



*Asesorías y Tutorías para la Investigación Científica en la Educación Puig-Salabarría S.C.
José María Pino Suárez 400-2 esq a Lerdo de Tejada. Toluca, Estado de México. 7223898475*

RFC: ATI120618V12

Revista Dilemas Contemporáneos: Educación, Política y Valores.

<http://www.dilemascontemporaneoseducacionpoliticayvalores.com/>

Año: XIV Número: 1 Artículo no.:45 Período: 1 de septiembre al 31 de diciembre del 2026

TÍTULO: Modelando la felicidad en adultos en formación académica aplicando algoritmos de aprendizaje automático.

AUTORA.

1. Dra. Argelia Berenice Urbina-Nájera.

RESUMEN: En este estudio se presentan modelos del bienestar subjetivo en estudiantes universitarios mexicanos mediante algoritmos de clasificación supervisada, con el fin de identificar los factores que predicen la felicidad y evaluar la capacidad de distintos modelos para clasificar los niveles de bienestar. Mediante un diseño cuantitativo, longitudinal y explicativo, utilizando microdatos del BIARE del 2019 al 2025. La muestra incluyó 6,793 instancias y 52 atributos de adultos en formación universitaria mayores de 18 años. Los resultados evidencian que el algoritmo SimpleLogistic alcanzó el mejor desempeño, con una exactitud de 99.72%. Los factores positivos asociados a la felicidad incluyen la solvencia mensual, la seguridad financiera, el bienestar presente y anual, el propósito y el logro diario.

PALABRAS CLAVES: bienestar subjetivo, formación académica, aprendizaje automático, felicidad, algoritmos de clasificación.

TITLE: Modeling happiness in adults in academic training by applying machine learning algorithms.

AUTHOR:

1. PhD. Argelia Berenice Urbina-Nájera.

ABSTRACT: This study presents models of subjective well-being among Mexican college students using supervised classification algorithms, with the aim of identifying factors that predict happiness and

evaluating the ability of different models to classify well-being levels. The study employs a quantitative, longitudinal, and explanatory design, utilizing BIARE microdata from 2019 to 2025. The sample included 6,793 observations and 52 attributes for university students aged 18 and older. The results show that the SimpleLogistic algorithm achieved the best performance, with an accuracy of 99.72%. Positive factors associated with happiness include monthly solvency, financial security, present and annual well-being, purpose, and daily achievement.

KEY WORDS: subjective well-being, academic training, machine learning, happiness, classification algorithms.

INTRODUCCIÓN.

En los últimos años, el aprendizaje automático se ha aplicado cada vez más en diversas áreas del conocimiento; una de sus aplicaciones típicas es el análisis económico, especialmente en la detección de los fraudes financieros; no obstante, su potencial para identificar patrones ocultos en grandes volúmenes de datos la convierte en una herramienta poderosa para el estudio del bienestar subjetivo, al buscar patrones que expliquen cómo distintos factores influyen en la percepción de la satisfacción y de la calidad de vida.

El estudio del bienestar subjetivo en personas mayores de 18 años que asisten a la escuela merece especial atención, pues constituye una dimensión clave para comprender la calidad de vida y las condiciones que favorecen la felicidad en la población joven adulta. En un contexto marcado por tensiones económicas, sociales y emocionales, identificar los factores que influyen en la forma en que los estudiantes evalúan su satisfacción resulta fundamental para diseñar políticas educativas y sociales más efectivas; además, la creciente disponibilidad de datos longitudinales, como los del Módulo Bienestar Autorreportado (Instituto Nacional de Estadística y Geografía [INEGI], 2025) (que, a partir de este momento, denominaremos BIARE), posibilita el análisis de tendencias históricas y de patrones ocultos que antes permanecían invisibles.

El enfoque de esta investigación se centra en aplicar algoritmos de clasificación supervisada, regresión logística, regresión logística simple, y máquinas de soporte vectorial para modelar el bienestar subjetivo de universitarios. El propósito es evaluar la capacidad predictiva de cada modelo para identificar el mejor algoritmo que clasifique con mayor precisión las variables del bienestar subjetivo en distintos niveles. La necesidad de esta investigación radica en que aunque existen estudios sobre el bienestar en jóvenes, pocos han explorado sistemáticamente el uso de algoritmos de aprendizaje supervisado para identificar patrones complejos a partir de encuestas nacionales.

El bienestar subjetivo ha sido ampliamente estudiado en la psicología positiva y en las ciencias sociales. Diener (1984) lo definió como la evaluación cognitiva y afectiva que las personas realizan de su vida, lo cual sentó las bases para medir la satisfacción vital. Posteriormente, Ryff (1989) propuso un modelo multidimensional del bienestar psicológico que incluye la autonomía, el propósito vital y las relaciones positivas. Seligman (2011) amplió el enfoque hacia la eudaimonía, integrando dimensiones como el logro, el compromiso y el sentido. Estos referentes son los más citados y constituyen el marco teórico de la investigación sobre la felicidad y la calidad de vida.

En el contexto educativo, el bienestar subjetivo se ha vinculado con el rendimiento académico, la resiliencia y la autoeficacia. Estudios como el de Codina-Lorente, Gil-Madrona y Gómez (2025) muestran que el afecto positivo y el compromiso académico predicen significativamente el desempeño, mientras que el afecto negativo presenta una baja correlación.

En México, Contreras (2024) realizó una revisión sistemática que evidenció el uso predominante de la escala de Ryff en estudiantes de nivel superior, destacando la necesidad de metodologías más robustas para captar la diversidad de experiencias. Estos hallazgos refuerzan la importancia de estudiar la felicidad en jóvenes adultos en formación académica.

La literatura nacional ha vinculado el bienestar subjetivo con la calidad educativa y la pertinencia curricular. Cervera, Martí y Alejo (2024) demostraron que la satisfacción estudiantil es un componente

clave en la evaluación de programas de posgrado, lo que evidencia la relevancia del bienestar subjetivo en la acreditación y mejora de la educación superior; asimismo, investigaciones coordinadas por Millán y Castellanos (2018) subrayan la relación entre el capital humano, la calidad de vida, y la felicidad en México, y aportan un marco contextual para el análisis de encuestas nacionales como el BIARE.

