

Jornadas de Automática

¿Puede un algoritmo de *Machine Learning* ayudarnos en la evaluación de informes de prácticas?

Cabrera, Pedro^{a*}

^a Departamento de Ingeniería Mecánica de la Universidad de Las Palmas de Gran Canaria, 35017, Las Palmas, España.

To cite this article: Cabrera, P. 2024. Can a Machine Learning algorithm assist in the evaluation of practices reports?. *Jornadas de Automática*, 45. <https://doi.org/10.17979/ja-cea.2024.45.10902>

Resumen

La evaluación del aprendizaje en ingeniería requiere de múltiples recursos y actividades formativas prácticas que fomenten el desarrollo de habilidades. Las prácticas de laboratorio y computacionales son esenciales, pero evaluar su efectividad supone un desafío debido a la carga de trabajo y la subjetividad del proceso. Este estudio propone un método de evaluación mediante el uso de *Machine Learning*, aplicado a informes de prácticas en la asignatura de Diseño de Máquinas de la ULPGC. El método propuesto incluye la recopilación y evaluación manual de informes, seguido de la extracción de indicadores como la puntuación *Flesch Reading Ease*, el análisis de sentimiento y el análisis de las palabras frecuentemente utilizadas por el estudiante. Se desarrolla un algoritmo en *Python* para estructurar digitalmente estos parámetros y se implementan técnicas de *Machine Learning*, entre las que destacan las Redes Neuronales Artificiales, para modelar las calificaciones. Los resultados muestran que el modelo logra estimaciones precisas, con métricas de evaluación MAE = 0,72 puntos y MAPE = 8,94%. Este enfoque proporciona una herramienta útil para la evaluación, capaz de mejorar la objetividad del profesorado.

Palabras clave: Machine Learning, Educación, Evaluación, Prácticas, Ingeniería

Can a Machine Learning algorithm assist in the evaluation of practices reports?

Abstract

Evaluating learning in engineering requires multiple resources and practical training activities that foster the development of skills. Laboratory and computer-based practices are essential, but evaluating their effectiveness poses a challenge due to the workload and subjectivity of the process. This study proposes an evaluation method using Machine Learning, applied to practice reports in the Machine Design course at ULPGC. The proposed method includes the collection and manual evaluation of reports, followed by the extraction of indicators such as the Flesch Reading Ease score, sentiment analysis, and the analysis of the words frequently used by the student. A Python algorithm is developed to digitally structure these parameters, and Machine Learning techniques, particularly Artificial Neural Networks, are implemented to model the grades. The results show that the model achieves precise estimates, with evaluation metrics MAE = 0.72 points and MAPE = 8.94%. This approach provides a useful tool for evaluation, capable of improving the objectivity of the faculty.

Keywords: Machine Learning, Education, Evaluation, Practices, Engineering

1. Introducción

El proceso para medir adecuadamente el nivel de desempeño/aprendizaje alcanzado por un estudiante de ingeniería exige la utilización de múltiples y variados recursos. En esta tarea resulta de gran utilidad el uso de varias

actividades formativas. La adquisición práctica del conocimiento, así como las actividades educativas centradas en el desarrollo efectivo de habilidades y competencias en los estudiantes, desempeñan un papel fundamental en este ámbito (Tristacho Ortiz, Contreras Bravo and Vargas Tamayo, 2011). Tanto las prácticas de laboratorio como las prácticas

