

Por
Liliana Ochoa Echeverri
CC 1036927400

Descripción del problema predictivo a resolver

La clasificación ATC es un sistema europeo de codificación de sustancias farmacéuticas y medicamentos con arreglo al sistema u órgano efector y al efecto farmacológico, las indicaciones terapéuticas y la estructura química de un fármaco [1p], de allí la importancia de estudiar el índice de ventas de sus ocho diferentes categorías para lograr obtener datos que indiquen el mayor consumo por órgano afectado ya que el ATC, recoge el sistema u órgano sobre el que actúa [3w]. En base a esto, se desea predecir si la cantidad vendida y exportada desde el sistema de punto de vista en el individuo farmacia es óptima y suficiente para la demanda que tendrá [2k].

Dataset que se va a utilizar

El dataset que se va a utilizar fue extraído de Kaggle y se llama Datos de Ventas Farmacéuticas [2k], este dataset contiene cerca de consta de 600 000 datos transaccionales recopilados en 6 años (período 2014-2019), que indican la fecha y la hora de venta, la marca del medicamento farmacéutico y la cantidad vendida, exportados desde el sistema de punto de venta en el individuo farmacia. El grupo seleccionado de medicamentos del conjunto de datos (57 medicamentos) se clasifica en las siguientes categorías del Sistema de clasificación químico terapéutico anatómico (ATC):

M01AB - Productos antiinflamatorios y antirreumáticos, no esteroides, derivados del ácido acético y sustancias relacionadas

M01AE - Productos antiinflamatorios y antirreumáticos, no esteroides, derivados del ácido propiónico

N02BA - Otros analgésicos y antipiréticos, ácido salicílico y derivados

N02BE/B - Otros analgésicos y antipiréticos, Pirazolonas y Anilidas

N05B - Medicamentos psicodélicos, Medicamentos ansiolíticos

N05C - Drogas psicodélicas, hipnóticas y sedantes

R03 - Medicamentos para enfermedades obstructivas de las vías respiratorias

R06 - Antihistamínicos para uso sistémico

Los datos de ventas son remuestreados a los períodos horarios, diarios, semanales y mensuales. Los datos ya están preprocesados, donde el procesamiento incluyó la detección y el tratamiento de valores atípicos y la imputación de datos faltantes.

Además, el dataset contiene 4 archivos con datos diarios, semanales, anuales y saleshourly contenidos en 44 columnas donde se podrán predecir mejor las estadísticas según su predicción.

Métricas de desempeño requeridas (machine learning y negocio)

Ya que se trabajará con un modelo de clasificación se decidió usar las métricas de desempeño más comunes en este tipo de modelos; se quiere monitorear el modelo partiendo de una matriz de confusión, y analizar los resultados que se obtienen de hacer cálculos con los datos que arroja dicha matriz. Las métricas que se usarán son las siguientes:

1. Machine Learning:

- *Matriz de confusión o error* (Confusión o error Matrix es una tabla que describe el rendimiento de un modelo supervisado de Machine Learning en los datos de prueba, donde se desconocen los verdaderos valores.) [4]
- *Precisión* (Es el porcentaje total de elementos clasificados correctamente)
- *Recall o sensibilidad o TPR (Tasa positiva real)*
- *Exactitud* (En general, ¿con qué frecuencia es correcto el clasificador?) [4]

2. Negocio

- *Margen de utilidad bruta:* el margen de utilidad bruta se puede obtener restando el costo de los bienes vendidos a las ventas netas de la empresa.

Primer criterio sobre el desempeño deseable en producción

Según el foro de Salud Digital FDS, es importante tener presente, además de las categorías, los 5 niveles de la clasificación jerárquica [5] para tener como primer criterio el orden de venta según esta clasificación y así evaluar más precisamente el órgano que a menudo sufre más afectación según su jerarquía

Referencias

- [1] Kaggle (2019). Datos de ventas farmacéuticas. Recuperado de:
<https://www.kaggle.com/datasets/milanzdravkovic/pharma-sales-data>
- [2] Tremedica(2022) . El sistema de clasificación ATC de sustancias farmacéuticas para uso humano. Recuperado de :
https://www.tremedica.org/wp-content/uploads/n15_tribuna-Saladrigas.pdf
- [3] Foro Salud Digital. Para qué sirve la clasificación ATC y para qué no sirve (para recetar o uso clínico) Recuperado de :
<https://discourse.forosaluddigital.cl/t/para-que-sirve-la-clasificacion-atc-y-para-que-no-sirve-para-recetar-o-uso-clinico/777>
- [4] Machine Learning: Seleccin Metricas de clasificación. Recuperado de:
[https://sitiobigdata.com/2019/01/19/machine-learning-metrica-clasificacion-parte-3/#:~:text=Presentemos%20primero%20las%20principales%20m%C3%A9tricas,o%20TPR%20\(Tasa%20positiva%20real\)](https://sitiobigdata.com/2019/01/19/machine-learning-metrica-clasificacion-parte-3/#:~:text=Presentemos%20primero%20las%20principales%20m%C3%A9tricas,o%20TPR%20(Tasa%20positiva%20real))
- [5] Wikipedia. Código ATC. Recuperado de:
https://es.wikipedia.org/wiki/C%C3%B3digo_ATC#:~:text=El%20c%C3%B3digo%20ATC%20o%20Sistema,medicamentos%2C%20organizados%20seg%C3%BAn%20grupos%20terap%C3%A9uticos.