



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

González Casals, Pol; Montenegro Ruiz, Victoria, tut. Monitorización de Tendencias en Redes Sociales. 2025. (Enginyeria Informàtica)

This version is available at <https://ddd.uab.cat/record/317557>

under the terms of the  license

Monitorització de Tendències a les Xarxes Socials

Pol González Casals

29 de juny de 2025

Resum– El projecte consisteix en el desenvolupament d'una aplicació web que permet fer cerques sobre temes concrets a Reddit i analitzar les opinions de la comunitat. A partir d'una paraula clau (la cerca feta per l'usuari), el sistema extreu comentaris, filtra els que son realment rellevants mitjançant un model de classificació entrenat amb DistilBERT, que descarta automàticament els comentaris que no estan relacionats amb el tema de búsqueda o que no donen opinions. Un cop filtrats, es fa un anàlisi de sentiments (positius, negatius o neutres) utilitzant l'API de ChatGPT, que processa els textos i els classifica en funció del seu to i contingut. També es generen resums amb les opinions més destacades i es mostren gràfics i comentaris més votats. L'objectiu és oferir a l'usuari una visió clara i visual sobre què pensa la gent sobre un tema determinat, com ara un jugador de futbol, una destinació turística o un videojoc. L'aplicació busca ser útil per a persones interessades a conèixer tendències d'opinions de manera ràpida i intuïtiva.

Paraules clau– Anàlisi d'opinions, Reddit, classificació de sentiments, visualització de dades, aplicació web, PLN, DistilBERT, ChatGPT API

Abstract– The project consists of developing a web application that allows users to search for specific topics on Reddit and analyze the community's opinions. Based on a keyword (the user's search term), the system extracts comments and filters only those that are truly relevant using a classification model trained with DistilBERT. This model automatically discards comments unrelated to the search topic or lacking opinions. Once filtered, a sentiment analysis (positive, negative or neutral) is performed using the ChatGPT API, which processes the texts and classifies them based on tone and content. The system also generates summaries of the most notable opinions and displays charts and the most upvoted comments. The goal is to give users a clear and visual overview of what people think about a specific topic, such as football player, a travel destination or a video game. The application aims to be useful for those interested in quickly and intuitively understanding opinion trends.

Keywords– Opinion analysis, Reddit, sentiment classification, data visualization, web application, NLP, DistilBERT, ChatGPT API

1 INTRODUCCIÓ

AQUEST projecte consisteix en el desenvolupament d'una aplicació web que permet fer cerques sobre temes concrets a Reddit i analitzar les opinions dels usuaris. A partir d'una paraula clau introduïda

per l'usuari, el sistema extreu comentaris rellevants, els filtra mitjançant un model de classificació entrenat amb DistilBERT i descarta aquells que no estan relacionats amb el tema o que no aporten cap opinió significativa.

A continuació, es duu a terme un anàlisi de sentiments (positiu, negatiu o neutre) usant l'API de ChatGPT, que classifica els textos segons el seu to i context. També es generen resums amb les opinions més destacades, i es mostren gràfics i comentaris seleccionats amb més puntuació.

L'objectiu és oferir a l'usuari una visió clara i visual

• E-mail de contacte: polglezcasals@gmail.com
• Menció realitzada: Enginyeria del Software
• Treball tutoritzat per: Victoria Montenegro Ruiz (Departament de Ciències de la Computació)
• Curs 2024/25

sobre què pensa la comunitat de Reddit sobre un tema determinat, com pot ser un jugador de futbol, una destinació turística o un videojoc. L'aplicació pretén ser una eina útil per a persones que volen obtenir una síntesi d'opinions d'una manera ràpida, visual i fiable.

Finalment, aquest informe descriu el desenvolupament del projecte, les tecnologies usades, els resultats obtinguts i les possibles línies de millora.

2 FONAMENTS I ESTAT DE L'ART

Per desenvolupar aquest projecte ha estat necessari entendre diferents conceptes i tecnologies ja existents dins l'àmbit del tractament del llenguatge natural, la mineria d'opinions i l'anàlisi de tendències en xarxes socials. A més, s'ha tingut en compte l'existència de serveis ja creats que fan anàlisis semblants, amb l'objectiu de detectar què aporta de nou aquest sistema.

2.1 Anàlisi d'opinions i sentiments

L'anàlisi de sentiments és una tècnica de processament del llenguatge natural (PLN) que permet detectar l'emoció o la valoració subjacent d'un text [1][4]. Aquest tipus d'anàlisi s'ha aplicat amb èxit en àmbits com l'estudi de productes, la reputació de marques o l'anàlisi d'impacte social [3].

En aquest projecte, aquesta anàlisi s'ha aplicat sobre comentaris de Reddit per detectar si són positius, negatius o neutres, i per extreure opinions rellevants sobre temes com videojocs, futbol o viatges.

2.2 Serveis existents de social listening

Actualment existeixen diverses eines comercials dedicades a l'escolta activa a xarxes socials ("social listening"), l'anàlisi de sentiments i la monitorització de reputació en línia. A continuació es presenten algunes de les més destacades:

- **Brandwatch** [19]: plataforma reconeguda internacionalment, amb més de 7500 clients. Ofereix anàlisi multicanal de reputació, tendències i sentiment.
- **Sprout Social** [?]: molt utilitzat per empreses, permet analitzar més de 10 fonts (Twitter, YouTube, Instagram...) i fer benchmarking entre marques.
- **Mention** [21]: eina SaaS enfocada a pimes i agències, amb alertes i monitoratge en temps real de xarxes, blogs i fòrums.
- **Brand24** [22]: destaca per la seva facilitat d'ús i preu assequible. Permet seguiment de sentiment en xarxes i fòrums.
- **Awario** [23]: coneguda pel seguiment de Reddit i altres plataformes en temps real, útil per detectar tendències i crisis.
- **Mentionlytics** [24]: orientada a pimes, ofereix seguiment de competència i anàlisi de sentiment.

