



**UNIVERSIDAD AUTÓNOMA DEL ESTADO  
DE HIDALGO**

---

---

**INSTITUTO DE CIENCIAS BÁSICAS E INGENIERÍA  
ÁREA ACADÉMICA DE COMPUTACIÓN Y ELECTRÓNICA**

**Análisis de los aspectos académicos en la Educación Superior  
ocasionados por la Pandemia COVID-19 mediante técnicas  
descriptivas de Minería de datos.**

**T E S I S**

Que para obtener el título de  
**LICENCIADA EN CIENCIAS COMPUTACIONALES**

Presenta  
**ERICA GARCÍA GARCÍA**

Directores de tesis  
**DRA. MA. DE JESÚS GUTIÉRREZ SÁNCHEZ**  
**DRA. ANILU FRANCO ARCEGA**

Mineral de la Reforma, Hidalgo 2024



Universidad Autónoma del Estado de Hidalgo  
 Instituto de Ciencias Básicas e Ingeniería  
*School of Engineering and Basic Sciences*

Mineral de la Reforma, Hgo., a 2 de septiembre de 2024

Número de control: ICBI-D/1041/2024  
 Asunto: Autorización de impresión.

**MTRA. OJUKY DEL ROCÍO ISLAS MALDONADO**  
**DIRECTORA DE ADMINISTRACIÓN ESCOLAR DE LA UAEH**

Con fundamento en lo dispuesto en el Título Tercero, Capítulo I, Artículo 18 Fracción IV; Título Quinto, Capítulo II, Capítulo V, Artículo 51 Fracción IX del Estatuto General de nuestra Institución, por este medio le comunico que el Jurado asignado a la Egresada de la Licenciatura en Ciencias Computacionales **Erica García García**, quien presenta el trabajo de titulación **“Análisis de los aspectos académicos en la Educación Superior ocasionados por la Pandemia COVID-19 mediante técnicas descriptivas de Minería de Datos”**, después de revisar el trabajo en reunión de Sinodales ha decidido autorizar la impresión del mismo, hechas las correcciones que fueron acordadas.

A continuación, firman de conformidad los integrantes del Jurado:

**Presidente:** Dra. Theira Irasema Samperio Monroy

**Secretario:** MID. Alberto Suárez Navarrete

**Vocal:** MID. Ma. de Jesús Gutiérrez Sánchez

**Suplente:** Dra. Anilú Franco Arcega

Sin otro particular por el momento, reciba un cordial saludo.

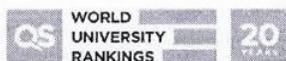
Atentamente  
 “Amor, Orden y Progreso”

Mtro. Gabriel Vergara Rodríguez  
 Director del ICBI



GVR/YCC

Ciudad del Conocimiento, Carretera Pachuca-Tulancingo Km. 4.5 Colonia Carboneras, Mineral de la Reforma, Hidalgo, México. C.P. 42184  
 Teléfono: 771 71 720 00 Ext. 40001  
 direccion\_icbi@uaeh.edu.mx,  
 vergarar@uaeh.edu.mx



# Índice

Resumen	v
Introducción	vi
Planteamiento del problema	vii
Propuesta de solución	viii
Justificación	ix
Objetivo general	x
Objetivos específicos	x
Alcances y limitaciones	xi
Estructura del documento	xi
<b>CAPÍTULO 1 MARCO TEÓRICO</b>	<b>1</b>
1.1 Antecedentes	1
1.2 Educación-andragogía	2
1.2.1 Historia	3
1.2.2 Teoría de Knowles	4
1.2.3 Características de la andragogía	5
1.3 Educación antes de la pandemia	6
1.4 Educación durante la pandemia	6
1.5 Educación después de la pandemia	7
1.6 Cuestionarios de seguimiento a la educación	8
1.7 Aprendizaje automatizado	15
1.7.1 Usos del aprendizaje automático	15
1.8 Minería de datos	16
1.8.1 Beneficios de la minería de datos	16
1.9 Aprendizaje supervisado	17
1.9.1 Naive Bayes	18
1.9.2 Árbol de decisión	18
1.10 Aprendizaje no supervisado	19
1.10.1 Aplicaciones del aprendizaje no supervisado	20
1.10.2 K- means	20
1.10.3 Clustering jerárquico	23
1.11 Coeficiente alfa de Cronbach	24

1.12 Índices de validez	25
1.12.1 Calinski harabasz	26
1.12.2 Davies Bouldin	27
1.12.3 Silhouette	28
1.13 Escala de Likert	29
1.13.1 Ventajas	30
1.13.2 Desventajas	30
1.14 WEKA	31
1.14.1 Breve historia	31
1.14.2 Características	31
CAPITULO 2 ESTADO DEL ARTE	32
2.1 El uso de mecanismos de innovación en la educación de adultos	32
2.2. Retos de la educación virtual en el proceso enseñanza aprendizaje durante la pandemia de COVID-19	33
2.3 Factores asociados al éxito de los estudiantes en modalidad de aprendizaje en línea un análisis en minería de datos	34
2.4 Relación entre el rendimiento escolar y la ansiedad en estudiantes de la licenciatura en ciencias de la educación de la UAEH durante la pandemia, semestre enero-junio 2022	35
2.5 Ansiedad, estrés y percepción de riesgo por COVID-19 en estudiantes universitarios	36
CAPÍTULO 3 METODOLOGÍA	38
3.1 KDD	38
3.2 CRISP-DM	40
3.3 SEMMA	41
3.4 Diseño y Desarrollo de Metodología	44
3.4.1 Etapa de selección	44
3.4.2 Etapa pre-procesamiento o limpieza	59
3.4.3 Etapa de transformación	60
3.4.4 Etapa Minería de datos	61
3.4.5 Interpretación y evaluación	64
Conclusiones	70
Referencias	71

# Índice de figuras

Figura 1 Sentir de los docentes ante los cambios que implica el ajuste de clases presenciales a virtuales (a) (Miguel, 2020).	8
Figura 2 Obstáculo más significativo de los docentes ante los cambios que implica el ajuste de clases presenciales a virtuales (b) (Miguel, 2020).	9
Figura 3 Competencias desarrolladas por los docentes ante los cambios que implica el ajuste de clases presenciales a virtuales (c) (Miguel, 2020).	9
Figura 4 ¿Cómo te sientes con el ajuste de clases presenciales a virtuales debido a la contingencia sanitaria COVID-19? (Miguel, 2020).	10
Figura 5 ¿Cuál consideras que ha sido el obstáculo más significativo de este ajuste de clases presenciales a virtuales debido a la contingencia sanitaria: COVID-19? (Miguel, 2020).	11
Figura 6 ¿Cuáles son las competencias que has tenido que desarrollar para solucionar los obstáculos del presente contexto y adaptarte? (Miguel, 2020)	12
Figura 7 Dificultades de los estudiantes (IESALC, 2020)	14
Figura 8 Alfa de Cronbach (Pérez León, 2022.)	24
Figura 9 Calinski-Harabasz (Dey, 2022)	27
Figura 10 Silhouette (Prado, 2022)	28
Figura 11 Escala de Likert (Muguira, 2020).	30
Figura 12 Diagrama KDD (Hitpass, 2014)	39
Figura 13 Diagrama CRISP-DM (Bellini Saibene, 2021)	41
Figura 14 Diagrama SEMMA (Jun, 2017)	42
Figura 15 Vista desde Google Forms 1. Elaboración Propia.	50
Figura 16 Vista desde Google Forms 2. Elaboración Propia.	51
Figura 17 Vista desde Google Forms 3. Elaboración Propia.	52
Figura 18 Vista desde Google Forms 4. Elaboración Propia.	53
Figura 19 Vista desde Google Forms 5. Elaboración Propia.	54
Figura 20 Grafica género. Elaboración Propia.	55
Figura 21 Grafica Edad. Elaboración Propia.	56
Figura 22 Zonas. Elaboración Propia.	56
Figura 23 Grafica elementos de aprendizaje. Elaboración Propia.	57
Figura 24 Grafica dificultades. Elaboración Propia.	57
Figura 25 Mayor conflicto. Elaboración Propia.	58
Figura 26 Mayor estrés. Elaboración Propia.	58
Figura 27 Limpieza de datos. Elaboración Propia.	59
Figura 28 Transformación de Información. Elaboración Propia.	60
Figura 29 Datos con 2 Clústeres. Elaboración Propia.	61
Figura 30 Datos con 3 Clústeres. Elaboración Propia.	62
Figura 31 Datos con 4 Clústeres. Elaboración Propia.	63
Figura 32 Distribución de la muestra 2 clústeres. Elaboración Propia.	64
Figura 33 Distribución de la muestra 3 clústeres. Elaboración Propia.	65
Figura 34 Distribución de la muestra 4 clústeres. Elaboración Propia.	66

## Índice de tablas

Tabla 1 Comparación de metodologías	43
Tabla 2 Elementos de aprendizaje	45
Tabla 3 Elementos personales que causaron mayor conflicto	46
Tabla 4 Elementos técnicos que causaron mayor estrés	46

## Índice de ecuaciones

Ecuación 1 Distancia Euclidea	21
Ecuación 2 Distancia de Manhattan	22
Ecuación 3 Coeficiente alfa de Cronbach	24
Ecuación 4 Índice Calinski-Harabasz	26

## Resumen

Constantemente, la tecnología evoluciona a gran velocidad lo que permite procesar datos de manera rápida y eficiente, mediante la implementación de algoritmos especializados en el trato de datos, para ello se tiene a la minería de datos que reúne una serie de técnicas, algoritmos que permiten analizar grandes cantidades de información con el propósito de encontrar patrones en ellos obteniendo resultados de manera certera y efectiva.

La educación, como se conoce es el proceso de enseñanza-aprendizaje en el estudiante, el cual siempre está en constante modificación para crear/fortalecer las habilidades, creencias, hábitos, del estudiante, con el propósito de tener una educación integral llena de valores y respeto por el medio ambiente.

Es por ello, que el presente trabajo se describe el proceso de recopilación, análisis y procesamiento de datos que ayuden a determinar qué aspectos académicos del estudiante resultados impactados por la pandemia que surgió el 30 de enero del 2020, denominada por la OMS COVID-19, por lo que las Instituciones de Educación Superior (IES), se vieron en la necesidad de modificar sus métodos de enseñanza tradicionales, a métodos virtuales.

Se presenta, el análisis de las dificultades que enfrentaron los estudiantes de la Licenciatura en Ciencias Computacionales, de séptimo semestre julio - diciembre 2023, en su educación durante el periodo de pandemia del COVID-19, para conocer el impacto de estas en el rendimiento académico mediante el uso de técnicas descriptivas de minería de datos.

## Introducción

La tecnología en tiempos de pandemia ha desempeñado un papel fundamental para que millones de estudiantes pudieran dar continuidad del proceso educativo haciendo que millones de estudiantes se vieran obligados a tomar clases en línea, donde la tecnología no sólo fue un medio de conexión, también una herramienta esencial para el desarrollo del autoaprendizaje y la comunicación efectiva entre docentes y alumnos.

El uso de plataformas educativas como Classroom, Moodle, Khan Academy, entre otras, no sólo ha facilitado la administración y distribución de actividades académicas, sino que también han permitido a los educadores medir el desempeño de sus estudiantes de manera más efectiva. La tecnología ha facilitado el acceso a recursos, permitiendo que los alumnos puedan aprender a su propio ritmo.

La minería de datos implica el análisis de información recopilada previamente, datos que, en el contexto educativo, pueden abarcar desde el rendimiento de los estudiantes hasta sus preferencias de aprendizaje. Esta herramienta permite identificar patrones, predecir el desarrollo de eventos, evaluar riesgos y tomar decisiones estratégicas que optimizan el proceso educativo.

En este proyecto se presentará un análisis de aquellos contratiempos que enfrentaron los estudiantes de séptimo semestre julio - diciembre 2023 durante su educación en línea, ocasionando problemas en su aprendizaje.

Obteniendo como resultado la división de datos en 2 grupos acorde a sus similitudes tomando aspectos como la autonomía, disciplina, dificultades personales y técnicas etc. donde se obtuvo que ambos clústeres coincidieron que la pandemia les había

afectado negativamente a su desempeño académico, a uno de los grupos ligeramente más que a otro, viéndose reflejado en su promedio, pero reprobar una materia no lo retribuyen directamente a las clases en línea, sino a otros aspectos como el desinterés académico, problemas económicos y personales entre otras.

## Planteamiento del problema

La Universidad Autónoma del Estado de Hidalgo es la máxima casa de estudios de la entidad hidalguense que ha ofrecido una educación con autonomía. Cuenta con un total de 6 institutos y 9 escuelas superiores, así como preparatorias, que se encuentran distribuidas al interior del Estado, de las cuales egresan cada año un total de 3 mil alumnos aproximadamente (UAEH - Universidad Autónoma del Estado de Hidalgo, 2017).

La epidemia de COVID-19 fue declarada por la OMS una emergencia de salud pública de preocupación internacional el 30 de enero de 2020 (Organización Panamericana de la salud, 2020), y como medida de prevención la Universidad Autónoma del Estado de Hidalgo (UAEH) suspendió las clases presenciales, tomando por sorpresa a docentes y alumnos, sin embargo la universidad ya contaba con la plataforma garza en el cual los docentes durante la pandemia subían su contenido educativo con actividades programadas en un periodo establecido junto con las evaluaciones de cada parcial, también buscaron otras alternativas para dar continuidad a la educación de miles de estudiantes como classroom, meet, google drive entre otras.

Durante estas clases en línea los alumnos enfrentaron múltiples dificultades, desde personales, tecnológicas, económicas, por mencionar algunas. Tras 2 años de confinamiento, la UAEH anunció el regreso a clases el 21 de febrero del 2022 de

manera escalonada, respetando las medidas sanitarias pertinentes para evitar la propagación del virus.

La problemática que se encontró, es que en el Instituto de Ciencias Básicas e Ingeniería la cual cuenta con la carrera en Ciencias Computacionales, con un plan de estudios de 9 semestres, distribuida en 4 grupos, de aproximadamente 35 alumnos al inicio. Durante esta pandemia el semestre de séptimo, del periodo julio - diciembre 2023 pasó sus primeros 2 años en clases virtuales (primero, segundo, tercero y cuarto semestre), en donde se desconocen cuáles fueron las principales desafíos de los alumnos, la falta de información dificulta la planificación de actividades en un posible escenario similar o en la consideración de realizar encuadres a los estudiantes al inicio de las clases que ayuden a identificar el nivel de sus conocimientos de un tema específico, para un mejor aprovechamiento.

## **Propuesta de solución**

Se visualizó que, para dar una solución al planteamiento del problema, se realizará una encuesta donde se incluirán diferentes secciones de grupos de contratiempo, así como elementos de aprendizaje, personales, técnicos y dificultades implementado en Google Forms, la cual se aplicará a 53 alumnos de séptimo semestre julio - diciembre 2023 de la licenciatura en Ciencias Computacionales, para conocer las principales dificultades que tuvieron durante su educación en la pandemia de COVID-19.

Una vez recolectada la información se procederá a analizar los datos con técnicas descriptivas de minería de datos como el K-Means, el cual es útil para la segmentación de información de acuerdo a las similitudes entre sí, lo que ayudará a categorizar las dificultades con mayor impacto en los estudiantes, siguiendo la

metodología KDD siendo una de sus características principales la extracción de conocimiento útil.

Posteriormente, se implementará el análisis con el algoritmo K- Means, que genere y/o busque grupos de identificación afines, para poder realizar una interpretación de los resultados obtenidos, durante este paso se prestará atención a las características y patrones que se definan a cada grupo.

## Justificación

El presente trabajo de investigación pretende dar a conocer los diversos desafíos tanto educativos, tecnológicos como personales; que se enfrentaron los alumnos en su educación durante la pandemia del COVID-19, generando conocimiento a la institución para tomar en consideración aspectos que pasaron inadvertidos, generando nuevas estrategias eficaces para futuras situaciones similares, siendo la educación un pilar importante para la formación de profesionales.

Gracias a las herramientas tecnológicas, como la minería de datos se puede realizar un análisis que identifique factores y consecuencias en diversos campos como: la agrupación, abstracción y filtrado de información que sea útil a la investigación, haciendo posible este tipo de proyectos eficazmente, en el cual se aplicara la Metodología KDD que cuenta con 5 etapas, las cuales son: selección, pre-procesamiento, transformación, minería de datos e interpretación.

Se cuenta con los conocimientos y herramientas necesarias para llevar a cabo este proyecto que es factible realizar dentro del Instituto de Ciencias Básicas e Ingeniería en la Licenciatura de Ciencias Computacionales, para conocer las dificultades de la

pandemia en la educación superior, dando un panorama más amplio de éstas, que se podrían evitar en escenarios futuros.

## Objetivo general

Analizar los aspectos que se enfrentaron los estudiantes de la Licenciatura en Ciencias Computacionales, de séptimo semestre en su educación durante el periodo de pandemia del COVID-19 mediante las técnicas descriptivas de minería de datos, para conocer el impacto de estas en el rendimiento académico de los estudiantes.

## Objetivos específicos

- Realizar una investigación previa sobre las dificultades que enfrentaron estudiantes en sus clases virtuales mediante la indagación de encuestas ya realizadas por instituciones para tener un panorama general de estas. para ser incorporadas en un cuestionario de Google Forms.
- Aplicar la encuesta a los alumnos de séptimo semestre de la Licenciatura en Ciencias Computacionales mediante el cuestionario de Google Forms para la recopilación de información.
- Revisar (procesar) los datos a través del algoritmo de K-Means, mediante la metodología KDD para llevar a cabo la identificación de las dificultades más significativas para los estudiantes.

## Alcances y limitaciones

La investigación está enfocada en el análisis de 35 estudiantes de la licenciatura en Ciencias Computacionales de la Universidad Autónoma del Estado de Hidalgo del semestre de 7o, tomando en cuenta los 4 grupos que lo componen en el semestre julio-diciembre 2023.

Los estudiantes no toman con la seriedad necesaria contestar el cuestionario, su falta de compromiso puede deberse a que consideren el formulario como una actividad de poca relevancia para su educación ya que no conlleva una calificación, pudiendo afectar a la calidad y validez de los datos para esta investigación obstaculizando obtener una mejor certeza de la información.

## Estructura del documento

### Capítulo 1 Marco Teórico

El primer capítulo proporciona el marco teórico para comprender esta investigación. Se presentan los conceptos fundamentales relacionados con el tema en estudio para así establecer una base conceptual para dicho trabajo. Además de exponer cuestionarios aplicados por otras instituciones siendo un punto de partida importante para dicho trabajo desarrollado.

