



ANÁLISIS COMPARATIVO DE TÉCNICAS DE CLASIFICACIÓN BINARIA APLICADAS A LA APROBACIÓN DE CRÉDITOS

Verónica Paz Pinto Ibáñez, Universidad Diego Portales, veronica.pinto_i@mail.udp.cl
Alejandra León Zapata, Universidad Diego Portales, Alejandra.leon@udp.cl

RESUMEN

Este estudio presenta un análisis comparativo entre dos enfoques de aprendizaje supervisado utilizados para predecir la aprobación de créditos bancarios: la regresión logística y una red neuronal multicapa. A partir de una base de datos simulada compuesta por 3000 observaciones y múltiples variables financieras, sociodemográficas, derivadas y ambientales, se entrenaron ambos modelos utilizando técnicas de balanceo de clases, codificación categórica y escalamiento. La red neuronal fue probada con distintas arquitecturas ocultas (compacta, equilibrada y profunda), mientras que la regresión logística fue entrenada sobre el mismo conjunto de datos para asegurar una comparación equitativa. Las métricas de rendimiento obtenidas, como la exactitud y la matriz de confusión, mostraron que la red neuronal superó consistentemente a la regresión logística en su capacidad predictiva. Así, los resultados evidencian el potencial de las redes neuronales para capturar patrones complejos y no lineales en contextos de evaluación crediticia, lo que las posiciona como una herramienta robusta y eficaz en la toma de decisiones bancarias.

PALABRAS CLAVE: Aprendizaje supervisado, red neuronal, regresión logística, predicción crediticia, clasificación binaria.

INTRODUCCIÓN

La inteligencia artificial (IA), se ha utilizado en las últimas décadas para resolver problemas en ciencias e ingeniería, en particular para optimizar diseños en ingeniería civil (Jiang & Adeli, 2008; Zhang, Wang & Wang, 2018), en diagnóstico médico y bioinformática (Esteva et al., 2017; Libbrecht & Noble, 2015), modelado climático y ambiental (Reichstein et al., 2019), robótica y automatización (Kober, Bagnell & Peters, 2013) y ciencias de los materiales (Raccuglia et al., 2016). En la actualidad, con la masificación y democratización de estas herramientas, en nuestra Facultad de Ingeniería y Ciencias, hemos incorporado algunas metodologías basadas en IA, en cursos formales y en instancias finales del proceso de formación de nuestros ingenieros. En esta tesis de grado de magíster en Ciencias de la Ingeniería, hemos implementado un modelo basado en redes neuronales para abordar un problema complejo y de alto impacto práctico: la evaluación del riesgo crediticio.

La evaluación del riesgo crediticio constituye un elemento clave en la gestión financiera de las instituciones bancarias, ya que permite reducir pérdidas por morosidad y mejorar la asignación de recursos. Tradicionalmente, este proceso ha sido abordado mediante modelos estadísticos como la regresión logística, debido a su eficiencia, simplicidad e interpretabilidad. No obstante, con el avance de la inteligencia artificial y la ciencia de datos, han surgido enfoques más



XXXVII CONGRESO CHILENO DE EDUCACIÓN EN INGENIERÍA 2025

PROYECCIÓN DE LAS TECNOLOGÍAS DIGITALES EN LA FORMACIÓN EN INGENIERÍA:
LA EDUCACIÓN EN MODALIDAD PRESENCIAL, HÍBRIDA Y VIRTUAL

Concepción, 8 al 10 de octubre 2025

sofisticados capaces de ofrecer un mayor poder predictivo, entre ellos las redes neuronales artificiales.

Este estudio compara el desempeño de una red neuronal multicapa con el de una regresión logística en la predicción de aprobación de créditos, utilizando una base de datos simulada de 3000 observaciones que incluye variables financieras, sociodemográficas, derivadas y ambientales. La base fue construida a partir de criterios provenientes de fuentes oficiales del sistema financiero chileno (Comisión para el Mercado Financiero, 2023, 2024; Banco Central de Chile, 2022), lo que refuerza su validez y aplicabilidad en contextos reales. Dado que los factores que influyen en la aprobación crediticia están profundamente interrelacionados, se hace especialmente relevante el uso de modelos capaces de representar relaciones no lineales y adaptarse a estructuras complejas de datos. Así, este análisis busca determinar cuál de los dos enfoques entrega mejores resultados en tareas de clasificación binaria, aportando evidencia empírica útil para investigadores y profesionales del área.

