
SPDE: Sistema Predictivo de Deserción Estudiantil

Miguel Tasende - Sebastián Laborde - Juan Pablo Montesano

Motivación

Deserción estudiantil tiene una alta incidencia en Uruguay.

La pobreza y la inequidad de oportunidades son motivos importantes que conducen a la deserción estudiantil.

Necesidad de políticas y planes de acción específicos que vayan en la dirección de revertir los niveles de deserción escolar.



Objetivos

Dotar al estado de una herramienta que permita predecir el riesgo de deserción de los estudiantes del nivel básico y medio en el sistema educativo uruguayo.

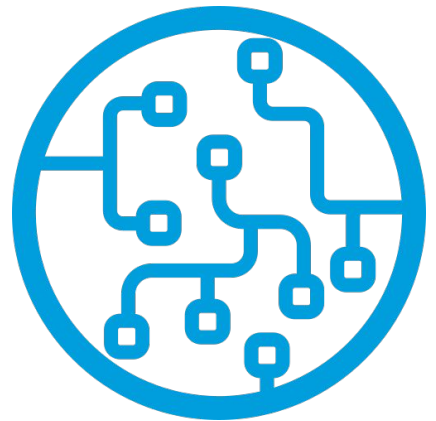
Contribuir a tomar acciones correctivas y preventivas para reducir el índice de deserción estudiantil.

Mejorar la calidad de la educación evitando deserciones intermitentes y altas tasas de inasistencias.



Propuesta

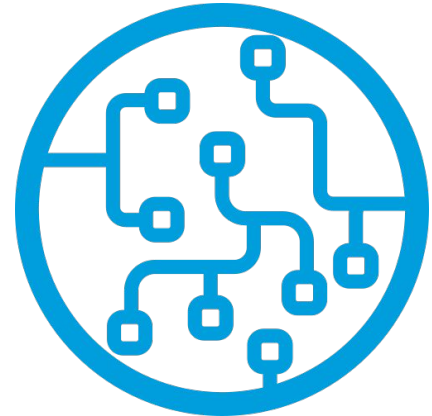
Diseñar un sistema IA (machine learning) que aprende (supervised learning) en base a la historia académica del estudiante (**ANEP-CODICEN**), junto a los factores socio económicos y ambientales (**Censo-INE**), que determinan su condición de potencial desertor y permite predecir para el siguiente periodo escolar si un estudiante desertara (1) o no (0) del sistema educativo con la asociada probabilidad de ocurrencia, [deserción(0,1), riesgo(%)].



Propuesta

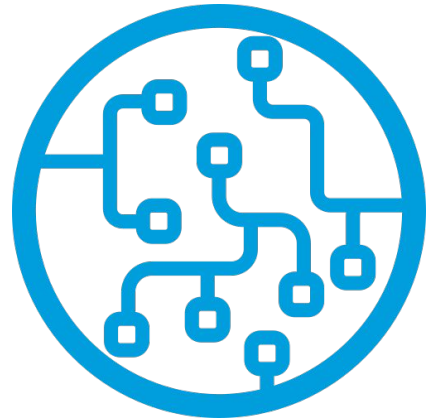
Prototipo Escalable: Modelo predictivo a aplicar sobre un conglomerado de centros educativos con características demográficas similares.

- Abordar el problema completo no es viable, conviene mantener la solución simple.
- Mayor cantidad de deserción en centros socio-económicos bajo.
- Rapidez de desarrollo para obtener resultados inmediatos.
- No perdemos generalidad, podemos escalar llevando la solución a otros conglomerados re-entrenando el modelo.



