UNIVERSIDAD EUROPEA DE MADRID

ESCUELA DE ARQUITECTURA, INGENIERÍA Y DISEÑO

MÁSTER UNIVERSITARIO EN ANÁLISIS DE DATOS MASIVOS (BIG DATA)

TRABAJO FIN DE MÁSTER

SISTEMA DE RECOMENDACIÓN DE VIAJES PERSONALIZADOS BASADO EN LA DETECCIÓN EMOCIONAL MEDIANTE PLN

FRANCISCO JAVIER CRISTÓBAL FERNÁNDEZ

Dirigido por

[Dr.] ÁLVARO MANUEL RODRÍGUEZ RODRÍGUEZ

CURSO 2025-2026

TÍTULO: SISTEMA DE RECOMENDACIÓN DE VIAJES PERSONALIZADOS BASADO EN LA DETECCIÓN EMOCIONAL MEDIANTE PLN

AUTOR: FRANCISCO JAVIER CRISTÓBAL FERNÁNDEZ

TITULACIÓN: MÁSTER UNIVERSITARIO EN ANÁLISIS DE DATOS MASIVOS (BIG DATA)

DIRECTOR/ES DEL PROYECTO: [Dr.] ÁLVARO MANUEL RODRÍGUEZ RODRÍGUEZ

FECHA: 13 de MARZO de 2026

RESUMEN

El presente Trabajo de Fin de Máster se enmarca en el ámbito del análisis de datos y la inteligencia artificial aplicada a la experiencia del usuario, y tiene como objetivo el diseño de un **sistema de recomendación de viajes personalizado basado en la detección emocional** mediante técnicas de Procesamiento del Lenguaje Natural (PLN). El proyecto parte de un **dataset real de reseñas en español** de Filmaffinity obtenido de la plataforma Kaggle, transformado y etiquetado manualmente para crear un **Gold Standard de 500 textos**, equilibrado en cinco emociones principales (tristeza, ansiedad, calma, curiosidad y alegría). El proceso de etiquetado se realizó de forma **dual y calibrada**, utilizando tres métricas de consistencia (porcentaje de acuerdo, Kappa ponderado cuadrático y discrepancia máxima de dos saltos) para garantizar la fiabilidad del corpus.

Sobre este conjunto de datos se desarrolló un **pipeline exhaustivo de preprocesamiento** (tokenización, lematización, eliminación de stopwords y términos cinematográficos, normalización y neutralización de entidades), a partir del cual se entrenaron **modelos supervisados** de clasificación emocional utilizando técnicas BoW y TF-IDF combinadas con algoritmos SVM y Naive Bayes.

Con el fin de evaluar la capacidad de generalización del sistema, se construyó adicionalmente un **corpus out-of-domain compuesto por 25 perfiles ficticios heterogéneos** y 300 textos con características propias del lenguaje informal de redes sociales. Este entorno permitió validar el rendimiento del modelo en condiciones más exigentes y realistas.

El resultado constituye la **base metodológica y técnica del sistema de recomendación emocional TRIP(U)**, que combinará el estado afectivo detectado en los textos del usuario con variables de perfil, intereses y presupuesto, promoviendo una experiencia de viaje más empática, personalizada y contextualizada.

Palabras clave: *Procesamiento del Lenguaje Natural, análisis emocional, sistema de recomendación, computación afectiva, TF-IDF, BoW, SVM, Naive Bayes.*

ABSTRACT

This Master's Thesis is framed within the field of data analysis and artificial intelligence applied to user experience. Its main goal is the design of a **personalized travel recommendation system** based on emotional detection using Natural Language Processing (NLP) techniques. This project builds upon a real dataset of Spanish-language film reviews originally sourced from *FilmAffinity* and made available via Kaggle, which was transformed and manually labeled to create a **Gold Standard of 500 texts**, balanced across five main emotions (sadness, anxiety, calmness, curiosity, and joy). The labeling process was conducted in a **dual and calibrated** manner, employing three consistency metrics (agreement rate, weighted quadratic kappa, and a maximum discrepancy threshold of two steps) to ensure corpus reliability.

On this dataset, an **exhaustive preprocessing pipeline** was developed —including tokenization, lemmatization, stopword and film-related term removal, normalization, and entity neutralization— from which **baseline supervised models** for emotional classification were trained using TF-IDF & BoW techniques combined with SVM and Naive Bayes algorithms.

In order to evaluate the system's generalization capability, an **additional out-of-domain corpus was constructed, composed of 25 heterogeneous fictional profiles** and 300 texts reflecting informal social media language characteristics. This environment enabled the validation of model performance under more demanding and realistic conditions.

The result establishes the **methodological and technical foundation of the TRIP(U) emotional recommendation system**, which combines the affective state inferred from user texts with profile, interest, and budget variables, thereby promoting a more empathetic and emotionally adaptive travel experience.

Keywords: *Natural Language Processing, emotional analysis, recommendation system, affective computing, TF-IDF, BoW, SVM, Naive Bayes.*

AGRADECIMIENTOS

A lo largo de este trabajo he contado con el apoyo y la comprensión de muchas personas que, de una u otra forma, hicieron posible su desarrollo.

En primer lugar, quiero agradecer a mi tutor académico, el Dr. Álvaro Manuel Rodríguez Rodríguez por su orientación, confianza y disposición para guiarme durante todo el proceso.

Agradezco también a Diana, la anotadora colaboradora, su mirada complementaria y sensibilidad social, que fueron esenciales para construir el proceso de etiquetado dual y alcanzar un corpus equilibrado y riguroso.

A mis padres, por ser el mejor ejemplo posible de amor incondicional, por estar en las duras y en las maduras, por ayudarme a ser una persona de bien y por apoyarme sin importar las consecuencias a forjar el camino que me ha traído hasta aquí. A pesar de que unas pocas líneas no hagan justicia a todo lo que tengo que agradecerles, por todo ello y por mucho más, este trabajo es por y para ellos.

Y a mi dulce Hanita, la pequeña flor del jardín, por su complicidad e inocencia, la ternura hecha ser, le dedico estas líneas que desearía pudiera leer.

Finalmente, dedico este trabajo a todos aquellos que creen en el poder de los datos para mejorar la experiencia humana, no solo desde la tecnología, sino desde la empatía.

"No hay datos sin emoción, ni emoción sin contexto. Solo cuando entendemos ambos, podemos crear tecnología verdaderamente humana."

F. J. Cristóbal Fernández

TABLA RESUMEN

	DATOS
Nombre y apellidos:	Francisco Javier Cristóbal Fernández
Título del proyecto:	Sistema de recomendación de viajes personalizados basado en la detección emocional mediante PLN
Directores del proyecto:	[Dr.] Álvaro Manuel Rodríguez Rodríguez
El proyecto se ha realizado en colaboración de una empresa o a petición de una empresa:	NO
El proyecto ha implementado un producto: (esta entrada se puede marcar junto a la siguiente)	NO
El proyecto ha consistido en el desarrollo de una investigación o innovación: (esta entrada se puede marcar junto a la anterior)	SI
Objetivo general del proyecto:	Diseñar y desarrollar un sistema de recomendación de viajes personalizados que integre el análisis emocional del usuario, utilizando técnicas de Procesamiento del Lenguaje Natural para generar experiencias de viaje más empáticas y contextualizadas

Índice

Índice de Figuras

Índice de Tablas

Capítulo 1. RESUMEN DEL PROYECTO

1.1 Contexto y justificación

El presente proyecto se enmarca en el ámbito del análisis de datos y la Inteligencia Artificial aplicada a la experiencia del usuario, con un enfoque centrado en la detección emocional a partir del lenguaje. En un entorno cada vez más orientado a la personalización, comprender la influencia de las emociones en la toma de decisiones resulta fundamental. En este contexto surge TRIP(U), un sistema de recomendación de viajes que integra el análisis emocional como variable central del proceso. Su lema, *"La variable eres tú, el viaje es tu destino"*, sintetiza su filosofía: cada usuario genera una experiencia distinta, moldeada por su estado afectivo.

1.2 Planteamiento del problema

Las plataformas de recomendación tradicionales se basan en datos objetivos sin considerar la dimensión afectiva del usuario. Este proyecto plantea cómo incorporar el estado emocional, detectado mediante *Procesamiento del Lenguaje Natural (PLN)*, como componente clave en el proceso de recomendación, combinando análisis de sentimiento y computación afectiva.

1.3 Objetivos del proyecto

El objetivo principal es desarrollar un sistema de recomendación de viajes que integre el análisis emocional del usuario para ofrecer sugerencias más empáticas. Para ello, se construye un corpus emocional en español, se entrenan y validan modelos de clasificación in-domain y out-of-domain, y se diseña una arquitectura multicriterio que incorpora emoción, intereses, perfil viajero y presupuesto. Los objetivos generales y específicos se detallan en el *Capítulo 3*.

1.4 Resultados obtenidos

El proyecto culmina con la implementación de un sistema de recomendación emocional capaz de clasificar el estado afectivo del usuario y generar rankings personalizados de destinos. El modelo seleccionado mostró solidez en validación in-domain y capacidad de generalización out-of-domain, incorporando además filtros éticos de protección contextual.

1.5 Estructura de la memoria

La memoria se estructura en 9 capítulos: justificación, objetivos y metodología, desarrollo técnico, resultados, discusión y conclusiones, incluyendo referencias bibliográficas y anexos.

Capítulo 2. ANTECEDENTES / ESTADO DEL ARTE

2.1 Estado del arte

La cuestión central de este proyecto es **cómo incorporar el estado emocional inferido del lenguaje natural como señal contextual y ordinal** dentro de un sistema de recomendación de viajes, más allá de los atributos objetivos clásicos (historial, demografía, preferencias).

En psicología, el **modelo circunflejo de la emoción de Russell** [1] organiza los afectos en dos ejes continuos: **valencia** (positiva/negativa) y **activación** (alta/baja), proporcionando una geometría conceptual que **ordena** las categorías emocionales y permite tratar sus etiquetas con **relaciones de cercanía/lejanía**. Este marco teórico resulta especialmente adecuado para tareas computacionales, ya que habilita codificaciones **ordinales** (p. ej., 0–4) y fundamenta la idea de que **no todos los errores de clasificación emocional tienen el mismo coste**.

En el ámbito del *Procesamiento del Lenguaje Natural* y la **computación afectiva**, la literatura ha transitado desde enfoques léxicos y *Bag of Words* hacia **modelos de aprendizaje profundo** como las **redes neuronales recurrentes (RNN)**, capaces de capturar dependencias secuenciales, y los **modelos *transformer***, que aprenden relaciones de largo alcance. Estas arquitecturas han permitido avances sustanciales en la detección emocional en texto, mejorando la comprensión contextual [2], [3].

La investigación también subraya dos retos clave: (i) la **subjetividad** del etiquetado emocional, que exige protocolos rigurosos (guías claras, doble anotación, resolución de discrepancias), y (ii) la necesidad de **métricas de calidad** acordes a la naturaleza ordinal de las emociones. En este sentido, el **kappa ponderado cuadrático** de **Cohen** [4] resulta especialmente apropiado, al penalizar más los desacuerdos lejanos que los próximos.

Por otro lado, diversos autores advierten sobre los **sesgos** presentes tanto en los datos como en los procesos de anotación (por ejemplo, sesgos de frecuencia, de saliencia o sociales), que pueden distorsionar la interpretación afectiva y su generalización. Para mitigarlos, se recomiendan conjuntos balanceados, **asimetrías controladas** que eviten el conteo de etiquetas, y la **neutralización de entidades** (nombres propios, títulos) para favorecer la **transferencia fuera del dominio** [5].

Desde la perspectiva algorítmica, los **modelos supervisados** como **TF-IDF combinados con SVM o Naive Bayes** siguen siendo referencias sólidas por su interpretabilidad y eficiencia con corpus de tamaño moderado. Sobre esta base, los **Transformers pre-entrenados** ofrecen un potencial adicional para capturar matices contextuales y variaciones semánticas, aunque a mayor coste computacional y menor interpretabilidad.

En síntesis, el estado del arte respalda: (a) la **ordinalidad** de las etiquetas emocionales, (b) la evaluación de concordancia con **Kappa ponderado cuadrático**, (c) protocolos de anotación y **mitigación de sesgos** para asegurar la fiabilidad, y (d) el **uso de modelos supervisados** interpretables, dejando abierta la incorporación futura de arquitecturas más expresivas como los Transformers para reforzar la capacidad contextual del sistema.

2.2 Contexto y justificación

El presente proyecto se sitúa en la intersección entre el **Procesamiento del Lenguaje Natural (PLN)**, la **computación afectiva** y los **sistemas de recomendación personalizados**, con el propósito de explorar cómo la **dimensión emocional** puede incorporarse de forma estructurada y medible en los procesos de decisión automatizados. En un contexto donde la inteligencia artificial se orienta cada vez más hacia la **personalización de experiencias**, comprender la **influencia de las emociones en la conducta humana** se convierte en un factor esencial para avanzar hacia sistemas más empáticos y adaptativos.

Los sistemas de recomendación actuales (ya sean basados en contenido, filtrado colaborativo o modelos híbridos) operan principalmente sobre variables estáticas como el historial de interacción, la similitud de preferencias o la demografía del usuario. Sin embargo, estos enfoques omiten una dimensión clave: el **estado emocional del usuario**, que condiciona tanto la percepción como la elección. El proyecto **TRIP(U)** propone precisamente introducir esta variable dinámica como **eje de personalización**, integrando emociones inferidas del lenguaje natural en la recomendación de destinos de viaje.

Desde el punto de vista académico, el proyecto contribuye al campo del **análisis emocional en español**, mediante la construcción de un **corpus calibrado**, generado a través de un proceso de **etiquetado dual con métricas de acuerdo ordinal**, que busca garantizar la consistencia y transferibilidad de los resultados. Este corpus no solo permite entrenar modelos de detección

emocional, sino también sentar las bases para futuras investigaciones sobre **sistemas de recomendación afectiva**.

En el plano aplicado, **TRIP(U)** funciona como un **prototipo conceptual** de recomendador emocionalmente consciente, donde las emociones detectadas modulan el peso de otras variables de decisión. Su diseño persigue generar experiencias de viaje **contextualizadas y empáticas**, que integren la racionalidad de los datos con la subjetividad de las emociones.

Así, el proyecto se justifica como un **puente entre ciencia de datos y experiencia humana**, aportando una aproximación innovadora a la personalización mediante IA afectiva. En el siguiente apartado se desarrolla el **planteamiento del problema**, que sintetiza las limitaciones identificadas en la literatura y los vacíos que este trabajo busca abordar.

2.3 Planteamiento del problema

A partir del análisis del estado del arte, se identifica una carencia clara en la integración efectiva de la **dimensión emocional** dentro de los **sistemas de recomendación personalizados**. Aunque la literatura en *Procesamiento del Lenguaje Natural (PLN)* y *computación afectiva* ha avanzado significativamente en la **detección automática de emociones**, su aplicación práctica en motores de recomendación sigue siendo limitada. En la mayoría de los casos, las emociones se tratan como un metadato secundario o se reducen a una polaridad binaria (positiva/negativa), lo que impide capturar la complejidad real del estado afectivo del usuario.

Por otro lado, aunque existen datasets en español con contenido potencialmente emocional (como reseñas o comentarios en línea), la mayoría **carecen de un etiquetado manual riguroso, métricas de validación inter-anotador y control de sesgos**. Esta falta de recursos **calibrados** dificulta el desarrollo de modelos fiables y su transferencia a contextos distintos del dominio de origen. En este sentido, el proyecto **TRIP(U)** aborda dicha limitación mediante la **construcción de un Gold Standard emocional en español**, derivado de un dataset real (FilmAffinity) y validado mediante un proceso dual de anotación y métricas de consistencia ordinal.

De forma complementaria, se detecta la **ausencia de una conexión directa** entre la emoción inferida a partir del lenguaje y los algoritmos de recomendación, lo que impide incorporar el componente afectivo como variable contextual en la personalización de experiencias.

Así, el problema que este trabajo aborda puede sintetizarse en la siguiente pregunta de investigación aplicada:

¿Cómo puede incorporarse el estado emocional, inferido mediante técnicas de PLN, como una variable dinámica y contextual dentro de un sistema de recomendación de viajes?

Responder a esta pregunta requiere articular un enfoque metodológico que combine **rigurosidad en la creación del corpus**, **evaluación transparente de modelos emocionales** y **aplicabilidad práctica** en un entorno de recomendación realista, avanzando hacia sistemas más **humanizados y empáticos**.

Capítulo 3. OBJETIVOS

3.1 Objetivos generales

El objetivo general del presente trabajo consiste en diseñar y desarrollar un sistema de recomendación de viajes personalizados que integre el análisis emocional del lenguaje del usuario mediante técnicas de Procesamiento del Lenguaje Natural (PLN), con el fin de generar sugerencias contextualizadas y empáticas basadas en su estado afectivo.

3.2 Objetivos específicos

O1. Curar y estructurar un corpus emocional en español de alta calidad a partir de reseñas reales, garantizando su equilibrio y fiabilidad mediante un proceso de etiquetado dual ciego y métricas de consistencia inter-evaluador.

O2. Implementar un pipeline de preprocesamiento lingüístico que optimice los textos para su análisis, incluyendo tokenización, lematización y enmascaramiento de términos in-domain.

O3. Entrenar y evaluar varios modelos supervisados de clasificación emocional en un entorno in-domain basados en representaciones vectoriales y algoritmos tradicionales.

O4. Construir un corpus out-of-domain sintético compuesto por 25 perfiles heterogéneos y 300 textos con características propias del lenguaje de redes sociales, sobre el que poner a prueba el sistema de recomendación.

O5. Validar el rendimiento del modelo seleccionado en escenarios out-of-domain para analizar su capacidad de generalización.

O6. Diseñar e integrar un sistema de recomendación emocional basado en variables de perfil, intereses, presupuesto y estado afectivo.

O7. Aplicar filtros éticos para evaluar el comportamiento del sistema en situaciones sensibles y contextos de vulnerabilidad del usuario.

O8. Explorar casos de uso representativos y patrones emocionales agregados mediante herramientas de Business Intelligence.

3.3 Beneficios del proyecto

El desarrollo de este proyecto aporta beneficios en tres dimensiones complementarias: técnica, científica y social.

En primer lugar, **a nivel técnico**, contribuye al avance del Procesamiento del Lenguaje Natural en español, mediante la creación de un corpus emocional equilibrado y validado, y la evaluación del rendimiento *out-of-domain* de modelos de clasificación emocional.

En segundo lugar, **a nivel metodológico y científico**, el trabajo demuestra la viabilidad de incorporar el análisis emocional dentro de sistemas de recomendación, integrando métricas afectivas en un contexto realista y multidimensional que combina factores de perfil, intereses y viabilidad económica.

Finalmente, **a nivel social y ético**, el sistema propone un enfoque de recomendación empático y responsable, que incorpora filtros de advertencia no intrusivos orientados a contextos sensibles (por ejemplo, personas con movilidad reducida, miembros de la comunidad LGBT+, o usuarios con detección persistente de tristeza como indicador de vulnerabilidad emocional).

En conjunto, el proyecto **TRIP(U)** representa un paso hacia la construcción de tecnologías más humanas, donde los datos no solo describen comportamientos, sino que también interpretan emociones para generar experiencias más conscientes, seguras y personalizadas.

Capítulo 4. DESARROLLO DEL PROYECTO

4.1 Planificación del proyecto

El desarrollo del presente Trabajo de Fin de Máster se ha planificado de forma escalonada entre **noviembre de 2025 y marzo de 2026**, siguiendo el calendario de seguimiento establecido por la Universidad Europea y garantizando una progresión coherente del proyecto.

La planificación se estructura en **cuatro hitos principales**, que organizan el proyecto desde la preparación de los datos hasta la integración final del sistema de recomendación TRIP(U). Cada hito agrupa un conjunto coherente de tareas alineadas con los objetivos específicos del proyecto y produce resultados verificables que sirven de base para la fase siguiente. Esta estructura permite un control progresivo del avance del trabajo y facilita la validación incremental de las decisiones técnicas adoptadas.

A continuación, se describen los hitos que articulan la planificación del proyecto:

HITO 1: Fundamentación, corpus in-domain y preprocesamiento *(noviembre–diciembre 2025)*

Este primer hito está orientado a la construcción de la base metodológica y de datos del proyecto. Incluye la creación del Gold Standard emocional en español, a partir de reseñas reales, mediante un proceso de doble etiquetado ciego.

Asimismo, se desarrolla el pipeline completo de preprocesamiento lingüístico, que comprende normalización, lematización, eliminación de ruido semántico y neutralización de entidades, garantizando que los textos sean adecuados para su análisis automático. Este hito sienta las bases sobre las que se entrenarán y evaluarán los modelos posteriores.

HITO 2: Modelos de clasificación y validación out-of-domain *(diciembre 2025–enero 2026)*

El segundo hito se centra en el entrenamiento y evaluación de modelos supervisados de clasificación emocional, empleando representaciones vectoriales clásicas y algoritmos de Machine Learning.

Además de la evaluación in-domain, se construye un corpus fuera de dominio (out-of-domain) y se analiza la capacidad de generalización del modelo seleccionado, identificando la pérdida de rendimiento y las limitaciones derivadas del cambio de dominio. Este análisis permite validar la robustez del sistema en contextos lingüísticamente más exigentes.

HITO 3: Diseño e integración del sistema de recomendación *(febrero 2026)*

En este hito se aborda el diseño del sistema de recomendación emocional TRIP(U), integrando las predicciones del clasificador con variables de perfil, intereses y presupuesto. Se define el sistema de puntuación multicriterio y se establecen los mecanismos de cálculo que permiten generar un ranking personalizado de destinos para cada usuario.

Asimismo, se implementan filtros éticos y mecanismos de advertencia orientados a contextos sensibles, garantizando un comportamiento responsable del sistema. Este hito consolida la transición desde el modelo predictivo hacia la aplicación funcional del recomendador.

HITO 4: Visualización, análisis y cierre *(marzo 2026)*

El último hito corresponde al análisis agregado del comportamiento del sistema y a la generación de visualizaciones como aplicación de Business Intelligence. En esta fase se examinan patrones emocionales y resultados derivados de los casos de uso representativos.

Finalmente, se realiza la interpretación global de los resultados y la redacción de la memoria final del proyecto, consolidando conclusiones y proponiendo líneas futuras de investigación.

CRONOGRAMA DEL PROYECTO

El cronograma del proyecto se organiza en **cuatro hitos**, desglosados en **ocho tareas (T1–T8)** alineadas con los **objetivos específicos** del trabajo. Esta planificación estructura el desarrollo del proyecto desde la construcción de los datos y el preprocesamiento hasta la validación de modelos y la integración final del sistema TRIP(U). La secuencia de tareas sigue una lógica progresiva, en la que cada fase se apoya en los resultados de la anterior. La **descripción detallada de las tareas**, junto con su duración y alcance, se presenta en el **apartado 4.2**.

Tabla 1. Cronograma general del proyecto TRIP(U).

HITO	COD.	OBJ.	TAREA	Nov 25	Dic 25	Ene 26	Feb 26	Mar 26
1	T1	O1	Construcción del corpus emocional in-domain	X	X			
	T2	O2	Pipeline de preprocesamiento lingüístico exhaustivo		X			
2	T3	O3	Entrenamiento y validación in-domain de modelos		X			
	T4	O4	Construcción del corpus out-of-domain		X	X		
	T5	O5	Validación out-of-domain del modelo			X		
3	T6	O6	Diseño e integración del sistema de recomendación				X	
	T7	O7	Implementación de filtros éticos				X	
4	T8	O8	Análisis de casos de uso y visualización de resultados					X

Fuente: elaboración propia.

4.2 Descripción de la solución, metodologías y herramientas empleadas

El presente proyecto se plantea como una investigación aplicada centrada en el diseño de un sistema de recomendación de viajes personalizado basado en detección emocional mediante PLN. La metodología se organiza en tareas secuenciales que cubren la preparación de datos, el entrenamiento y validación de modelos emocionales y su integración en el sistema TRIP(U).

Metodología general de trabajo en 8 tareas secuenciales:

T1. Construcción del corpus emocional in -domain *(apartado 4.2.1)*

T2. Pipeline de preprocesamiento lingüístico *(apartado 4.2.2)*

T3. Entrenamiento y validación in-domain de modelos *(apartado 4.2.3)*

T4. Construcción del corpus out-of-domain *(apartado 4.2.4)*

T5. Validación out-of-domain del modelo *(apartado 4.2.5)*

T6. Diseño e integración del sistema de recomendación *(apartado 4.2.6)*

T7. Implementación de filtros éticos *(apartado 4.2.7)*

T8. Análisis de casos de uso y visualización de resultados *(apartado 4.2.8)*

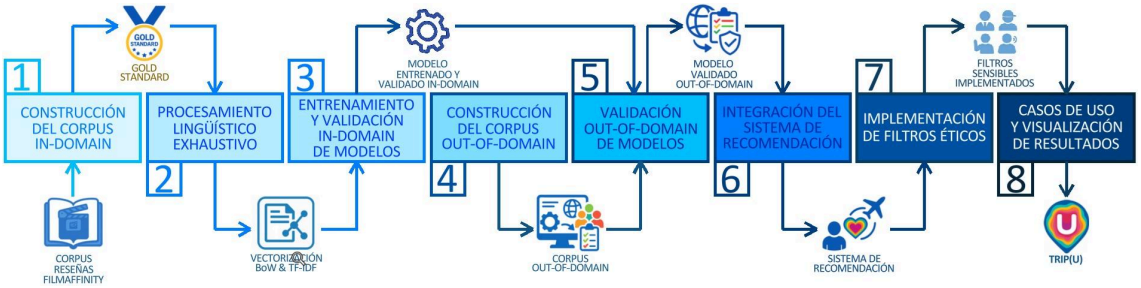

Figura 1. Secuencia de fases metodológicas del proyecto TRIP(U).

