



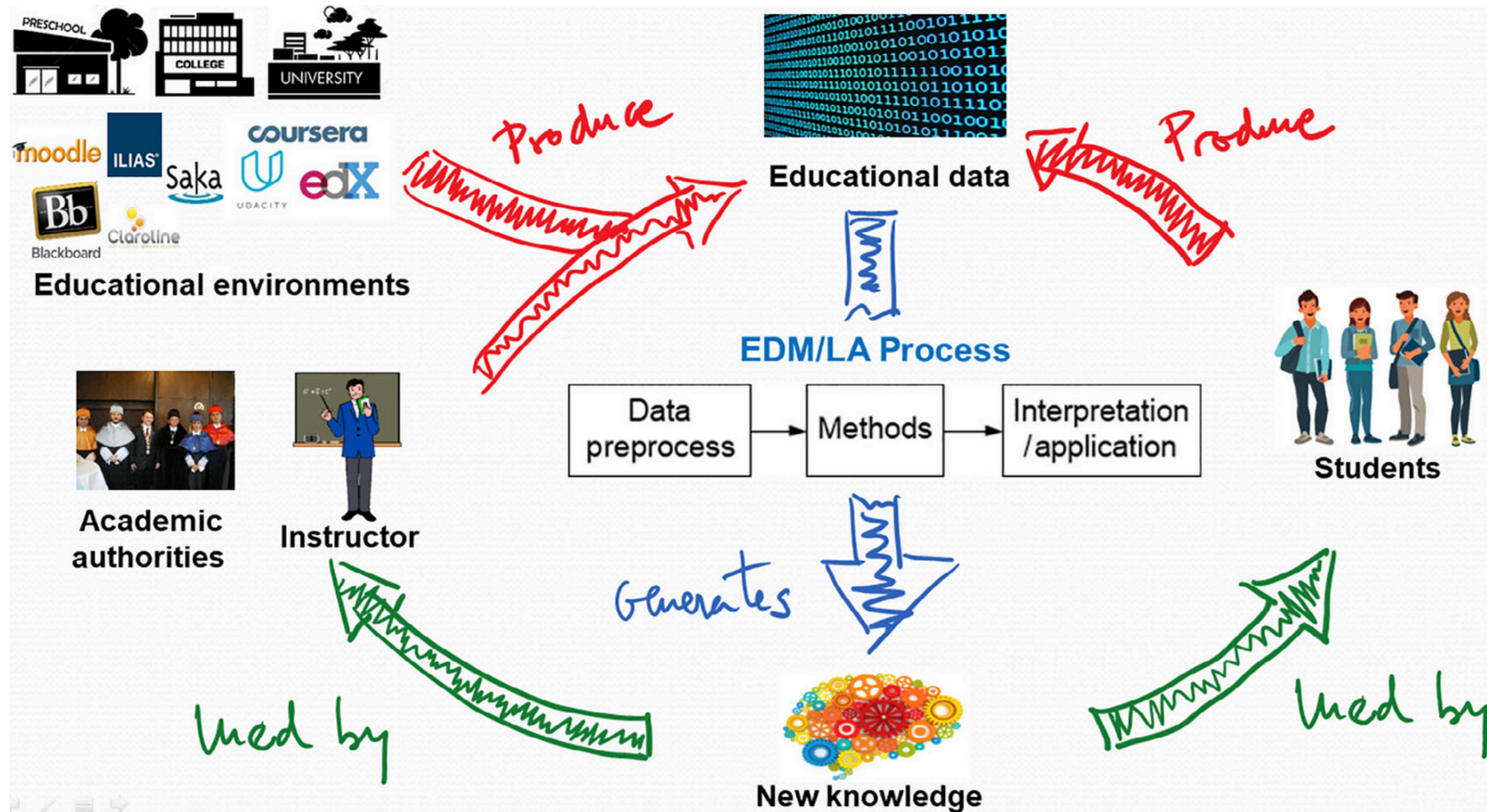
# **Educational Data Mining: Comparative Analysis of the Prediction of the Academic Performance of Entering University Students Using Decision Tree and Random Forest**

