretén, que a partir de l'aprenentatge les màquines actuïn intel·ligentment.

Un problema d'aprenentatge consta d'una tasca, unes dades i una mesura de millora. La tasca és el comportament o el procediment que s'està millorant. Les dades són el que dona experiència al agent, poden ser una bossa d'exemples, en cas que no importi l'ordre, o una seqüència temporal. La mesura de millors és com es mesura les noves habilitats o coneixements

que ha adquirit l'agent, pot ser la precisió de la predicció, la velocitat o d'altres (Poole & Mackworth, 2023).

L'aprenentatge automàtic es pot classificar en tres tipus: l'aprenentatge supervisat, l'aprenentatge no supervisat i l'aprenentatge de reforç. Cadascun té unes finalitats i tècniques diferents.

En l'**aprenentatge supervisat** es basa en un grup de dades prèviament etiquetades. La màquina ha d'aprendre a interpreta una funció f, això significa que tenim unes dades d'entrada i unes dades de sortida. Per exemple, donada una bossa de dades d'entrenament T, la màquina ha d'aprendre el seu valor en el domini de la funció f :

$$T=\{\langle x_1,f(x_1)\rangle,\langle x_2,f(x_2)\rangle,\ldots,\langle x_n,f(x_n)\rangle\}$$

Dit d'un altra manera es proporcionen a l'algorisme dades que estan etiquetades amb el seu destí (amb la resposta correcta). I llavors, s'utilitzarà el model d'aprenentatge automàtic per a predir aquesta resposta en dades de les quals no coneix la resposta. A més a més, se li dona al algorisme les característiques que s'utilitzen per identificar els patrons i predir les respostes destí. Per exemple, si volem predir els pesos de les persones les dades d'entrenament seria donar les altures, el sexe i la edat, i dir a quin pes corresponen (Bringsjord, & Govindarajulu, 2024). Visualment seria així:

PES (kg)	ALTURA (cm)	EDAT	SEXE
70	175	19	Home
65	168	20	Dona
86	169	35	Dona
77	180	45	Home
55	160	15	Dona
93	175	27	Home

Taula 1. Exemple de dades etiquetades

De manera que per cada pes, que és la variable que es vol predir, tenim assignades unes característiques. L'algorisme aprendrà aquest valors, que un hom de 19 anys que mesura 175cm pesa 70 kg, que una dona de 20 anys que mesura 168 cm pesa 65 kg, i així amb gran quantitat de dades, tenir en compte molts cassos possibles. Així doncs, un cop ha aprés totes aquestes dades, si li donem una altura, una edat i un sexe, serà capaç de dir-nos quin pes hauria de tenir aquesta persona. En aquest cas, $x_1= \langle 175, 19, \text{Home}\rangle$ i $f(x_1) = 70$.

Un exemple d'aprenentatge supervisat seria com l'algoritme del correu classifica els correus en "brossa", "important" o els grups que s'escullin. Per dur a terme la tasca, l'algorisme revisa el correu i el pot classificar entre un nombre limitat d'opcions.

La supervisió de l'aprenentatge supervisat es presenta en funció del valor f(x) en diversos punt de X en alguna part del domini de la funció. Si diem que h és la funció apresa per l'agent, l'objectiu es que hi hagi la mínima diferencia entre f i h. L'error es defineix com:

$$\text{error}=\sum_{x\in T}\delta(f(x)-h(x))$$

sobre les dades d'entrenament T.

Seguint amb l'exemple del pes, si per un home de 19 anys que mesura 175 cm, l'algorisme ens prediu que ha de pesar 73kg i el valor desitjat és 70kg, l'error de x_1 seria 73-70= 3. I per veure l'error de l'algorisme es sumarien les diferencies de $x_1, x_2, x_3, \ldots x_n$.

L'**aprenentatge no supervisat** està basat en dades no etiquetades, de manera que tenim unes dades d'entrada $\{x_1,x_2,\ldots,x_n\}$, però no unes dades de sortida. No hi ha cap funció que aprendre, sinó que l'agent ha de trobar patrons o informació que es troba oculta entre les dades. Els algorismes d'aprenentatge automàtic no supervisat descobreixen patrons de dades o agrupacions ocultes sense la necessitat d'intervenció humana. Aquests agents tenen la capacitat de trobar similituds i diferències en la informació. S'utilitza principalment per tres tasques: agrupació de clústers, associació i reducció de dimensionalitat. S'utilitza per exemple en la detecció d'anomalies en compres amb targeta de crèdit, on els bancs utilitzen algorismes per detectar patrons de comportament esperats en el client, i intentar detectar desviacions de l'estàndard (Bringsjord & Govindarajulu, 2024).

En l'**aprenentatge per reforç** hi ha un agent que interactua amb l'entorn constantment, de manera que actuï i percebi, i ocasionalment rep retroalimentació del seu comportament. En aquest tipus d'aprenentatge s'imita el procés d'aprenentatge per prova i error que utilitzen els humans. L'agent té un nombre possible d'opcions que pot prendre, quan pren una decisió i actua, si és la decisió correcta es recompensat, altrament és penalitzat. A partir d'aquestes recompenses i càstigs aprenen quines són les millors rutes de processament per aconseguir els millors resultats. Un exemple d'aprenentatge per reforç són els videojocs, on es té un entorn ja programat on se simula un ambient i en el que tenen llocs diferents esdeveniments al mateix temps (Aprendizaje por Refuerzo, 2020).

La principal diferència entre el l'aprenentatge automàtic supervisat i el no supervisat, és que el primer necessita que les dades estiguin prèviament etiquetades, necessiten més intervenció humana que els d'aprenentatge no supervisat. Aquest últim no requereix que les dades estiguin prèviament etiquetades, sinó que buscarà patrons ocults entre totes les dades. L'aprenentatge per reforç normalment és utilitzat per dur a terme l'entrenament d'alguns algorismes, no tant per fer les prediccions.

D'alguna manera podem dir que l'aprenentatge supervisat entén el que fa a partir de la intervenció humana, ja que se li dona les dades objectiu, en canvi, en l'aprenentatge no supervisat, l'algorisme entén el que fa per sí mateix, sense intervenció humana.

3.4.Aprenentatge automàtic profund

L'aprenentatge automàtic profund és una subàrea de l'aprenentatge automàtic que es basa en xarxes neuronals artificials amb múltiples capes, conegudes com a xarxes neuronals profundes. Aquestes xarxes tenen la capacitat d'aprendre representacions complexes de dades mitjançant el processament de grans volums d'informació.

Les xarxes neuronals profundes estan inspirades en l'estructura del cervell humà, on les neurones estan connectades per sinapsis. En el context del aprenentatge automàtic profund, una xarxa neuronal està composta per una sèrie de capes: una capa d'entrada, diverses capes ocultes i una capa de sortida. Cada capa està formada per nodes o neurones artificials, i cada connexió entre neurones té un pes que es va ajustant durant l'entrenament per minimitzar l'error en les prediccions del model.

El procés d'aprenentatge en les xarxes neuronals profundes es realitza mitjançant l'algoritme de descens de gradient, que ajusta els pesos de les connexions en funció de l'error entre les prediccions del model i els valors reals. Un tipus especial d'aquest algoritme és el descens de gradient estocàstic, que actualitza els pesos basant-se en mostres aleatòries del conjunt de dades d'entrenament, permetent una convergència més ràpida i evitant minimitzacions locals.

Una característica clau del aprenentatge automàtic profund és la seva capacitat per a realitzar aprenentatge no supervisat i supervisat. En l'aprenentatge supervisat, el model és entrenat amb dades etiquetades, mentre que en l'aprenentatge no supervisat, el model intenta descobrir patrons subjacents en dades sense etiquetar. Això fa que el aprenentatge automàtic sigui

especialment útil en àrees com el reconeixement d'imatges, el processament del llenguatge natural i la detecció de fraus.

El aprenentatge automàtic ha demostrat un rendiment superior en moltes tasques, gràcies a la seva capacitat per a aprendre característiques jeràrquiques i abstractes directament a partir de dades brutes. No obstant això, també requereix una gran quantitat de dades i recursos computacionals per a entrenar els model.

4. TÈCNIQUES ESTADÌSTIQUES DE L'APRENETATGE AUTOMÀTIC SUPERVISAT

Com s'ha vist abans, l'aprenentatge automàtic supervisat es caracteritza per tenir unes dades d'entrada i unes dades objectiu o de sortida. Un *input* o *feature* (una característica) és una funció de exemples a un valor. Si t és un exemple, i f és una característica, f (t) és el valor de la característica f. El domini d'una característica les el conjunt de valors que es poden tornar. Una característica booleana és aquella que el domini és {*fals, vertader*}.

Per tant, per entrenar l'algorisme es necessita:

- Unes característiques d'entrada $X_1,\ldots,X_n$,
- Un conjunt de característiques objectiu $Y_1,\ldots,Y_k$
- Una bossa d'exemples d'entrenament on cada exemple t és un parell (x_n,y_n), on $x_t=(X1(t),\ldots,Xn(t))$ és un vector d'un valor per cada característica d'entrada i $y_t=(Y1(t),\ldots,Yk(t))$ és un vector d'un valor per cada característica de sortida.

La sortida de l'algorisme és un predictor, és a dir, una funció que prediu les Y de les X, de manera, que l'objectiu és predir valors de les característiques objectius per a dades exemple que l'agent no ha vist.

Hi ha dos tipus d'aprenentatge supervisat, la **regressió** i la **classificació**. Es parla de regressió quan el domini de l'objectiu és un subconjunt dels nombres reals, és a dir, es pretén obtenir un valor numèric. Així doncs, l'agent genera un model en funció de les dades d'exemple. En la classificació el domini de l'objectiu és un conjunt de finit fixes, per exemple, talles de roba (XS,S, M,L...). La classificació pretén identificar a quina categoria pertany una mostra del problema, entre un nombre limitat de categories (Poole & Mackworth, 2023).

Hi ha diverses tècniques estadístiques utilitzades en el machine learning, algún son particulars de la regressió, i algunes són particulars de la classificació pero moltes es poden utilitzar en els dos tipus d'aprenentatge supervisat.

Tècnica	**Regressió**	**Classificació**
Regressió lineal	X	
Regressió logística		X
Màquines de vectors de suport (SVM)	X	X
Arbres de decisió	X	X
Boscos aleatoris	X	X
Naive Bayes		X
k-NN		X

Taula 2. Tècniques d'aprenentatge automàtic supervisat

4.1. Regressió lineal

La regressió lineal s'utilitza únicament en tasques de regressió. Consisteix en ajustar una funció lineal a un conjunt de dades d'entrenament, on les dades d'entrada i les dades objectiu son nombres reals. Si suposem que les característiques d'entrada (*feature inputs*) $X1,\ldots,Xn$, són totes nombres reals i hi ha una única característica objectiu Y, una funció lineal té la forma següent:

$$\widehat{Y_W}\ (t) = w_0 + w_1 * X_1(t) + \ldots + w_m * X_m(t) = \sum_{i=0}^{n} (w_i * X_i(t)),$$

On $w=\langle w_0, w_1, \ldots, w_n \rangle$ és el vector de pesos, i X és la característica que es vol mesurar, que sempre és igual a 1.

