

Análisis de clúster de las PYMES de Ecuador

Title: Analysis of clusters of SMEs in Ecuador

Autor: Roberto Carlos Masabanda Pilataxi

Lugar de Trabajo: RML Consultores

Dirección: Reina Victoria N25-33 y Av. Colón. Edificio Banco Guayaquil. Piso 10 oficina 10 – 08 A. Quito. Ecuador.

E-mail: gerencia@rmlconsultores.com

Teléfono: 593- 2908588

Resumen

Se empleó el RapidMiner Studio para agrupar en cinco clústeres, 156 registros que corresponden PYMES y sectores económicos, de acuerdo con la clasificación, hasta el tercer dígito, según el Registro de Actividades Económicas vigente en Ecuador. El análisis de clústeres realizado ofrece una perspectiva más clara y multidimensional al agrupar: Número de Micro Empresas, Personal Ocupado, Remuneraciones, Producción y Consumo Intermedio, para los diferentes sectores económicos y permitió considerar las características particulares de algunos de estos y las correlaciones entre los factores analizados.

El software empleado (Rapidminer) resultó apropiado y de fácil implementación por su interfase gráfica permitiendo realizar gráficos ilustrativos y establecer correlaciones entre los parámetros considerados que pueden ser de gran utilidad para una mejor comprensión del estado actual y peso de las PYMES en la economía actual de Ecuador.

Palabras claves: Minería de datos, análisis clúster, Rapidminer, pymes, k-media clustering

Abstract

The RapidMiner Studio was used to group into five clusters, 156 records corresponding to SMEs and economic sectors according to the classification, up to the third digit, of the current Registry for Economic Activities in Ecuador. The performed cluster analysis provides a clearer and multidimensional perspective, grouping five factors: Number of Micro Enterprises, Occupied Personal, Earnings, Intermediate Production and Consumption, and allowed to consider the particular characteristics of some economic sectors under consideration and the correlations among them.

The Rapid Miner's graphical interface was appropriate and easily implemented allowing making illustrative graphs and establishing correlations between the considered parameters that can be useful for a better understanding of the current status and weight of SMEs in the current economy Ecuador.

Key words: Data mining, cluster analysis, Rapidminer, SMEs, k-means clustering,

Introducción

Las Pequeñas y Medianas Empresas desempeñan un papel, cada día más importante, en América Latina por su papel como generadoras de empleo y en la producción. En el caso particular de Ecuador desde hace años se han generado diferentes políticas para el apoyo a este tipo de Empresas [1]. El empleo de técnicas de minería de datos, como el agrupamiento en clústeres, puede brindar nuevos elementos para la clasificación de las mismas y ganar claridad en cuánto a su papel en la economía..

El análisis de las PYMES ha sido igualmente objeto de diferentes publicaciones y se han reportado 149 artículos sobre este tipo de empresas en revistas de 8 países latinoamericanos [2].

Esta comunicación tiene como objetivo el realizar un análisis de clústeres para los datos reportados [3] en relación con las PYMES ecuatorianas.

El análisis de clústeres consiste en asignar al conjunto de objetos seleccionados grupos denominados clústeres. Este tipo de análisis es típico de las etapas exploratorias de la minería de datos [4]. Existen para ellos diferentes softwares [5] y de estos seleccionamos el Rapid Miner [6] que brinda una interfase gráfica que permite manejar fácilmente la gran cantidad de operadores que ofrece este software para la minería de datos y ha sido empleado en otros casos [7], [8] para el estudio de datos relacionados con referencias bibliográficas, para lograr el agrupamiento de estas y facilitar su análisis.

