

Application of artificial intelligence in the management of a public university in Peru: A case of supervised machine learning using neural networks to classify if an engineering student would graduate in 5 years

Paul Tocto, Doctor¹, Gloria Teresita Huamani, Doctora², Eduardo Villacorta, Alumno³

¹Universidad Nacional de Ingeniería, Perú, ptocto@uni.edu.pe ²Universidad Nacional de Ingeniería, Perú, gloria.huamani.h@uni.edu.pe ³Universidad Nacional de Ingeniería, Perú, eduardo.villacorta.c@uni.pe

Abstract— Considering that Artificial intelligence is applied more frequently in educational management; this article presents the results of a case of supervised machine learning using neural networks to classify if an engineering student finishes his degree in five years at the Universidad Nacional de Ingeniería from Peru. The TensorFlow library is applied with a 5-layer neural network in the Google Collaboratory platform for training and validation, obtaining a classification model with an accuracy of 90%, which could help as a monitoring and follow-up tool for students with academic problems

Keywords—Machine Learning; Neural Network; Classification; graduate.

Digital Object Identifier (DOI):

<http://dx.doi.org/10.18687/LACCEI2022.1.1.565>

ISBN: 978-628-95207-0-5 **ISSN:** 2414-6390

Aplicación de Inteligencia artificial en la gestión de una universidad pública en el Perú: Un caso de Aprendizaje automático supervisado usando redes neuronales para clasificar si un alumno de Ingeniería egresaría en 5 años

Paul Tocto, Doctor¹, Gloria Teresita Huamani, Doctora², Eduardo Villacorta, Alumno³

¹Universidad Nacional de Ingeniería, Perú, ptocto@uni.edu.pe ²Universidad Nacional de Ingeniería, Perú, gloria.huamani.h@uni.edu.pe ³Universidad Nacional de Ingeniería, Perú, eduardo.villacorta.c@uni.pe

Resumen— Considerando que con mayor frecuencia se aplica la Inteligencia artificial en la gestión educativa, en este artículo se presenta los resultados de: un caso de Aprendizaje automático supervisado usando redes neuronales para clasificar si un alumno de Ingeniería terminaría su carrera en 5 años en la Universidad nacional de Ingeniería de Perú. Se aplica la librería Tensorflow con una red neuronal de 5 capas en la plataforma Colaboratory de Google para el entrenamiento y validación, obteniéndose un modelo de clasificación con una precisión aproximada de 90%, que podría ayudar como una herramienta de monitoreo y seguimiento de los alumnos con problemas académicos.

Palabras clave-- Aprendizaje automático, red neuronal, clasificación, egresar.

Abstract—Considering that Artificial intelligence is applied more frequently in educational management; this article presents the results of a case of supervised machine learning using neural networks to classify if an engineering student finishes his degree in five years at the Universidad Nacional de Ingeniería from Peru. The TensorFlow library is applied with a 5-layer neural network in the Google Collaboratory platform for training and validation, obtaining a classification model with an accuracy of 90%, which could help as a monitoring and follow-up tool for students with academic problems

Keywords—Machine Learning; Neural Network; Classification; graduate.

I. INTRODUCCIÓN

La educación pública, sobre todo las de nivel superior introducen en la gestión académica las buenas prácticas de la inteligencia de negocios, inteligencia artificial (IA). Para lo cual se van identificando soluciones en base a plataformas tecnológicas, también existen modelos, herramientas, técnicas de inteligencia artificial como el aprendizaje automático “machine learning” y estos permiten construir, crear modelos en la nube, porque da soporte de almacenamiento y cálculo de datos. Los resultados dan soporte a la toma de decisiones de las autoridades correspondientes [1]. La investigación sobre el

aprendizaje automático se desarrolla incrementalmente. En la solución de diversos problemas el aprendizaje automático proporciona confiabilidad y eficiencia. [2,3].

