

PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ

Escuela de Posgrado



Inteligencia Artificial para la mejora de la evaluación
docente en una universidad privada peruana

Tesis para obtener el grado de Maestro en Integración e Innovación
Educativa de las Tecnologías de la Información y la Comunicación que
presenta:

Arturo Gustavo Díaz Rosemberg

Asesora:

Carol Rivero Panaqué


Lima, 2024

Informe de Similitud

Yo, Carol Rivero Panaqué, docente de la Escuela de Posgrado de la Pontificia Universidad Católica del Perú, asesora de la tesis de investigación titulado: Inteligencia Artificial para la mejora de la evaluación docente en una universidad privada peruana del autor Arturo Gustavo Díaz Rosemberg, dejo constancia de lo siguiente:

- El mencionado documento tiene un índice de puntuación de similitud de **15%**. Así lo consigna el reporte de similitud emitido por el software Turnitin el 05/11/2024.
- He revisado con detalle dicho reporte y la Tesis, y no se advierte indicios de plagio.
- Las citas a otros autores y sus respectivas referencias cumplen con las pautas académicas.

Lima, 1 de diciembre de 2024

<u>Apellidos y nombres de la asesora: Carol Rivero Panaqué</u>	
DNI: 29734633	Firma
ORCID: 0000-0002-0123-8999	

DEDICATORIA

A mi querida Lu, cada página lleva tu luz y tu sonrisa. Desde que naciste, has sido mi mayor inspiración y fortaleza.



AGRADECIMIENTOS

A Pame, Lu, mi mamá, mi hermana Dora y toda mi familia: gracias por entender mis momentos de estrés y las horas que dediqué a este proyecto.

Un agradecimiento especial a mi asesora Carol, quien además de guiarme me acompañó con gran paciencia en este proceso. Su orientación y apoyo fueron claves para completar esta tesis.



RESUMEN

El presente trabajo se desarrolla en una universidad peruana con sedes a nivel nacional y que actualmente cuenta con más de 72,000 estudiantes y 3,666 docentes. Para esta institución, al tener una plana docente y estudiantil considerable, se vuelve prioritario el contar con información cuantificable de la percepción de los alumnos hacia el desempeño de los docentes. Para solucionar esta necesidad se propone el análisis de las encuestas de los alumnos, haciendo uso de un análisis de sentimiento y la clasificación de las respuestas de una pregunta abierta, a partir de la inteligencia artificial.

Los objetivos de esta propuesta de innovación se enfocan en optimizar la interpretación de los resultados de encuestas de evaluación docente a partir de la inteligencia artificial, mediante la automatización del análisis de las preguntas abiertas de las encuestas utilizando procesamiento de lenguaje natural y con estos resultados proveer métricas cualitativas referenciales. Finalmente, utilizando la clasificación se plantea entrenar un modelo que pueda identificar respuestas resaltantes que ayuden a identificar un problema particular con un curso o un docente.

La solución planteada busca ser una herramienta que permita actuar con rapidez, por parte de los gestores de la universidad y poder considerar las respuestas abiertas de las encuestas. Esta información, frecuentemente susceptible de pasar inadvertida, cobra especial relevancia en entidades de gran envergadura como la presente institución académica.

Palabras clave: evaluación docente, inteligencia artificial, procesamiento de lenguaje natural.

ABSTRACT

This work is carried out at a Peruvian university with multiple campuses, which currently has more than 72,000 students and 3,666 teachers. For this institution, having a considerable teaching staff and student body, it becomes a priority to have quantifiable information on students' perception of teachers' performance. To address this need, the analysis of student surveys is proposed, using sentiment analysis and classification of responses to an open-ended question, based on artificial intelligence.

The objectives of this innovation proposal focus on optimizing the interpretation of teacher evaluation survey results using artificial intelligence. Specifically, it seeks to automate the analysis of open-ended survey questions using natural language processing and provide qualitative reference metrics with these results. Finally, using classification, it is proposed to train a model that can identify outstanding responses that help identify a particular problem with a course or a teacher.

The proposed solution aims to be a tool that allows for quick action by university administrators and to consider the open-ended responses of surveys, which is often information that can go unnoticed, particularly in large institutions like this one.

Keywords: teacher evaluation, artificial intelligence, natural language processing.

ÍNDICE

INTRODUCCIÓN	1
CAPÍTULO 1. DISEÑO DE LA PROPUESTA DE INNOVACIÓN	3
1.1 Información general de la propuesta de innovación	3
1.1.1 Institución promotora y responsable de la ejecución de la propuesta	3
1.1.2 Ámbito de intervención de la propuesta	4
1.1.3 Duración de la propuesta	4
1.1.4 Población objetiva directa e indirecta	4
1.2 Justificación y antecedentes de la propuesta de innovación educativa	4
1.3 Fundamentación teórica	6
1.3.1 Evaluación Docente	7
1.3.2 Métodos de evaluación docente	8
1.3.3 La Inteligencia Artificial y sus principales características	9
1.3.4 La evolución de la IA en la educación	11
1.3.5 Oportunidades y beneficios de la IA en la Educación	15
1.3.6 El futuro de la IA en la Educación	18
1.3.7 Casos de estudio de la IA en la Educación	20
1.4 Caracterización del contexto	22
1.5 Objetivos	22
1.6 Metas	23
1.7 Estrategia operativa y plan de trabajo	24
1.7.1 Recursos humanos	26
1.7.2 Monitoreo y evaluación	27
1.7.3 Sostenibilidad	27
1.7.4 Presupuesto	27
1.7.5 Cronograma	29
CAPÍTULO 2. INFORME DE LA EXPERIENCIA PILOTO	30
2.1 Objetivos	30
2.2 Metas	31
3.1 Fases de la propuesta	31
3.1.1 Primera fase: Planificación	32
3.1.2 Segunda fase: Diseño de la experiencia piloto	32
3.1.3 Tercera fase: Implementación del piloto	36
3.1.4 Cuarta fase: Toma de la encuesta a los estudiantes	37

3.1.5 Quinta fase: Entrega de resultados y evaluación	37
3.2 Recursos humanos	37
3.3 Monitoreo y evaluación de la experiencia piloto	38
3.4 Sostenibilidad	39
3.5 Riesgos y contingencias	39
3.6 Cronograma del piloto	40
3.7 La experiencia piloto	40
3.7.1 Instrumentos utilizados para la recolección de datos	41
3.7.2 Ejecución de la experiencia piloto	41
3.7.3 Resultados de la experiencia piloto	48
CONCLUSIONES	52
RECOMENDACIONES	53
REFERENCIAS BIBLIOGRÁFICAS	54
ANEXOS	59



ÍNDICE DE TABLAS

Tabla 1 Presupuesto de la propuesta de innovación	28
Tabla 2 Comparación de Principales Proveedores de AI Cloud.....	35
Tabla 3 Ejemplo de parámetros para el entrenamiento de la segmentación...	42
Tabla 4 Ejemplo de resultados obtenidos para dos de los textos	45
Tabla 5 Ejemplo de resultados extraídos para dos de los textos	46
Tabla 6 Resultados obtenidos luego del segundo entrenamiento.....	48



ÍNDICE DE FIGURAS

Figura 1	Etapas del proyecto	24
Figura 2	Cronograma del proyecto.....	29
Figura 3	Fases del proyecto.....	32
Figura 4	Magic Quadrant for Cloud AI Developer Services.....	34
Figura 5	Fases y áreas del proyecto	36
Figura 6	The Scrum Framework.....	39
Figura 7	Pantalla inicial de la encuesta a alumnos	41
Figura 8	Información cargada al script para ser procesada	44
Figura 9	Promedios de puntajes de sentimiento obtenido	49
Figura 10	Porcentajes correspondientes a cada clasificación entrenada.....	50
Figura 11	Distribución de puntaje de sentimientos de acuerdo a categorías ..	51



INTRODUCCIÓN

La inteligencia artificial (IA) es el campo de estudio que se ocupa del comportamiento inteligente, especialmente en lo que se refiere a las máquinas. La IA existe desde la década de 1960, pero los desarrollos recientes en el poder de procesamiento de la computadora y el análisis de datos han llevado a un rápido progreso en el campo.

La IA, en sus inicios, estuvo orientada a desarrollar la capacidad de predicción y toma de decisiones de los gobiernos y milicias, a partir del uso de una computadora. Actualmente, su uso se ha extendido a una amplia gama de tareas como el reconocimiento y traducción de voz, procesamiento de lenguaje natural, análisis de sentimientos de un texto, reconocimiento de imágenes, diagnóstico de afecciones médicas, entre muchos otros. Todos estos sub campos de la inteligencia artificial son aplicables a contextos más cotidianos de la sociedad, como la Medicina, el Marketing, Banca y Finanzas, etc.

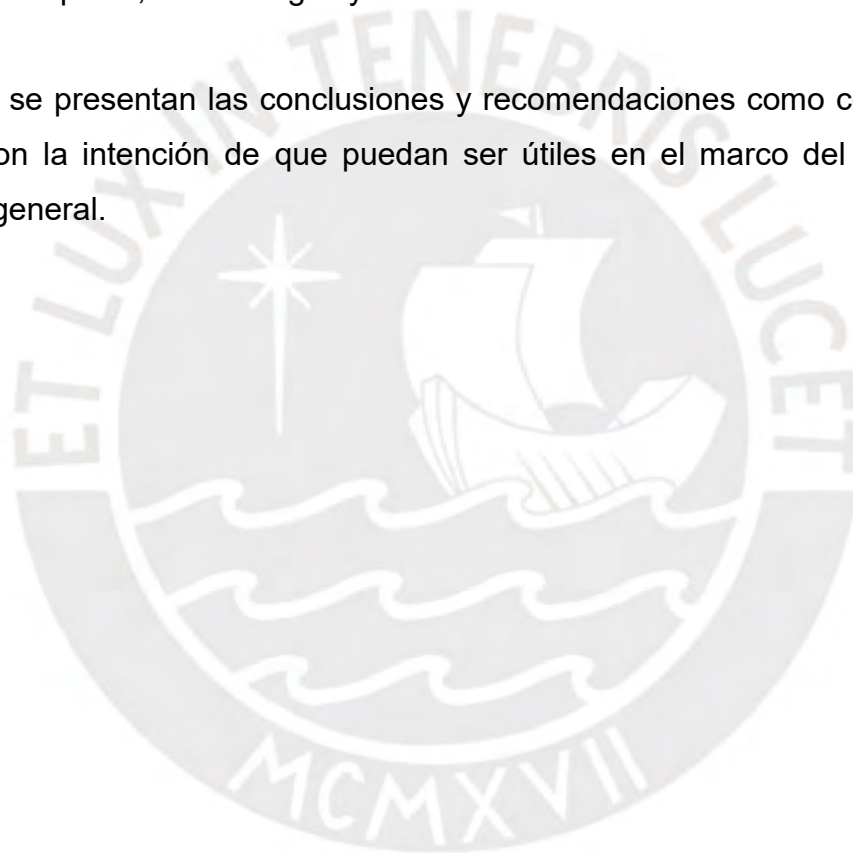
Sin embargo, en el sector Educación aún no se aprovecha al máximo esta tecnología, la cual podría mejorar los procesos, identificar oportunidades de mejora de los actores o analizar con mayor velocidad la información masiva. De manera específica, el procesamiento de lenguaje natural se ha desarrollado mucho en los últimos años y podría potenciar, por ejemplo, aspectos clave de la evaluación docente. En ese sentido, existe poca información sobre reportes o publicaciones de Instituciones Educativas latinoamericanas que hayan aprovechado esta tecnología para la mejora de sus procesos de evaluación.

En este contexto, en el año 2015 se desarrolló un primer piloto en el Perú, el cual realizaba el análisis de sentimiento (subcampo de la dimensión del procesamiento de lenguaje natural) en cada una de las respuestas de las encuestas a alumnos sobre el desempeño o valoración de sus docentes. Sobre este diseño y resultados, el presente proyecto busca potenciar este análisis incluyendo el entrenamiento de un motor adaptado a las necesidades particulares de la institución, la segmentación de respuestas a partir de variables prioritarias, la identificación de las temáticas más recurrentes en las respuestas (por ejemplo, el diseño de la clase, retrasos en el inicio

de la sesión, falta de sesiones prácticas, etc.) a fin de poder brindar retroalimentación rápida y efectiva a los docentes evaluados e ir generando una base de información histórica para la institución.

Por lo tanto, el presente documento se divide en dos capítulos, en el primero se diseña la propuesta y se presenta la fundamentación teórica, así como la justificación, el contexto, los objetivos y las metas de la misma. El segundo capítulo detalla la experiencia piloto que se planteó alineada a la propuesta de innovación, mencionando sus peculiaridades y detallando sus objetivos, metas y riesgos, así como el desarrollo de las fases del piloto, los hallazgos y resultados encontrados en esta experiencia.

Finalmente, se presentan las conclusiones y recomendaciones como consecuencia del piloto con la intención de que puedan ser útiles en el marco del proyecto de innovación general.



CAPÍTULO 1. DISEÑO DE LA PROPUESTA DE INNOVACIÓN

1.1 Información general de la propuesta de innovación

La presente propuesta de innovación busca mejorar las capacidades de evaluación docente en una universidad privada peruana mediante la implementación de un sistema que haga uso de técnicas avanzadas de análisis de datos. Específicamente, se propone utilizar el procesamiento de lenguaje natural (PLN), una rama de la inteligencia artificial que se enfoca en la interacción entre las computadoras y el lenguaje humano en su forma natural y cotidiana. Los detalles de la institución y del enfoque metodológico se describen en las siguientes secciones.

1.1.1 Institución promotora y responsable de la ejecución de la propuesta

La institución universitaria tiene cinco sedes en Lima, en los distritos de Ate, Centro de Lima, Los Olivos, San Juan de Lurigancho y Villa El Salvador. A nivel nacional, cuenta con sucursales en Arequipa, Chiclayo, Huancayo, Ica, Nuevo Chimbote, Piura y Trujillo. Además, esta universidad privada cuenta con 63 programas de estudio entre pregrado y posgrado (UTP, 2021).

Actualmente tiene cerca de 72,000 estudiantes, de los cuales, más de la mitad están inscritos en programas del campo de la ingeniería y el 29% en programas de Administración y Negocios de acuerdo con la Sunedu (2019).

La plana docente está compuesta por 3,666 docentes, siendo el área de desarrollo y comunicación docente, la que tiene dentro de sus labores la evaluación de los profesores de la institución. Cada ciclo se realiza una encuesta a los estudiantes, en la que estos evalúan a cada docente en sus respectivos cursos, considerando la cantidad de alumnos y los cursos en los que estos se encuentran matriculado, esta práctica da como resultado miles de respuestas a evaluar en cada ciclo lectivo.

1.1.2 Ámbito de intervención de la propuesta

El ámbito de intervención es institucional, ya que se aplicará en todas las sedes y a todos los alumnos de la institución obteniendo resultados de cada uno de los docentes durante los ciclos de la propuesta.

1.1.3 Duración de la propuesta

La propuesta se plantea con una duración inicial de dos años durante los cuales se evaluarán por lo menos 4 ciclos, a la vez que se van complementando funcionalidades en el sistema haciendo un mejor uso de la inteligencia artificial (IA) y realizando mejoras a nivel de interfaz y de la forma de mostrar los resultados. La meta final es que lo desarrollado pueda permanecer como una herramienta base sobre la cual se podrá realizar mejoras conforme las nuevas tecnologías lo permitan.

1.1.4 Población objetiva directa e indirecta

Población objetiva directa: Aproximadamente 3,666 docentes.

Población objetiva indirecta: Aproximadamente 72,000 estudiantes de pregrado.

1.2 Justificación y antecedentes de la propuesta de innovación educativa

Srinivasa-Desikan (2018) menciona que, si bien los esfuerzos para dotar de inteligencia a las máquinas se remontan a 1966 en el MIT, los avances en el área de la inteligencia artificial y en el análisis de textos, ha crecido mucho en los últimos años. Actualmente encontramos formas de inteligencia artificial en nuestro día a día, desde las sugerencias de qué video ver o qué música escuchar, asistentes virtuales, filtros de correos no deseados, entre muchos ejemplos de servicios y dispositivos cotidianos.