### Capítulo 2 Estado del arte

En el presente capítulo aborda 5 estados del arte de los cuales se analizan detalladamente buscando similitud y relación con este trabajo. Sirviendo para proporcionar un contexto más amplio del tema a tratar como lo es la pandemia por COVID-19 y comprender mejor los diferentes aspectos a los que se enfrentaron los estudiantes.

### **Capítulo 3 Metodología**

En este tercer capítulo, se menciona 3 metodologías que podrían aplicarse a este proyecto, cada metodología se describen los pasos de esta, para posteriormente su comparación en un cuadro comparativo. Para finalmente elegir la mejor para este trabajo, además del desarrollo de la metodología elegida, desarrollada paso por paso.

### **Capítulo 4 Resultados**

En este capítulo se exponen los resultados tras el análisis de los datos recopilados durante la investigación, donde se detallan aquellos aspectos más significativos después del procesamiento de datos, proporcionado una interpretación de los mismos

# CAPÍTULO 1 MARCO TEÓRICO

En el presente capítulo, se dan a conocer conceptos relevantes de la presente investigación, así como artículos científicos que ayudaron a sustentar este trabajo. Además de presentar un antecedente relacionado donde se aborda la influencia que tuvo la pandemia por COVID-19 en la educación, que a diferencia de esta tesis fue realizado en educación básica, se pretende ampliar la comprensión sobre cómo la crisis sanitaria que se vivió a nivel internacional desde el punto de vista educativo.

Es importante tomar en cuenta diferentes contextos y niveles educativos, ya que los impactos pueden variar significativamente. Al explorar este antecedente, se pueden identificar patrones y tendencias que podrían aplicarse o adaptarse al contexto específico de la educación superior, proporcionando así una perspectiva más amplia sobre el tema.

## 1.1 Antecedentes

El COVID-19 ha afectado en diversas áreas de nuestra vida cotidiana y una de ellas es en la educación, donde se han realizado investigaciones para conocer su impacto, una de ellas fue publicada en la revista multidisciplinar Ciencia Latina, teniendo como nombre Pandemia, educación virtual y su impacto en la educación de la región Puno - Perú, que tuvo como objetivo describir el impacto de la COVID-19 en la educación básica, en la mencionada región, donde realizó un análisis del abandono escolar, como también de los problemas en el regreso a clases presenciales.

Este estudio tuvo un enfoque cuantitativo, a nivel exploratorio, en el cual realizaron reportes de asistencia del periodo 2020, y aplicaron un cuestionario a una muestra de 657 directores de instituciones escolares públicas de educación básica.

Dando como resultado que los factores con mayor medida en el abandono escolar, fueron la carencia o limitado acceso a internet, limitado acceso a recursos tecnológicos, poco apoyo de padres en las actividades académicas y problemas económicos en las familias, por otra parte, los problemas educativos por del confinamiento fueron estudiantes con bajos niveles de aprendizaje, poco interés para aprender y problemas socio emocionales como aburrimiento y frustración (Hurtado et al., 2022).

Este estudio está relacionado con el presente proyecto, ya que se realiza un análisis del impacto de la pandemia a nivel educación básica, a diferencia del presente proyecto que fue aplicado a nivel superior en la Universidad Autónoma del Estado de Hidalgo del mismo modo, enfocándose en las dificultades que presentaron los alumnos.

## **1.2 Educación-andragogía**

La andragogía es una disciplina que se enfoca en los procesos educativos de adultos, con el propósito de fomentar el crecimiento de sus habilidades, la expansión de sus conocimientos, la adquisición de nuevas tecnologías y, en términos generales, el mejoramiento de su desempeño en ámbitos personales, profesionales y sociales. (Sierra Fontalvo, 2006, p.100).

La andrología está orientada hacia la obtención de un modelo que promueva la capacidad de asumir cambios con juicio, la capacidad valorativa para tomar decisiones acordes a las necesidades de la sociedad actual

Según Sierra (2006), comenta que para la educación en adultos es recomendable tomar en cuenta los siguientes objetivos:

- Obtener y mantener al día la educación fundamental, así como facilitar la entrada a diversos niveles del sistema educativo.
- Fomentar su habilidad para involucrarse en la vida social, cultural, política y económica.
- Elevar su nivel de competencia profesional o adquirir la formación necesaria para desempeñarse en otras ocupaciones.

### 1.2.1 Historia

El término andragogía surgió en 1833 por un maestro de secundaria alemán llamado Alexander Kapp, el cual lo utilizó en un libro acerca de las teorías educativas de Platón. Knowles (2006), decía que los adultos tenían que seguir aprendiendo, también creía que la andragogía tenía que contemplar el aprendizaje a reflexionar, pero esta idea no fue muy reconocida ya que había otras teorías sobre la educación para adultos.

Knowles (2006) señaló las distintas formas en que los niños y los adultos aprenden, propuso la distinción entre la Pedagogía dirigida a niños y la Andragogía orientada a adultos. Sugirió que al igual que no se ajustan procesos andragógicos para niños, tampoco deberían aplicarse procesos pedagógicos para adultos. La distinción principal entre ambas disciplinas radica en que la Pedagogía implica un proceso educativo dirigido y estructurado, con un énfasis en la enseñanza por parte del educador, mientras que la andragogía se enfoca en un aprendizaje más autodirigido por parte del adulto, sin un énfasis tan marcado en la instrucción directa.

Lindeman dijo que la educación de adultos es cuando los estudiantes toman conciencia de sus experiencias más significativas conduciendo a una evaluación, donde se comprende el significado de las experiencias y su impacto en la personalidad. El cual mencionó algunos puntos sobre el aprendizaje en adultos:

- Los adultos encuentran motivación para aprender cuando tienen necesidad e interés que el aprendizaje satisfará.
- El aprendizaje de los estudiantes hacia el aprendizaje se centra en la vida.
- La experiencia es lo más valioso para el aprendizaje de los adultos.
- Los adultos tienen la necesidad de ser autónomos.
- Con el paso del tiempo las diferencias individuales son más notables.

### 1.2.2 Teoría de Knowles

Según Knowles (1984), la andragogía es el arte del aprendizaje adulto, considerándose el término pedagogía como equivalente de éste y a su vez, indicando 4 principios para aprendizaje en adultos, en donde se refleja la importancia de de la autonomía, experiencia y relevancia que tiene la educación en adultos.

#### Principios de la andragogía de Knowles

- Es importante que los adultos participen en la elaboración y análisis de su enseñanza.
- La experiencia incluyendo los errores representa la base de las tareas de aprendizaje.
- Los adultos muestran mayor interés en adquirir conocimientos que tengan aplicaciones inmediatas y repercusiones en su trabajo o vida personal.

*análisis de los aspectos académicos en la educación superior ocasionados por la pandemia mediante técnicas descriptivas de minería de datos*

- La educación de adultos se enfoca en la resolución de problemas en lugar de enfocarse solamente en contenidos (Knowles, 2020).

### 1.2.3 Características de la andragogía

1. **Tiene un auto concepto:** A diferencia de los niños los adultos cuentan con una necesidad para ser auto dirigidos, al tener mayor madurez le permite saber que él es responsable de su crecimiento y de su vida.
2. **Tiene experiencia:** Los adultos tienen conocimientos, así como experiencias acumuladas por el tiempo y eventos vividos.
3. **Prisa en aprender:** Los adultos están dispuestos a aprender cosas que necesitan saber o saber hacer, para así cumplir con los papeles en la sociedad.
4. **Orientación para el aprendizaje:** Los adultos tienen mayor tendencia a mantener el enfoque en situaciones, problemas, decisiones y mejoras permanentes buscando desarrollar las habilidades que necesitan aplicar a situaciones o problemas a los que se enfrentan en la vida real.
5. **Motivación para aprender:** La motivación viene de la propia persona, para algunos puede ser dinero, poder, autoridad o para otros el éxito, voluntad, valor y disfrute (La teoría del aprendizaje adulto de Malcolm Knowles, 2020).

### 1.3 Educación antes de la pandemia

Durante mucho tiempo, profesores como estudiantes desempeñaron un papel educativo predecible, sumergidos en una rutina que siempre parecía ser la misma; en la cual el estudiante acudía a un aula, el docente impartía sus clases, el alumno tomaba notas, planteaba preguntas, compartía con sus compañeros, salía del aula, realizaba tareas, y este ciclo se repetía siendo así el sistema educativo en el cual niños, adolescentes y adultos habían participado como parte de su formación académica.

Antes del inicio de la pandemia, la mayoría de los estudiantes solían experimentar un acceso similar al aprendizaje al asistir a la escuela para sus clases. Aunque las diferencias económicas y sociales siempre han estado presentes en todo el mundo, la mayoría de los alumnos que asistían regularmente a clases de manera presencial tenían la oportunidad de recibir sus clases como habitualmente se hace y poder continuar con sus estudios. Sin embargo, con la llegada de la contingencia por COVID-19, las diferencias económicas se hicieron más evidentes, lo que dificultó que todos los estudiantes tuvieran igualdad de oportunidades para el aprendizaje a larga distancia (Padilla & Abreu, 2021).

### 1.4 Educación durante la pandemia

Durante el inicio de la pandemia, los 36.5 millones de mexicanos que estaban cursando el ciclo escolar 2019-2020 se vieron enfrentados a la necesidad de tomar clases desde sus casas. La SEP implementó una estrategia la cual se basó en utilizar la televisión como medio de transmisión en donde los alumnos sintonizaron programas específicos para su nivel escolar a través de este o internet, y estos contenidos se complementan con actividades disponibles en línea o distribuidas en los hogares.

Con el fin de ajustarse de manera adecuada las escuelas incorporaron plataformas tecnológicas para facilitar la enseñanza y evaluación de los alumnos. Las familias también hicieron esfuerzos económicos ya que aproximadamente el 43% de los hogares adquiriendo dispositivos y el 26% contratando servicios de internet (García Dobarganes, 2021).

## 1.5 Educación después de la pandemia

Catorce meses después del cierre total de las escuelas, la Secretaría de Educación Pública anunció el reinicio de las clases presenciales para los primeros días de junio de 2021 (Torres Ramírez, 2021)

El secretario de Educación Pública Esteban Moctezuma Barragán anunció que la Estrategia de Regreso a Clases de manera gradual se implementaría a partir del 10 de agosto en todas las escuelas y niveles del Sistema Educativo Nacional, siempre y cuando el semáforo de riesgo se encontrara en verde. Algunas medidas que se tomaron para reducir el riesgo de contagio fueron:

- Garantizar el acceso a jabón y agua.
- Cuidado de docentes en grupos de riesgo.
- Uso obligatorio de cubrebocas.
- Maximizar el uso de espacios al aire libre.
- Suspensión de ceremonias y reuniones.
- Detección temprana de síntomas.
- Proporcionar apoyo emocional tanto a docentes como a estudiantes. (IESPE, 2020, párrafo séptimo).

## 1.6 Cuestionarios de seguimiento a la educación

Diferentes organizaciones dieron seguimiento a instituciones de diferentes niveles educativos sobre el impacto de la pandemia en su educación, de los cuales aplicaron test a los alumnos

En el artículo, “La educación superior en tiempos de pandemia: una visión desde dentro del proceso formativo”, escrito por José Antonio Miguel Román en el año 2020 en la Universidad José Vasconcelos de Oaxaca, México, donde se da a conocer un test que fue aplicado a estudiantes, docentes y administrativos participantes que se encontraban adscritos a IES de los Valles Centrales de Oaxaca (Miguel Román, 2020), en cuanto a las respuestas por parte de los maestros se visualiza lo siguiente:

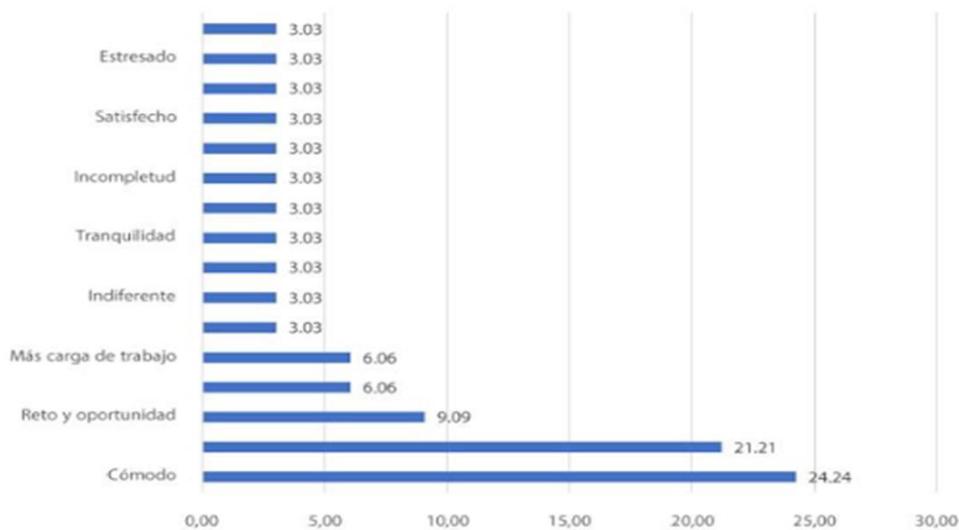


Figura 1 Sentir de los docentes ante los cambios que implica el ajuste de clases presenciales a virtuales (a) (Miguel, 2020).

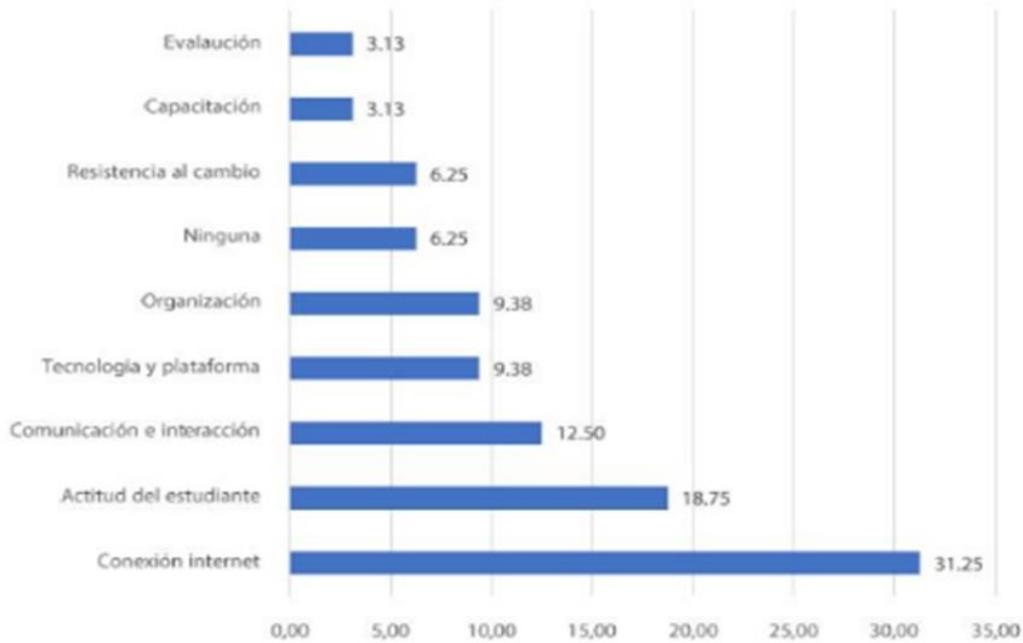


Figura 2 Obstáculo más significativo de los docentes ante los cambios que implica el ajuste de clases presenciales a virtuales (b) (Miguel, 2020).

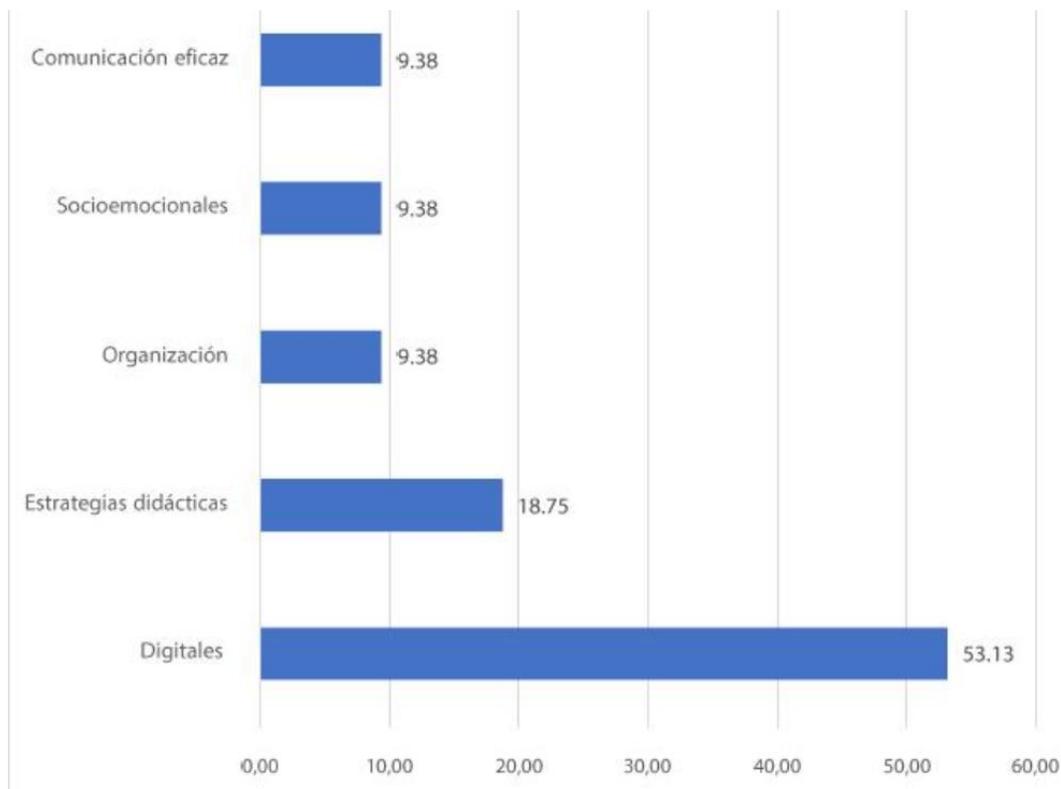


Figura 3 Competencias desarrolladas por los docentes ante los cambios que implica el ajuste de clases presenciales a virtuales (c) (Miguel, 2020).

