



UNS
ESCUELA DE
POSGRADO

**Programa de Maestría en Ingeniería de Sistemas e Informática
mención Gestión de Tecnologías de la Información**

**“Machine Learning para predecir la cantidad de alumnos
de Pregrado a matricularse por asignatura en el periodo
Académico 2019-II en la Universidad Nacional del Santa”**

**Tesis para optar el grado de Maestro en Ingeniería
de Sistemas e informática mención Gestión de
Tecnologías de la Información**

Autor:

Br. Ninaquispe Matame, Humberto Angel

Asesor:

Dr. Urrelo Huiman, Luis Vladimir

**Nuevo Chimbote - PERÚ
2022**



UNS
ESCUELA DE
POSGRADO

CONSTANCIA DE ASESORAMIENTO DE LA TESIS

Yo, Dr. Luis Vladimir Urrelo Huiman, mediante la presente certifico mi asesoramiento de la Tesis de Maestría titulada: **MACHINE LEARNING PARA PREDECIR LA CANTIDAD DE ALUMNOS DE PREGRADO A MATRICULARSE POR ASIGNATURA EN EL PERIODO ACADÉMICO 2019-II EN LA UNIVERSIDAD NACIONAL DEL SANTA**, elaborada por el bachiller Humberto Ángel Ninaquispe Matame, para obtener el Grado Académico de Maestro en Ingeniería de Sistemas e Informática mención Gestión de Tecnologías de la Información en la Escuela de Posgrado de la Universidad Nacional del Santa.

Nuevo Chimbote, octubre del 2022

DR. LUIS VLADIMIR URRELO HUIMAN

ASESOR

CODIGO ORCID: 0000-0003-1523-2640

DNI N° 40010219



UNS
ESCUELA DE
POSGRADO

CONFORMIDAD DEL JURADO EVALUADOR

“MACHINE LEARNING PARA PREDECIR LA CANTIDAD DE ALUMNOS DE PREGRADO A MATRICULARSE POR ASIGNATURA EN EL PERIODO ACADEMICO 2019-II EN LA UNIVERSIDAD NACIONAL DEL SANTA”

TESIS PARA OPTAR EL GRADO DE MAESTRO EN INGENIERIA DE SISTEMAS E INFORMATICA MENCION GESTION DE TECNOLOGIA DE LA INFORMACION

Revisado y Aprobado por el Jurado Evaluador:

Dr. HUGO ESTEBAN CASELLI GISMONDI

PRESIDENTE

CODIGO ORCID: 0000-0002-2812-6727

DNI N° 32819296

Ms. YIM ISAIAS APESTEGUI FLORENTINO

SECRETARIO

CODIGO ORCID: 0000-0003-2873-1748

DNI N° 32541215

Dr. LUIS VLADIMIR URRELO HUIMAN

VOCAL

CODIGO ORCID: 0000-0003-1523-2640

DNI N° 40010219



UNS
ESCUELA DE
POSGRADO

ACTA DE EVALUACIÓN DE SUSTENTACIÓN DE TESIS

A los veintidós días del mes de octubre del año 2022, siendo las 10:00 horas, en la Escuela Profesional de Ingeniería de Sistemas e Informática de la Universidad Nacional del Santa, se reunieron los miembros del Jurado Evaluador designados mediante Resolución Directoral N° 548-2021-EPG-UNS de fecha 03 de diciembre de 2021, conformado por los docentes: Dr. Hugo Esteban Caselli Gismondi (Presidente), Ms. Yim Isaias Apestegui Florentino (Secretario) y Dr. Luis Vladimír Urrelo Huiman (Vocal), con la finalidad de evaluar la tesis titulada: **MACHINE LEARNING PARA PREDECIR LA CANTIDAD DE ALUMNOS DE PREGRADO A MATRICULARSE POR ASIGNATURA EN EL PERIODO ACADÉMICO 2019-II EN LA UNIVERSIDAD NACIONAL DEL SANTA**; presentado por el tesista **Humberto Ángel Ninaquispe Matame**, egresado del programa de **Maestría en Ingeniería de Sistemas e Informática** mención **Gestión de Tecnologías de la Información**.

Sustentación autorizada mediante Resolución Directoral N° 533-2022-EPG-UNS de fecha 29 de setiembre de 2022 y Resolución Directoral N° 540-2022-EPG-UNS de fecha 05 de octubre de 2022.

El presidente del jurado autorizó el inicio del acto académico; producido y concluido el acto de sustentación de tesis, los miembros del jurado procedieron a la evaluación respectiva, haciendo una serie de preguntas y recomendaciones al tesista, quien dio respuestas a las interrogantes y observaciones.

El jurado después de deliberar sobre aspectos relacionados con el trabajo, contenido y sustentación del mismo y con las sugerencias pertinentes, declara la sustentación como: APROBADO asignándole la calificación de: DIECIOCHO

Siendo las 11:30 horas del mismo día se da por finalizado el acto académico, firmando la presente acta en señal de conformidad.

Dr. Hugo Esteban Caselli Gismondi
Presidente

Ms. Yim Isaias Apestegui Florentino
Secretario

Dr. Luis Vladimír Urrelo Huiman
Vocal

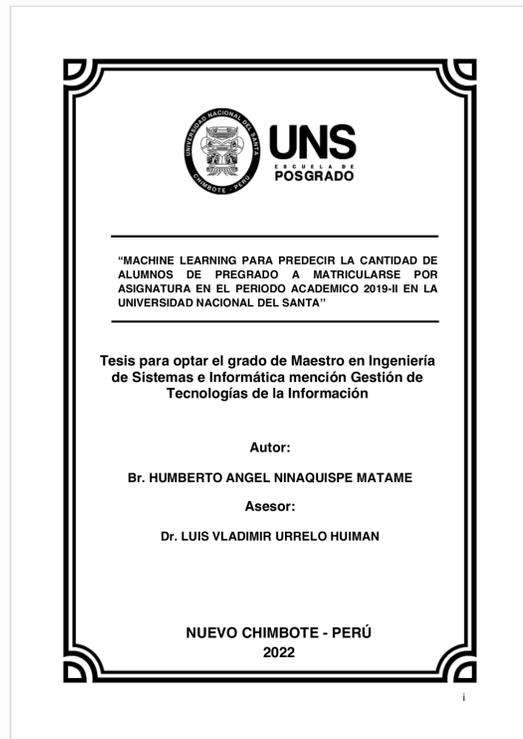


Recibo digital

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: Humberto Angel Ninaquispe Matame
Título del ejercicio: Posgrado
Título de la entrega: MACHINE LEARNING PARA PREDECIR LA CANTIDAD DE ALUM...
Nombre del archivo: tesis_ninaquispe_082022.docx
Tamaño del archivo: 9.22M
Total páginas: 95
Total de palabras: 17,104
Total de caracteres: 87,842
Fecha de entrega: 06-dic.-2022 12:30p. m. (UTC-0500)
Identificador de la entre... 1973353569



PRESENTACIÓN

Señores Miembros del Jurado:

Cumpliendo con los requerimientos estipulados en el reglamento de Grados y Títulos de la Escuela de Postgrado de la Universidad Nacional del Santa, para obtener el grado de Maestro en Ingeniería de Sistemas e Informática mención Gestión de Tecnologías de la Información, pongo a vuestra disposición la presente tesis titulada: Machine Learning para predecir la cantidad de alumnos de pregrado a matricularse por asignatura en el periodo académico 2019-II en la Universidad Nacional del Santa.

Gracias.

Chimbote, octubre del 2022

Br. Humberto Angel Ninaquispe Matame

Dedicado:

A Dios, por darme la fuerza para el logro de mis metas trazadas.

A mi familia, en especial a mi madre Paula que me enseñó que todo esfuerzo tiene una recompensa.

A mis profesores, por su ayuda y aporte en mi formación académica y profesional.

AGRADECIMIENTOS

A mi familia, que ha puesto toda su confianza en mí para el logro de un objetivo más en mi vida.

A mi institución de trabajo, que me a permitido el uso de la información para el logro de esta investigación.

A todas las personas que de una u otra manera han contribuido con la realización del presente trabajo de investigación.

ÍNDICE

ACREDITACIÓN	ii
PRESENTACIÓN.....	iv
AGRADECIMIENTOS	vi
RESUMEN	xii
ABSTRACT	xiii
INTRODUCCIÓN	14
CAPÍTULO I.....	16
PROBLEMA DE INVESTIGACIÓN	16
1.1. Planteamiento y fundamentación del problema de investigación.	16
1.2. Antecedentes de la investigación.	23
1.3. Formulación del problema de investigación.....	26
1.4. Delimitación del estudio.....	26
1.5. Justificación e importancia de la investigación.	26
1.6. Objetivos de la investigación: General y específicos.	27
CAPÍTULO II.....	28
MARCO TEÓRICO.....	28
2.1. Fundamentos teóricos de la investigación.....	28
APRENDIZAJE AUTOMATIZADO (MACHINE LEARNING):	28
APRENDIZAJE AUTOMATIZADO EN LA NUBE	29
MATRÍCULA DE ASIGNATURAS	30
2.2. Marco conceptual	33
CAPÍTULO III.....	34
MARCO METODOLÓGICO	34
3.1. Hipótesis central de la investigación.....	34
3.2. Variables e indicadores de la investigación.....	34
3.3. Métodos de la investigación.	35
3.4. Diseño o esquema de la investigación.	36
3.5. Población y muestra.	36
3.6. Actividades del proceso investigativo	37
3.7. Técnicas e instrumentos de la investigación.	38

3.8. Procedimiento para la recolección de datos (Validación y confiabilidad de los instrumentos).....	38
3.9. Técnicas de procesamiento y análisis de los datos.	38
CAPÍTULO IV	39
RESULTADOS Y DISCUSIÓN.....	39
CAPÍTULO V	81
CONCLUSIONES Y RECOMENDACIONES	81
5.1. Conclusiones.	81
5.2. Recomendaciones.....	83
REFERENCIAS BIBLIOGRÁFICAS.....	84
ANEXOS	86
1. Formato para Lista de expertos.....	86
2. Formato para evaluación de expertos	87
3. Encuesta Machine Learning	88
4. Análisis de fiabilidad de encuesta Machine Learning	89
5. Variables para número de matriculados	93
6. Análisis de fiabilidad de encuesta Variables.....	94

LISTA DE CUADROS

Tabla 01.	Porcentaje de matriculados modo web	17
Tabla 02.	Cálculo del MAPE en asignaturas de Ingeniería en el 2018-II	18
Tabla 03.	Lista de reajustes de una carga lectiva según normatividad	20
Tabla 04.	Reajuste de carga lectiva 2018-II según reglamento	21
Tabla 05.	Resumen de asignaturas procesadas 2018-II	22
Tabla 06.	Checklist modelos y técnicas ML	39
Tabla 07.	Algoritmos según docentes UNS	40
Tabla 08.	Características para infraestructura ML	44
Tabla 09.	Comparación de plataformas en la nube.....	45
Tabla 10.	Plataforma según docentes UNS	47
Tabla 11.	Lista de variables predictoras.....	50
Tabla 12.	Cuadro de promedios de grado de influencia de variables	52
Tabla 13.	Variables predictoras definitivas.....	57
Tabla 14.	Resultados de testeo periodo académico 2018-II	69
Tabla 15.	Resumen de asignaturas procesadas predicción 2018-II.....	71
Tabla 16.	Resultados de proyección según carga lectiva del periodo académico 2019-II	73
Tabla 17.	Resumen de asignaturas sobre proyección 2019-II	75
Tabla 18.	Resultados de predicción periodo académico 2019-II.....	75
Tabla 19.	Resumen de asignaturas sobre predicción 2019-II	77

LISTA DE GRÁFICOS

Gráfico 1. Cuadrante de Gartner para servicios de desarrollador de IA en la nube	30
Gráfico 2. Algoritmos sugeridos por docentes UNS.....	41
Gráfico 3. Gráfico de torta para algoritmos seleccionados	42
Gráfico 4. Cuadrante de Gartner para servicios de desarrollador de IA en la nube	45
Gráfico 5. Gráfico de barras para Plataformas según docentes UNS.....	47
Gráfico 6. Diagrama de torta para Plataformas según docentes UNS.....	48
Gráfico 7. Diagrama de dependencia	49
Gráfico 8. Gráfico de barras para el grado de influencia por variable.....	53
Gráfico 9. Atributos del modelo en Weka	55
Gráfico 10. Vista de data en Weka	56
Gráfico 11. Vista de atributos seleccionados por Weka.....	57
Gráfico 12. Creando proyecto Matrícula	58
Gráfico 13. Agregando dataset	59
Gráfico 14. Asignando Dataset a Proyecto	60
Gráfico 15. Agregando Dataset en Experimento	60
Gráfico 16. Vista de datos del Dataset	61
Gráfico 17. Agregando componentes de manipulación	61
Gráfico 18. Agregando columnas como categoría.....	62
Gráfico 19. Eliminando valores nulos	62
Gráfico 20. Definiendo valor de salida	63
Gráfico 21. Eliminación de picos.....	63
Gráfico 22. Split data	64
Gráfico 23. Algoritmo ML y entrenamiento	64
Gráfico 24. Score model Regresión Lineal	65
Gráfico 25. Score model Redes Neuronales	65
Gráfico 26. Score model en cada algoritmo.....	66
Gráfico 27. Vista General del modelo	66
Gráfico 28. Vista del servicio generado	67
Gráfico 29. Vista de la información del servicio	68

Gráfico 30. Vista de la página de testeo	68
Gráfico 31. Vista de la información obtenida	69
Gráfico 32. Fiabilidad de los datos de matriculas	72
Gráfico 33. Vista de la estadística de fiabilidad	78
Gráfico 34. Vista de la prueba de normalidad.....	79
Gráfico 35. Vista SPSS para análisis de la hipótesis.....	80
Gráfico 36. Vista de variables Encuesta Machine Learning.....	91
Gráfico 37. Vista de datos Encuesta Machine Learning	91
Gráfico 38. Alfa de Crombach Encuesta Machine Learning	92
Gráfico 39. Vista de datos Encuesta Variables.....	94
Gráfico 40. Vista de variables Encuesta Variables	94
Gráfico 41. Alfa de Cronbach Encuesta Variables.....	95

RESUMEN

Machine Learning para predecir la cantidad de alumnos de pregrado a matricularse por asignatura en el periodo académico 2019-II en la Universidad Nacional del Santa

La investigación tuvo como objetivo predecir la cantidad de alumnos de pregrado a matricularse por asignatura en el periodo académico 2019-II en la Universidad Nacional del Santa, haciendo uso de Machine Learning, para lo cual, se identificó el algoritmo apropiado para realizar el análisis predictivo investigado, luego se identificó una adecuada plataforma de implementación de algoritmos de Machine Learning en la nube, posteriormente se determinó las variables que intervienen en la predicción de la cantidad de alumnos de pregrado a matricularse por asignatura durante un periodo académico y posteriormente se elaboró una solución de Machine Learning para predecir dicha cantidad. Al identificar el algoritmo apropiado, se observó que se puede utilizar varios algoritmos y luego compararlos en la solución de Machine Learning, también se identificó a Microsoft Azure como plataforma adecuada para implementar la solución de Machine Learning, posteriormente se logró determinar diez variables predictoras como las más apropiadas para la investigación y finalmente, haciendo uso del algoritmo, plataforma y variables seleccionadas, se elaboró una solución de Machine Learning para predecir la cantidad de alumnos de pregrado a matricularse por asignatura; al hacer uso de esta solución con los valores para el periodo académico 2019-II, se observó que existe similitud con la cantidad real de alumnos que se matricularon en cada asignatura en dicho semestre. Por lo que se concluye que Machine Learning permite predecir la cantidad de alumnos de pregrado a matricularse por asignatura en el periodo académico 2019-II en la Universidad Nacional del Santa.

Por Br. Humberto Angel Ninaquispe Matame

Palabras clave: Machine Learning en Microsoft Azure, gestión académica, modelo predictivo con algoritmo regresivo.

ABSTRACT

Machine Learning to predict the number of undergraduate students to enroll per subject in the academic period 2019-II at the National University of Santa

The objective of the research was to predict the number of undergraduate students to enroll by subject in the 2019-II academic period at the National University of Santa, using Machine Learning, for which the appropriate algorithm was identified to perform the predictive analysis. investigated, then an adequate platform for the implementation of Machine Learning algorithms in the cloud was identified, later the variables that intervene in the prediction of the number of undergraduate students to enroll per subject during an academic period were determined and subsequently a solution was developed. Machine Learning to predict that amount. When identifying the appropriate algorithm, it was observed that several algorithms can be used and then compared in the Machine Learning solution, Microsoft Azure was also identified as an adequate platform to implement the Machine Learning solution, later it was possible to determine ten predictor variables such as more appropriate for research and finally, using the algorithm, platform and selected variables, a Machine Learning solution was developed to predict the number of undergraduate students to enroll per subject; When using this solution with the values for the 2019-II academic period, it was observed that there is a similarity with the actual number of students who enrolled in each subject in said semester. Therefore, it is concluded that Machine Learning allows predicting the number of undergraduate students to enroll by subject in the 2019-II academic period at the National University of Santa.

By Br. Humberto Angel Ninaquispe Matame

Keywords: Machine Learning in Microsoft Azure, academic management, predictive model with regressive algorithm.

INTRODUCCIÓN

En cumplimiento de la ley universitaria 30220, las universidades, en cada periodo académico, deben llevar a cabo el proceso de matrícula de sus estudiantes, acción que conlleva a elaborar, con un tiempo de anticipación, la distribución de asignaturas (carga lectiva) y sus respectivos horarios de clases, en los cuales un alumno puede registrar matrícula. La ley indica también, que cada universidad cuenta con su propio estatuto y reglamentos en los cuales se establecen las normas y procedimientos que deben cumplirse en los procesos académicos.

En la Universidad Nacional del Santa, la distribución de la carga lectiva es un proceso por el cual se establecen, entre otras cosas, el número de grupos de teoría y/o de laboratorio para cada asignatura, respetando la normatividad vigente que regula el proceso, para ello los responsables de elaborar la carga lectiva deben proyectar una cantidad de alumnos a matricularse por cada asignatura buscando evitar, que posterior al proceso de matrícula, las asignaturas sufran algún tipo de reajustes; en el contexto académico, podemos definir proyectar cantidad de alumnos a matricular como la estimación o predicción que se espera tener de dicha cantidad.

Paja Dominguez, H. E. (2017), logra predecir el rendimiento académico mediante regresión y redes neuronales en los estudiantes de la Escuela Profesional de Ingeniería Estadística e Informática de la Universidad Nacional del Altiplano.

Yamao, E. (2018), logra predecir, a través de minería de datos, el rendimiento académico de los estudiantes del primer ciclo de la escuela profesional de Ingeniera de Computación y Sistemas de la Universidad de San Martin de Porras.

Zúñiga Ugalde, M. A. (2019), realiza un estudio técnico de las características y la oferta de servicios tecnológicos digitales basadas en tecnología de Machine Learning en la gran área metropolitana de Costa Rica en el periodo mayo 2017 a junio 2018.

Rodríguez Suárez, H. Z. (2018), hace un estudio de las herramientas basadas en IA Cloud y su aplicación en el desarrollo de las actividades académicas de la carrera de Telemática de la Universidad de Guayaquil.

Márquez Vera, C. (2015), logra predecir el fracaso y abandono escolar mediante técnica de minería de datos.

Menacho Chiok, C. H. (2017), predice el rendimiento académico aplicando técnicas de minería de datos.

Candia Oviedo, D. I. (2019), logra predecir el rendimiento académico de los estudiantes de la UNSAAC a partir de sus datos de ingreso utilizando algoritmos de aprendizaje automático.

Gamarra Gómez, F. (2019), establece un modelo basado en Machine Learning para el neurorendimiento académico de estudiantes universitarios.

Entonces, ¿Cómo predecir la cantidad de alumnos de pregrado a matricularse por asignaturas en el semestre 2019-II en la Universidad Nacional del Santa haciendo uso de Sistemas Inteligentes?, la hipótesis es que Machine Learning predice la cantidad de alumnos de pregrado a matricularse por asignatura en el periodo académico 2019-II en la Universidad Nacional del Santa, para lo cual se trazó el objetivo general de predecir la cantidad de alumnos de pregrado a matricularse por asignatura en el periodo académico 2019-II en la Universidad Nacional del Santa, haciendo uso de Machine Learning, siguiendo los objetivos específicos: a. Identificar el algoritmo de Machine Learning apropiado para realizar el análisis predictivo investigado, mediante el estudio y comparación de algoritmos. b. Identificar una adecuada plataforma de implementación de algoritmos de Machine Learning para realizar el análisis predictivo investigado, mediante el estudio y comparación de al menos tres plataformas. c. Determinar las variables predictoras que intervienen en la cantidad de alumnos de pregrado a matricularse por asignatura durante un periodo académico, mediante el estudio, análisis y selección de las más apropiadas. d. Elaborar una solución de Machine Learning para predecir la cantidad de alumnos de pregrado a matricularse por asignatura, mediante la implementación, entrenamiento y testeo de un modelo utilizando las variables, algoritmos y plataforma seleccionados, siendo justificada la investigación debido a que permite determinar de una manera más rápida y precisa la cantidad de alumnos de pregrado a matricularse por cada asignatura en un periodo académico, permitiendo reducir el número de reajustes en la carga lectiva posterior al proceso de matrícula.

Así, en el capítulo I: se describe el problema de la investigación, en el capítulo II se estudia el marco teórico, en el capítulo III se revisa el marco metodológico, en el capítulo IV se muestran los resultados y discusión, para finalmente, en el capítulo V se muestran las conclusiones y recomendaciones.

CAPÍTULO I

PROBLEMA DE INVESTIGACIÓN

1.1. Planteamiento y fundamentación del problema de investigación.

La Universidad Nacional del Santa, en cada periodo académico, debe elaborar la carga lectiva propuesta, actividad que consiste en la distribución de las asignaturas ofertadas en grupos de teoría y/o laboratorio, dependiendo de la cantidad de alumnos que se estima y/o proyecta van a matricularse; con esta carga lectiva propuesta se elaboran los horarios de clases para las diferentes escuelas profesionales, horarios que son utilizados por los alumnos para elegir los grupos de teoría y/o laboratorio al que desean matricularse teniendo en cuenta la normatividad vigente en el reglamento del estudiante de pregrado.

La Universidad Nacional del Santa, en el 2006, implementó el Sistema de Información Integral de Gestión Académica y Administrativa (SIIGAA) de modalidad desktop, el cual actualmente está a cargo de la Unidad de Desarrollo, Evaluación y Mantenimiento de Sistemas de Información (UDEMSI), dicho sistema cuenta con el módulo de matrícula desktop de pregrado y desde entonces la institución cuenta con una base de datos alimentada con la información de las matrículas de los alumnos en cada periodo académico, por otro lado también cuenta con información de las características de las escuelas profesionales, planes curriculares, asignaturas, convalidaciones e historial de notas de cada alumno de pregrado; desde el 2016, y en cumplimiento de la ley universitaria 30220, la institución estableció un nuevo Reglamento General y Reglamento del Alumno de Pregrado, normas que afectan en la toma de decisiones durante el proceso de registro de matrícula.

Desde el periodo académico 2018-II, la Universidad Nacional del Santa, implementó la aplicación web para el registro de matrícula en línea de los estudiantes de pregrado, sin embargo, uno de los factores que ha dificultado el uso de dicha aplicación es el hecho de no contar a tiempo con la carga lectiva propuesta y sus respectivos horarios; como se aprecia en la Tabla 01, en los dos últimos semestres, menos del 21% de los alumnos matriculados lo hicieron

por la aplicación web debido a que solo se habilitaron, para matrícula en línea, las especialidades que contaban con el registro en la base de datos de sus horarios completos, los demás alumnos debieron matricularse por el sistema de escritorio, el cual es operado por un docente en cada escuela profesional y deja al operador la tarea de dar cumplimiento de la normatividad vigente, en cambio, el sistema en línea, da cumplimiento de manera autónoma a dicha normatividad, como por ejemplo, lo referente a la prohibición de cruce de horario entre dos o más asignaturas (Art. 18 del Reglamento del Estudiante de Pregrado)

Tabla 01. Porcentaje de matriculados modo web

Periodo	Matricula regular	Matricula web	Porcentaje web (%)
2018-II	3817	301	07.89%
2019-I	4315	869	20.13%

Fuente: (UDEMSI, 2019)

En la actualidad el proceso de elaboración de la carga lectiva, según el Reglamento General de la Universidad Nacional del Santa en su artículo 32 inciso 9, está a cargo de los Departamentos Académicos; para esta tarea, es necesario que se haga una proyección de la cantidad de alumnos a matricularse, esta proyección se hace en base a la información con la que cuenta el Departamento, como número de matriculados en la asignatura en el semestre anterior o el número de alumnos que aprobaron el prerrequisito de cada asignatura. En el periodo académico 2018-II, que es el último periodo académico con las mismas características académicas que el periodo académico en estudio, los departamentos de la facultad de ingeniería en su proyección de la cantidad de alumnos matriculados para las asignaturas de los tres últimos ciclos pares y tomando en cuenta solo las asignaturas no electivas, han tenido un error porcentual absoluto [EPA] máximo de 156.52% y un error porcentual absoluto medio (MAPE) de 23.37%, tal como se observa en la Tabla 02.