Per mesurar l'error d'una regressió lineal s'utilitza **error quadràtic mitjà**. De manera que si tenim les dades d'entrenament T, l'error per la predicció $\hat{Y}$ es calcula de la següent manera:

$$ECM(T, w) = \frac{1}{T} \sum_{t \in T} (\widehat{Y_W}\ (t)\ -\ Y_t)$$

Per tant, l'objectiu de l'agent o de la màquina és trobar la recta que minimitzi aquest error, passa a ser un problema d'optimització. Per fer-ho es busca minimitzar l'error quadràtic mitjà,

i existeix un mínim únic, que es troba amb les derivades parcials respecte els pesos quan tots són zero. Aquesta derivada parcial respecta al pes w_i és:

$$\frac{\partial}{\partial w_i} error\ (T, w)\ =\ \frac{1}{|T|}\sum_{t \in T} 2 * \delta(T)\ *\ X_i(t)$$

De manera que $\delta(T)\ =\ (\widehat{Y_W}\ (t)\ -\ Y_t)$, una funció lineal dels pesos. Per calcular analíticament quins son els pesos que minimitzen l'error, s'igualen les derivades parcials a zero i es resolen les equacions resultants (Poole & Mackworth, 2023) .

4.2. Funcions logística

Per tasques de classificació la regressió lineal no és suficient. Si considerem una variable objectiu binaria, i que per tant, el seu domini és {0,1}, veiem que no es pot predir la resposta mitjançant una regressió lineal. Pot ser per exemple, per saber si un usuari que navega per una pàgina web li donarà al botó de comprar (1) o no (0). A partir d'una variable explicativa, es vol estimar la probabilitat d'una resposta binaria.

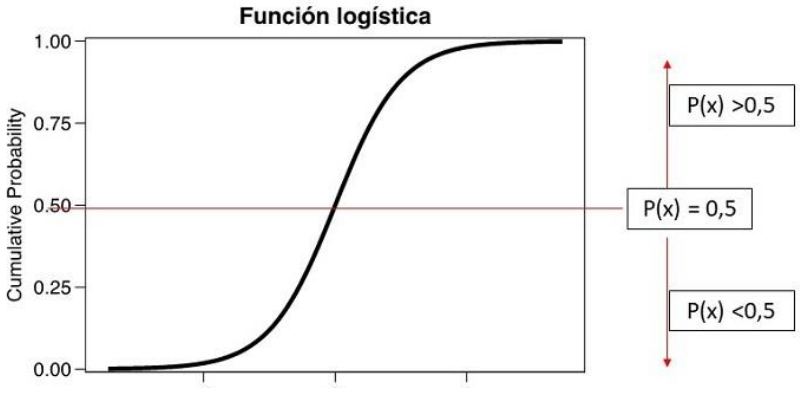
Per dur a terme la regressió logística s'utilitza, en primer lloc una funció d'activació, una de les que s'utilitza és la funció logística que és,

$$\phi\ =\ \frac{1}{a+\ e^{-x}}$$

Aquesta funció té un valor entre [0,1]. A partir d'aquesta obtenim la funció lineal aplastada:

$$\widehat{Y_W}\ (t) = \phi\ (w_0{+}w_1{*}X_1(t){+}...{+}w_m{*}X_m(t)) = \phi\ (\sum_{i=0}^{n}(w_i * X_i(t)))$$

D'aquesta manera s'obté una funció, la qual si en representem el gràfic, s'obté una forma de *s:*



Il·lustració 2. Funció logística

Mai es farà una predicció per damunt de 1 o per sota de 0. Si la corba va al infinit positiu, la predicció es convertirà en 1, en canvi, si la corba va cap al infinit negatiu, predicció es convertirà en 0. Si el valor per exemple és 0'75, significa que hi ha un 75% de probabilitat que la resposta sigui 1.

A l'igual que en la regressió lineal, en la logística també es pot calcular l'error, s'anomena pèrdua logarítmica mitjana, la funció és la següent:

$$LL(T,w) = \frac{1}{|E|} * \sum_{t \in T} (Y(t) * log(\hat{Y}(t)) + (1 - Y(t)) * log(1 - log(\widehat{Y(t)}))$$

Per minimitzar l'error , també s'han de calcular les derivades dels pessos w_i, de manera que tenim:

$$\frac{\partial}{\partial w_i} LL(T,w) = \frac{1}{|T|} \sum_{t \in T} \delta(T) * X_i(t),$$

on $\delta(T) = (\widehat{Y_W}(t) - Y_t)$. La diferència amb el cas de la regressió lineal es que el predictor en aquest error no és lineal, ja que els paràmetres no són lineals, complicant una mica el càlcul.

Descens del gradient estocàstic

Els dos tipus de regressió pretenen minimitzar els errors dels pesos o paràmetres, de manera que ens trobem davant d'un problema d'optimització. Se sol utilitzar el descens del gradient per resoldre. Aquest procediment per trobar el mínim local comença amb uns pesos inicials. A cada pas, es disminueix el pes en proporció a la derivada parcial:

$$w_i = w_i - \eta * \frac{\partial}{\partial w_i} error(T,w),$$

on η és la taxa d'aprenentatge. Aquesta taxa d'aprenentatge és la mida del pas de descens del gradient, i es proporciona a l'entrada de l'algoritme, igual que les dades d'entrenament i les variables.

En el cas de la regressió amb error quadràtic i la regressió logística amb pèrdua logística, el gradient té un petit canvi:

$$w_i = w_i - \eta * \frac{1}{|T|} * \sum_{t \in T} \delta(T) * X_i(t),$$

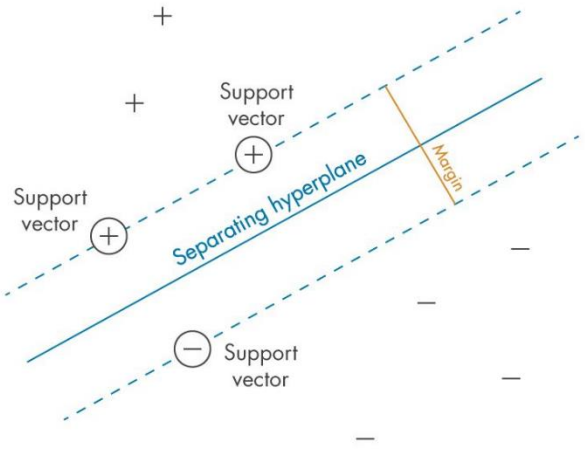
on $\delta(T) = (\widehat{Y_W}(t) - Y_t)$.

El gradient no actualitza el pes fins que ha considerat totes les dades donades, de manera que pot ser molt llarg i costos per grans quantitats de dades. Molt sovint, la solució és calcular mitjanes de forma aproximada calculant una mostra aleatòria. Si s'agafa una mostra aleatòria s'anomena descens del gradient estocàstic. El nombre de dades exemple utilitzades, és a dir, la mostra s'anomena *minibatch* o *batch*. L'algoritme actualitza els pesos cada cop que ha acabt d'explorar un *batch*, de manera que la taxa d'aprenentatge ɲ és per *minibatch*, de manera que l'actualització es divideix entre la mida de la mostra.

Els *batch* més petits permeten aprendre més ràpidament, ja que requereixen menys exemples per a una actualització. No obstant això, poden no convergir a una solució òptima local, mentre que el *batch* més grans, fins a incloure totes les dades, sí que ho faran. Això és perquè les mostres més petits poden utilitzar paràmetres no òptims i derivades incorrectes. És habitual començar amb mostres petites i augmentar-ne la mida fins a obtenir convergència o un rendiment suficient. El descens de gradient incremental, o en línia, és un cas especial de descens de gradient estocàstic que utilitza lots d'una mida d'1, actualitzant els pesos directament sense emmagatzemar valors intermedis. Això és útil per a dades de transmissió, però pot patir **d'oblit catastròfic** si els exemples no es seleccionen a l'atzar. L'oblit catastròfic és que els exemples no se seleccionen a l'atzar i el model s'entrena només amb dades més recents, pot ajustar-se molt be a les dades noves, però s'oblida de les que ha aprés anteriorment, de manera que li manca informació (Poole & Mackworth, 2023).

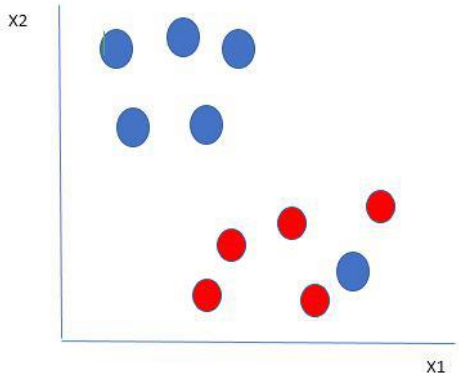
4.3. Màquines de vector de suport (SVM)

L'algoritme de màquines de vectors de suport s'utilitza tant per regressió com per classificació, tot i que és més indicat per la classificació, però també per la detecció de valors atípics (*outliers*). Els SVM són molt útils quan es té una gran quantitat de característiques, és a dir, de variables explicatives, i amb relacions no lineals. L'objectiu dels SVM és trobar l'hiperplà òptim en un espai N-dimensional que pugui separar los punts de dades en diferents classes en l'espai de característiques. L'hiperplà pretén que el marge entre els punts més propers de diferents classes sigui el màxim. El marge és l'amplada màxima de la llosa paral·lela a l'hiperplà que no té punts de dades interiors. De manera que si la sortida de la funció es major que 1, identifiquem una classe, si la sortida és -1, la identifiquem com un altra classe.

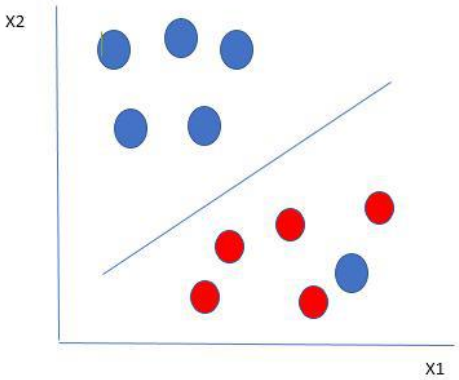


Il·lustració 3. Hiperplà del SVM

Aquest hiperplà no es pot trobar sempre, únicament pels problemes linealment separables, per la majoria de casos l'agent maximitza el marge suau permeten una petita quantitat d'errors. En la imatge per següent per exemple, SVM considerarà que la bola blava entre boles vermelles és un valor atípic, de manera que la segona figura és l'hiperplà més optimitzat (Support vector machine (SVM), n.d.).



Il·lustració 5. Dades sense classificar



Il·lustració 4. Dades classificades

En SVM les característiques es transformen mitjançant funcions de nucli, conegudes com funcions de *kernel*. Les funcions de *kernel* assignen les dades a un espai diferent, per tal de que les classes siguin més fàcils de separar un cop les dades estan transformades. Això es gràcies a que aquestes funcions simplifiquen els límits d'una decisió no lineal complexa a límits lineals en l'espai de característiques mapades de dimensions més altes (GeeksforGeeks, 2020).

