

UNIVERSIDAD NACIONAL DEL SANTA
FACULTAD DE INGENIERIA
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS E
INFORMÁTICA



UNS
UNIVERSIDAD
NACIONAL DEL SANTA

“Modelo predictivo del rendimiento académico de estudiantes universitarios asociado con la influencia de los videojuegos usando Machine Learning”

Tesis para obtener el Título Profesional de Ingeniero de Sistemas e Informática

Autores:

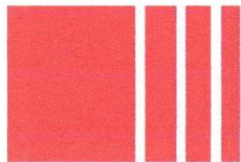
Bach. Castañeda Chávez, Diego Alberto
Código ORCID: 0009-0004-8622-8512

Bach. Perez Chang, Valeria Isabel
Código ORCID: 0000-0003-1673-153X

Asesor:

Dr. Caselli Gismondi, Hugo Esteban
DNI N° 32819296
Código ORCID: 0000-0002-2812-6727

NUEVO CHIMBOTE – PERÚ
2024



FACULTAD DE INGENIERÍA
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS E INFORMÁTICA

UNIVERSIDAD NACIONAL DEL SANTA
FACULTAD DE INGENIERIA
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS E
INFORMÁTICA



UNS
UNIVERSIDAD
NACIONAL DEL SANTA

“Modelo predictivo del rendimiento académico de estudiantes universitarios asociado con la influencia de los videojuegos usando Machine Learning”

Tesis para Optar el Título Profesional de Ingeniero de Sistemas e Informática

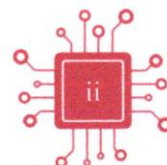
Revisado y Aprobado por Asesor:

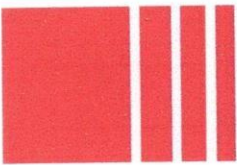
Dr. Caselli Gismondi, Hugo Esteban
Asesor

DNI N° 32819296
Código ORCID: 0000-0002-2812-6727

BACH. CASTAÑEDA CHÁVEZ DIEGO ALBERTO

BACH. PEREZ CHANG VALERIA ISABEL





UNIVERSIDAD NACIONAL DEL SANTA
FACULTAD DE INGENIERIA
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS E
INFORMÁTICA

MODELO PREDICTIVO DEL RENDIMIENTO ACADÉMICO DE
ESTUDIANTES UNIVERSITARIOS ASOCIADO CON LA
INFLUENCIA DE LOS VIDEOJUEGOS USANDO MACHINE
LEARNING

Tesis para Optar el Título Profesional de Ingeniero de Sistemas e
Informática

Revisado y Aprobado por Jurado Evaluador:

Dr. Juan Pablo Sánchez Chávez
Presidente

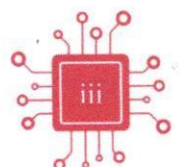
Código ORCID: 0000-0002-3521-7037

Dr. Hugo Esteban Caselli Gismondi
Secretario

Código ORCID: 0000-0002-2812-6727

Ms. Whiston Kendrick Borja Reyna
Miembro

Código ORCID: 0000-0002-5966-3859



ACTA DE EVALUACIÓN PARA SUSTENTACIÓN DE TESIS

En el Campus Universitario de la Universidad Nacional del Santa, siendo las 10:00 a.m. del día viernes 25 de octubre de 2024, en el Aula S2 del Pabellón nuevo de la EPISI, en atención a la Transcripción de Resolución Decanal N° 669-2024-UNS-FI de Declaración de Expedito de fecha 24.10.2024; se llevó a cabo la instalación del jurado Evaluador, designado mediante Transcripción de Resolución N° 562- 2024 -UNS-CFI de fecha 03.09.2024, integrado por el Dr. Juan Pablo Sánchez Chávez (Presidente), Dr. Hugo Esteban Caselli Gismondi (Secretario), Ms. Whiston Kendrick Borja Reyna (Integrante), para dar inicio a la sustentación del Informe Final de Tesis, cuyo título es: "MODELO PREDICTIVO DEL RENDIMIENTO ACADEMICO DE ESTUDIANTES UNIVERSITARIOS ASOCIADO CON LA INFLUENCIA DE LOS VIDEOJUEGOS USANDO MACHINE LEARNING" perteneciente a los Bachilleres: CASTAÑEDA CHÁVEZ DIEGO ALBERTO, con código de matrícula N° 0201514033 y PEREZ CHANG VALERIA ISABEL, con código de matrícula N° 0201614046 y tienen como ASESOR al Dr. Hugo Esteban caselli Gismondi, asesor del PT, designado mediante T/R. D. N° 519-2022-UNS-CFI de fecha 06.09.2022

Terminada la sustentación, el tesista respondió a las preguntas formuladas por los miembros del Jurado Evaluador y el público presente.


El Jurado después de deliberar sobre aspectos relacionados con el trabajo, contenido y sustentación del mismo y con las sugerencias pertinentes y en concordancia con el artículo 71° y 111° del Reglamento General de Grados y Títulos, vigente de la Universidad Nacional del Santa (T/Res. N° 337-2024-CU-R-UNS DEL 12.04.2024); considera la siguiente nota final de Evaluación:

BACHILLER	CALIFICACIÓN	CONDICIÓN
CASTAÑEDA CHÁVEZ DIEGO ALBERTO	19	Excedente

Siendo la 10:50 pm se dio por terminado el Acto de Sustentación y en señal de conformidad, firma el Jurado la presente Acta.

Nuevo Chimbote, 25 de octubre de 2024


DR. JUAN PABLO SANCHEZ CHAVEZ
PRESIDENTE


DR. HUGO ESTEBAN CASELLI GISMONDI
SECRETARIO


MS. WHISTON KENDRICK BORJA REYNA
INTEGRANTE

ACTA DE EVALUACIÓN PARA SUSTENTACIÓN DE TESIS

En el Campus Universitario de la Universidad Nacional del Santa, siendo las 10:00 a.m. del día viernes 25 de octubre de 2024, en el Aula S2 del Pabellón nuevo de la EPISI, en atención a la Transcripción de Resolución Decanal N° 669-2024-UNS-FI de Declaración de Expedito de fecha 21.10.2024; se llevó a cabo la instalación del jurado Evaluador, designado mediante Transcripción de Resolución N° 562- 2024 -UNS- CFI de fecha 03.09.2024, integrado por el Dr. Juan Pablo Sánchez Chávez (Presidente), Dr. Hugo Esteban Caselli Gismondi (Secretario), Ms. Whiston Kendrick Borja Reyna (Integrante), para dar inicio a la sustentación del Informe Final de Tesis, cuyo título es: "MODELO PREDICTIVO DEL RENDIMIENTO ACADEMICO DE ESTUDIANTES UNIVERSITARIOS ASOCIADO CON LA INFLUENCIA DE LOS VIDEOJUEGOS USANDO MACHINE LEARNING" perteneciente a los Bachilleres: CASTAÑEDA CHÁVEZ DIEGO ALBERTO, con código de matrícula N° 0201514033 y PEREZ CHANG VALERIA ISABEL, con código de matrícula N° 0201614046 y tienen como ASESOR al Dr. Hugo Esteban caselli Gismondi, asesor del PT, designado mediante T/R. D. N° 519-2022-UNS-CFI de fecha 06.09.2022



Terminada la sustentación, el tesista respondió a las preguntas formuladas por los miembros del Jurado Evaluador y el público presente.

El Jurado después de deliberar sobre aspectos relacionados con el trabajo, contenido y sustentación del mismo y con las sugerencias pertinentes y en concordancia con el artículo 71° y 111° del Reglamento General de Grados y Títulos, vigente de la Universidad Nacional del Santa (T/Res. N° 337-2024-CU-R-UNS DEL 12.04.2024); considera la siguiente nota final de Evaluación:

BACHILLER	CALIFICACIÓN	CONDICIÓN
PEREZ CHANG VALERIA ISABEL	19	Excelente

Siendo la 10:50 pm se dio por terminado el Acto de Sustentación y en señal de conformidad, firma el Jurado la presente Acta.

Nuevo Chimbote, 25 de octubre de 2024


DR. JUAN PABLO SANCHEZ CHAVEZ
PRESIDENTE
DR. HUGO ESTEBAN CASELLI GISMONDI
SECRETARIO
MS. WHISTON KENDRICK BORJA REYNA
INTEGRANTE



Recibo digital

Este recibo confirma que su trabajo ha sido recibido por **Turnitin**. A continuación podrá ver la información del recibo con respecto a su entrega.

La primera página de tus entregas se muestra abajo.

Autor de la entrega:	Diego Castañeda Chávez
Título del ejercicio:	Tesis
Título de la entrega:	TESIS FINAL - CASTANEDA - PEREZ.pdf
Nombre del archivo:	TESIS_FINAL_-_CASTANEDA_-_PEREZ.pdf
Tamaño del archivo:	1.55M
Total páginas:	110
Total de palabras:	17,693
Total de caracteres:	95,798
Fecha de entrega:	29-oct.-2024 09:09p. m. (UTC-0500)
Identificador de la entrega...	2411759587

UNIVERSIDAD NACIONAL DEL SANTA
FACULTAD DE INGENIERIA
ESCUELA PROFESIONAL DE INGENIERIA DE SISTEMAS E
INFORMÁTICA



UNS
UNIVERSIDAD
NACIONAL DEL SANTA

Modelo predictivo del rendimiento académico de estudiantes
universitarios asociado con la influencia de los videojuegos usando
Machine Learning

Tesis para obtener el Título Profesional de Ingeniero de Sistemas e
Informática

Autores:

Bach. Castañeda Chávez, Diego Alberto
Código ORCID: 0009-0004-8622-8512

Bach. Perez Chang, Valeria Isabel
Código ORCID: 0000-0003-1673-153X

Asesor:

Dr. Caselli Gismondi, Hugo Esteban
DNI N° 32819296
Código ORCID: 0000-0002-2812-6727

NUEVO CHIMBOTE – PERÚ
2024

TESIS FINAL - CASTANEDA - PEREZ.pdf

INFORME DE ORIGINALIDAD

19%

INDICE DE SIMILITUD

16%

FUENTES DE INTERNET

4%

PUBLICACIONES

10%

TRABAJOS DEL
ESTUDIANTE

FUENTES PRIMARIAS

1	repositorio.unsaac.edu.pe Fuente de Internet	2%
2	hdl.handle.net Fuente de Internet	1%
3	repositorio.uns.edu.pe Fuente de Internet	1%
4	Submitted to Universidad Nacional Federico Villarreal Trabajo del estudiante	1%
5	Submitted to Centro Europeo de Postgrado - CEUPE Trabajo del estudiante	1%
6	Submitted to Universidad Internacional de la Rioja Trabajo del estudiante	1%
7	aws.amazon.com Fuente de Internet	1%
8	huru.unsj.edu.ar Fuente de Internet	1%



DEDICATORIA

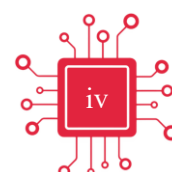
A Dios, por ser mi guía constante, fuente inagotable de fortaleza, y por la sabiduría que me da cada día. Gracias por iluminar mi camino en los momentos de duda, por darme la perseverancia para alcanzar mis metas y por llenar mi vida de propósito y esperanza.

A mis padres, Jorge y Elvia, por su amor incondicional, su apoyo inquebrantable, por enseñarme a nunca rendirme, y por todo lo que me han dado. Gracias por ser mi mayor inspiración y mi sostén en cada paso de este camino.

A mi hermano Gabriel, por ser mi compañero de vida, mi amigo fiel y mi mayor apoyo en los momentos difíciles. Gracias por compartir conmigo risas, retos y sueños. Tu presencia ha hecho de este viaje uno más significativo y lleno de valor.

A mis amigos, por estar a mi lado en cada paso de este camino, por sus risas, consejos y apoyo incondicional. En especial, a mi amiga Valeria, por su amistad sincera y su presencia constante en mi vida. Gracias por ser un pilar fundamental y por hacer de esta experiencia algo inolvidable.

Diego Alberto Castañeda Chávez





DEDICATORIA

Dedico este trabajo a Dios, cuya guía y fortaleza me han permitido superar cada desafío. Su amor y guía han sido mi luz en los momentos de oscuridad, sin él nada de esto habría sido posible.

A mis padres, por su amor, sacrificio y apoyo incondicional, que me motiva a cumplir todos mis logros. A ellos, que siempre han creído en mí, han estado conmigo, me han enseñado el valor del esfuerzo y me han motivado a seguir adelante en cada etapa de mi vida.

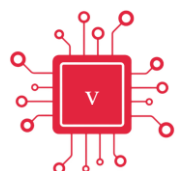
A mis hermanos, por su constante ánimo, locuras, compañía y fe en mí, ellos que son el motor de mi vida para seguir adelante.

A mis amigos, que estuvieron siempre al tanto del desarrollo de este proyecto y me motivaron constantemente para su culminación. En especial, a Diego, cuyo apoyo incondicional y amistad fueron cruciales para superar los desafíos y alcanzar este objetivo.

Valeria Isabel Perez Chang

BACH. CASTAÑEDA CHÁVEZ DIEGO ALBERTO

BACH. PEREZ CHANG VALERIA ISABEL





AGRADECIMIENTO

La culminación de este proyecto de tesis es el resultado de un arduo camino que no hemos recorrido solos. A lo largo de este proceso, hemos contado con el apoyo, la orientación y motivación de muchas personas a quienes deseamos expresar nuestro más profundo agradecimiento.

En primer lugar, agradecemos a Dios, por darnos la fuerza, la sabiduría y la perseverancia para lograrlo.

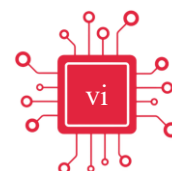
A nuestro asesor, el Dr. Hugo Caselli Gismondi, queremos expresar nuestro más sincero agradecimiento. Su orientación, paciencia y conocimiento han sido fundamentales en la realización de este trabajo. Su apoyo constante y sus valiosas aportaciones nos han permitido crecer académicamente y culminar este proyecto con éxito.

A nuestros amigos, quienes han sido nuestro soporte emocional y motivacional durante todo este proceso. Gracias por su compañía, sus palabras de aliento.

Bach. Castañeda Chávez Diego Alberto y Bach. Perez Chang Valeria Isabel

BACH. CASTAÑEDA CHÁVEZ DIEGO ALBERTO

BACH. PEREZ CHANG VALERIA ISABEL





ÍNDICE GENERAL

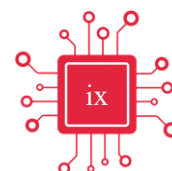
DEDICATORIA.....	iv
AGRADECIMIENTO	vi
RESUMEN	xiv
ABSTRACT.....	xv
CAPÍTULO I	1
INTRODUCCIÓN	1
1.1. DESCRIPCIÓN Y FORMULACIÓN DEL PROBLEMA.....	1
1.1.1. Descripción del problema	1
1.1.2. FORMULACIÓN DEL PROBLEMA.....	9
1.1.3. OBJETIVOS	10
1.2. FORMULACIÓN DE LA HIPÓTESIS.....	10
1.2.1. Hipótesis General.....	10
1.2.2. Hipótesis Específicas	11
1.3. JUSTIFICACIÓN E IMPORTANCIA	11
1.3.1. Justificación	11
CAPÍTULO II	15
MARCO TEÓRICO	15
2.1. ANTECEDENTES	15
2.1.1. Antecedentes Internacionales	15
2.1.2. Antecedentes Nacionales	17
2.1.3. Antecedentes Locales	20
2.2. MARCO CONCEPTUAL	22
2.2.1. Bases Teóricas.....	22
2.2.2. Metodología	33
2.2.3. Contexto Tecnológico	37
CAPÍTULO III	40



METODOLOGÍA	40
3.1. DISEÑO DE INVESTIGACIÓN	40
3.2. MÉTODOLOGÍA DE LA INVESTIGACIÓN	40
3.3. POBLACIÓN	41
3.4. MUESTRA.....	41
3.5. OPERACIONALIZACIÓN DE VARIABLE	42
3.6. TÉCNICAS E INSTRUMENTOS DE RECOLECCIÓN DE DATOS	43
CAPÍTULO IV	44
RESULTADOS	44
4.1. FASE I: Comprensión del Negocio.....	44
4.1.1. Comprensión del contexto y determinación de objetivos	44
4.1.2. Evaluación de la situación	45
4.1.3. Determinación de los objetivos de Machine Learning aplicados al proyecto	46
4.1.4. Desarrollo del plan del proyecto	47
4.2. FASE II: Comprensión de los Datos	48
4.2.1. Obtención de datos preliminares.....	48
4.2.2. Identificación de los datos	52
4.2.3. Análisis preliminar de los datos.....	53
4.2.4. Evaluación de la calidad de los datos	69
4.3. FASE III: Preparación de los datos	70
4.3.1. Selección de los datos relevantes	70
4.3.2. Limpieza de datos	71
4.3.3. Transformación de datos.....	74
4.3.4. Consolidación de datos	76
4.3.5. Reducción de la dimensionalidad de la encuesta.....	77
4.4. FASE IV: Modelado.....	79
4.4.1. Modelados por Regresión y métricas de comprobación	79



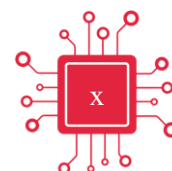
4.4.2.	Modelados por Clasificación y métricas de comprobación	87
4.4.3.	Comparación de modelos y selección del optimo	95
4.5.	FASE V: Evaluación	96
4.5.1.	Evaluación de resultados de Modelo de Regresión	96
4.5.2.	Evaluación de resultados de Modelo de Clasificación	98
4.6.	FASE VI: Despliegue.....	100
CAPÍTULO V		101
DISCUSIÓN		101
5.1.	Contrastación de la hipótesis general.....	103
5.2.	Contrastación de las hipótesis específicas	103
CAPÍTULO VI		106
CONCLUSIONES Y RECOMENDACIONES.....		106
6.1.	CONCLUSIONES	106
6.2.	RECOMENDACIONES.....	108
CAPÍTULO VII		110
REFERENCIAS BIBLIOGRÁFICAS		110
CAPÍTULO VIII		114
ANEXOS		114





ÍNDICE DE TABLAS

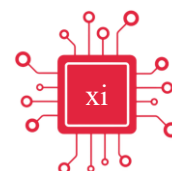
Tabla 1: Operacionalización de Variables	42
Tabla 2: Plan del proyecto	47
Tabla 3: Evaluación del modelo de Regresión Linear	79
Tabla 4: Valores reales vs valores predichos del modelo de Regresión Lineal	81
Tabla 5: Evaluación del modelo de Regresión Múltiple.....	82
Tabla 6: Valores reales vs valores predichos del modelo de Regresión Lineal	84
Tabla 7: Evaluación del modelo de Random Forest	85
Tabla 8: Valores reales vs valores predichos del modelo Random Forest	87
Tabla 9: Evaluación del modelo de Regresión Logística.....	88
Tabla 10: Evaluación del modelo de Naive Bayes	92
Tabla 11: Comparación de modelos y selección del óptimo.....	95
Tabla 12: Margen de error de dataset de prueba del modelo Random Forest.....	97
Tabla 13: Margen de error de dataset de prueba del modelo Naive Bayes.....	99





ÍNDICE DE GRÁFICAS

Gráfica 1: Tiempo de consumo de videojuegos	3
Gráfica 2: Egresados que recibieron atención de consejería y/o tutoría	7
Gráfica 3: Gráfico de barras: Pregunta 01	53
Gráfica 4: Gráfico de barras: Pregunta 02	54
Gráfica 5: Gráfico de barras: Pregunta 03	54
Gráfica 6: Gráfico de barras: Pregunta 04	55
Gráfica 7: Gráfico de barras: Pregunta 05	56
Gráfica 8: Gráfico de barras: Pregunta 06	56
Gráfica 9: Gráfico de barras: Pregunta 07	57
Gráfica 10: Gráfico de barras: Pregunta 08	57
Gráfica 11: Gráfico de barras: Pregunta 09	58
Gráfica 12: Gráfico de barras: Pregunta 10	59
Gráfica 13: Gráfico de barras: Pregunta 11	59
Gráfica 14: Gráfico de barras: Pregunta 12	60
Gráfica 15: Gráfico de barras: Pregunta 13	61
Gráfica 16: Gráfico de barras: Pregunta 14	61
Gráfica 17: Gráfico de barras: Pregunta 15	62
Gráfica 18: Gráfico de barras: Pregunta 16	63
Gráfica 19: Gráfico de barras: Pregunta 17	64
Gráfica 20: Gráfico de barras: Pregunta 18	64
Gráfica 21: Gráfico de barras: Pregunta 19	65
Gráfica 22: Gráfico de barras: Pregunta 20	66
Gráfica 23: Gráfico de barras: Pregunta 21	66
Gráfica 24: Gráfico de barras: Pregunta 22	67
Gráfica 25: Gráfico de barras: Pregunta 23	68
Gráfica 26: Gráfico de barras: Pregunta 24	68





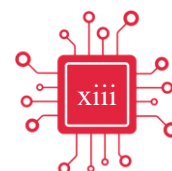
ÍNDICE DE FIGURAS

Figura 1: Consumo de videojuegos a través de descargas e internet	4
Figura 2: Metadatos del dataset	71
Figura 3: Metadatos del dataset de las columnas seleccionadas	72
Figura 4: Metadatos del dataset limpio.....	73
Figura 5: Vista Previa del dataset con variables categóricas	74
Figura 6: Vista Previa del Dataset con variables numéricas.....	74
Figura 7: Transformación de valores categóricos a numéricos según la escala de Likert en Python	75
Figura 8: Generación de dimensiones en Python	76
Figura 9: Metadata de Dataset final	77
Figura 10: Vista previa de Dataset final.....	77
Figura 11: Mapa de Calor entre las 6 dimensiones y el promedio ponderado de los estudiantes.....	78
Figura 12: Valores Reales vs Predicciones del Modelo de Regresión Lineal.....	80
Figura 13: Valores Reales vs Predicciones del Modelo de Regresión Múltiple	83
Figura 14: Valores Reales vs Predicciones del Modelo Random Forest	86
Figura 15: Matriz de confusión del modelo de la Regresión Logística.....	89
Figura 16: Matriz de confusión del modelo de Naive Bayes.....	92
Figura 17: Código Python para generar margen de error del modelo Random Forest .	96
Figura 18: Código Python para generar margen de error del modelo Naive Bayes	98



ÍNDICE DE ANEXOS

Anexo 1: Cuestionario.....	114
Anexo 2: Validación de Instrumento por Licenciada en Psicología.....	116
Anexo 3: Validación de Instrumento por Médico Psiquiatra	121
Anexo 4: Validación de Instrumento por Docente Universitario	126
Anexo 5: Solicitud de información a rectora de la UNS.....	131
Anexo 6: Derivación de solicitud de información al área encargada.....	132
Anexo 7: Muestra de la información recibida	133
Anexo 8: Código Python del Modelo de Regresión Lineal.....	134
Anexo 9: Código Python del Modelo de Regresión Múltiple	136
Anexo 10: Código Python del Modelo Random Forest	138
Anexo 11: Código Python del Modelo de Regresión Logística	140
Anexo 12: Código Python del Modelo Naive Bayes	141





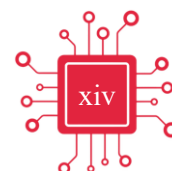
RESUMEN

El presente trabajo de investigación tiene como objetivo determinar en qué medida el rendimiento académico de los estudiantes universitarios puede ser predicho a partir del consumo de videojuegos utilizando técnicas de Machine Learning. La idea surgió a partir de las vivencias anecdóticas y conversaciones con estudiantes de la Escuela de Ingeniería de Sistemas de la Universidad Nacional del Santa y se sustentó con los diversos antecedentes e información encontrada en relación con el tema de investigación, la muestra es de tipo no probabilístico e intencional, se escogieron a los estudiantes matriculados en el ciclo académico 2023-I de la Escuela Profesional de Ingeniería de Sistemas e Informática, lo cual según lo recopilado dan un total aproximado de 300 estudiantes.

Utilizando modelos predictivos basados en seis dimensiones clave ("Capacidad Económica", "Tiempo de Consumo de Videojuegos", "Bienestar Físico-Mental", "Impacto sobre Actividades Cotidianas", "Relaciones Sociales" y "Desarrollo Cognitivo"), se obtuvieron resultados significativos. Los modelos óptimos, Random Forest y Naive Bayes, demostraron ser particularmente efectivos. El modelo de Random Forest alcanzó un porcentaje de predicción del 87.5%, mientras que el modelo de Naive Bayes alcanzó un 93%. Estos resultados confirman la hipótesis del estudio, mostrando que el rendimiento académico puede ser predicho con una precisión superior al 80%.

Estos hallazgos subrayan que los estudiantes con un mayor consumo de videojuegos tienden a presentar una mayor variabilidad y, en general, un rendimiento académico inferior. Por lo tanto, es crucial implementar intervenciones y programas de orientación que ayuden a los estudiantes a equilibrar adecuadamente el tiempo dedicado a los videojuegos y al estudio, promoviendo un uso saludable para optimizar su rendimiento académico.

Palabras clave: RENDIMIENTO ACADÉMICO, VIDEOJUEGOS, MACHINE LEARNING, ESTUDIANTES UNIVERSITARIOS, MODELO PREDICTIVO





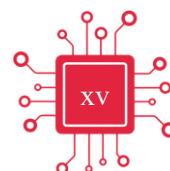
ABSTRACT

The present research work aims to determine to what extent the academic performance of university students can be predicted from the consumption of video games using Machine Learning techniques. The idea arose from the anecdotal experiences and conversations with students of the School of Systems Engineering of the Universidad Nacional del Santa and was supported by the various background and information found in relation to the research topic, the sample is non-probabilistic and intentional, students enrolled in the academic year 2023-I of the Professional School of Systems Engineering and Computer Science were chosen, which according to what was collected gives a total of approximately 300 students.

Using predictive models based on six key dimensions ("Economic Capacity", "Video Game Consumption Time", "Physical-Mental Well-Being", "Impact on Daily Activities", "Social Relationships" and "Cognitive Development"), significant results were obtained. The optimal models, Random Forest and Naive Bayes, proved to be particularly effective. The Random Forest model achieved a prediction rate of 87.5%, while the Naive Bayes model achieved 93%. These results confirm the hypothesis of the study, showing that academic performance can be predicted with an accuracy of over 80%.

These findings underline that students with higher video game consumption tend to present greater variability and, in general, lower academic performance. Therefore, it is crucial to implement interventions and guidance programs that help students to properly balance time spent playing video games and studying, promoting a healthy use to optimize their academic performance.

Keywords: ACADEMIC PERFORMANCE, VIDEO GAMES, MACHINE LEARNING, UNIVERSITY STUDENTS, PREDICTIVE MODELING





CAPÍTULO I

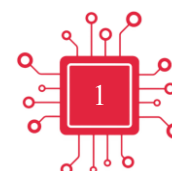
INTRODUCCIÓN

1.1. DESCRIPCIÓN Y FORMULACIÓN DEL PROBLEMA

1.1.1. Descripción del problema

La tecnología ha revolucionado nuestra calidad de vida, siendo la educación un aspecto relevante que se encuentra en constante cambio gracias a los recursos y herramientas que estos avances le ofrecen, siendo así, que en la mayoría de los casos la integración de la tecnología en la educación influye potencialmente en el rendimiento académico de los estudiantes. Al mencionar “rendimiento académico” para Tacilla Cardenas et al. (2020) es un constructo que va desde lo básico, complejo y multidimensional, presente en el proceso de enseñanza-aprendizaje; permitiendo reconocer el avance académico-intelectual del estudiante y emitir un juicio de valor. Lo cual conlleva a crear metodologías adecuadas para la obtención de conocimiento y fortalecer el rendimiento académico de los estudiantes.

Por otro lado, el avance constante y acelerado de la tecnología ha transformado significativamente el sector del entretenimiento. Los videojuegos, en particular, han emergido como una actividad predominante entre los estudiantes, generando un intenso debate sobre su impacto en el rendimiento académico. Según Rivera & Torres (2018) afirman que: “Los videojuegos ayudan a los usuarios a desarrollar habilidades del pensamiento, competencias y además generan una diversidad de conocimientos. Sin embargo, también existen desventajas, como el uso





excesivo (adicción) y en lo físico el sedentarismo, lo que puede llevar a producir diversos efectos negativos en los estudiantes”.

Alrededor del mundo, estos conceptos han sido motivo de diversas investigaciones, por ejemplo, en un estudio reciente realizado por Chen (2024) a 45 estudiantes de posgrado de una prestigiosa universidad taiwanesa, se examinó la relación entre la adicción a los videojuegos y el rendimiento académico mediante una prueba de Chi-cuadrado, usando las categorías de calificación GPA. Los estudiantes autoevaluaron sus niveles de adicción y se encontró una asociación significativa ($p = 0,013$) entre estos niveles y sus calificaciones. Los resultados mostraron que los estudiantes no adictos tienen un rendimiento académico consistente, mientras que los niveles de adicción más altos se correlacionaban con una mayor variabilidad y un rendimiento académico inferior.

Así mismo, según Araúz M. (2020) quien realizó una investigación enfocada a una muestra escogida por conveniencia de 23 estudiantes de dos universidades de la provincia de Chiriquí en Panamá en su primer año de estudios. Estos estudiantes respondieron un cuestionario de forma voluntaria y honesta referente a la relación existente entre los videojuegos y su rendimiento académico en su primer año de universidad. Dando como resultado un importante índice estadístico, el cual indica que la mayoría de los jóvenes (un 61%) dedican más de 4 horas al uso de videojuegos como se muestra en la *Gráfica 1*, comentando además que dicha actividad les consume mucho tiempo y sienten una gran satisfacción al realizarla. En lo que se refiere al rendimiento académico, estos jóvenes en su mayoría obtuvieron una baja calificación en puntualidad, entregar sus tareas a

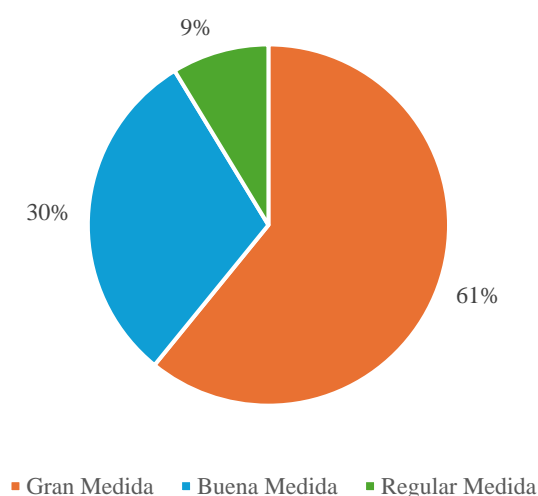




tiempo y obtener buenas calificaciones. Al calcular la correlación entre el uso de los videojuegos y el rendimiento académico utilizando la prueba de Chi-cuadrado, se halló que hay una significativa relación entre estas variables ($p = 0,000$), comprobando el grado de correlación entre ambas.

