

**A DATA MINING APPROACH TO
ENHANCING BIRTH AND DEATH
REGISTRATION PROCESSES**

ERFAN HASMIN

**ASIA e UNIVERSITY
2025**

A DATA MINING APPROACH TO ENHANCING BIRTH AND DEATH
REGISTRATION PROCESSES

ERFAN HASMIN

A Thesis Submitted to Asia e University in
Fulfilment of the Requirements for the
Degree of Doctor of Philosophy

March 2025

ABSTRACT

Accurate and timely birth and death registration is crucial for effective policymaking and public service delivery. However, Indonesia's current population administration system faces challenges such as centralized registration processes and low public awareness, leading to delays and incomplete records. This study explores the use of data mining techniques to enhance registration efficiency by analyzing birth and death records from Makassar city's population and civil registration office. Using k-means clustering, apriori association rules, and C5 decision trees, this research identifies key patterns influencing late registrations. The optimal number of clusters for birth and death data is determined as three using elbow and silhouette validation methods. The apriori algorithm refines registration data by identifying associations that reduce inconsistencies, while decision tree analysis highlights critical factors contributing to registration delays. A total of 45 decision trees were generated, leading to policy recommendations aimed at improving data collection and public compliance. This study contributes to ICT governance and public administration by demonstrating how data-driven approaches can optimize civil registration services. The findings offer actionable insights for policymakers to enhance registration models, reduce delays, and improve public accessibility. Future research may explore the integration of deep learning models for further automation of the registration process and enhance predictive accuracy.

Keywords: Civil registration, data mining model, k-means, apriori, C5.0

APPROVAL

This is to certify that this thesis conforms to acceptable standards of scholarly presentation and is fully adequate, in quality and scope, for the fulfilment of the requirements for the degree of Doctor of Philosophy.

The student has been supervised by: **Prof Ts Dr Aedah Abd Rahman**

The thesis has been examined and endorsed by:

Ts Dr Amna Saad,

Asia e University

Examiner 1

Prof Madya Ts Dr Zulkefli Mansor,

Universiti Kebangsaan Malaysia

Examiner 2

This thesis was submitted to Asia e University and is accepted as fulfilment of the requirements for the degree of Doctor of Philosophy.



Professor Dr Siow Heng Loke

Asia e University

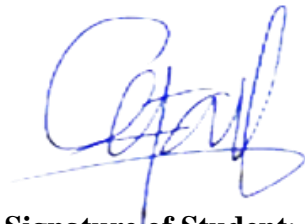
Chairperson, Examination Committee

(21 March 2025)

DECLARATION

I hereby declare that the thesis submitted in fulfilment of the requirements for the Degree of Doctor of Philosophy is my own work and that all contributions from any other persons or sources are properly and duly cited. I further declare that the material has not been submitted either in whole or in part, for a degree at this or any other university. In making this declaration, I understand and acknowledge any breaches in this declaration constitute academic misconduct, which may result in my expulsion from the programme and/or exclusion from the award of the degree.

Name: Erfan Hasmin



Signature of Student:

Date: 21 March 2025

ACKNOWLEDGEMENTS

My profound gratitude extends to Yayasan Dipanegara and Dipa Makassar University for their invaluable scholarship support, which provided the foundational means to pursue and complete my research. I am equally indebted to Asia e University for cultivating an exceptional academic environment that fostered intellectual growth and facilitated the extensive work required for this dissertation. A very special acknowledgment goes to my supervisor, Prof Ts Dr Aedah Abd Rahman, whose unwavering dedication, profound insights, and tireless guidance were truly indispensable throughout every stage of this demanding journey. Finally, to my incredible cohort of PhD ICT students, your steadfast encouragement, shared understanding, and genuine camaraderie provided invaluable emotional support and motivation, making the challenges far more manageable and the successes more meaningful.

