

# **Aprendizaje No Supervisado aplicado a la Clasificación de Competencias de Maestros del Sistema Educativo Colombiano**

**Autores:** Edgar Javier Abril Duran - Rebeca Gamboa Venegas - Jorge Esteban Caballero - Jesús Parada Perez

## **Resumen**

El rápido avance tecnológico y económico global presenta nuevos retos a la sociedad, uno de ellos es, la mejora en la calidad de la educación. Este estudio pretende evaluar mediante técnicas de Inteligencia Artificial el desempeño de los docentes colombianos en cinco competencias y utilizar esta información para orientar su desarrollo profesional. Se cuenta con un conjunto de datos de la Fundación Future Education que contiene respuestas a 20 preguntas sobre habilidades de los docentes y características sociodemográficas. Los resultados obtenidos muestran seis agrupaciones principales de docentes, donde se distinguen aquellos con baja capacidad para la resolución de conflictos y pensamiento crítico, se muestran posibles dos posibles agrupaciones de competencias 1) Comunicación y Trabajo Colaborativo, 2) Creatividad y Pensamiento Crítico, potencialmente estas agrupaciones permiten una capacitación más efectiva en los grupos de interés. Finalmente, se concluye que en la encuesta posee preguntas con información que no aporta a la varianza, por lo que es posible mejorar la confiabilidad del instrumento reevaluando algunas de las preguntas realizadas.

## **Introducción**

Los sistemas educativos latinoamericanos presentan una amplia diversidad de enfoques y desafíos. A lo largo de la región se observan diferencias en la calidad y accesibilidad de la población a la educación. El avance tecnológico, la globalización y una profunda transformación de la sociedad y los mercados, coloca a la educación latinoamericana en aprietos, ya que para hacer frente a estos desafíos requiere una transformación de los sistemas educativos, que son el motor de desarrollo y movilidad social en América Latina.

La organización colombiana Future Education, tiene como propósito promover prácticas innovadoras en la educación. Dentro de sus proyectos se encuentra la aplicación de una encuesta para medir la percepción de los docentes frente a las “*Competencias del Siglo XXI*”. Para identificar el nivel de desempeño alcanzado por el docente, un grupo de expertos de Future Education diseñó un cuestionario de 20 preguntas asociadas a 5 competencias: creatividad e innovación, pensamiento crítico, resolución de problemas, comunicación y trabajo colaborativo. La organización no ha verificado la pertinencia del constructo para cada competencia y actualmente emplea un promedio simple que asigna pesos equiprobables a cada pregunta para medir el nivel de desempeño de los docentes, lo que ha sido objeto de críticas por parte de los aliados.

Este proyecto emplea las respuestas al cuestionario de una muestra compuesta por 5270 docentes. Inicialmente se evalúa la robustez del instrumento empleado para recopilar los datos, siguiendo los criterios definidos por Fabila, Minami & Izquierdo (2013). Para optimizar la calidad de la información, se aplica la técnica de reducción de dimensionalidad llamada Análisis de Componentes Principales (PCA). Esto identifica la posibilidad de ajustes en la encuesta para obtener datos de mayor calidad.

En una segunda etapa de este proyecto, se identifican agrupaciones a los docentes en función de sus niveles de competencia, mediante la asignación de puntajes basados en sus respuestas a las preguntas, lo que permitirá adaptar la educación según el "perfil típico" de cada grupo.

El alcance de este proyecto se limita a la identificación y validación de estos grupos, con el propósito de responder a la pregunta crucial: ¿Existen agrupamientos claramente distinguibles basados en las respuestas y características sociodemográficas de los docentes, que permitan en el futuro ofrecer recomendaciones o asignar niveles de competencias multidisciplinarias para brindar educación especializada?

El potencial impacto de este enfoque radica en la optimización de la capacitación de los docentes en competencias específicas, lo que podría acelerar significativamente la mejora en la calidad de la educación en

América Latina. Al definir con precisión las competencias necesarias para el futuro y proporcionar una formación adecuada a los docentes y tutores, se facilita el acceso de los estudiantes a las herramientas esenciales para su desarrollo con mayor rapidez y calidad. Este enfoque innovador, aunque incipiente en América Latina, presenta un caso prometedor para la aplicación de estas metodologías.

