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

Hernández Garrido, Víctor; Antens, Coen Jacobus , dir. Portal immobiliari basat en IA per a l'optimització de procediments i detecció d'oportunitats d'inversió. 2024. (Enginyeria Informàtica)

This version is available at <https://ddd.uab.cat/record/298969>

under the terms of the  license

Portal immobiliari basat en IA per a l'optimització de procediments i detecció d'oportunitats d'inversió

Víctor Hernández Garrido

Resum — Aquest article explora el desenvolupament d'una portal immobiliari web enriquit per la Intel·ligència Artificial, amb un fort enfocament en millorar tant l'eficiència de la gestió per als propietaris com l'experiència dels usuaris a l'hora de buscar bones llars i/o oportunitats d'inversió. Destaca l'ús de tecnologies com OpenCLIP i ChatGPT per generar descripcions detallades a l'hora de crear anuncis de les propietats, i l'aprofitament d'un dataset extens per a la predicció de preus mitjançant l'algorisme de Machine Learning, Random Forest. A més, s'analitza la integració de sistemes de recomanació personalitzada per millorar la interacció dels usuaris, contribuint a una experiència més intuïtiva i eficaç.

Paraules clau— Intel·ligència Artificial, Machine Learning, Plataforma Immobiliària, OpenCLIP, Model de Subtítols d'Imatges, ChatGPT, Predicció de Preus, Random Forest, Recomanació Personalitzada.

Abstract — This article explores the development of a web real estate portal enriched by Artificial Intelligence, with a strong focus on improving both management efficiency for property owners and the user experience in finding suitable homes and/or investment opportunities. It highlights the use of technologies such as OpenCLIP and ChatGPT to generate detailed descriptions when creating property listings and leverages an extensive dataset for price prediction using the Random Forest machine learning algorithm. Furthermore, it analyses the integration of personalized recommendation systems to enhance user interaction, contributing to a more intuitive and effective experience.

Index Terms — Artificial Intelligence, Machine Learning, Real Estate Platform, OpenCLIP, Image Captioning Model, ChatGPT, Price Prediction, Random Forest, Personalized Recommendation.



1 INTRODUCCIÓ - CONTEXT DEL TREBALL

Trobar bons pisos per llogar a Barcelona amb una relació qualitat-preu adequada és cada vegada més complicat... [1] Per això, he creat aquesta plataforma innovadora que ajuda als inversors i clients a trobar oportunitats en llars amb bons preus. La plataforma no només simplifica la recerca de propietats, sinó que també ofereix eines i recursos per prendre decisions informades.

Tant si ets un inversor que busca rendibilitzar el seu capital, un client que vol trobar la llar ideal o un propietari d'un portal immobiliari que vol optimitzar el seu negoci, aquí trobaràs tot el que necessites per fer una bona elecció. Amb un mercat immobiliari tan competitiu, tenir accés a informació actualitzada i fiable és essencial, i això és justament el que s'ofereix.

El projecte se centra en la creació d'una pàgina web moderna i funcional, utilitzant Angular pel Front-End i Node.js pel Back-End. Integrant tecnologies innovadores, l'objectiu és oferir un servei eficient i complet que beneficia tant als inversors com als propietaris de portals immobiliaris.

El portal immobiliari desenvolupat es presenta com un software modular que integra diversos plugins clau. Un dels plugins destacats és la **generació d'anuncis automàtics mitjançant IA**. Aquesta funcionalitat permet als propietaris del portal carregar les imatges de les seves propietats mentre la IA genera automàticament títols, descripcions i preus suggerits per als anuncis. Aquest procés no només optimitza els recursos, sinó que també assegura la coherència i qualitat de la informació presentada.

A més, el portal inclou una **comparativa de propietats mitjançant IA**, que permet als usuaris comparar de manera eficient dues propietats seleccionades. Aquesta funcionalitat utilitza algorismes d'Intel·ligència Artificial per generar informació detallada i significativa, facilitant així la presa de decisions informades i ràpides.

-
- E-mail de contacte: 1597028@uab.cat
 - Menció realitzada: Enginyeria de Computació
 - Treball tutoritzat per: Coen Antens (Àrea de Ciències de la Computació i Intel·ligència Artificial)
 - Curs 2023/24

Un altre aspecte innovador és la **predicció de preus** en funció de les característiques de l'habitatge. Es tracta d'una funcionalitat avançada que utilitza les dades específiques de les llars juntament amb algorismes d'IA per fer prediccions acurades. Això proporciona als usuaris un avantatge competitiu significatiu a l'hora de prendre decisions d'inversió informades.

Finalment, el portal ofereix **recomanacions personalitzades** basades en les preferències i el comportament de cerca de l'usuari, augmentant les possibilitats de conversió. En proporcionar suggeriments altament rellevants i adaptats a les necessitats individuals, es millora l'experiència de l'usuari, s'incrementa la satisfacció i es fomenta una major interacció amb el portal, cosa que pot traduir-se en un augment significatiu en les taxes de conversió.

En resum, aquesta iniciativa està dissenyada per transformar la manera en què es gestionen i presenten les propietats immobiliàries, oferint major transparència i eficiència en el mercat immobiliari actual de Barcelona.

2 OBJECTIUS

El projecte té com a finalitat desenvolupar una plataforma immobiliària innovadora que maximitzi l'eficiència, millori l'experiència de l'usuari i ofereixi solucions avançades mitjançant la Intel·ligència Artificial.

Els objectius específics del projecte són:

- **Implementació Tecnològica:** Desenvolupar una plataforma web interactiva amb una interfície d'usuari intuïtiva que permeti als usuaris accedir fàcilment a totes les seccions i funcionalitats del portal immobiliari.
- **Generació Automàtica d'Anuncis:** Crear un sistema intern que permeti als propietaris del portal generar automàticament anuncis de noves propietats mitjançant IA. Els propietaris només hauran de pujar les imatges, i l'algoritme generarà automàticament títols, descripcions i preus suggerits, optimitzant els recursos i reduint la càrrega de treball.
- **Comparativa de Propietats:** Implementar una eina de comparativa de propietats mitjançant IA que permeti als usuaris comparar de manera eficient dues propietats seleccionades. Aquesta funcionalitat utilitzarà algorismes d'Intel·ligència Artificial per generar informació detallada i significativa, facilitant la presa de decisions informades i ràpides.
- **Predicció de Preus:** Desenvolupar una funcionalitat avançada que utilitzi dades detallades i algorismes d'IA per fer prediccions acurades de preus en funció de les característiques de l'habitatge, proporcionant un avantatge competitiu

significatiu a l'hora de prendre decisions d'inversió informades.

