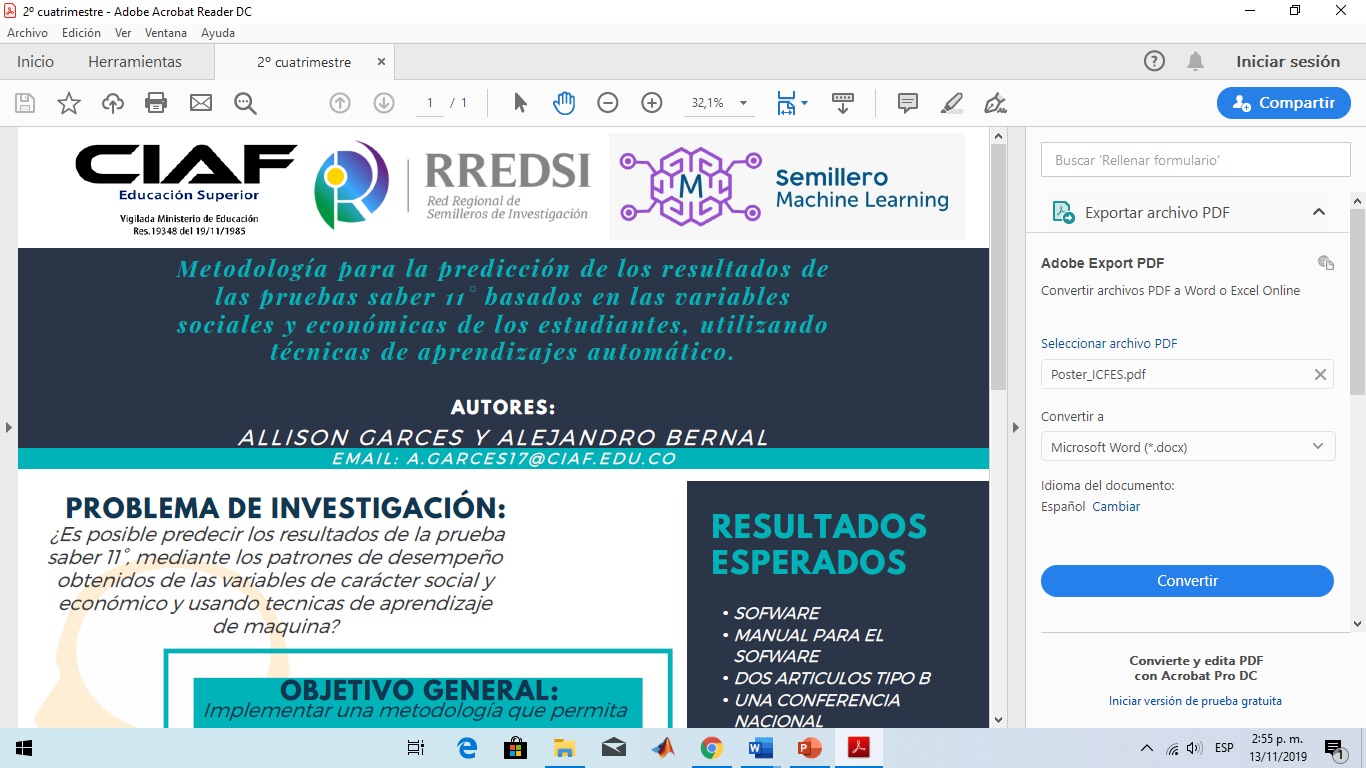
Corporación instituto de administración y finanzas

facultad de ingenieras



metodología para la predicción de los resultados de las pruebas saber 11° basados en las variables sociales y económicas de los estudiantes de Risaralda, utilizando técnicas de aprendizajes automático.

Alejandro Bernal Mora

Allison Garcés Castaño

Pereira Risaralda

**13 de noviembre de 2019**

Tabla de contenido

[1. Resumen 3](#_Toc24547963)

[2. Introducción 4](#_Toc24547964)

[3. Planteamiento del problema 6](#_Toc24547965)

[3.1 Formulación de problema 9](#_Toc24547966)

[4. Objetivo del proyecto 9](#_Toc24547967)

[4.1. Objetivo general 9](#_Toc24547968)

[4.2. Objetivos específicos 9](#_Toc24547969)

[5. Justificación 10](#_Toc24547970)

[7. Metodología del proyecto 11](#_Toc24547971)

[7. Resultados 13](#_Toc24547972)

[8. Conclusiones y recomendaciones 17](#_Toc24547973)

[9. Referencias 18](#_Toc24547974)

# 1. Resumen

La educación es un derecho fundamental que promueve el desarrollo social y económico de una nación, además de fortalecer a los ciudadanos en sus dimensiones personales, sociales y culturales. Es el medio por el cual un país puede alcanzar sus objetivos de desarrollo sustentable, mediante las personas que son capaces de obtener una educación de mejor calidad y así poder romper con el ciclo de la pobreza. La educación, por lo tanto, ayuda a reducir las desigualdades y alcanzar la equidad de género, también permite que las personas tengan una vida más sana y sostenible, e igualmente es crucial para fomentar la tolerancia entre las personas y contribuye a una sociedad más pacífica. A causa de esto, diversos organismos internacionales (Unicef, UNESCO, El Banco Mundial, la OCDE) están promoviendo la cobertura de la educación en los países subdesarrollados [1], mientras que por otro lado proyectos como PISA (Programa para la Evaluación de Estudiantes) cuyo objetivo es realizar evaluaciones anuales a más de 30 países pertenecientes a este ente y clasificarlos según el desempeño en las áreas de lectura, matemáticas y ciencias. Sirviendo como referencia a los países evaluados, e impulsando a las naciones a mejorar la calidad de la educación [2].

