

## **ANALISIS DE LOS RAZONAMIENTOS DE ESTUDIANTES EN MODELACIONES MATEMATICAS MEDIANTE TECNICAS DE MINERIA DE DATOS: UN ESTUDIO DE CASO**

**Pablo CARRANZA<sup>1</sup>, Evangelina PALÓPOLO<sup>2</sup>, Natalia CLAROTTI<sup>3</sup>, Mathias TEJERA<sup>4</sup>**

### **ANALYSE DES RAISONNEMENTS DES ÉLÈVES DANS DES MODÈLES MATHÉMATIQUES À L'AIDE DE TECHNIQUES DE FOUILLE DE DONNÉES : UNE ÉTUDE DE CAS**

### **ANALYSIS OF STUDENTS' REASONINGS IN MATHEMATICAL MODELS THROUGH DATA MINING TECHNIQUES: A CASE STUDY**

#### **RESUMEN**

En un contexto de aprendizaje basado en proyectos, donde estudiantes trabajan en equipo, resulta necesario conocer el desempeño individual de cada estudiante. En esta comunicación compartimos un estudio basado en dos enfoques de análisis de datos con la intención de contar con información sobre cómo los estudiantes razonan en el marco de un modelo matemático donde se abordan cuestiones referidas a un proyecto de construcción de un sistema potabilizador de agua. Los datos corresponden a respuestas de los estudiantes a un cuestionario y fueron analizados principalmente con la técnica de cluster (algoritmo knn) y el enfoque de análisis implicativo (ASI). La comunicación muestra un conjunto de tendencias en las respuestas que permiten a los docentes comprender no solo las dificultades por las que atraviesan los estudiantes sino también individualizar a cada estudiante según su nivel de comprensión respecto de la problemática abordada.

*Palabras claves: aprendizaje basado en proyectos, k-means, análisis estadístico implicativo*

#### **RÉSUMÉ**

Dans un contexte d'apprentissage par projet, où les étudiants travaillent en équipe, il est nécessaire de connaître la performance individuelle de chaque étudiant. Dans cette communication, nous partageons une étude basée sur deux approches d'analyse de données dans le but d'avoir des informations sur la façon dont les étudiants raisonnent dans le cadre d'un modèle mathématique où les problèmes liés à un projet de construction d'un système de traitement de l'eau sont abordés. Les données correspondent aux réponses des étudiants à un questionnaire et ont été analysées principalement avec la technique de cluster (algorithme knn) et l'approche d'analyse implicative (ASI). La communication montre un ensemble de tendances dans les réponses qui permettent aux enseignants non seulement de comprendre les difficultés que traversent les élèves, mais aussi d'individualiser chaque élève selon son niveau de compréhension par rapport à la problématique abordée.

*Mots-clés : apprentissage par projet, k-means, analyse statistique implicative*

#### **ABSTRACT**

In a project-based learning context, where students work as a team, it is necessary to know the individual performance of each student. In this communication we share a study based on two data

1 Universidad Nacional de Río Negro. Primeros Pobladores 1224. 8328. Allen. Río Negro. Argentina. pfcarranza@gmail.com

2 Universidad Nacional de Río Negro. Isidro Lobos 516. 8332. Gral. Roca. Río Negro. Argentina. e.palopolo@gmail.com

3 Universidad Nacional de Río Negro. Isidro Lobos 516. 8332. Gral. Roca. Río Negro. Argentina. nataliaclarotti@gmail.com

4 Johannes Kepler University. Altenbergerstraße 69. Linz Austria. matutejera@gmail.com

analysis approaches with the intention of getting information on how students reason within the framework of a mathematical model where issues related to a construction project of a water treatment system are addressed. The data correspond to student responses to a questionnaire and were analyzed mainly with the cluster technique (knn algorithm) and the implicative analysis (ASI) approach. The communication shows a set of trends in the answers that allow teachers to understand not only the difficulties that students are going through, but also to individualize each student according to their level of understanding regarding the addressed problem.

*Key words : project-based learning, k-means, implicative statistical analysis*

### **1. 1 Introducción**

La presente comunicación se inscribe en el interés por disponer de mayor información sobre la actividad de estudiantes en el marco de una propuesta del tipo aprendizaje basado en proyectos.

El proyecto que motivó tal interés consistió en el diseño, construcción y testeo de sistemas potabilizadores de agua para pobladores rurales de la Patagonia Argentina y se desarrolló con estudiantes de primer año de una carrera universitaria en el marco de la cátedra anual de Matemática. Este proyecto se basó en una tecnología desarrollada por investigadores de la Universidad de Buffalo (Gan & Zhang, 2017) y básicamente consiste en destilar agua para luego potabilizarla mediante sales apropiadas. La destilación del agua se produce provocando su evaporación y posterior condensación mediante el aprovechamiento de energía solar.

Con financiamiento del Ministerio de Educación Nacional y en el marco de un proyecto de extensión universitaria, el presupuesto permitía la construcción de cinco prototipos de sistemas potabilizadores. Así, se le propuso a los estudiantes agruparse en cinco equipos, debiendo cada uno llegar a desarrollar un prototipo de sistema potabilizador.

En ese contexto, cada equipo de estudiantes debía realizar un número importante de actividades que se compartirían luego en sucesivas presentaciones a modo de avance al resto de la clase.

En esa dinámica de trabajo colectivo y de presentaciones en clase también colectivas, resultaba difícil a los docentes tener información de la evolución y de la comprensión de las tendencias de los estudiantes en las diferentes actividades que el proyecto demandaba. Es así que se consideró de manera exploratoria la utilización de cuestionarios individuales que serían luego analizados con un conjunto de técnicas de análisis de datos, entre ellas el análisis implicativo.

En particular interesaban los razonamientos que los estudiantes pudiesen realizar al momento de analizar modelaciones producidas en clase. Estas modelaciones resultaban ser instancias importantes en el proyecto pues en ellas se construían los argumentos que daban sustento a las toma de decisiones a efectuar en el proyecto.

