**UNIVERSIDAD MAYOR DE SAN ANDRÉS**

**FACULTAD DE CIENCIAS PURAS Y NATURALES**

**POSTGRADO EN INFÓRMATICA**

**DIPLOMADO EN ANÁLISIS DE DATOS Y BUSINESS INTELLIGENCE**

**MODALIDAD VIRTUAL**

**GESTIÓN 2024**

Logotipo

Descripción generada automáticamente

**MONOGRAFÍA DE GRADO**

**ANÁLISIS DE EXPORTACIÓN DE PRODUCTOS NO TRADICIONALES MEDIANTE MODELOS DE CLUSTERIZACIÓN Y REGLAS DE ASOCIACIÓN PARA ABORDAR LA ESCASEZ DE DIVISAS**

**POR:** LIC. YAMIL GAMARRA LOPEZ

**TUTOR:** M. SC. MARCELO PALMA SALAS

LA PAZ – BOLIVIA

Noviembre, 2024

**ÍNDICE**

[CAPÍTULO I 1](#_Toc183903329)

[MARCO INTRODUCTORIO 1](#_Toc183903330)

[CAPÍTULO 1 MARCO INTRODUCTORIO 2](#_Toc183903331)

[1.1. Antecedentes del Problema 2](#_Toc183903332)

[1.2. Problema de Investigación 2](#_Toc183903333)

[1.2.1. Formulación del Problema 2](#_Toc183903334)

[1.3. Justificación 3](#_Toc183903335)

[1.3.1. Justificación Teórica 3](#_Toc183903336)

[1.3.2. Justificación Práctica 3](#_Toc183903337)

[1.3.3. Pertinencia Social 3](#_Toc183903338)

[1.4. Objeto de Estudio 3](#_Toc183903339)

[1.5. Objetivos 4](#_Toc183903340)

[1.5.1. Objetivo General 4](#_Toc183903341)

[1.5.2. Objetivos Específicos 4](#_Toc183903342)

[CAPÍTULO II 5](#_Toc183903343)

[MARCO TEÓRICO 5](#_Toc183903344)

[CAPÍTULO 2 MARCO TEÓRICO 6](#_Toc183903345)

[2.1. Contexto económico en Bolivia y la Escasez de Divisas 6](#_Toc183903347)

[2.2. Crecimiento económico y exportaciones No tradicionales 7](#_Toc183903348)

[2.2.1. Diversificación de Exportaciones 8](#_Toc183903349)

[2.2.2. Exportación de productos No tradicionales 9](#_Toc183903350)

[2.2.3. Relevancia del análisis de exportaciones no tradicionales 10](#_Toc183903351)

[2.3. Minería de Datos y Machine Learning en el Análisis Económico 11](#_Toc183903352)

[2.3.1. Importancia en el Análisis Económico 13](#_Toc183903353)

[2.3.2. Modelos de Minería de Datos y la diversificación de Exportaciones 14](#_Toc183903354)

[2.3.3. Aplicaciones de la Minería de Datos en la Diversificación de Exportaciones 15](#_Toc183903355)

[2.4. Modelos de Clusterización 16](#_Toc183903356)

[2.4.1. Algoritmo K-means 17](#_Toc183903357)

[2.4.2. Método del codo 18](#_Toc183903358)

[2.4.3. Coeficiente de Silueta (Silhouette Index) 19](#_Toc183903359)

[2.5. Métodos de reducción de dimensionalidad 20](#_Toc183903360)

[2.5.1. PCA (Análisis de componentes principales) 21](#_Toc183903361)

[2.5.2. Análisis de correlación de variables 22](#_Toc183903362)

[2.6. Reglas de Asociación 22](#_Toc183903363)

[2.6.1. Algoritmo Apriori 24](#_Toc183903364)

[2.6.2. Validación de las reglas de asociación 25](#_Toc183903365)

[CAPÍTULO III 28](#_Toc183903366)

[MARCO APLICATIVO 28](#_Toc183903367)

[CAPÍTULO 3 MARCO APLICATIVO 29](#_Toc183903368)

[3.1. Configuración del entorno de trabajo 29](#_Toc183903370)

[3.2. Descripción del conjunto de datos 29](#_Toc183903371)

[3.3. Preprocesamiento y selección de los datos 30](#_Toc183903372)

[3.3.1. Verificación y asignación del tipo de datos 31](#_Toc183903373)

[3.3.2. Valores faltantes 31](#_Toc183903374)

[3.3.3. Campos corruptos 32](#_Toc183903375)

[3.3.4. Outliers (valores sesgados de los datos) 32](#_Toc183903376)

[3.3.5. Selección de columnas relevantes 33](#_Toc183903377)

[3.3.6. Transformación y escalado 35](#_Toc183903378)

[3.4. Minería de datos 35](#_Toc183903379)

[3.4.1. Método del codo 35](#_Toc183903380)

[3.4.2. Algoritmo: K-Means 36](#_Toc183903381)

[3.4.3. Análisis de clústeres 37](#_Toc183903382)

[3.5. Reglas de asociación 38](#_Toc183903383)

[3.5.1. Algoritmo Apriori 39](#_Toc183903384)

[3.5.2. Filtrado de las reglas de asociación 40](#_Toc183903385)

[3.6. Evaluación de los modelos 44](#_Toc183903386)

[3.6.1. Coeficiente de silueta 44](#_Toc183903387)

[CAPÍTULO IV 45](#_Toc183903388)

[CONCLUSIONES Y RECOMENDACIONES 45](#_Toc183903389)

[CAPÍTULO 4 CONCLUSIONES Y RECOMENDACIONES 46](#_Toc183903390)

[4.1. Conclusiones 46](#_Toc183903392)

[4.2. Recomendaciones 47](#_Toc183903393)

[Bibliografía 48](#_Toc183903394)

[ANEXOS 51](#_Toc183903395)

[ANEXO I 52](#_Toc183903396)

[ANEXO II 53](#_Toc183903397)

[ANEXO III 54](#_Toc183903398)

[ANEXO IV 55](#_Toc183903399)

**ÍNDICE DE FIGURAS**

[**Figura 1** Exportaciones públicas y privadas 6](#_Toc183903400)

[**Figura 2** Actividad Económica enero - junio 2024 7](#_Toc183903401)

[**Figura 3** Principales productos No Tradicionales (Expresados en millones de dólares) 10](#_Toc183903402)

[**Figura 4** ¿Qué es la minería de datos? 12](#_Toc183903403)

[**Figura 5** Machine learning 12](#_Toc183903404)

[**Figura 6** Antes y despues de aplicar K-means 18](#_Toc183903405)

[**Figura 7** Método del codo 19](#_Toc183903406)

[**Figura 8** Ítems y transacciones 23](#_Toc183903407)

[**Figura 9** Visualización de tipos de datos del dataset 31](#_Toc183903408)

[**Figura 10** Visualización de valores nulos 31](#_Toc183903409)

[**Figura 11** Visualización de fila [9451] corrupta 32](#_Toc183903410)

[**Figura 12** Visualización de fila [9451] corregida 32](#_Toc183903411)

[**Figura 13** Histograma de distribuciones sesgadas 33](#_Toc183903412)

[**Figura 14** Matriz de Correlación 34](#_Toc183903413)

[**Figura 15** Varianza acumulada PCA 35](#_Toc183903414)

[**Figura 16** Método del codo de la data final 36](#_Toc183903415)

[**Figura 17** Clúster Valor de exportación VS Peso bruto de productos 37](#_Toc183903416)

[**Figura 18** Distribución de VALOR por cada clúster 39](#_Toc183903417)

[**Figura 19** Reglas de Asociación 40](#_Toc183903418)

[**Figura 20** Reglas de Asociación Filtradas 42](#_Toc183903419)

**ÍNDICE DE TABLAS**

[**Tabla 1** Breve Descripción del conjunto de datos 30](#_Toc183903420)

**Resumen**

Esta monografía tiene como objetivo analizar las exportaciones de Bolivia, en particular en los productos no tradicionales, con la aplicación de modelos de clusterización y reglas de asociación para identificar patrones que puedan contribuir para el análisis y mitigar la escasez de divisas en el país. En la actualidad Bolivia ha estado enfrentando una creciente escasez de divisas, lo que ha afectado negativamente a varios sectores económicos. La monografía utiliza modelos de minería de datos, como la clusterización y reglas de asociación, para explorar patrones en los datos de exportación, identificando asociaciones entre productos que podrían captar agrupaciones con características similares y para tomar decisiones estratégicas para optimizar las exportaciones. Mediante la aplicación del modelo K-means, el estudio agrupa los productos según características relevantes, y las reglas de asociación con el algoritmo Apriori se utilizaron para descubrir relaciones clave entre los productos exportados. Los resultados sugieren enfocarse en mercados para productos no tradicionales ya que estos tienen un buen potencial y podría ayudar a aumentar la eficiencia de las exportaciones y así reducir la dependencia de los sectores tradicionales, no solucionando pero si contribuyendo a la estabilización de la economía de Bolivia y las reservas de divisas.

**Palabras clave:** Análisis de exportaciones, Escasez de divisas, Modelos de clusterización, Reglas de asociación, K-means, Algoritmo A priori, Productos no tradicionales, Minería de datos, Bolivia.

**Abstrac**

This monograph aims to analyze Bolivia's exports, particularly in non-traditional products, with the application of clustering models and association rules to identify patterns that can contribute to the analysis and mitigate the shortage of foreign exchange in the country. Currently Bolivia has been facing a growing shortage of foreign exchange, which has negatively affected several economic sectors. The monograph uses data mining models, such as clustering and association rules, to explore patterns in export data, identifying associations between products that could capture clusters with similar characteristics and to make strategic decisions to optimize exports. By applying the K-means model, the study clusters products according to relevant characteristics, and association rules with the Apriori algorithm were used to discover key relationships between exported products. The results suggest focusing on markets for non-traditional products as these have good potential and could help increase export efficiency and thus reduce dependence on traditional sectors, not solving but contributing to the stabilization of Bolivia's economy and foreign exchange reserves.

**Keywords:** Exportation analysis, Foreign currency shortage, Clustering models, Association rules, K-means, Apriori algorithm, Non-traditional products, Data mining, Bolivia

**CAPÍTULO I**

**MARCO INTRODUCTORIO**

**INTRODUCCIÓN**

La economía de Bolivia se encuentra en una situación de inestabilidad a raíz de la falta de divisas, entre otros factores adicionales, situación que imposibilita sostener la estabilidad económica y satisfacer las demandas del comercio exterior. Las exportaciones de productos tradicionales como el gas natural y los minerales han exhibido una disminución tanto desde el punto de vista de los volúmenes como desde el punto de vista del valor; las exportaciones del primer trimestre del 2024 registraron una baja del 28%, asociado a la variación de los precios internacionales, situaciones climáticas y restricciones logísticas entre otros factores (INE, 2024).

