

Analítica del aprendizaje sobre clases sincrónicas en entorno virtual mediante técnicas de procesamiento de lenguaje natural

Learning Analytics on Synchronous Classes in Virtual Environments Using Natural Language Processing Techniques

Marcelo Dante Caiafa⁽¹⁾, Alejandro Bevilacqua⁽²⁾, Gustavo Mauricio Rosales⁽³⁾

⁽¹⁾Departamento de Ingeniería e Investigaciones Tecnológicas
Universidad Nacional de La Matanza, Argentina
mcaiafa@unlam.edu.ar

⁽²⁾Departamento de Ingeniería e Investigaciones Tecnológicas
Universidad Nacional de La Matanza, Argentina
abevilacqua@unlam.edu.ar

⁽³⁾Departamento de Ingeniería e Investigaciones Tecnológicas
Universidad Nacional de La Matanza, Argentina
grosales@unlam.edu.ar

Resumen:

La analítica del aprendizaje se refiere a la medición, recopilación, análisis e informes de datos relacionados con los estudiantes y sus entornos de aprendizaje.

La medición del desempeño sobre ambientes virtuales de aprendizaje se ha convertido en una interesante herramienta para comprender y optimizar la experiencia educativa en plataformas digitales.

El enfoque aborda la implementación progresiva de analítica avanzada de aprendizaje para comprender y mejorar el proceso de aprendizaje a partir del análisis de los datos generados por las interacciones de los estudiantes en clases virtuales sincrónicas.

Este trabajo se basa en la utilización de técnicas de procesamiento de lenguaje natural para identificar las distintas variables a partir de las transcripciones de los registros grabados durante las clases y la aplicación de modelos de aprendizaje automático para medir su desempeño.

El estudio propone analizar las interacciones mediante la identificación de variables y la elaboración de los indicadores de desempeño relacionados a las dimensiones conductual, emocional y cognitiva.

Los resultados pretenden aportar valor en la mejora de los procesos educativos en línea para optimizar la experiencia de aprendizaje.

Abstract:

Learning analytics refers to the measurement, collection, analysis, and reporting of data about learners and their learning environments.

Performance measurement in virtual learning environments has become a valuable tool for understanding and optimizing the educational experience on digital platforms.

This paper focuses on the progressive implementation of advanced learning analytics to understand and enhance the learning process through the analysis of data generated by student interactions in synchronous virtual classes.

The study is based on the use of natural language processing techniques to identify various variables from transcripts of recorded class sessions, along with the application of machine learning models to evaluate performance.

The approach aims to analyze student interactions by identifying variables and developing performance indicators related to behavioral, emotional, and cognitive dimensions.

The results seek to contribute to the improvement of online educational processes by optimizing the learning experience.

Palabras Clave: *aprendizaje automático, procesamiento de lenguaje natural, ambiente virtual de aprendizaje, analítica de aprendizaje*

Key Words: *machine learning, natural language processing, virtual learning environment, learning analytics*

I. INTRODUCCIÓN

La inteligencia artificial (IA) ha avanzado significativamente desde que John McCarthy fundó el primer laboratorio en la universidad de Stanford en 1956 [1]. El aprendizaje automático, conocido por sus siglas en inglés como ML (machine learning), es un subcampo de la IA que implica el uso de algoritmos y modelos estadísticos para permitir que las máquinas aprendan de los datos, dándole capacidad para hacer predicciones, detectar patrones y tomar decisiones sin necesidad de estar programadas explícitamente.

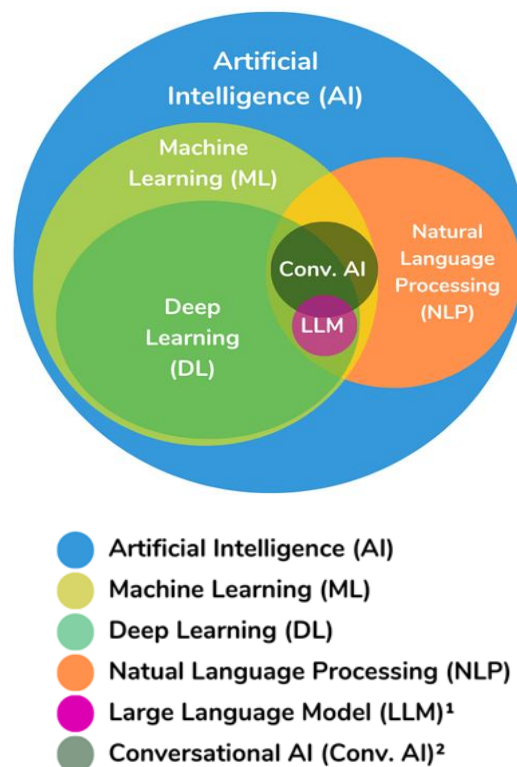
El procesamiento de lenguaje natural, conocido por sus siglas en inglés como NLP (Natural Language Processing), es un tipo de tecnología de machine learning que brinda a las computadoras la capacidad de interpretar, manipular y comprender el lenguaje humano.

