**UNIVERSIDAD PERUANA LOS ANDES (UPLA)**

**FACULTAD DE INGENIERÍA DE SISTEMAS Y COMPUTACIÓN**

PROGRAMA ACADÉMICO: INGENIERÍA DE SISTEMAS Y COMPUTACIÓN

**Plataforma de Inteligencia Documental con IA para Análisis Estadístico y Generación Automática de Informes**

**TESIS**

Para optar el Título Profesional de Ingeniero de Sistemas y Computación

Autor: Oscar Salas Contreras

Asesor: Carlos Burgos

Huancayo, 2025

# Tabla de contenidos

# Lista de Figuras

# Lista de Tablas

# Declaración de originalidad

El presente trabajo es original y no ha sido presentado previamente para la obtención de un grado o título. Las fuentes consultadas se citan conforme a APA 7.

# Agradecimientos

A la UPLA, a la Facultad y al asesor por su orientación; y a los usuarios piloto por su participación en la validación del sistema.

# Dedicatoria

A mi familia, por su apoyo incondicional.

# Resumen

Agente\_IA\_OyP\_5.0 integra Procesador de Documentos, Servicio de IA, Analizador Estadístico y Generador de Reportes, orquestados por un Gateway. Permite ingesta multiformato, selección/entrenamiento de modelos por proveedor, análisis estadístico completo con contrato estándar (tables/figures/summary) y generación de informes en APA 7 (PDF/DOCX). La arquitectura de microservicios y jobs asíncronos asegura escalabilidad y trazabilidad. La precisión del modelo se reporta como entero (p. ej., 93).

# Abstract

Agente\_IA\_OyP\_5.0 is an end‑to‑end document intelligence platform unifying ingestion/cleaning, provider‑selectable AI training/inference, a full statistical analysis suite with a standardized output, and APA‑compliant reporting to PDF/DOCX. Microservices and asynchronous jobs provide scalability and traceability. Model accuracy is reported as an integer KPI (e.g., 93).

# 1. Introducción

Los flujos de ingesta, análisis y reporte suelen ser fragmentados. Esta tesis propone una plataforma unificada que automatiza el proceso y garantiza la reproducibilidad mediante contratos estables y trazabilidad extremo a extremo.

# 2. Marco Teórico

Gestion documental y OCR; NLP para clasificación, resumen y NER; estadística aplicada (descriptiva, inferencial, correlaciones y regresiones); normas APA 7 para la comunicación de resultados.

# 3. Metodología

Diseño experimental (grupo control vs. experimental). Variables: tiempo (min), errores (%), satisfacción (Likert 1–5), precisión del modelo (entera). Instrumentos: rúbricas, SUS/NPS, logs del sistema.

# 4. Arquitectura y Diseño del Sistema

Cuatro módulos orquestados por Gateway: Procesador de Documentos, Servicio de IA, Analizador Estadístico y Generador de Reportes. Persistencia de artefactos y versionado. Contratos de API para datasets, modelos, análisis y reportes.

# 5. Implementación

Frontend con Tabler 1.4 (SPA, estilo SPSS), selector de proveedor/modelo IA, wizard de 4 pasos y Centro de Jobs. Backend en microservicios Flask; Redis para colas; SQLite (dev) y Postgres (prod).

# 6. Resultados

## 6.1 Configuración experimental

Ambiente local con Docker Compose; Redis para colas; SQLite en desarrollo. Documentos multiformato categorizados (tesis, informes, cotizaciones, plantillas, fotos, otros).