Los productos no tradicionales aparecen como una posibilidad de diversificar el mercado y fortalecer la resiliencia económica, sin embargo, el sector no tradicional también atraviesa problemáticas, como ejemplo la falta de estrategias para identificar mercados potenciales y obtener el impacto que tienen en la generación de divisas. La falta de dólares también ha intensificado las limitaciones para importar insumos, afectando la competitividad de las cadenas de exportación (IBCE, 2024), de tal manera que exportaciones e importaciones caminan de la mano.

La aplicación de técnicas de minería de datos, tales como clusterización y reglas de asociación, aparecen como una solución de análisis para resolver este problema, capaces de obtener patrones ocultos en los datos de exportación, como por ejemplo relaciones entre productos y agrupamiento en clusters en base a ciertas características que permitirían optimizar la diversificación de mercados y maximizar el ingreso por exportaciones (Ramos Serrano, 2021).

La presente monografía tiene como objetivo fundamental el análisis de las exportaciones bolivianas a partir de las técnicas de clusterización y reglas de asociación en productos no tradicionales, análisis que permitirá no sólo analizar el impacto que tiene la falta de divisas, sino que también apunta a una vía de aportación para la toma de decisiones en el comercio exterior. La importancia de este análisis está determinada por la lógica de que una economía más diversificada puede contribuir a la estabilidad económica y a la reducción de la dependencia con sectores vulnerables a las variaciones internacionales, de tal forma que se garantice la sostenibilidad del desarrollo nacional (Smith, 2022).

CAPÍTULO 1 **MARCO INTRODUCTORIO**

* 1. Antecedentes del Problema

En los últimos años, Bolivia se ha visto sometida a una creciente escasez de divisas, que afecta para mal a la economía de muchos sectores económicos. Las exportaciones bolivianas para el primer trimestre del 2024 llegaron a un total de 1.902 millones de dólares, lo que supuso una baja del 28% en el valor y de 11% en el volumen respecto al 2023 (INE, 2024)

Esta baja se explica a través de varios factores como la caída de precios, la sequías y bloqueos, lo que profundiza la crisis económica y representan desafíos para revertir el daño a la economía boliviana (Infobae, 2024).

La escasez de dólares ha propiciado dificultades en la importación para los insumos y bienes de capital para la producción y exportaciones de bienes no tradicionales. Esta situación ha derivado una crisis que afecta a pequeños y grandes productores, mostrando la vulnerabilidad del país a las condiciones del mercado internacional (Ips Noticias, 2024).

* 1. Problema de Investigación

La escasez de divisas afecta a los sectores tradicionales y a los no tradicionales, impidiendo que el país participe fácilmente en mercados internacionales, además de llevar un plazo en la satisfacción de la demanda de insumos importados. El hecho de que el país no cuente con un análisis sistemático de la información disponible sobre exportaciones obstaculiza la posibilidad de construir mejores estrategias para contrarrestar esta situación (Los Tiempos, 2024).

* + 1. Formulación del Problema

¿Cómo pueden las técnicas de clusterización y reglas de asociación optimizar las exportaciones bolivianas y contribuir a mitigar la escasez de divisas?

* 1. Justificación
     1. Justificación Teórica

La implementación de técnicas de minería de datos, como la clusterización y las reglas de asociación, permite descubrir patrones ocultos en grandes volúmenes de datos. En las exportaciones bolivianas, su aplicación permite descubrir tales asociaciones entre productos y mercados que guardan relación en las decisiones estratégicas del país para diversificar y optimizar las exportaciones (El deber, 2024).

* + 1. Justificación Práctica

La implementación de estas técnicas en el análisis de las exportaciones bolivianas permitirá a los políticos, hombres de negocios, y personas en general, determinar oportunidades de mercado no aprovechadas y aprovechar las cadenas de suministro. Sin duda, esto es importante ya que, en la actualidad, la falta de divisas limita la capacidad de las importaciones y la producción nacional (El deber, 2024).

* + 1. Pertinencia Social

Intentar resolver la falta de divisas a través de una mejor exportación de los productos facilitará la estabilidad de la economía del país.

La economía diversificada y en crecimiento puede generar empleo, facilitar el bienestar y mejorar la calidad de vida, al tiempo que minimizará la dependencia de las actividades económicas clásicas, que son más vulnerables a las oscilaciones del mercado internacional (El deber, 2024).

* 1. Objeto de Estudio

El enfoque de la investigación de esta monografía es el análisis de las exportaciones bolivianas centrada en productos no tradicionales, y de qué manera la aplicación de técnicas de clusterización y reglas de asociación pueden ayudar a identificar patrones, así como también asociaciones que permitan disminuir la falta de divisas en el país.

* 1. Objetivos
     1. Objetivo General

Analizar las exportaciones bolivianas mediante modelos de clusterización y reglas de asociación para identificar patrones que contribuyan a mitigar la escasez de divisas en el país.

* + 1. Objetivos Específicos
* Recolectar información sobre las exportaciones de Bolivia para su análisis exploratorio, estadístico y la limpieza de datos.
* Aplicar modelos de clusterización para agrupar productos de exportación en base a características relevantes del dataset de exportaciones de Bolivia.
* Utilizar reglas de asociación para poder identificar y examinar patrones de exportación relevantes que tienen un alto potencial para la generación de divisas.
* Plantear estrategias que tengan en cuenta el conocimiento adquirido y que permitan diversificar y optimizar las exportaciones para reducir la dependencia de los productos no tradicionales.

CAPÍTULO II

MARCO TEÓRICO

CAPÍTULO 2 MARCO TEÓRICO

1. 1. Contexto económico en Bolivia y la Escasez de Divisas

Bolivia enfrenta una serie de problemas económicos. La disminución en la producción de gas, la escasez de dólares y la caída en las exportaciones han puesto a su población en una crisis que se refleja en el aumento del costo de vida y en la desesperación por obtener la divisa estadounidense al ver al boliviano (Bs), su moneda nacional, perder su valor.

El país tiene reservas internacionales cada vez más bajas (de US$15.000 millones en 2014 a US$1.800 millones en 2024, según el Banco Central de Bolivia (BCB)), situación que ha provocado una escasez de dólares en el mercado nacional y un aumento del riesgo país, según las calificadoras internacionales (infobae, 2024).

Desde ese año, el país ha experimentado unas salidas netas de divisas consecutivas en ocho años y, al no mover el tipo de cambio, el ajuste externo se realizó vía pérdida de divisas. Por tanto, la insuficiencia de divisas está asociada al hecho de que las reservas internacionales en moneda extranjera casi se agotaron cubriendo las salidas netas de divisas, como producto de que la demanda de divisas fue persistentemente mayor a la oferta de divisas, donde solo el sector público, YPFB, tiene la obligación de entregar (Loza, 2024). (ver Figura 1).

**Figura 1**Exportaciones públicas y privadas



*Fuente*: (Loza, 2024)

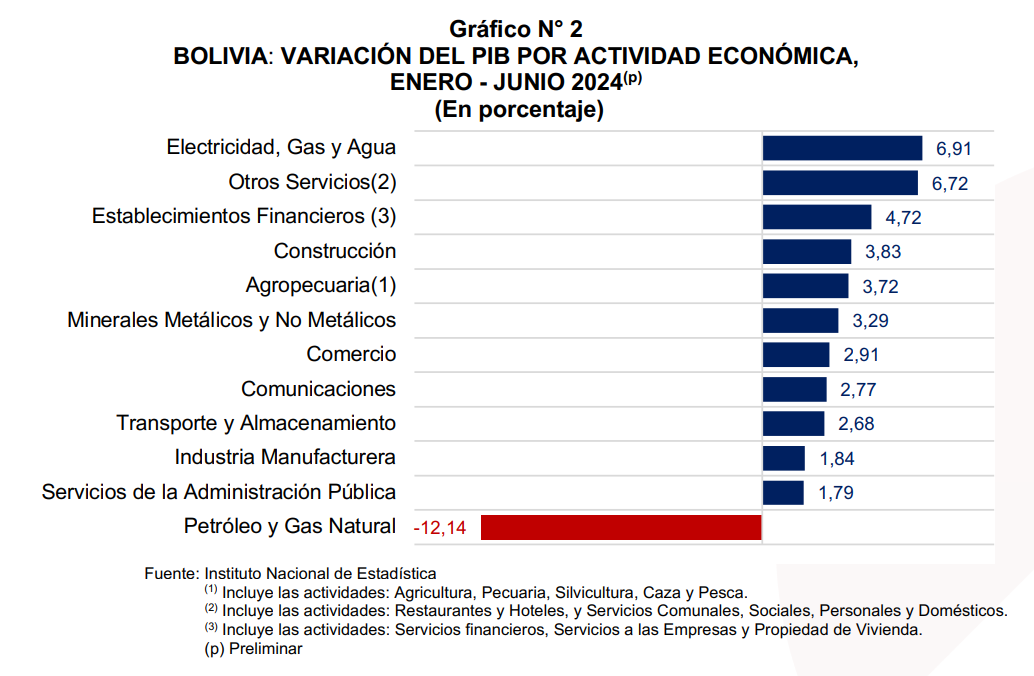
* 1. Crecimiento económico y exportaciones No tradicionales

En un contexto internacional adverso caracterizado por la desaceleración de la actividad económica mundial, traducido en problemas inflacionarios, las condiciones financieras desfavorables y la crisis climática global, la actividad económica de Bolivia registró un crecimiento de 2,58% entre enero y junio de 2024, en comparación con el mismo período del año anterior. Esta tasa es superior en 1,27 puntos porcentuales a la registrada en el primer semestre de 2023, las importaciones de bienes y servicios y la formación bruta de capital fijo registraron disminuciones de 21,25% y 8,12%, respectivamente.

Si bien las exportaciones muestran una caída de 3,85% acumulada al segundo trimestre de 2024, esta disminución es menor en comparación con la observada en el primer trimestre, debido a la recuperación del sector soyero, que enfrentó un desfase en su calendario agrícola como consecuencia de la crisis climática mundial.

