UNIVERSIDAD AUSTRAL

UNIVERSIDAD AUSTRAL

Deserción Universitaria

Maestría en Data Mining - Trabajo Final

Gastón Gadea

La deserción en números

- 27% de los alumnos se gradúan, 40% en universidades privadas (La Nación, 18 de junio de 2013)
- La mayoría de las bajas se dan entre el 1º y 2º año de estudios: 58% en el 1º año (Interuniversidades.com, 2012)
- Causas más mencionadas en U. Austral: vocacionales, exigencia académica, compatibilidad de horarios.

UNIVERSIDAD AUSTRAL

Objetivo

- Académico
- Elaborar modelos que permitan
  - Detección temprana de candidatos a darse de baja
  - Análisis de posibles causas
- Probar Data Mining Multi-Relacional

UNIVERSIDAD AUSTRAL

- 5 Carreras de Grado de Sedes Bs. As. de la U. Austral con más de 10 cohortes de graudados
  - Lic. en Comunicación
  - Ing. Industrial
  - Ing. en Informática
  - Derecho
  - Medicina
- Información disponible en el SIA
- Interés: Secretaría Académica – Dirección de Estudios

Límite y Alcance

UNIVERSIDAD AUSTRAL

UNIVERSIDAD NACIONAL DE MISIONES

- *Identificación de Causales de Abandono de Estudios Universitarios. Uso de Procesos de Explotación de Información* (Kuna, García Martínez, Villatoro, 2010, UNAM)
  - Árboles de decisión
  - Parte de información demográfica y académica
  - Encuentra 2 reglas que concentran el 33% de las bajas:
    - Regularizó no más de una materia en el 1º año, costea los estudios con su trabajo y tiene título Bachiller.
    - Regularizó no más de una materia en el 1º año, costea los estudios con el aporte de familiares u otros, tiene 3 o menos finales desaprobados/ausentes en el 1º año, pasaron entre 8 y 15 años entre el secundario y el ingreso a la universidad, viaja para ir a clases.

Antecedentes locales

UNIVERSIDAD AUSTRAL

PURDUE UNIVERSITY®

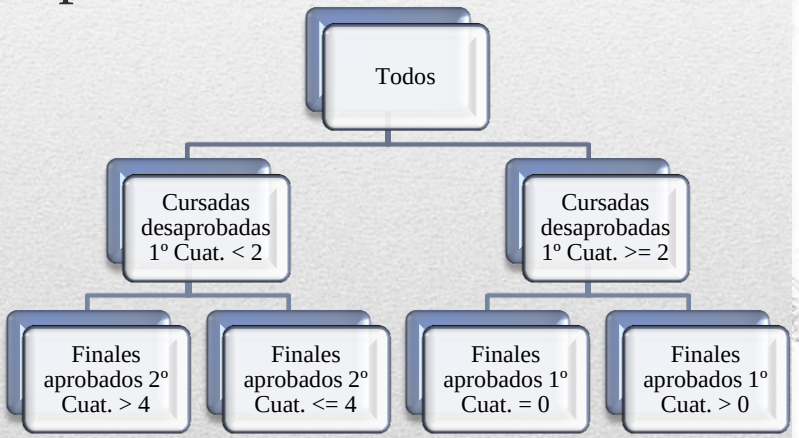
- *Course Signals at Purdue: Using Analytics to Increase Student Success* (Arnold y Pistilli, 2012, Purdue University)
  - Interacción de alumnos en *Blackboard Vista*
  - Implementa acciones preventivas:
    - Sistema de semáforos
    - Seguimiento de instructores
  - Mejora de retención del 82% al 96%

