



**ESCUELA UNIVERSITARIA DE POSGRADO**

ANÁLISIS DE LAS IMPLICANCIAS DE LAS CUALIDADES DE LOS DOCENTES EN  
EL RENDIMIENTO ACADÉMICO DE LOS ESTUDIANTES EN LAS  
UNIVERSIDADES PÚBLICAS DE LA REGIÓN CUSCO UTILIZANDO TÉCNICAS  
DE CIENCIA DE DATOS

**Línea de investigación:**

**Ingeniería de software, simulación y desarrollo de TICs**

Tesis para optar el grado académico de Doctora en Ingeniería Sistemas

**Autora**

Acurio Usca, Nila Zonia

**Asesor**

Quispe Prado, Wilber

ORCID: 0000-0003-2452-3669

**Jurado**

Rodriguez Rodriguez, Ciro

Tafur Anzualdo, Irene Vicenta

Sobrado Gómez, Angel

**Lima - Perú**

**2025**



# ANÁLISIS DE LAS IMPLICANCIAS DE LAS CUALIDADES DE LOS DOCENTES EN EL RENDIMIENTO ACADÉMICO DE LOS ESTUDIANTES EN LAS UNIVERSIDADES PÚBLICAS DE LA REGIÓN CUSCO UTILIZANDO TÉCNICAS DE CIENCIA DE DATOS

## INFORME DE ORIGINALIDAD

25%

INDICE DE SIMILITUD

23%

FUENTES DE INTERNET

9%

PUBLICACIONES

9%

TRABAJOS DEL ESTUDIANTE

## FUENTES PRIMARIAS

1	<a href="https://repositorio.unfv.edu.pe">repositorio.unfv.edu.pe</a> Fuente de Internet	2%
2	<a href="https://sedici.unlp.edu.ar">sedici.unlp.edu.ar</a> Fuente de Internet	1%
3	<a href="https://issuu.com">issuu.com</a> Fuente de Internet	1%
4	<a href="https://worldwidescience.org">worldwidescience.org</a> Fuente de Internet	1%
5	<a href="https://researchonline.ljmu.ac.uk">researchonline.ljmu.ac.uk</a> Fuente de Internet	1%
6	<a href="https://hdl.handle.net">hdl.handle.net</a> Fuente de Internet	1%
7	Submitted to Universidad TecMilenio Trabajo del estudiante	1%
8	<a href="https://www.slideshare.net">www.slideshare.net</a> Fuente de Internet	1%



Universidad Nacional  
**Federico Villarreal**

**VRIN** | VICERRECTORADO  
DE INVESTIGACIÓN

ESCUELA UNIVERSITARIA DE POSGRADO

ANÁLISIS DE LAS IMPLICANCIAS DE LAS CUALIDADES DE LOS DOCENTES EN  
EL RENDIMIENTO ACADÉMICO DE LOS ESTUDIANTES EN LAS UNIVERSIDADES  
PÚBLICAS DE LA REGIÓNCUSCO UTILIZANDO TÉCNICAS DE CIENCIA DE  
DATOS

Línea de Investigación:  
Ingeniería de software, simulación y desarrollo de TICs

Tesis para optar el grado académico de:  
Doctora en Ingeniería Sistemas

Autora:  
Acurio Usca, Nila Zonia

Asesor (a):  
Quispe Prado, Wilber  
(ORCID: 0000-0003-2452-3669)

Jurado:  
Rodriguez Rodriguez, Ciro  
Tafur Anzualdo, Irene Vicenta  
Sobrado Gómez, Angel

Lima – Perú

2025

### **Dedicatoria**

A Arturo, quien ha sido mi compañero de sueños y mi roca en cada paso del camino. A mis queridos hijos Nattaly, Melina y Javier, quienes han sido y son la luz de nuestras vidas y la razón de nuestro esfuerzo constante.

A mi familia, cuyo amor incondicional y apoyo inquebrantable han sido mi mayor fortaleza. A mis padres, cuya presencia sigue guiándome desde lo más profundo de mi corazón.

A mis hermanos, familia y amigos, quienes han compartido risas, lágrimas y momentos inolvidables, llenando mi vida de alegría y optimismo.

Esta tesis está dedicada a ustedes, quienes han sido mi inspiración y mi motivación para alcanzar mis metas. Con todo mi amor y gratitud,

Zonia.

## ÍNDICE GENERAL

<b>RESUMEN .....</b>	<b>10</b>
<b>ABSTRACT.....</b>	<b>11</b>
<b>I. INTRODUCCION .....</b>	<b>12</b>
1.1 Planteamiento del problema .....	13
1.2 Descripción del problema.....	16
1.3 Formulación del problema .....	17
1.3.1 Problema general .....	17
1.3.2 Problemas específicos.....	18
1.4 Antecedentes .....	18
1.5 Justificación.....	24
1.6 Limitaciones.....	25
1.7 Objetivos .....	26
1.7.1 Objetivo general.....	26
1.7.2 Objetivos específicos .....	26
1.8 Hipótesis.....	27
1.8.1 Hipótesis general.....	27
1.8.2 Hipótesis específicas.....	27
<b>II. MARCO TEÓRICO .....</b>	<b>28</b>
2.1 Marco Conceptual .....	28
2.1.1 Ciencia de datos .....	28
2.1.1.1 Sub Áreas de la ciencia de datos.....	29

2.1.1.2	Minería de datos.....	32
2.1.1.3	Python para ciencia de datos.....	34
2.1.1.4	SQL para ciencia de datos .....	34
2.1.2	Procesamiento determinístico y estocástico de datos .....	35
2.1.2.1	Procesamiento determinístico de datos.....	35
2.1.2.2	Procesamiento estocástico de datos .....	36
2.1.3	Aprendizaje automático .....	37
2.1.3.1	Clasificación .....	39
2.1.3.2	Herramientas y librerías.....	40
2.1.4	Rendimiento académico.....	44
2.1.5	Perfil del docente .....	46
<b>III.</b>	<b>MÉTODO .....</b>	<b>49</b>
3.1	Tipo de investigación .....	49
3.2	Población y muestra .....	50
3.2.1	Población .....	50
3.2.2	Muestra .....	50
3.3	Operacionalización de variables.....	51
3.4	Instrumentos.....	53
3.5	Procedimientos .....	53
3.6	Análisis de datos .....	54
<b>IV.</b>	<b>RESULTADOS .....</b>	<b>56</b>
4.1	Procesamiento determinístico .....	57
4.1.1	Influencia del grado académico y la experiencia de los docentes.....	57
4.1.1.1	Influencia del grado académico de los docentes.....	58
4.1.1.2	Influencia de la experiencia de los docentes.....	60

4.1.1.3	Influencia del grado académico y la experiencia de los docentes.....	61
4.1.2	Influencia del grupo etario y el género de los docentes.....	64
4.1.2.1	Influencia del grupo etario.....	64
4.1.2.2	Influencia del género de los docentes en el rendimiento académico.....	67
4.1.2.3	Influencia del género y el grupo etario de los docentes.....	68
4.1.3	Influencia de la categoría y el régimen de los docentes.....	71
4.1.3.1	Influencia de la categoría de los docentes.....	71
4.1.3.2	Influencia del régimen de los docentes.....	73
4.1.3.3	Influencia de la categoría y el régimen de los docentes.....	74
4.2	Procesamiento estocástico.....	76
4.2.1	Relevancia de las características distintivas de los docentes.....	79
4.2.2	Relevancia de las preguntas de las encuestas.....	80
4.3	Validación de la hipótesis.....	83
4.3.1	Validación de la hipótesis general.....	83
4.3.2	Validación de las hipótesis específicas.....	84
<b>V.</b>	<b>DISCUSIÓN DE RESULTADOS.....</b>	<b>88</b>
<b>VI.</b>	<b>CONCLUSIONES.....</b>	<b>96</b>
<b>VII.</b>	<b>RECOMENDACIONES.....</b>	<b>97</b>
<b>VIII.</b>	<b>REFERENCIAS.....</b>	<b>99</b>
<b>IX.</b>	<b>ANEXOS.....</b>	<b>105</b>

## ÍNDICE DE FIGURAS

Figura 1. Comparativa en la Universidad Nacional De San Antonio Abad del Cusco.....	15
Figura 2. Diagrama de Venn de Ciencia de Datos.....	28
Figura 3. Script para Pre-procesar y Resumir los Datos por Grado académico.....	58
Figura 4. Script para determinar los ratios de la influencia del Grado académico .....	59
Figura 5. Script para Pre-procesar y Resumir los Datos por Experiencia.....	60
Figura 6. Script para determinar los ratios de la influencia de la experiencia .....	61
Figura 7. Script para Pre-procesar y Resumir los Datos por Grado Académico y Experiencia.....	62
Figura 8. Script para determinar los ratios de la influencia del Grado Académico y la experiencia ....	63
Figura 9. Script para Pre-procesar y Resumir los Datos por Grupo Etario.....	65
Figura 10. Script para determinar los ratios de la influencia del Grupo etario .....	66
Figura 11. Script para Pre-procesar y Resumir los Datos por Género .....	67
Figura 12. Script para determinar los ratios de la influencia del Género.....	68
Figura 13. Script para Pre-procesar y Resumir los Datos por Género y Grupo Etario .....	69
Figura 14. Script para determinar los ratios de la influencia del Género y Grupo Etario.....	70
Figura 15. Script para Pre-procesar y Resumir los Datos por Categoría .....	71
Figura 16. Script para determinar los ratios de la influencia del Régimen .....	72
Figura 17. Script para Pre-procesar y Resumir los Datos por Régimen .....	73
Figura 18. Script para determinar los ratios de la influencia del Régimen .....	74

Figura 19. Script para Pre-procesar y Resumir los Datos por Categoría y Régimen .....	75
Figura 20. Script para determinar los ratios de la influencia de la Categoría y Régimen .....	75
Figura 21. Encuesta de valoración del desempeño de los docentes .....	77
Figura 22. Librerías utilizadas .....	78
Figura 23. Carga de datos y pre procesamiento .....	79
Figura 24. Procesamiento para determinar la relevancia de las características distintivas .....	79
Figura 25. Relevancia de las características distintivas de los docentes .....	80
Figura 26. Procesamiento para determinar la relevancia de las preguntas de la encuesta .....	81
Figura 27. Relevancia de las preguntas de las encuestas de desempeño docente .....	82

**ÍNDICE DE TABLAS**

Tabla 1. Operacionalización de variables .....	51
Tabla 2. Estructura del archivo CSV con datos originales.....	56
Tabla 3. Influencia del Grado Académico en el rendimiento de los estudiantes .....	59
Tabla 4. Influencia de la experiencia en el rendimiento de los estudiantes .....	61
Tabla 5. Influencia del Grado Académico y la Experiencia en el Rendimiento de los Estudiantes .....	63
Tabla 6. Influencia del Grupo etario en el rendimiento de los estudiantes .....	66
Tabla 7. Influencia del Género en el rendimiento de los estudiantes.....	68
Tabla 8. Influencia del Género y Grupo Etario en el Rendimiento de los Estudiantes.....	70
Tabla 9. Influencia de la Categoría en el rendimiento de los estudiantes .....	72
Tabla 10. Influencia del Régimen en el rendimiento de los estudiantes .....	74
Tabla 11. Influencia de la Categoría y Régimen en el Rendimiento de los Estudiantes.....	76
Tabla 12. Relevancia de las características distintivas de los docentes .....	80
Tabla 13. Relevancia de las preguntas de las encuestas de desempeño docente .....	82
Tabla 14. Influencia del Grado Académico y la experiencia de los docentes.....	84
Tabla 15. Influencia del Grupo Etario y Género.....	85
Tabla 16. Influencia de la Categoría y Régimen.....	86
Tabla 17. Análisis de la Influencia de la Característica Grado Académico .....	88
Tabla 18. Análisis de la Influencia de la Característica Experiencia .....	89

Tabla 19. Análisis de la Influencia de las Características Grado Académico y Experiencia.....	89
Tabla 20. Análisis de la Influencia de la Característica Grupo Etario .....	90
Tabla 21. Análisis de la Influencia de la Característica Género .....	91
Tabla 22. Análisis de la Influencia de las Características Grupo Etario y Género .....	91
Tabla 23. Análisis de la Influencia de la Característica Categoría.....	92
Tabla 24. Análisis de la Influencia de la Característica Régimen.....	92
Tabla 25. Análisis de la Influencia de las Características Categoría y Régimen .....	93

## RESUMEN

El objetivo fue determinar qué implicancias existe entre las características distintivas de los docentes como el grado académico, grupo etario, categoría y régimen, género, experiencia, así como la percepción o valoración que tiene los estudiantes de los docentes, en el rendimiento académico de los estudiantes en las universidades públicas de la región Cusco, utilizando técnicas de ciencia de datos. Se realizó dos tipos de procesamiento; procesamiento determinístico, para los casos de relaciones simples entre los atributos, lo cual se efectuó con SQL (Structured Query Language); y el procesamiento estocástico, para los casos de relaciones más complejas entre los atributos, para lo cual se utilizó algoritmos de aprendizaje automático. Se logró determinar las características distintivas de los docentes que tienen mayor influencia en el rendimiento académico; así, la característica de grado académico de Doctorado tiene los mejores promedios (14.44) y mayor porcentaje de aprobados (83.96); la característica etaria de mayores de 70 tiene los mejores promedios (14.47) y el mayor porcentaje de aprobados (84.64); la característica género femenino tiene los mejores promedios (14.62) y el mayor porcentaje de aprobados (86.60). En conclusión, se determinó las implicancias que existe entre las características distintivas de los docentes en el rendimiento académico de los estudiantes en las universidades públicas de la región Cusco, utilizando técnicas de ciencia de datos; lo que permite plantear estrategias de mejora en el proceso de enseñanza-aprendizaje.

*Palabras clave:* Rendimiento académico, ciencia de datos, procesamiento determinístico y estocástico, SQL, modelos aprendizaje automático.

## ABSTRACT

The objective was to determine what implications exist between the distinctive characteristics of teachers such as academic degree, age group, category and regime, gender, experience, as well as the perception or assessment that students have of teachers, on the academic performance of the teachers. Students in public universities in the Cusco region, using data science techniques. Two types of processing were carried out; deterministic processing, for cases of simple relationships between attributes, which was carried out with SQL (Structured Query Language); and stochastic processing, for cases of more complex relationships between attributes, for which machine learning algorithms were used. It was possible to determine the distinctive characteristics of teachers that have the greatest influence on academic performance; Thus, the academic degree characteristic of Doctorate has the best averages (14.44) and the highest percentage of passes (83.96); The age characteristic of those over 70 has the best averages (14.47) and the highest percentage of passes (84.64); The characteristic female gender has the best averages (14.62) and the highest percentage of passes (86.60). In conclusion, the implications that exist between the distinctive characteristics of teachers on the academic performance of students in public universities in the Cusco region were determined, using data science techniques; which allows us to propose improvement strategies in the teaching-learning process.

*Keywords:* Academic performance, data science, deterministic and stochastic processing, SQL, machine learning models.

## I. INTRODUCCION

El trabajo de tesis se enfoca en el análisis de los datos de las bases de datos transaccionales que contienen los resultados de los procesos de enseñanza-aprendizaje de la universidad pública de la Región del Cusco, que incluyen datos de las dimensiones estudiante, escuela profesional y docente, de éste último principalmente sus características distintivas. Se considera el manejo de los grandes volúmenes de datos generados por la transformación digital en los procesos automatizados de la educación superior. Se enfatiza, el aumento en el número de atributos, los cuales presentan en algunos casos relaciones obvias y simples, en otros casos relaciones no obvias y complejas. Por lo tanto, desde la perspectiva del procesamiento y análisis de datos, se identifica dos contextos; primero, la necesidad de utilizar herramientas determinísticas convencionales como la estadística o el lenguaje SQL para casos de relaciones entre atributos simples u obvias; y segundo, la necesidad de técnicas y métodos estocásticos, por ejemplo de ciencia de datos, para casos con relaciones entre atributos más complejas.

La tesis se estructuró conforme lo señala la normatividad vigente de la universidad; considerándose en el primer capítulo el contexto de la problemática que se intenta resolver en el trabajo, la formulación del problema, algunos antecedentes asociados al presente proyecto, la justificación que sustenta el por qué es importante este trabajo, las limitaciones, para finalmente plantear los objetivos y las respectivas hipótesis planteadas.

El siguiente capítulo, considera el marco conceptual requerido para el desarrollo y entendimiento del trabajo. Se consideran fundamentalmente conceptos como Bases de datos transaccionales y técnicas de ciencia de datos.

El siguiente capítulo, considera los aspectos metodológicos del trabajo, enfatizando el tipo de investigación, la población de datos utilizada, la identificación y la definición operacional de las variables, y los modelos utilizados para el proceso de análisis de datos.

En el siguiente capítulo, se realiza el proceso de ciencia de datos, analizando los datos considerando dos enfoques, el determinístico basado en el lenguaje SQL y el estocástico basado en modelos de minería de datos. También se realiza el compendio de los resultados así como la validación de las hipótesis.

En el siguiente capítulo, se efectúa la discusión de resultados. Se enfatiza dos criterios, el análisis a detalle de los hallazgos de la investigación y la comparación de los resultados con los antecedentes referenciados en el trabajo.

Finalmente, se exponen las conclusiones; se plantean algunas recomendaciones sobre posibles trabajos futuros; y se lista las referencias bibliográficas. En la parte final, se expone la matriz de consistencia como parte de los anexos.

## **1.1 Planteamiento del problema**

Actualmente se percibe un entorno social condicionado por las grandes transformaciones motivadas por las Tecnologías de la Información y la Comunicación (TIC), donde conceptos como “La cuarta revolución industrial” propuesta por Schwab (2016) que hace referencia a la aplicación de tecnologías de inteligencia artificial, robótica, internet de las cosas y biotecnología; por otra parte, Harari (2018) en su libro “21 lecciones para el siglo XXI” nos relata que las innovaciones propiciadas por las TIC, ha transformado, viene transformando y transformará el mundo de maneras que antes no eran posibles, y que estas transformaciones implicarán impactos trascendentales en los diferentes paradigmas del futuro de la humanidad. Esta tendencia se viene dando desde varias décadas atrás, es así que, Laney (2001), definió el concepto de “BIG DATA” como la gestión de grandes conjuntos de datos que superan las capacidades de los sistemas de gestión de bases de datos convencionales. Esta generación de ingentes volúmenes de datos se da en menor o mayor grado en las organizaciones actuales. Todo este contexto conlleva un problema, las organizaciones subutilizan las tecnologías y

generalmente desperdician los grandes volúmenes de datos que tienen; en consecuencia, no se diseñan estrategias que permitan obtener ventajas competitivas de estos aspectos.

Orientando el análisis hacia el sector de la educación superior universitaria, el panorama es similar; más aún, si se considera que estas instituciones cuentan con grandes volúmenes de datos, correspondientes a registros oficiales académicos a lo largo de los semestres y años, datos asociados a los procesos de enseñanza-aprendizaje.

Las políticas de gobierno del Perú, en su intento de modernizar y mejorar la calidad de la educación superior, en base a la promulgación de la nueva ley universitaria (2014), marco normativo aprobado mediante Resolución N° 30220, propugna el “Licenciamiento de las universidades” y la “Acreditación de las escuelas profesionales o programas académicos”. Muchas de las escuelas profesionales ya han logrado la respectiva acreditación, encontrándose actualmente en el proceso de “Mejora continua”. Es en este proceso, que se requiere de información y conocimiento que coadyuve al diseño de estrategias que permitan el fortalecimiento de las actividades de Enseñanza-aprendizaje. Es en este contexto, que en la universidad peruana se subutilizan la información de sus bases de datos transaccionales y no se explotan exhaustivamente esos datos; algunas veces, condicionados por la falta de herramientas apropiadas como las técnicas de la ciencia de datos.

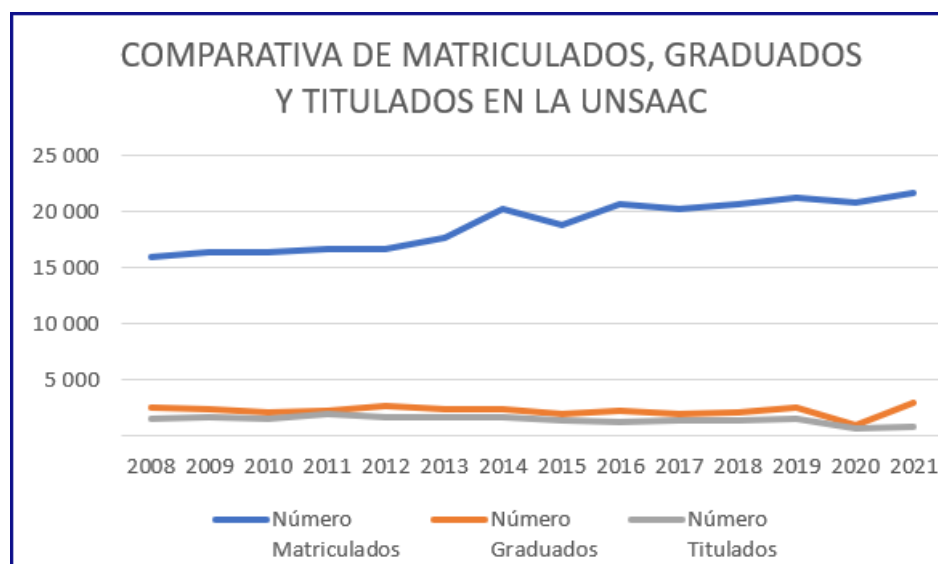
En el abanico de grandes volúmenes de información tanto de corte académico como administrativo que existe a la interna de las universidades, una dimensión poco explorada es, la implicancia de las características distintivas de los docentes universitarios en el rendimiento académico de los estudiantes; esto debido a que los datos referidos a estos aspectos presentan relaciones complejas y generalmente ocultas. Se debe considerar, que en las bases de datos transaccionales se cuenta con registros oficiales de información referida a las características distintivas del docente, así como, de información del rendimiento académico de los estudiantes asociados a las asignaturas impartidas por estos docentes. Dentro de las características

distintivas de los docentes se consideran aspectos que corresponden al perfil profesional del docente, características personales y la valoración del docente por parte de los estudiantes.

La universidad pública de la región del Cusco no es ajena a esta problemática, se tienen muchos aspectos en los que se pueden implementar acciones de mejora continua; por ejemplo, de acuerdo con los datos del Instituto Nacional de Estadística e Informática (INEI) la región Cusco, el número de matriculados tiene una tendencia creciente, mientras que el número de graduados no presenta esa tendencia creciente, peor aún, el número de titulados presenta una tendencia decreciente, tal como se puede apreciar en la figura 1:

### Figura 1

*Comparativa en la Universidad Nacional De San Antonio Abad del Cusco*



*Nota.* Gráfico elaborado con datos obtenidos del INEI de la siguiente dirección:

<https://m.inei.gob.pe/estadisticas/indice-tematico/university-tuition/>.

En este contexto, se hace imperativo utilizar herramientas tecnológicas de vanguardia como las técnicas de ciencia de datos para analizar grandes volúmenes de datos y obtener información y conocimiento relevante que permita implementar procesos de mejora continua. Si bien, el contexto universitario es bastante complejo y tiene varias dimensiones de análisis,

una dimensión poco analizada es la dimensión docente, desde la perspectiva de la implicancia de sus características distintivas en el rendimiento académico de los estudiantes.

## **1.2 Descripción del problema**

En la actual sociedad del conocimiento, se está experimentando una continua reestructuración en los diferentes paradigmas socioculturales como resultado de la transformación digital y el avance vertiginoso de las tecnologías de la información y la comunicación (TIC). Este cambio continuo y la digitalización de los procesos está produciendo una tendencia al crecimiento exponencial de los datos en las organizaciones. Sin embargo, a menudo, esta cantidad de datos generada internamente es ignorada o subutilizada debido a la falta de herramientas tecnológicas, la complejidad y los costos asociados al análisis y extracción de conocimientos valiosos. Esto puede propender a disminuir la calidad en la toma de decisiones y reducir la eficacia de la organización en general.