Durante los primeros seis meses de 2024, las actividades económicas que más contribuyeron al crecimiento fueron electricidad, gas y agua, y otros servicios. En el primer caso, la variación fue del 6,91%, principalmente impulsada por el aumento del 8,08% en la generación de energía eléctrica. En el sector de otros servicios, la variación del 6,72% estuvo vinculada al crecimiento de los servicios comunales, sociales y personales, así como de restaurantes y hoteles, que registraron aumentos del 7,45% y 5,78%, respectivamente. Otra actividad económica destacada fue la de los establecimientos financieros, que mostraron una variación del 4,72%, impulsada principalmente por el crecimiento del 5,32% en los servicios financieros.

**Figura 2**   
Actividad Económica enero - junio 2024



*Fuente:* (INE, 2024)

Por el contrario, durante este período, la actividad económica que registró una variación de -12,14% fue la de petróleo crudo y gas natural, debido a la menor producción de gas (INE, 2024).

* + 1. Diversificación de Exportaciones

La diversificación de exportaciones es un factor clave para fortalecer la estabilidad económica de un país, especialmente en economías en desarrollo que dependen en gran medida de sectores tradicionales, como los recursos naturales. En este contexto, la minería de datos emerge como una herramienta estratégica para identificar oportunidades de diversificación y mitigar los riesgos asociados a la dependencia económica de mercados o productos específicos (World Bank, 2020).

Una economía que tiende a exportar más de lo que importa del resto del mundo, muestra una estructura productiva lo suficientemente diversificada e industrializada como para asumir los incrementos en el ingreso y el bienestar de su mercado y del resto del mundo.

Las exportaciones serán diversificadas como reflejo de su aparato productivo y las importaciones se concentrarán en las materias primas que necesite para su producción. Por tanto, la tasa de crecimiento de esta economía será mayor a la del resto del mundo; así en la dinámica del mercado mundial podrá posicionarse como una economía desarrollada.

En contraste una economía que tiende a exportar menos de lo que importa en el comercio mundial, muestra una estructura productiva escasamente diversificada para responder a incrementos en el ingreso y bienestar de su mercado. Las exportaciones son concentradas, básicamente en materia prima, y las importaciones buscan la diversidad en el mercado mundial ausente en el suyo. Esta economía genera dependencia de la demanda de materias primas del aparato productivo de las economías diversificadas. Luego, su tasa de crecimiento económico será menor que la del resto del mundo. Entonces en el mercado mundial será una economía no desarrollada (Ibañez Choque, Gavincha Lima, & Llapaco Ávila, 2024).

La dependencia de productos de exportación tradicionales puede exponer a un país a fluctuaciones en precios internacionales y limitaciones de demanda. Según Hausmann (2014), las economías que diversifican sus exportaciones hacia sectores no tradicionales tienden a experimentar mayor resiliencia económica y crecimiento sostenido. Sin embargo, identificar nuevos productos o mercados potenciales requiere análisis detallados de grandes volúmenes de datos económicos y comerciales, una tarea para la que la minería de datos es particularmente adecuada.

* + 1. Exportación de productos No tradicionales

Las Exportaciones forma parte de la balanza comercial se la define “básicamente la demanda del resto del mundo por los bienes nacionales, como cualquier demanda, dependerán del precio y el ingreso. Si el precio de los bienes nacionales baja, el mundo demandara más de ellos. Esto es, cuando el Tipo de Cambio real sube, se necesitan menos unidades del bien extranjero para adquirir un bien nacional y si el ingreso del mundo sube, el mundo demandara más de los bienes nacionales” (FAO, 1999).

Existen varias definiciones de Exportación No Tradicional, en los años sesenta la CEPAL, señala que son aquellos considerados como productos de exportación manufacturados y nuevos, definidos conforme a la Política Económica de cada País. La Exportación No Tradicional requieren de un proceso con alto valor agregado, para los países con características extractivas en sus recursos naturales exportar este tipo de productos es importante porque estos requieren Insumos, Inversión, proceso logístico y una mayor cantidad de personas que generan puestos de trabajo directos e indirectos (FAO, 1999)

**Principales productos no tradicionales exportados**

Los productos no tradicionales estás son referentes a los bienes de industria manufactura, agroindustria y agropecuaria. Entre los principales productos No Tradicionales se enumerar de la siguiente manera: Productos agropecuarios como las Nueces de Brasil, Quinua, Bananas, Chía, Frijoles, Maní. Industria manufacturera: se encuentran la Soya y los productos de Soya, Alcohol etílico, Maderas y manufacturas de Madera, Girasol y productos de Girasol, Azúcar, Cueros y manufactura de Cuero, Productos Textiles, Carne de la especie Bovina, Palmito entre los principales. Mientras que entre los productos tradicionales de exportación estarían los de extracción de hidrocarburos y los de extracción de minerales entre los principales. (Guillermo, 2020)

**Figura 3**  
 Principales productos No Tradicionales (Expresados en millones de dólares)

Gráfico, Gráfico de líneas

Descripción generada automáticamente

* + 1. Relevancia del análisis de exportaciones no tradicionales

El análisis de las exportaciones no tradicionales, al igual que el de otros productos así calificados anteriormente, revela que su importancia para las exportaciones y la economía del país en cuestión está relacionada con el tamaño de éste y el valor de sus exportaciones. Cuantos menores sean el tamaño del país y el valor de sus exportaciones, tanto más puede ser y aumentar la importancia de las exportaciones de algunos productos no tradicionales, y viceversa. Por ejemplo, las exportaciones de camarones en Ecuador y de vino en Chile han modificado la estructura de exportación de esos países. Por otra parte, las crecientes ventas externas de café soluble no han incidido en la de Brasil. Asimismo, las dinámicas exportaciones de vino de Argentina apenas se destacaron en su estructura de exportación. Además, los países de mayor tamaño tienen, como regla, mayor número de productos que venden en el exterior y su oferta exportadora está más diversificada. (Kouzmine, 2000)

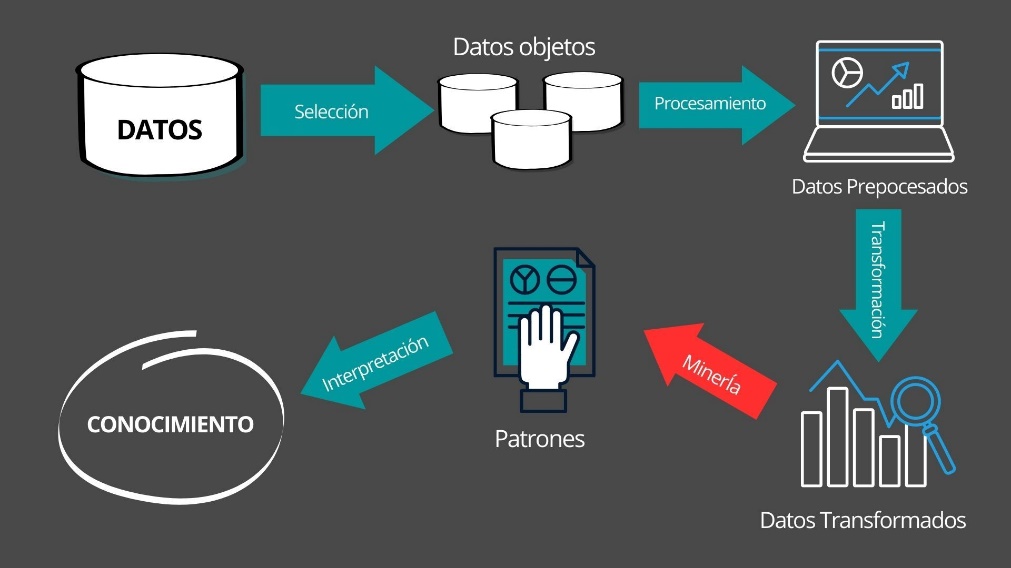
La venta externa de productos considerados no tradicionales es un elemento importante en el proceso de inserción internacional de los países exportadores de esos rubros, así como en el de la globalización de mercados. A la vez, este es un proceso en el cual se aprovechan las ventajas comparativas de los recursos naturales de cada país de la región. Por otra parte, la inserción internacional se lleva a cabo con productos de bajo nivel del valor agregado y que requieren tecnologías relativamente simples (Kouzmine, 2000).

Por otra parte, en torno de algunos productos básicos (recursos naturales) pueden estructurarse estrategias de desarrollo basadas en complejos productivos (clusters). Al mismo tiempo, esos productos suponen un proceso consecutivo de transformación de la materia prima, así como la producción de varios bienes de capital y la adopción de tecnologías modernas para llevar a cabo esa transformación, mejorar la calidad del producto y prestar servicios. Sin embargo, el proceso de transformación de los cuatro productos analizados tiene pocas etapas y actividades conexas. Por lo tanto, es improbable que se pueda crear un complejo productivo centrado en alguno de ellos (Kouzmine, 2000)

* 1. Minería de Datos y Machine Learning en el Análisis Económico

La minería de datos es un proceso que permite explorar y analizar grandes volúmenes de datos con el objetivo de descubrir patrones útiles (ver Figura 4), tendencias y relaciones significativas (Tan , Steinbach, & Kumar, 2019). Este enfoque combina diversas disciplinas, como la estadística, el aprendizaje automático y la inteligencia artificial, para transformar los datos en información accionable, relevante para la toma de decisiones, utiliza métodos de aprendizaje automático (Machine Learning) para transformar los datos en información accionable. En este contexto, Machine Learning es una subdisciplina de la inteligencia artificial (IA) que permite a las máquinas aprender de los datos, identificar patrones y hacer predicciones sin ser explícitamente programadas (Witten I & Hall, 2016).

**Figura 4**   
¿Qué es la minería de datos?



Machine Learning (ML) se puede dividir principalmente en dos tipos de aprendizaje, supervisado y no supervisado como se puede ver en la Figura 5, ambos enfoques son fundamentales para el análisis de las exportaciones bolivianas, especialmente en la diversificación de exportaciones.

**Figura 5**   
Machine learning

Diagrama

Descripción generada automáticamente

*Fuente:* (Martínez Ballesteros, 2024)

**Aprendizaje Supervisado:** En este enfoque, los modelos son entrenados con datos etiquetados (donde las respuestas son conocidas) para predecir o clasificar nuevas observaciones. Por ejemplo, modelos de regresión y clasificación pueden predecir el valor de exportaciones futuras o clasificar productos según su nivel de demanda en el mercado internacional (IBM, 2024).