### Revisión preliminar de antecedentes en la literatura

En la literatura, se han tenido acercamientos más que todo enfocados a estudiantes, cómo en el documento desarrollado por Bagunaid et al en el 2022. Aquí, los autores utilizan diversas metodologías para llegar a un puntaje que se usa clustering por DBSCAN con distancia Mhalanobis. Luego, se hace un ajuste para el desempeño individual del estudiante y esta información procede a un sistema de recomendación para brindar opciones de aprendizaje adicional a los estudiantes. Este puede ser un buen acercamiento al problema en cuestión y el agrupamiento de competencias por medio de DBSCAN.

Revisando el Systematic Review para el Student Assessment de González-Calatayud, se observa que no existen contribuciones significativas en América Latina sobre el uso de inteligencia artificial para evaluar al estudiante, hasta su publicación en junio del 2021. No obstante, en 15 de los 22 papers que eligieron se tocan temas de Inteligencia Artificial para el assessment formativo, además muestran 10 papers que tratan del scoring automático de competencias afines a las evaluadas en la encuesta usada en este proyecto (por ejemplo, en Kaila et al. se evalúa el trabajo colaborativo).

### Materiales y Métodos

La base de datos suministrada por Future Education tiene un total de 5270 encuestas realizadas entre 2022 y 2023 por los docentes de las instituciones educativas adscritas a la organización. A continuación, se presentan las 33 variables suministradas:

Variable	Descripción	Variable	Descripción
ID	Identificador de la encuesta	departamento	Departamento de aplicación
Year	Año de aplicación	ubicacioninstitucion	Indica el área (rural/urbana) de la IE
Mes	Mes de aplicación	cargo	Cargo de la persona encuestada
marcatemporal	Fecha y hora de aplicación	edades	Rango de edad de la persona encuestada
formulario	Tipo de formulario aplicado	formacion	Nivel de formación
solución	Programa implementado en la organización	sexoinscrito	sexo (hombre/mujer)
regional	Ubicación geográfica (Caribe, Antioquia, Suroccidente, Centro, Nacional)	p1 a p20	preguntas para la evaluación de competencias

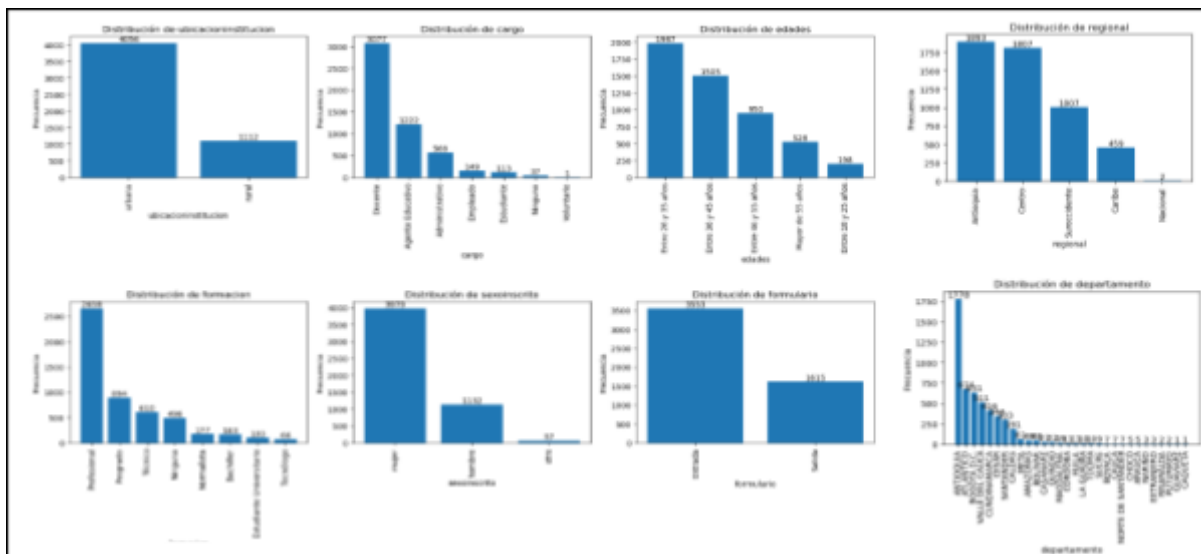
### Limpieza de datos: evaluación de datos faltantes

Se encuentran datos faltantes en las variables: regional, departamento, ubicacioninstitucion, cargo, edades, formacion, sexoinscrito; se procede a eliminar los datos faltantes en la variable regional, ya que para la organización es importante saber a qué región pertenecen los docentes. La base de datos después de eliminar los datos faltantes tiene 5186 observaciones y 33 variables.