DESARROLLO

• BASE DE DATOS Y VARIABLES

El presente estudio se fundamenta en una base de datos simulada, elaborada con el objetivo de representar de manera realista el comportamiento de solicitudes de crédito en el sistema bancario. La base contiene un total de 3000 observaciones, cada una correspondiente a un individuo que ha solicitado un crédito, junto con sus características personales, financieras y contextuales, en línea con criterios observados en encuestas nacionales como la EFH del Banco Central de Chile (Banco Central de Chile, 2022). La variable dependiente binaria, Aprobado_Complejo, indica si la solicitud fue aprobada (1) o rechazada (0), de acuerdo con criterios multifactoriales.

Las variables independientes fueron agrupadas en cuatro categorías principales, lo que permitió abordar el problema desde una perspectiva integral:

- **VARIABLES FINANCIERAS:** Ingreso mensual, nivel de endeudamiento, historial de pagos, antigüedad laboral, score crediticio, monto solicitado, número de cuotas solicitadas y presencia de garantía. Estas variables permiten caracterizar el perfil financiero del solicitante y su capacidad de pago, considerando elementos analizados en estudios recientes sobre sostenibilidad del endeudamiento en Chile (Lemus & Pulgar, 2023).
- **VARIABLES SOCIODEMOGRÁFICAS:** Edad, estado civil, nivel educacional, presencia de hijos y tipo de crédito solicitado. Se incluyen por su influencia en la estabilidad percibida del solicitante y el contexto social asociado.
- **VARIABLES DERIVADAS:** Se construyeron combinaciones no triviales a partir de las variables anteriores para capturar relaciones complejas, como ingreso por cuota, score ajustado por deuda, carga financiera por hijos, edad relativa, educación ponderada por score, cuotas por antigüedad, entre otras. Estas variables enriquecen la representación



del solicitante mediante relaciones no lineales que pueden ser relevantes para modelos más complejos.

- **Variable ambiental:** Se incorporó un factor denominado Shock_Económico, que simula el impacto de condiciones externas adversas, tales como crisis económicas, sobre la probabilidad de aprobación del crédito. Esto permite evaluar la robustez de los modelos ante condiciones macroeconómicas inciertas.

El diseño de esta base permite explorar el rendimiento de distintos modelos bajo condiciones cercanas a la realidad, incorporando tanto datos directamente observables como transformaciones complejas derivadas del análisis financiero y sociodemográfico.

● PREPROCESAMIENTO DE DATOS

Previo al entrenamiento de los modelos, se aplicó un proceso riguroso de preparación de los datos para asegurar su compatibilidad con los algoritmos y reducir posibles sesgos o distorsiones. El preprocesamiento se llevó a cabo en tres etapas principales:

- **Codificación de variables categóricas:** Las variables no numéricas, como estado civil, nivel educacional y tipo de crédito, fueron transformadas mediante codificación one-hot, generando columnas binarias para cada categoría. Esta técnica permite que los modelos procesen variables cualitativas sin asignar un orden implícito que no corresponde.
- **Escalamiento de variables numéricas:** Se aplicó escalamiento estándar (z-score) a las variables cuantitativas, normalizando su media en 0 y su desviación estándar en 1. Esto es especialmente relevante para la red neuronal, dado que la diferencia de magnitudes entre variables puede afectar la convergencia del entrenamiento.
- **Balanceo de clases:** Dado que la variable objetivo presentaba una distribución ligeramente desbalanceada entre casos aprobados y no aprobados, se aplicó una técnica de submuestreo aleatorio (undersampling) para asegurar que ambas clases estuvieran igualmente representadas durante el entrenamiento. Esta medida evita que el modelo favorezca la clase mayoritaria y garantiza un aprendizaje equilibrado.