Los entornos educativos no son ajenos a estos cambios y la aplicación de la inteligencia artificial y su uso se vuelve cada vez más relevante. Andriessen y Sandberg (1999) mencionan un cambio fundamental de énfasis en la educación y su relación con la tecnología, en particular con la inteligencia artificial. Sin embargo, es importante mencionar que aún no existe una lista consensuada sobre los subcampos de la inteligencia artificial, sino que han ido identificándose uno por uno mientras se ha ido perfeccionando o mejorando el entrenamiento de las computadoras, como el

reconocimiento facial, el procesamiento del lenguaje natural, la inferencia a partir de datos, el reconocimiento de patrones, entre otros.

Por su parte, Khaled (2014) menciona que el procesamiento del lenguaje natural, un subcampo identificado de la inteligencia artificial, es un enfoque eficaz para mejorar el entorno educativo. Es decir, implementar estas técnicas implica iniciar el proceso de aprendizaje, a través de la adquisición de información en los sistemas educativos. El procesamiento del lenguaje natural está ampliamente integrado con una gran cantidad de contextos educativos, como la investigación, la ciencia, la lingüística, el aprendizaje electrónico, el sistema de evaluaciones y contribuye a obtener resultados positivos en entornos educativos como escuelas, institutos y universidades.

Por ejemplo, Litman (2016) menciona que especialmente en casos en los que se tiene mucha diferencia entre la cantidad de evaluadores versus la cantidad de evaluados se suelen buscar herramientas adicionales como la evaluación de pares, evaluaciones automáticas, ponderaciones de evaluadores, etc. para optimizar procesos de evaluación, este es el caso de los cursos masivos en línea (MOOC por sus siglas en inglés) y también es el caso de las encuestas masivas, en las que el ratio entre revisores y encuestados es particularmente bajo. En este último entorno es el que se desarrolla el presente trabajo, que busca optimizar el proceso de interpretación de los resultados de encuestas en una institución educativa superior peruana.

En estas encuestas, la gran mayoría de preguntas son cerradas, en particular de opción múltiple. Sin embargo, también se tiene por lo menos una pregunta de respuesta abierta, lo cual hace que sea particularmente complicado evaluar estos resultados, uno a uno, o poder encontrar respuestas particularmente relevantes, que denoten un aspecto positivo o negativo de algún docente o curso.

Por lo tanto, dada la gran cantidad de datos, poder revisar cada respuesta abierta se convierte en un proceso largo, laborioso y propenso al error humano. De esta manera, como se muestra en el Anexo 1, el problema principal de esta universidad privada peruana en la que se realizará la propuesta de innovación consiste en el insuficiente análisis de las encuestas de evaluación docente, la cual se sustenta en diversos argumentos interrelacionados.

En primer lugar, la institución se enfrenta a un volumen considerable de datos provenientes de las respuestas abiertas de las encuestas, lo que hace que el análisis manual sea una tarea abrumadora y poco práctica. Esta situación se agrava por las limitaciones de tiempo, ya que las encuestas se realizan al final de cada ciclo académico, dejando poco margen para analizar los resultados antes de tomar las decisiones para el siguiente periodo. Además, los recursos humanos disponibles para esta tarea son limitados, resultando en un bajo ratio entre revisores y encuestados.

El proceso actual de revisión manual no solo es lento e ineficiente, sino que también está sujeto a errores humanos, inconsistencias y posibles sesgos por parte de los revisores. Esto puede llevar a interpretaciones inconsistentes entre diferentes evaluadores, dificultando la comparación de resultados. Como consecuencia, existe un alto riesgo de que comentarios relevantes o *insights* importantes pasen desapercibidos, resultando en una pérdida de información valiosa que podría ser crucial para la mejora de los cursos y el desempeño docente.

La falta de herramientas adecuadas para el análisis de estas respuestas abiertas también dificulta la identificación de patrones o tendencias a lo largo del tiempo o entre diferentes cursos y docentes. Esto, sumado a la lentitud del proceso, puede resultar en que las sugerencias o problemas identificados no se aborden a tiempo para el siguiente ciclo académico, retrasando la implementación de mejoras necesarias.

En última instancia, a pesar de recopilar una gran cantidad de retroalimentación cualitativa valiosa, la institución no está aprovechando plenamente esta información para la toma de decisiones y la mejora continua de su oferta educativa. Todo esto supone una carga administrativa significativa que podría optimizarse con herramientas más eficientes, justificando así la necesidad de explorar soluciones como la presentada en este trabajo, basada en inteligencia artificial y procesamiento de lenguaje natural para abordar este desafío.

1.3 Fundamentación teórica

La retroalimentación por parte de los alumnos es una fuente muy rica para poder optimizar el proceso de enseñanza-aprendizaje, establecer puntos de mejora

inmediata e incluso puede ser útil para identificar líneas de capacitación docente. Sin embargo, establecer procesos que promuevan la recopilación y centralización de esta retroalimentación y su posterior procesamiento puede tener un gran impacto en la evaluación docente. En esta parte nos centraremos en la fundamentación teórica del procesamiento de la retroalimentación docente haciendo uso de la inteligencia artificial.

1.3.1 Evaluación Docente

La evaluación docente tiene el potencial de ayudar a mejorar la calidad de enseñanza de cada profesor. Sin embargo, Stiggins y Duke (1988) comentan que, en la práctica, si bien hay una evaluación, esta usualmente no ayuda a los profesores pues se presentan dos problemas principales. Por un lado, la evaluación suele ser superficial y, por otro lado, en muchos casos, los resultados no se suelen analizar o no generan ninguna consecuencia. En ese sentido, Skedsmo y Huber (2018) resaltan tres puntos primordiales para la evaluación docente:

- La necesidad de que la evaluación no interfiera con la autonomía del profesor ni con la cultura de la institución.
- Es importante contar con un esquema que incluya la observación sistémica del profesor en el aula, a lo largo del tiempo y en múltiples escenarios, sin embargo, esto conlleva un alto costo asociado.
- Los profesores deben estar involucrados en el diseño de los sistemas de evaluación de forma que se pueda priorizar y obtener información relevante para poder dar una retroalimentación al docente.

De esta manera buscamos tener la mayor cantidad de retroalimentación posible. Esta información permite al docente tomar conciencia de cómo es percibido, reforzar sus aspectos positivos y trabajar en aquellos que requieren mejora. Igualmente, Gutiérrez (2010) posiciona al docente como un agente de cambio, que además de transmitir el conocimiento, aporta al desarrollo del alumno, en ese sentido se resalta aún más la importancia de la mejora continua del docente.

Los resultados de estas evaluaciones no solo benefician a los docentes, sino que también constituyen una valiosa fuente de información para la institución educativa. Estos datos pueden utilizarse para identificar y priorizar a los docentes con mejores

desempeños, así como para analizar las áreas de oportunidad y desarrollar estrategias que eleven la calidad de la enseñanza en general. De esta manera, la evaluación docente se convierte en una herramienta fundamental para el desarrollo profesional de los educadores y la mejora continua de la institución, contribuyendo así a la excelencia académica y al éxito de los estudiantes.

1.3.2 Métodos de evaluación docente

Los sistemas de evaluación docente usualmente sirven para dos propósitos. Por un lado, un propósito sumativo, este provee información para el manejo del personal, asignación de horas, aumentos, entre otros. El segundo propósito es promover el desarrollo profesional de los profesores, para esto debe ser posible identificar las fortalezas y debilidades de cada profesor.

Tuytens et al. (2020), así como Skedsmo y Huber (2018) también mencionan la importancia de definir y medir la evaluación docente. Esto se relaciona con la necesidad de establecer estándares claros de desempeño docente, desarrollar instrumentos válidos y confiables de evaluación y utilizar múltiples medidas para una evaluación integral. Todo esto mientras se considera el contexto del profesor.

Por otro lado, Shinkfield y Stufflebeam (2012) listan 15 metodologías para la evaluación docente, las cuales incluyen la evaluación de pares, la observación de clases, la evaluación por parte de los alumnos, la autoevaluación, entre otras. Por su parte, Gutiérrez (2010) hace énfasis en medir la satisfacción de los alumnos y propone un modelo de evaluación, en el cual, el 30% del puntaje es dedicado a este parámetro, mencionando que:

La opinión de los alumnos respecto a su proceso de enseñanza y aprendizaje es de suma importancia para la valoración del desempeño docente, tomando en cuenta que un buen proceso de aprendizaje está directamente vinculado a la capacidad del profesor de generar un proceso de enseñanza diversificado, que considere las características de los estudiantes, sus conocimientos previos y sus estilos de aprendizaje. Es por ello, que se aplica la encuesta de Opinión del Estudiante, que valora las características del proceso de enseñanza y aprendizaje. (p. 8)

Peterson (2000) también menciona el rol central de los estudiantes en la evaluación, dado que estos tienen una perspectiva única y distinta a la que puede ser percibida por el personal administrativo u otros profesores. Sin embargo, en muchos casos y sobre todo en instituciones de Educación Básica Regular (EBR) no se suele tomar en cuenta la evaluación de los estudiantes. Se menciona además a los costos como uno de los principales problemas y este autor hace más énfasis en encuestas con opciones múltiples. Sin embargo, las nuevas tecnologías hacen posible que los costos asociados sean más bajos y además abren la posibilidad al análisis de otro tipo de preguntas abiertas.

El tema de los costos también está presente en la observación de clases, idealmente estas se deben dar en varias ocasiones por un externo, sin embargo, como esto no siempre es factible en algunos escenarios se recurre a la autoevaluación, aunque Rollett et al. (2021) mencionan que la información obtenida de la autoevaluación puede ser no tan válida, pues los profesores ven la enseñanza desde su propia perspectiva.

Como alternativa se menciona, una vez más, el uso de la perspectiva de los alumnos para juzgar la calidad de la enseñanza, ya que una de las consecuencias adicionales de esto es que los alumnos observan la enseñanza en cada clase y además se cuentan con múltiples perspectivas de la misma lección.

Finalmente, si bien se pueden tener reservas sobre la validez de percepción del estudiante en relación a la calidad de la enseñanza, múltiples estudios muestran que la información obtenida puede ser válida para retroalimentar al profesor (Rollett et al., 2021).

Muchas veces la implementación de estas metodologías de evaluación docente pueden tener un costo asociado que resulte prohibitivo para la institución, es por esto que, en las siguientes secciones, se evaluará el uso de la Inteligencia Artificial para la evaluación como una alternativa de costo eficiente.

1.3.3 La Inteligencia Artificial y sus principales características

El campo de la inteligencia artificial estudia y diseña agentes racionales, es decir programas de computadora entrenados para la toma de decisiones. Russell y Norvig (2009) mencionan que estos agentes perciben impulsos del entorno y toman

decisiones. Por ello, la racionalidad de estas decisiones va a depender de experiencias pasadas y el conocimiento del agente. El objetivo es que estos agentes puedan tomar las mejores decisiones de acuerdo al entrenamiento recibido.

En los últimos años, el estudio y aplicación de la IA en distintos campos ha pasado por un gran desarrollo, y el campo de la educación no es ajeno a estos nuevos avances. En ese sentido, Roll y Wylie (2016) identificaron que, el énfasis del uso de la inteligencia artificial en educación, está relacionado al apoyo en la resolución de problemas, como por ejemplo en los campos de la ciencia, tecnología, ingeniería y matemáticas (STEM por sus siglas en inglés). Sin embargo, se plantea un cambio en la aplicación de la inteligencia artificial, buscando un enfoque más amplio que impacte en el servicio de formación, como la evaluación docente. Para ello, uno de los subcampos que podría brindar un mayor beneficio al sector educación, es el procesamiento de lenguaje natural.

a) Procesamiento de lenguaje natural

Se define al procesamiento de lenguaje natural como el uso de programas o máquinas para manipular el lenguaje. Esta es una tarea particularmente complicada y Kasliwal (2018) resalta que nosotros como humanos somos la única especie conocida que ha desarrollado lenguajes escritos; siendo tan difícil esta tarea que los niños no aprenden a leer o escribir por sí solos, por lo que entrenar a una máquina a interpretar el lenguaje ha sido un gran logro para la inteligencia artificial.

En este documento nos centraremos en el lenguaje escrito, ya que es la forma más común de obtener retroalimentación de estudiantes. Asimismo, Srinivasa-Desikan (2018) resalta que el texto escrito es el medio al que más estamos expuestos en el día a día, desde el material de estudio, los correos electrónicos, etc.

La omnipresencia del Internet hace que las cantidades de información en texto crezcan día a día, como ejemplo Google maneja más de 40 000 búsquedas cada segundo y Twitter más de 1 600 millones de mensajes al día, por lo que esto forma una gran cantidad de información. Además, varias empresas utilizan estos datos para entender

a sus clientes e identificar tendencias. De manera equivalente, las instituciones educativas pueden utilizar la información que tienen a su disposición.

En este contexto, el procesamiento de lenguaje natural tiene como objetivo que la máquina interprete cantidades grandes de texto, las relacione y logre identificar la información más relevante, tal como lo haría un humano, pero con una gran velocidad. Esta información, además, puede ser clasificada por temáticas o por intensidad de sentimiento, dependiendo del uso que se desee dar a la información.

b) Clasificación de textos y análisis de sentimientos

Kasliwal (2018) define la clasificación de textos como la tarea en la que, partiendo de un texto, se decide a qué clase pertenece y esta decisión tiene asociada a un grado de certidumbre. En el procesamiento de lenguaje natural, se tienen aplicaciones de clasificación como la detección de idioma, detección de *spam* y el análisis de sentimientos.

Particularmente, el análisis de sentimientos quiere evaluar si un texto, comentario o reseña es positivo o negativo, además de considerar qué tan negativa o qué tan positiva es la carga del texto. Además, un caso más complejo de la clasificación es considerar parámetros propios por medio de clasificaciones propias, en este caso se debe entrenar un modelo *ad hoc* que permita evaluar estas clasificaciones.

Por lo tanto, valorar o indicar si el texto tiene un sentimiento positivo o negativo y un nivel definido de intensidad es una tarea muy compleja. Así, este subcampo utiliza la inteligencia artificial para analizar los textos, identificar oraciones y buscar palabras que tengan un significado positivo o negativo alineado a un "diccionario" previamente entrenado del tema, de esta manera brinda pesos a las palabras y las valora.

1.3.4 La evolución de la IA en la educación

La integración de la IA en la educación no es un fenómeno repentino, sino el resultado de décadas de avance tecnológico e investigación pedagógica. Los primeros sistemas de instrucción asistida por computadora de los años 60 y 70 sentaron las bases para las sofisticadas herramientas educativas impulsadas por IA de hoy. El rápido progreso

en el aprendizaje automático, el procesamiento del lenguaje natural y el análisis de la *big data* ha acelerado el desarrollo y la adopción de la IA en entornos educativos durante los últimos años.

Por ello, las tecnologías de la IA se están utilizando en múltiples facetas de la educación, entre las que se puede destacar:

a. Aprendizaje Personalizado

Los sistemas de IA pueden analizar datos individuales de los estudiantes para proporcionar experiencias de aprendizaje personalizadas y adaptadas a las necesidades y estilo de aprendizaje de cada estudiante (Ng et al., 2023). Las plataformas de aprendizaje adaptativo utilizan algoritmos de IA para ajustar la dificultad y presentación del contenido basándose en el rendimiento del estudiante y asegurando que los estudiantes sean constantemente desafiados sin llegar a ser agobiados.

Varios estudios han demostrado la efectividad de los sistemas de aprendizaje personalizado impulsados por la IA. Por ejemplo, Thai et al. (2021) realizaron un ensayo de *My Math Academy*, un programa adaptativo de aprendizaje digital de matemáticas para estudiantes de preescolar. Los resultados mostraron que los estudiantes que usaron *My Math Academy* durante 12 a 14 semanas superaron significativamente a los estudiantes del grupo control en las evaluaciones de matemáticas. El programa fue capaz de adaptar el contenido al nivel de cada estudiante y proporcionar rutas de aprendizaje personalizadas.

Además de los resultados académicos, el aprendizaje personalizado utilizando IA también puede aumentar el interés y compromiso de los estudiantes. Pratama et al. (2023) encuestaron a estudiantes sobre sus percepciones de la IA en la educación y encontraron que el 88% estaba muy de acuerdo en que la IA era importante para ayudar a su aprendizaje. En ese sentido, los estudiantes apreciaron cómo las herramientas de IA podían proporcionar apoyo y retroalimentación personalizados adaptados a sus necesidades individuales.

