APLICACIÓN DE PROCESAMIENTO DE LENGUAJE NATURAL (PLN) PARA IDENTIFICAR RIESGOS OPERACIONALES EN EL SECTOR SALUD

Alejandro Mejía Delgado
Código 8975175

Proyecto aplicado para optar al título de
Magíster en Ciencia de Datos

Director
Leonairo Pencue-Fierro

FACULTAD DE INGENIERÍA Y CIENCIAS
MAESTRÍA EN CIENCIA DE DATOS
SANTIAGO DE CALI, MAYO DE 2026

TABLA DE CONTENIDO

LISTA DE FIGURAS

LISTA DE TABLAS

LISTA DE ANEXOS

Anexo 1: Cronograma del proyecto.xlsx

Anexo 2: Notebook del proyecto: Proyecto_PNL_PQRD.ipynb

Anexo 3: Metodología etiquetado manual.docx

Anexo 4: muestra_etiquetado_v1.xlsx

INTRODUCCIÓN

En las empresas colombianas de prestación de servicios de salud, los errores operacionales, deficiencias en procesos administrativos, logísticos, y de prestación de servicios generan inconformidades y afectaciones en la salud de los usuarios.

En el año 2024, la Superintendencia Nacional de Salud recibió un total de 1.728.069 PQRD [1] referentes a distintos factores relacionados con la prestación del servicio de salud. El alto flujo de insatisfacciones puede constituir un indicador relevante dentro del deterioro de la gestión y atención al usuario.

Esto sugiere, que existe un limitado análisis de las PQRD, y que las soluciones planteadas por las entidades competentes no han sido eficaces, seguramente porque resuelven las situaciones como “casos aislados” y no controlan la causa de raíz que puede estar generando las fallas manifiestas por los usuarios.

El presente proyecto realiza el análisis de las descripciones textuales y categóricas asociadas a las PQRD interpuestas ante la Superintendencia de Salud de Colombia, aplicando técnicas de PLN y aprendizaje automático de manera que se pueda identificar, clasificar y asociar riesgos operacionales con las etapas del proceso de atención en el sistema de salud y así generar información que contribuya a mejorar la prestación de servicio en este sistema de salud.

Para su desarrollo se tomó como insumo la Base de Datos de PQRD publicada en el portal de Datos Abiertos de la página Web de la Superintendencia de Salud [2]. El proyecto se estructuró bajo la metodología CRISP-DM [3] para las etapas de entendimiento del contexto del sector salud, comprensión de los datos, preparación de los datos, modelado, evaluación y despliegue del modelo.

Como resultado de esta metodología, se consolidó una base de datos depurada, se construyó un corpus textual para modelado, se elaboró una muestra etiquetada de riesgos operacionales, se entrenaron y evaluaron diferentes modelos de clasificación supervisada, y se seleccionó un modelo para aplicar sobre la base depurada. Adicionalmente, los riesgos clasificados se relacionaron con procesos de la prestación de servicios de salud, generando insumos para el análisis de patrones, la priorización de procesos críticos y el diseño posterior de visualizaciones.

Se espera que las recomendaciones derivadas del análisis contribuyan a orientar a las entidades a priorizar la causa raíz del problema e implementar controles efectivos para prevenir la materialización del riesgo.

1. CONTEXTUALIZACIÓN DEL PROYECTO

1.1. Definición del problema

1.1.1. Planteamiento del problema

En las empresas colombianas de prestación de servicios de salud, los errores operacionales en procesos administrativos, logísticos y de prestación de servicios pueden generar inconformidades y afectaciones en la salud de los usuarios ocasionadas por demoras en la atención, negación de servicios, tratamientos no pertinentes, inoportunidad en la asignación y cancelación de citas. Estas fallas afectan la calidad del servicio de salud y la seguridad de los pacientes al no ser detectadas a tiempo para ser gestionadas. Según datos de la Superintendencia Nacional de Salud, durante el año 2024 se recibieron un total de 1.728.069 de peticiones, quejas, reclamos y denuncias (PQRD) [1] en el sistema de salud, siendo la negación en la asignación de citas o consultas, falta de oportunidad en las citas o consultas, negación para la entrega de tecnologías en salud, servicios autorizados y falta de oportunidad en la atención de servicios de salud, las más recurrentes. La coyuntura actual, derivada de la deficiente prestación de los servicios de salud en el país, motivó la intervención administrativa de algunas EPS por parte de la Superintendencia Nacional de Salud; uno de los elementos bajo los cuales se justificaron estas intervenciones es el deterioro de los indicadores de atención al usuario, particularmente el alto volumen de PQRD reportados por los usuarios. Dos de las entidades con mayor número de afiliados en el país fueron intervenidas por la Superintendencia Nacional de Salud. La primera, según la Resolución No. 2024160000003002-6 de 2024 [4], pasó de una tasa de 126,43 a 263,61 reclamos por cada 10.000 afiliados entre 2017 y diciembre de 2023. La segunda, de acuerdo con la Resolución No. 2024160000003012-6 de 2024 [5], registró en 2023 un total de 185.634 reclamos y una tasa de 321,25 por cada 10.000 afiliados. Estas cifras superan el promedio nacional y evidencian un deterioro en la capacidad de atención y gestión, comprometiendo la sostenibilidad operativa de dichas entidades. Estas problemáticas suelen estar asociadas a fallas recurrentes en procesos como autorizaciones, gestión de citas, atención al usuario, entrega de medicamentos, prestación de servicios y continuidad de la atención. En este sentido, las PQRD constituyen una fuente relevante de información para aproximar la identificación de riesgos operacionales, en la medida en que permiten observar patrones de inconformidad asociados a procesos internos, personas, sistemas, tecnología, infraestructura o factores externos. No obstante, el análisis limitado de estas fuentes puede ocasionar que los reportes sean tratados como casos aislados, sin identificar tendencias, recurrencias o posibles causas comunes. El problema radica en la ausencia de soluciones basadas en ciencia de datos que permitan transformar de manera sistemática el volumen de PQRD en información útil para la toma de decisiones. En este contexto, el PLN permite convertir descripciones textuales y categóricas asociadas a las PQRD en representaciones analíticas que pueden ser utilizadas por modelos supervisados de clasificación. A partir de este enfoque, es posible identificar categorías de riesgo operacional, analizar patrones recurrentes y relacionar los resultados con procesos de prestación de servicios de salud. Por lo anterior, este proyecto aborda la necesidad de aplicar técnicas PLN y aprendizaje automático para clasificar las PQRD en categorías de riesgo operacional, con el fin de generar insumos que apoyen la gestión preventiva y la priorización de procesos críticos.

1.1.2. Formulación del problema

¿Cómo puede el procesamiento de lenguaje natural (PLN), por medio de la aplicación y evaluación comparativa de modelos supervisados de clasificación de texto, permitir identificar, clasificar y analizar riesgos operacionales a partir de las descripciones textuales y categóricas asociadas a las PQRD consolidadas por la Superintendencia Nacional de Salud? Además, ¿cómo pueden los resultados de estos modelos ser utilizados para asociar los riesgos identificados con procesos de la prestación de servicios de salud y formular recomendaciones que contribuyan a prevenir la recurrencia de fallas operacionales y mejorar la calidad de la atención en el sector salud colombiano?

1.1.3. Sistematización

a. ¿Cómo se pueden explorar depurar y procesar las descripciones textuales y categóricas asociadas a las PQRD, utilizando técnicas de procesamiento de lenguaje natural como normalización, tokenización, lematización, eliminación de stopwords, normalización léxica y representación textual mediante TF-IDF, para construir un corpus analítico útil en la identificación de riesgos operacionales?

b. ¿Qué tipos de riesgos operacionales se pueden identificar y clasificar a partir del contenido textual asociado a las PQRD, teniendo en cuenta su naturaleza, frecuencia y relación con posibles fallas en los procesos de prestación de servicios de salud?

c. ¿De qué forma pueden asociarse los riesgos operacionales detectados mediante modelos de PLN con las distintas etapas del proceso de prestación de servicios de salud, para lograr una interpretación operativa e integral de los hallazgos identificados?

d. ¿Qué modelos de clasificación supervisada resultan más adecuados para clasificar las PQRD en categorías de riesgo operacional, considerando su desempeño predictivo, interpretabilidad, estabilidad, costo computacional y capacidad de aplicación sobre grandes volúmenes de información?

e. ¿Qué recomendaciones operativas y estratégicas pueden derivarse del análisis automatizado de las PQRD y la clasificación de riesgos operacionales, considerando los patrones detectados por los modelos de PLN, su relación con los procesos de la prestación de los servicios de salud y cómo estas pueden contribuir a la mejora de la calidad del servicio y la prevención de fallos recurrentes?

f. ¿Qué recomendaciones operativas y estratégicas pueden derivarse del análisis automatizado de las PQRD y la clasificación de riesgos operacionales, considerando los patrones detectados por los modelos de PLN, su relación con los procesos de la prestación de los servicios de salud y cómo estas pueden contribuir a la mejora de la calidad del servicio y la prevención de fallos recurrentes?

1.2. Objetivos del proyecto

1.2.1. Objetivo General

Aplicar técnicas basadas en procesamiento de lenguaje natural (PLN) para analizar las descripciones textuales y categóricas asociadas a las PQRD, con el fin de identificar, clasificar y asociar riesgos operacionales con las etapas del proceso de atención en el sistema de salud colombiano, evaluando diferentes enfoques de modelado y generando insumos que fortalezcan la gestión preventiva de dichos riesgos y contribuyan a la mejora de la calidad en la prestación del servicio.

1.2.2. Objetivos Específicos

a. Desarrollar una metodología de procesamiento de datos para explorar depurar y preparar las descripciones textuales y categóricas asociadas a las PQRD del sistema de salud colombiano, aplicando técnicas de PLN que permitan comprender su estructura y extraer información relevante para la detección de riesgos operacionales.

b. Comparar y evaluar diferentes modelos de clasificación de texto basados en PLN, incluyendo modelos clásicos de aprendizaje automático y enfoques modernos de representación textual, con el objetivo de seleccionar el modelo más adecuado según desempeño predictivo, interpretabilidad, estabilidad y viabilidad computacional.

c. Evaluar el desempeño del modelo de clasificación de riesgos mediante métricas estadísticas que validen su precisión, cobertura, estabilidad y utilidad para la gestión preventiva.

d. Identificar y clasificar los riesgos operacionales denotados en las PQRD, según su naturaleza, frecuencia y relación con posibles fallos en la prestación de servicios de salud.

e. Asociar los riesgos operacionales identificados con las distintas etapas del proceso de atención en salud, permitiendo una interpretación operativa y sistémica de los hallazgos.

f. Formular recomendaciones basadas en los hallazgos del análisis, orientadas a prevenir la recurrencia de los riesgos operacionales y mejorar la calidad en la atención en salud.

1.3. Marco de Referencia

1.3.1. Marco Teórico

El presente proyecto desarrolla un modelo basado en el procesamiento de lenguaje natural para identificar riesgos operacionales en empresas prestadoras de servicios de salud en Colombia, a partir del análisis de datos provenientes de las peticiones, quejas, reclamos y denuncias (PQRD) reportadas ante la Superintendencia Nacional de Salud.

En el Marco Teórico se exponen los fundamentos del riesgo operacional en salud y el concepto de PQRD en el ámbito sanitario. Desde el enfoque de ciencia de datos, se abordan los conceptos fundamentales de PNL, describiendo sus principales técnicas y métricas de desempeño.

1.3.1.1. Fundamentos del riesgo operacional en salud

El Comité de Supervisión Bancaria de Basilea, (Acuerdo Basilea II, 2006), introduce y define formalmente el Riesgo Operativo como "el riesgo de pérdida derivado de procesos internos inadecuados o fallidos, personas y sistemas, o de eventos externos" [6], si bien, la gestión de este riesgo inicialmente se extendió para el sector bancario, en Colombia varios sectores de la economía han incorporado la gestión de este riesgo en procesos estratégicos.

En el sector de la salud en Colombia, las Circulares Externas 45 de 2021 [7] y 51 de 2022 [8], de la Superintendencia de Salud establecen los lineamientos y exige a las Instituciones Prestadoras de Servicios de Salud (IPS) y Entidades Promotoras de Salud (EPS), la implementación del Sistema de Administración de Riesgos y sus Subsistemas. Acorde con las circulares, se define riesgo operacional como la posibilidad de que una entidad presente desviaciones de sus objetivos misionales como consecuencia de deficiencias, inadecuaciones o fallas en los procesos, en el recurso humano, en los sistemas tecnológicos, legales y biomédicos, en la infraestructura, ya sea por causa interna o por la ocurrencia de acontecimientos externos, entre otros, los cuales pueden tener impactos negativos en el cumplimiento de los objetivos institucionales. La materialización de estos eventos afecta la prestación de los servicios de salud y la satisfacción de los usuarios [9].

1.3.1.2. Peticiones, Quejas, Reclamos y Denuncias (PQRD) como fuente de datos

En el sector salud, las peticiones, quejas, reclamos y denuncias, son un mecanismo formal que le permite a los usuarios manifestar inquietudes, solicitudes o inconformidades por la prestación de servicios de salud [10], las definiciones de cada uno según la Superintendencia de Salud son:

a. Petición

Solicitud a través de la cual una persona por motivos de interés general o particular solicita la intervención de la entidad para la resolución de una situación, la prestación de un servicio, la información o requerimiento de copia de documentos, entre otros.

b. Queja

Manifestación de una persona, a través de la cual expresa inconformidad con el actuar de un funcionario de la entidad.

c. Reclamo

Solicitud a través de la cual los usuarios dan a conocer su insatisfacción con la prestación del servicio por parte de un actor del Sector Salud o solicita el reconocimiento del derecho fundamental a la salud.

d. Denuncia

Manifestación de una persona a través de la cual pone en conocimiento hechos que revisten las características de un delito, y, por lo tanto, debe investigarse de oficio por parte de la Fiscalía General de la Nación.

El Portal de datos abiertos de la Superintendencia de Salud, agrupa las PQRD reportadas por los usuarios del sistema de salud colombiano. Esta fuente constituye un insumo relevante para los fines de este proyecto dado que contiene variables estructuradas y descripciones textuales asociadas a los motivos de inconformidad, las cuales permiten analizar patrones recurrentes, clasificar riesgos operacionales y relacionarlos con procesos de la prestación de servicios de salud mediante técnicas de procesamiento de lenguaje natural.

1.3.1.3. Ciencia de datos y análisis de texto aplicado a salud

a. Procesamiento de lenguaje natural (PLN)

IBM describe el procesamiento de lenguaje natural (PLN) como un subcampo de la informática y la inteligencia artificial que emplea el aprendizaje automático para que las computadoras comprendan y se comuniquen con el lenguaje humano [11].

El Procesamiento de lenguaje natural integra dos elementos: el aprendizaje automático que es una tecnología que entrena a una computadora con datos de muestra, en el contexto del PLN enseña a las aplicaciones para reconocer elementos complejos del lenguaje natural y estructuras gramaticales irregulares facilitando su comprensión, y el aprendizaje profundo es un campo específico del aprendizaje automático que enseña a las computadoras a aprender y pensar como humanos, facilitando reconocer, clasificar y correlacionar patrones complejos en los datos de entrada [12].

El PLN tiene diversas aplicaciones, entre las más relevantes se destaca la detección de temas, extracción de frases clave, clasificar y analizar datos de texto [13]. En este contexto, permite procesar grandes volúmenes de información textual, identificar patrones lingüísticos, representar documentos mediante variables numéricas y entrenar modelos capaces de clasificar contenidos de acuerdo con categorías previamente definidas.

b. Técnicas para el análisis de texto

Las técnicas para el análisis de texto son métodos aplicados dentro del procesamiento de lenguaje natural (PLN) que permiten extraer, transformar y analizar información contenida en textos, además permiten identificar patrones, clasificar contenidos y generar conocimiento útil para la toma de decisiones. En este proyecto, se emplean principalmente técnicas de normalización textual, tokenización, lematización, eliminación de stopwords, normalización léxica, representación mediante TF-IDF y clasificación supervisada de texto.

c. Tokenización

Técnica que divide un texto en unidades más pequeñas que pueden ser palabras, frases o caracteres individuales. La tokenización es el primer paso en el PLN, porque facilita el análisis y procesamiento posterior del texto [14].

d. Lematización

Proceso para reducir una palabra a su forma base o "lema", que es su forma canónica en el diccionario. Este proceso considera el contexto y las reglas gramaticales para convertir diferentes formas de una palabra en su forma raíz [14].

En este proyecto, esta técnica permite disminuir la variabilidad del lenguaje y mejorar la consistencia del corpus textual.

e. Eliminación de stopwords

La eliminación de stop words es una técnica que implica la eliminación de palabras muy frecuentes en un idioma que no aportan un significado relevante al análisis del texto [14].

En el contexto de este proyecto, esta técnica se utiliza para reducir ruido en el corpus; sin embargo, se debe aplicar con criterio, evitando eliminar términos que puedan tener importancia semántica en las PQRD, especialmente palabras asociadas a negaciones o ausencia de prestación del servicio.

f. Análisis de sentimiento

Técnica que permite determinar emociones, opiniones y actitudes expresadas en un texto. El objetivo es clasificar las expresiones en categorías como positivas, negativas o neutras, y en algunos casos, identificar emociones más específicas como alegría, tristeza o enojo [14].

Aunque esta técnica es relevante en el análisis de quejas y experiencia del usuario, en el presente proyecto se considera como referencia conceptual, dado que el enfoque principal se orienta a la clasificación supervisada de riesgos operacionales.

g. Detección de tópicos

Consiste en el procesamiento automático no supervisado de textos con el objetivo de identificar los asuntos o motivos sobre los que dichos textos tratan [15].

h. TF-IDF (Term Frequency-Inverse Document Frequency)

Es una técnica que permite evaluar la importancia de una palabra en un documento. Este método ayuda a identificar qué palabras son más relevantes en un texto específico, permitiendo una mejor comprensión y análisis de datos textuales [14].

En este proyecto, TF-IDF se utiliza como representación numérica del corpus textual para entrenar modelos supervisados de clasificación de riesgos operacionales.

i. Word2Vec

Es una técnica que utiliza un modelo de red neuronal para aprender asociaciones de palabras a partir de un gran corpus de texto, representando cada palabra con un vector que captura su significado semántico [14].

Esta técnica se considera dentro del marco conceptual del procesamiento de lenguaje natural, aunque el modelo seleccionado del proyecto se basa en una representación TF-IDF por su desempeño, interpretabilidad y viabilidad computacional.

j. BERT (Bidirectional Encoder Representations from Transformers).

Es un modelo de lenguaje desarrollado por Google que permite captar el contexto bidireccional en una oración, mejorando la comprensión contextual en diversas tareas de PLN [14].

En el contexto de este proyecto, los modelos basados en transformers se consideran como enfoques modernos de referencia para el análisis textual; sin embargo, su aplicación debe evaluarse considerando el costo computacional, la disponibilidad de datos etiquetados y la escalabilidad requerida para procesar grandes volúmenes de PQRD.

k. Clasificación supervisada de texto.

La clasificación supervisada de texto es una técnica de aprendizaje automático que permite asignar documentos o registros textuales a categorías previamente definidas, a partir de ejemplos etiquetados. En este proyecto, esta técnica se utiliza para clasificar las PQRD en categorías de riesgo operacional R1–R7.

Para ello, se construye una muestra etiquetada manualmente, se transforma el texto en representaciones numéricas y se entrenan modelos capaces de aprender patrones asociados a cada categoría de riesgo. La evaluación de estos modelos se realiza mediante métricas de clasificación, las cuales se presentan en el siguiente numeral.

1.3.1.4. Métricas de evaluación de modelos de clasificación.

En el contexto de este proyecto, se aplican algoritmos de aprendizaje supervisado para entrenar modelos de clasificación que permitan identificar tipos de riesgos operacionales en las PQRD. La evaluación de los modelos se realizó mediante métricas de desempeño como accuracy, precisión macro, recall macro, F1 macro, coeficiente de correlación de Matthews (MCC) y log loss. Estas métricas permitieron comparar los modelos desde diferentes perspectivas: desempeño global, equilibrio entre clases, capacidad de generalización y calidad probabilística [16].

La matriz de confusión constituye la base para calcular varias métricas de clasificación. En ella se comparan las clases reales frente a las clases predichas por el modelo.

a. Accuracy

La exactitud o accuracy mide la proporción total de predicciones correctas frente al total de observaciones evaluadas. Esta métrica permite obtener una visión general del desempeño del modelo, dado que compara los aciertos del clasificador con el total de registros analizados.

La fórmula de accuracy se expresa de la siguiente manera:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Donde:

TP: Representa los verdaderos positivos.

TN: Representa los verdaderos negativos.

FP: Representa los falsos positivos.

FN: Representa los falsos negativos.

En esta métrica, el numerador corresponde al total de clasificaciones correctas realizadas por el modelo, mientras que el denominador representa el total de observaciones evaluadas. Por tanto, un valor alto de accuracy indica que el modelo clasifica correctamente una proporción elevada de registros.

b. Precisión

La precisión mide la proporción de registros que fueron clasificados por el modelo como pertenecientes a una clase determinada y que realmente corresponden a dicha clase. Esta métrica permite evaluar la confiabilidad de las predicciones realizadas por el modelo para cada categoría.

La fórmula de precisión se expresa de la siguiente manera:

$$Precision = \frac{TP}{TP + FP}$$

Donde:

TP: Representa los verdaderos positivos.

FP: Representa los falsos positivos.

En esta métrica, el numerador corresponde al total de registros correctamente clasificados por el modelo dentro de una clase específica, mientras que el denominador representa el total de registros que el modelo predijo como pertenecientes a esa clase. Por tanto, un valor alto de precisión indica que el modelo asigna pocos registros incorrectos a la categoría evaluada.

c. Recall

El recall, también denominado sensibilidad o exhaustividad, mide la proporción de registros que pertenecen realmente a una clase y que fueron correctamente identificados por el modelo. Esta métrica

permite evaluar la capacidad del modelo para recuperar los casos reales de cada categoría.

$$Recall = \frac{TP}{TP + FN}$$

Donde:

TP: Representa los verdaderos positivos.

FN: Representa los falsos negativos.

En esta métrica, el numerador corresponde al total de registros correctamente clasificados por el modelo dentro de una clase específica, mientras que el denominador representa el total de registros que realmente pertenecen a esa clase. Por tanto, un valor alto de recall indica que el modelo identifica adecuadamente una proporción elevada de los casos reales de la categoría evaluada.

d. F1-score

El F1-score mide el equilibrio entre la precisión y el recall de un modelo de clasificación. Esta métrica corresponde a la media armónica entre ambas medidas, por lo que resulta útil cuando se requiere evaluar simultáneamente la confiabilidad de las predicciones y la capacidad del modelo para recuperar los casos reales de cada clase.

La fórmula de F1-score se expresa de la siguiente manera:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Donde:

Precisión: Representa la proporción de registros predichos como pertenecientes a una clase que realmente corresponden a dicha clase.

Recall: Representa la proporción de registros reales de una clase que fueron correctamente identificados por el modelo.

En esta métrica, el numerador corresponde al producto entre precisión y recall multiplicado por dos, mientras que el denominador corresponde a la suma de ambas métricas. Por tanto, un valor alto de F1-score indica que el modelo mantiene un equilibrio adecuado entre clasificar correctamente los registros asignados a una clase y recuperar los casos reales de dicha categoría.

e. Métricas macro

Las métricas macro permiten evaluar el desempeño promedio del modelo entre todas las clases, sin ponderar por el número de registros de cada categoría. En problemas de clasificación multiclase con clases desbalanceadas, estas métricas son relevantes porque evitan que la evaluación esté dominada por las clases con mayor número de observaciones.

La fórmula de precisión macro se expresa de la siguiente manera:

$$Precision_{macro} = \left(\frac{1}{K}\right) \times \sum_{k=1}^{k} Precisionk$$

La fórmula de recall macro se expresa de la siguiente manera:

$$Recall_{macro} = \left(\frac{1}{K}\right) \times \sum_{k=1}^{k} Recall_k$$

La fórmula de F1 macro se expresa de la siguiente manera:

$$F1_{macro} = \left(\frac{1}{K}\right) \times \sum_{K=1}^{K} f1K$$

Donde:

K: Representa el número total de clases evaluadas.

k: Representa cada una de las clases del problema de clasificación.

Precision_k: Representa la precisión calculada para la clase k.

Recall_k: Representa el recall calculado para la clase k.

F1_k: Representa el F1-score calculado para la clase k.

En estas métricas, el numerador corresponde a la suma de los resultados obtenidos en cada clase, mientras que el denominador representa el número total de clases evaluadas. Por tanto, un valor alto en las métricas macro indica que el modelo mantiene un desempeño equilibrado entre las diferentes categorías [16].

f. Coeficiente de correlación de Matthews (MCC)

El coeficiente de correlación de Matthews, conocido como MCC, es una métrica que evalúa la calidad global de un modelo de clasificación considerando simultáneamente los verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos. Esta métrica es especialmente útil en problemas con clases desbalanceadas, porque no depende únicamente de la proporción total de aciertos [17].

La fórmula del MCC se expresa de la siguiente manera:

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

Donde:

TP: Representa los verdaderos positivos.

TN: Representa los verdaderos negativos.

FP: Representa los falsos positivos.

FN: Representa los falsos negativos.

En esta métrica, el numerador corresponde a la diferencia entre las clasificaciones correctas cruzadas y las clasificaciones incorrectas cruzadas, mientras que el denominador normaliza el resultado considerando todos los componentes de la matriz de confusión. Por tanto, un valor alto de MCC indica que el modelo mantiene una clasificación global consistente entre aciertos y errores.

g. Log loss

El log loss es una métrica probabilística que evalúa la calidad de las probabilidades asignadas por el modelo a cada clase. A diferencia de las métricas basadas únicamente en la clase predicha, el log loss penaliza con mayor severidad las predicciones incorrectas cuando el modelo asigna una probabilidad alta a una clase equivocada [18].

La fórmula de log loss en clasificación multiclase se expresa de la siguiente manera:

$$LogLoss = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} y_{i,k}\log(p_{i,k})$$

Donde:

N: Representa el número total de observaciones evaluadas.

K: Representa el número total de clases.

i: Representa cada observación evaluada.

k: Representa cada una de las clases del problema de clasificación.

yi,k: Representa la clase real de la observación i para la clase k. Toma el valor de 1 si la observación pertenece realmente a la clase k y 0 en caso contrario.

pi,k: Representa la probabilidad predicha por el modelo de que la observación i pertenezca a la clase k.

En esta métrica, la sumatoria compara las probabilidades predichas por el modelo frente a la clase real observada. Por tanto, valores más bajos de log loss indican una mejor calidad probabilística del modelo.

1.3.2. Antecedentes

a. Análisis con Machine Learning de Peticiones (PQRS) del Servicio Nacional de Aprendizaje

El trabajo de Ayala, Salamanca, Durán y González (2023) analiza las PQRS del SENA mediante regresión logística para predecir incumplimientos normativos, utilizando variables categóricas y la metodología SEMMA [19]. Se relaciona con este proyecto en el uso de fuente de datos de PQRS ; sin embargo, se diferencia en los modelos empleados, en el caso del presente proyecto se aplicará NLP y en el trabajo citado se utilizó un modelo de Regresión logística.

b. Asistente Virtual Académico utilizando Tecnologías Cognitivas de PLN

El trabajo de Manjarrés-Betancur y Echeverry-Torres (2020) desarrolla un asistente virtual académico utilizando tecnologías cognitivas basadas en el procesamiento de lenguaje natural para mejorar la atención en procesos educativos [20]. Este estudio se relaciona con este proyecto porque ambos aplican técnicas de NLP sobre texto no estructurado; sin embargo, difieren en el enfoque, en el caso del presente proyecto se busca identificar riesgos operacionales en salud, mientras que el objetivo de dicho trabajo es mejorar la experiencia del usuario en el ámbito académico.

c. Clasificación de Quejas de los Clientes Empleando Procesamiento de Lenguaje Natural

El trabajo de Flores Poma (2023) presenta una revisión sistemática sobre la clasificación automática de quejas utilizando procesamiento de lenguaje natural (NLP), destacando algoritmos como SVM, BERT y LSTM [21]. Este estudio está relacionado con este proyecto en el uso de NLP para clasificar textos de queja; sin embargo, se diferencia en el enfoque. En el presente proyecto se aborda esta clasificación en el sector salud en un contexto de riesgo operacional, mientras que el estudio mencionado se centra en las quejas y su impacto en la satisfacción del cliente en diversos sectores.

d. Computational Methods for Information Processing from NLP - A Systematic Review

El trabajo de Blandón Andrade et al. (2025) revisa métodos computacionales para procesar quejas en lenguaje natural, usando técnicas híbridas entre NLP y aprendizaje automático [22]. Se relaciona con este proyecto porque también extrae información de quejas mediante NLP. Sin embargo, el presente proyecto se enfoca en el sector salud colombiano, mientras que este estudio es una síntesis bibliográfica multisectorial.

e. PLN Impulsado por Inteligencia Artificial para Empresas de Servicios Públicos

Navia (2025) presenta un módulo de NLP con IA para clasificar registros textuales en empresas de servicios públicos [23]. Aunque ambos proyectos usan NLP para estructurar información y generar alertas, en este proyecto el enfoque se centra en el análisis las PQRD del sector salud colombiano. Este trabajo clasifica reportes técnicos del sector energético.

f. Prototipo de herramienta para la mejora en los procesos de designación de PQRSD de la Alcaldía de Bucaramanga

Gómez Bueno y Gómez Cárdenas (2023) desarrollaron una herramienta que usa modelos de clasificación y procesamiento de lenguaje natural para mejorar la asignación de PQRSD en la Alcaldía de Bucaramanga [24].

Utilizan técnicas de NLP y modelos predictivos sobre texto no estructurado para optimizar la gestión institucional. Aunque su enfoque es la clasificación eficiente de solicitudes administrativas municipales, este proyecto se centra en identificar riesgos operacionales en salud.

g. Comparación de metodologías PEFT aplicadas a modelos de lenguaje grandes enfocados en generar resúmenes de textos

El trabajo de Flórez Pazos et al. (2024) compara metodologías PEFT aplicadas a modelos de lenguaje grande (LLM) como T5 y GPT-2 para generar resúmenes de texto de forma eficiente y con bajo costo

Computacional [25]. Se relaciona con este proyecto en el uso de NLP y modelos de lenguaje para procesar texto no estructurado; sin embargo, se diferencia en que su enfoque es la optimización del fine-tuning para tareas de resumen, mientras que en el presente proyecto se pretende aplicar NLP para identificar riesgos operacionales en el sector salud.

2. EXPLORACIÓN Y PROCESAMIENTO DE LOS DATOS

2.1 Obtención y carga de los datos

Para la construcción de la base analítica del proyecto, se recolectaron los archivos históricos de PQRD disponibles en el portal de Datos Abiertos de la Superintendencia Nacional de Salud de Colombia. Estos archivos contienen los reportes de peticiones, quejas, reclamos y denuncias interpuestas por los usuarios del sistema de salud colombiano durante el periodo comprendido entre 2017 y 2024.

Los archivos fueron descargados manualmente y alojados en una carpeta de Google Drive definida para el proyecto. Posteriormente, se montó el repositorio en el entorno de Google Colab mediante el módulo drive de la librería google.colab, lo cual permitió habilitar la lectura, validación y procesamiento de los archivos desde este entorno de trabajo.

