

Càmeres de seguretat amb IA

Treball d'innovació

Intel·ligència Artificial

Àlex Domínguez, Manel Murillo, Joan Sales

June 2023



UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

Facultat d'Informàtica de Barcelona



Índex

1	Introducció	4
2	Història de l'ús d'aquestes càmeres	5
2.1	Funcionament acutal del reconeixement facial	5
2.2	Funcionament actual del reconeixement de patrons	5
3	Bases del Deep Learning	6
3.1	Perceptró o neurona artificial	6
3.2	Xarxes neuronals	7
3.3	Tipus de Xarxes neuronals	7
4	DP en les càmeres de seguretat	9
4.1	Xarxes convolucionals	9
4.2	Xarxes convolucionals reflexives	11
5	Capacitats de les càmeres amb IA	12
5.1	Merodeig	12
5.2	Creuament de límits	12
5.3	Detecció d'intrusió	12
5.4	Detecció de reunions	13
5.5	Reconeixement facial	13
6	Tipus de càmeres	14
6.1	EVA	14
6.2	IVA	14
6.3	Smart Tracking	14
6.4	Smart Video	14
7	Sectors que s'han beneficiat	15
7.1	Smart Cities	15
7.2	Reconeixement facial per a negocis	15
7.3	Logística	15
7.4	Agricultura i ramaderia	15
7.5	Aparcaments	16
7.6	Grans esdeveniments	16
8	Ús per part de països	17
8.1	Xina	17
8.2	Estats Units	17
8.3	Regne Unit	17
8.4	Alemanya	17
8.5	Japó	18
9	Limitacions de la tecnologia	19
9.1	Esbiaixos d'entrenament	19

9.1.1	Definició del problema	19
9.1.2	Recollida de dades	19
9.1.3	Preparació de les dades	19
9.2	Explicabilitat en Deep Learning	19
9.2.1	Exemple de tècnica post-hoc agnòstica	20
9.2.2	Exemple de tècnica post-hoc específica	20
9.3	Mal ús de la tecnologia	21
10	Impacte en la societat	22
11	Bibliografia	23

1 Introducció

El dia 6 de març de 2023 ens vam llevar amb la notícia que el govern francès ha autoritzat l'ús de càmeres de seguretat amb anàlisis de patrons de comportament, a través d'intel·ligència artificial per als Jocs Olímpics de 2024. La notícia va córrer com la pólvora i s'han fet ressò molts mitjans i entitats. Com podem llegir a la notícia publicada per la revista francesa RFI, el dia 21 de març d'aquest mateix any:

“...esta ley preocupa a unos 40 cargos electos europeos que han escrito a los diputados franceses pidiéndoles que se opongan a ella. Los autores de esta carta piden el rechazo de este proyecto de ley, por considerar que permitiría una vigilancia constante y desproporcionada de la población...”. En la mateixa notícia afegeixen: “Esta opinión es compartida por 38 organizaciones de la sociedad civil, entre ellas Amnistía Internacional, que señaló ,... , que ya se ha desplegado en muchos países del mundo para luchar contra la delincuencia, nunca ha aportado la menor prueba de su eficacia.”

Es va obrir el debat en moltes tertúlies, on gent més o menys experta, expressava la seva opinió respecte a la seva efectivitat, la vulneració de les llibertats individuals, el dilema del Big Brother, etc. Ja que, com ha passat amb altres grans canvis en la història, a vegades, un canvi per a una situació puntual, es converteix en permanent.

Com a estudiants hem volgut fer una anàlisi de la tecnologia que hi ha darrere d'aquest avens en la tecnologia de la seguretat, per veure quines són les seves debilitats i les seves mancances, per poder treure una conclusió objectiva respecte a la moralitat de l'ús d'aquestes tècniques.

2 Història de l'ús d'aquestes càmeres

En els inicis, les càmeres de seguretat eren principalment dispositius passius que enregistraven les imatges per a un ús posterior en cas de necessitat. Amb l'avanç de la tecnologia, la visió per computador i l'aprenentatge automàtic van començar a aplicar-se a les càmeres de seguretat per afegir funcionalitats més intel·ligents.

A mitjans i finals de la dècada de 2010, es van començar a utilitzar algorismes d'intel·ligència artificial per a la detecció de moviment, la identificació de persones, la segmentació d'objectes i altres tasques relacionades. Això permet a les càmeres de seguretat alertar o prendre accions en temps real d'acord amb les imatges capturades, com enviar notificacions o activar alarmes en cas d'intrusió o comportament sospitos.

En els últims anys, amb l'augment del poder de càlcul i l'aparició de les xarxes neuronals profundes, s'ha aconseguit un gran progrés en la detecció i el reconeixement d'objectes en temps real en les càmeres de seguretat. S'han desenvolupat arquitectures com les xarxes neuronals convolucionals (CNN) i les xarxes neuronals recurrents (RNN) per a tasques de vigilància i anàlisi de vídeo.

