Sistema Inteligente de Consejería Académica

Erich Giusseppe, Santiago Martínez, David Fuquen y Nicolás Klopstock

El Problema Actual

- Miles de estudiantes con perfiles diversos
- Recomendaciones basadas en experiencia, no es datos
- Falta de consistencia y personalización de recomendaciones

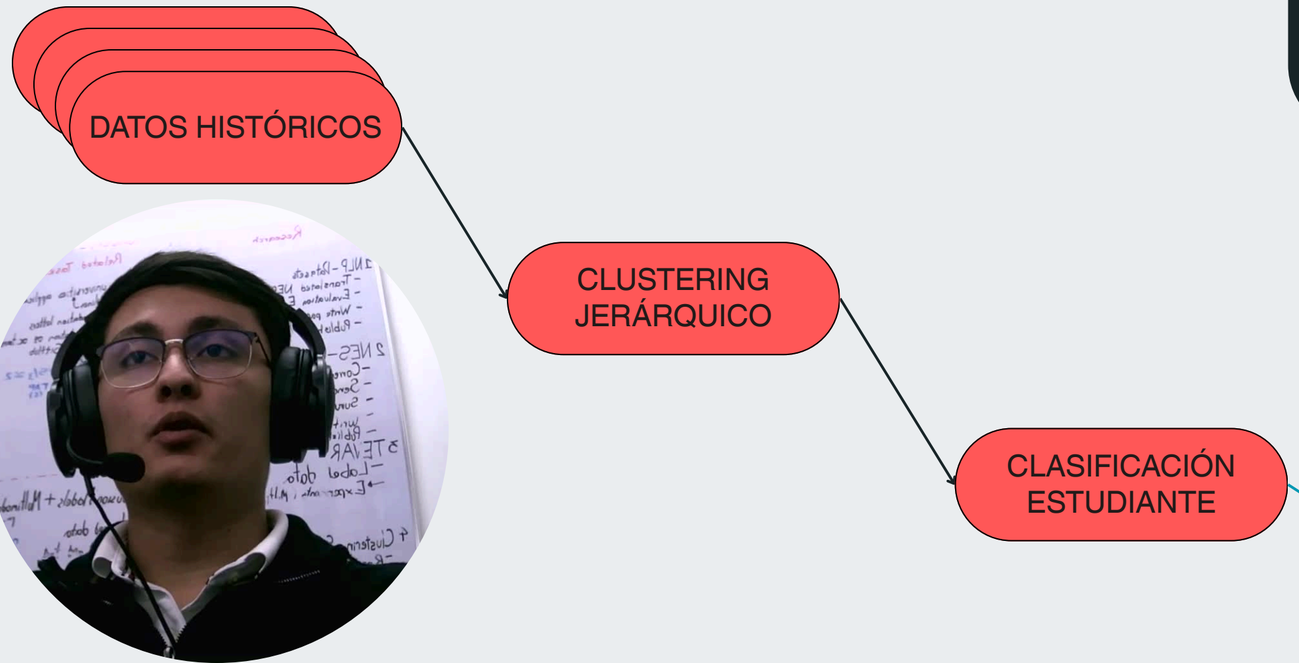
Propuesta de Solución

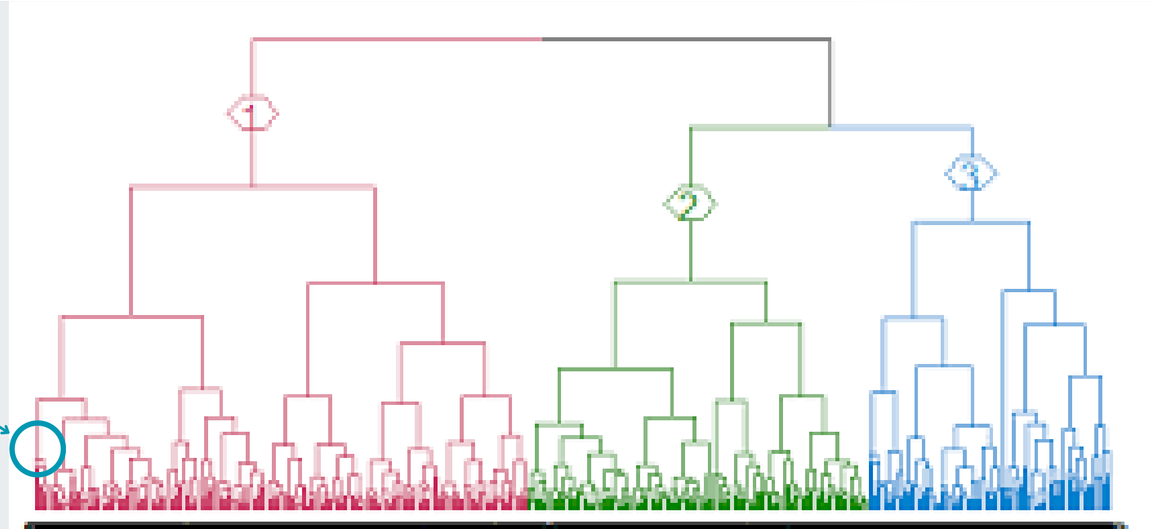
Agrupar estudiantes según patrones históricos de desempeño usando +600,000 trayectorias académicas.