L'algorisme SVM busca maximitzar el marge entre els punts i l'hiperplà, això es fa mitjançant la funció de pèrdua, que permet maximitzar la "*Hinge Loss*". De forma més matemàtica, es busca un hiperplà definit per:

$$w^t x + b = 0$$

On, $\boldsymbol{w^t}$ és el vector de pesos trasposta; $\mathbf{x}$ és el vector de característiques i $\mathbf{b}$ és el terme de biaix.

Es necessita una funció de pèrdua que mesuri com de bé el model està separant les classes, aquí es on entra la "*Hinge Loss*":

$$L(w; b; x_i, y_i) = max\big(0,1 - y_i(w^t x + b)\big)$$

On, y_i és la etiqueta de la mostra *i* (típicament $y_i \in$ {-1,1} per classificació binaria); $w^t x + b$ és la sortida del model per la mostra *i*. La "*Hinge Loss*" es comporta de la següent manera:

- Si $y_i(w^t x + b) \geq 1$, la pèrdua es cero, lo que indica que la mostra està correctament classificada i fora del marge.
- Si $y_i(w^t x + b) < 1$, la pèrdua és positiva, lo que indica que la mostra està mal classificada.

L'objectiu de SVM és minimitzar la pèrdua total en tot el conjunt d'entrenament, la qual cosa inclou la "*Hinge Loss*" i un terme de regularització per evitar el sobreajustament. La funció objectiu és:

$$min_{w,b} \frac{1}{2} ||w||^2 + C \sum_{i=1}^{n} max(0, 1 - y_i(w^t x + b))$$

On:

- $\frac{1}{2}||w||^2$ és el terme de regularització que controla la complexitat del model. Minimitzar aquest ajuda a evitar coeficient molt grans, reduint el risc de sobreajustament.
- C és un hiperparàmetre que controla l'equilibri entre la maximització del marge i la minimització de la pèrdua *hinge.*
- $\sum_{i=1}^{n} max(0, 1 - y_i(w^t x + b))$ és la suma de pèrdues *hinge* sobre totes les mostres.

Per tal de minimitzar la pèrdua i optimitzar es deriva la funció objectiu per obtenir els gradients:

$$\frac{\delta}{\delta w_k}\lambda||w||^2 = 2\lambda w_k$$

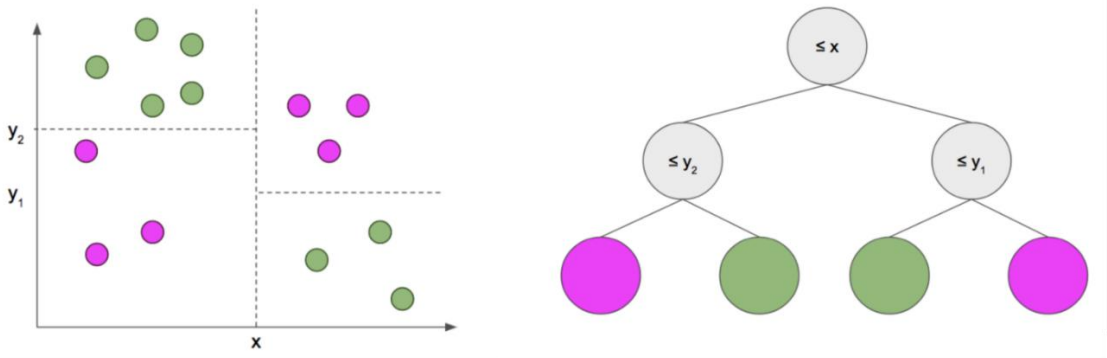
$$\frac{\delta}{\delta w_k}(1 - y_i < x_i, w >) = \begin{cases} 0, & si \langle x_i, w \rangle \geq 1 \\ -y_i x_{ik}, & else \end{cases}$$

De manera que, si hi ha una mostra mal classificada, s'actualitza el vector de pesos utilitzant el gradient d'ambdós termes; si la classificació és correcta, nomes s'actualitza w per al gradient del regularitzador. Aquest regularitzador controla la compensació entre aconseguir un error d'entrenament baix i un error de proba baix que es la capacitat de generalitzar el seu classificador a dades invisibles (Rodriguez, C. C., 2020).

4.4. Arbres de decisió

Els arbres de decisió són una representació simple per classificar exemples. S'anomenen arbres de classificació quan l'objectiu (fulla) és una classificació, i arbres de regressió quan l'objectiu és un valor real. Un arbre de decisió és un arbre en el que cada node intern està etiquetat amb una condició, una funció booleana d'exemples. Cada node intern té dos rames, una etiquetada com a "*vertader*" i l'altra etiquetada com a "*fals*". Cada fulla està etiquetada com una estimació puntual.

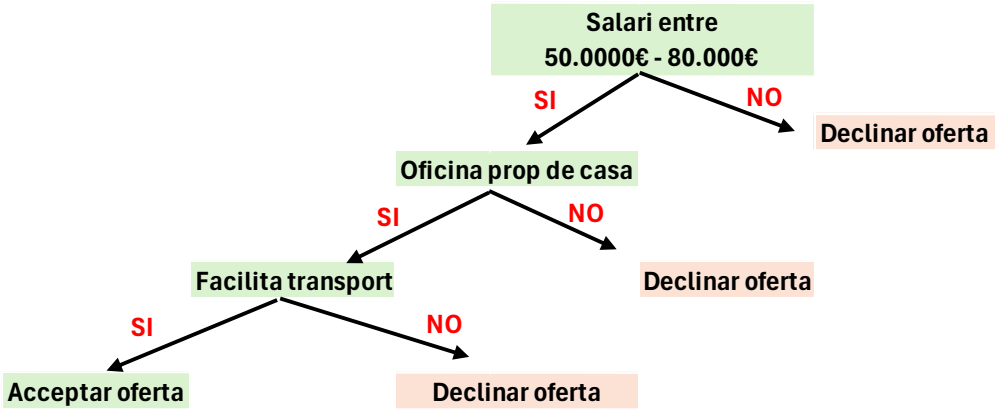
Per classificar un exemple, s'avalua la primera condició trobada a l'arbre, si se segueix l'arc fins arribar al resultat corresponent, i quan s'arriba a una fulla, es torna la classificació corresponent a aquella fulla. Un arbre de decisió correspon a una estructura niuada (Poole & Mackworth, 2023).



Il·lustració 6. Arbres de decisió

La complexitat dels arbres de decisió apareix a l'hora de buscar quin arbre s'hauria de generar i com hauria de procedir l'agent per construir un arbre de decisions. Un arbre de decisió pot representar qualsevol funció basada en característiques d'entrada discretes, per tant, s'introdueix un biaix al preferir un arbre de decisió sobre un altre. Aquest biaix pot ser a favor de l'arbre més petit que s'ajusti a les dades, és a dir, un arbre amb menor profunditat o menys nodes. Determinar quins arbres són millors per predir dades no vistes es fa a través d'experimentació i validació empírica. Tot i que l'ideal és buscar l'arbre de decisió més petit que s'ajusti a les dades, les possibilitats són enormes. Una solució pràctica és fer una cerca cobejosa, escollint la millor opció a cada pas per a minimitzar l'error (Asana, T. , 2024).

Un exemple:



Il·lustració 7. Exemple d'arbre de decisió

El problema principal es com seleccionar el millor atribut pel nodes. S'utilitza la mesura ASM, i existeixen dues tècniques:

1. **Guany d'informació**. Es mesuren els canvis en la entropia després de la segmentació d'un conjunt de dades en funció d'un atribut. Segons el valor de guany, es divideixen els nodes i es construeix un arbre de decisió. Sempre es busca el node amb més valor de guany, i el node que té el màxim valor es divideix primer. Es calcula amb la fórmula següent:

$$Guany\ d'informaci = Entropía(S) - [(Promig\ ponderat) * Entropia\ (cada\ característica)]$$

On la entropia, que mesura la impuresa d'un atribut, es calcula com:

$$Entropia(S) = -P(si) * log_2 P(si) - P(no) * log_2 P(no)$$

On S = nombre total de mostres; P(si) = probabilitat de si; P(no) = probabilitat de no.

2. **Índex de Gini**. També és una mesura de puresa que s'utilitza per crear divisions binaries. Es preferible un atribut amb un index de Gini baix a un atribut amb un índex alt. Es calcula de la forma següent:

$$Index\ de\ Gini\ = \sum_k P_k^2$$

En funció d'aquestes mesures s'escull aquell algoritme que amb les millors característiques i punts de divisió. Igual que els algoritmes vistos fins ara, necessita unes dades d'entrenament, a partir de les quals millora la seva precisió (Decision tree algorithm in machine learning – Javatpoint, n.d.).

4.5. Boscos aleatoris

Els boscos aleatoris són un algoritme basat en la creació de múltiples arbres de decisió i la combinació dels seus resultats per millorar la seva robustesa. Aconsegueix atenuar alguns dels problemes dels arbres de decisió com ara el sobreajustament, l'alta variabilitat o el biaix.

En els boscos aleatoris hi ha diversos arbres de decisió, cadascun dels quals fa una predicció sobre un subconjunt de les dades, de manera que hi ha tants arbres com subconjunts. Un cop els arbres han estat entrenats, les prediccions dels diferents arbres es combinen utilitzant un mètode apropiat per al criteri d'optimització. Alguns criteris són la moda de les prediccions del arbres per maximitzar la precisió o la mitjana de les prediccions dels arbres per minimitzar l'error quadrat mitjançant la validació creuada (Poole & Mackworth, 2023).

La combinació dels arbres de decisió es fa mitjançant dos tipus de mètodes, el *Bagging* i el *Boosting*. El *bagging*, és una agregació Bootstrap, que aconsegueix la agregació de diversos models a partir d'una família inicial. Segueix els passos següents:

1. Selecció de mostres. En primer lloc s'escullen mostres de la base de dades. Aquestes mostres s'agafen mitjançant *Bootstrap*, és a dir, es prenen dades originals amb reemplaçament.

2. Per cada mostra es crea un arbre de decisió, s'entrena i de cadascun s'obtenen les prediccions.

3. A continuació es realitzarà una votació per cada resultat previst, en una problema de classificació s'utilitzarà la moda, i en un problema de regressió s'utilitzarà la mitjana.

4. Finalment,, l'agent seleccionarà el resultat de predicció més votat com a predicció final, això es coneix com agregació.

El *boosting*, en canvi és un mètode d'aprenentatge lent. En aquest mètode es combinen una gran varietat de models que s'obtenen d'un mètode amb poca predicció, per tal d'aconseguir un millor predictor. D'alguna manera es pot dir que el *boosting* intenta arreglar els errors de predicció dels models anteriors. Segueix els passos següents:

1. Selecció d'un subconjunt de dades (n) i un subconjunt d'atributs o característiques (m) que tenen k número de registres.

2. Per cada mostra, es construeix un arbre de decisió que generarà un resultat.

3. Mitjançant votació majoritària o la mitjana, s'escollirà el resultat final.

En els boscos aleatoris es poden introduir una gran quantitat de variables, i l'agent identificarà aquells que son més significatius, mitjançant el mètode de reducció de dimensionalitat. Tot això fa que la seva interpretació sigui com complicada, és a dir, resultaria molt difícil explicar perquè pren les decisions que pren, el funcionament de l'algorisme és com una caixa negra (R, S. E., 2021).