A la dècada del 2010 el reconeixement facial es va popularitzar amb l'aparició de tecnologies com la del desbloqueig facial en els telèfons mòbils i els sistemes de vigilància amb reconeixement facial. Fins aquell moment, els algorismes de reconeixement facial i els de reconeixement de patrons seguien el mateix camí, però va ser aleshores que es van separar.

2.1 Funcionament actual del reconeixement facial

El reconeixement facial va començar a treballar amb algorismes SMV (Support Vector Machines), un algorisme d'aprenentatge automàtic supervisat utilitzat per a la classificació.

2.2 Funcionament actual del reconeixement de patrons

En el cas del reconeixement de patrons, en canvi, es va optar per treballar amb una tècnica de Deep Learning, anomenades xarxes convolucionals reflexives, també conegudes com a xarxes convolucionals recurrents (RCN). És una arquitectura que combina elements de les xarxes convolucionals i les xarxes recurrents per processar dades amb una dimensió espacial i temporal. Aquest tipus de xarxa és especialment útil quan s'han d'analitzar seqüències de dades, com ara en tasques de processament de vídeo, reconeixement d'acció o reconeixement de discurs. La seva estructura permet capturar dependències temporals i patrons seqüencials en les dades d'entrada.

3 Bases del Deep Learning

El Machine Learning i el Deep Learning estan estretament relacionats i són subcampos de la intel·ligència artificial.

Quan parlem del significat de Machine Learning, ens referim a una àrea de coneixement dins de la Intel·ligència Artificial on els ordinadors apliquen tècniques d'aprenentatge estadístic amb l'objectiu d'identificar automàticament patrons en les dades. Els algoritmes generats són capaços de generalitzar comportaments a partir de les dades subministrades en forma d'exemples. La característica principal d'aquest tipus d'algoritmes és que són capaços de reajustar-se automàticament per millorar el seu rendiment en funció del nombre d'encerts i d'errors produïts en un procés d'entrenament previ a la seva aplicació i durant l'execució en temps real d'aquest.

D'altra banda, el deep learning és una subcategoria del machine learning que es basa en xarxes neuronals artificials profundes. Aquestes xarxes neuronals estan compostes per múltiples capes de nodes interconnectats que s'assemblen a les neurones del cervell humà. L'aprenentatge profund utilitza algoritmes que poden aprendre representacions i característiques d'alt nivell directament de les dades, sense necessitat d'una extracció manual de característiques.

3.1 Perceptró o neurona artificial

Un perceptró és un model de neurona artificial que s'utilitza en el camp de l'aprenentatge automàtic i la intel·ligència artificial. Va ser proposat pel psicòleg i científic de la computació Frank Rosenblatt el 1958.

El perceptró es basa en el funcionament bàsic d'una neurona biològica. Rep una sèrie d'entrades, les processa i produeix una sortida. Cada entrada està associada a un pes, que indica la importància o contribució d'aquesta entrada en la sortida final. La neurona també té un llindar, que determina el nivell d'activació necessari perquè la neurona s'activi i produeixi una sortida.

El procés de càlcul en un perceptró és el següent: es multipliquen les entrades pels seus pesos corresponents, se sumen aquests productes ponderats i es compara el resultat amb el llindar. Si la suma supera el llindar, la neurona s'activa i produeix una sortida; si no, roman inactiva i no produeix cap sortida.

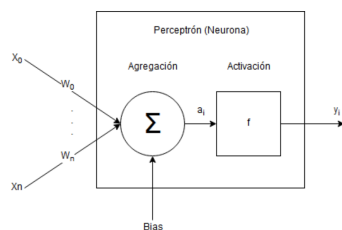


Figura 1: Esquema d'un perceptró

$$Y = f\left(\sum_{i=1}^n w_i x_i\right)$$

Figura 2: Funció de calcul d'un perceptró

Tot i que els perceptrons són models de neurona senzills i lineals, han assentat les bases per al desenvolupament de xarxes neuronals més complexes i potents, com les xarxes neuronals profundes.

3.2 Xarxes neuronals

Les xarxes neuronals artificials estan conformades per 3 tipus de nodes o neurones:

- Nodes d'entrada: reben la informació des de l'exterior de la xarxa (input).
- Nodes de sortida: envien la informació cap a l'exterior de la xarxa (output).
- Nodes ocults: transmeten la informació entre els nodes de la xarxa. Per tant, es troben enmig dels nodes d'entrada i de sortida i no tenen contacte amb l'exterior.

