

Tarea: Creación de un Modelo de Clasificación para Predecir Problemas en Procesos de un Servidor

Objetivo

Desarrollar un modelo de clasificación que pueda predecir si un proceso en ejecución en un servidor causará problemas, basándose en datos históricos de los procesos. Este modelo ayudará en la detección temprana de procesos potencialmente problemáticos, permitiendo tomar acciones preventivas para asegurar la estabilidad y el rendimiento del servidor.

Datos

El conjunto de datos proporcionado incluye las siguientes características para cada proceso en el servidor:

- **ID_Proceso**: Identificador único del proceso.
- **Uso_CPU**: Porcentaje del CPU utilizado por el proceso.
- **Uso_Memoria**: Porcentaje de memoria utilizada por el proceso.
- **Numero_Hilos**: Número de hilos del proceso.
- **Tiempo_Ejecucion**: Tiempo de ejecución del proceso en horas.
- **Numero_Errores**: Número de errores generados por el proceso en las últimas 24 horas.
- **Tipo_Proceso**: Categoría del proceso (Servicio, Aplicación, Sistema).

La variable objetivo, **Estado**, indica si un proceso es problemático (1) o no (0), basada en el uso de recursos, número de errores, tipo de proceso, y otros factores relevantes.

Tareas a Realizar

1. **Análisis Exploratorio de Datos (EDA)**: Realice un análisis preliminar para entender la distribución y las características de los datos. Esto incluye verificar valores faltantes, la distribución de las variables numéricas, y la frecuencia de las categorías en variables categóricas.
2. **Preprocesamiento de Datos**:
 - Limpieza: Maneje valores faltantes y elimine duplicados si los hay.
 - Codificación de variables categóricas: Utilice técnicas como One-Hot Encoding para convertir **Tipo_Proceso** en variables numéricas.
 - Escalado de características: Normalice o escale las características numéricas para que tengan el mismo rango de valores, lo cual es importante para algunos modelos de clasificación.
3. **Selección y División del Conjunto de Datos**: Divida el conjunto de datos en entrenamiento, validación y prueba para evaluar la efectividad del modelo.
4. **Construcción y Evaluación de Modelos**:
 - Pruebe diferentes algoritmos de clasificación (por ejemplo, Regresión Logística, Árboles de Decisión, Bosques Aleatorios, y Máquinas de Soporte Vectorial) para encontrar el más adecuado.
 - Utilice validación cruzada para optimizar los hiperparámetros y evaluar la robustez del modelo.
 - Evalúe el modelo final utilizando el conjunto de prueba y métricas adecuadas (precisión, recall, puntuación F1, y curva ROC).

5. **Interpretación de Resultados y Conclusiones:** Interprete los resultados del modelo en el contexto del problema. Identifique las características más importantes que influyen en la predicción de procesos problemáticos y discuta cómo se puede utilizar el modelo en un entorno de producción.
6. **Reporte:** Prepare un informe detallado que cubra todos los pasos realizados, desde el EDA hasta la interpretación de los resultados, incluyendo cualquier insight relevante obtenido durante el proceso.

Entrega

La tarea final consistirá en un Jupyter Notebook que contenga todo el código utilizado para el análisis, preprocesamiento de datos, modelado, y evaluación, junto con un reporte escrito (dentro del mismo notebook o como un documento separado) que resuma los hallazgos y las recomendaciones basadas en el modelo desarrollado.