



Modelo Analítico para el Estudio de la Retención Estudiantil en la Facultad de Ciencias y Tecnologías de la Universidad Nacional de Caaguazú

Autor: Fabrizio Villar Ferreira

Tutor: Prof. Ing. Victor Manuel Melgarejo Riveros



Introducción y Contexto

1) La retención estudiantil es un problema relevante caracterizado por trayectorias discontinuas y bajas tasas de culminación

2) Abordaje desde perspectiva analítica utilizando datos históricos y técnicas de aprendizaje automático

3) Objetivo: comprender patrones de permanencia y detectar estudiantes en riesgo

Problema y Definición Operativa



Grandes Volúmenes de Datos

Datos académicos disponibles sin análisis sistemático



Definición de Retención

Estudiante retenido registra actividad dentro de máximo dos años



Clasificación de Riesgo

Diferenciación entre retrasos temporales y abandono consolidado

Datos y Población Analizada

75,937

Registros Académicos

1,422

Estudiantes

100%

Trayectorias Completas

Dataset Longitudinal

Registros académicos históricos de eventos como materias cursadas y exámenes realizados por estudiantes.

- Múltiples registros por estudiante individual
- Eventos académicos detallados y cronológicos
- Cobertura completa de trayectorias estudiantiles

Proceso de Consolidación

Automatización para agrupar registros por estudiante y generar observación única por alumno.

- Agrupación automatizada de eventos académicos
- Generación de métricas agregadas por estudiante
- Análisis de trayectoria académica completa