- **Recomanacions Personalitzades:** Utilitzar IA per oferir recomanacions personalitzades basades en les preferències i el comportament de cerca de l'usuari. Analitzar patrons de navegació i cerca per identificar propietats que s'ajustin als interessos específics de cada usuari, millorant l'experiència de l'usuari i augmentant les possibilitats de conversió.

A través d'aquests objectius, la plataforma no només aspira a oferir una eina innovadora per a la comparativa immobiliària, sinó també a destacar la contribució crucial de la Intel·ligència Artificial a la simplificació i millora d'aquest procés fonamental.

3 PLANIFICACIÓ

Per garantir l'èxit del projecte, s'estableix una planificació detallada que abasta les diverses activitats necessàries, especificant els temps i recursos associats a cadascuna. A continuació, es detallen les activitats clau i la seva progressió temporal, juntament amb la justificació dels recursos involucrats (vegeu resum a la Taula 1 en Apèndix):

1. **Desenvolupament de la Plataforma Web**
2. **Desenvolupament d'Algoritmes d'Intel·ligència Artificial**
3. **Integració de Components i Testos**
4. **Optimització i Ajustos**
5. **Seguiment i Control**

Es realitzarà una revisió setmanal dels programes de tasques i es faran ajustos segons els resultats i imprevistos. Es mantindrà una comunicació regular amb el tutor per avaluar el progrés i identificar possibles desafiaments.

S'establirà una plataforma de seguiment en línia per a l'entrada de dades i la revisió de la progressió dels objectius mitjançant l'entorn JIRA.

Aquest és el diagrama de Gantt que representa el conjunt de les activitats programades (vegeu la Figura 1 en Apèndix).

Amb aquesta planificació en detall i el seguiment regular, es busca assegurar que els recursos s'utilitzin de manera eficient, i que les activitats es compleixin dins dels terminis establerts, garantint així l'acompliment exitós dels objectius del projecte.

4 METODOLOGIA

Per optar a un enfocament més fluid i adaptatiu, s'ha decidit implementar la metodologia Kanban per al desenvolupament del projecte. La metodologia Kanban destaca per la seva flexibilitat i la seva capacitat per gestionar de manera eficient tasques variades. A continuació, es detalla la

implementació específica:

1. Tauler Kanban: Es crea un tauler Kanban virtual, dividit en columnes que representen les diferents etapes del procés, com ara "Backlog", "En Curs", "En Revisió" i "Complet" (vegeu la Figura 2)..

2. Cards del Tauler: Cada tasca es representa com un "card" al tauler Kanban, amb informació sobre la tasca, la seva prioritat i l'estat actual de desenvolupament.

3. Fluida Gestió del Backlog: El Backlog conté totes les tasques pendents, i el Product Owner prioritza les tasques basant-se en les necessitats immediates.

4. Reunions d'Optimització: S'organitzen reunions periòdiques per avaluar el rendiment del tauler Kanban, identificar possibles problemes i optimitzar el flux de treball.



Figura 2. Kanban Board exemple

Kanban proporciona una visió clara de totes les tasques en curs i permet una adaptació ràpida a canvis en les prioritats. El límit de tasques en procés ajuda a mantenir una càrrega de treball fàcil de gestionar, evitant sobrecàrregues i assegurant una progressió estable.

5 ESTAT DE L'ART

L'estat de l'art d'aquest projecte s'enfoca en cinc àrees principals:

➤ Generació d'anuncis de propietats immobiliàries

La Intel·ligència Artificial ha experimentat un creixement significatiu en els últims anys, especialment en l'àmbit del processament d'imatges.

Algoritmes i Tècniques Clau:

- **Models de Visió per Computador:** Algoritmes com YOLO (You Only Look Once) [2] i SSD (Single Shot Multibox Detector) [3] són capaços d'analitzar imatges de propietats, identificant elements clau com habitacions, mobles i altres característiques importants, que després s'utilitzen per generar descripcions detallades i precises.
- **Open CLIP:** CLIP (Contrastive Language-Image Pre-Training) és un model de IA desenvolupat per OpenAI

que pot associar imatges i descripcions textuais. Això és especialment útil per a crear títols i descripcions d'anuncis basats en les imatges de les propietats, assegurant que la informació sigui rellevant i atractiva per als possibles compradors o llogaters (vegeu la Figura 3) [4].

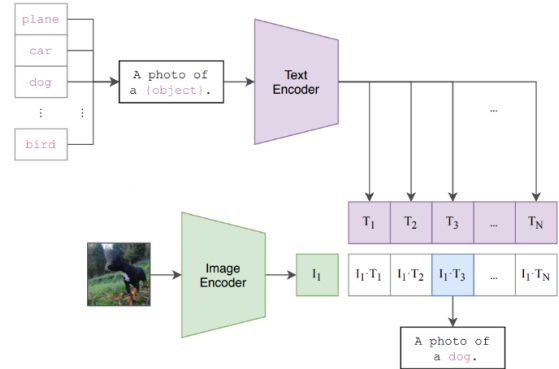


Figura 3. Esquema funcionament OpenCLIP

- **SAM (Segment Anything Network):** Aquesta xarxa de segmentació és capaç de dividir imatges en regions o segments significatius. Aplicada a les imatges de propietats immobiliàries, SAM pot identificar i separar diferents àrees com cuines, salons, dormitoris, etc. Això facilita l'extracció d'informació específica sobre cada part de la propietat [5].
- **DINO (DIstillation with NO labels):** DINO és un algorisme que funciona sense necessitat de dades etiquetades, permetent l'aprenentatge autònom de les característiques visuals importants. En el context de la generació d'anuncis, DINO pot ajudar a identificar les característiques úniques i valorades de cada propietat, contribuint a crear descripcions que ressalten els punts forts de les propietats sense requerir una gran quantitat de dades etiquetades manualment [6].

Un cas exemplar d'aplicació d'IA és Seeing AI de Microsoft [7], una aplicació que utilitza IA per a descriure l'entorn visual a persones amb discapacitat visual.

- Imatge (vegeu la Figura 4):

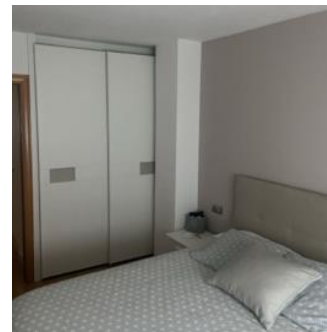


Figura 4. Fotografia d'un dormitori amb un llit

- Descripció generada:
Una cama con almohada y un armario.