De todas maneras, si bien la IA muestra un gran potencial para personalizar la educación, los investigadores advierten que debe usarse cuidadosamente para

umentar en lugar de reemplazar a los profesores. Como señalan Bang et al. (2023), los aspectos emocionales, sociales y morales de la educación son difíciles de replicar solo con tecnología. La conexión humana y la orientación proporcionada por los profesores sigue siendo crucial para el desarrollo integral de los estudiantes. A medida que la tecnología continúa avanzando, tiene aún más potencial de hacer que la educación sea más personalizada, atractiva y efectiva para todos los estudiantes.

b. Sistemas de Tutoría Inteligente

Los programas de tutoría utilizando IA pueden proporcionar a los estudiantes orientación individualizada, respondiendo preguntas y ofreciendo explicaciones de manera conversacional (Chen et al., 2023). Estos sistemas pretenden simular tutores humanos, ofreciendo apoyo personalizado a gran escala. Pueden identificar temas que requieren refuerzo en cada estudiante y con esto, proporcionar una práctica dirigida y ofrecer retroalimentación inmediata, mejorando así el proceso de aprendizaje.

c. Calificación y Retroalimentación automatizadas

El procesamiento del lenguaje natural y el aprendizaje automático permiten a la IA evaluar el trabajo de los estudiantes, proporcionando retroalimentación rápida sobre escritura, resolución de problemas y otras evaluaciones (Naismith et al., 2023). Esto puede ahorrar tiempo a los educadores en tareas rutinarias de calificación, permitiéndoles centrarse en aspectos más complejos de la enseñanza. Los sistemas impulsados por IA pueden evaluar ensayos, resolución de problemas matemáticos e incluso tareas de programación, ofreciendo retroalimentación detallada para ayudar a los estudiantes a mejorar.

d. Tareas administrativas

La IA asiste en diversas funciones administrativas como programación de clases, gestión de matrículas y asignación de recursos educativos, liberando tiempo para que los educadores se concentren en la enseñanza (Wu et al., 2024). En este contexto, los chatbots impulsados por IA pueden manejar consultas rutinarias de los estudiantes, agilizando la comunicación entre instituciones y aprendices.

e. Creación de Contenido

Las herramientas de IA generativa ayudan a los educadores a desarrollar materiales curriculares, planes de lecciones y actividades de aprendizaje de manera más eficiente (Chen & Wu, 2024). Esto incluye la generación de preguntas de práctica, resúmenes y contenido multimedia. La IA también puede asistir en la creación de materiales educativos multilingües, haciendo el contenido más accesible para poblaciones estudiantiles diversas.

f. Análisis de datos y Analítica predictiva

La IA analiza grandes conjuntos de datos sobre el rendimiento de los estudiantes para identificar patrones y predecir resultados, permitiendo intervenciones tempranas (Wu et al., 2024). Estos conocimientos pueden ayudar a los educadores a identificar estudiantes en riesgo y adaptar, en consecuencia, estrategias de apoyo. La analítica predictiva también puede informar la toma de decisiones institucional, desde la oferta de cursos hasta la asignación de recursos.

g. Herramientas de Accesibilidad

La IA impulsa herramientas que mejoran la accesibilidad para estudiantes con discapacidades, como subtítulo en tiempo real, texto a voz y tecnologías de asistencia visual (Wu et al., 2024). Estas herramientas pueden mejorar significativamente la experiencia de aprendizaje para estudiantes con diversas necesidades, promoviendo la inclusividad en la educación.

h. Realidad virtual y realidad aumentada

La IA mejora las experiencias educativas de realidad virtual y realidad aumentada, creando entornos de aprendizaje inmersivos. Estas tecnologías pueden simular conceptos científicos complejos, eventos históricos o incluso excursiones virtuales, haciendo las ideas abstractas más tangibles para los estudiantes.

1.3.5 Oportunidades y beneficios de la IA en la Educación

La integración de la IA en la educación ofrece varios beneficios potenciales, como los que observamos a continuación:

a. Mejora de los Resultados de Aprendizaje

El aprendizaje personalizado habilitado por la IA puede llevar a un mejor rendimiento académico y desarrollo de habilidades (Bond et al., 2024). Al adaptar la instrucción a las necesidades individuales, la IA puede ayudar a abordar las brechas de aprendizaje y acelerar el progreso.

b. Mayor motivación

Los entornos de aprendizaje interactivos impulsados por la IA pueden aumentar la motivación y participación de los estudiantes (Firat, 2023). Los elementos de gamificación y los desafíos adaptativos mantienen a los estudiantes comprometidos e involucrados en su proceso de aprendizaje.

c. Ahorro de tiempo para los educadores

Al automatizar tareas rutinarias, la IA permite a los profesores dedicar más tiempo a actividades de mayor valor como una instrucción individualizada y la tutoría (Perezchica-Vega et al., 2024). Esto puede llevar a interacciones más significativas entre profesor y estudiante y a experiencias educativas mejoradas.

d. Toma de decisiones basada en datos

La analítica de IA proporciona a educadores y administradores conocimientos para tomar decisiones más informadas sobre el currículo, los métodos de enseñanza, la asignación de recursos y la evaluación (Wu et al., 2024). Este enfoque basado en datos puede llevar a estrategias educativas más efectivas y a un mejor rendimiento institucional.

e. Acceso expandido

Las plataformas de aprendizaje en línea habilitadas por IA pueden aumentar el acceso a la educación para poblaciones remotas o desatendidas (Bond et al., 2024). Esto es particularmente crucial para abordar las brechas educativas globales y proporcionar nuevas oportunidades de aprendizaje.

f. Preparación para una futura labor impulsada por la IA

Exponer a los estudiantes a tecnologías de IA en entornos educativos ayuda a prepararlos para futuras carreras donde la IA jugará un papel significativo (Perezchica-Vega et al., 2024). Esta exposición puede fomentar la alfabetización en IA y las habilidades de pensamiento crítico esenciales para el futuro mercado laboral.

g. Mejora continua

Los sistemas de IA pueden aprender y mejorar continuamente basándose en las interacciones y retroalimentación de los usuarios, llevando a herramientas y contenido educativo en constante evolución y refinamiento.

Ahora bien, a pesar de su potencial, la adopción de la IA en la educación también enfrenta varios desafíos como los que mencionamos a continuación:

a. Preocupaciones éticas

Se necesitan abordar cuestiones en torno a la privacidad de datos, el sesgo de los algoritmos y los límites apropiados de la IA en la toma de decisiones educativas (Chan, 2023; Wu et al., 2024). Hay preocupaciones sobre la recopilación y uso de datos de los estudiantes, sobretodo de menores de edad, así como el potencial de que los sistemas de IA agraven los sesgos existentes.

b. Integridad académica

Existe preocupación en relación a los estudiantes que utilizan herramientas de IA, como por ejemplo, modelos extensos de lenguaje o LLM (siglas en inglés para *Large Language Model*), entre ellos, el ChatGPT con la finalidad de completar tareas de manera deshonestas, potencialmente tergiversando los objetivos de aprendizaje (Perezchica-Vega et al., 2024). Esto plantea cuestiones sobre cómo se debería evaluar el conocimiento y las habilidades de los estudiantes en un mundo rodeado de muchas herramientas que hacen uso de la IA.

c. Formación y adopción del profesorado

Muchos educadores pueden carecer de las habilidades y conocimientos necesarios para integrar eficazmente la IA en sus prácticas de enseñanza (García-Cuevas et al.,

2023; Ruediger et al., 2024). Por consiguiente, esta brecha de habilidades puede obstaculizar la implementación efectiva de las tecnologías de IA en entornos educativos.

d. Equidad y acceso

Las disparidades en el acceso a las tecnologías de IA podrían exacerbar las desigualdades educativas existentes (Wu et al., 2024). Hay preocupaciones de que los beneficios de la IA en la educación puedan no distribuirse equitativamente, potencialmente ampliando la brecha entre instituciones educativas bien dotadas y con pocos recursos.

e. Calidad y fiabilidad de los sistemas de IA

Asegurar la precisión, efectividad y adecuación del contenido y las recomendaciones generadas por IA es un desafío continuo (Wu et al., 2024). Hay preocupaciones sobre la naturaleza de la "caja negra" de algunos algoritmos de IA, es decir que no se sepa bien, el proceso interno que manejan los algoritmos. Esto dificulta entender y validar sus procesos de toma de decisiones.

f. Dependencia excesiva de la tecnología

Existen preocupaciones sobre los posibles impactos negativos en las habilidades de pensamiento crítico si los estudiantes se vuelven excesivamente dependientes de las herramientas de IA (Perezchica-Vega et al., 2024). Es crucial encontrar un equilibrio entre aprovechar las capacidades de la IA y fomentar el pensamiento independiente.

g. Diversidad cultural y lingüística

Asegurar que las herramientas educativas de IA sean culturalmente sensibles y lingüísticamente diversas sigue siendo un desafío. Muchos sistemas de IA se desarrollan principalmente en inglés y pueden no servir adecuadamente a poblaciones que no hablan inglés o a contextos culturales diversos.

1.3.6 El futuro de la IA en la Educación

A medida que la IA continúa evolucionando e integrándose en la educación, varias áreas clave requieren atención:

a. Desarrollo de políticas

Las instituciones educativas y los responsables políticos necesitan desarrollar directrices integrales para el uso ético y efectivo de la IA en la educación (UNESCO, 2021; Chan, 2023). Estas políticas deben abordar cuestiones de privacidad de datos, transparencia algorítmica y uso ético de la IA.

b. Desarrollo profesional de los docentes

Se necesitan programas de formación robustos para generar en los educadores habilidades que permitan aprovechar las herramientas de IA y, en consecuencia, adaptar sus prácticas de enseñanza (Ruiz-Mendoza et al., 2024; Ayuso-del Puerto & Gutiérrez-Esteban, 2022). Esto incluye no solo habilidades técnicas, sino también enfoques pedagógicos para integrar la IA de manera efectiva.

c. Investigación y evaluación

La investigación permanente es crucial para evaluar los impactos a largo plazo de la IA en los resultados de aprendizaje, las prácticas de enseñanza y la equidad educativa (Bond et al., 2024). Se necesitan estudios longitudinales y evaluaciones rigurosas para comprender los verdaderos efectos de la IA en la educación.

d. Desarrollo colaborativo

Educadores, tecnólogos y responsables políticos deben trabajar juntos para diseñar herramientas educativas de IA que se alineen con las mejores prácticas pedagógicas y aborden las necesidades reales del aula (Wu et al., 2024). Este enfoque colaborativo puede asegurar que las herramientas de IA sean tanto técnicamente sofisticadas como pedagógicamente sólidas.

e. Enfoque en la Alfabetización en IA

Los planes de estudio deben actualizarse para incluir la alfabetización en IA, asegurando que los estudiantes comprendan las capacidades, limitaciones e

implicaciones éticas de las tecnologías de IA (Ng et al., 2021). Esto preparará a los estudiantes para ser usuarios y creadores informados de tecnologías de IA en el futuro.

f. Marco ético de la IA

Implementar marcos como el propuesto por Floridi et al. (2018), enfatizando principios de beneficencia, no maleficencia, autonomía, justicia y explicabilidad en las aplicaciones educativas de IA. Estas directrices éticas pueden ayudar a navegar el complejo panorama de la IA en la educación.

g. Abordar la brecha digital

Se deben hacer esfuerzos para asegurar un acceso equitativo a las herramientas y recursos educativos impulsados por la IA. Esto puede implicar inversiones en infraestructura, programas de acceso a dispositivos e iniciativas para llevar la educación mejorada por IA a comunidades desatendidas.

h. Estrategias de evaluación adaptativa

Desarrollar nuevas estrategias que puedan evaluar eficazmente el aprendizaje de los estudiantes en un entorno educativo mejorado por IA. Esto puede incluir más evaluaciones basadas en proyectos, evaluaciones de desempeño en tiempo real y métodos de prueba adaptativos.

i. Desarrollo de la IA transcultural

Fomentar el desarrollo de herramientas educativas de IA que sean culturalmente receptivas y lingüísticamente diversas. Esto incluye involucrar a diversos interesados en el proceso de diseño y desarrollo y crear sistemas de IA capaces de operar eficazmente en diferentes idiomas y contextos culturales.

j. Enfoque interdisciplinario

Promover un enfoque interdisciplinario de la IA en la educación, reuniendo a expertos de campos como la educación, la informática, la psicología y la ética. Este esfuerzo colaborativo puede conducir a soluciones educativas de IA más holísticas y efectivas.

1.3.7 Casos de estudio de la IA en la Educación

Como último punto de esta fundamentación teórica, se presentan tres casos que permitirán identificar claramente como la IA se ha utilizado en la educación.

a. MathIA de Carnegie Mellon:

La universidad ha desarrollado un sistema de tutoría híbrida humano-IA para mejorar el aprendizaje de matemáticas en escuelas intermedias de bajos ingresos. El sistema combina software de matemáticas adaptativo con apoyo de tutores humanos remotos, proporcionando una experiencia de aprendizaje personalizada que se adapta a las necesidades de cada estudiante. Los tutores utilizan paneles de control en tiempo real para identificar a los estudiantes que requieren asistencia, permitiendo una intervención oportuna (Thomas et al., 2024).

Un estudio cuasi-experimental realizado en tres escuelas urbanas demostró que este enfoque híbrido aumenta significativamente el tiempo que los estudiantes dedican a practicar matemáticas y mejora su progreso en las lecciones. Notablemente, los estudiantes con menor rendimiento previo se beneficiaron más de esta intervención (Thomas et al., 2024).

b. Chatbot de apoyo estudiantil de la Universidad Estatal de Georgia:

El chatbot de IA "Pounce" implementado por la Universidad Estatal de Georgia (GSU) es un excelente ejemplo de cómo la inteligencia artificial conversacional (IAC) puede mejorar los procesos de gestión del conocimiento y la ventaja competitiva en instituciones de educación superior (Wu et al., 2024).

Inicialmente, Pounce se desarrolló para ayudar a los nuevos estudiantes con información sobre registro, finanzas y vida en el campus. Posteriormente, se integró con el contenido de los cursos para proporcionar recordatorios personalizados sobre tareas y exámenes, permitir a los estudiantes hacer preguntas a sus instructores y ofrecer cuestionarios breves antes de los exámenes (Mdavis, 2022).

En ese sentido, Pounce ha demostrado mejoras tangibles en varios aspectos clave en la creación de conocimiento. Por ejemplo, al analizar las preguntas y conversaciones de los estudiantes, Pounce puede identificar áreas comunes de dificultad y transformar este conocimiento tácito en conocimiento explícito útil para mejorar el plan de estudios (Perezchica-Vega et al., 2024). Además, Pounce proporciona respuestas rápidas y en tiempo real a las preguntas de los estudiantes, facilitando un intercambio eficiente de conocimientos las 24 horas del día y los 7 días de la semana. También dirige a los estudiantes a foros en línea y grupos de estudio para fomentar el intercambio de conocimientos entre pares (Perezchica-Vega et al., 2024). Finalmente, basándose en los perfiles individuales y conversaciones pasadas, Pounce ofrece preguntas personalizadas, consejos y recordatorios para ayudar a los estudiantes a aplicar el conocimiento adquirido de manera práctica (Perezchica-Vega et al., 2024).

Como resultado de estas mejoras en los procesos de gestión del conocimiento, la GSU ha observado un aumento significativo en el rendimiento de los estudiantes. Además, Pounce ha liberado tiempo y recursos de los educadores al aprender de las preguntas y experiencias de los estudiantes, ampliando su base de conocimientos para responder preguntas frecuentes (Mdavis, 2022).

Por lo tanto, la implementación de Pounce demuestra cómo las herramientas de IAC pueden mejorar los procesos de gestión del conocimiento en instituciones de educación superior, aumentando la eficiencia operativa, los resultados educativos y el rendimiento institucional general. Esto, a su vez, contribuye a mejorar la ventaja competitiva de la institución en el panorama de la educación superior (Tsoi & Strønen, 2024).

c. Aprendizaje adaptativo de idiomas de Duolingo:

La popular aplicación de aprendizaje de idiomas utiliza IA para personalizar lecciones, ajustar la dificultad y proporcionar práctica dirigida. Su efectividad en la adquisición de idiomas la ha convertido en una herramienta ampliamente utilizada tanto en entornos educativos formales como informales. El sistema de IA de Duolingo, llamado Birdbrain, utiliza algoritmos sofisticados basados en décadas de investigación en psicología educativa y avances recientes en aprendizaje automático para mejorar continuamente la experiencia del alumno. Birdbrain predice la probabilidad de que un alumno

responda correctamente a un ejercicio, estimando tanto la dificultad del ejercicio como la competencia actual del alumno, y utiliza estas predicciones para seleccionar dinámicamente nuevos ejercicios para la siguiente lección (Bicknell et al., 2023).

