

¿Cómo estimar el riesgo de la exclusión educativa?

Dr. Patricio Rodríguez

Profesor asistente
Instituto de Estudios Avanzados en Educación
Universidad de Chile

 prodriguez@uchile.cl

 www.linkedin.com/in/prodrigv



FONDEF
Fondo de Fomento al Desarrollo Científico y Tecnológico

Santiago, Chile
Abril 15, 2021

Objetivos

- Entender el **comportamiento del fenómeno** de la exclusión escolar de los estudiantes en Chile a lo largo del **tiempo**.
- Identificar los **factores** que más **inciden** en que se produzca la exclusión escolar de estudiantes en Chile .
- Estimar el riesgo que tiene cada niña, niño y joven de excluirse del sistema educativo regular.

Trayectorias escolares

Trayectorias escolares

Definiciones

Exclusión escolar:

Situación de una niña, niño o joven (**NNJ**) que habiendo estado matriculada(o) en algún nivel de enseñanza durante el periodo escolar de un año t , al año siguiente $t + 1$:

- **Regular**: no está matriculada(o) en ningún establecimiento (de cualquier tipo) o está en educación de adultos.
- **Global**: no está matriculada(o) en ningún establecimiento de cualquier tipo.

Trayectorias escolares

Datos utilizados

Centro de Estudios Mineduc (CEM):

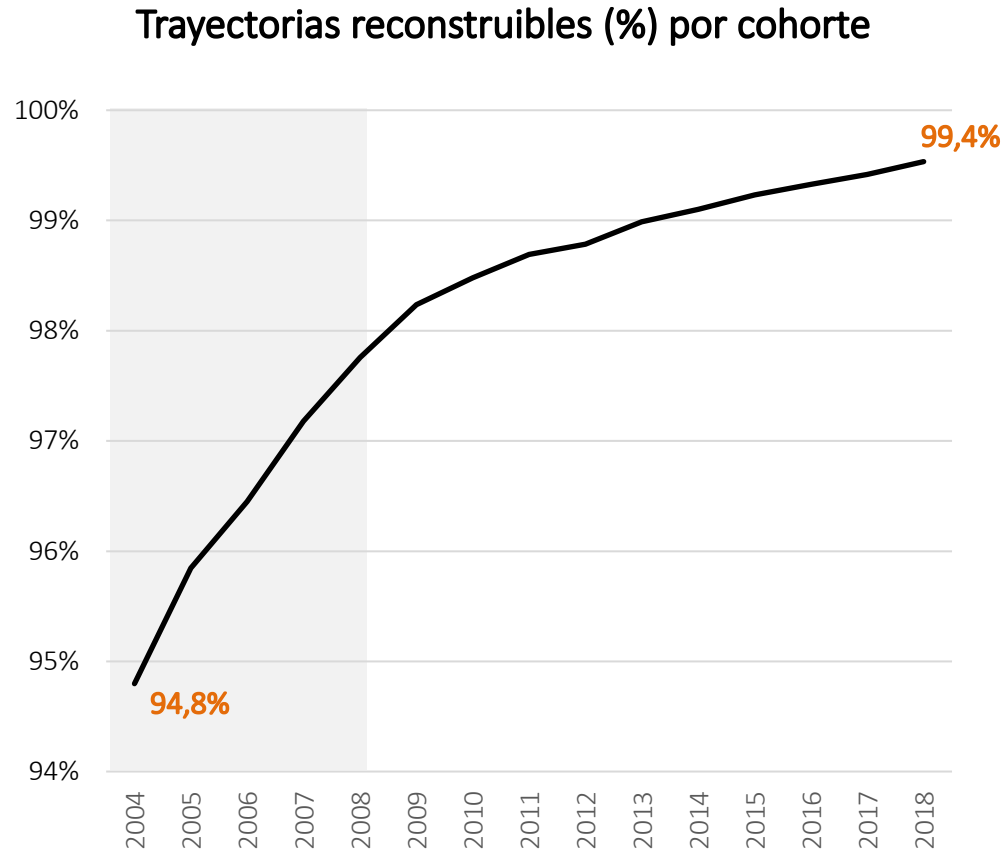
- Bases de **matrícula y rendimiento individuales** 2004 – 2020.
- Directorio de **establecimientos abiertos** 2004 – 2020.
- Datos de estudiantes **graduados y fallecidos**.

Agencia de la Calidad de la Educación (ACE)

- Resultados [SIMCE](#).
- Cuestionario padres y apoderados SIMCE.

Trayectorias escolares

Análisis y creación de trayectorias individuales



Trayectorias escolares

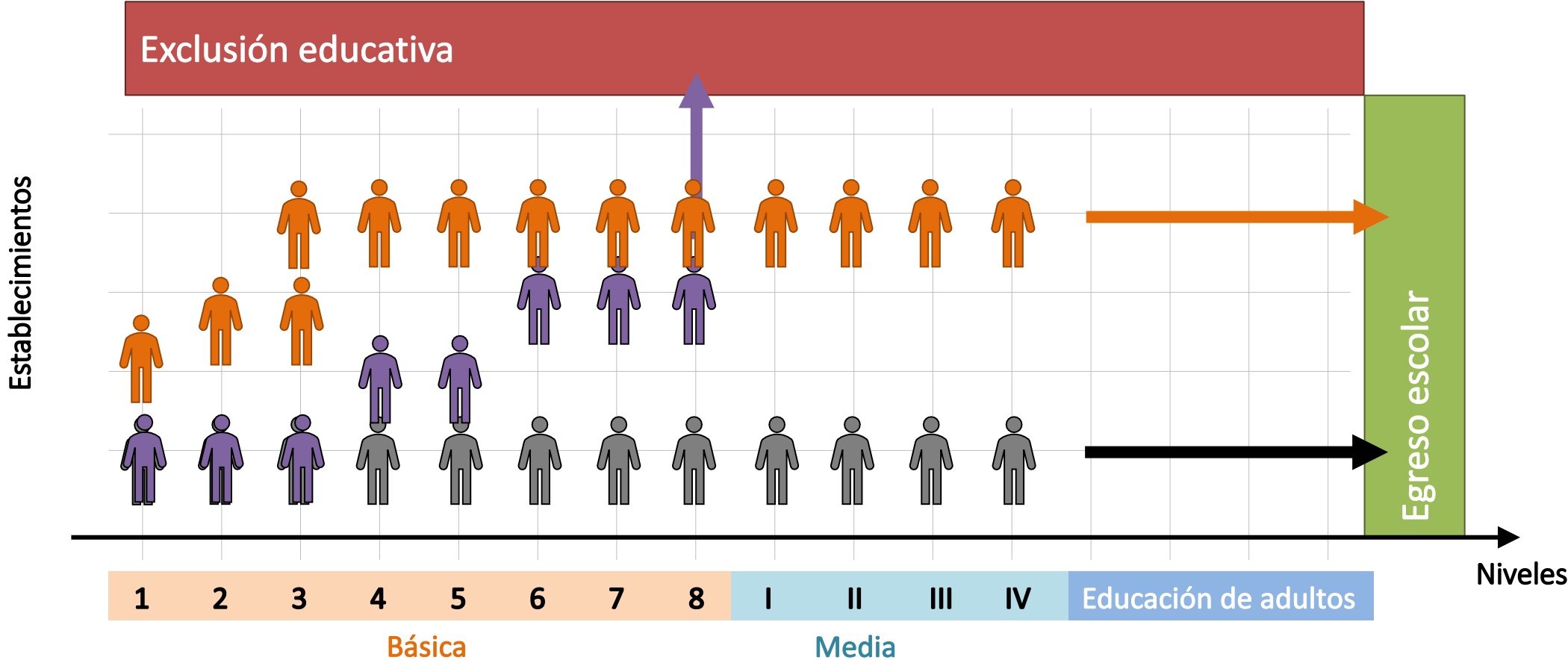
Caracterización de trayectorias

Para ello, se codifica la trayectoria de cada estudiante en un conjunto de **variables** que las resumen en las siguientes **dimensiones**:

- **Características individuales del estudiante:** fecha de nacimiento, género, etnia, nacionalidad, nivel socioeconómico, beneficiarios SEP, si estuvo embarazada.
- **Historial de desempeño:** asistencia (inasistencia crónica), rendimiento (puntaje z relativo en a su nivel), resultados en estándar aprendizaje SIMCE, cambios de curso.
- **Historial de residencia:** cambios de residencia (a nivel comunal).
- **Historial de establecimientos:** cambios de EE (inter/intra año), si hubo cambio por cierre/ o se acabó su escuela, datos del establecimiento final del estudiante (graduado o con deserción/exclusión).
- **Fenómeno que (no) le afecta:** exclusión (regular, global), reingresos, graduación, defunción.
- **Variables del EE:** cantidad y rotación de docentes, trayectorias de desempeño escolar (IDE).

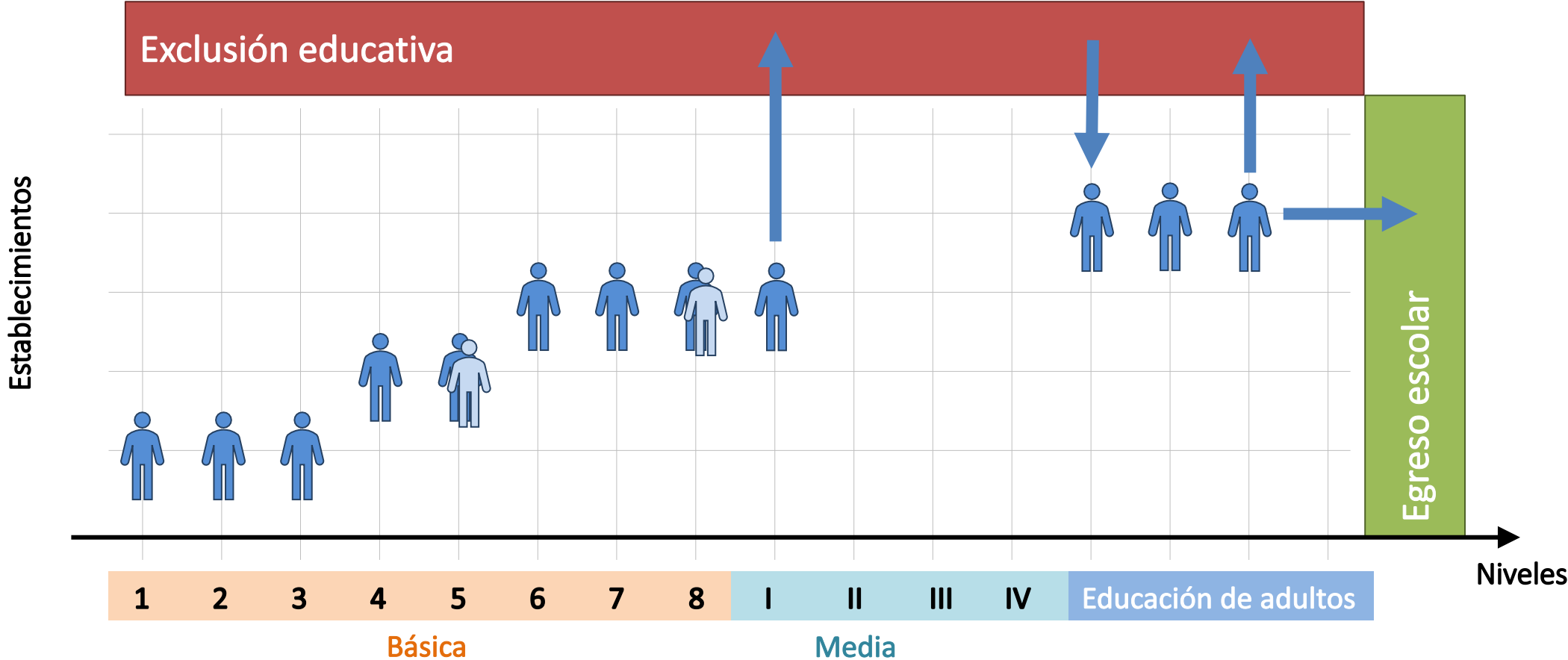
Trayectorias escolares

Caracterización de trayectorias



Trayectorias escolares

Caracterización de trayectorias



Trayectorias escolares

Trayectorias y exclusión escolar: Cohortes 2004 – 2008



Modelo predictivo basado en trayectorias escolares

Predicción exclusión escolar

Qué predecir

- Queremos **evitar** que cualquier estudiante se **desvincule** de su establecimiento, con todas las implicancias que ello conlleva.
- **Por lo tanto**, nos concentraremos en predecir el fenómeno de la **deserción regular**:
 - Detectar el riesgo que cualquier estudiantes de educación regular no continúe por **primera vez** sus estudios.
 - Dentro del próximo año (**$t + 1$**) o el subsiguiente (**$t + 2$**).

Predicción exclusión escolar

Variables:

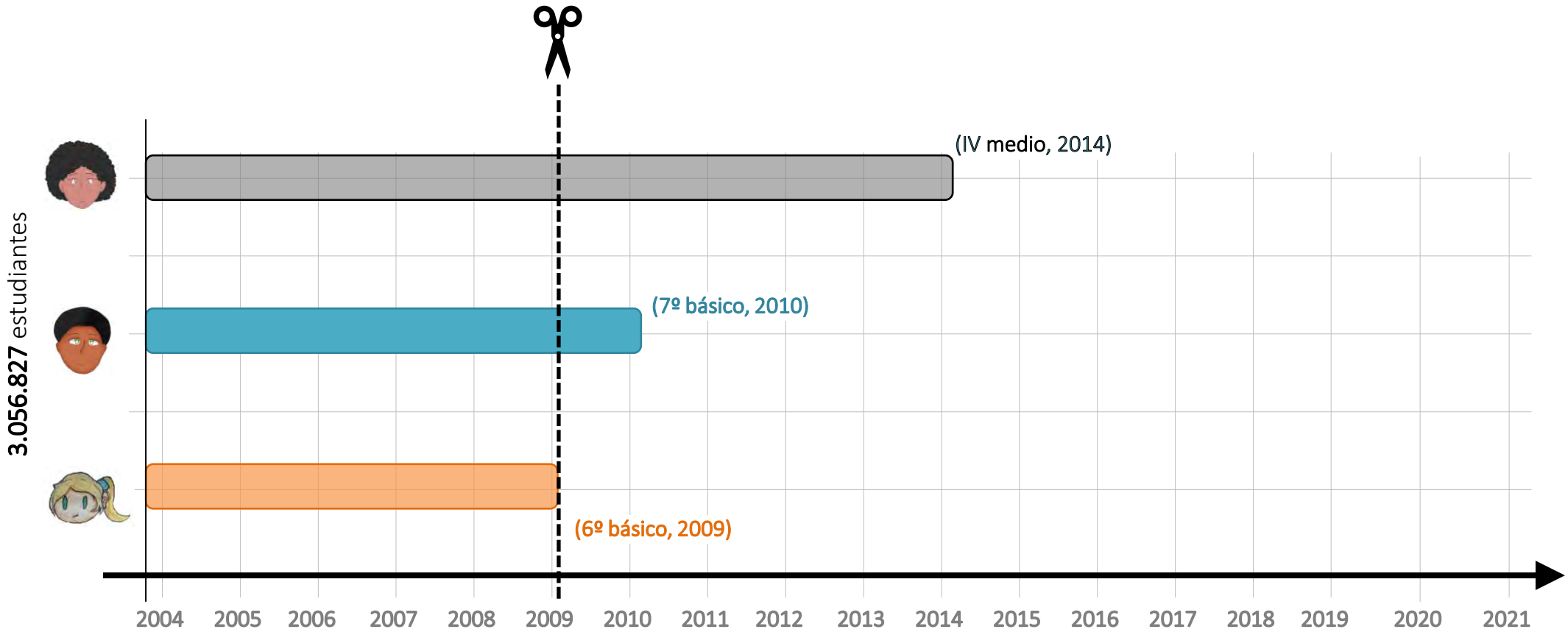
Totalizan **114** variables para Exclusión $t+1$ y Exclusión $t+2$

Algunos ejemplos de variables:

Variable	Descripción	Variable	Descripción	Variable	Descripción
NSE	Nivel socio-económico (ingreso + escolaridad padres)	Q_CAMB_CERR	Cantidad de cambios por cierre establecimiento	Q_DEGR_PUB	Años en dependencia pública (Municipal + SLE)
GEN_ALU	Género estudiante	NIVEL_PARC	Último nivel donde se observa la trayectoria	Q_DEGR_SUB	Años en dependencia subvencionada
CAMB_EB	Cambio por fin de nivel en escuela	Q_RBD	Total de EE por los cuales pasó el estudiante	Q_DEGR_PAG	Años en dependencia particular pagada
EXTR_ALU	Estudiante es extranjero(a)	Q_COM_ALU	Cantidad de comunas diferentes reportadas como domicilio	Q_DEGR_ADE	Años en dependencia administración delegada
EMB_ALU	Estudiante estuvo embarazada	Q_RUR	Cantidad EE rurales en que se estudió	IC_TOTAL	Años de inasistencia crónica (< 90%)
ALU_INT	Estudiante estuvo en programa integración	Q_CAMB_INTRA	Cantidad de años donde hubo cambios dentro del año	Q_ABN_NJ	Cantidad de abandonos a mediados de año en educación regular
ETNIA_ALU	Etnia del estudiante	Q_CAMB_INTER	Cantidad de años donde hubo cambios entre años	ULT_PZ	Puntaje Z último ciclo disponible
ULT_EDAD	Última edad registrada	Q_REPIT	Cantidad de repitencias	ULT_PA	Promedio asistencia ciclo disponible

