

Un algoritmo de clustering difuso para el análisis de las causas que afectan el desarrollo, productividad y competitividad de la industria de calzado, cuero y marroquinería con relación a su producción y mano de obra

Karol Juliana Benavidez Robles y Liz Bleidy Tatiana Jiménez Wandurraga

Trabajo de grado para optar por el título de Ingeniero Industrial

Director

Henry Lamos Díaz

Ph. D en Física - Matemática

Codirector

David Esteban Puentes Garzón

Msc. en Ingeniería Industria

Universidad Industrial de Santander

Facultad de Ingenierías Fisicomecánicas

Escuela de Estudios Industriales y Empresariales

Bucaramanga

2022

Agradecimientos

A Dios, gracias por permitirnos culminar una de las etapas más importantes de nuestras vidas, por estar presente en cada paso, regalarnos fortaleza y sabiduría para seguir adelante.

Al profesor Henry Lamos por todos su conocimiento y acompañamiento, a todos los profesores que hicieron que este momento fuese realidad, en especial nos regocijamos de gratitud con el profesor David Puentes Garzón por su alegría, su motivación, sus sabias palabras y su apoyo incondicional, mil gracias por su valioso tiempo y todos los momentos que compartió con nosotras creando mejores personas y profesionales; por todo su apoyo y creer en nosotras. Lo recordamos con cariño y admiración.

A nuestros padres y hermanos y por apoyarnos y alentarnos a no desfallecer, por su dedicación y amor que llenó de valentía cada día de este arduo camino, gracias por su comprensión y paciencia.

A nuestros amigos por permitirnos hacerlos parte de nuestras vidas y compartir sin fines de experiencias.

Contenido

	Pág.
Introducción	11
1. Generalidades del proyecto.....	14
1.1 Objetivos.....	14
1.1.1 Objetivo general.....	14
1.1.2 Objetivos específicos	14
2. Planteamiento y/o justificación del problema.....	15
3. Marco de referencias.....	17
3.1 Marco de antecedentes.....	17
3.2 Marco teórico.....	20
3.2.1. Inteligencia artificial.....	20
3.2.2. Aprendizaje automático.....	21
3.2.3. Modelamiento de clústers.....	21
3.2.4. Sector cuero, calzado y marroquinería.....	22
4. Factores asociados	24
5. Metodología para la investigación.....	27
5.1 Etapa 1. Búsqueda y análisis de literatura	28
5.2 Etapa 2. Definición de variables y selección / recolección, limpieza y preprocesamiento de datos	30
5.3 Etapa 3. Elección de algoritmos y minería de datos	30

5.4 Etapa 4. Aplicación de algoritmos clustering	31
5.5 Etapa 5. Análisis e interpretación de datos y documentación.....	31
6. Revisión de literatura	32
6.1 Ecuación de búsqueda.....	32
6.2 Revisión de literatura preliminar	34
6.3 Identificación de variables	40
7. Preprocesamiento y análisis de datos.....	43
7.1 Preprocesamiento de datos.....	44
7.1.1. Eliminación de variables y grupos industriales.	45
7.1.2. Valores faltantes.....	46
7.1.3. Imputación de datos.	46
7.1.4. Detección y eliminación de datos atípicos.....	47
7.1.5. Agrupación (categorización) de variables.	48
7.1.6. Normalización.....	54
7.2 Análisis descriptivo.....	55
8. Aplicación de algoritmos de agrupamiento	60
8.1 Algoritmo clustering K-means:.....	60
8.2 Algoritmo clustering fuzzy K-means:.....	69
9. Análisis de algoritmos.....	75
9.1. Algoritmo K-Means.....	76
9.1.1 Caracterización de conglomerados K-means.....	77
9.1.1.1 Agrupaciones año 2015.....	77
9.1.1.2 Agrupaciones año 2016.....	79

9.1.1.3 Agrupación año 2017	80
9.1.1.4 Agrupación año 2018.....	82
9.1.1.5 Agrupación año 2019.....	83
9.1.2 Comparativa de conglomerados K-Means.....	85
9.2 Fuzzy K-Means.....	93
9.2.1 Caracterización de conglomerados Fuzzy K-means	95
9.2.1.1 Agrupaciones año 2015.....	95
9.2.1.2 Agrupaciones año 2016.....	98
9.2.1.3 Agrupaciones año 2017.....	100
9.2.1.4 Agrupaciones año 2018.....	102
9.2.1.5 Agrupaciones año 2019.....	104
10. Análisis de productividad, competitividad y desarrollo	106
11. Comparaciones de algoritmos de clustering	112
12. Conclusiones.....	115
13. Recomendaciones	117
Referencias Bibliográficas	120

Lista de Figuras

	Pág.
Figura 1. Metodología.....	27
Figura 2. Diagrama de flujo revisión de literatura.....	29
Figura 3. Tipo de establecimientos según el tipo de empleados.....	56
Figura 4. Empleabilidad según la ubicación del establecimiento.....	56
Figura 5. Participación de los establecimientos en la encuesta según la actividad que realiza	57
Figura 6. Participación de la actividad económica a través del tiempo.....	58
Figura 7. Personal ocupado según las ventas totales.....	59
Figura 8. Diagrama de flujo algoritmo K-means.....	61
Figura 9. Diagrama de flujo Fuzzy K-means.....	69
Figura 10. Valor de compras nacionales en los diferentes años.....	85
Figura 11. Valor de compras internacionales en los diferentes años.....	86
Figura 12. Inversión bruta en los diferentes años.....	87
Figura 13. Salario pago a personal en los diferentes años.....	89
Figura 14. Producción bruta en los diferentes años.....	90
Figura 15. Valor en ventas en los diferentes años.....	91
Figura 16. Consumo intermedio en los diferentes años.....	91
Figura 17. Energía comprada en los diferentes años.....	92

Lista de Tablas

	Pág.
Tabla 1. Cumplimiento de objetivos.....	14
Tabla 2. Palabras claves y términos asociados	32
Tabla 3. Ecuación de búsqueda.....	33
Tabla 4. Agrupación de variables	48
Tabla 5. Correlación grupo 1 – mano de obra	51
Tabla 6. Correlación grupo 2 – producción bruta.....	51
Tabla 7. Correlación grupo 3 – consumo intermedio	52
Tabla 8. Correlación grupo 4 – activos fijos.....	53
Tabla 9. Correlación grupo 5 – Inversión bruta.....	53
Tabla 10. Resultados codo de jambú	62
Tabla 11. Variación de variables a través del tiempo	65
Tabla 12. Selección de variables.....	66
Tabla 13. Agrupación K-Means.....	67
Tabla 14. Graficas coeficiente de partición difuso	71
Tabla 15. Graficas agrupamiento Fuzzy K-Means	74
Tabla 16. Agrupamiento realizado por el algoritmo K-Means.....	76
Tabla 17. Agrupamiento realizado por el algoritmo Fuzzy K-Means	94

Apéndices

Ver en documentos adjuntos y adicionalmente pueden ser visualizados en la Base de Datos de la biblioteca de la Universidad Industrial de Santander

Apéndice 1. Microdatos DANE

Apéndice 2. Base de datos procesadas

Apéndice 3. Clasificación de Variables

Apéndice 4. Análisis Exploratorio

Apéndice 5. Resultados y Análisis de datos K-means por años

Apéndice 6. Resultado y Análisis de datos Fuzzy K-means por año

Apéndice 7. Análisis detallado por años K-means

Apéndice 8. Análisis detallado por años Fuzzy K-means

Apéndice 9. Análisis Bibliométrico

Apéndice 10. Resultado datos normalizados Fuzzy K-means por año

Apéndice 11. Comparación de algoritmos

Apéndice 12. Artículo académico de investigación.

Resumen

Título: un algoritmo de clustering difuso para el análisis de las causas que afectan el desarrollo, productividad y competitividad de la industria de calzado, cuero y marroquinería con relación a su producción y mano de obra*

Autores: Karol Juliana Benavidez Robles, Liz Bleidy Tatiana Jiménez Wandurraga**

Palabras claves: aprendizaje automático, agrupamiento, mano de obra, producción, manufactura.

Descripción:

El sector de cuero, calzado y marroquinería en Colombia es una actividad industrial especializada en el tratamiento del cuero y la manufactura de varios productos a base de este insumo y demás materiales. Este sector abarca diferentes actividades como la recolección y manipulación del cuero, diseño y elaboración de los productos de calzado y marroquinería y/o su venta y distribución. De acuerdo con el estado actual de esta industria, se tienen problemáticas que afectan directa e indirectamente su producción y venta como lo es la escasez en la mano de obra calificada, competencia desleal, variedad de productos sustitutos, falencias en la automatización en sus procesos de producción, limitadas adquisiciones de tecnología innovadora entre otras.

La presente investigación ejecuta un algoritmo de clustering difuso y nítido por medio de software Python el cual agrupa las diferentes empresas del sector de cuero calzado y marroquinería a partir de una base de datos proporcionada por el DANE en su encuesta anual manufacturera en el periodo de tiempo 2015/2019. Estos algoritmos permiten analizar y comparar el resultado de cada uno para observar el comportamiento a través de los años y así determinar aquellos factores que influyen directa e indirectamente en el desarrollo, la productividad y la competitividad del sector y a su vez plantear propuestas de mejora con la finalidad de reestablecer y buscar el crecimiento de este.

* Trabajo de grado

** Facultad de Ingeniería Físico Mecánicas. Escuela de Estudios Industriales y Empresariales. Director: Henry Lamos Díaz, Ph. D en Física - Matemática. Codirector: David Esteban Puentes Garzón, M.Sc. Ingeniería Industrial.

Abstract

Title: A fuzzy clustering algorithm for the analysis of the causes affecting the development, productivity and competitiveness of the footwear, leather and leather goods industry in relation to its production and workforce*

Authors: Karol Juliana Benavidez Robles, Liz Bleidy Tatiana Jiménez Wandurraga**

Keywords: machine learning, clustering, workforce, production, manufacturing.

Description:

The leather, footwear and leather goods sector in Colombia is an industrial activity that involves the processing of leather and the manufacture of various leather-based products and the production of leather footwear and other materials. This sector covers different activities such as the collection and manipulation of leather, design and manufacture of footwear and leather goods and/or their sale and distribution. According to the current state of this industry, there are problems that directly and indirectly affect its production and sales, such as the shortage of skilled labor, unfair competition, variety of substitute products, lack of automation in its production processes, limited acquisitions of innovative technology, among others.

The present investigation executes a fuzzy and soft clustering algorithm by means of a software called Python which clusters the different companies of the leather goods sector from a dataset provided by DANE on their annual manufacturing survey in the timeline 2015/2019. These algorithms make it possible to analyze and compare the results of each to observe the behaviors through the year and determine those factors that directly and indirectly influence the development, productivity and competitiveness of the sector and, in addition, to make recommendations for improvement in order to reestablish and seek the growth of this industry

* Bachelor thesis

** Faculty of Physical Mechanical Engineering. School of Industrial and Business Studies. Advisor: Henry Lamos Díaz, Ph. D in Physics - Mathematics. Co-advisor: David Esteban Puentes Garzón, M.Sc. Industrial Engineering

Introducción

El sector de cuero, calzado y marroquinería colombiano es considerado como uno de los sectores de crecimiento activo y permanente de la economía colombiana debido a las diferentes oportunidades de desarrollo y sólidas vocaciones regionales para el crecimiento el cual favorece los procesos de valor agregado, diseño e innovación. Sin embargo, para seguir creciendo, el sector debe realizar mejoras en los procesos productivos e implementar buenas prácticas dentro de los mismos para alcanzar un excelente entorno económico donde no exista el comercio desleal y se consolide un aumento estable dentro de su producción (Quintero Castillo, 2020).

Este sector tiene diferentes actividades que van desde la adquisición y manejo del cuero, pasando por el diseño y fabricación de los productos de calzado y marroquinería para terminar en su venta y distribución. En Colombia, el sector es conformado por una amplia cantidad de empresas, que en su mayoría se encuentran clasificadas como micro y pequeñas empresas y por lo general, dependientes del producto del cuero como insumo principal.

De acuerdo con la actividad económica de este sector a través de los años se ha visto que existe la necesidad de implementar “elementos” que aporten diferencia respecto a la competencia como la calidad de la fabricación del producto, el diseño, el posicionamiento de marca, los procesos y técnicas para la elaboración del producto final. A su vez, se requiere la conformación de redes empresariales, el crecimiento de clústers de producción, articular la cadena de confecciones y utilizar las zonas francas y los distritos industriales con la finalidad de sacarle provecho a las economías de ventajas tributarias y acumulación de actividades. En la cadena de curtiembres y marroquinería las actividades en su mayoría son delegadas al trabajo realizado por

terceros puesto que en las curtiembres se adquieren los insumos y en la marroquinería se delegan los procesos de armado, costura y modelaje (Barrientos et al., 2007).

El grupo de empresas que se dedican a la fabricación y producción de calzado y cuero tienen problemáticas que impiden el buen desempeño económico del mismo, estas dificultades se ven reflejadas en la débil gestión del sistema productivo, los altos costos de transacción, pocos niveles de asociatividad entre los empresarios y el deficiente manejo de los recursos humanos. Algunas de las razones por las que este sector presenta estos inconvenientes es porque existe informalidad en las relaciones entre proveedores y productores, falta manejo de los desechos del cuero de forma amigable con el medio ambiente, no todas las empresas tienen elementos de protección industrial, tienen poca innovación en los métodos de producción y diseño de los productos, existe una alta inestabilidad laboral y poco interés por las personas para involucrarse en actividades productivas del sector (SENA, 2016).

Colombia es un país que requiere de importaciones de productos netos de calzado y marroquinería debido a que la producción nacional no da abasto para cubrir toda la demanda que requiere el sector. Este comportamiento crea un déficit en la balanza comercial puesto que las importaciones son mayores a las exportaciones que se realizan. Una de las razones por la que se presenta este comportamiento es porque el sector se encuentra conformado por micro y pequeñas empresas que no cuentan con el potencial y el conocimiento para producir mercancía y exportarla a los demás países, adicionalmente, esta distribución de productos terminados de calzado y marroquinería se encuentra controlada por aquellas grandes organizaciones internacionales lo cual viene siendo un problema para aquellos establecimientos locales.

Adicionalmente, Colombia cuenta con instituciones que tienen el potencial para apoyar el proceso que realizan las micro y pequeñas empresas del sector, sin embargo, los empresarios no le

dan la importancia ni les sacan el provecho a estos servicios para así poder potencializar la producción, esto posiblemente ocurre por el desconocimiento que se tiene de los servicios que estas instituciones prestan o del proceso para poder adquirirlo. A su vez, se le suma el bajo desarrollo que tiene el país respecto a la infraestructura vial, férrea y portuaria el cual dificulta el respectivo movimiento de los insumos y productos terminados a lo largo del territorio independientemente si es para su comercialización o exportación (SENA, 2016).

Mencionado lo anterior, el presente trabajo de investigación busca las circunstancias (causas) que permiten comprender el comportamiento que ha presentado el sector a través de los años teniendo en cuenta su producción y mano de obra por medio de técnicas de aprendizaje automático más específicamente mediante técnicas de clustering nítido y difuso para conocer e interpretar aquellos patrones que impactan y definen el sector de cuero, calzado y marroquinería colombiano.

La estructura de este documento es como sigue: inicialmente se realiza una búsqueda y análisis de literatura para profundizar en temas relacionados a la temática y a la aplicación de los algoritmos, de acuerdo a la base de datos brindada por el DANE se identifican las variables de interés para el proyecto y se hace una limpieza y preprocesamientos de datos, teniendo la base de datos consolidada se procede al análisis de las variables mediante diferentes técnicas de aprendizaje automático, una vez se obtienen los resultados se analizan e interpretan. Finalmente, se realiza una comparación entre los algoritmos, se presenta una serie de hipótesis sobre la tendencia del sector y se proponen algunas recomendaciones para el mejoramiento de este en base a la taxonomía hallada.

Conforme a lo anteriormente descrito, el cumplimiento de los objetivos planteados para el presente trabajo de investigación se presenta en la tabla 1.

Tabla 1.*Cumplimiento de objetivos*

Objetivos Específicos	Cumplimiento
Identificar aplicaciones de clustering difuso en el análisis de la producción y mano de obra en el subsector de calzado cuero y marroquinería, a partir de una revisión de literatura	Capítulo 6
Desarrollar algoritmos de clustering difuso y nítido para la clasificación de patrones que impactan en la producción y mano de obra en la industria colombiana de cuero, calzado y marroquinería	Capítulo 8
Comparar los grupos obtenidos mediante un clasificador nítido y un clasificador difuso	Capítulo 11
Describir los grupos conformados a partir de la aplicación de las técnicas de clustering	Capítulo 9 y 10
Elaborar un artículo académico publicable a partir de la investigación realizada	Apéndice 12

1. Generalidades del proyecto

1.1 Objetivos

1.1.1 Objetivo general

Construir un algoritmo de clustering difuso para el análisis de los factores que influyen significativamente en el desarrollo, productividad y competitividad en la industria colombiana de calzado cuero y marroquinería en relación con su producción y mano de obra.

1.1.2 Objetivos específicos

Identificar aplicaciones de clustering difuso en el análisis de la producción y mano de obra en el subsector de calzado cuero y marroquinería, a partir de una revisión de literatura

Desarrollar un algoritmo de clustering difuso para la clasificación de patrones que impactan en la producción y mano de obra en la industria colombiana de cuero, calzado y marroquinería

Comparar los grupos obtenidos mediante un clasificador nítido y un clasificador difuso

Describir los grupos conformados a partir de la aplicación de las técnicas de clustering

Elaborar un artículo académico publicable a partir de la investigación realizada

2. Planteamiento y/o justificación del problema

La industria de calzado, cuero y marroquinería en Colombia es considerada un generador de empleo potencial, según la Asociación Colombiana de Industriales de Calzado – ACICAM, en el 2020 este subsector generó aproximadamente 500,000 empleos directos e indirectos, convirtiéndolo en uno de los más importantes en la industria manufacturera, contando con una participación del 3 % del empleo del país (Forero, 2016). Así mismo, en la última década este subsector mostró aumentos significativos, logró consolidar cerca de 30,000 empresas dedicadas a actividades relacionadas con calzado, cuero y marroquinería (Colombia productiva, 2020). Para el 2018 la marroquinería demostró un aumento del 5 % en su producción y un 9,7 % en ventas totales, crecimientos que se pueden asociar al aumento de las exportaciones hacia Estados Unidos y China (ACICAM, 2019).

En la actualidad, este subsector juega un papel fundamental en el crecimiento económico de ciudades como Bucaramanga, Bogotá, Cali y Medellín esto se debe a que son las zonas que conforman los clústers más importantes de todo el país. En el 2017, los colombianos tuvieron un consumo per cápita estimado de 2,8 pares anuales lo que representa un estimado de 4 a 5 billones de pesos colombianos (ACICAM, 2018). Aunque la producción y comercialización de calzado, cuero y marroquinería es crucial para la economía nacional, las empresas que conforman esta industria se han visto en jaque debido a varios problemas, como: contrabando y falta de un entorno económico estable, (competencia desleal y la gran variedad de productos sustitutos); así mismo, en cuanto a fabricación existe debilidades en lo que tiene que ver con automatización en sus procesos de producción y poca adquisición de tecnología innovadora. Por último, en relación con la materia prima cuenta con poca disponibilidad de insumos de alta calidad que generalmente se encuentran a elevados costos, sobreproducción de cuero por el aumento descontrolado de exportaciones a Brasil, y la creciente de informalidad en las curtiembres a causa de la poca tecnificación con la que se cuenta para la obtención de la materia prima en el subsector.

Sumado a los problemas descritos, la industria presenta una escasez de mano de obra calificada, convirtiéndola en un factor crucial para la sostenibilidad del subsector, a causa de que la disponibilidad de ésta presenta una amplia incidencia en la productividad, competitividad y desarrollo de las empresas que conforman la industria. En este orden de ideas, se propone este trabajo de grado con el objetivo de: entender, interpretar y caracterizar la dinámica de la producción y mano de obra por medio de técnicas de aprendizaje automático. La aplicación de un algoritmo de clustering difuso y nítido resulta eficiente para el agrupamiento de datos, determinando parámetros representativos en los mismos. Así, este recurso brindará la oportunidad

de indagar, organizar, extraer información y descubrir patrones o tendencias que sean de utilidad para apoyar la toma de decisiones y hacer frente a los problemas que se han venido identificando.

Los datos consultados y analizados serán tomados de la Encuesta Anual Manufacturera realizada por el Departamento Administrativo Nacional de Estadísticas - DANE, para la ventana de tiempo de 2015 a 2019 en todo el territorio colombiano. Esta fuente de información resulta ser de gran ayuda para el correcto modelado del algoritmo que se propone realizar, ya que cuenta con datos importantes en cuanto a mano de obra y productividad de la industria. En el apéndice 1 se encuentran los microdatos proporcionados por el DANE.

3. Marco de referencias

En este capítulo se presenta el marco teórico en función de las temáticas de interés que aportan a la comprensión, interpretación y caracterización de la investigación propuesta.

3.1 Marco de antecedentes

En esta sección se estudian algunos artículos recopilados de la revisión de literatura que han implementado la técnica de clustering en diferentes ámbitos en los últimos 5 años.

Julián M y Petra Unterberger (2021), realizan una agrupación de K-Means con el propósito de detectar grupos de empresas que se diferencian por sus incentivos de relocalización; utilizaron una muestra de 94 empresas industriales alemanas que realizan actividades globales de

abastecimiento y producción. El objetivo de este estudio fue indagar que incentivos de relocalización son necesarios para los diferentes grupos de empresas manufactureras alemanas por medio del aprendizaje no supervisado dentro del contexto de reshoring para agrupar y clasificar las empresas. En este estudio se utilizó el algoritmo de agrupamiento de K-Means con el fin de mantener abierta la posibilidad de reconocer aquellos datos que no son etiquetados y distinguir la estructura por medio de datos. La finalidad de la aplicación de K-Means es agrupar aquellas empresas que tienen resultados muy similares y a su vez mantener las empresas con diferentes datos en diferentes grupos. En el caso del preprocesamiento de datos lo realizaron por medio de los términos de estandarización y selección de características, para indagar el número óptimo de grupos aplicaron el método del codo y los resultados los ajustaron de acuerdo con los requisitos de solidez. Asimismo, se obtuvo como resultado grupos muy robustos que difieren en aspectos como la priorización de la incertidumbre respecto a la evolución económica y política, el incremento en los costes laborales en aquellos países con bajos salarios, los altos costes de coordinación y comunicación y las mejoras en las posibilidades de seguimiento y control.

Por otra parte, Lee y colaboradores (2021) en su estudio titulado “A hybrid approach combining fuzzy c-means-based genetic algorithm and machine learning for predicting job cycle times for semiconductor manufacturing” proponen un enfoque híbrido el cual acopla el algoritmo genético (GA) apoyado en c-means y una red de retro propagación (BPN) con el fin de predecir el tiempo de ciclo de trabajo en las fábricas de fabricación de semiconductores y mejorar la efectividad de la predicción basado en las relaciones entre los atributos de trabajo y el tiempo del ciclo del trabajo. Para el desarrollo de este enfoque tuvieron en cuenta todos los registros de trabajo los cuales se segmentan en dos: el primero se usa para entrenar y agrupar y el segundo para probar. Para preordenar el primer conjunto de datos en distintos grupos lo desarrollaron por medio del

método de clasificación (GA) centrado en c-means y para predecir el tiempo del ciclo y de igual forma compararlo con el segundo conjunto de datos utilizaron el predictor BPN. Los resultados que se obtienen por medio de GA basado en c-means se usan como datos de entrenamiento y se incluyen en el predictor de BPN el cual se entrena para pronosticar el tiempo de ciclo de los nuevos trabajos con diferentes atributos. Para la verificar efectividad del enfoque propuesto los autores comparan los resultados con un caso de estudio en donde concluyen que el método de clasificación GA permite preclasificar correctamente el primer conjunto de datos de registro de trabajo el cual deja que el predictor genere predicciones precisas para el respectivo tiempo de ciclo de trabajo, igualmente, los resultados de clasificación de registros de trabajo permanente y difuso obtenidos facilitan el trabajo que realiza el predictor BPN para que este pronostique el tiempo del ciclo de trabajo con una alta precisión.

Desde otro punto de vista, Duan y colaboradores (2020) en su artículo “Modelo de predicción de la demanda de Big Data de energía basado en un conjunto aproximado difuso” plantean un modelo de predicción de la demanda para aquellos datos masivos de energía apoyado de un modelo de conjunto de aproximación difusa para dar respuesta a aquellas predicciones que generan gran incertidumbre a la hora de pronosticar la demanda de la energía. Para el desarrollo de este modelo predictivo inicialmente partieron de aquellos datos proporcionados para determinar los factores que influyen en la demanda de energía; seguidamente, hicieron uso del algoritmo Fuzzy-K-Means para discriminar los datos a partir de las características del conjunto aproximado difuso. Adicionalmente a esto, establecieron unos parámetros de decisiones y de igual forma calcularon la importancia que debe tener cada atributo para luego realizar una reducción de atributos, una vez realizado lo mencionado anteriormente, separaron las reglas de correlación para poder establecer un modelo de predicción. Finalmente, pudieron concluir que la demanda

energética es el centro y el obstáculo en el campo de la predicción energética de los últimos años a partir del modelo para pronosticar la demanda de big data de energía apoyado en un modelo clustering difuso. Los autores esperan indagar el camino de desarrollo de la modificación de la estructura energética de China y el reemplazo de la energía eléctrica por medio de la predicción de la demanda de la energía futura.

3.2 Marco teórico

En este fragmento, se determina el marco conceptual con base en las temáticas de utilidad que aportan a la comprensión, interpretación y caracterización de la presente investigación.

3.2.1. Inteligencia artificial.

La Inteligencia Artificial es considerada como esa capacidad que tienen las máquinas para emplear algoritmos, estudiar los datos y aprovechar esa información recolectada en la toma de decisiones, así como lo haría un ser humano. Esta herramienta cuenta con una proporción de errores significativamente baja, la cual permite obtener beneficios en las mejoras y aprovechar de una mayor eficiencia en los procesos o situaciones (Rouhiainen, 2018).

Según Salazar (2021) la IA es el presente, puesto que cada vez son más las organizaciones que se suman a la aplicación de la IA como herramienta para aumentar la eficiencia de sus diferentes procesos. La automatización, rapidez y exactitud en el análisis de datos complejos son componentes clave en estos sistemas que dominan a la perfección, además, este tipo de tecnología

promueve el incremento de los ingresos y la minimización de gastos debido a la realización de predicciones de alta precisión centrado en patrones.

3.2.2. Aprendizaje automático.

El aprendizaje automático es una rama de la inteligencia artificial, la cual tiene como objetivo desarrollar procedimientos que permitan a las computadoras estudiar y aprender para que se conviertan en un soporte primordial para el manejo de datos a gran escala (Hinestroza & Cárdenas, 2018). Este tipo de aprendizaje es más conocido como “Machine Learning” o “ML” y se fundamenta en alcanzar el aprendizaje autónomo de las máquinas para que solucionen los problemas o situaciones actuales.

Para esta investigación el tipo de aprendizaje a implementar es el aprendizaje no supervisado, se usa cuando el conjunto de datos no se encuentra etiquetado y por esta razón se desea construir una serie de grupos de “individuos” con características similares mediante el uso de métodos de clustering.

3.2.3. Modelamiento de clústers.

Este modelamiento es una fuerte herramienta en el tratamiento de datos debido a la capacidad que tiene para incorporar la información de diversas fuentes como modelos empíricos, mediciones o conocimientos de expertos. Estos conjuntos difusos trabajan como interfaz entre aquellos datos numéricos, variables cualitativas y las entradas y salidas del sistema (González & Barato, n.d.).

En esta investigación se requiere trabajar con una metodología de identificación nítida y difusa en específico, esta es la técnica de “Clustering” puesto que se fundamenta en el manejo de datos de entrada y salida del sistema y su agrupamiento en subconjuntos (clústeres). Las técnicas de clustering tienen como objetivo hallar subgrupos que tienen cierto grado de similitud dentro de un conjunto de datos; además de esto, determina aquellos parámetros que representan a cada uno de los subconjuntos. La aplicación de dos técnicas de clustering permitirán realizar comparaciones más acertadas, cabe resaltar que la máxima diferencia entre las dos metodologías es el grado de pertenencia fuerte en los algoritmos nítidos y la flexibilidad de pertenencia (suave) en los algoritmos difusos.