Gráfica 1:

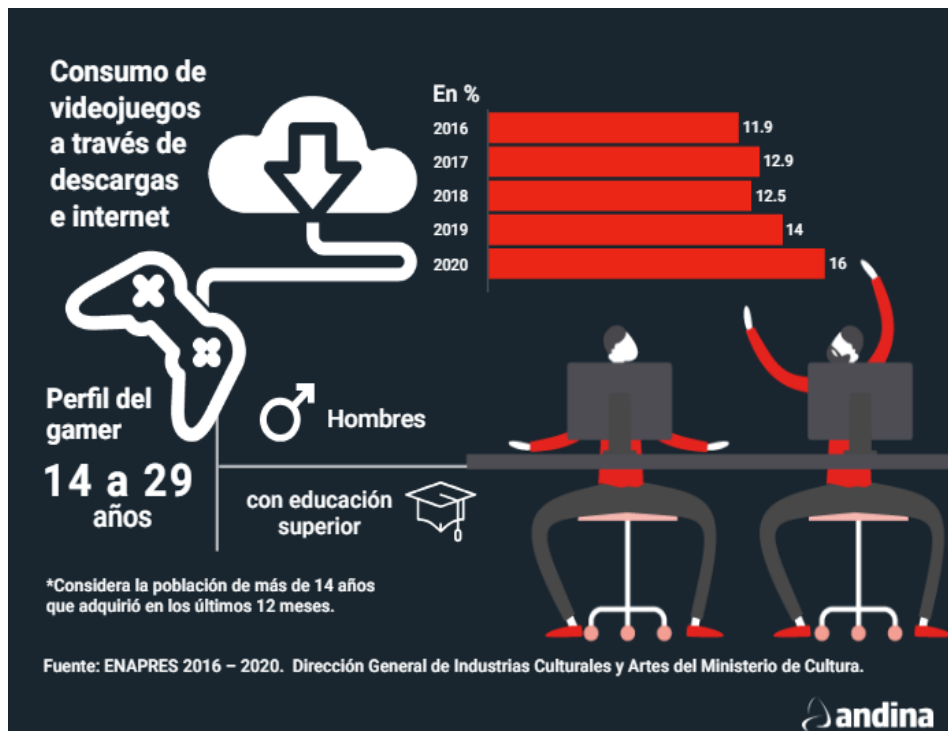
Tiempo de consumo de videojuegos



En el Perú, los especialistas del Hospital Víctor Larco Herrera (HVLH), perteneciente al Ministerio de Salud (Minsa), han observado un incremento en la adicción a los videojuegos entre niños, adolescentes y jóvenes desde el comienzo de la pandemia de COVID-19. Esto se atribuye al confinamiento y a la escasez de oportunidades para socializar con individuos de su misma edad. Así mismo, la empresa peruana de servicios editoriales S. A. EDITORA PERÚ (2021), publicó un artículo en donde mostraban el porcentaje del consumo de videojuegos incrementado desde el año 2016 hasta el 2020 como se muestra en la *Figura 01* alegando que más de 600 mil personas adquirieron al menos 01 videojuego en ese último año según la Encuesta Nacional de Programas Presupuestales (ENAPRES)

Figura 1:

Consumo de videojuegos a través de descargas e internet



Nota. Extraído de *Consumo de videojuegos en el Perú creció durante la pandemia*, por Empresa Peruana de Servicios Editoriales S. A. EDITORA PERÚ, 2021 (<https://andina.pe/agencia/noticia-consumo-videojuegos-el-peru-crecio-durante-pandemia-859350.aspx>)

Como se menciona en párrafos anteriores, este fenómeno podría tener repercusiones en el rendimiento académico: por un lado, la excesiva dedicación a los videojuegos podría distraer a los estudiantes y reducir el tiempo destinado al estudio; por otro lado, es posible que el desarrollo de ciertas habilidades cognitivas y de resolución de problemas asociadas con los videojuegos tenga un efecto positivo en el aprendizaje académico.

En el año 2022, en la Universidad Privada de Chiclayo, el estudio realizado por Ludeña a 272 estudiantes de psicología reveló distintos niveles de adicción a los videojuegos: bajo (43.01%), medio (37.87%) y alto (19.12%). Ludeña logró identificar una correlación significativa entre el consumo de



videojuegos y la calidad del sueño, demostrada mediante un coeficiente de correlación de Spearman con un valor de $P < 0.05$. Este estadístico subraya la influencia de la adicción a los videojuegos en el bienestar psicosocial de los estudiantes, destacando su impacto en él.

Así mismo, en un estudio realizado a adolescentes de cinco instituciones educativas de diferentes distritos de la ciudad de Lima el cual intenta validar el Test de dependencia de videojuegos (Choliz & Marco, 2011) concluyó que los adolescentes de género masculino son los más propensos a un mayor consumo de videojuegos, siendo la edad un dato trivial para la investigación.

Tras los estudios referidos anteriormente, se sugiere una posible correlación entre el consumo de videojuegos y el rendimiento académico. En el contexto de la Escuela Profesional de Ingeniería de Sistemas e Informática (EPISI) de la Universidad Nacional del Santa (UNS), además de estos estudios, se han observado patrones similares a través de vivencias anecdóticas y conversaciones con compañeros. Estas interacciones han revelado que muchos estudiantes perciben que el consumo excesivo de videojuegos puede tener un impacto en su rendimiento académico.

Las historias compartidas por los estudiantes de EPISI sugieren que el tiempo dedicado a los videojuegos puede desplazar el tiempo destinado al estudio, afectando así el desempeño en exámenes y la entrega de trabajos. Además, algunos compañeros han mencionado que los videojuegos pueden ser una fuente de distracción que interfiere con la concentración y la productividad académica.





Con el fin de validar estas observaciones anecdóticas y obtener una comprensión más precisa de la relación entre el consumo de videojuegos y el rendimiento académico en nuestro contexto, se decidió llevar a cabo una encuesta. Esta encuesta fue aplicada a 86 egresados de las promociones de ingreso 2015 y 2016 de la EPISI. A través de este estudio, se busca cuantificar la percepción y las experiencias de los egresados respecto a cómo los videojuegos pueden haber influido en su rendimiento académico, proporcionando así datos concretos que puedan ser analizados y comparados con los resultados de investigaciones previas, teniendo como resultado que el 77 % de los egresados jugaba a videojuegos, distribuyéndose en los niveles de la escala de Likert de la siguiente manera: Normal (24 %), Leve (27 %), Moderado (48 %) y Severo (1 %). Al disponer de estos datos preliminares, hemos podido identificar la prevalencia del consumo de videojuegos entre los estudiantes de la EPISI, lo cual nos proporciona una base sólida para profundizar en nuestra investigación. Estos resultados indican que una mayoría significativa de los estudiantes dedica tiempo a los videojuegos, con diferentes niveles de intensidad según la escala de Likert. Esta información es crucial, ya que nos permite comprender la magnitud del fenómeno dentro de nuestra población de estudio.

Adicionalmente, esta encuesta permite diferenciar 7 dimensiones que influyen en la actitud de los encuestados (Relevancia, Tolerancia, Cambios en el estado de ánimo, Recaída, Retirada, Conflicto y Problemas), que tienen una alta relación con el rendimiento académico.



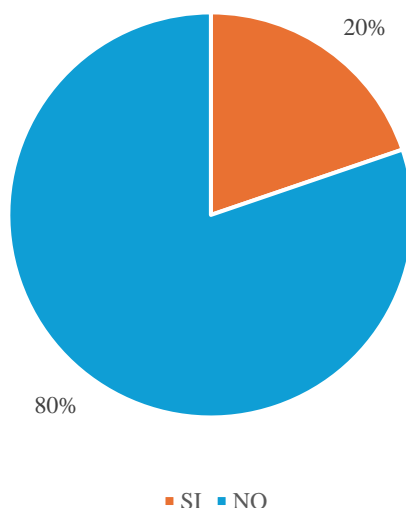


Como se mencionó, los resultados de la encuesta revelaron un alto nivel de consumo de videojuegos en la EPISI, lo que puede tener tanto efectos positivos como negativos en el rendimiento académico de los estudiantes.

En este caso, como se observa en la Gráfica 2, un 80% alega que no ha recibido atención de consejería y tutoría. Esta carencia puede atribuirse a varios factores, como la falta de conocimiento sobre el tema, recursos limitados y la escasa colaboración entre las distintas áreas involucradas.

Gráfica 2:

Egresados que recibieron atención de consejería y/o tutoría



De acuerdo con el reglamento de tutoría académica y consejería de pregrado de la Universidad Nacional del Santa, aprobado mediante la resolución N° 377-2021-CU-R-UNS, se brinda orientación y acompañamiento a los estudiantes. Sin embargo, según como reflejan los resultados de la encuesta, la mayoría de los participantes no consideran haber recibido este beneficio. La encuesta arroja datos significativos sobre el uso de videojuegos entre los egresados universitarios, evidenciando que esta actividad pudo acarrear una

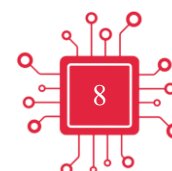


serie de desafíos, como la reducción del tiempo dedicado al estudio, distracciones durante las horas académicas y una disminución en la calidad del sueño. Sin embargo, es importante reconocer que no todos los videojuegos tienen un impacto negativo. De hecho, algunos de ellos pueden contribuir de manera positiva al desarrollo de habilidades cognitivas y sociales, ofreciendo oportunidades para la resolución de problemas, el trabajo en equipo y la creatividad.

Aun así, ante este panorama, es esencial mantener un seguimiento continuo del rendimiento académico de los estudiantes, es por ello que se propone implementar medidas de consejería y tutoría adaptadas a su consumo de videojuegos, esto a su vez puede ser replicada para cualquier otra actividad perjudicial en la que se exceda el estudiante. Sin un seguimiento constante, existe el riesgo de pasar por alto las señales tempranas de dificultades académicas debido al consumo de videojuegos, lo que podría desembocar en problemas más graves a largo plazo.

En atención al diagnóstico de la problemática, y con el objetivo de proyectar el rendimiento académico de los estudiantes de la Escuela Profesional de Ingeniería de Sistemas e Informática (EPISI), esta investigación se orienta a encontrar posibles soluciones.

Se desarrollará un modelo predictivo que analizará la influencia de los videojuegos en el rendimiento académico. Este modelo se basará en los datos recopilados con el objetivo de identificar a los estudiantes en riesgo. Se correlacionarán diversas dimensiones asociadas al consumo de videojuegos que puedan afectar negativamente el rendimiento académico.





El modelo proporcionará información valiosa y de gran utilidad para la Dirección de Bienestar Universitario, ya que permitirá diseñar e implementar estrategias orientadas a mitigar el consumo excesivo de videojuegos y reducir la tasa de desaprobación entre los alumnos.

1.1.2. FORMULACIÓN DEL PROBLEMA

a. Problema General

Teniendo en cuenta lo anterior se formuló el siguiente enunciado interrogativo: ¿En qué medida puede el rendimiento académico de los estudiantes universitarios ser predicho a partir del consumo de videojuegos utilizando técnicas de Machine Learning?

b. Problemas Específicos

- ¿Cuál es la relación entre el consumo de videojuegos por parte de los estudiantes universitarios con su rendimiento académico?
- ¿Cómo afecta la capacidad económica de los estudiantes universitarios al consumo de videojuegos?
- ¿Qué algoritmo de Machine Learning es el más eficiente para predecir el rendimiento académico de los estudiantes de la Escuela Profesional de Ingeniería de Sistemas e Informática (EPISI) en base al consumo de videojuegos?
- ¿Cuáles son los factores de entrada clave para el modelo predictivo que predecirán el rendimiento académico de los estudiantes universitarios en base al consumo de videojuegos?
- ¿Es conveniente usar la metodología CRISP-DM para el desarrollo del Modelo Predictivo?



1.1.3. OBJETIVOS

a. Objetivo General

Determinar en qué medida el rendimiento académico de los estudiantes universitarios puede ser predicho a partir del consumo de videojuegos utilizando técnicas de Machine Learning.

b. Objetivos Específicos

- Analizar la relación entre el tiempo de consumo de videojuegos por parte de los estudiantes universitarios, para predecir su rendimiento académico.
- Evaluar si el consumo de videojuegos se ve afectado por la capacidad económica de los estudiantes universitarios.
- Elegir el algoritmo de Machine Learning más eficiente para predecir el Rendimiento Académico de los estudiantes de la EPISI en base al consumo de videojuegos.
- Identificar los factores de entrada clave para el modelo predictivo que predecirán el rendimiento académico de los estudiantes universitarios en base al consumo de videojuegos
- Usar la metodología CRISP-DM para el desarrollo del Modelo Predictivo.

1.2. FORMULACIÓN DE LA HIPÓTESIS

1.2.1. Hipótesis General

El consumo de videojuegos tiene una influencia significativa en el rendimiento académico de los estudiantes universitarios, y utilizando



técnicas de Machine Learning, es posible predecir dicha influencia con una precisión de más del 80%.

1.2.2. Hipótesis Específicas

- Existe una correlación significativa entre el tiempo de consumo de videojuegos por parte de los estudiantes universitarios con el rendimiento académico.
- La capacidad económica de los estudiantes universitarios afecta al consumo de videojuegos
- Se elige el algoritmo de Machine Learning más eficiente para predecir el Rendimiento Académico de los estudiantes de la EPISI en base al consumo de videojuegos.
- Se logra identificar los factores de entrada clave del modelo predictivo para predecir rendimiento académico de los estudiantes universitarios en base al consumo de videojuegos.
- El uso de la metodología CRISP-DM facilita el desarrollo del modelo predictivo.

1.3. JUSTIFICACIÓN E IMPORTANCIA

1.3.1. Justificación

a. Justificación Social:

- Mejorar la imagen de la organización educativa “Universidad Nacional del Santa”
- Fortalecer la relación con los estudiantes mediante una atención de calidad hacia sus problemas o necesidades.



- Se busca mejorar la Oficina de Bienestar Estudiantil mediante la integración de psicólogos y docentes para atender a alumnos con bajo rendimiento académico.

b. Justificación Teórica

- Este proyecto de investigación sirve como base de conocimiento para futuras investigaciones.
- El uso de Machine Learning en lugar de un análisis estadístico tradicional en este proyecto de investigación se justifica por su capacidad para manejar grandes volúmenes de datos y detectar patrones complejos no lineales que los métodos estadísticos podrían pasar por alto. Además, Machine Learning permite la actualización y adaptación continua del modelo, mejorando su precisión con el tiempo, y ofrece predicciones personalizadas basadas en características individuales.

c. Justificación Metodológica

- En este estudio, el uso de la metodología CRISP-DM garantiza que la investigación sea transparente y se realice con las mejores prácticas para reproducirse en futuras investigaciones promoviendo un enfoque metodológico estandarizado y eficiente.
- La metodología para este estudio se justifica por su estructura organizada, flexibilidad y enfoque integral en la preparación de los datos, modelado, evaluación y despliegue, facilitando el análisis, así como garantizando la calidad los resultados obtenidos y respaldando la aplicabilidad en otros contextos prácticos.



d. Justificación Tecnológica:

- Promover la adopción y usabilidad de tecnologías de la información y comunicación en el ámbito académico
- Brinda una herramienta para mejorar estrategias educativas y de apoyo a universitarios.

e. Justificación Práctica

- La flexibilidad del modelo permite su aplicación en otros contextos, como la influencia de factores socioeconómicos, hábitos de estudio, uso de tecnologías digitales, actividades extracurriculares, e incluso aspectos relacionados con la salud mental y el bienestar emocional de los estudiantes. Cada uno de estos ámbitos puede ser adaptado en el modelo predictivo mediante la recolección y el análisis de datos específicos, proporcionando así una herramienta robusta para la identificación de patrones y la elaboración de intervenciones personalizadas.

1.3.2. Importancia

El presente proyecto de investigación tiene como finalidad el desarrollo de un modelo predictivo usando Machine Learning para proyectar el rendimiento académico de los estudiantes universitarios en función a la influencia de los videojuegos. Este modelo facilitará una mejor relación entre la UNS y sus estudiantes, permitiendo diseñar intervenciones específicas para promover un equilibrio saludable entre el consumo de videojuegos y el rendimiento académico de los estudiantes universitarios.



Asimismo, proporcionará información valiosa para que el personal correspondiente y los docentes tomen decisiones informadas. La aplicación de este modelo predictivo contribuirá a una mejor gestión en el apoyo estudiantil, y fomentará un entorno educativo que se adapte a las necesidades y hábitos contemporáneos de los estudiantes, para mejorar su bienestar y su éxito académico.



CAPÍTULO II

MARCO TEÓRICO

2.1.ANTECEDENTES

2.1.1. Antecedentes Internacionales

a. Artículo Científico 01

Autor : Fabito S., et al.

Título : Exploring Mobile Game Addiction, Cyberbullying, and its Effects on Academic Performance among Tertiary Students in one University in the Philippines

Revista : IEEE

Año : 2018

Resumen u Objetivo

Según (Fabito y otros, 2018) en su trabajo de investigación desarrollada en la Universidad Nacional de Filipinas determinó si la adicción a los videojuegos móviles y el ciberacoso afectan negativamente el rendimiento académico. El investigador consideró como muestra de estudio a 85 estudiantes de la Universidad Nacional de Filipinas de los ciclos I y II, de los cuales se obtuvo su información a través de dos cuestionarios (Adicción a los Juegos y Ciberacoso) usando de Google Forms. Del resultado en la regresión, sólo se aceptaron la hipótesis H5 (Abstinencia por los videojuegos) y la hipótesis H19 (Victimización del ciberacoso) con un nivel de significancia menor a 0.05. Tras la investigación se concluye que la adicción a los videojuegos móviles y el



ciberacoso no son factores causales del rendimiento académico de los estudiantes.

Correlación

Este trabajo de investigación ayudará en nuestro estudio a poder establecer algunas de las dimensiones de nuestra variable independiente (influencia de los videojuegos).

b. Artículo Científico 02

Autor : Xu X., et al.

Título : Prediction of academic performance associated with internet usage behaviors using machine learning algorithms.

Revista : Computers in Human Behavior

Año : 2019

Resumen u Objetivo

(Xu y otros, 2019) en su trabajo de investigación desarrollada en la Universidad de Beijing reveló la asociación entre los comportamientos de uso de Internet y el rendimiento académico, y predecir el rendimiento académico de los estudiantes universitarios a partir de los datos de uso mediante aprendizaje automático. El investigador consideró como muestra de estudio a 4000 estudiantes de la Universidad de Beijing del semestre de otoño del 2016, de los cuales se obtuvo su información a través Técnicas de extracción, transformación y carga de datos. Obteniéndose que los valores más altos de precisión media para la



Máquina de Vectores de soporte fueron $72,75\% \pm 2,14\%$ (Media \pm SD) y $69,55\% \pm 1,38\%$ (Media \pm SD) de los dos experimentos, respectivamente. Estos valores fueron un 9,20%-10,45% y un 0,85%-1,80% superiores a las precisiones con el Árbol de Decisión y la Red Neuronal, respectivamente. Tras la investigación se concluye que existen algunas diferencias significativas de los comportamientos en Internet entre los grupos de rendimiento académico. Además, las características de los comportamientos en Internet están significativamente correlacionadas con el rendimiento académico.

Correlación

Este antecedente, debido a su relevancia y metodología asociada, nos proporcionará una base sólida para relacionar de manera efectiva nuestras variables de investigación en nuestro estudio, que son esenciales para un análisis profundo y significativo de nuestros hallazgos".

2.1.2. Antecedentes Nacionales

a. Tesis 01

Autor : Candia Oviedo, Dennis Iván

Título : Predicción del rendimiento académico de los estudiantes de la UNSAAC a partir de sus datos de ingreso utilizando algoritmos de aprendizaje automático

Institución : Universidad Nacional de San Antonio Abad del Cusco

Grado : Maestría en Ciencias mención Informática

Año : 2019



Resumen u Objetivo

(Candia, 2019) en su trabajo de investigación predijo el rendimiento académico utilizando algoritmos de aprendizaje automático. El investigador estimó una muestra probabilística conformada por 12,698 estudiantes ingresantes desde el semestre 2014-I hasta 2018-I, de quienes se obtuvo su información a través de las bases de datos del Centro de Cómputo de la UNSAAC adicionando técnicas de extracción, transformación y carga de datos. Se determinó que el algoritmo de árboles de decisión Random Forest, es el que tuvo el mayor porcentaje de predicción acertada con un total de 69.35%. Tras la investigación se concluye que, si se puede llegar a predecir el rendimiento académico a través de los datos de ingreso o de admisión a la UNSAAC, utilizando algoritmos de aprendizaje automático hasta en un 69% de efectividad.

Correlación

Este trabajo de investigación nos permitirá definir nuestra metodología de desarrollo, denominada CRISP-DM (Cross-Industry Standard Process for Data Mining), brindándonos una estructura sólida y documentada para abordar nuestra investigación.



b. Tesis 02

Autor : Vega García, Javier Fernando

Título : Modelo de pronóstico de rendimiento académico de alumnos en los cursos del programa de estudios básicos de la Universidad Ricardo Palma usando algoritmos de Machine Learning

Institución : Universidad Ricardo Palma

Grado : Maestría en Ciencia de los Datos

Año : 2019

Resumen u Objetivo

(Vega, 2019) a través de su trabajo de investigación pronóstico la cantidad de alumnos aprobados y desaprobados mediante el uso de algoritmos de Machine Learning. El investigador, consideró una población de 9118 estudiantes pertenecientes a los períodos 2015-I al 2019-0, de quienes se obtuvo su información aplicando técnicas de recolección de datos. Obteniéndose una diferencia del 4.9% de estudiantes aprobados y un 28% de alumnos desaprobados en relación con la predicción y los datos reales. Al concluir la investigación se demostró que si es posible aplicar un modelo de Machine Learning para la determinación del rendimiento académico.

Correlación

Esta tesis desempeñará un papel muy fundamental a nuestra tesis, ya que nos proporcionará una guía sólida en relación con nuestra variable dependiente (rendimiento académico), teniendo en cuenta el contexto específico en el que se desarrolla nuestra investigación.



2.1.3. Antecedentes Locales

a. Tesis 03

Autor : Espinoza, G. & León, E.

Título : Modelo de Machine Learning para la clasificación de estudiantes de acuerdo con su rendimiento académico en el Centro de Idiomas de la Universidad Nacional del Santa

Institución : Universidad Nacional del Santa

Grado : Ingeniero de Sistemas e Informática

Año : 2020

Resumen u Objetivo

(Espinoza, 2020) en su trabajo de investigación logró mejorar el proceso de clasificar a los estudiantes del centro de idiomas utilizando Aprendizaje Automático. Los investigadores tomaron como muestra a los estudiantes de idiomas matriculados en el mes de diciembre del 2018, de quienes se consiguió su información a través técnicas de extracción, transformación y carga de datos. Obteniéndose como resultado la reducción del Tiempo Promedio de clasificación de los estudiantes según su rendimiento académico en un 74.60% (de 218.19 segundos a 55.42 segundos). Tras la investigación se concluye que el modelo de Machine Learning facilitó al personal del CEIDUNS en la identificación de estudiantes, asignación de aulas e incrementando el número docentes.

Correlación

Este trabajo de investigación jugará un papel importante en la orientación de la selección de modelos predictivos más resaltantes para considerar en nuestro trabajo de investigación.



b. Tesis 04

Autor : Caselli Gismondi, Hugo Esteban

Título : Modelo predictivo basado en Machine Learning como soporte para el seguimiento académico del estudiante universitario

Institución : Universidad Nacional del Santa

Grado : Doctor en Ingeniería de Sistemas e Informática

Año : 2021

Resumen u Objetivo

(Caselli, 2021) en su trabajo de investigación obtuvo un modelo predictivo basado en Machine Learning para mejorar la gestión del seguimiento académico. El investigador consideró como población a los estudiantes de la Universidad Nacional del Santa, de los cuales se obtuvo su información a través de las diferentes oficinas de información de la Universidad Nacional del Santa, aplicando técnicas de extracción, transformación y carga de datos. Obteniendo la validación de que el modelo del experimento N° 13, tiene una menor diferencia de precisión entre el conjunto de entrenamiento (98.97%) y el conjunto de prueba (81,73%) de 14,24 puntos porcentuales. Tras la investigación se concluye que al aplicar el modelo de predicción basado en Machine Learning, se logró mejorar el seguimiento académico de los estudiantes.

Este trabajo de investigación permitirá tener una idea más clara referente al diseño del modelo predictivo.



Correlación

Esta tesis nos proporcionará una visión más clara sobre los criterios utilizados para seleccionar la población y muestra en nuestra investigación, dado que se llevará a cabo en la misma entidad.

2.2.MARCO CONCEPTUAL

2.2.1. Bases Teóricas

a. Rendimiento Académico

El rendimiento académico es considerado “un indicador y una medida del desempeño académico de los estudiantes y una meta central de la educación” (Tacilla Cardenas y otros, 2020)

(Tacilla Cardenas y otros, 2020) asimismo, lo define como el producto de la interacción educativa y pedagógica entre profesores y alumnos. El cual se obtiene durante un período de aprendizaje específico, en el que se realiza una evaluación cuantitativa y cualitativa de alcanzar los objetivos planteados. (p. 56)

- Dimensiones del Rendimiento Académico

Dimensión Académica

(Tacilla Cardenas y otros, 2020) lo define como un proceso en el que los estudiantes se desarrollan académica y creativamente. Se refiere a variables que inciden directamente en el logro de los resultados del proceso; se cuantifica el rendimiento académico. Las calificaciones que se obtendrán se considerarán predictores de qué tan bien le irá al estudiante en la especialización. (p. 56)



Dimensión Económica

“Son las condiciones que los estudiantes deben cumplir para mantenerse durante sus estudios. Estos incluyen vivienda, comida, ropa, transporte, materiales de aprendizaje y entretenimiento. Si estos son buenos, el resultado será positivo” (Tacilla Cardenas y otros, 2020)

Dimensión Familiar

(Tacilla Cardenas y otros, 2020) considera que esta dimensión implica las situaciones familiares de los estudiantes. Ya sea que promueva o limite su potencial humano y social. Esto se debe a que, en casa, se forma una base sólida para el estudiante. En el futuro, se fortalecerá la actitud hacia la investigación y el éxito académico. Es aquí (el entorno familiar) donde se crean patrones de comportamiento, valores y sistemas de relación entre los miembros. Este aspecto describe perfectamente el proceso de enseñanza y aprendizaje de manera significativa. (p. 56)

- Factores del Rendimiento Académico

Factor Endógeno

Estos incluyen perfil demográfico, antecedentes familiares, antecedentes educativos e inclinación a campos académicos. Estas son las características del docente, las estrategias metodológicas y las relaciones sociales dinámicas. Los buenos resultados del aprendizaje dependen de la motivación (superación personal) y la influencia (autorregulación). (Tacilla Cardenas y otros, 2020)



Factor Exógeno

Este factor se relaciona con las variables familiares, sociales y económicas del estudiante que inciden en el rendimiento académico.

Los estudiantes asocian las buenas notas con las prácticas educativas, la presencia significativa de adultos, la educación de los padres y las condiciones de trabajo. (Tacilla Cardenas y otros, 2020)

b. Educación Universitaria

“Etimológicamente, educación proviene del latín educare que significa instruir. Y educere significa guiar. Al principio estos conceptos se utilizaron para definir el cuidado o pastoreo del ganado, y más tarde a la educación y cuidado de los niños” (Julca Meza, 2016)

Hay dos motivaciones principales para la educación individual, llamadas heteroeducación y autoeducación. El primero consiste en el proceso educativo impuesto por el cual se forman los individuos. La autoeducación, por otro lado, ocurre cuando los individuos toman información por sí mismos y tratan de incorporarla a su cuerpo de conocimiento. Idealmente, los dos 'motivos' se complementan para que la educación sea un proceso que proporcione a los individuos las herramientas, los medios y las técnicas necesarias para la formación del conocimiento. (Julca Meza, 2016)

c. Videojuegos

Un videojuego es un tipo de juego electrónico en el que interactúa más de una persona. Lo experimentas a través de una pantalla, y continúa evolucionando y volviéndose más complejo y poderoso a medida que



avanza la tecnología. Pueden implementarse en una o más plataformas como, computadoras, consolas, dispositivos móviles (teléfonos, tabletas), arcades, etc. (Rivera Arteaga & Torres Cosío, 2018)

- **Clasificación de Videojuegos**

Juegos de acción

Ofrecen actividades que inspiran a los jugadores a reaccionar con precisión, decisión y velocidad. Estos son juegos que no requieren que planifiques tus acciones, sino que interactúas en el entorno lo más rápido posible a través de acciones simples (decisiones rápidas) como disparar y golpear. (Rivera Arteaga & Torres Cosío, 2018)

Árcade (plataformas, laberintos, aventuras)

En este tipo de videojuegos, “los jugadores son rápidos y requieren un tiempo de reacción mínimo. Requieren atención enfocada y memoria. Contribuciones al desarrollo psicológico y la conciencia espacial”. (Rivera Arteaga & Torres Cosío, 2018)

Juegos de estrategia

Enfatiza la necesidad de planificar y establecer una estrategia para avanzar en el juego. Esto desarrolla especialmente el pensamiento lógico y la resolución de problemas. Necesitan mantener la concentración, saben administrar recursos, pensar y definir estrategias, desarrollar planes de acción y anticipar las acciones de sus competidores. Ayudan a desarrollar una cantidad de organización mental y espacial. (Rivera Arteaga & Torres Cosío, 2018)



Juegos de aventura.

“Es un tipo de videojuego en donde el término aventura es uno de los elementos fundamentales, que incluye un alto grado de interactividad y la constante necesidad de tomar decisiones” (Rivera & Torres, 2018, p. 5).