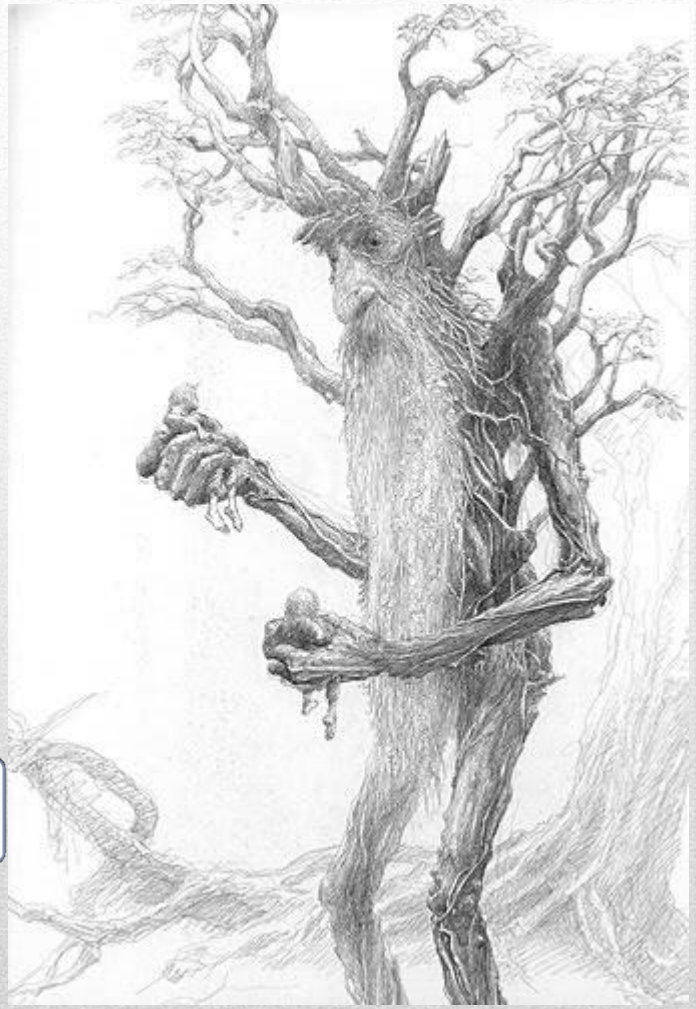
Antecedentes internacionales

UNIVERSIDAD AUSTRAL

La decisión de usar Árboles de Decisión

- Buena clasificación
- Fácil interpretación





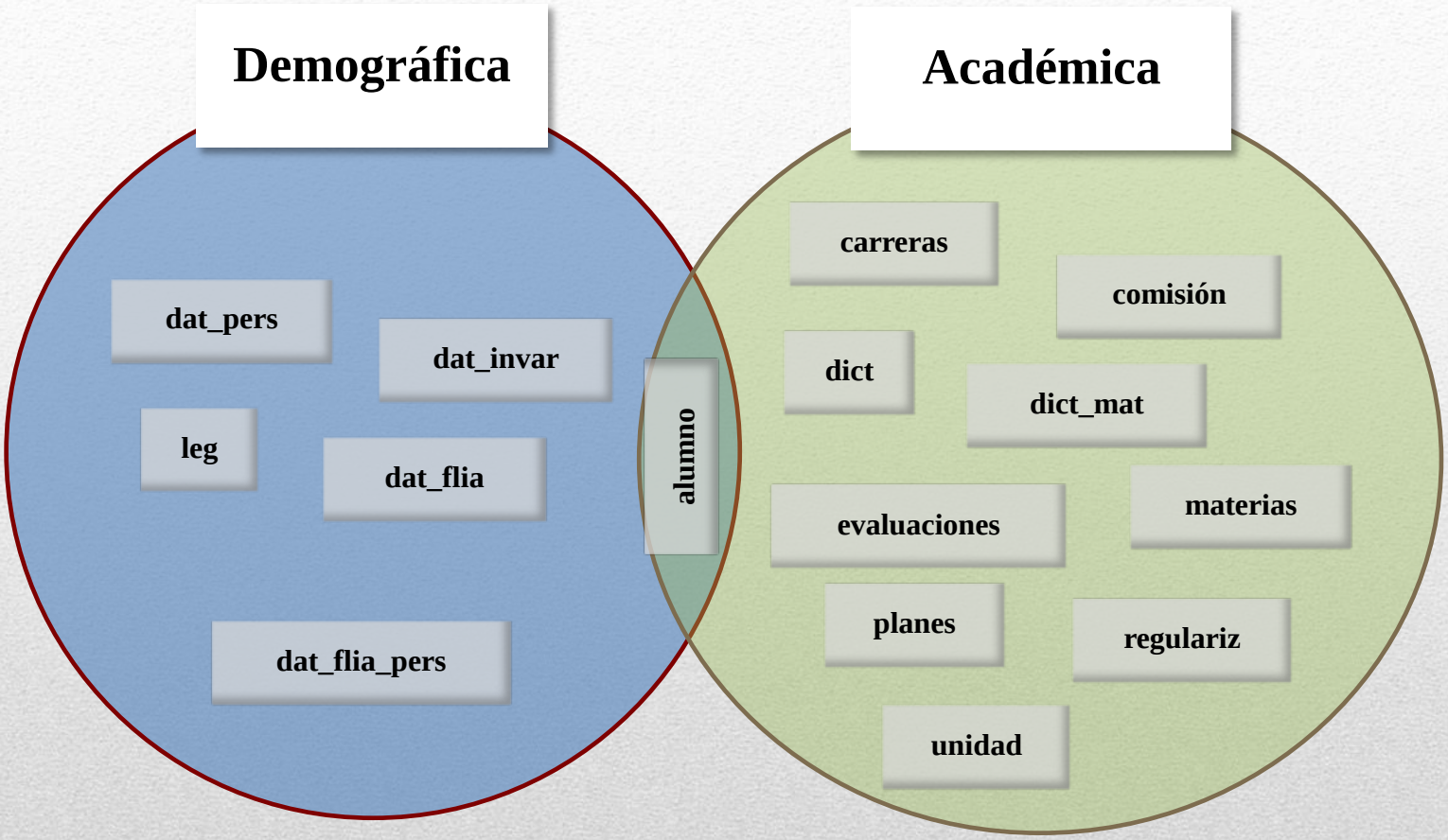
Método de clasificación

Herramienta

Microsoft

- Business Intelligence Development Studio
  - Analysis Services – Data Mining Multi-relacional