El aprendizaje automático ha ganado terreno en la investigación social por su capacidad para identificar patrones ocultos y mejorar la clasificación de categorías minoritarias. Chen, Wy, Hu, et al. (2021) destacaron su potencial para la inferencia causal y la evaluación de políticas públicas. En educación, la revisión sistemática de Forero-Corba y Negre (2024) identificó más de 30 técnicas aplicadas en contextos pedagógicos, mostrando su impacto en la innovación; además, Taboada (2024) evidenció que el procesamiento del lenguaje natural y los algoritmos supervisados permiten analizar fenómenos emergentes en espacios virtuales, ampliando las posibilidades metodológicas en las ciencias sociales.

A pesar de los avances, pocos trabajos han integrado de manera sistemática el análisis del bienestar subjetivo en universitarios mexicanos mediante algoritmos de aprendizaje supervisado. La mayoría de los estudios se centran en mediciones tradicionales o en poblaciones generales. La investigación aporta evidencia metodológica al demostrar que algoritmos como SimpleLogistic pueden equilibrar precisión y explicabilidad, ofreciendo un modelo confiable para clasificar tanto categorías mayoritarias como minoritarias de bienestar; por ello, las preguntas de investigación se orientan a responder:

- a. ¿Qué factores predicen mejor la felicidad en estudiantes universitarios?
- b. ¿Qué modelo ofrece mayor capacidad para clasificar los distintos niveles de satisfacción?
- c. ¿Cómo varían estos patrones a lo largo del tiempo?

Las contribuciones teóricas, de este estudio, se sitúan en la intersección entre el análisis social y el aprendizaje automático. Al demostrar que algoritmos como SimpleLogistic pueden equilibrar precisión y explicabilidad, se aporta evidencia de que la inteligencia artificial puede enriquecer la investigación en ciencias sociales. El impacto previsto se basa en ofrecer un modelo confiable para predecir el bienestar

subjetivo de universitarios mexicanos, así como en generar insumos prácticos para diseñar intervenciones educativas, comunitarias y psicológicas, que promuevan una experiencia de vida más satisfactoria en la edad adulta.

Algoritmos de clasificación basados en regresión.

La regresión constituye uno de los enfoques más utilizados en el análisis estadístico y en el aprendizaje automático, pues permite modelar relaciones entre variables y estimar probabilidades de pertenencia a categorías específicas. En el ámbito de la clasificación supervisada, los algoritmos basados en regresión como la regresión logística y las máquinas de soporte vectorial ofrecen una combinación valiosa de precisión y facilidad de interpretación, lo que los convierte en herramientas idóneas para estudios sociales y psicológicos (Martínez y Pérez, 2024; Jurafsky y Martin, 2026).

La regresión logística, en particular, es un método de clasificación que estima la probabilidad de que una persona se encuentre en determinados niveles de bienestar. Su utilidad radica en la construcción de fronteras de decisión lineales, y frente a la regresión lineal, ofrece ventajas notables: es más sencilla de implementar, requiere menos esfuerzo de entrenamiento, permite clasificar nuevos registros con eficacia y puede ampliarse a problemas multiclase (Whig et al., 2024); además, su formulación matemática se deriva directamente de la regresión lineal, lo que facilita su comprensión y aplicación en estudios sociales y psicológicos.

En tanto que la regresión logística simple mejora la robustez y la precisión frente a datos con ruido o variabilidad, como ocurre en encuestas como BIARE, es especialmente útil para capturar patrones más finos y mejorar la clasificación de clases minoritarias, asegurando que las conclusiones no estén sesgadas hacia la mayoría (Christensen, 2025).

Por último, las máquinas de soporte vectorial (SMO en Weka) constituyen una de las herramientas más potentes para analizar patrones complejos. Su lógica consiste en ubicar los datos; por ejemplo, respuestas de estudiantes sobre satisfacción, salud o apoyo social en un espacio matemático para trazar una frontera

que maximice la separación entre distintos niveles de bienestar. A diferencia de otros modelos, las máquinas de soporte vectorial pueden detectar relaciones no lineales mediante kernels, lo que les permite descubrir combinaciones de factores de la variable dependiente que no siguen patrones simples.

En síntesis, los SVM aportan una perspectiva analítica que complementa la regresión logística y la regresión logística simple; mientras estas últimas ofrecen interpretabilidad y robustez, los SVM permiten capturar estructuras ocultas y no lineales, enriqueciendo la comprensión integral del bienestar subjetivo (Sabry, 2023); de esta manera, este estudio no solo se centra en identificar el modelo que mejor predice, sino también en cómo cada uno de ellos contribuye a una visión panorámica y matizada del bienestar de los universitarios.

Metodología.

El estudio se centró en estudiantes universitarios mexicanos, con el propósito de identificar los factores asociados al bienestar subjetivo y estimar la probabilidad de ubicarse en distintos niveles de satisfacción mediante modelos de regresión logística, regresión logística simple y máquinas de soporte vectorial. En esta sección se detalla la estrategia metodológica que posibilita el análisis de tendencias longitudinales, la evaluación de interacciones y la captura de variaciones históricas vinculadas a la identificación del grado de felicidad de los universitarios (Figura 1).

Figura 1. Proceso para modelar la felicidad con algoritmos de regresión.



Diseño de la investigación.

Se realizó un estudio cuantitativo, longitudinal y explicativo utilizando los microdatos del Bienestar Autorreportado (BIARE) del Instituto Nacional de Estadística y Geografía (INEGI) correspondientes al periodo 2019–2025. El objetivo fue identificar los factores asociados al bienestar subjetivo en estudiantes universitarios mexicanos y modelar la probabilidad de encontrarse en distintos niveles de satisfacción mediante una regresión logística ordinal.

Fuente de datos.

El BIARE recoge información representativa sobre satisfacción con la vida, bienestar económico, salud mental, estados de ánimo y de salud física, eudemonía, soledad y satisfacción con dominios específicos (trabajo, vivienda, relaciones), y es de acceso público. Se utilizaron los microdatos anonimizados, con un diseño muestral probabilístico estratificado y por conglomerados correspondientes del 2019 - 2025 (INEGI, 2025).