Para las técnicas clustering nítido (K-Means), cada dato tiene exclusividad de un solo conglomerado. Este algoritmo se encarga de escoger un centroide de manera inteligente para compararlo con los demás puntos de datos en función de su similitud para calcular la distancia entre los mismos, así, los puntos de datos semejantes al centroide se asignan al grupo de su centroide; para la formación de los grupos K se itera el algoritmo calculando nuevos centroides (Khalid et al., 2014).

3.2.4. Sector cuero, calzado y marroquinería.

Este sector industrial hace referencia a la actividad industrial que está relacionada con el tratamiento de cuero, la manufactura de distintos productos que son a base de este insumo y la producción de calzado en cuero y otros materiales. Este tipo de sector adjunta diferentes actividades que se basan desde la recolección y tratado del cuero, seguidamente por el diseño y manufactura de los productos de calzado y marroquinería hasta su distribución y venta. Por el

comportamiento que tiene el sector respecto al PIB, se puede distinguir su variación siendo esta una característica de un sector que es enormemente dependiente del comportamiento referente a la demanda por los productos que ofrece al mercado, que se transforma a partir de las tendencias de moda o gustos personales de la población y adicionalmente, es afectada directamente por el ingreso de productos al país con bajos costos de producción y por los productos de contrabando (ACICAM, 2019).

En Colombia, este sector industrial se ha venido expandiendo durante la última década y se encuentra conformado por un aproximado de 30.000 empresas formales respecto a insumos, transformaciones y comercializaciones, siendo altamente intensivo en la mano de obra. Su potencial exportación es enorme debido a los tratados de libre comercio firmados por el país y a la calidad, creatividad y diferenciación en el diseño de los productos nacionales. Actualmente, el país cuenta con grandes vocaciones regionales que contribuyen con el desarrollo de este sector los cuales facilitan la reproducción de procesos de gran valor agregado, diseño e innovación, por esta razón, se avanza con el fortalecimiento del tejido empresarial, el crecimiento de participación en el mercado interno, la batalla contra el comercio desleal, el fortalecimiento de un crecimiento estable en la producción y así lograr que la “marca país” se introduzca primordialmente con los productos manufacturados a los mercados cercanos (Productiva, 2021).

4. Factores asociados

Este capítulo se crea con la intención de presentar factores asociados a mano de obra y producción que inciden significativamente en la productividad, desarrollo y competitividad de las empresas.

En la actualidad las organizaciones han adoptado la idea que a la medida que una empresa es altamente productiva su capacidad de competencia aumenta, por esto, hablar de competitividad implica directamente el término de productividad (Díaz Muñoz et al., 2020). Así, la competitividad constituye la base fundamental del crecimiento y desarrollo para las organizaciones del futuro que se preocupan por ser cada vez más productivas. En el mismo orden de ideas, Porter (1985) sugiere que la competitividad incrementa la productividad mediante el uso de los recursos ya que una buena administración permite responder rápidamente a todos los requerimientos del mercado.

A razón de lo anterior, se parte de la premisa que las empresas colombianas dedicadas a actividades relacionadas con cuero, calzado y marroquinería caracterizadas por ser competitivas son altamente productivas. Se entiende por productividad la relación positiva entre los recursos disponibles y los productos obtenidos, es decir, se busca la optimización del consumo de recursos como tierra, capital, materiales, energía, entre otros para obtener mejores resultados (Martínez De Ita, 2019). También, dadas las altas definiciones de productividad para agregar precisión se hablan de tres medidas de producción: total, laboral parcial y capital parcial.

En el mismo orden de ideas, la competitividad se atribuye al conjunto de recursos y capacidades que marcan una diferencia significativa entre las empresas. Porter destaca la competitividad como la capacidad de las empresas de producir y vender productos diferenciados

de alta calidad, en buenas condiciones monetarias y con mayor oportunidad. En otras palabras, el éxito competitivo se alude a los diversos recursos y capacidades que soportan a las empresas y las hacen diferentes a los demás competidores del sector (Porter, 1985).

Con la evolución de los mercados, obligando a las empresas a ser más competitivas en visión a su desarrollo se crea la necesidad trabajar y estudiar algunos factores que podrían tener alta incidencia en el aumento de la productividad. Algunos elementos internos principales que se observan en el comportamiento de una organización para aumentar su presencia en el mercado respecto a sus rivales por medio del aumento de la productividad y desarrollo se destacan: tecnología, innovación, recursos humanos, producción, compras, calidad y logística. Perez Zúñiga et al., (2014) proponen la existencia de una relación altamente positiva entre las inversiones monetarias destinadas a investigación y desarrollo con el aumento de la productividad.

Así mismo, dentro de las empresas surge la necesidad de incrementar inversiones de capital físico, tecnológico y humano para el desarrollo de la competitividad como componente fundamental de la productividad; cuando las organizaciones logran optimizar sus recursos en calidad de gestión empresarial la ganancia de competitividad es mayor.

Aragón y Sánchez destacan la influencia de una acertada gestión de recursos humanos como mecanismo de productividad resaltando que resulta ser un efecto aún más favorable en organizaciones pequeñas ya que se logra mejor comunicación con la alta dirección respecto a los objetivos propuestos por cada organización (Aragon & Sanchez, 2005). En cuanto a capital humano, el conocimiento, habilidades, competencias y demás atributos constituyen la capacidad de las empresas para elevar y aumentar su productividad gracias a la capacidad de innovación y mejora de los seres humanos.

Por otro lado, la tecnología con la que cuentan las empresas influye en la productividad abriendo camino a la automatización de los procesos que ofrecen la oportunidad de alcanzar altos niveles de producción con mayores estándares de calidad siempre y cuando se proporcione el adecuado manejo de las maquinarias implementadas en el desarrollo del producto (Herrera Fontalvo et al., 2018). Cuando las empresas apuntan a la innovación, esta se destaca por convertirse en una fuente de valor constante permitiéndoles permanecer en la industria.

Los productos con los que cada empresa ingresa en el mercado influyen en la productividad por diversos motivos como: el proceso de elaboración y la calidad del producto final deben adaptarse a los sistemas de producción de las organizaciones. Cuando en el mercado se lanza un producto de calidad con altos estándares de productividad las organizaciones serán competitivas generando mayores ingresos como consecuencia de la entrega de productos de calidad con su debida optimización de recursos (consumo de materias primas, costos de producción, habilidades de los colaboradores, requisiciones de materias primas, otros gastos, etc.)

La gestión del flujo de materias primas y el eficiente manejo de inventarios o el factor de tiempo en la cadena de suministro demuestran desarrollo en los procesos que a su vez generan productos con estándares de calidad. También, Molina y Hernández (1999) destacan la capacidad competitiva de las empresas en función de crecimiento y desarrollo para fortalecer las organizaciones a través del tiempo como consecuencia de estrategias adecuadas de inversión en recursos que faciliten sus procesos. La decisión de inversión o innovación de las empresas resulta un elemento fundamental de desarrollo y competitividad gracias a la gestión de recursos que brinda haciendo empresas cada vez más productivas, así mismo, determinar los objetivos básicos de consumo bajo una planeación estratégica resulta fundamental para tener éxito en la optimización de recursos.

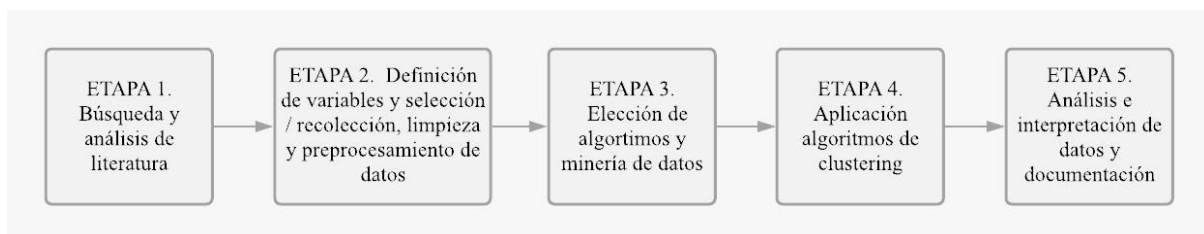
Por otro lado, cabe destacar la existencia de diversos aspectos que contribuyen al decaimiento de la competitividad empresarial cerrando la posibilidad de desarrollo, algunos errores se aluden malas prácticas detectadas al final de la cadena productiva que implican pérdidas de tiempo, materia prima y recurso humanos; generalmente lo anterior es ocasionado por falta de comunicación, mala ejecución de procesos productivos, falta de control en los estándares de calidad, requisiciones de materia prima de baja calidad, retrasos en las ordenes de materiales, equivocaciones en el personal operativo, entre otros.

5. Metodología para la investigación

El desarrollo de la investigación propuesta se ejecutará a lo largo de cinco etapas como se muestra en la figura 1

Figura 1.

Metodología

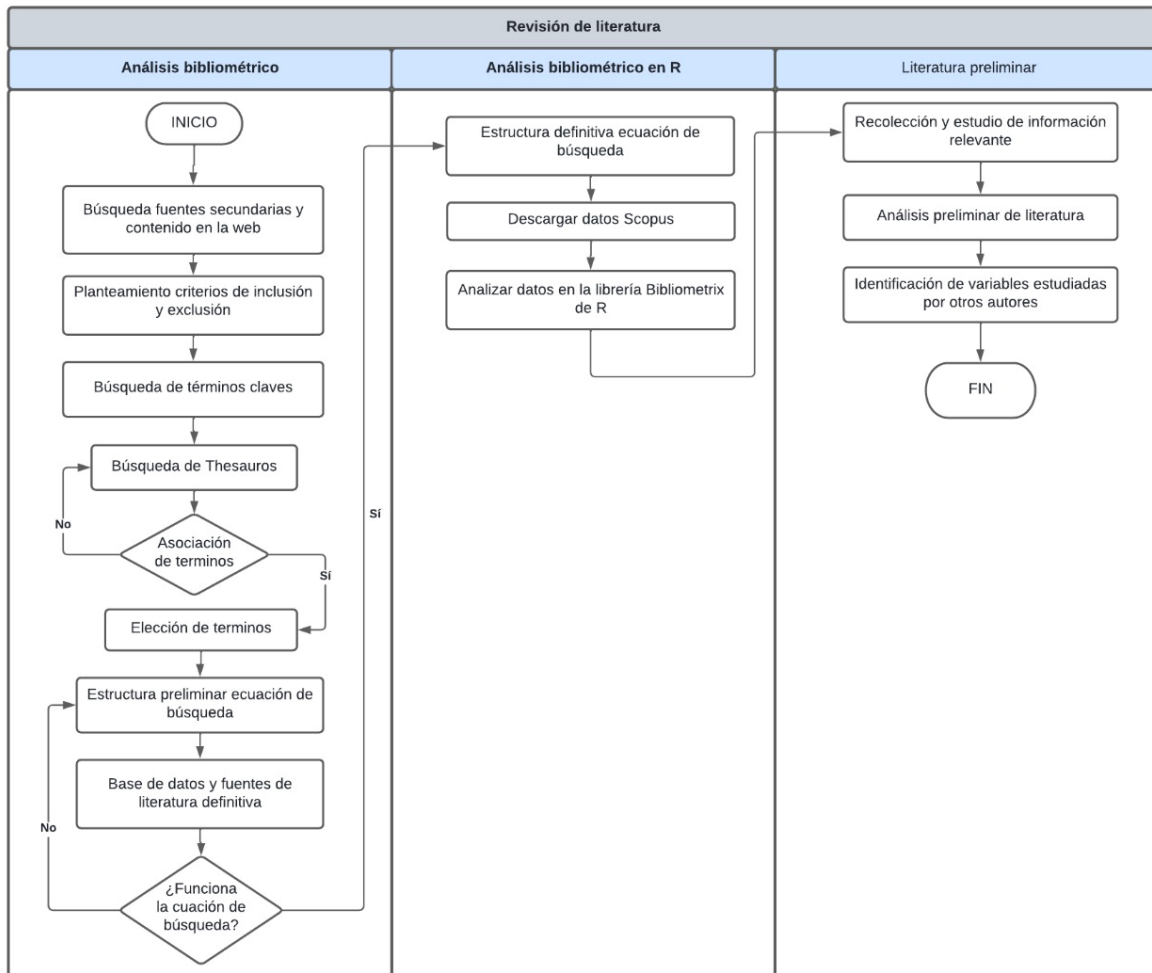


5.1 Etapa 1. Búsqueda y análisis de literatura

En esta etapa se proponen diferentes actividades que permiten un acercamiento a las temáticas de interés para la investigación que se propone. Inicialmente, a través del motor de búsqueda de Google académico se hace una primera búsqueda para ampliar el conocimiento previo a la temática. Luego, se procede con la investigación preliminar para el planteamiento de la ecuación de búsqueda que nos proporcionará información relevante a los temas de estudios realizados en los últimos años acerca de Machine Learning con aplicaciones en aprendizaje no supervisado relacionados (TITLE-ABS-KEY ((“fuzzy clustering” OR “clustering” OR “k-means clustering” AND “machine learning” OR “unsupervised learning”) AND (Productivity OR development OR competitive OR evolution) AND (workforce OR manpower OR “human resources” OR manufacturing OR production OR manufacture)))

Para el planteamiento de la ecuación de búsqueda en primer lugar se proponen algunos términos claves que faciliten la obtención de documentos, luego se establecen algunos criterios de inclusión y exclusión para tener en cuenta. Por medio de las palabras claves definidas se realiza una nueva búsqueda de términos relacionados o tesauros para una mejor segmentación de las temáticas a consultar.

Mencionado lo anterior, se cuenta con los recursos necesarios para realizar la búsqueda especializada en las bases de datos proporcionadas por la biblioteca virtual y el análisis de la documentación requerida como se muestra en la Figura 2 la cual representa el diagrama de flujo propuesto para esta etapa.

Figura 2.*Diagrama de flujo revisión de literatura*

De la información recolectada, se realiza una lectura profunda con el objetivo de establecer de manera acertada el problema de la aplicación de aprendizaje no supervisado con técnicas de agrupamiento. Cabe mencionar que es de suma importancia realizar un análisis bibliométrico a través del software R con el objetivo de descubrir las tendencias de estudios en los últimos años.

5.2 Etapa 2. Definición de variables y selección / recolección, limpieza y preprocesamiento de datos

A partir de la documentación realizada por medio de la ecuación de búsqueda se identifican **algunas variables** que han sido de interés para otros autores que han realizado investigaciones con relación al estudio de la influencia de factores como producción y mano de obra en diferentes campos de acción y se proponen otras que pueden ser relevantes para la presente investigación.

El departamento administrativo nacional de estadística DANE proporciona microdatos de información referentes al sector manufacturero colombiano, por lo que resulta necesario la ejecución de una recolección y limpieza preliminar para separar exclusivamente aquellos datos pertenecientes a las empresas que conforman la industria de cuero calzado y marroquinería en el país. Una vez realizado el paso anterior, se procede con la validación y selección de aquellas variables de interés según el estudio de otros autores que han investigado temas relacionados, por lo que resulta preciso llevar a cabo una limpieza de variables a conveniencia.

Finalmente, se realiza un preprocesamiento de los datos obtenidos con el objetivo de eliminar ruido y recopilar únicamente aquella información necesaria para el desarrollo del algoritmo propuesto. El apéndice adjunto 2 presenta la base de datos preprocesada.

5.3 Etapa 3. Elección de algoritmos y minería de datos

En esta etapa, se realiza un análisis exploratorio para la selección de algoritmos de extracción de datos y métodos a implementar para la búsqueda de patrones con ayuda de la información ya mencionada. Seguido a esto, se procede a la elección apropiada de minería de

datos. De acuerdo con el objetivo de descubrimiento (Fayyad & Piatetsky-Shapiro G Smyth, 1996) que consta de la predicción de comportamiento y la descripción de estos a través de la aplicación de algoritmos bajo las técnicas de clustering establecidas.

Del mismo modo, a partir de la literatura base se ejecutan los algoritmos decantados en el lenguaje de programación de Python para el posterior estudio de la información de salida de los algoritmos implementados.

5.4 Etapa 4. Aplicación de algoritmos clustering

A partir de los algoritmos propuestos para la predicción y descripción del comportamiento de la industria de calzado cuero y marroquinería colombiana se establecen métricas de comparación de agrupamientos por medio de clasificadores diferentes para estudiar las respectivas variaciones y la evaluación de cada algoritmo.

Adicionalmente, es importante analizar y describir la influencia de las variables anteriormente propuestas por los autores con las variables que fueron resultado del algoritmo de clustering aplicado con el objetivo de caracterizar las variables con más influencia en la industria para un posterior contraste con el panorama actual de la industria.

5.5 Etapa 5. Análisis e interpretación de datos y documentación

En esta etapa final se propone la revisión de resultados por medio de análisis e interpretación de variables de salida obtenidas con la aplicación de clústers difusos y nítidos para el entendimiento de la dinámica del sector en estudio.

Por último, se presenta la respectiva documentación de la investigación en la que se destacan los análisis e interpretaciones anteriormente expuestas junto con sus respectivas conclusiones y recomendaciones que serán plasmados en el libro final del trabajo de investigación junto con un artículo académico de carácter publicable.

6. Revisión de literatura

6.1 Ecuación de búsqueda

En la construcción de la ecuación de búsqueda se proponen tres grupos con términos claves buscando encontrar documentación que se encuentre bastante relacionado con el tema de investigación propuesto. Los términos asociados se presentan en la siguiente tabla.

Tabla 2.

Palabras claves y términos asociados

Palabra clave	Inglés	Español
Inteligencia artificial	Artificial intelligence /machine learning o ML	Inteligencia artificial
Mano de obra	Work market / human resources o HR / job / forces human / staff / workforce	Mercado de trabajo / recursos humanos / empleo / fuerzas humanas / personal
Clustering difuso / fuzzy clustering	K-means (K-means clustering)	K-medias
Productividad	Productivity / industrial productivity / labor productivity / efficiency / work performance	Productividad / productividad industrial / productividad laboral

Palabra clave	Inglés	Español
		/ eficiencia / rendimiento del trabajo
Competitividad	Competence / competing / capability / comparative advantage / competitive	Competencia / compitiendo / capacidad / ventaja comparativa / competitividad
Desarrollo	Industrial development / integral development / development	Desarrollo industrial / desarrollo integral / desarrollo
Producción	Factor of production / manufacturing / production / manufacture	Factor de producción / fabricación / producción / fabricación

Planteados los términos claves y sus respectivos tesauros, a través de la herramienta de búsqueda Scopus se fortalecen todos los temas relacionados a la temática por medio de dos operadores lógicos esenciales, el primero busca asociar todos los términos que contienen similitud a los términos asociados “OR” y el segundo con el objetivo de agrupar los conjuntos propuestos a los términos claves “AND”. En el mismo sentido, se lleva a cabo una limpieza a la ecuación de búsqueda ejecutando algunos criterios de inclusión de interés con la ayuda de los filtros brindados por Scopus. En la tabla 3 a continuación se presenta la ecuación de búsqueda propuesta para el proyecto de investigación.

Tabla 3.

Ecuación de búsqueda

Fecha	Ecuación	Resultado
Diciembre 27, 2021	TITLE-ABS-KEY ((“fuzzy clustering” OR “clustering” OR “k-means clustering” AND “machine learning” OR “unsupervised learning”) AND (Productivity OR development OR competitive OR evolution) AND (workforce OR manpower OR “human resources” OR manufacturing OR production OR manufacture))	190

6.2 Revisión de literatura preliminar

Machine learning se encarga de la construcción de modelos y/o sistemas capaces de aprender de una serie de datos de entrada con el objetivo de encontrar predicciones precisas o acertadas (Atiku & Obagbuwa, 2021). La presente revisión de literatura pretende indagar sobre investigaciones relacionadas con aplicaciones de algoritmos de predicción en diferentes campos de aplicación que pueden ser de gran utilidad para la comprensión del comportamiento de la industria de cuero calzado y marroquinería colombiana en cuanto a la influencia de factores como mano de obra y producción.

Entender, interpretar e identificar patrones para predecir el comportamiento del sector manufacturero resulta ser una tarea compleja de realizar a través de técnicas sencillas, de este modo se crea el desarrollo de nuevas investigaciones motivadas en encontrar soluciones eficientes a través de técnicas avanzadas como la manipulación automática de bases de datos a través de la minería de datos resulta ser una solución óptima.

La minería de datos también caracterizada por minería de conocimiento o extracción de conocimiento cuenta con la habilidad de extraer información de gran relevancia de extensas bases de datos sobre situaciones reales de sectores o empresas específicas (Wang & Wang, 2021). A su vez, junto con el desarrollo de las economías, el uso de algoritmos de aprendizaje automático para analizar y estudiar datos contribuye al mejoramiento de las operaciones empresariales de manera eficiente (Xie, 2020). En este orden de ideas, para algunas empresas manufactureras el aprendizaje automático mejora significativamente el desarrollo de las organizaciones resolviendo con mayor precisión problemas del sistema de recursos humanos y producción. Un ejemplo sencillo de lo

anterior resulta ser la dificultad en contratación de personal adecuado y las fallas en producción causadas por la falta de mantenimiento en maquinaria.

De este modo, diversos contenidos relacionados con IA desde el entendimiento del aprendizaje automático se encargarán de brindar una contextualización más acertada sobre las temáticas de interés de diferentes investigadores. Desde la ingeniería, IA se caracteriza por la experiencia de aprendizaje con respecto a una tarea proporcionada y el rendimiento de esta. De igual forma, es una realidad el crecimiento significativo del volumen de datos disponible, por lo que ha surgido la necesidad de dividir el machine learning (ML) en diferentes ejes fundamentales que cada vez más atraen la atención de la industria y la academia, tales como: aprendizaje no supervisado, aprendizaje supervisado, aprendizaje por refuerzo, entre otros para un excelente desarrollo de la ciencia de la computación y a su vez de la humanidad (Diaz Ramírez, 2021).

En el aprendizaje supervisado algunos autores han demostrado gran interés, Hussain y colaboradores (2021) presentaron un estudio de la correlación entre la experiencia laboral de las organizaciones y su desarrollo con el objetivo de predecir el potencial de los empleados a través de datos proporcionados por una empresa estatal de China con técnicas de clasificación de bosque aleatorio mejorada y aplicación de muestreo aleatorio con el propósito de mejorar las funciones implementadas. En el área de producción, Huertas (2020) plantea una revisión de detección de anomalías potenciales de maquinaria, pronosticando el tiempo de falla entre operaciones y minimizando la participación de recursos humanos en el proceso de diagnóstico y mantenimiento por medio de la automatización o programación, el trabajo que se propone abarca temáticas interesantes relacionadas con clasificaciones, regresiones y redes neuronales. De igual forma, Zhang y colaboradores (Zhang et al., 2021) trabajaron en la detección de defectos en una planta procesadora de peras bajo un modelo de red neuronal artificial supervisada para detectar defectos

en producto trayendo beneficios en disminución de costos a operaciones encargadas de supervisar calidad en producto terminado.

Por otro lado, en materia de aprendizaje no supervisado, el cual es el tema de interés principal en esta investigación se presentan distintos investigadores en diversas áreas de estudio en los últimos años. En cuanto a mano de obra Wu y colaboradores plantean el estudio de conceptos, procesos y algoritmos básicos de la minería de datos (2021) asociación basada en las agrupaciones rápidas en clústers para la asignación de recursos humanos bajo el concepto de algoritmos de minería de datos. Igualmente, **proponen otros autores que han realizado investigaciones en relación con la minería de datos tales como X. Chen & Wang** (Xiaofan & Fengbin, 2010) que presenta un algoritmo de clustering K-Means mejorado y su respectivo análisis para los sistemas de gestión de personal en las organizaciones. Del mismo modo, exponen a Lin (2020) como una introducción en la selección y retención de talentos corporativos bajo la aplicación de árboles de decisiones combinada con una investigación empírica con el fin de optimizar la gestión de recursos humanos en las empresas.

Así mismo, en la actualidad, con el fácil acceso a la información y el gran avance de las organizaciones, se han originado grandes volúmenes de datos que requieren de una inmediata atención para la correcta gestión y control de las industrias. Para ejemplo de lo anterior, el aprendizaje por experiencia en el departamento de recursos humanos proporciona diferentes estudios de interés dentro de los que se puede destacar la predicción de rotación del personal a partir de una red neuronal artificial, la búsqueda de nuevos candidatos basados en el conocimiento, la programación de tareas del personal basado en algoritmo genético y el análisis de las emociones de recursos humanos (Zhu, 2021). Así mismo, este tipo de aplicaciones dentro de las empresas es considerado un factor diferenciador para la toma de decisiones, la productividad, control de calidad

de productos y/o servicios, convirtiéndolas cada vez más competitivas y eficientes (Díaz Ramírez, 2021).

En el mismo orden de ideas, W. Chen & Panahi (2017) presentan la combinación de algunos métodos y principios de la minería de datos con la inteligencia artificial para diseñar un modelo híbrido entre **árboles de decisiones cruzadas** y árboles de decisiones cruzadas múltiples para el diagnóstico y detección de fallas en el área de recursos humanos. Liu y colaboradores (2021), se encargaron de estudiar el impacto de la experiencia laboral en el desarrollo de los empleados con métodos basados en aprendizaje no supervisado en donde se enfocaron en dividir en conjuntos la información proporcionada por recursos humanos en diferentes estados por medio de k-means.

La gestión del talento humano se considera de suma importancia en el momento de adquirir ventaja competitiva frente a la actividad emergente e innovación del mercado, por lo que resulta de gran beneficio la implementación de software o tecnologías para el mejoramiento de las operaciones, Gurbuz & Mert, (2011) presentan una nueva investigación en la que pretenden estudiar los conceptos, procesos y algunos algoritmos de aprendizaje no supervisado de agrupación en clústeres difusos para la asignación de recursos humanos con el objetivo de descubrir los tipos de talentos dentro de las empresas como aporte a la toma de decisiones organizacionales en referencia a la mano de obra que las conforman. Dentro de este documento se explican las diferentes herramientas de datos utilizadas como el análisis de conglomerados, reglas de asociación y evaluación integral difusa para el modelamiento y parametrización del juicio subjetivo del área directiva.

Zhu (2021) expone un método de agrupamiento difuso basado en un algoritmo genético adaptativo con el objetivo de evaluar el desempeño de los empleados en la gestión de los recursos

humanos de empresas nuevas. Del mismo modo, Atiku & Obagbuwa (2021) destacan la labor de la minería de datos en la extracción de patrones interesantes para la tendencia futura y el comportamiento del sistema, por lo que presentan un estudio con la implementación de ocho algoritmos para predecir la influencia del desempeño de los bancos comerciales en Nigeria en función de variables como las habilidades, actitudes o comportamiento de los empleados. Este estudio surge de la necesidad de obtener ventaja competitiva sostenible a través del conocimiento en un entorno empresarial cambiante y competitivo que requiere de desarrollar el potencial interno a partir de la gestión de la fuerza laboral.

En segundo lugar, dentro de la investigación que se desea realizar, también se tiene en cuenta a estudiar el factor producción dado que la IA ha ganado mayor participación en la mejora de la productividad industrial desde el área de producción. Las herramientas y técnicas avanzadas de análisis de datos suelen ser apropiadas dentro de las empresas para obtener mayor beneficio. Particularmente, estas herramientas brindan predicción de rendimiento a partir de datos organizacionales que contribuyen a la mejora de la eficiencia operativa y minimizan algunos costos de producción.

Por medio de esta revisión literaria se logra identificar que muchos autores en los últimos años han trabajado en la optimización de producción de las organizaciones. A causa de que el aprendizaje no supervisado posibilita la oportunidad de interpretar la realidad a partir de unos datos de entrada y tomar decisiones acertadas para el crecimiento empresarial estas aplicaciones han incursionado en la industria manufacturera. Dicho interés de los investigadores en producción abarca diferentes temas de discusión que comúnmente se generalizan en detección de fallas o anomalías, prevención de mantenimiento, mantenimiento correctivo, predicciones de demanda,

tiempos promedios en producción o demanda, pronósticos de utilización de materia prima, caracterización de clientes importantes y demás investigaciones.

