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

Folch Aloy, Ona; Espinosa Morales, Antonio, dir. Anàlisi i revisió del flux de dades d'aplicacions mòbils. 2023. (Enginyeria de Dades)

This version is available at <https://ddd.uab.cat/record/281549>

under the terms of the  license

Anàlisi i revisió del flux de dades d'aplicacions mòbils

Ona Folch Aloy

4 de juliol del 2023

Resum– Les aplicacions mòbils són de summa importància en l'actualitat, ja que cada vegada més persones opten per realitzar les seves activitats de manera digital. Aquestes juguen un paper fonamental en la comunicació amb els clients, sent crucial el seu correcte funcionament, ja que les valoracions i opinions dels usuaris tenen un fort impacte en les decisions dels consumidors a l'hora de triar una aplicació, influenciant les seves decisions de descàrrega i ús. És vital, per tant, analitzar els comentaris i millorar el funcionament per atraure i fidelitzar els usuaris, generant confiança i augmentant la visibilitat de l'aplicació. Mitjançant l'ús de Power BI es visualitzen les dades de valoracions i puntuacions dels usuaris, permetent identificar problemes i oportunitats, i prendre decisions informades per millorar l'experiència dels usuaris i assegurar l'èxit del negoci.

Paraules clau– Visualització de dades, Power BI, aplicació mòbil, ressenyes, puntuacions

Abstract– Mobile applications are extremely important nowadays, as an increasing number of people choose to carry out their activities digitally. They play a fundamental role in communicating with customers, with their proper functioning being crucial. User ratings and reviews have a strong impact on consumers' decisions when choosing an application, influencing their download and usage decisions. It is therefore vital to analyze user feedback and improve performance to attract and retain users, building trust and increasing the visibility of the application. Reviews and ratings are visualized using Power BI to identify issues and opportunities and make informed decisions to improve the user experience and ensure business success.

Keywords– Data visualization, Power BI, mobile application, reviews, ratings



1 INTRODUCCIÓ

LA visualització de dades és una eina fonamental en l'anàlisi de dades, ja que permet als usuaris veure i entendre la informació amb major claredat. En la nostra era digital, les empreses, organitzacions i individus generen quantitats massives de dades, però sense una manera de comprendre i comunicar-les, la informació es perd o es malinterpreta.

En aquest projecte s'estudien les dades d'aplicacions mòbils bancàries, tant d'Android com iOS, analitzant els *ratings* i valoracions dels usuaris classificades en diferents categories. Així mateix, també s'analitza el volum de co-

mentaris i les operacions realitzades en la plataforma.

Les valoracions i les opinions dels usuaris són factors clau en la presa de decisions dels consumidors a l'hora de triar una aplicació mòbil. Les persones confien en les experiències compartides i en les opinions expressades pels altres usuaris per avaluar la qualitat i la fiabilitat d'una aplicació, ja que proporcionen una perspectiva realista i imparcial de l'experiència d'ús. Busquen informació sobre com es compara una aplicació amb altres competidores en termes de funcionalitat, rendiment i facilitat d'ús, i aquestes comparatives els permeten prendre una decisió més informada i triar l'aplicació que millor s'adapti a les seves necessitats.

L'estudi i millora d'aquests rànquings és molt rellevant per a l'adopció de l'aplicació per part dels usuaris, ja que una alta puntuació genera confiança i augmenta la visibilitat, atraient nous usuaris i fidelitzant els ja existents. A més, la bona funcionalitat és fonamental per garantir que els usuaris estiguin satisfets. Han de ser intuïtives i fàcils d'utilitzar, i funcionar sense problemes en tots els disposi-

• E-mail de contacte: ona.folch@autonoma.cat
• Treball tutoritzat per: Antonio Espinosa Morales (Departament d'Arquitectura de Computadors i Sistemes Operatius)
• Curs 2022/23

tius mòbils, ja que si no funciona correctament, els usuaris poden abandonar-la ràpidament i cercar una alternativa.

Com a eina principal per a l'anàlisi i visualització de dades es fa servir *Power BI*, una plataforma de *business intelligence* que permet a les empreses i organitzacions recopilar, analitzar i visualitzar dades en temps real per a prendre decisions informades i estratègiques [1]. És una eina potent que habilita als usuaris connectar-se a múltiples fonts de dades, com ara bases de dades, fulls de càlcul, aplicacions en línia i fitxers. Això permet analitzar dades de diferents fonts i integrar-les en un sol lloc.

2 OBJECTIUS

L'objectiu és obtenir una comprensió més profunda dels patrons de comportament dels usuaris i les seves preferències d'ús per a les aplicacions mòbils, utilitzant tècniques i eines avançades d'anàlisi de dades. A través de la modelització d'aquestes preferències i la gestió de les aplicacions tant per a Android com per a iOS, es pot millorar significativament l'experiència dels usuaris i augmentar la seva satisfacció.

Aconseguir una visió més completa del comportament dels usuaris per entendre com aquests responen davant les funcionalitats i característiques de l'aplicació, a través de la creació de visualitzacions específiques que permetran respondre a una sèrie de preguntes.

Analitzar la percepció dels usuaris de les aplicacions davant d'altres del mateix sector, mitjançant les valoracions i comentaris disponibles a la *Play Store* i l'*App Store*. Es busca identificar les seves àrees de fortaleza i debilitat, i proposar recomanacions concretes per millorar l'experiència dels usuaris i guanyar terreny davant la competència.

Utilitzar tècniques de processament del llenguatge natural per classificar els temes recurrents i proporcionar una visió general de la satisfacció dels usuaris i la reputació de l'aplicació, analitzant les dades i interpretant-les per a la presa de decisions.

2.1 Objectius Desenvolupament Sostenible

Aquest projecte incorpora dos dels objectius de les Nacions Unides per al Desenvolupament Sostenible.

- **Objectiu 8.** Treball digne i creixement econòmic
- **Objectiu 9.** Indústria, innovació i infraestructures

3 ESTAT DE L'ART

La bona funcionalitat i utilitat de les aplicacions mòbils són fonamentals per garantir la satisfacció de l'usuari i fomentar l'ús continu de l'aplicació, i és per aquesta raó que actualment s'utilitzen programes per visualitzar les dades generades per aquestes aplicacions i poder extreure'n un valor beneficiós.

3.1 Estudis previs

En aquest apartat es revisen els estudis previs en visualització de dades per presentar una visió general de les tècniques més comunes, les eines més populars i els mètodes emprats per avaluar l'efectivitat de les visualitzacions.

Henry Okam, analista de dades i enginyer de *software*, ha fet una anàlisi de les aplicacions de Google Play Store amb l'objectiu de descobrir les tendències i les categories més rellevants [2]. Per fer-ho, ha fet servir diferents eines com *SQL* per fer consultes sobre les dades, les llibreries *pandas* i *numpy* de *Python* per processar-les i *Power BI* per visualitzar-les de manera clara i efectiva. Entre les tècniques emprades destaquen els gràfics de barres per mostrar dades quantitatives, un *wordcloud* per analitzar el sentiment dels usuaris i un gràfic de línies per mostrar l'evolució mensual de les dades.