- Enterprise Miner



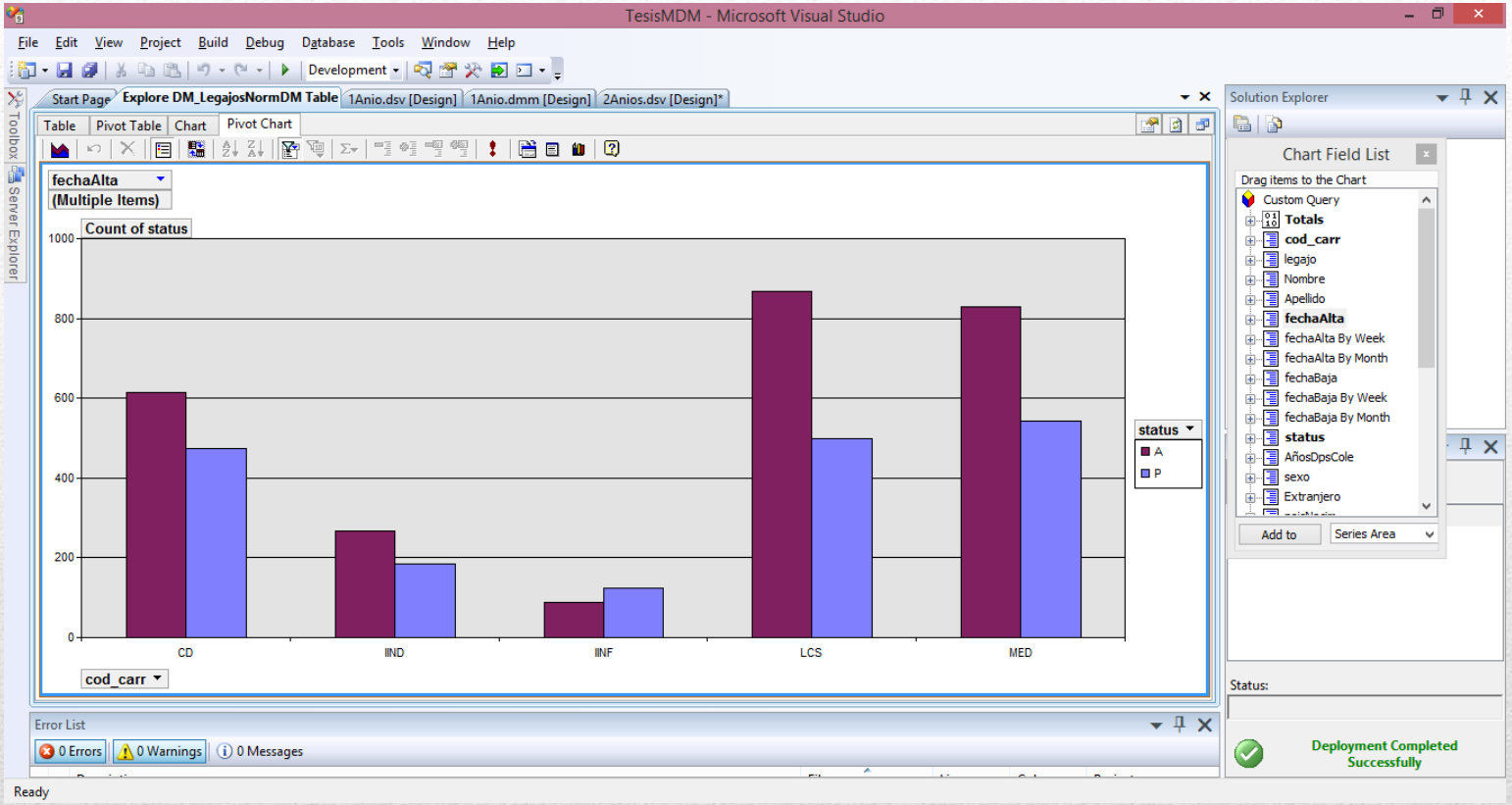
Fuente de datos

UNIVERSIDAD AUSTRAL

Variables disponibles

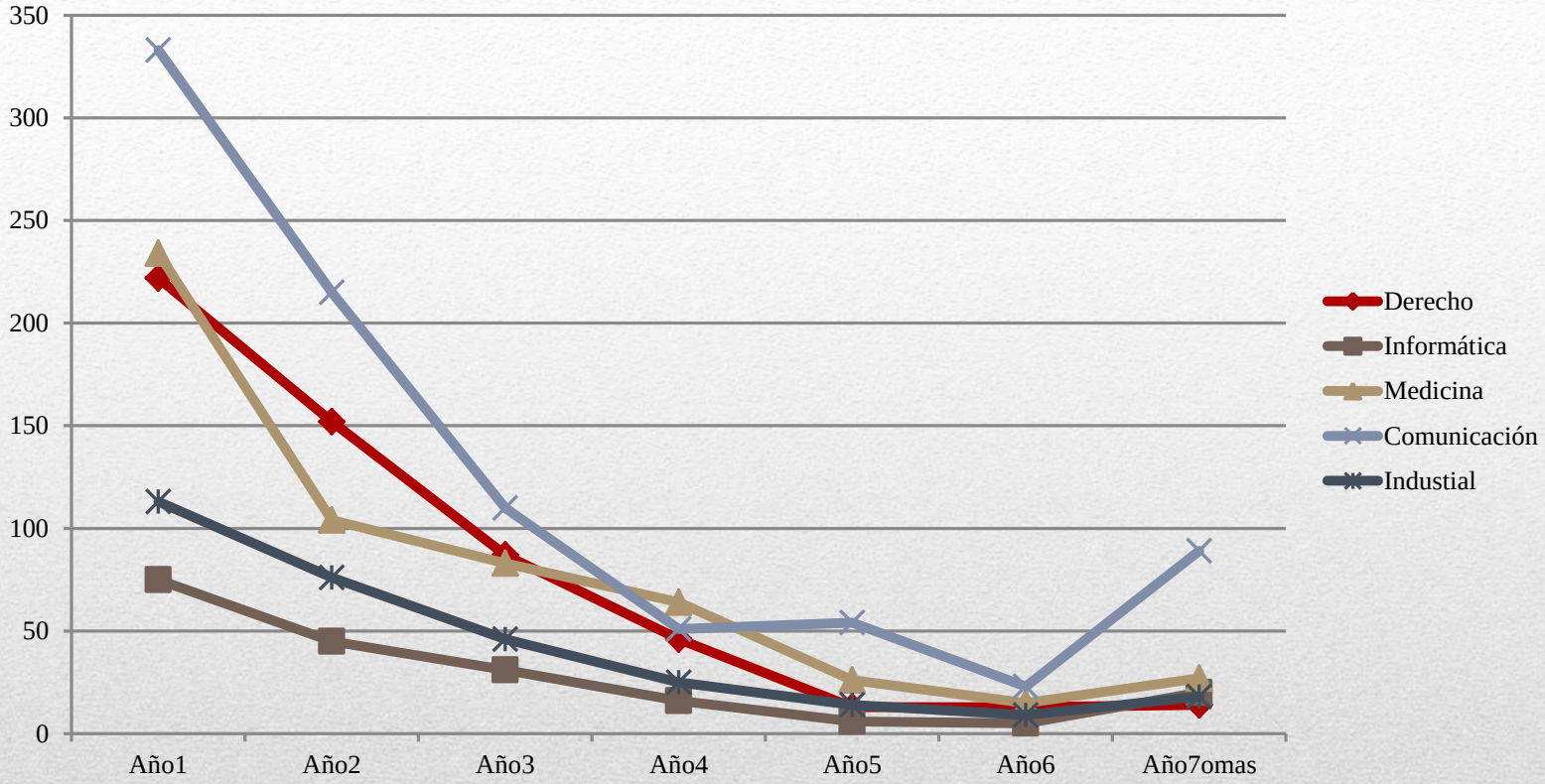
- Dificultades con información demográfica:
- Gran cantidad de datos ausentes
  - Ej. Nivel de estudios del padre/madre
- Información no tipificada
  - Ej. título del colegio

UNIVERSIDAD AUSTRAL



Variable Objetivo

UNIVERSIDAD AUSTRAL



Análisis Univariado: Bajas por antigüedad

UNIVERSIDAD AUSTRAL

Ventana de tiempo

- 1 o 2 años de vida académica
  - No se consideran bajas dentro de la ventana de tiempo
  - Anticipación vs Precisión

- ¿Considerando fecha de baja o no?
  - Bajas próximas
    - Bajas tardías tienen causas diversas de bajas tempranas
  - Bajas en general
    - Bajas tardías se manifiestan en el comienzo
- Alumnos pasivos
- Graduados ≠ Desertores

Identificación de la baja

UNIVERSIDAD AUSTRAL

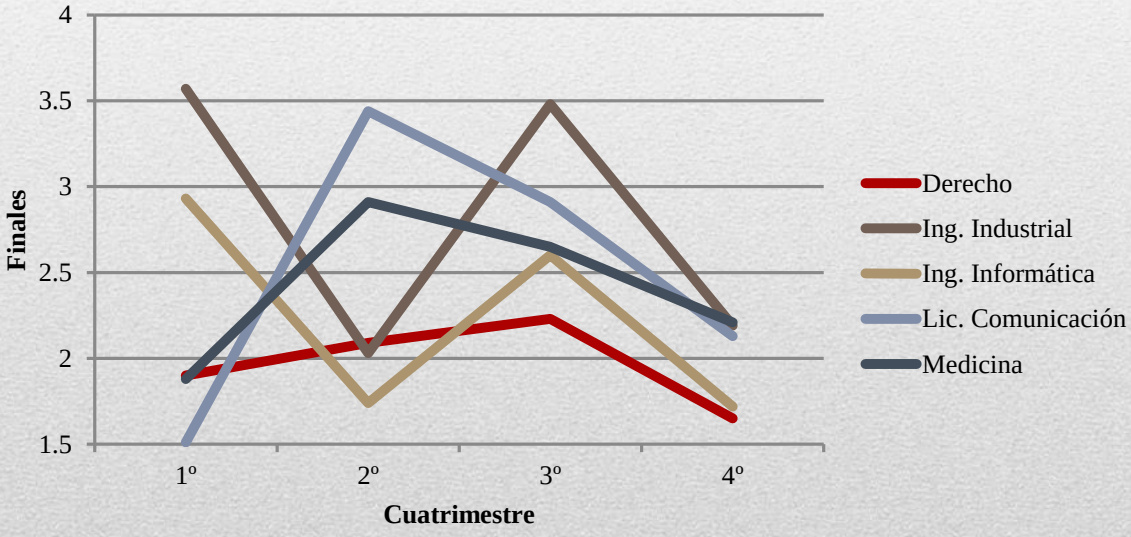
- Fecha de baja
- Fecha de alta
- Edad al ingresar
- Años entre secundario e ingreso
- País de origen
- Tiempo que duró en la carrera
- Información de familiares