## 6.2 Descriptiva general

Table . Estadística descriptiva por variable

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variable** | **n** | **Media** | **DE** | **Mín** | **Máx** |
| Tiempo (min) |  |  |  |  |  |
| Errores (%) |  |  |  |  |  |
| Satisfacción (1–5) |  |  |  |  |  |
| Precisión (entera) |  |  |  |  |  |

## 6.3 Supuestos

Table . Pruebas de supuestos (normalidad y homocedasticidad)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable/Grupo** | **Prueba** | **Estadístico** | **p** | **Conclusión** |
| Tiempo | Shapiro–Wilk |  |  |  |
| Tiempo | Levene |  |  |  |
| Errores | Shapiro–Wilk |  |  |  |
| Errores | Levene |  |  |  |

## 6.4 Pruebas de hipótesis

Table . Pruebas paramétricas y no paramétricas

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Prueba** | **Estadístico** | **gl** | **p** | **Tamaño del efecto** | **IC** |
| t (ind/pareada) / ANOVA / Welch |  |  |  |  |  |
| Mann–Whitney / Kruskal–Wallis |  |  |  |  |  |
| Chi² / Fisher |  |  |  |  |  |

## 6.5 Correlaciones

Table . Correlaciones (Pearson/Spearman/Kendall)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Par** | **Método** | **r/ρ/τ** | **p** | **n** |
| Var1–Var2 | Pearson |  |  |  |
| Var1–Var3 | Spearman |  |  |  |
| Var2–Var3 | Kendall |  |  |  |

## 6.6 Modelado

Table . Regresiones y métricas de ajuste

|  |  |  |  |
| --- | --- | --- | --- |
| **Modelo** | **Parámetros clave** | **Métricas (R²/AUC)** | **Diagnósticos** |
| Lineal |  |  |  |
| Logística |  |  |  |
| Poisson |  |  |  |