L'aplicació d'aquestes tècniques en la generació d'anuncis permet als agents immobiliàries reduir significativament el temps dedicat a crear contingut, permetent-los centrar-se en altres aspectes del negoci.

➤ Comparativa de Propietats Immobiliàries

Les plataformes web per a la comparativa de propietats han crescut notablement, oferint eines sofisticades per a la recerca i comparació de llars. Aquestes plataformes faciliten la cerca de propietats i ofereixen informació detallada i eines de comparació. A continuació, es llisten les plataformes més destacades [8]:

- 1) Idealista: Ofereix serveis de cerca i comparació de propietats per criteris com preu, ubicació, i metres quadrats. Proporciona tendències de preus i estadístiques de demanda, ajudant a detectar propietats amb preus baixos (vegeu la Figura 5).

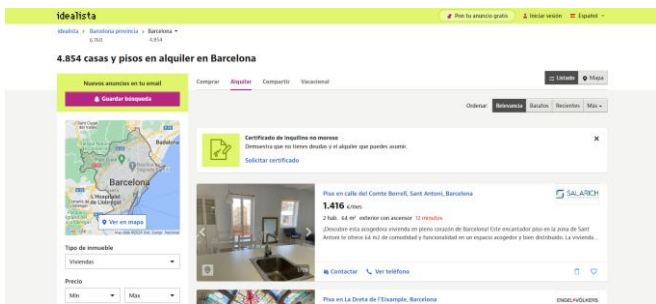


Figura 5. Llistat de propietats en portal Idealista

- 2) Fotocasa: Similar a Idealista, permet comparar opcions immobiliàries per preu, ubicació i característiques de la propietat. Ofereix informació actualitzada sobre els preus del mercat i alertes de preus.
- 3) Zillow: Als Estats Units, Zillow proporciona dades sobre el mercat immobiliari, incloent-hi valoracions estimades de les propietats (Zestimates), i eines de comparativa per a decisions informades.

Aquestes plataformes han revolucionat la manera com els compradors accedeixen a la informació i prenen decisions, millorant la precisió i l'eficiència de la recerca immobiliària.

➤ Predicció de Preus

Comprovar si el preu actual d'una propietat està sobrevalorat o infravalorat és crucial per a inversors, compradors i venedors. Per abordar aquest problema, s'utilitzen diverses tècniques de modelització i algoritmes de Machine Learning. A continuació, es presenten les tècniques més rellevants en l'actualitat:

- 1) Regressió: és una tècnica estadística que es basa en la relació entre una variable dependent i una o més variables independents [9].
- 2) Random Forest: és un algoritme d'ensemble que utilitza múltiples arbres de decisió per realitzar prediccions més robustes i precises. És especialment efectiu

gràcies a la seva capacitat de manipular grans conjunts de dades amb moltes variables [10].

- 3) Xarxes Neuronals: són models inspirats en el cervell humà, compostes per capes de neurones artificials. Són especialment potents per capturar relacions complexes i no lineals en les dades [11].
- 4) AutoML: simplifica el procés de desenvolupament de models de Machine Learning automatitzant la selecció, l'entrenament i la validació dels models. Això és especialment útil, ja que pot identificar automàticament els millors models i hiperparàmetres per a un conjunt de dades determinat [12].

Tècnica	Avantatges	Desavantatges
Regressió	Simple d'interpretar, fàcil d'implementar	Pot ser massa simplista, susceptible a l'overfitting
Random Forest	Alta precisió, robustesa, versatilitat	Pot ser lent en prediccions en temps real
Xarxes Neuronals	Captura relacions complexes i no lineals	Requereix grans conjunts de dades, temps de càlcul alt
AutoML	Automatitza el procés de selecció de models	Pot ser una "caixa negra", requeriments computacionals

Taula 2. Avantatges i desavantatges algorismes ML

Aquestes tècniques proporcionen diverses opcions per abordar el problema de la detecció de sobrevaloració o infravaloració dels preus en el mercat immobiliari. La selecció de la tècnica adequada depèn dels recursos disponibles, la complexitat de les dades i els requisits específics del projecte (vegeu comparativa en Taula 2).

➤ Eines de Gestió Immobiliària

Les eines de gestió immobiliària permeten gestionar eficientment les propietats, millorar la comunicació amb els clients i optimitzar els processos administratius. A continuació, es llisten les eines principals [13]:

- 1) MRI Software: Utilitzat per grans corporacions per a la gestió de propietats, amb funcionalitats com la gestió financera i el seguiment de contractes (vegeu la Figura 6).

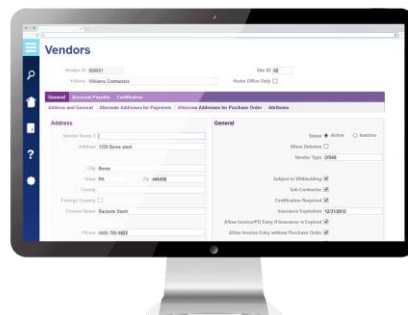


Figura 6. Pàgina principal MRI Software

- 2) Buildium: Adaptat per a mitjanes empreses, ofereix gestió de lloguers, manteniment, comptabilitat i comunicació amb els inquilins.
- 3) TenantCloud: Solució gratuïta i basada en el núvol per a petites empreses i empreses familiars, coneguda per la seva simplicitat i cost efectiu.

Aquestes eines permeten a les organitzacions optimitzar els seus processos, millorant l'eficiència operativa i proporcionant un millor servei als clients. Tot i l'existència de solucions parcials en cadascuna d'aquestes àrees, encara hi ha una mancança d'eines que integrin completament la descripció avançada d'imatges, la comparació de propietats i la gestió immobiliària en una única plataforma.

➤ Motors de Recomanació Personalitzada

Els motors de recomanació personalitzada són essencials per ajudar als usuaris a trobar propietats que s'ajustin millor a les seves necessitats i preferències específiques. A continuació, es llisten les tècniques clau més utilitzades [14]:

1) Filtratge basat en Contingut:

Recomana elements similars als que l'usuari ha mostrat interès anteriorment. Aquesta tècnica analitza les característiques dels elements i el perfil de l'usuari per generar recomanacions.

En primer lloc, es crea un perfil de l'usuari basat en les característiques dels elements que ha valorat positivament o amb els quals ha interactuat.

A continuació, les recomanacions es generen identificant elements que comparteixen característiques similars amb els elements preferits per l'usuari.