Juegos deportivos

Consisten en la interacción entre jugadores reales, permite el comercio y la gestión de equipos de diferentes ligas. Requiere habilidad, velocidad y precisión. Su futuro y presente inmediatos están consagrados a los juegos de estrategia y juegos de acción, donde comparten su potencial en el entrenamiento de habilidades, el procesamiento de la información y el desarrollo de habilidades sensoriales. (Rivera Arteaga & Torres Cosío, 2018)

Juegos de simulación

La simulación es uno de los principales retos a los que se enfrentan los desarrolladores de videojuegos y, aunque es un juego en sí mismo, también se ha integrado como elemento de transición en otros géneros. Nos permiten experimentar y estudiar la manipulación de máquinas, fenómenos y situaciones, y controlarlos (como simulaciones de vuelo y vuelo de aeronaves). Se requiere una estrategia compleja. Identificar y proporcionar conocimientos específicos. (Rivera Arteaga & Torres Cosío, 2018) ***Juegos de rol***

Similar a los juegos de aventuras, pero basado en el desarrollo del personaje en lugar de resolver acertijos. Su éxito se basa en su destreza técnica sin precedentes, que te permite sumergirte por



completo en el juego. El juego de roles no solo desarrolla la aritmética mental, el vocabulario y estimula la creatividad, sino que también estimula actitudes y valores sociales específicos como la empatía, la tolerancia, la conciencia y la responsabilidad, y el equipo de Dios de la Espiritualidad (Rivera Arteaga & Torres Cosío, 2018)

Juegos masivos

El MMORPG (Multiplayer Online Role Playing Game) por sus siglas en inglés, también conocido como video juegos de rol, es un juego de rol que permite a miles de personas jugar e interactuar simultáneamente en un mundo virtual a través de Internet. Jugando videojuegos. El uno al otro (Rivera Arteaga & Torres Cosío, 2018)

Sobrevivencia o supervivencia

“El mismo nombre lo explica: el protagonista debe escapar o resolver problemas ante situaciones y enemigos para poder sobrevivir y continuar el juego. La mayoría de los juegos de terror cubren este género”. (Rivera Arteaga & Torres Cosío, 2018)

c. Inteligencia Artificial (IA)

La inteligencia artificial, o IA, se puede definir como una rama de informática que ofrece un modelo computacional que aprenden en función a redes neuronales biológicas humanas. Es por ello por lo que se han propuesto diversos modelos de IA, y los avances en la tecnología informática han permitido el desarrollo de sistemas “inteligentes” que procesan más datos en menos tiempo, agilizando el proceso de toma de decisiones (Márquez Díaz, 2020)



Es especialmente importante ser consciente de la idea de que las computadoras o el software pueden aprender y tomar decisiones porque sus operaciones crecen exponencialmente con el tiempo. Gracias a estas dos capacidades, los sistemas de inteligencia artificial ahora pueden realizar muchas tareas que antes solo realizaban los humanos. (Márquez Díaz, 2020)

d. Ciencia de Datos

La ciencia de datos es el estudio de los datos para obtener información significativa para su negocio. Es un enfoque interdisciplinario que combina principios y prácticas de los campos de las matemáticas, la estadística, la inteligencia artificial y la ingeniería informática para analizar grandes cantidades de datos. Este análisis permite a los científicos de datos hacer y responder preguntas como "qué pasó", "por qué sucedió esto", "qué sucede" y "qué podemos hacer con los resultados" (AWS, s.f.)

- ¿En qué usar Ciencia de Datos?

Análisis Descriptivo

El análisis descriptivo permite examinar datos para obtener información sobre lo que estaba o está sucediendo en su entorno de datos. Cuenta con visualizaciones de datos como gráficos circulares, gráficos de barras, gráficos de líneas, tablas e historias generadas. Por ejemplo, un servicio de reserva de vuelos registra datos como la cantidad de boletos reservados por día. El análisis descriptivo revela picos y valles preestablecidos y un alto rendimiento del servicio durante meses. (AWS, s.f.)



Análisis de Diagnóstico

El análisis de diagnóstico es un examen detallado o en profundidad de los datos para comprender por qué sucedió algo. Presenta técnicas como el análisis en profundidad, el descubrimiento, la extracción de datos y la correlación. Se pueden realizar muchas manipulaciones y transformaciones de datos en un conjunto determinado para descubrir patrones únicos para cada una de estas técnicas. (AWS, s.f.)

Análisis Predictivo

El análisis predictivo utiliza datos históricos para hacer predicciones precisas sobre la probabilidad de futuros patrones de datos. Presenta técnicas como el aprendizaje automático, la predicción, la coincidencia de modelos y el modelado predictivo. En cada una de estas técnicas, las computadoras están entrenadas para desentrañar relaciones causales en los datos. Por ejemplo, un equipo de servicios de vuelo puede usar la ciencia de datos al comienzo de cada año para predecir los patrones de reserva de vuelos del próximo año. Un programa informático o algoritmo puede analizar datos históricos para predecir el pico de reservas de mayo para un destino en particular. Anticipándose a las futuras necesidades de viaje de sus clientes, la compañía puede comenzar a anunciarse específicamente para estas ciudades a partir de febrero. (AWS, s.f.)

Análisis Prescriptivo

El análisis prescriptivo lleva los datos predictivos al siguiente nivel. No solo predice lo que es probable que suceda, sino que también sugiere la respuesta óptima a sus consecuencias. Puede analizar el



posible impacto de diferentes alternativas y recomendar el mejor curso de acción. Utilice motores de recomendación para análisis de gráficos, simulación, procesamiento de eventos complejos, redes neuronales y aprendizaje automático. Volviendo al ejemplo de la reserva de la aerolínea, el análisis regulatorio puede analizar campañas de marketing anteriores para maximizar los beneficios del próximo aumento de reservas. Los científicos de datos pueden predecir ajustes preestablecidos para diferentes niveles de gasto en diferentes canales de marketing. Estos datos predictivos brindan a las agencias de reservas de aerolíneas más confianza en sus decisiones de marketing. (AWS, s.f.)

e. Machine Learning

Está comprendido por diferentes algoritmos generados para resolver problemas de datos. A los científicos de datos les gusta señalar que no existe un algoritmo único que sea mejor para resolver un problema. El tipo de algoritmo que se utiliza depende del tipo de problema que se está resolviendo, el número de variables, qué tipo de modelo es el más adecuado, etc. A continuación, se muestra una breve descripción de algunos de los algoritmos más utilizados en el aprendizaje automático. (Batta, 2020)

- Aprendizaje Supervisado

Es una tarea de aprendizaje automático para aprender una función que asigna una entrada a una salida en función de pares. Extrayendo una característica de los datos de entrenamiento etiquetados que consisten en un conjunto de ejemplos. Los algoritmos de aprendizaje



automático supervisado son aquellos que requieren asistencia externa. (Batta, 2020)

Árbol de Decisiones

“Un árbol de decisiones es una representación gráfica de diversas opciones y sus resultados. Los nodos representan eventos o elecciones, las aristas representan reglas o condiciones de decisión y las ramas representan un valor que el nodo puede recibir”. (Batta, 2020)

Naive Bayes

Esta es una técnica de clasificación basada en el teorema de Bayes que asume la independencia entre predictores. En pocas palabras, los clasificadores ingenuos de Bayes asumen que la presencia de una característica particular en una clase no está relacionada con la presencia de otra característica. Naive Bayes se dedica principalmente al campo de la clasificación de textos. Esto se usa principalmente para fines de agrupación y clasificación en función de las probabilidades condicionales de que algo suceda. (Batta, 2020)

Máquinas de vectores de soporte

Una de las técnicas de aprendizaje automático más utilizadas es Support Vector Machines (SVM). En el aprendizaje automático, las máquinas de vectores de soporte son modelos de aprendizaje supervisado con algoritmos de aprendizaje asociativo para el análisis de datos utilizados para la clasificación y el análisis de regresión. (Batta, 2020)



- **Aprendizaje no supervisado**

Esto se llama aprendizaje no supervisado porque, a diferencia del aprendizaje supervisado anterior, no hay una respuesta correcta ni un maestro. El algoritmo es auto explorador y presenta estructuras interesantes en los datos. Los algoritmos de aprendizaje no supervisados aprenden algunas características de los datos. Cuando se ingresan nuevos datos, las características previamente aprendidas se utilizan para identificar las clases de datos. Se utiliza principalmente para encapsulación y reducción de funciones. (Batta, 2020)

K-medias

K-medias es uno de los algoritmos de aprendizaje no supervisado más simples que ayudan a resolver problemas de agrupamiento bien conocidos. Este proceso sigue un método simple y directo de clasificar un conjunto de datos dado en un grupo dado. La idea principal es definir k centros, uno para cada clúster. Estos centros deben colocarse sabiamente, ya que diferentes ubicaciones conducen a diferentes resultados. Así que la mejor opción es colocarlos lo más separados posible. (Batta, 2020)

- **Aprendizaje Semisupervisado**

El aprendizaje automático semisupervisado es una combinación de métodos de aprendizaje automático supervisado y no supervisado. Puede ser fructífero en aquellas áreas del aprendizaje automático y la minería de datos en las que los datos no etiquetados ya están presentes y obtener los datos etiquetados es un proceso tedioso. Con los métodos de aprendizaje automático supervisado más comunes, se



entrena un algoritmo de aprendizaje automático en un conjunto de datos "etiquetados" en el que cada registro incluye la información del resultado. A continuación, se describen algunos de los algoritmos de aprendizaje semisupervisado (Batta, 2019, p. 384).

f. Modelo Predictivo

Los modelos predictivos son un conjunto de técnicas destinadas a predecir resultados futuros a través de los campos de aprendizaje automático, recopilación de datos históricos, big data y reconocimiento de patrones. Su objetivo es esclarecer la toma de decisiones a través de técnicas de análisis de datos. En los últimos años, el dominio predictivo ha desempeñado un papel importante en una amplia gama de campos, como los negocios, la atención médica, los servicios financieros, las políticas gubernamentales, la publicidad, el marketing y las redes sociales. (edX, s.f.)

2.2.2. Metodología

a. Metodología CRISP – DM (Cross Industry Standard Process for Data Mining)

“El método CRISP-DM es uno de los métodos más utilizados en la actualidad para el desarrollo de proyectos de minería de datos” (Espinosa-Zuñiga, 2020)



La metodología CRISP-DM consta de seis etapas:

- **Etapas de la Metodología CRISP – DM**

Comprensión del Problema o Negocio

“Este es el paso más importante, porque el siguiente paso es inútil si no conoce muy bien el problema o negocio” (Espinosa-Zuñiga, 2020)

Dentro de sus principales actividades se encuentran:

Identificación del Problema

“Esto incluye comprender y definir el problema e identificar los requisitos, suposiciones, limitaciones y beneficios del proyecto” (Espinosa-Zuñiga, 2020)

Determinación de objetivos

Destaca objetivos a alcanzar proponiendo soluciones basadas en modelos de minería de datos. El propósito de este estudio fue obtener un modelo de comportamiento de las entidades en la República Mexicana de acuerdo con las unidades económicas ubicadas dentro de cada entidad (Espinosa-Zuñiga, 2020)

Evaluación de la situación actual

“Debido a que especifica el estado actual antes de implementar la solución de minería de datos propuesta, existe un punto de comparación para medir el éxito del proyecto” (Espinosa-Zuñiga, 2020)



Comprensión de Datos

Dentro de sus principales actividades se encuentran:

Recolección de datos

“Incluye recopilar los datos que se utilizarán en el proyecto, identificar las fuentes, las técnicas utilizadas para recopilarlos, los problemas encontrados al recopilarlos y cómo resolverlos” (Espinosa-Zuñiga, 2020)

Descripción de datos

“Esta actividad incluye la definición de tipos de datos, formatos, volúmenes y semántica de cada uno de ellos” (Espinosa-Zuñiga, 2020)

Exploración de datos

“Consiste en la aplicación de pruebas estadísticas básicas para ayudarlo a conocer las características de sus datos para comprenderlos mejor (Espinosa-Zuñiga, 2020)

Preparación de Datos

“Esta suele ser la fase del proyecto que consume más tiempo debido a que los datos transformados en función de los resultados de las fases anteriores se seleccionan para su uso en la fase de modelado” (Espinosa-Zuñiga, 2020)

Dentro de sus principales actividades se encuentran:

Limpieza de datos

“Consiste en aplicar diversas técnicas, como la normalización de datos, la discretización de campos numéricos, el manejo de valores perdidos, el manejo de



duplicados y la derivación de datos” (Espinosa-Zuñiga, 2020)

Creación de indicadores

“En esta actividad se busca generar métricas a partir de datos existentes que ayuden a mejorar su previsibilidad y revelar comportamientos interesantes para modelar” (Espinosa-Zuñiga, 2020)

Transformación de datos

“Implica el cambio del formato o la estructura de algunos datos sin cambiar su significado para aplicar ciertas técnicas durante la etapa de modelado” (Espinosa-Zuñiga, 2020)

Modelado

En esta etapa se logra el modelo propiamente dicho.

Dentro de sus principales actividades se encuentran:

Selección de técnica de modelado

“La elección de la técnica adecuada depende del problema a resolver, los datos disponibles, las herramientas de minería de datos disponibles y el dominio de la técnica elegida” (Espinosa-Zuñiga, 2020)

Selección de datos de prueba

“Algunos tipos de modelos requieren que la muestra se divida en datos de entrenamiento y validación” (Espinosa-Zuñiga, 2020)



Obtención del modelo

“Implica generar el mejor modelo a través de un proceso iterativo de cambio de parámetros” (Espinosa-Zuñiga, 2020)

Evaluación del Modelo

En este paso, la calidad del modelo se determina en función del análisis de ciertas estadísticas, ya sea comparando los resultados con resultados anteriores o analizando los resultados con la ayuda de expertos en la materia. Con base en el resultado de este paso, se decide si se continúa con las etapas finales del método, se regresa a uno de los pasos anteriores o se comienza de nuevo con un nuevo proyecto. (Espinosa-Zuñiga, 2020) ***Implementación del Modelo***

Este paso utiliza el conocimiento adquirido a través del modelo a través de acciones específicas. De nuevo, hay que documentar claramente los resultados para el usuario final, documentar correctamente los pasos metodológicos y revisar el proyecto buscando lecciones aprendidas. También se monitorea inventario para encontrar áreas de oportunidad e incluso nuevos problemas. (Espinosa-Zuñiga, 2020)

2.2.3. Contexto Tecnológico

a. Lenguaje de Programación

- Python

Python es un lenguaje de programación ampliamente utilizado en aplicaciones web, desarrollo de software, ciencia de datos y



aprendizaje automático (ML). Los desarrolladores usan Python porque es eficiente, fácil de aprender y puede ejecutarse en una variedad de plataformas. El software Python se puede descargar gratis, se integra bien con todo tipo de sistemas y acelera el desarrollo. (AWS, s.f.)

Bibliotecas de Python

Una biblioteca es una colección de código de uso frecuente que los desarrolladores pueden incluir en sus programas de Python para que no tengan que escribir el código desde cero. Por defecto, Python incluye una biblioteca estándar con muchas funciones reutilizables. Además, hay disponibles más de 137 000 bibliotecas de Python para diversas aplicaciones, como desarrollo web, ciencia de datos y aprendizaje automático (ML) (AWS, s.f.)

Matplotlib

Los desarrolladores usan Matplotlib para trazar datos en gráficos bidimensionales y tridimensionales (2D y 3D) de alta calidad. Se utiliza comúnmente en aplicaciones científicas. Matplotlib permite que los datos se muestren y visualicen en varios gráficos, como gráficos de barras y gráficos de líneas. También puede trazar múltiples gráficos a la vez y es portátil para todas las plataformas. (AWS, s.f.)

Pandas

Pandas proporciona estructuras de datos flexibles y optimizadas que se pueden usar para manipular datos de series temporales y datos estructurados, como tablas y matrices. Por ejemplo, puede usar



Pandas para leer, escribir, unir, filtrar y agrupar datos. Mucha gente lo usa para tareas de ciencia de datos, análisis de datos y ML. (AWS, s.f.)

NumPy

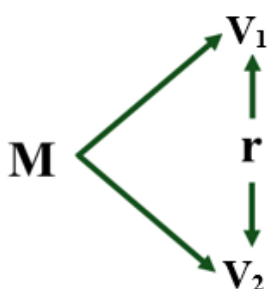
“NumPy es una biblioteca popular utilizada por los desarrolladores para crear y administrar fácilmente matrices, manipular diagramas lógicos y realizar operaciones de álgebra lineal. NumPy admite la integración en muchos lenguajes como C y C++” (AWS, s.f.)

CAPÍTULO III

METODOLOGÍA

3.1. DISEÑO DE INVESTIGACIÓN

El diseño de esta investigación es transversal de tipo correlacional-causal. En la que se busca describir la relación entre el consumo de videojuegos y el rendimiento académico, además de la causal de estas variables.



Donde:

M: Muestra

V₁: Influencia de los videojuegos

V₂: Rendimiento académico

r: Correlación entre dichas variables

3.2. METODOLOGÍA DE LA INVESTIGACIÓN

- Identificar la realidad problemática y definir objetivos e hipótesis de la investigación.
- Recopilar y explorar los datos para entender su estructura, calidad y relevancia.
- Limpiar y transformar los datos recolectados para hacerlos adecuados para el análisis.
- Seleccionar y aplicar técnicas de modelado de datos para crear modelos predictivos.
- Evaluar los modelos para asegurar su calidad y eficiencia.



- Implementar el modelo en un entorno real y comunicar los resultados para su aplicación práctica.
- Resultados y discusión
- Conclusiones y recomendaciones

3.3. POBLACIÓN

En la población se han considerado a los estudiantes matriculados en el ciclo académico 2023-I de la Universidad Nacional del Santa.

3.4. MUESTRA

La muestra es de tipo no probabilístico e intencional, bajo criterio de los investigadores, con el fin de que sea uniforme. Se escogió a los estudiantes matriculados en el ciclo académico 2023-I de la Escuela Profesional de Ingeniería de Sistemas e Informática, lo cual, según los datos recopilados da un total aproximado de 300 estudiantes.



3.5. OPERACIONALIZACIÓN DE VARIABLE

Tabla 1:

Operacionalización de Variables

VARIABLES	DIMENSIONES	INDICADORES
Influencia de los videojuegos (V ₁)	Capacidad Económica	<ul style="list-style-type: none">• Cartera de videojuegos• Gastos mensuales• Inversión en hardware• Relación entre situación económica y videojuegos
	Tiempo de Consumo	<ul style="list-style-type: none">• Días de juego a la semana• Horario de juego• Horas diarias de juego• Noción del tiempo
	Bienestar Físico/Mental	<ul style="list-style-type: none">• Autoevaluación de bienestar general• Pérdida de horas de sueño• Niveles de estrés• Resiliencia
	Impacto sobre actividades cotidianas	<ul style="list-style-type: none">• Autoevaluación del impacto del consumo de videojuegos• Dificultad para realizar actividades• Percepción en el rendimiento académico• Impacto en la productividad académica
	Relaciones Sociales	<ul style="list-style-type: none">• Cambios en la capacidad de comunicación e interacción social• Conflictos interpersonales• Autoevaluación sobre las habilidades de colaboración y trabajo en equipo• Descuido hacia personas del entorno
Rendimiento académico (V ₂)	Desarrollo Cognitivo	<ul style="list-style-type: none">• Desafíos cognitivos en los videojuegos• Adaptación a entornos virtuales• Desarrollo de habilidades multitarea• Transferencia de habilidades a la vida real
	Promedio de notas	<ul style="list-style-type: none">• Escala de valor: Excelente: 18 – 20 Bueno: 14 -17 Regular: 11 -13 Deficiente: 00 - 10



3.6. TÉCNICAS E INSTRUMENTOS DE RECOLECCIÓN DE DATOS

La encuesta es una técnica de recolección de datos que a través de un cuestionario ya sea físico o virtual, se puede obtener la perspectiva de los involucrados principales de la investigación sobre la forma en la que se llevan a cabo los procesos y/o actividades del área de trabajo en interés.

El instrumento será un cuestionario de alternativas múltiples, dirigido a los estudiantes matriculados a la carrera profesional de Ingenierías de Sistemas e Informática en el ciclo académico 2023-I, con el propósito de recopilar los datos cuantitativos de cada uno de nuestros indicadores.

Requerimiento de información

Para aplicar esta técnica, se realizó una solicitud al Rectorado de la UNS, quien emitió el Memorando N° 02679-2023-UNS-R a la Dirección de Evaluación y Desarrollo Académico (DEDA), autorizando la entrega del historial académico de los estudiantes matriculados en la Carrera Profesional de Ingeniería de Sistemas e Informática durante el ciclo académico 2023-I.



CAPÍTULO IV

RESULTADOS

4.1.FASE I: Comprensión del Negocio

4.1.1. Comprensión del contexto y determinación de objetivos

Esta etapa se centra en entender el contexto y los objetivos del proyecto. A partir de este conocimiento, se define un problema y se elabora un plan para lograr los objetivos.

a. Contexto

El objetivo principal de las universidades es asegurar la calidad de la educación en sus diversas carreras, lo cual usualmente se refleja en el rendimiento académico de los estudiantes. Sin embargo, a menudo no se logra este objetivo debido a la falta de técnicas y/o herramientas que permitan hacer seguimiento al rendimiento académico de los estudiantes, quienes son los actores principales en las universidades.

b. Objetivos

Este trabajo de investigación tiene como finalidad de determinar en qué medida el rendimiento académico de los estudiantes universitarios puede ser predicho a partir del consumo de videojuegos utilizando técnicas de Machine Learning, con estos resultados el personal administrativo y docentes brindar el apoyo correspondiente a los estudiantes para mejorar sus calificaciones.

Para desarrollar un modelo que prediga el rendimiento académico de los estudiantes de la EPISI de la UNS, se recopiló información de los



estudiantes matriculados en el periodo académico 2023-I. Esta información fue proporcionada por la Dirección de Evaluación y Desarrollo Académico y se complementó con datos obtenidos a través de una encuesta virtual realizada mediante Google Forms.

4.1.2. Evaluación de la situación

a. Desde el punto de vista de la UNS

Teniendo en cuenta que la UNS es una universidad licenciada y en proceso de su acreditación institucional, es crucial disponer de una herramienta que permita realizar el seguimiento del rendimiento académico de sus estudiantes y tomar decisiones informadas que promuevan el apoyo a aquellos que tienen un rendimiento deficiente.

b. Desde el punto de vista de Machine Learning

Según lo mencionado en el marco teórico, Machine Learning se clasifica en algoritmos supervisados y no supervisados. Dentro de los algoritmos supervisados, se distinguen las técnicas de regresión y clasificación. Es en la técnica de clasificación donde se utilizan los algoritmos para determinar rendimiento óptimo. Además, la estadística descriptiva es fundamental para evaluar los resultados obtenidos en las primeras etapas del proceso.

La metodología seleccionada para aplicarse en este proyecto es la CRISP-DM, por ser flexible y porque brinda una estructura de trabajo ordenada.



4.1.3. Determinación de los objetivos de Machine Learning aplicados al proyecto

Los propósitos de Machine Learning de esta tesis se sintetizan en los siguientes pasos:

- Proceder con la limpieza de los datos obtenidos mediante las técnicas de recolección de datos.
- Identificar los indicadores más pertinentes que influyen en el rendimiento académico.
- Evaluar y seleccionar el modelo más adecuado para realizar la predicción.
- Realizar la predicción del rendimiento académico utilizando el modelo más eficiente.

La predicción está sustentada por dos aspectos principales: en primer lugar, se realiza un análisis estadístico de los datos para identificar los factores que impactan en el rendimiento académico. En segundo lugar, se emplea Python para evaluar el desempeño de los algoritmos y seleccionar el modelo que prediga de manera óptima el rendimiento académico.



4.1.4. Desarrollo del plan del proyecto

El tiempo estimado realizar este proyecto mediante la metodología CRISP-DM, cumpliendo cada una de sus fases se muestra en la siguiente tabla:

Tabla 2:

Plan del proyecto

FASE	TIEMPO	RECURSOS	RIESGOS
Fase I Comprensión del negocio	20 días	<ul style="list-style-type: none">• Encuesta de consumo de videojuegos a egresados (Google Form)	No se consideran
Fase II Comprensión de los datos	15 días	<ul style="list-style-type: none">• Microsoft Excel	Datos vacíos Datos incongruentes
Fase III Preparación de los datos	15 días	<ul style="list-style-type: none">• Microsoft Excel• Python 3• Google Colab	Datos atípicos
Fase IV Modelado	45 días	<ul style="list-style-type: none">• Python• Librerías de ML• Google Colab	No encontrar un modelo óptimo
Fase V Evaluación	15 días	<ul style="list-style-type: none">• Python• Librerías de ML• Google Colab	Obtener un coeficiente de determinación irrelevante
Fase VI Despliegue	20 días	<ul style="list-style-type: none">• Informe Final• Cuaderno de código	No se consideran



4.2. FASE II: Comprensión de los Datos

Esta etapa inicia con la recolección de datos preliminares y las actividades que facilitan la familiarización de los datos. Durante esta fase, se identifican problemas de calidad, se descubre conocimiento inicial sobre los datos y permite formular hipótesis sobre la información no visible.

4.2.1. Obtención de datos preliminares

En este proceso se tomaron en consideración las siguientes técnicas de recolección de datos.

a. Requerimiento de información

Se solicitó la información del historial académico de los estudiantes de la EPISI matriculados de los periodos académicos 2022-I, 2022-II, 2023-I, con los siguientes datos:

- Escuela Profesional
- Código de estudiante
- Nombre del estudiante
- Asignaturas
- Créditos
- Calificaciones por asignatura
- Promedio ponderado por ciclo

b. Encuesta

- Datos demográficos
- Capacidad Económica
- Tiempo de Consumo de Videojuegos



- Bienestar Físico-Mental
- Impacto Sobre Actividades Cotidianas
- Relaciones Sociales
- Desarrollo Cognitivo

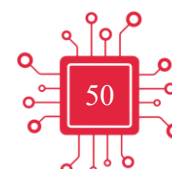
La predicción estará en base al promedio ponderado del ciclo académico 2023-I y la fuente principal de obtención de la información en bruto se obtuvo del cuestionario aplicado a los estudiantes de la EPISI matriculados en el periodo académico 2023-I (*Anexo 01*)

A continuación, se muestran el listado de los puntajes y alternativas por preguntas del cuestionario.

- **Preguntas 1**
- **Alternativas**
 - 0. S/. 0.00 - S/. 250.00
 - 1. S/. 250.00 - S/. 500.00
 - 2. S/. 500.00 - S/. 750.00
 - 3. S/. 750.00 - S/. 1000.00
 - 4. S/. 1000.00 a más
- **Preguntas 2**
- **Alternativas**
 - 0. S/. 0.00 - S/. 100.00
 - 1. S/. 100.00 - S/. 200.00
 - 2. S/. 200.00 - S/. 300.00
 - 3. S/. 300.00 - S/. 400.00
 - 4. S/. 500.00 a más

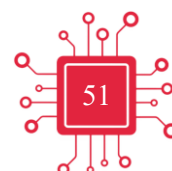


- **Preguntas 5**
- **Alternativas**
 0. Entre 1 a 3 días
 1. Entre 3 a 4 días
 2. Entre 4 a 5 días
 3. Entre 5 a 6 días
 4. Todos los días
- **Preguntas 6**
- **Alternativas**
 0. Menos de 1 hora
 1. De 1 a 2 horas
 2. De 2 a 4 horas
 3. De 4 a 6 horas
 4. Más de 6 horas
- **Preguntas 7**
- **Alternativas**
 0. Mañana (05 am – 10 am)
 1. Mediodía (10 am - 03 pm)
 2. Tarde (03 pm - 08 pm)
 3. Noche (08 pm - 12 am)
 4. Madrugada (12 am - 05 am)
- **Preguntas 3, 8, 10, 11, 12, 14, 15, 18, 20, 24**
- **Alternativas**
 0. Nunca
 1. Raramente





2. Ocasionalmente
 3. Frecuentemente
 4. Siempre
- **Preguntas 4, 21, 22, 23**
 - **Alternativas**
 0. Totalmente en desacuerdo
 1. En desacuerdo
 2. Neutral
 3. De acuerdo
 4. Totalmente de acuerdo
 - **Preguntas 9**
 - **Alternativas**
 0. Muy positivo
 1. Positivo
 2. Neutral
 3. Negativo
 4. Muy negativo
 - **Preguntas 13**
 - **Alternativas**
 0. Insignificadamente
 1. Levemente
 2. No sabe / No respondes
 3. Moderadamente
 4. Significativamente

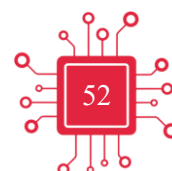




- **Preguntas 16, 17**
- **Alternativas**
 0. Ningún cambio
 1. Cambios mínimos
 2. Cambios moderados
 3. Cambios significativos
 4. Cambios muy significativos
- **Preguntas 19**
- **Alternativas**
 0. Ningún impacto
 1. Impacto mínimo
 2. Impacto moderado
 3. Impacto significativo
 4. Impacto muy significativo

4.2.2. Identificación de los datos

- a. **Escuela Profesional:** En este caso, este proyecto está enfocada en los estudiantes de la escuela profesional de Ingeniería de Sistemas e Informática.
- b. **Código de estudiante:** Identificador asignado a cada estudiante de la UNS para la gestión y seguimiento de su historial académico.
- c. **Nombre del estudiante:** Identificación según datos de RENIEC de los estudiantes
- d. **Asignaturas:** Materias o cursos que los estudiantes deben aprobar según su currículo vigente.





- e. **Créditos:** Valoración que se le da a cada asignatura en función a sus horas de teoría y práctica.
- f. **Calificaciones por asignatura:** Se establece en escala vigesimal, la condición para ser considerada asignatura aprobada es que la calificación sea ≥ 11 .
- g. **Promedio ponderado por ciclo:** Es la calificación final obtenida en el ciclo, se calcula a través del promedio de la multiplicación de los créditos por la nota de la asignatura.

4.2.3. Análisis preliminar de los datos

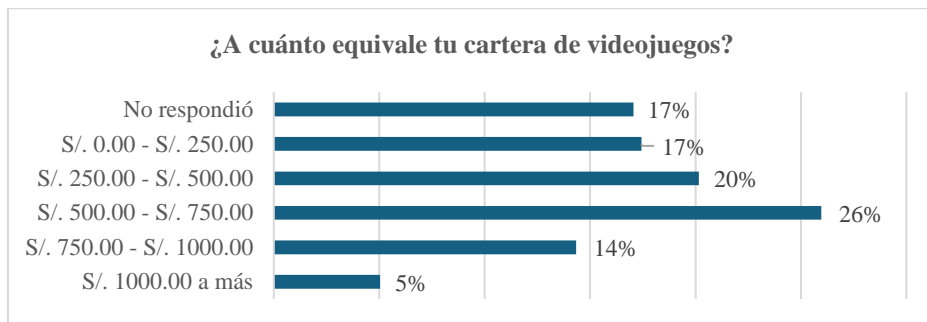
Para este proceso emplearemos estadísticos de las respuestas al cuestionario sobre la influencia de los videojuegos aplicado a los estudiantes de la EPISI matriculados en el periodo académico 2023-I, el cual cuenta con 24 preguntas divididas en 6 dimensiones.

El cuestionario fue dirigido a 258 estudiantes, de los cuales 214 respondieron correctamente, dando como resultado lo siguiente:

- a. **Pregunta 01:** Según tú, ¿a cuánto equivale tu cartera de videojuegos?

Gráfica 3:

Gráfico de barras: Pregunta 01



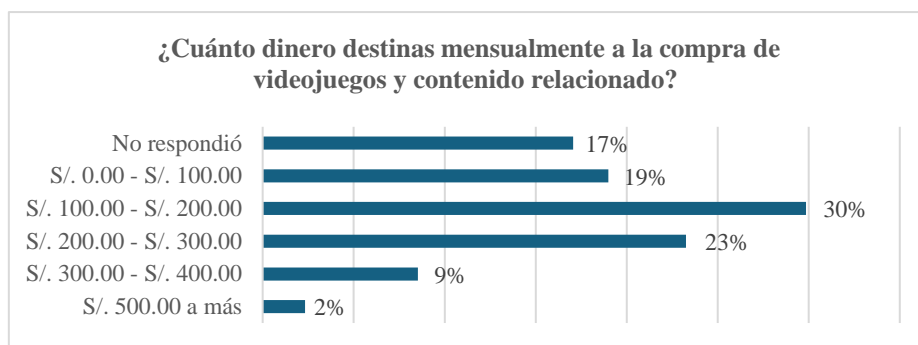


- El 26% de los estudiantes dice que su cartera de videojuegos equivale a S/. 500.00 - S/. 750.00.

b. Pregunta 02: ¿Cuánto dinero destinas mensualmente a la compra de videojuegos y contenido relacionado?