**Aprendizaje No Supervisado:** En este caso, los modelos se entrenan con datos no etiquetados y se enfocan en descubrir patrones, agrupaciones o relaciones dentro de los datos. Técnicas como clusterización (por ejemplo, K-Means) y reglas de asociación (como el algoritmo Apriori) son utilizadas para segmentar mercados, identificar productos con alto potencial y descubrir patrones entre productos exportados y destinos comerciales (IBM, 2024).

En el contexto económico, la minería de datos se ha convertido en una herramienta clave para la competitividad y sostenibilidad de las organizaciones y los gobiernos. Su capacidad para identificar correlaciones y predecir tendencias a partir de grandes conjuntos de datos ha demostrado ser especialmente útil en áreas como la planificación de políticas públicas, la optimización de recursos y la identificación de oportunidades en mercados emergentes (Tan , Steinbach, & Kumar, 2019).

* + 1. Importancia en el Análisis Económico

La importancia de la minería de datos en el análisis económico se puede dividir en estas características:

**Detección de Patrones y Tendencias:** La minería de datos permite analizar datos históricos para identificar patrones económicos, como el comportamiento de los mercados, el consumo de recursos o las tendencias de exportación e importación. Esto facilita el diseño de políticas económicas basadas en evidencia (Witten I & Hall, 2016).

**Segmentación y Clusterización:** Técnicas como la clusterización permiten agrupar datos económicos según características similares, lo que puede ser útil para clasificar regiones, sectores o mercados en función de su desempeño económico o potencial de crecimiento (Pei, Han, & Tong, 2011).

**Predicción y Modelado:** Los algoritmos predictivos ayudan a prever el comportamiento del mercado o el impacto de cambios en variables económicas clave. Estas herramientas son esenciales para mitigar riesgos y aprovechar oportunidades en un entorno global dinámico (Chakrabarti & Indyk, 2009).

**Análisis de Relaciones y Dependencias:** Las reglas de asociación descubren relaciones entre variables económicas, como los vínculos entre el crecimiento de sectores específicos y las exportaciones, lo que permite tomar decisiones más informadas y estratégicas (Agrawal & Swami, 1993).

En el análisis económico de países en desarrollo, como Bolivia, la minería de datos es particularmente relevante para superar desafíos relacionados con la escasez de divisas, la volatilidad de los mercados internacionales y la falta de diversificación económica. La implementación de estas técnicas permite a los responsables de la toma de decisiones diseñar estrategias informadas que optimicen los recursos disponibles y mejoren la competitividad global.

* + 1. Modelos de Minería de Datos y la diversificación de Exportaciones

La minería de datos ha ganado relevancia en el análisis del comercio internacional debido a su capacidad para procesar y extraer patrones de grandes volúmenes de datos relacionados con transacciones, flujos comerciales, y mercados globales. Los modelos de minería de datos aplicados al comercio internacional no solo permiten identificar tendencias y oportunidades, sino también optimizar las estrategias comerciales, reduciendo riesgos y fomentando la competitividad.

Los modelos de minería de datos son métodos sistemáticos y computacionales diseñados para explorar, analizar y modelar datos. Estos modelos utilizan algoritmos para identificar patrones y relaciones entre variables, con aplicaciones específicas en el comercio internacional, tales como la segmentación de mercados, la predicción de tendencias comerciales y la optimización de cadenas de suministro (Pei, Han, & Tong, 2011).

Entre los modelos más utilizados como se puede ver en la en el comercio internacional se encuentran:

* **Modelos de Clasificación:** Utilizados para categorizar productos, mercados o socios comerciales en grupos predefinidos basados en características compartidas.
* **Modelos de Clusterización:** Herramientas para agrupar mercados o productos según similitudes, facilitando la identificación de nuevas oportunidades comerciales.
* **Reglas de Asociación:** Aplicadas para descubrir relaciones entre productos y comportamientos de consumo en diferentes mercados.
* **Análisis Predictivo:** Usado para anticipar tendencias de mercado y proyectar flujos comerciales.
  + 1. Aplicaciones de la Minería de Datos en la Diversificación de Exportaciones

La minería de datos permite extraer información valiosa de grandes bases de datos, lo que resulta esencial para tomar decisiones informadas en la planificación de exportaciones. Entre sus principales aplicaciones se destacan:

* **Identificación de Nuevos Mercados:** Utilizando técnicas como el análisis de clusterización, los datos de exportaciones pueden agruparse en mercados con características similares, permitiendo identificar regiones con alto potencial para productos no tradicionales (Pei, Han, & Tong, 2011). Esto ayuda a focalizar esfuerzos en mercados específicos que maximicen el impacto económico.
* **Clasificación de Productos con Potencial de Exportación:** Mediante algoritmos predictivos y reglas de asociación, es posible analizar el desempeño histórico de productos en mercados internacionales y correlacionarlo con características de éxito, como la demanda creciente en sectores específicos o la complementariedad con productos ya exportados (Agrawal & Swami, 1993).
* **Análisis de Competitividad:** La minería de datos puede comparar el desempeño de productos nacionales frente a sus competidores en términos de precio, calidad y accesibilidad. Esto permite diseñar estrategias para mejorar la competitividad de productos en mercados seleccionados (Tan , Steinbach, & Kumar, 2019).
* **Optimización de Recursos:** La minería de datos también puede ser utilizada para analizar cadenas de suministro y procesos logísticos, identificando cuellos de botella y optimizando la asignación de recursos. Esto es especialmente importante para productos no tradicionales, que pueden requerir ajustes en infraestructura o logística para su exportación eficiente (Witten I & Hall, 2016).
* **Clusterización para Segmentación de Mercados:** La clusterización permite agrupar mercados internacionales en función de características económicas, demográficas y geográficas. Por ejemplo, un análisis de clusterización puede ayudar a identificar grupos de países que comparten demandas similares para productos específicos, optimizando así las estrategias de exportación (Tan , Steinbach, & Kumar, 2019).
* **Reglas de Asociación para Comercio Multilateral:** Las reglas de asociación permiten identificar patrones entre productos o mercados. Por ejemplo, se pueden descubrir asociaciones entre la exportación de un producto y el aumento en la demanda de bienes complementarios en un mercado internacional (Agrawal & Swami, 1993).
* **Modelos Predictivos para Pronósticos Comerciales:** Los algoritmos predictivos, como árboles de decisión y redes neuronales, se emplean para predecir la demanda de productos en mercados emergentes, ayudando a los exportadores a ajustar su producción y logística con base en proyecciones precisas (Witten I & Hall, 2016).
* **Optimización de Cadenas de Suministro:** En el comercio internacional, los modelos de minería de datos también se utilizan para analizar y optimizar cadenas de suministro, identificando ineficiencias y posibles ahorros en costos logísticos (Chakrabarti & Indyk, 2009).
  1. Modelos de Clusterización

El clustering, también conocido como análisis de conglomerados, es una tarea general que consiste en agrupar un conjunto de objetos de manera que los objetos con características similares permanezcan en el mismo grupo. Es un aspecto vital de la minería de datos y una técnica comúnmente empleada en varios campos (Hartigan, 1975).

En “Data Mining: Concepts and Techniques” (Pei, Han, & Tong, 2011), se describe una variedad de algoritmos de clustering empleados en el análisis de datos. Estos incluyen métodos basados en centroides, donde se utilizan puntos centrales para representar los clústers; técnicas basadas en la densidad, que agrupan datos según la concentración de puntos en una región específica; algoritmos jerárquicos, que construyen clusters mediante la creación de jerarquías basadas en la proximidad de los datos; y métodos basados en distribución, que emplean modelos estadísticos para identificar la distribución de los datos. Cada uno de estos enfoques tiene sus propias ventajas y se utiliza según el tipo de datos y el objetivo del análisis (Pei, Han, & Tong, 2011) .

* + 1. Algoritmo K-means

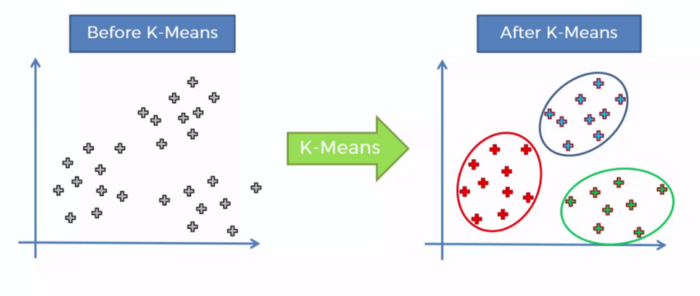
K-means es un algoritmo de clasificación no supervisada (clusterización) que agrupa objetos en k grupos basándose en sus características. El agrupamiento se realiza minimizando la suma de distancias entre cada objeto y el centroide de su grupo o cluster. Se suele usar la distancia cuadrática.

El algoritmo consta de tres pasos:

1. **Inicialización:** una vez escogido el número de grupos, k, se establecen k centroides en el espacio de los datos, por ejemplo, escogiéndolos aleatoriamente.
2. **Asignación objetos a los centroides:** cada objeto de los datos es asignado a su centroide más cercano.
3. **Actualización centroides:** se actualiza la posición del centroide de cada grupo tomando como nuevo centroide la posición del promedio de los objetos pertenecientes a dicho grupo.

Se repiten los pasos 2 y 3 hasta que los centroides no se mueven, o se mueven por debajo de una distancia umbral en cada paso y se formen grupos o cluster como se puede ver en la Figura 6.

**Figura 6**   
Antes y despues de aplicar K-means



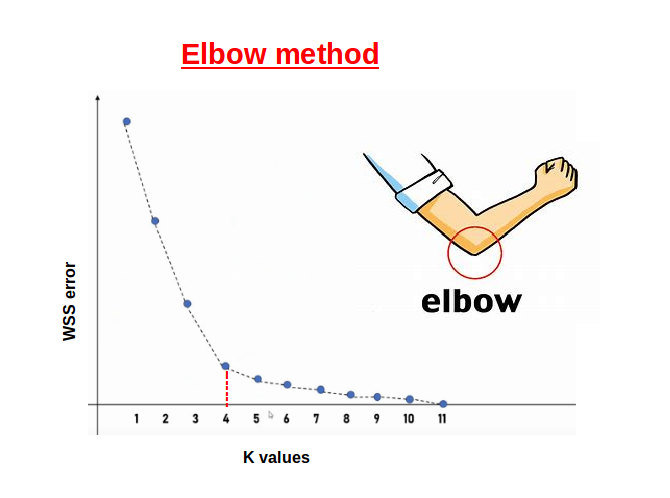
El algoritmo k-means resuelve un problema de optimización, siendo la función a optimizar (minimizar) la suma de las distancias cuadráticas de cada objeto al centroide de su cluster (Unioviedo, 2024)

* + 1. Método del codo

El método del codo es una técnica que se emplea para determinar el número óptimo de clústeres para el algoritmo de K-means. La idea detrás de este método es bastante sencilla. Identificar el número de clústeres para el que se observa un cambio significativo en la tasa de disminución de la varianza intra-cluster (también conocido como suma total de las distancias al cuadrado (Rodriguez, 2023).