# 2. Introducción

En el último siglo, la educación se ha convertido en el referente de los países para determinar el nivel económico, social y cultural de cada nación. Pues el índice de calidad de vida está fuertemente ligado al nivel educativo por país. Según investigaciones, los países que más invirtieron en educación en las 3 últimas décadas crecieron exponencialmente y hoy en día son los países más ricos del mundo y con una mejor calidad de vida que el resto, por ejemplo, USA, CHINA, CANADA, etc. Demostrando que el nivel educativo de cada país es una de las armas fundamentales para el desarrollo de este. Diferentes estudios han demostrado que el aporte de la educación a un país está ligado a cualquier área del mismo, desde lo político, educativo, económico, bélico, comercial, tecnológico, etc. Pues, mediante la inversión educativa es que surgen los estudiantes con un altísimo nivel intelectual incluso salen investigadores encargados desde los avances tecnológicos o militares por ejemplo USA, ALEMANIA, JAPÓN, etc. Por este motivo, Colombia ha ido en busca del mejoramiento educativo de su país, aumentando el presupuesto al ministerio de educación, entre otras formas de elevar el desarrollo educativo de los estudiantes. Como parte de solución a este problema, el ministerio ha estado realizando pruebas con las bases de datos del examen de estado ICFES saber 11°, para conocer cuáles son los factores de desempeño que más afectan el rendimiento de los estudiantes y los colegios. Diferentes investigaciones se han realizado sobre el tema, teniendo resultados muy buenos pero que aún no solucionan todas las problemáticas presentes en el país, pues aún queda un camino muy amplio por recorrer en el aprendizaje de máquinas, el cual podrá dotar al país y a las instituciones de educación media de herramientas que les permitirá tener un mejor control sobre el nivel educativo de sus estudiantes y poder realizar las acciones pertinentes que el caso amerite.

En este orden de ideas, El objetivo principal de este proyecto es desarrollar una metodología que pueda predecir los resultados de las pruebas saber 11° para cada estudiante de la educación media colombiana. Basándose en los patrones de desempeño encontrados de las variables sociodemográficas y socioeconómicas y utilizando técnicas de aprendizaje de máquina. Específicamente, se propone un enfoque basado en aprendizaje supervisado y no supervisado, para evaluar las ventajas y dificultades de cada metodología el cual podrá aportar a las investigaciones que viene realizando el gobierno sobre los factores de desempeño de los estudiantes, pero además contribuirá a este nuevo camino en el cual poco a poco Colombia debería afianzarse el cual es el aprendizaje de máquina, específicamente en softwares que puedan predecir o clasificar estos estudiantes. Para la realización de este proyecto se propuso varias etapas a seguir, como primera medida obtener todas las bases de datos de las pruebas saber 11° desde el 2015 hasta el 2018 y realizar el preprocesamiento pertinente para adecuar las variables y demás valores y características perdidas entre año y año. Además, dejar las bases listas por departamentos, prueba, etc. Y tener los datos sin missing values ni outliers. A continuación, se entrenará los datos con técnicas de aprendizaje no supervisado para determinar y tener claro cuáles son los factores de desempeño que más afectan a los estudiantes, además se hará un informe completo el cual le servirá al país para tener claro cuáles son estos factores. También, se utilizarán técnicas de aprendizaje supervisado para entrenar un modelo que pueda predecir si un estudiante ganará o perderá las pruebas de estado. Como último, se elaborará un software el cual tendrá como objetivo ensamblar todas estas metodologías utilizadas y poder dotar al país de una herramienta estandarizada que pueda predecir los resultados de las pruebas saber 11° y además pueda diagnosticar como tratar cada caso.

# 3. Planteamiento del problema

La educación es un derecho fundamental que promueve el desarrollo social y económico de una nación, además de fortalecer a los ciudadanos en sus dimensiones personales, sociales y culturales. Es el medio por el cual un país puede alcanzar sus objetivos de desarrollo sustentable, mediante las personas que son capaces de obtener una educación de mejor calidad y así poder romper con el ciclo de la pobreza. La educación, por lo tanto, ayuda a reducir las desigualdades y alcanzar la equidad de género, también permite que las personas tengan una vida más sana y sostenible, e igualmente es crucial para fomentar la tolerancia entre las personas y contribuye a una sociedad más pacífica. A causa de esto, diversos organismos internacionales (Unicef, UNESCO, El Banco Mundial, la OCDE) están promoviendo la cobertura de la educación en los países subdesarrollados [1], mientras que por otro lado proyectos como PISA (Programa para la Evaluación de Estudiantes) cuyo objetivo es realizar evaluaciones anuales a más de 30 países pertenecientes a este ente y clasificarlos según el desempeño en las áreas de lectura, matemáticas y ciencias. Sirviendo como referencia a los países evaluados, e impulsando a las naciones a mejorar la calidad de la educación [2].