Les xarxes neuronals artificials solen estar conformades per múltiples capes de nodes ocults, a les quals se les denomina "capas d'aprenentatge". A més capes, major és la profunditat de la xarxa i major és la capacitat d'aprenentatge.

En aquest context, els nodes d'entrada reben una sèrie de dades des de l'exterior, aquestes dades són enviades cap a l'interior de la xarxa als nodes ocults. Els nodes ocults van processant, modificant i transferint la informació d'una capa a una altra. Aquest procés és el que es coneix com a "aprenentatge", ja que cada capa de nodes ocults va aprenent de les capes més externes.

3.3 Tipus de Xarxes neuronals

Avui en dia, hi ha diversos tipus de xarxes neuronals artificials que s'han desenvolupat i s'utilitzen en diferents àmbits de l'aprenentatge automàtic i la intel·ligència artificial. Algunes de les xarxes neuronals més comunes són les següents:

- Xarxes neuronals de propagació (feedforward neural networks): Aquestes xarxes són les més bàsiques i estan compostes per múltiples capes de neurones interconnectades. La informació flueix en una direcció, des de la capa d'entrada fins a la capa de sortida, sense realimentació. Són utilitzades per tasques com la classificació, la regressió i el reconeixement de patrons.
- Xarxes neuronals recurrents (recurrent neural networks): Aquestes xarxes tenen connexions cícliques entre les neurones, permetent-los tenir una memòria interna i processar seqüències de dades. Són utilitzades per tasques com el processament del llenguatge natural, la traducció automàtica i el reconeixement de veu.
- Xarxes neuronals convolucionals (convolutional neural networks): Aquestes xarxes estan especialment dissenyades per processar dades amb una estructura espacial, com imatges. Utilitzen capes de convolució per detectar patrons visuals i capes de submostreig per reduir la dimensionalitat. Són molt efectives en tasques com la visió per computador, el reconeixement d'objectes i la segmentació d'imatges.
- Xarxes neuronals generatives (generative neural networks): Aquestes xarxes són capaces de generar mostres noves a partir d'un conjunt de dades d'entrenament. Inclouen models com les xarxes generatives adversàries (GANs) i les xarxes generatives estocàstiques (VAEs). Són utilitzades en tasques de generació d'imatges, síntesi de veu i creació de contingut creatiu.

A part d'aquestes, hi ha altres tipus de xarxes neuronals com ara les xarxes autoassociatives, les xarxes de Kohonen, les xarxes neuronals líquid, entre d'altres. Cada tipus de xarxa neuronal té les seves característiques, avantatges i aplicacions específiques en diferents àmbits.

4 DP en les càmeres de seguretat

La tecnologia que utilitzen les càmeres de seguretat que compten amb reconeixement de patrons, com ja hem esmentat abans, es basen en xarxes neuronals. En un inici es van desenvolupar les xarxes convolucionals tradicionals, però el mètode que s'ha estandarditzat avui en dia, és el que fan servir les xarxes convolucionals reflexives.

4.1 Xarxes convolucionals

Una xarxa neuronal convolucional (CNN) és un tipus de xarxa neuronal que està especialment dissenyada per al processament d'imatges i altres dades amb una estructura espacial similar, com ara senyals de veu o seqüències temporals.

En una CNN, cada neurona d'una capa està connectada només a una petita regió de la capa anterior. Aquestes connexions locals i compartides permeten a la xarxa aprendre característiques locals i invariants espacials, reduint la quantitat de paràmetres necessaris per entrenar la xarxa i afavorint l'extracció de característiques rellevants de la imatge.

A través de la Figura 3 explicarem cadascuna de les capes que estan dins d'una xarxa convolucional. En aquest esquema es descriu com funciona una xarxa convolucional dissenyada per classificar nombres escrits a mà.

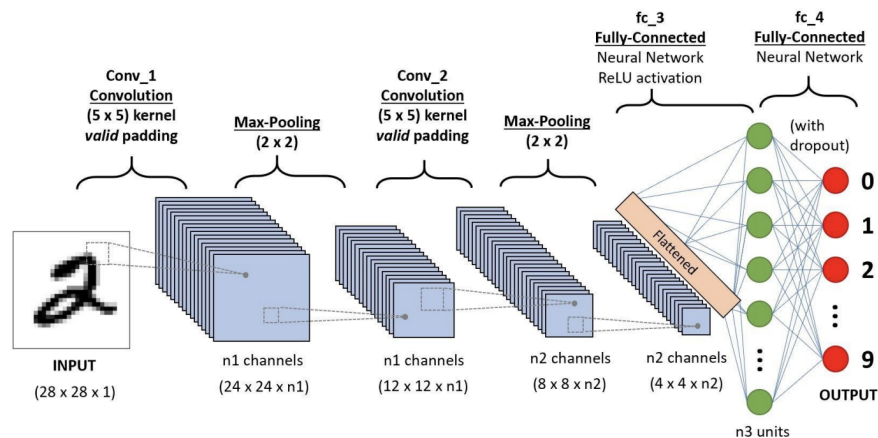
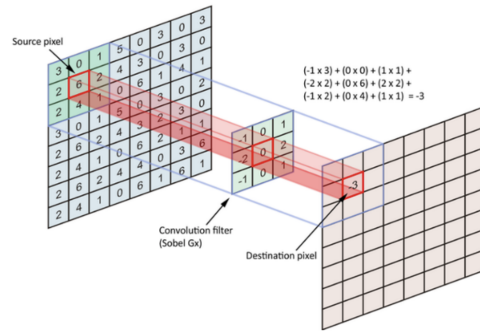