2) Filtratge Col·laboratiu

El filtratge col·laboratiu fa recomanacions basades en el comportament de la multitud. Hi ha dues subcategories principals: el filtratge col·laboratiu basat en usuaris i el filtratge col·laboratiu basat en elements (vegeu la Figura 7).

- Basat en Usuaris: Troba usuaris similars a l'usuari objectiu i recomana elements que aquests usuaris han valorat positivament.
- Basat en Elements: Troba elements similars als que l'usuari objectiu ha valorat positivament i recomana aquests elements similars.

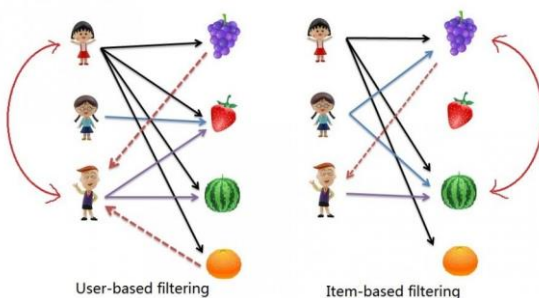


Figura 7. Funcionament basat en usuari vs elements

6 DESENVOLUPAMENT

Per garantir l'èxit del projecte, s'ha dut a terme un detallat procés de desenvolupament que engloba diversos aspectes:

6.1 Desenvolupament del Front-End amb AngularJS

En aquesta part s'ha desenvolupat una interfície d'usuari molt fàcil d'utilitzar que faciliti als usuaris el procés de buscar una llar per rentar.

S'ha utilitzat el framework Angular per desenvolupar la interfície d'usuari de la plataforma web, on podem observar diferents components i pàgines per gestionar les diverses funcionalitats de la plataforma (vegeu la Figura 8).

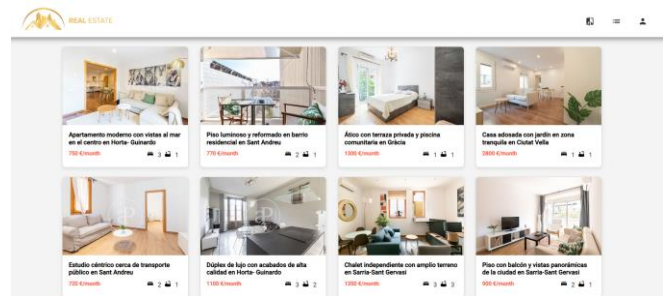


Figura 8. Llistat de propietats disponibles

En primer lloc, tenim el navbar o barra de navegació, on l'usuari podrà accedir a les diferents seccions de l'aplicació. Aquesta secció donarà accés principalment a tres components:

- Comparativa de propietats: conté dos seleccionables que permeten comparar dues propietats immobiliàries. Una vegada seleccionades les propietats, es generen de forma independent dues descripcions en un llistat de punts que, posteriorment, es compararan mitjançant l'ús de ChatGPT (vegeu la Figura 9).

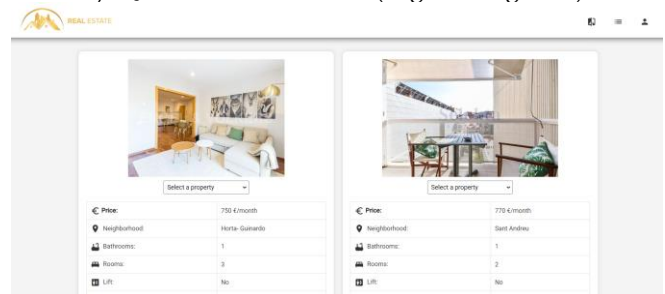


Figura 9. Vista pàgina comparativa de propietats

- Generació d'anuncis: a partir de la selecció de diverses imatges d'una propietat immobiliària, es genera una descripció exhaustiva de les característiques principals de la propietat. A continuació, obtindrem el títol de l'anunci, la descripció general de la propietat i un preu orientatiu per l'anunciant (vegeu la Figura 10).

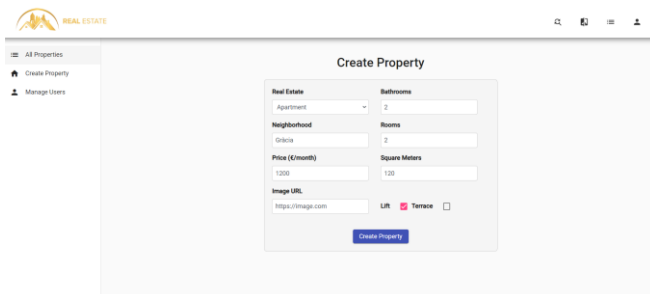


Figura 10. Vista generació d'anuncis automàticament

- Predicció de preus: mitjançant un gran dataset amb les característiques d'un conjunt de propietats immobiliàries, s'obtidran prediccions acurades de preus de noves propietats (vegeu la Figura 11).

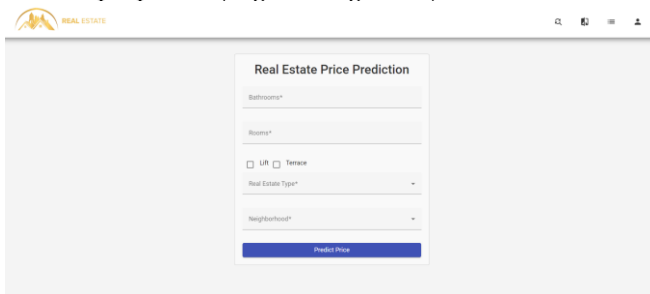


Figura 11. Vista predicció de preus

Cal mencionar que, durant les reunions de seguiment, es va plantejar l'idea de fer ús de la llibreria de python Gradio. El motiu pel qual es va descartar aquesta llibreria va ser per la gran versatilitat que té Angular a l'hora de personalitzar la part visible del lloc web.

6.2 Desenvolupament del Back-End amb NodeJS

Aquesta part se centra en el desenvolupament d'un Back-End robust, realitzat amb el framework NodeJS. El motiu principal pel qual s'ha decidit utilitzar aquesta tecnologia és perquè ofereix una estructura lleugera i flexible que s'adapta perfectament a les necessitats del projecte.

L'arquitectura està dissenyada per ser robusta i escalable. S'han implementat patrons de disseny com MVC (Model-View-Controller) per mantenir una estructura clara i modular. Això permet una millor gestió del codi i facilita la integració de nous components en el futur (vegeu la Figura 12).