La analítica de aprendizaje y la IA han crecido junto al desarrollo de algoritmos para hacer recomendaciones, tomar decisiones y aprender en diferentes contextos. Estas diferentes técnicas emergen como herramientas para mejorar la calidad educativa de los ambientes virtuales de aprendizaje [2].

El desempeño en las clases en línea adquiere particular relevancia en los desarrollos de clases sincrónicas virtuales. La disponibilidad de datos posibilita el uso de técnicas analíticas de IA para evaluar distintos niveles de interacción [3].

La efectividad de los entornos educativos digitales es abordada por algunos autores en dimensión conductual, emocional y cognitiva [4]. La participación en línea puede evaluarse desde dichas dimensiones, ya sea la interacción con el contenido, la cantidad y calidad de las respuestas o el nivel de compromiso con la propuesta, entre otras. Estas también dependen de los factores emocionales que influyen en el desempeño académico [5].

Respecto de las técnicas analíticas se muestran, a continuación las diferentes categorías de IA, destacando al NLP y ML como parte de ello.



¹LLM is an intersection of DL and NLP

²Conversational AI is a combination of ML and NLP. It may include DL and LLM.

Figura 1: Categorías de inteligencia artificial
Fuente: Gennadiy Bezko, blog MiaRec

El NLP se enfoca en la interacción entre las computadoras y el lenguaje humano, permitiendo a las máquinas interpretar texto, el ML se centra en algoritmos que permiten aprender de datos. Ambos utilizan modelos matemáticos para emular aspectos cognitivos [6] y ser capaces de analizar datos para extraer patrones [7] y tienen la capacidad de procesar textos, identificar emociones, evaluar la calidad de las respuestas [8].

El NLP se aplica para examinar las respuestas textuales de los estudiantes, analizar las transcripciones de las clases grabadas, y evaluar las interacciones de los estudiantes [9]. Facilitar la identificación de emociones es útil para medir el compromiso y el enfoque de los estudiantes [10]. Las universidades han adoptado plataformas digitales que permitan visualizar el avance, la

interacción y el compromiso del alumnado [11]. El uso de modelos puede revelar tendencias que serían difíciles de identificar [12]. Los modelos de *Redes Neuronales Recurrentes* y *Transformers* (como BERT) son capaces de captar relaciones complejas y patrones de secuencia, permiten una evaluación de las interacciones [13].

II. PROBLEMÁTICA PARA RESOLVER

Según McMurtrie, la IA se está utilizando con más frecuencia en la educación, especialmente en las instituciones de educación superior que buscan hacer que los cursos con un alumnado numeroso sean más interactivos [14].

A partir de las plataformas de educación en línea se generan grandes volúmenes de datos derivados de las interacciones como foros, chats, y transcripciones de clases grabadas. La disponibilidad de datos para analizarlos objetivamente genera las siguientes preguntas:

- ¿Cuáles son los parámetros a identificar para realizar un análisis cuantitativo del desempeño en ambientes virtuales de aprendizaje?
- ¿Cómo se pueden utilizar técnicas de NLP y ML para generar herramientas que permitan medir el desempeño de estudiantes en las clases virtuales sincrónicas?

III. OBJETIVO DEL TRABAJO

El objetivo del presente trabajo es el desarrollo de un sistema automatizado basado en técnicas de ML y NLP para analizar y medir la participación de docentes y estudiantes en las clases, utilizando como base los registros correspondientes a la grabación de clases sincrónicas en entornos virtuales.