Figura 3: Exemple d'una xarxa convolucional amb dues capes convolucionals i dues de pooling

Observant la Figura 3 d'esquerra a dreta, la primera capa que veiem és una capa convolucional. En aquesta capa s'apliquen dues operacions a la imatge.

En primera instancia, se li aplica un *padding*, és el procés d'afegir vores (píxels o valors) a una imatge d'entrada. Això es fa per preservar la informació en les vores d'una imatge i evitar la pèrdua d'informació.

En segon lloc, se li aplica un *convolutional filter* per fer una combinació lineal dels píxels de l'entrada, de tal manera aconseguir condensar la informació de la imatge d'entrada.



0	0	0	0	0	0
0	35	19	25	6	0
0	13	22	16	53	0
0	4	3	7	10	0
0	9	8	1	3	0
0	0	0	0	0	0

Figura 4: Esquema de l'aplicació d'un *convolutional filter*

Figura 5: Explicació del procés de *padding*

Si continuem recorrent la Figura 3, ens trobem amb la capa de *Pooling*. El procés que posa nom a la capa es un altre tipus de compressió de la informació de la imatge, però en aquest cas molt més agressiu. L'objectiu del "pooling" és reduir la dimensionalitat de les característiques obtingudes a través de l'extracció de característiques locals. Hi ha quatre tipus de *Pooling*: *MaxPooling*, *MinPooling*, *MeanPooling*, *GlobalPooling*. Depenent de l'objectiu que es persegueixi, s'utilitzen diversos un d'aquests quatre mètodes, tot i que el més comú i àmpliament utilitzat és el *MaxPooling*, sobretot en el nostre cas.

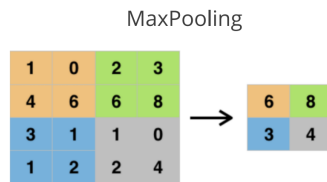


Figura 6: Esquema de l'aplicació del *MaxPooling*

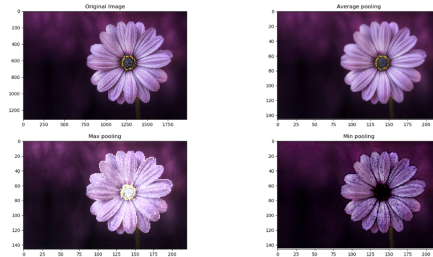


Figura 7: Tractament d'una imatge amb tres tipus de *pooling* diferents

La repetició d'aquest dos tipus de capes esmentades anteriorment ens permet l'extracció progressiva de característiques i la reducció de dimensionalitat i del soroll en la imatge, a més que de la construcció d'una jerarquia de característiques que ajuda a millorar la capacitat de la xarxa per a tasques de visió per computadora.

En últim lloc, el procés de *flatten* consisteix a convertir una matriu de dades multi-dimensional en un vector unidimensional, preparant les dades per a la seva utilització en capes denses de les xarxes neuronals.

4.2 Xarxes convolucional reflexives

La principal diferència entre les dues, és que la xarxa convolucional reflexiva afegeix capes recurrents. Això permet que la informació es propagui en múltiples passades a través dels mateixos pesos, permetent la memòria temporal en les dades d'entrada.

La xarxa convolucional reflexiva té una arquitectura en forma de bucle, on cada pas de temps processa una finestra temporal de les dades d'entrada. Les capes convolucional s'utilitzen per extreure característiques locals en cada pas de temps. A continuació, les capes recurrents processen aquestes característiques i propaguen la informació a través dels passos de temps.

Un dels tipus més comuns de capa recurrent usada en les xarxes convolucional reflexives és la capa de cel·les LSTM (Long Short-Term Memory). Les cel·les LSTM permeten la memòria a llarg termini i poden aprendre a recordar patrons temporals importants en les seqüències de dades.

5 Capacitats de les càmeres amb IA

5.1 Merodeig

Les càmeres poden analitzar patrons de comportament i detectar si una mateixa persona passa recurrentment per davant seu.