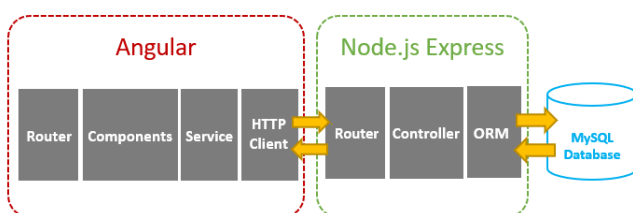


Figura 12. Esquema arquitectura projecte web

6.3 Obtenció i Tractament del Dataset

El dataset és una peça fonamental en el desenvolupament de la plataforma immobiliària, ja que proporciona les

dades necessàries per entrenar els models d'Intel·ligència Artificial utilitzats en les funcionalitats de comparació, generació d'anuncis i predicció de preus.

Concretament, va ser creat per Imanol Laconcha [16], un usuari de Kaggle, que va desenvolupar un algorisme de scraping per obtenir les dades del portal web Fotocasa. Aquest dataset conté les següents columnes:

- Preu: El preu de la propietat, expressat en unitats monetàries.
- Habitacions: El nombre d'habitacions de la propietat.
- Banys: El nombre de banys de la propietat.
- Ascensor: Un valor booleà que indica si la propietat disposa d'ascensor.
- Terrassa: Un valor booleà que indica si la propietat disposa de terrassa.
- Metres quadrats: La superfície de la propietat en metres quadrats.
- Tipus de propietat: El tipus de propietat (per exemple, pis, casa, etc.).
- Barri: El barri on es troba la propietat.
- Preu per metre quadrat: El preu per metre quadrat de la propietat.

A més, s'ha realitzat una anàlisi exhaustiu de les dades per comprendre millor les relacions entre les diferents característiques. Aquest anàlisi ha revelat els següents insights:

- Distribució dels Preus: S'ha observat que les propietats situades en barris centrals tendeixen a tenir preus més elevats en comparació amb les propietats situades en barris perifèrics (vegeu Taula 3).
- Impacte de les Característiques: Les propietats amb més habitacions, banys, i amb característiques addicionals com ascensors i terrasses, tenen preus significativament més alts.
- Variació del Preu per Metre Quadrat: El preu per metre quadrat varia considerablement segons la ubicació i el tipus de propietat, sent més alt en zones de major demanda.

	price	rooms	bathroom	lift	terrace	square_meters	square_meters_price
real_state							
flat	8565511	16532	9826	4822	1275	556272	102492.335884
attic	651024	760	547	254	223	32654	6269.341779
apartment	2067610	1998	1428	374	113	76140	28708.872892
study	80223	67	116	60	15	4480	2127.030131

Taula 3. Resum dades dataset

Es preveu una actualització contínua del dataset per incorporar noves dades del mercat immobiliari. Aquesta actualització permetrà mantenir els models al dia amb les tendències actuals, assegurant que les prediccions i descripcions generades siguin sempre rellevants i precises.

6.4 Generació de Descripcions

L'objectiu d'aquesta tasca és facilitar i optimitzar el procés

de creació d'anuncis als propietaris del portal immobiliari.

Mitjançant les descripcions generades pels algorismes de Machine Learning i les característiques de la propietat, la plataforma genera, de forma automàtica, un anunci complet per publicar la propietat al portal immobiliari (vegeu resultats la Taula 4).

	Step into this cozy kitchen with warm wooden cabinets and sleek white countertops, the perfect blend of rustic and modern vibes. (per fer anunci)
	A living room with a couch, television, and a fish tank. (només descripció)
	A bedroom with a bed, dresser, and mirror. (només descripció)

Taula 4. Descripcions obtingudes amb OpenCLIP

En primer lloc, s'ha utilitzat el model OpenCLIP per generar descripcions detallades a partir d'imatges de propietats immobiliàries. Aquesta arquitectura analitza i processa les imatges per identificar característiques com la forma, decoració i detalls estructurals de les propietats. A través d'aquest anàlisi d'imatges, OpenCLIP proporciona descripcions textuais que captiven les característiques visuals essencials de cada propietat [20].

Seguidament, mitjançant una connexió via API, les descripcions elaborades per OpenCLIP són refinades i millorades utilitzant ChatGPT. Aquest model de llenguatge natural utilitza el context i la coherència per oferir una narrativa fluida i informativa sobre les característiques i avantatges comparatius de les propietats (vegeu la Figura 14).



Figura 14. Esquema funcionament ChatGPT

6.5 Predicció de Preus en Funció de les Característiques

Per implementar la funcionalitat de predicció de preus a la web basada en les característiques proporcionades per l'usuari, s'ha utilitzat un model d'aprenentatge automàtic conegut com a Random Forest.

El procés comença amb el preprocessament de les dades per garantir que estiguin en el format adequat per al model. A continuació, s'entrena el model Random Forest amb aquestes dades preprocessades. Aquest entrenament permet al model aprendre els patrons del conjunt de dades per poder fer prediccions acurades de preus de propietats en base a les característiques introduïdes per l'usuari.

Quan un usuari introdueix les característiques d'una propietat a través de la interfície de la web, aquesta informació es transmet al Back-End, on el model Random Forest processa les dades i genera una predicció del preu estimat de la propietat.

6.6 Comparativa de Propietats

Per implementar la funcionalitat de comparativa de propietats mitjançant l'ús d'Intel·ligència Artificial, la plataforma utilitza un algorisme específic per estimar un preu de referència per a cadascuna de les propietats seleccionades. A continuació, es detalla com s'ha dut a terme aquest procés:

En primer lloc, quan l'usuari selecciona dues propietats a través de la interfície de la plataforma, les característiques numèriques d'aquestes propietats (com ara el nombre d'habitacions, banys, metres quadrats, etc.) s'envien al Back-End de la web.

Al Back-End, mitjançant l'ús de l'algorisme Random Forest, pren aquestes dades com a entrada i calcula un preu estimat per a cada propietat. Aquest preu estimat serveix com a punt de referència per determinar la millor relació qualitat-preu entre les dues propietats comparades.

Finalment, la plataforma retorna a l'usuari la comparativa entre les dues propietats, indicant quina d'elles ofereix una millor relació qualitat-preu segons el preu estimat calculat per l'algorisme.

6.7 Recomanacions Personalitzades segons Patrons de l'Usuari

La plataforma recull i analitza les interaccions dels usuaris amb la web, incloent-hi les propietats vistes, les comparatives realitzades, les preferències de cerca i altres comportaments rellevants mitjançant l'ús de Google Analytics (vegeu la Figura 15).