Asimismo, se tiene referencia que la UNESCO [4] “garantiza que la utilización de las tecnologías de la Inteligencia artificial (IA) en el contexto educativo esté regida por los principios fundamentales de inclusión y equidad”, además, “está decidida a ayudar a los Estados Miembros para que saquen provecho del potencial de las tecnologías de la IA con miras a la consecución de la Agenda de Educación 2030”. Por otra parte, la incorporación de la IA en los sistemas de gestión de las instituciones que administran los sistemas educativos permitiría aprovechar la enorme cantidad de datos disponibles en muchos países para predecir situaciones de riesgo y fortalecería las capacidades de monitoreo y toma de decisiones de las autoridades en los distintos niveles [5,6]. Este trabajo de investigación usa para entrenar a una red neuronal, la data de egresados de al menos diez años, que, según la duración de sus estudios, terminan en 5 años según la oferta académica o no. Los principales beneficiarios serían los tomadores de decisiones, los resultados obtenidos permitirían asignación de recursos para el acompañamiento y seguimiento de los ingresantes y garantizar una mayor tasa de titulación, presencia en el colegio de ingenieros del Perú y la sociedad peruana. Se aplica IA en un caso de Aprendizaje automático supervisado usando redes neuronales para clasificar si un alumno de Ingeniería termina su carrera en 5 años, la arquitectura de la red neuronal considerada en el presente estudio es de tres capas ocultas. Los resultados de este trabajo permitirían identificar a los alumnos que tienen riesgo académico y serviría de referencia para los programas de tutoría académica. También existen servicios en la nube que permiten el aprendizaje automático de las redes neuronales, en este trabajo se utiliza la plataforma Colaboratory de Google. La inteligencia artificial es una línea de investigación que se viene trabajando en la Facultad de Ingeniería Industrial y de Sistemas (FIIS) de la Universidad Nacional de Ingeniería (UNI), en el Perú. Algunos investigadores se dedican a la aplicación de modelos matemáticos en la gestión universitaria,

Digital Object Identifier (DOI):

<http://dx.doi.org/10.18687/LACCEI2022.1.1.12>

ISBN: 978-628-95207-0-5 ISSN: 2414-6390

a quienes les resulta fácil la implementación de proyectos de uso de datos [1].

Este artículo tiene la siguiente estructura: en la sección 2 se hace una revisión bibliográfica para dar cuenta del “aprendizaje automático”, la arquitectura de la red neuronal y el flujograma de entrenamiento de la red neuronal. A continuación, en la sección 3, se explica la metodología utilizada. Luego en la sección 4 se presenta el experimento y las pruebas. Y, por último, en la sección 5 se presenta las conclusiones y recomendaciones para futuras investigaciones.

II. MARCO DE REFERENCIA

Para la predicción del tiempo de permanencia en la universidad, hay escasas referencias al respecto y las que se abordan, se focalizan principalmente en la deserción estudiantil. Por eso se recurre a algunas investigaciones que respaldan el uso de las redes neuronales artificiales para la predicción de otras situaciones.[1]

El aprendizaje de la red neuronal requiere seguimiento. Muy de acuerdo con [7], ya que el valor del aprendizaje automático es que le permite continuamente aprender de los datos y predecir el futuro. Este poderoso conjunto de algoritmos y modelos se está utilizando en todas las industrias para mejorar procesos y obtener información sobre patrones y anomalías dentro de los datos. Pero el aprendizaje automático no es un esfuerzo solitario, es un proceso de equipo que requiere científicos de datos, ingenieros de datos, analistas de negocios, y líderes empresariales para colaborar.

En la figura 1 se muestra la metodología usada, desde la subida de datos, una primera acción ha sido obtener los datos de los egresados de la Facultad de Ingeniería Industrial y de Sistemas (FIIS) de la Universidad Nacional de Ingeniería (UNI), en un periodo de 10 años. Se ha utilizado para almacenar los datos la plataforma *Colaboratory* de Google, a continuación se identificaron las variables que tienen correlación con la variable objetivo, luego se ha normalizando los datos, se entrenó el modelo de red neuronal de 5 capas, con el 80% de los datos obtenidos y se validaron con el 20% de los datos restantes; seleccionando los datos de forma aleatoria.

