UAB
Universitat Autònoma de Barcelona

Dipòsit digital de documents de la UAB

This is the **published version** of the bachelor thesis:

Palma Garro, Jill Areny. *OutfitMatch: Un model que et troba la combinació ideal per destacar amb estil*. Treball de Final de Grau (Universitat Autònoma de Barcelona), 2025 (Enginyeria Informàtica)

This version is available at https://ddd.uab.cat/record/308787

under the terms of the BY-NC-ND license.

OutfitMatch: Un model que et troba la combinació ideal per destacar amb estil

Jill Areny Palma Garro

Resum—OutfitMatch és un sistema de recomanació de moda que utilitza aprenentatge profund per suggerir conjunts de roba estilísticament compatibles. El model combina imatges d'alta resolució i metadades de les peces (color, ús, temporada, etc.) per generar recomanacions precises. Es desenvolupen dues arquitectures: un autoencoder i xarxes siameses amb pseudo-parells. L'avaluació mostra que l'autoencoder excel·leix en la consistència amb regles de moda, mentre que les xarxes siameses aconsegueixen una alta similitud cosinus. A més, es proposen millores futures com la integració de ChatGPT i models avançats com CLIP.

Paraules clau—Recomanació de moda, xarxes neuronals, autoencoder, xarxes siameses, aprenentatge profund.

Abstract—OutfitMatch is a fashion recommendation system leveraging deep learning to suggest stylistically compatible outfits. The model integrates high-resolution images and garment metadata (color, usage, season, etc.) to generate accurate recommendations. Two architectures are developed: an autoencoder and siamese networks. Evaluation results show that the autoencoder excels in rule-based fashion consistency, while siamese networks achieve high cosine similarity. Future improvements include integrating ChatGPT and advanced models such as CLIP.

Index Terms—Fashion recommendation, neural networks, autoencoder, siamese networks, deep learning

1 Introducció - Context del treball

La moda sempre ha estat part integral de l'expressió humana, ja que reflecteix la personalitat, la cultura, i en ocasions, fins i tot l'estatus social. Amb l'auge del comerç electrònic i les compres en línia, la personalització de les recomanacions de moda s'ha convertit en un repte cada vegada més rellevant per a la tecnologia. Tot i que els sistemes de recomanació tradicionals han avançat significativament en el suggeriment d'articles basats en les preferències de l'usuari, la moda, amb la seva complexa interacció de colors, estils i tendències, planteja un desafiament únic. Per això, l'ús de l'aprenentatge profund en les recomanacions de moda ha captat l'atenció del món, oferint la possibilitat d'entendre els patrons d'estil i generar conjunts que no només siguin visualment atractius, sinó també contextualment rellevants.

En aquest projecte, el meu objectiu final és desenvolupar un sistema de recomanació dinàmic que utilitzi o aprengui les regles de compatibilitat de la moda, oferint bones recomanacions que agradin usuaris. El model aprofitarà tant dades visuals (imatges d'alta resolució de peces) com a dades categòriques (atributs com a color, tipus, ús, estació, estil i generé) per a produir suggeriments personalitzats de conjunts de roba.

1.1 Objectius

Com he dit abans l'objectiu principal d'aquest TFG és desenvolupar un model que pugui recomanar peces de roba. Els següents subobjectius estan ordenats en funció de la importancia per aconseguir l'objectiu final:

- **Construir un model capaç de recomanar peces de vestir a partir de l'entrada d'una altra:** Utilitzant imatges i dades categoriques sobre peces de vestir, el model aprendrà quines peces es complementen entre si. Aquesta tasca implica tant el tractament d'imatges com el modelatge de dades i coneixement en sistemes de recomanació avançats.

- **Investigar sistemes de recomanació mitjançant aprenentatge profund:** La recerca aprofundirà en com les xarxes neuronals modernes, en particular les xarxes neuronals convolucionals i els perceptrons multicapa, es poden entrenar i/o utilitzar en els sistemes de recomanació de moda. Per a això serà fonamental comprendre les recerques anteriors i les tendències actuals [2][3].

- **Desenvolupar una interfície/demo on provar el model:** Es desenvoluparà un prototip funcional o una plataforma de demostració per a mostrar el funcionament del sistema de recomanció.

- **Fer comparativa entre la recomanació humana i amb altres models:** Es fara una recol·lecta de outfits amb una prenda especificada fetes per humans i per altres models, i es compara els resultats.

1.2 Metodologia

1.2.1 Desenvolupament de la planificació

En aquest projecte, utilitzaré la metodologia Kanban, ja que permet gestionar de manera visual les tasques pendents i el flux de treball, especialment quan es tracta d'un únic desenvolupador. El tauler de Kanban estarà dividit en quatre columnes principals que ajudaran a monitoritzar el progrés del projecte:

- **Per fer:** Tasques que encara no han començat.
- **En progrés:** Tasques que s'estan realitzant.
- **Bloquejat:** Tasques que han trobat alguna dificultat o dependència externa, no es pot continuar fins que aquesta es resolgui.
- **Completat:** Tasques ja finalitzades.

A més, el tauler comptarà amb quatre tipus de targetes, que reflecteixen les diferents fases del projecte, diferenciades pel seu color:

- **Recerca inicial i exploració del conjunt de dades:** Color vermell.
- **Disseny i desenvolupament del model:** Color blau.
- **Desenvolupament de la demo i proves finals**: Color verd.
- **Redacció i finalització de l'informe final:** Color gris.

Cada targeta inclourà un nom descriptiu de la tasca i la data inici i data límit. També, s'assignarà una prioritat a cadascuna de les targetes:

- **Prioritat Alta:** Tasques que han de ser completades amb urgència per avançar el projecte.
- **Prioritat Media:** Tasques importants, però sense urgència immediata.
- **Priorita Baixa:** Tasques que es poden realitzar en un moment posterior.

Així mateix, faré ús d'un diagrama de Gannt per la planificació a llarg termini. Aquestes dues eines em permetran una gestió flexible i adaptable, fent visible l'estat de cada tasca i ajudant a mantenir un flux de treball constant i sense interrupcions.