Fuente: elaboración propia.

4.2.1 Construcción del corpus emocional in-domain (nov-dic 2025)

Duración: 6 semanas / Periodo: Nov-Dic 25 / Tarea alineada con: objetivo específico O1.

El propósito de esta fase fue construir un Gold Standard en español, equilibrado y fiable, para el entrenamiento y evaluación de modelos de clasificación emocional. El corpus final consta de 500 reseñas (100 por emoción), para mantener un estado balanceado entre clases, y con etiquetado dual y criterios de consenso explícitos.

1A. Selección del dataset de partida y justificación

Se partió de un conjunto de reseñas en español de **FilmAffinity** (Kaggle) [6], por tres motivos principales: (i) disponibilidad de textos **de longitud media** que contienen suficiente señal afectiva y contexto semántico (más informativos que tuits muy breves), (ii) **dominio rico en emociones** (el cine activa reacciones afectivas expresivas), y (iii) presencia de **metadatos prescindibles** (títulos de películas, puntuaciones, etc.) que pueden ser eliminados para evitar contaminación del aprendizaje. Adicionalmente, el dataset **NO incluye identificadores personales del autor**, lo que simplifica la gestión ética y evita procesos de anonimización.

1B. Transformación inicial y columnas del corpus

A partir del dataset bruto, se preparó un único texto por reseña (cuerpo), y se eliminaron columnas que no utilizará el futuro modelo (título de la película, puntuaciones agregadas, etc.), debido a que contaminarían el entrenamiento con un contexto cinematográfico no presente out-of-domain. Se generó la columna **review_id** como indexación y se añadieron las columnas:

- **etiqueta_emocional** (0–4 = emociones codificadas),
- **anotador_A** (etiquetas individuales)
- **anotador_B** (etiquetas individuales)
- **consenso** (0-1 = No/Sí),
- **lote** (1, 2 o 3, explicación a continuación).

1C. Marco emocional y codificación

Se adoptó una taxonomía de **cinco emociones** inspirada en el **modelo circunflejo** en el cual encontramos un **eje x dominante** con la **valencia** emocional (positivas, negativas o neutras) y un **eje y secundario** con la energía de **activación** (baja, media o alta).

Se decidió trasladar dichas emociones a una **codificación ordinal 0–4** para reflejar cercanías/lejanías afectivas dentro del eje dominante de las valencias.

- **0 = Tristeza** (valencia negativa -2, activación media),
- **1 = Ansiedad** (valencia negativa -1, activación alta),
- **2 = Calma** (valencia neutra 0, activación baja),
- **3 = Curiosidad** (valencia positiva 1, activación alta),
- **4 = Alegría** (valencia positiva 2, activación media).

La **Figura 2** muestra la adaptación del modelo circunflejo de Russell (1980) [1] al sistema TRIP(U), situando las cinco emociones según sus coordenadas de valencia y activación. El diagrama ilustra la continuidad entre emociones vecinas y la simetría entre los polos negativos y positivos del espacio afectivo. El eje X de las valencias emocionales sería el dominante, por lo que constituye la base principal sobre la que se cimenta la codificación anteriormente descrita.

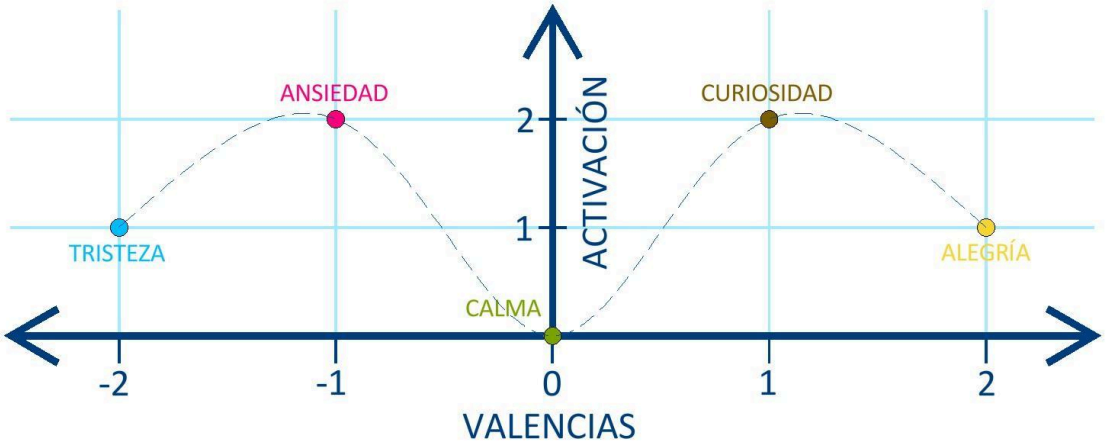

Figura 2. Adaptación del modelo circunflejo de las emociones de Russell (1980).
Fuente: elaboración propia.

1D. Criterios de muestreo de reseñas

Para maximizar la claridad afectiva y la consistencia del etiquetado:

- Longitud **máxima 150 palabras** (para evitar ruido por reseñas excesivamente largas).
- Longitud **mínima de 10 palabras** (para poder capturar cierta riqueza emocional).
- **Variedad léxica** y expresiva (evitando duplicados cercanos).
- Señales emocionales **explícitas o implícitas claras**.
- Cobertura balanceada por emoción en el total (100 reseñas por emoción).

1E. Diseño del etiquetado dual y perfil de anotadores

El etiquetado se realizó en doble ciego para controlar la subjetividad inherente emocional.

- **anotador_A:** hombre, 44 años, arquitecto español (autor del presente TFM).
- **anotador_B:** mujer, 35 años, educadora social costarricense.

Esta divergencia de perfiles fue deliberada para capturar sensibilidades distintas (formación técnica vs. psicosocial; España vs. Costa Rica), reduciendo el riesgo de sesgos homogéneos.

El autor del proyecto (anotador_A), con formación en análisis estructural y enfoque metodológico, aportó una mirada analítica y sistemática al etiquetado. La colaboradora (anotador_B), contribuyó con una perspectiva más empática y orientada a la interpretación emocional. Esta combinación buscó enriquecer la sensibilidad del corpus y reducir el sesgo.

1F. Calibración del manual de codificación

Con el fin de garantizar la consistencia inter-anotador y la fiabilidad del corpus, se desarrolló un proceso de **calibración iterativa del manual de codificación emocional**, basado en tres lotes sucesivos de 50 reseñas cada uno (150 en total). El propósito fue ajustar progresivamente los criterios de interpretación emocional y converger hacia un sistema de etiquetado reproducible.

1G. Estructura por lotes (3 × 50) y control de sesgo de frecuencia

Se organizaron **tres lotes de 50** reseñas para calibración progresiva:

- **Batch 1 (50)**: **perfectamente balanceado** (10 por emoción). Objetivo: fijar un punto de partida y comprobar el acuerdo básico.
- **Batch 2 (50)**: **ligeramente asimétrico** (p. ej., distribuciones 9–9–11–11–10). Objetivo: evitar el sesgo de frecuencia (el anotador "espera" igual reparto y corrija por conteo).
- **Batch 3 (50)**: **asimetría rotada** (las emociones con 9 pasan a 11 y viceversa). Objetivo: test de consistencia frente a cambios leves de distribución.

Tras cada lote, se calcularon las métricas de acuerdo y se revisaron los casos de desacuerdo para refinar las definiciones y ejemplos incluidos en el manual. Las observaciones derivadas de cada iteración se integraron en nuevas versiones del documento: **v1**, **v2** y **v3**.

La versión final (**v3**) (Anexo 9.1) representa el consenso definitivo entre ambos anotadores.

1H. Métricas de calidad inter-anotador y umbrales

Para cuantificar el grado de concordancia entre anotadores, se aplicaron 3 métricas complementarias:

1. **Kappa ponderado cuadrático (κ)** [4]: Evalúa el nivel de acuerdo teniendo en cuenta la *distancia ordinal* entre etiquetas. Penaliza más los desacuerdos amplios (por ejemplo, confundir "tristeza" con "alegría") que los cercanos (p. ej., "tristeza" con "ansiedad"). Se eligió esta variante por permitir una interpretación más sensible del error.
    - Umbral establecido: **$\kappa \geq 0.90$** (acuerdo excelente).

2. **Porcentaje de acuerdos exactos**: Mide la proporción de coincidencias idénticas sobre el total de reseñas. Se adoptó como métrica de control más intuitiva y fácilmente interpretable, complementaria al Kappa.
   Umbral establecido: **≥ 80% de coincidencias exactas**.

3. **Distribución de desacuerdos por magnitud**: Clasifica las diferencias entre anotadores según la *distancia en saltos ordinales* (1, 2, 3 o 4 puntos).
   Umbrales establecidos:
   - No se permiten desacuerdos de **4 puntos**.
   - Máximo **1 desacuerdo de 3 puntos** (por lote).
   - Máximo **3 desacuerdos** con diferencia **≥ 2 puntos** en total (por lote).

Estas métricas conjuntas ofrecieron una visión holística del proceso de calibración: el Kappa ponderado evaluó la solidez estadística del acuerdo, el porcentaje de coincidencias reflejó la claridad práctica de las etiquetas, y la distribución de errores permitió identificar las categorías más propensas a confusión.

1I. Resultados obtenidos

Los resultados obtenidos a lo largo de las tres iteraciones mostraron una mejora progresiva en todas las métricas, alcanzando valores que superaron los umbrales definidos:

Tabla 2. Resultado de métricas de calibración por lote

Iteración	κ ponderado cuadrático	% de acuerdos exactos	Desacuerdos ≥2 puntos
Lote 1	0.864	78%	4 casos
Lote 2	0.897	84%	3 casos
Lote 3	0.959	90%	1 caso

Fuente: elaboración propia.

Esta evolución confirma la eficacia del proceso iterativo y la maduración del manual de codificación. En la tercera y última iteración se alcanzó una **estabilidad inter-anotador óptima**, validando la fiabilidad del sistema de etiquetado y permitiendo extender su uso al resto del corpus (350 reseñas adicionales) con una sola persona (anotador_A).

Los resultados de las métricas anteriormente mencionadas y las matrices de confusión, pueden observarse graficadas y analizadas en detalle en el Anexo 2.

1J. Criterio de consenso y consolidación de etiquetas

- Si **anotador_A == anotador_B** → **consenso = Sí** y **etiqueta_emocional** = esa etiqueta.
- Si **|anotador_A − anotador_B| = 1** → se revisó con el **manual**; si persistió, se adoptó la **etiqueta intermedia** cuando conceptualmente tenía sentido.
- Si **|anotador_A − anotador_B| ≥ 2** → **revisión obligatoria** y discusión guiada por manual; si no hubo acuerdo, prevaleció una **tercera lectura mediada**.

Las **150 reseñas** de calibración quedaron con **consenso explícito** (y trazabilidad por lote).

1K. Consolidación del corpus final

Una vez completado el proceso de calibración y validado el manual de codificación (versión v3), se procedió a la **etiquetación definitiva del corpus completo**, compuesto por 500 reseñas.

Las 150 reseñas utilizadas en las iteraciones de calibración se conservaron como parte del conjunto final, dado que sus etiquetas ya cumplían con los criterios de fiabilidad establecidos. Las 350 reseñas restantes fueron etiquetadas exclusivamente por el anotador_A siguiendo el protocolo definitivo, replicando las decisiones y ejemplos del manual calibrado.

El corpus completo se **balanceó** a **100 reseñas por emoción**, verificando:

- Longitud media y varianza **similares** por clase (para evitar sesgos triviales de longitud).
- Distribución temporal y temática **heterogénea**.
- **No inclusión** de términos cinematográficos en **labeling cues**.

1L. Gold Standard

Se obtuvo un Gold Standard emocional en español, equilibrado (100×5), compuesto por 500 reseñas etiquetadas en 5 emociones. De ellas, 150 fueron consensuadas en doble anotación y las 350 restantes de forma unitaria tras la estabilización del manual de codificación.

El resultado final constituye un conjunto **homogéneo, verificable y reproducible**, que combina diversidad temática con consistencia emocional. La calidad inter-anotador, las matrices de confusión y la coherencia ordinal fueron validadas en anexos técnicos, asegurando que las categorías reflejen transiciones emocionales graduales más que rupturas discretas. Este corpus servirá como **punto de partida confiable** para las fases posteriores de **preprocesamiento, modelado supervisado y validación out-of-domain**, donde se evaluará su capacidad de generalización hacia contextos comunicativos no cinematográficos.

Para asegurar la trazabilidad y la transparencia del proceso, se mantuvieron las siguientes columnas finales en el archivo maestro.

Tabla 3. Columnas del archivo Gold Standard

Columna	Descripción
review_id	Identificador único de la reseña
texto	Cuerpo de la reseña preprocesada
etiqueta_emocion	Valor numérico (0–4) asignado según el modelo circunflejo adaptado
anotador_A	Etiqueta emocional correspondiente al anotador_A
anotador_B	Etiqueta emocional correspondiente al anotador_B
consenso	Variable binaria (1 = consenso; 0 = revisado durante la calibración)
lote	Indica el lote de calibración al que pertenece la reseña (si aplica)

Fuente: elaboración propia.

4.2.2 Pipeline de preprocesamiento lingüístico

Duración: 4 semanas / Periodo: Dic 25 / Tarea alineada con: objetivo específico O2.

Sobre la obtención del corpus gold standard, se constituye la base para la siguiente fase: el preprocesamiento textual. Dado que las reseñas proceden de un dominio cinematográfico, pero el objetivo es generalizar hacia textos más espontáneos (redes sociales), esta etapa cumple un papel esencial de **limpieza, estandarización y control del sesgo de dominio** [7].

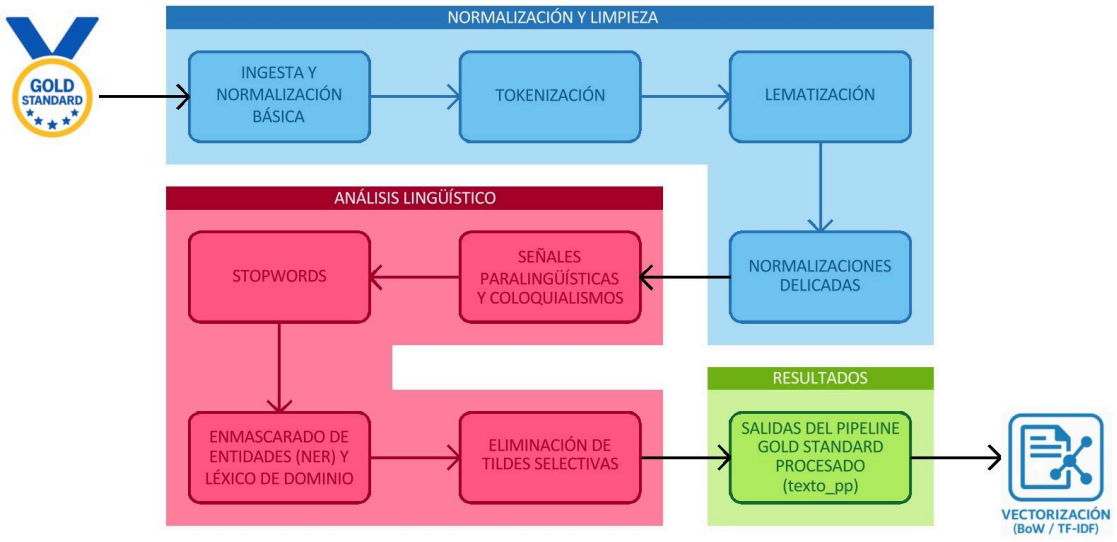

Figura 3. Pipeline de preprocesamiento del corpus emocional TRIP(U).

Fuente: elaboración propia.

En esta fase, se implementa un **pipeline exhaustivo de normalización**, con dos objetivos principales: (i) preparar la señal lingüística para que los modelos base (TF-IDF + SVM / Naive Bayes) operen sobre un vocabulario estable, y (ii) neutralizar dependencias de dominio (nombres propios, títulos o léxico de crítica) que dificultarían la transferencia *out-of-domain*.

La **Figura 3** resume de forma visual las fases que integran el pipeline de preprocesamiento:

2A. Ingesta, saneamiento y normalización básica

- Conversión de todo el corpus a **UTF-8**, normalización **Unicode NFC** (tildes y diéresis compuestas), y estandarización de comillas, guiones y espacios no-separables.
- Detección de anomalías por caracteres no ASCII (emojis, símbolos inusuales).
- Sustitución de URLs por el token **[URL]**; emails por **[EMAIL]**; menciones por **[USER]**; números largos (≥5 dígitos) por **[NUM]**; fechas por **[FECHA]**
- Espacios múltiples y puntuaciones duplicadas simplificadas a simples.

2B. Tokenización, lematización y tratamiento morfológico

- Tokenización con **spaCy** (pipeline es_core_news_lg) [8].
- No se segmentan contracciones (del/al), dejando que el lematizador las procese.
- Lematización a **forma canónica** (verbos, sustantivos, adjetivos), sin stemming. En español, la lematización evita perder matices (“tristes”, “tristeza” → “triste”).
- En verbos con clíticos (p. ej., “decírselo”), spaCy segmenta morfológicamente
- Guardamos POS tags para análisis descriptivo y posibles **features paralingüísticas.**

2C. Criterios de caja negra/blanca para normalizaciones delicadas

- **No se mantienen tildes** salvo en un conjunto reducido de excepciones (*té, más, sí, sé, tú, él, aún*) que podrían generar **homógrafos ambiguos** si se desacentúan.
- Tras el enmascarado NER (Reconocimiento de Entidades Nombradas), se realiza la minusculización, para **reducir el tamaño del vocabulario** y mejorar su robustez.
- No se corrigen automáticamente faltas (riesgo de introducir sesgos).

2D. Señales paralingüísticas y coloquialismos

- Normalización de elongaciones (“bueeeno” → “bueno”).
- Emojis → mapeo a tokens neutrales **[EMO_POS]**, **[EMO_NEG]**, **[EMO_SURP].**
- Normalización de **risas** en sus múltiples variantes (jajaja, jejee, xd, lol, etc.) → **[RISAS]**

2E. Stopwords y léxico "de dominio"

- Se eliminan artículos, preposiciones, pronombres átonos y conjunciones de poco valor informativo (lista base NLTK/spaCy).
- **Neutralización por enmascarado** para términos claramente de dominio (p. ej., "guión", "fotografía", "banda sonora", "taquilla", "metraje", "rodaje", "director", "actriz", "cameo", "cámara", "oscar", "goya", "cartelera"…), que mapeamos a **[DOM_CINE]**.
- **Enmascarado por categorías semánticas específicas**: Los tokens se agrupan y sustituyen según su función: tecnicismos y roles del cine (**[DOM_CINE]**), títulos de películas (**[ENT_OBRA]**), instituciones (**[ENT_ORG]**) y nombres propios del sector (**[ENT_PER]**). Esto permite preservar la estructura reduciendo el sesgo de dominio.
- Ventaja: evitamos que el clasificador aprenda "atajos" del dominio, pero **no borramos estructura** (el token contribuye a TF-IDF como *un solo término*).
- Conjunto completo documentado en Anexo 9.3: Lexicón de dominio cinematográfico.

2F. Enmascarado de entidades (NER)

- Mapeo: PERSON → [ENT_PER], ORG → [ENT_ORG], LOC/GPE → [ENT_LOC], WORK_OF_ART → [ENT_OBRA].
- Este paso es clave para **out-of-domain**: evitamos que "Almodóvar", "Bardem" o "El laberinto del fauno" sean *features* predictivas.

2G. Salidas del pipeline

- **Entrada**: review_ID, texto.
- **Salida preprocesada**: texto_pp (texto listo para vectorización), y **metadatos**: n_RISAS, n_EMO_POS/NEG/SURP, n_ENT_PER/ORG/LOC/OBRA, n_DOM_CINE.

	review_id	texto	texto_annot	texto_pp	etiqueta_emocion
420	421	El ambiente de las marismas, con su luz lúgubre, es un reflejo de la pena social. La película no te ofrece ni un atisbo de esperanza en la justicia. El pesimismo es el aire que se respira. "La isl...	El ambiente de las marismas, con su luz lúgubre, es un reflejo de la pena social. La [DOM_CINE] no te ofrece ni un atisbo de esperanza en la justicia. El pesimismo es el aire que se respira. "[ENT...	ambiente marisma luz lugubre reflejo pena social ofrecer atisbo esperanza justicia pesimismo aire respirar cronica abatimiento moral rural	0
398	399	En el minuto en que el agua entra, no hay vuelta atrás. Recuerdo agarrar el brazo de mi pareja sin darme cuenta. No era miedo, era una ansiedad absoluta como consecuencia del desastre. Cuando la p...	En el minuto en que el agua entra, no hay vuelta atrás. [ENT_ORG] agarrar el brazo de mi pareja sin darme cuenta. No era miedo, era una ansiedad absoluta como consecuencia del desastre. Cuando la ...	minuto agua entrar vuelta atra agarrar brazo pareja miedo ansiedad absoluto consecuencia desastre pantalla lleno barro descubri respirar rapido salir	1
237	238	Vi la película después de una jornada muy larga, y fue como sumergirme en agua templada. No por la trama, sino por la cadencia. Ninguna escena se impone, todo tiene una armonía suave, casi terapéu...	Vi la [DOM_CINE] después de una jornada muy larga, y fue como sumergirme en agua templada. No por la trama, sino por la cadencia. Ninguna [DOM_CINE] se impone, todo tiene una armonía suave, casi t...	jornada agua templado trama cadencia imponer armonia suave terapeutico	2

Figura 4. Muestra de salida con texto enmascarado y texto procesado

Fuente: elaboración propia.

2H. Consideraciones específicas para TF-IDF y evaluación

- **TF-IDF de palabras** con ngram_range=(1,2), min_df conservador para evitar ruido.
- Comparar top-20 términos por *chi²* **antes/después** del enmascarado: una lista "saludable" debe mostrar preferencia por **rasgos afectivos**.
- Para el análisis de vocabulario y visualizaciones detalladas, acudir al anexo 9.4.

2I. Ejemplos "antes/después"

- ***Ej. 1*:** "¡Bardem está increíble en *Mar adentro*! La fotografía es preciosa" → "[ENT_PER] estar increíble en [ENT_OBRA] ! [DOM_CINE] ser precioso y dejar muy feliz"
- ***Ej. 2*:** "Lloré, pero no por la película, sino por mí. 😢" → "llorar , pero no por [ENT_OBRA] , sino por yo . [EMO_NEG]"

4.2.3 Entrenamiento y validación in-domain de modelos

Duración: 2 semanas / Periodo: Dic 25 / Tarea alineada con: objetivo específico O3.

3A. Construcción preliminar de las representaciones vectoriales

Una vez completado el preprocesamiento, se generaron dos representaciones numéricas del corpus para alimentar los modelos supervisados:

- **Bag-of-Words (BoW)**: matriz dispersa documento-término (500×1701) con unigramas y bigramas filtrados con min_df=2. La inspección de sparsity (≈99%) confirmó la típica distribución altamente dispersa de textos cortos [9].
- **TF-IDF**: matriz equivalente en tamaño, pero ponderando términos según su relevancia global. La distribución es más equilibrada, útil en tareas de clasificación emocional [10].

Ambas exploraciones validaron que, pese al filtrado semántico aplicado, el corpus conserva suficiente diversidad léxica para un entrenamiento estable

3B. Estrategia de modelado supervisado

Para definir un modelo sólido, se evaluaron dos enfoques clásicos y complementarios:

- **Multinomial Naive Bayes (NB):** Modelo probabilístico muy eficiente con datos discretos. Tiende a funcionar bien con BoW/TF-IDF en textos cortos y multiclase [11].
- **Support Vector Machines (SVM, kernel lineal):** Modelo discriminativo robusto ante la alta dimensionalidad y la sparsity. Especialmente consistente con TF-IDF [12].

La comparación NB–SVM permite analizar enfoques estructuralmente distintos sobre un mismo espacio vectorial.

3C. Procedimiento experimental

Para cada combinación vectorizador-clasificador se aplicó el mismo marco experimental:

- **Train/test estratificado (80/20)** para mantener equilibrio entre emociones.
- **Pipelines sklearn** para evitar leakage (el vocabulario sólo se ajusta en *train*).
- **Búsqueda de hiper-parámetros**: max_features ∈ {1000, 1500, 2000, 2500, 3000}.
- **N-gramas (1,2)** para capturar patrones emocionales clave incluyendo bigramas.