4.6. K-NN

L'algorisme de k veïns més propers, o més conegut com K-NN, s'utilitza per comparar un punt de dades amb un set de dades amb el qual l'agent es va entrenar i va memoritzar per realitzar prediccions. El K-NN es coneix com un "algorisme mandrós", ja que no s'entrena amb totes les dades que se li donen, sinó que n'emmagatzema un conjunt en lloc de passar per una etapa d'entrenament com a tal. Aquest algorisme actua sota el supòsit de que els punts que son similars es troben prop entre ells.

L'algorisme funciona assignant una etiqueta de classe sobre la base d'un vot majoritari, és a dir, s'utilitza la etiqueta que es representa amb més freqüència al voltant d'un conjunt de punts. En el cas de la regressió, en lloc d'escollir-se l'etiqueta més freqüent s'escull la mitjana dels k veïns més propers per fer una predicció sobre una classificació. A partir d'aquí, l'agent ha de trobar els veïns més propera a un punt de consulta donat. Per determinar això es calcula la distancia entre el punt de consulta i la resta de punts, aquesta distancia permet dividir els punts en diferents seccions. La principal mesura de distància és la distància euclidiana:

- **Distancia euclidiana**. És la mesura més utilitzada i està limitada a vectors de valor real, mesura la distància amb una línia recta entre el punt de consulta i l'altre punt. La seva formula és la següent:

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^{p} (x_{ri} - x_{rj})^2}$$

on **p** representa el nombre d'atributs considerats per a la classificació, de manera que els valors dels atributs del i-èssim exemple es representen pel vector p-dimensional $X_i = (x_{1i}, x_{2i}, \ldots, x_{pi})$ € X (Arriagada, M, 2015).

S'utilitzen tots el atributs, de manera que hi ha molts d'ells que possiblement son irrellevants, això pot provocar que alguns atribuïts rellevant acabin tenint el mateix pes que els irrellevants. Per solucionar aquest biaix es pot assignar un pes a les distancies de cada atribut. Abans però, cal identificar i eliminar aquells atributs que siguin irrellevants.

El K-NN té dos algorismes, un d'entrenament i un de classificació.

- Algorisme d'entrenament:

Per cada exemple <x, f(x)>, on x € X, agregar l'exemple a la estructura representant els exemples d'aprenentatge.

- Algorisme de classificació:

Donat un exemple X, que ha de ser classificat, en $(y_1, y_2, \ldots, y_p)$ els K veïns més propers a X en els exemples d'aprenentatge, ha de retornar

$$\widehat{f(x)} \leftarrow argmax_{v \in V} \sum_{i=1}^{k} \delta(v, f(x_i))$$

On $\delta(a, b) = 1, si\ a = b;\ i\ 0$ altrament. El valor $\widehat{f(x)}$.

El valor $\widehat{f(X_q)}$ retornar per l'algorisme com un estimador de $f(X_q)$ és només el valor més comú de f entre els K veïns més propers a X_q.

La selecció òptima de K depèn principalment de les dades; en general, valors alts de K disminueixen l'impacte del soroll en la classificació, però poden generar límits entre classes

similars. Si el valor es molt baix, potser que quedi sobreajustat. Es pot trobar un bon valor de K mitjançant tècniques d'optimització. El cas especial en què es prediu la classe més propera a l'exemple d'entrenament (quan K = 1) s'anomena Algorisme del Veí Més Proper.

4.7. Naive Bayes

Els models de *Naive Bayes* són una classe especial d'algorismes de classificació que es basen en el "Teorema de Bayes". En aquest algorisme s'assumeix que les característiques son independents entre sí. A més a més, pressuposa que totes les característiques contribueixen de la mateixa manera al model.

Els models *Naive Bayes*, construeixen models a partir de calcular la probabilitat a "posteriori" de que tingui lloc un esdeveniment A, donades algunes probabilitats d'esdeveniments "anteriors" (Roman, V., 2019).

$$P(A|R) = \frac{P(R|A)P(A)}{P(R)}$$

P(A): Probabilitat de A

P(R|A): Probabilitat condicional de que es doni R donat A

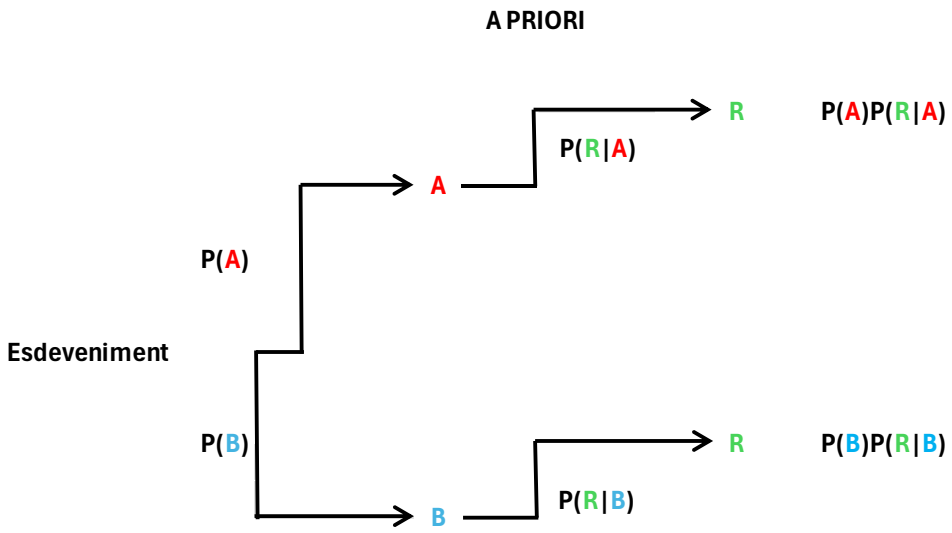
P(R): Probabilitat de R

P(A|R): Probabilitat posterior de que es doni A donat R

L'agent calcula les probabilitats a posteriori i retorna la més alta, de manera que:

$$\hat{y} = arg\max_{y \in Y} P(y|x) = argmax \frac{P(x|y)P(y)}{P(x)}$$

L'esquema que obtenim, pel qual pren decisions l'agent és el següent:



Il·lustració 8. Probabilitats a priori

POSTERIORI

$$P(B|R) = \frac{P(B)P(R|B)}{P(A)P(R|A)+P(B)P(R|B)}$$

$$P(A|R) = \frac{P(A)P(R|A)}{P(A)P(R|A)+P(B)P(R|B)}$$

Il·lustració 9. Probabilitats a posteriori

Un cop han estat analitzats tots els algorismes d'aprenentatge automàtic supervisat, en el següent apartat es plantejaran els problemes que aquests presenten.

5. PROBLEMES DELS ALGORISMES

En l'apartat de conceptes bàsics, s'ha vist que una de les definicions dels algorismes era que pensen/actuen com humans, i per tant, el seu coneixement, a l'igual que l'humà, pot estar esbiaixat. Cal descartar la concepció de que els algorismes pensen racionalment, ja que al analitzar-los ha quedat a la vista que tots contemplen errors, i que de fet, la majoria funcionen intentant disminuït l'efecte d'aquests errors.

A més a més, un cop examinats els algorismes d'aprenentatge automàtic supervisat, es fàcil veure que contenen una gran part estadística, i que de fet, són generats fonamentalment per

tècniques estadístiques. Des de regressions i teories probabilístiques, fins a problemes d'optimització d'errors. Es per tant, necessari analitzar quins problemes poden presentar aquest algorismes des d'un punt de vista estadístic.

Es poden classificar els problemes que poden presentar els algorismes en 3 grans grups. En primer lloc les dades d'entrenament utilitzades, això inclou si estan esbiaixades, la procedència d'aquestes dades, i per tant la seva privacitat, i l'oblit catastròfic. En segon lloc, aquells errors que van intrínsecs en els algorismes, formen part d'aquest grup tots aquells errors com l'error quadràtic mitjà de les regressions lineals, el sobreajustament, la confusió, etc. En tercer lloc cal considerar aquells errors que introdueix el factor humà, ja que l'aprenentatge automàtic supervisat requereix d'intervenció humana, i per tant, mai desapareix l'error causat per aquell que manipula l'algorisme. Aquest seria la correlació entre variables, no escollir bé les variables introduïdes, la K en el cas del K-NN per exemple, entre d'altres.

5.1. Dades d'entrenament

S'ha vist que una part fonamental i present en tots els algorismes són les dades d'entrenament. Es podria dir que un algorisme és com un nen que ha d'aprendre a fer una tasca, de manera que aprendrà allò que se li ensenyi. Això vol dir que si a un agent que s'encarrega de la llista d'espera li ensenyem que cal prioritzar les persones d'ètnia blanca sobre les de ètnia negra, així ho farà. En les persones és molt fàcil veure que tot allò que hem viscut marcar la nostra manera de pensar i actuar, i cal comprendre que d'igual manera passa amb els algorismes.

El coneixement es adquirit pels algorismes mitjançant la experiència de les dades que els introduïm, si aquestes dades estan esbiaixades tot el model ho estarà. Per tant, en aquest punt es presenta el primer problema de les dades d'entrenament, cal assegurar-se que no són esbiaixades i que no son discriminatòries. Per exemple, posem el cas d'un algorisme que gestiona la llista d'espera d'un hospital, i prioritza aquelles persones que tenen més probabilitats de morir. Té una variable que és el barri, que passaria si a causa dels diferents nivells econòmics, en les dades consta que aquelles persones que son de barris amb major nivell econòmic, sobreviuen més que les que viuen en barris amb nivells econòmics més baixos, tenen realment més probabilitat de sobreviure les persones que provenen d'un barri amb més nivell econòmic després de l'operació? No son altres els factors que afecten en el nivell de supervivència d'aquells que viuen en barris mes baixos?

Un altre tipus d'error, es la perspectiva més global que un agent pot perdre d'unes dades. Un exemple que poden causar les dades és el següent. Imaginem que tenim algunes característiques com les següent sobre una persona:

- Home
- Nascut el 1948
- Criat a Regne Unit
- Casat dues vegades
- Residència d'alt cost econòmic
- Famós

Aquesta persona, podria ser el rei Carles III d'Anglaterra, però també podria ser Ozzy Osbourne, això deixa entreveure que dues persones amb dades idèntiques poden ser realment molt diferents entre ells. Les grans diferències entre el rei i el cantant de rock son molt fàcils de veure per un humà, però no podrien ser detectables per una algorisme que només té aquestes dades.

A més a més, alguns dels algorismes, com ara la regressió logística o els boscos aleatoris no utilitzen totes les dades d'entrenament, sinó que en seleccionen algunes mostres; ja que d'altra manera el cost computacional seria molt alt. Cal assegurar-se de com s'agafen aquestes mostres, i això és una part fonamental de l'estadística. Posem el cas que en una base de dades, hi ha una variable que és el sexe, i és una variable significativa, les conseqüències de no agafar un dels dos sexes en les mostres o que no es fes el mostreig correctament podria tenir conseqüències molt negatives. Per exemple, en l'àmbit mèdic, on els efectes de les malalties o la seva prevenció pot variar entre sexes, si això no es detectés i es tractes a ambdós sexes igual podrien haver greus conseqüència. O el cas contrari, en cas que el mostreig provoques que es detectés unes diferències entre sexes que no existeix o no es rellevant en les dades al complet, però si en la mostra. O anant més enllà, potser que un grup o una classe, ni tant sols estigui contemplat en les mostres.