Preprocesamiento de los datos.

El conjunto de datos fue sometido a un proceso de limpieza y curación para garantizar su calidad analítica. En los casos de valores faltantes que impedían calcular el promedio del bienestar subjetivo se eliminaron, así como los con inconsistencias o fuera de rango; posteriormente, la variable de bienestar subjetivo se codificó en siete categorías, ordenadas de “completamente insatisfecho” a “completamente satisfecho”, a fin de preservar la naturaleza ordinal de la medición y facilitar su modelado estadístico.

El conjunto de datos resultante contiene 6,793 instancias, 52 atributos, y una variable categórica que clasifica el bienestar subjetivo en siete niveles, codificados como categorías ordenadas (Tabla 1). Se realizó un proceso de balanceo de clases para corregir la distribución desigual de observaciones entre ellas, a fin de evitar que los modelos se sesguen hacia la categoría mayoritaria; para ello, en Weka 3.8.6 se utilizó el filtro *Resample*.

Descripción de las variables sociodemográficas.

La muestra combinada total del periodo 2016–2025 supera las 25 mil observaciones. Para este estudio se construyó una submuestra de personas mayores de 18 años que continúan con su formación académica, definida como personas con nivel educativo superior (licenciatura o posgrado) y con asistencia escolar vigente. Se excluyeron los casos con valores faltantes o inconsistentes en la variable dependiente o en los predictores centrales (bienestar subjetivo), dado que solo para el año 2025 se obtuvieron resultados sobre salud mental, soledad y bienestar económico. En la Tabla 1 se presentan las características sociodemográficas de la muestra total de instancias a analizar (6,793), cuyas variables son: entidad, sexo, edad, nivel de instrucción (i_niv), estado conyugal (e_con), ocupación (ocu), actividad económica (a_eco) e ingresos (ing).

Tabla 1. Características sociodemográficas del conjunto de datos.

	Endidad	sexo	edad	i_niv	e_con	ocu	a_eco	ing
Media	16	2	29	6	5	4	5	\$1,619.43
Mediana	15	2	22	6	6	5	5	\$-
Moda	14	2	18	7	6	1	5	\$-
Desviación estándar	8	0	16	3	2	3	3	\$3,479.67
Varianza de la muestra	64	0	260	8	4	7	7	\$-
Coefficiente de asimetría	0	0	2	15	-1	0	0	\$6.92
Mayor	32	2	87	8	9	9	9	\$80,000.00
Menor	1	1	18	0	0	0	0	0
Nivel de confianza		95%						

En la Tabla 1 se presentan los datos estadísticos descriptivos que ofrecen una visión general de las características demográficas y económicas de la muestra. Cabe señalar, que en la encuesta del BIARE, el 99 significa “no sé”, por lo que los datos se interpretarán en función de lo reportado en el conjunto de datos como válido y codificado como “no sé”. Se observa que el sexo predominante es el de las mujeres, lo que evidencia una tendencia clara en la composición del grupo, que en promedio, tiene 29 años. En cuanto al nivel educativo, la mayoría de los participantes alcanzó la preparatoria o el bachillerato, aunque existe una amplia dispersión que abarca desde personas sin escolaridad hasta quienes cuentan con estudios superiores hasta el grado de maestría.

Respecto a la situación conyugal, los datos reflejan una media de 5, que corresponde al estado “casado”, aunque el rango y la asimetría negativa sugieren que hay más participantes en situación de separación o de soltería. La variable ocupación presenta diversidad, siendo más frecuente la participación en actividades relacionadas con gestiones o trámites para iniciar un negocio propio.

En la actividad económica, los valores indican que gran parte de los participantes se involucra en tareas como los quehaceres del hogar remunerados o el cuidado de infantes, lo que refleja un perfil de actividades no siempre formales.

Finalmente, la variable ingreso presenta una media baja de \$1,619.43, acompañada de una desviación estándar muy alta y de un rango de \$80,000. Esto revela una marcada desigualdad de ingresos, confirmada

por el coeficiente de asimetría positivo, que indica que pocos casos con ingresos muy elevados distorsionan la media, mientras que la mayoría percibe ingresos bajos.

Variable dependiente (Clase).

El bienestar subjetivo se midió en una escala de 7 categorías (de *completamente insatisfecho* a *completamente satisfecho*). Se utilizaron tres modelos: regresión logística, regresión logística simple, y máquinas de soporte vectorial para conservar el orden natural de las categorías y estimar la probabilidad de ubicarse en niveles superiores de satisfacción.

Variables independientes (atributos).

Se incluyeron los dominios centrales del BIARE considerados en la fuente de datos para los años 2019-2024; en los dominios de salud mental, soledad y bienestar económico se asignó un valor de cero a todas las observaciones. Cada dominio se operacionaliza como un índice promedio estandarizado en una escala del 0–10. En total, el conjunto de datos está etiquetado con 52 variables, distribuidas en 10 categorías (Tabla 2).

Tabla 2. Atributos asociados a cada categoría del bienestar subjetivo.

Categoría.	No.	Descripción de los atributos.
Periodo	1	2019-2025
(1) Variables sociodemográficas.	8	Entidad, sexo, edad, Nivel_instrucción, Edo_conyugal, Ocupación, Actividad_económica, Ingresos.
(2) Satisfacción con la vida actual y durante el año pasado.	2	Bienestar_hoy, Bienestar_año.
(3) Eudemonía.	11	Bienestar_personal, Optimismo_proyectivo, Autonomía_vital, Resiliencia, Propósito,

		Identidad_afortunada, locus_control, misión_personal, Religión, Logro_diario, Resiliencia_emocional.
(4) Estados de ánimo y físicos.	10	Buen_humor, Tranquilidad, Energía, Concentración, Alegre, Mal_humor, Preocupación, Cansancio, Sin_interés, Tristeza.
(5) Dominios de satisfacción.	5	Nivel_vida, Salud_física, Logros_vida, Relaciones_personales, Perspectiva_futuro.
(6) Dominio de satisfacción (tiempo libre).	1	Satis_tiempo.
(7) Dominio de satisfacción (ámbitos cotidianos).	6	Seguridad_ciudadana, Satis_vivienda, Satis_vecindario, Satis_ciudad, Satis_país, Satis_comunidad.
(8) Salud mental.	5	Tensión_nerviosa, Preocupación_constante, Interés_reducido, Desánimo, Trastorno_sueño.
(9) Soledad.	1	Soledad_percibida.
(10) Bienestar económico.	2	Solvencia_mensual, Seguridad_financiera.

