

PRÁCTICA FINAL

PIPELINE DEL PROYECTO

Pipeline del Proyecto

- 1. Definición del Problema
- 2. Recolección de Datos
- 3. Preprocesamiento de Datos
- 4. Análisis Exploratorio
- 5. Selección y Entrenamiento del Modelo
- 6. Evaluación del Modelo
- 7. Optimización y Ajuste
- 8. Despliegue y Puesta en Producción
- 9. Monitoreo y Mantenimiento
- 10. Documentación

DEFINICIÓN DEL PROBLEMA

1. Descripción de la Práctica

 Poner en práctica todos los conocimientos del bootcamp simulando una situación real de desarrollo en equipo.

 Desarrollo usando SCRUM, coordinando trabajo propio y del equipo.

 Proyecto real con plazo ajustado. Entregar el mejor y más completo prototipo posible.

 Backlog de historias de usuario. No se espera completar todo, se valora

Ideas de Proyecto (I)

A. Asistente virtual para atención al cliente

Responder consultas en servicios financieros o e-commerce.

B. Análisis automático de sentimientos en RRSS

Analizar sentimiento en publicaciones usando LLMs, ML,...

C. Generación de contenido para marketing

Crear contenido persuasivo para campañas publicitarias. A partir de datos, y generando informes con LLMs,...

D. Sistema de recomendación de productos

Sugerir productos según preferencias del usuario.

Ideas de Proyecto (II)

F. Detección de patologías en resonancias magnéticas torácicas

G. Sistema agente para ayuda en documentos

H. Personalización de contenido educativo

I. Análisis automático de noticias y tendencias

J. Automatización de algunos procesos de la empresa



Temática libre. Escoger modelo considerando escalabilidad y costes.

DATOS Y PREPROCESAMIENTO

3. Recolección de Datos

📌 Identificación de datos relevantes

Datos estructurados o no estructurados según el problema.

📌 Fuentes de recopilación

BD internas, datasets públicos, web scraping, Hugging Face, datos sintéticos.

📌 Evaluación de calidad

Verificar integridad, precisión, consistencia y actualidad.

📌 Cumplimiento normativo

GDPR, CCPA. Anonimizar datos y obtener consentimiento.

4. Preprocesamiento de Datos

Limpieza de datos

Eliminar ruido, valores atípicos, duplicados. Imputación.

Transformaciones

Normalización, escalado, codificación categórica (one-hot, ordinal).

Datos desbalanceados

Submuestreo/sobremuestreo. Algoritmos específicos.

Selección de características

PCA, correlación, árboles de decisión, regularización.

5. Análisis Exploratorio de Datos

Exploración y visualización

Gráficos, scatter plots, heatmaps, boxplots.

Patrones y tendencias

Relaciones lineales/no lineales, clusters, tendencias temporales.

Análisis de correlaciones

Pearson, Spearman, pruebas de hipótesis entre variables.

Comprensión de sesgos

Sesgos heredados del dataset y su impacto en la solución.

MODELADO Y EVALUACIÓN

6. Selección y Entrenamiento del Modelo



Selección del algoritmo

Clasificación: SVM, Random Forest, k-NN, redes neuronales.

Regresión: lineal, logística, árboles de decisión.

LLMs: entrenar de 0 o finetunear modelo instructed.



División de datos

Conjuntos de entrenamiento, validación y prueba.



Entrenamiento del modelo

7. Evaluación del Modelo



Métricas de evaluación

Clasificación: precisión, recall, F1, AUC-ROC.

Regresión: MSE, MAE, R². LLMs: MMLU, HELM, HellaSwag.



Validación cruzada

Dividir, entrenar, evaluar por separado, promediar resultados.



Estudio de ablación / Cohortes

Evaluar robustez sobre subgrupos del dataset. Crítico en medicina y seguridad.

8. Optimización y Ajuste

Selección de características

PCA, eliminación de redundancia, ingeniería de características.

Ajuste de hiperparámetros

Búsqueda en cuadrícula, optimización bayesiana.

Iteración en el proceso

Ciclos de entrenamiento y evaluación hasta rendimiento óptimo.

9. Despliegue y Puesta en Producción



Implementación del modelo

Integración con sistemas actuales. APIs e interfaces de usuario.

Garantizar buen funcionamiento en condiciones reales.



Pruebas adicionales

Tests de integración, carga y estrés.

Detectar problemas y garantizar estabilidad del sistema.

10. Monitoreo y Mantenimiento



Sistema de monitoreo

Vigilar constantemente el desempeño en producción.

Métricas en tiempo real, detección de degradaciones, alertas.



Actualizaciones periódicas

Reentrenar con datos actualizados.

Agregar funciones para mayor precisión.

Ajustar ante cambios en el entorno.

11. Documentación



Documentar todo el proceso

Desde recolección hasta implementación. Transparencia y reproducibilidad.



Estructura del informe

Suposiciones iniciales. Métricas y justificación. Métodos utilizados.

Análisis de sesgos. Lecciones aprendidas y conclusiones.



Comunicación de hallazgos

Resultados con datos cuantitativos y visualizaciones.

12. Logística y Trabajo en Equipo

Scrum Board en Trello

Historias de usuario en Trello. Burndown Chart actualizado.

Coordinación de equipo

Planificación por Sprint. Scrum Meetings diarios si es posible.

Rotación de roles

Tras cada sprint, roles rotan para practicar todas las tecnologías.

13. Formación de Grupos

Mínimo 2 personas, máximo 4

Se permite trabajo individual pero no se recomienda.

Libertad para decidir formación de grupos.

14. Evaluación y Presentación Final



Presentación ante instructores (10-15 min + 10-15 min Q&A)



Orden: Equipo → Pitch → Demo → Reflexiones



Criterios de evaluación:

Calidad del backlog y código. Estabilidad del prototipo.

Viabilidad económica. Calidad y claridad del pitch.

keep coding

keep coding



www.keepcoding.io



cursos@keepcoding.io



(+34) 916 33 1779