3D. Evaluación comparada de los cuatro modelos

Se entrenaron y analizaron cuatro variantes:

BoW + NB // TF-IDF + NB // BoW + SVM // TF-IDF + SVM

Para cada una se midieron accuracy, macro-precision, macro-recall y macro-F1, además del rendimiento por clase. La siguiente figura resume las métricas macro obtenidas:

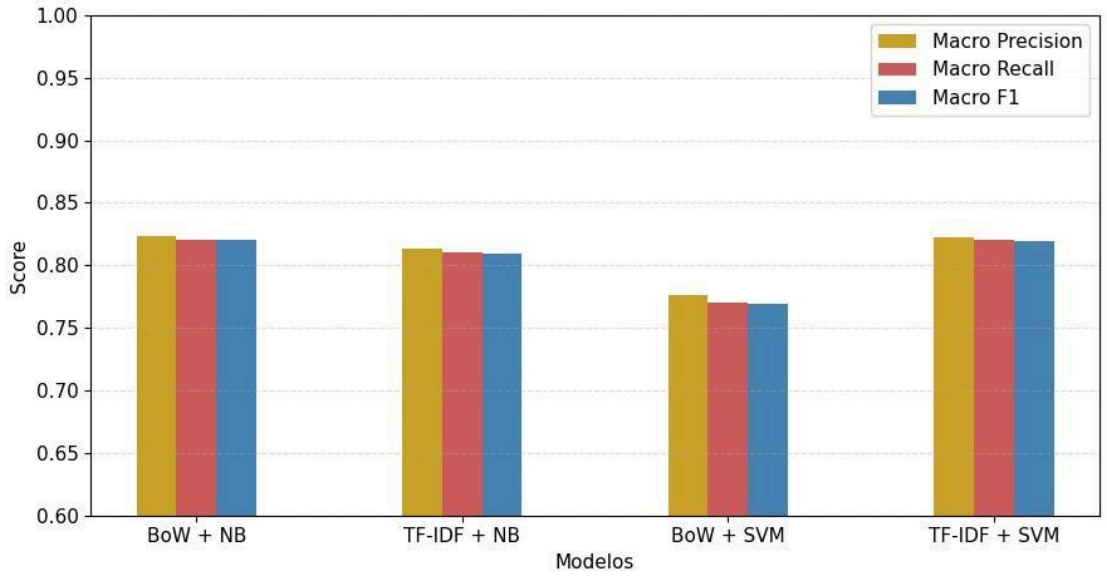

Figura 5. Comparación de métricas globales por modelo.

Fuente: elaboración propia.

Para consultar en detalle todas las métricas individuales, tablas comparativas y visualizaciones gráficas (incluyendo curvas ROC, matrices de confusión y análisis por clase), puede acudirse al Anexo 9.5., donde se recoge la evaluación completa de cada modelo.

3E. Resultados obtenidos (in-domain)

Los 4 modelos mostraron un rendimiento estable, con **macro-F1 0.77-0.82**, coherente con:

- corpus homogéneo (reseñas cinematográficas),
- clases equilibradas
- preprocesamiento exhaustivo
- ausencia de ruido semántico.

En términos comparativos:

- **BoW + NB** sorprendió por su solidez, beneficiándose del corpus limpio y del carácter discreto del vectorizado.
- **TF-IDF + NB** mantuvo un rendimiento casi idéntico, confirmando la afinidad natural entre TF-IDF y NB.
- **Ambas variantes de SVM** fueron ligeramente menos estables, aunque competitivas y consistentes, especialmente TF-IDF + SVM en *recall* para clases bien definidas.

3F. Consideraciones sobre el rendimiento in-domain vs out-of-domain

Estos resultados reflejan un escenario in-domain, donde los textos comparten estilo, longitud y temática. Es esperable que en contextos out-of-domain (textos ruidosos de redes sociales) el rendimiento sea menor. Precisamente por ello, contar con métricas altas in-domain (>0.80) ofrece un margen adecuado para evaluar la degradación posterior.

3G. Selección del modelo

Tras comparar las cuatro combinaciones, el modelo seleccionado es:

Multinomial Naive Bayes con vectorización Bag-of-Words (BoW + NB)

La elección se fundamenta en:

- mejor equilibrio global entre precisión, recall y F1;
- mayor estabilidad en todas las clases emocionales;
- menor sensibilidad a n-gramas ruidosos;
- robustez probada en escenarios multiclase con textos cortos.

Este modelo servirá como referencia para evaluar la degradación out-of-domain y valorar, en su caso, la incorporación de arquitecturas más avanzadas en futuras líneas de investigación.

4.2.4 Construcción del corpus out-of-domain

Duración: 4 semanas / Periodo: Dic 25-Ene 26 / Tarea alineada con: objetivo específico O4.

4A. Justificación de un corpus OOD sintético

El modelo emocional fue entrenado originalmente utilizando un corpus in-domain (ID), compuesto por reseñas cinematográficas de la plataforma FilmAffinity. Sin embargo, el sistema de recomendación planteado en este trabajo tiene como objetivo procesar textos reales de usuarios en sus redes sociales, lo cual implica un **cambio de dominio** importante.

Dado que las redes sociales presentan textos con características lingüísticas distintas (jerga, ruido textual, informalidad), es necesario evaluar la **capacidad de generalización del modelo** más allá del dominio cinematográfico. Este proceso de validación requiere de un **corpus OOD**, es decir, textos que simulen este nuevo entorno de uso.

El uso de un corpus OOD sintético se justifica por las siguientes razones:

1) Los datasets reales de redes sociales suelen ser anónimos y no permiten relacionar los textos con perfiles de usuarios completos.

2) El sistema de recomendación diseñado en este trabajo necesita contar con textos vinculados a variables sociodemográficas y contextuales del usuario (edad, intereses, presupuesto, etc.).

3) Un corpus sintético permite simular un entorno controlado pero realista, en el que se puede definir y analizar la diversidad en el comportamiento emocional de cada usuario de forma bastante precisa.

4B. Diseño y normalización de la base de datos OOD

Para garantizar la riqueza contextual del entorno simulado, se diseñó una base de datos relacional normalizada que permite cruzar perfiles de usuario, emociones, intereses y destinos turísticos recomendables.

La base de datos se compone de las siguientes tablas principales:

- **usuarios_info**: contiene información sociodemográfica de cada usuario ficticio (edad, género, ciudad de origen, presupuesto, perfil viajero, etc.).

- **usuarios_textos**: recoge los textos mensuales generados por cada usuario (uno por mes), con sus respectivas etiquetas de emoción predicha, emoción verdadera y score de confianza.

- **usuarios_emo_check**: tabla que almacena los check-ins emocionales (emociones autopercibidas por el usuario) para compararlas con las inferidas.

- **poi_info**: base de datos de puntos de interés turístico con información geográfica, accesibilidad, costo de vida, y nivel de inclusión LGBT.

- Tablas auxiliares como **poi_perfiles**, **poi_emo**, **poi_intereses** y **poi_costo_transporte**, que permiten personalizar las recomendaciones según distintos perfiles.

La normalización alcanza la **Tercera Forma Normal (3FN)**, utilizando claves primarias simples (usuario_id, poi_id, etc.) y foráneas que permiten mantener la integridad referencial.

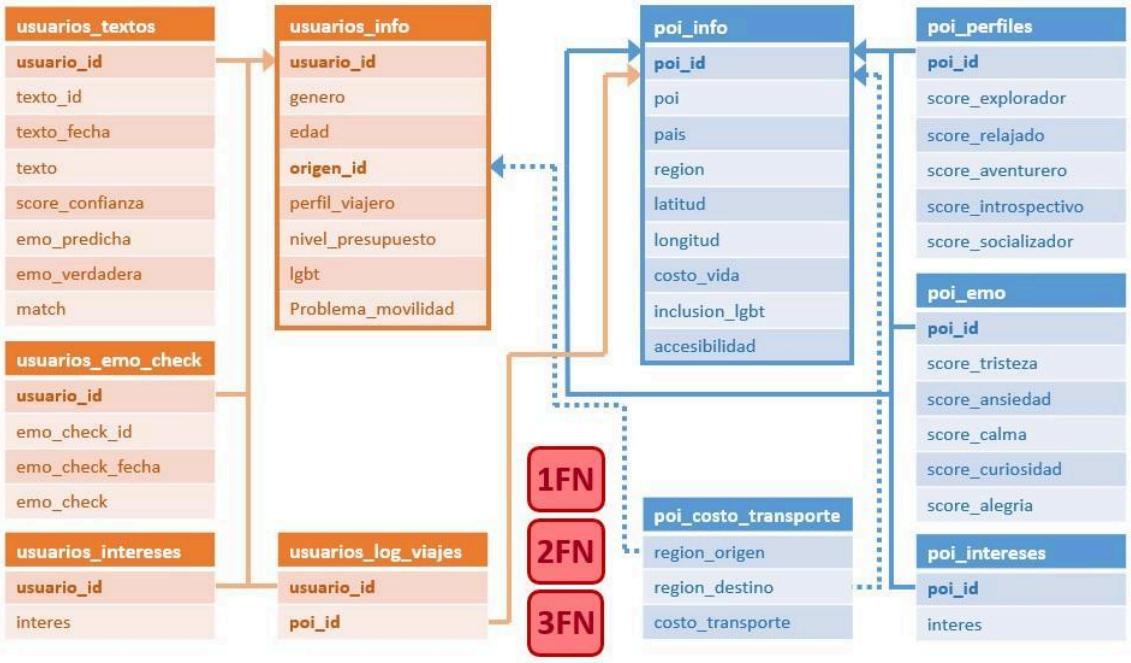

Figura 6. Diagrama de normalización de la base de datos OOD

Fuente: elaboración propia.

4C. Generación de los textos OOD

Se generaron un total de **300 textos sintéticos** (12 por cada uno de los 25 usuarios), distribuidos de forma que todas las emociones estén **balanceadas** (60 textos por emoción).

Cada usuario tiene un texto por mes (de enero a diciembre), lo que permite construir una **trayectoria emocional a lo largo del año**. Esta trayectoria puede luego analizarse combinando las emociones inferidas automáticamente y las emociones autopercibidas del usuario.

Los textos fueron **elaborados manualmente**, teniendo en cuenta:

- Jerga local, edad y contexto del usuario
- El ruido propio de redes sociales: faltas ortográficas, abreviaciones, emoticonos, y uso informal del lenguaje
- Variabilidad temática, para simular usos realistas (comentarios sobre el día a día, noticias, relaciones sociales, etc.)

Este corpus sintético OOD será utilizado tanto para **evaluar el rendimiento del modelo emocional** entrenado in-domain (fase de validación cruzada de métricas), como para **alimentar el sistema de recomendación** que se construirá sobre la base de estos perfiles.

4.2.5 Validación out-of-domain del modelo

Duración: 2 semanas / Periodo: Ene 26 / Tarea alineada con: objetivo específico O5.

En este apartado se evalúa el rendimiento del **modelo seleccionado** (basado en *Bag of Words* y un clasificador **Naive Bayes Multinomial**) en un escenario *out-of-domain*. El objetivo es analizar su capacidad de generalización cuando se aplica a textos procedentes de un dominio distinto al utilizado durante el entrenamiento.

El modelo, entrenado exclusivamente sobre el corpus *in-domain*, se aplica al corpus *out-of-domain* **sin reentrenamiento**, utilizando el mismo pipeline de preprocesamiento y representación textual. Este enfoque permite medir de forma directa el impacto del cambio de dominio (*domain shift*) sobre el rendimiento del clasificador.

5A. Evaluación del rendimiento por emoción en el corpus *out-of-domain*

La siguiente figura muestra la **tasa de aciertos por emoción** obtenida sobre el corpus *out-of-domain*, junto con el **score medio de confianza** de las predicciones. Se incluye el **baseline aleatorio teórico** (20% de aciertos en un problema de clasificación con cinco clases).

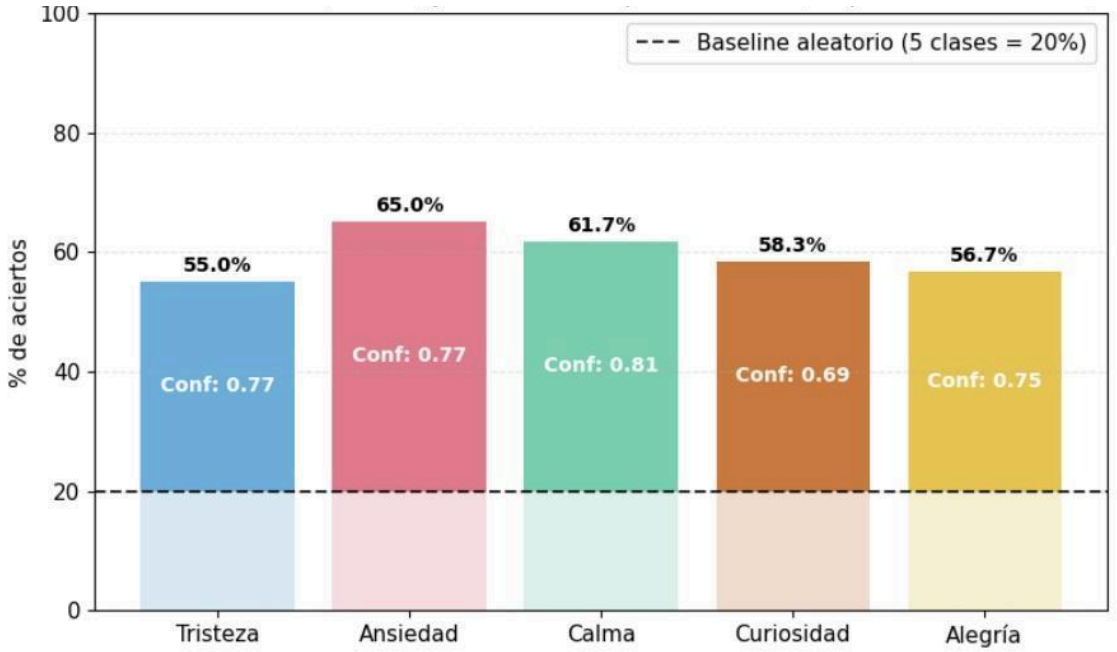

Figura 7. Tasa de aciertos (Recall) y score medio por emoción (corpus OOD)

Fuente: elaboración propia.

La figura muestra una **disminución del rendimiento del modelo al evaluarse en un entorno *out-of-domain***, un comportamiento coherente con el mayor ruido y la variabilidad lingüística propios de los textos de redes sociales. El cambio de dominio introduce un lenguaje más informal y una mayor diversidad estilística, lo que incrementa la dificultad de clasificación respecto al escenario *in-domain*, donde los textos presentan una estructura más homogénea.

A pesar de esta degradación, el modelo mantiene **tasas de acierto entre el 55% y el 65% en todas las emociones**, superando ampliamente el baseline aleatorio del 20% correspondiente a un problema multiclase con cinco categorías. Estos valores indican que el modelo conserva una **capacidad relevante de discriminación emocional**, incluso cuando se enfrenta a un dominio más complejo y heterogéneo, lo que resulta especialmente relevante en el contexto de textos generados por usuarios.

En conjunto, los resultados sugieren que, aunque el *domain shift* tiene un impacto claro en el rendimiento, el modelo sigue capturando patrones emocionales significativos de forma consistente. Este comportamiento justifica la comparación directa entre los escenarios *in-domain* y *out-of-domain*, que se aborda de manera más detallada en el apartado siguiente.

5B. Comparación *in-domain* vs *out-of-domain*

Para analizar de forma más precisa el efecto del cambio de dominio, la siguiente figura presenta una **comparación directa del rendimiento por emoción entre los escenarios *in-domain* y *out-of-domain***, utilizando el *recall* como métrica principal.

Se observa una **reducción consistente del rendimiento en todas las emociones**, con caídas moderadas atribuibles al *domain shift*. A pesar de ello, los resultados *out-of-domain* se mantienen en niveles que permiten realizar **predicciones razonables y estables**, lo que valida el uso del baseline como componente funcional del sistema de recomendación propuesto.

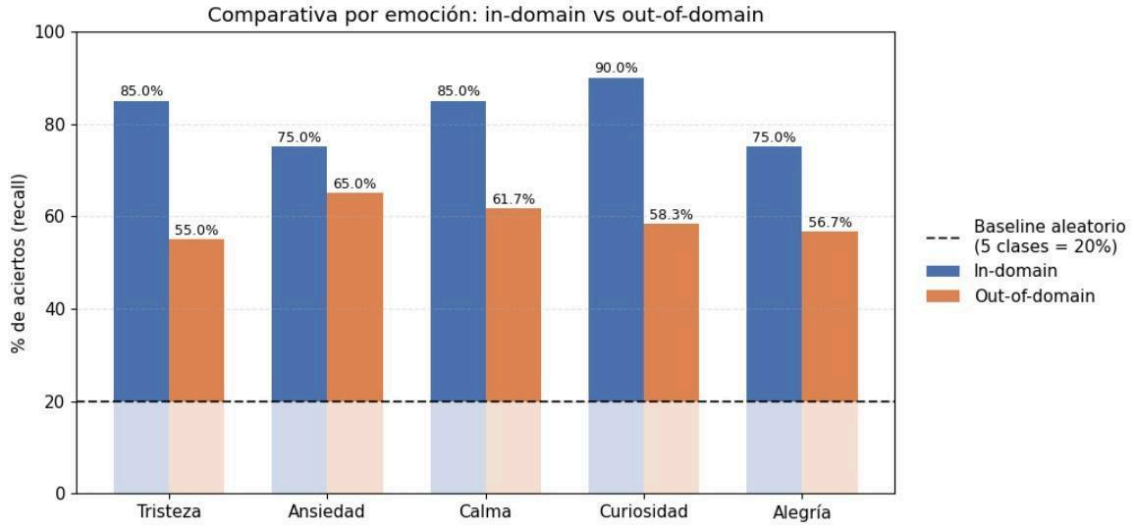

Figura 8. Comparativa por emoción: in-domain vs out-of-domain

Fuente: elaboración propia.

Se observa una **reducción consistente del rendimiento en todas las emociones**, con caídas moderadas atribuibles al *domain shift*. A pesar de ello, los resultados *out-of-domain* se mantienen en niveles que permiten realizar **predicciones razonables y estables**, lo que valida el uso del modelo como componente funcional del sistema de recomendación propuesto.

Con el objetivo de no sobrecargar el cuerpo principal del documento, en este apartado se han incluido únicamente las visualizaciones más representativas. Los resultados completos de la validación *out-of-domain*, incluyendo matrices de confusión y curvas ROC y Precision-Recall, se presentan en el Anexo 9.6.

4.2.6 Diseño e integración del sistema de recomendación

Duración: 3 semanas / Periodo: Feb 26 / Tarea alineada con: objetivo específico O6.

El sistema de recomendación desarrollado integra el análisis emocional del lenguaje en una arquitectura multicriterio orientada a la personalización contextual del viaje. A diferencia de los enfoques tradicionales basados solo en preferencias históricas o similitud entre usuarios, el modelo incorpora el estado afectivo como variable estructural del proceso de decisión, alineándose con la evolución hacia sistemas híbridos capaces de combinar múltiples fuentes de información para mejorar la calidad de las recomendaciones [13].

6A. Arquitectura general y flujo de integración

La arquitectura del sistema se estructura en cuatro fases principales:

1. Construcción del estado emocional reciente del usuario.
2. Generación de la matriz completa usuario × destino.
3. Cálculo independiente de cuatro pilares de evaluación.
4. Normalización, ponderación y generación del ranking final.

Este diseño responde a una estrategia de agregación ampliamente utilizada en sistemas de recomendación multicriterio por su equilibrio entre interpretabilidad y modularidad [14].

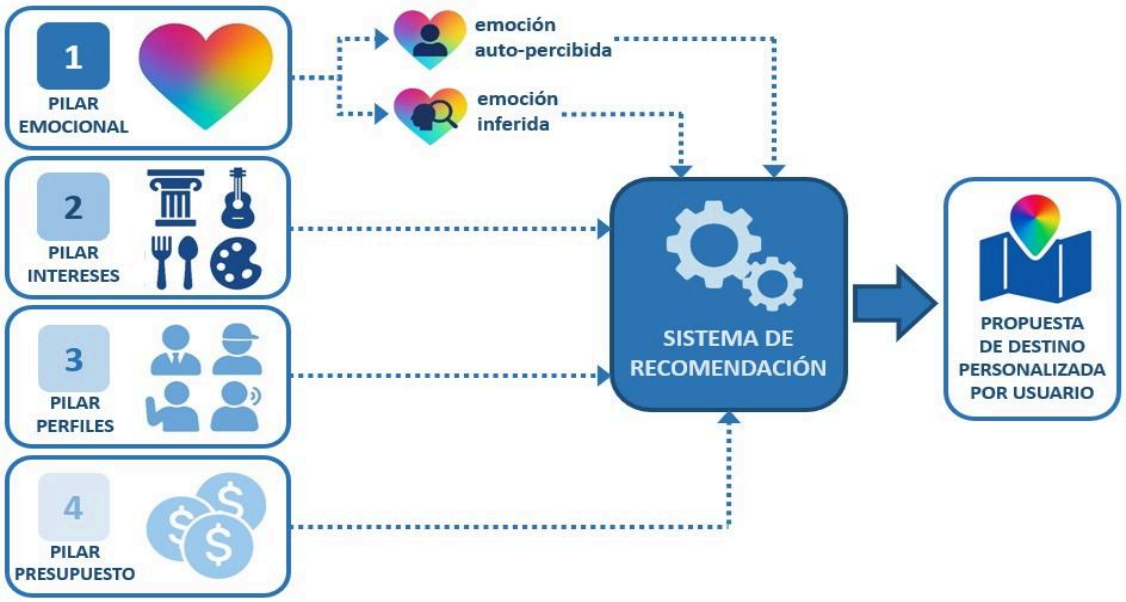

Figura 9. Diagrama de diseño del sistema de recomendación multicriterio

Fuente: elaboración propia.

6B. Construcción del snapshot emocional

El estado emocional actual del usuario se define combinando dos señales:

- Emoción autopercibida más reciente (emo-check).
- Emoción inferida mediante el clasificador PLN entrenado previamente.

Esta dualidad permite capturar tanto la dimensión declarativa como la dimensión lingüística del estado afectivo. La integración de señales heterogéneas constituye un principio central de los sistemas híbridos, donde distintas fuentes de evidencia se combinan para reducir incertidumbre y mejorar robustez del modelo [15].

El resultado es un snapshot compuesto por:

- emo_check_reciente
- emo_texto_reciente
- Fechas asociadas a cada señal

Este snapshot actúa como variable dinámica que alimenta el primer pilar del sistema.

6C. Pilar 1 — Score emocional

Cada destino (POI) dispone de una representación emocional estructurada en cinco dimensiones: tristeza, ansiedad, calma, curiosidad y alegría.

Para cada par usuario–destino se calculan:

- score_emo_check: score del destino asociado a la emoción autopercibida.
- score_emo_texto: score del destino asociado a la emoción inferida.

El score emocional final se obtiene como:

score_emocional = score_emo_check + score_emo_texto

Este diseño permite reforzar la coherencia cuando ambas señales coinciden y amortiguar el efecto cuando divergen, evitando depender de una única fuente afectiva. Además, introduce redundancia informativa controlada, lo que mejora la robustez del componente emocional frente a posibles errores de clasificación o sesgos de autoevaluación. Desde el punto de vista metodológico, esta combinación mantiene el equilibrio entre señales dinámicas y estabilidad estructural, característica relevante en modelos híbridos [13].

6D. Pilar 2 — Score de intereses

El segundo pilar evalúa la coincidencia temática entre los intereses declarados por el usuario y los atributos del destino.

Cada usuario y cada POI se modelan como conjuntos (sets) de intereses. La puntuación se asigna en función del número de intersecciones:

- 3 coincidencias ⟶ 5 puntos
- 2 coincidencias ⟶ 3 puntos
- 1 coincidencia ⟶ 1 punto
- 0 coincidencias ⟶ 0 puntos

Este componente representa una señal explícita de afinidad y actúa como mecanismo estabilizador frente a variaciones emocionales circunstanciales. La combinación de preferencias explícitas con señales contextuales dinámicas es una práctica habitual en sistemas híbridos modernos [15].

6E. Pilar 3 — Score por perfil viajero

El perfil viajero introduce una dimensión conductual relativamente estable (explorador, relajado, aventurero, introspectivo, socializador).

Cada destino dispone de cinco scores asociados a estos perfiles. El sistema selecciona automáticamente el score correspondiente al perfil del usuario.

Este pilar añade coherencia experiencial y reduce la probabilidad de recomendaciones emocionalmente compatibles pero estilísticamente inadecuadas. La integración de múltiples dimensiones del usuario dentro del proceso de recomendación se considera una evolución natural hacia sistemas más personalizados y explicables [13].

6F. Pilar 4 — Score de presupuesto

El componente económico se modela como una restricción gradual, no como exclusión binaria.

Se integran:

- Nivel de presupuesto del usuario.
- Coste de vida del destino.
- Coste de transporte estimado según región origen–destino.

Se define un coste compuesto:

costo_total = 0.40 * costo_vida + 0.60 * costo_transporte

Posteriormente se calcula la diferencia entre presupuesto y coste total, transformándola en un score por tramos (0–5).

Este enfoque permite equilibrar viabilidad económica con personalización sin convertir el presupuesto en una condición excluyente absoluta.