Con mayor frecuencia el uso de computación de la nube facilita los cálculos y almacenamiento, los predictores más importantes en la adopción de la computación en la nube, según señalan en [8] son: “*confianza, el riesgo de seguridad de TI percibido y el estilo de gestión*”. Por otra parte, la plataforma seleccionada es TensorFlow [9].

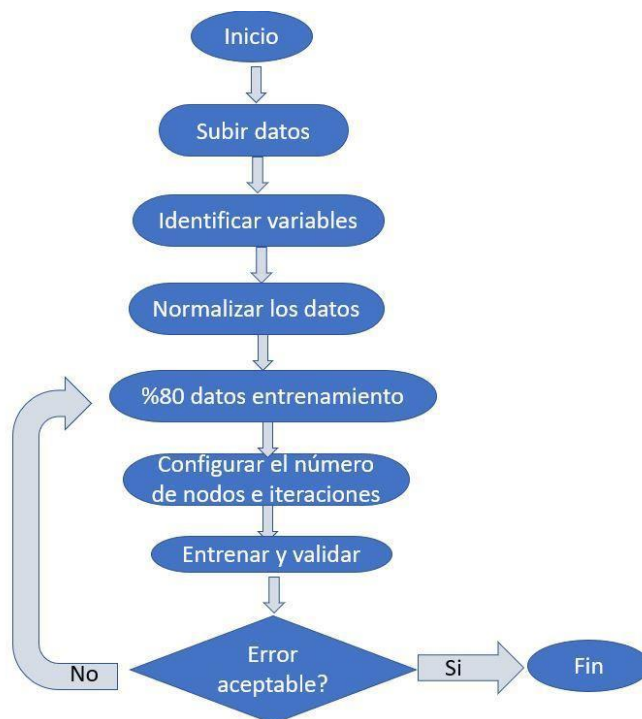


Fig. 1. Flujograma de entrenamiento de la red neuronal y validación

III. COLECCIÓN DE DATOS Y DEFINICIÓN DE VARIABLES

Se ha considerado los mismos datos presentados en el estudio [1], adicionalmente se introdujo una nueva variable código del alumno, el que nos ha permitido determinar: orden de ingreso, modalidad de ingreso y semestre de ingreso. La Tabla I especifica las 10 variables independientes y la variable dependiente “CLASIFICACIÓN”.

TABLA I
VARIABLES DEL MODELO

Nro	Atributo	Descripción
1	ESPECIA	Especialidad (I1, I2)
2	SEXO	Sexo (MASCULINO, FEMENINO)
3	MODALIDADING	Modalidad de ingreso
4	ORDENING	Orden de ingreso.
5	TIPCOLEGIO	Tipo de colegio
6	DATOSLABOR	Datos laborales (No trabaja, Trabaja, Practicante)
7	SEMESTREINGRESO	Semestre de ingreso
8	VIVECONPADR	Vive con sus padres (SI, NO)
9	EDAD INICIAL	Edad al iniciar sus estudios
10	PROMEDIOPRIMER	Promedio de notas en el primer ciclo
11	CLASIFICACIÓN	1: Acaba en 5 años, 0 acabaría en más de 5 años de estudio)

IV. METODOLOGÍA

Se utilizó una Red Neuronal de clasificación, para predecir si el alumno termina sus estudios en cinco años. La

arquitectura es una red neuronal de tres capas ocultas (ver Fig. 2).