Jesús Aguilar-Ruiz, Edgar Taya-Acosta and Edgar Taya-Osorio

Presentado por:  
**Edgar Taya Acosta**



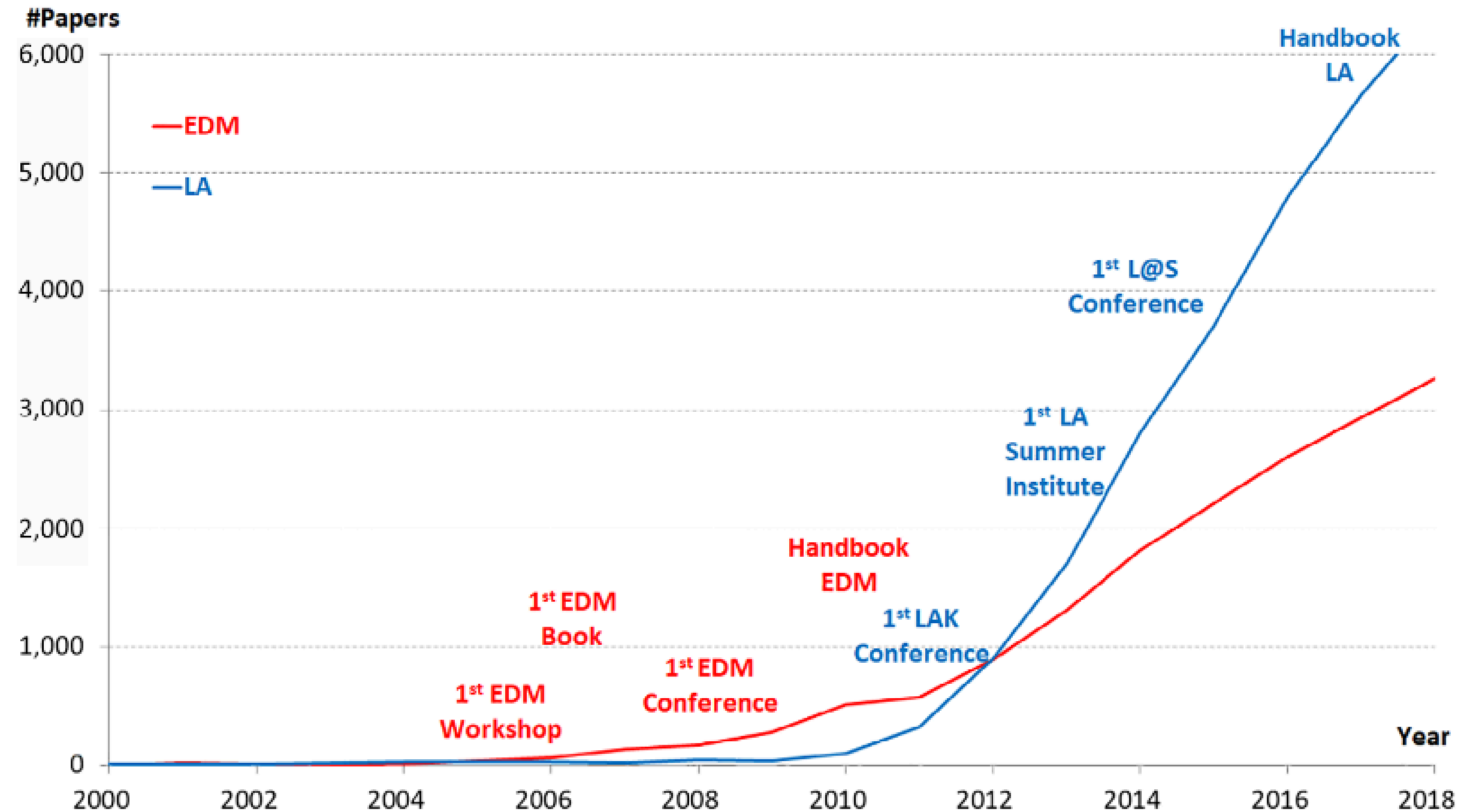
# Introducción



Source: (Romero et al. 2020)