6G. Normalización e integración multicriterio

Dado que los 4 pilares operan en escalas distintas, se realiza una normalización al rango 0–1: Emoción / Intereses / Perfil / Presupuesto

El score total se define como combinación ponderada:

score_total = 0.45 x Emoción + 0.20 x Intereses + 0.20 x Perfil + 0.15 x Presupuesto

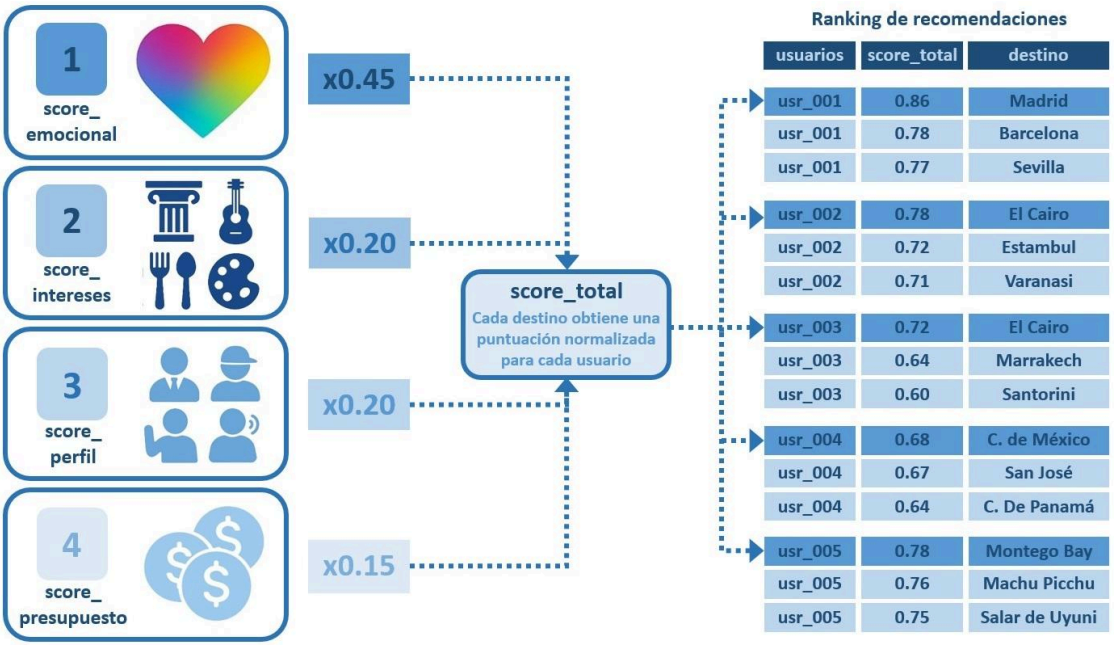

Figura 10. Desglose del sistema multicriterio y ejemplo de recomendaciones

Fuente: elaboración propia.

La mayor ponderación del componente emocional responde al objetivo central del proyecto: situar el estado afectivo como variable nuclear del proceso de recomendación. La agregación lineal ponderada constituye una estrategia ampliamente documentada por su simplicidad, transparencia y capacidad de extensión [14].

6H. Propiedades estructurales del sistema

El sistema presenta 3 características fundamentales:

- *Explicabilidad*: cada recomendación se descompone en 4 aportes independientes.
- *Modularidad*: los pesos pueden ajustarse o ampliarse con nuevos pilares.
- *Robustez contextual*: combina señales dinámicas (emoción) con señales estables.

Desde una perspectiva conceptual, el modelo puede considerarse una implementación de sistemas híbridos multicriterio con integración explícita de contexto emocional [13].

4.2.7 Implementación de filtros éticos

Duración: 1 semana / Periodo: Feb 26 / Tarea alineada con: objetivo específico O7.

El sistema de recomendación desarrollado incorpora un módulo específico de filtros éticos y advertencias sensibles cuyo objetivo no es restringir la autonomía del usuario, sino reforzar una aproximación responsable y contextualizada al proceso de recomendación. Este diseño parte de la premisa de que los sistemas basados en datos pueden generar impactos no deseados si no se consideran dimensiones sociales y contextuales del usuario [16].

Es fundamental distinguir entre:

- Filtro lógico estructural (duro)
- Filtros de advertencia sensible (no excluyentes)

Figura 11. Desglose de filtros éticos implementados

Fuente: elaboración propia.

7A. Filtro lógico estructural: exclusión del destino de origen

El único filtro duro implementado en el sistema es la exclusión automática del destino que coincide con el lugar de origen del usuario (poi_id ≠ origen_id).

Este filtro responde a un criterio de coherencia semántica y sentido común dentro del dominio turístico: recomendar como destino el propio lugar de residencia carece de utilidad funcional.

No se trata de una decisión ética en sentido estricto, sino de una regla estructural para preservar la lógica del sistema.

7B. Filtros de advertencia sensible

El núcleo de este módulo lo constituyen 3 filtros de advertencia que no bloquean ni alteran el ranking generado, sino que añaden mensajes personalizados de cuidado y reflexión.

Este enfoque evita el paternalismo algorítmico y se alinea con el principio de respeto a la autonomía del usuario, uno de los pilares de la IA responsable [17].

7B1. Alerta por tristeza persistente

Este filtro se activa cuando el sistema detecta una emoción persistente de tristeza en el usuario, derivada del análisis longitudinal de sus registros emocionales. La alerta se activa independientemente del destino recomendado y no asume diagnóstico alguno.

El objetivo es introducir un mensaje reflexivo de cuidado emocional:

💛 *"Sabemos que hay momentos en los que todo cuesta un poco más. Si últimamente has sentido que la tristeza se mantiene, quizá este sea un momento para escucharte con más atención. Viajar puede ser maravilloso, pero buscar apoyo cuando lo necesitamos también lo es. No tienes que gestionarlo todo solo."*

El diseño se basa en la necesidad de mitigar posibles impactos psicológicos negativos derivados de automatizar decisiones en contextos sensibles, tal como advierten Mittelstadt et al. [16]. El diseño se basa en la necesidad de mitigar posibles impactos psicológicos negativos derivados de automatizar decisiones en contextos sensibles, tal como advierten Mittelstadt et al

El sistema no sugiere soluciones médicas ni sustituye apoyo profesional; simplemente introduce una capa de responsabilidad contextual.

7B2. Alerta LGBT

Cuando un usuario se identifica como perteneciente al colectivo LGBT+ y el destino recomendado presenta baja inclusión en derechos o aceptación social (inclusion_lgbt == "No"), se genera una advertencia informativa.

Mensaje asociado:

🌈 *"Tu bienestar es importante. Este destino puede tener limitaciones en materia de derechos o aceptación del colectivo LGBT+. Si decides visitarlo, hacerlo con información previa puede ayudarte a disfrutar la experiencia con mayor seguridad."*

Este filtro no excluye el destino. No impone restricciones. No sustituye la decisión del usuario. Su función es ofrecer información relevante que pueda afectar a la experiencia personal.

7B3. Alerta por problemas de movilidad

Cuando un usuario declara problemas de movilidad y el destino presenta limitaciones de accesibilidad (accesibilidad == "No"), el sistema genera una advertencia preventiva:

♿ *"Queremos que tu experiencia sea cómoda y accesible. Este destino podría presentar ciertas limitaciones de accesibilidad en infraestructuras o desplazamientos. Te sugerimos revisar las condiciones para asegurarte de que el viaje se adapte a tus necesidades."*

Este enfoque se alinea con el principio de no maleficencia, los sistemas automatizados deben anticipar posibles riesgos prácticos derivados de decisiones generadas algorítmicamente [17].

7C. Evitación del paternalismo algorítmico

El diseño ético adoptado garantiza que ninguno de los filtros implementados elimina destinos ni modifica el *score_total* del sistema. No se bloquean recomendaciones ni se reordena el ranking generado por el modelo; la arquitectura decisional permanece intacta y los avisos actúan únicamente como información adicional ante situaciones potencialmente sensibles.

Su función es exclusivamente informativa y protectora. El sistema no sustituye el criterio del usuario ni condiciona su libertad de elección, sino que incorpora una capa de acompañamiento reflexivo alineada con los principios de IA confiable promovidos por la Comisión Europea (transparencia, autonomía, supervisión humana y no discriminación) [18]. El objetivo es acompañar la decisión, no tomarla por el usuario.

7D. Síntesis conceptual

El módulo ético convierte el sistema en una herramienta no solo personalizada, sino también consciente del contexto humano en el que opera. Las advertencias sensibles introducen una dimensión de responsabilidad que complementa la optimización algorítmica, situando al usuario en el centro del diseño y reconociendo la influencia de factores emocionales y personales en la toma de decisiones.

Lejos de limitar la libertad individual, el sistema ofrece información contextual cuando puede resultar relevante, actuando como acompañamiento reflexivo y evitando tanto la neutralidad acrítica como el intervencionismo excesivo. De este modo, integra principios de IA responsable en un entorno aplicado de recomendación turística, en coherencia con los marcos contemporáneos de gobernanza ética [18].

	usuario_id	poi	score_total	alerta_lgbt	alerta_movilidad	alerta_emocional	alertas
0	usr_001	Madrid	0.86000	False	False	False	
1	usr_002	El Cairo	0.78000	False	False	True	Sabemos que hay momentos en los que todo cuesta un poco más. Si últimamente has sentido que la tristeza se mantiene, quizá este sea un momento para escucharte con más atención. Viajar puede ser ...
2	usr_003	Ciudad de Panamá	0.72000	False	False	False	
3	usr_004	Marrakech	0.68000	False	False	False	
4	usr_005	Ciudad de México	0.76000	False	False	False	
5	usr_006	Salar de Uyuni	0.86000	False	False	False	

Figura 12. Ejemplo de alerta generada tras la recomendación

Fuente: elaboración propia.

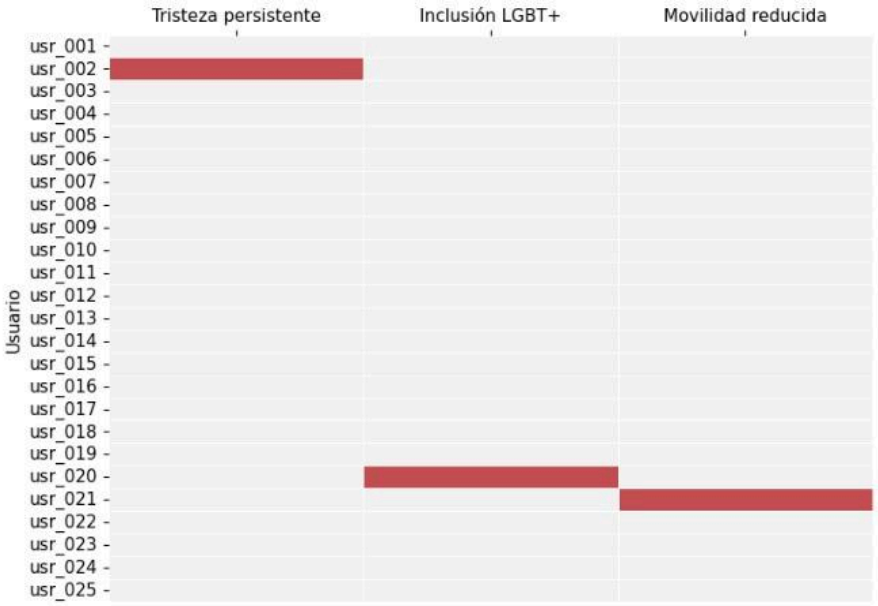

Figura 13. Activación de alertas sensibles por usuario

Fuente: elaboración propia.

4.2.8 Análisis de casos de uso y visualización de resultados

Duración: 2 semanas / Periodo: Mar 26 / Tarea alineada con: objetivo específico O8.

Una vez implementado el sistema de recomendación y los filtros éticos, se procedió a analizar el comportamiento mediante el estudio de casos de uso representativos y la construcción de visualizaciones orientadas a la interpretación de patrones emocionales agregados. Este análisis no persigue solo validar el funcionamiento técnico del sistema, sino explorar su potencial como herramienta de comprensión del comportamiento afectivo de los usuarios en el tiempo.

La estrategia adoptada se apoyó en técnicas de visual analytics, entendidas como la integración de análisis y representación visual interactiva para facilitar la interpretación de fenómenos complejos [19]. En este contexto, se seleccionaron dos visualizaciones principales por su capacidad explicativa: la evolución emocional individual y el mapa de cuadrantes CCE–CDE.

8A. Evolución emocional individual

La primera visualización representa, para varios usuarios, la evolución temporal de 2 señales emocionales: la emoción autopercibida (roja) y la inferida a partir del texto (azul). Ambas se representan como series temporales, con sus convergencias y divergencias a lo largo del año.

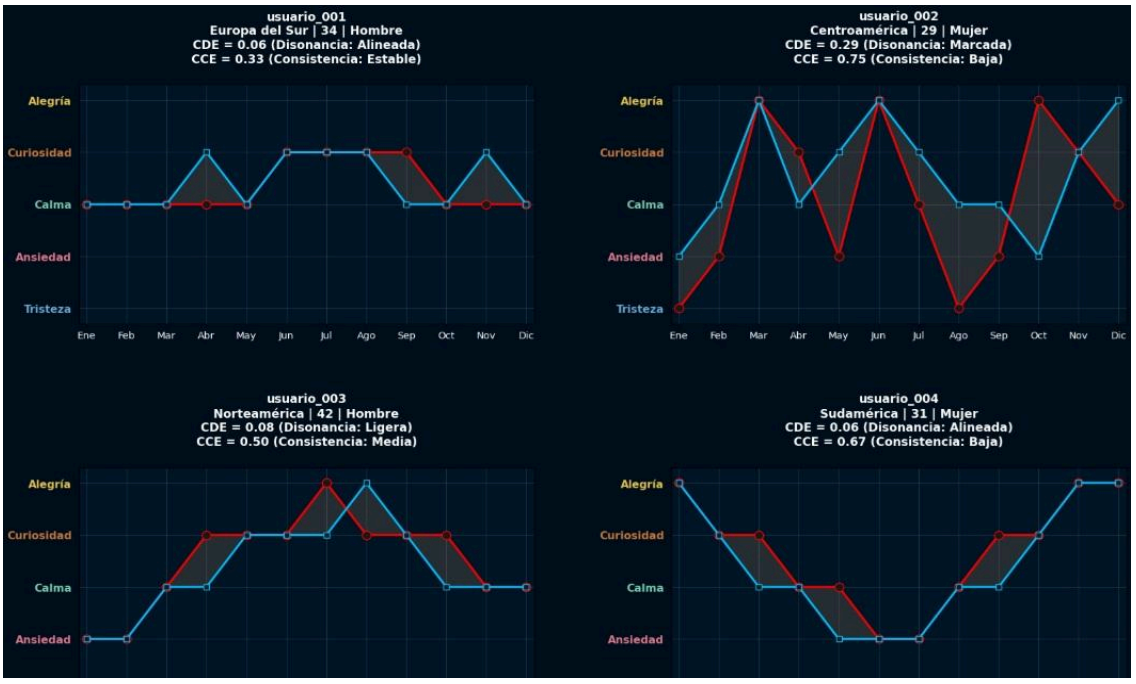

Figura 14. Gráficas de evoluciones emocionales de usuarios

Fuente: elaboración propia.

Además de la representación visual, se definieron 2 métricas cuantitativas de creación propia:

- CCE (Coeficiente de Consistencia Emocional)
- CDE (Coeficiente de Disonancia Emocional)

8B. Coeficiente de Consistencia Emocional (CCE)

El CCE se calcula como:

CCE = 1 − (desv. estándar de la emoción autopercibida / desviación estándar máxima teórica)

donde la desviación estándar de la emoción autopercibida mide cuánto varía la emoción declarada por el usuario a lo largo del tiempo, y la desviación estándar máxima teórica corresponde al rango máximo posible dentro de la escala emocional utilizada (0–4).

Esta métrica captura la estabilidad emocional del usuario: valores cercanos a 1 indican alta consistencia, mientras que valores bajos reflejan mayor fluctuación emocional.

8C. Coeficiente de Disonancia Emocional (CDE)

Por su parte, el CDE (Coeficiente de Disonancia Emocional) se define como:

CDE = (media de la diferencia absoluta entre emoción autopercibida y emoción inferida) / diferencia máxima posible

En este caso, se calcula la diferencia media absoluta entre la emoción que el usuario declara sentir y la emoción que el modelo infiere a partir de su lenguaje. La diferencia máxima posible en la escala 0–4 se utiliza para normalizar el resultado entre 0 y 1.

Valores altos de CDE indican divergencia sistemática entre lo que el usuario declara sentir y lo que su lenguaje sugiere, mientras que valores bajos reflejan coherencia entre ambas señales.

8D. Utilidad de la doble métrica

Estas métricas permiten transformar una observación visual en indicadores cuantificables, con potencial aplicación en sistemas adaptativos. Por ejemplo, un CDE elevado podría activar estrategias de recomendación más conservadoras o mayor contextualización en la interacción.

Más allá del uso en este experimento concreto, la combinación de análisis longitudinal y métricas derivadas abre la puerta a modelos dinámicos capaces de adaptar recomendaciones no solo al estado puntual, sino al patrón emocional del usuario.

8E. Mapa de perfiles emocionales (Cuadrantes CCE–CDE)

La segunda visualización seleccionada sintetiza la información anterior en un plano bidimensional donde cada usuario se posiciona según su nivel de consistencia (CCE) y disonancia (CDE).

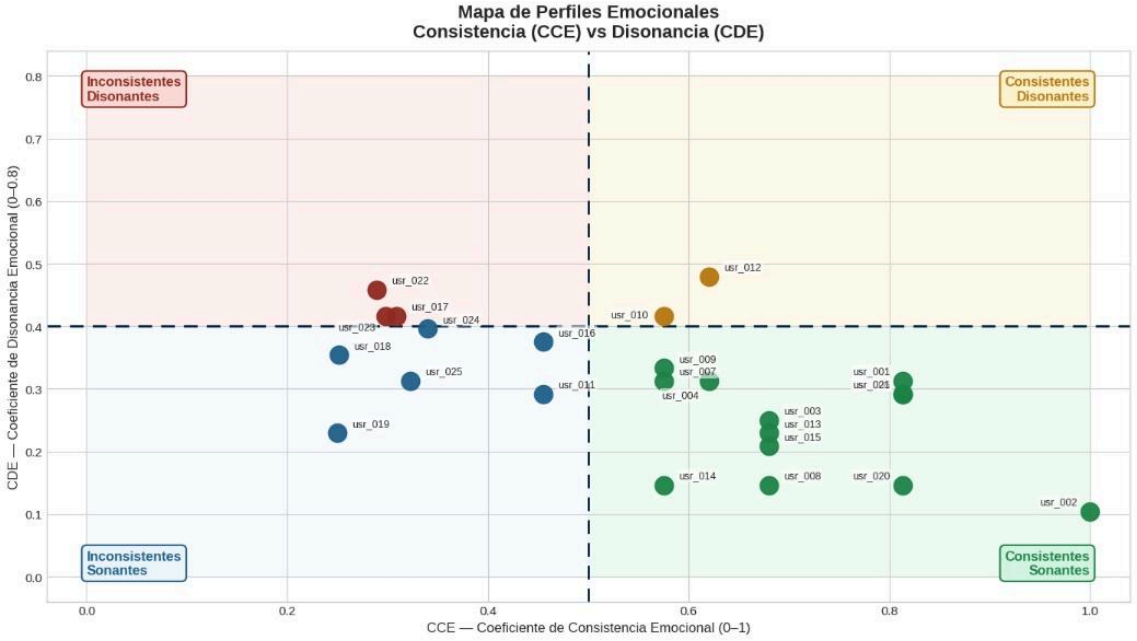

Figura 15. Mapa de perfiles emocionales (CCE vs CDE)

Fuente: elaboración propia.

El plano se divide en cuatro cuadrantes definidos por los umbrales teóricos CCE = 0.5 y CDE = 0.4, generando cuatro perfiles emocionales potenciales que permiten interpretar el comportamiento afectivo de los usuarios desde una perspectiva estructurada.

1. **Consistentes–Sonantes** (alto CCE, bajo CDE)

   Usuarios emocionalmente estables cuya emoción declarada coincide con la inferida. Constituyen el escenario ideal para el recomendador, ya que la señal afectiva es clara, coherente y fácilmente integrable en el cálculo del score emocional.

2. **Consistentes–Disonantes** (alto CCE, alto CDE)

   Usuarios estables en el tiempo, pero con divergencia sistemática entre lo declarado y lo expresado lingüísticamente. Este perfil puede reflejar diferencias entre autoimagen emocional y forma de comunicación, lo que exige mayor cautela en la interpretación del estado afectivo.

3. **Inconsistentes–Sonantes** (bajo CCE, bajo CDE)

   Aquí encontramos usuarios con elevada variabilidad emocional temporal, pero coherencia entre señal declarada e inferida en cada momento concreto. Se trata de perfiles emocionalmente cambiantes, aunque internamente alineados. Desde la perspectiva del recomendador, estos usuarios requieren una adaptación dinámica, ya que su estado afectivo puede modificarse con rapidez, pero la interpretación de cada instante resulta fiable.

4. **Inconsistentes–Disonantes** (bajo CCE, alto CDE)

   Este es el perfil más complejo: combina alta variabilidad emocional con divergencia entre señales. Desde el punto de vista del sistema, representa el mayor nivel de incertidumbre interpretativa. La doble inestabilidad (temporal y semántica) sugiere que la emoción no puede considerarse una señal robusta sin contextualización adicional, lo que abre la puerta a futuras líneas de investigación sobre detección de ambivalencia o emociones mixtas.

Esta representación permite identificar tipologías emocionales más allá del destino recomendado. El sistema deja de ser únicamente un generador de rankings para convertirse en una herramienta de segmentación afectiva con potencial estratégico.

8F. Valor analítico y proyección futura

La combinación de visualización longitudinal y segmentación en cuadrantes demuestra que el sistema no solo produce recomendaciones, sino que permite explorar patrones emocionales agregados, identificar perfiles diferenciados y evaluar coherencia entre señales afectivas.

Estas herramientas podrían integrarse en futuras versiones como mecanismos de adaptación dinámica del peso emocional en el ranking, como indicadores de riesgo contextual o incluso como base para sistemas de *matching* entre usuarios con sintonía emocional similar (o complementaria, por ejemplo perfiles ansiosos con perfiles calmados). Este último aspecto se desarrolla con más profundidad en el apartado 7.2, dentro de las futuras líneas de trabajo.

Asimismo, el enfoque visual facilita la interpretación humana del comportamiento del sistema, reforzando transparencia y explicabilidad. Para un despliegue completo de visualizaciones complementarias, se remite al Anexo 9.7.

4.3 Recursos requeridos

Para la ejecución del proyecto se han empleado los siguientes recursos:

Recursos técnicos y físicos:

- Material fungible (consumibles de oficina empleados durante el proyecto).
- Ordenador portátil con procesador Intel Core i7, 16 GB de RAM y SSD de 1 TB.
- Conexión a internet de banda ancha para ejecución de entornos colaborativos y sincronización de modelos en la nube.
- Acceso a bibliografía científica y documentación técnica especializada.
- Plataforma Google Colab Pro para la ejecución de notebooks en entorno cloud.
- Suscripción a Microsoft 365 Básico para el almacenamiento, sincronización de archivos y gestión de copias de seguridad del proyecto (Datasets y Memoria).

Recursos de software:

- **Lenguaje de programación:** Python 3.x.
- **Entornos de desarrollo:** Jupyter Notebook / Google Colab Pro
- **Librerías principales:** pandas, NumPy, scikit-learn, NLTK, spaCy, matplotlib, seaborn.
- **Herramientas de edición/documentación:** Microsoft 365 Básico (Word -> memoria, Excel -> datasets, PowerPoint -> presentación) y Adobe Acrobat (PDF final).

Recursos humanos:

- **Autor del proyecto:** Francisco Javier Cristóbal Fernández (300 h).
- **Colaboradora para etiquetado:** Diana Mora López (15 h).

Recursos de datos:

- Dataset de reseñas en español obtenido de la plataforma *FilmAffinity* (Kaggle) [6].

4.4 Presupuesto

El presente apartado recoge la estimación económica global del proyecto, considerando tanto los recursos materiales como el tiempo de dedicación invertido en su desarrollo. Aunque muchas de las herramientas empleadas no han supuesto un coste directo, se ha realizado una valoración aproximada de mercado para reflejar el valor real de los recursos utilizados. Esta estimación permite contextualizar el esfuerzo técnico y humano requerido.

Tabla 4. Estimación del presupuesto en base a los recursos necesarios

Tipo de coste	Valor (€)	Comentarios
Horas de trabajo (autor)	7.500 €	300 h × 25 €/h
Horas de trabajo (colaboradora etiquetado)	225 €	15 h × 15 €/h
Equipo utilizado (portátil i7, 16GB RAM, 1TB SSD)	1.200 €	Valor estimado de mercado si se adquiriera nuevo.
Material fungible	50 €	Consumibles de oficina empleados durante la elaboración y revisión del proyecto
Conexión a internet	50 €	Cuota prorrateada (25% de uso para el proyecto sobre 50 €/mes durante 4 meses)
Almacenamiento cloud (Microsoft 365 Básico)	10 €	4 meses × 2.5 €/mes
Google Colab Pro	40 €	4 meses × 10 €/mes
Software de desarrollo (Python y librerías open source)	100 €	Software Open Source. Importe imputado por horas de configuración y mantenimiento del entorno técnico.
Software ofimático (Suscripción mensual Microsoft 365)	28 €	4 meses × 7 €/mes
Dataset Kaggle (FilmAffinity)	0 €	Acceso libre bajo condiciones de uso
Acceso a bibliografía científica	0 €	Acceso a bases de datos y documentación técnica mediante recursos académicos institucionales
TOTAL	**9203 €**	

Fuente: elaboración propia.

