**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:** Examen Saber, data mining, CRISP-DM methodology

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

**INTRODUCCIÓN**

Este apartado debe contener los antecedentes, justificación, sistematización, alcance y planteamiento del problema a abordar en el marco del proyecto de investigación de maestría y debe cerrar o con el objetivo central del trabajo o con la pregunta de investigación que planteare el estudiante y el planteamiento de los objetivos específicos derivados de ello, se sugiere que su extensión no sea mayor a 1800 palabras. Hay que cuidar que el tema de estudio en lo posible refleje las siguientes características:

1. **Ser novedoso**: se debe revisar que tenga un aporte teórico, un aporte metodológico o que aborde un ámbito distinto a trabajos anteriores, con respecto a la revisión de literatura efectuada, de forma que los resultados del proyecto sean valiosos y atractivos para los diferentes escenarios de divulgación (revistas, ponencias, congresos, etc.)
2. **Orientación**: estar orientado a que sus resultados sirvan a la solución de problemáticas concretas y actuales en las áreas de conocimiento propias de la Maestría y del grupo de investigación de la Facultad de Ingeniería y Ciencias Básicas.
3. **Pertinencia**: evidenciar el impacto que tendrá la ejecución del proyecto en un contexto determinado.
4. **Factible**: prever que el proceso de investigación sea posiblemente ejecutado en términos de tiempo (un año) y recursos.
5. **Evidenciable**: prever y dimensionar la forma en que se lograría evidenciar el logro (entregables) de cada uno de los objetivos planteados.

Para el planteamiento y formulación del problema se sugiere que tenga en cuenta:

1. El problema expresa la relación entre dos o más variables
2. El problema y los objetivos relacionados dejan claro el contexto en el cual se desarrolla el proyecto.
3. El problema se debe formular con claridad y evitando ambigüedades.
4. El problema debe ser susceptible de validación o comprobación empírica
5. Los objetivos y el problema deben ser coherentes.

**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.

**Sobre el uso de tablas y figuras:**

Las tablas y figuras deben estar enumeradas consecutivamente y deben ser explicadas y referenciadas en el marco del texto, se sugiere usar el editor automático de *Word*; asimismo, al final de cada una de ellas se debe dar cuenta de su fuente. Ejemplo:

Tabla 1. Cantidad de Investigadores de Colombia según la evaluación de desempeño de actividades científicas de 2013.

|  |  |
| --- | --- |
| Categoría | Cantidad de Investigadores |
| Senior | 693 |
| Asociado | 1823 |
| Junior | 5495 |
| Total | **8011** |

Fuente: Elaboración Propia a partir de Colciencias (2014)

**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:

1. El enfoque metodológico
2. El diseño muestral o de abordaje de participantes.
3. La descripción del universo, población, muestra o participantes, según aplique, dando cuenta de los criterios o mecanismos de inclusión o exclusión y los aspectos éticos para abordarles y para el manejo de su información.
4. La descripción de las variables y las relaciones a verificar en ellas, y si aplica planteamiento de hipótesis.
5. Los instrumentos teniendo en cuenta los ejes de indagación, categorías orientadoras, relación y operacionalización de variables e indicadores, según aplique.
6. Los mecanismos para dar validez y rigor a los instrumentos, trabajos de campo y en general la ejecución metodológica.
7. Las maneras y aplicativos para procesar la información detallando el procedimiento para tal fin.

**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/>