

Técnicas de análisis de datos aplicadas a encuesta docente en Ingeniería en Sistemas de Información

Dominguez Blanda, Tomás ¹; Mosca, Sebastián Jesús ¹; Ardiles, Micaela ²; Avilar, Pilar ²; Gras, Luisina ³

1 - Universidad Tecnológica Nacional, Facultad Regional Villa María - UTN FRVM

2 - Universidad Tecnológica Nacional, Facultad Regional Córdoba- UTN FRC

3 - Universidad Tecnológica Nacional, Facultad Regional Concepción del Uruguay - UTN FRCU

tdominguez15td@gmail.com, sebastian.mosca@outlook.com, micardiles@hotmail.com,
piliavila01@gmail.com, luisinagras@gmail.com

Resumen

En el marco del Proyecto de Investigación Asociativo y Multi Facultad, se diseñó una encuesta docente en distintas facultades en la carrera de Ingeniería en Sistemas de Información. En función de éstas se propuso aplicar técnicas de análisis de datos a los fines de colaborar en la selección de casos y caracterización de la población estudiada. Se codificaron los datos para poder realizar un análisis estadístico descriptivo que consta de tablas de frecuencia y gráficos. Se diseñó una matriz de correlación en busca de relaciones fuertes entre variables, para luego avanzar con minería de datos, concretamente con la técnica de reglas de asociación que permitieron visibilizar la información de un modo diferente y mostrar nuevas relaciones entre categorías de la encuesta. Por último, se le asignó un puntaje a cada módulo o bloque y se generó un listado de todo el plantel docente con sus respectivos puntajes, lo que permitió seleccionar los primeros casos de estudio, con relativa objetividad, para poder avanzar en otras técnicas de investigación.

Palabras Clave

Codificación de Encuesta - Estadística Descriptiva - Python - Reglas de Asociación- Educación en Ingeniería

1 Introducción

El nuevo escenario social, globalizado y tecnológico, requiere de una redefinición y reconstrucción de las prácticas docentes en la educación universitaria. Por otro lado las carreras de ingeniería se encuentran atravesadas por un proceso de adecuación curricular que significó la incorporación de competencias de egreso, de acuerdo a lo establecido por el Ministerio de Educación de la Nación Argentina en la Resolución 1557/2021 [1-2]. En este contexto y en el marco del Proyecto de Investigación (PID),

Asociativo y Multi Facultad:: TEAIBCO0008216, “Aportes para la formación en competencias desde las prácticas de enseñanza mediadas tecnológicamente, en el marco de la carrera de Ingeniería en Sistemas de Información en cuatro facultades regionales de la Universidad Tecnológica Nacional”. Este trabajo presenta el desarrollo de algunas técnicas de análisis de datos con la finalidad de colaborar con el objetivo de dicho proyecto que es visibilizar prácticas de enseñanza/aprendizaje centradas en el estudiante, con inclusión de TIC en la mediación pedagógica y en la evaluación, en la medida que favorezcan el desarrollo de competencias generales y específicas en quienes estudian Ingeniería en Sistemas de Información.

2 Desarrollo

De acuerdo a lo expresado en el apartado anterior, el equipo de investigación decidió comenzar con el diseño de una encuesta docente, que se implementó en las cuatro Facultades Regionales, es decir, Córdoba, Villa María, Concepción del Uruguay y San Francisco a través de un formulario de Google. La población fue el plantel docente, de todos los niveles e independientemente de su cargo y rol en las asignaturas. Se obtuvieron así 251 respuestas.

Dicha encuesta se estructuró en cuatro bloques o módulos: Generalidades, Mediación Pedagógica, Mediación Tecnológica y Evaluación.

Los tipos de preguntas utilizados en la encuesta son los siguientes:

- De opción múltiple y lista desplegable: admiten seleccionar una única opción.
- Casillas de verificación: admiten la selección de múltiples opciones, como por ejemplo, diversas alternativas de evaluación en simultáneo.
- Preguntas cortas/largas: admiten una respuesta alfanumérica, de texto libre a los efectos de permitir el ingreso de datos como nombres, comisiones o alguna alternativa que no se hubiese considerado al momento del diseño de la encuesta.
- De cuadrículas o matriciales: permiten relacionar la frecuencia con diferentes opciones, tal como se muestra en la Figura 1.