Un últim punt sobre les dades que s'hauria de tenir en compte és la privacitat i la governança de dades. La privacitat va sorgir al principi com un dret a estar sol, i per tant, com un dret per

excloure a la resta dels fets personals. Més endavant aquest dret, va passar de la exclusió del fets personals, al seu control. Avui en dia Ara ens enfrontem a una tercera fase, que respon a la pregunta "qui sóc jo?" i la resposta no depèn del subjecte de les dades sinó de les patrons seleccionats per tercers que creen un perfil analític i porten a repensar el dret a la identitat. Ja no només s'obtenen dades d'enquestes, sinó que constantment es generen una gran quantitat de dades, en xarxes socials, cerques per internet, compres, etc. Moltes empreses utilitzen les dades dels seus clients per millorar el seu servei, però fins a quin punt utilitzar les dades dels clients per oferir-li allò que vol, a la hora que vol, no es manipular? I no només això, creuant dades es pot obtenir una gran quantitat d'informació, per exemple, amb les dades d'electricitat d'una casa, es podrien descobrir patrons de comportament de les persones que hi viuen, quan hi son, quan dormen, si s'aixequen per la nit. Totes les dades es poden utilitzar?

Cal tenir molt en compte la procedència de les dades, i fer un processament de dades i anàlisis molt curós, per mirar de no eliminar informació que podria esbiaixar les dades.

A continuació, es parlarà sobre els errors intrínsecs dels algorismes.

5.2. Errors intrínsecs dels algorismes

S'ha vist al analitzar els models que hi ha alguns errors estadístics en els models que ja es contemplen en ells, de manera que l'objectiu de l'algorisme és reduir aquest error el màxim possible. En aquest apartat m'agradaria esmentar els errors, deixant una classificació més clara. La intel·ligència artificial no pot ser confiable al 100%, si a l'igual que l'ésser humà contempla l'error.

Alguns dels algorismes mesuren l'error mateix, es el cas, per exemple de la regressió lineal i la regressió logística. La primera calcula error quadràtic mitjà, i la segona la pèrdua logarítmica mitjana. En aquests algorismes l'objectiu de l'agent en ambdós algorismes és reduir aquest error que ja es contempla en la regressió. L'error quadràtic mitjà és molt sensible als valor atípic (outliers) a causa de que es troba elevat al quadrat, de manera que alguns *outliers* poden afectar significativament l'EQM. Això pot provocar o que l'error sigui pitjor del que sembla per la majoria de dades, i a més a més, que les estimacions no siguin tant exactes com haurien de ser si aquests valors atípics no hi fossin. Aquí entra en debat que fer amb aquests outliers, ja que per exemple, en un algorisme de diagnòstic mèdic, pot fer que les decisions siguin subòptimes per la majoria de pacients. Però es clar que no es poden eliminar els *outliers*, perquè formen

part de la població, i també podria tenir greus conseqüències eliminar-los, ja que una part de la població quedaria totalment descartada.

L'EQM a més a més no detecta el biaix entre grups que pot haver, és a dir, pot ser que el valor de l'error sigui baix, però hi hagi errors sistemàtics per grups minoritaris. Per tant, s'han d'utilitzar més mesures que avaluïn l'equitat del model.

Un altre problema de l'error quadràtic mitjà, és que de vegades per reduir-lo cal afegir variables o característiques al model, de manera que es cada cop més complex d'interpretar. Prioritzar un baix EQM a un model transparent pot comportar que aquells que han dissenyat el model no el puguin interpretar. El reduir tant l'error quadràtic també pot provocar que el model sigui sobreajustat, tenint massa en compte els valors atípics (Madrigal, E., 2022).

Les problemàtiques amb la pèrdua logarítmica mitja son semblants a les de l'error quadràtic mitjà. En primer lloc penalitza molt les dades incorrectes amb alta confiança, de manera que el model pot quedar sobreajustat si s'intenta minimitzar massa. En segon lloc, tampoc té en compte quan la freqüència d'un grup és molt més alta que la d'un altre grup. Aquest que sembla un problema menor, però més endavant es parlarà dels problemes de sobre ajustar el model. Per últim, el models logístics optimitzats per la pèrdua sovint son molt difícils d'interpretar, i això pot causar grans problemes en models on l'explicabilitat és molt important, com diagnòstics mèdics. Aquesta poca transparència en el model també passa en altres algorismes, un cas on això és un problema freqüent son els boscos aleatoris, on la complexitat del model provocar que sigui com una caixa negra. Cal plantejar-se, si és més important la precisió de les prediccions d'un model, o és preferible que sigui interpretable i fàcil d'entre.

A més a més, el model logístic té el que es coneix com oblit catastròfic, que és la tendència del model d'oblidar informació prèviament apresa al intentar aprendre'n de nova. Això es provocat perquè si el model ha estat entrenat amb un dades i es exposat a nova informació, pot sobreescriure els pesos que eren importants per la informació prèviament apresa. Això provoca que la confiança cap al model disminueixi molt, ja que la seva capacitat de predicció disminueix amb el pas del temps. Això pot provocar per exemple, que un model oblidi informació que podia ser rellevant per un grup, provocant un tracte favorable cap a l'altre. O en un cas de diagnòstic mèdic, pot tenir greus conseqüències si per exemple oblida símptomes d'una malaltia. Cal afegir, que és força complicat entendre quina és la informació que el model ha oblidat a causa de la seva complexitat i poca transparència (Poole & Mackworth, 2023).

Així doncs, es veu que un dels problemes que sovint apareix és el sobreajustament, que és quan un agent fa prediccions molt exactes sobre les dades d'entrenament, però no sobre el mon real. Això es causat perquè el model ha aprés molt bé els detalls i el soroll del conjunt de dades d'entrenament, però aquests no es veuen relaxats en la població. Normalment passa en models molt complexos, amb gran quantitat de variables i paràmetres en relació al nombre d'observacions, provant que es memoritzi les dades i no els paràmetres. El problema és que solen ser models molt difícils d'interpretar i això dificultar la interpretació dels errors. A més a més, solent provocar molta falsa confiança ja que fan prediccions molt bones sobre les dades d'entrenament, fent que les mètriques resultin amb molt pocs errors; però a la realitat no fan bones prediccions. A més a més, pot ocórrer que si les dades d'entrenament estan una mica esbiaixades, aquest biaix s'amplifiqui molt més al fer prediccions en la realitat (Amazon, n.d.).

En l'error quadràtic mitjà s'ha vist que els *outliers*, poden significar un problema si no es tracten com cal. En els SVM, els valors atípics també representen un greu problema perquè, l'objectiu de l'agent és minimitzar el marge entre les classes i l'error de classificació, i els *outliers* poden distorsionar aquest marge, causant que no quedi ben reflectida la distribució general de les dades. En aquest algorisme, igual que en els anteriors, els valors atípics, poden provocar que no es tinguin en compte grups minoritaris. Al mateix temps, algunes tècniques per tractar-los provoquen que el model sigui molt complex, de manera que es molt difícil d'interpretar. A més a més, el SVM, contempla un terme (**b**), que és el biaix (Sickit- learn, n.d.).

Per últim, alguns dels algorisme no tenen en compte la dependència entre variables (multicol·linealitat), l'algorisme on principalment això és una problemàtica és el Naive Bayes, que s'assumeix que les característiques són independents entre sí. Això pot tenir greus conseqüències, ja que suposar que algunes característiques són independents, si no ho son, provocar que les probabilitats conjuntes no estiguin ben calculades, de manera, que tot l'algorisme seria erroni o imprecís. En els models de regressió, tant lineal com logística, és força important tenir en compte la multicol·linealitat, ja que pot provocar que els models siguin inestables i dificultar la interpretació dels models (Roman, V., 2019).

En tots els algorismes, cal a més a més, tenir en compte el soroll, que és l'error causat perquè les dades depèn de característiques no modelades. El soroll provoca que la variabilitat no explicada del model augmenti, i sigui difícil identificar per què es causada (Tilves, M., 2014).

La següent taula és una síntesi dels problemes que poden contenir els algorismes, de manera que el triangle verd significa que es robust respecte aquest error i ofereix bones prediccions. El diamant taronja significa que l'algorisme té certes limitacions i que per tant s'ha de tenir molt en compte com es tracten aquests problemes. El vermell significa que el problema representa un gran repte per l'algorisme.

Problema	Models lineals	Models logístics	SVM	Arbres de decisió	Boscos aleatoris	Naive Bayes	K-NN
Error contemplat algorisme	▲	▲	▲	▲	▲	◆	◆
Explicabilitat	▲	▲	◆	▲	◆	▼	◆
Outliers	▼	▼	◆	▲	▲	▼	▼
Sobreajustament	▼	◆	▲	▼	◆	▲	◆
Multicolinealitat	▼	▼	◆	▲	◆	▼	◆
Soroll	◆	◆	◆	▲	▲	▼	▼

Il·lustració 10. Fortaleses i debilitats dels algorismes

Per acabar aquest punt, només afegir que està clar que els algorismes no tenen un funcionament perfecte, i ben lluny d'això cal empra moltes eines i anàlisis per intentar evitar, almenys millorar tots el problemes esmentats. Aquests són els errors més bàsics dels algorismes, però si els analitzem amb molta més profunditat trobaríem molts més problemes com els tipus d'error, les hipòtesis, les validacions creuades, etc. Cal veure quins errors pot afegir el factor humà

5.3. Errors introduïts pel factor humà

L'aprenentatge automàtic supervisat requereix de la intervenció humana, i per tant, és més susceptible que la resta d'aprenentatges automàtics ha ser esbiaixat per l'ésser humà. Tot i això tots els algorismes, incloent els d'aprenentatge automàtic no supervisat, el de reforç i l'aprenentatge profund són susceptibles a tenir errors o biaixos causat pel factor humà.

El principal problema que pot introduir una persona en un algorisme és el fet de que ell quan el dissenya ho fa en base a una tesis o hipòtesis, és a dir, busca una resposta; i mai es trobarà allò que no es busca. Això significa que sovint, tota una investigació està esbiaixada des d'un

principi, per exemple, Cristòfol Colón, no buscava el continent americà, sinó que buscava la Asia, i ell va morir pensant que ho havia assolit. Si Cristòfol Colón hagués dissenyat un algorisme d'IA sobre aquest descobriment, hagués estat incorrecte perquè el coneixement que tenia era erroni. Al cap i a la fi, els agents de la IA, són com un nen al que l'ésser humà ensenya, i pot aprendre tots els biaixos i discriminacions que aquest li transmeti. És evident, que les conseqüències d'això poden ser molt greus. En el cas dels algorismes d'aprenentatge automàtic, un exemple són els arbres de decisió, ja que és l'ésser humà qui decideix quines característiques introdueix, potser no en té en compte alguna que es rellevant o decisiva, que canvia totalment el resultat.

