



Title: Propuesta de un sistema simple de recomendación para pequeñas y medianas empresas para la toma de decisiones basado en el aprendizaje no supervisado

Authors: URUETA-HINOJOSA, Daniel Edahi LARA-VELÁZQUEZ, Pedro, GUTIÉRREZ-ANDRADE, Miguel Ángel y DE LOS COBOS-SILVA, Sergio Gerardo

Editorial label ECORFAN: 607-8695

BCONIMI Control Number: 2019-016

BCONIMI Classification (2019): 050319-0016

Pages: 16

RNA: 03-2010-032610115700-14

ECORFAN-México, S.C.

143 – 50 Itzopan Street

La Florida, Ecatepec Municipality

Mexico State, 55120 Zipcode

Phone: +52 1 55 6159 2296

Skype: ecorfan-mexico.s.c.

E-mail: contacto@ecorfan.org

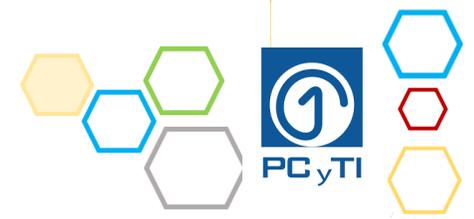
Facebook: ECORFAN-México S. C.

Twitter: @EcorfanC

www.ecorfan.org

Holdings

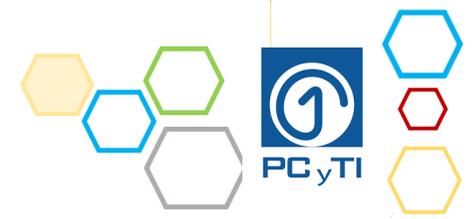
| | | |
|---------|-------------|------------|
| Mexico | Colombia | Guatemala |
| Bolivia | Cameroon | Democratic |
| Spain | El Salvador | Republic |
| Ecuador | Taiwan | of Congo |
| Peru | Paraguay | Nicaragua |



Agenda

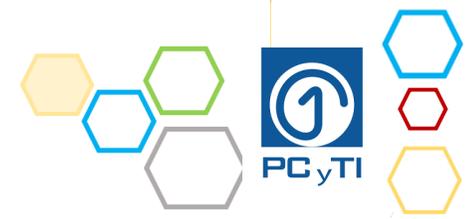
- Introducción
- Objetivos
- Estado del arte
- Metodología del modelo propuesto
- Resultados
- Conclusiones





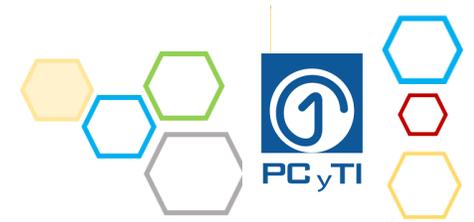
Introducción

- Los sistemas de recomendación generalmente son complicados porque combinan diversos enfoques de la inteligencia artificial principalmente del aprendizaje supervisado
- Una desventaja de estos sistemas es que debe existir una etiqueta previa para poder entrenarlo y así pueda clasificar
- La tarea de clasificación suele ser costosa



Objetivos

- Desarrollar una propuesta de un sistema simple de recomendación basado en aprendizaje no supervisado que ayude a las pequeñas y medianas empresas en la toma de decisiones para clasificación de sus productos
- Crear una propuesta que permita aminorar costos en la tarea de clasificación de sus productos



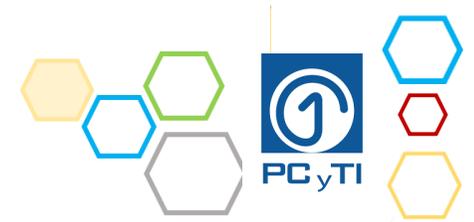
Estado del arte (I)

***k*-medias**

La meta es particionar los puntos (observaciones) en k clusters basándose en la similitud de sus características

Inicio

1. Elige aleatoriamente k centros de cluster
 2. Mientras los puntos dejen de ser asignados a los centroides
Asignar cada punto de los datos al centro del cluster más cercano
Establecer los centroides del cluster basándose en la posición promedio(media)de cada punto de centroide
 3. Fin Mientras
- Fin

