Designing an Alert Correlation Engine using Mutual Information values

# Project Progress – Dated: Nov 14th, 2021

Below is how we are progressing towards our tasks in the project.

1. Which tasks have been completed?

|  |  |  |
| --- | --- | --- |
| **Tasks** | **%Complete** | **Notes** |
| Implement scoring function | 100% | We have finalized our scoring function using mutual information and it has been coded for consumption for other modules. Code is uploaded [here](https://github.com/swatinanda/CS410-Fall2021-TeamAlpha/tree/main/SourceCode/alarm_correlation_mutual_information).  High level notes on how do we calculate the mutual information is presented [here](https://github.com/swatinanda/CS410-Fall2021-TeamAlpha/blob/main/SourceCode/alarm_correlation_mutual_information/how_to_calculate_mi.png) |

1. Which tasks are pending?

|  |  |  |
| --- | --- | --- |
| **Tasks** | **%Complete** | **Notes** |
| Data cleanup / preparation  (Dataset) | 80% | We are almost wrapping up the dataset cleanup and preparation. It is being built with real alert monitoring scenarios based on App, Infra and Network monitoring. Some examples are   * High Average response time, * High CPU usage on the Host, * Router not reachable, * Kafka Lag breached a configured threshold etc.   Our utility will replay the dataset message to generate real time alert template data. Currently scoring function is built based on alarm Id but populating the alarms with additional information will help us to enrich our “Knowledge Graph” |
| Create a bag of alerts | 80% | We have a corpus of Alert messages, Hostname and Source. This will be used to generate random alerts. Each alarm will have this 3 metadata information selected randomly from the corpus. |
| Build Knowledge graph with limited datasets | 10% | <Neo4j **Sanjeev**> Plan to wrap it up by Nov 18th |
| Develop REST API for frontend interaction  Implement user interface | <Swati> | Framework to train data for chatbot in progress.  Will club the two tasks to present data into the chatbot.  Plan to wrap it up by Nov 18th |
| Implement Feedback loop |  | Need to evaluate if we’ll be able to accomplish it optimally. |
| Integration, Testing & Evaluation |  | Data for evaluation is being created as part of Dataset preparation. We will add some rules to baseline this data, against which we can evaluate/verify our algorithm output.  Will start the Integration and Testing post that (tentatively by Nov 24th or 25th) |

1. Are you facing any challenges?

We had to do our dataset from scratch, so that is taking some time.