



**Universidad Nacional
Federico Villarreal**

**Vicerrectorado de
INVESTIGACIÓN**

ESCUELA UNIVERSITARIA DE POSGRADO

**“MODELO BASADO EN EL ANALISIS MULTIVARIADO DE DATOS PARA EL
ESTUDIO DEL RENDIMIENTO ACADEMICO DE LAS ESCUELAS
PROFESIONALES UNIVERSITARIAS PÚBLICAS DE LA REGION DEL CUSCO”**

**TESIS PARA OPTAR EL GRADO ACADÉMICO DE:
DOCTOR EN INGENIERIA DE SISTEMAS**

**AUTOR
ROBERT WILBERT ALZAMORA PAREDES**

**ASESORA
DRA. JACKELINE ROXANA HUAMÁN FERNÁNDEZ**

**JURADO:
DR. DÍAZ DUMONT JORGE RAFAEL
DR. FRANCO DEL CARPIO CARLOS MIGUEL
DRA. JURADO FALCONI EULALIA**

LIMA – PERÚ

2020

INDICE

INDICE DE GRÁFICOS	IV
INDICE DE TABLAS	VI
RESUMEN.....	VII
ABSTRACT	VIII
I. INTRODUCCION	9
1.1 Planteamiento del problema	10
1.2 Descripción del problema.....	11
1.3 Formulación del problema	12
-Problema general.....	12
-Problemas específicos	12
1.4 Antecedentes	13
1.5 Justificación de la Investigación.....	17
1.6 Limitaciones de la Investigación.....	17
1.7 Objetivos	18
- Objetivo General	18
- Objetivo Específicos	18
1.8 Hipótesis.....	18
1.8.1. Hipótesis General	18
1.8.2. Hipótesis Específicas.....	19
II. MARCO TEÓRICO.....	20
2.1 Marco conceptual	20
III. MÉTODO.....	58
3.1 Tipo de investigación	58
3.2 Población y muestra	58

3.3	Operacionalización de variables.....	61
3.4	Instrumentos	65
3.5	Procedimientos	69
3.6	Análisis de datos.....	69
IV.	RESULTADOS	70
4.1	Análisis descriptivo por dimensión	71
4.2	Contrastación de hipótesis.....	77
V.	DISCUSIÓN DE RESULTADOS	86
VI.	CONCLUSIONES	89
VII.	RECOMENDACIONES	91
VIII.	REFERENCIAS	93
IX.	ANEXOS.....	97
	ANEXO 1. Matriz de consistencia.....	97
	ANEXO 2. Documentación de usuario para aplicar el modelo	101
	ANEXO 3. Código en R para prueba del modelo	115
	ANEXO 4: Cuestionario	131
	ANEXO 5: Tablas de valoración de instrumento por expertos.....	133

INDICE DE GRÁFICOS

Gráfico 1. Clasificación de los algoritmos de formación de conglomerados.....	26
Gráfico 2. Diferentes cortes en un dendograma	27
Gráfico 3. Visualización clásica de un clúster mediante coordenadas paralelas.....	28
Gráfico 4. Pseudocódigo para el clasificador k-NN.....	31
Gráfico 5. Ejemplo de clasificación mediante k-NN	31
Gráfico 6. Diagrama BPM de la vista general del modelo.....	35
Gráfico 7. Vista parcial de los datos iniciales en hoja de cálculo	36
Gráfico 8. Diagrama BPM del subproceso de imputación de datos faltantes	38
Gráfico 9. Vista de calificaciones sin valores perdidos	38
Gráfico 10. Comparación entre métodos de imputación.....	39
Gráfico 11. Diagrama BPM del subproceso Análisis de Correspondencias	40
Gráfico 12. Vista parcial de la tabla de contingencias	41
Gráfico 13. Análisis de correspondencias con/sin outliers.....	42
Gráfico 14. Heatmap de la distribución de frecuencia de notas	42
Gráfico 15. Varianza explicada en factores del análisis de correspondencias con/sin elementos outliers.....	43
Gráfico 16. Diagrama BPM para el subproceso Análisis de Componentes Principales	44
<i>Gráfico 17.</i> Distribución de los estudiantes en base al ACP.....	44
<i>Gráfico 18.</i> Varianza de los componentes obtenidos mediante ACP.....	45
<i>Gráfico 19.</i> Relación entre el primer componente y el rendimiento académico.....	45
<i>Gráfico 20.</i> Diagrama BPM para el subproceso Análisis de Conglomerados	46
<i>Gráfico 21.</i> Histograma del rendimiento académico por clases de estudiantes	47
<i>Gráfico 22.</i> Dendograma de asignaturas mediante algoritmo Ward.....	47
<i>Gráfico 23.</i> Siluetas de los clusters de alumnos y asignaturas.....	48

<i>Gráfico 24.</i> Diagrama BPM para el subproceso Ordenamiento por clusters	49
<i>Gráfico 24.</i> Particionamiento de asignaturas y estudiantes.....	49
<i>Gráfico 26.</i> Particionamiento de asignaturas y estudiantes considerando el promedio de cada sector	50
<i>Gráfico 27.</i> Diagrama BPM para el subproceso Aplicación de árboles de clasificación y regresión	51
<i>Gráfico 28.</i> Árbol de clasificación de estudiantes en función de asignaturas.....	51
<i>Gráfico 29.</i> Árbol de regresión del rendimiento académico en función de las asignaturas	52
<i>Gráfico 30.</i> Árbol de clasificación de clusters de estudiantes en función de los clusters de las asignaturas.....	52
<i>Gráfico 31.</i> Coordenadas paralelas de estudiantes frente a clusters de asignaturas.....	53
<i>Gráfico 32.</i> Modelo de calidad de producto de Datos de la ISO/IEC 25012.....	54
<i>Gráfico 32.</i> Validez de los datos de entrada del modelo	71
<i>Gráfico 33.</i> Usabilidad del modelo.....	72
<i>Gráfico 34.</i> Funcionalidad del modelo	73
<i>Gráfico 35.</i> Capacidad explicativa de los gráficos del modelo	74
<i>Gráfico 36.</i> Modelo basado en el análisis multivariado de datos.....	75
<i>Gráfico 37.</i> Valoración del estudio del rendimiento académico	76

INDICE DE TABLAS

Tabla 1 <i>Operacionalización de variables</i>	61
Tabla 2 <i>Validación del instrumento</i>	66
Tabla 3 <i>Rangos de interpretación del Alpha de Cronbach</i>	67
Tabla 4 <i>Confiabilidad para las dimensiones</i>	68
Tabla 5 <i>Validez de los datos de entrada del modelo</i>	71
Tabla 6 <i>Usabilidad del modelo</i>	72
Tabla 7 <i>Funcionalidad del modelo</i>	73
Tabla 8 <i>Capacidad explicativa de los gráficos del modelo</i>	74
Tabla 9 <i>Modelo basado en el análisis multivariado de datos</i>	75
Tabla 10 <i>Modelo Valoración del estudio del rendimiento académico</i>	76
Tabla 11 <i>Prueba de Kolmogorov-Smirnov para una muestra</i>	77
Tabla 12 <i>Correlación entre la Validez de los datos de entrada del modelo y Valoración del estudio del rendimiento académico</i>	78
Tabla 13 <i>Correlación entre la usabilidad del modelo y valoración del estudio del rendimiento académico</i>	80
Tabla 14 <i>Correlación entre la funcionalidad del modelo y valoración del estudio del rendimiento académico</i>	81
Tabla 15 <i>Correlación entre la capacidad explicativa de los gráficos del modelo y valoración del estudio del rendimiento académico</i>	83
Tabla 16 <i>Correlación entre el modelo basado en el análisis multivariado de datos y valoración del estudio del rendimiento académico</i>	84

RESUMEN

La presente tesis plantea un modelo basado en técnicas de análisis multivariado de datos para estudiar el rendimiento académico de las escuelas profesionales universitarias. La materia prima para el funcionamiento de este modelo, son las calificaciones obtenidas por los egresados de una escuela profesional en sus asignaturas de corte obligatorio. El modelo plantea una serie de pasos ordenados a seguir, con el fin de observar el comportamiento de las variables de estudio desde diferentes ángulos, permitiendo al usuario una visión gráfica de los mismos que le posibilite un mejor entendimiento del rendimiento académico de su escuela profesional. El estudio es de naturaleza cuantitativo-deductivo, aplicado; es de nivel correlacional, con un diseño no experimental. La técnica utilizada fue la encuesta a través de un cuestionario de tipo Likert. La población estuvo formada por personal con cargos ligados a la parte académica de la UNSAAC (decanos, secretarios académicos, directores de escuelas profesionales, directores de departamentos académicos, personal del centro de cómputos y de áreas relacionadas) y especialistas en ingeniería informática y de sistemas, que en total suman 131 individuos. Un muestreo probabilístico aleatorio sin reemplazos, con un margen de error del 5% y un nivel de confianza del 95% arrojó una muestra de 54 individuos. La contrastación de las hipótesis se realizó utilizando el estadístico χ^2 de Pearson, debido a que las pruebas de Kolmogorov-Smirnov determinaron que los datos tenían una distribución normal.

PALABRAS CLAVE: Análisis multivariado de datos, rendimiento académico universitario, modelo, valoración de un modelo.

ABSTRACT

This thesis proposes a model based on multivariate data analysis techniques to study the academic performance of university professional schools. The raw material for the operation of this model, are the qualifications obtained by the graduates of a professional school in their compulsory court subjects. The model proposes a series of ordered steps to follow, in order to observe the behavior of the study variables from different angles, allowing the user a graphic view of them that allows a better understanding of the academic performance of their professional school. The study is quantitative-deductive, applied; It is correlational level, with a non-experimental design. The technique used was the survey through a Likert questionnaire. The population was made up of personnel with positions linked to the academic part of UNSAAC (deans, academic secretaries, directors of professional schools, directors of academic departments, staff of the computer center and related areas) and specialists in computer and systems engineering , which total 131 individuals. A random probabilistic sampling without replacements, with a margin of error of 5% and a confidence level of 95% yielded a sample of 54 individuals. The hypothesis was tested using the Pearson χ^2 statistic, because the Kolmogorov-Smirnov tests determined that the data had a normal distribution.

KEY WORDS: Multivariate data analysis, university academic performance, model, valuation of a model.

I. INTRODUCCION

El presente trabajo plantea un modelo basado en técnicas multivariantes de datos que se pueda aplicar al estudio del rendimiento académico de las escuelas profesionales de las universidades públicas. Para ello se hace uso de las calificaciones obtenidas por los egresados de una escuela profesional en sus asignaturas de corte obligatorio. Esto permite uniformizar los datos iniciales de tal forma que se minimicen los elementos perdidos o no existentes. El modelo plantea una serie de pasos documentados de tal forma que el usuario pueda llevarlos a cabo sin mayor dificultad y pueda interpretar los resultados que se muestran en formato de tablas y gráficas apropiadas.

La estructura del documento consta de nueve secciones. En la primera se explica el problema materia de la investigación; este primer apartado lleva el título de introducción y abarca información referente al planteamiento, descripción y formulación del problema, así como a los antecedentes, la justificación, limitaciones, objetivos e hipótesis de la investigación. En la segunda parte se documenta el marco teórico correspondiente a las principales técnicas multivariantes de datos, el marco conceptual del rendimiento académico y, finalmente, se plantea el modelo y sus variables de medición asociadas. La tercera sección describe el método de investigación utilizado; se establece el tipo de investigación, la población y muestra, la forma de operacionalizar las variables, los instrumentos de recolección de datos, así como los procedimientos para realizar el análisis de datos. La cuarta sección documenta los resultados obtenidos en cada una de las dimensiones y la contrastación de las hipótesis de investigación. En el capítulo quinto, se realiza la discusión de los resultados obtenidos, mientras que, en la sexta y séptima secciones, se incluyen las conclusiones y recomendaciones fruto de la investigación. La sección ocho contiene las fuentes bibliográficas consultadas para el presente trabajo. Finalmente, la novena sección incluye los anexos, dentro de las cuales se pueden observar: la matriz de consistencia, la documentación de usuario, el código en R de la

implementación del modelo, el cuestionario utilizado para la recolección de datos y su validación por parte de los expertos.

1.1 Planteamiento del problema

El sistema universitario peruano en la actualidad vive una etapa de transición. La Ley Universitaria 23733 dada el año de 1983 ha sido derogada, y en su lugar se ha promulgado la Ley 30220 cuyos esfuerzos se orientan a la búsqueda de la calidad universitaria. Este cambio en la legislación obliga a realizar modificaciones en las universidades y, consecuentemente, en las escuelas profesionales.

El proceso de licenciamiento ha puesto en evidencia las alarmantes condiciones en las cuales funcionan muchos de los centros de estudios universitarios, tanto en infraestructura como en cuestiones de corte académico. Una de las principales falencias que se han evidenciado es la falta del currículo apropiado y actualizados para la formación de profesionales que respondan al perfil que el mercado laboral y académico de los tiempos actuales exige.

La actualización constante del plan de estudios es una labor que se debe realizar periódicamente; pero no es una labor que se realiza al azar, por el contrario, debe responder a estudios serios del comportamiento académico estudiantil y de las exigencias del mercado. Sin herramientas apropiadas para estudiar el rendimiento académico es muy poco lo que se puede hacer en esta materia de manera adecuada.

La materia prima para realizar estos estudios son las calificaciones obtenidas por los estudiantes a lo largo del tiempo. Estos datos son capaces de caracterizar el comportamiento académico de una escuela profesional y de permitir descubrir información que se torne relevante para la toma de decisiones. Lamentablemente, el estudio de estos datos no es una práctica habitual en las universidades estatales. Los pocos estudios realizados no son fácilmente

ubicables o yacen empolvados en las bibliotecas de las universidades sin cumplir su principal propósito: servir de retroalimentación al proceso educativo universitario.

Brindar la posibilidad de contar con herramientas que faciliten estas labores es una obligación para quienes estamos relacionados al mundo de la ciencia y la tecnología. En este contexto, un modelo que permita realizar el estudio del rendimiento académico utilizando técnicas multivariadas de datos, puede ser de inmensa ayuda para quienes se encuentran en posición de tomar decisiones académicas dentro de las universidades, pero también para el docente de a pie que quiera observar el comportamiento de las asignaturas a su cargo dentro del contexto general de la escuela profesional a la cual brinda servicio.

1.2 Descripción del problema

Vivimos tiempos en el que el volumen de la información es abrumador. Transformar dicha información en conocimiento es de vital importancia para entender mejor la realidad y, a partir de ello, tomar decisiones. En el ámbito educativo, el estudio del rendimiento académico puede ayudarnos a mejorar distintos aspectos relacionados con la formación de los estudiantes.

Las escuelas profesionales universitarias presentan características propias dentro de sus planes de estudio: algunas son de corte rígido, otras de corte semi-rígido, y otras de corte flexible, El procesamiento de las notas no es uniforme: muchas universidades cuentan con sistemas automatizados que datan de hace muchos años, mientras que otras aún enfrentan problemas en el almacenamiento de las mismas como por ejemplo, pérdida total o parcial de la información, almacenamiento deficiente de notas antiguas, etc. Otro punto a considerar es que los planes de estudio son diferentes de una escuela a otra, e incluso dentro de la misma escuela debido al carácter dinámico de la misma.

En este escenario, realizar el análisis del rendimiento académico de todo el universo de estudiantes de una universidad se torna inviable.

Por estos motivos, el presente estudio se centra en realizar un diseño basado en el modelo de análisis multivariado de datos que permita realizar el estudio del rendimiento académico de los estudiantes de una carrera profesional a la vez, considerando para ello exclusivamente las calificaciones registradas en sus bases de datos institucionales.

1.3 Formulación del problema

Problema general

¿De qué manera las diferentes dimensiones de un modelo basado en el análisis multivariado de datos se relacionan con la valoración del estudio del rendimiento académico de las escuelas profesionales universitarias públicas de la región del Cusco?

Problemas específicos

- a) ¿De qué manera la validez de los datos de entrada del modelo basado en el análisis multivariado se relaciona con la valoración del estudio del rendimiento académico?
- b) ¿De qué manera la usabilidad del modelo basado en el análisis multivariado de datos se relaciona con la valoración del estudio del rendimiento académico?
- c) ¿De qué manera la funcionalidad del modelo basado en el análisis multivariado de datos se relaciona con la valoración del estudio del rendimiento académico?
- d) ¿De qué manera la capacidad explicativa de los gráficos que genera el modelo basado en el análisis multivariado de datos se relaciona con la valoración del estudio del rendimiento académico?

1.4 Antecedentes

Alzamora (2014) en su tesis para optar al grado de maestría en Ingeniería Computacional y Sistemas Inteligentes realiza un estudio exploratorio del rendimiento académico de la escuela profesional de Ingeniería Informática y de Sistemas de la Universidad Nacional de San Antonio Abad del Cusco, utilizando técnicas de Análisis Multivariado de Datos.

En el trabajo se presenta el uso de diversas técnicas y se estudia la pertinencia de cada una de ellas en su aplicación al estudio del rendimiento académico. Los productos obtenidos son validados paso a paso mediante el uso de estadísticos apropiados. Las conclusiones de la tesis están dirigidas a la evaluación de los resultados obtenidos para la escuela profesional en estudio.

Camborda (2014), en su tesis de Maestría de Ingeniería de Sistemas, hace uso de los árboles de decisión como un modelo predictivo del rendimiento de los estudiantes de Ingeniería Civil de la Universidad Continental de Huancayo, identificando las variables de los factores que más influyen en su rendimiento académico.

El trabajo reúne información académica y socio económica de los estudiantes y, aplicando la herramienta WEKA, construye árboles de decisión predictivos, intentando obtener un nivel de exactitud superior al 80% y una concordancia superior a 0.6.

La autora concluye que su trabajo:

permitió validar que la técnica de árboles de decisión aplicados a los factores demográficos, socioculturales, académicos, institucionales predicen el rendimiento académico de los estudiantes de los primeros ciclos de la carrera de Ingeniería civil de la Universidad Continental. (Camborda, 2014)

Además, hace hincapié en que las variables académicas son las más relevantes al momento de hacer predicciones del rendimiento académico.

Lizares (2017) hace una comparación entre dos modelos de clasificación: el modelo de regresión logística y los árboles de clasificación (CHAID) con el fin de evaluar el rendimiento académico de estudiantes de Matemática. La forma de medir el poder de clasificación estuvo definida por los indicadores: Sensibilidad, Curva ROC, Índice de GINI e Índice de Kappa.

La muestra consistió en 3600 registros de estudiantes con información sociodemográfica y académica, considerando un 70% de ellos para el entrenamiento y un 30% para de prueba. El software utilizado fue IBM SPSS 21.

Los resultados sugieren que ambos modelos identifican un grupo de variables explicativas similares: Tipo de colegio, horas de estudio y satisfacción del curso. Además, señalan que los árboles de clasificación lograron mejores resultados con una sensibilidad del 77.6%, AUC=90,1%, Gini =80,2% y Kappa=0,589. La autora manifiesta que los árboles de decisión demuestran ser una herramienta eficaz para obtener un modelo que permitan predecir el rendimiento académico

Chue, Barrerno y Millones (2007) presentan un estudio en el que utilizan técnicas del Analisis Multivariado de Datos como Análisis de Correspondencias Simple, análisis clúster, análisis de covariancia, análisis discriminante, regresión logística binaria y regresión logística ordinal y nominal con el fin de proporcionar un sistema que permita analizar las posibles relaciones entre las notas de las asignaturas obtenidas por los alumnos y algunas variables socioeconómicas de los mismos.

Al finalizar el estudio, los autores asocian cada técnica utilizada con los resultados del rendimiento académico y concluyen que las técnicas multivariadas son útiles para analizar el rendimiento académico de los alumnos.

Ferreiros, Santos y Villares (2017), publican una investigación sobre el Análisis Multivariable de Datos para analizar la influencia de variables como el género, el hábitat, la titularidad del centro o la formación de los padres en rendimiento académico de estudiantes de educación secundaria. La prueba se aplica a 1522 estudiantes que aspiran obtener el premio extraordinario de la Educación Secundaria Obligatoria (ESO) en Galicia, España, en el periodo del 2008 al 2013.

El estudio permitió determinar que las variables en mención influyen en el rendimiento de la población estudiada, y destaca el papel del Análisis Multivariable de Datos (con técnicas como Análisis de Varianza Unifactorial y Factorial) para lograr los objetivos.

Martelo, Acevedo y Martelo (2018), realizan un trabajo cuantitativo, no experimental, transeccional y correlacional sobre 59 alumnos desertores del programa de Ingeniería de Sistemas de la Universidad de Cartagena de Colombia, con el fin de encontrar los factores que inciden en la deserción universitaria.

Para ello utilizan el modelo de Análisis Multivariado de Datos (técnica de análisis factorial), pudiendo identificar 10 componentes relevantes para el estudio de la deserción universitaria. El estudio además resalta la importancia de el Análisis Multivariado para poder reducir la dimensionalidad de las variables a unas pocas con mínima pérdida de información.

Rodríguez (2007), en su tesis doctoral, propone un indicador del desempeño académico estudiantil (alternativo al tradicional promedio de notas) basado en los créditos académicos obtenidos por el estudiante. La autora utiliza técnicas de Análisis Multivariado de Datos que relacionan variables sociodemográficas, académicas y motivacionales y las comparan con ambos modelos.

El trabajo concluye que el comportamiento de ambos indicadores es similar. La correlación fuerte entre ambos modelos sugiere que el rendimiento académico universitario (desde cualquiera de las dos perspectivas) se ve impactado fuertemente por el rendimiento previo del alumno.

De igual manera, el trabajo resalta la importancia de técnicas de Análisis Multivariado de Datos (tales como regresión lineal, regresión logística y modelos de ecuaciones estructurales) para validar los resultados del indicador propuesto y compararlo con el indicador clásico.

Seijas (2000), utiliza técnicas de Análisis Multivariado de Datos, tales como el Análisis de Componentes Principales, Análisis Factorial de Correspondencias y el Análisis Factorial Discriminante en el estudio de variables psicológicas que permiten predecir el desempeño universitario.

El estudio se enfoca en realizar el análisis de variables clásicas de la psicometría (habilidad verbal, habilidad matemática, y el Coeficiente Intelectual), variables cognoscitivas, los estilos de aprendizaje, la autoestima y, finalmente variables transculturales como son el individualismo y el colectivismo para validar y confiabilizar los constructos que la psicología del conocimiento utiliza para explicar los fenómenos que le competen estudiar

El trabajo concluye afirmando que el análisis multivariado de datos es de gran utilidad tanto para la generación como para la comprobación de modelos que ayuden a consolidar las teorías conductuales que rodean a ramas como la psicología.

1.5 Justificación de la Investigación

El rendimiento académico de una escuela profesional, representado por las notas finales obtenidas por sus estudiantes en las diferentes asignaturas de sus currículas, constituye una fuente de datos que, adecuadamente modelada y tratada mediante técnicas de análisis de datos, puede aportar una visión diferente del estudio del rendimiento académico. La evolución de ramas como la Ingeniería de Sistemas, la Informática, la Estadística, la Inteligencia Artificial, la Ingeniería del Software y otras afines, ha permitido explorar la información que permanece oculta tras los datos, facilitando la obtención de conocimiento no trivial y relevante para la toma de decisiones. Relaciones entre variables que antes no se podían conseguir de manera manual, ahora son factibles de descubrir y evaluar, procesando grandes volúmenes de información

En un mundo que busca la calidad, la excelencia y la mejora continua, contar con herramientas de análisis de información es vital para la toma de decisiones. El presente trabajo pretende agregar una visión diferente al estudio del rendimiento académico que, sumado al estudio desde otras perspectivas, puedan contribuir al entendimiento del rendimiento académico de la población estudiantil de las escuelas profesionales y permita tomar decisiones con el fin de mejorar, corregir o cambiar las currículas u otros elementos de estudio de las mismas.

1.6 Limitaciones de la Investigación

Las limitaciones que tendrá el siguiente proyecto de investigación son:

- La falta de trabajos con enfoque similar al presente estudio.
- El tiempo, que se encuentra compartido con actividades laborales.
- La reserva de la información, siempre existente en las organizaciones.
- Calidad de la información histórica registrada en las bases de datos.

1.7 Objetivos

- Objetivo General

Determinar si las diferentes dimensiones de un modelo basado en el análisis multivariado de datos se relacionan con la valoración del estudio del rendimiento académico de los estudiantes de las escuelas profesionales universitarias públicas de la región del Cusco.

- Objetivo Específicos

- a) Determinar si la validez de los datos de entrada del modelo basado en el análisis multivariado se relaciona con la valoración del estudio del rendimiento académico.
- b) Determinar si la usabilidad del modelo basado en el análisis multivariado de datos se relaciona con la valoración del estudio del rendimiento académico.
- c) Determinar si la funcionalidad del modelo basado en el análisis multivariado de datos se relaciona con la valoración del estudio del rendimiento académico.
- d) Determinar si la capacidad explicativa de los gráficos que genera el modelo basado en el análisis multivariado de datos se relaciona con la valoración del estudio del rendimiento académico.

1.8 Hipótesis

1.8.1. Hipótesis General

Las diferentes dimensiones del modelo basado en el análisis multivariado de datos se relacionan con la valoración del estudio del rendimiento académico de las escuelas profesionales de las universidades públicas de la región del Cusco.

1.8.2. Hipótesis Específicas

- a) La validez de los datos de entrada del modelo basado en el análisis multivariado de datos se relaciona con la valoración del estudio del rendimiento académico.
- b) La usabilidad del modelo basado en el análisis multivariado de datos se relaciona con la valoración del estudio del rendimiento académico.
- c) La funcionalidad del modelo basado en el análisis multivariado de datos se relaciona con la valoración del estudio del rendimiento académico.
- d) La capacidad explicativa de los gráficos del modelo basado en el análisis multivariado de datos se relaciona con la valoración del estudio del rendimiento académico.

II. MARCO TEÓRICO

2.1 Marco conceptual

2.1.1 *Análisis multivariante de datos*

2.1.1.1 *Análisis de datos*

El análisis de datos es el proceso de transformar un conjunto de datos con el fin de extraer información útil, facilitar la formulación de conclusiones y, de esta manera, apoyar la toma de decisiones. Se trata de procesar la información con el fin de describir las características más importantes de un conjunto de datos.

Los datos se representan en una matriz X de dimensiones $n \times p$ que representan a dos variables: la primera se representa por filas (individuos) y tiene n categorías y la segunda por columnas (variables) y tiene p categorías. Cada una de estas p variables se denomina una variable escalar o univariante y el conjunto de las p variables forman una variable vectorial o multivariante. El elemento genérico x_{ij} de la matriz X representa el valor de la variable escalar j sobre el individuo i .

2.1.1.2 *Análisis multivariante de datos*

En palabras de Cuadras (1996) “el análisis multivariante es la rama de la estadística y del análisis de datos que estudia, interpreta, y elabora el material estadístico sobre un conjunto de $n > 1$ variables que pueden ser cuantitativas, cualitativas, o una mezcla”

Los individuos (u objetos estudiados) son medidos a través de múltiples variables. El análisis multivariante procesa dichas variables con el fin de tener un mejor entendimiento del fenómeno u objeto de estudio. Según Daniel Peña (2002) el análisis multivariante pretende los siguientes objetivos:

- Resumir el conjunto de variables en unas pocas nuevas variables, construidas como transformaciones de las originales, con la mínima pérdida de información.
- Encontrar grupos en los datos, si existen.
- Clasificar nuevas observaciones en grupos definidos.
- Relacionar dos conjuntos de variables.
- El análisis multivariante puede plantearse a dos niveles (Ayunga, 2010):
- A nivel exploratorio: cuando se desea extraer la información que contiene los datos.
- A nivel inferencial: Cuando se busca obtener conclusiones sobre la población que ha generado los datos, lo que requiere construir un modelo que explique su obtención y permita prever valores futuros.

El primer paso en el análisis de datos multivariantes es describir cada variable y comprender la estructura de dependencia que existe entre ellas. En seguida se debe condensar la información construyendo nuevas variables que sintetizen la información contenida en los datos originales. Los componentes principales nos indican las dimensiones necesarias para representar adecuadamente los datos.

Si los datos se pueden representar en forma de una tabla de contingencias, entonces es posible aplicar lo que se conoce con el nombre de Análisis de Correspondencias, que permite, además, cuantificar de manera objetiva atributos cualitativos.

También es posible averiguar si el conjunto de datos de la muestra forma un grupo homogéneo o no. Las técnicas de agrupamiento (clusterización) permiten clasificar los elementos en grupos, además de identificar qué individuos forman parte de cada uno de ellos.

2.1.1.3 Análisis de correspondencias

Es una técnica descriptiva o exploratoria cuyo objetivo es resumir una gran cantidad de datos en un número reducido de dimensiones, con la menor pérdida de información posible (De la Fuente, 2011).

