Recomendador editorial para el Blog de *Hernán Casciari*

Adrián Arturo Suárez García
202123771
a.suarezg@uniandes.edu.co

Luis Alejandro Rubiano Guerrero
202013482
la.rubiano@uniandes.edu.co

Gabriel Alejandro Moreno Riveros
202014583
g.morenor@uniandes.edu.co

Gianluca Cicco
202020881
g.cicco@uniandes.edu.co

Juan Sebastián Sierra
202020881
j.sierrat@uniandes.edu.co

I. Introducción

Hernán Casciari, reconocido escritor argentino y pionero de la literatura digital, ha construido a lo largo de más de una década un vasto universo narrativo compuesto por cientos de cuentos que documentan la experiencia del inmigrante argentino en España, las reflexiones sobre la paternidad, el fútbol como pasión cultural, y las complejidades de la vida moderna. Su obra, caracterizada por un registro conversacional íntimo y referencias culturales distintivas, presenta un desafío particular para los lectores: cómo navegar eficientemente a través de esta extensa colección para descubrir nuevas lecturas que resonarán con sus intereses específicos.

Para abordar esta necesidad, desarrollamos un sistema de recomendación dual que explora dos perspectivas complementarias del análisis textual. La primera metodología emplea TF-IDF combinado con similitud coseno para identificar cuentos que comparten vocabulario específico, capturando similitudes basadas en el uso literal de palabras y expresiones. La segunda aproximación utiliza modelado de tópicos LDA para descubrir estructuras temáticas latentes que conectan textos que, aunque expresados mediante vocabularios diferentes, abordan preocupaciones conceptuales similares. Al contrastar ambas metodologías sobre el corpus completo de Casciari, pretendemos ilustrar cómo diferentes técnicas de procesamiento de lenguaje natural capturan dimensiones distintas de la similitud textual, proporcionando información valiosa sobre los mecanismos que conectan experiencias narrativas aparentemente dispares.

II. Limpieza de datos

Antes de armar recomendaciones miramos el corpus para entender su forma. Observamos la distribución de longitudes y vimos que la mayoría de los cuentos se concentra en unos pocos miles de caracteres y aparece una cola hacia la derecha con muy pocos relatos largos (Ver figura 1). Eso confirma que el conjunto es heterogéneo en tamaño y nos obliga a controlar el efecto de la longitud para que un texto no parezca más similar a otro solo por tener más palabras.

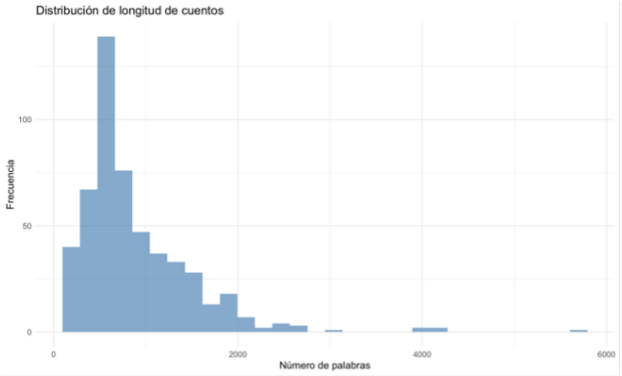

Fig. 1. Distribución de la longitud de los cuentos en palabras.

El corpus del blog de Hernán Casciari contiene 520 cuentos publicados entre septiembre de 2003 y noviembre de 2015. Como se observa en las Figuras 1 y 2, el análisis exploratorio revela la evolución narrativa y productiva. Distribución de longitud (Figura 1.1): Los cuentos presentan una distribución con un promedio de 882 palabras y mediana de 702 palabras, concentrándose la mayoría entre 500-800 palabras. Sin embargo, existe gran variabilidad que va desde cuentos muy breves de 95 palabras hasta extensos de 5,603 palabras. La distribución muestra una pronunciada cola derecha con 26 outliers largos (>percentil 95), incluyendo cuentos de más de 4,000 palabras que representan piezas más ambiciosas narrativamente pero que podrían sesgar los algoritmos de similitud léxica.

