

Informe Académico: Predicción del Abandono Estudiantil mediante Árboles de Decisión

El presente informe describe el proceso de análisis y modelado realizado para predecir el abandono estudiantil utilizando un algoritmo de Árbol de Decisión. El objetivo fue identificar los factores que más influyen en la continuidad académica de los estudiantes, a partir de un conjunto de datos institucionales. El trabajo se dividió en dos etapas principales: entrenamiento del modelo y generación de predicciones finales.

En la primera etapa, implementada en el script **ArbolEntrenamiento.py**, se realizó la carga y limpieza de los datos provenientes del archivo **TablaPrediccionAbandono-Entrenamiento.xlsx**. Posteriormente, se codificaron las variables categóricas para su tratamiento por el modelo y se separaron las variables predictoras (X) de la variable objetivo (*estado_final*). Se entrenó un Árbol de Decisión con el criterio de entropía, profundidad máxima de 6 y un mínimo de 5 observaciones por hoja. El modelo fue evaluado mediante una matriz de confusión, mostrando una alta precisión al diferenciar entre estudiantes que continuaron y los que abandonaron. También se analizaron la distribución del estado final y la importancia de las variables, identificando los factores más determinantes en la predicción del abandono.

En la segunda etapa, desarrollada en el script **ArbolFinal.py**, se reutilizó el modelo entrenado para realizar predicciones sobre nuevos datos contenidos en el archivo **TablaPrediccionAbandono-DatosFinal.xlsx**. El proceso incluyó la codificación de las variables de entrada de forma coherente con el entrenamiento y la generación de un nuevo archivo Excel (**Predicciones_Final.xlsx**), que contiene las predicciones individuales para cada estudiante. El modelo demostró una capacidad adecuada para generalizar los patrones aprendidos, con un reporte de clasificación que confirma su buen desempeño.

Entre los principales resultados, se destaca que las variables con mayor importancia en la predicción fueron aquellas relacionadas con la asistencia, desempeño académico y cantidad de materias aprobadas. La interpretación del árbol permite observar reglas claras de decisión que podrían ayudar a la institución a detectar tempranamente a los estudiantes con mayor riesgo de abandono.

Recomendaciones para la institución: Monitorear de forma constante a los estudiantes con bajo rendimiento o baja asistencia. Implementar programas de apoyo y tutorías virtuales. Desarrollar estrategias de retención temprana basadas en los predictores identificados. Prestar especial atención a quienes aprueban una o menos materias en el primer cuatrimestre.

En conclusión, el modelo de Árbol de Decisión permitió obtener una herramienta predictiva de utilidad práctica para el ámbito educativo. La integración de análisis exploratorio, visualizaciones y predicciones finales genera una base sólida para la toma de decisiones institucionales orientadas a reducir el abandono académico.