INTERACCION
Revista Digital de AIPO
Asociación Interacción Persona-Ordenador

© Interacción, Revista Digital de AIPO, Volumen 7, Número 2, 2026, ISSN: 2695-6578
DOI: https://doi.org/10.65234/interaccion.148

Análisis de características de usuario con *Machine Learning*

Analysis of user characteristics using Machine Learning

[1] Alberto Gaspar [0000-0002-9236-4391], [1] José Ignacio Panach [0000-0002-7043-6227], [1] Miriam Gil [0000-0002-2987-1825], [1] Verónica Romero [0000-0002-1721-5732] y [2] Damiano Distante [0000-0002-8467-535X]

[1] Departament d'informàtica, Escola Tècnica Superior d'Enginyeria, Universitat de València, Avenida de la universidad s/n, 46100 Burjasot, España, algasvi@alumni.uv.es, {joigpana, miriam.gil, veronica.romero}@uv.es

[2] UnitelmaSapienza University of Rome, Piazza Sassari, 4, Rome, 00161, Italy, damiano.distante@unitelmasapienza.it

Resumen. El diseño de interfaces inteligentes que se adaptan a las características del usuario es un factor clave para mejorar la experiencia de uso. Sin embargo, existe una carencia de estudios empíricos que comparen la extracción automática de características de usuario con la percepción que los propios usuarios tienen de dichas características. En este trabajo se evalúan cuatro algoritmos de aprendizaje automático para clasificar a los usuarios en función de su nivel de habilidad y conocimiento: Perceptrón Multicapa (MLP), Naive Bayes (NB), Máquina de Soporte Vectorial (SVM) y Random Forest (RF). Estos algoritmos son adecuados para modelar relaciones complejas entre los datos de interacción y los perfiles de usuario. Los resultados muestran que MLP y SVM ofrecen el rendimiento más consistente en términos de precisión. No obstante, todos los modelos presentan dificultades para clasificar correctamente las clases minoritarias, especialmente en los usuarios con bajos niveles de habilidad y conocimiento.

Palabras clave: Interfaces Inteligentes de Usuario, Clasificación de Usuarios, Machine Learning, Interfaz gráfica de Usuario.

Abstract. The design of intelligent interfaces that adapt to user characteristics is a key factor in improving the user experience. However, there is a lack of empirical studies comparing the automatic extraction of user characteristics with users' own perceptions of those characteristics. This study evaluates four machine learning algorithms for classifying users based on their skill and knowledge levels: Multilayer Perceptron (MLP), Naive Bayes (NB), Support Vector Machine (SVM), and Random Forest (RF). These algorithms are well-suited for modeling complex relationships between interaction data and user profiles. The results show that MLP and SVM offer

[Recibido: 19-04-26 | Aceptado: 01-06-26 | Publicado: 29-09-26]

the most consistent performance in terms of accuracy. However, all models have difficulty correctly classifying minority classes, especially among users with low levels of skill and knowledge.

Keywords: Intelligent User Interfaces, User Classification, Machine Learning, Graphical User Interface.

1. INTRODUCCIÓN

Las interfaces inteligentes de usuario integran un conjunto de métodos y reglas orientados a personalizar el contenido de la interfaz gráfica de usuario según las características de la persona que interactúa con el sistema, para mejorar la usabilidad y la experiencia de usuario. No obstante, la mayoría de los enfoques existentes se centran en soluciones específicas de cada dominio y no en perfiles de usuario genéricos y reutilizables, lo que limita su capacidad de adaptación a entornos dinámicos y a usuarios con comportamientos heterogéneos. Asimismo, trabajos de revisión sistemática sobre interfaces adaptativas (Kristić et al., 2025) muestran que los enfoques basados en el aprendizaje automático permiten a los sistemas inferir patrones de interacción y adaptar dinámicamente la interfaz en función del estado, del comportamiento y de las necesidades del usuario. Esto implica la realización de procesos de clasificación o de modelado de usuarios para personalizar la interacción.

En este contexto, diversos trabajos han abordado la clasificación de usuarios a partir de la información contextual de la aplicación. Por ejemplo, Duan et al. (2024) proponen un enfoque multimodal basado en transformadores y en técnicas de aprendizaje profundo para clasificar a los usuarios según su estado emocional a partir de señales visuales, acústicas y textuales. Sin embargo, este tipo de aproximaciones se centra en estados afectivos transitorios y no en características persistentes del usuario, como habilidades, conocimientos u objetivos. Zahoor et al. (2024) analizan comentarios de una aplicación móvil para identificar tendencias y preferencias de los usuarios, con el fin de ofrecer recomendaciones de contenido. A pesar de sus contribuciones, estos trabajos no definen perfiles de usuario estructurados que permitan una clasificación sistemática y reutilizable, un aspecto clave para avanzar hacia soluciones de personalización más generales y escalables.