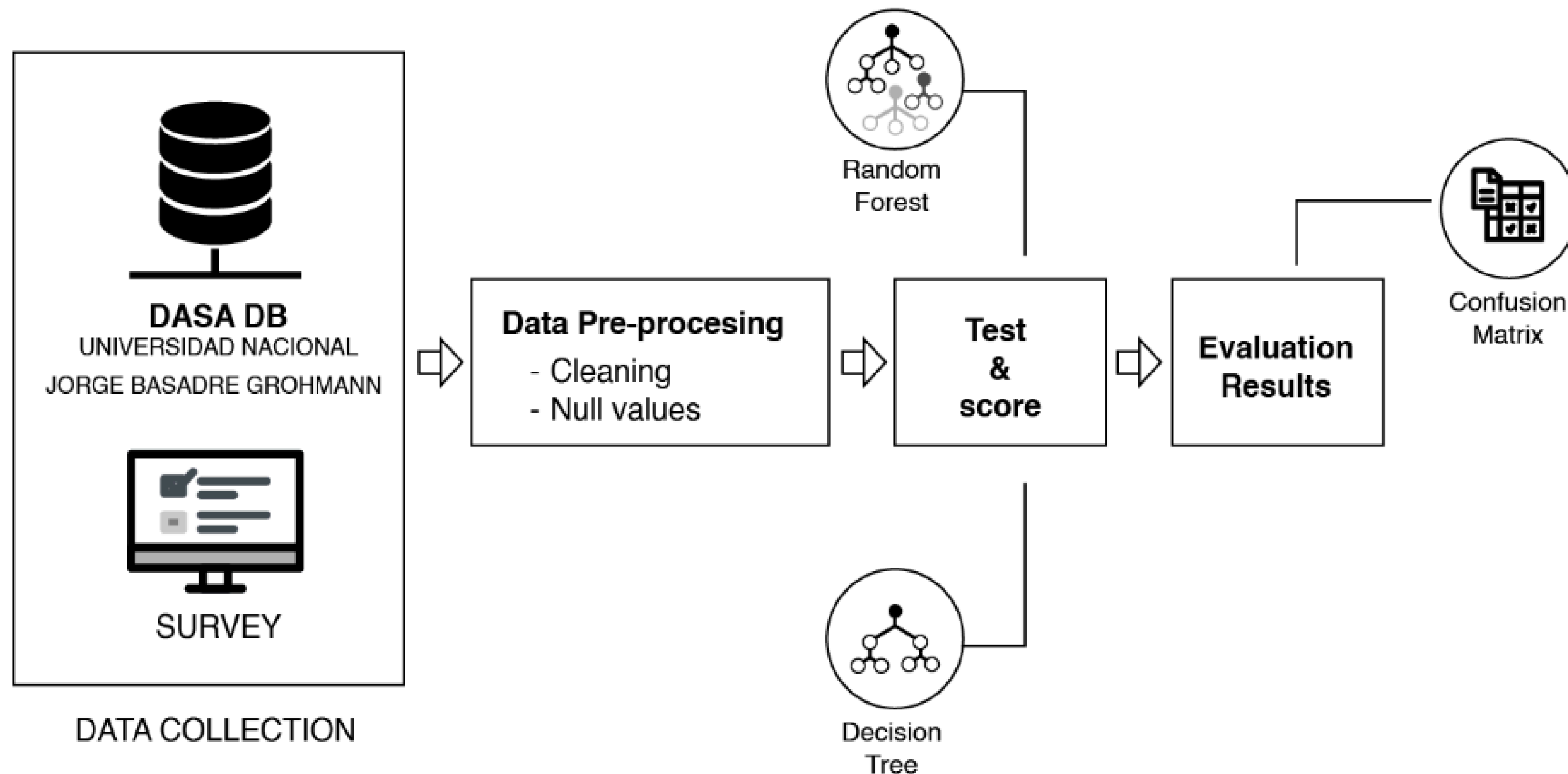
# Trabajos relacionados

**FIGURE 2** Number of papers and main events about Educational Data Mining/ Learning Analytics terms in Google Scholar by year (January 1, 2019)