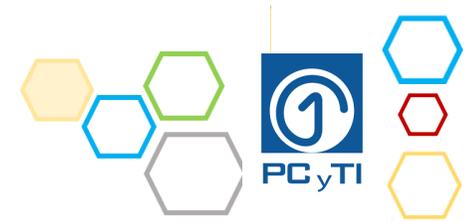


Estado del arte (II)

Método del codo

- Consiste en la ejecución del algoritmo k -medias para una instancia dada en un rango definido
- En cada iteración se calcula la suma de los errores cuadrados (SSE)
- SSE es la suma de las diferencias cuadradas entre cada observación y la media de su grupo

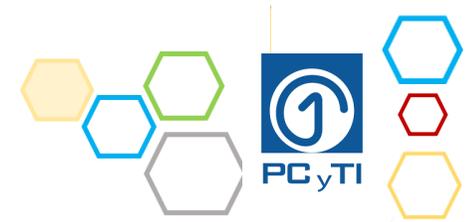
$$SSE = \sum_{i=1}^n (x_i - \bar{x})^2$$



Estado del arte (II)

Método del codo

- Finalmente, los resultados se grafican
- Si el gráfico de líneas se parece a un brazo, entonces el "codo" en el brazo es el valor de k que es el mejor



Estado del arte (III)

Matriz de confusión

- Es una matriz cuadrada de $n \times n$ donde las filas son las clases reales, mientras que las columnas son las clases encontradas por el modelo
- Cada campo en la matriz muestra un tipo de error diferente



Estado del arte (III)

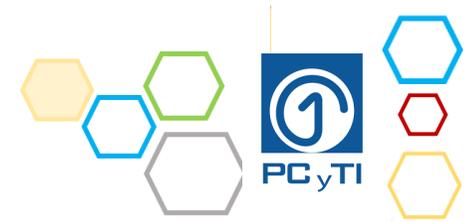
| | Negativo (Modelo) | Positivo (Modelo) |
|-----------------|-------------------------|-------------------------|
| Negativo (Real) | Verdadero Negativo (VN) | Falso Positivo (FP) |
| Positivo (Real) | Falso Negativo (FN) | Verdadero Positivo (VP) |

Tabla 1. Matriz de confusión de 2 clases

$$\textit{Precisión} = \frac{VN + VP}{VN + FP + FN + VP}$$

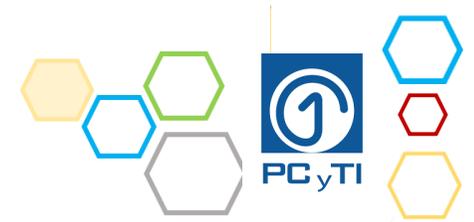
$$\textit{Error tipo I} = \frac{FP}{VN + FP}$$

$$\textit{Error tipo II} = \frac{FN}{FN + VP}$$



Metodología del modelo propuesto

1. Cargar la base de datos
2. Eliminar filas con datos faltantes
3. Cambiar valores alfanuméricos a numéricos
4. Normalizar los datos
5. Aplicar el método del codo para obtener el número de clases
6. Aplicar k -medias
7. Mostrar los resultados mediante una matriz de confusión
8. Tomar un elemento de cada grupo y clasificarlo para obtener la clasificación de todos los demás



Resultados (I)

Base de datos Wine

- Análisis químico de 178 instancias de vinos
- 13 atributos
- No hay datos faltantes, todos los valores son numéricos
- Distribuidos en 3 clases:
 - Vinos malos: 59
 - Vinos regulares: 71
 - Vinos buenos: 48



Resultados (I)