Gráfica 4:

Gráfico de barras: Pregunta 02

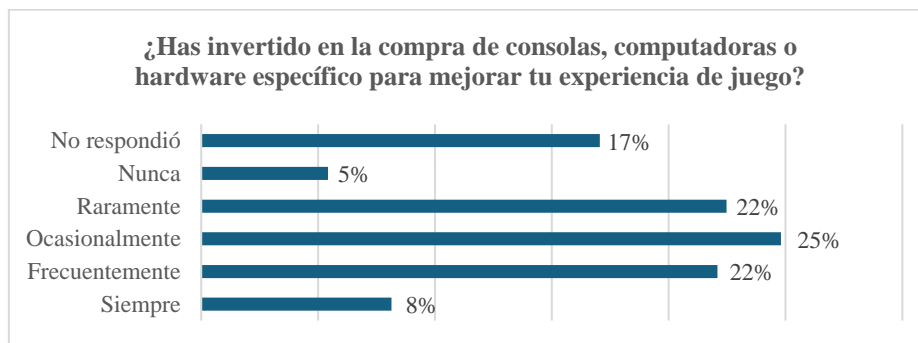


- El 23% de los estudiantes destina S/. 200.00 - S/. 300.00 a la compra de videojuegos y contenido relacionado de manera mensual, mientras que un 2% destina más de S/. 500.00.

c. Pregunta 03: ¿Has invertido en la compra de consolas, computadoras o hardware específico para mejorar tu experiencia de juego?

Gráfica 5:

Gráfico de barras: Pregunta 03



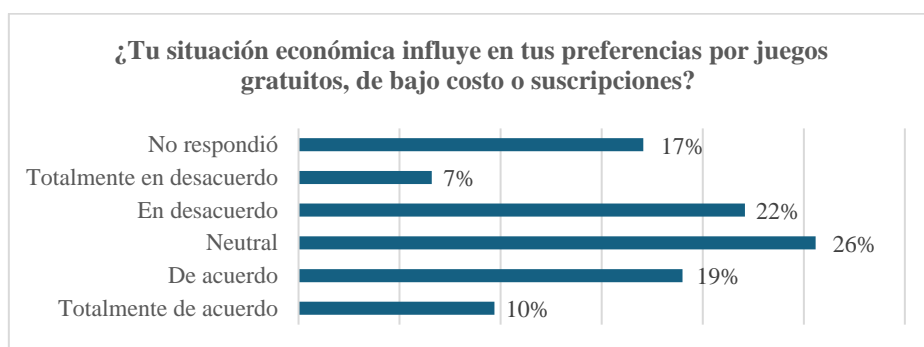


- El 25% de los estudiantes ha invertido ocasionalmente en la compra de consolas, computadoras o hardware específico, mientras que un 8% lo hace siempre.

d. **Pregunta 04:** ¿Tu situación económica influye en tus preferencias por juegos gratuitos, de bajo costo o suscripciones?

Gráfica 6:

Gráfico de barras: Pregunta 04



- El 19% y el 10% de los estudiantes están de acuerdo y totalmente de acuerdo respectivamente en que su situación económica influye en sus preferencias por videojuegos gratuitos, de bajo costo o suscripciones, mientras que el 7% considera que no influye y el otro 26% se mantienen en una posición neutra (no responde, no opina).



e. **Pregunta 05:** ¿Cuántos días a la semana juegas videojuegos?

Gráfica 7:

Gráfico de barras: Pregunta 05

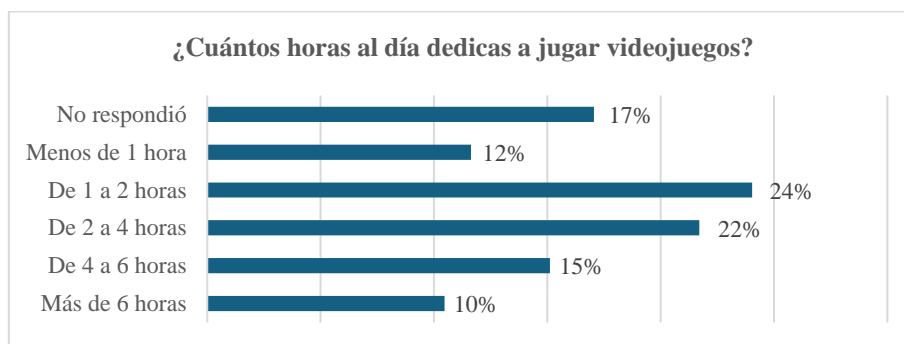


- El 18% de los estudiantes respondió que juega todos los días videojuegos, un acumulado de 18% juega entre 3 a 5 días y el 11% juega menos de 1 día por semana.

f. **Pregunta 06:** ¿Cuántos horas al día dedicas a jugar videojuegos?

Gráfica 8:

Gráfico de barras: Pregunta 06



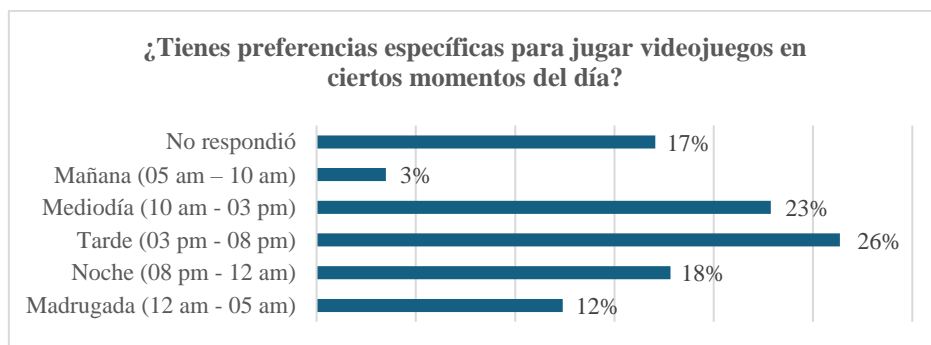
- Un acumulado de 24% de estudiantes juega menos de 2 horas al día, mientras, el 22% juega entre 2 a 4 horas, un 15% juega entre 4 a 6 horas y el 10% respondió que juega más de 6 horas por día.



g. **Pregunta 07:** ¿Tienes preferencias específicas para jugar videojuegos en ciertos momentos del día?

Gráfica 9:

Gráfico de barras: Pregunta 07

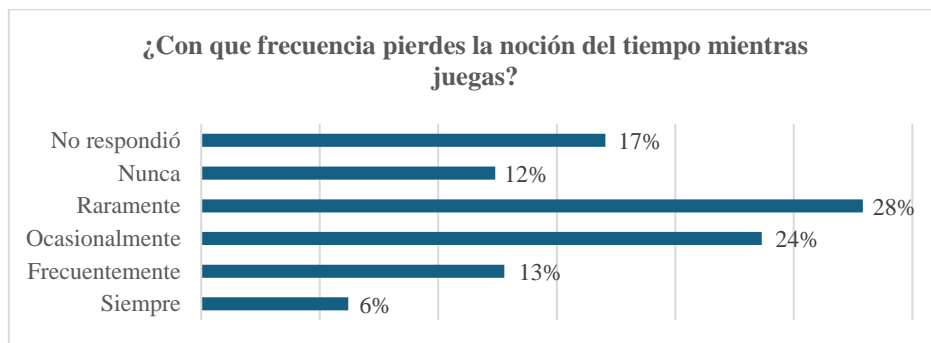


- El 26% de los estudiantes tienen una preferencia de jugar videojuegos por las tardes, así mismo, un 23% prefiere jugar al mediodía, un 12% lo hace de madrugada, un 18% por la noche, y un mínimo del 3% prefiere jugar por las mañanas.

h. **Pregunta 08:** ¿Con que frecuencia pierdes la noción del tiempo mientras juegas?

Gráfica 10:

Gráfico de barras: Pregunta 08



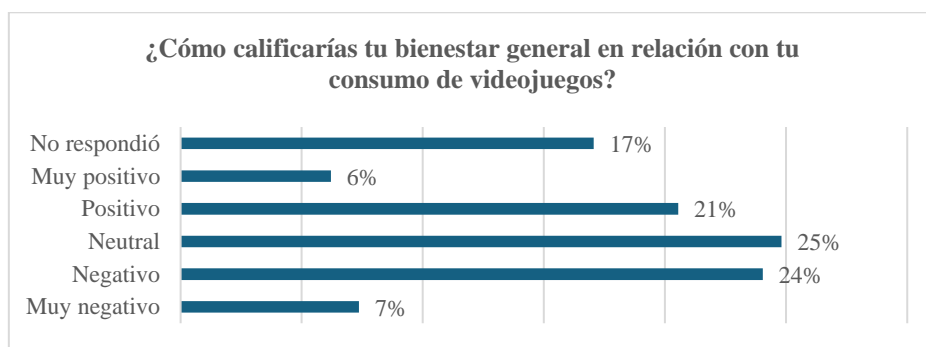


- El 12% de los estudiantes indicaron que nunca han perdido la noción del tiempo mientras juega, un 28% indicó que raramente le pasa, el 24% que le ocurre de manera ocasional, el 13% le ocurre de manera frecuente y el 6% indicó que siempre lo hace.

i. **Pregunta 09:** ¿Cómo calificarías tu bienestar general en relación con tu consumo de videojuegos?

Gráfica 11:

Gráfico de barras: Pregunta 09



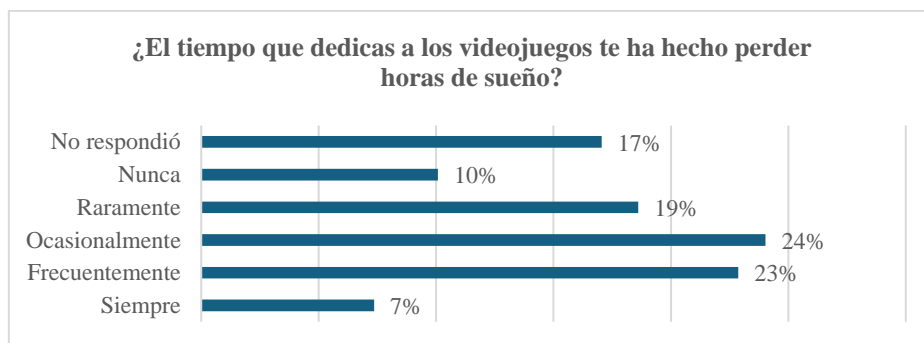
- Sólo el 7% de los estudiantes considera muy negativa la relación de su bienestar general en relación con su consumo de videojuegos, mientras que un 21% lo considera positivo y un 7% lo considera muy positiva.



- j. **Pregunta 10:** ¿El tiempo que dedicas a los videojuegos te ha hecho perder horas de sueño?

Gráfica 12:

Gráfico de barras: Pregunta 10



- El 7% de los estudiantes siempre pierde horas de sueño por dedicar tiempo a los videojuegos, mientras que el 23% lo hace de manera frecuente, el 24% de manera ocasional, el 19% raramente y el 10% indica que nunca han perdido horas de sueño.

- k. **Pregunta 11:** ¿Experimentas estrés o frustración al momento de jugar videojuegos?

Gráfica 13:

Gráfico de barras: Pregunta 11



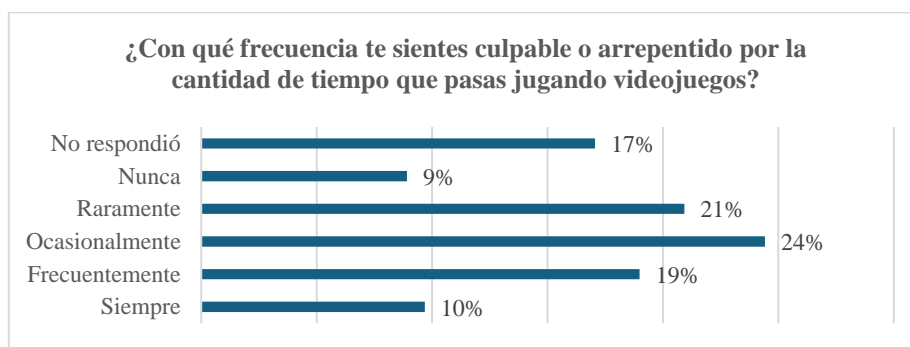


- El 5% de los estudiantes siempre experimentan estrés o frustración al momento de jugar videojuegos, mientras que un 26% lo experimenta frecuentemente, el 22% de manera ocasional, el 28% raramente y el 2% nunca ha lo ha experimentado.

I. Pregunta 12: ¿Con qué frecuencia te sientes culpable o arrepentido por la cantidad de tiempo que pasas jugando videojuegos?

Gráfica 14:

Gráfico de barras: Pregunta 12



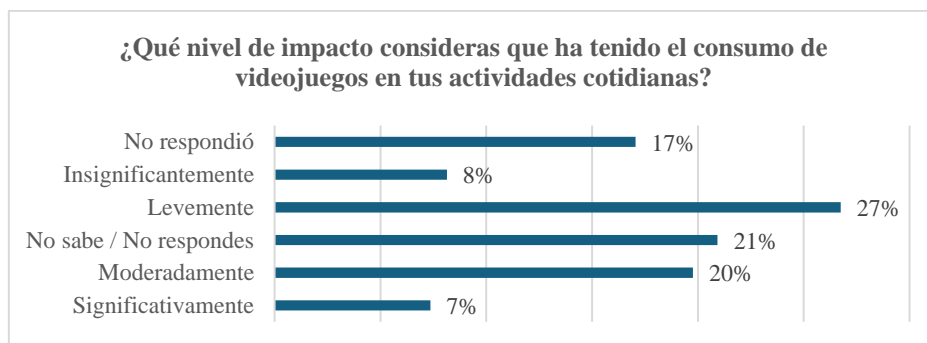
- El 10% de los estudiantes indicaron que siempre se sienten culpables o arrepentidos por la cantidad de tiempo que pasa jugando, el 19% de manera frecuente, el 24% de manera ocasional, el 21% raramente y el 9% nunca se ha sentido culpable o arrepentido.



m. Pregunta 13: ¿Qué nivel de impacto consideras que ha tenido el consumo de videojuegos en tus actividades cotidianas?

Gráfica 15:

Gráfico de barras: Pregunta 13

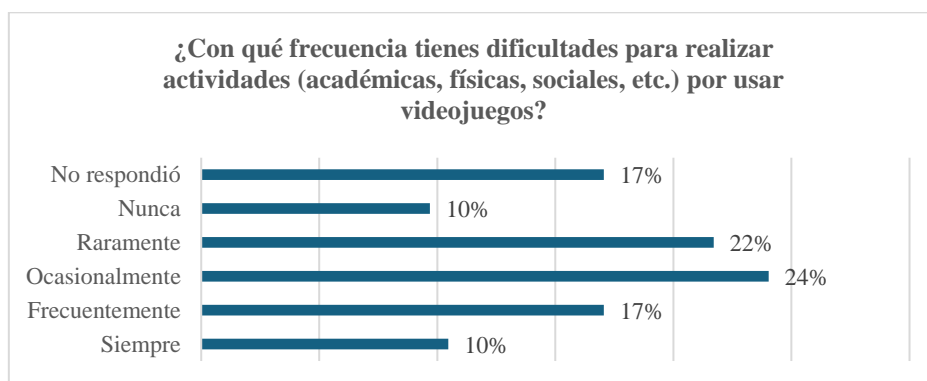


- El 7% de los estudiantes considera un impacto significativo del consumo de videojuegos en su vida cotidiana, mientras que el 20% considera un impacto moderado, el 21% no sabe/no responde, el 27% considera un impacto leve y un 8% indica un impacto insignificante.

n. Pregunta 14: ¿Con qué frecuencia tienes dificultades para realizar actividades (académicas, físicas, sociales, etc.) por usar videojuegos?

Gráfica 16:

Gráfico de barras: Pregunta 14



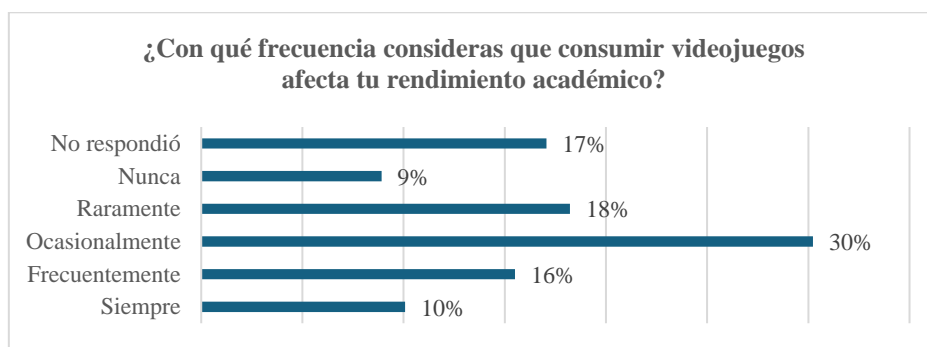


- El 10% de los estudiantes siempre experimentan estrés o frustración al momento de jugar videojuegos, mientras que un 17% lo experimenta frecuentemente, el 24% de manera ocasional, el 22% raramente y el 10% nunca ha lo ha experimentado.

o. **Pregunta 15:** ¿Con qué frecuencia consideras que consumir videojuegos afecta tu rendimiento académico?

Gráfica 17:

Gráfico de barras: Pregunta 15



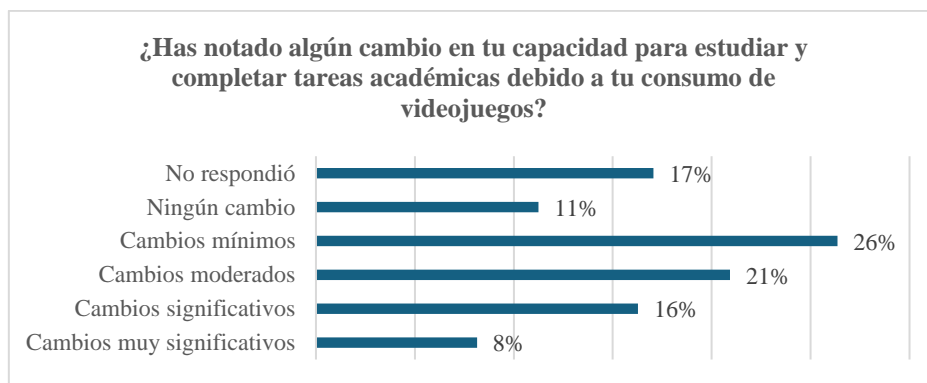
- El 18 % de los estudiantes indicó que los videojuegos rara vez afectan su rendimiento académico, el 9 % dijo que nunca lo han afectado, el 30 % señaló un impacto ocasional, el 16 % reportó un efecto frecuente y el 10 % afirmó que siempre han afectado su rendimiento.



- p. **Pregunta 16:** ¿Has notado algún cambio en tu capacidad para estudiar y completar tareas académicas debido a tu consumo de videojuegos?

Gráfica 18:

Gráfico de barras: Pregunta 16



- El 11% de los estudiantes considera que no ha notado ningún cambio para estudiar debido a su consumo de videojuegos, el 26% ha detectado cambios mínimos, el 21% cambios moderados, el 16% cambios significativos y el 8% de estudiantes ha notado cambios muy significativos.



- q. **Pregunta 17:** ¿Has experimentado cambios en tu capacidad para comunicarte o interactuar socialmente como resultado del consumo de videojuegos?

Gráfica 19:

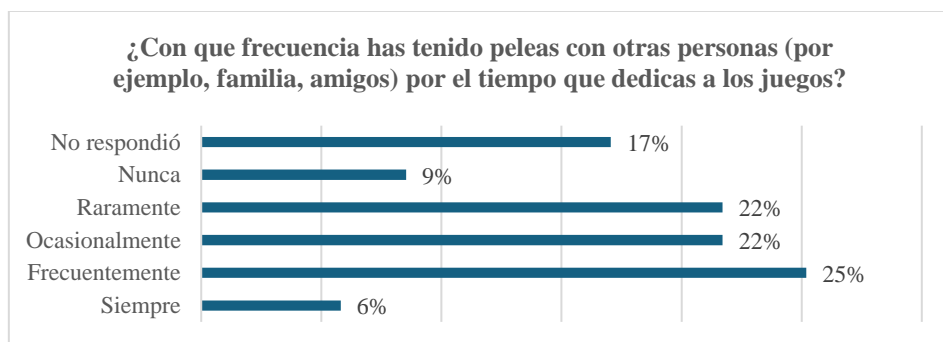
Gráfico de barras: Pregunta 17



- El 7% de los estudiantes considera que no ha notado ningún cambio, el 18% ha detectado cambios mínimos, el 24% cambios moderados, el 23% cambios significativos y el 11% de los estudiantes ha notado cambios muy significativos.
- r. **Pregunta 18:** ¿Con que frecuencia has tenido peleas con otras personas (por ejemplo, familia, amigos) por el tiempo que dedicas a los juegos?

Gráfica 20:

Gráfico de barras: Pregunta 18

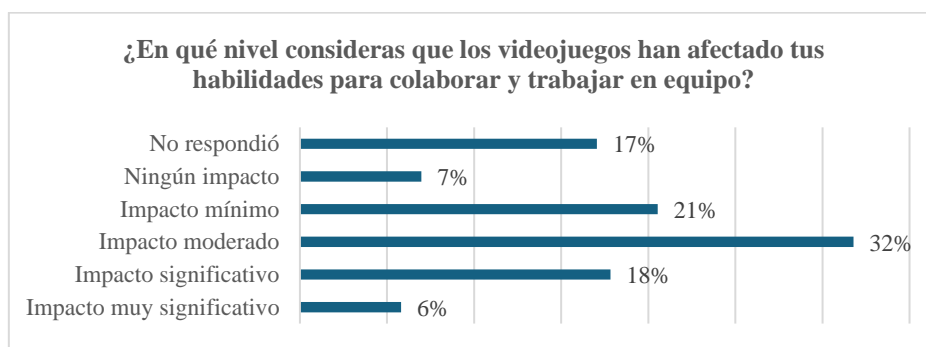




- El 9% de los estudiantes indicaron que nunca han tenido peleas con otras personas por el tiempo que dedica a los videojuegos, mientras que un 22% y un 25% indican que lo han tenido de manera ocasional y frecuente respectivamente.
- s. **Pregunta 19:** ¿En qué nivel consideras que los videojuegos han afectado tus habilidades para colaborar y trabajar en equipo?

Gráfica 21:

Gráfico de barras: Pregunta 19



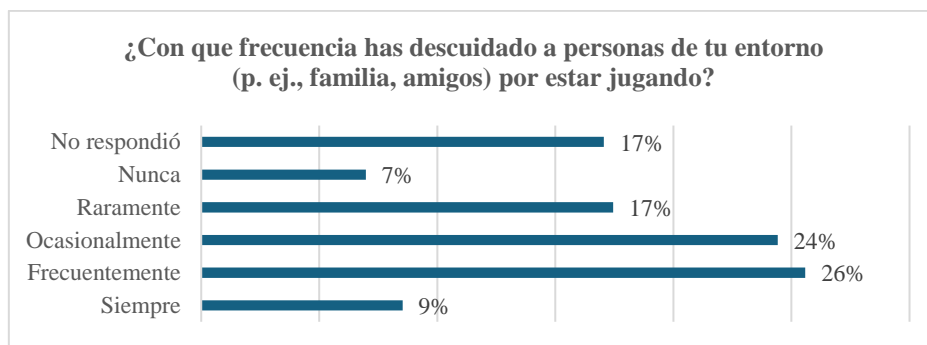
- El 7% de los estudiantes considera que los videojuegos no han tenido ningún impacto en sus habilidades para colaborar y trabaja en equipo, el 21% que ha tenido un impacto mínimo, el 32% un impacto moderado, el 18% un impacto significativo, y el 6% de los estudiantes considera un impacto muy significativo.



- t. **Pregunta 20:** ¿Con que frecuencia has descuidado a personas de tu entorno (p. ej., familia, amigos) por estar jugando?

Gráfica 22:

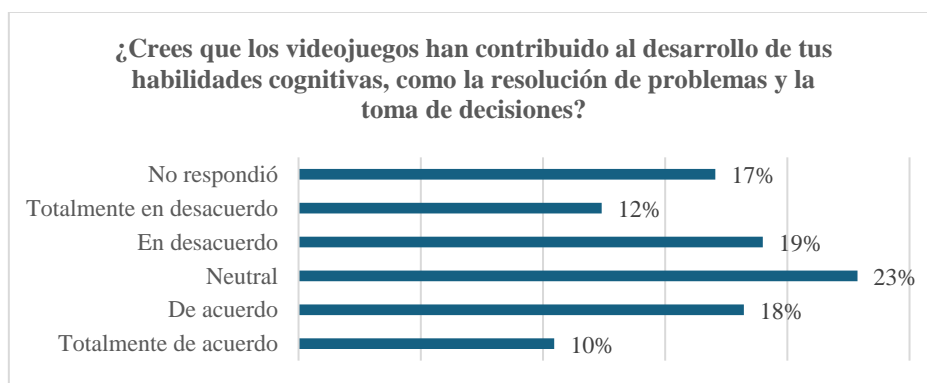
Gráfico de barras: Pregunta 20



- El 17% de los estudiantes indicó que 'raramente' han descuidado a personas de su entorno, un 17% dijo que esto ocurre 'ocasionalmente', y un 26% afirmó que sucede 'frecuentemente', por otra parte, el 7% dijo que 'nunca' ha descuidado a las personas por jugar videojuegos
- u. **Pregunta 21:** ¿Crees que los videojuegos han contribuido al desarrollo de tus habilidades cognitivas, como la resolución de problemas y la toma de decisiones?

Gráfica 23:

Gráfico de barras: Pregunta 21



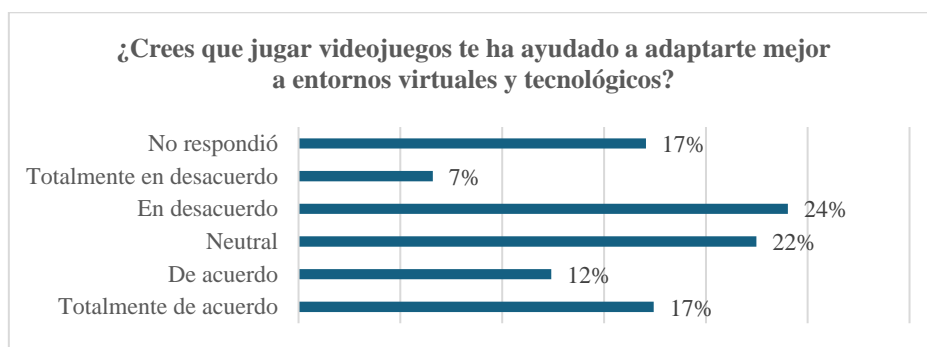


- El 10% de los encuestados está “totalmente de acuerdo” en que jugar videojuegos ha contribuido significativamente al desarrollo de sus habilidades cognitivas, mientras que el 18% está “de acuerdo” con la premisa.

v. **Pregunta 22:** ¿Crees que jugar videojuegos te ha ayudado a adaptarte mejor a entornos virtuales y tecnológicos?

Gráfica 24:

Gráfico de barras: Pregunta 22



- El 12% de los encuestados está de acuerdo en que jugar videojuegos les ha ayudado a adaptarse mejor a entornos virtuales y tecnológicos, mientras que el 17% está totalmente de acuerdo



w. **Pregunta 23:** ¿Crees que jugar videojuegos ha mejorado tu capacidad para realizar tareas múltiples eficientemente?

Gráfica 25:

Gráfico de barras: Pregunta 23



- El 16% de los estudiantes está de acuerdo en que los videojuegos han mejorado su capacidad para realizar múltiples tareas de manera eficiente, mientras que el 10% está totalmente de acuerdo.

x. **Pregunta 24:** ¿Has proyectado las habilidades aprendidas al jugar videojuegos a situaciones de la vida real?

Gráfica 26:

Gráfico de barras: Pregunta 24



- La mayoría de los estudiantes (29%) respondió que “raramente” han aplicado las habilidades aprendidas al jugar videojuegos en situaciones



de la vida real, mientras que un 11% afirmó que las utiliza con frecuencia y un 8% que siempre las emplea en la vida cotidiana.

4.2.4. Evaluación de la calidad de los datos

En esta etapa se realiza una evaluación de la calidad de los datos, abordando las siguientes preguntas:

- **¿Los datos que se obtuvieron están completos?** La información solicitada a la Dirección de Evaluación y Desarrollo Académico (DEDA), está completa ya que lleva el historial académico de los estudiantes de las diferentes escuelas profesionales de la UNS. En relación con la encuesta aplicada a los estudiantes de la EPISI matriculados en el periodo académico 2023-I, el 17% de la muestra prevista según el reporte alcanzado por DEDA no respondieron. Las dificultades encontradas fueron: algunos estudiantes se habían retirado, estudiantes que no tenían matrícula vigente, o decidieron no contestar la encuesta.
- **¿Hay duplicados o inconsistencias entre registros similares?** E La información del historial académico de los estudiantes proporcionada por DEDA no presenta duplicidad, ya que se extrae directamente del sistema académico de la UNS. Asimismo, los datos obtenidos de la encuesta están libres de duplicados, gracias a la configuración de Google Forms que recopila los correos electrónicos y limita las respuestas a una por persona. Además, se verificó que los códigos y datos que los estudiantes ingresaron fueran únicos.



- **¿Los datos recolectados son útiles y pertinentes para el análisis?**

Sí, los datos recolectados son útiles y pertinentes para el análisis. Serán de gran ayuda en el desarrollo del proyecto, ya que proporcionan la información necesaria para obtener resultados precisos y relevantes.

4.3. FASE III: Preparación de los datos

4.3.1. Selección de los datos relevantes

En este proceso se tomaron en consideración las siguientes técnicas de recolección de datos.

a. Requerimiento de información

De los datos mencionados en la caracterización, de los cuales, los que fueron identificados como relevantes para este estudio, son el código de estudiante y el promedio ponderado por ciclo.

b. Encuesta

Los datos obtenidos a través de la encuesta fueron procesados y puntuados utilizando la escala de Likert. Posteriormente, se exploraron diversos métodos estadísticos para analizar la relación entre las dimensiones (variable categórica) y el promedio ponderado (variable numérica).



