

Laboratorio 7

Predicción de Diabetes con AutoGluon

Descripción:

El objetivo de este ejercicio es construir, entrenar y evaluar un modelo clasificador para predecir diabetes en pacientes usando AutoGluon. Además de la práctica técnica, se busca reflexionar sobre las ventajas, limitaciones y aplicaciones del AutoML en problemas del mundo real, especialmente en contextos de salud donde la interpretación y confiabilidad de los resultados son fundamentales.

Instrucciones:

1. Carga del Conjunto de Datos:

- Cree un DataFrame con el contenido del archivo `diabetes.csv`.
- Revise y describa brevemente las variables del conjunto de datos (ej. glucosa, presión sanguínea, IMC, edad, etc.).
- Explique el contexto: la columna 'Outcome' representa si el paciente tiene (1) o no tiene (0) diabetes.

2. Análisis Exploratorio de Datos (EDA):

- Obtenga estadísticas descriptivas básicas del conjunto de datos.
- Visualice la distribución de las variables con histogramas y boxplots.
- Verifique la presencia de valores nulos o atípicos y decida cómo manejarlos.
- Analice el balance de clases de la variable 'Outcome'.
- Genere una matriz de correlación y un mapa de calor para identificar relaciones entre variables.

3. Entrenamiento con AutoGluon:

- Separe los datos en conjuntos de entrenamiento y prueba.
- Utilice la opción ``best_quality`` del parámetro “preset” y la métrica ``accuracy``.
- Limite el tiempo de entrenamiento (``time_limit``) para evitar ejecuciones largas.
- Entrene varios modelos de clasificación con AutoGluon para predecir la columna 'Outcome'.
- Compare los resultados con un modelo base (ej. regresión logística).
- Revise la importancia de las características (feature importance).

4. Evaluación del Modelo:

- Despliegue una tabla de los mejores modelos en función de la precisión y destaque el mejor modelo.
- Visualice la matriz de confusión del mejor modelo.
- Calcule métricas adicionales: precisión, recall y F1-score.
- Genere y analice la curva ROC y el AUC.

5. Reflexión:

- Redacte una reflexión crítica sobre su experiencia:
- ¿Qué ventajas y desventajas encontró al usar AutoGluon y AutoML en general?
- ¿Qué métricas considera más relevantes en este problema?
- ¿Qué precauciones deberían tomarse al aplicar este tipo de herramientas en salud?
- ¿Cómo compara esta experiencia con construir un modelo manualmente?

Rúbrica de Evaluación:

Criterio	Puntaje Máximo
Carga y preparación del conjunto de datos	15 puntos
EDA adecuado (incluyendo visualizaciones y correlaciones)	25 puntos
Uso correcto de AutoGluon (parámetros y comparación con baseline)	20 puntos
Evaluación adecuada del los modelos (incluyendo visualización de la tabla de líderes y matriz de confusión)	25 puntos
Reflexión crítica y análisis de AutoML	10 puntos
Claridad y documentación (Markdown, visualizaciones)	5 puntos
Puntaje Total	100 puntos

Recomendaciones:

- Documente cada paso en un cuaderno Jupyter (usando celdas Markdown para explicar).
- Utilice gráficos y visualizaciones para apoyar sus hallazgos y decisiones.
- Interprete y justifique sus decisiones basadas en los resultados.
- Considere aspectos técnicos, de usabilidad y de aplicabilidad en situaciones reales.
- Puede usar Google Colab para facilitar la ejecución del laboratorio.