5.2 Creuament de límits

Les càmeres poden detectar si alguna persona està superant uns límits arbitraris, com pot ser una propietat privada.



5.3 Detecció d'intrusió

Es pot detectar si una persona que no és habitual, està passant per una zona restringida.



5.4 Detecció de reunions

Aquestes càmeres són capaces d'identificar en multituds, i saber la quantitat de gent que les compon.



5.5 Reconeixement facial

Aquestes càmeres són capaces d'identificar cares i comparar-les amb una base de dades.



6 Tipus de càmeres

6.1 EVA

Les càmeres EVA poden:

- Detectar intrusions en ambients petits, en condicions d'il·luminació controlada o a curta distància
- Detectar i classificar objectes en diferents situacions, incloent-hi persones.
- Contejos simples
- Detectar files, cues i aglomeracions

6.2 IVA

Les càmeres IVA poden:

- Protecció perimetral
- Vigilància en entorns al aire lliure i en condicions climàtiques adverses
- Rangs de detecció grans
- Seguiment de vaixells
- Vigilància en exhibicions i museus
- Conteig de aglomeracions

6.3 Smart Tracking

Aquesta és una altra tecnologia que s'aplica a càmeres per millorar sistemes de videovigilància.

El seguiment intel·ligent es basa en EVA per detectar persones o vehicles.

6.4 Smart Video

Aquestes tecnologies d'IA en les càmeres de seguretat van aconseguir superar l'àmbit de la vigilància per a convertir-se en una font de dades. Poden capturar i analitzar dades convertint-les en informació rellevant.

Aquestes càmeres posseïxen un element informàtic i un dispositiu d'emmagatzematge.

La nova generació d'aplicacions d'IA es denomina Smart Video i utilitza tant l'IA com algoritmes de Big Data per brindar informació.

7 Sectors que s'han beneficiat

7.1 Smart Cities

Sens dubte, una de les majors contribucions de l'IA en les càmeres de seguretat és en la vigilància d'una ciutat. Gràcies a aquest avanç tecnològic, existeixen aplicacions com:

- **Mobilitat.** Estadístiques i mètriques obtingudes mitjançant contejos.
- **Gestió del tràfic.** Detecció i sanció d'infraccions. Detecció d'accidents.
- **Seguretat.** Detecció de vehicles buscats per delictes.
- **Reconeixement.** Tant de persones, com de vehicles, com d'objectes.

7.2 Reconeixement facial per a negocis

El reconeixement facial pot no ser només per qüestions de seguretat. També proporciona informació per accions de màrqueting com:

- Recopilar informació d'edat i gènere.
- Quantitat de visites
- Recorregut dels clients

D'aquesta manera es poden captar tendències, crear targets i en general, oferir millors serveis als clients.

7.3 Logística

La logística és una de les més beneficiades amb la utilització d'aquestes càmeres, ja que a part de seguretat, ofereix:

- Documentació detallada dels processos per identificar errors en una cadena.
- Rastreig de mercaderies.
- Control perimetral dels centres d'emmagatzematge.
- Protecció de les mercaderies i dels empleats.

7.4 Agricultura i ramaderia

Amb drons equipats amb aquest tipus de càmeres es pot:

- Inspeccionar i vigilar àrees grans de terreny.
- Diagnosticar l'estat de la vegetació i animals.
- Controlar l'aigua necessària pels regs.

7.5 Aparcaments

- Es poden determinar les places que hi ha disponibles.
- Tenir control del bon ús de les places d'aparcament.
- Recollir dades sobre les tendències de la gent que l'ocupa.

7.6 Grans esdeveniments

Tant en estadis de futbol com en concerts massius aquesta nova tecnologia aporta les següents avantatges:

- Ajuda amb la seguretat de grans multituds.
- Reconeixement facial i de matrícules.
- Poden detectar situacions de risc
- Registre d'aforament

8 Ús per part de països

La realitat és que moltes de les aplicacions que tenen aquestes càmeres són realment útils per a empreses i particulars. Però quan es tracta de situar càmeres al carrer per part dels estats, això genera una controvèrsia enorme.

Aquesta és només una llista parcial, ja que hi ha molts altres països que també utilitzen càmeres amb intel·ligència artificial a la via pública per a diverses finalitats. És important destacar que l'adopció d'aquestes tecnologies pot variar d'un país a un altre i estar subjecta a regulacions específiques.

Seguidament, citarem un a sèrie de països que ja han aplicat la videovigilància massiva als carrers amb càmeres que incorporen IA.

8.1 Xina

La Xina és un país conegut per utilitzar extensivament càmeres amb reconeixement facial i altres tecnologies d'intel·ligència artificial en la via pública. S'implementen en àrees urbanes, estacions de tren, aeroports i altres llocs per a la vigilància i el control de la població.

