Cookr Hackathon

Problem statement 3 : Predictive maintenance for Infrastructure

Team members : Dhikshitha R (21PD26)

Sharmila J (21PD33)

Introduction

Technologies such as artificial intelligence and machine learning and Internet of Things(IoT) has revolutionised the way industry manages their assets and operations. One such field is predictive maintenance. This report explores the usage of AI models in predictive maintenance.

Problem Statement

Create AI models or research necessary steps to create AI models that can predict potential failures or issues in infrastructure components based on historical data.

Approach

Infrastructure of a software company consists of many components, but the two most vital components are servers and network. A server is a program or a device that provides services to its users. A network is an interconnection of computers that can exchange information with each other.

Predictive maintenance refers to the process of using data analytics to identify operational anomalies and potential equipment defects, enabling timely repairs before failures occur. The most important step in predictive maintenance is to predict anomalies in the system. This can be achieved using a mix of machine learning and statistics.

Solution Description

As a solution to this problem, two AI models to predict the failures of both servers and network has been proposed.

System log files are historical records to record events of a computer. While syslog manages the log file structure, the content is predominantly generated by the applications themselves. Each log message comprises six fields: host (representing the sending machine's IP), facility (source of the message), level (severity), tag (a combination of facility and level for sorting), time (when the message was recorded), and the message itself (describing the event).

Server failure prediction can be done using physical metrics such as CPU utilisation percentage, RAM utilisation percentage and hard disk utilisation percentage, which can be calculated or extracted from the server logs. This is modelled as a binary classification problem in which the classes are: the server will not fail and the server will fail.

This data can be used as input to multiple machine learning models such as SVM (Support vector machine), naive Bayes, or random forest classification, which can be used to classify the given data. According to the research paper, “Predicting Computer System failures using SVM” by Fulp et all, the accuracy of predicting failure before 60 hours is 73% using SVM model.

Another approach is to use deep learning models to predict failure.The data obtained from syslog is then given as input to a 2 layer Long Short term Memory (LSTM) model which is used to predict the failure probability of the system. The last layer output is then passed to a logistic regression function which is used to classify whether server may fail or not.

The second machine learning model is designed to detect network equipment failure using various metrics.The dataset consists of datetime objects representing dates, node numbers, and various risk factors associated with network components, including manufacture discontinuation, power supply, CPU,

memory, location, misconfiguration, average link, card, and port risks, each ranging from 0 to 10. Additionally, there are failure categories with integer values from 0 to 9 denoting different failure types such as CPU, memory, card, port, power supply, location, OS upgrade, misconfiguration, and link failures. The dataset comprises three different network sizes with corresponding numbers of data points and memory usage per data frame: small (100 nodes), medium (200 nodes), and large (500 nodes), with memory usage ranging from 136.6 MB to 683.0 MB.

Machine learning models can be used on this data to predict failure.

References

1) https://ir.lib.uwo.ca/cgi/viewcontent.cgi?article=12104&context=etd

2) https://www.usenix.org/legacy/event/wasl08/tech/full_papers/fulp/fulp.pdf

3) https://towardsdatascience.com/system-failure-prediction-using-log-analysis-8eab84d56d1