IV. OBJETIVOS SECUNDARIOS

1. Construir los diferentes modelos que permitan la identificación de variables para un análisis

cuantitativo del desempeño de los estudiantes en clases virtuales sincrónicas.

2. Elaborar los indicadores de desempeño en los ámbitos cognitivo, emocional y social a partir de la medición de las variables identificadas.

3. Desarrollar una herramienta de software para medir, calcular y representar los resultados de los distintos indicadores de desempeño.

V. DESARROLLO DEL TRABAJO

El trabajo se estructura en cinco etapas ejecutadas secuencialmente. Se utiliza Python como lenguaje de programación por considerarse adecuado por sus características. Se trata de un lenguaje interpretado, de alto nivel, de tipado dinámico, multiplataforma, de código abierto, orientado a objeto, gran comunidad, diversidad librerías, etc [15].

Etapa 1: Recolección de los Datos

Se trabajó sobre un conjunto de datos no estructurados, provenientes de los textos resultantes de las transcripciones de la recopilación de las diez clases grabadas de física, correspondientes al primer año de ingeniería de la UNLAM del curso de verano.

Como herramienta de grabación de los registros de audio, se aprovechó la plataforma de colaboración MS Teams, que es la misma que regularmente utiliza la universidad para dictar las clases sincrónicas virtuales. Esta genera los registros de las grabaciones en formato de video mp4. A partir de allí se extrajo el audio para su posterior procesamiento en formato mp3.

El primer paso consistió en transcribir las automáticamente para construir el conjunto de datos a analizar, proceso conocido como STT por sus siglas en inglés (speech to text). Para ello se utilizó una integración con la librería www.assemblyai.com mediante API pública.

```
import assemblyai as aai
import pandas as pd
import nltk from nltk.tokenize
import sent_tokenize
#Defino llave de API
aai.settings.api_key = "87808a54e93945f2b2632cb01e740bd7"
#Seteo idioma y activo la detección de hablantes. Agrego rutas de files
config = aai.TranscriptionConfig(language_code="es", speaker_labels=True)
audiofile_path1="C:\Users\rosal\Downloads\ClaseFisica1_prueba.mp3"
audiofile_path2="C:\Users\rosal\Downloads\Clase2Fisica1_prueba.mp3"
audiofile_path3="C:\Users\rosal\Downloads\Clase3Fisica1_prueba.mp3"
audiofile_path5="C:\Users\rosal\Downloads\clase5Fisica1_prueba.mp3"
audiofile_path6="C:\Users\rosal\Downloads\clase6Fisica1_prueba.mp3"
audiofile_path9="C:\Users\rosal\Downloads\clase9Fisica1_prueba.mp3"
audiofile_path10="C:\Users\rosal\Downloads\clase10Fisica1_prueba.mp3"
#Listas con las que vamos a trabajar
listaPreguntas=[]
listaRespuestas=[]
listaHablaantes=[]
listaDiscursos=[]
#Configuraciones iniciales
def generarTranscripcion(audiofile_path,config,nombreArchiv):
    transcriber = aai.Transcriber(config=config)
    transcript = transcriber.transcribe(audiofile_path)
    transcripcion_txt=transcript.text #Genera texto(string)
    #Almacenamos los textos y los hablantes en listas
    for utterance in transcript.utterances:
        speaker = utterance.speaker
        text = utterance.text
        listaHablaantes.append(speaker)
        listaDiscursos.append(text)
```

Figura 2: Parte del script de integración API
Fuente: Elaboración propia

El resultado se contrastó con un procesamiento similar, mediante el algoritmo speech-recognition de la librería de Python.

Etapas 2: Preprocesamiento de los Datos

El preprocesamiento de datos es una etapa donde se adecúa el texto para el análisis posterior, esto incluye los siguientes pasos:

Limpieza de texto: elimina errores, caracteres especiales, no deseados, y otros que afecten.

Tokenización: segmenta el texto en unidades más pequeñas palabras, frases o párrafos.

Eliminación de stopwords: elimina palabras vacías que no aportan información relevante.

Lematización/Stemming: transforma cada palabra a su forma raíz, para identificar las variaciones de una misma palabra.

Identificación de turnos del habla: segmenta las intervenciones identificando al hablante.