Colombia, país que no es ajeno a la causa, ha sido partícipe de la evaluación PISA en los últimos años con un historial poco favorable, en el año 2018 su rendimiento no fue satisfactorio debido a que se ubica por debajo del promedio de otros estados pertenecientes a este proyecto, lo cual pone en evidencia el trabajo que aún queda por hacer en materia educativa en comparación a otras naciones [3], a pesar de esto en las últimas décadas Colombia ha sufrido múltiples procesos de transformación que han impulsado entre otros aspectos, el sistema de educación, el acceso a la educación ha sido una prioridad, con políticas ambiciosas que buscan incrementar el número de estudiantes matriculados en todos los niveles y llevar los servicios educativos a todos los rincones del país [4].

Para los colombianos, alcanzar altos niveles de educación es vital para lograr el éxito y prosperidad en diferentes aspectos, desde el aumento de ingresos y productividad hasta el compromiso político y la igualdad social [5]. Sin embargo, las enormes brechas de inequidad que existen en el país perjudican a los estudiantes de grupos o regiones con carencias socioeconómicas, impidiendo que el estudiante desarrolle sus actividades académicas con normalidad, aumentando así la deserción [4]. Debido a esto, el país busca evolucionar en este aspecto y lograr estar conformado por ciudadanos competentes, formados académicamente, que puedan satisfacer los requisitos impuestos por la oferta laboral [6].

Actualmente el ingreso a la educación superior en Colombia está determinado en gran medida por el desempeño en las pruebas saber 11, prueba que evalúa las áreas fundamentales del conocimiento como matemáticas, lectura crítica, ciencias naturales, lengua inglesa y ciencias sociales y ciudadanas, además de recoger información de tipo sociodemográfico y económica de los estudiantes de educación media [7]. Estas pruebas no solo sirven como criterio de ingreso para la educación superior, adicional a esto también aportan al desarrollo de investigaciones de carácter cultural, social y educativo [5].

Investigaciones previas indican que las variables socioeconómicas y sociodemográficas de los estudiantes son un factor de gran importancia en el contexto de la educación y están directamente relacionadas con el desempeño académico [8,9,10,11], variables como la ocupación o el nivel educativo de los padres [12], al igual que el estrato socioeconómico o la posesión de algún capital suelen estar ligadas al rendimiento del estudiante [13]. Por tal motivo, estas variables sociales permiten un mejor acercamiento al estudiante en el ámbito educativo, permitiendo así implementar alguna medida que puedan brindar una solución a los jóvenes que representaran futuros problemas con las pruebas de estado, permitiéndole al estado una mejora en el nivel educativo de la nación.

Dado que el examen de estado se convierte en un indicador de las capacidades intelectuales de los evaluados y que este es un requisito fundamental para la continuidad de la formación académica de los ciudadanos, se vuelve determinante identificar las variables tipo sociodemográfico que influyen de forma directa o indirecta sobre los resultados finales de las pruebas saber 11, ya que a la larga se convertirá en un factor que interviene tanto en la calidad de vida de los colombianos, como en el desarrollo del país.

Es por esto que se han realizado diferentes investigaciones en relación a la problemática antes mencionada. Como por ejemplo el ministerio de educación, que haciendo uso del banco de datos de las pruebas saber 11°, realizan análisis estadísticos para determinar relaciones entre variables sociodemográficas y los resultados de las pruebas [14], sin embargo, los análisis son muy superficiales y poco rigurosos ya que se limitan al uso de la estadística descriptiva y no permiten encontrar correlaciones de variables subyacentes que se puedan expresar en términos de las originales.

Por otro lado, la Universidad Distrital Francisco José de Caldas se propuso hallar en su trabajo los determinantes individuales de desempeño en las pruebas de estado, y para ello realizó un análisis comparativo de dos técnicas de regresión, modelo de regresión multinivel (HLM) y aprendizaje computacional basados en Kernel (SVM), en la cual hubo una notable superioridad por parte del modelo HLM tanto en su expresividad como en el tiempo de procesamiento. Sin embargo, los métodos utilizados no permitieron describir las relaciones individuales generadas y los resultados obtenidos no eran los esperados [15].

La minería de datos no es un tema nuevo dentro del campo educativo, y sus técnicas se pueden utilizar para explicar y predecir fenómenos dentro del sistema educativo [16,17], un ejemplo de esto es el uso de técnicas de aprendizaje automático para predecir la probabilidad de deserción con un alto grado de confiabilidad [18]. Por otro lado, las instituciones educativas hacen uso de estas técnicas para realizar caracterizaciones de los perfiles académicos de sus estudiantes [19], debido a esto se hace conveniente el uso de las herramientas computacionales que proporciona la minería de datos.

De acuerdo a lo planteado anteriormente es evidente que existe un problema metodológico en relación a la determinación de las variables no académicas que influyen en los resultados de las pruebas saber 11, y siendo consecuente con el auge actual de las técnicas de aprendizaje automático en el campo educativo, se hace pertinente plantear la siguiente pregunta de investigación ¿Es posible predecir los resultados de las pruebas saber 11, mediante los patrones de desempeño obtenidos de las variables de carácter social y económico y usando técnicas de aprendizaje de máquina?.