Destacamos que el trabajo de campo aquí presentado fue realizado en el segundo año de la pandemia Covid19 en Argentina (año 2021), por lo que los resultados se encuentran fuertemente condicionados por las excepcionales circunstancias en la que se produjeron los aprendizajes. Sin embargo y a pesar de la singularidad de la situación, retenemos elementos de interés tanto en lo metodológico como en los resultados que nos parecen trascender al momento de excepción en el cual se desarrolló la experimentación.

### **2. 2 Problemática**

El caso que compartimos en esta publicación se refiere a la modelación de las proporciones que debería tener el gabinete donde se realizaría el proceso de evaporación y posterior condensación del agua. Ese gabinete de plástico transparente debía dejar ingresar la radiación solar para que la misma sea utilizada al momento de la evaporación del agua.

Luego de desarrollar en clase de manera conjunta la modelación en cuestión y con la intención de disponer de información referida a los razonamientos de los estudiantes en el marco de la modelación en cuestión, les solicitamos entonces responder un cuestionario con preguntas referidas a esa modelación. Este documento presenta un análisis de las respuestas al cuestionario mencionado y sus posibles potencialidades.

Al momento de realizar el cuestionario, y por las condiciones arriba descritas, pocas eran las hipótesis formuladas a priori, por lo que el estudio entonces se encuadra en un enfoque exploratorio. Básicamente se deseaba obtener información sobre tendencias en los estudiantes sobre:

- a) La comprensión de las características generales de la modelación
- b) La interpretación de relaciones internas en la modelación
- c) La interpretación de relaciones entre la modelación y el contexto
- d) La producción de análisis en la modelación que impacten en decisiones a tomar en el proyecto

La **¡Error! No se encuentra el origen de la referencia.** representa un esquema básico del gabinete plástico de 1 m<sup>2</sup> de superficie de la base, donde se produce la condensación del agua.

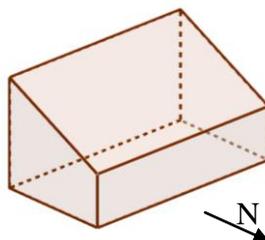


Figura 1 – Gabinete plástico transparente.

Fuente: Elaboración propia

Una pregunta en particular motivó la modelación sobre las proporciones del potabilizador en cuestión: “Si la radiación solar ingresa principalmente por el techo y si la energía se disipa principalmente por las paredes laterales, ¿cuáles serían las dimensiones ideales que minimicen la superficie de las caras laterales?”

Para abordar esta pregunta, se realizó una modelación en Geogebra ([www.geogebra.org](http://www.geogebra.org)) con tres deslizadores como parámetros. La modelación fue guiada por los docentes de la cátedra y se desarrolló en dos etapas. En la primera etapa se modeló considerando principalmente relaciones geométricas entre las dimensiones, en la segunda etapa se la abordó de una manera más analítica (Carranza, Letrouneau, & Navarro, 2022). La Figura 2 representa la modelación en cuestión.

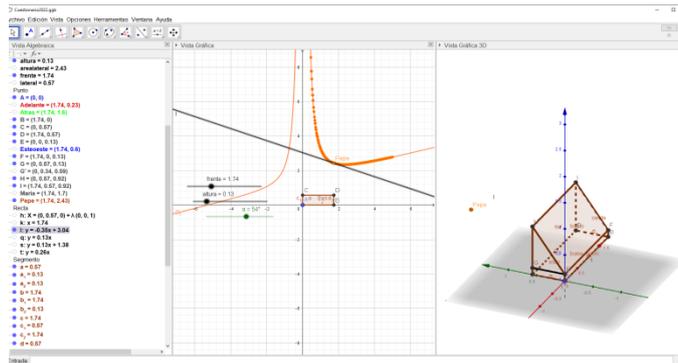


Figura 2 – Modelación en Geogebra

Fuente: Elaboración propia

### 3. 3 Métodos y Resultados

El cuestionario propuesto constó de 43 preguntas, todas abiertas. Algunas de las preguntas ya habían sido analizadas en clase, otras no. Cincuenta y un estudiantes respondieron el cuestionario de manera digital mediante el formulario Google Form.

Las respuestas a las 43 preguntas abiertas fueron analizadas por los 3 docentes de manera independiente según la pertinencia de las respuestas. Cada respuesta se asoció a una y solo una de las tres opciones siguientes: respuesta correcta, respuesta incorrecta y no responde. Las diferencias observadas entre las apreciaciones a las respuestas (15 de un total de 2193 respuestas) fueron debatidas al interior del equipo docente y consensuadas en una apreciación final.

A modo ilustrativo, reproducimos algunas de las preguntas formuladas:

- 2) Indica el equipo al que perteneces
- 3) ¿En qué unidad de medida está expresado el deslizador llamado “frente”?
- 35) ¿Qué relación hay entre la función  $p(x)$  y el rastro del punto Pepe?
- 36) ¿Qué información te brinda la recta llamada tangente?

Las respuestas de los estudiantes no fueron analizadas individualmente sino en función del tipo de pregunta formulada. Para ello, se consideró una tipificación determinada por tres categorías: descriptiva, interpretativa o analítica. Una pregunta era descriptiva si satisfacía alguno de los dos criterios siguientes:

- 1) se podía responder con una simple observación de la modelación
- 2) se refería a la semántica de un elemento en la modelación

Por su lado, una pregunta era interpretativa si:

- 1) se solicitaba un análisis al interior de la modelación ó
- 2) se demandaba una semántica compleja, relacionando el interior de la modelación con el contexto ó
- 3) se requerían razonamientos

Por último, una pregunta era analítica si

- 1) resultaba necesario justificar elecciones ó
- 2) demandaba producir argumentaciones ó
- 3) solicitaba poner en relación de manera compleja varios elementos de la modelación

A los efectos de poder tratar los datos estadísticamente con la intención de encontrar tendencias, las respuestas expresadas ya como categóricas (respuesta correcta, respuesta incorrecta y no responde) fueron codificadas como variables dicotómicas (valores 0 o 1 según el caso). La Tabla 1 ilustra un ejemplo de codificación para la pregunta número 31, en este caso, se trata de una respuesta asignada como incorrecta. Esta conversión triplicó entonces la cantidad de variables (más de 120 variables dicotómicas).