Etapas 3: Elaboración de indicadores

El **indicador conductual** refiere al grado de involucramiento de los estudiantes con las actividades propuestas. Los modelos de NLP

pueden medir el compromiso mediante el análisis de la frecuencia de interacción y la pertinencia de las intervenciones dentro de la clase [16].

```
display(df_Preguntas)
display(df_Original)
#display(df_respuestas)
```

	Hablante	Preguntas
0	A	¿Cuál es el valor de la aceleración normal de ...
1	A	¿Cuál es la velocidad de la pelota al impactar...
2	A	¿Esto está visto desde arriba, no?
3	A	¿Si yo corto la cuerda acá, para dónde se va l...
4	A	¿Estoy entendiendo bien el enunciado?
...	...	...
2195	A	¿Está bien?
2196	A	¿Alguna duda, alguna inquietud, algo que quieran?
2197	A	¿Quieren parar 5 min o siguen derecho?
2198	A	¿Cómo hacerlo normalmente?
2199	A	¿Cuánto tiempo ponemos?

[2200 rows x 2 columns]

	Hablante	Discurso
0	A	Ya el dibujito medio que me asusta. Una pelota...
1	B	¿Tipo en diagonal con la tangencial?
2	A	¿Para allá, no? Está bien. Entonces, esta velo...
3	B	No lo estaba entendiendo porque no lo veía des...
4	A	Lo leo de nuevo por las dudas. La pelota acá, ...
...	...	...
1315	A	Ahí en algún otro, en alguna otra instancia.
1316	C	Buenísimo, profe, muchas gracias.
1317	A	Muchas gracias.

Figura 3: Muestra de intervenciones
Fuente: Elaboración propia

Se contabiliza la cantidad de preguntas detectadas, ya sean formuladas y/o respondidas, y se obtiene un nivel de compromiso de cada alumno. Dado que cada una de estas métricas tiene diferentes escalas, es preciso normalizarlas para generar un indicador consolidado a partir de ellas.

Se mide la cantidad de interacciones [17], se utiliza la normalización Min-Max [18] y luego, se calcula el valor del indicador de compromiso. Es importante aclarar que suele hacerse referencia a los indicadores de desempeño como KPI por sus siglas en inglés (Key Performance Indicator)

Cálculo de KPI compromiso

Este KPI de compromiso evalúa el nivel de participación activa del estudiante durante la clase.

Se define KPI de Compromiso = KPI_C

$$KPI_C = (N_int + N_preg + N_resp) / T_tot$$

- N_int: Número de intervenciones del estudiante.
- N_preg: Número de preguntas formuladas.

- N_resp: Número de respuestas dadas.
- T_tot: Total de turnos de habla registrados en clase.

Rango estimado de interpretación

- < 0.30: Participación baja
- 0.31 – 0.60: Participación moderada
- > 0.60: Participación alta

Se observa en la siguiente gráfica el indicador compromiso de los estudiantes con mayor participación en las diferentes clases.

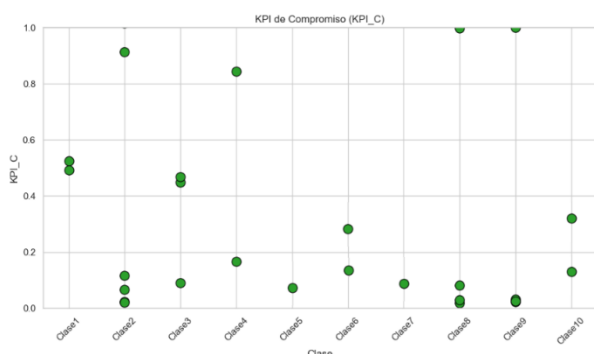


Figura 4: Distribución del índice KPI_C
Fuente: Elaboración propia

Etap 4: Modelado de Machine Learning

Para el tratamiento del indicador emocional y cognitivo se utilizó ML con distintos tipos de modelo analítico.