Evolución temporal (Figura 2): Se identifica un boom inicial 2004-2005 con aproximadamente 200 cuentos, período que coincide con el reconocimiento internacional al recibir el premio Deutsche Welle al "Mejor blog del mundo" en 2005. Posteriormente se observa un declive gradual 2006-2008 con 100-150 cuentos anuales, seguido de actividad intermitente 2009-2012 con 50-100 cuentos que sugiere diversificación hacia otros proyectos editoriales. Una pausa casi total desde el 2013-2014 con menos de 20 cuentos marca una pausa significativa, finalizada por una reactivación moderada hacia 2015-2016 con 50-75 cuentos anuales. Esta evolución temporal

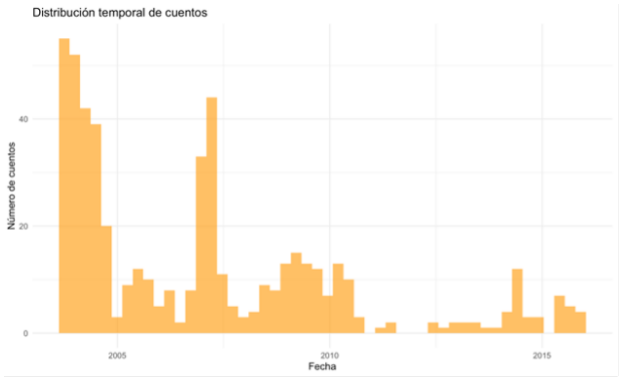

Fig. 2. Distribución temporal de los cuentos publicados.

es crucial para el sistema de recomendación, ya que refleja no solo cambios en productividad sino potencialmente en estilo, temas y madurez narrativa del autor.

Implementamos un pipeline de estandarización sistemático considerando que el objetivo es crear un sistema de recomendación para retener lectores del blog, donde tanto la longitud como la temporalidad emergen como factores críticos. La longitud refleja patrones de consumo y tiempo disponible del lector, mientras que los outliers identificados (cuentos muy cortos o muy largos) pueden sesgar los resultados de similitud en ambos enfoques propuestos. La dimensión temporal, de cuando fueron redactados los cuentos también es relevante, aunque los métodos léxico (TF-IDF) y temático (LDA) empleados son agnósticos temporalmente, por lo que no es un tema que tendremos en cuenta excepto para las limitaciones de nuestro sistema de recomendación.

El procesamiento inició con una lipieza básica: eliminación de tildes y caracteres especiales mediante transformación para evitar inconsistencias de encoding, remoción de saltos de línea y normalización de espacios, seguido de la eliminación de las últimas 5 palabras identificadas como firmas del autor. Posteriormente se aplicó limpieza lexical mediante conversión a minúsculas, eliminación de puntuación y caracteres numéricos, y normalización de espacios múltiples.

A continuación se ejecutó la lematización utilizando el modelo udpipe español. La lematización convierte las palabras a su forma base considerando el contexto gramatical (por ejemplo, ”niñas” y ”niños” se reducen a ”niño”), proceso que se realizó DESPUÉS de la limpieza básica pero ANTES de la eliminación de stopwords para capturar adecuadamente las variaciones morfológicas en el texto limpio.

El filtrado de contenido combinó stopwords estándar en español de dos fuentes complementarias (Snowball + NLTK) con stopwords específicas del dominio. Las stopwords específicas se construyeron mediante web scraping de expresiones argentinas, que fueron posteriormente lematizadas usando el mismo modelo udpipe para capturar variaciones morfológicas. Este enfoque preserva la esencia temática del corpus mientras elimina tanto ruido gramatical estándar como expresiones regionales específicas.

Una etapa distintiva del pipeline es la generación de n-gramas combinados: se crearon unigramas (palabras individuales), bigramas (secuencias de 2 palabras) y trigramas (secuencias de 3 palabras) a partir del corpus lematizado y filtrado. Estos tres tipos de tokens se combinaron en un solo texto por documento, enriqueciendo la representación semántica para capturar tanto términos individuales como expresiones compuestas frecuentes, mejorando la calidad de las recomendaciones basadas en similitud léxica.

