# Design of Personalized Asthma Management System With Data Mining Methods

Cut Fiarni[1]
Department of Information System,
InstitutTeknologiHarapanBangsa
DipatiUkur 80-84, Bandung, Indonesia
cutfiarni@ithb.ac.id

***Abstract*— Asthma is a chronic lung disease that inflames and narrows the airways. Asthma is a multifactorial chronic illness, there is no uniformity of triggers factors of asthma attack on asthma patients with different degree of asthma severity. The better understanding of the trigger factors could lead a better asthma management for asthma patient. in this paper, we proposed a design of personalized asthma management system. in this system we used pattern sequential mining, clustering and classification to predicting the patient's asthma attack based on extracted rules from data mining. the main methodology is to identify the most important trigger factors based on the clustering and classification rules to be extracted. The contribution to this research area is to analyze the suitable data mining technique and algorithm for the proposed system.**

***Keywords—Asthma; data mining; personalized management system.***

## I. INTRODUCTION

Asthma is a chronic lung disease that inflames and narrows the airways. It characterized by airway swollen and hyper-responsiveness, which causes airflow limitation. Asthma is a major cause of chronic morbidity and mortality throughout the world and there is evidence that its prevalence has increased considerably over the past 20 years, especially in children [1]. In Indonesia, asthma prevalence is 4.5% per mile, with the highest prevalence is in Sulawesi Tengah (7,8%), then in Nusa Tenggara Timur (7,3%), DI Yogyakarta (6,9%), and Sulawesi Selatan (6,7%) [2].

Chronic asthmatic sufferers need to be constantly observed to prevent sudden attacks. Of all asthma patients, 50% have symptoms on a daily basis and almost all patients report limitations to daily activities. A severe asthma attack may require emergency care, and it can cause death. Despite the high quality of available medications and treatment regimens that are being simplified on a regular basis, asthma is still not sufficiently controlled in many cases [3]. Sometime asthma symptoms are mild and go away on their own or after minimal treatment with an asthma medicine. Other times, symptoms continue to get worse [4]. It is important to treat symptoms of early staged, in order to prevent the symptoms from worsening and causing a severe asthma attack.

Asthma is a multifactorial chronic illness, there is no uniformity of triggers factors of asthma attack. each patient has different trigger factors that lead to different degree of asthma severity. The better understanding of the trigger factors could lead a better asthma management for asthma patient. But, the problem is because of the vast and complex trigger factors for asthma attacks, to there will be huge amounts of data that need to be processed and analyzed. Data mining provides the methodology and technology to extract these data into knowledge. One major advantage of data mining over a traditional statistical approach is its ability to deal directly with heterogeneous data fields, which a usually contained in medical data sets [5]. In medical research, data mining begins with the hypothesis and results are adjusted accordingly. With the use of data mining methods, useful patterns of information can be found in this data, that will later be used for further research and report evaluation [6].

The purpose of this study is to explore the applicability of data mining technique in the efforts of personalized asthma management and asthma severe attack prevention with particular emphasis to build a model that could help to extract pattern of asthma induced factors. With this objective, data mining technique was employed to classify patient daily activity records on the basis of the values of attributes asthma attack, the the trigger factors and asthma attack degree of severity.

In this proposed Personalized Asthma Management System, daily activities, external and internal trigger factors of asthma attack are collected, then two data mining methods clustering and classification are adopted for predicting trigger factors on asthma attack in different degree of severity. The first methodology objective is to extract the significant information of trigger factors of asthma attacks by clustering user's daily activity over some period of time. Then system will build a classifier by rules that extracted from the clustering method. With this proposed system, the patients hopefully could gain better understanding of their trigger factors of asthma attack, that could lead to better asthma management and a better quality of life. The rest of this paper is organized as follows. In section 2, we review several related studies. In section 3 we explained the propose system architecture in detail.. Finally, we summarize our research and list some future work in section 4.