Tabla 02. Cálculo del MAPE en asignaturas de Ingeniería en el 2018-II

Nro.	ESCUELA	ASIGNATURA	CICLO	PROYECCION	MATRICULA	APE
1	EPEF	MAQUINAS TERMICAS I	VI	45	41	9.76%
2	ENER	LABORATORIO DE ENERGIA I	VI	46	42	9.52%
3	ENER	TRANSFERENCIA DE CALOR	VI	45	44	2.27%
4	ENER	LABORATORIO DE ELECTRICIDAD	VI	51	47	8.51%
5	ENER	MAQUINAS ELECTRICAS	VI	52	48	8.33%
6	ENER	CONTROL AUTOMATICO	VI	46	54	14.81%
7	ENER	CIENCIA DE MATERIALES	VI	38	57	33.33%
8	ENER	CENTRALES HIDROELECTRICAS	VIII	30	32	6.25%
9	ENER	CAPT. Y ALMACEN. ENERGIA SOLAR	VIII	41	47	12.77%
10	ENER	SISTEMAS ELECTRICOS POTENCIA	VIII	44	47	6.38%
11	ENER	CENTRALES TERMOELECTRICAS	VIII	42	49	14.29%
12	ENER	TECNICAS CONSERV. USO EFIC. ENER	VIII	30	66	54.55%
13	ENER	GESTION ENERGETICA	X	30	27	11.11%
14	ENER	INDUSTRIALIZACION DEL GAS NATURAL	X	46	42	9.52%
15	ENER	PROYECTOS DE INGENIERIA	X	46	42	9.52%
16	ENER	SEMINARIO DE INVESTIGACION	X	46	47	2.13%
17	AGRO	INGENIERIA DE BIOPROCESOS	VI	50	33	51.52%
18	AGRO	REFRIGERACION Y CONG. DE PROD. AGROIND	VI	51	38	34.21%
19	AGRO	OPERACIONES UNITARIAS AGROIND. I	VI	43	41	4.88%
20	AGRO	NUTRICION Y PLANIFICACION	VI	50	42	19.05%
21	AGRO	ANALISIS INSTRUMENTAL DE PROD. AGROIND	VI	45	45	0.00%
22	AGRO	CONTROL Y AUTOMATIZACION DE PROC. IND.	VIII	45	49	8.16%
23	AGRO	LAB. DE OPERACIONES UNITARIAS	VIII	45	50	10.00%
24	AGRO	ING. DE PROCESOS AGROINDUSTRIALES II	VIII	45	52	13.46%
25	AGRO	FORMULACION Y EVAL. DE PROY. AGRO	X	47	42	11.90%
26	AGRO	SEMINARIO DE TESIS II	X	50	43	16.28%

27	EPIC	MECANICA DE SUELOS II	VI	58	47	23.40%
28	EPIC	MECANICA DE FLUIDOS I	VI	53	49	8.16%
29	EPIC	RESISTENCIA DE MATERIALES II	VI	49	52	5.77%
30	EPIC	PLANEAMIENTO URBANO Y REGIONAL	VIII	47	41	14.63%
31	EPIC	HIDROLOGIA	VIII	38	46	17.39%
32	EPIC	CONCRETO ARMADO II	VIII	53	50	6.00%
33	EPIC	ABASTECIMIENTO DE AGUA Y ALCANTAR	VIII	32	53	39.62%
34	EPIC	CONTRUCCIONES II	VIII	22	53	58.49%
35	EPIC	ANALISIS ESTRUCTURAL II	VIII	29	69	57.97%
36	EPIC	PROGRAMACION DE OBRAS	X	31	22	40.91%
37	EPIC	PROYECTOS DE INGENIERIA	X	60	30	100.00%
38	EPISI	MARKETING PARA TI	VI	59	23	156.52%
39	EPISI	SISTEMAS EXPERTOS	VI	42	30	40.00%
40	EPISI	SISTEMAS DE INFORMACION II	VI	43	31	38.71%
41	EPISI	INGENIERIA DE SOFTWARE	VI	39	33	18.18%
42	EPISI	COMUNICACIÓN DE DATOS	VI	55	48	14.58%
43	EPISI	GESTION DEL CONOCIMIENTO	VIII	39	40	2.50%
44	EPISI	REDES INFORMATICAS EMPRESARIALES II	VIII	31	45	31.11%
45	EPISI	SIMULACION DE SISTEMAS	VIII	36	57	36.84%
46	EPISI	PROYECTOS DE INVERSION EN TI	X	21	20	5.00%
47	EPISI	AUDITORIA DE SISTEMAS	X	33	34	2.94%
48	EPISI	SEMINARIO DE TESIS II	X	31	39	20.51%
MAPE						23.37%

Fuente: (UDEMSI, 2019)

Luego del proceso de matrícula, los departamentos académicos, elaboran y establecen lo que se denomina carga lectiva reajustada con el número real de alumnos matriculados por asignatura, actividad por la cual y en cumplimiento de la normatividad vigente en la Universidad Nacional del Santa, una asignatura puede sufrir algún tipo de los reajustes mostrados en la Tabla 03, esta nueva distribución afecta tanto en la estructura de los horarios de los alumnos como en la de los docentes. Para el periodo 2018-II, y respetando la normatividad de la Institución, el 56.25% de las asignaturas de los tres últimos

ciclos pares, tomando en cuenta solo las asignaturas no electivas que ofrecen los departamentos de la facultad de ingeniería, debieron sufrir al menos un tipo de reajuste, tal como se muestra en la tabla 04 y en el resumen de la Tabla 05.

Tabla 03. Lista de reajustes de una carga lectiva según normatividad

Norma	Descripción	Acción
Reglamento del estudiante de pregrado Art 68 inciso a	El proceso de enseñanza y aprendizaje en una asignatura se realiza de acuerdo al Sílabo y el número mínimo y máximo exigido de estudiantes matriculados: ... Mínimo ocho (8) y máximo 60 para una asignatura;	Si la asignatura tiene menos de ocho matriculados, se elimina la asignatura y se retiran las matrículas
Reglamento General Art. 81	El desarrollo de las clases teóricas se realiza con un máximo de 60 estudiantes por aula.	Si la asignatura tiene más de 60 alumnos matriculados se crea nuevo grupo de teoría
Reglamento del estudiante de pregrado Art 68	...Si el número de estudiantes es más de 60; la DEDA, de oficio los divide en dos grupos, en el orden alfabético de la matrícula y en forma equitativa.	
Reglamento General Art. 49	El número mínimo de estudiantes por grupo de práctica en laboratorio es de quince (15) ...	Según comparación del número redondeado superior de la cantidad proyectada entre 15 y de la cantidad real de matriculados, se realiza: 1. Si la proyección es mayor, se reduce el número de grupos de laboratorio. 2. Si la proyección es menor, se aumenta el número de grupos de laboratorio.
Reglamento del estudiante de pregrado Art. 68 inciso b	El proceso de enseñanza y aprendizaje en una asignatura se realiza de acuerdo al Sílabo y el número mínimo y máximo exigido de estudiantes matriculados: ... Para laboratorio, máximo quince (15); o de acuerdo a la capacidad del mismo ...	

Fuente: (Normatividad de la UNS, 2016)

Tabla 04. Reajuste de carga lectiva 2018-II según reglamento

Nro.	ESCUELA	ASIGNATURA	CICLO	PROYECCION	MATRICULA	REAJUSTE
1	EPEF	MAQUINAS TERMICAS I	VI	45	41	Sin cambio
2	ENER	LABORATORIO DE ENERGIA I	VI	46	42	Con reajuste
3	ENER	TRANSFERENCIA DE CALOR	VI	45	44	Sin cambio
4	ENER	LABORATORIO DE ELECTRICIDAD	VI	51	47	Sin cambio
5	ENER	MAQUINAS ELECTRICAS	VI	52	48	Sin cambio
6	ENER	CONTROL AUTOMATICO	VI	46	54	Sin cambio
7	ENER	CIENCIA DE MATERIALES	VI	38	57	Con reajuste
8	ENER	CENTRALES HIDROELECTRICAS	VIII	30	32	Con reajuste
9	ENER	CAPT. Y ALMACEN. ENERGIA SOLAR	VIII	41	47	Con reajuste
10	ENER	SISTEMAS ELECTRICOS POTENCIA	VIII	44	47	Con reajuste
11	ENER	CENTRALES TERMOELECTRICAS	VIII	42	49	Con reajuste
12	ENER	TECNICAS CONSERV. USO EFIC. ENER	VIII	30	66	Con reajuste
13	ENER	GESTION ENERGETICA	X	30	27	Sin cambio
14	ENER	INDUSTRIALIZACION DEL GAS NATURAL	X	46	42	Con reajuste
15	ENER	PROYECTOS DE INGENIERIA	X	46	42	Con reajuste
16	ENER	SEMINARIO DE INVESTIGACION	X	46	47	Sin cambio
17	AGRO	INGENIERIA DE BIOPROCESOS	VI	50	33	Con reajuste
18	AGRO	REFRIGER. Y CONG. DE PROD. AGRO	VI	51	38	Con reajuste
19	AGRO	OPERACIONES UNITARIAS AGROIND. I	VI	43	41	Sin cambio
20	AGRO	NUTRICION Y PLANIFICACION	VI	50	42	Con reajuste
21	AGRO	ANALISIS INSTRUM. DE PROD AGRO	VI	45	45	Sin cambio
22	AGRO	CONTROL Y AUTOMAT. DE PROC IND.	VIII	45	49	Con reajuste
23	AGRO	LAB. DE OPERACIONES UNITARIAS	VIII	45	50	Con reajuste
24	AGRO	ING. DE PROC. AGROINDUSTRIALES II	VIII	45	52	Con reajuste
25	AGRO	FORMULACION Y EVAL. DE PROY. AGRO	X	47	42	Con reajuste

26	AGRO	SEMINARIO DE TESIS II	X	50	43	Con reajuste
27	EPIC	MECANICA DE SUELOS II	VI	58	47	Sin cambio
28	EPIC	MECANICA DE FLUIDOS I	VI	53	49	Sin cambio
29	EPIC	RESISTENCIA DE MATERIALES II	VI	49	52	Sin cambio
30	EPIC	PLANEAMIENTO URBANO Y REGIONAL	VIII	47	41	Con reajuste
31	EPIC	HIDROLOGIA	VIII	38	46	Con reajuste
32	EPIC	CONCRETO ARMADO II	VIII	53	50	Sin cambio
33	EPIC	ABASTECIMIENTO DE AGUA Y ALCANTA	VIII	32	53	Con reajuste
34	EPIC	CONSTRUCCIONES II	VIII	22	53	Con reajuste
35	EPIC	ANALISIS ESTRUCTURAL II	VIII	29	69	Con reajuste
36	EPIC	PROGRAMACION DE OBRAS	X	31	22	Con reajuste
37	EPIC	PROYECTOS DE INGENIERIA	X	60	30	Con reajuste
38	EPISI	MARKETING PARA TI	VI	59	23	Con reajuste
39	EPISI	SISTEMAS EXPERTOS	VI	42	30	Con reajuste
40	EPISI	SISTEMAS DE INFORMACION II	VI	43	31	Sin cambio
41	EPISI	INGENIERIA DE SOFTWARE	VI	39	33	Sin cambio
42	EPISI	COMUNICACIÓN DE DATOS	VI	55	48	Sin cambio
43	EPISI	GESTION DEL CONOCIMIENTO	VIII	39	40	Sin cambio
44	EPISI	REDES INFORMAT.EMPRESARIALES II	VIII	31	45	Sin cambio
45	EPISI	SIMULACION DE SISTEMAS	VIII	36	57	Con reajuste
46	EPISI	PROYECTOS DE INVERSION EN TI	X	21	20	Sin cambio
47	EPISI	AUDITORIA DE SISTEMAS	X	33	34	Sin cambio
48	EPISI	SEMINARIO DE TESIS II	X	31	39	Sin cambio

Fuente: (UDEMSI, 2019)

Tabla 05. Resumen de asignaturas procesadas 2018-II

Tipo	Cantidad	Porcentaje (%)
Con reajuste	27	56.25%
Sin reajuste	21	43.75%
Total	48	100.00%

1.2. Antecedentes de la investigación.

Título: Predicción de rendimiento académico mediante regresión y redes neuronales en los estudiantes de la escuela profesional de ingeniería estadística e informática de la universidad nacional del altiplano - puno, 2015

Autor: Paja Dominguez, H. E.

Año: 2017

Conclusiones más relevantes:

Tras aplicar técnicas de redes neuronales y modelos de regresión, le dio como resultado que las técnicas de redes neuronales tienen menor grado de error frente a la técnica de regresión; así, obtuvo que los modelos de RNAs mostraron mejor desempeño con promedios de 0.041635 de error en el modelo y 0.0008120 error en predicción frente a los modelos de regresión con 0.535218 error de modelo y 0.1880641 de error para las predicciones, de acuerdo al objetivo general de este trabajo de investigación se afirmó las RNA's es la mejor técnica en predicción; puesto que tienen una diferencia de 0.534 error en modelo y 0.1307 error en predicción respecto a la regresión en los estudiantes de la Escuela Profesional de Ingeniería Estadística e Informática de la Universidad Nacional del Altiplano - Puno 2015. (Paja Dominguez, 2017)

Título: Predicción del rendimiento académico mediante minería de datos en estudiantes del primer ciclo de la Escuela Profesional de Ingeniería de Computación y Sistemas, Universidad de San Martín de Porres, Lima-Perú

Autor: Yamao, E.

Año: 2018

Conclusiones más relevantes:

Tras aplicar tres técnicas de minería de datos para realizar la predicción de rendimiento académico, el algoritmo de árbol de decisiones es el que obtuvo los mejores resultados con una exactitud de predicción de 82.87%; así mismo, la facilidad de interpretación de los resultados de la técnica de árbol de decisiones lo convierten en la mejor técnica a utilizar para este tipo de estudio. (Yamao, 2018)

Título: Estudio técnico de las características y la oferta de servicios tecnológicos digitales basadas en tecnología de Machine Learning en la gran área metropolitana de Costa Rica en el período mayo 2017 a junio 2018.

Autor: Zúñiga Ugalde, M. A.

Año: 2019

Conclusiones más relevantes:

Logran concluir que el estado del Machine Learning se encuentra en una fase inicial y de crecimiento, sin llegar a su completo desarrollo, según las características encontradas en el nivel técnico, funcional y organizacional.

Con respecto de los estilos de Machine Learning, se concluye que el estilo más utilizado es el supervisado, pues permite un seguimiento y control de los resultados que el modelo arroje y así lograr el objetivo que la empresa estipule.

Recomienda que la infraestructura basada en la nube brinda cualidades convenientes como un alto poder computacional como CPU, GPU y almacenamiento elástico, que permite incluir más información en las bases de datos, sin necesidad de hacer cambios de hardware, pues las grandes granjas de servidores mantenidas por grandes empresas como Google, Amazon o Microsoft generan un gran respaldo, además de que se evita el mantenimiento de equipo tecnológico, destacando en su trabajo a Microsoft Azure ML Studio uno de los framework cubre la mayoría de las tareas relacionadas con la ML. (Zúñiga Ugalde, 2019)

Título: Estudio de herramientas basadas en IA CLOUD y su aplicación en el desarrollo de las actividades académicas de la carrera de Telemática de la Universidad de Guayaquil.

Autor: Rodríguez Suárez, H. Z.

Año: 2018

Conclusiones más relevantes:

Concluyó que las herramientas IA Cloud analizadas en su investigación poseen medios accesibles para los estudiantes de la carrera de Ingeniería Telemática que están acordes a los requerimientos académicos, orientados a tecnología utilizada actualmente.

A partir del análisis de las principales herramientas de IA Cloud, seleccionó la plataforma de Microsoft Azure pues cuenta con mayor nivel de utilidad. (Rodríguez Suarez, 2018)

Título: Predicción del fracaso y abandono escolar mediante técnica de minería de datos

Autor: Márquez Vera, C.

Año: 2015

Conclusiones más relevantes:

Durante su pre-procesado de las variables, y con el uso de la aplicación Weka, realizó un análisis y selección de los mejores atributos para predecir la alta dimensionalidad de los datos. Y obtuvo mejores resultados usando solamente los mejores atributos en lugar de todos los disponibles. (Márquez Vera, 2015)

Título: Predicción del rendimiento académico aplicando técnicas de minería de datos

Autor: Menacho Chiok, C. H.

Año: 2017

Conclusiones más relevantes:

En su trabajo de demostración de eficacia de las técnicas de minería de datos logra obtener mayor precisión con la técnica de red naive de bayes (71% de correcta clasificación), pero para lograr esto, selecciona las variables predictoras que sean estadísticamente significativas haciendo uso del programa Weka. (Menacho Chiok, 2017)

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

Autor: Candía Oviedo, D. I.

Año: 2019

Conclusiones más relevantes:

El algoritmo de árboles de decisión, fue el algoritmo que tuvo mejor performance para la predicción del rendimiento académico de los ingresantes en los primeros semestres a la UNSAAC con un 69% de predicción, el segundo

algoritmo con mejor performance fue el algoritmo de regresión logística con un 68%. (Candia Oviedo, 2019)

Título: Modelo basado en Machine Learning para el neurorendimiento académico de estudiantes universitarios.

Autor: Gamarra Gómez, F.

Año: 2019

Conclusiones más relevantes:

El mejor modelo predictivo es la regresión lineal con una métrica de coeficiente de determinación de 0.988.

Para la aplicación de su trabajo utiliza el software de Microsoft Azure, el cual se encuentra en la nube. (Gamarra Gómez, 2019)

1.3. Formulación del problema de investigación.

En la presente investigación se planteó la siguiente pregunta:

¿Cómo predecir la cantidad de alumnos de pregrado a matricularse por asignaturas en el semestre 2019-II en la Universidad Nacional del Santa haciendo uso de Sistemas Inteligentes?

1.4. Delimitación del estudio.

La presente investigación está delimitada para:

- Asignaturas ofertadas en el proceso de matrícula de los alumnos de pregrado de la Universidad Nacional del Santa de la ciudad de Nuevo Chimbote, durante el periodo académico 2019-II.
- Asignaturas de los últimos ciclos de cada escuela, debido a que las especialidades han cambiado sus planes curriculares y con ellos sus asignaturas, no contando con mucha información histórica.

1.5. Justificación e importancia de la investigación.

El desarrollo de la presente investigación se justifica por los siguientes criterios:

- El presente trabajo le va a permitir a la institución contar de manera más precisa y rápida con la cantidad de alumnos de pregrado a matricularse por cada asignatura en un periodo académico, demostrando que en la proyección de cantidad de matriculados no solo se debe tener en cuenta el

histórico de los alumnos matriculados en los últimos periodos académicos, y la cantidad de alumnos que han aprobado los prerrequisitos en el periodo académico anterior, sino que hay otros factores que no son tomados en cuenta.

- Establecido el modelo para el análisis predictivo de la cantidad de alumnos de pregrado a matricularse por asignatura, este mismo podrá servir para establecer un modelo predictivo para el ámbito de posgrado, respetando los criterios y normatividad que varía respecto al ámbito de pregrado.
- La investigación servirá de guía para implementar y entender un modelo predictivo de una manera sencilla, sin necesidad de tener un conocimiento profundo sobre Machine Learning y sus algoritmos.

1.6. Objetivos de la investigación: General y específicos.

El objetivo general es: Predecir la cantidad de alumnos de pregrado a matricularse por asignatura en el periodo académico 2019-II en la Universidad Nacional del Santa, haciendo uso de Machine Learning y, los objetivos específicos son:

- a. Identificar el algoritmo de Machine Learning apropiado para realizar el análisis predictivo investigado, mediante el estudio y comparación de algoritmos.
- b. Identificar una adecuada plataforma de implementación de algoritmos de Machine Learning para realizar el análisis predictivo investigado, mediante el estudio y comparación de al menos tres plataformas.
- c. Determinar las variables predictoras que intervienen en la cantidad de alumnos de pregrado a matricularse por asignatura durante un periodo académico, mediante el estudio, análisis y selección de las más apropiadas.
- d. Elaborar una solución de Machine Learning para predecir la cantidad de alumnos de pregrado a matricularse por asignatura, mediante la implementación, entrenamiento y testeado de un modelo utilizando las variables, algoritmos y plataforma seleccionados.

CAPÍTULO II MARCO TEÓRICO

2.1. Fundamentos teóricos de la investigación

APRENDIZAJE AUTOMATIZADO (MACHINE LEARNING):

Primero debemos entender que aprender significa que se genera nuevo conocimiento a partir de las observaciones y que este conocimiento se utiliza para lograr objetivos definidos. (Kramer, 2016)

En el ámbito de la informática y específicamente en el campo de la inteligencia artificial se puede definir al aprendizaje automático como sistemas o algoritmos que se mejoran a través de la experiencia de datos sin depender de una programación explícita, es decir aprenden sin necesidad de ser programados constantemente. (Sotiropoulos y Tsihrintzis, 2017)

• **Características del Machine Learning**

- ✓ Capacidad para adoptar y modificar automáticamente el comportamiento en función de las necesidades de los usuarios. Por ejemplo, correo electrónico personalizado o noticias.
- ✓ Capacidad para descubrir nuevos conocimientos en base de datos de hechos.
- ✓ Capacidad para ayudar a los humanos y reemplazar tareas monótonas, que requieren algo de inteligencia.
- ✓ Capacidad de generar "visión" operando iterativamente en datos y aprendiendo de los errores.

(Kashyap, 2017)

• **Algoritmos Machine Learning**

La característica fundamental de estos algoritmos es su extraordinaria capacidad para extraer patrones que se escapan a la mayoría de métodos aplicados históricamente. Resultan, en gran medida, eficaces para abordar problemas de segmentación, clasificación e incluso predicción.

Sin embargo, para crear y alimentar las variables que formarán parte de los algoritmos y poder ponderarlas adecuadamente durante el proceso de

entrenamiento, aún resulta imprescindible contar con un conocimiento previo del contexto y de las dinámicas de los procesos de negocio. (Kuperman et al., 2019)

Según el tipo de variable de respuesta que tenemos en los datos de entrenamiento, el aprendizaje se puede clasificar en:

- ✓ Aprendizaje supervisado: Los datos contienen una variable de respuesta (también llamada etiqueta) o es posible generar una.
- ✓ Aprendizaje no supervisado: Las etiquetas no están disponibles.
- ✓ Aprendizaje semisupervisado: No hay etiquetas para todas las observaciones en el conjunto de datos.
- ✓ Aprendizaje por refuerzo: Los datos para el aprendizaje no están disponibles o se actualizarán rápidamente con el tiempo

(Ramasubramanian y Singh, 2017)

APRENDIZAJE AUTOMATIZADO EN LA NUBE

Actualmente ya no es necesario contar con un equipo de cómputo potentes cuyo problema para el procesamiento de gran información solía ser la capacidad de procesamiento, en la actualidad el análisis de los datos puede hacerse en la nube, aunque existe como contra la necesidad de elegir correctamente la configuración y el costo de usar estos servicios en la nube.

Los principales proveedores de la nube son los siguientes:

- Google Cloud Platform (GCP)
- Microsoft Azure
- Amazon Web Services (AWS)

(Ayyadevara, 2018)

Cada proveedor tiene una lista variable de servicios con diferentes niveles de personalización. Por ejemplo, algunos servicios emitirán el modelo final que se puede implementar en un entorno local o en el borde. Algunos proveedores solo exponen un punto final API que es accesible solo en Internet público. (Gartner, 2020)

- **Cuadrante mágico (MQ) de Gartner**

El informe MQ de Gartner, de febrero del 2020, analiza el enfoque de consumir servicios de IA basados en la nube ofrecidos por los proveedores.



Gráfico 1. Cuadrante de Gartner para servicios de desarrollador de IA en la nube

Fuente: (Gartner, 2019)

MATRÍCULA DE ASIGNATURAS

La matrícula es el acto de inscripción formal y voluntario en un periodo lectivo, derivado de la admisión o promoción del estudiante en la asignatura y/o ciclo o año académico correspondiente. Por la matrícula vigente se adquiere la condición de estudiante de pregrado, comprometiéndose a cumplir la Ley Universitaria, el Estatuto y los reglamentos de la UNS. (art. 15 Reglamento del alumno de pregrado de la UNS)

Según indica el reglamento, el estudiante debe matricularse en la fecha que la UNS señale, siendo el único responsable de determinar por dicho acto los cursos que, entre los debidamente autorizados desarrollará, durante el ciclo

y/o año académico correspondiente; estos cursos (asignaturas) están limitados por prerrequisitos que deben de cumplirse, así como otros parámetros requeridos para que un alumno pueda registrar matrícula en una asignatura.