Adicionalmente, el conjunto de datos fue dividido en muestras de distinto tamaño, 900 (30% - 70%), 1500 (50% - 50%) y 2100 observaciones (70% - 30%) para evaluar cómo varía el rendimiento de los modelos según la cantidad de datos disponibles. Este análisis permite identificar la sensibilidad de cada enfoque a distintos volúmenes de entrenamiento y su capacidad de generalización.

● MODELOS DE CLASIFICACIÓN

Para la predicción de la aprobación de créditos bancarios, se implementaron dos enfoques de clasificación binaria: una red neuronal multicapa (MLP) y una regresión logística. Ambos modelos fueron entrenados sobre los mismos datos preprocesados, y su configuración técnica se diseñó para permitir una comparación posterior en condiciones equivalentes.



1. RED NEURONAL MULTICAPA

La red neuronal empleada en este estudio corresponde a un perceptrón multicapa (MLP), modelo de aprendizaje supervisado ampliamente utilizado para clasificación no lineal. Esta arquitectura se basa en una estructura de capas conectadas que transforman las entradas mediante funciones de activación no lineales, permitiendo representar relaciones complejas entre variables.

Se evaluaron tres configuraciones arquitectónicas distintas:

1. Arquitectura compacta: tres capas ocultas con 16, 8 y 4 neuronas respectivamente.
2. Arquitectura equilibrada: dos capas ocultas con 64 y 32 neuronas.
3. Arquitectura profunda: tres capas ocultas con 128, 64 y 32 neuronas.

En todos los casos, se utilizó la función de activación ReLU en las capas ocultas y sigmoide en la capa de salida. El entrenamiento se realizó utilizando el optimizador Adam, con función de pérdida binary crossentropy. Se implementaron técnicas de early stopping para detener el entrenamiento cuando no se observaban mejoras en el conjunto de validación, y se trabajó con conjuntos de entrenamiento de distinto tamaño (900, 1500 y 2100 observaciones), lo que permitió analizar el efecto del volumen de datos sobre el comportamiento del modelo.

2. REGRESIÓN LOGÍSTICA

La regresión logística es un modelo estadístico clásico que permite estimar la probabilidad de ocurrencia de un evento binario a partir de un conjunto de variables explicativas. El modelo aplica una transformación logística a una combinación lineal de las variables de entrada, generando salidas en el intervalo $[0, 1]$, interpretables como probabilidades.

Este modelo fue implementado como un modelo de referencia (baseline), siguiendo buenas prácticas para su configuración. Se entrenó utilizando los mismos datos codificados y escalados que la red neuronal, con inclusión de términos de regularización para reducir el riesgo de sobreajuste. Asimismo, se aplicaron técnicas de validación cruzada y se evaluó el comportamiento del modelo bajo los mismos tamaños de muestra que en la red neuronal, asegurando la comparabilidad metodológica entre ambos enfoques.

Ambos modelos fueron configurados y entrenados bajo condiciones homogéneas, con el objetivo de garantizar una comparación rigurosa y metodológicamente válida. La elección de estas dos técnicas responde a su amplio uso en el ámbito financiero, tanto desde una perspectiva tradicional como lo es la regresión logística, hasta un enfoque más moderno y flexible como la red neuronal, cuya aplicación en el ámbito crediticio chileno ha sido explorada previamente por Salazar (2022).



- **EVALUACIÓN DEL RENDIMIENTO**

Para evaluar el desempeño de los modelos de clasificación implementados, se utilizaron métricas estándar comúnmente empleadas en problemas de clasificación binaria. Estas métricas permiten analizar tanto la capacidad predictiva global como el comportamiento específico en cada clase (aprobado o no aprobado), considerando el posible desbalance de clases en los datos.