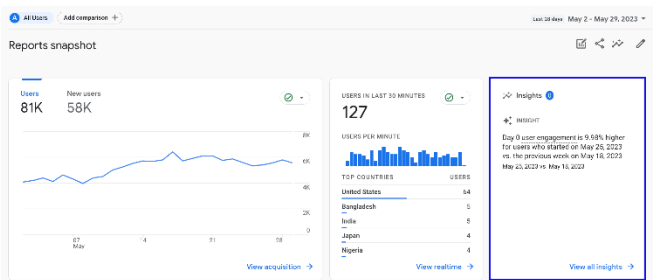


Figura 15. Monitoreig accions usuaris amb Google Analytics

Una de les biblioteques més utilitzades en Python és Surprise [15]. Aquesta biblioteca facilita la creació, entrenament i avaluació de sistemes de recomanació.

Per utilitzar Surprise, primer s'ha d'instal·lar la biblioteca. Un cop instal·lada, es poden carregar les dades d'interacció dels usuaris, com ara valoracions de propietats o altres preferències rellevants. Amb aquestes dades, Surprise permet entrenar diferents models de recomanació que analitzen les similituds entre usuaris o ítems per suggerir propietats que podrien ser d'interès per als usuaris.

La biblioteca inclou diversos algorismes, com ara el filtratge col·laboratiu basat en usuaris i en ítems, així com mètodes més avançats com la descomposició matricial (SVD). Aquests models s'entrenen i s'avaluen constantment per millorar la precisió de les recomanacions. Així, la plataforma pot adaptar-se als canvis en les preferències i comportaments dels usuaris, oferint suggeriments més rellevants i personalitzats.

7 EXPERIMENTS I RESULTATS

Els resultats obtinguts serveixen per validar les innovacions i demostrar el seu impacte en l'optimització de la gestió de propietats i millora de l'experiència d'usuari.

7.1 Generació d'anuncis de propietats

Per avaluar l'eficàcia de la generació d'anuncis de propietats utilitzant OpenCLIP i ChatGPT, es va realitzar una avaluació qualitativa de les descripcions generades per determinar la precisió, la claredat i la idoneïtat del contingut.

A més, es va recollir el feedback d'usuaris externs al projecte per validar la utilitat i l'adequació de les descripcions automàtiques per a les seves necessitats. Aquestes avaluacions multidimensionals han proporcionat una visió completa de l'eficàcia i el rendiment de les tecnologies implementades en la plataforma web immobiliària.

7.2 Comparativa de propietats immobiliàries

Els resultats de l'estudi sobre la comparativa de propietats revelen una valuosa perspectiva sobre les prediccions de preus i la relació qualitat-preu de les propietats analitzades.

Es va dur a terme una predicció del preu real de cada propietat en base a diverses característiques clau, com ara la ubicació, la superfície, el nombre de habitacions i altres factors rellevants.

Els resultats mostren que les prediccions de preus obtingudes amb aquestes tècniques van ser generalment precises, amb una variància mitjançada baixa respecte als preus reals del mercat. Això indica una bona capacitat del model per capturar les relacions entre les característiques de les propietats i els preus observats.

7.3 Predicció de preus

Abans d'entrenar els models, es van dur a terme les següents etapes de preprocessament de dades:

- Revisió i neteja de les dades, incloent el tractament de valors nuls.
- Conversió de variables categòriques en variables dummy per representar característiques com ara el barri i el tipus de propietat.

A continuació, les dades es van dividir en conjunts d'entrenament i prova en una proporció del 80/20, respectivament. Es va aplicar escalatge de característiques utilitzant StandardScaler per assegurar una convergència eficient dels models.

Els models es van avaluar utilitzant les següents mètriques de rendiment:

- Mean Absolute Error (MAE): La diferència mitjana entre les prediccions i els valors reals.
- Mean Squared Error (MSE): La mitjana de les diferències al quadrat entre les prediccions i els valors reals.
- R-squared (R²): La proporció de la variància en la variable dependent que és previsible a partir de les variables independents.
- Accuracy: percentatge de prediccions correctes sobre el total de prediccions realitzades.

A continuació es presenten els resultats obtinguts per cada model:

- **Regressió Lineal:**

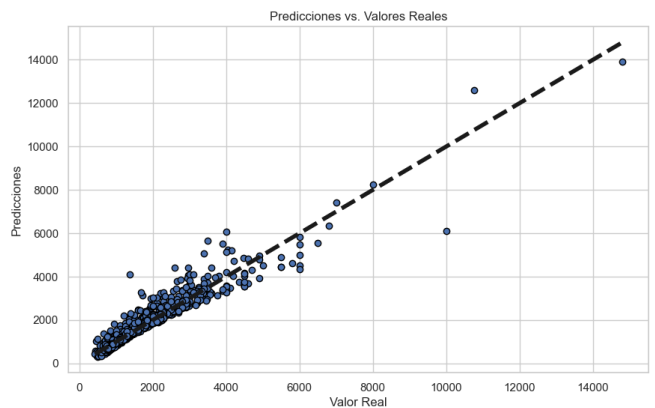


Figura 16. Predicció vs Valor Real (Linear Regression)

La Regressió Lineal mostra un bon ajust del model, amb un R² de 0.911, indicant que el model explica aproximadament el 91.1% de la variància en les dades. No obstant, els valors del MAE i MSE són elevats, suggerint que les prediccions poden estar força desviades dels valors reals (vegeu la Figura 16).

• Random Forest:

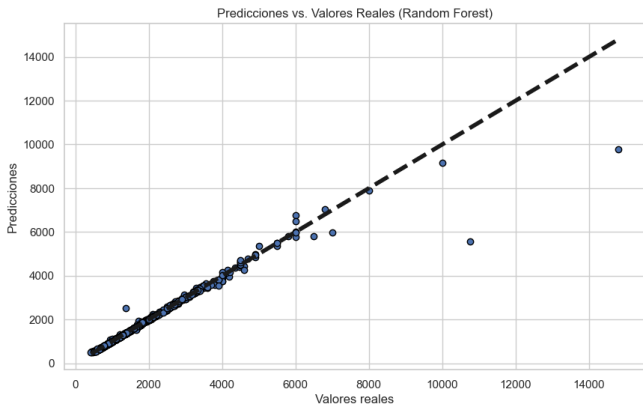


Figura 17. Predicció vs Valor Real (Random Forest)

El model Random Forest destaca per tenir els valors de MAE i MSE més baixos, el que indica que les prediccions són molt més properes als valors reals en comparació amb els altres models. Amb un R^2 de 0.96, aquest model explica el 96% de la variància en les dades, una mètrica molt alta que mostra la seva efectivitat. L'Accuracy és de 0.99, suggerint que el 99% de les prediccions van ser correctes, fet que reflecteix la gran capacitat d'aquest model per fer prediccions precises (vegeu la Figura 17).