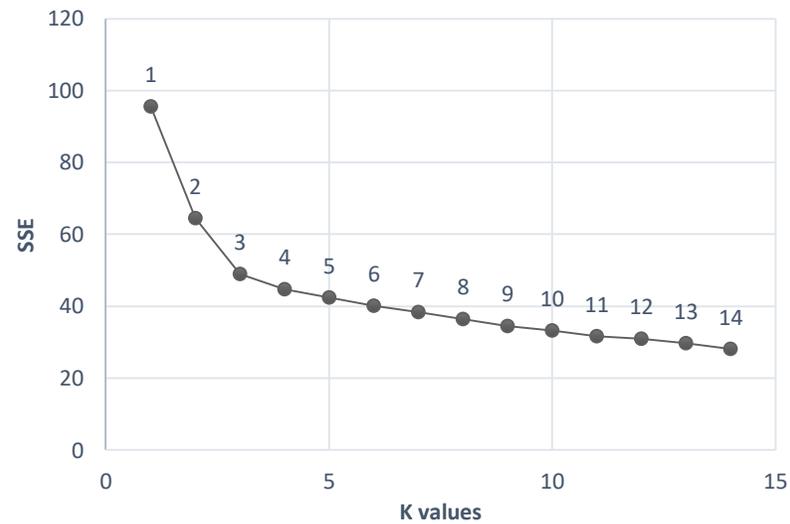


Figura 1. Método del codo para la instancia Vinos

| | | |
|----|----|----|
| 51 | 0 | 0 |
| 0 | 65 | 0 |
| 0 | 0 | 62 |

Figura 2. Matriz de confusión para la instancia Vinos usando el modelo propuesto

Resultados (I)

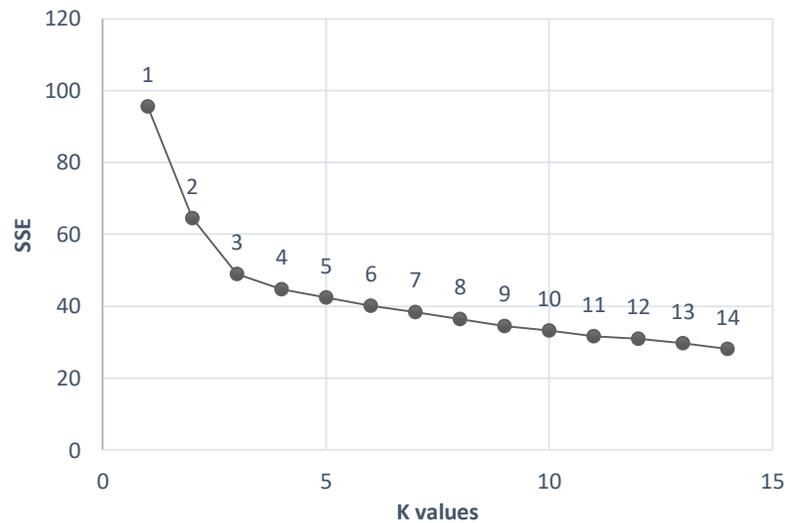
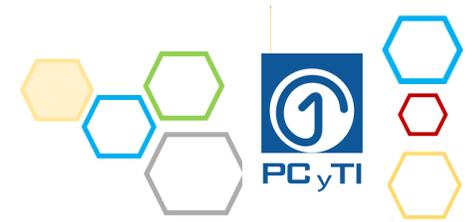


Figura 3. Método del codo para la instancia Vinos con clasificación previa

| | | |
|----|----|----|
| 59 | 0 | 0 |
| 3 | 65 | 3 |
| 0 | 0 | 48 |

Figura 4. Matriz de confusión para la instancia Vinos con clasificación previa usando el modelo propuesto



Resultados (II)

Base de datos Automóviles

- Análisis químico de 205 automóviles
- 26 atributos
- Hay datos faltantes, valores alfanuméricos
- Diversos atributos pueden ser usados como etiqueta
- Los actuarios sugieren una escala de 7 tomando en cuenta el factor de riesgo asociado a su precio

Resultados (II)

