2022 PRACTICUM

CHANGE POINT DETECTION OF TIME SERIES DATA

FOR COMNET MOBILE DATA PING TIME

BY

ALEX WALTON

(GT-ID: 903398136)

OMSA PRACTICUM

# Abstract

The idea behind this thesis is to detect significant change points in ping response time from cell towers around the US. The approach taken needs to be able to deal with significantly noisy data, and be able to distinguish between noise and real underling changes in the time-series data. A range of methods are explored.

The data is from 300+ cell towers, with 4 variables. In this project, identifying that there is a change in mean, and that when there is a delay in ping time, we prioritise voice data over data at all times. The algorithm not only has to be accurate in detecting change points within the time series data, but also has a time constraint component. The data is pulled every 10 minutes from 2 servers, therefore we need this algorithm to be able to scan all the data within the time frame, as well as be able to show results, to allow engineers to highlight issues in reasonable time.

# Table of Contents

[Abstract 1](#_Toc107150654)

[Table of Contents 2](#_Toc107150655)

[1. Introduction 3](#_Toc107150656)

[2. Methods 3](#_Toc107150657)

[2.1. Method 1 3](#_Toc107150658)

[2.2. Method 2 3](#_Toc107150659)

[References 4](#_Toc107150660)

# Introduction

For this project, data from a plethora of mobile sites around the US, each with a timeseries lookback of ping times (in Ms) for 4 variables (EFlarge, EFsmall, BElarge, BEsmall). Each of these variables represents a type of data packet, and the response time of each of these packets of data, for every sample interval, is represented on a time series graph as shown below (Fig1.1)

