# 1. Introduction

## 1.1. Background

Change point analysis is the method to detect changes in time series data when the time instant is unknown (Basseville et al., 1993) and has a bottom line for discovering and estimating the time point where the changes in time series occur. Serveral terms such as breakpoint or turning point is also used to denote the respective events. However, the commonly used term when the point in time series takes place is *change-point* and the term *regime switch* is for referring to the different of regime after the occurence of change point (Weskamp and Hochstotter, 2010). It has been studied over decades as it is a problem of interest in many applications in which the characteristics of data is collected over time. Change point analysis can be utilized in medical condition monitoring (Staudacher et al., 2005), to evaluate the sleep of patients based on their hearth rate variability. It can be found in crimate analysis (Reeves et al., 2007; Beaulieu et al., 2012), where temperature or climate variations is detected. This has been arised since the greenhouse gas emission rapidly increases for several decades and the effects from global warming (Aminikhanghahi and Cook, 2016). Applications are extensively implemented in the field of quality control (Page, 1954)in which the industrial process is monitored over a continuous period and should be able to identify when something went wrong (i.e., defective products, machine breakdowns). Methods are also useful in signal processing to detect the significant change in the streaming data (Basseville et al., 1993), identify fraud as soon as it is commited when fraud prevention is failed (Bolton and Hand, 2002), and detect anomalies in the market price (Gu et al., 2013). The change should be flagged when it occured in order to properly deal with such changes in time and reduce the consequences that could happen (Sharkey and Killick, 2014).

In this study, change point analysis will be used to identify the changes in performance of Ericsson's software product. 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 does 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 performance metrics (e.g., memory usage, latency).

Recently, method which is now becoming more popular for addressing the change in time series is a hidden Markov model. It used a concept of Markov process where the system treats data as observations and tries to model an underlying segmentation

as states. Hence, it is able to identify the switch in hidden states when change-point is most likely to occur (Luong et al., 2012). A hidden Markov model is widely used in almost all current system in speech recognition (Rabiner, 1989) and found to be important in climatology such as describing the state in the wind speed time series (Ailliot and Monbet, 2012) and in biology (Stanke and Waack, 2003) where the prediction of gene is being made. It has been extensively applied in the field of economics and finance and have a large literature about it. For instance, business cycle can be seen as hidden states with seasonal changes and modeled as a switching process to uncover the state over expansions and recessions and also the duration of each period (Hamilton, 1994). Financial time series is modeled in order to investigate how stock markets are doing in general i.e., bull or bear markets (Kim et al., 1998). Markov regime switching model is what econometrician and people who study related to this field always address when refering to hidden Markov model.

The term Markov regime switching model will be used throughout the thesis. Markov regime switching model is one of the most well-known non linear time series models. It takes the behavior of shifting regime in time series into account and models multiples structures that can explain this characteristic in different states at different time. The shift between state or regime comes from the switching mechanism which is assumed to follow an unobserved Markov chain. Thus, the model is able to capture complex dynamic patterns, identify the change of location and regime switch in time series. The historical behavior of the observation sequence in this study is not completely independent of each other (i.e., performance of current software package depends on the performance evaluated from the past version of software package). Therefore, additional dependencies at observation level with the first order autoregression is taken into consideration when modeling the Markov regime swiching model.

## 1.2. Objective

The main idea of this thesis is to reduce the laborer work to do visual inspection whenever they want to analyze the performance of update software package. With the rise of data generation coming from a large number of test runs, this work becomes more difficult and perhaps inefficient to do it manually. The main objectives for this thesis is to implement machine learning, the algorithm that has an ability to learn from data, to analyze the performance of the system. The algorithm will help indicate whether the performance of software package is in a degradation, improvement or steady state. There is also a case when the changes in the test environment affect performance even though there is no change in the software package. The implemented algorithm should be able to also detect when the test environment is altered. CPU utilization is the performance metric that will be looked upon in this thesis as it is one of the main focus in the test run that need to be optimized for the released of upgrade software package.

To summarize, this thesis aims to:

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

The thesis is structured as follows: Chapter 2 provides detail and description of datasets used in the analysis. Chapter 3 presents methodology. Results from the analysis is shown in Chapter 4 along with tables and plots. Chapter 5 discusses about the outcome and the obtained results, and conclusion can be found in Chapter 6.