Además, a partir de un análisis de los dominios de las variables, se encuentran los siguientes errores en Departamento, Ubicacioninstitucion, Edades y Sexoinscrito, en muchos casos era necesaria la eliminación o imputación de los datos. Para corregirlos se realiza imputación univariada con la moda; esta técnica consiste en utilizar únicamente la información que posee la variable que se imputa y establecer valores en las observaciones correspondientes.

### Visualización:

Para conocer a profundidad los datos se realiza un análisis visual de la distribución de las variables. La mayoría de las personas encuestadas son mujeres, profesionales, docentes, ubicadas en regiones urbanas; las cuales se ubican entre los 26 y 35 años. Además, la mayoría de los encuestados son de Antioquia.



Por otro lado, las variables que representan las respuestas de las preguntas son variables categóricas con valores que pueden ser transformados para su análisis, estos valores únicos son: “Nunca”, “A veces”, “Constantemente”, “Siempre”, antes de realizar la transformación es necesario invertir el valor de 9 preguntas, específicamente las preguntas 2, 4, 7, 8, 11, 14, 16, 18 y 20 para los cuales “Nunca” tendrá un valor de 1, “A veces” será igual a 2, “Constantemente”, será igual a 3 y “Siempre” igual a 4. Con ello, da para verificar si existe algún tipo de correlación entre las respuestas de las preguntas evaluadas, pero no se observa correlación alguna entre las variables, para una correcta visualización del gráfico se adjunta en el Anexo 1.

## Resultados

### Consistencia interna del instrumento / Coeficiente Alfa de Cronbach

Este coeficiente es comúnmente utilizado para evaluar la confiabilidad y la consistencia de investigaciones de cuestionarios o entrevistas. Shavelson (2009) indica que dicho coeficiente tiene es beneficioso pues 1) provee una medida razonable de confiabilidad en un único test, 2) la fórmula general del coeficiente permite su aplicación a cuestionarios de opción múltiple de escalas dicotómicas o de variables categóricas politómicas y 3) puede ser calculado con principios estadísticos básicos.

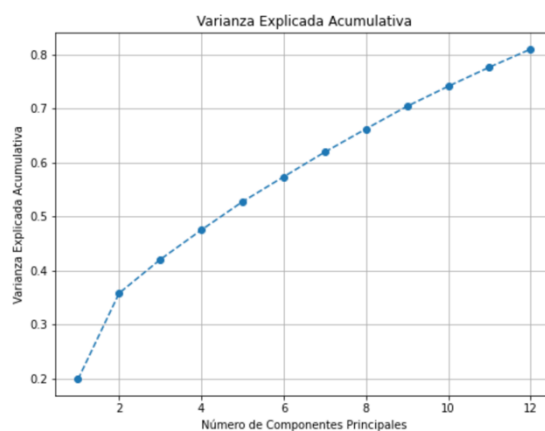
En términos generales la aplicación del coeficiente a la encuesta da como resultados la obtención de un valor general de Alpha de 0,726 lo que indica una confiabilidad buena. Los ítems de la escala están razonablemente correlacionados y la medida es considerada confiable para su uso en la mayoría de los casos. Sin embargo, las preguntas que miden cada competencia no son consistentes, tienen una baja confiabilidad, lo que implica que no parece ser la mejor manera de agruparlas.