## 3.1 Formulación de problema

¿Es posible predecir los resultados de las pruebas saber 11, mediante los patrones de desempeño obtenidos de las variables de carácter social y económico y usando técnicas de aprendizaje de máquina?

# 4. Objetivo del proyecto

## 4.1. Objetivo general

Implementar una metodología que permita la incorporación de un sistema predictivo de las pruebas saber 11° en las instituciones de educación media en el país de Colombia, a través de la identificación de perfiles de los estudiantes con técnicas de aprendizaje de máquina.

## 4.2. Objetivos específicos

1. Desarrollar un protocolo de preprocesamiento y reorganización a la base de datos suministrada por el ministerio de educación superior, pruebas saber 11°.
2. Diseñar una metodología basada en técnicas de aprendizaje de máquina supervisado y no supervisado para la identificación de perfiles de los estudiantes de acuerdo con el nivel educativo de Colombia. La metodología se validará con métricas de rendimiento aplicadas en las máquinas de aprendizaje y análisis de independencia estadística.
3. Desarrollar un aplicativo software que prediga que estudiantes aprobarán o perderán el examen de estado saber 11°.

# 5. Justificación

El mundo de hoy en día evoluciona a pasos agigantados en materia de industria y cada vez se hace más recurrente que las empresas o el sector laborar en general sean exigentes con la formación educativa de sus empleados a causa de la competencia que existe en la industria [20], las capacidades, habilidades y destrezas de un ciudadano dependen ampliamente del nivel de formación académico recibido en el transcurso de su vida, y como se mencionaba anteriormente, uno de los requisitos más importantes para poder acceder a la educación superior en Colombia es obtener un desempeño aceptable en las pruebas de estado.

Por otro lado, la identificación de las variables de tipo económico y social que intervienen en el rendimiento de las pruebas de estado se ha vuelto un tema de estudio en los últimos años [8,9,10,11], investigaciones empeñadas en buscar las relaciones que existen entre este tipo de características y los resultados finales de la prueba, contribuyen al sistema educativo a fomentar metodologías más eficientes a la hora de buscar mejorar la calidad de la educación en el país [7].

En consecuencia, la implementación de una metodología que permita la incorporación de un sistema predictivo de las pruebas saber 11 se convertiría en una herramienta de auxilio para las instituciones educativas, las cuales podrían evaluar la probabilidad de éxito de sus alumnos y determinar si estos necesitan un apoyo de índole económico o social, según la deficiencia que el sistema indique.

# 7. Metodología del proyecto

Es importante y pertinente aclarar que la información de las encuestas que se tomaron del ministerio de educación que se utilizaran durante la ejecución de este proyecto, no será divulgada y será de uso confidencial. De acuerdo con los objetivos específicos establecidos, se proponen los siguientes pasos metodológicos:

**Actividades para lograr el objetivo específico 1: “**Desarrollar un protocolo de preprocesamiento y reorganización a la base de datos suministrada por el ministerio de educación superior, pruebas saber 11°”.

**Etapa 1 – Revisión del estado del arte:** Se realizará una cuidadosa y refinada revisión del estado del arte relacionado a la educación colombiana, así mismo a lo referente al aprendizaje de máquina aplicado al reconocimiento de perfiles de los estudiantes partir de información de la encuesta obtenida por el ministerio de educación.

**Etapa 2 – Organización y preprocesamiento de la base de datos:** Se organizará una base de datos etiquetada a partir de las encuestas realizadas a los estudiantes colombianos sobre las pruebas saber 11°. Los estudiantes se dividen en los que ganaron y los que perdieron el examen de estado. Luego se hará un respectivo preprocesamiento para eliminar missing values entre otros inconvenientes y los resultados obtenidos se utilizarán para el entrenamiento de los algoritmos de reconocimiento de patrones.

**Actividades para lograr el objetivo específico 2:** “Diseñar una metodología basada en técnicas de aprendizaje de máquina supervisado y no supervisado para la identificación de perfiles de los estudiantes de acuerdo al nivel educativo de Colombia. La metodología se validará con métricas de rendimiento aplicadas en las máquinas de aprendizaje y análisis de independencia estadística.”.

**Etapa 3 – Aprendizaje supervisado:** Se desarrollará un módulo para el reconocimiento de perfiles de los estudiantes basado en aprendizaje supervisado. Se requiere que la base de datos esté etiquetada (rotulada) por un experto. Se aplicarán algoritmos básicos como el clasificador Bayesiano (lineal y cuadrático), k-nearest neighbor y otros más avanzados como las máquinas de soporte vectorial y procesos Gaussianos.

**Etapa 4 – Aprendizaje no supervisado:** Se desarrollará un módulo para el reconocimiento de perfiles de los estudiantes basado en aprendizaje no supervisado. Específicamente se aplicarán métodos validados en el estado del arte como el algoritmo k-means y mezclas de Gaussianas con inferencia Bayesiana: Expectation-Maximization e inferencia variacional.

