

# 2024 Spring Session 7: Qualitative Analytics Capabilities Using Natural Language Processing (NLP)

**Date:** Thursday, 5/16 2:00-3:00 pm ET

**Speaker:** Nikhil Nirhale

1. **Q- How generalizable would this model be from one rule to the next?**

A- It depends. The context changes whenever a new docket comes into picture that is something we have observed for some classification types of tasks. If you are referring to fast models, it would be generalizable. If we are talking about identifying recommendations, recommendations have a structure to it that do things in a certain way. There is a certain language at least that is used as part of the comments to say that something is a recommendation or agreement so that is where the context does not help as much. You always look at the model and can understand the structure of the statement to decide about whether it is an agreement or a recommendation. Context incorporating additional context can help with improvement in the model. If we see a need for it, that something will incorporate. That is an open possibility for us, and we have not ruled that out. But these models understand the structure of language and that is where we have seen decent performances.

2. **Q- What is FAS? I apologize for the question if you have explained it already.**

A- Feedback Analysis System. A use case for machine learning (ML) being utilized to analyze comments submitted to proposals.

3. **Q- You mentioned about a nursing home business use case, where recommendations were provided around meeting staffing standards. Can you elaborate a bit on the problem statement/ business requirement as well, which specific problems you were trying to solve regarding nursing homes?**

A- The recommendations provided regarding meeting staffing standards fall under the Feedback Analysis System (FAS). FAS is a fully developed application that enables efficient review of proposed policy feedback. A proposed policy related to minimum staffing standards was analyzed with the help of FAS in the latter half of 2023. This was an interesting overlap between the FAS and 2567 nursing home use cases that are part of Artificial Intelligence for Machine Learned Modeling (AIMM).

AIMM has a 2567 nursing home use cases where we are building an application to review 2567 deficiencies for Quality, Safety & Oversight Group (QSOG)/Survey & Operations Group (SOG) within CMS that will include features to navigate 600,000+ deficiencies collected as part of the survey process. The application will be powered by ML/natural language processing (NLP) models that can identify deficiencies that are under/over cited for severity levels, identifying key elements within a deficiency, and generating summaries for 2567 surveys using the identified key elements.

4. **Q- How do you determine the right performance metrics for an Artificial Intelligence (AI)/ML solution?**

A- The importance of your metrics depends on the stage of your project. If you are in the initial stages, your focus should be on collecting the right data and appropriate test data. At this point, having rigid metrics does not make sense until you have a decent amount of testing and training data. Once you have collected sufficient test data, you can set goals such as achieving 95 percent accuracy. However, it is crucial that your test data is well-represented and free of bias. It should not favor certain features or concepts in the training dataset.

After addressing these considerations, you can set a realistic benchmark. High accuracy with poorly balanced test data is less valuable than slightly lower accuracy with well-balanced data, which provides more meaningful insights. This approach is essential in the initial stages. As you continue to build your feedback systems and add components to your model, you will gather more data and gain a better understanding of your use case. This process will help you refine your metrics and determine what kind of data should be included in your test dataset. As your solution improves and your understanding of the business processes deepens, your metrics will naturally evolve.

5. **Q- How do you go about the process of choosing a model to solve a particular use case?**

A- It depends on what your business process is in terms of that use case and that is what helps you.

6. **Q- How do you generalize existing workflows for new use cases?**

A- It is critical to determine what a complete workflow encompasses. A ML workflow typically consists of the following steps - Data Prep, Model Training, Model Deployment, Model Monitoring, Orchestration mechanism.

All the steps are reusable for a new use case. The various libraries, tools, services, and technologies can be applied to a new use case with minor modifications that are required due to demands of a new use case. Although a trained model might not be directly used for a new use case due to the data it has been trained on could be very different from the data for a new use case.

Hence an evaluation process is recommended for a new case to determine the reusable components from a previous workflow. The evaluation can be conducted on specific areas for a ML workflow (Data Prep, Model Training, Model Deployment, Model Monitoring, Orchestration mechanism).

7. **Q- How do you ensure your AI/ML solutions are free of hallucination, inaccuracies, errors, and/or bias?**

A- The first thing to understand is that hallucinations are inevitable. To address inaccuracies, a test-driven approach and using different metrics can be helpful. Numerous open-source repositories are available that integrate with your models to show metrics related to hallucinations or inaccuracies in responses. These tools can provide the right metrics before deploying your large language models. While there are many tools

available for handling inaccuracies or errors, qualitative analysis is also valuable.

For addressing bias, you can use explainability tools like Clarify. If you detect a specific statement, you can run it through a Clarify explainability job, which will highlight the words or areas the model focuses on to make a prediction. This helps you understand how the model interprets data and how it processes new data.

Reducing bias and inaccuracies can be achieved by improving training processes, conducting thorough validation, and using proper validation techniques. In machine learning, data is the most critical component. Investing time in acquiring and refining data can prevent many of these problems from occurring.

8. **Q- Can you explain the details of the integration of SageMaker with Databricks? Does it work with Python only or does it also work with R and Spark?**

A- SageMaker is not integrated and available with Databricks. Currently, SageMaker is not a tool/application that is available to the CCSQ community as it resides within AWS and most non-ADO contracts/groups do not have access to AWS. As of right now, we are not aware of SageMaker becoming available nor becoming integrated with Databricks. There are other features that Databricks offers around AI/ML tooling that will eventually be opened to the CCSQ community, but we must get through some prerequisite work first before we can get to that point.

9. **Q- What type of framework governs your AI/ML solution?**

A- A framework that governs an AI/ML solution will ensure transparency on the criterion for choice of models and performance benchmarks, provides visibility into the training data and experiments through detailed documentation, integrates tools and technologies that can enable model explain ability, bias, and data drift detection. The framework should also incorporate a feedback mechanism to ensure you are collecting feedback on your model predictions to ensure up to date performance of your models. A robust ML ops process to track model experiments, versioning and enable effortless deployments into a production environment is a critical requirement.

The framework that is mentioned above is the ideal objective. The AIMM team is currently working on all the individual areas of this framework.