Graduation Project Proposal

Medical Representative

Team Members

AbdelRahman AbdelHalem Helal Ahmed

Mazen Mohamed Elsayed Elsafty

Mazen Mahmoud Shaban Mohamed

Rashad Mohamed AboElmkaram Shehab

Mazen Mohamed Mohamed Ahmed

Karim Ahmed Farhat Eid

Supervisor

Eng Sherif Said

Track

IBM Data Science

Organization

Digital Egypt Pioneers

Abstract

Medical representatives serve as the crucial link between pharmaceutical and medical companies and healthcare professionals, responsible for promoting and selling products such as pharmaceutical drugs and medical equipment. Their primary challenge is persuading healthcare providers, such as doctors, nurses, and pharmacists, to choose their company's products over those of competitors, particularly when multiple companies offer drugs with the same active ingredients. This project aims to determine whether a doctor, based on certain features, will prescribe one of six drugs from a local pharmaceutical company that produces generic versions of multinational drugs. By analyzing specific patterns and features in prescribing behavior, the project seeks to enhance the effectiveness of medical representatives in promoting their company's products and navigating the highly competitive pharmaceutical market.

1 Introduction

Medical representatives play a vital role in the pharmaceutical industry, acting as the primary point of contact between pharmaceutical companies and healthcare professionals. Their main responsibility is to promote and sell products, including drugs and medical equipment, to doctors, nurses, and pharmacists. Through interactions with healthcare professionals, medical representatives work to increase product awareness, answer questions, and generate demand for their company's products.

1.1 Problem Statement

One of the primary challenges faced by medical representatives is the difficulty of convincing doctors to prescribe their company's drugs when several competing products with the same active ingredients are available. Despite the identical composition of these drugs, doctors may have different prescribing behaviors based on factors such as the representative's relationship-building efforts, product familiarity, or perceived trust in the brand. The core problem lies in understanding the specific factors that influence a doctor's decision to prescribe a drug from a particular company when multiple similar alternatives exist in the market.

1.2 Objectives of the Study

The objective of this project is to determine if a doctor, based on specific characteristics (such as their specialty, experience, and past prescribing patterns), will choose to prescribe one of six drugs produced by a local pharmaceutical company. This company produces generic drugs that contain the same active ingredients as those produced by multinational pharmaceutical companies. The project aims to analyze the key features of doctors' profiles and prescribing behaviors to predict whether they will opt for the local company's drug over competitors' products.

2 Graduation Project Organization

2.1. Roles and Responsibilities

The team will be responsible for conducting the research, data collection, analysis, and implementation of the project. This includes preparing the data, applying machine learning models, and interpreting results.

2.2 Project Phases

This project will be divided into several phases to ensure systematic completion. The phases include:

- **Phase 1: Data Collection and Exploration**
    - Collect the data from a SQL database file and organize data from the two tables (medicine_table and doctor_table).
    - Perform initial data exploration to understand the distribution of features such as specialty, price, area, and prescription behavior.
    - Calculate basic statistics like total number of doctors, the number of doctors prescribing the drugs, and percentages of doctors prescribing versus not prescribing.
- **Phase 2: Data Cleaning and Preprocessing**
    - Clean the dataset by handling missing values, duplicated data, correcting inconsistencies, and ensuring data integrity.

    - Use techniques such as imputation for missing values (e.g., mean, median, mode).
    - Transform the data into a suitable format for analysis, ensuring that categorical variables (e.g., specialty, area) are encoded correctly.

- **Phase 3: Data Analysis and Visualization**
    - Analyze the data to find hidden relations between the data (e.g., correlation).
    - Visualizing the data to see the patterns and relations between the data by creating statistical plots (e.g., bar charts, histograms, scatter plots).
- **Phase 4: Feature Engineering and Data Clustering**
    - Create new features that may enhance model performance by classifying the data into clusters using clustering methods (e.g., K-means).
    - Understand which features have high effect on doctors' prescriptions
- **Phase 5: Data Transformation and Preparation**
    - Transform the data into a suitable format for machine learning stage by normalizing the data using Sklearn library (e.g., MinMax Scaler, Standard Scaler).
    - Encoding categorical variables (e.g., specialty, area) to be in a numerical format that is suitable for model training and fitting.
    - Spilt the dataset into training set and testing set for model training and testing.
- **Phase 6: Model Selection and Implementation**
    - Apply various classification models to predict whether a doctor will prescribe one of the six drugs based on their features.
    - Evaluate models such as decision trees, SVC, and ensample methods to determine which model provides the best predictive accuracy.
    - Perform model validation using techniques such as cross-validation.
- **Phase 7: Model Evaluation and Optimization**

    - Evaluate the performance of the selected models using metrics like accuracy, precision, recall, and F1-score.
    - Tune model parameters to optimize predictive performance.
- **Phase 8: Final Report and Presentation**
    - Compile the results and analysis into a comprehensive report.
    - Prepare a presentation that outlines the project's goals, methodology, results, and conclusions.

2. Project Planning:

2.1 Tentative Timetable:

- Data Collection and Exploration: 1 day
- Data Cleaning and Preprocessing: 1 day
- Data Analysis and Feature Engineering: 2 days
- Data Transformation and Preparation: 1 day
- Model Selection: 1 day
- Model Evaluation and Optimization: 2 days
- Final Report and Presentation: 1 day

3. Resources:

- **Hardware**: A laptop or desktop computer with sufficient processing power for data analysis and machine learning model training.
- **Software**: Python (with libraries such as Pandas, NumPy, Scikit-learn), Jupyter Notebook for coding and documentation, and any additional machine learning tools for model building.
- **Data**: The dataset provided, which includes two connected tables (medicine_table and doctor_table) containing details about drugs, prices, and doctors.