En canvi, Shehzada Alam, científic de dades, fa servir Tableau i Qlik per analitzar les vendes d'una empresa de subministraments generals [3]. Utilitza *tree maps* i gràfs per representar visualment les dades de vendes i així poder obtenir una comprensió més clara i ràpida de les tendències.

Vamsi Vempati, enginyer de *software*, compara les ressenyes de l'aplicació *Google Maps* i *Waze*, utilitzant eines com *Python* per la preparació de les dades, *Excel* per fer algunes visualitzacions i *Thematic* per crear gràfiques i visualitzacions interactives. A través de l'anàlisi de patrons d'ús, la satisfacció dels usuaris, els problemes comuns i les característiques més populars d'ambdues aplicacions, s'identifiquen les diferències i les similituds entre elles.

En general, tots els projectes esmentats anteriorment tenen un objectiu comú: processar les dades, analitzar-les i extreure conclusions a través de representacions gràfiques per a identificar tendències i comparar resultats.

3.2 Eines de visualització de dades

Gartner, una empresa de consultoria que ajuda a avaluar i comparar proveïdors en un determinat mercat, va publicar el març del 2022 un quadrant amb les empreses més rellevants en aquest sector [5].



Fig. 1: Quadrant per a plataformes d'anàlisi i intel·ligència empresarial

Representa una matriu de dues dimensions que avalua les empreses en funció de la capacitat d'execució (habilitat de l'empresa per a dur a terme productes i serveis d'alta qualitat) i la visió estratègica (capacitat per anticipar les tendències del mercat). En el quadrant es pot veure que *Power BI* està ressaltada en color vermell. Totes les empre-

ses que apareixen són importants, ja que la selecció depèn dels requisits específics que es necessiten. No obstant això, les que es troben a la secció de líders, com seria *Power BI*, *Tableau* i *Qlik*, són les que tenen una posició dominant al mercat i una sòlida base de clients.

4 METODOLOGIA I PLANIFICACIÓ

En aquesta secció es descriu la metodologia utilitzada per a la visualització de les dades, juntament amb altres eines utilitzades per a la gestió del projecte.

Per a la visualització de les dades, es fa ús de la plataforma *Power BI*, i s'empren eines de visualització per la creació de gràfics i vistes. Les dades es processen prèviament en *Python* per a la seva neteja i estructuració.

Pel que fa al seguiment del projecte, s'utilitza Microsoft Teams, juntament amb reunions quinzenals per a revisar el progrés del projecte i establir les tasques següents.

Per fer aquest treball i assolir els objectius esmentats, se segueixen una sèrie de passos definits a continuació:

1. **Identificació del problema i definició:** s'exposa la necessitat d'analitzar les dades d'aplicacions mòbils. Definir el problema específic que es vol abordar i com l'empresa pot beneficiar-se d'aquesta anàlisi. A més a més, també és important destacar els avantatges que pot aportar la visualització de dades en aquest procés.
2. **Extracció i neteja de les dades:** Es dissenyen fluxos *ETL* (*Extract, Transform, Load*) per l'extracció de les dades i la preparació de l'anàlisi. Es tenen en compte els diferents formats de les dades, i com estan emmagatzemades pel seu posterior processament, i es realitza una neteja i transformació per poder ser utilitzades en les visualitzacions, i estructurar-les de manera que siguin fàcils de consultar i analitzar.
3. **Anàlisi de les dades:** s'elaboren anàlisis de les dades processades que siguin rellevants per l'estudi de visualització.
4. **Visualitzacions:** es duen a terme visualitzacions de les dades rellevants, amb la finalitat d'identificar patrons i tendències en la informació. Per aquestes visualitzacions, s'escullen les millors eines de visualització per presentar les dades de manera efectiva.
5. **Conclusions:** s'avaluen els resultats obtinguts en relació amb els objectius del treball i s'extreuen conclusions.

És un treball que requereix una planificació rigorosa i un enfocament metodològic ben definit per tal d'assegurar que se segueixen aquests passos i durant el termini establert. Per aquesta raó, s'ha planificat el projecte amb *Scrum*, un enfocament de gestió de projectes àgil que implica la divisió de tasques en *sprints*, períodes de temps definits que solen durar entre 2 i 3 setmanes.

En la taula 1 podem observar les diferents tasques que s'han de fer per dur a terme el projecte, juntament amb l'*sprint* corresponent, tot això tenint en compte les possibles contingències que puguin sorgir durant el procés.

Aquests *sprints* es poden definir en un diagrama de *Gantt* (Figura 2), una eina de gestió de projectes per planificar i

TAULA 1: TASQUES PER CADA SPRINT

Sprint 1	- Identificació i definició del problema - Revisió exhaustiva de la literatura existent - Plantejament dels objectius i metodologia
Sprint 2	- Comprensió de les dades proporcionades - Tècniques per mostrar les dades - Esbossos de visualitzacions
Sprint 3	- Transformació i neteja de les dades - Selecció de les dades rellevants - Anàlisi de les dades proporcionades
Sprint 4	- Experimentació de la plataforma <i>Power BI</i> - Visualitzacions de les dades rellevants
Sprint 5	- Interpretació de les gràfiques - Donar resposta a les preguntes formulades
Sprint 6	- Conclusions i resposta dels objectius inicials - Revisió de l'informe final
Sprint 7	- Preparació de la presentació final

controlar el temps que es dedica a cada *sprint* i les seves corresponents tasques, mostrant la durada prevista de cada una d'elles i la seva posició en el temps.

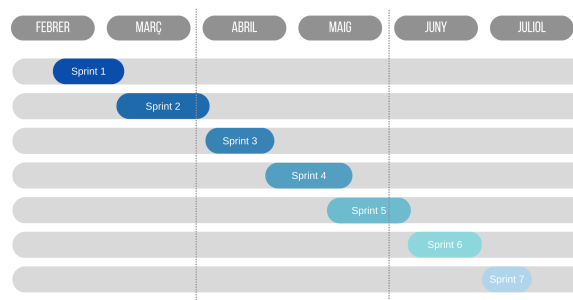


Fig. 2: Diagrama de *Gantt* de la planificació del projecte

Es documenta tot el procés seguit, des de l'extracció de les dades, la creació de les visualitzacions i l'anàlisi dels resultats, i es realitzen revisions periòdiques del progrés.

5 TIPOLOGIA DE LES DADES

En aquest projecte s'estudien les dades recopilades de les aplicacions mòbils d'Android i iOS durant 24 mesos. Les dades consisteixen en les puntuacions que els usuaris han atorgat a cada aplicació, que varien d'una escala de l'1 a 5.

A més dels *ratings*, també disposem d'una gran quantitat de comentaris dels usuaris, amb els que podem comprendre els motius que els han portat a valorar l'aplicació d'una manera determinada. A la taula 2 es pot veure el nombre de comentaris i la longitud d'aquests per cada aplicació.

La longitud mitjana dels comentaris amb una puntuació d'1 és de 30 paraules, mentre que els comentaris amb la valoració més alta és de 8-10 paraules. Per tant, els comentaris negatius són més extensos per ambues aplicacions, mentre que els comentaris amb una puntuació més elevada no proporcionen tanta informació. A més a més, hi ha una quantitat més alta de comentaris per la primera aplicació, de tal manera que s'hauran de normalitzar les dades.