El **indicador emocional** refiere a la percepción subjetiva que los estudiantes tienen de la clase. El análisis de los comentarios y de sus intervenciones proporciona información sobre su nivel de satisfacción. Las técnicas de *topic modeling* (modelo de temas) permite identificar los temas de interés [19]. En este caso el algoritmo utilizado es el modelo de análisis de sentimientos BERT (Bidirectional Encoder Representations from Transformers) que dispone de la capacidad de

comprender el contexto bidireccional de las palabras en una secuencia [20].

```
#Realizamos análisis de sentimientos de estos dos dataframes

from transformers import pipeline

# Cargamos el modelo preentrenado de análisis de sentimientos
classifier = pipeline('sentiment-analysis',
                      model="nlpTown/bert-base-multilingual-uncased-
                      sentiment",
                      truncation=True) # Truncar textos largos a 512
                      tokens

def obtener_polaridad_score(pregunta):
    result = classifier(pregunta) # Realiza el análisis de
    sentimiento
    # Devolvemos el label (polaridad) de la respuesta, puede ser
    'LABEL_0', 'LABEL_1', etc.
```

Figura 5: Parte del script de análisis de emociones
Fuente: Elaboración propia

Cálculo de KPI experiencia

Este indicador pretende medir la calidad emocional en las interacciones del estudiante durante la clase.

Se define KPI de Experiencia = KPI_E

$$KPI_E = ((S_{pos} - S_{neg}) / S_{tot})$$

- S_pos: Sentimientos positivos detectados.
- S_neg: Sentimientos negativos detectados.
- S_tot: Total de frases analizadas en lo emocional

Rango estimado de interpretación

- < 0.5: Experiencia deficiente
- 0.51 – 1: Experiencia aceptable
- > 1: Experiencia enriquecida

Se observa en la siguiente gráfica los KPI_E de los estudiantes con los valores más altos en las diferentes clases.

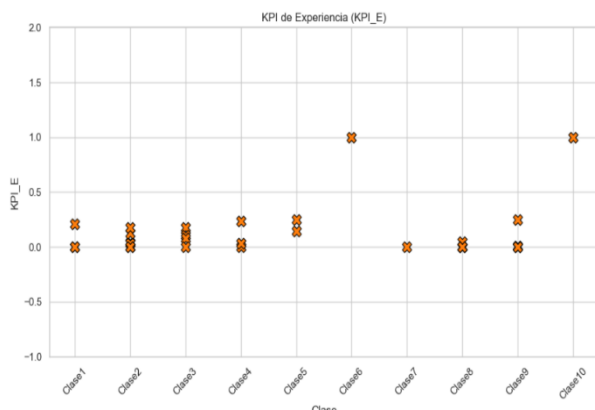


Figura 6: Distribución del índice KPI_E
Fuente: Elaboración propia

El **indicador cognitivo** mide la atención que los estudiantes mantienen sobre el contenido de la clase. A través de las interacciones, permite inferir si un estudiante está siguiendo la temática del curso. Se construye mediante la relevancia de intervenciones de los estudiantes en relación con unidades temáticas [21]. Se clasifican los tipos de preguntas para luego relacionarlas con temáticas particulares.

Se realizó el modelado de ML mediante la librería scikitlearn de Python con algoritmo de Naives Bayes para clasificar los tipos de preguntas.

```
# Iterar sobre las preguntas en el DataFrame
for i, pregunta in enumerate(df_Preguntas["Preguntas"]):
    doc = nlp(pregunta)
    texto_encontrado = None
    tematica_encontrada = None

    # Iterar sobre las entidades reconocidas
    for ent in doc.ents:
        # Verificar si la entidad coincide con las claves del
        # diccionario
        if ent.text in tematicas_dict:
            texto_encontrado = ent.text
            tematica_encontrada = tematicas_dict[ent.text]
            break # Detener la búsqueda al encontrar la primera
            coincidencia

    # Asignar los valores encontrados a la fila actual del DataFrame
    df_Preguntas.loc[i, "Texto"] = texto_encontrado
    df_Preguntas.loc[i, "Tematica"] = tematica_encontrada

# Mostrar el DataFrame actualizado
df_preg_cinematica=df_Preguntas[df_Preguntas["Tematica"] ==
'Cinematica']
df_preg_dinamica=df_Preguntas[df_Preguntas["Tematica"] == 'Dinamica']

#Cantidad de preguntas realizadas de cada tematica
cantpreg_x_tematica=df_Preguntas.groupby('Tematica').size()
print(cantpreg_x_tematica)

display(df_preg_cinematica)
display(df_preg_dinamica)
```

Figura 7: Segmento del script para identificar unidades temáticas
Fuente: Elaboración propia

Cálculo de KPI enfoque

Este indicador pretende medir la alineación temática de las intervenciones con los objetivos curriculares de la clase

Se define KPI de Enfoque = KPI_F

$$KPI_F = (C_clave / C_tot) * R_coh$$

- C_clave: Número de conceptos clave mencionados.
- C_tot: Total de conceptos tratados en clase.
- R_coh: Índice de coherencia semántica entre segmentos del discurso (calculado comparado con diccionario de términos).