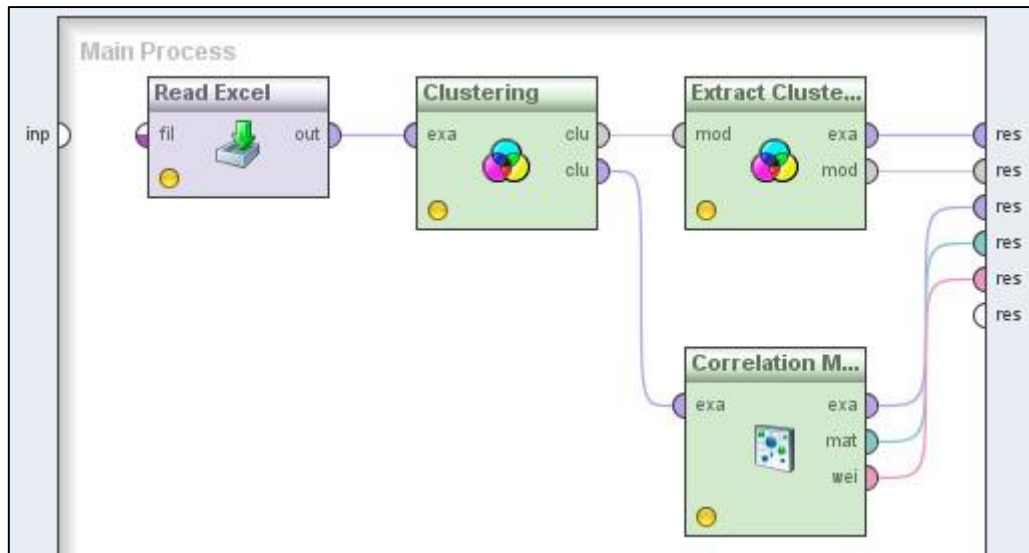
Métodos

La data a analizar se tomó de [3]. Para la clasificación de las Empresas se utilizó la Clasificación Nacional de Actividades Económicas [9], (CIUU). Se utilizó hasta el tercer dígito y con ello se conformó una base de 156 registros, con los datos siguientes para cada uno:

- Código CIUU
- Actividad Económica
- Número de Micro Empresas
- Personal Ocupado
- Remuneraciones
- Producción
- Consumo Intermedio

Para la minería de datos se utilizó el diagrama de procesos siguientes realizado en RapidMiner Estudio Versión 6.3.0:

Figura 1. Diagrama de procesos en RapidMiner para el agrupamiento en Clústeres



Como se observa los procesos se conectan y cada uno tributa al otro:

- 1) Lectura del Fichero de Datos en Excel. Permite a la vez fijar los atributos y parámetros a considerar.
- 2) Clusterización. El algoritmo empleado fue el de k-Medias, con mediciones numéricas y distancias euclidianas.
- 3) Extracción del Prototipo de clústeres obtenidos. Este proceso permite a la vez obtener los valores de los centroides de los clústeres obtenidos.
- 4) Determinación de la Matriz de correlación. Toma datos del módulo de clusterización y obtiene los coeficientes de correlación para las variables consideradas.

Resultados y discusión

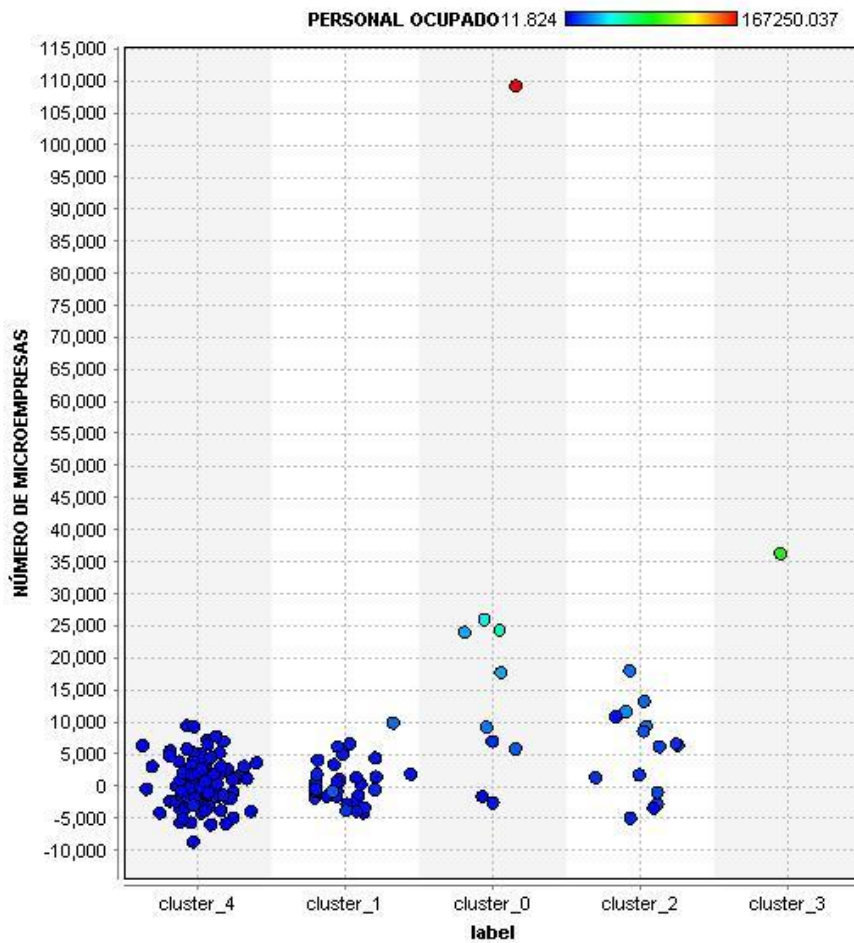
La aplicación del modelo de clusterización empleado clasificó los datos en los cuatro grupos que se muestran en la Tabla 1:

Tabla 1. Total de elementos por clúster

Clúster No.	Ítems
0	10
1	32
2	16
3	1
4	97
Total	156

La figura 2 a continuación ilustra claramente una nueva perspectiva en la clasificación de los datos, en este caso de acuerdo con el Número de Empresas y el Personal Ocupado.

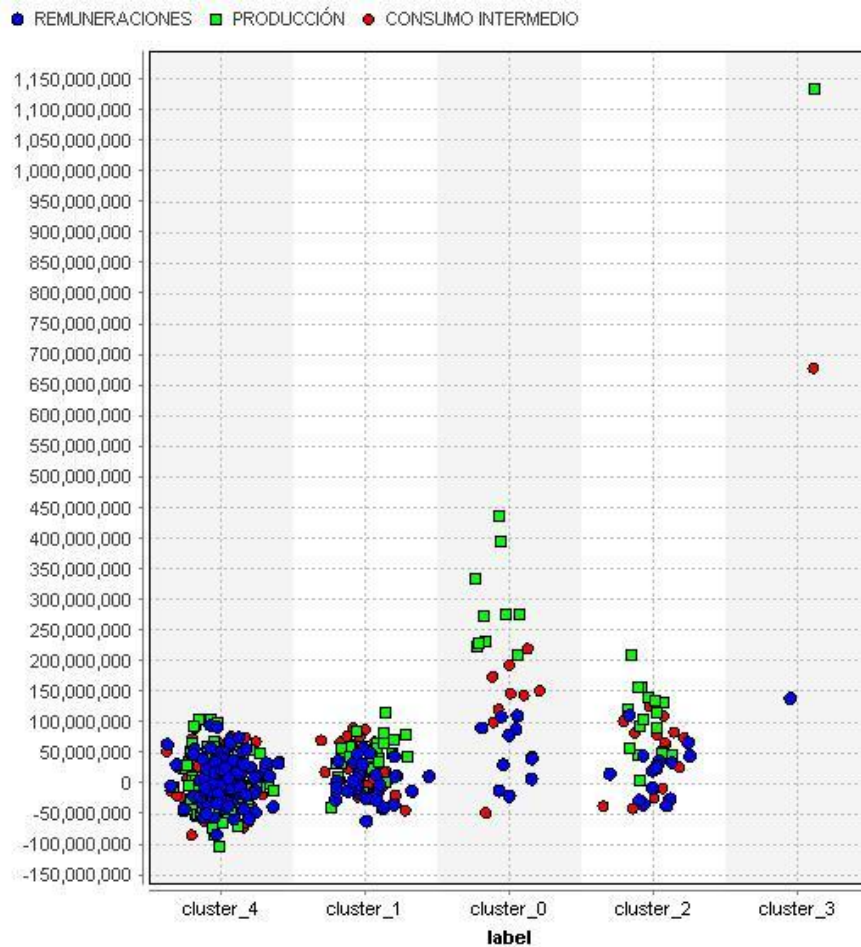
Figura 2. Número de Empresas y Personal Ocupado por Clústeres



Destaca el caso del Clúster 3 que corresponde a las Actividades de Restaurantes y de Servicio Móvil de Comidas. De igual forma en el Clúster 0, la singularidad que presenta el Sector de Venta al por Menor en Comercios no Especializados (punto rojo en la Figura 2).