*Análisis de los aspectos académicos en la educación superior ocasionados por la pandemia mediante técnicas descriptivas de minería de datos*

En este proyecto, las preguntas van orientadas para los alumnos, en donde:

Las preguntas que utilizaron en este cuestionario, fueron las siguientes:

1. *¿Cómo te sientes con el ajuste de clases presenciales a virtuales debido a la contingencia sanitaria COVID-19? (ver Figura 4).*

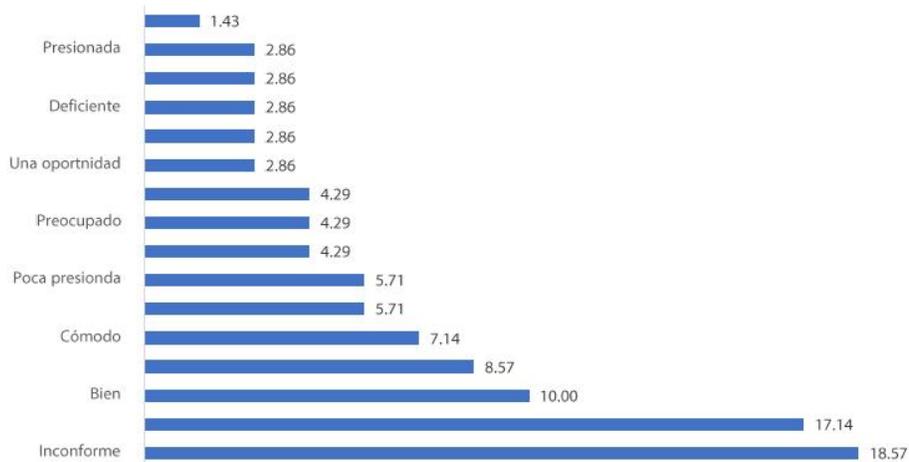


Figura 4 *¿Cómo te sientes con el ajuste de clases presenciales a virtuales debido a la contingencia sanitaria COVID-19? (Miguel, 2020).*

Se puede observar que en las respuestas fueron que no se encuentran conformes o satisfechos, en donde:

- Inconformidad 18.57%
- Estresado 17.14%
- Cómodo 10%

2. ¿Cuál consideras que ha sido el obstáculo más significativo de este ajuste de clases presenciales a virtuales debido a la contingencia sanitaria: COVID-19? (ver Figura 5).

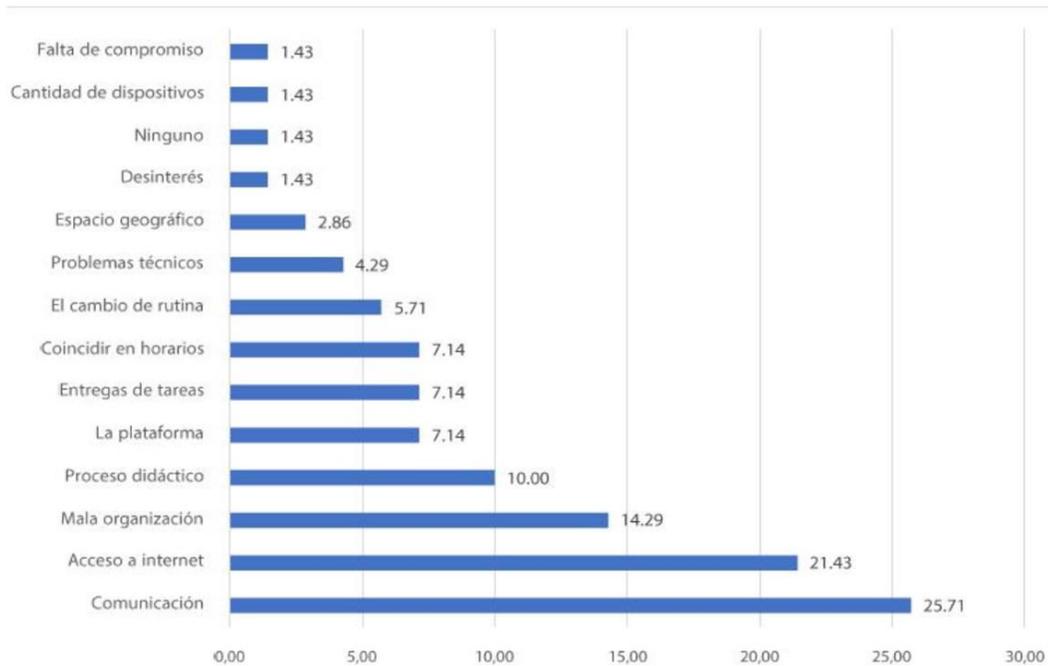


Figura 5 ¿Cuál consideras que ha sido el obstáculo más significativo de este ajuste de clases presenciales a virtuales debido a la contingencia sanitaria: COVID-19? (Miguel, 2020).

Las tres respuestas con mayor ponderación fueron:

- Comunicación 25.71%
- Acceso a internet 21.43%
- Mala organización 14.49%

Para finalizar, la última pregunta que se planteó a continuación fue:

3. *¿Cuáles son las competencias que has tenido que desarrollar para solucionar los obstáculos del presente contexto (ajuste de clases presenciales a virtuales debido a la contingencia sanitaria: COVID-19) y adaptarte? (ver Figura 6).*

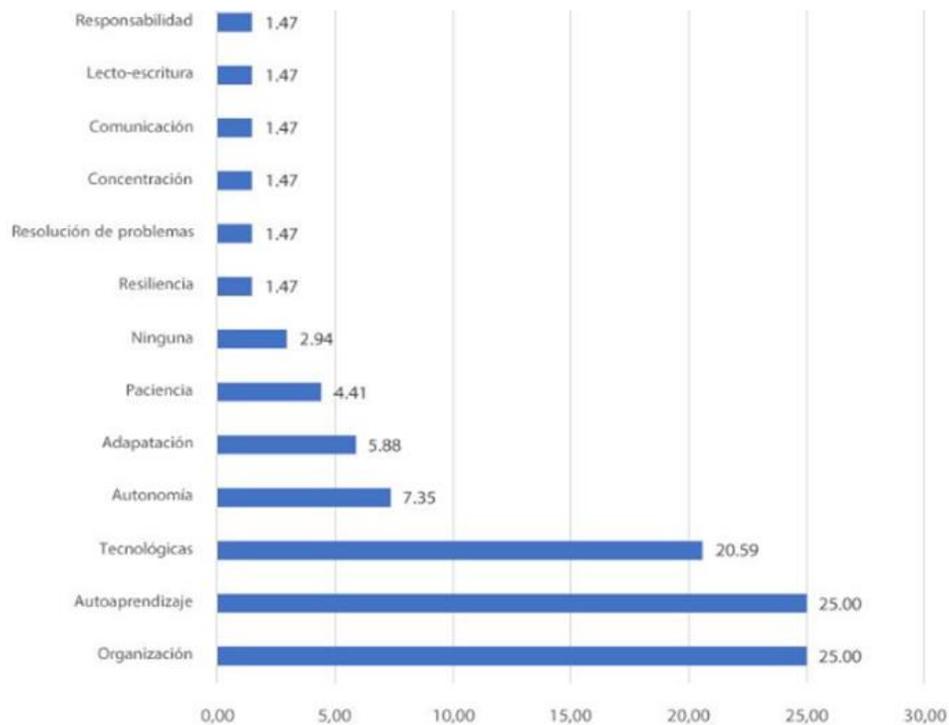


Figura 6 *¿Cuáles son las competencias que has tenido que desarrollar para solucionar los obstáculos del presente contexto y adaptarte? (Miguel, 2020)*

Las respuestas con mayor ponderación para los estudiantes fueron las siguientes:

- Organización 25%
- Autoaprendizaje 25%
- Tecnologías 20.59%

Concluyendo que “es preocupante el hecho de que tanto estudiantes como docentes advierten que es necesario ser autónomos, aprender de manera independiente, además de que exijan de ellos tener competencias socioemocionales.” (Miguel Román, 2020).

Otro formulario aplicado por el equipo técnico del Instituto Internacional de la UNESCO para la Educación Superior (IESALC) en el cual examina cuáles están siendo los impactos inmediatos de la pandemia en el sector de la educación superior universitaria. En este test hace una comparativa entre las cátedras iberoamericanas y las del resto del mundo. La pregunta más relevante que se ha tomado en cuenta para este proyecto es:

¿Cuáles son las principales dificultades de los estudiantes de educación superior durante la pandemia según los profesores? (ver Figura 7)

Teniendo como respuesta de manera global, las principales preocupaciones son:

- Aislamiento social.
- Preocupaciones económicas.
- Conectividad a internet
- Ansiedad general con respecto al COVID-19.

Mientras que en Iberoamérica el orden de las preocupaciones es diferente:

- Conectividad a internet.
- Preocupaciones económicas.
- Dificultades para mantener un horario regular.

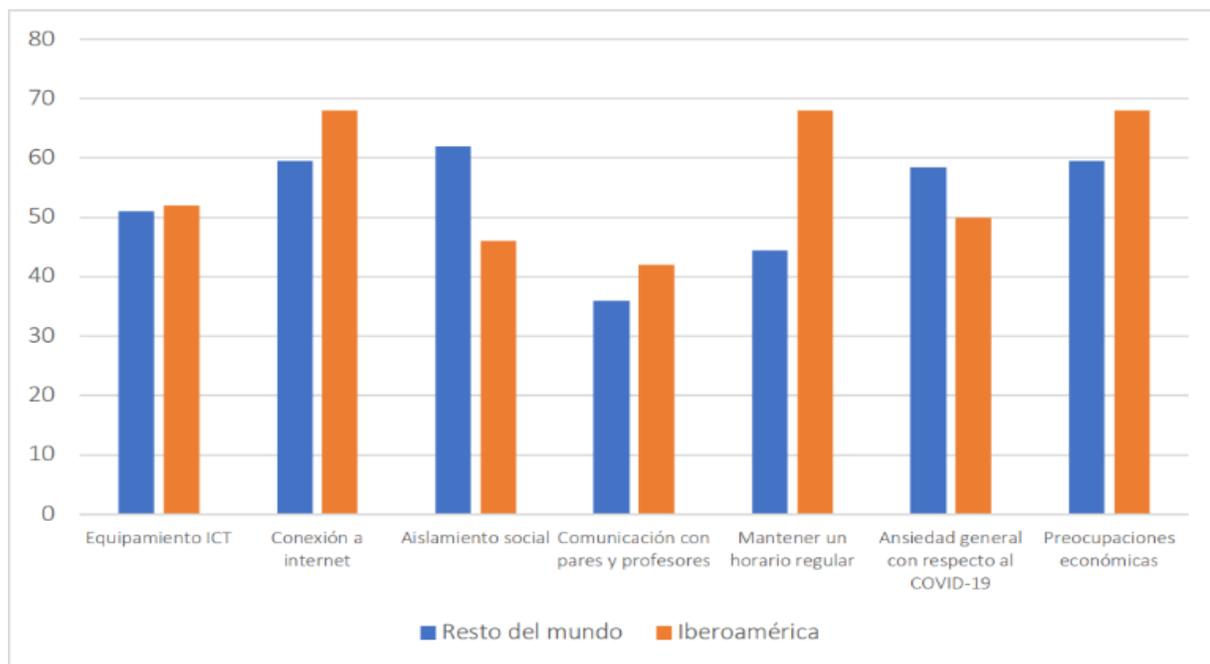


Figura 7 Dificultades de los estudiantes (IESALC, 2020)

Para finalizar han concluyeron que “la situación generada es particularmente preocupante respecto de los estudiantes más vulnerables que ingresaron a la educación superior en condiciones más frágiles, puede convertir esa fragilidad en abandono” (IESALC, 2020).

Estas encuestas fueron tomadas en cuenta para la elaboración del formulario que se ha aplicado a los alumnos de la licenciatura en Ciencias Computacionales, considerando las respuestas con mayor puntualización y relevancia.

## 1.7 Aprendizaje automatizado

El aprendizaje automático permite utilizar los datos para enseñarle a un sistema con ayuda de algoritmos la resolución de problemas y cómo perfeccionarse con el transcurso del tiempo. Estos algoritmos identifican patrones en los datos y esos patrones se utilizan para crear modelos que pueden hacer predicciones.

El uso del aprendizaje automático se ha destacado como una herramienta altamente eficaz en el procesamiento y análisis de grandes cantidades de datos. En donde la disponibilidad de datos ha tenido un gran crecimiento, las técnicas de aprendizaje automático se vuelven indispensables para la extracción de información valiosa y relevante. Estos algoritmos tienen la capacidad de analizar y aprender de los datos, lo que capacita a las empresas para tomar decisiones estratégicas y fundamentadas con mayor precisión (Descifrando la importancia del aprendizaje automático en la tecnología moderna, 2023).

### 1.7.1 Usos del aprendizaje automático

Los usos del aprendizaje automático se han expandido de manera impresionante en diferentes áreas como la salud, comercio, la industria, seguridad, entre otros. Esta tecnología está transformando la forma en que se realizan tareas y toma de decisiones de dichos campos, proporcionando nuevas oportunidades para la innovación y la mejora continua.

1. **Predecir valores:** Los algoritmos de regresión ayudan a reconocer la causa y efecto de las variables, generan un modelo basado en estos datos, el cual se implementará para realizar predicciones.

2. **Detecta repeticiones inusuales:** Los algoritmos de detección de anomalías, usualmente aplicados para identificar posibles situaciones riesgosas, reconocen datos que se desvían de las expectativas habituales.
3. **Predice categorías:** Los algoritmos de categorización contribuyen a establecer la categoría precisa para los datos, aunque es similar a la agrupación en clústeres se distingue por su aplicación en el aprendizaje supervisado, en el cual se asignan etiquetas predefinidas (Aprendizaje automático, 2017).

## 1.8 Minería de datos

Es la actividad de obtener información relevante de un conjunto de datos recopilados, los cuales suelen estar almacenados en una base de datos. Para llevar a cabo esta tarea, se utilizan herramientas de minería de datos con capacidades estadísticas, matemáticas y analíticas de gran alcance, con el objetivo de analizar grandes conjuntos de datos para descubrir tendencias, patrones y relaciones que puedan respaldar la toma de decisiones informadas y la planificación estratégica (SAP, 2023).

### 1.8.1 Beneficios de la minería de datos

- **Encontrar fácilmente los datos esenciales.** Aunque un conjunto grande de datos tiene información valiosa, también está repleto de datos que no son necesarios. La minería de datos hace más fácil identificar automáticamente la información valiosa y luego presentarla en informes prácticos.
- **Comprender al cliente:** La minería de datos permite la recolección de datos de clientes que provienen de diferentes fuentes, lo que permite la creación de perfiles detallados y abundantes en información. Esto proporciona conocimiento acerca de las tendencias, preferencias, comportamientos, similitudes y diferencias entre los consumidores. Esta valiosa información se

convierte en un recurso esencial para mejorar la experiencia del cliente y optimizar la comunicación en todos los puntos de contacto.

- **Tomar decisiones más rápidas y automatizadas:** En lugar de depender de una revisión manual de datos y la toma de decisiones por parte de una persona, es posible automatizar algunas.
- **Desarrollar campañas de marketing:** La minería de datos permite la creación de campañas de marketing eficientes y adaptadas a nivel personal. Aprovechando la información recopilada, de esta manera el marketing puede realizar ajustes con mayor precisión a las preferencias y conductas de los consumidores, personalizando el contenido como las recomendaciones de productos (María, 2022).

## 1.9 Aprendizaje supervisado

El aprendizaje supervisado, es un algoritmo Machine Learning más utilizado, consiste en utilizar datos conocidos (datos de entrenamiento), estos datos contienen información de entrada que está etiquetada y se relaciona con los valores de salida deseados a través del uso de este conjunto de datos, el algoritmo de aprendizaje supervisado construye un modelo que establezca conexiones entre las características y los datos de salida. Este modelo luego se emplea para predecir los valores de respuesta en un conjunto de datos nuevo (¿Qué es el aprendizaje supervisado?, 2017)

### 1.9.1 Naive Bayes

Naive Bayes es un método de aprendizaje automático probabilístico empleado en la clasificación de datos, es decir, en predecir a qué categoría pertenece un conjunto de datos. Este se basa en el Teorema de Bayes, el cual calcula probabilidades considerando información previa. Aunque es un modelo simple, Naive Bayes puede generar modelos muy precisos, especialmente en situaciones donde se manejan datos categóricos.

Aplicaciones de Naive Bayes:

- Filtrado de Spam
- Clasificación de documentos
- Análisis de sentimientos
- Diagnostico medico (Pompas Gutiérrez, 2024).

### 1.9.2 Árbol de decisión

Un árbol de decisión, es un algoritmo que se utiliza para tareas de clasificación. Lo cual tiene una estructura de árbol jerárquico, donde se encuentra un nodo raíz, ramas, nodos internos y nodos hoja (¿Qué es un árbol de decisión?, 2022)

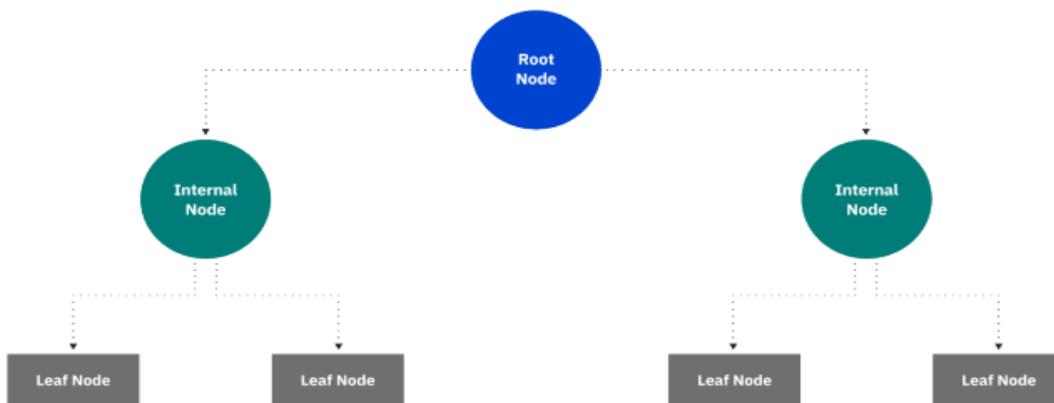


Figura 5 Árbol de decisión (¿Qué es un árbol de decisión?, 2022.)

Como se observa en la figura 5, un árbol de decisión se inicia con un nodo principal, el cual no tiene ramificaciones que lleguen hacia él. Las ramificaciones que se desprenden de este nodo principal conducen a los nodos internos, que también se llaman nodos de decisión. Tanto los nodos principales como los internos realizan evaluaciones para formar subconjuntos, que se indican mediante nodos hoja o nodos terminales. Estos últimos representan todas las posibles conclusiones dentro del conjunto de información (¿Qué es un árbol de decisión?, 2022)

## 1.10 Aprendizaje no supervisado

El término “no supervisado” se refiere a que el algoritmo no está guiado como en el aprendizaje supervisado y se utiliza para procesar datos que no se encuentran estructurados de acuerdo a sus similitudes y patrones diferentes en un conjunto de datos.