Nota: La clasificación de la tabla se toma del [*Módulo de Bienestar Autorreportado del INEGI \(2025\)*](#).

Escenarios de experimentación.

En la Figura 1 se muestran los escenarios de experimentación que conllevan el análisis en tres etapas:

La primera es la fase descriptiva longitudinal, que consiste en estimar promedios anuales de cada dominio del bienestar para identificar tendencias y rupturas en las personas mayores de 18 años que asisten a la escuela.

La segunda representa los modelos de regresión logística ordinal que favorecen la estimación de modelos acumulativos y por año, interpretando coeficientes mediante *odds ratios* y efectos marginales.

La tercera, la comparación temporal e interacciones de términos de interacción (dominio $\times$ año, dominio $\times$ condición universitaria) para evaluar cambios en la fuerza explicativa de los predictores en distintos contextos históricos.

Se organizan en tres escenarios: 1) experimentación con los parámetros de cada algoritmo y la normalización de los datos, 2) modificación de los parámetros de cada algoritmo, normalización de los datos y balanceo de clases, y por último, 3) se detallan los resultados por clase, destacando el comportamiento de cada modelo frente a categorías mayoritarias y minoritarias.

Consideraciones éticas.

Los datos del BIARE son de acceso público, están anonimizados, y cumplen con los estándares éticos en materia de protección de datos personales. No se requirió aprobación adicional del INEGI (2025).

Resultados.

El propósito de este apartado es presentar los resultados de la experimentación descritos con el fin de identificar el algoritmo que clasifique con mayor exactitud los distintos niveles del bienestar subjetivo; asimismo, busca determinar cuál es el modelo más adecuado para predecir la felicidad y responder a las preguntas de investigación. En este apartado se presentan los resultados obtenidos en cada uno de los escenarios de experimentación descritos en la Figura 1.

Escenario 1: Configuración inicial.

En la Tabla 3 se muestra el desempeño de los tres algoritmos; el análisis con regresión logística mostró una exactitud del 98.69%, con una precisión del 98.8% y un índice Kappa de 0.976, lo que indica una concordancia casi perfecta entre las predicciones y las categorías reales. Este modelo permitió identificar con claridad los factores asociados al bienestar subjetivo y resultó especialmente sólido en las categorías

de “Satisfecho” y “Muy satisfecho”; sin embargo, presenta limitaciones en la clasificación de estudiantes “Muy insatisfechos”, evidenciando la dificultad de capturar patrones en los extremos negativos del bienestar (Figura 2).

Tabla 3. Métricas del desempeño de los tres algoritmos.

Métricas	Escenario 1			Escenario 2		
	Logistic	SimpleLogistic	SMO	Logistic	SimpleLogistic	SMO
Exactitud	98.69%	99.65%	97.38%	99.63%	99.72%	97.95%
Precisión	98.80%	99.60%	97.30%	99.60%	99.70%	99.70%
Recall	98.70%	99.60%	97.40%	99.60%	99.70%	99.70%
F-measure	98.70%	99.60%	97.30%	99.50%	99.70%	99.70%
Kappa Statistic	0.9769	0.9938	0.9531	0.9935	0.9951	0.9634
ROC	0.998	0.999	0.984	1	1	1
MCC	98.20%	99.50%	95.80%	99.50%	99.60%	99.60%

El modelo *SimpleLogistic*, al incorporar boosting, alcanzó el mejor desempeño global con una precisión del 99.6% y un índice *Kappa* de 0.994. Este algoritmo demostró gran robustez frente a la variabilidad en los datos, y logró mejorar la clasificación en categorías minoritarias. Destaca por su capacidad para identificar correctamente a estudiantes “muy insatisfechos” con un recall de 99.6%, lo que lo convierte en el modelo más equilibrado y confiable para representar la diversidad de percepciones sobre el bienestar; por otro lado, el modelo SMO (*Support Vector Machines*) ofreció una perspectiva distinta, orientada a captar relaciones no lineales entre los factores asociados a la felicidad. Aunque su precisión global fue elevada (97,3%) mostró un rendimiento inferior en categorías intermedias y minoritarias como “Neutral” y “Muy insatisfecho”. Esto sugiere, que en este caso, los datos se ajustan mejor a modelos lineales robustos que a aproximaciones basadas en la separación en espacios de alta dimensión.

Escenario 2: Balanceo de clases.

El ajuste de parámetros, la normalización de los datos, y el balanceo de clases permitieron mejorar significativamente el desempeño de los modelos de clasificación supervisada (Tabla 3). En términos generales, los tres algoritmos alcanzaron niveles de exactitud superiores al 99%, lo que confirma que la

combinación de normalización y balanceo de clases favorece la estabilidad y la capacidad predictiva de los modelos.

El modelo de regresión logística obtuvo una exactitud de 99,63% y un índice *Kappa* de 0,9935, lo que refleja una concordancia casi perfecta entre las predicciones y las categorías reales. Este resultado confirma la solidez del enfoque lineal y su utilidad para interpretar los factores asociados al bienestar subjetivo, aunque se mantiene ligeramente por debajo del desempeño alcanzado por *SimpleLogistic* que se posicionó como el modelo más robusto y equilibrado en este escenario con una exactitud de 99.72% y un *Kappa* de 0.9951. El uso de boosting mejoró tanto la precisión como el recall, lo que evidenció un rendimiento superior en todas las métricas. Este hallazgo sugiere, que cuando los datos se normalizan y las clases se balancean, *SimpleLogistic* capta mejor la diversidad de patrones en el bienestar subjetivo de los estudiantes.

En contraste, el modelo SMO, aunque alcanzó valores altos de precisión y recall (99,7%), presentó una exactitud global menor (97,95%) y un *Kappa* de 0,9634. Esto indica, que pese a su capacidad para explorar relaciones no lineales, en este escenario los datos respondieron mejor a modelos lineales optimizados. El balanceo de clases y la normalización favorecieron su desempeño, pero no lograron compensar por completo la menor concordancia con Logistic y SimpleLogistic.

