
APUNTES LIBRO

Introducción a la Minería de Datos

Alvarado Becerra Ludwig
Libro de José Hernández Orallo
24 de mayo de 2024

Índice

1. ¿Qué es la minería de datos?	3
1.1. Determinar grupos diferenciados de empleados	3
1.2. Tipos de datos	5

1. ¿Qué es la minería de datos?

La minería de datos tiene su fundamentación en la estadística y en el *Machine Learning*. Utiliza grandes cantidades de datos para generar conocimiento a partir de los datos, mientras que en otras ramas de estadística la idea era organizar los datos y obtener información a partir de ella a partir de consultas como lo era en el caso de SQL. En la minería de datos se buscan patrones que puedan ayudarnos a generar conocimiento a partir de esos datos, es ampliamente útil en todo tipo de industria, ingeniería, campo de la salud, todo lado!

Según el autor, el objetivo de la minería de datos es convertir datos en conocimiento. Para eso mismo, expone unos ejemplos, de los cuales voy a detallar sólo en uno que fue el que más me gustó ;)

1.1. Determinar grupos diferenciados de empleados

El departamento de recursos humanos de una gran empresa desea categorizar a sus empleados en distintos grupos con el objetivo de entender mejor su comportamiento y tratarlos de manera adecuada. Para ello dispone en sus bases de datos de información sobre los mismos (sueldo, estado civil, si tiene coche, número de hijos, si su casa es propia o de alquiler, si está sindicado, número de bajas al año, antigüedad y sexo). La tabla 1 muestra algunos de los registros de su base de datos.

id	Suel-do	Ca-sa-do	Co-che	Hi-jos	Alq/-prop	Sindi-cado	Baja-s/Año	An-tigüedad	Sexo
1	1.000	Sí	No	0	Alquiler	No	7	15	H
2	2.000	No	Sí	1	Alquiler	Sí	3	3	M
3	1.500	Sí	Sí	2	Prop	Sí	5	10	H
4	3.000	Sí	Sí	1	Alquiler	No	15	7	M
5	1.000	Sí	Sí	0	Prop	Sí	1	6	H
6	4.000	No	Sí	0	Alquiler	Sí	3	16	M
7	2.500	No	No	0	Alquiler	Sí	0	8	H
8	2.000	No	Sí	0	Prop	Sí	2	6	M
9	2.000	Sí	Sí	3	Prop	No	7	5	H
10	3.000	Sí	Sí	2	Prop	No	1	20	H
11	5.000	No	No	0	Alquiler	No	2	12	M
12	800	Sí	Sí	2	Prop	No	3	1	H
13	2.000	No	No	0	Alquiler	No	27	5	M
14	1.000	No	Sí	0	Alquiler	Sí	0	7	H
15	800	No	Sí	0	Alquiler	No	3	2	H

⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
---	---	---	---	---	---	---	---	---	---

Cuadro 1: Datos de los empleados

Un sistema de minería de datos podría obtener tres grupos con la siguiente descripción:

Grupo 1:

- Sueldo: 1.535,2 €.
- Casado: No -> 0,777 — Sí -> 0,223
- Coche: No -> 0,82 — Sí -> 0,18
- Hijos: 0,05
- Alq/Prop: Alquiler -> 0,99 — Propia -> 0,01
- Sindic.: No -> 0,8 — Sí -> 0,2
- Bajas/Año: 8,3
- Antigüedad: 8,7
- Sexo: H -> 0,61 — M -> 0,39

Grupo 2:

- Sueldo: 1.428,7 €.
- Casado: No -> 0,98 — Sí -> 0,02
- Coche: No -> 0,01 — Sí -> 0,99
- Hijos: 0,3
- Alq/Prop: Alquiler -> 0,75 — Propia -> 0,25
- Sindic.: Sí -> 1,0
- Bajas/Año: 2,3
- Antigüedad: 8
- Sexo: H -> 0,25 — M -> 0,75

Grupo 3:

- Sueldo: 1.233,8 €.
- Casado: Sí -> 1,0
- Coche: No -> 0,05 — Sí -> 0,95
- Hijos: 2,3
- Alq/Prop: Alquiler -> 0,17 — Propia -> 0,83
- Sindic.: No -> 0,67 — Sí -> 0,33
- Bajas/Año: 5,1
- Antigüedad: 8,1
- Sexo: H -> 0,83 — M -> 0,17

Estos grupos podrían ser interpretados por el departamento de recursos humanos de la siguiente manera:

- Grupo 1: sin hijos y con vivienda de alquiler. Poco sindicados. Muchas bajas.
- Grupo 2: sin hijos y con coche. Muy sindicados. Pocas bajas. Normalmente son mujeres y viven en casas de alquiler.
- Grupo 3: con hijos, casados y con coche. Mayoritariamente hombres propietarios de su vivienda. Poco sindicados.

1.2. Tipos de datos

La minería de datos se puede aplicar a cualquier tipo de información, lo que cambia son las técnicas que se aplican a los datos. Se diferencian entre datos relacionados y no estructurados.