2.3 Comparativa i valor diferencial del projecte

Tot i que hi ha múltiples serveis de social listening consolidats com Brandwatch, Sprout Social o Brand24, el projecte desenvolupat té una sèrie de característiques diferencials que aporten valor específic per a l'usuari:

- Mentre que la majoria d'eines comercials estan orientades al seguiment de marques i estratègies empresarials globals, aquest projecte se centra en temes d'interès social.
- S'utilitzen dades obtingudes de Reddit, sense necessitat de subscripcions ni accés restringit, a diferència de moltes eines comercials que requereixen llicències costoses o estan limitades a certes plataformes.
- L'usuari pot fer una cerca i obtenir els resultats en una sola sessió, sense haver d'aprendre a utilitzar dashboards avançats o navegar per informes complexos.
- Gràcies a l'ús de la API de ChatGPT, el sistema genera un resum clar de les opinions més rellevants, una funcionalitat poc habitual en plataformes comercials.

Aquest punt de vista més concret i accessible permet arribar a un públic més ampli i oferir una experiència més especialitzada en àmbits on les eines existents poden resultar massa generalistes i amb un enfocament massa empresarial.

3 METODOLOGIA DE TREBALL

3.1 Planificació i fases del projecte

El desenvolupament del projecte ha seguit una metodologia iterativa i incremental, inspirada en pràctiques del desenvolupament àgil i dividida en diferents fases que han permès avançar de la següent manera:

- **Fase 1: Anàlisi i definició dels requisits**
S'ha fet una recerca prèvia sobre l'anàlisi d'opinions a les xarxes socials i els mètodes més utilitzats. A partir d'aquest punt, es van establir els objectius del projecte i es van definir els requisits funcionals (com ara fer cerques, mostrar comentaris, analitzar sentiments) i els no funcionals (com l'eficiència, la modularitat o l'accessibilitat).
- **Fase 2: Disseny de l'arquitectura**
S'ha decidit com s'organitzaria tot el sistema. S'ha optat per separar la lògica (backend) de la interfície web (frontend), per facilitar el manteniment i les possibles millores. També s'ha definit el flux de dades complet: des que l'usuari fa una cerca, fins que veu els resultats amb gràfics i comentaris. Aquest disseny ha servit de base per desenvolupar cada mòdul.
- **Fase 3: Implementació tècnica**
Es van programar primer els mòduls de backend, encarregats d'extreure dades, filtrar comentaris i classificar sentiments. Seqüencialment, s'ha construït el frontend amb una estructura clara i una navegació intuïtiva. Un cop cada part funcionava per separat, es van

connectar mitjançant peticions HTTP. Així s'ha aconseguit una comunicació fluida i eficient entre client i servidor.

- **Fase 4: Entrenament del model de classificació**

Es van recollir centenars de títols i comentaris de Reddit i es van etiquetar manualment per distingir els que eren útils dels que no aportaven informació rellevant. Amb aquestes dades s'ha entrenat un model basat en DistilBERT, capaç de filtrar automàticament els comentaris que no ajuden a entendre l'opinió general sobre un tema.

- **Fase 5: Validació i proves**

Un cop integrades totes les parts, es van fer diferents proves per assegurar que el sistema funcionava correctament. Es van fer cerques de prova, s'ha verificat la qualitat dels comentaris filtrats, la precisió en la classificació de sentiments i el bon funcionament dels gràfics i la navegació.

3.2 Tecnologies i eines usades

Durant el projecte s'han utilitzat diferents eines i tecnologies, escollides per la seva eficiència, documentació i compatibilitat amb projectes open source.

Aquest subapartat resumeix les tecnologies emprades en les diferents fases del projecte. L'ús concret de cada eina es detalla més endavant al Capítol 4.

- **Python** ha estat el llenguatge principal per programar tota la lògica del servidor (backend), ja que és molt flexible i té moltes llibreries útils per a tractament de dades i IA.
- **FastAPI** ha servit per crear els endpoints del backend d'una forma més ràpida i clara, ha fet accessibles funcionalitats com pasar la cerca al backend o l'enviament de resultats al frontend.
- **PRAW** [10] s'ha utilitzat per connectar amb Reddit, permetent obtenir fils i comentaris relacionats amb les cerques de l'usuari.
- **HuggingFace Transformers** i el model **DistilBERT** [6] han servit per entrenar el model que decideix si un comentari és útil o no. Aquest model s'ha entrenat amb dades reals del projecte.
- **ChatGPT API** [9] ha servit per analitzar el to dels comentaris (positiu, negatiu, neutre) i generar resums clars amb les opinions més representatives.
- **React** ha permès construir una interfície web amb components reutilitzables i interactius, i **Tailwind CSS** [11] ha facilitat aplicar un disseny net i flexible sense gaire complexitat.
- **React Router DOM** [14], els **hooks** [13] i els **contextos** s'han utilitzat per gestionar la navegació entre pantalles, l'estat de la cerca i la retenció de dades durant la sessió.
- **Git** i **GitHub** s'han utilitzat per al control de versions i gestió del projecte.

