

# RECONOCIMIENTO DE PATRONES RELACIONADOS CON EL DESEMPEÑO ACADÉMICO DE LOS ESTUDIANTES DE LA ESCUELA DE INGENIERÍA INFORMÁTICA

Ricardo Bocaz L.

Pontificia Universidad Católica de Valparaíso

Profesor Guía: Rodrigo Alfaro Arancibia.

Profesor Correferente: Alexandru Cristian Rusu.



PONTIFICIA UNIVERSIDAD  
CATOLICA  
DE VALPARAISO



# Contenido de la presentación

- 1 Objetivos
  - Objetivo General
  - Objetivos Específicos
- 2 Marco Teórico
  - Marco Conceptual
  - Marco Referencial
- 3 Planificación
  - Alcance
  - Programa
- 4 Datos de entrada
- 5 Técnica a utilizar
- 6 Conclusiones

# Introducción

- Continua búsqueda de mejorar los procesos formativos y la calidad de la educación.
- Adelantarse a los hechos.
- Gran cantidad de datos históricos.
- Aporte de la informática, apoyo a la toma de decisiones.
- Se pretende identificar características de los estudiantes relacionadas con su desempeño (Procedencia, Nivel Socioeconómico, etc.).

# Contenido de la Presentación

- 1 **Objetivos**
  - Objetivo General
  - Objetivos Específicos
- 2 Marco Teórico
  - Marco Conceptual
  - Marco Referencial
- 3 Planificación
  - Alcance
  - Programa
- 4 Datos de entrada
- 5 Técnica a utilizar
- 6 Conclusiones

# Contenido de la Presentación

- 1 **Objetivos**
  - **Objetivo General**
  - Objetivos Específicos
- 2 Marco Teórico
  - Marco Conceptual
  - Marco Referencial
- 3 Planificación
  - Alcance
  - Programa
- 4 Datos de entrada
- 5 Técnica a utilizar
- 6 Conclusiones

# Objetivo General del Proyecto

Realizar un reconocimiento de patrones, que permitan el pronóstico del desempeño universitario de los estudiantes de la Escuela de Ingeniería Informática de la Pontificia Universidad Católica de Valparaíso, mediante el uso del Análisis Inteligente de Datos.

# Contenido de la Presentación

- 1 **Objetivos**
  - Objetivo General
  - **Objetivos Específicos**
- 2 Marco Teórico
  - Marco Conceptual
  - Marco Referencial
- 3 Planificación
  - Alcance
  - Programa
- 4 Datos de entrada
- 5 Técnica a utilizar
- 6 Conclusiones

# Objetivos Específicos

- Definir indicadores que permitan la medición del desempeño y rendimiento de los estudiantes.
- Aplicar un Análisis Inteligente de Datos para el reconocimiento de los patrones relacionados con el desempeño estudiantil.
- Crear la base para la implementación de una herramienta que permita el pronóstico del desempeño universitario.



# Objetivos Específicos

- Definir indicadores que permitan la medición del desempeño y rendimiento de los estudiantes.
- Aplicar un Análisis Inteligente de Datos para el reconocimiento de los patrones relacionados con el desempeño estudiantil.
- Crear la base para la implementación de una herramienta que permita el pronóstico del desempeño universitario.

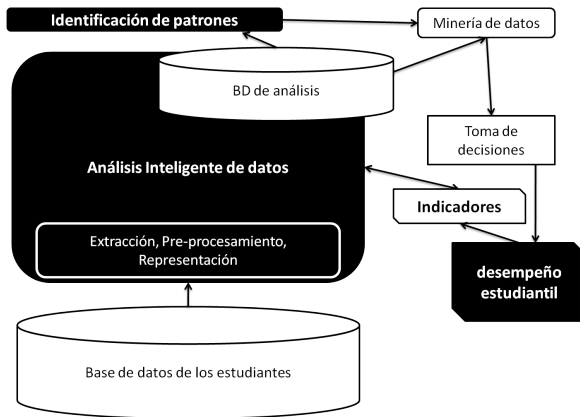
# Objetivos Específicos

- Definir indicadores que permitan la medición del desempeño y rendimiento de los estudiantes.
- Aplicar un Análisis Inteligente de Datos para el reconocimiento de los patrones relacionados con el desempeño estudiantil.
- Crear la base para la implementación de una herramienta que permita el pronóstico del desempeño universitario.