El contexto de la educación superior no es ajeno a este problema; es decir, se cuenta con grandes volúmenes de información, pero, en el general de los casos solo se explota un segmento de ellos. Específicamente, por ejemplo, para evaluar y extraer conocimiento que coadyuve a plantear estrategias de mejora continua en el rendimiento académico de los estudiantes, generalmente solo se analizan las dimensiones estudiante e infraestructura; la dimensión docente no se analiza o solo se analiza mínimamente desde la perspectiva del perfil académico del docente, dejándose de lado otras características distintivas, como segmento etario al que pertenece, categoría y régimen que ostenta, género al que pertenece, la percepción o valoración que tienen los estudiantes del docente, etc.

Así mismo, la naturaleza de los datos corresponde en algunos casos a datos objetivos, por ejemplo las notas obtenidas por los estudiantes en los diferentes procesos de evaluación; sin embargo, también existen datos subjetivos, por ejemplo los obtenidos en base a encuestas

aplicadas a los estudiantes, donde ellos realizan una evaluación subjetiva del desempeño de sus docentes en el dictado de sus respectivas asignaturas.

En este contexto, generalmente solo se procesan los datos objetivos y con relaciones simples o claras entre ellos, normalmente utilizando análisis estadístico, siendo éstos claramente procesos determinísticos. Respecto a los datos objetivos y subjetivos, con relaciones complejas entre ellos, se efectúan procesos mínimos o no se efectúa ningún proceso de análisis, esto principalmente porque los procesos determinísticos no son viables en estos casos o simplemente ofrecen resultados muy pobres. Para estos casos se requiere de procesos estocásticos asociados a técnicas o herramientas que permitan, por ejemplo descubrir patrones de comportamiento, asociaciones entre variables, etc.

En suma, considerando específicamente las universidades públicas de la región Cusco, se ignora o subutiliza segmentos de los grandes volúmenes de datos existentes; sobre todo aquellos que presentan relaciones complejas entre ellos; en consecuencia, los procesos de mejora continua en las que están inmersas las escuelas profesionales cuentan con poca información valiosa que permita diseñar mejores estrategias orientadas a mejorar la calidad de la enseñanza.

### **1.3 Formulación del problema**

#### ***1.3.1 Problema general***

¿Qué implicancia existe entre las características distintivas de los docentes (grado académico, grupo etario, categoría y régimen, género, experiencia, así como la percepción o valoración que tienen los estudiantes de los docentes) en el rendimiento académico de los estudiantes en las universidades públicas de la región Cusco, utilizando técnicas de ciencia de datos?

### **1.3.2 Problemas específicos**

- ¿En qué medida influyen el grado académico y la experiencia de los docentes en el rendimiento académico de los estudiantes en las universidades públicas de la región del Cusco?
- ¿En qué medida influyen el grupo etario y el género de los docentes en el rendimiento académico de los estudiantes en las universidades públicas de la región del Cusco?
- ¿En qué medida influyen la categoría y régimen al que pertenecen los docentes en el rendimiento académico de los estudiantes en las universidades públicas de la región del Cusco?
- ¿En qué medida tiene relevancia las características distintivas de los docentes en la valoración o percepción que tienen los estudiantes de los docentes, así como la pertinencia de las encuestas aplicadas a los estudiantes para evaluar el desempeño de los docentes, en las universidades públicas de la región del Cusco?

### **1.4 Antecedentes**

Existen diferentes trabajos relacionados al análisis y/o evaluación del rendimiento académico en la educación superior, también, existen trabajos sobre perfil docente universitario. Se seleccionaron los más relevantes para el presente trabajo, los cuales se citan a continuación.

Molero-Castillo et al. (2018) en el artículo "Interactive System for the Analysis of Academic Achievement at the Upper-Middle Education in Mexico", presenta un análisis del rendimiento académico de los estudiantes de educación media superior en México, utilizando un sistema interactivo que integra datos educativos, indicadores socioeconómicos y georreferenciados. El objetivo del estudio es comprender mejor los factores que influyen en el rendimiento académico y proponer políticas públicas que aborden estos desafíos. Molero-Castillo destaca la importancia de la calidad y equidad en la educación, especialmente en el

contexto de México, donde existen desigualdades regionales y económicas que afectan a la educación.

El sistema interactivo propuesto por Molero-Castillo se basa en un enfoque multidisciplinario y utiliza técnicas de análisis espacial, visualización de datos y estadísticas descriptivas e inferenciales. Este enfoque permite identificar y analizar patrones y tendencias en el rendimiento académico, así como relaciones entre variables socioeconómicas y educativas. El sistema se alimenta de diversas fuentes de información, incluyendo datos de evaluaciones estandarizadas, censos escolares y datos socioeconómicos, para crear un panorama completo de las condiciones educativas en México. El autor también aborda las limitaciones del sistema, reconociendo que puede haber factores no medidos o difíciles de cuantificar que influyen en el rendimiento académico.

A partir del análisis realizado mediante el sistema interactivo, Molero-Castillo identifica diversos factores que afectan el rendimiento académico en México, tales como el nivel socioeconómico, la infraestructura escolar, la formación docente y las condiciones del entorno en el que viven los estudiantes. El autor propone diversas políticas públicas y acciones que podrían mejorar la calidad y equidad de la educación en México, como la inversión en infraestructura y recursos educativos, la formación continua de docentes, el apoyo a estudiantes en situación de vulnerabilidad y la promoción de la participación comunitaria en la gestión escolar. El estudio de Molero-Castillo es un importante aporte al debate sobre cómo mejorar el rendimiento académico y promover una educación más equitativa en México y otros países en desarrollo.

Martínez et al. (2016) en el artículo "Data-mining Techniques in Detecting Factors Linked to Academic Achievement", los autores exploran cómo las técnicas de minería de datos pueden ser utilizadas para identificar factores asociados al rendimiento académico de los estudiantes. El objetivo principal del estudio es aprovechar el potencial de la minería de datos

para analizar grandes conjuntos de datos educativos y descubrir patrones y relaciones que puedan informar a educadores y responsables políticos en la toma de decisiones.

Martínez Abad destaca la importancia de utilizar enfoques metodológicos innovadores y sólidos para mejorar la comprensión de los factores que contribuyen al éxito académico y, en última instancia, informar la implementación de intervenciones educativas efectivas.

El estudio de Martínez Abad emplea varias técnicas de minería de datos, como el análisis de clusters, la regresión logística y los árboles de decisión, para identificar patrones y relaciones entre variables relevantes para el rendimiento académico. Estas técnicas permiten analizar grandes volúmenes de datos y detectar asociaciones complejas y no lineales entre variables, lo cual es esencial para comprender la naturaleza multifacética del rendimiento académico.

Al aplicar estas técnicas, el autor examina diversos factores asociados al rendimiento académico, incluyendo características demográficas, socioeconómicas, escolares y familiares, así como el acceso a recursos educativos y el apoyo recibido por los estudiantes.

A partir de los resultados obtenidos en su análisis, Martínez Abad identifica varios factores clave que están relacionados con el rendimiento académico de los estudiantes, como el nivel socioeconómico, el apoyo familiar, la calidad de la enseñanza y la motivación del estudiante. El autor sugiere que las intervenciones educativas y las políticas públicas deben abordar estos factores de manera integral para mejorar el rendimiento académico y reducir las desigualdades educativas. En conclusión, el estudio de Martínez Abad demuestra el potencial de las técnicas de minería de datos para enriquecer la investigación en el ámbito educativo y proporcionar información valiosa sobre los factores que influyen en el rendimiento académico de los estudiantes, lo que puede ser de gran utilidad para diseñar estrategias y políticas educativas más efectivas.

Chávez et al. (2020) en el artículo "A Data Analytics Approach to Contrast the Performance of Teaching (only) vs. Research Professors", los autores presentan un estudio que utiliza técnicas de análisis de datos para comparar el rendimiento de profesores dedicados exclusivamente a la enseñanza y aquellos que también se dedican a la investigación. El objetivo del estudio es entender si la participación en actividades de investigación afecta la calidad de la enseñanza y el aprendizaje de los estudiantes. Chávez aborda la importancia de evaluar el impacto de diferentes perfiles de profesorado en el rendimiento académico de los estudiantes, lo que puede tener implicaciones significativas en la formulación de políticas educativas y la asignación de recursos en las instituciones de educación superior.

El enfoque de análisis de datos empleado por Chávez incluye la recopilación y el análisis de datos cuantitativos y cualitativos relacionados con el desempeño académico de los estudiantes y la calidad de la enseñanza de profesores de diferentes perfiles. Entre las variables consideradas se encuentran las calificaciones de los estudiantes, las evaluaciones de los cursos y los profesores, y las publicaciones y proyectos de investigación de los docentes.

El autor utiliza técnicas estadísticas y de minería de datos para identificar patrones y tendencias en el rendimiento académico de los estudiantes, así como para evaluar el impacto de la participación en actividades de investigación en la enseñanza y el aprendizaje.

Los resultados del análisis de Chávez sugieren que, en general, no hay diferencias significativas en el rendimiento académico de los estudiantes y la calidad de la enseñanza entre profesores dedicados exclusivamente a la enseñanza y aquellos que también se dedican a la investigación. Sin embargo, el autor también identifica ciertas áreas y disciplinas en las que la participación en actividades de investigación parece tener un impacto positivo en la enseñanza y el aprendizaje. Chávez concluye que las instituciones de educación superior deben considerar la diversidad de perfiles de profesorado y su impacto en el rendimiento académico de los estudiantes, al diseñar políticas y asignar recursos. Además, el estudio destaca la importancia

de fomentar la colaboración entre profesores de diferentes perfiles y la integración de la enseñanza y la investigación para mejorar la calidad de la educación superior.

Nafie et al. (2021) en el artículo "Using Data Mining Techniques in Building a Model to Determine the Factors Affecting Academic Data for Undergraduate Students", se presenta un estudio en el que se emplean técnicas de minería de datos para construir un modelo que identifica los factores que afectan el rendimiento académico de los estudiantes universitarios. Se enfoca en la importancia de comprender estos factores para diseñar intervenciones y políticas educativas que promuevan el éxito académico y reduzcan las desigualdades entre los estudiantes. El artículo hace hincapié en que la minería de datos es una herramienta poderosa para analizar grandes conjuntos de datos y descubrir patrones y relaciones ocultas en el ámbito educativo.

Se utilizan diversas técnicas de minería de datos, como el análisis de regresión, árboles de decisión y redes neuronales, para construir un modelo predictivo del rendimiento académico de los estudiantes universitarios. Se analizan variables demográficas, socioeconómicas, familiares y académicas, así como variables relacionadas con el entorno de aprendizaje y los recursos educativos disponibles para los estudiantes. El modelo se basa en un conjunto de datos que incluye información sobre los estudiantes, sus familias y las instituciones educativas, lo que permite un análisis integral de los factores que influyen en el rendimiento académico.

Los resultados del estudio revelan varios factores clave que afectan el rendimiento académico de los estudiantes universitarios, incluyendo el nivel socioeconómico, el apoyo familiar, la motivación y el compromiso del estudiante, y la calidad de la enseñanza. Se sugiere que las instituciones de educación superior y los responsables políticos deben considerar estos factores al diseñar intervenciones y políticas educativas, con el fin de mejorar el rendimiento académico y reducir las desigualdades entre los estudiantes. En resumen, el estudio demuestra el potencial de las técnicas de minería de datos para identificar factores clave que afectan el

rendimiento académico en el ámbito universitario, proporcionando información valiosa para la toma de decisiones en el diseño de políticas y prácticas educativas.

Batool et al. (2022) en el artículo "Educational Data Mining to Predict Students' Academic Performance: A Survey Study", se presenta una revisión exhaustiva de los estudios que utilizan técnicas de minería de datos educativos para predecir el rendimiento académico de los estudiantes. La minería de datos educativos es un enfoque interdisciplinario que combina métodos de aprendizaje automático, estadísticas y análisis de datos para explorar y analizar grandes conjuntos de datos educativos con el objetivo de descubrir patrones y relaciones útiles. La revisión se centra en cómo las técnicas de minería de datos se han aplicado para predecir el rendimiento académico de los estudiantes y cómo estos enfoques pueden ser utilizados para mejorar las políticas y prácticas educativas.

El estudio revisa una amplia gama de técnicas de minería de datos utilizadas en la predicción del rendimiento académico, incluyendo árboles de decisión, redes neuronales, algoritmos de agrupamiento y técnicas de regresión. Se analizan las ventajas y desventajas de cada enfoque, así como los desafíos y oportunidades asociados con su implementación en entornos educativos. Además, se examinan las variables y factores que se han investigado en estudios anteriores, como características demográficas, socioeconómicas, familiares, académicas y psicológicas, y se discute su relevancia y utilidad en la predicción del rendimiento académico de los estudiantes.

En conclusión, la revisión destaca la importancia y el potencial de las técnicas de minería de datos educativos para predecir el rendimiento académico de los estudiantes y proporcionar información valiosa para la toma de decisiones en el diseño de políticas y prácticas educativas. A partir de los estudios analizados, se sugiere que la minería de datos educativos puede desempeñar un papel crucial en la identificación temprana de estudiantes en riesgo de fracaso académico, la evaluación de la efectividad de intervenciones educativas y la

promoción de la equidad y la calidad en la educación. Sin embargo, también se enfatiza la necesidad de abordar los desafíos éticos y de privacidad asociados con el uso de datos personales y sensibles en la minería de datos educativos, así como la importancia de desarrollar enfoques metodológicos sólidos y validados en futuras investigaciones

### **1.5 Justificación**

Este trabajo de investigación es importante por su relevancia académica, científica, técnica y social, y tiene como objetivo generar conocimiento que permita mejorar la calidad de la educación superior en la región Cusco y contribuir al desarrollo del país.

Desde la perspectiva de la justificación académica, el trabajo de investigación permitirá obtener conocimiento valioso de la relación que existe entre las características distintivas de los docentes y el rendimiento académico de los estudiantes en el contexto de las universidades públicas de la región del Cusco; determinando las debilidades y fortalezas del docente, que permitirá diseñar estrategias de mejora continua también en la dimensión docente, en el proceso de enseñanza-aprendizaje.

Desde la perspectiva de la justificación científica, el presente trabajo de investigación se realizará completamente con el rigor científico requerido; dada su naturaleza cuantitativa, se utilizará técnicas determinísticas y estocásticas que incluirán técnicas de ciencia de datos, para analizar los datos y proporcionar información que permita generar nuevo conocimiento.

Se propenderá a obtener información clara, precisa, válida y confiable; que contribuyan a una gestión de mejora continua, sustentada en conocimiento.

Desde la perspectiva tecnológica, el avasallante avance de las Tecnologías de la Información y la Comunicación, cada vez ofrece nuevas e innovadoras herramientas tecnológicas; éstas se deben difundir y promover su uso en las organizaciones, de modo que, contribuyan a obtener ventajas competitivas. En este sentido, este trabajo enfatizará la necesidad de considerar en los sistemas procesos determinísticos y estocásticos que incluyan

técnicas y herramientas de ciencia de datos, lo que permitirá demostrar la aplicación de tecnología de vanguardia en el descubrimiento de nuevo conocimiento en el contexto universitario de la región Cusco.

Desde la perspectiva de la justificación social, uno de los fines de la universidad es la formación de profesionales idóneos para entender y resolver los problemas de la sociedad en general; por tanto, se debe considerar la formación de profesionales con pensamiento crítico, capacidad de abstracción, creativos y sobre todo identificados con el desarrollo de su región y del país. El presente trabajo de investigación tiene el propósito de plantear el uso de herramientas tecnológicas como las técnicas de ciencia de datos, que permitan obtener información que contribuyan al proceso de mejora continua en las universidades públicas de la región Cusco.

## **1.6 Limitaciones**

El análisis del rendimiento académico de los estudiantes en las instituciones de educación superior es muy complejo, requiere que se analicen diferentes variables como: aspectos personales, familiares y socioeconómicos de los estudiantes; currículo de estudios pertinentes y actualizados a estándares internacionales; infraestructura proporcionada por la universidad, tal las aulas, talleres, laboratorios, bibliotecas, etc.; y no solamente la implicancia de las características distintivas de los docentes.

Sin embargo, este trabajo se limita al análisis de la dimensión docente y las notas dadas por ellos en las diferentes asignaturas regentadas, en la perspectiva que la información y conocimiento obtenido mediante procesos determinísticos y estocásticos que incluyen técnicas de ciencia de datos, se utilice como recurso para diseñar diferentes estrategias que soporte el proceso de mejora continua en el que están inmersas las escuelas profesionales de las universidades públicas.

También los datos utilizados se circunscriben al contexto de las universidades públicas de la región del Cusco, limitado a los últimos 3 años.

Un aspecto relevante es que cada universidad es una realidad distinta; no obstante, las técnicas y modelos de ciencia de datos planteados en el presente trabajo, se pueden aplicar a cualquier realidad universitaria y a cualquier rango de años, reformulando simplemente los conjuntos de datos utilizados.

## **1.7 Objetivos**

### **1.7.1 *Objetivo General***

Determinar qué implicancias existe entre las características distintivas de los docentes (grado académico, grupo etario, categoría y régimen, género, experiencia, así como la percepción o valoración que tiene los estudiantes de los docentes) en el rendimiento académico de los estudiantes en las universidades públicas de la región Cusco, utilizando técnicas de ciencia de datos.

### **1.7.2 *Objetivos Específicos***

- Determinar en qué medida influyen el grado académico y la experiencia de los docentes en el rendimiento académico de los estudiantes en las universidades públicas de la región del Cusco.
- Determinar en qué medida influyen el grupo etario y el género de los docentes en el rendimiento académico de los estudiantes en las universidades públicas de la región del Cusco.
- Determinar en qué medida influyen la categoría y régimen al que pertenecen los docentes en el rendimiento académico de los estudiantes en las universidades públicas de la región del Cusco.
- Determinar en qué medida tiene relevancia las características distintivas de los docentes en la valoración o percepción que tienen los estudiantes de los docentes, así como la

pertinencia de las encuestas aplicadas a los estudiantes para evaluar el desempeño de los docentes, en las universidades públicas de la región del Cusco.

## **1.8 Hipótesis**

### **1.8.1 Hipótesis General**

Si se utilizan técnicas de ciencia de datos, se determinan las implicancias que existen entre las características distintivas de los docentes (grado académico, grupo etario, categoría y régimen, género, experiencia, así como la percepción o valoración que tiene los estudiantes de los docentes) en el rendimiento académico de los estudiantes en las universidades públicas de la región Cusco.

### **1.8.2 Hipótesis Específicas**

- Si se analizan los datos referidos al grado académico y la experiencia de los docentes se determina la medida en que influyen en el rendimiento académico de los estudiantes en las universidades públicas de la región del Cusco.
- Si se analizan los datos referidos al grupo etario y el género de los docentes se determina la medida en que influyen en el rendimiento académico de los estudiantes en las universidades públicas de la región del Cusco.
- Si se analizan los datos referidos a la categoría y régimen al que pertenecen los docentes se determina la medida en que influyen en el rendimiento académico de los estudiantes en las universidades públicas de la región del Cusco.
- Si se analizan las características distintivas de los docentes en la valoración o percepción que tienen los estudiantes de los docentes, así como la pertinencia de las encuestas aplicadas a los estudiantes para evaluar el desempeño de los docentes, en las universidades públicas de la región del Cusco, se determinan en qué medida tiene relevancia estas características y la pertinencia de las encuestas.

## II. MARCO TEÓRICO

### 2.1 Marco conceptual

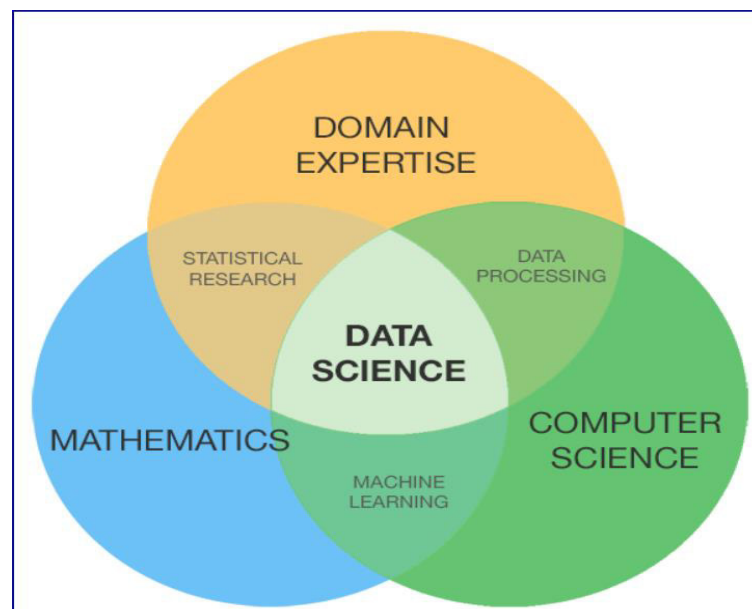
#### 2.1.1 Ciencia de datos

La ciencia de datos es una disciplina que utiliza técnicas de estadística, matemáticas e informática para analizar grandes cantidades de datos y extraer conocimientos útiles de ellos. Una definición de Ciencia de Datos se tiene Kelleher (2018), donde se entiende que es un área interdisciplinaria que considera conceptos estadísticos, el dominio del negocio y algorítmica; que permite extraer patrones no obvios de grandes volúmenes de datos.

Hoy en día, la ciencia de datos es cada vez más importante debido a la enorme cantidad de datos que se generan en diversos campos, desde la industria hasta la investigación científica. Palmer (2015) ilustra con un diagrama de Venn las ramas del conocimiento que involucra la ciencia de datos.

#### Figura 2

*Diagrama de Venn de Ciencia de Datos*



La ciencia de datos tal como lo describe Kelleher (2018), permite detectar patrones y relaciones en los datos que, de otra manera, serían difíciles o imposibles de identificar. Estos

patrones y relaciones pueden utilizarse para predecir comportamientos futuros, optimizar procesos y tomar decisiones informadas. Por ejemplo, en la industria, la ciencia de datos se utiliza para mejorar la eficiencia de los procesos de producción, reducir los costos y mejorar la calidad de los productos. En la investigación científica, la ciencia de datos se utiliza para analizar grandes conjuntos de datos, como los datos de secuenciación genómica, y para descubrir patrones que pueden ser relevantes para la comprensión de enfermedades y para el desarrollo de tratamientos.

Además, la ciencia de datos ha dado lugar a nuevas tecnologías y servicios, como los motores de búsqueda, los sistemas de recomendación y las aplicaciones de inteligencia artificial. Estas innovaciones están transformando la forma en que interactuamos con el mundo y están abriendo nuevas oportunidades para la innovación y el desarrollo económico.

En suma, la ciencia de datos es una disciplina crucial en la actualidad debido a la gran cantidad de datos que se generan y la necesidad de analizarlos para extraer conocimiento útil. La ciencia de datos es esencial para la toma de decisiones informadas, la optimización de procesos y el desarrollo de nuevas tecnologías y servicios que impulsan el progreso en múltiples áreas.