3.3 Estratègia de desenvolupament

El projecte s'ha pensat seguint un desenvolupament modular, dividint les funcionalitats en blocs independents: extracció de dades, filtratge, classificació, visualització i interacció amb l'usuari. Aquesta divisió ha permès treballar cada part per separat i provar i millorar cada mòdul de forma separada i controlada.

A més, s'ha donat molta importància a l'experiència d'usuari (UX). Durant el desenvolupament s'han incorporat funcionalitats útils com:

- Missatges i indicadors de càrrega mentre es processa una cerca.
- Mecanismes per evitar duplicació de dades o errors en la navegació.
- Persistència temporal de dades, per no perdre els resultats si es recarrega la pàgina o es torna enrere.

3.4 Criteris de validació

La validació del sistema s'ha realitzat en diversos nivells:

- **Validació funcional:** S'ha comprovat que cada acció (fer una cerca, carregar resultats, mostrar gràfics, etc.) funcionava tal com s'esperava. També s'ha provat la comunicació entre backend i frontend amb diferents cerques.
- **Validació de resultats:** S'ha revisat si els comentaris mostrats estaven ben filtrats i si la classificació de sentiments era coherent amb el contingut real dels textos. També es van analitzar els resums generats per l'API per assegurar que fossin clars i representatius.
- **Proves d'ús real:** S'han simulat casos d'ús amb temes habituals com futbolistes, destinacions turístiques o videojocs, per comprovar que el sistema era capaç de trobar opinions útils i mostrar-les d'una forma clara i visualment entenedora.

4 DESENVOLUPAMENT DEL SISTEMA

Aquesta secció descriu com està construït el sistema des del punt de vista tècnic i funcional. Es presenten els components principals, l'arquitectura general, el procés d'extracció de dades de Reddit, el filtratge i classificació d'opinions, així com la visualització dels resultats. També es reflexiona sobre el paper que tenen aquests mòduls en l'assoliment dels objectius del projecte i les possibilitats d'ampliació futura.

4.1 Arquitectura General

El sistema desenvolupat segueix una arquitectura basada en dues capes: frontend i backend. Aquesta separació permet millorar l'organització, la mantenibilitat i les futures possibles ampliacions.

El frontend s'ha desenvolupat amb React, utilitzant Tailwind CSS per ajustar disseny i l'estructura dels components i funcionalitats (cercador, gràfics, opinions, etc).

L'objectiu és oferir una interfície intuïtiva i visualment clara, que permeti a l'usuari interpretar de manera fàcil els resultats de la cerca.

El backend s'ha implementat amb FastAPI (Python), gestiona totes les operacions per processar les dades. Les seves funcions principals són:

1. Rep les paraules introduïdes per l'usuari
2. Fa crides a Reddit a través de PRAW per extreure comentaris relacionats
3. Filtra comentaris mitjançant un model entrenat amb DistilBERT
4. Classifica els sentiments de cada comentari utilitzant l'API de ChatGPT 3.5 Turbo
5. Retorna al frontend un objecte JSON amb la informació estructurada per ser visualitzada

Aquesta estructura permet que cada part del sistema es desenvolupi de manera independent, permetent optimitzar cada part de manera individual i fer modificacions sense la necessitat de readaptar la resta de l'estructura.

La figura 1 mostra aquesta arquitectura general, amb els actors implicats i el flux de dades des de la interacció inicial fins a la visualització final:

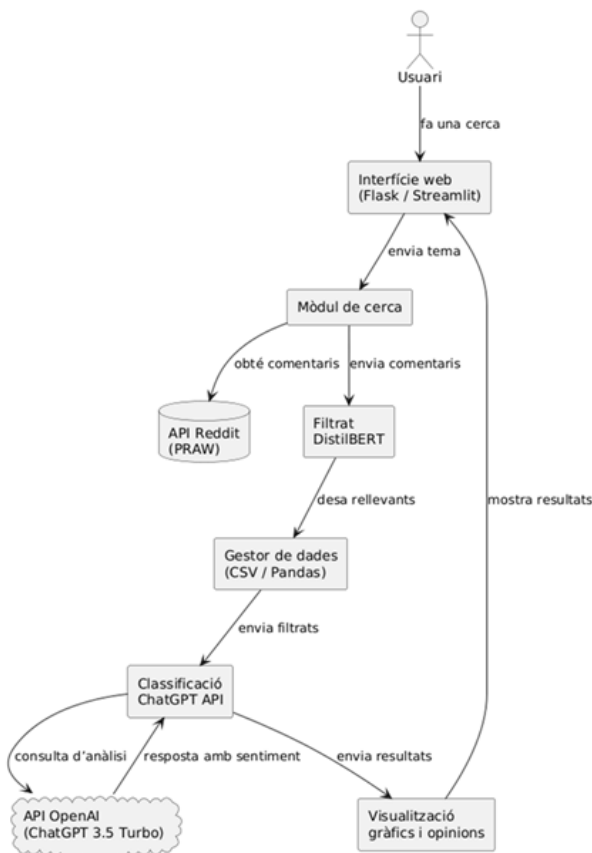


Fig. 1: Diagrama de l'arquitectura del sistema

4.2 Extracció de dades de Reddit

Per extreure comentaris rellevants sobre un tema específic, el sistema utilitza PRAW (Python Reddit API Wrapper),

una llibreria de Reddit que permet accedir de forma automatitzada al contingut de la xarxa social.

Quan l'usuari fa una cerca, s'executa una consulta global al subreddit `all` amb el paràmetre `sort=relevance`, per tal d'obtenir els fils més relacionats amb la paraula clau. De cada fil seleccionat, es recuperen els comentaris principals, limitant el nombre a 100 per consulta per garantir un temps de resposta raonable.