Para ello se debe ejecutar el algoritmo de k-means con diferentes números de clústeres (k) y calcular la suma de las distancias al cuadrado de cada punto respecto a su centroide. Usando los resultados para crear una gráfica con los valores de k en el eje x y la suma de las distancias al cuadrado en el eje y. En esta gráfica se busca el punto donde se produce un cambio brusco en la disminución de la suma de las distancias al cuadrado. Punto en el que la curva muestra la forma de un codo (Ver figura 7). El número óptimo de clústeres se corresponde con este punto que se conoce como el “codo” (Rodriguez, 2023)

**Figura 7**   
Método del codo



*Fuente:* (Tovar, 2023)

* + 1. Coeficiente de Silueta (Silhouette Index)

La validación de clústeres es un paso esencial para evaluar los de resultados de la agrupación. Existen tres tipos de procedimientos de validación de clústeres: validación externa de clústeres, validación interna de clústeres y validación relativa de clústeres (Halkidi, Batistakis, & Vazirgiannis, 2001).

El coeficiente de Silueta es una métrica para evaluar la calidad del agrupamiento obtenido con algoritmos de clustering. El objetivo de Silueta es identificar cuál es el número óptimo de agrupamientos.

En los algoritmos de aprendizaje no supervisado, la cantidad de grupos puede ser un parámetro de entrada del algoritmo o puede ser determinado automáticamente por el algoritmo. En el primer caso, como ocurre con el algoritmo de K-Mean, la determinación del número óptimo de clusters tiene que ser realizado mediante alguna medida externa al algoritmo. El coeficiente de silueta es indicador del número ideal de clusters. Un valor más alto de este índice indica un caso más deseable del número de clusters (UNLU, 2015).

El coeficiente de Silueta para una observación i se denota como s(i) y se define como:

Donde:

* **a** es el promedio de las disimilitudes (o distancias) de la observación i con las demás observaciones del cluster al que pertenece i
* **b** es la distancia mínima a otro cluster que no es el mismo en el que está la observación i. Ese cluster es la segunda mejor opción para i y se lo denomina vecindad de i.

El valor de s(i) puede ser obtenido combinando los valores de a y b como se muestra a continuación:

El coeficiente de Silueta es un valor comprendido entre -1 y 1.

* s(i) ≈ 1, la observación i está bien asignada a su cluster
* s(i) ≈ 0, la observación i está entre dos cluster
* s(i) ≈ −1, la observación i está mal asignada a su cluster (UNLU, 2015)
  1. Métodos de reducción de dimensionalidad

Los métodos de reducción de la dimensionalidad se pueden categorizar principalmente en dos grupos: lineales y no lineales. Los métodos lineales, como el análisis de componentes principales (ACP), buscan construir nuevas variables colectivas mediante la realización de combinaciones lineales de las variables de entrada. Por el contrario, los métodos no lineales, como el método de incrustación de vecinos estocásticos distribuidos (t SNE) y el método de aproximación y proyección de variedad uniforme (UMAP), construyen nuevas variables colectivas mediante el mapeo las variables de entrada a una función no lineal (Jolliffe, 2002).

En las estrategias de análisis de datos, es fundamental considerar las orientaciones proporcionadas por (Jolliffe, 2002). En este texto se enfatiza la importancia de seleccionar cuidadosamente el método de reducción de dimensionalidad más adecuado para cada conjunto de datos. Jolliffe explica que la elección debe basarse en las características ´únicas y los objetivos específicos del análisis de datos en cuestión. Por lo tanto, recomienda evaluar distintas técnicas de reducción de dimensionalidad, comparando sus resultados para determinar cuáles más efectivo en revelar las estructuras subyacentes del conjunto de datos o en simplificar la información sin sacrificar aspectos críticos. Este enfoque comparativo es esencial para facilitar un análisis de datos más eficiente y preciso (Jolliffe, 2002).

* + 1. PCA (Análisis de componentes principales)

El análisis de componentes principales (ACP) reduce la dimensionalidad de los datos proyectando cada punto sobre algunos componentes principales. Estos representan una versión de menor dimensionalidad de los datos originales mientras preservan la variación de estos. Los componentes en ACP son combinaciones lineales de las variables de entrada y son ortogonales entre sí. Dadas dos variables, y , su covarianza muestral mide como estas dos variables varían conjuntamente de sus medias y , basado en observaciones.

En ACP, se construye una matriz de covarianza C de tamaño , para un conjunto de datos con p variables. Cada elemento representa la covarianza entre dos variables, xi y, es decir, . Los vectores propios de C son los componentes del ACP. Los valores propios de C indican la importancia de cada componente en el conjunto de datos; a mayor magnitud del valor propio, mayor es la contribución de su correspondiente componente principal. En general, los vectores propios con los valores propios más altos se seleccionan como componentes principales para formar un espacio 2D o 3D para la proyección de los datos (Jolliffe, 2002).

* + 1. Análisis de correlación de variables

El análisis de correlación es una técnica estadística utilizada para medir y evaluar la fuerza y la dirección de la relación entre dos o más variables. Ayuda a identificar si los cambios en una variable están asociados a cambios en otra y cuantifica el grado de esta asociación (Appinio Research, 2024).

Cuando se trabaja con grandes conjuntos de datos que incluyen múltiples variables, es común que algunas de estas variables estén fuertemente correlacionadas, lo que puede llevar a problemas de multicolinealidad. La multicolinealidad ocurre cuando dos o más variables predictoras están altamente correlacionadas, lo que puede generar resultados erróneos o difíciles de interpretar en los modelos analíticos. (Kendall & Stuart, 1961)

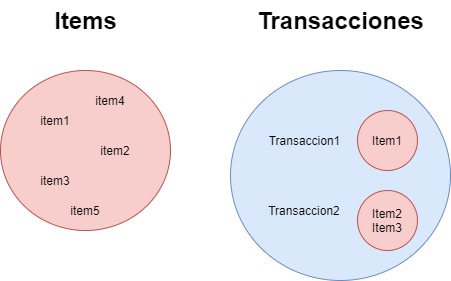
Para evitar y mejorar la calidad de los modelos, es necesario aplicar técnicas de reducción de dimensionalidad. Una forma de hacerlo es a través del análisis de correlación, que nos permite identificar variables que ofrecen información redundante y que, por lo tanto, pueden ser eliminadas o combinadas. Esta reducción de dimensionalidad es crucial en los análisis de clusterización y minería de reglas de asociación, ya que permite trabajar con un conjunto de datos más manejable y eficiente (Kendall & Stuart, 1961).

La matriz de correlación y la reducción de dimensionalidad son herramientas esenciales para asegurar que los datos sean óptimos para el análisis y que el modelo sea capaz de identificar patrones significativos, sin ser influenciado por ruido o redundancia. Este paso prepara los datos para el uso de técnicas avanzadas como clusterización y reglas de asociación, mejorando la calidad de los resultados obtenidos y la toma de decisiones (Kendall & Stuart, 1961).

* 1. Reglas de Asociación

Las reglas de asociación se utilizan para identificar patrones frecuentes entre variables dentro de grandes conjuntos de datos (ver Figura 8). En el contexto de las exportaciones, las reglas de asociación permiten encontrar relaciones entre los productos exportados y los destinos. El algoritmo Apriori es uno de los más utilizados para esta tarea, ya que permite encontrar combinaciones frecuentes de productos y destinos con un soporte y una confianza determinados (Agrawal & Swami, 1993).

**Figura 8**  
 Ítems y transacciones



Las reglas de asociación, a un nivel básico, implica el uso de modelos de Machine Learning para analizar los datos en busca de patrones, o coocurrencias, en una base de datos. Identifica las asociaciones frecuentes, que a su vez son las reglas de asociación.

Una regla de asociación tiene dos partes:

un antecedente (si) y

un consecuente (entonces)

Un antecedente es un elemento que se encuentra dentro de los datos. Un consecuente es un elemento que se encuentra en combinación con el antecedente (Gonzáles, 2021).

Las reglas de asociación se crean buscando en los datos patrones frecuentes de “if then” y utilizando los criterios de apoyo y confianza para identificar las relaciones más importantes. El apoyo es una indicación de la frecuencia con que los elementos aparecen en los datos. La confianza indica el número de veces que las afirmaciones del tipo “if-then” se consideran verdaderas. Se puede utilizar una tercera métrica, llamada fit para comparar la confianza con la confianza esperada, o cuántas veces se espera que una afirmación del tipo “if-then” se considera cierta (Gonzáles, 2021).

Las reglas de asociación se calculan a partir de conjuntos de elementos, que se componen de dos o más elementos. Si las reglas se construyen a partir del análisis de todos los conjuntos de elementos posibles, podría haber tantas reglas que éstas tendrían poco significado. Con eso, las reglas de asociación se crean típicamente a partir de reglas bien representadas en los datos (Gonzáles, 2021).

* + 1. Algoritmo Apriori

El algoritmo Apriori se basa en la idea de que, si una combinación de elementos ocurre con frecuencia en el dataset, es probable que haya una asociación significativa entre ellos. Al aplicar Apriori a los datos de exportación, podemos descubrir patrones de productos que son frecuentemente exportados a determinados destinos. Este tipo de análisis es crucial para identificar mercados con alto potencial y optimizar las estrategias de exportación (Pei, Han, & Tong, 2011).

Los conjuntos de elementos candidatos se generan usando solo los grandes conjuntos de elementos de la pasada anterior. El conjunto de ítems grandes de la pasada anterior se une a sí mismo para generar todos los conjuntos de ítems con un tamaño que es mayor por uno. Cada conjunto de ítems generado con un subconjunto que no es grande es entonces eliminado. El resto de los conjuntos de elementos son candidatos (Gonzáles, 2021).

