



ESCUELA DE INGENIERÍA
FACULTAD DE INGENIERÍA

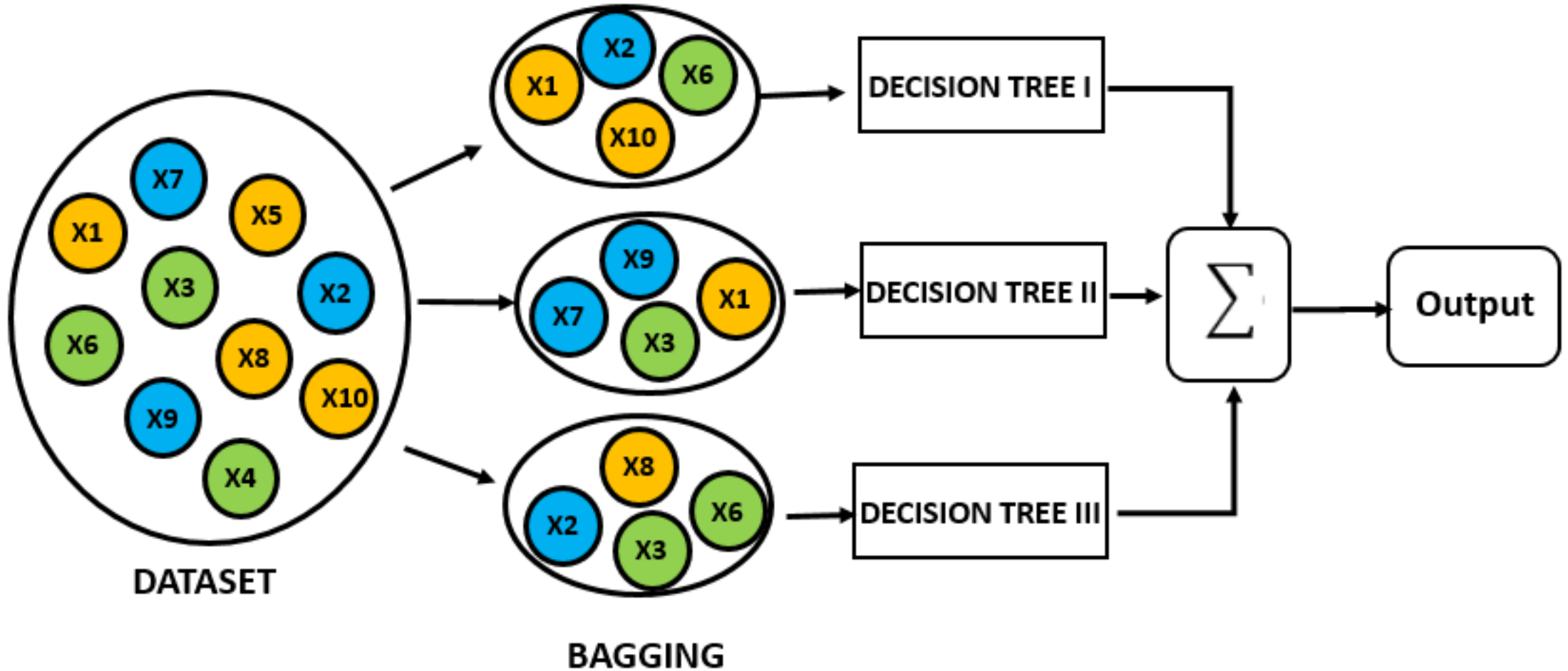
IIC2433 Minería de Datos

Random Forest

Profesor: Mauricio Arriagada

Minería de Datos

Random Forest



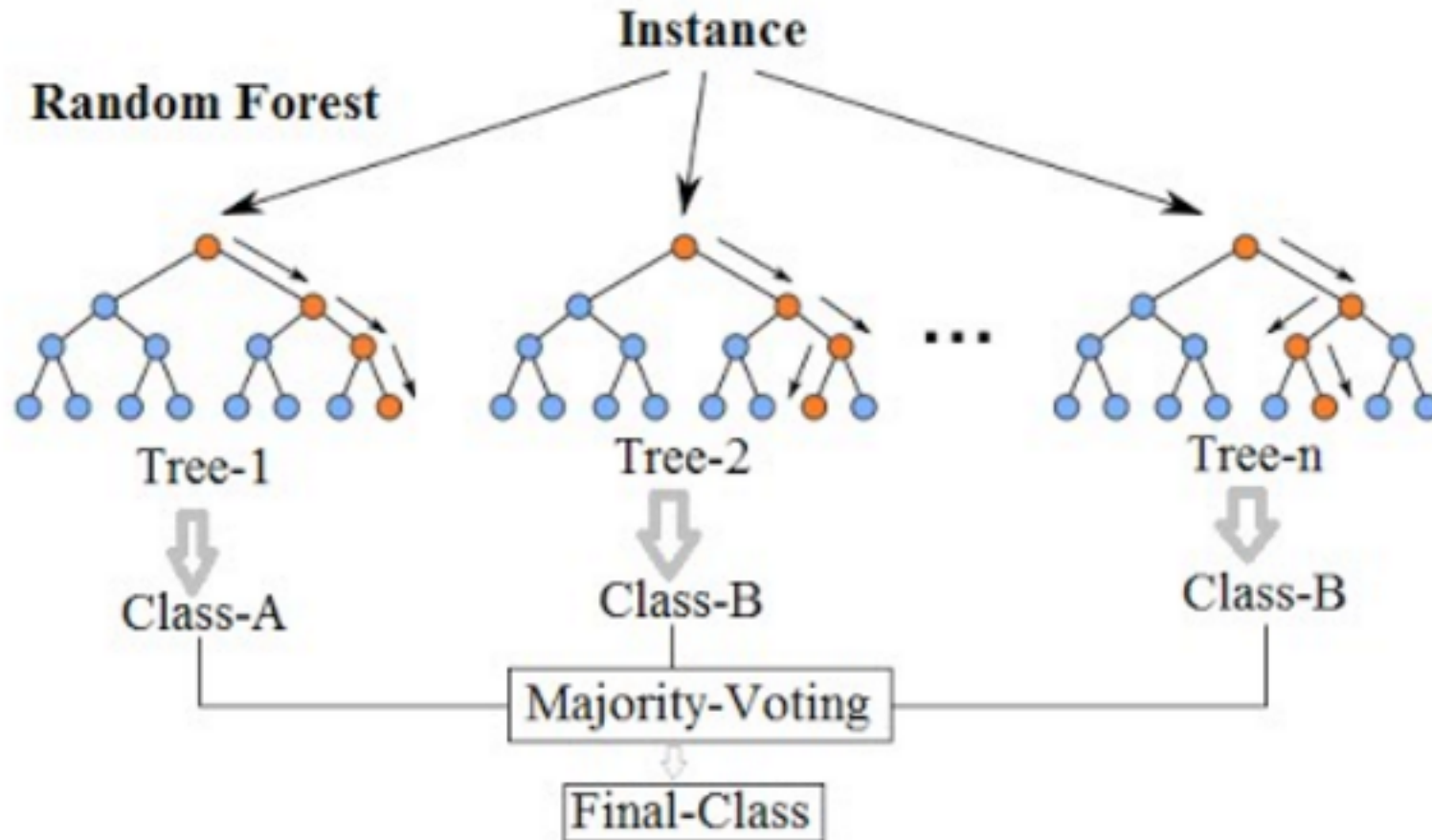
¿Qué es Random Forest?

Random Forest es un método versátil de aprendizaje automático capaz de realizar tanto tareas de regresión como de clasificación. También lleva a cabo métodos de reducción dimensional, trata valores perdidos, valores atípicos y otros pasos esenciales de exploración de datos

¿Cómo trabaja Random Forest?

- En Random Forest se ejecutan varios algoritmos de árbol de decisiones en lugar de uno solo. Para clasificar un nuevo objeto basado en atributos, cada árbol de decisión da una clasificación y finalmente la decisión con mayor “votos” es la predicción del algoritmo.

Random Forest Simplified



Ejemplos de uso

- ▶ Los bosques aleatorios se han implementado con éxito en diferentes áreas. Algunas aplicaciones incluyen:
 - ▶ Reconocimiento de imágenes
 - ▶ Biología Molecular (Análisis de secuencias de aminoácidos)
 - ▶ Remote sensing (reconocimiento de patrones)
 - ▶ Astronomía (clasificación de galaxias, etc.)

Ventajas

- ▶ Evita el sobreajuste
- ▶ Uno de los modelos de decisión más precisos
- ▶ Funciona bien en grandes conjuntos de datos
- ▶ Puede manejar miles de variables de entrada e identificar las variables más significativas

Desventajas

- ▶ Sobreajuste en caso de datos ruidosos.
- ▶ A diferencia de los árboles de decisión, los resultados son difíciles de interpretar.
- ▶ Tarda más tiempo en crear. Es computacionalmente costoso en comparación con un árbol de decisión.

REFERENCIAS

- ▶ Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.
- ▶ Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.
- ▶ Hand, D. J. (2006). Data Mining. *Encyclopedia of Environmetrics*, 2.