Actualmente, la clasificación de usuarios suele basarse en algoritmos de aprendizaje automático (Machine Learning, ML). Estos algoritmos han sido ampliamente utilizados para tareas de clasificación en diversos dominios, incluyendo sistemas de recomendación, modelado de usuario y adaptación de interfaces (Schaible et al., 2024). A pesar de su uso extendido, existe una escasez de estudios comparativos que analicen empíricamente el grado de concordancia entre la clasificación obtenida mediante estos algoritmos y la realizada por los propios usuarios. En este contexto, la principal contribución de este artículo es la comparación entre la clasificación de los atributos del usuario obtenida mediante algoritmos de ML y la realizada por los propios usuarios según su percepción. Se ha estudiado el rendimiento de diferentes algoritmos de ML, en concreto: Perceptrón Multicapa (MLP), Naive Bayes (NB), Máquinas de Soporte Vectorial (SVM) y Random Forest (RF).

Las características utilizadas en este trabajo se basan en las presentadas en (Gaspar et al., 2024), donde se define un modelo de usuario con las características que afectan a la personalización de la interfaz. De entre todas las características del modelo, este trabajo se centra en el **conocimiento** y la

habilidad. Esta selección se justifica porque son las características que presentan mayores implicaciones en el proceso de adaptación de las interfaces. Para cada una de estas características, el algoritmo de ML realiza una estimación automática del valor asociado a un usuario a partir de un conjunto de variables observadas durante la interacción con el sistema, como los clics en la interfaz y el tiempo invertido en las distintas tareas. A partir de esta información, el algoritmo asigna un valor de Bajo, Medio o Alto a cada característica del usuario. Paralelamente, cada usuario proporciona una autoevaluación del valor que considera que posee para esas mismas características, lo que permite comprobar si hay correlación entre ambos valores.

El artículo se estructura de la siguiente manera. La Sección 2 revisa el estado del arte. La Sección 3 describe la metodología empleada, incluyendo los participantes, el conjunto de datos y el algoritmo de clasificación. La Sección 4 presenta y analiza los resultados. Finalmente, la Sección 5 recoge las conclusiones y las líneas futuras de trabajo.

2. ESTADO DEL ARTE

La clasificación de los usuarios ha sido ampliamente analizada en la literatura. Estos trabajos han utilizado algoritmos de ML para establecer grupos de usuarios y clasificarlos. Algunos trabajos emplean el algoritmo BERT (Bidirectional Encoder Representations from Transformers) para clasificar a los usuarios a partir de sus patrones de interacción y comportamiento en sistemas digitales. Siguiendo esta línea, Saranya et al. (2025) proponen una variante mejorada, denominada IBERT, para clasificar a los usuarios de redes cognitivas 5G como malintencionados o legítimos en función de sus características de actividad. Los autores concluyen que la propuesta alcanza una precisión cercana al 99,74 %. Siguiendo esta línea, Siino et al. (2022) emplean modelos Transformer como BERT, RoBERTa y DistilBERT para clasificar usuarios que difunden noticias falsas en Twitter y concluyen que el rendimiento depende del contexto del problema.

Otros trabajos emplean el algoritmo SVM para clasificar a los usuarios según su comportamiento en sistemas digitales. Siguiendo esta idea, Yuvapriya et al. (2025) utilizan SVM para clasificar a profesionales de tecnologías de la información en un entorno de e-learning adaptativo según su rendimiento, considerando patrones de navegación, participación y resultados en evaluaciones. Concluyen que la combinación de técnicas de agrupamiento y SVM permite separar eficazmente a los usuarios en distintos niveles de desempeño. De forma similar, Alwafi y Fakieh (2024) aplican el SVM para clasificar a usuarios de redes sociales en función de su nivel de fatiga de privacidad, utilizando rasgos de personalidad y de conciencia sobre la privacidad como variables predictoras. Los autores concluyen que el SVM alcanza una precisión cercana al 78 %, lo que demuestra una alta capacidad para identificar correctamente a usuarios con niveles elevados de fatiga.