Por otro lado, el reglamento en su artículo 46 señala que: El estudiante seleccionará las asignaturas a llevar y los grupos de teoría o práctica, según corresponda; estos grupos se asignan en la carga lectiva provisional misma que es establecida desde los departamentos académicos.

- **Consideraciones pre y post matrícula:**

- a) Elaboración de la carga lectiva inicial**

Actividad a cargo de los Departamentos académicos, y que consiste en determinar las asignaturas a cargo de un docente, estas asignaturas se miden en horas lectivas que están divididas en horas teóricas y prácticas.

Para realizar esta actividad los departamentos cuentan con la siguiente información:

- Número de alumnos matriculados en el semestre anterior
- Número de alumnos que aprobaron el pre requisito en el semestre anterior
- Número de alumnos que han desaprobado la asignatura en el semestre anterior.
- Número de horas lectivas que debe tener cada docente dependiendo de su categoría, encargatura o tipo de contrato.
- Número máximo de matriculados para grupos de teoría y laboratorio (con el reglamento actual son 50 alumnos para teoría y 15 para practico cuando hay necesidad de laboratorio)

Con estos criterios el departamento determina el número de grupos teóricos y de practica que tendrá cada asignatura y cuales le corresponde a cada docente.

- b) Elaboración de la carga horaria**

Proceso por el cual se establece el día, hora de inicio y fin de una clase de estudio de cada asignatura, por ciclo y especialidad.

Esta actividad lo realiza una comisión de elaboración de horarios con la carga lectiva inicial que es alcanzado por los departamentos académicos. (Art. 77 del Reglamento del Estudiante de Pregrado, 2017)
Los horarios son elaborados por ciclo académico para cada especialidad, en cada horario se identifica los grupos tanto de teoría como de práctica.

c) Proceso de registro de matrícula

Es el proceso por el cual el alumno decide en que asignaturas inscribirse para el periodo académico vigente.

Entre los requisitos para poder registrar matrícula, debe cumplirse que no debe existir cruce de horarios entre las asignaturas seleccionadas para matrícula, sea en sus horas teóricas o prácticas. (Art. 18 del Reglamento del Estudiante de Pregrado, 2017)

d) Elaboración de carga lectiva reajustada

Es el proceso que se ejecuta culminada la matrícula, por medio de este proceso los departamentos elaboran la carga lectiva definitiva con el número real de alumnos matriculados por asignatura. (Art. 74 del Reglamento del Estudiante de Pregrado, 2017)

En la elaboración de la carga lectiva definitiva o reajustada existen criterios para la división de un grupo teórico o de laboratorio, así como de la anulación de una asignatura dependiendo del número mínimo de alumnos matriculados. (Art. 68 y 69 del Reglamento del Estudiante de Pregrado, 2017)

Debido a una incorrecta decisión de realizar la carga lectiva inicial para matrícula puede suceder estas acciones:

- Si un grupo de teoría tiene más de 60 alumnos, el grupo de teoría se divide en dos grupos con la misma cantidad de alumnos en promedio.
- Si la asignatura no tiene el número mínimo de alumnos matriculados, la asignatura es descartada y se anula la matrícula de los alumnos en esa asignatura.

2.2. Marco conceptual

Análisis predictivo:

El análisis predictivo es el proceso de extracción de información de grandes conjuntos de datos con el fin de hacer predicciones y estimaciones sobre los resultados futuros. (Larose y Larose, 2015)

Machine Learning:

El aprendizaje automático (Machine Learning) es una rama de la inteligencia artificial. Usando la computación, se diseñan sistemas que pueden aprender de los datos de una manera que se entrena. Los sistemas pueden aprender y mejorar con la experiencia, y con el tiempo, refinar un modelo que se puede usar para predecir los resultados de las preguntas basadas en el aprendizaje anterior. (Bell, 2015)

Matricula:

La matrícula es el acto de inscripción formal y voluntario en un periodo lectivo, en las asignaturas de un ciclo o año académico correspondiente, por la matrícula regular el individuo obtiene la condición de alumno de la UNS. (Reglamento del alumno de pregrado de la UNS)

CAPÍTULO III MARCO METODOLÓGICO

En el presente capítulo se muestra el marco metodológico teniendo en cuenta que la formulación del problema es:

¿Cómo predecir la cantidad de alumnos de pregrado a matricularse por asignaturas en el semestre 2019-II en la Universidad Nacional del Santa haciendo uso de Sistemas Inteligentes?

En tal sentido,

3.1. Hipótesis central de la investigación.

H₀: Machine Learning no permite predecir la cantidad de alumnos de pregrado a matricularse por asignatura en el periodo académico 2019-II en la Universidad Nacional del Santa

H₁: Machine Learning permite predecir la cantidad de alumnos de pregrado a matricularse por asignatura en el periodo académico 2019-II en la Universidad Nacional del Santa

3.2. Variables e indicadores de la investigación.

Matriz de operacionalización de variables.

Variable	Definición Conceptual	Indicador	Unidad de medida	Técnica	Instrumento
Independiente: Machine Learning	Solución que pretende predecir tendencias y patrones de comportamiento	Algoritmos a comparar en la solución	Cantidad	Modelado	Plataforma Machine Learning
		Coefficiente de determinación de la solución	Cantidad	Medición	Plataforma Machine Learning
		Fiabilidad del modelo predictivo final	%	Medición	Software estadístico

Dependiente: Cantidad de alumnos de pregrado a matricularse por asignatura en el periodo académico 2019-II en la UNS	Número de alumnos de pregrado que se matriculan en una asignatura en el periodo académico 2019-II	Número de matriculados en el modelo predictivo	Cantidad	Medición	Hoja de datos
		Número real de matriculados	Cantidad	Medición	Hoja de datos
		Error porcentual absoluto medio	%	Medición	Hoja de datos

3.3. Métodos de la investigación.

- Se aplicó el método analítico en la revisión del estado del arte y las variables problemáticas encontradas.
- Para el desarrollo del objetivo específico uno: *“Identificar el algoritmo de Machine Learning apropiado para realizar el análisis predictivo investigado, mediante el estudio y comparación de algoritmos”*, se aplicó el método de análisis hipotético deductivo.
- En el caso del objetivo específico dos: *“Identificar una adecuada plataforma de implementación de algoritmos de Machine Learning para realizar el análisis predictivo investigado, mediante el estudio y comparación de al menos tres plataformas”*, se aplicó el método hipotético deductivo.
- Para el desarrollo del objetivo específico tres: *“Determinar las variables predictoras que intervienen en la cantidad de alumnos de pregrado a matricularse por asignatura durante un periodo académico, mediante el estudio, análisis y selección de las más apropiadas”* se aplicó el método cualitativo sistémico.
- Para el desarrollo del objetivo específico cuatro: *“Elaborar una solución de Machine Learning para predecir la cantidad de alumnos de pregrado a matricularse por asignatura, mediante la implementación, entrenamiento y testeo de un modelo utilizando las variables, algoritmos y plataforma seleccionados”* se aplicó el método cuantitativo de experimentación y de simulación

- Se aplicó el método inductivo para el análisis de los resultados y la contrastación de la hipótesis.

3.4. Diseño o esquema de la investigación.

- Tipo de Estudio:
Cuasiexperimental, con caso de estudio.
- Diseño del estudio:
Correlacional.

RG1 O1 X O2

RG1: Grupo de asignaturas ofertadas para el proceso de matrícula en el periodo 2019-II.

O1: Observación y medición de la cantidad de alumnos de pregrado matriculados por asignatura real sin aplicar modelo predictivo.

O2: Observación y medición de la cantidad de alumnos de pregrado a matricularse por asignatura aplicando el modelo predictivo.

3.5. Población y muestra.

- Objeto de estudio
Cantidad de alumnos de pregrado matriculados por asignatura
- Población
Asignaturas ofertadas por periodo académico para matrícula regular de pregrado.
- Muestra
Caso de estudio:
Asignaturas ofertadas para matrícula regular en el periodo académico 2019-II en la Universidad Nacional del Santa de los últimos ciclos pares de las especialidades cuyo plan curricular tiene más de 4 años vigentes.
(Muestreo intencional o por conveniencia)

- Unidad de Análisis

Predicción de la cantidad de alumnos de pregrado a matricularse por asignatura en el periodo académico 2019-II

3.6. Actividades del proceso investigativo

- a. Revisión de bibliografía de los diferentes algoritmos para la elaboración de un modelo predictivo para Machine Learning; se va a evaluar los diferentes algoritmos supervisados y elegir el que más se adecue para nuestro caso de estudio.
- b. Revisión bibliográfica acerca de las plataformas en la nube para uso de algoritmos de Machine Learning; se va a evaluar las propiedades de al menos tres plataformas en la nube y elegir la más apropiada para nuestro caso de estudio.
- c. Revisión de las variables que pueden intervenir en la formulación de nuestro modelo predictivo; se van a analizar y seleccionar las variables más influyentes para nuestro caso de estudio.
- d. Recolección y preparación de la información necesaria para la elaboración del modelo predictivo en base a las variables seleccionadas. Elaboración e implementación del modelo predictivo en la plataforma seleccionada, comparación del modelo con los algoritmos seleccionados para encontrar la mejor solución predictiva y se seleccionará el que muestre mayor coeficiente de determinación, luego se va a testar el modelo utilizando los valores para el semestre 2018-II.
- e. Validación y aplicación del modelo predictivo al periodo académico 2019-II para obtener la predicción de la cantidad de alumnos de pregrado a matricularse por asignatura en la Universidad Nacional del Santa y contrastación de la hipótesis.

3.7. Técnicas e instrumentos de la investigación.

Técnica	Instrumento
Análisis de la literatura sobre algoritmos y plataformas para Machine Learning	Material bibliográfico, revistas y artículos
Encuesta a juicio de experto [para la selección de algoritmos, plataforma Machine Learning y Variables]	Cuestionarios
Tabulación de la información obtenida	Registros

3.8. Procedimiento para la recolección de datos (Validación y confiabilidad de los instrumentos).

Técnica	Instrumento	Validez	Confiabilidad
Análisis de la literatura	Material bibliográfico, revistas y artículos	100%	100%
Encuesta [algoritmos]	Cuestionario	90%	70%
Encuesta [plataforma]	Cuestionario	90%	70%
Encuesta [variables]	Cuestionario	90%	90%
Tabulación de información	Registro	100%	100%

3.9. Técnicas de procesamiento y análisis de los datos.

3.9.1. Procesamiento de datos

Se realizó a través de estadística básica, así como la utilización del Alfa de Crombach para medir la fiabilidad de los datos obtenidos, con el uso del software SPSS.

3.9.2. Procesamiento de resultados.

Para el procesamiento, análisis e interpretación de resultados se realizó el análisis estadístico de datos apareados mediante Estadística Correlacional, con pruebas no paramétricas de muestras relacionadas.

CAPÍTULO IV

RESULTADOS Y DISCUSIÓN

A continuación, se presentan los resultados y discusión por cada objetivo específico, luego de haber desplegado una solución Machine Learning para predecir el número de alumnos de pregrado matriculados en una asignatura.

4.1. Identificar el algoritmo de Machine Learning apropiado para realizar el análisis predictivo investigado, mediante el estudio y comparación de algoritmos

Para la identificación del algoritmo se realizaron las siguientes actividades:

- Checklist sobre la información requerida, que permitió determinar el tipo de algoritmo a utilizar.
- Validez y aplicación del cuestionario sobre los algoritmos Machine Learning que se podrían usar en el modelo predictivo.
- Prueba de fiabilidad del cuestionario, para determinar si el instrumento produce resultados consistentes y coherentes.

a) Checklist sobre la información requerida

Tabla 06. Checklist modelos y técnicas ML

Información		Revisado		Análisis
		Si	No	
01	Tipos de Técnicas para Machine Learning	X		Debido a que contamos con un conjunto de datos históricos ya clasificados para el entrenamiento, se usará el tipo Supervisado .
02	Tipos de algoritmos Supervisados	X		Debido a que se va a predecir un valor, los de tipo Regresión son los más adecuados

03	Algoritmos de regresión	X	Por tiempo de entrenamiento y linealidad podemos usar Regresión lineal , mientras que por precisión podría ser Red neuronal
----	-------------------------	---	---

Interpretación: El resultado del Checklist nos indicó que los algoritmos de regresión son los más indicados para nuestro caso de estudio, para ser más precisos nos indicó que la Regresión lineal y Red neuronal son los más indicados.

- b) Se validó y aplicó un cuestionario a un grupo de docentes de la UNS
La validez del instrumento se realizó a través de juicio de experto
(véase formato en anexos)

Se aplicó una encuesta a 10 docentes de la escuela de Ingeniería de Sistemas e Informática que de alguna manera hay aplicado los algoritmos de machine Learning, para este objetivo se utilizó el ítem 01 de la encuesta Machine Learning (anexo 03) en la que se pide al docente que elija tres algoritmos que podrían lograr la predicción del número de alumnos a matricularse en una asignatura.

Luego de obtener los resultados de la encuesta, se agruparon por algoritmos seleccionados, como se muestra en la tabla siguiente.

Tabla 07. Algoritmos según docentes UNS

Nro.	Algoritmo	Cantidad
01	Agrupación	1
02	Árbol de decisiones	8
03	Bayesiana lineal	0
04	Máquina de vectores	1
05	Red neuronal	5
06	Redes bayesianas	4

07	Regresión logística	3
08	Regresión lineal	8
09	Selva de decisión	0
10	Otros	0

Interpretación: La tabla anterior nos mostró que según los docentes encuestados sobre los algoritmos Machine Learning, los algoritmos de Árbol de decisiones y Regresión lineal fueron los más sugeridos para dar solución a nuestro objeto de predicción.

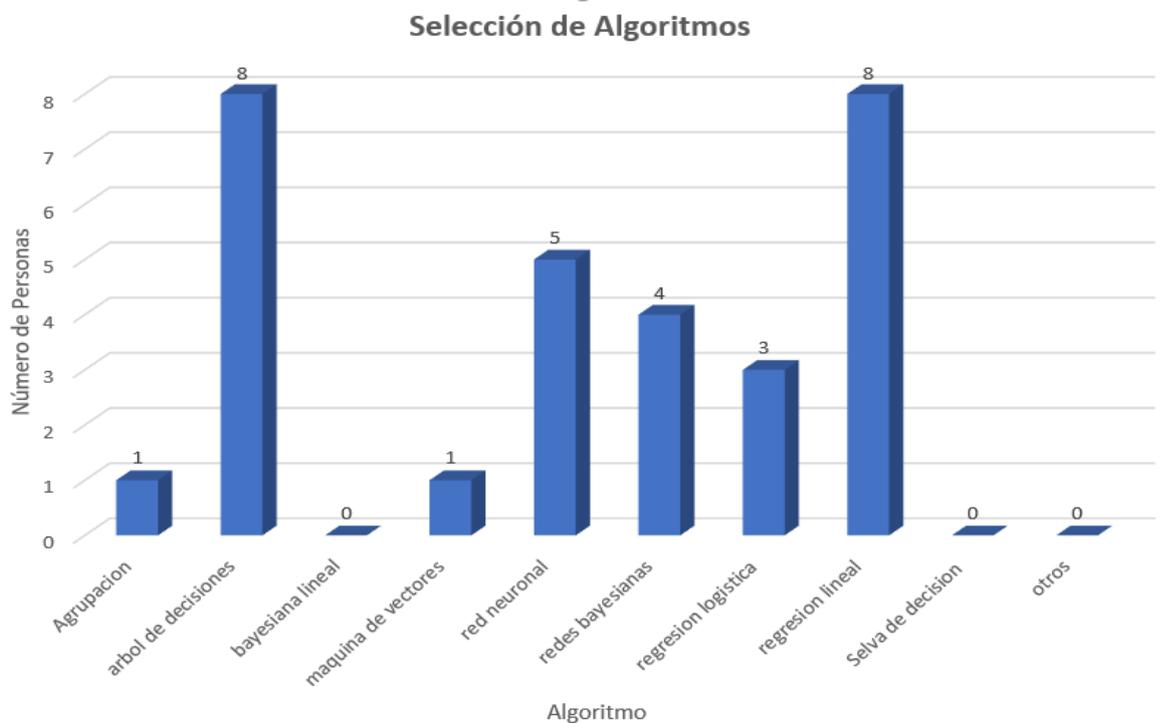


Gráfico 2. Algoritmos sugeridos por docentes UNS

Interpretación: El gráfico anterior nos mostró que los docentes de la UNS sugieren, que se puede utilizar Arboles de decisiones o Regresión Lineal, tal como se puede apreciar en la frecuencia de los algoritmos seleccionados, sin embargo, otros algoritmos como red neuronal o redes bayesianas también tienen buena aceptación.

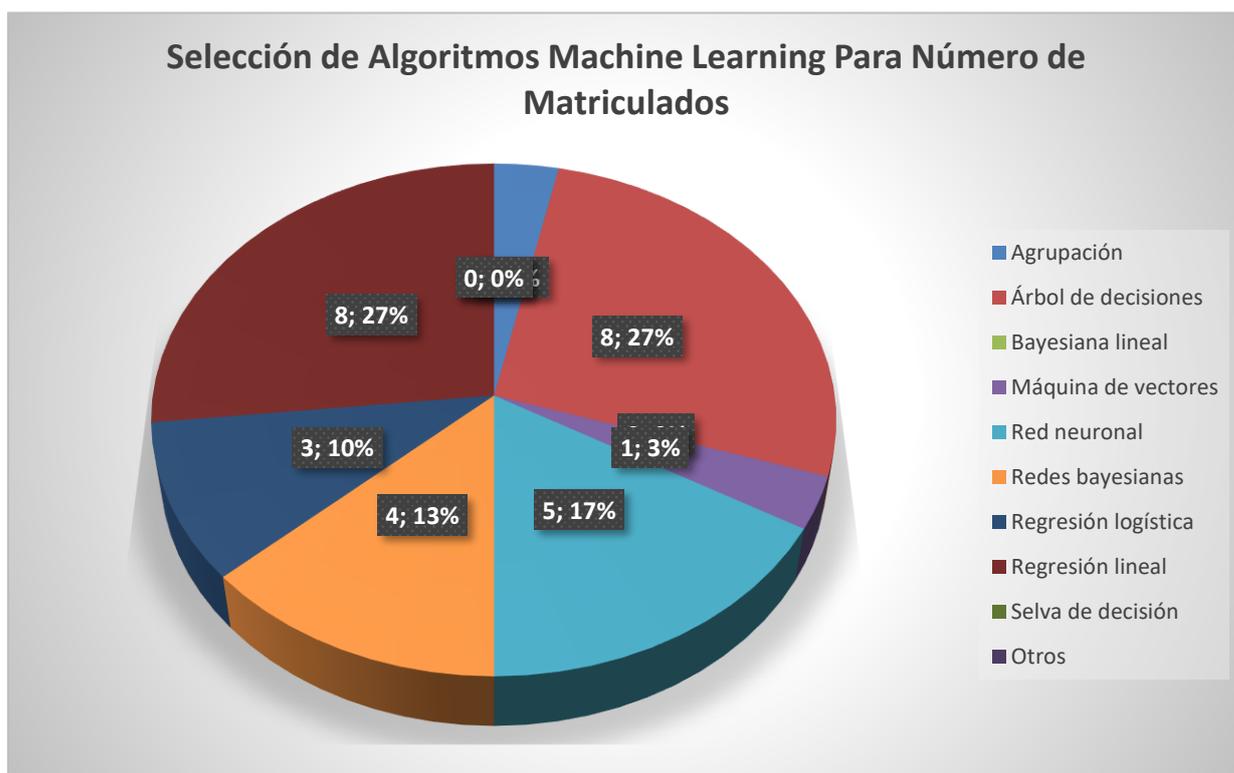


Gráfico 3. Gráfico de torta para algoritmos seleccionados

Interpretación: Del gráfico se aprecia que el 27% de las decisiones dicen que tanto árbol de decisiones y regresión lineal son algoritmos adecuados para lograr la predicción del número de alumnos a matricularse en una asignatura, así mismo, podemos apreciar que 8 de 10 docentes indican que estos dos algoritmos son apropiados para nuestro caso de estudio.

- c) Con la aplicación SPSS Vr. 25 se pudo demostrar que la fiabilidad de la encuesta es de 0.962 lo que involucra el ítem 01, ítem que se utilizó para este objetivo.

Interpretación: Con el resultado de la fiabilidad para el instrumento (encuesta) que nos dio una significancia del 0.962 podemos decir que existe un grado de fiabilidad muy alta para el instrumento utilizado.

Discusión: Concluido el análisis, observamos que no solo un algoritmo es determinado como el indicado para poder utilizarlo en nuestro modelo predictivo, así como también se puede observar un sesgo en cuanto al conocimiento de los algoritmos machine Learning, pues se introdujo valores no reales en el nombre de los algoritmos y esos fueron seleccionados, por lo tanto, se pretende utilizar o comparar, al menos, los algoritmos Árbol de decisiones, Regresión Lineal y red neuronal para determinar cuál algoritmo nos da mayor precisión al momento de predecir el número de alumnos a matricularse en una asignatura; así como lo hace Paja Domínguez (2017) quien concluyó que con red neuronal obtuvo un mejor desempeño en su modelo predictivo para el rendimiento académico; por otro lado, Yamao (2018), determina que regresión y arboles de decisiones le permitieron encontrar mejor predicción en rendimiento académico y recomienda también el uso de redes neuronales.

4.2. Identificar una adecuada plataforma de implementación de algoritmos de Machine Learning para realizar el análisis predictivo investigado, mediante el estudio y comparación de al menos tres plataformas

Para identificar la adecuada plataforma Machine Learning se realizó:

- Evaluación de Características de infraestructura para machine Learning
- Comparación de plataformas en la nube para Machine Learning
- Encuesta a docentes sobre que plataforma usar para nuestro objeto de estudio.
- Evaluación de fiabilidad con SPSS

A continuación, se detalla cada actividad realizada:

- a) Se estudió las características de la infraestructura para Machine Learning, de lo cual se aprecian algunas características más importantes.

Para este análisis se considera como:

Infraestructura física: a una infraestructura dentro de nuestras instalaciones, que incluye equipos, programas, accesorios y profesionales necesarios para implementar y sostener este tipo de infraestructura.

Infraestructura en la nube: que consiste en servicios cloud (en la nube), que incluiría el pago de alquiler y otros gastos mínimos para su implementación.

Tabla 08. Características para infraestructura ML

Nro.	Característica	Infraestructura física	Infraestructura en la nube
1	Hardware	Inversión elevada	Baja inversión
2	Software	Existen libres	Costo según utilización
3	Mantenimiento	Necesaria periódicamente	No requiere
4	Espacio físico	Según hardware y personal	No requiere
5	Tiempo procesamiento	Elevado	Reducido
6	Profesionales	Mas de un profesional	Mínimo un profesional

Interpretación: En el cuadro anterior se apreció que el uso de una infraestructura en la nube nos ofrece una opción muy aceptable por la baja inversión requerida tanto en hardware, software y mantenimiento, así como en el número de profesionales que deberían estar involucrados para la implementación de la infraestructura requerida.

- b) Se evaluó el cuadrante mágico para servicios de desarrollador de IA en la nube y se hizo cuadro comparativo de tres plataformas en la nube para Machine Learning.



Gráfico 4. Cuadrante de Gartner para servicios de desarrollador de IA en la nube

fuelle: (Gartner, 2020)

Tabla 09. Comparación de plataformas en la nube

		Amazon ML	Microsoft Azure	IBM Watson
Supervisado	Clasificación	x	x	x
	Regresión	x	x	x
No supervisado	Agrupación	x	x	
Detención de anomalías			x	
Algoritmos incorporados		x	x	x

Ventajas resaltantes	Elige el método ML automáticamente después de ver los datos proporcionados.	Interfaz gráfica, y gran variedad de algoritmos disponibles.	Procesamiento de datos automatizado e interfaz de construcción de modelos.
Framework soportados	TensorFlow, MXNet, Keras, Gluon, Pytorch, caffe2, Chainer, Torch	TensorFlow, scikit-learn, Microsoft Cognitive Toolkit, spark ML	TensorFlow, Spark MLlib, scikit-learn, XGBoost, PyTorch, IBM SPSS, PMML

Fuente: (Altexsoft, 2019)

Interpretación: Según el cuadrante de Gartner (febrero del 2020), las plataformas líderes, para servicios de desarrollador de IA en la nube, en la actualidad son AWS, Microsoft, Google e IBM; de los cuales procedimos a comprar sus características más resaltantes, y en base a que necesitaremos hacer comparativas con los algoritmos determinados en el objetivo 01, podríamos elegir cualquiera de los descritos en la tabla, sin embargo, siendo que Microsoft Azure ofrece una interfaz gráfica más amigable y gran variedad de algoritmos que pueden ser comparados de una manera más fácil e interactiva, es una excelente elección.