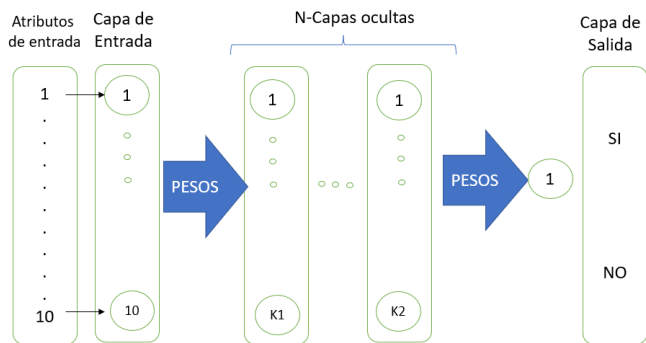


Fig. 2 Arquitectura de la red neuronal para hallar clasificar a un alumno

El experimento se realizó en la plataforma *Colaboratory* de Google, mediante los siguientes pasos:

Paso 1. Subir los datos depurados a la plataforma de Colaboratory en la nube.

Paso 2. Identificar las categorías por cada variable no numérica: ESPECIA, SEXO, TIPCOLEGIO, DATOSLABOR y VIVECONPADR ver Fig.3 y luego se realiza la codificación respectiva.

```
lista_categorias_especia = train_features["ESPECIA"].unique()
lista_categoriassexo = train_features["SEXO"].unique()
lista_categorias_tipcolegio = train_features["TIPCOLEGIO"].unique()
lista_categorias_datoslabor = train_features["DATOSLABOR"].unique()
lista_categorias_viveconpadr = train_features["VIVECONPADR"].unique()
```

Fig. 3 Identificación de las categorías

Paso 3. Analizar los datos a partir del histograma, para identificar variables con valores extremos. La Fig. 4 muestra el histograma de una de las variables dependientes EdadInic. La Fig. 5 muestra los estadísticos para esta variable, donde tenemos valores extremos, valores inferiores como el 14 y valores superiores como los mayores a 25.

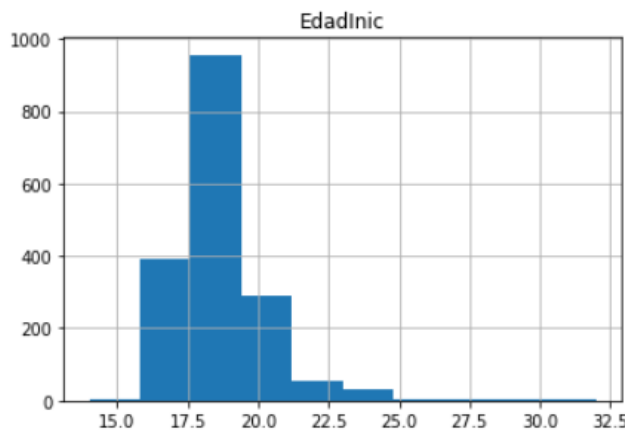


Fig. 4 Histograma de la variable Edad de Inicio

TABLA II
ESTADÍSTICOS DE LA VARIABLE

Media	18.66876791
Error típico	0.040600513
Mediana	18
Moda	18
Desviación estándar	1.696013244
Varianza de la muestra	2.876460924
Curtosis	8.402620475
Coefficiente de asimetría	1.864630638
Rango	18
Mínimo	14
Máximo	32
Suma	32577
Cuenta	1745

Paso 4. Configurar la red neuronal de clasificación con las capas internas, configuradas con la función relu, como se muestra en la Fig. 5

```
model = tf.keras.Sequential([
    tf.keras.layers.Input((None,11,)),
    tf.keras.layers.Dense(64, activation='relu'),
    tf.keras.layers.Dropout(0.2),
    tf.keras.layers.Dense(128, activation='relu'),
    tf.keras.layers.Dense(1, activation='sigmoid')
])
```

Fig. 5 Configuración de la red neuronal con 11 nodos de entrada y 1000 iteraciones.