Otros trabajos usan el algoritmo RF para clasificar a los usuarios en sistemas inteligentes en función de patrones de comportamiento y variables contextuales. Siguiendo esta línea, Jaber y Csonka(2023) analizan distintos grupos de usuarios de sistemas de bike-sharing mediante modelos de minería de datos, incluidos los modelos de RF, considerando variables temporales y meteorológicas, como la temperatura, la radiación solar o el tipo de día. Los autores concluyen que RF permite identificar diferencias significativas entre perfiles de usuario y mejorar la predicción del uso del sistema. De forma similar, Abourobea et al.(2025) aplican RF para la clasificación de usuarios de la vía pública a partir de datos de radar obtenidos de infraestructuras inteligentes, empleando características

geométricas y cinemáticas extraídas de nubes de puntos. Los resultados muestran que los modelos basados en RF alcanzan valores cercanos al 97 %, lo que destaca su robustez frente a la variabilidad del entorno y su capacidad para clasificar usuarios en tiempo real.

Algunos trabajos emplean árboles de decisión (DT) para clasificar usuarios a partir de características demográficas, conductuales o de patrones de interacción en sistemas inteligentes. Basuni y Siregar (2023) comparan diferentes modelos de aprendizaje automático para clasificar individuos como usuarios o no usuarios de sustancias a partir de variables psicológicas, demográficas y de consumo. Concluyeron que, aunque los DT ofrecen una interpretación clara del proceso de toma de decisiones y facilitan el análisis de atributos relevantes, su rendimiento es inferior al de los modelos de ensemble. De forma similar, Shannaq (2024) emplea DT en un sistema de ciberseguridad para clasificar a los usuarios según patrones de sus contraseñas mediante TF-IDF, para diferenciar entre usuarios legítimos e intrusos potenciales. Los resultados muestran que DT permite modelar reglas interpretables basadas en características textuales del comportamiento del usuario, aunque con menor precisión que otros modelos probabilísticos y basados en márgenes.

Los trabajos descritos en esta sección emplean diversas metodologías de ML para la clasificación de usuarios. Estos trabajos presentan varias limitaciones. En primer lugar, muchas propuestas se centran en dominios muy específicos, como redes 5G o la detección de noticias falsas, lo que limita su generalización a otros contextos. En segundo lugar, la mayoría de los enfoques prioriza patrones de comportamiento o datos contextuales sin considerar características persistentes del usuario, como conocimientos o habilidades. Además, algunos modelos avanzados, como los basados en Transformers, presentan alta complejidad computacional y dependencia del conjunto de datos, lo que puede afectar su rendimiento en distintos escenarios. Por otra parte, varios trabajos se centran únicamente en evaluar la precisión del algoritmo, sin contrastar los resultados con la percepción o la autoevaluación de los usuarios. Además, muchos enfoques no emplean modelos de usuario estructurados y reutilizables, lo que dificulta su integración en sistemas adaptativos más generales. En este contexto, nuestra propuesta consiste en comparar la clasificación realizada por un algoritmo de aprendizaje automático con la percepción que cada usuario tiene de sus propias características.

3. CLASIFICACIÓN DE USUARIOS POR MACHINE LEARNING

En esta sección se analizarán las características de usuario utilizadas en la clasificación de usuarios, indicando los diferentes valores que pueden adoptar, el conjunto de datos utilizado para clasificar a los usuarios y los algoritmos de clasificación empleados.

3.1. Características de usuario evaluadas

Esta subsección describe las características de usuario utilizadas para definir los perfiles de usuario: habilidad y conocimiento. De todas las características del modelo de usuario (Gaspar et al., 2024), este artículo se centra en estas dos porque son las que más influyen en la adaptación de las interfaces (Brusilovsky, 1996).

La característica **conocimiento** permite identificar los conceptos que el usuario comprende en el contexto de la aplicación (Abri et al., 2020). No obstante, su principal dificultad para estimarla radica en su carácter dinámico, ya que evoluciona de forma continua: el usuario adquiere y olvida

conceptos en cada interacción con el sistema. Para su estimación, el sistema clasifica esta característica en tres niveles: (1) conocimiento bajo, (2) conocimiento medio y (3) conocimiento alto.

La característica **habilidad** permite identificar las competencias técnicas que posee el usuario, como el uso de la búsqueda avanzada para acceder a un género específico dentro de una biblioteca digital (Tachie-Donkor & Ezema, 2023). Para su estimación, el sistema clasifica esta característica en tres niveles: (1) habilidad baja (novato), (2) habilidad media y (3) habilidad alta (experto).

