



Universidad de  
**San Andrés**

**Ciencia de Datos**

**Propuesta de Investigación:**

**Modelo predictivo de Jet Lag Social en adolescentes: identificación  
de factores de riesgo para mejorar el bienestar y rendimiento  
académico**

**Profesora:**

Maria Noelia Romero

**Tutor:**

Ignacio Spiousas

**Autores:**

Gabriel Alejandro Díaz y Gabriela Belen Sanchez

**Fecha de entrega:**

7 de diciembre de 2024

## Introducción

Desde hace mucho tiempo que se tiene un amplio conocimiento de los beneficios y la importancia del sueño, tanto para el correcto desempeño diario, como para el desarrollo a lo largo de la vida. Este toma una mayor relevancia durante la niñez y la adolescencia, cuando el desarrollo cognitivo y emocional se encuentra latente, junto con mayores demandas académicas y sociales. De esta forma, la falta de sueño puede tener efectos perjudiciales en la salud física, mental y emocional de los jóvenes, así como en su rendimiento escolar.

Una de las principales causas de esta carencia y/o la presencia de *jetlag social* (es decir, la diferencia del tiempo de sueño entre los días de semana y los libres) es el desajuste entre los tiempos de sueño biológicos (cronotipos) y las obligaciones sociales, donde se resaltan los horarios escolares. Es aquí donde se presenta una paradoja, ya que, frente a este desajuste y a una incapacidad de cambiar los horarios institucionales, los estudiantes enfrentan dos posibles caminos: a) cumplir con los horarios preestablecidos del colegio, o b) perder partes de las clases para dormir siguiendo su ritmo biológico. Ambos casos pueden resultar en un bajo rendimiento académico; junto con los efectos cognitivos negativos del primero.

En este sentido, identificar fácilmente a los estudiantes que sufren de jetlag social toma un rol central. No solo permitiría detectar de manera temprana estos desajustes, sino que también daría pie al desarrollo e implementación de posibles intervenciones institucionales que promuevan una mejor adaptación a los horarios, o, inclusive, un replanteamiento de estos, ya dados por sentado hace años.

Es bajo este contexto que este trabajo plantea la pregunta de investigación: ¿es posible predecir la presencia de jet lag social en adolescentes empleando variables fácilmente observables en el contexto educativo? Para ello, se plantea la generación

de dos modelos estadísticos: RandomForest y una regresión logística utilizando variables de fácil acceso (como el género, edad, horario de entrada, notas, entre otras), para posteriormente compararlos y escoger el que tenga un mejor desempeño. Así, se espera desarrollar una herramienta útil y fácil de emplear por docentes u otros miembros de la institución que permita identificar a posibles estudiantes en riesgo.

### **Literatura previa**

El cronotipo se refiere a las preferencias biológicas individuales de las personas para el sueño y la vigilia, influenciadas por su reloj biológico interno (Goldin et al., 2020). Durante la adolescencia, este cronotipo suele desplazar hacia un patrón más nocturno, lo que implica que los adolescentes se sienten más despiertos y alertas durante la noche, y prefieren dormir más tarde por la mañana (Goldin et al., 2020). Este fenómeno se debe al desarrollo neurobiológico de los adolescentes, que involucra un retraso en la producción de melatonina, una hormona clave en la regulación del sueño (Roenneberg et al., 2007). Sin embargo, el cronotipo no siempre está alineado con las demandas sociales, como el horario escolar temprano, lo que genera un desajuste conocido como jetlag social.

El jetlag social se refiere al desajuste entre el ritmo circadiano interno de una persona (es decir, su cronotipo) y los horarios impuestos por la sociedad, como el horario escolar, que obliga a los adolescentes a despertarse temprano, a pesar de que su reloj biológico indica que deberían dormir más tarde. Este desajuste tiene efectos negativos en la salud de los adolescentes, incluidas consecuencias como somnolencia diurna, fatiga, alteraciones cognitivas y una disminución en el rendimiento académico (Rodríguez Ferrante et al., 2022). Además, se ha observado

que el jetlag social puede estar relacionado con problemas de salud mental, como depresión y ansiedad (Goldin et al., 2020). Estos efectos no solo afectan el bienestar general de los adolescentes, sino que también impactan su capacidad para rendir de manera óptima en el ámbito académico.