1.4 Caracterización del contexto

La organización donde se llevará a cabo la presente propuesta de innovación es una institución educativa superior privada, con sedes en varios distritos de Lima y en varios departamentos del país. En total se tienen 3 666 docentes y más de 72 000 estudiantes. La evaluación docente está a cargo de la Dirección de Calidad Educativa central en Lima y se realiza cada ciclo, en la antepenúltima semana de clases, de manera censal y virtual en todas las sedes del país.

Los resultados de esta evaluación son considerados para la renovación de carga académica del siguiente ciclo y la tasa de respuesta es alta (mayor al 70%), ya que es un requisito para la previsualización de las notas de exámenes.

La evaluación es una encuesta virtual que consta principalmente de preguntas cerradas y una pregunta abierta por curso. Es sobre esta última pregunta en la que se ejecutará el análisis de sentimientos y uso de inteligencia artificial. Los resultados de ambos tipos de preguntas y el análisis forman parte del informe que recibirá el director de carrera a la que pertenece cada docente.

1.5 Objetivos

El proyecto de innovación propuesto tiene los siguientes objetivos:

Objetivo general

- Optimizar la interpretación de los resultados de encuestas de evaluación docente a partir de la inteligencia artificial en una universidad privada peruana.

Objetivos específicos

- Implementar un sistema de fácil uso para automatizar el análisis de las preguntas abiertas de las encuestas, utilizando un procesamiento de lenguaje natural.

- Exportar reportes con métricas cualitativas referenciales a partir del análisis de sentimiento de las respuestas de la encuesta.
- Identificar respuestas resaltantes que puedan implicar un problema singular en un curso o con un profesor particular.

1.6 Metas

Metas de ocupación

- Un project manager.
- Un especialista en experiencia de usuario y diseño de interfaces.
- Un desarrollador *fullstack*¹.
- Un programador con experiencia en AI.
- Un especialista en visualización de datos.
- Una persona encargada de realizar las pruebas a los sistemas.
- Un comunicador.

Metas de atención

- 100% de los profesores de todas las sedes de la institución.
- 100% de los alumnos de todas las sedes de la institución.

Metas de capacitación

- Capacitación al área de desarrollo y comunicación docente en el uso del sistema y la interpretación de resultados.

Metas de implementación

- Una plataforma web que permita realizar el análisis cualitativo automatizado de las encuestas.

Metas de producción

- Reportes de los resultados de cada ciclo.

¹ Hace referencia a un desarrollador que trabaja tanto en la parte de interfaz del usuario como en los datos y el tratamiento de los mismos.

- Campaña de sensibilización al alumnado sobre la importancia del llenado de las encuestas.

1.7 Estrategia operativa y plan de trabajo

Estrategia operativa

Para el desarrollo de la propuesta se plantean 4 etapas mencionadas en la siguiente figura.

Figura 1

Etapas del proyecto



En las siguientes subsecciones se describe a detalle cada una de estas etapas y su relación con la experiencia piloto.

- **Planificación del proyecto de innovación:** La primera fase tuvo por objetivo planificar la experiencia y para ello se tomó en cuenta un proyecto anterior desarrollado por el mismo tesista en la institución que sentó las bases para la presente propuesta. En el anterior piloto se realizó el análisis de sentimientos de las respuestas abiertas de las encuestas, por parte de los estudiantes, lo cual ayudó en la rapidez del análisis cualitativo y en poder aterrizar a números, las respuestas que de otra forma podrían no ser consideradas. Finalmente, se alinearon las fechas a los ciclos de la universidad, teniendo especial énfasis en la fecha predefinida de la toma de encuestas a los alumnos.

- **Diseño de la propuesta de innovación:** Considerando lo analizado en la sección anterior se diseñó la propuesta, la cual tomó en cuenta una mejora adicional que apuntaba a reforzar el análisis de esas respuestas abiertas. Concretamente, dado que el uso de estrategias de procesamiento de lenguaje natural para el análisis de sentimientos fue bien recibido por la institución, se propuso la iniciativa de agregar la funcionalidad de clasificación de textos, de forma que se puedan identificar los principales comentarios de los estudiantes y así tomar acciones al respecto. Además, se consideraron los pasos necesarios para que la propuesta pueda ser sostenible en el tiempo.
- **Diseño e implementación del piloto:** Teniendo en cuenta la propuesta anterior se presentó el piloto donde se validaron las nuevas funcionalidades y se sentaron las bases para el completo desarrollo de la propuesta. La experiencia piloto se planteó de manera que el inicio del análisis coincidiera con la fecha de obtención de resultados de la encuesta realizada a los estudiantes, con la finalidad de que se pudieran analizar los textos de forma automatizada y lo más rápido posible para facilitar la interpretación de los resultados.
Ahora bien, para el desarrollo del presente piloto, el principal instrumento de recolección de datos fue la encuesta en el portal de alumnos, la misma que tenía preguntas cerradas con escala Likert y una pregunta abierta pidiendo comentarios sobre la clase llevada. De esta manera, el enfoque de este piloto fue el análisis de esta última pregunta y la evaluación de los resultados.
- **Presentación de resultados del piloto:** Los resultados obtenidos fueron presentados a las distintas áreas involucradas para tener una retroalimentación y así poder realizar las mejoras de ser necesario.
- **Sistematización de la experiencia:** Una vez que se tengan los modelos de clasificación entrenados, estos pueden ser utilizados directamente sin necesidad de volver a entrenarlos. Así también, los modelos pueden ser reentrenados para pulir los resultados y ser probados con resultados de encuestas de distintos ciclos. Al tener los modelos finales se propondrá el diseño de un sistema web que permita al área de

desarrollo y comunicación docente aplicar tanto, el análisis de sentimientos, como los modelos de clasificación sin requerir intervención externa. Luego de tener el sistema implementado se capacitará al área en su utilización.

- **Sensibilización:** En paralelo con la etapa anterior se propone mostrar al alumnado cómo es que se utilizan los resultados de las encuestas, ya que al ser una encuesta obligatoria se ha encontrado algo de rechazo por su llenado. Se propone por ello, una campaña que muestre cómo se procesan los resultados y las consecuencias directas de los comentarios que se dejan, de forma que los alumnos están más motivados en llenar los comentarios y en brindar la mayor cantidad de información posible en las futuras encuestas.

1.7.1 Recursos humanos

Los recursos humanos necesarios para el desarrollo de la propuesta se mencionaron previamente en las metas de ocupación. Sin embargo, a continuación, se detalla con mayor detalle de su rol en la presente propuesta:

- Un *project manager* para liderar el desarrollo del proyecto, coordinando los recursos y avances.
- Un especialista en experiencia de usuario y diseño de interfaces.
- Un desarrollador *fullstack*, encargado de implementar la plataforma web y las bases de datos asociadas.
- Un programador con experiencia en inteligencia artificial, para entrenar los modelos y realizar la integración para el despliegue de los mismos.
- Un especialista en visualización de datos, para diseñar la información más relevante a ser mostrada.
- Una persona encargada de realizar las pruebas a los sistemas, para asegurar el correcto funcionamiento del sistema y validar los casos de uso del mismo.
- Comunicador para realizar una campaña de sensibilización de cómo se utilizan los resultados de las encuestas.

1.7.2 Monitoreo y evaluación

El principal punto de control se dará al finalizar el piloto y mostrar sus resultados, la viabilidad del proyecto depende de que estos puedan ser entregados con relativa rapidez y validar que los resultados sean congruentes con las respuestas.

Una vez se tengan los resultados satisfactorios se procederá con la etapa de sistematización, que tiene una duración de un año. Durante este tiempo, se ejecutarán de manera manual los mismos procesos del piloto, y se podrá realizar modificaciones al modelo entrenado.

Al finalizar la sistematización y el entrenamiento se pasará a monitorear el correcto funcionamiento del sistema por parte del área usuaria.

1.7.3 Sostenibilidad

Para que el proyecto sea sostenible se proponen las siguientes estrategias:

- Es necesario contar con los resultados de la encuesta, cada ciclo y en formato virtual. Por ello, es imprescindible asegurar que las preguntas se realicen en línea o se procesen digitalmente. Además, se necesita tener calendarizada las fechas de la toma y procesamiento de encuestas.
- El desarrollo se realizará utilizando metodologías ágiles, las cuales tienen como una de sus características, que se debe involucrar a las áreas usuarias en el desarrollo, esto permitirá tener un resultado muy alineado a las necesidades y expectativas de los usuarios.
- Se deben tener espacios de retroalimentación de los resultados al finalizar el ciclo, luego de haber realizado el análisis de las encuestas.
- Se debe contar con la etapa de sensibilización que busca involucrar a los alumnos e incentivarlos a una más provechosa participación en las encuestas.

1.7.4 Presupuesto

La siguiente tabla muestra los presupuestos involucrados para el desarrollo de todo el proyecto de innovación:

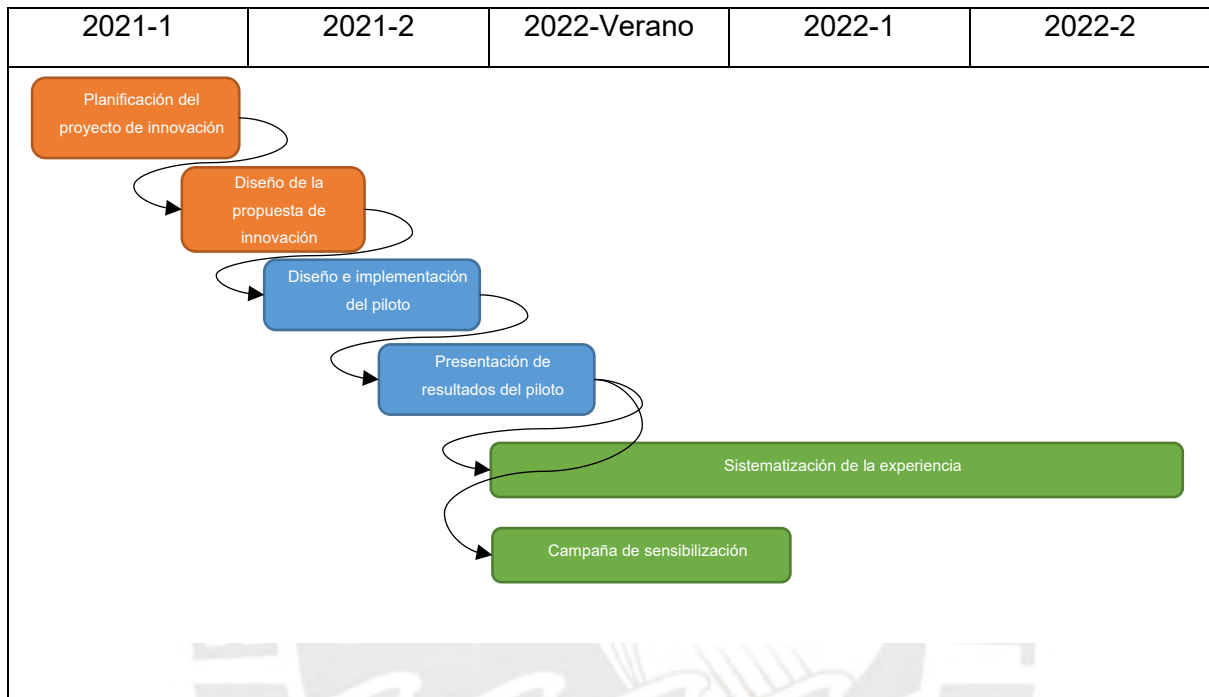
Tabla 1*Presupuesto de la propuesta de innovación*

Etapas y actividades	Concepto	Costo aproximado
Planificación del proyecto de innovación	-	-
Diseño de la propuesta de innovación	-	-
Diseño e Implementación del piloto	Programador de AI y Project Manager	S/. 10 000
	Gastos en servidores y servicio cloud	S/. 500
Presentación de resultados del piloto	-	-
Sistematización de la experiencia.	Diseño de interfaces	S/. 8 000
	Implementación de sistema web y base de datos	S/. 25 000
	Pruebas de calidad	S/. 3 000
	Despliegue del servicio y costos de mantenimiento por la duración del proyecto	S/. 5 000
Sensibilización	Campaña de sensibilización	S/. 3 000
Total		S/. 54 500

1.7.5 Cronograma

La siguiente figura muestra un diagrama de Gantt con la proyección de tiempos para la propuesta de innovación.

Figura 2
Cronograma del proyecto



CAPÍTULO 2. INFORME DE LA EXPERIENCIA PILOTO

El presente capítulo se centra primero en el diseño realizado de la experiencia piloto, previo a la puesta en marcha de la misma. Esta experiencia se planteó con una duración de dos meses, teniendo como hito la toma de la encuesta para su posterior procesamiento y en una segunda parte se describe la ejecución de la experiencia y los resultados.

Esta experiencia es fundamental para el desarrollo del proyecto pues demostró su viabilidad y produjo los primeros resultados de lo que se podría esperar al tener el proyecto finalizado. Durante esta primera etapa se considera el análisis de las encuestas de un ciclo, para esto se implementó y utilizaron programas que hacen uso de modelos de inteligencia artificial entrenados para identificar patrones claves que permitan resaltar comentarios que tengan temas recurrentes en la pregunta abierta de la encuesta a los alumnos.

2.1 Objetivos

Tomando en cuenta los objetivos del proyecto general se plantearon los siguientes objetivos para el piloto:

Objetivo general

- Validar la viabilidad del uso de procesamiento de lenguaje natural para optimizar la interpretación de los resultados de encuestas de evaluación docente.

Objetivos específicos

- Automatizar el análisis de las preguntas abiertas de las encuestas utilizando procesamiento de lenguaje natural.
- Proveer métricas cualitativas referenciales a partir del análisis de sentimiento de las respuestas de la encuesta.
- Analizar las evidencias sobre la efectividad de la propuesta bajo el modelo considerado.

2.2 Metas

Metas de ocupación

- Un programador con experiencia en AI.

Metas de atención

- 100 % de los alumnos de las sedes de Lima.
- 100 % de los profesores de las sedes de Lima.

Dado que la encuesta se viene tomando ciclo a ciclo a todos los alumnos de la universidad no existe impedimento en considerar la totalidad de alumnos y profesores desde el piloto planteado.

Metas de capacitación

- Durante este piloto no se realizará ninguna capacitación.

La encuesta se realizará de la misma manera como se viene realizando y todo el procesamiento y análisis estará a cargo del equipo que presenta el proyecto, por esto no es necesaria ninguna capacitación adicional.

Metas de implementación

- Programas que se integran con los sistemas de procesamiento de lenguaje natural. Es decir, se implementará un código que interactúa con el servicio de terceros y que permitirá el análisis de los textos obtenidos por la encuesta.

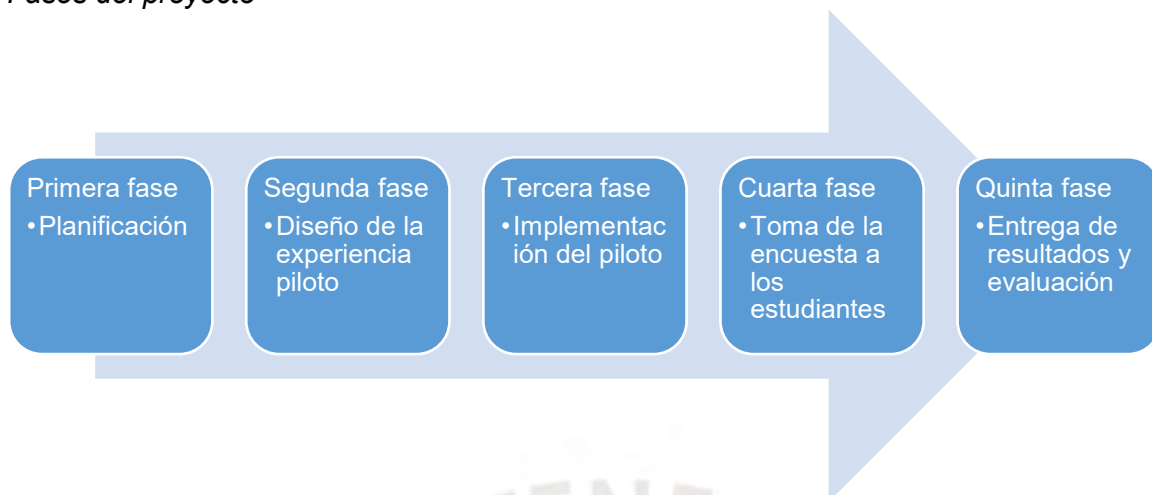
Metas de producción

- Reportes de los resultados del presente ciclo para las sedes analizadas. Se considera dentro del piloto un reporte general que presente los resultados obtenidos.