Las métricas seleccionadas fueron las siguientes:

- **Precisión (Accuracy):** Representa el porcentaje total de predicciones correctas. Es útil como primera aproximación al rendimiento, aunque puede ser engañosa si las clases están desbalanceadas.
- **Precisión positiva (Precision):** Indica qué proporción de las predicciones positivas (aprobados) fueron realmente correctas. Es especialmente relevante cuando el costo de un falso positivo es alto.
- **Recall (Sensibilidad o Tasa de Verdaderos Positivos):** Mide la proporción de casos verdaderamente positivos que fueron correctamente identificados. Es útil en contextos donde es importante minimizar los falsos negativos.
- **F1-score:** Corresponde a la media armónica entre precisión y recall, equilibrando ambas métricas. Es particularmente útil cuando existe un desbalance entre clases.
- **Matriz de confusión:** Proporciona una visualización detallada del rendimiento del modelo, mostrando la cantidad de verdaderos positivos, falsos positivos, verdaderos negativos y falsos negativos. Esta herramienta permite interpretar los errores cometidos y entender el comportamiento del clasificador en ambos grupos.

Ambos modelos fueron evaluados utilizando métricas estándar sobre subconjuntos de validación, generados a partir de divisiones estratificadas del conjunto de datos para conservar la proporción original de clases. Además, se comparó su rendimiento en tres escenarios distintos según el tamaño del conjunto de entrenamiento (900, 1500 y 2100 observaciones), lo que permitió analizar la evolución del desempeño en función de la cantidad de datos disponibles y determinar la robustez de cada enfoque. Las métricas fueron calculadas sobre los conjuntos de prueba correspondientes, y se registraron para cada combinación de modelo, arquitectura (en el caso de la red neuronal) y tamaño de muestra. Esta evaluación cuantitativa constituye la base para el análisis comparativo que se presenta en la sección de resultados.

RESULTADOS

Tras entrenar ambos modelos en las condiciones descritas, se procedió a evaluar su desempeño utilizando las métricas definidas previamente. Los resultados obtenidos permiten analizar el comportamiento de cada modelo según su configuración y volumen de datos, y extraer conclusiones sobre su capacidad predictiva en el contexto de evaluación crediticia.



• **DESEMPEÑO POR ARQUITECTURA DE RED NEURONAL**

La evaluación del rendimiento de la red neuronal multicapa se realizó considerando tres configuraciones arquitectónicas: una compacta (16, 8 y 4 neuronas), una equilibrada (64 y 32 neuronas) y una profunda (128, 64 y 32 neuronas). Cada modelo fue entrenado sobre tres tamaños distintos de muestra (900, 1500 y 2100 observaciones), lo que permitió analizar tanto su precisión como su capacidad de generalización en función del volumen de datos.

Tabla 1: Comparación entre arquitecturas de redes neuronales.

Arquitectura	Accuracy	Precisión (Clase 1)	Recall (Clase 1)	F1-Score (Clase 1)	AUC - ROC
32, 16 y 8	0,95	0,88	0,97	0,93	0,9936
16, 8 y 4	0,96	0,90	0,99	0,94	0,9957
128, 64 y 32	0,95	0,90	0,98	0,94	0,9941

Fuente: Elaboración propia.