Creación, modificación y corrección de variables

UNIVERSIDAD AUSTRAL

Modelo único vs modelos por carrera:

- Planes de estudios
- Comportamiento académico
- Tamaño de la muestra vs Especificidad



Carrera	Casos
Derecho	1.368
Industrial	808
Informática	392
Comunicación	2.519
Medicina	1.566
Total	**6.653**

Información académica

UNIVERSIDAD AUSTRAL

- Información disponible: cursadas y finales por alumno
- Problema de cambios de Planes de Estudios
  - Materias que cambian de ubicación
  - Materias que desaparecen
  - Nuevas materias
  - Cambios en duración de la carrera

Información académica: Nivel de detalle

UNIVERSIDAD AUSTRAL

- **Modelos sumarizados:** *Totales por cuatrimestre*
  - Cursadas aprobadas
  - Cursadas desaprobadas
  - Finales aprobados
  - Finales desaprobados

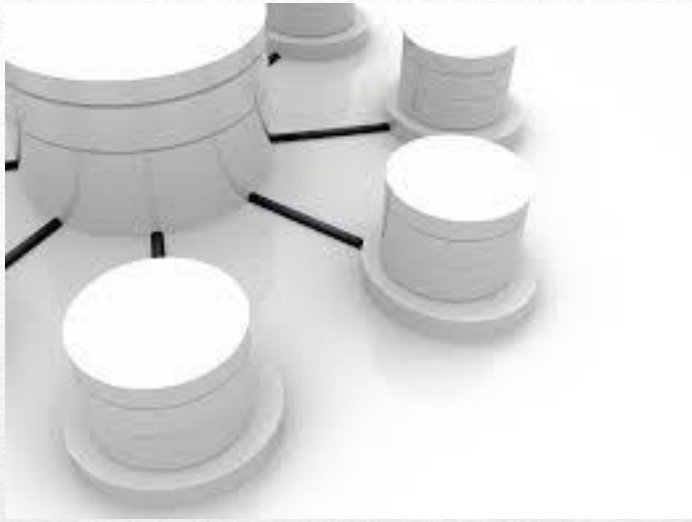
Información académica: Nivel de detalle

UNIVERSIDAD AUSTRAL

- **Modelos detallados:** *Sólo Medicina*
  - 3 planes de estudios en 15 años
  - Mínimas modificaciones en los primeros 2 años
  - 2ª carrera con mayor cantidad de casos

Información académica: Nivel de detalle

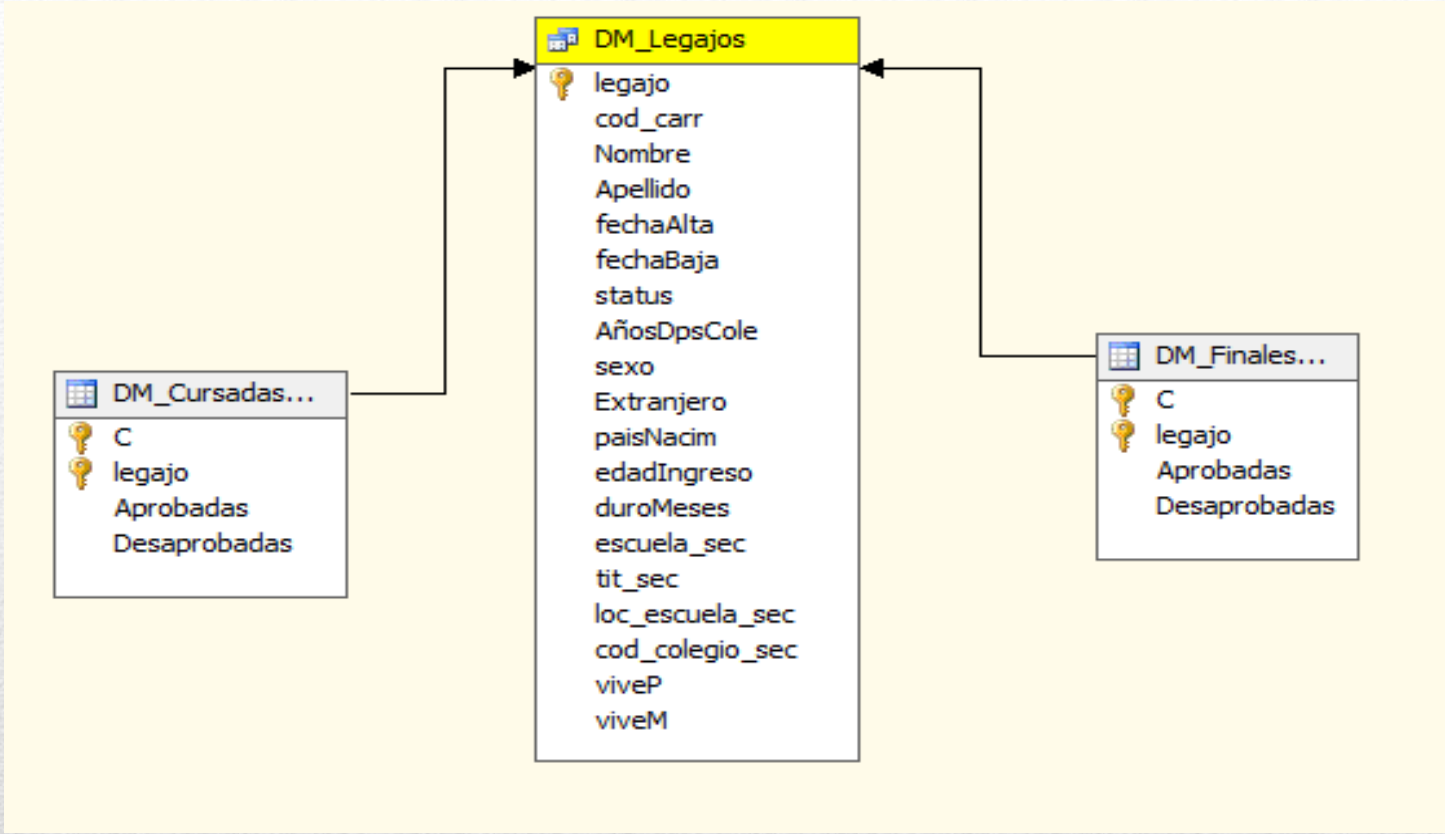
UNIVERSIDAD AUSTRAL



- Enfoque tradicional: una fila por caso
- Enfoque multi-relacional:
  - Case Table Una fila por caso
  - Nested Tables Una fila por variable