3.1 Fases de la propuesta

Para el desarrollo del piloto se definieron las fases presentadas en la siguiente figura:

Figura 3
Fases del proyecto



En las siguientes secciones se desarrollan a detalle cada una de las fases.

3.1.1 Primera fase: Planificación

Tomando en cuenta el proyecto en general, se planificó el piloto de manera que se pudiera cumplir con los objetivos del mismo, haciendo énfasis en probar las mejoras que puede traer el uso del procesamiento de lenguaje natural para el proceso de las encuestas.

3.1.2 Segunda fase: Diseño de la experiencia piloto

La segunda fase buscó concretar la mejor opción para realizar la implementación entre los distintos proveedores de soluciones *cloud* con funcionalidades de inteligencia artificial. El uso de estos proveedores hace posible, que no sea necesaria una inversión en infraestructura especializada, pues en particular el entrenamiento de modelos customizados requiere el uso de equipamiento como tarjetas de video especializadas en procesamiento en paralelo. Al utilizar estos proveedores optimizamos entonces los recursos y, particularmente en la etapa del piloto, se pueden realizar pruebas con relativa facilidad.

En la actualidad existen múltiples proveedores que ofrecen servicios de este tipo, por lo que la elección de uno sobre otro puede no ser tan sencilla. Por lo tanto, según Bresciani et al. (2008), el *Magic Quadrant de Gartner* es una gran herramienta gráfica para la evaluación y toma de decisiones de estos productos, ya que se presenta en

un gráfico, dos dimensiones² donde se ubican los proveedores de una solución determinada. De esta manera, Gartner utiliza criterios, tanto objetivos como subjetivos, para evaluar a cada proveedor. Los líderes en teoría cuentan tanto con la habilidad de proveer los servicios de mejor manera que la competencia y además una visión apropiada.

La figura 4, muestra el *Magic Quadrant de Gartner* (Baker et al., 2021), que permite evaluar las capacidades en procesamiento de lenguaje natural (NLP, por sus siglas en inglés) de los líderes en servicios de desarrollo en nube orientados a IA. De esta manera, considerando como referencia a los líderes, se tomó una decisión más informada sobre el proveedor a utilizar buscando que la propuesta sea lo más eficiente posible.

Si bien la solución anterior utilizaba Google Cloud, el panorama de servicios ofrecidos por los diferentes proveedores, así como la facilidad de uso y costo han variado desde la primera implementación.

² Las dimensiones del gráfico son: Habilidad para ejecutar (Ability to Execute) e Integridad de la visión (Completeness of Vision), con estas dos dimensiones se forma un cuadro con cuatro cuadrantes.

Figura 4
Magic Quadrant for Cloud AI Developer Services



Nota. Baker et al. (2021)

Tomaremos como punto de inicio de este análisis a los líderes del sector según Gartner, para pasar a explorar las funcionalidades relevantes para el proyecto.

En la Tabla 2 se muestra un resumen de las principales funcionalidades evaluadas en los proveedores.

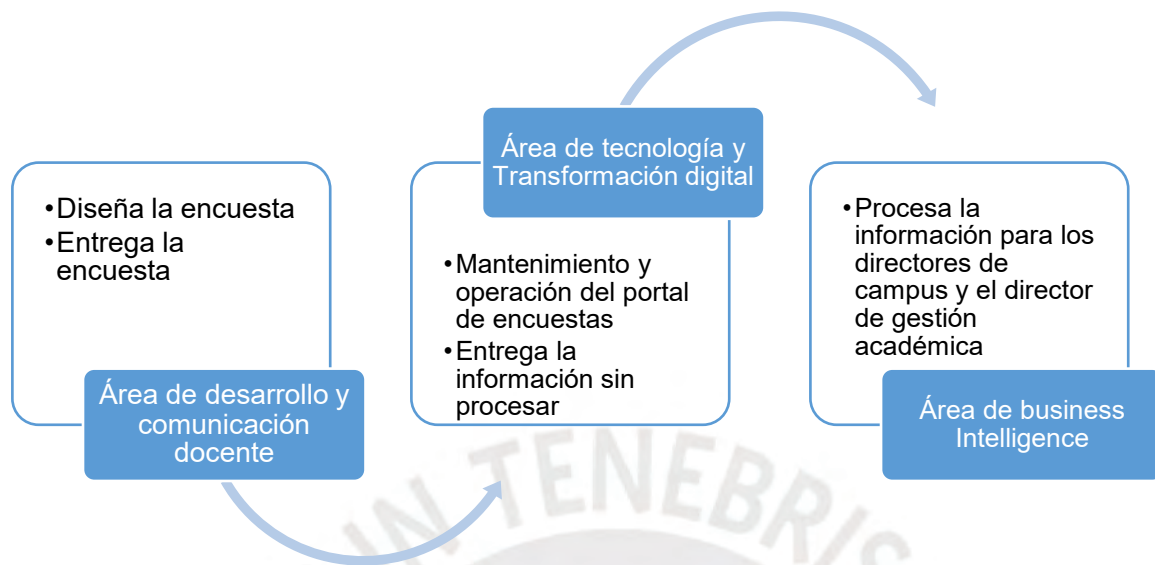
Tabla 2*Comparación de Principales Proveedores de AI Cloud*

Parámetro	Microsoft	Google	Amazon	IBM
Nombre de servicio	Azure Cognitive Services	Cloud Natural Language	Comprehend	Watson Natural Language Understanding
Análisis de sentimiento	Si	Si	Si	Si
Clasificación	Si	Si	Si	Si
Funcionalidades en idioma español	Si	Si	Si	Si

Como se muestra en la Tabla 2, actualmente cualquiera de los proveedores puede suplir las necesidades técnicas del proyecto, así que la elección de uno u otro puede reducirse al costo y/o facilidad de uso de los API. El análisis de costo de cada proveedor no forma parte del alcance del presente trabajo y las diferencias son mínimas, por lo que la elección del proveedor, en este caso, se consideró a partir del análisis de la facilidad para integrar la solución. En ese sentido, para este piloto, se eligió *Watson Natural Language Understanding*.

Por otro lado, se analizaron las necesidades, en cuanto a los recursos humanos. Para ello, la institución ya cuenta con el personal y el proceso definido para la toma normal de la encuesta. La figura 5 muestra el procedimiento actual en líneas generales, identificando las áreas involucradas.

Figura 5
Fases y áreas del proyecto



Dentro de las áreas mencionadas se tiene el personal clave para el correcto funcionamiento del proceso de encuesta a los alumnos y, por ende, para el funcionamiento del presente proyecto, el mismo que será detallado en la sección de recursos humanos.

3.1.3 Tercera fase: Implementación del piloto

Antes de iniciar con la distribución de la encuesta se implementaron los pasos necesarios para el análisis de sentimientos y el entrenamiento del motor para la clasificación de resultados. En el presente piloto se utilizó *Watson Natural Language Understanding* y para los desarrollos y entrenamiento necesario se utilizó *Watson Knowledge Studio*.

El entrenamiento inicial del modelo será genérico y deberá ser pulido con los resultados iniciales para adaptarlo mejor a las necesidades particulares de la encuesta y clasificaciones presentes en las respuestas. Es por esto que, luego de un primer análisis de los datos, se procederá a entrenar nuevamente el modelo y a realizar un nuevo análisis.

3.1.4 Cuarta fase: Toma de la encuesta a los estudiantes

En las últimas semanas de cada ciclo se activa la opción para tomar la encuesta en el portal de los estudiantes. Ellos, al entrar al LMS del curso deben responder unas preguntas sobre el profesor y el curso. También, al final de esta encuesta, hay una pregunta abierta para comentarios y es esta pregunta la más relevante para el desarrollo del presente proyecto, ya que se analiza la carga de sentimientos, se interpreta el texto, se reconocen las frases y oraciones presentes y se clasifica la respuesta en grupos predefinidos utilizando un motor de clasificación entrenado para este caso.

Toda la información llenada por los estudiantes se almacena en una base de datos en un servidor administrado por la Gerencia de tecnología y transformación digital, y ellos son también los encargados de entregar la información de estas respuestas.

3.1.5 Quinta fase: Entrega de resultados y evaluación

La información es entregada para su análisis en un archivo en formato CSV, para que pueda ser interpretada y analizada por el área de *Business intelligence*. De esta manera, en paralelo al análisis, por parte de esta área, se realizó el trabajo del piloto.

Finalmente, los resultados se entregaron al área de desarrollo y comunicación docente para su análisis y retroalimentación a los profesores y directores.

3.2 Recursos humanos

Para la adecuada realización del proyecto de innovación se requiere el trabajo conjunto de varios grupos de trabajo, presentados a continuación:

- Equipos del área de desarrollo y comunicación docente, encargados de la elaboración de la encuesta, de la interpretación y presentación de los resultados.
- Gerencia de tecnología y transformación digital, encargada de hacer que la encuesta esté disponible en el portal y que los alumnos tengan acceso a la

misma. Así mismo esta área entregará la base de datos con los resultados de la encuesta una vez acabada la aplicación de la misma.

- Desarrollador para el módulo de análisis de sentimientos, quien realizó la programación necesaria para la interpretación de los resultados y la entrega de los mismos al área de desarrollo y comunicación docente.

3.3 Monitoreo y evaluación de la experiencia piloto

Se realizará un seguimiento del porcentaje de cobertura de las encuestas en función del total de alumnos y la matrícula, esperando llegar al mayor porcentaje posible.

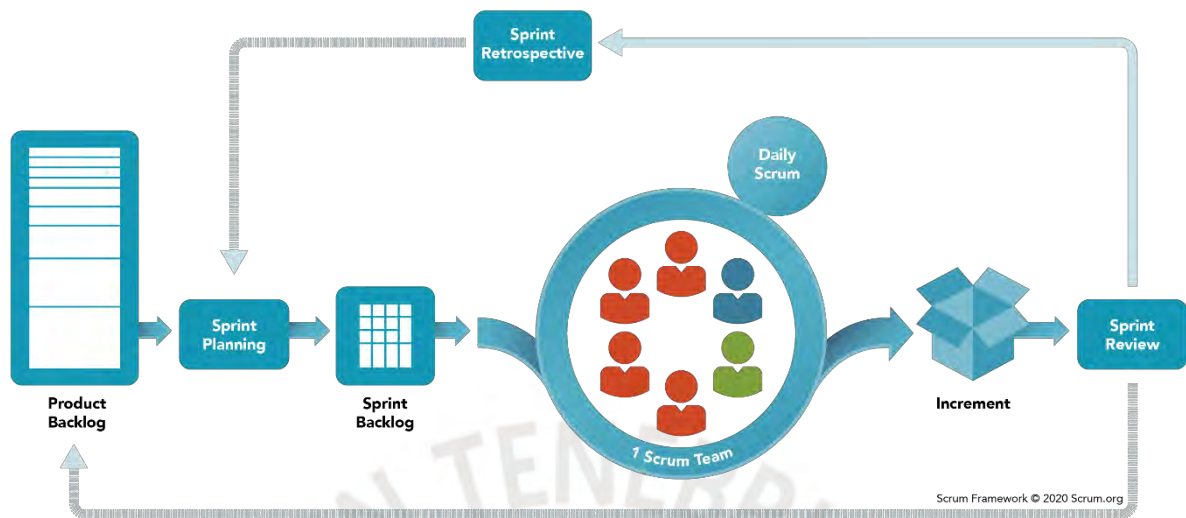
También, se espera procesar todas las respuestas válidas. Además, se evaluarán las respuestas que no se pueden procesar para implementar estrategias de preprocesado o filtros necesarios y así los resultados sean lo más confiables posible.

Para el entrenamiento del motor de clasificación se realizará por lo menos dos veces, de manera que se puedan utilizar los primeros resultados para poder reentrenar al motor con las respuestas obtenidas.

El desarrollo de la solución se realizará en constante comunicación con un representante del área de desarrollo y comunicación docente, el cual será el *product owner* dentro del esquema de desarrollo ágil adoptado para este desarrollo.

La figura 6 muestra un resumen de la metodología *Scrum*, y dentro de la misma se considera también el *feedback* de los involucrados, de manera que se puedan identificar posibles oportunidades para mejorar el producto final.

Figura 6
The Scrum Framework



Fuente: Scrum.org (s.f.).

3.4 Sostenibilidad

La continuidad de la presente propuesta pasa por contar con el presupuesto necesario con la finalidad de arrendar los servicios del procesamiento de lenguaje natural. La obtención de dicho presupuesto será consecuencia directa de los resultados obtenidos en el presente piloto. Se consideran además las estrategias mencionadas para el proyecto general, haciendo énfasis en que el uso de metodologías ágiles involucra al área usuaria permitiendo que los desarrollos se realicen en completa coordinación con ellos.

3.5 Riesgos y contingencias

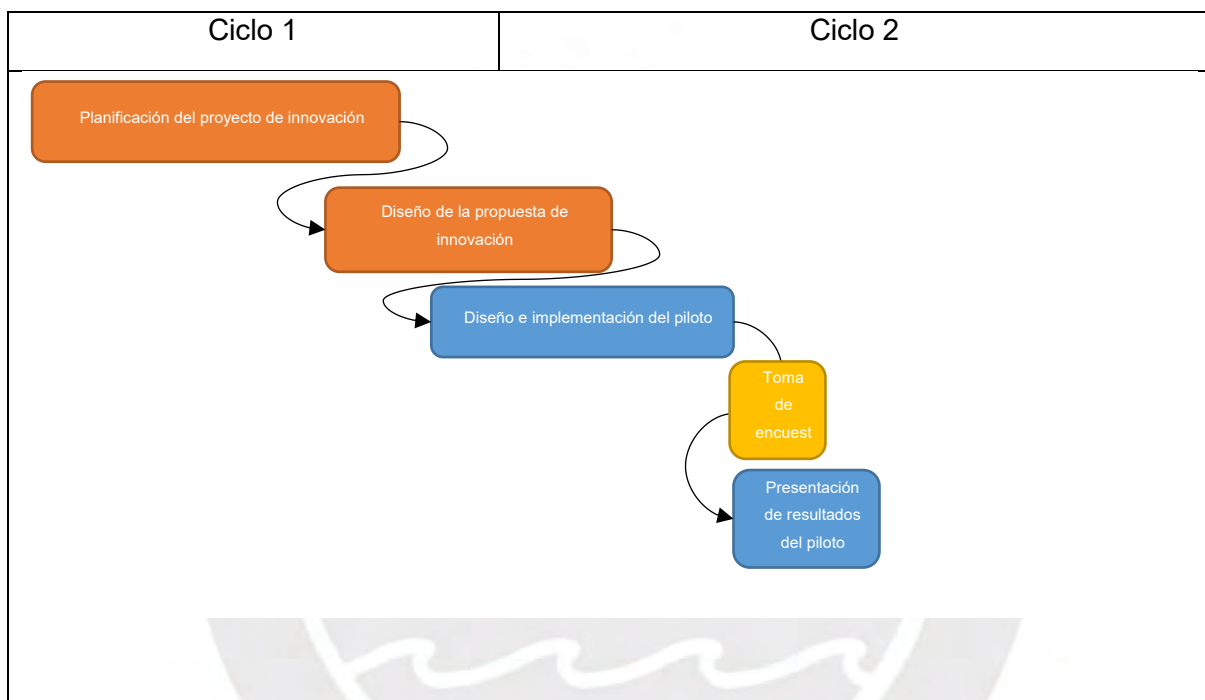
Dado que la materia prima para el análisis del piloto son los resultados de la encuesta, el principal riesgo es que los estudiantes no respondan masivamente la encuesta. Sin embargo, por la forma en la que es presentada y la obligatoriedad de la misma, el riesgo de que esto pase es muy bajo.

En caso la cantidad de encuestas respondidas fuera muy bajo, la principal complicación se dará para el entrenamiento del motor de clasificación y la contingencia será que se podrá entrenar al motor tomando como base una encuesta de ciclo anterior y así proceder al análisis de las encuestas respondidas.