Paso 5. Se configura la compilación de la red neuronal de clasificación, ver figura 6.

```
[ ] model.compile(optimizer='adam',
    loss=tf.keras.losses.binary_crossentropy,
    metrics=['accuracy'])
```

Fig. 6 Configuración de la red neuronal con 11 nodos de entrada y 1000 iteraciones.

V. EXPERIMENTO

Se realizó un conjunto de experimentos buscando mejorar la precisión y la pérdida. En este artículo, se muestra el experimento con los mejores resultados.

El experimento se realizó utilizando los modelos definidos en el paso 4, con 80% de datos para el

entrenamiento y 20% de datos para la validación, obteniendo los siguientes resultados. Ver figuras 7, 8, 9, 10, 11.

Se realizó el entrenamiento, considerando 1000 iteraciones, ver figura 7.

```
history=model.fit(train_features_numeric,
                  train_labels, validation_split=0.2,
                  epochs=1000)
```

Fig. 7 Configuración del entrenamiento

Se validó el modelo con los datos de verificación, que son el 20% del total. Ver figura 8 y el resultado ver figura 9.

```
test_loss, test_acc = model.evaluate(train_features_numeric,
                                     train_labels, verbose=2)
```

Fig. 8 Configuración de la evaluación

54/54 - 0s - loss: 0.5639 - accuracy: 0.8954 - 99ms/epoch - 2ms/step

Fig. 9 Resultado del entrenamiento de la red neuronal de clasificación

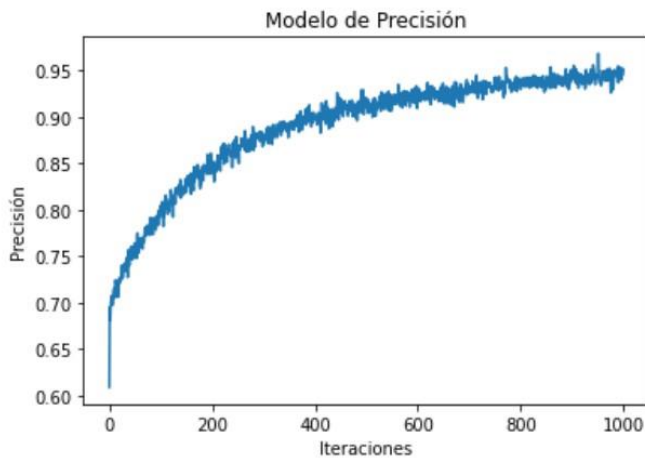


Fig 10 Comportamiento del valor de la precisión en el entrenamiento

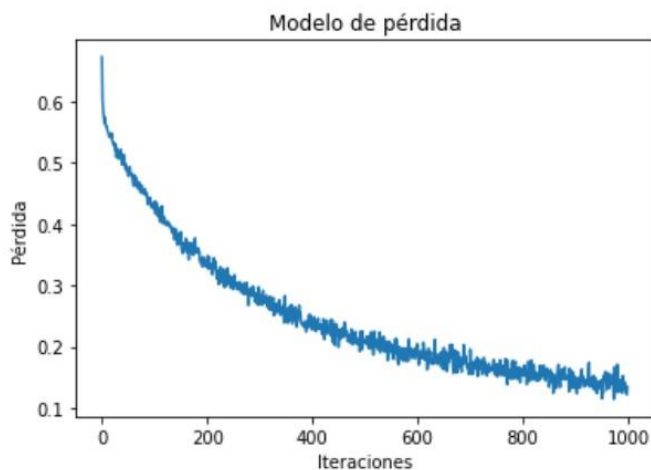


Fig 11 Comportamiento del valor de la pérdida en el entrenamiento

VI. ANÁLISIS DE RESULTADOS

El modelo hallado es una red neuronal de 3 capas internas, además de la capa de entrada y de salidas, se ha considerado 1000 iteraciones de entrenamiento, obteniéndose como resultado una precisión de 90% aproximadamente, ver figura 9; así como su comportamiento con una tendencia a crecer (figura 10) y el de la pérdida de 0.5639, ver figura 9; Se observó también una tendencia a disminuir en su comportamiento, ver figura 11. Este modelo permitirá clasificar a un alumno que ha terminado el primer ciclo, si terminará en el periodo de cinco años o si terminará en más de cinco años.