**2.1.1.1 Sub Áreas de la ciencia de datos.** Desde la perspectiva de que la ciencia de datos no se considera como parte de la ciencia formal en el sentido estricto de la definición de Mario Bunge (1974); sino más bien, una disciplina interdisciplinaria que conjuga conocimientos y técnicas de varias áreas, como son las matemáticas, ciencias de la computación y el dominio de negocio, tal como lo manifiesta VanderPlas (2016), tiene un alto componente de ciencia formal; por lo que, existen diferentes propuestas de taxonomía de sub áreas de ciencia de datos. Aquí se presenta una taxonomía ampliamente utilizada en la literatura de ciencia de datos:

- **Análisis de datos:** Prioritariamente se enfoca a determinar tendencias en los datos con el objetivo de identificar patrones, también a efectuar predicciones, que permitan tomar decisiones informadas. En esta sub área se hace uso de conceptos estadísticos y de aprendizaje automático, con el propósito de datos e implementar modelos predictivos. El análisis de datos se puede aplicar a muchos campos: La medicina, las finanzas, la ciencia, el marketing, la industria, Etc. Esta sub área se enfoca en el estudio de los datos para extraer información y conocimientos útiles.
- **Ingeniería de datos:** El propósito de esta sub área es garantizar la calidad y la disponibilidad de los datos para su respectivo análisis. Además del diseño, es responsable de los procesos de extracción o recolección, transformación y almacenamiento de los datos, incluyendo también la limpieza y preparación para su posterior utilización. Esta sub área es crucial en un proyecto de ciencia de datos, debido a que obtener resultados fiables depende de la calidad y la completitud de los datos.
- **Visualización de datos:** El propósito de esta sub área es hacer que la información sea más accesible y comprensible, enfocándose prioritariamente en la representación visual de los datos. Utiliza herramientas gráficas para representar los datos de manera clara y efectiva. Esta sub área es importante porque puede visualmente apoyar a identificar patrones y tendencias en los datos, y a comunicar la información de manera más efectiva.
- **Aprendizaje automático:** Tal lo manifestado por Goodfellow (2016), el propósito del aprendizaje automático es desarrollar modelos que permitan efectuar predicciones a partir de los datos e identificar patrones y tendencias no obvias; en base a algoritmos que hacen, que el computador aprenda autónomamente. Se utilizan técnicas como redes neuronales, árboles de decisión y algoritmos de clustering para desarrollar modelos predictivos y de clasificación.

- Minería de datos: El propósito de esta sub área se centra en descubrir patrones, relaciones no obvias y reglas de asociación en grandes volúmenes de datos. Se utiliza una variedad de técnicas, tal como análisis de correlación, análisis de agrupamiento y análisis de componentes principales. La aplicación de la minería de datos abarca varios campos, dentro de los que se tiene la medicina, las finanzas, el marketing, la ingeniería y también en investigación científica.
- Big Data: El propósito de esta sub área se enfoca en el manejo y análisis de grandes volúmenes de datos estructurados y no estructurados, que deben ser procesados en tiempos relativamente aceptables. Generalmente las herramientas y técnicas convencionales no son suficientes para trabajar con Big Data; por lo que se utilizan técnicas específicas, como procesamiento distribuido, Apache Hadoop y Apache Spark. La tendencia actual es que diversas organizaciones cuenten cada vez con mayores cantidades de datos.
- Ciencia de datos social: El propósito de esta sub área se focaliza en el uso de técnicas que permitan estudiar, analizar y entender los fenómenos sociales y culturales. Utiliza técnicas como procesamiento de lenguaje natural aplicado a la minería de texto, análisis de los datos de las redes sociales, así como el análisis de sentimientos, con el objetivo de entender las manifestaciones sociales y culturales de la sociedad. Se aplica en muchos campos, prioritariamente a las ciencias sociales, la psicología y la ciencia política.
- Ética de la ciencia de datos: El propósito de esta sub área se orienta a garantizar el manejo y uso responsable de los datos, sobre todo de los datos sensibles, respetando los derechos de privacidad de las personas. En este contexto, tiene que ver con los aspectos éticos y legales relacionados con el compendio, almacenamiento y divulgación de los datos sobre todo personales. La ética en la ciencia de datos, también tiene por finalidad, prevenir los posibles sesgos en los algoritmos de aprendizaje automático, el fomento de políticas y

regulaciones para la protección de datos, la promoción de la transparencia en el uso y divulgación de los datos.

**2.1.1.2 Minería de datos.** La conceptualización de minería de datos se atribuye a Piatetsky-Shapiro (1989), se dio en un artículo publicado en la revista AI Magazine. La minería de datos es una sub área de la ciencia de datos que se enfoca en descubrir patrones y relaciones en grandes conjuntos de datos. Para lograr esto, se utilizan diversas técnicas y algoritmos de aprendizaje automático, como el análisis de correlación, análisis de clustering y análisis de componentes principales. El objetivo de la minería de datos es identificar información valiosa y conocimientos útiles en los datos, lo que puede llevar a la toma de decisiones informadas y mejorar la eficiencia en diversos campos.

También se tiene definiciones, como la propuesta por Bramer (2016), que define como el proceso de descubrir de manera automáticamente información de mucha utilidad. Utiliza conceptos estadísticos, técnicas computacionales, aprendizaje automático y gestión de bases de datos con la finalidad de descubrir patrones y relaciones no obvias en los datos.

La aplicación de la minería abarca áreas como la medicina, las finanzas, la ingeniería, el marketing y la ciencia. En la medicina, por ejemplo, la minería de datos se utiliza para procesar y analizar grandes volúmenes de datos de historias clínicas que permitan descubrir patrones que pueden ayudar a predecir enfermedades que coadyuven a tratamientos más efectivos. En el marketing, se puede personalizar las estrategias analizando los datos del comportamiento del consumidor.

Uno de los mayores desafíos de la minería de datos es la enorme cantidad de datos que deben ser procesados. Para superar este desafío, se utilizan técnicas de procesamiento distribuido, como el uso de clusters y bases de datos distribuidas. Además, la minería de datos requiere un conjunto de habilidades y conocimientos especializados, como la comprensión de

los algoritmos de aprendizaje automático y la capacidad de interpretar los resultados del análisis de datos.

En síntesis, la minería de datos es una técnica valiosa en la ciencia de datos para identificar patrones y relaciones en grandes conjuntos de datos. La minería de datos se aplica en muchos campos y requiere habilidades especializadas y técnicas de procesamiento distribuido para superar el desafío de la gran cantidad de datos.

Herramientas Fundamentales. La ciencia de datos utiliza diferentes tipos de herramientas, sin embargo éstas se pueden resumir en tres grandes grupos: Estadística, lenguajes de programación y procesamiento de datos. Desde la perspectiva del presente trabajo, se hace énfasis en estadística, Python como lenguaje de programación y SQL para el procesamiento de datos. Estos conceptos y herramientas son fundamentales en ciencia de datos, porque posibilitarán los procesos de: Recopilación de datos; pre procesamiento de datos, considerando aspectos de extracción, transformación y limpieza de datos; visualización, análisis e interpretación de datos. El científico de datos podrá o tener información que coadyuvará a tomar decisiones informadas.

Estadística para Ciencia de Datos. La estadística es fundamental en los procesos de ciencia de datos, porque haciendo uso de conceptos estadísticos se posibilita tareas como pre-procesamiento, análisis, interpretación y presentación de datos.

- La estadística es crucial en las etapas de pre-procesamiento y análisis de datos. Dentro de los conceptos más relevantes se consideran:
  - Medidas de tendencia central, tal la media, la mediana y la moda.
  - Medidas de dispersión tal la desviación estándar y el rango intercuartílico.
  - Análisis de regresión. Pruebas de hipótesis.
  - Distribuciones de probabilidad.
  - Intervalos de confianza.

La importancia de la estadística en la ciencia de datos, radica en que se constituye la base matemática que posibilita los algoritmos de identificación de patrones, los procesos de inferencia sobre grandes conjuntos de datos, construir y validar modelos.

**2.1.1.3 Python para ciencia de datos.** Python es un lenguaje de programación de alto nivel, multi paradigma (imperativo, funcional y orientado a objetos).

El propósito en esta sección, no es describir las características más relevantes del lenguajes, sino, hacer énfasis en que es el lenguaje ampliamente utilizado en inteligencia artificial y ciencia de datos, debido a los más de 10 millones de archivos librería con las que cuenta, tal como se puede apreciar en el repositorio oficial del lenguaje, el Índice de paquetes de Python (PyPI) (2024); además de la versatilidad, legibilidad, sintaxis y semántica clara.

Existen muchas herramientas y librerías orientadas a ciencia de datos, dentro de las librerías más relevantes se puede mencionar:

- NumPy como librería para procesos matemáticos y de tipo numérico.
- Pandas para la gestión de datos similar a los *gestores de datos convencionales*.
- PySpark utilizar comandos de SQL utilizando Python para manipular y analizar grandes volúmenes de datos, incluso en entornos distribuidos.
- Scikit-learn para implementar modelos de aprendizaje automático.
- Matplotlib como graficadores o visualizadores de datos.
- Seaborn también como graficador o visualizador de datos.

Una característica adicional de Python es que su curva de aprendizaje no es muy pronunciada, por el contrario es constante y fácil de aprender; además de lo ya mencionado que cuenta con una inmensidad de librerías que dan soporte a la diversidad de áreas y sub áreas de la inteligencia artificial.

**2.1.1.4 SQL para ciencia de datos.** Los gestores de bases de datos del paradigma relacional, utilizan como lenguaje de programación para gestionar y sobre todo obtener

información de las bases de datos transaccionales, el Lenguaje Estructurado de Consulta (Structured Query Language - SQL).

Las organizaciones cuentan con grandes volúmenes de datos, gran parte de ellos en bases de datos transaccionales del tipo relacional; es ahí, que se hace necesario como una de las fases de la ciencia de datos, extraer y pre procesar los datos haciendo uso de SQL.

Un aspecto relevante de SQL es que la gran mayoría de lenguajes de programación de alto nivel, proveen librerías o interfaces para interactuar con las bases de datos relacionales.

### **2.1.2 *Procesamiento determinístico y estocástico de datos***

Dada la tendencia a que las instituciones cada vez cuentan con mayores volúmenes de datos, tanto de naturaleza objetiva y subjetiva; así como la incesante innovación en el campo de las tecnologías de la información y las comunicaciones en especial la inteligencia artificial, hacen imperativo que el procesamiento de datos sea enfocado desde dos perspectivas, procesamiento determinístico y procesamiento estocástico.

**2.1.2.1 Procesamiento determinístico de datos.** Se caracterizan por producir los mismos resultados para las mismas entradas o condiciones iniciales de los datos; por tanto, éstos pueden ser representados de manera explícita con modelos matemáticos e implementarse con algoritmos sencillos. Desde la perspectiva del análisis de datos, el procesamiento determinístico implica utilizar métodos y técnicas sobre atributos con relaciones claras y simples. Un ejemplo de procesamiento determinístico es el procesamiento realizado sobre bases de datos con el lenguaje SQL.

Dentro de las ventajas más relevantes del procesamiento determinístico, se tiene: Una comprensión evidente y clara; facilidad en la implementación; eficiencia en el manejo de diferentes volúmenes de datos, incluso grandes volúmenes; posibilidad de efectuar procesos complejos de consulta de manera clara y sencilla; y principalmente precisión en los resultados obtenidos.

Dentro de las desventajas se puede mencionar las siguientes: No maneja de manera apropiada los datos generados en contextos donde se tienen incertidumbre y aleatoriedad; presenta limitaciones para modelar relaciones complejas y no lineales entre atributos; generalmente no es efectivo en la identificación de patrones de comportamiento no evidentes en los datos; en algunos casos requiere de un amplio dominio de las herramientas utilizadas, como SQL.

**2.1.2.2 Procesamiento estocástico de datos.** Se caracterizan por producir resultados que varían incluso para las mismas entradas o condiciones iniciales de los datos; por tanto, éstos no pueden ser representados de manera sencilla y explícita con modelos matemáticos. Desde la perspectiva del análisis de datos, el procesamiento estocástico utiliza técnicas y métodos que permiten manejar relaciones complejas entre los atributos, permitiendo modelar la incertidumbre y la variabilidad inherentes a los datos, lo que facilita la identificación de patrones no evidentes y relaciones no lineales entre los atributos. Un ejemplo de procesamiento estocástico es el realizado en minería de datos, para identificar grupos de datos con patrones de comportamiento similares o reglas de asociación entre atributos.

Dentro de las ventajas se puede mencionar las siguientes: Maneja de manera apropiada los datos generados en contextos donde se tienen incertidumbre y aleatoriedad; es adecuado para modelar relaciones complejas y no lineales entre atributos; es apropiado para la identificación de patrones de comportamiento y reglas de asociación no evidentes en los datos.

Dentro de las desventajas más relevantes del procesamiento estocástico, se tiene: Mayor complejidad en la comprensión e implementación de los modelos de procesamiento; genera resultados con márgenes de aproximación, que implica análisis más cuidadosos de los resultados para evitar interpretaciones erróneas; generalmente suelen ser procesos computacionalmente costosos, más aún, cuando se procesan grandes volúmenes de datos.

### 2.1.3 *Aprendizaje automático*

Una definición formal de aprendizaje automático se tiene Murphy (2021). Sin embargo, desde una perspectiva práctica se puede definir como el conjunto de conceptos y técnicas encargadas de crear sistemas que puedan aprender autónomamente de los datos. El proceso de aprendizaje implica mejorar en alguna tarea, considerando cierta medida de rendimiento. La taxonomía básica del aprendizaje automático considera el aprendizaje supervisado y el aprendizaje no supervisado. Es más fácil de entender el aprendizaje automático con el aprendizaje supervisado, donde el sistema recibe un conjunto de datos en base a los cuales aprende el modelo; el algoritmo de aprendizaje encuentra una función que se aproxima a la verdadera relación entre las entradas y las salidas, de modo que la función se puede utilizar para predecir el resultado cuando se da un nuevo valor de entrada.

El aprendizaje automático (Machine Learning, ML) es una rama de la inteligencia artificial (IA) que se centra en el desarrollo de algoritmos y modelos que permiten a las computadoras aprender y mejorar su rendimiento en tareas específicas a partir de datos, sin necesidad de ser programadas explícitamente para cada tarea. En lugar de seguir reglas específicas, los sistemas de ML utilizan datos para encontrar patrones y tomar decisiones.

Es una disciplina amplia y en constante evolución que abarca una variedad de técnicas y enfoques para el desarrollo de sistemas inteligentes que pueden aprender y mejorar con la experiencia. Su aplicación se encuentra en una amplia gama de campos, desde la medicina y la biología hasta la industria y la tecnología.

Los proyectos de aprendizaje automático implican varias fases en su implementación, estas fases tienen que ser definidas y estructuradas de manera que garanticen un desarrollo efectivo, y resultados con un margen adecuado de fiabilidad. Las fases típicas en los proyectos de aprendizaje automático son:

***Definición del problema***, fase en la que se debe identificar completamente los objetivos del negocio, los requisitos funcionales y no funcionales, las restricciones y las interrogantes que debe responder el proyecto de aprendizaje automático. En suma, un entendimiento cabal del problema.

***Recolección y comprensión de los datos***, fase que consiste en identificar posibles fuentes de datos requeridos en el proyecto. Generalmente, estas fuentes corresponden a los datos operacionales o transaccionales de la organización (Fuentes internas) y a datos que se puede recopilar de internet, de organismos oficiales, mediante encuestas o cualquier otra fuente (Fuentes externas). Para una cabal aplicación del aprendizaje automático, se debe tener una comprensión completa de los datos y tener claro el significado de los mismos.

***Preparación de los datos***, fase que considera tres aspectos: Limpieza de datos, que se encarga de la gestión de valores faltantes, duplicados, valores inconsistentes o incongruentes. Transformación de datos, que se encarga de la codificación de valores categóricos, estandarización y normalización de los datos. Finalmente, Ingeniería de características, que se encarga de la posibilidad de generar nuevas características a partir de las existentes, en la intención de aportar mejores y mayores niveles de información al modelo, así como, mejorar la semántica de alguna de las características, que ayude a mejorar el rendimiento del modelo.

***Selección del modelo***, fase en la que se debe elegir el modelo, en función de la naturaleza de los datos y la pertinencia de la aplicabilidad del modelo, de modo que se seleccione el algoritmo de aprendizaje automático más apropiado para el problema específico. En esta fase también se efectúa la división del conjunto de datos en subconjuntos de entrenamiento evaluación y testeo.

***Entrenamiento del modelo***, que consiste en ajustar el modelo en base a los datos de entrenamiento; luego, dependiendo de la confiabilidad o rendimiento alcanzado por el modelo, determinar la necesidad de efectuar procesos de ajuste de hiperparámetros.

***Evaluación del modelo***, fase encargada de ponderar el rendimiento del modelo, utilizando métricas propias del aprendizaje automático supervisado y no supervisado (precisión, recall, F1-score, RMSE, etc.). Se considera también, validar la confiabilidad o rendimiento del modelo a través de técnicas como k-fold cross-validation.

***Interpretación y explicación del modelo***, fase encargada de interpretar los resultados y el comportamiento del modelo, en base a la identificación y explicación de la importancia de las características, así como, visualizaciones.

***Implementación y despliegue***, fase encargada de poner el modelo en el entorno de producción para su uso real. También es responsable de supervisar el rendimiento del modelo en producción y realizar las adecuaciones requeridas.

***Documentación y comunicación***, fase encargada de documentar todas las fases del proceso, tanto documentación técnica como documentación de usuario, de modo que finalmente, permita comunicar los hallazgos y el impacto del modelo a las partes interesadas a través de informes y presentaciones.

***Retroalimentación y mejora continua***, fase que consiste en recoger retroalimentación de los usuarios y las partes interesadas; para efectuar un proceso iterativo que permita mejorar el modelo con nuevos datos y técnicas mejoradas.

**2.1.3.1 Clasificación.** La taxonomía del aprendizaje automático considera varias clases, siendo las principales:

- **Aprendizaje Supervisado.** Consiste de algoritmos al que se proporciona al algoritmo un conjunto de datos de entrada, donde asociado a cada instancia se tiene una etiqueta (resultado o valor de salida) que representa la semántica de los datos. El algoritmo aprende a realizar predicciones en base a datos conocidos. Se tiene dos tipos: clasificación, que predice un valor discreto asociado a la clase a la que pertenece la instancia; y regresión que predice un valor continuo que corresponde a la ponderación numérica de la instancia.

Los modelos más populares de este tipo son: Regresión Lineal, Regresión Logística, Máquinas de Vectores de Soporte (SVM), Árboles de Decisión, Random Forest, Etc.

- Aprendizaje No Supervisado. Corresponde a los casos en los que se tiene datos sin etiquetar. El objetivo es explorar la estructura oculta en los datos y encontrar patrones de comportamiento, como la segmentación de clientes en grupos basados en sus costumbres de compra. Los modelos más son: K-Means y Mean Shift, DBSCAN para agrupamiento, Análisis de Componentes Principales (PCA), Análisis de Componentes Independientes (ICA), Algoritmo de Agrupamiento Espectral, Etc. Para propósitos de Reducción de Dimensionalidad.
- Aprendizaje por Refuerzo. Corresponde a los casos en los que el modelo interactúa con un entorno del cual recibe retroalimentación en forma de penalizaciones o recompensas, que le permite determinar qué acciones son más beneficiosas y reforzar de esta manera su proceso de aprendizaje.

Además de estos tipos principales, existen otras áreas y técnicas en el campo del aprendizaje automático, como el aprendizaje profundo (Deep Learning), que utiliza redes neuronales profundas para aprender de datos no estructurados como imágenes, texto o voz, y el aprendizaje por transferencia (Transfer Learning), que aprovecha el conocimiento adquirido en una tarea para mejorar el rendimiento en otra tarea relacionada.

**2.1.3.2 Herramientas y librerías.** Existen lenguajes de programación enriquecidos con una gran cantidad de librerías para aprendizaje automático; también existen herramientas de usuario final como Weka, Orange, Etc.

En esta sección, se hace mención a algunas librerías utilizadas en el presente trabajo.

- Scikit-learn: Es una de las librerías más importantes del aprendizaje automático en Python, que proporciona una gran cantidad de funcionalidades para realizar análisis de datos y elaborar modelos predictivos. Considera funcionalidades para pre

procesamiento de datos, así como para aprendizaje automático supervisado y no supervisado.

- **StringIndexer y OneHotEncoder:** Son herramientas de pre procesamiento de datos utilizadas para convertir variables categóricas en variables numéricas en un conjunto de datos. El primero, asigna un índice numérico único a cada valor categórico de una columna de tipo categórico. El segundo, transforma una columna de variables categóricas en múltiples columnas binarias, donde cada columna representa una categoría distinta y tiene un valor binario que indica la presencia o ausencia de esa categoría en cada fila. Ambas librerías, permiten codificar valores categóricos para algoritmos de aprendizaje automático que requieren variables numéricas.
- **NumPy, Pandas y Matplotlib:** Que vienen a ser librerías genéricas, la primera para procesamiento numérico como álgebra lineal; la segunda para procesar datos de manera muy parecida a un gestor de bases de datos; y el último, para efectuar diversos tipos de gráficos.

#### **2.1.4 *Modelo Random Forest***

En esta sección, se revisa el modelo de aprendizaje automático “Random Forest”, debido a que éste se utiliza para el procesamiento estocástico del presente trabajo.

Este modelo fue propuesto y desarrollado por Breiman y Cutler en 2001. Es una técnica de aprendizaje automático utilizada para modelos supervisados de clasificación y regresión; así como para determinar las características más importantes de un conjunto de datos. Tiene por objetivo reducir el riesgo de sobreajuste, basándose principalmente en el hecho de componer los resultados de múltiples árboles de decisión. Durante el entrenamiento del modelo se generan múltiples árboles de decisión, produciendo la clase que es el modo de las clases (clasificación) o la media de las predicciones (regresión) de los árboles individuales.

El aspecto más relevante del modelo Random Forest está orientado a reducir el sobreajuste (overfitting) propendiendo a mantener una alta precisión en las predicciones. Para este propósito, hace uso de dos técnicas importantes: La primera técnica consiste en el uso del **Bagging**, mediante el cual genera múltiples subconjuntos aleatorios de datos de entrenamiento mediante muestreo con reemplazo, y entrenar un árbol de decisión con cada subconjunto, luego, el modelo determina el promedio de las predicciones de estos árboles; lográndose de esta manera, reducir la varianza del modelo y mejorar la generalización a nuevos datos. La segunda técnica consiste en la selección aleatoria de características (Random Subspaces) para cada nodo en cada árbol de decisión, en lugar de considerar todas las características disponibles. Este hecho, permite agregar diversidad adicional en los diferentes árboles, lo que reduce la correlación entre ellos y mejora la robustez del modelo.

Entonces, el hecho de que el Random Forest sea especialmente robusto y preciso, se debe a la combinación de estas técnicas, que propenden a disminuir el riesgo de sobreajuste; mejorando de esta manera la generalización.

De acuerdo al contexto del presente trabajo, tiene mayor relevancia la **estimación de la importancia de las características**, dado que se tiene datos subjetivos de la percepción de los estudiantes del desempeño académico de sus docentes (datos recogidos en las encuestas aplicadas a los estudiantes de manera obligatoria al finalizar cada semestre), y se trata de determinar cuáles de estas características tienen mayor importancia desde la perspectiva de los estudiantes.

Los conceptos básicos requeridos para tener una cabal comprensión del modelo Random forest, son los siguientes:

**Árboles de decisión**, que es un modelo predictivo que se utiliza para tomar decisiones basadas en la división de los datos en subconjuntos homogéneos en función de un conjunto de reglas de decisión. Cada nodo interno del árbol representa una prueba en una característica,

cada rama representa el resultado de la prueba y cada hoja representa una etiqueta de clase (en clasificación) o un valor (en regresión).

**Bagging (Bootstrap Aggregating)**, pertenece a las técnicas denominadas de ensamble, consiste en entrenar varios modelos de manera independiente con diferentes subconjuntos de datos, generar cada subconjunto mediante la técnica de bootstrap (muestreo con reemplazo); luego, utilizar la mayoría del conteo de votos para modelo supervisado de clasificación o el promedio de las predicciones para modelos supervisados de regresión.