3.6 Cronograma del piloto

El piloto toma como base lo elaborado durante el ciclo 1 para ser implementado durante el ciclo 2 y alinear la finalización de los programas necesarios con la toma de la encuesta a los alumnos.

Figura 7
Cronograma del piloto



3.7 La experiencia piloto

La presente propuesta busca implementar un mecanismo para la rápida interpretación de los resultados de las encuestas a alumnos, por lo que el piloto estuvo enfocado en los primeros pasos para contar con esta retroalimentación de manera rápida y eficiente.

La experiencia piloto se planteó de manera que el inicio de análisis coincidiera con la fecha de obtención de resultados de la encuesta. El análisis de los textos se realizó de forma automatizada y lo más rápido posible para facilitar la interpretación de los resultados. El proceso que demoró más fue el entrenamiento del motor de clasificación y todo el procedimiento debió tomar como máximo una semana.

3.7.1 Instrumentos utilizados para la recolección de datos

Para el desarrollo del presente piloto, el principal instrumento de recolección de datos fue la encuesta en el portal de alumnos, la misma que contaba con preguntas de opción múltiple y una pregunta abierta en la que se pedía comentarios sobre la clase llevada. En ese sentido, el enfoque de este piloto fue el análisis de esta última pregunta y la evaluación de los resultados.

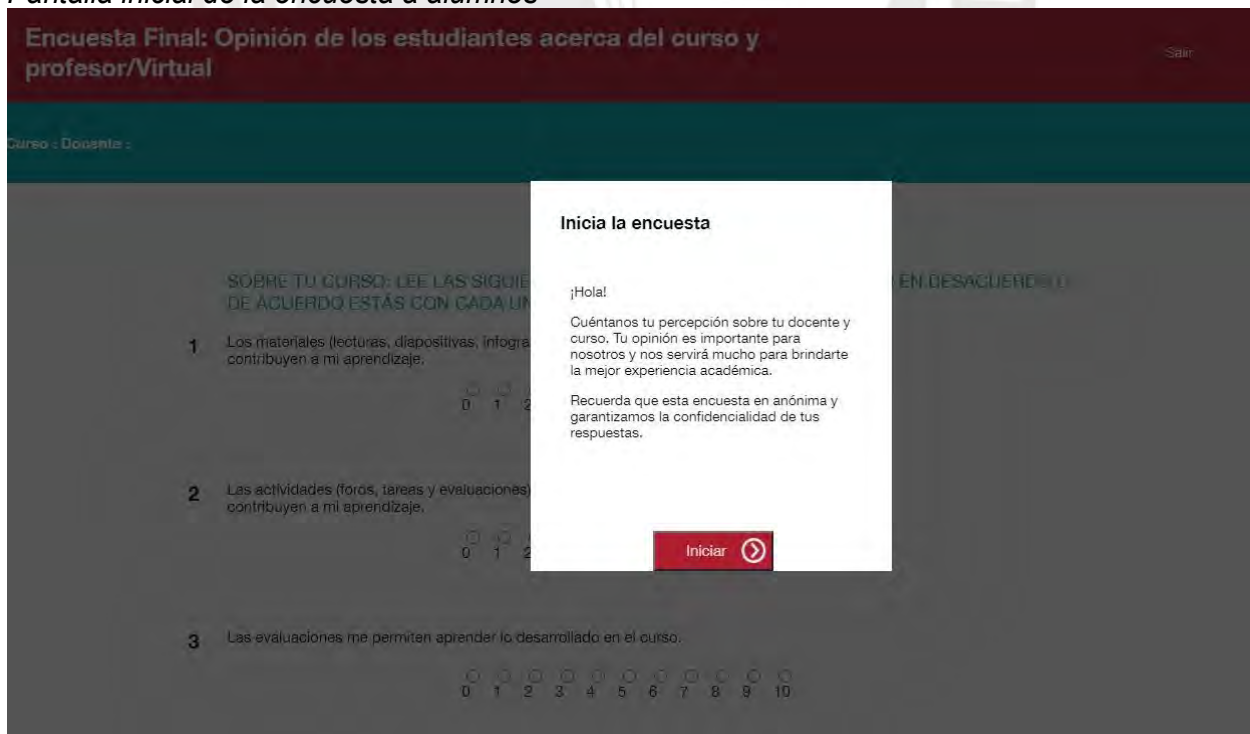
3.7.2 Ejecución de la experiencia piloto

Para la ejecución del piloto se tuvieron cuatro momentos importantes, descritos en las siguientes secciones.

Toma de encuesta en línea

En las últimas semanas del ciclo 2021-2 se realizó la encuesta en línea en el portal de los alumnos y en la siguiente figura se muestra la pantalla inicial de la encuesta.

Figura 8
Pantalla inicial de la encuesta a alumnos



Esta encuesta estaba segmentada por curso y por los horarios dentro de cada curso, obteniendo finalmente información del profesor que estaba dictando ese curso en particular.

La encuesta se coloca al inicio del portal del estudiante y su llenado es, por política de la institución, requisito para poder acceder al portal, lo cual garantiza tener un porcentaje elevado de participación.

Entrenamiento del modelo de segmentación

Para poder tener la segmentación de las respuestas, más allá de los términos genéricos, se debe entrenar previamente un modelo.

Este entrenamiento se realiza en base a criterios de segmentación predefinidos y dando ejemplos de oraciones que pertenecen a estas categorías. La siguiente tabla muestra un ejemplo de oraciones con las categorías asignadas.

Tabla 3

Ejemplo de parámetros para el entrenamiento de la segmentación

Oración de ejemplo	Categoría
Buena metodología	metodología
Buenos ejemplos	metodología
Muchas tareas	metodología
Incentiva participación	metodología
Clases didácticas	metodología
Clases virtuales	virtualidad

Oración de ejemplo	Categoría
Uso de zoom	virtualidad
Conexión a internet	virtualidad
Videoconferencias	virtualidad
Fallas de conexión	virtualidad

La tabla muestra un extracto de la información utilizada para entrenar el modelo, mientras más ejemplos se utilice para entrenar mejor será la segmentación. El Anexo 3 muestra el *script* programado para realizar el entrenamiento utilizando un modelo de “Watson Natural Language Understanding”.

Con el modelo entrenado se procedió a realizar una primera prueba, de manera que revisando los resultados se pueda agregar ejemplos o categorías de ser necesario y re-entrenar al modelo.

Análisis de sentimientos y segmentación

Los resultados de la encuesta son recibidos como un archivo de Microsoft Excel y este se convierte en un archivo csv (archivo de valores separados por comas, por sus siglas en inglés). Este archivo no cuenta con ningún formato y fue manipulado utilizando *dataframes*³ de la librería pandas.

La siguiente figura muestra un extracto de la tabla importada, lista para ser procesada.

³ Los *dataframes* son una estructura interna de la librería, para el almacenamiento y manipulación de datos de dos dimensiones.

Figura 9
 Información cargada al script para ser procesada

```

df_data_1 = pd.read_csv("encuesta.csv")
df_data_1 = df_data_1.drop(columns='Apellidos_Nombres_Docentes')
df_data_1.head()

```

te	Seccion	Cod_Curso	Curso	Matriculados	Encuestados	Fecha_Registro	...	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18
69	19347	100000G29T	FORMACION PARA LA EMPLEABILIDAD	17	13	26/11/2021	...	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	Bien
69	19347	100000G29T	FORMACION PARA LA EMPLEABILIDAD	17	13	24/11/2021	...	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	buen docente
69	19347	100000G29T	FORMACION PARA LA EMPLEABILIDAD	17	13	16/11/2021	...	9.0	9.0	9.0	9.0	9.0	9.0	9.0	9.0	8.0	Es buen docente, sus clases son interesantes y...
25	19330	100000G78T	INVESTIGACION OPERATIVA	18	7	04/12/2021	...	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	La docente complementa con material lo publica...

Como se ha comentado previamente, el énfasis de la interpretación se dió en la pregunta abierta (columna P18 en la figura anterior). Consecuentemente, para nuestro análisis, únicamente mantendremos las columnas: Sección, Cod_Curso y P18. Ahora bien, para mayor detalle, el Anexo 4 contiene el programa desarrollado para realizar el análisis de sentimiento y segmentación de las respuestas.

Además, por cada respuesta se tiene un resultado en formato JSON (notación de objeto de JavaScript, por sus siglas en inglés). Este formato es un estándar que hace posible que se tenga múltiple información estructurada y, en nuestro caso, sirve para que con una sola respuesta se tenga información tanto del análisis de sentimientos como de la segmentación.

La tabla siguiente muestra un ejemplo del resultado obtenido para unos de los textos analizados.

Tabla 4

Ejemplo de resultados obtenidos para dos de los textos

Texto analizado	Resultado
<p>No me gusta mucho su método de enseñanza. Sin embargo, considero que la virtualidad es el principal problema en el correcto aprendizaje de este curso.</p>	<pre> "usage":{ "text_units":1, "text_characters":148, "features":2 }, "sentiment":{ "document":{ "score":-0.782193, "label":"negative" } }, "language":"es", "classifications":[{ "confidence":0.276313, "class_name":"material" }, { "confidence":0.25808, "class_name":"metodología" }, { "confidence":0.22537, "class_name":"virtualidad" }] </pre>
<p>Muy bueno el profesor, deja casos aplicativos y nos de trabajos para complementar la enseñanza.</p>	<pre> "usage":{ "text_units":1, "text_characters":94, "features":2 }, "sentiment":{ "document":{ "score":0.841601, "label":"positive" } }, "language":"es", "classifications":[{ "confidence":0.459433, "class_name":"metodología" }, { "confidence":0.181103, "class_name":"material" }, { "confidence":0.160692, "class_name":"virtualidad" }] </pre>

Si bien los resultados obtenidos en la tabla anterior tienen toda la información recuperada, esta no se encuentra en un formato apropiado para su interpretación. Por

ello, en la siguiente sección estos resultados son procesados para poder ser presentados de manera apropiada.

Una consideración a tomar en cuenta es que la calidad de los resultados depende de la calidad de los textos analizados. Srinivasa-Desikan (2018) menciona la importancia de tener información correctamente formateada, ya que en el caso de las encuestas se encuentran textos sin sentido, emoticones y jergas que se filtran o reemplazan para que los resultados no se vean afectados. En particular se notó que el motor de análisis de sentimientos no reconocía “ok” como una palabra, por lo que esta fue reemplazada en las respuestas por “bien” de manera que se mantenga la carga positiva asociada coloquialmente a esa expresión.

Preparación de los resultados

El siguiente proceso fue preparar los resultados para que puedan ser presentados de una manera más amigable, ya sea para el usuario final o para un programa.

Siguiendo el mismo ejemplo anterior se tienen para las dos oraciones, los resultados siguientes:

Tabla 5

Ejemplo de resultados extraídos para dos de los textos

Texto analizado	Análisis de sentimiento	Clasificación
No me gusta mucho su método de enseñanza. Sin embargo, considero que la virtualidad es el principal problema en el correcto aprendizaje de este curso.	-0.782193	
Muy bueno el profesor, deja casos aplicativos y nos de trabajos para complementar la enseñanza.	0.841601	metodología (0.459433)

El análisis de sentimientos nos da un número de -1 a 1, mientras más negativo sea el número se tiene un sentimiento más negativo detectado, 0 es un sentimiento neutro y mientras más cercano a 1 sea el número, más positivo será el contenido.

En los ejemplos de la tabla se observa el correcto funcionamiento del análisis de sentimientos de las oraciones mostradas.

Para los valores de clasificación, el modelo devuelve un resultado en un rango de 0 a 1 y cuanto más cercano a 1 sea este valor, mayor es la confianza del modelo en la clasificación otorgada. Para mitigar el efecto del ruido⁴ y mejorar la calidad de los resultados, se ha establecido un umbral de confianza: se han descartado todas las clasificaciones con un valor inferior a 0.4. Esta medida ayuda a filtrar los datos potencialmente poco fiables, permitiendo centrarse en las clasificaciones en las que el modelo muestra una mayor seguridad.

Para este ejemplo se muestra que el primer texto tuvo una clasificación de virtualidad de 0.22 por lo que ha sido descartado. Sin embargo, claramente el texto pertenece a esta categoría, por lo que esto se puede tomar como un ejemplo de que el modelo requiere una mayor cantidad de ejemplos para una correcta clasificación. Se muestra, además, que esta clasificación no depende de la presencia del término, pues el mismo término "virtualidad" está presente en el texto.

Para resolver problemas con el mencionado texto se volvió a entrenar el modelo y aplicarlo por segunda vez a los datos. En esta segunda corrida para las oraciones del ejemplo se obtuvieron los resultados mostrados en la siguiente tabla.

⁴ En este contexto, el término "ruido" se refiere a las clasificaciones con baja confiabilidad, es decir, aquellas que el modelo asigna con un nivel de certeza reducido. Estas clasificaciones de baja confianza pueden introducir imprecisiones o distorsiones en el análisis general.

Tabla 6

Resultados obtenidos luego del segundo entrenamiento

Texto analizado	Análisis de sentimiento	Clasificación
No me gusta mucho su método de enseñanza. Sin embargo, considero que la virtualidad es el principal problema en el correcto aprendizaje de este curso.	-0.782193	virtualidad (0.461484)
Muy bueno el profesor, deja casos aplicativos y nos de trabajos para complementar la enseñanza.	0.841601	metodología (0.459433)

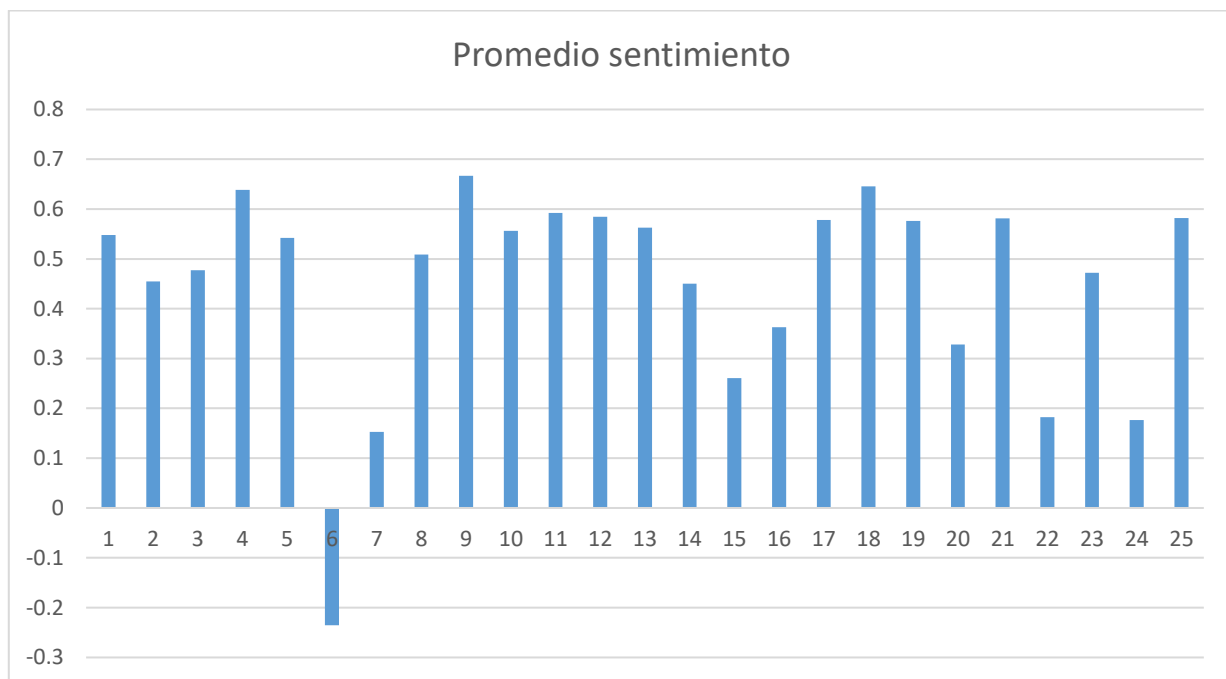
Como se aprecia en la tabla anterior los resultados del análisis de sentimiento no han variado, pero si ha variado la clasificación (producto del reentrenamiento del modelo), en particular la clasificación del primer texto, que ahora muestra correctamente la virtualidad como categoría.

