# THE DATA MINING APPLICATIONS OF SHOULDER PAIN PATIENTS TREATMENT: PHYSICAL THERAPY EQUIPMENT USAGE APPROACHES

**KITTISAK KAEWBOODDEE**

**A THESIS SUBMITTED IN PARTIAL FULFILLMENT**
**OF THE REQUIREMENTS FOR**
**THE DEGREE OF MASTER OF SCIENCE**
**(TECHNOLOGY OF INFORMATION SYSTEM MANAGEMENT)**
**FACULTY OF GRADUATE STUDIES**
**MAHIDOL UNIVERSITY**
**2015**

Copyright by Mahidol University

**COPYRIGHT OF MAHIDOL UNIVERSITY**

Thesis
entitled

# THE DATA MINING APPLICATIONS OF SHOULDER PAIN PATIENTS TREATMENT: PHYSICAL THERAPY EQUIPMENT USAGE APPROACHES

.....
Mr. Kittisak Kaewbooddee
Candidate

.....
Lect. Sotarat Thammaboosadee,
Ph.D. (Information Technology)
Major advisor

.....
Asst. Prof. Supaporn Kiattisin,
Ph.D. (Electrical and Computer Engineering)
Co-advisor

.....
Asst. Prof. Waranyu Wongseree,
Ph.D. (Electrical Engineering)
Co-advisor

.....
Prof. Patcharee Lertrit,
M.D., Ph.D. (Biochemistry)
Dean
Faculty of Graduate Studies
Mahidol University

.....
Asst. Prof. Supaporn Kiattisin,
Ph.D. (Electrical and Computer Engineering)
Program Director
Master of Science Program in Information Technology Management
Faculty of Engineering
Mahidol University

Copyright by Mahidol University

Thesis
entitled

# THE DATA MINING APPLICATIONS OF SHOULDER PAIN PATIENTS TREATMENT: PHYSICAL THERAPY EQUIPMENT USAGE APPROACHES

was submitted to the Faculty of Graduate Studies, Mahidol University
for the degree of Master of Science
(Technology of Information System Management)
on
June 4, 2015

.....
Mr. Kittisak Kaewbooddee
Candidate

.....
Lect. Taweesak Samanchuen,
Ph.D. (Electrical Engineering)
Chair

.....
Lect. Sotarat Thammaboosadee,
Ph.D. (Information Technology)
Member

.....
Asst. Prof. Kairoek Choeychuen,
Ph.D. (Electrical and Computer
Engineering)
Member

.....
Asst. Prof. Supaporn Kiattisin,
Ph.D. (Electrical and Computer
Engineering)
Member

.....
Prof. Patcharee Lertrit,
M.D., Ph.D. (Biochemistry)
Dean
Faculty of Graduate Studies
Mahidol University

.....
Lect. Worawit Israngkul,
M.S. (Technical Management)
Dean
Faculty of Engineering
Mahidol University

Copyright by Mahidol University

# ACKNOWLEDGEMENTS

It is a pleasure to thank the many people who made this thesis possible. I wish to express my sincere appreciation to those who have contributed to this thesis and supported me in one way or the other during this amazing journey.

First of all, I am extremely grateful to my main advisor, Dr. Sotarat Thammaboosadee for his guidance and all the useful discussions and brainstorming sessions, especially during the difficult conceptual development stage. His deep insights helped me at various stages of my research. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my master program study.

My sincere gratitude is reserved for Asst. Prof. Dr. Waranyu Wongseree and Asst. Prof. Dr. Supaporn Kiattisin for their invaluable insights and suggestions. I really appreciate their willingness to meet me at short notice every time.

Very special thanks to the Faculty of Medicine Siriraj Hospital, Mahidol University for giving me the opportunity to carry out my master degree and for their financial support. Heartfelt thanks goes to my mentor, Tang, Joy, Seree, Jiew, Buya and all of my friends for providing a stimulating and fun environment.

Words cannot express the feelings I have for my family members for their constant unconditional support - both emotionally and financially. I would not be here if it not for you. Special thanks are also due to my brother James for their amazing encouragement. It is amazing to have family close by so far away from home.

Finally, I would like to acknowledge the most important people in my life – my father and my mother. They have been a constant source of strength and inspiration. There were times during the past 30 years when everything seemed hopeless and I didn't have any hope. I can honestly say that it was only their determination and constant encouragement that ultimately made it possible for me to see this thesis research through to the end.

Copyright by Mahidol University

Kittisak kaewbooddee

THE DATA MINING APPLICATIONS OF SHOULDER PAIN PATIENTS TREATMENT: PHYSICAL THERAPY EQUIPMENT USAGE APPROACHES

KITTISAK KAEWBOODDEE 5437842 EGTI/M

M.Sc. (TECHNOLOGY OF INFORMATION SYSTEM MANAGEMENT)

THESIS ADVISORY COMMITTEE: SOTARAT THAMMABOOSADEE, Ph.D., WARANYU WONGSEREE, Ph.D.