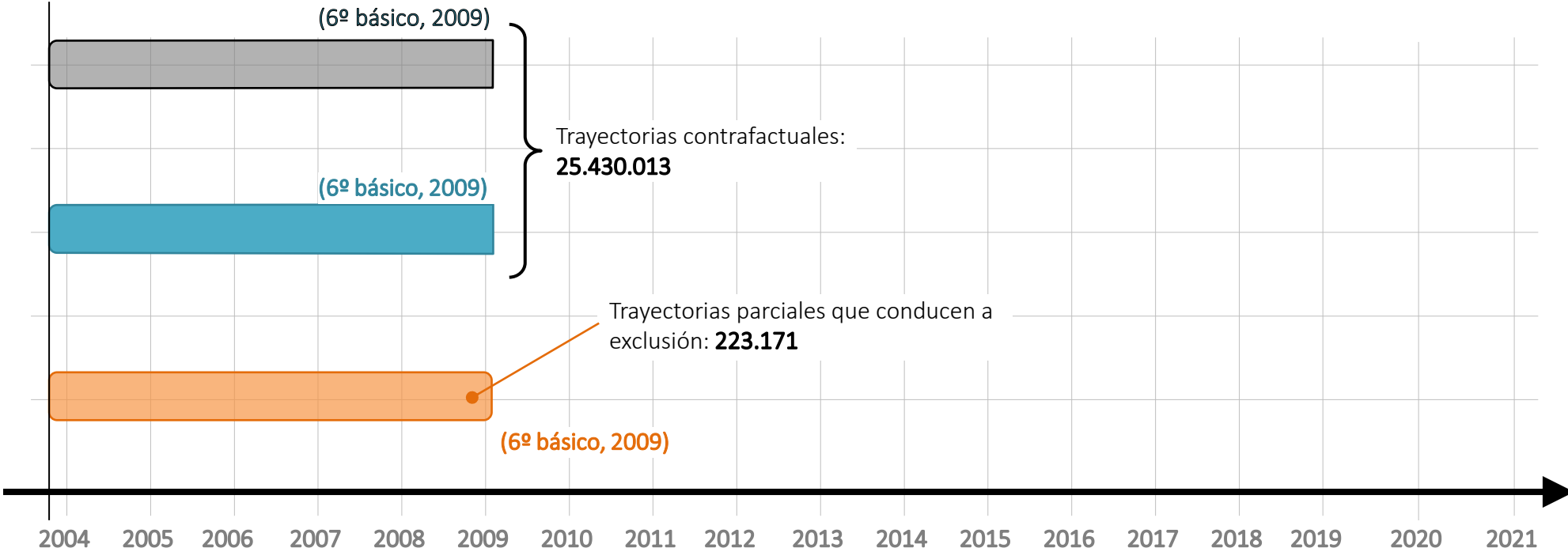
Predicción exclusión escolar

Generación de muestra en base a trayectorias: **contrafactuales**



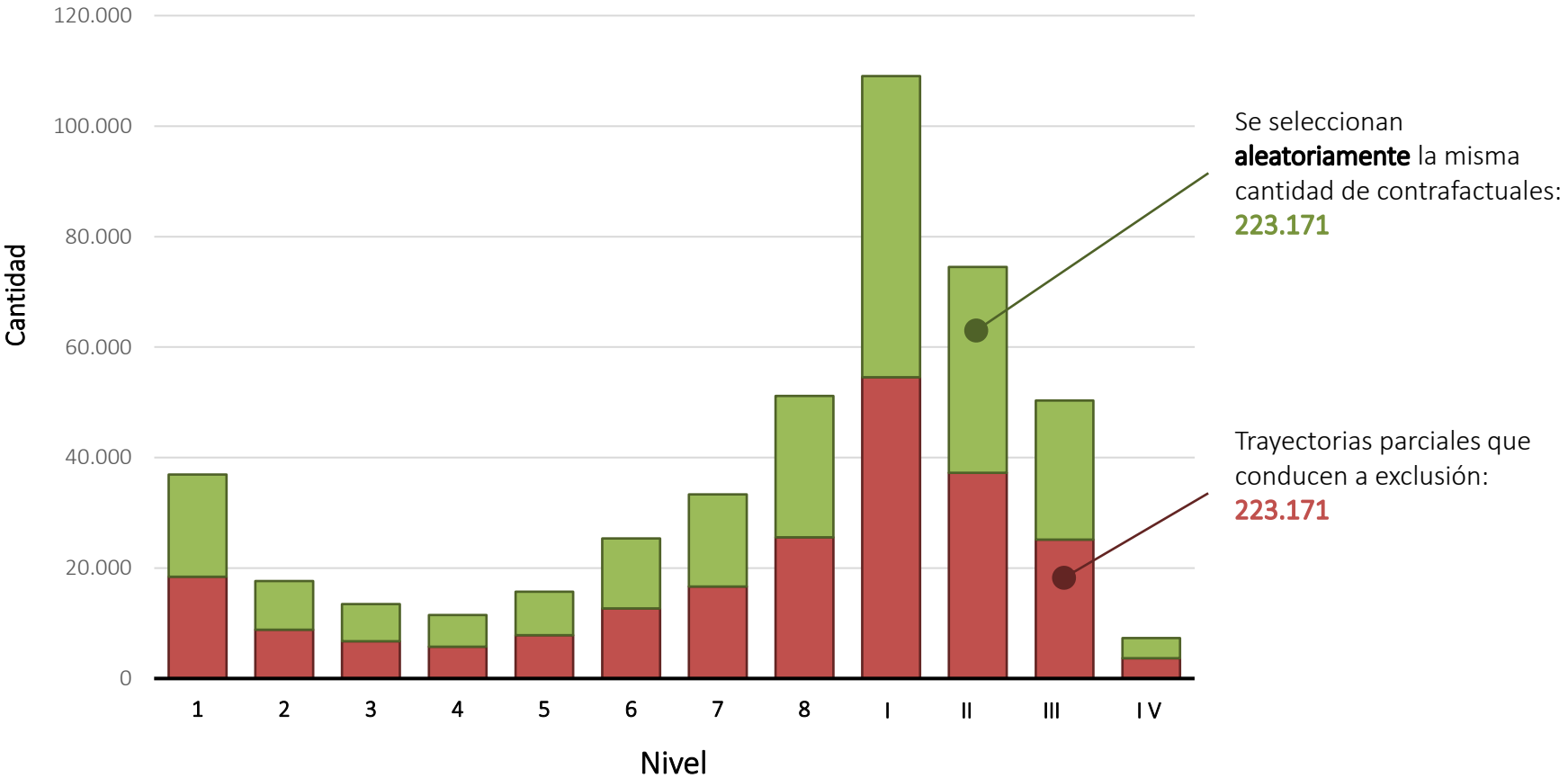
Predicción exclusión escolar

Generación de muestra en base a trayectoria



Predicción exclusión escolar

Generación de muestra en base a trayectoria



Predicción exclusión escolar

Algoritmos

Algoritmos a utilizar:

Se prefirió utilizar algoritmos de aprendizaje supervisado de la familia de árboles de decisión, dada su **mayor explicabilidad**:

- [Decision Tree](#)
- [XGBoost](#)
- [CatBoost](#)
- [LightGBM](#)

Predicción exclusión escolar

Generación de modelo predictivo



Muestra de trayectorias

year	id_inst	id_codi_cem_1	id_codi_cem_2	nom_com_alu	lat_com_alu	lon_com_alu	nom_com_ced	lat_com_ced	lon_com_ced
2014	1201	1310	Puerto Aho	-13.5007862	-76.5795721	La Florida	-13.5205941	-76.0124905	144
2014	1301	1310	Puerto Aho	-13.5007862	-76.5795721	La Florida	-13.5205941	-76.0124905	149
2014	1301	1310	Puerto Aho	-13.5007862	-76.5795721	La Florida	-13.5205941	-76.0124905	154
2014	1301	1310	Puerto Aho	-13.5007862	-76.5795721	La Florida	-13.5205941	-76.0124905	159
2014	1301	1310	Puerto Aho	-13.5007862	-76.5795721	La Florida	-13.5205941	-76.0124905	164

Muestra de entrenamiento

year	id_inst	id_codi_cem_1	id_codi_cem_2	nom_com_alu	lat_com_alu	lon_com_alu	nom_com_ced	lat_com_ced	lon_com_ced	year
2014	1201	1310	Puerto Aho	-13.5007862	-76.5795721	La Florida	-13.5205941	-76.0124905	144	2014
2014	1301	1310	Puerto Aho	-13.5007862	-76.5795721	La Florida	-13.5205941	-76.0124905	149	2014
2014	1301	1310	Puerto Aho	-13.5007862	-76.5795721	La Florida	-13.5205941	-76.0124905	154	2014
2014	1301	1310	Puerto Aho	-13.5007862	-76.5795721	La Florida	-13.5205941	-76.0124905	159	2014

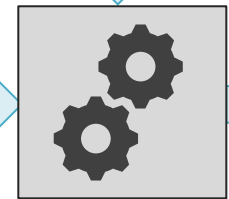
$$S_e \cap S_v = \emptyset$$

Muestra de validación

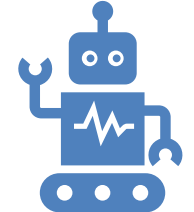
year	id_inst	id_codi_cem_1	id_codi_cem_2	nom_com_alu	lat_com_alu	lon_com_alu	nom_com_ced	lat_com_ced	lon_com_ced	year
2014	1301	1310	Puerto Aho	-13.5007862	-76.5795721	La Florida	-13.5205941	-76.0124905	144	2015
2014	1301	1310	Puerto Aho	-13.5007862	-76.5795721	La Florida	-13.5205941	-76.0124905	149	2015
2014	1301	1310	Puerto Aho	-13.5007862	-76.5795721	La Florida	-13.5205941	-76.0124905	154	2015
2014	1301	1310	Puerto Aho	-13.5007862	-76.5795721	La Florida	-13.5205941	-76.0124905	159	2015