Los estudios han demostrado que un retraso en el inicio de las clases escolares podría mitigar algunos de los efectos negativos del jetlag social. De hecho, las investigaciones han reportado que un horario escolar más tardío está asociado con mayores cantidades de sueño, menores niveles de jetlag social y una mejora en el bienestar de los adolescentes (Rodríguez Ferrante et al., 2023). Sin embargo, no solo el horario de inicio de las clases tiene un impacto en los patrones de sueño y rendimiento académico, sino que también deben considerarse otros factores como la exposición a la luz, el género, la edad y el rendimiento académico (Goldin et al., 2020; Rodriguez Ferrante & Leone, 2023).

En este contexto, el desarrollo de un modelo predictivo para identificar a los adolescentes en riesgo de sufrir jetlag social es esencial para diseñar intervenciones que puedan mitigar estos efectos negativo

### **Base de datos**

La base de datos proviene de un análisis longitudinal en una secundaria de la Ciudad de Buenos Aires realizado por Goldin et al. (2020). Contiene 712 observaciones de estudiantes de primer año (2015) y de quinto año (2019), de los turnos mañana, tarde y vespertino. Cuenta con variables demográficas como edad y género, así como las notas escolares y variables sobre el comportamiento de sueño, como la hora a la que se duerme, a la que se despierta y la duración del sueño, tanto para días de semana como libres.

La muestra está conformada por un 47.01% de mujeres ( $n = 334$ ); y presenta una edad media de 13.47 ( $DE = 0.34$ ) en primer año y de 17.51 años ( $DE = 0.38$ ) en quinto. Los estudiantes se encuentran distribuidos un 35.49% ( $n = 538$ ) en el turno mañana, un 35.17% ( $n = 534$ ) en la tarde, el resto en vespertino (29.34%,  $n = 434$ ).

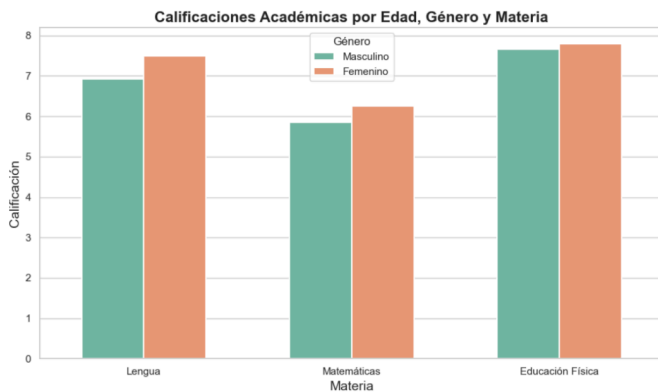
Con respecto al desempeño académico, como se puede apreciar en las figuras 1 y 2, las mujeres muestran un mejor desempeño generalizado en todas las asignaturas (Lengua:  $M = 7.48$ ,  $DE = 1.24$ ; Matemáticas:  $M = 6.26$ ,  $DE = 1.97$ ; Ed. Física:  $M = 7.79$ ,  $DE = 1.32$ ) en comparación con los hombres (Lengua:  $M = 6.92$ ,  $DE = 1.26$ ; Matemáticas:  $M = 5.84$ ,  $DE = 1.99$ ; Ed. Física:  $M = 7.65$ ,  $DE = 1.15$ ). En los tres turnos, la materia con peor desempeño en matemática ( $M = 6.04$ ,  $DE = 1.99$ ), mientras que aquella con el mejor es Educación Física ( $M = 7.72$ ,  $DE = 1.23$ ).