**Etapa 5 - Validación:** Los métodos propuestos de aprendizaje supervisado y no supervisado se validarán empleando métricas de rendimiento como la precisión, la dispersión, curvas ROC, especificidad y sensibilidad sobre los resultados de identificación de los estudiantes.

**Etapa 6 - Redacción de informe final:** Se redactará un informe final

**Actividades para lograr el objetivo específico 3:** “Desarrollar un aplicativo software que prediga que estudiantes aprobarán o perderán el examen de estado saber 11°”.

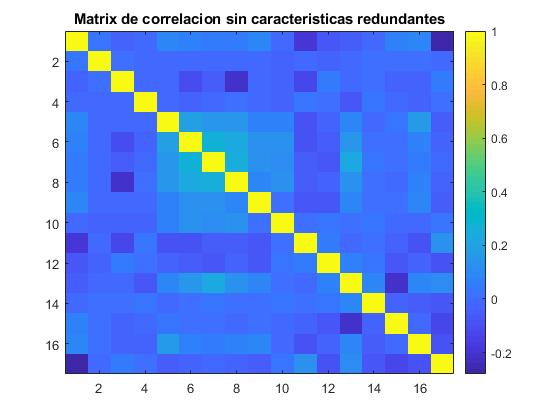
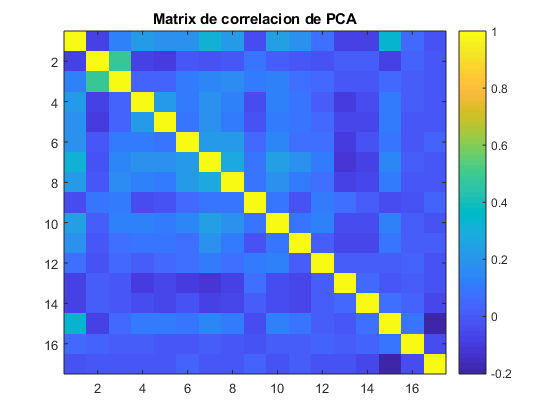
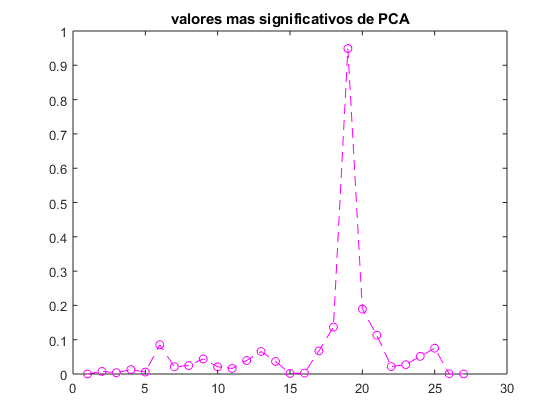
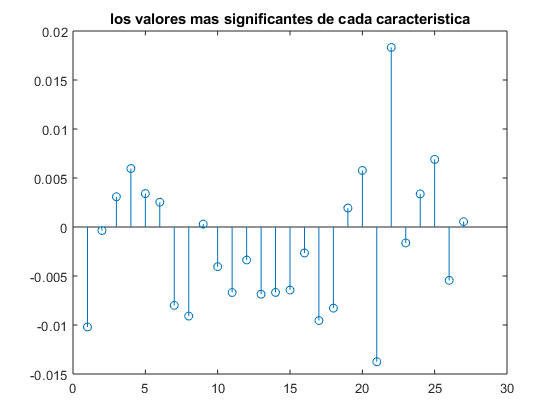
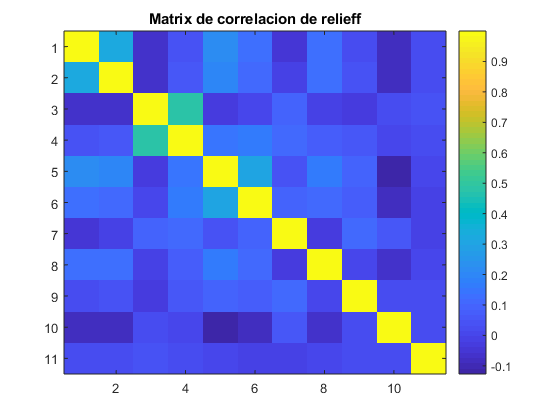
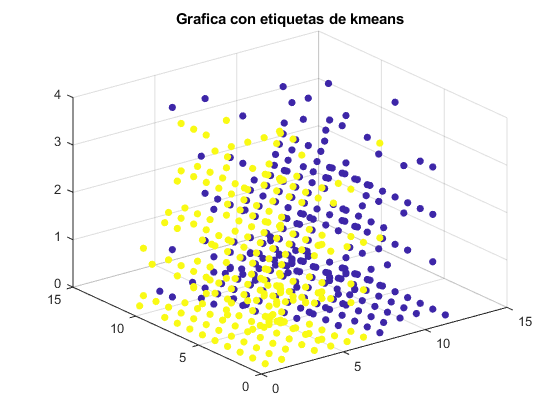
**Etapa 7 - Desarrollo de aplicativo software y artículo de investigación:** Se desarrollará un aplicativo en Matlab o Python que articule los módulos de reconocimiento de patrones. El software será utilizado y validado por el ente regulador encargado. Además, se presentará un artículo a un congreso internacional.

