



LICENCIATURA EN CIENCIAS DE DATOS  
Propuesta de tema de tesis

## **MODELOS DE PREDICCIÓN DEL ABANDONO UNIVERSITARIO EN ARGENTINA**

*En Argentina, se estima que en las universidades nacionales sólo el 12% (Período 2010-2016) de los estudiantes que ingresan se gradúa, muy por debajo de otros países de la región, como en Chile, en donde se gradúa alrededor del 57%. Entendemos que el abandono estudiantil es, tal vez, el factor individual más importante que explica este fenómeno. La Universidad Nacional de Hurlingham en Argentina (UNAHUR) ha estado desarrollando modelos de predicción de abandono utilizando técnicas de Aprendizaje Automático, con el objetivo de prevenir el abandono estudiantil de manera temprana. Estos modelos se basan en los datos del Sistema de Información Universitaria Guaraní. Una vez entrenados, son capaces de detectar estudiantes con alto riesgo de abandono, a la vez que permiten indagar en algunos de los motivos subyacentes. La identificación temprana de estudiantes en riesgo de abandono permite desarrollar intervenciones más eficaces.*

**Palabras clave:** Modelo de Predicción, Ingeniería de Atributos, Abandono Universitario, Aprendizaje Automático

### Conocimientos deseables

*Técnicas de Aprendizaje Automático, Estadística*

¿Qué podría aprender quien realice esta tesis?

*-Como se mide y se trata el problema del abandono universitario en Argentina. - Como se puede predecir el abandono antes de que ocurra -Desarrollar modelos de predicción en general (aprendizaje supervisado) a través de la metodología CRISP-DM. -Ingeniería de atributos. -Tratamiento de datos faltantes -Métodos para la evaluación de modelos -Variables latentes.*

---

### Dirección de la tesis

*Pustilnik, Martin  
Universidad Nacional de Hurlingham*

Contacto: [martin.pustilnik@unahur.edu.ar](mailto:martin.pustilnik@unahur.edu.ar)

Más información en el pdf a continuación.

# Propuesta “MODELOS DE PREDICCIÓN DEL ABANDONO UNIVERSITARIO EN ARGENTINA”

La propuesta tiene tres etapas:

-Capacitar al alumno en el estado del arte, sugiriendo bibliografía y trabajos relacionados:

- [1] Síntesis de información estadísticas universitarias 2020-2021. <http://bibliotecadigital.cin.edu.ar/handle/123456789/2835>, accessed: 2023-04-01.
- [2] Istvan R.M., Falco M., Antonini S.A.: Análisis de los modelos educativos de universidades estatales chilenas. Revista Latinoamericana de Educación Comparada 53 (19), 1109–1113 (2022), <http://sedici.unlp.edu.ar/handle/10915/61343>.
- [3] Tinto, V.: Limits of theory and practice in student attrition. The Journal of Higher Education 53(6), 687–700 (1982), <http://www.jstor.org/stable/1981525>.
- [4] Pustilnik M., Ndukanma G.: Modelos para la predicción del abandono en la Universidad Nacional de Hurlingham (TEyET 2023). <https://teyet2023.unahur.edu.ar/> (en prensa).
- [5] Sistema de Información Universitaria Guaraní: <https://www.siu.edu.ar/>.
- [6] Lundberg, S., Lee, S.I.: A unified approach to interpreting model predictions. NIPS2017 (2017), <https://doi.org/10.48550/arXiv.1705.07874>.
- [7] Pustilnik, M. et al.: Estrechando el contacto entre universidades y estudiantes: comunicación ante posibles casos de abandono, propuestas para la inscripción. In: XXIV Workshop de Investigadores en Ciencias de la Computación (WICC 2022, Mendoza). pp. 734–738 (2022), <http://sedici.unlp.edu.ar/handle/10915/145216>.

Entro otros.

-Implementar modelos nuevos a partir de las variables existentes o nuevas variables generadas por el alumno:

Es importante distinguir que los motivos de abandono que predice el modelo están basados en los datos disponibles, y en general existen otros motivos subyacentes (variables latentes) a analizar para cada alumno. Por ejemplo, un tiempo de viaje elevado y no abandono podría estar midiendo la variable latente ‘El alumno tiene auto’.

Uno de los objetivos es poder sugerir cuales de estas variables son importantes para incluir en el próximo censo.

Por último, poder comparar dichos modelos con los actuales.

-Si es posible, es deseable publicar los resultados en algún congreso como Agranda, Wicc, TeYEt, entre otros.