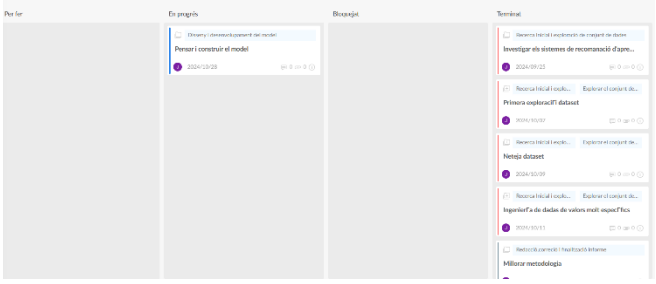


Fig 1 Imatge del KANBAN

1.2.2 Font de Dades

Com a font de dades, el projecte utilitzarà un dels conjunts de dades de productes de moda més utilitzats en el món, *Fashion Product Images Dataset* (la versió completa)[1], aquest presenta una gran varietat de peces juntament amb imatges d'alta resolució i atributs clau com la categoria del producte, el color i l'estil. Tot això, ens proporciona una base ideal per a entrenar un model d'aprenentatge profund capaç de comprendre i predir la compatibilitat de la roba.



Fig 2 Imatges del dataset

A aquestes dades, òbviament, li farem una mica d'enginyeria de dades, que, a grosso modo, consistirà en preprocessar les imatges ,i , al mateix temps, estructurar i netejar les dades tabulars. A més d'extreure que colors poden combinar amb uns altres segons la teoria de color [4] i que peces de roba no poden anar juntes.

1.2.3 Arquitectures del Model

De manera sencilla, l'idea de l'arquitectura final per a aquest projecte és un sistema de recomanció híbrid que combina la entrada d'una imatge i la entrada de metadades de la peça de roba a recomanar. La part CNN extraurà les característiques visuals de les imatges, mentre que el MLP processarà les metadades (color, estil, marca, etc.). Aquests dos components es fusionaran per a formar una representació rica en característiques de la peça, a tot això se'l diu encoder. A partir d'aquí, el model predirà una recomanació adequada, és a dir, quines peces complementen millor l'article d'entrada.

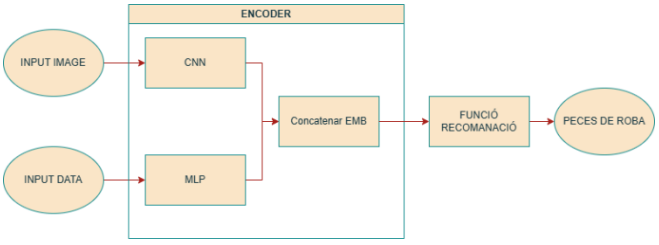


Fig 3 Arquitectura del Model Recomanador

Per entrenar un encoder eficient, explorarem dues estratègies prometedores:

- **Els autoencoders**: que permeten condensar i sintetitzar informació rellevant de les peces.
- **Les xarxes siameses**: pensades per identificar similituds entre diferents elements.

Aquest entrenament també necesitara l'implementació de regles de compatibilitat sintètiques per garantir recomanacions coherents, com assegurar la combinació de colors harmònics i evitar suggerir peces que no tinguin sentit juntes.

1.3 Pla de Treball

Per simplicitat, en el pla de treball només s'especificara les tasques grans. Les subtasques d'aquestes estaran descrites en el diagramma de Gannt.

Recerca inicial i exploració del conjunt de dades (25/09 – 27/10)

- Investigar els sistemes de recomanació d'aprenentatge profund existents, centrant-se en aplicacions relacionades amb la moda.
- Explorar el conjunt de dades i preprocessar les imatges i les dades tabulars, a més de fer l'enginyeria de dades.

Disseny i desenvolupament del model (28/10 – 02/12)

- Pensar i construir el model.
- Integrar regles de moda y adaptar dataset pel model escollit.
- Entrenar el model i avaluar el seu rendiment en conjunts de validació.

Desenvolupament de la demo i proves finals (16/12 – 13/01)

- Desenvolupament d'una interfície d'usuari/demo que permeti als usuaris provar el model.
- Dur a terme proves amb els usuaris per a recaptar les seves opinions, perfeccionar el model i fer una comparativa amb altres models.

Redacció,correció i finalització informe (05/10 – 03/02)

- Redactar l'informes de progrés, fer correccions i redactar informe final.

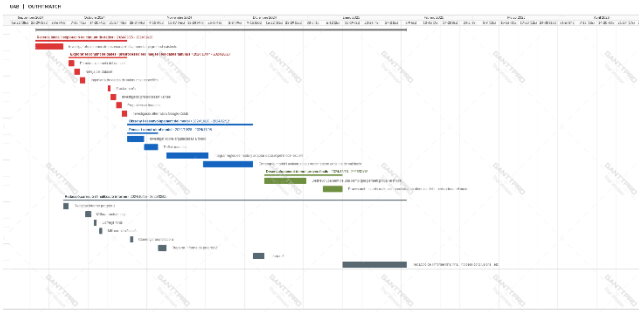


Fig 4 Diagrama de Gantt

2 SISTEMES DE RECOMANACIÓ

Per desenvolupar sistemes de recomanació efectius, és fonamental entendre la base d'aquestos i com aquestos usan l'intel·ligència artificial (IA). Aquesta ultima es la responsable, més concretament l'aprenentatge profund, dels avenços més significatius en aquest àmbit, permetent recomanacions més precises i rellevants. Tot això, ha fet millora als sistemes de recomanció, fent-los capaços d'identificar patrons complexos i captar certs aspectes de les preferències de l'usuari, cosa que fa que els sistemes de recomanació siguin cada cop més adaptatius i personalitzats.

2.1 Tipus de Sistemes de Recomanació i les seves Aplicacions

Hi ha diversos tipus de sistemes de recomanació, cadascun adequat per a diferents necessitats. En aquest apartat, presentaré els enfocaments més comuns a la indústria i la seves aplicacions [5] [6]:

- **Sistemes de Popularitat**: és l'enfocament més bàsic i antic dels sistemes de recomanació. Simplement recomanen els articles més comprats o millor valorats dins d'una categoria. Tot i que és senzill, aquest mètode pot ser efectiu, ja que es basa en la saviesa de la multitud; no obstant això, no té en compte els interessos personals de cada usuari. Per aquesta raó, el seu ús sol ser limitat, encara que segueix sent útil en situacions en què es busca maximitzar la visibilitat de productes populars.