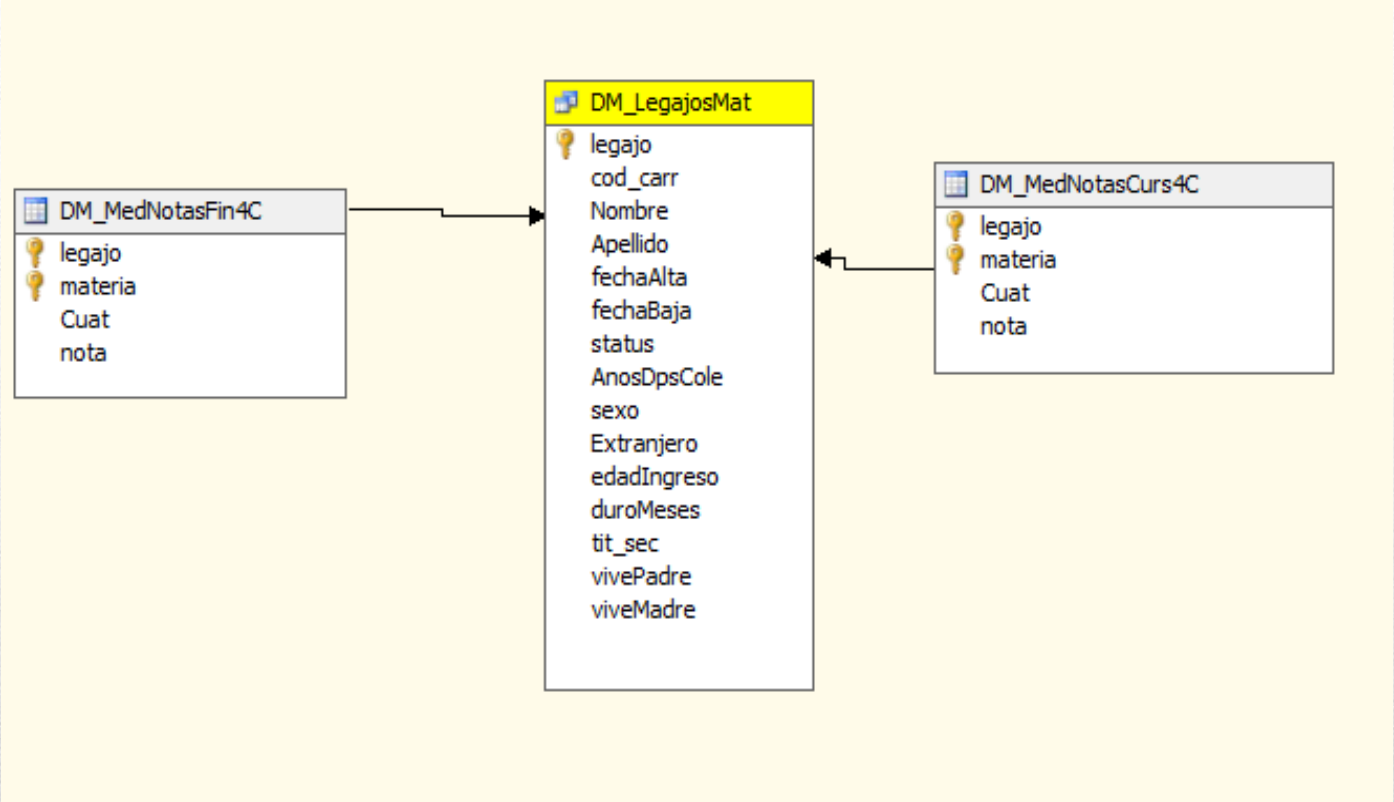
Data Mining Multi-relacional

UNIVERSIDAD AUSTRAL



Estructura de Datos: Modelo sumarizado

UNIVERSIDAD AUSTRAL



Estructura de Datos: Modelo detallado

UNIVERSIDAD AUSTRAL

Sistema de claves

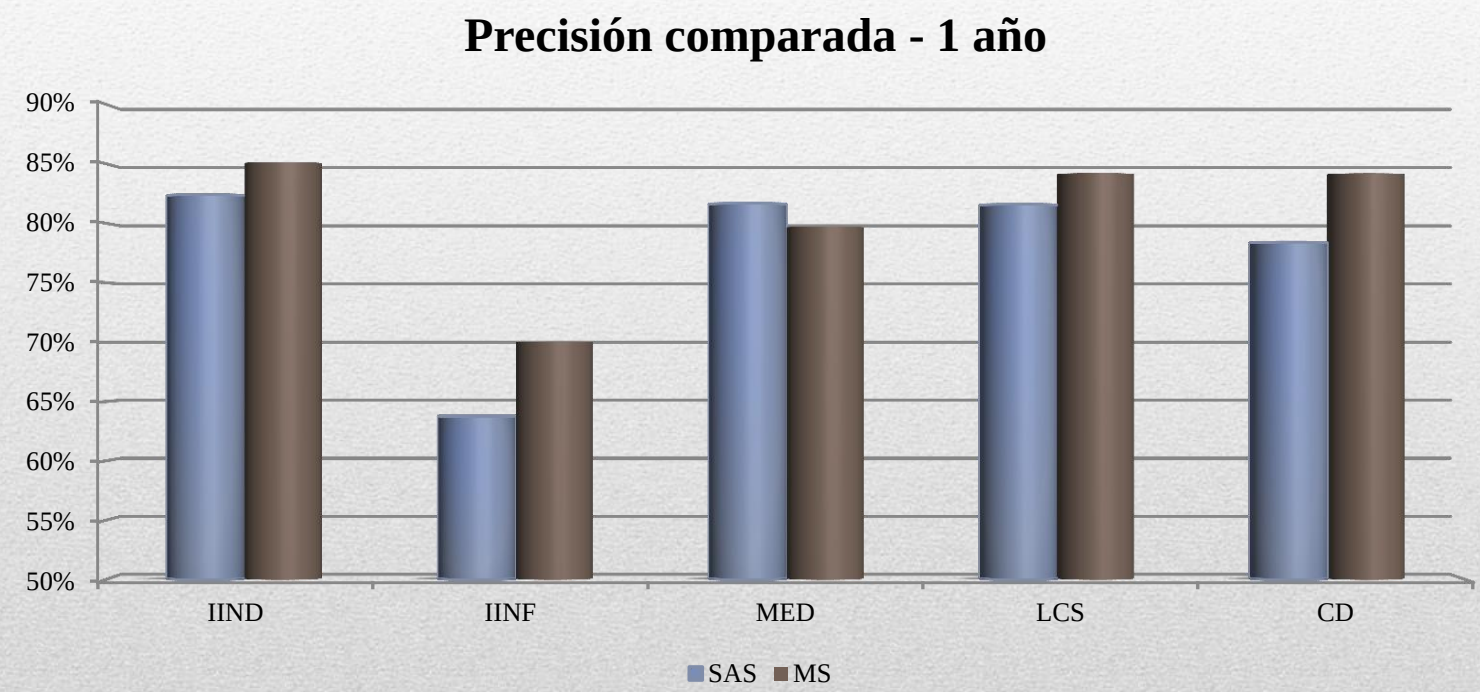
- Analysis Services sólo permite claves dobles
  - ¿Legajo + Materia?
  - Cada materia se puede cursar y rendir varias veces
  - Solución: “Materia” -> Materia & Instancia