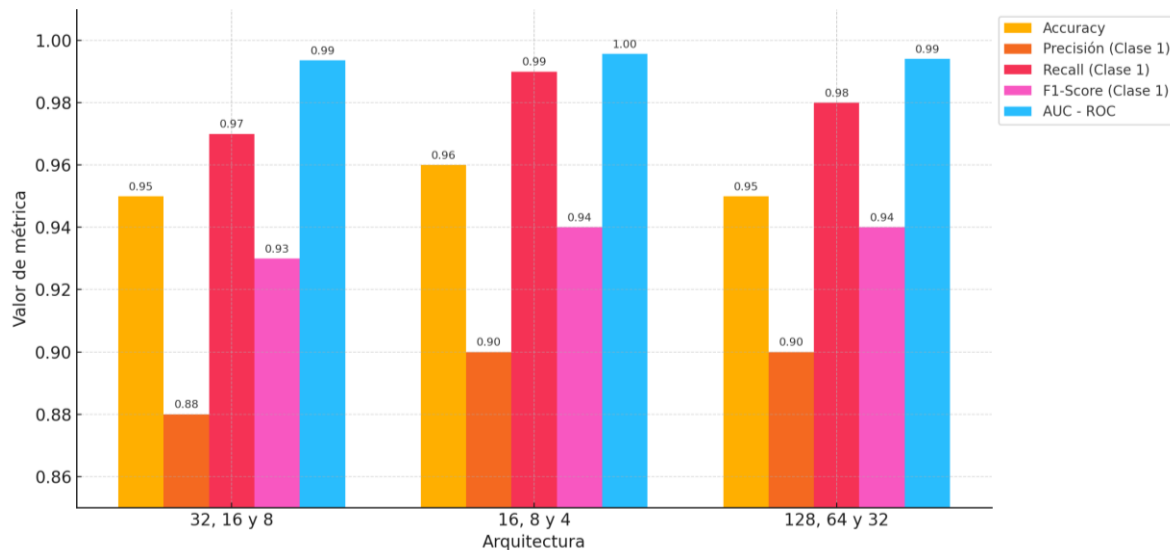


Figura 1: Gráfico comparativo de arquitecturas de redes neuronales.

Fuente: Elaboración propia

Los resultados obtenidos mostraron que la arquitectura compacta (16-8-4) fue la que presentó un desempeño más sólido y equilibrado a lo largo de todos los tamaños de muestra evaluados. Este modelo destacó por mantener un buen balance entre precisión, recall y F1-score, al tiempo que evidenció una estabilidad notable durante el proceso de entrenamiento, incluso al incrementarse el volumen de datos. A diferencia de las arquitecturas más complejas, la red compacta no mostró señales de sobreajuste, lo que sugiere una buena capacidad de generalización con menor carga computacional.

En contraste, la arquitectura profunda (128-64-32), si bien logró altos niveles de precisión en etapas iniciales, mostró un comportamiento inestable en conjuntos de menor tamaño, como el de 900 observaciones, lo que sugiere una sobre parametrización que la hizo más



vulnerable al sobreajuste. Por otra parte, la arquitectura equilibrada (64-32), ofreció un rendimiento intermedio, con mejoras progresivas al aumentar el tamaño del entrenamiento, pero sin alcanzar la solidez ni la consistencia observadas en la red compacta, especialmente en métricas como el recall. En conjunto, estos hallazgos resaltan cómo una arquitectura más sencilla puede ofrecer mejores resultados en tareas prácticas, siempre que logre capturar las relaciones relevantes sin incurrir en una complejidad innecesaria.

- **COMPARACIÓN ENTRE REDES NEURONALES Y REGRESIÓN LOGÍSTICA**

Al comparar directamente el desempeño de la red neuronal y la regresión logística bajo condiciones de entrenamiento equivalentes, se observó una ventaja sostenida a favor del modelo neuronal, tanto en precisión como en métricas asociadas al equilibrio entre clases. La red neuronal, especialmente en su configuración compacta (16-8-4), logró capturar relaciones complejas entre las variables de entrada que la regresión logística, por su naturaleza lineal, no fue capaz de modelar de forma efectiva. Esta diferencia fue particularmente evidente en conjuntos de entrenamiento de mayor tamaño, donde la red aprovechó mejor la información disponible para optimizar sus predicciones.

La regresión logística, en cambio, mostró un rendimiento razonablemente estable y aceptable, con mejoras proporcionales al aumento del tamaño de muestra. Sin embargo, su capacidad predictiva se vio limitada en comparación con la red neuronal, reflejando dificultades para adaptarse a interacciones no lineales o efectos cruzados entre variables, como aquellos introducidos por las variables derivadas y ambientales del conjunto de datos. A pesar de estas limitaciones, su principal fortaleza siguió siendo la interpretabilidad: la posibilidad de examinar coeficientes y comprender el peso relativo de cada variable en la predicción, lo que puede resultar valioso en contextos donde la trazabilidad del proceso de decisión es una prioridad.

A continuación se puede observar una tabla comparativa de los resultados obtenidos:

Tabla 2: Comparación entre modelos

Métrica	RN 900	RL 900	RN 1500	RL 1500	RN 2100	RL 2100
Accuracy	0,96	0,88	0,93	0,89	0,88	0,89
Precisión (Clase 1)	0,90	0,82	0,84	0,82	0,75	0,81
Recall (Clase 1)	0,99	0,82	0,98	0,85	0,94	0,87
F1-Score (Clase 1)	0,94	0,87	0,90	0,83	0,84	0,84
AUC - ROC	0,9957	0,9591	0,9871	0,9606	0,9627	0,9602

Fuente: Elaboración propia.