	31cor	31inco	31nores
Estudiante x	0	1	0

Tabla 1 – Ejemplo de codificación

Fuente: Elaboración propia

En este documento, y por razones de espacio, compartiremos un resumen de dos de los análisis realizados y vinculados entre sí. Uno se refiere a la distribución de frecuencias relativas de las respuestas al cuestionario, el otro a relaciones entre las respuestas. Respecto al primero, retendremos algunos resultados obtenidos al agrupar los estudiantes ya no en función de sus equipos de trabajo sino por cercanía en las respuestas. Para ello, se recurrió al método de conformación de grupos denominado knn (Hastie, Tibshirani, & Friedman, 2001; Piryonesi & El-Diraby, 2020). El mismo se realizó en lenguaje `python` con las bibliotecas `sklearn`, `pandas` y `numpy`.

En lo referido a la relación entre variables, presentaremos los resultados obtenidos al utilizar el método de análisis implicativo disponible en el software CHIC. El método permite explicitar relaciones del tipo “si...entonces...” con niveles de significación determinados por el usuario (Gras et al., 1996; Gras & Bailleul, 2000; Gras & Régnier, 2009). También se utilizó hoja de cálculo para estadística descriptiva básica y representación gráfica.

Por razones de espacio, presentaremos aquí de manera muy resumida algunos de los resultados obtenidos. Uno de ellos se relaciona a la percepción de los docentes sobre la supuesta homogeneidad de los equipos de trabajo: Con el transcurrir de las clases, el equipo docente se había construido una imagen de cada uno de los cinco equipos de estudiantes: “El equipo ... trabaja mejor que el equipo...”, “Definitivamente el equipo... produce muy buenas reflexiones y análisis de los modelos”, “Lamentablemente el equipo ... no logra avanzar como deseamos”, etc. Expresiones de ese tipo se compartían en las reuniones de trabajo y progresivamente se iban consolidando entre los docentes.

Sin embargo, las primeras exploraciones de los datos no resultaban en consonancia con esa percepción. El primer análisis de los datos mediante indicadores descriptivos y la búsqueda de reglas cuasi-implicativas no indicaron tendencias significativas en función de los equipos. Por razones de espacio no desarrollamos aquí ese primer estudio exploratorio.

## Análisis de los razonamientos de estudiantes en modelaciones matemáticas

Se descartó entonces el agrupamiento de estudiantes por pertenencia a los respectivos equipos y se exploró la posibilidad de una generación de grupos por aprendizaje inicialmente no supervisado (K-means).

El criterio de constitución de los clusters fue en base a la cercanía de los individuos. Esta cercanía está basada en la coincidencia o no en las respuestas (dos estudiantes son cercanos si sus respuestas en conjunto se asemejan). Para ello, nos servimos del algoritmo k-means y en lo que respecta a la cantidad de grupos (clusters) a retener, se testearon dos métodos: uno en función de la varianza intra cluster (Figura 3) y el otro en función de la evolución de los índices llamado silhoutte (Figura 4).

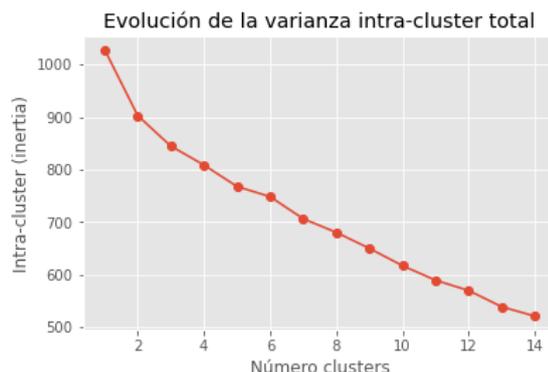


Figura 3 – Evolución de la varianza intra-cluster en función del número de clusters  
Fuente: Elaboración propia

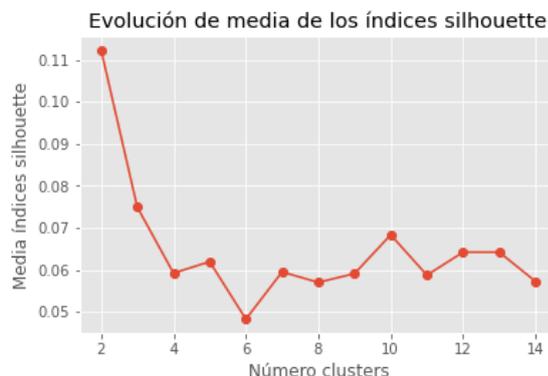


Figura 4 – Evolución de media de los índices silhoutte  
Fuente: Elaboración propia

La ausencia de “un codo” en el gráfico de la Figura 3 pone en evidencia la inexistencia de un número de clusters que represente una disminución significativa de la varianza intraclusters. En la figura 4 (método silhoutte) (Das, 2021) sugiere considerar 4 clusters. Sin embargo y a los fines de no complejizar las interpretaciones, se retuvieron tres clusters.

La Figura 5 (y Tabla 2) muestran la distribución de los estudiantes para estos tres clusters:

	Clusters		
	G0	G1	G2
Cantidad estudiantes	12	20	17

Tabla 2 – Distribución de Estudiantes por clusters

Fuente: Elaboración propia

## Análisis de los razonamientos de estudiantes en modelaciones matemáticas

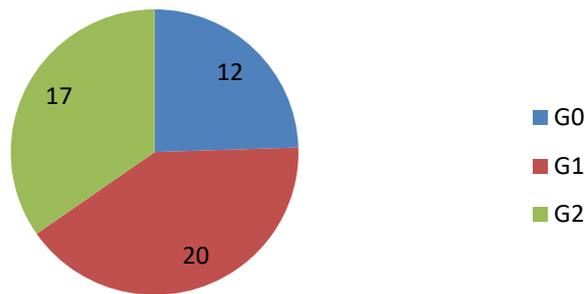


Figura 5 – Cantidad de estudiantes por cluster para una configuración de 3 clusters.  
Fuente elaboración propia

La Figura 6 muestra la distribución de los estudiantes para cada cluster (G0, G1 y G2) en función del equipo al cual pertenecían originalmente (Arctan, Pit, Loscar, LosPi y Eucli).