4.3.2. Limpieza de datos

Como se puede apreciar en la *Figura 3*, al analizar la información del dataset utilizando Python, se observa que el dataset consta de 31 columnas que representan diferentes variables del estudio. Así mismo, se detecta la presencia de registros vacíos en algunas de estas columnas, lo que indica la necesidad de un proceso de limpieza de datos antes de proceder con el análisis.

Figura 2:

Metadatos del dataset

```
1 df_analisisTesis.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 258 entries, 0 to 257
Data columns (total 31 columns):
#   Column                                                                 Non-Null Count  Dtype
---  -
0   Codigo Estudiante                                                    258 non-null    int64
1   Apellidos y Nombres                                                  258 non-null    object
2   Ponderado                                                            240 non-null    float64
3   Sexo                                                                214 non-null    object
4   Ingresa tu edad                                                      214 non-null    float64
5   ¿Actualmente trabajas?                                              214 non-null    object
6   Si en la pregunta anterior tu respuesta fue "Sí": ¿Cuál es tu jornada laboral? 214 non-null    object
7   Pregunta 1                                                           214 non-null    object
8   Pregunta 2                                                           214 non-null    object
9   Pregunta 3                                                           214 non-null    object
10  Pregunta 4                                                           214 non-null    object
11  Pregunta 5                                                           214 non-null    object
12  Pregunta 6                                                           214 non-null    object
13  Pregunta 7                                                           214 non-null    object
14  Pregunta 8                                                           214 non-null    object
15  Pregunta 9                                                           214 non-null    object
16  Pregunta 10                                                          214 non-null    object
17  Pregunta 11                                                          214 non-null    object
18  Pregunta 12                                                          214 non-null    object
19  Pregunta 13                                                          214 non-null    object
20  Pregunta 14                                                          214 non-null    object
21  Pregunta 15                                                          214 non-null    object
22  Pregunta 16                                                          214 non-null    object
23  Pregunta 17                                                          214 non-null    object
24  Pregunta 18                                                          214 non-null    object
25  Pregunta 19                                                          214 non-null    object
26  Pregunta 20                                                          214 non-null    object
27  Pregunta 21                                                          214 non-null    object
28  Pregunta 22                                                          214 non-null    object
29  Pregunta 23                                                          214 non-null    object
30  Pregunta 24                                                          214 non-null    object
dtypes: float64(2), int64(1), object(28)
memory usage: 62.6+ KB
```

a. Selección de datos relevantes en Python

Se puede observar en la *Figura 4* que de las 31 columnas del dataset original, nos quedamos con 25, las cuales son relevantes para el análisis.

Figura 3:

Metadatos del dataset de las columnas seleccionadas

```
1 #SELECCION DE COLUMNAS CON DATOS RELEVANTES
2 columnas_seleccionadas = ['Ponderado', 'Pregunta 1',
3                             'Pregunta 2', 'Pregunta 3', 'Pregunta 4', 'Pregunta 5',
4                             'Pregunta 6', 'Pregunta 7', 'Pregunta 8', 'Pregunta 9', 'Pregunta 10',
5                             'Pregunta 11', 'Pregunta 12', 'Pregunta 13', 'Pregunta 14',
6                             'Pregunta 15', 'Pregunta 16', 'Pregunta 17', 'Pregunta 18',
7                             'Pregunta 19', 'Pregunta 20', 'Pregunta 21', 'Pregunta 22',
8                             'Pregunta 23', 'Pregunta 24']
9 df_analisisTesisLimpio = df_analisisTesis[columnas_seleccionadas]
10 df_analisisTesisLimpio.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 258 entries, 0 to 257
Data columns (total 25 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   Ponderado              240 non-null   float64
1   Pregunta 1            214 non-null   object
2   Pregunta 2            214 non-null   object
3   Pregunta 3            214 non-null   object
4   Pregunta 4            214 non-null   object
5   Pregunta 5            214 non-null   object
6   Pregunta 6            214 non-null   object
7   Pregunta 7            214 non-null   object
8   Pregunta 8            214 non-null   object
9   Pregunta 9            214 non-null   object
10  Pregunta 10           214 non-null   object
11  Pregunta 11           214 non-null   object
12  Pregunta 12           214 non-null   object
13  Pregunta 13           214 non-null   object
14  Pregunta 14           214 non-null   object
15  Pregunta 15           214 non-null   object
16  Pregunta 16           214 non-null   object
17  Pregunta 17           214 non-null   object
18  Pregunta 18           214 non-null   object
19  Pregunta 19           214 non-null   object
20  Pregunta 20           214 non-null   object
21  Pregunta 21           214 non-null   object
22  Pregunta 22           214 non-null   object
23  Pregunta 23           214 non-null   object
24  Pregunta 24           214 non-null   object
dtypes: float64(1), object(24)
memory usage: 50.5+ KB
```

b. Eliminación de registros vacíos en Python

Se puede observar en la *Figura 5* que los registros que figuraban con valores vacíos se han eliminado, quedando así para el análisis un total de 214 entradas.

Figura 4:

Metadatos del dataset limpio

```
1 # REMOVER REGISTROS VACÍOS  
2 df_analisisTesisLimpios = df_analisisTesisLimpios.dropna()  
3 df_analisisTesisLimpios.info()  
4 df_analisisTesisLimpios.head()
```

```
<class 'pandas.core.frame.DataFrame'>  
Index: 214 entries, 0 to 239  
Data columns (total 25 columns):  
#   Column          Non-Null Count  Dtype  
---  ---  
0   Ponderado        214 non-null    float64  
1   Pregunta 1       214 non-null    object  
2   Pregunta 2       214 non-null    object  
3   Pregunta 3       214 non-null    object  
4   Pregunta 4       214 non-null    object  
5   Pregunta 5       214 non-null    object  
6   Pregunta 6       214 non-null    object  
7   Pregunta 7       214 non-null    object  
8   Pregunta 8       214 non-null    object  
9   Pregunta 9       214 non-null    object  
10  Pregunta 10      214 non-null    object  
11  Pregunta 11      214 non-null    object  
12  Pregunta 12      214 non-null    object  
13  Pregunta 13      214 non-null    object  
14  Pregunta 14      214 non-null    object  
15  Pregunta 15      214 non-null    object  
16  Pregunta 16      214 non-null    object  
17  Pregunta 17      214 non-null    object  
18  Pregunta 18      214 non-null    object  
19  Pregunta 19      214 non-null    object  
20  Pregunta 20      214 non-null    object  
21  Pregunta 21      214 non-null    object  
22  Pregunta 22      214 non-null    object  
23  Pregunta 23      214 non-null    object  
24  Pregunta 24      214 non-null    object  
dtypes: float64(1), object(24)  
memory usage: 43.5+ KB
```

4.3.3. Transformación de datos

En la *Figura 6* se muestran los 5 primeros registros del dataset en donde se observan los datos categóricos de la encuesta.

Figura 5:

Vista Previa del dataset con variables categóricas

Ponderado	Pregunta 1	Pregunta 2	Pregunta 3	Pregunta 4	Pregunta 5	Pregunta 6	Pregunta 7	Pregunta 8	Pregunta 9	...	Pregunta 15	Pregunta 16
0.0000	S/. 500.00 - S/. 750.00	S/. 200.00 - S/. 300.00	Frecuentemente	Neutral	Todos los días	De 4 a 6 horas	Madrugada (12 am - 05 am)	Frecuentemente	Muy negativo	...	Frecuentemente	Cambios muy significativos
0.2222	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN
0.9524	S/. 1000.00 a más	S/. 300.00 - S/. 400.00	Siempre	Totalmente de acuerdo	Todos los días	Más de 6 horas	Noche (08 pm - 12 am)	Siempre	Negativo	...	Siempre	Cambios muy significativos
1.8947	S/. 1000.00 a más	S/. 500.00 a más	Siempre	Totalmente de acuerdo	Todos los días	Más de 6 horas	Madrugada (12 am - 05 am)	Siempre	Muy negativo	...	Siempre	Cambios muy significativos
2.1905	S/. 1000.00 a más	S/. 500.00 a más	Siempre	Totalmente de acuerdo	Todos los días	Más de 6 horas	Madrugada (12 am - 05 am)	Siempre	Muy negativo	...	Siempre	Cambios muy significativos

Sin embargo, para un correcto análisis, los datos fueron transformados mediante normalización y discretización, de variables categóricas a numéricas, según la puntuación establecida en la FASE II: Comprensión de los Datos.

Figura 6:

Vista Previa del Dataset con variables numéricas

Ponderado	Pregunta 1	Pregunta 2	Pregunta 3	Pregunta 4	Pregunta 5	Pregunta 6	Pregunta 7	Pregunta 8	Pregunta 9	...	Pregunta 15	Pregunta 16
0.0000	2	2	3	2	4	3	4	3	4	...	3	4
0.9524	4	3	4	4	4	4	3	4	3	...	4	4
1.8947	4	4	4	4	4	4	4	4	4	...	4	4
2.1905	4	4	4	4	4	4	4	4	4	...	4	4
3.2700	4	3	4	4	4	4	4	2	2	...	4	4



Figura 7:

Transformación de valores categóricos a numéricos según la escala de Likert en Python

```
1 #CAMBIAR LOS DATOS
2
3 # Definir el mapeo según la escala de Likert
4 mapeo = {
5     'Nunca': 0, 'Totalmente en desacuerdo': 0, 'S/. 0.00 - S/. 250.00': 0,
6     'Muy positivo': 0, 'Insignificadamente': 0, 'Ningún cambio': 0,
7     'S/. 0.00 - S/. 100.00': 0, 'Entre 0 a 1 días': 0,
8     'Menos de 1 hora': 0, 'Ningún impacto': 0, 'Mañana (05 am - 10 am)': 0,
9
10    'Raramente': 1, 'En desacuerdo': 1, 'Positivo': 1, 'Levemente': 1,
11    'Cambios mínimos': 1, 'S/. 250.00 - S/. 500.00': 1,
12    'S/. 100.00 - S/. 200.00': 1, 'Entre 1 a 3 días': 1,
13    'De 1 a 2 horas': 1, 'Impacto mínimo': 1, 'Mediodía (10 am - 03 pm)': 1,
14
15    'Ocasionalmente': 2, 'Neutral': 2, 'No sabe / No respondes': 2,
16    'Cambios moderados': 2, 'S/. 500.00 - S/. 750.00': 2,
17    'S/. 200.00 - S/. 300.00': 2, 'Entre 3 a 5 días': 2,
18    'De 2 a 4 horas': 2, 'Impacto moderado': 2, 'Tarde (03 pm - 08 pm)': 2,
19
20    'Frecuentemente': 3, 'De acuerdo': 3, 'Negativo': 3,
21    'Moderadamente': 3, 'Cambios significativos': 3,
22    'S/. 750.00 - S/. 1000.00': 3, 'S/. 300.00 - S/. 400.00': 3,
23    'Entre 5 a 6 días': 3, 'De 4 a 6 horas': 3, 'Impacto significativo': 3,
24    'Noche (08 pm - 12 am)': 3,
25
26    'Siempre': 4, 'Totalmente de acuerdo': 4, 'Muy negativo': 4,
27    'Significativamente': 4, 'Cambios muy significativos': 4,
28    'S/. 1000.00 a más': 4, 'S/. 500.00 a más': 4, 'Todos los días': 4,
29    'Más de 6 horas': 4, 'Impacto muy significativo': 4,
30    'Madrugada (12 am - 05 am)': 4
31 }
32
33 # Aplicar el mapeo a todo el DataFrame
34 df_analisisTesisLimpios.replace(mapeo, inplace=True)
35 df_analisisTesisLimpios.info()
```

4.3.4. Consolidación de datos

En este proceso a través de Python se generaron las dimensiones a través de agregaciones, como se establece en la operacionalización la observación de la variable 1 (Consumo de videojuegos).

Figura 8:

Generación de dimensiones en Python

```
1 #SUMA DE LAS COLUMNAS Y GENERACIÓN DE LAS DIMENSIONES DEL ESTUDIO
2 #CAPACIDAD ECONOMICA
3 df_analisisTesisLimpios['CAPACIDAD ECONOMICA'] = df_analisisTesisLimpios['Pregunta 1'] +
4 df_analisisTesisLimpios['Pregunta 2'] +
5 df_analisisTesisLimpios['Pregunta 3'] +
6 df_analisisTesisLimpios['Pregunta 4']
7
8 #TIEMPO DE CONSUMO DE VIDEOJUEGOS
9 df_analisisTesisLimpios['TIEMPO DE CONSUMO DE VIDEOJUEGOS'] = df_analisisTesisLimpios['Pregunta 5'] +
10 df_analisisTesisLimpios['Pregunta 6'] +
11 df_analisisTesisLimpios['Pregunta 7'] +
12 df_analisisTesisLimpios['Pregunta 8']
13
14 #BIENESTAR FISICO-MENTAL
15 df_analisisTesisLimpios['BIENESTAR FISICO-MENTAL'] = df_analisisTesisLimpios['Pregunta 9'] +
16 df_analisisTesisLimpios['Pregunta 10'] +
17 df_analisisTesisLimpios['Pregunta 11'] +
18 df_analisisTesisLimpios['Pregunta 12']
19
20 #IMPACTO SOBRE ACTIVIDADES COTIDIANAS
21 df_analisisTesisLimpios['IMPACTO SOBRE ACTIVIDADES COTIDIANAS'] = df_analisisTesisLimpios['Pregunta 13'] +
22 df_analisisTesisLimpios['Pregunta 14'] +
23 df_analisisTesisLimpios['Pregunta 15'] +
24 df_analisisTesisLimpios['Pregunta 16']
25
26 #RELACIONES SOCIALES
27 df_analisisTesisLimpios['RELACIONES SOCIALES'] = df_analisisTesisLimpios['Pregunta 17'] +
28 df_analisisTesisLimpios['Pregunta 18'] +
29 df_analisisTesisLimpios['Pregunta 19'] +
30 df_analisisTesisLimpios['Pregunta 20']
31
32 #DESARROLLO COGNITIVO
33 df_analisisTesisLimpios['DESARROLLO COGNITIVO'] = df_analisisTesisLimpios['Pregunta 21'] +
34 df_analisisTesisLimpios['Pregunta 22'] +
35 df_analisisTesisLimpios['Pregunta 23'] +
36 df_analisisTesisLimpios['Pregunta 24']
37
38 #Seleccionamos las columnas del dataset final
39 columnas_seleccionadas = ['Ponderado',
40 'CAPACIDAD ECONOMICA', 'TIEMPO DE CONSUMO DE VIDEOJUEGOS',
41 'BIENESTAR FISICO-MENTAL', 'IMPACTO SOBRE ACTIVIDADES COTIDIANAS',
42 'RELACIONES SOCIALES', 'DESARROLLO COGNITIVO']
43
44 df_analisisTesisLimpios = df_analisisTesisLimpios[columnas_seleccionadas]
45 df_analisisTesisLimpios.info()
46 df_analisisTesisLimpios.head()
```




Figura 9:

Metadata de Dataset final

```
<class 'pandas.core.frame.DataFrame'>
Index: 214 entries, 0 to 239
Data columns (total 7 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0   Ponderado                                 214 non-null    float64
1   CAPACIDAD ECONOMICA                     214 non-null    int64
2   TIEMPO DE CONSUMO DE VIDEOJUEGOS        214 non-null    int64
3   BIENESTAR FISICO-MENTAL                 214 non-null    int64
4   IMPACTO SOBRE ACTIVIDADES COTIDIANAS    214 non-null    int64
5   RELACIONES SOCIALES                    214 non-null    int64
6   DESARROLLO COGNITIVO                   214 non-null    int64
dtypes: float64(1), int64(6)
memory usage: 13.4 KB
```

Figura 10:

Vista previa de Dataset final

Ponderado	CAPACIDAD ECONOMICA	TIEMPO DE CONSUMO DE VIDEOJUEGOS	BIENESTAR FISICO-MENTAL	IMPACTO SOBRE ACTIVIDADES COTIDIANAS	RELACIONES SOCIALES	DESARROLLO COGNITIVO
0.0000	9	14	16	15	12	16
0.9524	15	15	15	15	15	16
1.8947	16	16	16	16	16	16
2.1905	16	16	16	15	16	16
3.2700	15	14	9	15	14	10

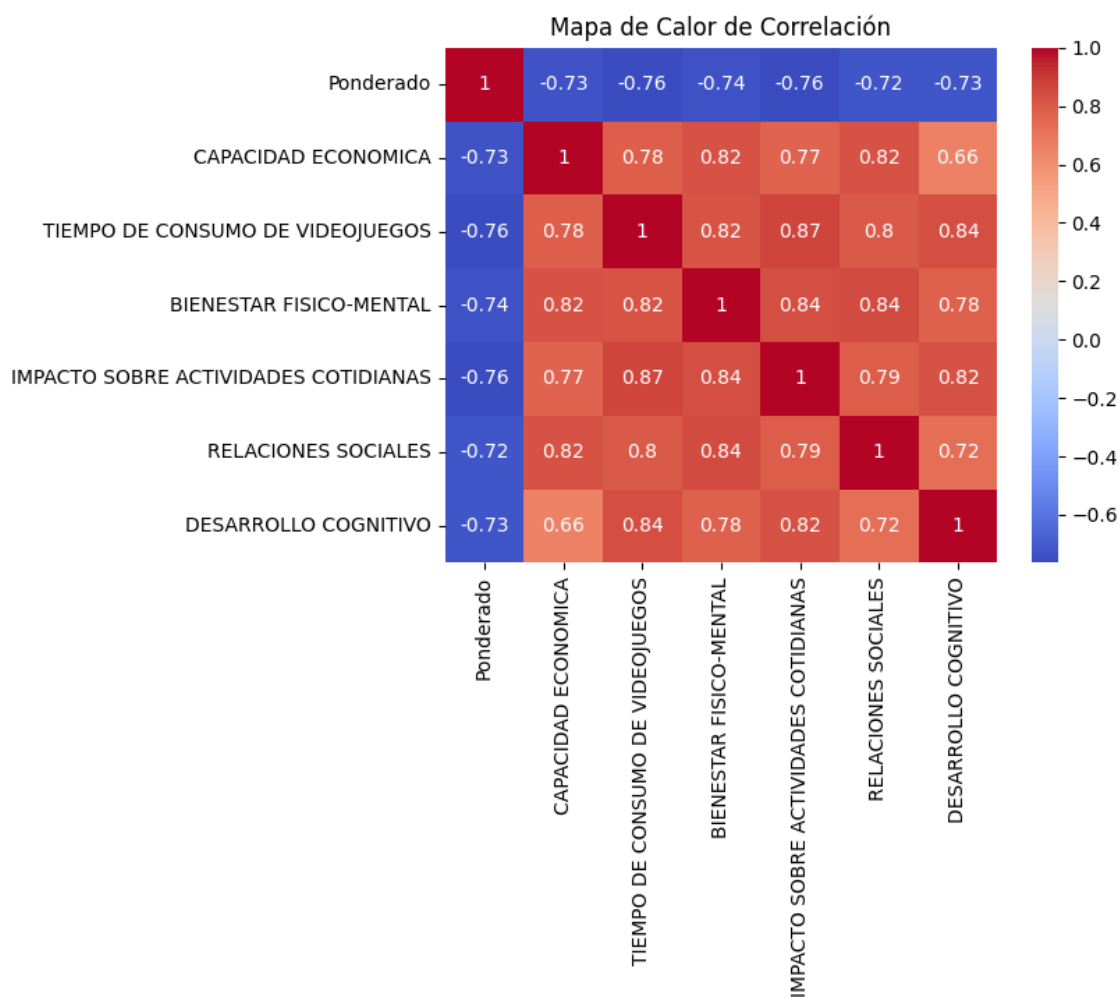
4.3.5.Reducción de la dimensionalidad de la encuesta

Se empleó la correlación de Pearson como método estadístico, el cual permite comparar la relación de las medidas de las seis dimensiones simultáneamente e identificar la intensidad en la que se relacionan con el promedio ponderado. Para este análisis, se utilizó Python a través de las librerías, Numpy, Matplotlib, Pandas y Seaborn, utilizando las puntuaciones de cada dimensión y el promedio ponderado de los estudiantes, de dicho análisis se obtuvieron los siguientes resultados:



Figura 11:

Mapa de Calor entre las 6 dimensiones y el promedio ponderado de los estudiantes



Como se puede observar en el mapa de calor, todos los resultados de la correlación todas las dimensiones son altas y se relacionan de manera inversa con el promedio ponderado. Esto sugiere que dimensiones como la capacidad económica, el tiempo de consumo de videojuegos, el bienestar físico-mental, el impacto en las actividades cotidianas, las relaciones sociales y el desarrollo cognitivo son importantes para entender las variaciones en los promedios ponderados de los estudiantes encuestados.



4.4. FASE IV: Modelado

En este proceso se realizará la elección y aplicación de las diferentes técnicas de modelados que sean pertinentes.

Para el desarrollo de los modelos se empleará Python con sus diferentes librerías, a través de Google Colab. Donde se entrenarán 5 modelos supervisados compatibles con los datos recolectados.

4.4.1. Modelados por Regresión y métricas de comprobación

a. Algoritmo de Regresión Lineal

Para poder entrenar este algoritmo necesitamos acumular todos los valores de las 6 dimensiones (“Capacidad Económica”, Tiempo de Consumo de Videojuegos”, Bienestar Físico-Mental”, “Impacto sobre actividades cotidianas”, “Relaciones sociales” y “Desarrollo Cognitivo”) de nuestra encuesta en una sola, la cual denominaremos “Puntaje Encuesta”, este puntaje obtenido nos servirá para predecir el promedio ponderado de los estudiantes como se puede observar en la *Tabla 4*.

Después de haber entrenado el modelo, se obtuvo la ecuación lineal ideal que cumple el criterio de los mínimos cuadrados, en donde: $y' = w + xb$.

Tabla 3:

Evaluación del modelo de Regresión Linear

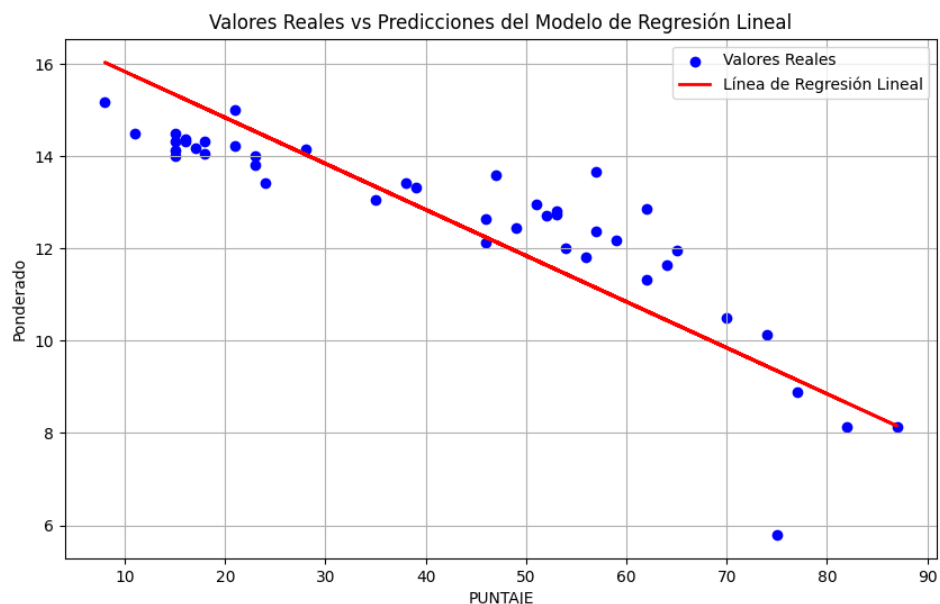
Evaluación del modelo de Regresión Linear	Resultados
Intercepto (w)	16.83345331583186
Pendiente (b)	-0.0997846
MSE	1.28
R ²	0.67

Así mismo, con el modelo de regresión lineal se obtuvo que el cuadrado de la diferencia entre los valores predichos y los valores reales (MSE) es de 1.28, siendo un valor razonable para la predicción, con una tasa de error no tan grande y un R^2 de 0.67 o un porcentaje de predicción del 67% (generalmente considerado aceptable).

A continuación, en la *Figura 12*, se muestra la gráfica de dispersión con la línea ideal obtenida por el modelo.

Figura 12:

Valores Reales vs Predicciones del Modelo de Regresión Lineal



Así mismo, hacemos la comparativa del dataset de prueba (Valor real) con los promedios ponderados predichos por el modelo, en la *Tabla 4*, se pueden observar los 15 primeros registros que nos arroja el modelo de Regresión Lineal.



Tabla 4:

Valores reales vs valores predichos del modelo de Regresión Lineal

Valor Real	Valores predichos	Margen de error
5.8000	9.349608	3.549608
14.3600	15.236900	0.8769
11.9500	10.347454	-1.602546
14.3200	15.236900	0.9169
13.0455	13.340992	0.295492
12.9500	11.744439	-1.205561
8.1333	8.152193	0.018893
12.4545	11.944008	-0.510492
12.1818	10.946162	-1.235638
13.4100	13.041638	-0.368362
14.1818	15.137115	0.955315
14.0000	15.336684	1.336684
13.4091	14.438623	1.029523
12.1364	12.243362	0.106962
14.1400	12.243362	-1.896638

b. Algoritmo de Regresión Múltiple

Para generar el algoritmo del modelo de regresión múltiple se emplearon los valores de las 6 dimensiones (“Capacidad Económica”, Tiempo de Consumo de Videojuegos”, Bienestar Físico-Mental”, “Impacto sobre actividades cotidianas”, “Relaciones sociales” y “Desarrollo Cognitivo”) de nuestra encuesta, las cuales nos servirán para predecir el promedio ponderado de los estudiantes.

Después de haber entrenado el modelo, se obtuvo la ecuación lineal ideal que cumple el criterio de los mínimos cuadrados, en donde:

$$y' = w + \beta_1 \cdot \chi_1 + \beta_2 \cdot \chi_2 + \beta_3 \cdot \chi_3 + \beta_4 \cdot \chi_4 + \beta_5 \cdot \chi_5 + \beta_6 \cdot \chi_6$$

Donde:

- X_1 = Capacidad Económica
- X_2 = Tiempo de consumo de Videojuegos
- X_3 = Bienestar Físico - Mental
- X_4 = Impacto sobre Actividades Cotidianas
- X_5 = Relaciones Sociales
- X_6 = Desarrollo Cognitivo

Tabla 5:

Evaluación del modelo de Regresión Múltiple

Evaluación del modelo de Regresión Múltiple	Resultados
Intercepto (w)	-0.0997846
MSE	1.29
R ²	0.67
Coeficiente (β) por dimensión	
Capacidad Económica (β_1)	-0.190329
Tiempo de consumo de Videojuegos (β_2)	-0.063407
Bienestar Físico-Mental (β_3)	-0.032532
Impacto sobre actividades cotidianas (β_4)	-0.105626
Relaciones Sociales (β_5)	-0.063706
Desarrollo Cognitivo (β_6)	-0.149730

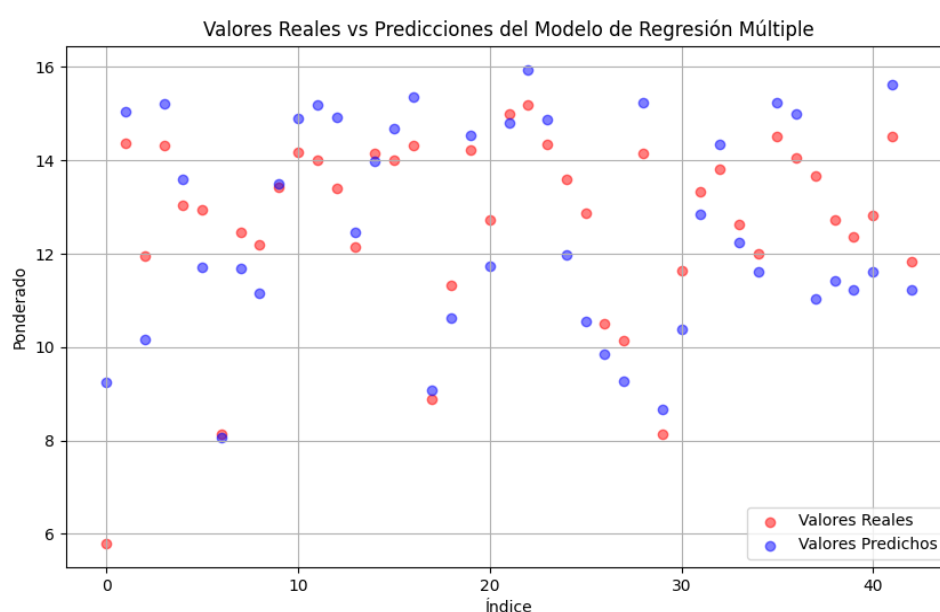
Así mismo, como se muestra en la *Tabla 5* con el modelo de regresión múltiple se obtuvo que el cuadrado de la diferencia entre los valores predichos y los valores reales (MSE) es de 1.29, ligeramente más alto que el valor obtenido en la regresión lineal, siendo un valor razonable para la

predicción, con una tasa de error no tan grande y al igual que en la regresión lineal se obtuvo un R^2 de 0.67 o un porcentaje de predicción del 67% (generalmente considerado aceptable).

A continuación, en la Figura 12, se muestra la gráfica de dispersión con la línea ideal obtenida por el modelo de Regresión Múltiple.

Figura 13:

Valores Reales vs Predicciones del Modelo de Regresión Múltiple



Eje X (Índice):

Representa los índices de las muestras en el conjunto de prueba.

Cada punto en el eje X corresponde a una observación específica del conjunto de prueba.

Eje Y (Ponderado):

Representa los valores de la variable objetivo (Ponderado).

Puntos Rojos (Valores Reales):

Estos puntos representan los valores reales de la variable objetivo en el conjunto de prueba.



Puntos Azules (Valores Predichos):

Estos puntos representan los valores predichos por el modelo de regresión múltiple para las mismas observaciones.

La gráfica sugiere que el modelo tiene un rendimiento aceptable, ya que muchas de las predicciones están cerca de los valores reales.

Así mismo, hacemos la comparativa del dataset de prueba (Valor real) con los promedios ponderados predichos por el modelo, en la *Tabla 6*, se pueden observar los 15 primeros registros que nos arroja el modelo de Regresión Múltiple.

Tabla 6:

Valores reales vs valores predichos del modelo de Regresión Lineal

Valor Real	Valores predichos	Margen de error
5.8000	9.246952	3.446952
14.3600	15.037976	0.677976
11.9500	10.172522	-1.777478
14.3200	15.206818	0.886818
13.0455	13.594008	0.548508
12.9500	11.712954	-1.237046
8.1333	8.063886	-0.069414
12.4545	11.678435	-0.776065
12.1818	11.152856	-1.028944
13.4100	13.491469	0.081469
14.1818	14.889567	0.707767
14.0000	15.175798	1.175798
13.4091	14.907588	1.498488
12.1364	12.453239	0.316839
14.1400	13.976028	-0.163972



c. **Random Forest**

Para desarrollar el algoritmo del modelo de Random Forest, se utilizaron los valores de las 6 dimensiones de nuestra encuesta: "Capacidad Económica", "Tiempo de Consumo de Videojuegos", "Bienestar Físico-Mental", "Impacto sobre Actividades Cotidianas", "Relaciones Sociales" y "Desarrollo Cognitivo". Estas dimensiones se considerarán como variables predictoras del promedio ponderado de los estudiantes.