## II. RELATED WORK

Definition Data mining or is the non-trivial extraction of implicit, previously unknown and potentially useful information from the data. This encompasses a number of technical approaches, such as clustering, data summarization,

classification, finding dependency networks, analyzing changes, and detecting anomalies [6]. Today, data mining is one of the best technique to gain knowledge from huge amounts of data that has been used by many organizations. Especially in healthcare, which their transaction data are really complex and sometime need to be recorded on very long period of time, consequently it is resulting a voluminously dataset that is vital to gain better decision on diagnosis or treatment. Data mining brings a set of tools and techniques that can be applied to discover hidden patterns that provide healthcare professionals an additional source of knowledge for making important decisions.There are vast potential for data mining applications in healthcare. Generally, these can be grouped as the evaluation of treatment effectiveness; management of health care; customer relationship management; and detection of fraud and abuse. In this paper we focus data mining for Healthcare management. The main focus of data mining for healthcare management are to aid healthcare management on the better way to identify and track chronic disease states and high-risk patients, design appropriate interventions, and reduce the number of hospital admissions and claims.

There are already many studies which focus on the etiology and diagnosis of asthma. O'Leary study emphasized patients' DNA and found the allergen of asthma [7]. Jan et al. proposed a system named as Blue Angle Asthma Guardian System, which is designed for children asthma monitoring [8]. O' Sullivan et al. propose to incorporate formalized external expert knowledge in building a prediction model for asthma exacerbation severity for pediatric patients in the emergency department [9]. Tseng.proposed an integrated bio-signal data mining system with two data mining methods for predicting asthma attacks [10]. Meanwhile, Kudyba stated that strong disease management programs such as asthma management, depend on data mining technique[11].

## III. Personalized Asthma Management System With K-Mean And Rule Based Classification

In this section, we will explain the problem definition, system architecture and data mining methods for the proposed system. This section will also explain how well asthma management system and data mining are integrated and also describes the datasets undertaken for theproposed system.

### A. Problem Definition

Asthma is a multifactorial chronic illness, there are no uniformity of triggers factors of asthma attack. each patients has different trigger factors that lead to different degree of asthma severity. Triggers are things that may bother the airways, making it hard to breathe or cause asthma signs or symptoms. Ideally, any asthma patient should know their trigger factors, so they can avoid their triggers whenever possible. But the problem is, triggers factors of asthma attack can be different from person to person. The triggers varies from gender, age, BMI, environmental and living condition, variations in lifestyle, allergic factors, daily activity and exercise, eating habits, hormonal factors, and stress condition.

On this proposed system we divided the trigger factors of asthma attack into three major categories:

1. User profile: gender, age, BMI
2. InternalFactors : Exercise, alergic food. Infections(respiratory infections --like a cold, the flu, or sinusitis), Stress and emotion, Food Allergies, Diet with low in vitamin and high fat
3. External : Dust Mites and Dust Control, animals exposure , Environmental Tobacco Smoke, poor Moisture and Ventilation, Air Pollution, Household Products and Air Fresheners with strong odor and weather condition.

### B. System Architecture

In this research, in order to increase asthma management, we proposed a personalized asthma management system to records all significant information about patient regarding to its asthma condition, over some period of time. Then we will adopt data mining technique discover important information of patient asthma attacks with asthma datasets. The knowledge that extracted from data mining process will be used by system to give recommendations and restriction regarding to personal trigger factor of asthma attack. The knowledge also can be used to predict the possibility of asthma attack and its degree severity from patient activity and condition. The architecture of the proposed system illustrate on Figure 1.