Usualmente $\varphi = 0,2$

Algoritmo



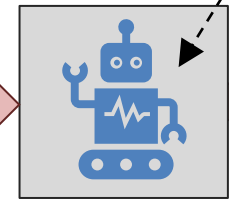
Entrenamiento



Modelo ajustado

Validación

Modelo predice los resultados de la muestra de prueba: conocidos



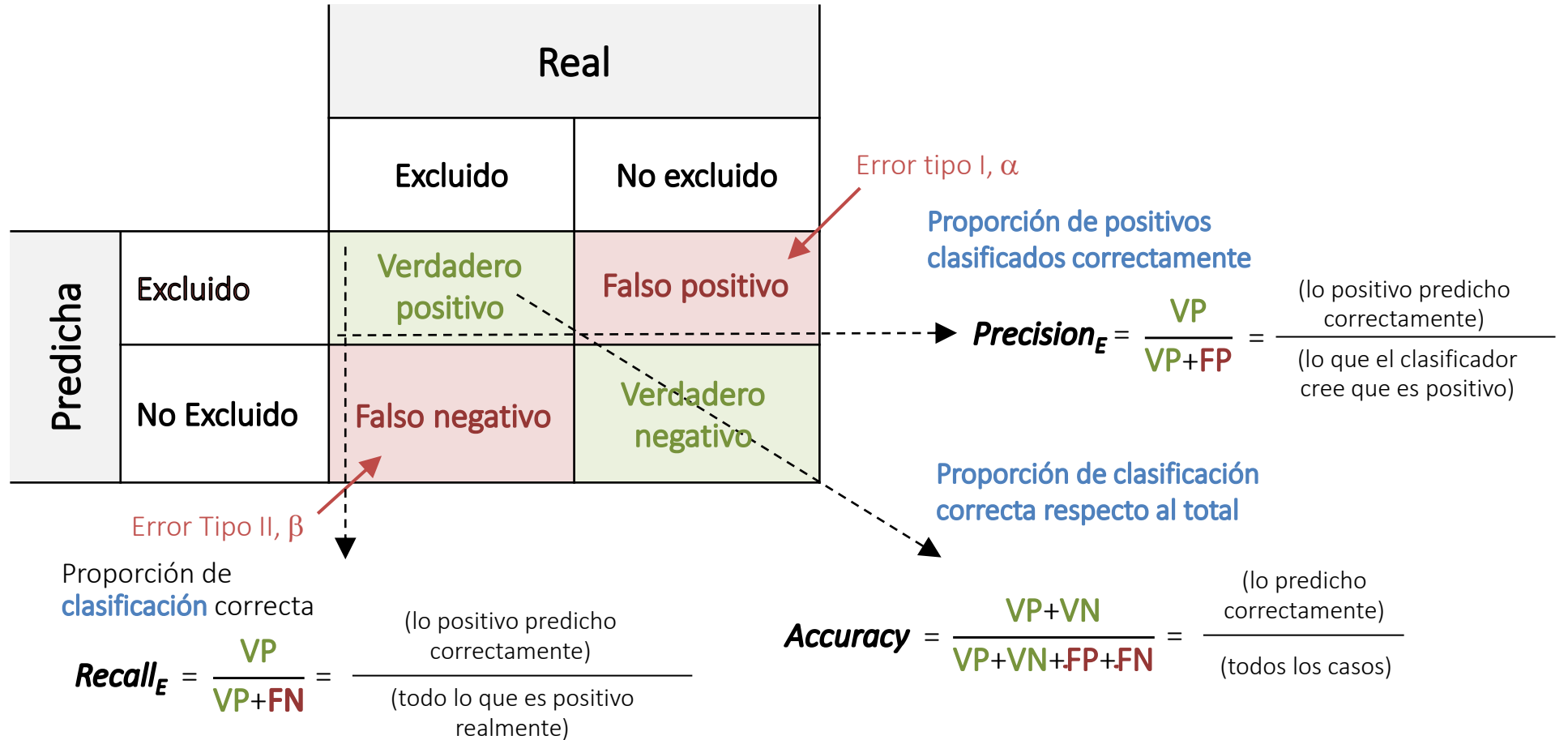
Modelo predice

Desempeño

Compara predicción y resultados reales

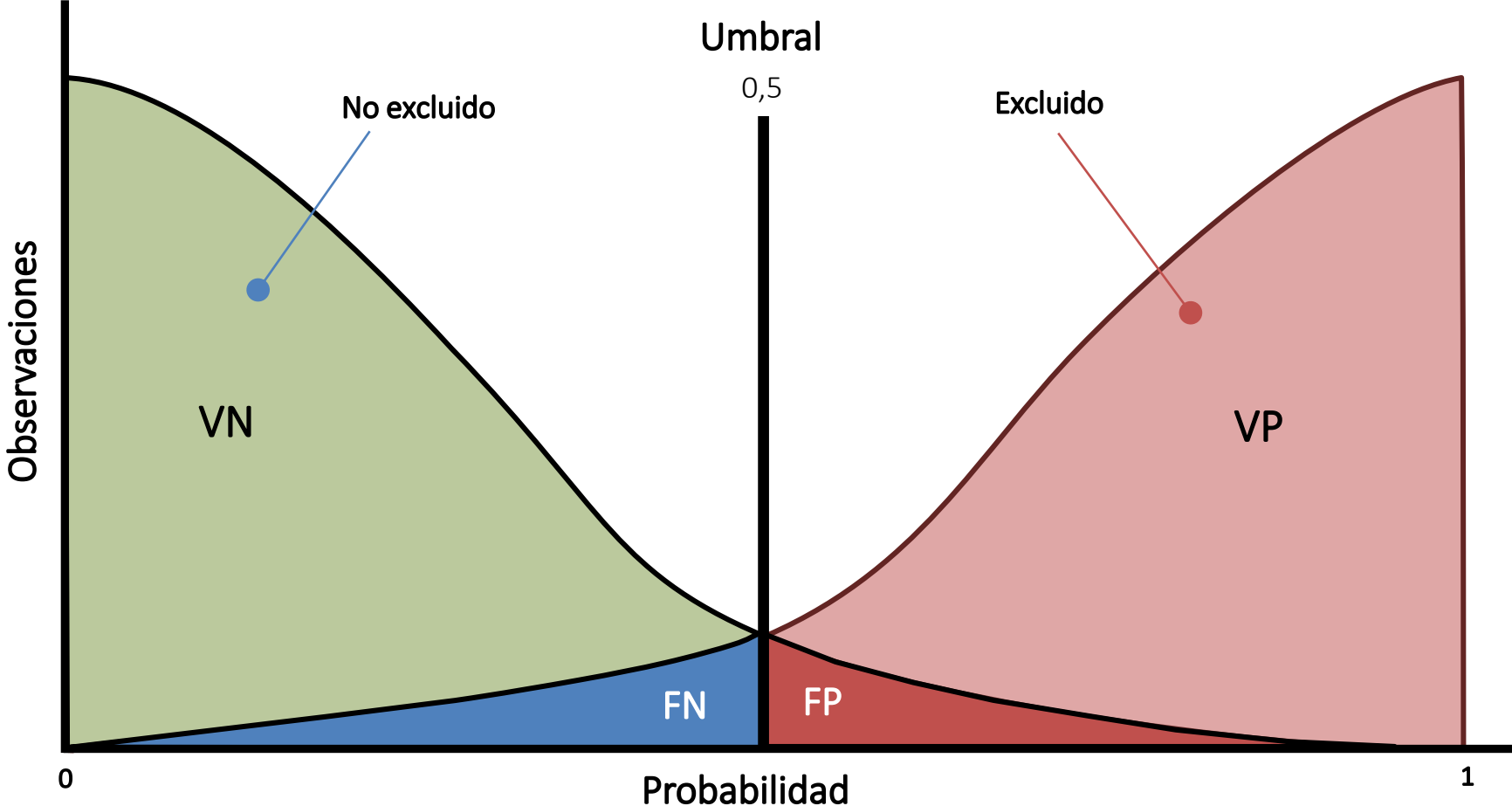
Predicción exclusión escolar

Métricas de desempeño: matriz de confusión



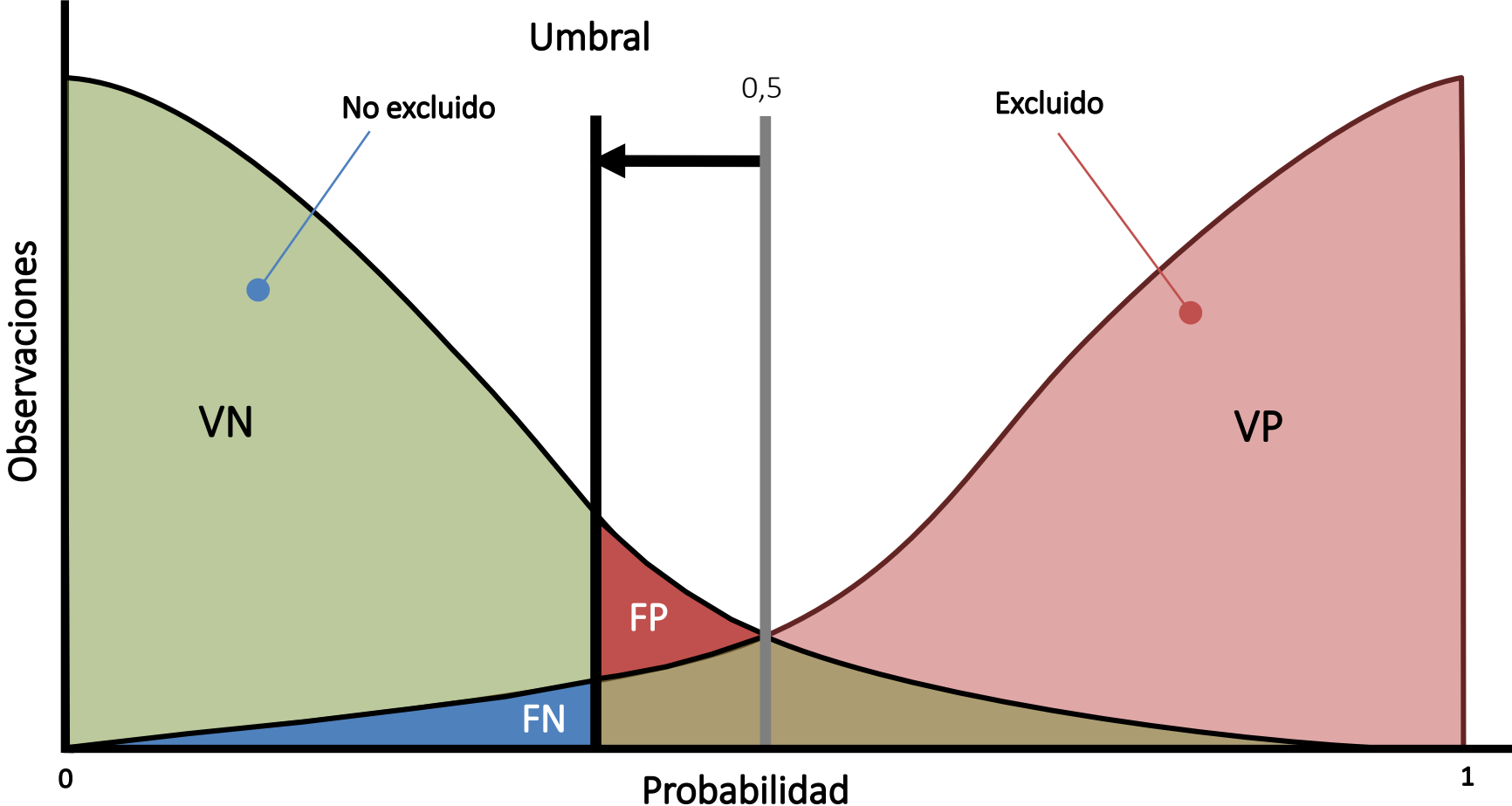
Predicción exclusión escolar

Métricas de desempeño: **umbral**