- **Sistemes de Filtrat Col·laboratiu**: utilitzen dades d'usuaris amb comportaments similars per predir les seves preferències. Aquesta tècnica analitza patrons d'interacció entre usuaris i productes per fer recomanacions, com les compres o les valoracions. Hi ha dues variants principals:

  - **Filtrat per Usuari:** Compara un usuari amb altres usuaris de gustos similars per suggerir productes.
  - **Filtrat per Ítem:** Recomana articles basats en la seva similitud amb altres que l'usuari ha valorat o comprat prèviament.

  Encara que el filtrat col·laboratiu és efectiu en moltes aplicacions, pot no ser suficient en situacions on els estils i les combinacions són molt personals. A més, necessita una quantitat considerable de dades d'usuari, per la qual cosa no és la millor opció quan les dades són limitades o quan no hi ha un historial d'interacció.

- **Sistemes de Filtrat Basat en Contingut**: aquest enfocament es centra en les característiques dels productes, en lloc de dependre de les interaccions dels usuaris. Això significa analitzar atributs com el color, el tipus de peça, etc.; Y, el sistema recomanarà articles amb característiques similars als que l'usuari ha vist o valorat. Aquest enfocament és útil en situacions on es vol recomanar sense dependre de dades d'usuari. Tanmateix, té les seves limitacions, ja que captar preferències d'estil complexes (per exemple, combinacions inusuals o preferències personals) pot ser difícil o impossible només amb dades de productes.

- **Models Híbrids**: aquest tipus de sistema combina el filtrat col·laboratiu i el filtrat basat en contingut, aprofita els punts forts dels dos mètodes. Els models híbrids poden, per exemple, iniciar una recomanació basada en popularitat i ajustar-la segons les característiques del producte o el comportament de l'usuari. Aquests sistemes solen ser els més robusts i efectius a la pràctica, ja que minimitzen les

limitacions de cada enfocament quan s'utilitzen de forma aïllada. Tanmateix aquests també són els més complexos i requereixen una major quantitat de dades i processament.

- **Filtrat Contextual**: és el més recent en els sistemes de recomanació i proposa incorporar informació contextual sobre l'usuari i el seu entorn. Aquests sistemes poden recomanar productes tenint en compte factors com el país, la temporada, l'hora del dia o fins i tot la seqüència de productes que l'usuari ha vist o comprat en una mateixa sessió. Son utils quan les preferències poden variar en funció de factors externs a l'atribut i/o l'usari. Per exemple, un sistema de recomanació contextual de roba pot suggerir roba lleugera i de colors clars si detecta que l'usuari es troba en un país amb clima càlid. Aquest enfocament requereix moltes dades contextuals, però pot oferir recomanacions altament personalitzades i oportunes.[5]

Atès que aquest projecte no compta amb dades d'usuari ni dades contextuals, s'emprarà **un sistema de filtrat basat en contingut**.

3 Dataset

Com he dita bans, el dataset base utilitzat en aquest projecte és el *Fashion Product Images Dataset* [1]. Aquest conjunt de dades original conté imatges i metadades detallades de productes de moda, amb categories com tipus de peça, temporada, ús i color. No obstant això, degut a la gran varietat d'elements i a la presència de valors nuls o categories poc representatives, he realitzat una sèrie de passos de neteja i transformació per millorar la qualitat de les dades i l'adequació al model.

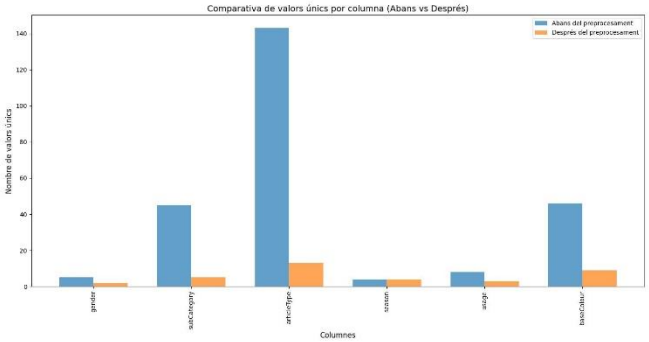


Fig 5 Abans i després dels valors cada columna

3.1 Procés de Neteja i Simplificació

Per preparar el dataset, he pres decisions específiques que facilitarien l'entrenament del model i reduirien la complexitat de les dades. En primer lloc, vaig eliminar certs tipus d'accessoris i peces que podrien confondre el model. Això incloïa elements com tirants o banyadors, que no aportaven valor en el context de predicció desitjat. A més, els valors amb poca representació (menys de cinc instàncies), com els conjunts complets, es van a combinar amb categories més generals o es van a eliminar si no hi havia cohesio amb altres categories. Aquest criteri va ajudar a unificar el dataset i a mantenir únicament categories que oferissin una representació útil.

Un altre ajust realitzat va a ser la simplificació d'algunes categories massa específiques. Per exemple, en el camp *articleType*, vaig unificar els valors "sport sandals" i "sandals" sota "sandals". Aquesta simplificació es va a aplicar de manera consistent en tots els camps rellevants per reduir el soroll en les dades.

A més, vaig reorganitzar els valors que tenien el camp subCategory. La categoria Topwear es va subdividir en dues: Topwear, que inclou shirts i T-shirts, i Over-Topwear, formada per jackets i sweaters. Aquesta divisió es va fer amb l'objectiu de fer més fácil les recomancións.

3.2 Ompliment de Valors Nuls

A més de la neteja, es va dur a terme una fase d'enginyeria de dades per omplir valors nuls en els camps *season*, *usage* i *color*. Com que la quantitat de valors nuls en els camps *usage* i *color* era baixa, aquests es completaren manualment. En el cas de *season*, on el nombre valors nuls era més alt, vaig utilitzar el valor més freqüent en funció del tipus de peça (*articleType*), assignant un valor adequat en funció del context de cada article.