**Random Subspaces**, que se refiere al hecho de generar una mayor diversidad de árboles, mediante la selección de un subconjunto aleatorio de características para cada división de nodo en los árboles de decisión. Se logra reducir la correlación entre los árboles.

El funcionamiento del modelo Random Forest, considera básicamente dos fases: la construcción del bosque y la predicción en base al modelo entrenado. A continuación se detalla ambas fases:

La **construcción del bosque**, implica primero seleccionar múltiples subconjuntos aleatorios de los datos de entrenamiento mediante muestreo con reemplazo (bootstrap); luego, para cada subconjunto, se construye un árbol de decisión, donde, en cada nodo, se selecciona un subconjunto aleatorio de características y se elige la mejor característica para dividir los datos en ese nodo; finalmente, cada árbol se construye hasta la máxima profundidad posible sin poda.

La predicción en el modelo Random Forest se determina en función al tipo de modelo supervisado y las predicciones realizadas por cada árbol individual. Si es un modelo de clasificación, se determina en base a la mayoría de votos de las predicciones entre los árboles individuales. Si es un modelo de regresión, se determina en base al promedio de las predicciones individuales de los árboles.

El modelo Random Forest ofrece una serie de ventajas, tal como las siguientes: La robustez al sobreajuste, lo que hace que el modelo sea más eficiente respecto al modelo de árboles de decisión simples; esto, debido a que se usa múltiples árboles de decisión. Otra de sus ventajas es que permite el manejo de grandes conjuntos de datos y con muchas características. Otra ventaja es la capacidad de adaptarse tanto a problemas de clasificación como a los de regresión. Finalmente, desde la perspectiva del presente trabajo, la ventaja más relevante es la estimación de la importancia de las características, porque proporciona una medida de la importancia de cada característica en la predicción.

El modelo Random Forest presenta también desventajas, siendo la más relevante la complejidad temporal, dado que, el proceso de construcción y predicción de múltiples árboles de decisión implica mayores tiempos de procesamiento. También se considera como una desventaja la interpretabilidad, en el sentido de que los resultados se obtienen de las predicciones de los árboles individuales, entonces puede dificultar la interpretabilidad del modelo global.

### **2.1.5 Rendimiento académico**

El rendimiento académico en las universidades es un tema de gran importancia, tanto para los estudiantes como para la misma universidad. Hace referencia al desempeño de los estudiantes en los procesos de enseñanza - aprendizaje, evaluándose en función a los resultados obtenidos por los estudiantes en los diferentes procesos de evaluación en las diferentes asignaturas, los porcentajes de aprobados y de deserción. El rendimiento académico es crucial en el proceso de formación profesional de los estudiantes, ya que un mejor rendimiento implicará mejores oportunidades a futuro.

Para las universidades, la calidad de los procesos de enseñanza – aprendizaje y el éxito de las escuelas profesionales, se miden generalmente en base al rendimiento. En consecuencia, se deben implementar estrategias adecuadas para garantizar la formación profesional de los

estudiantes; estrategias, como el uso de modelos educativos por competencias, tutoría académica, la implementación de tecnologías educativas, personalización de la enseñanza, Etc.

Es importante considerar que el rendimiento académico debe tener en cuenta Aspectos adicionales a la sola evaluación de los conocimientos adquiridos por el estudiante, sino también, debe considerar otros aspectos, como investigación formativa, actividades extracurriculares, elaboración de proyectos que permitan resolver problemas, Etc. Se debe considerar también, el desarrollo de habilidades blandas, como el trabajo en equipo, capacidad de comunicación, Etc. Igualmente, se propender a fomentar la creatividad y la innovación.

Las universidades también deben tener en cuenta otros factores que pueden incidir en el rendimiento académico, tales como factores personales, nivel socioeconómico, estado emocional y salud mental, contexto familiar, Etc. Para estos casos, se debe considerar estrategias de apoyo a los estudiantes (Por ejemplo, apoyo psicológico), que les permita a los estudiantes garantizar su éxito académico.

En síntesis, es vital para las universidades el rendimiento académico de los estudiantes en el proceso de formación profesional; en consecuencia, se deben efectuar procesos de evaluación en base a indicadores, que consideren la actividad de enseñanza – aprendizaje desde una perspectiva holística; teniendo en cuenta aspectos netamente académicos, como aspectos personales y sociales de los estudiantes. Por lo tanto, es fundamental implementar mecanismos que coadyuven a garantizar el éxito académico de todos los estudiantes.

Se tiene amplia información sobre este concepto en el trabajo de García (2022) que considera un análisis exploratorio en las universidades, respecto a los factores asociados al rendimiento académico. Igualmente, en el trabajo de Rodríguez (2020) se plantean estrategias de apoyo para mejorar el rendimiento de los estudiantes universitarios.

### **2.1.6 Perfil del docente**

En la educación superior, el papel del profesor universitario es fundamental en el desarrollo académico y personal de los estudiantes; por ello, resulta importante contar con un perfil docente que reúna las habilidades y competencias necesarias para cumplir con éxito este rol. En este sentido, un marco teórico para el perfil docente de los profesores universitarios puede resultar de gran utilidad, ya que nos permite identificar las características fundamentales que deben poseer estos profesionales para desempeñar su labor de manera eficiente y eficaz.

Información sobre el perfil del docente universitario se puede revisar en el trabajo de Aguilar (2021) y en el trabajo de Chicaiza (2019), quienes enfatizan el perfil del docente universitario para el siglo XXI, así como las competencias que deben tener los docentes en los escenarios actuales de la educación, donde se tiene modelos educativos por competencias. Los siguientes párrafos hacen referencia a los trabajos citados.

Entre las principales habilidades y competencias que debe poseer un profesor universitario, encontramos su sólido conocimiento disciplinar en su área de especialidad, el dominio de competencias pedagógicas que le permitan transmitir eficazmente el conocimiento, habilidades de comunicación que le permitan interactuar efectivamente con sus estudiantes y colegas, así como una actitud ética y de compromiso social que le permita contribuir al desarrollo de una sociedad más justa y equitativa.

Así mismo, resulta fundamental que los profesores universitarios sean capaces de adaptarse a los cambios y a las nuevas tendencias educativas y tecnológicas, promoviendo la innovación y actualización constante en su práctica docente. También es importante que estén dispuestos a trabajar en equipo y colaborar con sus colegas, y que sean flexibles y capaces de adaptarse a las necesidades de sus estudiantes y del entorno.

En síntesis, el perfil docente de los profesores universitarios debe contemplar una combinación de habilidades y competencias que les permitan desempeñarse de manera efectiva

en su rol docente y contribuir al desarrollo de una sociedad más justa, equitativa y sostenible. Este marco teórico puede ser de gran utilidad para el diseño de políticas y estrategias que fomenten la formación y el desarrollo de estos profesionales.

- **Conocimiento de la disciplina:** Es vital que los docentes de las diferentes escuelas profesionales de las universidades tengan una formación sólida en su especialidad.
- **Competencias pedagógicas:** Adicionalmente al conocimiento de la especialidad, es necesario que los docentes universitarios cuenten también con formación pedagógica, lo que les permitirá una fluida interacción con sus estudiantes, transmitir adecuadamente el conocimiento, diseñar sesiones más eficientes, realizar actividades de evaluación y retroalimentación, incentivar la participación y el pensamiento crítico.
- **Habilidades de comunicación:** Otro aspecto fundamental en los docentes universitarios es la capacidad de comunicarse de manera efectiva, que le permita desarrollar habilidades de oratoria, escritura y escucha activa.
- **Innovación y actualización constante:** El conocimiento científico es dinámico, lo que condiciona a los docentes universitarios a estar siempre actualizados en su especialidad, como también en tecnologías y tendencias educativas. Propiciará en ellos fomentar la innovación en su actividad educativa, así como una fácil adaptación a los cambios.
- **Actitud ética y compromiso social:** Considerando que la influencia de los docentes en sus estudiantes es determinante, los docentes deben tener siempre una actitud ética y de compromiso social; promoviendo principios y valores como la honestidad, la solidaridad y la responsabilidad. También deben estar comprometidos en inculcar a los futuros profesionales conciencia social que contribuya al bienestar de la sociedad.
- **Trabajo en equipo y colaboración:** El avance vertiginoso de la ciencia, exige que los docentes universitarios tengan que trabajar en equipo con sus colegas; efectuar un

intercambio de ideas y experiencias, impulsar proyectos incluso multidisciplinares; actividades que sin duda redundará en una mejora significativa del ambiente académico.

- Flexibilidad y adaptación: Considerando la frase que lo único constante es el cambio, los profesores universitarios deben tener la capacidad de adaptarse al cambio, asimilar las nuevas tendencias y estrategias, que en algunos casos plantean la necesidad de adaptarse a las necesidades y particularidades de cada estudiante y grupo; lo que les permitirá afrontar de manera efectiva a los retos del nuevo contexto de la labor docente.

### III. MÉTODO

#### 3.1 Tipo de investigación

De acuerdo con lo planteado por Hernández y Mendoza (2018), el presente trabajo corresponde a una investigación cuantitativa, se utilizará métodos y técnicas de ciencia de datos para el análisis de los datos, que permitirá someter a contrastación las hipótesis planteadas, que luego, posibilitará obtener e interpretar los resultados que conducirán a las conclusiones.

Respecto a la finalidad de la investigación, el trabajo pertenece a un tipo de investigación aplicada; porque, se utilizará conceptos tecnológicos determinísticos y estocásticos de vanguardia como es la ciencia de datos, en la solución de problemas inherentes al contexto académico de universidades públicas de la región del Cusco.

Desde la perspectiva del alcance de la investigación, ésta corresponde al tipo explicativo; porque busca identificar y explicar las relaciones causales entre variables; específicamente, se relacionan variables como las diferentes características distintivas de los docentes tal el grado académico, experiencia, segmento etario, género, categoría, régimen, percepción o valoración de los docentes por parte de los estudiantes; para determinar el grado de asociación e identificar las que tienen relaciones preponderantes con el rendimiento académico de los estudiantes.

Desde la perspectiva del diseño de la investigación, se considera ésta como una investigación no experimental, porque no se manipularán variables, sino, sólo se utilizará registros oficiales de datos académicos de bases de datos transaccionales que corresponden al comportamiento real de los estudiantes y docentes de las universidades públicas de la región del Cusco.

## **3.2 Población y muestra**

### **3.2.1 Población**

Las universidades públicas de la región Cusco cuentan con datos de registros oficiales en las bases de datos transaccionales; por tanto, para determinar la implicancia de las características distintivas de los docentes en el rendimiento académico de los estudiantes mediante técnicas de ciencia de datos, se utilizará la totalidad de los datos de estos registros oficiales; más aún, si se considera que estas técnicas implican procesos computacionales.

Por tanto, la población está constituida por la totalidad de los datos de registros oficiales subyacentes en las bases de datos transaccionales, considerando sólo los semestres correspondientes a las clases presenciales, semestres entre el 2021-1 y el 2023-2.

### **3.2.2 Muestra**

Una ventaja de los trabajos de investigación computacionales, es que permiten trabajar con gran cantidad de datos; por tanto, se considera para la muestra la totalidad de la población definida, es decir, los datos académicos de las bases de datos transaccionales pertinentes para determinar las implicancias de las características distintivas de los docentes en el rendimiento académico de los estudiantes en las universidades de la región del Cusco.

### 3.3 Operacionalización de variables

El detalle de la operacionalización de variables se muestra en la tabla 1.

**Tabla 1**

*Operacionalización de variables*

<b>VARIABLES</b>	<b>DIMENSIONES</b>	<b>INDICADORES</b>	<b>ESCALA</b>	<b>MEDIDA</b>
<b>VARIABLE 1</b>	1. Características distintivas personales	1.1 Grado académico	Nominal	Maestría/doctorado
Características distintivas de los docentes		1.2 Experiencia	Razón	Años
		1.3 Grupo etario	Nominal	Número grupo
		1.4 Género	Nominal	Femenino/masculino
		1.5 Categoría	Nominal	Principal/Asociado/Auxiliar
		1.6 Régimen	Nominal	Dedicación exclusiva/tiempo completo/tiempo parcial
	2. Valoración de los docentes por parte de los estudiantes	2.1 Ponderación numérica	Razón	Valor numérico entre 0 y 10

Rendimiento académico	3. Rendimiento académico	3.1 Porcentaje de aprobados	Razón	Porcentaje
		3.2 Nota promedio asignada por el profesor	Razón	Valor numérico entre 0 y 20

## VARIABLE 2

Implicancia de las características distintivas de los docentes en el rendimiento académico de los estudiantes	4. Implicancia de las características distintivas de los docentes en el rendimiento académico de los estudiantes
---	--

### **3.4 Instrumentos**

Como instrumentos para la recolección de datos se usan documentos técnicos utilizados en el campo de la inteligencia artificial y ciencia de datos denominados DataSets (Conjuntos de datos) en formato CSV (Comma Separated Values); en este documento, se recupera los datos de los registros oficiales subyacentes en las bases de datos transaccionales. Como se mencionó anteriormente, se limita el trabajo a los 3 últimos años, semestres comprendidos entre el 2021-1 hasta el 2023-2, de los que se utiliza la totalidad de los datos.

Los datos corresponden al registro oficial de las notas asignadas por los docentes a los estudiantes en cada asignatura, semestre a semestre a lo largo de los semestres indicados. También se cuenta con los datos de encuestas oficiales aplicadas a los estudiantes al finalizar cada semestre, donde se consigna la valoración que efectúa cada estudiante del desempeño del docente en cada asignatura; situación que, garantiza la validez y la confiabilidad de los datos.

Es necesario poner a relieve, que este trabajo de investigación se realizará ceñido a los principios éticos de la investigación científica, considerando sobre todo no utilizar información altamente sensible como nombres, direcciones, documentos de identidad, etc. que garanticen la privacidad y confidencialidad de los participantes.

### **3.5 Procedimientos**

Para la primera fase del trabajo se considera lo referente a la investigación documental, tomando en cuenta aspectos como los antecedentes y el marco teórico requerido.

Para la fase de recolección de datos se utiliza un proceso de gestión de datos denominado ETL (Extract, transform and Load – Extraer, transformar y cargar); para lo cual, se implementan procesos computacionales que permiten: Extraer datos de las bases de datos transaccionales, uniformar y adecuar los datos a los requerimientos del presente trabajo y finalmente cargar a los denominados dataset's en formato CSV.

En la fase del procesamiento de los datos, para determinar la relación existente entre las características distintivas de los docentes y la valoración de los estudiantes hacia los docentes con el rendimiento académico de los estudiantes, se aplica técnicas determinísticas como SQL y técnicas estocásticas con modelos de aprendizaje automático, identificando para ello claramente las variables independientes y dependientes; lo que permitirá coadyuvar a plantear estrategias más adecuadas para fortalecer el proceso de mejora continua.

La presentación de resultados se realiza en tablas y gráficos, sobre las que se efectúa la interpretación y discusión de resultados; para finalmente, determinar las conclusiones.

### **3.6 Análisis de datos**

El análisis de datos en esta investigación se efectúa considerando una perspectiva metodológica a dos niveles de procesamiento: uno determinístico y otro estocástico; para lo cual, se considera primordialmente la complejidad de las relaciones entre los atributos del conjunto de datos. Esta estrategia permite realizar un análisis exhaustivo de la información, posibilitando la extracción o generación de conocimiento relevante en el contexto de la investigación.

*Análisis en base a procesos determinísticos*, éste se efectúa sobre los datos originados de manera objetiva y que presentan relaciones entre atributos obvias y claramente identificables, tal como las notas obtenidas por los estudiantes y las características distintivas más relevantes de los docentes. Se utilizan métodos determinísticos basados en lenguajes de consulta como SQL. Este enfoque permite extraer información en base a consultas estructuradas y eficientes. El uso de SQL permite transformar y gestionar los datos de manera precisa, lo que posibilita un mejor análisis e interpretación de las relaciones entre los atributos.

*Análisis en base a procesos estocásticos*, éste se efectúa sobre los datos que tienen un origen en base a criterios subjetivos, como son las encuestas realizadas a los estudiantes sobre la evaluación a sus docentes. Se aplica técnicas de aprendizaje automático de datos para un

análisis exploratorio de los datos que permite entender las características y la naturaleza de los datos; luego, como consecuencia de este análisis exploratorio se determina los atributos o características distintivas que tienen relevancia en la perspectiva de los estudiantes respecto a sus docentes; así como, la pertinencia de las diferentes preguntas de la encuesta en la intención de una evaluación más apropiada de los docentes por parte de los estudiantes.

## IV. RESULTADOS

Este capítulo plasma los resultados obtenidos. Los tres primeros objetivos se alcanzan con procesamientos determinísticos, para lo cual se utiliza el lenguaje de bases de datos relacionales SQL; mientras, que el último objetivo se alcanza con un procesamiento estocástico, para lo cual se utiliza un modelo de ciencia de datos.

También cabe mencionar, que los datos originales recopilados del Centro de Cómputo de la Universidad Nacional, corresponden a datos de bases de datos transaccionales. Estos datos fueron pre procesados y convertidos a formato CSV, donde no se consideraron datos sensibles como nombres o documentos de identificación. La estructura del referido archivo se muestra en la tabla 2:

**Tabla 2**

*Estructura del archivo CSV con datos originales*

Nro	Atributo	Tipo	Descripción
1	Docente	texto	Identificador de docente
2	IdDepartamento	texto	Identificador de departamento académico
3	Facultad	texto	Identificador de facultad
4	Regimen	texto	Régimen del docente
5	Categoria	texto	Categoría del docente
6	Experiencia	entero	Número de años de experiencia del docente
7	GradoAcademico	texto	Grado académico del docente
8	Edad	entero	Edad del docente
9	Genero	texto	Genero del docente
10	Semestre	texto	Semestre de evaluación
11	Curso	texto	Asignatura de evaluación
12	Creditos	entero	Créditos de la asignatura de evaluación
13	Alumno	texto	Identificador del alumno
14	CP	texto	Identificador de Carrera profesional
15	P1	decimal	Promedio del primer periodo parcial
16	P2	decimal	Promedio del segundo periodo parcial
17	P3	decimal	Promedio del tercer periodo parcial
18	promedio	entero	Promedio final
19	Resultado	entero	Reprobado, Desaprobado y Aprobado
20	P01	entero	Preg. 01 de encuesta de desempeño docente
21	P02	entero	Preg. 02 de encuesta de desempeño docente
22	P03	entero	Preg. 03 de encuesta de desempeño docente

23	P04	entero	Preg. 04 de encuesta de desempeño docente
24	P05	entero	Preg. 05 de encuesta de desempeño docente
25	P06	entero	Preg. 06 de encuesta de desempeño docente
26	P07	entero	Preg. 07 de encuesta de desempeño docente
27	P08	entero	Preg. 08 de encuesta de desempeño docente
28	P09	entero	Preg. 09 de encuesta de desempeño docente
29	P10	entero	Preg. 10 de encuesta de desempeño docente
30	P11	entero	Preg. 11 de encuesta de desempeño docente
31	P12	entero	Preg. 12 de encuesta de desempeño docente
32	P13	entero	Preg. 13 de encuesta de desempeño docente
33	P14	entero	Preg. 14 de encuesta de desempeño docente
34	P15	entero	Preg. 15 de encuesta de desempeño docente
35	P16	entero	Preg. 16 de encuesta de desempeño docente
36	P17	entero	Preg. 17 de encuesta de desempeño docente
37	P18	entero	Preg. 18 de encuesta de desempeño docente
38	P19	entero	Preg. 19 de encuesta de desempeño docente
39	P20	entero	Preg. 20 de encuesta de desempeño docente
40	Puntaje	decimal	Puntaje global del desempeño docente

Sobre estos datos se efectuó el procesamiento determinístico como el estocástico. Para el procesamiento determinístico los datos fueron pre-procesados y cargados a una base de datos relacional, donde se aplicó sentencias SQL para obtener los resultados determinísticos. Mientras que, para el procesamiento estocástico, los datos se utilizaron en el formato CSV.

#### **4.1 Procesamiento determinístico**

##### **4.1.1 *Influencia del grado académico y la experiencia de los docentes***

Para determinar en qué medida influyen el grado académico y la experiencia de los docentes en el rendimiento académico de los estudiantes en las universidades públicas de la región del Cusco, se consideró los atributos:

Grado Académico.- Grado académico del docente (Doctorado, Maestría o Título profesional).

Experiencia.- Expresado en quinquenios de haber ingresado a la docencia.

Promedio.- Nota final obtenida por un estudiante en una asignatura regentada por un docente.

Puntaje.- Valoración dada por un estudiante al desempeño del docente en una asignatura.

Luego, se efectuó un procesamiento determinístico, para lo cual se implementó sentencias SQL, las que se detallan a continuación.

**4.1.1.1 Influencia del grado académico de los docentes.** Para determinar en qué medida influye el grado académico de los docentes en el rendimiento académico de los estudiantes, se implementó dos funciones en SQL, la primera para pre procesar y resumir los datos, cuyo código se muestra en la figura 3:

### Figura 3

*Script para Pre-procesar y Resumir los Datos por Grado académico*

```

/* *****
Nombre:          fn_Resumen_GradoAcademico
Descripción:     funcion para recuperar el resumen por grado académico
Parametros Entrada: Ninguno
OUTPUT:
Ejemplo:
      Select * from fn_Resumen_GradoAcademico()
***** */
IF OBJECT_ID (N'fn_Resumen_GradoAcademico', N'IF') IS NOT NULL
  drop function fn_Resumen_GradoAcademico
Go
create function fn_Resumen_GradoAcademico()
returns Table
as
return(
  SELECT  [GradoAcademico]
          ,avg(Promedio*1.0) as Promedio
          ,sum(case when Promedio >= 14 then 1 else 0 end) as Aprobados
          ,sum(case when Promedio >= 9 and Promedio < 14 then 1 else 0 end) as Desaprobados
          ,sum(case when Promedio < 9 then 1 else 0 end) as Reprobados
          ,avg([Puntaje]) as Puntaje
          ,count(*) as NroAlumnos
  FROM [BD_Docentes].[dbo].[Info_Academica_]
  group by [GradoAcademico]
)
go

```

El segundo script, calcula los ratios que determinan la medida en que el grado académico de los docentes influye en el rendimiento de los estudiantes. El respectivo script se muestra en la figura 4:

## Figura 4

### Script para determinar los ratios de la influencia del Grado académico

```

/* *****
Nombre:          fn_Resumen_GradoAcademico_Ratio
Descripción:     funcion para recuperar el ratio por grado académico
Parametros Entrada: Ninguno
OUTPUT:
Ejemplo:
    Select * from fn_Resumen_GradoAcademico_Ratio()
***** */
IF OBJECT_ID (N'fn_Resumen_GradoAcademico_Ratio', N'IF') IS NOT NULL
drop function fn_Resumen_GradoAcademico_Ratio
Go
create function fn_Resumen_GradoAcademico_Ratio()
returns Table
as
return(
    SELECT [GradoAcademico]
           ,cast(Promedio as numeric(5,2)) as Promedio
           ,cast(Aprobados*100.0/NroAlumnos as numeric(5,2)) as Aprobado
           ,cast(Desaprobados*100.0/NroAlumnos as numeric(5,2)) as Desaprobado
           ,cast(Reprobados*100.0/NroAlumnos as numeric(5,2)) as Reprobado
           ,cast(Puntaje as numeric(5,2)) as Puntaje
           ,NroAlumnos
    FROM fn_Resumen_GradoAcademico()
)
go

```

Ejecutando este segundo script que a su vez invoca a la primera función, se tiene la influencia del grado académico de los docentes en el rendimiento académico de los estudiantes.

Los resultados obtenidos se muestran en la tabla 3.

