

Research internship project proposal

Satyanarayan Nayak, Swastik
s.nayak.1@student.rug.nl
S4151968

March 31, 2020

1 Project

Technologies used: Machine learning, ELK stack, Magenta.

The proposed project is a machine learning research-oriented task, that is focused to provide the company Magement the **ability to automatically predict anomalies and classify them**. An anomaly is the deviation of their system from its normal working condition like network slow down, exceptions, out of memory, large input file and so on.

Figure 1 shows a high-level design of this project, the section marked with (1) is the existing system, and (2) is the functionalities that will be implemented as a part of this research internship. The *ELK stack* stores the information of the data that is being transferred between server A and server B, the proposed system will receive data from the ELK stack at regular intervals and will apply *Machine learning* techniques to predict anomalies and classify the anomalies based on their classes. The output is written back to the ELK stack and will be undertaken by another module that will scale or tweak Magement's system accordingly to suppress or resolve the anomaly.

Training data is generated through the help of ELK stack from the logs accumulated from the historical data transfers. Magement can generate any realistic metric, this will aid in the process of feature selection.

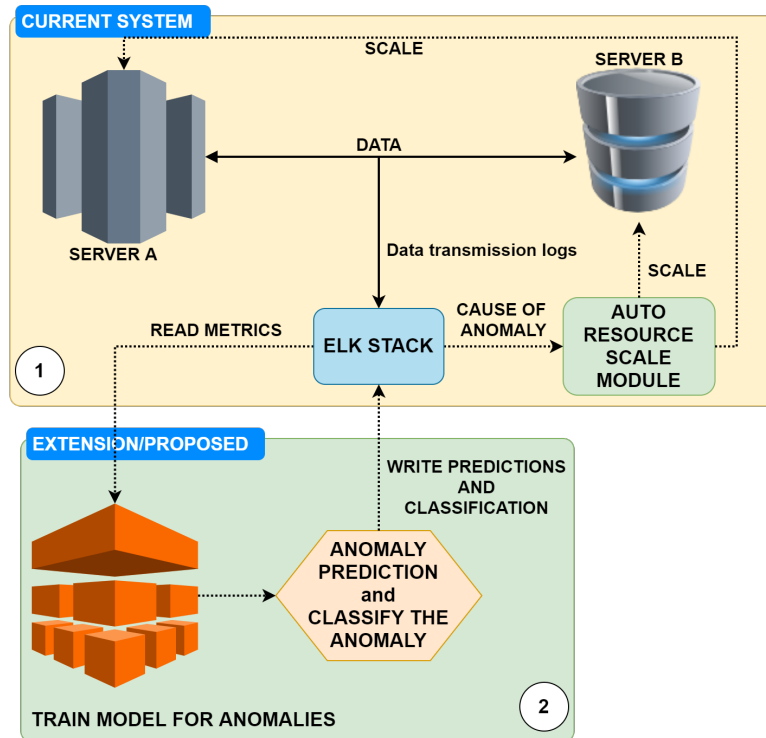


Figure 1: Architectural diagram of the proposed project.

2 Research Questions

- Feature selection: The training data contains many features and this demands for a crucial step in determining only the feature that contribute most towards the anomaly predictions.
- Prediction: Need to determine and implement the correct prediction algorithm based on the requirement.
- Classification: Need to determine and implement the correct classification algorithm based on the requirement.
- Testing to validate the prediction and classification capabilities of the machine learning model.