3.3 Preprocessament d'Imatges

Finalment, vaig realitzar un preprocessament de les imatges per preparar-les per al model ResNet50. Aquest preprocessament incloïa ajustos en mida, normalització i format, adaptant cada imatge als requisits d'entrada de l'arquitectura ResNet50. Aquesta etapa assegura que les imatges siguin compatibles i optimés per a l'entrenament del model de recomanació.

3.4 Observacions

En l'anàlisi del dataset, s'han detectat enllaços incorrectes que porten a imatges que no corresponen amb el tipus de peça indicat. No es disposa del nombre exacte d'errors presents, i actualment no hi ha cap manera sencilla de corregir aquesta incidència.

4 Models de entrenament

4.1 AutoEncoder

4.1.1 Concepte

Un autoencoder és una xarxa neuronal que pren dades complexes, com conjunts amplis d'imatges, i les converteix en una representació interna que busca conservar l'essència del contingut original. La idea és que la xarxa aprengui a generar un resum intern sense perdre els elements essencials. Això ho fa intentant que la sortida s'assembli a l'entrada, la xarxa va polint quins elements irrelevants i quins són relevants. Tot i que, en ocasions, les representacions poden resultar semblar complexes d'interpretar, quan tot encaixa, l'autoencoder és capaç d'ordenar aquestos d'una manera comprensible.

L'estructura del autoencoder es compon de dues parts

principals: el codificador, que compacta la informació, i el descodificador, que tracta de recuperar l'aspecte original. Aquesta concepte va més enllà de la mera representació de dades, ja que, per exemple, pot netejar soroll en les imatges, recrear patrons similars als existents o detectar pautes internes sense una guia externa. La qualitat final del resultat depèn del tipus de dades, de l'arquitectura emprada i dels objectius plantejats.7][8]

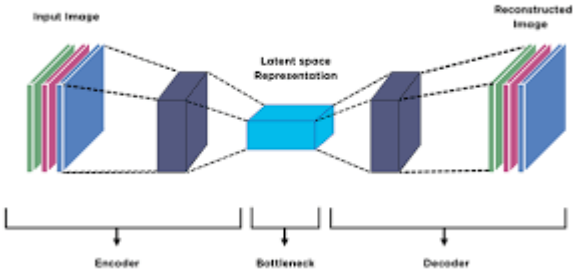


Fig 6 AutoEncoder

4.1.2 OutifMatch: AutoEncoder

En el meu cas, l'autoencoder no es limitara en processar imatges de peces de roba, sinó que també incorporara dades tabulars que descriuen cada article. Per exemple, en el cas d'una jaqueta, doncs, el model generara una represnetació interna que combina detalls visuals amb atributs com categoria, color o temporada. Aquesta unió creara una base que orienta la recomanació d'altres peces, intentant que encaixin amb el producte inicial.

També, per la reconstrucción visual, el descodificador aprofita una arquitectura del tipus U-Net. Aquesta configuració interna, amb els seus enllaços entre capes, recupera detalls subtils que tendeixen a perdre's. D'aquesta manera, l'autoencoder no només condensa una peça, sinó que també reconstrueix la seva imatge amb un nivell de detall superior. D'altra banda, el codificador visual es basa en una ResNet50, reconeguda per la seva capacitat per extreure trets complexos, cosa que aporta a la representació un sentit més profund de la peça.

Amb la unió de l'arquitectura visual i una arquitectura sencilla per les dades tabulars, l'autoencoder combinat sembla que assoleix una sintonia interessant entre la informació visual i la numèrica, aconseguint recomanacions amb més coherència.

Durant l'entrenament, les pèrdues marquen el camí a seguir. En comptes de comparar píxels, resulta més fructífer emprar pèrdues perceptuals basades en xarxes ja entrenades, com la ResNet50. Aquestes pèrdues capten qualitats visuals més internes, incrementant la qualitat de la reconstrucció sense quedar-se amb diferències molt superficials. Per a les dades tabulars, una pèrdua del tipus *Sparse Categorical Crossentropy* ajusta millor l'assignació de categories, guiant la xarxa a representar la informació sense desorientar-se.

Per el que es el sistema de recomanció, després de codificar l'entrada, el sistema cerca al catàleg les peces més similars. Al mateix temps, aplica regles sobre el tipus d'ús, el gènere, l'estació, la classe d'article i el color. Aquesta combinació, entre informació numèrica, detalls visuals i condicions específiques, encamina cap a recomanacions més consistents. Si una opció no encaixa, es descarta i es passa a la següent, formant així conjunts amb més harmonia. No hi ha una fórmula màgica, però aquest mètode obre una via per a resultats més estables, barrejant intuïció humana (regles tabulars) amb la comprensió interna del model.

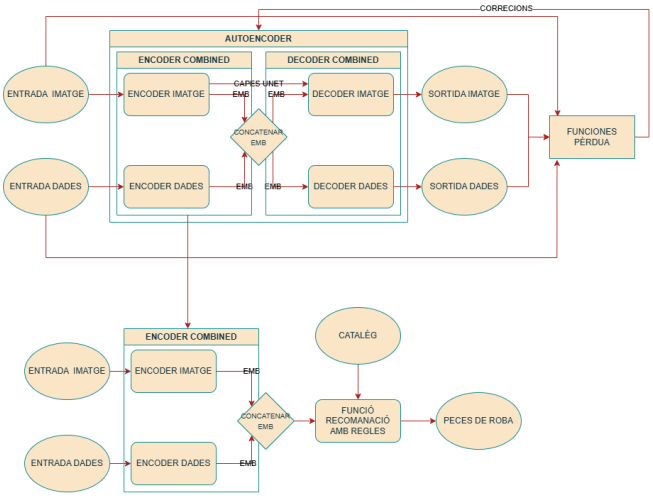


Fig 7 OutfitMatch: AutoEncoder

A continuació, presento una foto d'un exemple d'outfit generat pel meu model autoencoder, que demostra com la combinació de diverses peces del model resulta en un conjunt estilísticament cohesionat i atractiu