El algoritmo de Apriori aprovecha el hecho de que cualquier subconjunto de un conjunto de ítems frecuente es también un conjunto de ítems frecuente. Por lo tanto, el algoritmo puede reducir el número de candidatos que se están considerando, explorando solo los conjuntos de elementos cuyo número de apoyos es mayor que el número mínimo de apoyos. Todos los conjuntos de elementos infrecuentes pueden ser podados si tienen un subconjunto infrecuente con algunos puntos a considerar:

* Tiene gran importancia en la minería de datos.
* Las reglas resultantes son intuitivas y fáciles de comunicar a un usuario final.
* Es fácil de implementar.
* No requiere de datos etiquetados ya que está totalmente desprovisto de supervisión.
* Puede ser muy lento.
* Si el conjunto de datos es pequeño puede encontrar muchas asociaciones falsas que ocurren por casualidad.
* Requiere muchos escaneos de la base de datos (Gonzáles, 2021).
  + 1. Validación de las reglas de asociación

La validación de las reglas de asociación es importante en el proceso de minería de datos, ya que ayuda a determinar si las reglas generadas realmente representan patrones significativos y útiles para la toma de decisiones. La validación garantiza que las reglas encontradas no sean simplemente el resultado de coincidencias aleatorias, sino que tengan un valor real en el contexto del análisis (Agrawal & Srikant , 1995)

Existen varias formas de validar las reglas de asociación, y cada una de ellas permite asegurar que las reglas sean relevantes, precisas y útiles para el análisis. A continuación, se explican las métricas de validación más comunes utilizadas en minería de reglas de asociación:

* **Soporte (Support)**

**Definición:** El soporte mide la frecuencia con la que ocurre una regla en el conjunto de datos. Es decir, qué tan frecuente es una combinación de productos (o ítems) en las transacciones.

**Cálculo:** El soporte se calcula como el número de transacciones que contienen tanto el antecedente como el consecuente de la regla, dividido por el número total de transacciones.

**Interpretación:** Un soporte alto indica que la regla es común y frecuente en los datos, lo que significa que puede ser relevante. Sin embargo, un soporte demasiado bajo puede indicar que la regla no es lo suficientemente significativa.

* **Confianza (Confidence)**

**Definición:** La confianza mide la probabilidad de que el consecuente de una regla ocurra, dado que el antecedente ha ocurrido. En otras palabras, ¿cuán probable es que la acción B (consecuente) se realice si ya ha ocurrido la acción A (antecedente)?

**Cálculo:** La confianza se calcula como el número de transacciones que contienen tanto A como B, dividido por el número de transacciones que contienen solo A.

**Interpretación:** Un valor de confianza alto (generalmente superior al 70%) indica que la regla es fuerte y tiene un alto grado de fiabilidad. Sin embargo, una alta confianza no garantiza que la regla sea útil si el soporte es bajo.

* **Levantamiento (Lift)**

**Definición:** El levantamiento mide cuán grande es la probabilidad de que los elementos de la regla ocurran juntos, en comparación con si los elementos fueran independientes. El lift ayuda a identificar reglas que no solo son frecuentes, sino que también tienen una relación más fuerte de lo que sería esperado por azar.

**Cálculo:** El levantamiento se calcula dividiendo la confianza de la regla entre el soporte de los elementos A y B.

**Interpretación:** Si el lift es mayor que 1, significa que A y B ocurren más juntos de lo que se esperaría si fueran independientes, lo que indica una fuerte asociación. Si el lift es igual a 1, no hay una relación significativa. Si el lift es menor que 1, indica que A y B tienden a ocurrir menos juntos de lo esperado.

* **Convicción (Conviction)**

**Definición:** La convicción mide cuánta más probabilidad tiene la regla de ser válida en comparación con una ocurrencia aleatoria. Esta métrica toma en cuenta la probabilidad de que A ocurra sin B y la probabilidad de que B ocurra sin A.

**Cálculo:** La convicción se calcula como:

**Interpretación:** Un valor de convicción alto indica que la regla es más confiable y que el antecedente A es realmente útil para predecir el consecuente B, por lo que la regla es sólida (Agrawal & Srikant , 1995)

CAPÍTULO III

MARCO APLICATIVO

CAPÍTULO 3 MARCO APLICATIVO

El desarrollo del marco aplicativo en esta monografía describe el procedimiento y las herramientas metodológicas para implementar modelos de minería de datos, clusterización y reglas de asociación, en el análisis y aplicación de exportaciones no tradicionales para identificar patrones y asociaciones que permitan optimizar las exportaciones y contribuir a la generación de divisas en Bolivia.

Se utilizó un enfoque cuantitativo basado en minería de datos para identificar patrones y asociaciones en las exportaciones no tradicionales de Bolivia, la metodología para el análisis se basa en un enfoque estructurado de minería de datos.

1. 1. Configuración del entorno de trabajo

El entorno de ejecución se llevó a cabo en una laptop con las siguientes características:

Laptop Dell G7 7588 con procesador Intel(R) Core (TM) i7-8750H CPU @ 2.20GHz 2.21 GHz, 32 GB SDRAM, NVIDIA GeForce GTX 1060 MaxQ, 4GB GDDR5 VRAM, Windows 11 (64 bits).

El análisis se realizó en Google Colab para facilitar la ejecución remota y aprovechar los recursos de procesamiento en la nube.

Se utilizaron las siguientes librerías clave, como Pandas, Scikit-learn, Matplotlib y MLxtend, todas las cuales proporcionan herramientas poderosas para el preprocesamiento de datos, y los datos cargados se leyeron desde el archivo de Excel en el office 365.

* 1. Descripción del conjunto de datos

El dataset analizado contiene 9851 registros y 38 atributos proveniente Instituto Nacional de Estadística (INE) que incluyen datos sobre productos exportados de enero a agosto 2024, países de destino, medios de transporte, peso neto y bruto, y valores económicos. La Tabla 1 resume los atributos clave.

**Tabla 1**   
Breve Descripción del conjunto de datos

|  |  |
| --- | --- |
| **ATRIBUTO** | **VALOR** |
| GESTION y MES | Año y mes de importación. |
| PAIS y DESPAIS | Código y descripción del país. |
| KILBRU y KILNET | Peso bruto y peso neto. |
| FINO Y VALOR | Valores asociados a los derechos aduaneros en bs y dólares |
| DESNAN, NANDINA, CAP, DESCAP, SECC, DESSEC | Código y descripción de los países miembros de la comunidad andina. |
| AREA Y DESAREA | Código y descripción de la zona geoeconómica. |
| OTROS | Mercosur, comunidad andina, etc. |
| MEDI, DESMEDI, VIASAL, DESVIA | Código y descripción medios de transporte. |
| DEPART, DESDEP | Código y descripción código departamento origen de las exportaciones |
| CLACT, CODACT, DESACT | Código y descripción del producto de la actividad económica |
| TNT, DESTNT | Código y descripción código del producto tnt (productos tradicionales y no tradicionales) |

* 1. Preprocesamiento y selección de los datos

Se identificaron los valores faltantes y fueron rellenados por el promedio, se manejaron atributos categóricos y numéricos. Las visualizaciones permitieron observar relaciones clave entre los atributos como el valor de las exportaciones, país de origen y tipo de productos. El preprocesamiento se realizó en Google Colab.

* + 1. Verificación y asignación del tipo de datos

Las columnas ['CUCI3', 'GCE3', 'CODACT2', 'CIIUR3', 'CODACT2'] fueron convertidas a formatos numéricos, como se puede ver en la Figura 9, y se procedió a corregirlos.

**Figura 9**   
Visualización de tipos de datos del dataset

Un conjunto de letras blancas en un fondo blanco

Descripción generada automáticamente con confianza media

* + 1. Valores faltantes

Los valores ausentes en columnas categóricas se reemplazaron con "NA", mientras que los atributos numéricos fueron imputados con el promedio, ver la siguiente Figura 10.

**Figura 10**   
Visualización de valores nulos

**Texto

Descripción generada automáticamente**

* + 1. Campos corruptos

Se detectaron filas con datos corruptos, como se puede ver en la figura 11.

**Figura 11**   
Visualización de fila [9451] corrupta

Imagen que contiene Texto

Descripción generada automáticamente

Se procedió a limpiar con la función str.replace() buscando todos los posibles valores que no sean ASCII haciendo su respectiva conversión y eliminación, como se puede observar en la Figura 12.

**Figura 12**   
Visualización de fila [9451] corregida

Texto

Descripción generada automáticamente

* + 1. Outliers (valores sesgados de los datos)

Se procedió a hacer un análisis exploratorio basado en la distribución de los datos, para observar y determinar que columnas del dataset necesitaban una transformación logarítmica, utilizando el método skew() de scipy, para valores grandes a la derecha como se puede ver en la Figura 13.

**Figura 13**   
Histograma de distribuciones sesgadas

Diagrama

Descripción generada automáticamente con confianza media

Se evidencia que las columnas ['KILBRU', 'KILNET', 'VALOR', 'FINO'] necesitan transformación logarítmica y se hará por el método de Z-score para afectar solo a los datos sesgados.

* + 1. Selección de columnas relevantes

Se tomaron en cuenta estos factores clave:

* **Alto valor total de exportaciones:** Representa una contribución significativa a la generación de divisas.
* **Mayor diversidad de categorías únicas:** Indica un rango amplio de productos exportados.
* **Mayor diversidad de países únicos:** Refleja mercados diversificados con potencial de crecimiento.

Se aplicó la matriz de correlación de Pearson para las variables numéricas con el objetivo de identificar las variables redundantes, cómo se puede observar en la Figura 14

**Figura 14**   
Matriz de Correlación

Gráfico, Gráfico de rectángulos

Descripción generada automáticamente

Se identificaron las columnas con alta correlación con ['VALOR']

* **VALOR:** 1.000000
* **KILBRU:** 0.813823
* **KILNET:** 0.795076
* **FINO:** 0.453490
* **MEDI:** 0.408369

Para las variables categóricas se utilizo el PCA y se identificaron los componentes principales mediante la varianza acumulada como se puede ver en la Figura 15.

**Figura 15**   
Varianza acumulada PCA

Gráfico

Descripción generada automáticamente

Y se seleccionaron 6 componentes principales:

['ADUDES', 'GESTION', 'MES', 'NANDINA', 'CAP', 'SECC'], dtype='object').

* + 1. Transformación y escalado

Se aplicó escalado a las columnas numéricas ['VALOR', 'KILNET'] usando StandardScaler (), y se codificaron las variables categóricas con LabelEncoder en las columnas ['DESPAI', ' DESTNT '] para el análisis posterior.

