Persistency of a drug

Name:Srilatha S

Email:ssiri0189@gmail.com

Country: USA

Problem Statement:

Pharmaceutical companies face the challenge of understanding how long patients continue to take their medications as prescribed by physicians, which is known as persistency. To overcome this challenge, ABC pharma has sought the help of an analytics company to develop an automated system that can identify patterns of persistency in the data they collect. By automating this process, they aim to gain insights that can help them improve patient outcomes and achieve better business results in a competitive industry. Essentially, the business understanding is that leveraging advanced analytics can provide valuable insights and advantages for pharmaceutical companies.

Business understanding:

The business understanding here is that pharmaceutical companies need to have a thorough understanding of patient medication adherence in order to improve patient outcomes and achieve success in a highly competitive industry. By partnering with an analytics company to develop an automated system to identify patterns of medication persistency, ABC pharma is seeking to gain valuable insights that can help them make informed decisions and drive better business results.

In addition, the use of advanced analytics can provide pharmaceutical companies with a competitive advantage by enabling them to identify and address issues with medication adherence more quickly and accurately. Ultimately, the goal is to improve patient outcomes, which can lead to increased sales and revenue for the company. Therefore, the business understanding is that leveraging data analytics is a crucial component of success in the pharmaceutical industry.

Project lifecycle:

The project lifecycle for the pharmaceutical project described would involve the following steps:

1. Project Initiation: In this phase, the goals, objectives, and scope of the project would be defined. The primary goal of this project would be to develop can identify patterns of persistency in the data collected by ABC pharma.
2. Planning: In this phase, the project plan would be developed. I would create a work breakdown structure (WBS) to break down the project into smaller, more manageable tasks. The project schedule would also be developed, identifying timelines and deadlines for each task. I would identify the risks and challenges that may arise during the project and create mitigation plans to address them.
3. Execution: In this phase, the project plan would be put into action. I would develop the automated system that can identify patterns of persistency in the data collected by ABC pharma. The system would be tested and validated to ensure that it meets the project requirements.
4. Monitoring and Control: In this phase, I would track the progress of the project, identify any deviations from the project plan, and take corrective action as required. I would also ensure that the project is on track to meet the project timeline and budget.
5. Closing: In this phase, the project would be formally closed. I would present the results to the stakeholders and obtain their approval. I would also conduct a post-project review to identify lessons learned and opportunities for improvement for future projects.

Overall, the project lifecycle would involve a structured approach to planning, executing, and monitoring the project, with a focus on delivering the desired results on time and within budget. The project would enable ABC pharma to gain valuable insights that can help improve patient outcomes and achieve better business results in a competitive industry by leveraging advanced analytics.

Data Understanding

Data Source:

The data source for this project is a dataset provided by ABC Pharma. The dataset contains patient demographics, provider attributes, clinical factors, disease/treatment factors, and adherence information. The dataset is stored in a excel format.

Data Size

The dataset contains a total of 3425 rows and 69 columns.

Data Types

The dataset contains both categorical and numerical data. The data types of the variables are as follows:

- Bucket:
- Unique Row Id: numerical
- Patient ID: categorical
- Target Variable:
- Persistency_Flag: categorical
- Demographics:
- Age: numerical
- Race: categorical
- Region: categorical
- Ethnicity: categorical
- Gender: categorical
- IDN Indicator: categorical
- Provider Attributes:
- NTM - Physician Specialty: categorical
- Clinical Factors:
- NTM - T-Score: numerical
- Change in T Score: categorical
- NTM - Risk Segment: categorical
- Change in Risk Segment: categorical
- NTM - Multiple Risk Factors: categorical
- NTM - Dexa Scan Frequency: numerical
- NTM - Dexa Scan Recency: categorical
- Dexa During Therapy: categorical
- NTM - Fragility Fracture Recency: categorical
- Fragility Fracture During Therapy: categorical

- NTM - Glucocorticoid Recency: categorical
- Glucocorticoid Usage During Therapy: categorical
- Disease/Treatment Factor:
- NTM - Injectable Experience: categorical
- NTM - Risk Factors: categorical
- NTM - Comorbidity: categorical
- NTM - Concomitancy: categorical
- Adherence:
- Adherence for the therapies: numerical

Data Quality

Upon preliminary inspection, the dataset contains missing values and inconsistent formatting. The missing values are mainly in the form of null values and NaN values. Some of the variables contain inconsistent formatting, such as the use of abbreviations and different spellings for the same category. These issues will need to be addressed during the data cleaning and feature engineering process.

Data Cleaning and Feature Engineering

The data cleaning and feature engineering process will involve the following steps:

- Handling missing values
- Handling inconsistent formatting
- Encoding categorical variables
- Scaling numerical variables
- Feature selection

Conclusion

The data intake report provides an overview of the dataset and outlines the steps that will be taken during the data cleaning and feature engineering process. By addressing the issues in the dataset and engineering relevant features, we aim to build an accurate classification model that can provide insights into the factors impacting persistency in the pharmaceutical industry.