**Título del trabajo**Avance de proyecto de grado  
Línea de investigación del grupo FICB-PG: Investigación aplicada  
Grupo de Investigación FICB-PG **Nombre apellidos 1 (código), Nombre apellidos 2 (código),**Asesor temático: Isabel Mahecha

*Junio 2019*

**\_\_\_\_\_\_\_\_\_\_\_\_**

**Resumen**

Este trabajo presenta la aplicación de la minería de datos a la educación, basada en un conjunto de datos de los resultados de las pruebas Saber 3,5,9 correspondiente a los años 2017, 2016 y 2015 en Colombia. En la investigación se usa técnicas de clasificación para identificar las características más influyentes en los resultados del examen. El documento presenta la descripción de los datos, la metodología usada, los algoritmos de agrupamientos aplicados, los diferentes experimentos realizados y los resultados.

**Palabras clave:** Pruebas Saber, minería de datos, metodología CRISP-DM

**Abstract**

This paper presents the application of data mining to education, based on a set of data from the Saber 3,5,9 exam results corresponding to the years 2017, 2016 and 2015 in Colombia. In the research, classification techniques that use to identify the most influential characteristics in the results of the examination. The document presents the description of the data, the methodology used, the algorithms classifications applied, the different experiments carried out and the results.

**Key words:** Exam Saber, data mining, CRISP-DM methodology

**\_\_\_\_\_\_\_\_\_\_\_\_**

**INTRODUCCIÓN**

El Instituto Colombiano para la Evaluación de la Educación (ICFES) es una entidad autónoma adscrita al Ministerio de Educación Nacional de Colombia, que se encarga de la evaluación de la educación en todos los niveles y realiza investigaciones sobre los factores que inciden en la calidad educativa.

Desde el 2009, El ICFES realiza la prueba Saber 3,5,9 anualmente, es estandarizada, con el fin de evaluar la educación primaria y secundaria, esta prueba evalúa a los estudiantes que están cursando grados tercero y quinto sobre dos áreas en común lenguaje y matemáticas y dos opcionales que se intercalan cada año que son ciencias naturales y competencias ciudadanas.

La prueba Saber 3,5,9 se ha constituido como un pilar para medir el progreso académico, también para diseñar planes de mejoramiento en los colegios y las secretarias de educación, desde el 2009 hasta el 2017 se han recolectado los resultados de dichas pruebas, para dicho fin.

La minería de datos educativa (EDM) es un campo de la minería de datos especializada en la educación, que se centra en el desarrollo de métodos para la exploración de datos de un entorno educativo.

Este trabajo buscar aplicar técnicas de minería de datos sobre los datos de las prueba Saber 3,5,9 del 2015 al 2017 en Colombia para proponer un modelo que permita encontrar relaciones entre diferentes factores que impactan el desempeño académico de los estudiantes que presentan el examen.

**REVISIÓN DE LITERATURA**

Este apartado debe contener la literatura relevante y adecuada que expresa las posturas teóricas respecto al problema planteado en el apartado anterior. En coherencia con lo anterior, debe contener las variables que se pretenden a abordar en el marco del proyecto de investigación de maestría, dando cuenta de su relación y relevancia en el contexto o ámbito en el que se desarrollaría la investigación, se sugiere que su extensión no sea mayor a 6.000 palabras y que lleve títulos y subtítulos que lleven y orienten al lector de lo general a lo particular. De igual forma, la literatura que se referencie debe estar en una ventana de observación de máximo 10 años con respecto a la fecha de escritura del documento.

**ESTRATEGIA METODOLÓGICA**

Este apartado debe contener la manera en la cual se abordará el problema en coherencia con su naturaleza, se sugiere que su extensión no sea mayor a 2000 palabras. Se debe exponer, sustentar y justificar

**DESARROLLO E IMPLEMENTACIÓN (AVANCES)**

En esta sección se presentan y explican, a modo de avances, las herramientas desarrolladas y particularidades de la implementación según los pasos presentados en la metodología. Si aplica, se presenta el escenario particular de aplicación de la metodología, y la forma en que dicho escenario o contexto modifica o afecta la forma de implementación. Puede así presentarse en esta sección el diseño de experimentos y el detalle del conjunto de datos y variables de entrada.

Se presentan también aspectos técnicos que sean relevantes en el proyecto y que hayan de alguna manera afectado (de manera positiva o negativa) el procesamiento o la obtención de los resultados.

**RESULTADOS PRELIMINARES**

Este apartado debe contener los hallazgos y resultados preliminares de haber implementado los pasos de la metodología, de preferencia en el orden del planteamiento que llevan los objetivos específicos del trabajo. En este apartado no se hacen conclusiones o sugerencias, sino se presenta la información tal cual fue obtenida de las fuentes de las que se dio cuenta en el apartado de ESTRATEGIA METODOLÓGICA..

**REFERENCIAS**

Este apartado debe dar cuenta de TODOS los textos a los que efectivamente se hizo referencia en el marco del texto, no se deben incluir textos a lo que no se haya hecho alusión. Se sugiere usar el editor automático de *Word* o *Mendeley*. El formato de las referencias para esta entrega es el más reciente indicado por la Asociación Americana de Psicología (APA, por sus siglas en inglés).

Colciencias. (2014). Modelo de Medición de Grupos, de Investigacion, Desarrollo Tecnológico o de Innovación y reconocimiento de investigadores del Sistema Nacional de Ciencia, tecnología e Innovación 2014. Retrieved from http://www.colciencias.gov.co/sites/default/files/ckeditor\_files/files/DOCUMENTO MEDICI%C3%93N GRUPOS - INVESTIGADORES VERSI%C3%93N FINAL 15 10 2014 (1).pdf

Pérez-Rave, J., Álvarez-Jaramillo, G., & Henao-Velásquez, J. (2012). ¿ Cómo identificar la literatura “poco - vital” y cuantificar su representatividad ? Aplicación al problema del vendedor viajero < TSP >. *Actualidad Y Nuevas Tendencias*, (8), 51–60.

**ADVERTENCIA**

Toda idea, texto (frases o palabras) o cifra de cualquier fuente institucional, persona natural, persona jurídica u otros, debe tener la adecuada referencia y por tanto acreditación de manera explícita, para el caso del cuerpo del texto y el listado de bibliografía, inclusive si es una traducción propia. Así mismo, se deben usar adecuadamente las citas directas e indirectas de acuerdo con las ideas que se quieren expresar en el marco de las entregas para los trabajos o avances del programa. Se sugiere que antes de hacer cualquier envío del documento se utilicen plataformas libres para verificar el adecuado crédito, referencia y uso de los aportes propios y de terceros. Para mayor información consulte:

i. <http://www.apa.org/>

ii. <http://normasapa.com/>

iii. <http://repository.poligran.edu.co/handle/10823/328>

iv. <http://www.plagium.com/>

v. <http://www.plagscan.com/es/>

vi. <http://plagiarisma.net/es/>

vii. <http://www.plagtracker.com/es/>