

Proyecto de Inteligencia Artificial

Bootcamp IAp191

Ejecutor: Víctor Aguilar

17 de septiembre de 2025

Índice

Estructura de Carpetas del Proyecto	3
1. Introducción	4
2. Definición del Problema	4
3. Revisión de Literatura	4
4. Metodología	5
4.1. Datos	5
4.2. Modelos	5
4.3. Evaluación	5
5. Resultados	5
6. Discusión	6
7. Conclusiones	6

8. Referencias

6

Estructura de Carpetas del Proyecto

Para mantener orden y reproducibilidad, se recomienda la siguiente estructura:

```
project_name/  
  README.md          # Explica propósito, uso y pasos de ejecución  
  requirements.txt    # Librerías necesarias (p.ej. scikit-learn, keras)  
  
  data/  
    raw/              # Datos originales (no se modifican nunca)  
    processed/        # Datos limpios listos para modelos  
  
  notebooks/          # Jupyter para análisis exploratorio y prototipos  
  results/  
    metrics/          # Reportes en CSV/JSON con accuracy, F1, etc.  
    figures/           # Gráficas exportadas (curvas, matrices de confusión)  
  docs/               # Documento final en LaTeX/PDF
```

Nota para estudiantes: - Nunca sobrescribir los datos en `raw/`. - Documentar cambios de preprocesamiento en `processed/`. - Guardar gráficos y métricas en `results/`. - En `notebooks/` mantener experimentos claros con celdas ordenadas.

1 Introducción

Aquí el estudiante explica el problema en un lenguaje sencillo:

- ¿Qué tema eligió?
- ¿Por qué le interesa?
- ¿A quién beneficia resolverlo?

Ejemplo: “El análisis de sentimientos en reseñas de películas es útil para que plataformas de cine automaticen la clasificación de opiniones y entiendan la experiencia del usuario. Este proyecto explora un modelo de aprendizaje automático para predecir si una reseña es positiva o negativa”.

2 Definición del Problema

Se formula el problema de forma clara y medible.

- Objetivo general.
- 2–3 objetivos específicos.

Ejemplo: - **Objetivo general:** Desarrollar un clasificador de reseñas de películas en positivas o negativas. - **Objetivos específicos:** 1. Preparar un conjunto de datos de reseñas en español e inglés. 2. Entrenar un modelo clásico (Naive Bayes). 3. Entrenar un modelo profundo (LSTM) y comparar resultados.

3 Revisión de Literatura

Los estudiantes deben consultar al menos 2 fuentes (libros, artículos, manuales). **Ejemplo de redacción sin cita:** “Diversos estudios han demostrado que los modelos de redes neuronales recurrentes capturan mejor las dependencias en texto que los algoritmos basados en conteo de palabras. Por otro lado, los modelos probabilísticos simples siguen siendo competitivos cuando los datos son limitados”.

4 Metodología

4.1 Datos

- Indicar origen del dataset.
- Explicar cómo se dividieron (train/test).
- Describir pasos de limpieza.

Ejemplo: “El dataset IMDB contiene 50 mil reseñas, divididas en 25 mil para entrenamiento y 25 mil para prueba. Se realizó tokenización y truncamiento a 200 palabras por reseña”.

4.2 Modelos

Los estudiantes deben entrenar: - Un modelo clásico de ML. - Un modelo de Deep Learning.

Ejemplo: “Se entrenó un Naive Bayes con suavizado de Laplace y un modelo LSTM con embedding de 128 dimensiones y 64 unidades recurrentes”.

4.3 Evaluación

Explicar métricas y cómo se usaron.

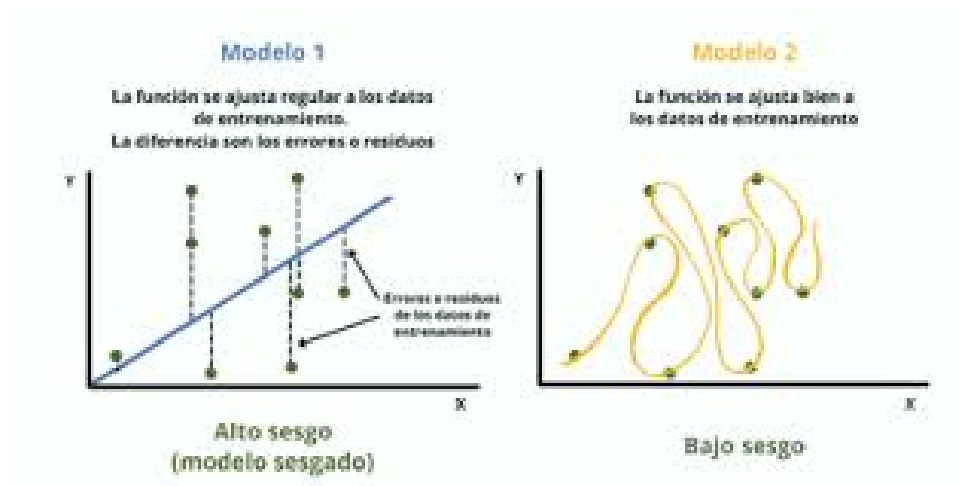
Ejemplo de tabla:

Modelo	Accuracy	Precisión	Recall	F1
Naive Bayes	0.85	0.84	0.85	0.85
LSTM	0.90	0.89	0.91	0.90

5 Resultados

Se muestran gráficas y métricas finales.

Ejemplo: “El modelo LSTM superó al Naive Bayes en todas las métricas, especialmente en F1”.



Curva de aprendizaje del LSTM.

6 Discusión

Aquí los estudiantes analizan críticamente:

- ¿Por qué un modelo funcionó mejor?
- ¿Qué limitaciones tuvo el experimento?
- ¿Qué mejoras proponen?

Ejemplo: “El LSTM obtuvo mejor rendimiento porque captura dependencias entre palabras. Sin embargo, requiere más tiempo de entrenamiento y recursos computacionales”.

7 Conclusiones

Se resumen aprendizajes en 4–6 líneas.

Ejemplo: “Se logró clasificar reseñas con un nivel aceptable de precisión. Los algoritmos de Deep Learning muestran mejores resultados que los clásicos, aunque requieren mayor infraestructura. Una futura extensión es probar transformers”.

8 Referencias

Aquí los estudiantes deben redactar sus fuentes según el formato que se indique (APA 7 u otro).