En cuanto a labores de producción, se realizan actividades que en la mayoría de las industrias aún dependen de mano de obra, largas y exhaustivas horas de trabajos, altos costos y baja eficiencia. Razón por la cual la IA ha proporcionado ventajas significativas a la producción, inspirados en esto, Amany et al (2022) ilustran un nuevo modelo de simulación que proporciona una visión integradas de varios aspectos del sistema de mantenimiento en donde se introduce la adquisición de activos, la planificación de la fuerza laboral dedicada a mantenimiento y la programación de actividades de mantenimiento preventivos como una simulación que intenta imitar el proceso de mantenimiento de activos por medio de un algoritmos de evolución diferencial mejorado con métodos de agrupamiento K-Means.

Mantenimiento en producción resulta ser una inversión adecuada dado que previene y evita pérdidas de producción a causa de retraso o disponibilidad de maquinaria. Mencionado lo anterior, una actividad relevante en el proceso de mantenimiento preventivo o correctivo es establecer los estados funcionales de todo el sistema productivo. Es así como Ramírez y colaboradores (2017) exponen el diseño de un clasificador difuso a partir de datos obtenidos en los procesos de monitoreo a variables críticas para la estimación de los estados funcionales útiles de un sistema de fabricación de aire medicinal; esta clasificación resulta ser una alternativa eficiente a la ingeniería de mantenimiento a partir del diagnóstico en la programación de intervenciones a maquinaria.

Generalmente la implementación de herramientas de predicción de rendimiento en complemento a los datos organizacionales de producción aporta en gran medida la eficiencia operativa y la disminución de costos de producción. La anterior de puede evidenciar con Jiang y colaboradores (2020) al presentar un estudio de predicción del rendimiento de pruebas finales en

la fabricación de semiconductores bajo la implementación de técnicas de aprendizaje automático. Asimismo, Meiners y colaboradores (2019) proponen un método de visualización sencilla y holística de las posibles interrelaciones existentes en la cadena de procesos mediante métodos de agrupación que amplían el conocimiento de los procesos y mejoran la producción en el largo plazo.

En resumen, realizar diagnósticos y comparaciones partiendo de la clasificación resulta útil en el momento de estudiar el comportamiento y la influencia de factores determinantes a través del tiempo del sector. La lógica difusa y nítida con ayuda de la minería de datos dan creación a los clasificadores, combinación que hace posible así mismo la clasificación de patrones. De este modo, estos conjuntos realizan interconexión entre variables cualitativas, datos numéricos, entradas y salida del sistema que funcionan acertadamente bajo metodologías de técnicas de clustering.

Por otro lado, en la literatura estudiada se identifica que la mayoría de los investigadores han centrado sus estudios en el campo de la influencia de la mano de obra desde diferentes perspectivas y por el contrario en relación con la producción han sido temas más delimitados a los procesos internos de las empresas y no en relación con el avance del sector. Mencionado lo anterior, la presente investigación resulta conveniente ya que abordara temas que han sido poco estudiados y que podrían demostrar importancia significativa a la evolución de las empresas fabriles dedicadas a actividades relacionadas con cuero, calzado y marroquinería en el país.

6.3 Identificación de variables

En esta sección de revisión de literatura, se realiza una investigación sobre las variables que han sido estudiadas dentro de las temáticas de interés de diversos autores. Por medio del análisis preliminar de literatura se identifican variables que se encuentran directamente

relacionadas con el estudio de la influencia de la mano de obra y la producción en diferentes **cambios de acción**.

En relación con la influencia de la mano de obra y la producción de las organizaciones, diferentes autores han expuesto su interés en el estudio en estos dos factores por aportar al desarrollo, productividad y sostenibilidad de las empresas. De igual modo, el DANE dentro de su investigación para el estudio del sector fabril colombiano abarca otras variables en función de los factores previamente mencionados que se proponen para esta investigación a razón de que resultan interesantes para la temática en estudio.

Presentado lo anterior, Yamila Catela y colaboradores (2012) en su interés de analizar la evolución de la productividad y heterogeneidad estructural de las empresas manufactureras de Brasil proponen un algoritmo K-Means con el objetivo de identificar algunos conglomerados existentes. De lo anterior, se identificaron variables como aprendizaje acumulativo, en donde se estudia la influencia de la experiencia y la acumulación de conocimiento de los empleados al igual que la capacidad de innovación medida por la cantidad de empleados innovadores, el nivel de escolaridad de estos y la cantidad de empleados activos dentro de las organizaciones. Por otro lado, en cuanto a producción destacan la importancia del estudio de variables como la innovación en productos proporcionados por la empresa, la participación de las exportaciones en las ventas totales de la empresa y a su vez la participación de la empresa en los ingresos totales del sector en el que realiza sus actividades; del mismo modo, se plantea la importancia de las inversiones realizadas por la empresa a fondos de investigación y desarrollo (I + D) e inversiones a maquinaria y equipos novedosos en el mercado.

De igual forma, Akl y colaboradores (2022) motivados en proporcionar diferentes **puntos de vista** del sistema de mantenimiento en donde se incluye toda adquisición de activos,

planificación de recursos humanos encargados de mantenimiento y algunas actividades de mantenimiento preventivo; mediante un algoritmo de evolución diferencial mejorado con técnicas de aprendizaje automático por agrupamiento de K-Means adoptaron variables como el número total de técnicos contratos, número de técnicos proporcionado a nivel de competencia y número total de técnicos separados del nivel de competencia.

Para las empresas fabriles dedicadas al cuero, calzado y marroquinería el análisis de la mano de obra representa eficiencia en sus áreas productivas, razón por la cual (Barrientos et al., 2007) trabajaron en la caracterización de la eficiencia de un conjunto de empresas del sector en diferentes ciudades de Colombia. Lo anterior, con el objetivo de estipular la incidencia de la mano de obra en el área productiva a través de variables como la cantidad de empleados, innovación en empleados y durabilidad de estos dentro de las empresas y demás variables relacionadas directamente con el área productiva como la inversión a maquinaria, cantidad de producción, innovación al I + D, ubicación de la empresa (plantas) y la producción de la industria tal como insumos, integración de actividades, eficiencia, innovación y exportaciones.

En el país la asociatividad de las empresas que conforman esta industria ha jugado un papel importante en el crecimiento del sector, por tal motivo Calderón & Tomas (2018) y González (2017) han estudiado la dinámica del sector en ciudades como Bucaramanga y Bogotá ya que son algunas de las ubicaciones con mayor asociatividad de empresas. De estas investigaciones, se concluye que el estudio de variables como la calidad y cantidad de mano de obra, el nivel de exportaciones y producción, la empleabilidad, el crecimiento de la industria manufacturera, la capacidad de producción y la inversión son determinantes en el desarrollo significativo del sector.

El presente trabajo aborda la comprensión de la problemática usando fuentes secundarias mediante la recopilación de información brindada por el DANE. La información se obtuvo de la

encuesta anual manufacturera la cual dispone de una información anual sobre los diferentes comportamientos, cambios y evolución económica del sector manufacturero de Colombia (Maldonado et al., 2009). A partir de un análisis cuidadoso se extraen una serie de variables que según los autores pueden ayudar a entender el comportamientos de los sectores con base a variables como el personal ocupado, producción bruta, empleo, remuneración con respecto a sueldos, salarios y prestaciones, inversiones en los activos fijos, número de establecimientos, valor agregado, costos y gastos de producción, gastos de administración y ventas, consumo de materia prima, inventarios, costos laborales, productividad, coeficientes técnicos de producción, entre otras. Las anteriores variables se encuentran descritas en el *apéndice 3* y cabe resaltar que pueden estar sujetas a modificaciones, puesto que al realizar el proceso de la implementación de algoritmos se puede considerar incluir o eliminar algunas variables.

7. Preprocesamiento y análisis de datos

Para asegurar que los algoritmos de Machine Learning propuestos cumplan con los objetivos de la presente investigación, se establece un preprocesamiento de datos para limpiar la estructura y calidad de estos con el propósito de obtener una sólida base de datos que genere óptimas agrupaciones y buenos resultados para ejecutar un correcto análisis. Adicionalmente, en este capítulo se lleva a cabo un primer acercamiento a los patrones que presenta el sector de cuero, calzado y marroquinería por medio de un análisis descriptivo.

A continuación, se describen las fases implementadas para realizar el preprocesamiento y análisis de datos:

7.1 Preprocesamiento de datos

El preprocesamiento de datos es una etapa fundamental previa a la transformación de los datos, se considera un paso vital durante el proceso de minería de datos puesto que se encarga de transformar los datos brutos en datos que incluyan formatos que sean más sencillos de manipular. Normalmente, los datos son recolectados y se presentan en microdatos, la información que se tiene de estos frecuentemente no está limpia, un ejemplo de esto es la existencia de celdas que no registran datos, algunos datos presentan inconsistencias que resultan ser causales de ruido dentro del ML o en otras ocasiones los datos contienen errores o valores atípicos. Si a los datos no se les realiza el respectivo preprocesamiento todas las posibles inconsistencias en la base de datos permanecerán y la probabilidad de éxito en la calidad de la minería de datos disminuirá significativamente (POWERDATA, 2016).

Para realizar un buen preprocesamiento de datos se requiere realizar actividades como: rellenar aquellas celdas que presenten valores faltantes, estudiar y eliminar los datos que pueden generar ruido en la información recolectada, revisar y corregir las inconsistencias que eviten el buen manejo de los datos y ajustar aquellas observaciones que generen redundancia.

En este trabajo de investigación se analizan los microdatos brindados por el Departamento Administrativo Nacional de estadística (DANE) en la ventana de tiempo 2015 - 2019 referente a las empresas que conforman el sector industrial de cuero, calzado y marroquinería en el territorio colombiano; los cuales deben ser preprocesados para ser empleados en los algoritmos de clustering

nítido y difuso. Inicialmente, se revisa y se estudia la estructura de los datos con la finalidad de empezar a limpiar y organizar los datos que se requieren para la implementación del clustering.

En este preprocesamiento de datos se realiza lo siguiente:

7.1.1. Eliminación de variables y grupos industriales.

La eliminación de variables hace parte del proceso de limpieza de datos la cual es sumamente importante para asegurar la integridad de los datos. Cuando la información de la base de datos es confiable permite tomar decisiones o conclusiones certeras. En el paso inicial se realiza la eliminación manual de aquellas empresas que no pertenecen al sector industrial de cuero, calzado y marroquinería.

La encuesta anual manufacturera es un instrumento que recopila información básica del sector fabril colombiano a través de una clasificación de actividades económicas por proceso productivo (CIIU), cuenta con información de todos los establecimientos dedicados a actividades manufactureras en el territorio nacional. En la presente investigación se requiere de información netamente del sector de cuero, calzado y marroquinería, es por esta razón que se realiza una selección manual únicamente de las empresas con los siguientes CIIU: *1921* - fabricación de calzado de cuero y piel, con cualquier tipo de suela, excepto el calzado deportivo, *1922* - fabricación de calzado con materiales textiles, con cualquier tipo de suela, excepto calzado deportivo, *192* – fabricación de calzado y *1910* - curtido y abono de cueros.

7.1.2. Valores faltantes.

El dato o valor faltante (N/A, 0) se produce cuando la variable no se guarda o no se almacena alguna información correspondiente siendo un problema frecuente en los estudios. Los datos faltantes se determinan como aquellos valores que no están disponibles pero que aportarían información útil para el análisis de los resultados.

Existe variabilidad de tipos de datos faltantes y demasiadas razones por las cuales estos ocurren, en esta situación principalmente se debe decidir si esta ausencia es aleatoria, en otras palabras, si afecta por igual a todas las observaciones de la base de datos o si por el contrario tiene razones puntuales que pueden generar sesgos que invaliden la información (Dagnino S, n.d.).

De acuerdo con los microdatos obtenidos por el DANE, se encuentra la existencia de casillas cuyo valor es cero los cuales pueden interpretarse como valores faltantes, pero, luego de una exhausta investigación se conoce que el DANE cuenta con una serie de análisis y revisión rigurosa de la consistencia de estos que aseguran la calidad de la información que proporcionan en su información recolectada. A razón de lo anterior, se determina que la base de datos no contiene observaciones faltantes y no es modificada.

7.1.3. Imputación de datos.

La imputación de datos es un proceso que cambia los datos faltantes en datos simulados con la finalidad de minimizar de la manera más adecuada la influencia en el resultado que se espera. Este proceso suele generar una o más versiones de los datos simulados los cuales se basan en la cantidad de iteraciones que se especifiquen.

Cada una de estas iteraciones produce una versión de los datos las cuales deben analizarse y ser estimadas dependiendo del interés que se quiera de los datos. Finalmente, el total de estas iteraciones son agrupadas en un solo conjunto que permite la comprobación del dato imputado con relación al grupo de datos totales (Araneda, 2021).

De acuerdo con los microdatos en estudio, se encuentra una cantidad considerable de observaciones con valores en cero las cuales proporcionan información importante, por esta razón, estos datos se mantienen con el objetivo de no perder información que contribuya a la identificación de comportamientos del sector a través de los años. Según lo anterior, no se realiza imputación de datos.

7.1.4. Detección y eliminación de datos atípicos.

Los datos atípicos son observaciones que contienen valores diferentes al resto de datos de la muestra. Este tipo de datos causan inconvenientes en la interpretación de la base de datos debido a que distorsionan el análisis de los datos, en consecuencia, es necesario realizar una detección de datos atípicos para limpiar la base de datos y a su vez eliminar aquellos datos que no aportan o desvían información relevante al estudio.

En este estudio, se realiza una revisión rigurosa y se logra apreciar que 18 establecimientos no brindan información completa de las variables en estudio y que presentan valores atípicos puesto que contienen valores que presentan inconsistencia entre las variables. Por esta razón, se eliminan estas observaciones; así, la base de datos preliminar a trabajar contará con 36 variables que constan de 510 observaciones.

7.1.5. Agrupación (categorización) de variables.

El principal objetivo de realizar una agrupación de datos en intervalos es facilitar el análisis y la interpretación de información. Dada la complejidad del manejo de los microdatos, se toma la decisión de categorizar las variables buscando una aproximación para realizar un análisis descriptivo que proporcione información relevante sobre el comportamiento de las variables con el objetivo de obtener un primer acercamiento y conocimiento sobre el comportamiento del sector. La categorización de las variables se encuentra en el apéndice 4.

Para realizar los respectivos análisis, inicialmente se agrupan las variables en 5 grupos importantes como se detallan a continuación:

Tabla 4.

Agrupación de variables

Grupo de variables	Variables
Mano de obra	<ul style="list-style-type: none"> - Prestaciones de permanente (PRESSPER) - Salario de permanentes (SALARPER) - Prestaciones de permanente y temporal directo (PRESPTYTE) - Prestaciones de permanente y temporal (SALPEYTE) - Personal permanente + propietarios + personal temporal directo + temporal empresas (PERTOTAL)
Producción bruta	<ul style="list-style-type: none"> - Costos y gastos de productos y servicios industriales elaborados por terceros – producción (c3r14c1) - Costos y gastos por servicios contratados con terceros outsourcing – producción (c3r41c1) - Valor de la energía vendida (c5r1c5) - Existencias en productos en proceso – al final del año anterior (c6r2c1) - Existencias en productos en proceso – al final del año actual (c6r2c3) - Producción bruta (PRODBR2)

Grupo de variables	Variables
Consumo intermedio	- Producción industrial (PRODBIND)
	- Valor de las ventas (VALORVEN)
	- Honorarios y servicios técnicos – producción (c3r15c1)
	- Arrendamiento de bienes inmuebles – producción (c3r16c1)
	- Arrendamiento sin opción de compra de maquinaria y equipo – producción (c3r17c1)
	- Mantenimiento, reparaciones, accesorios y repuestos consumidos – producción (c3r23c1)
	- Costos y gastos de transporte de materias primas – producción (c3r42c1)
	- Valor energía comprada (c3r19c3)
	- Consumo de materias (CONSMATE)
	- Otros gastos (CONSIN)
	- Consumo intermedio (CONSIN2)
	- Valor de la materia prima comprada (VALORCOM)
	- Valor compras de materia prima en el exterior (VALORCX)
	Activos fijos
- Valor en libros del año anterior – Edificios y estructuras (c7r5c1)	
- Valor en libros del año anterior – Maquinaria y equipo industrial (c7r6c1)	
- Costos y gastos de producidos o construidos para uso propio – edificios y estructuras (c7c2r4)	
- Costos y gastos de producidos o construidos para uso propio – Maquinaria y equipo industrial (c7c3r4)	
- Valor causado en el año por la maquinaria en montaje – Maquinaria y equipo industrial (c7c3r6)	
- Valor causado por mejoras y reformas a los activos – Edificios y estructuras (c7c2r7)	
- Valor causado por mejoras y reformas a los activos – Maquinaria y equipo industrial (c7c3r7)	
- Valor en libros del año actual – Terrenos (c7r1c6)	
- Valor en libros del año actual – Edificios y estructuras (c7r5c6)	
- Valor en libros del año actual – Maquinaria y equipo (c7r6c6)	
- Existencias materias primas, materiales y empaques – al final del año anterior (c6r1c1)	

Grupo de variables	Variables
	<ul style="list-style-type: none"> - Existencias materias primas, materiales y empaques – al final del año actual (c6r1c3) - Existencias productos en proceso – al final del año anterior (c6r2c1) - Existencias productos en proceso – al final del año actual (c6r2c3) - Existencias productos terminados – al final del año anterior (c6r3c1) - Existencias productos terminados – al final del año actual (c6r3c3) - Existencias productos no fabricados por el establecimiento – al final del año anterior (c6r4c1) - Existencias productos no fabricados por el establecimiento – al final del año actual (c6r4c3)
Inversión bruta	<ul style="list-style-type: none"> - Valor compra de activos – usados – Terrenos (c7c1r3) - Valor compra de activos – usados – Edificios y estructuras (c7c2r3) - Valor compra de activos – usados – Maquinaria y equipo (c7c3r3) - Total, inversiones en activos fijos – Terrenos (c7r1c2) - Total, inversiones en activos fijos – Edificios y estructuras (c7r5c2) - Total, inversiones en activos fijos – Maquinaria y equipo (c7r6c2) - Valor en libro de los activos vendidos – Terrenos (c7c1r12) - Valor en libro de los activos vendidos – Edificios y estructuras (c7c2r12) - Valor en libro de los activos vendidos – Maquinaria y equipo (c7c3r12) - Valor de los activos trasladados en el año a otros establecimientos de la empresa – Edificios y estructuras (c7r17c2) - Valor de los activos trasladados en el año a otros establecimientos de la empresa – Maquinaria y equipo (c7r17c3) - Valor de los activos recibidos en el año por traslado de otros establecimientos de la empresa – Edificios y estructuras (c7r18c2)

Grupo de variables	Variables
	- Valor de los activos recibidos en el año por traslado de otros establecimientos de la empresa – Maquinaria y equipo industrial (c7r18c3)
	- Inversión bruta (INVEBRTA)

Teniendo en cuenta la información anterior, se realiza un estudio de la correlación de cada grupo de variables con la finalidad de determinar que variables aportan información relevante al estudio. A continuación, se determina que variables tienen relación entre sí:

Tabla 5.

Correlación grupo 1 – mano de obra

	PRESSPER	SALARPER	PRESPYTE	SALPEYTE	PERTOTAL
PRESSPER	1.000000	0.962952	0.997924	0.980380	0.593853
SALARPER	0.962952	1.000000	0.961202	0.992902	0.635167
PRESPYTE	0.997924	0.961202	1.000000	0.978569	0.590461
SALPEYTE	0.980380	0.992902	0.978569	1.000000	0.629140
PERTOTAL	0.593853	0.635167	0.590461	0.629140	1.000000

Teniendo en cuenta los datos anteriores, se puede observar que existe una correlación fuerte entre las variables de SALARPER, PRESSPER y SALPEYTE, por ende, no es necesaria la implementación de las tres. Estos datos pueden ser redundantes a la hora de analizarlos (brindan la misma información), a razón de lo anterior, las variables PRESSPER y SALARPER son eliminadas.

Tabla 6.

Correlación grupo 2 – producción bruta

	c3r14c1	c3r41c1	c5r1c5	c6r2c1	c6r2c3	PRODBR2	PRODBIND	VALORVEN	VALAGRI
c3r14c1	1.000000	0.057095	0.010415	0.179728	0.213875	0.063947	0.063947	0.202279	-0.052130
c3r41c1	0.057095	1.000000	0.238216	0.355225	0.299889	0.489374	0.489374	0.545785	0.240187
c5r1c5	0.010415	0.238216	1.000000	0.799323	0.781439	0.577440	0.577440	0.487454	0.242193
c6r2c1	0.179728	0.355225	0.799323	1.000000	0.910854	0.577207	0.577207	0.535557	0.308683
c6r2c3	0.213875	0.299889	0.781439	0.910854	1.000000	0.578938	0.578938	0.540407	0.281036
PRODBR2	0.063947	0.489374	0.577440	0.577207	0.578938	1.000000	1.000000	0.900653	0.380644
PRODBIND	0.063947	0.489374	0.577440	0.577207	0.578938	1.000000	1.000000	0.900653	0.380644
VALORVEN	0.202279	0.545785	0.487454	0.535557	0.540407	0.900653	0.900653	1.000000	0.332875
VALAGRI	-0.052130	0.240187	0.242193	0.308683	0.281036	0.380644	0.380644	0.332875	1.000000

A partir de la tabla anterior, se puede ver que las variables VALORVEN, PRODBIND y PRODBR2 tienen una correlación fuerte respecto a las demás, por esta razón, se determina que la implementación de la variable PRODBIND referente a la producción industrial no aporta nueva información a la base de datos en estudio y se elimina.

Tabla 7.

Correlación grupo 3 – consumo intermedio

	c3r15c1	c3r16c1	c3r17c1	c3r23c1	c3r42c1	c3r19c3	CONSMATE	CONSIN	CONSIN2	VALORCOM	VALORCX	VALAGRI
c3r15c1	1.000000	0.142831	0.342714	0.720186	0.560651	0.581873	0.528014	0.630606	0.526801	0.433128	0.399407	0.339618
c3r16c1	0.142831	1.000000	0.219985	0.068219	0.060384	0.231270	0.302669	0.157726	0.299982	0.258129	0.256890	0.083732
c3r17c1	0.342714	0.219985	1.000000	0.350223	0.169514	0.678214	0.216716	0.250470	0.215145	0.370482	0.282774	0.268883
c3r23c1	0.720186	0.068219	0.350223	1.000000	0.562002	0.588152	0.554545	0.758909	0.551939	0.469758	0.418392	0.331241
c3r42c1	0.560651	0.060384	0.169514	0.562002	1.000000	0.474393	0.241154	0.614818	0.240250	0.127439	0.159008	0.150030
c3r19c3	0.581873	0.231270	0.678214	0.588152	0.474393	1.000000	0.400968	0.632667	0.401149	0.494874	0.382688	0.283563
CONSMATE	0.528014	0.302669	0.216716	0.554545	0.241154	0.400968	1.000000	0.540351	0.997831	0.777155	0.678831	0.352854
CONSIN	0.630606	0.157726	0.250470	0.758909	0.614818	0.632667	0.540351	1.000000	0.542099	0.459940	0.452506	0.275525
CONSIN2	0.526801	0.299982	0.215145	0.551939	0.240250	0.401149	0.997831	0.542099	1.000000	0.775966	0.676659	0.353819
VALORCOM	0.433128	0.258129	0.370482	0.469758	0.127439	0.494874	0.777155	0.459940	0.775966	1.000000	0.341218	0.584825
VALORCX	0.399407	0.256890	0.282774	0.418392	0.159008	0.382688	0.678831	0.452506	0.676659	0.341218	1.000000	-0.165358
VALAGRI	0.339618	0.083732	0.268883	0.331241	0.150030	0.283563	0.352854	0.275525	0.353819	0.584825	-0.165358	1.000000

En este caso existe una correlación casi perfecta con las variables CONSMATE y CONSIN2, por lo que se infiere que el consumo intermedio y el consumo de materias primas brindan la misma información, por consiguiente, se excluye la variable de CONSMATE en el estudio.

Tabla 8.

Correlación grupo 4 – activos fijos

c6r1c1	c6r1c3	c6r2c1	c6r2c3	c6r3c1	c6r3c3	c6r4c1	c6r4c3	c7r1c1	c7r5c1	c7r6c1	c7c2r4	c7c3r4	c7c3r6	c7c2r7	c7c3r7	c7r1c6	c7r5c6	c7r6c6
1.000000	0.824088	0.448839	0.448817	0.313685	0.298525	0.483461	0.386195	0.431311	0.433234	0.442534	0.104634	-0.005207	0.134199	0.157132	0.039224	0.446486	0.452807	0.442534
0.824088	1.000000	0.438590	0.482828	0.310924	0.316558	0.485437	0.483252	0.449893	0.476113	0.453209	0.093192	-0.009332	0.151381	0.225574	0.125501	0.463067	0.499139	0.453209
0.448839	0.438590	1.000000	0.910854	0.687175	0.693228	0.526109	0.468476	0.752556	0.728038	0.793865	-0.026795	0.020674	0.256157	0.245206	0.064538	0.743277	0.729438	0.793865
0.448817	0.482828	0.910854	1.000000	0.672753	0.687616	0.498091	0.511204	0.751099	0.726646	0.795811	-0.027679	0.023494	0.311942	0.126563	0.071359	0.748972	0.728570	0.795811
0.313685	0.310924	0.687175	0.672753	1.000000	0.929087	0.414120	0.389527	0.628389	0.623884	0.738035	-0.028487	0.050414	0.332983	0.241631	0.136032	0.617532	0.609040	0.735020
0.298525	0.316558	0.693228	0.687616	0.929087	1.000000	0.395679	0.379388	0.622680	0.628950	0.716000	-0.025209	0.055511	0.217200	0.230010	0.181842	0.612290	0.614559	0.712921
0.483461	0.485437	0.526109	0.498091	0.414120	0.395679	1.000000	0.813836	0.538040	0.565147	0.480457	0.061791	-0.025743	0.265086	0.131162	0.127152	0.530588	0.561927	0.480457
0.386195	0.483252	0.468476	0.511204	0.389527	0.379388	0.813836	1.000000	0.523538	0.530214	0.458906	0.055637	-0.027180	0.289626	0.132728	0.089730	0.532106	0.526660	0.458906
0.431311	0.449893	0.752556	0.751099	0.628389	0.622680	0.538040	0.523538	1.000000	0.834915	0.864586	0.012932	-0.018315	0.284729	0.249126	0.032117	0.985844	0.830251	0.864586
0.433234	0.476113	0.728038	0.726646	0.623884	0.628950	0.565147	0.530214	0.834915	1.000000	0.809169	-0.003722	-0.035166	0.234440	0.257944	0.176752	0.832735	0.983036	0.812119
0.442534	0.453209	0.793865	0.795811	0.738035	0.716000	0.480457	0.458906	0.864586	0.809169	1.000000	0.003259	0.003379	0.359810	0.259815	0.005086	0.858842	0.807121	0.987944
0.104634	0.093192	-0.026795	-0.027679	-0.028487	-0.025209	0.061791	0.055637	0.012932	-0.003722	0.003259	1.000000	0.128133	-0.003830	0.017250	-0.026017	0.010431	0.068333	0.003259
-0.005207	-0.009332	0.020674	0.023494	0.050414	0.055511	-0.025743	-0.027180	-0.018315	-0.035166	0.003379	0.128133	1.000000	0.058483	0.098248	-0.026980	-0.020680	-0.038347	0.003379
0.134199	0.151381	0.256157	0.311942	0.332983	0.217200	0.265086	0.289626	0.284729	0.234440	0.359810	-0.003830	0.058483	1.000000	-0.043766	-0.035349	0.290825	0.227193	0.359810
0.157132	0.225574	0.245206	0.126563	0.241631	0.230010	0.131162	0.132728	0.249126	0.257944	0.259815	0.017250	0.098248	-0.043766	1.000000	0.034478	0.244232	0.252994	0.259815
0.039224	0.125501	0.064538	0.071359	0.136032	0.181842	0.127152	0.089730	0.032117	0.176752	0.005086	-0.026017	-0.026980	-0.035349	0.034478	1.000000	0.028130	0.171207	0.005086
0.446486	0.463067	0.743277	0.748972	0.617532	0.612290	0.530588	0.532106	0.985844	0.832735	0.858842	0.010431	-0.020680	0.290825	0.244232	0.028130	1.000000	0.827633	0.858842
0.452807	0.499139	0.729438	0.728570	0.609040	0.614559	0.561927	0.526660	0.830251	0.983036	0.807121	0.068333	-0.038347	0.227193	0.252994	0.171207	0.827633	1.000000	0.810062
0.442534	0.453209	0.793865	0.795811	0.735020	0.712921	0.480457	0.458906	0.864586	0.812119	0.987944	0.003259	0.003379	0.359810	0.259815	0.005086	0.858842	0.810062	1.000000

Dentro de esta agrupación de variables no existe correlación entre las mismas, por ende, no se excluye ninguna variable de esta agrupación.

