INSTITUTIONEN FÖR DATAVETENSKAP | LUNDS TEKNISKA HÖGSKOLA | PRESENTERAD 2022-08-25

EXAMENSARBETE Identification of relevant error descriptions in build logs using machine learning
STUDENTER Lykke Axlin, Klara Broman
HANDLEDARE Martin Höst (LTH), Marcus Klang (LTH), Gustaf Lundh (Axis), Ola Söder (Axis)
EXAMINATOR Elin Anna Topp (LTH)

Utilizing artificial intelligence to help troubleshoot failed software builds

POPULÄRVETENSKAPLIG SAMMANFATTNING **Lykke Axlin, Klara Broman**

When software builds fail, software engineers often need to manually go through large unstructured log files to troubleshoot the failure cause. In our thesis, we investigate the possibility of using machine learning to automatically identify the relevant error descriptions in the log files.

Have you ever wondered how certain emails end up in your spam folder? The technique used to determine whether an email is spam or not is called text classification. In this study, we used text classification of rows in log files to help engineers find relevant error descriptions.

Many companies use the open source automation server Jenkins to facilitate continuous integration of new code in large code bases. When a build is run on Jenkins, a build log is created which contains information about the build, e.g. whether a test succeeded or not. The log files can sometimes become very large, and going through log files in order to troubleshoot why a build failed is a tedious task for an engineer. At Axis Communications, an open source plugin called the *Build Failure Analyzer* (BFA) is used to assist the engineers with this task. The BFA scans failed build logs for certain patterns in order to identify relevant error descriptions. These patterns are handcrafted by the engineers and new patterns need to be added continuously as new types of errors appear in the builds.

In our Master's thesis we investigated the possibility of using a machine learning model to automatically identify relevant error descriptions in the Jenkins build logs. The overall approach was to train a model on a dataset annotated with the findings from the BFA, and then evaluate if the model could detect error descriptions the BFA had not been able to detect. In the cases when the BFA does not find any relevant error descriptions in the build log, it would be very helpful if the AI-model could point the engineer in the right direction.

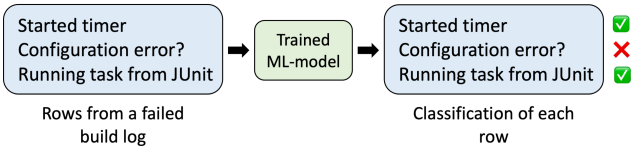

Figure 1: Prediction by a trained model on three build log lines

The results show that it is possible to train an AI-model to predict new relevant error descriptions in failed log files. However, the model predicts many more lines as relevant error descriptions than what is present in the file. This means that the engineer still needs to look through many log lines manually to find the relevant error descriptions. Therefore, further improvements to the model, as well as the dataset, are needed in order for the model to be useful in the industry.