3.2. Conjuntos de datos

El conjunto de datos se generó mediante un proceso de recopilación manual de variables observadas a partir de la interacción de los usuarios, con el objetivo de entrenar un modelo de aprendizaje automático capaz de clasificar sus características (habilidad y conocimiento). Para ello, se desarrollaron varios prototipos en *QuantUX* para dar soporte a 2 aplicaciones (coche inteligente y sistema domótico) y 16 tareas. El conjunto de tareas puede ser accedido en (Gaspar et al., 2026).

La recogida de datos se basó en las interacciones de los usuarios con los prototipos. Para estimar la característica de conocimiento, se analizaron las siguientes variables: la duración total de la interacción, el número de clics, las pausas sensibles (> 5 segundos), los clics erróneos y el tiempo promedio por página. Por otro lado, para estimar la característica de habilidad, se consideraron las siguientes variables observadas: el tiempo total de interacción (en segundos), el número de clics, los clics erróneos, el número de páginas visitadas y el porcentaje de acciones especiales realizadas, como usar el comando Ctrl+C en lugar de hacer clic con el ratón para copiar un elemento. Asimismo, se diseñó un cuestionario estructurado donde cada participante indicaba el valor que consideraba poseer en conocimiento y habilidad. El valor proporcionado por este cuestionario se considera el Ground Truth (GT) del experimento, es decir, el valor de referencia que se asume como correcto y frente al cual se comparan las predicciones del modelo de aprendizaje automático. Se reclutaron 60 estudiantes de Ingeniería Informática (20–24 años). Para cada participante se recopilaron las variables descritas previamente y la valoración de preferencia utilizada como GT, empleándose la media de las respuestas como estimación de cada característica. Aunque los participantes reportaron un conocimiento previo moderado de los sistemas evaluados (3,0/5 y 2,8/5), el perfil técnico de la muestra podría introducir un sesgo de selección.

3.3. Algoritmo de clasificación

En esta sección se describen los algoritmos de ML empleados para clasificar las características del usuario: habilidad y conocimiento. Para analizar el rendimiento desde distintos enfoques, se han evaluado cuatro algoritmos de clasificación supervisada ampliamente utilizados en la literatura: Perceptrón Multicapa (MLP), Naive Bayes (NB), Máquinas de Soporte Vectorial (SVM) y Random Forest (RF) (Kadhim, 2019; Murphy, 2012). Estos algoritmos han sido seleccionados por su diversidad en cuanto a principios de funcionamiento y capacidad para modelar relaciones lineales y no lineales en los datos.

El MLP es un modelo basado en redes neuronales artificiales capaz de capturar relaciones complejas mediante la composición de múltiples capas de neuronas. Por su parte, NB es un clasificador probabilístico basado en el teorema de Bayes que asume la independencia condicional entre las

variables de entrada. Las SVM permiten encontrar un hiperplano óptimo que maximiza la separación entre clases, siendo especialmente eficaces en espacios de alta dimensión. Finalmente, RF es un método de tipo ensemble basado en la construcción de múltiples árboles de decisión, cuya combinación permite mejorar la capacidad de generalización y reducir el sobreajuste.

Para cada una de las características analizadas, los modelos fueron entrenados utilizando las variables observadas descritas en la sección anterior y evaluados comparando sus predicciones con el Ground Truth proporcionado por los usuarios. Así, se analizó el rendimiento de cada algoritmo en términos de su capacidad para estimar correctamente el conocimiento y la habilidad de los usuarios.

3.4. Resultados

Los resultados de los modelos de clasificación para las características de habilidad y conocimiento se evaluaron mediante validación cruzada Leave-One-Out, dejando a un usuario fuera en cada iteración y entrenando el modelo con el resto. Este protocolo permite evaluar a cada usuario de forma independiente y aprovechar al máximo el conjunto de datos disponible. Por otro lado, en el caso del MLP, se probaron varias configuraciones de número de neuronas en la capa oculta. Finalmente, se seleccionó una capa con 10 neuronas para la habilidad y otra con 9 neuronas para el conocimiento, ya que estas configuraciones ofrecieron el mejor desempeño global según las métricas evaluadas.