informáticas son recursos activos y valiosos para adquirir diversas habilidades necesarias en la formación de ingenieros (Tristacho Ortiz, Contreras Bravo and Vargas Tamayo, 2011). Shuman et al. enumeran varias de estas habilidades en (Shuman, Besterfield-Sacre and McGourty, 2005), las cuales pueden ser alcanzadas mediante un diseño adecuado de las actividades prácticas en el contexto de las asignaturas ingeniería. Ruiz-Valencia et al. (Ruiz-Valencia, Magallón-Gudiño and Muñoz-Díaz, 2006) muestran la importancia de recurrir a simulaciones de experiencias reales y tareas prácticas reales para aportar un valor significativo y que los estudiantes retengan la información deseada (Ricardo, 1998). Sin embargo, evaluar adecuadamente la adquisición de habilidades conseguidas a través de experiencias prácticas en laboratorio o en el aula de informática no siempre es sencillo. La puesta en práctica, así como la evaluación de dichas actividades supone un gran reto para el profesorado, debido a la alta carga de trabajo que supone evaluar múltiples informes generados por los estudiantes. Debido a ello, el profesorado se ve limitado en tiempo y capacidad para llevar a cabo nuevas, más variadas y mejores actividades formativas. Esto redundaría en un menor aprovechamiento de los recursos. Asimismo, ante el reto de evaluar numerosas actividades formativas, la objetividad del proceso de evaluación puede verse afectada. Por tanto, el desarrollo de herramientas que permitan complementar esta tarea podría contribuir en la mejora y en la satisfacción del profesorado y del propio alumnado. Por otra parte, algunas técnicas de *Machine Learning* como los árboles de decisión permiten extraer información de los datos imperceptibles al ojo humano y detectar relaciones complejas entre las variables analizadas (Hopfield, 1982). Esto, sin duda, puede contribuir a mejorar el proceso de aprendizaje de los estudiantes debido a que se podrán profundizar en aspectos de la materia que tengan más relación con el desempeño de los estudiantes.

En este trabajo se presenta un caso de estudio real realizado con los estudiantes de la asignatura de Diseño de Máquinas del cuarto curso del Grado en Ingeniería Mecánica de la ULPGC. El trabajo trata de mejorar el método de evaluación empleando un sistema de aprendizaje automático que permita obtener patrones de los informes de prácticas realizados por los estudiantes. Esto le brinda más información al profesorado, a la vez que le permite aplicar un método objetivo de evaluación. La idea se basa en analizar múltiples informes de prácticas y desarrollar una herramienta de *Machine Learning* que ayude a determinar qué ítems de evaluación son más relevantes, así como para obtener una valoración precisa del informe atendiendo a algunas variables/patrones detectados en el mismo.

2. Método

A continuación, se describen los pasos llevados a cabo en este trabajo de investigación para aplicar el método propuesto:

2.1. Evaluación manual de los informes de prácticas.

En primer lugar, se ha recopilado un total de 44 informes de prácticas de estudiantes diferentes, mediante el campus

virtual. Para ello, se les ha solicitado que sigan la estructura científica habitual de presentación de documentos (Figura 1). Estos han entregado en un fichero con formato PDF los distintos informes. A partir de ese momento, el profesor ha evaluado de forma manual cada informe aplicando los criterios de evaluación que tratan de valorar las competencias de la asignatura indicados en la Tabla 1.

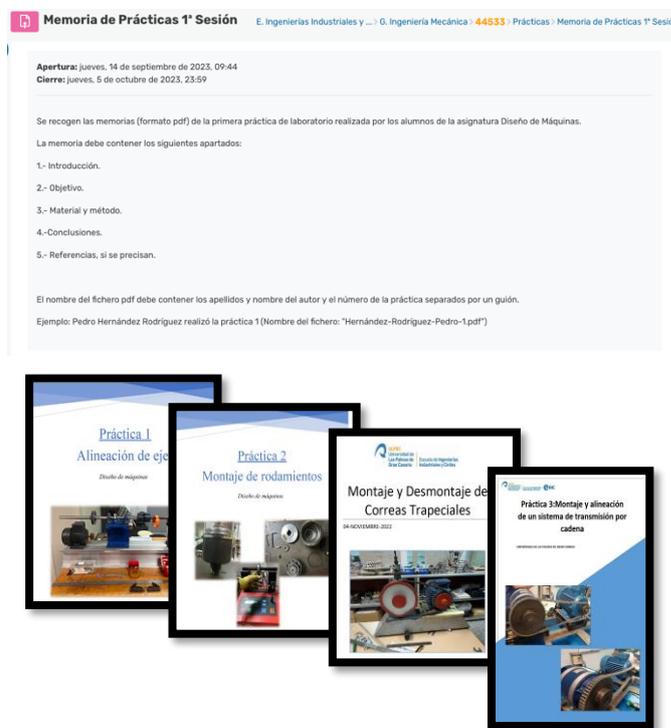


Figura 1: Ejemplo de solicitud de entrega e informes de prácticas.