Para asegurar la integridad del conjunto de datos, se realizó una validación inicial de archivos mediante Python. Este procedimiento permitió listar, verificar y contabilizar los archivos disponibles en el directorio de trabajo. Adicionalmente, se emplearon librerías como pandas, openpyxl, pyarrow y duckdb para la lectura de archivos Excel, transformación a formato Parquet, estructuración de datos y consulta analítica eficiente. Como resultado, se identificaron correctamente 11 archivos con extensión .xlsx, correspondientes a las bases históricas de PQRD utilizadas en el proyecto:

- PQRD_2017.xlsx
- PQRD_2018.xlsx
- PQRD_2019.xlsx
- PQRD_2020.xlsx
- PQRD_2021.xlsx
- PQRD_2022_1.xlsx
- PQRD_2022_2.xlsx
- PQRD_2023_1.xlsx
- PQRD_2023_2.xlsx
- PQRD_2024_1.xlsx
- PQRD_2024_2.xlsx

2.2 Análisis preliminar del volumen, estructura y peso de los archivos

Como parte del diagnóstico inicial del corpus, se evaluó el peso en megabytes (MB) de cada archivo, identificando que los archivos de mayor tamaño corresponden al año 2022 (154,1 MB) y al año 2021 (153,4 MB), seguidos por los reportes de 2020 (131,1 MB), 2024_2 (147,1 MB) y 2023_2 (114,6 MB). En total, la suma de los archivos representa más de 1,3 GB de información, lo cual justificó la utilización de formatos y herramientas eficientes para el almacenamiento, consulta y procesamiento de datos.

2.3 Consolidación y estructura del conjunto de datos

Una vez validados y cargados los archivos, se unificaron y transformaron al formato Parquet, con el objetivo de agilizar las operaciones de lectura, consulta y procesamientos posteriores. Adicionalmente, se creó una base analítica en DuckDB para facilitar la consulta eficiente del conjunto de datos desde Google Colab., como resultado de esta etapa, se obtuvo una base inicial consolidada compuesta por 8.064.736 registros y 34 variables.

La estructura de 34 columnas abarca campos administrativos, temporales, demográficos, geográficos, de entidad vigilada y de clasificación del motivo de la PQRD. Dentro de estos campos se incluyen variables como año, mes, entidad, régimen de afiliación, género, edad, macromotivo, motivo general y motivo específico. Estas variables permitieron caracterizar el conjunto de datos, realizar análisis exploratorios y conservar trazabilidad para las etapas posteriores de preparación textual, modelado y despliegue.

Posteriormente, como parte de la depuración, se construyó la tabla pqrd_clean, excluyendo duplicados identificados al comparar todas las variables excepto el identificador objectid. Como resultado, la base final depurada quedó conformada por 8.039.810 registros, que constituyen el universo utilizado para las fases posteriores del proyecto.

2.4 Revisión de calidad de los datos

La revisión de calidad de los datos tuvo como propósito verificar la integridad, consistencia, completitud y validar la presencia de información sensible dentro del conjunto de datos. Para ello, se aplicaron validaciones para identificar valores faltantes, inconsistencias en variables clave, duplicados y presencia de datos personales.

Inicialmente, se implementó una verificación sobre la existencia de registros completamente duplicados, considerando la totalidad de las columnas de la base. Esta validación no arrojó coincidencias con ocurrencias idénticas, concluyendo que no existían duplicados exactos cuando se incluía el identificador del registro.

No obstante, con el fin de realizar una validación más funcional, se ejecutó una segunda revisión excluyendo la variable objectid, dado que esta corresponde a un identificador único y puede ocultar registros equivalentes en las demás variables. Como resultado de esta depuración, se eliminaron 24.926 registros duplicados funcionales, equivalentes aproximadamente al 0,31 **%** de la base inicial. De esta forma, la base pasó de 8.064.736 a 8.039.810 registros, consolidándose como la base depurada para las siguientes fases del proyecto.

Con el objetivo de identificar datos faltantes, se elaboró una matriz de porcentaje de valores nulos por variable y por año. Esta matriz fue visualizada mediante un mapa de calor, lo que ayudó a detectar campos críticos y analizar la evolución de la completitud de las variables durante el periodo 2017–2024.

Figura 2.1: Matriz de valores faltantes por campo y año (Elaboración propia)

El análisis evidenció que los valores ausentes se concentran principalmente en variables asociadas al afectado y su caracterización. En particular, afec_genero presenta porcentajes elevados de nulos entre 2017 y 2019, con una mejora progresiva a partir de 2020, hasta alcanzar niveles cercanos a 0 % en los años siguientes. De forma similar, las variables de ubicación del afectado (afec_cod_depto y afec_cod_mpio) registran niveles altos de ausencia entre 2018 y 2021, pero desde 2022 muestran una completitud consistente.

Adicionalmente, se observa un faltante moderado y recurrente en la ubicación del peticionario (pet_cod_depto y pet_cod_mpio) a lo largo del periodo 2018–2024, con valores cercanos al 10%–20%, lo que sugiere una ausencia estructural en esa información. En contraste, las variables operativas y de clasificación del caso (campos de entidad y motivos) mantienen una completitud cercana al 100% en todos los años analizados.

Se aplicaron reglas de validación para detectar valores fuera de rango o inconsistentes en variables numéricas y temporales, como resultado se identificó que:

- No se identificaron edades negativas ni superiores a 120 años en la variable afec_edadr.
- Los valores de mes se encuentran correctamente codificados entre 1 y 12.
- La variable periodo solo contiene años válidos dentro del rango 2017–2024.

Adicionalmente, se verificaron los campos que presentan mayor susceptibilidad de contener información sensible de los usuarios, tales como patologia_1, patologia_tipo, motivo_especifico, entre otros. Se generaron las validaciones para detectar posibles patrones de nombres, números de documento, correos electrónicos, teléfonos o direcciones, obteniendo como resultado que no se identificaron registros que incluyeran datos sensibles o identificadores personales.

2.5 Análisis de las variables del conjunto de datos

Esta fase tuvo como objetivo principal conocer la estructura del conjunto de datos, identificar el tipo de cada campo y clasificar las variables según su naturaleza analítica. Este análisis permitió definir posteriormente las transformaciones, validaciones y técnicas de procesamiento más adecuadas para cada variable, diferenciando entre campos de identificación, variables temporales, información del peticionario y afectado, datos de entidad vigilada, clasificación del motivo y variables relacionadas con condiciones clínicas o de severidad.

Con base en los nombres técnicos y su correspondencia con el diccionario oficial publicado por la Superintendencia Nacional de Salud, se elaboró un diccionario de datos, incorporando para cada variable: su nombre técnico, alias o nombre funcional, tipo de dato, descripción y categoría de información.

No. variable	Nombre del campo	Descripción del campo
1	**objectid**	Identificador único del registro de PQRD.
2	**year**	Año calendario en el que se interpone la PQRD.
3	**mes**	Mes calendario en el que se interpone la PQRD (1–12).
4	**periodo**	Clave año–mes de la PQRD en formato YYYYMM.
5	**pqr_canal**	Canal por el cual se interpone la PQRD (presencial, telefónico, web, etc.).
6	**pet_cod_depto**	Código DANE del departamento donde se encuentra el peticionario.
7	**pet_cod_mpio**	Código DANE del municipio donde se encuentra el peticionario.
8	**id_afec**	Identificador anonimizado de la persona afectada por la PQRD.
9	**afec_parentesco**	Relación entre el peticionario y la persona afectada.
10	**afec_genero**	Género de la persona afectada.
11	**afec_edadr**	Rango de edad de la persona afectada.
12	**afec_educ**	Nivel educativo alcanzado por la persona afectada.
13	**afec_regafiliacion**	Régimen de afiliación en salud de la persona afectada (contributivo, subsidiado, etc.).
14	**afec_getnico**	Grupo étnico al que pertenece la persona afectada, si aplica.
15	**afec_pobespecial**	Condición de población de protección especial de la persona afectada, si aplica.
16	**afec_cod_depto**	Código DANE del departamento asociado a la situación que afecta el derecho en salud.
17	**afec_cod_mpio**	Código DANE del municipio asociado a la situación que afecta el derecho en salud.
18	**ent_nombre**	Nombre de la entidad vigilada (EPS, IPS u otra) sobre la cual se interpone la PQRD.
19	**ent_tipovig_sns**	Clasificación de la entidad vigilada según la Superintendencia Nacional de Salud.
20	**ent_cod_sns**	Código asignado por la Superintendencia a la entidad vigilada.
21	**ent_alias_sns**	Nombre corto o alias de la entidad vigilada.
22	**ent_cod_depto**	Código DANE del departamento donde se ubica la sede de la entidad vigilada.
23	**ent_cod_mpio**	Código DANE del municipio donde se ubica la sede de la entidad vigilada.
24	**cod_macromot**	Código de la categoría de macromotivo asignada a la PQRD.
25	**macromotivo**	Descripción de la categoría de macromotivo de la PQRD.
26	**cod_motgen**	Código de la categoría de motivo general de la PQRD.
27	**motivo_general**	Descripción del motivo general por el cual se interpone la PQRD.
28	**cod_motesp**	Código del motivo específico de la PQRD.
29	**motivo_especifico**	Descripción del motivo específico por el cual se interpone la PQRD.
30	**patologia_1**	Descripción de la patología relacionada con el motivo de la PQRD, si aplica.
31	**patologia_tipo**	Tipo o grupo de patología relacionada con la PQRD.
32	**cie_10**	Código de diagnóstico según la clasificación internacional de enfermedades CIE-10.
33	**alto_costo**	Indicador de si la PQRD se relaciona con una patología catalogada como de alto costo.
34	**riesgo_vida**	Indicador de si la situación reportada representa riesgo de vida para la persona afectada sí/no.

Tabla 2.1: Diccionario de datos (Elaboración propia con base en la estructura de datos de la Superintendencia Nacional de Salud.)

Posteriormente, se identificaron los nombres de las columnas y los tipos de datos registrados en DuckDB. Se validó que el conjunto de datos utilizado para el análisis está conformado por 34 variables, incluyendo campos originales y variables derivadas como periodo, construida a partir del año y el mes de la PQRD. El tipo de dato predominante fue VARCHAR, con 30 variables categóricas o textuales; seguido por 3 variables de tipo INTEGER y una variable de tipo BIGINT, correspondiente al identificador técnico objectid, como se observa en la Figura 2.2.

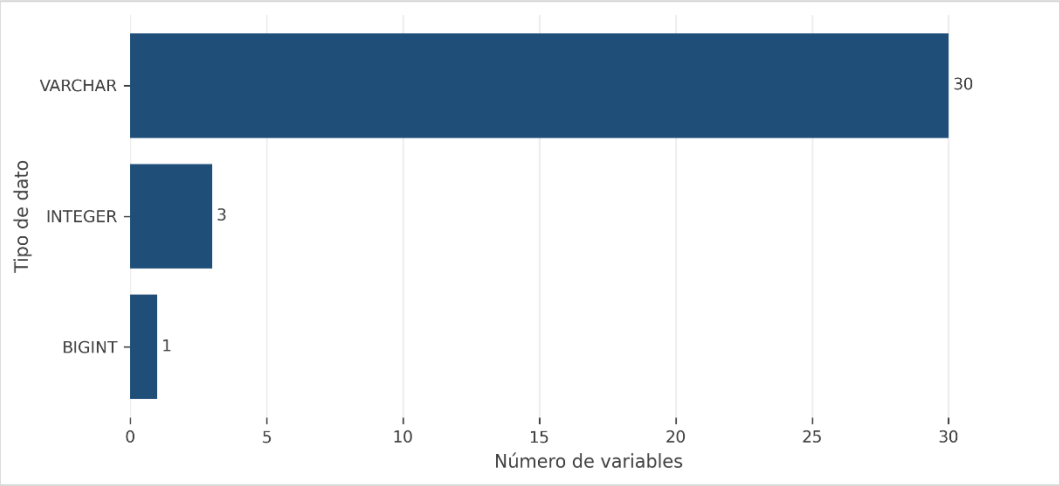


Figura 2.2: Distribución de variables por tipo de dato (Elaboración propia)

En la Figura 2.2, el tipo de dato predominante en el conjunto de datos es VARCHAR, lo que indica que la base está compuesta principalmente por variables categóricas y textuales. En menor proporción se identifican variables numéricas de tipo INTEGER y BIGINT, asociadas principalmente a campos temporales y de identificación.

Posteriormente, se clasificaron las variables según su naturaleza (cuantitativa o cualitativa), su tipo (discreta, continua, binaria, nominal), su escala de medición (nominal, ordinal, agrupada)

- Variables como pqr_canal, afec_genero, afec_regafiliacion, ent_nombre, ent_alias_sns, macromotivo, motivo_general y motivo_especifico fueron identificadas como variables cualitativas nominales. Estas variables resultan relevantes para segmentaciones, análisis de frecuencia, caracterización de los reportes y conservación del contexto durante el modelado supervisado.

- Variables como afec_edadr y afec_educ fueron clasificadas como variables cualitativas ordinales o agrupadas, dado que representan rangos de edad o niveles educativos con un orden conceptual definido.

- Otras como alto_costo y riesgo_vida fueron reconocidas como variables cualitativas, representando indicadores críticos de severidad o complejidad clínica de los casos reportados.

No. variable	Nombre de la variable	Tipo DuckDB	Tipo de variable	Clase / Escala de medición
1	**objectid**	BIGINT	Cuantitativa	Discreta (ID técnico)
2	**year**	INTEGER	Cuantitativa	Discreta - Ordinal (año)
3	**mes**	INTEGER	Cuantitativa	Discreta - Ordinal (1–12)
4	**periodo**	INTEGER	Cuantitativa	Discreta - Ordinal (YYYYMM)
5	**pqr_canal**	VARCHAR	Cualitativa	Nominal
6	**pet_cod_depto**	VARCHAR	Cualitativa	Nominal (código)
7	**pet_cod_mpio**	VARCHAR	Cualitativa	Nominal (código)
8	**id_afec**	VARCHAR	Cualitativa	Nominal (identificador anonimizado)
9	**afec_parentesco**	VARCHAR	Cualitativa	Nominal
10	**afec_genero**	VARCHAR	Cualitativa	Nominal
11	**afec_edadr**	VARCHAR	Cualitativa	Agrupada / Ordinal (rangos de edad)
12	**afec_educ**	VARCHAR	Cualitativa	Ordinal (nivel educativo)
13	**afec_regafiliacion**	VARCHAR	Cualitativa	Nominal
14	**afec_getnico**	VARCHAR	Cualitativa	Nominal
15	**afec_pobespecial**	VARCHAR	Cualitativa	Nominal
16	**afec_cod_depto**	VARCHAR	Cualitativa	Nominal (código)
17	**afec_cod_mpio**	VARCHAR	Cualitativa	Nominal (código)
18	**ent_nombre**	VARCHAR	Cualitativa	Nominal
19	**ent_tipovig_sns**	VARCHAR	Cualitativa	Nominal
20	**ent_cod_sns**	VARCHAR	Cualitativa	Nominal (código)
21	**ent_alias_sns**	VARCHAR	Cualitativa	Nominal
22	**ent_cod_depto**	VARCHAR	Cualitativa	Nominal (código)
23	**ent_cod_mpio**	VARCHAR	Cualitativa	Nominal (código)
24	**cod_macromot**	VARCHAR	Cualitativa	Nominal (código)
25	**macromotivo**	VARCHAR	Cualitativa	Nominal
26	**cod_motgen**	VARCHAR	Cualitativa	Nominal (código)
27	**motivo_general**	VARCHAR	Cualitativa	Nominal
28	**cod_motesp**	VARCHAR	Cualitativa	Nominal (código)
29	**motivo_especifico**	VARCHAR	Cualitativa	Nominal
30	**patologia_1**	VARCHAR	Cualitativa	Nominal
31	**patologia_tipo**	VARCHAR	Cualitativa	Nominal
32	**cie_10**	VARCHAR	Cualitativa	Nominal (código CIE-10)
33	**alto_costo**	VARCHAR	Cualitativa	Nominal
34	**riesgo_vida**	VARCHAR	Cualitativa	Binaria (Sí/No)

Tabla 2.2: Clasificación de variables del conjunto de datos (Elaboración propia)

2.6 Análisis descriptivo de los Datos

A continuación, se presenta la descripción del análisis exploratorio realizado sobre las variables clave del conjunto de datos. Este análisis caracterizó el comportamiento general de las PQRD durante el periodo 2017–2024, identificar las entidades con mayor volumen de reportes, analizar la distribución por variables

demográficas y geográficas, y conocer los motivos más frecuentes de las inconformidades de los usuarios.

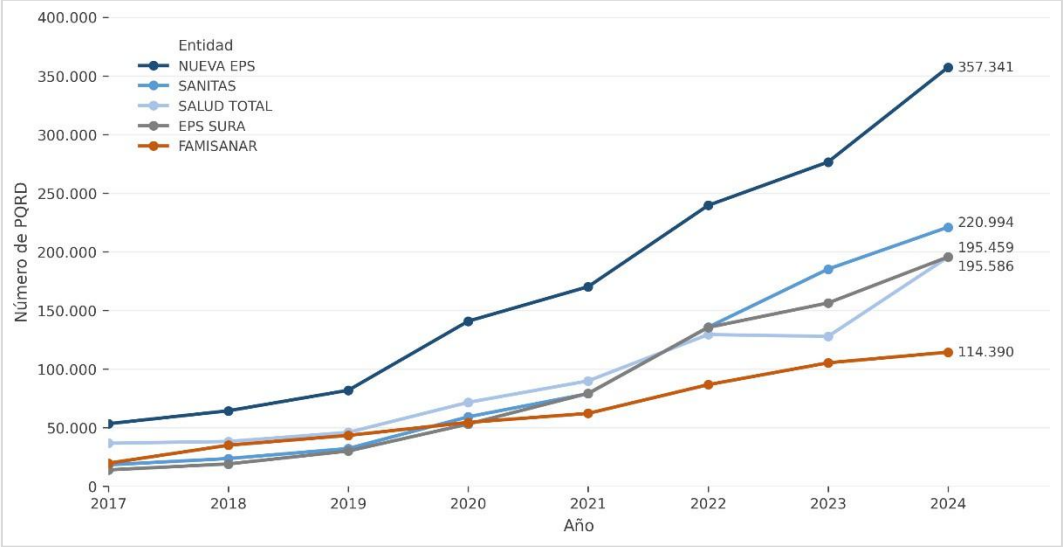


Figura 2.3: Tendencia anual de reportes PQRD (Elaboración propia)

Se observó un incremento sostenido en el número de registros reportados por la EPS NUEVA EPS a lo largo del periodo 2017–2024, alcanzando su punto máximo en 2024. SANITAS y SALUD TOTAL también muestran un comportamiento creciente, con aumentos notables a partir del año 2021. En el caso de SURAMERICANA EPS, se evidenció un crecimiento abrupto entre los dos últimos años, duplicando el volumen de reportes.

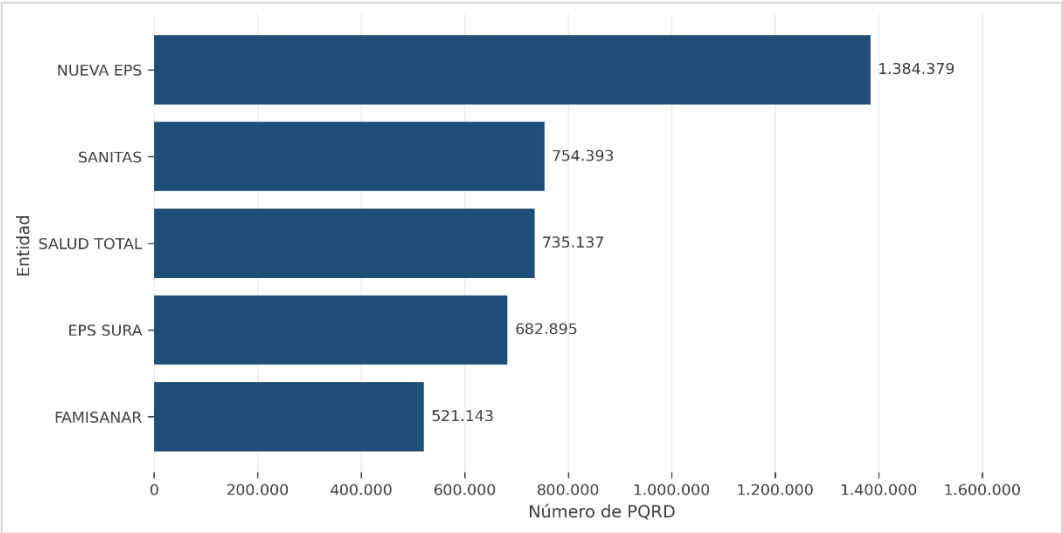


Figura 2.4: Top 5 de EPS con mayor número de PQRD (Elaboración propia)

Las entidades con mayor número acumulado de PQRD fueron NUEVA EPS, SANITAS y SALUD TOTAL. NUEVA

EPS lidera con 1.384.379 PQRD, seguida por SANITAS (754.393) y SALUD TOTAL (735.137). En el top 5

también se ubican EPS SURA (682.895) y FAMISANAR (521.143). Este resultado evidencia una alta concentración de reportes en un grupo reducido de entidades.

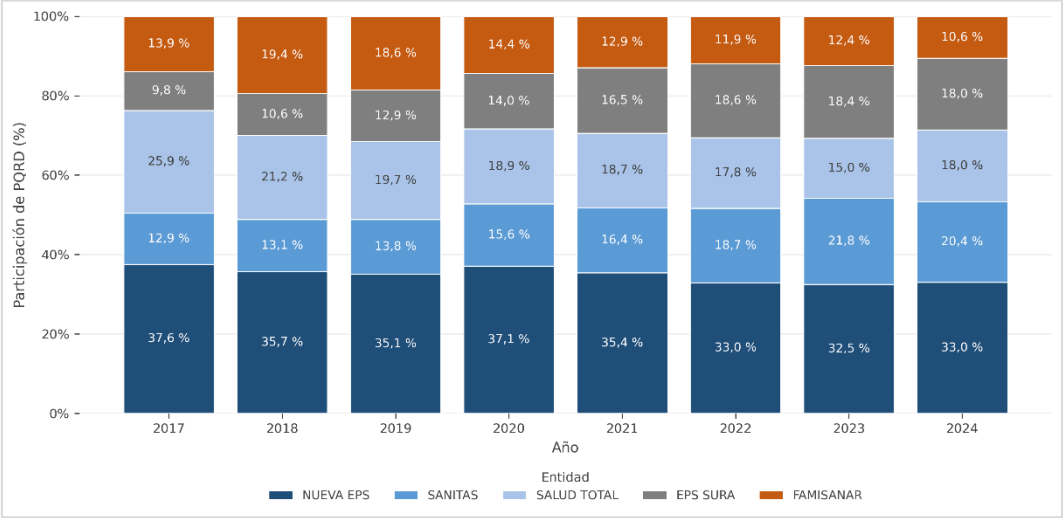


Figura 2.5: Participación porcentual de PQRD por año (Elaboración propia)

La participación relativa de las cinco EPS más representativas aumentó progresivamente durante el periodo analizado, representando más del 60 % de las PQRD totales en 2024. Esto indica una creciente concentración del volumen de reportes en pocas entidades aseguradoras.

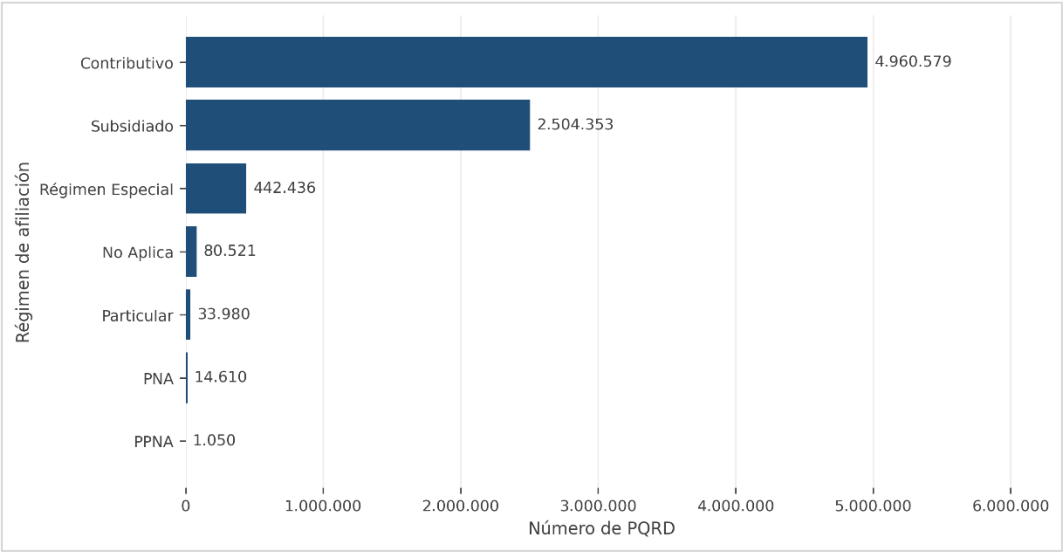


Figura 2.6: Distribución de PQRD por tipo de régimen (Elaboración propia)

La mayoría de las PQRD pertenece al régimen contributivo, con más de 4,9 millones de casos. El régimen subsidiado ocupa el segundo lugar, con cerca de 2,5 millones de registros. Los demás regímenes presentan

una participación menor en el conjunto de datos.

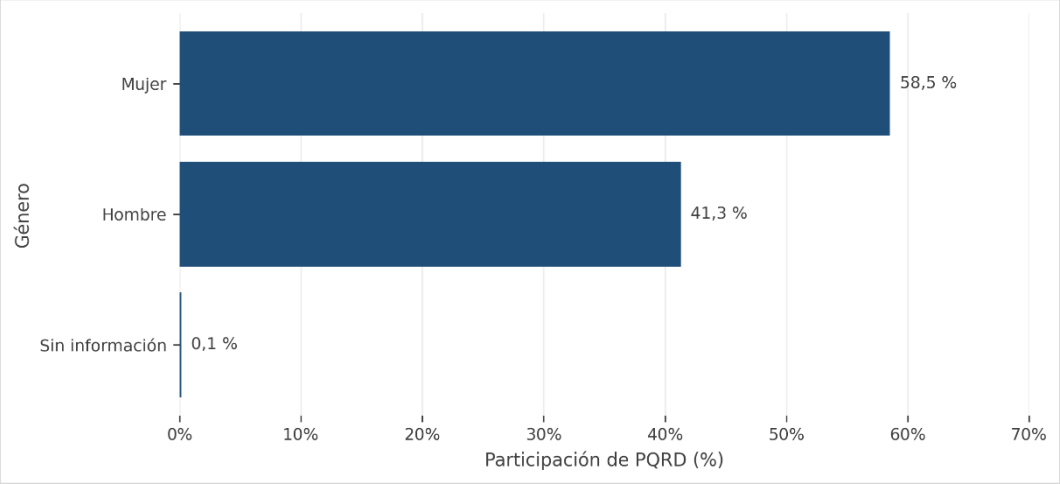


Figura 2.7: Distribución de PQRD por género (Elaboración propia)

Se registraron más reportes asociados a mujeres que a hombres. Aproximadamente el 58,6 % del total corresponde a mujeres, mientras que los hombres representan cerca del 41 %. Se identificaron más de 12 mil registros sin información definida de género.

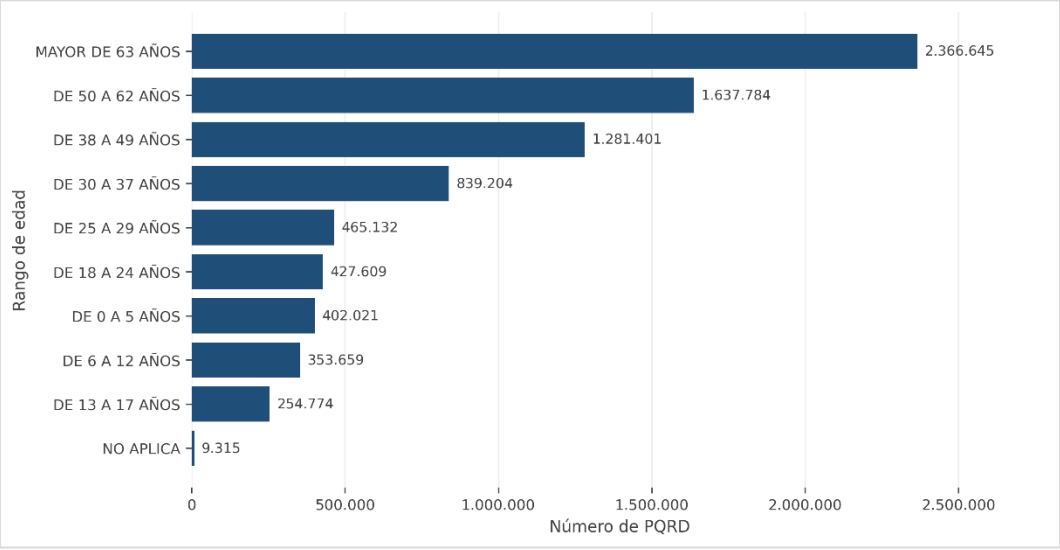


Figura 2.8: Distribución de PQRD por rango de edad (Elaboración propia)

La mayor proporción de reportes son realizados por personas mayores de 63 años y grupo entre los 50 a los 62 años, seguidas de los y de 38 a 49 años. Los grupos de edad pediátricos presentan una participación menor.

Este resultado sugiere que las personas adultas y mayores concentran una proporción significativa de las inconformidades reportadas, posiblemente por su mayor demanda de servicios.

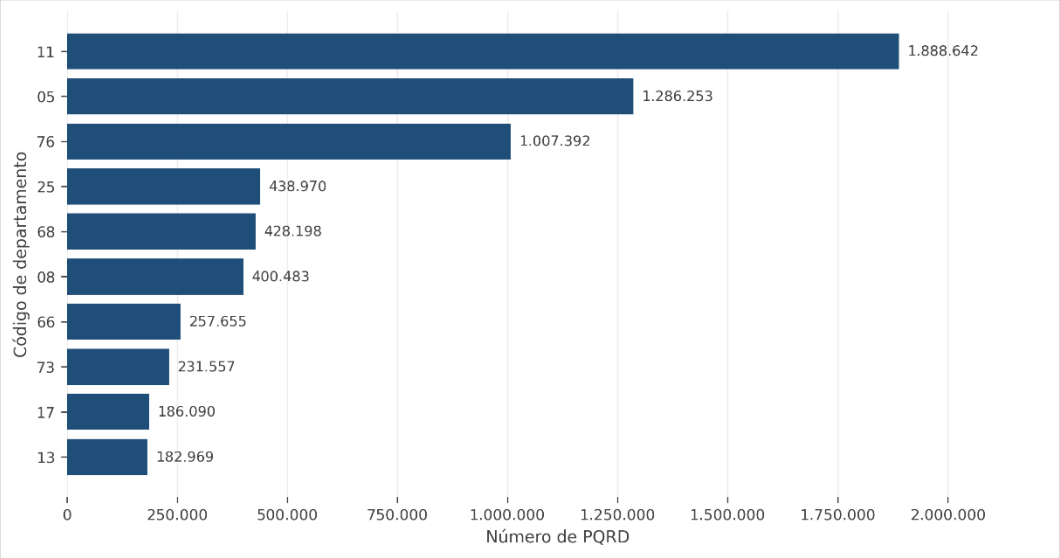


Figura 2.9: Top 10 de PQRD por departamento (Elaboración propia)

Bogotá es el territorio con mayor volumen de PQRD, superando los 1,8 millones de registros. Otros municipios con alta carga de reportes corresponden a los principales centros urbanos del país, como Medellín, Cali y Barranquilla, lo cual evidencia una concentración territorial asociada a zonas con alta densidad poblacional y mayor volumen de afiliados y servicios de salud.