Para medir el rendimiento de los modelos se utilizó el *accuracy* (Geraci et al., 1991), que indica la proporción de predicciones correctas respecto al total, así como las puntuaciones F1 (una métrica que combina precisión y exhaustividad en una única medida) en dos variantes: *macro F1*, que calcula el promedio de la F1 de cada clase sin ponderar por el tamaño de la clase, reflejando un desempeño equilibrado entre todas las categorías; y *weighted F1*, que pondera la F1 de cada clase según su número de instancias, reflejando el desempeño global considerando la distribución de clases (Manning et al., 2008). Estas métricas permiten evaluar no solo la capacidad del modelo para predecir correctamente la mayoría de los usuarios, sino también su desempeño en clases menos representadas, como los usuarios con conocimiento bajo. La Tabla 1 resume las métricas globales de accuracy. macro F1 y weighted F1 para todas las combinaciones de modelo y característica.

Tabla 1: Resultados comparativos de los modelos de clasificación

Característica	Modelo	Accuracy	Macro F1	Weighted F1
Habilidad	MLP (10 neuronas)	0,61	0,55	0,60
	Naive Bayes	0,53	0,49	0,51
	Random Forest	0,54	0,48	0,53
	SVM	0,60	0,46	0,56
Conocimiento	MLP (9 neuronas)	0,65	0,57	0,64
	Naive Bayes	0,60	0,47	0,57
	Random Forest	0,58	0,48	0,57
	SVM	0,67	0,54	0,64

En la clasificación de la habilidad, el MLP con 10 neuronas logró un accuracy de 0,61 y un weighted F1 de 0,60. La matriz de confusión (Figura 1) evidencia que MLP clasifica bien a usuarios con habilidad media y alta (expertos), pero tiene dificultades con habilidad baja (novatos),

DOI: https://doi.org/10.65234/interaccion.148

probablemente por su menor representación y la similitud en los patrones de interacción. El NB alcanzó un accuracy de 0,53 y weighted F1 de 0,51. Su matriz de confusión refleja que favorece patrones más claros, como los usuarios expertos, pero sacrifica precisión en clases minoritarias. RF mostró un accuracy de 0,54 y un weighted F1 de 0,53. La matriz de confusión revela que 9 de 28 novatos fueron correctamente clasificados, 13 se confundieron con habilidad media y 6 con expertos. El SVM alcanzó un accuracy de 0,60 y un weighted F1 de 0,56. La matriz muestra que solo 2 de 28 novatos fueron correctamente identificados.

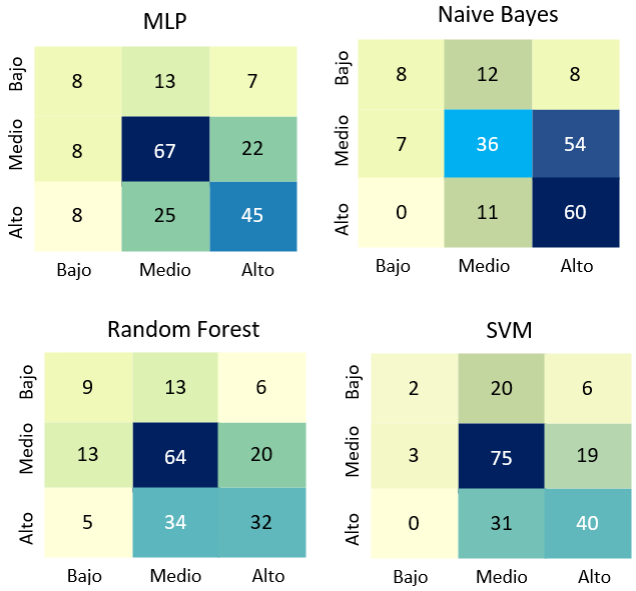

Figura 1: Matrices de confusión de la característica habilidad

Para la característica conocimiento, el MLP con nueve neuronas alcanzó un accuracy de 0,65 y weighted F1 de 0,64. NB obtuvo un accuracy de 0,60 y weighted F1 de 0,57. La matriz de confusión muestra que solo 2 de 25 usuarios con bajo conocimiento fueron correctamente identificados. El SVM alcanzó el mejor desempeño para conocimiento, con un accuracy de 0,67 y weighted F1 de 0,64. La matriz de confusión (Figura 2) muestra que 3 de 25 usuarios con bajo conocimiento fueron correctamente clasificados, 14 se confundieron con usuarios intermedios y 8 con usuarios expertos. Finalmente, RF mostró un accuracy de 0,58 y un weighted F1 de 0,57.