Cronbach's Alpha general: (0.7657319479217339, array([0.756, 0.775]))  
 Cronbach's Alpha creatividad: (0.3817759051798002, array([0.354, 0.409]))  
 Cronbach's Alpha resolución: (0.21097282270979179, array([0.175, 0.246]))  
 Cronbach's Alpha pensamiento critico: (0.5193599078446379, array([0.498, 0.54 ]))  
 Cronbach's Alpha trabajo colaborativo: (0.34554037563285445, array([0.316, 0.374]))  
 Cronbach's Alpha comunicacion: (0.5190298579137435, array([0.497, 0.54 ]))

### Reducción de dimensionalidad /Análisis de Componentes Principales (PCA)

Se aplica el PCA para conocer si la información suministrada por la encuesta puede ser representada por un conjunto más pequeño de variables que simplifiquen el análisis y la interpretación de los datos. Además, de conocer cuales son las preguntas más representativas de la encuesta. Se prefiere el PCA ante otros métodos de reducción de dimensionalidad por sus características de conservación de la varianza y por su capacidad de preservar gran parte de la información.

Los resultados arrojan que es posible explicar el 80% de la varianza de los datos con utilización de 12 de los 20 componentes analizados. Las preguntas más representativas son las 7, 12, 5, 15, 9, 2, 17, 18, 20, 6, respectivamente. Observando la relación de las preguntas con las competencias, se concluye que las competencias Resolución de conflictos (p5-p8) y Comunicación (p17-p20) aportan 3 de sus 4 preguntas a la explicación del 80% de la varianza. Por otra parte, se puede intuir que Creatividad e Innovación, Pensamiento Crítico y Trabajo Colaborativo no necesariamente están siendo evaluadas correctamente, ya sea porque las personas no contestan esta pregunta o tienden a responder lo mismo. Por tanto, es posible que se deba realizar una diferenciación, eliminación o revisión de las preguntas que no agregan mucha variabilidad a las respuestas del instrumento.

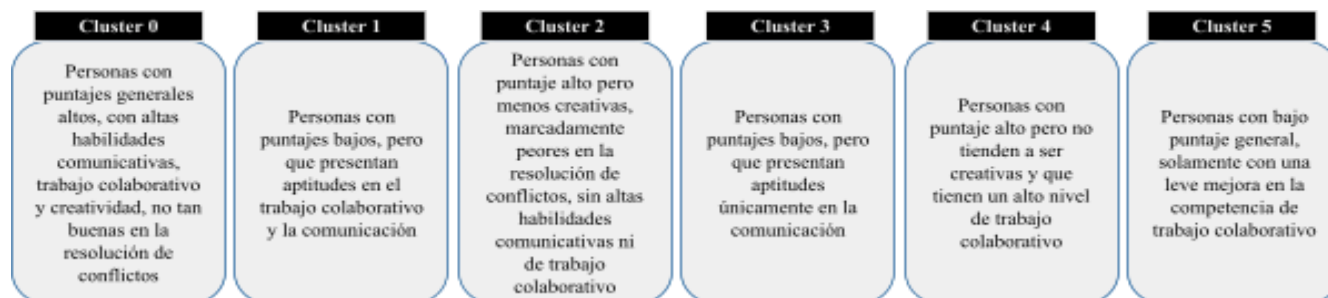
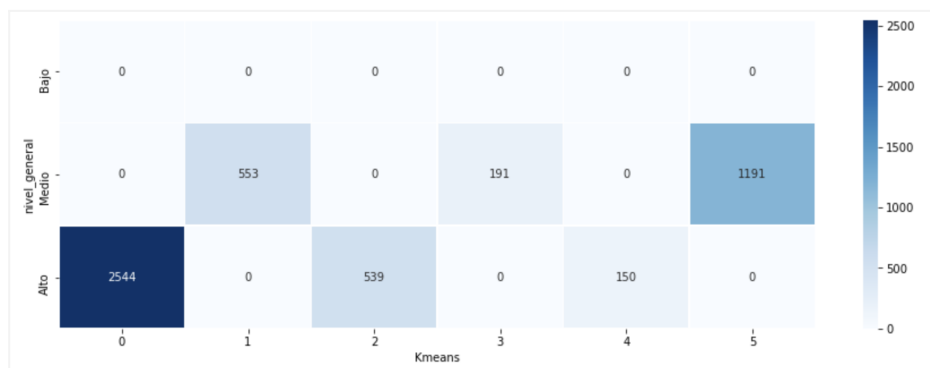


