

SENSOR

METODOLOGÍA

● Selección de variables

La base de datos original de CompraNet cuenta con 44 variables, el equipo de Sensor la pudo reducir a 27, que son las que se pueden observar en el apartado de **Contratos**:

1. Institución
2. Clave de la UC
3. Nombre de la UC
4. Código del expediente
5. Clave CUCOP
6. Título del expediente
7. Número del procedimiento
8. Fecha de fallo
9. Fecha de publicación
10. Fecha de apertura
11. Carácter del procedimiento
12. Tipo de contratación
13. Tipo de procedimiento
14. Forma de participación
15. Código del contrato
16. Título del contrato
17. Descripción del contrato
18. Fecha de inicio del contrato
19. Fecha de fin del contrato
20. Importe del contrato
21. Moneda del contrato

22. Estatus del contrato
23. Fecha de firma del contrato
24. Compra consolidada
25. Proveedor o contratista
26. Estratificación de la empresa
27. Dirección del anuncio

No todas son útiles para el análisis, ya que debido al tipo de modelo que se utilizará se necesitan variables numéricas o con pocas categorías, de estas 27, solamente 14 cuentan con las características para realizar el cruce de información en el modelo.

1. Fecha de fallo
2. Fecha de publicación
3. Fecha de apertura
4. Carácter del procedimiento
5. Tipo de contratación
6. Tipo de procedimiento
7. Forma de participación
8. Fecha de inicio del contrato
9. Fecha de fin del contrato
10. Importe del contrato
11. Moneda del contrato
12. Estatus del contrato
13. Compra consolidada
14. Estratificación de la empresa

De estas 14, solamente 11 se tienen sospechas que **puedan detectar indicios de corrupción**.

1. Fecha de fallo (Conclusión del proceso de licitación)
2. Fecha de publicación (Cuando se publicó el anuncio en CompraNet)
3. Fecha de apertura (Cuando se recibirán las propuestas)
4. Fecha de inicio de contrato (Cuando entra en vigor el contrato después de la firma)
5. Fecha de fin de contrato (Cuando expira el contrato)

6. Importe del contrato
7. Tipo de contratación
8. Tipo de procedimiento
9. Forma de participación
10. Compra consolidada
11. Estratificación de la empresa

Específicamente de estas 14 se desprenden tres **constantes**:

- Tiempo que se tardan en tomar una decisión (fecha de fallo-fecha de apertura)
- Tiempo de duración del contrato (fecha de fin del contrato-fecha de inicio del contrato)
- Importe del contrato.

Y se cruzaron estas tres constantes con 4 variables, para detectar indicios de corrupción.

- Tipo de contratación
- Tipo de procedimiento
- Forma de participación
- Estratificación de la empresa

● **Modelos de aprendizaje**

Ahora bien, para la detección de indicios de corrupción. Se utilizaron **modelos de aprendizaje de máquinas automáticos** (Machine Learning) que corresponden a una sub-área de la inteligencia artificial.

El primer algoritmo corresponde al de **K-Means**. Que es un algoritmo diseñado para determinar la semejanza entre puntos en n-dimensiones a partir de características o variables numéricas. Dicho esto, se crearon variables calculadas a partir de otras variables que servirán como parámetros del modelo:

1. Pesos al día: Que hace referencia al promedio de pesos que se gastaron al día durante el contrato. Es la división del importe del contrato entre la duración en días del contrato.

2. Días apertura de la firma: Es la cantidad de días entre el inicio del contrato y la apertura de la empresa.
3. Duración de apertura: Es la cantidad de días entre la fecha de fallo del contrato y la fecha de apertura del contrato.

Cabe recalcar que estas variables tuvieron que seguir un proceso de normalización estadística con el **método de Box-Cox** para que el modelo funcionara correctamente.

Siguiendo un procedimiento similar, se implementó un modelo de **Bosques Isolados** que funciona para detectar anomalías (o en nuestro caso, indicios de corrupción) en un conjunto de datos. El objetivo de este modelo no es realizar predicciones o clasificaciones si no, **detectar anomalías en base a particiones de los datos originales** y no requiere de que los datos estén previamente clasificados, por ende, su eficiencia en detectar anomalías.

De esta manera, la combinación de estos dos modelos nos dan una precisión de cuáles contratos pueden tener indicios de corrupción.