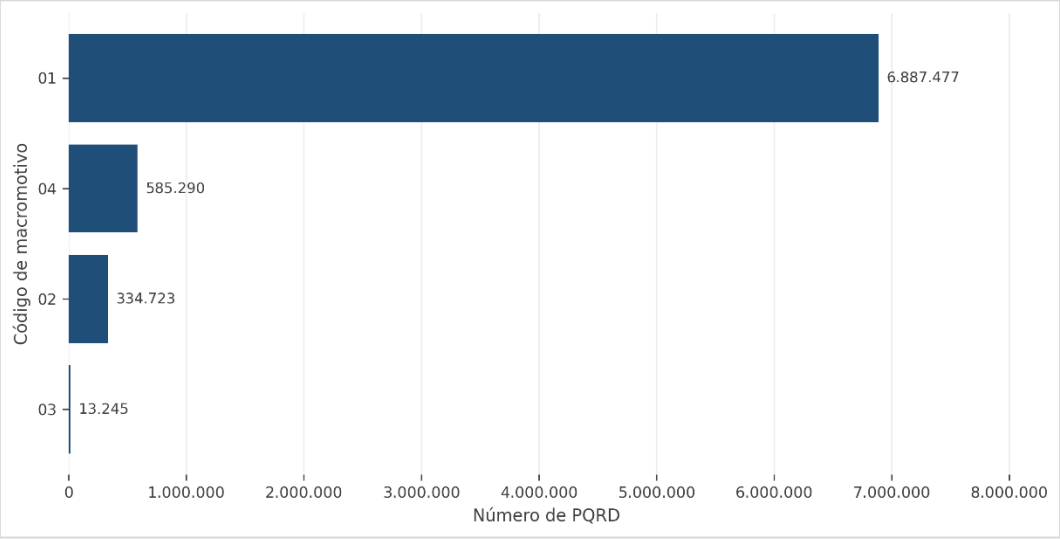


Figura 2.10: Top 5 de PQRD por macromotivo (Elaboración propia)

El macromotivo 01 (Barreras en el acceso a tecnologías y servicios de salud; y otros elementos complementarios para la atención del usuario) concentra más del 83 % del total de registros. Las demás categorías presentan un volumen considerablemente menor, lo cual evidencia una alta concentración temática de las PQRD en aspectos relacionados con el acceso y la oportunidad en la prestación de servicios y tecnologías en salud.

- **01:** Barreras en el acceso a tecnologías y servicios de salud; y otros elementos complementarios para la atención del usuario.
- **02:** Insatisfacción relacionada con la atención en salud.
- **03:** Insatisfacción relacionada con infraestructura y logística.
- **04:** Insatisfacción del usuario con el proceso administrativo.

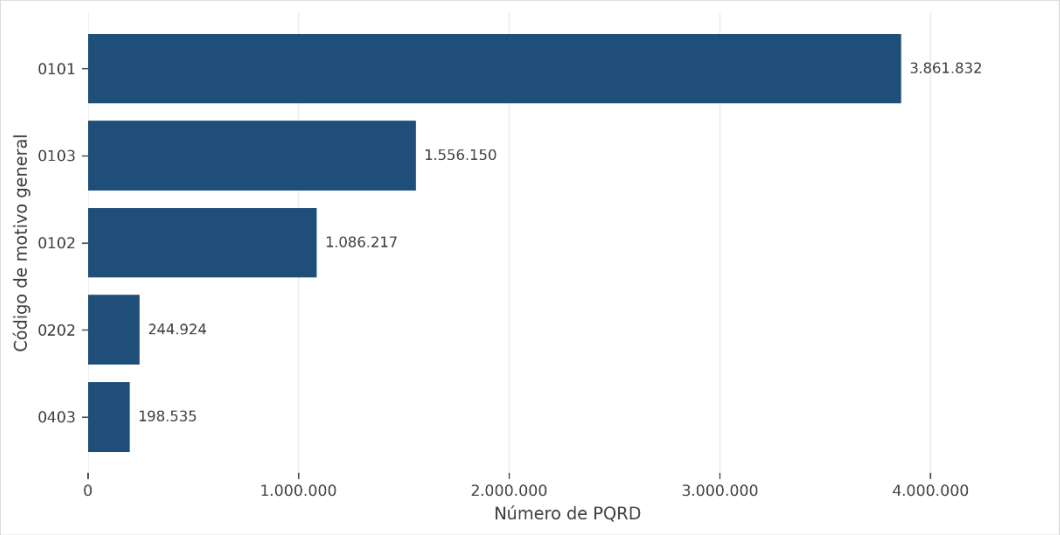


Figura 2.11: Top 5 de PQRD por motivo general (Elaboración propia)

El motivo general 0101 (falta de oportunidad en la prestación de tecnologías en salud y otros elementos complementarios para la atención del usuario) es el más frecuente en el conjunto de datos, seguido por 0103 (Falta de oportunidad en la autorización de tecnologías en salud y otros elementos complementarios para la atención del usuario) y 0102 (Negación de tecnologías en salud y otros elementos complementarios para la atención del usuario). Estos motivos se encuentran vinculados al macromotivo 01, lo cual refuerza la concentración de las PQRD en barreras de acceso, oportunidad y autorización de servicios o tecnologías en salud.

- **0101:** Falta de oportunidad en la prestación de tecnologías en salud y otros elementos complementarios para la atención del usuario.
- **0102:** Negación de tecnologías en salud y otros elementos complementarios para la atención del usuario.
- **0103**: falta de oportunidad en la autorización de tecnologías en salud y otros elementos complementarios para la atención del usuario.
- **0202**: Insatisfacción relacionada con la atención del personal en salud.
- **0403**: Insatisfacción relacionada con la atención del personal administrativo.

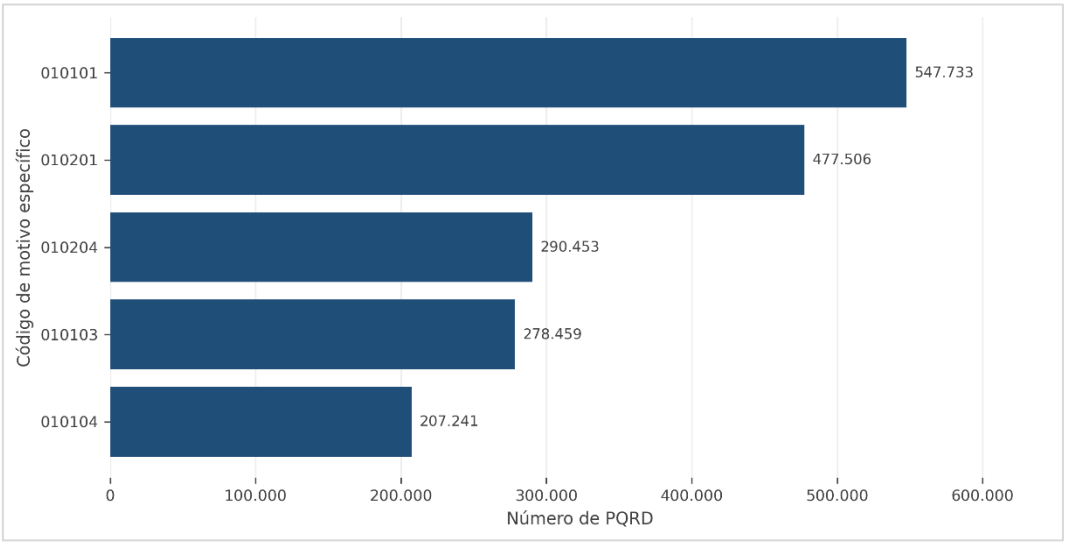


Figura 2.12: Top 5 de PQRD por motivo específico (Elaboración propia)

Como se presenta en la Figura 2.12, los motivos específicos con mayor frecuencia permiten identificar con mayor detalle los temas que concentran las inconformidades de los usuarios. Este nivel de clasificación complementa el análisis de macromotivos y motivos generales, y aporta información relevante para la construcción del corpus textual y la posterior identificación de riesgos operacionales.

- **010101:** Falta de oportunidad en las citas o consultas
- **010201:** Negación en la asignación de citas de cita o consulta
- **010204**: Negación para la entrega de tecnologías en salud y/o de otros servicios autorizados
- **010103**: Falta de oportunidad en la atención en otros servicios de salud
- **010104**: Falta de oportunidad en la entrega o entrega incompleta de tecnologías en salud y/o prestación de otros servicios

En conjunto, el análisis descriptivo evidencia una alta concentración de las PQRD en ciertas entidades, territorios, grupos poblacionales y motivos asociados principalmente a barreras de acceso, oportunidad, autorizaciones y tecnologías en salud. Estos hallazgos justifican la necesidad de complementar el análisis exploratorio con técnicas de procesamiento de lenguaje natural y clasificación supervisada, con el fin de reinterpretar los reportes desde una perspectiva de riesgo operacional y facilitar su posterior asociación con procesos de prestación de servicios de salud.

2.7 Depuración y preparación del corpus textual

En esta fase se construyó el corpus textual de trabajo a partir de los campos descriptivos asociados a la clasificación de las PQRD, específicamente macromotivo_clean, motivo_general_clean y motivo_especifico_clean. Estos campos contienen información textual estandarizada sobre la naturaleza de la inconformidad reportada, por lo cual constituyen una fuente relevante para el análisis mediante técnicas de procesamiento de lenguaje natural. A partir de estos campos se generó una variable denominada texto_full, mediante la concatenación de las descripciones normalizadas del macromotivo,

motivo general y motivo específico.

El subconjunto de trabajo conservó, junto con el texto, las variables estructuradas mínimas necesarias para el análisis posterior, tales como el identificador del registro, año, entidad, macromotivo, motivo general y motivo específico. Esto logró mantener la trazabilidad entre el registro original, las transformaciones aplicadas y las fases posteriores de etiquetado, modelado y despliegue.

Como parte de la depuración y preparación del corpus, se implementaron las siguientes actividades:

- **Eliminación de registros sin contenido textual útil:** se filtraron los casos con campos vacíos, nulos o conformados exclusivamente por textos genéricos del tipo “SIN DESCRIPCIÓN”, “NO APLICA”, “NA” u otros patrones equivalentes, que no aportan información relevante para la identificación de riesgos operacionales.

- **Normalización básica del texto:** se estandarizó el contenido textual mediante transformación a minúsculas, eliminación de saltos de línea, tabulaciones, caracteres no imprimibles, espacios redundantes y signos de formato que podían interferir con los procesos posteriores de tokenización y modelado.

- **Construcción de la variable texto_basic:** a partir de texto_full, se generó una versión depurada denominada texto_basic, conservando letras, números y espacios, y eliminando símbolos o caracteres no alfanuméricos irrelevantes. Esta variable permitió contar con una representación textual homogénea para las etapas posteriores de procesamiento de lenguaje natural.

- **Conservación de variables de contexto:** se mantuvieron variables estructuradas como objectid, year, ent_alias_sns, macromotivo_clean, motivo_general_clean y motivo_especifico_clean, con el propósito de garantizar trazabilidad analítica y facilitar la interpretación de los resultados del modelo.

Como resultado de estas transformaciones, se obtuvo un corpus textual depurado, constituido por registros con campos descriptivos de motivos válidos para el análisis textual. Este corpus mantiene el vínculo con las variables relevantes y se convierte en el insumo para las etapas de etiquetado, preprocesamiento lingüístico y construcción de modelos supervisados.

3. PREPARACIÓN DE LOS DATOS

En esta fase se preparó el corpus textual de las PQRD para el uso en modelos supervisados de clasificación, Para esto, se aplicaron técnicas de procesamiento de lenguaje natural para transformar los campos textuales. El proceso incluyó normalización básica del texto, tokenización, lematización, eliminación de stopwords y normalización léxica. Posteriormente, se construyó una muestra etiquetada manualmente como referencia para el entrenamiento y evaluación de modelos. A partir de esta base, se identificaron,

evaluaron y seleccionaron algoritmos de clasificación textual, considerando criterios de desempeño, interpretabilidad, estabilidad y viabilidad computacional en el contexto de riesgos operacionales en salud.

3.1 Normalización básica previa del texto

Esta actividad tuvo como objetivo generar una versión textual homogénea y depurada del corpus construido a partir de los campos descriptivos de las PQRD. Para ello, se aplicaron reglas de normalización básica sobre el texto, para estandarizar minúsculas, eliminar saltos de línea, tabulaciones, caracteres no imprimibles, espacios redundantes, símbolos y puntuación irrelevante para el modelado. Se conservaron letras, números y espacios, dado que estos elementos aportan información útil para la representación textual posterior.

Como resultado de esta fase, se generó la variable texto_basic, la cual constituye una versión depurada del texto y sirve como insumo para las etapas posteriores de tokenización, lematización, eliminación de stopwords y normalización léxica.

Texto original (texto_full)	Texto normalizado (texto_basic)
restricción en el acceso a los servicios de salud restricción en el acceso por falta de oportunidad para la atención falta de oportunidad en la asignación de citas de consulta médica general	restricción en el acceso a los servicios de salud restricción en el acceso por falta de oportunidad para la atención falta de oportunidad en la asignación de citas de consulta médica general
no reconocimiento de las prestaciones económicas incumplimiento de las prestaciones económicas (incapacidades) no reconocimiento y pago de las prestaciones económicas derivadas de licencia de enfermedad general	no reconocimiento de las prestaciones económicas incumplimiento de las prestaciones económicas incapacidades no reconocimiento y pago de las prestaciones económicas derivadas de licencia de enfermedad general
insatisfacción del usuario con el proceso administrativo disconformidad manifiesta seguimiento a derechos de petición	insatisfacción del usuario con el proceso administrativo disconformidad manifiesta seguimiento a derechos de petición
restricción en el acceso a los servicios de salud restricción en el acceso por demoras en la autorización demora de la autorización de procedimientos pos	restricción en el acceso a los servicios de salud restricción en el acceso por demoras en la autorización demora de la autorización de procedimientos pos
restricción en el acceso a los servicios de salud restricción en el acceso por demoras en la autorización demora de la autorización de hospitalizaciones	restricción en el acceso a los servicios de salud restricción en el acceso por demoras en la autorización demora de la autorización de hospitalizaciones
restricción en el acceso a los servicios de salud restricción en el acceso por falta de oportunidad para la atención falta de oportunidad en la prestación de servicios de promoción y prevención	restricción en el acceso a los servicios de salud restricción en el acceso por falta de oportunidad para la atención falta de oportunidad en la prestación de servicios de promoción y prevención

Tabla 3.1: Ejemplo de normalización básica del texto (Elaboración propia)

3.2 Tokenización

La aplicación de esta técnica consistió en segmentar el texto en tokens. Debido al volumen de la base, no se procesaron inicialmente los más de ocho millones de registros de forma completa para esta validación; en su lugar, se tomó una muestra aleatoria de 100.000 PQRD desde la vista v_pqrd_nlp_basic, con el fin de evaluar el comportamiento del corpus y revisar la longitud de los textos.

Para la tokenización del campo texto_basic se empleó la librería spaCy con modelo en español. Este procedimiento divide cada registro textual en tokens y calcula el número de unidades por texto, facilitandoel análisis preliminar de la estructura lingüística del corpus y su adecuación para las siguientes fases de procesamiento.

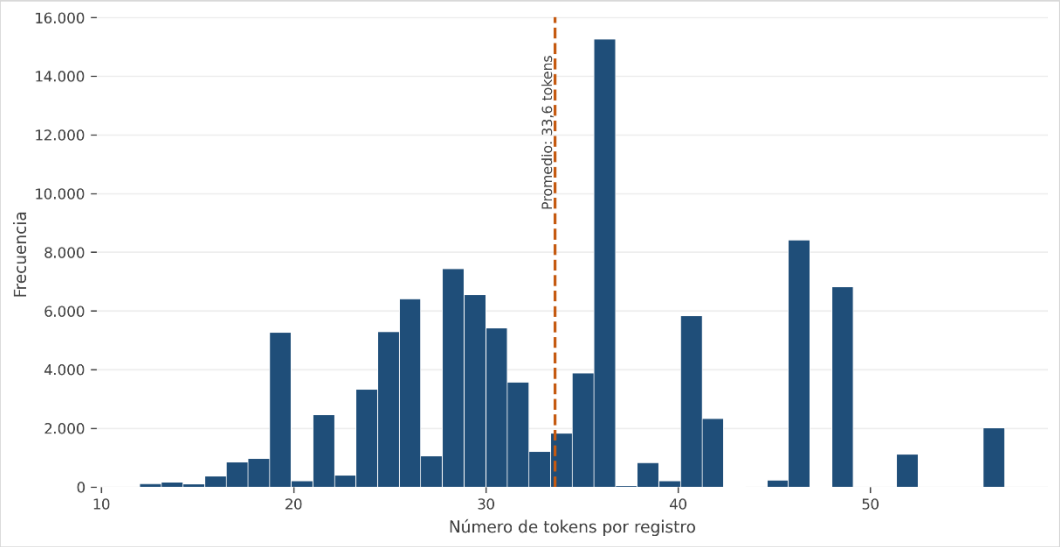


Figura 3.1. Distribución del número de tokens por registro (Elaboración propia)

Como se muestra la Figura 3.1, la mayoría de los registros presenta una longitud textual corta o media, con un promedio cercano a 34 tokens por registro. Este comportamiento es consistente con la estructura del corpus construido a partir de campos normalizados de macromotivo, motivo general y motivo específico, y permite confirmar que los textos tienen una extensión manejable para las etapas posteriores de procesamiento y modelado.

Para evaluar el efecto del preprocesamiento textual, se comparó el número promedio de términos por registro antes y después de aplicar la lematización, eliminación de palabras vacías y depuración básica del texto. Esta comparación permite observar la reducción del contenido textual y el efecto de la limpieza sobre la representación del corpus.

Tamaño de la muestra	100.000	**Promedio de tokens**	33,6
Mínimo	12	**Mediana**	33
Percentil 25	27	**Percentil 75**	41
Máximo	58	**Desviación estándar**	9,61

Texto básico	Tokens generados (muestra)	No. tokens
deficiencia en la efectividad de la atención en salud ineficacia en la atención presuntas fallas en la ética médica	deficiencia, en, la, efectividad, de, la, atención, en, salud, ineficacia, ...	19
barreras en el acceso a tecnologias y servicios de salud y otros elementos complementarios para la atención del usuario falta de oportunidad en la prestación de tecnologías en salud y otros elementos complementarios para la atención del usuario falta de oportunidad en las citas o consultas	barreras, en, el, acceso, a, tecnologías, y, servicios, de, salud, ...	46
restricción en el acceso a los servicios de salud restricción en el acceso por falta de oportunidad para la atención falta de oportunidad en la entrega de medicamentos no pos	restricción, en, el, acceso, a, los, servicios, de, salud, restricción, ...	30
restricción en el acceso a los servicios de salud restricción en el acceso por falta de oportunidad para la atención falta de oportunidad en la asignación de citas de consulta médica especializada de otras especialidades médicas	restricción, en, el, acceso, a, los, servicios, de, salud, restricción, ...	36
barreras en el acceso a tecnologías y servicios de salud y otros elementos complementarios para la atención del usuario negación de tecnologías en salud y otros elementos complementarios para la atención del usuario negación en la asignación de citas o consultas	barreras, en, el, acceso, a, tecnologías, y, servicios, de, salud, ...	41
restricción en el acceso a los servicios de salud restricción en el acceso por demoras en la autorización demora de la autorización de servicios de alto costo para enfermedades raras o huérfanas	restricción, en, el, acceso, a, los, servicios, de, salud, restricción, ...	32
restricción en el acceso a los servicios de salud restricción en el acceso por falta de oportunidad para la atención falta de oportunidad en la programación de cirugía	restricción, en, el, acceso, a, los, servicios, de, salud, restricción, ...	28
barreras en el acceso a tecnologías y servicios de salud y otros elementos complementarios para la atención del usuario falta de oportunidad en la prestación de tecnologías en salud y otros elementos complementarios para la atención del usuario falta de oportunidad en las citas o consultas	barreras, en, el, acceso, a, tecnologías, y, servicios, de, salud, ...	46
restricción en el acceso a los servicios de salud restricción en el acceso por falta de oportunidad para la atención falta de oportunidad en la asignación de citas de consulta médica especializada de otras especialidades médicas	restricción, en, el, acceso, a, los, servicios, de, salud, restricción, ...	36
barreras en el acceso a tecnologías y servicios de salud y otros elementos complementarios para la atención del usuario negación de tecnologías en salud y otros elementos complementarios para la atención del usuario negación para la entrega de tecnologías en salud y/o de otros servicios autorizados	barreras, en, el, acceso, a, tecnologías, y, servicios, de, salud, ...	48

Tabla 3.2: Ejemplo de tokenización del corpus textual (Elaboración propia)

3.3 Lematización y eliminación de stopwords

Se aplicó la lematización y el tratamiento de stopwords para normalizar el texto y reducir la variabilidad lingüística de las PQRD. La lematización transforma las palabras a su forma base o lema, agrupando variantes gramaticales bajo una representación común. Posteriormente, se aplicó una lista de stopwords en español para eliminar términos frecuentes que no aportan información relevante al análisis textual.

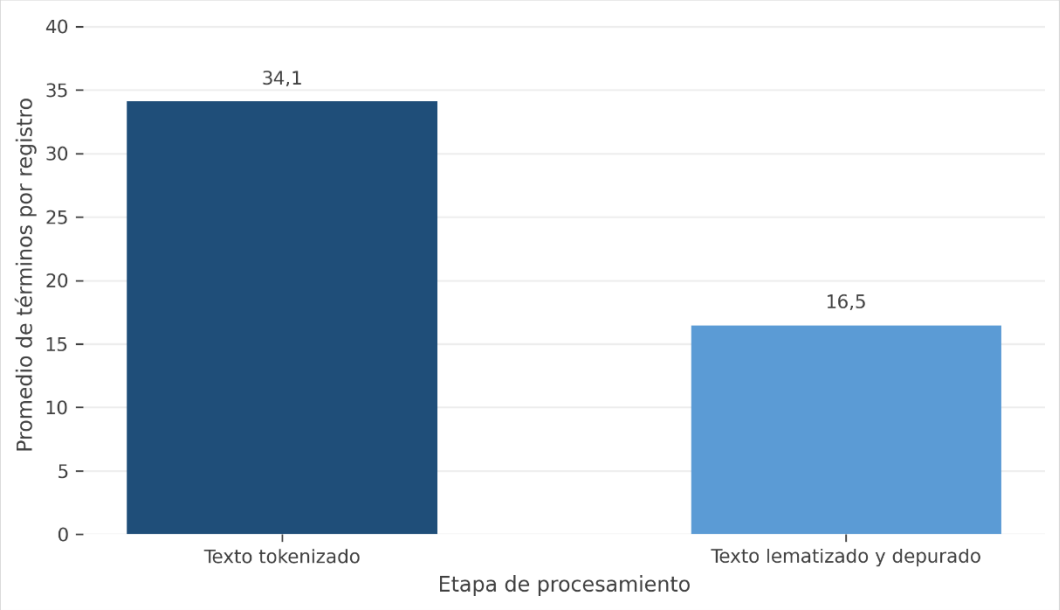


Figura 3.2. Comparación número promedio términos antes y después del preprocesamiento (Elaboración propia)

Como se observa en la Figura 3.2, el número promedio de términos disminuye después del preprocesamiento, pasando de un texto tokenizado más extenso a una versión lematizada y depurada más compacta. Esta reducción evidencia que el proceso permitió disminuir ruido textual y conservar términos con mayor valor semántico para la posterior vectorización y clasificación

La lista de stopwords fue ajustada, conservando palabras de negación como “no”, “nunca”, “jamás”, “sin” y “ni”, debido a su importancia semántica en el análisis de inconformidades relacionadas con negación de servicios, ausencia de atención, falta de oportunidad o barreras de acceso. Como resultado de esta etapa, se generó una versión lematizada del texto, representada en la variable texto_lem.

Texto básico	Texto lematizado y depurado	No. tokens	No. lemas
restricción en el acceso a los servicios de salud restricción en el acceso por fallas en la afiliación solicitud del usuario de traslado de eps por fallas en la prestación del servicio	restricción acceso servicio salud restricción acceso falla afiliación solicitud usuario traslado eps falla prestación servicio	32	15
restricción en el acceso a los servicios de salud restricción en el acceso por demoras en la autorización demora de la autorización de cirugía pos	restricción acceso servicio salud restricción acceso demora autorización demora autorización cirugía pos	25	12
restricción en el acceso a los servicios de salud restricción en el acceso por falta de oportunidad para la atención falta de oportunidad en la asignación de citas de consulta médica especializada de otras especialidades médicas	restricción acceso servicio salud restricción acceso falta oportunidad atención faltar oportunidad asignación cita consulta médico especializado especialidad médico	36	18
restricción en el acceso a los servicios de salud restricción en el acceso por fallas en la afiliación negación del servicio cuando no se ha hecho efectivo el traslado a otra eps	restricción acceso servicio salud restricción acceso falla afiliación negación servicio efectivo traslado eps	32	13

restricción en el acceso a los servicios de salud restricción en el acceso por falta de oportunidad para la atención falta de oportunidad en la asignación de citas de consulta médica general	restricción acceso servicio salud restricción acceso falta oportunidad atención faltar oportunidad asignación cita consulta médico general	32	16
insatisfacción del usuario con el proceso administrativo limitaciones en la información deficiente información sobre derechos deberes y trámites	insatisfacción usuario proceso administrativo limitación información deficiente información derecho deber trámite	18	11
barreras en el acceso a tecnologías y servicios de salud y otros elementos complementarios para la atención del usuario falta de oportunidad en la autorización de tecnologías en salud y/o de otros servicios	barrera acceso tecnología servicio salud elemento complementario atención usuario faltar oportunidad autorización tecnología salud servicio	54	24
restricción en el acceso a los servicios de salud restricción en el acceso por demoras en la autorización demora de la autorización de medicamentos no pos	restricción acceso servicio salud restricción acceso demora autorización demora autorización medicamento	26	11

Tabla 3.3: Ejemplo de lematización y eliminación de stopwords (Elaboración propia)

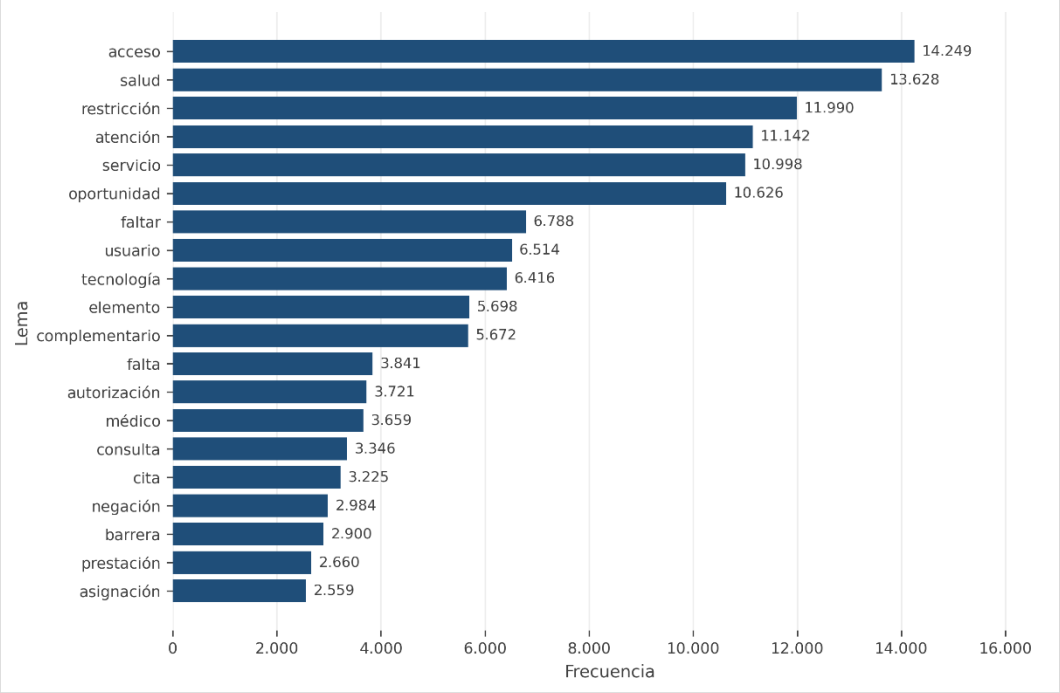


Figura 3.3. Top 20 lemas más frecuentes después de la lematización (Elaboración propia)

Los resultados en la Figura 3.3, muestra que los lemas más frecuentes del corpus depurado se relacionan con conceptos centrales de la prestación de servicios de salud, entre ellos acceso, salud, atención, servicio, oportunidad, tecnología y autorización. Este resultado indica que, una vez aplicada la lematización y eliminación de palabras vacías, el corpus conserva términos con alta carga semántica y relevancia para la identificación de riesgos operacionales.

3.4 Normalización léxica y definición del texto final para modelar.

A partir de los lemas obtenidos en la fase anterior, se aplicaron reglas de normalización léxica orientadas a unificar variantes ortográficas, términos equivalentes y expresiones frecuentes en el dominio de salud. Por ejemplo, se normalizaron términos asociados a autorizaciones, tecnologías, medicamentos, prestaciones y usuarios, con el fin de reducir ruido y mejorar la consistencia del vocabulario.

Con el fin de examinar el efecto del preprocesamiento sobre la diversidad léxica del corpus, se comparó el tamaño del vocabulario antes y después de la depuración textual. Para ello, se contabilizó el número de términos únicos presentes en el texto tokenizado inicial y en el texto final preparado para el modelado.

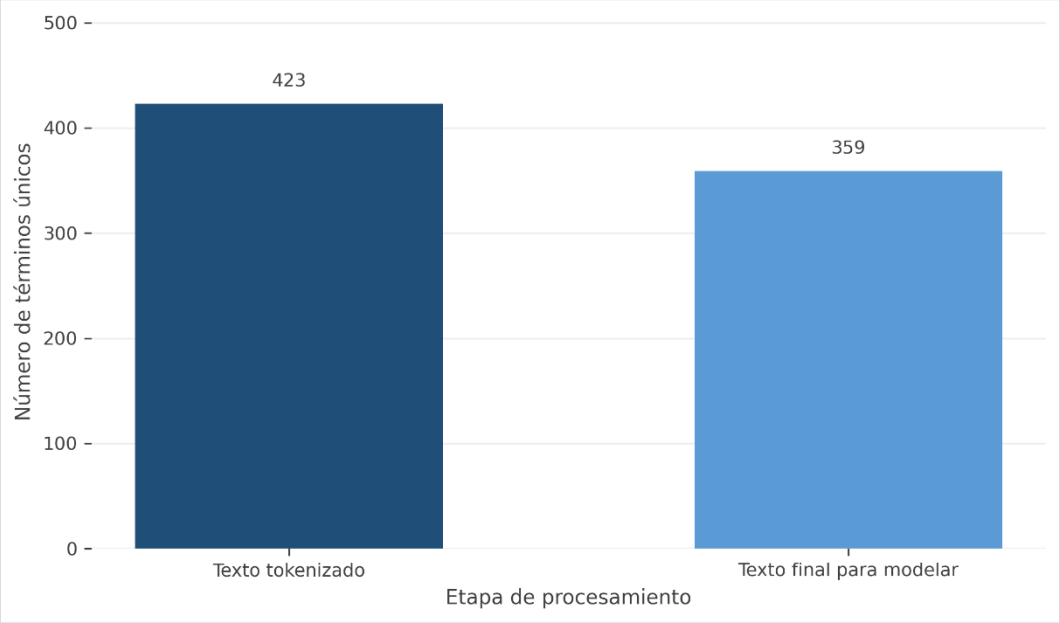


Figura 3.4. Comparación del tamaño del vocabulario por etapa de procesamiento (Elaboración propia)

Como se evidencia en la Figura 3.4, el tamaño del vocabulario se reduce después del preprocesamiento, al pasar de 423 términos únicos en el texto tokenizado a 359 términos en el texto final para modelar. Esta disminución refleja la eliminación de variantes redundantes, palabras vacías y términos con menor valor analítico. Aunque la reducción no es amplia, el resultado es consistente con la naturaleza del corpus, construido a partir de campos estructurados de clasificación de las PQRD.