Els comentaris extrets es guarden inicialment en un fitxer .CSV, que inclou dades com el text del comentari, el nombre de vots positius (upvotes) i el nombre de respostes.

4.3 Filtrat i Classificació de les Dades

Un cop recollits els comentaris, el sistema aplica un procés de filtratge per seleccionar només aquells comentaris que són realment rellevants. Per aconseguir-ho, s'utilitza un model de classificació entrenat amb DistilBERT [5], una versió lleugera del model BERT [7] optimitzada per mantenir el rendiment amb menys recursos computacionals.

El model ha estat entrenat amb un conjunt de dades etiquetades manualment, en què es van distingir:

- **Útils:** aquells que expressen una opinió clara o aporten informació relacionada amb el tema.
- **No Útils:** comentaris genèrics, fora de context, amb enllaços o sense contingut útil (com contextos polítics no relacionats, aportacions vagues, ...).

El model aprèn a identificar quins textos aporten valor real a l'anàlisi d'opinió i descarta automàticament la resta.

Després del filtratge, s'aplica un sistema de classificació de sentiments mitjançant l'API de ChatGPT (model 3.5-turbo) [9]. Aquesta model permet identificar si el to de cada comentari és positiu, negatiu o neutre, tenint en compte el contingut complet del text. També s'utilitza per generar un resum dels comentaris i de les opinions més repetides o representatives, evitant duplicats i comentaris redundants.

4.4 Integració Frontend-Backend (FastAPI)

La integració entre la interfície web i el sistema s'ha realitzat seguint un model d'arquitectura client-servidor [18] basat en peticions HTTP [15], on el servei web (frontend) actua com a client interaccionant amb el servidor backend implementat amb FastAPI.

Quan un usuari fa una cerca, es realitzen dues crides principals:

• Inici del procés d'anàlisi

El frontend envia una petició GET a l'endpoint [17] `/api/search` del backend, passant la cerca feta com a paràmetre. Aquesta crida activa tot el procés intern:

- Extracció de dades de Reddit amb PRAW.

- Filtrat dels comentaris amb el model entrenat (DistilBERT).
- Anàlisi de sentiments i generació del fitxer `.json` amb resultats.

• Recuperació de resultats

Un cop el backend ha generat la resposta, el frontend realitza una nova petició GET a `/search/{keyword}` per obtenir el fitxer `.json`. Aquest fitxer conté els comentaris filtrats, els resums d'opinions i la classificació de sentiments.

Aquestes comunicacions s'implementen mitjançant funcions asincròniques [16] a React utilitzant `fetch(...)`. El flux complet d'interacció entre frontend i backend s'ha representat en un diagrama inclòs a l'Apèndix (Figura 3).

4.5 Servei web i Visualització

El servei web s'ha desenvolupat amb React [12], aprofitant la seva capacitat per crear interfícies interactives mitjançant components i per a l'estil del servei web s'ha utilitzat Tailwind CSS [11], que ha facilitat la creació d'un disseny modern.

El sistema organitza la navegació en diverses rutes amb React Router DOM [14], amb seccions diferenciades com:

- **Home:** pàgina principal amb el cercador i una breu descripció del projecte.
- **Search:** pàgina de resultats on es mostren els comentaris rellevants, les opinions destacades i els gràfics de sentiment.
- **About i Contact:** pàgines informatives sobre el projecte i l'autor.

La lògica de cerca i visualització s'ha implementat mitjançant hooks [13] i contextos personalitzats, que permeten gestionar l'estat global de l'aplicació i mantenir els resultats entre pantalles. Això garanteix una experiència d'usuari fluida, on les cerques es poden repetir o comparar fàcilment.

A la pàgina de resultats, el frontend carrega el fitxer `.json` generat pel backend i mostra la informació de manera visual i estructurada. Els resultats es divideixen en:

- Un gràfic (de barres o circular) representa la distribució de sentiments (positius, negatius, neutres).
- Opiniions destacades en format resumit.
- Comentaris individuals més votats.

L'usuari pot escollir el format de visualització que prefereixi i iniciar noves cerques des de la mateixa pàgina de resultats. Aquesta disposició permet una comprensió àgil, ràpida i efectiva de l'opinió general recollida sobre qualsevol tema.

Una vista detallada de la pàgina de resultats, amb els diferents blocs visuals (gràfic de sentiments, opinions i comentaris), es pot consultar a l'Apèndix (Figura 4).

5 RESULTATS I AVALUACIÓ

5.1 Funcionalitats assolides

El sistema desenvolupat compleix amb tots els requisits funcionals plantejats inicialment. Entre les funcionalitats més destacades s'inclouen:

- **Cerca per paraula clau:** l'usuari pot introduir qualsevol terme i el sistema inicia automàticament la cerca a Reddit, recollint fils i comentaris rellevants.
- **Filtrat automàtic de comentaris:** es descarten automàticament comentaris que no aporten informació útil mitjançant un model personalitzat basat en DistilBERT, entrenat amb dades reals i etiquetades manualment.
- **Classificació de sentiments:** cada comentari rellevant es classifica com a positiu, negatiu o neutre, amb a l'ús de l'API de ChatGPT, que permet identificar el to i la intenció del text.
- **Generació de resums:** les opinions més destacades es presenten de forma resumida per facilitar la comprensió general del tema, agrupant comentaris similars en missatges clars.
- **Visualització gràfica:** els resultats es mostren de manera visual amb gràfics de barres o circulars per representar la distribució dels sentiments, a més d'una llista d'opinions destacades i comentaris rellevants amb metadades.
- **Interfície web:** tota la informació es presenta a través d'una pàgina web interactiva, intuïtiva i responsiva, permetent la realització de múltiples cerques sense perdre la fluïdesa de navegació.

