

Argentinos por la Educación:
¿Se puede predecir el desempeño en las pruebas
APRENDER?

Hackaton de Datos Educativos, 3ra Edición

Gianmarco Cafferata, Matias Scakosky, Nayla Nahmod,
Federico del Mazo, Agustina Luquez, Agustina Garcia Orsi

<https://github.com/matiscakosky/argxed>

Junio 2020

Índice

1	Introducción	2
2	Marco teórico	2
2.1	Marco educativo	2
2.2	Análisis exploratorio de datos	5
2.2.1	Variables aleatorias	6
2.2.2	Distribución de probabilidad	6
2.3	Machine Learning	7
2.3.1	Árboles de decisión	7
2.3.2	Gradient Boosted Trees	7
2.3.3	Shapley values	7
2.3.4	UMAP	8
3	Herramientas utilizadas	8
4	Datos utilizados	8
5	Análisis exploratorio	9
5.1	Variables de interés en la predicción	9
5.1.1	El uso del ponderador	9
5.1.2	Variables deprecadas del análisis	9
5.1.3	Variables de peso estadístico considerable	10
5.2	Índice socioeconómico medio	11
5.3	Smartphones	13
5.4	Repitencia primaria	15
5.5	ICSE	16
5.6	Gestión	18
6	Machine Learning	19
6.1	Construcción de árboles de decisión	19
6.2	Construcción de LightGBM	20
6.3	UMAP sobre escuelas	22
7	Conclusiones y orientaciones	27

1. Introducción

El análisis que detallamos a continuación se enmarca en el desafío propuesto acerca de las pruebas Aprender reuniendo los datos presentados por el Ministerio de Educación de la Nación entre los años 2013 y 2018 en todo el país.

Al reflexionar acerca de los factores que inciden en los resultados de las evaluaciones, debemos procurar analizar las evaluaciones en su contexto para evitar sesgos previos. Es por esto, que nos resulta importante remitir a los aportes teóricos que ponen en cuestión el escenario en el cual transcurren las pruebas Aprender.

Observamos que el objetivo de estas pruebas es mejorar la calidad educativa de nuestro país a través de la medición del desempeño de lxs estudiantxs, docentxs, directivxs y toda la comunidad educativa.

Nos resulta relevante destacar la importancia de que el análisis de resultados debe venir acompañado de una intención de mejora, en vistas a todos los factores, tanto aquellos que dependen del sistema educativo, como los que quedan enmarcados dentro de las políticas nacionales.

Es nuestra intención a lo largo de este análisis, desplazar el frecuente foco en lxs estudiantes al analizar los resultados de estas evaluaciones en especial desde los medios masivos de comunicación, que en ocasiones desprenden los datos de su contexto, haciendo difícil que se ponga el énfasis en las orientaciones para la mejora de la calidad educativa.

Para comprender en profundidad el objetivo de estas evaluaciones, nos remitimos a los aportes de Hebe Roig, quien retoma a Ernest House al explicar que de todo proceso evaluativo en la educación, es necesario que se desprendan orientaciones que fomenten los aprendizajes por parte de lxs actorxs involucradxs en la situación.

No es menor el énfasis que ponen estxs teóricos al referirse a la situación, como un escenario en donde se enmarcan las acciones que llevan adelante lxs actores del sistema educativo. En síntesis, cada situación denota una propia singularidad que es muy difícil de apreciar en evaluaciones estandarizadas que establecen criterios comunes para cada estudiante en cada rincón de nuestro país.

Es por esto, que nuestro interrogante mayor radica, no sólo en los factores que inciden en las evaluaciones, sino también en el uso de los datos y en los aprendizajes para la mejora que se alcanzan a través de los mismos. Para resumir, las evaluaciones sobre los aprendizajes no pueden tener como objetivo un informe descriptivo, sino que deben tener una invitación a poner el foco en la mejora para lxs actorxs que conforman el sistema educativo.

2. Marco teórico

2.1. Marco educativo

Retomando el desafío planteado, se nos propone analizar los factores que intervienen en los resultados de las pruebas aprender, la identificación de los factores endógenos y exógenos al sistema educativo y las posibles líneas de acción para los ministerios y las escuelas a raíz de lo analizado.

En cuanto a los factores, debemos mencionar en primer lugar al instrumento

de evaluación en sí, es decir, el examen que es tomado a lxs estudiantes como un factor endógeno que sin dudas determina en parte los resultados posibles. Cuando se establecen contenidos mínimos y formas de medir el aprendizaje en lxs estudiantes, se deja de lado el factor de que no en todas las instituciones vamos a encontrar la misma enseñanza, ni la misma infraestructura, ni la misma formación docente. Incluso estaríamos ignorando que no todxs lxs estudiantes tienen un mismo contexto e iguales tiempos de aprendizaje de los contenidos.

En las pruebas aprender, se destaca el conocimiento que pueda ser directamente medible por sobre las producciones de lxs estudiantes, notamos que con una modalidad de opción múltiple, se pretende medir el nivel de escritura de lxs niñxs, cuando nunca se les asigna realizar una producción escrita, se parte desde otros indicios que den cuenta de la adquisición de la escritura, la coherencia y la cohesión en los textos. Roig sobre esto explica que “La relación entre leer, escribir y aprender, que requiere ser cultivada desde la educación primaria, elude una visión integral de la alfabetización ya que la evaluación queda limitada a la comprensión lectora.” (Roig, 2019: 4)

Por otro lado y siguiendo en esta línea, en los últimos años ha variado el instrumento de evaluación, eliminando progresivamente las consignas abiertas y trayendo en consecuencia, un recorte de la currícula medible en estas pruebas. Felfeber y Puiggros dan cuenta de esto al mencionar:

“El Aprender no sólo utiliza exclusivamente ítems de respuesta cerrada, sino que reduce los contenidos de las pruebas a dos áreas: Lengua y Matemática, en el nivel primario y en 2/3 año del nivel secundario. Sólo se mantienen las cuatro áreas, Lengua, Matemática, Ciencias Sociales y Ciencias Naturales, en 5/6 año del nivel secundario, donde se aplica en forma censal. Posteriormente, en el Aprender 2017, las áreas evaluadas en 5/6 año del nivel secundario también se reducen a dos: Lengua y Matemática.” (Felfeber y Puiggros, 2018: 82)

En segundo lugar, pero en sintonía con lo ya expresado, es necesario observar el factor institucional, existen diversas posturas acerca de la implementación de estas evaluaciones en las escuelas debido a los fines a los que se destinan los resultados. Como ya mencionamos, no suelen estar orientados a la mejora, sino a la descripción y análisis, lo que en ocasiones fomenta la mediatisación. Por el contrario, también se pueden encontrar posturas centradas en la enseñanza para el examen, es decir, la enseñanza en las escuelas se orienta al aumento en el desempeño, por lo tanto, aquellxs estudiantes que aprenden para el examen, están más familiarizados con el instrumento de evaluación que aquellxs que abordaron los contenidos de una manera diferente. Esto se evidencia en los resultados de los exámenes y es frecuente que aquellas escuelas mejor puntuadas apoyen la implementación del Plan, mientras que las que no obtienen buenos resultados incluso lleguen a afectar negativamente la transparencia de los resultados mediante la ausencia del estudiantado por ejemplo.

En relación a lo que detallamos, no sólo las instituciones pueden llegar a variar en su manera de enseñar los contenidos, sino que también ha ocurrido un cambio importante en los libros de texto escolares, que variaron su enfoque priorizando contenidos que se evalúan en las pruebas estandarizadas, cuando en realidad las evaluaciones deberían medir los aprendizajes y no regularlos.

Al reflexionar sobre la participación docente y del resto de la comunidad educativa en el diseño del instrumento de evaluación, Felfeber y Puiggros exponen

que “a los docentes se los asume como objeto de la evaluación y no como sujetos de la misma, en tanto se los involucra sin incorporar una mirada superadora que incluya la dimensión político-pedagógica de todo el proceso de evaluación.” (Felfeber y Puiggros, 2018: 83) sobre esto ampliaremos luego pero no queremos dejar de mencionar la importancia de acortar la distancia entre lo que ocurren al interior de las aulas y la evaluación que lxs estudiantes completan y consideramos que incluir la mirada docente en todo el proceso de diseño, aporta a superar esta problemática.

Por otro lado, si bien el examen toma lugar en las instituciones, cabe aclarar que existen factores contextuales asociados como el nivel socioeconómico, el acceso a las nuevas tecnologías, a los servicios básicos, etc. Asimismo, las instituciones también son evaluadas bajo los mismos parámetros siendo diferentes los contextos de cada institución. Las teóricas a quienes ya mencionamos, destacan que la inclusión de este enfoque integral aporta una mirada más amplia de las condiciones en las que se producen los procesos de enseñanza, aunque a su vez, deja en evidencia la arbitrariedad de las pruebas estandarizadas.

Si bien todas estas cuestiones que detallamos, ponen en la mira las acciones propias que se tienen que llevar adelante desde el sistema educativo, es importante destacar que intervienen a su vez factores contextuales a cada estudiante en particular. Mencionamos esto como eje central, ya que al analizar los datos, no nos deja de llamar la atención la influencia del nivel socioeconómico y el acceso a la tecnología en el desempeño de lxs estudiantes en las evaluaciones. En el informe del 2016 sobre las pruebas Aprender que hace referencia al acceso a las tecnologías, se expresa la amplia brecha en el acceso entre aquellxs que asisten a escuelas de gestión estatal (donde un 55 % en el Nivel Primario y 64 % en el Nivel Secundario tienen acceso a Internet en sus hogares) y quienes asisten a escuelas de gestión privada (donde un 84 % en acceden a Internet en el Nivel Primario y un 88 % en el Nivel Secundario). Sobre esto ampliaremos luego, el aspecto considerable a remarcar, es que en los últimos años y teniendo en cuenta el desarrollo de las nuevas tecnologías, el acceso a Internet y a la tecnología se torna una necesidad básica. Lxs estudiantes no sólo acceden a la tecnología en las instituciones sino que también se relacionan con los contenidos abordados en la escuela a través de sus propios dispositivos. Sin ir más allá, en contexto de pandemia, se evidencian las diferencias en cuanto al acceso a los recursos por parte de todxs lxs estudiantes del país.