El coste total estimado del proyecto asciende a **9.203 €**, incluyendo la valoración del tiempo, los recursos técnicos y el material fungible empleado. En el siguiente apartado se analiza su viabilidad económica y sostenibilidad futura.

4.5 Viabilidad

Desde el punto de vista económico, el coste estimado del proyecto (9.203 €) se corresponde principalmente con horas de desarrollo y diseño del sistema, lo que indica que se trata de una solución basada fundamentalmente en capital intelectual y no en infraestructuras costosas. Esto favorece su escalabilidad y adaptación a distintos entornos empresariales sin requerir inversiones iniciales elevadas.

En términos de sostenibilidad, el sistema se ha desarrollado utilizando herramientas open source y tecnologías ampliamente consolidadas, lo que reduce costes de mantenimiento y dependencia de licencias propietarias. Asimismo, su arquitectura modular permite incorporar mejoras futuras (como nuevos modelos de clasificación o ajustes dinámicos en las ponderaciones del recomendador) sin necesidad de rediseñar el sistema completo.

Desde una perspectiva de mercado, la integración de la dimensión emocional en sistemas de recomendación representa una oportunidad diferencial frente a soluciones tradicionales basadas exclusivamente en datos objetivos. Esto refuerza la viabilidad estratégica del proyecto en sectores como el turismo, el comercio electrónico o los asistentes digitales personalizados.

4.6 Resultados del proyecto

Conforme a los objetivos específicos planteados, el proyecto ha dado lugar a los siguientes resultados finales:

- Construcción y validación de un corpus emocional en español equilibrado y etiquetado mediante doble anotación, garantizando su fiabilidad como gold standard.
- Desarrollo de un pipeline de preprocesamiento lingüístico optimizado, incluyendo estrategias de mitigación de sesgo de dominio.
- Entrenamiento y evaluación comparativa de modelos supervisados de clasificación emocional en entorno in-domain, seleccionando como modelo final una arquitectura Bag of Words + Naive Bayes por su rendimiento y estabilidad.
- Construcción de un corpus sintético out-of-domain estructurado en tercera forma normal, permitiendo evaluar la capacidad de generalización del modelo ante lenguaje tipo redes sociales.

- Implementación de un sistema de recomendación multicriterio basado en cuatro pilares: estado emocional, intereses, perfil viajero y presupuesto.
- Integración de filtros éticos y mecanismos de advertencia en contextos sensibles, sin excluir destinos de forma automática.
- Desarrollo de herramientas de análisis visual que permiten identificar patrones de consistencia y disonancia emocional en los usuarios.

El plan de pruebas incluyó validación cuantitativa del clasificador, evaluación comparativa in-domain y out-of-domain, y verificación funcional del sistema de recomendación mediante generación de rankings personalizados y análisis de coherencia de resultados.

Durante el desarrollo se produjo una ampliación del alcance inicial, incorporando la dimensión ética y el análisis exploratorio avanzado como parte integral del sistema, enriqueciendo el resultado final respecto al planteamiento original.

Para una exposición detallada de las métricas obtenidas, así como para el análisis pormenorizado de los casos de uso y las herramientas de visualización generadas, se remite al lector al apartado 4.2.8 de esta memoria y, de forma complementaria, al desglose gráfico extendido presente en los respectivos Anexos.

Capítulo 5. DISCUSIÓN

El presente trabajo ha tenido como finalidad integrar técnicas de Procesamiento del Lenguaje Natural con un sistema de recomendación personalizado basado en el análisis emocional del usuario. Los resultados permiten afirmar que la propuesta es viable y funcional, si bien su desarrollo ha implicado decisiones metodológicas que merecen ser discutidas.

5.1 Adecuación de la metodología empleada

La metodología inicialmente planteada (curado de corpus in-domain, entrenamiento supervisado, validación out-of-domain y posterior integración en un sistema de recomendación multicriterio) ha demostrado ser coherente con los objetivos establecidos y adecuada para abordar el problema desde una perspectiva integral.

La construcción de un corpus emocional propio ha sido una decisión metodológica determinante. Aunque implicó un esfuerzo considerable en términos de etiquetado dual y validación inter-evaluador, permitió garantizar la calidad del *gold standard* y mantener el equilibrio de clases, constituyendo una base sólida para el rendimiento del clasificador.

El pipeline de preprocesamiento lingüístico también resultó fundamental. La incorporación de estrategias como el enmascaramiento de vocabulario cinematográfico (orientado a mitigar el riesgo de transferencia de dominio) redujo sesgos semánticos específicos del corpus original, convirtiéndose en un componente crítico del sistema más allá de su planteamiento inicial.

En cuanto al modelado, la comparación de distintos algoritmos y representaciones confirmó que, en este contexto, modelos tradicionales como Bag of Words + Naive Bayes pueden ofrecer un rendimiento competitivo, estable y computacionalmente eficiente frente a alternativas más complejas. Esta conclusión refuerza la importancia de la adecuación metodológica al problema concreto por encima de la sofisticación técnica.

Asimismo, la creación de un corpus sintético out-of-domain, estructurado y normalizado en 3FN, permitió simular escenarios heterogéneos y evaluar la capacidad de generalización del modelo en condiciones más realistas. Esta base posibilitó la posterior integración coherente del clasificador dentro de un sistema de recomendación multicriterio funcional y estructurado.

5.2 Adaptaciones y cambios durante el desarrollo

A lo largo del proyecto se produjeron varios ajustes respecto al planteamiento inicial.

En primer lugar, el análisis del fenómeno de domain shift cobró mayor relevancia de la prevista originalmente. La necesidad de evaluar el modelo en un entorno más cercano a redes sociales condujo a la construcción de un corpus out-of-domain sintético, compuesto por perfiles heterogéneos. Esta decisión permitió observar las limitaciones reales del modelo fuera del entorno de entrenamiento y enriqueció significativamente el alcance del trabajo.

En segundo lugar, el sistema de recomendación evolucionó desde una idea centrada únicamente en el estado emocional hacia una arquitectura multicriterio basada en cuatro pilares: emoción, intereses, perfil viajero y presupuesto. Este cambio respondió a la necesidad de evitar recomendaciones excesivamente reduccionistas y construir un sistema más contextualizado y realista.

Finalmente, la incorporación de filtros éticos no estaba inicialmente planteada con el nivel de profundidad alcanzado. Sin embargo, durante el desarrollo del sistema surgió la reflexión sobre la responsabilidad que implica recomendar experiencias en función del estado emocional del usuario. Esto llevó a integrar mecanismos de advertencia en situaciones sensibles, ampliando la dimensión ética del proyecto.

5.3 Limitaciones del estudio y de la tecnología empleada

A pesar de los resultados positivos, el proyecto presenta limitaciones que deben mencionarse.

En primer lugar, el corpus out-of-domain utilizado para la validación es sintético. Aunque se diseñó para simular perfiles heterogéneos y lenguaje ruidoso y complejo propio de redes sociales, no sustituye completamente a datos reales obtenidos en entornos naturales. Esto puede limitar la extrapolación directa de los resultados a contextos reales de uso.

En segundo lugar, el modelo seleccionado, si bien eficiente y estable, se basa en representaciones vectoriales tradicionales que no capturan matices semánticos profundos ni dependencias contextuales complejas. La incorporación futura de modelos basados en embeddings podría mejorar la sensibilidad del sistema ante ambigüedades lingüísticas.

Asimismo, el sistema de recomendación implementado utiliza ponderaciones fijas para combinar los cuatro pilares. Si bien estas ponderaciones fueron definidas de forma razonada, no han sido optimizadas mediante técnicas de aprendizaje automático o validación con usuarios reales, lo que representa una oportunidad de mejora futura.

Por último, el análisis emocional se basa en cinco categorías discretas. Esta simplificación facilita el modelado, pero no captura la complejidad continua y multidimensional de la experiencia afectiva humana.

5.4 Impacto del proyecto

Más allá de sus limitaciones, el proyecto demuestra la viabilidad de integrar computación afectiva en sistemas de recomendación personalizados. La propuesta no se limita a optimizar preferencias explícitas, sino que introduce el contexto emocional como variable central en la toma de decisiones algorítmica.

Este enfoque tiene implicaciones relevantes en ámbitos como el turismo, el comercio electrónico y los sistemas de asistencia digital. Además, la incorporación de filtros éticos sugiere una línea de desarrollo donde la personalización no solo persigue eficiencia, sino también responsabilidad.

En conjunto, el trabajo evidencia que la combinación de análisis emocional, modelado supervisado y arquitectura multicriterio puede generar sistemas más empáticos y contextualizados, abriendo nuevas posibilidades para el diseño de tecnologías centradas en la experiencia humana.

Capítulo 6. CONCLUSIONES

6.1 Conclusiones del trabajo

El presente Trabajo Fin de Máster ha tenido como objetivo general el diseño y desarrollo de un sistema de recomendación de viajes personalizados que integra el análisis emocional del lenguaje del usuario mediante técnicas de Procesamiento del Lenguaje Natural (PLN), con el propósito de generar sugerencias contextualizadas y empáticas basadas en su estado afectivo.

A partir de los resultados obtenidos, puede afirmarse que dicho objetivo ha sido alcanzado satisfactoriamente. Respecto a los objetivos específicos, se han cumplido los siguientes hitos:

En primer lugar, se ha curado y estructurado un corpus emocional en español de alta calidad **(O1)**, construido a partir de reseñas reales y sometido a un proceso de etiquetado dual ciego. La consistencia inter-evaluador ha sido evaluada mediante métricas estadísticas, garantizando la fiabilidad del gold standard empleado para el entrenamiento del clasificador.

En segundo lugar, se ha implementado un pipeline de preprocesamiento lingüístico exhaustivo **(O2)**, que incluye normalización, limpieza, tokenización, lematización y estrategias específicas de mitigación del sesgo de dominio, como el enmascaramiento de vocabulario cinematográfico. Este proceso ha permitido optimizar la representación textual para su posterior modelado.

En tercer lugar, se han entrenado y evaluado diversos modelos supervisados de clasificación emocional en un entorno in-domain **(O3)**, comparando distintas representaciones vectoriales y algoritmos tradicionales. El modelo basado en Bag of Words con Naive Bayes ha demostrado el mejor equilibrio entre rendimiento y estabilidad, siendo seleccionado como modelo final.

Posteriormente, se ha construido un corpus out-of-domain sintético compuesto por 25 perfiles heterogéneos y 300 textos con características propias del lenguaje de redes sociales **(O4)**, lo que ha permitido evaluar el comportamiento del sistema en un escenario más cercano a un contexto real de aplicación. La validación del modelo en este entorno **(O5)** ha evidenciado su capacidad de generalización, aunque también ha permitido identificar los límites inherentes al fenómeno del domain shift.

A partir de esta base, se ha diseñado un sistema de recomendación emocional multicriterio **(O6)**, estructurado en 4 pilares fundamentales: estado emocional, intereses, perfil viajero y presupuesto. La combinación ponderada de estos factores ha permitido generar rankings personalizados de destinos coherentes con el contexto afectivo y sociodemográfico del usuario.

Asimismo, se han incorporado filtros éticos y contextuales **(O7)** que no excluyen recomendaciones, pero introducen mecanismos de advertencia en situaciones potencialmente sensibles, tales como condiciones de vulnerabilidad emocional o restricciones relacionadas con accesibilidad e inclusión. Este enfoque refuerza la dimensión ética y responsable del sistema.

Finalmente, se han explorado casos de uso representativos y patrones emocionales agregados mediante herramientas de análisis visual y métricas globales **(O8)**, permitiendo identificar perfiles de comportamiento, niveles de consistencia emocional y patrones de disonancia entre emoción autopercibida e inferida.

En conjunto, el trabajo no solo demuestra la viabilidad técnica de integrar análisis emocional y sistemas de recomendación, sino que también propone un enfoque éticamente consciente para el desarrollo de sistemas personalizados basados en el lenguaje afectivo del usuario.

6.2 Conclusiones personales

La realización de este Trabajo Fin de Máster ha supuesto para mí mucho más que la culminación académica de un programa formativo; ha representado un punto de inflexión personal y profesional.

Mi formación de base como arquitecto implicó que, al iniciar este Máster en Big Data, me enfrentara a una curva de aprendizaje especialmente pronunciada. Muchos de los conceptos y herramientas me resultaban completamente nuevos. Los primeros meses estuvieron marcados por la sensación constante de estar explorando un territorio desconocido. Sin embargo, precisamente esa dificultad inicial convirtió el proceso en un reto profundamente estimulante. Cada avance, cada modelo entrenado, suponía una pequeña conquista personal.

Con el paso del tiempo, la complejidad dejó de ser intimidante y comenzó a transformarse en oportunidad. Este proyecto me ha permitido comprender que el análisis de datos y la

inteligencia artificial no son únicamente herramientas técnicas, sino lenguajes capaces de reinterpretar cualquier disciplina. Como arquitecto, siempre he concebido los espacios como experiencias humanas; este proyecto me ha demostrado que los datos también pueden diseñar experiencias, pero desde una dimensión intangible: la emocional.

El desarrollo de este TFM ha sido especialmente gratificante porque me ha permitido unir dos dimensiones que forman parte de mi identidad: la pasión por viajar y la inquietud por comprender cómo la tecnología puede mejorar la vida de las personas. Diseñar un sistema de recomendación de viajes que tenga en cuenta el contexto emocional del usuario no ha sido únicamente un ejercicio técnico, sino una reflexión sobre cómo nos relacionamos con nuestras propias emociones y cómo la tecnología podría acompañarnos de manera más empática.

He intentado que el sistema no se limite a recomendar destinos, sino que se aproxime a la figura de un amigo que conoce tu momento vital, que entiende tu estado de ánimo y que, desde esa comprensión, sugiere lo que puede ser mejor para ti. Incluso que sea capaz de advertirte cuando un destino quizá no sea el más adecuado, incorporando una dimensión ética y protectora. Esa intención (que la tecnología no solo optimice, sino que cuide) ha sido uno de los motores personales de este trabajo.

Creo firmemente que la computación afectiva representa una de las vías más prometedoras para integrar la inteligencia artificial en nuestra vida cotidiana de forma responsable y humana. Si la tecnología aprende a escuchar y comprender nuestras emociones, podrá convertirse en una herramienta más consciente, más contextual y más respetuosa.

Este proyecto me ha enseñado no solo a entrenar modelos o diseñar sistemas de recomendación, sino a mirar la tecnología desde una perspectiva más amplia. Me ha abierto los ojos a las infinitas posibilidades que surgen cuando se combinan disciplinas aparentemente alejadas. Y, sobre todo, me ha reafirmado en la convicción de que el aprendizaje continuo y la capacidad de reinventarse son, probablemente, las herramientas más valiosas en cualquier trayectoria profesional.

Este TFM no marca un final, sino el comienzo de una nueva forma de entender mi propio camino.

Capítulo 7. FUTURAS LÍNEAS DE TRABAJO

7.1 Mejora del clasificador emocional mediante Transformers

Una posible evolución del sistema desarrollado en este trabajo consiste en la incorporación de arquitecturas Transformer preentrenadas, especialmente modelos adaptados al lenguaje informal en español. En este contexto, un candidato especialmente adecuado sería RoBERTuito, basado en RoBERTa [20] y entrenado sobre grandes volúmenes de datos procedentes de Twitter en español. Su especialización en abreviaturas, variaciones ortográficas, emojis y fenómenos propios de redes sociales lo convierte en una alternativa particularmente relevante para escenarios out-of-domain como el corpus sintético construido en este proyecto.

A diferencia de los enfoques basados en BoW o TF-IDF, los modelos Transformer generan representaciones contextualizadas capaces de capturar dependencias semánticas complejas y matices emocionales más sutiles. Esto permite modelar mejor ambigüedades léxicas, ironía ligera o variaciones pragmáticas frecuentes en textos informales, lo que potencialmente podría reducir la caída de rendimiento observada al trasladar el modelo desde el dominio cinematográfico hacia entornos con mayor variabilidad lingüística.

Para adaptar RoBERTuito a la tarea de clasificación emocional con cinco categorías (tristeza, ansiedad, calma, curiosidad y alegría), sería necesario realizar un proceso de fine-tuning [21]. Este procedimiento consiste en añadir una capa de clasificación sobre el embedding global del texto (habitualmente asociado al token especial de agregación) y reajustar los parámetros del modelo utilizando el corpus etiquetado del proyecto, optimizando una función de pérdida multiclase como la entropía cruzada. De este modo, el modelo conserva el conocimiento adquirido durante el preentrenamiento y lo especializa en la detección de emociones en español dentro del marco específico del sistema propuesto.

Si bien esta aproximación podría mejorar la capacidad de generalización y la robustez frente al domain shift, su implementación implica mayores requerimientos computacionales, necesidad de infraestructura adecuada y un proceso experimental más complejo, por lo que se plantea como una línea futura orientada a la evolución metodológica del sistema.

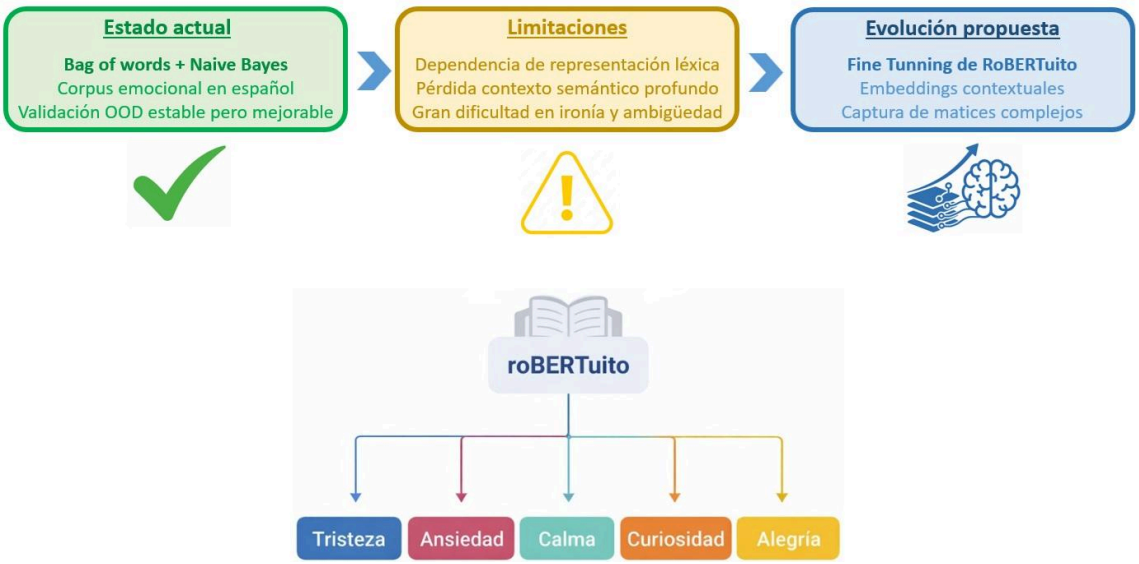

Figura 16. Líneas futuras - Mejora del clasificador mediante Transformers
Fuente: elaboración propia.

7.2 Matching efímero mediante segmentación dinámica

Una segunda línea futura de investigación propone evolucionar el sistema hacia un modelo de **segmentación emocional dinámica**, coherente con el carácter transitorio tanto de las emociones como de la experiencia turística. A diferencia de los enfoques clásicos de segmentación —basados en variables demográficas o preferencias relativamente estables—, este planteamiento asume que el estado afectivo constituye una variable no estacionaria, dependiente del contexto y del momento.

Las métricas de consistencia y disonancia emocional definidas previamente ofrecen una base cuantitativa para esta reconfiguración dinámica. En lugar de describir rasgos estructurales del individuo, los clusters representarían **instantáneas temporales de su estado afectivo**, en línea con los modelos de la computación afectiva que conciben la emoción como un fenómeno contextual, dinámico y situado [22].

A partir de estas métricas podrían emplearse algoritmos de clustering ampliamente documentados en la literatura (como K-Means o métodos jerárquicos) [23], generando agrupaciones cuya pertenencia no sería fija, sino actualizable y reversible en función de la

evolución emocional del usuario. El resultado no sería una segmentación permanente, sino una organización adaptativa sensible al momento.

Este marco permitiría implementar un **matching efímero entre usuarios**, basado en la coincidencia simultánea de estado emocional, localización y ventana temporal de viaje. Dado que los desplazamientos turísticos son, por naturaleza, experiencias acotadas en el tiempo, la conexión resultante no buscaría afinidades duraderas, sino sincronías contextuales intensas pero transitorias. De este modo, el sistema ampliaría su alcance hacia una dimensión social dinámica, alineada con la naturaleza cambiante de las emociones y de los propios viajes.

Aunque su implementación requeriría una base de usuarios más amplia para garantizar robustez estadística, esta línea futura extiende de forma coherente la arquitectura conceptual del sistema hacia modelos adaptativos y contextualmente conscientes.

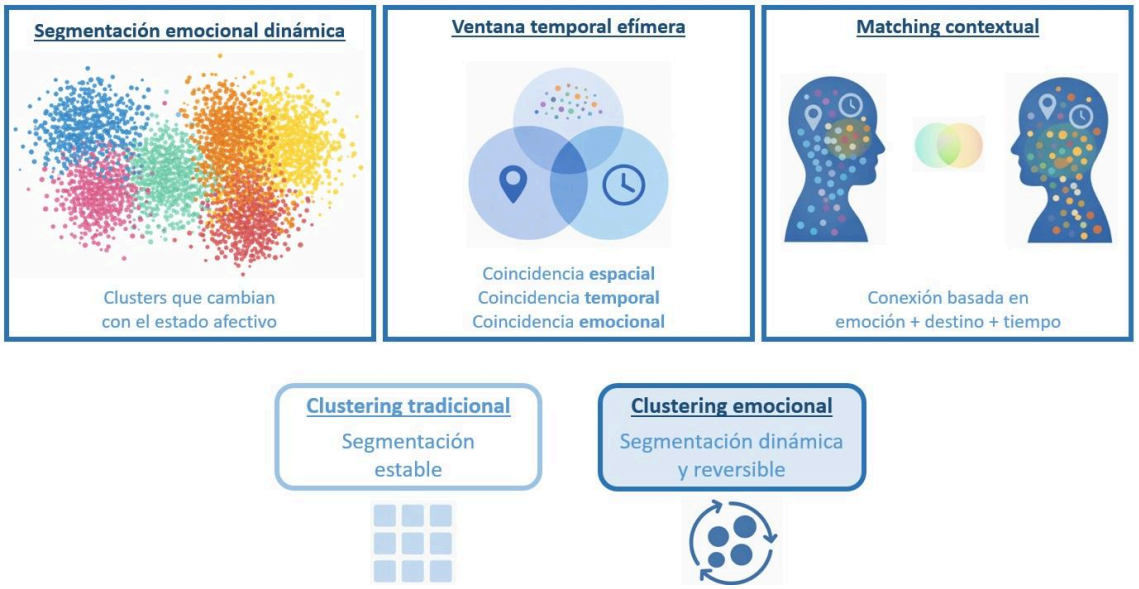

Figura 17. Líneas futuras - Matching efímero con segmentación dinámica
Fuente: elaboración propia.

7.3 Propuesta de arquitectura técnica escalable y despliegue

Más allá de las evoluciones metodológicas en el procesamiento de lenguaje natural y en la lógica de recomendación, la transición de TRIP(U) hacia un sistema de producción real requiere una infraestructura técnica robusta. Se propone una arquitectura de **microservicios desacoplados** diseñada para garantizar la escalabilidad y la baja latencia.

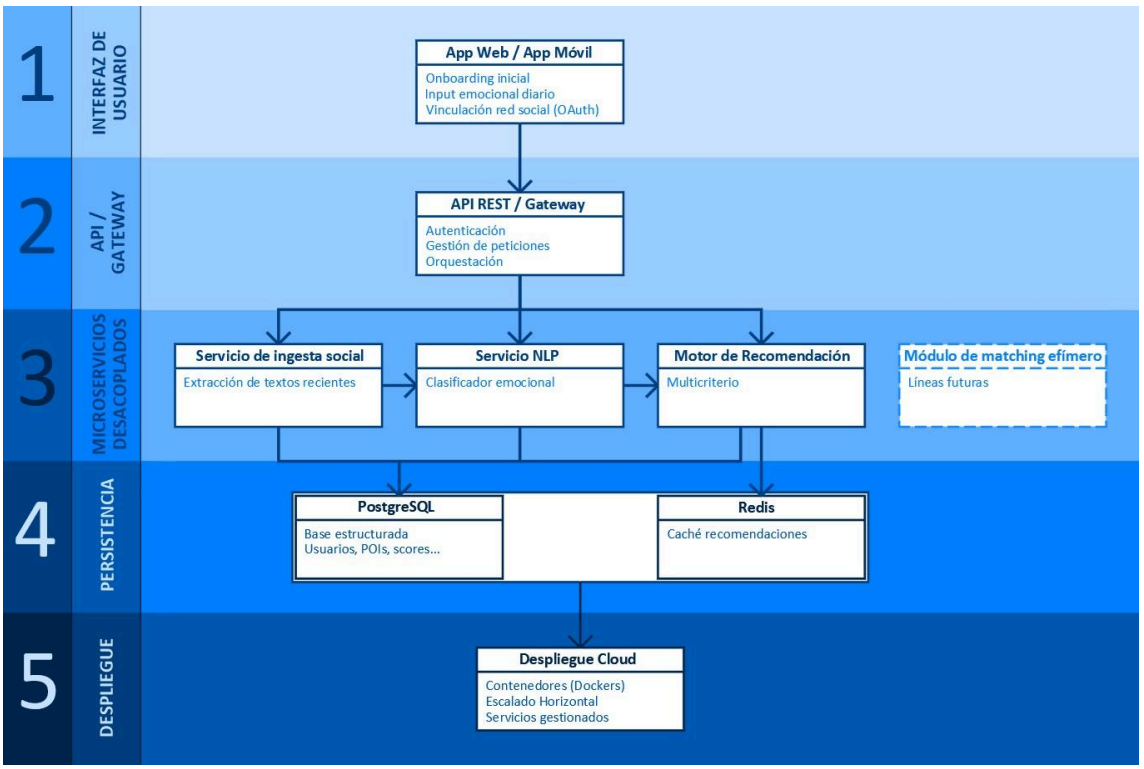

Figura 18. Líneas futuras - Diagrama de arquitectura técnica

Fuente: elaboración propia.