En síntesis, los resultados muestran que el preprocesamiento adecuado, como la normalización y el balanceo de clases, potencia el rendimiento de los modelos lineales, confirmando que SimpleLogistic es la alternativa más confiable y precisa para modelar la felicidad y el bienestar subjetivo en universitarios mexicanos, mientras que Logistic aporta interpretabilidad y el SMO ofrece un contraste metodológico para explorar patrones complejos.

Escenario 3: Comportamiento de los modelos.

A partir de los resultados obtenidos en el *Escenario 2* (Figura 1), se identificó que el algoritmo con mejor desempeño para clasificar el bienestar subjetivo de los estudiantes universitarios mexicanos fue

SimpleLogistic, seguido muy de cerca por la regresión logística; no obstante, para comprender con mayor detalle cómo cada modelo logró distinguir entre las categorías de bienestar, tanto mayoritarias como minoritarias, resulta fundamental analizar la información contenida en las matrices de confusión (Figura 2). Estas permiten observar no solo el nivel de acierto global, sino también cómo cada algoritmo enfrentó los retos de clasificación en categorías menos representadas, como “Muy insatisfecho”, en contraste con las más frecuentes, como “Satisfecho” y “Muy satisfecho”.

Figura 2. Matriz de confusión de cada algoritmo.

	Logistic					
	a	b	c	d	e	
a= Satisfecho	3676	13	9	4	1	a= Satisfecho
b= Muy satisfecho	9	2450	0	0	0	b= Muy satisfecho
c= Neutral	12	0	497	11	4	c= Neutral
d= Insatisfecho	0	0	12	79	7	d= Insatisfecho
e= Muy insatisfecho	0	0	0	7	2	e= Muy insatisfecho

	SimpleLogistic					
	a	b	c	d	e	
a= Satisfecho	3699	1	1	1	1	a= Satisfecho
b= Muy satisfecho	7	2453	0	0	0	b= Muy satisfecho
c= Neutral	0	0	524	0	0	c= Neutral
d= Insatisfecho	0	0	10	88	0	d= Insatisfecho
e= Muy insatisfecho	0	0	0	3	6	e= Muy insatisfecho

	SMO					
	a	b	c	d	e	
a= Satisfecho	3690	6	4	3	0	a= Satisfecho
b= Muy satisfecho	10	2449	0	0	0	b= Muy satisfecho
c= Neutral	89	0	433	2	0	c= Neutral
d= Insatisfecho	0	0	22	75	1	d= Insatisfecho
e= Muy insatisfecho	0	0	0	2	7	e= Muy insatisfecho

Como se observa, los tres modelos presentan una diagonal principal sólida, lo que indica una precisión global elevada, superior a 99% (Tabla 3). Aunque también revela con detalle, que el algoritmo que mejor clasifica a casi todas las categorías de la clase es *SimpleLogistic*, en las clases minoritarias como d (Insatisfecho) y e (Muy insatisfecho), logra capturar más aciertos (88 y 6 respectivamente) con menos confusiones que los otros dos; en tanto, que *logistic* muestra un buen desempeño, pero tiene ligeras dificultades para distinguir entre "Insatisfecho" y "Muy insatisfecho" (confunde 7 casos de 'e' con 'd'). Aunque SMO es robusto, presenta ruido en la clase c (Neutral), donde clasifica erróneamente 89 casos como la clase a (Satisfecho). Esto sugiere, que para el SMO, la frontera de decisión entre "Neutral" y "Satisfecho" es borrosa.

Para comprender el comportamiento de las clases en los tres algoritmos se observa lo siguiente:

- Las categorías a y b (Satisfecho y Muy satisfecho) son identificadas casi a la perfección por los tres modelos. Esto ocurre, porque incluso con el balanceo aplicado, estas clases presentan características muy distintivas en las encuestas de bienestar, lo que facilita su clasificación.
- La categoría c (Neutral), el modelo SimpleLogistic muestra el desempeño más consistente con 524 aciertos, mientras que SMO es el que más errores comete, desviando varios casos hacia la categoría “Satisfecho”.
- Finalmente, en las categorías d y e (Insatisfecho y Muy insatisfecho) se revela la verdadera capacidad de cada algoritmo. El SimpleLogistic alcanza una precisión casi perfecta en la clase “Muy insatisfecho”, mientras que el modelo Logistic apenas identifica dos casos correctamente en esa misma categoría, lo que evidencia sus limitaciones frente a clases minoritarias.

El análisis de los tres clasificadores, bajo condiciones de normalización y balanceo de datos, revela un desempeño sobresaliente en los algoritmos evaluados con niveles de exactitud superiores al 97,9%; por ejemplo, el algoritmo SimpleLogistic se posiciona como el más robusto, alcanzando la mayor exactitud (99,72%) y un *Kappa Statistic* de 0,9951. Su matriz de confusión respalda estas cifras al mostrar una clasificación casi perfecta en las categorías de mayor vulnerabilidad como "Insatisfecho" (88 aciertos) y "Muy insatisfecho" (6 aciertos) con un margen de error mínimo en comparación con sus pares.

También el algoritmo Logistic presenta métricas de precisión y recall de 99,60%. Si bien su desempeño es alto, la matriz de confusión revela una ligera dificultad para discriminar entre los grados extremos de insatisfacción con 7 casos de la clase "e" clasificados erróneamente en la clase "d". Al mismo tiempo, el algoritmo SMO muestra limitaciones específicas aun a pesar de tener un ROC de 1 y un MCC de 99,60%, registrando la exactitud más baja del grupo (97,95%). Esto se atribuye directamente a la confusión detectada en la matriz, en la que 89 individuos con percepción "Neutral" fueron clasificados erróneamente como "Satisfechos". Este hecho sugiere, que, en términos de variables (ingresos, salud mental, eudemonía,

dominios de satisfacción), los individuos "neutrales" comparten un perfil casi idéntico al de los "satisfechos".