ABSTRACT

This research is to apply data mining techniques for discovery and prediction of equipment usage amount from physical therapy equipment usage patterns based on a classification system and establish selection rules of physical therapy techniques based on the association rule discovery method. Both data mining aspects aim to support the decision making for physical therapists in the treatment of shoulder pain patients. The prediction system was driven by the usage patterns of physical therapy equipment, and the association rule discovering method was applied for studying the association with the amount of physical therapy equipment. The classification system was tasted and compared with the Naïve Bayes, Neural Network, and Decision Tree. The best result of Artificial Neural Network was 92.30% accuracy. In addition, the top five interesting discovered association rules are demonstrated. Both data mining applications of this research could support the decision making in the treatment of shoulder pain patients.

KEY WORDS: SHOULDER PAIN / PHYSICAL THERAPY / DATA MINING / CLASSIFICATION / ASSOCIATION RULE

44 pages

Copyright by Mahidol University

การประยุกต์ใช้เทคนิคการทำเหมืองข้อมูลในการรักษาผู้ป่วยโรคปวดไหล่: กรณีศึกษารูปแบบการใช้อุปกรณ์ในการรักษาทางกายภาพบำบัด

THE DATA MINING APPLICATIONS OF SHOULDER PAIN PATIENTS TREATMENT: PHYSICAL THERAPY EQUIPMENT USAGE APPROACHES

กิตติศักดิ์ แก้วบุตรดี 5437842 EGTI/M

วท.ม. (เทคโนโลยีการจัดการระบบสารสนเทศ)

คณะกรรมการที่ปรึกษาวิทยานิพนธ์: โษฑศ์รัตต ธรรมบุษดี, Ph.D., วรัญญ วงษ์เสรี, Ph.D.

บทคัดย่อ

การประยุกต์ใช้เทคนิคการทำเหมืองข้อมูลเพื่อศึกษาจำนวนการใช้เครื่องมือ โดยการศึกษาจากรูปแบบการใช้อุปกรณ์ในการรักษาทางกายภาพบำบัดด้วยวิธีการจำแนกประเภท (Classification system) และการสร้างกฎความสัมพันธ์เพื่อใช้ในการเลือกเทคนิคการรักษาทางกายภาพบำบัดด้วยวิธีการสร้างกฎความสัมพันธ์ (Association rules) เพื่อช่วยสนับสนุนการตัดสินใจการเลือกเทคนิคในการรักษาผู้ป่วยโรคปวดไหล่สำหรับนักกายภาพบำบัด ในการทำนายระยะเวลาที่ใช้ในการรักษาด้วยรูปแบบการใช้งานของอุปกรณ์การรักษาทางกายภาพบำบัดและการค้นพบองค์ความรู้จากการสร้างกฎความสัมพันธ์เพื่อศึกษาถึงความสัมพันธ์ของจำนวนการใช้อุปกรณ์ในการรักษาทางกายภาพบำบัด ซึ่งในการศึกษาด้วยวิธีการจำแนกประเภทประกอบด้วย 3 วิธี ดังนี้ ทฤษฎีของเบย์ (Bayesian theorem), โครงข่ายประสาทเทียม (Artificial Neural Networks) และ ต้นไม้ตัดสินใจ (Decision Tree) ผลการทดลองที่ดีที่สุดของแต่ละวิธีมีค่าความถูกต้องโครงข่ายประสาทเทียมคือ 92.30% นอกจากนี้ยังนำเสนอวิธีในการค้นหาองค์ความรู้จากวิธีการสร้างกฎความสัมพันธ์เพื่อการศึกษาความสัมพันธ์ในการใช้จำนวนของเครื่องมือสำหรับการรักษาทางกายภาพ ซึ่งจากการทดลองพบว่ามี 5 กฎที่น่าสนใจในการใช้เครื่องมือในการรักษา ซึ่งการศึกษาทั้งสองวิธีคือการจำแนกประเภท (Classification system) และ การสร้างกฎความสัมพันธ์ (Association rules) สามารถประยุกต์ใช้ในการสนับสนุนการตัดสินใจในการเลือกเทคนิคในการรักษาผู้ป่วยที่มีอาการปวดไหล่ได้