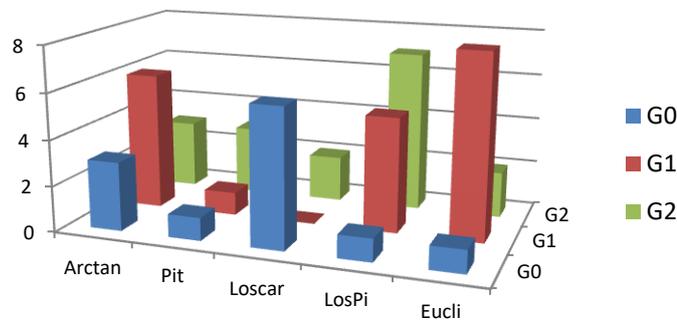


Figura 6 – Distribución de estudiantes por clusters en función del equipo de pertenencia  
Fuente: Elaboración propia

Una primera caracterización de cada uno de estos tres clusters (G0, G1 y G2) puede observarse en la Figura 7. Allí, las respuestas se encuentran categorizadas según la tipificación antes indicada en función de la validez de la respuesta. La Tabla 3 muestra la codificación asociada a cada tipificación y su significado:

Código de tipificación	Significado
Dcor	Pregunta Descriptiva respondida correctamente
Dinco	Pregunta Descriptiva respondida incorrectamente
Dnores	Pregunta Descriptiva no respondida
Icor	Pregunta Interpretativa respondida correctamente
Iinco	Pregunta Interpretativa respondida incorrectamente
Inores	Pregunta Interpretativa no respondida
Acor	Pregunta Analítica respondida correctamente
Ainco	Pregunta Analítica respondida incorrectamente
Anores	Pregunta Analítica no respondida
Just	Se solicita una justificación, respondida correctamente
Nojust	Se solicita una justificación y no se responde
Per	Se solicita una justificación, respuesta aceptable aunque

confusa

**Tabla 3.** Código de variables y tipificación atribuida

Fuente: Elaboración propia

A su vez, debido a la cantidad de estudiantes dispar en cada cluster, se promediaron la cantidad de respuestas en función del número de estudiantes en cada cluster. Así, por ejemplo, el cluster G0 se caracteriza por estudiantes donde no se han respondido 11 preguntas analíticas (Anores) por estudiante, en promedio. El mismo cluster (G0) se caracteriza también por estudiantes que en promedio no han respondido 4 preguntas interpretativas (valores redondeados a los fines interpretativos). La Figura 3 reúne esa información para cada uno de los tres clusters:

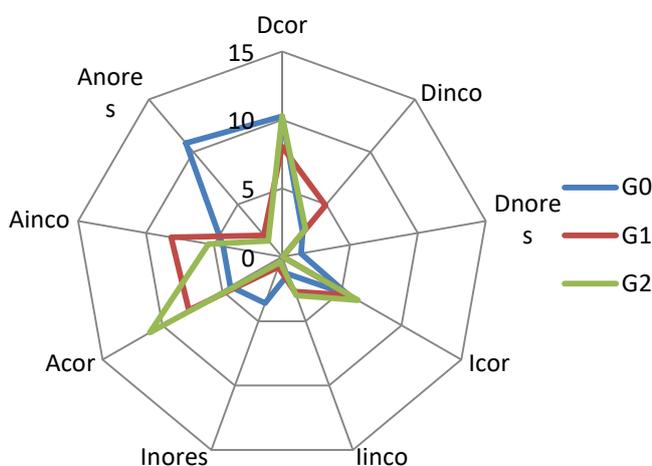


Figura 7 – Proporción de respuestas por estudiante según cluster  
Fuente: Elaboración propia

Esta representación pone en evidencia:

Respecto a las respuestas a preguntas interpretativas (Icor, Inco, Inores)

Los tres clusters se caracterizan por una tendencia parecida en lo que respecta a las preguntas del tipo interpretativas, aunque destacamos al cluster G0 que se caracteriza por preferir no responder una pregunta interpretativa (Inores cercano a 4) y así evitar las respuestas incorrectas ( $Inco < 1$ ).

Los estudiantes tienden a responder correctamente la misma cantidad de preguntas interpretativas (Icor): entre 5 y 6 por estudiante. Son escasas las preguntas interpretativas no respondidas (Inores); menos de una por estudiante, salvo para el cluster G0 como detallamos arriba.

Respecto a las respuestas a las preguntas descriptivas (Dcor, Dinco, Inores)

a) se esperaba (y se constató) una alta tasa de respuestas correctas (Dcor) por el hecho que este tipo de preguntas eran relativamente sencillas. En general, la tarea a realizar para responderlas consistía en observar algún fenómeno de escasa o nula complejidad de la modelación y responder en consecuencia. En esta categoría se incluyeron las preguntas referidas a la determinación de unidades de medida de una variable dada (m, o m<sup>2</sup>, según el caso). Precisamente los errores observados (Dincor) en el cluster G1 (aproximadamente 5 incorrectas por estudiante) se explican en buena medida por asociar incorrectamente una

unidad de medida a un número dado. Se interpreta esto como una dificultad en asociar elementos básicos de la modelación con el contexto de referencia.

Respecto a las respuestas a las preguntas analíticas (Acor, Aincor, Anores)

Es en este tipo de preguntas donde se observan las mayores diferencias entre los clusters. El cluster G0 había mostrado una tendencia a no responder las preguntas tanto directas (Dnores entre 1 y 2) como indirectas (Inores entre 3 y 4), al menos de manera más acentuada que los otros clusters. Con las preguntas analíticas, este cluster (G0) acentúa la tendencia: 11 preguntas analíticas no respondidas por estudiante. Mientras que las y los estudiantes de los otros dos clusters, se resisten a no responder este tipo de preguntas (entre 1 y 2 preguntas analíticas sin responder). Se desconoce si las y los estudiantes del cluster G0 son “prudentes” y prefieren no responder a responder erróneamente o, si por el contrario, carecen de elementos para responder.