Se utiliza para analizar, desde un punto de vista gráfico, las relaciones de dependencia e independencia de un conjunto de variables categóricas a partir de los datos de una tabla de contingencia.

Para ello asocia a cada una de las modalidades de la tabla, un punto en el espacio R^n (habitualmente $n=2$) de forma que las relaciones de cercanía/lejanía entre los puntos calculados reflejen las relaciones de dependencia y semejanza existentes entre ellas (Salvador, 2003).

El Análisis de Correspondencias tiene dos objetivos básicos:

- Asociación entre categorías de columnas o filas: Medir la asociación de solo una fila o columna, para ver, por ejemplo, si las modalidades de una variable pueden ser combinadas.
- Asociación entre categorías de filas y columnas: Estudiar si existe relación entre categorías de las filas y columnas.

Se distingue habitualmente el Análisis de correspondencias simples del Análisis de correspondencias múltiples, según sean dos variables cualitativas estudiadas (Análisis de correspondencias simple) o más de dos (Análisis de correspondencias múltiple).

El análisis de correspondencias toma como punto de partida las tablas de contingencia, es decir, tablas donde se recogen las frecuencias de aparición de dos o más variables cualitativas en un conjunto de elementos. La información de partida es una matriz de dimensiones $I \times J$, que representa las frecuencias absolutas observadas de dos variables cualitativas en n elementos. La primera variable se representa por filas, y toma I valores posibles, y la segunda se representa por columnas, y toma J valores posibles (Peña, 2002)

El análisis de correspondencias consiste en resumir la información presente en las filas y columnas de manera que pueda proyectarse sobre un subespacio reducido, y representarse simultáneamente los puntos fila y los puntos columna, pudiéndose obtener conclusiones sobre relaciones entre las dos variables nominales u ordinales de origen

2.1.1.4 *Análisis de componentes principales*

El Análisis de Componentes Principales (ACP) es una técnica estadística de síntesis de la información, o reducción de la dimensión (número de variables). Es decir, ante un banco de datos con muchas variables, el objetivo será reducirlas a un menor número perdiendo la menor cantidad de información posible. Según Daniel Peña (2002) "...con variables con alta dependencia es frecuente que un pequeño número de nuevas variables (menos del 20% de las originales) expliquen la mayor parte (más del 80%) de la variabilidad original".

El mismo autor refiere que el ACP tiene una doble utilidad, a saber:

1. Permite representar óptimamente en un espacio de dimensión pequeña, observaciones de un espacio general p -dimensional. En este sentido componentes principales es el primer paso para identificar posibles variables "latentes" o no observadas, que están generando la variabilidad de los datos.
2. Permite transformar las variables originales, en general correladas, en nuevas variables incorreladas, facilitando la interpretación de los datos.

Tanto el Análisis de Correspondencias como el Análisis de Componentes Principales son dos tipos de análisis factoriales. Se diferencian en la métrica usada, adaptada al tipo de datos (en correspondencias son frecuencias y la métrica es el chi cuadrado; en componentes principales es la distancia euclideana habitual cuadrática).

El objetivo inicial del ACP es la representación en dos o tres dimensiones de unos puntos que originariamente están en muchas dimensiones. Como consecuencia de su propio

procedimiento, el ACP consigue crear unos objetos matemáticos denominados componentes principales o factores que vienen a ser una combinación lineal de las variables originales, y además son incorreladas linealmente entre sí. Estos factores son importantes porque nos permiten establecer relaciones entre las variables, ver cómo se asocian, cómo se distancian, etc.

Reducir la dimensionalidad de un problema implica seleccionar las componentes que poseen la mayor cantidad de información. El concepto de mayor información se relaciona con el de mayor variabilidad o varianza. Cuanto mayor sea la variabilidad de los datos (varianza) se considera que existe mayor información.

Del total de factores se elegirán aquellos que recojan el porcentaje de variabilidad que se considere suficiente. A éstos se les denominará componentes principales.

2.1.1.5 Análisis de conglomerados

El Análisis de Conglomerados, también conocido como Análisis de Clusters, es una técnica estadística multivariante cuya finalidad es dividir un conjunto de objetos en grupos (cluster en inglés) de forma que los perfiles de los objetos en un mismo grupo sean muy similares entre sí (cohesión interna del grupo) y los de los objetos de clusters diferentes sean distintos (aislamiento externo del grupo).

El número posible de combinaciones de grupos y de elementos que integran los posibles grupos se hace intratable desde el punto de vista computacional, aún con un número escaso de observaciones. Se hace necesario, pues, encontrar métodos o algoritmos que infieran el número y componentes de los conglomerados más aceptable, aunque no sea el óptimo absoluto. Para ello es necesario considerar el concepto de medida de similitud/disimilitud.

Las medidas de proximidad, similitud o semejanza (por ejemplo, el coeficiente de correlación) miden el grado de semejanza entre dos objetos de forma que, cuanto mayor es su

valor, mayor es el grado de similaridad existente entre ellos y con más probabilidad los métodos de clasificación tenderán a ponerlos en el mismo grupo.

Las medidas de disimilitud, desemejanza o distancia (por ejemplo, la distancia euclídea) miden la distancia entre dos objetos de forma que, cuanto mayor sea su valor, más diferentes son los objetos y menor la probabilidad de que los métodos de clasificación los pongan en el mismo grupo.

2.1.1.6 Algoritmos de formación de conglomerados

Los algoritmos de formación de conglomerados se agrupan en dos categorías:

1. Algoritmos no jerárquicos (de partición): Método de dividir el conjunto de observaciones en k conglomerados (clusters), en donde k lo define inicialmente el usuario.
2. Algoritmos jerárquicos: Método que entrega una jerarquía de divisiones del conjunto de elementos en conglomerados.
 - Un método jerárquico asociativo parte con una situación en que cada observación forma un conglomerado y en sucesivos pasos se van uniendo, hasta que finalmente todas las situaciones están en un único conglomerado.
 - Un método jerárquico disociativo sigue el sentido inverso, parte de un gran conglomerado y en pasos sucesivos se va dividiendo hasta que cada observación queda en un conglomerado distinto.

En la Figura 1 se puede observar la clasificación de los métodos de formación de conglomerados:



Gráfico 1. Clasificación de los algoritmos de formación de conglomerados
Fuente: De la Fuente (2011)

2.1.1.6.1 *El dendograma*

Un dendograma es una representación gráfica en forma de árbol (Dendro=árbol) que resume el proceso de agrupación en un análisis de clusters. Este tipo de representación permite apreciar claramente las relaciones de agrupación entre los datos e incluso entre grupos de ellos, aunque no las relaciones de similitud o cercanía entre categorías

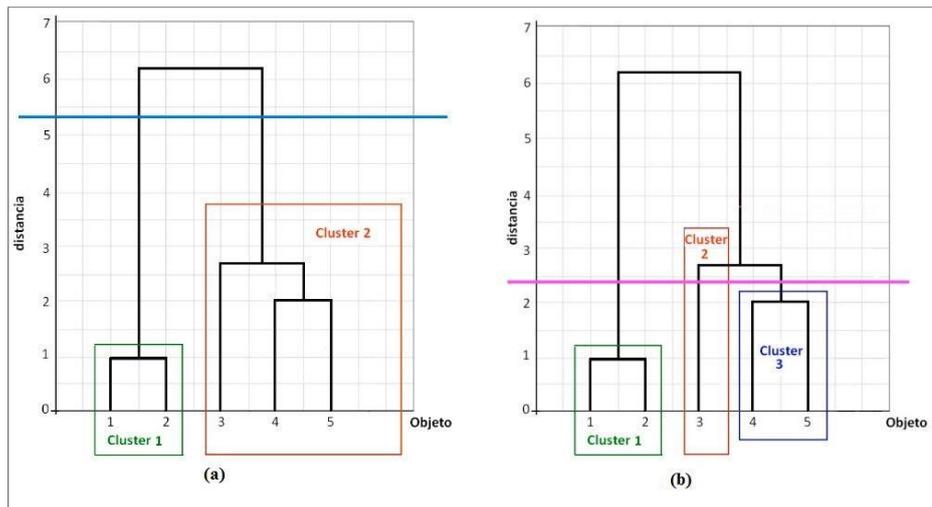


Gráfico 2. Diferentes cortes en un dendrograma
 (a) Corte que produce dos clusters (b) Corte que produce 3 clusters
 Fuente: De la Fuente (2011)

Los objetos similares se conectan mediante enlaces cuya posición en el diagrama está determinada por el nivel de similitud/disimilitud entre los objetos.

El número de clúster puede determinarse trazando una línea horizontal en algún punto del eje de las distancias, tal como se observa en la Figura 2.

La decisión sobre el número óptimo de clusters es subjetiva, especialmente cuando se incrementa el número de objetos pues si se seleccionan pocos, los clusters resultantes son heterogéneos y artificiales, mientras que, si se seleccionan demasiados, la interpretación de los mismos suele resultar complicada.

2.1.1.7 *Coordenadas paralelas*

Las Coordenadas Paralelas son un sistema de visualización que permite representar n-dimensiones en un sistema bidimensional. En este sistema, cada eje vertical (ordenada) representa un atributo (dimensión) que puede ser continuo o categórico. Cada uno de los ejes verticales de un sistema de Coordenadas Paralelas puede tener su propia escala o definirse todos con una sola escala. La primera forma nos permite la visualización de hipersuperficies y

el análisis del comportamiento del conjunto de datos, con la segunda podemos hacer un análisis de las relaciones entre las variables (Araya, 2011).

En general, las Coordenadas Paralelas son una técnica de visualización donde las dimensiones son simbolizadas como una serie de ejes paralelos perpendiculares, con la misma separación entre ellos (equidistantes), en el que se encuentran representados los valores. Cada eje representa una coordenada en la dimensión correspondiente. Uniendo con líneas los ejes, podemos simbolizar los puntos en n-dimensiones. Asimismo, un punto en un espacio n-dimensional es transformado en una línea poligonal a través de n ejes paralelos como segmentos de línea. De tal forma, el vector $x = [x_1, x_2, \dots, x_n]$, es representado por medio de x_1 en la coordenada 1, x_2 en la coordenada 2 y así sucesivamente, hasta x_n en la coordenada n. A partir de la representación resultante, podemos sacar conclusiones al respecto, por ejemplo, sobre la relación entre las variables. Un grupo de líneas proyectas bastante próximas una con otra nos indicará un grado de asociación positiva entre las variables que la componen. La **¡Error! No se encuentra el origen de la referencia.** nos muestra un ejemplo de la representación de múltiples variables en coordenadas paralelas.

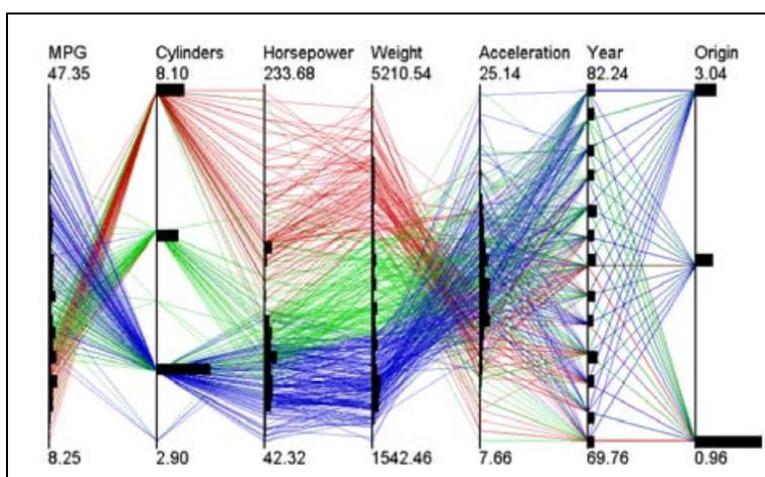


Gráfico 3. Visualización clásica de un clúster mediante coordenadas paralelas
Fuente: McDonnell y Mueller (2008)

2.1.1.8 Algoritmos de clasificación supervisada

2.1.1.8.1 Árboles de decisión

Los métodos basados en árboles (o árboles de decisión) son bastante populares en data mining, pudiéndose usar para clasificación y regresión. Son útiles para la exploración inicial de datos y apropiados cuando hay un número elevado de datos, y existe incertidumbre sobre la manera en que las variables explicativas deberían introducirse en el modelo.

Un árbol de decisión tiene unas entradas las cuales pueden ser un objeto o una situación descrita por medio de un conjunto de atributos y a partir de esto devuelve una respuesta la cual en últimas es una decisión que es tomada a partir de las entradas. Los valores que pueden tomar las entradas y las salidas pueden ser valores discretos o continuos. Cuando se utilizan valores discretos en las funciones de una aplicación se denomina clasificación y cuando se utilizan los continuos se denomina regresión.

2.1.1.8.2 Árboles de clasificación

La tarea de aprendizaje para la cual los árboles de decisión se adecuan mejor es la clasificación. Debido al hecho que la clasificación trata con clases o etiquetas disjuntas, un árbol de decisión conducirá un ejemplo hasta una y sólo una hoja, asignando por tanto una única clase al ejemplo (Hernández, Ramírez y Ferri, 2004).

Los árboles de clasificación se construyen de arriba hacia abajo. Se parte de un nodo raíz añadiendo particiones (que constituyen los hijos del nodo partido). En cada partición los ejemplos se van dividiendo entre los hijos. Finalmente, se llega al punto en que todos los ejemplos que caen en los nodos inferiores son de la misma clase y esa rama ya no sigue creciendo.

Los puntos más importantes para obtener un buen árbol de clasificación son: las particiones a considerar y el criterio de selección de particiones. Esto es lo que diferencia a los distintos

algoritmos de partición existentes como CART, ID3, ID4, ID5, C4.5, C4.8, C5.0, ASSISTANT, etc.

2.1.1.8.3 Árboles de regresión

Los árboles de decisión se han adaptado para tareas como la regresión, el agrupamiento o la estimación de probabilidades.

Un árbol de regresión se construye de manera similar a un árbol de clasificación, pero con las siguientes diferencias (Hernández, Ramírez y Ferri, 2004):

- La función aprendida tiene dominio real y no discreto, como en los clasificadores.
- Los nodos hoja del árbol se etiquetan con valores reales, de tal forma que una cierta medida de calidad se maximice, por ejemplo, la varianza de los ejemplos que caen en ese nodo respecto al valor asignado.

Entre los árboles de regresión más populares se tienen a CART, M5, M5', etc.

2.1.1.9 K-Vecinos más cercanos

El algoritmo de clasificación k-NN (K- Nearest Neighbour) es un clasificador supervisado basado en criterios de vecindad. Parte de la idea de que una nueva muestra será clasificada en la clase más frecuente a la que pertenecen sus K vecinos más cercanos.

Este es un método de clasificación no paramétrico, que estima el valor de la función de densidad de probabilidad o directamente la probabilidad a posteriori de que un elemento x pertenezca a la clase C_j a partir de la información proporcionada por el conjunto de prototipos. En el proceso de aprendizaje no se hace ninguna suposición acerca de la distribución de las variables predictoras.

El algoritmo básico de los vecinos más cercanos lo podemos apreciar en la Figura 4:

```

COMIENZO
Entrada:  $D = \{(x_1, c_1), \dots, (x_N, c_N)\}$ 
 $x = (x_1, \dots, x_n)$  nuevo caso a clasificar
PARA todo objeto ya clasificado  $(x_i, c_i)$ 
    calcular  $d_i = d(x_i, x)$ 
Ordenar  $d_i (i = 1, \dots, N)$  en orden ascendente
Quedarnos con los  $K$  casos  $D_x^K$  ya clasificados más cercanos a  $x$ 
Asignar a  $x$  la clase más frecuente en  $D_x^K$ 
FIN

```

Gráfico 4. Pseudocódigo para el clasificador k-NN

Fuente: <http://www.dia.fi.upm.es/~concha/B.3-knn.pdf>

Existen múltiples variantes del algoritmo básico como, por ejemplo:

- K-NN con rechazo
- K-NN con distancia media
- K-NN con distancia mínima
- K-NN con pesado de vecino
- K-NN con pesado de variables

Escoger el mejor valor de k es importante. En el ejemplo de la Figura 5 se desea clasificar el círculo verde. Para $k = 3$ éste es clasificado con la clase triángulo, ya que hay solo un cuadrado y 2 triángulos dentro del círculo que los contiene. Si $k = 5$ este es clasificado con la clase cuadrado, ya que hay 2 triángulos y 3 cuadrados, dentro del círculo externo.

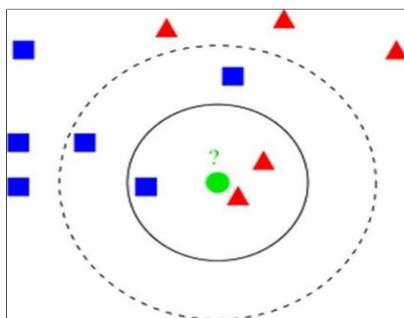


Gráfico 5. Ejemplo de clasificación mediante k-NN

Fuente: <http://es.wikipedia.org/wiki/Knn>

Si K es muy pequeño el modelo será muy sensitivo a puntos que son atípicos o que son ruido (datos corruptos); Si K es muy grande, el modelo tiende a asignar siempre a la clase más grande.

Cabe indicar que, si la variable fuera cuantitativa, el algoritmo asigna al nuevo elemento la media de los k vecinos más cercanos.

2.1.2 Estudio del rendimiento académico

2.1.2.1 Definición y aspectos relacionados al rendimiento académico

Existen múltiples definiciones sobre lo que es, o debería ser, el rendimiento académico. En este punto citaremos algunas de ellas.

Pizarro (1985) conceptualiza el rendimiento académico como una “medida de las capacidades respondientes o indicativas que manifiesta, en forma estimativa, lo que una persona ha aprendido como consecuencia de un proceso de instrucción o formación”.

Novárez (1986) sostiene que el rendimiento académico es el “quantum obtenido por el individuo en determinada actividad académica”.

Martínez-Otero (2002) define el rendimiento académico como “el producto que da el alumnado en los centros de enseñanza y que habitualmente se expresa a través de las calificaciones”.

Mientras tanto, otros autores como Páez (1987) señala que el rendimiento académico “es el grado en que cada estudiante ha alcanzado los objetivos propuestos y las condiciones bajo las cuales se produjo ese logro”.

Se han establecido un sinnúmero de indicadores que permitan medir el rendimiento académico, desde el individual, hasta el grupal. Algunos de ellos son:

- Calificaciones por materia.
- Media de calificaciones.
- Créditos acumulados

- Tasa de rendimiento
- Índice de deserción,
- Número de intentos antes de aprobar,
- Número de egresados
- Número de graduados
- Otros.

La materia prima para la presente investigación es el indicador tradicional más utilizado: las calificaciones por materia.

2.1.2.2 *Estudios sobre el rendimiento académico*

Se han hecho un sinnúmero de estudios respecto al rendimiento académico. Aquellos que estudian los factores asociados al rendimiento académico (sociales, económicos, psicológicos, ambientales, pedagógicos, etc.), otros que intentan afirmar el concepto de rendimiento académico como un conjunto de variables complejas que van más allá de una simple calificación (y estudian y analizan cada una de esas variables componentes) y, aquellos que ponen el rendimiento académico como punto de partida para estudiar sus efectos en, por ejemplo, el campo laboral.

Para la presente tesis conceptualizaremos el estudio del rendimiento académico como el conjunto de trabajos en los que se estudia, se investiga o reflexiona sobre el rendimiento académico, sus indicadores y /o variables asociadas.

2.1.2.3 *El análisis multivariado de datos y el estudio del rendimiento académico*

Los estudios mencionados anteriormente se realizan apoyados en múltiples técnicas estadísticas que buscan confirmar la relación entre distintas variables con el rendimiento académico.

Es recientemente, con el desarrollo de técnicas de análisis de datos e inteligencia artificial, y con la creciente capacidad de los computadores para procesar grandes volúmenes de información, que se busca encontrar patrones ocultos dentro de los datos ya existentes en las organizaciones. En el caso de la educación superior, la existencia de calificaciones registradas en las bases de datos de universidades, son un excelente insumo para describir la realidad de una escuela profesional en base al rendimiento académico de sus estudiantes.

La aplicación de modelos complejos como el Anàlisis Multivariado de Datos podría responder preguntas como ¿Cómo se agrupan las asignaturas de acuerdo a las calificaciones obtenidas por el estudiante? ¿Existen asignaturas que tengan un comportamiento diferente del resto en cuanto a las calificaciones de los estudiantes? ¿Cuáles son las asignaturas que son capaces de explicar de mejor manera el rendimiento académico alto/bajo de los estudiantes?

Las respuestas a estas interrogantes pueden ser bastante ilustrativas y útiles para la toma de decisiones de aquellos que están encargados de tan delicado rol. Desde ese punto de vista, la formalización de un modelo de esta naturaleza aportaría información útil al ya basto estudio del rendimiento académico.

2.1.2.4 Aspectos a medir en el estudio del rendimiento académico

La valoración del impacto del modelo sobre el estudio del rendimiento académicos se hace en base a los siguientes aspectos:

- Capacidad del modelo de brindar información de las asignaturas que imparte un docente.
- Concordancia de los resultados obtenidos por el modelo con la experiencia empírica del docente.
- La valoración del encuestado sobre los resultados del modelo frente a los recursos que tiene a su disposición.

- Valoración de la relevancia de la información obtenida mediante el modelo para la toma de decisiones
- Valoración del modelo del encuestado para recomendar su implementación en las unidades académicas de su universidad.

2.1.3 Un modelo basado en el análisis multivariado para el estudio del rendimiento académico

Vista general del modelo

El modelo basado en el Análisis Multivariado de Datos contará con una etapa previa orientada a la correcta preparación de los datos y 6 subprocesos encargados de procesar la información y mostrar los resultados adecuadamente. Este proceso se puede visualizar en la siguiente figura:

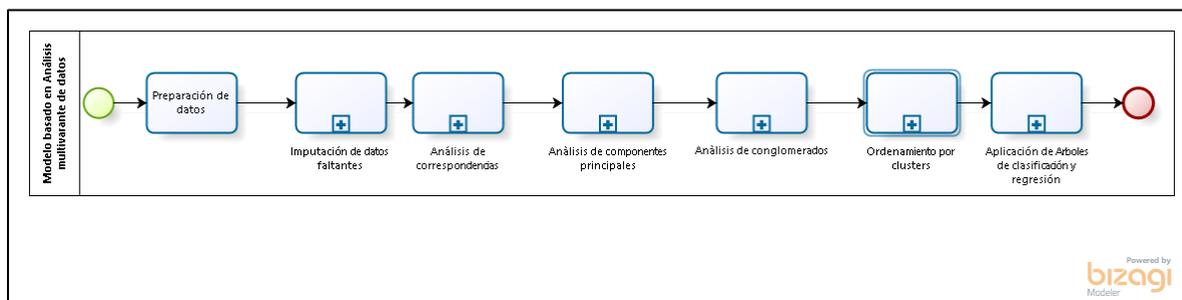


Gráfico 6. Diagrama BPM de la vista general del modelo
Fuente: Elaboración propia

1. Fase inicial: Preparación de datos

1.1. Descripción

La preparación de los datos es una etapa fundamental dentro del modelo. Una preparación y un preprocesamiento adecuados de los datos permitirán que el modelo funcione de manera correcta, y que los resultados sean de utilidad.

Para ello se requiere de un conjunto de datos que sirvan de materia prima para el estudio del rendimiento académico. Para el presente trabajo, los datos a considerar serán las calificaciones

de aquellas materias de corte obligatorio que hayan sido cursados por los egresados de la escuela profesional, preparados dentro de una hoja Excel en un formato de tabla cruzada, en la cual, las columnas corresponderán a las asignaturas (variables), las filas a los estudiantes (casos) y las celdas de intersección a las calificaciones obtenidas en un espacio de tiempo determinado. Se consideran sólo las calificaciones aprobadas por los egresados. Estos datos deben corresponder a un periodo de tiempo en las cuales las condiciones de las calificaciones puedan ser consideradas homogéneas (el mismo plan de estudios, la misma nota mínima de aprobación, entre otros). En caso de tener valores perdidos dentro de la matriz, se los considerará con el valor de NULL.

El motivo de esta caracterización es contar con un conjunto de datos que sean uniformes como materia prima del modelo y que provean de unos resultados interpretables de manera coherente con los datos iniciales.

1.2. Secuencia

Los datos iniciales deben ser proporcionados por la instancia académica-administrativa competente de la universidad (Centro de Cómputos o su equivalente) en formato Excel. Este documento debe contar con dos hojas, la primera de las cuales contendrá las calificaciones en formato de tabla cruzada (asignaturas x estudiantes), y la segunda, la descripción de las asignaturas (como mínimo código, nombre y créditos).

Un ejemplo de ambas hojas del libro Excel se puede apreciar en el siguiente gráfico:

	A	B	C	D	E	F	G
1	ALUMNO	IF411	FI357	IF425	EC171	ME305	IF410
2	10003	11	12	13	11	11	
3	10013	13	13	12	13	18	
4	10026	13	11	12	11	NULL	
5	10028	14	12	11	12	11	
6	10085	13	12	11	12	12	
7	10090	13	11	13	14	11	
8	10137	13	12	14	13	14	
9	10164	11	11	12	12	11	
10	10181	15	11	13	15	11	
11	102	11	11	12	13	11	
12	10211	12	13	15	13	12	
13	10212	14	15	14	12	14	
14	10227	12	13	11	14	11	
15	10231	15	12	11	12	12	

	A	B	C
1	Curso	Asignatura	Creditos
2	DE151	CONSTITUCION POLITICA DEL PERU Y DERECHOS HUMANOS	2
3	EC171	TEORIA ECONOMICA	3
4	ED153	METODOS Y TECNICAS DE APRENDIZAJE	3
5	FI261	FISICA I	4
6	FI263	FISICA III	4
7	FI264	FISICA IV	4
8	FI357	MICROELECTRONICA	4
9	FP179	FILOSOFIA DE LA CIENCIA	3
10	IF001	SEMINARIO EN INFORMATICA	3
11	IF011	PRACTICAS PRE-PROFESIONALES EN INFORMATICA	3
12	IF053	ACTIVIDADES DE PRODUCCION DE BIENES Y PRESTACION DE SERVICIOS	1
13	IF101	INTRODUCCION A LA INFORMATICA	4

Gráfico 7. Vista parcial de los datos iniciales en hoja de cálculo
Fuente: Elaboración propia

En caso de ser adoptada por la universidad, el modelo debe extraer la información directamente de la base de datos institucional, respetando los formatos establecidos tanto para las calificaciones como para el diccionario de asignaturas.

2. Subproceso: Imputación de datos faltantes

2.1. Descripción

Este subproceso es el encargado de asegurar que los datos a procesar se encuentren completos. De acuerdo a la caracterización de los datos de entrada no deberían presentarse valores faltantes; pero, estos pueden estar presentes debido a diferentes factores propios de la institución como inadecuado sistema de ingreso de calificaciones, fallas técnicas en los repositorios de datos electrónicos, digitalización parcial de las calificaciones, etc. Debido a ello, y para garantizar un adecuado procesamiento de los datos se contempla una etapa de imputación de datos faltantes, con el fin de eliminar los valores faltantes. El método de imputación será el de la media aritmética.

2.2. Secuencia

El proceso parte por importar los datos de la hoja Excel hacia el modelo. Una vez importados y ya sabiendo la existencia de valores nulos en la etapa previa, se debe decidir si es necesario o no realizar la imputación de valores faltantes. Al finalizar el subproceso la información estará completa.