**Tabla 3**

### *Influencia del Grado Académico en el rendimiento de los estudiantes*

Influencia del Grado Académico de los Docentes en el Rendimiento Académico de los Estudiantes						
Semestres: 2021-1 al 2023-2						
GradoAcademico	Promedio	Aprobado	Desaprobado	Reprobado	Puntaje	NroAlumnos
DOCTORADO	14.54	85.65	7.47	6.88	7.9	102010
MAESTRIA	14.11	81.9	10.61	7.49	7.68	80269
TITULO	14.11	82.35	10.07	7.58	7.66	44022

Donde, el atributo o columna *Promedio* corresponde a la nota final obtenida por todos los estudiantes en las asignaturas regentadas por los docentes con diferentes grados académicos; así mismo, los atributos *Aprobado*, *Desaprobado* y *Reprobado* se expresan en porcentajes según corresponda; finalmente, el atributo *Puntaje* corresponde al promedio de la valoración realizada por los estudiantes a sus respectivos docentes.

**4.1.1.2 Influencia de la experiencia de los docentes.** Para determinar en qué medida influye la experiencia de los docentes en el rendimiento académico de los estudiantes, se consideró el número de años dedicados a la enseñanza por cada docente, años que se agruparon en quinquenios. En este caso, también se implementó dos funciones en SQL, la primera para pre-procesar y resumir los datos por quinquenio de experiencia; el respectivo script se muestra en la figura 5:

### Figura 5

#### *Script para Pre-procesar y Resumir los Datos por Experiencia*

```

/* *****
Nombre:          fn_Resumen_Experiencia
Descripción:     funcion para recuperar el resumen por Experiencia
Parametros Entrada: Ninguno
OUTPUT:
Ejemplo:
    Select * from fn_Resumen_Experiencia()
***** */
IF OBJECT_ID (N'fn_Resumen_Experiencia', N'IF') IS NOT NULL
drop function fn_Resumen_Experiencia
Go
create function fn_Resumen_Experiencia()
returns Table
as
return(
    WITH InfoExperiencia(Experiencia, Promedio, Puntaje)
    as
    (
        SELECT case
            when Experiencia < 5 then '<5'
            when Experiencia >= 5 and Experiencia < 10 then '05-10'
            when Experiencia >= 10 and Experiencia < 15 then '10-15'
            when Experiencia >= 15 and Experiencia < 20 then '15-20'
            when Experiencia >= 20 and Experiencia < 25 then '20-25'
            when Experiencia >= 25 then '>25'
        end as Experiencia, Promedio, Puntaje
        FROM [BD_Docentes].[dbo].[Info_Academica_]
    )
    SELECT [Experiencia]
        ,avg(Promedio*1.0) as PromedioExperiencia
        ,sum(case when Promedio >= 14 then 1 else 0 end) as Aprobados
        ,sum(case when Promedio >= 9 and Promedio < 14 then 1 else 0 end) as Desaprobados
        ,sum(case when Promedio < 9 then 1 else 0 end) as Reprobados
        ,avg([Puntaje]) as Puntaje
        ,count(*) as NroAlumnos
    FROM InfoExperiencia
    group by [Experiencia]
)

```

El segundo script, calcula los ratios que determinan la medida en que la experiencia de los docentes influye en el rendimiento académico. El script se muestra en la figura 6:

**Figura 6**

*Script para determinar los ratios de la influencia de la experienci*

```

/* *****
Nombre:          fn_Resumen_Experiencia_Ratio
Descripción:     funcion para recuperar el ratio por Experiencia
Parametros Entrada: Ninguno
OUTPUT:
Ejemplo:
    Select * from fn_Resumen_Experiencia_Ratio() order by Puntaje desc
***** */
IF OBJECT_ID (N'fn_Resumen_Experiencia_Ratio', N'IF') IS NOT NULL
drop function fn_Resumen_Experiencia_Ratio
Go
create function fn_Resumen_Experiencia_Ratio()
returns Table
as
return(
    SELECT  [Experiencia]
            ,cast(PromedioExperiencia as numeric(5,2)) as Promedio
            ,cast(Aprobados*100.0/NroAlumnos as numeric(5,2)) as Aprobado
            ,cast(Desaprobados*100.0/NroAlumnos as numeric(5,2)) as Desaprobado
            ,cast(Reprobados*100.0/NroAlumnos as numeric(5,2)) as Reprobado
            ,cast(Puntaje as numeric(5,2)) as Puntaje
            ,NroAlumnos
    FROM fn_Resumen_Experiencia()
)
go

```

Ejecutando los scripts, se tiene la influencia de la experiencia de los docentes en el rendimiento académico de los estudiantes. Los resultados obtenidos se muestran en la tabla 4.

**Tabla 4**

*Influencia de la experiencia en el rendimiento de los estudiantes*

Influencia de la Experiencia de los Docentes en el Rendimiento Académico de los Estudiantes						
Semestres: 2021-1 al 2023-2						
Experiencia	Promedio	Aprobado	Desaprobado	Reprobado	Puntaje	NroAlumnos
05-10	12.59	68.55	13.38	18.07	7.58	1644
10-15	14.19	82.90	8.75	8.35	7.89	26393
15-20	14.51	84.83	8.16	7.02	7.87	45215
20-25	14.27	82.76	10.01	7.23	7.89	41802
>25	14.28	83.96	9.14	6.90	7.67	111247

Donde, la experiencia se expresa en quinquenios, los demás atributos tienen la misma semántica que en el caso anterior.

**4.1.1.3 Influencia del grado académico y la experiencia de los docentes.** Para determinar en qué medida influye de manera conjunta el grado académico y la experiencia de los docentes en el rendimiento académico de los estudiantes, se consideró estos dos atributos

como dimensiones de resumen; también se implementaron dos scripts en SQL, el primero para el pre-procesamiento y resumen de los datos, cuyo código se muestra en la figura 7:

### Figura 7

*Script para Pre-procesar y Resumir los Datos por Grado Académico y Experiencia*

```

/* *****
Nombre:          fn_Resumen_GradoAcademico_Experiencia
Descripción:     funcion para recuperar el resumen por Grado Académico y Experiencia
Parametros Entrada: Ninguno
OUTPUT:
Ejemplo:
    Select * from fn_Resumen_GradoAcademico_Experiencia()
***** */
IF OBJECT_ID (N'fn_Resumen_GradoAcademico_Experiencia', N'IF') IS NOT NULL
drop function fn_Resumen_GradoAcademico_Experiencia
Go
create function fn_Resumen_GradoAcademico_Experiencia()
returns Table
as
return(
    WITH InfoGradoAcademicoExperiencia(GradoAcademico, Experiencia, Promedio, Puntaje)
    as
    (
        SELECT GradoAcademico,
            case
                when Experiencia < 5 then '<5'
                when Experiencia >= 5 and Experiencia < 10 then '05-10'
                when Experiencia >= 10 and Experiencia < 15 then '10-15'
                when Experiencia >= 15 and Experiencia < 20 then '15-20'
                when Experiencia >= 20 and Experiencia < 25 then '20-25'
                when Experiencia >= 25 then '>25'
            end as Experiencia, Promedio, Puntaje
        FROM [BD_Docentes].[dbo].[Info_Academica_]
    )
    SELECT GradoAcademico
        ,[Experiencia]
        ,avg(Promedio*1.0) as PromedioExperiencia
        ,sum(case when Promedio >= 14 then 1 else 0 end) as Aprobados
        ,sum(case when Promedio >= 9 and Promedio < 14 then 1 else 0 end) as Desaprobados
        ,sum(case when Promedio < 9 then 1 else 0 end) as Reprobados
        ,avg([Puntaje]) as Puntaje
        ,count(*) as NroAlumnos
    FROM InfoGradoAcademicoExperiencia
    group by GradoAcademico, [Experiencia]
)

go

```

El segundo script, calcula los ratios que determinan la medida en que el grado académico y la experiencia de los docentes influyen en el rendimiento de los estudiantes. El script se muestra en la figura 8:

## Figura 8

*Script para determinar los ratios de la influencia del Grado Académico y la experiencia*

```
-- Select * from fn_Resumen_GradoAcademico_Experiencia() order by Experiencia, GradoAcademico

/* *****
Nombre:          fn_Resumen_GradoAcademico_Experiencia_Ratio
Descripción:     funcion para recuperar el ratio por Grado Académico y Experiencia
Parametros Entrada: Ninguno
OUTPUT:
Ejemplo:
      Select * from fn_Resumen_GradoAcademico_Experiencia_Ratio order by Puntaje desc
***** */
IF OBJECT_ID (N'fn_Resumen_GradoAcademico_Experiencia_Ratio', N'IF') IS NOT NULL
drop function fn_Resumen_GradoAcademico_Experiencia_Ratio
Go
create function fn_Resumen_GradoAcademico_Experiencia_Ratio()
returns Table
as
return(
      SELECT  GradoAcademico
            , [Experiencia]
            , cast(PromedioExperiencia as numeric(5,2)) as Promedio
            , cast(Aprobados*100.0/NroAlumnos as numeric(5,2)) as Aprobado
            , cast(Desaprobados*100.0/NroAlumnos as numeric(5,2)) as Desaprobado
            , cast(Reprobados*100.0/NroAlumnos as numeric(5,2)) as Reprobado
            , cast(Puntaje as numeric(5,2)) as Puntaje
            , NroAlumnos
      FROM    fn_Resumen_GradoAcademico_Experiencia()
)
go
```

Ejecutando los scripts, se tiene la influencia conjunta del grado académico y la experiencia de los docentes en el rendimiento académico de los estudiantes. Los resultados obtenidos se muestran en la tabla 5.

**Tabla 5**

*Influencia del Grado Académico y la Experiencia en el Rendimiento de los Estudiantes*

Influencia del Grado Académico y la Experiencia de los Docentes en el Rendimiento Académico de los Estudiantes						
Semestres: 2021-1 al 2023-2						
Experiencia	Promedio	Aprobado	Desaprobado	Reprobado	Puntaje	NroAlumnos
DOCTORADO						
05-10	14.20	83.91	7.39	8.70	8.61	230
10-15	14.47	85.50	6.87	7.63	7.99	12263
15-20	14.66	86.21	6.83	6.96	7.97	23082
20-25	14.60	85.68	7.23	7.10	8.03	19802
>25	14.48	85.40	8.05	6.55	7.79	46633
MAESTRÍA						
05-10	12.42	66.12	15.73	18.15	7.69	1036.00
10-15	14.40	84.32	9.83	5.85	7.92	10862.00

15-20	14.34	83.56	9.64	6.80	7.85	14560.00
20-25	13.86	78.40	13.62	7.98	7.80	12792.00
>25	14.06	82.15	10.10	7.75	7.51	41019.00
TÍTULO						
05-10	12.10	65.87	10.58	23.54	6.62	378.00
10-15	12.47	68.39	12.24	19.37	7.39	3268.00
15-20	14.36	83.02	9.36	7.62	7.62	7573.00
20-25	14.12	82.53	11.00	6.47	7.71	9208.00
>25	14.28	84.26	9.63	6.10	7.72	23595.00

#### 4.1.2 *Influencia del grupo etario y el género de los docentes*

Para determinar en qué medida influyen el grupo etario y el género de los docentes en el rendimiento académico de los estudiantes en las universidades públicas de la región del Cusco, se consideró los atributos:

Edad.- Del docente expresado en decenios.

Género.- 0 = Femenino, 1 = Masculino.

Promedio.- Nota final obtenida por un estudiante en una asignatura regentada por un docente.

Puntaje.- Valoración dada por un estudiante al desempeño del docente en una asignatura.

Luego, se efectuó un procesamiento determinístico, para lo cual se implementó sentencias SQL, las que se detallan a continuación.

**4.1.2.1 Influencia del grupo etario.** Para determinar en qué medida influye el grupo etario de los docentes en el rendimiento académico de los estudiantes, se implementó dos funciones en SQL, la primera para pre-procesar y resumir los datos por grupo etario; el respectivo script se muestra en la figura 9:

## Figura 9

### Script para Pre-procesar y Resumir los Datos por Grupo Etario

```

/* *****
Nombre:          fn_Resumen_GrupoEtario
Descripción:     funcion para recuperar el resumen por GrupoEtario
Parametros Entrada: Ninguno
OUTPUT:
Ejemplo:
      Select * from fn_Resumen_GrupoEtario()
***** */
IF OBJECT_ID (N'fn_Resumen_GrupoEtario', N'IF') IS NOT NULL
  drop function fn_Resumen_GrupoEtario
Go
create function fn_Resumen_GrupoEtario()
returns Table
as
return(
  WITH InfoGrupoEtario(GrupoEtario, Promedio, Puntaje)
  as
  (
    SELECT case
      when Edad < 30 then '<30'
      when Edad >= 30 and Edad < 40 then '30-40'
      when Edad >= 40 and Edad < 50 then '40-50'
      when Edad >= 50 and Edad < 60 then '50-60'
      when Edad >= 60 and Edad < 70 then '60-70'
      when Edad >= 70 then '>70'
    end as GrupoEtario, Promedio, Puntaje
    FROM [BD_Docentes].[dbo].[Info_Academica_]
  )
  SELECT [GrupoEtario]
    ,avg(Promedio*1.0) as PromedioGrupoEtario
    ,sum(case when Promedio >= 14 then 1 else 0 end) as Aprobados
    ,sum(case when Promedio >= 9 and Promedio < 14 then 1 else 0 end) as Desaprobados
    ,sum(case when Promedio < 9 then 1 else 0 end) as Reprobados
    ,avg([Puntaje]) as Puntaje
    ,count(*) as NroAlumnos
  FROM InfoGrupoEtario
  group by [GrupoEtario]
)
go

```

El segundo script calcula los ratios que determinan la medida en que el grupo etario de los docentes influye en el rendimiento de los estudiantes. El respectivo script se muestra en la figura 10:

## Figura 10

*Script para determinar los ratios de la influencia del Grupo etario*

```

/* *****
Nombre:          fn_Resumen_GrupoEtario_Ratio
Descripción:     funcion para recuperar el ratio por GrupoEtario
Parametros Entrada: Ninguno
OUTPUT:
Ejemplo:
  Select * from fn_Resumen_GrupoEtario_Ratio() order by Puntaje desc
***** */

IF OBJECT_ID (N'fn_Resumen_GrupoEtario_Ratio', N'IF') IS NOT NULL
  drop function fn_Resumen_GrupoEtario_Ratio
Go

create function fn_Resumen_GrupoEtario_Ratio()
returns Table
as
  return(
    SELECT  [GrupoEtario]
            ,cast(PromedioGrupoEtario as numeric(5,2)) as Promedio
            ,cast(Aprobados*100.0/NroAlumnos as numeric(5,2)) as Aprobado
            ,cast(Desaprobados*100.0/NroAlumnos as numeric(5,2)) as Desaprobado
            ,cast(Reprobados*100.0/NroAlumnos as numeric(5,2)) as Reprobado
            ,cast(Puntaje as numeric(5,2)) as Puntaje
            ,NroAlumnos
    FROM    fn_Resumen_GrupoEtario()
  )
go

```

Ejecutando este segundo script que a su vez invoca a la primera función, se tiene la influencia del grupo etario de los docentes en el rendimiento académico de los estudiantes. Los resultados obtenidos se muestran en la tabla 6.

**Tabla 6**

*Influencia del Grupo etario en el rendimiento de los estudiantes*

Influencia del Grupo Etario de los Docentes en el Rendimiento Académico de los Estudiantes						
Semestres: 2021-1 al 2023-2						
GrupoEtario	Promedio	Aprobado	Desaprobado	Reprobado	Puntaje	NroAlumnos
30-40	12.30	65.80	15.99	18.22	7.92	269
40-50	13.93	80.62	8.89	10.49	7.86	9628
50-60	14.23	83.10	8.37	8.53	8.00	46665
60-70	14.20	82.63	10.16	7.21	7.74	96885
>70	14.53	85.91	8.13	5.96	7.67	72854

Donde, el atributo o columna *Promedio* corresponde a la nota final obtenida por todos los estudiantes en las asignaturas regentadas por los docentes que pertenecen a diferentes

grupos etarios; así mismo, los atributos *Aprobado*, *Desaprobado* y *Reprobado* se expresan en porcentajes según corresponda; finalmente, el atributo *Puntaje* corresponde al promedio de la valoración realizada por los estudiantes a sus respectivos docentes.

**4.1.2.2 Influencia del género de los docentes en el rendimiento académico.** Para determinar en qué medida influye el género de los docentes en el rendimiento académico de los estudiantes, se consideró el atributo Género de cada docente. En este caso, también se implementó dos funciones en SQL, la primera para pre-procesar y resumir los datos, cuyo código se muestra en la figura 11:

**Figura 11**

*Script para Pre-procesar y Resumir los Datos por Género*

```

/* *****
Nombre:          fn_Resumen_Genero
Descripción:     funcion para recuperar el resumen por género
Parametros Entrada: Ninguno
OUTPUT:
Ejemplo:
      Select * from fn_Resumen_Genero()
***** */
IF OBJECT_ID (N'fn_Resumen_Genero', N'IF') IS NOT NULL
drop function fn_Resumen_Genero
Go
create function fn_Resumen_Genero()
returns Table
as
return(
      SELECT  [Genero]
            ,avg(Promedio*1.0) as PromedioGenero
            ,sum(case when Promedio >= 14 then 1 else 0 end) as Aprobados
            ,sum(case when Promedio >= 9 and Promedio < 14 then 1 else 0 end) as Desaprobados
            ,sum(case when Promedio < 9 then 1 else 0 end) as Reprobados
            ,avg([Puntaje]) as Puntaje
            ,count(*) as NroAlumnos
      FROM [BD_Docentes].[dbo].[Info_Academica_]
      group by [Genero]
)
go

```

El segundo *script*, calcula los ratios que determinan la medida en que el género de los docentes influye en el rendimiento de los estudiantes. El script se muestra en la figura 12:

**Figura 12**

*Script para determinar los ratios de la influencia del Género*

```

/* *****
Nombre:          fn_Resumen_Genero_Ratio
Descripción:     funcion para recuperar el ratio por género
Parametros Entrada: Ninguno
OUTPUT:
Ejemplo:
    Select * from fn_Resumen_Genero_Ratio()
***** */
IF OBJECT_ID (N'fn_Resumen_Genero_Ratio', N'IF') IS NOT NULL
    drop function fn_Resumen_Genero_Ratio
Go
create function fn_Resumen_Genero_Ratio()
returns Table
as
return(
    SELECT case when Genero = 0 then 'Femenino' else 'Masculino' end as Genero
    ,cast(PromedioGenero as numeric(5,2)) as Promedio
    ,cast(Aprobados*100.0/NroAlumnos as numeric(5,2)) as Aprobado
    ,cast(Desaprobados*100.0/NroAlumnos as numeric(5,2)) as Desaprobado
    ,cast(Reprobados*100.0/NroAlumnos as numeric(5,2)) as Reprobado
    ,cast(Puntaje as numeric(5,2)) as Puntaje
    ,NroAlumnos
    FROM fn_Resumen_Genero()
)
go

```

Ejecutando los scripts, se tiene la influencia del género de los docentes en el rendimiento académico de los estudiantes. Los resultados obtenidos se muestran en la tabla 7.

**Tabla 7**

*Influencia del Género en el rendimiento de los estudiantes*

Influencia del Género de los Docentes en el Rendimiento Académico de los Estudiantes						
Semestres: 2021-1 al 2023-2						
Género	Promedio	Aprobado	Desaprobado	Reprobado	Puntaje	NroAlumnos
Femenino	14.43	84.38	8.71	6.90	7.89	59335
Masculino	14.25	83.42	9.23	7.35	7.74	166966

Donde, además de Género, los demás atributos tienen la misma semántica que en el caso anterior.

**4.1.2.3 Influencia del género y el grupo etario de los docentes.** Para determinar en qué medida influye de manera conjunta el género y el grupo etario de los docentes en el rendimiento académico de los estudiantes, se consideró estos dos atributos como dimensiones

de resumen; también se implementaron dos scripts en SQL, el primero para el pre-procesamiento y resumen de los datos, cuyo código se muestra en la figura 13:

**Figura 13**

*Script para Pre-procesar y Resumir los Datos por Género y Grupo Etario*

```

/* *****
Nombre:          fn_Resumen_Genero_GrupoEtario
Descripción:     función para recuperar el resumen por Género y GrupoEtario
Parametros Entrada: Ninguno
OUTPUT:
Ejemplo:
    Select * from fn_Resumen_Genero_GrupoEtario()
***** */
IF OBJECT_ID (N'fn_Resumen_Genero_GrupoEtario', N'IF') IS NOT NULL
drop function fn_Resumen_Genero_GrupoEtario
Go
create function fn_Resumen_Genero_GrupoEtario()
returns Table
as
return(
    WITH InfoGeneroGrupoEtario(Genero, GrupoEtario, Promedio, Puntaje)
    as
    (
        SELECT Genero,
            case
                when Edad < 30 then '<30'
                when Edad >= 30 and Edad < 40 then '30-40'
                when Edad >= 40 and Edad < 50 then '40-50'
                when Edad >= 50 and Edad < 60 then '50-60'
                when Edad >= 60 and Edad < 70 then '60-70'
                when Edad >= 70 then '>70'
            end as GrupoEtario, Promedio, Puntaje
        FROM [BD_Docentes].[dbo].[Info_Academica_]
    )
    SELECT Genero
        ,GrupoEtario
        ,avg(Promedio*1.0) as PromedioGrupoEtario
        ,sum(case when Promedio >= 14 then 1 else 0 end) as Aprobados
        ,sum(case when Promedio >= 9 and Promedio < 14 then 1 else 0 end) as Desaprobados
        ,sum(case when Promedio < 9 then 1 else 0 end) as Reprobados
        ,avg([Puntaje]) as Puntaje
        ,count(*) as NroAlumnos
    FROM InfoGeneroGrupoEtario
    group by Genero, GrupoEtario
)
go

```

El segundo *script*, calcula los ratios que determinan la medida en que el género y el grupo etario de los docentes influyen en el rendimiento de los estudiantes. El script se muestra en la figura 14:

**Figura 14**

*Script para determinar los ratios de la influencia del Género y Grupo Etario*

```

/* *****
Nombre:          fn_Resumen_Genero_GrupoEtario_Ratio
Descripción:     funcion para recuperar el ratio por Genero y GrupoEtario
Parametros Entrada: Ninguno
OUTPUT:
Ejemplo:
  Select * from fn_Resumen_Genero_GrupoEtario_Ratio() order by Puntaje desc
***** */

IF OBJECT_ID ('fn_Resumen_Genero_GrupoEtario_Ratio', N'IF') IS NOT NULL
drop function fn_Resumen_Genero_GrupoEtario_Ratio
Go

create function fn_Resumen_Genero_GrupoEtario_Ratio()
returns Table
as
return(
    SELECT Genero
    ,GrupoEtario
    ,cast(PromedioGrupoEtario as numeric(5,2)) as Promedio
    ,cast(Aprobados*100.0/NroAlumnos as numeric(5,2)) as Aprobado
    ,cast(Desaprobados*100.0/NroAlumnos as numeric(5,2)) as Desaprobado
    ,cast(Reprobados*100.0/NroAlumnos as numeric(5,2)) as Reprobado
    ,cast(Puntaje as numeric(5,2)) as Puntaje
    ,NroAlumnos
    FROM fn_Resumen_Genero_GrupoEtario()
)

go

```

Ejecutando los scripts, se tiene la influencia conjunta del género y grupo etario de los docentes en el rendimiento académico de los estudiantes. Los resultados obtenidos se muestran en la tabla 8.