Otro factor que atraviesa de manera transversal al operativo Aprender son las normativas a nivel Nacional y Provincial. Como se ha mencionado en la introducción a los desafíos, el sistema educativo en nuestro país sigue las normativas nacionales fomentando la autonomía de las provincias en cuanto al presupuesto y la implementación de políticas de mejora. Desde los años 90', la Ley de Transferencia marca un hito central que se reafirma con la Ley Federal de Educación. La LFE fue la primera Ley Nacional que abarcó a todo el sistema educativo y allí se establece el Sistema Nacional de Evaluación de la Calidad marcando el paso de un Estado Docente a un Estado Evaluador, o en palabras de Pavigliantti (Felfeber y Gluz, 2018) a un “Ministerio sin Escuelas”.

En este sentido “El papel del Estado se presenta en forma ambigua: en algunos aspectos aparece como principal e indelegable y en otros como subsidiario o secundario otorgando un rol importante a las familias, la Iglesia y el sector

privado. Por otro, concentra las funciones de control en manos del gobierno central.” (Felfeber y Gluz, 2011: 4)

A su vez, ampliando en el marco normativo, la Ley de Educación Nacional que permite la creación del Consejo Federal de Educación posibilita la representación a nivel jurisdiccional e incorpora a las escuelas de gestión social y cooperativas. Así se reconocen las experiencias educativas diversas que incluyen a los movimientos sociales. En 2009, el CFE resuelve la elaboración de una agenda de trabajo que posibilite que todxs lxs niñxs y jóvenes del sistema cumplan la educación obligatoria y a su vez “logren una experiencia escolar valiosa que les permita el ejercicio pleno de su ciudadanía y el acceso a los bienes culturales disponibles” (Resolución n. 89/09).

Retomando en la importancia del acceso a las nuevas tecnologías, en 2010 el Plan Conectar Igualdad establece la entrega de netbooks a cada alumno y docente de educación secundaria de escuela pública, educación especial y de Institutos de Formación Docente; junto con el desarrollo de contenidos digitales que se utilicen en propuestas didácticas. Lamentablemente para el 2016 se observó un fuerte deterioro en la implementación de este Plan que luego fue retomado en 2020 de cara a un cambio de gestión a nivel nacional.

Especialmente en contexto de pandemia, esto garantiza la posibilidad de que más niñxs y jóvenes accedan a un dispositivo personal para continuar las clases en tiempo de Aislamiento Social, preventivo y obligatorio. En el contexto de surgimiento de este plan, se suman también el Plan Nacional de Educación Obligatoria y el Plan de Mejora Institucional para la Educación Secundaria, lo cual nos resulta importante al entender al sistema educativo en conjunto y no como variables disociadas que no reflejan la situación en la que se enmarca la problemática analizada.

Desde el año 2015, experimentamos la vuelta de la evaluación al centro de las políticas educativas, teniendo como prioridad la valoración de resultados de las pruebas estandarizadas. Las mismas fueron utilizadas como estrategia principal para justificar las transformaciones educativas, tendiendo a la meritocracia y la mercantilización. El Banco Mundial entre sus orientaciones, afirma que “las prioridades educacionales deben establecerse teniendo en cuenta los resultados, utilizando análisis económicos, estableciendo normas y midiendo los resultados a través de la evaluación del aprendizaje” (Banco Mundial, 1996:10).

Durante los últimos años se depositó en lxs docentes y estudiantes la responsabilidad del fracaso escolar y la deserción, tendiendo a la privatización y en consecuencia, la mercantilización del sistema educativo.

Esto puede notarse fuertemente en la creación del Instituto de Evaluación de Calidad y Equidad Educativa por el CFE, dedicado a la evaluación continua y periódica de los procesos y logros de aprendizaje, los proyectos y programas educativos, la formación y las prácticas de docentes, directivos y supervisores, las unidades escolares, los contextos socioculturales del aprendizaje y los propios métodos de evaluación.” (Art. 2). y la resistencia especialmente por parte de los sindicatos docentes a la implementación del mismo.

2.2. Análisis exploratorio de datos

Previo a la generación de cualquier modelo predictivo, es necesario hacer una interpretación de las variables que luego el modelo decidirá tomar y ejecutar una

predicción.

Este análisis permite entender de una forma estadística, cómo se comporta el modelo y por qué toma las decisiones que toma.

El análisis sobre las variables es una cuestión mera de estadística, por la cual, es importante ser cuidadosos con las inferencias que se realizan, ya que un mal uso de las variables puede inducir a inferencias incorrectas, los resultados de eso pueden derivar en una mejora errónea del sistema educativo o incluso la destinación de dinero en áreas que pueden no ser determinantes a la hora de mejorar el nivel de desempeño de las alumnas y los alumnos.

Para ello se presentan a continuación una serie conceptos que serán de utilidad para entender el análisis y la correlación con la sección que explican los datos y consecuentemente la parte de Machine Learning

2.2.1. Variables aleatorias

Se conoce en estadística a una variable aleatoria como una función que asigna un valor numérico a cada posible resultado de un experimento.

De forma intuitiva, una variable aleatoria puede tomarse como una cantidad cuyo valor no es fijo pero puede tomar diferentes valores como por ejemplo, un alumno que está por rendir la prueba APRENDER y será evaluado en una asignatura o más.

El resultado es incierto, hasta la realización del experimento, luego tenemos la distribución de los valores que puede tomar y consecuentemente su función de distribución.

Una distribución de probabilidad se usa para describir la probabilidad de que se den los diferentes valores.

2.2.2. Distribución de probabilidad

Las variables aleatorias tienen lo que se define como un soporte que es el conjunto de valores posibles que puede tomar, por ejemplo el soporte de una moneda es cara y seca y el soporte del resultado en matemática de la prueba APRENDER para un alumno cualquiera será: Por debajo del básico, Básico, Satisfactorio y Avanzado. A priori, lo que llamamos distribución de esta última variable aleatoria sería muy similar a los porcentajes de los alumnos que entraron en esas categorías, pero hay otras muchas distribuciones que se consiguen cuando uno tiene más información, las distribuciones condicionadas. Las distribuciones condicionadas son aquellas cuyas probabilidades se obtienen con más información que lo que sabemos a priori, por ejemplo, la distribución del desempeño de un alumno cualquiera dado que no repitió la primaria, o dado que además ganó las olimpiadas de matemática. Cualquier dato extra que tengamos puede darnos una nueva distribución.

Cuando dividimos los desempeños según categorías lo que estamos viendo son esas distribuciones condicionadas, y mucho de los procesos relacionados con machine learning tienen por objetivo encontrar distribuciones condicionadas que dado ciertos datos nos ayude a predecir para nueva información, muchas de estas búsquedas terminan en relaciones sumamente complejas que son difíciles de entender y visualizar.

Durante la primer parte, el análisis exploratorio, veremos aquellas distribuciones que más sencillas y fáciles de visualizar, pero como nuestro objetivo

es encontrar relaciones complejas, también entrenaremos modelos pero no para predecir sino con el solo objetivo de comprender que hicieron luego.

2.3. Machine Learning

2.3.1. Árboles de decisión

Un árbol de decisión es un modelo que intenta predecir una variable objetivo, en este caso la categoría, dividiéndose por preguntas binarias de Sí/No, por ejemplo: ¿Le alumnx tiene smartphone?. Al hacerse estas preguntas divide el set de datos de forma tal que pueda separar los grupos de las distintas categorías lo mejor posible.

Para saber qué división hacer y cómo hacerla usa una teoría matemática formulada principalmente por Claude Shannon con la intención de transmitir mensajes por circuitos electrónicos de forma eficiente. Para esto se define un número denominado *entropía de Shannon* que mide la cantidad de aleatoriedad o "desorden" (análogo al concepto de entropía en química) que tiene una variable.

Un árbol de decisión mide la aleatoriedad de todas las variables respecto del objetivo y elige aquellas que le saquen la mayor aleatoriedad posible a los datos que se tienen, que puedan separar el conjunto de datos de forma eficiente distinguiendo los de un tipo de otro.

2.3.2. Gradient Boosted Trees

Los árboles de decisión optimizados por el gradiente funcionan de la misma forma que los árboles de decisión tradicionales pero eligen por qué variable dividir con principios de otra rama de las matemáticas, la optimización. Utilizan una función de error en los datos sobre los que se quiere aprender y por medio de un *gradiente*, un concepto del análisis multivariable, logran elegir aquellos cortes que más chico hagan al error de predicción. Típicamente se construyen miles de estos árboles con distintas características y variaciones aleatorias y a la hora de predecir se los hace votar para obtener la mejor predicción.

2.3.3. Shapley values

Los shapley values son un concepto de un área de la matemática llamada teoría de juegos que tiene por objetivo entender cómo se comportan distintos agentes en un determinado "juego" con unas reglas definidas. Muchas interacciones de la vida entre agentes, muchas veces individuos, pueden modelarse desde la teoría de juegos, por ejemplo estrategias militares o modelos económicos.

Una de las cuestiones de interés en teoría de juegos son los *payouts*, esto es, en un "juego" donde los jugadores cooperan, una vez que los jugadores ganan o pierden algo, cómo saber cuánto aportó cada uno para pagarle su parte. Uno de estos métodos de *payouts* son los Shapley Values, a mayor Shapley Value tenga un jugador, mayor recompensa le corresponderá.

Estos shapley values se usan en modelos de Machine Learning para determinar cuánto hay que "pagarle" por su aporte a cada *feature*, por lo que si uno tiene un buen predictor, estos shapley values indicarán qué es aquello que el modelo considera más importante.

2.3.4. UMAP

UMAP es un algoritmo utilizado para realizar *embeddings*, esto es, representaciones numéricas en un espacio de los elementos de un set de datos con una distancia definida. Es un algoritmo novedoso que utiliza teoría de probabilidad y estadística, topología algebraica y teoría de categorías. Se suele utilizar para visualizar objetos que están en una cantidad de dimensión no visibles y verlos en dos dimensiones encontrando relaciones interesantes.

3. Herramientas utilizadas

El análisis exploratorio y los modelos desarrollados fueron realizados en Python utilizando como principales librerías las siguientes:

- **matplotlib** para la realización de visualizaciones
- **umap-learn** para realizar embeddings de datos
- **sklearn** para el entrenamiento de árboles de decisión
- **lightgbm** para el entrenamiento de gradient boosted trees
- **shap** librería para el uso de shapley values como medio de interpretación

Además se utilizó Adobe Illustrator para graficar un árbol de decisión y editar detalles en algunos de los gráficos de matplotlib.