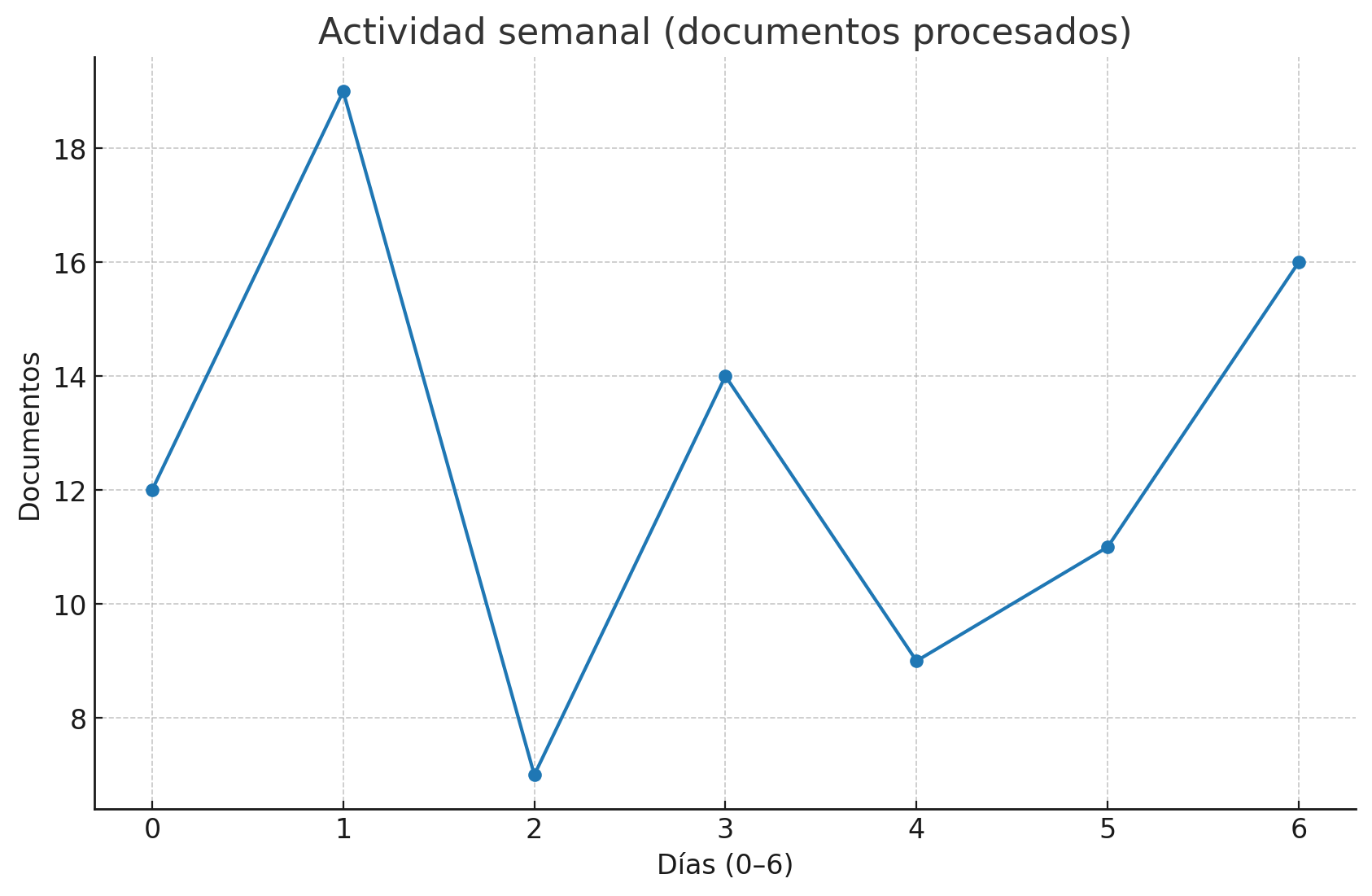


Figure . Actividad semanal de procesamiento de documentos (simulada).