# Contenido de la Presentación

- 1 Objetivos
  - Objetivo General
  - Objetivos Específicos
- 2 **Marco Teórico**
  - Marco Conceptual
  - Marco Referencial
- 3 Planificación
  - Alcance
  - Programa
- 4 Datos de entrada
- 5 Técnica a utilizar
- 6 Conclusiones

# Marco Teórico



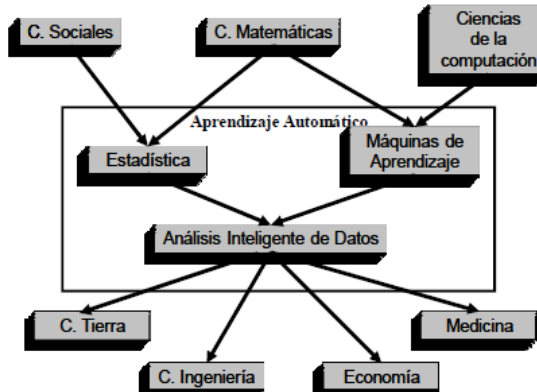
# Contenido de la Presentación

- 1 Objetivos
  - Objetivo General
  - Objetivos Específicos
- 2 Marco Teórico
  - Marco Conceptual
  - Marco Referencial
- 3 Planificación
  - Alcance
  - Programa
- 4 Datos de entrada
- 5 Técnica a utilizar
- 6 Conclusiones

# Desempeño Académico

- Objetivos del estudiante.
- Calidad de la educación.
- Diferencia entre desempeño y rendimiento.
- Indicadores.

# Análisis Inteligente de Datos



# Reconocimiento de Patrones





# Reconocimiento de Patrones

«Dado algunos ejemplos de señales complejas y las decisiones correctas para ellas, realizar decisiones automáticamente para un flujo de ejemplos futuros»[4]

- Características.
- Clasificador
  - Clasificación.
  - Outliers.
  - Rechazos.
- Modelos y Patrones[2]
  - Un modelo es una descripción global, de alto nivel.
  - Un patrón es una característica local de los datos.

# Contenido de la Presentación

- 1 Objetivos
  - Objetivo General
  - Objetivos Específicos
- 2 **Marco Teórico**
  - Marco Conceptual
  - **Marco Referencial**
- 3 Planificación
  - Alcance
  - Programa
- 4 Datos de entrada
- 5 Técnica a utilizar
- 6 Conclusiones

# Determinantes de desempeño universitario. ¿Importa la habilidad relativa?

- Estudio del Consejo Superior de Educación en el 2009[1].
- Revisar si una medida de habilidad relativa aporta información, aparte de la batería de selección tradicional.
- Cuatro universidades estudiadas (PUCV, UCT, UFT,USACH).
- Ranking de egreso del alumno.
- Análisis de estudiantes por ingreso especial.
- Resultados
  - Estar entre los mejores alumnos de la escuela implica mejor desempeño en primer año.
  - Ingresos especiales tienen rendimiento igual o superior.
  - Batería de selección tradicional no es el único predictor del rendimiento universitario.

# Estudio de validez predictiva de la PSU y comparación con el sistema PAA

- Tesis de magíster de S. Prado en 2008 (UTAL) [3].
- Análisis de la validez predictiva de la PSU.
- Alcance: Ingeniería Civil 2001-2006 - PUC, UCH .
- Comparación con la PAA.
- Análisis por prueba (Lenguaje, Matemática, etc.).
- Resultados
  - Validez predictiva menor que la reportada en un estudio del CRUCH para la PUC.
  - PSU tiene 2 % más de reprobaciones totales en primer año comparados con la PAA.