La secuencia se observa en la siguiente figura:

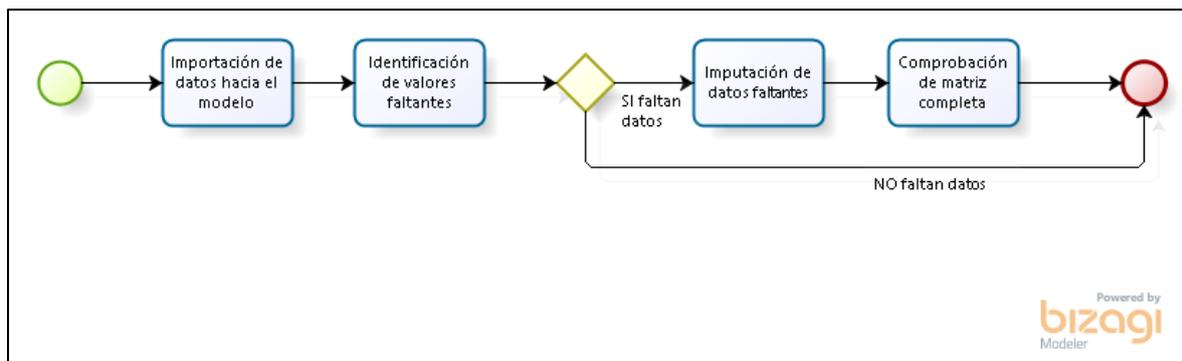


Gráfico 8. Diagrama BPM del subproceso de imputación de datos faltantes
Fuente: Elaboración propia

2.3. Productos para el usuario

El usuario puede comprobar que la tabla de notas no contiene datos perdidos; es decir, los campos marcados como NA han sido reemplazados por calificaciones calculadas mediante la media aritmética. Un ejemplo de ello se aprecia en la siguiente figura:

```

> (Notas.Inicial[c(1:30),c(1:10)])
      IF411 FI357 IF425 EC171 ME305 IF410 IF202 IF053 IF412 LC154
10003  11    12    13    11    11    13    11    18    13    11
10013  13    13    12    13    18    12    13    16    11    13
10026  13    11    12    11    12    12    13    13    11    14
10028  14    12    11    12    11    12    11    16    13    14
10085  13    12    11    12    12    11    12    14    12    11
10090  13    11    13    14    11    12    13    17    14    16
10137  13    12    14    13    14    14    13    15    15    16
10164  11    11    12    12    11    12    13    14    12    11
10181  15    11    13    15    11    11    11    11    12    12
102    11    11    12    13    11    11    12    13    17    11
10211  12    13    15    13    12    12    12    13    13    11
10212  14    15    14    12    14    17    13    16    15    14
10227  12    13    11    14    11    11    12    20    13    12
  
```

Gráfico 9. Vista de calificaciones sin valores perdidos
Fuente: Elaboración propia

2.4. Aspectos adicionales

Existen muchos métodos para realizar la imputación de los valores faltantes. Alzamora (2014) realiza una prueba comparativa entre dos de ellos (medias y KNN) permitió observar que no existe una gran diferencia de imputación de los valores faltantes; es más, a medida que el parámetro N del método KNN subía (3, 6, 9, 12) los valores imputados se parecían más a los calculados por la media aritmética. El gráfico siguiente muestra el resultado de dicha prueba:

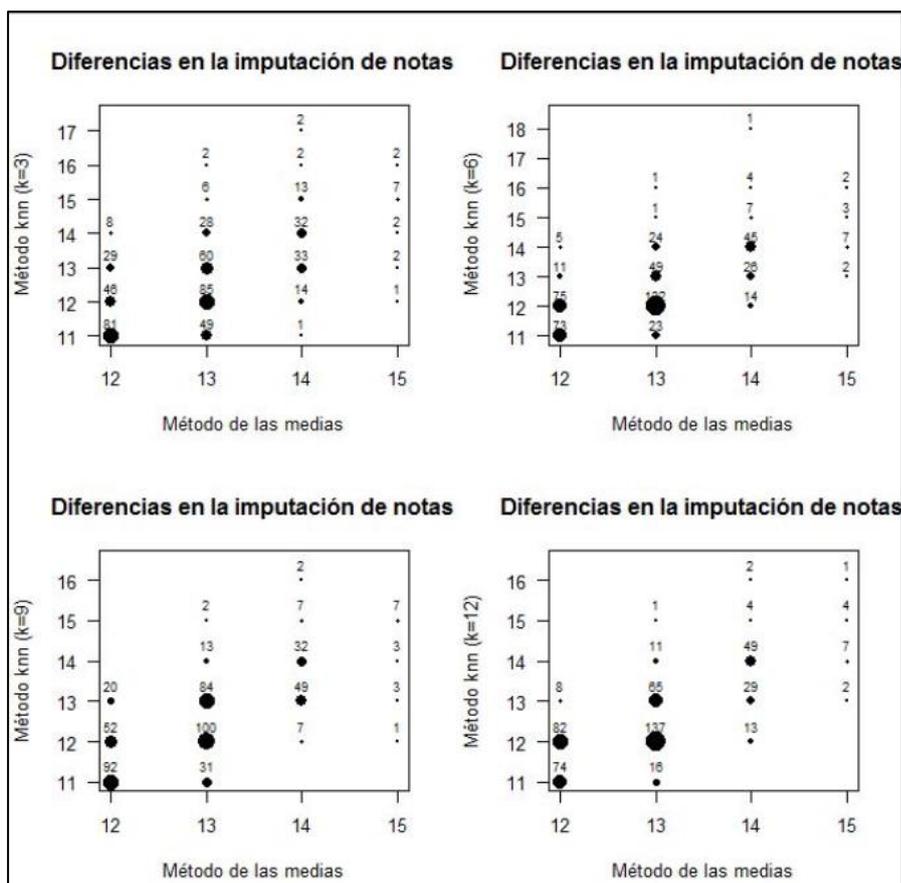


Gráfico 10. Comparación entre métodos de imputación
Fuente: Alzamora (2014)

Dado que el concepto de promedio está relacionado al tema de las calificaciones, el modelo adopta la imputación de datos faltantes por el método de las medias.

3. Subproceso: Análisis de Correspondencias

3.1. Descripción

Esta etapa nos permitirá identificar elementos con comportamiento atípico (outliers), estudiar las inercias existentes entre calificaciones y asignaturas en dos escenarios diferentes: con la presencia de outliers, y sin la presencia de outliers.

Una gráfica que permita observar un orden en base al primer factor nos dará una idea de las asignaturas que presentan mayor cantidad de estudiantes con alto y bajo rendimiento académico.

3.2. Secuencia

Este proceso inicia con la construcción de una tabla de contingencias con el fin de observar el número de ocurrencias de cada calificación (11-20) por cada una de las asignaturas. Esta tabla es la materia prima para realizar el Análisis de Correspondencias.

Al graficar notas y asignaturas en un plano cartesiano en base a los dos primeros componentes del análisis, se podrá visualizar las inercias existentes entre dichos elementos. Esta gráfica permitirá identificar a elementos que tienen un comportamiento diferente que el resto de asignaturas (outliers). Los elementos outliers pueden ser bastante perjudiciales para los algoritmos de clasificación, por lo que, se hace un ejercicio de retirarlos y volver a realizar el Análisis de Correspondencias. Si el gráfico mejora habrá que considerar la posibilidad de eliminarlos definitivamente para continuar con el estudio.

La construcción de un mapa de calor permitirá observar la concentración de estudiantes en cada asignatura relacionada con cada calificación. Este mapa de calor es construido a partir de la matriz reordenada de acuerdo al primer factor del Análisis de Correspondencias.

En este punto del estudio se deberá tomar la decisión de mantener o de eliminar los elementos outlier para continuar con el estudio.

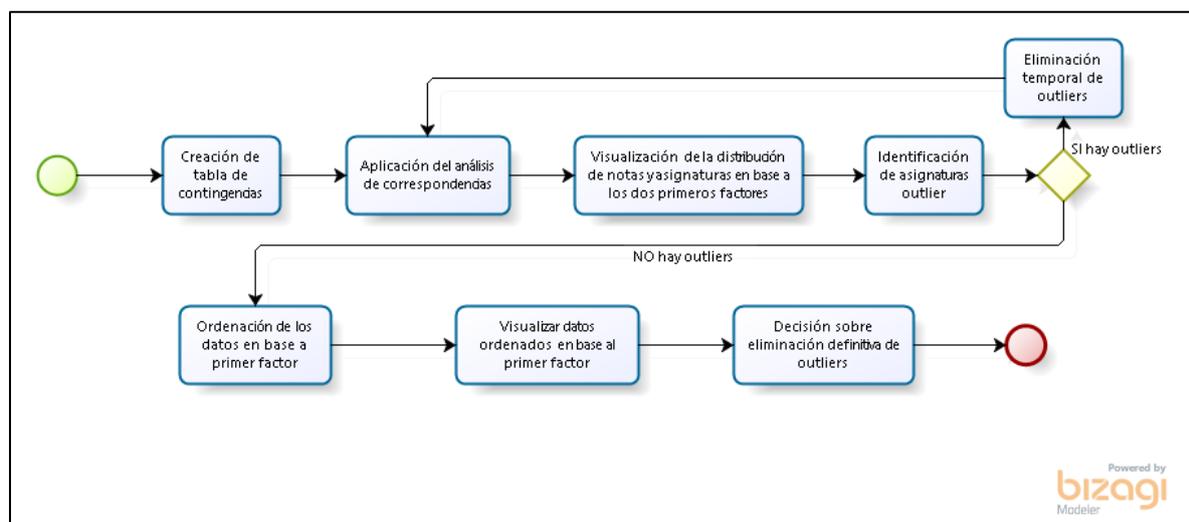


Gráfico 11. Diagrama BPM del subproceso Análisis de Correspondencias

Fuente: Elaboración propia

3.3. Productos para el usuario

El usuario puede visualizar una tabla de contingencias y buscar asignaturas que presenten valores muy altos o muy bajos en alguna de las columnas. Estas asignaturas tendrán especial interés más adelante.

Un ejemplo de la tabla de contingencias se ve en la siguiente figura:

```
> (Notas.Contingencia[,])
      11  12  13  14  15  16  17  18  19  20
IF411 138 152 201 162 103  61  25  14  5  5
FI357 354 236 135  87  40  10  3  1  0  0
IF425 217 208 163 119  72  50  27  10  0  0
EC171 187 222 172 128  90  32  17  10  7  1
ME305 412 240 105  53  35  15  3  3  0  0
IF410 156 186 162 141 104  73  29  9  5  1
IF202 200 225 185 154  75  21  6  0  0  0
IF053  31  34  89 143 113 155  78  20 10 193
IF412 111 134 179 170 114  71  49  27  8  3
LC154 158 166 191 138  97  55  28  17 11  4
IF413  89 150 174 144 130 101  53  15  9  1
IF301 212 218 185 131  65  34  13  4  4  0
FI264 337 241 156  79  37  11  4  1  0  0
ME660 393 214 125  60  37  18  10  2  5  2
ME359 461 198  99  58  23  15  7  5  0  0
ME253 400 217 125  69  35  8  5  6  1  0
FI263 371 245 145  55  27  19  2  1  0  0
IF302 231 253 196 108  57  15  6  0  0  0
```

Gráfico 12. Vista parcial de la tabla de contingencias

Fuente: Elaboración propia

Más adelante podrá visualizar un gráfico en dos dimensiones para analizar las inercias entre calificaciones y asignaturas. Los elementos atípicos (si los hubiese) se verán alejados del conjunto de asignaturas. Luego de eliminar temporalmente dichos elementos podrá visualizar nuevamente el gráfico de inercias y comparar los resultados. Si hubiese otros outliers podrá repetir esta sección tantas veces como sea necesario. Un ejemplo de esta gráfica con/sin outliers se observa en la siguiente ilustración:

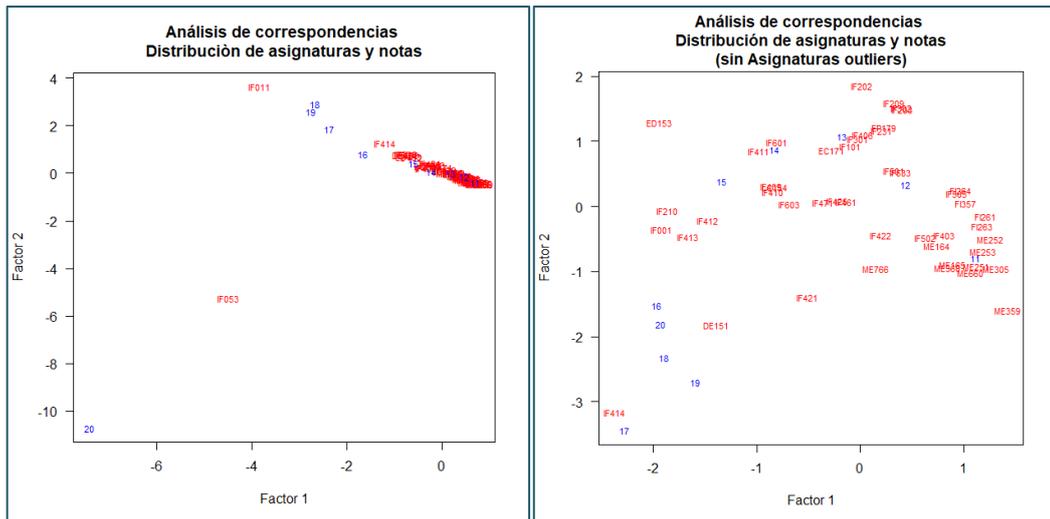


Gráfico 13. Análisis de correspondencias con/sin outliers
Fuente: Elaboración propia

Un mapa de calor (heatmap) permitirá observar un ordenamiento de las asignaturas de acuerdo a la concentración de alumnos en las diferentes calificaciones (1-20), tal como se ve en la siguiente ilustración:

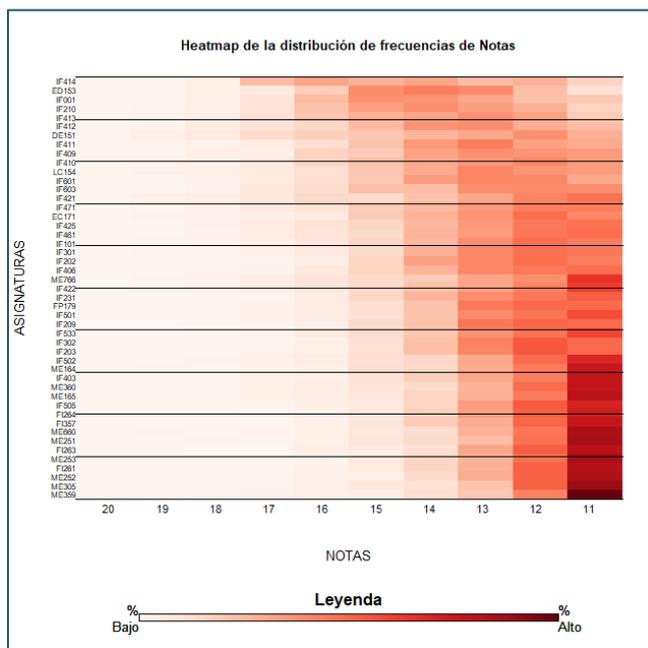


Gráfico 14. Heatmap de la distribución de frecuencia de notas
Fuente: Elaboración propia

3.4. Aspectos adicionales

Alzamora (2014) realiza el Análisis de Correspondencias y evidencia que los dos primeros factores absorben más del 90% de la variabilidad expresada en función de la varianza, tanto en el escenario de análisis con elementos outliers, como en el estudio sin elementos outliers, tal como se aprecia en la siguiente figura:

Principal inertias (eigenvalues):						Principal inertias (eigenvalues):					
	1	2	3	4	5		1	2	3	4	5
Value	0.134079	0.028059	0.004465	0.004232	0.002719	Value	0.115996	0.017639	0.00458	0.002903	0.001942
Percentage	75.28%	15.75%	2.51%	2.38%	1.53%	Percentage	78.73%	11.97%	3.11%	1.97%	1.32%
	6	7	8	9			6	7	8	9	
Value	0.001723	0.001439	0.000991	0.000389		Value	0.001518	0.001266	0.001057	0.000428	
Percentage	0.97%	0.81%	0.56%	0.22%		Percentage	1.03%	0.86%	0.72%	0.29%	

Gráfico 15. Varianza explicada en factores del análisis de correspondencias con/sin elementos outliers
Fuente: Alzamora (2014)

El primer factor es, de lejos, el más significativo y, por lo tanto, es utilizado para realizar el ordenamiento de las asignaturas y generar el mapa de calor.

4. Subproceso: Análisis de componentes principales

4.1. Descripción

Este subproceso permite analizar las correlaciones entre las asignaturas y estudiar el peso de cada una de las componentes logradas tras el análisis. La relación entre el primer componente y el rendimiento académico es el paso más importante ya en base a ella se realiza el resto del estudio.

4.2. Secuencia

Tras aplicar el análisis de correspondencias, y verificar que el primer componente es el predominante, se pasa a visualizar un gráfico que permita comprobar la relación entre rendimiento académico y dicho componente.

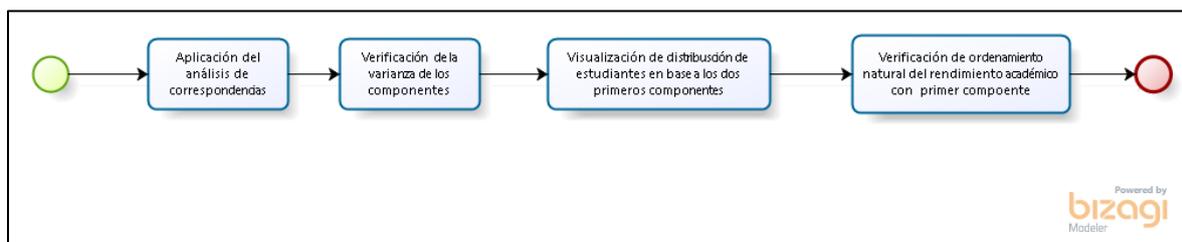


Gráfico 16. Diagrama BPM para el subproceso Análisis de Componentes Principales
Fuente: Elaboración propia

4.3. Productos para el usuario

El usuario puede apreciar una nube de puntos en un plano cartesiano cuyos ejes corresponden a los dos primeros componentes del análisis. El gráfico etiqueta a los estudiantes que ocupan los cuatro extremos del gráfico junto a sus calificaciones con el fin de estudiar la relación de orden entre sus promedios y el primer componente (cosa que no ocurre con el segundo componente). La forma de la nube de puntos también es de interés para los procesos de clasificación de las siguientes etapas. Un ejemplo de este gráfico se ve en la siguiente figura:

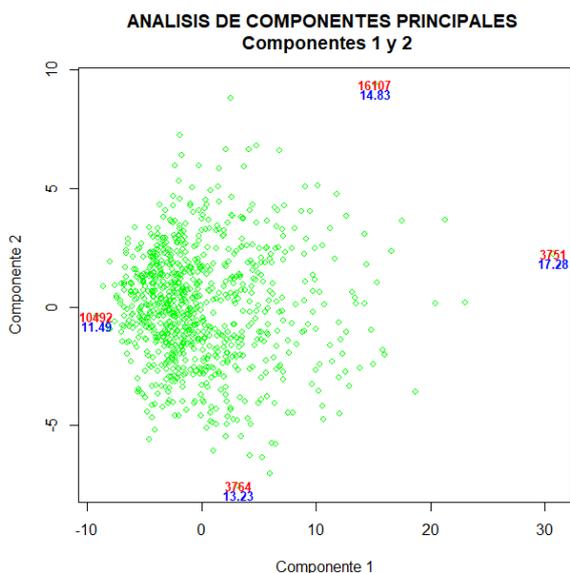


Gráfico 17. Distribución de los estudiantes en base al ACP
Fuente: Elaboración propia

4.4. Aspectos adicionales

Alzamora (2014) realiza un estudio de la varianza de los componentes y muestra claramente que el primer componente es el más importante de todos, tal como se ve en la figura siguiente:

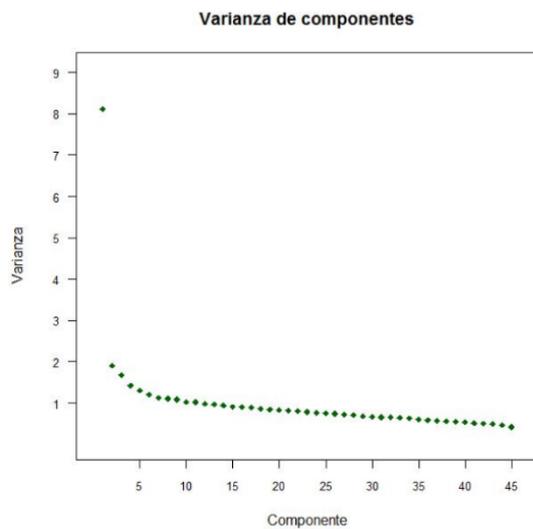


Gráfico 18. Varianza de los componentes obtenidos mediante ACP
Fuente: Alzamora (2014)

Además, también se demuestra la alta relación que existe entre el primer componente y el rendimiento académico, tal como se muestra en la siguiente figura:

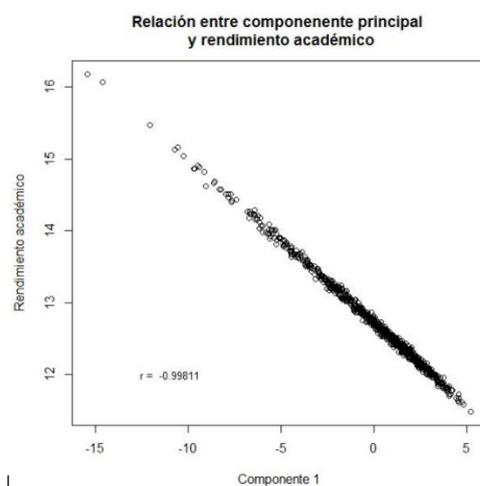


Gráfico 19. Relación entre el primer componente y el rendimiento académico
Fuente: Alzamora (2014)

5. Subproceso: Análisis de conglomerados

5.1. Descripción

Esta etapa del estudio pretende lograr una clasificación de los estudiantes en base a su rendimiento académico, agrupándolos en 4 conglomerados: estudiantes con rendimiento bajo, medio-bajo, medio-alto y alto; y luego, realizar un agrupamiento jerárquico de las asignaturas en 5 clusters.

5.2. Secuencia

Esta etapa es sencilla. El modelo construye conglomerados de estudiantes y asignaturas en base al rendimiento académico y muestra los resultados en formato gráfico. La secuencia se observa en la siguiente figura:

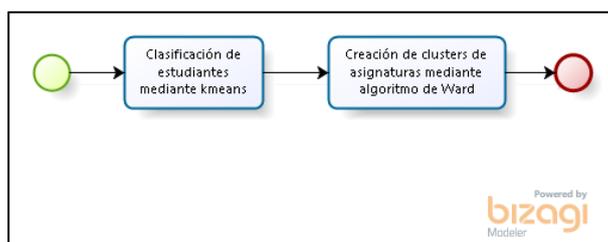


Gráfico 20. Diagrama BPM para el subproceso Análisis de Conglomerados
Fuente: Elaboración propia

5.3. Productos para el usuario

El usuario podrá observar un histograma coloreado con los conglomerados de los estudiantes y la frecuencia máxima de cada uno de los clusters, tal como se observa en la siguiente figura:

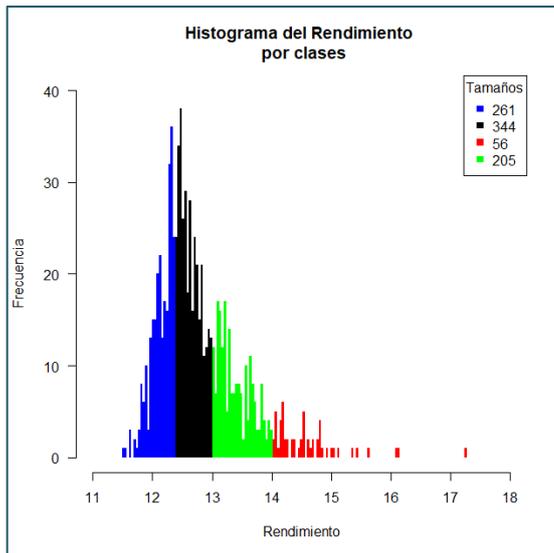


Gráfico 21. Histograma del rendimiento académico por clases de estudiantes
Fuente: Elaboración propia

Luego, podrá visualizar un dendograma con los conglomerados obtenidos mediante el algoritmo de clasificación jerárquico de Ward, tal como se observa en la siguiente figura:

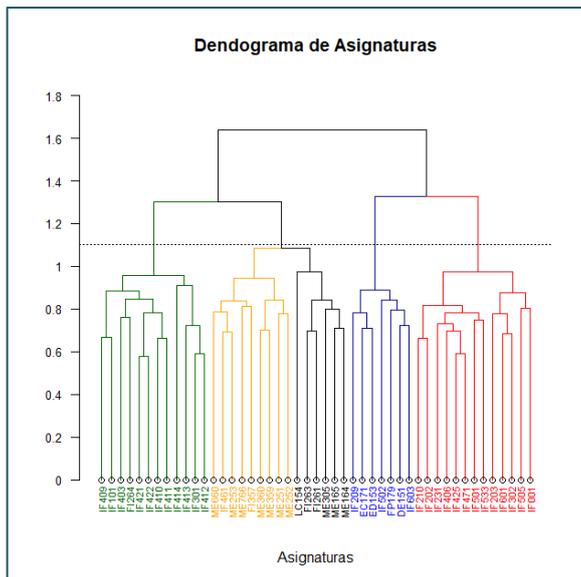


Gráfico 22. Dendograma de asignaturas mediante algoritmo Ward
Fuente: Elaboración propia

5.4. Aspectos adicionales

Alzamora (2014) realiza una validación de los conglomerados a través de gráficos de siluetas tal como se aprecia en la siguiente figura:

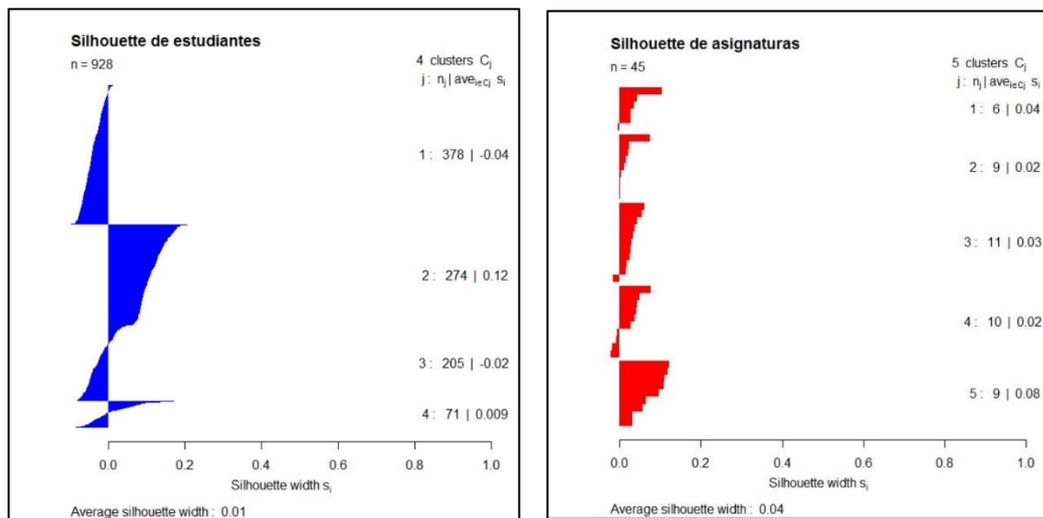


Gráfico 23. Siluetas de los clusters de alumnos y asignaturas
Fuente: Alzamora (2014)

Estas pruebas muestran que la clasificación de estudiantes no es muy clara y que se obtienen mejores resultados en los clusters de asignaturas.

6. Subproceso: Ordenamiento por clusters

6.1. Descripción

El objetivo de esta etapa es ordenar la matriz de notas de acuerdo a los clúster obtenidos en el subproceso anterior y construir un mapa de calor que permita analizar el cruce entre clúster de estudiantes y clúster de asignaturas.

6.2. Secuencia

Primero se ordena la matriz de notas de acuerdo a los clusters de alumnos; esta acción cambia el orden de las filas de la matriz. En el segundo paso se reordena la matriz en función a los clusters de asignaturas obtenidos; esta acción cambia el orden de las columnas de la matriz. Finalmente se visualiza el resultado en un mapa de calor inicial y en un mapa de calor mejorado.

La secuencia se observa en la siguiente figura:

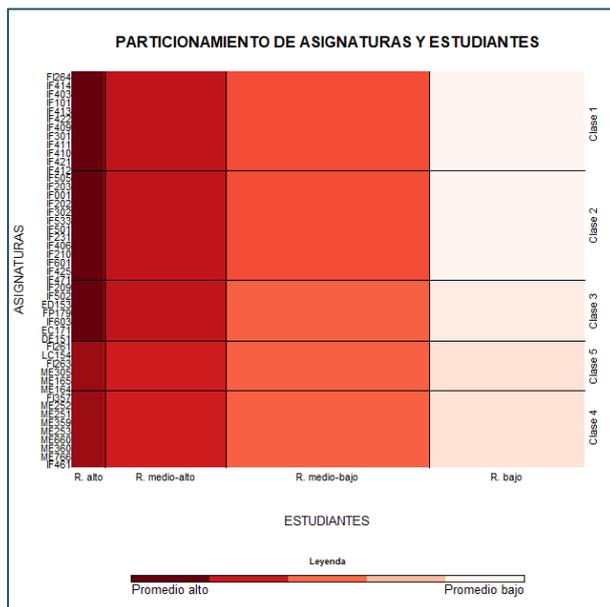


Gráfico 26. Particionamiento de asignaturas y estudiantes considerando el promedio de cada sector
Fuente: Elaboración propia

7.

8. Subproceso: Aplicación de árboles de clasificación y regresión

8.1. Descripción

Esta etapa tiene como finalidad explicar la conformación de los clusters de alumnos y asignaturas haciendo uso de árboles de clasificación y regresión.