Es parla de l'existència d'un sistema de punts on cada ciutadà és considerat millor o pitjor segons la puntuació que té. Aquestes càmeres ajuden a detectar comportaments contraris al règim, com reunions clandestines o delictes menors i a aplicar reduccions en aquest sistema de punts.

8.2 Estats Units

En diversos llocs dels Estats Units s'utilitzen càmeres amb intel·ligència artificial per a la vigilància i el control del trànsit (com per exemple, radars de velocitat). Ciutats com Nova York, Chicago i Los Angeles ja han implementat sistemes de càmeres amb reconeixement facial per a la seguretat pública.

8.3 Regne Unit

El Regne Unit també utilitza càmeres amb intel·ligència artificial en la via pública per a la seguretat i el control del trànsit. Ciutats com Londres tenen sistemes de vigilància àmpliament desplegats amb càmeres amb reconeixement facial.

8.4 Alemanya

Alemanya utilitza càmeres amb intel·ligència artificial per al control del trànsit, la seguretat pública i la prevenció del crim. Hi ha càmeres amb tecnologies avançades com la detecció de matrícules i el reconeixement facial en llocs clau.

8.5 Japó

En diverses ciutats japoneses, s'utilitzen càmeres amb intel·ligència artificial per al control del trànsit, la seguretat i la prevenció del crim. Aquests sistemes fan servir tecnologies com el reconeixement de matrícules i la detecció de comportaments sospitosos.

9 Limitacions de la tecnologia

9.1 Esbiaixos d'entrenament

És habitual culpar a les dades d'entrenament que té un algoritme com a culpable de l'existència dels esbiaixos. Però la realitat és que l'esbiaix pot aparèixer molt abans que les dades es recopilin i una vegada creats aquests esbiaixos són molt complicats de corregir, sobretot pel mateix grup que ha creat l'algoritme.

Per analitzar el fenomen el millor possible, assenyalarem tres etapes clau:

9.1.1 Definició del problema

El primer que es fa quan es vol crear un model d'aprenentatge és decidir quin serà el seu objectiu. En aquesta etapa es pot proposar un objectiu massa difús i que l'algoritme l'interpreti d'una forma que els creadors no contemplaven i esdevingui un problema.

9.1.2 Recollida de dades

Es poden produir dos casos, les dades recollides no són representatives de la realitat o es reflecteixen prejudicis ja existents. Això es podria exemplificar de forma clara amb els diferents tons de pell que existeixen. Una empresa podria agafar una base de dades amb la majoria de les persones caucàsiques, i l'algoritme interpretaria que aquell to de pell és millor o més important que la resta. També pot passar amb el cas de gènere, de la mateixa forma, un algoritme que rebí més homes que dones pot esdevenir en una millor consideració de l'home que de la dona.

9.1.3 Preparació de les dades

És possible introduir esbiaixos a l'hora de seleccionar aquells atributs que es vol que l'algoritme tingui en compte. Elegir quins atributs considerar o ignorar pot influir significativament en la precisió de la predicció d'un model.

9.2 Explicabilitat en Deep Learning

Quan parlem d'explicabilitat en la intel·ligència artificial, i en concret del deep learning, el que estem intentant és intentar trobar formes d'explicar, d'una manera comprensible, la forma d'*enraonar* d'uns algoritmes dels quals veiem les seves entrades i les seves sortides, però que no veiem el procés pel que passa internament per anar d'una a l'altra.

Aquella persona, o grup de persones que creen l'algorisme saben perfectament com funciona allò que han programat, però això no guarda relació directa amb el raonament que executa l'algorisme programat.

Tan creixent és la necessitat d'explicar els models d'intel·ligència artificial, que existeix un camp d'investigació conegut com a XAI (*Explainable Artificial Inte-*

lligence). Aquest camp comprèn tècniques i mètodes orientats a l'explicació de la intel·ligència artificial, tenint en compte tres factors fonamentals:

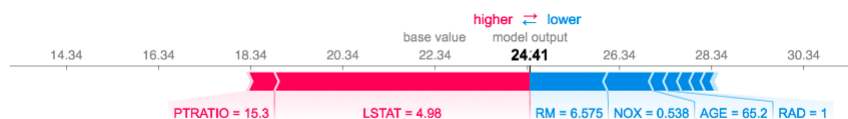
1. **La naturalesa del model a explicar**, que pot tractar-se des d'un model transparent, com pot ser un arbre de decisió, fins a un model opac, com en el cas del deep learning.
2. **L'audiència objectiu**, que en un cas poden ser directius que volen valorar el funcionament de la IA, com usuaris finals que volen conèixer les implicacions de l'ús d'aquesta tecnologia, com la privacitat de les seves dades.
3. **La forma en la qual l'explicació serà fabricada**, que dependrà del grau de coneixement previ sobre el tema i les capacitats cognitives de l'audiència.