3.7.3 Resultados de la experiencia piloto

El programa da como resultado una hoja de cálculo con los valores del análisis de sentimientos y la segmentación y esta información es utilizada por el área de *Business Intelligence* para el tratamiento y procesamiento de los datos. Ahora bien, dada la gran cantidad de datos se presentan los resultados de 25 cursos como ejemplos de los datos obtenidos.

La siguiente figura muestra el promedio de puntajes obtenidos por un grupo de cursos analizados.

Figura 10
Promedios de puntajes de sentimiento obtenido

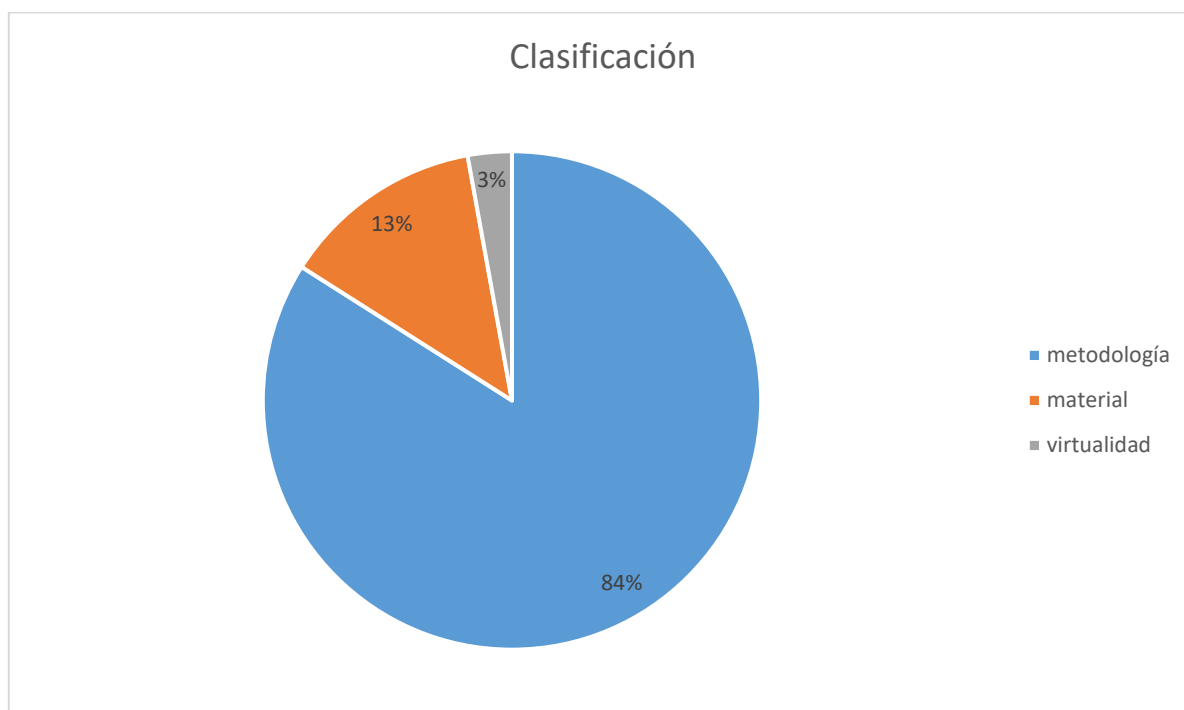


Dentro de la muestra resalta el resultado con promedio de sentimiento negativo, revisando las respuestas se verifica que muchos de los comentarios hacen referencia a que la profesora ha cancelado muchas clases durante el ciclo y comentarios sobre su metodología. Este tipo de hechos podría ser más difícil de identificar dentro de las miles de respuestas obtenidas.

Por otro lado, en la muestra podemos ver la cantidad de comentarios de acuerdo a cada una de las clasificaciones entrenadas. La siguiente figura muestra los porcentajes de cada categoría.

Figura 11

Porcentajes correspondientes a cada clasificación entrenada

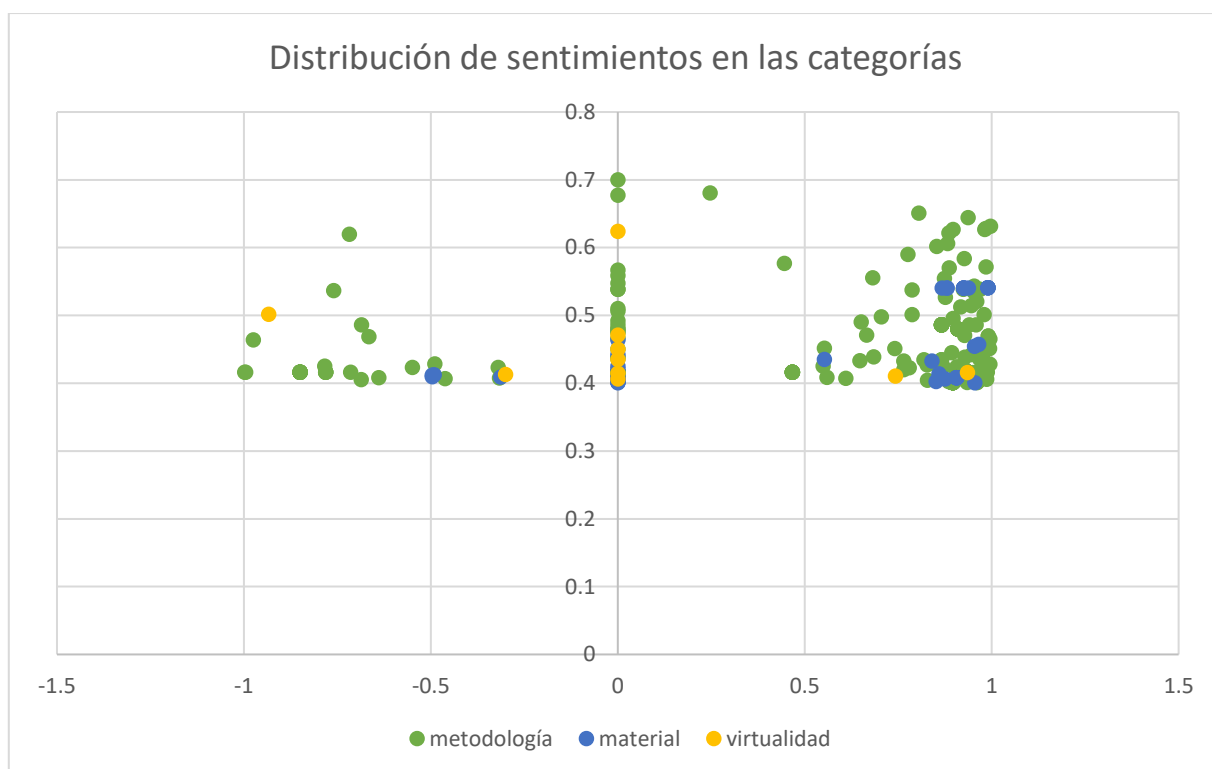


En la muestra obtenida, la gran cantidad de comentarios han sido clasificados como que son relacionados a la metodología utilizada en la clase. Cabe resaltar que estas categorías fueron elegidas para el piloto y se pueden considerar otras categorías o una mayor cantidad de acuerdo al tipo de análisis que se quiera realizar.

Finalmente combinando puntajes de sentimiento con la clasificación, se puede tener una idea adicional del sentimiento expresado hacia cada una de las categorías. Un ejemplo de este tipo de resultados se muestra en la siguiente figura.

Figura 12

Distribución de puntaje de sentimientos de acuerdo a categorías



El gráfico anterior confirma la distribución de los resultados en categorías y nos permite interpretar que la mayoría de problemas puntuales son identificados como que tienen que ver con la metodología, sin embargo, en general la metodología se percibe como buena.

CONCLUSIONES

A partir del diseño de la propuesta de innovación y de la ejecución de la experiencia piloto es posible plantear las siguientes conclusiones:

Respecto al diseño de la propuesta

- La incorporación de la Inteligencia Artificial y, en particular, el procesamiento de lenguaje natural en la educación es un fenómeno que ya se está dando y es importante involucrar a este tipo de tecnologías para ayudar en la toma de decisiones.
- Mientras se tengan respuestas más completas, por parte de los alumnos, el sistema podrá procesarlas de mejor manera, por lo que se requiere una etapa de sensibilización a los alumnos sobre la importancia de la encuesta.

Respecto a la ejecución de la experiencia piloto

- A partir de los resultados obtenidos en la experiencia piloto, el uso del modelo de procesamiento de lenguaje natural nos da una herramienta muy eficiente para el análisis de respuestas a preguntas abiertas.
- El análisis de sentimientos nos da un indicador general importante de la perspectiva de los alumnos hacia un curso o profesor y permite además identificar patrones y agrupar respuestas de una manera eficiente, un proceso que, si bien podría ser ejecutado por un recurso humano, dada la gran cantidad de datos, esto implicaría un esfuerzo enorme y tiempos elevados.

RECOMENDACIONES

Considerando la presente propuesta y los resultados obtenidos se realizan las siguientes recomendaciones y trabajos futuros:

- Perfeccionar el Modelo de Análisis de Sentimientos. Aunque el proyecto piloto demostró la eficacia del análisis de sentimientos para evaluar las percepciones de los estudiantes, es posible mejorar la precisión y el valor del modelo entrenándolo con un conjunto de datos más amplios y diversos de respuestas estudiantiles, incluyendo comentarios de diferentes cursos, disciplinas y perfiles demográficos de estudiantes. Asimismo el modelo no es capaz de comprender jergas y matices del lenguaje y el sentimiento en el contexto educativo específico.
- Mejorar el Modelo de Clasificación. El proyecto piloto se centró en un conjunto limitado de categorías predefinidas. Expandir y refinar el modelo de clasificación permitirá obtener resultados más ricos, por ejemplo, incorporando nuevas categorías de interés y planteando sub-categorías para pasar de temas amplios a algunos más específicos.
- Mejorar la Interfaz de Usuario y los Informes. Un uso más constante de la herramienta hará que sea más importante contar con un diseño más amigable e intuitivo para que se pueda usar esta sin requerir de soporte técnico especializado.

REFERENCIAS BIBLIOGRÁFICAS

- Andriessen, J., & Sandberg, J. (1999). Where is education heading and how about AI. *International Journal of Artificial Intelligence in Education*, 10(2), 130-150. <https://www.scirp.org/reference/referencespapers?referenceid=3391255>
- Ayuso-del Puerto, D., & Gutiérrez-Esteban, P. (2022). La Inteligencia Artificial como recurso educativo durante la formación inicial del profesorado. *Revista Iberoamericana de Educación a Distancia*, 25(2). <https://doi.org/10.5944/ried.25.2.32332>
- Bang, H. J., Li, L., & Flynn, K. (2023). Efficacy of an adaptive game-based math learning app to support personalized learning and improve early elementary school students' learning. *Early Childhood Education Journal*, 51, 717-732.
- Baker, V., Elliot, B., Sicular, S., Mullen, A., & Brethenoux, E. (2021). *Magic Quadrant for Cloud AI Developer Services*. Gartner. <https://www.gartner.com/doc/reprints?id=1-25C36W9W&ct=210226&st=sb>
- Bresciani, S., Eppler, M., & Eppler, M. J. (2008). *Gartner's magic quadrant and Hype cycle*. <https://www.researchgate.net/publication/265407275>
- Bicknell, K., Brust, C., & Settles, B. (2023). How Duolingo's AI Learns what you Need to Learn: The language-learning app tries to emulate a great human tutor. *IEEE Spectrum*, 60(3), 28–33. <https://doi.org/10.1109/mspec.2023.10061631>
- Floridi, L., Cows, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines*, 28(4), 689–707. <https://doi.org/10.1007/s11023-018-9482-5>
- García-Cuevas, J., Alor-Dávila, L., & Cisneros-Del Toro, Y. (2023). Percepción de los tutores virtuales sobre el impacto de la inteligencia artificial en la educación

- universitaria. *Company Games & Business Simulation Academic Journal*, 3(1), 49-58.
- Gutiérrez, E. (2010). Un modelo de evaluación del desempeño docente que contribuye en la mejora de la calidad de los servicios educativos. *Congreso Iberoamericano de Educación* (pp. 1-11). Buenos Aires.
- Kasliwal, N. (2018). *Natural Language Processing with Python Quick Start Guide: Going from a Python developer to an effective Natural Language Processing Engineer*. Packt Publishing.
- Khaled, D. (2014). Natural Language Processing and its Use in Education. *International Journal of Advanced Computer Science and Applications*, 5(12). <https://doi.org/10.14569/ijacsa.2014.051210>
- Litman, D. (2016). Natural Language Processing for Enhancing Teaching and Learning. *AAAI Conference on Artificial Intelligence*. <https://www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/view/12310/12208>
- Mdavis. (2022). Classroom chatbot improves student performance, study says. *Georgia State News Hub*. <https://news.gsu.edu/2022/03/21/classroom-chatbot-improves-student-performance-study-says/>
- Ng, D. T. K., Lee, M., Tan, R. J. Y., Hu, X., Downie, J. S., & Chu, S. K. W. (2023). A review of AI teaching and learning from 2000 to 2020. *Education and Information Technologies*, 28, 8445–8501. <https://doi.org/10.1007/s10639-022-11491-w>
- Perezchica-Vega, J. E., Sepúlveda-Rodríguez, J. A., & Román-Méndez, A. D. (2024). Inteligencia artificial generativa en la educación superior: usos y opiniones de los profesores. *European Public & Social Innovation Review*, 9, 01-20.
- Peterson, K. D. (2000). Student Surveys for School Teacher Evaluation. *Journal of Personnel Evaluation in Education*, 14(2).

- Pratama, M. P., Sampelolo, R., & Lura, H. (2023). Revolutionizing education: Harnessing the power of artificial intelligence for personalized learning. *Klasikal: Journal of Education, Language Teaching and Science*, 5(2), 350-357.
- Roll, I., & Wylie, R. (2016). Evolution and Revolution in Artificial Intelligence in Education. *International Journal of Artificial Intelligence in Education*, 26(2), 582–599. <https://doi.org/10.1007/s40593-016-0110-3>
- Rollett, W., Bijlsma, H., & Röhl, S. (2021). *Student Feedback on Teaching in Schools: Using Student Perceptions for the Development of Teaching and Teachers*. Springer.
- Ruediger, D., Blankstein, M., & Love, S. (2024). *Generative AI and Postsecondary Instructional Practices: Findings from a National Survey of Instructors*. <https://doi.org/10.18665/sr.320892>
- Ruiz-Mendoza, K. K., Miramontes-Arteaga, M. A., & Castillo-Villapudua, K. Y. (2024). Experiences in training teachers at universities in Baja California on generative AI. In *CS & IT Conference Proceedings* (Vol. 14, No. 12). <https://aircconline.com/csit/papers/vol14/csit141203.pdf>
- Russell, S., & Norvig, P. (2009). *Artificial Intelligence: A Modern Approach* (3rd ed.). Pearson.
- Scrum.org. (n.d.). What is Scrum? <https://www.scrum.org/resources/what-is-scrum/>
- Skedsmo, G., & Huber, S. G. (2017). Evaluation of educators' performance—balancing various measures to improve practice. *Educational Assessment, Evaluation and Accountability*, 29(2), 107–110. <https://doi.org/10.1007/s11092-017-9262-4>
- Skedsmo, G., & Huber, S. G. (2018). Teacher evaluation: the need for valid measures and increased teacher involvement. *Educational Assessment, Evaluation and Accountability*, 30(1), 1–5. <https://doi.org/10.1007/s11092-018-9273-9>