Utilitzant l'aplicació *Appfollow*, una plataforma

TAULA 2: NOMBRE DE COMENTARIS I LONGITUD

	Rating	1	2	3	4	5	Total
App 1	# reviews	8049	1356	1187	1989	20752	33333
	# words	33.55	34.68	28.57	19.91	10.10	18.01
App 2	# reviews	3068	342	538	2014	17198	23160
	# words	29.82	31.84	19.39	13.54	7.44	11.57

d'anàlisi i gestió d'aplicacions mòbils, se selecciona l'interval de dates i es descarreguen les dades en forma d'arxius *Excel*. Per a cada aplicació, tenim un fitxer per a les puntuacions i un altre per a les opinions, tant per a *iOS* com per a *Android*.

Dins de l'arxiu d'opinions, es troba cada comentari realitzat per un usuari, juntament amb la informació de l'usuari i la puntuació associada. També s'inclou la data en què es va publicar el comentari, la versió de l'aplicació que es va fer servir en el moment de fer el comentari, així com la resposta de l'empresa i la data en què es va proporcionar.

6 TIPUS DE VISUALITZACIONS

És fonamental comprendre que la qualitat de les visualitzacions depèn en gran manera de la selecció adient del tipus de gràfic i de les dades que s'utilitzin per a cada cas, ja que no totes les visualitzacions són igualment adequades per a tots els tipus de dades i situacions.

Un cop es tenen les visualitzacions, s'han de poder respondre una sèrie de preguntes. Però, abans de començar a crear les representacions és important tenir en compte quines són aquestes preguntes per poder saber quins són els millors mètodes per fer-les. Cal tenir una comprensió clara de les dades que es tenen a disposició, les variables que s'estan analitzant i les hipòtesis que s'intueixen.

6.1 Preguntes

Com s'ha comentat, és necessari respondre a un seguit de preguntes, les quals es detallen a continuació:

1. Quina és la percepció general dels usuaris respecte a l'experiència amb l'aplicació?
2. Quins són els aspectes que els usuaris valoren més positivament i negativament de l'aplicació?
3. En quins àmbits les aplicacions analitzades destaquen o tenen punts forts i febles respecte a la competència?
4. Quins aspectes específics de l'aplicació haurien de ser prioritzats en la inversió de recursos per millorar la satisfacció dels usuaris i superar la competència?
5. En quina mesura les valoracions dels usuaris reflecteixen la seva satisfacció amb els serveis?

6.2 Estudi de mètriques i visualitzacions

En aquesta secció es relaciona cada pregunta plantejada amb les dades que s'utilitzaran per a la seva visualització i amb el tipus de representació visual que s'aplicarà.

S'han de recopilar les puntuacions dels usuaris per a cada aplicació que s'està analitzant, i generar un gràfic de barres amb una escala de l'1 al 5 per a representar els *ratings*. Cada barra representarà una puntuació específica, i la seva alçada indicarà la proporció o nombre de vegades que aquesta puntuació ha estat donada. Es pot incloure una línia discontinua que representa la puntuació mitjana de cada aplicació, la qual ens proporcionarà una visió general de la distribució de les puntuacions per obtenir una primera impressió de la percepció general dels usuaris sobre l'experiència amb cada aplicació. A la figura 3 en podem veure un exemple.

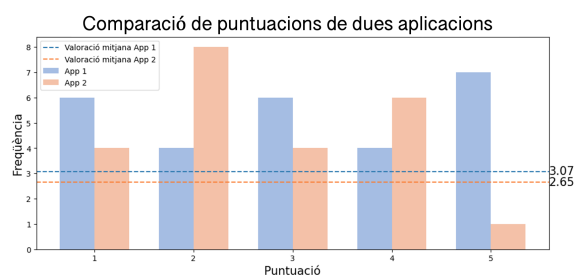


Fig. 3: Gràfic de barres per comparar puntuacions

Amb aquesta representació gràfica podem saber quina aplicació té una millor percepció per part dels usuaris. No obstant això, és important tenir en compte que aquesta és només una primera aproximació per a tenir una idea general. Per aconseguir conclusions més precises es realitzaran altres tipus d'anàlisis estadístiques i gràfiques sobre els comentaris i les puntuacions de l'aplicació.

Per la segona pregunta es pot utilitzar la tècnica de visualització de *wordcloud*. Implica crear un núvol de paraules per a cada aplicació, on la grandària de cada una reflecteix la seva freqüència d'ús. És necessari fer servir els comentaris que els usuaris han fet sobre l'aplicació, els quals han de ser prèviament processats i netejats per eliminar les paraules no rellevants.

A través d'aquesta visualització, es pot obtenir una comprensió ràpida i clara dels aspectes importants per als usuaris en relació amb cada aplicació. Per identificar els aspectes positius i negatius, es pot establir un *threshold* i utilitzar la puntuació associada als comentaris per filtrar i crear dues visualitzacions per a cada aplicació: una amb els comentaris amb una puntuació superior al límit establert, per comprendre què valoren els usuaris i detectar els punts forts, i una altra amb els comentaris amb una puntuació inferior al límit, considerats negatius, per identificar les àrees d'insatisfacció dels usuaris i punts a millorar (Figura 4).



Fig. 4: Wordclouds de comentaris positius i negatius

Per poder avaluar les aplicacions en diferents àmbits i comparar-les, es pot crear un gràfic de radar o de barres, permetent-nos identificar els punts forts i febles de cada una i comparar-les. Es poden categoritzar els comentaris en diferents aspectes rellevants i agrupar-los, com podrien

ser la seguretat, l'accessibilitat i la usabilitat. Les dades es representen mitjançant punts que es connecten per formar una forma poligonal, la qual ens facilita la identificació del rendiment de cada aplicació en cada aspecte, i d'aquesta manera obtenir una visió general de les diferències entre les aplicacions i identificar tant les seves fortaleeses com les seves debilitats en comparació amb la competència.

A la figura 5 es mostra un gràfic de radar amb 5 aspectes per comparar dues aplicacions, on s'avaluen de l'1 al 5.

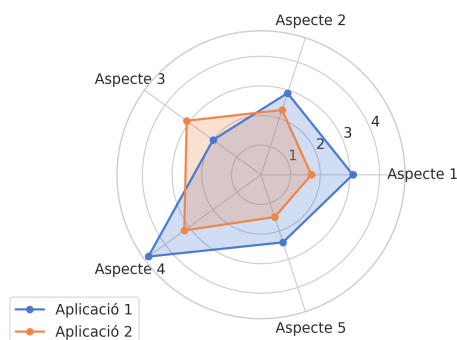


Fig. 5: Gràfic de radar dels aspectes de dues aplicacions

La següent pregunta està molt relacionada amb l'anterior i podria generar confusió, ja que aparentment és similar. L'anterior agrupa els comentaris per aspectes i els compara amb els de la competència, però no té en compte la importància relativa de cada un. És a dir, no considera la importància que els usuaris atorguen als diferents aspectes. Per aconseguir-ho, és necessari fer una anàlisi per a identificar els aspectes més rellevants que es repeteixen amb més freqüència i assignar-los un valor, i d'aquesta manera determinar en quines àrees és necessari invertir per tal de millorar la satisfacció dels usuaris i superar la competència. Es pot fer un diagrama de barres per poder comparar les dues aplicacions, ordenant els aspectes segons el seu nivell d'importància.

