

Índice

1. Introducción	8
2. Problematización y Objetivos	10
2.1. Planteamiento del problema	10
2.2. Pregunta de investigación	10
2.3. Justificación de la investigación	11
2.4. Delimitación de la investigación	11
2.5. Alcance de la investigación	12
2.6. Objetivos de la investigación	12
2.6.1. Objetivo general	12
2.6.2. Objetivos específicos	12
3. Marco Teórico	12
3.1. Historia del combustible	12
3.1.1. Uso de combustibles en el mundo	14
3.2. Historia de la industria manufacturera en Chile	15
3.3. Estado del Medio Ambiente	20
3.3.1. Contaminación del aire	22
3.4. Clúster como concepto general	27
3.4.1. Aplicación de la metodología de clustering en el País Vasco	29
3.5. Minería de datos	31
3.6. Clustering en minería de datos	34
3.6.1. Algoritmos de clústering	35
3.6.2. Parámetros de clusters	37
3.6.3. Normalización de datos	41
3.7. Metodología CRISP-DM	42
3.8. R como lenguaje de programación	44
3.8.1. RStudio como Software	45
4. Marco Metodológico	47
4.1. Elaboración de hipótesis	47
4.2. Diseño de la investigación	47
4.3. Muestreo	47
4.4. Recolección de datos	48
4.5. Descripción de la base de datos	49
5. Resultados	53
5.1. Estadístico de Hopkins	53
5.2. Número óptimo de clusters	53
5.3. Algoritmo de Clustering	53
5.4. Método PAM	54
5.5. Método Jerárquico	55
5.6. Método K-Means	56

6. Discusión y conclusión	59
6.1. Discusión	59
6.2. Conclusión	61
7. Bibliografía	64
Anexos	67
A. Anexo I: Tabla de tamaños según el número de trabajadores	67
B. Anexo II: Tabla de centroides	67

Índice de figuras

1.	Producto Interno Bruto de la industria manufacturera en Chile en %. 2019	20
2.	Metodología del Estado del Medio Ambiente	21
3.	Resumen de la Contaminación del Aire	23
4.	Concentración de la emisión de gases dañinos para el ser humano y el medio ambiente así como su principal origen	24
5.	Emisiones al aire de SO ₂ por región y según el tipo de fuente del año 2018	25
6.	Emisiones al aire de MP ₂₅ por región y tipo de fuente del año 2018.	26
7.	Emisión del gas NO año 2018.	27
8.	Clúster del cuero Italiano para el mercado de la moda.	29
9.	Proceso KDD.	32
10.	Proceso SEMMA.	33
11.	Ejemplo del método KMeans.	35
12.	Ejemplo del método PAM.	36
13.	Ejemplo del método ward.d2.	37
14.	Ejemplo de la técnica PCA.	38
15.	Ejemplo del método Silhouette.	39
16.	Ejemplo de Distancia Euclídea.	41
17.	Ejemplo de aplicaciones de clústeres en minería de datos.	42
18.	Mapa conceptual CRISP-DM.	44
19.	Visión esquemática del funcionamiento de R.	45
20.	Interface de RStudio.	46
21.	Número óptimo de clústeres según promedio silhouette.	53
22.	Determinación de algoritmo de clustering a utilizar.	54
23.	Clusters PAM.	55
24.	Medoides de los clústeres PAM.	55
25.	Clusters representados en un dendograma.	56
26.	Cluster K-means.	57
27.	Centroides de los clústeres k-means.	57
28.	Gráfico de centroides k-means.	58

Índice de tablas

1.	Participación de los distintos tipos de establecimientos en el número total de establecimientos, el empleo y el valor agregado en 1979	16
2.	Participación en el Empleo de los Distintos Tipos de Establecimientos en los Diversos Sectores Productivos en (%).	17
3.	Participación en el Empleo de los Distintos Tipos de Establecimientos en los Diversos Sectores Productivos en (%).	18
4.	Valor Bruto, Consumo Intermedio y Valor Agregado Según Tramo por Número de Trabajadores.	19
5.	Participación en el Empleo de los Distintos Tipos de Establecimientos en los Diversos Sectores Productivos (%).	30
6.	Cantidad de establecimientos según columna CIU-4.	50
7.	Número de Establecimientos por región.	51
8.	Asignación de intervalo de trabajadores ENIA 2019	67
9.	Tabla de centroides, elaborada con RStudio.	67