ANOMALY DETECTION IN TIME SERIES DATA

N G Dulanjalee S Niralgama

199542 R

Degree of Master of Financial Mathematics

Department of Mathematics

University of Moratuwa

Sri Lanka

May 2024

ANOMALY DETECTION IN TIME SERIES DATA

Niralgama Gamaethige Dulanjalee Samanthika Niralgama

199542 R

Thesis/Dissertation submitted in partial fulfillment of the requirements for the

Degree of Master of Science in Financial Mathematics

Department of Mathematics

University of Moratuwa

Sri Lanka

May 2024

Declaration

I declare that this is my own work and this thesis/dissertation does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my thesis/dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature: Date:

The above candidate has carried out research for the Masters Dissertation under my supervision.

Name of the supervisor:

Signature of the supervisor: Date :

Abstract

Fraudulent activities have been affecting the world's day-to-day business, resulting losses over millions to its users. Due to the advancement of technology, fraudsters have many platforms to stay undetected and fraud detection has become critical for organizations and stakeholders such as bankers, investors, credit card users, financial auditors and insurance agents. In addition to detecting fraud, anomaly detections such as diagnosing certain sicknesses, identifying climate change disasters, early warnings on floods and volcano activations are important for many stakeholders. Time series data, in any field can provide significant information when it comes to detecting anomalous activities. Bayesian network, neural network, logistic method, K-nearest neighbor and regression are the most popular data mining approaches in anomaly detection that are being introduced so far. Finding anomalies among time series data is the key approach of detecting fraudulent activities and identifying striking real-world scenarios such as economic recessions. The existing methods and approaches have limitations such as manual thresholds to determine anomaly points, lack of theoretical foundation and lack of attention to class imbalance problem and multivariate data. Approaches such as interquartile rang (IQR) methods assumes that time series data are normally distributed. This assumption does not comply with practical data as they could be positively or negatively skewed. Therefore, to overcome these limitations, this research project introduced a framework to detect anomalies in both univariate and multivariate time series data using extreme value theory. Results of this framework demonstrates that this method's succeeds identifying the anomalies better than the methods implemented in the two R software packages, anomalize and timetk for detecting anomalies in univariate timeseries data. Further, this framework visualizes the identified anomalies which can be utilized by all stakeholders in for both univariate and multivariate anomalies.

Keywords: anomaly detection, fraud detection, univariate time series, multivariate time series, extreme value theory, data visualization.

Acknowledgement

I want to express my profound gratitude Dr Priyanga D Thalagala, who served as my supervisor and helped me write this thesis. Her guidance, patience, insightful feedback, continuous checks on the progress of my thesis and support to fill the gaps in my knowledge is invaluable. Even when I faced challenges, she provided unwavering support, helping me navigate the complexities of my research.

Many thanks to my husband, who has helped me throughout this research work, for his continuous inspiration, encouragement and for patience, late-night pep talks, and belief in my abilities kept me going when the research felt overwhelming. Without his constant dedication and assistance in every single moment, this thesis would not have been possible.

And most importantly, heartiest thanks to my parents for their unwavering encouragement and belief in me, which fueled my determination. From my childhood dreams to this very moment, they have been my unwavering cheerleaders. This achievement is as much theirs as it is mine.

Table of Contents

Table of Figures