El modelo de clusterización empleado permite a la vez una visualización interactiva con respecto, por ejemplo, a los factores: remuneraciones, volumen de producción y consumo interno, en dólares, que permite un análisis clasificatorio de las diferentes variables, para los 156 sectores considerados:

Figura 3. Total de Remuneraciones, Producción y Consumo Interno por Clústeres



Los valores para los Centroides para el modelo de clústeres obtenidos se presentan en la Tabla 2 a continuación:

Tabla 2. Centroides calculados para los atributos considerados y clústeres obtenidos

Atributo	Clúster 0	Clúster 1	Clúster 2	Clúster 3	Clúster 4
NÚMERO DE MICROEMPRESAS	20637.95839	1671.95673	5407.269903	34425.30491	209.4649949
PERSONAL OCUPADO	36504.55515	3835.43139	10637.41405	87558.24367	550.5961044
REMUNERACIONES	3.92E+07	8397119.45	1.83E+07	1.20E+08	1393892.888
PRODUCCIÓN	2.88E+08	4.39E+07	1.21E+08	1.08E+09	6180041.632
CONSUMO INTERMEDIO	1.15E+08	1.62E+07	4.97E+07	6.81E+08	2700100.284

La correlación entre las diferentes variables se muestra en la Tabla 3:

Tabla 3. Matriz de Correlación determinada

	NÚMERO DE MICROEMPRESAS	PERSONAL OCUPADO	REMUNERACIONES	PRODUCCIÓN	CONSUMO INTERMEDIO

NÚMERO DE MICRO-EMPRESAS	1.00	0.98	0.55	0.61	0.55
PERSONAL OCUPADO	0.98	1.00	0.68	0.73	0.68
REMUNERACIONES	0.55	0.68	1.00	0.88	0.83
PRODUCCIÓN	0.61	0.73	0.88	1.00	0.95
CONSUMO INTERMEDIO	0.55	0.68	0.83	0.95	1.00

Algunas de estas correlaciones son esperadas, por ejemplo el valor cercano a uno para la correlación entre personal ocupado y número de microempresas, o la correlación entre volumen de producción y consumo intermedio (0.95).

Otras como la baja correlación entre personal ocupado y remuneraciones (0.68) o entre personal ocupado y volumen de producción (0.73) o que este factor (volumen de producción), está más ligado a las remuneraciones (0.88) e ilustran el valor de poder agrupar en clústeres como un paso inicial de la minería de datos.

Resulta igualmente interesante el agrupamiento de sectores productivos diferentes de acuerdo con el Calificador CIUU pero que se agrupan de acuerdo con el modelo empleado que consideró los cinco factores mencionados, por ejemplo:

Tabla 4. Sectores agrupados en Clúster 0

C107	ELABORACIÓN DE OTROS PRODUCTOS ALIMENTICIOS
C251	FABRICACIÓN DE PRODUCTOS METÁLICOS PARA USO ESTRUCTURAL, TANQUES, DEPÓSITOS, RECIPIENTES DE METAL Y GENERADORES DE VAPOR
G452	MANTENIMIENTO Y REPARACIÓN DE VEHÍCULOS AUTOMOTORES
G463	VENTA AL POR MAYOR DE ALIMENTOS, BEBIDAS Y TABACO
G471	VENTA AL POR MENOR EN COMERCIOS NO ESPECIALIZADOS
G475	VENTA AL POR MENOR DE OTROS ENSERES DOMÉSTICOS EN COMERCIOS ESPECIALIZADOS
G477	VENTA AL POR MENOR DE OTROS PRODUCTOS EN COMERCIOS ESPECIALIZADOS
H492	OTRAS ACTIVIDADES DE TRANSPORTE POR VÍA TERRESTRE
H532	ACTIVIDADES DE MENSAJERÍA
S960	OTRAS ACTIVIDADES DE SERVICIOS PERSONALES

O para el clúster 2:

Tabla 4. Sectores agrupados en Clúster 2

C141	FABRICACIÓN DE PRENDAS DE VESTIR, EXCEPTO PRENDAS DE PIEL
C181	IMPRESIÓN Y ACTIVIDADES DE SERVICIOS RELACIONADOS CON LA IMPRESIÓN