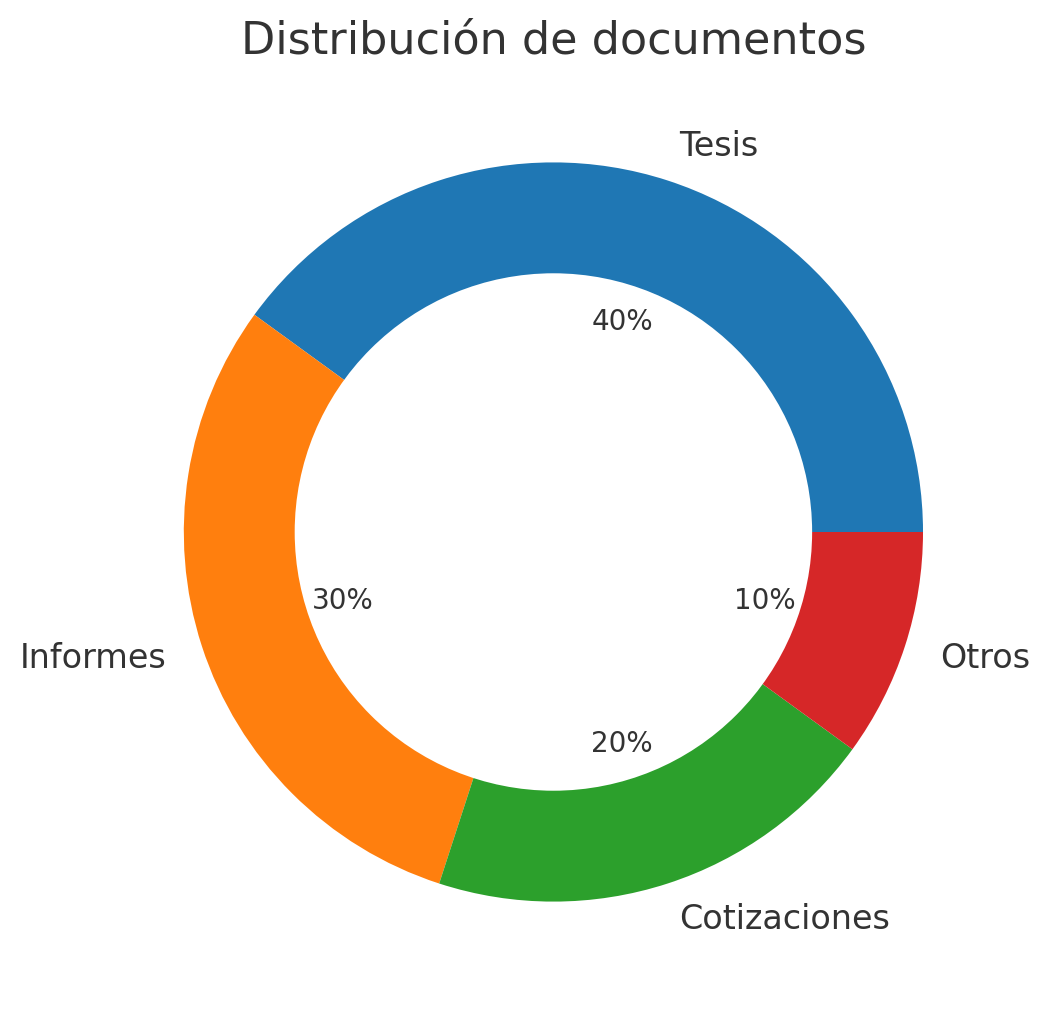


Figure . Distribución de documentos por tipo (simulada).