- c) Se encuestó a docentes de la UNS sobre, según ellos, cual consideran que es la mejor plataforma en la nube, para nuestro caso de estudio. Esta encuesta fue realizada a los docentes de la escuela de Ingeniería de Sistemas e informática que indicaban tener un grado de experiencia en el uso de plataformas en la nube, resultado del cual se elaboró la siguiente tabla.

Tabla 10. Plataforma según docentes UNS

Plataforma	Total, docentes
Amazon	1
Microsoft Azure	4
IBM Watson	2
Google Cloud	3
Otros	0

Interpretación: Se pudo apreciar que 4 de los 10 docentes encuestados nos indicaron que Microsoft Azure es la plataforma más adecuada para nuestro objeto de estudio.

También se obtuvo gráficos en la que se puede observar la selección de las plataformas según la encuesta realizada a los docentes

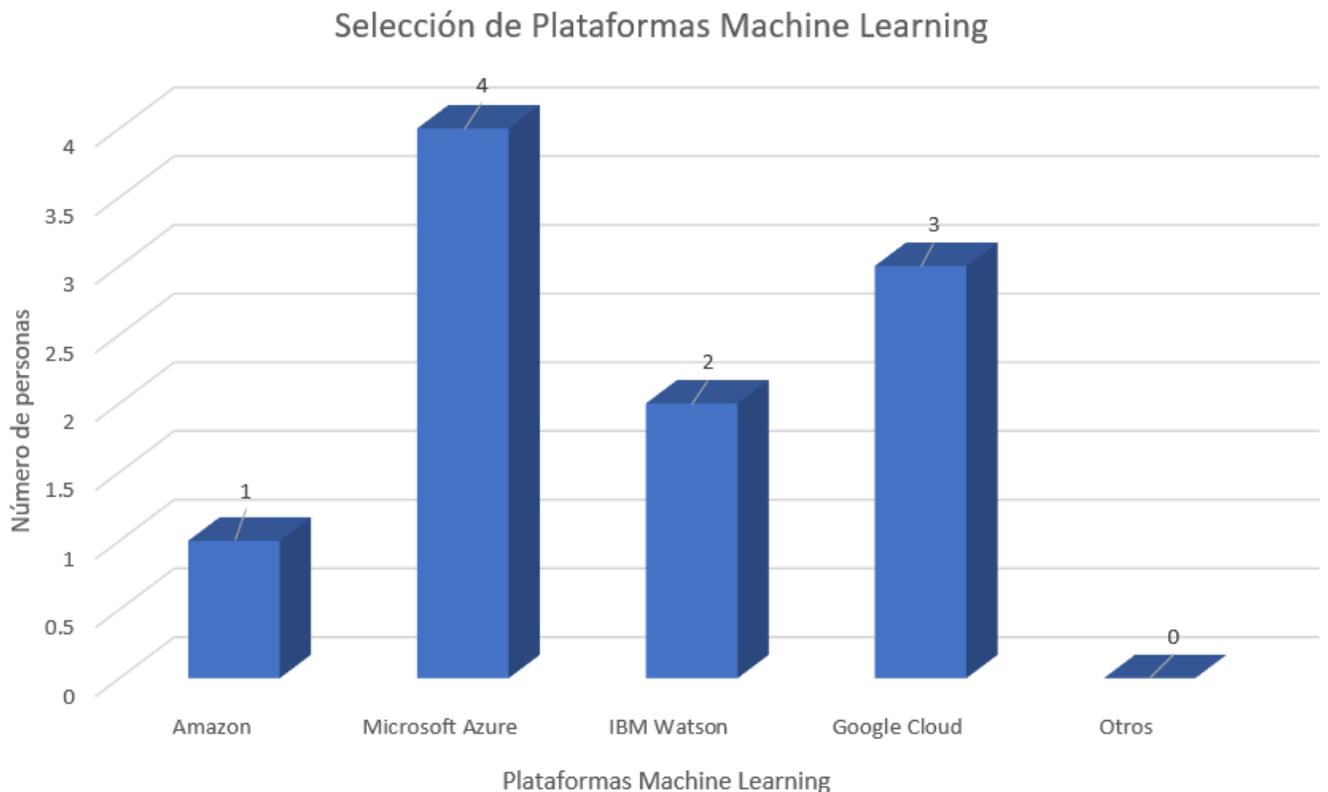


Gráfico 5. Gráfico de barras para Plataformas según docentes UNS

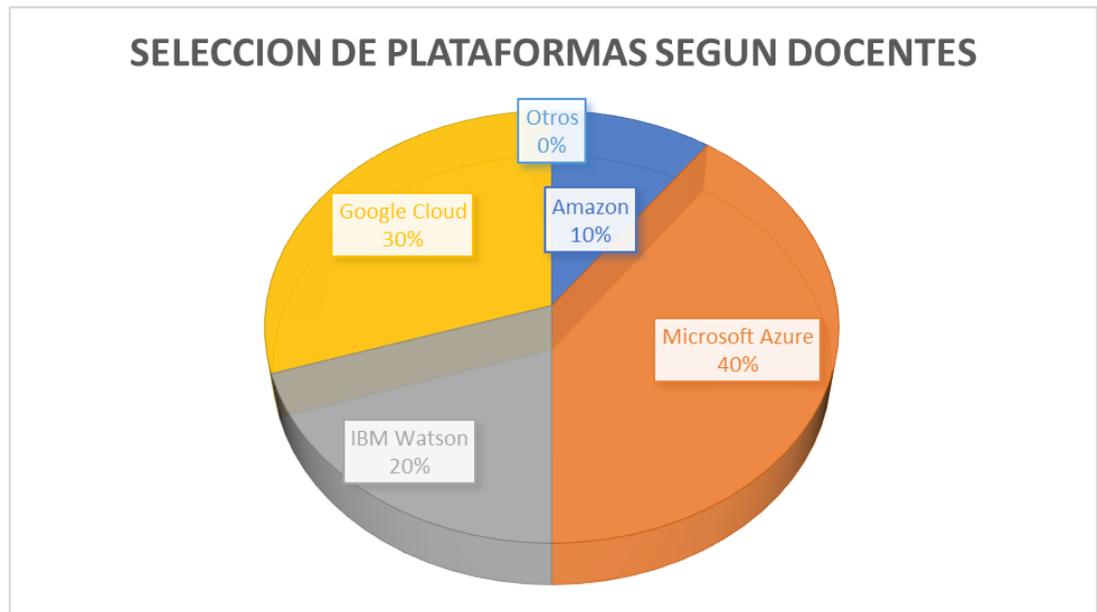


Gráfico 6. Diagrama de torta para Plataformas según docentes UNS

Interpretación: según los gráficos anteriores, el 40% de los docentes de la UNS encuestados, recomendaron que la plataforma de Microsoft Azure es apropiada para realizar la predicción de nuestro caso de estudio.

- d) Con la aplicación SPSS se realizó el estudio de fiabilidad para la encuesta, obteniendo un valor de significancia de 0.962.

Interpretación: Con una significancia del 0.962 para la encuesta, que incluye el ítem 02 para la elección de la plataforma, podemos decir que es un instrumento con alta fiabilidad.

Discusión: En base al estudio realizado a las plataformas de servicio Machine Learning, podemos ver que el 40% de profesionales reconocen a Microsoft Azure como la plataforma aceptable para realizar el análisis predictivo para nuestro caso de estudio, y siendo considerada por la empresa Gartner una de las plataformas líderes en el 2020, así como los trabajos de Zúñiga Ugalde (2019), quien destaca a Microsoft Azure ML Studio como un framework con las

características más variadas, cubriendo la mayoría de las tareas relacionadas con el Machine Learning.

4.3. Determinar las variables predictoras que intervienen en la cantidad de alumnos de pregrado a matricularse por asignatura durante un periodo académico, mediante el estudio, análisis y selección de las más apropiadas

Para determinar las variables predictoras que interviene se realizó:

- Análisis de las variables predictoras
- Encuesta de influencia de variable predictoras
- Recopilación de data sobre las variables predictoras
- Selección de mejores variables predictoras usando Weka

A continuación, se detallan las acciones requeridas para el logro del objetivo:

- a) Se realizó un análisis de las variables predictoras a través de un diagrama de dependencia y descripción de las variables

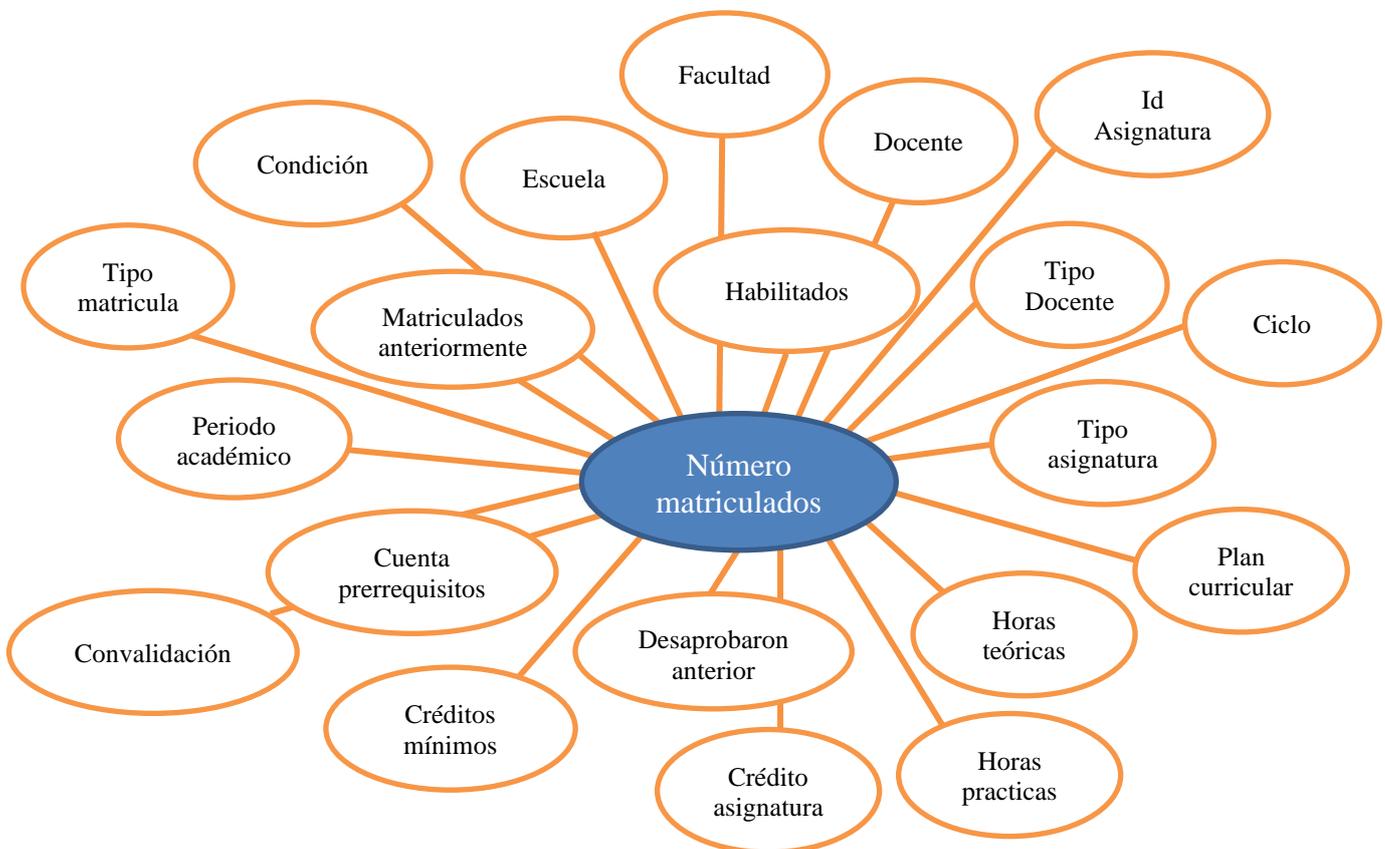


Gráfico 7. Diagrama de dependencia

- Se realizó una tabla para describir cada variable predictora, y se pueda tener un mejor entendimiento de los valores de las variables, su escala y categoría.

Tabla 11. Lista de variables predictoras.

Variable	Tipo	Escala de medición	Categoría	
01	Id Asignatura	Cuantitativo	Ordinal	Id de identificación de asignatura
02	Escuela	Cualitativo	Ordinal	110: Ing. En Energía 120: Ing. Agroindustrial 130: Ing. Civil 140: Ing. Sistemas e Informática 150: Ing. Mecánica 160: Ing. Agrónoma
03	Facultad	Cualitativo	Nominal	1: Ingeniería 2: Ciencias 3: Educación
04	Docente	Cualitativo	Ordinal	Según el id en la base de datos
05	Tipo docente	Cualitativo	Nominal	0: Contratado 1: Nombrado
06	Ciclo	Cualitativo	Ordinal	Del ciclo 1 al 10
07	Tipo asignatura	Cualitativo	Nominal	1: Regular 2: Electivo
08	Plan curricular	Cualitativo	Ordinal	1: Plan activo 2: Plan no activo
09	Crédito	Cuantitativo	Ordinal	Entre 0 a 20 créditos
10	Horas teóricas	Cuantitativo	Escalar	Entre 0 a 30 horas semanal
11	Horas practicas	Cuantitativo	Escalar	Entre 0 a 30 horas semanal
12	Convalidación	Cualitativo	Nominal	0: sin convalidación 1: con convalidación

13	Periodo académico	Cualitativo	Ordinal	Id del periodo académico
14	Tipo matricula	Cualitativo	Nominal	0: Solo regular 1: Otras matriculas
15	Condición	Cuantitativo	Nominal	0: Solo por primera 1: Con más de una vez
16	Matricula anterior	Cuantitativo	Escalar	Número de alumnos matriculados en el semestre anterior
17	Prerrequisito	Cualitativo	Nominal	0: sin prerrequisitos 1: con prerrequisitos
18	Habilitados	Cuantitativo	Escalar	Número de alumnos que aprobaron prerrequisito en el semestre anterior
19	Desaprobados	Cuantitativo	Escalar	Número de alumnos que desaprobaron la asignatura semestre anterior.
20	Matriculados	Cuantitativo	Escalar	Número de alumnos matriculados

Fuente: UDEM SI

Interpretación: Con el estudio de las tablas del Sistema de Información Integral de Gestión Académica y Administrativa (SIIGAA) de la UNS, referentes al ámbito académico, se pudo establecer las variables involucradas para determinar el número de alumnos matriculados en una asignatura, entre las que destacan variables que identifican a la asignatura, periodo académico, ciclo, plan curricular, así como algunas variables para las cuales se tendrán que calcular, como matriculados en el semestre anterior, número de desaprobados o número de habilitados para matrícula en la asignatura.

b) Se realizó una encuesta a los jefes de departamentos para determinar el grado de influencia de las variables predictoras para proyectar el número de alumnos a matricularse en una asignatura

- Se realizó la encuesta a los jefes de departamento consultándoles por el grado de influencia de cada variable y se obtuvo el siguiente resultado

Tabla 12. Cuadro de promedios de grado de influencia de variables

Nro	Variables	Grado de influencia						Promedio
		1	2	3	4	5	6	
1	ID de asignatura	1	0	2	0	0	0	0.50
2	Escuela	1	0	3	3	1	4	2.00
3	Facultad	1	0	2	0	2	3	1.33
4	Docente	1	1	4	1	5	0	2.00
5	Tipo docente	0	1	5	1	4	0	1.83
6	Ciclo	1	2	2	2	0	2	1.50
7	Tipo asignatura	3	2	5	3	4	3	3.33
8	Plan curricular	3	2	3	2	4	4	3.00
9	Crédito	0	3	2	1	0	3	1.50
10	HT	0	1	2	1	3	3	1.67
11	HP	0	1	2	1	3	3	1.67
12	Convalidación	2	2	2	3	4	2	2.50
13	Periodo académico	0	2	2	0	0	1	0.83
14	Tipo de matrícula	0	2	3	3	0	3	1.83
15	Condición	2	5	4	2	2	4	3.17
16	Numero matricula anterior	5	4	5	5	4	4	4.50
17	Tener pre requisito	5	5	4	4	5	4	4.50
18	Numero de habilitados	5	5	5	5	4	5	4.83
19	Numero de desaprobados	5	5	4	5	4	4	4.50

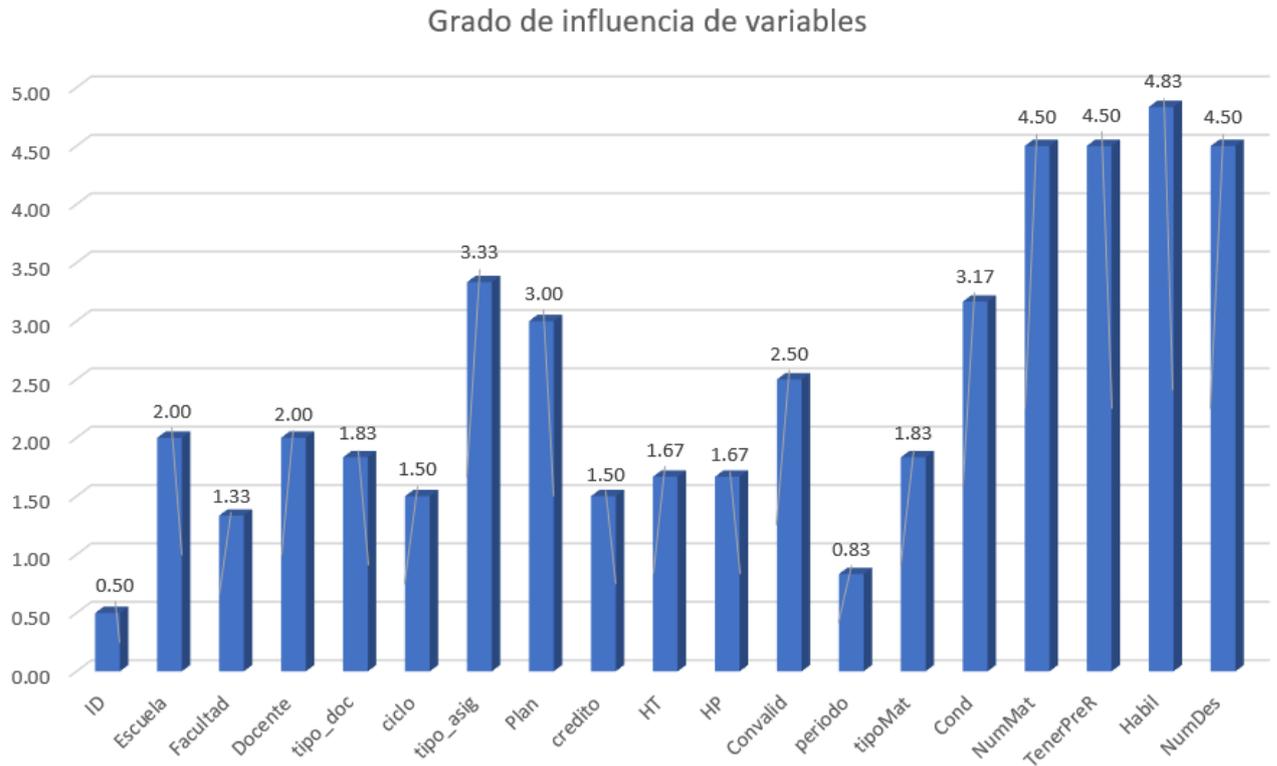


Gráfico 8. Gráfico de barras para el grado de influencia por variable

- La validez de la encuesta se realizó con la técnica a juicio de experto y la fiabilidad se obtuvo con un valor de significancia del 0.880 (véase anexo 6).

Interpretación: Se evaluó cada variable y se realizó la encuesta determinándose que las variables como número de matriculados en el semestre anterior, que la asignatura tenga o no pre requisito, el número de alumnos hábiles y el número de alumnos desaprobados en el semestre anterior, son variables que influyen en gran medida en el número de alumnos a matricularse, pero las demás variables también influyen de alguna manera; sin embargo, por la experiencia en el ámbito académico se dejarán de lado las variables que identifican a la facultad que también viene representada por su escuela; así como las variables que identifican al docente y tipo de docente porque hasta el momento de matrícula no estará totalmente definido.

c) Recopilación de la data

Debido a que hasta el punto anterior se determinó que las variables predictoras están influenciando de alguna manera en la proyección del número de alumnos a matricularse en una determinada asignatura, se va a utilizar la herramienta Weka para limpiar mejor las variables predictoras.

La Universidad Nacional del Santa cuenta con el Sistema de Información Integral de Gestión Académica y Administrativa que desde el 2006 recopila información del número de alumnos matriculados semestre a semestre y por asignatura.

Para lo cual se elaboró un procedimiento almacenado para la recopilación total de la información que esta almacenado en la Base de Datos de la institución, puesto que se identificó variables que deberían ser calculados (como número de matriculados, número de desaprobados y número de matriculados en el semestre anterior) y no era posible obtenerlo simplemente con sentencias SQL o vistas; dicho procedimiento dio como resultado un total de 4624 filas de información.

Interpretación: De la información total obtenida, se filtró solo las asignaturas que se dictan en el ciclo par, puesto que para nuestro objeto de estudio se intenta predecir el número de matriculados en el ciclo 2019-II

d) Selección de variables involucradas

Con el apoyo del software Weka 3.9 se validaron los datos proporcionados por la recopilación de data

- Primero se carga la información del procedimiento almacenado a Weka.

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Open file... | Open URL... | Open DB... | Gene...

Filter

Choose **None**

Current relation

Relation: dataWeka | Instances: 4625 | Attributes: 17 | Sum of weights: 4625

Attributes

All | None | Invert | Pattern

No.	Name
1	<input checked="" type="checkbox"/> idAsignatura
2	<input checked="" type="checkbox"/> escuela
3	<input checked="" type="checkbox"/> ciclo
4	<input checked="" type="checkbox"/> tipoAsig
5	<input checked="" type="checkbox"/> planCur
6	<input checked="" type="checkbox"/> credito
7	<input checked="" type="checkbox"/> ht
8	<input checked="" type="checkbox"/> hp
9	<input checked="" type="checkbox"/> convalidacion
10	<input checked="" type="checkbox"/> periodoAcad
11	<input checked="" type="checkbox"/> tipoMatricula
12	<input checked="" type="checkbox"/> condicion
13	<input checked="" type="checkbox"/> matAnterior
14	<input checked="" type="checkbox"/> prerrequisito
15	<input checked="" type="checkbox"/> habilitados
16	<input checked="" type="checkbox"/> desaprobados
17	<input checked="" type="checkbox"/> matriculados

Remove

Status

OK

Gráfico 9. Atributos del modelo en Weka

Interpretación: Para el logro de lo graficado se tuvo que hacer la conversión del resultado obtenido por el procedimiento almacenado a un formato CSV el cual puede leer el software Weka, aceptando todos los valores proporcionados como se muestra en la gráfica.