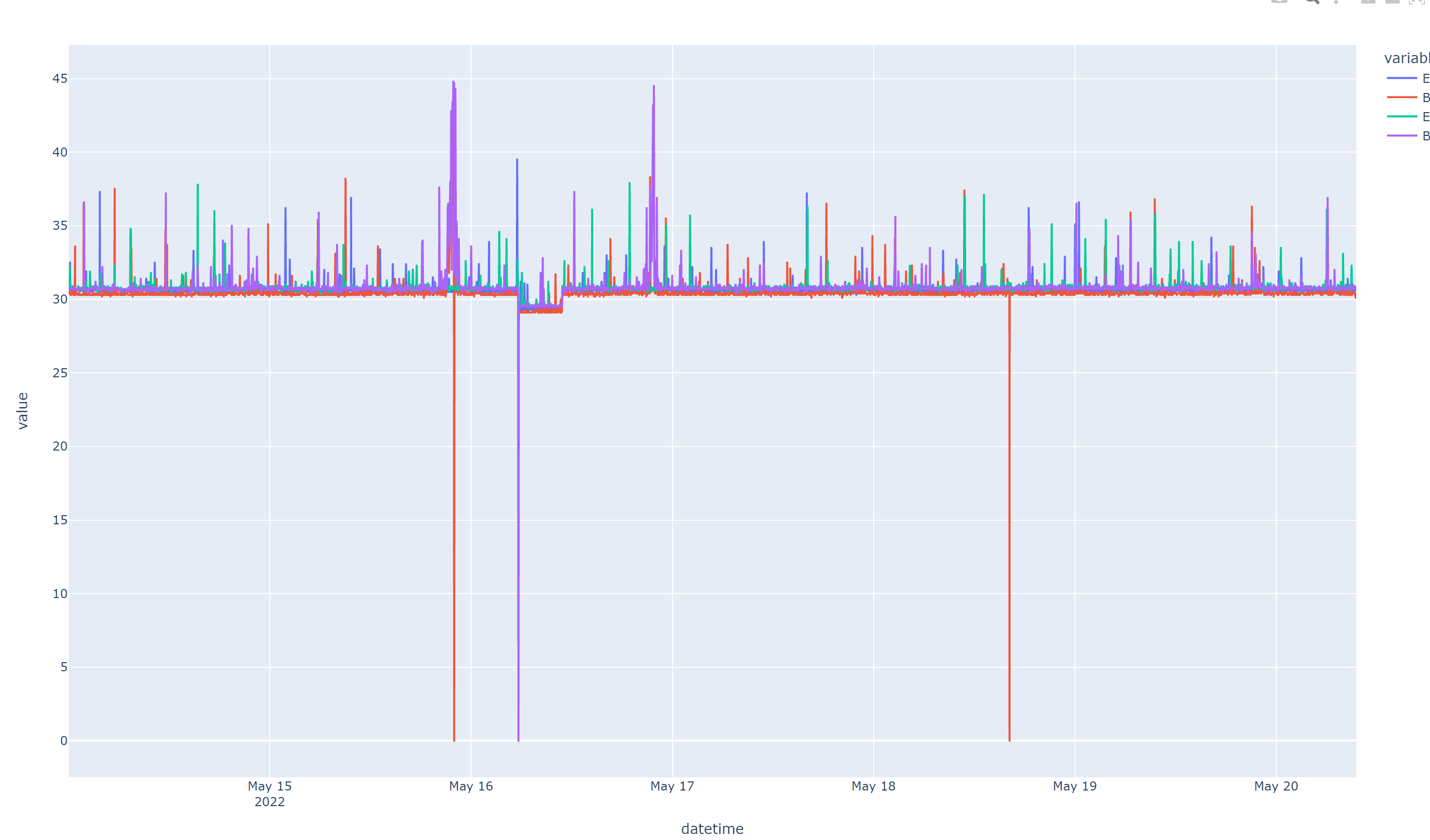


Figure .

The aim of the project is to correctly identify when a statistically significant change in the mean of the time series for each variable occurs. Alongside this, we want to ensure that when there is lag in the time series data, always we need to make sure that voice data is prioritised over data packets. The reasoning is simple: When on a mobile call, we can’t afford data to drop as the voice quality will sounds “computerised” and may not even get through at all. On the other hand, most mobile/internet devices have some form of local buffer, and therefore there is some tolerance if data is buffering/ delayed.

Changepoint Detection is a well-researched area of time series analysis, for example, Truong et al investigate different offline change detection methods for multivariate time series (Charles Truong, 2020). In our particular use case, we can investigate offline changepoint detection methods as we don’t need real time monitoring. The different between offline and online change points is essentially as follows. Offline changepoint detects changes when all samples are received, whereas online changepoint attempts to detect changepoints in real time settings. If we can spot a real and significant change in the time series, as well as highlight to relevant areas in-between data being redownloaded, this is perfectly acceptable.

Now we know that we can use offline change point detection, the next question is how do we approach detecting significant changes in a time-series? There are many approaches to address this question, simple CUMSUM can be used to detect changes in data, however in this instance we attempt to reduce the number of false positives that are flagged to the user.

Of note, the data supplied is unlabelled. This provided a challenge in itself, as to how to measure the accuracy of the model without manually analysing all plots to see if algorithm tested was detecting the correct changepoints.

For this project, attempting change point detection in the given data, a number of more advanced detection algorithms were used, and evaluated.

# Problem Statement

There are many difference change point detection algorithms available, via open-source python libraries. Some of these libraries offer a range of different detection methods for changepoint detection in time series data, while others offer similar algorithms wrapped in different libraries. One issue initially having to face was the fact that the data was unlabelled. This meant that either the data could be hand labelled, or a method that didn’t require data to be labelled should be used. For this project, and given the fact that the data was only a small window of the timeseries data generated, I opted for an unlabelled approach, that would then be assessed graphically. This saved time initially in that labelling data was a manual process and was too time consuming, however, it meant that algorithms suitable for unsupervised learning had to be used. For example, one initially looked at was a library called sktime (Alan Turing Institure) Many were tried and tested to assess their usefulness in detecting changepoints in the data highlighted above, until one was settled on and tested further. These will be outline below

# Methods

test

# PELT

# Binseg

# Window

(Sales Force)

# References

Alan Turing Institure, n.d. *sktime.* [Online]   
Available at: https://github.com/alan-turing-institute/sktime  
[Accessed 24 06 2022].

Charles Truong, L. O. N. V., 2020. *Selective review of offline change point detection methods.* [Online]   
Available at: https://arxiv.org/abs/1801.00718  
[Accessed 05 06 2022].

NA, n.d. [Online]   
Available at: https://paperswithcode.com/task/change-point-detection/codeless  
[Accessed 10 06 2022].

S. F., n.d. [Online]   
Available at: https://github.com/salesforce/Merlion  
[Accessed 01 05 2022].