En la tipificación del tipo analítica se encuentra la mayor diferencia entre los clusters G1 y G2. El cluster G1 balancea sus tasas de correctas e incorrectas (entre 7 y 8 respuestas por estudiante cada una). Por su parte, el cluster G2 logra mejores resultados con una tasa de respuestas correctas (Acor aproximadamente 11) que duplica a las incorrectas (Aincor entre 5 y 6). Existe entonces claramente tres grupos de estudiantes que se van progresivamente diferenciando en la medida que las preguntas van ascendiendo en complejidad: desde las descriptivas hacia las analíticas.

Ante esto, nos preguntamos si existía algún tipo de relación entre las respuestas, algún tipo de tendencia que caracterice posibles relaciones entre variables, más allá de la tipificación en descriptivas, interpretativas y analíticas. En particular nos interesaba analizar la existencia de relaciones (cuasi) implicativas que puede expresarse bajo la forma “si...entonces...”.

Para ello, exploramos los datos con el método conocido como análisis estadístico implicativo disponible en el software CHIC. Se utilizó la versión 6 del software, parametrizando el mismo en el tipo de implicación según la teoría clásica con el tipo de ley Binomial, esto principalmente debido a la escasez de datos en cada cluster (Gras et al., 1996; Gras & Bailleul, 2000; Gras, Régnier, & Guillet, 2009).

Se realizó una primera exploración considerando de manera conjunta la totalidad de las y los estudiantes, es decir los tres clusters reunidos de manera indistinta. Por razones de espacio, no presentamos aquí los resultados de esa primera exploración por considerarlos con escaso aporte a los fines del estudio. Una segunda exploración se realizó con los tres clusters de manera separada, por lo que se realizaron tres análisis implicativos, uno para cada cluster (G0, G1 y G2).

Todos los análisis fueron realizados con un índice implicativo de 0.85. Este índice, que no representa un porcentaje, puede interpretarse como un indicador del nivel de permisividad de contraejemplos admitidos en una regla retenida. En la medida que se desciende del límite superior (valor máximo 1) el algoritmo se torna relativamente menos exigente en la generalización de las reglas, admitiendo así más casos que contradicen las reglas en cuestión. La literatura considera al valor 0.85 como uno admisible para situaciones como la de este estudio, donde se cuenta con una cantidad pequeña de datos y una importante cantidad de variables.

Los gráficos que siguen muestran relaciones cuasi implicativas entre las variables. A modo ilustrativo y a los fines interpretativos, si por ejemplo se observa la regla de la figura 8 ella puede interpretarse como la (cuasi) implicación siguiente: “quienes no responden a la pregunta 22 (22nores) entonces responden de manera incorrecta a la pregunta 35 (35incor)”.

El término “cuasi” es utilizado aquí para destacar la existencia de casos (raros probabilísticamente hablando) que contradicen la implicación en sentido estricto. La rareza de los casos que la contradicen justifica entonces retener la regla en cuestión.



Figura 8 – Regla ilustrativa  
Fuente: Elaboración propia

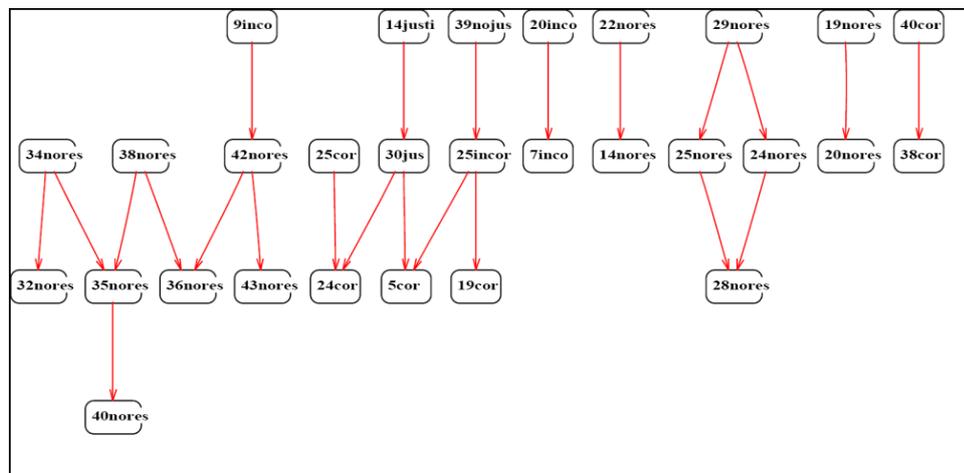


Figura 9 – Grafo implicativo cluster G0. Índice implicativo 0.85  
Fuente: Elaboración propia

Como se bosquejó anteriormente, el cluster G0 se caracteriza por la no respuesta a preguntas del cuestionario. El gráfico de la Figura 7 muestra las relaciones cuasi implicativas entre esas no respuestas. Así, aparece un conjunto importante de variables (XXnores) que se relacionan entre sí. Pocas son las reglas que relacionan respuestas correctas. La Figura 10 representa una parte de la Figura 9, la que contiene reglas relacionando respuestas correctas (14just→30jus→24cor; 25cor→24cor).

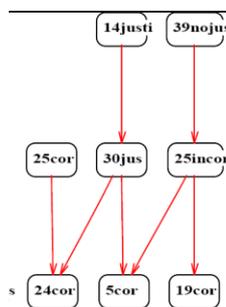


Figura 10 – Recorte de Figura 7  
Fuente: Elaboración propia

## Análisis de los razonamientos de estudiantes en modelaciones matemáticas

Las variables antes mencionadas (14just; 30just; 24cor y 25cor) son respuestas correctas (o justificadas correctamente) a preguntas analíticas. Todas tienen dos denominadores en común: por un lado, ellas preguntan sobre cuestiones relacionadas a un deslizador (frente), los efectos que produce el cambio de valor de ese deslizador en el gabinete y las implicancias en el modelo; es decir aspectos que si bien son analíticos en la tipificación, están relacionados entre sí. Por otro lado, esa cuestión fue abordada en clase en reiteradas oportunidades. En este sentido, podemos decir que estas preguntas no resultan nuevas u originales para las y los estudiantes.