L'explicació de models no transparents s'aborda amb l'ús de tècniques XAI post-hoc, capaces de lidiar amb models amb complexitat que impedeix l'enteniment directe del procés pel que prenen decisions, distingim entre:

- **Tècniques post-hoc agnòstiques.** Dissenyades per abordar qualsevol mena de model no transparent sense tenir en compte la seva estructura interna.
- **Tècniques post-hoc específiques.** Són per a un model determinat, el que permet una major profunditat a l'hora d'analitzar-lo.

9.2.1 Exemple de tècnica post-hoc agnòstica

Imaginem un model que observa els valors de certes variables i aprèn a generar un valor a la sortida segons una sèrie d'exemples representatius del problema en qüestió. Utilitzant la tècnica XAI coneguda com a LIME (*Local Interpretable Model-Agnostic Explanations*), el diagrama de forces mostrat a continuació, mostra gràficament la intensitat amb la qual un canvi en una de les entrades al model mou la sortida cap a un cantó o cap a l'altre. Així l'audiència és capaç d'apreciar visualment la importància que té cada una de les variables a l'hora de generar una sortida.



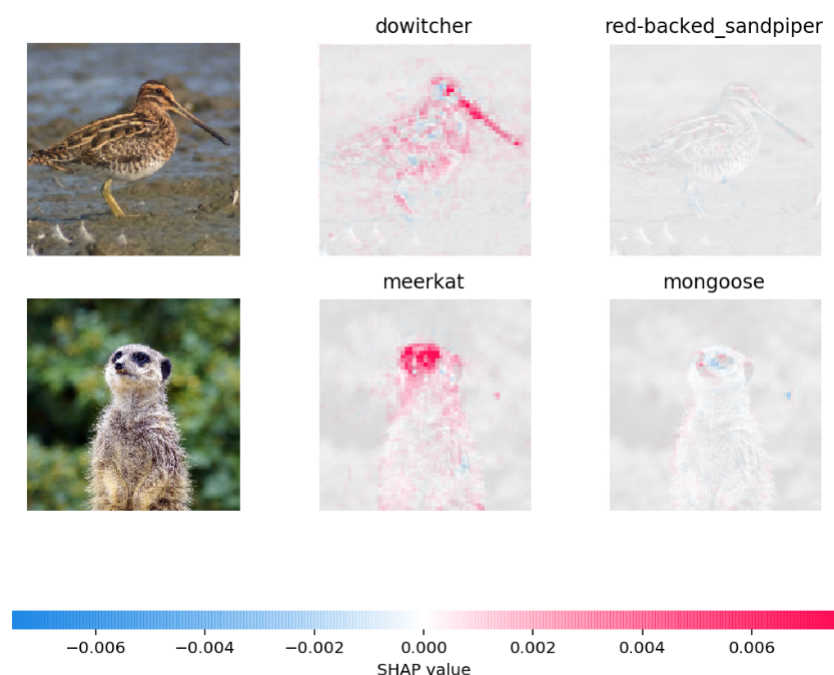
9.2.2 Exemple de tècnica post-hoc específica

En la següent imatge es mostra el resultat de l'aplicació de la tècnica XAI post-hoc coneguda com a PRP (*Pixel-wise Relevance Propagation*) per a un model d'aprenentatge profund capaç de determinar si una fotografia es correspon amb una sèrie d'espècies d'animals.

Per a aconseguir-ho, el model ha sigut alimentat amb fotos que ja especifiquen l'animal que hi apareix en elles.

PRP proporciona un mapa de calor que indica les regions on el model s'ha fixat per produir la seva sortida. En aquest cas es pot veure com el model es fixa en el bec de l'ocell per determinar el tipus d'au de la imatge.

Aquest mapa de calor, serveix per mostrar a l'audiència quin patró segueix el model per identificar cada animal concret i generar confiança en la qualitat de les sortides que genera.



9.3 Mal ús de la tecnologia

Ja hem parlat dels països que disposen d'una gran xarxa de càmeres de videovigilància equipades amb intel·ligència artificial, i això, si no es gestiona correctament, passa de ser una tecnologia que ajuda a la població i fa més fàcils les seves vides a una tecnologia que controla i oprimeix a la gent per les seves idees polítiques, per la seva orientació sexual, per la seva religió, per la seva aparença... Aquesta desconfiança és el que ha portat molta gent a manifestar-se en contra d'aquestes càmeres i el seu ús per part dels governs.