La estructura se organizaría en las cinco capas funcionales que se observan en el diagrama:

Capa 1 - Interfaz de Usuario: Representa el punto de interacción (App Web/Móvil) donde se gestiona el *onboarding* y la vinculación con redes sociales mediante protocolos OAuth.

Capa 2 - API / Gateway: Actúa como punto de entrada único que centraliza la autenticación, la gestión de peticiones y la orquestación de los flujos de datos hacia los servicios internos.

Capa 3 - Microservicios Desacoplados: Corresponde al núcleo lógico del sistema. El *Servicio de Ingesta Social* extrae los textos recientes; el *Servicio NLP* ejecuta el clasificador emocional (permitiendo futuras mejoras); y el *Motor de Recomendación* aplica la lógica multicriterio.

Capa 4 - Persistencia: Implementa un modelo híbrido. **PostgreSQL** almacena la base estructurada (usuarios, destinos, intereses), mientras que **Redis** funciona como una caché de recomendaciones para asegurar una respuesta instantánea al usuario (si nada ha cambiado).

Capa 5. Despliegue: Define la infraestructura final en la nube. Se basa en la contenedorización con **Docker** y el escalado horizontal, permitiendo que el sistema crezca de forma flexible.

Capítulo 8. REFERENCIAS

[1] J. A. Russell, "A circumplex model of affect," *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, 1980. doi: https://doi.org/10.1037/h0077714

[2] S. Poria, E. Cambria, N. Howard, G. Huang, and A. Hussain, "Fusing audio, visual and textual clues for sentiment analysis from multimodal content," *Neurocomputing*, vol. 174, pp. 50–59, 2016. doi: https://doi.org/10.1016/j.neucom.2015.01.095

[3] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proc. Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, Minneapolis, USA, 2019, pp. 4171–4186. doi: https://doi.org/10.18653/v1/N19-1423

[4] J. Cohen, "Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit," *Psychological Bulletin*, vol. 70, no. 4, pp. 213–220, 1968. doi: https://doi.org/10.1037/h0026256

[5] D. Hovy and S. L. Spruit, "The Social Impact of Natural Language Processing," in *Proc. 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Berlin, Germany, 2016, pp. 591–598. doi: https://doi.org/10.18653/v1/P16-2096

[6] R. Moya, "Críticas de películas - FilmAffinity en español," Kaggle, 2020. [Online]. Available: https://www.kaggle.com/datasets/ricardomoya/criticas-peliculas-filmaffinity-en-espaniol [Accessed: 20-Nov-2025].

[7] C. P. Chai, "Comparison of text preprocessing methods," *Natural Language Engineering*, vol. 29, no. 3, pp. 509–553, May 2023. doi: https://doi.org/10.1017/S1351324922000213

[8] M. A. Palomino and F. Aider, "Evaluating the Effectiveness of Text Pre-Processing in Sentiment Analysis," *Applied Sciences*, vol. 12, no. 17, p. 8765, Aug. 2022. doi: https://doi.org/10.3390/app12178765

[9] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," *Information Processing & Management*, vol. 24, no. 5, pp. 513–523, 1988. doi: https://doi.org/10.1016/0306-4573(88)90021-0

[10] K. S. Jones, "A statistical interpretation of term specificity and its application in retrieval," *Journal of Documentation*, vol. 28, no. 1, pp. 11–21, 1972. doi: https://doi.org/10.1108/eb026526

[11] A. McCallum and K. Nigam, "A comparison of event models for Naive Bayes text classification," *AAAI Workshop on Learning for Text Categorization*, 1998. https://cdn.aaai.org/Workshops/1998/WS-98-05/WS98-05-007.pdf

[12] T. Joachims, "Text categorization with Support Vector Machines: Learning with many relevant features," in *Proc. 10th European Conf. Machine Learning (ECML)*, 1998, pp. 137–142. doi: https://doi.org/10.1007/BFb0026683

[13] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 6, pp. 734–749, 2005. doi: https://doi.org/10.1109/TKDE.2005.99

[14] F. Ricci, L. Rokach, and B. Shapira, "Recommender Systems: Introduction and Challenges," in *Recommender Systems Handbook*, 2nd ed., Boston, MA, USA: Springer, 2015, pp. 1–34. doi: https://doi.org/10.1007/978-1-4899-7637-6_1

[15] R. Burke, "Hybrid recommender systems: Survey and experiments," *User Modeling and User-Adapted Interaction*, vol. 12, no. 4, pp. 331–370, 2002. doi: https://doi.org/10.1023/A:1021240730564

[16] B. D. Mittelstadt, P. Allo, M. Taddeo, S. Wachter, and L. Floridi, "The ethics of algorithms: Mapping the debate," *Big Data & Society*, vol. 3, no. 2, 2016. doi: https://doi.org/10.1177/2053951716679679

[17] C. Cath, "Governing artificial intelligence: ethical, legal and technical opportunities and challenges," *Philosophical Transactions of the Royal Society A*, vol. 376, no. 2133, 2018. doi: https://doi.org/10.1098/rsta.2018.0080

[18] European Commission, "Ethics Guidelines for Trustworthy AI," High-Level Expert Group on Artificial Intelligence, 2019. doi: https://doi.org/10.2759/346720

[19] D. A. Keim, J. Kohlhammer, G. Ellis, and F. Mansmann, "Mastering the information age: Solving problems with visual analytics," *Eurographics Association*, 2010. doi: https://doi.org/10.2312/14803

[20] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "RoBERTa: A robustly optimized BERT pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019. doi: https://doi.org/10.48550/arXiv.1907.11692

[21] C. Sun, X. Qiu, Y. Xu, and X. Huang, "How to fine-tune BERT for text classification?" in *China National Conference on Chinese Computational Linguistics (CCL)*, 2019, pp. 194–206. doi: https://doi.org/10.1007/978-3-030-32381-3_16

[22] R. A. Calvo and S. D'Mello, "Affect detection: An interdisciplinary review of models, methods, and their applications," *IEEE Transactions on Affective Computing*, vol. 1, no. 1, pp. 18–37, 2010. doi: https://doi.org/10.1109/T-AFFC.2010.1

[23] R. Xu and D. Wunsch, "Survey of clustering algorithms," *IEEE Transactions on Neural Networks*, vol. 16, no. 3, pp. 645–678, 2005. doi: https://doi.org/10.1109/TNN.2005.845141

Declaración sobre el uso de Inteligencia Artificial

En cumplimiento de la normativa de integridad académica, se informa del uso de los siguientes recursos como asistencia técnica y de refinamiento en la formalización de este trabajo:

Redacción y síntesis de contenidos:

Herramienta: ChatGPT (v. 4o).
Uso: Soporte en la síntesis de secciones extensas de la memoria para la elaboración de resúmenes y apoyo en la revisión de la redacción académica para mejorar la fluidez y cohesión.
Interacción: Consultas iterativas para el refinamiento de estilo y la adaptación de contenidos técnicos al formato de la memoria.

Asistencia en el desarrollo de código:

Herramienta: Gemini (Google).
Funciones: Asistente integrado en el entorno Jupyter Notebook para la depuración de errores de código (*debugging*), optimización de funciones de preprocesamiento y apoyo en la generación de scripts para la visualización de resultados.
Interacción: Peticiones de optimización lógica y resolución de errores sintácticos en el pipeline de procesamiento de datos.

Gestión de referencias bibliográficas:

Herramienta: Mendeley Reference Manager.
Funciones: Organización del catálogo bibliográfico, gestión de metadatos de las fuentes consultadas y automatización del formato de citación bajo la norma IEEE.
Interacción: Uso funcional para la normalización de la bibliografía.

Búsqueda y exploración de literatura académica:

Herramienta: Perplexity AI.
Funciones: Localización preliminar de literatura científica relacionada con computación afectiva y sistemas de recomendación híbridos, facilitando la revisión del estado del arte.
Interacción: Búsquedas temáticas dirigidas a validar el estado del arte en modelos de personalización emocional.

Se hace constar que todas las decisiones metodológicas, diseño experimental, análisis de resultados y redacción final del trabajo son responsabilidad exclusiva del autor.

Capítulo 9. ANEXOS

9.1 Manual de codificación

- **Versión:** v3
- **Elaborado por:** Francisco Javier Cristóbal Fernández *(anotador_A)*
- **Colaboradora:** Diana Mora López *(anotador_B)*
- **Fecha:** Noviembre, 2025

Objetivo del manual

Este manual tiene como finalidad garantizar la consistencia en la clasificación emocional de un conjunto de reseñas de películas obtenidas de la base de datos de FilmAffinity (8.603 reviews).

El propósito es asignar a cada review una emoción predominante entre cinco posibles categorías. Las emociones deben identificarse a partir del tono general, el contenido semántico y las expresiones lingüísticas del texto, no del contenido emocional de la película en sí.

Importante: No clasificamos cómo es la película, sino cómo se siente el autor de la review al escribir sobre ella. Se evalúa la emoción expresada (lo inferido del texto), no la emoción que la película busca provocar.

Escala de emociones y códigos

Código 0 — Tristeza
Expresa decepción, desánimo, frustración o negatividad. El tono es pesimista o melancólico.
Polaridad: Negativa. *Nivel de energía:* Media

Código 1 — Ansiedad
Expresa nerviosismo, inquietud, tensión, incomodidad, agobio, estrés o frustración activa.
Polaridad: Negativa. *Nivel de energía:* Alta.

Código 2 — Calma
Expresa serenidad, tranquilidad, equilibrio, relax o sensación de paz.
Polaridad: Neutra / Positiva. *Nivel de energía:* Baja.

Código 3 — Curiosidad

Expresa interés, admiración, excitación, sorpresa, asombro o fascinación ante algo inesperado.

Polaridad: Positiva. *Nivel de energía:* Alta

Código 4 — Alegría

Expresa diversión, placer, felicidad o buen humor. El tono es optimista, social y de disfrute.

Polaridad: Positiva. *Nivel de energía:* Media

Instrucciones generales de codificación

- Cada review completa equivale a una unidad de análisis. No se codifican fragmentos.
- Se debe identificar la emoción predominante que transmite el autor. En caso de emociones mixtas, debe elegirse la dominante o más persistente en el texto.
- Si una review presenta dos emociones cercanas (por ejemplo, curiosidad y alegría), prevalece la de menor energía (alegría) para mantener consistencia.
- No existe una categoría “neutral” o “sin emoción”. Si el texto es muy descriptivo pero con tono relajado, se clasifica como calma (2).
- En caso de identificarse sarcasmo, debe detectarse la emoción implícita en el mismo.
- El etiquetado se basa en lenguaje explícito e implícito. Ejemplos:
- Lenguaje explícito: palabras emocionalmente cargadas (“me encantó”, “me aburrió”).
- Lenguaje implícito: expresiones valorativas o modales (“fue demasiado agotadora”).

Guía detallada de cada emoción

Código 0 — Tristeza

Expresa decepción, tristeza, desánimo, amargura o desilusión. El tono es pesimista o fatalista. Indicadores lingüísticos frecuentes:

- Léxicos negativos: malo, decepcionante, aburrido, vacío, sin sentido, insoportable, detesto, horrendo, lento, tedioso, pésimo…
- Adverbios de frustración: demasiado, poco, nunca, nada…
- Verbos negativos: odiar, fallar, defraudar, arruinar…
- Puntuación o estilo: frases cortas, exclamaciones negativas, juicios categóricos.
- Uso del pasado como resignación: “Pensé que sería mejor”

Código 1 — Ansiedad

Expresa tensión, agobio, incomodidad, miedo o nerviosismo. El tono es activo y saturado.

- Léxicos de incomodidad: tenso, caótico, confuso, agobiante, desesperante, ansioso…
- Verbos de sobrecarga: cansar, saturar, no soportar, abrumar, agotar, tensar…
- Ritmo acelerado, frases largas, exceso de exclamaciones o repeticiones.
- Valoraciones ansiosas: “me puso nervioso”, “no pude disfrutarla”, “temblaba”.

Código 2 — Calma

Expresa serenidad, satisfacción, equilibrio o contemplación. Puede ser una crítica positiva o simplemente reflexiva, con tono pausado.

- Léxicos suaves: tranquila, sencilla, pausada, armoniosa, agradable, sutil, humana…
- Verbos introspectivos: reflexionar, conectar, recordar, disfrutar, apreciar…
- Estructuras narrativa fluida, frases medianas, sin exceso de exclamaciones.
- Expresiones tipo: “una experiencia serena”, “transmite paz”.

Código 3 — Curiosidad

Expresa asombro, admiración, interés o fascinación ante algo inesperado. Suele tener un tono analítico o exploratorio.

- Léxicos de descubrimiento: sorprendente, original, innovadora, intrigante, fascinante.
- Verbos de exploración: descubrir, aprender, cuestionar, abrir los ojos.
- Estructura de asombro y excitación, frases usualmente cortas con exclamaciones.
- Expresiones tipo: “me dejó impactado”, “quedé en shock”, “ahora estoy intrigado”.

Código 4 — Alegría

Expresa entusiasmo, diversión, placer o energía positiva. El tono es optimista, festivo o emocionalmente expansivo.

- Léxicos positivos: divertida, encantadora, brillante, inspiradora, maravillosa, adorable…
- Verbos de disfrute: encantar, reír, disfrutar, amar, fascinar…
- Uso de exclamaciones positivas, adjetivos superlativos (increíble, espectacular).
- Expresiones tipo: “me hizo feliz”, “no paré de sonreír”, “una experiencia maravillosa”.

Este esquema emocional se inspira en los **modelos circunflejos de Russell** [1], adaptados a la práctica de análisis de sentimiento y emociones en texto.

9.2 Métricas de acuerdo entre anotadores

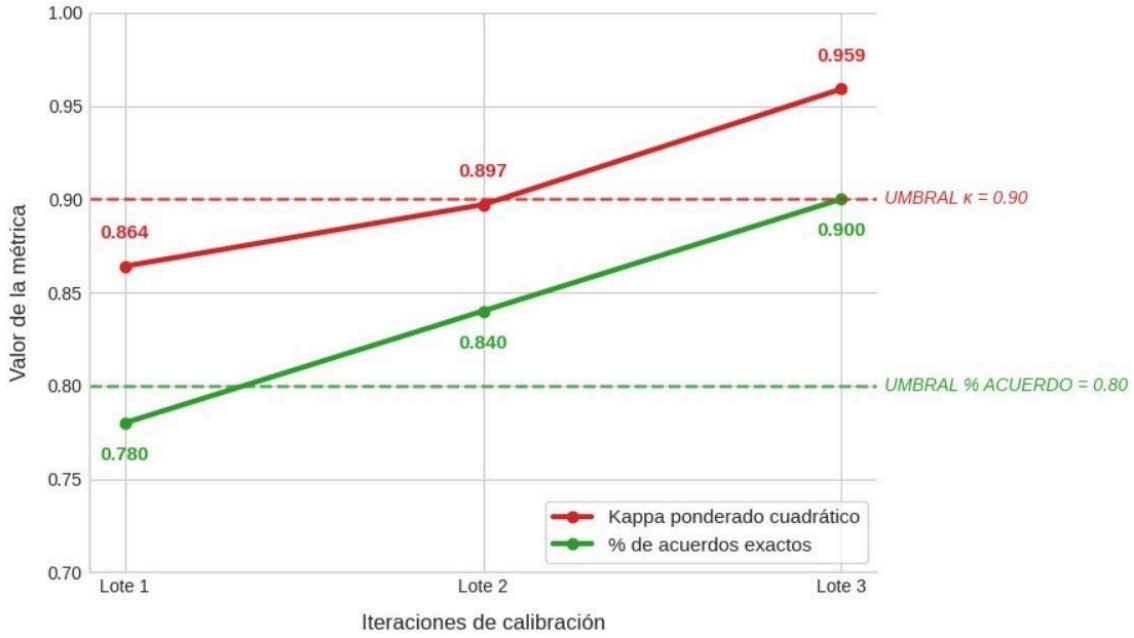

Figura A. Anexo 9.2. Evolución del acuerdo inter-anotador

Fuente: elaboración propia.

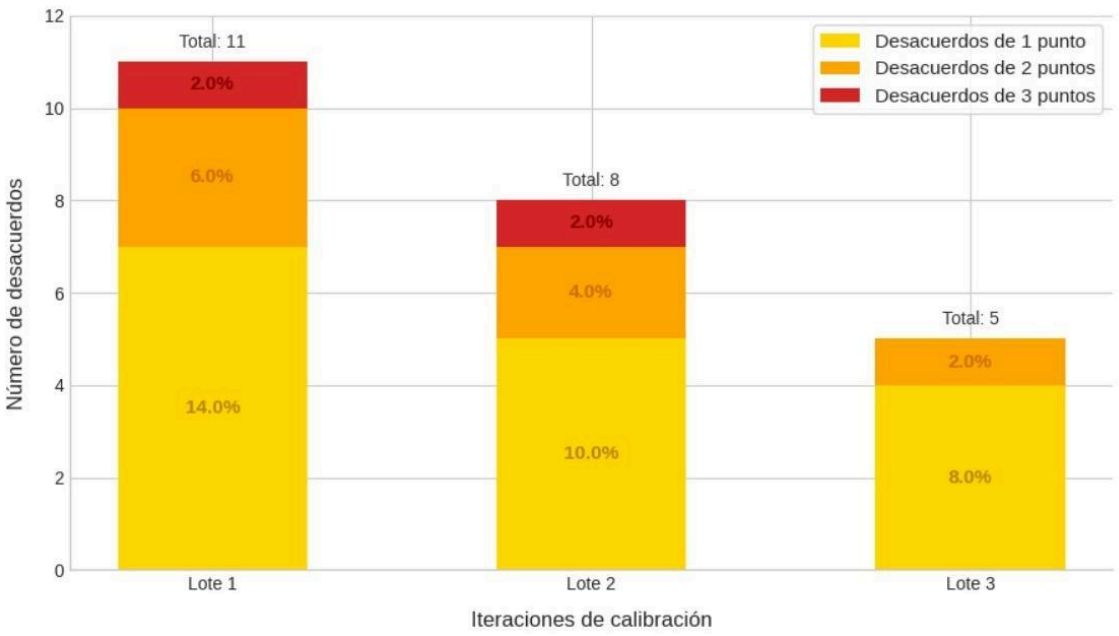

Figura B. Anexo 9.2. Evolución de los desacuerdos por tipo de salto

Fuente: elaboración propia.

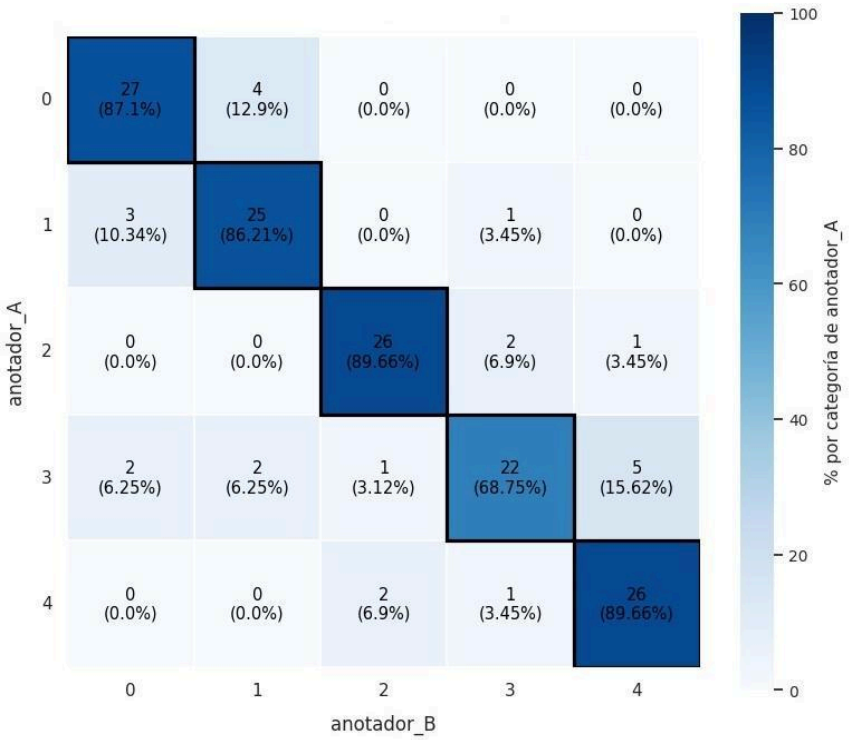

Figura C. Anexo 9.2. Matriz de confusión agregada (3 lotes combinados)

Fuente: elaboración propia.

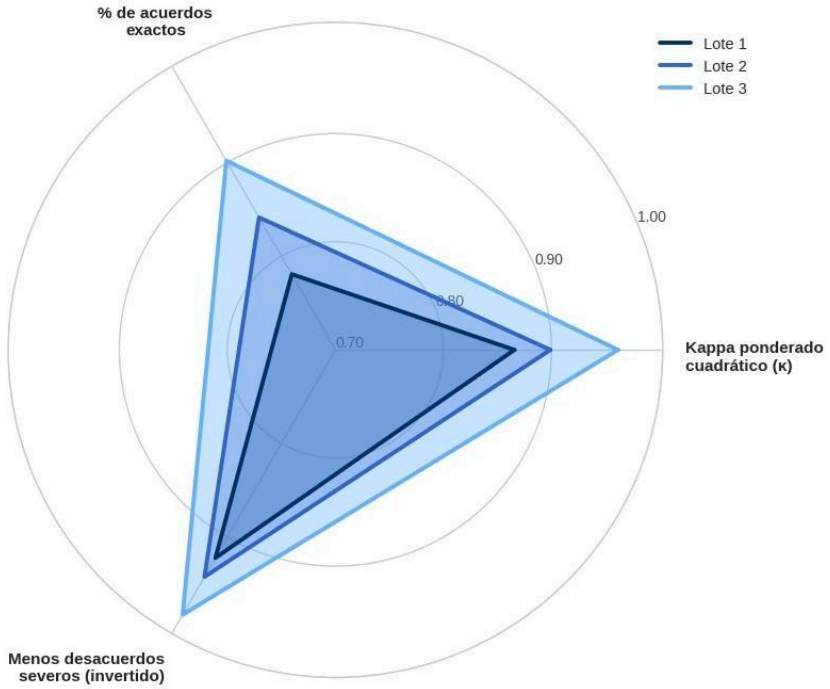

Figura D. Anexo 9.2. Desempeño global de calibración por lote

Fuente: elaboración propia.

Los resultados agregados confirman que el proceso de calibración alcanzó plenamente los objetivos planteados. El **valor global del kappa ponderado cuadrático (κ=0.906)** supera el umbral mínimo de 0.90, evidenciando un nivel de **acuerdo casi perfecto** entre los anotadores. Este valor promedio refleja la consolidación progresiva observada a lo largo de los tres lotes, donde κ fue incrementándose de 0.864 (lote 1) a 0.897 (lote 2) y finalmente a 0.959 (lote 3).

El **porcentaje de acuerdos exactos (84%)** mantiene la tendencia ascendente iniciada en la segunda iteración, situándose significativamente por encima del mínimo del 80% establecido como referencia. Este indicador, junto con la mejora del kappa, muestra que los desacuerdos remanentes son escasos y tienden a concentrarse en casos límite, sin impacto relevante sobre la consistencia global del etiquetado.

La **distribución de los desacuerdos** confirma esta estabilidad:

- Los errores leves (±1 punto) representan apenas el 10,7% del total (16 de 150).
- Los desacuerdos moderados (±2 puntos) son muy poco frecuentes (4%).
- Los desacuerdos severos (±3 puntos) son prácticamente anecdóticos (2 casos, 1.3%).
- No existen desacuerdos de 4 puntos, garantiza la ausencia de divergencias extremas.

En la **matriz de confusión global** se observa un patrón fuertemente diagonal, con la mayor concentración de acuerdos exactos en las categorías 0, 1, 2 y 4. Las categorías intermedias (2–3) presentan una ligera dispersión, coherente con los límites más sutiles del continuo emocional, pero sin generar asimetrías ni sesgos sistemáticos entre anotadores. En particular:

- Las clases bajas (0–1–2) presentan una **consistencia casi total**, con desvíos mínimos y siempre hacia categorías contiguas.
- Las clases altas (3–4) concentran los pocos desacuerdos de mayor magnitud, pero sin patrones recurrentes ni desplazamientos sistemáticos.

Tabla A. Anexo 9.2. Cumplimiento de métricas globales en el agregado de los 3 lotes

Métrica	Umbral mínimo	Resultado global	Cumplimiento
κ ponderado cuadrático	≥ 0.90	0.906	✅
% de acuerdos exactos	≥ 0.80	0.84	✅
Control de saltos graves	≤ 3 con diff ≥2, sin diff=4	6 de 2p, 2 de 3p, 0 de 4p	✅

Fuente: elaboración propia.

