

Introduction

In Ericsson, many test runs have been executed for testing the software packages in the simulation environment. Before launching product to customer, they need to test how well do the software package perform. The performance of these software packages are evaluated by considering on the CPU utilization, a percent of the CPU's cycle that spent on each process, and some other metrics (e.g., memory usage, latency). CPU utilization is one of the main focus in the test run that need to be optimized for the released of upgrade software package. They are visualized over a period of time.

Nowadays, whenever they need to inspect on the performance of new released software package, they have to visualize by themselves. With the rise of data generation and a large number of test runs, this work becomes more difficult and perhaps inefficient to do it manually. This is when Machine learning, the algorithm that has an ability to learn from data, comes into focus. The algorithm will help indicate whether the performance of software package is degradation, improvement or steady. There is also a case when there is no change in the software package but the changes in the test environment affect performance.

Background

Ericsson

Ericsson founded by Lars Magnus Ericsson since 1876 is one of the world's leading in telecommunication industry. It provides services, software and information and communications technology (ICT). Its head quater is located in Stockholm, Sweden. Nowadays, Ericsson starts to expand its service and product beyond telecoms sector.

LTE/4G

Objective

The main objective for this thesis is to implement machine learning to analyze the performance trend.

- Detect the degradation, improvement or steady state in CPU Utilization
- Detect whether there is some changes in test environment that impact on CPU Utilization

R vs Python

Both R and Python are powerful programming language for data analysis. Python has been known as a general purpose language with an easy to understand syntax.

R is developed by and for statistician. It provides huge number of essential packages in statistics and started to expand to different fields. R has a great reputation for data visualization.

In term of computation, R cannot compete with Python which is .

With strength and weakness between R and Python described earlier, R is chosen to be used in this study. It offers more algorithm to implement for the problem. Moreover, as and provides a great feature in visualizing data interactive graph

ecp package