Como resultado, se construyeron las variables tokens_nlp y texto_nlp. La primera conserva la lista de términos normalizados por registro, mientras que la segunda corresponde a la versión textual final utilizada como insumo para la representación numérica mediante TF-IDF y demás técnicas de modelado supervisado. Esta etapa permitió consolidar un corpus más estable, menos disperso y adecuado para la clasificación automática de riesgos operacionales.

Resumen longitud de texto final (n_tokens_nlp)	Muestra: 100.000 registros | Media: 16,4 | Mín: 8 | P25: 13 | Mediana: 16 | P75: 20 | Máx: 27

Texto básico	Texto lematizado	Texto final para modelar	No tokens	No lemas	No. términos finales
restricción en el acceso a los servicios de salud restricción en el acceso por fallas en la afiliación solicitud del usuario de traslado de eps por fallas en la prestación del servicio	restricción acceso servicio salud restricción acceso falla afiliación solicitud usuario traslado eps falla prestación servicio	restricción acceso servicio salud restricción acceso falla afiliación solicitud usuario traslado eps falla prestación servicio	32	15	15
restricción en el acceso a los servicios de salud restricción en el acceso por demoras en la autorización demora de la autorización de cirugía pos	restricción acceso servicio salud restricción acceso demora autorización demorar autorización cirugía pos	restricción acceso servicio salud restricción acceso demorar autorización demorar autorización cirugía pos	25	12	12
restricción en el acceso a los servicios de salud restricción en el acceso por falta de oportunidad para la atención falta de oportunidad en la asignación de citas de consulta médica especializada de otras especialidades médicas	restricción acceso servicio salud restricción acceso falta oportunidad atención faltar oportunidad asignación cita consulta médico especializado especialidad médico	restricción acceso servicio salud restricción acceso falta oportunidad atención faltar oportunidad asignación cita consulta médico especializado especialidad médico	36	18	18
restricción en el acceso a los servicios de salud restricción en el acceso por fallas en la afiliación negación del servicio cuando no se ha hecho efectivo el traslado a otra eps	restricción acceso servicio salud restricción acceso falla afiliación negación servicio efectivo traslado eps	restricción acceso servicio salud restricción acceso falla afiliación negación servicio efectivo traslado eps	32	13	13
restricción en el acceso a los servicios de salud restricción en el acceso por falta de oportunidad para la atención falta de oportunidad en la asignación de citas de consulta médica general	restricción acceso servicio salud restricción acceso falta oportunidad atención faltar oportunidad asignación cita consulta médico general	restricción acceso servicio salud restricción acceso falta oportunidad atención faltar oportunidad asignación cita consulta médico general	32	16	16
insatisfacción del usuario con el proceso administrativo limitaciones en la información deficiente información sobre derechos deberes y trámites	Insatisfacción usuario proceso administrativo limitación información deficiente información derecho deber trámite	insatisfacción usuario proceso administrativo limitación información deficiente información derecho deber trámite	18	11	11
barreras en el acceso a tecnologías y servicios de salud y otros elementos complementarios para la atención del usuario falta de oportunidad en la autorización de tecnologías en salud y/o de otros servicios	barrera acceso tecnología servicio salud elemento complementario atención usuario faltar oportunidad autorización tecnología salud servicio	barrera acceso tecnología servicio salud elemento complementario atención usuario faltar oportunidad autorización tecnología salud servicio	54	24	24
restricción en el acceso a los servicios de salud restricción en el acceso por demoras en la autorización demora de la autorización de medicamentos no pos	restricción acceso servicio salud restricción acceso demora autorización demorar autorización medicamento	restricción acceso servicio salud restricción acceso demorar autorización demorar autorización medicamento	26	11	11
barreras en el acceso a tecnologías y servicios de salud y otros elementos complementarios para la atención del usuario falta de oportunidad en la prestación de tecnologías en salud y otros elementos complementarios para la atención del usuario falta de oportunidad en la atención en otros servicios de salud	barrera acceso tecnología servicio salud elemento complementario atención usuario faltar oportunidad prestación tecnología salud elemento complementario atención usuario faltar oportunidad atención servicio salud	barrera acceso tecnología servicio salud elemento complementario atención usuario faltar oportunidad prestación tecnología salud elemento complementario atención usuario faltar oportunidad atención servicio salud	49	23	23

Tabla 3.4: Ejemplo de normalización léxica (Elaboración propia)

3.5 Construcción del corpus etiquetado de riesgos operacionales

Con el corpus depurado, se avanzó hacia la construcción de un conjunto de datos etiquetado específicamente para riesgos operacionales, alineado con el marco conceptual del proyecto y con la definición de riesgo operacional adoptada para el sector salud.

En primer lugar, se definió una taxonomía de categorías de riesgo operacional R1–R7, orientada a agrupar las PQRD según su relación con fallas o debilidades en procesos de la prestación de servicios de salud.

Estas categorías permitieron clasificar los registros en riesgos asociados, entre otros aspectos, a acceso y oportunidad, autorizaciones, medicamentos y tecnologías en salud, calidad técnico-asistencial, atención al usuario, aspectos económicos y otras situaciones no clasificadas en las categorías principales.

Sobre esta base, se seleccionó una muestra aleatoria de 4.000 registros desde la vista v_pqrd_nlp_basic, conservando variables de contexto como identificador del registro, año, entidad, macromotivo, motivo general, motivo específico y texto normalizado. Cada registro fue revisado manualmente y se le asignó una etiqueta de riesgo operacional primaria, siguiendo criterios de consistencia semántica y reglas de decisión definidas para la clasificación.

El proceso de etiquetado permitió construir una base supervisada para el entrenamiento y evaluación de modelos de clasificación. La etiqueta primaria fue validada para garantizar que estuviera diligenciada y que correspondiera a una de las categorías definidas entre R1 y R7. Adicionalmente, la distribución de clases identificó un desbalance natural en los datos, con mayor concentración en algunas categorías de riesgo y menor representación en otras, aspecto que fue considerado posteriormente en la fase de modelado.

Como producto de estas actividades se obtuvo un conjunto de datos etiquetado que, para cada PQRD incluida en la muestra, contiene:

- Identificador único del registro.
- Texto depurado de la PQRD.
- Variables estructuradas de contexto como año, entidad, macromotivo, motivo general y motivo específico.
- Etiqueta de riesgo operacional primaria asignada manualmente.
- Campos complementarios para observaciones o validaciones del etiquetado.

objectid	year	Entidad	Macromotivo	Motivo general	Motivo específico	Texto básico	Riesgo
783092	2022	NUEVA EPS	restricción en el acceso a los servicios de salud	restricción en el acceso por falta de oportunidad para la atención	falta de oportunidad en la entrega de medicamentos pos	restricción en el acceso a los servicios de salud restricción en el acceso por falta de oportunidad para la atención falta de oportunidad en la entrega de medicamentos pos	R3
129002	2019	COOMEVA	restricción en el acceso a los servicios de salud	restricción por razones económica o de capacidad de pago	contratos vencidos con ips que afectan la continuidad del servicio	restricción en el acceso a los servicios de salud restricción por razones económica o de capacidad de pago contratos vencidos con ips que afectan la continuidad del servicio	R6
880605	2024	SANITAS	barreras en el acceso a tecnologías y servicios de salud; y otros elementos complementarios para la atención del usuario	falta de oportunidad en la autorización de tecnologías en salud y otros elementos complementarios para la atención del usuario	falta de oportunidad en la autorización de otros servicios de salud	barreras en el acceso a tecnologías y servicios de salud y otros elementos complementarios para la atención del usuario falta de oportunidad en la autorización de tecnologías en salud y otros elementos complementarios para la atención del usuario falta de oportunidad en la autorización de otros servicios de salud	R2
611342	2023	SAVIA SALUD EPS	barreras en el acceso a tecnologías y servicios de salud; y otros elementos complementarios para la atención del usuario	negación de tecnologías en salud y otros elementos complementarios para la atención del usuario	negación en la asignación de citas o consultas	barreras en el acceso a tecnologías y servicios de salud y otros elementos complementarios para la atención del usuario negación de tecnologías en salud y otros elementos complementarios para la atención del usuario negación en la asignación de citas o consultas	R2
443176	2023	MUTUAL SER	restricción en el acceso a los servicios de salud	restricción en el acceso por falta de oportunidad para la atención	falta de oportunidad en la asignación de citas de consulta médica especializada de gastroenterología	restricción en el acceso a los servicios de salud restricción en el acceso por falta de oportunidad para la atención falta de oportunidad en la asignación de citas de consulta médica especializada de gastroenterología	R1
178771	2024	CAPITAL SALUD	barreras en el acceso a tecnologías y servicios de salud; y otros elementos complementarios para la atención del usuario	negación de tecnologías en salud y otros elementos complementarios para la atención del usuario	negación en la asignación de citas o consultas	barreras en el acceso a tecnologías y servicios de salud y otros elementos complementarios para la atención del usuario negación de tecnologías en salud y otros elementos complementarios para la atención del usuario negación en la asignación de citas o consultas	R2
481894	2020	NUEVA EPS	restricción en el acceso a los servicios de salud	restricción en el acceso por falta de oportunidad para la atención	falta de oportunidad en la asignación de citas de consulta médica especializada de otras especialidades médicas	restricción en el acceso a los servicios de salud restricción en el acceso por falta de oportunidad para la atención falta de oportunidad en la asignación de citas de consulta médica especializada de otras especialidades médicas	R1
820797	2024	EPS SURA	barreras en el acceso a tecnologías y servicios de salud; y otros elementos complementarios para la atención del usuario	negación de tecnologías en salud y otros elementos complementarios para la atención del usuario	negación en la asignación de citas o consultas	barreras en el acceso a tecnologías y servicios de salud y otros elementos complementarios para la atención del usuario negación de tecnologías en salud y otros elementos complementarios para la atención del usuario negación en la asignación de citas o consultas	R2
43880	2020	EPS SURA	restricción en el acceso a los servicios de salud	restricción en el acceso por falta de oportunidad para la atención	falta de oportunidad en la asignación de citas de consulta médica general	restricción en el acceso a los servicios de salud restricción en el acceso por falta de oportunidad para la atención falta de oportunidad en la asignación de citas de consulta médica general	R1
635257	2023	COOSALUD	barreras en el acceso a tecnologías y servicios de salud; y otros elementos complementarios para la atención del usuario	falta de oportunidad en la prestación de tecnologías en salud y otros elementos complementarios para la atención del usuario	falta de oportunidad en la entrega o entrega incompleta de tecnologías en salud y/o prestación de otros servicios	barreras en el acceso a tecnologías y servicios de salud y otros elementos complementarios para la atención del usuario falta de oportunidad en la prestación de tecnologías en salud y otros elementos complementarios para la atención del usuario falta de oportunidad en la entrega o entrega incompleta de tecnologías en salud y/o prestación de otros servicios	R3

Tabla 3.5: Muestra del corpus etiquetado de riesgos operacionales (Elaboración propia)

3.6 Identificación y evaluación de algoritmos

En esta etapa se identificaron y evaluaron diferentes algoritmos de clasificación de texto, con el propósito de seleccionar aquellos que fueran más adecuados para clasificar las PQRD en categorías de riesgo operacional R1–R7. La evaluación consideró tanto modelos clásicos de aprendizaje automático como modelos modernos basados en representaciones semánticas y arquitecturas transformer.

Para realizar esta evaluación preliminar, se tuvieron en cuenta los siguientes criterios:

- **Adecuación al tipo de dato:** capacidad del algoritmo para trabajar con texto procesado y representaciones vectoriales como TF-IDF o embeddings.
- **Desempeño esperado:** comportamiento esperado del modelo en tareas de clasificación de texto multiclase.
- **Manejo de alta dimensionalidad:** capacidad para procesar matrices dispersas y vocabularios amplios, propios de la representación textual.
- **Interpretabilidad:** facilidad para explicar el funcionamiento del modelo y justificar sus resultados en el contexto de gestión de riesgos.
- **Viabilidad computacional:** posibilidad de entrenar y evaluar el modelo dentro del entorno de Google Colab, considerando restricciones de memoria, tiempo de ejecución y disponibilidad de GPU.
- **Escalabilidad:** capacidad del modelo para ser aplicado posteriormente sobre un volumen masivo de registros PQRD.
- **Valor comparativo:** utilidad del algoritmo como línea base, modelo de contraste frente a otros enfoques.

A partir de estos criterios, los algoritmos se organizaron en dos grupos principales:

a. Modelos clásicos

Los modelos clásicos evaluados corresponden a algoritmos de aprendizaje automático que pueden trabajar con representaciones textuales como TF-IDF. Se caracterizan por su eficiencia, facilidad de implementación y menor costo computacional frente a modelos neuronales o transformers.

Dentro de este grupo se analizaron los siguientes algoritmos:

- **SVM lineal con TF-IDF:** modelo que busca un hiperplano de separación entre clases y suele tener buen desempeño en problemas de texto de alta dimensionalidad.
- **Regresión Logística con TF-IDF:** modelo lineal interpretable, estable y eficiente para clasificación textual multiclase.
- **Complement Naive Bayes y Multinomial Naive Bayes:** modelos probabilísticos rápidos y útiles como línea base inicial.
- **LightGBM y XGBoost:** modelos de boosting basados en árboles, útiles para evaluar relaciones no lineales y comparar con modelos lineales.

- **Random Forest:** ensamble de árboles de decisión utilizado como modelo de contraste metodológico.
- **fastText supervisado:** modelo basado en embeddings de palabras y subpalabras, evaluado como alternativa rápida, aunque no priorizado para la fase principal.

b. Modelos modernos

Los modelos modernos evaluados corresponden a enfoques basados en embeddings, redes neuronales o arquitecturas transformer. Estos modelos tienen mayor capacidad para capturar relaciones semánticas, contexto y polisemia, aunque suelen requerir más recursos computacionales.

Dentro de este grupo se analizaron los siguientes algoritmos:

- **BETO con fine-tuning:** modelo transformer pre entrenado en español, con alta capacidad para capturar contexto semántico.
- **Sentence Transformer y SBERT con clasificador lineal:** enfoque intermedio que genera embeddings semánticos densos y permite entrenar clasificadores más livianos.
- **SetFit:** enfoque basado en embeddings de frases y aprendizaje con pocos ejemplos etiquetados.
- **XLM y RoBERTa:** modelo transformer multilingüe, útil en contextos con mezcla de idiomas o ruido textual.
- **BiLSTM con embeddings:** red neuronal recurrente bidireccional capaz de procesar secuencias de texto.
- **CNN para texto:** red neuronal convolucional orientada a detectar patrones locales en secuencias textuales.

La evaluación permitió identificar qué algoritmos tenían mayor alineación con las características del proyecto, especialmente considerando que se trata de un problema de clasificación textual multiclase, con una muestra etiquetada moderada, clases desbalanceadas, necesidad de interpretabilidad y posterior aplicación sobre una base de PQRD.

3.7 Selección de los algoritmos más apropiados

Con base en la evaluación preliminar, se seleccionaron los algoritmos más adecuados para la fase de modelado, priorizando aquellos que presentaban mejor equilibrio entre desempeño esperado, interpretabilidad, viabilidad computacional y aplicabilidad al contexto del proyecto.

a. Algoritmos seleccionados

Se seleccionaron los siguientes modelos clásicos:

- **SVM lineal con TF-IDF:** se seleccionó por su alta adecuación al texto, buen desempeño esperado en matrices dispersas, robustez con vocabularios grandes y viabilidad de ejecución en Google Colab.

- **Regresión Logística con TF-IDF:** se seleccionó por ser un modelo lineal robusto, rápido, estable e interpretable. Su uso permite analizar pesos por clase y contar con una referencia metodológica clara frente a otros modelos.
- **Complement Naive Bayes y Multinomial Naive Bayes:** se seleccionaron como modelos probabilísticos de línea base, debido a su bajo costo computacional, rapidez de entrenamiento y utilidad para establecer un punto inicial de comparación.
- **LightGBM y XGBoost:** se seleccionaron como modelos de contraste frente a los enfoques lineales, dado que permiten evaluar si los métodos de ensamble y relaciones no lineales aportan valor adicional al problema.
- **Random Forest:** se seleccionó como modelo de comparación metodológica, aunque no suele ser la mejor opción para texto representado mediante TF-IDF de alta dimensionalidad.

También se seleccionaron los siguientes modelos modernos:

- **BETO con fine-tuning:** se seleccionó por ser un modelo transformer en español con alta capacidad para capturar contexto, polisemia y relaciones semánticas.
- **Sentence Transformer y SBERT con clasificador lineal:** se seleccionó como un enfoque intermedio entre modelos clásicos y transformers completos, debido a que permite obtener embeddings semánticos y entrenar clasificadores.

b. Criterios finales de selección

La selección final de algoritmos para la fase de modelado se fundamentó en los siguientes criterios:

- Modelos con buen desempeño esperado en clasificación textual.
- Capacidad para trabajar con textos en español.
- Viabilidad de entrenamiento en Google Colab.
- Adecuación a una muestra etiquetada de tamaño moderado.
- Balance entre desempeño, interpretabilidad y costo computacional.
- Utilidad para aplicar posteriormente el modelo sobre una base de PQRD.

Los algoritmos restantes fueron descartados para la fase de modelado, debido a que mayores exigencias de implementación en términos de costo computacional y escalado a una base de datos de más de 8 Millones de registros.

Con esta selección se estableció una estrategia de modelado comparativa, que ayuda a evaluar enfoques simples, robustos e interpretables frente a modelos modernos de mayor complejidad semántica. Esta comparación fue necesaria para seleccionar el modelo más adecuado para clasificar riesgos operacionales en el sector salud, garantizando que el modelo elegido no solo tuviera buen desempeño, sino que también fuera viable, explicable y aplicable sobre grandes volúmenes de información.

4. MODELADO

La fase de modelado tuvo como propósito entrenar, validar y comparar modelos de clasificación supervisada para asignar cada PQRD a una categoría principal de riesgo operacional entre R1 y R7. Para esta etapa se utilizó como insumo el corpus etiquetado manualmente y la variable texto_modelo, construida a partir del preprocesamiento textual desarrollado en las fases anteriores.

El problema se abordó como una tarea de clasificación multiclase y también cada registro debía asociarse con una única categoría principal de riesgo. Para ello, se evaluaron modelos clásicos basados en representación TF-IDF y modelos modernos basados en representaciones semánticas, con el fin de comparar alternativas con diferentes niveles de complejidad, capacidad predictiva, interpretabilidad y costo computacional.

El flujo general seguido durante esta fase se presenta en la Figura 4.1. Esta figura resume las principales actividades realizadas, desde la conformación del corpus etiquetado hasta la selección del modelo final, incluyendo la preparación del texto, la división de los datos, la representación numérica, la evaluación de modelos, la calibración de probabilidades y los controles metodológicos aplicados para reducir riesgos de fuga de información.

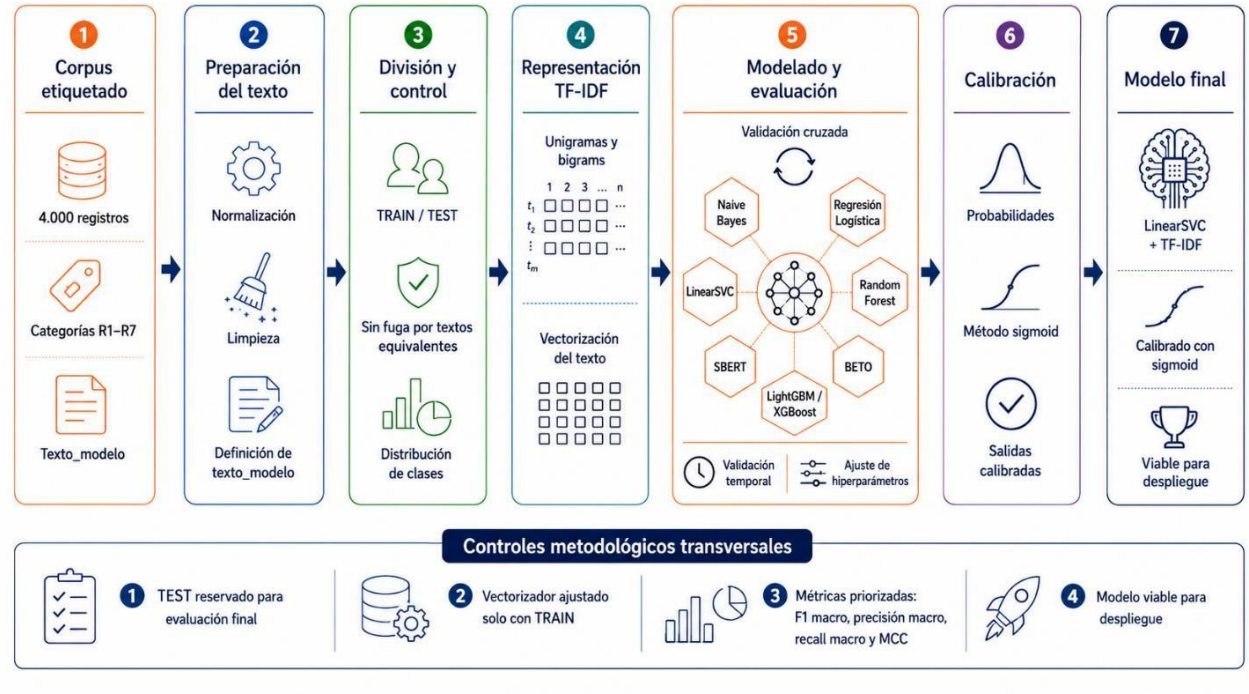


Figura 4.1: Flujo general del proceso de modelado (Elaboración propia)

La Figura 4.1 resume el flujo metodológico seguido durante la fase de modelado. El proceso inició con el corpus etiquetado de PQRD en las categorías de riesgo operacional R1–R7 y continuó con la preparación del texto, la división del conjunto de datos y la transformación mediante TF-IDF. La evaluación no se limitó a una única partición de entrenamiento y prueba. Se incorporaron controles metodológicos orientados a

reducir fuga de información, validar la estabilidad interna de los modelos y observar su comportamiento bajo una validación temporal. En particular, se priorizaron métricas macro, como F1 macro, precisión macro y recall macro, además del coeficiente de correlación de Matthews, debido al desbalance observado entre las categorías R1–R7.

La selección del modelo final se sustentó en la comparación cuantitativa de desempeño, estabilidad y viabilidad de implementación. Los resultados de esta comparación se presentan en las tablas y figuras siguientes, donde se contrastan los modelos evaluados bajo esquemas de validación cruzada y validación temporal.

4.1 Definición de la estrategia de modelado

La estrategia de modelado partió de la muestra etiquetada manualmente, compuesta inicialmente por 4.000 registros de PQRD clasificados en siete categorías de riesgo operacional. Estas categorías se definieron para representar los principales tipos de fallas o situaciones operativas identificables a partir de las inconformidades reportadas por los usuarios.

Las categorías de riesgo operacional utilizadas en el proyecto fueron las siguientes:

- **R1:** Acceso y oportunidad a servicios.
- **R2:** Autorizaciones y barreras administrativas.
- **R3:** Suministro de medicamentos y tecnologías en salud.
- **R4:** Calidad técnico asistencial y seguridad del paciente.
- **R5:** Atención al usuario, información y trato.
- **R6:** Temas económicos, facturación, recobros o cuotas moderadoras.
- **R7:** Otras situaciones no clasificadas en las categorías anteriores.

La distribución de la muestra etiquetada se presenta en la Tabla 4.1 y en la Figura 4.2. La tabla resume el número de registros y la participación porcentual por categoría, mientras que la figura permite observar visualmente el grado de concentración entre las clases.

Clase	Categoría de riesgo operacional	Número de registros	Participación (%)
R1	Acceso y oportunidad a servicios	1.633	40,8
R2	Autorizaciones y barreras administrativas	809	20,2
R3	Suministro de medicamentos y tecnologías en salud	690	17,2
R4	Calidad técnico-asistencial y seguridad del paciente	411	10,3
R5	Atención al usuario, información y trato	94	2,4
R6	Temas económicos, facturación, recobros o cuotas moderadoras	176	4,4
R7	Otras situaciones no clasificadas en las categorías anteriores	187	4,7
Total	**Corpus etiquetado**	**4.000**	**100,0**

Tabla 4.1: Distribución de clases del corpus etiquetado R1–R7 (Elaboración propia)

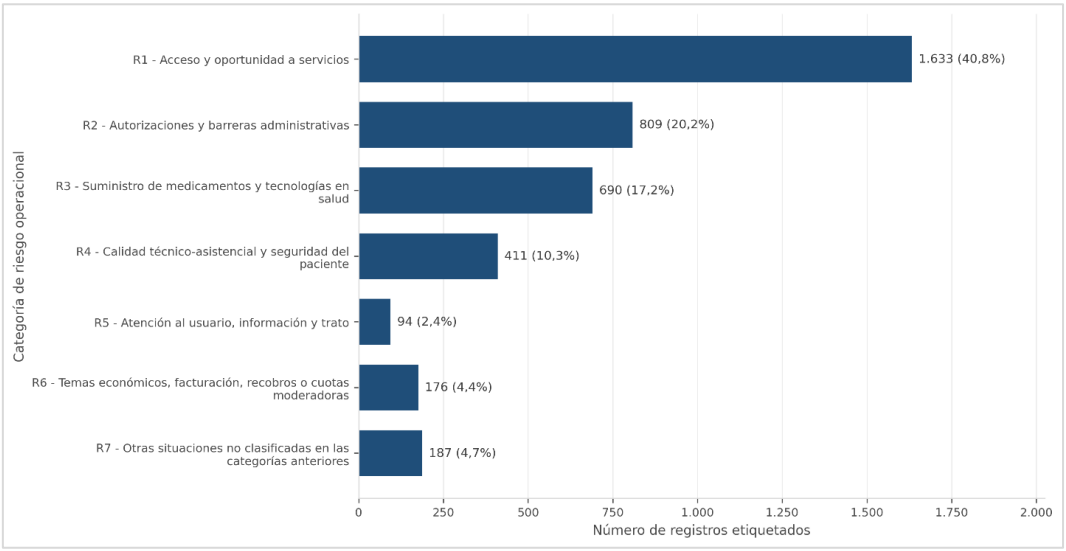


Figura 4.2: Distribución de clases del corpus etiquetado R1–R7. (Elaboración propia)

La distribución del corpus evidencia un desbalance importante entre las categorías de riesgo operacional. La clase R1 concentra 1.633 registros, equivalentes al 40,8% de la muestra, seguida por R2 con 809 registros, equivalentes al 20,2%, y R3 con 690 registros, equivalentes al 17,2%. En conjunto, estas tres categorías agrupan el 78,2% del corpus etiquetado, lo que indica una alta concentración de observaciones en los riesgos relacionados con acceso, autorizaciones y suministro de medicamentos o tecnologías en salud.

En contraste, las clases R5, R6 y R7 presentan una representación menor, con participaciones de 2,4%, 4,4% y 4,7%, respectivamente. La diferencia entre R1 y R5 es especialmente relevante, dado que la clase mayoritaria tiene aproximadamente 17,4 veces más registros que la clase con menor frecuencia. Este comportamiento podía afectar la evaluación del desempeño, ya que una métrica global como accuracy podría favorecer modelos con buen reconocimiento de las clases mayoritarias, pero con bajo desempeño en las categorías menos representadas.

Por esta razón, la evaluación no se basó únicamente en accuracy, sino que incorporó métricas macro, como precisión macro, recall macro y F1 macro, además del coeficiente de correlación de Matthews (MCC). Estas métricas permiten valorar el comportamiento del modelo de forma más equilibrada entre las categorías R1–R7, reduciendo el efecto dominante de las clases con mayor número de registros.

Además, la estrategia de modelado incorporó controles para reducir riesgos de sobreajuste y fuga de información. En particular, se separó un conjunto de prueba independiente, se ajustó el vectorizador únicamente sobre los datos de entrenamiento y se compararon los modelos bajo esquemas de validación cruzada estratificada y validación temporal. Esta aproximación permitió evaluar tanto la estabilidad interna de los modelos como su capacidad de generalización frente a registros de periodos posteriores.

4.2 División del conjunto de datos y controles de fuga de información

Después de definir la estrategia de modelado, el corpus etiquetado fue dividido en conjuntos de entrenamiento y prueba. Esta separación permitió entrenar los modelos con una parte de los datos y reservar un subconjunto independiente para la evaluación final del desempeño.

La partición incorporó dos controles principales. En primer lugar, se procuró conservar la representación de las siete categorías R1–R7 en los subconjuntos de entrenamiento y prueba, debido al desbalance observado en la muestra etiquetada. En segundo lugar, se aplicó un control anti-fuga basado en el texto normalizado, con el fin de evitar que registros con contenido textual equivalente quedaran simultáneamente en ambos subconjuntos.

Este control fue relevante porque la base contenía registros con textos repetidos o equivalentes después del proceso de normalización textual. Si estos textos se distribuían simultáneamente entre entrenamiento y prueba, el modelo podría reconocer patrones ya vistos durante el entrenamiento y generar una estimación sobrevalorada del desempeño. Por esta razón, el procedimiento utilizó texto_modelo_normalizado como criterio de agrupación y verificó posteriormente que no existieran textos compartidos entre los subconjuntos.

La división del conjunto de datos se realizó después de depurar los textos con conflictos exactos de etiquetado. Por esta razón, aunque la muestra etiquetada inicial estaba compuesta por 4.000 registros, la base efectiva de modelado quedó conformada por 3.984 registros. Esta base fue separada en conjuntos de entrenamiento y prueba, aplicando controles orientados a conservar la presencia de las siete categorías R1–R7 y reducir el riesgo de fuga de información por textos equivalentes.

Control metodológico	Resultado aplicado o verificado	Evidencia
Separación entrenamiento-prueba	TRAIN: 3.158 registros (79,3%) TEST: 826 registros (20,7%)	Total de registros modelados: 3.984
Estratificación por categoría de riesgo	TRAIN: 7 de 7 clases presentes TEST: 7 de 7 clases presentes	Soporte mínimo en TEST por clase: 22 registros
Conservación de la distribución de clases	Se comparó la proporción de cada clase frente a la distribución total	Máx. diferencia TRAIN vs total: 0,0294 Máx. diferencia TEST vs total: 0,1122
Control anti-fuga textual	Los textos equivalentes fueron asignados a un único subconjunto	Textos compartidos entre TRAIN y TEST: 0
Agrupación por texto normalizado	Se utilizó texto_modelo_normalizado como grupo anti-fuga	Textos únicos TRAIN: 133 textos únicos TEST: 53
Reserva del conjunto TEST	El conjunto TEST quedó reservado para evaluación final	No se utilizó en ajuste de parámetros, balanceo ni selección preliminar
Transformaciones posteriores al split	La vectorización TF-IDF y los ajustes de modelado se aplicaron después de dividir los datos	El vectorizador se ajustó únicamente con TRAIN

Tabla 4.2: División del conjunto de datos (Elaboración propia)

La Tabla 4.2, presenta que la base efectiva de modelado quedó conformada por 3.984 registros, distribuidos en 3.158 registros para entrenamiento, equivalentes al 79,3%, y 826 registros para prueba, equivalentes al 20,7%. Ambos subconjuntos conservaron representación de las siete categorías de riesgo operacional, con un soporte mínimo de 22 registros por clase en el conjunto TEST.