Algunos de los usos del aprendizaje no supervisado, son:

- El aprendizaje no supervisado identifica patrones ocultos en conjuntos de datos.
- Los métodos sin supervisión permiten la detección de atributos que podrían resultar valiosos para la clasificación.
- La obtención de datos sin etiquetar es una tarea menos complicada en comparación con la adquisición de datos que ya cuentan con etiquetas (González, 2022).

### 1.10.1 Aplicaciones del aprendizaje no supervisado

El aprendizaje no supervisado se caracteriza por utilizar conjuntos de datos pequeños y sin etiquetas para generar resultados. En este proceso, la máquina establece relaciones entre los elementos sin depender de etiquetas predefinidas, el aprendizaje no supervisado se centra en explorar cómo están organizados los datos para encontrar patrones y conexiones ocultas. Este enfoque tiene diversas aplicaciones, entre las cuales González (2022), concluye en:

- La técnica de agrupación en clústeres realiza una partición de los datos en grupos basándose en sus semejanzas.
- La detección de anomalías permite identificar elementos inusuales dentro de su conjunto de datos y resulta útil para descubrir transacciones fraudulentas.
- Los modelos de variables son empleados en el pre- procesamiento de datos, ya sea para reducir la cantidad de características en un conjunto de datos o para descomponerlo en múltiples componentes (González, 2022).

### 1.10.2 K- means

Es un método de agrupamiento de clasificación no supervisada que consiste en la segmentación un conjunto de datos en k grupos o clústeres, el proceso de agrupación se ejecuta de manera que los elementos dentro de un mismo clúster sean más similares entre sí que con los elementos pertenecientes a otros clústeres.

En el contexto del clustering, el concepto de distancia se utiliza para medir la similitud o disparidad entre observaciones. Cuando representamos las observaciones en un espacio de  $p$  dimensiones, donde  $p$  es el número de variables relacionadas con cada observación, las observaciones que son más similares se encontrarán más cercanas entre sí, lo que justifica el uso del término "distancia".

Una característica distintiva del clustering es su flexibilidad para utilizar diferentes tipos de medidas de distancia, lo que brinda al investigador la libertad de elegir aquella que mejor se adapte a los datos y al objetivo del análisis. Esto permite que el algoritmo de clustering se ajuste mejor a las características específicas del conjunto de datos y a las necesidades del análisis (Ramírez, 2023).

### **Distancia Euclídea:**

La distancia euclidiana entre dos puntos,  $p$  y  $q$ , se determina como la longitud de la línea recta que conecta estos puntos. En un sistema de coordenadas cartesianas, su cálculo se basa en la aplicación del teorema de Pitágoras (Giorgia & Cristina, 2020).

$$d(x,y) = \sqrt{\sum_{i=1}^n (y_i - x_i)^2}$$

*Ecuación 1 Distancia Euclídea*

### **Distancia de Manhattan**

Esta métrica también es conocida como métrica *de taxicab*, *distancia rectilínea* o *distancia L1*. Es una medida que define la distancia entre dos puntos,  $p$  y  $q$ , como la suma de las diferencias absolutas entre cada dimensión de esos puntos.

Esta métrica es ampliamente utilizada en diversos campos debido a su simplicidad y aplicabilidad.

$$d(x, y) = \left( \sum_{i=1}^m |x_i - y_i| \right)$$

*Ecuación 2 Distancia de Manhattan*

El proceso de agrupamiento del algoritmo k-means se realiza de la siguiente manera:

1. **Inicialización:** Se eligen aleatoriamente k centroides iniciales, uno para cada clúster. El centroide es simplemente un punto que representa el centro del clúster.
2. **Asignación de puntos:** Cada punto de datos se asigna al clúster cuyo centroide esté más cerca en función de una medida de distancia, generalmente la distancia euclidiana.
3. **Actualización de centroides:** Una vez que todos los puntos han sido asignados a clústeres, los centroides se recalculan tomando el promedio de las coordenadas de todos los puntos en ese clúster.
4. **Reasignación de puntos:** Los puntos se reasignan nuevamente a los clústeres según la distancia a los nuevos centroides.
5. **Los pasos 3 y 4:** Se repiten hasta que los centroides de los clústeres dejen de cambiar o hasta que llegue un alcance límite de iteraciones.

### 1.10.3 Clustering jerárquico

El Clustering Jerárquico es un método de minería de datos que tiene como finalidad la agrupación de datos llamados clústers. El procedimiento de este algoritmo de clúster es que agrupa los datos basándose en la distancia entre cada uno de ellos, con el fin de lograr que aquellos elementos situados en un mismo clúster presenten similitudes entre sí.

A diferencia del algoritmo k-means que divide los datos en un número fijo de clusters, el clustering jerárquico construye una estructura de árbol que muestra cómo los datos se agrupan a diferentes niveles de similitud.

El proceso de clustering jerárquico se lleva a cabo de la siguiente manera:

1. **Iniciación:** Cada dato se inicia como su propio clúster individual.
2. **Combinación:** En cada paso, los dos clústeres más cercanos (según una métrica de distancia, como la distancia euclidiana) se combinan en un nuevo clúster.
3. **Repeticiones:** El paso 2 se repite hasta que todos los datos estén en un solo clúster o hasta que se alcance un criterio de parada predefinido.

## 1.11 Coeficiente alfa de Cronbach

El coeficiente de confiabilidad de Cronbach fue creado por Lee Cronbach en el año 1951. Es una métrica estadística que se suele utilizar como una medida de consistencia interna o confiabilidad de un instrumento psicométrico principalmente cuando se emplea las escalas de Likert, evalúa que tan bien un conjunto de variables o ítems mide un aspecto que es único a través de la aplicación de un cuestionario (Ruiz Mitjana, 2019).

$$\alpha = \frac{N * \bar{c}}{\bar{v} + (N - 1) * \bar{c}}$$

*Ecuación 3 Coeficiente alfa de Cronbach*

Donde:

N = el número de elementos.

$\bar{c}$  = covarianza promedio entre pares de ítems.

$\bar{v}$  = varianza promedio.

**Interpretación:**

### Rangos del Alfa de Cronbach

Alfa de Cronbach	Consistencia Interna
$\alpha \geq 0,9$	Excelente
$0,8 \leq \alpha < 0,9$	Buena
$0,7 \leq \alpha < 0,8$	Aceptable
$0,6 \leq \alpha < 0,7$	Cuestionable
$0,5 \leq \alpha < 0,6$	Pobre
$\alpha < 0,5$	Inaceptable

*Figura 8 Alfa de Cronbach (Pérez León, 2022.)*

## 1.12 Índices de validez

La validación de un proceso de agrupación es importante para garantizar el éxito de las aplicaciones de agrupamiento. El resultado de un método de agrupamiento puede cambiar de acuerdo a sus parámetros, se realizaron enfoques denominados índices de validez que evalúan la calidad de las agrupaciones generadas.

La validación de agrupamientos se puede dividir principalmente en dos categorías: interna y externa. La diferencia entre la validación interna y externa es si se incorpora o no información externa durante el proceso de validación. Las técnicas de validación externa comparan una partición con la partición de referencia correcta, mientras que las técnicas de validación interna evalúan una partición únicamente a partir de los datos sin hacer referencia a información externa.

- **Compactación:** Mide la relación entre los elementos dentro de un grupo. Una medida frecuente de la compactación es la varianza, siendo una varianza baja indicativa de una compactación más sólida.
- **Separación:** Mide lo diferente o que tan bien está separado un grupo con respecto a otros grupos (Méndez Álvarez, 2019).

### 1.12.1 Calinski harabasz

También llamado criterio de relación de varianza, fue propuesto por Calinski y Harabasz en 1974 como una herramienta para evaluar modelos de agrupamiento. Este índice se utiliza para medir qué tan bien se ha llevado a cabo la agrupación, este índice mide la similitud de un objeto con su propio grupo en comparación con otros grupos (Rudeus, 2022).

El índice Calinski-Harabasz se basa en la comparación de la relación ponderada entre la suma de los cuadrados, que mide la separación entre clústeres, y la suma de los cuadrados dentro del clúster, que mide cuán cercanos están los puntos dentro de un clúster. (Herramienta de diagnóstico de K-centroides, 2019).

#### Cálculo del índice Calinski-Harabasz:

El índice CH para el número K de grupos en un conjunto de datos  $D = [d_1, d_2, d_3, \dots, d_N]$  (Dey, 2022).

$$CH = \left[ \frac{\sum_{k=1}^K n_k \|c_k - c\|^2}{K - 1} \right] / \left[ \frac{\sum_{k=1}^K \sum_{i=1}^{n_k} \|d_i - c_k\|^2}{N - K} \right]$$

Ecuación 4 Índice Calinski-Harabasz

Donde:

$n_k$  y  $c_k$  son el no. de puntos y centroide del k-ésimo grupo respectivamente,  $c$  es el centroide global,  $N$  es el número total de puntos de datos.

Un valor más alto del índice CH significa que los grupos son densos y se encuentran bien separados, aunque no hay un valor que se considere aceptable. Se necesita elegir la solución que dé un pico en el gráfico lineal de los índices CH (Dey, 2022).

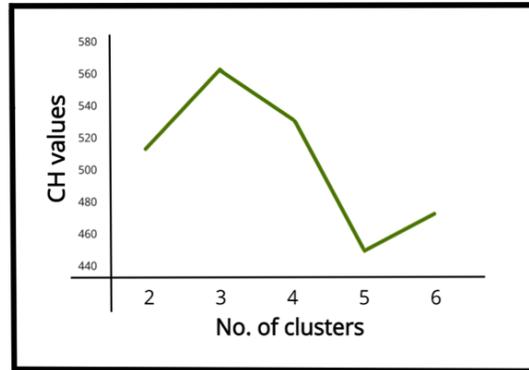


Figura 9 Calinski-Harabasz (Dey, 2022)

### 1.12.2 Davies Bouldin

Desarrollado por David L. Davies y Donald W. Bouldin en el año 1979, el índice de Davies-Bouldin (DBI) se utiliza como una medida para analizar la eficacia de los grupos generados por un algoritmo de clustering.

El índice de Davies-Bouldin se basa en la relación entre la dispersión dentro de los clústeres (intra-clúster) en donde se evalúa la proximidad de los puntos dentro de cada grupo, donde una dispersión intra-clúster baja quiere decir que los elementos en un grupo están muy cercanos entre sí, lo cual es un buen resultado en un clustering efectivo, también se tiene la separación entre los clústeres (inter-clúster) en el cual se mide la distancia entre los diferentes grupos, si hay una dispersión alta indica que los clústeres están muy separados, lo cual es bueno en un agrupamiento.

### 1.12.3 Silhouette

Este índice es una métrica utilizada para evaluar la consistencia en agrupamientos, lo cual ayuda a verificar que los grupos son tanto compactos como separados correctamente, considerando tanto la cohesión como la separación (Prado, 2022).

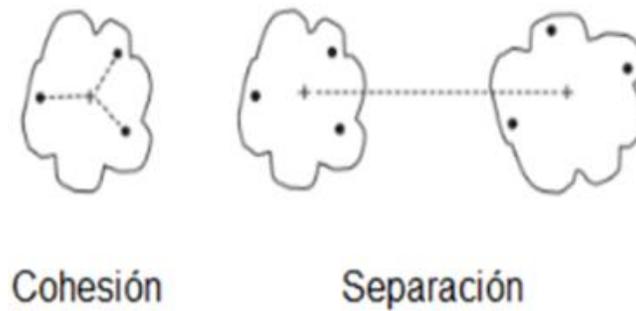


Figura 10 Silhouette (Prado, 2022)

La cohesión implica que todos los elementos dentro de un grupo deben estar ubicados cerca unos de otros, mientras que el de separación los grupos tienen que mantener una distancia significativa entre sí.

$$S(i) = \frac{b(i) - a(i)}{a(i), b(i)}$$

Ecuación 6 Silhouette

Donde:

a(i): Del punto objeto  $i$  con todos los demás puntos en el mismo clúster.

b(i): Del objeto  $i$  hacia todos los objetos en otro clúster más cercano.

El índice Silhouette puede tener valores de entre  $-1$  y  $1$ . Cuando este índice se acerca a  $1$ , quiere decir que el punto en cuestión está asignado a su grupo correspondiente. Un valor que está cercano a  $0$  sugiere que el punto podría ser parte de otro clúster, y cuando el valor se aproxima a  $-1$ , significa que el punto está mal clasificado y se encuentra en algún punto intermedio entre los grupos (Prado, 2022).

### 1.13 Escala de Likert

La Escala de Likert se emplea como un sistema de calificación para saber la opinión de un individuo respecto a su grado de conformidad o desacuerdo con una declaración. Esta herramienta resulta útil para evaluar las respuestas, actitudes y conductas de una persona.

Las escalas de frecuencia que adoptan el formato de Likert emplean conjuntos predefinidos de respuestas para evaluar actitudes y opiniones. Estas escalas posibilitan la identificación del nivel de acuerdo o desacuerdo entre los participantes encuestados.

Las respuestas tienen la posibilidad de presentarse en distintos niveles de medición, lo que posibilita realizar escalas de 5, 7 o 9 elementos. Es importante incluir un punto neutro para aquellos usuarios que no se posicionan ni a favor ni en contra.



Figura 11 Escala de Likert (Muguira, 2020).

### 1.13.1 Ventajas

- Una de sus principales características es su facilidad de uso y estructura.
- Crea mediciones de calidad y precisión que reducen el margen de errores en la medición.
- Facilita el análisis posibilitando cumplir con los objetivos de la investigación.
- Permite comparar con otras evaluaciones ya aplicadas o similares a esta.
- Es fácil de contestar

### 1.13.2 Desventajas

- Hay investigaciones científicas que señalan la presencia de un sesgo en esta escala, ya que las respuestas positivas tienden a ser más comunes que las negativas de manera constante.
- Los encuestados tienden a responder "de acuerdo", ya que esto implica un menor esfuerzo al contestar la encuesta.
- Resulta difícil determinar con exactitud la cantidad de respuestas tanto positivas como negativas (Muguira, 2021).

## 1.14 WEKA

Weka (Waikato Environment for Knowledge Analysis, en español entorno para análisis del conocimiento de la Universidad de Waikato) es un software de libre distribución desarrollado en JAVA, conformado por paquetes de código abierto con diferentes herramientas y algoritmos para la clasificación, agrupamiento, asociación entre otras, en una interfaz gráfica de fácil uso para su aplicación y análisis de datos (Gutiérrez, Albarran Fernández, & Rico Molina, 2017).

### 1.14.1 Breve historia

Weka fue desarrollado por la Universidad de Waikato de Nueva Zelanda en el año de 1993 utilizando TCL/TK y C fue diseñada principalmente para realizar análisis de datos provenientes de la agricultura. En 1997 fue reescrito en lenguaje JAVA además de implementar algoritmos de modelado, esta versión más reciente se ha utilizado en distintas áreas, particularmente en la docencia y la investigación.

### 1.14.2 Características

Weka cuenta con las siguientes características:

- Es accesible bajo los términos de licencia pública general de GNU.
- Portabilidad ya que está implementado en Java y permitiendo correr en casi cualquier plataforma.
- Tiene una gran colección de técnicas para pre procesamiento de datos y modelado.
- Fácil de utilizar por principiantes gracias a su interfaz gráfica de usuario intuitiva (Rico et al., 2019).

## CAPITULO 2 ESTADO DEL ARTE

### 2.1 El uso de mecanismos de innovación en la educación de adultos

El estudio se concentró en revisar las políticas e investigaciones alrededor de la educación de adultos para reconocer la forma como se podía potenciar el proceso, teniendo como objetivo identificar la incidencia en el proceso educativo de un grupo de adultos al implementar un mecanismo de innovación. La metodología de esta investigación tiene un enfoque cualitativo, orientada bajo el paradigma hermenéutico.

El trabajo que se desarrolló estuvo orientado a la Investigación – Acción, es decir identificar un problema y proponer una solución. Se trabajó en 4 ciclos que son: detectar el problema, elaborar el plan, implementar y evaluar el plan y retroalimentación. Dicha investigación se aplicó a 14 alumnos adultos del programa de educación secundaria y media con edades de entre 24 y 60 años.

Luego de que implementaron el mecanismo de “Design Thinking”, llegaron a la conclusión de que este permitía desarrollar y favorecer actividades en el salón de clase, en especial el trabajo por proyectos, debido que lograba que los estudiantes pudieran reconocer problemas reales y plasmarlas en acciones. (Cardozo & Herrera, 2019).

Después de analizar la investigación se concluyó que posee una relevancia significativa relacionar lo que se expone en clases con el entorno laboral y personal

de la misma manera es trascendental que los alumnos se sientan partícipes de su educación.

## **2.2. Retos de la educación virtual en el proceso enseñanza aprendizaje durante la pandemia de COVID-19**

Realizaron búsquedas bibliográficas en PubMed, Elsevier, Google Académico, Scielo, Scopus, Web of Science y Páginas Web seguras hasta septiembre del 2020 para identificar estudios que documenten los retos que la educación virtual, igualmente utilizaron los términos solos o combinados con los operadores booleanos.

Se centraron en la búsqueda en artículos de texto, considerando los resúmenes relevantes en la redacción del artículo. Se identificaron estudios adicionales a través de la evaluación precisa de la lista de referencias de los trabajos incluidos.

Determinaron que a pesar de los problemas que se puedan generar en clases virtuales, la educación en línea promete una enseñanza de calidad para todo tipo de estudiantes, minimizando sesgos y trabajando en conjunto con padres de familias e instituciones, aprovechando todas las herramientas virtuales que existen (Valero et al., 2020)

La reflexión final es que nadie estaba preparado para enfrentar una pandemia y contar con una estrategia estructurada correctamente para dar continuidad a la educación de millones de estudiantes, sin embargo, las herramientas tecnológicas fueron indispensables para dar un adecuado seguimiento.

## 2.3 Factores asociados al éxito de los estudiantes en modalidad de aprendizaje en línea un análisis en minería de datos

Desarrollada por Gabriela Mancilla Vela, Paola Leal Gatica, Aurora Ortiz y Cristian Vidal Silva del departamento de administración en la Universidad Católica del Norte, ubicada en Antofagasta, Chile. Con el fin de determinar las variables asociadas al éxito de los estudiantes en programas con modalidad de aprendizaje en línea (e-learning). Utilizando la metodología CRISP-DM consistiendo en 6 fases en donde las tareas generales se proyectan como específicas del mismo modo la utilización de técnicas y algoritmos de minería de datos como árbol de decisión, estadísticas descriptivas y redes neuronales. Apoyándose de la herramienta SPSS Statistics 22 (Ball-Rokeach y Hoyt, 2001).