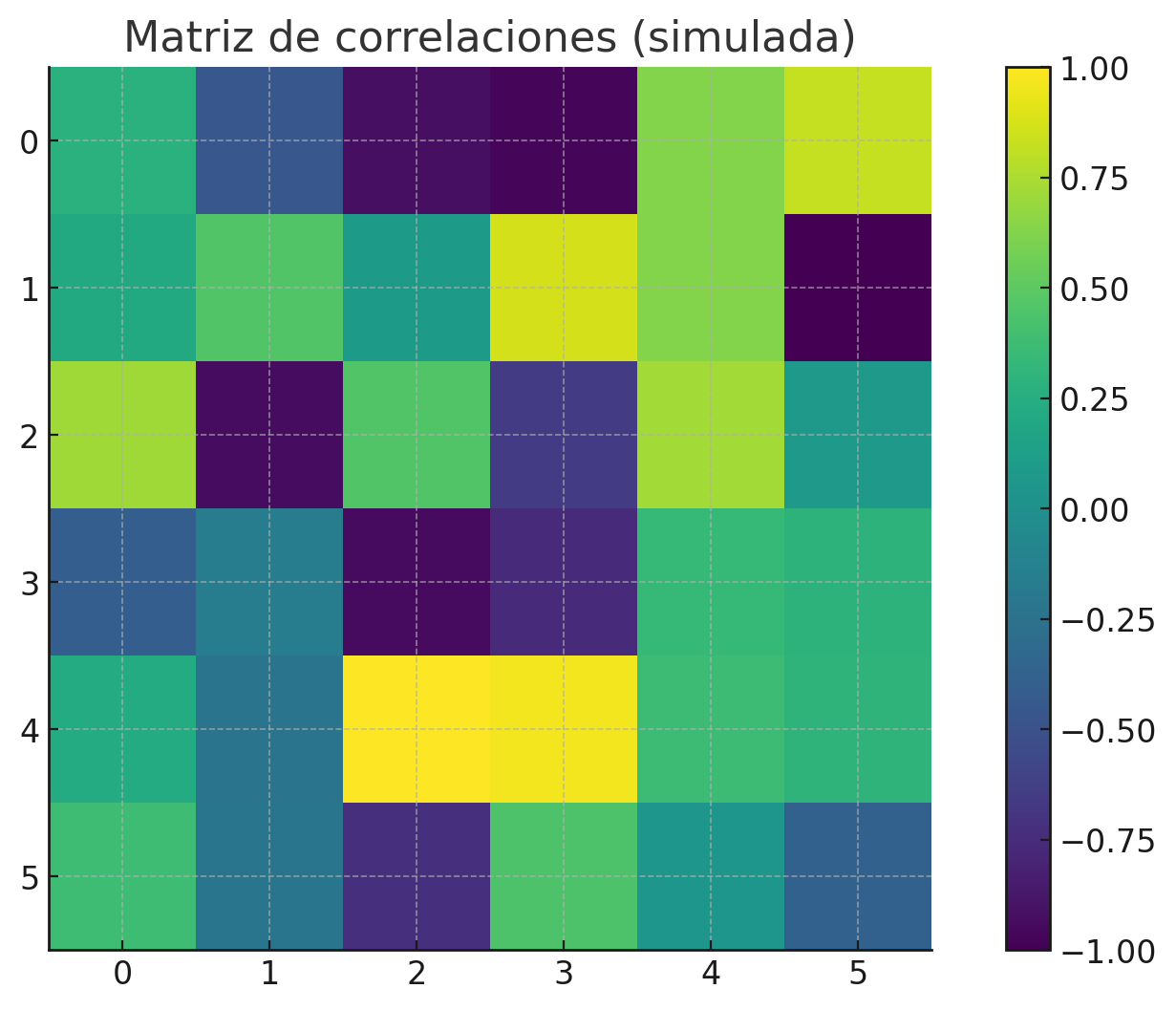


Figure . Matriz de correlaciones sobre variables operativas (simulada).

