

Text Analytics to Inform Deviation Root Cause Analysis in Biomanufacturing



BUSINESS PROBLEM

Whenever a major deviation from a defined process, product, or system requirement occurs, steps are taken to determine potential causal factors, investigate to find the root cause, and determine appropriate actions to prevent future issues. Data from historical deviation investigations could provide valuable insights to improve the efficiency and accuracy of future investigations. However, the unstructured nature of the reports means that it is difficult to aggregate and use effectively. The business would like to determine a way to structure this data and draw meaningful insights.

DATA SOURCES

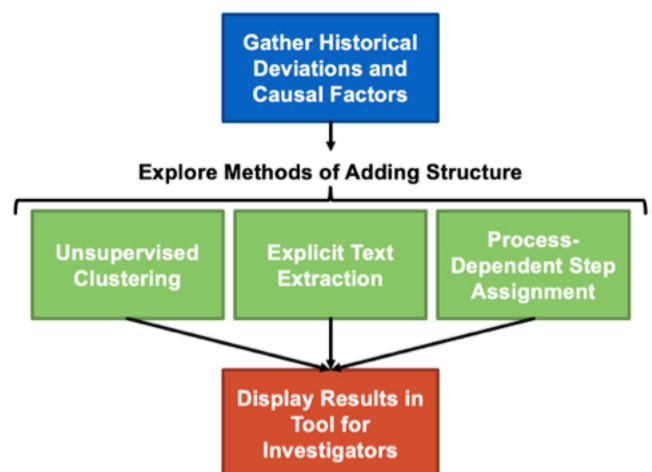
The main source of deviation data was Amgen's Quality Management System. This included structured metadata as well as free text fields describing the deviation. Additionally, root cause analysis reports from document management systems were sourced to access causal factors and investigation outcomes. Along with the core dataset, additional reference data was pulled in from supporting sources.

Data Types and Format

Unstructured text fields available in CSV files scraped from original PDF or Word Doc formats. Reference data stored in graph databases in Amgen's Enterprise Datalake.

APPROACH

To begin, we focused on the process of determining potential causal factors for new deviations as an area to be addressed. To enable this, we explored various analytical solutions to add structure to the unstructured dataset. These included unsupervised clustering, explicit text extraction, and process driven step assignment. The results of each method were combined in a report for investigators.



IMPACT

The output of the project is a variety of methods that could be used to add structure to deviations and causal factors as well as a proof-of-concept tool showcasing the results of the method. This kind of tool that centers on causal factors proposed in historical deviation has the potential to improve the efficiency and accuracy of future deviation investigations. If the methods we developed are fully implemented, investigators of a new deviation can use the tool to see information about similar deviations from the past including what causal factors were proposed, what kind of topics those causal factors centered on, the results of investigating those causal factors, and the final outcomes, all of which will inform their investigation. This can replace the current process where investigators might manually search for historical records and read through lengthy reports, if they seek out historical guidance at all. We hope that the solution will also help identify trends in deviations and root causes leading to durable fixes preventing future deviations from occurring.

DRIVERS

As a biopharma company, Amgen's processes are highly regulated. Technology and automation for quality assurance can be challenges in biomanufacturing where many steps are highly manual and process parameters necessarily exist within a range. Utilizing the deviation investigation reports was one of the only ways to holistically analyze Amgen's deviation management system.

BARRIERS

Working with unstructured and unlabeled data can be quite challenging. When working on a particular method, it was very difficult to determine whether or not the method had been successful since there was nothing to say what the "right" answer should be.

ENABLERS

Amgen was very open and supportive of this project throughout. The team did a great job of not only providing access to any data requested but also preparing most of the data beforehand so that I could hit the ground running on day one. Additionally, I am very grateful that the team was open to new methods and supportive of exploration even when some ideas did not work out.

ACTIONS



My project focused on exploring possible directions rather than implementing a final solution. Based on the outputs of several methods I explored, I developed a proof-of-concept tool that serves as a blueprint for the kind of tool that could be implemented at Amgen in the future.

INNOVATION

Some of the methods we used were developed specifically for this project such as those involved in explicit text extraction. Additionally, the idea of using documents describing a process to enhance the text classification is a novel approach. The implementation I worked on was unsuccessful, but there is promise in the idea and continues to be investigated by collaborators.

IMPROVEMENT

The potential improvements cannot easily be quantified. Rather, they provide the opportunity to improve the efficacy of deviation investigations. Hopefully, if these solutions are implemented, Amgen will see a reduction in the number of major deviations, and the time it takes to resolve deviations may decrease.

BEST PRACTICES

When working with unstructured free-text data describing deviation or non-conformance investigations, it is critical to understand your data and the system it describes. Think carefully about what kind of structure you can create that best fits with the data and the system to provide useful insights.

OTHER APPLICATIONS

Many of the methods explored can be applied to a variety of areas where the main source of data is unstructured text. In particular, the methods are not specific to biomanufacturing and could easily be applied in any manufacturing setting where quality assurance is critical and investigations of problems on the manufacturing line frequently occur.