La matriz documento-término final presenta dimensiones de 520 documentos $\times$ 504 términos tras aplicar un umbral de sparsity del 90%, eliminando términos presentes en menos del 10% de documentos. El vocabulario se redujo de 22,232 a 504 términos únicos, representando una disminución del 97.7% que optimiza la representación semántica sin sacrificar información temáticamente relevante. Los términos más frecuentes post-procesamiento reflejan el núcleo temático del autor: decir, hacer, él, mas, vez, evidenciando el estilo narrativo característico de Hernán Casciari centrado en la oralidad y la experiencia cotidiana.

III. TF-IDF ENFOQUE LÉXICO

TF-IDF (Term Frequency-Inverse Document Frequency) es una medida estadística que evalúa la importancia relativa de una palabra en un documento dentro de una colección textual. Combina dos componentes complementarios, por una parte la frecuencia del término, que es el número de veces que el término t aparece en el documento, y la frecuencia inversa del documento, que es el logaritmo del total de documentos que contienen el término t, y penaliza los términos ubicuos, es decir, que se encuentra en muchos lugares al mismo tiempo y prioriza las palabras distintivas. Los beneficios de este método es que es simple y fácil de interpretar, además que es robusto frente a la variación en longitud de diferentes documentos.

Este método comienza con la matriz documento previamente procesada con 520 documentos por 504 términos. Sobre esta representación numérica existente se aplicaron los pesos TF-IDF mediante la función `weightTfIdf()`, que calcula automáticamente las frecuencias de términos y frecuencias inversas de documentos para cada elemento de la matriz. Dado que el coseno es invariante a reescalamientos de los vectores, la medida de cercanía depende de la composición relativa de términos (orientación de los vectores) y no del volumen absoluto de texto. De manera opcional, se aplicó una normalización L^2 a los vectores TF–IDF únicamente para mejorar el acondicionamiento numérico y ubicar los documentos sobre la esfera unitaria. Esto no cambia el orden inducido por la similitud del coseno, pero facilita interpretaciones geométricas y ciertos cálculos que puedan surgir para proyectos futuros. El resultado es una matriz 520 $\times$ 504 donde cada fila representa un cuento como vector TF-IDF normalizado.

Posteriormente, se construye la similitud coseno, la cual se calculó entre todos los pares de vectores TF-IDF normalizados, generando una matriz simétrica 520 $\times$ 520. Cada elemento $[i, j]$ representa la similitud léxica entre los cuentos i y j, con

valores entre 0 (vocabularios completamente diferentes) y 1 (vocabularios idénticos).

Para demostrar el funcionamiento del sistema de recomendación, se seleccionó el cuento "Messi es un perro" como caso de estudio. Esta elección responde a criterios que lo hacen representativo del corpus de Casciari: presenta la temática autobiográfica característica del autor, combinando elementos futbolísticos, familiares y reflexiones sobre crisis económica que son recurrentes en su obra. El cuento emplea el registro coloquial y la oralidad típica del estilo de Casciari, con expresiones como "La respuesta rápida es por mi hija, por mi esposa" y referencias culturales específicas como "estoy a cuarenta minutos en tren del mejor fútbol de la historia". El texto desarrolla una reflexión personal sobre permanecer en Barcelona a pesar de las circunstancias adversas, proporcionando vocabulario distintivo que incluye términos relacionados con fútbol, familia, crisis y experiencias migratorias.

La aplicación del algoritmo TF-IDF sobre "Messi es un perro" generó una lista de diez recomendaciones ordenadas por similitud coseno decreciente. Los cinco cuentos más similares identificados por el sistema son: "Canelones" (0.631), "Instrucciones para crear mundos paralelos" (0.496), "El espectáculo de volar" (0.457), "¿Cuni... qué?" (0.435) y "Ni olvido ni perdón" (0.407). Esta selección revela similitudes significativas en el funcionamiento del algoritmo: los valores de similitud, que oscilan entre 0.407 y 0.631, demuestran superposición léxica sustancial, indicando que el sistema identifica efectivamente patrones vocabulario compartidos en el registro autobiográfico y la estructura narrativa reflexiva característica del corpus de Casciari.

