

Project Proposal



Menna Tarek - Salma Ahmad - Mahmoud Tag - Mahmoud Atef – OMAR ALI

GROUP CODE: ALX1\_AIS3\_G1e

**Predicting Doctor Prescription Behavior Using Machine Learning**

**Abstract**

This project aims to develop a machine learning model to predict whether a doctor will prescribe a recommended medication based on data from interactions between medical representatives and doctors. The dataset includes various features, such as doctor specialization, type of medication, interaction history, and prescription patterns. By analyzing this data, the project seeks to uncover the key factors that influence prescription decisions and develop an accurate predictive model that can assist in optimizing sales strategies for pharmaceutical companies.

**Introduction**

Problem Statement: Medical representatives often influence doctors' decisions to prescribe specific medications, but the factors contributing to these decisions are not well understood. A predictive model could help uncover the key elements that lead to prescription behavior, optimizing the efforts of medical representatives and improving healthcare outcomes.

Background: Previous research has shown that a variety of factors—such as doctor specialization, medication type, and clinic type—affect doctors’ prescription decisions. However, predicting these decisions with machine learning can provide deeper insights into this process.

**Objectives**

1. Build a machine learning model that predicts whether a doctor will prescribe a recommended medication.
2. Identify the key features influencing the prescription decision, such as doctor specialization, clinic type, and past interaction history.
3. Evaluate the performance of different machine learning models to find the most effective one for the prediction task.

**Methodology**

* **Data Collection:** The dataset used contains historical records of interactions between medical representatives and doctors. Key attributes include:
  + Doctor’s area of specialization (e.g., chest, surgery, general practice)
  + Type of medication (Type 1, 2,….,6)
  + Doctor’s Class (Class A (high patient rate) or B (medium patient rate) )
  + Clinic or hospital type
  + Prescription status (whether the doctor prescribed or not)
* **Data Preprocessing:** Data cleaning includes handling missing values, ensuring consistency in the records, and normalizing numerical features. Categorical variables will be encoded to prepare them for machine learning algorithms.
* **Feature Selection/Engineering:** Relevant features such as doctor specialization, interaction frequency, and medication type will be analyzed for their predictive power. Feature engineering will be applied to combine and create new insights from the data.
* Modeling: Multiple machine learning algorithms will be evaluated, including:
  + Logistic Regression
  + Decision Trees
  + Random Forests
  + Support Vector Machines (SVM)
* Evaluation: Models will be evaluated using performance metrics such as accuracy, precision, recall, and F1 score to determine the best-performing algorithm.

**Timeline**

**Week 1:**

* ***Data collection and preprocessing***: Clean the dataset, handle missing values, and encode categorical variables.
* ***Exploratory Data Analysis (EDA)***: Investigate patterns, distributions, and relationships in the dataset.

**Week 2:**

* ***Feature selection and engineering***: Identify the most relevant features and create new features if necessary.
* ***Initial model building***: Develop and train multiple machine learning models (e.g., Logistic Regression, Decision Trees).

**Week 3:**

* ***Model evaluation***: Test model performance using cross-validation and metrics like accuracy, precision, and recall.
* ***Model optimization***: Fine-tune the best-performing models by adjusting hyperparameters.

**Week 4:**

* ***Final model selection***: Choose the best model based on evaluation results.
* ***Prepare final report and presentation***: Summarize the findings, insights, and key factors influencing predictions.

**Expected Outcomes**

A machine learning model capable of accurately predicting whether a doctor will prescribe a medication based on a medical representative’s recommendation.

Identification of the key factors that influence prescription decisions.

Insights that can help pharmaceutical companies optimize their engagement strategies with doctors.

**Impact**

The model will allow pharmaceutical companies to target their efforts more effectively, potentially reducing costs and improving the quality of medical interactions. Additionally, the insights gained could assist in improving the efficiency of medical representative strategies, leading to better healthcare outcomes.

**Conclusion**

This project will provide a data-driven solution for predicting doctor prescription behavior, giving pharmaceutical companies and medical representatives valuable insights into how to optimize their interactions with healthcare providers.