8.2. Secuencia

En primer lugar, se crea un árbol de clasificación de los estudiantes en función de las asignaturas con el fin de observar las asignaturas que son más discriminantes al momento de clasificar a un estudiante.

Luego, se crea un árbol de regresión entre las asignaturas y el rendimiento académico con el fin de observar las asignaturas que discriminan mejor el rendimiento académico.

Finalmente, se crea un árbol de clasificación de los estudiantes en función de los conglomerados de las asignaturas con el fin de evaluar la capacidad discriminante de los clusters de asignaturas en la determinación de los clusters de estudiantes.

El proceso se puede observar en el siguiente gráfico:

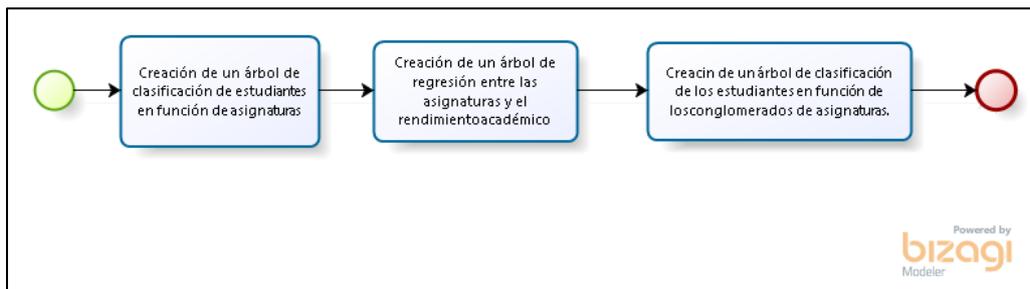


Gráfico 27. Diagrama BPM para el subproceso Aplicación de árboles de clasificación y regresión
Fuente: Elaboración propia

8.3. Productos para el usuario

El usuario puede observar cada uno de los árboles de manera gráfica y analizar las asignaturas que se encuentran en los niveles superiores e inferiores del árbol y contrastarlos con los mapas de calor anteriores. Estos árboles se ven en las siguientes figuras:

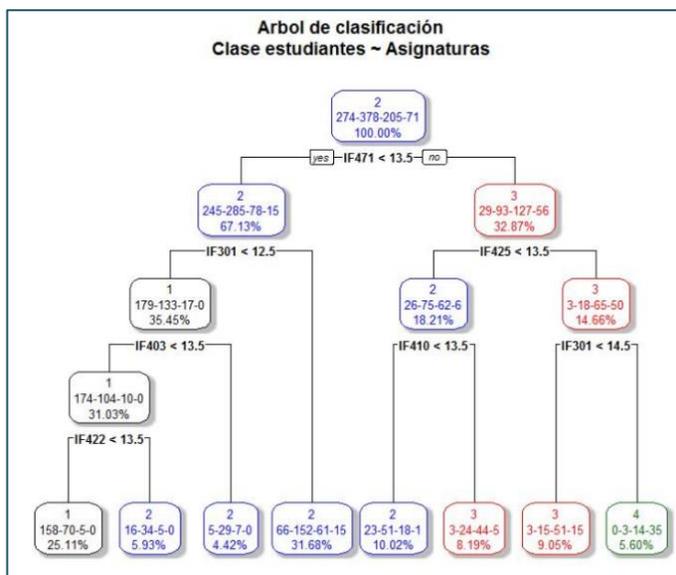


Gráfico 28. Árbol de clasificación de estudiantes en función de asignaturas
Fuente: Elaboración propia

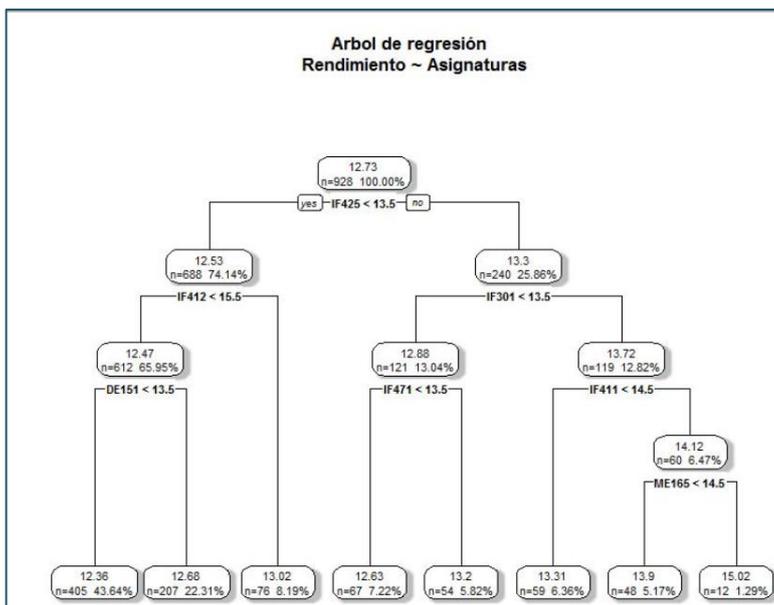


Gráfico 29. Árbol de regresión del rendimiento académico en función de las asignaturas
Fuente: Elaboración propia

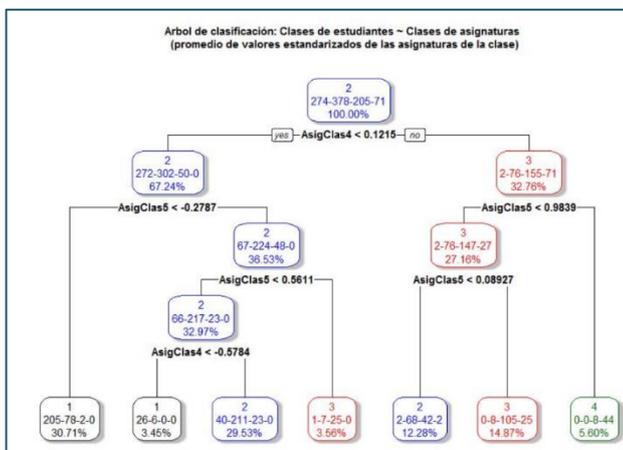


Gráfico 30. Árbol de clasificación de clusters de estudiantes en función de los clusters de las asignaturas
Fuente: Elaboración propia

8.4. Aspectos adicionales

Alzamora (2014) hace estudio mediante un gráfico de coordenadas paralelas y nos muestra que, como se había observado en la nube de puntos del análisis de componentes principales y en los gráficos de siluetas de los conglomerados, existe una sobreposición en la clasificación de estudiantes y que algunos clusters de asignaturas discriminan de mejor manera a los estudiantes que otros. Esto se evidencia en la siguiente gráfica:

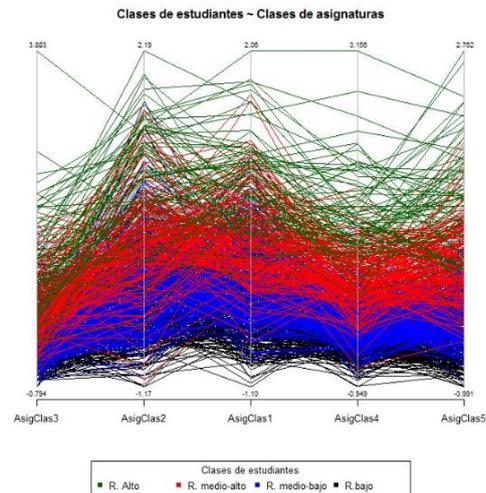


Gráfico 31. Coordenadas paralelas de estudiantes frente a clusters de asignaturas
Fuente: Alzamora (2014)

2.1.4 Aspectos a medir en el modelo

Al no contar con un estándar definido para medir modelos de la naturaleza de este estudio, se determinó tomar algunos aspectos relacionados a la valoración del software, a las características de calidad de los gráficos estadísticos y a la satisfacción del usuario.

Para mostrar la funcionalidad del modelo se programó una secuencia de líneas de código en el programa R, que el usuario puede ejecutar siguiendo un instructivo elaborado para ese fin. Como parte del proceso de valoración del modelo se evalúan aspectos relacionados a estos dos productos mencionados.

2.1.4.1 Validez de datos de entrada

Todo modelo que procesa información requiere de datos de entrada. Estos datos constituyen la base sobre la cual se construye el modelo y del cual depende el correcto funcionamiento del mismo.

La International Organization for Standardization (ISO) define 15 características que se deben evaluar dentro de su modelo de calidad de producto de datos. Estas características se pueden ver en la siguiente figura:

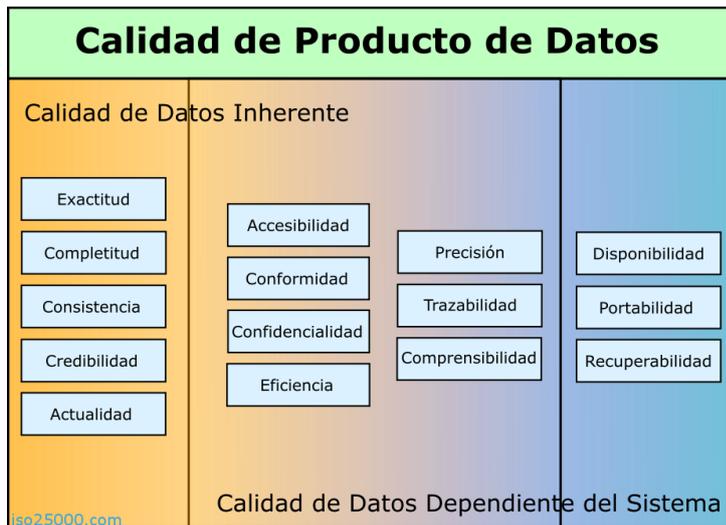


Gráfico 32. Modelo de calidad de producto de Datos de la ISO/IEC 25012

Fuente: ISO/IEC 25012 (s.f.)

La calidad de datos inherente se refiere al potencial propio que tienen los datos para satisfacer las necesidades de un espacio de problema bajo ciertas condiciones. La calidad de datos dependiente del sistema es el grado que alcanzan los datos cuando son utilizados bajo ciertas condiciones por un sistema informático.

“Desde el punto de vista dependiente del sistema, la Calidad de Datos depende del dominio tecnológico en el que los datos se utilizan, y se alcanza mediante las capacidades de los componentes del sistema informático tales como: dispositivos hardware (e.g., Respaldo Software para alcanzar la Recuperabilidad), y otro software” (ISO/IEC 25012, s.f.)

Dado que el modelo presentado en este trabajo trae consigo una implementación del modelo en una herramienta software (sin ser el modelo en sí, un software de usuario final), se tomarán y adaptarán algunos de los conceptos definidos por la ISO/IEC 25012, como son: exactitud, completitud, credibilidad y comprensibilidad.

- Exactitud: está relacionado al hecho de que si los datos representan de manera correcta el valor del atributo (por ejemplo, las calificaciones).
- Completitud: está relacionado a evaluar si todas las entidades poseen todos los valores para cada una de las variables definidas.

- Credibilidad: está relacionado a la valoración si todos los atributos considerados se consideran ciertos y creíbles para el espacio del problema planteado.
- Comprensibilidad: está relacionado al hecho de si los datos poseen “atributos que permiten ser leídos e interpretados por los usuarios y son expresados utilizando lenguajes, símbolos y unidades apropiados en un contexto de uso específico” (ISO/IEC 25012, s.f.)

2.1.4.2 Usabilidad

La usabilidad es un atributo cualitativo relacionado a la facilidad de uso de una herramienta informática que tenga la capacidad de interactuar con un usuario.

Según la ISO “la usabilidad se refiere a la capacidad de un software de ser comprendido, aprendido, usado y ser atractivo para el usuario, en condiciones específicas de uso” (ISO/IEC 25012, s.f.). Esta característica propia del software puede ser evaluada a través de aspectos como: adecuación a las necesidades del usuario, facilidad de aprendizaje, facilidad de uso, protección contra errores, estética de la interfaz, accesibilidad, etc.

Para el presente estudio tomarán en cuenta aspectos como:

- La facilidad del modelo de cargar los datos iniciales.
- La valoración de la parte estética de la interfaz de usuario
- La facilidad de uso
- La valoración que se le da a la documentación para guiar la ejecución del modelo y como apoyo a la interpretación de los resultados.

2.1.4.3 Funcionalidad

La funcionalidad es una característica relacionada a lo que una herramienta es capaz de hacer de manera práctica. La ISO contempla como parte de su modelo de calidad del producto de software la característica de Adecuación Funcional, la cual es definida como “la capacidad

del producto software para proporcionar funciones que satisfacen las necesidades declaradas e implícitas, cuando el producto se usa en las condiciones especificadas” (ISO/IEC 25010, s.f.). Las subcaracterísticas de este aspecto son: completitud, corrección y pertinencia. La primera de ellas se refiere a que el programa debe cubrir todas las necesidades y objetivos del espacio del problema; la segunda, se refiere a que los resultados obtenidos deben ser correctos y con un nivel de precisión adecuado y, la tercera, se refiere a que las funciones utilizadas sean las adecuadas para la consecución de los resultados.

En el presente estudio, se examinarán aspectos como:

- La rapidez del modelo para obtener resultados
- La capacidad del modelo para obtener los mismos resultados siempre que se lo ejecute con el mismo conjunto de datos.
- La idoneidad de los formatos en los que se presentan los resultados.
- La cantidad adecuada de información que se muestra en los resultados.
- La capacidad del modelo de evaluar diferentes situaciones.
- La coherencia del modelo con el eje temático estudiado.

2.1.4.4 *Calidad de gráficos estadísticos*

Reza un dicho popular que “un gráfico vale por mil palabras”. La ciencia no es, ni mucho menos, ajena a este hecho. Los esfuerzos científicos tienen como centro los datos; los investigadores recopilan datos de diferentes formas y en diferentes escalas. Representar adecuadamente esos datos es, y ha sido, de vital importancia para entender adecuadamente la información tratada. Tablas, gráficos y figuras juegan en este escenario un papel preponderante.

Cleveland y McGill (1985) publicaron un artículo en el que ponen en relieve la importancia de utilizar representaciones visuales estadísticas en función del objetivo de cada gráfico que se desee utilizar. Más adelante, Cleveland reflexiona sobre el hecho de escoger apropiadamente

las gráficas que ayuden a mostrar con claridad los resultados de un proceso de análisis de datos (Cleveland, 1993).

Los elementos que, en un primer momento, consideraron Cleveland y McGill son, por ejemplo, la posición de los elementos respecto a las escalas, el área del gráfico, el volumen y la curvatura, la saturación del color, el sombreado entre otros.

Esta escala valorativa ha ido evolucionando con el tiempo y se han ido añadiendo muchos parámetros con el fin de lograr gráficas que sean efectivas al momento de mostrar adecuadamente el resultado del análisis de datos.

En el presente estudio, se utilizarán algunos de estos parámetros con el fin de valorar la capacidad de los resultados gráficos que produce el modelo. Estos parámetros son:

- La capacidad de representar relaciones entre las variables.
- La capacidad de mostrar los elementos de manera clara: los principales y los secundarios.
- La consistencia de los gráficos entre sí.
- La suficiencia de la información mostrada para entender el rendimiento académico.
- La claridad con la que representan los gráficos los datos iniciales.
- La capacidad interpretativa que brindan los gráficos
- La utilidad de los gráficos en cuanto al objeto de estudio
- La validez de las escalas utilizadas.

III. MÉTODO

3.1 Tipo de investigación

De acuerdo a la naturaleza del estudio se considera un enfoque cuantitativo ya que, como indican Hernández, Fernández y Baptista (2014) inicia examinando la teoría para luego formular hipótesis que el investigador someterá a prueba.

El tipo de investigación es aplicada ya que está orientada a resolver problemas; en el caso del presente trabajo, se trata de enriquecer el estudio del rendimiento académico de las escuelas profesionales universitarias a través de técnicas multivariadas y un modelo planteado que facilite tal estudio.

El nivel de investigación es correlacional ya que se trata de vincular dos variables: el modelo basado en el análisis multivariado de datos y el estudio del rendimiento académico, estableciendo para cada una de ellas los indicadores que permitan medirlas adecuadamente y estudiando influencia de la primera sobre la segunda.

El diseño es no experimental ya que, como indica Hernández et al., (2014), no se realizará la manipulación deliberada de la variable independiente con el fin de observar el comportamiento de la variable dependiente.

3.2 Población y muestra

3.2.1. Población

La población objeto de investigación estará conformada por personal académico de la UNSAAC con cargos relacionados a la parte académica (vice rector académico, decanos, secretarios académicos docentes, directores de departamento y directores de escuelas profesionales), ingenieros del área de tecnologías de la información y centro de cómputo que

proveen servicios a las áreas académicas, y profesionales relacionados a la ing. de sistemas e informática que podrían asumir dichos cargos (docentes nombrados de esa especialidad)

La siguiente tabla detalla la población en mención:

FUNCION/CARGO	CANTIDAD
Vicerrector académico	1
Decanos	10
Secretarios académicos docentes de facultad	10
Directores de Escuelas Profesionales	42
Directores de Departamentos académicos	35
Docentes nombrados de Ing. Sistemas	23
Especialistas del Centro de Cómputo	5
Especialistas del Área de TICs	5
TOTAL	131

3.2.2. Muestra

Se utilizará el muestreo probabilístico aleatorio sin reemplazo, considerando los siguientes valores:

Margen de error: 5%

Nivel de confianza: 95%

Población: 131

Tamaño de muestra:

Para el cálculo del tamaño de la muestra se utilizará la siguiente fórmula

$$n = \frac{N \times Z_a^2 \times p \times q}{d^2 \times (N - 1) + Z_a^2 \times p \times q}$$

En donde, N = tamaño de la población Z = nivel de confianza, P = probabilidad de éxito, o proporción esperada Q = probabilidad de fracaso D = precisión (Error máximo admisible en términos de proporción).

Así, para N = 131, Z= 95%, p=q=50%, d=5%

Se obtiene una muestra de n= 54 individuos

3.3 Operacionalización de variables

Tabla

Operacionalización de variables

1

VARIABLES	DIMENSIONES	INDICADORES	ESCALA	MEDIDA
VARIABLE INDEPENDIENTE: Modelo basado en el análisis multivariado de datos	1. Validez de los datos de entrada del modelo	1.1 Facilidad de interpretación de los datos iniciales	Ordinal	Likert
		1.2 Facilidad de interpretar todo el volumen de información inicial	Ordinal	Likert
		1.3 Pertinencia de la información inicial para el estudio del rendimiento académico	Ordinal	Likert
	2. Usabilidad del modelo	2.1 Facilidad en la carga de datos iniciales	Ordinal	Likert
		2.2 Facilidad de uso del modelo	Ordinal	Likert
		2.3 Facilidad de uso de la interfaz	Ordinal	Likert

	2.4	Valoración de la documentación para ejecutar el modelo	Ordinal	Likert
	2.5	Valoración de la documentación para la interpretación de los resultados	Ordinal	Likert
3. Funcionalidad del modelo	3.1	Rapidez en la obtención de resultados	Ordinal	Likert
	3.2	Obtención de los mismos resultados con los mismos datos	Ordinal	Likert
	3.3	Variedad de formatos para analizar la información	Ordinal	Likert
	3.4	Sobrecarga de información	Ordinal	Likert
	3.5	Capacidad de analizar diferentes escenarios	Ordinal	Likert
	3.6	Coherencia del eje temático estudiado	Ordinal	Likert
4. Capacidad explicativa de los gráficos del modelo	4.1	Capacidad de mostrar las relaciones entre alumnos y asignaturas	Ordinal	Likert

4.2	Capacidad de mostrar las relaciones entre asignaturas y calificaciones	Ordinal	Likert
4.3	Identificación clara de los elementos de los gráficos	Ordinal	Likert
4.4	Identificación de los elementos más relevantes	Ordinal	Likert
4.5	Consistencia entre gráficos	Ordinal	Likert
4.6	Suficiencia de la información mostrada para relacionar los gráficos	Ordinal	Likert
4.7	Capacidad de representar los datos originales con claridad	Ordinal	Likert
4.8	Capacidad de los gráficos para generar múltiples interpretaciones	Ordinal	Likert
4.9	Valoración de la utilidad de los gráficos	Ordinal	Likert
4.1	Valoración de los gráficos en relación al rendimiento académico	Ordinal	Likert

		4.11	Presentación adecuada de los gráficos	Ordinal	Likert
		4.12	Valoración de la escala de los gráficos	Ordinal	Likert
VARIABLE DEPENDIENTE: Valoración del estudio del rendimiento académico	5. Valoración del estudio del rendimiento académico	5.1	Valoración de la capacidad de analizar las asignaturas a su cargo.	Ordinal	Likert
		5.2	Concordancia de los resultados con el conocimiento previo del usuario	Ordinal	Likert
		5.3	Valoración del aporte en la comprensión del rendimiento académico de la escuela profesional.	Ordinal	Likert
		5.4	Percepción sobre relevancia de la información	Ordinal	Likert
		5.5	Valoración del modelo para recomendar su implementación en su universidad.	Ordinal	Likert

3.4 Instrumentos

Para el presente estudio se utilizaron los siguientes instrumentos:

- El instrumento de exploración de datos que se empleó para probar el **modelo de análisis multivariado de datos es el lenguaje R**. R es un lenguaje de programación con enfoque en análisis estadístico y que es ampliamente utilizado para la investigación científica en los campos de exploración y análisis de datos, minería de datos, bioinformática, machine learning, deep learning, matemática financiera, etc. Esta herramienta de software libre cuenta con un gran número de librerías que facilitan la exploración y clasificación de datos. La versión a utilizar será R de 64 bits para Windows v. 3.5. Algunas de las librerías a utilizar son: VIM, ca, RColorBrewer, dendextend, cluster, rpart, MASS, etc.
- Para la presentación gráfica del modelo se empleó el software Bizagi Modeler (v. 3.2), que es una suite ofimática freeware para modelado de procesos (Bizagi Process Modeler). Mediante este programa se pudo diagramar adecuadamente el modelo mediante gráficos intuitivos y de fácil lectura. La documentación técnica y de usuario se hizo mediante el procesador de textos Microsoft Word 2016.
- Para la recolección de datos se utilizó el cuestionario, con un formato estructurado, que se le aplicó al personal académico de la UNSAAC con cargos relacionados a la parte académica y a docentes con especialidad en Ing. Informática y de Sistemas de acuerdo a la muestra establecida.

Los instrumentos se validaron con el criterio de juicio de expertos, quienes aseguraron la validez del contenido en coherencia a los indicadores que se utilizaron en el estudio. Sin embargo, también se realizó el estudio de confiabilidad del instrumento mediante el uso del paquete de software SPSS v.25.

3.4.1 Validez del instrumento

Para validar el instrumento, se consultó a 3 expertos en el área con grado académico de doctor (cuyas fichas de evaluación se pueden encontrar en los anexos del presente estudio). El resultado de las evaluaciones se ven en la siguiente tabla:

Tabla

2

Validación del instrumento

Ítem	Criterio de evaluación	Experto 1	Experto 2	Experto 3
1	Claridad	90	90	95
2	Objetividad	95	85	95
3	Actualización	100	95	90
4	Organización	95	100	95
5	Suficiencia	95	95	80
6	Intencionalidad	95	100	85
7	Consistencia	90	90	95
8	Coherencia	100	95	95
9	Metodología	100	100	90
10	Pertinencia	95	100	85
	PROMEDIO	95.5	95	90.5

El resultado final nos da una puntuación promedio de 93.6 de 100 puntos posibles, lo cual indica que, a juicio de los expertos, el instrumento es válido para medir los indicadores ligados a la investigación.

3.4.2 Confiabilidad del instrumento

Con el fin de evaluar la confiabilidad del instrumento mediante el uso de la estadística se utilizó el programa SPSS v.25.

El índice de consistencia mide si el instrumento tiene la confiabilidad suficiente para ser aplicado, si el valor está debajo de 0.60 el instrumento presenta una variabilidad heterogénea en sus ítems y por tanto llevará a conclusiones equivocadas.

Para calcular el valor de α , se utilizó la fórmula:

$$\alpha = \frac{K}{K - 1} \left[1 - \frac{\sum V_i}{V_t} \right]$$

Donde:

α = Alfa de Cronbach

K = Número de Ítems

V_i = Varianza de cada ítem

V_t = Varianza total

Los rangos de confiabilidad establecidos son:

Tabla
Rangos de interpretación del Alpha de Cronbach

3

Rango	Magnitud
0.01 - 0.20	Muy baja
0.21 - 0.40	Baja
0.41 - 0.60	Moderada
0.61 - 0.80	Alta
0.81 - 1.00	Muy alta

En la siguiente tabla podemos apreciar los valores obtenidos mediante el indicador de Alfa de Cronbach para cada una de las dimensiones del estudio.

Tabla 4
Confiabilidad para las dimensiones

	Alfa de Cronbach	N de elementos
Validez de los datos de entrada del modelo	0.762	3
Usabilidad del modelo	0.759	5
Funcionalidad del modelo	0.745	6
Capacidad explicativa de los gráficos del modelo	0.882	12
Valoración del estudio del rendimiento académico	0.817	5
Global	0,941	31

El valor mínimo obtenido es de 0.762, mientras que el valor máximo es de 0.882; por lo tanto, los valores de coeficiente de confiabilidad calculados se encuentran en el intervalo $0.61 < \text{Alfa} < 1$. El instrumento presenta entonces, una confiabilidad entre alto y muy alto. Además, el valor global del test arroja un valor bastante alto (0.941).

Por lo tanto, se concluye que el instrumento aplicado para determinar la relación de las diferentes dimensiones del modelo basado en el análisis multivariado de datos con la valoración del estudio del rendimiento académico de las escuelas profesionales universitarias públicas de la región de Cusco, es confiable

3.5 Procedimientos

Se aplicó el cuestionario al personal académico con cargos directivos relacionados a la actividad académica y a especialistas en ingeniería informática y de sistemas seleccionados en la muestra. Esto permitió validar los resultados del proceso de análisis de datos respecto a la utilidad y valor de la información obtenida.

Para la recolección de datos se reunió a quienes serían encuestados en grupos (sin orden en particular), y se les hizo una exposición del modelo en base al instructivo y el código R consignados en los anexos del presente trabajo. Luego, se procedió a solicitar a los participantes que llenaran la encuesta de valoración del modelo. Estas sesiones se repitieron en múltiples oportunidades hasta lograr el total de la muestra establecida.

3.6 Análisis de datos

Los datos obtenidos en los cuestionarios fueron tabulados, analizados y presentados en forma de tablas y gráficos utilizando para ello las técnicas estadísticas y las herramientas de procesamiento de datos de Excel 2016 y SPSS v. 25.0.

IV. RESULTADOS

En esta sección se muestran los resultados logrados para las variables en estudio. Los datos fueron obtenidos de las encuestas, las cuales se transcribieron en una hoja de cálculo de Microsoft Excel para luego ser copiados al software estadístico SPSS (v. 25). A pesar de que la presente investigación es de corte correlacional, el análisis estadístico se hará, en primer lugar, de forma descriptiva y luego de manera correlacional, para lo cual se mostrarán tablas, gráficos y la prueba de las hipótesis respectivas.

4.1 Análisis descriptivo por dimensión

4.1.1 Dimensión: Validez de los datos de entrada del modelo

Tabla

5

Validez de los datos de entrada del modelo

	Frecuencia	Porcentaje
Regular	5	9,3
Bueno	23	42,6
Muy bueno	26	48,1
Total	54	100,0

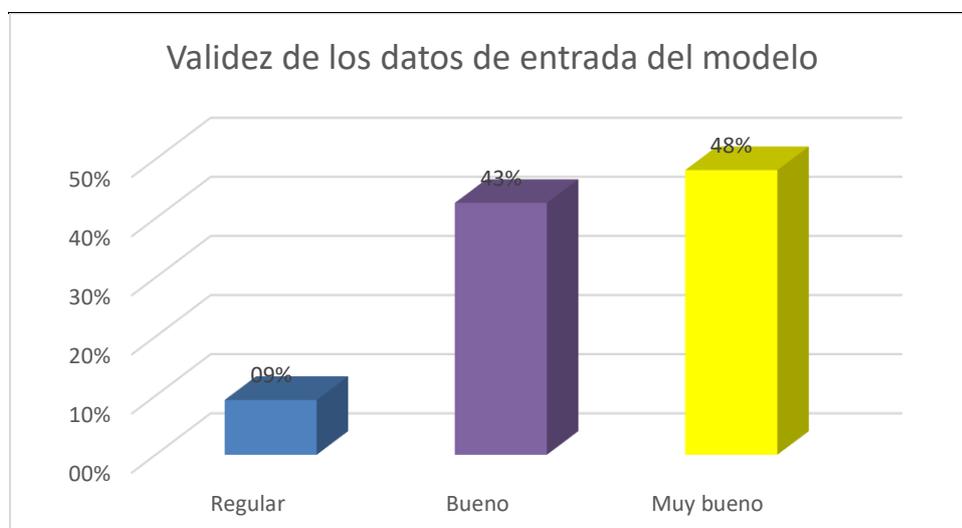


Gráfico 33. Validez de los datos de entrada del modelo

Fuente: Elaboración propia

En la tabla y figura anteriores se muestra la dimensión validez de los datos de entrada del modelo de los cuales el 9.3% (5) considera que es regular, el 42.6% (23) considera que es bueno, y el 48.1% (26) considera que es muy bueno.