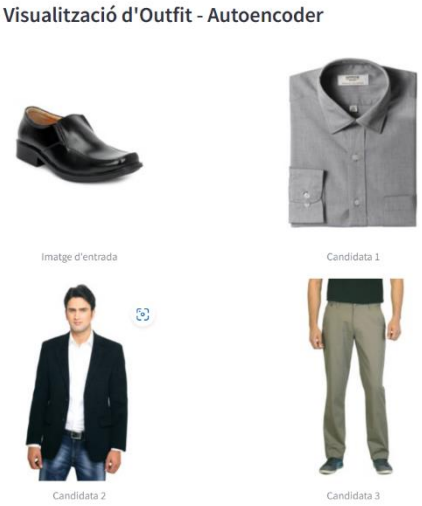


Fig 8 Exemple OutfitMatch: AutoEncoder

4.2 Xarxes Siameses

4.2.1 Concepte

Les xarxes siameses neixen amb la finalitat de mesurar semblances entre parelles de dades. Aquestes arquitectures tenen dues branques bessones (dos encoders), cadascuna processant una entrada diferent, i comparteixen els mateixos paràmetres per generar vectors representatius. Mitjançant el càlcul de la distància entre aquests vectors, es pot determinar si dos elements estan relacionats d'alguna manera. Aquest mecanisme resulta especialment útil en contextos com el reconeixement facial, la verificació de signatures o la comparació de productes. [9][10][11]

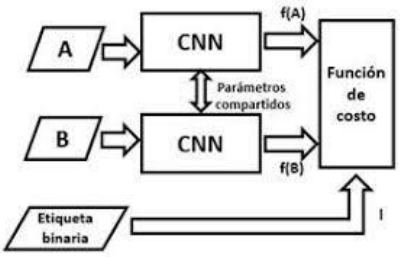


Fig 9 Xarxes Siameses

4.2.1 OutifMatch: Xarxes Siameses

En el nostre context, l'encoder, que inclou una ResNet-50 preentrenada per obtenir trets visuals, combina aquestes característiques amb atributs tabulars a través d'una MLP. Com a resultat, cada peça es projecta en un espai vectorial que integra aparença, textura, estació i altres factors rellevants.

Per generar els pseudo-parells necessaris durant l'entrenament, s'assigna un percentatge de compatibilitat a cada combinació de característiques entre peces (com ara estació, color o ús). Posteriorment, es calcula una mitjana ponderada d'aquests percentatges per determinar la compatibilitat global del parell. Si aquesta mitjana supera un llindar predefinit, el parell es classifica com a compatible; en cas contrari, s'etiqueta com a no compatible.

L'entrenament de la xarxa siamèsa inclou cicles continus on els pesos s'actualitzen amb cada lot de parells. Després de moltes iteracions, l'encoder es fa capaç de generar embeddings representatius, cosa que permet que el motor de recomanacions sorgeixi amb agilitat en la selecció de peces a partir de catàlegs molt amplis. El resultat final és un sistema que, a partir de la regla de que no em doni dos peces de la mateixa categoría i un model entrenat, troba combinacions coherents, integrant coneixement visual i atributs específics sense necessitat d'intervenció manual constant.

El sistema de recomanció com tal, parteix d'una peça inicial: l'encoder genera el seu embedding. A continuació, es calcula la distància cosinus respecte a altres opcions del catàleg, i les més afins superen un primer filtre. En triar una segona peça, el seu vector s'incorpora al conjunt i s'obté una mitjana que actualitza l'embedding global de l'outfit, reflectint la compatibilitat col·lectiva.

Aquest mètode no només se centra en la mesura de proximitat vectorial. Com he dit abans, hi ha regles que s'encarregen de filtrar categories per evitar redundàncies. Per exemple, si hi ha un vestit al conjunt, es prefereix no afegir peces que saturin la mateixa funció. Així, el procés no es limita a sumar ítems que s'assemblin a la peça inicial, sinó que construeix un conjunt més coherent.

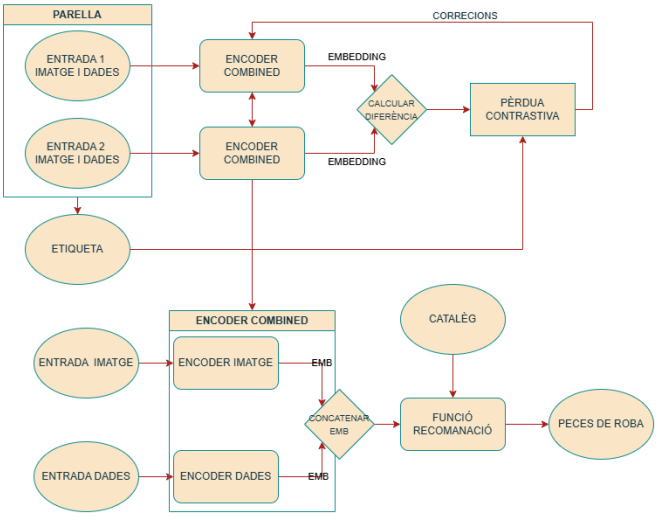


Fig 10 OutfitMatch: Xarxes Siameses

A continuació, presento una foto d'un exemple d'outfit generat pel meu model, que il·lustra com la combinació de diverses peces seleccionades pel model resulta en un conjunt estilísticament cohesionat:



Fig 11 Exemple OutfitMatch: Xarxes Siameses

Aquest exemple demostra la capacitat del sistema per integrar diferents atributs i trets visuals, creant combinacions compatibles.

4.3 Plan d'avaluació

4.3.1 Consistència amb Regles

Disseny

- **Preparació**: Definim regles específiques per a les combinacions, com emparellar peces formals amb altres formals o colors que es complementin.
- **Avaluació**: El model genera combinacions que després comparem amb les nostres regles predefinides. Calculem la taxa de compliment dividint les recomanacions que segueixen les regles entre el total de recomanacions.

4.3.2 Similitud de cosinus entre peces

Disseny

- **Preparació**: Definim un conjunt de prova i generem embeddings per a cada peça utilitzant el model.
- **Avaluació**: Per a cada peça en el conjunt de prova, el model suggereix les combinacions més compatibles. Calculem la similitud cosinus entre els embeddings de les peces combinades i la peça de prova, verificant si estan dins d'un llindar de similitud.
- **Resultats**: Percentatge de recomanacions que compleixen aquest llindar, proporcionant una mesura de precisió de les nostres suggerències.