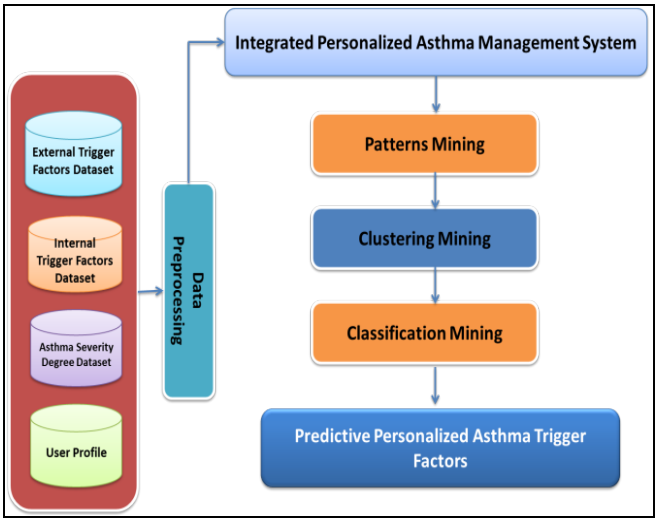


Fig. 1. Architecture Of Personalized Asthma Management System

As shown on Fig 1, the source data for data mining proces are come from four major database, with categorize as explained on problem definition. All data from all database would preprocessed before its could be extracted to gain pattern and knowledge, to minimize error and integrity of data.

### C. Data Mining Methods

This paper proposed personalized asthma management system with three data mining methods as illustrate on Fig.1. this following part will explained in detail how data mining techniques integrated to gather the predictive trigger factors of asthma attack with its degree of severity.

The first phase of this research is to build a personalized asthma monitoring application. The application will record all suspected asthma trigger factors with its degree of asthma severity daily. A dataset that will be used for data mining processes would be generated from data warehouse of personalized asthma monitoring application. The database would contain the following 4 major variables of particular interest to Asthma management: gender, age, body mass index (BMI), allergic status, the number of times a patient exercises per week, stress level, onset of an asthma attack and its possible excurbations. Decision support system is connected with the database management [12]. All trigger factors become target variable on the data mining process. It measured by a dichotomous variable indicating severity degree of the asthma attack. For this kind of integrated dataset, we adopted sequential pattern mining to discover high risk prefix sequential patterns of asthma attacks and to segmenting the data, based on the frequency and severity of asthma attack.

The next step in data mining process is to build clusters with the extracted patterns. In more detail, clustering the trigger factors that have the same asthma status and the degree of severity to help discovering asthma trigger rules. Clustering uses to make groups of objects, eachgroups has similar characteristics but has different characters from the objects in another group. Ussually the suitable number of clusters is not clearly shown. One difficulty in this proposed data mining process is that we don't have any prior knowledge about the structure of the data, or its labels, because clustering is considered to be an unsupervised learning problem [13]. The data flow of trigger factors of asthma attack with cluster analysis is illustrate on Figure 2.

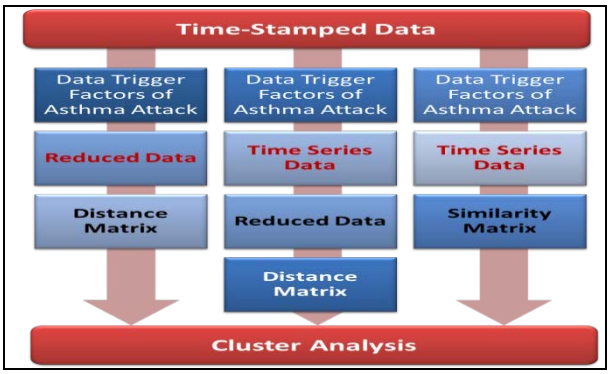


Fig. 1.Data Flow of Trigger Factor of Asthma Attack with cluster analysis

In this proposed system, data are clustered into three groups according to their asthma attack severity and we adopt K means algorithm for that purpose. K-means algorithm works optimally with categorical and numeric data so that this is the best for binary data clustering. It is simple and fairly fast [14],results are easy to interpret and it can work under a variety of conditions hence it stand as the standard algorithm for clustering. The fundamental idea is to find K average or mean values, about which the data can be clustered. The k-means algorithm is a simple iterative method to partitiona given dataset into a user specified number of clusters K. K-means is initialized from some random or approximate solution. The algorithm works as follows:

1. Define the number of clusters to be founded by k-means, which in this case is 3;
2. k-means starts by initializing a number of prototypes equal to number of desired clusters,
3. then the system will assigning each point to its closest center, then moving each prototype to the mean of its assigned points.
4. Reapeat step three (3) until it converges to a solution.