Las reglas entonces detectadas en G0 referidas a respuestas correctas a preguntas analíticas parecen indicar que las y los estudiantes de este cluster se caracterizan por haber retenido los análisis desarrollados en clase, pero que no pudieron extender ese tipo de estrategia a otras cuestiones planteadas.

La figura 11 representa las reglas cuasi implicativas (siempre con un índice de 0.85) para los 20 integrantes del cluster G1.

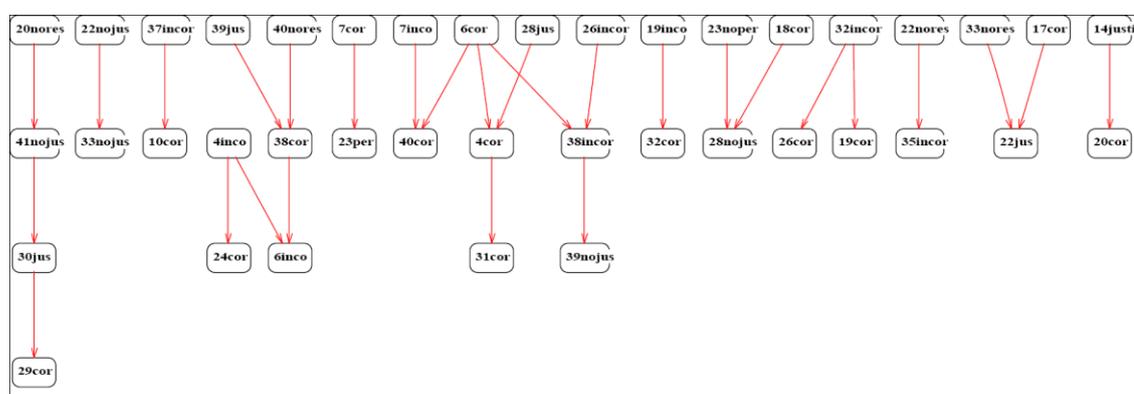


Figura 11 – Grafo implicativo cluster G1. Índice implicativo 0.85

Fuente: Elaboración propia

Este cluster resulta algo más homogéneo que el anterior pues aparece un número mayor de reglas a un mismo índice de significación (0.85). Las posibilidades de interpretación se multiplican y por ende las de información. Nótese el cruzamiento entre variables en el sentido de una respuesta correcta implica otra incorrecta (ejemplo: 6cor→38inco) y viceversa (37inco→10cor). Retendremos algunas reglas que no representan cruzamiento entre correcta e incorrecta, en particular cuando el antecedente y el consecuente son respuestas correctas.

Algunas de estas reglas permanecen en la tipificación descriptiva, como por ejemplo 6cor→4cor y 7cor→23per (antecedente y consecuente son respuestas correctas a variables descriptivas). Otras son casi triviales, como por ejemplo 30just→29cor y 39just→38cor. En ambos se trata de una pregunta interpretativa que se responde correctamente (29cor y 38cor) y de sus respectivas justificaciones que son analíticas y se responden correctamente (30just y 39just). Estas dos reglas resultan triviales pues si se respondió correctamente la justificación (por ejemplo 30just), es de esperar que se haya respondido bien la pregunta sobre la cual se pidió una interpretación (29cor).

La regla 14just→20cor indica que quienes pueden justificar una elección de su equipo (14just) comprenden el interés por encontrar el mínimo de caras laterales (20cor). Ambas preguntas son analíticas y sugieren que la comprensión de la decisión colectiva del equipo está asociada a la comprensión general que motiva la modelación: encontrar un mínimo de caras laterales.

Destacamos además en el cluster G1 la aparición de más reglas en comparación con el cluster G0. Sin embargo y como observamos anteriormente ellas sugieren una cierta inestabilidad en la comprensión de los conceptos pues en buena medida esas reglas relacionan respuestas correctas con respuestas incorrectas. Si observamos las reglas del tipo  $xcor \rightarrow yinco$  estas reglas estarían sugiriendo una suerte de frontera de comprensión donde el límite estaría dado por el antecedente: al avanzar de  $xcor$  en la dirección de  $yinco$ , se pierde la comprensión pues se pasa de una respuesta correcta a una incorrecta. Tal es el caso de la sucesión  $6cor \rightarrow 38incor \rightarrow 39nojust$ : la comprensión estaría dada en el plano de las preguntas descriptivas (6cor) y la frontera serían las preguntas analíticas. Así, en cuanto se pasa de preguntas descriptivas a preguntas analíticas, se pasaría también de la comprensión a la incomprensión.

La Figura 12 representa el conjunto de cuasi implicaciones para el cluster G2, siempre con un índice de 0.85.