Tabla 9.

Correlación grupo 5 – Inversión bruta

	c7c1r3	c7c2r3	c7c3r3	c7r1c2	c7r5c2	c7r6c2	c7c1r12	c7c2r12	c7c3r12	c7r17c2	c7r17c3	c7r18c2	c7r18c3	INVEBRTA
c7c1r3	1.000000	0.291474	0.176112	0.938159	0.220328	0.242852	0.168546	0.038735	0.096540	0.168931	-0.053007	0.101976	0.253158	0.252069
c7c2r3	0.291474	1.000000	0.546358	0.270703	0.124079	0.019019	-0.043556	0.048078	0.077212	0.058110	-0.040225	0.074568	0.018453	0.044021
c7c3r3	0.176112	0.546358	1.000000	0.162084	0.051077	0.052295	-0.026388	0.017718	0.007695	-0.004120	-0.037633	0.005352	-0.029451	0.038319
c7r1c2	0.938159	0.270703	0.162084	1.000000	0.204713	0.238138	0.182667	0.082349	0.082330	0.149613	-0.028765	0.087085	0.226671	0.247384
c7r5c2	0.220328	0.124079	0.051077	0.204713	1.000000	0.544572	0.138298	0.201633	0.347925	0.268693	0.255135	0.364626	0.412617	0.544362
c7r6c2	0.242852	0.019019	0.052295	0.238138	0.544572	1.000000	0.113630	0.088581	0.195249	0.390687	0.475794	0.502789	0.597165	0.751469
c7c1r12	0.168546	-0.043556	-0.026388	0.182667	0.138298	0.113630	1.000000	0.365113	0.195033	-0.026900	0.058744	0.071777	0.107745	-0.061552
c7c2r12	0.038735	0.048078	0.017718	0.082349	0.201633	0.088581	0.365113	1.000000	0.614053	-0.017976	0.077401	0.088836	0.014677	-0.178033
c7c3r12	0.096540	0.077212	0.007695	0.082330	0.347925	0.195249	0.195033	0.614053	1.000000	-0.052247	-0.002134	0.090293	-0.025355	0.020343
c7r17c2	0.168931	0.058110	-0.004120	0.149613	0.268693	0.390687	-0.026900	-0.017976	-0.052247	1.000000	0.142242	0.166043	0.778588	0.311444
c7r17c3	-0.053007	-0.040225	-0.037633	-0.028765	0.255135	0.475794	0.058744	0.077401	-0.002134	0.142242	1.000000	0.781454	0.213470	0.308153
c7r18c2	0.101976	0.074568	0.005352	0.087085	0.364626	0.502789	0.071777	0.088836	0.090293	0.166043	0.781454	1.000000	0.360890	0.386797
c7r18c3	0.253158	0.018453	-0.029451	0.226671	0.412617	0.597165	0.107745	0.014677	-0.025355	0.778588	0.213470	0.360890	1.000000	0.521879
INVEBRTA	0.252069	0.044021	0.038319	0.247384	0.544362	0.751469	-0.061552	-0.178033	0.020343	0.311444	0.308153	0.386797	0.521879	1.000000

En este grupo se presenta la misma situación del grupo anterior, debido a que no existe una correlación fuerte entre variables no se hace necesario eliminar alguna.

Después de realizar el análisis de correlación, en total se eliminan dos variables pertenecientes al grupo de personal ocupado y una variable perteneciente al grupo de producción bruta.

7.1.6. Normalización.

Las técnicas de normalización de datos se implementan para definir la base de datos de manera más natural y limpia mediante la reducción de tamaño simplificando su estructura con el fin de que estos sean más ~~sencillos~~ sencillos de contrastar, localizar y analizar. En relación con lo anterior para diferentes estimadores de aprendizaje automático la normalización se convierte en una fase de preprocesamiento crucial ya que muchas de estas técnicas asumen que las observaciones se centran en cero y cuentan con varianza similar, de este modo, la existencia de

características con varianzas muy diferentes podrían predominar en la función objetivo obteniendo resultados incorrectos (*SCIKIT LEARN*, n.d.-a).

En algoritmos de agrupamiento, la normalización es fundamental debido que estas técnicas comprenden métricas de distancias en su ejecución. Para la normalización de la información que presentan diferentes empresas del sector de cuero, calzado y marroquinería se realiza por medio del paquete preprocessing de la librería sklearn de Python.

7.2 Análisis descriptivo

El análisis descriptivo no se encarga de realizar predicciones del futuro, pero si ayuda a filtrar algunos datos que no aportan información al estudio que se requiere y/o realizar inferencias sobre el comportamiento del sector. Para realizar un buen análisis descriptivo se requiere una recopilación de datos, seguidamente realizar una limpieza de los datos recolectados y finalmente aplicar métodos estadísticos para extraer la información deseada. Los métodos que se escojan dependen de los datos que se estén manejando y de lo que se quiera determinar (*Tudashboard*, 2020).

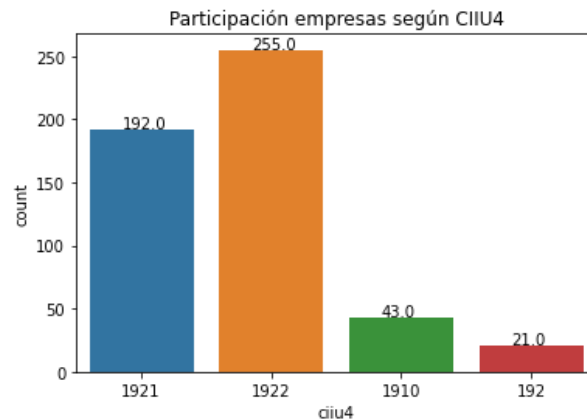
Mencionado lo anterior, en esta investigación resulta necesaria la ejecución de un análisis descriptivo con los datos anteriormente preprocesados y categorizados con la finalidad de obtener un análisis diagnóstico donde se logre observar la tendencia del sector de cuero, calzado y marroquinería colombiano a través de los años. Este capítulo recopila información sobre la variabilidad, la correlación, la dependencia de variables y los distintos comportamientos que presentan los datos para saber cómo se ha comportado el sector.

Del análisis descriptivo se destacan los siguientes aspectos:

De la figura anterior, se puede ver que la mayor empleabilidad del sector se genera en el departamento de Bogotá (azul oscuro – 11), Valle del Cauca (rojo – 76), Antioquia (verde oscuro – 5), Cundinamarca (café – 25) y Atlántico (gris – 8), respectivamente, departamentos que a través del tiempo han demostrado tener los clústers más importantes para el país. Por otra parte, de estos departamentos se resalta que en Santander (naranja - 68) se encuentra una empresa grande que genera una empleabilidad baja.

Figura 5.

Participación de los establecimientos en la encuesta según la actividad que realiza

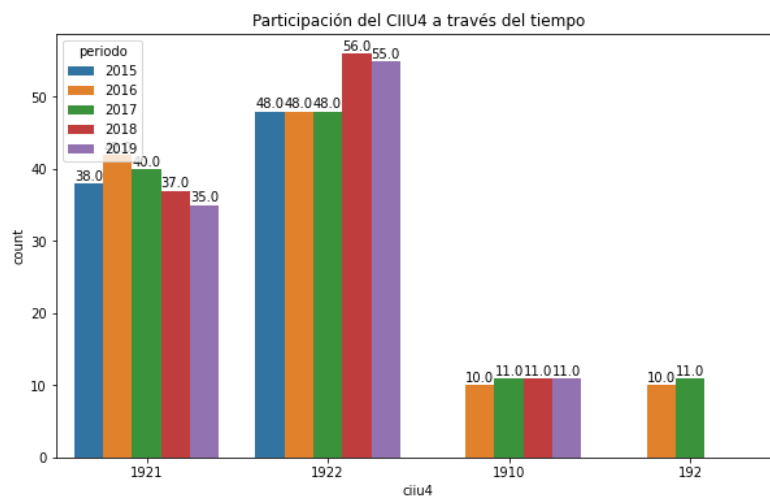


Con base a la información suministrada por la figura 5, la clasificación de las actividades económicas que realizan las empresas en este sector y las cuales tienen mayor participación con 255 organizaciones son las que se encargan de fabricar calzado con materiales textiles con cualquier tipo de suela excepto calzado deportivo (ciiu: 1922), estas empresas realizan compras de materia prima nacional y extranjera entre el rango de \$10.000.000 a 35.000.000. Seguidamente las empresas que fabrican calzado de cuero y piel con cualquier tipo de suela excepto el calzado

deportivo (ciiu: 1921) con una cantidad de 192 organizaciones, estas hacen compras de materia prima nacional y extranjera hasta un valor máximo de \$10.000.000. Las empresas que realizan curtido y abono de cueros en general (ciiu: 1910) presentan una participación baja puesto que la conforman 43 organizaciones en el sector y sus compras en materia prima también suelen ser bajas con un valor límite de 10.000.000. Finalmente, la participación de las empresas que realizan fabricación de calzado en general (ciiu: 192) es muy baja respecto a las demás clasificaciones puesto que solo participan en la encuesta anual manufacturera 21 empresas.

Figura 6.

Participación de la actividad económica a través del tiempo



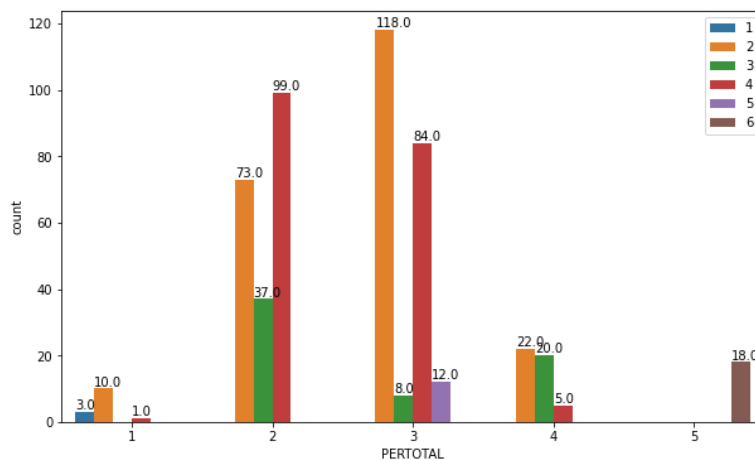
Por medio de la figura 6 se logra apreciar que la participación de la clasificación de las empresas respecto a fabricación de calzado de cuero y piel (ciiu: 1921) han tenido una participación alta en el año 2016 con 44 empresas. De acuerdo con la participación de fabricación de calzado de materiales textiles (ciiu: 1922) ha sido muy relevante puesto que de 2015 a 2017 la

participación fue alta con 48 organizaciones y se mantuvo, en 2018 aumentó significativamente la participación con 8 empresas más y en 2019 esta cantidad disminuyó con una empresa menos.

Respecto a la clasificación de curtido y abono de cueros (ciiu: 1910) la participación es baja respecto a las otras clasificaciones con una cantidad de 11 empresas y en la cual no presentó participación en el año 2015 y en relación con la fabricación de calzado en general (ciiu: 192) solo se reporta una participación baja de diez empresas en el año 2016 y 11 empresas en el año 2017.

Figura 7.

Personal ocupado según las ventas totales



Teniendo en cuenta la información brindada por la figura 7 se puede ver que las pequeñas empresas (PERTOTAL – 2) y microempresas (PERTOTAL – 1) que cuentan pocos empleados (entre 1 a 50 empleados) generan ventas totales bajas con un valor máximo de \$40.000.000, y caso contrario, las grandes (PERTOTAL – 5) algunas de ellas con más de 201 colaboradores generan ventas totales altas. De lo anterior, surge el siguiente dilema ¿Cuándo las empresas generan mayor empleabilidad aumentan su productividad en ganancias?

Finalmente, cabe resaltar que en este caso no se tiene el mismo número de empresas encuestadas para cada categoría, por consiguiente, la mayoría de los datos se concentran en las

microempresas y pequeñas empresas debido a que estas predominan. El sector manufacturero de cuero, calzado y marroquinería no abarca mayor cantidad de empresas grandes ya que el país no cuenta con la capacidad de apoyarlas en función de sus requerimientos.

8. Aplicación de algoritmos de agrupamiento

Entre las diferentes técnicas y acciones soportadas en la inteligencia artificial, los algoritmos de agrupación consisten en la clasificación de datos conformando grupos a raíz de una serie de características según su homogeneidad. Los algoritmos de clúster no supervisados trabajan a través de conceptos de proximidad, es decir, busca similitudes y diferencias que descubren la estructura de los datos.

El objetivo de estas técnicas es la búsqueda y análisis de grupos cuya semejanza entre ellos sea muy baja y la similitud entre los datos de cada grupo sea alta, de esta manera se conformarán grupos de alta calidad que brinden la oportunidad de explorar la base de datos.

Esta sección contiene información sobre la ejecución de los algoritmos de agrupamiento nítido y difuso implementados en este proyecto con el objetivo de descubrir patrones en la industria colombiana de cuero, calzado y marroquinería.

El modelado de los datos fue desarrollado utilizando el jupyter de Python y diversos módulos de paquete de sklearn. El código aplicado puede **encontrarse** en la plataforma Github (Benavidez & Jimenez, 2022).

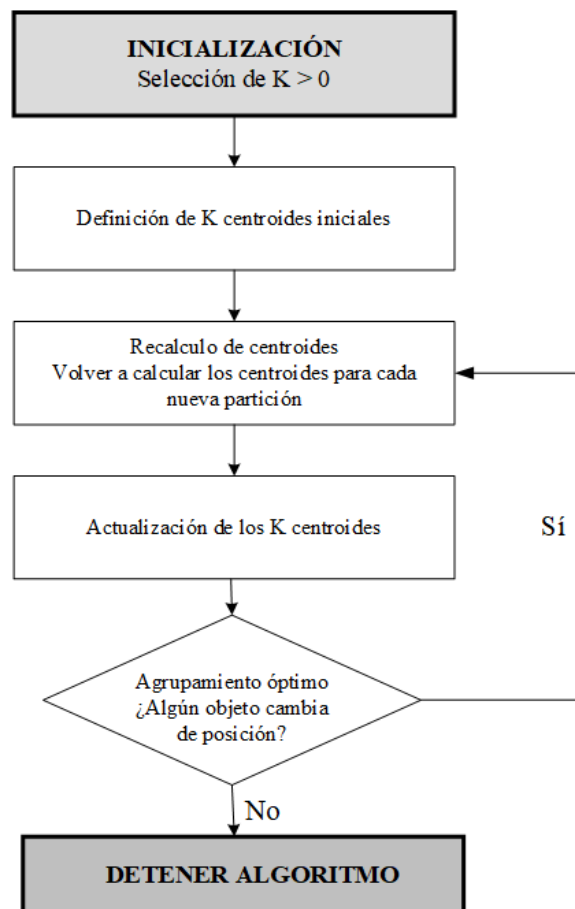
8.1 Algoritmo clustering K-means:

K-Means se encarga de encontrar un punto representativo para cada conjunto, es decir, un centroide con la facultad de caracterizar el clúster al que pertenece. así mismo al ser K-Means un algoritmo caracterizado por ser un modelo de aprendizaje duro, por medio de iteraciones, se estabiliza el punto representativo y cada punto de dato se irá asignado a uno y solo un grupo

La ejecución del algoritmo K-Means iterativamente cuenta con 4 pasos que se describen en el siguiente diagrama de flujo

Figura 8.

Diagrama de flujo algoritmo K-means



Nota. Adaptado de (Cazares Aguirre & Martínez Gómez, 2018)

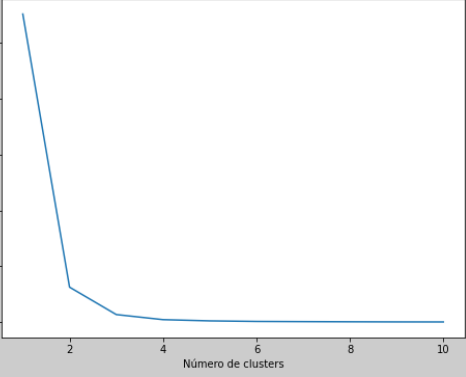
Este algoritmo necesitar tener conocimiento a priori sobre el número de clústers a conformar, en este orden de ideas, dentro de este proyecto se implementa el método del codo “Elbow method” que brinda información sobre los valores de inercia luego de ejecutar el mismo algoritmo con diferente cantidad de centros. La inercia es definida como la suma de las distancias al cuadrado de cada observación a su punto representativo.

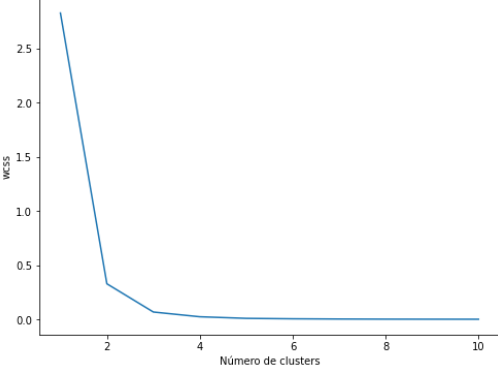
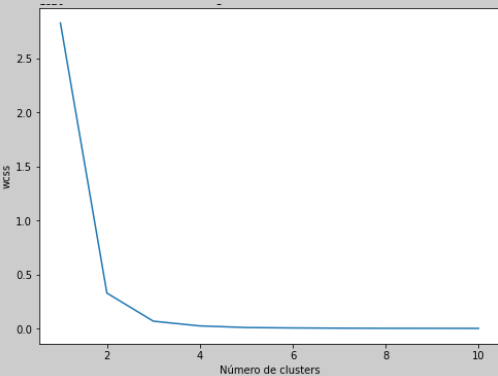
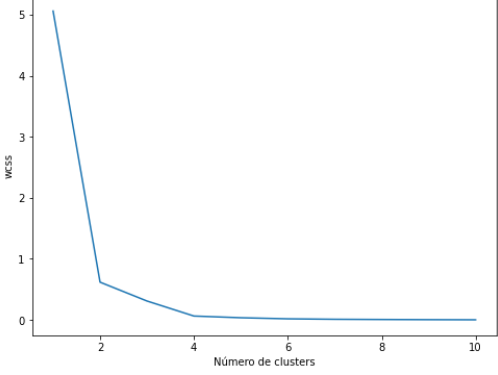
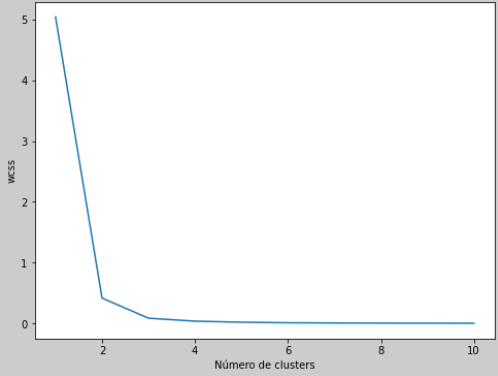
Una vez obtenidos los valores de inercia luego de aplicar el algoritmo de 1 a 11 clústers, se presentan los resultados a través de una gráfica buscando apreciar un cambio lo suficientemente brusco en la inercia que logre asemejar un brazo y su codo, de este modo, el punto donde se identifique lo anterior, infiere la cantidad de clústers que mejor describen los datos.

“Elbow method” encuentra tres clústers que describen de manera acertada los datos del sector en todos los años. A continuación, se presentan los resultados obtenidos.

Tabla 10.

Resultados codo de jambú

Periodo	Codo jambú	Declive de inercia
2015		Tres grupos – 6,768e+18

Periodo	Codo jambú	Declive de inercia
2016		Tres grupos – 6,797e+18
2017		Tres grupos – 2,510e+19
2018		Tres grupos - 3,102e+19
2019		Tres grupos - 8,541e+18

En el mismo sentido, una de las preocupaciones de este trabajo de investigación es encontrar las variables que mejor describen el comportamiento de la industria. Las variables preseleccionadas se segmentan en 5 clases que se describen así: en primer lugar, una clase conformada por variables en referencia a mano de obra dentro la cual guarda información sobre el total de empleados y el salario, la segunda con datos importantes sobre producción bruta contiene variables en relación con las existencias de productos en proceso, el valor de las ventas, entre otros. La tercera clase, consumo intermedio con variables como el arrendamiento, mantenimiento, valor en libros de activos, costos y gastos de producción, existencias de materias primas entre otras, seguidamente la cuarta clase comprende información sobre inversión bruta donde se asigna el valor de compra de activos fijos y el valor en libros de los activos vendidos, finalmente, la quinta clase abarca el valor agregado donde fijan variables con relación a producción bruta y consumo intermedio.

De acuerdo con la taxonomía de las clases expuestas, surge la necesidad de validar las variables con el objetivo de encontrar aquellas que brinden información verdaderamente significativa a la formación de los clústers eliminando el ruido o relevancia entre los datos. Con la ayuda del software Python se buscan las variables idóneas que conforman mejores agrupamientos. En relación con lo anterior, se obtienen los siguientes resultados: la información de los años 2015, 2016, 2017, 2018 y 2019 crea agrupaciones buenas con 33, 22, 21, 33 y 33 variables respectivamente, algunas de estas como producción bruta, consumo de materias, otros gastos, consumo intermedio, salario del personal, inversión bruta, total de personal, valor de la materia prima comprada nacional e internacional, valor de la energía comprada, mantenimiento, honorarios y servicios técnicos, valor de las ventas, existencias de productos terminados, existencias de productos no fabricados por el establecimiento, valor en libros de los activos

vendidos por maquinaria y equipo, valor en libros en terrenos y las existencias de productos en proceso se mantienen en el tiempo.

A continuación, se presentan las variables que contribuyen a crear mejores clústers.

Tabla 11.

Variación de variables a través del tiempo

Periodo	Variables
2015 / 2016 / 2018 / 2019	Valor en libros del año actual – edificios
	Valor en libros del año actual - maquinaria
	Valor de los activos trasladados en el año a otros establecimientos de la empresa - edificios
	Valor de los activos trasladados en el año a otros establecimientos de la empresa - maquinaria
2015 / 2017 / 2018 / 2019	Arrendamiento sin opción de compra de maquinaria
	Costos y gastos de productos y servicios industriales elaborados por terceros - producción
	Costos y gastos por servicios contratados con terceros (outsourcing) - producción
2015 / 2018 / 2019	Costos y gastos de producidos o construidos para uso propio – edificios
	Costos y gastos de producidos o construidos para uso propio – maquinaria
	Existencias materias primas, materiales y empaques - al final del año actual
	Total, inversiones en activos fijos – edificios
	Total, inversiones en activos fijos – maquinaria
	Costos y gastos de transporte de materias primas - producción
	Total, inversiones en activos fijos - terrenos
	Arrendamiento de bienes inmuebles – producción

(Benavidez et al., 2022) abarca la selección de las variables descritas en un fragmento de código por medio de la definición de una función con “k” cantidades de clústers a formar e “i” cantidad de variables que mejor describen los datos, obteniendo resultados que contienen el porcentaje de datos que se asigna a cada conglomerado. Los resultados se presentan en la siguiente tabla.

Tabla 12.*Selección de variables*

Periodo	Cantidad de clústers	Cantidad de variables	Porcentaje en cada clúster
2015	3	33	1 – 0,662
			2 – 0,302
			3 – 0,034
2016	3	22	1 – 0,678
			2 – 0,293
			3 – 0,027
2017	3	21	1 – 0,854
			2 – 0,090
			3 – 0,054
2018	3	33	1 – 0,663
			2 – 0,307
			3 – 0,028
2019	3	33	1 – 0,613
			2 – 0,356
			3 – 0,029

Llegando a este punto, el algoritmo de clustering K-Means se lleva a cabo en Python con la ayuda del paquete K-Means disponible en la librería scikit-learn (*SCIKIT LEARN*, n.d.-b). Dentro de los parámetros de agrupación se ponen en marcha para los cinco años: tres conglomerados a conformar establecidos en “Elbow method”, método de inicializar el modelo con k-means++ buscando seleccionar los centroides iniciales de cada subgrupo de una forma inteligente para acelerar la convergencia de este, 150 como número máximo de iteraciones, n_init en 20 para definir el número de veces que se ejecutará el algoritmo con semillas de centroides diferentes, y finalmente el parámetro random_state en cero para determinar la inicialización de los centroides en cada iteración.