En general más del 90% de encuestados valora positivamente la validez de los datos de entrada que plantea el modelo para un adecuado procesamiento estudio del rendimiento académico.

4.1.2 Dimensión: Usabilidad del modelo

Tabla
Usabilidad del modelo

6

	Frecuencia	Porcentaje
Regular	6	11,1
Bueno	37	68,5
Muy bueno	11	20,4
Total	54	100,0

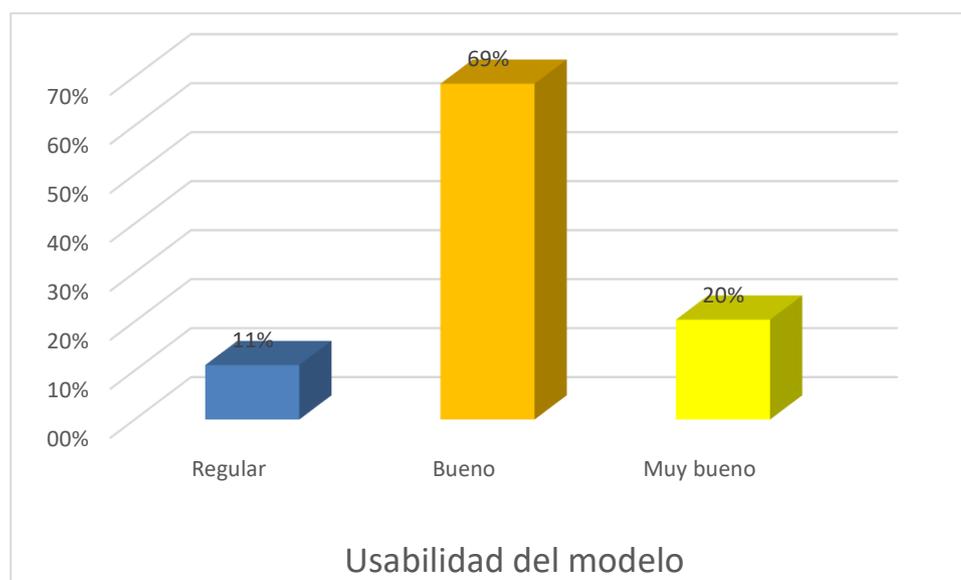


Gráfico 34. Usabilidad del modelo
Fuente: Elaboración propia

En la tabla y figura anteriores se muestra la dimensión usabilidad del modelo de los cuales el 11.1% (6) considera que es regular, el 68.5% (37) considera que es bueno, y el 20.4% (11) considera que es muy bueno.

Se puede apreciar que cerca al 90 % de los encuestados validan positivamente la usabilidad del modelo, aun cuando no se trata de una herramienta de software de usuario final.

4.1.3 Dimensión: Funcionalidad del modelo

Tabla

7

Funcionalidad del modelo

Funcionalidad del modelo	Frecuencia	Porcentaje
Regular	1	1,9
Bueno	33	61,1
Muy bueno	20	37,0
Total	54	100,0

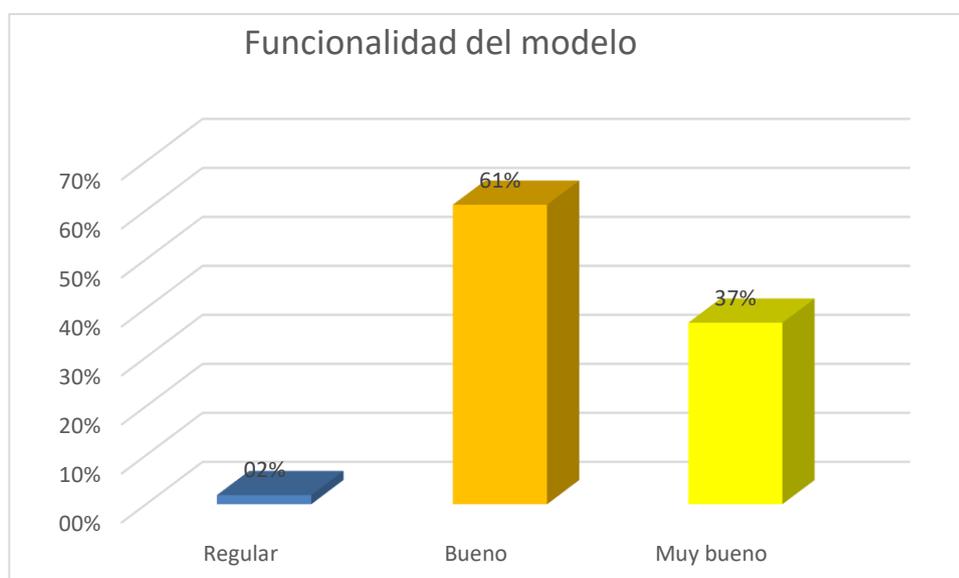


Gráfico 35. Funcionalidad del modelo

Fuente: Elaboración propia

En la tabla y figura anteriores se muestra la dimensión funcionabilidad del modelo de los cuales el 1.9% (1) considera que es regular, el 61.1% (33) considera que es bueno, y el 37% (20) considera que es muy bueno.

Esta dimensión está bastante bien valorada por los usuarios, ya que casi el total de los encuestados considera que el modelo tiene una buena funcionalidad.

4.1.4 Dimensión: Capacidad explicativa de los gráficos del modelo

Tabla
Capacidad explicativa de los gráficos del modelo

8

Capacidad explicativa de los gráficos del modelo	Frecuencia	Porcentaje
Regular	4	7,4
Bueno	38	70,4
Muy bueno	12	22,2
Total	54	100,0

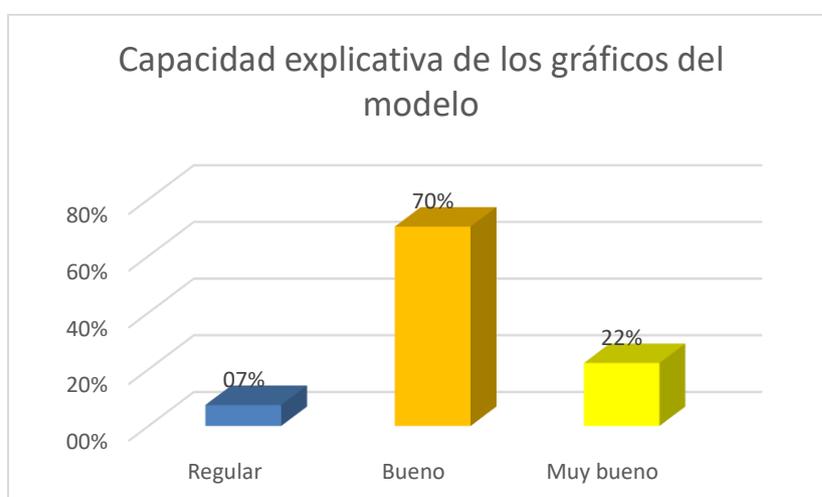


Gráfico 36. Capacidad explicativa de los gráficos del modelo
Fuente: Elaboración propia

En la tabla y figura se muestra la dimensión capacidad explicativa de los gráficos del modelo de los cuales el 7.4% (4) considera que es regular, el 70.4% (38) considera que es bueno, y el 22.2% (12) considera que es muy bueno.

Más del 90% de encuestados valora como bueno o muy bueno el aporte analítico que brindan los gráficos del modelo para el estudio del rendimiento académico.

4.1.5 Variable: Valoración general del modelo basado en el análisis multivariado

de datos

Tabla
Modelo basado en el análisis multivariado de datos

9

	Frecuencia	Porcentaje
Regular	4	7,4
Bueno	38	70,4
Muy bueno	12	22,2
Total	54	100,0

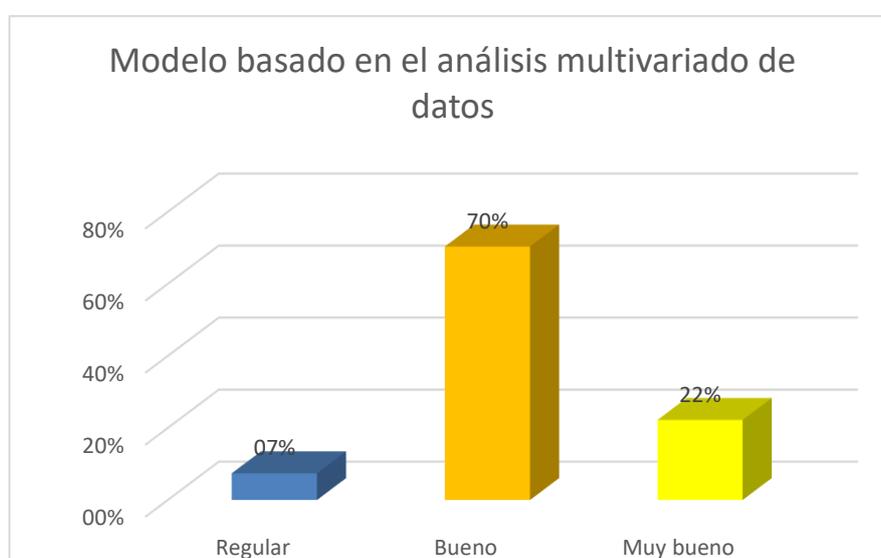


Gráfico 37. Modelo basado en el análisis multivariado de datos
Fuente: Elaboración propia

En la tabla y figura anteriores se muestra la Variable modelo basado en el análisis multivariado de datos de los cuales el 7.4% (4) considera que es regular, el 70.4% (38) considera que es bueno, y el 22.2% (12) considera que es muy bueno.

En general, los usuarios consideran que el modelo basado en el análisis multivariado de datos presenta características muy buenas para poder aportar al estudio del rendimiento académico.

4.1.6 Dimensión: Valoración del estudio del rendimiento académico

Tabla 10
Modelo Valoración del estudio del rendimiento académico

Valoración del estudio del rendimiento académico	Frecuencia	Porcentaje
Regular	3	5,6
Bueno	31	57,4
Muy bueno	20	37,0
Total	54	100,0

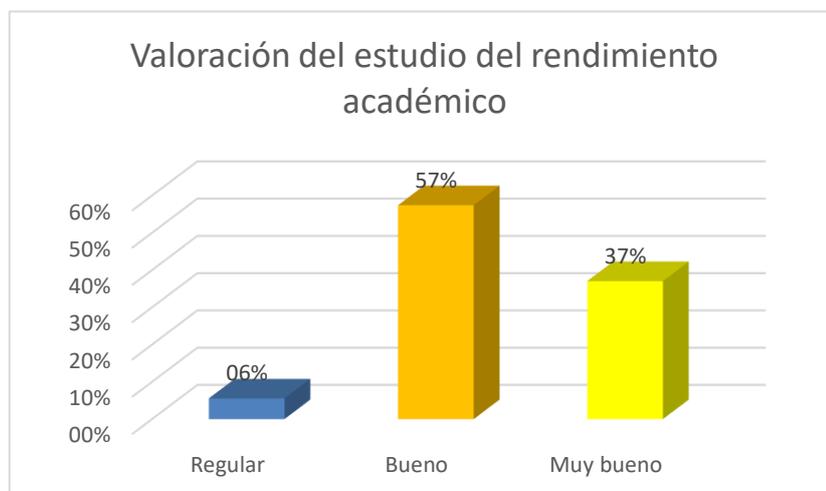


Gráfico 38. Valoración del estudio del rendimiento académico
Fuente: Elaboración propia

En la tabla y figura se muestra la variable valoración del estudio del rendimiento académico de los cuales el 5.6% (3) considera que es regular, el 57.4% (31) considera que es bueno, y el 37% (20) considera que es muy bueno.

La valoración del estudio del rendimiento académico es bastante buena. Más del 90% de los encuestados consideran muy positivamente las características que se aportan en esta variable.

4.2 Contrastación de hipótesis

Para el análisis de los resultados obtenidos se determinó inicialmente el tipo de distribución que presentan las variables, para ello se utilizó la prueba Kolmogorov-Smirnov. Su objetivo es señalar si las variables provienen de una población que tiene la distribución normal o no lo tiene. Una vez realizada esta prueba, se decidirá entre usar una prueba paramétrica r de Pearson o no paramétricos Rho de Spearman.

Para determinar la normalidad de los datos mediante la prueba de Kolmogorov-Smirnov, se considera el siguiente criterio:

$P \geq \alpha$ = Los datos provienen de una distribución normal

$P < \alpha$ = Los datos no provienen de una distribución normal

Tabla 11
Prueba de Kolmogorov-Smirnov para una muestra

	Z de Kolmogorov-Smirnov	P
Validez de los datos de entrada del modelo	1,28	0,075
Usabilidad del modelo	1,11	0,173
Funcionalidad del modelo	1,27	0,078
Capacidad explicativa de los gráficos del modelo	1,29	0,071
Modelo basado en el análisis multivariado de datos	1,03	0,243
Valoración del estudio del rendimiento académico	1,13	0,153

En la tabla se observa que todos los valores de P son mayores que $\alpha = 0.05$, lo que indica que las variables tienen distribución normal y, por lo tanto, corresponde aplicar, para el estudio, una prueba paramétrica de correlación de Pearson, como se muestra a continuación.

4.2.1 *Contrastación de la hipótesis específica 1*

La hipótesis específica 1 se encuentra planteada de la siguiente forma:

“La validez de los datos de entrada del modelo basado en el análisis multivariado de datos se relaciona con la valoración del estudio del rendimiento académico”

Los resultados de la aplicación de la prueba de Pearson se observan en la siguiente tabla:

Tabla 12
Correlación entre la Validez de los datos de entrada del modelo y Valoración del estudio del rendimiento académico

	Valoración del estudio del rendimiento académico		
	N	Correlación de Pearson	P
Validez de los datos de entrada del modelo	54	,666	,000

Procedimiento de la prueba de hipótesis

Formulación de las Hipótesis estadísticas	Ho: No existe correlación entre la validez de los datos de entrada del modelo y la valoración del estudio del rendimiento académico.	
	Ha: Existe correlación entre la validez de los datos de entrada del modelo y la valoración del estudio del rendimiento académico.	
Nivel de significación	$\alpha = 0,05 = 5\%$	
Estadígrafo de contraste	$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$	Valor calculado
		$t = 6.44$
Valor p calculado	$p = 0,000$	

Conclusión

Como $p < 0,05$, se rechaza la hipótesis nula y se concluye que existe correlación entre la validez de los datos de entrada del modelo y la valoración del estudio del rendimiento académico. También podemos observar que la validez de los datos de entrada del modelo está correlacionado directamente con la valoración del estudio del rendimiento académico, esto es a menor validez de los datos de entrada del modelo, menor será la valoración del estudio del rendimiento académico y a mayor validez de los datos de entrada del modelo, mayor será la valoración del estudio del rendimiento académico; además la correlación de Pearson es 0.666, lo que indica que esta correlación es alta.

4.2.2 *Contrastación de la hipótesis específica 2*

La hipótesis específica 2 se encuentra planteada de la siguiente forma:

“La usabilidad del modelo basado en el análisis multivariado de datos se relaciona con la valoración del estudio del rendimiento académico”

Los resultados de la aplicación de la prueba de Pearson se observan en la siguiente tabla:

Tabla 13
Correlación entre la usabilidad del modelo y valoración del estudio del rendimiento académico

	Valoración del estudio del rendimiento académico		
	N	Correlación de Pearson	P
Usabilidad del modelo	54	,615	,000

Procedimiento de la prueba de hipótesis

Formulación de las Hipótesis estadísticas	Ho: No existe correlación entre la usabilidad del modelo y la valoración del estudio del rendimiento académico.	
	Ha: Existe correlación entre la usabilidad del modelo y la valoración del estudio del rendimiento académico.	
Nivel de significación	$\alpha = 0,05 = 5\%$	
Estadígrafo de contraste	$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$	Valor calculado
		$t = 5.62$
Valor p calculado	$p = 0,000$	

Conclusión

Como $p < 0,05$, se rechaza la hipótesis nula y se concluye que existe correlación entre la usabilidad del modelo y la valoración del estudio del rendimiento académico. También podemos observar que la usabilidad del modelo está correlacionada directamente con la valoración del estudio del rendimiento académico, esto es, a menor usabilidad del modelo, menor será la valoración del estudio del rendimiento académico y a mayor usabilidad del modelo, mayor será la valoración del estudio del rendimiento académico; además, la correlación de Pearson es 0.615, lo que indica que esta correlación es alta.

4.2.3 *Contrastación de la hipótesis específica 3*

La hipótesis específica 2 se encuentra planteada de la siguiente forma:

“La funcionalidad del modelo basado en el análisis multivariado de datos se relaciona con la valoración del estudio del rendimiento académico.

Los resultados de la aplicación de la prueba de Pearson se observan en la siguiente tabla:

Tabla 14
Correlación entre la funcionalidad del modelo y valoración del estudio del rendimiento académico.

	Valoración del estudio del rendimiento académico		
	N	Correlación de Pearson	P
Funcionalidad del modelo	54	,722	0,000

Procedimiento de la prueba de hipótesis

Formulación de las Hipótesis estadísticas	Ho: No existe correlación entre la funcionalidad del modelo y la Valoración del estudio del rendimiento académico.	
	Ha: Existe correlación entre la funcionalidad del modelo y la Valoración del estudio del rendimiento académico.	
Nivel de significación	$\alpha = 0,05 = 5\%$	
Estadígrafo de contraste	$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$	Valor calculado
		$t = 7.52$
Valor p calculado	$p = 0,000$	

Conclusión:

Como $p < 0,05$, se rechaza la hipótesis nula y se concluye que existe correlación entre la funcionalidad del modelo y la valoración del estudio del rendimiento académico. También podemos observar que la funcionalidad del modelo está correlacionada directamente con la valoración del estudio del rendimiento académico, esto es, a menor funcionalidad del modelo, menor será la valoración del estudio del rendimiento académico y, a mayor la funcionalidad del modelo, mayor será la valoración del estudio del rendimiento académico, además la correlación de Pearson es 0.722, lo que indica que esta correlación es alta.

4.2.4 Contrastación de la hipótesis específica 4

La hipótesis específica 4 se encuentra planteada de la siguiente forma:

“La capacidad explicativa de los gráficos del modelo basado en el análisis multivariado de datos se relaciona con la valoración del estudio del rendimiento académico”

Los resultados de la aplicación de la prueba de Pearson se observan en la siguiente tabla:

Tabla 15
Correlación entre la capacidad explicativa de los gráficos del modelo y valoración del estudio del rendimiento académico.

	Valoración del estudio del rendimiento académico		
	N	Correlación de Pearson	P
Capacidad explicativa de los gráficos del modelo	54	,682	,000

Procedimiento de la prueba de hipótesis

Formulación de las Hipótesis estadísticas	Ho: No existe correlación entre la capacidad explicativa de los gráficos del modelo y la valoración del estudio del rendimiento académico.	
	Ha: Existe correlación entre la capacidad explicativa de los gráficos del modelo y la valoración del estudio del rendimiento académico.	
Nivel de significación	$\alpha = 0,05 = 5\%$	
Estadígrafo de contraste	$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$	Valor calculado
		$t = 6.72$
Valor p calculado	$p = 0,000$	

Conclusión

Como $p < 0,05$, se rechaza la hipótesis nula y se concluye que existe correlación entre la capacidad explicativa de los gráficos del modelo y la valoración del estudio del rendimiento académico. También podemos observar que la capacidad explicativa de los gráficos del modelo está correlacionada directamente con la valoración del estudio del rendimiento académico; esto es, a menor capacidad explicativa de los gráficos del modelo, menor será la valoración del estudio del rendimiento académico y, a mayor capacidad explicativa de los gráficos del modelo, mayor será la valoración del estudio

4.2.5 *Contrastación de la hipótesis general*

La hipótesis general se encuentra planteada de la siguiente forma:

“Las diferentes dimensiones del modelo basado en el análisis multivariado de datos se relacionan con la valoración del estudio del rendimiento académico de las escuelas profesionales de las universidades públicas de la región del Cusco.

Los resultados de la aplicación de la prueba de Pearson se observan en la siguiente tabla:

Tabla 16
Correlación entre el modelo basado en el análisis multivariado de datos y valoración del estudio del rendimiento académico.

	Valoración del estudio del rendimiento académico		
	N	Correlación de Pearson	P
Modelo basado en el análisis multivariado de datos	54	,770	,000

Procedimiento de la prueba de hipótesis

Formulación de las Hipótesis estadísticas	Ho: No existe correlación entre las diferentes dimensiones del modelo basado en el análisis multivariado de datos y la valoración del estudio del rendimiento académico.	
	Ha: Existe correlación entre las diferentes dimensiones del modelo basado en el análisis multivariado de datos y la valoración del estudio del rendimiento académico.	
Nivel de significación	$\alpha = 0,05 = 5\%$	
Estadígrafo de contraste	$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$	Valor calculado
		$t = 8.70$
Valor p calculado	$p = 0,000$	

Conclusión

Como $p < 0,05$, se rechaza la hipótesis nula y se concluye que existe correlación entre las diferentes dimensiones del modelo basado en el análisis multivariado de datos y la valoración del estudio del rendimiento académico. Esta relación es directa y tiene un valor muy alto, lo cual confirma la hipótesis planteada en el presente estudio.

V. DISCUSIÓN DE RESULTADOS

Los resultados obtenidos permiten determinar la correlación positiva existente entre las diferentes dimensiones del modelo basado en el análisis multivariante de datos y la valoración del estudio del rendimiento académico. La utilidad del análisis multivariante de datos relacionado al estudio de factores socioeconómicos relacionados al estudio del rendimiento académico fue resaltado por Chue, Barreno y Millones (2007) quienes utilizaron técnicas como el Análisis de Correspondencias y el análisis de clusters para su estudio. De la misma forma otros estudios permiten valorar las técnicas multivariadas de datos en estudios del rendimiento académico, como, Camborda (2014) con los árboles de decisión, Lizares (2017) con las técnicas de clasificación y de Martelo, Acevedo y Martelo (2018) con el análisis factorial.

Los resultados obtenidos para la evaluación de la primera dimensión respecto a la validez de los datos de entrada, permite observar que existen aspectos de inicio que son muy importantes para empezar a utilizar el modelo. La posibilidad de observar los datos antes de ser procesados, hace que el usuario pueda entender el formato inicial, que los pueda reconocer y que pueda familiarizarse con ellos; la herramienta que se usa para este fin (Microsoft Excel) facilita esta labor, ya que los usuarios están familiarizados con su entorno de trabajo. Del mismo modo, esta vista previa permite que el usuario tome conciencia de la magnitud de la información que se le presenta y de lo difícil que sería procesarla por los medios con los que cuenta. También permite que el usuario valide la información en cuanto a las escalas de calificación y la selección de las asignaturas; el hecho de la elección de las asignaturas de corte obligatorio es de suma importancia para contextualizar el análisis, tal como se puede apreciar en el estudio realizado a la escuela profesional de Ingeniería Informática y de Sistemas de la UNSAAC (Alzamora, 2014). Todos estos factores tienen influencia en la valoración final que

le dan al modelo, ya que pueden observar el tránsito y la transformación de estos datos iniciales en información que le resulte valiosa.

Respecto a la usabilidad del modelo se debe precisar que, la herramienta utilizada (R) para probar la funcionalidad de un modelo, no es una herramienta destinada a crear aplicaciones para usuarios finales; esto es, no tiene dentro de sus funcionalidades, la creación de Interfaces Gráficas de Usuario; básicamente es un programa que ejecuta scripts y al que los especialistas en informática y sistemas están acostumbrados, pero el resto de usuarios no. Con el fin de contrarrestar esta debilidad se diseñó un instructivo de uso que guíe paso a paso la forma de trabajo y que permita obtener resultados favorables sin necesidad de ser un experto en la materia. Los resultados de este trabajo fueron favorables para los usuarios, ya que aspectos como la facilidad de carga de datos desde la hoja electrónica hacia la herramienta R, la facilidad de ejecutar el modelo paso a paso, la interfaz de usuario del entorno de análisis de datos, la documentación para ejecutar el modelo con facilidad, la documentación para leer e interpretar los resultados del modelo y otros aspectos adicionales contemplados por la International Organization for Standardization (ISO/IEC 25012, s.f), fueron altamente valorados en el estudio del rendimiento académico de una escuela profesional. Aun así, es deseable que el modelo cuente con una mejor interfaz que oculte al usuario final el léxico, la sintaxis y semántica de un lenguaje de programación. Hay que resaltar, sin embargo, que este último aspecto no forma parte de los objetivos de la presente tesis.

La dimensión de la funcionalidad del modelo es uno de los más importantes, ya que modelo que no es funcional, simplemente es un fracaso. Los resultados obtenidos permiten asegurar que el modelo presenta una gran funcionalidad en el procesamiento de los datos. La idea de que el usuario pueda ver evolucionar los datos iniciales a través de todo el modelo y que pueda ligar los resultados obtenidos en cada etapa de manera natural, fueron logrados satisfactoriamente. Aspectos contemplados por la International Organization for Standardization

(ISO/IEC 25010, s.f) como la rapidez en el procesamiento de la información, la verificación del hecho que se obtienen los mismos resultados si se ejecuta el modelo con los mismos datos iniciales, la valoración de los formatos de tabla y gráficos para analizar la información, la adecuada cantidad de información mostrada en los resultados, la capacidad de analizar diferentes escenarios y la coherencia del eje temático estudiado fueron sometidos al juicio de los usuarios. Queda claro, luego del estudio, que estos aspectos tienen una gran valoración para realizar el estudio del rendimiento académico de una escuela profesional.

Respecto a la calidad que tienen los gráficos para permitir interpretar la información presentada por el modelo, el estudio demuestra que este es uno de los aspectos mejor valorados por los usuarios. Tal como lo establecieron Cleveland-McGill (1985) el uso de los gráficos adecuados es muy favorable para simplificar la complejidad del volumen de información mostrada. El modelo hace uso de esta premisa y lo convierte en uno de sus puntos fuertes. Para evaluar la capacidad de los gráficos se evaluaron aspectos como la capacidad de los gráficos para mostrar la relación entre las diferentes variables de estudio (asignaturas-estudiantes, asignaturas-calificaciones), la identificación clara de los elementos de los gráficos (genéricos y relevantes), la consistencia de los resultados entre los diferentes gráficos, la capacidad de representar los datos originales con claridad, la capacidad de los gráficos para analizar la información y permitir múltiples interpretaciones, la valoración de la utilidad de los gráficos en el análisis de la información y, específicamente en el estudio del rendimiento académico y, la validez de la escala de los gráficos. Todos estos aspectos fueron muy apreciados al momento de realizar un estudio del rendimiento académico de una escuela profesional y, definitivamente, demostraron que se encuentran relacionados entre sí.

VI. CONCLUSIONES

1. Las dimensiones del modelo basado en técnicas del análisis multivariado de datos se relacionan con la valoración del estudio del rendimiento académico de las escuelas profesionales universitarias. Un 92.6% de los encuestados califican el modelo como bueno o muy bueno en relación con el estudio del rendimiento académico. Este hecho se evidencia también en una alta correlación entre el modelo planteado (a través de un proceso de valoración de sus características) y la valoración que se le da en el estudio del rendimiento académico, lo cual permite afirmar que el modelo mejora la observación y análisis de las calificaciones de los estudiantes y su relación con las asignaturas de su currículo de estudios.
2. La validez de los datos de entrada es sumamente importante para el modelo planteado y se encuentra relacionado con la valoración del estudio del rendimiento académico. El 90.7% de los encuestados percibe la validez de los datos de entrada del modelo como buena o muy buena. El estudio estadístico demuestra que existe una correlación directa entre la validez de los datos de entrada y la valoración del estudio del rendimiento académico. La confianza en la información que se utiliza como materia prima para el estudio permite al usuario una línea de análisis articulada a lo largo del proceso con un significado claro y sencillo.
3. La usabilidad del modelo está relacionado con la valoración del estudio del rendimiento académico. El 88.9% de los encuestados califica los aspectos de usabilidad del modelo como buenos o muy buenos a pesar de que el programa utilizado para probar el funcionamiento del modelo no se puede considerar como una herramienta de usuario final. Aun cuando el entorno de R no es el mejor, los usuarios pudieron valorar de manera positiva su facilidad de uso en base al instructivo de usuario presentado como parte del

modelo. Los resultados estadísticos permitieron establecer una correlación directa entre la usabilidad del modelo y la valoración del estudio del rendimiento académico de las escuelas profesionales.