Este estudio fue realizado aplicado a los programas de e-learning impartidos por el Centro de Educación a Distancia de la Universidad Católica del Norte (CED-UCN) en Chile. La muestra utilizada fue de 18.610 sujetos participantes en dichos programas durante 19 años. Los datos en bruto de todos los estudiantes y los datos históricos se localizan en un servidor local (LICANCABUR) con acceso restringido por personal autorizado mediante autenticación de usuario y contraseña. Este servidor da servicios al sistema de base de datos Oracle Developer 2000 de nombre ANTEC, base de datos oficial del CED-UCN. Otros datos complementarios están almacenados en archivos e impresos manejados por los encargados de las distintas áreas.

Como conclusión la utilización de la minería de datos educacional y de la metodología CRISP-DM es un gran aporte a la sistematización y eficiencia en la identificación de patrones en los datos en la educación a distancia. Sus resultados

creen que ayudarán a la admisión, así como de retención de estudiantes potencialmente más expuestos al abandono (Mancilla et al., 2020).

La percepción obtenida ha sido que gracias a la minería de datos y utilizando diferentes algoritmos han podido determinar patrones de comportamiento que llevaron a los estudiantes al éxito en modalidad en línea al igual que los fracasos.

## **2.4 Relación entre el rendimiento escolar y la ansiedad en estudiantes de la licenciatura en ciencias de la educación de la UAEH durante la pandemia, semestre enero-junio 2022**

Tesis fue realizada en la Universidad Autónoma del Estado de Hidalgo, la cual tuvo como objetivo poder identificar la relación entre la ansiedad por modalidad virtual a causa de la pandemia de COVID-19 y el rendimiento académico de los alumnos, utilizando como metodología un enfoque cuantitativo de tipo exploratorio descriptivo no experimental de corte transversal.

En esta investigación aplicaron una prueba piloto para después realizar 3 cuestionarios con ayuda de la herramienta de Google Forms enviándoselo a sus correos electrónicos, primero a 123 estudiantes para posteriormente 18 más, además de 6 docentes pidiendo la autorización de la coordinadora de su programa educativo.

Obteniendo como resultado que el 40.14% de su población presentaron ansiedad o estrés académico, pero sin encontrar una relación entre un trastorno mental con la disminución al rendimiento académico.

Por ultimo esta investigación tiene similitud con el presente trabajo que pretende conocer el impacto de la pandemia en la educación, para conocer a lo que se enfrentaron miles de estudiantes, por otra parte, esta investigación ofreció una perspectiva de parte de los docentes que tuvieron que adaptarse a las nuevas tecnologías para poder dar continuidad a sus alumnos.

## **2.5 Ansiedad, estrés y percepción de riesgo por COVID-19 en estudiantes universitarios**

Trabajo realizado en el Instituto de Ciencias de la Salud en la UAEH, teniendo como objetivo analizar la relación entre la ansiedad, estrés y percepción por COVID-19 en los estudiantes en el 2021 la cual es una investigación de tipo no experimental, transversal de alcance correlacional.

Para la recolección de datos aplicaron un cuestionario en GoogleForms el cual se les hizo llegar por reuniones en la Plataforma Zoom y Google Meet, el cual esta conformaba conformado por el inventario de ansiedad de Beck, la escala de estrés percibido, en el que aplicaron el alfa de Cronbach obteniendo en los dos 0.83, en otro cuestionario de percepción de riesgos por covid 19 obtuvieron 0.597 a 0.52.

Logrando resultados tales como en la escala de ansiedad un nivel leve, del mismo modo que los alumnos frecuentemente se encuentran estresados, concluyendo que los aspectos negativos a disminuir son miedos, preocupaciones y presión escolar.

Sin duda no importa de qué escuela o instituto se provenga, los alumnos coinciden en que la educación en línea durante la pandemia fue un gran desafío para muchos donde algunos se quedaron a mitad del camino, los docentes desempeñaron un papel importante para poder dar continuidad, en el que ambas partes aprendieron a adaptarse durante la pandemia del COVID-19.

## CAPÍTULO 3 METODOLOGÍA

En esta investigación la selección de una adecuada metodología es fundamental para poder garantizar la obtención de resultados significativos para llegar a una correcta conclusión sobre estos. Es por ello que en este capítulo se explora las diferentes metodologías evaluando cuál es la que mejor se acopla para este proyecto.

### 3.1 KDD

La metodología KDD (Knowledge Discovery in Databases) es un proceso que se utiliza para el análisis en grandes volúmenes de datos, buscando identificar patrones y “convertirlos en conocimiento, filtrando y descartando aquellos hallazgos que no resulten de utilidad para los objetivos fijados”.

El proceso KDD consta de 5 etapas: (ver figura 9)

1. Etapa de selección: En esta se seleccionan los datos relevantes para el análisis. Se decide qué variables o atributos se utilizarán y se eligen los conjuntos de datos adecuados. Es importante elegir datos representativos y apropiados para el análisis, ya que esto sentará las bases para las etapas posteriores del proceso KDD
2. Etapa pre-procesamiento o limpieza: Se lleva a cabo una homogeneización de los registros, limpiando y normalizando para que no haya problemas con el procesamiento por parte de las herramientas de Big Data.
3. Etapa de transformación: En esta etapa, los datos pre procesados se transforman en una forma adecuada para la aplicación de técnicas de minería de datos. Esto puede incluir la reducción de la dimensionalidad, la normalización de los datos o la creación de nuevas variables.
4. Etapa Minería de datos: Aquí se aplican técnicas de minería de datos para descubrir patrones, relaciones o conocimientos interesantes en los datos.

*Análisis de los aspectos académicos en la educación superior ocasionados por la pandemia mediante técnicas descriptivas de minería de datos*

Esto puede implicar el uso de algoritmos de aprendizaje automático, técnicas estadísticas o métodos de visualización.

5. Interpretación y evaluación: Los resultados de la minería de datos se interpretan y evalúan para determinar su relevancia y utilidad. Se analizan los patrones descubiertos y se evalúa su impacto en el problema o el objetivo del análisis (Gómez, 2020).

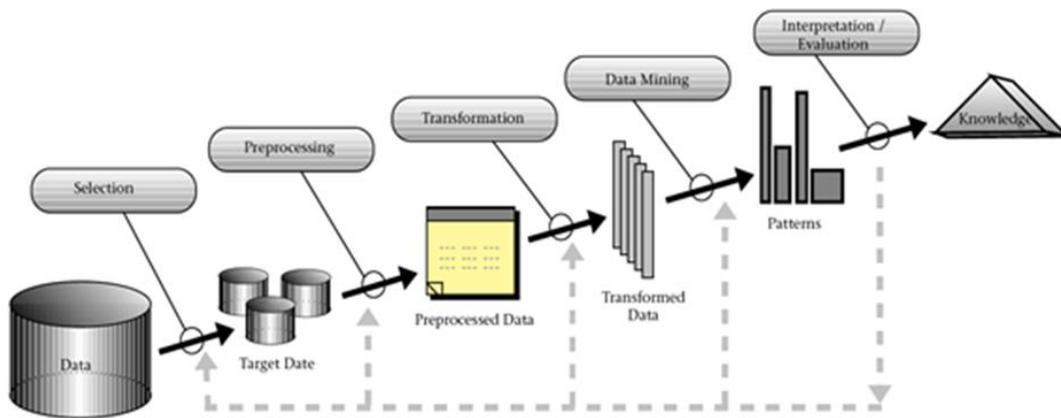


Figura 12 Diagrama KDD (Hitpass, 2014)

## 3.2 CRISP-DM

La metodología sigue una estructura estructurada en un proceso jerárquico que consta de tareas distribuidas en cuatro niveles distintos de abstracción, abarcando desde lo más general a lo más específico.

CRISP-DM, propone seis fases para el proceso de minería de datos. (ver figura 10)

1. **Comprensión del negocio:** Se establecen los objetivos y requisitos del proyecto, lo que implica la definición del problema en minería de datos y la planificación de las actividades a realizar.
2. **Comprensión de los datos:** Se lleva a cabo la recopilación de los datos que serán empleados en el proyecto. En esta etapa, es posible que surjan las primeras suposiciones o hipótesis sobre la información que podría estar oculta en los datos.
3. **Preparación de los datos:** Se realizan actividades de tratamiento de los datos para construir un conjunto de datos final sobre el cual se aplicarán las técnicas de minería de datos.
4. **Modelado:** Se emplean distintas técnicas y algoritmos de minería de datos en el conjunto de datos con el propósito de descubrir la información y los patrones que se encuentran en ellos.
5. **Evaluación:** Se analizan los patrones obtenidos de acuerdo con los objetivos establecidos en la etapa comprensión del negocio.
6. **Implementación:** consiste en la comunicación e implementación del nuevo conocimiento, el cual debe ser representado de forma entendible para el usuario (Gómez, 2020).

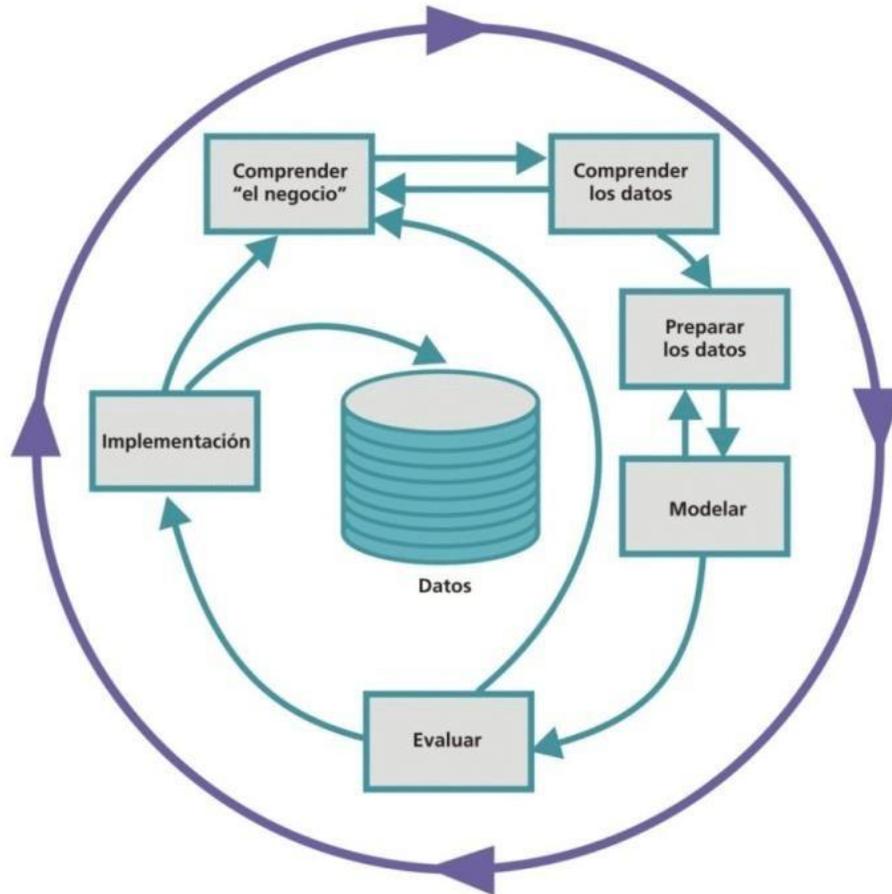


Figura 13 Diagrama CRISP-DM (Bellini Saibene, 2021)

### 3.3 SEMMA

Se centra principalmente en la creación del modelo de minería, excluyendo otros aspectos del proyecto como la comprensión del problema que se está investigando o la planificación de la implementación.

1. Muestreo: En esta etapa se toma una muestra del conjunto de datos disponible, esta muestra debe tener un tamaño adecuado para incluir la información importante, pero al mismo tiempo, debe ser lo bastante reducida como para permitir un procesamiento ágil.

*Análisis de los aspectos académicos en la educación superior ocasionados por la pandemia mediante técnicas descriptivas de minería de datos*

2. Exploración: Se exploran los datos buscando relaciones y tendencias al igual que formular nuevas hipótesis a partir de su análisis.
3. Modificación: Se preparan los datos, limpiando los valores anómalos y se seleccionan, crean y modifican las variables con las que se trabajarán.
4. Modelado: Se crea el modelo que permitirá predecir las variables de respuesta a partir de las variables explicativas, utilizando técnicas predictivas.
5. Evaluación: En esta fase se evalúa la utilidad y la exactitud de los modelos obtenidos en el proceso de minería de datos (Gómez, 2020).

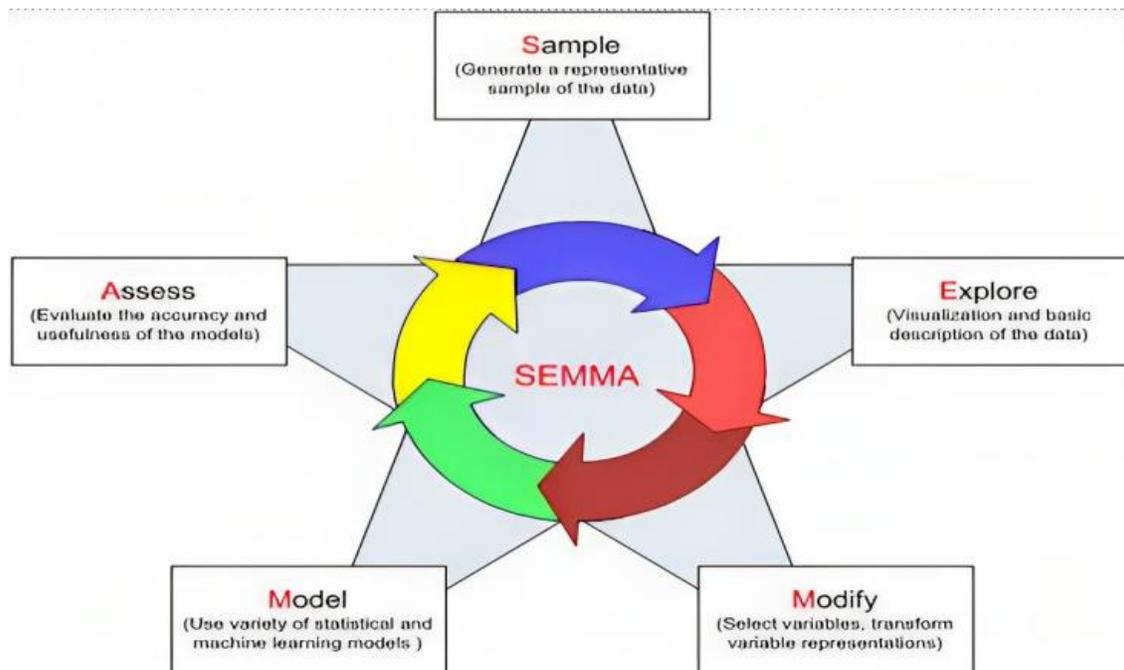


Figura 14 Diagrama SEMMA (Jun, 2017)

Tabla 1 Comparación de metodologías

	KDD	SEMMA	CRISP-DM
Definición	Metodología para el descubrimiento de conocimiento a partir de grandes conjuntos de datos	Metodología para el descubrimiento de patrones de negocios desconocidos	Metodología estructurada y detallada para el proceso de minería de datos
Etapas	Selección de datos, pre procesamiento, transformación, minería de datos, interpretación y evaluación de resultados.	Muestreo, exploración, modificación, modelado y evaluación.	Comprensión del negocio, comprensión de los datos, preparación de los datos, modelado.
Aplicación	Descubrimiento de conocimiento en diversas áreas	Descubrimiento de patrones de negocios para mejorar la toma de decisiones y la eficiencia operativa.	utilizado en proyectos de minería de datos.
Flexibilidad	Se adapta a diferentes situaciones y requisitos específicos.	Diseñada para aplicaciones empresariales utilizando sus herramientas y tecnologías.	Metodología flexible y adaptable que puede ser utilizada con diferentes herramientas y tecnologías de minería de datos.

Se eligió la metodología KDD para el análisis de datos ya que una de sus características es el descubrimiento de conocimiento y relaciones entre los datos, además de que es más flexible a distintos escenarios a diferencia de SEMMA y CRISP- DM que son más rígidas y lineales que es lo que se busca saber de cuáles son las principales dificultades que se enfrentaron los alumnos

## 3.4 Diseño y Desarrollo de Metodología

En el presente capítulo se desarrollará la metodología KDD (Knowledge Discovery in Databases), que es utilizada para descubrir información útil y conocimiento. La cual se desarrollará paso a paso para extraer y analizar la información, para lograr los objetivos de esta investigación.

### 3.4.1 Etapa de selección

Con el propósito de recopilar información crucial para un análisis detallado, se tiene la intención de desarrollar un cuestionario específico destinado al alumnado del séptimo semestre de la licenciatura en Ciencias Computacionales. El diseño de este cuestionario se enfocará en abordar aspectos significativos que permitan comprender su vivencia en las clases en vivo.

Durante esta pandemia de COVID-19 el semestre de séptimo pasó sus primeros 2 años en clases en línea en donde se desconoce cuáles fueron las principales afectaciones de dichos alumnos.

Todos los datos recopilados de las 53 personas se han tomado en cuenta para el análisis, además de que las preguntas realizadas y mostradas anteriormente son los atributos

Para la recolección de datos se consideró la manera más viable la aplicación de un cuestionario con la herramienta Google Forms. Para el diseño de éste, se realizó una previa búsqueda de información acerca de cuestionarios ya aplicados por diferentes instituciones para dar seguimiento a la educación de los estudiantes,

*Análisis de los aspectos académicos en la educación superior ocasionados por la pandemia mediante técnicas descriptivas de minería de datos*

tomando como referencia las respuestas con mayor relevancia como se expresa en el capítulo 1 apartado 1.5 Cuestionarios de seguimiento a la educación, donde se formularon las preguntas en secciones como: elementos de aprendizaje, elementos que causaron mayor conflicto, elementos técnicos que causaron mayor estrés, siendo estas de opción múltiple en escala de Likert como se muestra.

Se realizó una prueba piloto de este cuestionario con 19 preguntas de opción múltiple en escala de Likert, aplicado a 17 alumnos de la comunidad en general de dicha licenciatura que hayan cursado algún semestre durante el confinamiento por COVID-19.

Las preguntas que se realizaron se describen a continuación a través de las siguientes tablas:

*Tabla 2 Elementos de aprendizaje*

<b>Elementos de aprendizaje</b>	Muy buena (5)	Buena (4)	Regular (3)	Mala (2)	Muy mala (1)
¿Cómo consideras que fue tu autoaprendizaje?					
¿Tu planificación para realizar tus actividades fue?					
¿Tu concentración en tus actividades consideras que fue?					
Mi disciplina durante mis clases virtuales fue...					
¿Cómo consideraste tu autonomía?					