• Xarxa Neuronal:

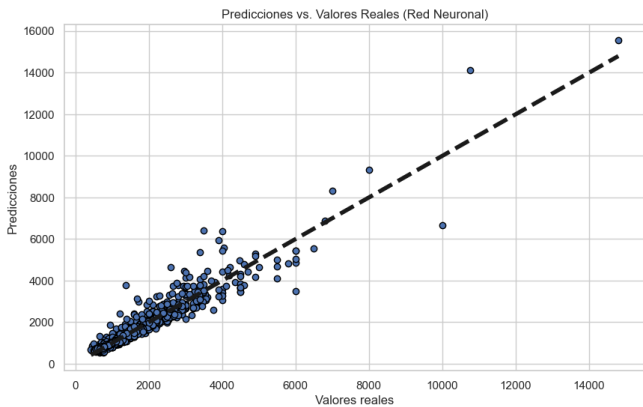


Figura 18. Predicció vs Valor Real (Xarxa Neuronal)

La Xarxa Neuronal ofereix un R^2 de 0.90, indicant que el model explica el 90% de la variància en les dades. Els valors de MAE i MSE són més alts que els del model Random Forest, però més baixos que els de la Regressió Lineal, suggerint un rendiment intermedi. L'Accuracy d'aquest model és del 72%, millor que la de la Regressió Lineal però inferior a la de Random Forest (vegeu la Figura 18).

• AutoML:

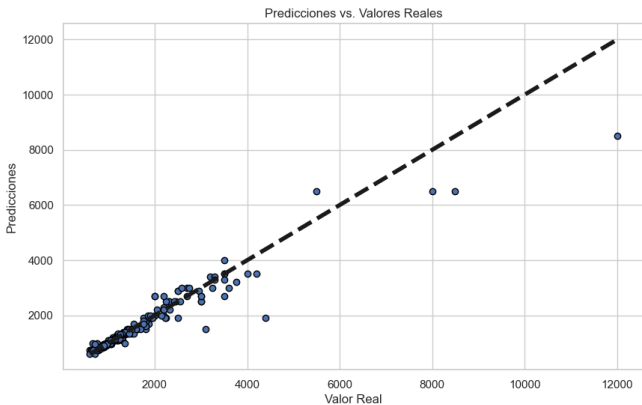


Figura 19. Predicció vs Valor Real (AutoML)

L'AutoML destaca per tenir valors competitius en totes les mètriques avaluades, oferint una solució robusta per a la predicció de preus de propietats immobiliàries. Amb un MAE de 170.685 i un MSE de 176996.275, el model indica que les prediccions són raonablement properes als valors reals. Aquestes mètriques suggereixen que les prediccions de l'AutoML són menys desviades respecte als valors reals en comparació amb altres models més senzills, com la Regressió Lineal (vegeu la Figura 19).

	MAE	MSE	R2	Accu-racy
Regressió li-neal	160.03	86061.52	0.911	-
Random Fo-rest	20.65	35499.37	0.96	0.99
Xarxa Neu-ronal	145.28	93313.99	0.90	0.72
AutoML	170.685	176996.275	0.89	0.77

Taula 5. Resultats Machine Learning

En comparació, el model Random Forest sobresurt en totes les mètriques, amb els valors més baixos de MAE i MSE, un R^2 alt de 0.96, i una Accuracy excel·lent del 99%. Això el converteix en el model més eficaç per a aquest conjunt de dades. La Regressió Lineal i la Xarxa Neuronal també ofereixen bons resultats en termes de R^2 , però les seves prediccions no són tan precises com les del Random Forest, especialment en termes de MAE i MSE (vegeu la Taula 5).

7.4 Motors de recomanació personalitzada

Els experiments amb motors de recomanació basats en contingut i col·laboratius han estat realitzats per avaluar la seva capacitat per millorar la personalització i la satisfacció de l'usuari. S'han recollit dades sobre les preferències i interaccions dels usuaris amb el sistema per adaptar les recomanacions. Els resultats han indicat una millora significativa en la rellevància i l'engagement dels usuaris amb les propostes de propietats recomanades. Aquests usuaris no han format part del desenvolupament per assegurar la integritat de les seves opinions.

8 CONCLUSIONS

Basant-nos en els objectius establerts al principi del projecte, destaquen els assoliments i la direcció futura de la plataforma immobiliària enriquida amb Intel·ligència Artificial:

Millora de l'Experiència de l'Usuari

Des de l'inici, l'objectiu va ser desenvolupar una interfície intuïtiva que facilités la cerca i la comparació de propietats. Amb la implementació d'Angular i la integració d'algorismes avançats d'IA, hem aconseguit proporcionar als usuaris una experiència fluida i altament personalitzada. La resposta inicial ha estat positiva, destacant la facilitat d'ús i la precisió de les recomanacions.

Optimització de Recursos mitjançant IA

L'objectiu d'automatitzar la generació d'anuncis immobiliaris s'ha complert amb èxit. La utilització de models com OpenCLIP i ChatGPT ha permès no només reduir el temps dedicat a la creació de contingut, sinó també millorar la qualitat i la rellevància de les descripcions i avaluacions de les propietats. Això ha conduït a una major eficiència operativa i a una millor satisfacció tant dels propietaris com dels usuaris finals.

Innovació en la Comparació de Propietats

El nostre objectiu era proporcionar eines avançades de comparació basades en IA que permetessin als usuaris prendre decisions informades. Les tècniques implementades han superat les expectatives, oferint anàlisis detallats i precisos que faciliten la identificació de propietats amb la millor relació qualitat-preu. Aquesta innovació ha diferenciat la nostra plataforma en un mercat competitiu.

Futur del Projecte

Mirant cap al futur, el projecte continuarà evolucionant per mantenir-se alineat amb les tendències del mercat i les necessitats canviant dels usuaris. Planegem incorporar noves tecnologies i funcionalitats que millorin encara més l'experiència de l'usuari i optimitzin la gestió immobiliària. La retroalimentació positiva rebuda fins ara reforça el nostre enfocament i ens motiva a seguir innovant.

En resum, el desenvolupament d'aquesta plataforma immobiliària impulsada per IA ha assolit els objectius establerts inicialment, establint una base sòlida per al seu creixement futur. Els resultats obtinguts subratllen el potencial de la plataforma per continuar liderant en innovació dins del sector immobiliari, proporcionant solucions avançades i eficaces que satisfan les necessitats de tots els usuaris implicats.