En segon lloc, un problema que presenta l'ésser humà respecte a la IA, és com aquest l'utilitza, és a dir, la finalitat. Pot passar que algunes persones facin un mal ús dels algorismes. Per exemple, els algorismes que s'encarreguen de fer les recomanacions, poden portar a la radicalització. El cas d'una persona que hagi buscat informació sobre terrorisme, i els algorismes li comencin a recomanar pàgines o fòrums on s'incita al terrorisme. Per tant, el mal ús dels algorismes pot comportar greus problemes. Un altre mal ús podria ser obtenir informació que hauria de ser privada. Té molt a veure amb la privacitat de les dades que s'ha comentat prèviament.

A més a més, també s'ha de tenir en compte el context, o la falta de context a l'hora de que s'interpreti o s'utilitzi un algorismes. De vegades una persona dissenya un algorisme, però és un altra la que l'utilitza, de manera que pot ser que o bé no tingui el coneixement necessari per comprovar que no hi ha biaix o que les dades estan correctes; o bé que amb les dades del nou usuari els resultats no siguin tant bons. Pot ser que el nou usuari tingui coneixements estadístics però a causa de la falta de transparència dels models, com ara el boscos aleatoris, sigui difícil per aquesta persona interpretar-lo o explicar perquè es comporta com es comporta.

6. ÈTICA

La filosofia i la intel·ligència artificial tenen una gran vinculació, tot i que la demostració d'aquesta no és adient en aquest treball. Però, al cap i a la fi, una rama molt extensa de la filosofia és la Teoria del Coneixement, on es pretén definir què és el coneixement, que es pot considerar coneixement, com s'adquireix i moltes altres qüestions. I la finalitat de la intel·ligència artificial és replicar aquest coneixement humà en màquines, caldria doncs, definir

què és el coneixement humà, abans de saber si aquest pot replicar en màquines. Alguns filòsofs grecs, com Aristòtil (384-322 a.C.), ja es plantejaven un conjunt de lleis formals per tal de governar la part racional de la intel·ligència, la lògica aristotèlica permetia treure conclusions a partir d'unes premisses. Thomas Hobbes (1588-1679), més endavant va proposar que el pensament era com una computació numèrica, de manera que el pensament era com sumar i resta idees. Altres, com Leonardo da Vinci (1452-1519), Pascal (1623-1662) o Leibniz (1646-1716), van arribar a dissenyar, i construir calculadores i màquines que replicaven càlculs.

René Descartes (1596-1650), per la seva banda va ser el primer en plantejar la discussió sobre la separació entre la ment i la matèria, i tots els problemes que això comporta. A partir d'aquí van sorgir diferents corrents, el dualisme, el materialisme, l'empirisme o el positivisme lògic. Aquesta última doctrina sosté que tot el coneixement es pot caracteritzar mitjançant teories lògiques. I finalment, la teoria de la confirmació de Carnap (1891-1970) i Carl Hempel (1905-1997) pretén explicar que tot el coneixement s'obté a partir de l'experiència. Aquesta, ja podria ser un altra rama molt important de la filosofia que seria la Lògica, aquesta també està força relacionada amb els algorismes d'intel·ligència artificial, ja que molts segueixen raonaments lògics per tal d'arribar a les seves prediccions. Tot i això aquestes rames, no son les que interessen en aquest treball, sinó l'ètica (Russell & Norvig, 2004).

L'apartat anterior era una síntesis d'alguns problemes que poden tenir els algorismes d'aprenentatge automàtic supervisat. Els problemes son diversos, i es necessari tenir-los en compte a l'hora d'utilitzar i dissenyar algorismes. Igual que la resta de disciplines científiques, la Intel·ligència Artificial, és i ha de ser analitzada des d'una perspectiva ètica. Primer cal definir dos conceptes:

- Ètica: Estudia el comportament humà i la seva relació amb les nocions de bé i mal. Pretén definir les principis que guien el comportament, analitza quines serien les accions correctes tot i que vagin en contra de la tradició (Equipo de Enciclopedia Significados. 2013).

- Moral: És allò pertanyent o relatiu a les accions de les persones, des del punt de vista de l'obrar en relació al bé i al mal i en funció de la seva vida individual, i sobretot, col·lectiva (Rae - Asale., n.d.). També es pot entendre com el conjunt de normes, costums, creences i valors que formen part de la tradició d'un individu o una societat.

L'ètica i la moral es relacionen de manera que la primera és una disciplina filosòfica que s'encarrega d'estudiar els principis que regules el comportament moral. D'alguna manera, l'ètica és la disciplina i la moral l'objecte d'estudi. Totes les persones i societat tenen una moral, en tant que fan judici de les accions, considerant que algunes son bones i d'altres dolentes. La societat requereix de codis morals per tal de garantir la pau ciutadà i l'harmonia social. Cada societat té una moral diferent, una mateixa acció pot ser ben vista en una cultura cristina, i en canvi, se mal vista en una cultura asiàtica. La moral evoluciona al llarg del temps, com a conseqüència de canvis històrics, innovacions, etc. Sovint en les disciplines es defineixen codis morals que totes aquelles persones que en formen part han de complir, en la biologia, la medicina o l'advocacia en son exemples.

En aquest punt es pretén dur a terme un anàlisis ètic, sobre si la IA, hauria de tenir o no un codi moral, i quin és el paper de l'estadístic en tot això. Al llarg dels diferents punts s'han vist els diferents tipus de problemes. D'una banda tenim aquells que d'alguna manera son "inevitables", els algorismes contenen errors, i de fet, l'objectiu dels agents no és tant executar l'algorisme, sinó optimitzar l'efecte dels errors que conté. S'ha vist que els efectes d'aquests errors poden ser molt perjudicials, per exemple en casos mèdics, on si l'algorisme està esbiaixat alguns diagnòstics poden ser erronis. És necessari analitzar de qui seria la responsabilitat en cas d'errors, pot ser una màquina responsable dels seus actes? No es parla d'un mal ús de la tecnologia, sinó del seu disseny i desenvolupament. No es pot confiar plenament amb els sistemes d'Intel·ligència artificial, no se sap si les decisions que prendran seran sempre les correctes. Que cal fer al respecte? Es pot fer implementar un codi moral en els agents per tal de que actuïn correctament?

Alguns autors ja han escrit sobre temàtiques semblants. Isaac Asimov (1950), un dels primers pesadors, va dissenyar les conegudes Les Lleis de la Robòtica de Asimov, que són les següents:

1. Un robot no pot danyar un ser humà ni, per interacció, permetre que un ésser humà pateixi sofriment.

2. Un robot ha d'obeir les ordres que li donen els éssers humans, excepte quan aquestes entrin en conflicte amb la primera llei.

3. Un robot ha de protegir la seva pròpia existència, sempre que aquesta protecció no entri en conflicte amb la Primera o la Segona llei.

La proposta d'Asimov proposa que totes les màquines segueixin aquestes lleis, es podria considerar una mena de codi ètic que tot fabricant de robòtica hauria de garantir. (Poole & Mackworth, 2023) Aquest podria ser un molt bon inici, elaborar un seguit de lleis que tot algorisme hauria de complir per ser utilitzat o comercialitzat. És potser, una tasca més complexa que en el cas dels robots, perquè cal contemplar, els errors intrínsecs, que ja s'ha comentat que son inevitables. Tot i això, alguns dels aspecte que es podrien contemplar com un codi que s'hauria d'assegurar que els algorismes complissin són:

- La explicabilitat de la IA, ja que les decisions basades en aquesta poden afectar àmbits molt sensibles, i és necessari poder veure d'on provenen les decisions presses.
- Evitar biaixos, tant en la recopilació de les dades com en el seva gestió, i en l'entrenament dels agents.
- Garantir la privacitat de dades.

Es clar, que hi ha molts més aspectes a tenir en compte, com el soroll, les variables utilitzades, i altres que no entren dins dels límits d'aquest treball i que caldria implementar, aquest és només un plantejament.

Els estadístics son qui, moltes vegades, en un futur es dedicaran a desenvolupar els algorismes, i per tant, qui haurà d'analitzar i trobar una forma d'implementar algunes lleis morals per tal que la IA actuï moralment. No seria, per tant, necessari que les persones dedicades al desenvolupament d'aquests algorismes, i dels codis ètics que seguiran, tinguin algunes nocions mínimes sobre ètica i moral? Com si no, es podrà sinó assegurar que aquests codis tinguin sentit?

D'altra banda, s'ha comentat els factors introduïts pel factor humà. La influència de la IA es cada cop més extensa, els algorismes d'aprenentatge automàtic són utilitzats en una gran quantitat de disciplines, des de la gestió de reserves en un hotel, fins al diagnòstic de malalties en hospitals. Hi ha diverses preguntes que qüestionar-se respecte al codi ètic que haurien de seguir aquelles persones que es dediquen a desenvolupar la intel·ligència artificial.

La ètica kantiana sembla ser molt adequada per analitzar com s'hauria de comportar una persona dedicada al desenvolupament d'intel·ligència artificial, ja que és una teoria fàcilment universalitzable. És una de les teories que s'ha utilitzat per analitzar la moralitat de la IA. D'una

forma molt senzilla, es pot dir que per Kant una de les màximes[1] més importants és que cal utilitzar als altres sempre com un fi per ells mateixos, i mai com un mitjà. Si seguim aquest criteri, que passa amb totes les dades que les companyies utilitzen per treure'n benefici? Està vendre les dades? O recomanar contingut per què és el que convé a la companyia? Tota la ètica kantiana es troba dirigida contra la moral de l'amor propi, els actes, per Kant han d'estar dirigits cap al bé comú (Malishev, M. , 2014). De fet, Kant diferencia tres tipus d'actes:

- Acte en contra del deure: Són actes immorals, són aquells considerats dolents, per exemple matar o robar.
- Actes conforme al deure: Són aquells actes que són morals, però no son duts a terme perquè siguin morals, sinó per altres inclinacions, com per exemple la por al càstig, o el benefici econòmic.
- Actes per deure: Són aquells actes bons per ells mateixos, aquelles accions que es fan perquè és el correcte, i no el que li convé.

Sovint, és molt difícil diferenciar quan una persona actua conforme al deure i quan per deure. És fàcil pensar que sempre que es faci el correcte hauria de donar igual, però posem un cas en Intel·ligència Artificial. Imaginem un algorisme que fa prediccions mèdiques, aquest algorisme és dissenyat per una persona que el fa, no perquè salvi la vida dels pacients, sinó perquè sap que els hospitals els compraran a un alt preu, perquè és molt eficient. Ara posem el cas, que l'hospital A, li diu al programador que si ven un algorisme esbiaixat al hospital B, li pagarà el doble. És possible que el programador ho faci, perquè ell no ha fet l'algorisme perquè salvarà més vides, sinó per guanyar diners, i la opció que li donarà més diners ara serà vendre a l'hospital B un algorisme esbiaixat, que faci males prediccions, per tal de que l'hospital A li pagui el doble. Sembla un cas molt rebuscat, però queda clara la situació. Una decisió com aquesta pot tenir greus conseqüències.