En escalas como ésta (categorías de la clase), la frontera entre sentirse "ni bien ni mal" (Neutral) y "Bien" (Satisfecho) suele ser difusa. El SMO, al ser un modelo de vectores de soporte, parece encontrar una frontera de decisión que obliga a clasificar los indecisos en la categoría positiva, lo que podría reflejar una tendencia natural en los datos, según la cual lo neutro tiende a lo satisfactorio en la vida académica de los adultos. Este comportamiento indica, que para la población mayor de 18 años que asiste a la escuela, la diferencia real entre estar en un estado "Neutral" o "Satisfecho" no está determinada por las variables cuantitativas del estudio, sino que posiblemente lo está determinada por factores cualitativos o emocionales, que el modelo SMO no logra separar con la misma precisión que la regresión logística simple.

Por último, la alta correlación de Matthews ($MCC > 99,5\%$) en todos los modelos confirma que el tratamiento previo de los datos permitió identificar las clases minoritarias (insatisfacción) con la misma eficacia que las mayoritarias. La baja tasa de error indica que las características del conjunto de datos (posiblemente ingresos, salud mental, bienestar económico, eudemonía) están muy bien alineadas con la percepción de satisfacción del individuo; de esta manera, se determina que el algoritmo SimpleLogistic es el modelo ideal para predecir el bienestar subjetivo en esta población, dado que maximiza la concordancia (Kappa) y minimiza los falsos positivos en las categorías neutrales y negativas.

Modelo de la felicidad.

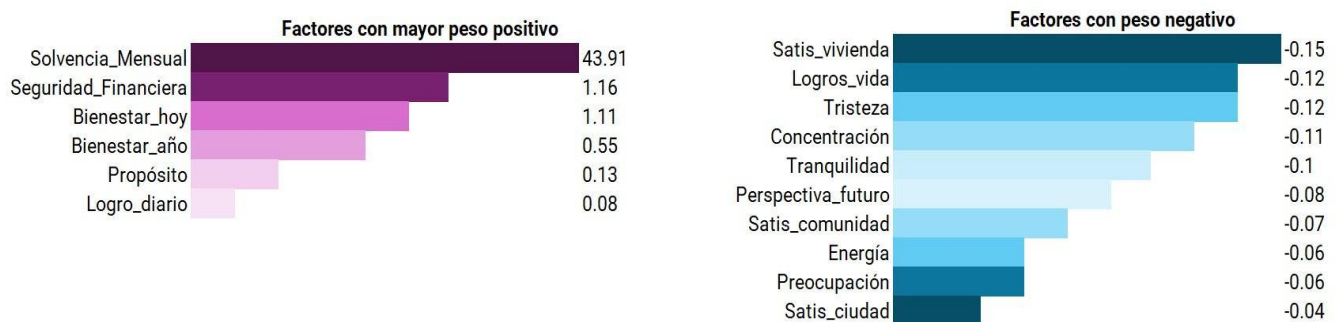
El modelo de predicción de la felicidad, que aplica el algoritmo SimpleLogistic, se destaca por su capacidad para equilibrar la precisión estadística con la interpretabilidad de los factores que subyacen al bienestar autorreportado. En el modelo generado para clasificar la probabilidad de pertenecer a la clase "Satisfactorio", el algoritmo ha seleccionado un conjunto específico de variables que actúan como predictores, asignándoles coeficientes que determinan su peso en el bienestar subjetivo (Modelo 2). Un

coeficiente positivo aumenta la probabilidad de estar en la clase, mientras que un coeficiente negativo la disminuye.

[Modelo 2] Satisfactorio= $-9.06 + [\text{Bienestar_hoy}] * 1.11 + [\text{Bienestar_año}] * 0.55 + [\text{Bienestar_personal}] * 0.08 + [\text{Propósito}] * 0.13 + [\text{Logro_diario}] * 0.08 + [\text{Tranquilidad}] * -0.1 + [\text{Energía}] * -0.06 + [\text{Concentración}] * -0.11 + [\text{Preocupación}] * -0.06 + [\text{Tristeza}] * -0.12 + [\text{Logros_vida}] * -0.12 + [\text{Perspectiva_futuro}] * -0.08 + [\text{Satis_vivienda}] * -0.15 + [\text{Satis_ciudad}] * -0.04 + [\text{Satis_comunidad}] * -0.07 + [\text{Solvencia_mensual}] * 43.91 + [\text{Seguridad_financiera}] * 1.16$

El Modelo 2 define el bienestar subjetivo autorreportado como “Satisfactorio” y revela una estructura clara de cómo se configura la percepción de satisfacción muy alta. Los factores positivos incluyen la Solvencia_Mensual, Seguridad_Financiera, Bienestar_hoy, Bienestar_año, Propósito, Logro_diario. Estos elementos actúan como impulsores clave de la satisfacción con especial énfasis en el bienestar diario y en la estabilidad económica; por otro lado, los factores negativos del modelo 2, como la tristeza, la preocupación, el cansancio, el mal humor, y la insatisfacción con la vivienda, la ciudad y la comunidad, indican que los estados emocionales adversos y el entorno físico-social pueden erosionar la percepción de satisfacción (Figura). Aunque su impacto es menor que el de los factores positivos, su presencia constante sugiere que el malestar emocional y la falta de conexión con el entorno deben abordarse para evitar que interfieran en el bienestar general (Figura 3).

Figura 3. Coeficientes del Modelo 2 Clase categórica “satisfactorio”.



En síntesis, el modelo sugiere que la satisfacción muy alta en este grupo poblacional se construye a partir del bienestar subjetivo y de la estabilidad financiera, pero también requiere atención a las emociones negativas y al entorno cotidiano. En síntesis, para elevar la satisfacción, se debe fortalecer el bienestar emocional diario, asegurar condiciones económicas estables, y mejorar la calidad del entorno físico y social. El modelo 2 ofrece una guía valiosa para diseñar intervenciones educativas, comunitarias y psicológicas, que promuevan una experiencia de vida más satisfactoria para quienes estudian en la edad adulta.

En la Figura se presentan los modelos correspondientes a las otras categorías de las clases, en los que se destaca que las variables Bienestar_hoy y Bienestar_año constituyen la cúspide del bienestar autorreportado.

Figura 4. Modelos de la felicidad en sus diversas categorías.

