LOGLEARN: PREDICTING COMPUTER NODE FAILURES USING CONTINUOUS MACHINE LEARNING

Kumararatnam Kabilesh

219347D

Master of Science in Computer Science

Department of Computer Science
Faculty of Engineering

University of Moratuwa
Sri Lanka

July 2023

LOGLEARN: PREDICTING COMPUTER NODE FAILURES USING CONTINUOUS MACHINE LEARNING

Kumararatnam Kabilesh

219347D

Thesis submitted in partial fulfillment of the requirements for the degree
Master of Science in Computer Science

Department of Computer Science
Faculty of Engineering

University of Moratuwa
Sri Lanka

July 2023

DECLARATION

I declare that this is my own work and this Thesis 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. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature: Date: 29-07-2023

The supervisor should certify the Thesis with the following declaration.

The above candidate has carried out research for the Master of Science in Computer Science Thesis under my supervision. I confirm that the declaration made above by the student is true and correct.

Name of Supervisor: Prof. Indika Perera

Signature of the Supervisor: Date:

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to everyone who contributed to the completion of this thesis.

First and foremost, I would like to thank my supervisors, Dr. Gayashan Amarasinghe and Prof. Indika Perera, for their guidance, patience, and support throughout my research. Their invaluable advice and feedback were instrumental in shaping my research and improving my writing skills.

I would thank all the members of the faculty and staff of the Department of Computer Science, Faculty of Engineering, University of Moratuwa for providing a supportive and stimulating academic environment that allowed me to pursue my research interests.

Finally, I would like to express my heartfelt gratitude to my friends and family for their unwavering encouragement and support throughout my academic journey. Their words of encouragement and motivation kept me going during the challenging times, and I am forever indebted to them.

ABSTRACT

Ensuring reliability, availability, and fault-tolerance is crucial in modern computer systems. Despite the substantial efforts put into the development, testing, and operation, failures still occur during runtime, leading to significant consequences. To address this issue, a proactive approach is necessary to predict and prevent failures before they happen. System and software logs provide essential data for monitoring systems and their performance during runtime. However, processing this information in real-time poses a unique challenge for machine learning because of the properties of streaming big data such as logs.

Therefore, this study utilizes the continuous machine learning paradigm to develop a failure prediction model called LogLearn, which uses system log data. The design and development of LogLearn consider the drawbacks and limitations of current continuous machine learning models to provide a more efficient and accurate approach to predicting computer node failures and their potential root cause with a high lead time.

The LogLearn model is implemented with an online failure prediction method, which is evaluated using multiple algorithms. Logistic regression showed the best performance in prediction. The LogLearn model outperformed previous studies' models in terms of accuracy, precision, recall, and f1-score. Additionally, an online time-series prediction model using the SNARIMAX algorithm was implemented to forecast the potential time of failure. Although previous studies have shown promising results, their lead times were insufficient to fix the underlying cause of failure in advance. Thus, LogLearn provides a viable alternative approach for failure prediction in computer systems.

Keywords: System logs, data streams, failure prediction, anomaly detection, continuous machine learning

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ALGORITHMS