**Tabla 8**

*Influencia del Género y Grupo Etario en el Rendimiento de los Estudiantes*

Influencia del Género y el Grupo Etario de los Docentes en el Rendimiento Académico de los Estudiantes						
Semestres: 2021-1 al 2023-2						
GrupoEtario	Promedio	Aprobado	Desaprobado	Reprobado	Puntaje	NroAlumnos
FEMENINO						
40-50	14.29	83.73	8.34	7.93	7.81	3657.00
50-60	14.17	83.18	8.65	8.16	7.98	15032.00
60-70	14.51	83.96	9.34	6.70	7.92	28498.00
>70	14.62	87.04	7.43	5.52	7.74	12148.00
MASCULINO						
30-40	12.30	65.80	15.99	18.22	7.92	269.00
40-50	13.70	78.71	9.23	12.06	7.90	5971.00
50-60	14.26	83.06	8.24	8.70	8.00	31633.00
60-70	14.08	82.07	10.50	7.43	7.66	68387.00
>70	14.51	85.68	8.27	6.04	7.66	60706.00

### 4.1.3 *Influencia de la categoría y el régimen de los docentes*

Para determinar en qué medida influyen la categoría y el régimen de los docentes en el rendimiento, se consideró los atributos:

Categoría.- Se consideran tres categorías: PRINCIPAL, ASOCIADO y AUXILIAR.

Régimen.- Tiempo Parcial de 10 y 20 horas (TP10H y TP20H), Tiempo completo (TC) y Dedicación exclusiva (DE).

Promedio.- Nota final obtenida por un estudiante en una asignatura regentada por un docente.

Puntaje.- Valoración dada por un estudiante al desempeño del docente en una asignatura.

Luego, se efectuó un procesamiento determinístico mediante sentencias SQL.

**4.1.3.1 Influencia de la categoría de los docentes.** Para este caso, se implementó dos funciones en SQL, la primera para pre-procesar y resumir los datos por categoría; el respectivo script se muestra en la figura 15:

#### **Figura 15**

##### *Script para Pre-procesar y Resumir los Datos por Categoría*

```

/* *****
Nombre:          fn_Resumen_Categoria
Descripción:     funcion para recuperar el resumen por categoría
Parametros Entrada: Ninguno
OUTPUT:
Ejemplo:
    Select * from fn_Resumen_Categoria()
***** */
IF OBJECT_ID (N'fn_Resumen_Categoria', N'IF') IS NOT NULL
drop function fn_Resumen_Categoria
Go
create function fn_Resumen_Categoria()
returns Table
as
return(
    SELECT [Categoria]
           ,avg(Promedio*1.0) as PromedioCategoria
           ,sum(case when Promedio >= 14 then 1 else 0 end) as Aprobados
           ,sum(case when Promedio >= 9 and Promedio < 14 then 1 else 0 end) as Desaprobados
           ,sum(case when Promedio < 9 then 1 else 0 end) as Reprobados
           ,avg([Puntaje]) as Puntaje
           ,count(*) as NroAlumnos
    FROM [BD_Docentes].[dbo].[Info_Academica_]
    group by [Categoria]
)
go

```

El segundo script calcula los ratios que determinan la medida en que el régimen de los docentes influye en el rendimiento de los estudiantes. El script se muestra en la figura 16:

**Figura 16**

*Script para determinar los ratios de la influencia del Régimen*

```

/* *****
Nombre:          fn_Resumen_Categoria_Ratio
Descripción:     funcion para recuperar el ratio por categoría
Parametros Entrada: Ninguno
OUTPUT:
Ejemplo:
    Select * from fn_Resumen_Categoria_Ratio()
***** */
IF OBJECT_ID (N'fn_Resumen_Categoria_Ratio', N'IF') IS NOT NULL
    drop function fn_Resumen_Categoria_Ratio
Go
create function fn_Resumen_Categoria_Ratio()
returns Table
as
    return(
        SELECT  [Categoria]
                ,cast(PromedioCategoria as numeric(5,2)) as Promedio
                ,cast(Aprobados*100.0/NroAlumnos as numeric(5,2)) as Aprobado
                ,cast(Desaprobados*100.0/NroAlumnos as numeric(5,2)) as Desaprobado
                ,cast(Reprobados*100.0/NroAlumnos as numeric(5,2)) as Reprobado
                ,cast(Puntaje as numeric(5,2)) as Puntaje
                ,NroAlumnos
        FROM    fn_Resumen_Categoria()
    )
go

```

Ejecutando este segundo script, se tiene la influencia de la categoría de los docentes en el rendimiento académico de los estudiantes. Los resultados se muestran en la tabla 9.

**Tabla 9**

*Influencia de la Categoría en el rendimiento de los estudiantes*

Influencia de la Categoría de los Docentes en el Rendimiento Académico de los Estudiantes						
Semestres: 2021-1 al 2023-2						
Categoría	Promedio	Aprobado	Desaprobado	Reprobado	Puntaje	NroAlumnos
AUXILIAR	13.93	81.05	9.66	9.29	7.65	16218
ASOCIADO	14.23	83.22	9.39	7.39	7.77	68849
PRINCIPAL	14.38	84.20	8.88	6.92	7.80	141234

Donde, el atributo o columna *Promedio* corresponde a la nota final obtenida por todos los estudiantes en las asignaturas regentadas por los docentes con diferentes categorías; así mismo, los atributos *Aprobado*, *Desaprobado* y *Reprobado* se expresan en porcentajes según

corresponda; finalmente, el atributo *Puntaje* corresponde al promedio de la valoración realizada por los estudiantes a sus docentes por categorías.

**4.1.3.2 Influencia del régimen de los docentes.** Para determinar en qué medida influye el régimen de los docentes en el rendimiento académico de los estudiantes, se consideró el atributo Régimen de cada docente. En este caso, también se implementó dos funciones en SQL, la primera para pre-procesar y resumir los datos, cuyo código se muestra en la figura 17:

**Figura 17**

*Script para Pre-procesar y Resumir los Datos por Régimen*

```

/* *****
Nombre:          fn_Resumen_Regimen
Descripción:     funcion para recuperar el resumen por Régimen
Parametros Entrada: Ninguno
OUTPUT:
Ejemplo:
    Select * from fn_Resumen_Regimen()
***** */
IF OBJECT_ID (N'fn_Resumen_Regimen', N'IF') IS NOT NULL
    drop function fn_Resumen_Regimen
Go
create function fn_Resumen_Regimen()
returns Table
as
    return(
        SELECT  [Regimen]
                ,avg(Promedio*1.0) as PromedioRegimen
                ,sum(case when Promedio >= 14 then 1 else 0 end) as Aprobados
                ,sum(case when Promedio >= 9 and Promedio < 14 then 1 else 0 end) as Desaprobados
                ,sum(case when Promedio < 9 then 1 else 0 end) as Reprobados
                ,avg([Puntaje]) as Puntaje
                ,count(*) as NroAlumnos
        FROM [BD_Docentes].[dbo].[Info_Academica_]
        group by [Regimen]
    )
go

```

El segundo script, calcula los ratios que determinan la medida en que el régimen de los docentes influye en el rendimiento de los estudiantes. El script se muestra en la figura 18:

**Figura 18***Script para determinar los ratios de la influencia del Régimen*

```

/* *****
Nombre:          fn_Resumen_Regimen_Ratio
Descripción:     funcion para recuperar el ratio por Régimen
Parametros Entrada: Ninguno
OUTPUT:
Ejemplo:
      Select * from fn_Resumen_Regimen_Ratio()
***** */
IF OBJECT_ID (N'fn_Resumen_Regimen_Ratio', N'IF') IS NOT NULL
drop function fn_Resumen_Regimen_Ratio
Go
create function fn_Resumen_Regimen_Ratio()
returns Table
as
return(
      SELECT  [Regimen]
            , cast(PromedioRegimen as numeric(5,2)) as Promedio
            , cast(Aprobados*100.0/NroAlumnos as numeric(5,2)) as Aprobado
            , cast(Desaprobados*100.0/NroAlumnos as numeric(5,2)) as Desaprobado
            , cast(Reprobados*100.0/NroAlumnos as numeric(5,2)) as Reprobado
            , cast(Puntaje as numeric(5,2)) as Puntaje
            , NroAlumnos
      FROM    fn_Resumen_Regimen()
)
go

```

Ejecutando los scripts, se tiene la influencia del régimen de los docentes en el rendimiento académico de los estudiantes. Los resultados se muestran en la tabla 10.

**Tabla 10***Influencia del Régimen en el rendimiento de los estudiantes*

Influencia del Régimen de los Docentes en el Rendimiento Académico de los Estudiantes						
Semestres: 2021-1 al 2023-2						
Régimen	Promedio	Aprobado	Desaprobado	Reprobado	Puntaje	NroAlumnos
TP 10 H	14.79	86.14	9.94	3.92	7.65	5049
TP 20 H	14.87	86.28	9.26	4.46	7.79	10578
TC	14.33	84.01	8.20	7.79	7.90	79522
DE	14.22	83.17	9.59	7.25	7.71	131152

Donde, el valor de régimen de la base de datos corresponde a DE=Dedicación exclusiva, TC= Tiempo completo y TP a tiempo parcial de 10 y 20 horas.

**4.1.3.3 Influencia de la categoría y el régimen de los docentes.** Para determinar en qué medida influye de manera conjunta la categoría y el régimen de los docentes en el rendimiento académico de los estudiantes, se consideró estos dos atributos como dimensiones

de resumen; también se implementaron dos scripts en SQL, el primero para el pre-procesamiento y resumen de los datos, cuyo código se muestra en la figura 19:

**Figura 19**

*Script para Pre-procesar y Resumir los Datos por Categoría y Régimen*

```

/* *****
Nombre:          fn_Resumen_Categoria_Regimen
Descripción:     funcion para recuperar el resumen por Categoría y Régimen
Parametros Entrada: Ninguno
OUTPUT:
Ejemplo:
    Select * from fn_Resumen_Categoria_Regimen()
***** */
IF OBJECT_ID (N'fn_Resumen_Categoria_Regimen', N'IF') IS NOT NULL
drop function fn_Resumen_Categoria_Regimen
Go
create function fn_Resumen_Categoria_Regimen()
returns Table
as
return(
    SELECT  Categoria
           ,Regimen
           ,avg(Promedio*1.0) as PromedioRegimen
           ,sum(case when Promedio >= 14 then 1 else 0 end) as Aprobados
           ,sum(case when Promedio >= 9 and Promedio < 14 then 1 else 0 end) as Desaprobados
           ,sum(case when Promedio < 9 then 1 else 0 end) as Reprobados
           ,avg([Puntaje]) as Puntaje
           ,count(*) as NroAlumnos
    FROM [BD_Docentes].[dbo].[Info_Academica_]
    group by Categoria, Regimen
)
go

-- Select * from fn_Resumen_Categoria_Regimen()

```

El segundo script, se determina la medida en que la categoría y el régimen de los docentes influyen en el rendimiento de los estudiantes. El script se muestra en la figura 20:

**Figura 20**

*Script para determinar los ratios de la influencia de la Categoría y Régimen*

```

/* *****
Nombre:          fn_Resumen_Categoria_Regimen_Ratio
Descripción:     funcion para recuperar el ratio por Categoría y Régimen
Parametros Entrada: Ninguno
OUTPUT:
Ejemplo:
    Select * from fn_Resumen_Categoria_Regimen_Ratio()
***** */
IF OBJECT_ID (N'fn_Resumen_Categoria_Regimen_Ratio', N'IF') IS NOT NULL
drop function fn_Resumen_Categoria_Regimen_Ratio
Go
create function fn_Resumen_Categoria_Regimen_Ratio()
returns Table
as
return(
    SELECT  Categoria
           ,Regimen
           ,cast(PromedioRegimen as numeric(5,2)) as Promedio
           ,cast(Aprobados*100.0/NroAlumnos as numeric(5,2)) as Aprobado
           ,cast(Desaprobados*100.0/NroAlumnos as numeric(5,2)) as Desaprobado
           ,cast(Reprobados*100.0/NroAlumnos as numeric(5,2)) as Reprobado
           ,cast(Puntaje as numeric(5,2)) as Puntaje
           ,NroAlumnos
    FROM fn_Resumen_Categoria_Regimen()
)
go

```

Ejecutando los scripts, se tiene la influencia conjunta de la categoría y el régimen de los docentes en el rendimiento académico de los estudiantes. Los resultados obtenidos se muestran en la tabla 11.

**Tabla 11**

*Influencia de la Categoría y Régimen en el Rendimiento de los Estudiantes*

Influencia de la Categoría y el Régimen de los Docentes en el Rendimiento Académico de los Estudiantes						
Semestres: 2021-1 al 2023-2						
Regimen	Promedio	Aprobado	Desaprobado	Reprobado	Puntaje	NroAlumnos
<b>AUXILIAR</b>						
TC	13.83	80.29	9.62	10.09	7.65	12527.00
TP 10 H	14.91	83.33	14.12	2.54	8.14	1062.00
TP 20 H	13.99	83.76	8.06	8.18	7.44	2629.00
<b>ASOCIADO</b>						
DE	13.96	81.54	11.00	7.46	7.56	24924.00
TC	14.20	83.01	8.84	8.14	7.89	36791.00
TP 10 H	14.70	84.73	10.32	4.95	7.75	1938.00
TP 20 H	15.52	92.22	5.18	2.60	7.89	5196.00
<b>PRINCIPAL</b>						
DE	14.28	83.55	9.25	7.20	7.74	106228.00
TC	14.70	86.77	6.83	6.40	8.00	30204.00
TP 10 H	14.81	88.92	7.42	3.66	7.29	2049.00
TP 20 H	14.50	77.48	18.09	4.43	7.94	2753.00

#### 4.2 Procesamiento estocástico

En esta sección se determina la relevancia de las características distintivas de los docentes desde la perspectiva de los estudiantes; así como, la relevancia de las preguntas de las encuestas aplicadas a los estudiantes sobre el desempeño académico de sus respectivos docentes.

En esta sección se analizan los datos en base a las encuestas virtuales aplicadas por la universidad a todos los estudiantes al finalizar cada semestre. En la figura 21 se tiene un ejemplo con el formato de la encuesta.

Figura 21

## Encuesta de valoración del desempeño de los docentes



**UNIVERSIDAD NACIONAL DE SAN ANTONIO ABAD DEL CUSCO**  
 EVALUACION DEL ALUMNO SOBRE EL DESEMPEÑO DOCENTE  
 SEMESTRE 2023-2  
 (COPIA NO VALIDA PARA TRAMITES)

ESCUELA PROFESIONAL: INGENIERÍA INFORMÁTICA Y DE SISTEMAS

Puntaje Maximo:10 Puntaje Minimo:1

ASIGNATURA : IF450AIN ABSTRACCIÓN DE DATOS Y OBJETOS

DOCENTE : ACURIO-USCA-NILA ZONIA

Alumnos Matriculados : 43

Alumnos Encuestados : 41

Promedio de Curso : 8.99

CONSULTA	PROMEDIO
1. El silabo es sustentado por el profesor y contiene la informacion pertinente (Objetivos, dosificacion de temas, programacion de actividades, bibliografias, etc.) para el desarrollo de la asignatura	9.02
2. El profesor comienza y termina las clases a la hora fijada en el horario y lugar	8.95
3. El profesor explica los motivos de su inasistencia	9.05
4. El profesor muestra dominio de la asignatura	8.98
5. El profesor inicia cada tema exponiendo los objetivos del Mismo	9.00
6. Existe una comunicacion fluida entre el profesor y los estudiantes	9.05
7. El profesor destaca los aspectos fundamentales de cada tema	9.07
8. El profesor realiza retroalimentacion cuando los alumnos tienen dificultad en el aprendizaje	8.95
9. El profesor motiva a los alumnos para que se interesen por el tema	9.05
10. El profesor cumple con los horarios de tutoria establecidos	8.93
11. El profesor propone actividades para favorecer la investigacion (Busqueda de informacion complementaria, trabajos, articulos, etc)	9.05
12. El profesor utiliza recursos didacticos (Medios audiovisuales, maquetas y en general TICS) que ayuden a comprender los contenidos	8.93
13. El profesor utiliza la metodologia de ensenaza adecuada a las caracteristicas del grupo y de la asignatura	9.00
14. El profesor utiliza un lenguaje claro durante el proceso de ensenaza y aprendizaje	9.00
15. El profesor informa sobre el sistema de evaluacion al inicio del curso	9.10
16. El profesor utiliza diferentes procedimientos (exámenes, trabajos, practicas, intervenciones, etc) para evaluar el aprendizaje de los alumnos	8.98
17. El profesor toma en consideracion la propuesta de los alumnos sobre el desarrollo de la asignatura	8.71
18. El profesor tiene un trato igualitario con todos los alumnos en sesiones de clase, calificacion de evaluacion y/o trabajos	8.88
19. El profesor responde con respeto e interes las intervenciones de los alumnos	8.90
20. En forma integral, valore globalmente al profesor de la asignatura	9.12

Estos datos corresponden a valoraciones subjetivas por parte de los estudiantes respecto al desempeño académico de sus docentes; en consecuencia, cualquier análisis a partir de estos datos implica relaciones bastante complejas, por tal razón, se recurren a técnicas estocásticas como el aprendizaje automático.

En esta sección, se utiliza el modelo *Random Forest* porque en las pruebas previas es la que mejor rendimiento ofreció. Se toma las siguientes consideraciones en el detalle de la implementación:

- Corresponde a un modelo supervisado, por tanto, los atributos considerados son: *Categoría, Régimen, Experiencia, Grado Académico, Edad y Género*. Para la variable

objetivo o etiqueta se utiliza el atributo *Puntaje final*, que corresponde a la valoración que efectúan los estudiantes del desempeño académico de sus respectivos docentes.

- Los datos se dividen en dos sub conjuntos, 80% para la fase de entrenamiento y un 20% para la fase de evaluación.
- Se utiliza como hiperparámetros los siguientes: *n\_estimators=1000* y *random\_state=333*.
- Como se trata de determinar qué características son las más relevantes, se hace uso específicamente del método *feature\_importances\_*.

Para la implementación se utiliza el entorno de desarrollo de *Google Colaboratory* con el lenguaje de programación Python. A continuación se desarrolla las fases más importantes de esta implementación:

El lenguaje de programación Python ofrece una amplia gama de librerías para aprendizaje automático y ciencia de datos; en la figura 22 se muestra el segmento de código que ilustra las librerías utilizadas en la parte del procesamiento estocástico.

## Figura 22

### *Librerías utilizadas*

```

1 from sklearn.ensemble import RandomForestClassifier
2 from sklearn.model_selection import train_test_split
3 from sklearn.preprocessing import LabelEncoder
4 from sklearn.metrics import accuracy_score
5 import pandas as pd
6 import numpy as np
7 import matplotlib.pyplot as plt

```

El dataset utilizado corresponde a un archivo *Datos.csv*, generado a partir de los datos de las bases de datos transaccionales proporcionado por la universidad. En cuanto al pre procesamiento de los datos, sólo se efectúa la codificación de los valores categóricos y la discretización de algunos atributos. Se toma en cuenta que los modelos de clasificación de aprendizaje automático trabajan sólo sobre información numérica.

El segmento de código relacionado a la carga de datos y pre procesamiento de los mismos, se ilustra en la figura 23.

### Figura 23

#### *Carga de datos y pre procesamiento*

```

1 # Leer los datos
2 df = pd.read_csv('Datos.csv', encoding='latin1')
3 # Codificar todas las variables categóricas
4 # Considerando que los clasificadores solo entienden datos numéricos
5 categorical_feature_mask = df.dtypes==object
6 categorical_cols = df.columns[categorical_feature_mask].tolist()
7 le = LabelEncoder()
8 df[categorical_cols] = df[categorical_cols].apply(lambda col:le.fit_transform(col))
9 df.head()

```

#### 4.2.1 *Relevancia de las características distintivas de los docentes*

Determinar qué características de los docentes, son realmente relevantes para los estudiantes, implica implementar el modelo *Random Forest* detallado en el acápite anterior. El segmento de código correspondiente, se ilustra en la figura 24.

### Figura 24

#### *Procesamiento para determinar la relevancia de las características distintivas*

```

1 # Recuperar del dataset original sólo los atributos de características distintivas de los docentes
2 df['PuntajeFinal'] = df['Puntaje'].apply(transformar_puntaje_2)
3 df['Promedio'] = df['promedio']
4 df_final = df[['Categoria', 'Regimen', 'Experiencia', 'GradoAcademico', 'Edad', 'Genero', 'PuntajeFinal']]
5 datos = df_final
6 # Dividir los datos en características (X) y la variable objetivo (y)
7 X = datos.drop('PuntajeFinal', axis=1)
8 y = datos['PuntajeFinal']
9 # Dividir los datos en conjunto de entrenamiento y conjunto de prueba
10 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=30)
11 # Entrenar el modelo de Random Forest
12 rf_model = RandomForestClassifier(n_estimators=1000, random_state=333)
13 rf_model.fit(X_train, y_train)
14 # Calcular la precisión del modelo en el conjunto de prueba
15 y_pred = rf_model.predict(X_test)
16 precision = accuracy_score(y_test, y_pred)
17 print("Precisión del modelo:", precision, "\n")
18 ## Determinar el nivel de importancia de las características
19 importancias = rf_model.feature_importances_
20 caracteristicas = X.columns
21 # Crear un DataFrame para mostrar las importancias de las características
22 importancia_df = pd.DataFrame({'Característica': caracteristicas, 'Importancia': importancias})
23 importancia_df = importancia_df.sort_values(by='Importancia', ascending=False)
24 print(importancia_df)
25 # Graficar
26 plt.figure(figsize=(10, 6))
27 plt.barh(importancia_df['Característica'], importancia_df['Importancia'], color='skyblue')
28 plt.xlabel('Importancia')
29 plt.ylabel('Característica')
30 plt.title('Importancia de las características')
31 plt.gca().invert_yaxis()
32 plt.show()

```

Luego de ejecutar este código, se obtienen los resultados mostrados en la tabla 12 y la figura 25.

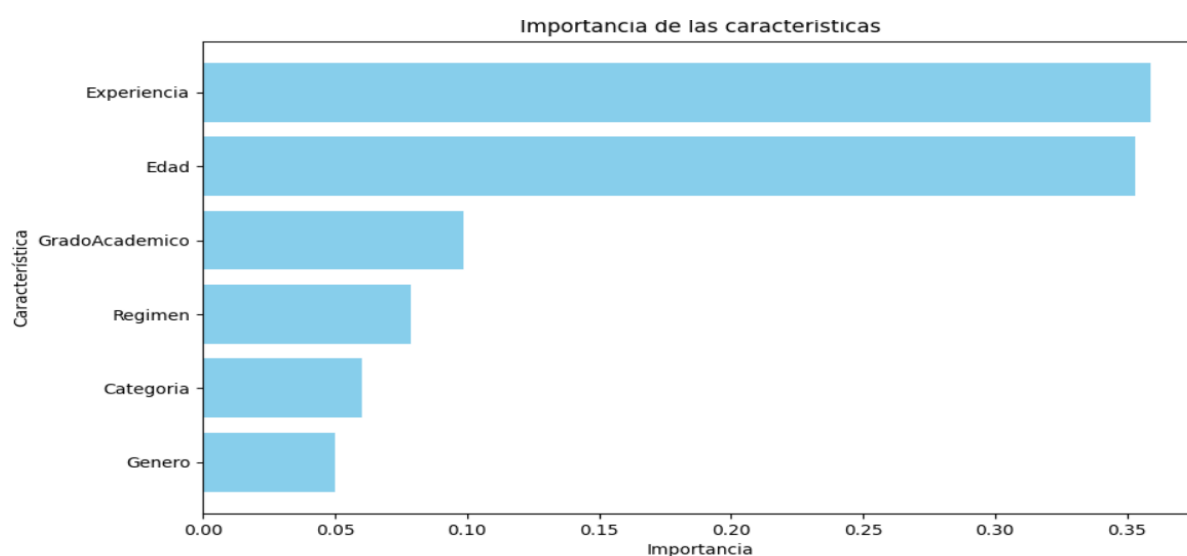
**Tabla 12**

*Relevancia de las características distintivas de los docentes*

Relevancia de las características distintivas de los docentes		
Nro	Característica	Importancia
1	Experiencia	0.359034
2	Edad	0.352979
3	Grado Académico	0.098760
4	Regimen	0.078904
5	Categoría	0.060361
6	Género	0.049961

**Figura 25**

*Relevancia de las características distintivas de los docentes*



La precisión del modelo es: 0.724619

#### 4.2.2 Relevancia de las preguntas de las encuestas

Determinar la relevancia de cada una de las 20 preguntas de las encuestas implica también relaciones complejas. Por tanto, en esta sección también se utiliza procesamiento

estocástico. De manera similar al anterior acápite, se utiliza el modelo *Random Forest* de aprendizaje automático, detallado al inicio de esta sección.