	Siempre	Muy Frecuentemente	Frecuente	Poco Frecuente	Nunca
Cuestionarios	<input type="radio"/>				
Documentos escritos individuales	<input type="radio"/>				
Presentaciones digitales	<input type="radio"/>				
Participación en foros	<input type="radio"/>				
Trabajos prácticos Integradores	<input type="radio"/>				
Mapas conceptuales	<input type="radio"/>				
Realización de videos	<input type="radio"/>				
Podcast	<input type="radio"/>				
Documentos colaborativos	<input type="radio"/>				
Participación de Wiki	<input type="radio"/>				
Técnica de cambio de roles	<input type="radio"/>				
Otro	<input type="radio"/>				

Figura 1: Pregunta tipo cuadrícula o matricial.

Las preguntas de tipo matriciales utilizaron escala likert [3,4]: escala de calificación que se utiliza para cuestionar a una persona sobre su nivel de acuerdo o desacuerdo con una declaración. Esta se utilizó con el fin de cuantificar aspectos que no son numéricos, es decir, que permita ponderarlos y tratarlos con herramientas de análisis y estadística de datos. Por otro lado otorgan gradualidad a las respuestas evitando preguntas dicotómicas.

2.1 Codificación

La codificación consistió en transformar variables de texto en números que pudieran facilitar el procesamiento estadístico. Para ello se definieron etiquetas utilizando una combinación de caracteres alfanuméricos representativos de cada una de ellas [5]. Esta combinación comienza con una letra “P” que representa la palabra “pregunta”, más el número de la pregunta. Se genera una tabla de referencias para cada opción, como lo muestra la Figura 2.

P1 - Facultad	
1.	UTN - Facultad Regional Concepción del Uruguay
2.	UTN - Facultad Regional Córdoba
3.	UTN - Facultad Regional San Francisco
4.	UTN - Facultad Regional Villa María

Figura 2: Modelo de Tablas de referencias para codificar.

En el caso de las preguntas tipo casillas de verificación, se generó un código para cada combinación de opciones, tal como se muestra en la Figura 3 donde se conjugaron diversas formas de evaluación.

P28 - ¿Quién evalúa los aprendizajes en su asignatura?	
1.	Docente – Heteroevaluación
2.	Docente - Heteroevaluación, Estudiante - Autoevaluación
3.	Docente - Heteroevaluación, Estudiante - Autoevaluación, Grupos de Estudiantes – Coevaluación
4.	Docente - Heteroevaluación, Estudiante - Autoevaluación, Externa (docentes externos a la asignatura)
5.	Docente - Heteroevaluación, Estudiante - Autoevaluación, Grupos de Estudiantes - Coevaluación, Externa (docentes externos a la asignatura)
6.	Docente - Heteroevaluación, Grupos de Estudiantes – Coevaluación
7.	Docente - Heteroevaluación, Grupos de Estudiantes - Coevaluación, Externa (docentes externos a la asignatura)
8.	Docente - Heteroevaluación, Externa (docentes externos a la asignatura)
9.	Estudiante - Autoevaluación
10.	Estudiante - Autoevaluación, Externa (docentes externos a la asignatura)

Figura 3: Tabla de referencias para preguntas tipo casillas de verificación.

En el caso de las matriciales, se agregó al final de la etiqueta una letra ordenada alfabéticamente que representa la fila, es decir las distintas opciones. Esto se puede apreciar en la Figura 1 donde las opciones

de dicha pregunta son: documentos colaborativos, podcast, trabajos prácticos, etc. (P20A, P20B,...). Las frecuencias se codificaron en las tablas de referencia del 1 al 5 de acuerdo a la gradualidad.

Por último, tanto las preguntas de escritura libre como las opciones “otro” de cada matriz fueron analizadas manualmente. Esto se realizó con el objetivo de traducir lo escrito en nuevas opciones para la base de datos o para reclasificarlas en las opciones ya establecidas, entendiendo que muchas de ellas eran aclaraciones sobre las opciones ya elegidas. En cuanto a campos como nombres o comisiones se decidió no codificarlos.