Adicionalmente, el análisis comparativo con "Canelones" permite observar cómo el algoritmo detecta similitudes temáticas específicas con una similitud coseno de 0.631, la más alta de las recomendaciones. Ambos textos comparten elementos autobiográficos fundamentales: "Canelones" narra experiencias de infancia y adolescencia en Mercedes con referencias específicas como "Con el Chiri nos convertimos en expertos cuando promediábamos el secundario", mientras que "Messi es un perro" incluye reflexiones personales sobre la vida en Barcelona con menciones familiares argentinas. Ambos cuentos emplean un registro conversacional directo, incorporan argentinismos y referencias culturales específicas, y desarrollan narrativas introspectivas que conectan experiencias pasadas con reflexiones presentes. Esta alta similitud (0.631) demuestra que TF-IDF captura efectivamente patrones léxicos asociados al estilo narrativo distintivo de Casciari, caracterizado por la oralidad, las referencias culturales argentinas y la estructura autobiográfica reflexiva.

IV. LDA Enfoque temático

Latent Dirichlet Allocation (LDA) es un modelo probabilístico generativo que descubre tópicos latentes en colecciones de documentos mediante la asignación de distribuciones de probabilidad sobre vocabularios temáticos. A diferencia de TF-IDF que se basa en frecuencias léxicas directas, LDA modela cada documento como una mezcla de tópicos subyacentes, donde cada tópico se caracteriza por una distribución de probabilidades sobre palabras. Este enfoque permite identificar similitudes conceptuales profundas más allá del vocabulario superficial, capturando la estructura temática latente del corpus mediante dos matrices fundamentales: theta (distribuciones palabra-tópico) y omega (distribuciones documento-tópico).

La implementación de LDA sobre el corpus de Casciari comenzó con la evaluación sistemática de diferentes valores de K tópicos (3, 5, 7, 10) para determinar el número óptimo mediante criterios de interpretabilidad temática. Se seleccionó K=7 tópicos tras analizar la coherencia y diferenciación de los grupos temáticos resultantes, equilibrando la granularidad suficiente para capturar la diversidad narrativa del autor sin fragmentar excesivamente los patrones temáticos. El modelo con 7 tópicos demostró generar agrupaciones interpretables que reflejan las principales líneas temáticas de Casciari: autobiografía familiar, crítica social, experiencias migratorias, reflexiones sobre fútbol, narrativas de infancia, comentarios culturales y relatos cotidianos.

Los tópicos generados revelan distribuciones coherentes que explican las similitudes detectadas por LDA. Los documentos recomendados comparten patrones distribucionales claros: todos presentan al Tópico 6, caracterizado por términos como "año", "escribir", "libro", "día", "fot" y "primero", representa la dimensión autobiográfica y narrativa temporal del corpus de Casciari. Como componente dominante, con probabilidades entre 43.5% y 55.9%. "Messi es un perro" (43.5% Tópico 6, 31.6% Tópico 2), "Donar los órganos" (54.5% Tópico 6, 44.3% Tópico 2), "Fútbol, fervor e independencia" (55.9% Tópico 6, 33.6% Tópico 2) y "¿Cuni... qué?" (54.1% Tópico 6, 37.7% Tópico 2) muestran composiciones temáticas similares dominadas por estos dos tópicos principales. "Los dos comodines" presenta una distribución ligeramente diferente (43.9% Tópico 6, 21.1% Tópico 5, 18.3% Tópico 2) pero mantiene al Tópico 6 como elemento central. Esta convergencia distribucional explica las altas similitudes coseno (0.94-0.96) obtenidas por LDA: los vectores de probabilidad gamma son estructuralmente similares independientemente del vocabulario específico de cada texto, demostrando que LDA identifica efectivamente patrones de composición temática latente que trascienden las diferencias léxicas superficiales del corpus de Casciari.