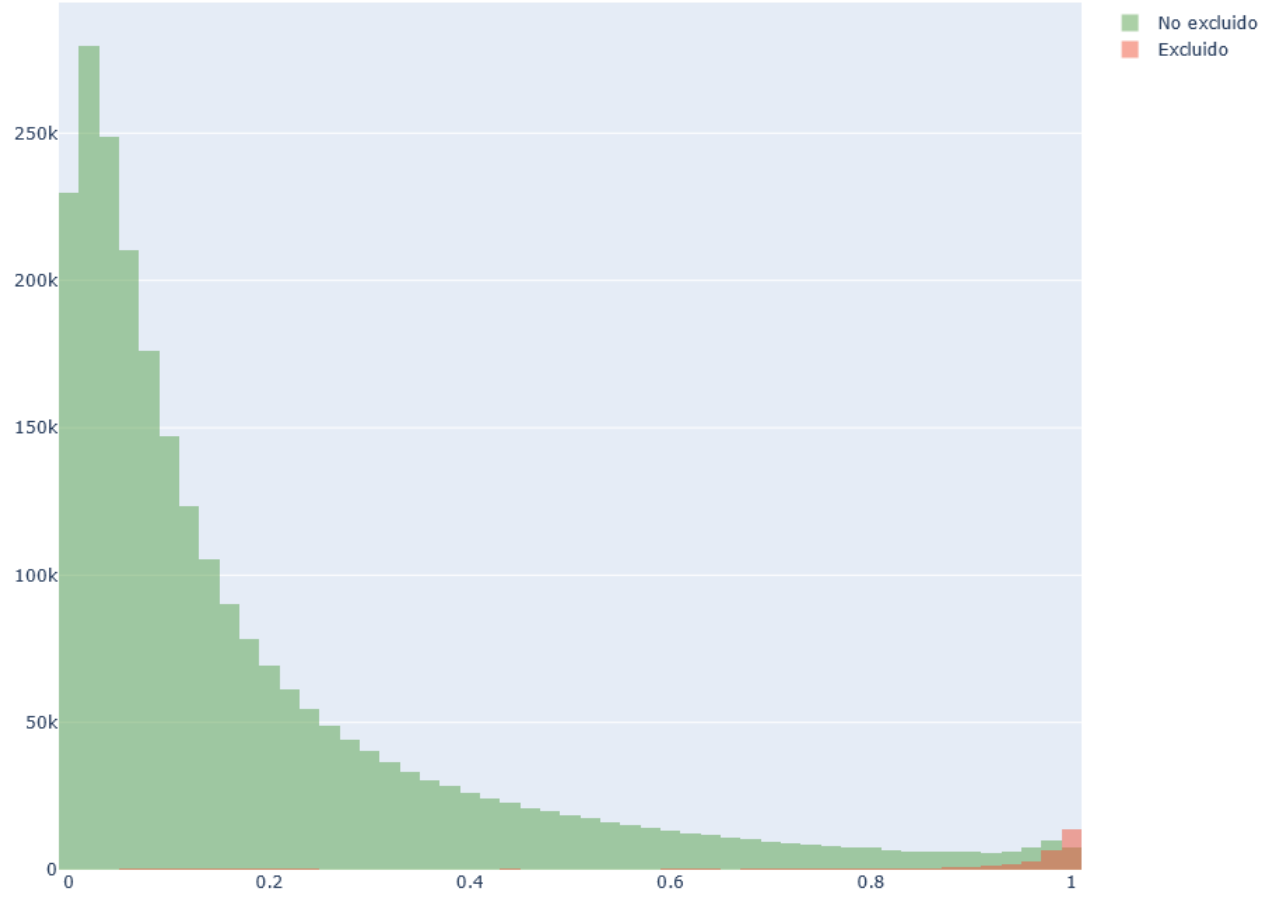
Predicción exclusión escolar

Métricas de desempeño: **umbral**



Predicción exclusión escolar

Métricas de desempeño: **umbral**



Predicción exclusión escolar

Ajuste modelo $t + 1$

$Exclusión_{t+1}$		<i>Precision</i>		<i>Recall</i>	
Algoritmo	<i>Accuracy</i>	No deserta	Si deserta	No deserta	Si deserta
<i>DecisionTree</i>	0,8228	0,8248	0,8208	0,8197	0,8258
<i>XGBoost</i>	0,8825	0,8630	0,9041	0,9092	0,8557
<i>CatBoost</i>	0,8849	0,8662	0,9055	0,9104	0,8594
<i>LightGBM</i>	0,8800	0,8607	0,9014	0,9067	0,8532