# 7. Resultados

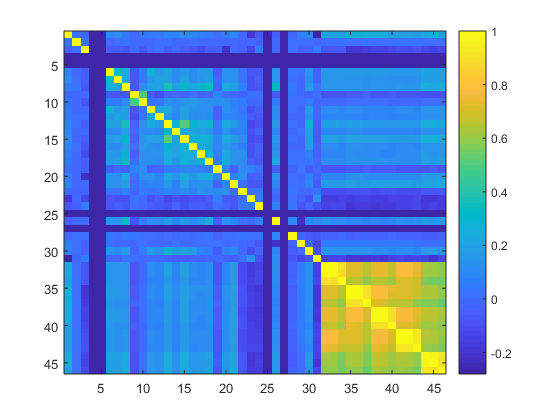
El proyecto se encuentra en proceso, con resultados parciales. De los cuales se destaca el artículo de investigación que se esta elaborando para presentarlo en la revista scientia et technica. En este orden de ideas, se hizo un preprocesamiento de las bases de datos del ICFES 2018 calendario B para determinar qué características sociales o sociodemográficas afectaban el resultado de los estudiantes. Como primera medida, se eliminaron todos los outliers y missing values que tenía la base de datos. Posteriormente, se utilizaron métodos de aprendizaje no supervisado como CORRELACIÓN, PCA y RELIEF. Donde RELIEF obtuvo resultados óptimos en cuanto a la matriz de correlación. A continuación, se utilizó una técnica llamada C-S, la cual tiene como objetivo mapear de un espacio discreto a uno continuo mediante la medida de disimilitud Chi-Square. la finalidad de esta es aumentar la dimensionalidad de la base de datos para permitir una adecuada separación de los datos donde se comparó con los métodos anteriormente mencionados utilizando k-means como referencia principal. Se compararon los algoritmos para ver cual aumentaba el acierto de las etiquetas vs las etiquetas reales utilizando la matriz de confusión para obtener los resultados. Como resultado, se obtuvo que con la técnica C-S el acierto fue mucho mayor que las demás técnicas. además, la separabilidad visual era optima pues se podía casi separar los grupos de una forma lineal.

Con los datos obtenidos, se pudo saber cuáles fueron las características que más afectaron en los resultados de los estudiantes de Risaralda. Además, con el estudio se puede ver qué cosas debería mejorar el departamento para que sus estudiantes puedan progresar en las pruebas ICFES saber 11°.

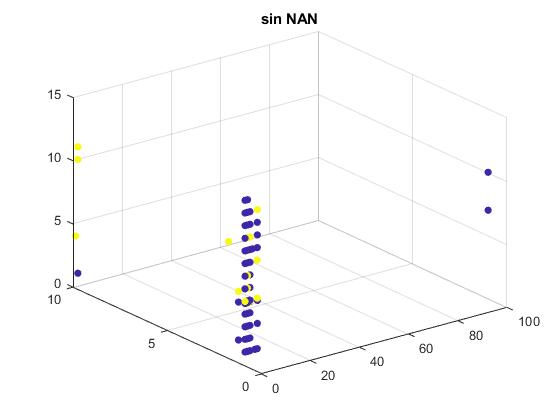
A futuro, con estos datos se espera utilizar técnicas de aprendizaje supervisado para poder predecir los estudiantes que ganaran o perderán las pruebas de estado y posteriormente embeberlo en un software que le permita al gobierno identificar estos factores y darle una solución adecuada.