Per abordar la darrera pregunta, és possible explorar la relació entre les valoracions dels usuaris i les puntuacions assignades als seus comentaris utilitzant un model d'anàlisi de sentiment. Si es troba una correlació positiva significativa, indicaria que les valoracions dels usuaris estan vinculades a la seva satisfacció amb els serveis de l'aplicació.

Es pot fer servir un diagrama de barres apilades, on l'eix Y mostra la freqüència relativa de les puntuacions de l'1 al 5, i l'eix X representa les diferents puntuacions (Figura 6).

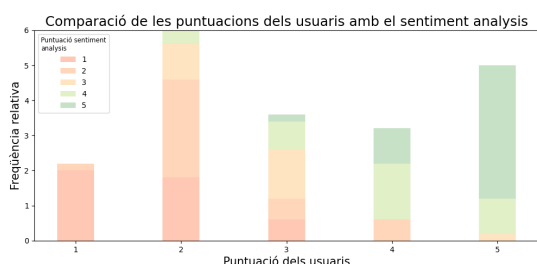


Fig. 6: Gràfic de barres apilades de les puntuacions dels usuaris i les de l'anàlisi de sentiment

Per avaluar de manera més precisa si existeix una correlació entre aquestes variables, es pot utilitzar una tècnica

estadística com el coeficient de correlació de *Pearson*, juntament amb una taula de contingència. Si els resultats indiquen que no hi ha una correlació positiva significativa, pot indicar una discrepància entre les percepcions dels usuaris i les puntuacions assignades. Si per exemple les puntuacions reals són més baixes que les predites, podria indicar que des d'un punt de vista objectiu, el problema no és tan greu com ho perceben alguns usuaris, però aquests consideren el problema com a molt important. En aquest cas, seria necessari analitzar més a fons aquests comentaris per identificar què és exactament allò que consideren tan greu, i així poder comprendre millor les expectatives i les necessitats dels usuaris per identificar àrees de millora.

7 PREPARACIÓ DE LES DADES

En aquesta secció es descriurà el flux de treball de l'ETL (*Extract, Transform i Load*) que s'ha utilitzat per al procés de neteja i preparació de les dades per dur a terme una anàlisi posterior.

Les dades que es fan servir es basen en les puntuacions assignades pels usuaris a l'aplicació i els comentaris que han escrit sobre ella. Aquests poden presentar diversos problemes que dificulten la seva anàlisi; la variabilitat en el format i l'idioma dels comentaris. Algunes persones els poden haver escrit en idiomes diferents o bé utilitzat paraules o frases que no segueixen les normes gramaticals.

Per aquesta raó, la primera tasca consisteix en la neteja i la normalització d'aquestes dades. Això implica diverses accions; la traducció dels comentaris i l'eliminació de la informació no rellevant, com ara emoticones, hashtags, símbols, entre d'altres.

És important realitzar la traducció a un idioma comú, ja que si estan escrits en diferents idiomes, la seva comprensió i avaluació posterior es dificulta. S'utilitza la API de *googletrans* [6] en *Python*, que proporciona accés al servei de traducció de Google. Mitjançant el mètode *translate* i especificant l'idioma de destí, es duu a terme la traducció de les dades. S'estan analitzant les dades d'Espanya i hi pot haver comentaris en castellà, català i anglès, entre d'altres, però per decidir en quin idioma traduïm, s'ha dut a terme un estudi dels percentatges de comentaris en cada idioma. El 88.75% dels comentaris estan en castellà, per tant, el castellà serà l'idioma de destí per tenir-los tots en la mateixa llengua.

També és important netejar-los [7] per assegurar que les dades estiguin en un format consistent. En primer lloc, es realitza la conversió del text a minúscules per evitar la distinció entre majúscules i minúscules en les paraules, i a continuació s'usa el mòdul d'expressions regulars de *Python*, conegut com a *"re"* [8], per eliminar símbols com *·*?;* que no aporten informació significativa. Utilitzant el mateix mòdul, s'eliminen els signes de puntuació i els dobles espais que puguin aparèixer en el comentari per tal de no afectar negativament l'anàlisi de text ni la detecció de paraules.

Un pas següent és l'eliminació de les emoticones, ja que no són considerades com a paraules significatives. A continuació, s'eliminen les paraules comunes o genèriques (*"stopwords"*) per centrar-se en les paraules que realment importen i que poden proporcionar informació valuosa. Per acabar, s'ha de dividir el text en unitats més petites, conegu-

des com a "tokens", per permetre l'anàlisi més detallada del contingut del text i identificar patrons específics de paraules que puguin ser rellevants.

Aquest apartat descriu el procés de neteja i normalització de les dades recopilades dels comentaris de les aplicacions per a l'estudi, el qual és essencial per assegurar que les dades estiguin en un format homogeni i consistent, i perquè la generació d'anàlisis i visualitzacions permetin una comprensió més detallada de les dades recopilades.

8 RESULTATS

En aquest apartat es presenta de manera detallada els resultats obtinguts i s'explica el procés seguit per aconseguir-los.

8.1 Percepció dels usuaris: Aspectes positius i negatius

Per a un primer contacte amb les dades s'ha utilitzat l'*excel* amb els *ratings*, i d'aquesta manera veure la percepció general dels usuaris respecte a l'experiència amb l'aplicació.

Com ja s'ha comentat en l'apartat 6.2, s'utilitza un gràfic de barres amb la mitjana global de cada aplicació. És important tenir en compte que el nombre de puntuacions disponibles per a cada aplicació és diferent, i per aquest motiu s'ha realitzat una normalització, de manera que cada aplicació tingui un pes equitatiu en la mitjana global, independentment del nombre de puntuacions. Es pot observar el gràfic final obtingut en la figura 7.

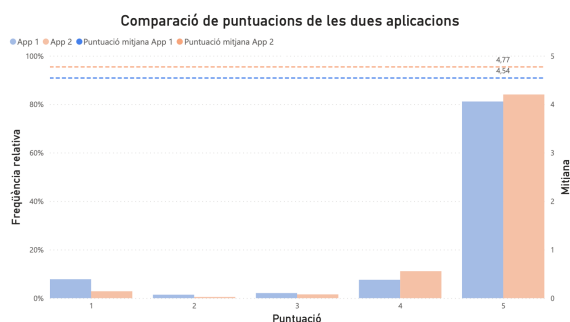


Fig. 7: Gràfic de barres dels *ratings* de les aplicacions

Tant l'aplicació 1 (d'estudi) com l'aplicació 2 (competència) presenten una concentració majoritària de puntuacions de 5, amb una distribució similar. No obstant això, hi ha una lleugera diferència en les freqüències relatives; l'aplicació 2 obté un percentatge més elevat d'aquestes puntuacions de 5 en comparació amb la 1.