Copyright by Mahidol University

44 หน้า

# CONTENTS


| | Page |
|---|---|
| **ACKNOWLEDGEMENTS** | **iii** |
| **ABSTRACT (ENGLISH)** | **iv** |
| **ABSTRACT (THAI)** | **v** |
| **LIST OF TABLES** | **viii** |
| **LIST OF FIGURES** | **ix** |
| **CHAPTER I INTRODUCTION** | **1** |
| 1.1 Background and Statement of Problems | 1 |
| 1.2 Objective of Study | 2 |
| 1.3 Scope of Work | 3 |
| 1.4 Expected Results | 3 |
| **CHAPTER II LITERATURE REVIEW** | **4** |
| 2.1 Shoulder Pain | 5 |
| 2.2 Current process of the physical therapy | 8 |
| 2.3 Knowledge Discovery from very large Database (KDD) | 11 |
| 2.4 Data Mining | 14 |
| 2.5 Related works | 18 |
| **CHAPTER III RESEARCH METHODOLOGY** | **21** |
| 3.1 Data preparation and Processing | 21 |
| 3.2 Preliminary experiment | 26 |
| 3.3 Application mining | 28 |
| 3.4 Knowledge presentation | 32 |
| 3.5 Research Schedule | 33 |
| **CHAPTER IV EXPERIMENTAL RESULTS AND DISCUSSION** | **34** |
| 4.1 Experimental results of classification system | 34 |
| 4.2 Experimental results of Association Rule | 37 |
| 4.3 Knowledge Presentation | 38 |


Copyright by Mahidol University

# CONTENTS (cont.)


| | Page |
|---|---|
| **CHAPTER V CONCLUSION** | **39** |
| 5.1 Conclusion | 39 |
| 5.2 Limitation | 40 |
| 5.3 Future works | 40 |
| **REFERENCES** | **41** |
| **BIOGRAPHY** | **44** |


Copyright by Mahidol University

# LIST OF TABLES


| Table | | Page |
|---|---|---|
| 3.1 | The example input data | 25 |
| 3.2 | The definitions of equipment | 25 |
| 3.3 | Equation and parameter for pattern simulation | 28 |
| 3.4 | The example of data used in ANN experiment | 31 |
| 3.5 | Thesis research schedule | 33 |


Copyright by Mahidol University

# LIST OF FIGURES


| Figure | | Page |
|---|---|---|
| 2.1 | Step of patient services | 9 |
| 2.2 | Physiotherapists work flow | 10 |
| 2.3 | An over view of the steps that compose the KDD process | 12 |
| 2.4 | The Data Information Knowledge | 13 |
| 3.1 | The patient medical record form | 22 |
| 3.2 | The physical therapy treatment medical record form | 23 |
| 3.3 | The Neural Network structure | 27 |


Copyright by Mahidol University

# CHAPTER I
# INTRODUCTION

## 1.1 Background and Statement of Problems

Rehabilitation work was done to humans on the assessment, diagnosis and treatment of disorders, including the promotion of physical and mental health by Physical therapy. In addition, physical therapy will work to promote the health and quality of life of the general public and prevent deficiency disorders, limitations and complications associated with movement in the physical environment [1]. However, these diseases are caused by pathology of bone that occurs by increasing age. Unfortunately, physical therapy can alleviate the symptoms, but cannot cure the symptoms. Among of these diseases, shoulder pain is caused by the wrong posture or position, such as the over-weighted carried bag, using the hand or arm on one side too frequent. Shoulder pain is found in people age 40-60 years and females rather than males. In addition, certain types of diseases of the internal organs can be the cause of shoulder pain, such as heart disease, gallbladder disease and liver disease [2]. The patient still have a normal living as a parasite quo. Important function of the shoulders is to control the arms or hands placed in various positions, and impacts on daily life [3]. The aim of the treatment of shoulder pain is to make the shoulder range of motion increase and patients to be able to live normally. However, because of ethical constraints, there is not enough data to compare the effects of various treatments with no treatment, but letting nature take its course through natural progression [4]. This research gathers medical information from the physical therapy treatment. The purpose of the physiotherapy aims to make patients feel more comfortable, increasing their range of motion, so the patients can use the shoulder as usual.

The details of physical therapy tool and special healing techniques are as follows. Physical therapy equipment is the equipment to maintain the theoretical science, such as ultrasound, electrical stimulation, short wave diathermy, laser, hot pack and cold pack, Transcutaneous Electrical Nerve Stimulation (TENS) physical

Copyright by Mahidol University

therapy intervention such as manual therapy, massage, mobilization. Physical therapy and exercise treatment is the rehabilitation, which protects and promotes the physical movement for the purpose of development to enhance flexibility, strength and maintaining a body [5]. Since there are a variety of treatment techniques, the physiotherapists could select how to use several techniques to treat patients. The criterion depends on the physiotherapist's experience, knowledge, and understanding in deciding treatment techniques.

