UNIVERSITAS
GADJAH MADA
IMPLEMENTATION OF DATA SCIENCE TECHNIQUES IN STOCKS PORTFOLIO MANAGEMENT
M R B E NOERRAHMAN, Prof. Dr. Eduardus Tandelilin, M.B.A.
Universitas Gadjah Mada, 2021 | Diunduh dari http://etd.repository.ugm.ac.id/

TABLE OF CONTENTS

ACKNOWLEDGEMENT ..... **i**
TABLE OF CONTENTS ..... **ii**
LIST OF TABLES ..... **v**
TABLE OF FIGURES ..... **vii**
ABSTRACT ..... **x**
INTISARI ..... **xi**
1. PRELIMINARY **1**
1.1. Background 1
1.2. Problem Formulation 3
1.3. Research Questions 4
1.4. Research Objective 4
1.5. Research Contributions 4
1.6. Research Scope 5
1.7. Systematic of Writing 7
2. LITERATURE REVIEW **9**
2.1. Modern Portfolio Theory (MPT) 9
2.1.1. Expected Return 10
2.1.2. Variance as Risk Estimation 10
2.2 Hierarchical Risk Parity (HRP) 11
2.3. Capital Asset Pricing Model (CAPM) 13
2.4. Data Science Techniques 15
2.4.1. Data Visualisation 16
2.4.2. Feature Engineering 16
2.4.3. Principal Component Analysis (PCA) 16
2.4.4. Cluster Analysis 17
2.4.5. Monte Carlo Method 22
2.5. Previous Studies 23
2.6. Research Framework 25
3. RESEARCH METHOD **26**

UNIVERSITAS
GADJAH MADA
IMPLEMENTATION OF DATA SCIENCE TECHNIQUES IN STOCKS PORTFOLIO MANAGEMENT
M R B E NOERRAHMAN, Prof. Dr. Eduardus Tandelilin, M.B.A.
Universitas Gadjah Mada, 2021 | Diunduh dari http://etd.repository.ugm.ac.id/

3.1. Research Design 26
3.2. Data and Sample 27
3.3. Variables and Measurements 29
3.3.1. Return and Rate of Return 31
3.3.2. Correlation Coefficients of Stocks 32
3.3.3. Variance and Standard Deviation of Returns 33
3.3.4. Covariance of Stocks and Market Prices 34
3.3.5. Security Characteristic Line (SCL) Regression 34
3.3.6. Risk-adjusted Return Ratios 35
3.3.7. Principal Component Analysis (PCA) 36
3.3.8. Distance Matrix 36
3.3.9. k-means Clustering 36
3.3.10. Hierarchical Agglomerative Clustering 37
3.4. Data Analysis Method 38
3.4.1. Feature Engineering 38
3.4.2. Portfolio Construction 40
3.4.3. Mean-Variance Portfolio Optimisation 40
3.4.4. Hierarchical Parity Risk (HRP) Portfolio 42
3.4.5. Portfolio Assessment 42
3.5. Case Profile 43
4. FINDINGS AND DISCUSSION 44
4.1. Data Descriptions 44
4.1.1. Stocks Selection 44
4.1.2. Raw Prices Data 45
4.2. Data Processing and Model Construction 49
4.2.1. Model Construction 49
4.2.2. Daily Returns Data 49
4.2.3 Market Premium and Risk Premiums 54
4.2.4. Expected Returns Testing 55
4.2.5. Expected Volatility Testing 56
4.2.6. Features from Model Construction Training Data 57
4.2.7. K-means Clustering Stocks Selection 61

UNIVERSITAS
GADJAH MADA
IMPLEMENTATION OF DATA SCIENCE TECHNIQUES IN STOCKS PORTFOLIO MANAGEMENT
M R B E NOERRAHMAN, Prof. Dr. Eduardus Tandelilin, M.B.A.
Universitas Gadjah Mada, 2021 | Diunduh dari http://etd.repository.ugm.ac.id/

4.2.8. Agglomerative Clustering Stocks Selections 64
4.2.9. Random Portfolios for Model Construction 67
4.2.10. Market Performance 2019 to 2020 68
4.2.11. Backtesting: Equal Weights Allocation 69
4.2.12. Backtesting: Mean-Variance Optimisation (MVO) 72
4.2.13. Backtesting: Hierarchical Parity Risk Portfolios (HRP) 75
4.3. Analysis and Model Validation 78
4.3.1. Model Validation 78
4.3.2. Features from Model Validation Training Data 79
4.3.3. k-means Clustering on Model Validation Dataset 80
4.3.4. Agglomerative Clustering on Model Validation Dataset 82
4.3.5. Market Performance 2020 to 2021 85
4.3.6. Validation Backtesting: Equal Weights Allocation 86
4.3.7. Validation Backtesting: Optimised Portfolios 89
4.4. Discussions 92
5. CONCLUSION 95
5.1. Conclusion 95
5.2. Limitations 96
5.3. Recommendations 96
BIBLIOGRAPHY 98
APPENDIX 102