Estas fórmulas pueden ser ajustadas/ponderadas según el curso, estilo docente o criterios institucionales.

Rango estimado de interpretación:

- < 0.40: Bajo enfoque
- 0.41 – 0.70: Enfoque intermedio
- > 0.70: Enfoque elevado

Se observa en la siguiente gráfica los KPI_F de los estudiantes con los valores más altos en las diferentes clases.

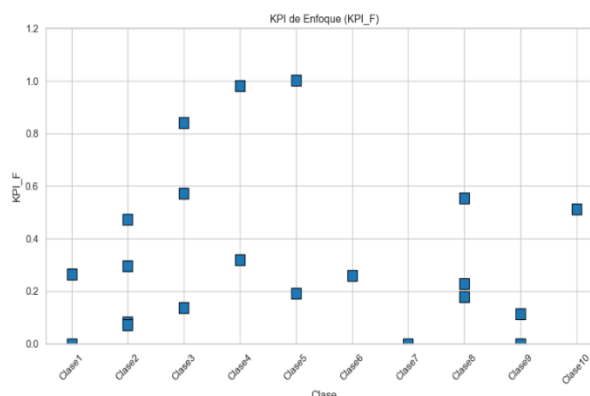


Figura 8: Distribución del índice KPI_F
Fuente: Elaboración propia

Etapa 5: Evaluación y Validación

Para garantizar la precisión y fiabilidad de los modelos, se emplea validación cruzada (cross-validation), que permite dividir los datos en subconjuntos de entrenamiento y prueba.

VI. RESULTADOS

En la distribución de la variable preguntas formuladas por clase se observa una clara preponderancia del hablante A, identificado como el docente. El resto de los hablantes, desde B, C, etc en adelante, corresponden a diferentes alumnos.

La participación de los estudiantes en términos de preguntas formuladas, alcanza su máximo en la clase 9, precisamente la que corresponde como clase previa a la fecha prevista de examen.

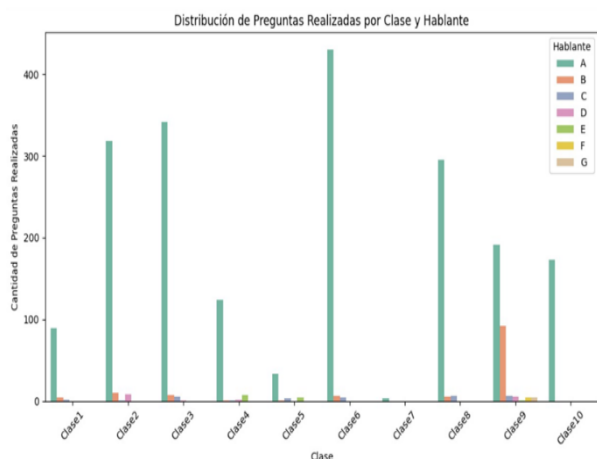


Figura 9: Distribución de preguntas realizadas
Fuente: Elaboración propia

En el siguiente gráfico se observa la distribución de respuestas y se destaca cierto balance entre las respuestas del docente y la de los alumnos en la mayoría de las clases. En la clase 9, se destaca la participación del alumno identificado como hablante B siendo su cantidad de intervenciones aún mayor a la del docente (hablante A).

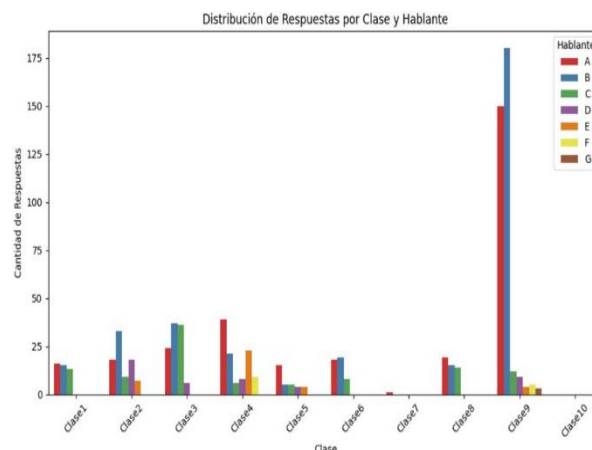


Figura 10: Distribución de respuestas formuladas
Fuente: Elaboración propia

Para enriquecer el análisis resulta interesante realizar un cruce entre preguntas y las respuestas de docente y estudiante respectivamente.