4. Datos utilizados

Se utilizaron los siguientes archivos de tipo csv provistos para el desafío:

- app.alumno.csv
- app_departamento.csv
- app_escuela.csv
- app_nivel.csv
- app_provincia.csv

Uniéndolos en un solo csv se consiguieron las siguientes variables:

- Sexo del alumne
- Índice socioeconómico
- Nivel de desempeño en todas las asignaturas
- Ponderadores de todas las asignaturas
- Variables relacionadas con el acceso a TICs (notebook, pc, smartphone, etc.)
- Año del examen

- Nivel (primario o secundario)
- Ámbito (Rural o Urbano)
- Gestión (Privada o Estatal)
- ICSE: Índice de contexto social de educación, es un número entre 0 y 1 que indica privaciones que enfrentan lxs niñxs en su contexto, siendo 0 ninguna privación y 1 privación severa en todos los indicadores.
- ICSE categorizado: El ICSE dividido en categorías
 - Índice socioeconómico medio: el índice socioeconómico correspondiente medio correspondiente a la escuela
 - Índice socioeconómico medio categorizado
- Escuela técnica: indica si el alumnx está en una escuela técnica
- Si la escuela tiene internet
- Subvención de la escuela
- Dependencia
- Provincia
- Departamento en la provincia
- Tasa con NBI: Tasa de necesidades básicas insatisfechas para la escuela

5. Análisis exploratorio

5.1. Variables de interés en la predicción

5.1.1. El uso del ponderador

Para poder medir correctamente las distribuciones de probabilidad no se realizó análisis unitario de cada alumnx por igual, ya que este indicador muestra que según diferentes parámetros, su desempeño y respuesta a cada pregunta es tan valiosa como indique la medida ponderada, de esa manera podemos asegurarnos de no solo contemplar las respuesta de cada alumnx, sino que además podemos ofrecer a la distribución una medida ponderada de su desempeño, extrapolando así a posibles otras respuestas de alumnxs parecidos en características a quien el ponderador resultó un número mayor a 1.

5.1.2. Variables deprecadas del análisis

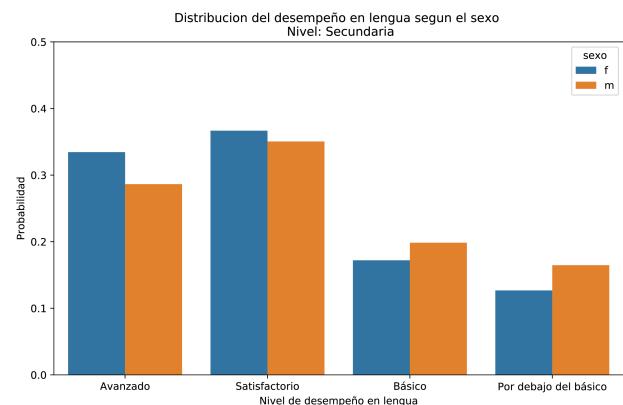
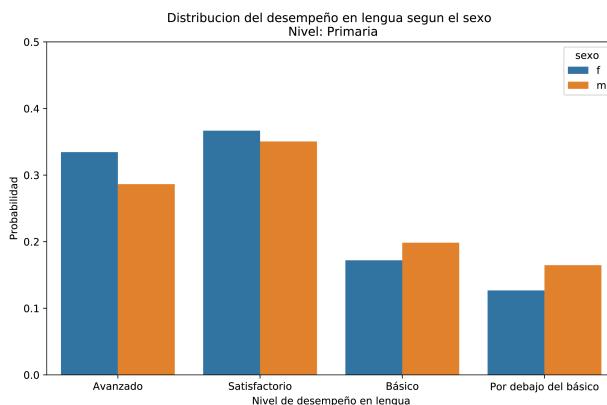
Se decidió hacer un análisis de aquellas variables binarias y no binarias, cuyas distribuciones no presenten una probabilidad de desempeño cercana a la distribución uniforme, donde cada variable aleatoria sólo puede tomar valores comprendidos entre dos extremos, siendo que todos los intervalos discretos tienen la misma probabilidad.

Informalmente esto significa que cada una de estas variables no hacen la diferencia a la clasificación del desempeño alumnx. Es más, algunas distribuciones

son tan cercanas al 50 % en features binarias que merece la pena preguntarse si conviene seguir preguntando acerca de estas variables en futuras evaluaciones. A continuación, se verán ejemplos de estas variables.

- **Sexo** Se encontró que la probabilidad de encontrar que los desempeños en lengua marcados según aquellxs que se identificaron con alguno de los géneros “m” o “f” no varía para ninguno de los niveles, en el gráfico podemos observar el caso de primaria, el de secundaria es muy parecido.

A nivel estadístico, conocer el sexo del alumnx no genera un impacto muy grande para inclinar la balanza hacia alguna respuesta en concreto.



- **Tiene cable** De la misma forma otro ejemplo de variable que dados los datos se encontraron distribuciones parecidas, es con la feature de si el alumnx tiene cable o no.

Una pequeña tabla con las probabilidades ponderadas puede explicar esto, idénticamente al ejemplo anterior se tomó alumnxs de primaria y secundaria, y se midió su desempeño en lengua.

	No tiene cable	Tiene cable
Avanzado	20 %	24 %
Satisfactorio	39 %	40 %
Básico	24 %	20 %
Por debajo del básico	17 %	16 %

5.1.3. Variables de peso estadístico considerable

El apartado describe las distribuciones de probabilidad de aquellas variables que encontramos que afectan de mayor forma al desempeño de lxs estudiantes de ambos niveles.

Es importante aclarar que estas inferencias se hicieron previas a la corrida del modelo, que luego, terminó de alguna manera de confirmar que estas variables son realmente importantes para el desafío propuesto, pues utilizando en alguna

proporción adecuada una combinación no lineal de los valores de estas variables, logramos determinar resultados de interés.

A continuación cuáles son estas variables, sus distribuciones y algunas suposiciones que permiten explicar por qué estas variables tienen peso.

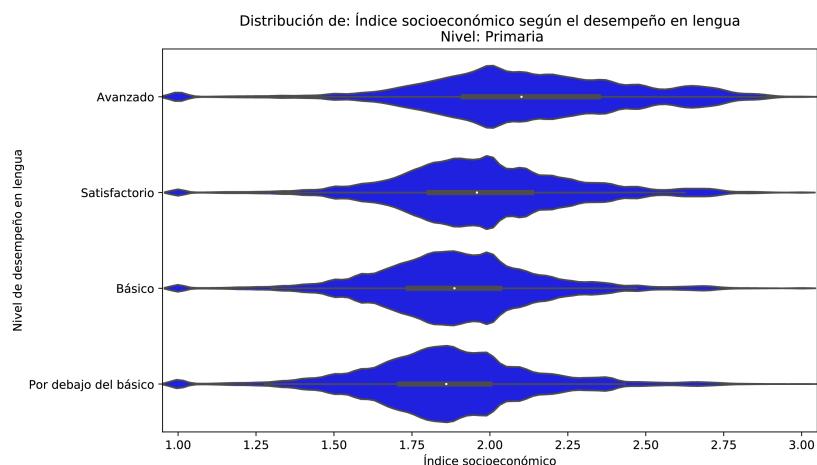
5.2. Índice socioeconómico medio

El índice socioeconómico pondera y permite clasificar a lxs estudiantes bajo ciertos parámetros cuan establecidx se encuentra respecto de alguna posición económica, y/o estrato social. Para medirlo se establecen diferentes variables que componen el índice, basándonos en los informes de las pruebas Aprender 2016, el índice socioeconómico está compuesto por:

1. Nivel educativo de los padres.
2. Hacinamiento en el hogar (relación entre la cantidad de habitaciones de la vivienda en la que habita el estudiante y el número de miembros del hogar).
3. Recepción de la Asignación Universal por Hijo (AUH) en el hogar.
4. Tenencia de equipamiento informático en el hogar (Internet, consolas de videojuegos, televisión y celular)

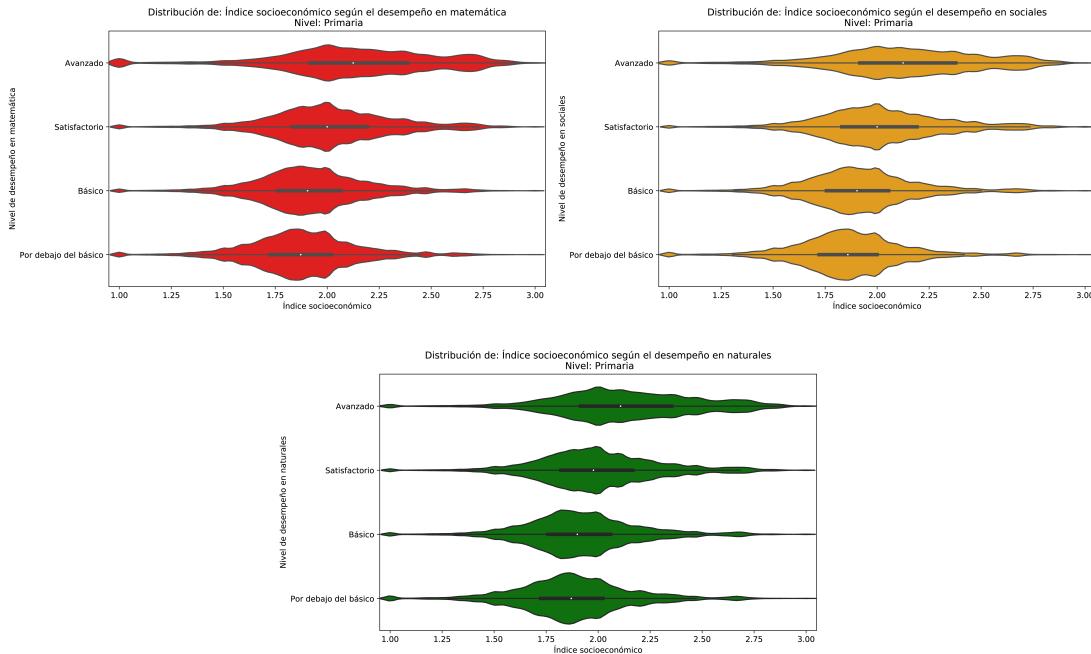
Es importante aclarar que esta variable es continua, por lo que utilizamos un gráfico en violín para mostrar los resultados. Ésta es una forma muy atractiva de mostrar la distribución de un conjunto de datos con respecto a una variable categórica, hay que tener en cuenta que, para conjuntos de datos pequeños, la estimación de densidad puede no transmitir con detalle el perfil real de los datos, mostrándolo mucho más suave de lo que realmente es, aunque considerando el tamaño del dataset, podemos afirmar que es una aproximación de la realidad muy buena.