Tabla 1: Criterios de evaluación utilizados y competencias a evaluar en los informes de prácticas de forma manual.

Criterios de evaluación	
1. Presentación, orden y claridad en la exposición de ideas (T4).	20%
2. Descripciones de los conceptos vistos en la práctica (MTEM2, T3, T5, T6)	50%
3. Comprensión, interpretación y aportación individual del alumno (G6).	30%
Competencias que tienen asignadas las prácticas	
MTEM2: Conocimientos y capacidades para el cálculo, diseño y ensayo de máquinas.	
T3: Conocimiento en materias básicas y tecnológicas, que les capacite para el aprendizaje de nuevos métodos y teorías, y les dote de versatilidad para adaptarse a nuevas situaciones.	
T4: Capacidad de resolver problemas con iniciativa, toma de decisiones, creatividad, razonamiento crítico y de comunicar y transmitir conocimientos, habilidades y destrezas.	
T5: Conocimientos para la realización de mediciones, cálculos, valoraciones, tasaciones, peritaciones, estudios, informes, planes de labores y otros trabajos análogos.	
T6: Capacidad para el manejo de especificaciones, reglamentos y normas de obligado cumplimiento.	
G6: Detectar deficiencias en el propio conocimiento y superarlas mediante la reflexión crítica y la elección de la mejor actuación para ampliar este conocimiento.	

2.2. Identificar patrones e indicadores en los informes de prácticas

Como segundo paso del método aplicado, se desarrolla un algoritmo en Python que recoge toda la información posible de cada informe de práctica y la estructura digitalmente para extraer de dichos informes distintos parámetros (Figura 2). De entre los indicadores extraídos de los informes destacan los siguientes:

a) **La puntuación Flesch Reading Ease** (índice de facilidad de lectura Flesch) es una medida que se utiliza para evaluar cuánto de fácil o difícil es leer un texto. Esta puntuación se basa en dos factores principales: la longitud de las palabras y la longitud de las oraciones en el texto. Cuando se obtiene una puntuación Flesch Reading Ease, generalmente se obtiene un número entre 0 y 100. Cuanto más alto sea el número, más fácil será de leer el texto:

- Una puntuación alta (por ejemplo, 80-100) significa que el texto es fácil de entender, como un texto simple o un libro para niños.
- Una puntuación intermedia (por ejemplo, 50-80) indica que el texto es moderadamente fácil de leer y es adecuado para la mayoría de los adultos.
- Una puntuación baja (por ejemplo, por debajo de 50) sugiere que el texto es difícil de entender y está dirigido a un público con un alto nivel de educación.

b) **El análisis de polaridad de sentimiento** trata de buscar si un texto tiene un tono positivo, negativo o neutral. El análisis de sentimiento obtiene un número entre -1 y 1, donde:

- Un valor cercano a 1 indica que el texto tiene un sentimiento muy positivo.
- Un valor cercano a -1 indica que el texto tiene un sentimiento muy negativo.
- Un valor cercano a 0 sugiere que el texto es neutral o carece de un sentimiento fuerte

Esto se usa para automatizar el análisis de grandes cantidades de texto para comprender la opinión general o el sentimiento detrás de ellos. Esto es útil en diversas aplicaciones, como la minería de opiniones en redes sociales, la evaluación de reseñas de productos o servicios, y la clasificación de comentarios de clientes.

c) **Palabras más frecuentes y número total de palabras.** Se cuentan las tres palabras más frecuentes para validar que el informe refuerza aquellas partes vistas durante la práctica.