## Generación y evaluación de Clusters

La literatura consultada muestra la utilidad de aplicar un sistema de puntuación o score en el contexto educativo con el propósito de lograr una agrupación efectiva de estudiantes y, de este modo, descubrir patrones relevantes en su rendimiento y comportamiento académico, para obtener este score se utilizó un modelo de clasificación Random Forest, no detallado a continuación ya que el interés primordial radica en la evaluación de las competencias de los docentes en función de sus respuestas a preguntas específicas. Se han empleado dos técnicas de agrupación ampliamente reconocidas: el algoritmo k-means y el algoritmo DBSCAN. Estas técnicas permiten la identificación y clasificación de educadores en grupos principalmente divididos por sus puntuaciones, para luego observar los componentes que más aportan a dicha división, lo que puede resultar de gran utilidad en la toma de decisiones y en el fortalecimiento del desarrollo profesional de los educadores.

### K-Means

La aplicación de este algoritmo de aprendizaje no supervisado sobre el nivel general de la población encuestada, da como resultado la agrupación de los docentes en 6 clusters, definidos utilizando el método de Silueta que da 0.69. Al observar el nivel general parece ser que los 6 están bien definidos entre puntajes medios y altos, sin embargo, las competencias de dichos grupos poblacionales muestran características diferentes, más específicamente se definen clusters como:

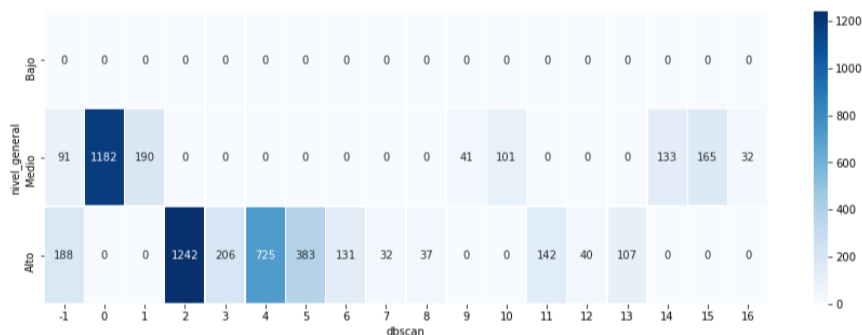


Se observan resultados a creatividad y el pensamiento crítico tienden a tener un comportamiento similar, por lo que se recomienda una reestructuración o unificación de ambas competencias. También las competencias de comunicación y trabajo colaborativo presentan comportamientos similares, estas podrían ser agrupadas en un componente denominado habilidades sociales, o se podrían reestructurar las preguntas de trabajo colaborativo ya que solamente una aporta a la explicación del 80% de la varianza. En todos los clusters se encuentran desafíos en la competencia resolución de conflictos, ya que ningún cluster relaciona un puntaje general alto, con un alto nivel de resolución de conflictos.

## DBSCAN

Para el método DBSCAN se determina que el mejor valor de epsilon es 0.03 a través de revisión manual de valores entre 0.01 y 0.1 y el método sugerido por Rahmah y Sukaesih (2012). Con dicho parámetro se aplica el algoritmo a los datos. Se determina el min\_samples