Como se observa se puede agrupar según los desempeños medidos por nivel, y encontramos los primeros resultados interesantes de esta variable.



El gráfico muestra cómo se agrupan alrededor de la mediana, los diferentes índices socioeconómicos de todxs lxs estudiantes, podemos observar que a medida que crece el índice socioeconómico encontramos grupos mayoritarios de personas que van obteniendo mejores calificaciones.

Podemos observar que lo mismo ocurre para el nivel primario en matemática.



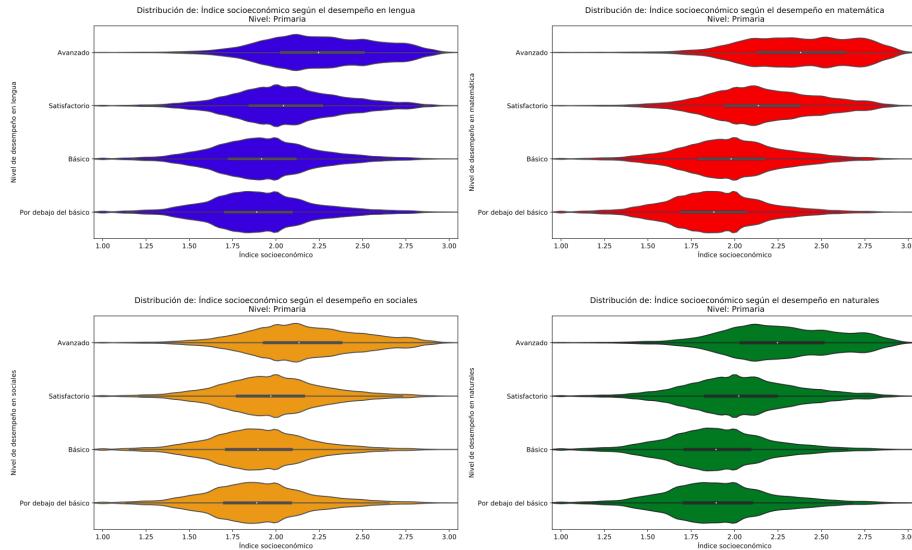
Otra vez, a mayor índice socioeconómico, la probabilidad de estar en un grupo de mayor nivel de desempeño es más grande, mientras se reduce la probabilidad de estar en un grupo de desempeño menor.

En secundario, observamos un comportamiento parecido.

Si prestamos atención a cómo se reparte el índice socioeconómico, podemos ver puntos de acumulación mas grandes a medida que leemos el gráfico de izquierda a derecha y de abajo hacia arriba

Estos puntos de acumulación representan cantidades de estudiantes, que conservan un índice socioeconómico superior.

Podemos observar el comportamiento se mantiene en primaria y en secundaria con una distribución muy parecida en el gráfico. Esto nos permite afirmar que es un parámetro realmente importante, que conocido su valor, nos da una idea para dónde orientar nuestra clasificación.

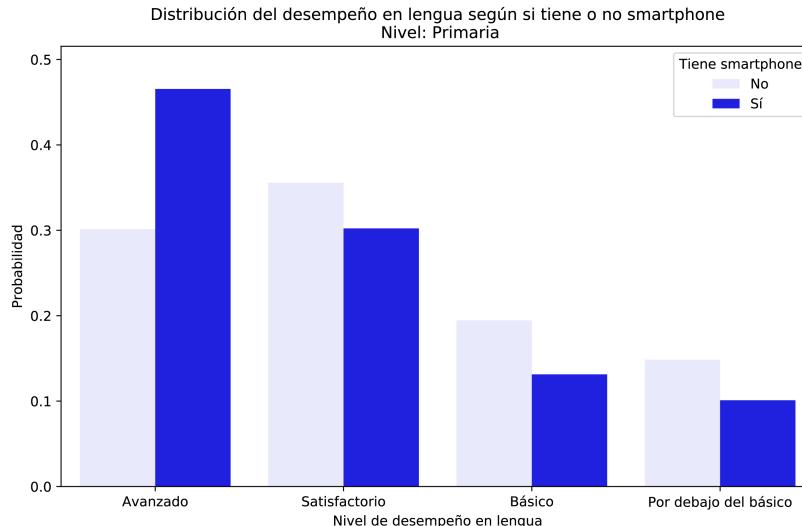


5.3. Smartphones

Un resultado interesante aparece cuando se analiza la tenencia de celulares smartphone por parte de lxs alumnxs. debido a que de alguna manera el acceso a estos dispositivos genera un impacto positivo en algunos de los resultados de las pruebas.

Para hacer un análisis fiel respecto de los datos, y manteniendo un nivel de rigurosidad estadística, cada uno de estos datos debido a la varianza de respuesta por nivel, se analiza por un lado primaria y por otro secundaria.

En el siguiente gráfico se observa la distribución de probabilidad de desempeño dado un alumnx que rindió la asignatura lengua en primaria.

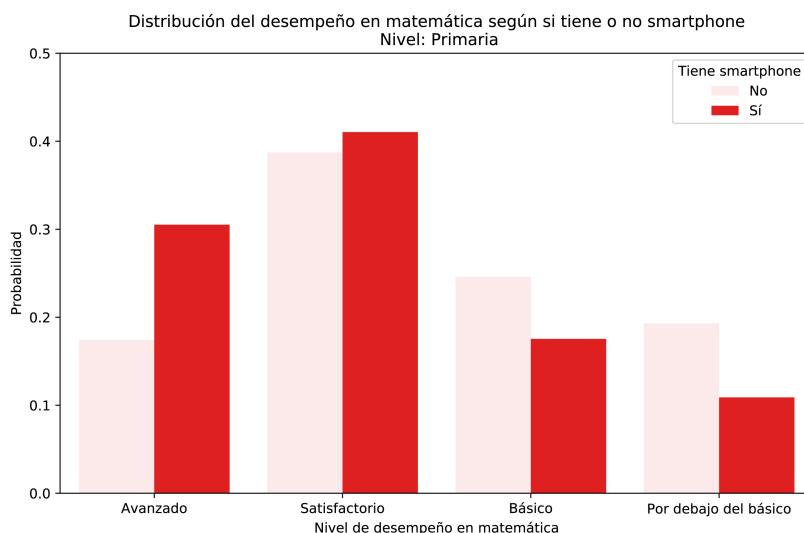


Este primer resultado es bastante interesante, puesto que encontramos que la

probabilidad acumulada entre “Avanzado” y “Satisfactorio” es cercana al 80 %, si nos encontramos en el caso de que el alumnx tiene smartphone.

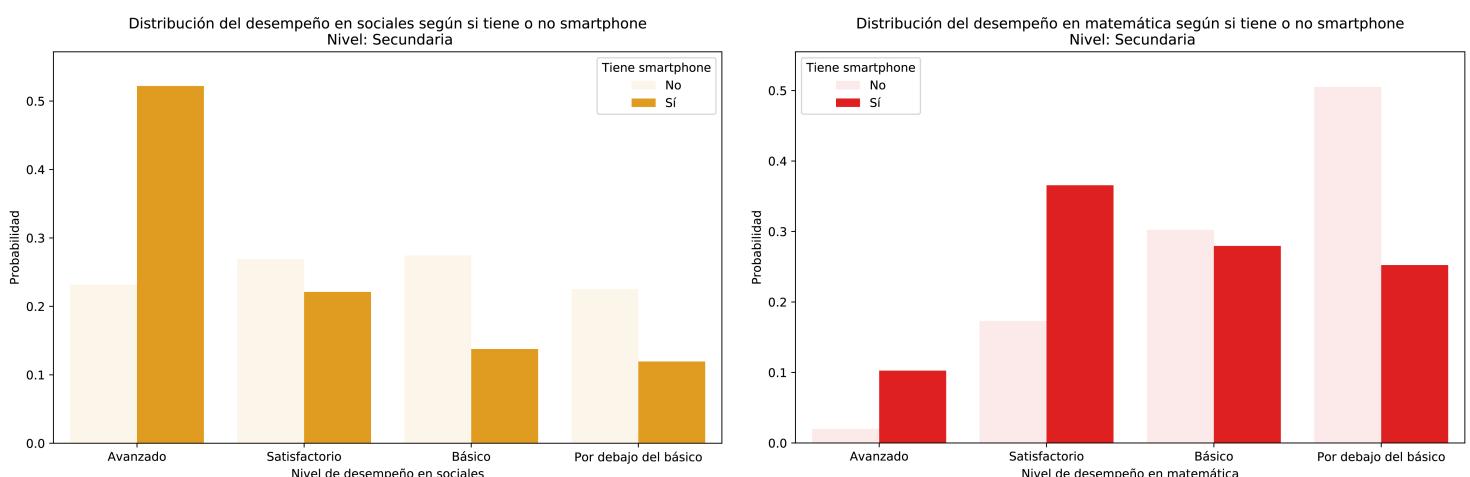
Se pueden hacer varias preguntas al respecto.

Para matemática la situación es menos marcada, puesto que las probabilidades de que si el estudiante tiene smartphone superan en desempeño a los que no, el resultado apunta más a un resultado satisfactorio que avanzado.



En secundaria el comportamiento de esta variable es realmente sugerente, ya que según los siguientes gráficos, el desempeño de los chicxs, se clasifica más cerca de los extremos de los desempeños.

Observemos los casos de ciencias sociales y matemática en secundaria según la variable **tiene smartphone**.



En el caso de sociales, tener un celular acumula casi el 50 % de probabilidades de obtener un “Avanzado”.

Pero vemos que al no tener un celular, para esta asignatura, no es determinante para la clasificación.