5.2 Casos d'ús demostrats

Per validar el funcionament global del sistema, s'han dut a terme diverses proves amb temes diferents. Aquestes cerques han permès comprovar que el sistema pot extreure informació rellevant, filtrar-la adequadament i representar-la de manera visual i entenedora. Alguns exemples destacats inclouen:

- **Futbol: “Cristiano Ronaldo fitxatge”**
Els gràfics de sentiments mostraven una polarització significativa, mentre que les opinions resumides reflectien temes recurrents com l'edat, el lideratge o la comparació amb altres jugadors.
- **Videojocs: “Logitech g502 hero”**
En aquest cas, els usuaris expressaven punts de vista positius sobre la precisió del ratolí i la seva relació qualitat-preu, però també apareixien crítiques sobre el pes i la forma.
- **Viatges: “Viatge a Malta”**
Es van obtenir comentaris sobre experiències reals, amb mencions a l'ambient, el cost, les activitats turístiques i el transport públic. Els resums destacaven aspectes com la seguretat, el bon clima i les recomanacions de llocs a visitar, etc.

Tots aquests casos d'ús han servit per validar no només les funcionalitats tècniques del sistema, sinó també la seva capacitat per oferir informació útil i comprensible a l'usuari final.

5.3 Impacte del model de filtratge

Una de les aportacions més rellevants del projecte ha estat la incorporació d'un model de classificació per filtrar els comentaris irrelevants amb DistilBERT. Aquest model permet filtrar automàticament els comentaris que no aporten una opinió clara o que no estan relacionats amb el tema cercat. Això evita mostrar a l'usuari informació poc útil o fora de context.

Durant les proves reals, s'ha observat que una gran part dels comentaris extrets inicialment no eren rellevants: contenien només enllaços, respostes buides, preguntes sense opinió o comentaris genèrics. El model ha estat capaç de descartar tots aquests casos, deixant únicament els comentaris que contenien valor informatiu.

El sistema és capaç de:

- Descartar automàticament comentaris que no són útils per analitzar l'opinió general.
- Agilitzar l'anàlisi de sentiments, ja que només es processen comentaris que tenen cert grau d'opinió o valoració.
- Millorar la interpretació final per part de l'usuari, mostrant només informació rellevant i fàcil de llegir.
- Reduir el soroll de les dades obtingudes, millorant tant la precisió com la qualitat dels resultats posteriors

Aquest comportament es pot observar clarament al gràfic següent 2, on es mostren els resultats obtinguts per a tres temes diferents. En tots els casos, es veu com el model amb DistilBERT descarta una part significativa dels comentaris extrets inicialment, conservant només els que contenen una opinió clara i rellevant.

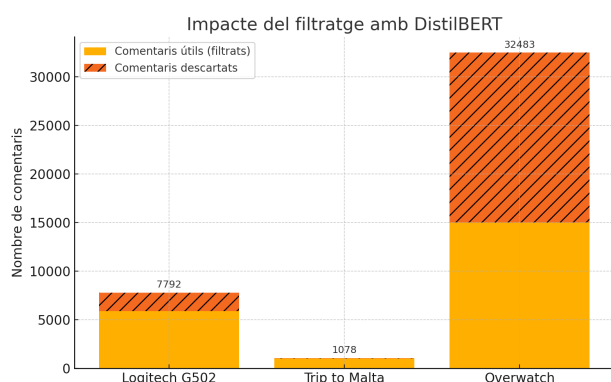


Fig. 2: Impacte del filtratge amb DistilBERT: comparativa entre comentaris útils i descartats

5.4 Limitacions detectades

Tot i que el sistema funciona correctament i compleix amb els objectius establerts, s'han detectat algunes limitacions

que poden afectar la precisió o eficiència del resultat final en determinades situacions:

- **Dependència de l'API de Reddit i ChatGPT:** Tant l'extracció de dades com l'anàlisi de sentiments depenen de serveis externs. Qualsevol canvi en les condicions d'ús, limitacions de crides o errors temporals poden afectar el funcionament del sistema.
- **Comentaris amb ambigüitat o ironia:** Tot i utilitzar un model avançat com ChatGPT 3.5 Turbo, alguns comentaris amb sarcasme, ironia o dobles sentits poden ser classificats incorrectament, especialment en contextos culturals concrets.
- **Dades desequilibrades segons el tema:** Temes molt concrets o amb poca activitat a Reddit poden generar pocs comentaris útils, dificultant l'anàlisi. A l'inrevés, temes molt amplis poden generar massa dades, algunes amb poca profunditat.
- **Rendiment i escalabilitat:** Tot i que el sistema funciona bé per a un ús individual, una càrrega molt elevada (molts usuaris fent cerques alhora) podria requerir optimitzacions al backend i una millor gestió dels recursos.

Aquestes limitacions no comprometen la funcionalitats bàsiques del sistema, però poden ser punts clau a tenir en compte de cara a futures millores i escenaris d'ús més avançats.