Tabla 7:

Evaluación del modelo de Random Forest

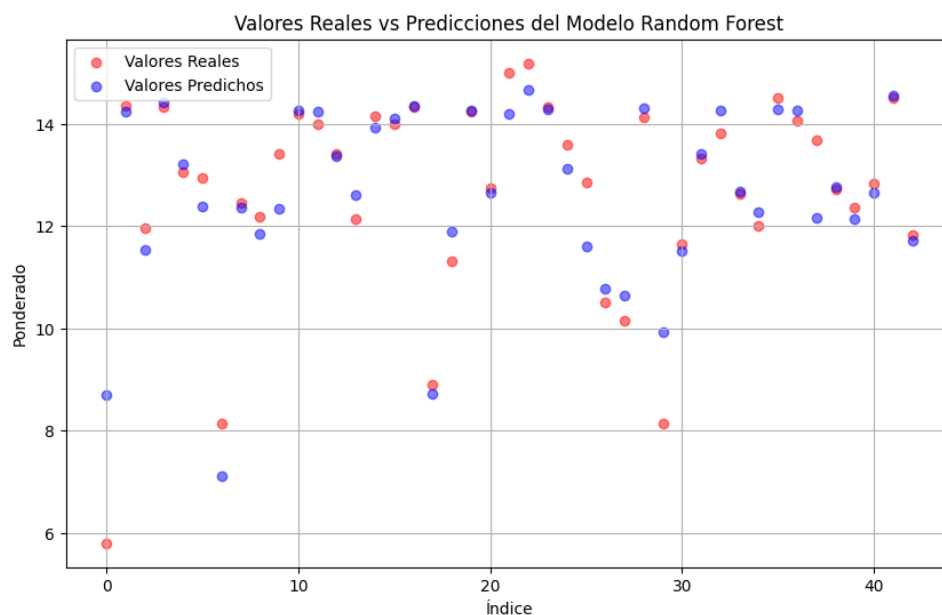
Evaluación del modelo de Random Forest	Resultados
MSE	0.48
R ²	0.875
Importancia de cada dimensión	
Impacto sobre actividades cotidianas	0.278998
Desarrollo Cognitivo	0.241202
Tiempo de consumo de Videojuegos	0.162419
Relaciones Sociales	0.145197
Bienestar Físico-Mental	0.103265
Capacidad Económica	0.068918

El modelo Random Forest ha demostrado ser bastante eficaz en la predicción del ponderado, con un alto coeficiente de determinación (0.875) lo cual indica un porcentaje de predicción del 87.5%, así mismo, presenta un bajo error cuadrático medio (0.48) cercano a 0. Las características más importantes destacan la relevancia del impacto en actividades cotidianas y

el desarrollo cognitivo en el rendimiento académico, sugiriendo que estas áreas son cruciales para entender y mejorar el desempeño académico de los estudiantes.

Figura 14:

Valores Reales vs Predicciones del Modelo Random Forest



Como se muestra en la *Figura 14*, el modelo Random Forest realiza predicciones precisas, con los valores reales y predichos del "Ponderado" generalmente cercanos, reflejando un buen rendimiento del modelo. La cercanía de la mayoría de los puntos rojos (reales) y azules (predichos) valida visualmente la alta precisión indicada por las métricas, destacando la eficacia del modelo en predecir el rendimiento académico basado en las características proporcionadas. Aunque hay algunas diferencias menores, el modelo demuestra una robusta capacidad predictiva en general.

Así mismo, hacemos la comparativa del dataset de prueba (Valor real) con los promedios ponderados predichos por el modelo, en la *Tabla 7*, se



pueden observar los 15 primeros registros que nos arroja el modelo Random Forest.

Tabla 8:

Valores reales vs valores predichos del modelo Random Forest

Valor Real	Valores predichos	Margen de error
5.8000	8.688788	2.888788
14.3600	14.238322	-0.121678
11.9500	11.527300	-0.4227
14.3200	14.425454	0.105454
13.0455	13.205112	0.159612
12.9500	12.383378	-0.566622
8.1333	7.102498	-1.030802
12.4545	12.357334	-0.097166
12.1818	11.841793	-0.340007
13.4100	12.335326	-1.074674
14.1818	14.261216	0.079416
14.0000	14.233666	0.233666
13.4091	13.366555	-0.042545
12.1364	12.601807	0.465407
14.1400	13.917812	-0.222188

4.4.2. Modelados por Clasificación y métricas de comprobación

Para generar los siguientes modelos, clasificaremos el promedio ponderado con la condición de ser 0 o desaprobado si es menor o igual a 11 y 1 o aprobado si es mayor.

a. Regresión Logística

Para generar el algoritmo del modelo de regresión logística se emplearon los valores de las 6 dimensiones (“Capacidad Económica”, Tiempo de Consumo de Videojuegos”, Bienestar Físico-Mental”, “Impacto sobre



actividades cotidianas”, “Relaciones sociales” y “Desarrollo Cognitivo”) de nuestra encuesta, las cuales se considerarán como variables predictoras del promedio ponderado de los estudiantes.

Tabla 9:

Evaluación del modelo de Regresión Logística

Evaluación del modelo de Regresión Logística	Resultados
Coefficientes por variable predictora	
Impacto sobre actividades cotidianas	-0.406083
Desarrollo Cognitivo	-0.235949
Tiempo de consumo de Videojuegos	-0.819954
Relaciones Sociales	-0.140963
Bienestar Físico-Mental	0.097621
Capacidad Económica	0.142997

El modelo de regresión logística sugiere que el rendimiento académico (ponderado) de los estudiantes, puede ser influenciado por varios factores. Un mayor tiempo de consumo de videojuegos y un mayor impacto sobre las actividades cotidianas están asociados con una disminución en el rendimiento académico, mientras que una mayor capacidad económica y un mejor bienestar físico-mental muestran una ligera asociación positiva con el rendimiento académico. Las relaciones sociales y el desarrollo cognitivo también afectan negativamente el rendimiento académico.

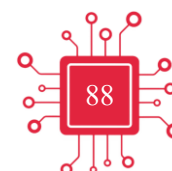
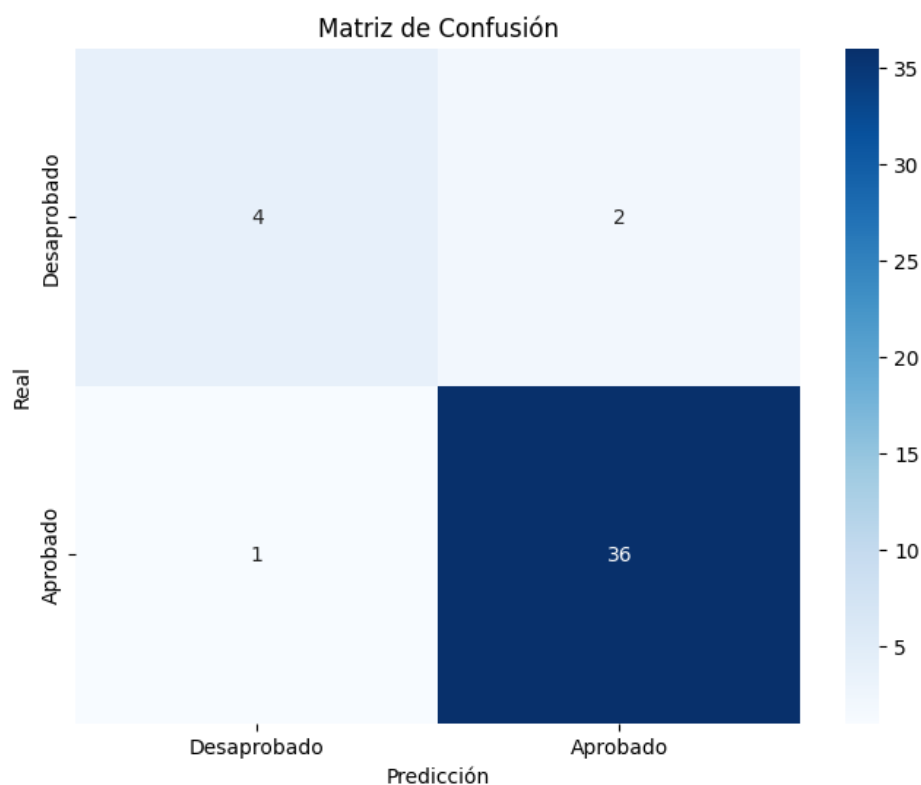


Figura 15:

Matriz de confusión del modelo de la Regresión Logística



Desglose:

- **Verdaderos Negativos (TN): 4**

Casos donde el modelo predijo "Desaprobado" y realmente eran "Desaprobado".

- **Falsos Positivos (FP): 2**

Casos donde el modelo predijo "Aprobado" pero realmente eran "Desaprobado".

- **Falsos Negativos (FN): 1**

Casos donde el modelo predijo "Desaprobado" pero realmente eran "Aprobado".



- **Verdaderos Positivos (TP): 36**

Casos donde el modelo predijo "Aprobado" y realmente eran "Aprobado".

Métricas de Evaluación:

- **Precisión:**

$$P = \frac{TP + TN}{Total} = \frac{36 + 4}{43} = 0.9302 \approx 93\%$$

La precisión indica que el modelo clasifica correctamente el 93% de los casos.

- **Precisión para Aprobado:**

$$P = \frac{TP}{TP + FP} = \frac{36}{36 + 2} = 0.9474 \approx 95\%$$

La precisión mide la proporción de verdaderos positivos sobre el total de predicciones positivas.

- **Recall para Aprobado**

$$R = \frac{TP}{TP + FN} = \frac{36}{36 + 1} = 0.9729 \approx 97\%$$

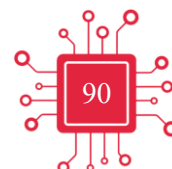
El recall mide la capacidad del modelo para encontrar todos los positivos reales.

- **F1 Score para Aprobado**

$$F1 = 2 * \frac{P * R}{P + R} = 2 * \frac{0.9474 * 0.9729}{0.9474 + 0.9729} = 0.96 \approx 96\%$$

El F1 score es la media armónica de la precisión y el recall, proporcionando un balance entre ambos.

- El modelo de regresión logística tiene un buen rendimiento general, con una alta precisión del 93%.





- La precisión y el recall para la clase "Aprobado" son altos, indicando que el modelo es efectivo en la identificación de estudiantes aprobados.
- La baja cantidad de falsos positivos (2) y falsos negativos (1) muestra que el modelo no tiende a cometer muchos errores de clasificación.

Este análisis sugiere que el modelo es confiable para predecir la aprobación de los estudiantes en base a las variables incluidas.

b. Naive Bayes

Para generar el algoritmo del modelo de Naive Bayes se emplearon los valores de las 6 dimensiones (“Capacidad Económica”, Tiempo de Consumo de Videojuegos”, Bienestar Físico-Mental”, “Impacto sobre actividades cotidianas”, “Relaciones sociales” y “Desarrollo Cognitivo”) de nuestra encuesta, las cuales se considerarán como variables predictoras del promedio ponderado de los estudiantes.

En la *Tabla 10* se puede observar que las 6 dimensiones tienen una fuerte influencia en la predicción del rendimiento académico de los estudiantes. Específicamente, se observa que mayores niveles de consumo de videojuegos, un impacto negativo en las actividades cotidianas y un bienestar físico-mental reducido están más asociados con la desaprobación. Por otro lado, mejores relaciones sociales y un mayor desarrollo cognitivo favorecen la probabilidad de aprobación.



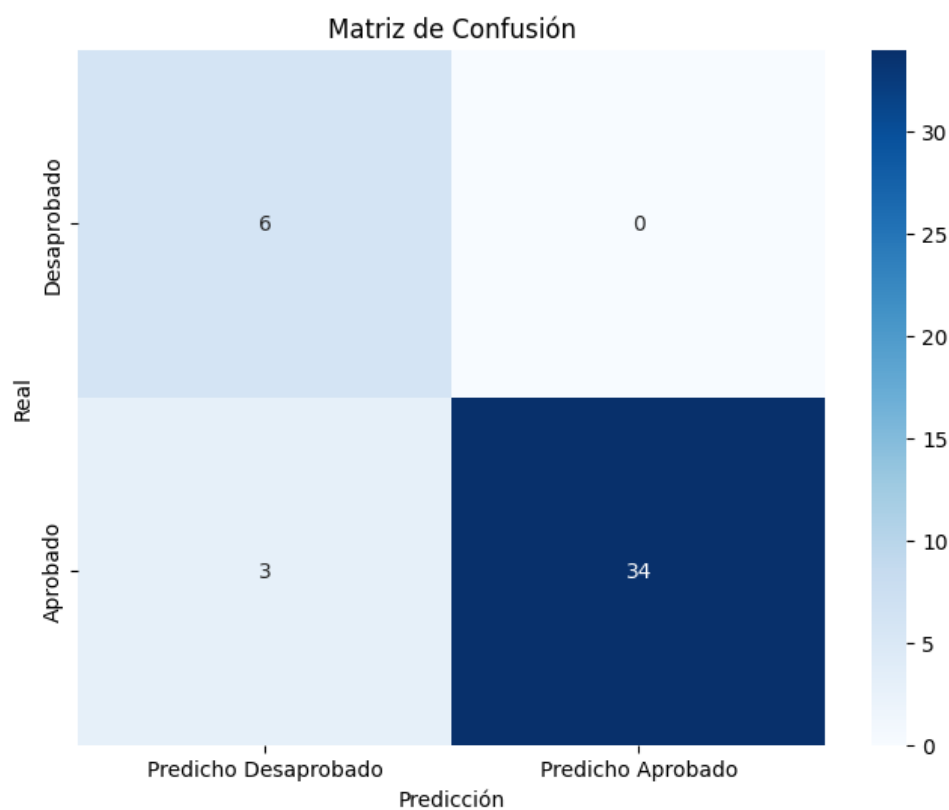
Tabla 10:

Evaluación del modelo de Naive Bayes

Evaluación del modelo de Naive Bayes		
Coefficientes por variable predictora	Desaprobado	Aprobado
Capacidad Económica	11.58823529	6.10218978
Tiempo de consumo de Videojuegos	13.79411765	6.55474453
Relaciones Sociales	12.91176471	7.18248175
Impacto sobre actividades cotidianas	13.11764706	6.5620438
Bienestar Físico-Mental	12.73529412	7.37226277
Desarrollo Cognitivo	13.29411765	6.48905109

Figura 16:

Matriz de confusión del modelo de Naive Bayes





Desglose:

- **Desaprobado (Real) y Predicho Desaprobado (6):**

El modelo predijo correctamente 6 estudiantes como desaprobados.

- **Desaprobado (Real) y Predicho Aprobado (0):**

No hay casos en los que el modelo haya predicho incorrectamente a un estudiante desaprobado como aprobado.

- **Aprobado (Real) y Predicho Desaprobado (3):**

El modelo predijo incorrectamente a 3 estudiantes aprobados como desaprobados.

- **Aprobado (Real) y Predicho Aprobado (34):**

El modelo predijo correctamente a 34 estudiantes como aprobados.

Para evaluar el rendimiento del modelo, se pueden utilizar varias métricas calculadas a partir de la matriz de confusión:

Métricas de Evaluación:

- **Precisión:**

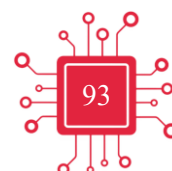
$$P = \frac{\text{Predicciones correctas}}{\text{Total de Predicciones}} = \frac{6 + 34}{6 + 0 + 3 + 34} = 0.930 \approx 93\%$$

Aproximadamente el 93% de las predicciones del modelo son correctas.

- **Precisión para Aprobado:**

$$P = \frac{TP}{TP + FP} = \frac{34}{34 + 3} = 0.919 \approx 91.9\%$$

Alrededor del 91.9% de las predicciones positivas del modelo son correctas.





- **Recall para Aprobado**

$$R = \frac{TP}{TP + FN} = \frac{34}{34 + 0} = 1 \approx 100\%$$

El modelo identifica correctamente el 100% de los estudiantes aprobados.

- **F1 Score para Aprobado**

$$F1 = 2 * \frac{P * R}{P + R} = 2 * \frac{0.919 * 1}{0.919 + 1} = 0.958 \approx 95.8\%$$

El balance entre precisión y recall es de aproximadamente 95.8%, indicando un buen rendimiento general del modelo en la clasificación de estudiantes como aprobados o desaprobados.

- La precisión del modelo es alta (aproximadamente 93%), lo que indica que el modelo clasifica correctamente la mayoría de los casos.
- Hay 3 casos en los que el modelo predijo incorrectamente a los estudiantes aprobados como desaprobados.
- No hay falsos negativos, lo que significa que el modelo no clasificó erróneamente a ningún estudiante desaprobado como aprobado.

En resumen, el modelo Naive Bayes muestra un buen rendimiento en la clasificación de estudiantes como aprobados o desaprobados, con una alta precisión y un F1-Score elevado. Sin embargo, se debe tener en cuenta la presencia de algunos falsos positivos que podrían afectar a los estudiantes aprobados.



4.4.3. Comparación de modelos y selección del óptimo

Para generar los modelos de predicción, utilizaremos los algoritmos descritos anteriormente y en donde el dataset se segmentó de manera aleatoria en un 80 % para los datos de entrenamiento y en 20% para los datos de prueba.

Tabla 11:

Comparación de modelos y selección del óptimo

	Modelo	Precisión
Modelos de Regresión	Lineal	67%
	Múltiple	67%
	Random Forest	87.5%
Modelos de Clasificación	Regresión Logística	93%
	Naive Bayes	93%

Comparando los modelos de regresión, el más eficiente es el Random Forest, con un 87.5% de precisión en su predicción. Además, como se detalló anteriormente, este modelo presenta un MSE de 0.48, un valor mucho menor que el modelo de regresión lineal (1.28) y el de regresión múltiple (1.29).

En el caso de los modelos de clasificación, podemos observar una igualdad en sus porcentajes de predicción. En este caso, se consideraron las características de los modelos en base a su eficiencia, siendo Naive Bayes más eficiente. Esto se debe a que Naive Bayes tiene una menor complejidad computacional y un menor uso de memoria en comparación con la regresión logística. Además, en términos de velocidad, Naive Bayes es significativamente mejor.

4.5. FASE V: Evaluación

En esta fase se evaluará el porcentaje de precisión demostrado anteriormente, haciendo una comparativa entre los valores reales y los valores predichos por los modelos.

4.5.1. Evaluación de resultados de Modelo de Regresión

La evaluación se realizó con el código Python como se muestra en la *Figura 17*, tomando los 42 valores del dataset de prueba autogenerados del 20% del total de los datos.

Figura 17:

Código Python para generar margen de error del modelo Random Forest

```
1 # Calcular el margen de error
2 error_margin = abs(y_test - y_pred)
3
4 # Crear la tabla de resultados
5 results_df = pd.DataFrame({
6     'Valor Real': y_test,
7     'Valor Predicho': y_pred,
8     'Margen de Error': error_margin
9 })
10
11 # Mostrar la tabla de resultados
12 results_df.reset_index(drop=True, inplace=True)
13 results_df
```

En la *Tabla 12* se observa el margen de error de los 42 registros de prueba del modelo Random Forest, en ellos se puede evidenciar que los valores reales no se encuentran tan distantes de los valores predichos, corroborando el alto porcentaje de predicción (87.5%) determinado en la fase de modelado.



Tabla 12:

Margen de error de dataset de prueba del modelo Random Forest

Nº	Valor Real	Valor Predicho	Margen de Error	Nº	Valor Real	Valor Predicho	Margen de Error
0	5.80	8.69	2.89	22	15.18	14.65	0.53
1	14.36	14.24	0.12	23	14.33	14.28	0.05
2	11.95	11.53	0.42	24	13.58	13.13	0.45
3	14.32	14.43	0.11	25	12.86	11.61	1.25
4	13.05	13.21	0.16	26	10.50	10.77	0.27
5	12.95	12.38	0.57	27	10.14	10.65	0.51
6	8.13	7.10	1.03	28	14.14	14.29	0.16
7	12.45	12.36	0.10	29	8.14	9.93	1.79
8	12.18	11.84	0.34	30	11.64	11.51	0.12
9	13.41	12.34	1.07	31	13.33	13.41	0.08
10	14.18	14.26	0.08	32	13.82	14.26	0.44
11	14.00	14.23	0.23	33	12.64	12.67	0.03
12	13.41	13.37	0.04	34	12.00	12.27	0.27
13	12.14	12.60	0.47	35	14.50	14.29	0.21
14	14.14	13.92	0.22	36	14.05	14.26	0.21
15	14.00	14.10	0.10	37	13.67	12.17	1.50
16	14.32	14.34	0.02	38	12.72	12.77	0.05
17	8.89	8.71	0.18	39	12.36	12.13	0.23
18	11.32	11.89	0.57	40	12.82	12.65	0.17
19	14.23	14.26	0.03	41	14.50	14.55	0.05
20	12.73	12.66	0.07	42	11.82	11.72	0.10
21	15.00	14.19	0.81	-	-	-	-

4.5.2. Evaluación de resultados de Modelo de Clasificación

La evaluación se realizó con el código Python como se muestra en la *Figura 18*, tomando los 42 valores del dataset de prueba autogenerados del 20% del total de los datos.

Figura 18:

Código Python para generar margen de error del modelo Naive Bayes

```
1 # Calcular el margen de error
2 error_margin = abs(y_test - y_pred)
3
4 # Crear la tabla de resultados
5 results_df = pd.DataFrame({
6     'Valor Real': y_test,
7     'Valor Predicho': y_pred,
8     'Margen de Error': error_margin
9 })
10
11 # Mostrar la tabla de resultados
12 results_df.reset_index(drop=True, inplace=True)
13 results_df
```

En la *Tabla 13* se observa el margen de error de los 42 registros de prueba del modelo Naive Bayes, solo 3 de ellos presentan una predicción errónea, cumpliendo así el porcentaje de predicción del 93%.



Tabla 13:

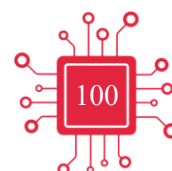
Margen de error de dataset de prueba del modelo Naive Bayes

Nº	Valor Real	Valor Predicho	Margen de Error	Nº	Valor Real	Valor Predicho	Margen de Error
0	0	0	0	22	1	1	0
1	1	1	0	23	1	1	0
2	1	0	1	24	1	1	0
3	1	1	0	25	1	1	0
4	1	1	0	26	0	0	0
5	1	1	0	27	0	0	0
6	0	0	0	28	1	1	0
7	1	1	0	29	0	0	0
8	1	1	0	30	1	0	1
9	1	1	0	31	1	1	0
10	1	1	0	32	1	1	0
11	1	1	0	33	1	1	0
12	1	1	0	34	1	1	0
13	1	1	0	35	1	1	0
14	1	1	0	36	1	1	0
15	1	1	0	37	1	1	0
16	1	1	0	38	1	1	0
17	0	0	0	39	1	1	0
18	1	0	1	40	1	1	0
19	1	1	0	41	1	1	0
20	1	1	0	42	1	1	0
21	1	1	0	-	-	-	-



4.6. FASE VI: Despliegue

En el presente documento se detalla de manera ordenada la metodología CRISP-DM, la cual está orientada para el desarrollo de proyectos de Inteligencia Artificial. Así mismo los notebooks del desarrollo de los 5 modelos de predicción generados para esta investigación están detallados en los anexos del presente informe.





CAPÍTULO V

DISCUSIÓN

- En la realidad del contexto estudiado en este trabajo de investigación, se menciona que, según patrones similares a través de vivencias anecdóticas y conversaciones con compañeros, se podría predecir el rendimiento académico de los estudiantes de la EPISI en función a su consumo de videojuegos. Tras culminar con la investigación, se demostró que el rendimiento académico puede ser predicho significativamente (con más del 80% de precisión) basado en el consumo de videojuegos. Estos resultados confirman que los estudiantes con un mayor consumo de videojuegos tienden a tener un rendimiento académico inferior. Aunque los videojuegos pueden desarrollar habilidades cognitivas, su uso excesivo tiene efectos negativos. Esto subraya la importancia de intervenciones y programas de orientación que ayuden a los estudiantes a equilibrar adecuadamente el tiempo dedicado a los videojuegos y al estudio, promoviendo un uso saludable para optimizar el rendimiento académico.
- Fabito S., et al. (2018), investigó si la adicción a los videojuegos móviles y el ciberacoso afectan negativamente el rendimiento académico. Concluyendo que ambos no son factores causales del rendimiento académico de los estudiantes. A diferencia de nuestro trabajo de investigación, en donde se encontró que el rendimiento académico de los estudiantes universitarios puede verse afectado de manera positiva o



negativa dependiendo de su consumo de videojuegos, permitiendo así mismo su predicción.

- Xu X., et al. (2019), en su investigación logró realizar predicciones eficaces con un dataset de 4000 estudiantes de la Universidad de Beijing, además, Caselli (2021), también realizó predicciones eficaces con un dataset que incluía información de estudiantes de 4 escuelas profesionales de la Universidad Nacional del Santa entre los años 2004-2018. Gracias a esta cantidad de datos, pudo desarrollar modelos más robustos, como el de redes neuronales. Gracias a estas investigaciones, tenemos como noción el poder mejorar la precisión de nuestros algoritmos y desarrollar nuevos con técnicas más avanzadas ampliando el tamaño de nuestro dataset, ya que la muestra recopilada en comparación es más pequeña.
- Candia (2019) y Vega (2019) lograron realizar predicciones aceptables del rendimiento académico, al igual que nuestra investigación. Estas tres investigaciones contribuyen positivamente al campo de la predicción del rendimiento académico a partir de enfoques distintos, los cuales, al combinarse, podrían ofrecer un enfoque más completo y efectivo para predecir y mejorar el rendimiento académico de los estudiantes.



5.1. Contrastación de la hipótesis general

H1: *El consumo de videojuegos tiene una influencia significativa en el rendimiento académico de los estudiantes universitarios, y utilizando técnicas de Machine Learning, es posible predecir dicha influencia con una precisión de más del 80%.*

A pesar de las limitaciones que tuvimos al momento de recabar los datos para nuestro proyecto de investigación, los resultados obtenidos apoyan la hipótesis alternativa validando que el consumo de videojuegos tiene una influencia significativa en el rendimiento académico de los estudiantes universitarios. Los modelos seleccionados como óptimos proporcionan una herramienta potencialmente útil para educadores y profesionales de las áreas correspondientes con porcentajes por encima del 80% (Random Forest: 87.5% y Naive Bayes: 93%). Se pueden mejorar las predicciones de los modelos propuestos en esta investigación, incorporando variables adicionales, expandiendo el dataset o en todo caso explorando otros métodos de aprendizaje automático que se adapten al contexto de la evaluación.

5.2. Contrastación de las hipótesis específicas

H1a: *Existe una correlación significativa entre el tiempo y la frecuencia del uso de videojuegos por parte de los estudiantes universitarios, para predecir su rendimiento académico.*

Esta hipótesis se apoya en la prueba de correlación de Pearson, que mostró una relación estadísticamente significativa (-0,76), indicando que a medida que aumenta el tiempo de consumo de videojuegos, el rendimiento académico tiende a disminuir de manera predecible. Además, la dimensión del tiempo de consumo de



videojuegos demostró ser una variable importante en los modelos, contribuyendo significativamente a su capacidad de predicción, mejorando la precisión y la eficacia de las predicciones realizadas.

H1b: *La capacidad económica de los estudiantes universitarios afecta al consumo de videojuegos*

Esta hipótesis no se confirmó plenamente, ya que los resultados de la encuesta mostraron que, aunque la capacidad económica tuvo alguna influencia, no fue un impedimento significativo para el consumo de videojuegos; los datos indicaron que tanto estudiantes con mayor capacidad económica como aquellos con recursos limitados dedicaron tiempo significativo a los videojuegos, sugiriendo que otros factores pueden ser más determinantes.

H1c: *Se elige el algoritmo de Machine Learning más eficiente para predecir el Rendimiento Académico de los estudiantes de la EPISI en base al consumo de videojuegos.*

Esta hipótesis fue válida en el sentido de que se pudo seleccionar un algoritmo de Machine Learning para hacer predicciones sobre el rendimiento académico basado en el consumo de videojuegos, su validez podría haber sido aún mayor con más datos. Esto habría permitido no solo mejorar la precisión de las predicciones, sino también validar de manera robusta la relación entre el consumo de videojuegos y el rendimiento académico.



H1d: *Se logra identificar los factores de entrada clave del modelo predictivo para predecir el rendimiento académico de los estudiantes universitarios en base al consumo de videojuegos.*

Esta hipótesis es validada en este trabajo de investigación, esto se debe a que se han calculado tanto el porcentaje de correlación de cada factor de entrada (como se muestra en la *Figura 11 “Mapa de Calor entre las 6 dimensiones y el promedio ponderado de los estudiantes”*) como la importancia y el coeficiente en los modelos de predicción seleccionados, como Random Forest y Naive Bayes (Como se muestran en la *Tabla 64* y la *Tabla 66*). Estas métricas permiten determinar cuáles variables tienen un mayor impacto en el rendimiento académico, proporcionando una base sólida para construir un modelo predictivo eficaz y preciso.

H1e: *El uso de la metodología CRISP-DM facilita el desarrollo del modelo predictivo.*

Según lo observado en el presente informe, la metodología CRISP-DM nos brinda una estructura ordenada y enfocada en el contexto de la evaluación, se centra en la comprensión de los objetivos y los requisitos del negocio y nos otorga una fase dedicada al desarrollo de los modelos, la cual nos ha permitido seleccionar y aplicar las mejores técnicas, y una fase de evaluación exhaustiva para garantizar que el modelo cumple con los objetivos antes de su implementación. Esto permite que el modelo seleccionado esté alineado y sea eficaz.



CAPÍTULO VI

CONCLUSIONES Y RECOMENDACIONES

6.1. CONCLUSIONES

Después de realizar un estudio exhaustivo utilizando técnicas de Machine Learning, se demostró que el rendimiento académico de los estudiantes universitarios puede ser predicho en función de su consumo de videojuegos. Los modelos predictivos propuestos, basado en las 6 dimensiones (“Capacidad Económica”, Tiempo de Consumo de Videojuegos”, Bienestar Físico-Mental”, “Impacto sobre actividades cotidianas”, “Relaciones sociales” y “Desarrollo Cognitivo”) seleccionadas como claves de entrada, mostró una precisión significativa en los modelos óptimos (Random Forest y Naive Bayes) cumpliendo así con el objetivo general de la investigación.