9.3 Lexicón de dominio cinematográfico

Este lexicón agrupa los términos y entidades propias del dominio cinematográfico detectadas en el corpus base de *FilmAffinity*. Su función dentro del pipeline de preprocesamiento es **neutralizar la información no transferible** (nombres de películas, actores, tecnicismos del cine, referencias a premios o géneros) que podrían sesgar el modelo emocional hacia patrones de dominio. Estas ocurrencias fueron enmascaradas bajo el token **[DOM_CINE]**, complementadas con el sistema de reconocimiento de entidades nombradas (*spaCy es_core_news_lg*) que asigna etiquetas específicas: **[ENT_PER]**, **[ENT_ORG]**, **[ENT_LOC]**, **[ENT_OBRA].**

Títulos de películas del corpus base de reviews -> [ENT_OBRA]

[8 apellidos vascos, 8 apellidos catalanes, Ágora, Airbag, Alatriste, Ahora o nunca, Atrapa la bandera, “Campeones”, Celda 211, Días de fútbol, El bola, El laberinto del fauno, El mejor verano de mi vida, El orfanato, El otro lado de la cama, Es por tu bien, Juana la Loca, La gran aventura de Mortadelo y Filemón, La isla mínima, Las 13 rosas, Las aventuras de Tadeo Jones, Lo dejo cuando quiera, Lo imposible, Los crímenes de Oxford, Los dos lados de la cama, Mar adentro, Mientras dure la guerra, Palmeras en la nieve, Perfectos desconocidos, Perdiendo el norte, Planet 51, REC, Regresión, Superlópez, Tadeo Jones 2, Tengo ganas de ti, Todo sobre mi madre, Torrente (saga completa), Tres metros sobre el cielo, Villaviciosa de al lado, “Volver”].

Profesiones y roles cinematográficos -> [DOM_CINE]

[director, directora, guionista, actor, actriz, intérprete, reparto, elenco, protagonista, coprotagonista, deuteragonista, secundarios, figurante, personaje, papel, cameo, doblaje, voz, especialista, figuración, producción, realizador, crítico, espectador, espectadora].

Lenguaje técnico de crítica y realización -> [DOM_CINE]

[guión, montaje, dirección, cinematografía, iluminación, fotografía, plano, encuadre, cámara, escena, secuencia, ritmo, narración, argumento, historia, clímax, estructura, desarrollo, final, desenlace, créditos, metraje, banda sonora, música, diálogo, interpretación, actuación, sobreactuación, efectos especiales, CGI, vestuario, decorado, ambientación, edición, posproducción, rodaje, guión técnico, obra maestra, película, film, cinta, peli, producción, historia real, basada en hechos reales, predecible, spoiler, cliché, estereotipo, comedia, terror, thriller, ciencia ficción, género, homenaje, sátira, parodia].

Premios, instituciones y referencias industriales -> [ENT_ORG]

[Goya, Oscar, Cannes, Academia, nominación, galardón, premio, jurado, taquilla, recaudación, estreno, cartelera, distribuidora, productora, festival, remake, secuela, precuela, trilogía, saga, doblada, subtítulos, filmografía].

Nombres propios y entidades del dominio -> [ENT_PER]

[Almodóvar, Amenábar, Álex de la Iglesia, Santiago Segura, Javier Fesser, Javier Bardem, Guillermo del Toro, Luis Tosar, Penélope Cruz, Belén Rueda, Dani Rovira, Karra Elejalde, Carmen Machi, Clara Lago, Blanca Suárez, Ernesto Alterio, Javier Gutiérrez, Naomi Watts, Ethan Hawke, Jaume Balagueró, Paco Plaza, Rachel Weisz, Yon González, Daniel Monzón, Chaplin, Romero, Medeiros, Hipatia, Malamadre].

9.4 Análisis de vocabulario y términos dominantes

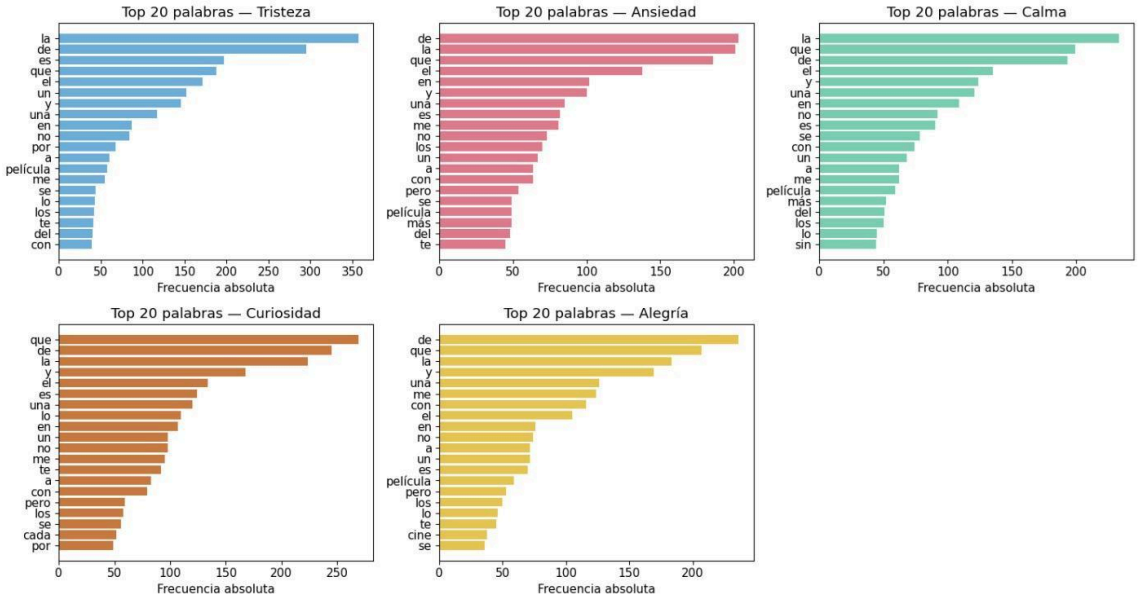

Figura E. Anexo 9.4. Top 20 palabras por emoción (sin filtrado)

Fuente: elaboración propia.

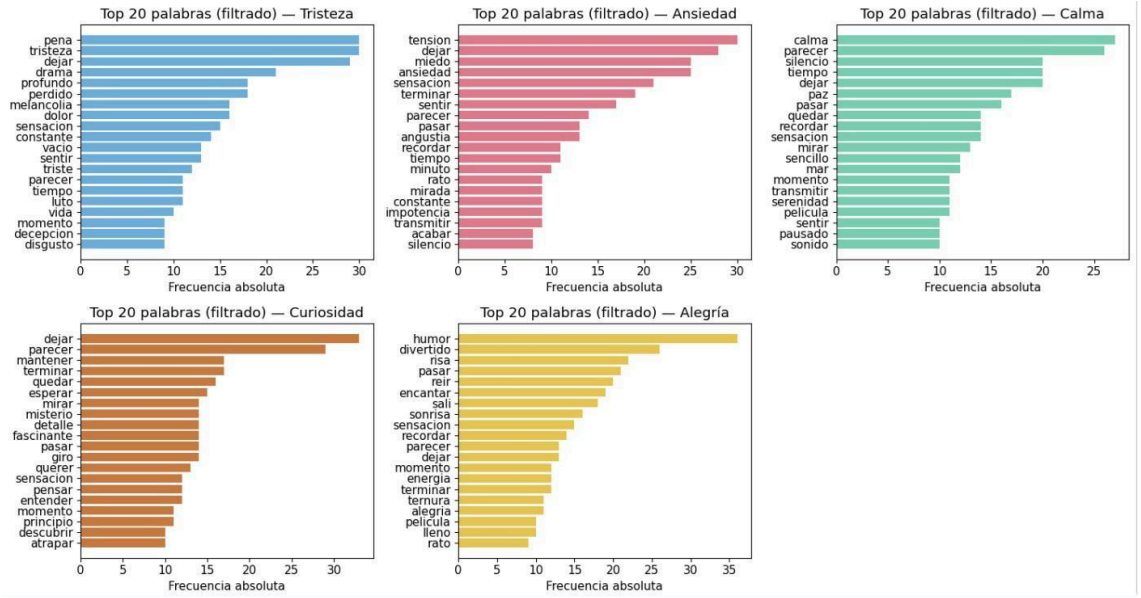

Figura F. Anexo 9.4. Top 20 palabras por emoción (filtrado)

Fuente: elaboración propia.

Figura G. Anexo 9.4. Wordcloud por emoción (filtrado)

Fuente: elaboración propia.

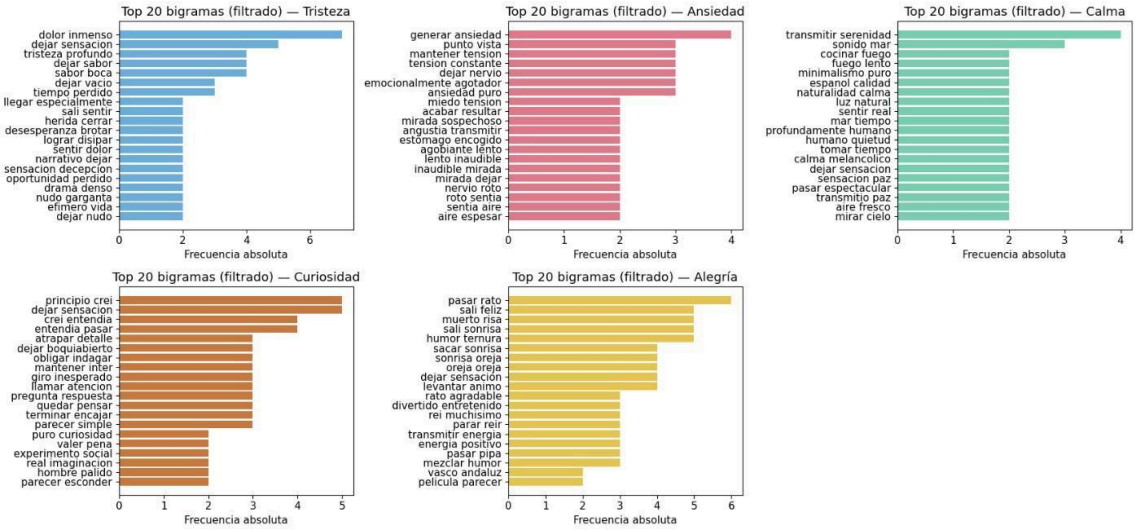

Figura H. Anexo 9.4. Top 20 bigramas por emoción (filtrado)

Fuente: elaboración propia.

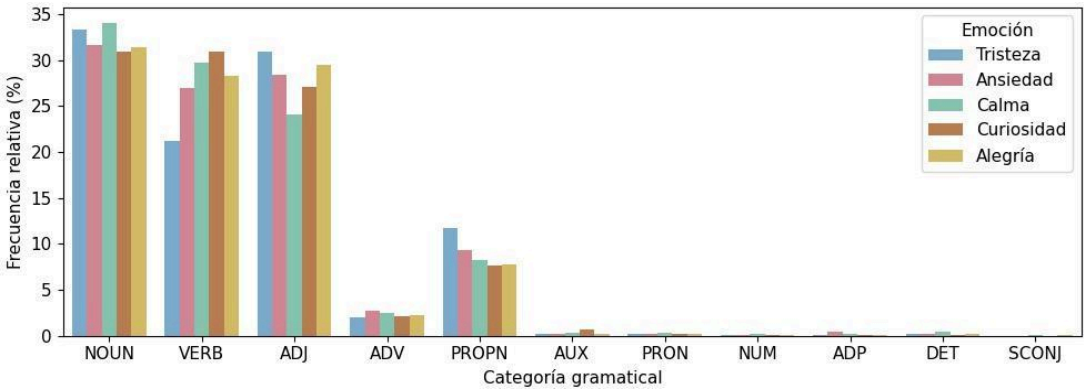

Figura I. Anexo 9.4. Distribución gramatical (POS) por emoción

Fuente: elaboración propia.

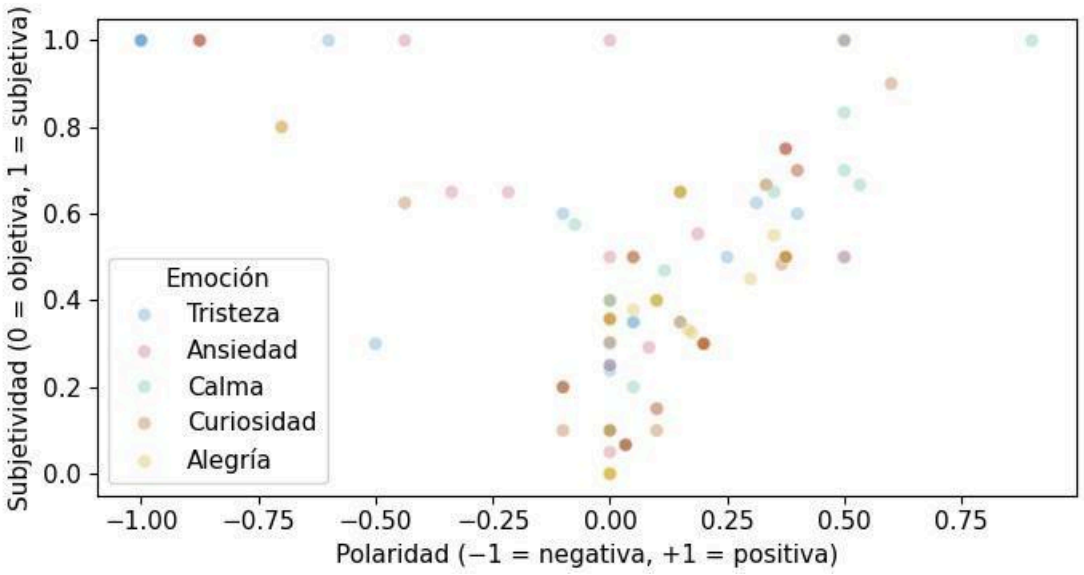

Figura J. Anexo 9.4. Dispersión de polaridad y subjetividad por emoción

Fuente: elaboración propia.

Tabla B. Anexo 9.4. Top 20 términos más distintivos entre emociones

Término	Chi2	p_value
tension	109.250000	1.051802e-22
tristeza	108.937500	1.226227e-22
ansiedad	100.000000	9.836624e-21
calma	97.379310	3.553182e-20
pena	90.666667	9.503357e-19
divertido	88.413793	2.860181e-18
humor	81.269231	9.377030e-17
miedo	76.000000	1.224262e-15
encantar	65.904762	1.658949e-13
paz	62.555556	8.416612e-13
reir	61.000000	1.787841e-12
melancolia	58.588235	5.742284e-12
sonrisa	54.222222	4.727941e-11
perdido	53.909091	5.498473e-11
drama	53.586207	6.424511e-11
risa	50.451613	2.906031e-10
vacio	46.714286	1.748832e-09
angustia	46.714286	1.748832e-09
dolor	46.000000	2.462851e-09
serenidad	44.000000	6.415777e-09

Fuente: elaboración propia.

9.5 Resultados completos de evaluación y comparación de modelos

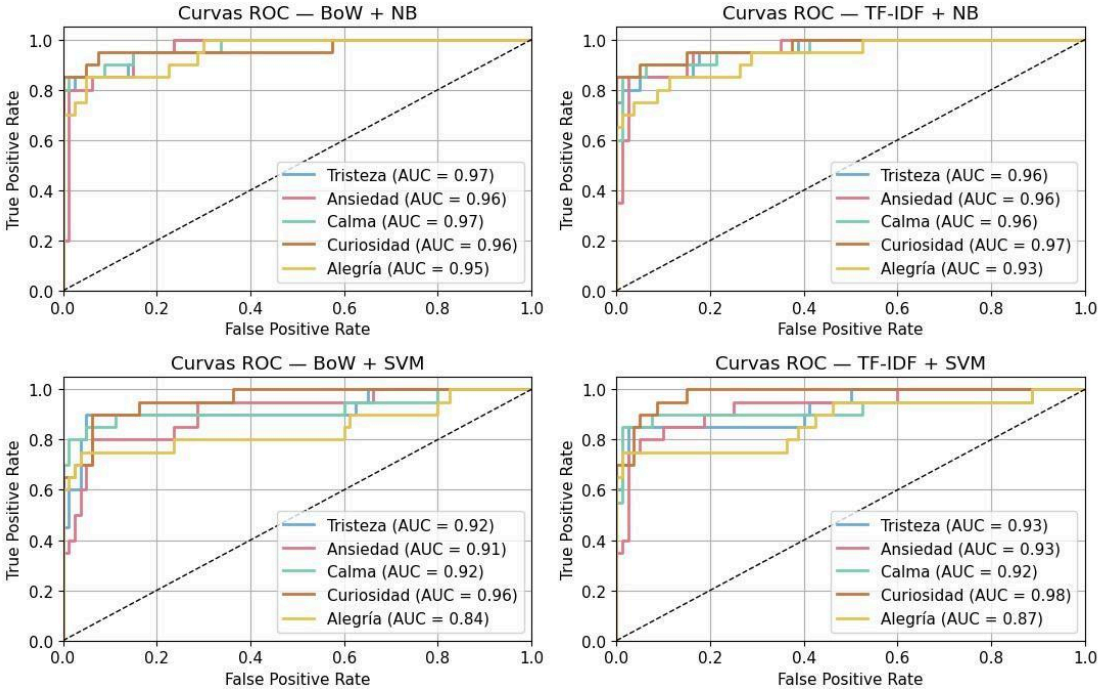

Figura K. Anexo 9.5. Curvas ROC por emoción en los modelos

Fuente: elaboración propia.

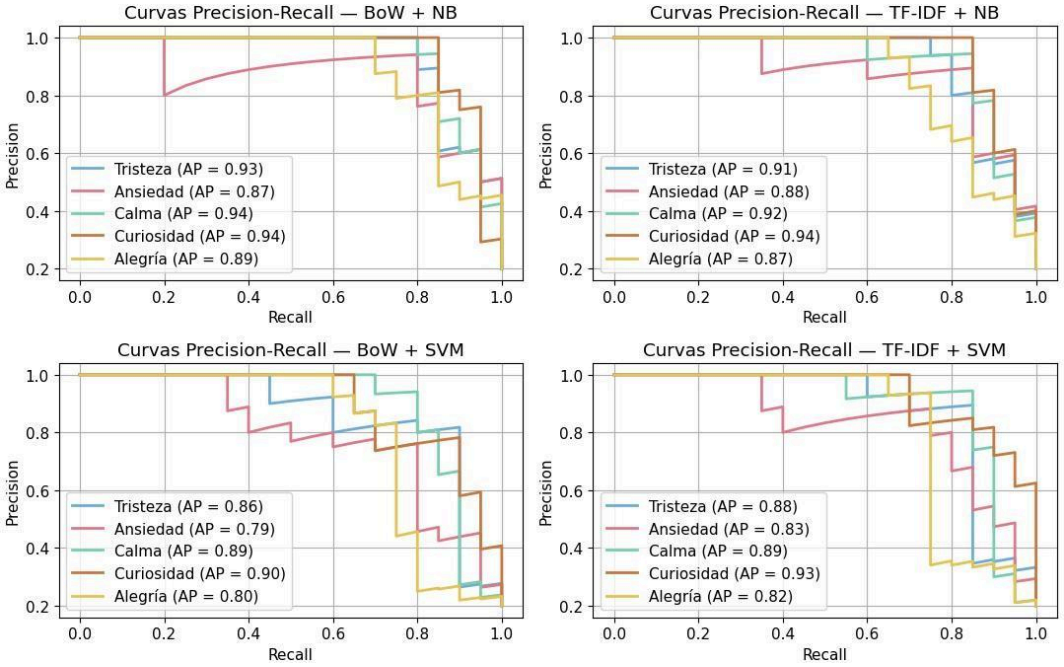

Figura L. Anexo 9.5. Curvas Precision-Recall por emoción en los modelos

Fuente: elaboración propia.

9.6 Resultados completos de validación out-of-domain

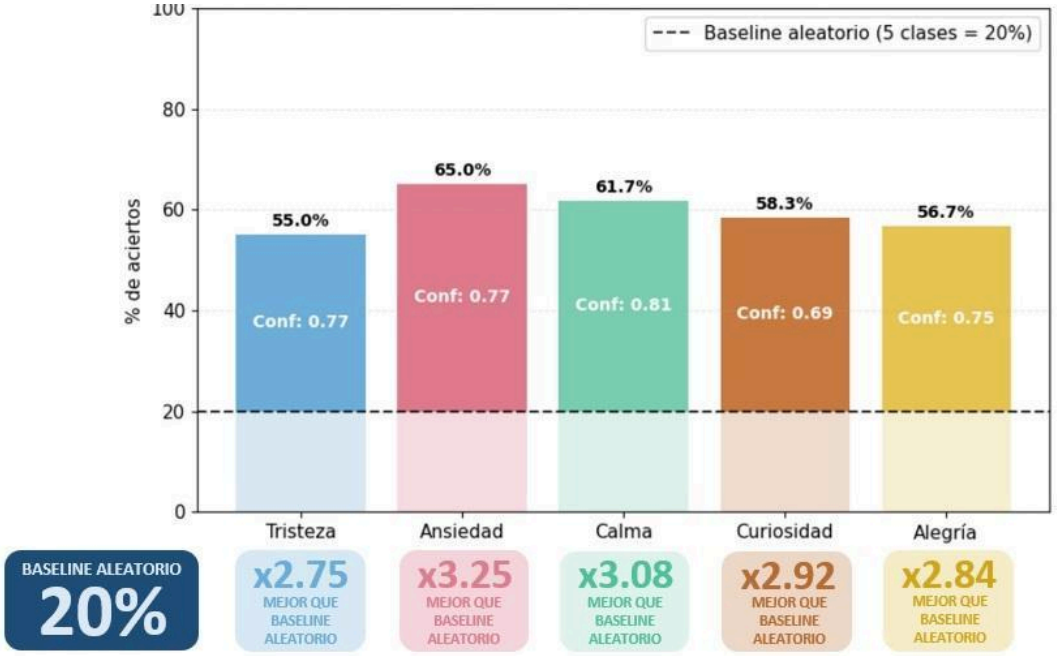

Figura M. Anexo 9.6. Tasa de aciertos y comparación respecto al baseline aleatorio

Fuente: elaboración propia.

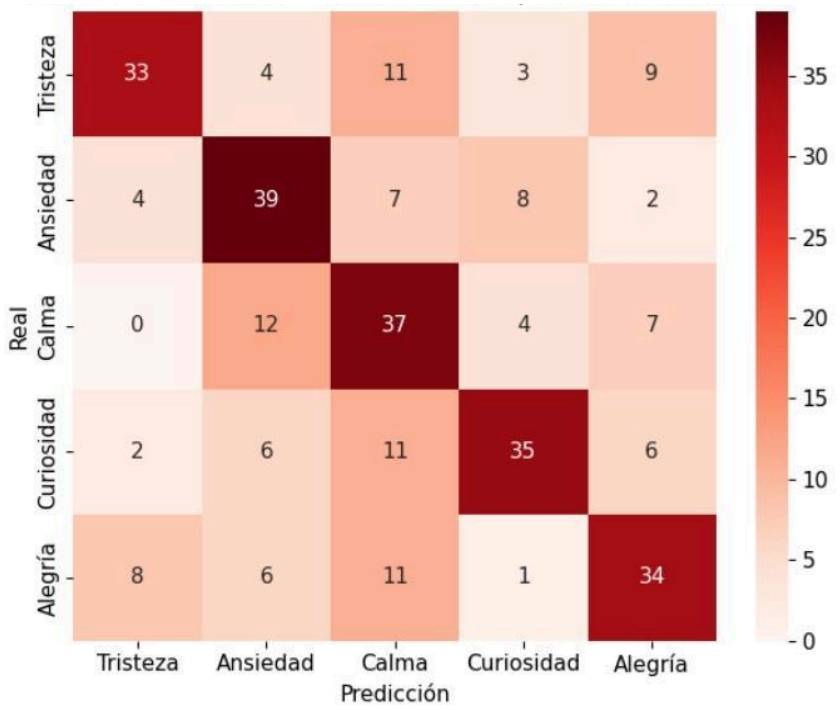

Figura N. Anexo 9.6. Matriz de confusión - BoW + NB (corpus out-of-domain)

Fuente: elaboración propia.