La validación anti-fuga confirmó que no existían textos compartidos entre entrenamiento y prueba, lo cual fue un control metodológico relevante para reducir el riesgo de sobreestimar el desempeño del modelo.

Además, el vectorizador TF-IDF se ajustó únicamente con los registros de entrenamiento, evitando que información del conjunto de prueba influyera en la construcción de la representación textual.

Aunque se procuró conservar la distribución de clases, la diferencia máxima frente a la distribución total fue de 0,0294 en TRAIN y de 0,1122 en TEST. Esta diferencia muestra que la partición estuvo condicionada por el control anti-fuga basado en texto normalizado, por lo que el objetivo no fue obtener una estratificación perfecta, sino balancear la representación de clases con la separación estricta de textos equivalentes entre entrenamiento y prueba.

Como resultado, quedaron definidos los archivos oficiales de entrenamiento y prueba que sirvieron como insumo para la vectorización TF-IDF, el entrenamiento de modelos, la validación, el ajuste de hiperparámetros y la evaluación final. Esta separación permitió mantener aislado el conjunto de prueba y reducir el riesgo de sobreestimar el desempeño del modelo.

4.3 Representación textual mediante TF-IDF

Después de dividir el conjunto de datos, los textos fueron transformados en una representación numérica para que pudieran ser utilizados por los modelos de clasificación. Para esta tarea se empleó TF-IDF, una técnica que pondera los términos de acuerdo con su frecuencia dentro de cada documento y su relevancia relativa dentro del corpus.

En este proyecto, TF-IDF fue pertinente porque el corpus de PQRD contiene expresiones recurrentes asociadas con acceso, autorizaciones, medicamentos, tecnologías en salud, oportunidad y atención al usuario. Esta representación permitió convertir cada registro en un vector de características textuales, manteniendo una estructura eficiente para modelos clásicos de clasificación como ComplementNB, MultinomialNB, LogisticRegression y LinearSVC.

La configuración utilizada para la representación textual se resume en la Tabla 4.3.

Componente	Configuración utilizada	Justificación metodológica
Columna de entrada	texto_modelo, derivada del texto normalizado del corpus	Texto preprocesado utilizado como insumo para el modelado supervisado.
Tipo de representación	TF-IDF	Convierte el texto en variables numéricas ponderadas según frecuencia e importancia relativa.
Analizador	analyzer = 'word'	Usa palabras completas como unidad de análisis.
Conversión a minúsculas	lowercase = False	Evita aplicar una transformación adicional, dado que el texto ya fue normalizado previamente.
Rango de n-gramas	ngram_range = (1, 2)	Incluye unigramas y bigramas para capturar términos individuales y combinaciones frecuentes.
Frecuencia mínima	min_df = 2	Excluye términos extremadamente raros dentro del conjunto de entrenamiento.
Frecuencia máxima	max_df = 0.95	Reduce el peso de términos frecuentes o poco discriminativos.
Escalamiento de frecuencia	sublinear_tf = True	Modera el efecto de términos repetidos dentro de un mismo documento.
Normalización	norm = l2	Estandariza la magnitud de los vectores y facilita la comparación entre registros
Control de fuga de información	Vectorizador ajustado solo sobre TRAIN	Evita que el vocabulario y los pesos TF-IDF se calculen usando información del conjunto de prueba

Tabla 4.3: Configuración de representación textual TF-IDF (Elaboración propia)

La representación TF-IDF se construyó a partir de la variable texto_modelo, derivada del texto previamente normalizado. La configuración incorporó unigramas y bigramas mediante ngram_range = (1, 2), lo que permitió capturar tanto términos individuales como combinaciones frecuentes dentro de las PQRD, así mismo, los parámetros min_df = 2 y max_df = 0.95 contribuyeron a excluir términos demasiado raros o excesivamente frecuentes, reduciendo ruido en la matriz de características.

El parámetro sublinear_tf = True permitió moderar el efecto de términos repetidos dentro de un mismo documento, mientras que norm = 'l2' estandarizó la magnitud de los vectores generados. Estas decisiones permitieron construir una representación textual compacta y adecuada para modelos clásicos de clasificación supervisada.

Para prevenir fuga de información, el vectorizador fue ajustado únicamente con los textos del conjunto de entrenamiento. Posteriormente, el conjunto de prueba fue transformado con ese mismo vectorizador, sin recalcular el vocabulario ni los pesos TF-IDF sobre los datos reservados para evaluación. Con ello, las matrices TF-IDF de entrenamiento y prueba quedaron disponibles para el entrenamiento, validación y comparación de los modelos clásicos.

4.4 Entrenamiento de modelos de clasificación

Con la representación textual definida mediante TF-IDF y los conjuntos de entrenamiento y prueba previamente separados, se procedió al entrenamiento de los modelos de clasificación. El objetivo de esta etapa fue comparar diferentes algoritmos supervisados para identificar cuál ofrecía mejor desempeño en

la asignación de las PQRD a las categorías de riesgo operacional R1–R7.

Los modelos evaluados incluyeron enfoques clásicos basados en TF-IDF, modelos de ensamble y modelos modernos basados en representaciones semánticas o arquitecturas tipo transformer. Esta combinación permitió contrastar alternativas con distintos niveles de complejidad, capacidad predictiva, interpretabilidad y costo computacional.

La Tabla 4.4 resume los modelos considerados en esta fase y el rol que cumplieron dentro del ejercicio comparativo.

No.	Modelo o enfoque evaluado	Representación textual	Tipo de enfoque	Propósito en la evaluación
1	Complement Naive Bayes / Multinomial Naive Bayes	TF-IDF	Modelo clásico probabilístico	Establecer una línea base para la clasificación textual multiclase.
2	LinearSVC	TF-IDF	Modelo clásico lineal	Evaluar un modelo lineal adecuado para matrices TF-IDF de alta dimensionalidad.
3	Regresión Logística	TF-IDF	Modelo clásico lineal probabilístico	Comparar un modelo interpretable y con salida probabilística frente a otros clasificadores lineales.
4	LightGBM / XGBoost	TF-IDF / embeddings	Modelo basado en boosting	Contrastar modelos de ensamble frente a representaciones textuales dispersas o densas.
5	Random Forest Classifier	TF-IDF	Modelo basado en árboles	Evaluar un enfoque no lineal y contrastarlo con los modelos lineales.
6	SBERT + LinearSVC	Embeddings semánticos	Modelo basado en embeddings	Evaluar representaciones semánticas densas mediante embeddings multilingües.
7	BETO + fine-tuning	Transformer en español	Modelo profundo contextual	Evaluar un modelo preentrenado en español ajustado al corpus etiquetado del proyecto.

Tabla 4.4: Modelos seleccionados para evaluación (Elaboración propia)

El entrenamiento se realizó respetando los controles definidos en las etapas anteriores. En particular, las transformaciones textuales se ajustaron únicamente sobre los datos de entrenamiento, y el conjunto de prueba se mantuvo reservado para la evaluación final. Asimismo, los ajustes relacionados con el tratamiento del desbalance de clases se aplicaron únicamente dentro del conjunto de entrenamiento o dentro de cada fold de validación, según correspondiera, evitando utilizar información del conjunto TEST durante el entrenamiento o la selección preliminar de modelos.

En esta etapa no se seleccionó directamente el modelo final. Los modelos entrenados sirvieron como insumo para las fases posteriores de validación cruzada, validación temporal, ajuste de hiperparámetros, calibración y evaluación en TEST. La selección definitiva se basó en la comparación cuantitativa de métricas como accuracy, precisión macro, recall macro, F1 macro y coeficiente de correlación de Matthews

4.5 Validación cruzada y validación temporal

Luego del entrenamiento inicial, los modelos fueron evaluados mediante dos esquemas complementarios: validación cruzada estratificada y validación temporal. El propósito de esta etapa fue analizar la estabilidad interna de los modelos y su capacidad de generalización frente a registros separados cronológicamente.

La validación cruzada estratificada permitió evaluar el desempeño en diferentes particiones del conjunto de entrenamiento, conservando la representación de las categorías de riesgo operacional R1–R7 en cada partición. Este control fue importante debido al desbalance observado en el corpus etiquetado, ya que redujo la dependencia de una única división de los datos.

De forma complementaria, se aplicó una validación temporal utilizando el año de la PQRD como criterio de separación. En este esquema, los modelos fueron entrenados con registros de años anteriores y evaluados sobre registros de años posteriores. Esta validación permitió simular un escenario más cercano al uso real del modelo, en el que las PQRD futuras pueden presentar variaciones frente a los datos utilizados durante el entrenamiento.

Durante ambos esquemas se mantuvieron controles para reducir fuga de información. En particular, las transformaciones textuales fueron ajustadas únicamente con los datos de entrenamiento de cada partición, y las evaluaciones se realizaron sobre subconjuntos no utilizados para ajustar los modelos. Este control fue relevante para evitar que el desempeño observado estuviera influenciado por información de los datos reservados para validación.

La comparación se realizó mediante métricas de clasificación multiclase. Aunque se calculó la exactitud global, el análisis priorizó precisión macro, recall macro, F1 macro y coeficiente de correlación de Matthews, debido a que estas métricas permiten evaluar de forma más equilibrada el desempeño entre categorías mayoritarias y minoritarias.

La Tabla 4.5 muestra los resultados de validación disponibles para los modelos evaluados. La comparación incluye métricas promedio y desviación estándar por esquema de validación, considerando accuracy, precisión macro, recall macro, F1 macro y MCC. Debido a que no todos los modelos generaron el mismo número de iteraciones válidas, estos resultados se interpretaron como evidencia preliminar de estabilidad y generalización, y no como criterio único para la selección definitiva del modelo.

Modelo	Esquema de validación	Iteraciones	Accuracy	Precisión macro	Recall macro	F1 macro	MCC
ComplementNB	Validación cruzada estratificada	5	0,7417 ± 0,1690	0,6770 ± 0,1992	0,7350 ± 0,1695	0,6596 ± 0,1879	0,6563 ± 0,2108
ComplementNB	Validación temporal	2	0,8616 ± 0,0870	0,8321 ± 0,1364	0,8801 ± 0,1191	0,8343 ± 0,1382	0,8476 ± 0,0927
MultinomialNB	Validación cruzada estratificada	5	0,7401 ± 0,1679	0,6781 ± 0,1977	0,7473 ± 0,1743	0,6627 ± 0,1916	0,6554 ± 0,2101
MultinomialNB	Validación temporal	2	0,8616 ± 0,0870	0,8321 ± 0,1364	0,8801 ± 0,1191	0,8343 ± 0,1382	0,8476 ± 0,0927
LinearSVC	Validación cruzada estratificada	5	0,7498 ± 0,1668	0,6486 ± 0,1314	0,7023 ± 0,1117	0,6325 ± 0,1288	0,6237 ± 0,1997
LinearSVC	Validación temporal	2	0,6593 ± 0,1991	0,5222 ± 0,0472	0,4880 ± 0,0873	0,4917 ± 0,0279	0,5880 ± 0,1071
Logistic Regression	Validación cruzada estratificada	5	0,6516 ± 0,3038	0,6484 ± 0,2051	0,6366 ± 0,1616	0,5664 ± 0,1559	0,5108 ± 0,3661
Logistic Regression	Validación temporal	2	0,6778 ± 0,1728	0,5138 ± 0,0591	0,5158 ± 0,0481	0,5026 ± 0,0039	0,6091 ± 0,0842
Random Forest Classifier	Validación cruzada estratificada	5	0,4867 ± 0,3084	0,5587 ± 0,1981	0,5583 ± 0,1698	0,4860 ± 0,1978	0,3176 ± 0,2925
SBERT + LinearSVC	Validación cruzada estratificada	5	0,9871 ± 0,0067	0,9877 ± 0,0053	0,9745 ± 0,0125	0,9808 ± 0,0086	0,9824 ± 0,0091
SBERT + LinearSVC	Validación temporal	5	0,9717 ± 0,0227	0,9500 ± 0,0653	0,9201 ± 0,0685	0,9329 ± 0,0661	0,9611 ± 0,0312
BETO + fine-tuning	Validación cruzada estratificada	3	0,9958 ± 0,0032	0,9953 ± 0,0039	0,9921 ± 0,0052	0,9936 ± 0,0033	0,9943 ± 0,0043
BETO + fine-tuning	Validación temporal	1	0,7972 ± 0,0000	0,8605 ± 0,0000	0,8226 ± 0,0000	0,8094 ± 0,0000	0,7379 ± 0,0000

Tabla 4.5: Resultados de validación cruzada y temporal disponibles por modelo (Elaboración propia)

A partir de los resultados consolidados en la Tabla 4.5, la Figura 4.4 resume visualmente el comportamiento de los modelos en dos métricas priorizadas para el análisis: F1 macro y MCC. Estas métricas permiten comparar el desempeño considerando el desbalance entre las categorías de riesgo operacional R1–R7.

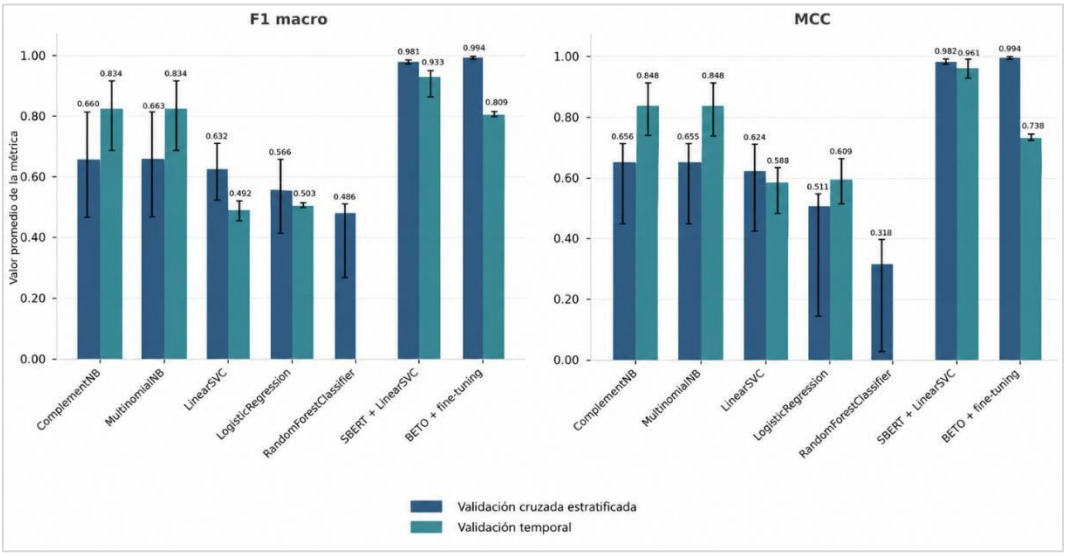


Figura 4.3: Comparación de F1 macro y MCC entre validación cruzada estratificada y validación temporal (Elaboración propia)

La comparación de la Tabla 4.5 y la Figura 4.4 muestra diferencias importantes entre los modelos evaluados. En la validación preliminar, los enfoques basados en representaciones semánticas y Transformers presentaron los valores más altos en algunas métricas. No obstante, estos resultados no se asumieron como criterio definitivo de selección, debido a que debían contrastarse con la evaluación final, la calibración de probabilidades, la estabilidad temporal, el costo computacional y la viabilidad de escalamiento sobre la base completa de PQRD.

En particular, SBERT + LinearSVC obtuvo un F1 macro de 0,9808 en validación cruzada estratificada y de 0,9329 en validación temporal, con valores de MCC de 0,9824 y 0,9611, respectivamente, esto evidencia un desempeño alto y relativamente estable entre ambos esquemas de validación.

Por su parte, BETO + fine-tuning alcanzó el mayor desempeño en validación cruzada, con un F1 macro de 0,9936 y un MCC de 0,9943. Sin embargo, en validación temporal el desempeño disminuyó a un F1 macro de 0,8094 y un MCC de 0,7379. Este resultado debe interpretarse con cautela, dado que la validación temporal de BETO se basó en una sola iteración, por lo que la desviación estándar de 0,0000 no representa estabilidad del modelo.

Los modelos clásicos presentaron comportamientos más variables. ComplementNB y MultinomialNB obtuvieron F1 macro cercanos a 0,66 en validación cruzada y de 0,8343 en validación temporal, aunque estos últimos resultados se calcularon sobre dos iteraciones. En contraste, LinearSVC redujo su F1 macro de 0,6325 en validación cruzada a 0,4917 en validación temporal, mientras que LogisticRegression presentó valores de 0,5664 y 0,5026, respectivamente.

Finalmente, RandomForestClassifier mostró el desempeño menos favorable, con un F1 macro de 0,4860 y un MCC de 0,3176 en validación cruzada estratificada, sin contar con una validación temporal válida disponible. En conjunto, estos resultados permitieron identificar diferencias entre estabilidad interna y generalización cronológica; sin embargo, debido al número desigual de iteraciones entre modelos, esta

etapa se consideró como un punto de control metodológico y no como el único criterio para seleccionar el modelo final.

Los resultados de SBERT + LinearSVC y BETO + fine-tuning reportados en la Tabla 4.5 deben interpretarse como resultados preliminares de validación, no como criterio definitivo de selección del modelo. Aunque estos enfoques obtuvieron métricas superiores en algunos escenarios, su comparación con los modelos clásicos requiere cautela, debido a diferencias en los esquemas de validación, número de iteraciones, costo computacional y condiciones de escalamiento.

En particular, los modelos basados en representaciones semánticas y Transformers pueden capturar mejor la similitud contextual entre textos, pero también pueden generar resultados optimistas cuando existen expresiones repetitivas o altamente similares, como ocurre en un corpus construido a partir de campos estandarizados de macromotivo, motivo general y motivo específico. Además, BETO redujo su desempeño en validación temporal y SBERT implicaba mayores costos de generación de embeddings e inferencia sobre millones de registros.

Por lo anterior, la Tabla 4.5 se asumió como una etapa de contraste metodológico entre familias de modelos. La selección final incorporó, además del desempeño predictivo, criterios de interpretabilidad, calibración de probabilidades, costo computacional, reproducibilidad, estabilidad y viabilidad de despliegue. Bajo estos criterios, se seleccionó LinearSVC + TF-IDF calibrado mediante sigmoid, por ofrecer un equilibrio más adecuado para la aplicación masiva sobre la base depurada de PQRD.

4.6 Ajuste de hiperparámetros

Una vez realizada la validación inicial, se ajustaron los hiperparámetros de los modelos con el propósito de mejorar el desempeño de las configuraciones candidatas y controlar el riesgo de sobreajuste. Esta etapa se desarrolló únicamente con datos de entrenamiento y validación interna, manteniendo aislado el conjunto TEST para la evaluación final.

El ajuste se realizó de acuerdo con la naturaleza de cada modelo. En los modelos clásicos basados en TF-IDF se evaluaron parámetros de regularización, suavizamiento, ponderación de clases y control de complejidad. En los modelos de ensamble se revisaron parámetros como número de estimadores, profundidad y regularización. En los enfoques basados en embeddings y transformers, el ajuste se concentró en el clasificador final o en el proceso de fine-tuning.

Durante el proceso se mantuvieron controles para evitar fuga de información. En particular, el vectorizador TF-IDF se ajustó únicamente dentro del conjunto de entrenamiento de cada partición o fold. Asimismo, cuando se aplicó balanceo o ponderación de clases, este procedimiento se realizó solo sobre los datos de entrenamiento correspondientes.

La Tabla 4.6 resume los principales hiperparámetros evaluados por modelo.

Modelo o enfoque	Hiperparámetros evaluados	Configuración	Esquema validación / ajuste	Criterio de selección
ComplementNB / MultinomialNB	alpha: 0,01; 0,05; 0,10; 0,25; 0,50; 1,00; 2,00. fit_prior: True / False. En ComplementNB también norm: False / True.	42 en total	StratifiedGroupKFold de 5 folds con grupo anti-fuga por texto normalizado	F1 macro, MCC y accuracy
LinearSVC + TF-IDF	C: 0,01; 0,05; 0,10; 0,25; 0,50; 1,00; 2,00; 5,00. loss: hinge / squared_hinge. max_iter: 10.000 / 20.000. class_weight: balanced.	32	StratifiedGroupKFold de 5 folds con grupo anti-fuga por texto normalizado	F1 macro, MCC y accuracy
LogisticRegression + TF-IDF	solver: lbfgs, newton-cg, saga. C: 0,10; 1,00; 3,00. max_iter: 3.000. multi_class: multinomial. Parámetros base: penalty='l2', class_weight='balanced'.	9	Validación cruzada sobre TRAIN con control de convergencia	Selection score, F1 macro, MCC y convergencia
LightGBM / XGBoost + TF-IDF	En XGBoost: n_estimators: 250, 300, 400; learning_rate: 0,03 / 0,05; max_depth: 4 / 6; min_child_weight: 1 / 3; subsample: 0,85 / 0,90; colsample_bytree: 0,80 / 0,90; gamma: 0,00 / 0,30; reg_lambda: 1,00 / 2,00; reg_alpha: 0,00.	8	Validación temporal expansiva	F1 macro, MCC y desempeño temporal
RandomForestClassifier + TF-IDF	n_estimators: 200, 300, 400. max_depth: None, 30, 40. min_samples_leaf: 1, 2, 4. max_features: sqrt, log2, 0,5. Parámetros base:class_weight='balanced_subsample', criterion='gini'.	Según grilla definida	StratifiedGroupKFold de 5 folds	F1 macro, MCC y accuracy
SBERT + LinearSVC	C: 0,01; 0,10; 1,00; 3,00; 10,00. class_weight: None / balanced. max_iter: 5.000.	10	StratifiedKFold de 5 folds sobre embeddings SBERT	F1 macro, MCC y accuracy
BETO + fine-tuning	learning_rate: 2e-5 / 3e-5. num_train_epochs: 2 / 3. per_device_train_batch_size: 8. per_device_eval_batch_size: 16. weight_decay: 0,01 / 0,05. warmup_ratio: 0,10. gradient_accumulation_steps: 1.	8	Ajuste sobre partición interna TRAIN / DEV	F1 macro, log loss y accuracy

Tabla 4.6. Hiperparámetros evaluados durante el ajuste de modelos (Elaboración propia)

La Tabla 4.6, presenta el ajuste que permitió evaluar configuraciones específicas por familia de modelos. La comparación priorizó métricas adecuadas para un problema multiclase y desbalanceado, especialmente F1 macro y MCC. Adicionalmente, la desviación estándar permitió revisar la estabilidad de los resultados

entre folds o cortes de validación.

Para sintetizar los resultados del ajuste, cada configuración fue evaluada mediante métricas promedio y desviación estándar. En particular, se priorizaron F1 macro y MCC, debido al desbalance entre las categorías R1–R7. La desviación estándar permitió observar la variabilidad entre folds o cortes de validación, evitando seleccionar configuraciones con buen desempeño promedio pero baja estabilidad.

La Tabla 4.7 muestra la mejor configuración identificada por modelo y el resultado de validación asociado.

Modelo o enfoque	Mejor configuración seleccionada	F1 macro validación	MCC validación	Criterio de selección
ComplementNB	alpha=0.01, fit_prior=True, norm=True	0,7141 ± 0,1354	0,7026 ± 0,1355	F1 macro, MCC y accuracy
MultinomialNB	alpha=0.1, fit_prior=True	0,6905 ± 0,1900	0,7069 ± 0,2027	F1 macro, MCC y accuracy
LinearSVC + TF-IDF	C=0.5, loss=squared_hinge, max_iter=10000	0,6325 ± 0,1288	0,6237 ± 0,1997	F1 macro, MCC y accuracy
LogisticRegression + TF-IDF	C=0.1, class_weight=balanced, max_iter=3000, multi_class=multinomial, penalty=l2, s olver=lbfgs	0,6043 ± 0,1170	0,5879 ± 0,2496	Selection score, F1 macro, MCC y convergencia
RandomForest + TF-IDF	n_estimators=300, max_depth=None, min_samples_leaf=2, max_features=sqrt, class_weight=balanced_subsample	0,5060	0,3454	F1 macro, MCC y accuracy
SBERT + LinearSVC	C=10.0, class_weight=balanced, max_iter=5000	0,9906 ± 0,0084	0,9926 ± 0,0051	F1 macro, MCC y accuracy
BETO + fine-tuning	learning_rate=3e-05, num_train_epochs=3, batch_size=8, weight_decay=0.01, warmup_ratio=0.1	0,7795	0,6588	F1 macro, log loss y accuracy

Tabla 4.7. Mejores configuraciones obtenidas durante el ajuste de hiperparámetros (Elaboración propia)

La Tabla 4.7 muestra que, dentro de los modelos Naive Bayes, ComplementNB obtuvo el mayor F1 macro de validación, con 0,7141 ± 0,1354, mientras que MultinomialNB presentó un MCC ligeramente superior, con 0,7069 ± 0,2027. En los modelos lineales basados en TF-IDF, LinearSVC superó a LogisticRegression, con un F1 macro de 0,6325 ± 0,1288 frente a 0,6043 ± 0,1170, y un MCC de 0,6237 ± 0,1997 frente a 0,5879

± 0,2496.

El mejor desempeño interno se observó en SBERT + LinearSVC, con un F1 macro de 0,9906 ± 0,0084 y un MCC de 0,9926 ± 0,0051. Sin embargo, estos resultados debían contrastarse posteriormente con la evaluación final, la calibración y la viabilidad de escalamiento. Por su parte, BETO + fine-tuning alcanzó un F1 macro de 0,7795 y un MCC de 0,6588, mientras que RandomForest + TF-IDF presentó el menor desempeño reportado, con F1 macro de 0,5060 y MCC de 0,3454.

En conjunto, el ajuste de hiperparámetros permitió reducir el conjunto de alternativas a configuraciones candidatas por familia de modelos. No obstante, estos resultados corresponden a validación interna y no constituyen por sí solos la selección final. Por esta razón, las configuraciones seleccionadas fueron evaluadas posteriormente mediante calibración, desempeño en TEST, costo computacional y viabilidad de despliegue sobre la base completa de PQRD.

4.7 Calibración de probabilidades

Después de comparar y ajustar los modelos, se realizó la calibración de probabilidades del modelo LinearSVC + TF-IDF, debido a que esta configuración fue priorizada para las etapas posteriores de evaluación final y despliegue. Esta etapa fue necesaria porque LinearSVC entrega una decisión de clasificación, pero no genera probabilidades calibradas de forma nativa.

Para el propósito del proyecto, contar con probabilidades era relevante porque permitía no solo asignar una categoría de riesgo operacional a cada PQRD, sino también estimar el nivel de confianza asociado a cada predicción. Esto resultaba útil para el despliegue posterior, ya que facilitaba identificar registros con alta confianza predictiva y diferenciar aquellos casos en los que el modelo presentaba mayor incertidumbre.

La calibración se realizó comparando los métodos sigmoid e isotonic. Además de las métricas de clasificación, se consideraron métricas de calidad probabilística, como log loss, Brier score y ECE, junto con controles metodológicos para evitar probabilidades degeneradas o excesivamente extremas.

Método de calibración	F1 macro	MCC	Log loss	Brier score	ECE	Probabilidades casi binarias	Guardrails	Decisión
sigmoid	0,5504	0,2924	1,0506	0,6222	0,3978	0,0%	Cumple	Seleccionado
isotonic	0,5504	0,2924	0,8169	0,5472	0,1382	65,0%	No cumple	No seleccionado

Tabla 4.8: Resultados de calibración de probabilidades del modelo LinearSVC + TF-IDF (Elaboración propia)

Como se visualiza en la Tabla 4.8, ambos métodos conservaron el mismo desempeño clasificatorio en F1 macro y MCC. La principal diferencia se presentó en las métricas probabilísticas. El método isotonic obtuvo menor log loss (0,8169) que sigmoid (1,0506), lo que inicialmente sugería una mejor calidad probabilística.

Sin embargo, la selección no se basó únicamente en log loss. El método isotonic generó un porcentaje elevado de probabilidades casi binarias (65,0%), por lo que no superó los controles definidos para evitar

salidas probabilísticas degeneradas. En contraste, el método sigmoid cumplió los controles metodológicos y no presentó probabilidades casi binarias, por lo que fue seleccionado como método de calibración del modelo final.

Una vez seleccionado el método de calibración, se consolidó la configuración final del modelo priorizado para el proyecto. La Tabla 4.9 resume los componentes principales del modelo calibrado, las métricas consideradas en su selección y su uso posterior en la inferencia masiva sobre la base de PQRD.

Componente	Configuración seleccionada	Justificación / uso en el proyecto
Modelo final	LinearSVC + TF-IDF	Se seleccionó por su equilibrio entre desempeño predictivo, F1 macro, MCC, costo de inferencia y viabilidad para ejecutar inferencia masiva.
Representación textual	TF-IDF sobre texto normalizado	Permite representar los textos como vectores dispersos, adecuados para modelos lineales de clasificación multiclase.
Clasificador	Linear SVC	Modelo lineal robusto para espacios de alta dimensionalidad, especialmente útil en problemas de clasificación textual.
Calibración de probabilidades	Sigmoid	Permite transformar las salidas del clasificador en probabilidades calibradas para cada clase de riesgo operacional.
Conjunto de evaluación	TEST	La selección se soportó en el desempeño observado sobre el conjunto de prueba, manteniendo separación frente al entrenamiento y validación.
Métricas priorizadas	F1 macro, MCC, log loss y viabilidad de escalamiento	Se priorizaron métricas que permiten evaluar balance entre clases, desempeño global y utilidad operativa del modelo.
Uso posterior	Inferencia masiva sobre PQRD 2017–2024	El modelo fue usado para clasificar la base depurada y relacionar las PQRD con categorías de riesgo operacional.

Tabla 4.9: Configuración del modelo final y calibración (Elaboración propia)

En la Tabla 4.9, la calibración no reemplaza la evaluación del modelo, sino que la complementa. Mientras las métricas de clasificación permiten valorar el desempeño predictivo, la calibración aporta una medida de confianza sobre las probabilidades asignadas a cada predicción.

En consecuencia, el modelo LinearSVC + TF-IDF calibrado con sigmoid quedó preparado para las etapas posteriores de evaluación final y despliegue. Esta calibración permitió que los resultados no se limitaran a una etiqueta de riesgo operacional, sino que incorporaran una medida de confianza útil para interpretar las predicciones y priorizar análisis sobre la base completa de PQRD.

4.8 Selección del modelo final

Con base en los resultados obtenidos durante el entrenamiento, validación, ajuste de hiperparámetros y calibración, se seleccionó como modelo final la combinación LinearSVC + TF-IDF calibrado mediante sigmoid. Esta decisión no se basó en una única métrica, sino en el equilibrio observado entre desempeño predictivo, calidad probabilística, costo computacional, interpretabilidad y viabilidad para procesar grandes volúmenes de PQRD.

La Tabla 4.10 se consolida la comparación integral de los modelos evaluados, considerando métricas de clasificación, log loss, costo computacional, capacidad de escalamiento y decisión final.