XXXVII CONGRESO CHILENO DE EDUCACIÓN EN INGENIERÍA 2025

PROYECCIÓN DE LAS TECNOLOGÍAS DIGITALES EN LA FORMACIÓN EN INGENIERÍA: LA EDUCACIÓN EN MODALIDAD PRESENCIAL, HÍBRIDA Y VIRTUAL

Concepción, 8 al 10 de octubre 2025



Figura 2: Gráfico comparativo entre red neuronal v/s regresión logística.
Fuente: Elaboración propia.

Como se observa en los resultados obtenidos, la red neuronal superó sistemáticamente a la regresión logística en todas las métricas y tamaños de muestra, demostrando una mayor capacidad para modelar relaciones complejas entre variables. Esta ventaja se vuelve especialmente relevante en escenarios como el de la aprobación crediticia, donde los factores involucrados interactúan de forma no lineal y multidimensional. Si bien la regresión logística sigue siendo una herramienta válida por su simplicidad y transparencia, particularmente útil como modelo base o explicativo, la red neuronal destacó por su precisión, adaptabilidad y capacidad de aprendizaje en contextos complejos. Estos hallazgos refuerzan la pertinencia de adoptar enfoques no lineales en la construcción de sistemas automatizados de decisión, especialmente cuando se dispone de una cantidad de datos suficiente para aprovechar su potencial predictivo.

CONCLUSIONES

Este estudio comparó el desempeño de una red neuronal multicapa y una regresión logística en la predicción de aprobación de créditos bancarios, utilizando una base simulada de 3000 observaciones con variables financieras, sociodemográficas, derivadas y ambientales. Ambos modelos fueron entrenados bajo condiciones metodológicas equivalentes y evaluados en distintos tamaños de muestra mediante métricas estándar de clasificación. Los resultados mostraron una clara superioridad de la red neuronal, especialmente en su arquitectura compacta (16-8-4), gracias a su capacidad para modelar relaciones no lineales y



XXXVII CONGRESO CHILENO DE EDUCACIÓN EN INGENIERÍA 2025

PROYECCIÓN DE LAS TECNOLOGÍAS DIGITALES EN LA FORMACIÓN EN INGENIERÍA:
LA EDUCACIÓN EN MODALIDAD PRESENCIAL, HÍBRIDA Y VIRTUAL

Concepción, 8 al 10 de octubre 2025

patrones complejos. Aunque la regresión logística destacó por su simplicidad e interpretabilidad, su rendimiento fue más limitado, funcionando mejor como modelo base.

Estos hallazgos refuerzan el valor de los modelos de aprendizaje profundo en la toma de decisiones crediticias, especialmente cuando se dispone de una cantidad suficiente de datos y se requiere alta precisión en la predicción. La arquitectura compacta, en particular, representa una alternativa eficiente en términos computacionales, lo que favorece su implementación en entornos de producción reales. Sin embargo, también se destaca que los modelos más complejos requieren monitoreo continuo, validación técnica y una adecuada gestión de riesgos algorítmicos para evitar decisiones opacas o no auditables.

Una de las principales limitaciones del estudio es el uso de una base de datos simulada, lo cual, si bien permite controlar las condiciones experimentales y probar escenarios específicos, reduce la capacidad de generalización de los resultados a contextos reales. Por ello, se propone como línea futura aplicar este enfoque a bases de datos provenientes de instituciones financieras reales, con el fin de evaluar su desempeño bajo condiciones más complejas, incluyendo la presencia de ruido, datos faltantes y distribuciones sesgadas. Asimismo, se sugiere incorporar variables temporales y secuenciales, así como explorar modelos híbridos o redes neuronales explicables que integren la capacidad predictiva de los métodos no lineales con la trazabilidad y transparencia requeridas en entornos regulados. En este sentido, resulta fundamental no solo mejorar el rendimiento técnico, sino también atender los aspectos éticos asociados a la implementación de inteligencia artificial en la toma de decisiones financieras. Dado que la evaluación crediticia automatizada impacta directamente en las oportunidades de las personas, todo modelo debe ser cuidadosamente auditado, validado y ajustado bajo principios de equidad, responsabilidad y no discriminación.