Data mining, as known as the Knowledge discovery in a large database, would be appropriate in the prediction of the number of equipment in the physical therapy from the equipment usage pattern in shoulder pain patients. From previously reviewed related researches in data mining with other diseases, they have not yet studied and analyzed to predict the equipment usage amount from equipment usage of shoulder pain patients.

Therefore, this research proposes the data mining in two individual aspect, classification system to predict the recovery time and predict the equipment usage amount and the another technical is association rules discovery to study the equipment usage to ensure that patients receive effective treatment from the physical therapy of Department techniques of Rehabilitation Medicine, Faculty of Medicine Siriraj Hospital. Treatment can be used and selected for the appropriate treatment decisions. The researcher has predicted the amount of equipment the treatment of shoulder pain patients with no underlying disease and not caused by accident.

## 1.2 Objective of Study

1) To construct the classifier in equipment usage pattern to predict the amount of equipment in the physical therapy.

2) To discover the association rules within equipment usage by Apriori Algorithm to study the association equipment usage of physical therapy equipment.

Copyright by Mahidol University

## 1.3 Scope of Work

The shoulder pain patient information is gathered and selected from the medical records of Department of Rehabilitation, Faculty of Medicine Siriraj Hospital between October 2012 to September 2013.

## 1.4 Expected Results

The classification model and association rules of physical therapy equipment usage patterns could possibly help decision support for physical therapists in the treatment of shoulder pain patients.

Copyright by Mahidol University

# CHAPTER II
# LITERATURE REVIEW

This research studied the principle and application of data mining in shoulder pain patients with no underlying disease and not caused by accident. The researchers conducted a study and research of theories and research related documents. The summaries are as follows.

2.1 Definitions Shoulder Pain

2.2 Current Process of Physical therapy

2.3 Knowledge Discovery from very large Database (KDD)

2.4 Evaluation

2.5 Related works

## 2.1 Shoulder Pain

Shoulder pains are caused by the wrong posture or poor hygiene. They usually occur after a shoulder injury or after using the shoulders for a long time. The resulting range of motion decreased significantly in both active and passive (active and passive motion). This condition is called the adhesive causalities which has the same meaning as frozen shoulder, but is used to convey the meaning. Pathology found in the condition is better than that is a chronic inflammation of the bladder due to ankle. Winter makes a bold Retractable leash and attached to the humorous [5]. The incidence of the shoulder in the general population 2% were common in the age range 35 to 60 years old and found more in women than men.

### 2.1.1 Cause of Shoulder pain

Copyright by Mahidol University

There are nine possible causes of Shoulder pain as follows:

1) The wrong posture or over-using the hand or arm on one side is given as: writing or using computer for long a time,. Wearing a high heels shoe, etc.

2) Psychological stress: e.g., muscle spasms in the neck and shoulders.

3) Injure in shoulder area.

4) Cold weather accumulates to obstruct the circulation of the blood, muscles, and tendons around the shoulder have not been fed adequately causing pain in the shoulder.

5) The deterioration of the body likes natural bone, muscle, and ligaments.

6) Injury accidents cause broken bones, torn muscle, torn ligament or joint dislocation.

7) Shoulder arthritis is an inflammatory joint diseases, such as rheumatoid and gout.

8) Tendinitis and calcification on x-ray will also see a loaf of white limestone around the shoulder joint.

9) Shoulder pain as a result of symptoms from inflammation of the area, such as bone degeneration, Gallbladder disease, heart disease, liver cancer, etc [6].

**2.1.2 Classification of Shoulder pain can be classified into three types with the following points**

Shoulder pain can be classified into three types [6] as follows:

1) Acute shoulder pain Inflammation within the shoulder muscles and arms will be more pain.

2) Chronic shoulder pain includes disorders of adhesive capsulitis, shoulder instability and shoulder arthritis and found in people of working age.

3) Inflammation of the shoulder joint. Common in those aged over 50 years most of whom have chronic shoulder pain usually caused by degeneration of the body. Shoulder pain may range from minor aches and being more aggressive at the nape. Shoulder and upper arm it may also be numbness in the fingertips to the shoulder. The pain is more severe. If not treated muscles will atrophy and become membrane then will be a shoulder frozen.

Copyright by Mahidol University

**2.1.3 Signs and symptoms of Shoulder Pain**

The Shoulder pain Signs and symptoms are as follows.

1) Pain in the shoulder joint. While the patient is moving the pain will be greater. The pain is more severe at night. After about 4 months the pain is reduced.

2) There is limited movement of the shoulder joint. Which the patient cannot raise his arms as usual. Movement is restricted to add more after about 4 months of limited mobility is increasing every year. Degree of movement

3) The shoulder joint cannot move freely in all directions.

4) Muscles surrounding the shoulder weakness and muscle atrophy especially are muscle.