5 AVALUACIÓ

5.1 Objectiu

En aquest apartat es vol veure fins a quin punt les diferents aproximacions de recomanació de moda dels models funcionen correctament. S'observarà si els conjunts suggerits segueixen les regles definides, si es mantenen prou a prop de la peça de referència i com es distribueixen els embeddings. Cal dir que la descripción del proces d'avaluació esta definida en el apartat 4.3 i que s'han fet servir les mateixes dades de prova en tots els casos.

5.2 Consistència amb regles

S'ha recollit la taxa de combinacions acceptables (llindar: 0,6) segons les normes de moda.

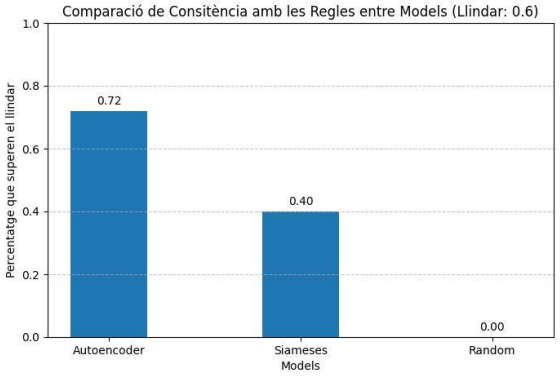


Fig 12 Consistència amb regles

Les xifres mostren que l'autoencoder obté un percentatge molt destacat, mentre que les Xarxes Siameses se situen al voltant del 40%. Tot i trobar-se lluny del nivell de l'autoencoder, aquest resultat és clarament superior a les propostes aleatòries, que no superen ni el 1%.

Si veiem la distribució de la consistència de regles dels models podem veure que els percentatges de combinacions acceptables tenen sentit, i encara que sembli que els aleatoris tenen una bona distribució tenim que tindre en compte que les regles estan amb probabilitats de que combinin, on el màxim es 1, encara que sigui molt difícil que passi del 0.8, i el mínim es 0 ,encara que sigui molt difícil de que passi del 0.2, tot això però la assignació de probabilitats de combianció entre valors.

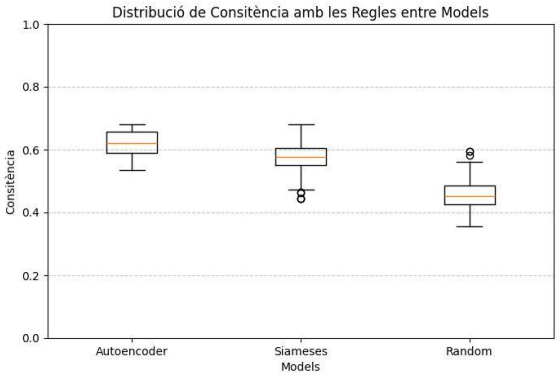


Fig 13 Distribució de la consistència amb regles

5.3 Similitud de cosinus entre peces

En aquesta metrica s'avalua la similitud cosinus entre els vectors generats i la peça original.

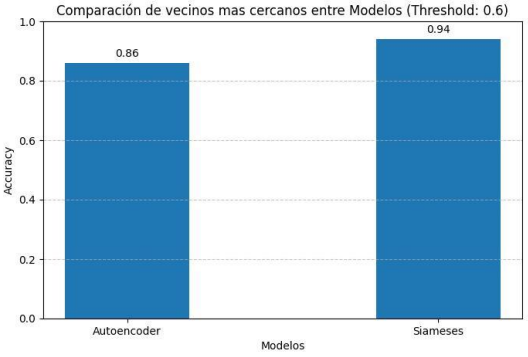


Fig 14 Similitud de cosinus entre peces

Es pot veure que l'autoencoder arriba a un 86%, mentre que les Xarxes Siameses assoleixen fins al 94%. Això indica que els dos models capturen les característiques internes que fan que les peces suggerides siguin compatibles amb la peça inicial.

Si veiem la distribució de la similitud cosinus podem veure que els resultats tenen sentit, però també veiem que hi ha alguns valors atipics en el model de les Xarxes Siameses.

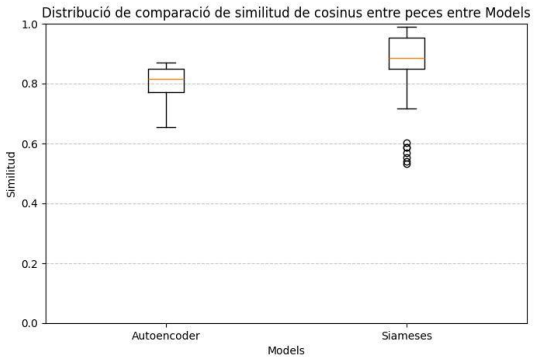


Fig 15 Distribució de similitud de cosinus entre peces

5.3 Visualització d'Embeddings

Aquest apartat pretén oferir una perspectiva més clara sobre la manera com l'Autoencoder i les Xarxes Siameses distribueixen les peces de roba en l'espai latent. Amb la representació en 2D, es pot identificar fins a quin punt les peces similars queden properes entre elles o si s'escampen en regions més y túdisperses.

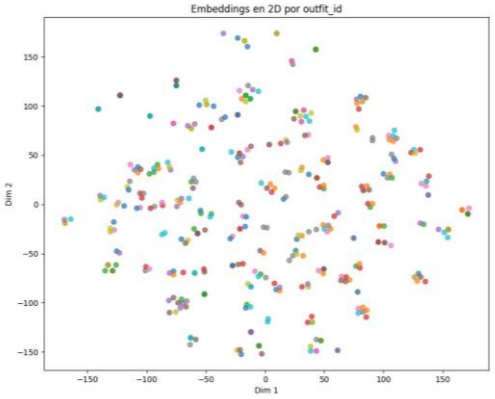


Fig 16 Embeddings AutoEncoder

La tendència és agrupar peces, que encara que no siguin del mateix outfit, tenen característiques semblants, fet que indica un cert grau de coherència interna. Això pot afavorir la cerca de conjunts que segueixin una lògica més consistent.