4. La funcionalidad del modelo se relaciona con la valoración del estudio del rendimiento académico. Un 98.1% de los encuestados manifiesta que la funcionalidad del modelo es buena o muy buena. La rapidez con que el modelo presenta resultados, la variedad de los formatos sin sobrecarga de información y la coherencia en los resultados de cada paso son bastante bien valorados a lo largo de todo el proceso de análisis de la información. Los resultados estadísticos permiten determinar una correlación positiva entre la funcionalidad del modelo y la valoración del estudio del rendimiento académico de las escuelas profesionales.
5. La capacidad explicativa que muestra el modelo a través de sus gráficos se relaciona con la valoración del estudio del rendimiento académico. La capacidad de poder mostrar de manera adecuada los resultados del análisis, mediante gráficos que sean fácilmente interpretables y que se encuentren concatenados a través de la secuencia establecida, es un factor que fue bastante bien valorado por los encuestados, ya que el 92% de los mismos calificó este aspecto como bueno o muy bueno. Los estudios estadísticos permitieron establecer una correlación positiva entre la capacidad explicativa de los gráficos del modelo y la valoración del estudio del rendimiento académico de las escuelas profesionales.

VII. RECOMENDACIONES

1. Con el fin de mejorar el aspecto de la entrada de datos, se recomienda que el modelo pueda extraer la información directamente del repositorio institucional (centro de cómputo o el que haga sus funciones). Esto será posible a través de la decisión de las autoridades universitarias quienes deberán adoptar el modelo como herramienta de apoyo a la toma de decisiones académicas de los directores de las escuelas profesionales.
2. El aspecto de la usabilidad del modelo puede ser mejorado a través de la construcción de una interfaz de usuario final (aplicación de escritorio, página web, etc.) que brinde todas las posibilidades de manipulación visual que actualmente no tiene. Eso se logrará luego de que el modelo sea adoptado por la universidad y que se encargue al área correspondiente su implementación.
3. Para mejorar la funcionalidad del modelo se puede ampliar su capacidad para incluir la variable docente dentro del estudio de forma que se pueda analizar su rol dentro del rendimiento académico. Este aspecto se puede lograr en el proceso de implementación por parte de la universidad, ya que la información sobre docentes y demás variables se encuentran bajo custodia de la institución. El aspecto relacionado a la confidencialidad de los datos, debe ser normado previamente a través de instrumentos de gestión universitaria (resoluciones, reglamentos, directivas o similares), para evitar el mal uso de la información.
4. Otra forma de ampliar la funcionalidad del modelo es articularlo con otras investigaciones como aquellas que ligan los factores socio-económicos, psicológicos, de rendimiento escolar, etc. y las asocian con el rendimiento académico universitario, o con aquellas que estudian la valoración estudiantil al desempeño docente, o con las que buscan estudiar el efecto del rendimiento académico universitario con el desempeño laboral de sus egresados,

entre otras. Este aspecto sólo será posible asumiendo una política universitaria de puesta en marcha de los trabajos de investigación, ya que, muchos de ellos, una vez presentados, pasan a formar parte del repositorio de trabajos y no se llegan a implementar para explotar su utilidad.

5. La capacidad explicativa de los gráficos puede ser mejorada utilizando librerías que permitan mostrar la información desde diferentes ángulos. Este aspecto forma parte de la implementación e implantación del modelo como herramienta de usuario final.

VIII. REFERENCIAS

- Alzamora Paredes, R. (2014). *Estudio del rendimiento académico mediante análisis multivariado de datos. Ingeniería Informática y de Sistemas – UNSAAC 1993-2010*. Universidad del País Vasco – España.
- Araya Alpizar, C. (2011). *Análisis de datos multivariantes con coordenadas paralelas*. Pensamiento actual. Universidad de Costa Rica. Vol 11 – N°16-17.
- Ayunga, E. (2010). *Análisis de datos multivariantes*. Recuperado de http://ocw.upm.es/estadistica-e-investigacion-operativa/matematicas-y-estadistica-aplicada/contenidos/OCW/Anal_Multivar/Mat_Clase/anal_mult_1.pdf
- Camborda Zamudio y María Gabriela (2014). *Aplicación de árboles de decisión para la predicción del rendimiento académico de los estudiantes de los primeros ciclos de La Carrera de Ingeniería Civil de la Universidad Continental*. Huancayo – Perú
- Chue Gallardo, Jorge, Barreno Vereau, Emma y Millones Rivalles, Rosa (2007). *Sistema para el análisis estadístico de técnicas multivariadas del rendimiento académico de los estudiantes de una institución de enseñanza superior*. Universidad de Lima- Perú.
- Cleveland, William y McGill, Robert. (1985). Graphical perception and graphical methods for analyzing scientific data. *Science*, 229(4716), 828-833.
- Cleveland, W.S. (1993). *Visualizing Data*. At&T Bell Laboratories.
- Cuadras, C. M. (1996). *Métodos de análisis multivariante*. Eunibar.
- De la Fuente Fernández, S. (2011). *Análisis de correspondencias simples y múltiples*. Universidad Autónoma de Madrid.
- Dürsteler, Juan C. (2010). *Coordenadas paralelas*. Revista digital InfoVis.net. Recuperado de <http://www.infovis.net/printMag.php?num=201&lang=1>

Ferreiro Seoane, Francisco, Santos, Marta y Villares, Manuel. (2017). *Análisis multivariable del rendimiento académico del alumnado más sobresaliente de la educación secundaria obligatoria en Galicia, España*.

Garnica, E. (1997). *El Rendimiento Estudiantil: Una Metodología para su Medición*. Revista de Economía, N°. 13, pp. 7-26. Recuperado de <http://www.saber.ula.ve/handle/123456789/19178>

Greenacre, M. (2008). *La práctica del Análisis de correspondencias*. Fundación BBVA.

Härdle, Wolfgang & Simar, Léopold (2003). *Applied Multivariate Statistical Analysis*. MD Tech.

Hernández Orallo, José, Ramírez Quintana, M^a José y Ferri Ramírez, César (2004). *Introducción a la Minería de Datos*. Pearson Prentice Hall.

Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, P. (2014). *Metodología de la investigación* (6a. ed.) México D.F.: McGraw-Hill.

ISO/IEC 25010. Modelo de Calidad del producto de software (s.f.). Recuperado de <https://iso25000.com/index.php/normas-iso-25000/iso-25010?limit=3&limitstart=0>

ISO/IEC 25012. Calidad de datos (s.f.). Recuperado de <https://iso25000.com/index.php/normas-iso-25000/iso-25012>

Lizares Castillo, Mónica (2017). *Comparación de modelos de clasificación: regresión logística y árboles de clasificación para evaluar el rendimiento académico*. Universidad Nacional Mayor de San Marcos. Lima – Perú.

Llopis Pérez, J. (s.f.). *Análisis de Correspondencias*. Recuperado de <http://estadisticaorquestainstrumento.wordpress.com/2013/07/06/tema-27-analisis-de-correspondencias/>

Marín Diazaraque, Juan Miguel (s.f.). *Análisis de componentes principales*. Recuperado de <http://halweb.uc3m.es/esp/Personal/personas/jmmarin/esp/AMult/tema3am.pdf>

- Martelo, Raúl J., Acevedo, Diofanor y Martelo, Piedad M. (2018). *Análisis Multivariado aplicado a determinar factores clave de la deserción universitaria*. Revista digital Espacios. Vol. 39 (Nº 10) Año 2018. Pág. 13.
- Martínez Otero Pérez (2002). *Claves del rendimiento escolar*. Artículo en Tribuna Libre. Ministerio de Educación, Cultura y Deporte.
- McDonnell K. T. & Mueller K. (2008). *Illustrative Parallel Coordinates*. Eurographics/ IEEE-VGTC Symposium on Visualization 2008. Volume 27 (2008), Number 3. Recuperado de http://dbonline.igroupnet.com/ACM.TOOLS/Rawdata/130103_0839/fulltext/2430000/2421829/p1031-mcdonnell.pdf
- Muñoz Rosas, J. F. y Alvarez Verdejo, E. (2009). *Métodos de imputación para el tratamiento de datos faltantes: aplicación mediante R/Splus*. Revista de métodos cuantitativos para la economía y la empresa. Recuperado de <http://www.upo.es/RevMetCuant/pdf/vol7/art25.pdf>
- Murtagh, F. & Heck, A. (1987). *Multivariate Data Analysis*. Editorial Board
- Nováez, M. (1986). *Psicología de la actividad*. México. Editorial iberoamericana
- Páez de Marín, A (1987). *Rendimiento estudiantil en química en el primer año de ciencias del nivel de educación media diversificada y profesional: distrito Maracaibo. Sector Público*. Universidad de Zulia. Maracaibo.P/s.p.
- Peña, D. (2002). *Análisis de datos multivariantes*. Mcgraw-Hill / Interamericana de España.
- Pizarro, R. (1985). *Rasgos y actitudes del profesor efectivo*. Universidad de Chile. Chile.
- Rencher, Alvin C. (2002). *Methods of Multivariate Analysis*. Wiley-Interscience
- Rodríguez, M (2007), *Análisis multivariado del desempeño académico de estudiantes universitarios de Química*. Universidad Autónoma de Madrid. España.

Salvador Figueras, M (2003). *Análisis de Correspondencias*. Recuperado de <http://www.5campus.com/leccion/correspondencias>

Seijas, Carlos R. (2000). *El análisis multivariado de datos como herramienta en la exploración de modelos en la comprensión de variables psicológicas*. Revista de Matemática: Teoría y Aplicaciones. Ejemplar 2000 7(1-2) : 229–244 cimpa – ucr issn: 1409-2433. Costa Rica.

IX. ANEXOS

ANEXO 1. Matriz de consistencia

PROBLEMAS	OBJETIVO	HIPOTESIS	VARIABLES, DIMENSIONES E INDICADORES	METODOLOGIA
<p>GENERAL</p> <p>¿De qué manera las diferentes dimensiones de un modelo basado en el análisis multivariado de datos se relacionan con la valoración del estudio del rendimiento académico de las escuelas profesionales universitarias públicas de la región del Cusco?</p>	<p>GENERAL</p> <p>Determinar si las diferentes dimensiones de un modelo basado en el análisis multivariado de datos se relacionan con la valoración del estudio del rendimiento académico de los estudiantes de las escuelas profesionales universitarias públicas de la región del Cusco.</p>	<p>GENERAL</p> <p>Las diferentes dimensiones del modelo basado en el análisis multivariado de datos se relacionan con la valoración del estudio del rendimiento académico de las escuelas profesionales de las universidades públicas de la región del Cusco.</p>	<p>VARIABLE INDEPENDIENTE</p> <p>Modelo basado en el análisis multivariado de datos.</p> <p>Dimensiones (D) e Indicadores (I)</p> <p>D - 1. Validez de los datos de entrada</p> <p>I - 1.1. Facilidad de interpretación de los datos iniciales.</p> <p>I - 1.2. Dificultad en el análisis de datos</p> <p>I - 1.3. Pertinencia de los datos para el estudio</p> <p>D - 2. Usabilidad</p> <p>I - 2.1. Facilidad de carga de datos</p> <p>I - 2.2. Facilidad de ejecutar el modelo</p> <p>I - 2.3. Valoración de la interfaz de usuario</p>	<p>Enfoque:</p> <p>Cuantitativo</p> <p>Tipo:</p> <p>Aplicada</p> <p>Nivel:</p> <p>Correlacional</p> <p>Diseño</p> <p>No Experimental</p> <p>Técnicas</p> <p>Encuesta</p>

<p>ESPECIFICOS</p> <p>a) ¿De qué manera la validez de los datos de entrada del modelo basado en el análisis multivariado se relaciona con la valoración del estudio del rendimiento académico?</p> <p>b) ¿De qué manera la usabilidad del modelo basado en el análisis multivariado de datos se relaciona con la valoración del estudio del rendimiento académico?</p>	<p>ESPECIFICOS</p> <p>a) Determinar si la validez de los datos de entrada del modelo basado en el análisis multivariado se relaciona con la valoración del estudio del rendimiento académico.</p> <p>b) Determinar si la usabilidad del modelo basado en el análisis multivariado de datos se relaciona con la valoración del estudio del rendimiento académico.</p>	<p>ESPECIFICOS</p> <p>a) La validez de los datos de entrada del modelo basado en el análisis multivariado de datos se relaciona con la valoración del estudio del rendimiento académico.</p> <p>b) La usabilidad del modelo basado en el análisis multivariado de datos se relaciona con la valoración del estudio del rendimiento académico.</p>	<p>I - 2.4. Valoración de la documentación para ejecutar el modelo</p> <p>I - 2.5. Valoración de la documentación en la interpretación de resultados.</p> <p>D - 3. Funcionalidad</p> <p>I - 3.1. Rapidez en la obtención de resultados</p> <p>I - 3.2. Obtención de los mismos resultados con los mismos datos</p> <p>I - 3.3. Variedad de formatos para analizar la información.</p> <p>I - 3.4. Sobrecarga de información</p> <p>I - 3.5. Capacidad de analizar diferentes escenarios.</p> <p>I - 3.6. Coherencia del eje temático estudiado.</p> <p>D - 4. Capacidad explicativa de los gráficos</p> <p>I - 4.1. Capacidad de mostrar la relación entre asignaturas y estudiantes.</p> <p>I - 4.2. Capacidad de mostrar la relación entre asignaturas y calificaciones</p> <p>I - 4.3. Identificación clara de los elementos de los gráficos</p>	<p>Instrumento</p> <p>Cuestionario de valoración sobre el modelo y el estudio del rendimiento académico de las escuelas profesionales universitarias (personal académico con cargos directivos relacionados a la actividad académica)</p> <p>Guía de entrevista (expertos en ing. de sistemas e informática)</p> <p>Población:</p> <p>Personal académico con cargos directivos</p>
---	---	--	--	--

<p>c) ¿De qué manera la funcionalidad del modelo basado en el análisis multivariado de datos se relaciona con la valoración del estudio del rendimiento académico?</p>	<p>c) Determinar si la funcionalidad del modelo basado en el análisis multivariado de datos se relaciona con la valoración del estudio del rendimiento académico.</p>	<p>c) La funcionalidad del modelo basado en el análisis multivariado de datos se relaciona con la valoración del estudio del rendimiento académico.</p>	<p>I - 4.4. Identificación de los elementos más relevantes. I - 4.5. Consistencia entre gráficos I - 4.6. Suficiencia de la información mostrada para relacionarla con otros gráficos. I - 4.7. Capacidad de representar los datos originales con claridad I - 4.8. Capacidad de los gráficos para generar múltiples interpretaciones I - 4.9. Valoración de la utilidad de los gráficos. I - 4.10. Valoración de los gráficos en el estudio del rendimiento académico. I - 4.11. Presentación adecuada de los gráficos. I - 4.12. Valoración de la escala de los gráficos.</p>	<p>relacionados a la actividad académica y expertos en Ing. de Sistemas e Informática: N=131 Muestra Probabilística $n = \frac{N \times Z_a^2 \times p \times q}{d^2 \times (N-1) + Z_a^2 \times p \times q}$ N=131 p=0.5 q=0.5 Z=95% d=5% n=54 individuos Procesamiento y análisis de datos</p>
<p>d) ¿De qué manera la capacidad explicativa de los gráficos que genera el modelo basado en el análisis multivariado de datos se relaciona con la valoración del estudio del rendimiento académico?</p>	<p>d) Determinar si la capacidad explicativa de los gráficos que genera el modelo basado en el análisis multivariado de datos se relaciona con la valoración del estudio del rendimiento académico.</p>	<p>d) La capacidad explicativa de los gráficos del modelo basado en el análisis multivariado de datos se relaciona con la valoración del estudio del rendimiento académico.</p>	<p>VARIABLE DEPENDIENTE Valoración del estudio del rendimiento académico D - 5. Valoración del estudio del rendimiento académico</p>	

			<p>I - 5.1. Valoración de la capacidad de analizar las asignaturas a su cargo.</p> <p>I - 5.2. Concordancia de los resultados con el conocimiento previo del usuario</p> <p>I - 5.3. Valoración del aporte en la comprensión del rendimiento académico de la escuela profesional.</p> <p>I - 5.4. Percepción sobre relevancia de la información</p> <p>I - 5.5. Valoración del modelo para recomendar su implementación en su universidad.</p>	<ul style="list-style-type: none"> • Recopilación de datos • Tabulación • Coeficiente alfa de Cronbach • Kolmogorov-Smirnov • r de Pearson
--	--	--	--	---

ANEXO 2. Documentación de usuario para aplicar el modelo

INSTRUCTIVO DE APLICACIÓN EN R

A. Fase preliminar: Instalación de R y sus librerías

Instalar el programa R (v.3.5) desde la URL:

<https://cran.r-project.org/bin/windows/base/old/3.5.0/>

Una vez instalado R, se procede a instalar las librerías necesarias para el presente estudio. Se puede realizar fácilmente, ejecutando el siguiente script desde un nuevo documento de R:

```
install.packages('readxl', repos="http://cran.rstudio.com/", dependencies=TRUE)
install.packages('ca', repos="http://cran.rstudio.com/", dependencies=TRUE)
install.packages('RColorBrewer', repos="http://cran.rstudio.com/", dependencies=TRUE)
install.packages('scatterplot3d', repos="http://cran.rstudio.com/", dependencies=TRUE)
install.packages('rgl', repos="http://cran.rstudio.com/", dependencies=TRUE)
install.packages("shiny", repos="http://cran.rstudio.com/", dependencies=TRUE)
install.packages('installr', repos="http://cran.rstudio.com/", dependencies=TRUE)
install.packages('devtools', repos="http://cran.rstudio.com/", dependencies=TRUE)
install.packages('dendextend', repos="http://cran.rstudio.com/", dependencies=TRUE)
install.packages('talgalili', repos="http://cran.rstudio.com/", dependencies=TRUE)
install.packages('rpart.plot', repos="http://cran.rstudio.com/", dependencies=TRUE)
```

Una vez instaladas todas las librerías, ya se puede ejecutar el modelo siguiendo los siguientes pasos:

B. CONFIGURACIÓN DEL ESPACIO DE TRABAJO

- Crear una carpeta con un nombre apropiado. Esta será su carpeta de trabajo.
- Copiar a dicha carpeta el archivo Excel que contiene los datos. El archivo debe tener el nombre de DATOS.xls. Dentro de este archivo deberá tener dos hojas: la primera con las calificaciones de los estudiantes (denominada NOTAS) y la segunda con los datos completos de las asignaturas (denominada ASIGNATURAS)
- Copie a su carpeta de trabajo el ModeloUsuario.R.
- Direcione el espacio de trabajo a la carpeta creada (Menú Archivo → Cambiar dir)
- Abra el archivo ModeloUsuario.R

C. LOS DATOS

Para importar los datos desde el archivo Excel, ejecute la sección de código en R delimitada por el separador

```
# -----
```

```
# 2. CARGA DE DATOS
```

```
# -----
```

En la consola verifique que se han importado los datos de las notas y de las asignaturas, de forma similar a los siguientes gráficos:

IP411	FI357	IF425	EC171	ME305	IF410	IF202	IF053	IF412	LC154	> Asignaturas	Asignatura	Creditos		
10093	11	12	13	11	11	13	11	18	13	11	1	DE151	CONSTITUCION POLITICA DEL PERU Y DERECHOS HUMANOS	2
10013	13	13	12	13	19	12	13	16	11	13	2	EC171	TEORIA ECONOMICA	3
10026	13	11	12	11	NA	12	13	13	11	14	3	ED153	METODOS Y TECNICAS DE APRENDIZAJE	3
10028	14	12	11	12	11	12	11	16	13	14	4	FI261	FISICA I	4
10085	13	12	11	12	12	11	12	14	12	11	5	FI263	FISICA III	4
10090	13	11	13	14	11	12	13	17	14	16	6	FI264	FISICA IV	4
10137	13	12	14	13	14	14	13	15	15	16	7	FI357	MICROELECTRONICA	4
10164	11	11	12	12	11	12	13	14	12	11	8	FF179	FILOSOFIA DE LA CIENCIA	3
10181	15	11	13	15	11	11	11	11	12	12	9	IF001	SEMINARIO EN INFORMATICA	3
102	11	11	12	13	11	11	12	13	17	11	10	IF011	PRACTICAS PRE-PROFESIONALES EN INFORMATICA	3
10211	12	13	15	13	12	12	12	13	13	11	11	IF053	ACTIVIDADES DE PRODUCCION DE BIENES Y PRESTACION DE SERVICIOS	1
10212	14	15	14	12	14	17	13	16	15	14	12	IF101	INTRODUCCION A LA INFORMATICA	4
10227	12	13	11	14	11	11	12	20	13	12	13	IF202	SISTEMAS DE INFORMACION II	4
10231	15	12	11	12	12	11	13	13	14	13	14	IF203	ADMINISTRACION DE CENTROS DE COMPUTO	4
10238	13	13	15	11	11	13	11	13	14	14	15	IF209	INGENIERIA DE SOFTWARE	4
1024	17	13	15	12	12	17	11	17	15	12	16	IF210	PLANEAMIENTO ESTRATEGICO	4
10283	11	11	11	13	13	11	12	17	12	15	17	IF231	SISTEMAS DE INFORMACION I	4
10293	12	12	12	11	11	13	14	13	13	12	18	IF301	SISTEMAS DE BASES DE DATOS I	4

Ilustración 1. Vista parcial de la importación de notas y asignaturas

Observe con detenimiento si los datos de la matriz de notas están completos o si presentan datos faltantes, signados con el marcador NA.

D. IMPUTACION DE DATOS FALTANTES

Si la matriz de notas presentara datos NA, ejecute el código marcado con el separador

```
# -----
```

```
# 3. IMPUTACION DE DATOS FALTANTES
```

```
# -----
```

Las notas faltantes serán reemplazadas por las medias de las asignaturas a las que pertenezcan cada una de dichas notas.

La instrucción (`Notas.Inicial[c(1:30),c(1:10)]`) permite verificar que el proceso se haya realizado correctamente. Los valores 30 y 10 sirven para recortar la visualización de la matriz (30 alumnos y 10 asignaturas). Si desea, puede modificar dichos valores para ampliar la visibilidad de la matriz.

Un ejemplo de la matriz corregida se ve en la siguiente ilustración:

```
> (Notas.Inicial[c(1:30),c(1:10)])
```

	IF411	FI357	IF425	EC171	ME305	IF410	IF202	IF053	IF412	LC154
10003	11	12	13	11	11	13	11	18	13	11
10013	13	13	12	13	18	12	13	16	11	13
10026	13	11	12	11	12	12	13	13	11	14
10028	14	12	11	12	11	12	11	16	13	14
10085	13	12	11	12	12	11	12	14	12	11
10090	13	11	13	14	11	12	13	17	14	16
10137	13	12	14	13	14	14	13	15	15	16
10164	11	11	12	12	11	12	13	14	12	11
10181	15	11	13	15	11	11	11	11	12	12
102	11	11	12	13	11	11	12	13	17	11
10211	12	13	15	13	12	12	12	13	13	11
10212	14	15	14	12	14	17	13	16	15	14
10227	12	13	11	14	11	11	12	20	13	12

Ilustración 2 Vista parcial de la matriz de notas sin valores perdidos

E. ANALISIS DEL RENDIMIENTO ACADEMICO

El primer paso será realizar un análisis de correspondencias que nos permitirá visualizar una primera relación entre las asignaturas y las calificaciones consignadas en la matriz de notas. Para ello se construirá, como primer paso, una tabla de contingencias.

Ejecute el código signado con el separador:

```
# -----
# 3.1. Creación de tabla de contingencias
# -----
```

Al finalizar, verá en la consola la tabla de contingencias semejante al de la siguiente gráfica

```
> (Notas.Contingencia[,])
```

	11	12	13	14	15	16	17	18	19	20
IF411	138	152	201	162	103	61	25	14	5	5
FI357	354	236	135	87	40	10	3	1	0	0
IF425	217	208	163	119	72	50	27	10	0	0
EC171	187	222	172	128	90	32	17	10	7	1
ME305	412	240	105	53	35	15	3	3	0	0
IF410	156	186	162	141	104	73	29	9	5	1
IF202	200	225	185	154	75	21	6	0	0	0
IF053	31	34	89	143	113	155	78	20	10	193
IF412	111	134	179	170	114	71	49	27	8	3
LC154	158	166	191	138	97	55	28	17	11	4
IF413	89	150	174	144	130	101	53	15	9	1
IF301	212	218	185	131	65	34	13	4	4	0
FI264	337	241	156	79	37	11	4	1	0	0
ME660	393	214	125	60	37	18	10	2	5	2
ME359	461	198	99	58	23	15	7	5	0	0
ME253	400	217	125	69	35	8	5	6	1	0
FI263	371	245	145	55	27	19	2	1	0	0
IF302	231	253	196	108	57	15	6	0	0	0

Ilustración 3 Vista parcial de la tabla de contingencias

Obsérvela cuidadosamente tratando de encontrar datos “curiosos” (asignaturas con mayor cantidad de calificaciones bajas, asignaturas con mayor cantidad de calificaciones altas, asignaturas que no presentan valor significativo en alguna de las calificaciones, etc.) En la ilustración anterior se presenta una asignatura (IF053) con una cantidad inusual de notas altas.

Luego, procedemos a realizar el análisis de correspondencias, y a visualizarlo en un plano cartesiano de dos dimensiones, con el fin de analizar visualmente las relaciones entre notas y asignaturas.

Ejecute el código signado con el separador:

```
#-----
# 3.2 Análisis de correspondencias: Asignaturas y Notas
#-----
```

Se mostrará un gráfico similar al siguiente:

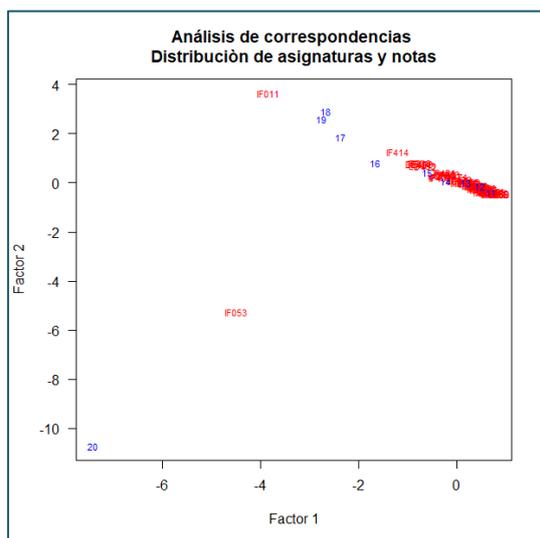


Ilustración 4. Análisis de correspondencias entre asignaturas y notas

Observe que los datos que resaltaban en la tabla de contingencias, son fácilmente identificables en el gráfico, debido a que se encuentran alejados del resto de valores. A estos valores se los conoce como outliers (valores atípicos). En la ilustración anterior se pueden identificar claramente dos outliers: IF053 y IF011.

Analice la importancia de dichas asignaturas y esboce hipótesis del por qué tienen ese comportamiento diferenciado. Este análisis le servirá para decidir, más adelante, si las elimina del estudio o las conserva, ya que los elementos atípicos suelen tener efectos negativos en los resultados del modelo.

Por esa razón, repetiremos el análisis de correspondencias eliminando los outliers.

Para ello digite los valores outliers en la primera línea de código de la sección 3.3, tal como se muestra en el siguiente ejemplo (que recoge los valores outliers del gráfico anterior):

```
Outliers<-c('IF053', 'IF011')
```

Una vez realizado este cambio, ejecute la sección de código delimitada por el separador:

#-----
 # 3.4 Construcción de un mapa de calor para visualizar los datos reordenados
 # en base al primer factor.
 #-----

Podrá visualizar un gráfico similar al siguiente:

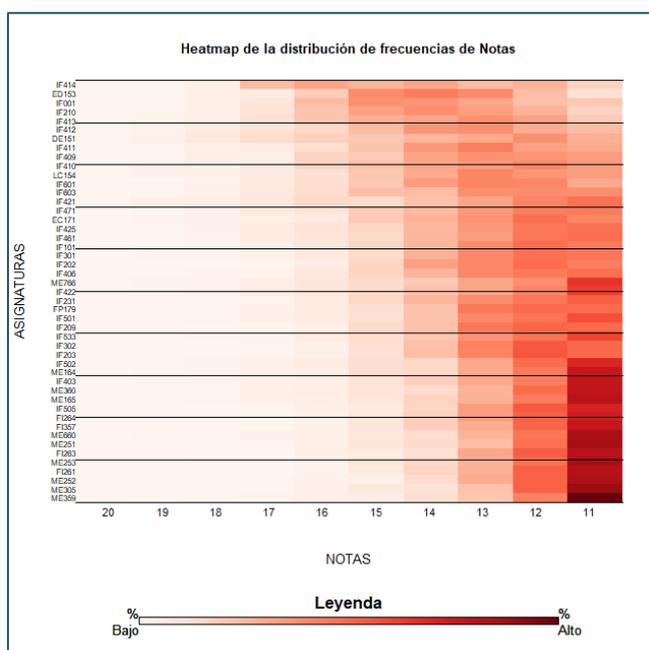


Ilustración 6. Heatmap de la distribución de frecuencia de notas

El color rojo varía de intensidad de claro (baja cantidad de estudiantes) a oscuro (alta concentración de estudiantes). El gráfico muestra en la parte inferior las asignaturas con una alta cantidad de estudiantes con rendimiento bajo. A medida que se escala hacia arriba, encontraremos asignaturas asociadas a mejores calificaciones.