[Modelo 1] Muy_satisfactorio = 57.42 + [periodo] * -0.11 + [Bienestar_hoy] * 8.93 + [Bienestar_año] * 8.99 + [Optimismo_proyectivo] * -0.2 + [Misión_personal] * -0.2 + [Resiliencia_emocional] * -0.06 + [Buen_humor] * 0.16 + [Tranquilidad] * 0.08 + [Energía] * 0.18 + [Concentración] * 0.06 + [Mal_humor] * 0.28 + [Preocupación] * 0.2 + [Cansancio] * 0.25 + [Nivel_vida] * 0.15 + [Salud_física] * 0.14 + [Logros_vida] * 0.08 + [Relaciones_personales] * 0.22 + [Perspectiva_futuro] * 0.15 + [Satis_Tiempo] * 0.13 + [Satis_vecindario] * 0.3

[Modelo 4] Insatisfecho = 52.66 + [ocupación] * 0.1 + [Bienestar_hoy] * -4.34 + [Bienestar_año] * -5.07 + [Resiliencia_emocional] * 0.07 + [Satis_vecindario] * -0.1 + [Soledad_percibida] * -24.18

[Modelo 3] Neutral = 36.07 + [ocupación] * -0.03 + [ingresos] * -0 + [Bienestar_hoy] * -2.32 + [Bienestar_año] * -2.99 + [Propósito] * 0.06 + [Identidad_afortunada] * -0.08 + [Sin_interés] * 0.05 + [Nivel_vida] * -0.06 + [Logros_vida] * -0.06 + [Satis_vivienda] * -0.06 + [Satis_pais] * -0.05 + [Solvenica_mensual] * -4.79 + [Seguridad_financiera] * -0.44

[Modelo 5] Muy_instatisfecho = 591.88 + [periodo] * -0.26 + [Bienestar_hoy] * -6.67 + [Bienestar_año] * -6.26 + [Resiliencia_emocional] * -0.28 + [Tranquilidad] * -0.5 + [Cansancio] * -0.19 + [Tristeza] * -0.24 + [Nivel_vida] * -0.19 + [Satis_pais] * 0.51 + [Soledad_percibida] * -18.1 + [Seguridad_financiera] * -2.12

Al analizar los cinco modelos obtenidos, se observa un patrón consistente en el bienestar subjetivo autorreportado de personas mayores de 18 años que asisten a la escuela. En los niveles altos de satisfacción, el bienestar_hoy y el Bienestar_año aparecen como los predictores más fuertes y positivos; es decir, quienes reportan sentirse bien hoy y haber tenido un buen año son los que más tienden a ubicarse en la categoría de “muy satisfactorio o Satisfactorio”. Este hallazgo confirma que la percepción subjetiva inmediata y la evaluación retrospectiva del bienestar constituyen el núcleo de la felicidad en este grupo.

En contraste, en los niveles de insatisfacción y de muy insatisfecho, el bienestar subjetivo se convierte en un predictor negativo. Los coeficientes muestran que cuando las personas reportan un bajo bienestar hoy o a lo largo del año, la probabilidad de insatisfacción aumenta significativamente; además, en estos modelos, la soledad percibida emerge como un factor crítico con un peso muy elevado, lo que indica que la falta de vínculos sociales es uno de los principales determinantes de la insatisfacción de las personas mayores de 18 años que asisten a la escuela.

El modelo 3, que representa la categoría neutral, refuerza la idea de que el bienestar subjetivo también se asocia con coeficientes negativos, así como con una menor solvencia económica y una menor seguridad financiera. Esto sugiere que quienes se ubican en un nivel neutral tienden a experimentar una combinación de carencias materiales y emocionales que impiden que su experiencia escolar se traduzca en una satisfacción plena.

En conjunto, los cinco modelos muestran que el bienestar subjetivo es el eje central de la experiencia de satisfacción. Cuando es alto, se asocia directamente con la felicidad; cuando es bajo, se convierte en el principal predictor de la insatisfacción. La solvencia económica y la seguridad financiera refuerzan la satisfacción, mientras que la soledad percibida y las emociones negativas (tristeza, preocupación, cansancio) erosionan el bienestar. En suma, para promover la felicidad en adultos que estudian, es necesario fortalecer el bienestar emocional diario, garantizar condiciones económicas estables, y sobre todo, construir redes sociales que reduzcan la soledad y refuercen el sentido de pertenencia.

CONCLUSIONES.

El estudio abordó la dificultad de identificar con precisión las clases minoritarias de insatisfacción en el bienestar subjetivo de estudiantes universitarios mexicanos; un aspecto que suele quedar invisibilizado en los análisis tradicionales. Se planteó responder qué factores predicen mejor la felicidad en este grupo y qué modelo ofrece la mayor capacidad para clasificar los distintos niveles de satisfacción.

Los resultados muestran que el algoritmo SimpleLogistic alcanzó el mejor desempeño global con una exactitud de 99.72% y un índice Kappa de 0.9951, destacando por su capacidad para clasificar correctamente categorías minoritarias como “Insatisfecho” y “Muy insatisfecho”. La regresión logística también obtuvo métricas altas y aportó interpretabilidad, aunque presentó limitaciones en los extremos negativos del bienestar.

El modelo SMO ofreció una perspectiva no lineal, aunque presentó dificultades para clasificar categorías intermedias como “Neutral”. Los hallazgos confirman que los modelos lineales optimizados son más confiables para este tipo de datos y que la felicidad en personas universitarias no es solo una variable individual medible, sino un proceso relacional, territorial y situado, atravesado, entre otras cosas, por violencias estructurales, economías del miedo y precarias, pero también por formas de resistencia y cuidado en los territorios.

En este sentido, desde una perspectiva situada, la vivencia del bienestar de personas mayores de 18 años que continúan su formación académica no puede desvincularse de condiciones como la precariedad económica, la violencia estructural, la desigualdad territorial, las violencias extremas o las formas de exclusión o expulsión social que atraviesan la vida cotidiana en diversos contextos. Es ahí donde las variables como la solvencia económica, la estabilidad emocional o la percepción de propósito de vida no solo operan como predictores individuales, sino también como expresiones de configuraciones sociales más amplias.