Imagen que contiene reloj

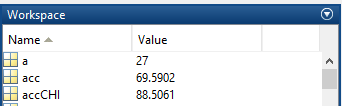
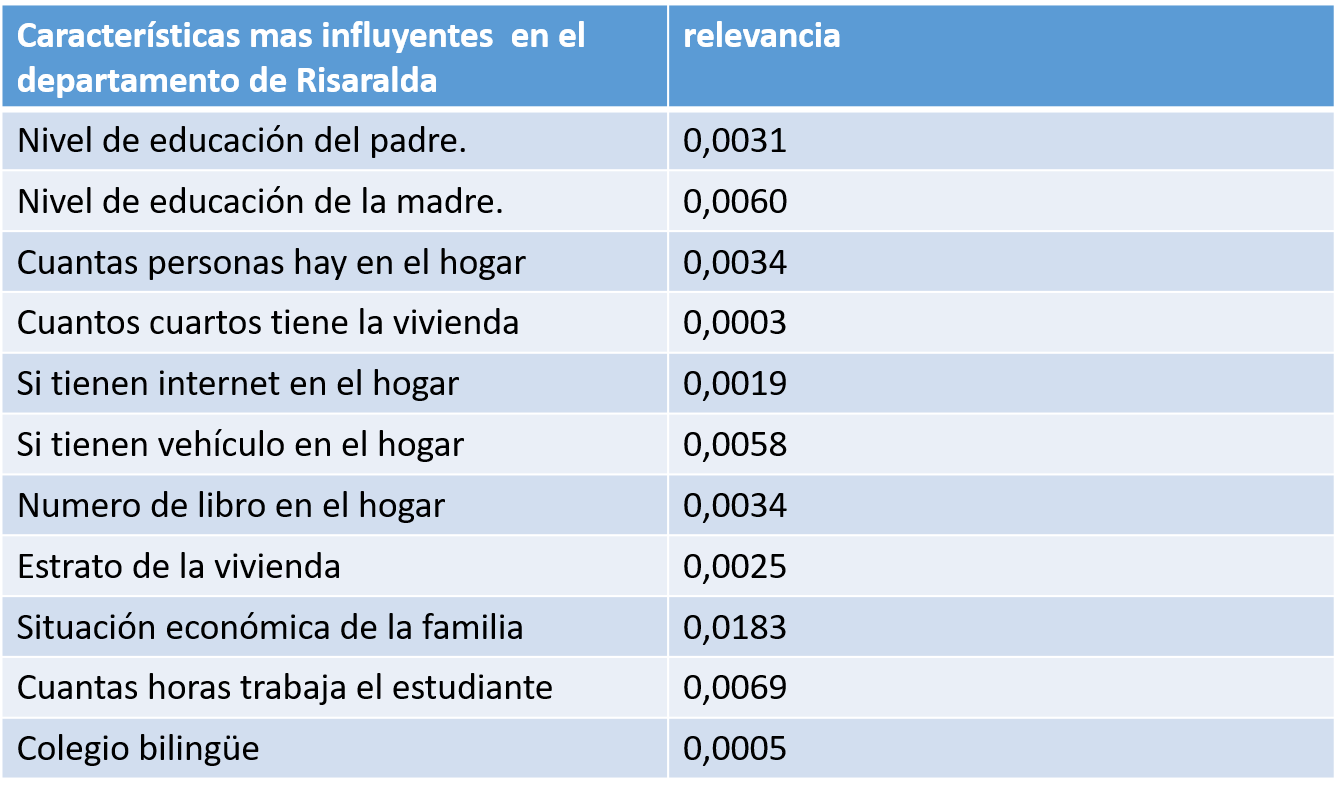
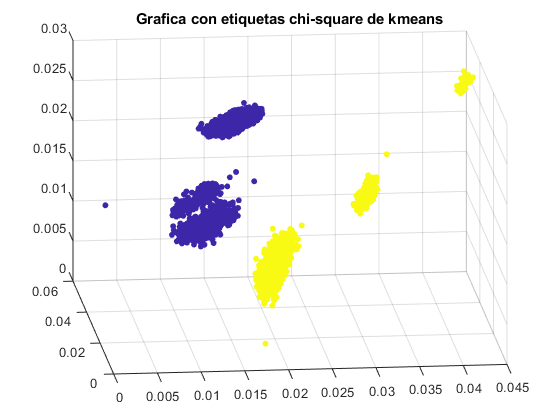
Descripción generada automáticamenteImagen que contiene luz

Descripción generada automáticamenteImagen que contiene texto, mapa, llenado, tabla

Descripción generada automáticamente

Imagen que contiene texto, mapa

Descripción generada automáticamente



# 8. Conclusiones y recomendaciones

Como conclusiones parciales del proyecto, se destaca la eficiencia de la distancia Chi-square para datos de tipo categórico, pues demuestra una gran eficiencia a la hora de agrupamiento de datos, clasificación, etc. Pues esta medida de disimilitud permite mapear datos de un espacio cualitativo a uno cuantitativo permitiendo en general una separabilidad visual optima.

Como parte del desarrollo del proyecto se evidencia una gran falencia en el departamento de la guajira en las pruebas saber 11°, es de resaltar que es el departamento con peores resultados y las características más influyentes son del factor socioeconómico y demográfico como se esperaba. los resultados del departamento de Risaralda aún están en procesamiento, pero se puede evidenciar que tiene carencias con respeto a Antioquia, Bogotá, etc.

A futuro, con estos datos se espera utilizar técnicas de aprendizaje supervisado para poder predecir los estudiantes que ganaran o perderán las pruebas de estado y posteriormente embeberlo en un software que le permita al gobierno identificar estos factores y darle una solución adecuada.

# 9. Referencias

[1] Acevedo, S., & Jaramillo, A. (2007). Perfil socioeconómico de los estudiantes de pregrado EAFIT. Trabajo de Grado). Universidad EAFIT, Medellín, Antioquia, Colombia.

[2] Armento, N. G., Pacheco, C., & Pineda, E. D. (2008). Factores socioeconómicos que intervienen en el desempeño académico de los estudiantes de la Facultad de Ciencias Humanas de la Universidad Autónoma de Baja California.

[3] Ayala-García, J. (2015). Evaluación externa y calidad de la educación en Colombia. Documentos de Trabajo Sobre Economía Regional y Urbana; No. 217.

[4] Azoumana, K. (2013). Análisis de la deserción estudiantil en la Universidad Simón Bolívar, facultad Ingeniería de Sistemas, con técnicas de minería de datos. Revista Pensamiento Americano, 6(10).

[5] Caso-Niebla, J. & Hernández, L. (2007), `Variables que inciden en el rendimiento académico de adolescentes mexicanos', Revista Latinoamericana de Psicóloga 39(3), 487-501.