Els estadístics, tenen i tindran en un futur un paper molt important, les seves decisions, com la que ha pres el venedor de l'algorisme de l'exemple anterior, poden tenir greus conseqüències. Sense parlar, de totes aquelles conseqüències que es poden cometre involuntàriament, a causa dels errors dels algorismes. És un mercat que està prenent molta influència, que genera molts diners, i que possiblement en generarà més. Està clar que les persones que es dedicaran a la investigació i desenvolupament dels algorismes hauran de tenir en compte si aquests han de

[1] Idea, norma o designi a la que s'ajusta la manera d'obrar (Rae - Asale, s.f., definició 3).

tenir un codi ètic. Però aquells que comercialitzaran amb ells, o els replicaran, també tenen una gran responsabilitat. Anant un pas encara més enllà, és possible, que en un futur una mica més llunyà, siguin les IA qui es dediquin a desenvolupar els algorismes, de manera que allò que podrem aportar les persones serà la perspectiva ètica, la perspectiva humanitària que es allò que encara ens diferencia de les màquines.

El punt al que es vol arribar és que, ara mateix, en Estadística i altres graus que es dediquen a formar als futurs treballadors amb intel·ligència artificial, no contenen formació ètica. No es prepara als alumnes per tenir un pensament crític, per aportar aquesta humanitat, per decidir davant d'aquests dilemes morals que segur, que es plantejaran. Entrant concretament al grau d'Estadística, no faria encara més competents als seus alumnes, davant d'altres graus que es dedicaran a la mateixa professió, donar una certa formació ètica?

III. CONCLUSIONS

A l'inici del treball es plantejava la idea de que l'estadística i la Intel·ligència Artificial tenen una forta relació, i és més, que la primera és una part fonamental de la segona. Els objectius del treball eren demostrar el paper clau dels estadístics en la Intel·ligència Artificial, trobar els possibles problemes que poden presentar els algorismes i la necessitat d'una formació ètica per als estadístics.

A causa de la gran quantitat d'algorismes d'Intel·ligència Artificial que hi ha avui en dia, i a causa de les limitacions d'aquest treball es va optar per centrar l'estudi en els algorismes d'aprenentatge supervisat, ja que amb la definició d'aquest tipus d'aprenentatge ja era suficient per trobar la interacció entre la IA i les tècniques estadístiques . A partir d'aquí es van desglossar els 7 algorismes, de manera que quedava força clara la base i els fonaments estadístics que tenen, i per tant la forta relació que tenen amb l'estadística.

D'una banda els models de regressió lineal i els models de regressió logística, estan basat en la regressió de models estadístics i els seus errors, intentant reduir-los mitjançant tècniques estadístiques. En els SVM es busca l'optimització d'un hiperplà mitjançant, també, diferents tècniques estadístiques. Els K-NN es basen en la moda de la distància euclidiana, que són conceptes també estadístics. Els *naive bayes* és fonamenten en l'aplicació del Teorema de Bayes, que pertany a l'àmbit de la probabilitat. D'altra banda, en els arbres de decisió i els boscos aleatoris, tot i que per la seva implementació no s'utilitzen exactament tècniques estadístiques, la bondat de les prediccions es mesura mitjançant termes estadístics. Així doncs, queda clar que el paper de l'estadístic en el desenvolupament dels algorismes d'Intel·ligència Artificial, és i serà fonamental.

Seguidament, es van extreure els possibles problemes que els algorismes podien presentar, i van quedar classificats en tres grups: problemes en les dades d'entrenament, problemes intrínsecs dels algorismes i problemes causats pel factor humà. Tots tenien una forta influència en els efectes que els algorismes podien tenir. En primer lloc, les dades d'entrenament són el que nodrirà els algorismes, si aquestes estan esbiaixades, tot l'algorisme ho estarà. Però només això, a més a més, es força important tenir en compte d'on provenen les dades, la seva font ha de ser legítima, i respectar el dret a la privacitat. És molt important tenir en compte que les dades d'entrenament son com l'educació que se li dona a un nen, i seran les que educaran al

algorisme, que aprendrà el que les dades li ensenyin, les bones prediccions, però també les dolentes, les característiques de les observacions, però també els biaixos.

En segon lloc es trobaven els errors intrínsecs dels algorismes, aquests són en certa manera inevitables, tot i això, són, en certa manera, controlables si es coneix quins algorismes són millors per cada cas, quines son les seves flaqueses, on s'han de vigilar, o on fallaran. S'ha realitzat una taula que permet veure quines són les fortaleses i les debilitats dels algorismes. Pel simple fet d'escollir un algorisme sobre un altre, ja s'està creant un biaix, ja que s'escull en funció de les fortaleses i les debilitats, de manera que es prioritzen alguns problemes sobre els altres. Per exemple, si s'escull un arbre de decisió sobre un model logístic, s'està prioritzant tenir en compte els *outliers*, sobre el fet de no tenir sobreajustament. És molt important descartar la idea tant comú de que les intel·ligències artificials no cometen errors, i entendre tot el contrari, que sovint el fet de pensar que l'IA no s'equivocarà, provoca que es cometin més errors.

En tercer lloc, es troben els error introduïts pel factor humà que són causats principalment pels perjudicis que les persones tenim inevitablement. A més a més, de la intenció amb que els fan servir. Al principi del treball es van plantejar diferents definicions de la Intel·ligència artificial, de manera que podia ser un sistema que penses com un humà, que actués com un humà, que penses racionalment o que actués racionalment. Sabent que la racionalitat és la forma perfecta d'intel·ligència, i la IA pot cometre errors, no podem dir que pensa racionalment, sinó que ho fa com un humà. Els agents, en tant que pensen com un humà, poden estar esbiaixats. No es pot esperar que les prediccions de la Intel·ligència Artificial siguin perfectes, i això ho ha de tenir en compte, tot aquell que ho desenvolupi.

Serà per tant, necessari que es contempli la idea d'implementar un codi moral en els agents, per tal d'assegurar que tots els problemes ètics dels algorismes són contemplats pels agents. Seguint la proposta d'Asimov, de les lleis de la robòtica, es podria continuar amb unes lleis per al Intel·ligència Artificial. Caldria per tot això, una visió més general de la IA, no únicament com una ciència, sinó també com una disciplina de la filosofia.

Es plantejava també en l'últim punt, el d'ètica, que la vinculació de la filosofia amb la IA és també notòria, ja que alguns de les disciplines que la filosofia investiga sobre l'ésser humà, com la lògica i la teoria del coneixement, són precisament les que la IA vol plasmar de l'ésser humà en les màquines. Mitjançant la ètica kantiana, es plantejava un dilema que podia tenir greus

conseqüències, arribant a la conclusió que era necessari que els estadístics rebessin una certa formació per tal de tenir un pensament crítics. Cal no només fer més èmfasis en la formació ètica, sinó en la importància que els estadístics tenim en el desenvolupament de la Intel·ligència Artificial.

A partir de tota aquesta informació, queda a la vista la necessitat d'una visió holística de la intel·ligència Artificial. Fa temps, que des del camp de la filosofia es pregunten sobre l'origen del coneixement, sobre si és replicable en màquines, d'on prové, com es pot mesurar, o la complexa qüestió de què és el coneixement en sí mateix. Una qüestió tant complexa, com és aquesta, no pot ser mirada únicament d'una perspectiva si es vol crear sistemes intel·ligents, és necessària una mirada general.

Aquest treball pretenia generar un espai de reflexió per totes les interaccions que hi ha entre la IA i l'estadística, tenint en compte el component ètic. És clar, que el treball s'ha topat amb algunes limitacions. En primer lloc ha estat tot un repte definir l'estructura del treball i decidir que s'explicaria i que quedaria fora. La quantitat d'informació actual sobre la intel·ligència artificial és molt extensa, les definicions havien de ser clares, i no es podien explicar tots els algorismes existents a causa dels propis límits del treball. Seleccionar la informació de qualitat i sintetitzar-la va ser tot un repte. En segon lloc, la informació sobre els problemes que presenten els algorismes era menys i molt dispersa, de manera que posar tots els conceptes en ordre i classificar-los va ser una tasca difícil. Per últim, va ser un desafiament considerable buscar quines necessitats ètiques tindria un estadístic, ja que és fàcil trobar quins son els problemes ètics de la IA, però no tant, com aquests afectaran a un estadístic.

Després de tot això, queda clar que els estadístics tindran un paper important en, per al que alguns, serà una revolució tecnològica. La IA està present en molts aspectes de la nostra vida, però encara queda molta feina per davant, i cal encara fer-se moltes preguntes. En aquest treball només es plantegen els problemes de l'aprenentatge automàtic supervisat, però caldrà també analitzar els del no supervisat, el per reforç i l'aprenentatge automàtic profund, presentaran aquests algorismes els mateixos problemes? Es podran mitigar els biaixos i els errors? El paper de l'ésser humà passarà algun moment a ser únicament de supervisor? Com evolucionaran les polítiques de privacitat?

Les preguntes que queden per formular-se són encara moltes, com ja s'ha dit el treball volia generar un espai de reflexió sobre quin serà el futur de l'estadístic en la IA. Ara mateix tenen

un paper principalment d'implementació i investigació, però tal i com aquesta estigui més desenvolupada, aquest paper anirà canviant, la responsabilitat serà major, i la bona pràctica de la intel·ligència artificial serà necessària. Els estadístics estan preparats per aquesta responsabilitat? Ara mateix poden assegurar una bona pràctica?

Si els agents són sistemes que pensen com humans, qui diu, que d'aquí a uns anys, els propis agents no siguin capaços d'implementar nous algorismes, d'investigar, solucionar els erros, mesura la seva pròpia bondat, és a dir, fer tot allò que avui en dia fan les persones? Què passarà amb totes les persones que s'hi dediquen? Seran els sistemes capaços de tenir moral? Seran potser més justos que les persones? Ho seran potser menys?

És evident que el creixement de la intel·ligència artificial en els darrers anys ha estat exponencial. Des que Alan Turing va proposar la primera definició de la "màquina pensant" el 1950, van passar diverses dècades fins al desenvolupament de les primeres xarxes neuronals als anys 80. No obstant això, els avenços a partir del 2000 han estat molt més ràpids i constants. Fa només tres anys es va llançar el ChatGPT com una novetat, i ara ja és una eina comuna. Cada pocs mesos hi ha notícies sobre noves IA que repliquen veus, generen cançons i molt més, fent que sigui molt difícil estar al dia.

Aquest ràpid avanç tecnològic ha provocat que molts aspectes importants no es plantegin adequadament, com ara el mal ús potencial de la IA. La inseguretat a Internet és ara més gran que mai, amb estafes i notícies falses (*fake news*) cada cop més habituals, molt difícils de detectar i sense una regulació adequada degut a la velocitat dels avenços. Que pot arribar a passar amb aquelles Intel·ligències Artificials de domini públic, com ara el ChatGPT, on tothom els pot donar informació i ells l'aprenen? Que passa si se li introdueix informació discriminatòria i aquest l'aprèn, de manera que les seves futures respostes son discriminatòries? És perillós no plantejar-se la necessitat d'un pensament crític en aquells encarregats de gestionar i desenvolupar tots aquests avenços.