The popular heuristics for solving the k-means problem is based on a simple iterative scheme for finding a locally minimal solution. K-means algorithm works optimally with categorical and numeric data so that this is the best for binary dataclustering. It is simple and fairly fast [15], results are easy to interpret and it can work under a variety of conditions hence it stand as the standard algorithm for clustering.

The final step on data mining process is tobuilt a classifier for each cluster with the extracted patterns. Once the rules are learned from a clustering, they can be used for predicting the class type of previously unseen data, with classification technique. Classification is the most commonly applied data mining technique, which employs a set of pre-classified examples to develop a model that can classify the population of records at large. This approach frequently employs decision tree or neural network-based classification algorithms. The data classification process involves learning and classification. In Learning the training data are analyzed by classification algorithm. In classification test data are used to estimate the accuracy of the classification rules. If the accuracy is acceptable the rules can be applied to the new data tuples [16].

We used Classification trees to build decision rules in a hierarchical fashion to make a classification of severity degree of asthma attack as illustrate on Figure 3.

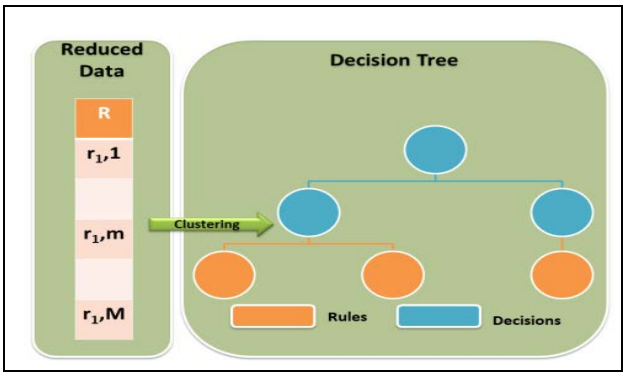


Fig. 2.Decision Tree analysis

A node in the decision tree represents a decision rule. A leaf in the decision tree represents one possible cluster as has been extracted from clustering methods in the previous step. The diagram below illustrates how the reduced data can provide one or more node statistics used in the decision tree. A decision tree will give requisite model which provides sufficient insights into the asthma management problem

The classification tree is used to group the severity of asthma attack into four major groups. We used classification as

has been done by Nagori [17], which are mild intermittent, mild persistent, moderate persistent and severe persistent.. The decision tree algorithm used for the proposed system based on following steps:

1. Check for basic symptoms of asthma attack
2. For each symptom find normalization of information gain and find the best attribute from external and internal factor from the highest normalizad of information gain.
3. Create a decision node that split into 4 groups

The severity group based on various symptoms when patient having asthma attack such as wheezing, coughing, shortness of breathing, chest tightness etc. Table 1 shows exactness of the four groups for asthma severity attack that will be used on classification trees on the proposed system.

TABLE I. GROUPS OF ASTHMA SEVERITY ATTACK

| Asthma Attack Severity Degree | Symptoms Of Airflow | Night Time Obstruction Symptoms | Peak Expiratory Flow* |
|---|---|---|---|
| Severe Persistent | • Continuous<br>• Limited physical activity | frequent | >60% of personal best |
| Moderate Persistent | • More than once a day<br>• Attack affect daily activity | More than once a week | Between 60% to 80% of personal best |
| Mild Persistent | • More than once a week but less than once a day | At least twice a month | More than 80% of personal best |
| Mild Intermittent | • Less than once a week | Less than twice a month | More than 80% of personal best |

*mandatory

Therefore, on this proposed system, to build comprehensive rules to gain good knowledge of a personal trigger factor of asthma attack,then we will validate the rules extracted with medical expert. The rule that's been used and validated from classifier process, will be used to predict asthma attacks by the matching patient's personal information, asthma allergy records, stress level and internal and external trigger factor as input data. When a prefix sequential pattern is matched in these data, system could remind users to take medicines or to seek medical treatment based on the prediction of asthma attack severity.