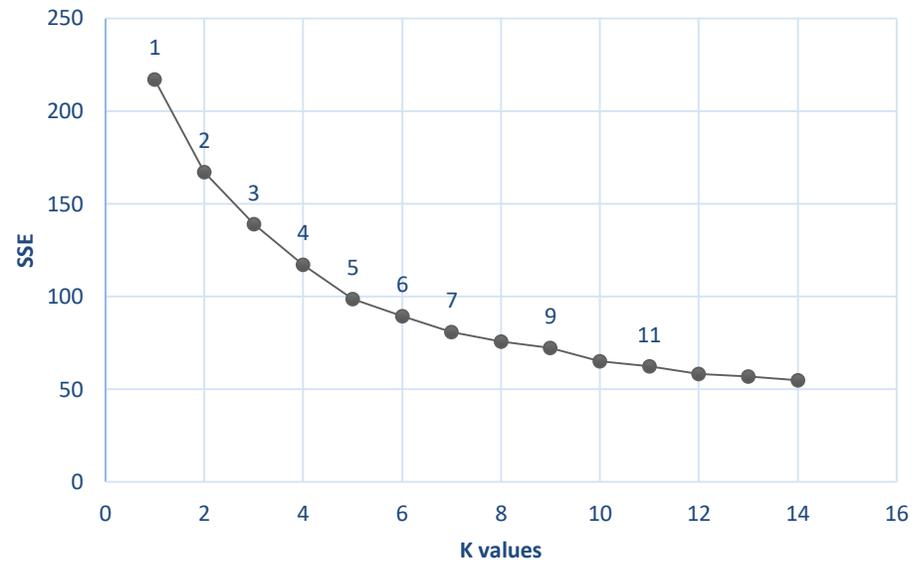
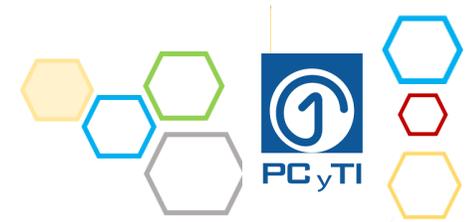


Figura 5. Método del codo para la instancia Automóviles

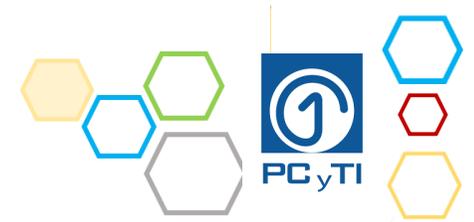
| | | | | | |
|---|----|----|----|----|---|
| 7 | 0 | 0 | 0 | 0 | 0 |
| 0 | 49 | 0 | 0 | 0 | 0 |
| 0 | 0 | 24 | 0 | 0 | 0 |
| 0 | 0 | 0 | 42 | 0 | 0 |
| 0 | 0 | 0 | 0 | 28 | 0 |
| 0 | 0 | 0 | 0 | 0 | 9 |

Figura 6. Matriz de confusión para la instancia Automóviles usando el modelo propuesto



Conclusiones (I)

- El modelo propuesto cumple la función de ser simple y, al mismo tiempo, poder clasificar las instancias de prueba
- Cuando el modelo obtiene el resultado del sistema, solamente es necesario analizar una de las muestras que forman un grupo para determinar la clasificación de todos los demás elementos que pertenecen al mismo grupo



Conclusiones (II)

- Del mismo modo, en el caso de que exista una clasificación previa, el modelo sirve para verificar estas clasificaciones y, si es necesario, realizar un análisis adicional para evitar productos posiblemente sobrevalorados o infravalorados



ECORFAN®

© Ecorfan-Mexico, S.C.

No part of this document covered by the Federal Copyright Law may be reproduced, transmitted or used in any form or medium, whether graphic, electronic or mechanical, including but not limited to the following: Citations in articles and comments Bibliographical, compilation of radio or electronic journalistic data. For the effects of articles 13, 162,163 fraction I, 164 fraction I, 168, 169,209 fraction III and other relative of the Federal Law of Copyright. Violations: Be forced to prosecute under Mexican copyright law. The use of general descriptive names, registered names, trademarks, in this publication do not imply, uniformly in the absence of a specific statement, that such names are exempt from the relevant protector in laws and regulations of Mexico and therefore free for General use of the international scientific community. BCONIMI is part of the media of Ecorfan-Mexico, S.C., E: 94-443.F: 008- (www.ecorfan.org/ booklets)