10 Impacte en la societat

Amb l'auge d'aquesta tecnologia arreu del món, tant empreses multinacionals que mouen mils de milions de dòlars a l'any, com petites i mitjanes empreses han començat a utilitzar-la per augmentar la seva seguretat i la seva eficiència.

Les capacitats que donen aquestes càmeres permeten als empresaris tenir un control major sobre el qual està passant a l'empresa, no només en el moment, sinó amb anàlisis mensuals de quanta gent que ha passat per les seves tendes, quina secció és la que més han visitat, o quina edat mitjana tenen els seus compradors més habituals.

Aquestes noves dades, que abans eren molt difícils de recollir, ajuden a l'empresari a augmentar els seus beneficis amb noves tècniques de màrqueting aplicades gràcies a les dades.

En el cas dels estats, aquesta tecnologia comporta un control major sobre la població, i això a la població li aporta pocs beneficis i molts riscos potencials.

Una ciutat plena de càmeres pot comportar una reducció de la criminalitat i un augment en la seguretat de la gent. Però és molta, la gent que rebutja l'ús d'aquesta tecnologia i reclama una legislació forta per part d'organismes com la Unió Europea per regular el seu ús i limitar la informació que poden recollir els estats sobre la gent, sense el seu consentiment.

11 Bibliografía

[1] EuroNews. *París quiere desplegar videovigilancia asistida por IA con motivo de los Juegos Olímpicos de 2024.*

<<https://es.euronews.com/next/2023/02/03/paris-quiere-desplegar-videovigilancia-asistida-por-ia-con-motivo-de-los-juegos-olimpicos>>

[2] SeguriTecnía. *¿Inteligencia Artificial vigilando los Juegos Olímpicos 2024?*

<https://www.seguritecnia.es/actualidad/inteligencia-artificial-vigilando-los-juegos-olimpicos-2024_20230314.html>

[3] Youtube. *Así es como funcionan las cámaras con inteligencia artificial [Medellín Digital] Teled Medellín.*

<<https://www.youtube.com/watch?v=uMwqfMJtLSo>>

[4] LisaInstitute. *Reconocimiento facial: Descubre cómo funciona y quién (y para qué) lo utiliza.*

<<https://www.lisainstitute.com/blogs/blog/reconocimiento-facial-como-funciona-quien-utiliza>>

[5] Juan Barrios. *Redes neuronales convolucionales.*

<<https://www.juanbarrios.com/redes-neurales-convolucionales/#Convoluciones>>

[6] USS. *¿Cómo funciona la Inteligencia Artificial en las cámaras de seguridad?*

<<https://uss.com.ar/tecnologia-y-equipamiento/inteligencia-artificial-en-las-camaras-de-seguridad/>>

[7] ArgSeguridad. *Cámaras de Seguridad con Inteligencia Artificial.*

<<https://site.argseguridad.com/blog/camaras-de-seguridad-con-inteligencia-artificial>>

[8] BusinessInsider. *Las cámaras de reconocimiento facial se están enfrentando a un inesperado obstáculo para poder inundar las calles de todo el mundo: las personas.*

<<https://www.businessinsider.es/ciudades-han-prohibido-camaras-reconocimiento-facial-485891>>

[9] TechnologyReview. *Cómo se produce el sesgo algorítmico y por qué es tan difícil detenerlo.*

<<https://www.technologyreview.es/s/10924/como-se-produce-el-sesgo-algoritmico-y-por-que-es-tan-dificil-detenerlo>>

[10] Wikipedia. *Sistema de reconocimiento facial.*

<https://es.wikipedia.org/wiki/Sistema_de_reconocimiento_facial#Aplicaciones>

[11] IBM. *El modelo de redes neuronales.*

<<https://www.ibm.com/docs/es/spss-modeler/saas?topic=networks-neural-model>>

- [12] Enzyme. *Cómo funcionan las redes neuronales artificiales*.
<<https://enzyme.biz/blog/redes-neuronales-artificiales-y-deep-learning>>
- [13] ThinkBig. *Historia de la IA: Frank Rosenblatt y el Mark I Perceptrón, el primer ordenador fabricado específicamente para crear redes neuronales en 1957*
<<https://empresas.blogthinkbig.com/historia-de-la-ia-frank-rosenblatt-y-e/>>
- [14] Ignacio Gavilan. *Dos métodos de explicabilidad del deep learning y un comentario*.
<<https://ignaciogavilan.com/dos-metodos-de-explicabilidad-del-deep-learning-y-un-comentario/>>
- [15] TecNALIA. *Explicabilidad e inteligencia artificial*
<<https://www.tecnalia.com/blog/explicabilidad-inteligencia-artificial>>
- [16] IEBSchool. *¿Qué es el Machine Learning? Aprendizaje supervisado vs no supervisado*
<<https://www.iebschool.com/blog/que-machine-learning-big-data/>>