Los tópicos generados revelan grupos temáticamente coherentes que justifican K=7: Tópico 1 ("decir", "mirar", "ver") agrupa interacciones sociales como "Belleza prohibida en televisión". Tópico 2 ("mundo", "poder", "nuevo") concentra reflexiones existenciales en "Mis abuelos". Tópico 3 ("día", "cosa", "siempre") reúne rutinas cotidianas como "La desidia". Tópico 4 ("caio", "zacario", "sofi") constituye narrativas familiares específicas en "La venganza del metegol". Tópico 5 ("casa", "noche") agrupa conversaciones domésticas como "El coleccionismo". Tópico 6 ("año", "escribir", "libro") representa el proceso autobiográfico-creativo en "Malos tiempos para el humor online". Tópico 7 ("argentino", "fútbol", "país") concentra identidad nacional en "Los dos rulfos". Esta segmentación demuestra que LDA identifica efectivamente las

líneas narrativas distintivas del corpus: interacciones sociales, reflexiones existenciales, cotidianidad, universo familiar, narrativa doméstica, proceso creativo e identidad argentina.

Aplicando el sistema de recomendación LDA sobre "Messi es un perro", se generó una lista de cinco cuentos más similares basada en la similitud coseno entre distribuciones de tópicos gamma: "Los dos comodines" (0.959), "Donar los órganos" (0.950), "Fútbol, fervor e independencia" (0.948), "Los payasos" (0.947) y "¿Cuni... qué?" (0.946). Estos valores de similitud, significativamente más altos que los obtenidos con TF-IDF (0.407-0.631), reflejan diferencias metodológicas fundamentales entre ambos enfoques.

El análisis de "Donar los órganos" revela por qué LDA detecta alta similitud temática (0.950) con "Messi es un perro" a pesar de campos léxicos completamente divergentes. Mientras "Messi es un perro" emplea vocabulario futbolístico ("Champions", "Liga", "Barcelona"), "Donar los órganos" utiliza terminología médica ("órganos", "doctorcito", "cirugías", "Ministerio de Salud"). Sin embargo, ambos textos comparten estructuras narrativas profundas: diálogos internos reflexivos, confrontación con sistemas burocráticos (fútbol profesional vs. sistema de salud), y protagonistas que cuestionan decisiones institucionales desde perspectivas personales. LDA captura estas afinidades conceptuales latentes que TF-IDF no puede detectar debido a la divergencia vocabulario. Esta divergencia entre métodos evidencia que TF-IDF privilegia la similitud léxica superficial mientras LDA detecta afinidades temáticas estructurales.

En la Tabla I se resumen las recomendaciones generadas por ambos métodos para "Messi es un perro", destacando las diferencias en títulos y valores de similitud.

V. Elección de k

Definir k es crucial porque determina la resolución temática del modelo: si k es muy pequeño, el LDA mezcla varios asuntos diferentes en un mismo tópico (temas demasiado generales); si k es muy grande, divide un mismo asunto en micro-tópicos redundantes o ruidosos. Por eso necesitamos un criterio cuantitativo que equilibre detalle e interpretabilidad.

Usamos coherencia de tópicos como métrica de calidad. En términos simples la coherencia indica qué tan bien van juntas las palabras principales de cada tópico. Si las palabras top coaparecen con frecuencia en los mismos documentos o ventanas y guardan relación semántica la coherencia es alta. Si casi no coaparecen o no están relacionadas la coherencia es baja. Esta métrica ayuda a escoger k porque evalúa directamente lo que buscamos en LDA. Queremos tópicos interpretables y útiles para explicar el contenido.

Probamos k en 3, 5, 7 y 10. La coherencia sube desde k igual a 3 hasta k igual a 7 y luego baja un poco en k igual a 10. Los valores fueron $k = 3$ con 0.383, $k = 5$ con 0.489, $k = 7$ con 0.529 y $k = 10$ con 0.519. Elegimos k igual a 7. Con ese valor los temas quedan más claros y fáciles de interpretar. No partimos un mismo tema en partes innecesarias y tampoco los mezclamos en bloques muy generales. Con k igual a 7 logramos un buen nivel de detalle y una lectura