Resultados

- $Precisión = \frac{Verdaderos\ Positivos + Verdaderos\ Negativos}{Casos\ totales}$

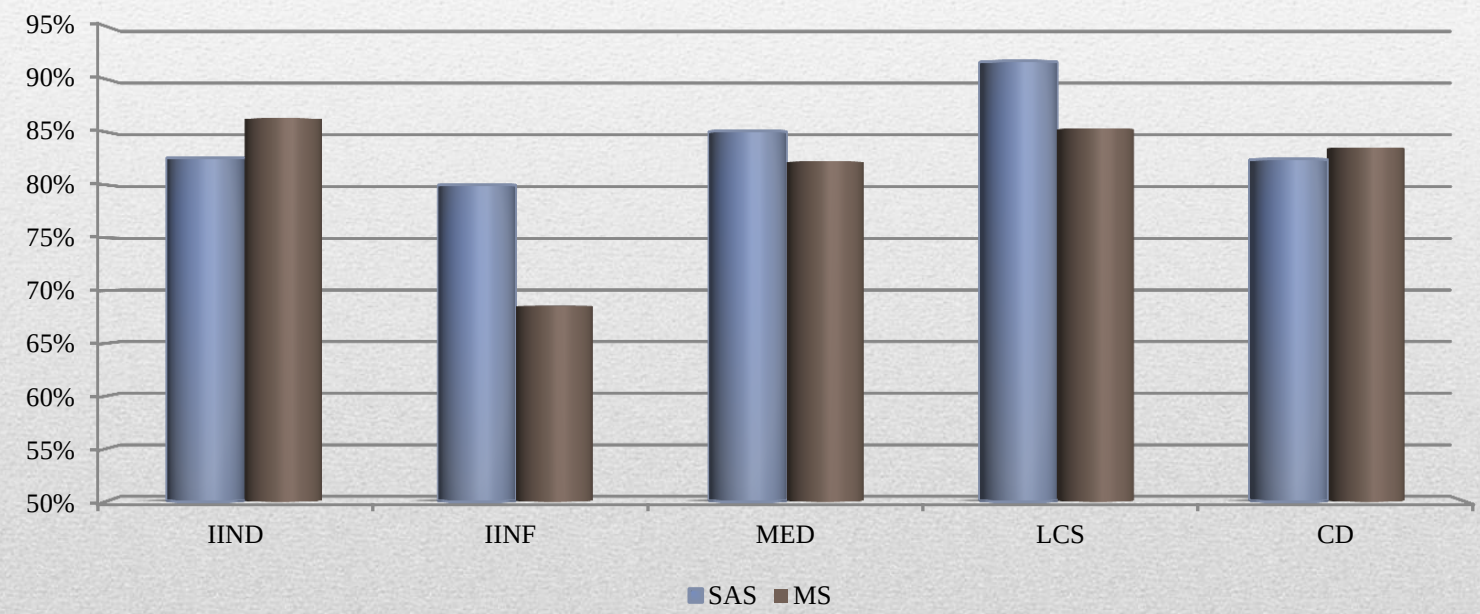
Precisión	**1 año**	**2 años**
Ing. Industrial	82%	77%
Ing. Informática	61%	74%
Medicina	79%	84%
Lic. en Comunicación	80%	88%
Derecho	79%	87%

Modelos sumarizados



Comparación de resultados

UNIVERSIDAD AUSTRAL



Comparación de resultados

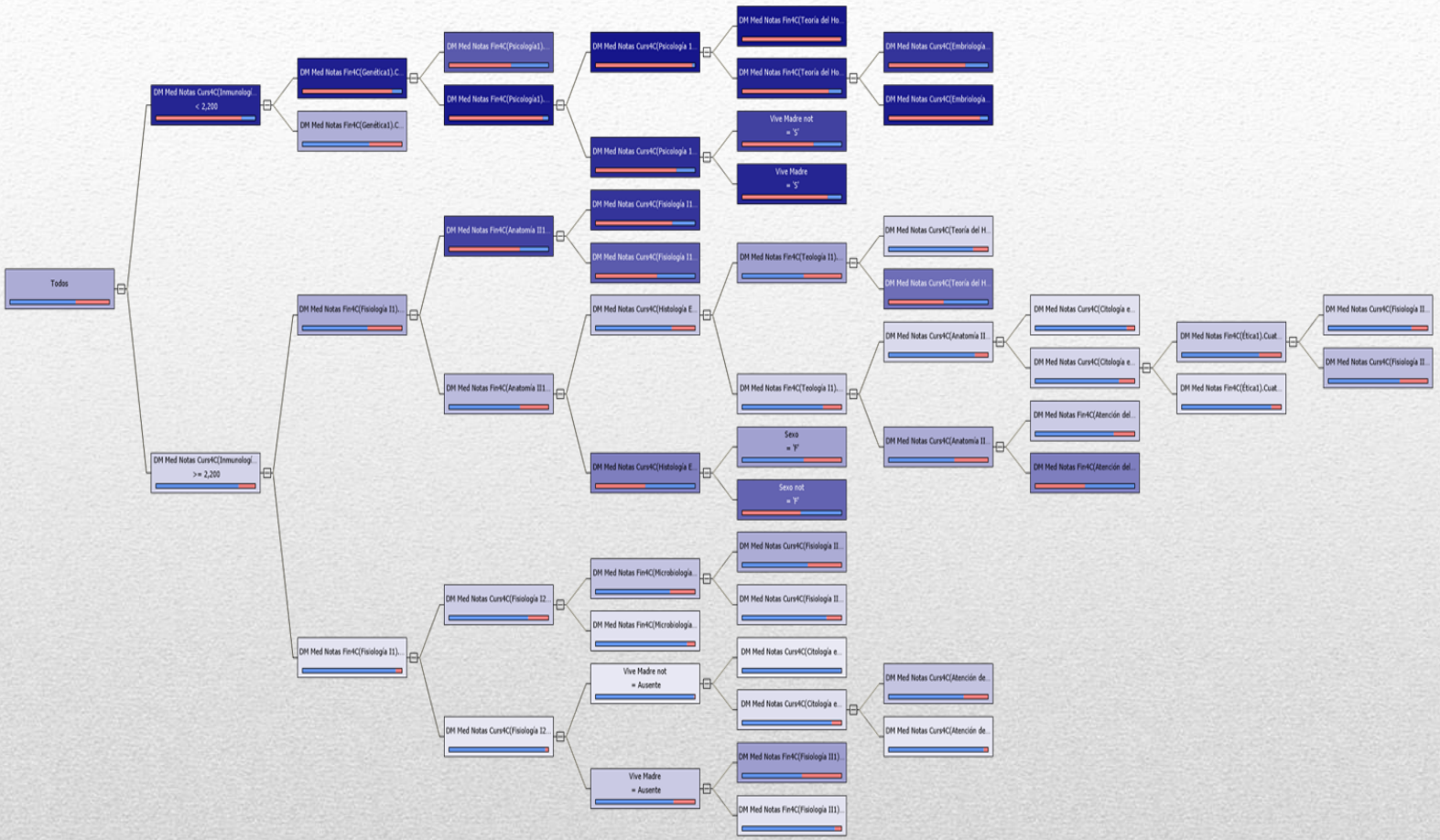
UNIVERSIDAD AUSTRAL

- Los Árboles no incluyen las **notas**, sino **dato ausente**
- Particularidad del dominio del problema: variables altamente correlacionadas