Estudiante	readability_score	sentiment_positive	sentiment_neutral	sentiment_negative	sentiment_compound	total_words	word1	word2	word3	Mark
Estudiante1	46.3	0.01	0.972	0.018	-0.8979	1796	alineacion	ilustracion		9.5
Estudiante2	47.18	0	0.986	0.014	-0.6808	516		3	alineacion	7.1
Estudiante3	41.19	0.009	0.98	0.011	-0.9516	1164	alineacion		laser	7.8
Estudiante4	47.12	0.018	0.978	0.004	-0.7845	601	ejes		alineacion	8.2
Estudiante5	31.96	0.028	0.958	0.014	-0.6486	401	laser	alineacion	ejes	6.8
Estudiante6	39.47	0.006	0.969	0.024	-0.9295	1016		titulo	maquinas	7.5
Estudiante7	42.45	0.047	0.953	0	-0.9705	541	ejes	sistema	mail	7.7
Estudiante8	38.79	0.014	0.971	0.015	-0.1027	1182		ejes	alineacion	8
Estudiante9	54.83	0	1	0	0	336	alineacion	maquina	ejes	6.6
Estudiante10	58.08	0.003	0.969	0.004	-0.0772	767		ilustracion	ejes	8.75
Estudiante11	34.33	0.009	0.991	0	-0.5423	409	ejes		posicion	5.75
Estudiante12	51.58	0.012	0.969	0.018	-0.6124	1120		ejes	mediante	9
Estudiante13	54.73	0.007	0.978	0.015	-0.5267	911	figura	alineacion	4	7
Estudiante14	54.52	0.01	0.981	0.009	0	1986	ejes		m	9.1
Estudiante15	52.29	0.006	0.961	0.019	-0.8225	1394	eje	alineacion		2
Estudiante16	35.07	0.002	0.99	0.008	-0.9432	1206		alineacion	maquinas	8.4
Estudiante17	53.21	0	0.994	0.006	-0.296	425	ejes		sistema	7
Estudiante18	52.9	0.005	0.995	0	-0.7003	1372	alineacion		ilustracion	9

Figura 2: Ejemplo de tabla de datos con indicadores digitales obtenidos de los informes de prácticas y las calificaciones manuales puestas por el docente.

2.3. Utilizar las calificaciones e indicadores para modelar una técnica de Machine Learning.

En tercer lugar, se programan varias técnicas de *Machine Learning* para detectar aquella que estima las calificaciones reales del estudiante en mejor medida. En concreto se programaron varias Redes Neuronales Artificiales, empleando distinto número de neuronas en sus distintas capas, así como la técnica conocida como Bosques Aleatorios (*Random Forest*) y Máquinas de Vector Soporte.

En este caso, la técnica que obtuvo mejores resultados fue las Redes Neuronales Artificiales compuestas por 10 neuronas y 6 neuronas en sus dos capas ocultas.

2.4. Evaluación del modelo.

Las métricas empleadas en este trabajo para evaluar las estimaciones numéricas del modelo propuesto han sido las siguientes:

- El error absoluto medio (MAE, de las siglas en inglés mean absolute error)
- El error absoluto porcentual medio (MAPE, de sus siglas en inglés mean absolute percentage error).

El MAE está definido por la Eq. (1) donde los n valores estimados están representados por la letra "e" y los n valores observados por la letra "o". El MAE está expresado en las mismas unidades de los parámetros que compara (Lozano, Cabrera and Blanco-Marigorta, 2020).

$$MAE = \frac{1}{n} \sum_{i=1}^n |o_i - \hat{e}_i| \quad (1)$$

El MAPE está definido por la Eq. (2) y es una medida relativa que expresa el error como porcentaje de los datos observados (Lozano, Cabrera and Blanco-Marigorta, 2020).

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{o_i - \hat{e}_i}{o_i} \right| \quad (2)$$

Método de entrenamiento y evaluación del modelo analizado

Para evaluar el modelo usado en este estudio se aplica la técnica de validación cruzada de 10 iteraciones (Hastie, Tibshirani and Friedman, 2009) para entrenar el modelo y ensayarlo 10 veces. Esta técnica sigue el procedimiento descrito a continuación:

1. Se reordenan aleatoriamente todas las muestras de la tabla de datos por filas, para garantizar la representatividad de todos los datos.
2. El conjunto completo de datos se divide en 10 partes de igual tamaño, *pliegues* o *folds*, cada uno de ellos con el mismo número aproximado de muestras.
3. Se escogen 9 de esos *folds* (que representan el 90% del total de los datos) y se consideran estos temporalmente como el *subconjunto de entrenamiento*, que servirá para ajustar el modelo. El *fold* restante (que representa el 10% restante de las muestras totales de los datos) se considera temporalmente como el *subconjunto de ensayo* y, con él, se llevará a cabo un ensayo del modelo.
4. Con el *subconjunto de entrenamiento* (90% de los datos) se entrena el modelo de *Machine Learning* escogido.
5. Este procedimiento se repite 10 veces para ir rotando el *subconjunto de ensayo* y obtener un total de 10 evaluaciones del modelo con subconjuntos diferentes de la misma tabla de datos. Como resultado, se obtienen 10 valores de error calculados mediante la aplicación de una métrica estadística. En este estudio, se emplean las dos métricas estadísticas mencionadas previamente (MAE y MAPE). Con ellas se puede cuantificar objetivamente el error entre el valor real de la variable objetivo (calificaciones de los informes de prácticas) y el valor estimado por el modelo.
6. Por último, se calculan los valores medios Eq. (3) y Eq. (4), y la desviación típica de los 10 valores obtenidos para cada métrica de error.

$$\overline{MAE} = \frac{1}{10} \sum_{i=1}^{10} MAE_i \quad (3)$$

$$\overline{MAPE} = \frac{1}{10} \sum_{i=1}^{10} MAPE_i \quad (4)$$

3. Resultados

Del proceso descrito se ha obtenido un modelo que logra los resultados representados en la Figura 3 y en la Tabla 2. En ambos se representan las notas reales frente a las notas estimadas por el algoritmo. Además, se representan los errores MAE (Mean Absolute Error) y MAPE (Mean Absolute Percentage Error) de dichas estimaciones.

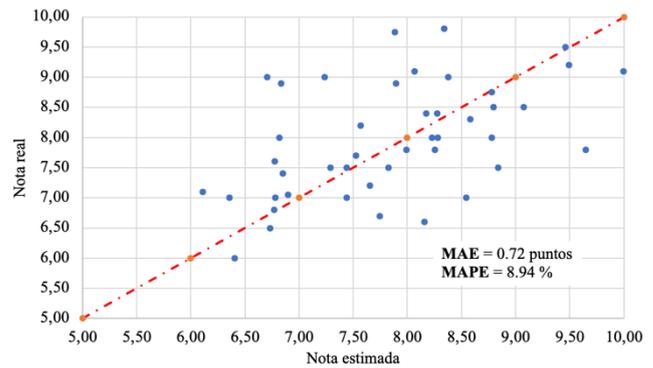


Figura 3: Desviaciones entre calificaciones reales puestas por el docente y calificaciones estimadas por el algoritmo.

Tabla 2: Resultados de las estimaciones realizada por el modelo

Nota real	Nota estimada	Error
9,00	7,24	-1,76
8,40	8,28	-0,12
7,80	7,99	0,19
9,80	8,34	-1,46
7,50	7,83	0,33
8,00	8,28	0,28
9,00	6,71	-2,29
7,40	6,85	-0,55
7,70	7,53	-0,17
8,40	8,18	-0,23
8,75	8,78	0,03
7,50	8,84	1,34
6,70	7,75	1,05
7,00	8,55	1,55
7,80	8,26	0,46
7,05	6,90	-0,15
8,90	6,84	-2,06
7,00	6,36	-0,64
7,60	6,77	-0,83
8,00	8,78	0,78
9,00	8,38	-0,62
7,80	9,65	1,85
8,00	8,23	0,23
6,60	8,16	1,56
6,50	6,73	0,23
6,80	6,77	-0,03
6,00	6,40	0,40
9,50	9,46	-0,04
8,90	7,90	-1,00
7,00	7,44	0,44
8,50	8,80	0,30
7,50	7,44	-0,06
8,30	8,58	0,28
7,00	6,78	-0,22
9,10	10,00	0,90
8,50	9,08	0,58
8,20	7,57	-0,63
9,75	7,88	-1,87
7,20	7,65	0,45
9,20	9,49	0,29
9,10	8,07	-1,03
7,10	6,11	-0,99
7,50	7,29	-0,21
8,00	6,82	-1,18
MAE		0,72
MAPE		8,94%