TABLE OF CONTENTS

ABSTRACT	ii
APPROVAL	iii
DECLARATION	iv
ACKNOWLEDGEMENTS	vi
TABLE OF CONTENTS	vii
LIST OF TABLES	ix
LIST OF FIGURES	xiv
LIST OF ABBREVIATION	xvii
CHAPTER 1 INTRODUCTION	1
1.0 Background of Study	1
1.1 Problem Statement	5
1.2 Research Objectives	8
1.3 Research Questions	8
1.4 Significance of the Study	13
1.5 Operational Definition of Terms	15
CHAPTER 2 LITERATURE REVIEW	18
2.0 Introduction	18
2.1 Theoretical Framework	19
2.1.1 Population Administration System	20
2.1.2 Challenges in Registration	27
2.1.3 Data Mining Clustering	31
2.1.4 Data Mining Association	40
2.1.5 Data Mining Decision Tree	54
2.2 Review of Research	66
2.3 Chapter Summary	90
CHAPTER 3 METHODOLOGY	91
3.0 Introduction	91
3.1 Research Framework	91
3.2 Research Activities	93
3.2.1 Clusters of Birth and Death Data	94
3.2.2 Determination of Association Rules Between Attributes	96
3.2.3 Recommendation Based on Decision Tree Method	98
3.2.4 Decision Tree Rules Validation	99
3.3 Data Collection	101
3.3.1 Data Collection Techniques	102
3.3.2 Data Processing Tools	103
3.3.3 Preprocessing of Birth and Death Data	104
3.3.4 Data Visualization	114
3.4 Chapter Summary	121
CHAPTER 4 RESULTS	123
4.0 Introduction	123
4.1 Birth and Death Data Clustering	123

4.1.1	Implementation of K-Means on Birth Data	124
4.1.2	Implementation of K-Means on Mortality Data	129
4.1.3	Determining the Optimal Number of Clusters	136
4.2	Data Cluster Association Rules	147
4.2.1	Converting Nominal Data into Binomial	147
4.2.2	Apriori Method on Birth Data	152
4.2.3	Apriori Method on Mortality Data	166
4.3	Decision Trees on Cluster Data	179
4.3.1	C5.0 Decision Tree on Birth Data	179
4.3.2	C5.0 Decision Tree on Mortality Data	201
4.3.3	Model Evolution	219
4.4	Descriptive Analysis	221
4.4.1	K-Means Clustering Results	221
4.4.2	Apriori Association Rule Result	223
4.4.3	C5.0 Decision Tree Result	226
4.5	Discussion of Findings	240
4.6	Conclusion	242
4.7	Chapter Summary	243
CHAPTER 5	DISCUSSION AND CONCLUSION	245
5.0	Introduction	245
5.1	Limitations of Research	246
5.2	Implications of Research	249
5.3	Contributions of Research	251
5.3.1	Theoretical Contributions	252
5.3.2	Practical Contributions	254
5.3.3	Contribution to Methodology	256
5.4	Recommendation for Future Research	257
	REFERENCES	259
	APPENDICES	268
	Appendix A: Rapid Miner Model	268
	Appendix B: Data Collection Sample	272
	Appendix C: Letter of Permission to Conduct Research	276
	Appendix D: Validation Document	277

LIST OF TABLES

Table	Page
Table 1.1 Mapping Relation Among of PS, RQ, and RO	10
Table 2.1 Comparison of Association Rule Algorithms	41
Table 2.2 Research Gap	74
Table 3.1 Amount of Data by Status	106
Table 3.2 Sub Criteria Place of Birth	107
Table 3.3 Sub Criteria Type of Birth	108
Table 3.4 Sub Criteria Birth Attendant	108
Table 3.5 Sub Criteria Demographics	108
Table 3.6 Sub Criteria Gender	109
Table 3.7 Sub Criteria the Birth was Registered	109
Table 3.8 Sub Criteria Age at Death	109
Table 3.9 Sub Criteria Cause of Death	110
Table 3.10 Sub Criteria Place of Death	110
Table 3.11 Sub Criteria Evidence of Death	111
Table 3.12 Sub Criteria Residential Demographics	111
Table 3.13 Sub Criteria Duration of Time at Death when Registered	111
Table 4.1 Starting Point of Each Birth Cluster is Random	124
Table 4.2 Starting Center Point of Each Birth Cluster	125
Table 4.3 Results of the Calculation Distance Birth Cluster First Iteration	126
Table 4.4 New Center Point of Each Cluster Birth Second Iteration	127
Table 4.5 Results of the Second Iteration Cluster Distance Calculation	127
Table 4.6 Results of the First and Second Iteration Birth Clusters	128
Table 4.7 Each Cluster of Death Data Points is Randomly Selected	129

