

Documentación para el Desarrollo de Productos Basados en Datos

Hecha

Yilena Mosquera Rivas

Enlace Operativo



Fecha

30 de agosto de 2023

Introducción

Esta documentación, elaborada por Yilena Mosquera Rivas, tiene como objetivo proporcionar una guía detallada sobre cómo acompañar y llevar a cabo actividades para el desarrollo de productos basados en datos utilizando Python y análisis de datos para realizar predicciones futuras. El proceso se dividirá en varias etapas, desde la adquisición y preparación de datos hasta la implementación de modelos de predicción.

Etapa 1: Adquisición y Preparación de Datos

RECÓPILACION DE DATOS :

- Identificar las fuentes de datos necesarias para el proyecto.
- Utilizar bibliotecas de Python como pandas y NumPy para importar y cargar los datos en un formato adecuado.

EXPLORACIÓN DE DATOS :

- Realizar un análisis exploratorio de datos (EDA) para comprender la estructura de los datos.
- Utilizar visualizaciones, como gráficos de dispersión y histogramas, para identificar patrones y tendencias.
- Tratar valores faltantes o atípicos según sea necesario.

LIMPIEZA Y PREPROCESAMIENTO :

- Limpiar los datos eliminando duplicados y registros irrelevantes.
- Aplicar técnicas de preprocesamiento como la codificación de variables categóricas y la normalización de datos numéricos.

LIMPIEZA Y PREPROCESAMIENTO :

- Limpiar los datos eliminando duplicados y registros irrelevantes.
- Aplicar técnicas de preprocesamiento como la codificación de variables categóricas y la normalización de datos numéricos.

Etapla 2: Análisis de Datos y Predicciones

SELECCIÓN DE CARACTERÍSTICAS :

- Identificar las características relevantes para el análisis y las predicciones.
- Utilizar técnicas como la selección de características y la ingeniería de características para mejorar la calidad de los datos.

MODELADO :

- Seleccionar el algoritmo de regresión logística de Python para el modelado de datos.
- Dividir los datos en conjuntos de entrenamiento y prueba utilizando bibliotecas como scikit-learn.
- Entrenar el modelo utilizando los datos de entrenamiento.

EVALUACIÓN DEL MODELO :

- Utilizar métricas de evaluación como la precisión, la exhaustividad y la puntuación F1 para evaluar el rendimiento del modelo.
- Realizar validación cruzada para asegurar la robustez del modelo.

Etapla 3: Implementación y Monitoreo

IMPLEMENTACIÓN :

- Implementar el modelo de regresión logística en una aplicación o sistema, utilizando bibliotecas como Flask o Django si es necesario.

MONITOREO CONTINUO :

- Establecer un sistema de monitoreo para supervisar el rendimiento del modelo en tiempo real.
- Actualizar y retrenar el modelo periódicamente para mantener su precisión.

CONCLUSIONES

Este proceso de desarrollo de productos basados en datos utilizando Python y análisis de datos es fundamental para tomar decisiones informadas y realizar predicciones precisas. Es importante recordar que el éxito de un proyecto de este tipo depende en gran medida de la calidad de los datos, la elección adecuada de características y la evaluación cuidadosa del modelo.