**2.1.4 Principles Examination of Shoulder Pain**

The doctor will perform a range of motion of the shoulder joint. Medical Imaging is the key to confirm the diagnosis of shoulder pain. Principle examinations are including subjective examination follow as history and characteristics of pain. Objective examinations include the muscles contraction test and motion observations. After the history and physical examination the physical approach to diagnosis for physical therapy and choose the physical therapy techniques.

**2.1.5 Treatment of Shoulder Pain**

The treatment aims to reduce pain. Increase a range of motion acting normally. The treatments of shoulder pain are operative treatment and conservative treatment.

Most treatments are physical therapy and exercise which do not cause injury. Surgical treatment is the alternative treatment in case of physical therapy methods ineffective after treatment for several months. Treatment by non-surgical methods is many ways as take the effective pain suppressing compounds and inflammation, an injected into the shoulder joint and physical therapy [7].

The main purpose of physiotherapy treatment with the following points:

- Decreased painful,
- Increased range of motion,
- Increased muscle strength,
- To use the shoulder as usual,

Copyright by Mahidol University

1) Treatment of Selection in Physical Therapy:

There are many types of Physical Therapy treatment techniques used in the treatment. Have to evaluate the treatment response of patients to treatment and Physical Therapy methods can be divided into 3 groups as follows:

2) Therapy by exercise and movement to treat the inadequacies of the muscle, joints and tissues with the design posture and exercise program: Passive exercise shoulder motion in patients, Pendulum exercise etc. A Survey on techniques of physical therapy was found physical exercise is treatment techniques are used. This is consistent with research that found exercise for treatment is the best method for treat in shoulder pain patient [8].

3) Treatment with 17 types of medical equipment is defined as physical therapy tools by Ministry of Public Health are given as: ultrasound, short wave diathermy, Electrical stimulation, Interferential current, transcutaneous electrical nerve stimulation (TENS), etc.

4) Massage therapy and Mobilization: The selection of treatment depends on the patient symptoms and deciding treatment techniques of physiotherapists. However the research in treatment for patients with shoulder pain has not concluded the most effective treatment techniques [9].

**2.1.6 Treatment period**

Physical therapy treatment doesn't have standard for treatments. It will ingin a week or months depends on the severity of the disease.

## 2.2 Current process of the physical therapy

**2.2.1 Scope of physical therapy Services**

Physical therapy provides the physical therapy services analysis plan for rehabilitation care, surgical, oncology, bones, and muscles.

**2.2.2 Scope of Physical therapist**

Copyright by Mahidol University

1) Examination to determine the cause and factors of the disorder.

Especially in terms of mobility, pain and other disorders.

2) Treatment of various diseases by physical therapy method, such as massage, heat therapy, cold therapy, electrical therapy, etc.

3) Rehabilitation after surgery, injury or disease. The patient can be back to normal life by fix the shortcoming or disabilities of the body parts.

4) Heath promotion.

5) Prevent disease and symptoms and complications.

**2.2.3 Conditions of treatment by physiotherapy**

The conditions of treatment by physiotherapy includes the conditions which cause by pain and movement disorders, broken bones, fracture bone and ligament or muscle tear. The inadequacies of the brain and spinal cord which muscles. The inadequacies of respiratory and lung diseases cause of fatigue, chest pain, shortness of breath.