# Contenido de la Presentación

- 1 Objetivos
  - Objetivo General
  - Objetivos Específicos
- 2 Marco Teórico
  - Marco Conceptual
  - Marco Referencial
- 3 **Planificación**
  - Alcance
  - Programa
- 4 Datos de entrada
- 5 Técnica a utilizar
- 6 Conclusiones

# Contenido de la Presentación

- 1 Objetivos
  - Objetivo General
  - Objetivos Específicos
- 2 Marco Teórico
  - Marco Conceptual
  - Marco Referencial
- 3 **Planificación**
  - **Alcance**
  - Programa
- 4 Datos de entrada
- 5 Técnica a utilizar
- 6 Conclusiones

# Alcance del Proyecto

- Estudiantes de la Escuela de Ingeniería Informática de la Pontificia Universidad Católica de Valparaíso.
- Se deben realizar dos estudios por separado
  - Dos carreras con diferentes datos.

# Contenido de la Presentación

- 1 Objetivos
  - Objetivo General
  - Objetivos Específicos
- 2 Marco Teórico
  - Marco Conceptual
  - Marco Referencial
- 3 **Planificación**
  - Alcance
  - **Programa**
- 4 Datos de entrada
- 5 Técnica a utilizar
- 6 Conclusiones





# Programación

		Nombre de tarea	Duración	Comienzo	Fin	Pre
1		Definir los objetivos	11 días	mar 15/03/11	mar 29/03/11	
2		<input type="checkbox"/> Estudiar el marco conceptual	9 días?	mié 30/03/11	lun 11/04/11	1
3		Investigar sobre el reconocimiento di	5 días	mié 30/03/11	mar 05/04/11	
4		Investigar sobre el analisis de datos	4 días?	mié 06/04/11	lun 11/04/11	
5		Investigar sobre el desempeño acad	4 días?	lun 04/04/11	jue 07/04/11	
6		Revisar informe de avance	7 días?	mar 12/04/11	mié 20/04/11	2
7		<b>Entregar informe de avance</b>	<b>1 día</b>	<b>jue 21/04/11</b>	<b>jue 21/04/11</b>	<b>6</b>
8		<input type="checkbox"/> Recolectar los datos	50 días	lun 04/04/11	vie 10/06/11	1
9		Definición de los datos a recolectar	20 días	lun 04/04/11	vie 29/04/11	
10		Solicitud de los datos	15 días	lun 02/05/11	vie 20/05/11	9
11		Extraer, limpiar, cargar	15 días	lun 23/05/11	vie 10/06/11	10
12		Definir los indicadores de desempeño	10 días	lun 02/05/11	vie 13/05/11	
13		Definir los modelos de DM	6 días	lun 13/06/11	lun 20/06/11	
14		Revisar informe final	7 días?	mié 22/06/11	jue 30/06/11	
15		<b>Entregar informe final</b>	<b>1 día</b>	<b>vie 01/07/11</b>	<b>vie 01/07/11</b>	<b>14</b>

# Contenido de la Presentación

- 1 Objetivos
  - Objetivo General
  - Objetivos Específicos
- 2 Marco Teórico
  - Marco Conceptual
  - Marco Referencial
- 3 Planificación
  - Alcance
  - Programa
- 4 Datos de entrada**
- 5 Técnica a utilizar
- 6 Conclusiones

# Datos de Entrada

## Dimensión Procedencia

Establecimiento de egreso.

Tipo de establecimiento.

Ranking de egreso.

Región del establecimiento.

Año de término de la educación media.

Modalidad estudios del estudiante (Científico-Humanista diurno, de adultos y Técnico-Profesional).

Notas de la enseñanza media (NEM).

# Datos de Entrada

## Dimensión Batería de Selección

Puntaje de la PSU/PAA.

Año de cada Prueba rendida.

Puntaje de las pruebas en específico (Lenguaje, Matemática, Ciencias, Historia).

# Datos de Entrada

## Dimensión Universitaria

Año de ingreso.

Asignaturas cursadas.

Calificaciones de las asignaturas cursadas.