En conjunto, los resultados muestran que MLP y SVM son los modelos más consistentes, logrando buen desempeño en las clases medias y altas, mientras que NB y RF presentan mayor dificultad en las clases minoritarias, especialmente en las de novatos y usuarios con bajo conocimiento. En cuanto a la habilidad, MLP y SVM alcanzan el mejor desempeño en términos de accuracy y weighted F1, mientras que NB y RF presentan un rendimiento ligeramente inferior. Todos los modelos presentan dificultades para clasificar correctamente a los usuarios novatos, probablemente debido a la menor representación de esta clase y a la complejidad de las variables observadas. Para la característica de conocimiento, SVM y MLP con nueve neuronas logran el mejor desempeño global, con accuracy cercano a 0,65–0,67 y weighted F1 de 0,64, mientras que RF y NB presentan resultados ligeramente inferiores, especialmente en la clase de usuarios con conocimiento bajo.

DOI: https://doi.org/10.65234/interaccion.148

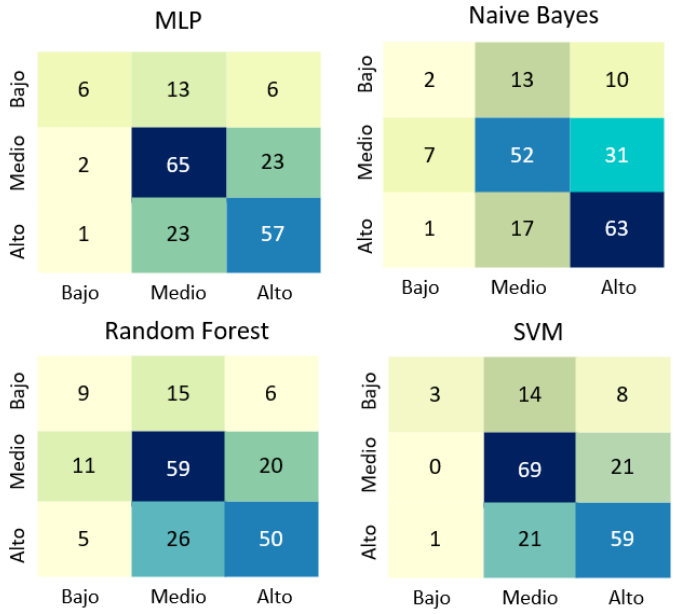

Figura 2: Matrices de confusión de la característica conocimiento

Como conclusión de los resultados, para la habilidad, el MLP (10 neuronas) y el SVM alcanzan los mejores resultados, con accuracy de 0,61 y 0,60, respectivamente, y weighted F1 de hasta 0,60. Sin embargo, todos los modelos presentan dificultades para clasificar correctamente a los usuarios novatos, debido principalmente a su menor representación y a la similitud de sus patrones con los de otras clases. En la clasificación del conocimiento, se obtienen resultados ligeramente superiores. El SVM logra el mejor rendimiento (accuracy de 0,67), seguido del MLP (9 neuronas), ambos con un weighted F1 de 0,64. A pesar de ello, los modelos presentan limitaciones para identificar a usuarios con bajo nivel de conocimiento, que tienden a confundirse con usuarios intermedios.

4. DISCUSIÓN

En esta sección se comparan los resultados obtenidos en la fase anterior con los de experimentos previos y con las expectativas de los investigadores. Respecto a la característica habilidad, nuestros resultados indican que los modelos basados en MLP y SVM lograron mayor accuracy en la clasificación de esta característica (0,61 y 0,60 de accuracy). En contraste, RF y NB mostraron un desempeño inferior. Esta jerarquía es consistente con estudios recientes. Por ejemplo, Çetinkaya et al. (2023) observaron que SVM alcanzó el mejor rendimiento (94,8%), superando a otros modelos. De forma similar, en nuestro estudio SVM presenta un rendimiento competitivo, aunque, a diferencia de dicho trabajo, MLP obtiene los mejores resultados, lo que podría explicarse por diferencias en los datos o en la formulación del problema.

Respecto a la característica conocimiento, nuestros resultados muestran valores de accuracy más moderados (entre 0,58 y 0,67). En contraste, Alruwais y Zakariah (2023) reportan precisiones de hasta el 98% utilizando distintos modelos de machine learning. Esta diferencia puede deberse al uso de variables más directamente relacionadas con el rendimiento académico en su estudio, lo que facilita la separabilidad entre clases, así como a posibles diferencias en la distribución de los datos. No obstante, se mantiene una tendencia similar: modelos más complejos, como MLP y SVM, ofrecen mejores resultados que enfoques más simples como NB.