El segmento de código correspondiente al procesamiento para determinar la relevancia de las preguntas de la encuesta, se ilustra en la figura 26.

## Figura 26

### *Procesamiento para determinar la relevancia de las preguntas de la encuesta*

```

1 # Recuperar del dataset original sólo los atributos relacionados a las encuestas
2 df['PuntajeFinal'] = df['Puntaje'].apply(transformar_puntaje)
3 df_final = df[['P01', 'P02', 'P03', 'P04', 'P05', 'P06', 'P07', 'P08', 'P09', 'P10',
4               'P11', 'P12', 'P13', 'P14', 'P15', 'P16', 'P17', 'P18', 'P19', 'P20', 'PuntajeFinal']]
5 datos = df_final
6 # Dividir los datos en características (X) y la variable objetivo (y)
7 X = datos.drop('PuntajeFinal', axis=1)
8 y = datos['PuntajeFinal']
9 # Dividir los datos en conjunto de entrenamiento y conjunto de prueba
10 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=30)
11 # Entrenar el modelo RF
12 rf_model = RandomForestClassifier(n_estimators=100, random_state=30)
13 rf_model.fit(X_train, y_train)
14 # Calcular la precisión del modelo
15 y_pred = rf_model.predict(X_test)
16 precision = accuracy_score(y_test, y_pred)
17 print("Precisión del modelo:", precision)
18 # Determinar el nivel de importancia de los atributos
19 importancias = rf_model.feature_importances_
20 caracteristicas = X.columns
21 # Crear un DataFrame para mostrar las importancias de las características
22 importancia_df = pd.DataFrame({'Característica': caracteristicas, 'Importancia': importancias})
23 importancia_df = importancia_df.sort_values(by='Importancia', ascending=False)
24 print(importancia_df)
25 # Graficar
26 plt.figure(figsize=(10, 6))
27 plt.barh(importancia_df['Característica'], importancia_df['Importancia'], color='skyblue')
28 plt.xlabel('Importancia')
29 plt.ylabel('Característica')
30 plt.title('Importancia de las características')
31 plt.gca().invert_yaxis()
32 plt.show()

```

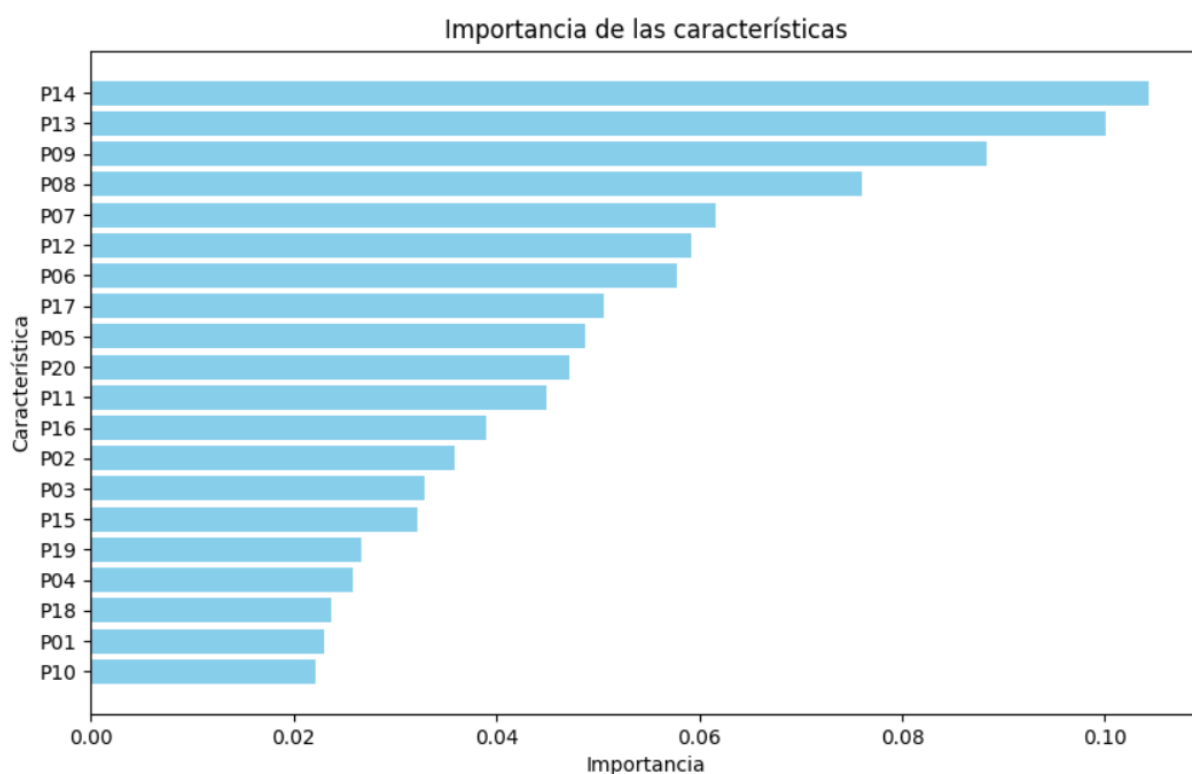
Luego de ejecutar este código, se obtienen los resultados mostrados en la tabla 13 y la figura 27.

**Tabla 13***Relevancia de las preguntas de las encuestas de desempeño docente*

Pertinencia de las preguntas de las encuestas de desempeño docente		
Nro. Pregunta	Pregunta	Importancia
P14	El profesor utiliza un lenguaje claro durante el proceso de enseñanza y aprendizaje	0.104397
P13	El profesor utiliza la metodología de enseñanza adecuada a las características del grupo y de la asignatura	0.100048
P09	El profesor motiva a los alumnos para que se interesen por el tema	0.088365
P08	El profesor realiza retroalimentación cuando los alumnos tienen dificultad en el aprendizaje	0.076001
P07	El profesor destaca los aspectos fundamentales de cada tema	0.061652
P12	El profesor utiliza recursos didácticos (Medios audiovisuales, maquetas y en general TICS) que ayuden a comprender los contenidos	0.059174
P06	Existe una comunicación fluida entre el profesor y los estudiantes	0.057855
P17	El profesor toma en consideración la propuesta de los alumnos sobre el desarrollo de la asignatura	0.050523
P05	El profesor inicia cada tema exponiendo los objetivos del Mismo	0.048691
P20	En forma integral, valore globalmente al profesor de la asignatura	0.047164
P11	El profesor propone actividades para favorecer la investigación (Búsqueda de información complementaria, trabajos, artículos, etc)	0.044978
P16	El profesor utiliza diferentes procedimientos (exámenes, trabajos, prácticas, intervenciones, etc) para evaluar el aprendizaje de los alumnos	0.038954
P02	El profesor comienza y termina las clases a la hora fijada en el horario y lugar	0.035937
P03	El profesor explica los motivos de su inasistencia	0.032943
P15	El profesor informa sobre el sistema de evaluación al inicio del curso	0.032133
P19	El profesor responde con respeto e interés las intervenciones de los alumnos	0.026623
P04	El profesor muestra dominio de la asignatura	0.025765
P18	El profesor tiene un trato igualitario con todos los alumnos en sesiones de clase, calificación de evaluación y/o trabajos	0.023635
P01	El silabo es sustentado por el profesor y contiene la información pertinente (Objetivos, dosificación de temas, programación de actividades, bibliografías, etc.) para el desarrollo de la asignatura	0.023009
P10	El profesor cumple con los horarios de tutoría establecidos	0.022155

**Figura 27**

*Relevancia de las preguntas de las encuestas de desempeño docente*



La Precisión del modelo es: 0.93890

### **4.3 Validación de la hipótesis**

#### **4.3.1 Validación de la hipótesis general**

La hipótesis general “*Si se utilizan técnicas de ciencia de datos, se determinan las implicancias que existen entre las características distintivas de los docentes (grado académico, grupo etario, categoría y régimen, género, experiencia, así como la percepción o valoración que tiene los estudiantes de los docentes) en el rendimiento académico de los estudiantes en las universidades públicas de la región Cusco*” se valida con los resultados obtenidos y mostrados en las respectivas tablas del acápite anterior, específicamente en las tablas 5, 8, 11, 12 y 13. Como se puede apreciar en las referidas tablas, se logra determinar las implicancias de las características distintivas de los docentes en el rendimiento académico de los estudiantes, que permitirán concluir que características distintivas de los docentes influyen en un mejor

rendimiento académico de los estudiantes, y qué características distintivas de los docentes deben ser fortalecidas con el planteamiento de estrategias de mejora continua. Se corrobora la hipótesis general, con la validación de las diferentes hipótesis específicas, que se detallan a continuación.

#### 4.3.2 Validación de las hipótesis específicas

La primera hipótesis específica, planteada como “*Si se analizan los datos referidos al grado académico y la experiencia de los docentes se determina la medida en que influyen en el rendimiento académico de los estudiantes en las universidades públicas de la región del Cusco*”, se valida con los resultados de la sección 4.1.1. Como se puede apreciar, analizando las características distintivas grado académico y experiencia de los docentes respecto al promedio final obtenido por cada estudiante, se logra determinar la medida en que estas características influyen en el rendimiento académico de los estudiantes. Se resume este análisis considerando, el promedio de las notas finales de los estudiantes y el porcentaje de alumnos aprobados. En consecuencia, la información mostrada en la tabla 14, valida suficientemente esta hipótesis específica.

**Tabla 14**

*Influencia del Grado Académico y la experiencia de los docentes*

Influencia del Grado Académico y la Experiencia de los Docentes en el Rendimiento Académico de los Estudiantes			
Semestres: 2021-1 al 2023-2			
Experiencia	Promedio	Aprobado	NroAlumnos
DOCTORADO			
05-10	14.20	83.91	230
10-15	14.47	85.50	12263
15-20	14.66	86.21	23082
20-25	14.60	85.68	19802
>25	14.48	85.40	46633
MAESTRÍA			
05-10	12.42	66.12	1036.00
10-15	14.40	84.32	10862.00

15-20	14.34	83.56	14560.00
20-25	13.86	78.40	12792.00
>25	14.06	82.15	41019.00
TÍTULO			
05-10	12.10	65.87	378.00
10-15	12.47	68.39	3268.00
15-20	14.36	83.02	7573.00
20-25	14.12	82.53	9208.00
>25	14.28	84.26	23595.00

La segunda hipótesis específica, planteada como “*Si se analizan los datos referidos al grupo etario y el género de los docentes se determina la medida en que influyen en el rendimiento académico de los estudiantes en las universidades públicas de la región del Cusco*”, se valida con los resultados de la sección 4.1.2. Como se puede apreciar, analizando las características distintivas grupo etario y género de los docentes respecto al promedio final obtenido por cada estudiante, se logra determinar la medida en que estas características influyen en el rendimiento académico de los estudiantes. Se resume este análisis considerando para cada par de características distintivas, el promedio de las notas finales de los estudiantes y el porcentaje de alumnos aprobados. En consecuencia, la información mostrada en la tabla 15, valida suficientemente esta hipótesis específica.

**Tabla 15**

*Influencia del Grupo Etario y Género*

Influencia del Género y el Grupo Etario de los Docentes en el Rendimiento Académico de los Estudiantes			
Semestres: 2021-1 al 2023-2			
GrupoEtario	Promedio	Aprobado	NroAlumnos
FEMENINO			
40-50	14.29	83.73	3657.00
50-60	14.17	83.18	15032.00
60-70	14.51	83.96	28498.00
>70	14.62	87.04	12148.00
MASCULINO			

30-40	12.30	65.80	269.00
40-50	13.70	78.71	5971.00
50-60	14.26	83.06	31633.00
60-70	14.08	82.07	68387.00
>70	14.51	85.68	60706.00

La tercera hipótesis específica, planteada como “*Si se analizan los datos referidos a la categoría y régimen al que pertenecen los docentes se determina la medida en que influyen en el rendimiento académico de los estudiantes en las universidades públicas de la región del Cusco*”, se valida con los resultados de la sección 4.1.3. Como se puede apreciar, analizando las características distintivas categoría y régimen de los docentes respecto al promedio final obtenido por cada estudiante, se logra determinar la medida en que estas características influyen en el rendimiento académico de los estudiantes. Se resume este análisis considerando para cada par de características distintivas, el promedio de las notas finales de los estudiantes y el porcentaje de alumnos aprobados. En consecuencia, la información mostrada en la tabla 16, valida suficientemente esta hipótesis específica.

**Tabla 16**

*Influencia de la Categoría y Régimen*

Influencia de la Categoría y el Régimen de los Docentes en el Rendimiento Académico de los Estudiantes			
Semestres: 2021-1 al 2023-2			
Regimen	Promedio	Aprobado	NroAlumnos
AUXILIAR			
TC	13.83	80.29	12527.00
TP 10 H	14.91	83.33	1062.00
TP 20 H	13.99	83.76	2629.00
ASOCIADO			
DE	13.96	81.54	24924.00
TC	14.20	83.01	36791.00
TP 10 H	14.70	84.73	1938.00
TP 20 H	15.52	92.22	5196.00
PRINCIPAL			
DE	14.28	83.55	106228.00

TC	14.70	86.77	30204.00
TP 10 H	14.81	88.92	2049.00
TP 20 H	14.50	77.48	2753.00

La cuarta hipótesis específica, planteada como *“Si se analizan las características distintivas de los docentes en la valoración o percepción que tienen los estudiantes de los docentes, así como la pertinencia de las encuestas aplicadas a los estudiantes para evaluar el desempeño de los docentes, en las universidades públicas de la región del Cusco, se determinan en qué medida tiene relevancia estas características y la pertinencia de las encuestas”*, se valida con los resultados obtenidos y mostrados en las tablas y gráficos de la sección 4.2.1 y 4.2.2.

Efectivamente, desde la perspectiva de los estudiantes, se tiene distintos niveles de relevancia en las características distintivas de los docentes. Así mismo, se determinan el nivel de relevancia de las preguntas de las encuestas.

## V. DISCUSIÓN DE RESULTADOS

Los resultados alcanzados en el presente trabajo muestran la importancia del análisis de datos, ya sea utilizando procesamientos determinísticos o estocásticos; más aún, en el contexto académico de las universidades, donde, se tiene una tendencia creciente en el volumen de datos, que hacen vital el análisis de datos que permita configurar un sistema nervioso digital, que coadyuve a la toma de mejores decisiones en la gestión académica. Entonces, se constituye en un elemento imprescindible que todo software de gestión académica debe incluir la mayor variedad posible de módulos que coadyuven al análisis de los datos, ya sea mediante procesos determinísticos y estocásticos.

El análisis de resultados se efectuará considerando cada característica distintiva de los docentes de manera individual y también por pares, de acuerdo a los objetivos planteados.

En la tabla 17, se resumen la influencia de la característica *Grado Académico* del docente.

**Tabla 17**

*Análisis de la Influencia de la Característica Grado Académico*

Influencia del Grado Académico de los Docentes en el Rendimiento Académico de los Estudiantes			
Semestres: 2021-1 al 2023-2			
GradoAcademico	Promedio	Aprobado	NroAlumnos
DOCTORADO	14.54	85.65	102010.00
MAESTRÍA	14.11	81.90	80269.00
TÍTULO	14.11	82.35	44022.00

Como se puede apreciar, los docentes con grado académico de Doctorado tienen mayor influencia en el rendimiento académico de los estudiantes, tanto a nivel de mejores promedios (14.54), así como también, a nivel del porcentaje de aprobados (85.65).

Así mismo, en la tabla 18, se resumen la influencia de la característica *Experiencia* del docente.

**Tabla 18**

*Análisis de la Influencia de la Característica Experiencia*

Influencia de la Experiencia de los Docentes en el Rendimiento Académico de los Estudiantes			
Semestres: 2021-1 al 2023-2			
Experiencia	Promedio	Aprobado	NroAlumnos
15-20	14.51	84.83	45215

En este punto, se da una inferencia discutible, que ameritaría un mayor análisis, considerando talvez más variables, porque son los docentes con una experiencia entre 15 y 20 años en el ejercicio de la docencia los que logran mejores resultados, tanto a nivel de mejores promedios (14.51) y del porcentaje de aprobados (84.83).

Considerando el par de características *Grado Académico* y *Experiencia*, en la tabla 19 se resumen la influencia de ambas características del docente.

**Tabla 19**

*Análisis de la Influencia de las Características Grado Académico y Experiencia*

Influencia del Grado Académico y la Experiencia de los Docentes en el Rendimiento Académico de los Estudiantes			
Semestres: 2021-1 al 2023-2			
Experiencia	Promedio	Aprobado	NroAlumnos
DOCTORADO			
15-20	14.66	86.21	23082
MAESTRÍA			
10-15	14.40	84.32	10862.00
TÍTULO			
15-20	14.36	83.02	7573.00
>25	14.28	84.26	23595.00

Considerando ambas características, los docentes con grado académico de *Doctorado* y con una experiencia de entre 15 y 20 años logran un mejor rendimiento respecto a las notas de los estudiantes (14.66) y porcentaje de estudiantes aprobados (86.21). Respecto a los docentes con el grado académico de *Maestría*, los que tienen entre 10 y 15 años de experiencia son los que logran que los estudiantes tengan un mejor rendimiento académico; en promedio de notas (14.40) y porcentaje de estudiantes aprobados (84.32). Respecto a los docentes con el grado académico de *Título*, los que tienen entre 15 y 20 años de experiencia son los que logran que los estudiantes tengan un mejor rendimiento académico; en promedio de notas (14.36), sin embargo no son los que logran el mayor porcentaje de aprobados, sino, los mayores a 25 años con un porcentaje de (84.26) de estudiantes aprobados.

Otro par de características distintivas de los docentes que tienen influencia en el rendimiento académico de los estudiantes son el grupo etario y el género. Se efectúa el análisis primero de cada característica independientemente, luego, la de ambas características a la vez.

En la tabla 20, se resumen la influencia de la característica *Grupo etario* del docente.

**Tabla 20**

*Análisis de la Influencia de la Característica Grupo Etario*

Influencia del Grupo Etario de los Docentes en el Rendimiento Académico de los Estudiantes			
Semestres: 2021-1 al 2023-2			
GrupoEtario	Promedio	Aprobado	NroAlumnos
>70	14.53	85.91	72854

Como se puede apreciar, los docentes con más de 70 años tienen mayor influencia en el rendimiento académico de los estudiantes, tanto a nivel de mejores promedios (14.53), así como también, a nivel del porcentaje de aprobados (85.91).

Así mismo, en la tabla 21, se resumen la influencia de la característica *Género* del docente.

**Tabla 21***Análisis de la Influencia de la Característica Género*

Influencia del Género de los Docentes en el Rendimiento Académico de los Estudiantes			
Semestres: 2021-1 al 2023-2			
Género	Promedio	Aprobado	NroAlumnos
Femenino	14.43	84.38	59335
Masculino	14.25	83.42	166966

En este punto, se puede apreciar que los docentes del género *femenino* tienen mayor influencia en el rendimiento académico de los estudiantes, tanto a nivel de mejores promedios (14.43), así como también, a nivel del porcentaje de aprobados (84.38).

Considerando el par de características *Grupo Etario* y *Género*, en la tabla 22 se resumen la influencia de ambas características del docente.

**Tabla 22***Análisis de la Influencia de las Características Grupo Etario y Género*

Influencia del Género y el Grupo Etario de los Docentes en el Rendimiento Académico de los Estudiantes			
Semestres: 2021-1 al 2023-2			
GrupoEtario	Promedio	Aprobado	NroAlumnos
FEMENINO			
>70	14.62	87.04	12148.00
MASCULINO			
>70	14.51	85.68	60706.00

Considerando ambas características, los docentes con grupo etario mayor a 70 años, tienen una mayor influencia en el rendimiento académico de los estudiantes.

Otro par de características distintivas de los docentes que tienen influencia en el rendimiento académico de los estudiantes son la categoría y el régimen. Se efectúa el análisis primero de cada característica independientemente, luego, la de ambas características a la vez.

En la tabla 23, se resumen la influencia de la característica *Categoría* del docente.

**Tabla 23**

*Análisis de la Influencia de la Característica Categoría*

Influencia de la Categoría de los Docentes en el Rendimiento Académico de los Estudiantes			
Semestres: 2021-1 al 2023-2			
Categoría	Promedio	Aprobado	NroAlumnos
AUXILIAR	13.93	81.05	16218
ASOCIADO	14.23	83.22	68849
PRINCIPAL	14.38	84.20	141234

Como se puede apreciar, los docentes de la categoría *Principal* tienen mayor influencia en el rendimiento académico de los estudiantes, tanto a nivel de mejores promedios (14.38), así como también, a nivel del porcentaje de aprobados (84.20).

Así mismo, en la tabla 24, se resumen la influencia de la característica *Régimen* del docente.

**Tabla 24**

*Análisis de la Influencia de la Característica Régimen*

Influencia del Régimen de los Docentes en el Rendimiento Académico de los Estudiantes			
Semestres: 2021-1 al 2023-2			
Régimen	Promedio	Aprobado	NroAlumnos
TP 10 H	14.79	86.14	5049
TP 20 H	14.87	86.28	10578
TC	14.33	84.01	79522
DE	14.22	83.17	131152

Como se puede apreciar, los docentes del régimen *Tiempo parcial a 20 horas (TP 20 H)* tienen mayor influencia en el rendimiento académico de los estudiantes, tanto a nivel de mejores promedios (14.87), así como también, a nivel del porcentaje de aprobados (86.28).

Considerando el par de características *Categoría y régimen*, en la tabla 25 se resumen la influencia de ambas características del docente.

**Tabla 25**

*Análisis de la Influencia de las Características Categoría y Régimen*

Influencia de la Categoría de los Docentes en el Rendimiento Académico de los Estudiantes			
Semestres: 2021-1 al 2023-2			
Regimen	Promedio	Aprobado	NroAlumnos
AUXILIAR			
TP 10 H	14.91	83.33	1062.00
TP 20 H	13.99	83.76	2629.00
ASOCIADO			
TP 20 H	15.52	92.22	5196.00
PRINCIPAL			
TP 10 H	14.81	88.92	2049.00

Considerando ambas características, los docentes de la categoría *Auxiliar* y del régimen a *Tiempo Parcial a 10 (TP 10 H)* y *20 Horas (TP 20 H)* tienen una mayor influencia en el rendimiento académico de los estudiantes; mientras que, los docentes en la categoría *Asociado* los que tienen mayor influencia son los de *Tiempo Parcial a 20 Horas (TP 20 H)*; finalmente, en la Categoría *Principal* los que tienen mayor influencia son los de *Tiempo Parcial a 10 Horas (TP 10 H)*

Respecto a la relevancia de las características distintivas de los docentes, se determina que efectivamente, para los estudiantes hay algunas características distintivas con mayor relevancia que otras. De acuerdo al gráfico se tiene que para los estudiantes es más relevante la característica *Experiencia* con una ponderación de 0.359034; seguido por, la característica *Edad* con una ponderación de 0.352979. Las otras características las consideran ínfimamente relevantes, sobre todo el género cuya ponderación es mínima con un 0.049961.