- Se pudo visualizar la data importada a Weka

Viewer X

Relation: dataWeka

No.	1: idAsignatura	2: escuela	3: ciclo	4: tipoAsig	5: planCur	6: credito	7: ht	8: hp	9: convalidacion	10: periodoAcad	11: tipoMatricula	12: condicion	13: matAnterior	14: prerequisite	15: habilitados	16: desaprobados	17: matriculados
	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric
1	431.0	210.0	6.0	1.0	1.0	10.0	5.0	15.0	1.0	20022.0	1.0	1.0	0.0	1.0	55.0	8.0	54.0
2	431.0	210.0	6.0	1.0	1.0	10.0	5.0	15.0	1.0	20032.0	1.0	1.0	54.0	1.0	46.0	11.0	63.0
3	431.0	210.0	6.0	1.0	1.0	10.0	5.0	15.0	1.0	20052.0	1.0	1.0	47.0	1.0	65.0	10.0	45.0
4	431.0	210.0	6.0	1.0	1.0	10.0	5.0	15.0	1.0	20042.0	1.0	1.0	63.0	1.0	42.0	7.0	47.0
5	434.0	210.0	6.0	1.0	1.0	4.0	2.0	4.0	1.0	20062.0	1.0	1.0	43.0	1.0	40.0	3.0	60.0
6	434.0	210.0	6.0	1.0	1.0	4.0	2.0	4.0	1.0	20022.0	1.0	1.0	0.0	1.0	55.0	5.0	59.0
7	434.0	210.0	6.0	1.0	1.0	4.0	2.0	4.0	1.0	20032.0	1.0	1.0	59.0	1.0	51.0	14.0	54.0
8	434.0	210.0	6.0	1.0	1.0	4.0	2.0	4.0	1.0	20052.0	1.0	1.0	58.0	1.0	44.0	9.0	43.0
9	434.0	210.0	6.0	1.0	1.0	4.0	2.0	4.0	1.0	20042.0	1.0	1.0	54.0	1.0	36.0	9.0	58.0
10	437.0	210.0	7.0	1.0	1.0	3.0	1.0	4.0	0.0	20061.0	1.0	1.0	52.0	1.0	46.0	14.0	39.0
11	437.0	210.0	7.0	1.0	1.0	3.0	1.0	4.0	0.0	20021.0	1.0	1.0	0.0	1.0	38.0	8.0	44.0
12	437.0	210.0	7.0	1.0	1.0	3.0	1.0	4.0	0.0	20031.0	1.0	1.0	44.0	1.0	39.0	3.0	50.0
13	437.0	210.0	7.0	1.0	1.0	3.0	1.0	4.0	0.0	20041.0	1.0	0.0	50.0	1.0	44.0	2.0	39.0
14	437.0	210.0	7.0	1.0	1.0	3.0	1.0	4.0	0.0	20051.0	1.0	1.0	39.0	1.0	53.0	1.0	52.0
15	437.0	210.0	7.0	1.0	1.0	3.0	1.0	4.0	0.0	20071.0	1.0	1.0	39.0	1.0	48.0	2.0	58.0
16	439.0	210.0	8.0	1.0	1.0	5.0	2.0	6.0	1.0	20022.0	1.0	1.0	0.0	1.0	63.0	8.0	48.0
17	439.0	210.0	8.0	1.0	1.0	5.0	2.0	6.0	1.0	20032.0	1.0	1.0	48.0	1.0	38.0	0.0	47.0
18	439.0	210.0	8.0	1.0	1.0	5.0	2.0	6.0	1.0	20052.0	1.0	1.0	56.0	1.0	54.0	6.0	42.0
19	439.0	210.0	8.0	1.0	1.0	5.0	2.0	6.0	1.0	20042.0	1.0	0.0	47.0	1.0	34.0	4.0	56.0
20	444.0	210.0	8.0	1.0	1.0	10.0	5.0	15.0	1.0	20062.0	1.0	1.0	57.0	1.0	1.0	4.0	44.0
21	444.0	210.0	8.0	1.0	1.0	10.0	5.0	15.0	1.0	20022.0	1.0	0.0	0.0	1.0	55.0	5.0	43.0
22	444.0	210.0	8.0	1.0	1.0	10.0	5.0	15.0	1.0	20032.0	1.0	1.0	43.0	1.0	46.0	7.0	42.0
23	444.0	210.0	8.0	1.0	1.0	10.0	5.0	15.0	1.0	20052.0	1.0	1.0	49.0	1.0	65.0	8.0	57.0
24	444.0	210.0	8.0	1.0	1.0	10.0	5.0	15.0	1.0	20042.0	1.0	1.0	42.0	1.0	42.0	6.0	49.0
25	445.0	210.0	9.0	1.0	1.0	5.0	2.0	6.0	1.0	20061.0	1.0	0.0	49.0	1.0	41.0	0.0	32.0
26	445.0	210.0	9.0	1.0	1.0	5.0	2.0	6.0	1.0	20021.0	1.0	0.0	0.0	1.0	48.0	5.0	45.0
27	445.0	210.0	9.0	1.0	1.0	5.0	2.0	6.0	1.0	20031.0	1.0	1.0	45.0	1.0	48.0	5.0	58.0
28	445.0	210.0	9.0	1.0	1.0	5.0	2.0	6.0	1.0	20041.0	1.0	0.0	58.0	1.0	43.0	2.0	39.0
29	445.0	210.0	9.0	1.0	1.0	5.0	2.0	6.0	1.0	20051.0	1.0	0.0	39.0	1.0	50.0	1.0	49.0
30	445.0	210.0	9.0	1.0	1.0	5.0	2.0	6.0	1.0	20071.0	1.0	1.0	32.0	1.0	46.0	3.0	50.0
31	445.0	210.0	9.0	1.0	1.0	5.0	2.0	6.0	1.0	20081.0	1.0	0.0	50.0	1.0	53.0	0.0	63.0
32	446.0	210.0	9.0	1.0	1.0	10.0	5.0	15.0	1.0	20061.0	1.0	0.0	52.0	1.0	54.0	2.0	40.0
33	446.0	210.0	9.0	1.0	1.0	10.0	5.0	15.0	1.0	20021.0	1.0	0.0	0.0	1.0	33.0	0.0	48.0
34	446.0	210.0	9.0	1.0	1.0	10.0	5.0	15.0	1.0	20031.0	1.0	0.0	48.0	1.0	54.0	0.0	36.0
35	446.0	210.0	9.0	1.0	1.0	10.0	5.0	15.0	1.0	20041.0	1.0	0.0	36.0	1.0	51.0	4.0	37.0
36	446.0	210.0	9.0	1.0	1.0	10.0	5.0	15.0	1.0	20051.0	1.0	1.0	37.0	1.0	33.0	3.0	52.0
37	452.0	210.0	10.0	1.0	1.0	5.0	0.0	10.0	1.0	20062.0	1.0	0.0	51.0	1.0	71.0	0.0	49.0
38	452.0	210.0	10.0	1.0	1.0	5.0	0.0	10.0	1.0	20022.0	1.0	0.0	0.0	1.0	44.0	0.0	43.0
39	452.0	210.0	10.0	1.0	1.0	5.0	0.0	10.0	1.0	20032.0	1.0	0.0	43.0	1.0	45.0	0.0	45.0
40	452.0	210.0	10.0	1.0	1.0	5.0	0.0	10.0	1.0	20052.0	1.0	0.0	26.0	1.0	59.0	0.0	51.0
41	452.0	210.0	10.0	1.0	1.0	5.0	0.0	10.0	1.0	20042.0	1.0	0.0	45.0	1.0	37.0	1.0	26.0
42	454.0	210.0	10.0	1.0	1.0	5.0	0.0	10.0	1.0	20062.0	1.0	0.0	52.0	1.0	40.0	0.0	55.0
43	454.0	210.0	10.0	1.0	1.0	5.0	0.0	10.0	1.0	20022.0	1.0	0.0	0.0	1.0	48.0	0.0	43.0
44	454.0	210.0	10.0	1.0	1.0	5.0	0.0	10.0	1.0	20032.0	1.0	0.0	43.0	1.0	32.0	0.0	39.0
45	454.0	210.0	10.0	1.0	1.0	5.0	0.0	10.0	1.0	20072.0	1.0	0.0	55.0	1.0	55.0	0.0	49.0
46	454.0	210.0	10.0	1.0	1.0	5.0	0.0	10.0	1.0	20052.0	1.0	0.0	24.0	1.0	50.0	0.0	52.0
47	454.0	210.0	10.0	1.0	1.0	5.0	0.0	10.0	1.0	20042.0	1.0	0.0	39.0	1.0	34.0	0.0	24.0
48	454.0	210.0	10.0	1.0	1.0	5.0	0.0	10.0	1.0	20082.0	1.0	0.0	49.0	1.0	58.0	0.0	57.0

Add instance Undo OK Cancel

Gráfico 10. Vista de data en Weka

Interpretación: Se verificó que Weka no eliminó ni observó ninguna fila de la importación de data hecha por el procedimiento almacenado

- Con Weka, se determinó la mejor selección de atributos para el modelo predictivo.

```

=== Attribute Selection on all input data ===

Search Method:
  Greedy Stepwise (forwards).
  Start set: no attributes
  Merit of best subset found: 0.793

Attribute Subset Evaluator (supervised, Class (numeric): 17 matriculados):
  CFS Subset Evaluator
  Including locally predictive attributes

Selected attributes: 2,3,4,5,9,12,13,15,16 : 9
  escuela
  ciclo
  tipoAsig
  planCur
  convalidacion
  condicion
  matAnterior
  habilitados
  desaprobados

```

Gráfico 11. Vista de atributos seleccionados por Weka

Interpretación: Con el apoyo de la herramienta Weka se pudo establecer que los mejores atributos para nuestro modelo predictivo son:

Tabla 13. Variables predictoras definitivas

	Variable	Tipo
01	Escuela	Entrada
02	Ciclo	Entrada
03	Tipo asignatura	Entrada
04	Plan curricular	Entrada
05	Convalidación	Entrada
06	Condición	Entrada
07	Matricula anterior	Entrada
08	Habilitados	Entrada
09	Desaprobados	Entrada
10	Matriculados	Salida

Discusión: El software libre Weka nos facilita la selección de atributos más representativos para un modelo predictivo, que luego del análisis, de las 20 variables predictoras con que partimos solo quedaron 10; Márquez Vera

(2015), Menacho Chiok (2016) en sus trabajos de investigación utilizan esta herramienta para el análisis y selección de las mejores variables predictoras.

4.4. Elaborar una solución de Machine Learning para predecir la cantidad de alumnos de pregrado a matricularse por asignatura, mediante la implementación, entrenamiento y testeo de un modelo utilizando las variables, algoritmos y plataforma seleccionados

Para elaborar una solución Machine Learning se realizó las siguientes actividades:

- Creación de proyecto matricula en Microsoft Azure
- Carga de dataset desde Weka
- Uso de Machine Learning de Microsoft Azure y comparando algoritmos
- Comparativa del modelo con el periodo 2018-II
- Fiabilidad de los datos obtenidos

a) Se inició la plataforma Microsoft Azure Machine Learning con la creación del proyecto Matrícula

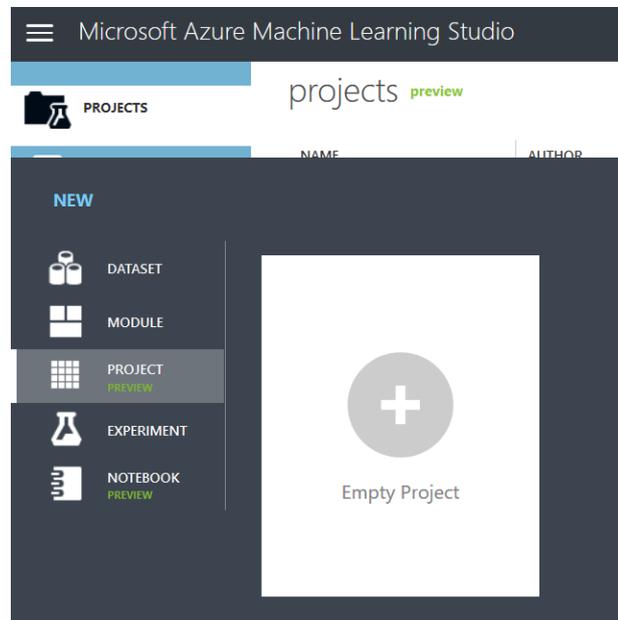


Gráfico 12. Creando proyecto Matrícula

Interpretación: Para iniciar el trabajo en Microsoft Azure hay que crear un proyecto, para lo cual previamente hay que crear un usuario para acceso,

el cual es gratuito por 30 días luego del cual hay un costo por procesamiento.

b) Se cargó dataset creado desde Weka

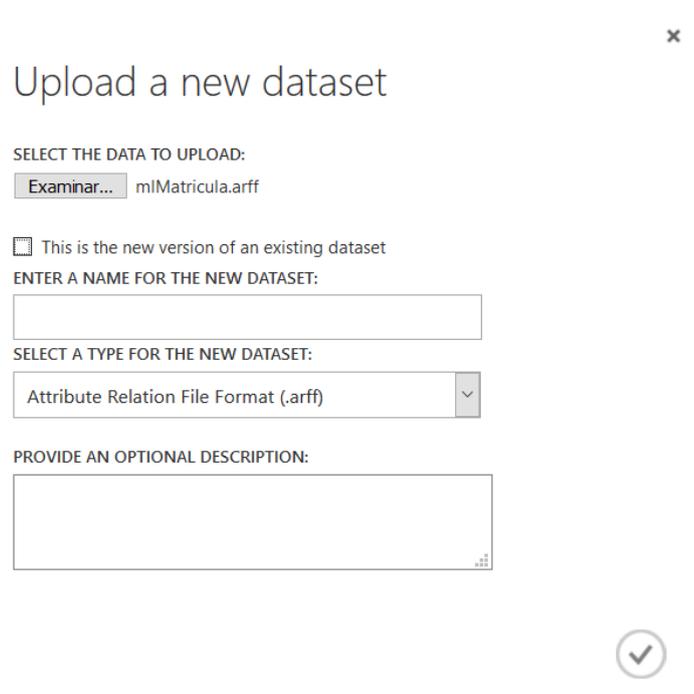


Gráfico 13. Agregando dataset

Interpretación: Hay que cargar el dataset filtrado desde Weka, que será ahora el dataset para nuestro modelo predictivo.

c) Se hizo uso de Microsoft Azure Machine Learning, y para lo cual se siguió los siguientes pasos a continuación descritos:

a. Asignación del conjunto de datos

Add to project

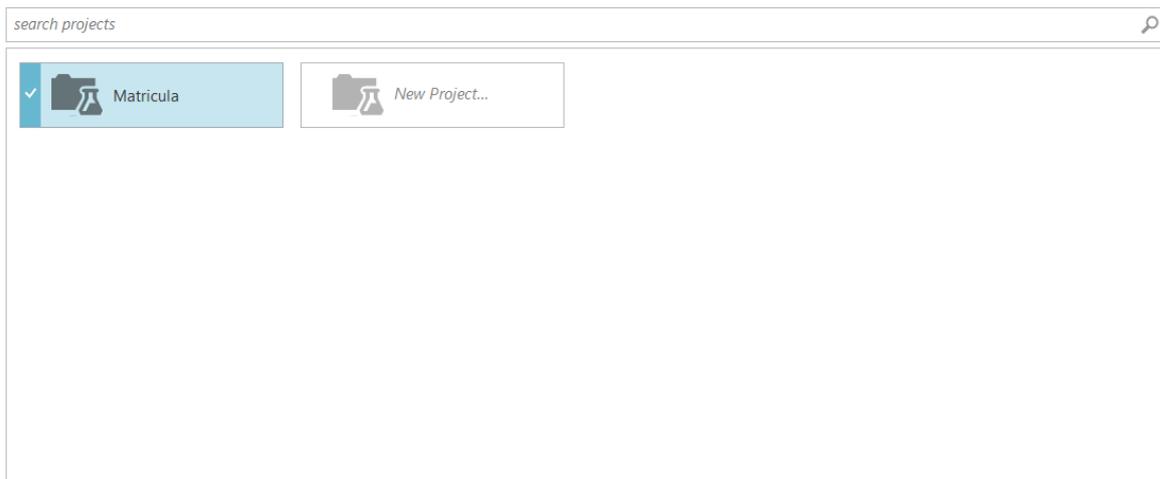


Gráfico 14. Asignando Dataset a Proyecto

b. Agregación del Dataset en experimento Matricula

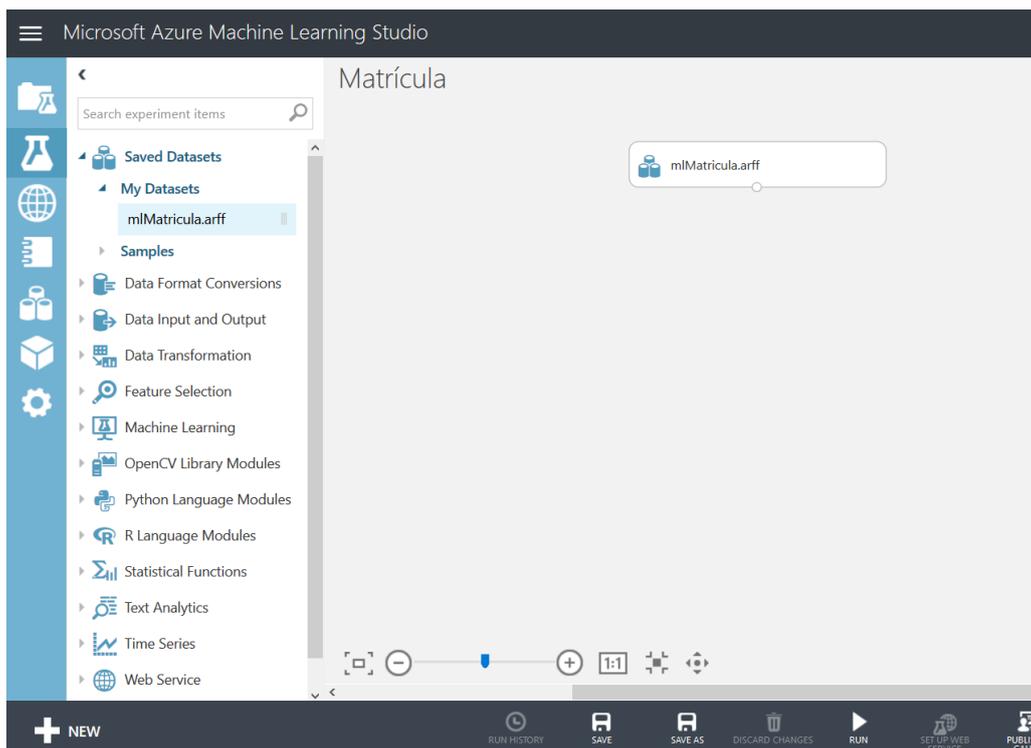


Gráfico 15. Agregando Dataset en Experimento

c. Se puede tener una vista de los datos para verificar que todo este correctamente cargado

Matrícula > mlMatricula.arff > dataset



Gráfico 16. Vista de datos del Dataset

d. Se agregaron los componentes de manipulación

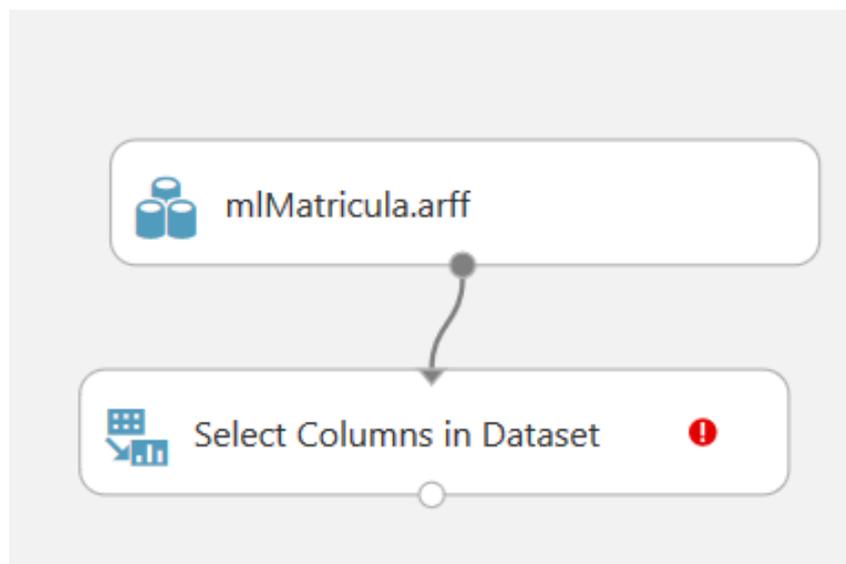


Gráfico 17. Agregando componentes de manipulación

e. Se agregan las columnas como categorías

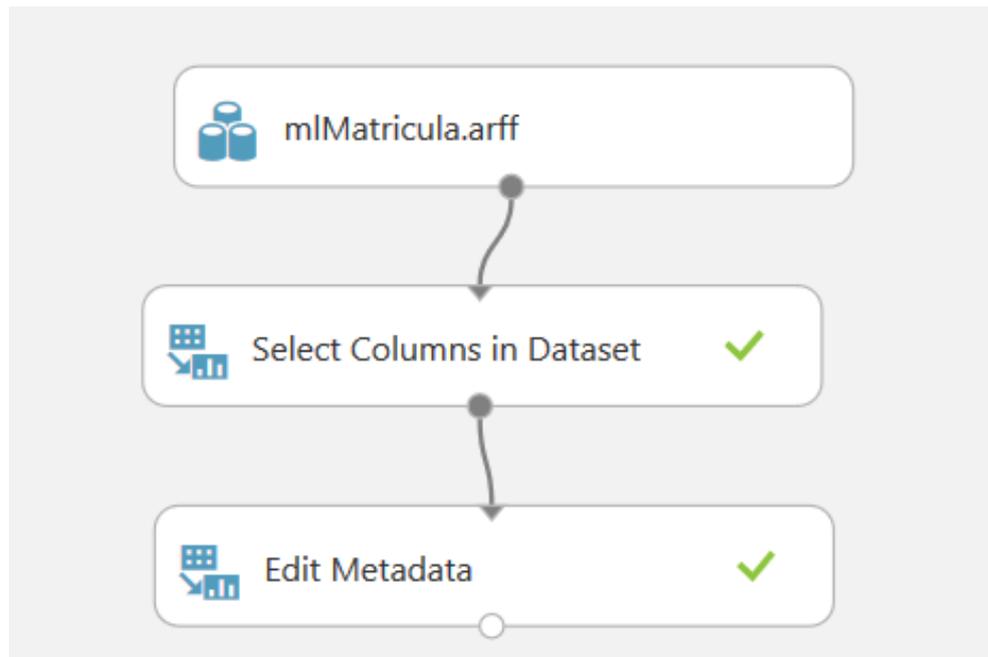


Gráfico 18. Agregando columnas como categoría

f. Se realizó la Limpieza y transformación de datos

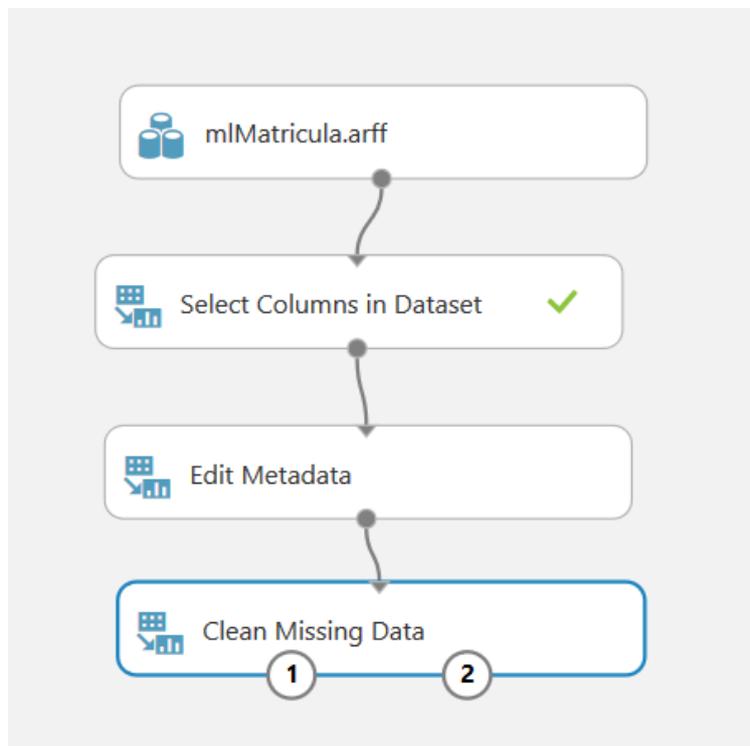


Gráfico 19. Eliminando valores nulos

g. Se seleccionaron los campos sobre los que se evaluará el resultado

Draft saved at 5:18:49

mlMatricula.arff

Select Columns in Dataset ✓

Edit Metadata ✓

Clean Missing Data ✓

Edit Metadata 1

Edit Metadata.
Edits metadata associated with columns in a dataset. Formerly known as Metadata Editor.