Los resultados obtenidos, en términos de accuracy, estos no se ajustan por completo a lo esperado, aunque sí esperábamos que los algoritmos MLP y NB obtuviesen los mejores resultados. Esto se debe a la baja variabilidad de la muestra, compuesta exclusivamente por estudiantes de un grado de ingeniería, lo que implica un nivel relativamente homogéneo y medio-alto de conocimientos y habilidades. Esta característica conlleva una infrarrepresentación de la clase *novato* y de *bajos conocimientos*, lo que genera un desequilibrio entre las clases. Esta distribución no permite al algoritmo de ML identificar correctamente esa clase, lo que provoca que se clasifiquen erróneamente. Las clases habilidad intermedia y experto, al estar más representadas, obtienen un mayor accuracy, lo cual sí corresponde a lo esperado.

5. CONCLUSIONES Y TRABAJO FUTURO

Este trabajo presenta como principal contribución la comparación entre la clasificación de las características del usuario obtenida mediante algoritmos de aprendizaje automático y la realizada directamente por los propios usuarios. Para ello, se han empleado los algoritmos MLP, NB, SVN y RF, entrenados con una clasificación previamente definida. La evaluación se realiza mediante el cálculo de la correlación entre los resultados de los algoritmos de ML y los valores asignados por los propios usuarios. De este modo, se pretende determinar la fiabilidad de los modelos automáticos para representar las características del usuario en relación con sus propias expectativas. Como discusión, cabe destacar que, dado que este experimento se ha llevado a cabo con estudiantes de la Universidad de Valencia, los niveles de conocimiento y habilidades son elevados, lo que ha dificultado la correcta clasificación de los usuarios con menores competencias.

Como trabajo futuro, se plantea ampliar la muestra de sujetos en los tres conjuntos de datos para obtener resultados más robustos. Por último, se propone implementar y validar los algoritmos de clasificación en un entorno de pruebas como OpenTelemetry.

Agradecimientos: Este trabajo ha sido desarrollado con la ayuda de la Generalitat Valenciana, mediante el proyecto TENTACLE (CIAICO/2023/089), y con el soporte del Ministerio de Ciencia, Innovación y Universidades, a través de los proyectos PHYLOVAR (PID2024-162114OB-I00) y EU4IRI (PID2024-161104OB-C22) financiados por MICIU/AEI/10.13039/501100011033 y por FEDER.

6. REFERENCIAS

Abourobea, M., Atia, M., & Ismail, K. (2025). Comparative Study of Machine Learning Approaches for Fixed Radar-Based Road User Classification. *2025 International Conference on Artificial Intelligence, Computer, Data Sciences and Applications (ACDSA)*, 1-6. https://doi.org/https://www.doi.org/10.1109/ACDSA65407.2025.11165993

Abri, S., Abri, R., & Çetin, S. (2020). A Classification on Different Aspects of User Modelling in Personalized Web Search. *NLPIR 2020: 4th International Conference on Natural Language Processing and Information Retrieval, Seoul, Republic of Korea, December 18-20, 2020*, 194-199. https://doi.org/https://doi.org/10.1145/3443279.3443291

Alruwais, N., & Zakariah, M. (2023). Evaluating Student Knowledge Assessment Using Machine Learning Techniques. *Sustainability*, *15*(7). https://doi.org/10.3390/su15076229

Alwafi, G., & Fakieh, B. (2024). A machine learning model to predict privacy fatigued users from social media personalized advertisements. *Scientific Reports*, *14*(1), 3685. https://doi.org/https://doi.org/10.1038/s41598-024-54078-w

Basuni, N. S., & Siregar, A. M. (2023). Comparison of the Accuracy of Drug User Classification Models Using Machine Learning Methods. *Journal RESTI*, *7*(6), 1348-1358. https://doi.org/https://www.doi.org/10.29207/resti.v7i6.5401

Brusilovsky, P. (1996). Methods and techniques of adaptive hypermedia. En *User Modeling and User-Adapted Interaction* (Vol. 6, Número 2, pp. 87-129). https://doi.org/10.1007/BF00143964

Çetinkaya, A., Baykan, Ö. K., & Kırgız, H. (2023). Analysis of Machine Learning Classification Approaches for Predicting Students' Programming Aptitude. *Sustainability*, *15*(17). https://doi.org/10.3390/su151712917