Construcción de Variables



Variables de Rendimiento Académico

- Calificaciones promedio
- Tasas de aprobación



Variables de Avance Curricular

- Materias cursadas
- Progreso en la carrera

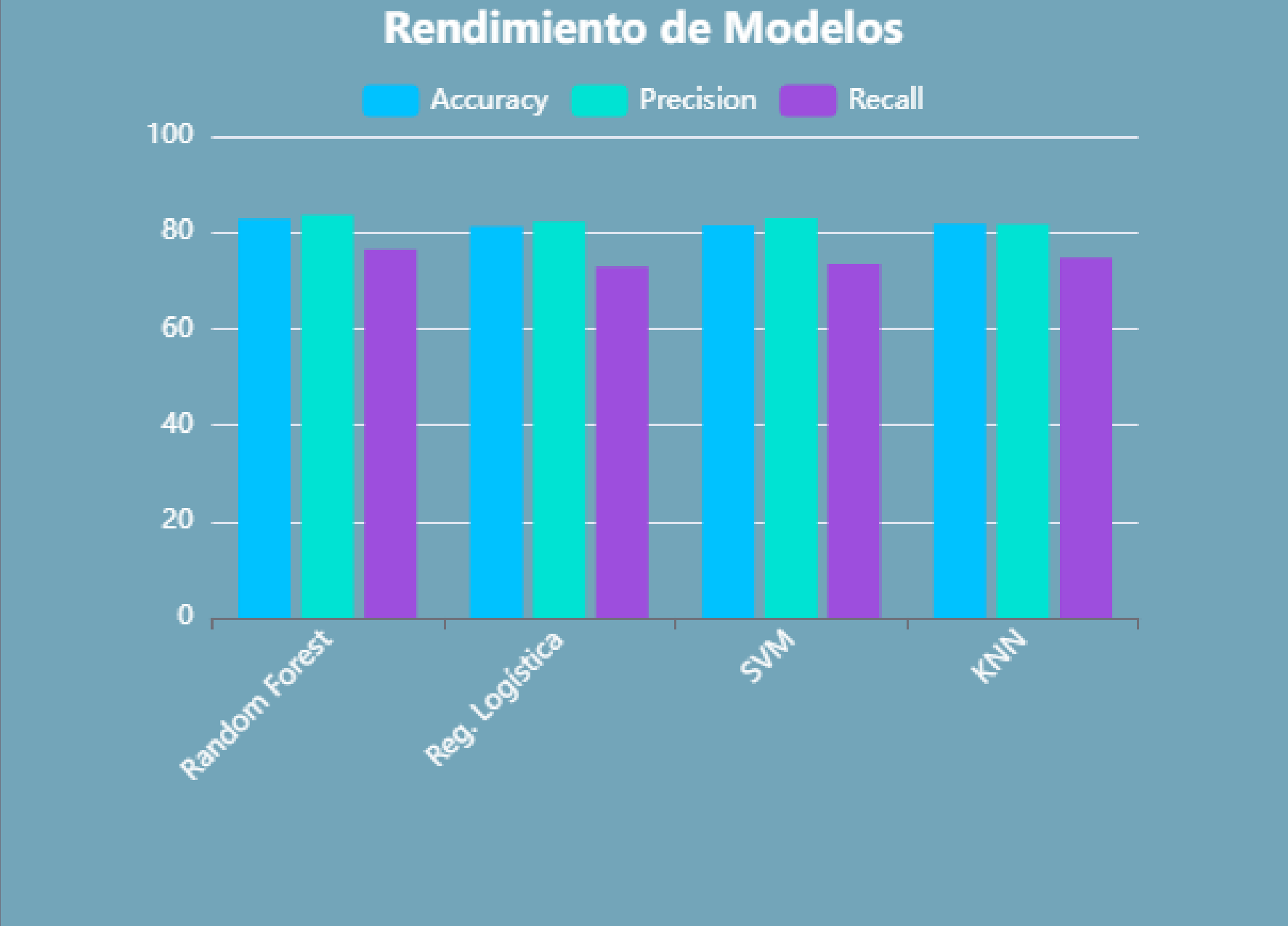


Variables Temporales

- Inactividad académica
- Permanencia

Nota: Todos los cálculos se realizan automáticamente mediante software desarrollado específicamente.

Modelos Utilizados y Evaluación



Modelos Evaluados

Random Forest,
Regresión Logística,
SVM y KNN con
métricas de
clasificación
supervisada