[6] Castro, A., & Victoria, M. (2010). Financiamiento de la educación superior en Colombia reflexiones para un próximo futuro. Revista de la educación superior, 39(156), 89-102.

[7] Charrys Llanos, A., Fuentes Gomez, M., Medina Rodriguez, J., & Prieto Diaz, N. (2016). Factores institucionales asociados al logro educativo de las pruebas saber 11 de ciencias naturales en colegios de la región caribe (Magister). Universidad del norte.

[8] Contreras, K., Caballero, C., Palacio, J. & Pérez, A. (2008), ‘Factores asociados al fracaso académico en estudiantes universitarios de Barranquilla (Colombia)’, psicóloga desde el Caribe 22, 110-135.

[9] Departamento nacional de planeación. (s.f.). La Educación en Colombia. Recuperado 25 abril, 2019, de https://www.dnp.gov.co/programas/desarrollo-social/subdireccion-de-educacion/Paginas/subdireccion-de-educacion.aspx

[10] Gakure, R. W., Mukuria, P., & Kithae, P. P. (2013). An evaluation of factors that affect performance of primary schools in Kenya: A case study of Gatanga district. Educational Research and Reviews, 8(13), 927-937.

[11] Hernández, E. & González, M. (2011), ‘Modelo de ecuación estructural que evalúa las relaciones entre el estatus cultural y económico del estudiante y el logro educativo’, Revista Electrónica de Investigación Educativa 13(2), 188-203.

[12] Instituto Colombiano para la Evaluación de la Educación Icfes (2018). Documentación del examen Saber 11. Ministerio de educación.

[13] Kotsiantis, S. B., Pierrakeas, C. J., & Pintelas, P. E. (2003, September). Preventing student dropout in distance learning using machine learning techniques. In International conference on knowledge-based and intelligent information and engineering systems (pp. 267-274). Springer, Berlin, Heidelberg.

[14] Luna, E. L., & Maldonado, M. Y. C. (2006). Competencias laborales del trabajador social vistas desde el mercado laboral. Tabula Rasa: revista de humanidades, 5, 261-296.

[15] Lykourentzou, I., Giannoukos, I., Nikolopoulos, V., Mpardis, G., & Loumos, V. (2009). Dropout prediction in e-learning courses through the combination of machine learning techniques. Computers & Education, 53(3), 950-965.

[16] McKenzie, K., & Schweitzer, R. (2001). Who succeeds at university? Factors predicting academic performance in first year Australian university students. Higher education research & development, 20(1), 21-33.

[17] Maldonado, A. (2000). Los organismos internacionales y la educación en México: el caso de la educación superior y el Banco Mundial. Perfiles educativos, 22(87), 51-75.

[18] OECD (2016), Education in Colombia, Reviews of National Policies for Education, OECD Publishing, Paris, https://doi.org/10.1787/9789264250604-en.

[19] Orea, S. V., Vargas, A. S., & Alonso, M. G. (2005). Minería de datos: predicción de la deserción escolar mediante el algoritmo de árboles de decisión y el algoritmo de los k vecinos más cercanos. Ene, 779(73), 33.

[20] Orjuela, J. (2012). Determinantes individuales de desempeño en las Pruebas de Estado para la educación media en Colombia. In ICFES, Estudios sobre la calidad de la educación en Colombia. Bogotá: Ministerio de Educación Nacional, 164-176.

[21] Pereira, R. T., Romero, A. C., & Toledo, J. J. (2013). Aplicación de la minería de datos en la extracción de perfiles de deserción estudiantil [Application of data mining in extracting student dropout profiles]. Ventana Informática, (28).

[22] Pereira, R. T. (2013, August). La Minería de Datos como un método innovador para la detección de patrones de deserción estudiantil en programas de pregrado en Instituciones de Educación Superior. In WEEF 2013 Cartagena.

[23] Resultados de investigaciones - Icfes Instituto Colombiano para la Evaluación de la Educación. (2018),http://www2.icfes.gov.co/investigadores-y-estudiantes-posgrado/resultados-de-investigaciones

[24] Rovai, A. P., Gallien Jr, L. B., & Wighting, M. J. (2005). Cultural and interpersonal factors affecting African American academic performance in higher education: A review and synthesis of the research literature. The Journal of Negro Education, 359-370.

[25] Turner, R. (2006). El Programa Internacional para la Evaluación de los Alumnos (PISA). Una perspectiva general. Revista de educación, 2006, 45-74.

[26] Un.org. (n.d.). [online] Available at: https://www.un.org/sustainabledevelopment/wp-content/uploads/2017/02/ENGLISH\_Why\_it\_Matters\_Goal\_4\_QualityEducation.pdf [Accessed 24 Apr. 2019].

[27] Vargas, G. M. G. (2013), ‘Factores asociados al rendimiento académico en estudiantes universitarios desde el nivel socioeconómico: Un estudio en la Universidad de Costa Rica’, Revista Electrónica Educare 17(3), 57-87.

[28] Vermunt, J. D. (2005). Relations between student learning patterns and personal and contextual factors and academic performance. Higher education, 49(3), 205.

[29] Zhu, X. (2015, March). Machine teaching: An inverse problem to machine learning and an approach toward optimal education. In Twenty-Ninth AAAI Conference on Artificial Intelligence.

4512689