6 FUTURES LÍNIES DE MILLORA

De cara al futur, s'han trobat diverses millores que podrien ampliar les funcionalitats del sistema, augmentar la precisió i optimitzar l'ús:

- **Optimització del rendiment:** Millorar la velocitat general del sistema, especialment en cerques complexes o amb moltes dades, optimitzant els temps de resposta del backend i la càrrega al frontend.
- **Noves fonts de dades:** Obrir el sistema a noves fonts d'informació més enllà de Reddit, com Twitter/X, YouTube, fòrums o blogs per obtenir una quantitat més àmplia de opinions.
- **Control d'ús i escalabilitat:** Preparar el sistema per donar servei a múltiples usuaris simultàniament.
- **Historial i usuari:** Afegir un sistema d'autenticació opcional que permeti guardar l'historial de cerques, marcar comentaris com útils o generar informes personalitzats.
- **Interfície multilinguatge:** Implementar la possibilitat de canviar l'idioma de la interfície web, amb suport inicial per català, castellà i anglès.
- **Millora dels resums:** El model actual s'ha entrenat amb un nombre limitat de dades. Una ampliació del dataset (amb més comentaris etiquetats manualment) ajudaria a millorar la fiabilitat del filtratge.

- **Desplegament en producció:** Actualment el sistema només funciona en local. Una millora seria publicar-lo en un entorn de producció, com ho és Vercel, per fer-lo accessible a qualsevol usuari via web.

7 CONCLUSIONS

El projecte desenvolupat ha permès desenvolupar un servei web funcional capaç de buscar, extreure i analitzar opinions de Reddit sobre qualsevol tema. A través d'una arquitectura modular i l'ús d'eines com FastAPI, React i models de processament del llenguatge natural, s'ha aconseguit una aplicació clara, flexible i eficient.

S'han assolit els objectius plantejats inicialment: el sistema permet fer cerques personalitzades, filtrar comentaris rellevants, analitzar el sentiment dels textos i mostrar els resultats en una interfície visual intuïtiva. També s'han aplicat criteris de disseny web pensats per garantir una bona experiència d'usuari.

Un dels punts més destacats ha estat l'entrenament del model de classificació amb dades pròpies, que ha permès millorar el filtratge de comentaris no útils. A més, la integració amb l'API de ChatGPT ha afegit un nivell d'anàlisi extra que millora la informació final.

Tot i això, també s'han detectat algunes limitacions: la qualitat dels resultats pot variar segons el tema cercat i l'ús de serveis externs com l'API de ChatGPT comporta una dependència que pot traduir-se en costos addicionals o restriccions de disponibilitat. Malgrat això, el sistema funciona correctament i compleix amb les funcionalitats previstes.

En conclusió, aquest projecte demostra que és possible construir una eina automàtica per entendre què opina la comunitat sobre un tema determinat i obre la porta a futures millores que podrien ampliar el seu abast o millorar encara més els resultats obtinguts i mostrats.

AGRAÏMENTS

Vull expressar el meu agraïment sincer a la meua tutora, la professora Victoria Montenegro Ruíz, per la seva orientació constant, el seu suport tècnic i el seu criteri constructiu al llarg de tot el desenvolupament del projecte.

També vull donar les gràcies als meus pares, per haver-me donat sempre suport. I a la meua parella, per la seva paciència, ànims i companyia durant els moments més intensos d'aquest procés.

Aquest treball no hauria estat possible sense vosaltres.

REFERÈNCIES

- [1] IBM, "What is NLP (natural language processing)?," IBM Think. [Online]. Available: <https://www.ibm.com/think/topics/natural-language-processing> [Accessed: Mar. 10, 2025].
- [2] DeepLearning.AI, "Natural Language Processing — Resources." [Online]. Available: <https://www.deeplearning.ai/resources/natural-language-processing/> [Accessed: Mar. 10, 2025].
- [3] B. Liu, *Sentiment Analysis and Opinion Mining*, Morgan & Claypool Publishers, 2012. [Online]. Available: <https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html> [Accessed: Jun. 11, 2025].
- [4] Wikipedia contributors, "Sentiment analysis," *Wikipedia, The Free Encyclopedia*. [Online]. Available: https://en.wikipedia.org/wiki/Sentiment_analysis [Accessed: Jun. 11, 2025].
- [5] V. Sanh, L. Debut, J. Chaumond and T. Wolf, "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter," 2019. [Online]. Available: <https://arxiv.org/abs/1910.01108> [Accessed: Mar. 18, 2025].
- [6] Hugging Face, "distilBERT Model Card." [Online]. Available: <https://huggingface.co/distilbert-base-uncased> [Accessed: Mar. 18, 2025].
- [7] Wikipedia contributors, "BERT (model de llenguatge)," *Viquipèdia, l'enciclopèdia lliure*. [Online]. Available: [https://ca.wikipedia.org/wiki/BERT_\(model_de_llenguatge\)](https://ca.wikipedia.org/wiki/BERT_(model_de_llenguatge)) [Accessed: Mar. 18, 2025].
- [8] Hugging Face, "Transformers Documentation." [Online]. Available: <https://huggingface.co/docs/transformers/index> [Accessed: Mar. 18, 2025].
- [9] OpenAI, "OpenAI API Documentation." [Online]. Available: <https://platform.openai.com/docs> [Accessed: Apr. 2, 2025].
- [10] PRAW Contributors, "PRAW Documentation – Python Reddit API Wrapper." [Online]. Available: <https://praw.readthedocs.io/en/stable/> [Accessed: Mar. 5, 2025].
- [11] Tailwind Labs, "Tailwind CSS Documentation." [Online]. Available: <https://tailwindcss.com> [Accessed: Apr. 20, 2025].
- [12] Meta Platforms, Inc., "React – Una biblioteca per crear interfícies d'usuari." [Online]. Available: <https://es.react.dev/> [Accessed: Apr. 18, 2025].
- [13] Meta Platforms, Inc., "Introducció a los Hooks – Documentació oficial de React." [Online]. Available: <https://es.reactjs.org/docs/hooks-intro.html> [Accessed: Apr. 21, 2025].