Pel que fa a les puntuacions assignades amb el *rating* més baix, l'aplicació 2 en rep una quantitat menor. A causa d'aquestes diferències, les mitjanes també varien, amb un desavantatge de 0,23 punts per a l'aplicació 1. Això indica que la segona aplicació ofereix una experiència més satisfactòria pels usuaris en comparació amb la 1.

Un cop s'han analitzat els *ratings* de cada aplicació, es busca comprendre les raons darrere d'aquestes puntuacions. Un mètode eficaç per identificar els aspectes més importants de cada aplicació és mitjançant un *wordcloud*.

S'implementa utilitzant la llibreria *wordcloud* de *Python* i es realitza un procés de neteja per eliminar aquelles pa-

raules que no aporten informació rellevant i podrien distorsionar la visualització de les paraules importants. A la figura 8 es mostren dos *wordclouds*: a l'esquerra, amb els termes rellevants dels comentaris negatius, i a la dreta on es destaquen els aspectes positius. Per tal de facilitar-ne la comparació, s'ha normalitzat la mida de les paraules.



Fig. 8: *Wordclouds* de comentaris positius i negatius de l'aplicació analitzada

En el primer *wordcloud*, les paraules més prominents són 'bizum' i 'entrar'. Això suggereix que els principals problemes estan relacionats amb el funcionament del servei Bizum a l'aplicació i l'accés a aquest (ja que també apareixen paraules com 'acceder', 'abrir cuenta' i 'abre' amb una mida destacable). Es ressalten també paraules com 'lenta', 'tiempo' i 'carga', que poden indicar que l'aplicació és lenta en el seu funcionament, a més a més de paraules relacionades amb l'atenció al client, i 'publicidad' i 'anuncio'.

D'altra banda, en examinar el segon núvol de paraules, la més destacada és 'fácil usar'. Acompanyada de termes com 'intuitiva', 'práctica' i 'sencilla', podem concloure que l'aplicació té un bon disseny i que és valorat positivament pels clients, ja que la majoria dels termes que apareixen estan relacionats amb la facilitat d'ús i la comoditat del funcionament de l'aplicació. A més, es pot veure la presència de la paraula 'Bizum'. Aquest terme també figura en un gran format al *wordcloud* dels aspectes negatius. Per tant, és una funcionalitat valorada pels usuaris quan funciona adequadament, però si s'experimenten problemes pot tenir un impacte significatiu en l'experiència global dels usuaris.

A la figura 9 es mostren dos nous *wordclouds* que representen els comentaris de l'aplicació competidora, a través dels quals s'intenta obtenir una visió més completa de les opinions dels usuaris sobre l'aplicació rival.

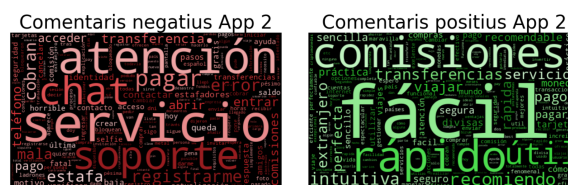


Fig. 9: *Wordclouds* de comentaris positius i negatius de la segona aplicació

Centrant-nos primer en els comentaris negatius, destaquen termes com 'estafa' i 'estafadores', suggerint que els usuaris han tingut experiències negatives relacionades amb possibles enganys o pràctiques fraudulentes. D'altra banda, sobresurt el tema 'atenció al client', ja que les paraules com 'servicio', 'chat', 'atención' i 'soporte' es repeteixen amb freqüència, coincidint també amb l'aplicació 1. Aquestes similituds ens permeten identificar problemes comuns o temes crítics que cal tenir en compte.

En el *wordcloud* de paraules positives es veu també les

paraules 'fácil', 'sencilla' i 'cómoda', indicant la facilitat d'ús de la interfície de l'aplicació. S'aprecia també la paraula 'comisiones' amb una gran mida, de manera que és un aspecte que agrada als usuaris, ja sigui per la seva absència o per tenir un cost raonable. Altres paraules que no són tan prominents són 'viajar', 'extranjero', 'divisa' i 'viaje'. Tenint en compte que hi ha diverses paraules d'aquest estil, podem inferir que l'aplicació destaca en temes relacionats amb viatges o transaccions internacionals.

8.2 Anàlisi comparativa: Àmbits destacats i àrees de millora

Un cop s'han examinat els temes principals tractats en els comentaris, és necessari classificar-los. Per dur a terme aquesta categorització temàtica, s'han provat diversos models. En primer lloc, es va intentar utilitzar el mètode de clusterització *K-means*, que consisteix a agrupar les dades en *clústers* (grups) basant-se en les seves característiques, però els resultats obtinguts no van ser els esperats, ja que no s'obtenien agrupacions adequades.

Vaig decidir provar un altre mètode de clusterització, l'algorisme LDA (Anàlisi de *Latent Dirichlet Allocation*), però tampoc va funcionar de manera correcta per les dades que hi ha. Això es deu al fet que LDA té dificultats per identificar els temes en textos breus, ja que a l'haver-hi una quantitat limitada de paraules es dificulta l'extracció d'informació significativa [9].

Finalment, es va crear un classificador. Per implementarlo, el primer pas va ser determinar quines categories s'ajustaven millor als comentaris. Es van considerar els resultats dels *wordclouds* explicats anteriorment i es van explorar els comentaris, i un cop les categories estaven clares vaig dedicar uns dies a classificar els comentaris de cada aplicació. En tenir un percentatge de comentaris classificats (aproximadament una quarta part), es va procedir a entrenar un classificador utilitzant aquestes dades etiquetades.

És necessari vectoritzar els comentaris per poder-los processar de manera adequada. En aquest cas, es va usar el mètode de vectorització TF-IDF (*Term Frequency-Inverse Document Frequency*), que assigna pesos als termes en funció de la seva freqüència relativa i importància en un comentari. Una vegada estan vectoritzats i es converteixen en representacions numèriques, s'utilitzen en un classificador. S'empra el model *MultiLabelBinarizer*, que permet tractar múltiples etiquetes de manera simultània. Aquesta és una opció útil per a quan hi ha més d'una categoria possible per a cada comentari.

Amb el model ja entrenat, s'aplica el classificador als comentaris sense etiqueta. La precisió ens serveix per saber el percentatge de temes classificats correctament respecte al total de temes avaluats, el qual en aquest cas ha donat d'un 78%.

Amb les dades classificades, s'han generat els gràfics de radar esmentats a la secció 6.2. A la figura 10 s'observen els resultats obtinguts. Per cada categoria, s'ha calculat la mitjana dels comentaris per cada aplicació, així com el nombre total de comentaris per categoria i aplicació. Tots aquests càlculs han estat normalitzats, tenint en compte que no hi ha el mateix nombre de comentaris en cada aplicació.