Tabla 3 Elementos personales que causaron mayor conflicto

Elementos personales que te causaron mayor conflicto	Mucho (5)	Lo suficiente (4)	Algo (3)	Poco (2)	Nada (1)
Problemas Económicos					
Problemas personales					
Cambios de rutina					
Problemas de salud mental					
Problemas de salud de familiares					
Problemas ambientales (lluvia, granizo, fuertes vientos)					

Tabla 4 Elementos técnicos que causaron mayor estrés

Elementos Técnicos que causaron mayor estrés	Muy significativa (5)	Significativa (4)	Algo Significativa (3)	Poco Significativa (2)	Irrelevante (1)
Los tiempos para la entrega de tareas					
La comunicación con tus profesores					
La organización con tus compañeros para realizar trabajos en equipo					
Problemas técnicos					
Organización de los tiempos					
Plataforma de aprendizaje					
Desinterés académico					
Conexión a internet					

Como complemento un apartado donde los estudiantes pudieran expresar otra dificultad que no se haya mencionado en el test.

- Menciona otro factor que consideras de impacto que causó durante la pandemia, que no se haya mencionado en este cuestionario.

Para validar la fiabilidad de este formulario de prueba se aplicó el coeficiente alfa de Cronbach, obteniendo un resultado de **0,8006667626**.

Lo que implica que de acuerdo al rango establecido se encuentra en “buena”, es decir que es fiable y representa calidad teniendo una buena consistencia interna.

Al analizar las respuestas proporcionadas por los estudiantes en la pregunta abierta, se agregaron más interrogantes acerca de estas dificultades mencionadas que no se había tomado en consideración. Además, se incorporó una sección adicional en la que se indagaba sobre su desempeño académico

Las preguntas que se agregaron son:

### **Elementos de Aprendizaje**

- ¿El seguimiento docente que te brindo, fue?
- ¿Cómo consideras que fue impartida la clase por los docentes?
- La empatía del docente hacia los contratiempos de los alumnos, ¿cómo consideras que fue?

### **Dificultades**

- ¿Crees que tu desempeño académico se vio afectado negativamente durante la pandemia?

- ¿Crees que tu promedio fue afectado de manera negativa durante la pandemia?
- Si en dado caso que hayas no acreditado una asignatura ¿crees que el motivo se debió por las clases en línea?
- ¿Crees que el número de personas conectadas en tu hogar, influyó en tu aprendizaje?
- ¿Negativamente?

### **Elementos personales que te causaron mayor conflicto**

- Falta de interacción social

Se abordó la mayor cantidad posible de dificultades en aspectos generales como de aprendizaje, personales, económicos y técnicos, intentando crear un formulario que no resultara tedioso de completar. Véase anexo 2

Teniendo un total de 31 preguntas que se enlistan a continuación:

1. ¿Cuál es tu género?
2. ¿Cuál es tu edad?
3. ¿Dónde vivías durante la pandemia?
4. ¿Cómo consideras que fue tu autoaprendizaje?
5. ¿Tu planificación para realizar tus actividades fue?
6. ¿Tu concentración en tus actividades consideras que fue?
7. Mi disciplina durante mis clases virtuales fue
8. ¿Cómo consideraste tu autonomía?
9. ¿El seguimiento docente que te brindo, fue?
10. ¿Cómo consideras que fue impartida la clase por los docentes?

11. La empatía del docente hacia los contratiempos de los alumnos, ¿cómo consideras que fue?
12. ¿Crees que tu desempeño académico se vio afectado negativamente durante la pandemia?
13. ¿Crees que tu promedio fue afectado de manera negativa durante la pandemia?
14. Si en dado caso que hayas no acreditado una asignatura ¿crees que el motivo se debió por las clases en línea?
15. ¿Crees que el número de personas conectadas en tu hogar, influyo en tu aprendizaje?
16. ¿Negativamente?
17. Problemas Económicos
18. [Problemas personales
19. Cambios de rutina
20. Problemas de salud mental
21. Problemas de salud de familiares
22. Problemas ambientales (lluvia, granizo, fuertes vientos)
23. Falta de interacción social
24. Los tiempos para la entrega de tareas
25. La comunicación con tus profesores
26. La organización con tus compañeros para realizar trabajos en equipo
27. Problemas técnicos
28. Organización de los tiempos
29. Plataforma de aprendizaje
30. Desinterés académico
31. Conexión a internet

Luciendo de siguiente manera en google forms:

## Aspectos académicos en la educación superior ocasionados por la pandemia

Este formulario tiene como finalidad el conocer las dificultades que han tenido los estudiantes durante su educación en la pandemia por el COVID-19

Indica que la pregunta es obligatoria

### Datos personales

¿Cuál es tu género? \*

Marca solo un óvalo.

- Femenino
- Masculino
- Indefinido
- Prefiero no decirlo

¿Cuál es tu edad? \*

Marca solo un óvalo.

- 19-20 años
- 21-22 años
- 23-24 años

¿Dónde vivías durante la pandemia? \*

Marca solo un óvalo.

- Ciudad (15 mil a 100 mil y de 100 mil y más habitantes)
- Semi Urbana (2,500 a 14,999 habitantes)
- Rural (menos de 100, de 100 a 499 y de 500 a 2 499 habitantes)

Figura 15 Vista desde Google Forms 1. Elaboración Propia.

Análisis de los aspectos académicos en la educación superior ocasionados por la pandemia mediante técnicas descriptivas de minería de datos

## Desempeño

- Lee cuidadosamente cada pregunta
- Selecciona el nivel con la que tienes más afín
- (5) Muy buena
- (4) Buena
- (3) Regular
- (2) Mala
- (1) Muy mala

### Elementos de Aprendizaje \*

Marca solo un óvalo por fila.

	Muy buena	Buena	Regular	Mala	Muy mala
¿Cómo consideras que fue tu autoaprendizaje?	<input type="radio"/>				
¿Tu planificación para realizar tus actividades fue?	<input type="radio"/>				
¿Tu concentración en tus actividades consideras que fue?	<input type="radio"/>				
Mi disciplina durante mis clases virtuales fue	<input type="radio"/>				
¿Cómo consideraste tu autonomía?	<input type="radio"/>				
¿El seguimiento docente que te brindo, fue?	<input type="radio"/>				
¿Como consideras que fue impartida la clase por los docentes?	<input type="radio"/>				
La empatia del docente hacia los contratiempos de los alumnos, ¿como consideras que fue?	<input type="radio"/>				

Figura 16 Vista desde Google Forms 2. Elaboración Propia.

Análisis de los aspectos académicos en la educación superior ocasionados por la pandemia mediante técnicas descriptivas de minería de datos

**Dificultades**

- Lee cuidadosamente cada pregunta
- Selecciona el nivel con la que tienes más afín
- (5) Mucho
- (4) Lo suficiente
- (3) Algo
- (2) Poco
- (1) Nada

**Dificultades**

*Marca solo un óvalo por fila.*

	Mucho	Lo suficiente	Algo	Poco	Nada
<b>¿Crees que tu desempeño académico se vio afectado negativamente durante la pandemia ?</b>	<input type="radio"/>				
<b>¿Crees que tu promedio fue afectado de manera negativa durante la pandemia?</b>	<input type="radio"/>				
<b>Si en dado caso que hayas no acreditado una asignatura ¿crees que el motivo se debió por las clases en línea ?</b>	<input type="radio"/>				
<b>¿Crees que el número de personas conectadas en tu hogar, influyo en tu aprendizaje?</b>	<input type="radio"/>				
<b>¿Negativamente?</b>	<input type="radio"/>				

Figura 17 Vista desde Google Forms 3. Elaboración Propia.

**Elementos personales que te causaron mayor conflicto**

*Marca solo un óvalo por fila.*

	Mucho	Lo suficiente	Algo	Poco	Nada
<b>Problemas Económicos</b>	<input type="radio"/>				
<b>Problemas personales</b>	<input type="radio"/>				
<b>Cambios de rutina</b>	<input type="radio"/>				
<b>Problemas de salud mental</b>	<input type="radio"/>				
<b>Problemas de salud de familiares</b>	<input type="radio"/>				
<b>Problemas ambientales (lluvia, granizo, fuertes vientos)</b>	<input type="radio"/>				
<b>Falta de interacción social</b>	<input type="radio"/>				

**Estres**

- Lee cuidadosamente cada pregunta
- Selecciona el nivel con la que tienes más afín
- (5) Muy significativo
- (4) Significativa
- (3) Algo significativa
- (2) Poco Significativa
- (1) Irrelevante

*Figura 18 Vista desde Google Forms 4. Elaboración Propia.*

*Análisis de los aspectos académicos en la educación superior ocasionados por la pandemia mediante técnicas descriptivas de minería de datos*

Elementos Técnicos que causaron mayor estrés \*

Marca solo un óvalo por fila.

	Muy significativa	Significativa	Algo significativa	Poco significativa	Irrelevante
<b>Los tiempos para la entrega de tareas</b>	<input type="radio"/>				
<b>La comunicacion con tus profesores</b>	<input type="radio"/>				
<b>La organización con tus compañeros para realizar trabajos en equipo</b>	<input type="radio"/>				
<b>Problemas técnicos</b>	<input type="radio"/>				
<b>Organización de los tiempos</b>	<input type="radio"/>				
<b>Plataforma de aprendizaje</b>	<input type="radio"/>				
<b>Desinterés académico</b>	<input type="radio"/>				
<b>Conexión a internet</b>	<input type="radio"/>				

Dificultades de los alumnos

Menciona otro factor que consideras de impacto que causo durante la pandemia, que no se haya mencionado en este cuestionario

Figura 19 Vista desde Google Forms 5. Elaboración Propia.

El cuestionario final se aplicó a los estudiantes del séptimo semestre del programa de Ciencias Computacionales, con un total previsto de 70 alumnos distribuidos en cuatro grupos, de los cuales 53 alumnos contestaron a lo que corresponde el 75.71% de la población muestra, dicho formulario se proporcionó a través del medio de comunicación WhatsApp.

Se aplicó nuevamente el coeficiente alfa de Cronbach para la corroboración de fiabilidad del test, obteniendo como resultado **0,8172180425**.

Obteniendo como resultados después de aplicar el cuestionario a la muestra objetivo:

¿Cuál es tu género?

53 respuestas

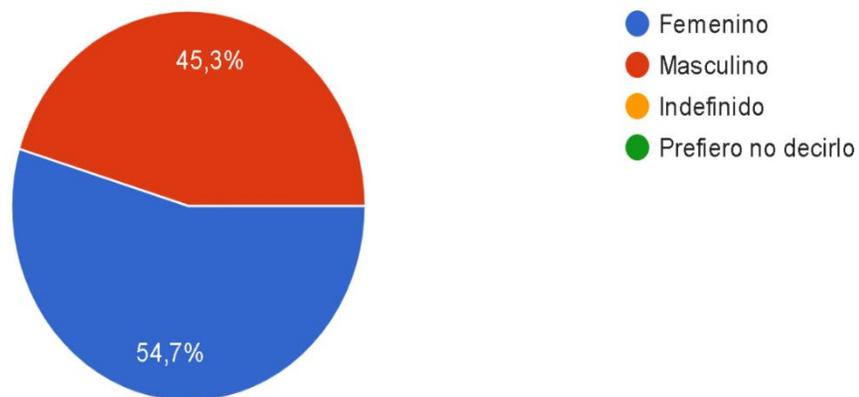


Figura 20 Grafica género. Elaboración Propia.

### ¿Cuál es tu edad?

53 respuestas

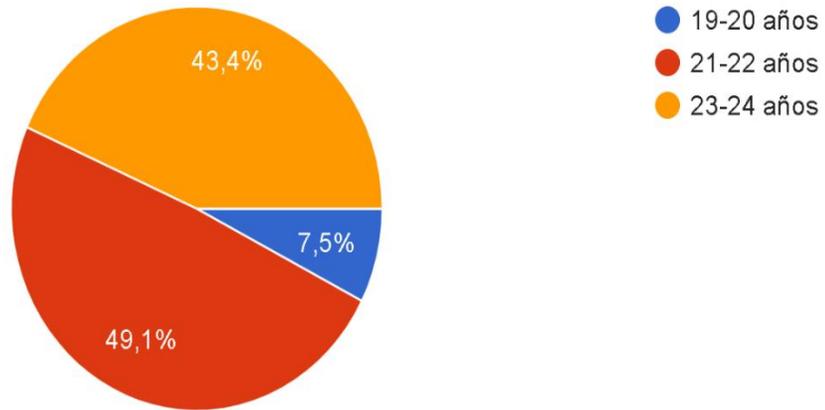


Figura 21 Grafica Edad. Elaboración Propia.

### ¿Dónde vivías durante la pandemia?

53 respuestas

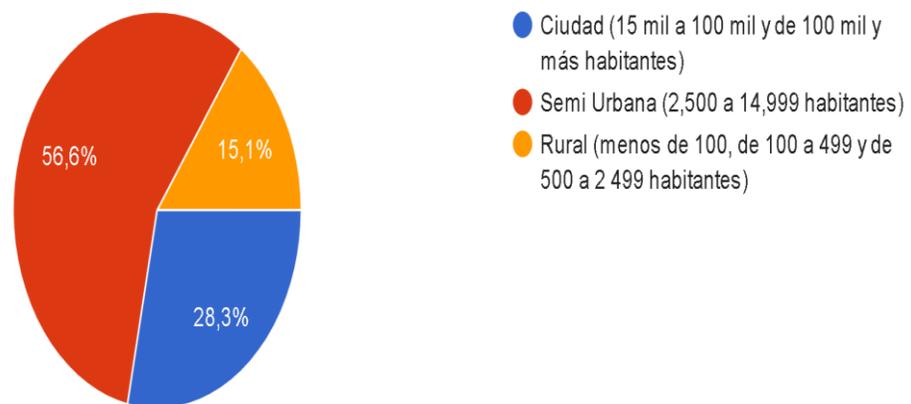


Figura 22 Zonas. Elaboración Propia.

Elementos de Aprendizaje

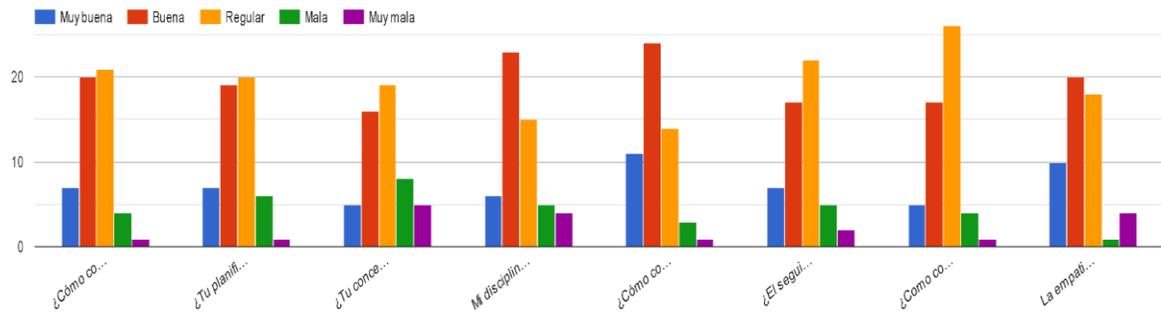


Figura 23 Grafica elementos de aprendizaje. Elaboración Propia.

Dificultades

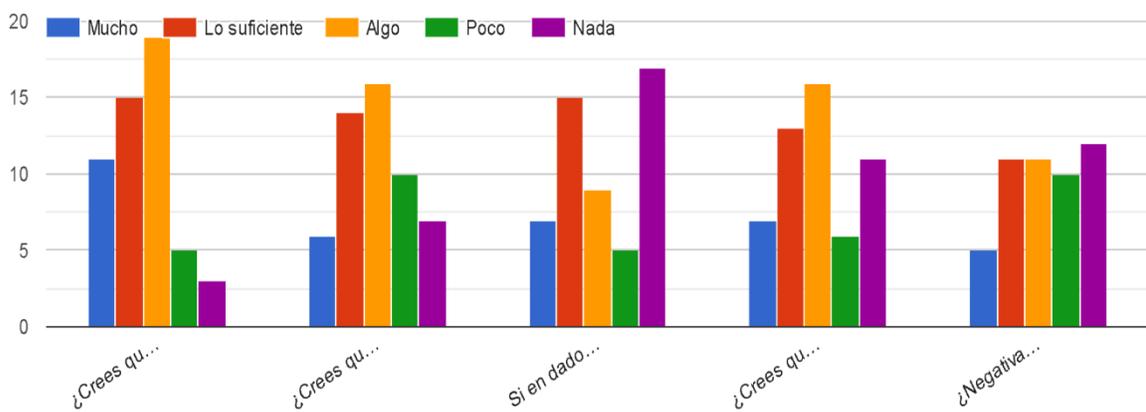


Figura 24 Grafica dificultades. Elaboración Propia.

Elementos personales que te causaron mayor conflicto

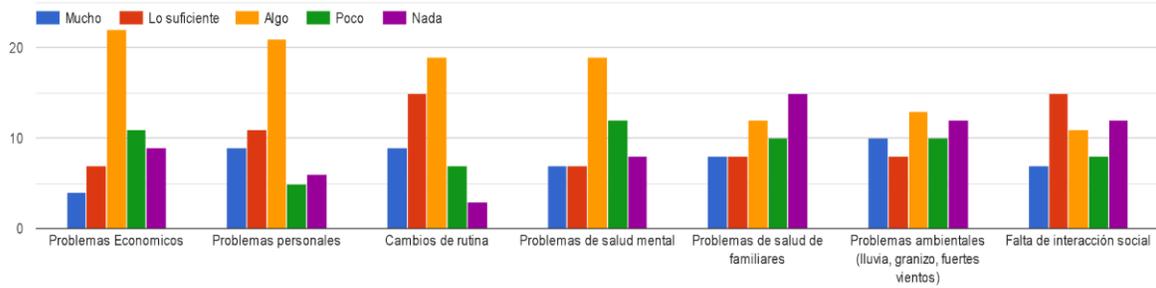


Figura 25 Mayor conflicto. Elaboración Propia.

Elementos Técnicos que causaron mayor estrés

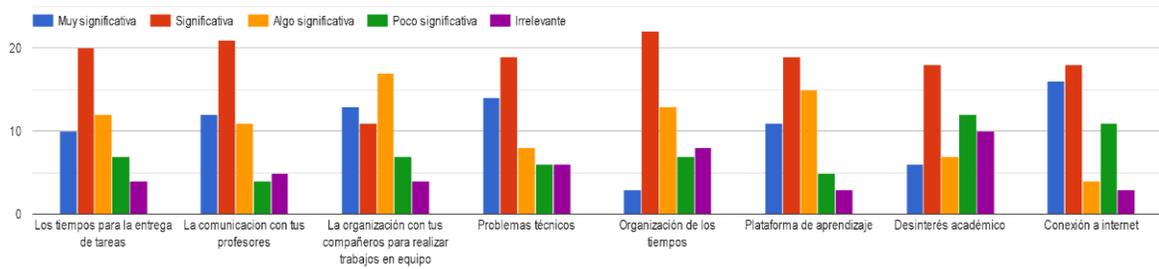


Figura 26 Mayor estrés. Elaboración Propia.