El caso inverso ocurre con matemática, no tener smartphone resulta en una distribución de la probabilidad de de desempeño, muy marcada donde casi un 50 % de lxs alumnxs van a resultar por debajo del básico.

De la misma forma, que el simple hecho de no tener un dispositivo smartphone, en secundaria, descarta completamente la posibilidad de que algún alumnx obtenga un “Avanzado” en matemática.

Surgen algunas cuestiones. Lo primero que es necesario aclarar es que el hecho de tener o no tener smartphone no está desligado del índice socioeconómico medio del alumnx por lo que de ninguna manera podemos afirmar ningún tipo de independencia entre estas variables.

Sin embargo, el análisis de esta variable nos abre un nuevo interrogante. ¿Ayuda el smartphone una herramienta de aprendizaje más allá del sistema educativo?

Por supuesto que resulta trivial que un estudiante no va a mejorar su desempeño por el simple hecho de tener o no un smartphone. Pero en contextos donde la tecnología y la innovación están en constante avance, y por lo mencionado en la sección desarrollo acerca que cada vez más, lxs estudiantes hacen uso de sus propias tecnologías, para mejorar su propio proceso educativo, no podemos evitar pensar que es cuestión de tiempo encontrar a esta variable como una de las más descriptivas en el proceso de clasificación de desempeño.

Hoy en día, donde nos encontramos en tiempos de pandemia, el acceso a la información resulta un elemento clave no solo para un desarrollo educativo/profesional sino, como un derecho de necesidad básica que según podemos observar gracias a los datos, todavía no está siendo otorgado de manera uniforme y homogénea a los chicxs de nuestro país. Lo que dificulta, en retrospectiva, al desarrollo educativo del alumnx.

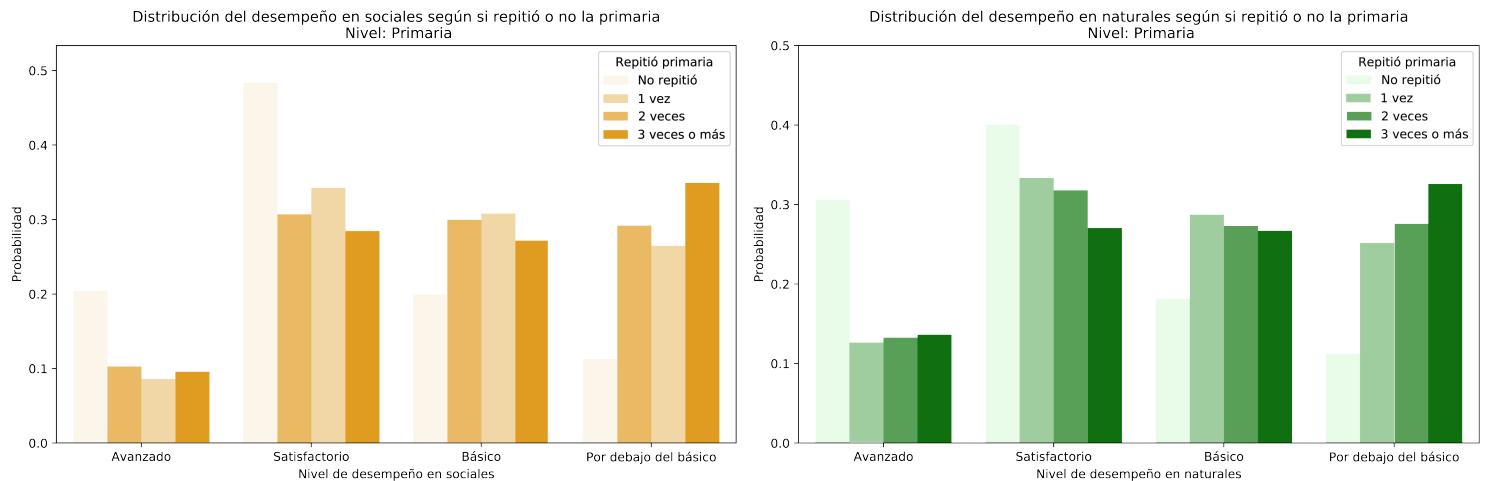
5.4. Repitencia primaria

En estos gráficos podemos observar la distribución de la probabilidad de obtener una calificación según si el alumnx no repitió nunca, repite una o más veces algún año de la primaria.

De todos los datos acerca de repitencias solamente tiene valor las respuestas de alumnxs de primaria debido a la gran cantidad de “No contesta” que se hallaron en secundaria.

	Repitio primaria	Repitio secundaria
Alumnx primaria	12 % no contestó	No aplica
Alumnx secundaria	91 % no contestó	92 % no contestó

Esta distribución de alumnxs que no respondieron esta pregunta llama particularmente la atención. Lamentablemente no se dispone de fuentes de datos que permitan entender los factores por los cuales esta pregunta es tan poco respondida entre lxs estudiantes.



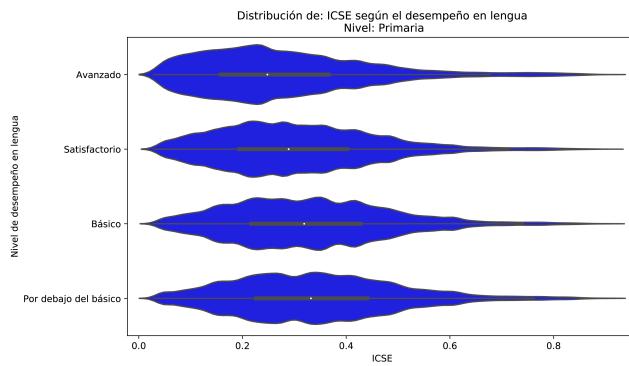
Los resultados de estas distribuciones son sugerentes, debido a que presenta una diferencia muy grande en las materias observadas el hecho de haber repetido, al menos una vez, la primaria, para la determinación de la calificación.

El análisis de esta variable se ve complejizado debido a que lxs estudiantes a medida que van repitiendo, comienzan a desertar del sistema educativo, podemos afirmar entonces, que esta variable se ve sesgada por el instrumento evaluativo en el que se recolectó el dato.

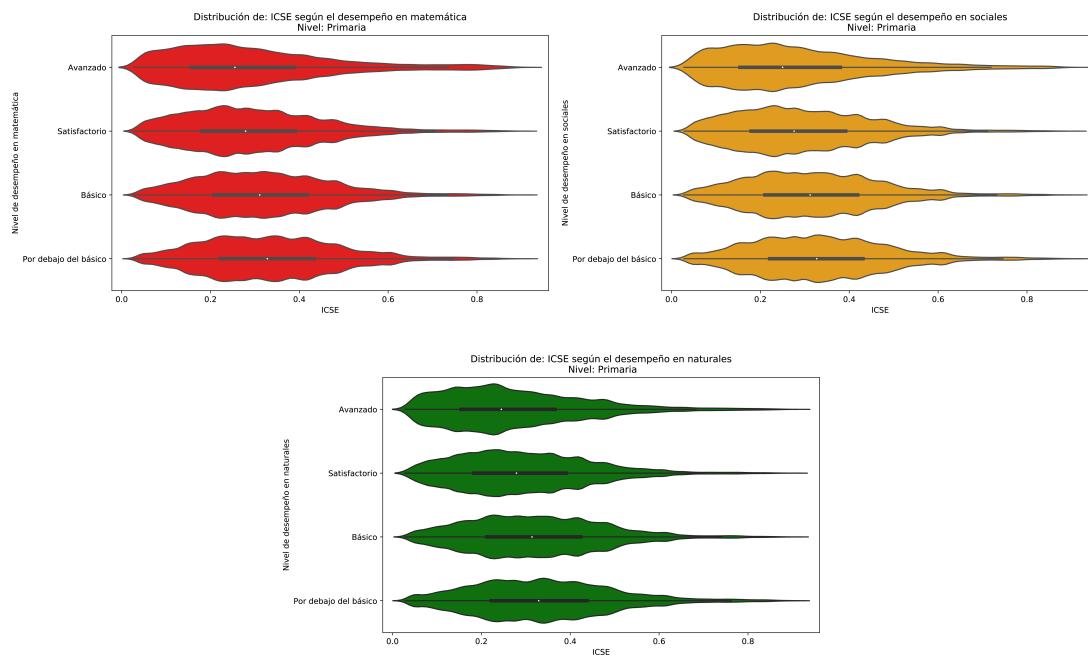
5.5. ICSE

El ICSE mide el índice de clasificación de unidades geográficas como función de las condiciones de vida y/o vulnerabilidad de aquellas casas donde residen estudiantes. Según su definición se construye de forma multidimensional con diferentes variables entre las que tenemos, características de vivienda, acceso agua, nivel de formación de adultos residentes en el hogar y otro tipo de factores que son desconocidos en estos dataset.

Los resultados arrojados por cómo se distribuye la probabilidad de desempeño en las Aprender cortada por ICSE resultó interesante, debido a que analizando ambos niveles por separado obtuvimos resultados inversos a los obtenidos por el índice socioeconómico.



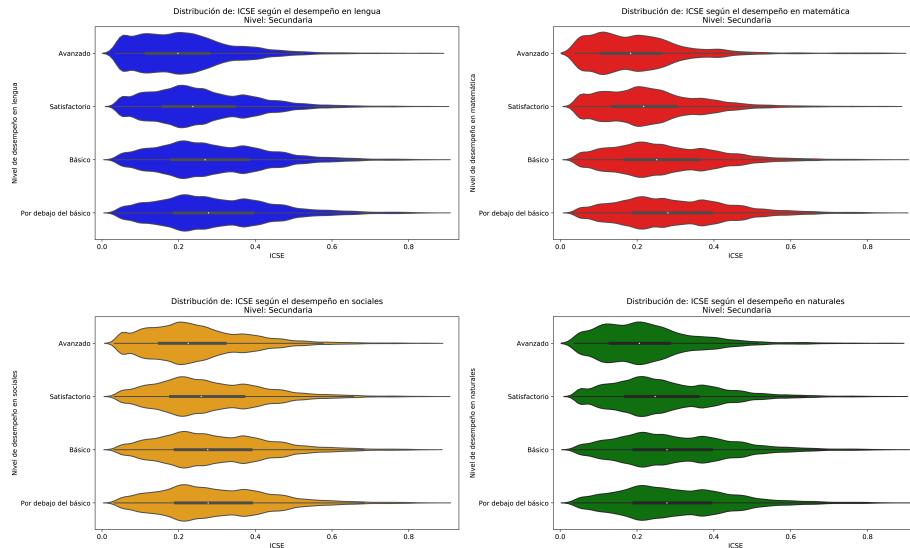
Podemos ver que en primaria, la mediana, aparece más a izquierda mientras que nos acercamos a los desempeños más avanzados, lo que significa que el ICSE es menor a medida que crecemos en desempeño, este gráfico lo demuestra para primaria en lengua, pero el mismo patrón se puede observar para el resto de las materias en este nivel.