A simple vista sembla que les dues aplicacions tenen una distribució similar, però això només és cert si considerem

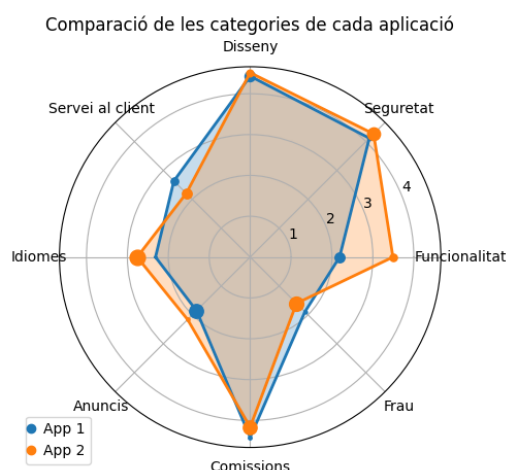


Fig. 10: Gràfic de radar de les categories de cada aplicació

la mitjana, sense tenir en compte el nombre de comentaris totals en cada una de les categories, el qual es pot identificar mirant la mida de les circumferències. La categoria 'disseny' és la que presenta més semblances entre les dues aplicacions; tant el nombre de comentaris com la mitjana dels comentaris relacionats amb aquest tema són gairebé iguals, amb una puntuació propera a 5. Això indica que els usuaris estan satisfets amb la interfície gràfica de les dues aplicacions, ja sigui perquè és intuïtiva o estètica.

Un altre aspecte que és relativament semblant és el 'servei al client', tot i que la mitjana de la segona aplicació és lleugerament inferior (prop del 2) en comparació amb la primera aplicació. A més, hi ha més comentaris sobre aquest tema en la segona, per tant, podríem afirmar que aquest aspecte és una debilitat per a ambdues aplicacions, però hi ha més queixes per part dels clients de la segona.

Fixant-nos en la categoria 'Seguretat', s'observa que la mitjana és propera a 4 per a les dues aplicacions. No obstant això, és evident que hi ha un nombre més gran de comentaris en l'aplicació 2. Aquesta alta quantitat de comentaris indica que la seguretat és una característica destacada i valorada positivament pels usuaris de la segona aplicació. De la mateixa manera, la categoria 'Comissions' té una mitjana propera a 5 per a les dues, però el nombre de comentaris en la segona és molt més elevat. En l'estudi dels comentaris sobre aquest tema, destaquen les referències a les baixes comissions o fins i tot la possibilitat d'efectuar pagaments en diferents divises sense càrrec addicional.

Els 'anuncis' són una característica mal valorada pels usuaris de l'aplicació principal. Quan obren l'aplicació, es troben amb molts *pop-ups* publicitaris que han de tancar constantment per tal de continuar utilitzant-la.

Finalment, pel que fa a la 'funcionalitat', hi ha més o menys el mateix nombre de comentaris, però a diferència de les altres, aquesta categoria té una mitjana molt diferent entre aplicacions. Per poder investigar més, dintre de 'funcionalitat' s'han categoritzat els comentaris en diferents subgrups. A la figura 11 es veuen els resultats obtinguts.

La distribució de les dues aplicacions en aquest gràfic mostra diferències significatives en comparació amb l'anterior. Hi ha tres subgrups en els quals l'aplicació 2 destaca de manera positiva respecte a la primera: 'Bizum', 'Estranger' i 'Lent'. El més clar i evident és el d'Estranger', ja

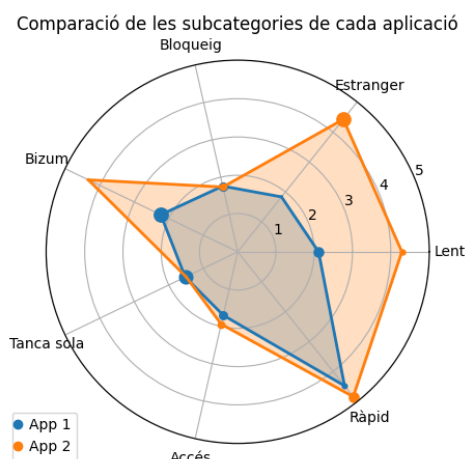


Fig. 11: Gràfic de radar dels subgrups de cada aplicació

que no només té una mitjana més elevada, sinó que també té un nombre considerablement més gran de comentaris. En investigar-los, ràpidament es pot concloure que una de les funcionalitats més ben valorades pels clients és la facilitat d'utilitzar la targeta en altres països durant els viatges, la qual és una característica que ha estat destacada en el núvol de paraules analitzat anteriorment.

En el subgrup de 'Bizum', l'aplicació principal té una mitjana propera a 2, mentre que la segona és aproximadament de 4, tot i que aquesta última no té la funcionalitat. A primera vista, podríem suposar que, si l'aplicació no té aquesta funcionalitat, els comentaris relacionats amb *bizum* serien negatius, simplement destacant la seva absència. No obstant això, en explorar els comentaris es pot observar que la majoria dels usuaris de la segona aplicació, quan ho mencionen, expressen comentaris positius i a part indiquen que aquesta és una funcionalitat que apreciarien tenir. Malgrat la seva absència, valoren l'aplicació positivament. Passa el mateix amb la categoria 'Lent', ja que consideren que és dolent, però no tan greu com per valorar negativament l'aplicació en conjunt.

Pel que fa a la primera aplicació, tenint en compte la mitjana i el nombre elevat de comentaris relacionats amb 'Bizum', és evident que aquesta funcionalitat presenta molts problemes, on destaquen principalment els problemes de funcionament d'aquest servei.

Finalment, una característica rellevant a comentar és el grup de 'Tanca sola'. Si bé les dues aplicacions tenen una mitjana pràcticament igual en aquest aspecte, la primera té un nombre més gran de comentaris, indicant que experimenta amb més freqüència problemes de tancament automàtic, cosa que afecta negativament la seva valoració per part dels usuaris. Així, aquesta característica esdevé un problema recurrent que té un impacte directe en la percepció global de l'aplicació.

És important identificar aquells aspectes que tenen un major impacte en la valoració global de cada aplicació. Especialment, si es tracta d'aspectes negatius que poden ser millorats, ja que aquestes tenen el potencial de tenir un impacte significatiu en l'avaluació global de l'aplicació si se'n realitza una millora.

A partir dels comentaris classificats, calculem la freqüència de les categories per a cada aplicació, identi-

ficant les que es repeteixen més sovint entre els usuaris. Aquest procés s'aplica a cada aplicació per tal de tenir les dades normalitzades, ja que sinó les categories de l'aplicació amb un nombre més gran de comentaris tindrien una influència desproporcionada en els resultats.

A la figura 12 es presenta un diagrama de barres amb les categories (dreta) i subcategories (esquerra) ordenades segons importància de major a menor. S'indica la mitjana de cada categoria per a cada aplicació, així com el percentatge de comentaris que fan referència a aquest aspecte.