Table 4.8	Starting Center Point of Each Death Cluster	129
Table 4.9	Calculation Distance Death Cluster First Iteration	131
Table 4.10	New Center Point of Each Second Iteration Death Cluster	131
Table 4.11	Calculation Distance Death Cluster Second Iteration	132
Table 4.12	First and Second Iteration Death Cluster Results	133
Table 4.13	New Center Point of Each Third Iteration Death Cluster	134
Table 4.14	Calculation Distance Death Cluster Third Iteration	135
Table 4.15	Second and Third Iteration Death Cluster Results	135
Table 4.16	K-Means Clustering Results on Birth Data	136
Table 4.17	K-Means Clustering Results on Death Data	137
Table 4.18	Birth Data Cluster Distance Performance Value	137
Table 4.19	Birth Data Silhouette Coefficient	139
Table 4.20	Comparison Birth Data Optimum Cluster	140
Table 4.21	Death Data Cluster Distance Performance Value	142
Table 4.22	Death Data Silhouette Coefficient	143
Table 4.23	Comparison Death Optimum Cluster with Elbow and Silhouette	145
Table 4.24	Place of Birth Binominal Distribution	147
Table 4.25	Type of Birth Binominal Distribution	148
Table 4.26	Birth Attendant Binominal Distribution	148
Table 4.27	Birth Order Binominal Distribution	148
Table 4.28	Demographics Binominal Distribution	149
Table 4.29	Gender Binominal Distribution	149
Table 4.30	The Age of the Child Binominal Distribution	150
Table 4.31	Age at Death Binominal Distribution	150
Table 4.32	Cause of Death Binominal Distribution	151

Table 4.33	Place of Death Binominal Distribution	151
Table 4.34	Evidence of Death Binominal Distribution	151
Table 4.35	Residential Demographics Binominal Distribution	152
Table 4.36	Duration of Death when Registered Binominal Distribution	152
Table 4.37	Birth Data Binominal	153
Table 4.38	Support Itemset Birth Data	154
Table 4.39	Support Two Itemset Birth Data	154
Table 4.40	Support Two Itemset with Minimum Support 0.2 Birth Data	155
Table 4.41	Confidence Value Birth Data	157
Table 4.42	Minimum Confidence 0.5 Birth Data	158
Table 4.43	Lift Ratio Birth Data	159
Table 4.44	Lift Ratio Minimum 1 Birth Data	160
Table 4.45	Eliminate Data Cluster Using Association Rule Birth Data	161
Table 4.46	Cluster_0 Birth Data Association Relationship	162
Table 4.47	Cluster_1 Birth Data Association Relationship	163
Table 4.48	Cluster_2 Birth Data Association Relationship	163
Table 4.49	Results of Elimination Birth Data with Apriori	164
Table 4.50	Mortality Data Binominal	166
Table 4.51	Support Itemset Death Data	167
Table 4.52	Support Two Itemset Death Data	168
Table 4.53	Support Two Itemset with Minimum Support 0.2 Death Data	168
Table 4.54	Confidence Value Death Data	170
Table 4.55	Minimum Confidence 0.5 Death Data	171
Table 4.56	Lift Ratio Death Data	173
Table 4.57	Lift Ratio Minimum 1	174

Table 4.58	Eliminate Death Data Cluster Using Association Rule	175
Table 4.59	Cluster_0 Death Data Association Relationship	176
Table 4.60	Cluster 1 Death Data Association Relationship	177
Table 4.61	Cluster 2 Death Data Association Relationship	177
Table 4.62	Results of Elimination Death Data with Apriori	177
Table 4.63	Criteria and Attributes in Birth Data	180
Table 4.64	Entropy Value of Each Attribute	182
Table 4.65	Gain Ratio Value for Each Criterion with Data	184
Table 4.66	Generates Rules from Birth Data Cluster 0	190
Table 4.67	Generates Rules from Birth Data Cluster 1	193
Table 4.68	Generates Rules from Birth Data Cluster 2	197
Table 4.69	Confusion Matrix Birth Data Cluster 0	198
Table 4.70	Confusion Matrix Testing on Birth Data for Each Cluster	199
Table 4.71	Birth Data Confusion Matrix Before and After Model	200
Table 4.72	Criteria and Attributes in Mortality Data	202
Table 4.73	Gain Ratio Value for Each Criterion of Death Data	204
Table 4.74	Gain Ratio for Each Criterion of Death Data	206
Table 4.75	Generates Rules from Death Data Cluster 0	212
Table 4.76	Generates Rules from Death Data Cluster 1	214
Table 4.77	Generates Rules from Birth Data Cluster 2	215
Table 4.78	Confusion Matrix Death Data Cluster 0	216
Table 4.79	Confusion Matrix Testing on Death Data for Each Cluster	217
Table 4.80	Death Data Confusion Matrix Before and After Model	218
Table 4.81	Data Distribution in Each Cluster	222
Table 4.82	Distribution of Label Criteria for Each Cluster	222