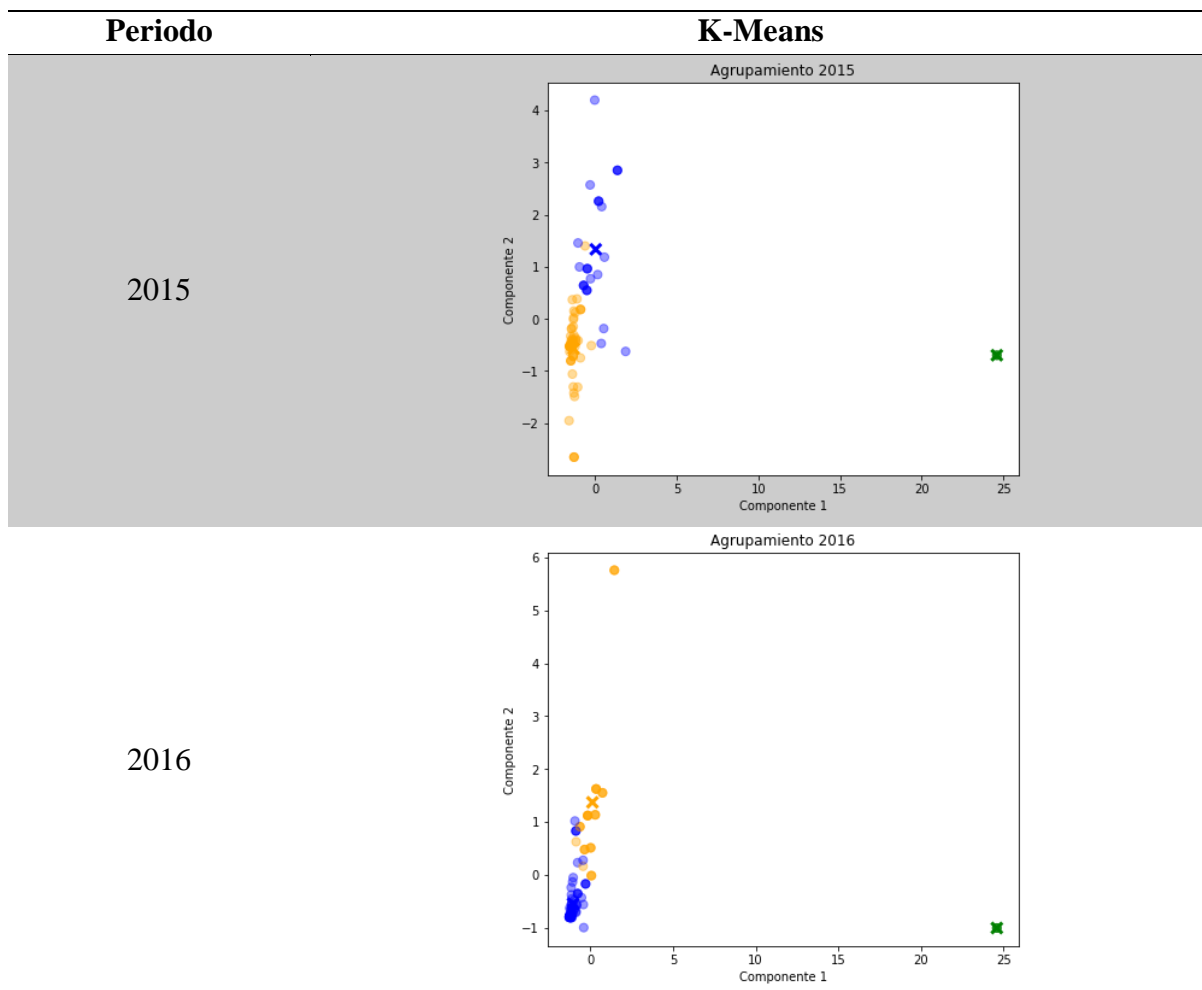
Tras la ejecución del modelado K-Means, se encuentran resultados relacionados con: etiquetas de los datos presentando a qué grupo pertenece cada observación, matriz que contiene

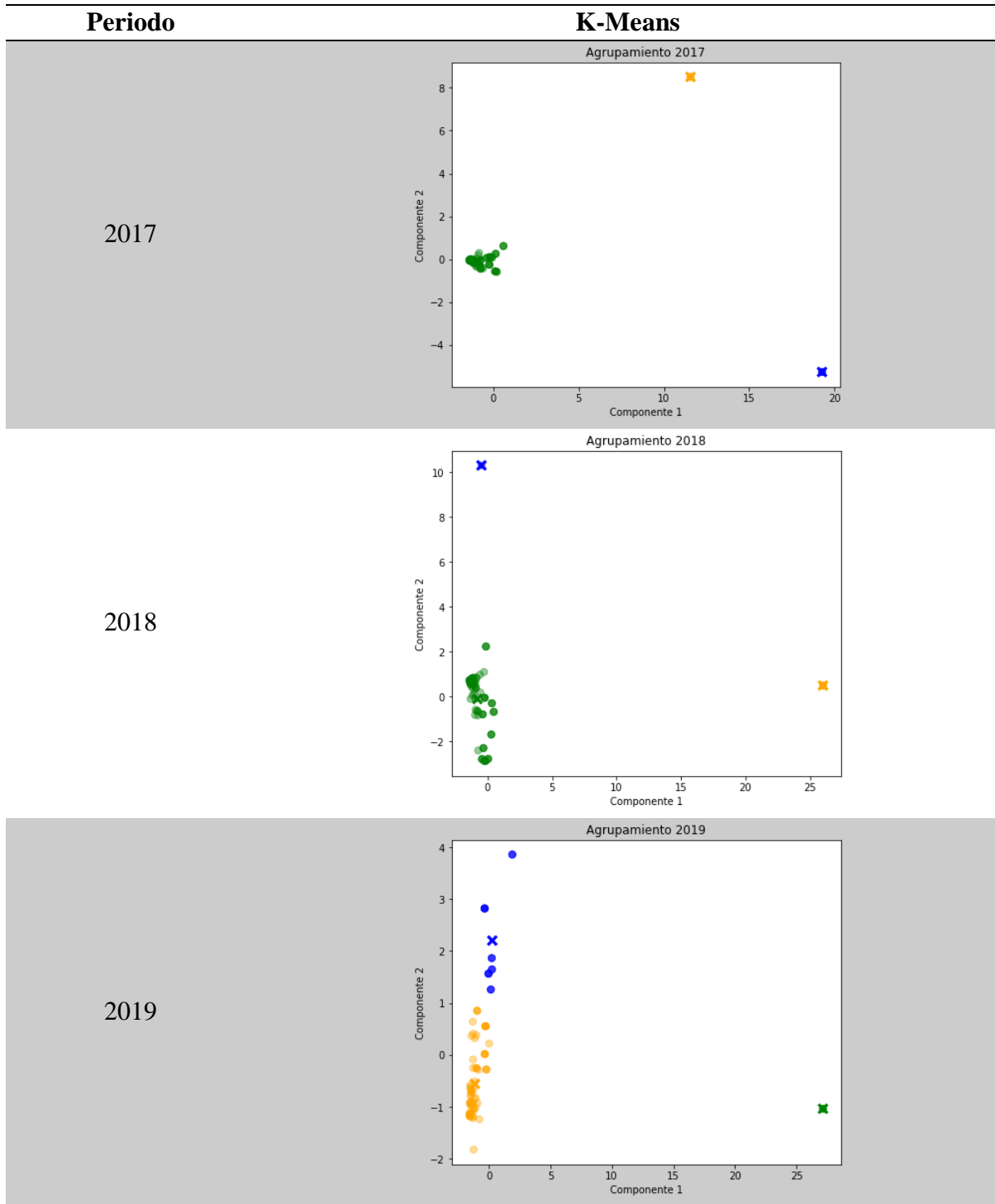
los centros de los conglomerados mencionados y una matriz que contiene información sobre la distancia desde cada observación a su punto representativo el centroide.

Para la representación gráfica de las etiquetas extraídas del modelo K-Means se hace uso de una técnica de análisis de componentes principales, PCA, buscando resumir información en dos componentes, en otras palabras, dicha técnica busca reducir la dimensionalidad transformando la base de datos en dos componentes que contienen la mayor cantidad de información, es decir, conservar la varianza tanto como sea posible. Las agrupaciones se presentan a continuación.

Tabla 13.

Agrupación K-Means





La descripción más detallada de la información obtenida sobre la estructura y posibles predicciones del sector colombiano de cuero calzado y marroquinería obtenidos en el

agrupamiento K-Means se realizada en el siguiente capítulo “Comparación de algoritmos de clustering” que busca contrastar la información con un agrupamiento Fuzzy K-Means.

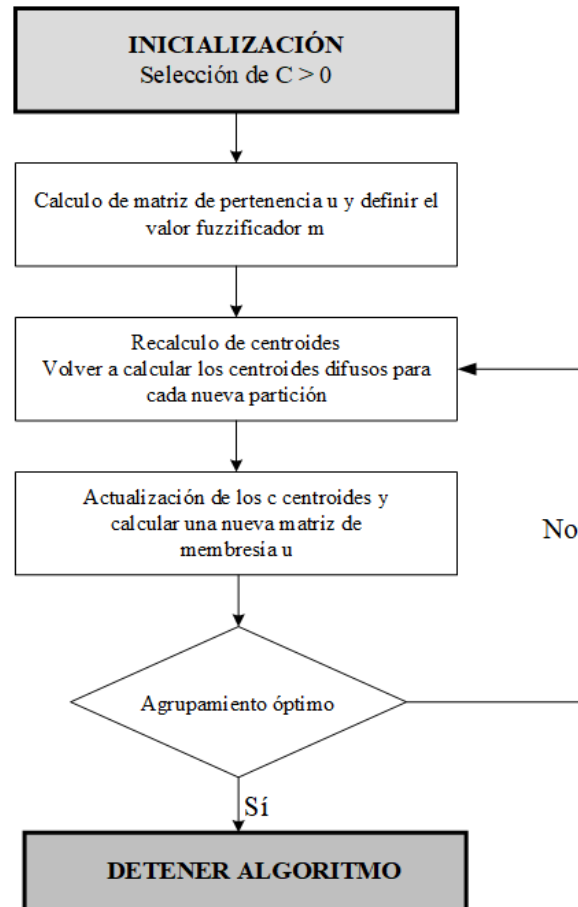
8.2 Algoritmo clustering fuzzy K-means:

Suele suceder en muchas ocasiones que una observación cuenta con características de varios grupos que resulta complejo etiquetarlo en un solo grupo. La técnica de aprendizaje no supervisado Fuzzy-K-Means resulta ser una extensión del algoritmo ya mencionado K-Means, se trata de un agrupamiento suave que encuentra una probabilidad de pertenencia a cada grupo, es decir, un algoritmo que crea grupos borrosos permitiendo que una observación pertenezca a varios grupos según un grado de membresía de cero a uno, siendo uno el mayor grado de pertenencia (Kumar, 2021).

El algoritmo Fuzzy-K-Means repite cuatro pasos hasta lograr la convergencia, formalmente se muestra en el siguiente diagrama de flujo.

Figura 9.

Diagrama de flujo Fuzzy K-means



Nota. Adaptado de (Cazares Aguirre & Martínez Gómez, 2018)

En el mismo orden de ideas, Fuzzy-K-Means requiere del conocimiento a priori de la cantidad de clústers a formar, conforme a lo anterior, se toma la decisión de aplicar **una métrica que busca** descifrar que tan limpiamente están descritos los datos, conocida como coeficiente de partición difusa “FPC”.

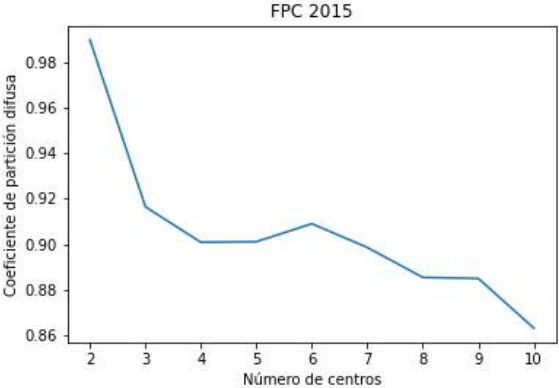
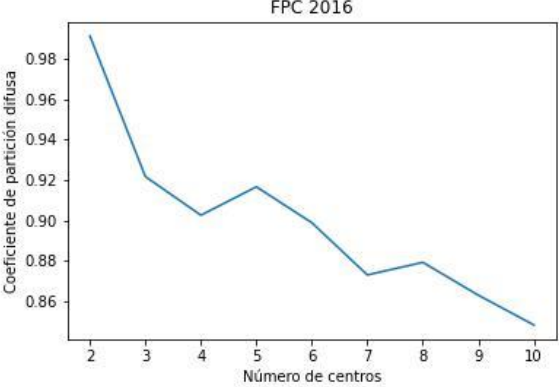
FPC se encarga de medir el índice de solapamiento entre la cantidad de clústers que se crean en rangos de cero a uno, de este modo agrupamientos con índices cercanos a uno indican un modelo con una cantidad de grupos buenos. Para trazar el índice de partición difusa es necesario ejecutar el algoritmo Fuzzy-K-Means para diferentes cantidades de clústers, es importante resaltar

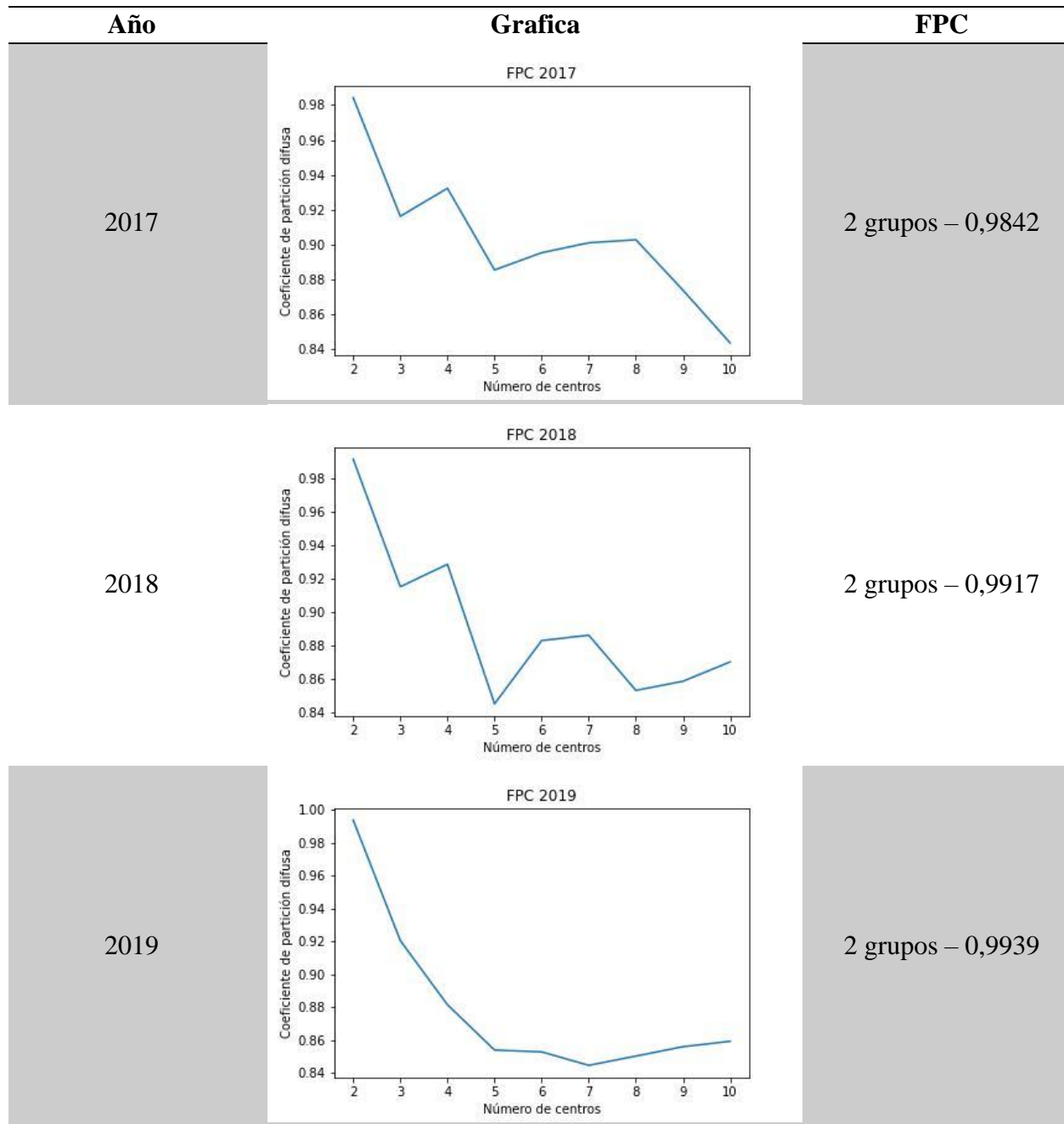
que el primer agrupamiento se inicializa en dos grupos ya que un grupo resulta ser una solución trivial.

En la siguiente tabla se representa el coeficiente de partición que presenta dos grupos como número óptimo de clústers borrosos a crear en la industria de cuero, calzado y marroquinería.

Tabla 14.

Graficas coeficiente de partición difuso

Año	Grafica	FPC
2015		2 grupos – 0,9897
2016		2 grupos – 0,9916



En sentido en paralelo se adoptan las variables analizadas en el algoritmo K-Means en la construcción de los agrupamientos difusos. Presentado lo anterior, la ejecución del clústers difuso a través del paquete skfuzzy (*SFIKIT FUZZY*, n.d.). Nuevamente, el algoritmo precisa los siguientes parámetros que fortalecen el agrupamiento: cantidad de clústers a conformar a partir de los resultados obtenidos en el coeficiente de partición difusa, parámetro de borrosidad m el cual se

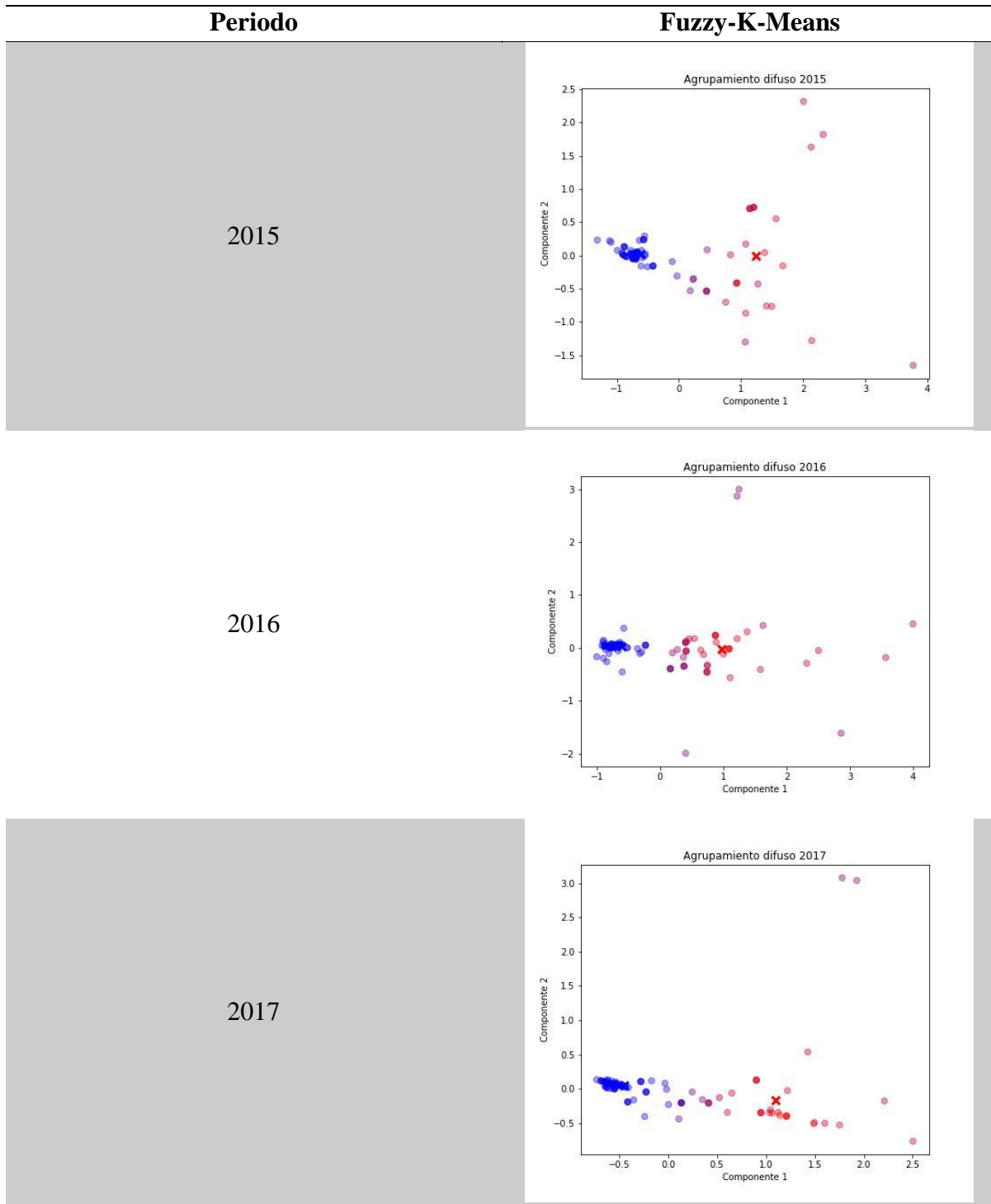
recomienda establecer en dos ya que en cuanto m se acerca a uno las agrupaciones resultan ser bastante diferenciadas, pero al acercarse al infinito el agrupamiento se convierte muy difuso, del mismo modo, se establece un error en 0,005 para indicar el cambio de los valores de centroides en cada iteración, un máximo de iteraciones establecido en 150 veces, un método de inicialización y una semilla en cero para establecer los puntos centrales de inicialización aleatorio; dados estos parámetros el algoritmo comienza a iterarse hasta que se cumpla el máximo de iteraciones o hasta que el cambio de los centroides sea menor al error, el algoritmo converge al alcanzar alguno de estos dos parámetros.

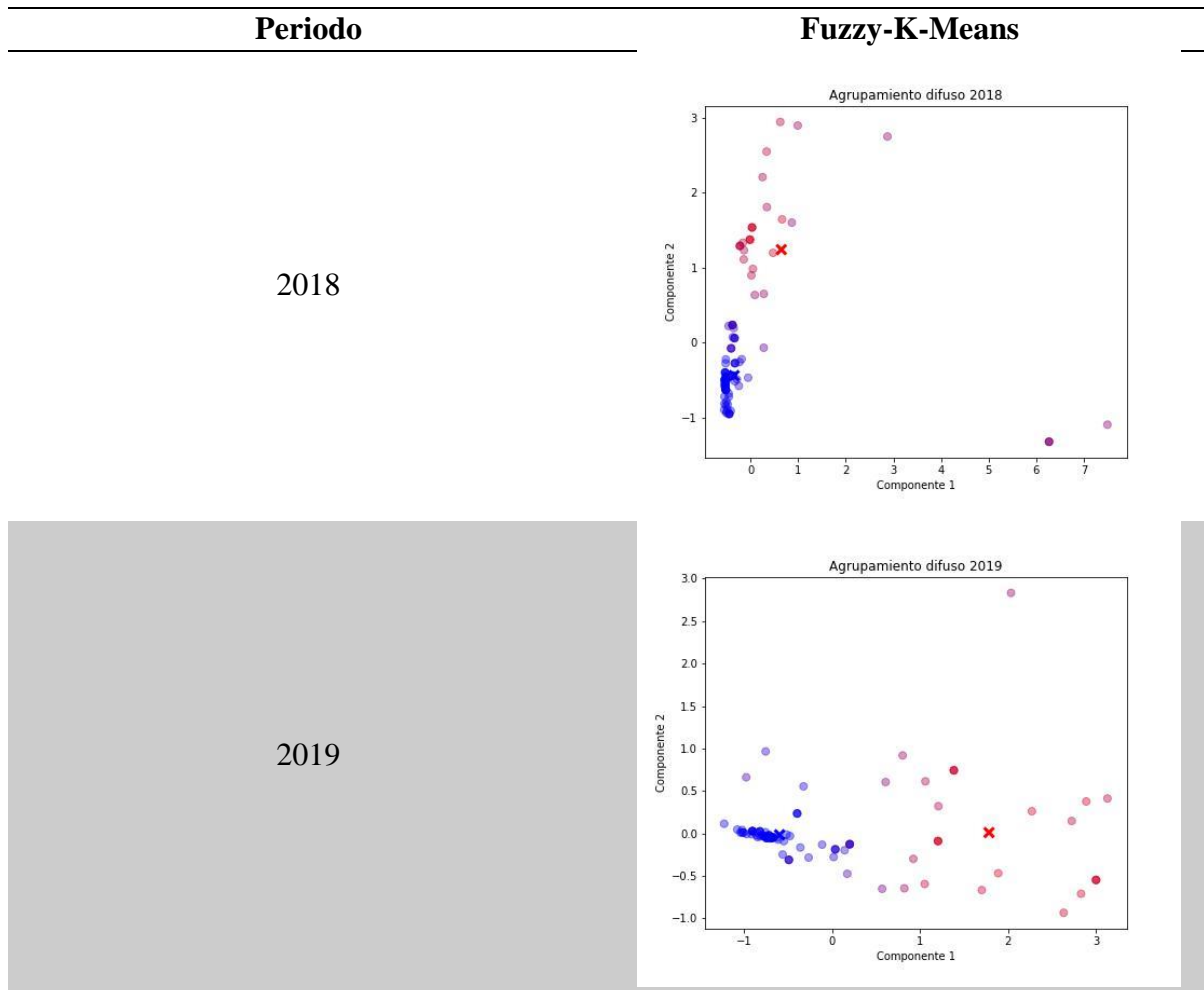
Fuzzy K-Means proporciona resultados relaciones con : matriz que contiene información sobre los centroides de cada conglomerado ($cntr$), una matriz de membresía (u) que contiene los valores asignados de cada observación a cada grupo conformado, matriz de membresía inicial (u_0), una matriz de distancia euclidiana final de cada punto al centro del conglomerado (d), valor de optimización de la función objetivo del algoritmo (jm) a partir de la optimización de u actualizado u_0 en cada iteración, número de iteraciones a la que el modelo converge (p), es decir, la cantidad de veces que el modelo de inicializa y encuentra jm hasta alcanzar su valor óptimo, finalmente, muestra resultados acerca de la construcción de un buen agrupamiento de cero a uno a través del índice de partición difusa el cual se encarga de medir el solapamiento o separación entre los grupos conformados (Córdoba, 2014).

Graficar los resultados obtenidos facilita la comprensión de la pertenencia de las observaciones a cada grupo, razón por la cual se implementa nuevamente un análisis de componentes principales que permita visualizar los datos en dos componentes guardando su variabilidad. Los resultados se presentan a continuación:

Tabla 15.

Graficas agrupamiento Fuzzy K-Means





9. Análisis de algoritmos

En este capítulo se realiza un análisis preliminar de los resultados obtenidos en cada uno de los algoritmos aplicados con el objetivo entender los patrones obtenidos, realizar un acercamiento más acertado de la información extraída y de este modo facilitar la comparación de los algoritmos que se propone en el próximo capítulo.

En los apéndices 5 y 6 se presentan los análisis de los resultados obtenidos en la ejecución de los algoritmos de agrupamiento propuestos en este proyecto de investigación.

9.1. Algoritmo K-Means

Con el objetivo de obtener un acercamiento al comportamiento del sector a través del tiempo, se determina realizar un algoritmo de K-Means para cada año. A partir del reporte extraído del software Python, la metodología del algoritmo de K-Means indica que las empresas del sector de cuero calzado y marroquinería de Colombia pueden agruparse de manera óptima en 3 clústers anualmente de la siguiente manera:

Tabla 16.

Agrupamiento realizado por el algoritmo K-Means

Año	Clúster	Descripción
2015 – 2016 – 2019	0	Empresas comercializadoras, micro, pequeñas y medianas
	1	Empresas micro, pequeñas y medianas
	2	Empresas grandes
2017	0	Empresas comercializadoras, micro, pequeñas y medianas
	1	Empresas grandes
	2	Empresas grandes
2018	0	Empresas pequeñas
	1	Empresas comercializadoras, micro, pequeñas y medianas
	2	Empresas grandes

Cada una de estas agrupaciones presenta características similares que hacen que pertenezcan a un clúster en específico, de la tabla anterior se puede detallar que, en los años 2015, 2016 y 2019 las agrupaciones de los clústers se mantuvieron a diferencia de los años 2017 y 2018 donde se obtuvieron agrupaciones particulares debido a los diferentes comportamientos de las

variables en esos años. Adicionalmente, es importante recalcar que el Departamento Administrativo Nacional de Estadística – DANE no encuestó al mismo número de empresas todos los años, en el año 2015 se encuestaron 83 establecimientos, en el año 2016 fueron 109 establecimientos, en el año 2017 fueron 110 establecimientos, en el año 2018 fueron 104 establecimientos y en el 2019 fueron 101 establecimientos.

9.1.1 Caracterización de conglomerados K-means

Dentro de esta sección se detallan las características que contiene cada agrupación con el fin de encontrar patrones significativos.

9.1.1.1 Agrupaciones año 2015

Clúster 0. En esta primera agrupación se encuentra el 30% de las organizaciones encuestadas en el año 2015 (26 empresas) las cuales se categorizan como micro, pequeñas y medianas. En este año las empresas alcanzaron una producción bruta medio alta con un valor máximo de \$1.035.984.842 y un mínimo de \$199.696.430, respecto a las compras de materia prima sólo realizaron compras nacionales con montos desde \$184.548.480 hasta \$935.746.031. Los costos y gastos de producidos para el uso propio de maquinaria y equipo de las empresas fueron nulos, en otras palabras, no pagaron nada por este servicio; de acuerdo con el servicio pagado para el traslado de las materias primas fue poco puesto que las compras de las materias se realizaron dentro del país. Estas organizaciones invierten bajos montos en edificios y maquinaria, esto se ve reflejado en que la máxima inversión realizada en este año fue de un total de \$17.812.016, por lo

anterior, estas empresas se ven en la necesidad de alquilar la maquinaria para su producción. Adicionalmente, esta agrupación adquiere trabajos realizados por terceros realizando pagos bajos por este servicio y a su vez presenta valores en productos terminados por un máximo de \$50.637.484.

Clúster 1. En este clúster se encuentra el 3% de las empresas encuestadas en el año (3 empresas) categorizadas como grandes empresas. Esta agrupación representa la producción bruta más alta del sector en el año 2015 con un valor máximo de \$5.366.718.950, también realizan compras de materia prima nacional e internacional, pero siendo mayor la compra nacional con un total de \$3.236.450.716 en comparación con \$974.030.213. En este año estas organizaciones invirtieron en activos fijos como maquinaria y equipo un total de \$407.159.776 y reflejándose en un valor alto en libros por este activo de \$2.165.425.148 siendo las empresas que mayores inversiones realizan respecto a este activo puesto que en terrenos y edificios las inversiones son bajas. Las grandes empresas de este sector no recurren al trabajo realizado por terceros a razón de que cuentan con el personal y maquinaria necesario para ejecutar su producción.