Se puede resaltar que las distribuciones tienen diferentes acumulaciones dependiendo la materia, esto significa que si bien existe una correlación entre cómo se ordenan las medianas de forma decreciente en el gráfico, las observaciones principales que podemos hacer es cuán suave es la curva llegando a los extremos del violín.

Podemos observar que en el extremo izquierdo, la calificación de por debajo del básico acumula muy pocos o en algunos casos, como ciencias naturales ningún ICSE cercano a cero generando una punta que no contiene valores bajos del indicador. El caso inverso se puede observar en los Satisfactorio/Avanzado.

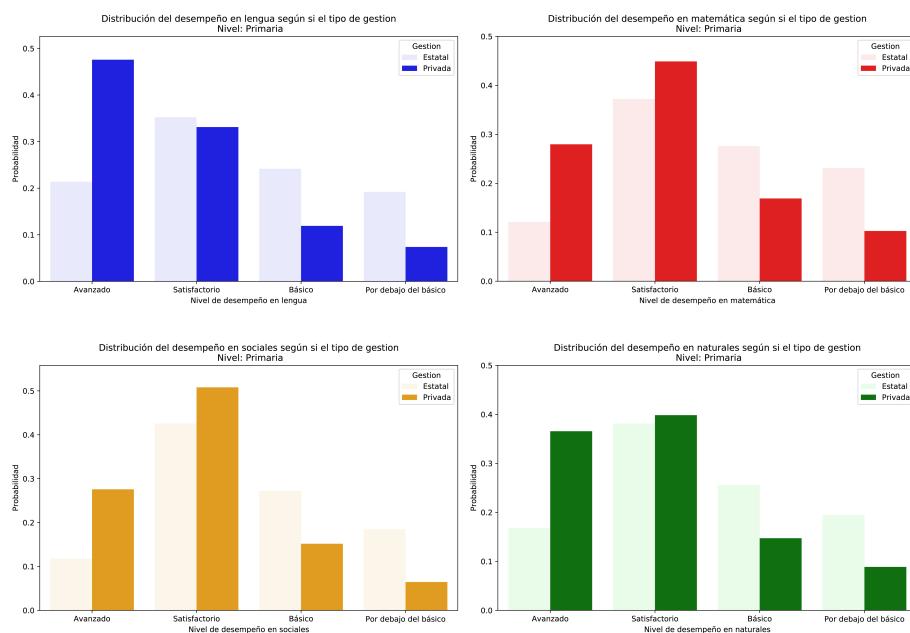
En secundaria esta idea continúa.



5.6. Gestión

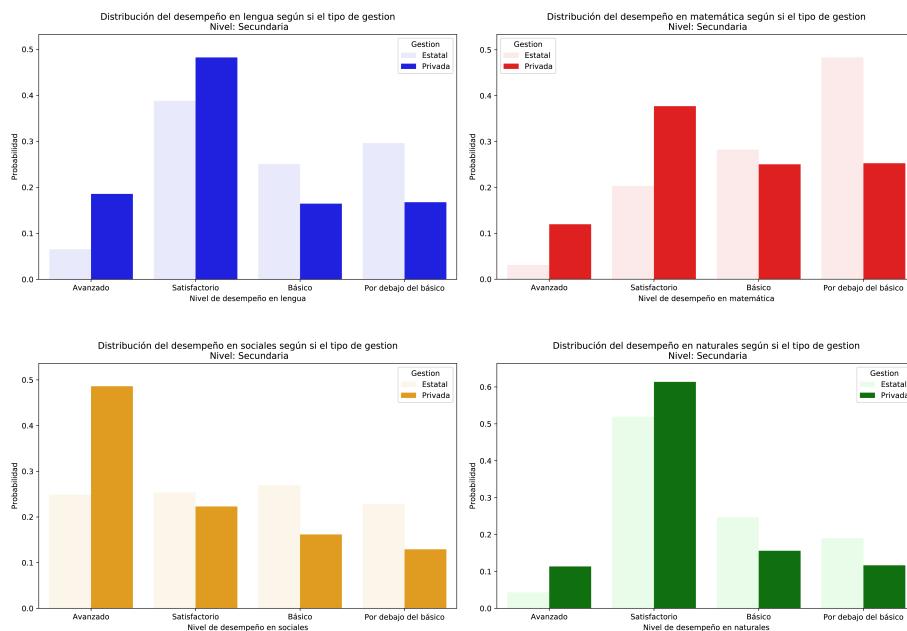
El análisis de la gestión resultó una variables muy influyentes para determinación de una clasificación. Podemos observar en los resultados de los niveles primarios por materia

Los exámenes de lengua, son una muestra interesante para el corte de esta variable, pues la distribución de la probabilidad concentra gran parte de la probabilidad de encontrar a un alumno entre los niveles de “Avanzado” o “Satisfactorio”.



Esta variable resultó de muchísimo interés ya que conocer el valor para un alumnx de primaria permite concentrar gran parte de las probabilidades en un subconjunto del desempeño general en primaria

En secundaria obtenemos los siguientes resultados.



6. Machine Learning

6.1. Construcción de árboles de decisión

Como primer modelo interpretable decidimos construir árboles de decisión de profundidad 4 y graficarlos para entender de forma sencilla cuáles eran las variables que más influyen.

Producto de la gran disparidad de distribuciones exhibidas en el análisis exploratorio, alcanza con un árbol de hasta 4 niveles para separar grupos en donde el 40% o más está por debajo del básico de otros grupos en donde el 40% está en el nivel avanzado. A modo de ejemplo tenemos como referencia los árboles de lengua para la secundaria¹ y matemática para la primaria² y la secundaria³ pero resultados similares se replican en todas las asignaturas y todos los niveles con un modelo muy sencillo y gran interpretabilidad.

Entre las variables destacables para la mayoría de los árboles: El índice socioeconómico, el ICSE, la tasa con necesidades básicas insatisfechas, acceso a

¹https://github.com/matiscakosky/argxed/blob/master/plots_mejorados/Arbol_Lengua.pdf

²https://github.com/matiscakosky/argxed/blob/master/plots_mejorados/Arbol_Mat_Primaria.pdf

³https://github.com/matiscakosky/argxed/blob/master/plots_mejorados/Arbol_Mat_Secundaria.pdf

TICs (Smartphone, Smart TV, Internet, etc.), si la escuela tiene internet y si lxs alumnxs repitieron ya sea la secundaria o la primaria.

Este modelo consideró al ponderador duplicando los puntos del set de datos para alumnos cuyo ponderador fuera mayor a 1, por ejemplo si el ponderador es 3.47 el dato de ese examen se repetirá 3 veces mínimo y una vez más con probabilidad 0.47. Esto cambia las probabilidades condicionadas y es efectivo al influir la *entropía de Shannon*.

6.2. Construcción de LightGBM

LightGBM es un modelo de *gradient boosted trees*, construimos un árbol para predecir el desempeño en cada asignatura para los dos niveles primario y secundario. Con estos modelos logramos nuestras mejores predicciones, que para el problema que estamos enfrentando y su complejidad, consideramos interesantes. Este modelo consideró al ponderador duplicando los puntos del set de datos para alumnos cuyo ponderador fuera mayor a 1, por ejemplo si el ponderador es 3.47 el dato de ese examen se repetirá 3 veces mínimo y una vez más con probabilidad 0.47. Como este modelo optimiza por medio de *gradientes* las cantidades sirven como ponderadores de los puntos en el entrenamiento.

Una matriz de confusión es una grilla de dos entradas que nos permite ver para una categoría determinada cuáles fueron las clases a las que nuestro modelo dijo que pertenecían. A continuación mostraremos algunas de las matrices que resultaron de los distintos modelos entrenados.

		Verdad			
		Por debajo del básico	Básico	Satisfactorio	Avanzado
Predicción	Por debajo del básico	0.40	0.27	0.14	0.05
	Básico	0.32	0.38	0.24	0.10
	Satisfactorio	0.22	0.29	0.42	0.36
	Avanzado	0.05	0.06	0.21	0.50

Figura 1: Matriz de confusión para el modelo que predice el desempeño en ciencias naturales en el nivel secundario normalizada por el valor de verdad.

En esta matriz podemos observar por ejemplo para los alumnxs con desempeño por debajo del básico (en su respectiva columna) la cantidad de ellos predichos por nuestro modelo en las distintas categorías, nuestro modelo predice de ellxs al 41 % correctamente en como “Por debajo del básico”, al 29 % de

ellxs como “Básico”, otro 29 % de ellxs como “Básico” y al 1 % como avanzados. Podemos ver que si bien nuestro modelo confunde a aquellxs alumnxs cuyo desempeño fue por debajo del básico con desempeños básicos y satisfactorios es muy difícil que confunda estos con un desempeño avanzado. Algo similar ocurre para aquellxs con desempeño avanzado ya que solo el 6 % de ellxs es clasificado erróneamente como “Por debajo del básico”.