Table 4.83	Data Pruning Results Based on Association Rules	224
Table 4.84	Scatter Plot Data Cluster Before after Apriori	225
Table 4.85	Decision Tree of Birth Data Policy and Activity Recommendations	228
Table 4.86	Decision Tree of Death Data Policy and Activity Recommendations	235

LIST OF FIGURES

Figure		Page
Figure 2.1	Theoretical Framework	19
Figure 2.2	Population Administration System Framework	21
Figure 2.3	Birth Registration Process	24
Figure 2.4	Death Registration Process	27
Figure 2.5	Cluster using K-Means	36
Figure 2.6	Association Rule with Apriori	49
Figure 2.7	Decision Trees Classify	56
Figure 2.8	Flowchart Algorithm C5.0	61
Figure 3.1	Research Framework	92
Figure 3.2	Research Activities	93
Figure 3.3	Research Activities for Birth and Death Data Cluster	94
Figure 3.4	Research Activities for Association Rules Between Attributes	96
Figure 3.5	Research Activities for Recommendation Based on Decision Tree	98
Figure 3.6	Missing Value Birth Attribute	113
Figure 3.7	Missing Value Death Attribute	113
Figure 3.8	Visualization Place of Birth Criteria	114
Figure 3.9	Visualization Type of Birth Criteria	115
Figure 3.10	Visualization Birth Attendant Criteria	115
Figure 3.11	Visualization Birth Order Criteria	116
Figure 3.12	Visualization Demographics of where Parents Live Criteria	116
Figure 3.13	Visualization Gender Criteria	117
Figure 3.14	Visualization Age of the Child when the Birth Criteria	117
Figure 3.15	Visualization Age at Death Criteria	118

Figure 3.16	Visualization Cause of Death Criteria	119
Figure 3.17	Visualization Place of Death Criteria	119
Figure 3.18	Visualization Evidence of Death Criteria	120
Figure 3.19	Visualization Residential Demographics Criteria	120
Figure 3.20	Visualization Criteria Duration of Time when Registered	121
Figure 4.1	Birth Data Elbow Graphics	138
Figure 4.2	Birth Data Silhouette Graphics	140
Figure 4.3	Distribution of Birth Data With 3 Clusters	141
Figure 4.4	Death Data Elbow Graphics	143
Figure 4.5	Death Data Silhouette Graphics	144
Figure 4.6	Distribution of Mortality Data with 3 Clusters	146
Figure 4.7	Birth Clustering with K-Means	165
Figure 4.8	Data Elimination using Apriori	165
Figure 4.9	Death Clustering with K-Means	178
Figure 4.10	Mortality Data Elimination Using Apriori	179
Figure 4.11	Branch Formation of Root Node Birth Data Simulation	186
Figure 4.12	Decision Tree the Birth Data Simulation	187
Figure 4.13	Decision Tree of Birth Data Cluster 0	189
Figure 4.14	Decision Tree of Birth Data Cluster 1	192
Figure 4.15	Decision Tree of Birth Data Cluster 2	196
Figure 4.16	Branch Formation of Root Node Mortality Data Simulation	209
Figure 4.17	Decision Tree from the Death Data Simulation	209
Figure 4.18	Decision Tree Death Data Cluster 0	211
Figure 4.19	Decision Tree Death Data Cluster 1	213
Figure 4.20	Decision Tree Death Data Cluster 2	215