El resultado del proceso de codificación fue una base de datos con formato csv, tsv, xlsx de tal manera que pudiera ser procesado en herramientas ofimáticas como Excel, softwares estadísticos y/o lenguajes de programación.

2.2 Análisis Estadístico Descriptivo

2.2.1 Tablas de frecuencia

Para una primera observación de los datos se realizó un análisis estadístico descriptivo utilizando el software de IBM “SPSS Statistics 21.0”. Se confeccionaron tablas de frecuencias que permitieron avanzar en la distribución de los datos, la centralidad y variabilidad de los mismos, así como conocer las técnicas, más y menos, empleadas para la mediación pedagógica, tecnológica y sobre los procesos de evaluación. A modo de ejemplo se muestra una tabla en la **Tabla 1**.

Tabla 1: Tabla de frecuencias

P3				
	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
1	22	8.8	8.8	8.8
2	31	12.4	12.4	21.1
3	75	29.9	29.9	51.0
4	123	49.0	49.0	100.0
Tota	251	100.0	100.0	

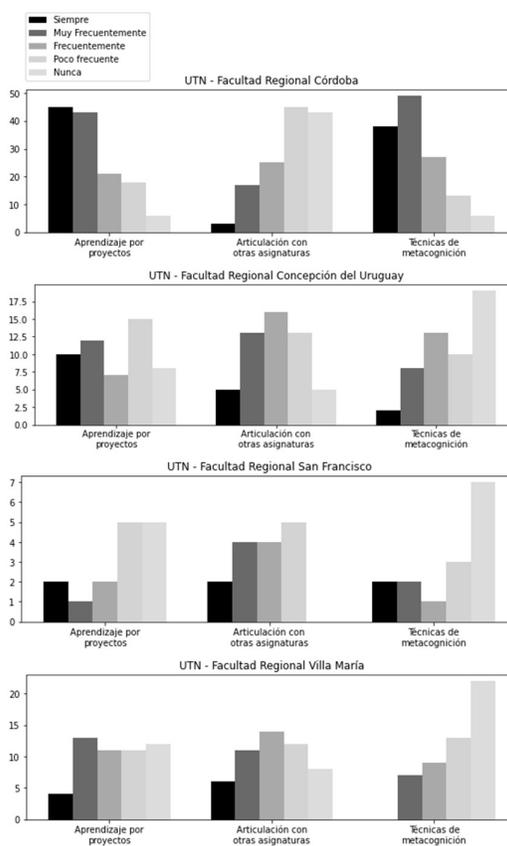


Figura 4: Distribución de frecuencias según la estrategia didáctica

La **Tabla 1** responde a la experiencia en años del cuerpo docente, y tal como se ve el 49% de ellos tienen más de 20 años y el 30% entre 10 y 20 años. La primera columna indica la frecuencia expresada en cantidad de docentes, la segunda el porcentaje que representa sobre el total encuestado, la tercera el porcentaje sobre el total sin incluir los nulos y la última, el porcentaje acumulado.

2.2.2 Gráficos

A los fines de visualizar gráficamente toda la información relevante, hacia el interior de cada facultad, de tal manera que cada Departamento pudiese conocer las características de las prácticas docentes, se particionó la base de datos por unidad académica. Se construyeron gráficos de barras y gráficos circulares como se muestran en la **Figura 4** y **5** respectivamente. En la **Figura 5** se pueden

ver cuatro gráficos circulares, uno por cada facultad donde puede observarse la distribución por cargo en cada una.

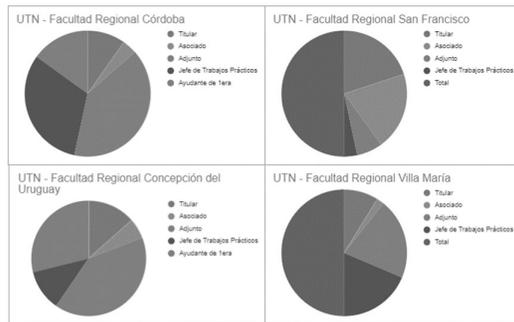


Figura 5: Distribución de cargos docentes por Facultad

El gráfico de barras compuesto fue de gran utilidad para la información de las preguntas matriciales, para ello se construyeron nuevas tablas de frecuencias por cada pregunta y por facultad. Para calcular la frecuencia de cada fila utilizamos la función de Excel “CONTAR.SI”, y a partir de dichas tablas se generó, de forma automática las distintas gráficas.