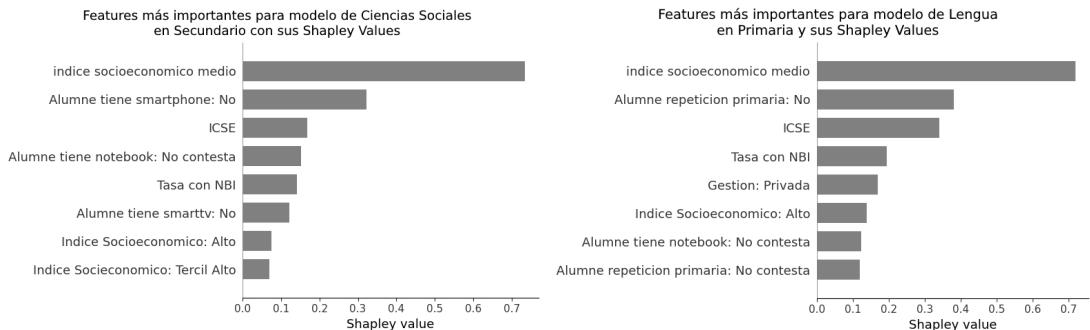
Algo muy similar ocurre para el resto de los modelos construidos por cada nivel y asignatura, he aquí algunos ejemplos de matrices de confusión para los otros modelos:



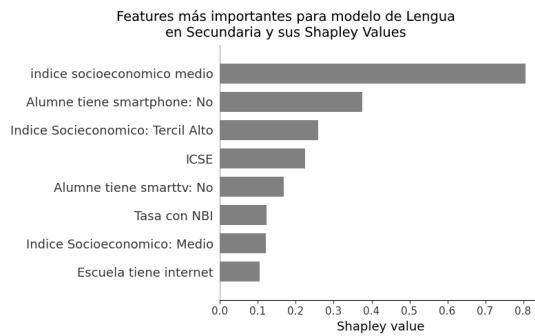
(a) Matriz de confusión para el modelo que predice el desempeño en ciencias sociales en el nivel primario normalizada por el valor de verdad (b) Matriz de confusión para el modelo que predice el desempeño en matemática en el nivel secundario normalizada por el valor de verdad

Podemos observar resultados similares en todos los modelos construidos, por lo que creemos que ante la pregunta: ¿Se puede predecir los resultados de las pruebas aprender? Una posible respuesta podría ser, se puede más de lo que se debería poder con los datos que tenemos. Utilizando sólo variables solo variables relacionadas al contexto de lxs estudiantes por fuera de las instituciones educativas, somos capaces de separar de forma muy precisa aquellos alumnos cuyo desempeño será avanzado de aquellos cuyo desempeño será por debajo del básico, algo que llama la atención ya que desde las políticas públicas, se destacan los valores de equidad e igualdad de oportunidades.

A continuación nos preguntamos, ¿Qué es lo que observa nuestro modelo para lograr estas predicciones? Por lo que recurrimos a los *shapley values* que asignan un valor de importancia a cada uno de los aspectos observados por el modelo.



(a) Magnitud del shapley value para las variables de mayor importancia observadas por el modelo que predice el desempeño de Ciencias Sociales en el nivel secundario
(b) Magnitud del shapley value para las variables de mayor importancia observadas por el modelo que predice el desempeño de Lengua en el nivel primario



(c) Magnitud del shapley value para las variables de mayor importancia observadas por el modelo que predice el desempeño de Lengua en el nivel secundario

Se observa que casi todo lo que nuestros modelos utilizan para predecir está relacionado a las variables previamente descritas en la introducción y en análisis. Todos los modelos construidos se explican con variables similares.

6.3. UMAP sobre escuelas

Al tener varias muestras sobre las mismas escuelas podemos estudiarlas a visualizar a modo de agregado, UMAP va a intentar llevar puntos en alta dimensión a puntos en dos dimensiones para poder ser observados. Lo que se hizo en este caso es, ya que los exámenes se dividen entre los de lengua y matemática y los de ciencias sociales y ciencias naturales tomamos aquellas escuelas cuya muestra consideramos más representativa (de 40 muestras o más, que coincide con la mediana de la cantidad de muestras para cada escuela, quedándonos con 15 mil escuelas) y calculamos para cada una la probabilidad de los desempeños en las dos asignaturas de interés, lo que nos deja con 8 valores para cada escuela: Probabilidad del desempeño de un alumnx por debajo del básico, básico, satisfactorio y avanzado para ambas asignaturas. Como estas 8 variables corresponden a dos *multinomiales* la elección de la distancia no es trivial, y terminamos usando la distancia coseno.

La salida del algoritmo de UMAP en dos dimensiones para cada punto (cada escuela) tendrá como puntos más cercanos a aquellas escuelas cuyos alumnos

tengan desempeños parecidos. Este fue el resultado que obtuvimos para el gráfico en dos dimensiones de las escuelas para matemática y lengua:

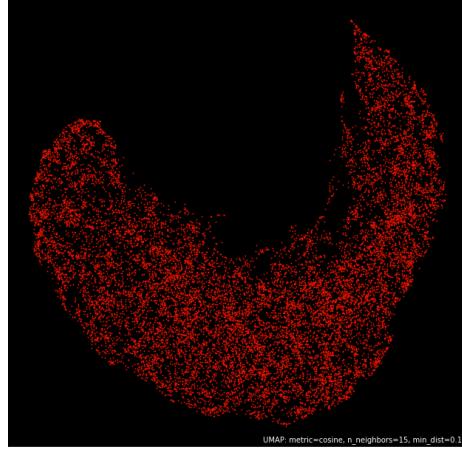


Figura 2: Embedding con UMAP para las escuelas según el desempeño de sus alumnxs en Lengua y Matemática.

Donde cada punto es una escuela, la relación que exigimos a UMAP se puede ver mucho más clara si coloreamos estas escuelas de acuerdo al desempeño promedio de sus alumnxs para cada asignatura. Lo que da por resultado los siguientes mapas para matemática, donde mientras más rojizo es el color más se concentra allí ese desempeño.

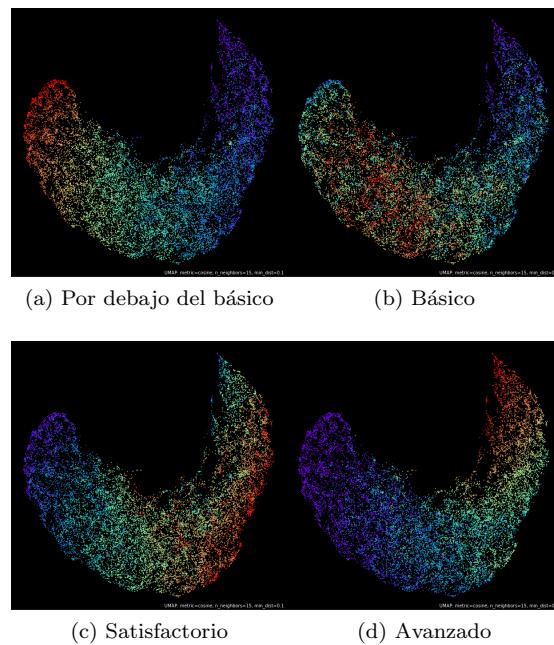


Figura 3: Desempeño en matemática por colores para las escuelas

Y el siguiente mapa de colores para los desempeños en lengua:

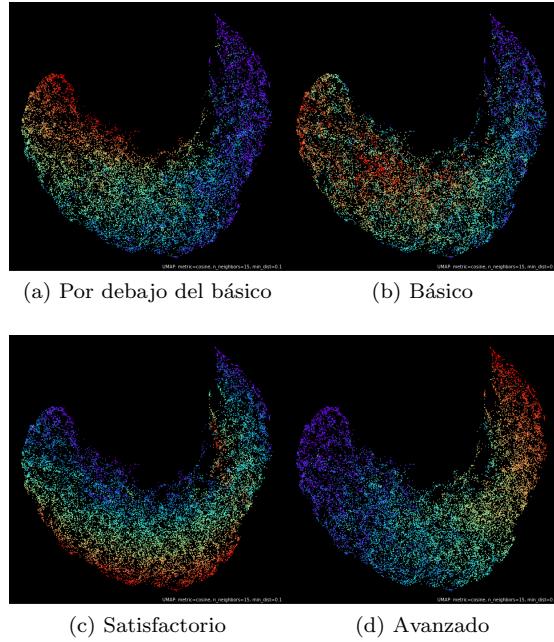
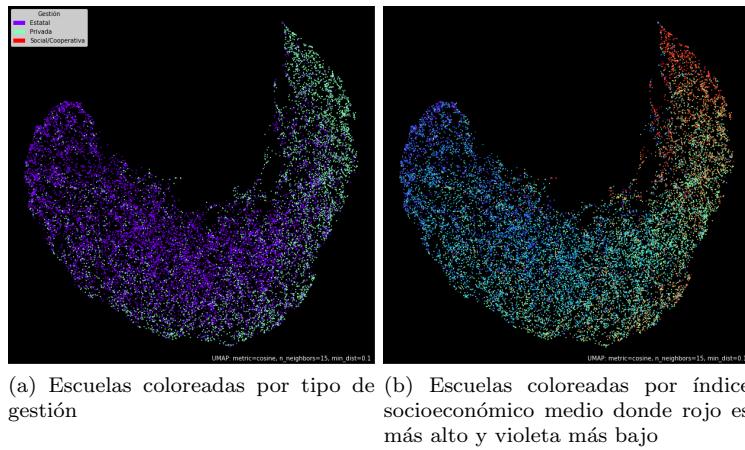


Figura 4: Desempeño en lengua por colores para las escuelas

Podemos ver entonces que nuestro *embedding* para escuela representa un gradiente de derecha a izquierda de que tan buen desempeño tienen sus alumnos. Intentamos luego colorear estas mismas escuelas por dos variables que no fueron utilizadas para la construcción de los puntos.



Podemos observar aquí también la relación entre el desempeño en las pruebas y las variables del contexto, pero a nivel de escuela.

Para el caso de los exámenes de ciencias sociales y naturales podemos observar el siguiente *embedding*:

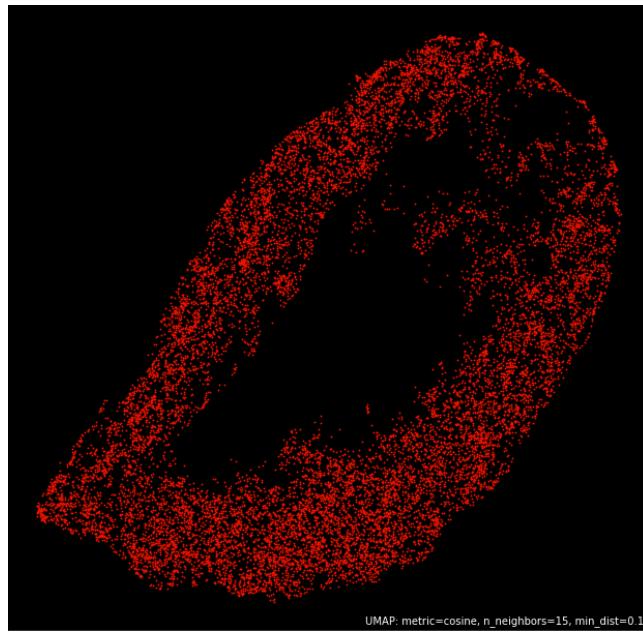


Figura 5: Embedding con UMAP para las escuelas según el desempeño de sus alumnxs en Ciencias Sociales y Ciencias Naturales.