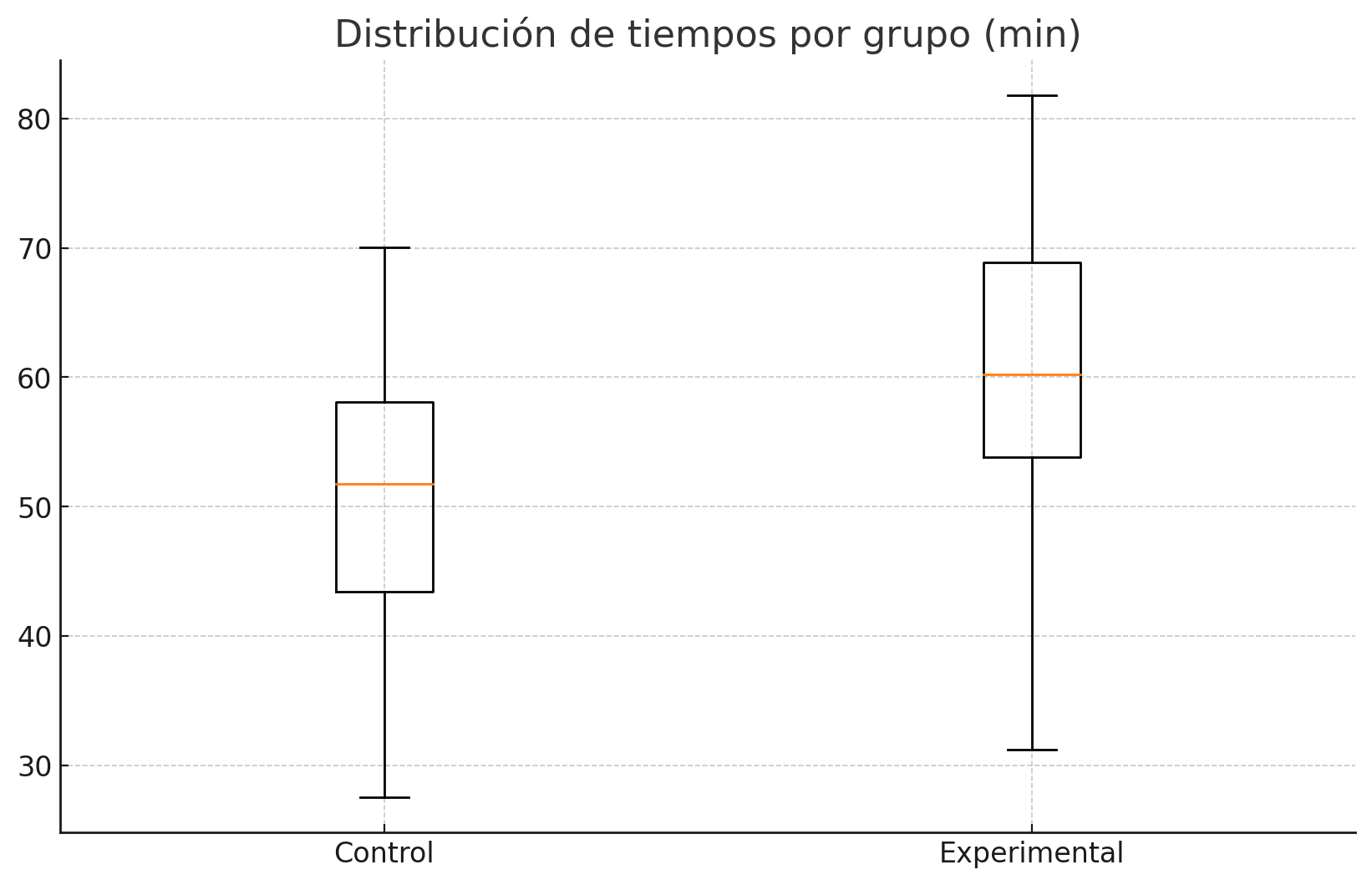


Figure . Distribución de tiempos por grupo (control vs. experimental).

# 7. Discusión

La arquitectura de 4 módulos y el contrato estándar mejoran eficiencia y reproducibilidad. Amenazas: tamaño muestral, heterogeneidad de documentos, sesgos. Limitaciones: SSO y plantillas avanzadas pendientes.

# 8. Conclusiones y Recomendaciones

La plataforma es viable y útil; se recomienda fortalecer el Centro de Jobs, ampliar plantillas APA y habilitar SSE/WebSocket y MLflow.

# 9. Trabajo Futuro

Progreso en tiempo real, editor visual de plantillas, MLflow para experimentos, SSO y políticas de retención, análisis avanzados.

# 10. Referencias (APA 7)

American Psychological Association. (2020). Publication manual of the American Psychological Association (7th ed.). APA.

Field, A. P. (2013). Discovering statistics using IBM SPSS statistics (4th ed.). SAGE.

Virtanen, P., et al. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. Nature Methods, 17, 261–272.

van der Walt, S., Colbert, S. C., & Varoquaux, G. (2011). The NumPy array: A structure for efficient numerical computation. Computing in Science & Engineering, 13(2), 22–30.

Flask documentation. https://flask.palletsprojects.com

Docker Compose documentation. https://docs.docker.com/compose/

Chart.js documentation. https://www.chartjs.org

Tabler UI documentation. https://tabler.io

# 11. Anexos

## Anexo A. Contratos de API (resumen)

GET /api/estadisticas/resumen, GET /api/estadisticas/graficos; POST /api/datasets; GET /api/datasets/:id; GET /api/ia/providers, GET /api/ia/models, POST /api/ia/test, POST /api/services/analyze-text; POST /api/analysis/jobs, GET /api/analysis/jobs/:id, GET /api/analysis/results/:id; GET /api/templates, GET /api/templates/:id/preview, POST /api/reports, GET /api/reports/:id.

## Anexo B. Modelo de datos (ER)

documents, datasets, dataset\_items, models (precision\_int), analysis\_jobs, analysis\_results, reports, templates, categories.

## Anexo C. Despliegue (Docker Compose)

Servicios: gateway, document\_processor, ai\_service, analysis, report\_generator, redis, db. Volúmenes: storage/, models/, reports/.