VII. CONCLUSIONES Y RECOMENDACIONES

7.1. Conclusiones

- Se creó un modelo de una red neuronal, que permite clasificar a un alumno, si terminará en cinco o mayor a cinco años en sus estudios.
- Es posible crear modelos de machine learning en la nube, que permite ayudar en la toma de decisiones a las autoridades correspondientes.
- Este modelo permitirá identificar a los alumnos que tienen riesgo académico.
- Esta herramienta puede ayudar en el seguimiento, monitoreo y predicción del avance de los estudiantes en su carrera.

7.2. Recomendaciones

- Capacitar a los docentes en herramientas, que están disponibles en la nube.
- Considerar el uso de esta herramienta como base para crear modelos que detecten los futuros abandonos de los estudios, por parte de los estudiantes.
- Comprobar que los estudiantes bajo la modalidad e-learning son proclives a la deserción “se encontró que sí es posible construir modelos predictivos exitosos y que el algoritmo que produjo los mejores modelos fue una red neuronal en tres de los cuatro cursos” [10].
- Analizar otros modelos para comparar la efectividad de las predicciones [11].
- Plantear investigaciones que incluyan variables de tipo sociocultural, socioeconómica, nivel educativo de los padres entre otros.

REFERENCIAS

[1] P. Tocto, G. Huamani, y L. Zuloaga. “Construcción de un modelo basado en redes neuronales para determinar la duración de los estudios de ingeniería en una universidad pública en el Perú”. LACCEI 2021.

- [2] M. Mohant and V. Pellakuri. “Innovative supervised machine learning techniques for classification data”. 2020 www.elsevier.com
- [3] H. Pallatadka, A. Wenda, E. Ramirez-Asis, M. Asís-, J. Flores-Albornoz and K. Phasinam. “Classification and prediction of student performance data using various machine learning algorithms”. 2021 www.elsevier.com
- [4] UNESCO Inteligencia artificial en la educación. <https://es.unesco.org/themes/tic-educacion/inteligencia-artificial>
- [5] UN Educational, Scientific and Cultural Organization (UNESCO). 2019b. “Artificial Intelligence in Education: Challenges and Opportunities for Sustainable Development”. Working Papers on Education Policy, 7
- [6] I. Jara y J. Ochoa. “Usos y efectos de la Inteligencia artificial en educación”. BID 2020
- [7] J. Hurwitz and D. Kirsh. “Machine learning for dummies”. IBM. Wileys and sons, Inc..2018
- [8] P. Priyadarshinee, R. Raut, M. Kumar and B. Gardas. “Understanding and predicting the determinants of cloud computing adoption: A two staged hybrid SEM - Neural networks approach”. Computers in Human Behavior, Nov 2017, pp. 341-362
- [9] Ankush. TensorFlow en 10 plataformas de inteligencia artificial para crear su aplicación. 2020 <https://geekflare.com/es/ai-plataformas/>
- [10] I. Urteaga y L. Siri. “Predicción temprana de deserción mediante aprendizaje automático en cursos profesionales en línea”. RIED: revista iberoamericana de educación a distancia, 23 (2) 2020. ISSN 1138-2783.
- [11] R. Sutradhar and L. Barbera. Comparing an Artificial Neural Network to Logistic Regression for Predicting ED Visit Risk Among Patients With Cancer: A Population-Based Cohort Study, Journal of Pain and Symptom Management, Volume 60, Issue 1, 2020, Pages 1-9, ISSN 0885-3924, <https://doi.org/10.1016/j.jpainsymman.2020.02.010>