Para eso se registra en la siguiente figura y se comprueba que no existe adecuado balance entre ambas. Es importante aclarar que se grafica la cantidad de respuestas de alumnos en función de preguntas de docente para las distintas clases. Se observa que la cantidad de preguntas del docente superan en un orden de magnitud a las respuestas de los estudiantes.

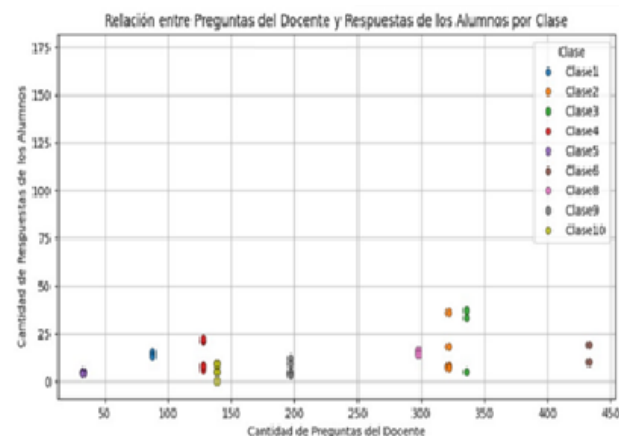


Figura 11: Cruce de preguntas de docente y respuestas de alumnos
Fuente: Elaboración propia

Se realiza ahora un segundo cruce de variables, de forma de graficar la cantidad de preguntas formuladas por los estudiantes en función de las

respuestas del docente. En este caso se asemeja a recta con pendiente de $m=0.5$, lo que refleja cierto equilibrio en la cantidad de participaciones, por cada pregunta de alumno hay dos respuestas del docente.

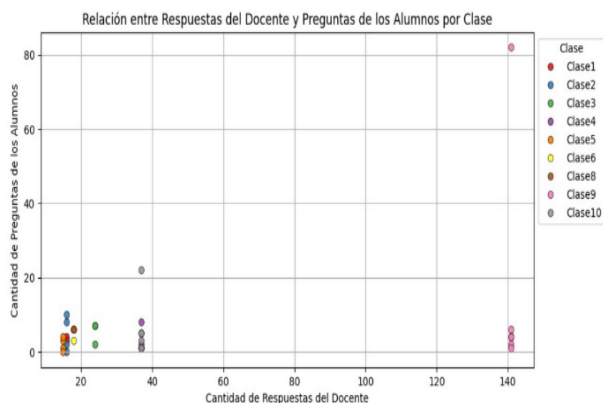


Figura 12: Cruce de preguntas de alumnos y respuestas de docente

Fuente: Elaboración propia

VII. CONCLUSIONES

El proyecto de investigación permitió valorar la utilización de modelos de ML dentro de la experiencia del proceso enseñanza-aprendizaje como una competencia necesaria dentro del entorno analítico de datos.

El trabajo ofreció una aproximación al contexto productivo de las técnicas del NLP con la limitación que depende de un postprocesamiento, si bien resulta útil para una evaluación no aplica en tiempo real.

Los resultados obtenidos han validado que es posible clasificar los niveles de participación de los estudiantes y docentes utilizando las tecnologías referenciadas y los indicadores propuestos tomados como referencia.

La combinación de NLP para el análisis de texto y ML para la clasificación de las interacciones ha permitido identificar patrones relevantes en el desempeño de los estudiantes, su compromiso en la participación activa, como la intensidad emocional

en las respuestas, y la calidad de la experiencia educativa.

Está previsto continuar con la recopilación de datos para seguir nutriendo al corpus con las transcripciones de las clases correspondientes al primer cuatrimestre de la cursada, para seguir entrenando al modelo y mejorar su comportamiento.