* 1. Minería de datos
     1. Método del codo

Generalmente, el "codo" en la curva indica el número óptimo de clusters, donde la reducción de la inercia comienza a disminuir significativamente, como se puede ver en la figura 16, en K=3 muestra un cambio notable en la pendiente donde la inercia comienza a reducir, y en cambio sí se agregarán más cluster no reduce significativamente la inercia.

**Figura 16**   
Método del codo de la data final

Gráfico, Gráfico de líneas

Descripción generada automáticamente

* + 1. Algoritmo: K-Means

Se aplico el algoritmo de Clusterización K-Means, para identificar patrones en los datos con k=3 ha asignado un cluster a cada producto en el dataset para analizar los clusters y entender las características distintivas de cada grupo como se puede ver en la Figura 17

**Figura 17**   
Clúster Valor de exportación VS Peso bruto de productos

Gráfico, Gráfico de dispersión

Descripción generada automáticamente

* + 1. Análisis de clústeres

Los resultados del análisis de clústeres indican que los productos no tradicionales se agrupan en tres categorías distintas. La siguiente es una descripción de cada clúster identificado:

**Clúster 0 (morado):**

**Características:** Los puntos del cluster morado se agrupan en la parte inferior izquierda del gráfico, donde tanto KILBRU como VALOR tienen valores relativamente bajos.

**Interpretación:** Este cluster probablemente representa productos con bajos volúmenes de exportación (indicado por valores bajos en KILBRU) y bajo valor de exportación (indicado por valores bajos en VALOR).

**Posibles aplicaciones:** Este grupo podría estar relacionado con productos menos demandados o de bajo valor en el mercado. Podría ser útil para identificar productos que requieren estrategias de promoción o diversificación para aumentar su competitividad.

**Clúster 1(amarillo):**

**Características:** El cluster amarillo se distribuye a lo largo del eje KILBRU, con valores más altos que el cluster morado, pero bajo en comparación con el cluster amarillo. El VALOR sigue una tendencia ascendente.

**Interpretación:** Este grupo podría representar productos que tienen un volumen moderado de exportación y un valor intermedio en comparación con otros productos. Esto puede incluir productos con una buena presencia en mercados internacionales, pero con espacio para mejorar.

**Posibles aplicaciones:** Los productos en este cluster podrían beneficiarse de estrategias de optimización de precios o expansión en mercados emergentes. También podría ser un buen objetivo para mejorar la calidad y valor agregado.

**Clúster 2 (verde):**

**Características:** El cluster verde se agrupa en la parte superior derecha del gráfico, con valores muy altos tanto en KILBRU como en VALOR. Esto sugiere que estos productos tienen un alto volumen de exportación y alto valor.

**Interpretación:** Este cluster representa productos que probablemente son los más exitosos en términos de volumen y valor de exportación. Es posible que estos productos sean líderes en el mercado y altamente demandados.

**Posibles aplicaciones:** Este grupo debería recibir estrategias de consolidación y expansión, ya que está generando un buen rendimiento. Se podría considerar fortalecer la relación con los principales mercados y buscar oportunidades para expandir en nuevos mercados.

* 1. Reglas de asociación

Se selecciono el Cluster 1 porque presenta una distribución más concentrada y homogénea como se puede ver en el Anexo I en la distribución por flujos, con valores que se agrupan alrededor de la mediana en las variables de VALOR y KILBRU como se puede ver en los Anexos II y III, y el cluster 1 refleja la mayor cantidad de productos no tradicionales, como se puede ver en la Figura 18.

**Figura 18**   
Distribución de VALOR por cada clúster

Gráfico

Descripción generada automáticamente

* **Diversidad de categorías**: La variedad de productos permite identificar combinaciones interesantes entre ellos, esto genera un mayor potencial para diversificar las exportaciones y encontrar patrones ocultos para generar reglas de asociación.
* **Diversidad de países:** El rango amplio de destinos brinda la oportunidad de entender las preferencias del mercado y crear reglas que fortalezcan estrategias comerciales.
* **Menor valor promedio:** Los valores individuales de exportación son menores, el volumen y la variedad compensan este aspecto, permitiendo explorar patrones únicos.
  + 1. Algoritmo Apriori

Se aplico el A priori para encontrar patrones frecuentes, con la siguiente línea de codigo: *frequent\_itemsets=fpgrowth(transaction\_df,min\_support=0.25,use\_colnames=True)*

Se aplicó escalado a las columnas numéricas ['VALOR', 'KILNET'] usando StandardScaler(), y se codificaron las variables categóricas con LabelEncoder en las columnas ['DESPAI', ' DESACT2 '] para el análisis posterior.

Se generaron reglas de asociación utilizando conjuntos frecuentes de ítems y se agruparon por destino de exportación y producto como se puede ver en el Anexo IV como se ve en la siguiente línea de código: rules = association\_rules(frequent\_itemsets, num\_itemsets=len(frequent\_itemsets), metric="lift", min\_threshold=1.0). Se exportaron las 61 reglas de asociación como se ve en la Figura 19

**Figura 19**   
Reglas de Asociación

Texto

Descripción generada automáticamente

Se realizo el ajuste del valor de min\_support según los datos, y el número total de transacciones y cómo el soporte afecta la selección de conjuntos frecuentes.

* Min Support = 0.05: Un conjunto debe aparecer al menos en 6 transacciones.
* Min Support = 0.2: Un conjunto debe aparecer al menos en 22 transacciones.

Se uso min\_support = 0.2, ya que capturará asociaciones significativas con buena frecuencia y se evitó la sobrecarga con análisis con asociaciones irrelevantes.

* + 1. Filtrado de las reglas de asociación

Se realizo el filtrado de las reglas para responder al objetivo general con los siguientes criterios:

* **Alta Confianza:** Seleccionar reglas con valores de máxima confianza elevados **(> 0.98)**, ya que indican asociaciones más consistentes.
* **Elevado Lift:** Reglas con lift **(> 1.45)** muestran relaciones muy fuertes entre antecedente y consecuente.
* **Frecuencia Relevante (Soporte):** Reglas con soporte significativo **(> 0.15)** para asegurar que las asociaciones no sean raras.
* **Alta Certeza:** Asegura asociaciones consistentes y útiles **(> 0.98)** para decisiones estratégicas.
* **Productos Estratégicos:** Filtrar reglas donde el antecedente o consecuente incluya productos no tradicionales de exportación presentes en las reglas de asociación:

1. Quinua.
2. Castaña.
3. Chía.
4. Productos Alimenticios.
5. Productos Textiles.
6. Prendas de Vestir, Adobo y Teñido de Pieles.
7. Maderas y Manufacturas de Madera.
8. Otros Metales Manufacturados.
9. Otras Manufacturas.

En base a esos criterios se generaron 8 reglas fuertes como se ve en la Figura 20

**Figura 20**   
Reglas de Asociación Filtradas

Interfaz de usuario gráfica, Texto

Descripción generada automáticamente

De las reglas filtradas, se seleccionaron 5 altamente significativas que destacan relaciones entre productos no tradicionales (quinua, castaña, textiles) y manufacturas. Estas asociaciones sugieren oportunidades para diversificar mercados y fortalecer exportaciones estratégicas.

***Regla 1 Si:*** *Quinua, Productos alimenticios* ***Entonces:*** *Otras manufacturas*

**Soporte:** 20.91%, lo que indica que casi una quinta parte de las transacciones contienen ambos elementos.

**Confianza:** 100%, mostrando una relación determinística.

**Lift:** 1.47, lo que sugiere que esta asociación es 47% más probable que por azar.

**Implicación:** Las exportaciones de quinua están integradas con otras manufacturas, posiblemente debido al empaquetado o al procesamiento.

***Regla 2 Si:*** *Castaña, Productos alimenticios* ***Entonces:*** *Otras manufacturas*

**Soporte:** 20%, similar al caso de la quinua.

**Confianza:** 100%, garantizando una asociación completa.

**Lift:** 1.47, igual que la Regla 1.

**Implicación:** Las castañas, al igual que la quinua, tienen una alta dependencia de las manufacturas para completar el proceso de exportación.

***Regla 3 Si:*** *Productos textiles* ***Entonces:*** *Otras manufacturas*

**Soporte:** 27.27%, lo que muestra una frecuencia ligeramente mayor que las reglas anteriores.

**Confianza:** 100%.

**Lift:** 1.47.

**Implicación:** Los productos textiles dependen de las manufacturas en la cadena de exportación, quizás debido al embalaje o complementos.

***Regla 4 Si:*** *Maderas y manufacturas de madera* ***Entonces:*** *Otras manufacturas*

**Soporte:** 20%.

**Confianza:** 100%.

**Lift:** 1.47.

**Implicación:** La madera y sus manufacturas se exportan consistentemente con otras manufacturas, destacando su rol como productos semi-procesados o finales.

***Regla 5 Si:*** *Otros metales manufacturados* ***Entonces:*** *Otras manufacturas*

**Soporte:** 26.36%.

**Confianza:** 100%.

**Lift:** 1.47.

**Implicación:** Este patrón refuerza que los metales manufacturados se exportan junto con otras manufacturas, posiblemente como parte de cadenas de valor agregado.

* 1. Evaluación de los modelos
     1. Coeficiente de silueta

Los resultados del análisis de los clusters con el algoritmo K-means son los siguientes:

* **Inercia (Suma de distancias dentro del cluster):** 6675.38
* **Silhouette Score:** 0.44

**Interpretación:**

**Inercia:** El valor de 6675.38 muestra que los clusters formados en este modelo K-Means tienen un cierto grado de dispersión. No es un valor bajo, lo que indica que los puntos dentro de los clusters no están perfectamente agrupados, pero tampoco es un valor muy alto, lo que podría implicar que los clusters no están bien definidos.

**Silhouette Score:** Un Silhouette Score de 0.44 es relativamente moderado, lo que indica que los clusters son razonablemente buenos, pero hay margen para mejorar. Este valor sugiere que los puntos dentro de cada cluster están bastante cerca de los centroides, pero que los clusters podrían estar un poco más separados. En este caso, el modelo no está proporcionando un agrupamiento extremadamente claro.

CAPÍTULO IV

CONCLUSIONES Y RECOMENDACIONES

CAPÍTULO 4 CONCLUSIONES Y RECOMENDACIONES

1. 1. Conclusiones

La elaboración de la monografía para un análisis de las exportaciones bolivianas mediante el uso de modelos de minería de datos, como clusterización y reglas de asociación, ha cumplido con los objetivos que se plantearon, y proporcionó información sobre cómo optimizar las exportaciones de productos no tradicionales y, con ello, reducir la escasez de divisas.