### 3.4.3 Etapa de transformación

Se realiza la transformación en documento Excel donde se almacenó la información de la encuesta previamente aplicada para poder ser procesado correctamente en el software Weka y aplicar la técnica de minería de datos.

Para que los datos pudieran ser correctamente procesados las respuestas obtenidas se tradujeron a valores numéricos de acuerdo a la escala de Likert además de que en la pregunta donde se cuestionaba acerca de la edad se sustituyó la letra “ñ” por “años” también preguntas se reemplazaron por “x1, x2..” sucesivamente para tener un control de los atributos evitando problemas con caracteres al introducir los datos al software WEKA permitiendo una correcta lectura de datos y finalmente reducir información donde se especificaba el número de habitantes por zona para evitar grandes espacios en blanco.

A partir de este documento **.xlsx** se convirtió a **.csv** ya que weka permite la lectura con esta extensión (Ver figura 28).

x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11
Femenino	21-22 años	Ciudad	5	4	4	4	5	3	3	4
Masculino	21-22 años	Semi Urbana	3	5	3	4	5	3	3	5
Femenino	21-22 años	Semi Urbana	4	4	4	4	5	4	5	3
Masculino	21-22 años	Semi Urbana	4	3	4	5	5	3	3	5
Masculino	21-22 años	Ciudad	4	3	3	2	4	3	3	4
Masculino	19-20 años	Semi Urbana	4	3	3	3	4	4	4	4
Masculino	23-24 años	Semi Urbana	3	3	2	4	4	4	3	4
Masculino	23-24 años	Semi Urbana	4	3	4	4	4	4	4	4
Masculino	21-22 años	Semi Urbana	5	5	5	5	5	5	5	5
Femenino	21-22 años	Semi Urbana	3	3	2	4	4	3	4	4
Masculino	21-22 años	Rural	5	2	1	2	4	2	3	3
Masculino	21-22 años	Semi Urbana	2	3	3	3	2	3	3	4
Masculino	21-22 años	Rural	3	4	3	2	3	3	3	3
Masculino	21-22 años	Semi Urbana	5	4	5	4	5	5	4	4
Masculino	21-22 años	Semi Urbana	3	3	3	3	3	3	3	3
Femenino	21-22 años	Ciudad	4	4	5	4	5	3	2	3
Femenino	21-22 años	Semi Urbana	3	2	2	1	3	2	3	3

Figura 28 Transformación de Información. Elaboración Propia.

### 3.4.4 Etapa Minería de datos

Se aplica la técnica de minería de datos para el análisis de los datos para posteriormente su interpretación de manera general, resaltando las preguntas con valores más relevantes, para mejor interpretación véase anexo 3 para visualizar el orden de las preguntas.

Se ejecutó el algoritmo K - means en el software WEKA teniendo las siguientes condiciones: 2 clústeres para el agrupamiento de los datos, utilizó la Distancia euclidiana para calcular la distancia entre los objetos y los centros, teniendo como máximo 500 iteraciones. Teniendo 53 objetos descritos por 31 atributos.

Consiguiendo 25 objetos es decir el 47% en el clúster 1 y los 28 objetos restantes representando el 53% en el clúster 2. (ver figura 29)

Final cluster centroids:			
Attribute	Full Data (53.0)	Cluster#	
		0 (25.0)	1 (28.0)
x1	Femenino	Masculino	Femenino
x2	21-22 años	21-22 años	23-24 años
x3	Semi Urbana	Semi Urbana	Semi Urbana
x4	3.5283	3.68	3.3929
x5	3.4717	3.6	3.3571
x6	3.1509	3.4	2.9286
x7	3.4151	3.56	3.2857
x8	3.7736	4.04	3.5357
x9	3.4151	3.56	3.2857
x10	3.3962	3.48	3.3214
x11	3.5849	3.92	3.2857
x12	3.4906	3.24	3.7143
x13	3.0377	2.76	3.2857
x14	2.8113	2.2	3.3571
x15	2.9811	2.52	3.3929
x16	2.7547	2.16	3.2857
x17	2.7358	2.4	3.0357
x18	3.2264	2.92	3.5
x19	3.3774	3.08	3.6429
x20	2.8679	2.36	3.3214
x21	2.6981	2.16	3.1786
x22	2.8868	2.24	3.4643
x23	2.9434	2.52	3.3214
x24	3.4717	3.32	3.6071
x25	3.5849	3.48	3.6786
x26	3.4528	3.16	3.7143
x27	3.5472	3.08	3.9643
x28	3.0943	2.72	3.4286
x29	3.566	3.4	3.7143
x30	2.9623	2.84	3.0714
x31	3.6604	3	4.25

Figura 29 Datos con 2 Clústeres. Elaboración Propia.

Nos encontramos otro escenario de 3 posibles clústeres en donde:

Se obtuvieron 21 objetos, lo que equivalía al 40%, en el primer clúster. En el segundo clúster, se encontraron 19 objetos, representando el 36% del total. Los últimos 13 objetos fueron asignados al tercer clúster con un 25% (ver figura 30).

Final cluster centroids:				
Attribute	Full Data (53.0)	Cluster#		
		0 (21.0)	1 (19.0)	2 (13.0)
x1	Femenino	Femenino	Femenino	Masculino
x2	21-22 años	21-22 años	23-24 años	23-24 años
x3	Semi Urbana	Semi Urbana	Semi Urbana	Semi Urbana
x4	3.5283	3.2381	3.5263	4
x5	3.4717	3.1429	3.5263	3.9231
x6	3.1509	2.8095	3.0526	3.8462
x7	3.4151	2.9048	3.5263	4.0769
x8	3.7736	3.5238	3.7895	4.1538
x9	3.4151	2.9048	3.5263	4.0769
x10	3.3962	3.0476	3.3684	4
x11	3.5849	3.1429	3.5789	4.3077
x12	3.4906	3.6667	3.6842	2.9231
x13	3.0377	3.2857	3.3158	2.2308
x14	2.8113	3.0952	3.1579	1.8462
x15	2.9811	3.1905	3.4211	2
x16	2.7547	3.0952	3.1053	1.6923
x17	2.7358	2.3333	3.1579	2.7692
x18	3.2264	2.9524	3.7368	2.9231
x19	3.3774	3.2857	3.8421	2.8462
x20	2.8679	2.7143	3.5789	2.0769
x21	2.6981	2.5714	3.2632	2.0769
x22	2.8868	2.3333	4.0526	2.0769
x23	2.9434	2.8571	3.5263	2.2308
x24	3.4717	3.5238	4	2.6154
x25	3.5849	3.7143	4.2632	2.3846
x26	3.4528	3.0476	4.2632	2.9231
x27	3.5472	3.2857	4.5263	2.5385
x28	3.0943	2.8571	3.8947	2.3077
x29	3.566	3.4762	4.0526	3
x30	2.9623	3	3.6842	1.8462
x31	3.6604	3.4762	4.5263	2.6923

Figura 30 Datos con 3 Clústeres. Elaboración Propia.

Por último, realizamos una prueba utilizando 4 clústeres, en la cual se observó que el 38% de los elementos (20 objetos) pertenecían al clúster 1, mientras que el 26% (14 objetos) se ubicaban en el clúster 2. El tercer clúster estaba compuesto por 13 estudiantes, lo que representaba el 25% del total, y el cuarto clúster contenía únicamente el 11% de los datos, con 6 elementos. (ver figura 31)

Final cluster centroids:					
Attribute	Cluster#				
	Full Data (53.0)	0 (20.0)	1 (14.0)	2 (13.0)	3 (6.0)
x1	Femenino	Femenino	Femenino	Masculino	Masculino
x2	21-22 años	21-22 años	23-24 años	23-24 años	23-24 años
x3	Semi Urbana	Semi Urbana	Semi Urbana	Semi Urbana	Semi Urbana
x4	3.5283	3.2	3.5714	4	3.5
x5	3.4717	3.15	3.5	3.9231	3.5
x6	3.1509	2.8	3.0714	3.8462	3
x7	3.4151	2.85	3.7857	4.0769	3
x8	3.7736	3.55	3.7143	4.1538	3.8333
x9	3.4151	2.9	3.4286	4.0769	3.6667
x10	3.3962	3.1	3.2143	4	3.5
x11	3.5849	3.25	3.5	4.3077	3.3333
x12	3.4906	3.65	3.5	2.9231	4.1667
x13	3.0377	3.3	2.9286	2.2308	4.1667
x14	2.8113	3.1	2.6429	1.8462	4.3333
x15	2.9811	3.2	3	2	4.3333
x16	2.7547	3.15	2.5	1.6923	4.3333
x17	2.7358	2.4	2.8571	2.7692	3.5
x18	3.2264	2.9	3.5	2.9231	4.3333
x19	3.3774	3.2	3.8571	2.8462	4
x20	2.8679	2.65	3.5	2.0769	3.8333
x21	2.6981	2.45	3.4286	2.0769	3.1667
x22	2.8868	2.25	3.7857	2.0769	4.6667
x23	2.9434	2.75	3.6429	2.2308	3.5
x24	3.4717	3.45	3.8571	2.6154	4.5
x25	3.5849	3.7	4.1429	2.3846	4.5
x26	3.4528	3.05	4.0714	2.9231	4.5
x27	3.5472	3.25	4.5	2.5385	4.5
x28	3.0943	2.8	3.8571	2.3077	4
x29	3.566	3.45	4	3	4.1667
x30	2.9623	3	3.6429	1.8462	3.6667
x31	3.6604	3.4	4.5	2.6923	4.6667

Figura 31 Datos con 4 Clústeres. Elaboración Propia.

### 3.4.5 Interpretación y evaluación

Una vez obtenidos los resultados del algoritmo K-Means en el software Weka se hará una interpretación de los datos más relevantes de los escenarios conseguidos de 2, 3 y clústeres, dado que se encuentran en decimal los resultados se realizará un redondeo hacia arriba a partir de 0.5 para un adecuado análisis, traduciendo a la escala de Likert previamente establecida en cada sección de preguntas como se observa en el punto 3.1 Etapa de Selección.

- 2 Clústeres (ver Figura 32): Los datos más significativos en el clúster 1 son referentes a las preguntas 4 y 8 cuestionando su autoaprendizaje y autonomía, considerándola buena al igual que la empatía del profesor, En clúster 2 los datos más relevantes se encuentran en los factores que causaron un estrés notable siendo su conexión a internet y problemas técnicos.

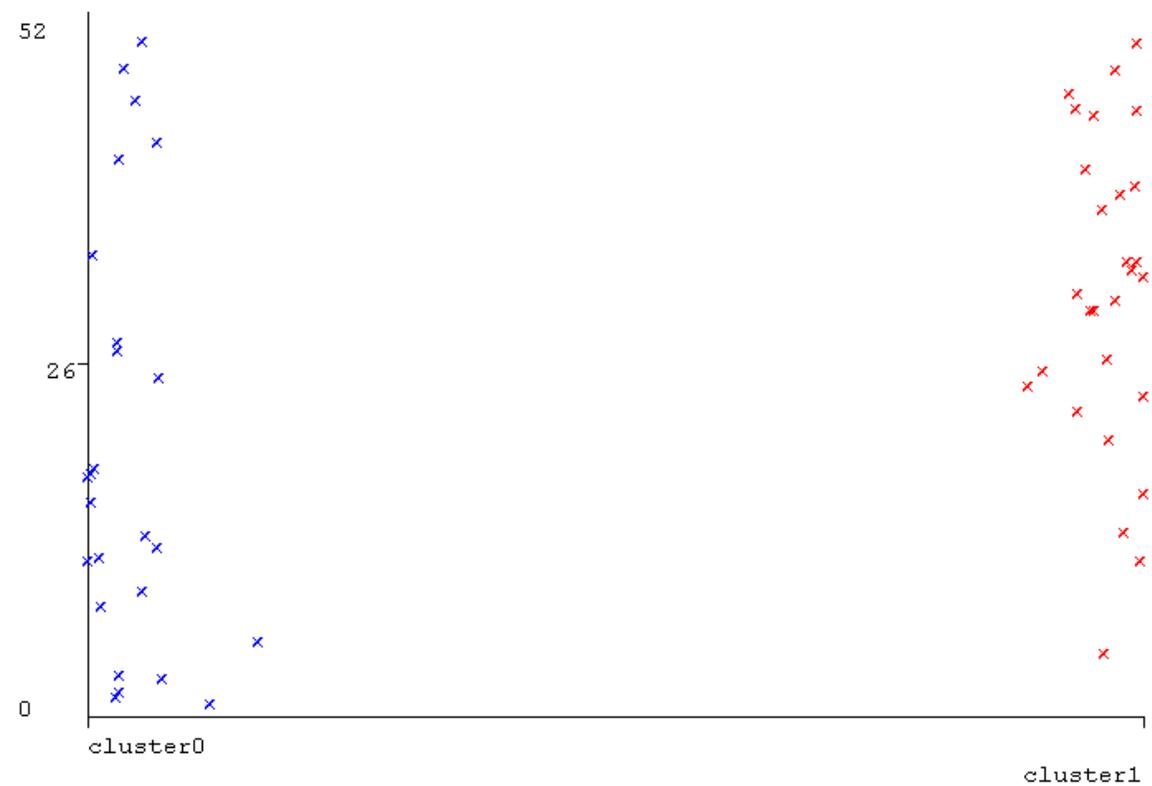


Figura 32 Distribución de la muestra 2 clústeres. Elaboración Propia.

- 3 Clústeres (ver Figura 33): Observamos el clúster número 1 tiene mayor concentración (40%), resalta la pregunta sobre si su desempeño se vio afectado por la pandemia donde en su mayoría consideró que esta afectó lo suficiente al igual que su mayor estrés fue la comunicación con sus profesores. En el clúster 2 gran número de los problemas técnicos causaron un estrés significativo durante sus clases virtuales, por último, en el tercer clúster las 2 preguntas relevantes son como consideraban su autoaprendizaje y como creían que fue impartidas sus clases puntuando como notables.

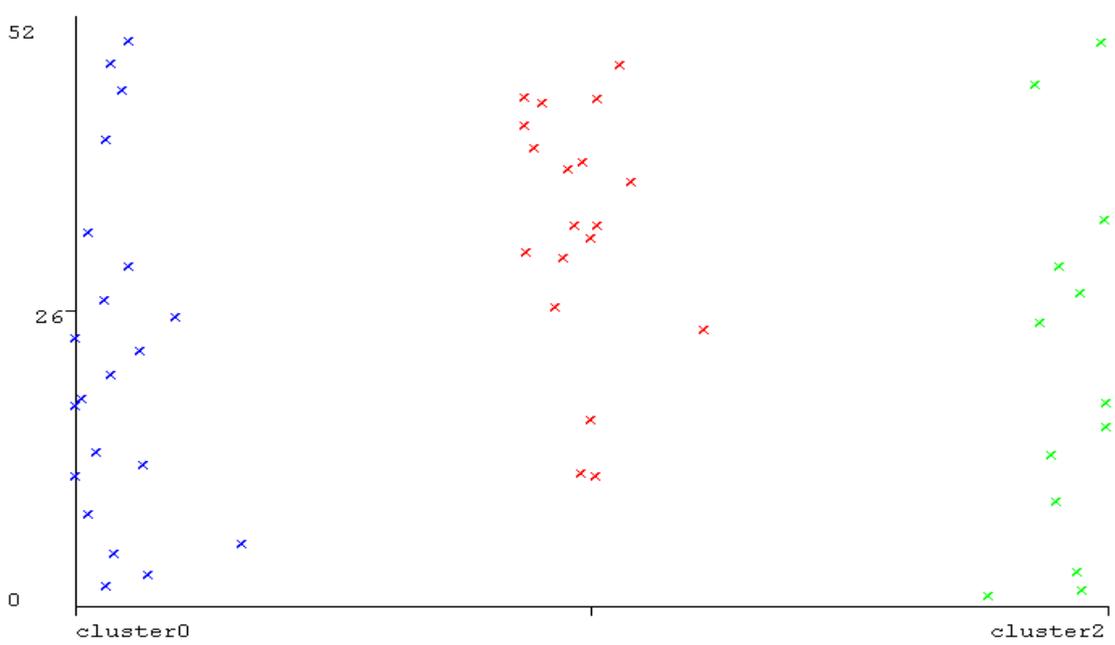


Figura 33 Distribución de la muestra 3 clústeres. Elaboración Propia.

- 4 Clústeres (ver figura 34): Se destaca el clúster número 4 que únicamente tiene 6 objetos de los cuales la mayoría de los datos tienen una puntuación más de 4 en la mayoría de las preguntas a excepción donde se cuestiona sobre cómo consideran su concentración y disciplina durante sus actividades, respondiendo como “regular”.

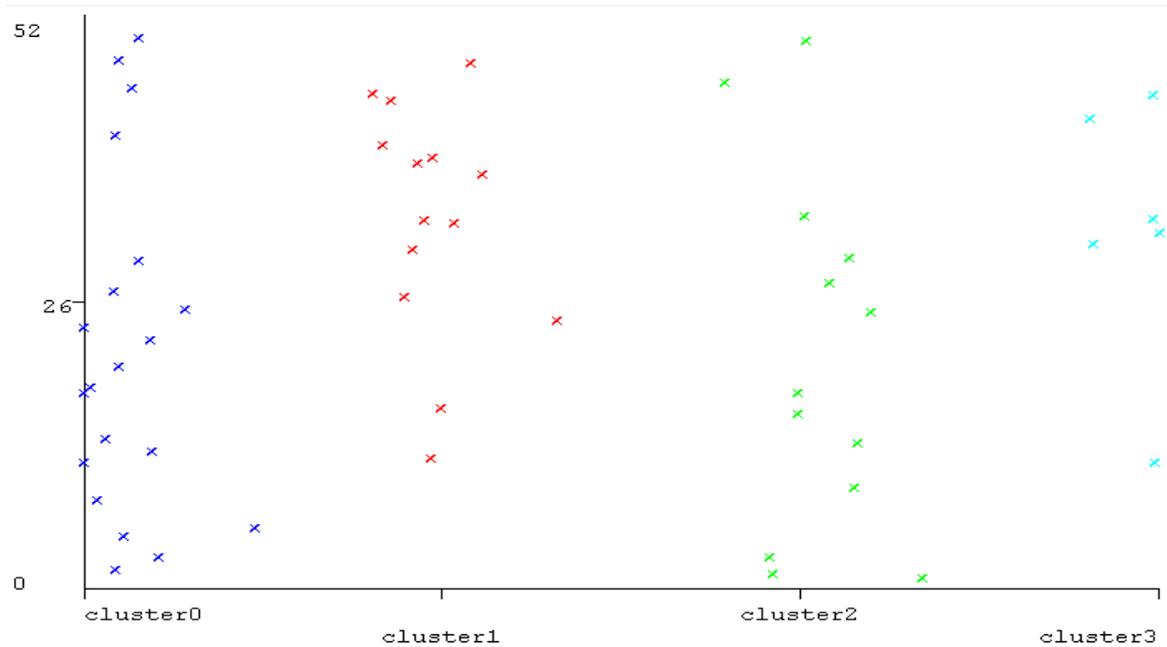


Figura 34 Distribución de la muestra 4 clústeres. Elaboración Propia.