## IV. CONCLUSION AND FUTURE WORK

This research is concerned with the study and analysis of data mining with data clustering and classification algorithms, analyzing the existing methods for predicting trigger factors of personalized asthma attack and to design an efficient and effective method for predicting severity of asthma attack based on personalized patient trigger factors. We proposed K-Means Clustering techniques used for segmenting severity degree of asthma attack. Then we used classification tree for predicting the likeliness of asthma attacks. This study will help asthma patients to have a better understanding of her/his asthma trigger factors, that could lead to improvements on their daily activity. It also to identify those trigger factors and taking appropriate medical action at the right time. The future work of this research is to build the prototype system and generated data for the mining proposed. We will also explore the most suitable data mining algorithm comparingseveral algorithms and find the best precision and recall.

## REFERENCES


[1] Global Initiative for asthma management and prevention, A pocket guide for physicians and nurses, GINA, 2012, pp 5-7

[2] Riset Kesehatan Dasar ,Badan Penelitian dan Pengembangan Kesehatan, Kementrian Kesehatan RI, 2013

[3] Gruffydd-Jones K. Measuring pulmonary function in practice. Practitioner 2002; 246: 445–449. Medline

[4] Global Initiative for Asthma (GINA). Global strategy for asthma management and prevention. Updated 2011. Cape Town: University of Cape Town Lung Institute; 2011

[5] Cios, K.J: Medical data mining and Knowledge Discovery. IEEE engineering in medicine and biology 7. (2000) pp 15-16

[6] Milovic Boris and Milan Milovic, Prediction and Decision Making in Healthcare Using Data Mining. International Journal of Public Health Science (IJPHS) Vol 1, No.2,December 2012,pp.69-78

[7] O'Leary, M. Koolpiruck, D. Balachandran, W. Emberlin, J. Lewis, R. "The role of electrostatic charge accumulated by respirable sized allergens with regard to thunderstorm asthma," Industry Applications Conference, 2005.

[8] R. L. Jan, J. Y. Wang, M. C. Huang, S. M. Tseng, H. J. Su, L.F. Liu. "An Internet-based interactive telemonitoring system for improving childhood asthma outcomes in Taiwan." Telemedicine and e-Health, 2006, pp. 1-28.

[9] Dympna O'Sullivan, William Elazmeh, Szymon Wilk, Ken Farion, Stan Matwin, Wojtek Michalowski, and Morvarid Sehatkar, "Using Secondary Knowledge to Support Decision Tree Classification of Retrospective Clinical Data", MCD 2007, LNAI 4944, pp. 238–251, 200

[10] Tseng Vincent S., Lee Chao-Hui,and Chen Jessie Chia-Yu. "integrated Bio-Signal Data Mining with Applications on Asthma Monitoring", National Cheng Kung University, Taiwan

[11] Kudyba, S: Managing data mining : advice from experts. (2004)

[12] Zhou Qishen, Z. Yin, Q. Ying, W. Shahnhui, Intelligent Data Mining and Decision System for Commercial Decision Making. TELEKOMNIKA Indonesian Journal of Electrical Engineering Vol.12, No 1, January 2014,pp.792-801

[13] G.Gan,C.Ma, J Wu. " Data Clustering Theory, Algorithm, and Applications". American Statistical Association Alexandria, Virginia, 2007

[14] Arun K Punjari, ―Data Mining Techniques‖, Universities (India) Press Private Limited, 2006

[15] P. Bradley, U. Fayyad and C. Reina, "Scaling Clustering Algorithms to Large Databases", In Proc. of the 4th ACM SIGKDD, pp. 9-15, 1998.

[16] Brijesh Kumar Baradwaj, Saurabh Pal, Data mining: machine learning, statistics, and databases, 1996.

[17] Meghana Nagori, Pawar Suvana and Vivek Kshirsagar, " Managing asthma in children and analyzing best possible treatment with data mining approach of classification", in Proc of WCECS 2009, San Francisco, USA.