Observe qué asignaturas se encuentran en la base de la gráfica, identifíquelas y asócielas a datos como el área formativa a la que pertenecen, semestre en el que se dicta, creditaje, etc. ¿Concuerda con el conocimiento previo que tenía del rendimiento académico de la escuela profesional? ¿Encuentra información nueva y que sea relevante para usted? ¿En qué parte del mapa se encuentran las asignaturas a su cargo?

Terminados estos procedimientos deberá decidir si continúa el estudio con la matriz de notas original o si lo hace con la matriz de notas que no tiene elementos atípicos. Para ello deberá modificar la primera línea de la siguiente sección de la siguiente manera: si decide realizar el estudio con la matriz original de notas consigne:

Notas<-Notas.Inicial

Si desea realizarlo sin elementos atípicos, consigne la instrucción:

```
Notas<-Notas.SinOut
```

Ejecute el código delimitado por los separadores

```
#-----
```

```
# 4. Punto de decisión para continuar el estudio
```

```
#-----
```

Una vez ejecutado el código de la sección 4 (sin resultados gráficos), realizaremos un análisis de componentes principales (ACP), con el fin de visualizar la distribución de los estudiantes en función de nuevas variables calculadas por el ACP.

Para ello ejecute el código asociado a la sección:

```
#-----
```

```
# 5. ANALISIS DE COMPONENTES PRINCIPALES
```

```
#-----
```

Podrá observar un gráfico de nube de puntos similar al siguiente:

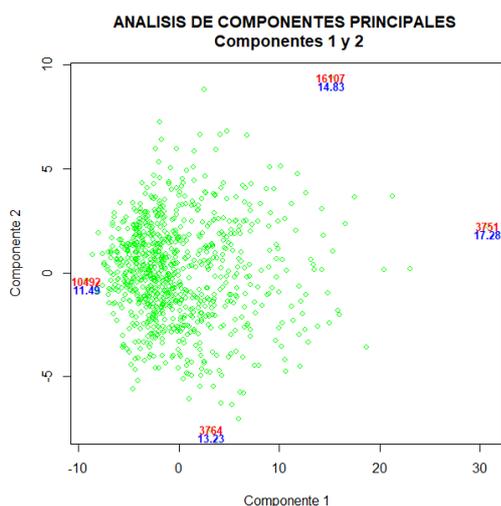


Ilustración 7. Distribución de los estudiantes en base al ACP

En la gráfica cada punto corresponde a un estudiante. Se han etiquetado a los alumnos que ocupan los extremos superior, inferior, izquierdo y derecho de la nube de puntos con sus respectivos promedios aritméticos de calificaciones. Note la relación que existe entre la distribución de las calificaciones de dichos estudiantes y la componente 1: las notas se encuentran ordenadas de izquierda a derecha en la misma dirección que dicha componente. Esto no ocurre con la componente 2 (este aspecto es importante, ya que el estudio continuará tomando como referencia la Componente 1 del ACP)

Observe la forma de la gráfica ¿se pueden observar grupos de puntos definidos? No. La nube de puntos nos da la idea de que una clasificación de estudiantes será complicada de lograr.

A continuación, el estudio prosigue con la sección de Análisis de Conglomerados de estudiantes y asignaturas. Para ello se ordenará la matriz de notas en dos pasos: en el primero se ordenarán los estudiantes (filas) de acuerdo a su promedio aritmético; en el segundo, se ordenarán las asignaturas (columnas) de acuerdo a su correlación con el rendimiento académico. Luego, aplicaremos el algoritmo k-means para lograr un agrupamiento de estudiantes en cuatro grupos.

Ejecute la sección de código que corresponde a la sección

```
#-----
# 6.1. Ordenamiento y clasificación de alumnos y asignaturas
#-----
```

Obtendrá como resultado un histograma similar al mostrado en la siguiente figura:

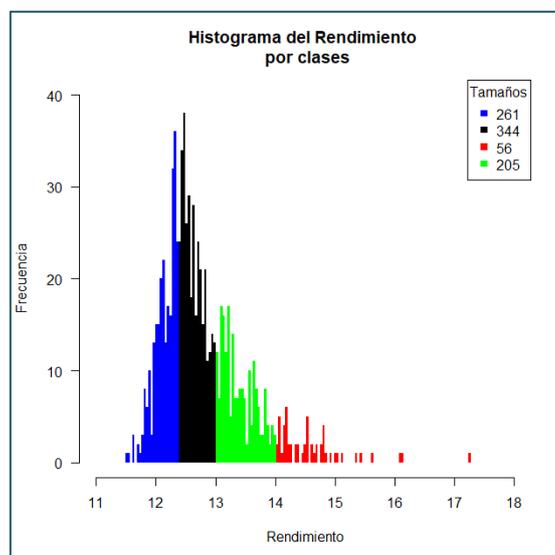


Ilustración 8. Histograma del rendimiento por conglomerados de estudiantes

La figura muestra a los estudiantes agrupados en cuatro categorías: estudiantes de rendimiento bajo (azul), medio bajo (negro), medio alto (verde) y alto (rojo).

En base a estos conglomerados (clusters), realizaremos el proceso de construir los conglomerados de asignaturas utilizando algoritmos de agrupamiento jerárquico.

Ejecute el código correspondiente al separador

```
#-----
# 6.2. Clasificación de Asignaturas a partir de las correlaciones
```

Podrá visualizar un dendograma similar al de la figura siguiente:

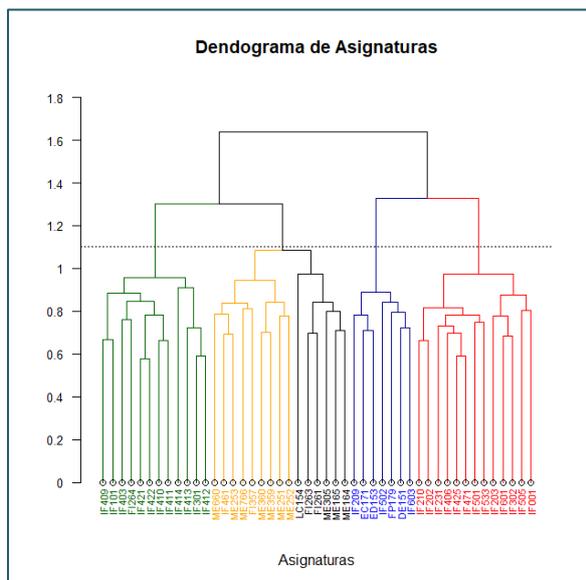


Ilustración 9. Dendograma de asignaturas

La figura nos muestra 5 grupos de asignaturas que tienen relación con el rendimiento académico de los estudiantes.

Observe cuidadosamente el gráfico ¿Qué asignaturas forman cada uno de los grupos? ¿Pertencen a la misma área formativa? ¿Con qué asignaturas se encuentran agrupadas las materias a su cargo?

El siguiente paso será ordenar la información de acuerdo a los conglomerados de asignaturas que se lograron en la etapa anterior.

Para ello ejecute el bloque correspondiente al separador

6.3. Un nuevo ordenamiento en base a los cluster de Asignaturas

El resultado será un nuevo mapa de calor que combina los clusters de asignaturas con los de alumnos. El gráfico debe ser similar al siguiente:

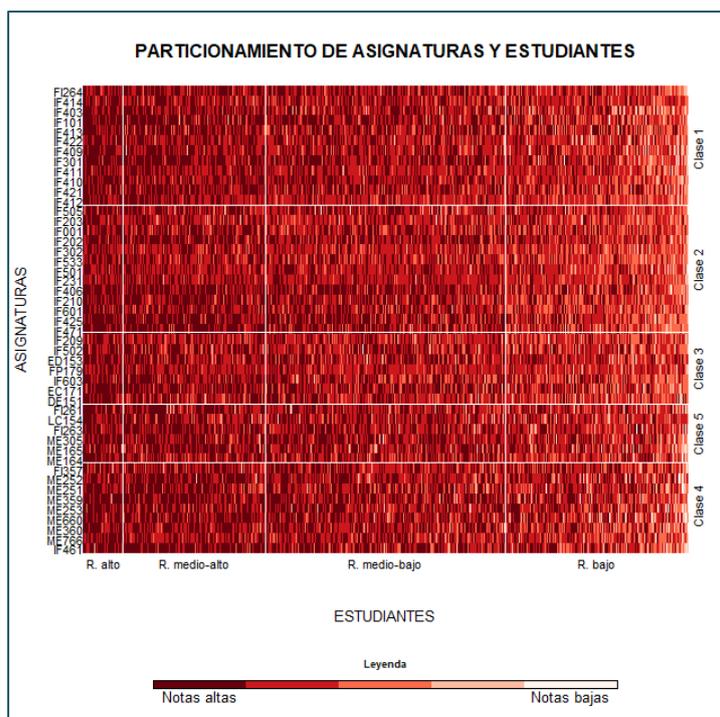


Ilustración 10. Heatmap de notas de acuerdo a clusters de alumnos y asignaturas

La figura muestra los conglomerados de asignaturas cruzados con los conglomerados de alumnos. Cada rectángulo coloreado corresponde al promedio aritmético de un estudiante. Se observa un ligero degradado de color de izquierda a derecha, aunque la cantidad de puntos dificulta su lectura.

Para mejorar la visibilidad del particionamiento, se construirá un gráfico de tamaño 4x5, asignando a cada sector la media de todos los estudiantes que la componen. Para lograr ello ejecute el código de la sección

```
#-----
# 6.4. Visualización de los clusters por sectores
#-----
```

El resultado será un gráfico análogo al anterior, pero con una uniformización de los colores de cada sector de acuerdo al promedio de los estudiantes que lo componen. El gráfico debe ser similar al siguiente:

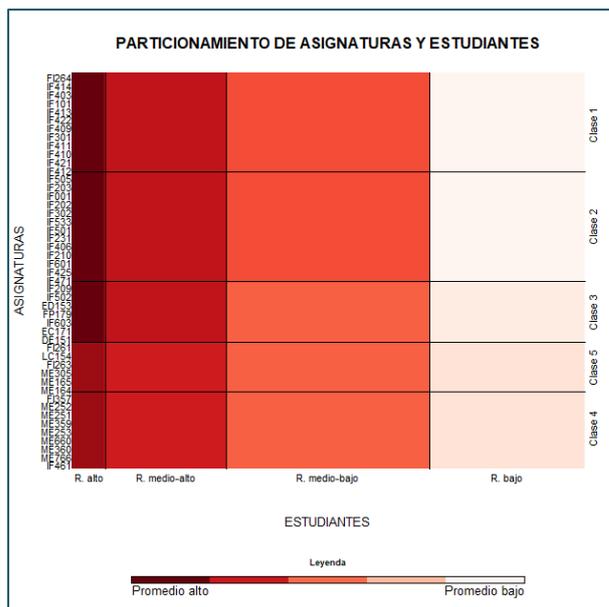


Ilustración 11. Heatmap de los clusters por sectores

Como se puede observar las asignaturas de la clase superior son las que aglutinan a la mayor cantidad de alumnos con buen rendimiento, mientras que las de la parte inferior tienen menos cantidad de esos alumnos. Lo mismo ocurre con el resto de franjas de rendimiento medio-alto, medio-bajo y bajo.

Observe cuidadosamente el gráfico e identifique a las asignaturas que se encuentran en la clase situada en la parte superior y repita el análisis de manera descendente con el resto de clases.

Identifique en qué parte de la gráfica se encuentran las asignaturas a su cargo.

El siguiente paso será aplicar los árboles de clasificación y regresión con el fin de entender la conformación de los clusters de alumnos y asignaturas.

En primer lugar, observaremos la clasificación de los estudiantes en función de las asignaturas. Para ello, ejecute la sección

```
# -----
# 7.1. Árbol de clasificación de estudiantes en función de asignaturas
# -----
```

Se observará un árbol similar al de la figura siguiente:

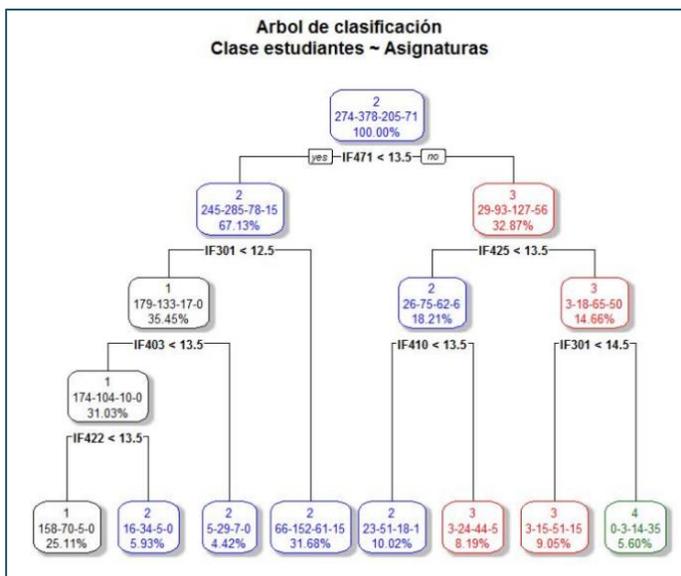


Ilustración 12. Árbol de clasificación de estudiantes en función de asignaturas

Observe las asignaturas que se encuentran en los niveles superiores, identifíquelas y contraste los resultados con el heatmap anterior ¿Cuáles son las asignaturas que son determinantes en la clasificación de los estudiantes?

Ahora, construiremos un árbol de regresión entre las asignaturas y el rendimiento académico.

Para ello ejecute la sección de código 7.2

```
# -----
# 7.2 Árbol de regresión del rendimiento
# -----
```

El resultado será similar a la siguiente figura:

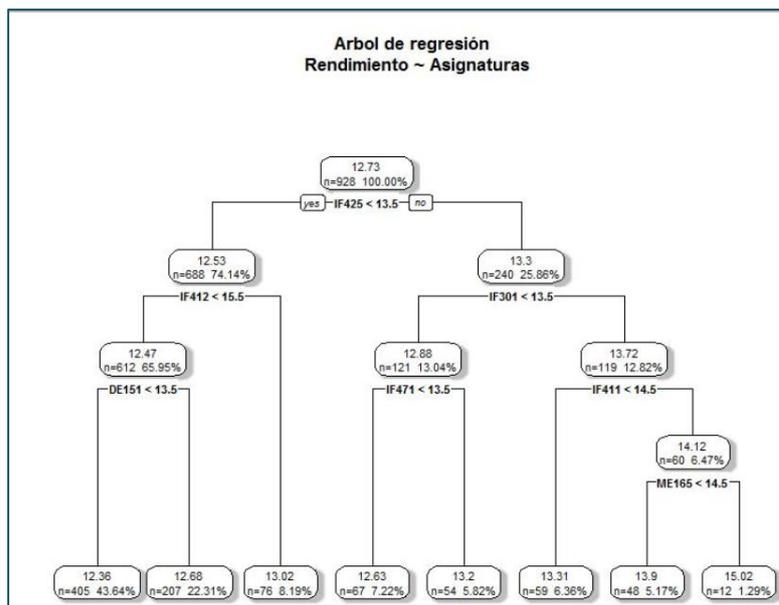


Ilustración 13. Árbol de regresión en base a las asignaturas

Nuevamente, observe la estructura del árbol e identifique las asignaturas que discriminan el rendimiento académico de los estudiantes. Compare los resultados con el último heatmap y con el árbol de clasificación anterior ¿Son las mismas asignaturas? ¿Pertencen en ambos árboles al mismo clúster? ¿Se encuentran sus materias dentro de las asignaturas más discriminantes?

Finalmente, construiremos un árbol de clasificación que nos permita observar el papel de los clusters de asignaturas en la determinación de los conglomerados de los estudiantes.

Ejecute el código de la sección

```
# -----
# 7.3 Árbol de clasificación - (CLASE_ASIGNATURAS + CLASE_ALUMNOS)
# -----
```

Obtendrá una gráfica similar a la siguiente:

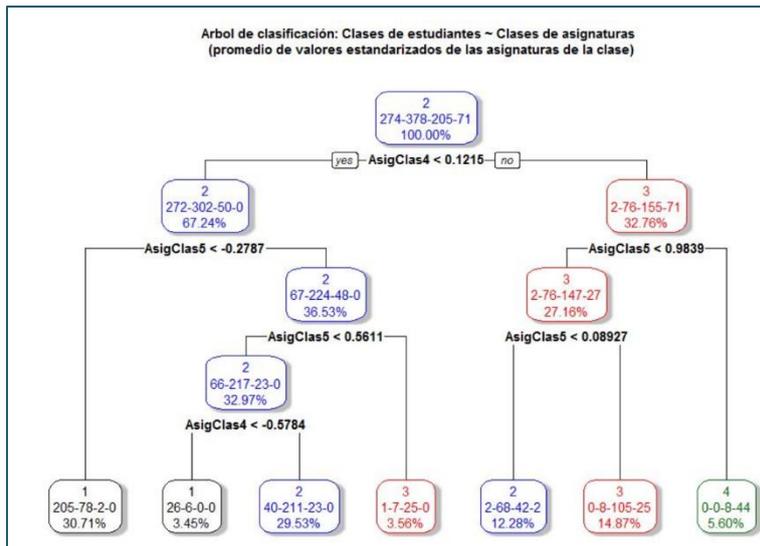


Ilustración 14. Árbol de clasificación de los estudiantes en función de los clusters de asignaturas

Realice un análisis similar al anterior ¿Cuáles son los clusters de asignaturas más determinantes en la clasificación de las estudiantes?

ANEXO 3. Código en R para prueba del modelo

```

# -----
# 1. CONFIGURACION DEL ESPACIO DE TRABAJO
# -----
# Dirreccione el directorio de trabajo hacia la carpeta que contiene el presente
# archivo y en el cual debe estar el archivo Excel con los datos.
# Opcion en el menú: Archivo\Cambiar dir...
# -----
# 2. CARGA DE DATOS
# -----
library(readxl)
Notas.Inicial<-read_excel("DATOS.xlsx", sheet = "NOTAS", col_names = TRUE, na =
"NULL")
CodAlumnos<-Notas.Inicial$ALUMNO
CodAsignaturas<-names(Notas.Inicial)
Notas.Inicial<-as.matrix(Notas.Inicial[,c(2:ncol(Notas.Inicial))])
names(Notas.Inicial)<-CodAsignaturas
rownames(Notas.Inicial)<-CodAlumnos
Asignaturas<-read_excel("DATOS.xlsx", sheet = "ASIGNATURAS", col_names = TRUE,
na = "NULL")
Asignaturas<-as.data.frame(Asignaturas)
#
(Notas.Inicial[c(1:30),c(1:10)])
Asignaturas
# -----
# 3. IMPUTACION DE DATOS FALTANTES
# -----
for(j in 1:ncol(Notas.Inicial))
{
  Prom<-round(mean(Notas.Inicial[,j], na.rm=T))
  Notas.Inicial[is.na(Notas.Inicial[,j]),j] <- Prom
}

```

```

#
(Notas.Inicial[c(1:30),c(1:10)])
#-----
#-- 3. ANALISIS DE CORRESPONDENCIAS
#-----
# 3.1. Creacion de tabla de contingencias
# -----

#-- Creacion de tabla de perfiles (dataframe) por Notas
Perfiles<- as.data.frame(matrix(rep(0,ncol(Notas.Inicial)*10), ncol=10))
for(j in 1:ncol(Notas.Inicial))
{
    x<-factor(Notas.Inicial[,j],levels=c(11:20))
    Perfiles[j,] <-as.factor(table(x))
}
#
# tabla de contingencias o perfiles (convertidos a valores numéricos)
#
Notas.Contingencia<-as.data.frame(matrix(rep(0,ncol(Notas.Inicial)*10), ncol=10))
for(j in 1:ncol(Notas.Inicial))
{
    Notas.Contingencia[j,] <-as.numeric(as.character(Perfiles[j,]))
}
rownames(Notas.Contingencia)<-colnames(Notas.Inicial)
colnames(Notas.Contingencia)<-c(11:20)
(Notas.Contingencia[,])
#-----
# 3.2 Anàlisis de correspondencias: Asignaturas y Notas
#-----
library(ca)
Notas.AC <- ca(Notas.Contingencia)
# gráfica de inercias
#coordenadas de Asignaturas

```

```

XMaxAsig<-max(Notas.AC$rowcoord[,1]); XMinAsig<-min(Notas.AC$rowcoord[,1]);
YMaxAsig<-max(Notas.AC$rowcoord[,2]); YMinAsig<-min(Notas.AC$rowcoord[,2])
#coordenadas de Notas
XMaxNotas<-max(Notas.AC$colcoord[,1]); XMinNotas<-min(Notas.AC$colcoord[,1]);
YMaxNotas<-max(Notas.AC$colcoord[,2]); YMinNotas<-min(Notas.AC$colcoord[,2])
#dimensiones del grafico
XMax<-max(XMaxAsig, XMaxNotas); XMin<-min(XMinAsig, XMinNotas); YMax<-
max(YMaxAsig, YMaxNotas); YMin<-min(YMinAsig, YMinNotas)
#
plot(Notas.AC$rowcoord[,1:2], type="n", xlim=c(XMin, XMax), ylim=c(YMin, YMax),
     las=1, xlab="Factor 1", ylab="Factor 2",
     main="Análisis de correspondencias \nDistribución de asignaturas y notas")
text(Notas.AC$rowcoord[,1:2], labels=rownames(Notas.Contingencia), col="red", cex= 0.7)
text(Notas.AC$colcoord[,1:2], labels=colnames(Notas.Contingencia), col="blue", cex= 0.7)
#-----
# visualizar información de outliers
Outliers<-c('IF053', 'IF011')
#Outliers<-c('IF053', 'IF011','IF414','ED153')
for(j in 1:length(Outliers)) print(Asignaturas[which(Asignaturas == Outliers[j]),])
for(j in 1:length(Outliers))
print(Notas.Contingencia[which(rownames(Notas.Contingencia)== Outliers[j]),])
#-----
# 3.3 Anàlisis de correspondencias sin elementos outlier
#-----
# eliminar elementos outliers
Outliers<-c('IF053', 'IF011')
Notas.SinOut <- Notas.Inicial
Notas.Cont.SinOut <- Notas.Contingencia
for(j in 1:length(Outliers))
{
    Notas.Cont.SinOut<-Notas.Cont.SinOut[-
which(rownames(Notas.Cont.SinOut)==Outliers[j]),]
    Notas.SinOut<-Notas.SinOut[,-which(colnames(Notas.SinOut)==Outliers[j])]
}

```

```

}
# análisis de correspondencias sin elementos outliers
#Notas.AC.SinOut <- ca(Notas.Cont.SinOut)
#-- gráfica de inercias sin elementos perturbadores
#-- coordenadas de Asignaturas
XMaxAsig<-max(Notas.AC.SinOut$rowcoord[,1]); XMinAsig<-
min(Notas.AC.SinOut$rowcoord[,1]); YMaxAsig<-max(Notas.AC.SinOut$rowcoord[,2]);
YMinAsig<-min(Notas.AC.SinOut$rowcoord[,2])
#-- coordenadas de Notas
XMaxNotas<-max(Notas.AC.SinOut$colcoord[,1]); XMinNotas<-
min(Notas.AC.SinOut$colcoord[,1]); YMaxNotas<-max(Notas.AC.SinOut$colcoord[,2]);
YMinNotas<-min(Notas.AC.SinOut$colcoord[,2])
#-- dimensiones del grafico
XMax<-max(XMaxAsig, XMaxNotas); XMin<-min(XMinAsig, XMinNotas); YMax<-
max(YMaxAsig, YMaxNotas); YMin<-min(YMinAsig, YMinNotas)
plot(Notas.AC.SinOut$rowcoord[,1:2], type="n", xlim=c(XMin, XMax), ylim=c(YMin,
YMax),
     las=1, xlab="Factor 1", ylab="Factor 2",
     main="Análisis de correspondencias \nDistribución de asignaturas y notas\n(sin
Asignaturas outliers)")
text(Notas.AC.SinOut$rowcoord[,1:2], labels=rownames(Notas.Cont.SinOut), col="red",
cex= 0.7)
text(Notas.AC.SinOut$colcoord[,1:2], labels=colnames(Notas.Cont.SinOut), col="blue",
cex= 0.7)
#-----
# 3.4 Construcción de un mapa de calor para visualizar los datos reordenados
#   en base al primer factor.
#-----
# Análisis de acuerdo a porcentajes y Factor 1
#-- Elaboración de tabla de porcentajes con asignaturas ordenadas de acuerdo a factor 1
ReOrdenado <- sort.int(Notas.AC.SinOut$rowcoord[,1],
                      , index.return=TRUE, decreasing=TRUE)
Notas2<-Notas.Cont.SinOut[ReOrdenado$ix,]

```

```

SumaFilas <- apply(Notas2, 1, sum)
(PorNotas<-round(Notas2/SumaFilas*100,2))
PorNotas[1:5,]
# -- Gráfico de la tabla de contingencias (ordenada) expresada en porcentajes
library("RColorBrewer")
#-- definir layout
def.par <- par(no.readonly = TRUE)
a<-matrix(c(1,1,2,2), 2, 2, byrow = TRUE)
lay<-layout(a,widths=c(1,8), heights=c(8,1), respect=TRUE)
#
tPorNotas<-t(rev(PorNotas))[1:ncol(PorNotas),1:nrow(PorNotas)]
NroColres<-200
Colores <- colorRampPalette(brewer.pal(9,"Reds"))(NroColres)
image(as.matrix(tPorNotas), axes=FALSE, ylab="ASIGNATURAS", xlab="NOTAS",
col=Colores)
title("Heatmap de la distribución de frecuencias de Notas", cex.main=1)
mtext(20:11, side=1, adj=(0:9)/10+0.05, cex=0.7)
posiciones <- seq(0.01,1.01, length.out=length(colnames(tPorNotas)))
mtext(colnames(tPorNotas), side=2, padj=1, at=posiciones, cex=0.5, las=2)
plineas <- seq(-0.01,1.01, length.out=length(colnames(tPorNotas))/5+1)
abline(h=plineas, col="black")
# leyenda
par(mar=c(2,8,2,6))
Colores <- colorRampPalette(brewer.pal(9,"Reds"))(NroColres)
image(as.matrix(1:NroColres),col=Colores, axes=FALSE, main = "Leyenda")
mtext(side=2,"%\nBajo", las=2, adj=c(1,1),cex=0.8)
mtext(side=4,"%\nAlto", las=2, adj=c(0,0),cex=0.8)
box(col="black")
par(def.par) #- reset to default
# para visualizar la información de una asignatura cambie ponga su código en la sig. línea:
#print(Asignaturas[which(Asignaturas == "CODIGO"),])
#-----

```

```

# 4. PUNTO DE DECISIÓN PARA CONTINUAR CON EL ESTUDIO
#-----
# Decidir el conjunto de datos a analizar y calcular promedios aritméticos de estudiantes
Notas<-Notas.SinOut
Promedios<- round(apply(Notas, 1, mean),2)
#-----

# 5. ANALISIS DE COMPONENTES PRINCIPALES
#-----

acp<-prcomp(Notas, scale=FALSE)
coord<-acp$x[,1:3]
#-- Distribución de alumnos respecto a CP1 y CP2 (con rendimiento de algunos estudiantes)
# selección de los alumnos en los extremos
AlumnoDerecha <- rownames(Notas)[which(coord[,1]==max(coord[,1]))]
AlumnoIzquierda <- rownames(Notas)[which(coord[,1]==min(coord[,1]))]
AlumnoArriba<- rownames(Notas)[which(coord[,2]==max(coord[,2]))]
AlumnoAbajo <- rownames(Notas)[which(coord[,2]==min(coord[,2]))]
Alumnos <- c(AlumnoDerecha , AlumnoIzquierda, AlumnoArriba, AlumnoAbajo)
# cálculo de los promedios y gráfico en función de componentes 1 y 2
Promedios<- round(apply(Notas, 1, mean),2)
Pos<-c(1:length(Alumnos))
for(j in 1:length(Alumnos)) { Pos[j]<-which(rownames(Notas)==Alumnos[j]) }
plot (coord[,1], coord[,2], xlab="Componente 1", ylab="Componente 2",
      main = "ANALISIS DE COMPONENTES PRINCIPALES \n Componentes 1 y 2",
      col="green", cex=0.8)
text (coord[Pos,1], coord[Pos,2], rownames(Notas[Pos,]), col="red", font=2, cex=0.8)
text (coord[Pos,1], coord[Pos,2]-0.4, Promedios[Pos], col="blue", font=2, cex=0.8)
#-----

# 6. ANALISIS DE CONGLOMERADOS DE ESTUDIANTES Y ASIGNATURAS
#-----
#-----

# 6.1. Ordenamiento y clasificación de alumnos y asignaturas
#-----

#-- ordenando estudiantes en base a Rendimiento

