Data Glacier

Your Deep Learning Partner

JANUARY 3, 2024.

DATA SCIENCE - HEALTHCARE

# PERSISTENCY OF A DRUG.

abc FARMA


# Jorge Jaramillo Bermudez

## HEALTH DATA SCIENTIST.


# PROBLEM STATEMENT REPORT FOR ABC PHARMA: AUTOMATING DRUG PERSISTENCY INSIGHTS

Understanding the factors influencing drug persistency, defined as the continued use of a prescribed medication, is a major challenge for pharmaceutical companies like ABC Pharma. This information is crucial for optimizing drug development, marketing, and patient adherence. Traditionally, this process has been manual and time-consuming, hindering timely insights.

Problem Statement Report for ABC Pharma: Automating Drug Persistency Insights

Problem:

Understanding the factors influencing drug persistency, defined as the continued use of a prescribed medication, is a major challenge for pharmaceutical companies like ABC Pharma. This information is crucial for optimizing drug development, marketing, and patient adherence. Traditionally, this process has been manual and time-consuming, hindering timely insights.

Solution:

ABC Pharma partnered with an analytics company to automate the identification of factors impacting drug persistency using machine learning. This report outlines the ML approach and insights gained from the project.

ML Problem:

The objective is to build a classification model that identifies patients likely to persist with their medication based on various factors. The target variable is a "Persistency Flag" indicating whether a patient continued medication for a specific duration.

Data and Tasks:

1. Problem Understanding:
    - Identify relevant business questions and metrics to assess the model's success.

- Analyze the data landscape and potential challenges.

2. Data Understanding:
   - Explore the dataset characteristics, including features, data types, and missing values.
   - Perform basic data visualization to understand data distribution and potential relationships between features.
3. Data Cleaning and Feature Engineering:
   - Handle missing values and outliers appropriately.
   - Create new features based on domain knowledge and data analysis to improve model performance.
4. Model Development:
   - Implement and train various classification algorithms like Logistic Regression, Random Forest, and Gradient Boosting Machines.
   - Optimize hyperparameters for each model to maximize performance.
5. Model Selection:
   - Compare the performance of different models using metrics like accuracy, precision, recall, and ROC-AUC.
   - Select the model with the best overall performance and interpretability.
6. Model Evaluation:
   - Evaluate the final model on unseen data to assess its generalizability.
   - Analyze the model's confusion matrix to understand misclassification patterns.
7. Reporting:

- Report accuracy, precision, recall, and ROC-AUC for both classes of the target variable.
- Provide visualizations of the model's performance and important features.

8. Deployment and Monitoring:
   - Deploy the selected model into a production environment to generate insights on drug persistency.
   - Continuously monitor the model's performance and retrain it as needed to maintain accuracy.

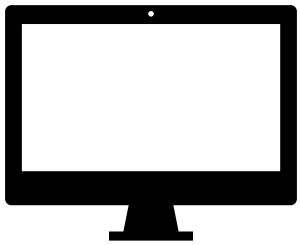
9. Challenges and Model Selection:
   - Discuss any challenges encountered during the project, such as data quality issues, feature selection, or model interpretability.
   - Explain the rationale behind choosing the final model based on its performance and suitability for the specific business needs.

Images:

To enhance your report, you can consider including the following images:

- Data distribution plots: visualize the distribution of key features for both persistent and non-persistent patients.
- Feature importance graphs: show how different features contribute to the model's predictions.
- Confusion matrix: illustrate the model's classification accuracy for both classes.
- ROC-AUC curve: demonstrate the model's ability to distinguish between persistent and non-persistent patients.

DATA

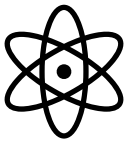


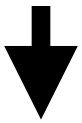
BUSINESS UNDERSTANDING.

DATA UNDERSTANDING

DATA PREPARATION

MODELING





EVALUATION

DEPLOYMENT