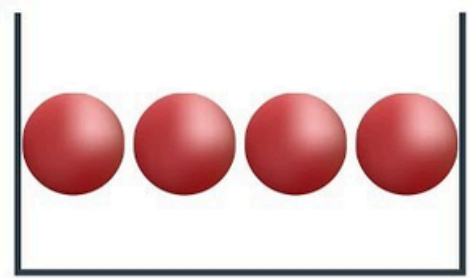
Source: (Romero et al. 2020)

# Metodología propuesta

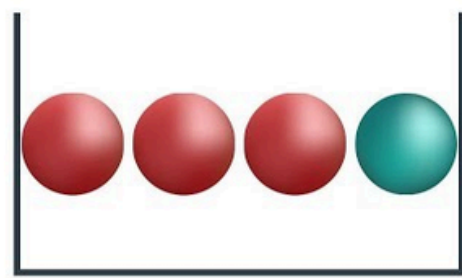


# Base teórica

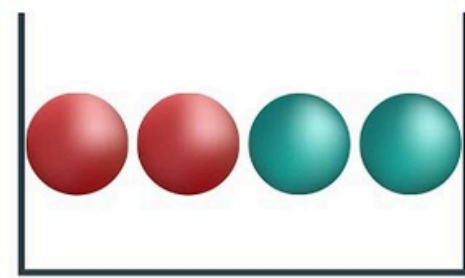
## Entropia



Low



Medium



High

¿Cómo se  
mide?

$$Entropy = - \sum_{i=1}^n p_i * \log(p_i)$$

$$Gini index = 1 - \sum_{i=1}^n p_i^2$$