LIST OF ABBREVIATION

DBI	Davies Bouldin Index
DBI	Davies Bouldin Index
IKI	Institut Kewarganegaraan Indonesia
SIAK	Sistem Informasi Administrasi Kependudukan
SSE	Sum of Square Error
UNICEF	United Nations Childrens Fund
UPT	Unit Pelaksana Teknis/Technical Implementation Unit
WHO	World Health Organization

CHAPTER 1

INTRODUCTION

1.0 Background of Study

The Population Administration System in Indonesia has lasted since 2011. The system is Population Administration Information System (SIAK: *Sistem Informasi Administrasi Kependudukan*); It involves the registration of both births and deaths. The addition of new data and updates is entered into the SIAK system, which is hosted on a decentralized server within the city. Periodically, this information will be synchronized with a central server located at the Ministry of Home Affairs (Departemen Dalam Negeri Indonesia, 2011). The provisions regarding birth registration in the system and making birth certificates are regulated in articles 27 and 49. Simultaneously, the data of death reports that require families, community leaders, or residents to report within 30 days of the incidents is regulated in article 44 of Law of the Republic of Indonesia Number 24, 2013. Meanwhile, the steps in government programs to improve the birth and death registration rates become the authority of each region's civil registration offices. Procedures for registering births and deaths by visiting the population service office by bringing birth and death events. This procedure has the constraint that the community will only process the document if needed at some point in the future, even though the document should be registered after the birth and death events occur. This is due to several factors, including access to registration centralized in the district capital, making it difficult for people to come to the office due to distances and long queues. There is low public awareness of reporting births to issue birth certificates; generally, people inform deliveries when the child is sick, requires health insurance, and is entering school age (Ritonga et al.,

2021). Likewise, in reporting death events, death events are generally reported to the civil registry office to issue death certificates to withdraw insurance funds or become heirs (Adi, 2022).

Various efforts to enhance birth and death registration are ongoing. The optimization of technology in civil registration aims to achieve more efficient outcomes. Simultaneously, technology assumes a critical function in bolstering population management services. It has the potential to facilitate inter-agency collaboration, expedite registration processes, and streamline procedures. Consequently, it's imperative to oversee innovations that harness technology for support effectively (Adomako & Nguyen, 2024). However, the involvement of other institutions, such as health institutions, in the birth registration process can improve birth registration but becomes a dilemma because medical institutions already have the primary function of providing health facilities (Siagian et al., 2019). Moreover, registering deaths can involve sub-district and village institutions. This will increase the number of births and deaths. The Civil Registration Office of Makassar has innovated by developing a birth registration system that involves health institutions and urban village offices named KUCATA'KI. This innovation makes it easier for the people of Makassar to get services from the government and collaborate with related parties in the reporting process of some health institutions and every district in the city of Makassar (Sulmiah et al., 2022). In addition to these technological innovations, efforts to improve birth and death registration are also carried out in the civil registration service policy with various activities such as direct services to sub-districts and increasing understanding and technical capabilities of civil registration in the form of seminars and training. The Three in One Service Program also completes the application for a Birth Certificate by including the Family Card and Child Identity

Card as program outputs. The program is running well and should be more optimal (Malik & Ammar Hadi, 2022). The full potential of discounts offered to holders of the Child Identity Card (KUMARA) has not yet been fully realized in terms of quantity. The promotion and oversight of this program have not been optimized, requiring practical observations and supervision by service providers when serving recipients (Wirata et al., 2022). In addition, to ensure the success of the Civil Registration System for registering births and deaths, it is crucial to enhance infrastructure through a reliable electricity supply, better building facilities, a well-trained workforce, and the implementation of an online system with the necessary software installation (Rane et al., 2020).

