

Data Understanding

Problem Description



Jorge Alberto Jaramillo Bermúdez
Health Data Scientist
18 de enero de 2024

ABC Pharma Company faces a challenge in comprehending drug persistency based on physician prescriptions. To address this issue, the company aims to automate the identification process. The primary objective is to construct a classification model for identifying drug persistency.

Data Understanding:

- Data Type: .xlsx
- Data Shape: 3424 rows by 69 columns
- Variable Types: 67 categorical variables, 2 numerical variables

Issues with the Data:

- Number of NA values: 0
- Number of duplicates: 0
 - Outliers:
 - Identified in numerical columns: Dex_Freq_During_Rx and Count_Of_Risk
 - Skewed data:
 - Present in numerical columns
 - Class imbalance in the target variable:
 - Slight imbalance detected

Solutions to the Problems:

- Outliers:
- Utilize the Inter-Quartile-Range method to drop outliers.
- Skewed Data:
- Drop outliers (successful)
- Apply log transformation to numerical features
- Implement min-max scaling on numerical features

- Class Imbalance:
- Oversample the minority class
- Downsample the majority class

Vocabularies:

- **Drug Persistency:** The act of continuing treatment for the prescribed duration.
- **Adherence:** The extent to which a patient follows the prescribed interval and dose of a dosing regimen.
- **IDN:** A network of healthcare providers and facilities offering a full range of healthcare services in a specific geographic region.
- **NTM:** Non-Tuberculosis Mycobacteria
- **Rx:** Medical prescription
- **DEXA Scan:** Also known as a bone density scan, used to measure calcium and other minerals in the bone.
- **Fragility Fracture:** A fracture resulting from a fall.
- **Glucocorticoid:** Steroid hormones used for treating inflammation, autoimmune diseases, and cancer.
- **Injectable Drugs:** Drugs administered through injections.
- **Comorbidity:** The condition of having two or more diseases simultaneously.
- **Concomitancy:** Existing or occurring together. Concomitant medication refers to other prescription drugs, over-the-counter drugs, or dietary supplements taken alongside the investigated drug.
- **T Score:** Measures how much a bone density is higher or lower than that of a healthy 30-year-old adult.