Al analizar la relación entre el tiempo de consumo de videojuegos y el rendimiento académico de los estudiantes universitarios revela la posibilidad de que el tiempo dedicado a esta actividad tenga un impacto significativo en cómo los estudiantes se desempeñan en sus estudios. Este análisis proporciona una base para explorar cómo gestionar mejor el equilibrio entre las actividades de ocio y las responsabilidades académicas, ayudando a informar políticas educativas y estrategias de apoyo personalizadas.

Al evaluar si el consumo de videojuegos se ve afectado por la capacidad económica de los estudiantes universitarios indica que, aunque el poder adquisitivo puede influir parcialmente en las decisiones de consumo hacia los videojuegos, es



evidente que los estudiantes de diferentes niveles económicos continúan participando en esta actividad. Además, los estudiantes con mejor situación económica tienden a invertir más en productos y servicios relacionados con los videojuegos, lo que destaca la diversidad de comportamientos de consumo dentro de la comunidad estudiantil universitaria.

Al elegir el algoritmo de Machine Learning más eficiente para predecir el rendimiento académico de los estudiantes de la EPISI basado en el consumo de videojuegos se fundamenta en el éxito alcanzado mediante la metodología CRISP-DM. Siendo así, que el modelo de Random Forest demostró un porcentaje de predicción del 87.5%, mientras que el modelo de Naive Bayes alcanzó un 93%. Estos resultados destacan la eficacia de ambos algoritmos para abordar el problema desde diferentes enfoques, proporcionando herramientas robustas y precisas para predecir el rendimiento académico y entender la influencia del consumo de videojuegos en el desempeño de los estudiantes. Además, la mejora continua de la metodología de recolección de datos y la expansión del conjunto de datos podrían potenciar aún más estos modelos, permitiendo la exploración de técnicas más avanzadas de predicción en el futuro.

En este estudio, se lograron identificar los factores de entrada clave para el modelo predictivo que predecirá el rendimiento académico de los estudiantes universitarios en base al consumo de videojuegos. Utilizando técnicas de Machine Learning, se evaluaron las 6 dimensiones relacionadas con el consumo de videojuegos y su impacto en el rendimiento académico. Mediante el cálculo del porcentaje de correlación, la importancia de las variables y los coeficientes en los modelos



seleccionados, se determinaron los factores más influyentes. Estos análisis permitieron destacar los aspectos críticos que deben considerarse para realizar predicciones precisas y efectivas del rendimiento académico y su nivel de impacto de cada una de ellas

Utilizando la metodología CRISP-DM, se logró demostrar con éxito la capacidad de predecir el rendimiento académico de los estudiantes universitarios. La metodología CRISP-DM, con su enfoque estructurado y sistemático, facilitó la comprensión de los datos, la preparación adecuada, la modelización precisa y la evaluación efectiva del modelo. Esta metodología permitió identificar patrones significativos y relaciones clave entre las dimensiones, seleccionando así los modelos óptimos, asegurando la calidad de este.

6.2. RECOMENDACIONES

Emplear métricas estadísticas que permitan clasificar la relación entre las variables predictoras y determinar las más relevantes las cuales tengan porcentajes significativos en relación con la variable a predecir.

Trabajar con un dataset mucho más amplio, por ejemplo, escalar el análisis a los estudiantes de toda la facultad de ingeniería, esto permitirá que el grupo de entrenamiento y de prueba sea mucho más grande y aumente la confiabilidad y robustez de la predicción de los modelos.



Asegurarse de que se sigan las mejores prácticas en la gestión de datos, incluyendo la calidad, seguridad y privacidad de los datos recopilados. Esto es crucial para mantener la integridad y confidencialidad de la información, así como para cumplir con las regulaciones de protección de datos.

Promover la colaboración entre diferentes dependencias dentro de la universidad para enriquecer el análisis de datos. Esto puede generar perspectivas más amplias y profundas sobre diversos problemas y facilitar la implementación de soluciones innovadoras basadas en datos.

La propuesta de este trabajo de investigación puede ser transformada para recibir variables de entrada referentes a otros contextos que pueden perjudicar el rendimiento académico de los estudiantes universitarios, validando así, su versatilidad e importancia académica y práctica.



CAPÍTULO VII

REFERENCIAS BIBLIOGRÁFICAS

- Araúz M., I. A. (2020). Los videojuegos y su relación con el rendimiento académico en Jóvenes Universitarios. Revista Espectro Investigativo Latinoamericano, 2(2), 9-12. <https://revista.isaeuniversidad.ac.pa/index.php/EIL/article/view/50>
- Arimetrics. (s.f.). Qué es Entorno de desarrollo. Glosario Digital Arimetrics: <https://www.arimetrics.com/glosario-digital/entorno-de-desarrollo>
- AWS. (s.f.). ¿Qué es la ciencia de datos? What is - AWS: <https://aws.amazon.com/es/what-is/data-science/>
- AWS. (s.f.). ¿Qué es Python? What is - AWS: <https://aws.amazon.com/es/what-is/python/>
- Batta, M. (2020). Machine Learning Algorithms - A Review. International Journal of Science and Research (IJSR), 9. <https://doi.org/10.21275/ART20203995>
- Candia, D. (2019). Predicción del rendimiento académico de los estudiantes de la UNSAAC a partir de sus datos de ingreso utilizando algoritmos de aprendizaje automático. Repositorio Institucional - UNSAAC: <https://repositorio.unsaac.edu.pe/handle/20.500.12918/4120>
- Caselli, H. (2021). Modelo predictivo basado en Machine Learning como soporte para el seguimiento académico del estudiante universitario. Repositorio Institucional Universidad Nacional del Santa: <https://repositorio.uns.edu.pe/handle/20.500.14278/3804>
- Chen, G. Y. (2024). Impact of Internet Usage and Video Gaming on the Academic Performance of Postgraduate Students in a Prestigious Taiwanese University. Research Square (Research Square). <https://doi.org/10.21203/rs.3.rs-3849566/v1>



Creswell, J. (2014). Research design. Qualitative, quantitative and mixed methods approaches. SAGE.

edX. (s.f.). Cursos de Modelo Predictivo. Aprende:
<https://www.edx.org/es/aprende/modelo-predictivo>

Empresa Peruana de Servicios Editoriales S. A. EDITORA PERÚ. (2021, 6 septiembre). Consumo de videojuegos en el Perú creció durante la pandemia.
<https://andina.pe/agencia/noticia-consumo-videojuegos-el-peru-crecio-durante-pandemia-859350.aspx>

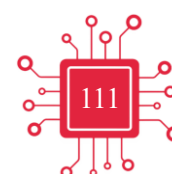
Espinosa-Zuñiga, J. (2020). Aplicación de metodología CRISP-DM para segmentación geográfica de una base de datos públicos. Ingeniería, Investigación y Tecnología, 21(1), 13. <https://doi.org/10.22201/fi.25940732e.2020.21n1.008>

Espinoza, G. L. (2020). Modelo de Machine Learning para la clasificación de estudiantes de acuerdo con su rendimiento académico en el Centro de Idiomas de la Universidad Nacional del Santa. Repositorio Institucional Universidad Nacional del Santa: <https://repositorio.uns.edu.pe/handle/20.500.14278/3588>

Fabito, B., Rodriguez, R., Diloy, M., Trillanes, A., Macato, L., & Octaviano, M. (2018). Exploring Mobile Game Addiction, Cyberbullying, and its Effects on Academic Performance among Tertiary Students in one University in the Philippines. IEEE Region 10 International Conference TENCON, 1859 - 1864.
<https://doi.org/10.1109/TENCON.2018.8650251>

Hernández, R., Fernández, C., & Baptista, M. (2014). Metodología de la Investigación. Mc Graw Hill.

Julca Meza, E. (2016). CONCEPTOS BÁSICOS DE LA EDUCACIÓN. Revista Cultura de la Asociación de Docentes de la USMP.





Lemmens, J., Valkenburg, P., & Peter, J. (2009). Development and Validation of a
Game Addiction. Media Psychology.

<https://doi.org/10.1080/15213260802669458>

Ludeña, C. (2022). Adicción a los videojuegos y calidad del sueño en estudiantes de
psicología de una universidad privada de Chiclayo, 2022.

<https://repositorio.uss.edu.pe/handle/20.500.12802/10562>

Márquez Díaz, J. (2020). Inteligencia artificial y Big Data como soluciones frente a
la COVID-19. Revista de Bioética y Derecho.

https://doi.org/https://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1886-58872020000300019

Minsa: La pandemia COVID-19 ha incrementado la adicción a los videojuegos en
niños y adolescentes. (s. f.). Noticias - Ministerio de Salud - Plataforma del Estado
Peruano. <https://www.gob.pe/institucion/minsa/noticias/348005-minsa-la-pandemiacovid-19-ha-incrementado-la-adiccion-a-los-videojuegos-en-ninos-y-adolescentes>

Rivera Arteaga, E., & Torres Cosío, V. (2018). Videojuegos y habilidades del
pensamiento. Revista Iberoamericana para la Investigación y el Desarrollo
Educativo, 8(16). <https://doi.org/10.23913/ride.v8i16.341>

Tacilla Cardenas, I., Vásquez Villanueva, S., Verde Avalos, E., & Colque Díaz, E.
(2020). Rendimiento académico: universo muy complejo para el quehacer
pedagógico. Revista Muro de la Investigación - Universidad Peruana Unión, 5(2).
<https://doi.org/10.17162/rmi.v5i2.1325>

Universidad Europea. (22 de Marzo de 2022). ¿Para qué sirve un gestor de base de
datos? Blog - Universidad Europea: <https://universidadeuropea.com/blog/para-que-sirve-gestor-base-datos/>



Vega, J. (2019). Modelo de pronóstico de rendimiento académico de alumnos en los cursos del programa de estudios básicos de la Universidad Ricardo Palma usando algoritmos de Machine Learning. Repositorio Institucional Universidad Ricardo Palma: <https://repositorio.urp.edu.pe/handle/20.500.14138/2914>

Xu, X., Wang, J., Peng, H., & Wu, R. (2019). Prediction of academic performance associated with internet usage behaviors using machine learning algorithms. *Computers in Human Behavior*, 98, 166-173.
<https://doi.org/10.1016/j.chb.2019.04.015>



CAPÍTULO VIII

ANEXOS

Anexo 1: Cuestionario

CUESTIONARIO SOBRE EL CONSUMO DE VIDEOJUEGOS

Esta encuesta nos ayudará a conocer los hábitos en relación con los videojuegos en cualquier tipo de plataforma (celular, PC, consola, etc) o si consume un juego en línea o fuera de línea. Por favor, responda a las siguientes preguntas colocando un aspa en la casilla correspondiente pensando en los últimos 6 meses.

P.D.: Valoramos mucho tu sinceridad. No hay respuestas mejores ni peores. Te pedimos que contestes con sinceridad para así lograr contribuir con propuestas adecuadas. Ten la seguridad que este cuestionario es muy confidencial.

PREGUNTAS			RESPUESTAS				
N°	Dimensiones	Items	MA	BA	A	PA	NA
1	Datos demográficos	¿Cuál es tu edad?					
2		¿Cuál es tu sexo?					
3		¿Cuál es tu código de matrícula?					
4		¿Actualmente trabajas?	SI			NO	
5		Si en la pregunta anterior tu respuesta fue "SI": ¿Cuál es tu jornada laboral?	Tiempo Parcial			Tiempo Completo	
	Dimensiones	Items	RESPUESTAS				
			0	1	2	3	4
6	Capacidad Económica	Según tú, ¿a cuánto equivale tu cartera de videojuegos?	S/. 0.00 - S/. 250.00	S/. 250.00 - S/. 500.00	S/. 500.00 - S/. 750.00	S/. 750.00 - S/. 1000.00	S/. 1000.00 a más
7		¿Cuánto dinero destinas mensualmente a la compra de videojuegos y contenido relacionado?	S/. 0.00 - S/. 100.00	S/. 100.00 - S/. 200.00	S/. 200.00 - S/. 300.00	S/. 300.00 - S/. 400.00	S/. 500.00 a más
8		¿Has invertido en la compra de consolas, computadoras o hardware específico para mejorar tu experiencia de juego?	Nunca	Raramente	Ocasionalmente	Frecuentemente	Siempre
9		¿Tu situación económica influye en tus preferencias por juegos gratuitos, de bajo costo o suscripciones?	Totalmente en desacuerdo	En desacuerdo	Neutral	De acuerdo	Totalmente de acuerdo
10	Tiempo de consumo de videojuegos	¿Cuántos días a la semana juegas videojuegos?	Entre 1 a 3 días	Entre 3 a 4 días	Entre 4 a 5 días	Entre 5 a 6 días	Todos los días
11		¿Cuántos horas al día dedicas a jugar videojuegos?	Menos de 1 hora	De 1 a 2 horas	De 2 a 4 horas	De 4 a 6 horas	Más de 6 horas
12		¿Tienes preferencias específicas para jugar videojuegos en ciertos momentos del día?	Mañana (05 am – 10 am)	Mediodía (10 pm – 03 pm)	Tarde (03 pm – 08 pm)	Noche (08 pm- 12 pm)	Madrugada (12 am – 05 am)
13		¿Con que frecuencia pierdes la noción del tiempo mientras juegas?	Nunca	Raramente	Ocasionalmente	Frecuentemente	Siempre
14	Bienestar Físico - Mental	¿Cómo calificarías tu bienestar general en relación con tu consumo de videojuegos?	Muy positivo	Positivo	Neutral	Negativo	Muy negativo
15		¿El tiempo que dedicas a los videojuegos te ha hecho perder horas de sueño?	Nunca	Raramente	Ocasionalmente	Frecuentemente	Siempre
16		¿Experimentas estrés o frustración al momento de jugar videojuegos?	Nunca	Raramente	Ocasionalmente	Frecuentemente	Siempre
17		¿Con qué frecuencia te sientes culpable o arrepentido por la cantidad de tiempo que pasas jugando videojuegos?	Nunca	Raramente	Ocasionalmente	Frecuentemente	Siempre



18	Impacto sobre actividades cotidianas	¿Qué nivel de impacto consideras que ha tenido el consumo de videojuegos en tus actividades cotidianas?	Insignificante mente	Levemente	No sabe / No respondes	Moderadamente	Significativamente
19		¿Con qué frecuencia tienes dificultades para realizar actividades (académicas, físicas, sociales, etc.) por usar videojuegos?	Nunca	Raramente	Ocasionalmente	Frecuentemente	Siempre
20		¿Con qué frecuencia consideras que consumir videojuegos afecta tu rendimiento académico?	Nunca	Raramente	Ocasionalmente	Frecuentemente	Siempre
21		¿Has notado algún cambio en tu capacidad para estudiar y completar tareas académicas debido a tu consumo de videojuegos?	Ningún cambio	Cambios minimos	Cambios moderados	Cambios significativos	Cambios muy significativos
22	Relaciones Sociales	¿Has experimentado cambios en tu capacidad para comunicarte o interactuar socialmente como resultado del consumo de videojuegos?	Ningún cambio	Cambios minimos	Cambios moderados	Cambios significativos	Cambios muy significativos
23		¿Con que frecuencia has tenido peleas con otras personas (por ejemplo, familia, amigos) por el tiempo que dedicas a los juegos?	Nunca	Raramente	Ocasionalmente	Frecuentemente	Siempre
24		¿En qué nivel consideras que los videojuegos han afectado tus habilidades para colaborar y trabajar en equipo?	Ningún impacto	Impacto minimo	Impacto moderado	Impacto significativo	Impacto muy significativo
25		¿Con que frecuencia has descuidado a personas de tu entorno (p. ej., familia, amigos) por estar jugando?	Nunca	Raramente	Ocasionalmente	Frecuentemente	Siempre
26	Desarrollo Cognitivo	¿Crees que los videojuegos han contribuido al desarrollo de tus habilidades cognitivas, como la resolución de problemas y la toma de decisiones?	Totalmente en desacuerdo	En desacuerdo	Neutral	De acuerdo	Totalmente de acuerdo
27		¿Crees que jugar videojuegos te ha ayudado a adaptarte mejor a entornos virtuales y tecnológicos?	Totalmente en desacuerdo	En desacuerdo	Neutral	De acuerdo	Totalmente de acuerdo
28		¿Crees que jugar videojuegos ha mejorado tu capacidad para realizar tareas múltiples eficientemente?	Totalmente en desacuerdo	En desacuerdo	Neutral	De acuerdo	Totalmente de acuerdo
29		¿Has proyectado las habilidades aprendidas al jugar videojuegos a situaciones de la vida real?	Nunca	Raramente	Ocasionalmente	Frecuentemente	Siempre
Total:							



Anexo 2: Validación de Instrumento por Licenciada en Psicología



UNIVERSIDAD NACIONAL DEL SANTA
Escuela Profesional de Ingeniería de Sistemas e Informática

PRESENTACIÓN A JUICIO DE EXPERTO

Estimado Validador: Lc. Alvarez Michola Blanca

Me es grato dirigirme a usted, a fin de solicitar su colaboración como experto para validar el instrumento que adjunto denominado:

Cuestionario Sobre el consumo de Videjuegos, diseñado por los bachilleres Diego Alberto Castañeda Chávez y Valeria Isabel Perez Chang, cuyo propósito es medir la influencia de los videjuegos, el cual será aplicado a estudiantes de la E.P. Ing. de Sistemas e Inform. por cuanto considero que sus observaciones, apreciaciones y acertados aportes serán de utilidad.

El presente instrumento tiene como finalidad recoger información directa para la investigación que se realiza actualmente, titulada:

"Modelo predictivo del rendimiento académico de estudiantes universitarios asociado con la influencia de los videjuegos usando Machine Learning"

Tesis que será presentada a la Universidad Nacional del Santa, como requisito para obtener el Título profesional de:

Ingeniería de Sistemas e Informática

Para efectuar la validación del instrumento, usted deberá leer cuidadosamente cada enunciado y sus correspondientes alternativas de respuesta, en donde se pueden seleccionar una, varias o ninguna alternativa de acuerdo al criterio personal y profesional del actor que responda al instrumento. Se le agradece cualquier sugerencia referente a redacción, contenido, pertinencia y congruencia u otro aspecto que se considere relevante para mejorar el mismo.

Gracias por su aporte


Bach. Castañeda Chávez Diego


Bach. Perez Chang Valeria

Modelo Predictivo del Rendimiento Académico de Estudiantes Universitarios Asociado con la Influencia de los Videjuegos usando Machine Learning.



JUICIO DE EXPERTO SOBRE LA PERTINENCIA DEL INSTRUMENTO

Instrucciones: Marque con una X en donde corresponde, que, según su criterio, Si cumple o No cumple, la coherencia entre dimensiones e indicadores de la variable en estudio.

Variable	Dimensiones	Indicadores	N° de ítem	COHERENCIA	
				SI	NO
Influencia de los videojuegos (V.I.)	Capacidad económica	- Ingresos económicos - Gastos mensuales - Inversión en hardware - Relación entre situación económica y videojuegos	6, 7, 8, 9	X	
	Tiempo de consumo	- Días de juego a la semana - Horas diarias de juego - Preferencias de tiempo - Noción del tiempo	10, 11, 12, 13	X	
	Bienestar físico/mental	- Autoevaluación del bienestar general - Pérdida de horas de sueño - Niveles de estrés - Resiliencia	14, 15, 16, 17	X	
	Impacto sobre actividades cotidianas	- Autoevaluación del impacto del consumo de videojuegos - Dificultad para realizar actividades - Percepción en el rendimiento académico - Impacto en la productividad académica	18, 19, 20, 21	X	
	Relaciones sociales	- Cambios en la capacidad de comunicación e interacción social - Conflictos interpersonales - Autoevaluación sobre las habilidades de colaboración y trabajo en equipo - Descuido hacia personas del entorno	22, 23, 24, 25	X	
	Desarrollo Cognitivo	- Desafíos cognitivos en los videojuegos - Adaptación a entornos virtuales - Desarrollo de habilidades multitarea - Transferencia de habilidades a la vida real	26, 27, 28, 29	X	

Modelo Predictivo del Rendimiento Académico de Estudiantes Universitarios Asociado con la Influencia de los Videojuegos usando Machine Learning.



Instrucciones de Evaluación de ítems: Coloque en cada casilla de valoración la letra o letras correspondiente al aspecto cualitativo que, según su criterio, cumple o tributa cada ítem a medir los aspectos o dimensiones de la variable en estudio. Las valoraciones son las siguientes:

MA= Muy adecuado / BA= Bastante adecuado / A = Adecuado / PA= Poco adecuado / NA= No adecuado

Categorías a evaluar: Redacción, contenido, congruencia y coherencia en relación a la variable de estudio. En la casilla de observaciones puede sugerir mejoras.

PREGUNTAS			VALORACIÓN					Observaciones
Nº	Dimensiones	Ítems	MA	BA	A	PA	NA	
1	Datos demográficos	¿Cuál es tu edad?	X					
2		¿Cuál es tu sexo?		X				
3		¿Cuál es tu código de matrícula?	X					
4		¿Actualmente trabajas?		X				
5		Indica tu tipo de jornada laboral	X					
6	Capacidad Económica	Según tú, ¿a cuánto equivale tu cartera de videojuegos?		X				
7		¿Cuánto dinero destinas mensualmente a la compra de videojuegos y contenido relacionado?	X					
8		¿Has invertido en la compra de consolas, computadoras o hardware específico para mejorar tu experiencia de juego?	X					
9		¿Tu situación económica influye en tus preferencias por juegos gratuitos, de bajo costo o suscripciones?		X				
10	Tiempo de consumo de videojuegos	¿Cuántos días a la semana juegas videojuegos?		X				
11		¿Cuántos horas al día dedicas a jugar videojuegos?	X					
12		¿Tienes preferencias específicas para jugar videojuegos en ciertos momentos del día?	X					
13		¿Con qué frecuencia pierdes la noción del tiempo mientras juegas?		X				
14	Bienestar Físico - Mental	¿Cómo calificarías tu bienestar general en relación con tu consumo de videojuegos?		X				
15		¿El tiempo que dedicas a los videojuegos te ha hecho perder horas de sueño?	X					
16		¿Experimentas estrés o frustración al momento de jugar videojuegos?		X				
17		¿Con qué frecuencia te sientes culpable o arrepentido por la cantidad de tiempo que pasas jugando videojuegos?	X					
18	Impacto sobre actividades cotidianas	¿Qué nivel de impacto consideras que ha tenido el consumo de videojuegos en tus actividades cotidianas?	X					
19		¿Con qué frecuencia tienes dificultades para realizar actividades (académicas, físicas, sociales, etc.) por usar videojuegos?		X				
20		¿Con qué frecuencia consideras que consumir videojuegos afecta tu rendimiento académico?	X					
21		¿Has notado algún cambio en tu capacidad para estudiar y completar tareas académicas debido a tu consumo de videojuegos?	X					

Modelo Predictivo del Rendimiento Académico de Estudiantes Universitarios Asociado con la Influencia de los Videojuegos usando Machine Learning.



UNIVERSIDAD NACIONAL DEL SANTA
Escuela Profesional de Ingeniería de Sistemas e Informática

22	Relaciones Sociales	¿Has experimentado cambios en tu capacidad para comunicarte o interactuar socialmente como resultado del consumo de videojuegos?		X			
23		¿Con que frecuencia has tenido peleas con otras personas (por ejemplo, familia, amigos) por el tiempo que dedicas a los juegos?		X			
24		¿En qué nivel consideras que los videojuegos han afectado tus habilidades para colaborar y trabajar en equipo?		X			
25		¿Con que frecuencia has descuidado a personas de tu entorno (p. ej., familia, amigos) por estar jugando?		X			
26	Desarrollo Cognitivo	¿Crees que los videojuegos han contribuido al desarrollo de tus habilidades cognitivas, como la resolución de problemas y la toma de decisiones?	X				
27		¿Crees que jugar videojuegos te ha ayudado a adaptarte mejor a entornos virtuales y tecnológicos?	X				
28		¿Crees que jugar videojuegos ha mejorado tu capacidad para realizar tareas múltiples eficientemente?	X				
29		¿Has proyectado las habilidades aprendidas al jugar videojuegos a situaciones de la vida real?	X				
Total:							

Evaluado por: (Apellidos y Nombres) Alvarez Minchola Blanca

D.N.I.: 08968535 Fecha: 10/01/24

Firma: [Firma]

GOBIERNO REGIONAL DE ANCASH
HOSPITAL REGIONAL
"ELEAZAR GUZMAN BARRON"
Psic. BLANCA ALVAREZ MINCHOLA
C.Ps.P. 10004
PSICOLOGIA

Modelo Predictivo del Rendimiento Académico de Estudiantes Universitarios
Asociado con la Influencia de los Videojuegos usando Machine Learning.



CONSTANCIA DE VALIDACIÓN

Yo, Blanca Alvarez Minchola, con Documento Nacional de Identidad N°
08268535, de profesión psicología, grado académico licenciada, con
código de colegiatura 10004, labor que ejerzo actualmente como
jefatura del servicio, en la Institución
Hospital Regional "Eleazar Guzman Barrón"

Por medio de la presente hago constar que he revisado con fines de Validación el Instrumento
denominado cuestionario Influencia Videjuegos cuyo propósito es medir
el consumo de videojuegos, a los efectos de su aplicación a
estudiantes de E.P. Ingeniería de Sistemas e informática.
Luego de hacer las observaciones pertinentes a los ítems, concluyo en las siguientes
apreciaciones.

Criterios evaluados	Valoración positiva			Valoración negativa	
	MA (3)	BA (2)	A (1)	PA	NA
Calidad de redacción de los ítems.	X				
Amplitud del contenido a evaluar.		X			
Congruencia con los indicadores.	X				
Coherencia con las dimensiones.	X				

Apreciación total:

Muy adecuado (X) Bastante adecuado () A= Adecuado () PA= Poco adecuado () No
adecuado ()

Nuevo Chimbote, a los 10 días del mes de enero del 2024

Apellidos y nombres: Alvarez Minchola Blanca DNI: 08268535 Firma: [Firma]



Psic. BLANCA ALVAREZ MINCHOLA
C. Ps. 10004
PSICOLOGÍA

Modelo Predictivo del Rendimiento Académico de Estudiantes Universitarios
Asociado con la Influencia de los Videojuegos usando Machine Learning.



Anexo 3: Validación de Instrumento por Médico Psiquiatra



UNIVERSIDAD NACIONAL DEL SANTA
Escuela Profesional de Ingeniería de Sistemas e Informática

PRESENTACIÓN A JUICIO DE EXPERTO

Estimado Validador: Dr. Martín Sabino Alva Díaz

Me es grato dirigirme a usted, a fin de solicitar su colaboración como experto para validar el instrumento que adjunto denominado:

Cuestionario sobre el consumo de videojuegos, diseñado por los bachilleres Diego Alberto Castañeda Chávez y Valeria Isabel Perez Chang, cuyo propósito es medir la influencia de los videojuegos el cual será aplicado a estudiantes de la EP Ingeniería de sistemas e informática por cuanto considero que sus observaciones, apreciaciones y acertados aportes serán de utilidad.

El presente instrumento tiene como finalidad recoger información directa para la investigación que se realiza actualmente, titulada:


"Modelo Predictivo del rendimiento académico de estudiantes universitarios asociado con la influencia de los videojuegos usando Machine Learning"

Tesis que será presentada a la Universidad Nacional del Santa, como requisito para obtener el Título profesional de:

Ingeniería de Sistemas e Informática

Para efectuar la validación del instrumento, usted deberá leer cuidadosamente cada enunciado y sus correspondientes alternativas de respuesta, en donde se pueden seleccionar una, varias o ninguna alternativa de acuerdo al criterio personal y profesional del actor que responda al instrumento. Se le agradece cualquier sugerencia referente a redacción, contenido, pertinencia y congruencia u otro aspecto que se considere relevante para mejorar el mismo.

Gracias por su aporte


Bach. Castañeda Chávez Diego


Bach. Perez Chang Valeria

Modelo Predictivo del Rendimiento Académico de Estudiantes Universitarios Asociado con la Influencia de los Videojuegos usando Machine Learning.



JUICIO DE EXPERTO SOBRE LA PERTINENCIA DEL INSTRUMENTO

Instrucciones: Marque con una X en donde corresponde, que, según su criterio, Si cumple o No cumple, la coherencia entre dimensiones e indicadores de la variable en estudio.

Variable	Dimensiones	Indicadores	N° de ítem	COHERENCIA	
				SI	NO
Influencia de los videojuegos (V.I.)	Capacidad económica	- Ingresos económicos - Gastos mensuales - Inversión en hardware - Relación entre situación económica y videojuegos	6, 7, 8, 9	X	
	Tiempo de consumo	- Días de juego a la semana - Horas diarias de juego - Preferencias de tiempo - Noción del tiempo	10, 11, 12, 13	X	
	Bienestar físico/mental	- Autoevaluación del bienestar general - Pérdida de horas de sueño - Niveles de estrés - Resiliencia	14, 15, 16, 17	X	
	Impacto sobre actividades cotidianas	- Autoevaluación del impacto del consumo de videojuegos - Dificultad para realizar actividades - Percepción en el rendimiento académico - Impacto en la productividad académica	18, 19, 20, 21	X	
	Relaciones sociales	- Cambios en la capacidad de comunicación e interacción social - Conflictos interpersonales - Autoevaluación sobre las habilidades de colaboración y trabajo en equipo - Descuido hacia personas del entorno	22, 23, 24, 25	X	
	Desarrollo Cognitivo	- Desafíos cognitivos en los videojuegos - Adaptación a entornos virtuales - Desarrollo de habilidades multitarea - Transferencia de habilidades a la vida real	26, 27, 28, 29	X	

Modelo Predictivo del Rendimiento Académico de Estudiantes Universitarios Asociado con la Influencia de los Videojuegos usando Machine Learning.



Instrucciones de Evaluación de ítems: Coloque en cada casilla de valoración la letra o letras correspondiente al aspecto cualitativo que, según su criterio, cumple o tributa cada ítem a medir los aspectos o dimensiones de la variable en estudio. Las valoraciones son las siguientes:

MA= Muy adecuado / BA= Bastante adecuado / A= Adecuado / PA= Poco adecuado / NA= No adecuado

Categorías a evaluar: Redacción, contenido, congruencia y coherencia en relación a la variable de estudio. En la casilla de observaciones puede sugerir mejoras.