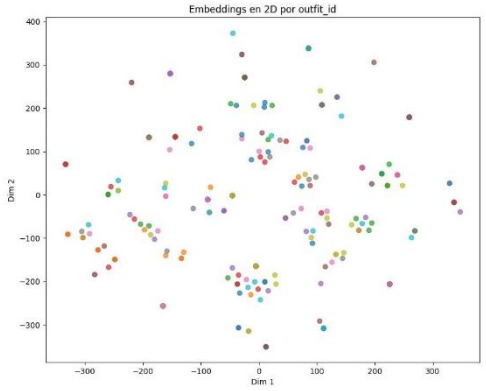


Fig 17 Embeddings Xarxes Siameses

La distribució és semblant a la de l'autoencoder pero menys compacta, fet que dóna cabuda a conjunts diversos que, tot i això, no es dispersen completament. Això pot ser útil si es volen recomanacions amb un ventall ampli d'estils. També sembla que hi han menys varietat de articles en els outfits.

5.4 Conclusions de l'Avaluació

Els resultats indiquen que l'autoencoder presenta un alt nivell de coincidència amb les regles i aconsegueix una similitud destacable en comparació amb els altres models.

Pel que fa a la consistència amb les normes de més de un 60%, supera de manera clara la resta de mètodes, mentre que les Xarxes Siameses es situen en un 40%, una dada que, malgrat ser inferior a la de l'autoencoder, queda per sobre de les combinacions aleatòries (amb percenetage inferior al 1%), tot i que no usen les regles directament com l'autoencoder. I, encara que la distribució sigui bastant properes entre models, hem de tindré en compte que les regles estan fetes perquè la probabilitats no sigui quasi mai més alta de 0.8 ni menys baixa de 0.2.

Pel que fa a la similitud cosinus, les Xarxes Siameses aconsegueixen que un 94% superi una similitud de un 60%, superant a l'autoencoder (86%) en aquesta mètrica. La distribució dels models en aquesta mètrica ens diu que les Xarxes Siameses genera combinacions de peces amb similituds més altes en promig i amb menys variabilitat, però amb alguns valors atípics.

En relació amb la disposició dels embeddings, l'autoencoder acostuma a generar agrupacions més concentrades de peces de roba amb caracteristiques semblants, mentre que les Xarxes Siameses presenten un patró menys compacte però igualment identificable.

I, tenint en compte que totes les proves s'han realitzat amb el mateix conjunt de dades, podem dir que la comparació és coherent i permet afirmar que l'autoencoder destaca en el compliment de les normes, i que les Xarxes Siameses mostren un valor alt en la semblança entre peces.

5.5 Observacions dels models

Les Xarxes Siameses tendeixen a repetir peces de roba en diferents combinacions, però, al meu parer, sense generar combinacions dolentes. En canvi, l'Autoencoder no tendeix a repetir peces de roba i es centra més en les regles(encara que aquestes només haurien de ser una guia), fet que, a vegades, pot donar lloc a combinacions dolentes, segons el meu criteri.

4 CONCLUSIÓ

El projecte OutfitMatch ha complert la majoria dels objectius proposats, desenvolupant un sistema de recomanació de moda multimodal que combina dades visuals i categòriques per generar outfits estilísticament coherents. A continuació, es destaquen els principals assoliments, aprenentatges del projecte i millores futures:

1- Compliment d'objectius:

- S'han desenvolupat models amb la base híbrida plantejada, xarxes neuronals convolucionals (CNN) i perceptrons multicapa (MLP), aconseguint un sistema capaç de recomanar peces compatibles amb una entrada inicial.
- S'ha investigat en profunditat l'ús de sistemes de recomanació basats en aprenentatge profund, avaluant l'efectivitat d'arquitectures com els Autoencoders i les Xarxes Siameses.
- S'ha creat una demo funcional per provar els sistemes en un entorn pràctic.
- S'han dissenyat i aplicat mètriques d'avaluació per valorar la consistència amb regles i la similitud entre peces dels models.
- No obstant, no s'ha pogut fer una comparativa amb recomanacions humanes a causa de les limitacions de temps.

2- Avaluació de models:

- **Autoencoder:** Aquest model ha destacat per la seva capacitat de generar recomanacions que complien, en almenys un 60%, amb les regles predefinides de moda, assolint un 72% de consistència dita.
- **Xarxes Siameses:** Aquest enfocament ha mostrat un rendiment excel·lent en la mètrica de similitud cosinus, arribant al 94%. Això evidencia la seva capacitat per capturar i representar les característiques intrínseques de les peces.
- **Resultats generals:** Ambdós models han demostrat capacitat per generar conjunts estilísticament harmònics, almenys en la teoria. La integració de regles de compatibilitat ha contribuït, directament o indirectament, a recomanacions relativament coherents i pràctiques.

3- Limitacions identificades:

- La manca de temps ha impedit implementar i avaluar una comparativa amb recomanacions humanes i una possible millora dels models.
- Accés a informació actualitzada i comprensible: Algunes fonts sobre intel·ligència artificial, especialment relacionades amb TensorFlow, presentaven explicacions complexes o informació desactualitzada, fet que ha dificultat aprofundir en alguns temes.

4- Millores futures:

Per tal de millorar el sistema, es suggereixen les següents millores:

- **Incorporar recomanacions múltiples i ChatGPT**: Adaptar els models per tal que generin les N millors opcions de peces compatibles. Aquestes opcions podrien ser enviades a un model de llenguatge com ChatGPT, que aplicaria criteris estilístics més complexos o personalitzats per seleccionar l'outfit final. Aquesta millora no es va implementar per manca de temps, però presenta un gran potencial per enriquir l'experiència de l'usuari.
- **Explorar models més avançats:** Models preentrenats com CLIP podrien millorar la qualitat de les recomanacions.
- **Validació amb usuaris reals:** Realitzar estudis comparatius entre les recomanacions del sistema i les d'estilistes humans per avaluar-ne l'efectivitat des d'una perspectiva pràctica.