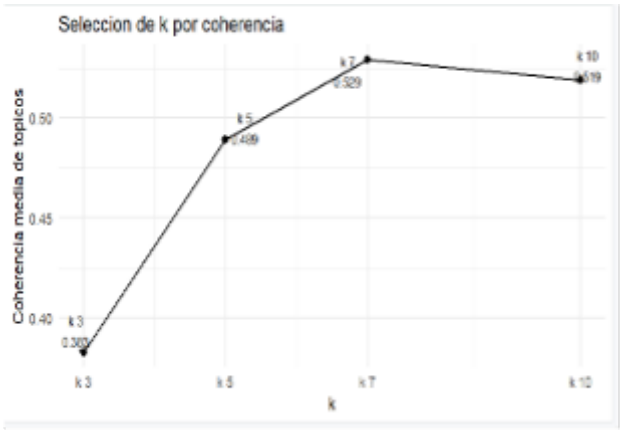

Fig. 3. Coherencia de tópicos vs. número de tópicos k.

nítida de los temas, lo que sirve mejor para el análisis y para las recomendaciones.

VI. Recomendacion y mecanismo de evaluación

Con base a los resultados obtenidos, se recomienda implementar un enfoque hibrido entre el modelo LDA y TF-IDF pues esta integración logra equilibrar el estilo del autor, la diversidad de temas en las recomendaciones dentro del blog. Este enfoque se fundamenta en la complementariedad de ambos modelos, juntos ofrecerían una visión más amplia de los cuentos. El modelo TF-IDF evidenció una notable capacidad para capturar el estilo narrativo propio de Hernán Casciari, caracterizado por la oralidad, el humor cotidiano y las referencias culturales argentinas. Esto se refleja en los altos niveles de similitud entre "Messi es un perro" y cuentos como "Canelones" (similitud 0.63), donde el sistema identificó coincidencias léxicas, tono autobiográfico y estructuras reflexivas típicas del autor. Este modelo logra encontrar la voz narrativa de Casciari. Sin embargo, su principal limitación radica en su enfoque estrictamente superficial del lenguaje, al depender netamente del vocabulario compartido tiende a recomendar textos con palabras similares, pero no necesariamente una diversidad narrativa

En contraste, el modelo LDA permite descubrir la estructura temática que subyace a la obra de Casciari. Este enfoque identifica patrones de coocurrencia entre palabras y los organiza en tópicos conceptuales, ofreciendo una representación más profunda de los contenidos. Gracias a este modelo, fue posible agrupar los cuentos según 7 temas recurrentes, teniendo mayor coherencia que otro número de temas. De esta manera la modelo promueve la variedad conceptual y facilitando el hallazgo de textos que, si bien pueden diferir en vocabulario, comparten el mismo trasfondo emocional o simbólico. Sin embargo, en algunos casos su abstracción excesiva dificulta la conexión entre la recomendación y la percepción directa del estilo narrativo del autor. Por tanto, su utilidad aumenta cuando se combina con un modelo que preserve la coherencia lingüística y tonal. A partir de esto se propone un modelo hibrido, este tipo de modelo permite aprovechar las ventajas

TABLE I
RECOMENDACIONES POR MÉTODO (TF–IDF VS. LDA)

Posición	TF–IDF		LDA	
	Título del cuento	Similitud	Título del cuento	Similitud
1	Messi es un perro	1.00	Messi es un perro	1.00
	Canelones	0.63	Los dos comodines	0.96
2	Instrucciones para crear mundos paralelos	0.49	Donar los órganos	0.95
3	El espectáculo de volar	0.46	Fútbol, fervor e independencia	0.95
4	**¿Cuni… qué?**	0.43	Los payasos	0.94
5	Ni olvido ni perdón	0.41	**¿Cuni… qué?**	0.94

Notas: la columna "Similitud" reporta la similitud del coseno.

de ambos modelos, dando recomendaciones resultantes no solo son coherentes con el estilo de Casciari, sino que también reflejan las múltiples dimensiones que atraviesan su narrativa. El enfoque hibrido asegura que las recomendaciones sean a la vez relevantes y variadas, evitando la redundancia y fomentando la exploración del catálogo completo del autor.