- [14] Remix Software Inc., “React Router Documentation.” [Online]. Available: <https://reactrouter.com> [Accessed: Apr. 21, 2025].
- [15] MDN Contributors, “HTTP Overview - MDN Web Docs.” [Online]. Available: <https://developer.mozilla.org/en-US/docs/Web/HTTP> [Accessed: May 8, 2025].
- [16] Mozilla Developer Network, “async function - JavaScript,” *MDN Web Docs*. [Online]. Available: https://developer.mozilla.org/en-US/docs/Web/JavaScript/Reference/Statements/async_function [Accessed: May 8, 2025].
- [17] Cloudflare, “What is an endpoint?,” *Cloudflare Learning Center*. [Online]. Available: <https://www.cloudflare.com/learning/security/glossary/what-is-endpoint/> [Accessed: May 8, 2025].
- [18] Wikipedia contributors, “Client-server model,” *Wikipedia, The Free Encyclopedia*. [Online]. Available: https://en.wikipedia.org/wiki/Clientserver_model [Accessed: Jun. 11, 2025].
- [19] Brandwatch, “Digital Consumer Intelligence Platform.” [Online]. Available: <https://www.brandwatch.com/> [Accessed: Jun. 11, 2025].
- [20] Sprout Social, “Social Media Management Solutions.” [Online]. Available: <https://sproutsocial.com/> [Accessed: Jun. 11, 2025].
- [21] Mention, “Real-time Media Monitoring.” [Online]. Available: <https://mention.com/> [Accessed: Jun. 11, 2025].
- [22] Brand24, “Internet Monitoring Tool.” [Online]. Available: <https://brand24.com/> [Accessed: Jun. 11, 2025].
- [23] Awario, “Social Media and Web Monitoring.” [Online]. Available: <https://awario.com/> [Accessed: Jun. 11, 2025].
- [24] Mentionlytics, “Social Media Monitoring and Analytics.” [Online]. Available: <https://www.mentionlytics.com/> [Accessed: Jun. 11, 2025].

APÈNDIX A. DIAGRAMES I CAPTURES DE PANTALLA

A.1 Diagrames

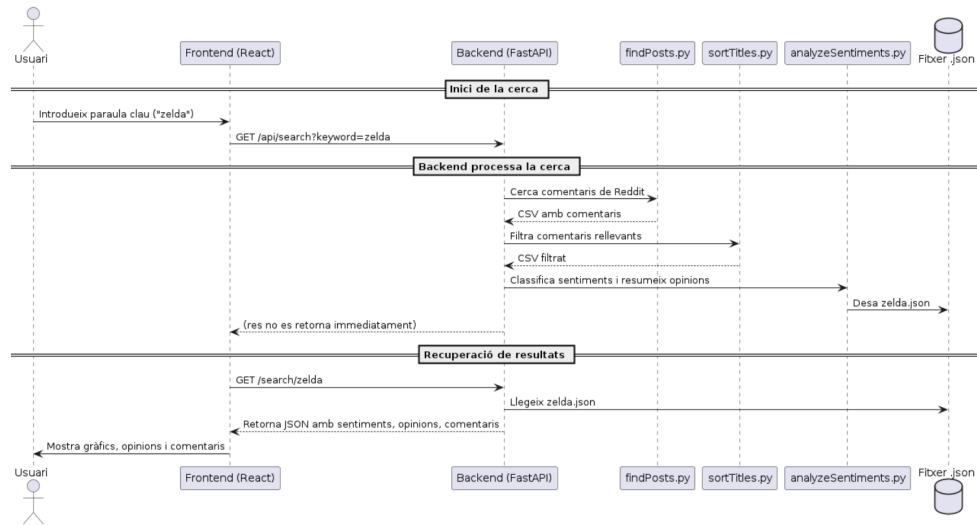


Fig. 3: Diagrama complet del flux del sistema: mostra el procés des de la cerca fins a la visualització.