Improvements in birth and death registration can be achieved by thoroughly understanding the data that has been recorded to date. An analytical approach is essential for understanding birth and death data. This necessity is underscored by the data mining analysis conducted on civil registration data. Use of data mining techniques on government data is proven to make better planning and decision-making (Musadad et al., 2023). As well as the birth data analysis using the Neural Network method emphasizes the need for data mining analysis on civil registration data to understand e-governance data better (Desai, 2019). Utilization of birth and death data for the development and determination of policies and activities needs to be done to increase the number of registered births and deaths. Data recorded for three years can be analyzed using data mining tools to find new information that can be used as a basis for determining policies for developing birth and death registration systems. The utilization of pre-established algorithms and association rules in decision-making holds importance for policymakers in public service (Ebenezer, 2019), and the use of data mining techniques on civil registry databases to find association rules using the a

priori method succeeded in presenting new regulations that can be used for decision-makers to do the right thing for certain groups of rules that are generated (Musa & Ahmad, 2019). This underscores the importance of collecting data for analysis to derive association rules that relate to the correlation between the area of residence and the timeliness of reporting birth and death events, thereby enabling the precise formulation of activities. Beyond identifying specific rules and correlations within collected data, it is also possible to categorize the data using relevant and suitable criteria. This categorization allows the government to formulate public policies that align with the actual situation (Viloria & Lezama, 2019). Data mapping is needed for policy adjustments related to increasing birth and death registration, and a policy will be adjusted to the characteristics of the existing data clusters. This way, the policies taken can be implemented with finances limited to each cluster (Jurun et al., 2017). This ensures that improvements to the birth and death registration process are implemented based on accurate cluster data.

Understanding the characteristics of birth and death data recording, along with the factors influencing community registration of these events, is essential. These factors may include social, economic, cultural, and geographical aspects that affect individual or family decisions to report births and deaths (Yokobori et al., 2021). Additionally, the community awareness of the importance of civil registration plays a crucial role. In this context, data mining analysis is vital. It can uncover patterns and trends that are not immediately apparent, such as specific times of the year when registrations are lower or demographic groups that are less likely to report these significant events. The insights gained from this analysis can inform government strategies, particularly for the population and civil registration office, enabling the design of more effective programs and activities. For example, if data indicates low

birth registration rates among individuals with lower education levels, the government can implement educational programs highlighting the importance of civil registration or establish direct service initiatives in hard-to-reach areas. By developing activities based on accurate data and thorough analysis, it is anticipated that improvements in birth and death registration will increase in the future. This will ensure that every individual is officially recognized by the state, a fundamental right of every citizen. Enhanced registration can also improve public services, such as health and education, as accurate data facilitates more targeted resource allocation.

1.1 Problem Statement

Efforts to improve birth and death registration through various activities conducted to date are not yet data-driven. Initiatives undertaken without a data foundation may hinder the community access to basic rights, such as education, health services, and inheritance rights (Yu et al., 2020). Efforts and initiatives aimed at improving birth and death registration across various scenarios, geographic areas, and social contexts should be formulated based on data, prioritizing the results of data analysis as a crucial factor in these endeavors. Current data mining techniques can analyze specific data, including civil registration data, more effectively than other data analysis methods (Tekieh & Raahemi, 2015). This serves as the foundation for this study to utilize data mining analysis methods on civil registration data to:

- i. Birth and death registration are fundamental events that must be recorded by the civil registration office. However, many regions in Indonesia, especially the city of Makassar, still face challenges in achieving optimal registration rates. One of the causes is the lack of understanding of the characteristics and patterns of registration in various community groups, which results in efforts

to improve birth and death registration being suboptimal. By understanding registration patterns in different community groups, service innovations can be tailored to these conditions. Additionally, by comprehending the characteristics and complex registration patterns, such as changes in population structure, migration, or changes in family composition, efforts to improve registration based on data can be more effective (Oktaviany et al., 2024). Therefore, a data mining approach is employed to categorize birth and death registration data using several criteria. This multi-criteria clustering enables the government to develop public policies that align with the actual situation (Viloria & Lezama, 2019). Clustering data based on relevant criteria, including birth and death events, child biodata, parental information, and regional demographics, is essential. The resulting data clusters can serve as a foundation for formulating activities aimed at improving birth and death registration.

- ii. Significant challenges remain in improving the registration of births and deaths. Data shows that more than half of births and deaths are reported more than thirty days after the event. This phenomenon reflects a gap in public understanding of the importance of official registration. Many individuals and families believe that obtaining a birth certificate and death certificate from the hospital is sufficient, without realizing that official registration at the population and civil registration office is a crucial step to ensure legal recognition and better access to public services. One approach to improving this registration is to understand the associations in the data that influence people's decisions to register births and deaths. By applying data mining techniques, especially association rule methods, we can identify patterns and relationships that exist in civil registration data. The association process is also