Las implicaciones de este trabajo son relevantes para la investigación social y educativa, pues demuestran que el aprendizaje supervisado puede equilibrar precisión y explicabilidad, lo que aporta insumos prácticos para diseñar políticas y programas que promuevan la felicidad entre los universitarios. Los modelos predictivos permiten atender tanto a las mayorías satisfechas como a las minorías insatisfechas, lo que abre la posibilidad de intervenciones más inclusivas en los ámbitos educativo, comunitario y psicológico.

Entre las limitaciones se encuentra la disponibilidad parcial de variables en el BIARE, especialmente en dominios como la salud mental y el bienestar económico, que solo están completos a partir del año 2025; además, el estudio se limita a estudiantes universitarios mexicanos, lo que limita la generalización a otros contextos. Finalmente, la frontera entre categorías como “Neutral” y “Satisfecho” sigue siendo difusa, lo que refleja la complejidad subjetiva del fenómeno.

Las líneas de investigación futuras incluyen ampliar el análisis a otras poblaciones jóvenes y adultas en formación académica fuera de México, incorporar variables longitudinales completas para fortalecer la capacidad explicativa de los modelos, explorar algoritmos híbridos que combinen la interpretabilidad y la detección de patrones no lineales, así como integrar análisis cualitativos que complementen la visión estadística.

En concreto, la investigación demuestra que la felicidad de los universitarios puede modelarse con alta precisión mediante algoritmos supervisados, siendo SimpleLogistic la alternativa más confiable para capturar tanto las mayorías satisfechas como las minorías insatisfechas. Este hallazgo confirma que los algoritmos de aprendizaje automático derivados de la inteligencia artificial, aplicados con rigor metodológico, pueden convertirse en una herramienta clave para comprender y promover el bienestar subjetivo en la educación superior, y que además, los hallazgos invitan a desbordar el valor analítico del enfoque fuera de horizontes más amplios, en los que la comprensión del bienestar subjetivo implique no solo su medición, sino también la problematización de las condiciones que lo hacen posible o lo limitan.

REFERENCIAS BIBLIOGRÁFICAS.

1. Cervera, C., Martí, M. M., y Alejo, A. (2024). El bienestar subjetivo de los estudiantes: Un componente para considerar en la evaluación curricular. Congreso Internacional de Educación. Centro de Investigación Educativa, Universidad Autónoma de Tlaxcala. <https://cie.uatx.mx/debates-en-evaluacion-y-curriculum/pdf2024/C143.pdf>

2. Chen, Y., Wu, X., Hu, A., et al. (2021). Social prediction: A new research paradigm based on machine learning. *Journal of Chinese Sociology*, 8, 15. <https://doi.org/10.1186/s40711-021-00152-z>
3. Christensen, R. (2025). Log-linear models and logistic regression. Springer Nature Switzerland. <https://link.springer.com/book/10.1007/978-3-031-69038-9>
4. Codina-Lorente, M. del P., Gil-Madrona, P., y Gómez B. I. M. (2025). Influence of self-perceived affection on school performance by primary education students. *SAGE Open*. <https://doi.org/10.1177/21582440251358555>
5. Contreras L., G. (2024). El bienestar psicológico de estudiantes de nivel superior: una revisión sistemática. *Ciencias Sociales y Humanas*, 8(4), 5755-5774. https://doi.org/10.37811/cl_rcm.v8i4.12798
6. Diener, E. (1984). Subjective well-being. *Psychological Bulletin*, 95(3), 542–575. https://labs.psychology.illinois.edu/~ediener/Documents/Diener_1984.pdf
7. Forero-Corba, W., y Negre B. F. (2024). Técnicas y aplicaciones del machine learning e inteligencia artificial en educación: una revisión sistemática. *RIED-Revista Iberoamericana de Educación a Distancia*, 27(1), 209–253. <https://revistas.uned.es/index.php/ried/article/view/37491>
8. Instituto Nacional de Estadística y Geografía (INEGI). (2025). Módulo de bienestar autorreportado. <https://www.inegi.org.mx/programas/biare/>
9. Jurafsky, D., y Martin, J. H. (2026). Speech and language processing (Chap. 4, “Logistic regression and text classification”). Pearson. <https://web.stanford.edu/~jurafsky/slp3/4.pdf>
10. Martínez P., J. A., y Pérez M., P. S. (2024). Regresión logística. *Semergen*, 50(1), 102086. <https://doi.org/10.1016/j.semereg.2023.102086>
11. Millán, R., y Castellanos, R. (2018). Bienestar subjetivo en México. UNAM. https://ru.iis.sociales.unam.mx/bitstream/IIS/5445/7/bienestar_subjetivo.pdf

12. Ryff, C. D. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of Personality and Social Psychology*, 57(6), 1069–1081.
<https://doi.org/10.1037/0022-3514.57.6.1069>
13. Sabry, D. (2023). Support vector machine: Enhancing machine learning algorithms for intelligent robotics (Vol. 59). One Billion Knowledgeable.
14. Seligman, M. E. P. (2011). *Flourish: A visionary new understanding of happiness and well-being*. Free Press.
<https://www.amazon.com/Flourish-Visionary-Understanding-Happiness-Well-being/dp/1439190763> <https://www.scirp.org/reference/referencespapers?referenceid=3868065>
15. Taboada, M. (2024). La desinformación desde el análisis computacional del discurso. Jornadas “La lingüística frente a la amenaza de la desinformación”, Universidad de Navarra.
https://www.sfu.ca/~mtaboada/docs/Taboada_UNavarra_2024.pdf
16. Whig, P., Silva, N., Elongar, A. A., y Sharma, P. (2024). Sustainable development through machine learning, AI and IoT: Second International Conference, ICSD 2024, virtual event, April 27–28, 2024, proceedings. Springer Nature Switzerland. <https://link.springer.com/book/10.1007/978-3-031-71729-1>

DATOS DE LA AUTORA.

1. **Argelia Berenice Urbina-Nájera.** Doctorado en Planeación Estratégica y Dirección de Tecnología. Universidad Iberoamericana Puebla. Responsable de Posgrados. México. Correo electrónico: abunajera@gmail.com

RECIBIDO: 18 de mayo del 2026.

APROBADO: 9 de junio del 2026.