	Modelo	Calibración	N test	Accuracy	Precisión macro	Recall macro	F1 macro	MCC	Log loss	Costo	Escalamiento	Decisión
1	LinearSVC + TF-IDF	sigmoid	0.8777	0.8880	0.8443	0.8512	0.8460	0.3024	Bajo	Alta	Seleccionado
2	LogisticRegression + TF-IDF	No disponible	0.8826	0.8094	0.8109	0.7972	0.8556	0.7880	Bajo	Alta	No seleccionado
3	MultinomialNB	sigmoid	0.4898	0.5637	0.7423	0.5878	0.3392	1.6249	Muy bajo	Alta	No seleccionado
4	SBERT + LinearSVC	isotonic	0.8161	0.6624	0.5575	0.5698	0.7287	0.5033	Bajo	Alta	No seleccionado
5	RandomForestClassifier	isotonic	0.5799	0.5934	0.5652	0.5424	0.5127	1.8125	Medio/Alto	Media	No seleccionado
6	BETO + fine-tuning	temperature_scaling	0.5333	0.6049	0.4391	0.4718	0.3098	1.0779	Alto	Baja	No seleccionado
7	SBERT + LogisticRegression	sigmoid	0.2897	0.4908	0.3795	0.4083	0.3229	2.7512	Bajo	Alta	No seleccionado

Tabla 4.10: Comparación integral de modelos evaluados para la selección final (Elaboración propia)

La Tabla 4.10, muestra que LogisticRegression + TF-IDF obtuvo una accuracy ligeramente superior a LinearSVC + TF-IDF (0,8826 frente a 0,8777) y un MCC marginalmente mayor (0,8556 frente a 0,8460). Sin embargo, LinearSVC + TF-IDF presentó mejor desempeño en métricas macro, con precisión macro de 0,8880, recall macro de 0,8443 y F1 macro de 0,8512, lo que evidenció un mejor balance entre clases.

Dado el desbalance de las categorías R1–R7, la selección no podía basarse únicamente en accuracy. Por ello, se priorizó el equilibrio entre F1 macro, MCC y calidad probabilística. En este último aspecto, LinearSVC + TF-IDF obtuvo el menor log loss (0,3024), lo que favoreció su uso en despliegue al permitir asociar cada predicción con una medida de confianza.

Los demás modelos presentaron limitaciones más evidentes en desempeño, estabilidad o viabilidad computacional. Por ejemplo, MultinomialNB alcanzó un F1 macro de 0,5878, RandomForestClassifier de 0,5424 y BETO + fine-tuning de 0,4718, además de mayores restricciones de escalamiento en algunos casos.

En consecuencia, se seleccionó LinearSVC + TF-IDF calibrado con sigmoid como modelo final, al ofrecer el mejor equilibrio entre desempeño predictivo, interpretabilidad, calidad probabilística, bajo costo computacional y viabilidad para aplicar inferencia masiva sobre la base depurada de PQRD 2017–2024.

5. EVALUACIÓN DEL MODELO SELECCIONADO

La evaluación del modelo tuvo como objetivo medir el desempeño de LinearSVC + TF-IDF calibrado con sigmoid sobre el conjunto de prueba. En esta fase se verificó si el modelo seleccionado clasificaba adecuadamente las PQRD en las categorías de riesgo operacional R1–R7 y si su desempeño era suficiente para sustentar su aplicación posterior sobre la base depurada completa.

La evaluación consideró métricas globales, métricas por clase, matriz de confusión y análisis de confianza de las predicciones. Esta revisión fue necesaria porque el problema no solo requería una buena exactitud general, sino también un comportamiento equilibrado entre clases, especialmente por el desbalance identificado en el corpus etiquetado.

5.1 Métricas globales de desempeño

Las métricas globales permitieron evaluar el comportamiento general del modelo sobre el conjunto TEST, para ello, se calcularon indicadores como accuracy, precisión macro, recall macro, F1 macro y coeficiente de correlación de Matthews (MCC). Estas métricas permiten analizar el desempeño del modelo desde diferentes perspectivas.

La exactitud mide la proporción general de predicciones correctas. Sin embargo, debido al desbalance entre clases, esta métrica no fue suficiente por sí sola. Por esta razón, se dio especial importancia al F1 macro, dado que calcula el desempeño promedio entre clases sin favorecer únicamente a las categorías con mayor número de registros. También se consideró el MCC, porque resume la calidad global de la clasificación en problemas multiclase y desbalanceados.

Métrica / elemento	Valor	Interpretación
Modelo final	LinearSVC + TF-IDF	Clasificador lineal con representación TF-IDF.
Método de calibración	Sigmoid	Transforma las salidas del modelo en probabilidades calibradas.
Registros evaluados en TEST	826	Observaciones utilizadas para la evaluación final.
Accuracy	0,8777	Proporción total de clasificaciones correctas.
Precision macro	0,8880	Precisión promedio entre clases, sin ponderar por soporte.
Recall macro	0,8443	Capacidad promedio para recuperar correctamente cada clase.
F1 macro	0,8512	Balance entre precisión y recall macro.
MCC	0,8460	Medida robusta de desempeño global en escenarios multiclase y desbalanceados.
Log loss	0,3024	Mide la calidad probabilística; valores menores indican mejor calibración.

Tabla 5.1: Métricas globales del modelo seleccionado (Elaboración propia)

Como se observa en la Tabla 5.1, el modelo LinearSVC + TF-IDF calibrado con sigmoid obtuvo un desempeño adecuado sobre el conjunto TEST, con un accuracy de 0,8777, F1 macro de 0,8512 y MCC de 0,8460. Estos resultados indican que el modelo no solo alcanzó una alta proporción de clasificaciones correctas, sino que también mantuvo un comportamiento equilibrado entre las categorías de riesgo operacional.
El valor de F1 macro es especialmente relevante porque permite evaluar el desempeño promedio entre las clases R1–R7 sin favorecer las categorías con mayor número de registros. A su vez, el MCC de 0,8460 respalda la calidad global de la clasificación en un escenario multiclase y desbalanceado.

Finalmente, el log loss de 0,3024 indica una adecuada calidad probabilística después de la calibración, aspecto importante para el despliegue posterior, ya que permite asociar cada predicción con una medida de confianza. Estos resultados sirvieron como primer punto de validación del modelo seleccionado y fueron complementados con el análisis por clase, la matriz de confusión y la revisión de la confianza de las predicciones.

5.2 Evaluación por clase de riesgo operacional

Además de las métricas globales, se evaluó el desempeño del modelo para cada una de las categorías de riesgo operacional R1–R7. Este análisis fue necesario porque la muestra etiquetada presenta desbalance entre clases, lo que puede hacer que el modelo tenga mejor desempeño en categorías con mayor número de registros y mayores dificultades en clases con menor representación.

Para esta evaluación se revisaron tres métricas principales por clase: precisión, recall y F1-score. La precisión indica qué tan confiables fueron las predicciones asignadas a cada categoría; el recall mide qué proporción de los casos reales de cada clase fue recuperada correctamente por el modelo; y el F1-score resume el equilibrio entre precisión y recall.

Los resultados por clase se ven en la Tabla 5.2.

Clase	Categoría de riesgo operacional	Precision	Recall	F1-score	Support
R1	Acceso y oportunidad a servicios	0.9000	0.8773	0.8885	277
R2	Autorizaciones y barreras administrativas	0.5610	0.9324	0.7005	74
R3	Suministro de medicamentos y tecnologías en salud	0.9495	0.9079	0.9283	228
R4	Calidad técnico-asistencial y seguridad del paciente	1.0000	0.6585	0.7941	82
R5	Atención al usuario, información y trato	0.8750	0.6364	0.7368	22
R6	Temas económicos, facturación y recobros	1.0000	0.8980	0.9462	49
R7	Otras situaciones no clasificadas en las categorías anteriores	0.9307	1.0000	0.9641	94
Promedio macro	Promedio no ponderado entre clases	0.8880	0.8443	0.8512	826
Promedio ponderado	Promedio ponderado por soporte de cada clase	0.9020	0.8777	0.8812	826

Tabla 5.2: Métricas por clase de riesgo operacional (Elaboración propia)

Como se evidencia en la Tabla 5.2, el modelo presentó mejores desempeños en R3, R6 y R7, con F1-score de 0,9283, 0,9462 y 0,9641, respectivamente, lo que evidencia un buen balance entre precisión y recall. La clase R1 también mostró un desempeño sólido, con F1-score de 0,8885 y el mayor soporte del conjunto TEST (277 registros).

En contraste, R4 y R5 registraron menores niveles de recall, con 0,6585 y 0,6364, lo que indica que parte de sus casos reales fue clasificada en otras categorías. Este resultado debe interpretarse con cautela en R5, debido a su bajo soporte en TEST (22 registros). Por su parte, R2 presentó un recall alto (0,9324), pero una precisión baja (0,5610), lo que sugiere que el modelo recuperó la mayoría de sus casos reales, aunque también asignó a esta clase registros pertenecientes a otras categorías semánticamente cercanas.

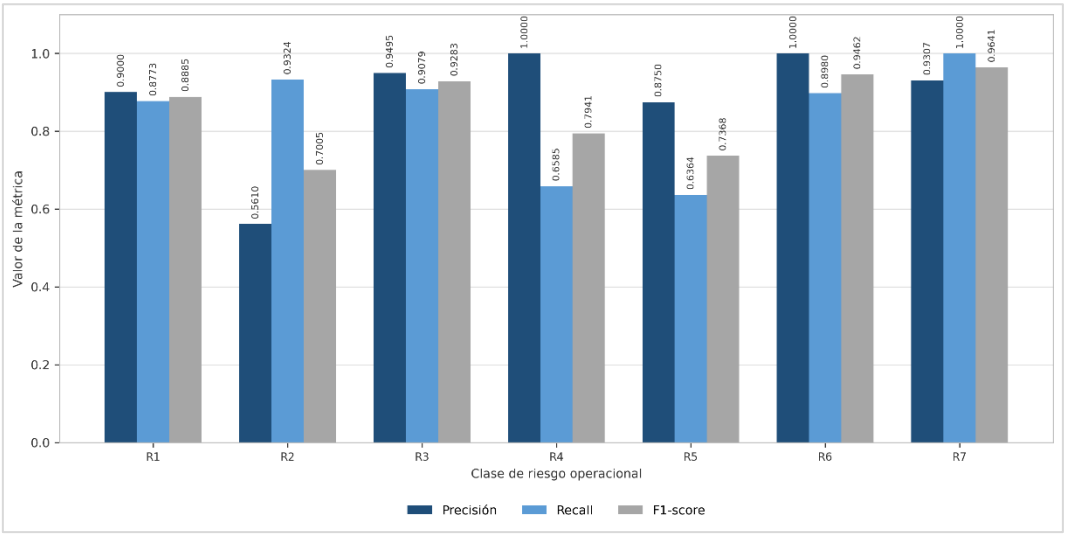


Figura 5.1. Comparación de precisión, recall y F1-score por clase de riesgo (Elaboración propia)

Como se indica en la Figura 5.3, la comparación visual de precisión, recall y F1-score facilita la identificación de clases con desempeño más sólido y de categorías que pueden requerir mayor cantidad de datos etiquetados, ajuste de reglas de clasificación o revisión del proceso de etiquetado en futuros ciclos del proyecto.

Cómo conclusión, la evaluación por clase permitió pasar de una lectura general del desempeño a una revisión más detallada del comportamiento del modelo frente a cada riesgo operacional. Esta información fue clave para interpretar las fortalezas y limitaciones del modelo antes de aplicarlo sobre la base depurada completa.

5.3 Matriz de confusión

Para complementar las métricas globales y por clase, se construyó la matriz de confusión del modelo seleccionado. Esta matriz permitió comparar las etiquetas reales de las PQRD con las etiquetas predichas, identificando tanto los aciertos como los errores de clasificación entre las categorías R1–R7.

La matriz de confusión absoluta se presenta en la Figura 5.4. En esta figura, la diagonal principal representa los casos correctamente clasificados, mientras que los valores ubicados por fuera de la diagonal muestran las confusiones entre clases.

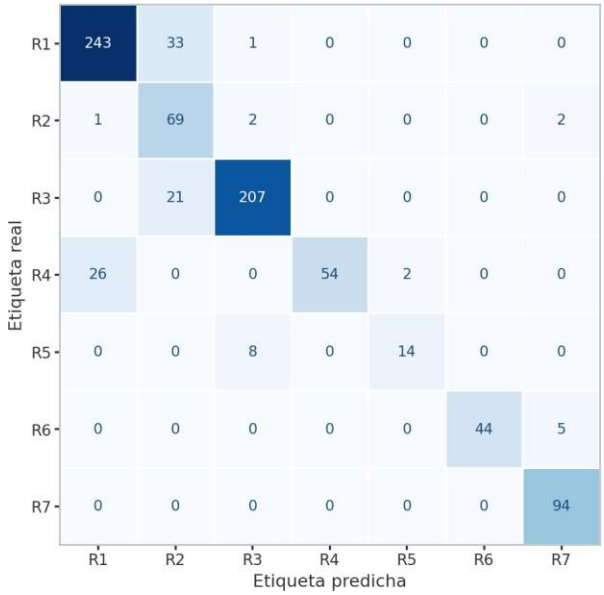


Figura 5.2: Matriz de confusión absoluta del modelo seleccionado (Elaboración propia)

En la Figura 5.4, se presenta como el modelo clasificó correctamente la mayoría de los registros en varias categorías. Los mayores aciertos se observaron en R1, con 243 registros correctamente clasificados; R3, con 207; y R7, con 94. También se identificaron confusiones relevantes, especialmente entre R1 y R2, donde 33 registros reales de R1 fueron clasificados como R2, y entre R3 y R2, donde 21 registros reales de R3 fueron clasificados como R2.

Adicionalmente, se construyó una matriz de confusión normalizada, presentada en la Figura 5.5. Esta versión permite interpretar los resultados en proporciones, lo cual facilita la comparación entre clases con diferente número de registros.

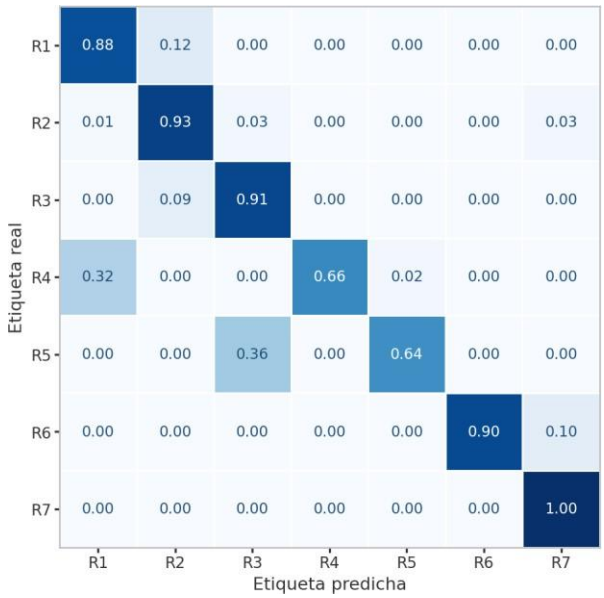


Figura 5.3: Matriz de confusión normalizada del modelo seleccionado (Elaboración propia).

La Figura 5.5 confirma que las clases con mejor recuperación fueron R7, con recall de 1,00; R2, con 0,93; R3, con 0,91; y R6, con 0,90. En contraste, las clases con menor recuperación fueron R4 y R5, con valores normalizados de 0,66 y 0,64, respectivamente. En R4, el principal error fue la clasificación de registros como R1 (0,32), mientras que en R5 el principal error fue la clasificación como R3 (0,36).

En conjunto, ambas matrices permitieron identificar los principales patrones de error del modelo. Las confusiones más relevantes se concentraron entre categorías semánticamente cercanas, como acceso, autorizaciones, suministro de tecnologías y calidad de la atención. Esta revisión es importante para interpretar las limitaciones del modelo y orientar futuros ajustes en el etiquetado, el entrenamiento o la definición de reglas complementarias de validación.

5.4 Análisis de errores y limitaciones del modelo

El análisis de errores permitió revisar los casos en los que el modelo no coincidió con la etiqueta real del conjunto de prueba. Esta revisión fue útil para identificar si las fallas de clasificación estaban asociadas con similitudes entre categorías, desbalance de clases o ambigüedad en los textos utilizados.

Una de las principales limitaciones fue la cercanía temática entre algunas categorías. Algunas PQRD pueden mencionar simultáneamente acceso, autorizaciones, medicamentos, continuidad de la atención o calidad del servicio, lo que puede generar confusión entre clases. La matriz de confusión evidenció errores relevantes, especialmente registros reales de R1 y R3 clasificados como R2, así como menor recuperación en R4 y R5.

Otra limitación corresponde al desbalance del corpus etiquetado. Las categorías con menor número de registros pueden ser más difíciles de aprender para el modelo, por lo que la interpretación no debe basarse únicamente en métricas globales, sino también en las métricas por clase y en los patrones de confusión.

También debe considerarse que el texto usado para el modelado provino de campos asociados al macromotivo, motivo general y motivo específico de la PQRD. Aunque estos campos aportan información relevante, no siempre contienen el detalle completo del caso reportado. Por tanto, los resultados del modelo deben entenderse como un insumo analítico y no como una valoración definitiva de cada PQRD individual.

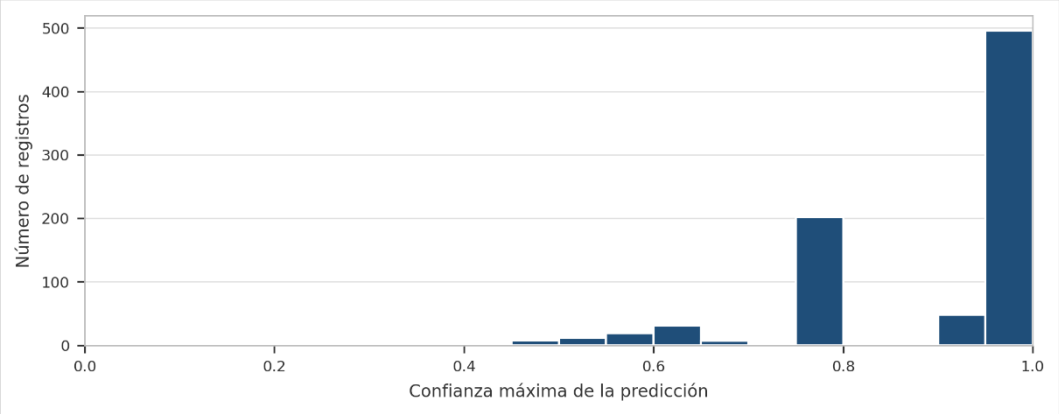


Figura 5.4: Distribución de confianza de las predicciones en TEST (Elaboración propia)

La confianza promedio de las predicciones fue de 0,8926 y la mediana de 0,9695. Además, 543 registros del conjunto TEST, equivalentes al 65,7%, tuvieron una confianza muy alta, igual o superior a 0,90. Sin embargo, 240 registros (29,1%) se ubicaron en confianza media y 39 registros (4,7%) en confianza baja, por lo que estos casos podrían requerir revisión adicional en usos operativos.

En resumen, el modelo es útil para clasificar y analizar grandes volúmenes de PQRD, pero sus resultados deben interpretarse considerando el desbalance de clases, la similitud temática entre riesgos, las limitaciones del etiquetado y la información disponible en los campos textuales.

5.5 Justificación del modelo seleccionado

A partir de la evaluación realizada, se confirmó la selección de LinearSVC + TF-IDF calibrado con sigmoid como alternativa final para la clasificación de riesgos operacionales en las PQRD. Esta decisión se sustentó en el equilibrio entre desempeño predictivo, interpretabilidad, calidad probabilística y viabilidad computacional para procesar grandes volúmenes de información.

El modelo resultó adecuado para el problema abordado porque la representación TF-IDF permite transformar los textos en matrices dispersas de alta dimensión, mientras que LinearSVC ofrece un clasificador lineal eficiente y robusto para tareas de clasificación textual. Esta combinación fue conveniente frente a la necesidad de aplicar posteriormente el modelo sobre millones de registros.

La selección también consideró que modelos más complejos, como los basados en transformers, pueden capturar relaciones semánticas más profundas, pero requieren mayores recursos computacionales, tiempos de entrenamiento más altos y una gestión más exigente del ajuste. En contraste, el modelo seleccionado ofreció una solución más viable, reproducible y suficiente para el objetivo del proyecto

Adicionalmente, la calibración mediante sigmoid permitió complementar la etiqueta predicha con una medida de confianza. De esta manera, cada PQRD clasificada no solo cuenta con una categoría de riesgo operacional, sino también con una probabilidad asociada que facilita la interpretación y priorización de resultados.

A continuación, se resume la consistencia probabilística del modelo la Tabla 5.3.

Indicador	Resultado	Interpretación
Modelo evaluado	LinearSVC + TF-IDF	Modelo final seleccionado para evaluación e inferencia masiva.
Método de calibración	sigmoid	Calibración de probabilidades mediante método sigmoide para obtener salidas probabilísticas interpretables.
Registros evaluados en TEST	826	Tamaño del conjunto de prueba utilizado para la evaluación final del modelo.
Clases evaluadas	R1, R2, R3, R4, R5, R6 y R7	Categorías de riesgo operacional consideradas en el modelo.
Suma mínima de probabilidades	1,0000	Las probabilidades por registro suman aproximadamente 1.
Suma máxima de probabilidades	1,0000	No se identifican desalineaciones relevantes en la suma de probabilidades.
Error máximo absoluto frente a 1	2,22e-16	El error numérico es prácticamente nulo, consistente con una salida probabilística bien formada.
Log loss	0,302445	Mide la calidad de las probabilidades calibradas; valores más bajos indican mejor ajuste probabilístico.
Brier score multiclase	0,159337	Resume el error cuadrático de las probabilidades predichas frente a la clase observada; menor es mejor.
Confianza promedio	0,892578	Promedio de la probabilidad máxima asignada por el modelo en las predicciones del TEST.
Confianza mediana	0,969453	Mediana de la probabilidad máxima, útil para observar el nivel típico de seguridad del modelo.

Tabla 5.3. Resumen de calibración y confianza del modelo seleccionado (Elaboración propia).

Como se observa en la Tabla 5.3, la calibración mediante sigmoid permitió obtener probabilidades consistentes para las siete categorías de riesgo operacional. Las probabilidades por registro sumaron aproximadamente 1, el error máximo frente a esta suma fue prácticamente nulo (2,22e-16) y las métricas probabilísticas, como el log loss de 0,3024 y el Brier score de 0,1593, respaldan la calidad de las salidas calibradas.

En conclusión, el modelo seleccionado fue adecuado para el propósito del proyecto porque permitió clasificar las PQRD de forma eficiente, trazable y con una medida de confianza asociada. Esta evaluación

sirvió como base para avanzar a la siguiente fase, en la cual el modelo fue aplicado sobre la base depurada completa.

6. DESPLIEGUE DEL MODELO Y ANÁLISIS DE RESULTADOS

La fase de despliegue tuvo como objetivo aplicar el modelo seleccionado sobre la base depurada de PQRD y transformar las predicciones en insumos para el análisis de riesgo operacional. En esta etapa, el modelo pasó de la evaluación sobre la muestra etiquetada a la aplicación sobre los registros disponibles en la base final del proyecto.

El despliegue no se limitó a generar una etiqueta de riesgo operacional. También permitió asociar cada predicción con una probabilidad o nivel de confianza, y relacionar los resultados con variables de contexto como año, entidad, motivo y proceso de atención. De esta manera, las salidas del modelo se convirtieron en una base analítica para identificar patrones recurrentes, procesos críticos y posibles focos de intervención.

A continuación, se describe el flujo:

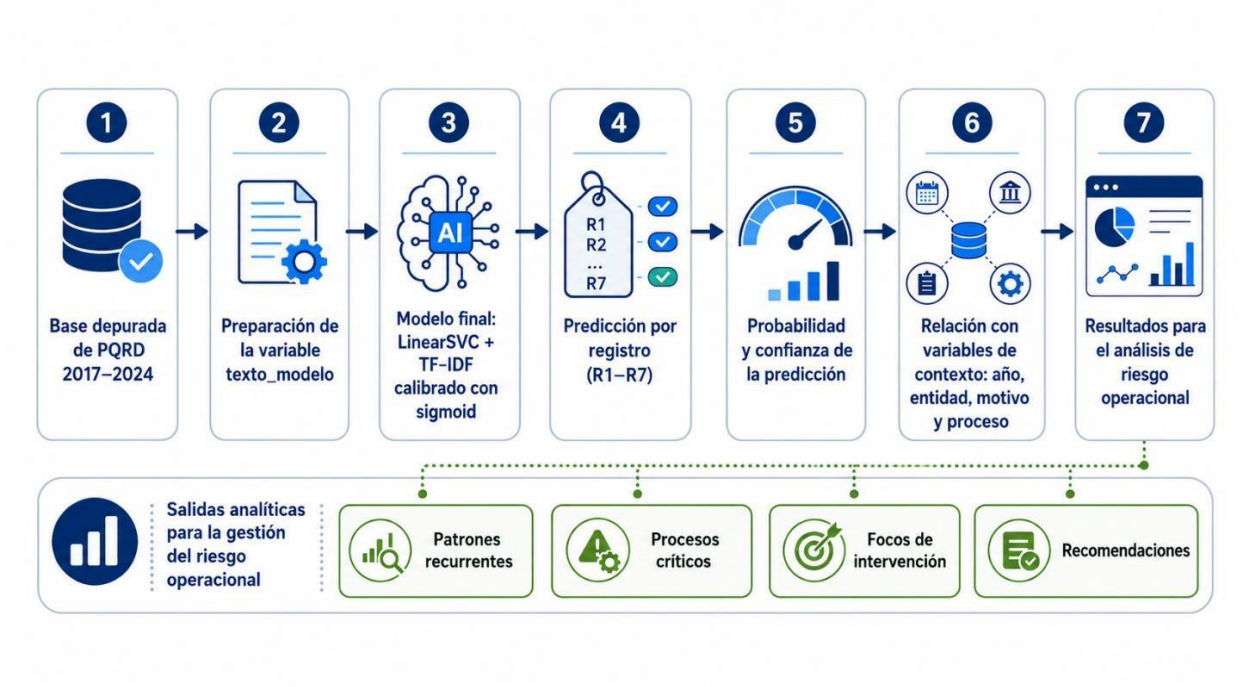


Figura 6.1: Flujo de despliegue sobre la base depurada (Elaboración propia)

La Figura 6.1, describe el despliegue que inició con la preparación de la base depurada, continuó con la aplicación del modelo final calibrado y la generación de predicciones por registro, y finalizó con la consolidación de resultados para el análisis de riesgos operacionales. Estos resultados sirvieron como punto de partida para el análisis de hallazgos, la relación con procesos de prestación de servicios de salud y la formulación de recomendaciones.

6.1 Preparación de la base

Para aplicar el modelo sobre el universo de datos disponible, se utilizó la base depurada del proyecto, conformada por 8.039.810 registros correspondientes al periodo 2017–2024. Esta base corresponde a la versión consolidada de las PQRD, después de los procesos de limpieza, depuración y conservación de las variables requeridas para la inferencia y el análisis posterior.

Antes del despliegue, se verificó que la base contara con los campos necesarios para construir el texto de entrada del modelo. En particular, se mantuvieron las variables relacionadas con macromotivo, motivo general y motivo específico, junto con campos de contexto como identificador del registro, año y entidad. La base preparada conservó la misma lógica de procesamiento utilizada durante la fase de modelado, de manera que los textos fueran normalizados y estructurados en una forma compatible con la representación TF-IDF del modelo final.

El resumen de la base utilizada para el despliegue se presenta en la Tabla 6.1.

Elemento	Valor / configuración	Descripción
Fuente de información	PQRD de la Superintendencia Nacional de Salud	Base pública utilizada como insumo para el análisis textual del proyecto.
Periodo cubierto	2017-2024	Ventana histórica considerada para la consolidación y análisis de registros.
Total, de registros procesados	8.039.810	Número de PQRD disponibles en la base depurada para la fase de inferencia.
Unidad de análisis	Registro individual de PQRD	Cada fila corresponde a una petición, queja, reclamo o denuncia registrada en la base.
Texto utilizado para clasificación	Texto normalizado para modelado	Campo construido a partir de la información temática de macromotivo, motivo general y motivo específico.
Modelo aplicado	LinearSVC + TF-IDF	Modelo final seleccionado por su equilibrio entre desempeño predictivo y viabilidad de escalamiento.
Método de calibración	Sigmoid	Calibración utilizada para obtener probabilidades por clase a partir del clasificador lineal.
Clases de riesgo operacional	R1-R7	Taxonomía de siete categorías definida para clasificar los textos de PQRD.
Salidas despliegue	Clase predicha y probabilidades por clase	Resultados usados para los análisis posteriores de distribución de riesgos y relación riesgo-proceso.

Tabla 6.1: Resumen de la base utilizada para el despliegue (Elaboración propia)

Como se presenta en la Tabla 6.1, el despliegue se realizó sobre la base depurada de PQRD del periodo 2017–2024, con un total de 8.039.810 registros procesados. Para esta fase se aplicó el modelo final LinearSVC + TF-IDF calibrado con sigmoid, generando como salida la categoría de riesgo operacional

predicha y las probabilidades asociadas a cada clase R1–R7.

6.2 Aplicación del modelo sobre la base depurada

Una vez preparada la base, se aplicó el modelo final LinearSVC + TF-IDF calibrado con sigmoid sobre los registros depurados de PQRD. El objetivo fue asignar a cada registro una categoría de riesgo operacional entre R1 y R7, manteniendo la misma lógica de procesamiento definida durante la fase de modelado.

La aplicación del modelo se realizó sobre el texto construido a partir de los campos normalizados de macromotivo, motivo general y motivo específico. Estos textos fueron transformados mediante el vectorizador TF-IDF previamente ajustado durante el entrenamiento, evitando modificar el vocabulario o recalcular los pesos sobre la base completa. De esta forma, se conservaron las mismas condiciones técnicas utilizadas para entrenar y evaluar el modelo seleccionado.

Como resultado, cada registro de la base depurada quedó asociado con una categoría de riesgo operacional predicha y con probabilidades estimadas para cada una de las clases R1–R7. Esta información permitió no solo clasificar los registros, sino también contar con una medida de confianza para interpretar la salida del modelo.

El despliegue permitió pasar de una muestra etiquetada de entrenamiento a una clasificación masiva sobre la base depurada de PQRD. Con ello, las predicciones se convirtieron en un insumo analítico para estudiar la distribución de riesgos operacionales por año, entidad, motivo y, posteriormente, por proceso de atención.

En síntesis, el proceso mantuvo una secuencia controlada: preparación de la base, construcción del texto de entrada, transformación mediante TF-IDF, aplicación del modelo calibrado y consolidación de los resultados para el análisis posterior.

6.3 Consolidación de resultados de clasificación

Después de aplicar el modelo, se consolidaron los resultados de clasificación en una estructura analítica que permitiera consultar, resumir y analizar las predicciones generadas. Esta base integró la información original de las PQRD con la categoría de riesgo operacional asignada por el modelo y las probabilidades estimadas para cada clase.

La consolidación de resultados permitió transformar una salida técnica del modelo en una base interpretable para el análisis de riesgos. En esta etapa se conservaron variables clave como el identificador del registro, año, entidad, macromotivo, motivo general, motivo específico, riesgo predicho, probabilidad máxima y probabilidades por categoría R1–R7.

La distribución general de las PQRD clasificadas por riesgo operacional se muestran en la Tabla 6.2 y en la Figura 6.2.

Clase de riesgo	Categoría de riesgo operacional	Número de PQRD	Participación (%)
R1	Acceso y oportunidad a servicios	3.102.959	38,6
R2	Autorizaciones y barreras administrativas	1.973.150	24,5
R3	Suministro de medicamentos y tecnologías en salud	1.388.697	17,3
R4	Calidad técnico-asistencial y seguridad del paciente	731.637	9,1
R5	Atención al usuario, información y trato	162.187	2,0
R6	Temas económicos, facturación y recobros	331.751	4,1
R7	Otras situaciones no clasificadas en las categorías anteriores	349.429	4,3
Total	Base clasificada	8.039.810	100,0

Tabla 6.2: Distribución de PQRD clasificadas por riesgo operacional (Elaboración propia.)