Duan, S., Wang, Z., Wang, S., Chen, M., & Zhang, R. (2024). Emotion-Aware Interaction Design in Intelligent User Interface Using Multi-Modal Deep Learning. *CoRR*, *abs/2411.06326*. https://doi.org/10.48550/ARXIV.2411.06326

Gaspar, A., Gil, M., Panach, J. I., & Romero, V. (2024). Towards a general user model to develop intelligent user interfaces. *Multim. Tools Appl.*, *83*(26), 67501-67534. https://doi.org/https://doi.org/10.1007/s11042-024-18240-w

Gaspar, A., Panach Navarrete, J. I., Gil, M., Romero Gómez, V., & Distante, D. (2026). *Análisis de Características de Usuario con Machine Learning: Material suplementario*. Zenodo. https://doi.org/10.5281/ZENODO.20848896

Geraci, A., Katki, F., McMonegal, L., Meyer, B., Lane, J., Wilson, P., Radatz, J., Yee, M., Porteous, H., & Springsteel, F. (1991). *IEEE Standard Computer Dictionary: Compilation of IEEE Standard Computer Glossaries*. IEEE Press. https://doi.org/https://dl.acm.org/doi/abs/10.5555/574566

Jaber, A., & Csonka, B. (2023). Investigating the temporal differences among bike-sharing users through comparative analysis based on count, time series, and data mining models. *Alexandria Engineering Journal*, *77*, 1-13. https://doi.org/https://doi.org/10.1016/j.aej.2023.06.087

Kadhim, A. I. (2019). Survey on supervised machine learning techniques for automatic text classification. *Artificial Intelligence Review*, *52*(1), 273-292. https://doi.org/10.1007/s10462-018-09677-1

Kristić, M., Zakarija, I., Škopljanac-Mačina, F., & Car, Ž. (2025). Machine Learning for Adaptive Accessible User Interfaces: Overview and Applications. *Applied Sciences*, *15*(23). https://doi.org/10.3390/app152312538

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. Cambridge University Press. https://doi.org/10.1017/CBO9780511809071

Murphy, K. P. (2012). *Machine learning—A probabilistic perspective*. MIT Press.

Saranya, S., Malligeswari, N., Graf, F. T., & Murugan, V. (2025). Malicious user classification in cognitive 5G networks using novel improved bidirectional encoder representations from transformers model. *Scientific Reports*, *15*(1), 43415. https://doi.org/https://doi.org/10.1038/s41598-025-19156-7

Schaible, M., Unger, H., Besier, A., & Mur, E. (2024). Adaptive Learning of User's Interests. *Proceedings of the 2023 7th International Conference on Natural Language Processing and Information Retrieval, NLPIR '23*, 261-265. https://doi.org/https://doi.org/10.1145/3639233.3639336

Shannaq, B. (2024). IMPROVING SECURITY IN INTELLIGENT SYSTEMS: HOW EFFECTIVE ARE MACHINE LEARNING MODELS WITH TF-IDF VECTORIZATION FOR PASSWORD-BASED USER CLASSIFICATION. *Journal of Theoretical and Applied Information Technology*, *102*(22), 8340-8355.

Siino, M., Di Nuovo, E., Tinnirello, I., & La Cascia, M. (2022). Fake News Spreaders Detection: Sometimes Attention Is Not All You Need. *Information*, *13*(9). https://doi.org/https://www.doi.org/10.3390/info13090426

Tachie-Donkor, G., & Ezema, I. J. (2023). Effect of information literacy skills on university students' information seeking behaviour and lifelong learning. *Heliyon*, *9*(8), e18427. https://doi.org/https://doi.org/10.1016/j.heliyon.2023.e18427

Yuvapriya, P., Subramanian, P., & Surendran, R. (2025). Performance Analysis of Information Technology Professionals Using Optimal User Clustering and Machine Learning Model for Adaptive e-learning. *2025 8th International Conference on Trends in Electronics and Informatics (ICOEI)*, 1159-1165. https://doi.org/https://www.doi.org/10.1109/ICOEI65986.2025.11013491

Zahoor, K., & Bawany, N. Z. (2024). Explainable artificial intelligence approach towards classifying educational android app reviews using deep learning. *Interact. Learn. Environ.*, *32*(9), 5227-5252. https://doi.org/https://doi.org/10.1080/10494820.2023.2212708