En cuanto a los horarios, durante la semana, los adolescentes más jóvenes del turno mañana se despiertan a las 06:17 a.m. ( $M = 6:17$ ,  $DE = 26$  min), mientras que los de los turnos tarde y noche lo hacen más tarde, a las 08:33 a.m. ( $M = 8:33$ ,  $DE = 64$  min) y 09:16 a.m. ( $M = 9:16$ ,  $DE = 87$  min), respectivamente. Los adolescentes mayores se despiertan a las 06:23 a.m. ( $M = 6:23$ ,  $DE = 25$  min) en el turno mañana, y a las 08:45 a.m. ( $M = 8:45$ ,  $DE = 71$  min) y 10:03 a.m. ( $M = 10:03$ ,  $DE = 88$  min) en los turnos tarde y noche. En cuanto al inicio del sueño, los más jóvenes se duermen a las 11:37 p.m. ( $M = 11:37$ ,  $DE = 66$  min) en el turno mañana, y más tarde a las 00:17 a.m. ( $M = 00:17$ ,  $DE = 83$  min) y 00:55 a.m. ( $M = 00:55$ ,  $DE = 77$  min) en los turnos tarde y noche. Los adolescentes mayores presentan patrones más tardíos, durmiéndose a la 01:19 a.m. ( $M = 01:19$ ,  $DE = 64$  min) y 01:52 a. m. ( $M = 01:52$ ,  $DE = 81$  min). Durante los fines de semana, tanto los jóvenes como los mayores se despiertan más tarde (11:22 a. m.,  $M = 11:22$ ,  $DE = 102$  min y 11:43

a.m.,  $M = 11:43$ ,  $DE = 119$  min, respectivamente). El inicio del sueño se retrasa y ocurre entre las 02:06 a.m. ( $M = 02:06$ ,  $DE = 108$  min) y 03:42 a.m. ( $M = 03:42$ ,  $DE = 104$  min), según el turno escolar.

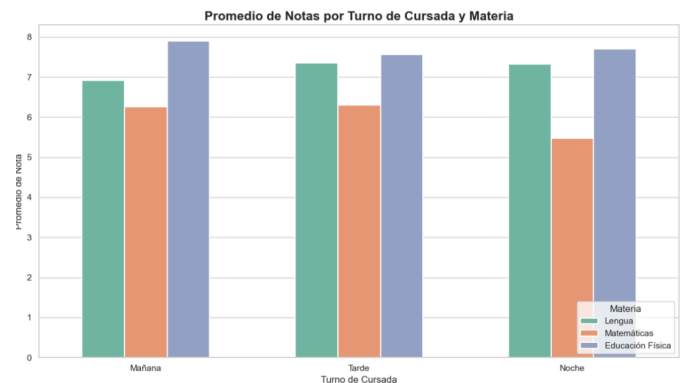
**Figura 1**

*Calificaciones académicas por Edad, Género y Materia*



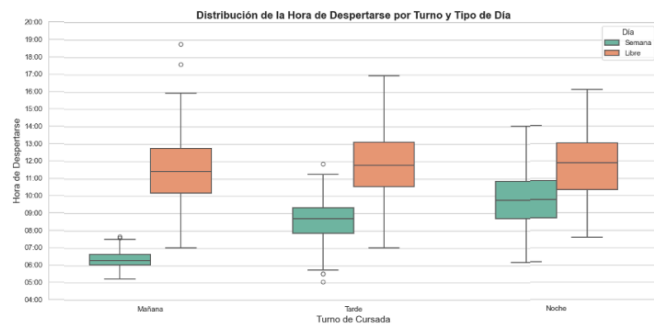
**Figura 2**

*Promedio de Notas por Turno de Cursada y Materia*



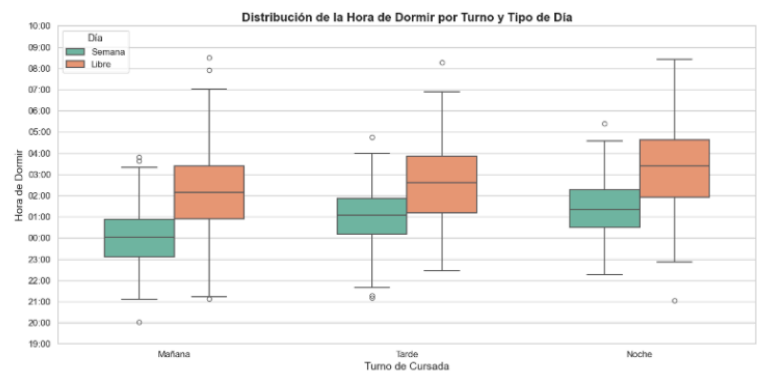
**Figura 3**

*Distribución de la Hora de Despertarse por Turno de Cursada y Dia*



**Figura 4**

*Distribución de la Hora de Dormir por Turno de Cursada y Dia*



## Metodología

Para llevar a cabo la predicción sobre la presencia o no de jetlag en estudiantes adolescentes, en primer lugar, se realizará una búsqueda de *outliers*, es decir, valores que se encuentran significativamente alejados del resto. Específicamente, estos serán analizados, y descartados siguiendo la metodología MAD (Median Absolute Deviation), que se basa en la mediana y su desviación absoluta. Una vez