A.2 Captures de pantalla

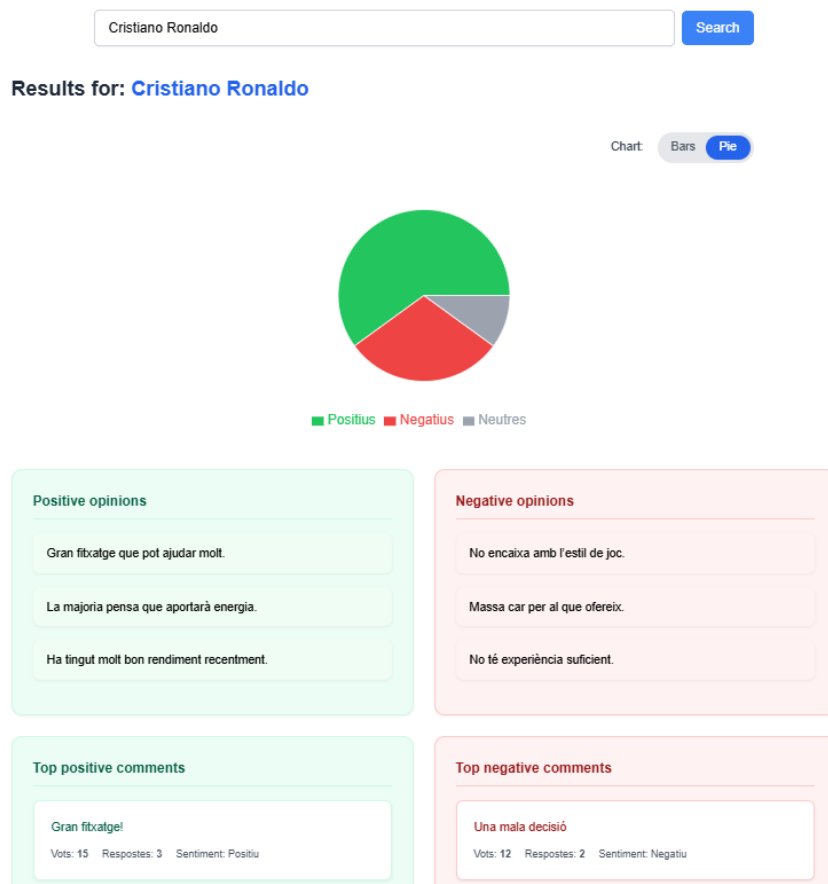


Fig. 4: Vista de la pàgina de resultats amb gràfic de sentiments, opinions i comentaris rellevants.

APÈNDIX B. EXEMPLE DE FITXER JSON GENERAT

A continuació es mostra un fragment representatiu del fitxer JSON generat pel sistema per una cerca sobre "Logitech G502". Aquest fitxer conté informació estructurada amb el nombre de comentaris per sentiment, els resums d'opinions i els comentaris destacats segons votació:

Listing 1: Exemple de fitxer JSON generat pel sistema

```
1 {
2   "sentiments": {
3     "positiu": 200,
4     "negatiu": 78,
5     "neutre": 172
6   },
7   "opinions": {
8     "positives": [
9       "I've had it for 2 years, the switches work perfectly, no double clicks...",
10      "For casual gaming and productivity yes",
11      "I just bought the white one. I love it. It would be perfect..."
12    ],
13    "negatives": [
14      "The dpi buttons are really badly placed and I changed the dpi multiple times accidentally.",
15      "Like all logitech mice with mechanical switches, this is prone to double-clicking.",
16      "Nice, but battery life is pretty awful if you have the RGB at all high."
17    ]
18  },
19  "comentaris": [
20    {
21      "text": "For Moba its actually the best option you can get...",
22      "sentiment": "positiu",
23      "vots": 27,
24      "respostes": 0
25    },
26    {
27      "text": "My 604 just died and I replaced it with a Razer... Going to return it for a 502x.",
28      "sentiment": "negatiu",
29      "vots": 10,
30      "respostes": 3
31    }
32  ]
33 }
```

Aquest format JSON és llegit pel frontend per generar els gràfics i llistes d'opinions mostrats a la interfície web.

APÈNDIX C. REQUISITS DEL SISTEMA I CASOS D'ÚS

Requisits Funcionals

ID	Nom del Requisit	Descripció
RF01	Cerca per paraules clau	L'usuari introdueix un terme i el sistema inicia la cerca de comentaris rellevants a <i>Reddit</i> .
RF02	Extracció de comentaris des de <i>Reddit</i>	El sistema fa servir l'API de <i>Reddit</i> per extreure posts i comentaris rellevants al tema cercat.
RF03	Filtrat automàtic de comentaris rellevants	El model ja entrenat filtra les opinions útils i descarta continguts no rellevants.
RF04	Classificació de sentiments	Les opinions seleccionades es classifiquen com a positives, negatives o neutres.
RF05	Visualització de resultats	Es generen gràfics i representacions visuals de sentiments i popularitat del tema.
RF06	Interfície d'usuari web clara i interactiva	El sistema ha de tenir una interfície gràfica senzilla per fer cerques i visualitzar resultats.

Requisits No Funcionals

ID	Nom del Requisit	Descripció
RNF01	Accessibilitat i usabilitat	L'aplicació ha de ser fàcil d'usar i accessible per a usuaris sense coneixements tècnics.
RNF02	Eficiència	L'extracció i classificació ha de ser àgil, i el filtratge ha de respondre en pocs segons.
RNF03	Escalabilitat	El sistema ha de poder ampliar-se a més temes o més dades sense redissenyar-lo completament.
RNF04	Compatibilitat tecnològica	Basat en tecnologies obertes: <i>Python</i> , <i>PRAW</i> , <i>Hugging Face</i> , <i>Plotly</i> , etc.
RNF05	Fiabilitat i qualitat de dades	Només s'analitzaran opinions útils, descartant publicacions de poc interès.
RNF06	Privacitat i seguretat	No s'ha de guardar informació personal ni exposar dades sensibles externament.
RNF07	Mantenibilitat i modularitat	El codi ha de ser estructurat i modificable per a evolucions futures del projecte.

Casos d'ús principals

ID	Cas d'ús	Actor	Descripció
CU01	Fer una cerca per tema	Usuari	L'usuari introdueix un tema i el sistema comença la recollida d'informació.
CU02	Extreure comentaris de <i>Reddit</i>	Sistema	El backend fa una crida a l'API de <i>Reddit</i> i obté posts i comentaris.
CU03	Filtrar comentaris rellevants	Sistema	El model selecciona només els comentaris que poden generar opinions globals útils.
CU04	Classificar sentiments	Sistema	El sistema aplica classificació positiva, negativa o neutra sobre els comentaris filtrats.
CU05	Visualitzar resultats	Usuari	L'usuari pot veure gràfics i resum de l'anàlisi.