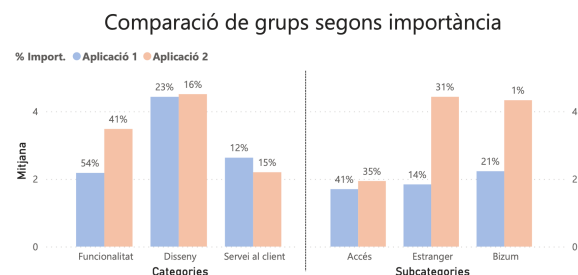


Fig. 12: Gràfic de barres de les categories i subcategories ordenades segons importància

Pel que fa a les categories, en primer lloc, tenim la de 'Funcionalitat', la qual segons els percentatges és la més important per ambdues aplicacions, representant més de la meitat dels comentaris en el cas de la primera aplicació. Aquesta té una puntuació mitjana de 2, mentre que la segona és de 3,5. Com que aquesta categoria es divideix en diverses subcategories, s'aprofundirà més en aquest aspecte amb la gràfica de la dreta.

En segon lloc, trobem la categoria 'Disseny', que també coincideix sent la segona més valuosa en totes dues aplicacions. En aquest cas, les dues mitjanes són altes, pràcticament arribant a 5. Això indica que el disseny és una característica molt ben valorada per la majoria dels usuaris.

Finalment, ens trobem amb la categoria 'Servei al client', valorada de manera negativa en ambdues aplicacions, tot i ser lleugerament millor la mitjana de la primera. Aquest és un servei que els usuaris consideren important, però que no funciona com desitjarien o no compleix les seves expectatives.

Si ens fixem en els subgrups, el més comentat és 'Accés' amb una puntuació molt baixa. Aquest aspecte és clau pels usuaris, però el seu mal funcionament contribueix a una valoració negativa de l'aplicació, afectant l'experiència global.

La següent funcionalitat, 'Estranger', és molt rellevant pels usuaris de la segona aplicació, amb un 31% dels comentaris. En canvi, per la primera aplicació quasi no se'n parla. La raó d'aquesta diferència és que la segona aplicació es distingeix per ser una solució financera líder en el sector dels viatges, oferint funcionalitats com el canvi de divises i els pagaments internacionals sense comissions. En canvi, la primera aplicació ofereix una àmplia gamma de serveis, però no funcionalitats específicament orientades als viatges.

Finalment, pel que fan els comentaris sobre 'Bizum', passa el contrari que l'anterior, on per la primera aplicació és molt comentat, però no per la segona. Això és degut al fet que aquesta segona no disposa d'aquesta característica, com bé ja s'ha comentat a la secció 8.2. És una característica molt important, però amb una valoració ne-

gativa. És similar al que passa amb 'Accés', ja que si els problemes relacionats amb el funcionament de 'Bizum' es resolguessin, el percentatge de valoracions negatives causades per aquest mal funcionament disminuiria, comportant un augment en la valoració global de l'aplicació.

8.3 Priorització d'aspectes

Gràcies als resultats anteriors podem identificar aquells aspectes que cal prioritzar en la inversió de recursos. S'ha creat una taula per mostrar els efectes de millorar aquells més importants. Pels tres primers (accés, bizum i servei al client), s'ha calculat una nova mitjana en el cas que aquest aspecte no presentés cap problema a l'aplicació, i per la categoria 'Estranger', s'ha calculat la nova mitjana en un escenari en què l'aplicació 1 disposés d'aquesta funcionalitat.

A la taula 3, es mostren les noves mitjanes obtingudes i la puntuació guanyada per la primera aplicació en relació amb la mitjana original. Cal tenir en compte que aquest valor és una aproximació, ja que no podem predir exactament què succeiria en una situació hipotètica.

TAULA 3: CATEGORIES A MILLORAR

Categories	Rating	Guanyat
Accés	3.96	0.19
Bizum	3.90	0.13
Servei al client	3.89	0.12
Estranger	3.83	0.06
Total	4.23	0.46

L'aspecte que té un impacte més significatiu en la mitjana és la funcionalitat 'Accés', seguida de 'Bizum'. A l'última fila es mostra la nova mitjana que s'obtingria si no hi hagués cap problema en els tres primers aspectes i si existís la funcionalitat relacionada amb els comentaris de l'estranger. Aquesta millora és de 0.46 punts en comparació amb la mitjana original de 3.77, superant així la mitjana de la segona aplicació, que és de 4.21.

8.4 Tractament de valoracions subjectives

Per tal de determinar si les puntuacions dels usuaris reflecteixen la seva satisfacció amb els serveis, s'ha realitzat un model d'anàlisi de sentiment. S'ha utilitzat un model pre-existent basat en BERT, un algorisme de processament de llenguatge natural (NLP) adaptat per l'anàlisi de sentiment de *reviews* de productes [10]. Aquest model assigna una puntuació de l'1 al 5 a cada text, on 1 representa una connotació negativa i 5 una positiva. A causa de l'alta precisió dels resultats del model en anglès, s'han traduït els comentaris a aquest idioma, seguint el mateix procés descrit a la secció de processament de dades (secció 7), però en aquest cas, a l'anglès.

Els resultats obtinguts per la primera aplicació es poden veure a la figura 13, en un gràfic de barres apilades amb la relació entre els valors reals de les puntuacions assignades pels usuaris i els valors assignats pel model d'anàlisi de sentiment. Cada barra del gràfic mostra la freqüència dels valors de *rating* i es divideix en diversos segments per mostrar la freqüència de cada valor de la puntuació del model.

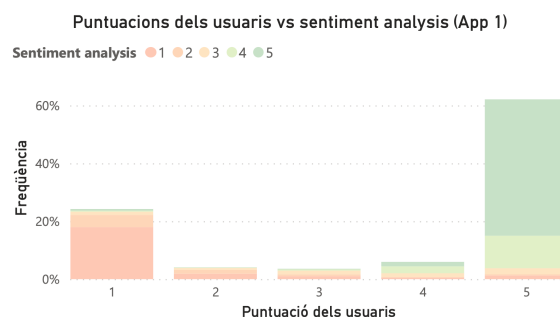


Fig. 13: Gràfic de barres dels *ratings* dels usuaris amb les puntuacions de l'anàlisi de sentiment (app 1)

A més del gràfic, també s'han calculat diverses mètriques per obtenir resultats numèrics. La mitjana de les puntuacions assignades pels usuaris és de 3.77, mentre que la de les puntuacions generades pel model és de 3.61. Això indica una disminució de 0.16 en les puntuacions estimades pel model respecte a les puntuacions dels usuaris.

A la figura 14 es presenta el mateix gràfic per a la segona aplicació, on la mitjana de les puntuacions dels usuaris en aquest cas és de 4.21, mentre que la mitjana del model d'anàlisi de sentiment és de 4.03, amb una diferència de 0.18.

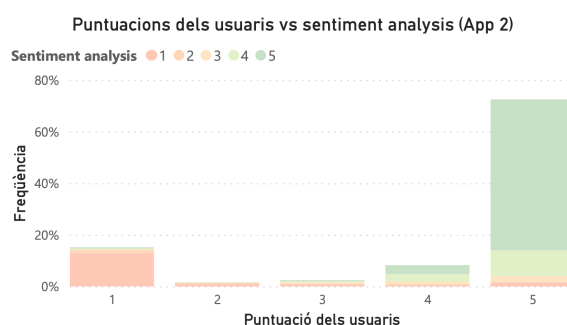


Fig. 14: Gràfic de barres dels *ratings* dels usuaris amb les puntuacions de l'anàlisi de sentiment (app 2)

Les prediccions del model d'anàlisi de sentiment no s'han allunyat gaire de la realitat, i mirant les mitjanes, només varia 0.16 i 0.18 respectivament. D'altra banda, analitzant les desviacions de la distribució de les puntuacions, de la primera aplicació és de 1.72, mentre que la segona té una desviació estàndard de 1.46. Això ens indica que els usuaris de la primera tendeixen a donar puntuacions més extremes, és a dir, hi ha una major concentració de puntuacions molt baixes o molt altes.