2.2.3 Matriz de correlación

Se decidió observar la medida de la relación lineal entre algunas variables de la encuesta, para conocer si las mismas estaban correlacionadas, es decir si variaban de manera conjunta.

A modo de ejemplo citamos algunas de las relación de interés para la investigación:

- Experiencia docente y estrategias para conocer a sus estudiantes
- Nivel o año de cursado de la asignatura y tipos de mediación pedagógica
- Diversidad de herramientas tecnológicas y antigüedad docente, entre otras.

Para ello, se construyó una Matriz de Correlación, herramienta que mide el grado de relación lineal entre cada par de elementos o variables. Los valores de correlación en cada celda se pueden ubicar entre -1 y +1. Si los dos elementos tienden a aumentar o disminuir al mismo tiempo, el

valor de correlación se aproxima hacia uno de los dos extremos [6], lo que nos facilitó comparar un gran volumen de datos en una matriz nxm. Para el desarrollo de la misma, se decidió aplicar una capa de procesamiento a los datos ya existentes. En los casos donde las respuestas tenían el formato mostrado en la **Figura 1**, se agruparon los campos “Siempre”, “Muy frecuentemente” y “Frecuentemente” asignándoles el valor 1, y se agruparon los campos “Poco Frecuente” y “Nunca” asignándoles el valor 0. Una vez que se realizó el procesamiento, se utilizó Microsoft Excel para construir la matriz de correlación y obtener los resultados en formato de tabla (**Figura 6**), donde tanto la fila superior como la columna en la izquierda muestran las variables que se necesitan correlacionar, mientras que la celda que una fila con columna muestra el valor de la correlación entre ambas preguntas.

Bloque 2	P23A-C	P23B-C	P23C-C	P23D-C	P23E-C
P18A-C	0.074802	-0.06714	-0.07143	0.024351	0.04217
P18B-C	0.100865	0.128998	0.24212	0.109292	0.142265
P18C-C	0.094732	-0.04123	-0.07726	0.130212	0.120096
P18D-C	0.026816	0.034337	0.104546	0.087393	0.079058
P18E-C	0.090422	0.146449	0.097122	0.251014	0.184091
P18F-C	0.096791	0.12005	0.104188	0.212353	0.144775
P18G-C	0.061903	-0.03676	0.04562	0.073408	0.093929
P18H-C	0.011085	0.037749	0.171754	0.12893	0.080528
P18I-C	0.056292	0.032129	0.12467	0.092093	0.214149

Figura 6: Parte de la matriz de correlación.

Analizando la matriz en búsqueda de alguna relación estadísticamente fuerte entre las preguntas, en términos generales, se detectó una baja correlación en la gran mayoría de los casos. En cada correlación, se presentaron puntajes por debajo de 0.50 en correlaciones directas, siendo que su rango posible va desde 0 hasta 1 punto. Por otro lado, en correlaciones inversas, se presentaron en su mayoría puntajes por encima de -0.50 puntos, siendo su rango posible desde 0 a -1.

2.3 Reglas de asociación

Dado que con la matriz de correlación no se encontraron relaciones fuertes, se avanzó con técnicas de minería de datos, como las

reglas de asociación [7,8], ya que esta permite obtener cuáles de las opciones de respuesta a las diferentes preguntas cerradas aparecen juntas normalmente en la base de datos.

Para esto se requiere que las opciones se estructuren como un ítem. Esto quiere decir que en una pregunta donde se pueda responder con la frecuencia con la que se hace algo, como utilizar una estrategia concreta de mediación pedagógica, cada una de las frecuencias para esa estrategia concreta es una fila en la tupla o un elemento de una lista.