```

```

Rendimientos<-Promedios
OrdenEst <- sort.int(Rendimientos,index.return=TRUE, decreasing=FALSE)
Notas3 <- Notas[OrdenEst$ix,]
Rendimientos<-Rendimientos[OrdenEst$ix]
#-- ordenando Asignaturas de acuerdo a su correlaciòn con el Rendimiento
cNotasasig<-cor(Notas3, Rendimientos)
OrdenAsig <- sort.int(cNotasasig, index.return = TRUE, decreasing=TRUE)
Notas3 <-Notas3[,OrdenAsig$ix]
# clasificaciòn de los estudiantes respecto al Rendimiento acadèmic (kmeans)
NroClusters<-4
Rend<-Rendimientos
kmRend <- kmeans(Rend,centers=NroClusters, nstart=10)
# histograma de clasificaciòn de estudiantes en base a Rendimiento (Promedio aritmetico)
intlength=150; maxy=40; maxX=18
#Colores<-ifelse(kmRend$cluster==1, "green", ifelse(kmRend$cluster==2,"red",
ifelse(kmRend$cluster==3,"blue", "black")))
Colores=c("blue","black","red","green")
par(mfrow=c(1,1))
hist(Rend,
      xlim=range(c(11:maxX)), ylim=c(0,maxy), las=1,
      main="Histograma del Rendimiento \n por clases",
      ylab="Frecuencia", xlab="Rendimiento",
      breaks=seq(min(Rend),max(Rend), length=intlength),
      col="transparent", border="transparent")
hist(Rend[kmRend$cluster==1],
      xlim=range(c(11:maxX)), ylim=c(0,maxy), axes=FALSE,
      breaks=seq(min(Rend),max(Rend), length=intlength),
      col=Colores[1], border="transparent", add=TRUE)
hist(Rend[kmRend$cluster==2],
      xlim=range(c(11:maxX)), ylim=c(0,maxy), axes=FALSE,
      breaks=seq(min(Rend),max(Rend), length=intlength),
      col=Colores[2], border="transparent", add=TRUE)
hist(Rend[kmRend$cluster==3],

```

```

xlim=range(c(11:maxX)), ylim=c(0,maxy), axes=FALSE,
breaks=seq(min(Rend),max(Rend), length=intlength),
col=Colores[3], border="transparent", add=TRUE)
hist(Rend[kmRend$cluster==4],
xlim=range(c(11:maxX)), ylim=c(0,maxy), axes=FALSE,
breaks=seq(min(Rend),max(Rend), length=intlength),
col=Colores[4], border="transparent", add=TRUE)
legend(x="topright", legend=kmRend$size,title="Tamaños",
pch=15, col=Colores)
# -----
# 6.2. Clasificación de Asignaturas a partir de las correlaciones
# -----
# determinar los estudiantes de cada clases (de Rendimiento alto a Rendimiento bajo)
PartEstudiantes<-sort(c(max(Rendimientos[kmRend$cluster==1]),
max(Rendimientos[kmRend$cluster==2]),
max(Rendimientos[kmRend$cluster==3]),
max(Rendimientos[kmRend$cluster==4])))
est1<-rownames(Notas3)[Rendimientos > PartEstudiantes[3]]
# Rendimiento alto
est2<-rownames(Notas3)[which(Rendimientos <= PartEstudiantes[3] & Rendimientos >
PartEstudiantes[2])] # Rendimiento medio-alto
est3<-rownames(Notas3)[which(Rendimientos <= PartEstudiantes[2] & Rendimientos >
PartEstudiantes[1])] # Rendimiento medio-bajo
est4<-rownames(Notas3)[which(Rendimientos <= PartEstudiantes[1])]
# Rendimiento bajo
OrdenEst<-c(rev(est1), rev(est2), rev(est3), rev(est4))
ListaEst<-list(rev(est1), rev(est2), rev(est3), rev(est4))
#
DistanciaV<- 1 - cor(Notas3)
DistanciaV<- as.dist(DistanciaV)
hclusassign <- hclust(DistanciaV, method="ward.D")
tree <- cutree(hclusassign, k=5)
#

```

```

# crear Dendograma coloreado
Dendograma = as.dendrogram(hclusasign)
Colores<-c("red","darkgreen","orange","blue","black")
ColEtiquetas<- function(n)
{
  if (is.leaf(n))
  {
    a <- attributes(n)
    labCol <- Colores[tree[which(names(tree) == a$label)]]
    attr(n, "nodePar") <- c(a$nodePar, list(lab.col = labCol, lab.cex=0.7))
  }
  n
}
ClusDendro = dendrapply(Dendograma , ColEtiquetas)
#-- colorear ramas
library('installr'); library('devtools'); library('dendextend'); require('tagalili')
#-- colorear ramas
d1=color_branches(ClusDendro,k=5, col =
c("darkgreen","orange","black","darkblue","red"))
plot(d1,cex=0.5, main = "Dendograma de Asignaturas", xlab="Asignaturas",
ylab="",ylim=c(0,1.8), axes=FALSE)
axis(side=2, at=seq(0,1.8,by=0.2), labels=seq(0,1.8,by=0.2), las=2, cex.axis=0.8)
abline(h=1.1, lwd = 1, lty = 3)
#-----
# 6.3. Un nuevo ordenamiento en base a los cluster de Asignaturas
#-----
(asighc1<-names(tree[which(tree==1)]))
(asighc2<-names(tree[which(tree==2)]))
(asighc3<-names(tree[which(tree==3)]))
(asighc4<-names(tree[which(tree==4)]))
(asighc5<-names(tree[which(tree==5)]))
# -- ordenar cada cluster
acp<-prcomp(Notas, scale=FALSE)

```

```

asighc1<-names(sort(acp$rotation[asighc1,1], decreasing=TRUE))
asighc2<-names(sort(acp$rotation[asighc2,1], decreasing=TRUE))
asighc3<-names(sort(acp$rotation[asighc3,1], decreasing=TRUE))
asighc4<-names(sort(acp$rotation[asighc4,1], decreasing=TRUE))
asighc5<-names(sort(acp$rotation[asighc5,1], decreasing=TRUE))
#-- ordenar clusters de Asignaturas (de acuerdo al Promedio de la primera componente)
Promedios.clusters<-
c(mean(acp$rotation[asighc1,1]),mean(acp$rotation[asighc2,1]),mean(acp$rotation[asighc3,
1]),mean(acp$rotation[asighc4,1]),mean(acp$rotation[asighc5,1]))
ListaAsig<-list(asighc1,asighc2,asighc3,asighc4,asighc5)
for (i in 1:(length(Promedios.clusters)-1))
{
  for (j in (i+1):length(Promedios.clusters))
  {
    if (Promedios.clusters[i]<Promedios.clusters[j])
    {
      aux<-Promedios.clusters[i]; Promedios.clusters[i]<-
Promedios.clusters[j]; Promedios.clusters[j]<-aux
      aux<-ListaAsig[i]; ListaAsig[i]<-ListaAsig[j]; ListaAsig[j]<-aux
    }
  }
}
OrdenAsig<-c(ListaAsig[[1]],ListaAsig[[2]],ListaAsig[[3]],ListaAsig[[4]],ListaAsig[[5]])
# ordenar nuevamente lo clusters (para obtener un mejor coloreado y significado).
# De arriba hacia abajo: clases 1-2-3-5-4
OrdenAsig<-c(ListaAsig[[4]],ListaAsig[[5]],ListaAsig[[3]],ListaAsig[[2]],ListaAsig[[1]])
ListaAsig<-list(ListaAsig[[4]],ListaAsig[[5]],ListaAsig[[3]],ListaAsig[[2]],ListaAsig[[1]])
#
Notas.Contingencia<-scale(Notas3)
Notas.Contingencia3 <- Notas.Contingencia[,OrdenAsig]
#
# gráfico Asignaturas x Estudiantes
#

```

```

library("RColorBrewer")
def.par <- par(no.readonly = TRUE)
a<-matrix(c(1,1,2,2), 2, 2, byrow = TRUE)
lay<-layout(a,widths=c(1,8), heights=c(8,1), respect=TRUE)
NroColres<-5
Colores <- colorRampPalette(rev(brewer.pal(9,"Reds")))(NroColres)
image(as.matrix(Notas.Contingencia3), axes=FALSE, ylab="ASIGNATURAS",
xlab="ESTUDIANTES", col=Colores, main="PARTICIONAMIENTO DE
ASIGNATURAS Y ESTUDIANTES", cex=0.8)
Posiciones <- seq(0.01,1.01, length.out=length(colnames(Notas.Contingencia3)))
mtext(colnames(Notas.Contingencia3), side=2, padj=1, at=Posiciones , cex=0.7, las=2)
# Particionamiento de estudiantes
labelRendimiento<-c("R. alto", "R. medio-alto", "R. medio-bajo", "R. bajo")
#ce<-
c(length(est1),length(est2),length(est3),length(est4))/length(rownames(Notas.Contingencia3)
)
ce<-
c(length(ListaEst[[1]]),length(ListaEst[[2]]),length(ListaEst[[3]]),length(ListaEst[[4]]))/lengt
h(rownames(Notas.Contingencia3))
ce<-c(ce[1], ce[1]+ce[2],ce[1]+ce[2]+ce[3])
abline(v=ce, col="white")
mtext(labelRendimiento, side=1, at=c(ce[1]/2,(ce[2]+ce[1])/2,(ce[3]+ce[2])/2,(1+ce[3])/2),
cex=0.7)
# particionamiento de Asignaturas
ca<-
c(length(ListaAsig[[1]]),length(ListaAsig[[2]]),length(ListaAsig[[3]]),length(ListaAsig[[4]]),
length(ListaAsig[[5]]))/length(colnames(Notas.Contingencia3))
ca<-c(ca[1], ca[1]+ca[2], ca[1]+ca[2]+ca[3]+0.008, ca[1]+ca[2]+ca[3]+ca[4]+0.01)-0.004
abline(h=ca, col="white")
labelAsignaturas<-c("Clase 4", "Clase 5", "Clase 3", "Clase 2", "Clase 1")
mtext(labelAsignaturas, side=4,
at=c(ca[1]/2,(ca[1]+ca[2])/2,(ca[2]+ca[3])/2,(ca[3]+ca[4])/2,(1+ca[4])/2), cex=0.7)
# leyenda

```

```

par(mar=c(2,8,2,6))
image(1:NroColres,1,as.matrix(1:NroColres), col=Colores, xlab="Leyenda", ylab="",
xaxt="n", yaxt="n", bty="n", ylim=c(0,10))
title("Leyenda",cex.main=0.7)
mtext(c("Notas altas","Notas bajas"), side=1, at=c(1,5), cex=0.8)
box(col="black")
par(def.par) #- reset to default
#-----
# 6.4. Visualización de los clúster por sectores
#-----
# gráfico coloreado por sectores (4x5)
CortesX<-c(0,length(ListaEst[[1]]),length(ListaEst[[1]])+length(ListaEst[[2]]),
           length(ListaEst[[1]])+length(ListaEst[[2]])+length(ListaEst[[3]]),
           length(ListaEst[[1]])+length(ListaEst[[2]])+length(ListaEst[[3]])+length(ListaEst[[4]]
           ))

CortesY<-c(0,length(ListaAsig[[1]]),length(ListaAsig[[1]])+length(ListaAsig[[2]]),
           length(ListaAsig[[1]])+length(ListaAsig[[2]])+length(ListaAsig[[3]]),
           length(ListaAsig[[1]])+length(ListaAsig[[2]])+length(ListaAsig[[3]])+length(ListaAsig[[4]]),
           length(ListaAsig[[1]])+length(ListaAsig[[2]])+length(ListaAsig[[3]])+length(ListaAsig[[4]])+length(ListaAsig[[5]])
           )
#-- calcular la media de cada sector y asignarlo a todas sus celdas individuales
Notas.Contingencia4<-Notas.Contingencia3
for (i in 1:(length(CortesX)-1))
{
  for (j in 1:(length(CortesY)-1))
  {
    media<-
mean(Notas.Contingencia3[(CortesX[i]+1):CortesX[i+1]],[CortesY[j]+1]:CortesY[j+1]])

    Notas.Contingencia4[(CortesX[i]+1):CortesX[i+1]],[CortesY[j]+1]:CortesY[j+1]]<-
media

```

```

# definir layout
def.par <- par(no.readonly = TRUE)
a<-matrix(c(1,1,2,2), 2, 2, byrow = TRUE)
lay<-layout(a,widths=c(1,8), heights=c(8,1), respect=TRUE)
#
#-- crear gráfico coloreado
#
NroColres<-20
Colores <- colorRampPalette(rev(brewer.pal(9,"Reds")))(NroColres)
image(as.matrix(Notas.Contingencia4), axes=FALSE, ylab="ASIGNATURAS",
col=Colores,xlab="ESTUDIANTES", main="PARTICIONAMIENTO DE ASIGNATURAS
Y ESTUDIANTES", cex=0.8)
Posiciones <- seq(0.01,1.01, length.out=length(colnames(Notas.Contingencia3)))
mtext(colnames(Notas.Contingencia3), side=2, padj=1, at=Posiciones , cex=0.7, las=2)
# Particionamiento de estudiantes
labelRendimiento<-c("R. alto", "R. medio-alto", "R. medio-bajo", "R. bajo")
ce<-c(length(ListaEst[[1]]),length(ListaEst[[2]]),length(ListaEst[[3]]),length(ListaEst[[4]])-
0.007)/length(rownames(Notas.Contingencia3))
ce<-c(ce[1], ce[1]+ce[2],ce[1]+ce[2]+ce[3])
abline(v=ce, col="black")
mtext(labelRendimiento, side=1, at=c(ce[1]/2,(ce[2]+ce[1])/2,(ce[3]+ce[2])/2,(1+ce[3])/2),
cex=0.7)
# particionamiento de Asignaturas
ca<-
c(length(ListaAsig[[1]]),length(ListaAsig[[2]]),length(ListaAsig[[3]]),length(ListaAsig[[4]]),
length(ListaAsig[[5]])/length(colnames(Notas.Contingencia3))
ca<-c(ca[1], ca[1]+ca[2], ca[1]+ca[2]+ca[3]+0.008, ca[1]+ca[2]+ca[3]+ca[4]+0.015)-0.004
abline(h=ca, col="black")
labelAsignaturas<-c("Clase 4", "Clase 5", "Clase 3", "Clase 2", "Clase 1")
mtext(labelAsignaturas, side=4,
at=c(ca[1]/2,(ca[1]+ca[2])/2,(ca[2]+ca[3])/2,(ca[3]+ca[4])/2,(1+ca[4])/2), cex=0.7)
# leyenda
par(mar=c(2,8,2,6))

```

```

NroColres<-5
Colores <- colorRampPalette(rev(brewer.pal(9,"Reds")))(NroColres)
image(1:NroColres,1,as.matrix(1:NroColres), col=Colores, xlab="Leyenda", ylab="",
xaxt="n", yaxt="n", bty="n", ylim=c(0,10))
title("Leyenda",cex.main=0.7)
mtext(c("Promedio alto","Promedio bajo"), side=1, at=c(1,NroColres), cex=0.8)
box(col="black")
par(def.par) #- reset to default
# -----
# 7. ÁRBOLES DE CLASIFICACION Y REGRESION
# -----
# 7.1 Árbol de clasificación de estudiantes en función de asignaturas
# -----
library(rpart)
library(rpart.plot)
N<-length(colnames(Notas.clasificacion))-1
Notas.clasificacion<-cbind(Notas3,clase=kmRend$cluster)
Nombres<-paste(colnames(Notas.clasificacion)[1:N])
form<-as.formula(paste("clase ~ ", paste(Nombres, collapse="+")))
Árbol.clasificacion<-rpart(formula = form, method = "class", data
=as.data.frame(Notas.clasificacion), cp=0.0001)
Poda.clasificacion<-prune(Árbol.clasificacion,cp= 0.025)
Colores<-ifelse(Poda.clasificacion$frame$yval==1, "black",
                ifelse(Poda.clasificacion$frame$yval==2,"blue",
                ifelse(Poda.clasificacion$frame$yval==3,"darkgreen",
                "red")))
rpart.plot(Poda.clasificacion, type=2, extra=101,
           digits=4,shadow.col="white",xsep="-", col=Colores, boxes.include.gap=F,
           nn=F, fallen.leaves=TRUE, main = "Árbol de clasificación\n Clase
estudiantes ~ Asignaturas ")
# -----
# 7.2 Árbol de regresión del rendimiento
# -----

```

```

ÁrbolRend<-rpart(Rend ~ ., method = "anova", data = as.data.frame(Notas3), cp=0.0115)
PodaRend<-prune(ÁrbolRend,cp= 0.03)
rpart.plot(PodaRend, type=2, extra=101, digits=4,shadow.col="gray",
           xsep="-", boxes.include.gap=F,nn=F, fallen.leaves=TRUE,
           main = "Árbol de regresión \n Rendimiento ~ Asignaturas")
# -----
# 7.3 Árbol de clasificación - (CLASE_ASIGNATURAS + CLASE_ALUMNOS)
# -----
#-- crear tabla Alumnos * Clase_Asignatura
Notas2 <- as.data.frame(matrix(0, ncol = 5, nrow =
length(rownames(Notas.Contingencia3))))
rownames(Notas2)<-rownames(Notas.Contingencia3)
colnames(Notas2)<-c("AsigClas3","AsigClas2","AsigClas1","AsigClas4","AsigClas5")
for (i in 1:length(rownames(Notas2)))
{
  for (j in 1:(length(CortesY)-1))
  {
    Notas2[i,j]<-(mean(Notas.Contingencia3[i,(CortesY[j]+1):CortesY[j+1]]))
  }
}
#-- agregar Clase_Alumno
#Notas2<-cbind(Notas2,EstClas=acp1km4[,2])
Notas2<-cbind(Notas2,EstClas=kmRend$cluster)
# crear el Árbol de clasificación
Nombres<-paste(colnames(Notas2)[1:5])
form<-as.formula(paste("EstClas ~ ", paste(Nombres, collapse="+")))
Árbol2<-rpart(formula = form, method = "class", data = Notas2, cp=0.0001)
Poda2<-prune(Árbol2, cp= 0.03)
Colores<-ifelse(Poda2$frame$yval==1, "black", ifelse(Poda2$frame$yval==2,"blue",
ifelse(Poda2$frame$yval==3,"red", "darkgreen")))
#Colores<-ifelse(Poda2$frame$yval==1, "red", ifelse(Poda2$frame$yval==2,"darkgreen",
ifelse(Poda2$frame$yval==3,"black", "blue")))

```

```
rpart.plot(Poda2, type=2, extra=101, digits=4,shadow.col="gray",xsep="-", col=Colores,  
nn=F, fallen.leaves=T, main = "")  
title("Árbol de clasificación: Clases de estudiantes ~ Clases de Asignaturas \n (Promedio de  
valores estandarizados de las Asignaturas de la clase)", cex.main=0.9)  
summary(Poda2)
```

ANEXO 4: Cuestionario

El presente cuestionario está orientado a evaluar distintos aspectos sobre el modelo basado en el análisis multivariado de datos para el estudio del rendimiento académico de las escuelas profesionales universitarias. Se le pide contestar con honestidad y cuidado cada una de las preguntas formuladas.

Desde ya, muchas gracias por su colaboración.

Datos generales

Indique su rol dentro de la universidad:

() Docente () Docente con cargo académico () Personal de áreas de TICS

Años de experiencia dentro de la universidad: ()

Instrucciones

Para cada uno de los ítems coloque una **X** en el casillero correspondiente a su nivel de acuerdo o desacuerdo.

Nro. Ítem	Ítem	Muy de acuerdo	De acuerdo	Ni de acuerdo, ni en desacuerdo	En desacuerdo	Totalmente en desacuerdo
1	VALIDEZ DE LOS DATOS DE ENTRADA DEL MODELO					
1.1	El formato inicial de los datos es fácilmente interpretable					
1.2	Los datos iniciales no son fácilmente analizables debido al volumen de información					
1.3	Los datos contemplados en la matriz son adecuados para realizar un estudio del rendimiento académico					
2	USABILIDAD DEL MODELO					
2.1	El proceso de carga de los datos iniciales es sencillo					
2.2	El modelo es fácil de utilizar					
2.3	La interfaz del modelo es amigable					
2.4	La documentación es adecuada para la ejecución del modelo					
2.5	La documentación del modelo permite interpretar adecuadamente sus resultados					
3	FUNCIONALIDAD DEL MODELO					
3.1	El modelo permite obtener resultados de forma rápida					
3.2	El modelo presenta los mismos resultados si se lo ejecuta en repetidas ocasiones con los mismos datos					
3.3	El modelo presenta diferentes formatos para el análisis de la información					
3.4	La información se presenta sin sobrecarga de datos					

Nro. Ítem	Ítem	Muy de acuerdo	De acuerdo	Ni de acuerdo, ni en desacuerdo	En desacuerdo	Totalmente en desacuerdo
3.5	El modelo es versátil y permite evaluar diferentes escenarios					
3.6	La información visualizada tiene el mismo eje temático (rendimiento académico)					
4	CAPACIDAD EXPLICATIVA DE LOS GRAFICOS DEL MODELO					
4.1	Los gráficos permiten analizar las relaciones entre asignaturas y estudiantes					
4.2	Los elementos presentados en los gráficos mejoran el entendimiento de la relación entre calificaciones y asignaturas.					
4.3	Los elementos presentados en los gráficos son claramente distinguibles					
4.4	Los elementos más importantes de los gráficos están claramente resaltados					
4.5	Los gráficos son consistentes unos con otros					
4.6	La información mostrada en los gráficos es suficiente para su entendimiento y relación con otros gráficos					
4.7	Los elementos presentes en los gráficos representan claramente a los datos del rendimiento académico					
4.8	Los tipos de gráficos presentados pueden servir a más de un propósito					
4.9	Los gráficos son una forma útil de mostrar la información					
4.10	Todos los gráficos presentados son necesarios para entender el rendimiento académico					
4.11	Los datos se encuentran adecuadamente colocados en la región de datos					
4.12	La escala mostrada en los datos permite representar adecuadamente los datos					
5	VALORACION DE RESULTADOS DEL ESTUDIO					
5.1	El modelo permite analizar el rendimiento académico de las asignaturas a su cargo					
5.2	Los resultados obtenidos por el modelo son concordantes con su conocimiento previo del rendimiento académico de su escuela profesional					
5.3	El modelo le permite mejorar su comprensión del rendimiento académico frente a otros recursos actualmente disponibles					
5.4	El modelo provee de información relevante para la toma de decisiones					
5.5	El modelo es recomendable para ser utilizado en las unidades académicas de la universidad					

ANEXO 5: Tablas de valoración de instrumento por expertos

TABLA DE VALIDACIÓN DE INSTRUMENTO POR EXPERTOS

Apellidos y Nombres del Informante	Institución donde labora	Nombre del Instrumento	Autor del Instrumento
Dr. Soncco Álvarez José Luis	UNSAAC	CUESTIONARIO	ROBERT WILBERT ALZAMORA PAREDES
Título de la Investigación: MODELO BASADO EN EL ANALISIS MULTIVARIANTE DE DATOS PARA EL ESTUDIO DEL RENDIMIENTO ACADEMICO DE LAS ESCUELAS PROFESIONALES DE LAS UNIVERSIDADES NACIONALES DE LA REGION DELCUSCO.			

I. ASPECTOS DE EVALUACIÓN

		DEFICIENTE				REGULAR				BUENA				MUY BUENA				EXCELENTE			
		5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	100
1. CLARIDAD	Esta formulado con lenguaje apropiado.																			X	
2. OBJETIVIDAD	Esta expresado en conductas observables.																				X
3. ACTUALIZACIÓN	Está adecuado al avance de la ciencia y la tecnología.																				X
4. ORGANIZACIÓN	Esta organizado en forma lógica.																				X
5. SUFICIENCIA	Comprende aspectos cuantitativos																				X
6. INTENCIONALIDAD	Es adecuado para valorar el modelo																				X
7. CONSISTENCIA	Está basado en aspectos teóricos científicos.																			X	
8. COHERENCIA	Entre las variables, dimensiones, Indicadores e ítems.																				X
9. METODOLOGÍA.	La estrategia responde al propósito de la investigación.																				X
10. PERTINENCIA	La escala es aplicable.																				X

II. OPINIÓN DE APLICACIÓN

El cuestionario es válido para medir la investigación

III. PROMEDIO DE VALORACIÓN:

95.5

LUGAR Y FECHA	DNI	FIRMA DEL EXPERTO INFORMANTE	Nº DE TELEFONO
Cusco, 02, Julio 2019	43199486		950358194

TABLA DE VALIDACIÓN DE INSTRUMENTO POR EXPERTOS

Apellidos y Nombres del Informante	Institución donde labora	Nombre del Instrumento	Autor del Instrumento
Dr. Enciso Rodas, Leandro	UNSAAC	CUESTIONARIO	ROBERT WILBERT ALZAMORA PAREDES
Título de la Investigación:			
MODELO BASADO EN EL ANALISIS MULTIVARIANTE DE DATOS PARA EL ESTUDIO DEL RENDIMIENTO ACADEMICO DE LAS ESCUELAS PROFESIONALES DE LAS UNIVERSIDADES NACIONALES DE LA REGION DEL CUSCO.			

I. ASPECTOS DE EVALUACIÓN

		DEFICIENTE				REGULAR				BUENA				MUY BUENA				EXCELENTE			
		5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	100
1. CLARIDAD	Esta formulado con lenguaje apropiado.																		X		
2. OBJETIVIDAD	Esta expresado en conductas observables.																	X			
3. ACTUALIZACIÓN	Está adecuado al avance de la ciencia y la tecnología.																			X	
4. ORGANIZACIÓN	Esta organizado en forma lógica.																				X
5. SUFICIENCIA	Comprende aspectos cuantitativos																		X		
6. INTENCIONALIDAD	Es adecuado para valorar el modelo																				X
7. CONSISTENCIA	Está basado en aspectos teóricos científicos.																	X			
8. COHERENCIA	Entre las variables, dimensiones, Indicadores e ítems.																		X		
9. METODOLOGÍA.	La estrategia responde al propósito de la investigación.																				X
10. PERTINENCIA	La escala es aplicable.																				X

II. OPINIÓN DE APLICACIÓN

El cuestionario es válido para medir la investigación

III. PROMEDIO DE VALORACIÓN:

90

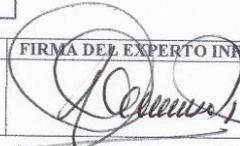
LUGAR Y FECHA	DNI	FIRMA DEL EXPERTO INFORMANTE	Nº DE TELEFONO
Cusco, 03 julio 2019	23853228		984002511

TABLA DE VALIDACIÓN DE INSTRUMENTO POR EXPERTOS

Apellidos y Nombres del Informante	Institución donde labora	Nombre del Instrumento	Autor del Instrumento
DR. VILLAFUERTE SERNA RONY	UNSAAC	CUESTIONARIO	ROBERT WILBERT ALZAMORA PAREDES
Título de la Investigación: MODELO BASADO EN EL ANALISIS MULTIVARIANTE DE DATOS PARA EL ESTUDIO DEL RENDIMIENTO ACADEMICO DE LAS ESCUELAS PROFESIONALES DE LAS UNIVERSIDADES NACIONALES DE LA REGION DEL CUSCO.			

I. ASPECTOS DE EVALUACIÓN

		DEFICIENTE				REGULAR				BUENA				MUY BUENA				EXCELENTE			
		5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	100
1. CLARIDAD	Esta formulado con lenguaje apropiado.																				X
2. OBJETIVIDAD	Esta expresado en conductas observables.																				X
3. ACTUALIZACIÓN	Está adecuado al avance de la ciencia y la tecnología.																		X		
4. ORGANIZACIÓN	Esta organizado en forma lógica.																				X
5. SUFICIENCIA	Comprende aspectos cuantitativos														X						
6. INTENCIONALIDAD	Es adecuado para valorar el modelo																X				
7. CONSISTENCIA	Está basado en aspectos teóricos científicos.																				X
8. COHERENCIA	Entre las variables, dimensiones, Indicadores e ítems.																				X
9. METODOLOGÍA.	La estrategia responde al propósito de la investigación.																				X
10. PERTINENCIA	La escala es aplicable.																	X			

II. OPINIÓN DE APLICACIÓN

El instrumento es válido para medir el modelo

III. PROMEDIO DE VALORACIÓN:

90.5

LUGAR Y FECHA	DNI	FIRMA DEL EXPERTO INFORMANTE	Nº DE TELEFONO
02 - JUL - 2019	23957778		984 74 98 18