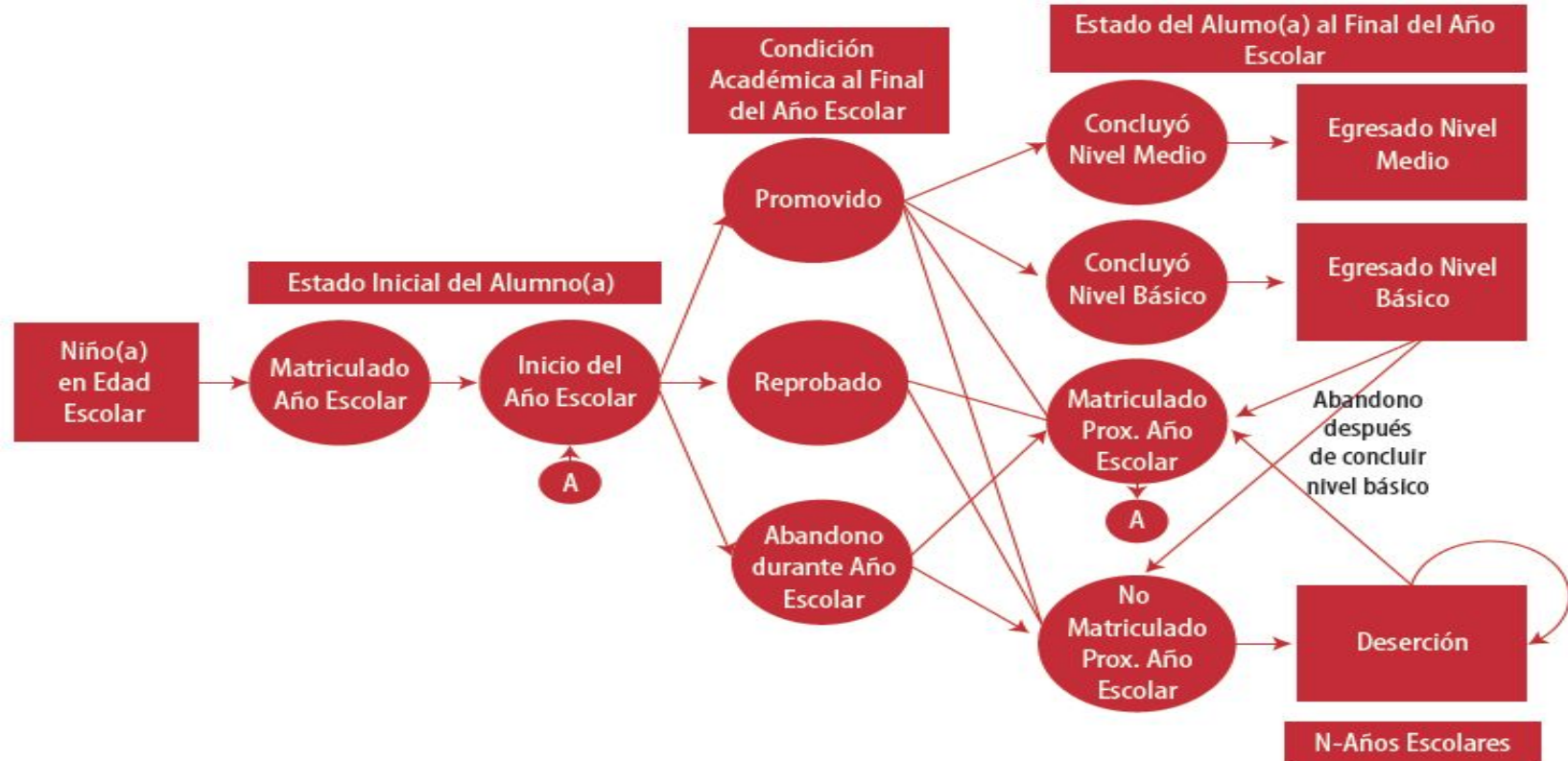
Metodología

- Entendimiento del problema
- Comprensión de los datos
- Preparación de los datos.
- **Feature Engineering.**
- **Modelado (Algoritmos y Entrenamiento).**
- **Evaluación y Mejora del Modelo.**
- Distribución del Modelo.



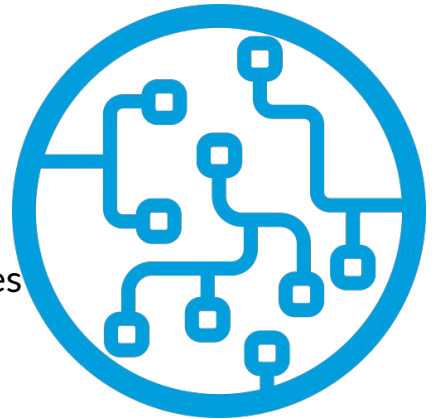
Variables (Feature Engineering)