Respecto a la relevancia de las preguntas de las encuestas aplicadas a los estudiantes sobre el desempeño académico de sus respectivos docentes, se determina que hay algunas preguntas que tienen bastante relevancia, sobre todo aquellas que tienen que ver con aspectos comunicacionales, como por ejemplo la pregunta P14 “*El profesor utiliza un lenguaje claro durante el proceso de enseñanza y aprendizaje*” con una ponderación de 0.104397, así como la pregunta P13 “*El profesor utiliza la metodología de enseñanza adecuada a las características del grupo y de la asignatura*” con una ponderación de 0.100048. Mientras que las preguntas que no tienen una relación directa con los procesos de enseñanza aprendizaje no implican una relevancia significativa para los estudiantes, tal la pregunta P10 “*El profesor cumple con los horarios de tutoría establecidos*” que tiene una ponderación de 0.022155. En consecuencia, en base a los resultados obtenidos en el presente trabajo, se podría reconsiderar la reformulación de las encuestas y tomar en cuenta por ejemplo sólo aquellas preguntas que tengan una ponderación mayor a 0.05.

Respecto a la comparación de los resultados obtenidos en el presente trabajo con los antecedentes referenciados en el presente trabajo, se puede manifestar, que la mayoría de trabajos orientados a analizar el rendimiento académico de los estudiantes analizan más datos asociados a la dimensión estudiante y/o infraestructura; muy pocos trabajos analizan datos de la dimensión docente.

El trabajo realizado por Chávez et al. (2020), presentan un estudio que utiliza técnicas de minería de datos para analizar datos correspondientes a la dimensión docente, similar al presente trabajo. Sin embargo, en el caso del trabajo de Chávez, se compara el rendimiento de profesores dedicados exclusivamente a la enseñanza y aquellos que también se dedican a la investigación; analizando la influencia de estas características en el rendimiento académico; mientras que en el presente trabajo, se analizan otras características distintivas de los docentes, como: grado académico, experiencia, género, categoría, etc. Lo relevante, es que ambos

trabajos sugieren que, en general, las características distintivas de los docentes no presentan diferencias significativas en el rendimiento académico de los estudiantes; aunque, a partir de esas pequeñas diferencias, se puede plantear estrategias y acciones de mejora en el proceso de formación profesional de los estudiantes.

Mientras que, con los trabajos de molero Molero et al. (2018), Martínez et al. (2016), Nafie et al. (2021) y Batool et al. (2022), hay coincidencia con algunas de las técnicas de minería de datos utilizada en el procesamiento de los datos; y en la intención que a partir de estos análisis se puedan plantear nuevas estrategias o políticas de mejora en el proceso de enseñanza-aprendizaje. Sin embargo, estos trabajos consideran datos más orientados a la dimensión estudiante y a la dimensión infraestructura, tales como: nivel socioeconómico, cuestiones demográficas, contexto familiar, aspectos psicológicos, condiciones del entorno en el que viven los estudiantes, ambientes, equipos, laboratorios, etc. Por tanto, la comparación entre estos trabajos y el presente trabajo, va más orientado a las técnicas utilizadas, que son similares.

## VI. CONCLUSIONES

- Se determinó las implicancias que existe entre las características distintivas de los docentes en el rendimiento académico de los estudiantes en las universidades públicas de la región Cusco, utilizando técnicas de ciencia de datos; lo que permite plantear estrategias de mejora en el proceso de enseñanza-aprendizaje.
- Se determinó la medida en que influyen el grado académico y la experiencia de los docentes en el rendimiento académico de los estudiantes en las universidades públicas de la región del Cusco, siendo lo más relevante que los docentes de grado académico de *Doctorado* y con relativa *Experiencia* son los que más influyen en el rendimiento académico de los estudiantes.
- Se determinó la medida en que influyen el grupo etario y el género de los docentes en el rendimiento académico de los estudiantes en las universidades públicas de la región del Cusco, donde lo más significativo es que las docentes del género *Femenino* y del *Grupo Etario* de mayor edad tienen ligeramente mayor influencia en el rendimiento académico de los estudiantes.
- Se determinó la medida en que influyen la categoría y régimen al que pertenecen los docentes en el rendimiento académico de los estudiantes en las universidades públicas de la región del Cusco, donde los docentes de la categoría *Asociado* y del régimen de *Tiempo Parcial* presentan mayor influencia en el rendimiento académico de los estudiantes.
- Se determinó que desde la perspectiva de los estudiantes, tienen más relevancia las características distintivas de los docentes que están asociadas a *Experiencia* y *Edad*. Así mismo, respecto a las encuestas, hay preguntas que tienen una relevancia significativa, como también hay preguntas con una significancia mínima, situaciones que podrían implicar la reformulación de dichas encuestas.

## VII. RECOMENDACIONES

- En el presente trabajo se consideró datos de las características distintivas más importantes de los docentes que influyen en el rendimiento académico de los estudiantes; sin embargo, hay muchas otras características distintivas que no se consideraron, por ejemplo, competencias pedagógicas, nivel de actualización, competencias tecnológicas, etc. Se recomienda incluir datos de estas otras características y aplicar técnicas de ciencia de datos que permita descubrir patrones de comportamiento y reglas de asociación con una perspectiva más integral.
- Un aspecto relevante en las universidades es la producción de conocimiento científico, en ese sentido, muchos docentes pertenecen a la categoría de docentes investigadores. Se recomienda también efectuar un estudio sobre la influencia que tienen los docentes investigadores RENACYT sobre el rendimiento académico de los estudiantes, comparado con los que no lo son.
- El avance vertiginoso de la tecnología, propone excelentes herramientas tecnológicas en el contexto educativo, en especial los denominados *Sistemas de Gestión de Aprendizaje* (LMS - Learning Management System); estas herramientas, requieren competencias adicionales de los docentes. Se recomienda investigar sobre el uso de estas herramientas, el impacto de ellas, el nivel de desempeño del docente y sobre todo las competencias necesarias para desarrollarse en este contexto.
- La actual era del conocimiento es consecuencia de las transformaciones digitales trascendentales propiciadas por la tecnología; sin embargo, en el plano educativo estas transformaciones han impactado mínimamente; es así, que aún se siguen utilizando paradigmas del siglo pasado, con sesiones de clase eminentemente expositivas. Se recomienda, trabajos multidisciplinarios, con profesionales en: educación, sociología, psicología y tecnologías de la información, para proponer nuevos paradigmas educativos,

donde el docente sea eminentemente un guía o mentor; por tanto, como parte del trabajo se debe determinar que otras características distintivas debe desarrollar el docente, de modo que influya positivamente en el rendimiento académico de los estudiantes.

## VIII. REFERENCIAS

- Aguilar, S., & Arreguín, M. (2021). *Perfil docente universitario para el siglo XXI: Competencias y habilidades necesarias en un entorno tecnológico*. Revista de Ciencias Sociales, (1), 23-35. <https://doi.org/10.36410/rcs.2021.01.02>.
- Batool, S., Rashid, J., Nisar, M., Kim, J., Kwon, Hyuk-Yoon & Hussain, A. (2022). *Educational data mining to predict students' academic performance: A survey study*. Education and Information Technologies. 28. <https://doi.org/10.1007/s10639-022-11152-y>.  
[https://www.researchgate.net/publication/361866122\\_Educational\\_data\\_mining\\_to\\_predict\\_students'\\_academic\\_performance\\_A\\_survey\\_study](https://www.researchgate.net/publication/361866122_Educational_data_mining_to_predict_students'_academic_performance_A_survey_study)
- Bramer, M. (2016). *Principles of Data Mining*. Springer.
- Bunge, M. (1974). *La ciencia, su método y su filosofía*. Siglo Veintiuno Editores.
- Chávez, M. & Ceballos, H. & Cantu-Ortiz, F. (2020). *A data analytics approach to contrast the performance of teaching (only) vs. research professors*. International Journal on Interactive Design and Manufacturing (IJIDeM). 14. 1-16.  
<https://doi.org/10.1007/s12008-020-00713-5>.  
[https://www.researchgate.net/publication/346437809\\_A\\_data\\_analytics\\_approach\\_to\\_contrast\\_the\\_performance\\_of\\_teaching\\_only\\_vs\\_research\\_professors](https://www.researchgate.net/publication/346437809_A_data_analytics_approach_to_contrast_the_performance_of_teaching_only_vs_research_professors)
- Chicaiza, D., & Pico, M. (2019). *Competencias docentes y perfil del docente universitario para el siglo XXI*. Revista Observatorio de Innovación Educativa en América Latina y el Caribe, (15), 13-26. <https://doi.org/10.26817/16641093.461>
- García, R. M., & Cepeda, A. (2022). *Factores asociados al rendimiento académico de los estudiantes universitarios: Un análisis exploratorio*. Revista Latinoamericana de Educación Superior, (1), 78-92. <https://doi.org/10.28992/rles2022.1.04>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

- Harari, Y. (2018). *21 lecciones para el siglo XXI* (1a. ed.). Pendin Random House Grupo Editorial.
- Hernández Sampieri, R., & Mendoza Torres, C. (2018). *Metodología de la investigación las rutas cuantitativa, cualitativa y mixta* (1a. ed.) McGraw-Hill.
- Instituto Nacional de Estadística e Informática [INEI]. (2021). *Datos estadísticos de Educación Universitaria*. <https://m.inei.gob.pe/estadisticas/indice-tematico/university-tuition/>
- Kelleher, J. D., & Tierney, B. (2021). *Ciencia de datos*. MIT Press.
- Laney, D. (2001). *3D Data Management: Controlling Data Volume, Velocity, and Variety*. META Group Research Note, 6, 70.
- Ley universitaria. (2014). Ley N° 30220: *Ley Universitaria*. *El Peruano*.  
<https://busquedas.elperuano.pe/normaslegales/ley-universitaria-ley-n-30220-1067350-1/>
- Martínez Abad, F. & Chaparro, A. (2016). *Data-mining techniques in detecting factors linked to academic achievement*. *School Effectiveness and School Improvement*. 1-17.  
10.1080/09243453.2016.1235591.
- Molero-Castillo, G. & Maldonado-Hernández, G. & Mezura-Godoy, C. & Benitez-Guerrero, E. (2018). *Interactive System for the Analysis of Academic Achievement at the Upper-Middle Education in Mexico*. *Computación y Sistemas*. 22.  
<https://doi.org/10.13053/cys-22-1-2773>.  
[https://www.researchgate.net/publication/324204456\\_Interactive\\_System\\_for\\_the\\_Analysis\\_of\\_Academic\\_Achievement\\_at\\_the\\_Upper-Middle\\_Education\\_in\\_Mexico](https://www.researchgate.net/publication/324204456_Interactive_System_for_the_Analysis_of_Academic_Achievement_at_the_Upper-Middle_Education_in_Mexico)
- Murphy, K. P. (2021). *Machine Learning: A Probabilistic Perspective*. MIT Press.

- Nafie, F. & Hamed, A. (2021). *Using Data Mining Techniques in Building a Model to Determine the Factors Affecting Academic Data for Undergraduate Students*.  
<https://doi.org/306.10.22937/IJCSNS.2021.21.4.38>.  
[https://www.researchgate.net/publication/353751722\\_Using\\_Data\\_Mining\\_Techniques\\_in\\_Building\\_a\\_Model\\_to\\_Determine\\_the\\_Factors\\_Affecting\\_Academic\\_Data\\_for\\_Undergraduate\\_Students](https://www.researchgate.net/publication/353751722_Using_Data_Mining_Techniques_in_Building_a_Model_to_Determine_the_Factors_Affecting_Academic_Data_for_Undergraduate_Students)
- PyPi (2024). *The Python Package Index*. <https://pypi.org/>
- Palmer, S. (2015). *Data Science for the C-Suite*. Digital Living Press.
- Piatetsky-Shapiro, G. (1989). *Knowledge Discovery in Databases*. *AI Magazine*, 11(5), 37-54.
- Rodríguez, R., & Díaz, J. (2020). *Estrategias de apoyo académico para mejorar el rendimiento estudiantil en la educación superior*. *Revista de Investigación Académica*, (15), 1-12. Recuperado de <https://ria.uaemex.mx/handle/123456789/9202>.
- Schwab, K. (2016). *La Cuarta Revolución Industrial*. Foro Económico Mundial.
- VanderPlas, J. (2016). *Python data science handbook: Essential tools for working with data*. O'Reilly Media, Inc.

## GLOSARIO

**Apache Hadoop:** Framework de software de código abierto utilizado para el almacenamiento y procesamiento distribuido de grandes conjuntos de datos en clústeres de computadoras.

**Apache Spark:** Framework de computación en clúster de código abierto que proporciona un motor unificado para el procesamiento de big data, con soporte para procesamiento en memoria y en disco.

**Bases de datos:** Conjunto de datos de entidades u objetos y sus relaciones, almacenados con redundancia controlada, cuyo propósito es servir a una o varias aplicaciones y satisfacer necesidades de información de los usuarios. También debe almacenar la estructura y las restricciones.

**BIG DATA:** Término que hace referencia al manejo y análisis de grandes conjuntos de datos que pueden ser de gran volumen, velocidad y variedad, y que requieren de herramientas especiales para su procesamiento.

**Ciencia de datos:** Disciplina que utiliza técnicas de estadística, matemáticas, informática y visualización para analizar grandes cantidades de datos y extraer conocimientos útiles de ellos.

**Clustering:** en el contexto del aprendizaje automático no supervisado, es el proceso de agrupar un conjunto de objetos en subconjuntos, o "clusters", de modo que los objetos dentro de un mismo cluster sean más similares entre sí que con los objetos en otros clusters.

**CSV:** (Comma-separated values ) Formato de archivo de texto en forma de tabla, donde los valores de cada fila están separados por comas.

**Data Lake:** Es un repositorio de grandes volúmenes de datos estructurados y no estructurados, que provienen de diversos orígenes. Generalmente se almacenan sin efectuar procesos de transformación de datos. Tiene como propósito fundamental, proveer la

estructura que permita la exploración y el descubrimiento de patrones, que coadyuve al análisis de datos y a la toma de decisiones informadas.

**Data Warehouse:** En el contexto informático es una colección o almacén de datos que se utiliza como soporte a la inteligencia de negocios.

**Data Mining:** Es un conjunto de técnicas encaminadas a la extracción de conocimiento procesable, implícito en las bases de datos.

**Dataset:** Conjunto de datos estructurados que se utilizan como entrada para un análisis o proceso específico.

**ETL:** Proceso que consiste en extraer datos de diversas fuentes, transformarlos para que cumplan con requisitos específicos y cargarlos en un sistema de almacenamiento de datos para su análisis.

**INEI:** Instituto Nacional de Estadística e Informática.

**Inteligencia de negocios:** Conjunto de técnicas y herramientas para convertir datos en información y ésta en conocimiento, como soporte para la toma de decisiones estratégicas.

**Inteligencia analítica:** Conjunto de técnicas y herramientas matemáticas, estadísticas y de procesamiento de datos, orientadas a identificar patrones, tendencias y relaciones ocultas; que permitan esbozar planes estratégicos para la organización.

**LMS:** (Learning Management System) Un sistema de gestión de aprendizaje es un software que coadyuva a la gestión de los diferentes actividades del proceso de enseñanza – aprendizaje.

**OLAP:** (Online Analytical Processing) Procesamiento analítico en línea. Tecnología de inteligencia de negocios, orientada al análisis multidimensional de datos como soporte a la toma de decisiones.

**OLTP:** (Online Transactional Processing) Procesamiento de transacciones en línea. Técnica referida al registro de datos en las bases de datos transaccionales.

**Paradigma:** Conjunto de teorías, estándares y métodos que juntos representan una forma de organizar el conocimiento.

**Procesamiento determinístico:** Procesamiento de datos que produce los mismos resultados para las mismas entradas o condiciones iniciales de los datos. Se considera que son predecibles y consistentes.

**Procesamiento estocástico:** Procesamiento de datos que produce resultados que varían incluso para las mismas entradas o condiciones iniciales de los datos, debido a que interviene en el proceso factores asociados a la aleatoriedad o incertidumbre.

**Script:** Fragmentos de código o instrucciones de computador, utilizado principalmente en el contexto de las bases de datos.

**SGBD:** Sistema de Gestión de Base de Datos. Software mediante el cual se interactúa con las bases de datos.

**SQL:** Lenguaje Estructurado de Consulta en sistemas de gestión de bases de datos relacionales (Structure Query Language).

**SQL Server:** Software de gestión de bases de datos de la empresa Microsoft.

**TIC:** Tecnologías de la Información y la Comunicación.

**Tupla:** Fila o registro de una base de datos.

## IX. ANEXOS

## ANEXO 1. Matriz de consistencia

PROBLEMAS	OBJETIVO	HIPOTESIS	VARIABLES, DIMENSIONES E INDICADORES	METODOLOGIA
<p>GENERAL</p> <p>¿Qué implicancia existe entre las características distintivas de los docentes (grado académico, grupo etario, categoría y régimen, género, experiencia, así como la percepción o valoración que tienen los estudiantes de los docentes) en el rendimiento académico de los estudiantes en las universidades públicas</p>	<p>GENERAL</p> <p>Determinar qué implicancias existe entre las características distintivas de los docentes (grado académico, grupo etario, categoría y régimen, género, experiencia, así como la percepción o valoración que tiene los estudiantes de los</p>	<p>GENERAL</p> <p>Si se utilizan técnicas de ciencia de datos, se determinan las implicancias que existen entre las características distintivas de los docentes (grado académico, grupo etario, categoría y régimen, género, experiencia, así como la percepción o valoración que</p>	<p>VARIABLE 1</p> <p>Características distintivas de los docentes</p> <p>D - 3. Características distintivas personales</p> <p>I - 3.1. Grado académico.</p> <p>I - 3.2. Experiencia.</p> <p>I - 3.3. Grupo etario.</p> <p>I - 3.4. Género</p> <p>I - 3.5. Categoría</p> <p>I - 3.6. Régimen</p>	<p><b>Enfoque</b></p> <p>Cuantitativo</p> <p><b>Tipo</b></p> <p>Aplicada</p> <p><b>Nivel</b></p> <p>Correlacional</p> <p><b>Diseño</b></p> <p>No experimental</p> <p><b>Técnicas</b></p> <p>Ciencia de datos</p>

<p>de la región Cusco, utilizando técnicas de ciencia de datos?</p> <p><b>ESPECÍFICOS</b></p> <p>a) ¿En qué medida influyen el grado académico y la experiencia de los docentes en el rendimiento</p>	<p>docentes) en el rendimiento académico de los estudiantes en las universidades públicas de la región Cusco, utilizando técnicas de ciencia de datos.</p> <p><b>ESPECÍFICOS</b></p> <p>a) Determinar en qué medida influyen el grado académico y la experiencia de los docentes en el rendimiento académico</p>	<p>tiene los estudiantes de los docentes) en el rendimiento académico de los estudiantes en las universidades públicas de la región Cusco.</p> <p><b>ESPECÍFICOS</b></p> <p>a) Si se analizan los datos referidos al grado académico y la experiencia de los docentes se determina la medida en que influyen</p>	<p>D - 4. Valoración de los docentes por parte de los estudiantes</p> <p>I - 4.1. Ponderación numérica.</p> <p>D - 5. Rendimiento académico</p> <p>I - 5.1. Porcentaje de aprobados.</p> <p>I - 5.2. Nota promedio asignada por docente</p> <p><b>VARIABLE 2</b></p> <p>Implicancia de las características distintivas de los docentes en el rendimiento académico de los estudiantes.</p> <p>D - 6. Implicancia de las características distintivas de los docentes en el rendimiento académico de los estudiantes</p>	<p><b>Instrumento</b></p> <p>Datasets en formato CSV y encuestas oficiales de valoración de los docentes por parte de los estudiantes.</p> <p><b>Población:</b></p> <p>Está constituida por la totalidad de los datos de registros oficiales subyacentes en las bases de datos transaccionales.</p> <p><b>Muestra</b></p> <p>Está constituida por la totalidad de los datos académicos de registros oficiales que serán utilizados para determinar las implicancias de las características distintivas de los docentes en el rendimiento académico de</p>
---	--	--	--	---

<p>académico de los estudiantes en las universidades públicas de la región del Cusco?</p> <p>b) ¿En qué medida influyen el grupo etario y el género de los docentes en el rendimiento académico de los estudiantes en las universidades públicas de la región del Cusco?</p>	<p>de los estudiantes en las universidades públicas de la región del Cusco.</p> <p>b) Determinar en qué medida influyen el grupo etario y el género de los docentes en el rendimiento académico de los estudiantes en las universidades públicas de la región del Cusco.</p>	<p>en el rendimiento académico de los estudiantes en las universidades públicas de la región del Cusco.</p> <p>b) Si se analizan los datos referidos al grupo etario y el género de los docentes se determina la medida en que influyen en el rendimiento académico de los estudiantes en las universidades públicas de la región del Cusco.</p>	<p>I - 6.1. Nivel en que influye el grado académico y experiencia de los docentes en el rendimiento académico de los estudiantes.</p> <p>I - 6.2. Nivel en que influye el grupo etario y el género de los docentes en el rendimiento académico de los estudiantes.</p> <p>I - 6.3. Nivel en que influye la categoría y el régimen de los docentes en el rendimiento académico de los estudiantes.</p> <p>I - 6.4. Patrones de comportamiento entre la valoración que tienen los estudiantes de los docentes y el rendimiento académico de los estudiantes.</p> <p>I - 6.5. Asociaciones entre la valoración que tienen los estudiantes de los docentes y el</p>	<p>los estudiantes en las universidades de la región del Cusco.</p> <p><b>Procesamiento y análisis de datos</b></p> <ul style="list-style-type: none"> <li>• Recolección de datos con procesos ETL (Extract, Transform and Load)</li> <li>• Preparación de los datos utilizando técnicas de ciencia de datos, como discretización.</li> <li>• Aplicación de modelos de aprendizaje automático como Random Forest, con sus respectivas métricas de confiabilidad</li> </ul>
--	--	--	---	--

<p>c) ¿En qué medida influyen la categoría y régimen al que pertenecen los docentes en el rendimiento académico de los estudiantes en las universidades públicas de la región del Cusco?</p>	<p>c) Determinar en qué medida influyen la categoría y régimen al que pertenecen los docentes en el rendimiento académico de los estudiantes en las universidades públicas de la región del Cusco.</p>	<p>c) Si se analizan los datos referidos a la categoría y régimen al que pertenecen los docentes se determina la medida en que influyen en el rendimiento académico de los estudiantes en las universidades públicas de la región del Cusco.</p>	<p>rendimiento académico de los estudiantes.</p>	<ul style="list-style-type: none"> <li>• Interpretación y discusión de resultados</li> </ul>
<p>d) ¿En qué medida tiene relevancia las características distintivas de los docentes en la</p>	<p>d) Determinar en qué medida tiene relevancia las características distintivas de los</p>	<p>d) Si se analizan las características distintivas de los docentes en la valoración o percepción que tienen los estudiantes</p>		

<p>valoración o percepción que tienen los estudiantes de los docentes, así como la pertinencia de las encuestas aplicadas a los estudiantes para evaluar el desempeño de los docentes, en las universidades públicas de la región del Cusco?</p>	<p>docentes en la valoración o percepción que tienen los estudiantes de los docentes, así como la pertinencia de las encuestas aplicadas a los estudiantes para evaluar el desempeño de los docentes, en las universidades públicas de la región del Cusco.</p>	<p>de los docentes, así como la pertinencia de las encuestas aplicadas a los estudiantes para evaluar el desempeño de los docentes, en las universidades públicas de la región del Cusco, se determinan en qué medida tiene relevancia estas características y la pertinencia de las encuestas.</p>		
--	---	---	--	--