Como futuros trabajos está previsto avanzar en el análisis predictivo de las tasas de graduación, las tasas de abandono y el rendimiento académico. Se considera la posibilidad de construir un gemelo digital áulico para conocer las potenciales respuestas de los estudiantes frente a situaciones educativas eventuales. También se propone un sistema que, además, permitiría la intervención en tiempo real de lo que está sucediendo en clase como ayuda en línea al docente.

VIII. BIBLIOGRAFÍA

- [1] Andresen S.L. (2002) John McCarthy: father of Artificial Intelligence. *IEEE Intelligent Systems* (vol 17, art 5, pag 84 - 85)
- [2] Chen, L., Chen, P., & Lin, Z. (2020). Artificial Intelligence in Education: a review. *IEEE Access*, 8, 75264-75278.
- [3] Martin, F., Sunley, R., & Turner, P. (2019). "Student engagement in online learning environments: A literature review." *The International Journal of Information and Learning Technology*, 36(4), 284-29.
- [4] Fredricks, Blumenfeld & Paris, A. H. (2004). "School engagement: Potential of the concept, state of the evidence." *Review of Educational Research*, 74(1), 59-109
- [5] Bower, M., Cavanagh, M., & Moloney, R. (2015). "Designing a Blended Learning Environment: A Case Study of Teaching Business Information Systems." *International Journal of*

Educational Technology in Higher Education,
12(1), 1-15

[6] Woolf, B. P. (2010). *Building Intelligent Interactive Tutors: Student-Centered Strategies for Learning*. Ed. Elsevier.

[7] Davenport, T. H., & Ronanki, R. (2018). "Artificial intelligence for the real world." *Harvard Business Review*, 96(1), 108-116

[8] Jurafsky, D., & Martin, J. H. (2020). "*Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*". Ed. Pearson.

[9] Baker, R. S., D'Mello, S. K., Rodrigo, M. M., & Graesser, A. C. (2019). "Better understanding student engagement and learning through data mining and machine learning techniques." *Computers in Human Behavior*, 100, 68-75.

[10] Machova K., Szaboova M., Paralic J., & Mikco (2023). "Detection of emotion by text analysis using machine learning." *Page 3*.

[11] Siemens, G. (2013). "Learning analytics: The emergence of a discipline." *American Behavioral Scientist*, 57(10), 1380-1400

[12] McMurtrie Beth (2020). Preparing for different kind of classroom. The Chronicle of higher education Inc. (art 1, pag 5-8). Washington

[13] Baker, R.S., & Yacef, K (2009). "The state of educational data mining in 2009". Proceedings of 2nd Conference on educational Data Mining, 3-13

[14] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., Kaiser, Ł., &

Polosukhin, I. (2017). "Attention is all you need." *Proceedings of the 31st International Conference on Neural Information Processing Systems* (NeurIPS 2017), 5998-6008.

[15] Bel W., *Algoritmos y estructuras de datos en Python*. Un enfoque ágil. Facultad Ciencia y Tec. (2020). pag17. Ed Uader.

[16] Garrison, D. R., Anderson, T., & Archer, W. (2000). "Critical thinking, cognitive presence, and computer conferencing in distance education." *American Journal of Distance Education*, 15, 7-23

[17] Freeman, S., et al. (2014). *Active learning increases student performance in science, engineering, and mathematics*. Proceedings of the National Academy of Sciences, 111(23), 8410-8415.

[18] Kotsiantis, S. B., & Pintelas, P. E. (2004). *Data Preprocessing for Supervised Learning*. International Journal of Computer Science, 1(1), 111-116.

[19] Zhao X., Jiang Z. & Gray J., (2020) "*Text Classification and Topic Modeling for Online Discussion Forums*". Trends and Applications of Text Summarization Techniques. (ch 6) 151-166

[20] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." *Proceedings of NAACL-HLT*, 4171-4186

[21] Blei, D. M., Ng, A. Y., & Lafferty, J. D. (2003). "Latent Dirichlet allocation." *Journal of Machine Learning Research*, 3, 993-1022



Recibido: 2025-04-20

Aprobado: 2025-07-25

Hipervínculo Permanente: <https://doi.org/10.54789/reddi.10.1.3>

Datos de edición: Vol. 10 - Nro. 1 - Art. 3

Fecha de edición: 2025-07-31