DIAGRAMA DE ESTADO DESERCIÓN ESCOLAR



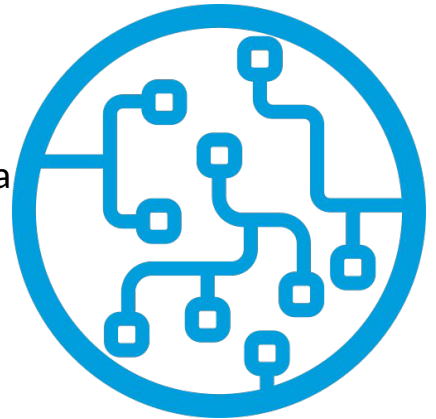
Variables (Feature Engineering)

- Condición académica del alumno(a) al final del año escolar.
- Tiempo de permanencia del alumno en el sistema educativo al momento del corte del estudio y antes de pasar a condición de deserción o de egresado.
- Último grado alcanzado antes de pasar a su condición de deserción.
- Cantidad de abandonos tenidos antes de pasar a su condición de deserción o de egresado.
- Tiempo de deserción transcurrido, es decir, la cantidad de años escolares sin retornar al sistema.



Variables (Feature Engineering)

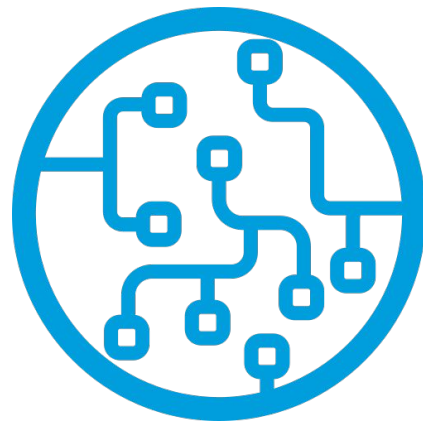
- Cantidad de reprobaciones tenidas antes de pasar a su condición de deserción o egresado.
- Cantidad de inasistencias antes de pasar a su condición de deserción o egresado.
- Cantidad de promociones tenidas antes de pasar a la condición de deserción o egresado.
- Si se ha transferido de centro educativo durante su estadía en el sistema antes de la condición de deserción o egresado (movilidad).



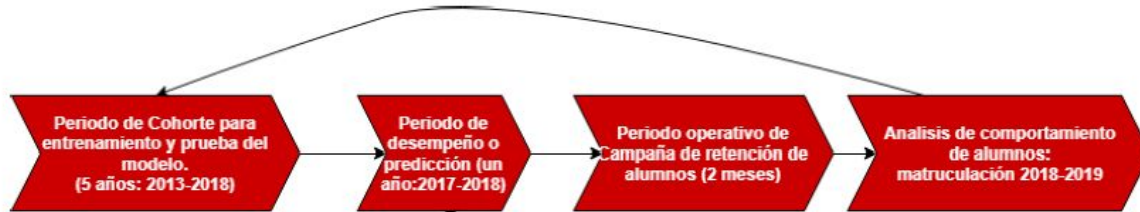
Variables (Feature Engineering)

LISTA DE ATRIBUTOS DE ENTRADA

NO.	NOMBRE ATRIBUTO	DESCRIPCIÓN	EJEMPLO
1	IdEstudiante		35
2	Año_Academico_first_1		2009-2010
3	Año_Academico_last_1		2013-2014
4	CodigoCentro_first		2358
5	CodigoCentro_last		359
6	Sector_last		PUBLICO
7	Tanda_last		MATUTINA
8	FechaNacimiento_last		29-Nov-04
9	Sexo_first_1		Masculin
10	ICV1	Pertenece o no al programa	0
11	nivel_grado_last		14
12	Abandono_sum		0
13	Promovido_sum		2
14	Reprobado_sum		0
15	Condicion_otra_sum		0
16	anios_acad		2
17	EdadEstudiante		10
18	ZonadelCentro		URBANA-M
19	IndicePobrezaCentro	vulnerabilidad del sector geográfico	69.88
20	IndiceProsoliCentro	Indice de cantidad de estudiantes en el programa	8.2
21	Condicion_Acad_last		Promov
22	Desercion		0 o 1



