

Informe de Resultados - Amazon Employee Access Challenge

Análisis y Justificación de la Elección del Modelo CatBoost

Este proyecto tiene como objetivo predecir si un empleado debería tener acceso a un recurso específico dentro de la empresa utilizando atributos relevantes.

Se ha utilizado un enfoque basado en Machine Learning para entrenar varios modelos, uno de ellos siendo CatBoost optimizado mediante Bayesian Optimization.

A continuación, se presentan los resultados y el análisis de las características más importantes.

Resultados del Modelo CatBoost Optimizado:

Mejores Hiperparámetros encontrados:

- Subsample: 1.0
- Scale_pos_weight: 3
- Learning_rate: 0.1
- Iterations: 200
- Depth: 5
- Colsample_bylevel: 0.8

Reporte de Clasificación:

- Precisión (Clase 0 - Acceso Permitido): 1.00
- Precisión (Clase 1 - Acceso Denegado): 0.91
- Recall (Clase 0): 1.00
- Recall (Clase 1): 0.80
- F1-Score (Clase 0): 1.00

- F1-Score (Clase 1): 0.85
- Accuracy Total: 1.00
- Macro Avg (F1-score): 0.93

Justificación de la Elección del Modelo CatBoost:

Se eligió CatBoost debido a su capacidad de manejar clases desbalanceadas de manera eficiente, lo cual es esencial en este contexto, donde los accesos denegados son mucho menos frecuentes que los permitidos. La optimización bayesiana permitió encontrar hiperparámetros que maximizan tanto la precisión como el recall, especialmente en la clase minoritaria (accesos denegados). Esto es crucial en un entorno donde es más importante evitar falsos negativos (accesos permitidos incorrectamente).

Además, CatBoost tiene ventajas computacionales importantes, incluyendo su manejo eficiente de características categóricas y su capacidad para trabajar con grandes volúmenes de datos sin requerir un preprocesamiento extenso. El uso de gráficos SHAP ha mostrado que las características clave, como 'V4', 'V14' y 'Time', tienen una alta influencia en las predicciones del modelo, lo que justifica su capacidad para interpretar patrones en los datos.

Gráficos SHAP:

Se generaron gráficos SHAP para evaluar la importancia de las características en las predicciones del modelo CatBoost.

A continuación se muestran los gráficos que destacan la influencia de las variables clave en el modelo.