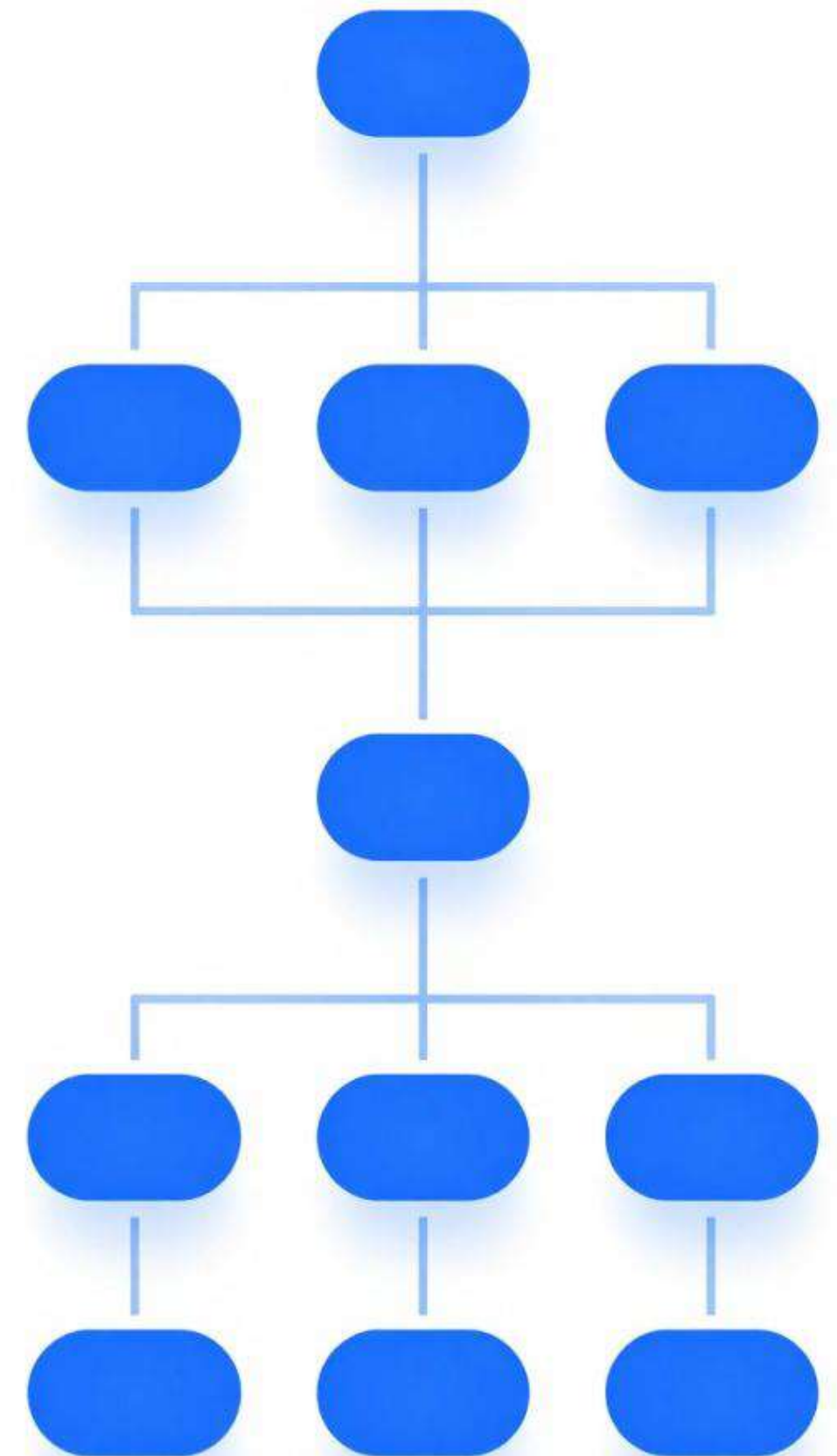


Recall Prioritario

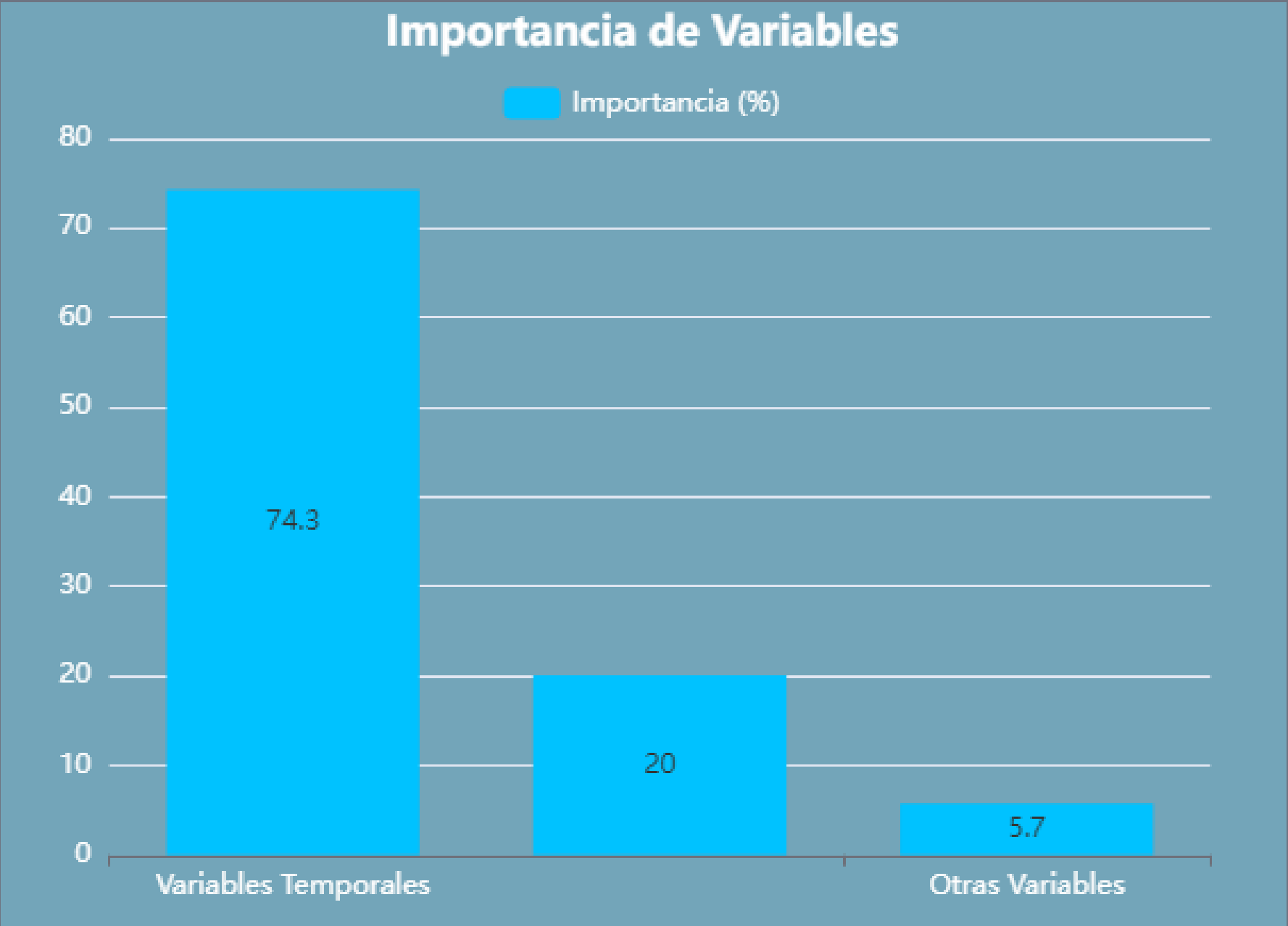
Métrica clave para
identificar
correctamente
estudiantes en
riesgo de abandono
académico

Importancia de Variables - Índice de Gini

El índice de Gini constituye una métrica fundamental para cuantificar la separación entre clases en algoritmos de árboles de decisión. Esta medida teórica evalúa la impureza de las particiones, donde valores cercanos a cero indican nodos puros. En la implementación práctica de Random Forest, el índice opera como criterio para seleccionar divisiones óptimas durante el crecimiento de los árboles. La ventaja clave radica en su capacidad de extracción automática, permitiendo la generación de rankings de importancia variable mediante scripts desarrollados en entornos como Python o R.



Resultados Principales



Hallazgos Clave

Random Forest mostró el mejor desempeño global. Las variables temporales de inactividad concentran más del 74% de la importancia total del modelo.

- Continuidad académica: factor principal de retención
- Rendimiento académico: impacto secundario en modelo
- Inactividad prolongada: principal indicador de riesgo

Conclusiones y Aplicaciones



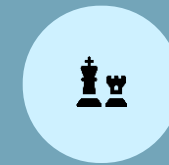
Viabilidad Analítica

Retención estudiantil analizable
mediante modelos predictivos con datos
reales



Factor Crítico Identificado

Inactividad académica como principal
indicador versus calificaciones
secundarias



Base para Implementación

Constitución de base analítica para
sistemas de alerta temprana y
estrategias institucionales

A 3D block with the letters 'AI' on top, set against a background of a glowing blue circuit board. The block is dark blue and has a glowing blue 'AI' on its top surface. The background is a dark blue circuit board with glowing blue lines and patterns.

Gracias por su Atención

Agradezco su participación y estaré complacido de responder sus preguntas sobre el modelo analítico desarrollado.