Edit Metadata

Column

Selected columns:
Column names:
matriculados

Launch column selector

Data type
Integer

Categorical
Unchanged

Fields
Label

New column names

Gráfico 20. Definiendo valor de salida

h. Se eliminaron los picos de data para la salida

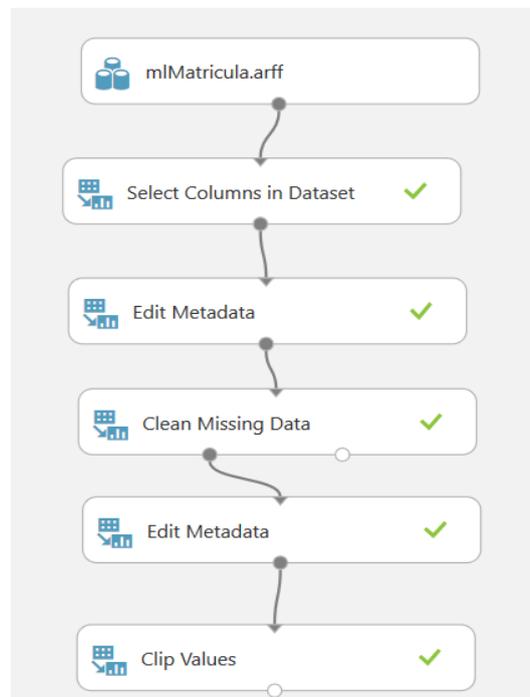


Gráfico 21. Eliminación de picos

i. Se seleccionaron los Split

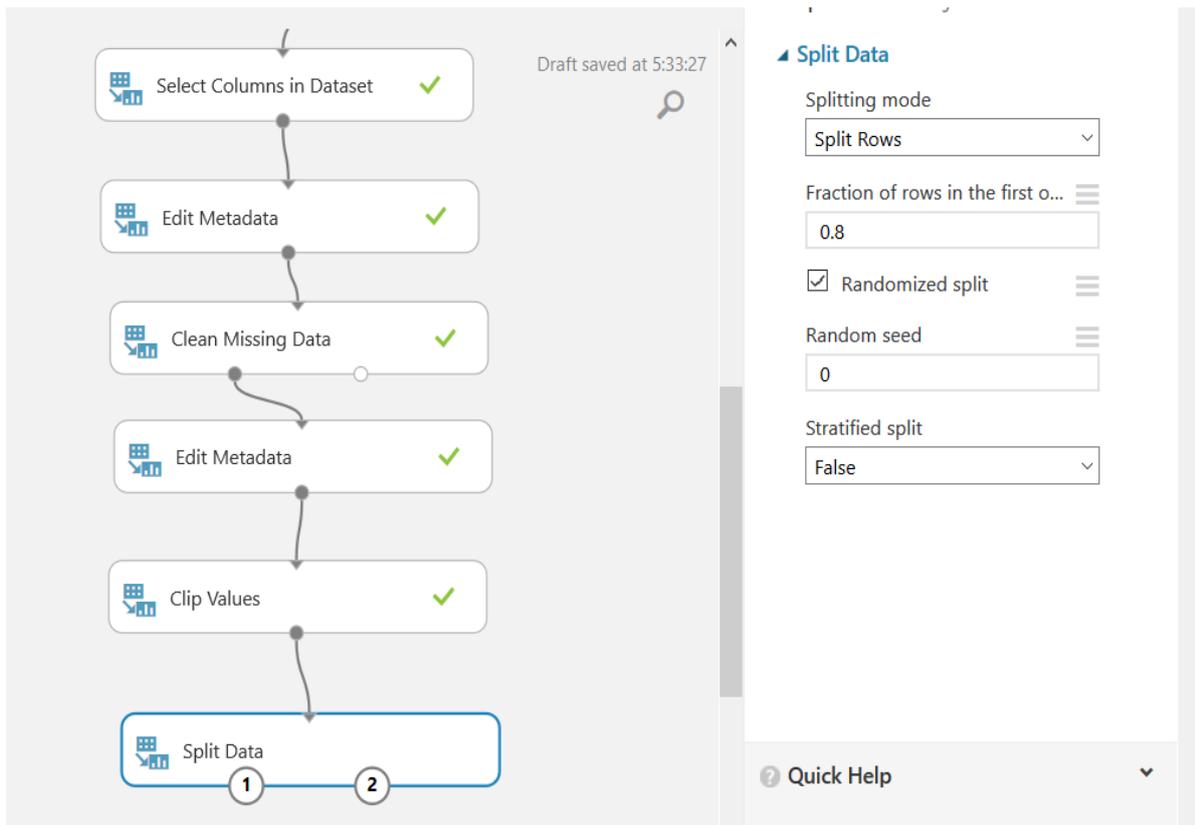


Gráfico 22. Split data

j. Se seleccionó un primer algoritmo (Regresión lineal)

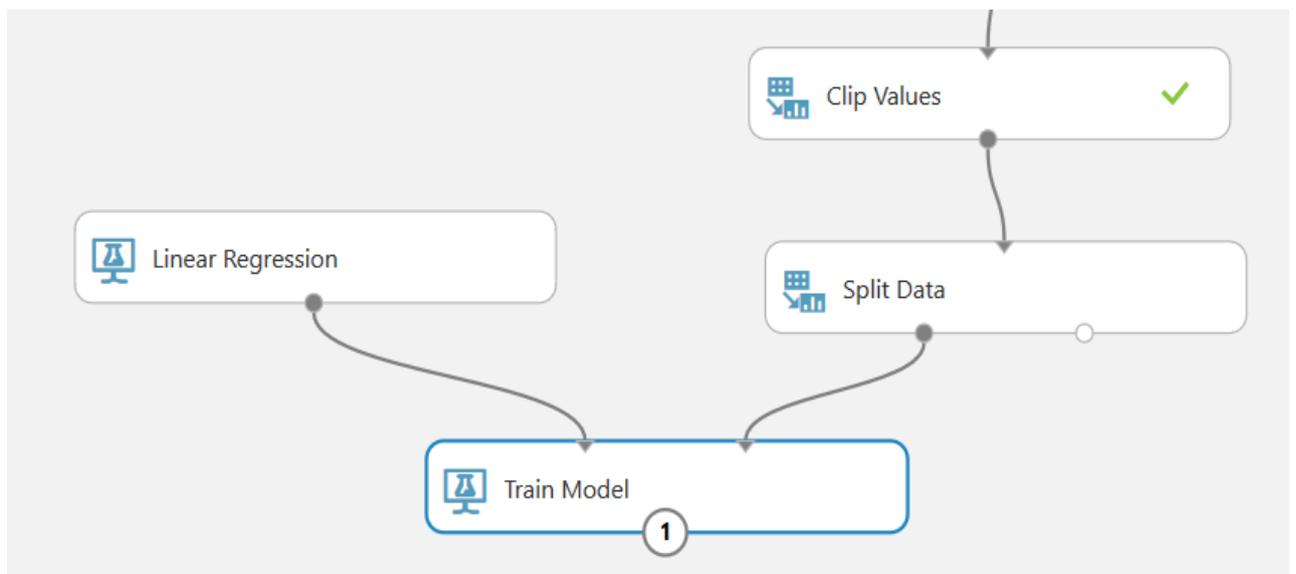


Gráfico 23. Algoritmo ML y entrenamiento

k. Se midió el modelo de regresión lineal

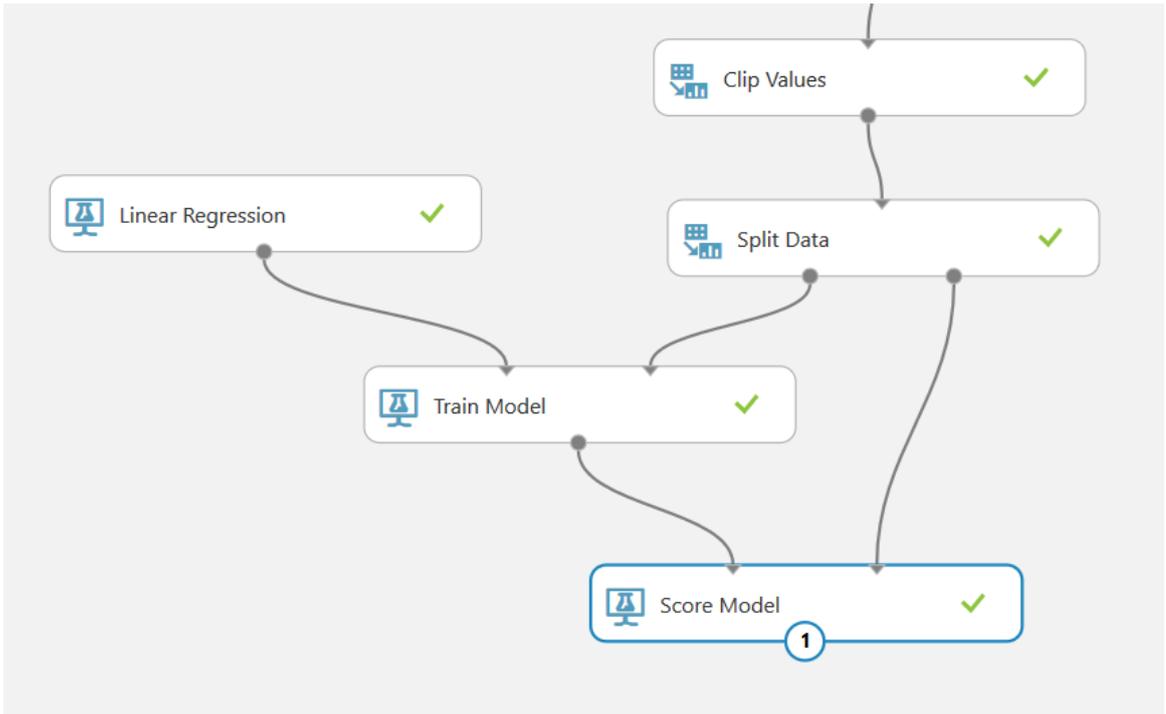


Gráfico 24. Score model Regresión Lineal

- I. Se agrego y midió el modelo de redes neuronales para hacer comparaciones

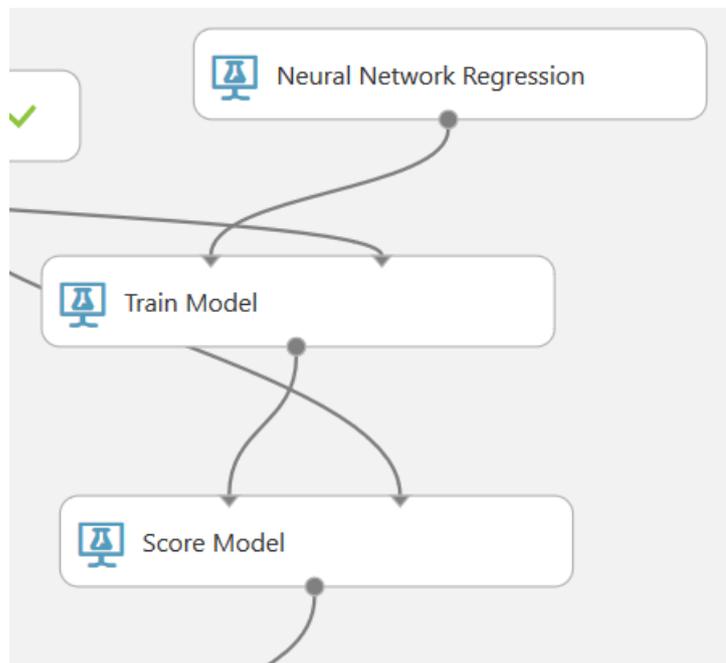


Gráfico 25. Score model Redes Neuronales

m. Se evaluaron y compararon los modelos

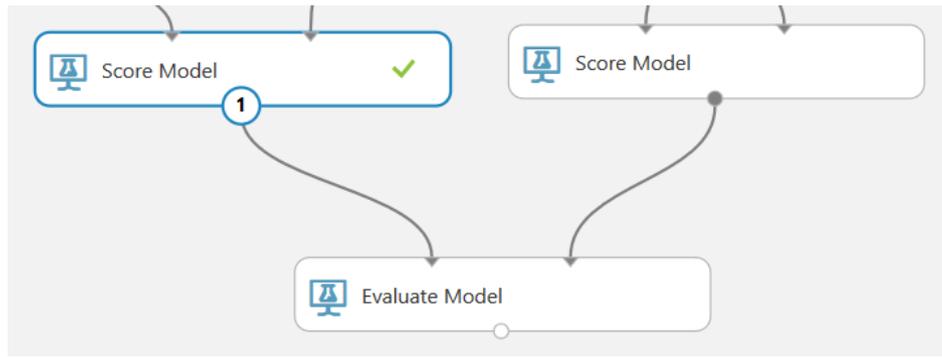


Gráfico 26. Score model en cada algoritmo

n. Se agregaron, midieron y evaluaron también los algoritmos de poisson, bayesiana, y árbol de decisiones

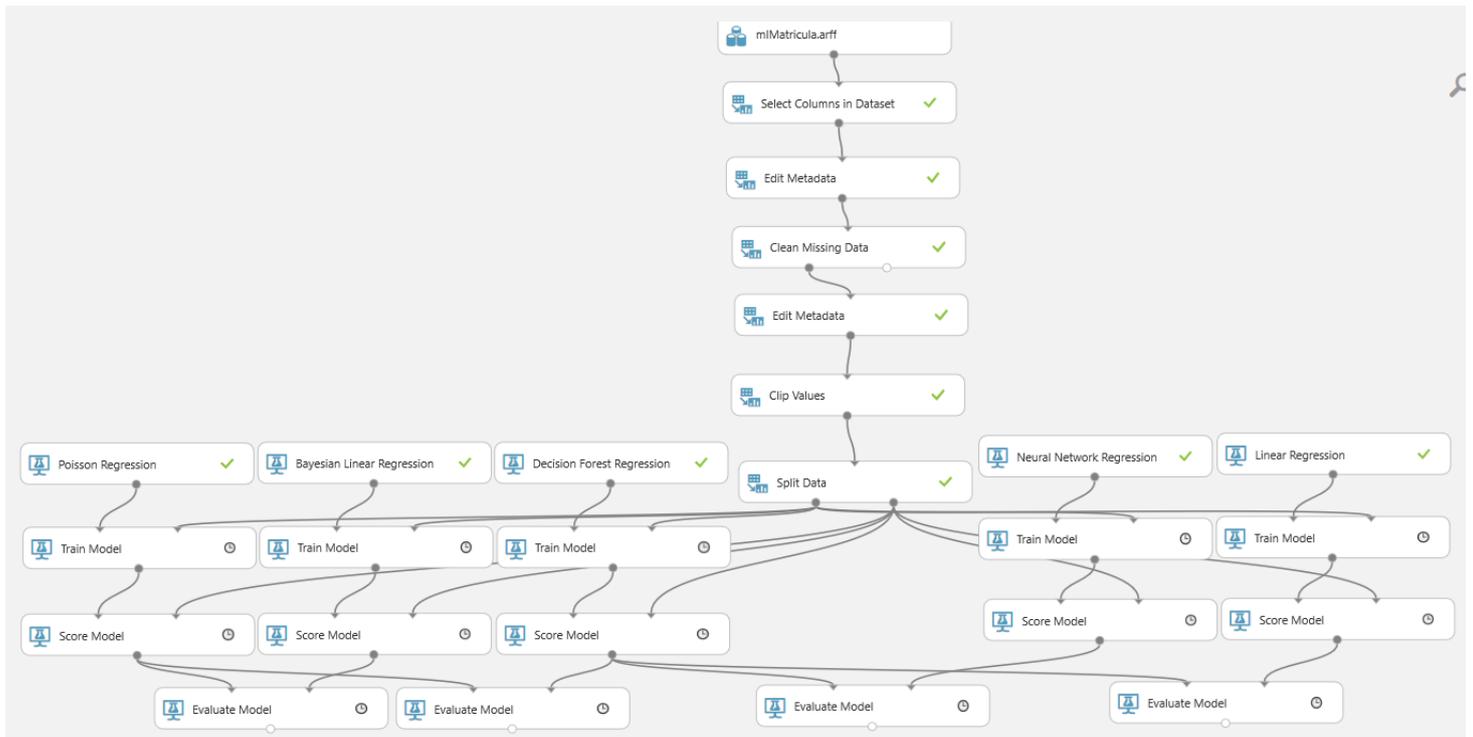


Gráfico 27. Vista General del modelo

Interpretación: Con el uso del objetivo 01 se realizó la comparación entre los algoritmos de regresión como Poisson, Bayesiano Lineal, Bosque de decisiones, Redes Neuronales y Lineal en la plataforma Machine Learning de Microsoft Azure, el cual dio como resultado que el mejor Coeficiente de

determinación fue de 0.7858 que corresponde al algoritmo Bosque de decisiones.

- d) Se realizó un muestreo con los valores del periodo académico 2018-II y los valores obtenidos por el modelo predictivo usando el algoritmo de Bosque de decisiones.

Para realizar esta actividad es necesario publicar nuestra solución en Microsoft Azure

- Se crea el servicio web con el botón “*SET UP WEB SERVICE* >

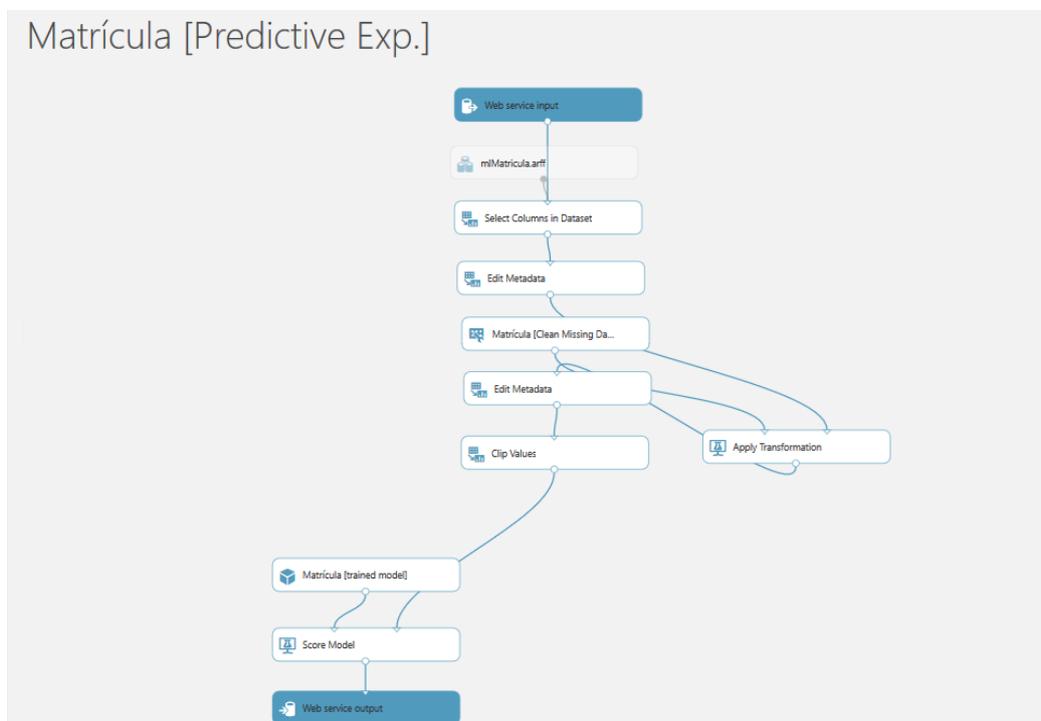


Gráfico 28. Vista del servicio generado

- Una vez creado el servicio, primero lo lanzamos con la opción “RUN” y luego lo publicamos desde la siguiente opción “DEPLOY WEB SERVICE”:



Finalmente obtendremos una pantalla con datos y acceso a la información del servicio

- En el último punto fuimos redirigidos a una página con la información del servicio, en se puede revisar la información más importante.

matrícula [predictive exp.]

DASHBOARD CONFIGURATION

General [New Web Services Experience](#) *preview*

Published experiment
[View snapshot](#) [View latest](#)

Description
 No description provided for this web service.

API key
 JF0Bg9Qc/GxHiqcPKsOz04NMiWflqXy3+ROs2C21IBGVRPkTlgHU6+C64YdWYO4DKkPtlbuuYNkbaIXWoIMXrA==

Default Endpoint

API HELP PAGE	TEST	APPS	LAST UPDATED
REQUEST/RESPONSE	Test Test preview	Excel 2013 or later Excel 2010 or earlier workbook	2/28/2020 6:46:29 PM
BATCH EXECUTION	Test Test preview	Excel 2013 or later workbook	2/28/2020 6:46:29 PM

Gráfico 29. Vista de la información del servicio

- Pulsamos el enlace “Test” se abre una página desde donde probamos el servicio. La página mostró los campos seleccionados en el modelo para que los rellenemos, una vez rellenos pulsamos en “Test Request-Response”.

input1

escuela

ciclo

tipoAsig

planCur

convalidacion

condicion

matAnterior

habilitados

desaprobados

matriculados

[Test Request-Response](#)

Gráfico 30. Vista de la página de testeo

- La página envió al servicio la información introducida, el modelo la procesó y devolvió las métricas obtenidas a través del servicio.



Gráfico 31. Vista de la información obtenida

- Con esta página se testeó 48 registros del periodo académico 2018-II, para las asignaturas de los últimos tres ciclos que los departamentos de la facultad de ingeniería que no son electivos, con lo cual se obtuvo la información detallada en la tabla 14.

Tabla 14. Resultados de testeo periodo académico 2018-II

escuela	ciclo	Tipo Asignatura	Plan Curricular	convalidación	condición	Mat. Anterior	habilitados	desaprobados	matriculados	Azure (redondeo)	APE	Reajuste
110	6	1	1	1	1	53	40	20	41	44	7.32%	Sin reajuste
110	6	1	1	1	0	51	40	2	42	41	2.38%	Sin reajuste
110	6	1	1	1	0	61	40	4	44	38	13.64%	Sin reajuste

110	6	1	1	1	1	50	46	4	47	48	2.13%	Sin reajuste
110	6	1	1	1	1	51	46	26	48	48	0.00%	Sin reajuste
110	6	1	1	1	1	66	46	17	54	49	9.26%	Sin reajuste
110	6	1	1	1	0	43	55	0	57	55	3.51%	Sin reajuste
110	8	1	1	1	0	40	33	1	32	37	15.63%	Sin reajuste
110	8	1	1	1	0	44	46	1	47	46	2.13%	Sin reajuste
110	8	1	1	1	1	40	43	5	47	47	0.00%	Sin reajuste
110	8	1	1	1	1	22	46	4	49	49	0.00%	Sin reajuste
110	8	1	1	1	0	21	44	0	66	55	16.67%	Con reajuste
110	10	1	1	1	1	45	26	4	27	28	3.70%	Sin reajuste
110	10	1	1	1	0	42	57	0	42	53	26.19%	Con reajuste
110	10	1	1	1	1	45	42	3	42	40	4.76%	Sin reajuste
110	10	1	1	1	0	46	47	2	47	43	8.51%	Con reajuste
120	6	1	2	1	0	50	44	5	33	37	12.12%	Sin reajuste
120	6	1	2	1	0	51	33	1	38	36	5.26%	Sin reajuste
120	6	1	2	1	1	54	40	9	41	44	7.32%	Sin reajuste
120	6	1	2	1	1	58	43	2	42	45	7.14%	Sin reajuste
120	6	1	2	1	1	58	43	8	45	46	2.22%	Con reajuste
120	8	1	2	0	0	41	49	2	49	50	2.04%	Sin reajuste
120	8	1	2	1	0	45	49	2	50	50	0.00%	Sin reajuste
120	8	1	2	1	0	46	51	1	52	50	3.85%	Sin reajuste
120	10	1	2	1	0	47	87	5	42	44	4.76%	Sin reajuste
120	10	1	2	1	0	60	45	4	43	42	2.33%	Sin reajuste
130	6	1	1	1	0	58	47	1	47	47	0.00%	Sin reajuste
130	6	1	1	1	0	52	49	0	49	51	4.08%	Sin reajuste
130	6	1	1	1	0	49	52	0	52	52	0.00%	Sin reajuste
130	8	1	1	1	1	47	41	1	41	41	0.00%	Sin reajuste
130	8	1	1	1	1	38	34	2	46	47	2.17%	Sin reajuste
130	8	1	1	1	0	53	49	13	50	48	4.00%	Sin reajuste
130	8	1	1	1	0	22	31	0	53	43	18.87%	Con reajuste
130	8	1	1	1	0	31	34	0	53	47	11.32%	Sin reajuste
130	8	1	1	1	1	28	56	4	69	62	10.14%	Sin reajuste

130	10	1	1	1	0	62	22	0	22	26	18.18%	Sin reajuste
130	10	1	1	1	1	60	22	1	30	28	6.67%	Sin reajuste
140	6	1	2	0	0	57	24	1	23	27	17.39%	Sin reajuste
140	6	1	2	0	0	42	36	0	30	32	6.67%	Con reajuste
140	6	1	2	1	0	43	29	3	31	32	3.23%	Sin reajuste
140	6	1	2	1	0	39	37	0	33	39	18.18%	Sin reajuste
140	6	1	2	1	1	55	40	11	48	46	4.17%	Sin reajuste
140	8	1	2	0	0	39	46	0	40	41	2.50%	Sin reajuste
140	8	1	2	0	1	31	42	9	45	44	2.22%	Sin reajuste
140	8	1	2	1	1	35	52	3	57	49	14.04%	Sin reajuste
140	10	2	2	1	0	21	18	0	20	17	15.00%	Sin reajuste
140	10	1	2	1	0	33	36	0	34	37	8.82%	Sin reajuste
140	10	1	2	0	0	31	38	0	39	39	0.00%	Sin reajuste
MAPE											6.89%	

Tabla 15. Resumen de asignaturas procesadas predicción 2018-II

Tipo	Cantidad	Porcentaje (%)
Con reajuste	6	12.50%
Sin reajuste	42	87.50%
Total	48	100.00%

Interpretación: La tabla nos muestra los resultados luego de ejecutar una muestra de registros del periodo académico 2018-II con nuestro modelo predictivo en Microsoft Azure y se apreció que el MAPE de los datos predichos son de 6.89%, un valor mucho menor al MAPE de la proyección del mismo periodo académico el cual fue de 23.37%, así mismo, el porcentaje de asignaturas que sufren reajuste con los datos de la predicción es de solo 12.5%, en comparación con los 56.25% que correspondió a los datos de la proyección de las mismas asignaturas.

e) Prueba de fiabilidad de los datos obtenidos

Se realizó la prueba de fiabilidad para los datos obtenidos, es decir entre los alumnos matriculados en el semestre 2018-II y los que nos arrojaba el modelo predictivo, esta prueba nos arroja una fiabilidad de 0.953, por lo que los datos son altamente fiables.