Clúster 2. En este conglomerado se localiza el 66% de las empresas encuestadas en el año (57 empresas) las cuales se clasifican como comercializadoras, micro, pequeñas y medianas. Al igual que las empresas del clúster uno, estas realizan mayor compra de materia prima nacional por un total de \$240.725.247 a comparación de las compras de materia prima internacional que las realizan por un total de \$11.964.904. En relación con las inversiones, este conglomerado no realiza inversiones en terrenos y aquellas inversiones en maquinaria las realizan por valores bajos por un total de \$78.598, esta agrupación cuenta con un total de \$301.919 en valores de producto terminado siendo menor a comparación de las otras agrupaciones.

9.1.1.2 Agrupaciones año 2016

Clúster 0. Esta agrupación está conformada por el 68% de las empresas analizadas en el año (74 empresas) caracterizadas por ser comercializadoras, micro, pequeñas y medianas. En el año 2016 estos establecimientos realizan ventas totales de \$50.711.160 siendo el valor más bajo respecto a los demás conglomerados, tienen valores de existencias de productos terminados por un total de \$265.691 y de productos en proceso por un total de \$14.928. Esta agrupación ejecuta mayores compras de materia prima internacional por un valor de \$50.711.160 a comparación de las compras de materia prima nacional por un valor de \$4.462.936, respecto a los valores en libros, estas empresas contienen \$778.057 en terrenos, \$2.413.801 en edificios y \$2.417.923 en maquinaria. Adicionalmente, es importante recalcar que esta agrupación no tiene gastos elevados en el servicio de mantenimiento y reparaciones puesto que no cuenta con maquinaria y equipo para adquirir este servicio y a su vez se refleja que el pago que realiza por el servicio de energía no es elevado a comparación de los demás clústers.

Clúster 1. En este conglomerado se encuentra el 3% de las empresas encuestadas en el año (3 empresas) que se clasifican dentro de las grandes empresas. Este clúster se destaca por contener los mayores valores en las variables respecto a las demás agrupaciones. Genera mayores ventas por un valor de \$4.931.513.988 en el año, a su vez, realiza mayores compras de materia prima internacional por un total de \$3.441.846.819 en comparación a las compras nacionales que se realizan por un total de \$1.251.160.197. Es relevante mencionar que estas empresas incurren en un gasto alto por el servicio de mantenimiento y reparaciones por un total de \$80.346.658 debido a que están dotados de varios equipos de maquinaria, adicionalmente, estos establecimientos suelen guardar existencias de productos en proceso por \$28.260.582 y de productos terminados

por \$65.528.605. De acuerdo con sus valores en libros, estas presentan en terrenos un total de \$293.971.374, en edificios un total de 297.130.855 y en maquinaria un total de \$2.211.865.04.

Clúster 2. En este clúster se sitúa el 29% de las organizaciones entrevistadas en el año (32 empresas) las cuales se caracterizan por ser micro, pequeñas y medianas. Este conglomerado generó ventas en el 2016 por un total de \$429.778.345 considerándose un valor medio entre las otras dos agrupaciones, teniendo en cuenta al servicio que paga por mantenimiento y reparaciones es proporcional al valor en máquinas que tiene ya que paga un total de \$285.095. De acuerdo con la materia prima que consume, realiza mayores compras internaciones por un valor de \$429.778.345 a comparación de las compras nacionales las cuales paga un total de \$30.282.020, se destaca que estas organizaciones no dejan gran cantidad de existencias de productos en proceso puesto que el total que tienen en el año es de \$61.947 y en productos terminados un total de \$1.532.383.

9.1.1.3 Agrupación año 2017

Clúster 0. En este conglomerado se agrupan el 3% de las empresas en estudio en el año 2017 (3 empresas) que son caracterizadas por ser grandes. Estas empresas se caracterizan por presentar los valores más elevados respecto a las otras organizaciones, de acuerdo con las ventas totales realizadas en el año 2017 este clúster vendió un total de \$5.221.782.151, efectuó un total de \$5.155.749.980 por compras de materia prima internacional siendo mayor comparado a las compras de materia prima nacional por un total de \$913.215.353. De acuerdo con los servicios que adquiere, paga un valor de \$306.705 por la energía y un valor de \$53.308.235 por mantenimiento y reparaciones, adicionalmente, presenta existencias de productos en proceso por un total de

\$58.021,852 y por productos terminados un total de \$88.220.374. Es importante destacar que estas organizaciones no adquieren el servicio contratado con terceros puesto que cuentan con el personal y la maquinaria necesaria para realizar su producción.

Clúster 1. Esta agrupación se localiza el 95% de las organizaciones encuestadas en el año (104 empresas) que se clasifican como comercializadoras, micro, pequeñas y medianas. Este conglomerado es el que genera menores valores en el año comparado a las otras dos agrupaciones, respecto a sus ventas totales realiza un total de \$163.931.082 en el año y efectúa menores compras de materia prima nacional por un valor de \$13.754.511 en comparación a las compras de materia prima internacional por un valor de \$151.239.798. Los valores que paga por servicios son proporcionales al tamaño de la empresa, puesto que, por energía cancelan un total de \$85.255 y por mantenimiento y reparaciones un total de \$153.045. Estas empresas suelen alquilar maquinaria sin opción de compra para ejecutar su producción por un total de \$14.577, al igual, incurren en gastos de servicios contratados por terceros por un valor de \$88.425.

Clúster 2. El 3% de las empresas encuestadas en el año (3 empresas) se asignan a este conglomerado, las cuales se caracterizan por ser grandes. En este año, ocurre algo particular y es que existen dos clústers que se encuentran conformados por grandes empresas que se diferencian porque uno contiene valores más altos respecto al otro. Este clúster se caracteriza por tener valores menores al clúster cero, puesto que, en ventas presenta un total de \$3.116.378.954 en comparación a la otra agrupación. De acuerdo con la compra de materia prima, este efectúa mayores compras de materia prima internacional por un total de \$2.565.415.014 comparado a las compras nacionales por un total de \$443.608.907, adicionalmente, se destaca que esta agrupación no requiere de arrendamiento de maquinaria y equipo puesto que cuenta con los activos fijos para realizar su producción. Asimismo, esta presenta existencias de productos en proceso por un valor de

\$98.558.043 valor que es más alto en comparación a el clúster cero, pero, en existencias de productos terminados si es menor puesto que tiene un valor de \$42.858.421.

9.1.1.4 Agrupación año 2018

Clúster 0. En esta agrupación se encuentra el 0,96% de las empresas analizadas en el año (1 empresa) caracterizada por ser pequeña. En este año 2018 pasa algo particular similar a las agrupaciones del año anterior, pero en este caso se genera una agrupación donde solo pertenece una empresa categorizada como pequeña, esto se debe a que esta organización genera menores valores a las empresas que pertenecen al clúster uno. De acuerdo con las ventas realizadas por esta agrupación genera un total de \$152.831.662 siendo el menor valor a comparación de los demás conglomerados, esta empresa realiza compras de materia prima internacional por \$133.360.798 siendo mayor a las compras de materia prima nacional por un valor de \$17.581.840. Se destaca que esta agrupación no incurre en gastos por servicios contratados por terceros ni en alquiler de maquinaria y equipo, no realiza inversiones en edificios y no gasta en el servicio de transporte de materia prima. Adicionalmente, invierte en terrenos un total de \$6.550.402 y en maquinaria un total de \$74.461, respecto al servicio de mantenimiento y reparaciones estas efectúan un pago de \$4.828.426 siendo mayor a comparación del clúster uno.

Clúster 1. En este conglomerado se sitúa el 96,1% de las organizaciones encuestadas en el año (100 empresas) las cuales se categorizan como comercializadoras, micro, pequeñas y medianas. Estas empresas realizan ventas totales de \$201.034.284 valor que se encuentra en medio de las demás agrupaciones, de acuerdo con las compras efectuadas para materia prima internacional estos conglomerados pagan un total de \$178.983.075 siendo mayor a el valor pagado

por materia prima nacional por un total de \$25.449.849. De acuerdo con los servicios que estas empresas utilizan se incurre \$115.335 por el servicio contratados con terceros, \$103.457 por el valor de energía comprada y \$276.374 por el servicio de mantenimiento y reparaciones, asimismo, este conglomerado invierte en maquinaria \$308.631 y en terrenos \$398.329. También, genera gastos en el servicio de arrendamiento de maquinaria por un total de \$31.813 y por arrendamiento de bienes inmuebles un total de \$227.408.

Clúster 2. Este clúster se conforma por el 2,8% de las empresas entrevistadas en el año (3 empresas) clasificadas como grandes empresas. Este conglomerado se caracteriza por generar los mayores valores en el sector, conforme a valores de ventas totales del año 2018 se generó un total de \$5.474.677.629, a su vez, las compras de materia prima internacional son de \$5.060.731.600 siendo un valor mayor a la compra de materia prima nacional presentando un comprado de \$1.100.831.014. Adicionalmente, estas empresas realizan un pago total de \$612.208 por el servicio de energía y un total de \$54.194.738 por el servicio de mantenimiento y equipos, a su vez, este clúster presenta mayor valor en productos en proceso por un total de \$74.261.618 y en productos terminados un total de \$94.597.531. Ejecutan una inversión alta en maquinaria por un valor de \$153.493.028 y en edificios un total de \$692.628.

9.1.1.5 Agrupación año 2019

Clúster 0. Este conglomerado se encuentra conformado por el 20,7% de las empresas encuestadas en el año (21 empresas) categorizadas como micro y pequeñas. Estas empresas realizan ventas totales por un total de \$746.488.853 valor que se encuentra en medio de los otros dos clústers formados en el presente año, a su vez, en este año estas ejecutan compras de materia

prima internacional por un total de \$698.306.833 siendo mayor al valor comprado de materia prima nacional por un total de \$72.936.170. Los pagos que se generan por servicios adquiridos en energía comprada son de \$171.512 y en mantenimiento y reparaciones es de \$404.419, de acuerdo con las existencias de productos en proceso estas tienen un total de \$4.439 siendo el valor más pequeño en comparación a las demás agrupaciones y en productos terminados un total de \$142.455. Respecto a las inversiones realizadas, estas invierten más en edificios por un total de \$1.194.888 en comparación de la inversión en maquinaria que fue por un total de \$295.557.

Clúster 1. Esta agrupación está formada por el 2,97% de las organizaciones analizadas en el año (3 empresas) las cuales se clasifican como grandes. Al igual que los demás años, estas empresas que entran dentro de esta clasificación son caracterizadas por generar los mayores valores respecto a los demás conglomerados, de acuerdo con sus ventas realizadas estas en el año 2019 realizan un total de \$5.693.597.171. Teniendo en cuentas las compras de materia prima efectuadas, estas ejecutan mayores compras de materia prima internacional por un valor de \$5.482.782.274 en comparación de las compras de materia prima nacional por un valor de \$797.718.486, así mismo, estas organizaciones incurren en gastos como mantenimiento y reparaciones por un total de \$53.931.095 y en gastos de transporte de materias primas por valor de \$11.503.230. De acuerdo con las inversiones realizadas estas efectúan mayores inversiones en maquinaria por un total de \$190.248.441 y en edificios un total de \$262.215.

Clúster 2. Este clúster se conforma por el 74,24% de las empresas entrevistadas en el año (77 empresas) clasificadas como comercializadoras, micro, pequeñas y medianas. De acuerdo con esta agrupación, estas generan los valores más pequeños en el año puesto que en ventas efectuaron un total de \$82.750.277 siendo bajo en comparación a los demás conglomerados. Respecto a las compras de materia prima realizadas el patrón que se repite en los otros dos clústeres se mantiene,

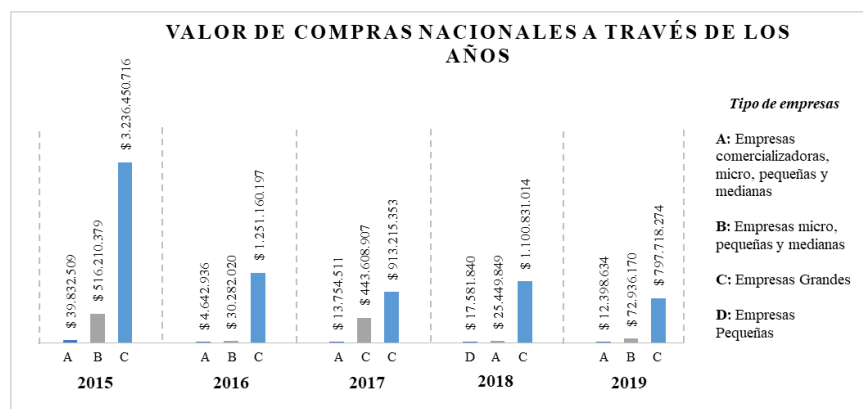
realizan mayores compras de materia prima internacional por un valor de \$69.468.530 en comparación a las compras de materia prima nacional por un valor de \$12.398.634, a su vez, estas empresas incurren en gastos bajos por servicios de energía por un total de \$94.355 y por el servicio de mantenimiento y reparaciones un total de \$269.440. Las inversiones que estas realizan son mínimas debido a que en edificios ejecutan en total una inversión de \$215.006 y en maquinaria de \$22.585, así mismo, este clúster adquiere el servicio de arrendamiento de maquinaria sin opción de compra por un valor de \$29.209 para efectuar su producción y adicionalmente el arrendamiento de bienes inmuebles por un total de \$200.393.

9.1.2 Comparativa de conglomerados K-Means

En esta sección se pretende conocer el comportamiento y los cambios de los clústers que se conforman a través del tiempo por medio de la ejecución del algoritmo K-Means, a continuación, se describen los conglomerados según algunas variables significativas que se presentan en todos los años en estudio.

Figura 10.

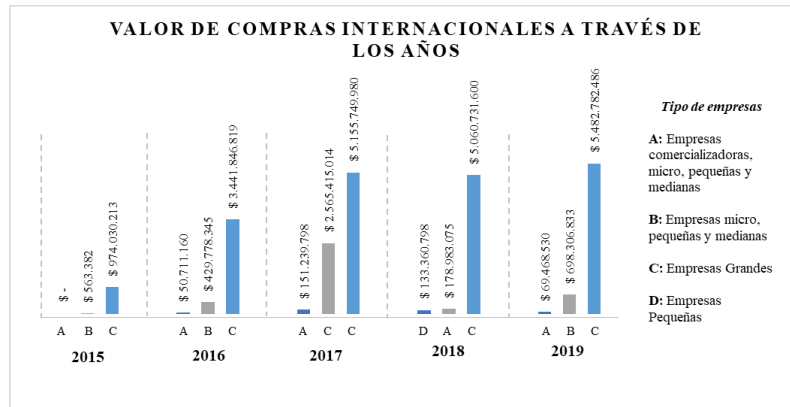
Valor de compras nacionales en los diferentes años



De acuerdo con la información suministrada por la figura 10 se determina el comportamiento que tuvieron las empresas del sector referente a la compra de materia prima nacional a través de los años. En el año 2015 se registró la compra más alta de materia prima realizada por las grandes empresas del sector, seguidamente por las empresas clasificadas como micro, pequeñas y medianas. Es posible detallar que las empresas que se encuentran en la categoría “A” son las que realizan bajas compras en materia prima nacional en valores desde \$4.642.936 hasta \$39.832.500, pero, adicionalmente ocurre algo particular en los años 2017 y 2018 en los cuales el comportamiento de las agrupaciones cambia debido a que en el año 2017 uno de los clústers que pertenece a las grandes empresas realiza a una menor compra nacional por el valor de \$443.608.907 en comparación a la otra agrupación de grandes empresas que pagan un valor promedio de \$913.215.353, algo similar de lo que sucede en el año 2018 donde el clúster de las pequeñas empresas efectúa una compra menor por un precio de \$17.581.840 de materia prima nacional respecto al grupo de la categoría A que realiza una compra de \$25.449.849 en promedio. Es importante destacar que las compras nacionales realizadas por este sector han venido disminuyendo través del tiempo como se puede observar en la figura 10.

Figura 11.

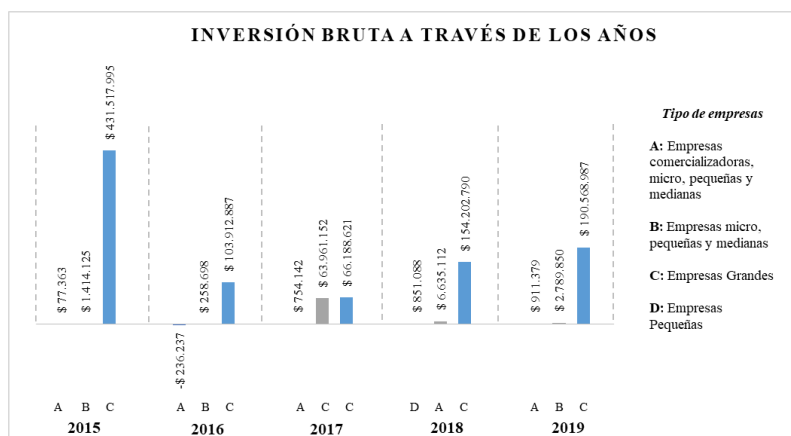
Valor de compras internacionales en los diferentes años



A partir de la figura 11 se contempla que en el año 2015 las empresas categorizadas en el grupo “A” no realizaron compras de materia prima internacional y que las grandes empresas realizaron pocas compras internacionales a comparación de los demás años con un valor promedio de \$974.030.213. En el 2019 las grandes empresas del sector de cuero, calzado y marroquinería colombiano realizaron la compra más alta respecto a los demás años en materia prima internacional por un valor de \$5.482.782.486 en promedio. El comportamiento de los clústers en el año 2017 y 2018 presentan características donde en un clúster el valor de la compra internacional es menor al valor comprado de la clasificación de empresas tipo “A” y “C”. Adicionalmente, al observar la figura se puede detallar la tendencia de alza de las compras internacionales realizadas por el sector a través de los años, que en su gran mayoría es realizada por las grandes empresas debido a que estas generan más producción y venta respecto a las demás clasificaciones de empresas.

Figura 12.

Inversión bruta en los diferentes años



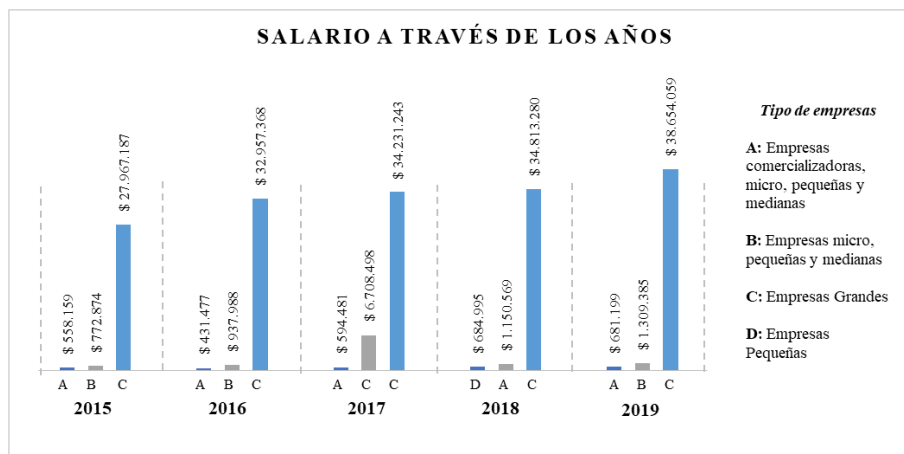
De acuerdo con la salida que se encuentra en la figura 12 se observa que el 2015 fue el año donde mayor inversión bruta alcanzaron las grandes empresas del sector de cuero, calzado y marroquinería colombiano puesto que el valor llegó a los \$431.517.993 en promedio. En este mismo año, las compras de materia prima en su mayoría se realizaron dentro del país, razón por la cual los costos de inversión bruta son elevados ya que los materiales de cuero y textil dentro del país resultan escasos y costosos.

En el lapso de 2016 a 2017 esta tuvo una disminución hasta llegar a un valor de \$66.188.621, sin embargo, en los años 2018 y 2019 estas organizaciones lograron aumentar de nuevo en sus inversiones. Es importante resaltar que la inversión bruta en este sector no es alta, se mantiene en promedios medio bajos entre \$100.000.000 a \$400.000.000 respecto a la información recolectada por el DANE, adicionalmente, es relevante el comportamiento que se presenta en los diferentes años, puesto que las empresas clasificadas dentro de las categorías “A”, “B” y “D” no realizan grandes inversiones de capital ya que los valores en los que invierten están entre los \$77.363 hasta \$2.789.950 en promedio debido a que el tamaño de estas organizaciones no les permite efectuar compras en proporciones grandes porque no tendrían ni el recurso ni el capital para generarlas. Dentro de este análisis, se puede observar que en el año 2016 se tiene un valor

negativo en inversión bruta por las empresas que se clasifican tipo “A”, este dato puede estar erróneo puesto que la inversión bruta es siempre positiva o nula (no se destruye el capital existente).

Figura 13.

Salario pago a personal en los diferentes años

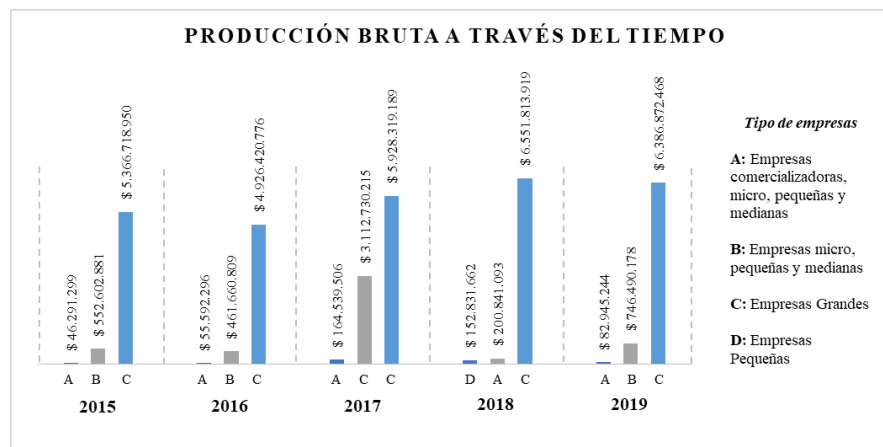


En base a la figura 13 se aprecia el aumento del salario a través de los años, siendo las grandes empresas las promotoras del pago mayor de salarios al personal debido a pagos en el 2015 por \$27.967.187 y en el año 2019 los valores se elevaron a \$38.654.059 en promedio. Las empresas que menores pagos realizan respecto a salarios por personal son las que se clasifican como empresas comercializadoras, micro, pequeñas y medianas ya que los valores que pagan están ente \$558.159 y \$681.199 en promedio. Lo anterior a razón de su reducido tamaño no generan alta empleabilidad en el sector. Así mismo, se puede deducir que no existe mayor diferencia entre los pagos que generan las empresas clasificadas entre la categoría “A” y “B” puesto que sus pagos se establecen entre los rangos de \$431.477 hasta \$684.995. En el año 2017 ocurre un cambio que es notorio entre el dinero pagado por una grande empresa la cual paga \$6.708.498 y por el clúster a

donde pertenecen las demás grandes empresas ya que estas pagan un valor de \$34.231.243 y debido a esta situación el algoritmo K-Means realiza un clúster donde solo pertenezca esa empresa que se sale de los rangos establecidos por las demás grandes empresas del sector.

Figura 14.

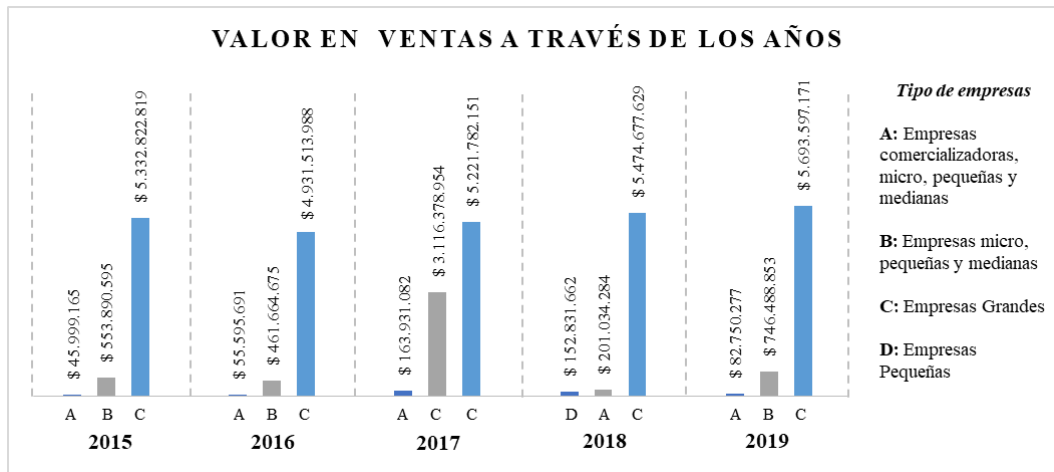
Producción bruta en los diferentes años



De acuerdo con la figura 14 es posible observar el comportamiento que han presentado las empresas del sector a través de los años, se puede destacar que la producción de estas empresas suele mantenerse en el transcurso del tiempo tendiendo a tratar de aumentar en su producción. El valor de la productividad que han generado estas empresas que pertenecen al sector se ve reflejado según el tipo de empresas que la conforman puesto que entre mayor sea la organización su producción será superior a las demás, lo anterior se puede ver reflejado en la producción que generan las grandes empresas, valores que se encuentran entre los rangos de \$4.926.420.776 hasta \$6.551.813.919. También se puede determinar que el volumen de productos que se realizaron en el año fue proporcional al tipo de empresas por clasificación.

Figura 15.

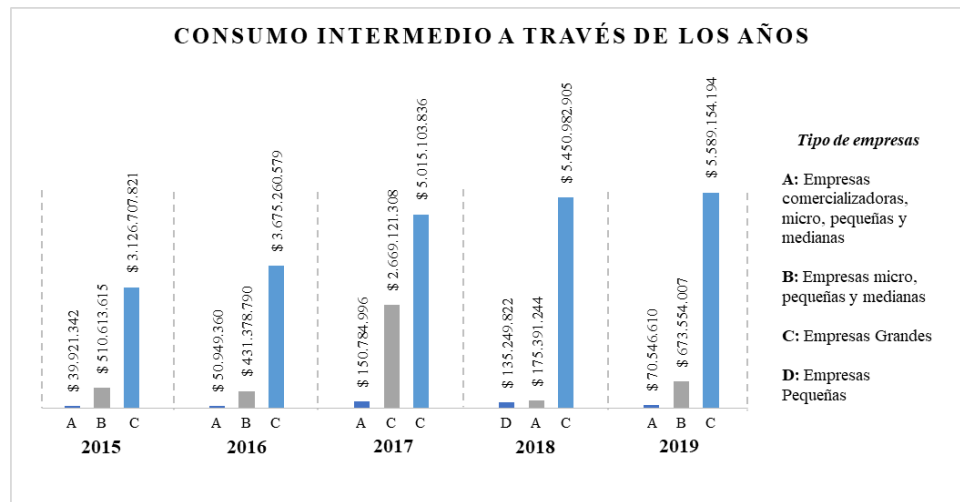
Valor en ventas en los diferentes años



En esta figura 15 se detalla que las ventas en el sector son proporcionales al tamaño de la empresa, de acuerdo con las grandes organizaciones en este caso obtuvieron ventas altas y las cuales alcanzaron un leve incremento a través del tiempo puesto que en el año 2015 obtuvieron \$5.332.822.819 y en el año 2019 el valor fue de \$5.693.597171. De las empresas clasificadas dentro de la categoría “A” y “B” se observa que estas generaron ventas bajas en valores entre \$152.831.662 hasta \$746.488.853 a comparación de las grandes empresas puesto que estas cuentan con poco personal y su área de trabajo es reducida para poder realizar producciones a gran escala. Adicionalmente, con estas observaciones se puede examinar que el sector de cuero, calzado y marroquinería no tiene potencializado el desempeño de la producción para poder generar mayores ventas y que al generar bajas cantidades de productos los clientes no tendrán variedad de mercancía para la compra.

Figura 16.