- Shinkfield, A. J., & Stufflebeam, D. L. (2012). *Teacher Evaluation: Guide to Effective Practice*. Springer.
- Srinivasa-Desikan, B. (2018). *Natural Language Processing and Computational Linguistics*. Van Haren Publishing.
- Stiggins, R. J., & Duke, D. L. (Eds.). (1988). *Case for Commitment to Teacher Growth, The: Research on Teacher Evaluation*. SUNY Press.
- Sunedu. (2019, June 13). Sunedu otorga el licenciamiento número 74 a la universidad tecnológica del Perú. <https://www.sunedu.gob.pe/sunedu-otorga-licenciamiento-numero-74-universidad-tecnologica-del-peru/>
- Thai, K. P., Bang, H. J., & Li, L. (2021). Accelerating early math learning with research-based personalized learning games: A cluster randomized controlled trial. *Journal of Research on Educational Effectiveness*, 14(4), 897-927.
- Thomas, D. R., Lin, J., Gatz, E., Gurung, A., Gupta, S., Norberg, K., Fancsali, S. E., Alevan, V., Branstetter, L., Brunskill, E., & Koedinger, K. R. (2024). Improving Student Learning with Hybrid Human-AI Tutoring: A Three-Study Quasi-Experimental Investigation. In *The 14th Learning Analytics and Knowledge Conference (LAK '24), March 18–22, 2024, Kyoto, Japan*. ACM. <https://doi.org/10.1145/3636555.3636896>
- Tsoi, J. C. H., & Strønen, F. (2024). Integration of Conversational AI Capabilities in Knowledge Management Processes for Higher Education. *Proceedings of the 25th European Conference on Knowledge Management, ECKM 2024*, 1026-1033.
- Tuytens, M., Vanhoof, J., Vekeman, E., Devos, G., & Van Petegem, P. (2020). An integral perspective on teacher evaluation: a review of empirical studies. *Educational Assessment, Evaluation and Accountability*. <https://doi.org/10.1007/s11092-020-09321-z>
- UNESCO (2021). *Inteligencia Artificial y Educación: Guía para las Personas a Cargo de Formular Políticas*. <https://unesdoc.unesco.org/ark:/48223/pf0000379376>

UTP. (2021). *Plana docente y docentes que realizan investigación 2021 - 1*. <https://www.utp.edu.pe/sites/default/files/transparencia/PLANA-DOCENTE2021-1.pdf>

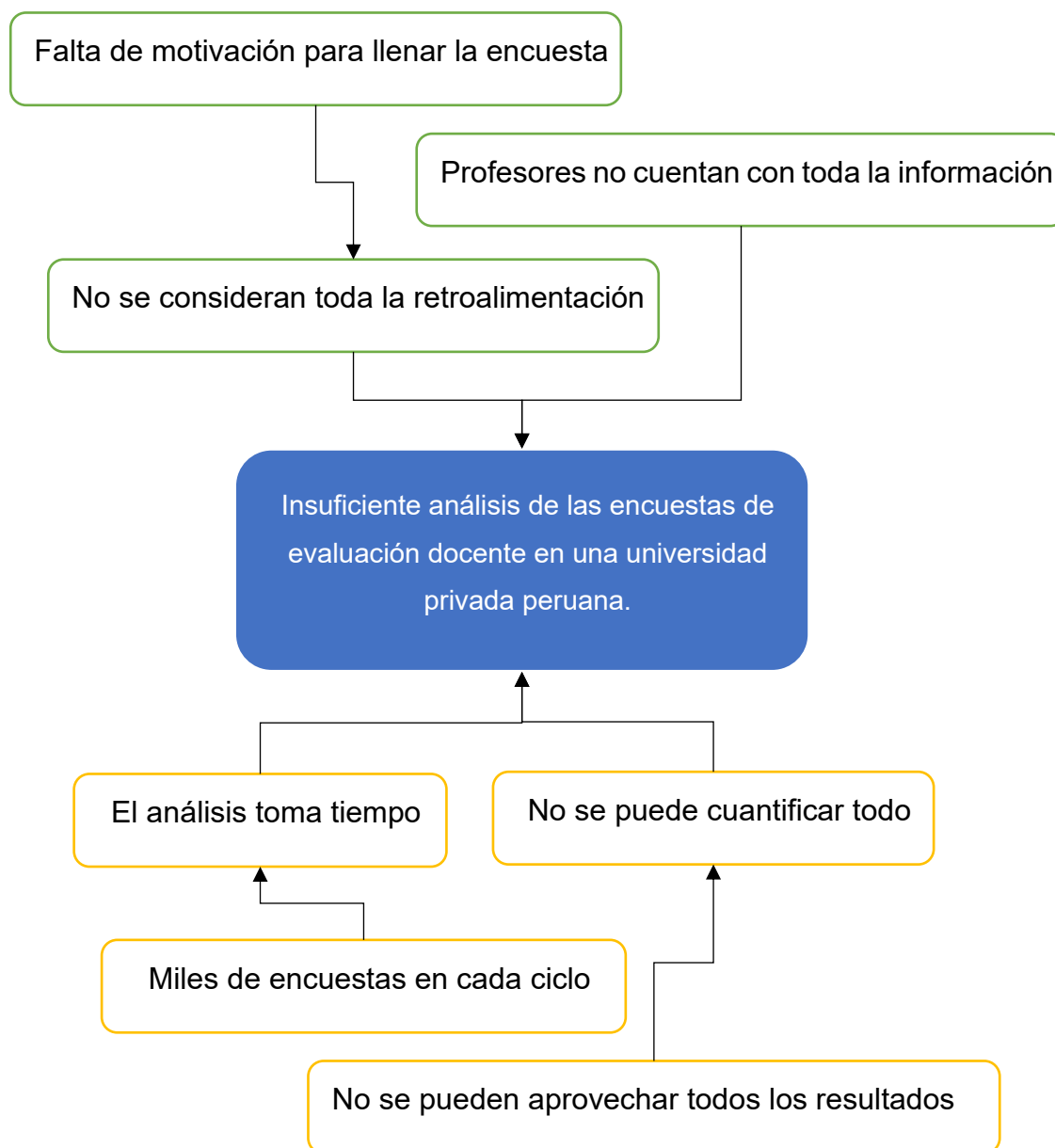
Wu, D., Chen, M., Chen, X., & Liu, X. (2024). Analyzing K-12 AI education: A large language model study of classroom instruction on learning theories, pedagogy, tools, and AI literacy. *Computers and Education: Artificial Intelligence*, 7, 100295.



ANEXOS

Anexo 1: Árbol de problemas

EFFECTOS



CAUSAS

Anexo 2: Encuesta realizada

A continuación, se presenta un listado de las preguntas de la encuesta:

¿Asiste puntualmente a clases?
¿Demuestra que conoce los temas del curso?
¿Motiva el interés de los estudiantes por el curso?
¿Explica de manera comprensible?
¿Propone actividades para aplicar lo aprendido?
¿Promueve en clase la participación de los estudiantes?
¿Hace uso adecuado de recursos tecnológicos para el aprendizaje?
¿Los materiales del curso favorecen el aprendizaje?
¿Las evaluaciones guardan relación con el contenido del curso?
¿Entrega puntualmente las evaluaciones corregidas?
¿Devuelve las evaluaciones con comentarios?
¿Establece un vínculo respetuoso con sus estudiantes?
En general:
¿Consideras que este curso te ha permitido aprender?
¿Cómo calificarías este curso?
Elige tu respuesta considerando 1 el mejor puntaje y 10 el mayor puntaje.
¿Recomendarías a este profesor(a) a otros compañeros?

Adicionalmente se tiene una pregunta de comentarios adicionales que será la pregunta abierta a analizar.

Anexo 3: Entrenamiento de modelos

```
api_key = ****
url = 'https://api.us-south.natural-language-
understanding.watson.cloud.ibm.com/instances/356aef86-d2de-4a41-8243-
6d8937f86717'

from ibm_watson import NaturalLanguageUnderstandingV1
from ibm_cloud_sdk_core.authenticators import IAMAuthenticator

auth = IAMAuthenticator(api_key)
nlu = NaturalLanguageUnderstandingV1(version='2021-03-25', authenticator=auth)

nlu.set_service_url(url)

training_data = [
    {
        "text": "buena metodología",
        "labels": ["metodología"]
    },
    {
        "text": "muchas tareas",
        "labels": ["metodología"]
    },
    {
        "text": "buenos ejemplos",
        "labels": ["metodología"]
    },
    {
        "text": "clases didácticas",
        "labels": ["metodología"]
    },
    {
        "text": "incentiva participación",
        "labels": ["metodología"]
    },
    {
        "text": "buenas dinámicas",
        "labels": ["metodología"]
    },
    {
        "text": "buen contenido del curso",
        "labels": ["material"]
    },
    {
        "text": "uso de canva",
        "labels": ["material"]
    }
]
```

```

    },
    {
        "text": "mal uso del material",
        "labels": ["material"]
    },
    {
        "text": "no comparte las diapositivas",
        "labels": ["material"]
    },
    {
        "text": "buen material",
        "labels": ["material"]
    },
    {
        "text": "clases virtuales",
        "labels": ["virtualidad"]
    },
    {
        "text": "uso de zoom",
        "labels": ["virtualidad"]
    },
    {
        "text": "conexión a internet",
        "labels": ["virtualidad"]
    },
    {
        "text": "fallas de conexión",
        "labels": ["virtualidad"]
    },
    {
        "text": "videoconferencias",
        "labels": ["virtualidad"]
    },
    {
        "text": "la virtualidad es un problema",
        "labels": ["virtualidad"]
    }
}
]

# Save Training data in a file
import json

training_data_filename = 'training_data.json'

with open(training_data_filename, 'w', encoding='utf-8') as f:
    json.dump(training_data, f, indent=4)

with open(training_data_filename, 'rb') as file:

```

```
model = nlu.create_classifications_model(language='es',
training_data=file, training_data_content_type='application/json',
name='TesisModel', model_version='1.0.1').get_result()

print("Created a NLU Classifications model:")
print(json.dumps(model, indent=4))

model_id = model['model_id']
model_to_view = nlu.get_classifications_model(model_id=model_id).get_result()
```



Anexo 4: Programa desarrollado para realizar el análisis de sentimiento y segmentación de las respuestas

```
!pip install pandas
!pip install ibm-cos-sdk==2.0.1
!pip install boto3
!pip install ibm_watson
import os, types
import pandas as pd
import numpy as np
from botocore.client import Config
import ibm_boto3
from ibm_watson import NaturalLanguageUnderstandingV1
from ibm_watson.natural_language_understanding_v1 import Features,
EntitiesOptions, KeywordsOptions, SentimentOptions
from ibm_cloud_sdk_core.authenticators import IAMAuthenticator
from tqdm.autonotebook import tqdm
tqdm.pandas()

df_data_1 = pd.read_csv("encuesta2.csv")
df_data_1 = df_data_1.drop(columns='Apellidos_Nombres_Docentes')
df_data_1.head()
```

Programa	Módulo	CodCampus	Codigo_Docente	Seccion	Cod_Curso	Curso	Matriculados	Encuestados	P18	
0	PREG	Presencial	CHICL	C17520	18767	100000107N	PRINCIPIOS DE ALGORITMOS	51	40	NaN
1	PREG	Presencial	CHICL	C17520	18767	100000107N	PRINCIPIOS DE ALGORITMOS	51	40	..
2	PREG	Presencial	CHICL	C17520	18767	100000107N	PRINCIPIOS DE ALGORITMOS	51	40	NaN
3	PREG	Presencial	CHICL	C17520	18767	100000107N	PRINCIPIOS DE ALGORITMOS	51	40	NaN
4	PREG	Presencial	CHICL	C19616	20609	100000CI06	FUNDAMENTOS DE DINÁMICA	49	40	Excelente

```
authenticator = IAMAuthenticator(****)

# Authentication via external config like VCAP_SERVICES
service = NaturalLanguageUnderstandingV1(
    version='2018-03-16', authenticator=authenticator)
# service.set_disable_ssl_verification(True)
service.set_service_url('https://api.us-south.natural-language-
understanding.watson.cloud.ibm.com/instances/356aef86-d2de-4a41-8243-
6d8937f86717')
"""
Consolidated method for Document level sentiment and Keyword level sentiment
analysis
OUTPUT: JSON
model-id = 8f9bde0e-4484-4059-9d02-91854d1f2609
"""
```

```

from ibm_watson.natural_language_understanding_v1 import Features,
ClassificationsOptions
model_id="05816c1b-8ac3-419f-a772-525806f5e553"

def sentiment_and_keyword(st,service=service):
    # return
    (service.analyze(text=st,language='es',features=Features(keywords=KeywordsOptions(sentiment= True,emotion= True,limit=
3),sentiment=SentimentOptions())).get_result())
    return
    (service.analyze(text=st,language='es',features=Features(classifications=ClassificationsOptions(model=model_id),keywords=KeywordsOptions(sentiment=
True,emotion= True,limit= 3),sentiment=SentimentOptions())).get_result())

"""
Extract sentiment score and append it to a new column in the dataframe
INPUT: JSON
OUTPUT: NUMERIC SCORE
"""

def sentiment_extractor(row_num):
    sentiment_scores=df2['sentiment_keyword_json'][row_num]['sentiment']['document']['score']
    return sentiment_scores

"""
Extract keyword score/s and append it to a new column in the dataframe
INPUT: JSON
OUTPUT: DICTIONARY [{Keyword1:Score},{Keyword2:Score},{Keyword3:Score}]
"""

def keyword_extractor(row_num):
    keyword_scores=[{df2['sentiment_keyword_json'][row_num]['keywords'][i]['text'] :
df2['sentiment_keyword_json'][row_num]['keywords'][i]['sentiment']['score']}
for i in range(0,len(df2['sentiment_keyword_json'][row_num]['keywords']))]
    return keyword_scores

"""
Extract classifications score/s and append it to a new column in the dataframe
INPUT: JSON
OUTPUT: DICTIONARY
[{'classifications:Score},{classifications:Score},{classifications:Score}]
"""

def classifications_extractor(row_num):
    classifications_scores=[{df2['sentiment_keyword_json'][row_num]['classifications'][i]['class_name'] :
df2['sentiment_keyword_json'][row_num]['classifications'][i]['confidence']}
for i in
range(0,len(df2['sentiment_keyword_json'][row_num]['classifications']))]

```

```

    return classifications_scores
df1['P18'] = df1['P18'].str.lower()
df1['P18'] = df1['P18'].replace({' ok ': ' bien '}, regex=True)
df1['P18'] = df1['P18'].replace({' okey': ' bien'}, regex=True)
df1['sentiment_keyword_json'] = df1.progress_apply(lambda row :
sentiment_and_keyword(str(row['P18'])),axis = 1)
df2 = df1[['Seccion', 'P18', 'sentiment_keyword_json']]

# Quitar vacíos
df2['P18'].replace('', np.nan, inplace=True)
df2.dropna(subset=['P18'], inplace=True)

"""
Extract document-level sentiment score and keyword-level sentiment score
and update it in 2 new columns, namely 'sentiment_score' and 'keyword_score'
"""
df2['sentiment_score'] = df2.apply(lambda row :
sentiment_extractor(row.name),axis = 1)
# df2['keyword_score'] = df2.apply(lambda row :
keyword_extractor(row.name),axis = 1)
df2['classifications'] = df2.apply(lambda row :
classifications_extractor(row.name),axis = 1)

```

	Seccion	P18	sentiment_keyword_json	sentiment_score	classifications
1	18767	..	{'usage': {'text_units': 1, 'text_characters':...	0.000000	{'metodología': 0.336281}, {'material': 0.324...
4	20609	excelente	{'usage': {'text_units': 1, 'text_characters':...	0.998324	{'material': 0.315591}, {'metodología': 0.308...
5	20609	buen docente e interesante curso	{'usage': {'text_units': 1, 'text_characters':...	0.953846	{'material': 0.45426}, {'metodología': 0.3146...
6	20609	las encuestas deben ser voluntarias	{'usage': {'text_units': 1, 'text_characters':...	0.000000	{'metodología': 0.538376}, {'virtualidad': 0...
8	20609	excelente curso	{'usage': {'text_units': 1, 'text_characters':...	0.989662	{'material': 0.540475}, {'metodología': 0.205...
...
994	9207	el aplica una metodología de enseñanza que pri...	{'usage': {'text_units': 1, 'text_characters':...	0.000000	{'material': 0.336218}, {'metodología': 0.197...
996	9207	buen docente	{'usage': {'text_units': 1, 'text_characters':...	0.865754	{'metodología': 0.485693}, {'material': 0.262...
997	9207	m	{'usage': {'text_units': 1, 'text_characters':...	0.000000	{'metodología': 0.416202}, {'material': 0.367...
998	9207	bajan la pension uu	{'usage': {'text_units': 1, 'text_characters':...	0.000000	{'virtualidad': 0.302941}, {'metodología': 0...
999	9207	bueno.	{'usage': {'text_units': 1, 'text_characters':...	0.969255	{'metodología': 0.38041}, {'material': 0.2397...

```

def classifications_extractor_metodologia(row_num):
    classifications_scores=[df2['classifications'][row_num]['classifications'
][i]['class_name'] :
df2['sentiment_keyword_json'][row_num]['classifications'][i]['confidence']]
for i in range(0,2)]
    return classifications_scores

df2['metodología'] = df2.apply(lambda row :
classifications_extractor_metodologia(row.name),axis = 1)

```