AGRAÏMENTS

Vull expressar el meu sincer agraïment a totes les persones que han contribuït al meu treball de final de grau. El seu suport continu, orientació i consells han estat de gran valor durant tot el procés de desenvolupament del projecte. Sense la seva ajuda, no hauria estat possible assolir els objectius establerts ni completar aquesta etapa acadèmica

amb èxit.

BIBLIOGRAFIA

- [1] Documental (Habitatge, dret o negoci): <https://www.ccma.cat/3cat/habitatge-dret-o-negoci/video/6287445/>
- [2] YOLO Object Detection: <https://towardsdatascience.com/yolo-object-detection-with-opencv-and-python-21e50ac599e9>
- [3] SSD Object Detection: <https://jonathan-hui.medium.com/ssd-object-detection-single-shot-multibox-detector-for-real-time-processing-9bd8deac0e06>
- [4] OpenCLIP documentation: https://python.langchain.com/v0.2/docs/integrations/text_embedding/open_clip/
- [5] Segment Anything Network: <https://github.com/facebookresearch/segment-anything>
- [6] DINO: Distillation with NO Labels: <https://medium.com/@ayyucedemirbas/distillation-with-no-labels-9e09b9b6dde2>
- [7] Seeing AI: Recognizing People, Objects and Scenes: <https://www.perkins.org/resource/seeing-ai-ios-app-recognizing-people-objects-and-scenes/>
- [8] Portals Immobiliaris: <https://inmogesco.com/blog/portales-inmobiliarios/>
- [9] Linear Regression Method: <https://www.freecodecamp.org/news/how-to-build-a-house-price-prediction-model/>
- [10] Random Forest ML: <https://www.kaggle.com/code/subhadeep88/house-price-predict-decision-tree-random-forest>
- [11] Neural Network for price prediction: <https://www.kaggle.com/code/tomasmantero/predicting-house-prices-keras-ann>
- [12] AutoML: <https://www.kaggle.com/code/aicentral/house-price-prediction-using-automl>
- [13] Eines gestió portal immobiliari: <https://www.doorloop.com/blog/real-estate-management-software>
- [14] Sistemes de recomanació (Suprise): https://medium.com/@jcjerez_77135/sistemas-de-recomendaci%C3%B3n-uso-de-la-librer%C3%ADa-surprise-de-python-bd3b6b72a3bb
- [15] Sistemes de recomanació use case: https://github.com/RicardoMoya/Recommender_Systems_Python
- [16] Dataset Barcelona Fotocasa Housing Prices Analysis: <https://www.kaggle.com/code/anacardells/barcelona-fotocasa-housing-prices-analysis>
- [17] Image Caption Generator using Deep Learning: <https://www.analyticsvidhya.com/blog/2021/12/step-by-step-guide-to-build-image-caption-generator-using-deep-learning/>
- [18] Create Captions with CoCa and OpenCLIP Notebook: https://colab.research.google.com/github/robgon-art/openclip/blob/main/Create_Captions_with_CoCa_and_OpenCLIP.ipynb#scrollTo=NSSrLY185j5f
- [19] Using OpenCLIP for Image Search and Automatic Captioning: <https://towardsdatascience.com/using-openclip-for-image-search-and-automatic-captioning-fa1cbbd48ce4>
- [20] ChatGPT API connection: <https://platform.openai.com/docs/api-reference/chat/create>
- [21] Price prediction with Python: <https://towardsdatascience.com/a-data-science-web-app-to-predict-real-estate-price-d2366df2a4fd>
- [22] Price prediction with Python use case: <https://thecleverprogrammer.com/2023/12/11/real-estate-price-prediction-using-python/>

APÈNDIX

A1. DIAGRAMA DE GANTT - PLANIFICACIÓ

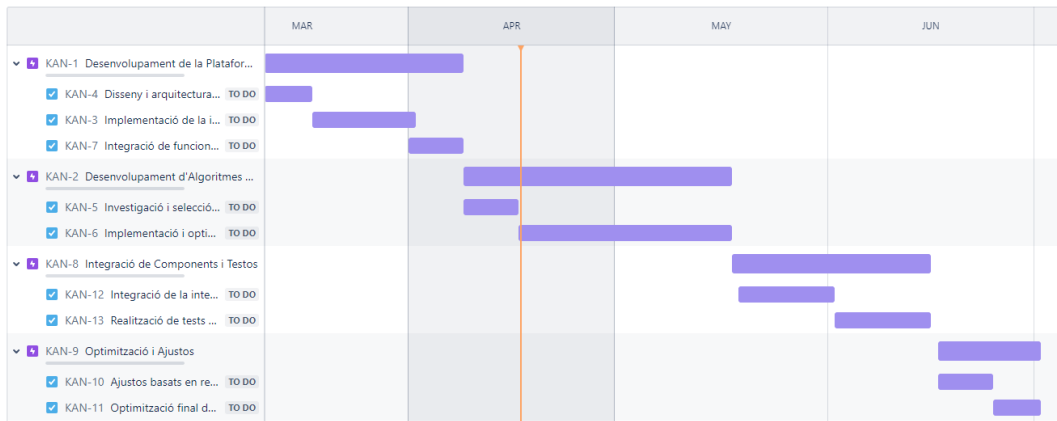


Figura 1. Diagrama de Gantt de la planificació general

A2. TASQUES - PLANIFICACIÓ

Act	Descripció	Temps	Recursos
1	Disseny i arquitectura de la plataforma web	2 setmanes	Desenvolupadors web, dissenyadors d'interfícies d'usuari.
2	Implementació de la interfície d'usuari interactiva (Angular)	3 setmanes	Desenvolupadors Front-End, dissenyadors d'interfícies d'usuari.
3	Integració de funcionalitats bàsiques de la plataforma (Python)	1 setmana	Desenvolupadors Back-End.
4	Investigació i selecció d'algoritmes d'IA per a l'anàlisi d'imatges	4 setmanes	Investigadors en intel·ligència Artificial
5	Implementació i optimització dels algoritmes d'IA (OpenCLIP & Image Captioning)	4 setmanes	Desenvolupadors especialitzats en Intel·ligència Artificial
6	Integració de la interfície amb els algoritmes d'IA	2 setmanes	Desenvolupadors full stack, testers
7	Realització de tests de funcionalitat i rendiment	1 setmanes	Equip de testeo, desenvolupadors
8	Ajustos basats en resultats de proves inicials	1 setmana	Desenvolupadors, testers
9	Optimització final del sistema	1 setmanes	Desenvolupadors, especialistes en rendiment

Taula 1. Taula amb resum de tasques