Predicting the Likelihood of Certain Types of Negative Events Occurring with Medical Devices

Problem:
This capstone project aims to identify and study factors that significantly influence certain types of negative events that occur with medical devices. These negative event types are recall, safety alert, field safety notice, and a combination of two or more of them.

Target Clients:
The primary clients this project targets are those that work in or deal with the medical field, particularly those that manufacture or utilize medical devices as well as the patients that receive them. Medical devices are known to save, extend, or made better millions of lives. However, according to the International Consortium of Investigative Journalists (ICIJ), more than 1.7 million injuries and nearly 83,000 deaths are suspected of being linked to medical devices over 10 years and reported to the U.S. alone. Additionally, the ICIJ noted that the U.S. had more than 26,700 device recalls while India — with more than a billion people — had just 14 from 2013 to 2017. It is important to consider though that this could be due to the U.S. having more rigorous testing and mandatory regulations and the possibility that medical devices were utilized more in the U.S. than in India. Patients often are the last to be informed about malfunctioning devices.

Thus, this project can help medical practitioners make better decisions on which medical devices to use with the most minimization of risks as well as patients to be better informed on which medical devices they want to receive. Also, the results of this project can give indications to manufacturers on how to better design medical devices with reduction to injury and mortality.

Data:
The International Medical Devices Database will be used for analysis in this project. It is licensed under the Open Database License and its contents under Creative Commons Attribution-ShareAlike license. It is also available for download by the link: https://medicaldevices.icij.org/p/download.

Proposed Approach:
Multiple steps will be taken to create the best possible prediction model for this project.

1. Ten to fifteen most relevant predictor variables will be manually selected from the dataset using objective metrics. The three csv files consisting of the data will be merged and cleaned via programming languages like python, SQL, etc.
2. The merged and cleaned dataset will be explored visually via graphs created from programming languages.
3. To determine which predictor variables are significantly associated with the multicategorical outcome, Event Type, a categorical regression model, such as a multinomial logistic regression model, will be built.
4. Once the significant predictor variables are identified, a machine learning method, like a classification decision tree, will be trained to create the best prediction model possible.

Deliverables:
The final draft of the project will be presented via PowerPoint slides and delivered via Jupyter Notebook detailing each step taken and code written for the analysis of the project. A Github repository for the project will be created as well.