## CAPITULO 4 RESULTADOS

Se hicieron diferentes particiones al conjunto de datos para determinar cuál es el número de clústeres indicado, para definir esto se utilizó el índice Calinski Harabasz el cual ayuda a evaluar qué tan bien se ha llevado a cabo la agrupación, cuanto más alto sea el número, mejor será la manera en que se agruparon. Así que se buscó el número de grupos que diera el índice Calinski-Harabasz más alto, indicando cuántos clústeres es el ideal para nuestros datos.

Teniendo como valores:

- **K=2 – CH(c)= 26.48**
- **K=3 – CH(c)= 15.51**
- **K=4 – CH(c)= 10.46**

De acuerdo a los resultados el número indicado de clústeres son **2**, teniendo el valor más alto, por lo tanto, la interpretación de dichos datos distribuidos en estos es:

Observamos en la **figura 29** que, en su mayoría, el Clúster 1 está compuesto por hombres de entre 21 y 22 años que residían en una zona semi urbana durante el período de confinamiento. Estos individuos consideraron que su autoaprendizaje y la planificación de actividades fueron buenas, aunque su nivel de concentración en las tareas se califica como regular. Además, valoraron positivamente su disciplina y autonomía, así como el seguimiento y la empatía proporcionados por su profesor. Sin embargo, difieren en sus clases impartidas por el docente, las cuales consideran que fueron aceptables.

En cuanto a sus dificultades, perciben que su desempeño y promedio se vieron suficientemente afectados, pero no consideran que reprobar una asignatura durante

este período sea directamente atribuible a las clases en línea. Además, suponen que la cantidad de personas conectadas en su hogar tuvo un impacto moderado en su rendimiento, al igual que sus problemas económicos. Creen que el factor que más estrés les generó fue la incidencia de problemas personales, así como la falta de interacción social. Dificultades relacionadas con problemas mentales, la salud de sus familiares y factores externos, como las condiciones climáticas (lluvia, vientos, granizo, etc.), consideran que causaron un nivel de estrés relativamente bajo.

Tanto los tiempos para la entrega de tareas, la comunicación con los profesores, la organización de su tiempo y con sus compañeros para realizar trabajos en equipo, problemas técnicos, la plataforma que se utilizó para dar seguimiento, su desinterés académico y su conexión a internet causaron un estrés considerable.

En el clúster 2 predominan mujeres de entre 23 y 24 años que vivían en zona semi urbana durante sus clases virtuales de las cuales consideran que su autoaprendizaje, planeación, concentración y disciplina fueron regulares, pero en cuanto a su autonomía creen que fue sobresaliente, en cuanto al seguimiento, clases y empatía que les brindó los docentes fue ordinaria.

Los alumnos creen que la pandemia les afectó de manera significativa a su rendimiento académico negativamente, viéndolo reflejado en su promedio y una posible causa de no acreditación de una materia, así como también el número de personas conectadas en su hogar fue influyente.

Problemas personales, cambios en la rutina fueron factores que causaron mayor conflicto a diferencia de los económicos, de salud mental, falta de interacción social

y aspectos ambientales. Elementos técnicos como la organización de tiempos y el desinterés académico no causaron tanto impacto a diferencia de la entrega de tareas, la comunicación con los profesores, organización con compañeros y el uso de la plataforma educativa que tuvieron mayor influencia negativa.

## Conclusiones

Este proyecto tuvo como objetivo analizar las dificultades que se enfrentaron los alumnos de séptimo de la Licenciatura en Ciencias Computacionales julio – diciembre 2023, y el impacto en su rendimiento académico.

Como se observó los datos se dividieron en 2 grupos acorde a sus características donde se obtuvo que ambos clústeres coincidieron que la pandemia había afectado negativamente a su desempeño académico, a uno de los grupos ligeramente más que a otro, viéndose reflejado en su promedio, pero reprobar una materia no lo retribuyen directamente a las clases en línea.

Los alumnos del clúster 1 en cuestión de su autoaprendizaje, su planificación de actividades y su disciplina durante sus clases virtuales se mostraron ligeramente más alto que el clúster 2, en donde ambos coinciden que su autonomía es buena.

Las principales dificultades con mayor peso para ambos clústeres fueron:

1. La plataforma de aprendizaje.
2. La comunicación con sus profesores.
3. Conexión a internet
4. Problemas técnicos.
5. La organización con los compañeros para realizar trabajos en equipo.
6. Problemas personales.
7. Cambios en la rutina.

Esta investigación aportara conocimiento a la Universidad Autónoma del Estado de Hidalgo a generar estrategias para futuros escenarios similares ya que no se está exento de que vuelva a suceder, con el fin de brindar una educación eficiente para la preparación de profesionales del mañana.

Para finalizar este trabajo se encontró una limitante que fue el paro del Instituto de Ciencias Básicas e Ingeniería que sucedió durante la recolección de datos.

## Referencias

- Aprendizaje automático. (1 de Junio de 2017). Obtenido de Hewlett Packard Enterprise: <https://www.hpe.com/lamerica/es/what-is/machine-learning.html#>
- Bellini Saibene, Y. N. (Septiembre de 2021). Desarrollo y uso de herramientas libres para la explotación de datos de los radares meteorológicos del INTA. Obtenido de ResearchGate : [https://www.researchgate.net/publication/306959832\\_Desarrollo\\_y\\_uso\\_de\\_herramientas\\_libres\\_para\\_la\\_explotacion\\_de\\_datos\\_de\\_los\\_radares\\_meteorologicos\\_del\\_INTA](https://www.researchgate.net/publication/306959832_Desarrollo_y_uso_de_herramientas_libres_para_la_explotacion_de_datos_de_los_radares_meteorologicos_del_INTA)
- Cardozo Contreras, D., & Herrera Rodríguez, Á. A. (2019). El uso de mecanismos de innovación en la educación de adultos. Repositorio Intitucional, Bogotá. Obtenido de <https://hdl.handle.net/10656/7969>
- Descifrando la importancia del aprendizaje automático en la tecnología moderna. (27 de Julio de 2023). Obtenido de CIS INFORMATICA: <https://www.cisinformatica.cat/es/que-es-el-aprendizaje-automatico/>
- Descifrando la importancia del aprendizaje automático en la tecnología moderna. (27 de Julio de 2023). Obtenido de CIS INFORMATICA: <https://www.cisinformatica.cat/es/que-es-el-aprendizaje-automatico/>
- Dey, D. (25 de abril de 2022). Índice Calinski-Harabasz – Índices de validez de conglomerados . Obtenido de Geeksforgeeks.: <https://www.geeksforgeeks.org/calinski-harabasz-index-cluster-validity-indices-set-3/>
- García Dobarganes, P. C. (2 de Junio de 2021). Educación en pandemia. Obtenido de IMCO: [https://imco.org.mx/wp-content/uploads/2021/06/20210602\\_Educacio%CC%81n-en-pandemia\\_Documento.pdf](https://imco.org.mx/wp-content/uploads/2021/06/20210602_Educacio%CC%81n-en-pandemia_Documento.pdf)
- Giorgia, C., & Cristina, V. (26 de Enero de 2020). Cluster aplicado el metodo K-means. Obtenido de Escuela Superior Politécnica de Chimborazo: [https://rstudio-pubs-static.s3.amazonaws.com/570289\\_dffab0be9f304cd3bccb3164031da504.html#1](https://rstudio-pubs-static.s3.amazonaws.com/570289_dffab0be9f304cd3bccb3164031da504.html#1)
- Gómez Caballero, L. M. (30 de Noviembre de 2020). Análisis en Big Data: métodos de minería de datos y herramientas. Obtenido de datahack: <https://www.datahack.es/analisis-big-data/#kdd>

- González, L. (2022). ¿Qué es el Aprendizaje no Supervisado? Obtenido de aprende IA: [https://aprendeia.com/aprendizaje-no-supervisado-machine-learning/#%C2%BFQue\\_es\\_el\\_Aprendizaje\\_no\\_Supervisado](https://aprendeia.com/aprendizaje-no-supervisado-machine-learning/#%C2%BFQue_es_el_Aprendizaje_no_Supervisado)
- Gutiérrez Cruz, D., Albarran Fernández, Y. A., & Rico Molina, R. [. (2017). MANUAL PARA PRÁCTICAS DEL: Taller con Weka. Repositorio Institucional. Obtenido de <http://hdl.handle.net/20.500.11799/69982>
- Herramienta de diagnóstico de K-centroides. (2019). Obtenido de alteryx: [https://help.alteryx.com/2020.2/es/K-Centroids\\_Diagnostics.htm](https://help.alteryx.com/2020.2/es/K-Centroids_Diagnostics.htm)
- Hitpass, B. (Noviembre de 2014). El nivel de Actividad en el Proceso Educativo como Indicador de Riesgo de Deserción Estudiantil medido en tiempo real con apoyo de tecnología BAM. Obtenido de ResearchGate: [https://www.researchgate.net/publication/268687479\\_El\\_nivel\\_de\\_Actividad\\_en\\_el\\_Proceso\\_Educativo\\_como\\_Indicador\\_de\\_Riesgo\\_de\\_Desercion\\_Estudiantil\\_medido\\_en\\_tiempo\\_real\\_con\\_apoyo\\_de\\_tecnologia\\_BAM](https://www.researchgate.net/publication/268687479_El_nivel_de_Actividad_en_el_Proceso_Educativo_como_Indicador_de_Riesgo_de_Desercion_Estudiantil_medido_en_tiempo_real_con_apoyo_de_tecnologia_BAM)
- IESALC. (2020). COVID-19 y educación superior: De los efectos inmediatos al día después Análisis de impactos, respuestas políticas y recomendaciones.
- IESPE. (3 de Julio de 2020). <https://www.iespe.mx/post/se-anuncio-la-estrategia-de-regreso-a-clases-escalonada>. Obtenido de IESPE: <https://www.iespe.mx/post/se-anuncio-la-estrategia-de-regreso-a-clases-escalonada>
- Jun. (11 de Octubre de 2017). SEMMA. Obtenido de Tistory: <https://yunstudybox.tistory.com/90>
- La teoría del aprendizaje adulto de Malcolm Knowles. (01 de Septiembre de 2020). Obtenido de Scribd: <https://es.scribd.com/document/474462589/La-teoria-del-aprendizaje-adulto-de-Malcolm-Knowles-docx#>
- Lichtenstein, N. (14 de Mayo de 2013). La-historia-de-la-Andragogia. Recuperado el 09 de 2023, de Scribd: <https://es.scribd.com/document/141458096/La-historia-de-la-Andragogia#>
- Mancilla Vela, G., Leal Gatica, P., Sánchez-Ortiz, A., & Vidal-Silva, C. (2020). Factores asociados al éxito de los estudiantes en modalidad. Formación Universitaria,, 23-36. doi:<http://dx.doi.org/10.4067/S0718-50062020000600023>
- Maria, C. (19 de Septiembre de 2022). Qué es la minería de datos: conceptos, técnicas y ejemplos. Obtenido de HubSpot: <https://blog.hubspot.es/marketing/mineria-datos>
- Méndez Álvarez, I. D. (2019). Método de agrupamiento basado en funciones kernel dirigido por índices de validez [Tesis de maestría, Centro de *Análisis de los aspectos académicos en la educación superior ocasionados por la pandemia mediante técnicas descriptivas de minería de datos*

- Investigación y de Estudios Avanzados]. Repositorio Institucional. Obtenido de  
[https://www.tamps.cinvestav.mx/descargables/tesis/2019/10\\_2019\\_iv%C3%A1n\\_Daniel\\_Mendez\\_%C3%81lvarez\\_M\\_2017.pdf](https://www.tamps.cinvestav.mx/descargables/tesis/2019/10_2019_iv%C3%A1n_Daniel_Mendez_%C3%81lvarez_M_2017.pdf)
- Miguel Román, J. A. (2020). La educación superior en tiempos de pandemia: una visión desde dentro del proceso formativo. Latinoamericana de Estudios Educativos (México).
- Muguira, A. (31 de Agosto de 2020). ¿Qué es la escala de Likert y cómo utilizarla? Obtenido de QuestionPro: <https://www.questionpro.com/blog/es/que-es-la-escala-de-likert-y-como-utilizarla/>
- Núñez, R. H., Masías, E. J., & Quispe, W. B. (5 de Mayo de 2022). Pandemia, educación virtual y su impacto en la educación de la región Puno- Perú. Científica Multidisciplinar, 1. Obtenido de <https://ciencialatina.org/index.php/cienciala/article/view/2265>
- Organización Panamericana de la salud. (11 de Marzo de 2020). La OMS caracteriza a COVID-19 como una pandemia. Obtenido de Organización Panamericana de la salud: <https://www.paho.org/es/noticias/11-3-2020-oms-caracteriza-COVID-19-como-pandemia#:~:text=La%20epidemia%20de%20COVID%2D19,un%20gran%20n%C3%BAmero%20de%20personas.>),
- Ortega, C. (2023). Investigación cuantitativa. Qué es y cómo realizarla. Obtenido de Questionpro : <https://www.questionpro.com/blog/es/que-es-la-investigacion-cuantitativa/>
- Padilla Guerra, Y. M., & Abreu, J. L. (Noviembre de 2021). El 2020: Un Antes y Un Después en la Educación. Obtenido de spentamexico: [http://www.spentamexico.org/v16-n3/A15.16\(3\)1-11.pdf](http://www.spentamexico.org/v16-n3/A15.16(3)1-11.pdf)
- Pérez León, G. (30 de Octubre de 2022). ¿Qué es y para qué sirve el Alfa de Cronbach? Obtenido de LinkedIn: <https://es.linkedin.com/pulse/qu%C3%A9-es-y-para-sirve-el-alfa-de-cronbach-gabriel-p%C3%A9rez-le%C3%B3n->
- Pérez León, G. (30 de Octubre de 2022). Coeficiente Alfa de Cronbach: ¿Qué es y para qué sirve el Alfa de Cronbach? Obtenido de GPL: <https://gplresearch.com/coeficiente-alfa-de-cronbach/>
- Pompas Gutiérrez, J. (6 de Marzo de 2024). Naive Bayes al completo. Obtenido de LinkedIn: <https://es.linkedin.com/pulse/naive-bayes-al-compelto-jordi-pompas-guti%C3%A9rrez-wrmrf>

- Pompas Gutiérrez, J. (6 de Marzo de 2024). Naive Bayes al completo. Obtenido de LinkedIn: <https://es.linkedin.com/pulse/naive-bayes-al-compelto-jordi-pompas-guti%C3%A9rrez-wrmrf>
- Prado, P. E. (2022). Estudio de técnicas de agrupamiento en procesos de datos a gran escala [Tesis de Especialización, Universidad Nacional de La Plata]. Repositorio Institucional. Obtenido de <http://sedici.unlp.edu.ar/handle/10915/145268>
- Qué es el aprendizaje supervisado (2 de Marzo de 2017). Obtenido de IBM: <https://www.ibm.com/mx-es/topics/supervised-learning>
- Qué es el aprendizaje supervisado (2 de Marzo de 2017). Obtenido de IBM: <https://www.ibm.com/mx-es/topics/supervised-learning>
- Qué es un árbol de decisión (23 de Mayo de 2022). Obtenido de IBM: <https://www.ibm.com/mx-es/topics/decision-tree>
- Qué es un árbol de decisión (23 de Mayo de 2022). Obtenido de IBM: <https://www.ibm.com/mx-es/topics/decision-trees>
- Ramírez, L. (5 de Enero de 2023). Algoritmo k-means: ¿Qué es y cómo funciona? Obtenido de IEBS: <https://www.iebschool.com/blog/algoritmo-k-means-que-es-y-como-funciona-big-data/>
- Rico Molina, R., Rodríguez Páez, C. L., & Duran Lopez, V. [. (2019). MANUAL PARA PRÁCTICAS DE LA UNIDAD DE APRENDIZAJE: Taller con Weka. Repositorio Institucional. Obtenido de <http://hdl.handle.net/20.500.11799/107965>
- Rudeus, G. (5 de Julio de 2022). Índice de Calinski-Harabasz: índices de validez de conglomerados. Obtenido de Barcelona Geeks: <https://barcelonageeks.com/indice-de-calinski-harabasz-indices-de-validez-de-conglomerados-conjunto-3/>
- Ruiz Mitjana, L. (22 de Mayo de 2019). Alfa de Cronbach ( $\alpha$ ): qué es y cómo se usa en estadística. Obtenido de Psicología y mente
- SAP. (11 de Mayo de 2023). ¿Qué es la minería de datos? Obtenido de SAP Business Technology Platform: <https://www.sap.com/latinamerica/products/technology-platform/hana/what-is-data-mining.html>
- Sierra Fontalvo, R. (2006). La Andragogía, modelo propicio para el desarrollo de la educación de adultos. PROSPECTIVA, 100-101.
- Torres Ramírez, C. (Mayo de 2021). Refundar la educación después de la pandemia. Obtenido de bibliodigitalibd:

[http://bibliodigitalibd.senado.gob.mx/bitstream/handle/123456789/5635/ADP\\_23\\_Refundar%20la%20educaci%C3%B3n%20F.pdf?sequence=1&isAllowed=y](http://bibliodigitalibd.senado.gob.mx/bitstream/handle/123456789/5635/ADP_23_Refundar%20la%20educaci%C3%B3n%20F.pdf?sequence=1&isAllowed=y)

UAEH - Universidad Autónoma del Estado de Hidalgo. (31 de Marzo de 2017).  
Obtenido de MEXTUDIA: <https://mextudia.com/universidades/uaeh/>

UAEH - Universidad Autónoma del Estado de Hidalgo. (31 de Marzo de 2017).  
Obtenido de MEXTUDIA: <https://mextudia.com/universidades/uaeh/>

Valero Cedeño, N. J., Castillo Matute, A. L., Rodríguez Pincay, R., Padilla Hidalgo, M., & Cabrera Hernández, M. (2020). Retos de la educación virtual en el proceso enseñanza aprendizaje durante la pandemia de COVID-19. *Dominio De Las Ciencias*, 1201-1220. doi: <https://doi.org/10.23857/dc.v6i4.1530>