Alfa de Cronbach	N de elementos
,953	2

Gráfico 32. Fiabilidad de los datos de matriculas

Discusión: Los resultados nos permitió observar que el número de matriculados por el modelo Machine Learning tienen un buen porcentaje de igualdad en comparación con el número real de matriculados, siendo el algoritmo árbol de decisiones el que se utilizó para lograr nuestro objetivo, pues se logró una mejor precisión; por otro lado, Yamao (2018), determinó que el algoritmo de árboles de decisiones le permitió encontrar mejor predicción en el rendimiento académico al encontrar una exactitud de predicción de 82.87%.

4.5. Contrastación de hipótesis:

Se realizó la evaluación de la hipótesis, y para lo cual se siguieron los siguientes pasos:

- Se preparó la información del periodo académico 2019-II
 - Se usó el modelo predictivo Machine Learning de Microsoft Azure
 - Se preparó la contrastación de hipótesis
 - Se interpretó la hipótesis
- a. Se preparó la información del periodo académico 2019-II, para lo cual se utilizó un procedimiento almacenado para la base de datos de SIIGAA UNS para encontrar la información necesaria para el modelo predictivo con el periodo académico 2019-II, utilizando asignaturas ofertadas de los últimos

ciclos para aquellas especialidades cuyo plan curricular está vigente por más de cuatro años, también se excluyeron las asignaturas que el procedimiento almacenado arrojó como valor cero en la columna de alumnos habilitados.

Así mismo, primero se revisó la información referente a la proyección de los alumnos a matricularse de las mismas asignaturas según la carga lectiva del semestre 2019-II previa a la matrícula, del cual se obtuvo la siguiente información

Tabla 16. Resultados de proyección según carga lectiva del periodo académico 2019-II

Escuela	Ciclo	Matriculados	Proyección	APE	Reajuste
110	8	44	45	2.27%	Sin Reajuste
110	8	37	42	13.51%	Sin Reajuste
110	8	49	45	8.16%	Con Reajuste
110	8	52	47	9.62%	Sin Reajuste
110	8	55	42	23.64%	Con Reajuste
110	10	39	44	12.82%	Sin Reajuste
110	10	42	52	23.81%	Con Reajuste
110	10	48	48	0.00%	Sin Reajuste
110	10	40	46	15.00%	Con Reajuste
120	8	44	40	9.09%	Sin Reajuste
120	8	47	45	4.26%	Con Reajuste
120	8	42	45	7.14%	Sin Reajuste
120	10	48	43	10.42%	Con Reajuste
120	10	51	42	17.65%	Con Reajuste
130	8	60	30	50.00%	Con Reajuste

130	8	41	30	26.83%	Con Reajuste
130	8	53	46	13.21%	Sin Reajuste
130	8	40	30	25.00%	Con Reajuste
130	8	51	50	1.96%	Sin Reajuste
130	10	57	22	61.40%	Con Reajuste
130	10	52	60	15.38%	Sin Reajuste
140	8	12	58	383.33%	Con Reajuste
140	8	23	30	30.43%	Sin Reajuste
140	8	31	21	32.26%	Con Reajuste
140	10	25	38	52.00%	Con Reajuste
140	10	22	30	36.36%	Sin Reajuste
140	10	42	16	61.90%	Con Reajuste
140	10	44	20	54.55%	Con Reajuste
210	8	31	39	25.81%	Sin Reajuste
210	8	27	36	33.33%	Con Reajuste
210	10	32	32	0.00%	Sin Reajuste
230	8	27	20	25.93%	Sin Reajuste
310	8	38	39	2.63%	Sin Reajuste
310	8	35	40	14.29%	Sin Reajuste
310	8	39	40	2.56%	Sin Reajuste
340	8	36	33	8.33%	Sin Reajuste
340	8	31	40	29.03%	Sin Reajuste
340	8	36	38	5.56%	Sin Reajuste
340	8	39	41	5.13%	Sin Reajuste
340	8	39	43	10.26%	Sin Reajuste
340	10	45	42	6.67%	Sin Reajuste
340	10	34	33	2.94%	Sin Reajuste
MAPE				27.96%	

Tabla 17. Resumen de asignaturas sobre proyección 2019-II

Tipo	Cantidad	Porcentaje (%)
Con reajuste	17	40.48%
Sin reajuste	25	59.52%
Total	42	100.00%

- b. Se usó el modelo predictivo Machine Learning de Microsoft Azure, el cual fue el resultante del objetivo específico cuatro, y con la información preparada como datos de entrada, se obtuvo la siguiente información

Tabla 18. Resultados de predicción periodo académico 2019-II

Escuela	Ciclo	Tipo Asignatura	Plan Curricular	Convalidación	Condición	Mat Anterior	Habilitados	Desaprobados	Matriculados	Azure	APE	Reajuste
110	8	1	1	1	0	49	41	15	44	39	11.36%	Sin reajuste
110	8	1	1	1	0	32	34	3	37	35	5.41%	Sin reajuste
110	8	1	1	1	1	47	44	4	49	51	4.08%	Sin reajuste
110	8	1	1	1	1	66	50	8	52	54	3.85%	Sin reajuste
110	8	1	1	1	0	47	53	0	55	56	1.82%	Sin reajuste
110	10	1	1	1	0	42	49	0	39	48	23.08%	Con reajuste
110	10	1	1	1	1	42	40	4	42	36	14.29%	Sin reajuste
110	10	1	1	1	1	47	45	1	48	53	10.42%	Sin reajuste
110	10	1	1	1	0	27	39	1	40	37	7.50%	Sin reajuste
120	8	1	2	0	1	49	47	1	44	51	15.91%	Con reajuste
120	8	1	2	1	1	50	47	1	47	51	8.51%	Sin reajuste
120	8	1	2	1	0	52	46	0	42	44	4.76%	Sin reajuste
120	10	1	2	1	0	43	46	0	48	39	18.75%	Con reajuste
120	10	1	2	1	1	42	88	2	51	47	7.84%	Sin reajuste

130	8	1	1	1	1	53	43	6	60	52	13.33%	Sin reajuste
130	8	1	1	1	0	53	40	0	41	46	12.20%	Con reajuste
130	8	1	1	1	1	46	43	1	53	49	7.55%	Sin reajuste
130	8	1	1	1	0	50	39	1	40	38	5.00%	Sin reajuste
130	8	1	1	1	1	69	50	15	51	52	1.96%	Sin reajuste
130	10	1	1	1	1	29	46	4	57	42	26.32%	Con reajuste
130	10	1	1	1	0	22	46	0	52	39	25.00%	Con reajuste
140	8	2	2	0	0	21	12	0	12	23	91.67%	Con reajuste
140	8	1	2	1	1	57	21	3	23	36	56.52%	Con reajuste
140	8	1	2	0	1	40	41	1	31	45	45.16%	Sin reajuste
140	10	2	2	1	1	17	22	1	25	22	12.00%	Sin reajuste
140	10	2	2	1	0	20	23	1	22	23	4.55%	Sin reajuste
140	10	1	2	0	1	39	41	2	42	44	4.76%	Sin reajuste
140	10	1	2	1	0	34	44	0	44	42	4.55%	Sin reajuste
210	8	1	2	1	1	34	34	1	31	37	19.35%	Sin reajuste
210	8	1	2	1	0	32	27	0	27	30	11.11%	Sin reajuste
210	10	1	2	1	0	34	33	1	32	35	9.38%	Sin reajuste
230	8	2	2	0	1	20	27	3	27	18	33.33%	Sin reajuste
310	8	1	2	1	0	56	34	2	38	33	13.16%	Sin reajuste
310	8	1	2	1	0	37	34	0	35	37	5.71%	Sin reajuste
310	8	1	2	1	0	39	39	0	39	39	0.00%	Sin reajuste
340	8	1	2	0	1	44	27	0	36	32	11.11%	Sin reajuste
340	8	1	2	1	0	39	33	1	31	33	6.45%	Sin reajuste
340	8	1	2	0	1	37	34	1	36	38	5.56%	Sin reajuste
340	8	1	2	1	1	42	34	5	39	41	5.13%	Sin reajuste
340	8	1	2	0	0	44	40	1	39	39	0.00%	Sin reajuste
340	10	1	2	0	1	34	42	2	45	42	6.67%	Sin reajuste
340	10	1	2	0	0	26	34	1	34	33	2.94%	Sin reajuste
MAPE											13.76%	

Tabla 19. Resumen de asignaturas sobre predicción 2019-II

Tipo	Cantidad	Porcentaje (%)
Con reajuste	8	19.05%
Sin reajuste	34	80.95%
Total	42	100.00%

El modelo predictivo logró obtener un MAPE de 13.76% siendo menor al 27.96% del MAPE obtenido sobre los valores proyectados en la elaboración de la carga lectiva del semestre 2019-II; así mismo, el número de asignaturas que debieron sufrir cambios en el número de grupos de acuerdo a los valores predichos es de 19.05% el cual es menor al 40.48% que se obtuvo en base a los valores proyectados.

c. Se preparó la contrastación de hipótesis, para ello se utilizó el software SPSS

- Definición de variables

CAM_A : Cantidad de alumnos matriculados en el periodo académico 2019-II real, sin aplicar Modelo predictivo.

CAM_P : Cantidad de alumnos matriculados en el periodo académico 2019-II aplicando Modelo predictivo.

- Hipótesis estadística

Debido a que las muestras no son independientes, dado que se obtienen muestras con las mismas variables en dos momentos determinados, se usará un contraste bilateral

H_0 : Machine Learning no permite predecir la cantidad de alumnos de pregrado a matricularse por asignatura en el periodo académico 2019-II en la Universidad Nacional del Santa

$$H_0 = CAM_A - CAM_P \neq 0$$

H₁: Machine Learning permite predecir la cantidad de alumnos de pregrado a matricularse por asignatura en el periodo académico 2019-II en la Universidad Nacional del Santa

$$H_1 = CAM_A - CAM_p = 0$$

- Prueba de fiabilidad:

Haciendo uso de la aplicación SPSS se busca la confiabilidad de los datos obtenidos para las variables involucradas

➔ **Fiabilidad**

[ConjuntoDatos4] C:\Users\arkan\Desktop\doc\hipotesis.sav

Escala: ALL VARIABLES

Resumen de procesamiento de casos

		N	%
Casos	Válido	42	100,0
	Excluido ^a	0	,0
	Total	42	100,0

a. La eliminación por lista se basa en todas las variables del procedimiento.

Estadísticas de fiabilidad

Alfa de Cronbach	N de elementos
,891	2

Gráfico 33. Vista de la estadística de fiabilidad

El cual nos arroja un alfa de Crombach de 0.891, encontrándose en la escala de muy aceptable.

- Prueba de normalidad:

Se debe usar:

Kolmogórov-Smirnov para muestras grandes (>30 individuos)

Chápiro Wilk para muestras pequeñas (< 30 individuos)

Criterio para determinar Normalidad:

P-valor $\geq \alpha$ Aceptar H_0 = Los datos provienen de una distribución normal

P-valor $< \alpha$ Aceptar H_1 = Los datos NO provienen de una distribución normal

Aplicando SPSS:

Se busca determinar si los datos son normales.

Pruebas de normalidad

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
CAMa	,080	42	,200 [*]	,985	42	,854
CAMp	,082	42	,200 [*]	,968	42	,276

*. Esto es un límite inferior de la significación verdadera.

a. Corrección de significación de Lilliefors

Gráfico 34. Vista de la prueba de normalidad**Verificando datos para normalidad:**

P-Valor (Matriculados reales) = 0.200 > $\alpha = 0.05$

P-Valor (Matriculados Según ML) = 0.200 > $\alpha = 0.05$

Conclusión:

Los datos del número de alumnos matriculados provienen de una distribución normal.

- Nivel de Significancia:

Se usó un nivel de significancia ($\alpha = 0.05$) del 5%. Por lo tanto, el nivel de confianza ($1-\alpha = 0.95$) será del 95%

- Valores Tabulados

Se estimó una muestra de 42 asignaturas ofertadas en periodo académico 2019-II, que corresponden a las asignaturas no electivas

de los últimos dos ciclos de las especialidades cuyos planes curriculares tienen vigencia en más de tres años, como se muestran en la Tabla 16.

- Resultados de la Hipótesis Estadística
Para esto se utilizó la aplicación SPSS, utilizando prueba T para muestras relacionadas, del cual se obtuvo el siguiente resultado.

➔ **Prueba T**

Estadísticas de muestras emparejadas

		Media	N	Desv. Desviación	Desv. Error promedio
Par 1	CAMa	40,0000	42	10,28141	1,58646
	CAMp	39,9933	42	9,05313	1,39693

Correlaciones de muestras emparejadas

		N	Correlación	Sig.
Par 1	CAMa & CAMp	42	,809	,000

Prueba de muestras emparejadas

		Diferencias emparejadas							
		Media	Desv. Desviación	Desv. Error promedio	95% de intervalo de confianza de la diferencia		t	gl	Sig. (bilateral)
					Inferior	Superior			
Par 1	CAMa - CAMp	,00667	6,08474	,93890	-1,88947	1,90280	,007	41	,994

Gráfico 35. Vista SPSS para análisis de la hipótesis

Criterio para decisión es:

Si la probabilidad obtenida P-Valor $> \alpha$, rechace H_0 (se acepta H_1)

Si la probabilidad obtenida P-Valor $\leq \alpha$, no rechace H_0 (se acepta H_0)

De la Prueba en SPSS:

$$P\text{-Valor} = 0.994 > \alpha = 0.05$$

- d. Se contrastó la hipótesis, dando como resultado un valor de 0.994, siendo este valor mayor a 0.05 se rechaza H_0 , es decir se acepta H_1 . Concluyendo que Machine Learning permite predecir la cantidad de alumnos de pregrado a matricularse por asignatura en el periodo académico 2019-II en la Universidad Nacional del Santa

CAPÍTULO V

CONCLUSIONES Y RECOMENDACIONES

5.1. Conclusiones.

El presente trabajo de investigación nos lleva a las siguientes conclusiones:

- Se identificó, según el planteamiento del objetivo 01, que los algoritmos Machine Learning más adecuados para elaborar un modelo predictivo que nos permita predecir la cantidad de alumnos matriculados por asignatura en un periodo académico son los algoritmos de regresión como: Regresión Lineal, Red neuronal y Árbol de decisiones, lo cual se logró al hacer un análisis de los tipos de técnicas y algoritmos, y de la realización de una encuesta, estadísticamente significativa, en la que el 80% de docentes encuestados opinaron que dichos algoritmos eran adecuados. El objetivo es similar al antecedente “*Predicción de rendimiento académico mediante regresión y redes neuronales en los estudiantes de la escuela profesional de Ingeniería Estadística e Informática de la Universidad Nacional del Altiplano*” en la que uso el algoritmo de red neuronal con el cual obtuvo mejor desempeño en su predicción del rendimiento académico, así como en el antecedente “*Predicción del rendimiento académico mediante minería de datos en estudiantes del primer ciclo de la escuela profesional de Ingeniería de Computación y Sistemas, Universidad de San Martín de Porres, Lima-Perú*” en la que regresión y arboles de decisiones fueron usados en la predicción del rendimiento académico y también se recomienda el uso de redes neuronales.

- Se identificó a Microsoft Azure, en el cumplimiento del objetivo 02, como la plataforma en la nube más adecuada para el estudio y desarrollo de modelos predictivos, identificando que la institución no cuenta con una adecuada infraestructura física y es necesario optar por una plataforma en la nube, además, que el 40% de los docentes encuestados lo consideraron como adecuado. El objetivo es similar al antecedente “*Estudio técnico de las características y la oferta de servicios tecnológicos digitales basadas en tecnología de Machine Learning en la gran área metropolitana de costa*”

rica en el período mayo 2017 a junio 2018” en la que destacan a Microsoft Azure para tareas de Machine Learning.

- Se cumplió el objetivo 03, en la cual se determinó, haciendo uso de Weka, que, de las 20 variables analizadas inicialmente con el apoyo de los Departamentos al consultar sobre las variables que intervienen para proyectar el número de alumnos a matricularse en una asignatura, solo 10 de ellas tienen una relación significativa para la construcción del modelo predictivo. Este objetivo utilizó la misma herramienta para el análisis y selección de las mejores variables predictoras usadas en los antecedentes *“Predicción del fracaso y el abandono Escolar mediante técnicas de minería de Datos”* y *“Predicción del rendimiento académico aplicando técnicas de minería de datos”*
- Se logró cumplir con el objetivo 04, para lo cual se elaboró una solución Machine Learning haciendo uso de la plataforma en la nube de Microsoft Azure y las variables predictoras proporcionadas por Weka, además, se hizo la comparación de los algoritmos identificados en el objetivo específico 01 así como del algoritmo bosque de decisiones, observándose que este último tiene un coeficiente de determinación de 0.7858, el cual es mejor que el de los otros algoritmos. Con esta solución se hizo el testeo del periodo académico 2018-II logrando obtener un MAPE del 6.89% y que un 12.5% de asignaturas que deberían sufrir reajuste luego de la matrícula, el cual es un porcentaje menor a los obtenidos sin la solución Machine Learning (MAPE de 23.37% y 56.25% de asignaturas que debieron sufrir reajuste en el semestre 2018-II). Este objetivo logró un enfoque similar al antecedente *“Predicción del rendimiento académico mediante minería de datos en estudiantes del primer ciclo de la escuela profesional de Ingeniería de Computación y Sistemas, Universidad de San Martín de Porres, Lima-Perú”* en el que con el algoritmo de árboles de decisiones obtuvieron una mejor predicción del rendimiento académico.
- Por último, con la solución Machine Learning obtenida en el objetivo específico 04, se logró predecir la cantidad de alumnos matriculados para

el periodo académico 2019-II, según se demuestra con la contrastación de la hipótesis y con un nivel de confianza del 95%, obteniéndose un MAPE del 13.76% y solo un 19.05% de asignaturas debieran sufrir reajuste luego del proceso de matrícula, en comparación con el MAPE de 27.96% que se obtuvo con los valores proyectados por los departamentos en las mismas asignaturas y hasta un 40.48% de las asignaturas deberán sufrir algún tipo de reajuste con el número real de alumnos matriculados, por lo que la solución Machine Learning logra una mejor predicción y reduce el número de asignaturas con reajuste de números de grupos.

5.2. Recomendaciones

- Un estudio más amplio del ámbito académico como de la reglamentación de cada escuela puede llevarnos a encontrar nuevas variables para el modelo predictivo, así como el uso de las convalidaciones para los planes curriculares, los cuales deberán estudiarse y determinar si afectan significativamente al modelo resultante del presente trabajo.
- La existencia de otras plataformas que brindan similar características y servicios que Microsoft Azure nos puede ampliar el estudio del presente trabajo al utilizar y comparar los resultados con por ejemplo IBM Watson o Google Cloud.
- Elaborar un sistema informático web que permita utilizar el servicio web resultante que proporciona Microsoft Azure y poder utilizarlo en la elaboración de la carga lectiva propuesta el cual utiliza la proyección del número de alumnos a matricularse para determinar el número de grupos de teoría y de laboratorio para las asignaturas ofertadas en un proceso de matrícula.

REFERENCIAS BIBLIOGRÁFICAS

- Altexsoft (2019). *Comparing Machine Learning as a Service: Amazon, Microsoft Azure, Google Cloud AI, IBM Watson*. Recuperado el 27 de septiembre 2019, de <https://www.altexsoft.com/blog/datascience/comparing-machine-learning-as-a-service-amazon-microsoft-azure-google-cloud-ai-ibm-watson/>
- Ayyadevara K. (2018). *Pro Machine Learning Algorithms*. Apress.
- Barga R., Fontama V. y Hyong Tok W. (2015). *Análisis predictivo con Microsoft Azure Machine Learning*. Apress.
- Candia Oviedo, D.I. (2019). Predicción del rendimiento académico de los estudiantes de la UNSAAC a partir de sus datos de ingreso utilizando algoritmos de aprendizaje automático [Tesis de maestría, Universidad Nacional de San Antonio Abad del Cusco]. <http://hdl.handle.net/20.500.12918/4120>
- Gartner (2020). *Magic Quadrant for Cloud AI Developer Services*. Recuperado el 24 de febrero 2020, de <https://www.gartner.com/doc/reprints?id=1-1YCWP1OB&ct=200213&st=sb>
- Grández Márquez, M. A. (2017). *Aplicación de minería de datos para determinar patrones de consumo futuro en clientes de una distribuidora de suplementos nutricionales* [Tesis título profesional, Universidad San Ignacio de Loyola]. <https://repositorio.usil.edu.pe/handle/usil/2763>
- Gamarra Gomez, F. (2019). Modelo basado en Machine Learning para el neurorendimiento académico de estudiantes universitarios. *Revista ciencia y tecnología para el desarrollo – UJCM*, 5(9), 10-18. <https://revistas.ujcm.edu.pe/index.php/rctd/article/view/137/121>
- Kashyap P. (2017). *Machine Learning for Decision Makers*. Apress.
- Khan S. (2016). *Análisis predictivo: el conocimiento es poder*. Recuperado el 24 de octubre 2018, de <https://www.itworldcanada.com/blog/predictive-analytics-knowledge-is-power/387243>
- Kramer O. (2016). *Machine Learning for Evolution Strategies*. Springer.
- Kuperman J., Naselli F. y Gabaldón J. (2019). Machine Learning, Inteligencia artificial y Big Data, Lo que todo directivo debe conocer. *Digital Basics* N°4, 2019, Asociación española de directivos. <https://www.diarioabierto.es/wp->

content/uploads/2019/02/Machine-Learning-Inteligencia-Artificial-y-Big-Data.-Lo-que-todo-directivo-debe-saber.pdf

- Márquez Vera, C. (2015). *Predicción del fracaso y el abandono Escolar mediante técnicas de minería de Datos* [Tesis doctoral no publicada]. Universidad de Córdoba.
- Menacho Chiok, C.H. (2016). *Predicción del rendimiento académico aplicando técnicas de minería de datos* [Trabajo académico, Universidad Nacional Agraria la Molina]. <https://doi.org/10.21704/ac.v78i1.811>
- Microsoft, (2020). *Selección de algoritmos para Azure Machine Learning*. Recuperado el 07 de mayo 2020, de <https://docs.microsoft.com/es-es/azure/machine-learning/how-to-select-algorithms>
- Paja Dominguez, H.D. (2017). *Predicción de rendimiento académico mediante regresión y redes neuronales en los estudiantes de la escuela profesional de Ingeniería Estadística e Informática de la Universidad Nacional del Altiplano* [Tesis título profesional, Universidad Nacional del Altiplano]. <http://repositorio.unap.edu.pe/handle/UNAP/6337>
- Ramasubramanian K. y Singh A. (2017). *Machine Learning Using R*. Apress.
- Rodríguez Suarez, H.Z. (2018). Estudio de herramientas basadas en IA Cloud y su aplicación en el desarrollo de las actividades académicas de la carrera de telemática de la Universidad de Guayaquil. [Tesis título profesional, Universidad de Guayaquil]. <http://repositorio.ug.edu.ec/handle/redug/42167>
- Sotiropoulos D. y Tsihrintzis G. (2017). *Machine Learning Paradigms*. Springer.
- Yamao, Eiriku (2018). *Predicción del rendimiento académico mediante minería de datos en estudiantes del primer ciclo de la escuela profesional de Ingeniería de Computación y Sistemas, Universidad de San Martín de Porres, Lima-Perú* [Tesis de maestría, Universidad de San Martín de Porres]. <https://hdl.handle.net/20.500.12727/3555>
- Zúñiga Ugalde, M.A. (2019). *Estudio técnico de las características y la oferta de servicios tecnológicos digitales basadas en tecnología de Machine Learning en la gran área metropolitana de costa rica en el período mayo 2017 a junio 2018* [Tesis título profesional, Universidad Técnica Nacional]. <https://repositorio.utn.ac.cr/handle/123456789/245?show=full>

ANEXOS

1. Formato para Lista de expertos

Nro	Apellidos y nombres	Grado académico	Especialidad	Cargo	Institución	Teléfono
01						
02						
03						
04						
05						
06						
07						
08						
09						
10						

2. Formato para evaluación de expertos

Estimado profesional, usted ha sido invitado a participar en el proceso de evaluación de un instrumento para investigación en humanos. En razón a ello se le alcanza el instrumento motivo de evaluación y el presente formato que servirá para que usted pueda hacernos llegar sus apreciaciones para cada ítem del instrumento de investigación. Agradecemos de antemano sus aportes que permitirán validar el instrumento y obtener información válida, criterio requerido para toda investigación. A continuación, sírvase identificar el ítem o pregunta y conteste marcando con un aspa en la casilla que usted considere conveniente y además puede hacernos llegar alguna otra apreciación en la columna de observaciones.