PREGUNTAS			VALORACIÓN					Observaciones
N°	Dimensiones	Ítems	MA	BA	A	PA	NA	
1	Datos demográficos	¿Cuál es tu edad?	X					
2		¿Cuál es tu sexo?	X					
3		¿Cuál es tu código de matrícula?		X				
4		¿Actualmente trabajas?	X					
5		Indica tu tipo de jornada laboral	X					
6	Capacidad Económica	Según tú, ¿a cuánto equivale tu cartera de videojuegos?	X					
7		¿Cuánto dinero destinas mensualmente a la compra de videojuegos y contenido relacionado?	X					
8		¿Has invertido en la compra de consolas, computadoras o hardware específico para mejorar tu experiencia de juego?	X					
9		¿Tu situación económica influye en tus preferencias por juegos gratuitos, de bajo costo o suscripciones?	X					
10	Tiempo de consumo de videojuegos	¿Cuántos días a la semana juegas videojuegos?	X					
11		¿Cuántos horas al día dedicas a jugar videojuegos?		X				
12		¿Tienes preferencias específicas para jugar videojuegos en ciertos momentos del día?		X				
13		¿Con qué frecuencia pierdes la noción del tiempo mientras juegas?		X				
14	Bienestar Físico - Mental	¿Cómo calificarías tu bienestar general en relación con tu consumo de videojuegos?		X				
15		¿El tiempo que dedicas a los videojuegos te ha hecho perder horas de sueño?		X				
16		¿Experimentas estrés o frustración al momento de jugar videojuegos?	X					
17	Impacto sobre actividades cotidianas	¿Con qué frecuencia te sientes culpable o arrepentido por la cantidad de tiempo que pasas jugando videojuegos?	X					
18		¿Qué nivel de impacto consideras que ha tenido el consumo de videojuegos en tus actividades cotidianas?		X				
19		¿Con qué frecuencia tienes dificultades para realizar actividades (académicas, físicas, sociales, etc.) por usar videojuegos?	X	X				
20		¿Con qué frecuencia consideras que consumir videojuegos afecta tu rendimiento académico?	X					
21		¿Has notado algún cambio en tu capacidad para estudiar y completar tareas académicas debido a tu consumo de videojuegos?	X					

Modelo Predictivo del Rendimiento Académico de Estudiantes Universitarios Asociado con la Influencia de los Videojuegos usando Machine Learning.



22	Relaciones Sociales	¿Has experimentado cambios en tu capacidad para comunicarte o interactuar socialmente como resultado del consumo de videojuegos?			X		
23		¿Con que frecuencia has tenido peleas con otras personas (por ejemplo, familia, amigos) por el tiempo que dedicas a los juegos?			X		
24		¿En qué nivel consideras que los videojuegos han afectado tus habilidades para colaborar y trabajar en equipo?		X			
25		¿Con que frecuencia has descuidado a personas de tu entorno (p. ej., familia, amigos) por estar jugando?		X			
26	Desarrollo Cognitivo	¿Crees que los videojuegos han contribuido al desarrollo de tus habilidades cognitivas, como la resolución de problemas y la toma de decisiones?			X		
27		¿Crees que jugar videojuegos te ha ayudado a adaptarte mejor a entornos virtuales y tecnológicos?		X			
28		¿Crees que jugar videojuegos ha mejorado tu capacidad para realizar tareas múltiples eficientemente?	X				
29		¿Has proyectado las habilidades aprendidas al jugar videojuegos a situaciones de la vida real?	X				
Total:							

Evaluado por: (Apellidos y Nombres) Alba DIAZ MARTIN SAAVEDRA

D.N.I.: 00956035

Fecha: 11/01/24

Firma:


Dr. Martin Alba Diaz
MEDICO PSICHIATRA
GMP: 26620 RNE: 36855

Modelo Predictivo del Rendimiento Académico de Estudiantes Universitarios Asociado con la Influencia de los Videojuegos usando Machine Learning.



CONSTANCIA DE VALIDACIÓN

Yo, Martín Sabino Alva Díaz, con Documento Nacional de Identidad N° 00956035, de profesión Médico, grado académico Doctor, con código de colegiatura 26620, labor que ejerzo actualmente como Médico Psiquiatra, en la Institución Hospital Regional Eleazar Guzmán Baños

Por medio de la presente hago constar que he revisado con fines de Validación el Instrumento denominado Cuestionario Influencia de Videojuegos, cuyo propósito es medir Consumo de Videojuegos, a los efectos de su aplicación a estudiantes de E.P. Ingeniería de Sistemas e Informática.
Luego de hacer las observaciones pertinentes a los ítems, concluyo en las siguientes apreciaciones.

Criterios evaluados	Valoración positiva			Valoración negativa	
	MA (3)	BA (2)	A (1)	PA	NA
Calidad de redacción de los ítems.	X				
Amplitud del contenido a evaluar.	X				
Congruencia con los indicadores.	X				
Coherencia con las dimensiones.	X				

Apreciación total:

Muy adecuado (X) Bastante adecuado () A= Adecuado () PA= Poco adecuado () No adecuado ()

Nuevo Chimbote, a los once días del mes de Enero del 20 24

Apellidos y nombres: Alva Díaz Martín Sabino DNI: 00956035 Firma: Martín Alva Díaz
MEDICO PSIQUIATRA
CMP: 26620 RNE: 36855

Modelo Predictivo del Rendimiento Académico de Estudiantes Universitarios
Asociado con la Influencia de los Videojuegos usando Machine Learning.



Anexo 4: Validación de Instrumento por Docente Universitario



UNIVERSIDAD NACIONAL DEL SANTA
Escuela Profesional de Ingeniería de Sistemas e Informática

PRESENTACIÓN A JUICIO DE EXPERTO

Estimado Validador: Dra. Maribel Alegre Jara

Me es grato dirigirme a usted, a fin de solicitar su colaboración como experto para validar el instrumento que adjunto denominado:

Cuestionario sobre el consumo de videojuegos, diseñado por los bachilleres Diego Alberto Castañeda Chávez y Valeria Isabel Perez Chang, cuyo propósito es medir la influencia de los videojuegos el cual será aplicado a estudiantes de la E.P. de Ing. de Sistemas e Inform., por cuanto considero que sus observaciones, apreciaciones y acertados aportes serán de utilidad.

El presente instrumento tiene como finalidad recoger información directa para la investigación que se realiza actualmente, titulada:

"Modelo Predictivo del rendimiento académico de estudiantes universitarios asociado con la influencia de los videojuegos usando Machine Learning"

Tesis que será presentada a la Universidad Nacional del Santa, como requisito para obtener el Título profesional de:

Ingeniería de Sistemas e Informática

Para efectuar la validación del instrumento, usted deberá leer cuidadosamente cada enunciado y sus correspondientes alternativas de respuesta, en donde se pueden seleccionar una, varias o ninguna alternativa de acuerdo al criterio personal y profesional del actor que responda al instrumento. Se le agradece cualquier sugerencia referente a redacción, contenido, pertinencia y congruencia u otro aspecto que se considere relevante para mejorar el mismo.

Gracias por su aporte


Bach. Castañeda Chávez Diego


Bach. Perez Chang Valeria

Modelo Predictivo del Rendimiento Académico de Estudiantes Universitarios Asociado con la Influencia de los Videojuegos usando Machine Learning.



JUICIO DE EXPERTO SOBRE LA PERTINENCIA DEL INSTRUMENTO

Instrucciones: Marque con una X en donde corresponde, que, según su criterio, Si cumple o No cumple, la coherencia entre dimensiones e indicadores de la variable en estudio.

Variable	Dimensiones	Indicadores	N° de ítem	COHERENCIA	
				SI	NO
Influencia de los videojuegos (V.I.)	Capacidad económica	- Ingresos económicos - Gastos mensuales - Inversión en hardware - Relación entre situación económica y videojuegos	6, 7, 8, 9	X	
	Tiempo de consumo	- Días de juego a la semana - Horas diarias de juego - Preferencias de tiempo - Noción del tiempo	10, 11, 12, 13	X	
	Bienestar físico/mental	- Autoevaluación del bienestar general - Pérdida de horas de sueño - Niveles de estrés - Resiliencia	14, 15, 16, 17	X	
	Impacto sobre actividades cotidianas	- Autoevaluación del impacto del consumo de videojuegos - Dificultad para realizar actividades - Percepción en el rendimiento académico - Impacto en la productividad académica	18, 19, 20, 21	X	
	Relaciones sociales	- Cambios en la capacidad de comunicación e interacción social - Conflictos interpersonales - Autoevaluación sobre las habilidades de colaboración y trabajo en equipo - Descuido hacia personas del entorno	22, 23, 24, 25	X	
	Desarrollo Cognitivo	- Desafíos cognitivos en los videojuegos - Adaptación a entornos virtuales - Desarrollo de habilidades multitarea - Transferencia de habilidades a la vida real	26, 27, 28, 29	X	

Modelo Predictivo del Rendimiento Académico de Estudiantes Universitarios Asociado con la Influencia de los Videojuegos usando Machine Learning.



Instrucciones de Evaluación de ítems: Coloque en cada casilla de valoración la letra o letras correspondiente al aspecto cualitativo que, según su criterio, cumple o tributa cada ítem a medir los aspectos o dimensiones de la variable en estudio. Las valoraciones son las siguientes:

MA= Muy adecuado / BA= Bastante adecuado / A = Adecuado / PA= Poco adecuado / NA= No adecuado

Categorías a evaluar: Redacción, contenido, congruencia y coherencia en relación a la variable de estudio. En la casilla de observaciones puede sugerir mejoras.

PREGUNTAS			VALORACIÓN					Observaciones
N°	Dimensiones	Ítems	MA	BA	A	PA	NA	
1	Datos demográficos	¿Cuál es tu edad?	X					
2		¿Cuál es tu sexo?	X					
3		¿Cuál es tu código de matrícula?	X					
4		¿Actualmente trabajas?	X					
5		Indica tu tipo de jornada laboral	X					
6	Capacidad Económica	Según tú, ¿a cuánto equivale tu cartera de videojuegos?	X					
7		¿Cuánto dinero destinas mensualmente a la compra de videojuegos y contenido relacionado?		X				
8		¿Has invertido en la compra de consolas, computadoras o hardware específico para mejorar tu experiencia de juego?		X				
9		¿Tu situación económica influye en tus preferencias por juegos gratuitos, de bajo costo o suscripciones?		X				
10	Tiempo de consumo de videojuegos	¿Cuántos días a la semana juegas videojuegos?		X				
11		¿Cuántos horas al día dedicas a jugar videojuegos?		X				
12		¿Tienes preferencias específicas para jugar videojuegos en ciertos momentos del día?		X				
13		¿Con qué frecuencia pierdes la noción del tiempo mientras juegas?		X				
14	Bienestar Físico - Mental	¿Cómo calificarías tu bienestar general en relación con tu consumo de videojuegos?		X				
15		¿El tiempo que dedicas a los videojuegos te ha hecho perder horas de sueño?		X				
16		¿Experimentas estrés o frustración al momento de jugar videojuegos?		X				
17		¿Con qué frecuencia te sientes culpable o arrepentido por la cantidad de tiempo que pasas jugando videojuegos?		X				
18	Impacto sobre actividades cotidianas	¿Qué nivel de impacto consideras que ha tenido el consumo de videojuegos en tus actividades cotidianas?		X				
19		¿Con qué frecuencia tienes dificultades para realizar actividades (académicas, físicas, sociales, etc.) por usar videojuegos?		X				
20		¿Con qué frecuencia consideras que consumir videojuegos afecta tu rendimiento académico?		X				
21		¿Has notado algún cambio en tu capacidad para estudiar y completar tareas académicas debido a tu consumo de videojuegos?		X				

Modelo Predictivo del Rendimiento Académico de Estudiantes Universitarios Asociado con la Influencia de los Videojuegos usando Machine Learning.



22	Relaciones Sociales	¿Has experimentado cambios en tu capacidad para comunicarte o interactuar socialmente como resultado del consumo de videojuegos?	X				
23		¿Con que frecuencia has tenido peleas con otras personas (por ejemplo, familia, amigos) por el tiempo que dedicas a los juegos?	X				
24		¿En qué nivel consideras que los videojuegos han afectado tus habilidades para colaborar y trabajar en equipo?	X				
25		¿Con que frecuencia has descuidado a personas de tu entorno (p. ej., familia, amigos) por estar jugando?	X				
26	Desarrollo Cognitivo	¿Crees que los videojuegos han contribuido al desarrollo de tus habilidades cognitivas, como la resolución de problemas y la toma de decisiones?	X				
27		¿Crees que jugar videojuegos te ha ayudado a adaptarte mejor a entornos virtuales y tecnológicos?		X			
28		¿Crees que jugar videojuegos ha mejorado tu capacidad para realizar tareas múltiples eficientemente?		X			
29		¿Has proyectado las habilidades aprendidas al jugar videojuegos a situaciones de la vida real?		X			
Total:							

Evaluado por: (Apellidos y Nombres) Alegre Tara Maribel Engida

D.N.I.: 32959163

Fecha: 11/01/24

Firma: [Firma]



CONSTANCIA DE VALIDACIÓN

Yo, Maribel Alegre Jara, con Documento Nacional de Identidad N° 32959163 de profesión Docente, grado académico Doctorado, con código de colegiatura _____, labor que ejerzo actualmente como Docente Universitaria, en la Institución Universidad Nacional del Santa.

Por medio de la presente hago constar que he revisado con fines de Validación el Instrumento denominado Cuestionario sobre el consumo de videojuegos cuyo propósito es medir la influencia de los videojuegos, a los efectos de su aplicación a estudiantes de E.P. Ingeniería de sistemas e Informática.

Luego de hacer las observaciones pertinentes a los ítems, concluyo en las siguientes apreciaciones.

Criterios evaluados	Valoración positiva			Valoración negativa	
	MA (3)	BA (2)	A (1)	PA	NA
Calidad de redacción de los ítems.	X				
Amplitud del contenido a evaluar.	X				
Congruencia con los indicadores.	X				
Coherencia con las dimensiones.	X				

Apreciación total:

Muy adecuado (X) Bastante adecuado () A= Adecuado () PA= Poco adecuado () No adecuado ()

Nuevo Chimbote, a los once días del mes de enero del 2024

Apellidos y nombres: Alegre Jara Maribel E.

DNI: 32959163

Firma:



Anexo 5: Solicitud de información a rectora de la UNS

Nuevo Chimbote, 08 de setiembre de 2023



Dra. America Odar Rosario
Rectora de la UNS

Reciba mi más cordial saludo, siendo egresados de la carrera de Ingeniería de Sistemas e Informática (Diego Alberto Castañeda Chávez con código de estudiante 0201514033 y Valeria Isabel Pérez Chang con código de estudiante 0201614046) y habiendo sido aprobado nuestro Proyecto de Tesis titulado "MODELO PREDICTIVO DEL RENDIMIENTO ACADÉMICO DE ESTUDIANTES UNIVERSITARIOS ASOCIADO CON LA INFLUENCIA DE LOS VIDEOJUEGOS USANDO MACHINE LEARNING", con Resolución N° 532-2023-UNS-CFI el 24 de agosto del presente, es que recurrimos a usted para que por medio de su despacho se oficialice nuestra solicitud referente a la información del historial académico de los estudiantes de la Escuela Profesional de Ingeniería de Sistemas e Informática del Periodo académico 2022 – I, 2022 – II y 2023 – I, requerida para el proceso de diseño, elaboración y aplicación de nuestro proyecto, con el compromiso de no publicar ni divulgar ninguno de los datos proporcionados.

Desde ya agradezco su atención y comprensión.

Saludos cordiales.

Atentamente.

Bach. Valeria Isabel Perez Chang
Cod Matrícula: 0201614046



Anexo 6: Derivación de solicitud de información al área encargada



LICENCIADA POR LA SUNEDU

"Año de la Unidad, la Paz y el Desarrollo"



MEMORANDO N° 02679 -2023-UNS-R.

A **Mg. MIRKO MANRIQUE RONCEROS**
Director Evaluación y Desarrollo Académico

ASUNTO **BRINDAR ATENCIÓN - SOLICITUD BACH. VALERIA PEREZ CHANG, OFICIALIZACIÓN DE PROYECTO DE TESIS**

REF. Solicitud de fecha 08.09.2023 SUT N 1385-23-V
RESOLUCIÓN n° 532-2023-UNS-CFI

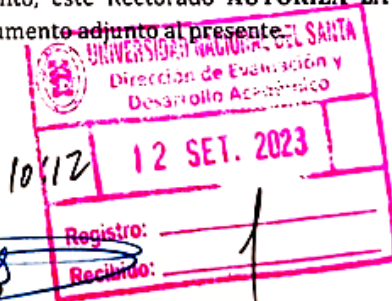
FECHA Nuevo Chimbote, 11 de setiembre de 2023.

En atención al documento de la referencia, presentado por los estudiantes **Diego Alberto Castañeda Chávez**, Código 0201514033 y **Valeria Isabel Pérez Chang** Código 201614046, mediante el cual solicitan información del historial académico de los estudiantes de la Escuela Profesional de Ingeniería de Sistemas e Informática del Periodo Académico 2022-I, 2022-II y 2023-I, requerida para el proceso de diseño y aplicación del Proyecto de Tesis, Titulado: "MODELO PREDICTIVO DEL RENDIMIENTO ACADÉMICO DE ESTUDIANTES UNIVERSITARIOS ASOCIADO CON LA INFLUENCIA DE LOS VIDEOJUEGOS USANDO MACHINE LEARNING". Por tanto, este Rectorado **AUTORIZA LA ATENCIÓN CORRESPONDIENTE**, del documento adjunto al presente.

Atentamente,



América Odar Rosario
Rectora UNS



Adj. Solicitud de fecha 08.09.2023 SUT N 1385-23-V (...2...fls).

cc: Archivo

OS/Dafila.

BACH. CASTAÑEDA CHÁVEZ DIEGO ALBERTO

BACH. PEREZ CHANG VALERIA ISABEL



Anexo 7: Muestra de la información recibida



UNIVERSIDAD NACIONAL DEL SANTA

DIRECCIÓN DE EVALUACIÓN Y
DESARROLLO ACADÉMICO

REGISTRO DE CALIFICACIONES

COD.ESP.	ESCUELA ACADÉMICO PROFESIONAL : ESPECIALIDAD				CICLO	SEM.ACAD.
140	INGENIERIA DE SISTEMAS E INFORMÁTICA				03	2023-01
PLAN CUR.	COD. ASIG.	NOMBRE DE ASIGNATURA				GRUPO CRED.
20181401	1411-0017	ESTADISTICA PARA INGENIEROS				01 04
PROFESOR	BOZA ROSARIO JOSE ANTONIO		DEPARTAMENTO ACAD.		MATEMATICAS	

UNID.	PESO	TIPO DE EVALUACIÓN	PESO
1	1	EVALUACIÓN DE UNIDAD (E.U.)	2
2	1	EVIDENCIAS DE APREND. (E.A.)	1
3	1	PRODUCTO (PR.)	1
---	---	EXAMEN SUSTITUTORIO (E.S.)	2

20/09/2023

03:16 PM

PAG. 1

Obs : ☒

MODC: SYS

Tipo : MR

Fase : PROC

PG DE 3

#	CODIGO	NOMBRE	PROM. FINAL	PROM. FINAL S.	TERCERA UNIDAD					SEGUNDA UNIDAD					PRIMERA UNIDAD				
					E.U.	E.A.	PR.	E.S.	P.U.	E.U.	E.A.	PR.	E.S.	P.U.	E.U.	E.A.	PR.	E.S.	P.U.
	0202214		INH	-											00.00	00.00	00.00		00.00
	0202214		10	11	09.00	12.00	13.00		11.00	05.00	13.00	12.00		10.00	07.00	08.00	11.00	14.00	11.00
	0201914		12	-	14.00	12.00	13.00		13.00	14.00	13.00	11.00		13.00	15.00	12.00	06.00		11.00
	0202214		12	-	10.00	13.00	12.00		12.00	08.00	13.00	12.00		11.00	09.00	14.00	12.00		12.00
	0202214		12	-	09.00	14.00	12.00		12.00	14.00	13.00	14.00		14.00	08.00	14.00	12.00		11.00
	0202214		13	-	14.00	13.00	12.00		13.00	15.00	14.00	13.00		14.00	13.00	14.00	13.00		13.00
	0202214		12	-	09.00	13.00	13.00		12.00	13.00	14.00	12.00		13.00	12.00	14.00	11.00		12.00
	0202214		15	-	14.00	14.00	15.00		14.00	18.00	17.00	17.00		17.00	11.00	14.00	17.00		14.00
	0202214		14	-	14.00	15.00	14.00		14.00	15.00	16.00	15.00		15.00	13.00	13.00	13.00		13.00
	0202214		INH	-											07.00	04.00	07.00		06.00
	0202214		12	-	11.00	14.00	12.00		12.00	15.00	12.00	13.00		13.00	12.00	13.00	12.00		12.00
	0202214		14	-	10.00	15.00	15.00		13.00	13.00	16.00	16.00		15.00	14.00	14.00	14.00		14.00
	0202214		11	-	08.00	14.00	11.00		11.00	13.00	12.00	11.00		12.00	11.00	11.00	12.00		11.00
	0202214		INH	-											07.00				02.00
	0202214		INH	-											08.00				03.00
	0202214		13	-	08.00	15.00	14.00		12.00	12.00	14.00	14.00		13.00	11.00	14.00	14.00		13.00
	0202214		13	-	09.00	14.00	14.00		12.00	14.00	14.00	14.00		14.00	12.00	13.00	16.00		14.00
	0202214		13	-	08.00	15.00	15.00		13.00	13.00	14.00	16.00		14.00	11.00	14.00	12.00		12.00
	0202214		13	-	09.00	15.00	15.00		13.00	13.00	13.00	16.00		14.00	11.00	12.00	15.00		13.00
	0201914		11	-	11.00	12.00	11.00		11.00	09.00	12.00	11.00		11.00	11.00	11.00	13.00		12.00
	0202314		INH	-						05.00					02.00		05.00		02.00
	0202214		INH	-						12.00					04.00	07.00			02.00
	0202214		14	-	15.00	14.00	12.00		14.00	14.00	16.00	13.00		14.00	13.00	14.00	15.00		14.00
	0202214		12	-	10.00	13.00	14.00		12.00	09.00	11.00	12.00		11.00	14.00	14.00	12.00		13.00
	0202214		12	-	10.00	12.00	13.00		12.00	14.00	13.00	12.00		13.00	13.00	13.00	11.00		12.00

BAR CODE.

MATRICULADOS : 56

APROBADOS : 48

DESAPROBADOS : 01

INHABILITADOS : 07

45557

___/___/2023

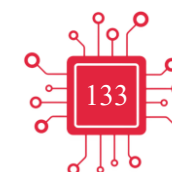
(DD/MM/YYYY)

FECHA DE CONFORMIDAD DEL PROFESOR

TERCERA UNIDAD	SEGUNDA UNIDAD	PRIMERA UNIDAD
PROFESOR RESPONSABLE	PROFESOR RESPONSABLE	PROFESOR RESPONSABLE

BACH. CASTAÑEDA CHÁVEZ DIEGO ALBERTO

BACH. PEREZ CHANG VALERIA ISABEL





Anexo 8: Código Python del Modelo de Regresión Lineal

```
#Importación de librerías a utilizar
import pandas as pd
import matplotlib.pyplot as plt

from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error, r2_score

# Cargar el dataset
df_dataRL = pd.read_excel('DATA FINAL LIMPIA RL.xlsx')
#Para generar datos de entrenamiento y test aleatorio
x = df_dataRL[['PUNTAJE']]
y = df_dataRL['Ponderado']

# Dividir el dataset en conjunto de entrenamiento (80%) y de prueba (20%)
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2, random_state=42)
df_dataRL.info()

#Generando el modelo de Regresión
modelo = LinearRegression()
modelo.fit(x_train, y_train)

# imprimiendo parámetros
intercepto = modelo.intercept_
pendiente = modelo.coef_
print(f"intercepto (b): {intercepto}")
print(f"pendiente (w): {pendiente}")

# Making predictions on the test set
y_pred = modelo.predict(x_test)

#Error Cuadrado Medio
print("Error Cuadrado Medio: %.2f" % mean_squared_error(y_test, y_pred))

#Puntaje de Varianza. El mejor puntaje es un 1.0
print("Varianza: %.2f" % r2_score(y_test, y_pred))

# Gráfica de dispersión con la línea de regresión lineal
plt.figure(figsize=(10, 6))
plt.scatter(x_test, y_test, color='blue', label='Valores Reales')
plt.plot(x_test, y_pred, color='red', lw=2, label='Línea de Regresión Lineal')
plt.xlabel('PUNTAJE')
plt.ylabel('Ponderado')
plt.title('Valores Reales vs Predicciones del Modelo de Regresión Lineal')
plt.legend()
```



```
plt.grid(True)  
plt.show()
```

```
# Crear un DataFrame para mostrar las predicciones
```

```
resultados = pd.DataFrame({"Valor Real": y_test, 'Predicción': y_pred})
```

```
# Mostrar los 15 primeros resultados
```

```
print(resultados.head(15))
```



Anexo 9: Código Python del Modelo de Regresión Múltiple

```
#Importación de librerías a utilizar
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score

#Cargar el dataset
data = pd.read_excel("DATA FINAL LIMPIA OTROS MODELOS.xlsx")

#Definir las variables predictoras y la variable objetivo
X = data.drop(columns=['Ponderado'])
y = data['Ponderado']

#Dividir el dataset en conjunto de entrenamiento (80%) y de prueba (20%)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

#Crear y entrenar el modelo de regresión múltiple
model = LinearRegression()
model.fit(X_train, y_train)

#Realizar predicciones sobre el conjunto de prueba
y_pred = model.predict(X_test)

#Calcular el MSE y R²
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

#Obtener los coeficientes (pendientes) y el intercepto del modelo
intercepto = model.intercept_
coeficientes = model.coef_

# Crear un DataFrame para mostrar los coeficientes y el intercepto
coef_df = pd.DataFrame({
    'Variable': X.columns,
    'Coeficiente': coeficientes
})

# Mostrar los resultados del MSE, R², intercepto y coeficientes
print("MSE:", mse)
print("R²:", r2)
print("Intercepto:", intercepto)
print(coef_df)

#Generar gráfico de dispersión
#Valores predichos por el modelo
```



```
y_pred = model.predict(X_test)

plt.figure(figsize=(10, 6))

#Valores reales
plt.scatter(range(len(y_test)), y_test, color='red', label='Valores Reales', alpha=0.5)

#Valores predichos
plt.scatter(range(len(y_pred)), y_pred, color='blue', label='Valores Predichos', alpha=0.5)

plt.xlabel('Índice')
plt.ylabel('Ponderado')
plt.title("Valores Reales vs Predicciones del Modelo de Regresión Múltiple")
plt.legend()
plt.grid(True)
plt.show()

#Crear un DataFrame para mostrar las predicciones
resultados = pd.DataFrame({'Valor Real': y_test, 'Predicción': y_pred})

#Mostrar los 15 primeros resultados
print(resultados.head(15))
```



Anexo 10: Código Python del Modelo Random Forest

```
#Importación de librerías a utilizar
import pandas as pd
import matplotlib.pyplot as plt

from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error, r2_score

#Cargar el dataset
data = pd.read_excel("DATA FINAL LIMPIA OTROS MODELOS.xlsx")

#Separar las variables predictoras y la variable objetivo
X = data.drop('Ponderado', axis=1)
y = data['Ponderado']

#Dividir el conjunto de datos en entrenamiento (80%) y prueba (20%)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

#Crear y entrenar el modelo Random Forest
model = RandomForestRegressor(random_state=42)
model.fit(X_train, y_train)

#Realizar predicciones
y_pred = model.predict(X_test)

#Calcular MSE y R²
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

#Obtener los coeficientes (importancias) del modelo
importancias = model.feature_importances_

#Crear un DataFrame para las importancias de las características
coef_df = pd.DataFrame({
    'Feature': X.columns,
    'Importance': importancias
}).sort_values(by='Importance', ascending=False)

#Función para realizar predicciones con el modelo entrenado
def predict_ponderado(model, input_data):
    return model.predict([input_data])[0]

#Resultados
print(f'MSE: {mse}')
print(f'R²: {r2}')
```




```
print('Importancias de las características:')
print(coef_df)

#Generar gráfico de dispersión
#Valores predichos por el modelo
y_pred = model.predict(X_test)

plt.figure(figsize=(10, 6))

#Valores reales
plt.scatter(range(len(y_test)), y_test, color='red', label='Valores Reales', alpha=0.5)

#Valores predichos
plt.scatter(range(len(y_pred)), y_pred, color='blue', label='Valores Predichos', alpha=0.5)

plt.xlabel('Índice')
plt.ylabel('Ponderado')
plt.title('Valores Reales vs Predicciones del Modelo Random Forest')
plt.legend()
plt.grid(True)
plt.show()
```



Anexo 11: Código Python del Modelo de Regresión Logística

```
#Importación de librerías a utilizar
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import mean_squared_error, r2_score, confusion_matrix

#Cargar los datos
data = pd.read_excel("DATA FINAL LIMPIA MODELOS DE CLASIFICACION.xlsx")

#Definir variables predictoras y la variable objetivo
X = data.drop(columns=['Ponderado'])
y = data['Ponderado']

#Dividir el dataset en conjunto de entrenamiento (80%) y prueba (20%)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

#Crear y entrenar el modelo de regresión logística
model = LogisticRegression(max_iter=10000)
model.fit(X_train, y_train)

#Predecir en el conjunto de prueba
y_pred = model.predict(X_test)

#Obtener coeficientes (interceptos y pendientes)
intercepto = model.intercept_
coeficientes = model.coef_

#Generar la matriz de confusión
conf_matrix = confusion_matrix(y_test, y_pred)

#Mostrar resultados
print(f"Intercepto: { intercepto }")
print(f"Coeficientes: { coeficientes }")
print(f"Matriz de Confusión:\n {conf_matrix}")
#Graficar la matriz de confusión
plt.figure(figsize=(8, 6))
sns.heatmap(conf_matrix, annot=True, fmt="d", cmap="Blues", xticklabels=['Desaprobado', 'Aprobado'],
yticklabels=['Desaprobado', 'Aprobado'])
plt.xlabel('Predicción')
plt.ylabel('Real')
plt.title('Matriz de Confusión')
plt.show()
```



Anexo 12: Código Python del Modelo Naive Bayes

```
#Importar las librerías necesarias
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB
from sklearn.metrics import mean_squared_error, r2_score, confusion_matrix, classification_report

#Cargar el dataset
data = pd.read_excel('DATA FINAL LIMPIA MODELOS DE CLASIFICACION.xlsx')

#Separar variables predictoras y objetivo
X = data.drop(columns=['Ponderado'])
y = data['Ponderado']

#Dividir el dataset en entrenamiento y prueba
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

#Crear el modelo Naive Bayes
model = GaussianNB()

#Entrenar el modelo
model.fit(X_train, y_train)

#Predecir en el conjunto de prueba
y_pred = model.predict(X_test)

#Generar la matriz de confusión
conf_matrix = confusion_matrix(y_test, y_pred)

#Crear un DataFrame para la matriz de confusión
conf_matrix_df = pd.DataFrame(conf_matrix, index=['Desaprobado', 'Aprobado'], columns=['Predicho Desaprobado', 'Predicho Aprobado'])

#Visualizar la matriz de confusión
plt.figure(figsize=(8, 6))
sns.heatmap(conf_matrix_df, annot=True, fmt='d', cmap='Blues')
plt.xlabel('Predicción')
plt.ylabel('Real')
plt.title('Matriz de Confusión')
plt.show()

# Imprimir resultados
print("Coeficientes del modelo:")
print(model.theta_)
print("Matriz de Confusión:")
print(conf_matrix_df)
```