Clustering multinivel:

- Segmenta por programa e identifica perfiles similares.

"Cada recomendación se fundamenta en evidencia de miles de casos similares, no solo en intuición"



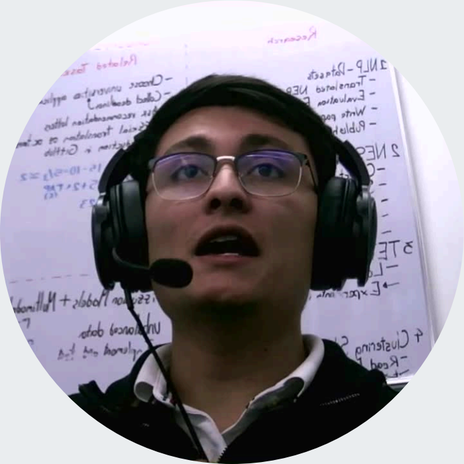


Impacto medible del sistema

CORTO PLAZO (Semestre 1-2)

Eficiencia Operativa:

- Reducción tiempo de consejería: -30%
- Recomendaciones fundamentadas: 90%
- Satisfacción estudiantil: >4.0/5.0



LARGO PLAZO (Año 1-2)

Resultados Académicos:

- Promedio Ponderado Acumulado: +0.2 puntos
- Tasas de aprobación: +5-8%
- Reducción deserción: -10%

El sistema

COMPONENTE ANALÍTICO

- Clustering multinivel adaptativo por programa
- Integración de 8 datasets del observatorio institucional

COMPONENTE TECNOLÓGICO

- API de clasificación en tiempo real
- Aplicación web intuitiva para consejeros

USUARIOS DEL SISTEMA

- Primarios: Consejeros académicos
- Indirectos: Estudiantes

OUTPUTS PRINCIPALES

- Cluster asignado con índice de confianza
- Exploración de escenarios alternativos
- Recomendaciones fundamentadas en casos similares

Mock-ups

Academic Advising
Data Analytics Platform

Students

Analytics

Student Search

Search by name, ID, or email...

EJ Emma Johnson
S001
Computer Science GPA: 3.65

MC Michael Chen
S002
Business GPA: 3.42

SW Sarah Williams
S003
Engineering GPA: 3.78

JM James Martinez
S004
Biology GPA: 2.95

OB Olivia Brown
S005
Psychology GPA: 3.51

Academic History

Semester	Course	Title	Credits	Grade	GPA
Fall 2024	CS 301	Data Structures	4	A	4.0
Fall 2024	MATH 210	Calculus II	4	B+	3.3
Fall 2024	ENG 202	Technical Writing	3	A-	3.7
Spring 2024	CS 201	Programming Fundamentals	4	A	4.0
Spring 2024	MATH 110	Calculus I	4	B	3.0
Spring 2024	PHYS 101	Physics I	3	B+	3.3

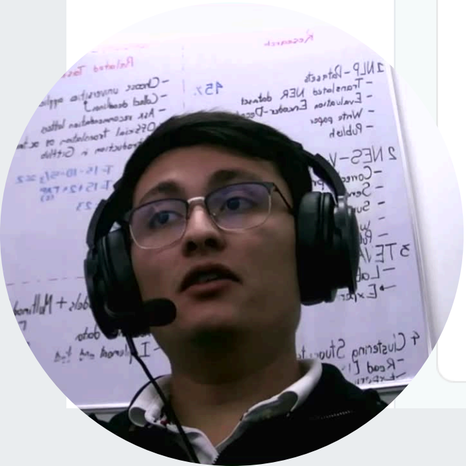
Proposed Course Load - Spring 2025

Edit

Course	Title	Credits	Difficulty
CS 401	Algorithms	4	High
CS 350	Database Systems	3	Medium
MATH 310	Linear Algebra	4	Medium
ENG 301	Professional Communication	3	Low