eliminados, se hará una partición aleatoria de la base de datos en una muestra de entrenamiento (70% de los datos), y una muestra de testeo (30% restante).

Se entrenaran dos modelos estadísticos para la clasificación: un modelo de RandomForest (RF) y una Regresión Logarítmica (Logit). Por un lado, RF es un algoritmo de Machine Learning que utiliza un conjunto de árboles de decisión para realizar predicciones. Cada nodo de estos es generado con una cantidad  $m$  de  $p$  predictores aleatorios (en este caso  $m = p/3$ ). Así, RF combina las respuestas de todos los árboles para obtener una predicción final, basada en el voto por mayoría. Por ser resultado de este “promedio” de árboles, su naturaleza no paramétrica, pierde en interpretabilidad en cuanto a sus predictores.

El modelo Logit, por su parte, permite predecir la probabilidad del evento. Por ello es necesario determinar un umbral de corte a partir del cual se considere a la observación como positiva (en este caso será de 0.5). A su vez, el modelo Logit se trata de un modelo paramétrico y más interpretable, según la fórmula:

$$p(X) = \frac{e^{\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p}}{1 + e^{\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p}}$$

Así, los coeficientes indican la influencia de las variables sobre la probabilidad, aunque esta, cabe aclarar, no es lineal, es decir, los

efectos no son proporcionales a los cambios en los coeficientes. De esta forma, el modelo Logit se trata de un modelo paramétrico y más interpretable: los coeficientes indican la influencia de las variables sobre la probabilidad. Aún así, cabe aclarar que esta influencia no es lineal, es decir, los efectos de las variables sobre la probabilidad de presentar jetlag social no son proporcionales a los cambios en los coeficientes, no es constante.

La comparación entre ambos se realizará empleando la muestra de testeo. Por un lado, se comparará la exactitud de los modelos, es decir, la proporción de

predicciones correctas respecto al total de predicciones:

$$Accuracy = \frac{Verdaderos\ Positivos + Verdaderos\ Negativos}{Total\ de\ Observaciones}$$

Esta métrica será complementada con un análisis de la matriz de confusión y de Curva ROC, para cada modelo. Específicamente, mediante el cálculo de la AUC (área bajo la curva) es posible medir la capacidad del modelo para clasificar correctamente entre las clases. Un valor AUC cercano a 1 indica una buena predicción, mientras que un AUC cercano a 0.5 sugiere un modelo con un rendimiento similar a uno aleatorio.

Considerando esta información, se seleccionará el modelo con la mejor capacidad predictiva, a fin de desarrollar una herramienta de fácil implementación.

## **Conclusiones y limitaciones**

### *Resultados esperados*

El propósito de este proyecto es desarrollar un modelo predictivo robusto para identificar a los adolescentes en riesgo de sufrir jetlag social, utilizando variables observacionales fácilmente accesibles en el entorno educativo. A través del análisis de los datos de estudiantes de secundaria, se utilizarán como variables de predicción del jetlag social a la edad, el género, el turno escolar, las horas de sueño total y el rendimiento académico. En particular, se anticipa la duración del sueño y el turno de entrada al colegio, especialmente si su cronotipo es más nocturno, serán las variables más relevantes.

Se espera que el modelo de Random Forest sea el más eficaz para clasificar a los estudiantes en riesgo. Este modelo podrá identificar patrones complejos no lineales



derivados de la interacción entre las horas de sueño, el turno escolar y el rendimiento académico. Por su parte, el modelo de regresión logística, aunque más sencillo y menos flexible, en términos de interacciones complejas, se destaca por su mayor interpretabilidad, lo que facilita la comprensión de cómo cada variable influye directamente en la probabilidad de sufrir jetlag social. Por ello, este modelo será especialmente útil para comunicar los resultados a los educadores.