Copyright by Mahidol University

### 2.2.4 Show step of patient services

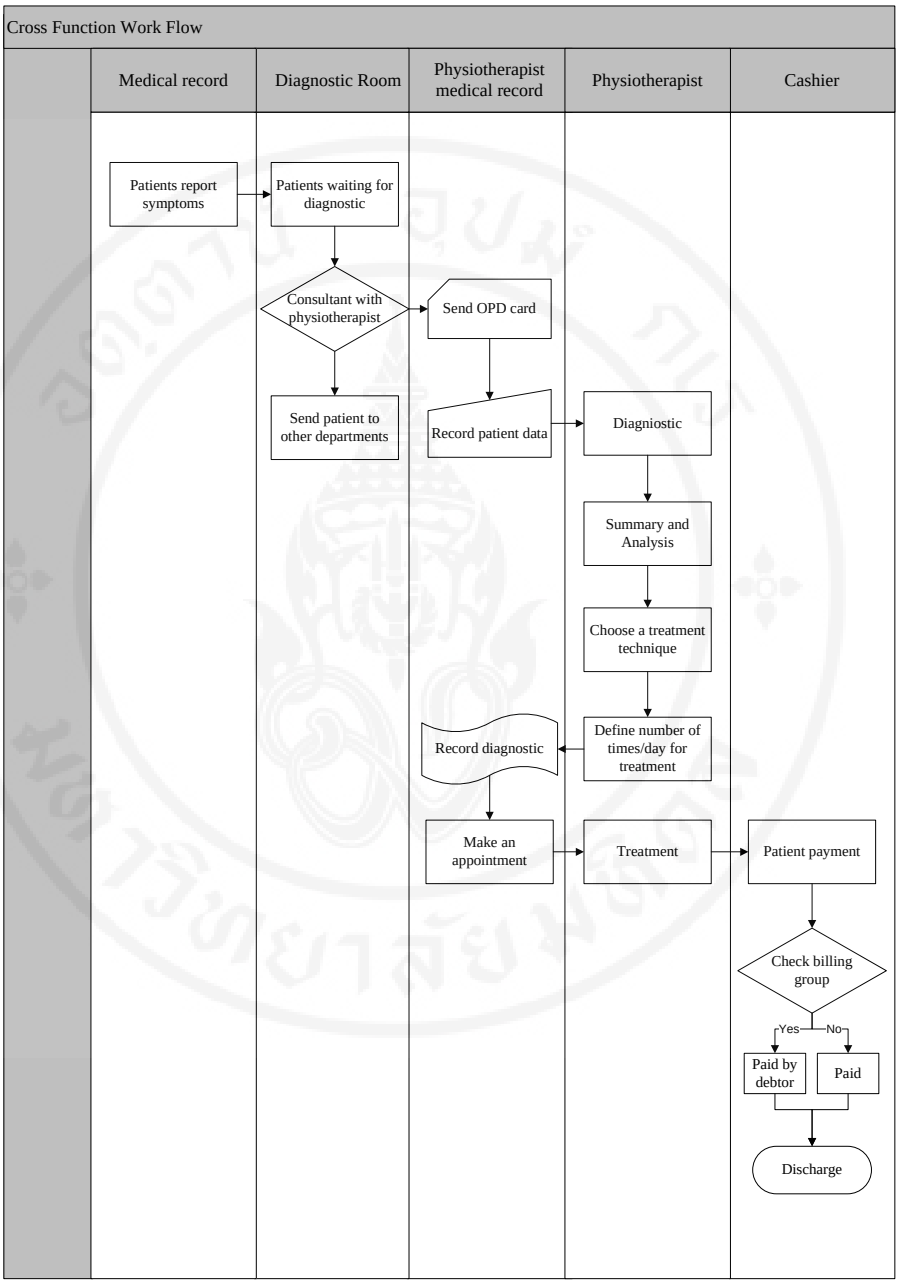



**Figure 2.1** Step of patient services

As shown in Figure 2.1 can be described the step of patient services are as follows.

Copyright by Mahidol University

1) The patient registers at medical record department to identify symptoms, and the hospital staffs send the patient into the examination room.

2) In case of consult with physical therapist, the patient are in the medical rehabilitation examination room. The hospital staff will send the OPD card to the medical record of medical Rehabilitation Department.

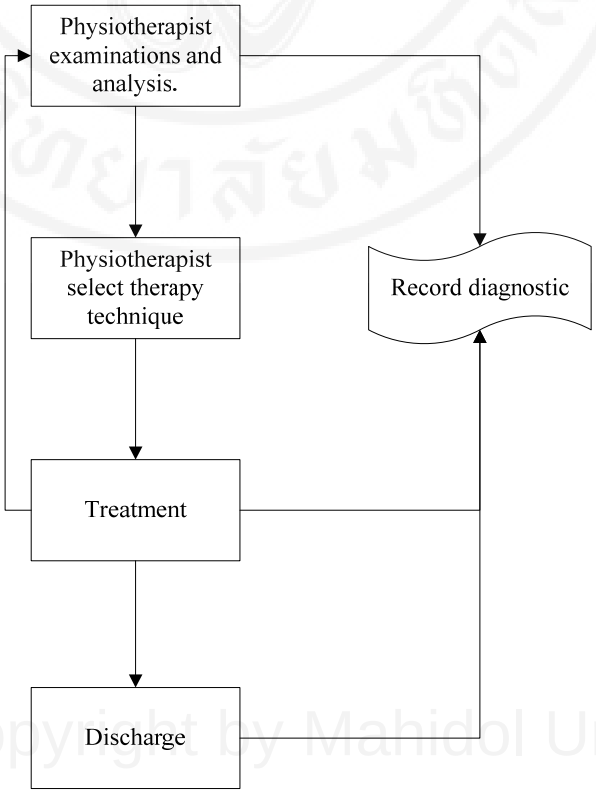
3) The patient history will be recorded to the OPD card by physical therapy medical record officer.

4) Physical examination was performed Analysis Choose a treatment techniques, defined the date and time for treatment, and make an appointment for the next treatment. Finally, the patient is treated by physiotherapist.

