

Documento de Ejemplo Extenso - PDF

Este es un documento PDF más largo, creado para demostrar la extracción de texto desde archivos PDF que contienen texto nativo. Su objetivo es permitir que los estudiantes del curso verifiquen cómo los modelos de Machine Learning procesan documentos reales o simulados antes de convertirlos en representaciones numéricas como TF-IDF.

El contenido incluye un análisis ficticio del comportamiento legislativo en una nación imaginaria, considerando estadísticas de proyectos de ley, indicadores de eficiencia parlamentaria y métricas de calidad regulatoria. Asimismo, se mencionan las dificultades comunes en la digitalización de documentos, la necesidad de sistemas OCR robustos y los desafíos técnicos asociados al procesamiento de archivos provenientes de múltiples formatos como CSV, PDF, Word y bases de datos externas.

Se discuten también temas relevantes como el uso de modelos de clasificación para categorizar artículos de prensa, la construcción de sistemas RAG para respuestas contextualizadas, y la integración de herramientas de IA generativa en flujos de trabajo institucionales.

El documento está diseñado intencionalmente con párrafos extensos para asegurar que los experimentos con TF-IDF produzcan vectores suficientemente densos y que los modelos puedan capturar patrones reales durante el entrenamiento.