Para saber cuán buena o significativa es una regla de asociación se evalúan 2 valores asociados a las mismas: el soporte y la confianza. El soporte es el porcentaje de tuplas que contiene a todos los ítem de una regla sobre el total de tuplas en la base de datos. La confianza es el porcentaje de tuplas que si contiene al ítem A también contiene al ítem B de la regla, como se puede observar en la **Tabla 2**.

Tabla 2: Tabla de reglas de asociación

Item A	Item B	Soporte	Confianza
-	Estrategia pedagógica , otra, nunca	0,84	0,84
Nivel de asignatura, 2do año	Estrategia pedagógica , guía de trabajos prácticos, de muy frecuentem ente a siempre	0,17	0,93

2.4 Valoración de Encuestas mediante Programación

En esta etapa se trabajó con el fin de establecer algún mecanismo que permitiera asignar, con cierto nivel de objetividad, una valoración a cada encuesta. El equipo de investigación seleccionó un conjunto de

preguntas por cada módulo, que fuesen representativas del tipo de práctica docente que requiere la formación por competencias. En este sentido, se establecieron puntuaciones para cada una de ellas de modo tal de obtener un puntaje total, representativo de cada encuestado y por otro lado que fuese posible considerar sus valores parciales, es decir por bloque.

El primer módulo, generalidades, no fue considerado por tratarse de información general, que fue trabajada en el análisis estadístico descriptivo.

El segundo módulo, Mediación Pedagógica, está comprendido por las preguntas que hacen alusión a las estrategias adoptadas por docentes para llevar a cabo un mayor acercamiento y conocimiento de sus estudiantes, junto a la utilización de diversas estrategias de mediación pedagógicas en sus clases.

El tercer módulo, Mediación Tecnológica, está compuesto por preguntas que referencian los diferentes recursos e instrumentos tecnológicos que utiliza el cuerpo docente para brindar soporte a sus estrategias de enseñanza.

Y por último, el módulo de Evaluación conforma las preguntas de las diferentes estrategias evaluativas que el cuerpo docente opta por llevar a cabo para corroborar los conocimientos adquiridos por sus estudiantes, comprendiendo las diferentes formas y participantes en el proceso de evaluación, el uso de rúbricas, y si se ofrece algún tipo de retroalimentación hacia el estudiantado luego de la evaluación.

2.4.1 Cálculo y Asignación de Puntajes

Una vez seleccionadas las preguntas representativas de cada bloque, se asignó un determinado valor o peso numérico a cada una de las respuestas. Los puntajes fueron asignados tomando en cuenta la regularidad con la que el profesorado afirma utilizar una determinada estrategia o recurso a lo largo de sus clases. Con el mismo criterio se

trabajó el resto de las preguntas. Si el ítem a valorar en la práctica docente se utiliza “Siempre”, se le asigna un punto. En cambio, si el mismo afirma utilizar dicha estrategia “Frecuentemente”, entonces se le suma 0.50 puntos, tal como se muestra en la **Tabla 3**.

Tabla 3: Valoraciones por frecuencia de uso

Frecuencia de uso	Valor Asignado
Siempre	1.00
Muy Frecuentemente	0.75
Frecuentemente	0.50
Poco Frecuente	0.25
Nunca	0.00

Las preguntas matriciales, al contar con varios ítems, su ponderación final es el resultado de la suma de cada ítem por su frecuencia.

Para las preguntas de opción múltiple se les asignó un puntaje según la respuesta seleccionada, sin necesidad de realizar otros cálculos posteriores, por ejemplo si realiza retroalimentación 1 punto, si no la realiza 0 puntos. Luego de haber asignado los diferentes pesos, se procedió a determinar el puntaje total obtenido por cada uno de los 251 docentes participantes de la encuesta. Cada uno de estos puntajes pudo desagregarse en 3 valores, siendo representativo de cada bloque bajo estudio.