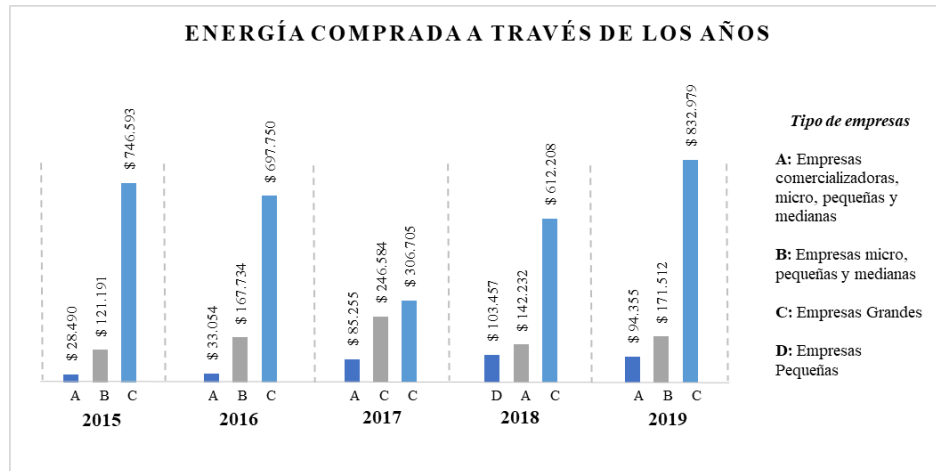
Consumo intermedio en los diferentes años



A partir de la figura 16 se determina que las empresas que conforman el sector de cuero, calzado y marroquinería en Colombia involucran en el proceso de producción bienes y servicios no durables para crear otros. Las empresas que mayor consumo intermedio utilizan son las grandes empresas, a través de los años su valor ha aumentado en comparación con el año 2015 que presenta pagos por \$3.126.707.821 por este servicio y en el año 2019 pagos de \$5.589.154.194, seguidamente las empresas que se clasifican como tipo “A” y “B” no emplean en su producción gran parte de este consumo puesto que al generar baja cantidad productos el consumo intermedio que requieren es mínimo. Es relevante destacar que en el año 2017 los clústers de las grandes empresas presentan una diferencia significativa entre el consumo intermedio que necesitaron para realizar la producción puesto que un clúster pago \$2.669.121.308 y el otro \$5.015.103.836.

Figura 17.

Energía comprada en los diferentes años



Teniendo en cuenta la figura 17 es posible estudiar la compra de energía que realiza cada grupo de empresas que pertenecen al sector, se determina que las organizaciones que más energía utilizan son las clasificadas como grandes empresas y las cuales tienen un comportamiento variable a través de los años, estos valores pagados varían entre \$306.705 hasta \$882.979. Es posible que estas organizaciones requieran de más energía porque en sus plantas de producción tienen maquinaria especializada para realizar sus productos a diferencia de las micro, pequeñas y medianas empresas puesto que estas ejecutan sus productos por medio de la mano de obra brindada por sus trabajadores.

9.2 Fuzzy K-Means

El análisis fuzzy se realiza con las mismas variables estudiadas, se procede a ejecutar un algoritmo de Fuzzy-K-means para cada año respectivamente, lo anterior con el objetivo de conseguir un resultado más detallado y una aproximación al comportamiento que presenta el sector de cuero calzado y marroquinería a través del tiempo por medio de las agrupaciones realizadas por el algoritmo borroso. De acuerdo con la solución obtenida del software Python, se determina que

el sector puede agruparse en dos conglomerados donde se obtienen las siguientes agrupaciones en los diferentes años desde 2015 hasta 2019:

Tabla 17.

Agrupamiento realizado por el algoritmo Fuzzy K-Means

Año	Agrupaciones	Empresas
2015	Clúster 1	<ul style="list-style-type: none"> - 1 empresa comercializadora - 4 microempresas - 12 pequeñas empresas - 4 medianas empresas - 3 grandes empresas
	Clúster 2	<ul style="list-style-type: none"> - 1 empresa comercializadora - 28 microempresas - 22 pequeñas empresas - 3 medianas empresas
	Existen 5 microempresas y dos pequeñas empresas que contienen pertenencia de las dos agrupaciones	
2016	Clúster 1	<ul style="list-style-type: none"> - 8 microempresas - 21 pequeñas empresas - 7 medianas empresas
	Clúster 2	<ul style="list-style-type: none"> - 1 empresa comercializadora - 39 microempresas - 22 pequeñas empresas - 1 mediana empresa
	Existen 2 empresas comercializadoras, 1 microempresa, 4 pequeñas empresas y 3 grandes empresas que contienen pertenencia de las dos agrupaciones	
2017	Clúster 1	<ul style="list-style-type: none"> - 2 microempresas - 18 pequeñas empresas - 7 medianas empresas
	Clúster 2	<ul style="list-style-type: none"> - 1 empresa comercializadora - 28 microempresas - 18 pequeñas empresas - 20 medianas empresas - 5 grandes empresas
	Existen 2 empresas comercializadoras, 2 microempresas y 1 pequeña empresa que contienen pertenencia de las dos agrupaciones	

Año	Agrupaciones	Empresas
2018	Clúster 1	<ul style="list-style-type: none"> - 1 empresa comercializadora - 3 microempresas - 13 pequeñas empresas - 6 medianas empresas
	Clúster 2	<ul style="list-style-type: none"> - 1 empresa comercializadora - 39 microempresas - 27 pequeñas empresas - 5 medianas empresas - 3 grandes empresas
Existe 1 empresa comercializadora y 5 pequeñas empresas que contienen pertenencia de las dos agrupaciones		
2019	Clúster 1	<ul style="list-style-type: none"> - 2 empresas comercializadoras - 2 microempresas - 13 pequeñas empresas - 8 medianas empresas
	Clúster 2	<ul style="list-style-type: none"> - 1 empresa comercializadora - 35 microempresas - 32 pequeñas empresas - 4 medianas empresas - 3 grandes empresas
Existe 1 microempresa que contiene pertenencia de las dos agrupaciones		

9.2.1 Caracterización de conglomerados Fuzzy K-means

En esta sección se describen las características que se logran identificar de cada conglomerado difuso según el comportamiento de las variables propuestas que inciden en la productividad, competitividad y desarrollo del sector.

9.2.1.1 Agrupaciones año 2015

Clúster 1. La agrupación se caracteriza por organizaciones que se categorizan como comercializadoras, micro, pequeñas, medianas y grandes empresas. En este año las mayores compras de materia prima que ejecutaron fueron nacionales desde \$1.648.276 hasta \$3.236.450.716 a comparación de las compras internacionales que se realizaron entre \$755.457 y \$974.030.213. Es importante resaltar que las empresas comercializadoras y micro no realizan compras de materia prima internacional en este año. El valor que pagan por el servicio de energía es proporcional al tamaño de la empresa, de este modo, aquellas empresas que mayor pago realizan por este servicio son las grandes empresas con \$746.593 en promedio, representando un valor alto comparado a las microempresas que pagan aproximadamente \$5.626. Esta agrupación genera mayores ventas a comparación del clúster dos, pero su producción bruta también es alta lo cual genera una desventaja ya que a mayor valor de producción bruta menores son las ganancias que generan. Adicionalmente, estas organizaciones tienen niveles de existencias de inventario que presentan al final del año, en productos en proceso tienen desde \$321 hasta \$56.147.996, por productos terminados abarcan desde \$87.451 hasta \$81.304.677 y por materia prima presentan valores desde \$122.291 hasta \$29.068.414 siendo valores más elevados en paralelo a la otra agrupación. El salario que estas empresas pagan por el personal es proporcional a las personas contratadas, por lo tanto, entre más cantidad de empleados tengan mayor salario pagan.

Asimismo, este conglomerado incurre en mayores costos y gastos en comparación al clúster dos conforme a los servicios de transporte de materia prima desde \$8.462 hasta \$96.531 y por el servicio de arrendamiento de inmuebles desde \$13.660 hasta \$1.096.525. También, es importante destacar que las empresas que tienen mayor pertenencia a esta agrupación son aquellas que generan mayores inversiones en cuanto a terrenos por \$3.627.302, en maquinaria alrededor de \$35.376 y \$407.159.350 y en edificios desde \$76.049 hasta \$9.931.350.

Clúster 2. En esta agrupación pertenecen en mayor membresía comercializadoras, micro, pequeñas y medianas empresas. Estas organizaciones generan menores ventas y producción bruta en comparación a las que pertenecen al clúster uno generando mejores utilidades. De acuerdo con las compras de materia prima realizadas por estas empresas en el año 2015, se realizaron compras de materia prima nacional entre valores de \$8.054.545 y \$381.445.312 siendo mayores respecto al tipo de empresa que lo realice en el clúster uno. En este año, las organizaciones categorizadas como comercializadoras, pequeñas y medianas no realizaron compras de materia prima internacional y aquellas que si lo realizaron por montos bajos.

Teniendo en cuenta los costos y gastos incurridos por servicios, estas organizaciones causan gastos menores a las empresas que se encuentran en el clúster uno, puesto que, por mantenimiento pagan alrededor de \$181 y \$143.008, por energía desde \$50.702 hasta \$109.380, por servicios contratados por terceros entre \$33.831 y \$112.326 y por el servicio de arrendamiento de inmuebles desde \$3.794 hasta \$417.462. Asimismo, al observar las existencias que tienen al final del año estas empresas presentan menores valores en comparación a los del clúster uno debido a que por existencias de productos terminados tienen valores desde \$71.461 hasta \$16.879.161, por productos en proceso desde \$4.339 hasta \$6.597 y por existencias de materia prima y materiales inicia en \$1.749.909 y terminan en \$6.130.921. Respecto a las inversiones que estas empresas realizan son menores a las realizadas por las del clúster uno, puesto que, por terrenos estas invierten en valores de \$608.333, por edificios desde \$1.645 hasta \$3.131.708 y por maquinaria desde \$15.719 hasta \$808.293.

Adicionalmente, existen empresas que presentan pertenencias similares en ambas agrupaciones “*clúster 1 y clúster 2*” las cuales se categorizan como micro, pequeñas y medianas empresas, al realizar el respectivo análisis estas organizaciones suelen pertenecer más al *clúster 1*

debido a que sus características tienen más similitud a las empresas que conforman ese conglomerado.

9.2.1.2 Agrupaciones año 2016

Clúster 1. Este conglomerado contiene empresas que se categorizan como micro, pequeñas y medianas, presenta bajas ventas y a su vez baja producción bruta lo cual genera que no se generen altas ganancias. De acuerdo con los costos y gastos que estas ocasionan son las más bajas en comparación al otro conglomerado dado que por el servicio de mantenimiento efectúan pagos desde \$1.610 hasta \$141.969 y por energía un valor desde \$4.781 hasta \$220.308. Esta agrupación cuenta con mayor cantidad de existencias en relación con las empresas que pertenecen al clúster dos ya que las existencias de productos en proceso tienen \$434 y \$227.251, por productos terminados tienen valores desde \$50.739 hasta \$5.218.854 y por productos no fabricados por el establecimiento alrededor de \$35 y \$187.757. Teniendo en cuenta los valores en libros que estas empresas tienen en contraste al clúster dos, presentan menores valores dado que por valores en libros en terrenos los valores oscilan desde \$3.963 hasta \$1.383.009, en edificios desde \$276 hasta \$7.209.009 y en maquinaria desde \$67.487 hasta \$187.757. De acuerdo con el comportamiento que estas organizaciones presentan respecto a las compras de materia prima son proporcionales tanto en compras internacionales como en compras nacionales puesto que por compras nacionales realizaron compras desde \$475.592 hasta \$16.691.658 y por compras internacionales entre \$542.41 y \$15.878.541.

Clúster 2. Este clúster se encuentre conformado por empresas comercializadoras, micro, pequeñas y medianas las cuales exponen buenas ventas y a su vez la producción bruta es alta lo

cual no permite que se obtengan buenas ganancias. Según los gastos y costos en los que incurren estas empresas son menores dado que por energía realizan pagos desde \$50.702 hasta \$123.363 y por mantenimiento y reparaciones desde \$66 hasta \$249.118. Adicionalmente, estas efectúan mayores compras de materia prima internacional por montos alrededor de \$8.435.302 y \$423.380.736 y en cuestión de materia prima nacional los montos comprados son desde \$254.537 hasta \$30.892.748. De acuerdo con las existencias que presentan estas organizaciones son menores a las del *clúster 1* debido a que por existencias de producto en proceso estas tienen rangos desde \$3.590 hasta \$4.743, por productos terminados rangos entre \$1.056 y \$1.257.980, en cuanto a productos no fabricados por el establecimiento se establecen desde \$104.933 hasta \$629.735. Los valores en libros de estas organizaciones suelen ser más altos en comparación al *clúster 1* puesto que por terrenos contienen valores desde \$39.728 hasta \$1.739.903, en edificios valores desde \$220.668 hasta \$9.801.864 y en maquinaria y equipos valores desde \$1.049.650 hasta \$8.521.9096. Los valores en libros de estas organizaciones suelen ser más altos en comparación al *clúster uno* puesto que por terrenos contienen valores desde \$39.728 hasta \$1.739.903, en edificios valores de \$220.668 y \$9.801.864, finalmente la maquinaria y equipos propone valores entre \$1.049.650 y \$8.521.9096.

Es importante destacar aquellas empresas que contienen una pertenencia proporcional al *clúster uno* y *clúster dos*, estas empresas se clasifican como comercializadoras, micro, pequeñas y grandes las cuales presentan altas ventas, pero una baja producción bruta generando mejores ganancias. Teniendo en cuenta lo anterior, determina que estas empresas pueden ser parte del *clúster dos*, pero a su vez, contienen una participación del *clúster uno* debido a que presentan altas existencias de producto en proceso, producto terminado, materia prima y productos no fabricados

por el establecimiento. Adicionalmente, estas organizaciones incurren en altos costos y gastos por los servicios que adquieren lo cual genera una afinidad por el clúster dos.

9.2.1.3 Agrupaciones año 2017

Clúster 1. El clúster conformado en este año presenta empresas que se denominan micro, pequeñas y medianas las cuales se caracterizan por contener mayores ganancias a razón de que a pesar de que sus ventas no sean las más altas no incurren en elevados costos de producción generando un mayor valor en ganancias. El clúster conformado en este año las empresas denominadas micro, pequeñas y medianas se caracterizan por contener mayores ganancias a razón aun cuando sus ventas no sean las más altas, del mismo modo, no incurren en elevados costos de producción generando un mayor valor en ganancias. Estas empresas presentan menores compras respecto a la materia prima comprada, a modo que, las compras de materia prima internacional varían entre los valores de \$527.986 hasta \$28.728.131 y referente a las compras de materia prima nacional estas varían desde \$481.330 hasta \$21.718.181. Según los costos y gastos que estas organizaciones pagan son menores en comparación al clúster dos dado que por el servicio de energía cancelan valores desde \$6.341 hasta \$246.982, por mantenimiento y reparaciones valores entre \$3.987 y \$242.238 y por arrendamiento de maquinaria valores desde \$1.240 hasta \$76.553. Teniendo en cuenta las existencias al final del año, estas empresas tienen valores bajos en relación con el clúster dos a razón de que las existencias de productos en proceso muestran rangos desde \$49.336 hasta \$252.197, por productos terminados desde \$32.273 hasta \$5.289.126 y por productos no fabricados por el establecimiento valores desde \$78.435 hasta \$336.633.

Clúster 2. Esta agrupación está compuesta por empresas que se categorizan como comercializadoras, micro, pequeñas, medianas y grandes, en relación con el clúster uno estas generan mayores ventas, pero incurren en más producción bruta lo que causa una menor ganancia. Esta agrupación está compuesta por empresas que se categorizan como comercializadoras, micro, pequeñas, medianas y grandes, en relación con el clúster uno que generan mayores ventas, pero incurren en más producción bruta lo que causa una menor ganancia. Al observar las compras de materia prima que realizan estas organizaciones son más altas en comparación a la otra agrupación, las compras de materia prima internacional son mayores a las compras de materia prima nacional puesto que por materia prima internacional efectúan pagos desde \$7.331.609 hasta \$3.601.549.00 y por materia prima nacional pagos desde \$711.882 hasta \$631.451.485. Al observar las compras de materia prima realizadas por estas organizaciones son más altas en comparación con otras empresas, las compras de materia prima internacional son mayores a las compras de materia prima nacional puesto que por materia prima internacional efectúan pagos desde \$7.331.609 hasta \$3.601.549.00 y por materia prima nacional pagos desde \$711.882 hasta \$631.451.485. Conforme a los gastos y costos a los que incurre este clúster se refleja que son mayores a los pagados por el clúster uno, de acuerdo con el servicio de energía van desde los \$49.115 hasta \$270.632, por el servicio contratados con terceros abarca desde \$6.954 hasta \$184.273, por mantenimiento y reparaciones \$22.870 y \$22.010.55, por arrendamiento de maquinaria sin opción de compra desde \$2.259 hasta \$26.456. Este conglomerado presenta mayores existencias de productos terminados por \$3.058 hasta \$61.003.202, de productos en proceso por \$1.828 hasta \$82.343.567 y por productos nos fabricados por el establecimiento valores desde \$159.142 hasta \$3.481.381.

De acuerdo a las empresas que presentan pertenencia al clúster uno y clúster dos formados en este año, se determina que una parte de estas organizaciones que se categorizan como

comercializadoras, micro y pequeñas pertenecen al clúster dos debido a que generan bajas ganancias a pesar de que su producción bruta y ventas son menores a las demás empresas evaluadas, la otra parte que pertenece al clúster uno se debe a que estas empresas incurren en menores gastos y existencias de producto en proceso, producto terminado y productos no fabricados por el establecimiento.

9.2.1.4 Agrupaciones año 2018

Clúster 1. En este conglomerado tienen un alto porcentaje de membresía empresas comercializadoras, micro, pequeñas y medianas. Este clúster se caracteriza por presentar organizaciones que realizan ventas bajas a comparación de las demás empresas en estudio pero que a su vez generan mayores ganancias debido a que no incurren en gastos elevados de producción bruta. Así mismo, esta agrupación incurre en menores costos y gastos por servicios de energía las cuales varían entre \$4.762 a \$307.545, por mantenimiento y reparaciones entre \$4.860 hasta \$691.892, por arrendamiento de maquinaria sin opción de compra entre \$1.155 hasta \$219.822 y por arrendamiento de inmuebles entre \$14.400 hasta \$247.114. Teniendo en cuenta las inversiones que este conglomerado realiza son menores a las del clúster 2 dado que por terrenos invierten por \$247.114. por edificios valores entre \$511.667 hasta \$1.108.696 y por maquinaria valores entre \$34.668 hasta \$304.451, es relevante mencionar que las micro y pequeñas empresas no realizan inversiones. En cuanto a la compra de materia prima que estas empresas realizan mayores compras de materia prima internacional dentro de los rangos de \$498.197 - \$30.448.033 en comparación a las compras nacionales que van desde \$341.564 hasta \$29.534.475, aunque, compren grandes cantidades son menores a las que realiza el clúster dos. En relación a las existencias que tienen

estas empresas al final del año son menores en comparación al clúster dos debido a que por existencias de productos en proceso tienen valores de \$1.446 - \$304.545, por productos terminados desde \$27.493 hasta \$5.940.086, por productos no fabricados por el establecimiento aproximadamente \$11.244 - \$310.761 y por existencias de materia prima y materiales valores desde \$99.668 hasta \$830.089, de lo anterior es importante mencionar que las empresas comercializadoras y micro no tienen existencias al final del año.

Clúster 2. Este clúster está compuesto por las empresas que se catalogan como comercializadoras, micro, pequeñas, medianas y grandes que en comparación a las del clúster uno estas generan mayores ventas, pero con una producción bruta alta la cual impide que se obtengan buenas ganancias. Al detallar las compras de materia prima que realizan estas organizaciones se determina que estas ejecutan mayores compras con relación a las demás ya que pues por materia prima internacional efectúan compras desde \$7.680.045 hasta \$5.060.731.600 y por materia prima nacional valores desde \$831.473 hasta \$1.100.831.014. De igual forma, estas organizaciones incurren en más costos y gastos en servicios como energía por un valor de \$48.873 hasta \$612.208, por mantenimiento y reparaciones entre \$113 - \$54.194.738, por servicios contratados con terceros desde \$12.055 hasta \$147.881, por arrendamiento de maquinaria sin opción de compra desde \$7.370 hasta \$35.523.669. y por arrendamiento de inmuebles valores de \$12.320 y \$2.517.538. Lo anterior, se debe a que estas empresas efectúan mayores inversiones en cuanto a terrenos por \$136.047, edificios por rangos alrededor de \$449.912 - \$692.628 y en maquinaria desde \$22.157 hasta \$153.493.028. Conforme a las existencias de productos al final del año 2018, este clúster suele tener mayores valores a causa de que por productos en proceso se tienen valores desde \$801 hasta \$74.261.618, por productos terminados valores desde \$2.069 hasta \$94.597.531, por

productos no fabricados por el establecimiento se presenta el rango de \$72.441 - \$5.659.459 y por materia prima desde \$854.007 hasta \$35.295.280.

En cuanto a las empresas que presentan porcentaje de membresía similar en ambas agrupaciones, estas se categorizan como comercializadoras y pequeñas. Estas organizaciones tienen pertenencia al clúster dos, puesto que al generar pocas ventas y baja producción no obtienen buenas ganancias, adicionalmente, estas realizan mayores compras de materia prima, pero, incurren en pocos gastos e inversiones lo cual genera que tenga pertenencia al clúster uno.

9.2.1.5 Agrupaciones año 2019

Clúster 1. La agrupación conformada en este año se denomina por aquellas empresas categorizadas como comercializadoras, micro, pequeñas y medianas que tienen un alto nivel de membresía. Se caracteriza por contener empresas organizaciones que generan ventas bajas en comparación a las demás en estudio y en donde la producción bruta es proporcional a las ventas lo cual genera que no obtengan ganancias. Dentro de las compras que estas realizan son bajas puesto que por materia prima internacional las compras varían desde \$661.255 hasta \$5.421.090 siendo valores bajos en comparación a la materia prima comprada nacionalmente ya que los valores varían entre \$471.940 y \$23.866.848, lo anterior en paralelo al clúster dos presenta menores valores. Teniendo en cuenta los gastos a los que incurren estas organizaciones son menores a los del clúster dos dado que por el servicio de energía cancelan un valor de \$6.588 y \$218.587, por mantenimiento se presente un rango de \$3.824 y \$1.102.122, por servicios contratados con terceros valores de \$6.217 - \$320.973, por arrendamiento de maquinaria pagan valores desde \$3.054 hasta \$120.877 y por arrendamiento de inmuebles valores desde \$40.551 hasta \$124.133, de lo anterior es

importante mencionar que las empresas comercializadores y micro no incurren en estos gastos. Conforme a las inversiones que estas organizaciones realizan son bajas en paralelo al clúster 2 dado que por terrenos invierten \$5.385, por edificios valores de \$136.194 y \$1.003.919 y por maquinaria \$58.166 y \$228.923. También, las existencias que estas empresas tienen son menores a razón de que por existencias de productos en procesos presentan valores desde \$4.174 hasta \$660.318, por productos terminados valores de \$50.869 - \$273.754 y por materia prima desde \$83.884 hasta \$1.082.929.

Clúster 2. Este conglomerado se caracteriza por contener empresas que se catalogan como comercializadoras, micro, pequeñas, medianas y grandes. Estas organizaciones generan mayores ventas, pero con una desventaja puesto que el costo de la producción bruta es alto y por ende no se obtienen buenas ganancias. Al observar las compras de materia prima que este clúster realiza se determina que son mayores a razón de que por materia prima internacional compran entre los valores de \$13.337.810 hasta \$5.482.782.486 y por materia prima nacional valores alrededor de \$1.159.533 y \$797.718.274. De acuerdo con los gastos incurridos por este conglomerado son elevados en comparación al clúster uno debido a los pagos de energía desde \$61.540 hasta \$832.979, por mantenimiento desde \$185 hasta \$53.931.098, por servicios contratados con terceros entre \$95.218 hasta \$11.503.230, por arrendamiento de maquinaria de \$5.419 - \$272.942 y por arrendamiento de bienes inmuebles desde \$13.715 hasta \$377.102. Los anteriores gastos se fundamentan en las inversiones realizadas por esta agrupación dado que en terrenos efectúan inversiones desde \$1.419 hasta \$895.518, en edificios por \$469 hasta \$727.490 y por maquinaria valores desde \$26.768 hasta \$190.248.441. Asimismo, las existencias que contienen estas empresas son mayores en paralelo al clúster dos a razón de que por existencias de productos en

proceso presentan valores desde \$4.496 hasta \$75.585.917, por productos terminados \$792.248 hasta \$121.311.933 y por materia prima valores alrededor de \$35.962 y \$9.184.735.

En este año existe una microempresa la cual presenta pertenencia similar al clúster uno y clúster dos, esto se genera debido a que una parte de la empresa contiene características similares al clúster uno ya que obtiene ganancias bajas con unas ventas y producción bruta bajas, adicionalmente, esta organización realiza bajas compras de materia prima y los gastos en los que incurre son proporcionales a las inversiones realizadas las cuales son bajas.

10. Análisis de productividad, competitividad y desarrollo

En este capítulo se presentan algunos de los factores asociados a la mano de obra y a la producción que inciden significativamente en la productividad, competitividad y desarrollo de las organizaciones de la industria del cuero, calzado y marroquinería según las características de la partición difusa. También, se mencionan aspectos relevantes referente a algunos factores que demuestran incidencia en una partición dura con el objetivo de contrastar un poco los resultados.

La partición Fuzzy-K-Means arroja dos grupos, en el primer grupo se tiene con las siguientes características relevantes: se obtienen ganancias aun cuando las ventas no son lo suficientemente grandes pero aun así solventan los costos y gastos de producción (consumo de energía, mantenimiento y reparaciones, arriendo de inmuebles y maquinaria, entre otras), los inventarios de productos en proceso, producto terminado y materia prima se mantienen en rangos

bajos, al igual que la inversión en maquinaria, equipo y edificios en comparación con el otro clúster borroso.

En el segundo clúster difuso se encuentran empresas con un alto porcentaje de membresía con características opuestas respecto al anterior grupo, este grupo presente altos rubros en inversiones de maquinaria, equipo y edificios, así mismo, incurre en costos más altos de mantenimiento, pero sus esfuerzos no se reflejan en sus ganancias. También, el consumo de energía y arrendamiento, mayor existencia de productos terminados, en proceso y en materia prima resultan ser valores elevados.

Es importante mencionar que este conglomerado presenta un factor que genera desventaja en las ganancias esperadas debido a que estas empresas poseen un alto precio de materias primas, materiales y empaques en bodega, lo anterior, se podría decir estas empresas no cuentan con una buena planeación de demanda para realizar el pedido óptimo de materia prima lo cual genera que se compren cantidades grandes del suministro y que no se procese en su totalidad, incrementando el volumen de producto almacenado en bodega causando elevados gastos y costos que no agregan valor.

Además, se encuentran diferentes aspectos asociados al desarrollo de las organizaciones en cuanto a su productividad y competitividad. Se logran inferir un mayor nivel de competitividad y productividad en las empresas que cuentan con características del clúster uno. Estas empresas resultan ser productivas al obtener ganancias de sus inversiones, aprovechar su cadena productiva y los insumos requeridos al generar ingresos por ventas que solventan los gastos y costos de toda la línea de producción.

Una ventaja competitiva del primer grupo respecto al segundo hace referencia al elevado consumo de energía en el que incurren algunas organizaciones en servicios de producción en el

uso de sus instalaciones. El clúster uno se enfoca en el desarrollo tecnológico y de innovación implementando maquinaria para agilizar, ahorrar dinero y optimizar los procesos productivos que se realizan para obtener un producto final con excelente calidad, razón por la cual se genera que el grupo de empresas sean más productiva.

Al mismo tiempo, se logra inferir que las empresas del grupo uno tiene mayor control sobre las inversiones destinadas a terrenos, edificios y maquinaria, razón por la cual resultan ser más productivas al contar con un nivel mayor de planeación de sus requerimientos. Los altos niveles de inversión no siempre generan efectos positivos dentro de las organizaciones, según el caso incurrir en gastos de arrendamiento de bienes y no en inversiones genera mejores efectos en las utilidades de las empresas. Altas inversiones no necesariamente conducen a altos niveles de desarrollo y competitividad, las empresas deben contar con la capacidad de maximizar la utilización de toda inversión realizada para lograr obtener ventaja respecto a la misma inversión, es decir, cualquier inversión debe ser proporcional a la utilidad obtenida.