# Detalles del experimento

## DATA

**305** Encuesta aplicada a todos los estudiantes que se matricularon en el periodo académico 2021-I.

**6,314** Registros de calificaciones de los estudiantes a los que se les aplicó la encuesta (Registro Académico)

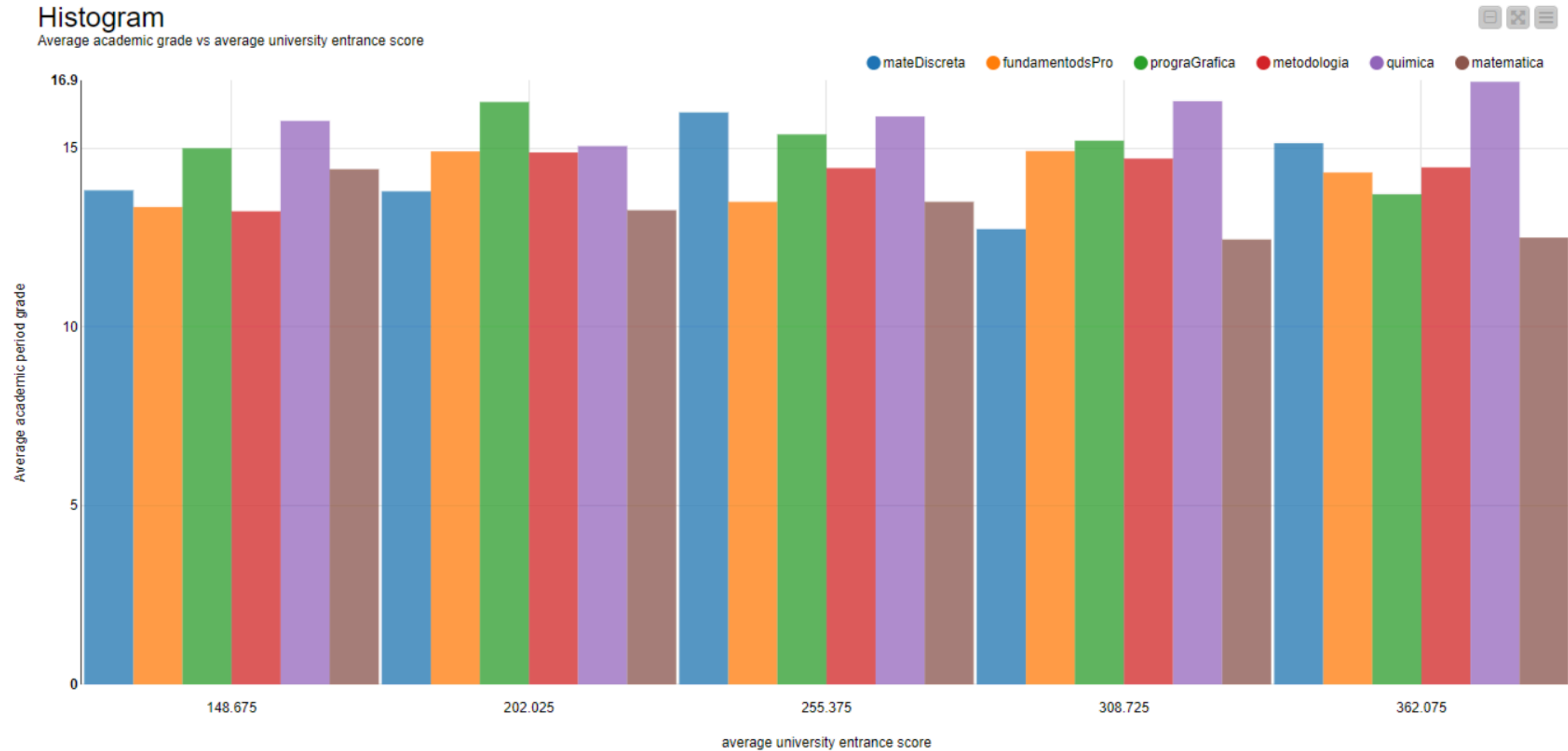
Nuestro conjunto de datos final estuvo compuesto por 77 columnas, algunas de ellas son:

Nro	Details
1	PROMEDIO_2021_1: Promedio de calificaciones del año académico 2021, período 1.
2	Frecuencia semanal de práctica de videojuegos.
3	Tipo de conexión a Internet.
4	Tipo de escuela donde terminaste la secundaria.
5	Género de lectura preferido.
6	Grado de acceso a la universidad.
7	Frecuencia semanal de visualización de vídeos de TV, Youtube, Netflix o Internet....
...	...

# Resultados y análisis

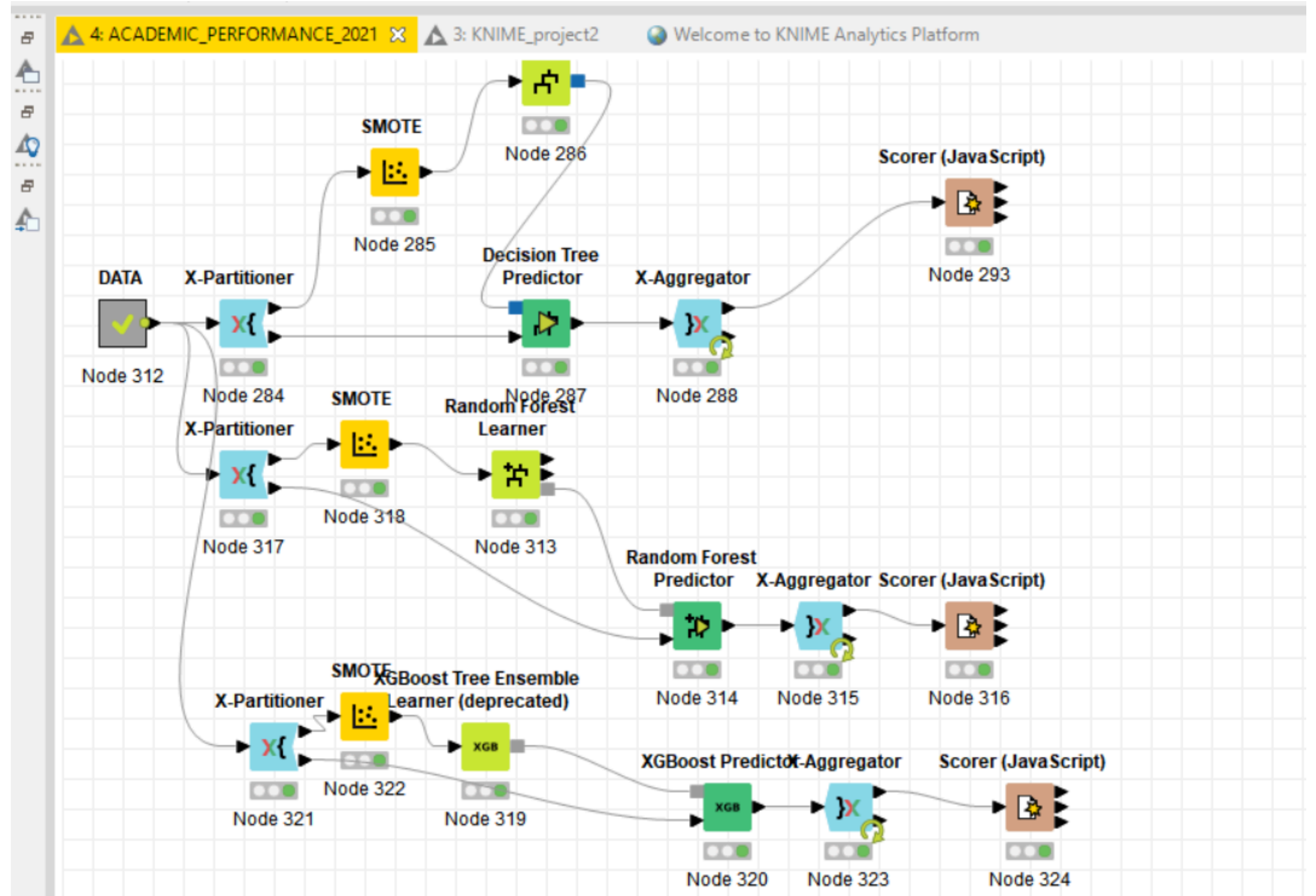
## Análisis

Promedio de nota académica (0 a 20) vs puntaje promedio de ingreso a la universidad (0 a 600).