Y vemos como están mapeados los desempeños en ciencias sociales:

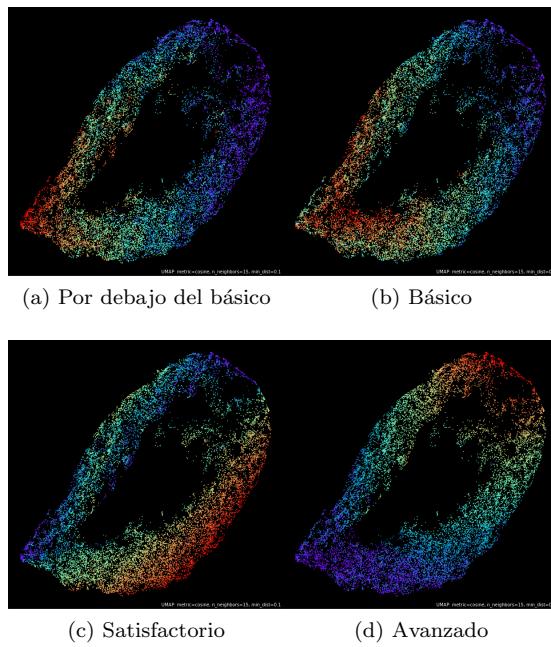


Figura 6: Desempeño en ciencias sociales por colores para las escuelas

Y los desempeños en ciencias naturales:

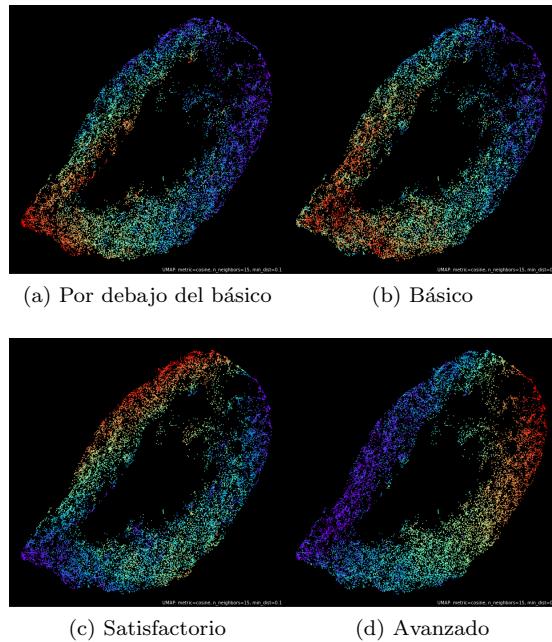
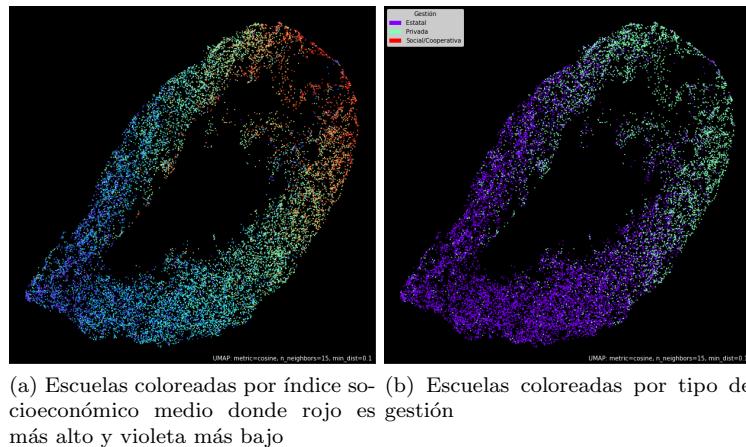


Figura 7: Desempeño en ciencias naturales por colores para las escuelas

Observando las mismas dependencias sobre las siguientes variables:



Estas relaciones en los agregados por colegios son consistentes con las predicciones a nivel alumno y con el análisis exploratorio realizado.

7. Conclusiones y orientaciones

Luego de diversos análisis sobre las variables, recortes de ruido del dataset y diversos procesos, se lograron resultados concretos y tangibles de modelos de datos que predicen, con una precisión considerablemente buena, el desempeño de un alumnx en alguna de las asignaturas.

El proceso de creación de dichos modelos, sin ahondar todavía en resultados, nos brindó cuestiones esperables acerca de las características que suponía más importantes a la hora de clasificar a un estudiante, puesto que tomamos en cuenta lo visto en el análisis exploratorio, no sorprendió el hecho de que el índice socioeconómico medio haya sido la feature que más peso brinda, tanto para el DecisionTree (árbol de decisión) como para LightGBM. Puesto que dentro del análisis mencionado, es de esperar que dadas las altas distribuciones de probabilidad de clasificación del desempeño, los modelos predictivos elijan estas variables para su entrenamiento y posterior testeo, en la predicción propiamente dicha.

Dentro de nuestro análisis exploratorio se identificaron variables, como la gestión, el ICSE, la tenencia de un teléfono smartphone y la repetición de primaria, entre otros, que consideramos, merecen ser tenidos en cuenta y discutidos en pos de mejorar la calidad de la educación Argentina, únicamente apelando a artilugios estadísticos. Luego utilizando un agrupamiento por escuela de acuerdo a sus probabilidades de desempeño logramos con UMAP una representación en bajas dimensiones que refuerza la relación entre los desempeños en las pruebas y los contextos, pero ya no a nivel individual sino también como agregado a nivel de escuela.

Lo que más nos impresionó fue la capacidad del modelo de LightGBM para discernir entre alumnxs avanzados y lxs que están por debajo del básico. Estas métricas en contraste con otros problemas de machine learning a los que nos enfrentamos podrían parecer bajas, pero es sorprendente y triste que se pueda determinar tan bien la calidad educativa de un alumno prácticamente sin conocerlo. Nuestro modelo tiene muy claro solo con el contexto del alumno si su probabilidad de ser avanzado es 5% o es 60% y lo mismo para su probabilidad de pertenecer a por debajo del básico, esto, como detallamos en la introducción desplaza el foco y la carga de los resultados de los exámenes de los estudiantes como individuos, deja ver una historia de profunda desigualdad de oportunidades y nos debe hacer reflexionar sobre nuestros privilegios.

Todos estos resultados, no son una casualidad, las variables medidas no son simplemente un número, detrás de ellas se encuentra una realidad que debemos afrontar. La educación depende de cuestiones que son exógenas a los alumnxs. Habiendo desarrollado los diversos factores que pueden tener incidencia en los resultados de las pruebas Aprender, queremos destacar siguiendo el aporte de Hebe Roig, que la solución de los problemas educativos no responde a una fórmula sino que se requiere de intervenciones que consideren todos los aspectos que se ponen en juego alrededor del examen. A su vez, es prioridad establecer como punto de partida la reflexión acerca del sentido de la educación que marca la diferencia en cómo comprendemos el problema que se presenta a partir de los datos observados. Cómo analicemos los mismos, determinará ampliamente las observaciones realizadas y las intervenciones posibles.

Volviendo sobre los datos analizados anteriormente, no sólo llegamos a la

conclusión de que los mismos pueden predecirse en gran parte, sino que este estudio reafirma lo que la comunidad educativa expresa hace ya varios años en relación a estas evaluaciones.

Las evaluaciones en este sentido, deben apuntar a poder comprender la situación para generar un impacto sobre la misma y por eso lo importante de todo proceso de evaluación son las orientaciones que de este se desprenden. No sirve describir el problema si no se actúa en consecuencia para mejorar la calidad educativa, entendiendo a la calidad como el foco en la realidad particular de cada jurisdicción/institución y no desde el enfoque de las escuelas eficaces que se buscó fomentar desde la reforma educativa de los años 90⁷.

Reiteramos la importancia de una mayor cercanía entre el sistema de evaluación y la toma de decisiones para encontrar soluciones. Para eso, es necesario incluir realmente a la comunidad educativa, conformada no solo por expertxs, sino también por docentxs, directivxs y gabinetes entre otrxs.

A su vez, queremos señalar que las transformaciones en el sistema educativo se dan en procesos que pueden llevar más de un año y es posible que en evaluaciones anuales no observemos un gran impacto en el desempeño de lxs estudiantes, con esto queremos reafirmar que el foco debe establecerse en la toma de decisiones y la implementación de políticas educativas que fomenten no sólo la calidad a través de la medición de los resultados, sino que también apunten a los factores contextuales que rodean a las pruebas Aprender.

Bibliografía

- [1] Feldfeber M., Puiggrós A., Robertson S. y Duhalde M. “*La privatización educativa en Argentina*”. *Investigaciones Internacionales de la Educación*, CTERA, Secretaría de Educación, 2018 [\[LINK\]](#)
- [2] Feldfeber M. y Gluz, N. *Las políticas educativas en Argentina: Herencias de los '90, contradicciones y tendencias de “NUEVO SIGNO”*, 2011 [\[LINK\]](#)
- [3] Roig, H. *Evaluar las evaluaciones. Un ejercicio necesario...*, Segundo Congreso Nacional de Educación, Universidad y Comunidad. ”Discursos y Prácticas sobre la Educación Pública”. San Luis, 5,6 y 7 de septiembre de 2019.
- [4] Aprender *Análisis de desempeños por capacidades y contemidos, Nivel primario*, Secretaría de Evaluación Educativa. Ministerio de Educación y Deportes, Presidencia de la Nación, 2016 [\[LINK\]](#)
- [5] Argentinos por la educación ¿*Cómo y cuánto se invierte en la educación argentina?*, 2019 [\[LINK\]](#)
- [6] Aprender *Acceso y uso de TIC en estudiantes y docentes*, Secretaría de Evaluación Educativa. Ministerio de Educación y Deportes, Presidencia de la Nación, 2016 [\[LINK\]](#)