Ítem	Validez de contenido		Validez de constructo		Validez de criterio		Observaciones
	El ítem corresponde a alguna dimensión de la variable		El ítem contribuye a medir el indicador planteado		El ítem permite clasificar a los sujetos en las categorías establecidas		
	Si	No	Si	No	Si	No	
01							
02							
03							
04							

Nombre:.....

DNI:.....

3. Encuesta Machine Learning

Ítem 01: Algoritmos Machine Learning

Instrucciones:

A continuación, se presenta un cuadro con algunos algoritmos, los mismos que se pretende utilizar para lograr la predicción de alumnos de pregrado matriculados por asignatura en un periodo académico, en la Universidad Nacional del Santa.

Seleccione tres algoritmos que considera serían más eficaces para lograr la predicción descrita anteriormente, para ello marque con una equis (X) en la columna Selección

[Nota: Las variables predictoras serán tipo numérica]

Algoritmos Machine Learning		
Nro.	Algoritmo	Selección
01	Agrupación	
02	Árbol de decisiones	
03	Bayesiana lineal	
04	Máquina de vectores	
05	Red neuronal	
06	Redes bayesianas	
07	Regresión logística	
08	Regresión lineal	
09	Selva de decisión	
10	Especifique:.....	

Ítem 02: Plataforma Machine Learning

Indicaciones:

A continuación, se presenta un cuadro con algunas plataformas que utilizan machine Learning, los mismos que se pretende utilizar para lograr la predicción de alumnos de pregrado matriculados por asignatura en un periodo académico, en la Universidad del Santa.

Seleccione una plataforma que considera serían más eficaces para lograr la predicción descrita anteriormente, para ello marque con una equis (X) en la columna Selección

Plataformas Machine Learning		
Nro.	Nombre	Selección
01	Amazon Machine Learning	
02	Microsoft Azure Machine Learning	
03	IBM Watson Machine Learning	
04	Google Cloud Machine Learning	
05	Especifique:.....	

4. Análisis de fiabilidad de encuesta Machine Learning

a) Respecto al ítem 01:

Para el estudio de la fiabilidad del ítem 01 se agruparán por algoritmos, debido a que de lo contrario el software SPSS encontrará varianza 0 para nuestro estudio de fiabilidad.

Para el objetivo específico 01 se utiliza el ítem 01 del cuestionario, para dar forma a nuestros datos y poder ingresarlo al software SPSS se define los siguientes valores:

- En caso sea escogido el algoritmo en la encuesta, se colocará un valor 1
- En caso no sea escogido el algoritmo en la encuesta, se colocará el valor 0

La siguiente tabla nos muestra el resultado de las 10 encuestas realizadas a docentes.

Tabla 1. Valores para algoritmos según encuestas

Algoritmos	Encuestas									
	E01	E02	E03	E04	E05	E06	E07	E08	E09	E10
Agrupación	0	0	0	1	0	0	0	0	0	0
Árbol de decisiones	0	1	1	1	1	1	1	0	1	1
Bayesiana lineal	0	0	0	0	0	0	0	0	0	0
Máquina de vectores	0	0	0	0	1	0	0	0	0	0
Red neuronal	1	1	0	0	0	1	0	1	0	1
Redes bayesianas	0	1	0	0	1	0	1	1	0	0
Regresión logística	1	0	1	0	0	0	0	0	1	0
Regresión lineal	1	0	1	1	0	1	1	1	1	1
Selva de decisión	0	0	0	0	0	0	0	0	0	0
Otros	0	0	0	0	0	0	0	0	0	0

b) Respecto al ítem 02:

Para el ítem 02, a cada plataforma se le asignó los siguientes valores

Tabla 2. Valores de asignación para plataformas

Plataforma	Valor
Amazon	1
Microsoft Azure	2
IBM Watson	3
Google Cloud	4
Otros	5

Por lo que habrá que agregar la siguiente fila para el tratamiento final

Tabla 3. Valores según encuesta

Ítem	Encuestas									
	E01	E02	E03	E04	E05	E06	E07	E08	E09	E10
Plataforma	4	4	3	2	3	1	2	4	2	2

c) Fiabilidad del cuestionario final:

Se utilizó el programa SPSS vr. 25 para el análisis de Alfa de Crombach para el estudio de fiabilidad del Objetivo específico 01.

Archivo Editar Ver Datos Transformar Analizar Gráficos Utilidades Ampliaciones Ventana Ayuda

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	Pregunta	Cadena	19	0		Ninguno	Ninguno	19	Izquierda	Nominal	Entrada
2	E01	Numérico	8	0		Ninguno	Ninguno	5	Derecha	Escala	Entrada
3	E02	Numérico	8	0		Ninguno	Ninguno	5	Derecha	Escala	Entrada
4	E03	Numérico	8	0		Ninguno	Ninguno	5	Derecha	Escala	Entrada
5	E04	Numérico	8	0		Ninguno	Ninguno	5	Derecha	Escala	Entrada
6	E05	Numérico	8	0		Ninguno	Ninguno	5	Derecha	Escala	Entrada
7	E06	Numérico	8	0		Ninguno	Ninguno	5	Derecha	Escala	Entrada
8	E07	Numérico	8	0		Ninguno	Ninguno	5	Derecha	Escala	Entrada
9	E08	Numérico	8	0		Ninguno	Ninguno	5	Derecha	Escala	Entrada
10	E09	Numérico	8	0		Ninguno	Ninguno	5	Derecha	Escala	Entrada
11	E10	Numérico	8	0		Ninguno	Ninguno	5	Derecha	Escala	Entrada
12											

Vista de datos Vista de variables

Gráfico 36. Vista de variables Encuesta Machine Learning

Archivo Editar Ver Datos Transformar Analizar Gráficos Utilidades Ampliaciones Ventana Ayuda

17 : Visible: 11 de 11 variables

	Pregunta	E01	E02	E03	E04	E05	E06	E07	E08	E09	E10
1	Agrupacion	0	0	0	1	0	0	0	0	0	0
2	Arbol de decisiones	0	1	1	1	1	1	1	0	1	1
3	Bayesiana lineal	0	0	0	0	0	0	0	0	0	0
4	Maquina de vectores	0	0	0	0	1	0	0	0	0	0
5	Red neuronal	1	1	0	0	0	1	0	1	0	1
6	Redes bayesianas	0	1	0	0	1	0	1	1	0	0
7	Regresion logistica	1	0	1	0	0	0	0	0	1	0
8	Regresion lineal	1	0	1	1	0	1	1	1	1	1
9	Selva de decision	0	0	0	0	0	0	0	0	0	0
10	Otros	0	0	0	0	0	0	0	0	0	0
11	Plataforma	4	4	3	2	3	1	2	4	2	2

Vista de datos Vista de variables

Gráfico 37. Vista de datos Encuesta Machine Learning

Estadísticas de fiabilidad

Alfa de Cronbach	Alfa de Cronbach basada en elementos estandarizados	N de elementos
,962	,969	10

Gráfico 38. Alfa de Cronbach Encuesta Machine Learning

El resultado del Alfa de Cronbach fue de 0.962 que es más próximo a 1 con lo que se puede concluir que el instrumento es muy fiable.

5. Variables para número de matriculados

Instrucciones:

A continuación, se describen una lista de variables del ámbito académico que se intenta determinar si tienen algún grado de influencia al momento de proyectar el número de alumnos que se matricularan en una asignatura y de esa manera determinar el número de grupos de teoría y de laboratorio para el proceso de matrícula.

El grado de influencia está representado por los siguientes valores:

Ninguna	Muy poco	Poco	Regular	Alto	Muy alto
0	1	2	3	4	5

[En donde cero (0) significa que no tiene ninguna influencia al momento de determinar el número de alumnos posibles para matricularse en una asignatura, y cinco (5) significa que tiene una muy alta influencia para determinar el número de alumnos posibles a matricularse].

Nro.	Variable	Influencia
01	Identificador de la asignatura	
02	Escuela profesional	
03	Facultad	
04	Docente de la asignatura	
05	Tipo de docente de la asignatura	
06	Ciclo de la asignatura	
07	Tipo de asignatura (los tipos de asignatura son: especialidad, electivo o autodesarrollo)	
08	Plan curricular de la asignatura	
09	Crédito de la asignatura	
10	Horas de teoría de la asignatura	
11	Horas prácticas de la asignatura	
12	Asignatura convalidable	
13	Periodo académico de matrícula	
14	Tipo de matrícula	
15	Condición de la matrícula	
16	Número de matriculados en el semestre anterior	
17	Pre requisitos de la asignatura	
18	Número de alumnos habilitados (Alumnos que pueden llevar la asignatura, es decir aprobaron pre requisitos)	
19	Alumnos que desaprobaron la asignatura en el semestre anterior	

6. Análisis de fiabilidad de encuesta Variables

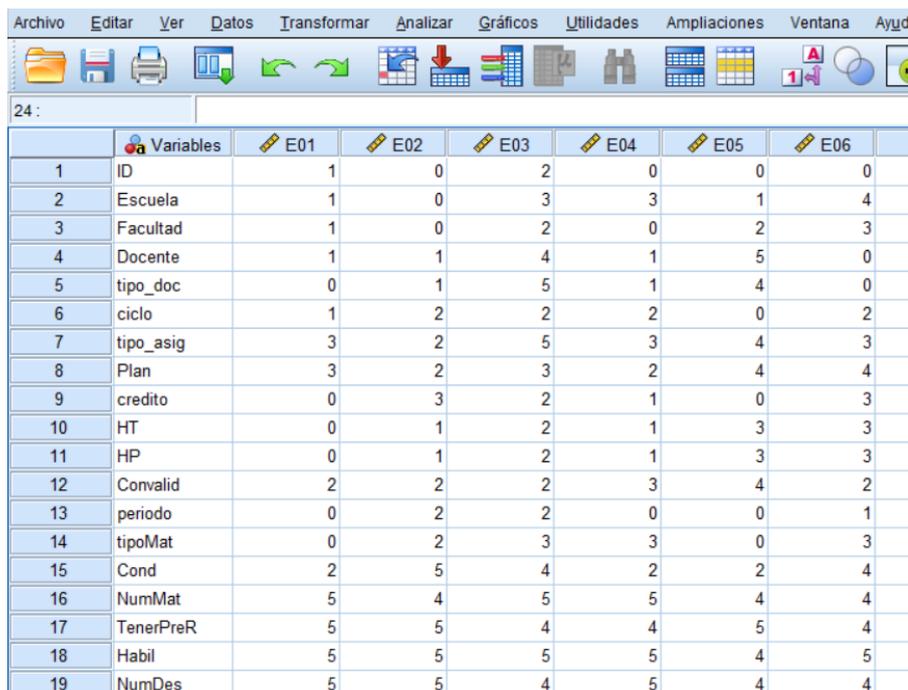
La encuesta se realizó a 6 jefes de departamento de la Universidad Nacional del Santa, quienes están involucrados con la elaboración de la carga lectiva propuesta para el proceso de matrícula y para lo cual deben proyectar el número de alumnos a matricularse en una asignatura y poder determinar el número de grupos de teoría y laboratorios para cada asignatura.

Para realizar la fiabilidad del instrumento usaremos SPSS v25, para lo cual se muestra la imagen que corresponde a la vista de variables



	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	Variables	Cadena	9	0		Ninguno	Ninguno	9	Izquierda	Nominal	Entrada
2	E01	Numérico	8	0		Ninguno	Ninguno	8	Derecha	Escala	Entrada
3	E02	Numérico	8	0		Ninguno	Ninguno	8	Derecha	Escala	Entrada
4	E03	Numérico	8	0		Ninguno	Ninguno	8	Derecha	Escala	Entrada
5	E04	Numérico	8	0		Ninguno	Ninguno	8	Derecha	Escala	Entrada
6	E05	Numérico	8	0		Ninguno	Ninguno	8	Derecha	Escala	Entrada
7	E06	Numérico	8	0		Ninguno	Ninguno	8	Derecha	Escala	Entrada
8											

Gráfico 40. Vista de variables Encuesta Variables



24 :	Variables	E01	E02	E03	E04	E05	E06
1	ID	1	0	2	0	0	0
2	Escuela	1	0	3	3	1	4
3	Facultad	1	0	2	0	2	3
4	Docente	1	1	4	1	5	0
5	tipo_doc	0	1	5	1	4	0
6	ciclo	1	2	2	2	0	2
7	tipo_asig	3	2	5	3	4	3
8	Plan	3	2	3	2	4	4
9	credito	0	3	2	1	0	3
10	HT	0	1	2	1	3	3
11	HP	0	1	2	1	3	3
12	Convalid	2	2	2	3	4	2
13	periodo	0	2	2	0	0	1
14	tipoMat	0	2	3	3	0	3
15	Cond	2	5	4	2	2	4
16	NumMat	5	4	5	5	4	4
17	TenerPreR	5	5	4	4	5	4
18	Habil	5	5	5	5	4	5
19	NumDes	5	5	4	5	4	4

Gráfico 39. Vista de datos Encuesta Variables

Estadísticas de fiabilidad

Alfa de Cronbach	Alfa de Cronbach basada en elementos estandarizados	N de elementos
,880	,882	6

Gráfico 41. Alfa de Cronbach Encuesta Variables

El resultado del Alfa de Cronbach fue de 0.880 que es más próximo a 1 con lo que se puede concluir que el instrumento es fiable.

MACHINE LEARNING PARA PREDECIR LA CANTIDAD DE ALUMNOS DE PREGRADO A MATRICULARSE POR ASIGNATURA EN EL PERIODO ACADÉMICO 2019-II EN LA UNIVERSIDAD NACIONAL DEL SANTA

Fecha de entrega: 06-dic-2022 12:30p.m. (UTC-0500)
por Humberto Angel Ninaquispe Matame

Identificador de la entrega: 1973353569

Nombre del archivo: tesis_ninaquispe_082022.docx (9.22M)

Total de palabras: 17104

Total de caracteres: 87842



MACHINE LEARNING PARA PREDECIR LA CANTIDAD DE ALUMNOS DE PREGRADO A MATRICULARSE POR ASIGNATURA EN EL PERIODO ACADÉMICO 2019-II EN LA UNIVERSIDAD NACIONAL DEL SANTA

INFORME DE ORIGINALIDAD

5%

INDICE DE SIMILITUD

%

FUENTES DE INTERNET

5%

PUBLICACIONES

%

TRABAJOS DEL ESTUDIANTE


DR. LUIS VLADIMÍR URRELO HUIMAN

FUENTES PRIMARIAS

- 1** Brenda Díaz-Landa, Rosana Meleán-Romero, William Marín-Rodríguez. "Rendimiento académico de estudiantes en Educación Superior: predicciones de factores influyentes a partir de árboles de decisión", Telos Revista de Estudios Interdisciplinarios en Ciencias Sociales, 2021

Publicación

1%
- 2** Eduardo Javier Pozo Burgos, Marco Rubén Burbano Pulles, Jack Iván Vidal Chica, Gabriela Elizabeth Revelo Salgado. "Sociocultural and demographic factors that influence academic performance: The pre-university case of the Universidad Politécnica Estatal del Carchi", Journal of Technology and Science Education, 2022

Publicación

1%

3

Fatima Farid Petiwala, Vinod Kumar Shukla, Sonali Vyas. "Chapter 15 IBM Watson: Redefining Artificial Intelligence Through Cognitive Computing", Springer Science and Business Media LLC, 2021

Publicación

4

Mario Aquino Cruz, Jose Abdon Sotomayor Chahuaylla, Leonardo Davila Huacoto, Edgar Eloy Carpio Vargas et al. "Yachay-Miray: web application for teaching learning of the multiplication of natural numbers", 2020 XV Conferencia Latinoamericana de Tecnologías de Aprendizaje (LACLO), 2020

Publicación

5

van Lidth de Jeude, Marije, Oliver Schütte, and Florencia Quesada. "The vicious circle of social segregation and spatial fragmentation in Costa Rica's greater metropolitan area", Habitat International, 2016.

Publicación

6

Pamela Mishelle Ríos-Carrión, Katina Vanessa Bermeo-Pazmiño, Cecilia Ivonne Narváez-Zurita. "Inteligencia de negocios como estrategia para la toma de decisiones en una empresa financiera", CIENCIAMATRIA, 2021

Publicación

7

Amelec Vilorio, Jesús García Guliany, William Niebles Núñez, Hugo Hernández Palma,

<1 %

<1 %

<1 %

<1 %

<1 %

Leonardo Niebles Núñez. "Data Mining Applied in School Dropout Prediction", Journal of Physics: Conference Series, 2020

Publicación

8

CHRISTOPHER R. COLE. "Correlation of Impedance Minute Ventilation with Measured Minute Ventilation in a Rate Responsive Pacemaker", Pacing and Clinical Electrophysiology, 6/2001

Publicación

<1 %

9

"Data Mining and Big Data", Springer Science and Business Media LLC, 2018

Publicación

<1 %

10

A. Menacho Villa, C. Perez Molina, Manuel Castro. "Students' Behavior When Connecting to the LMS: A Case Study at UNED", IEEE Revista Iberoamericana de Tecnologías del Aprendizaje, 2019

Publicación

<1 %

11

Daniel Lupp. "Effectuation, causation, and machine learning in co-creating entrepreneurial opportunities", Journal of Business Venturing Insights, 2022

Publicación

<1 %

12

Edisney Garcia Perdomo, Alvaro Hernan Alarcon Lopez, Julian Andres Quimbayo Castro, Juan Sebastian Garcia Vargas. "Teaching techniques based on data mining

<1 %

and machine learning at the Mercedes Perdomo de Lievano Children's Shelter", 2022 V Congreso Internacional en Inteligencia Ambiental, Ingeniería de Software y Salud Electrónica y Móvil (AmITIC), 2022

Publicación

13

Maria Luisa de La Fuente. "", IEEE Latin America Transactions, 6/2006

Publicación

<1 %

14

Jonathan Gustavo Castillo Sánchez, Jessica Estefanía Chimbo Solórzano. "Eficiencia en la remoción de materia orgánica mediante lombrifiltros (Eisenia foetida) en aguas residuales domésticas para zonas rurales", Enfoque UTE, 2021

Publicación

<1 %

15

Wolfram Lothar Laaser, Ulises Roman Concha. "MOOCs, A Phenomenon with Many Faces", International Journal of Smart Education and Urban Society, 2018

Publicación

<1 %

16

Norka Bedregal-Alpaca, Víctor Cornejo-Aparicio, Joshua Zárate-Valderrama, Pedro Yanque-Churo. "Classification Models for Determining Types of Academic Risk and Predicting Dropout in University Students", International Journal of Advanced Computer Science and Applications, 2020

Publicación

<1 %

17 "Tendencias en la Investigación Universitaria. Una visión desde Latinoamérica", Alianza de Investigadores Internacionales SAS, 2020 <1 %
Publicación

18 María José Rupérez Moreno. "MULTIDISCIPLINARY TECHNIQUES FOR THE SIMULATION OF THE CONTACT BETWEEN THE FOOT AND THE SHOE UPPER IN GAIT: VIRTUAL REALITY, COMPUTATIONAL BIOMECHANICS, AND ARTIFICIAL NEURAL NETWORKS", Universitat Politecnica de Valencia, 2011 <1 %
Publicación

19 Ricardo Forero Murcia, Lorena Castaño Álvarez, Carolina Mejía Corredor. "El estilo de aprendizaje en educación virtual: breve revisión de la literatura", Virtu@lmente, 2017 <1 %
Publicación

20 Rosario Romero-Martín, Antonio Fraile-Aranda, Víctor-Manuel López-Pastor, Francisco-Javier Castejón-Oliva. "The relationship between formative assessment systems, academic performance and teacher and student workloads in higher education / Relación entre sistemas de evaluación formativa, rendimiento académico y carga de trabajo del profesor y del alumno en la <1 %

docencia universitaria", Infancia y Aprendizaje, 2014

Publicación

21

Saida Margarita Cuadros Oria, Alex Santiago Uriarte Ortiz, Luz Alexandra Javier Silva. "Reducción del zinc mediante sulfato de aluminio y superfloc a-110: nivel laboratorio", Revista del Instituto de Investigación de la Facultad de Ingeniería Geológica, Minera, Metalúrgica y Geográfica, 2020

Publicación

<1 %

22

Rachel Corr. "Strange Seeds: Ethnohistorical Testimonies of the Clandestine Culture of Sacred Plants in Colonial Ecuador", Anthropology of Consciousness, 2022

Publicación

<1 %

23

"Frontier Applications of Nature Inspired Computation", Springer Science and Business Media LLC, 2020

Publicación

<1 %

24

Anubis Alberto Navarro-Rosas, Saul Antonio Obregón-Biosca. "Pedestrian behavior in signalized intersections: Santiago de Querétaro", Journal of Transport & Health, 2022

Publicación

<1 %

25

Christian Fieberg, Matthies Hesse, Thomas Loy, Daniel Metko. "Chapter 6 Machine

<1 %

Learning in Accounting Research", Springer
Science and Business Media LLC, 2022

Publicación

26

Yuri Nieto, Vicente Gacia-Diaz, Carlos Montenegro, Claudio Camilo Gonzalez, Ruben Gonzalez Crespo. "Usage of Machine Learning for Strategic Decision Making at Higher Educational Institutions", IEEE Access, 2019

Publicación

<1 %

27

"Soil Science: Fundamentals to Recent Advances", Springer Science and Business Media LLC, 2021

Publicación

<1 %

28

Armando Vieira, Bernardete Ribeiro. "Introduction to Deep Learning Business Applications for Developers", Springer Science and Business Media LLC, 2018

Publicación

<1 %

29

Carla Terrón Santiago. "Desarrollo e implementación de un banco de ensayos virtual de aerogeneradores para diferentes regímenes de funcionamiento y condiciones de fallo", Universitat Politecnica de Valencia, 2022

Publicación

<1 %

30

LEONARDO ANDRES SIERRA VARELA. "EVALUACIÓN MULTICRITERIO DE LA SOSTENIBILIDAD SOCIAL PARA EL

<1 %

DESARROLLO DE INFRAESTRUCTURAS.",
Universitat Politecnica de Valencia, 2017

Publicación

31

Nidia Danigza Lugo-López, María Del Carmen Pérez-Almagro. "Estrategia pedagógica interdisciplinar para la enseñanza en el primer ciclo de la educación básica en Colombia", Revista Electrónica Educare, 2022

Publicación

<1 %

32

Noa P. Cruz Díaz, Manuel J. Maña López, Jacinto Mata Vázquez, Victoria Pachón Álvarez. "A machine - learning approach to negation and speculation detection in clinical texts", Journal of the American Society for Information Science and Technology, 2012

Publicación

<1 %

33

Patricia Batista Grau. "Desarrollo de nanoestructuras de ZnO mediante anodizado electroquímico en diferentes condiciones para su aplicación en el área energética", Universitat Politecnica de Valencia, 2021

Publicación

<1 %

34

Sofia Aparisi Torrijo. "Los factores del liderazgo influyentes en el crecimiento y éxito del emprendimiento femenino", Universitat Politecnica de Valencia, 2022

Publicación

<1 %

35

Stefania Piedrahita Orozco, Alberto Ochoa-Zezzatti, Gustavo Delgado Lechuga. "Chapter 19 Using Machine Learning to Predict Online Buying Behaviour, Wholesale and Fashion Marketing at Zara, an Analysis Including Z Generation", Springer Science and Business Media LLC, 2021

Publicación

<1 %

36

Wendy Nelly Bada. "Kenebo shipibo-conibo y noción de espacio topológico en niños de 5 años", Praxis & Saber, 2021

Publicación

<1 %

37

por Eduardo Cavieres. "LA EDUCACIÓN CHILENA EN PERSPECTIVA HISTÓRICA, 1860 - 1973: PROBLEMAS, AVANCES Y LIMITACIONES", Paedagogica Historica, 2006

Publicación

<1 %

38

Hugo Esteban Caselli Gismondi, Luis Vladimir Urrelo Huiman. "Multilayer Neural Networks for Predicting Academic Dropout at the National University of Santa - Peru", 2021 International Symposium on Accreditation of Engineering and Computing Education (ICACIT), 2021

Publicación

<1 %

39

Victoria Cañal, María Ornella Beltrame. "Gastrointestinal parasite diversity of South American camelids (Artiodactyla: Camelidae):

<1 %

First review throughout the native range of distribution", International Journal for Parasitology: Parasites and Wildlife, 2022

Publicación

Excluir citas

Apagado

Excluir coincidencias

Apagado

Excluir bibliografía

Apagado