### *Aplicaciones prácticas*

El principal resultado esperado es una correcta identificación de estudiantes en riesgo, lo que permitirá el desarrollo de una herramienta predictiva accesible para docentes y personal educativo. De esta forma se facilitaría el desarrollo e implementación de intervenciones personalizadas tempranas, donde se proporcione a los estudiantes las condiciones necesarias para mejorar su salud física, su bienestar emocional y su rendimiento académico. Estas podrían incluir modificaciones en los horarios escolares, un inicio más tardío de las clases o la promoción de hábitos de higiene del sueño (como la concienciación sobre la importancia de horarios regulares para dormir).

### *Limitaciones del proyecto*

El proyecto enfrenta varias limitaciones que podrían influir en los resultados y su aplicación práctica. En primer lugar, la base de datos utilizada proviene de una única escuela secundaria en la Ciudad de Buenos Aires, lo que introduce un sesgo geográfico y cultural que podría dificultar la generalización de los resultados a otras regiones o tipos de instituciones educativas, dado que los patrones de sueño, actitudes hacia la educación y hábitos familiares pueden variar considerablemente. Además, las variables relacionadas con el sueño se obtienen mediante

auto-reportes de los estudiantes, lo cual podría introducir imprecisiones debido a errores en la percepción de los propios hábitos o sesgos en la información proporcionada, quienes podrían no ser completamente conscientes de sus patrones de sueño o preferir responder de manera socialmente deseable. Aunque se incluyen variables relevantes como edad, género, turno escolar y rendimiento académico, el modelo no contempla factores adicionales que podrían influir en la presencia de jetlag social, como el estrés, las actividades extracurriculares, el entorno familiar o la exposición a la luz, lo que podría limitar la capacidad del modelo para capturar toda la complejidad del fenómeno. Finalmente, dado que el cronotipo de los adolescentes puede cambiar con el tiempo, el modelo basado en datos de un solo momento podría no ser capaz de captar estas dinámicas a largo plazo, lo que afectaría la efectividad de las intervenciones si no se actualizan regularmente.

### *Conclusión final*

A pesar de estas limitaciones, el proyecto tiene un gran potencial para contribuir al bienestar de los adolescentes y mejorar la calidad educativa. El desarrollo de un modelo predictivo para identificar a los estudiantes en riesgo de sufrir jetlag social puede ser un paso importante para que las instituciones educativas ajusten sus prácticas a las necesidades biológicas de los estudiantes, favoreciendo su salud y su rendimiento académico. Aunque la implementación de intervenciones dependerá de recursos institucionales y de la disposición para realizar cambios, los resultados de este estudio podrían servir como base para futuras investigaciones y políticas educativas que busquen mejorar la calidad de vida y el desempeño académico de los adolescentes.

## Bibliografía

Goldin, A. P., Sigman, M., Braier, G., Golombek, D. A., & Leone, M. J. (2020). Interplay of Chronotype and School Timing Predicts School Performance. *Nature Human Behaviour*, 4(8), 743-752. <https://doi.org/10.1038/s41562-020-0853-6>

Rodríguez Ferrante, G., Goldin, A. P., Sigman, M., & Leone, M. J. (2022). Chronotype at the Beginning of Secondary School and School Timing Are Both Associated with Chronotype Development during Adolescence. *Scientific Reports*, 12, 8207. <https://doi.org/10.1038/s41598-022-11928-9>

Rodríguez Ferrante, G., Goldin, A. P., Sigman, M., & Leone, M. J. (2023). A better alignment between chronotype and school timing is associated with lower grade retention in adolescents. *npj Science of Learning*, 8, 21. <https://doi.org/10.1038/s41539-023-00171-0>

Rodríguez Ferrante, G., & Leone, M. J. (2023). Solar clock and school start time effects on adolescents' chronotype and sleep: A review of a gap in the literature. *Journal of Sleep Research*, 32(5), e13974. <https://doi.org/10.1111/jsr.13974>

Roenneberg, T., Kumar, C. J., & Merrow, M. (2007). The human circadian clock entrains to sun time. *Current Biology*, 17(2), R44-R45.