Aquest estudi és nomes un gra de sorra en una platja, els dubtes que es generen al voltant de la Intel·ligència Artificial són molts, i tot això, que els avenços són encara molt pocs. Cada cop, és més complicat distingir allò que fa una IA d'allò que fa un humà, com he dit les notícies falses, els àudios de veu falsos, imatges falsos, és gairebé impossible distingir els falsos dels vertaders. És clar que s'està elaborant una legislació per tal de regular la IA, però serà suficient per controlar una tecnologia que està al abast de tothom? Que fem amb aquelles persones que

no ho tenen al seu abast? S'ha parlat de la formació ètica cap aquells que desenvolupen i implementen la intel·ligència artificial, però que passa amb totes les persones que la utilitzen? Qui n'assegura el seu bon ús? Que passarà quan comenci a substituir les persones en alguns oficis? De totes les revolucions fins al moment, aquesta és la que avança més ràpid, la revolució industrial va ser molt lenta, la d'Internet va ser una mica més ràpida, aquesta és molt més ràpida i cada cop més, això pot fer que moltes persones quedin enrere, a causa de la dificultat d'estar al dia. Cal entendre que les persones van primer, està bé deixar enrere a aquells que no estan al dia a canvi del progrés? La Intel·ligència Artificial ha de comportar una millora en la vida de les persones, però sense abandonar-ne cap enrere. És molt important tenir en compte totes aquestes consideracions a l'hora de treballar amb la IA, i sobretot com estadístics, els quals coneixen millor que ningú els seus fonaments teòrics i tècnics. La intel·ligència Artificial és un arma poderosa, i el seu potencial és molt alt, però això també comporta un gran risc, està en les nostres mans decidir si volem fer-ne un bon o un mal ús.

IV. REFERÈNCIES

¿Qué es el algoritmo de k vecinos más cercanos? (n.d.). Retrieved June 25, 2024, from IBM website: https://www.ibm.com/es-es/topics/knn

¿Qué es el Aprendizaje mediante refuerzo? - Explicación del Aprendizaje mediante refuerzo - AWS. (n.d.). Retrieved June 25, 2024, from Amazon Web Services, Inc. website: https://aws.amazon.com/es/what-is/reinforcement-learning/

¿Qué es k vecino más cercano (kNN)? (n.d.). Retrieved June 25, 2024, from Elastic website: https://www.elastic.co/es/what-is/knn

Adán, S. (2023, October 18). La inteligencIA artificial: Transformamos nuestro modo de pensar. Retrieved June 22, 2024, from Casa de la Ciencia website: https://www.casadelaciencia.csic.es/es/blog/inteligencia-artificial-transformamos-nuestro-modo-pensar

Amazon. (n.d.). ¿Qué es el sobreajuste? - Explicación del sobreajuste en machine learning - AWS. Retrieved June 25, 2024, from Amazon Web Services, Inc. website: https://aws.amazon.com/es/what-is/overfitting/

Aprendizaje por Refuerzo. (2020, December 24). Retrieved June 25, 2024, from Aprende Machine Learning website: https://www.aprendemachinelearning.com/aprendizaje-por-refuerzo/

Arriagada, M. (2015). Comparación de métricas de distancia en el algoritmo K-Vecinos Más Cercanos para el problema de Reconocimiento ~1.pdf (pp. 7–9). Retrieved from http://opac.pucv.cl/pucv_txt/txt-3000/UCD3128_01.pdf

Asana, T. (2024, February 21). El árbol de decisiones: Un análisis de 5 pasos para tomar mejores decisiones [2024] • Asana. Asana. Retrieved from https://asana.com/es/resources/decision-tree-analysis

Bringsjord, S., & Govindarajulu, N. S. (2024). Artificial intelligence (E. N. Zalta & U. Nodelman, Eds.). Retrieved June 23, 2024, from Stanford Encyclopedia of Philosophy website: https://plato.stanford.edu/entries/artificial-intelligence/#toc

Content Studio. (2022, August 23). ¿Qué es la ética de los datos y de qué modo el almacenamiento puede mejorar las mejores prácticas éticas? Pure Storage. Retrieved from https://www.purestorage.com/es/knowledge/what-is-data-ethics.html

Daniel. (2022, January 25). Random Forest: Bosque aleatorio. Definición y funcionamiento. Retrieved June 25, 2024, from Formación en ciencia de datos | DataScientest.com website: https://datascientest.com/es/random-forest-bosque-aleatorio-definicion-y-funcionamiento

Decision tree algorithm in machine learning - Javatpoint. (n.d.). Retrieved June 25, 2024, from www.javatpoint.com website: https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm

Díez, T. A. (2021). ¿Por qué ética para la Inteligencia Artificial? Lo viejo, lo nuevo y lo espurio. Sociología y Tecnociencia: Revista Digital de Sociología Del Sistema Tecnocientífico, 11(2), 1–16.

Education, I. S. (2020, June 18). Tipos de datos: Datos estructurados, semiestructurados y no estructurados. Retrieved June 22, 2024, from Blog de Tecnología - IMF Smart Education website: https://blogs.imf-formacion.com/blog/tecnologia/tipos-de-datos-datos-estructurados-semiestructurados-y-no-estructurados/

Equipo de Enciclopedia Significados. (2013, September 18). Ética (Qué es, Concepto, Tipos y Ramas). Enciclopedia Significados. Retrieved from https://www.significados.com/etica/

Fernández Casal, R., Costa Bouzas, J., & Oviedo de la Fuente, M. (2021). Aprendizaje estadístico. Retrieved from https://rubenfcasal.github.io/aprendizaje_estadistico/ (Original work published 2021)

GeeksforGeeks. (2021b, January 20). Support vector machine (SVM) algorithm. GeeksforGeeks. Retrieved from https://www.geeksforgeeks.org/support-vector-machine-algorithm/

Gonzalez, D. (2023, February 2). Aprendizaje supervisado. Retrieved June 25, 2024, from Datarmony website: https://www.datarmony.com/blog/aprendizaje-supervisado-algoritmos-ejemplos/

Gonzalez, D. (2023a, January 31). Aprendizaje automático, qué es y cómo funciona. Retrieved June 25, 2024, from Datarmony website: https://www.datarmony.com/blog/aprendizaje-automatico-machine-learning-que-es-y-tipos/?_adin=12095330757

InnovaciónDigital360, R. (2024, February 26). Sistemas expertos, guía completa: Qué es, para qué sirven y clasificación. InnovaciónDigital360. Retrieved from https://www.innovaciondigital360.com/i-a/sistemas-expertos-que-son-su-clasificacion-como-funcionan-y-para-que-se-utilizan/

L. POOLE, D., & K. MACKWORTH, A. (2023). https://artint.info/3e/html/ArtInt3e.html. Retrieved June 22, 2024, from https://artint.info/3e/html/ArtInt3e.Ch1.S1.html

López, M. (2021, December 10). Métodos de bagging y de boosting: ¿Cuál es su diferencia? Retrieved June 25, 2024, from Immune website: https://immune.institute/blog/metodos-de-bagging-y-de-boosting-diferencia/

Madrigal, E. (2022, October 28). Conoce las métricas de precisión más comunes para Modelos de Regresión. Retrieved June 25, 2024, from Grow Up website: https://www.growupcr.com/post/metricas-precision

Malishev, M. (2014). Kant: Ética del imperativo categórico. La Colmena: Revista de La Universidad Autónoma Del Estado de México, (84), 9–21.

Mentes Abiertas Psicología S.L. (n.d.). Qué son los sesgos cognitivos - Psicólogos a tu alcance en Madrid Capital. Retrieved June 23, 2024, from Mentes Abiertas Psicología website: https://www.mentesabiertaspsicologia.com/blog-psicologia/blog-psicologia/que-son-los-sesgos-cognitivos

Pardo, L. (2022, February 10). ¿Qué es y para qué sirve una regresión lineal en machine learning? Platzi. Retrieved from https://platzi.com/blog/que-es-regresion-lineal/

R, S. E. (2021, June 17). Understand random forest algorithm (updated 2024). Retrieved June 25, 2024, from Analytics Vidhya website: https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/

Rae - Asale. (June 25, 2024). moral. In «Diccionario de la lengua española» - Edición del Tricentenario. Retrieved from https://dle.rae.es/moral

Rae - Asale. (June 25, 2024). máxima. In «Diccionario de la lengua española» - Edición del Tricentenario. Retrieved from https://dle.rae.es/máxima

Random Forest, el poder del Ensamble. (2019, June 17). Retrieved June 25, 2024, from Aprende Machine Learning website: https://www.aprendemachinelearning.com/random-forest-el-poder-del-ensamble/

Rodriguez, C. C. (2020b, September 2). Maquina de soporte vectorial (SVM) - César chique rodriguez. Medium. Retrieved from https://medium.com/@csarchiquerodriguez/maquina-de-soporte-vectorial-svm-92e9f1b1b1ac

Roman, V. (2019, April 29). Algoritmos Naive Bayes: Fundamentos e Implementación. Ciencia y Datos. Retrieved from https://medium.com/datos-y-ciencia/algoritmos-naive-bayes-fudamentos-e-implementaci%C3%B3n-4bcb24b307f

Rueda, J. F. V. (2019, August 4). Aprendizaje supervisado y no supervisado. Retrieved June 25, 2024, from healthdataminer.com website: https://healthdataminer.com/data-mining/aprendizaje-supervisado-y-no-supervisado/

Russell, S. J., & Norvig, P. (2004). Inteligencia artificial: Un enfoque moderno (2n ed., pp. 39–40). PRENTICE HALL. Retrieved from http://jdelagarza.fime.uanl.mx/IA/Libros/inteligencia-artificial-un-enfoque-moderno-stuart-j-russell.pdf

Sancho, J. M. C. (2023). Ética discursiva e inteligencia artificial. ¿Favorece la inteligencia artificial la razón pública? Daimon: Revista Internacional de Filosofía, (90).

Sick it-learn Developers. (n.d.). 1.9. Naive Bayes. Retrieved June 25, 2024, from scikit-learn website: https://scikit-learn.org/stable/modules/naive_bayes.html

Sotaquirá, M. (2018, July 18). La Regresión Lineal en el Machine Learning. Retrieved June 25, 2024, from Codificando Bits website: https://www.codificandobits.com/blog/regresion-lineal/

Sotaquirá, M. (2022, January 13). 2 - Ejemplos reales de aplicación del aprendizaje por refuerzo. Retrieved June 25, 2024, from Codificando Bits website: https://www.codificandobits.com/curso/aprendizaje-por-refuerzo-nivel-basico/2-ejemplos-reales-aplicacion-aprendizaje-por-refuerzo/#en-los-juegos-de-mesa-y-videojuegos

Support vector machine (SVM). (n.d.). Retrieved June 25, 2024, from MATLAB & Simulink website: https://www.mathworks.com/discovery/support-vector-machine.html

Tilves, M. (2014, October 28). Los datos con ruido y otros desafíos a los que se enfrenta Big Data. Silicon. Retrieved from https://www.silicon.es/los-datos-con-ruido-y-otros-desafios-los-que-se-enfrenta-big-data-68373