C239	FABRICACIÓN DE PRODUCTOS MINERALES NO METÁLICOS NCP
C310	FABRICACIÓN DE MUEBLES
C331	REPARACIÓN DE PRODUCTOS ELABORADOS DE METAL, MAQUINARIA Y EQUIPO
G453	VENTA DE PARTES, PIEZAS Y ACCESORIOS PARA VEHÍCULOS AUTOMOTORES
G466	OTRAS ACTIVIDADES DE VENTA AL POR MAYOR ESPECIALIZADA
G472	VENTA AL POR MENOR DE ALIMENTOS, BEBIDAS Y TABACO EN COMERCIOS ESPECIALIZADOS
H522	ACTIVIDADES DE APOYO AL TRANSPORTE
I551	ACTIVIDADES DE ALOJAMIENTO PARA ESTANCIAS CORTAS
I563	ACTIVIDADES DE SERVICIO DE BEBIDAS
J619	OTRAS ACTIVIDADES DE TELECOMUNICACIONES
L681	ACTIVIDADES INMOBILIARIAS REALIZADAS CON BIENES PROPIOS O ARRENDADOS
Q862	ACTIVIDADES DE MÉDICOS Y ODONTÓLOGOS
R932	OTRAS ACTIVIDADES DE ESPARCIMIENTO Y RECREATIVAS
S952	REPARACIÓN DE EFECTOS PERSONALES Y ENSERES DOMÉSTICOS

Estas tablas ilustran una nueva perspectiva en cuanto al criterio de clasificación por sectores económicos.

Conclusiones

El agrupamiento por clústeres debe considerarse como una etapa previa a otros procesos de minería de datos. En ese caso se logró agrupar en cinco clústeres 156 registros que corresponden a la agrupación, hasta el tercer dígito, de acuerdo con el clasificador de Actividades Económicas [9], (CIUU).

El análisis realizado ofrece una perspectiva más clara y multidimensional al agrupar considerando cinco factores: Número de Micro Empresas, Personal Ocupado, Remuneraciones, Producción y Consumo Intermedio para diferentes sectores económicos, permitió considerar las características particulares de algunos de los sectores económicos considerados y presentar su agrupamiento desde una perspectiva multidimensional.

El software empleado (Rapidminer) resultó apropiado y de fácil implementación por su interfase gráfica permitiendo realizar gráficos ilustrativos y establecer correlaciones entre los parámetros considerados que pueden ser de gran utilidad para una mejor comprensión del estado actual y peso de las PYMES en la economía actual de Ecuador.

Bibliografía

- [1]. Burneo, María del Carmen and Wilson Miño. "Políticas E Instituciones De Apoyo a Las Pyme En El Ecuador." En: Políticas de apoyo a las PYME en América Latina: entre avances innovadores y desafíos institucionales. Santiago: CEPAL, 2010. LC/G. 2421-P. p. 209-244, (2010)
- [2]. Masabanda Pilataxi, Roberto, & Jorge Alberto González Alonso. "Análisis del impacto de las publicaciones sobre PYMES en revistas latinoamericanas. Parte 1." *Revista Publicando* [En línea], 1.1 p. 27-36. Web. (2014)
- [3]. Directorio de Empresas. [En línea], <http://www.ecuadorencifras.gob.ec/directoriodeempresas/>. (2015).
- [4]. Jain, Anoop Kumar and Satyam Maheswari. "Survey of Recent Clustering Techniques in Data Mining." *Int. J. Comput. Sci. Manage. Res* 1, (2012): p. 72-78
- [5] Jovic, A, K Brkic and N Bogunovic. "An Overview of Free Software Tools for General Data Mining." In *Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2014 37th International Convention on*, p. 1112-1117: IEEE, (2014).
- [6] Kotu, Vijay and Bala Deshpande. *Predictive Analytics and Data Mining: Concepts and Practice with Rapidminer*: Morgan Kaufmann, (2014).
- [7] Pazmiño Santacruz, Mauro Roberto and Jorge Alberto González Alonso. "Análisis Exploratorio Sobre Las Publicaciones Relacionadas Con La Comunicación Organizacional En Pymes." *2014 1*, no. 1 (2014): 9.
- [8] Alonso, Jorge Alberto González and Yudeisy Pérez González. "Análisis De Las Revistas Latinoamericanas De Acceso Abierto. El Caso Ecuador." *Revista Publicando 2*, no. 1 p. 12-23. (2015):
- [9] Clasificación Ampliada de las Actividades Económicas (CIU REV. 4.0). [En línea], <http://www.inec.gob.ec/estadisticas/SIN/ciu4.php>



www.sociedadelainformacion.com

Edita:



Director: José Ángel Ruiz Felipe
 Jefe de publicaciones: Antero Soria Luján
 D.L.: AB 293-2001
 ISSN: 1578-326x