Precisión	**1 año**	**2 años**
Medicina Sumarizado	79%	84%
Medicina Detallado	79%	87%

Modelo detallado

UNIVERSIDAD AUSTRAL



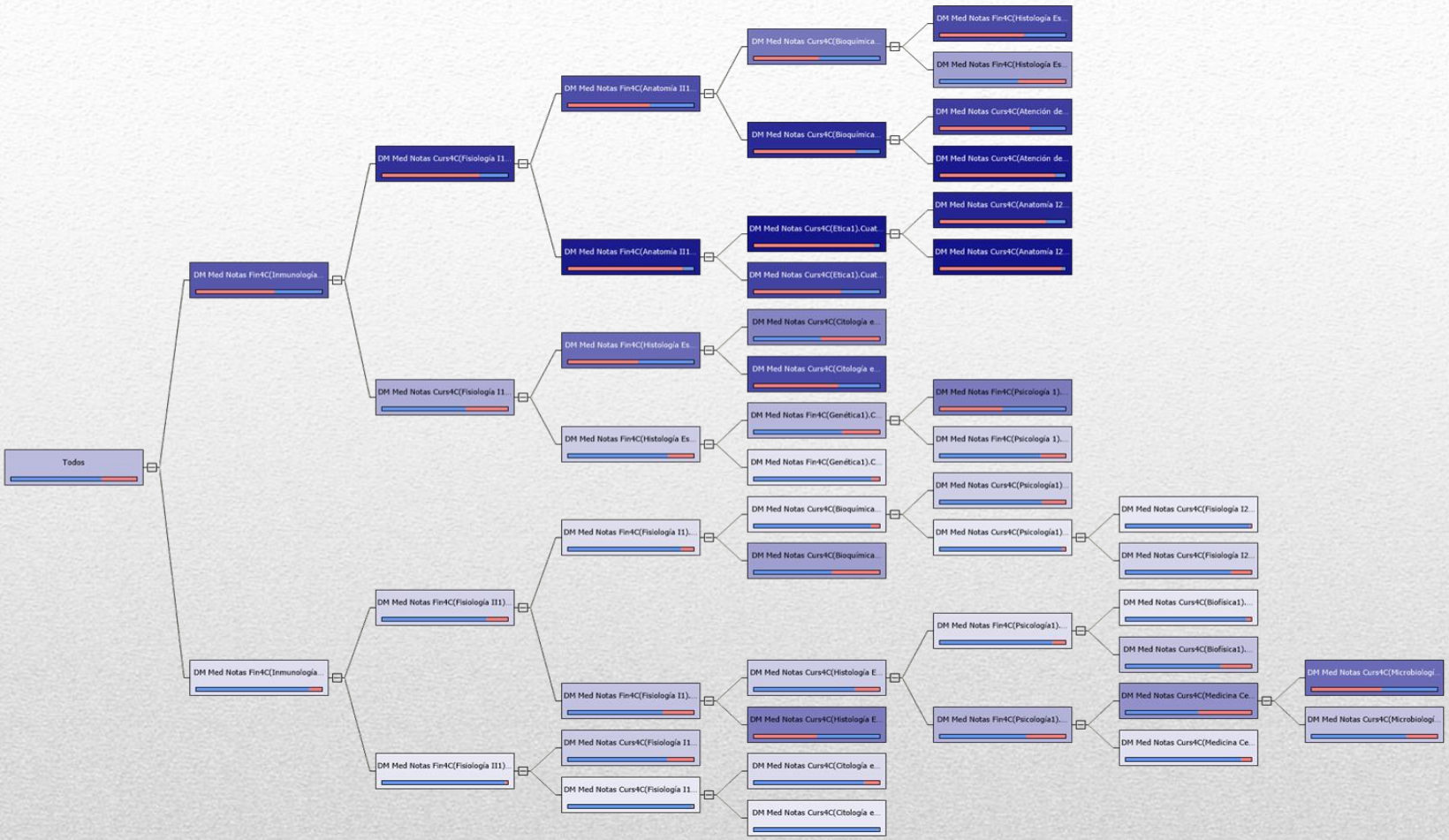
Modelo detallado 1 año

UNIVERSIDAD AUSTRAL

- No haber rendido el final de Anatomía II por primera vez (143 P, 54 N)
  - No haber rendido el final de Bioquímica II por primera vez (115 P, 17 N)
    - No haber rendido el final de Citología e Histología General por primera vez (82 P, 3 N)
      - No haber cursado Teología I por primera vez (57 P, 0 N)
      - Haber cursado Teología I por primera vez (25 P, 3 A)
        - Sexo Masculino (10 P, 3 N)
        - Sexo Femenino (15 P, 0 N)