# Resultados y análisis

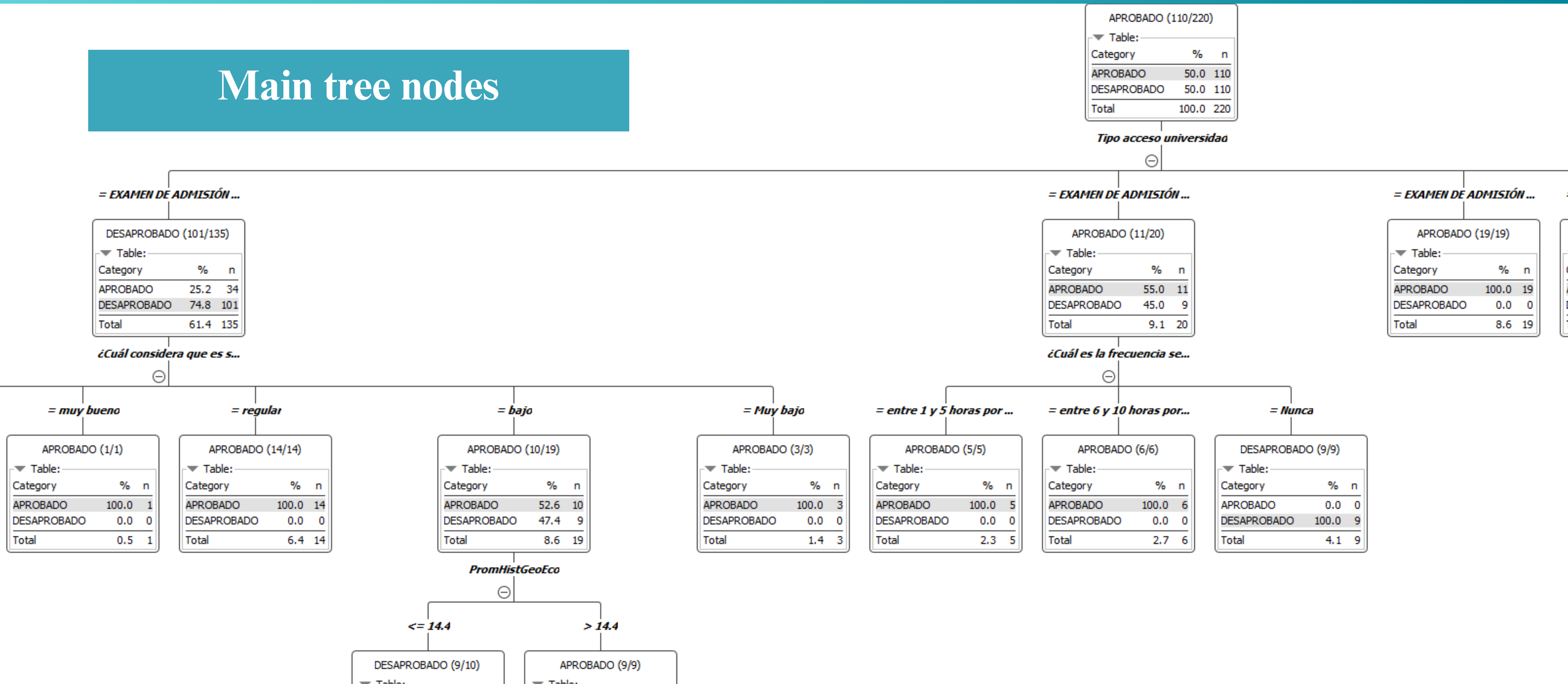
KNIME  
workflow





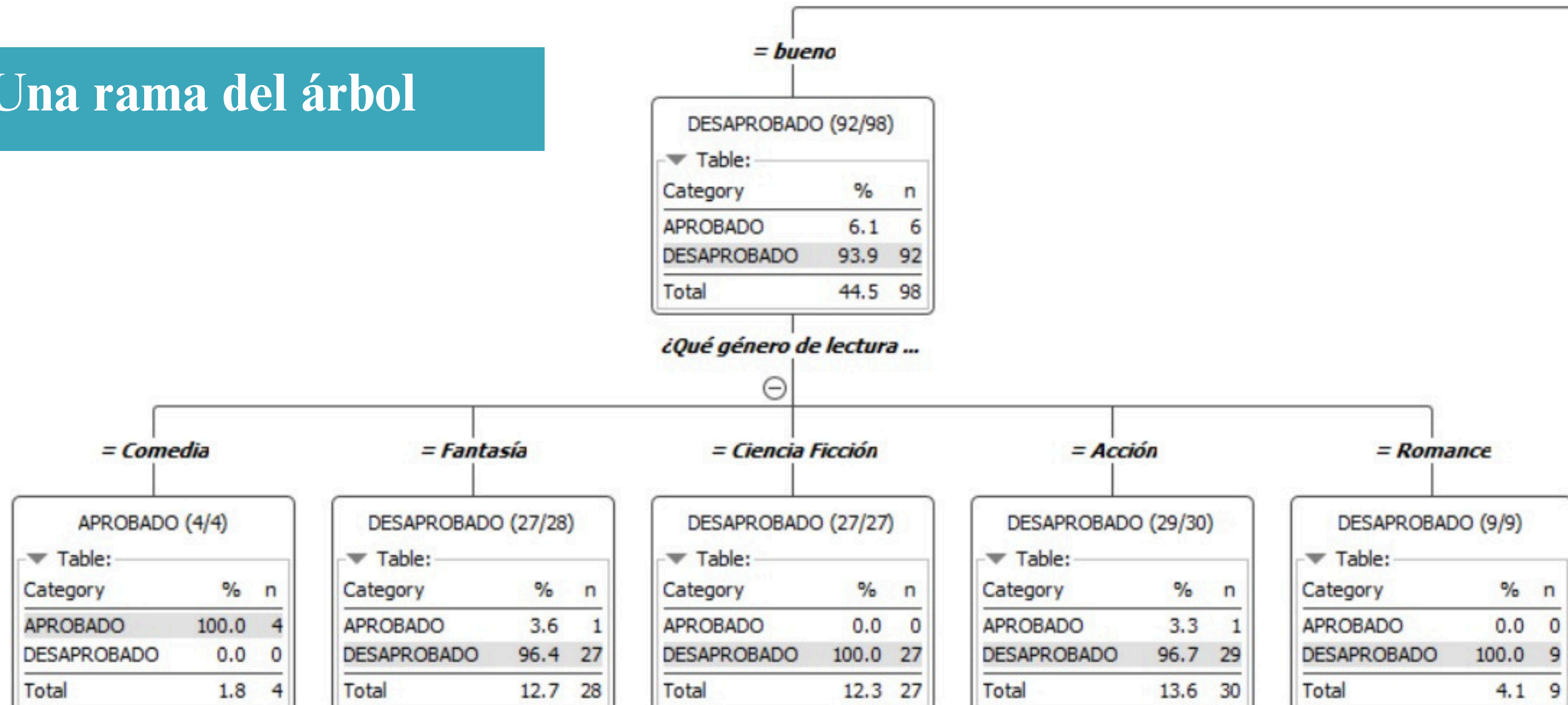
# Resultados y análisis

## Main tree nodes



# Resultados y análisis

## Una rama del árbol



# Resultados y análisis

## Decision Tree confusion matrix

		PREDICETED	
		Approved	Disapproved
ACTUAL	Approved	116	5
	Disapproved	4	9
	Total	96.69%	64.29%
	Overall Accuracy	93.28%	

# Resultados y análisis

## Random Forest confusion matrix

ACTUAL	PREDICETED		
	Approved	Disapproved	
	Approved	117	5
	Disapproved	4	9
	Total	96.69%	64.29%
Overall Accuracy		93.33%	

# Conclusiones

Se ha logrado identificar las principales variables que influyen en el rendimiento académico de los estudiantes de la facultad de ingeniería en la materia de matemáticas utilizando modelos de aprendizaje automático, especialmente árboles de decisión; **Se ha observado que el Bosque Aleatorio y el Árbol de Decisión tuvieron la misma precisión en la clasificación de nuestros datos, también se ha visto que las principales variables están relacionadas con las calificaciones escolares y los hábitos de lectura**, lo que consideramos de gran importancia para análisis posteriores de mucho trabajo más extenso.