		Real	
		E	NE
Predicha	E	VP	FP
	NE	FN	VN

$$Accuracy = \frac{(VP+VN)}{\Sigma Total}$$

$$Precision = \frac{VP}{(VN+FP)}$$

$$Recall = \frac{VP}{(VP+FN)}$$

Importancia de las variables en la predicción

Variable	Importancia (%)	Variable	Importancia (%)
ULT_APR	17,629	TOT_MATR	2,450
ULT_COMP	10,949	EDAD_FIN	2,360
NIVEL_PARC	6,339	ULT_Q_DOC	2,325
Q_REPIT	5,038	ULT_PZ	2,191
ULT_PA	4,207	CHIL_ALU	2,160
SOBREEDAD	4,138	ULT_IDE	2,060
FEC_NAC_ALU	2,986	ULT_COM	1,806
ULT_ABN_NJ	2,965	AÑO_PARC	1,641
Q_ALU_MED	2,503	ULT_RBD	1,620

[Libro de códigos](#)

Predicción exclusión escolar

Ajuste modelo $t + 2$

<i>Exclusión</i> _{$t+2$}		<i>Precision</i>		<i>Recall</i>	
Algoritmo	<i>Accuracy</i>	No deserta	Si deserta	No deserta	Si deserta
<i>DecisionTree</i>	0,7210	0,7212	0,7208	0,7205	0,7215
<i>XGBoost</i>	0,8123	0,8084	0,8162	0,8185	0,8060
<i>CatBoost</i>	0,8152	0,8132	0,8172	0,8184	0,8120
<i>LightGBM</i>	0,8100	0,8069	0,8132	0,8151	0,8049

		Real	
		E	NE
Predicha	E	VP	FP
	NE	FN	VN

$$Accuracy = \frac{(VP+VN)}{\Sigma Total}$$

$$Precision = \frac{VP}{(VN+FP)}$$

$$Recall = \frac{VP}{(VP+FN)}$$

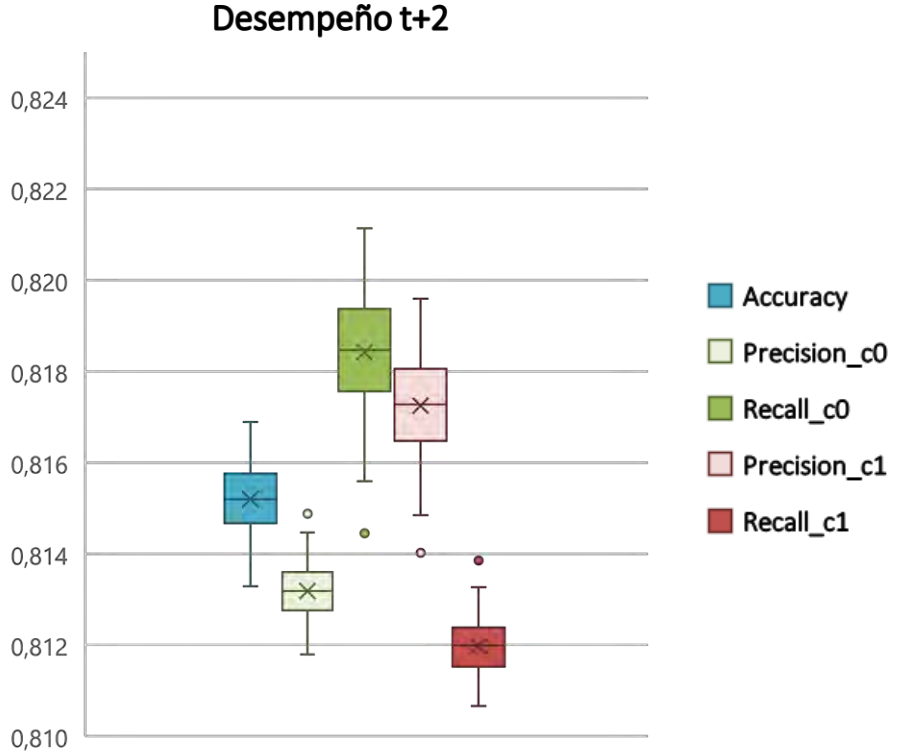
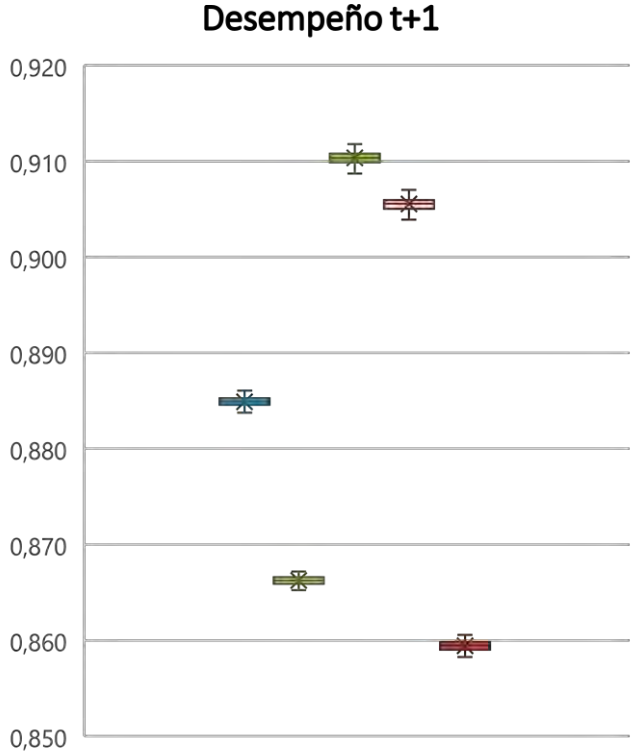
Importancia de las variables en la predicción

Variable	Importancia (%)	Variable	Importancia (%)
Q_REPIT	13,930	TOT_MATR	3,347
ULT_PA	8,448	ULT_IDE	3,331
ULT_PZ	6,019	CHIL_ALU	2,811
FEC_NAC_ALU	5,351	Q_ABN_NJ	2,703
NIVEL_PARC	4,705	ULT_Q_DOC	2,439
EDAD_FIN	3,821	ULT_COM	2,385
ABN_NJ	3,635	Q_ALU_BAS	2,069
ULT_APR	3,577	ULT_RBD	1,906
SOBREEDAD	3,374	PZ_TOTAL	1,875