5) The patient contacts the cashier to make a payment for treatments.

**2.2.5 Study physiotherapy systems**

Study the physiotherapy work process in Department of Rehabilitation, Faculty of Medicine Siriraj Hospital. This research study for analysis of treatment techniques are related to the work of physiotherapists. The Physiotherapists work flows are shown in Figure 3.2:




Copyright by Mahidol University

**Figure 2.2** Physiotherapists work flow.

The steps of patient services are as follows:

1) Physiotherapist examinations and analysis of symptoms.

2) Physiotherapist planning and select the treatment techniques by knowledge and understanding their experience.

3) The patient was treated until the next appointment. Physiotherapist examination of the symptoms and treatment until the patient has recovered or close to normal.

4) Record patient diagnostics, problem analysis, physical therapy techniques and the reason for discharge in OPD card.

5) Discharge or send the other department depends on patient symptoms.

**2.2.6 Requirement of physiotherapist**

Physiotherapist planning and select the treatment techniques by their knowledge, understanding and experience. From an interview the physiotherapist found need to know what factors affecting the treatment of shoulder pain patient and want to know an appropriate treatment technique to assist in decision making in the treatment of patients with shoulder pain. The objective of this research is to predict the number of equipment in the physical therapy for the shoulder pain patient.

After studying to predict the number of equipment in the physical therapy and the physiotherapist will be gain the information to analyze and select an appropriate treatment technique to decision support in the treatment of patients with shoulder pain.

## 2.3 Knowledge Discovery from very large Database (KDD)

Knowledge discovery from very large database is the process consisting of a lot of information to approach discovers patterns, and relationships are hidden in the database by using the principles of statistical machine learning, pattern recognition, and mathematical analysis. Its aim is to facilitate and effective handling a large database useful and creative solution in the future.

Copyright by Mahidol University

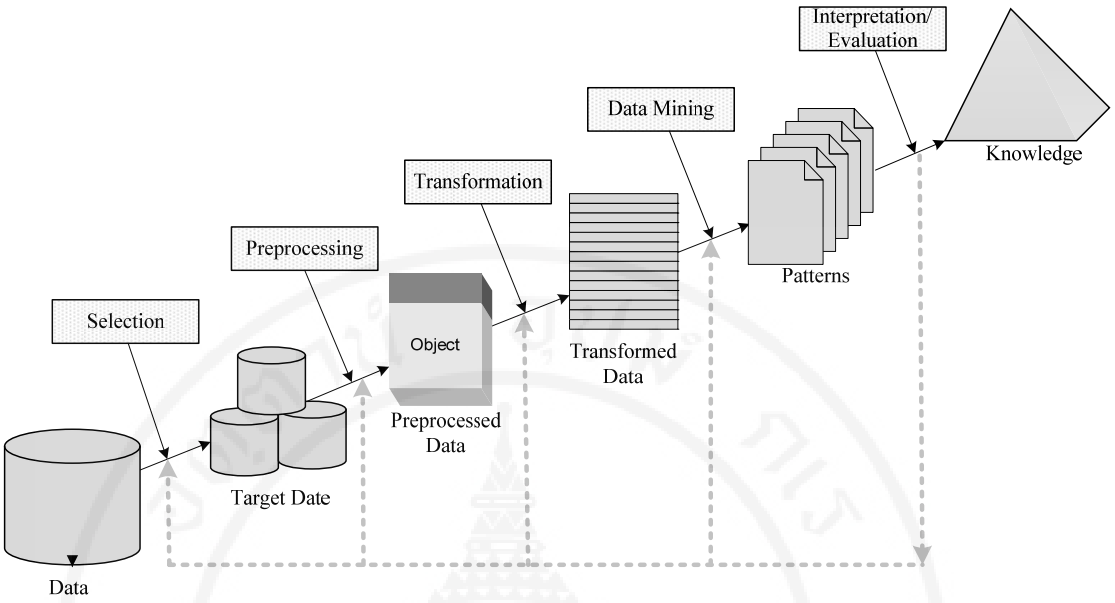



**Figure 2.3** An overview of the steps that compose the KDD process [10].

The KDD process, as shown in Figure 2.4, can be summarized as follows.

1) Data selection: the data includes the relevant prior knowledge and the goals of the application.

2) Data cleaning and preprocessing: includes basic operations, such as removing outliers if appropriate, collecting the necessary information to model.

3) Data reduction and projection: includes finding useful features to represent the data, depending on the goal of the task, and using dimensionality reduction or transformation methods to reduce the effective number of variables under consideration or to find invariant representations for the data.