Tot i això, abans de treure conclusions, s'ha fet per cada aplicació una taula de contingència, i s'ha calculat el coeficient de correlació de *Pearson*. El resultat es pot veure a la figura 15.

Els coeficients són significativament alts, amb un valor de 0.86 per a l'aplicació 1 i 0.84 per a l'aplicació 2. Això indica una forta correlació, de tal manera que podem afirmar que les puntuacions assignades pels usuaris reflecteixen la seva satisfacció amb els serveis proporcionats per les aplicacions, tenint en compte els comentaris que han realitzat.

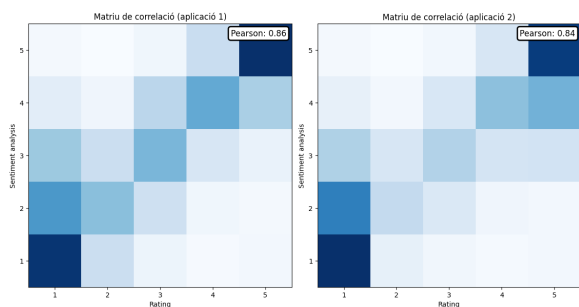


Fig. 15: Taules de contingència dels *ratings* dels usuaris amb les puntuacions de l'anàlisi de sentiment

9 CONCLUSIONS

En primer lloc, es pot dir que les dues aplicacions presenten puntuacions molt altes dels *ratings*, amb una diferència de tan sols 0.23 punts. Tot i que en alguns contextos aquesta diferència podria considerar-se no significativa, cal tenir en compte que és la mitjana global de l'aplicació i que és la primera cosa que els usuaris veuen en instal·lar-la. Per tant, aquesta discrepància adquireix una gran rellevància i resulta crucial millorar-la.

D'altra banda, s'han identificat els aspectes que els usuaris valoren de manera positiva i negativa. Per superar la competència, és imprescindible millorar els aspectes de l'aplicació principal que presenten deficiències, així com mirar d'implementar els serveis que són apreciats pels usuaris i que estan només disponibles en l'aplicació rival.

La mitjana dels *ratings* dels comentaris sol ser inferior a la puntuació global, ja que és més comú que els usuaris ressaltin els problemes en lloc dels aspectes positius. Dit això, el fet que la facilitat per realitzar tràmits i gestions mentre es viatja sigui tan aclamat en els comentaris de la segona aplicació, reflecteix la gran importància que té pels clients. A causa d'això, és possible que molts optin per altres aplicacions competidores només per aprofitar aquesta funcionalitat que no es troba disponible en l'aplicació principal.

Com ja s'ha comentat durant el treball, hi ha tres altres serveis de gran importància que són mal valorats pels clients: el funcionament del Bizum, l'accés a l'aplicació i el servei al client. Millorar el seu funcionament garantiria una experiència més fluida i sense obstacles per als usuaris, afavorint així la seva satisfacció i fidelitat. D'aquesta manera generaria una millor reputació per a l'aplicació, que a la llarga pot contribuir a l'increment de l'ús de l'aplicació i augmentar els beneficis.

10 PROBLEMES I MILLORES

Un dels principals reptes d'aquest projecte ha estat la dificultat de categoritzar les *reviews* dels usuaris, ja que la informació pot prendre diverses formes d'expressió i té una naturalesa complexa.

Pel que fa a les millores, es podria explorar l'ús d'un classificador més avançat amb un valor de precisió més elevat. Això permetria obtenir un millor rendiment en la identificació i agrupació de les dades en categories rellevants, abordant així els problemes detectats. A més, per aconseguir resultats més precisos, es pot fer una anàlisi més exhaustiva de les dades. Això pot implicar examinar altres dades dis-

ponibles, com ara tenir en compte les diferents versions de l'aplicació o les respostes de l'empresa als comentaris dels usuaris, tenint en compte també el temps de resposta.

AGRAÏMENTS

Agrair el suport del Dani Franco, i del meu tutor Toni Espinosa, per la seva col·laboració i orientació durant el desenvolupament d'aquest treball.

REFERÈNCIES

- [1] *¿Qué es Power BI?*, Power BI. [Online]. Disponible: <https://powerbi.microsoft.com/es-es/what-is-power-bi/>. [Accedit: Mar. 01, 2023]
- [2] H. OKAM, *Google Play Store Apps analysis and visualization*, GitHub, Gen. 31, 2023. [Online]. Disponible: <https://github.com/mchenryspag/Google-Play-Store-Apps-Analysis-Visualization>. [Accedit: Mar. 03, 2023]
- [3] S. ALAM, *Northwind-Sales-Analysis*, GitHub, Feb. 27, 2022. [Online]. Disponible: <https://github.com/ShehzadaAlam/Northwind-Sales-Analysis>. [Accedit: Mar. 03, 2023]
- [4] V. VEMPATI, *Google Maps vs Waze 2020 - Analyzing 10,000 App Reviews*, Thematic, Mar. 11, 2020. [Online]. Disponible: <https://getthematic.com/insights/google-maps-vs-waze/>. [Accedit: Abr. 07, 2023]
- [5] *Magic Quadrant for Analytics and Business Intelligence Platforms*, Gartner.com, Mar. 22, 2022. [Online]. Disponible: <https://www.gartner.com/doc/reprints?id=1-2916ZNCL&ct=220325&st=sb>. [Accedit: Mar. 04, 2023]
- [6] S. HAN, *googletrans: Free Google Translate API for Python*, PyPI, Juny 14, 2020. [Online]. Disponible: <https://pypi.org/project/googletrans/>. [Accedit: Abr. 11, 2023]
- [7] R. FERNANDEZ, *Cómo limpiar el texto manualmente usando NLTK*, Cursos de Programación de 0 a Experto, Jul. 24, 2018. [Online]. Disponible: <https://unipython.com/como-limpiar-el-texto-manualmente-usando-nltk/>. [Accedit: Abr. 11, 2023]
- [8] *re - Operaciones con expresiones regulares*, Python documentation. [Online]. Disponible: <https://docs.python.org/es/3/library/re.html>. [Accedit: Abr. 11, 2023]
- [9] K. ZHOU and Q. YANG, *LDA-PSTR: A Topic Modeling Method for Short Text*, *Advanced Data Mining and Applications*, vol. 11323, Nov. 2018. [Online]. Disponible: https://doi.org/10.1007/978-3-030-05090-0_29. [Accedit: Maig 04, 2023]
- [10] L. DEBUT, *bert base multilingual uncased sentiment*, Hugging Face. [Online]. Disponible: <https://huggingface.co/nlptown/bert-base-multilingual-uncased-sentiment>. [Accedit: Abr. 28, 2023]