AGRAÏMENTS

Vull expressar el meu profund agraïment a totes les persones e institucions que han contribuït al desenvolupament d'aquest projecte:

- A Sergio Bachiller Rubia, el meu tutor/a, pel seu suport, orientació i paciència al llarg de tot el procés. La seva experiència han estat fonamentals per a la finalització d'aquest treball.
- Als meus companys i amics del grau d'Enginyeria Informàtica , per la seva ajuda, ànims i suport emocional en els moments més difícils.
- A la meva família, especialment als meus pares i germans per creure sempre en mi i proporcionarme el suport necessari per seguir endavant.
- A la Universitat Autònoma de Barcelona, per oferirme l'oportunitat d'aprendre i desenvolupar-me professionalment en aquest entorn acadèmic tan enriquidor.

Finalment, vull agrair a tots els que, d'una manera o altra, han contribuït a fer realitat aquest projecte. Gràcies per fer aquest camí més lleuger i ple d'aprenentatges.

BIBLIOGRAFIA

[1] Fashion Product Images Dataset. (2019, 14 marzo). Kaggle. https://www.kaggle.com/datasets/paramaggarwal/fashion-product-images-dataset

[2] Gupta, M. (2024, 1 febrero). Recommendation Systems using Neural Collaborative Filtering (NCF) explained with codes. Medium. https://medium.com/data-science-in-your-pocket/recommendation-systems-using-neural-collaborative-filtering-ncf-explained-with-codes-21a97e48a2f7

[3] El sistema de recomendación de moda del que Cruella de Vil estaría orgullosa - Insight | Keyrus. (s. f.). https://keyrus.com/sp/es/insights/el-sistema-de-recomendacion-de-moda-del-que-cruela-de-vil-estaria-orgullosa

[4] Velázquez, M. G. C. (2024, 21 agosto). Cómo combinar los colores de tu outfit aplicando la teoría del color. HANEK. https://hanek.com/blogs/mundo-hanek/teoria-del-color-para-vestir

[5] What is a Recommendation System? (s. f.). NVIDIA Data Science Glossary. https://www.nvidia.com/en-us/glossary/recommendation-system/

[6] Na, & Na. (2020, 3 marzo). Sistemas de recomendación | Aprende Machine Learning. Aprende Machine Learning. https://www.aprendemachinelearning.com/sistemas-de-recomendacion/

[7] Codificando Bits. (2024, 26 agosto). Autoencoders ¡EXPLICADOS! [Vídeo]. YouTube. https://www.youtube.com/watch?v=ceZD_9vlGew

[8] Introducción a los codificadores automáticos. (s. f.-b). TensorFlow. https://www.tensorflow.org/tutorials/generative/autoencoder?hl=es-419

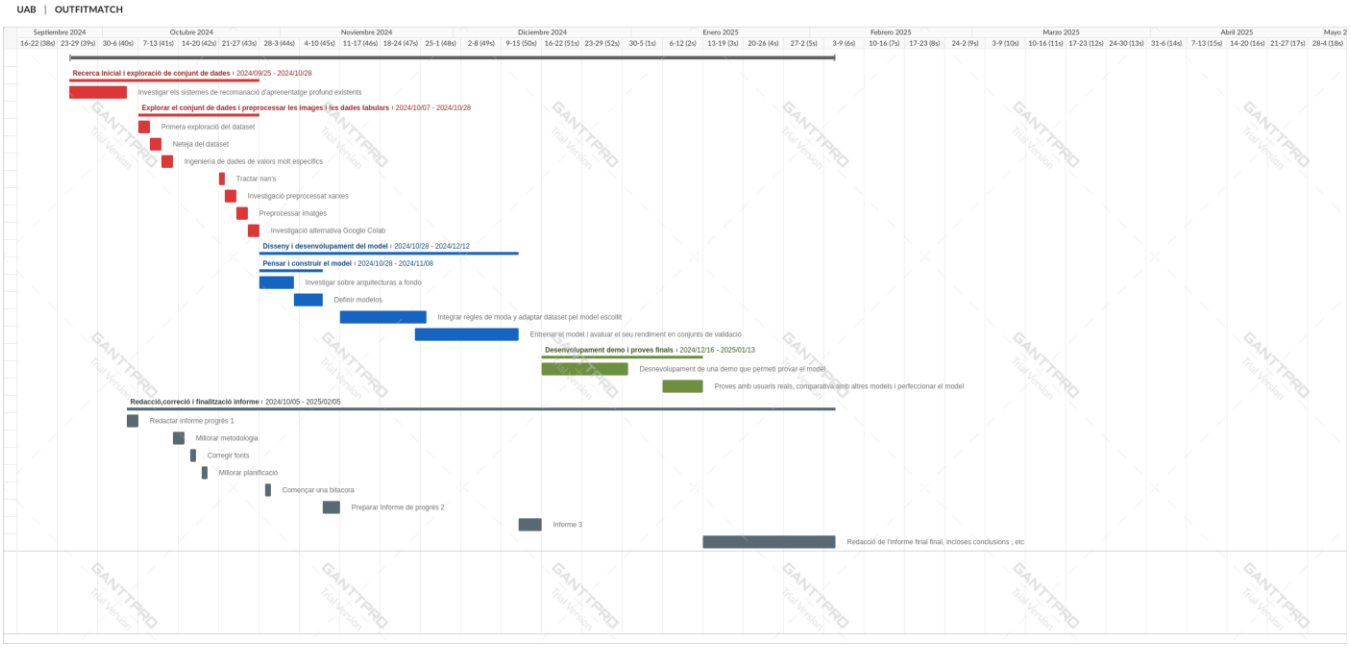
[9] ReconoSER ID. (2020, 7 octubre). Webinar | Redes Siamesas: aplicación de Deep learning para validación de identidad | ReconoSER ID [Vídeo]. YouTube. https://www.youtube.com/watch?v=jizv7MWXTVQ

[10] Cómo crear una Siamese Network para comparar imágenes, con TensorFlow. – Pere Martra. (s. f.). https://martra.uadla.com/como-crear-una-siamese-network-para-comparar-imagenes-con-tensorflow/

[11] Beastieux. (2023, 16 octubre). TensorFlow / Python: Comparación de Imágenes con Redes Neuronales Siamesas. Beastieux. https://beastieux.com/2022/03/17/tensorflow-python-comparacion-de-imagenes-con-redes-neuronales-siamesas/

APÈNDIX

A1. DIAGRAMA DE GANTT



A2. IMATGES STREAMLIT