Modelado: Ciclo de Vida del Modelo Predictivo

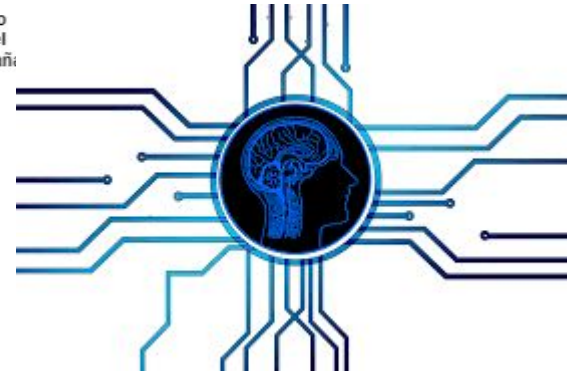


A partir de las bases de datos del historial académico y datos demográficos del alumno se analiza su comportamiento, se determinan los umbrales de deserción y se produce el modelo clasificador de deserción

Se preparan los datos y se realiza la predicción de deserción en base a su desempeño durante ese periodo

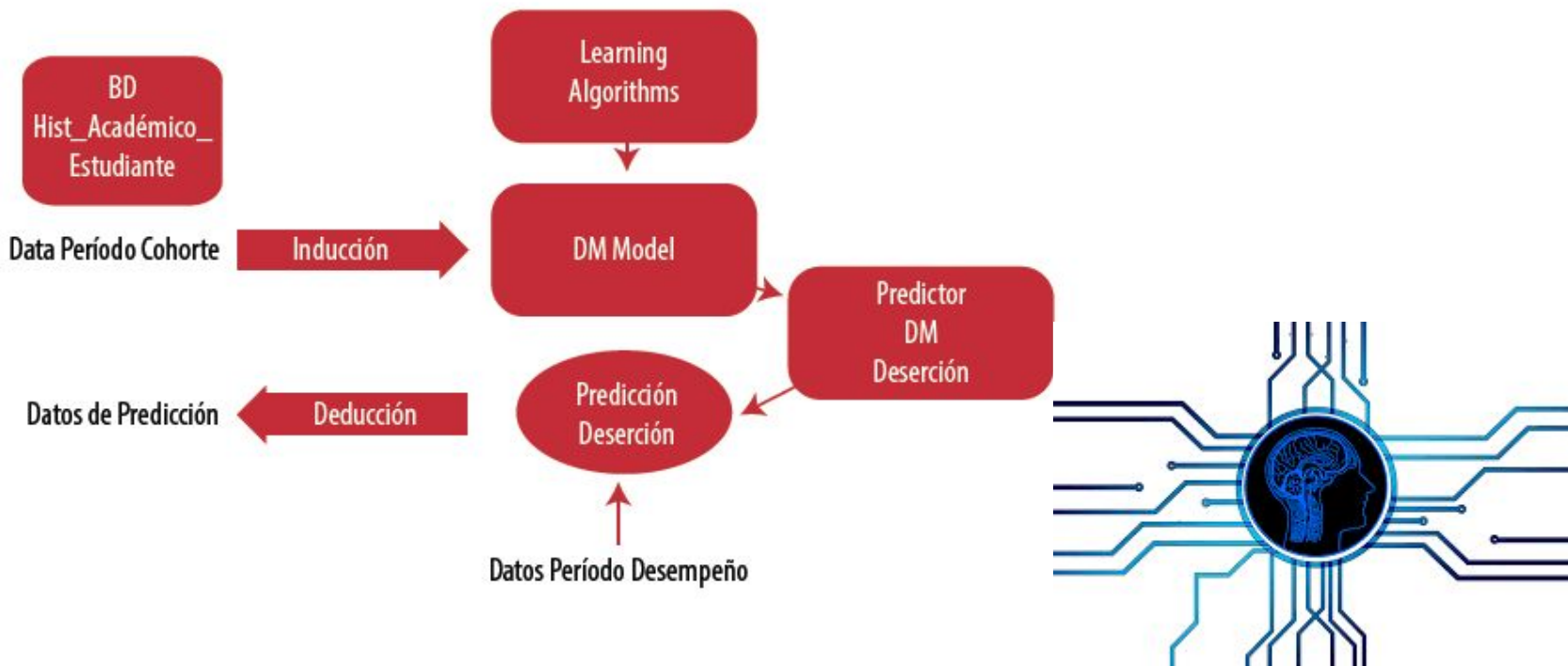
Se cargan los datos de predicción en la BD del sistema para realización de las campañas de retención