A. Enfoque híbrido y reglas de negocio

Fusión de señales.: Sea x_i el vector TF–IDF normalizado del cuento i y γ_i su vector de distribución documento–tópico (LDA). Para una consulta por cuento q, definimos:

$$s_{\text{lex}}(i \mid q) = \cos(x_i, x_q), \qquad s_{\text{lda}}(i \mid q) = \cos(\gamma_i, \gamma_q).$$

De esta manera definimos la combinación convexa de ambas señales:

$$s(i \mid q) = \alpha\, s_{\text{lex}}(i \mid q) + (1-\alpha)\, s_{\text{lda}}(i \mid q), \quad \alpha \in [0,1].$$

A medida que los usuarios interactúan con el sistema, ajustamos α dinámicamente mediante el proceso de Thompson Sampling con validación cruzada.

Reglas de negocio (restricciones editoriales).

- No auto-recomendación
- Cobertura temática: en top-5, al menos 2 tópicos distintos con peso agregado ≥ 0.25 cada uno.
- Descubrimiento guiado: reservar 1 lugar en top-5 para ítems con $s_{\text{lex}} < 0.35$ y $s_{\text{lda}} > 0.80$ (temas afines con léxico distinto).
- Priorización de nuevos cuentos: en top-5, al menos 1 cuento publicado en los últimos 2 años.

B. Riesgos y mitigaciones

El enfoque propuesto presenta riesgos tanto metodológicos como operativos que deben reconocerse para garantizar la validez y sostenibilidad del sistema de recomendación.

1) Riesgo de sobreajuste de α: El parámetro α define el peso relativo entre las señales léxica (TF–IDF) y temática (LDA). Un valor mal calibrado podría sesgar el sistema hacia recomendaciones redundantes (si α es demasiado alto) o hacia asociaciones conceptuales demasiado abstractas (si es demasiado bajo). *Mitigación:* la selección de α se realiza mediante un mecanismo de evaluación basado en experimentos controlados tipo A/B, donde el **KPI primario** es la *tasa de conversión de lectura* (CTR_{lectura}): la proporción de recomendaciones que el usuario efectivamente lee. La **unidad de análisis** es la *sesión*, para capturar decisiones de lectura contextuales. Cada sesión recibe un valor de α distinto (muestreado por Thompson Sampling), y el sistema ajusta su distribución posterior conforme observa resultados reales. Este proceso iterativo permite converger hacia el α que maximiza el KPI, garantizando equilibrio entre relevancia inmediata y diversidad narrativa.

2) Riesgo de dominio cerrado: El modelo se entrena sobre un corpus estático (520 textos), lo que puede limitar su capacidad para incorporar nuevo material o adaptarse a cambios en el estilo del autor. *Mitigación:* incluir mecanismos periódicos de reentrenamiento incremental y detección de deriva semántica mediante la comparación de distribuciones $p(\text{término} \mid \text{tópico})$ entre iteraciones del modelo. Un umbral basado en la métrica perplexity entre viejos y nuevos cuentos por debajo de cierto número activaría reentrenamiento completo.

3) Riesgo de homogeneización recomendacional: Sin restricciones adicionales, el sistema podría concentrar sus recomendaciones en unos pocos tópicos dominantes, reduciendo la variedad y por ende la exploración del catálogo. *Mitigación:* las reglas de negocio incorporan una cuota mínima de diversidad temática (Subsección VI-A), lo cual fuerza cobertura transversal y promueve descubrimiento guiado.

4) Riesgo de interpretación: LDA ofrece relaciones temáticas de alta abstracción, cuya coherencia conceptual no siempre es evidente para el lector. *Mitigación:* acompañar cada recomendación con indicadores de afinidad estilística y temática, expresados de manera transparente para el usuario final, de modo que la conexión entre textos sea explicable y verificable.

En conjunto, el ajuste adaptativo de α, el monitoreo continuo del modelo y las restricciones editoriales aseguran que el sistema mantenga un balance entre precisión, diversidad y coherencia, ofreciendo recomendaciones significativas tanto desde la perspectiva narrativa como desde la experiencia de usuario.