[Libro de códigos](#)

Predicción exclusión escolar

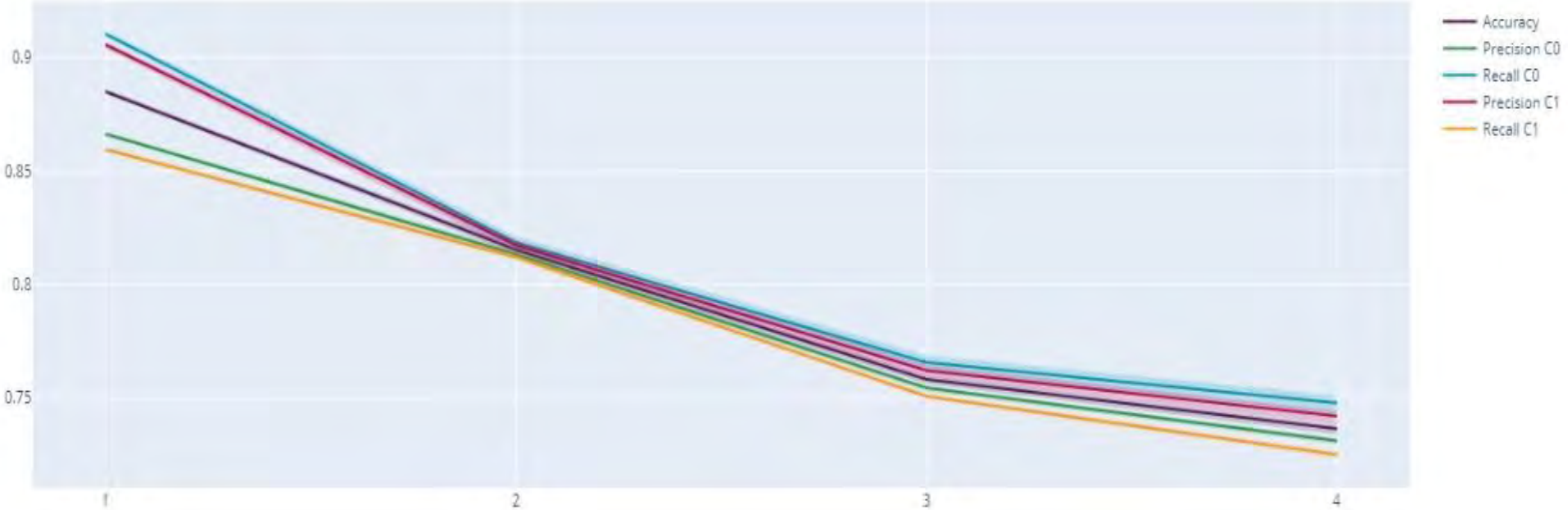
Varianza del desempeño



- Accuracy
- Precision_c0
- Recall_c0
- Precision_c1
- Recall_c1

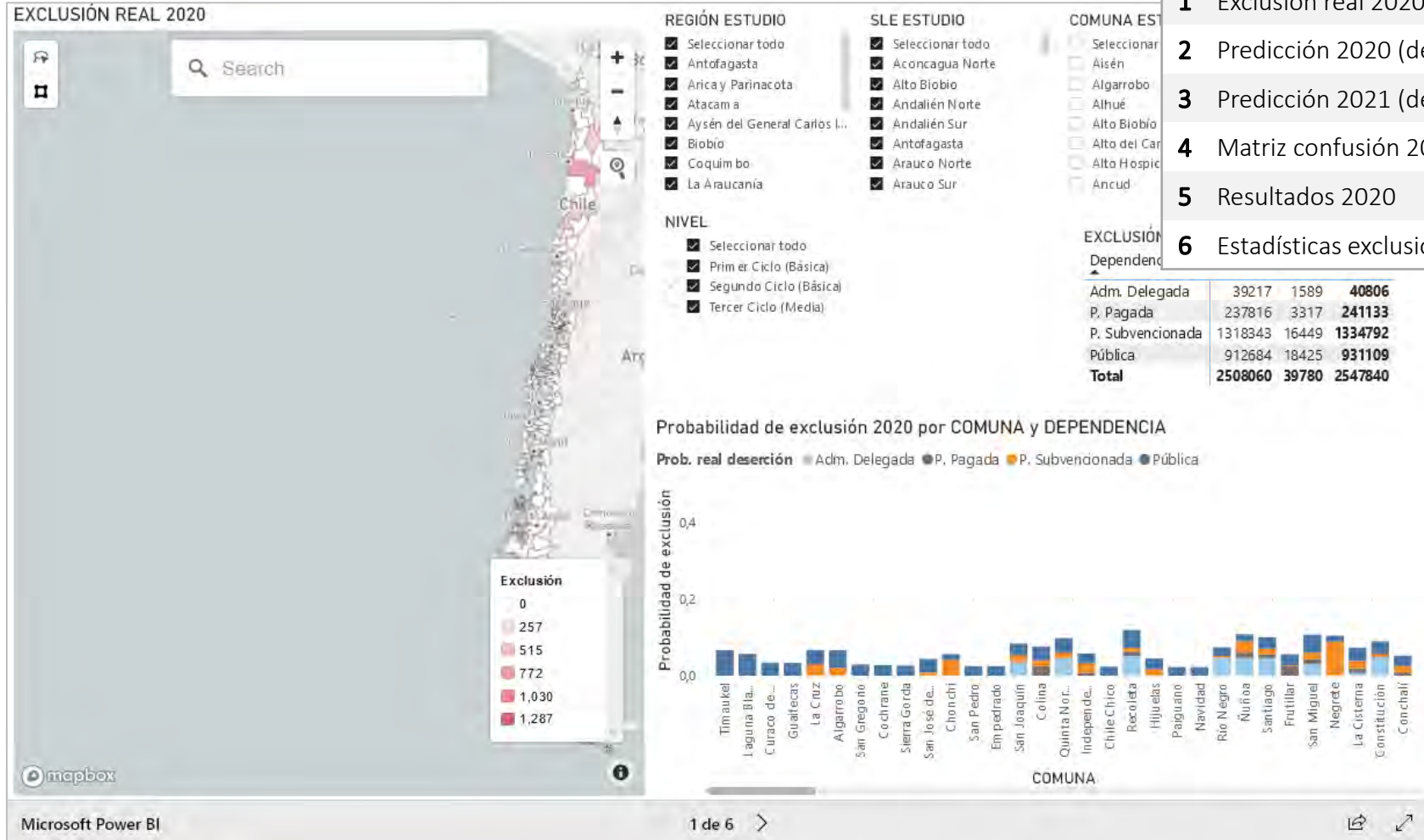
Predicción exclusión escolar

Varianza del desempeño



Predicción 2020 a partir de
2019

Predicción 2020



Hojas

- 1 Exclusión real 2020
- 2 Predicción 2020 (desde 2019)
- 3 Predicción 2021 (desde 2019)
- 4 Matriz confusión 2020
- 5 Resultados 2020
- 6 Estadísticas exclusión 2020

[Ir al dashboard](#)

2.547.840 trayectorias PowerBI