AGRADECIMIENTOS

Agradezco a mis papás, Angélica y Carlos, a la Mamá Sole, al Tata Mario, a la Monse, a la Yeye, a mi pololo Maximiliano, a mis amigos, a mi perrito Zac y a toda mi familia: gracias por estar, por sostenerme en cada paso y por hacerme sentir que nunca camino sola. A mis abuelos, que me cuidan desde el cielo, gracias por darme la fuerza para seguir adelante porque aunque no puedan conocer a la ingeniera en la que me estoy convirtiendo, sé que siempre supieron que lo lograría. Y gracias también a mi profesora de tesis, Alejandra León, por confiar en mí incluso cuando yo dudaba, su apoyo ha sido clave para que hoy este trabajo exista.

REFERENCIAS

Banco Central de Chile. (2022). *Encuesta Financiera de Hogares (EFH)*. Recuperado de <https://www.bcentral.cl/areas/encuestas-economicas/encuesta-financiera-de-hogares>

Comisión para el Mercado Financiero. (2023). *Informe de endeudamiento de los hogares: junio 2023*. Recuperado de https://www.cmfchile.cl/portal/principal/613/articles-77055_doc_pdf.pdf



XXXVII CONGRESO CHILENO DE EDUCACIÓN EN INGENIERÍA 2025
PROYECCIÓN DE LAS TECNOLOGÍAS DIGITALES EN LA FORMACIÓN EN INGENIERÍA:
LA EDUCACIÓN EN MODALIDAD PRESENCIAL, HÍBRIDA Y VIRTUAL
Concepción, 8 al 10 de octubre 2025

- Comisión para el Mercado Financiero.** (2024). Informe de endeudamiento de los hogares: junio 2024. Recuperado de https://www.cmfchile.cl/portal/principal/613/articles-89609_doc_pdf.pdf
- Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S.** (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115–118. <https://doi.org/10.1038/nature21056>
- Jiang, X., & Adeli, H.** (2008). Neural network-based adaptive control of nonlinear systems. *Engineering Applications of Artificial Intelligence*, 21(6), 805–819. <https://doi.org/10.1016/j.engappai.2007.11.004>
- Kober, J., Bagnell, J. A., & Peters, J.** (2013). Reinforcement learning in robotics: A survey. *The International Journal of Robotics Research*, 32(11), 1238–1274. <https://doi.org/10.1177/0278364913495721>
- Lemus, A., & Pulgar, C.** (2023). Endeudamiento máximo sostenible de los hogares en Chile. *Revista de Análisis Económico*, 38(1), 71–99. Recuperado de https://www.scielo.cl/scielo.php?pid=S0718-88702023000100071&script=sci_arttext
- Libbrecht, M. W., & Noble, W. S.** (2015). Machine learning applications in genetics and genomics. *Nature Reviews Genetics*, 16(6), 321–332. <https://doi.org/10.1038/nrg3920>
- Raccuglia, P., Elbert, K. C., Adler, P. D. F., Falk, C., Wenny, M. B., Mollo, A., ... & Norquist, A. J.** (2016). Machine-learning-assisted materials discovery using failed experiments. *Nature*, 533(7601), 73–76. <https://doi.org/10.1038/nature17439>
- Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat.** (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, 566(7743), 195–204. <https://doi.org/10.1038/s41586-019-0912-1>
- Salazar Vergara, J. G.** (2022). Diseño de un modelo predictivo para otorgar créditos. *Semestre Económico*, 24(57), 320–347. <https://doi.org/10.22395/seec.v24n57a15>
- Zhang, Y., Wang, J., & Wang, X.** (2018). Review on probabilistic forecasting of wind power generation. *Renewable and Sustainable Energy Reviews*, 32, 255–270. <https://doi.org/10.1016/j.rser.2014.01.033>