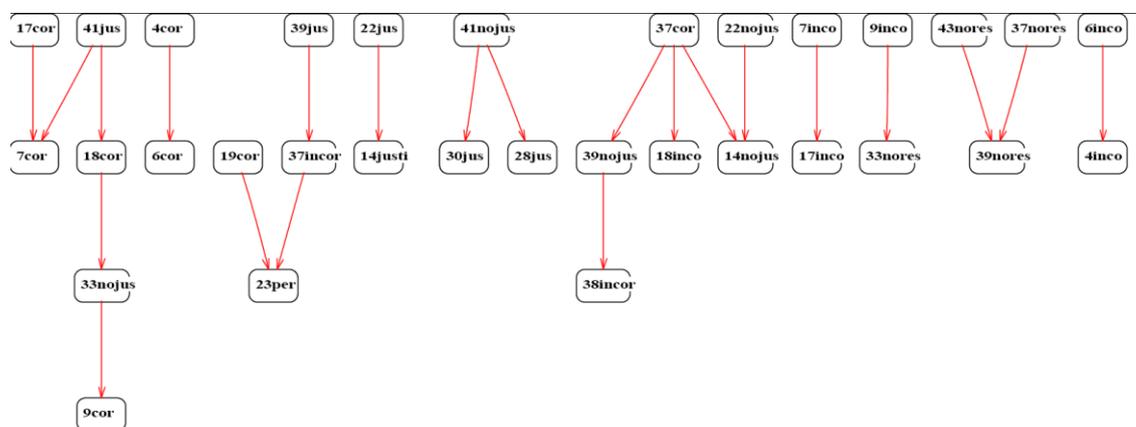


Figura 12 – Grafo implicativo cluster G2. Índice implicativo 0.85

Fuente: Elaboración propia

G2 es el cluster que obtuvo mejores resultados en términos de cantidad de respuestas correctas en las tres tipificaciones (descriptivas, interpretativas y analíticas). La expectativa era entonces encontrar entramados de reglas de respuestas correctas en las tres tipificaciones. Nótese que esto no resulta tan evidente. Las siguientes son las cuasi implicaciones de respuestas correctas donde tanto antecedente como consecuente son analíticos (dos únicamente):  $41just \rightarrow 18cor$  y  $22just \rightarrow 14just$ .

Este cluster, que tiene mejores resultados en término de cantidad de respuestas correctas a preguntas analíticas que los demás, no logra destacarse en términos de tendencias de reglas. Esto sugiere una suerte de inestabilidad en términos de grupo en la comprensión a cuestiones analíticas: algunos comprenden ciertas cuestiones analíticas, otros comprenderán otras cuestiones analíticas; pero no hay homogeneidad al interior del grupo. Se trataría de un conjunto de estudiantes con relativo buen nivel de respuestas correctas pero con dispersión en cuanto a lo que se comprende.

#### 4. 4 Conclusiones y Discusión

Una de las primeras conclusiones que destacamos es el haber descartado la representación que el grupo de docente se había construido respecto al desempeño de cada equipo. El análisis

de las respuestas puso en evidencia que tal construcción por parte de los docentes no era representativa de lo que realmente se producía al interior de cada equipo: resulta indispensable entonces observar al interior de cada uno y en todo caso considerar agrupación con criterios más objetivos, como es el caso del método k-medias utilizado. Esto representa un aprendizaje para las y los docentes.

Probablemente debamos proponer cuestionarios de este tipo con más frecuencia para obtener información más precisa de cada estudiante y a su vez longitudinal en el tiempo. Para ello sin duda, resulta necesario estandarizar este tipo de análisis para facilitar la tarea no tanto de implementación sino de obtención de resultados para una rápida extracción de información de los aprendizajes en las y los estudiantes.

Otra de las conclusiones se refiere a lo que podríamos denominar como la homogeneidad (o heterogeneidad) de los aprendizajes. Esta cuestión se observa más claramente en el tercer cluster (G2): Hubo mayor cantidad de respuestas correctas, pero ello no se manifestó en tendencias observables a través de reglas cuasi implicativas.

Entendemos que al menos dos elementos podrían explicar esta escasa homogeneidad: por un lado, la falta de hábito (y por lo tanto de desarrollo de habilidades y competencias) para realizar análisis de modelaciones. Destacamos que las y los estudiantes en cuestión eran ingresantes a primer año en una carrera universitaria y, según lo declarado por ellas y ellos de manera informal, las clases de matemática a las que estaban habituados eran del formato tradicional, con desarrollo intensivo de ejercicios.

Hay entonces aquí una dificultad proveniente de la falta de habilidades y competencias. Sin duda a ello debemos adicionar las relacionadas a la gestión de un nuevo contrato didáctico, donde lo que se espera de ellos es bastante diferente a lo que se esperaba en cursos de matemática anteriores.

Otro elemento que seguramente ha impactado en la escasa homogeneidad observada es el referido a la pandemia COVID 19. Al momento de realizar el proyecto de los potabilizadores y, más precisamente, al de la realización del cuestionario, los estudiantes llevaban ya un año en cuarentena estricta con clases virtuales y medio año en formato híbrido, con predominancia de clases virtuales. Sin dudas este singular fenómeno condicionó las posibilidades de aprendizaje de las y los estudiantes participantes.

No debemos descartar tampoco la calidad del cuestionario, donde ya percibimos posibles mejoras, tanto en la formulación de las preguntas como en el foco de las mismas a los efectos de un mejor direccionamiento de las mismas a los temas de interés de la investigación.

En los próximos estudios tenemos previsto tanto una mejora en la estandarización del proceso de análisis como una mejor focalización en las preguntas.

## 5. 5 Referencias

BEDNARZ, N. E. A. Contextualiser pour enseigner les mathématiques: un enjeu de formation. *Annales de didactique et de sciences cognitives*, v. 23, p. 69-105, 2018.

BLOMHØJ, M. 2019. Towards Integration of Modelling in Secondary Mathematics Teaching. In STILLMAN, G. A.; BROWN, J. P. (eds.). *Lines of Inquiry in Mathematical Modelling Research in Education: ICME-13 Monographs*: Springer, Cham., 2019. p. 37-52.

BLUM, W.; BORROMEO FERRI, R. Mathematical modelling: can it be taught and

learnt? *Journal of Mathematical Modelling and Application*, v. 1, n. 1, p. 45-58, 2009.

BOALER, J. The role of contexts in the mathematics classroom: Do they make mathematics more "real"? *For the learning of mathematics*, v. 13, n. 2, p. 12-17, 1993.

BOALER, J. Mathematical modelling and new theories of learning. Teaching mathematics and its applications. *International journal of the IMA*, v. 20, n. 3, p. 121-128, 2001.

BROWN, J.; IKEDA, T. Conclusions and future lines of inquiry in mathematical modelling research in education. In STILLMAN, G. A.; BROWN, J. P. (eds.). *Lines of Inquiry in Mathematical Modelling Research in Education: ICME-13 Monographs*: Springer, Cham., 2019. p. 233-253.

BROWN, R.; REDMOND, T. Privileging a contextual approach to teaching mathematics: a secondary teacher's perspective. *Proceedings of the 40th Annual Conference of the Mathematics Education Research Group, Australasia, Melbourne*. p.109-116, 2017.