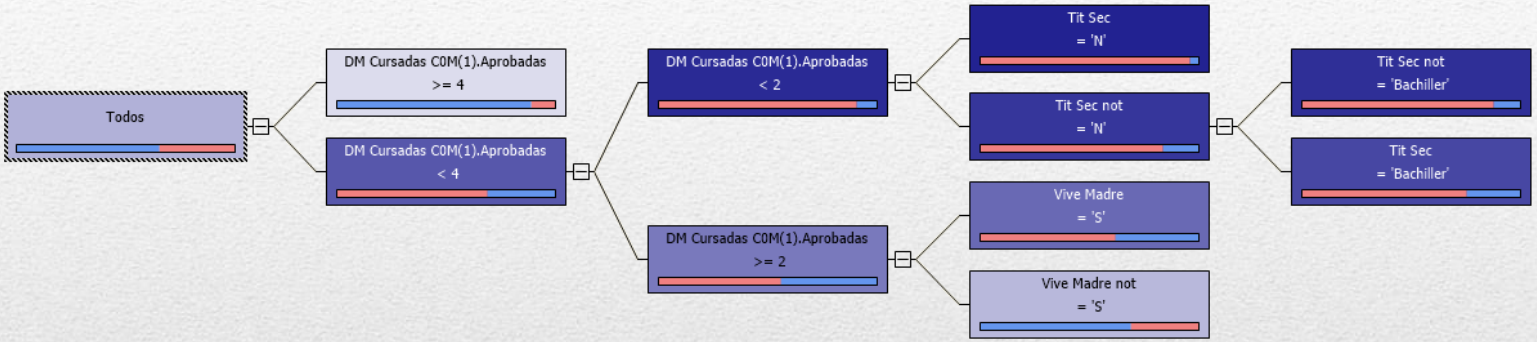
Ej. de Regla

UNIVERSIDAD AUSTRAL



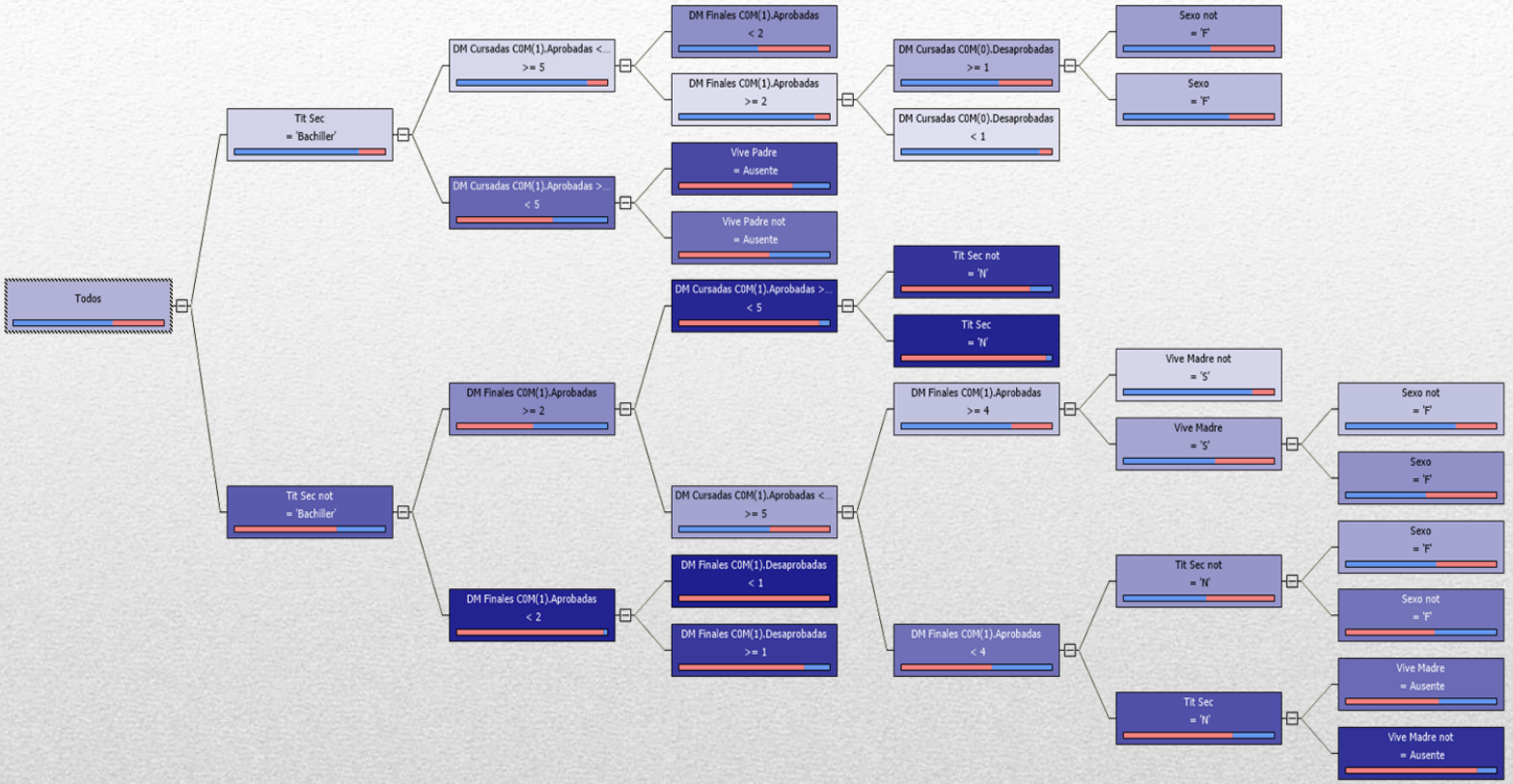
Modelo detallado 2 años

UNIVERSIDAD AUSTRAL



Árbol Ing. Industrial 1 año

UNIVERSIDAD AUSTRAL



Árbol Comunicación 1 año

UNIVERSIDAD AUSTRAL

- Es viable elaborar el modelo
- La ausencia de datos demográficos hace que estos casi no aparezcan
- Dentro de la información disponible, la académica es la más predictiva

Conclusiones: viabilidad del modelo

UNIVERSIDAD AUSTRAL

- El *modelo detallado* no es viable para todas las carreras
- Sólo son relevantes la presencia/ausencia de datos académicos (para cada materia-instancia de cursado y finales)
- La herramienta no permite un análisis multivariado de los datos incluidos en las tablas anidadas

Conclusiones: sobre los modelos detallados

UNIVERSIDAD AUSTRAL

- El uso de tablas anidadas no resultó tan útil como se esperaba
- La limitación de *claves dobles* puede salvarse, pero podría resolverlo la herramienta misma

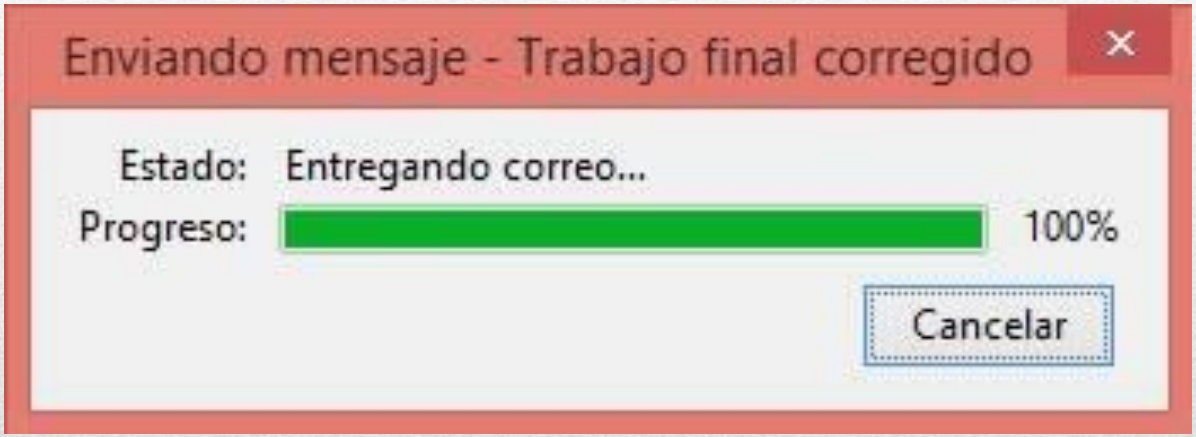
Conclusiones: sobre la herramienta

UNIVERSIDAD AUSTRAL

- Resultados útiles a estudios orientados a la calidad educativa y estrategias de retención.
- Otros estudios de interés académico: evaluar los cambios de plan de estudios, cambios de cátedra, cambios en el comportamiento académico de distintas generaciones, etc.
- Un mejor relevamiento de las variables demográficas.
- Incorporación del log del WebSIA / Moodle

Conclusiones: trabajos futuros

UNIVERSIDAD AUSTRAL



ggadea@austral.edu.ar

Gracias!

UNIVERSIDAD AUSTRAL