En conjunto, estas tres categorías concentraron el 80,4% de las PQRD clasificadas, lo que evidencia que la mayor parte de las inconformidades se relaciona con barreras de acceso, trámites administrativos y suministro de tecnologías en salud. En contraste, las categorías R5, R6 y R7 tuvieron una menor participación relativa, con 2,0%, 4,1% y 4,3%, respectivamente.

La Figura 6.2 presenta visualmente la distribución global de las PQRD clasificadas por categoría de riesgo operacional.

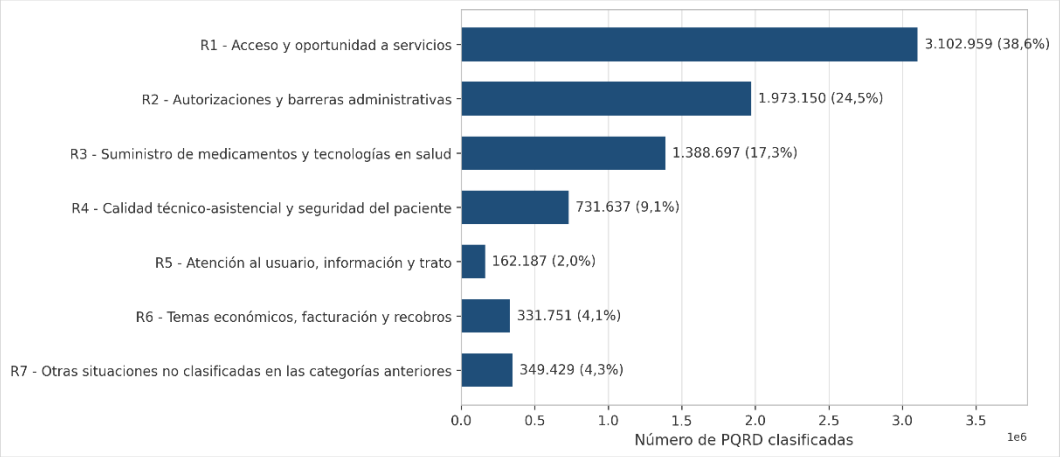


Figura 6.2: Distribución global de PQRD clasificadas por riesgo operacional (Elaboración propia)

Como se observa en la Figura 6.2, la clasificación masiva reproduce una concentración importante en las categorías R1, R2 y R3. Esta lectura permite priorizar los análisis posteriores sobre los riesgos de mayor recurrencia, sin perder de vista las categorías minoritarias, que pueden representar eventos menos frecuentes pero relevantes para la gestión operacional.

Además de la distribución global, se analizó la evolución anual de las categorías clasificadas durante el periodo 2017–2024. Este análisis permite observar cambios en el comportamiento de los riesgos operacionales a lo largo del tiempo.

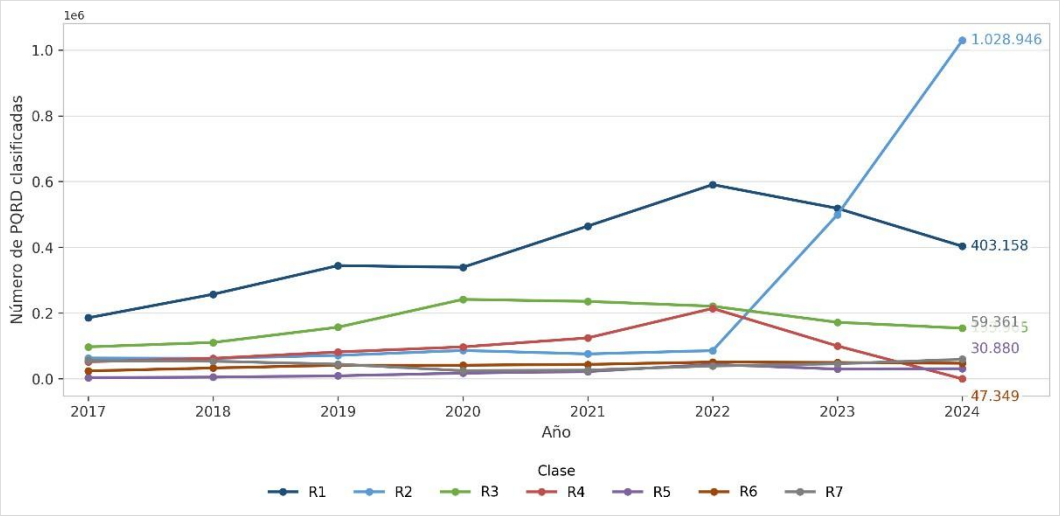


Figura 6.3: Evolución anual de los riesgos operacionales clasificados (Elaboración propia)

La Figura 6.3 muestra cambios relevantes en la evolución anual de las categorías clasificadas. En 2024, la categoría R2 presentó el mayor volumen, con 1.028.946 registros, superando a R1, que cerró el periodo con 403.158 registros. Esto sugiere un aumento importante de los riesgos asociados con autorizaciones y barreras administrativas en el último año analizado.

Por su parte, R3 registró 153.615 casos en 2024, manteniéndose como una categoría relevante, aunque por debajo de R1 y R2. Las demás clases presentaron volúmenes menores, lo que confirma que la mayor concentración de PQRD clasificadas se mantiene en los riesgos asociados con acceso, trámites administrativos y suministro de tecnologías en salud.

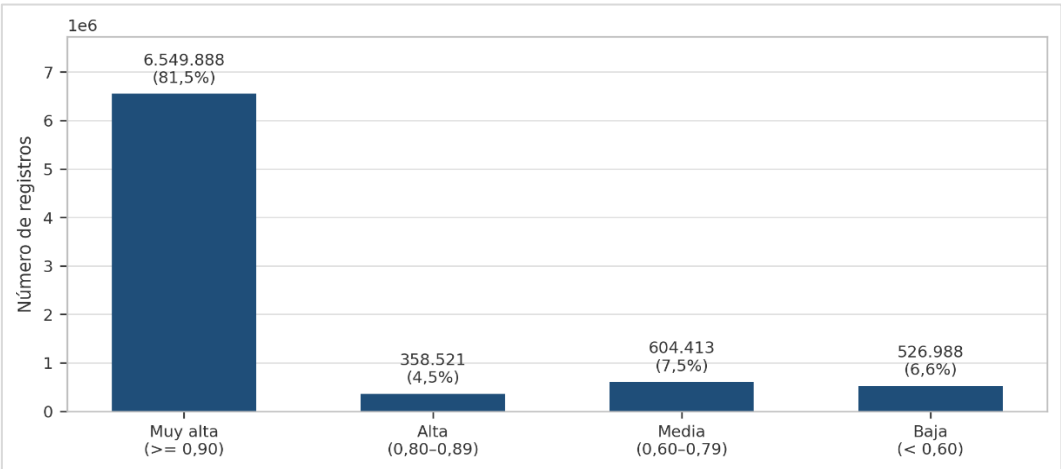


Figura 6.4: Distribución de confianza (Elaboración propia)

Como se ve en la Figura 6.4, una proporción importante de las predicciones se concentró en niveles altos de confianza, especialmente cerca del extremo superior de la escala. Esto indica que, para una parte relevante de la base, el modelo asignó la categoría de riesgo operacional con alta seguridad probabilística.

No obstante, también se observan registros con niveles de confianza intermedios o menores. Estos casos deben interpretarse con mayor cautela, especialmente si se utilizan para análisis operativos específicos, priorización de casos o revisión individual. En este sentido, la confianza de la predicción complementa la etiqueta asignada por el modelo y permite distinguir entre resultados más seguros y resultados que podrían requerir validación adicional.

La consolidación de resultados permitió transformar la inferencia masiva en una base analítica útil para el análisis operativo del proyecto. A partir de los 8.039.810 registros clasificados, fue posible identificar que R1, R2 y R3 concentran el 80,4% de las PQRD, revisar su comportamiento anual y disponer de una medida de confianza asociada a cada predicción. Estos insumos sirvieron como base para relacionar los riesgos clasificados con los procesos de prestación de servicios de salud y avanzar en la identificación de patrones recurrentes.

6.4 Mapa de procesos de la prestación de servicios de salud

Después de consolidar las predicciones del modelo, se diseñó un mapa de procesos de la prestación de servicios de salud. El propósito fue contar con una estructura operativa que permitiera relacionar las categorías de riesgo operacional con las principales etapas del ciclo de atención.

El mapa se construyó como una referencia para interpretar los resultados del modelo desde una perspectiva operativa. Para ello, se definieron procesos asociados con entrada y acceso, verificación de derechos, autorizaciones, agendamiento, admisión, prestación asistencial, entrega de medicamentos y tecnologías, continuidad de la atención, gestión administrativa, información al usuario y mejoramiento continuo.

A continuación, se visualiza el mapa de procesos:

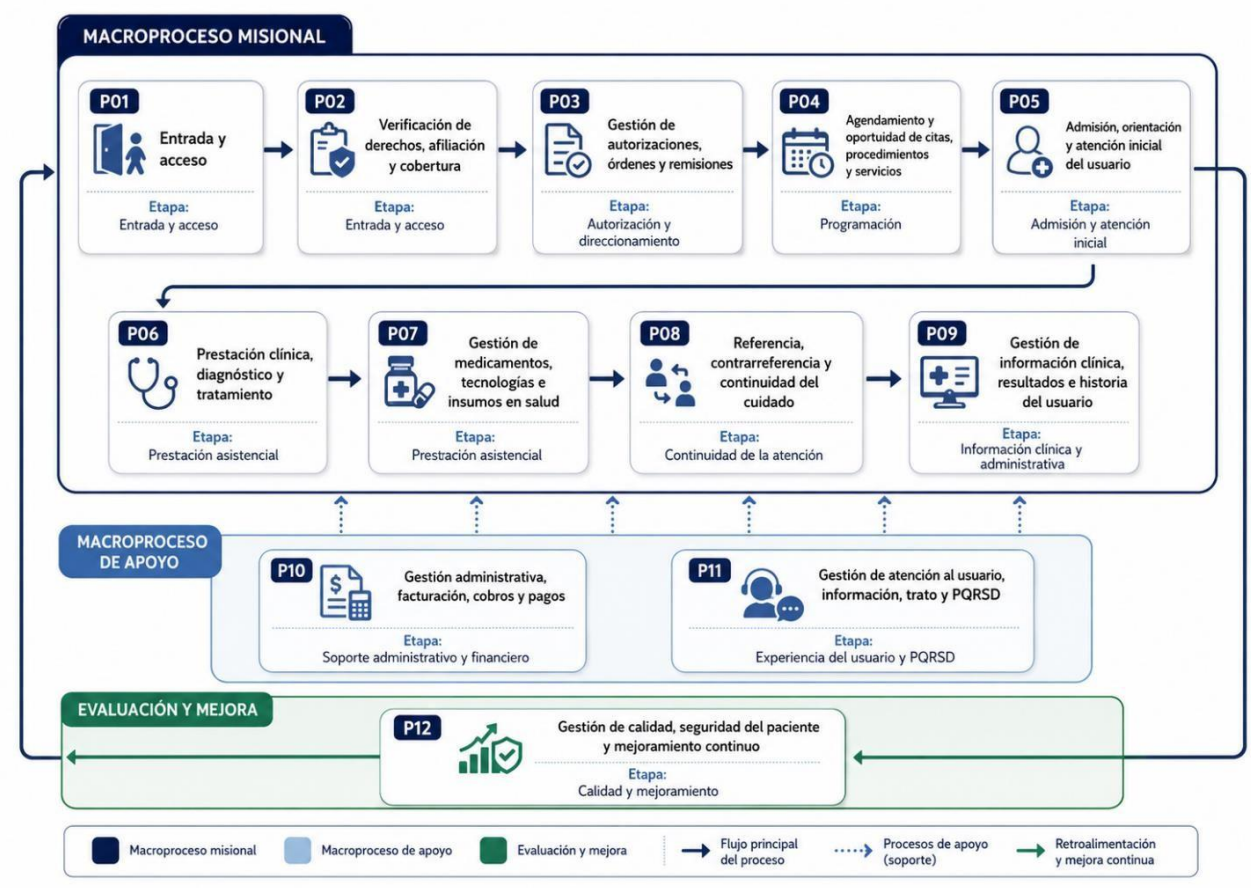


Figura 6.5: Mapa de procesos de la prestación de servicios de salud (Elaboración propia).

La Figura 6.5, muestra el mapa que permite organizar las etapas que puede recorrer un usuario dentro de la prestación de servicios de salud. Esta estructura facilita ubicar los riesgos operacionales clasificados por el modelo dentro de procesos concretos, pasando de una lectura estadística de las PQRD a una lectura operativa orientada a la gestión del riesgo.

El catálogo de procesos utilizado para esta asociación se resume en la Tabla 6.3.

Código	Macroproceso	Etapa del ciclo de atención	Proceso
P01	Misional	Entrada y acceso	Entrada y acceso
P02	Misional	Entrada y acceso	Verificación de derechos, afiliación y cobertura
P03	Misional	Autorización y direccionamiento	Gestión de autorizaciones, órdenes y remisiones
P04	Misional	Programación	Agendamiento y oportunidad de citas, procedimientos y servicios
P05	Misional	Admisión y atención inicial	Admisión, orientación y atención inicial del usuario
P06	Misional	Prestación asistencial	Prestación clínica, diagnóstico y tratamiento
P07	Misional	Prestación asistencial	Gestión de medicamentos, tecnologías e insumos en salud
P08	Misional	Continuidad de la atención	Referencia, contrarreferencia y continuidad del cuidado
P09	Misional	Información clínica y administrativa	Gestión de información clínica, resultados e historia del usuario
P10	Apoyo	Soporte administrativo y financiero	Gestión administrativa, facturación, cobros y pagos
P11	Apoyo	Experiencia del usuario y trato y PQRSD	Gestión de atención al usuario, información, trato y PQRSD
P12	Evaluación y mejora	Calidad y mejoramiento	Gestión de calidad, seguridad del paciente y mejoramiento continuo

Tabla 6.3: Catálogo de procesos de prestación de servicios de salud (Elaboración propia)

La Tabla 6.3 presenta el catálogo utilizado para relacionar las PQRD clasificadas por riesgo operacional con las etapas del ciclo de atención en salud. Este catálogo permitió estructurar la fase de despliegue y preparar los análisis posteriores de relación entre categorías de riesgo y procesos de prestación del servicio.

El mapa de procesos funcionó como un puente entre el resultado del modelo de PLN y la gestión operativa del sector salud. A partir de este insumo fue posible avanzar hacia el análisis de riesgos por proceso, desarrollado en el siguiente apartado.

6.5 Relación entre riesgos operacionales y procesos de atención

Una vez definido el mapa de procesos, se relacionaron las categorías de riesgo operacional predichas por el modelo con los procesos de prestación de servicios de salud, con el fin de identificar en qué etapas del ciclo de atención se concentran las principales señales de riesgo.

La relación se construyó a partir de las categorías R1–R7 y del catálogo de procesos definido en la sección anterior. Para cada riesgo operacional se identificaron los procesos con mayor afinidad, considerando la naturaleza del riesgo, los campos textuales de la PQRD y la lógica operativa de la prestación del servicio.

La matriz de relación entre riesgos operacionales y procesos de atención se relaciona en la Tabla 6.4.

Riesgo operacional	Código proceso	Proceso asignado	Etapa del ciclo de atención	No. PQRD	Participación (%)
R1	P01	Gestión de acceso a servicios de salud	Entrada y acceso	2.963.395	36,8590
R1	P08	Referencia, contrarreferencia y continuidad del cuidado	Continuidad de la atención	131.140	1,6311
R1	P03	Gestión de autorizaciones, órdenes y remisiones	Autorización y direccionamiento	8.424	0,1048
R2	P03	Gestión de autorizaciones, órdenes y remisiones	Autorización y direccionamiento	1.389.946	17,2883
R2	P02	Verificación de derechos, afiliación y cobertura	Entrada y acceso	528.668	6,5776
R2	P08	Referencia, contrarreferencia y continuidad del cuidado	Continuidad de la atención	54.536	0,6783
R3	P07	Gestión de medicamentos, tecnologías e insumos en salud	Prestación asistencial	902.780	11,2289
R3	P06	Prestación clínica, diagnóstico y tratamiento	Prestación asistencial	485.917	6,0449
R4	P03	Gestión de autorizaciones, órdenes y remisiones	Autorización y direccionamiento	395.632	4,9209
R4	P08	Referencia, contrarreferencia y continuidad del cuidado	Continuidad de la atención	257.968	3,2086
R4	P06	Prestación clínica, diagnóstico y tratamiento	Prestación asistencial	64.820	0,8062
R4	P09	Gestión de información clínica, resultados e historia del usuario	Información clínica y administrativa	13.217	0,1644
R5	P01	Gestión de acceso a servicios de salud	Entrada y acceso	69.546	0,8650
R5	P09	Gestión de información clínica, resultados e historia del usuario	Información clínica y administrativa	67.324	0,8374
R5	P11	Gestión de atención al usuario, información, trato y PQRSD	Experiencia del usuario	25.317	0,3149
R6	P10	Gestión administrativa, facturación, cobros y pagos	Soporte administrativo y financiero	196.051	2,4385
R6	P02	Verificación de derechos, afiliación y cobertura	Entrada y acceso	135.700	1,6879
R7	P09	Gestión de información clínica, resultados e historia del usuario	Información clínica y administrativa	306.743	3,8153
R7	P12	Gestión de calidad, seguridad del paciente y mejoramiento continuo	Calidad y mejoramiento	21.690	0,2698
R7	P11	Gestión de atención al usuario, información, trato y PQRSD	Experiencia del usuario	18.192	0,2263
R7	P08	Referencia, contrarreferencia y continuidad del cuidado	Continuidad de la atención	2.804	0,0349

Tabla 6.4: Relación entre riesgos operacionales y procesos de atención (Elaboración propia)

Como se presenta en la Tabla 6.4, la mayor concentración se observó en la relación entre R1 y P01 – Gestión de acceso a servicios de salud, con 2.963.395 PQRD (36,85%). De forma similar, R2 se concentró principalmente en P03 – Gestión de autorizaciones, órdenes y remisiones, con 1.389.946 registros (17,28%), mientras que R3 se asoció sobre todo con P07 – Gestión de medicamentos, tecnologías e insumos en salud, con 902.780 registros (11,22%).En R4, las mayores asociaciones se presentaron en P03

con 395.632 registros (4,92%) y en P08 – Referencia, contrarreferencia y continuidad del cuidado, con 257.968 registros (3,20%). En R5, la distribución fue más equilibrada entre P01 (69.546; 0,86%) y P09 – Gestión de información clínica, resultados e historia del usuario (67.324; 0,83%). Por su parte, R6 se concentró en P10 – Gestión administrativa, facturación, cobros y pagos, con 196.051 registros (2,43%), y R7 principalmente en P09, con 306.743 registros (3,81%).

Estos resultados muestran que los principales focos de riesgo operacional se ubicaron en los procesos de acceso, autorizaciones y gestión de medicamentos, lo que aporta una lectura más operativa de la clasificación obtenida por el modelo.

La distribución total de PQRD clasificadas por proceso de atención se presenta en la **Figura 6.6**.

Figura 6.6: Distribución de PQRD clasificadas por proceso de atención (Elaboración propia)

Como se muestra en la Figura 6.6, los procesos con mayor volumen de registros fueron P01, con 3.032.941; P03, con 1.794.002; y P07, con 902.780. En conjunto, estos tres procesos concentraron 5.729.723 registros, equivalentes al 71,3% de la base clasificada. Esta concentración confirma que las principales señales de riesgo se ubican en el acceso, las autorizaciones y la gestión de medicamentos y tecnologías en salud.

6.6 Análisis de hallazgos y patrones recurrentes de riesgo

A partir de la clasificación de las PQRD y de su relación con los procesos de atención, se analizaron los principales hallazgos y patrones recurrentes de riesgo operacional. El análisis se enfocó en tres dimensiones: la frecuencia de las categorías de riesgo, su comportamiento dentro de los procesos de atención y las combinaciones riesgo–proceso con mayor concentración de registros.

Con base en los resultados obtenidos, la Tabla 6.5 resume los principales hallazgos identificados. Esta síntesis permite pasar de los resultados cuantitativos a una lectura operativa de los riesgos más recurrentes.

Hallazgo / patrón recurrente	Evidencia en los resultados	Proceso(s) asociado(s)	Lectura operativa para la gestión del riesgo
Alta concentración de PQRD en acceso y oportunidad a servicios.	R1 agrupa 3.102.959 PQRD, equivalentes al 38,6% de la base clasificada. La mayor asociación se ubica en P01, con 2.963.395 registros.	P01 Gestión de acceso a servicios de salud; P08 Continuidad del cuidado; P03 Autorizaciones.	El patrón evidencia que las barreras de acceso y oportunidad son la principal señal operacional del análisis. Esto orienta la priorización de controles sobre disponibilidad de servicios, oportunidad de atención y seguimiento a demanda no resuelta.
Concentración relevante en autorizaciones y barreras administrativas.	R2 registra 1.973.150 PQRD (24,5%). La mayor asociación se ubica en P03, con 1.389.946 registros, y en P02, con 528.668 registros.	P03 Gestión de autorizaciones, órdenes y remisiones; P02 Verificación de derechos, afiliación y cobertura; P08 Continuidad del cuidado.	El resultado sugiere recurrencia de fallas asociadas con trámites, autorizaciones, verificación de derechos y direccionamiento del usuario.
Riesgos asociados al suministro de medicamentos y tecnologías.	R3 concentra 1.388.697 PQRD (17,3%). Su principal asociación se observa en P07, con 902.780 registros, y en P06, con 485.917 registros.	P07 Gestión de medicamentos, tecnologías e insumos en salud; P06 Prestación clínica, diagnóstico y tratamiento.	El patrón evidencia inconformidades relacionadas con entrega, disponibilidad o acceso a tecnologías en salud, con impacto en la continuidad de la atención.
Señales relacionadas con calidad técnico-asistencial y continuidad de la atención.	R4 registra 731.637 PQRD (9,1%). Sus asociaciones principales se ubican en P03, P08, P06 y P09.	P03 Autorizaciones; P08 Continuidad del cuidado; P06 Prestación clínica; P09 Información clínica.	Aunque su participación es menor que R1, R2 y R3, este riesgo es relevante por su relación con continuidad, atención clínica e información del usuario.
Riesgos económicos y administrativos concentrados en soporte financiero.	R6 agrupa 331.751 PQRD (4,1%), principalmente en P10, con 196.051 registros, y P02, con 135.700 registros.	P10 Gestión administrativa, facturación, cobros y pagos; P02 Verificación de derechos y cobertura.	El hallazgo evidencia señales asociadas con facturación, cobros, pagos, prestaciones económicas o validación de cobertura.
Casos no clasificados concentrados en información y calidad.	R7 registra 349.429 PQRD (4,3%). La mayor concentración se ubica en P09, con 306.743 registros.	P09 Información clínica y administrativa; P12 Calidad y mejoramiento; P11 Atención al usuario; P08 Continuidad del cuidado.	La concentración de R7 en información clínica y administrativa sugiere revisar la taxonomía, la calidad de los datos y los motivos que no quedaron representados en las categorías principales.
Señales de atención al usuario, información y trato con menor volumen.	R5 registra 162.187 PQRD (2,0%), distribuidas principalmente en P01, P09 y P11.	P01 Acceso; P09 Información clínica; P11 Atención al usuario y PQRSD.	Aunque su volumen es menor, este riesgo aporta señales sobre comunicación, orientación, trato y experiencia del usuario.

Tabla 6.5: Principales hallazgos y patrones recurrentes de riesgo (Elaboración propia)

Los hallazgos más relevantes se concentran en riesgos asociados con acceso y oportunidad, autorizaciones, suministro de medicamentos y tecnologías, calidad técnico-asistencial, continuidad del cuidado e información administrativa. Esta lectura permite priorizar procesos críticos y orientar el análisis hacia combinaciones específicas de riesgo y proceso.

La matriz de asociación entre riesgos y procesos, presentada en la Figura 6.7, complementa esta lectura al mostrar en dónde se concentran las combinaciones con mayor número de registros.

Figura 6.7: Matriz de asociación entre riesgos operacionales y procesos de atención (Elaboración propia)

Como se observa en la Figura 6.7, las asociaciones con mayor concentración fueron R1–P01, con 2.963.395 registros; R2–P03, con 1.389.946 registros; y R3–P07, con 902.780 registros. Estas combinaciones confirman que los principales focos de riesgo se ubican en el acceso a servicios, la gestión de autorizaciones y el suministro de medicamentos y tecnologías en salud.

En síntesis, esta sección transformó los resultados del modelo en hallazgos operativos. La utilidad principal del análisis está en identificar patrones que no deben interpretarse como casos aislados, sino como señales recurrentes que pueden apoyar la gestión preventiva de riesgos en el sector salud.

6.7 Formulación de recomendaciones por riesgo y proceso

Con base en los hallazgos identificados, se formularon recomendaciones orientadas a fortalecer la gestión preventiva de riesgos operacionales en los procesos de prestación de servicios de salud. Estas recomendaciones se construyeron a partir de la relación entre las categorías de riesgo R1–R7, los procesos de atención priorizados y los patrones recurrentes observados en la clasificación de PQRD.

La finalidad de esta etapa no fue generar acciones aisladas para cada queja o reclamo, sino proponer líneas de mejora aplicables a procesos donde se evidenciaron señales repetitivas de riesgo. En este sentido, la Tabla 6.6 conecta los resultados del modelo con posibles acciones de análisis, seguimiento y mejora operativa.

Riesgo operacional	Procesos priorizados	Recomendación formulada	Propósito de gestión	Información complementaria a validar
R1 Acceso y oportunidad a servicios	P01 P08 P03	Fortalecer el monitoreo de barreras de acceso, oportunidad de atención, disponibilidad de agenda y seguimiento a servicios pendientes o no prestados.	Reducir la recurrencia de inconformidades asociadas con acceso, oportunidad y continuidad del servicio.	Indicadores de oportunidad, tiempos de atención, seguimiento a servicios pendientes, remisiones y validación con responsables operativos.
R2 Autorizaciones y barreras administrativas	P03 P02 P08	Revisar el flujo de autorizaciones, órdenes, remisiones y verificación de derechos, incorporando controles sobre tiempos de respuesta, causales de negación y trazabilidad del trámite.	Disminuir demoras, reprocesos y barreras administrativas que afectan la continuidad del proceso de atención.	Tiempos de autorización, causales de devolución o negación, trazabilidad del trámite, validación de derechos y revisión con responsables del proceso.
R3 Suministro de medicamentos y tecnologías en salud	P07 P06	Implementar seguimiento operativo a la entrega oportuna y completa de medicamentos, tecnologías e insumos, articulando inventarios, contratación, prescripción y dispensación.	Prevenir interrupciones en tratamientos, retrasos en tecnologías autorizadas y fallas en la continuidad terapéutica.	Registros internos asociados a entrega de medicamentos y tecnologías, oportunidad de entrega, servicios pendientes y validación con responsables del proceso.
R4 Calidad técnico-asistencial y seguridad del paciente	P03 P08 P06 P09	Complementar los resultados del modelo con auditoría clínica, revisión de continuidad del cuidado, análisis de eventos y validación de información clínica disponible.	Identificar señales de riesgo asistencial que puedan afectar calidad, pertinencia, seguridad del paciente o continuidad de la atención.	Auditorías internas, reportes de eventos, indicadores de calidad y seguridad, información clínica disponible y validación con responsables asistenciales.
R5 Atención al usuario, información y trato	P01 P09 P11	Fortalecer los canales de atención al usuario, la claridad de la información entregada, la orientación sobre trámites y la resolutividad de las respuestas a PQRSD.	Mejorar la experiencia del usuario y disminuir inconformidades asociadas con información insuficiente, trato o falta de orientación.	Tiempos de respuesta PQRSD, registros de atención al usuario, canales de comunicación, casos reiterados y validación con responsables de atención.

R6 Temas económicos, facturación y recobros	P10 P02	Revisar controles sobre facturación, cobros, pagos, cuotas, prestaciones económicas y validación de cobertura o derechos del usuario.	Disminuir inconformidades económicas y administrativas que puedan generar barreras o afectaciones en la prestación del servicio.	Registros de facturación, cobros y pagos, validación de derechos o cobertura, prestaciones económicas y revisión con responsables administrativos.
R7 Otras situaciones no clasificadas	P09 P12 P01 P08	Analizar muestras de registros clasificados en R7 para identificar posibles subcategorías, ajustar reglas de clasificación y fortalecer la taxonomía de riesgos.	Reducir la concentración de casos residuales y mejorar la interpretabilidad de futuras clasificaciones.	Muestras manuales de registros R7, actualización de la taxonomía, revisión de calidad de datos, campos de texto y criterios de clasificación por proceso.

Tabla 6.6: Recomendaciones formuladas por riesgo y proceso (Elaboración propia)

En la Tabla 6.6, de relacional las recomendaciones que se concentran en los procesos con mayor recurrencia de señales de riesgo: acceso a servicios, autorizaciones, gestión de medicamentos y tecnologías, continuidad de la atención, información clínica y administrativa, y soporte administrativo y financiero.

Para R1, la recomendación se orienta a fortalecer el monitoreo de barreras de acceso, oportunidad de atención y disponibilidad de agenda, dado que esta categoría concentró 3.102.959 PQRD. En R2, asociada con 1.973.150 registros, se prioriza la revisión de autorizaciones, órdenes, remisiones y verificación de derechos. Para R3, con 1.388.697 PQRD, se plantea reforzar el seguimiento a la entrega de medicamentos, tecnologías e insumos en salud.

En las categorías restantes, las recomendaciones se enfocan en complementar la lectura del modelo con fuentes internas. Para R4, se propone validar señales de calidad y seguridad del paciente mediante auditorías e información clínica; para R5, fortalecer canales de atención, orientación y respuesta a PQRSD; para R6, revisar controles de facturación, cobros, pagos y cobertura; y para R7, analizar muestras manuales que permitan mejorar la taxonomía de riesgos y reducir la concentración de casos residuales.

Estas recomendaciones deben interpretarse como insumos para la toma de decisiones, no como conclusiones definitivas sobre la causa raíz de cada evento. Para avanzar hacia acciones correctivas concretas, es necesario complementar el análisis con información interna de las entidades, indicadores de proceso, tiempos de atención, auditorías, reportes de eventos y validación con los responsables operativos.

La formulación de recomendaciones cierra el ciclo analítico del proyecto al transformar la clasificación automática de PQRD en líneas de acción orientadas a la gestión preventiva. De esta manera, el modelo de PLN se convierte en una herramienta de apoyo para identificar focos de riesgo, priorizar procesos críticos y orientar mejoras en la prestación de servicios de salud.

6.8 Síntesis de hallazgos y recomendaciones

La fase de despliegue consolidó los principales resultados del proyecto, integrando la clasificación de PQRD,

la relación entre riesgos operacionales y procesos de atención, y la formulación de recomendaciones orientadas a la gestión preventiva. Este cierre permitió organizar los hallazgos de manera estructurada y facilitar su interpretación desde una perspectiva operativa.

La Tabla 6.7 resume los principales resultados del despliegue del modelo y las recomendaciones generales derivadas del análisis.