BUXTON, C. Creating Contextually Authentic Science in a "Low-Performing" Urban Elementary School. *Journal of research in science teaching*, v. 43, n. 7, p. 695-721, 2006.

CARON, F. Approaches to investigating complex dynamical systems. In STILLMAN, G. A.; BROWN, J. P. (eds.). *Lines of Inquiry in Mathematical Modelling Research in Education: ICME-13 Monographs*: Springer, Cham., 2019. p. 83-103.

CARRANZA, P.; CORDERO, F.; ROSA, M.; OREY, D. *La modelación en la vida de la gente (Extensión Científica)*. Editorial Gedisa: España. 2022. 117 pp.

CARRANZA, P.; LETROURNEAU, M.; NAVARRO, M. Analogical modelling and analytical modelling. Different approaches to the same context? In CARRANZA, P.; CORDERO, F.; ROSA, M.; OREY, D. (eds), *Mathematical modelling programs in Latinoamérica*: Springer: Cham., 2022. p. 221-244.

CARRANZA, P.; MORALES, A. Estudio del rotor en un proyecto de cálculo, construcción e instalación de molinos Savonius en clase de matemáticas. In CORDERO, F. (ed.), *Modelación y transversalidad de saberes. La matemática en la ingeniería: situaciones de aprendizaje*: Editorial Gedisa, España, 2021.

CARREJO, D.; MARSHALL, J. 2007. What is mathematical modelling? Exploring prospective teachers' use of experiments to connect mathematics to the study of motion mathematics. *Education Research Journal*, v. 19, n. 1, p. 45-76, 2007.

CZOCHER, J. Precision, priority, and proxies in mathematical modelling. In STILLMAN, G. A.; BROWN, J. P. (eds.). *Lines of Inquiry in Mathematical Modelling*

*Research in Education: ICME-13 Monographs*: Springer, Cham., 2019.p. 105-123.

DAS, A. *Unsupervised learning techniques using python — k means and silhouette score for clustering*. 2021. (URL:<https://towardsdatascience.com/unsupervised-learning-techniques-using-python-k-means-and-silhouette-score-for-clustering-d6dd1f30b660> – último acceso: 2022-10-22).

DE LOIOLA ARAÚJO, J. Toward a framework for a dialectical relationship between pedagogical practice and research. In STILLMAN, G. A.; BROWN, J. P.(eds.). *Lines of Inquiry in Mathematical Modelling Research in Education: ICME-13 Monographs*: Springer, Cham., 2019.p. 21-36.

DEVELAY, M. Le sens dans les apprentissages: du désir au passage a l'acte. *Pédagogie collégiale*, v. 7,n. 4, p. 23-26, 1993.

DEVELAY, M. *Donner du sens à l'école*. ESF sciences humaines, France. 2004. 128 pp.

GAN, Q.; ZHANG, N. Extremely cost-effective and efficient solar vapor generation under non-concentrated illumination using thermally isolated black paper. *Global Challenges*, v. 1, n. 2, p. 1600003, 2017.

GRAS, R.; AG ALMOULOU, S.; BAILLEUL, M.; LARHER, A.; POLO, M.; RATSIMBA-RAJOHN, H.; TOTOHASINA, A. *L'implication statistique*. Editions La Pensée Sauvage, France. 1996. 336 pp.

GRAS, R.; BAILLEUL, M. Les fondements de l'analyse statistique implicative et leur prolongement pour la fouille de données. *Mathématique et Sciences Humaines* ,v. 120, n. 154-155, p. 9-29, 2001.

GRAS, R.; RÉGNIER, J. C. Analyse implicative des variables binaires. Intensité implicative. Intensité entropique. In GRAS, R., RÉGNIER, J.C.; GUILLET, F. (eds.), *Analyse Statistique Implicative: Une méthode d'analyse de données pour la recherche de causalités*: Cépaudès Editions, Toulouse, 2009. p. 17-129.

GRAS, R.; RÉGNIER, J.C.; GUILLET, F. *Analyse Statistique Implicative: Une méthode d'analyse de données pour la recherche de causalités*. Cépaudès Editions: Toulouse, 2009. 493 pp.

HANKELN, C.; ADAMEK, C.; GEREEFRATH, G. Assessing sub-competencies of mathematical modelling—development of a new test instrument. In STILLMAN, G. A.; BROWN, J. P.(eds.). *Lines of Inquiry in Mathematical Modelling Research in Education: ICME-13 Monographs*: Springer, Cham., 2019.p. 143-160.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. H. *The elements of statistical*

*learning: data mining, inference, and prediction*. Springer, 2001. 758 pp.

ORTEGA, M.; ALBARRACÍN, L. The Influence of Technology on the Mathematical Modelling of Physical Phenomena. In STILLMAN, G. A.; BROWN, J. P.(eds.). *Lines of Inquiry in Mathematical Modelling Research in Education: ICME-13 Monographs*: Springer, Cham., 2019.p. 161-178.

PIRYONESI, M.; EL-DIRABY, T. E. Role of data analytics in infrastructure asset management: overcoming data size and quality problems. *Journal of transportation engineering*, v. 146, n. 2, p. 04020022, 2020.

SHOWALTER, D. Place-based mathematics education: a conflated pedagogy? *Journal of research in rural education*, v. 28,n. 6, p. 1-13, 2013.

SPANDAW, J. Modelling mathematics teachers' professional development. Proceedings of the *6th Congress of the European Society for Research in Mathematics Education*. Education–Working group, Lyon, France, v. 11, p. 2076-2085, 2009.

YUANITA, P.; ZULNALDI, H.; ZAKARIA, E. The effectiveness of realistic mathematics education approach: The role of mathematical representation as mediator between mathematical belief and problem solving. *PLoS One*, v. 13,n. 9, p. e0204847, 2018.

ZUBI, I.; PELED, I.; YARDEN, M. Modelling tasks and students with mathematical difficulties. In STILLMAN, G. A.; BROWN, J. P.(eds.). *Lines of Inquiry in Mathematical Modelling Research in Education:ICME-13 Monographs*: Springer, Cham., 2019.p. 213-231