**VIETNAM GENERAL CONFEDERATION OF LABOR**

**TON DUC THANG UNIVERSITY**

**FACULTY OF INFORMATION TECHNOLOGY**



**REPORT**

**MACHINE LEARNING**

*Instructor*: **LE ANH CUONG**

*Student*: **BUI HAI DUONG - 521H0220**

**BUI ANH PHU - 521H0508**

**VO KIEN NAM - 521H0000**

*Class*: **21H50302**

**HO CHI MINH CITY, 2023**

**VIETNAM GENERAL CONFEDERATION OF LABOR**

**TON DUC THANG UNIVERSITY**

**FACULTY OF INFORMATION TECHNOLOGY**



**REPORT**

**MACHINE LEARNING**

*Instructor*: **PhD.LE ANH CUONG**

*Student*: **BUI HAI DUONG - 521H0220**

**BUI ANH PHU - 521H0508**

**VO KIEN NAM - 521H0000**

*Class*: **21H50302**

**HO CHI MINH CITY, 2023**

ACKNOWLEDGEMENT

To complete this essay, besides our own efforts, we have received a lot of help in terms of knowledge, experience, and skills from the school and teachers. First and foremost, we would like to express my special gratitude to Nguyen Chi Thien - the lecturer of Design and Analysis of Algorithms course who has taught us valuable knowledge of the subject. That knowledge is the foundation for us to continue learning and effectively apply it to this essay. Additionally, We would like to thank the teacher for allowing us to complete this essay, which has helped us to further develop my understanding of the subject. Thank you for your guidance and support in helping us to complete this essay to the best of my ability. Moreover, we would also like to express our gratitude to the school and the teachers who have compiled the Design and Analysis of Algorithms materials, providing me with useful resources for research and essay writing. Thank you sincerely!

**COMPLETION OF THESIS**

**AT TON DUC THANG UNIVERSITY**

We here by certify that this thesis is my/our own work and was conducted under the guidance of Nguyen Chi Thien. The research and results presented in this thesis are truthful and have not been published previously in any form. The data presented in tables and figures used for analysis, comments, and evaluations were collected by the author from various sources and are clearly cited in the reference section.

Moreover, this thesis includes some comments, evaluations, and data from other authors and organizations, which are properly cited and referenced.

If any misconduct is detected, I fully take responsibility for the content of my thesis. Ton Duc Thang University is not liable for any copyright infringement that may occur during the thesis completion process.

*Ho Chi Minh City, October 22, 2023*

*Author*

*(signature and full name)*

ACKNOWLEDGEMENT AND EVALUATION SECTION BY INSTRUCTOR

**Instructor's Acknowledgement Section**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Ho Chi Minh City, 2023

(signature and full name)

**Instructor's Evaluation Section**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Ho Chi Minh City, 2023

(signature and full name)

SUMMARY

This document showing the theory based on the research document of Efficient weighted probabilistic frequent itemset mining in uncertain databases, explain the java code following the research then represent the result benchmark.

Table of Contents

[ACKNOWLEDGEMENT i](#_Toc153876958)

[ACKNOWLEDGEMENT AND EVALUATION SECTION BY INSTRUCTOR iii](#_Toc153876959)

[SUMMARY iv](#_Toc153876960)

[**REFERENCES** 1](#_Toc153876961)

**LIST OF SYMBOLS AND ABBREVIATIONS**

**Symbols**

Uncertain dataset of size n

A set of transaction identifiers

An itemset of size m

An existential probability for item appearing in the transaction

A weight table for the itemset I. Each item has a real-valued weight

An integer between (0, n], the minimum support for expected support- based FI mining

**msup** An integer between (0, n], the minimum support for probabilistic FI mining

A real value between (0, 1], the probabilistic frequent threshold for probabilistic FI mining

A real value between [0, 1], the scale factor

**Abbreviations**

**PFI** probability frequent itemset

**w-PFI** weighted probability frequent itemset

# CHAPTER 1: PERSONAL TASK

1. **Optimizers**
   1. **Definition**

Optimizers in machine learning are the algorithms used to adjust parameters in training section with the aim is to minimize the loss function and maximize the accuracy of the training model. Each optimizer has specific update rules, learning rates, and momentum to find optimal model parameters for improved performance.

* 1. **Types of optimizers**
     1. **Gradient Descent**

**Formula:**

This algorithm will update the weight after each loop(epoch) until the reaches the possible minimum value. The final possible minimum value is the expected result.

**Advantages:** Gradient Descent is easy to understand. It can solve the problem of optimizing the model by updating the weight value in each loop of the iteration.

**Disadvantages:**

* Gradient Descent is a simple algorithm. Therefore, it is dependent on the initial parameters.
  + - 1. **Stochastic Gradient Descent (SGD)**

This is the variant of Gradient Descent. The base Gradient Descent algorithm updates the weight value each epoch and the Stochastic Gradient Descent updates the weight value times each epoch with is the amount of data included in the epoch.

* + - 1. **Stochastic with Momentum**
      2. **Mini Batch**
    1. **Adagrad**
    2. **RMS Prop**
    3. **AdaDelta**
    4. **Adam**
  1. **Comparison**

1. **Continual learning and Test Production**

# CHAPTER 2: GROUP TASK

**REFERENCES**

1. Zhiyang Li | Fengjuan Chen | Junfeng Wu | Zhaobin Liu | Weijiang Liu (2020), Efficient weighted probabilistic frequent itemset mining in uncertain databases.