# Referencias

- [1] Katarya R, Gaba J, Garg A, Verma V (2021) A review on machine learning based student's academic performance prediction systems. In: Proceedings - International Conference on Artificial Intelligence and Smart Systems, ICAIS 2021. Institute of Electrical and Electronics Engineers Inc., pp 254–259
- [2] Daud A, Lytras MD, Aljohani NR, et al (2017) Predicting student performance using advanced learning analytics. In: 26th International World Wide Web Conference 2017, WWW 2017 Companion. International World Wide Web Conferences Steering Committee, pp 415–421
- Prenkaj B, Velardi P, Stilo G, et al (2020) A Survey of Machine Learning Approaches for Student Dropout Prediction in Online Courses. ACM Comput Surv 53
- [3] Adekitan AI, Salau O (2020) Toward an improved learning process: the relevance of ethnicity to data mining prediction of students' performance. SN Appl Sci 2:. <https://doi.org/10.1007/s42452-019-1752-1>
- [4] Romero C, Ventura S (2020) Educational data mining and learning analytics: An updated survey. Wiley Interdiscip Rev Data Min Knowl Discov 10:. <https://doi.org/10.1002/widm.1355>

# Preguntas y respuestas



# Gracias!

etayaa@unjbg.edu.pe

UNIVERSIDAD NACIONAL  
**JORGE BASADRE  
GROHMANN**