4. Conclusiones

En este trabajo se demuestra que las técnicas de *Machine Learning* pueden complementar la tarea de evaluación y contribuir en la mejora y en la satisfacción del profesorado y del propio alumnado. Se comprueba que estas técnicas pueden llegar a obtener errores de estimación de la calificación del docente menores al 10% (MAPE = 8,94%). Asimismo, algunas de estas técnicas permiten extraer información de los datos imperceptibles al ojo humano y detectar relaciones entre las variables analizadas. Esto, sin duda, constituye un gran paso para mejorar el proceso de aprendizaje de los estudiantes debido a que se podrán profundizar en aspectos de la materia que tengan más relación con el desempeño de estos. A partir de los resultados obtenidos, se recomienda profundizar en este estudio y en la integración de estas técnicas en los sistemas habituales de trabajo del docente (Moodle, campus virtual, etc). Asimismo, se cree que sería recomendable desarrollar a partir de esta experiencia una herramienta maquetada con una interfaz gráfica definida, para facilitar el uso docente.

Financiación

Esta contribución ha sido co-financiada por la Ayuda RED2022-134588-T financiada por MICIU/AEI /10.13039/501100011033.

Asimismo el trabajo ha sido co-financiado por los fondos europeos Next Generation EU (NGEU) bajo el “Real Decreto 641/2021, de 27 de julio, por el que se regula la concesión directa de subvenciones a universidades públicas españolas para la modernización y digitalización del sistema universitario español en el marco del plan de recuperación, transformación y resiliencia (UNIDIGITAL) - Proyectos de Innovación Educativa para la Formación Interdisciplinar (PIEFI) - Línea 3. Contenidos y programas de formación” en el seno del Proyecto de Innovación Educativa “¿Podría un algoritmo de *Machine Learning* ayudar al profesorado en la evaluación de algunas actividades formativas? (PIE2022-52)”.

Referencias

- Hastie, T., Tibshirani, R. and Friedman, J. (2009) *The Elements of Statistical Learning*. 2nd edn. New York: Springer-Verlag.
- Hopfield, J.J. (1982) ‘Neural networks and physical systems with emergent collective computational abilities.’, *Proceedings of the National Academy of Sciences of the United States of America*, 79(8), pp. 2554–2558. Available at: <https://doi.org/10.1073/pnas.79.8.2554>.
- Lozano, A., Cabrera, P. and Blanco-Marigorta, A.M. (2020) ‘Non-linear regression modelling to estimate the global warming potential of a newspaper’, *Entropy*, 22(5), p. 590. Available at: <https://doi.org/10.3390/E22050590>.
- Ricardo, C. (1998) ‘The Globalization and the Engineering Teaching for the XXI Century.’, *Primer Congreso Argentino de Enseñanza en la Ingeniería*, 8 October. Available at: <https://eric.ed.gov/?id=ED444861> (Accessed: 5 October 2021).
- Ruiz-Valencia, D.M., Magallón-Gudiño, J. and Muñoz-Díaz, E.E. (2006) ‘Herramientas de aprendizaje activo en las asignaturas de ingeniería estructural | Ingeniería y Universidad’, *Ingeniería y Universidad*, 10(1), pp. 97–115. Available at: <https://revistas.javeriana.edu.co/index.php/iyu/article/view/914> (Accessed: 5 October 2021).
- Shuman, L.J., Besterfield-Sacre, M. and McGourty, J. (2005) ‘The ABET “Professional Skills” — Can They Be Taught? Can They Be Assessed?’, *Journal of Engineering Education*, 94(1), pp. 41–55. Available at: <https://doi.org/10.1002/J.2168-9830.2005.TB00828.X>.
- Tristacho Ortiz, J.A., Contreras Bravo, L.E. and Vargas Tamayo, L.F. (2011) ‘Análisis y aplicación de técnicas de aprendizaje activo en mecánica aplicada’, *Revista Educación y Desarrollo Social*, 8(2), pp. 28–45.