Todo esto se desarrolló a través de un algoritmo en lenguaje de programación Python. Los datos se obtuvieron del archivo formato .xlsx, producto del proceso de codificación anteriormente mencionado. A los efectos de estandarizar el peso de cada bloque se multiplicó por un coeficiente determinado específicamente para cada módulo con la finalidad de que cada uno de estos alcance el valor 20. De esta forma se consolidó una ecuación final para determinar el puntaje total, como se muestra en la **Figura 7**.

$$M_Ped * 1.429 + M_Tec + Eva * 1.09 = P_Final$$

Figura 7: Ecuación para el cálculo de puntaje total

Como resultado, el algoritmo genera un listado con los puntajes para las 251 encuestas como se ve a modo de ejemplo en la **Tabla 4**

Tabla 4: Muestra de 4 casos del listado de puntajes por bloque

Med Ped	Med Tec	Eva	Puntaje total
11,07	7,75	7,47	26,29
9,65	8,5	7,09	25,24
8,57	12,25	8,99	29,81
8,93	8,75	8,23	25,91

Por último, se implementó dentro del algoritmo una funcionalidad encargada de generar gráficos de dispersión que ilustra todos los puntajes obtenidos. En la **Figura 8** se puede apreciar la dispersión de los puntajes totales. Si se observa la misma gráfica pero segmentada por módulo (ver **Figura 9**), los resultados no varían demasiado, es decir se encontró una gran dispersión de casos. Asimismo, en evaluación pareciera que existe un grado mayor de centralidad.

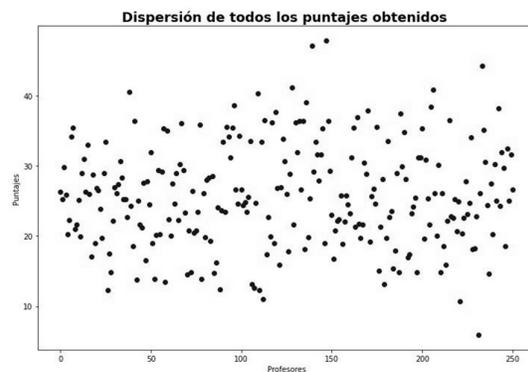


Figura 8: gráfica de dispersión de los puntajes totales

Considerando que el puntaje máximo que se podría obtener por módulo es de 20 puntos, y a sabiendas que dicho número resulta imposible ya que significaría utilizar todos los instrumentos y herramientas con

frecuencia “siempre”, se encontró que en mediación tecnológica 16 casos superan los 12 puntos, en cambio en la mediación pedagógica 66 superan ese puntaje y en evaluación 18 casos.

El análisis de estos resultados, en términos de práctica docente excede el objetivo del presente trabajo, aún así se posicionó el número 12 como referencia en virtud de superar ligeramente la media de los puntajes.

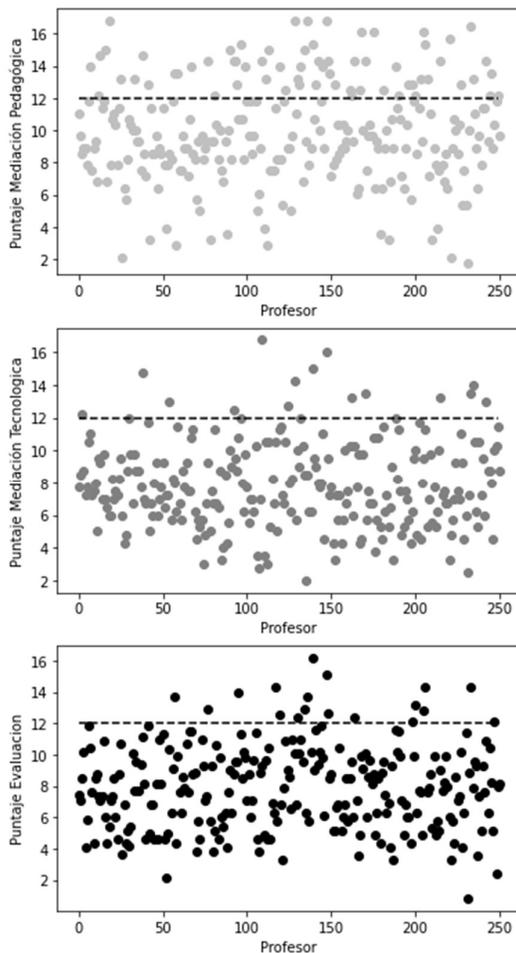


Figura 9: Gráficas de dispersión por bloque

3. Conclusión

En el marco del proyecto de investigación, y de la encuesta realizada, la aplicación de diversas técnicas resultó de utilidad para generar una aproximación a la forma en que cada docente diseña sus clases

y la variedad de instrumentos que utiliza en ellas.