Total Credits 14

Make Prediction



Mock-ups

Academic Advising
Data Analytics Platform

Students

Analytics

Predicted Student Cluster

High Achievers - STEM Focus

Cluster ID
Cluster 3

Similar Students
245 students

Key Characteristics:

- Strong performance in quantitative courses
- High engagement in technical electives
- Above average study time (15-20 hrs/week)
- Active participation in coding clubs

Confidence Index
Model accuracy

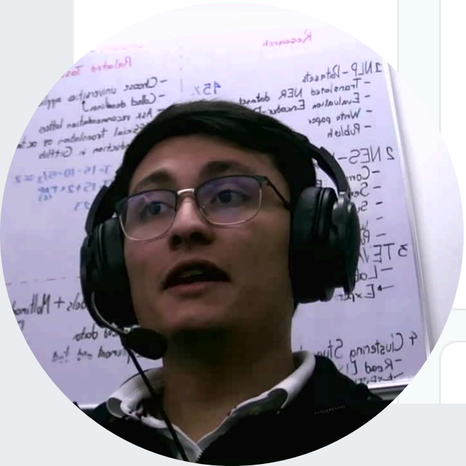
87%

High Confidence

Risk Assessment

Course Load	35%
Difficulty Mix	45%
Prerequisites Gap	15%
Time Management	25%

Overall Risk: Low-Moderate
Student shows strong foundation. Monitor course load balance for optimal performance.



Student Position in Cluster Distribution

Consideraciones Éticas

Constitución colombiana (*Artículo 15*) - *Ley 1581 de 2012* (protección de datos personales)

Datos anonimizados y encriptados

Transparencia algorítmica y auditorías periódicas

Estudiantes podrán conocer y refutar las clasificaciones

Enfoque analítico

Las 8 Fuentes de Datos

Modelo multinivel adaptativo para capturar variabilidad entre programas basado en enfoques de Clarke (2008) y Ochoa et al. (2016) para estructuras jerárquicas educativas.

8 datasets anonimizados (616k–4.9M registros).

Fuentes: rendimiento, materias, riesgos, finanzas, percentiles, horarios.

Dataset	Registros	Variables	Tipo
Historial Estados	616K	21	Temporal/Histórico
Historial Materias	4.9M	15	Temporal/Granular
Rendimiento Académico	611K	22	Temporal/Métricas
Info Actual Estudiantes	222K	30	Snapshot/Demográfico
Horarios Cursos	444K	16	Por periodo
Riesgos Pregrado	369K	12	Alertas/Binario
Info Financiera	793K	12	Histórico
Percentiles Académicos	158K	11	Relativo/Programa

Calidad de Datos - Hallazgos Principales

La calidad general de los datos es excelente, con menos del 1% de valores nulos en variables críticas de rendimiento y financiera. Se identificaron tres situaciones especiales: variables de segundo programa con más del 80% de nulos que fueron descartadas, datos de ICFES con 66% de nulos limitados a análisis específicos, y el dataset de riesgos donde los nulos son correctos por diseño, representando ausencia de factores de riesgo académico.

Variables Críticas Identificadas

Se identificaron variables clave en cuatro dimensiones estratégicas: capacidad académica mediante PERCENTIL_PGA_PROGRAMA que normaliza el rendimiento por programa, experiencia a través de GRUPO_SEMESTRES y total de semestres matriculados, indicadores de riesgo como tres o más materias retiradas y porcentaje de aprobación menor al 50%, y restricciones institucionales como créditos máximos y tipo de matrícula que limitan las recomendaciones.

CAPACIDAD ACADÉMICA	EXPERIENCIA	RIESGO	RESTRICCIONES
PERCENTIL_PGA_PROGRAMA	SEMESTRE_SEGÚN_CREDITOS	CREDITOS_REPROBADOS	CREDITOS_MAXIMOS
PGA	TOTAL_SEMESTRES	CREDITOS_RETIRADOS	TIPO_MATRICULA
% CRÉDITOS APROBADOS	GRUPO_SEMESTRES	3+ MATERIAS RETIRADAS	ESTADO_ACADÉMICO
PROMEDIO_SEMESTRAL	CREDITOS_PGA	% APROBACIÓN <50%	PROGRAMA_1

Conclusiones

Validación exploratoria confirma viabilidad técnica y estratégica.

Futuras fases:

- Limpieza e integración de datos.
- Modelamiento multinivel.
- Desarrollo tecnológico.
- Piloto y despliegue institucional.

Thank you
Danke