Dimensión de análisis	Resultado principal	Implicación para la gestión de riesgos	Recomendación de cierre
Inferencia masiva sobre la base depurada	El modelo Linear SVC + TF-IDF calibrado mediante sigmoid fue aplicado sobre 8.039.810 registros de PQRD del periodo 2017-2024.	El proyecto transforma una base masiva de inconformidades en una estructura analítica con categorías de riesgo operacional y probabilidades asociadas.	Mantener la trazabilidad del pipeline, versionar los artefactos del modelo y conservar las salidas de clasificación como insumo para análisis posteriores.
Distribución de riesgos operacionales	Las categorías R1, R2 y R3 concentran 6.464.806 PQRD, equivalentes al 80,4% de la base clasificada.	La mayor carga de señales operacionales se concentra en acceso, autorizaciones y suministro de medicamentos o tecnologías.	Priorizar estos riesgos en el análisis de procesos, sin excluir la revisión de categorías de menor frecuencia, pero con posible impacto en calidad o seguridad.
Relación entre riesgos y procesos de atención	La matriz riesgo-proceso muestra concentraciones relevantes en P01, P03, P07, P06, P08, P09 y P10, según la categoría de riesgo analizada.	La clasificación deja de ser únicamente estadística y se convierte en una lectura operativa orientada a procesos críticos de la prestación del servicio.	Usar la matriz de asociación como criterio inicial de priorización para análisis de causa raíz, evaluación de controles y seguimiento de planes de mejora.
Patrones recurrentes de riesgo	Los principales patrones se relacionan con barreras de acceso, oportunidad, autorizaciones, medicamentos, continuidad del cuidado, información y atención administrativa.	Estos patrones pueden interpretarse como señales tempranas de fallas recurrentes, no como conclusiones definitivas sobre cada caso individual.	Validar los patrones con responsables de proceso, indicadores internos, auditorías, eventos de riesgo y registros operativos de las entidades.
Uso de probabilidades y confianza del modelo	La calibración permite conservar, además de la clase predicha, una medida de confianza asociada a cada clasificación.	Las predicciones de mayor confianza pueden apoyar análisis agregados, mientras que los registros con menor confianza requieren lectura prudente o revisión adicional.	Definir umbrales de confianza para análisis operativos y utilizar registros inciertos como insumo para mejorar futuras muestras de entrenamiento.
Formulación de recomendaciones	Las recomendaciones se orientan a procesos donde se concentran señales de riesgo, especialmente acceso, autorizaciones, medicamentos, calidad, información y soporte administrativo.	El modelo aporta insumos para priorizar procesos, revisar controles y orientar acciones preventivas desde la gestión de riesgo operacional.	Implementar las recomendaciones como punto de partida para ejercicios de validación interna, análisis de causa raíz y fortalecimiento de controles.

Alcance y uso de los resultados	Los resultados del modelo corresponden a una aproximación analítica basada en campos textuales y categóricos disponibles en la fuente de PQRD.	La clasificación automatizada no reemplaza el juicio experto ni la evaluación específica de eventos, causas y controles.	Usar los resultados como insumo de gestión preventiva, complementándolos con información interna, validación operativa y seguimiento periódico.

Tabla 6.7: Síntesis de hallazgos y recomendaciones (Elaboración propia)

En la Tabla 6.7, el principal aporte del proyecto fue transformar 8.039.810 PQRD del periodo 2017–2024 en una base analítica con categorías de riesgo operacional, probabilidades asociadas y relación con procesos de atención. Esta estructura permitió identificar que las categorías R1, R2 y R3 concentraron 6.464.806 registros, equivalentes al 80,4% de la base clasificada.

Desde la perspectiva operativa, los resultados evidencian que las principales señales de riesgo se concentran en procesos asociados con acceso, autorizaciones, medicamentos y tecnologías, continuidad del cuidado, información clínica y administrativa, y soporte administrativo y financiero. Esta lectura permite priorizar procesos críticos y orientar análisis posteriores hacia la revisión de causas, controles y planes de mejora.

También se identificó que las probabilidades generadas por el modelo aportan una medida de confianza útil para interpretar los resultados. Las predicciones con mayor confianza pueden apoyar análisis agregados, mientras que los registros con menor confianza deben considerarse para revisión adicional, especialmente si se utilizan en decisiones operativas específicas.

En términos metodológicos, el proyecto permitió construir un flujo reproducible que inicia con la preparación de datos, continúa con el modelado supervisado y finaliza con la aplicación del modelo sobre la base depurada. No obstante, los resultados deben interpretarse como un insumo analítico y no como una conclusión definitiva sobre la causa raíz de cada evento.

En conclusión, el despliegue del modelo conectó el procesamiento de lenguaje natural con la gestión de riesgos operacionales en salud. La clasificación de PQRD, la asociación con procesos de atención y la formulación de recomendaciones constituyen una base útil para fortalecer el análisis preventivo, mejorar la trazabilidad de las inconformidades y orientar acciones de mejora en la prestación de servicios de salud.

7. CONCLUSIONES Y TRABAJOS FUTUROS

7.1 Conclusiones

El proyecto permitió desarrollar una metodología basada en procesamiento de lenguaje natural y aprendizaje automático para analizar las PQRD del sistema de salud colombiano desde una perspectiva de riesgo operacional. A partir de las bases históricas publicadas por la Superintendencia Nacional de Salud para el periodo 2017–2024, se consolidó, depuró y estructuró una base analítica que permitió transformar información textual dispersa en insumos útiles para la clasificación y análisis de riesgos.

En las fases de comprensión y preparación de los datos se evidenció que las PQRD constituyen una fuente relevante para identificar señales asociadas con fallas operacionales en la prestación de servicios de salud. El análisis exploratorio permitió reconocer concentraciones importantes por entidad, régimen, territorio y motivos de inconformidad, especialmente en temas relacionados con acceso, oportunidad, autorizaciones, tecnologías en salud y atención al usuario.

La construcción del corpus textual, a partir de los campos normalizados de macromotivo, motivo general y motivo específico, permitió convertir información no estructurada en una base apta para modelado supervisado. Además, la muestra etiquetada manualmente de 4.000 registros permitió definir una taxonomía de riesgos operacionales R1–R7, entrenar modelos de clasificación y evaluar su desempeño bajo un flujo controlado y reproducible.

En la fase de modelado se compararon diferentes algoritmos clásicos y modernos. Como resultado, se seleccionó el modelo LinearSVC + TF-IDF calibrado mediante sigmoid, debido a su equilibrio entre desempeño predictivo, estabilidad, interpretabilidad y viabilidad computacional. Esta elección permitió aplicar el modelo sobre una base depurada de 8.039.810 registros, generando para cada PQRD una categoría de riesgo operacional y una probabilidad asociada a la predicción.

El despliegue del modelo permitió pasar de una lectura individual de inconformidades a una visión agregada de riesgos y procesos. Al relacionar las categorías de riesgo con el mapa de procesos de prestación de servicios de salud, fue posible identificar patrones recurrentes, procesos críticos y posibles focos de intervención. En particular, los resultados mostraron una alta concentración de señales en riesgos asociados con acceso, autorizaciones y suministro de medicamentos y tecnologías en salud.

Desde la perspectiva de la ciencia de datos, el proyecto demuestra el valor de aplicar técnicas analíticas sobre grandes volúmenes de información textual. La metodología permitió procesar de forma sistemática y escalable millones de registros, identificar patrones que podrían pasar inadvertidos en revisiones manuales y generar información estructurada para apoyar la priorización de riesgos. Este valor agregado radica en convertir datos no estructurados en evidencia útil para la gestión preventiva y la toma de decisiones.

Asimismo, el proyecto aporta una base metodológica replicable. El flujo desarrollado puede actualizarse con nuevas bases de PQRD, ajustarse mediante nuevas muestras etiquetadas, integrarse con variables internas de las entidades y utilizarse como insumo para tableros de seguimiento. Esto fortalece la

capacidad de monitorear tendencias, anticipar focos de riesgo y orientar planes de mejora en los procesos de atención.

Finalmente, se concluye que el procesamiento de lenguaje natural puede ser utilizado como una herramienta de apoyo para fortalecer la gestión de riesgos operacionales en salud. Aunque el modelo no reemplaza el análisis experto ni la validación operativa de cada caso, sí ofrece una aproximación analítica útil para organizar grandes volúmenes de PQRD, priorizar riesgos, mejorar la trazabilidad de las inconformidades y orientar acciones preventivas en la prestación de servicios de salud.

7.2 Trabajos futuros

Como trabajo futuro, se recomienda ampliar y actualizar periódicamente el corpus etiquetado, especialmente en las categorías de riesgo con menor representación. Esto permitiría mejorar el aprendizaje del modelo en clases minoritarias, reducir posibles sesgos derivados del desbalance y fortalecer su capacidad de generalización frente a nuevos registros de PQRD.

También se sugiere incorporar fuentes de información complementarias a las variables utilizadas en este proyecto. Entre ellas se encuentran indicadores de oportunidad, tiempos de autorización, datos de entrega de medicamentos, auditorías internas, eventos de seguridad del paciente, reportes operativos y registros propios de las entidades. Esta integración permitiría contrastar las señales identificadas en las PQRD con información interna de los procesos y avanzar hacia análisis más completos de causa raíz, controles y planes de mejora.

Otra línea de trabajo consiste en profundizar la evaluación de modelos modernos de lenguaje en español, como RoBERTa, BETO u otros modelos basados en transformers. Esta exploración debería realizarse con una muestra etiquetada más amplia, mayor capacidad computacional y controles estrictos de validación, con el fin de determinar si estos modelos aportan mejoras relevantes frente a alternativas más simples, interpretables y escalables.

Adicionalmente, sería pertinente implementar mecanismos de monitoreo del desempeño del modelo en el tiempo. Esto incluye revisar cambios en la distribución de las categorías, evaluar la estabilidad de las predicciones, controlar los niveles de confianza y detectar posibles variaciones en los patrones de lenguaje de las PQRD. Este seguimiento permitiría identificar cuándo es necesario actualizar el modelo, ajustar la taxonomía o incorporar nuevos datos etiquetados.

Finalmente, se propone utilizar los resultados del modelo como insumo para ejercicios de gestión de riesgos en entidades del sector salud, integrando la clasificación de PQRD con matrices de riesgo, controles existentes, indicadores de calidad y planes de mejoramiento. De esta manera, el proyecto podría evolucionar desde un ejercicio analítico hacia una herramienta de apoyo para la gestión preventiva, la priorización de procesos críticos y el monitoreo continuo de riesgos operacionales.

8. REFERENCIAS BIBLIOGRÁFICAS

[1] Superintendencia Nacional de Salud, Informe de Gestión 2024. Bogotá D.C., Colombia, 2024. [En línea]. Disponible en: https://docs.supersalud.gov.co/PortalWeb/planeacion/InformesGestion/GG-67.pdf

[2] Superintendencia Nacional de Salud, Informe de Datos Abiertos 2023, Bogotá, Colombia, 2023. [En línea]. Disponible en: https://docs.supersalud.gov.co/PortalWeb/metodologias/Datos_Abiertos/Informe%20DA_2023.pdf

[3] IBM, "Descripción general de la ayuda de CRISP-DM," IBM Documentation – SPSS Modeler (SaaS), [En línea]. Disponible en: https://www.ibm.com/docs/es/spss-modeler/saas?topic=dm-crisp-help-overview. [Accedido: 2-jun-2025].

[4] Superintendencia Nacional de Salud, Resolución número 2024160000003002-6 de 2024. Bogotá D.C., Colombia, 2024. [En línea]. Disponible en: https://docs.supersalud.gov.co/PortalWeb/Juridica/Resoluciones/Resoluci%C3%B3n%20n%C3%BAmero%202024160000003002-6%20de%202024.pdf

[5] Superintendencia Nacional de Salud, Resolución número 2024160000003012-6 de 2024. Bogotá D.C., Colombia, 2024. [En línea]. Disponible en: https://docs.supersalud.gov.co/PortalWeb/Juridica/Resoluciones/Resoluci%C3%B3n%20n%C3%BAmero%202024160000003012-6%20de%202024.pdf

[6] Basel Committee on Banking Supervision, International Convergence of Capital Measurement and Capital Standards: A Revised Framework – Comprehensive Version, Bank for International Settlements, Basel, Switzerland, Jun. 2006. [Online]. Available: https://www.bis.org/basel_framework/

[7] Superintendencia Nacional de Salud, Circular Externa No. 045 de 2021, Bogotá, Colombia, 2021. [En línea]. Disponible en: https://docs.supersalud.gov.co/PortalWeb/Juridica/CircularesExterna/CIRCULAR%20EXTERNA%20202021170000000045.pdf

[8] Superintendencia Nacional de Salud, Circular Externa No. 051 de 2022, Bogotá, Colombia, 2021. [En línea]. Disponible en: https://docs.supersalud.gov.co/PortalWeb/Juridica/CircularesExterna/Circular%20Externa%20No.%20222151000000051-5%20de%202022.pdf

[9] F. Venegas-Martínez, L. C. Franco-Arbeláez, L. E. Franco-Ceballos y J. G. Murillo-Gómez, "Riesgo operativo en el sector salud en Colombia: 2013," eseconomía, vol. 10, no. 43, pp. 7–36, segundo semestre de 2015. [En línea]. Disponible en: https://yuss.me/revistas/ese/2015v10n43a01p007_036.pdf

[10] Superintendencia Nacional de Salud, Circular Externa No. 2023151000000010-5 de 2023, Bogotá, Colombia, 2023. [En línea]. Disponible en: https://docs.supersalud.gov.co/PortalWeb/Juridica/CircularesExterna/Circular%20Externa%20No.%2023151000000010-5%20de%202023.pdf

[11] IBM, "¿Qué es el procesamiento del lenguaje natural (PLN)?", IBM Think, 2024. [En línea]. Disponible en: https://www.ibm.com/mx-es/think/topics/natural-language-processing

[12] Amazon Web Services, "¿Qué es el procesamiento de lenguaje natural (PLN)?", AWS, [En línea]. Disponible en: https://aws.amazon.com/es/what-is/nlp/. [Accedido: 2-jun-2025].

[13] Microsoft, "Procesamiento de lenguaje natural: elecciones tecnológicas para arquitecturas de datos," Microsoft Learn, [En línea]. Disponible en: https://learn.microsoft.com/es-es/azure/architecture/data-guide/technology-choices/natural-language-processing. [Accedido:2-jun-2025].

[14] OpenWebinars, "Técnicas clave para el procesamiento de texto NLP," OpenWebinars, [En línea]. Disponible en: https://openwebinars.net/blog/tecnicas-clave-para-procesamiento-texto-nlp/#word-embeddings. [Accedido: 2-jun-2025].

[15] Instituto de Ingeniería del Conocimiento, "Topic Extraction: ¿sobre qué tratan los documentos?", IIC - Procesamiento del Lenguaje Natural, [En línea]. Disponible en: https://www.iic.uam.es/procesamiento-del-lenguaje-natural/topic-extraction-sobre-que-tratan-documentos/. [Accedido: 2-jun-2025].

[16] Scikit-learn developers, "Metrics and scoring: quantifying the quality of predictions," scikit-learn Documentation. [Online]. Disponible en: https://scikit-learn.org/stable/modules/model_evaluation.html. [Accedido: 10-Jun-2026].

[17] Scikit-learn developers, "sklearn.metrics.matthews_corrcoef," scikit-learn Documentation. [Online]. Disponible en: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.matthews_corrcoef.html. [Accedido: 10-Jun-2026].

[18] Scikit-learn developers, "sklearn.metrics.log_loss," scikit-learn Documentation. [Online]. Disponible en: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.log_loss.html. [Accedido: 10-Jun-

2026].

[19] Y.P. Ayala, J.A. Salamanca, J.M. Durán y E.F. González, “Análisis con Machine Learning de Peticiones Externas (PQRS) del SENA Regional Bolívar”, Fundación Universitaria Los Libertadores, Bogotá, 2023. [En línea]. Disponible en: https://repository.libertadores.edu.co/server/api/core/bitstreams/17882639-7ae2-4c81-b746-

[20] R. A. Manjarrés-Betancur y M. M. Echeverri-Torres, “Asistente Virtual Académico utilizando Tecnologías Cognitivas de Procesamiento de Lenguaje Natural,” Revista Politécnica, vol. 16, no. 30, pp. 59–67, 2020. [En línea]. Disponible en: https://revistas.elpoli.edu.co/index.php/pol/article/view/1701

[21] J. L. Flores Poma, “Clasificación de Quejas de los Clientes Empleando Procesamiento de Lenguaje Natural: Revisión Sistemática de la Literatura,” ResearchGate, 2023. [En línea]. Disponible en: https://www.researchgate.net/publication/377578962

[22] J. C. Blandón-Andrade, A. Castaño-Toro, A. Morales-Ríos, y D. Orozco-Ospina, “Computational Methods for Information Processing from Natural Language Complaint Processes—A Systematic Review,”Computers, vol. 14, no. 1, Art. no. 28, Jan. 2024. [En línea]. Disponible en: https://www.mdpi.com/2073-431X/14/1/28

[23] M. Navia Diazgranados, “Procesamiento de Lenguaje Natural Impulsado por Inteligencia Artificial para Empresas de Servicios Públicos”, Universidad Tecnológica de Pereira, Pereira, 2023. [En línea]. Disponible en: https://repositorio.utp.edu.co/items/ad85712a-1f8e-4ccc-bf6b-74cf44e7f5fb

[24] W. A. Gómez Bueno y E. A. Gómez Cárdenas, “Prototipo de herramienta para la mejora en los procesos de designación de PQRSD de la Alcaldía de Bucaramanga”, Pontificia Universidad Javeriana Cali, Maestría en Ciencia de Datos, Santiago de Cali, 2023. [En línea]. Disponible en: https://vitela.javerianacali.edu.co/items/7728c175-a9eb-4ff6-ac00-c3f30b3826ad

[25] B. S. Flórez Pazos, C. A. Arévalo Rodríguez y S. Barrera Sáenz, “Comparación de metodologías PEFT aplicadas a modelos de lenguaje grandes enfocados en generar resúmenes de textos”, Pontificia Universidad Javeriana Cali, Maestría en Ciencia de Datos, Santiago de Cali, 2024. [En línea]. Disponible en: https://vitela.javerianacali.edu.co/items/f5c8e245-3f58-4209-b227-95a08476043a

ANEXO 1

CRONOGRAMA

Avance	Gap
100%	0%

SEGUIMIENTO

No.	Actividad	Estatus	Año 2025						Año 2026					
			JUL	AGO	SEP	OCT	NOV	DIC	ENE	FEB	MAR	ABR	MAY	JUN
6.1	COMPRENSIÓN DE LOS DATOS													
6.1.1.	Obtener los datos históricos de PQRD desde el Portal de Datos Abiertos en la página web de la Superintendencia de Salud	Finalizado												
6.1.2.	Estructura del dataset e identificar datos faltantes, duplicados o inconsistentes.	Finalizado												
6.1.3.	Explorar de la base de datos mediante análisis exploratorio y visualización descriptiva	Finalizado												
6.1.4.	Depurar base de datos.	Finalizado												
6.1.5.	Normalizar las variables categóricas y estandarizar formatos de texto para su posterior análisis.	Finalizado												
6.2	PREPARACIÓN DE LOS DATOS													
6.2.1.	Preprocesar el texto aplicando técnicas de NLP como tokenización, lematización, eliminación de stopwords y normalización léxica	Finalizado												
6.2.2.	Etiquetar manualmente una muestra representativa de datos para entrenamiento supervisado.	Finalizado												
6.2.3.	Identificar y evaluar los algoritmos	Finalizado												
6.2.4.	Seleccionar los algoritmos de clasificación más apropiados para el proyecto	Finalizado												
6.3	MODELADO													
6.3.1.	Entrenar los modelos de clasificación	Finalizado												
6.3.2.	Dividir el conjunto de datos en conjuntos de entrenamiento y prueba.	Finalizado												
6.3.3.	Identificar y balancear las clases.	Finalizado												
6.3.4.	Evaluar y ajustar los parámetros de los modelos.	Finalizado												
6.3.5.	Ejecutar el modelo.	Finalizado												
6.4.	EVALUACIÓN													
6.4.1.	Comparar los resultados obtenidos para determinar la capacidad predictiva del modelo.	Finalizado												
6.4.2.	Calcular métricas de evaluación.	Finalizado												
6.4.3.	Evaluar métricas de desempeño.	Finalizado												
6.4.4.	Evaluar la estabilidad del modelo.	Finalizado												
6.4.5.	Analizar los errores de clasificación.	Finalizado												
6.4.6.	Identificar categorías de riesgo mal representadas.	Finalizado												
6.4.7.	Ajustar el módelo calibrando los hiperparámetros.	Finalizado												
6.4.8.	Documentar los resultados de desempeño del modelo.	Finalizado												
6.5.	DESPLIEGUE													
6.5.1.	Diseñar el mapa de procesos de la prestación en los servicios de salud.	Finalizado												
6.5.2.	Relacionar los textos clasificados con los procesos de la prestación de servicios de salud.	Finalizado												
6.5.3.	Diseñar las visualizaciones para presentar las correlaciones entre riesgos y los procesos de la prestación de servicios de salud.	Finalizado												
6.5.4.	Analizar los hallazgos relacionados con patrones recurrentes de riesgo.	Finalizado												
6.5.5.	Formular recomendaciones alineadas con estándares de calidad y normativas del sector salud.	Finalizado												
6.5.6.	Elaborar el documento final que consolide los hallazgos, el modelo y las recomendaciones operativas.	Finalizado												
6.5.7.	Diseñar las visualizaciones para presentar los resultados.	Finalizado												
6.6.	ELABORACIÓN DOCUMENTO FINAL													
6.6.1.	Consolidar documento	Finalizado												
6.6.2.	Estructurar documento	Finalizado												
6.6.3.	Revisar contenidos	Finalizado												
6.6.4.	Enviar para comentarios al Director	Finalizado												
6.6.5.	Ajustar comentarios	Finalizado												
6.6.6.	Elaborar versión final documento	Finalizado												

ANEXO 2

NOTEBOOK DEL PROYECTO: PROYECTO_PNL_PQRD.IPYNB

https://colab.research.google.com/drive/1f6HWUTnlM1WMjQ9LpON1Y0_NGuGWjyY4?usp=sharing

ANEXO 3

METODOLOGÍA DE ETIQUETADO MANUAL PQRD

1. Propósito y alcance

Esta metodología define el procedimiento de etiquetado manual de Peticiones, Quejas, Reclamos y Denuncias (PQRD) con el fin de construir un conjunto de entrenamiento supervisado para modelos de clasificación de riesgos operacionales. En esta fase se especifica qué se etiqueta, cómo se etiqueta y con qué reglas, garantizando consistencia, trazabilidad y reproducibilidad para su implementación.

2. Objetivo

Construir un conjunto de entrenamiento supervisado en el que cada registro de PQRD disponga de un texto apto para su análisis mediante técnicas de procesamiento de lenguaje natural (PLN) y de una etiqueta de riesgo operacional asignada manualmente bajo un esquema y reglas de decisión previamente definidos.

Este conjunto etiquetado servirá como insumo base para el entrenamiento, validación y evaluación de modelos de clasificación orientados a identificar y categorizar riesgos operacionales a partir del contenido textual de las PQRD.

En el conjunto de entrenamiento supervisado cada registro de PQRD debe contar con:

- Texto disponible para PLN y lectura.
- Una etiqueta de riesgo operacional primaria (R1–R7) asignada manualmente con base en la guía.
- Campos auxiliares para observaciones y, opcionalmente, una etiqueta secundaria.
- Un archivo final etiquetado que sirva como insumo para entrenar y evaluar modelos de clasificación de riesgos.

3. Unidad de análisis y objeto a etiquetar

- Unidad de análisis: un registro individual de PQRD.
- Objeto a etiquetar: el núcleo de la inconformidad expresada en el caso, identificado principalmente a partir del texto y complementado por el contexto temático.

4. Esquema de etiquetas de riesgo operacional (R1–R7)

Se adopta un esquema inicial de siete clases (R1–R7). En esta fase se utiliza un enfoque en el que cada PQRD recibe una etiqueta primaria única para simplificar el entrenamiento y el análisis. El campo de etiqueta secundaria se conserva como opcional para futuros refinamientos.

CÓDIGO	NOMBRE DE LA CLASE	DESCRIPCIÓN OPERATIVA	TEMAS TÍPICOS (ORIENTATIVOS)
R1	Acceso y oportunidad a servicios	Barreras o demoras para ingresar al sistema o acceder a un servicio específico.	cita, agenda, turno, demora, fila, restricción, cupo
R2	Autorizaciones y barreras administrativas	Trámites, autorizaciones, referencias, remisiones o trabas burocráticas.	autorización, negación, remisión, trámite, radicación
R3	Suministro de medicamentos y tecnologías en salud	Problemas de entrega, disponibilidad, sustitución u oportunidad de medicamentos, insumos o ayudas diagnósticas.	medicamento, fórmula, entrega, stock, desabastecimiento
R4	Calidad técnico-asistencial y seguridad del paciente	Errores o fallas en la atención clínica, procedimientos, diagnósticos, cuidados o complicaciones.	procedimiento, cirugía, diagnóstico, error, complicación
R5	Atención al usuario, información y trato	Mal trato, falta de información, comunicación deficiente u orientación inadecuada.	trato, maltrato, información, orientación, respeto
R6	Facturación, cobros y temas económicos	Cobros indebidos, facturación, copagos, cuotas, recobros, glosas u otros temas económicos/administrativos.	cobro, factura, cuota, copago, recobro, cartera
R7	Otros / No clasificable con la información disponible	Casos que no encajan claramente en las clases anteriores o con información insuficiente/ambigua.	texto corto, genérico, ambiguo o incongruente

Tabla 1. Clases de riesgo operacional para etiquetado manual (R1–R7)

5. Fuentes de información para asignar la etiqueta

Para asegurar consistencia, se debe revisar usando, como mínimo, los siguientes campos:

- texto_basic (texto principal para lectura).
- motivo_especifico_clean (motivo específico limpio).
- motivo_general_clean y macromotivo_clean (contexto temático).
- Variables de trazabilidad: objectid, year, ent_alias_sns.

6. Reglas de etiquetado

6.1 Regla general (etiqueta primaria)

Identificar el núcleo de la inconformidad (p. ej., acceso, autorizaciones, medicamentos, tecnologías, calidad clínica, trato o cobros) y asignar una única clase R1–R7 en el campo etiqueta_riesgo_primaria.

6.2 Priorización cuando hay múltiples temas

Si el texto incluye más de un problema, asignar la etiqueta primaria siguiendo este orden de priorización:

1. R4 si hay riesgo clínico evidente o aspectos de seguridad del paciente.
2. R3 si el problema central es suministro/entrega de medicamentos o tecnologías.
3. R1–R2 si lo crítico es la imposibilidad o demora para acceder al servicio o completar trámites.
4. R6 si el núcleo de la inconformidad es económico (cobros, facturación, copagos, etc.).
5. R5 si el núcleo es la experiencia del usuario (trato, información u orientación).
6. R7 si no es posible clasificar con seguridad.

6.3 Manejo de casos ambiguos

Cuando existan múltiples temas, se seleccionará el mejor descrito o el de mayor criticidad operacional. Si no hay información suficiente para decidir entre dos clases, se asignará R7 y se documentará el motivo en comentarios.

6.4 Uso del campo de comentarios

El campo comentarios se utiliza para registrar dudas, decisiones difíciles o casos borde. Este registro permite refinar la guía, identificar patrones de ambigüedad y soportar posibles reetiquetados.

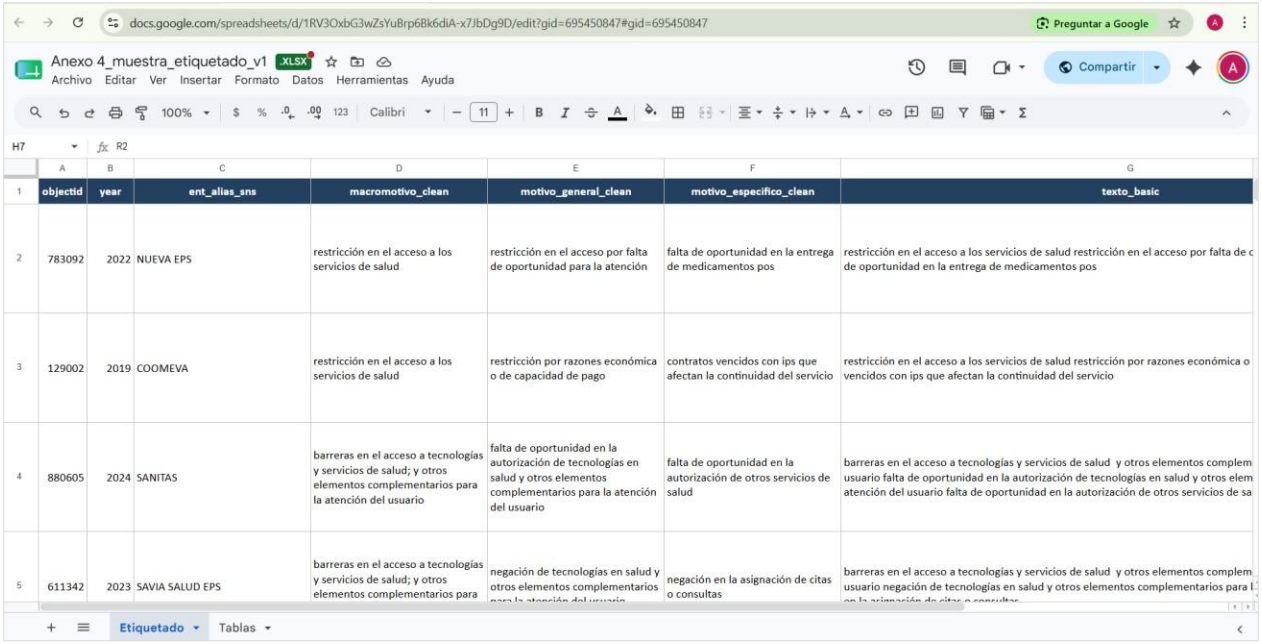
6.5 Etiqueta secundaria (opcional)

El campo etiqueta_riesgo_secundaria se mantiene como opción para registrar un segundo tema relevante cuando aplique; en la primera iteración puede dejarse en blanco.

ANEXO 4

MUESTRA ETIQUETADO

https://docs.google.com/spreadsheets/d/1RV3OxbG3wZsYuBrp6Bk6diA-x7JbDg9D/edit?gid=695450847#gid=695450847



	objectid	year	ent_alias_sns	macromotivo_clean	motivo_general_clean	motivo_especifico_clean	texto_basic
2	783092	2022	NUEVA EPS	restricción en el acceso a los servicios de salud	restricción en el acceso por falta de oportunidad para la atención	falta de oportunidad en la entrega de medicamentos pos	restricción en el acceso a los servicios de salud restricción en el acceso por falta de o de oportunidad en la entrega de medicamentos pos
3	129002	2019	COOMEVA	restricción en el acceso a los servicios de salud	restricción por razones económica o de capacidad de pago	contratos vencidos con ips que afectan la continuidad del servicio	restricción en el acceso a los servicios de salud restricción por razones económica o vencidos con ips que afectan la continuidad del servicio
4	880605	2024	SANITAS	barreras en el acceso a tecnologías y servicios de salud; y otros elementos complementarios para la atención del usuario	falta de oportunidad en la autorización de tecnologías en salud y otros elementos complementarios para la atención del usuario	falta de oportunidad en la autorización de otros servicios de salud	barreras en el acceso a tecnologías y servicios de salud y otros elementos complem usuario falta de oportunidad en la autorización de tecnologías en salud y otros elem atención del usuario falta de oportunidad en la autorización de otros servicios de sa
5	611342	2023	SAVIA SALUD EPS	barreras en el acceso a tecnologías y servicios de salud; y otros elementos complementarios para	negación de tecnologías en salud y otros elementos complementarios	negación en la asignación de citas o consultas	barreras en el acceso a tecnologías y servicios de salud y otros elementos complem usuario negación de tecnologías en salud y otros elementos complementarios para