El proceso de codificación de las variables hizo posible transformar los datos cualitativos propios de los procesos educativos en variables numéricas, las cuales fueran plausibles de ser procesadas en múltiples herramientas de análisis de datos.

El análisis estadístico descriptivo permitió informar acerca de las realidades de cada facultad y avanzar en la búsqueda de relaciones entre variables que a priori se pensaba estaban relacionadas. La matriz de correlación reveló que dichas correlaciones no eran lo suficientemente fuertes. Esto posibilitó avanzar sobre técnicas de minería de datos que pudiesen evidenciar relaciones que no son tan directas entre ítems de una misma categoría.

Por otro lado, el algoritmo programado en Python facilitó puntuar, de un modo más objetivo, cada encuesta. Con dichos valores se puede observar con claridad, no sólo los mejores puntajes totales, sino valorar aquellas prácticas que se destacan en cada uno de los bloques analizados. Esto es significativo porque permite rescatar casos que resultan relevantes en lo pedagógico, tecnológico y/o evaluativo, independientemente de que lo sea en los 3 bloques al mismo tiempo.

Adicionalmente, posibilita continuar con la investigación y avanzar con otras técnicas y herramientas sobre el objeto de estudio, esto es encontrar aquellos casos que promuevan el desarrollo de competencias en ingeniería.

Agradecimientos

Agradecemos a la Ing. Mercedes Soria, docente de la UTN - Facultad Regional Villa María por su colaboración como tutora durante toda la realización del trabajo y al Ing. Carlos Casanova, docente de la UTN - Facultad Regional Concepción del Uruguay, por su aporte en las técnicas de minería de datos.

Destacamos que la participación como becarios, en el proyecto de investigación, nos permitió trabajar en conjunto con estudiantes de otras facultades y de

diferentes niveles de la carrera, aportando positivamente en nuestra formación profesional.

Referencias

- [1] CONFEDI, Propuesta de Estándares de Segunda Generación para la Acreditación de Carreras de Ingeniería en la República Argentina - “Libro Rojo”. Rosario: Universidad FASTA Ediciones, 2018. [En línea]. Disponible en: <https://bit.ly/3PoHgaH>
- [2] Ministerio de Educación, Nuevos Estándares para acreditación de carreras comprendidas en el Artículo 43 de las Ley de Educación Superior. Resoluciones comprendidas entre la N° 1537/2021 y la N° 1566/21. 2021. [En línea]. Disponible en: <https://www.boletinoficial.gob.ar/detalleAviso/primera/244507/20210518>
- [3] A. F. Echaury, H. Minami, y J. Izquierdo, «La Escala de Likert en la evaluación docente: acercamiento a sus características y principios metodológicos», *Perspect. Docentes*, n.º 50, pp. 31-40, 2012.
- [4] S. M. Maldonado Luna, «Manual Práctico Para El Diseño De La Escala Likert», *Xihmai*, vol. 2, n.º 4, nov. 2012, doi: 10.37646/xihmai.v2i4.101.
- [5] E. Abascal y I. G. Esteban, *Análisis de encuestas*. ESIC Editorial, 2005.
- [6] Minitab Support V.21.0. Interpret all statistics and graphs for Item Analysis - Minitab Support, <https://support.minitab.com/en-us/minitab/21/help-and-how-to/statistical-modeling/multivariate/how-to/item-analysis/interpret-the-results/all-statistics-and-graphs/>
- [7] P.-N. Tan, M. Steinbach, A. Karpatne, y V. Kumar, *Introduction to Data Mining (2nd Edition)*, 2nd ed. Pearson, 2018.
- [8] Han, J., Kamber, M., & Pei, J. (2011, July 6). *Data Mining: Concepts and Techniques (The Morgan Kaufmann Series in Data Management Systems)* (3rd ed.). Morgan Kaufmann.(Han et al., 2011)