Se analiza el comportamiento de los alumnos y se evalúa el nivel de respuesta de la campaña para medir su efectividad



Implementación

COMO APRENDE Y PREDICE EL MODELO



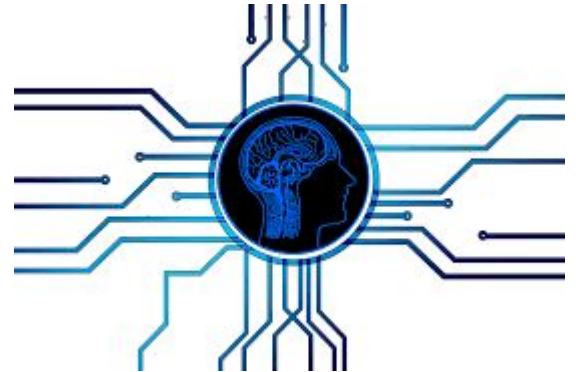
Implementación

Algoritmos de Aprendizaje:

Random Forest.

Gradient Boosting (XGBoost, LightGBoost).

Redes Neuronales.



Validación

Utilizamos el periodo escolar actual (2017) para predecir el próximo periodo escolar (2018). Cuando se registra la matriculación de los alumnos en (2018) es que se contrasta con la predicción realizada.

ID ESTU- DIANTE	AÑO ACA- DEMICO	CODIGO CENTRO	TANDA	NIVEL_ GRADO	CONDICIO ACADE- MICA	SEXO	ICV1	INDICE PROSOLI CENTRO	INDICE POBREZA CENTRO	EDAD ESTU- DIANTE	PROMOVI- DO_SUM	REPRO- BADO_ SUM	ABANDO- NO_SUM	CONDI- CION_ OTRA	CAM- BIO_ CENTRO	IL_ ANIOS_ ACAD	SR-DE- SERCION	SRC-DE- SERCION
35	2013-2014	359	MATUTINA	14	Promovido	Masculino	0	8.2	69.88	8.5	2	0	0	0	0	2	1	79.50%
134	2013-2014	226	VESPERTINA	17	Promovido	Femenino	1	24.3	37.04	12.53	5	0	0	0	0	5	0	59.50%
258	2013-2014	3061	VESPERTINA	15	Promovido	Femenino	1	12.5	36.32	10.02	3	0	0	0	0	3	1	75.80%
356	2013-2014	230	MATUTINA	16	Promovido	Femenino	0	13.5	37.04	11.02	5	0	0	0	0	5	0	71.00%
429	2013-2014	224	VESPERTINA	12	Promovido	Masculino	0	20.01	43.31	9.57	2	1	1	0	0	4	0	82.70%
1605	2013-2014	5710	MATUTINA	21	Promovido	Femenino	1	29.6	35.75	15.67	5	0	0	0	0	5	0	75.50%
1637	2013-2014	229	MATUTINA	16	Promovido	Femenino	0	10.5	37.04	11.79	2	0	0	0	0	2	1	72.90%
1777	2013-2014	13402	VESPERTINA	21	Promovido	Femenino	1	20.01	43.31	14.3	4	0	0	0	0	4	0	75.10%
1800	2013-2014	4852	MATUTINA	17	Promovido	Femenino	0	7.7	22.24	12.32	3	0	0	0	0	3	1	75.90%
1959	2013-2014	528	VESPERTINA	16	Promovido	Masculino	0	18	59.59	10.62	3	1	0	0	0	4	0	59.50%



Validacion

Se especifica un umbral para la medida de precisión (ejemplo 80%), tal que, si el resultado de la prueba de validación es menor que ese número, se procede a reajustar el modelo.



Medida de Exactitud

En los modelos de deserción escolar se presenta el fenómeno de que la cantidad de deserciones (1s o verdaderos positivos) es muy inferior a la de no deserciones (0s o verdaderos negativos).

Usaremos F1 Score como medida de precisión del modelo.

$$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$



Supuestos y Restricciones

Podemos contar con los datos y son suficientes para entrenar satisfactoriamente el modelo.

Restringimos a un piloto en un conglomerado de centros educativos.

Ayuda de un experto en educación.



Preguntas?



Gracias!