Aprobación o reprobación de las asignaturas cursadas.

Año de egreso.

Año de titulación.

# Contenido de la Presentación

- 1 Objetivos
  - Objetivo General
  - Objetivos Específicos
- 2 Marco Teórico
  - Marco Conceptual
  - Marco Referencial
- 3 Planificación
  - Alcance
  - Programa
- 4 Datos de entrada
- 5 **Técnica a utilizar**
- 6 Conclusiones

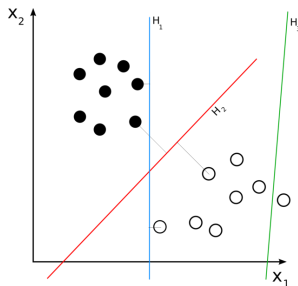
# Técnica

## Máquinas de Soporte Vectorial (SVM)

Serie de métodos de aprendizaje supervisado, usados para clasificación y reconocimiento de patrones.

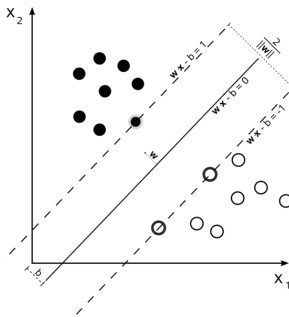
Toma un conjunto de datos de entrada y, para cada entrada, pronostica a cual de las dos diferentes clases pertenece.

Separa los datos mediante hiperplanos.



# Técnica

## Formalización



Ec. Hiperplano:  $\mathbf{w} \cdot \mathbf{x} - b = 0$ .

Margen: Distancia entre hiperplanos paralelos:  $\frac{2}{\|\mathbf{w}\|}$ .

Se busca maximizar el margen. Minimizar  $\|\mathbf{w}\|$ .

La salida es la clase de la entrada dada (+ ó -).



# Técnica

## Ventajas y Desventajas

### Ventajas

- Entrenamiento relativamente fácil.
- Si no es linealmente separable, el uso del Kernel no supone más costo computacional.

### Desventajas

- Alto costo computacional (QP)

# Técnica

## Ventajas y Desventajas

### Ventajas

- Entrenamiento relativamente fácil.
- Si no es linealmente separable, el uso del Kernel no supone más costo computacional.

### Desventajas

- Alto costo computacional (QP)

# Técnica

## Ventajas y Desventajas

### Ventajas

- Entrenamiento relativamente fácil.
- Si no es linealmente separable, el uso del Kernel no supone más costo computacional.

### Desventajas

- Alto costo computacional (QP)

# Técnica

## Ventajas y Desventajas

### Ventajas

- Entrenamiento relativamente fácil.
- Si no es linealmente separable, el uso del Kernel no supone más costo computacional.

### Desventajas

- Alto costo computacional (QP)

# Contenido de la Presentación

- 1 Objetivos
  - Objetivo General
  - Objetivos Específicos
- 2 Marco Teórico
  - Marco Conceptual
  - Marco Referencial
- 3 Planificación
  - Alcance
  - Programa
- 4 Datos de entrada
- 5 Técnica a utilizar
- 6 Conclusiones

# Conclusiones

- Estudio de los conceptos.
- Directrices del marco referencial.
- Resultados Esperados.

# Conclusiones

- Estudio de los conceptos.
- Directrices del marco referencial.
- Resultados Esperados.

# Conclusiones

- Estudio de los conceptos.
- Directrices del marco referencial.
- Resultados Esperados.



# Referencias



D. Contreras, S Gallegos, and F. Meneses.  
*Determinantes de Desempeño Universitario.*  
Consejo Superior de Educación, 2009.



D. Hand, H. Mannila, and P. Smyth.  
*Principles of Data Mining.*  
MIT Press, 2001.



Sebastián Prado.  
Estudio de validez predictiva de la psu y comparacion con el sistema  
paa.  
Master's thesis, Universidad de Chile, 2008.



B. Ripley.  
*Pattern Recognition and Neural Networks.*  
Cambridge University Press, 2000.

# Fin

## Muchas Gracias