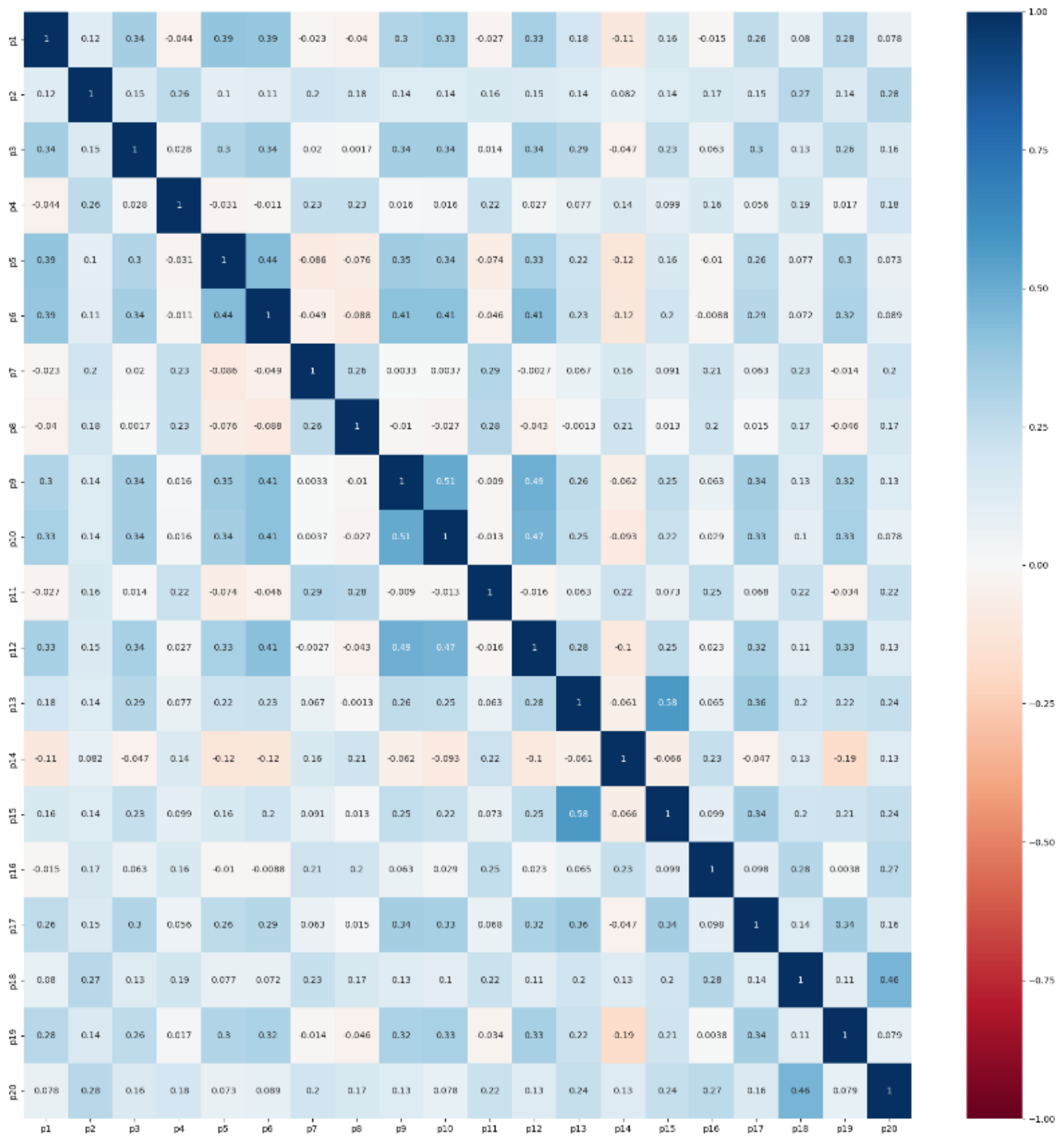
ajustando manualmente y revisando los clusters generados para los valores de 2 a 50, siendo el mejor el min\_samples 30 para la diferenciación de los datos en los clusters. El cluster se valida con el Silhouette score y con observación de los niveles, obteniendo un score de 0.72, siendo min\_samples=30 el equilibrio entre cluster diferenciables en el contexto de los niveles de la competencia y el máximo score de Silhouette. Dentro de los resultados se observan clusters adicionales, pero los más grandes son similares a los encontrados por K-Means (el cluster 0 es similar al cluster 1 de kmeans, o el cluster 4 similar al 0 de Kmeans).