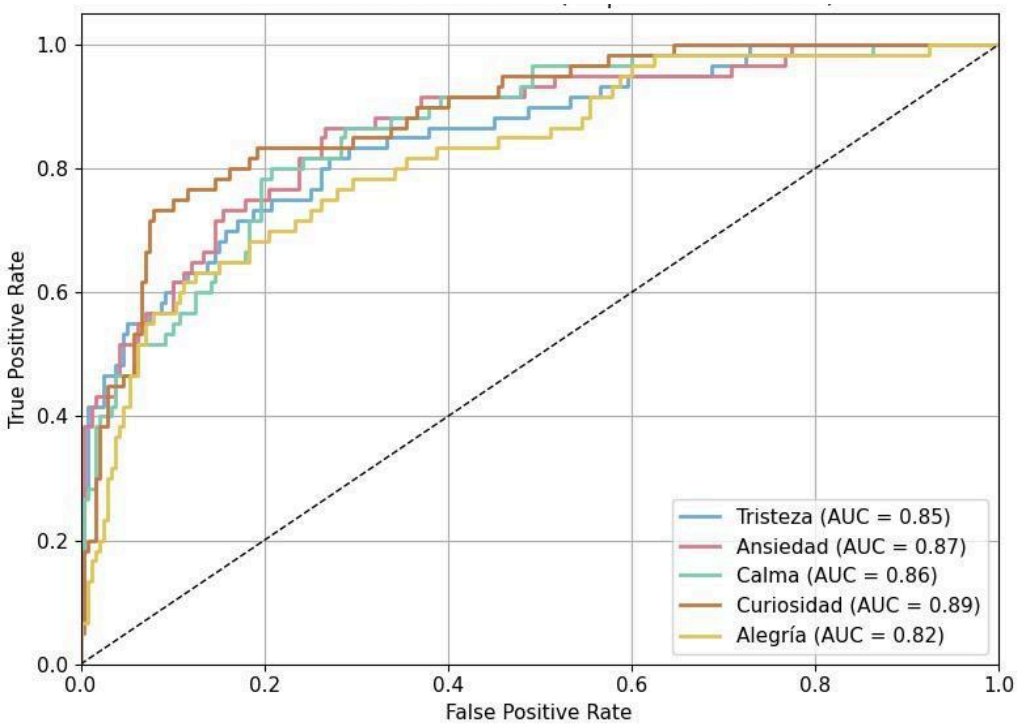

Figura O. Anexo 9.6. Curvas ROC - BoW + NB (corpus ot-of-domain)

Fuente: elaboración propia.

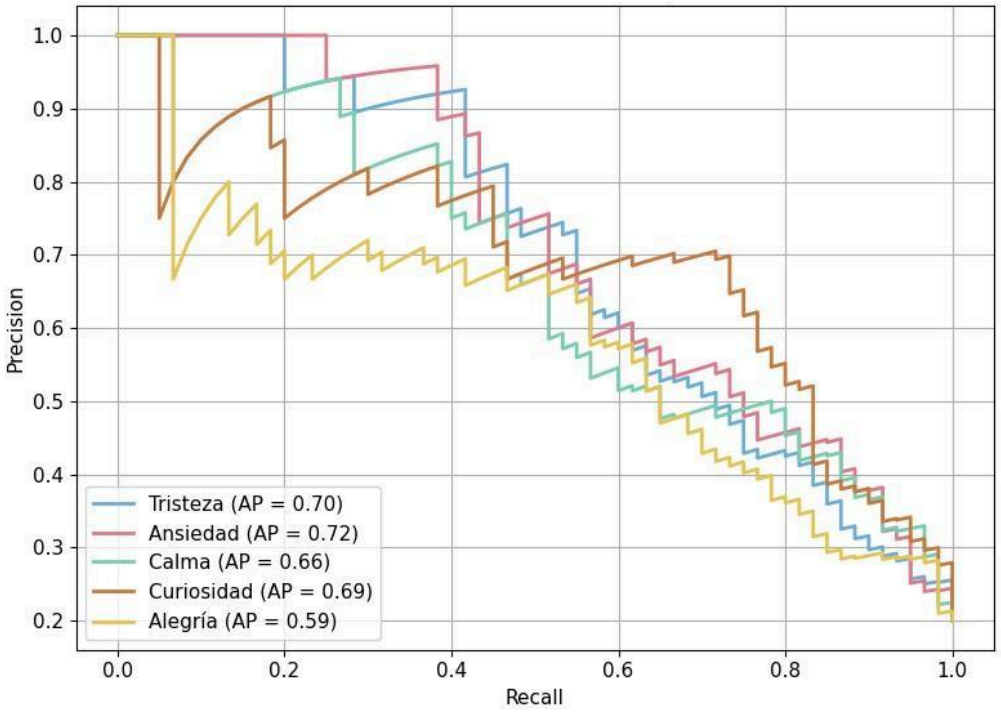

Figura P. Anexo 9.6. Curvas Precision-Recall - BoW + NB (corpus ot-of-domain)

Fuente: elaboración propia.

9.7 Análisis de resultados del sistema de recomendación

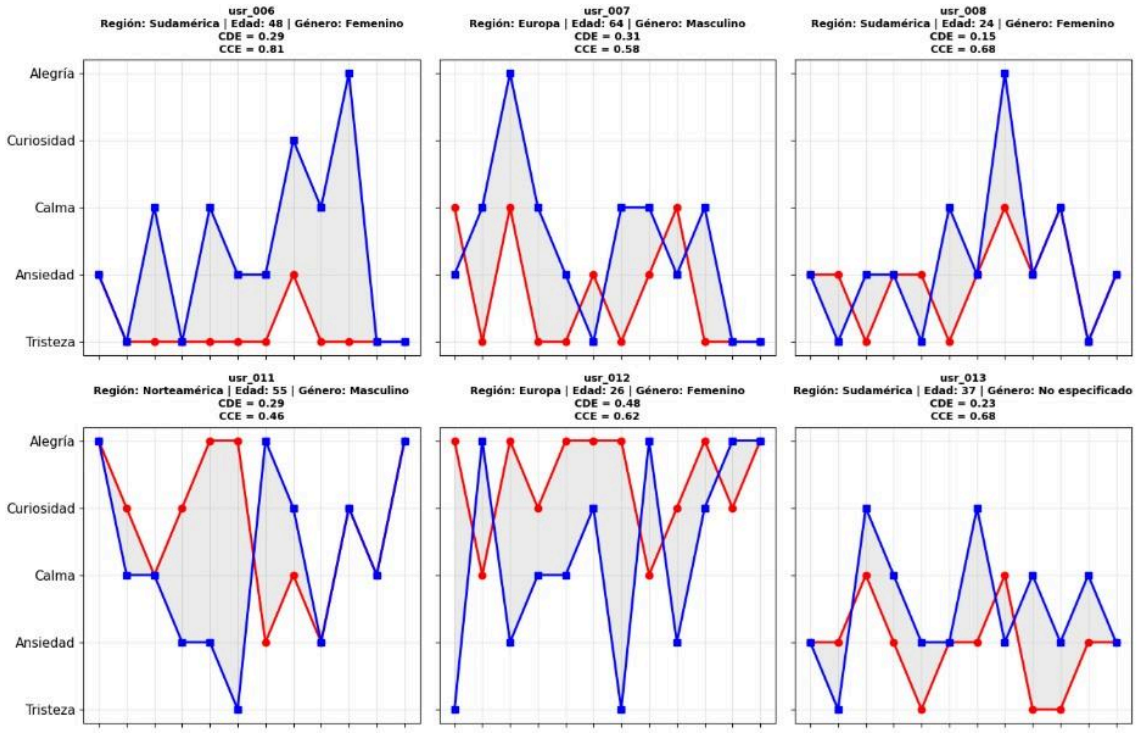

Figura Q. Anexo 9.7. Evolución emocional por usuario (líneas CCE y CDE)

Fuente: elaboración propia (visualización completa de 5x5 disponible en el notebook).

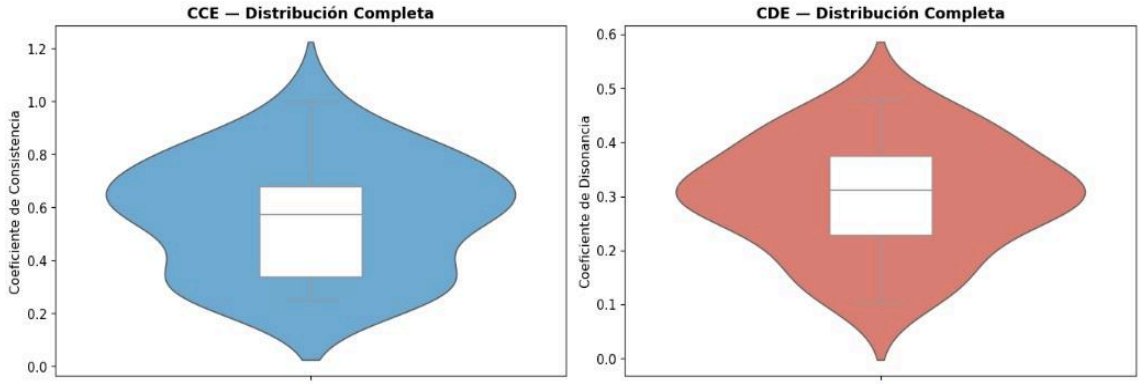

Figura R. Anexo 9.7. Distribución global de coeficientes emocionales (CCE y CDE)

Fuente: elaboración propia.

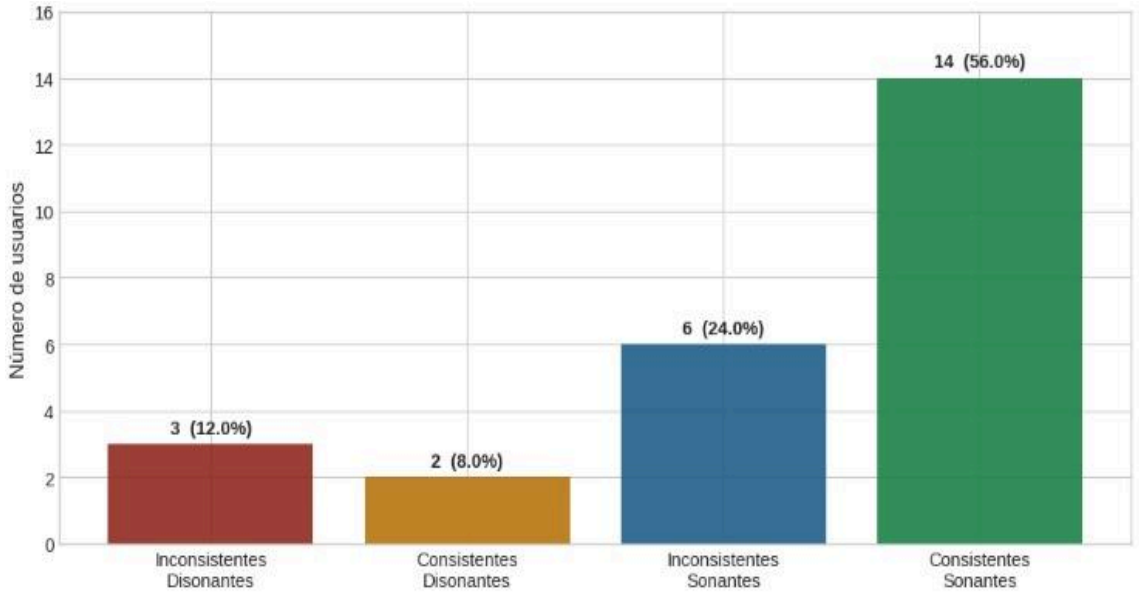

Figura S. Anexo 9.7. Distribución de los usuarios por cuadrante emocional

Fuente: elaboración propia.

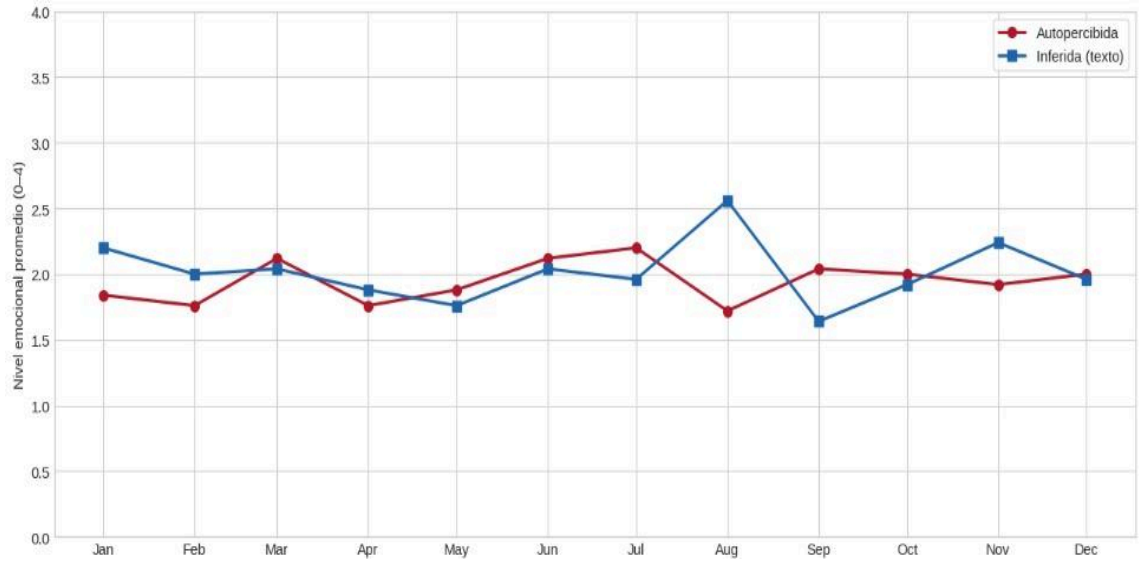

Figura T. Anexo 9.7. Evolución emocional promedio (autopercibida vs inferida)

Fuente: elaboración propia.

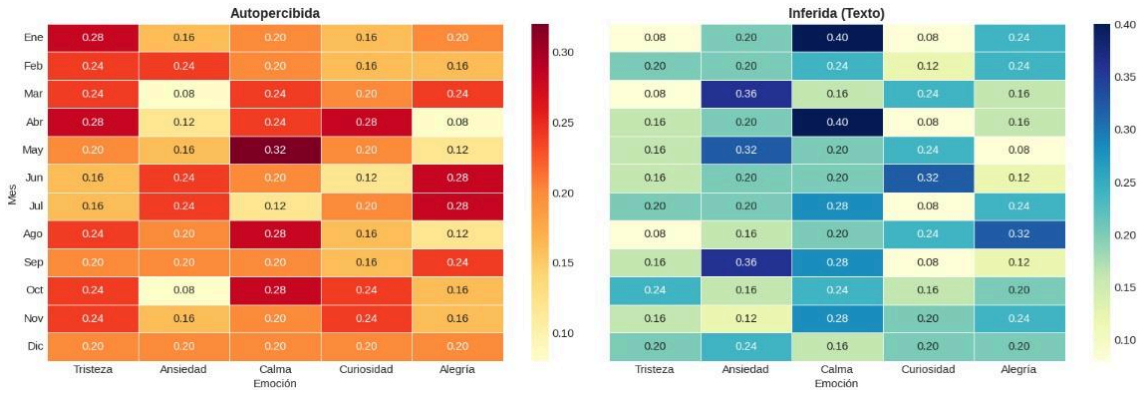

Figura U. Anexo 9.7. Distribución emocional mensual (autopercibida vs inferida)

Fuente: elaboración propia.

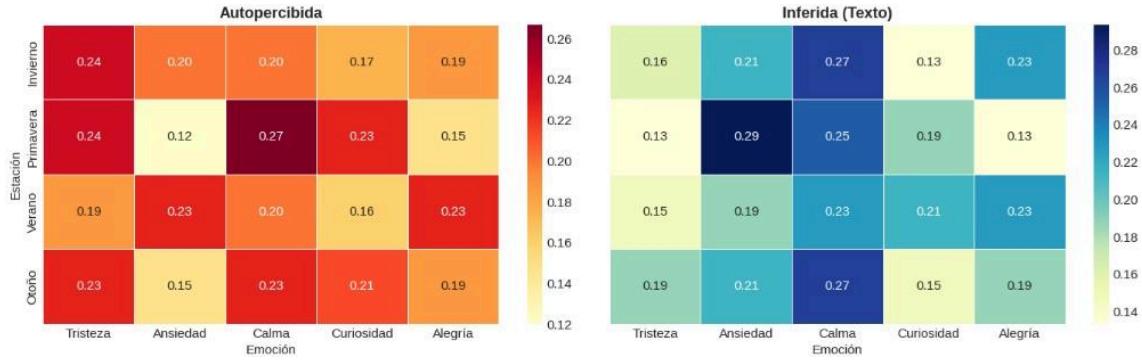

Figura V. Anexo 9.7. Distribución emocional estacional (autopercibida vs inferida)

Fuente: elaboración propia.

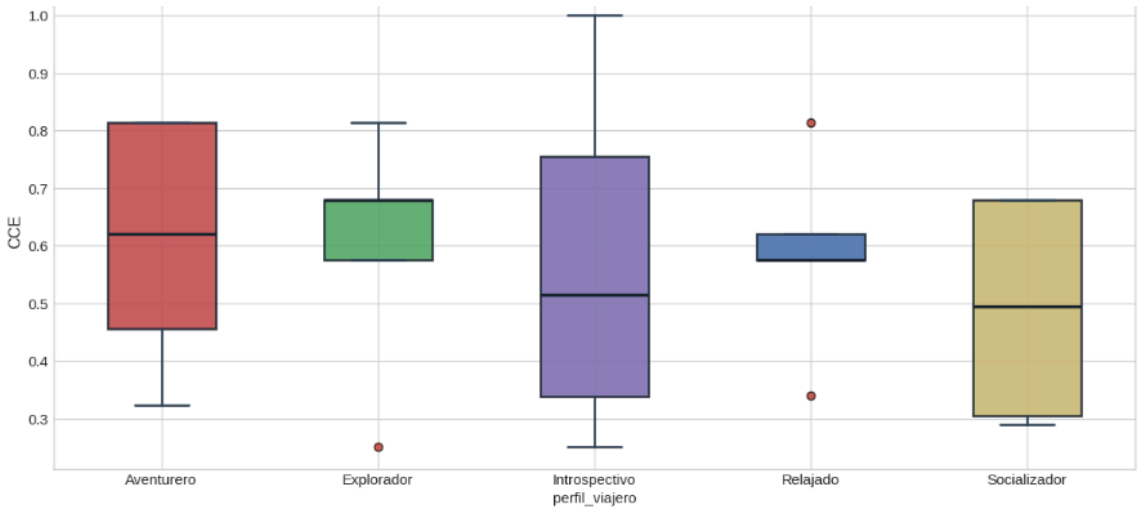

Figura W. Anexo 9.7. Distribución de CCE por tipo de perfil viajero

Fuente: elaboración propia.

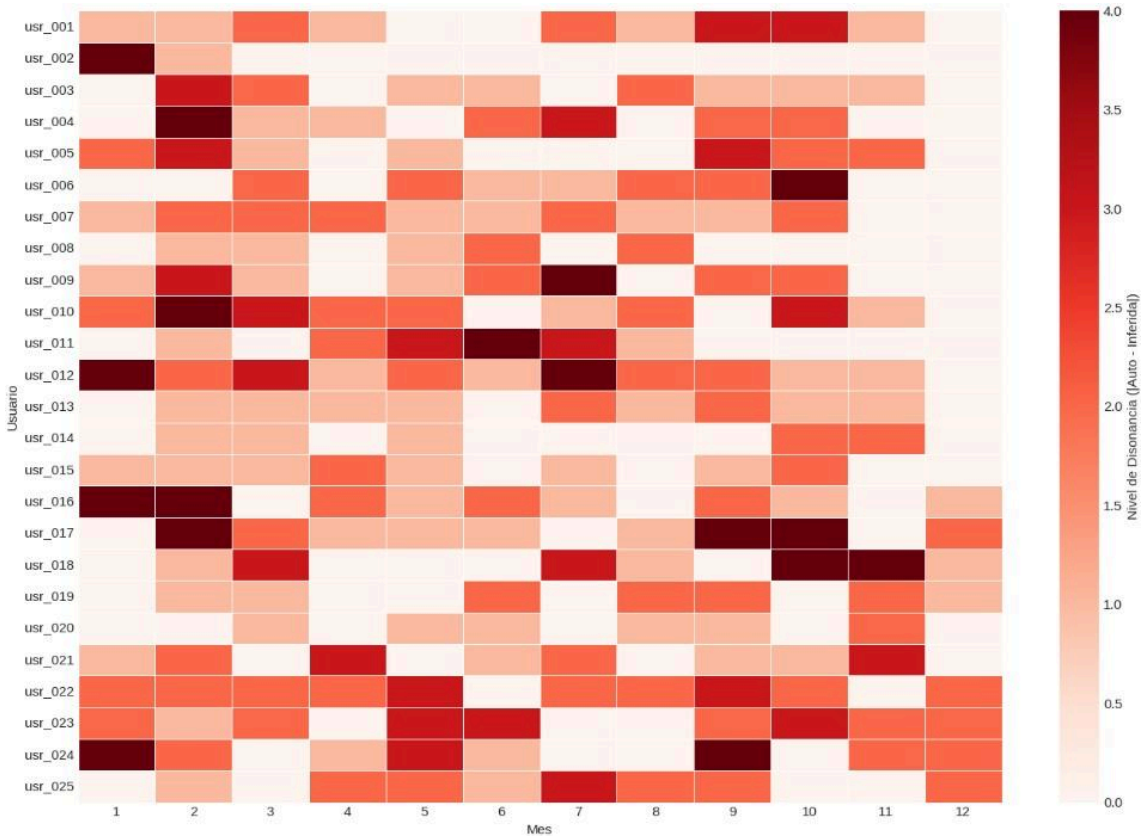

Figura X. Anexo 9.7. Matriz de disonancia emocional individual

Fuente: elaboración propia.

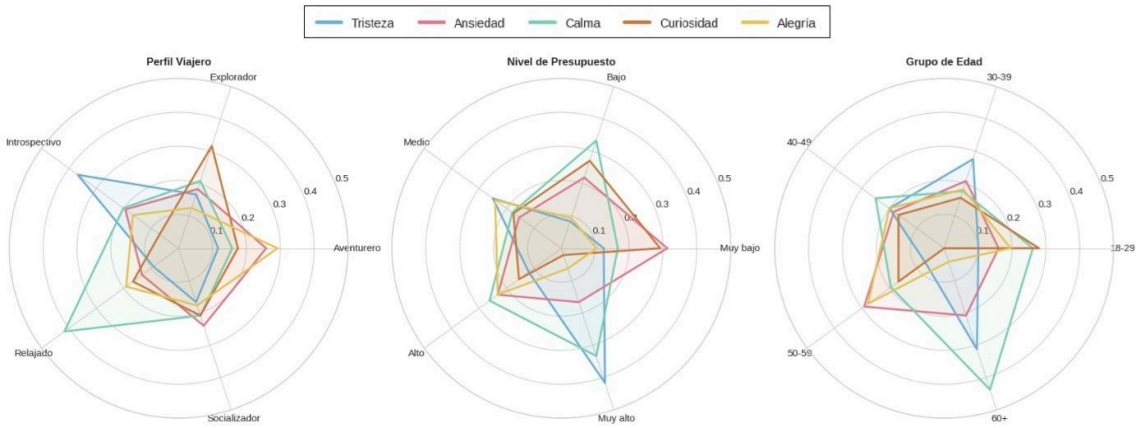

Figura Y. Anexo 9.7. Comparativa emocional según variables demográficas

Fuente: elaboración propia.

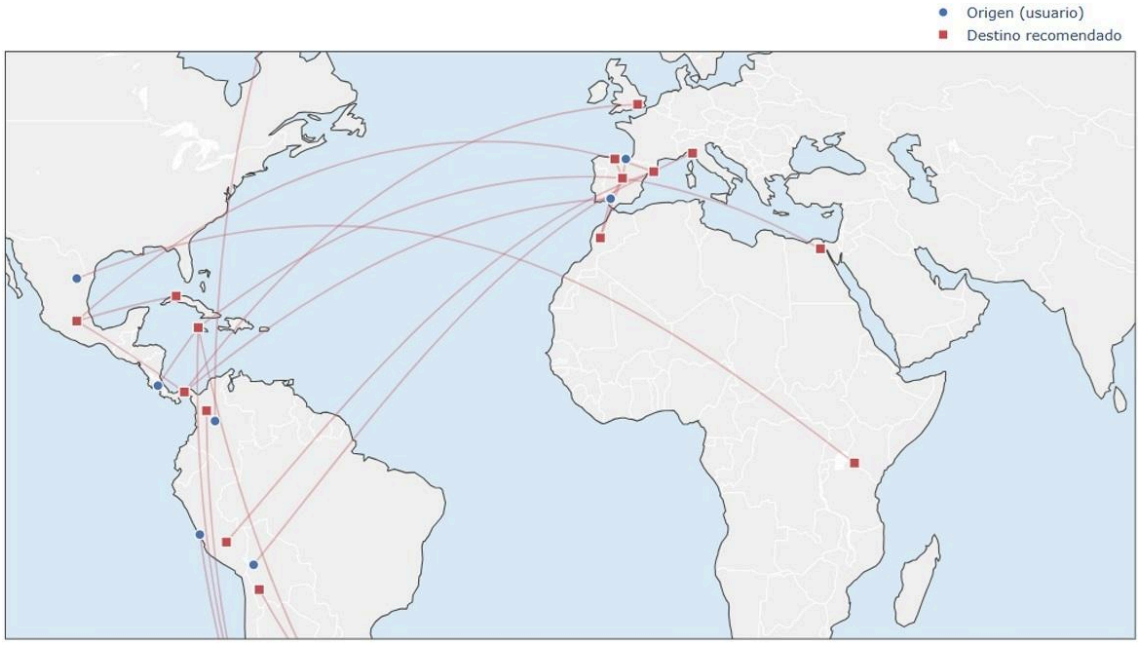

Figura Z. Anexo 9.7. Flujos geográficos de recomendación

Fuente: elaboración propia.

[PÁGINA INTENCIONADAMENTE EN BLANCO]