El resultado más significativo de la agrupación borrosa resulta ser las empresas que comparten características importantes de los grupos ya descritos, es decir, empresas con porcentaje de membresía similar en ambos grupos asignadas parcialmente a las dos agrupaciones conformadas. Cabe resaltar, que estas organizaciones resultan ser las **más competitivas y productividad de todo el sector**. Las empresas mencionadas se destacan por la obtención de ganancias implementando estrategias de consumo basadas en inversiones balanceadas de activos fijos, en otras palabras, la inversión resulta ser casi proporcional a la capacidad de utilización con la que cuentan las organizaciones según su tamaño.

También, cabe mencionar la problemática por la que atraviesa el sector en cuanto a los suministros de materia prima, para algunas empresas es muy difícil la obtención de materia prima

a raíz de la falta de proveedores cercanos y con estándares de calidad, por esta razón las empresas con pertenencia en ambos grupos deben prever la escasez de suministros buscando mantener el flujo de producción. En este orden de ideas, se aprecia el control sobre las existencias en inventarios de estas empresas evitando pausas de trabajo por falta de material (cuellos de botella) o causando altos costos por almacenamiento de inventario, creando competitividad y productividad en ellas.

El consumo de servicios de energía, mantenimiento y reparación, arrendamientos de maquinaria o inmuebles y servicios tercerizados resulta ser moderado para estas empresas. Un ejemplo de lo anterior, la prevención en mantenimiento y reparación aumenta la productividad del flujo de producción y a su vez en la competitividad de hacer más con menos, los daños en maquinaria por falta de mantenimiento puede generar problemas tan grandes como la suspensión de producción por falta de equipos en funcionamiento.

A través del tiempo se observa la migración e inclusión de las empresas que figuran con características de productividad y competitividad a través del tiempo. En 2015 expone la participación de ocho empresas categorizadas como micro, pequeñas y medianas, 2016 se resalta la participación de 3 grandes empresas, en el 2017 y 2018 solo se ubican 5 y 6 micro, pequeñas y medianas empresas respectivamente; para finalizar, en el 2019 existe la participación de una microempresa. Conforme a esto, se infiere el bajo desarrollo por el que atraviesa el sector de cuero, calzado y marroquinería en estos años, la participación de las empresas que figuras como productivas y competitivas resulta ser muy baja en comparación con las demás empresas, aunque no todas se encuentran generando perdidas resultan poco eficientes y conscientes en la utilización de sus recursos.

Otros aspectos importantes encontrados dentro del sector se descifran gracias a los resultados obtenidos en la ejecución del algoritmo K-Means, este modelo, expone resultados atractivos en cuando al comportamiento de las organizaciones según la dinámica del sector y el tamaño de la empresa. En el año 2015 el mayor consumo de suministros fue nacional, a partir del 2016 las empresas comenzaron a adquirir suministros internacionales, esto se debe a que la obtención de estos suministros resulta ineficiente, por un lado, las empresas colombianas dedicadas al tratamiento de la materia prima relacionada con cuero y piel no cuenta con procesos y desarrollo tecnológico para el tratamiento de estos lo cual afecta la competitividad de las empresas dedicadas a esta actividad económica.

Del mismo modo, existen pocas curtiembres legales que realicen un buen procedimiento en la obtención de materia prima sin generar impactos negativos en el medio ambiente y entregar un material con altos estándares de calidad, del mismo modo, las empresas de encurtido no representan competitividad o productividad relevante en el mercado ya que no cuentan con los estándares internacionales de diseño y cuentan con escasa inversión en maquinaria e innovación que aumente la productividad. Las pocas empresas que se dedican a los procesos manufactureros de tratamiento del cuero destinan sus productos para que sean exportados generando que el suministro nacional sea escaso y a altos costos ya que las empresas internacionales cuentan con la maquinaria especializada para el tratamiento de esta materia. Los suministros derivados de materiales textiles también presentan inconvenientes de obtención dentro del país puesto que Colombia no cuenta con la maquinaria y el proceso adecuado para generar este insumo. De lo anterior, se producen sobre costos y retrasos en la cadena productiva del sector generando baja competitividad y productividad respecto a los mercados más competentes.

En el mismo orden de ideas, K-Means presenta en sus resultados valores altos de productos terminados que generan percepción de no competitividad y productividad dentro de las empresas, ya que se demuestra su incapacidad participativa dentro de mercado nacionales e internaciones, las grandes empresas en Colombia no demuestran tener una competitividad fuerte frente a competidores líderes como Italia, España, Brasil, China, India, Vietnam o Indonesia. Así mismo, dentro del país, se cuenta con ingreso de altos productos de contrabando generando un ambiente de competencia desleal quitándole oportunidad de desarrollo y crecimiento a las empresas del sector.

Al mismo tiempo, las empresas categorizadas “grandes” por contar con más de 201 colaboradores exponen que sus valores en libros de maquinaria y edificios tienen tendencia de crecimiento y desarrollo a través del tiempo, todo esto a causa de que este tipo de organizaciones se encuentran en constante inversión y conservación de maquinaria, ya que son conglomerados que se están intentando competir en el mercado a través de su productividad y desarrollo.

También, empresas categorizadas como micro, pequeñas y medianas generan gastos referentes al arrendamiento de bienes inmuebles ya que no cuentan con inversiones en terrenos o edificios, desarrollo e innovación. Lo anterior resulta ser un factor que influye en el bajo desarrollo, potencial y productividad de estas empresas. Estas empresas demuestran baja competitiva frente a otras empresas ya que no buscan el crecimiento para migrar a un grupo de empresas productivas.

Dentro de los apéndices 7 y 8, se presenta un análisis detallado del comportamiento del sector conforme al algoritmo de agrupamiento ejecutado.

11. Comparaciones de algoritmos de clustering

En este capítulo se exponen las diferencias encontradas en los clústers formados en una partición dura como lo es el algoritmo K-Means y una partición suave por medio de un algoritmo Fuzzy-K-Means en el sector de cuero, calzado y marroquinería a nivel nacional.

K-Means proporciona clústers muy rígidos según los valores que suministra la base de datos, es decir, crea agrupaciones estándar en función de variables como la cantidad de empleados, ventas generadas, consumo intermedio, salario de sus colaboradores, inversiones en maquinaria, equipo, edificios, y/o gastos o servicios. Este tipo de clústers cumple con características propias según el tipo de empresa que se asigne a cada grupo, en otras palabras, según la categoría de las empresas (grande, mediana, pequeña, micro, comercializadora) que conforman cada conglomerado se definen las características de este.

En los años 2015, 2016 y 2019 el algoritmo K-Means genera agrupaciones similares que mantienen la siguiente tendencia: en primer lugar, se crea un agrupamiento exclusivamente con grandes empresas, un segundo agrupamiento con micro, pequeñas y medianas empresas y finalmente un tercer conglomerado con empresas comercializadoras, micro, pequeñas y medianas.

En el mismo orden de ideas, en el año 2017 se crean agrupaciones diferentes a las anteriormente mencionadas, un clúster agrupa comercializadoras, micro, pequeñas y medianas empresas, y los dos grupos restantes constan cada uno de grandes empresas; la causa de estas dos agrupaciones resulta de la diferencia en sus observaciones, en otras palabras, aunque los dos grupos son caracterizados por empresas grandes los puntos de cada conglomerado resultan ser muy diferentes entre sí.

También, el año 2018 muestra una agrupación diferente a la descrita para otros años en estudio. Conforman un primer clúster de grandes empresas, un segundo clúster con empresas comercializadoras, micro, pequeñas y medianas; un último grupo en el que se asigna una empresa pequeña con características diferentes a las pequeñas empresas pertenecientes al grupo dos, los valores de esta empresa resultan inferiores a la pertenecientes al segundo clúster expuesto.

En cuanto a las asignaciones realizadas por una agrupación suave a través de técnicas difusas como lo es el modelo Fuzzy-K-Means las empresas demuestran una tendencia distinta en comparación con los resultados obtenidos en el algoritmo K-Means descrito anteriormente. Una primera observación importante es la alta participación de una de empresa a un clúster independiente de la categorización de las empresas, en otras palabras, las empresas tienen un grado alto de membresía a un grupo independientemente al tamaño de esta. Por el contrario, el modelado difuso muestra relevancia en la observación de gastos, inversiones y /o ganancias de cada organización. En las siguientes líneas, se describen los dos grupos difusos creados con el algoritmo Fuzzy-K-Means en la ventana de tiempo de 2015 a 2019:

En el 2015 un primer grupo cuentan con alta membresía empresas comercializadores, micro, pequeñas y grandes que figuran con baja ganancia y altos niveles de inversiones, inventarios y gastos en servicios. También, se descifra el segundo clústers donde tienen alta participación empresas comercializadoras, micro, pequeñas y medianas que constan de características opuestas a las descritas en el anterior grupo. Se observan ocho empresas que comparten características con los dos grupos, sienta categorizadas como las más difusas de este año.

Para las agrupaciones conformadas en el año 2016 se encuentra un clúster en el que las micro, pequeñas y medianas empresas constan de ganancia medio altas y bajos niveles de inversión y gastos que tienen un alto porcentaje de membresía. Por otro lado, se describe un segundo

conglomerado que nuevamente figura con empresas comercializadoras, micro y medianas con observaciones posicionando las grandes empresas en la mitad de los dos grupos creados, es decir, este tipo de empresas tienen suficientes características de los grupos para no identificarse en mayor medida con uno solo clúster.

Cinco empresas comercializadoras, micro y pequeñas en el 2017 exponen un nivel de membresía similar a los dos grupos difusos construidos. Así mismo, algunas micro, pequeñas y medianas empresas presentan alto nivel de membresía en el grupo uno con características de bajas ganancias, costos, servicios e inversiones, del mismo modo el grupo dos, presenta alta participaciones de empresas de todas las categorías que no generan muchos ingresos por ganancias ya que los valores en ventas apenas cubren los costos y gastos de estas.

Finalmente, en 2018 y 2019 se crean dos grupos que tienen un nivel alto de participación empresas comercializadoras, micro, pequeñas y medianas, y otro grupo que consta de todo tipo de empresas. Estos dos grupos se diferencian el uno del otro por el nivel de ganancias según la cantidad de financiación utilizada para sus actividades. Respecto a las empresas que constan con características de ambos grupos, en el 2018 figuran empresas comercializadoras y pequeñas y en el 2019 se sitúan únicamente microempresas.

Los resultados de las particiones descritas en este capítulo se logran apreciar en los apéndices 6 y 8.

12. Conclusiones

La metodología propuesta en este proyecto de investigación facilita la comprensión de técnicas de agrupamiento en función de aprendizaje no supervisado para futuros proyectos que busquen entender el patrón de productividad, competitividad y desarrollo en diferentes sectores económicos del país enfocados en la búsqueda de diferentes factores que generan organizaciones más eficientes.

A partir de la revisión de literatura realizada se identifican diversos autores en diferentes campos de estudio que en los últimos años han tenido interés en aplicaciones de inteligencia artificial. Especialmente se hace énfasis en técnicas de aprendizaje no supervisado como herramienta de predicción de comportamiento en diferentes aspectos empresariales que contribuyen al desarrollo de las organizaciones incrementando su productividad y competitividad. Mediante el análisis preliminar de literatura se logra identificar las variables relevantes en las que algunos autores han centrado sus estudios en pro del desarrollo de las empresas.

El agrupamiento realizado a través de técnicas nítidas donde se realiza una partición dura como el K-Means resulta extraer información muy estándar, es decir, no proporciona mucha información referente al comportamiento del sector. Las agrupaciones presentan características similares y poco relevantes; en los años 2015, 2016 y 2019 las agrupaciones de los clústers se mantuvieron, el primero conformado se caracteriza por contener empresas comercializadoras, micro, y pequeñas, el segundo grupo contiene medianas empresas y el tercer grupo grandes empresas. A diferencia de los años 2017 y 2018 donde se obtuvieron agrupaciones particulares

debido a los diferentes comportamientos de las variables, en 2017 se crean dos grupos que contienen empresas grandes y el 2018 conforma dos grupos con empresas pequeñas.

Por parte del agrupamiento Fuzzy-K-Means, extrae resultados más informativos en cuanto a la dinámica de las empresas en el sector, brindando inferencias sobre tres tipos de empresas según el grado de membresía asignado a los dos agrupamientos difusos formados. En primer lugar, se encuentran dos grupos de empresas que figuran con baja y alta ganancia, pero contiene altos niveles de inversiones, inventarios y gastos en servicios. También, se destacan empresas que no se asignan con gran nivel de participación a ninguno de los dos clústers descritos, estas se caracterizan por ser las más productivas y competitivas del sector a raíz de la eficiencia en la utilización de sus recursos para producción.

Las agrupaciones resultantes de técnicas difusas y nítidas resultan ser muy diferentes entre sí, la principal característica de las agrupaciones K-Means resulta de la categoría o tamaño de las empresas mientras que los clústers conformados por el agrupamiento difuso Fuzzy-K-Means resultan de los valores de costos, gastos e inversiones de las empresas. También, una partición suave brinda la posibilidad de conocer mejor la estructura de los datos que permite conocer la tendencia del sector.

Finalmente, los resultados del algoritmo K-Means proporcionan conocimiento a rasgos generales de las empresas según su categorización, por el contrario, los resultados difusos exponen información sobre el comportamiento y tendencia de las empresas en el sector permitiendo identificar algunos factores de productividad y competitividad. De igual forma, los resultados borrosos proponen empresas cuya pertenencia es similar a ambos clústers que presentan características que se consideran favorables al desarrollo del sector.

A razón de lo anterior, se infiere un bajo grado de competitividad y productividad el sector de cuero, calzado y marroquinería en el territorio colombiano debido a la baja tasa de empresas que se localizan entre los dos clústers difusos propuestos.

13. Recomendaciones

Para futuras investigaciones se sugiere:

Impulsar este tipo de técnicas de agrupamiento desde la comparación de diferentes métricas de distancia o parámetros internos de inicialización del algoritmo para conocer el cambio de los resultados y de esta forma obtener un panorama más amplio a través del contraste.

Ejecutar diferentes técnicas para la validación de la cantidad de clústers a conformar, con el objetivo de verificar qué técnica se asocia más al tipo de datos implementados y de esta forma tener más certeza sobre la cantidad de clústers óptimos a conformar.

Los microdatos de la encuesta anual manufacturera se encuentran sujetos a la formalidad de las empresas y la participación de estas en la encuesta, razón por la cual ampliar la muestra con la colaboración de más empresas mostrará mejores resultados más acertados al mecanismo real del sector.

La encuesta anual manufacturera del DANE no propone mucha información referente a la mano de obra del sector, razón por la cual para siguientes investigaciones se recomienda fusionar diferentes data set que contengan más variables con relación al mecanismo de la mano de obra en

el sector tanto a nivel interno como externo y de esta forma obtener mejores resultados que proporcionen un mejor entendimiento de este.

Para un estudio con más información se recomienda que los autores se encarguen de la recolección de datos a través de la encuesta con el objetivo de realizar seguimiento a todas las empresas que proporcionen a la información.

las técnicas de aprendizaje no supervisado han demostrado su capacidad en el estudio de la tendencia en la industria manufacturera de cuero, calzado y marroquinería. Razón por la que se recomienda la integración de otras técnicas en algunos de los procesos internos de sus empresas buscando mejoramiento en toda su cadena de valor.

Para nuevas investigaciones relacionadas con el sector manufacturero en función de su desarrollo a través del incremento en su competitividad y productividad se recomienda indagar en nuevas variables no presentadas en este proyecto o en la implementación de técnicas de aprendizaje no supervisado en sectores diferentes al cuero, calzado y marroquinería. Lo anterior en busca de potencializar toda la industria manufacturera colombiana.

Con el objetivo de ampliar los resultados de esta investigación y proponer conclusiones más completas, se recomienda realizar una investigación donde se tengan en cuenta factores externos que permitan conocer la incidencia de la economía y las políticas del país en la industria.

Para el sector cuero, calzado y marroquinería en Colombia se sugiere:

En la comprensión de mecanismo de este sector dentro del territorio nacional se localizan diferentes puntos débiles internos y externos que le impiden desarrollar un alto grado de competitividad y productividad. Algunos datos relacionados con lo anterior resultan ser los productos de contrabando a bajo costo que ingresa el país, la dificultades para las empresas en la

adquisición de suministros, la falta de innovación en procesos, algunas políticas de tratado de libre comercio y demás.

Actualmente las empresas no cuentan con estandarización en la mayoría de sus procesos o con un nivel de desarrollo que las impulse en el mercado, a razón de esto, se recomienda incentivar la conformación de grupos empresariales con el objetivo de ayudarse entre sí a integrar toda su cadena productiva, comenzando desde la adquisición de materias primas hasta la comercialización de los productos terminados. Los conglomerados que se logren conformar abrirán camino a la inversión, retroalimentación de técnicas, información y/o habilidades. La conformación de clústers contribuye al crecimiento individual y conjunto, a su vez reduce el riesgo, consta de proyección internacional y establece eslabones de la cadena productiva más fuertes.

Referencias Bibliográficas

ACICAM. (2018). *Cómo va el sector? ACICAM*. <https://acicam.org/como-va-el-sector/>

ACICAM. (2019). *Cómo va el sector? Asociación colombiana de industriales del calzado, el cuero y sus manufacturas*. <https://acicam.org/como-va-el-sector/>

Amany, M. A., Sondoss, E. S., Ripon, K. C., & Hasan Hüseyin, T. (2022). A Joint Optimization of Strategic Workforce Planning and Preventive Maintenance Scheduling: A Simulation–Optimization Approach. *ELSEVIER*, 219. <https://doi.org/https://doi.org/10.1016/j.res.2021.108175>

Aragon, A., & Sanchez, M. (2005). Strategic Orientation, Management Characteristics, and Performance: A Study of Spanish SMEs. *Small Business Management*. https://www.researchgate.net/publication/227686269_Strategic_Orientation_Management_Characteristics_and_Performance_A_Study_of_Spanish_SMEs

Araneda, P. (2021). *RPubs by RStudio. Imputación de Datos*. <https://rpubs.com/paraneda/imputacion>

Atiku, S. O., & Obagbuwa, I. C. (2021). Machine Learning Classification Techniques for Detecting the Impact of Human Resources Outcomes on Commercial Banks Performance. *Ridha Ejwali*. <https://www.hindawi.com/journals/acisc/2021/7747907/>

Barrientos, J., Tobón, D., & Gutiérrez, E. (2007). Sobre la eficiencia de la industria del cuero en los sectores de calzado y marroquinería en Colombia. *Universidad Antioquia*, 26. https://www.banrep.gov.co/sites/default/files/eventos/archivos/eficienciaUdeA_0.pdf

- Benavidez, K., Juliana, & Jimenez, L. B. (2022). *Algoritmos de agrupamiento*.
<https://github.com/KarolJulianaBenavidez/Algoritmos-de-agrupamiento>
- Bernal Gonzalez, Y. (2017). *Clúster del sector del cuero, calzado y marroquinería en Bogotá*.
<https://repository.usta.edu.co/handle/11634/9931>
- Calderón Campos, J.A., & Santo Tomas, U. (2018). *Influencia de la Asociatividad para el Fomento del Clúster en la Industria de Cuero y Calzado en el Área Metropolitana de Bucaramanga*
<https://repository.usta.edu.co/bitstream/handle/11634/13060/2018johanacalderon.pdf?sequence=1&isAllowed=y>
- Catela, Y., Cimoli, M., & Porcile, G. (2012). Productivity and structural heterogeneity in the Brazilian manufacturing sector: trends and determinants Productivity and structural heterogeneity in the Brazilian manufacturing sector.
- Cazares Aguirre, M. S., & Martínez Gómez, P. (2018). *Guía para la implementación de un algoritmo de clustering con apache spark utilizando el lenguaje Scala en un entorno Big data: caso de estudio K-Means*. Universidad Autónoma de Ciudad Juárez, 91.
[http://erecursos.uacj.mx/bitstream/handle/20.500.11961/4660/PROYECTO K-MEANS.pdf?sequence=1&isAllowed=y](http://erecursos.uacj.mx/bitstream/handle/20.500.11961/4660/PROYECTO_K-MEANS.pdf?sequence=1&isAllowed=y)
- Chen, W., & Panahi, M. (2017). *Evolution*.
- Colombia productiva (2021). *¿Qué es el sector de cuero calzado y marroquinería?*
<https://www.colombiaproductiva.com/ptp-sectores/manufactura/cuero-calzado-marroquineria>
- Córdoba, M. A. (2014). *Herramientas estadísticas para el monitoreo y uso de la variabilidad espacial del rendimiento y propiedades de suelo intralote*.

- <https://1library.co/article/nálisis-cluster-fuzzy-means-aproximaciónmultivariada-análisis-espacial.y8p9m42z>
- Dagnino S, J. (n.d.). *Datos faltantes (missing values). datos faltantes (missing values)*.
<https://revistachilenadeanestesia.cl/datos-faltantes-missing-values/>
- Díaz Muñoz, D. A., Quintana Lombeida, M. D., & Fierro Mosquera, D. G. (2020). La competitividad como factor de crecimiento para las organizaciones. *INNOVA*, 17.
<https://dialnet.unirioja.es/servlet/articulo?codigo=7878906>
- Díaz Ramírez, J. (2021). *Aprendizaje Automático y Aprendizaje Profundo*. Editorial.
www.mckinsey.com/business-functions/sustainability/our-insights/artificial%02intelligence-and-the%02
- Duan, Z., Yuan, X., & Zhu, R. (2020). *Energy big data demand prediction model*.
<https://dl.acm.org/doi/abs/10.3233/JIFS-189014>
- Fayyad, U., & Piatetsky-Shapiro G Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *From Data Mining to Knowledge Discovery in Databases*.
<https://ojs.aaai.org//index.php/aimagazine/article/view/1230>
- Gurbuz, S., & Mert, I. (2011). Impact of the strategic human resource management on organizational performance: Evidence from Turkey. *International Journal of Human Resource Management*, 1803–1822.
<https://doi.org/https://doi.org/10.1080/09585192.2011.565669>
- Herrera Fontalvo, T., Granadillo, E. D. L. H., & Morelos Gómez, J. (2018). *La productividad y sus factores: incidencia en el mejoramiento organizacional*.
http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S1692-85632018000100047#B22

- Hinestroza, D., & Cárdenas, J. M. (2018). *El machine learning a través de los tiempos*.
<https://repository.unilibre.edu.co/bitstream/handle/10901/17289/EL%20MACHINE%20LEARNING.pdf?sequence=1&isAllowed=y>
- Huertas, A. (2020). *2020alexanderhuertas*.
- Jiang, D., Lin, W., & Raghavan, N. (2020). *Recommendation Algorithm Based on Machine Learning* <https://doi.org/10.1155/2021/8387277>
- Khalid, M., Rahmani, I., Pal, N., Arora, K., & Scholar, M. T. (2014). Clustering of Image Data Using K-Means and Fuzzy K-Means. *International Journal of Advanced Computer Science and Applications*. <https://thesai.org/Publications/IJACSA>
- Kumar, S. (2021). *Fuzzy C-Means Clustering —Is it Better than K-Means Clustering? Towards Data Science*. <https://towardsdatascience.com/fuzzy-c-means-clustering-is-it-better-than-k-means-clustering-448a0aba1ee7>
- Lee, G. M., & Gao, X. (2021). A Hybrid Approach Combining Fuzzy c-Means-Based Genetic Algorithm and Machine Learning for Predicting Job Cycle Times for Semiconductor Manufacturing. *MDPI*. <https://www.mdpi.com/2076-3417/11/16/7428>
- Liu, J., Huang, J., Wang, T., Xing, L., & He, R. (2021). *Data driven*. <https://ieeexplore.ieee.org/abstract/document/9319164>
- Maldonado, H., Abad, A. V., General, S., & Técnicos, D., Efraín, E Delgado, F. (2009). *Metodología PEE Sectorial*
https://www.dane.gov.co/files/planificacion/planificacion/metodologia/Metodologia_PEE_Sectorial_2009.pdf
- Martínes De Ita, M. E. (2019). El concepto de productividad en el análisis económico. *Academia*, 34. <https://d1wqtxts1xzle7.cloudfront.net/38054962/productividad-laboral-with-cover->

- page-
v2.pdf?Expires=1667234323&Signature=HmapAiLK163ftzhKs1O1ya24CXJENQTB4ky
8QpKATHhWnDJge0dVqz0AhQHmS57U0K4~a~ZDl0QUdO2eZTZ6ko7osB1myvlyET
ZinjJcViigK27zmOS7BZy7T4n5qMDsSn0ApW
- Meiners, M., Mayr, A., Lechler, T., & Franke, J. (2019). Accelerated Production Ramp-Up Utilising Clustering and Visualisation of Process Chain Interrelationships. *Eeie*.
<https://ieeexplore.ieee.org/document/9011827>
- Molina, M., & Hernández, U. (1999). Función financiera y estrategia competitiva de la empresa. *Investigaciones Europea*. <https://dialnet.unirioja.es/servlet/articulo?codigo=785054>
- Perez Zúñiga, R., Camacho Castillo, O., & rroyo Cervantes, G. (2014). el incremento de la productividad y competitividad en México: Innovación, conocimiento y desarrollo. *Revista de Tecnología y Sociedad*.
https://www.researchgate.net/publication/310005512_El_incremento_de_la_productivida_d_y_competitividad_en_Mexico_Innovacion_conocimiento_y_desarrollo
- Porter, M. (1985). *Ventaja competitiva: creación y sostenimiento de un desempeño superior*.
- Powerdata. (2016). *Calidad de datos en minería de datos a través del preprocesamiento*.
<https://blog.powerdata.es/el-valor-de-la-gestion-de-datos/calidad-de-datos-en-mineria-de-datos-a-traves-del-preprocesamiento>
- Quintero Castillo, Y. (2020). *Cadenas globales de valor en Colombia: un análisis desde el sector cuero, calzado y marroquinería* [Universidad de La Salle].
<https://ciencia.lasalle.edu.co/cgi/viewcontent.cgi?article=2655&context=economia>

- Raja, M., Vellaisama, P., Krishnan, S., & Rajendran, L. (2021). *Enhanced framework for ensemble effort estimation by using recursive-based classification*. <https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/sfw2.12020>
- Ramírez, J. A., Sarmiento, H. O., & López-Lezama, J. M. (2017). Diseño de un Clasificador Difuso para el Establecimiento de los Estados Funcionales de un Sistema de Producción de Aire Medicinal. *Scielo*. <https://doi.org/http://dx.doi.org/10.4067/S0718-07642017000600016>
- Rouhiainen, L. (2018). *Inteligencia artificial*. Alienta Editorial.
- Salazar, I. (2021). *El año de la inteligencia artificial*.
- SCIKIT LEARN. (n.d.-a). *Sklearn.Preprocessing.StandardScaler*. <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html>
- SCIKIT LEARN. (n.d.-b). *Sklearn.Cluster.KMeans*. <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html>
- SENA. (2016). *Boletín mesas sectoriales sector cuero y calzado*. https://observatorio.sena.edu.co/Content/pdf/mesas_sectoriales/cuero_calzado_marroquineria.pdf
- Sfikit Fuzzy. (n.d.). *Skfuzzy*. <https://pythonhosted.org/scikit-fuzzy/index.html>
- Tudashboard. (2020). *¿Qué Es Un Análisis Descriptivo?* <https://tudashboard.com/que-es-un-analisis-descriptivo/>
- Wang, Z., & Wang, S. (2021). *Human Resource Allocation Based on Fuzzy Data Mining Algorithm*. Zhihan Lv. <https://www.hindawi.com/journals/complexity/2021/9489114/>
- Xiaofan, C., & Fengbin, W. (2010). Application of Data Mining on Enterprise Human Resource Performance Management. *IEEE*. <https://www.ieee.org/sitemap.html>
- Xie, Q. (2020). *Machine learning in human resource system of intelligent manufacturing industry*.

Zhang, Y., Wa, S., Sun, P., & Wang, Y. (2021). Pear Defect Detection Method Based on ResNet and DCGAN. *MDPI*. <https://www.mdpi.com/2078-2489/12/10/397>

Zhu, H. (2021). Research on Human Resource Recommendation Algorithm Based on Machine Learning. *Scientific Programming for Smart Internet of Things*, 2021, 10. <https://www.hindawi.com/journals/sp/2021/8387277/>