La base de datos que se ha utilizado ha sido limpia y preprocesada Todos los valores vacíos han sido limpiados, corregidos y los tipos relevantes de tipo de datos han sido transformados para que la información estuviera adecuada para el análisis.

Se aplicó el algoritmo de K-Means para agrupar los productos exportados en tres clusters. A partir de los resultados fue posible discriminar productos por volumen y valor de agrupación. Esto proporciona una idea de que productos que están llenos de potencial y los que requieren atención para elevar su competitividad en los mercados globales.

El modelo K-Means presentó un Silhouette Score de 0.44, lo que indica que los clusters son razonablemente buenos, pero se puede mejorar la separación entre ellos. Este resultado sugiere que, aunque proporcionan información útil, se podría analizar la configuración del modelo y considerar la inclusión de más variables o la elección de otro número de K.

En cuanto a las Reglas de asociación, se utilizó para identificar asociaciones entre productos, como la relación entre quinua, castaña, productos textiles y otras manufacturas. Esta técnica identifico de productos con mayor potencial para la diversificación de mercados internacionales, lo que implica la formación de estrategias basadas en la combinación de productos que generen valor para las exportaciones.

En base a los resultados obtenidos una acción recomendada es promover productos con alto volumen y valor de exportación como quinua y castaña para aumentar el procesamiento de la producción en el país y agrega cadenas de valor a los productos no tradicionales para su exportación.

* 1. Recomendaciones

Se sugieren las siguientes recomendaciones:

**Ajuste del Número de Clusters:** El Silhouette Score indica que los clusters son bastante buenos, pero no óptimos. Se aconseja probar con diferentes valores K (por ejemplo, 𝐾=5) para lograr una mejor segmentación y encontrar más patrones de exportación de productos.

**Mejora del Preprocesamiento de Datos:** Aunque se hizo un buen preprocesamiento de los datos, se sugiere más ajustes en el manejo de outliers, transformación de variables sesgadas y el manejo de valores faltantes. Esto garantizará que el modelo se base en datos más representativos y que genere resultados más exactos.

**Diversificación de Productos y Mercados:** Es importante diversificar los productos que se exportan para no depender de los productos tradicionales. Incrementar la cantidad de productos no tradicionales y explorar nuevos mercados internacionales.

**Estrategias de Valor Añadido:** Productos como la quinua, la castaña y textiles están relacionados con otras manufacturas. Se sugiere fomentar la industria de valor añadido, como el procesamiento o el empaquetado de productos en el país, lo que ayudará a aumentar el valor de las exportaciones de Bolivia.

**Desarrollo de Políticas de Diversificación:** El análisis sugiere que, para mitigar la escasez de divisas, es crucial diversificar las exportaciones y expandir los mercados para los productos no tradicionales. Las políticas comerciales deben centrarse en incentivar el procesamiento interno de productos para agregar valor y aumentar la competitividad en el mercado internacional.

La aplicación de técnicas de minería de datos como la clusterización y las reglas de asociación ofrece una forma efectiva de optimizar las exportaciones bolivianas, permitiendo la identificación de patrones clave y asociaciones entre productos que contribuirán a la generación de divisas y a la diversificación de la economía.

# Bibliografía

Agrawal, R. I., & Swami, A. (1993). *Mining association rules between sets of items in large databases.* ACM SIGMOD .

Agrawal, R., & Srikant , R. (1995). Mining Sequential Patterns. *In Proceedings of the 1995 International Conference on Data Engineering (ICDE).*

Appinio Research. (25 de 06 de 2024). *Appinio Research.* Obtenido de Análisis de correlación: ¿Qué es? Definición, procedimiento, ejemplos: https://www.appinio.com/es/blog/investigacion-de-mercados/analisis-correlacion

Chakrabarti, S. D., & Indyk, P. (2009). *Enhanced mining of economic data using advanced clustering techniques* (Vol. Journal of Economic Analysis).

Chen, M. S., Han, J., & Yu, P. S. (1996). Data Mining: An Overview from a Database Perspective. *EEE Transactions on Knowledge and Data Engineering*, 866 - 883.

El deber. (23 de 08 de 2024). *EL DEBER*. Obtenido de La escasez de dólares y carburante se agudiza en Bolivia: https://eldeber.com.bo/pais/la-escasez-de-dolares-y-carburante-se-agudiza-en-bolivia\_382553

FAO. (1999). *Organización de las Naciones Unidas para la Agricultura y la Alimentación.* Roma.

Gonzáles, L. (13 de 04 de 2021). *Aprende IA.* Obtenido de Reglas de Asociación: https://aprendeia.com/reglas-de-asociacion/#google\_vignette

Guillermo, O. H. (2020). *Repositorio UMSA.* Obtenido de UNIVERSIDAD MAYOR DE SAN ANDRES.

Halkidi, M., Batistakis, Y., & Vazirgiannis, M. (2001). On clustering validation techniques. *Journal of intelligent information systems, 17*, 107-145.

Hartigan, J. (1975). *Clustering algorithms wiley.* New York.

Ibañez Choque, A. F., Gavincha Lima, M. I., & Llapaco Ávila, M. P. (10 de 11 de 2024). *CRECIMIENTO ECONÓMICO, CAMBIO ESTRUCTURAL Y DIVERSIFICACIÓN: EL CASO DE BOLIVIA.* Obtenido de https://www.bcb.gob.bo/webdocs/publicacionesbcb/revista\_analisis/ra\_vol24/articulo\_2\_v24.pdf

IBCE. (11 de 11 de 2024). *Instituto Boliviano de Comercio Exterior.* Obtenido de https://ibce.org.bo/images/publicaciones/ie-1-Informe-Especial-Cifras-del-Comercio-Exterior-Boliviano-1er-Semestre-2022.pdf

IBM. (11 de 11 de 2024). *IBM*. Obtenido de ¿Qué es el aprendizaje supervisado?: https://www.ibm.com/es-es/topics/supervised-learning

INE. (31 de 10 de 2024). *Instituto Nacional de estadistica*. Obtenido de https://www.ine.gob.bo/index.php/la-economia-de-bolivia-muestra-una-leve-recuperacion-con-un-crecimiento-del-258-al-segundo-trimestre-de-2024/

INE. (19 de 11 de 2024). *Instituto Nacional de Estadistica*. Obtenido de https://www.ine.gob.bo/index.php/estadisticas-economicas/comercio-exterior/exportacion-2/#

infobae. (8 de 8 de 2024). *infobae*. Obtenido de https://www.infobae.com/

Infobae. (19 de 11 de 2024). *Infobae*. Obtenido de Crisis en Bolivia: Luis Arce enfrenta un déficit comercial histórico de USD 329 millones: https://infobae.com

Ips Noticias. (15 de 11 de 2024). *IPS NOTICIAS*. Obtenido de Crisis alimentaria asoma en Bolivia y amenaza a grandes y pequeños productores: https://www.pressenza.com/es/2024/11/crisis-alimentaria-asoma-en-bolivia-y-amenaza-a-grandes-y-pequenos-productores

Jolliffe, I. T. (2002). *Principal component analysis for special types of data.* Springer.

Kendall, M. G., & Stuart, A. (1961). *The Advanced Theory of Statistics.* (2nd Edition ed., Vol. 1).

Kouzmine, V. (07 de 2000). *CEPAL.* Recuperado el 20 de 11 de 2024, de https://www.cepal.org/sites/default/files/publication/files/4413/S2000930\_es.pdf

Los Tiempos. (11 de 11 de 2024). *Los Tiempos*. Obtenido de Cuatro factores afectan la competitividad de las exportaciones bolivianas: https://www.lostiempos.com/actualidad/economia/20241111/cuatro-factores-afectan-competitividad-exportaciones-bolivianas

Loza, G. (1 de 9 de 2024). *La Razon*. Obtenido de https://www.la-razon.com/financiero/2024/09/01/por-que-hay-insuficiencia-de-divisas/

Martínez Ballesteros, M. (11 de 11 de 2024). *Universidad de Sevilla.* Obtenido de Departamento de Lenguajes y Sistemas Informáticos: https://www.aepia.org/wp-content/uploads/2023/06/Martinez\_EVIA2023.pdf

Pei, J., Han, J., & Tong, H. (2011). *Data Mining: Concepts and Techniques.* Morgan Kaufmann.

Ramos Serrano, C. (2021). *Minería de datos aplicada al análisis predictivo. Académica Española.*

Rodriguez, D. (9 de 07 de 2023). *ANALYTICS LANE*. Obtenido de https://www.analyticslane.com/2023/06/09/metodo-del-codo-elbow-method-para-seleccionar-el-numero-optimo-de-clusteres-en-k-means/

Smith, J. (2022). *Optimizing trade with data science techniques.*

Tan , P. N., Steinbach, M., & Kumar, V. (2019). *Introduction to Data Mining.* Pearson.

Tovar, J. C. (28 de 02 de 2023). *Huawei*. Obtenido de Método Elbow y método de siluetas para la determinación del número de clústeres en K-means: https://forum.huawei.com/enterprise/es/m%25C3%25A9todo-elbow-y-m%25C3%25A9todo-de-siluetas-para-la-determinaci%25C3%25B3n-del-n%25C3%25BAmero-de-cl%25C3%25BAsteres-en-k-means/thread/667237354135502848-667212895009779712

Unioviedo. (20 de 11 de 2024). *Escuela de ingeniería Informática*. Obtenido de Universidad de Oviedo: https://www.unioviedo.es/compnum/laboratorios\_py/kmeans/kmeans.html

UNLU. (10 de 2015). *Univerdidad Nacional de Lujan.* Obtenido de Departamento de ciencias básicas: https://www.labredes.unlu.edu.ar/sites/www.labredes.unlu.edu.ar/files/site/data/bdm/coeficiente-silueta.pdf

Witten I, H. F., & Hall, M. A. (2016). *Data Mining: Practical Machine Learning Tools and Techniques.* Morgan Kaufmann.

World Bank. (2020). *Economic Diversification in Emerging Markets: Strategies and Policies.*

ANEXOS

ANEXO I

Gráfico, Gráfico de cajas y bigotes

Descripción generada automáticamente

ANEXO II

Gráfico, Gráfico de cajas y bigotes

Descripción generada automáticamente

ANEXO III

Gráfico, Gráfico de rectángulos

Descripción generada automáticamente

ANEXO IV

Tabla

Descripción generada automáticamente