## Conclusiones

- El instrumento es confiable para medir los niveles de competencias generales, pero no los específicos. Dado lo anterior, se sugiere ajustar las preguntas de acuerdo a la unificación propuesta de las competencias, removiendo aquellas que aporten menos a la varianza. Se espera que el instrumento resultante tenga un mejor desempeño para los grupos de competencias específicas, además de ser más corto.
- Se definen 3 nuevos grupos de niveles de competencia específica: 1) Habilidades sociales (Trabajo en equipo + Comunicación) 2) Pensamiento y creatividad y 3) Resolución de conflictos. Esto puede dar a entender que se deba evaluar usar 3 competencias en lugar de 5 dentro del instrumento de medición, reduciendo así el número de preguntas.
- Dado que no se tiene un “real”, es difícil hacer la evaluación de los modelos de forma objetiva. Un mejor score de Silhouette no garantiza un mejor cluster para el caso de uso, e incluso dentro de los temas aprendidos, se hace énfasis en evaluaciones A/B o metodologías similares con un alto componente subjetivo. Se considera que los parámetros encontrados por evaluación manual y buscando la interpretabilidad de los clusters en función al contexto del problema son un buen primer acercamiento al problema.
- Se plantea en el largo plazo y luego de iniciar un proceso de ajustes de acuerdo con los hallazgos del proyecto, realizar un análisis similar al instrumento modificado para evaluar la pertinencia:
  - a. Concluir sobre la separación de los grupos y la viabilidad de a futuro aplicar sistemas de recomendación de contenido
  - b. Concluir sobre si existe información suficiente en la encuesta para dar resultados, o si por el contrario hay preguntas redundantes en la clasificación, esto por medio de la evaluación del instrumento de obtención de datos con técnicas de reducción de dimensionalidad, específicamente aplicar PCA a dicha información.
  - c. Dar conclusiones en relación con las ventajas y desventajas de no dar 5 puntajes sino la pertenencia a un grupo de competencias multidisciplinarias.

## Anexo 1. Gráfico de correlación



## Referencias

- Amat Rodrigo, Joaquín (2022). Clustering con Python. Available under a Attribution 4.0 International (CC BY 4.0) at <https://www.cienciadedatos.net/documentos/py20-clustering-con-python.html>. (Accedido 9 de Enero 2022)
- Bagunaid, W., Chilamkurti, N., & Veeraraghavan, P. (2022). AISAR: Artificial Intelligence-Based Student Assessment and Recommendation System for E-Learning in Big Data. *Sustainability*, 14(17), 10551. doi: 10.3390/su141710551
- Deisenroth, M. P., Faisal, A. A., & Ong, C. S. (2020). *Mathematics for machine learning*. Cambridge University Press.
- Fabila, Angelica & Minami, Hiroe & Izquierdo, Jesús. (2013). La escala de Likert en la evaluación docente: acercamiento a sus características y principios metodológicos. *perspectivas docentes*. 52.
- González-Calatayud, V., Prendes-Espinosa, P., & Roig-Vila, R. (2021). Artificial Intelligence for Student Assessment: A Systematic Review. *Appl. Sci.*, 11(12), 5467. doi: 10.3390/app11125467
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.
- Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction*. 2nd ed. New York: Springer.
- Hastie, T., Tibshirani, R., & Friedman, J. H. (2001). *The elements of statistical learning: Data mining, inference, and prediction*. New York: Springer.
- Jones, Aaron; Kruger, Christopher; Johnston, Benjamin. *The Unsupervised Learning Workshop: Get started with unsupervised learning algorithms and simplify your unorganized data to help make future predictions*. Packt Publishing. Kindle Edition.
- Kaila, E., Kurvinen, E., Lekkila, E., & Laakso, M.-J. (2016). Redesigning an Object-Oriented Programming Course. *ACM Trans. Comput. Educ.*, 16(4), 1–21. doi: 10.1145/2906362
- Kaufman, L. & Rousseeuw, P. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*, Wiley, New York.
- Macnaughton Smith, P., Williams, W., Dale, M. & Mockett, L. (1965). Dissimilarity analysis: a new technique of hierarchical subdivision, *Nature* 202: 1034–1035.
- Minn, S. (2022). AI-assisted knowledge assessment techniques for adaptive learning environments. *Computers and Education: Artificial Intelligence*, 3, 100050. doi: 10.1016/j.caeai.2022.100050
- Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT press.
- Peña, D. (2002). *Análisis de datos multivariantes* (Vol. 24). Madrid: McGraw-hill.
- Pedregosa, F. et al., 2011. Scikit-learn: Machine learning in Python. *Journal of machine learning research*, 12(Oct), pp.2825–2830.
- Shavelson, R.J. *Biographical memoirs: Lee J. Cronbach*. Washington, DC-USA: American Philosophical Society, v. 147, n. 4. p. 379-385, 2009.