4) Data mining: includes searching for patterns of interest in a particular representational form or a set of such representations, including classification rules or trees, regression, clustering, which can be divided into 2 categories as:

4.1) Predictive data mining is the predictions about the nature or value of the information that will be formed by use of the database in the past. Supervised learning as the classification and prediction.

The classification and predictive has created the classifier and the differentiate amount between the groups to predict values which were not known before.

Copyright by Mahidol University

4.2) Descriptive data mining is a model to explain certain aspects of the information available. Unsupervised Learning is the most of which would be characteristics of segmenting the data and generating association rules.

5) Interpretation: includes interpreting the discovered patterns and possibly returning to any of the previous steps [11].

### 2.3.1 Data-Information-Knowledge

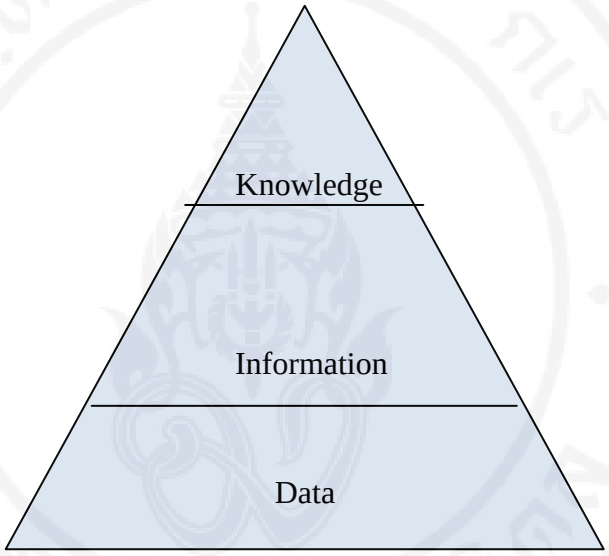



**Figure 2.4** The Data Information Knowledge.

The data used for decision support systems with 3 following points as data, Information and knowledge. There are details as follows:

- Data: collecting facts together without any processing such as patient details and treatment details etc.

- Information: Data collection, storage, processing and management to provide an accurate date and can be used immediately such as frequency of treatment each month.

- Knowledge: Information that can be used to solve the problem or guide for decision to be operational. This research used the knowledge to learn about selecting information used in the solution.

Copyright by Mahidol University

## 2.4 Data Mining

### 2.4.1 Classification Methods

Classification is the process of creating a data management in a defined group. In this research, three classification methods are as follows.

1) Naïve Bayes

The Naïve Bayes algorithm [12] is a classification based on Bayes's rule [13]. The model of classifier data using probability with strong independence assumptions between the features and it is often used as a baseline in text classification, medical diagnosis and systems performance management [14], because it is convenient to implement. Its assumptions make such efficiency possible adversely affect the quality of its results.

Bayes theorem provides a way of calculating the posterior probability, *P(a|b)*, from *P(a)*, *P(b)*, and P*(b|a)*. Naive Bayes classifiers assume that the effect of the value of a predictor (b) on a given class (a) is independent of the values of other predictors. This assumption is called class conditional independence and can be calculated according to the equation, given as:

$$P(a|b) = \frac{P(b|a)P(a)}{P(a)} \quad (2.1)$$

$$P(a|b) = P(b_1|a) \times P(b_2|a) \times \dots \times P(b_n|a) \times P(a) \quad (2.2)$$

Where,

$P(a|b)$ is the posterior probability of class (target) given predictor (attribute);

$P(a)$ is the prior probability of class;

$P(b|a)$ is the likelihood which is the probability of predictor given class;

$P(b)$ is the prior probability of predictor;

2) Decision Tree (C4.5)

Copyright by Mahidol University

C4.5 Decision Tree [15] is based on the task domain as one of classification and the decision support tools those decisions possible consequences by discrete value function which the idea behind decision tree is that it recursively divides

a training set until each division consists entirely or primary of examples from one class [16]. Each non-leaf node of the tree contains a split point, which is a test on one or more attributes and build an initial tree from the training data, such that each leaf node is pure, and then prune the tree to increase its generalization, and the prediction accuracy on test data. The decision tree construction is based on the concept of information entropy [17]. First, create a root node and assign all of the training data set then select the best splitting attribute. It creates the decision node using the selected node and the expected value of splitting. The algorithm recurs on the sublist of the obtained splitting on the selected attribute and adds these nodes as child nodes can be calculated according to the following equation:

$$gain(a) = info(T) - \sum_{i=1}^{s} \frac{|T_i|}{|T|} \times info(T_i) \quad (2.3)$$

The calculation of the information gain of an attribute *a* for *a* set of cases $T$ when *a* is discrete and $T_1, T_2, T_3, \ldots, T_{1s}$ are the subset of $T$ consisting of cases with distinct know values for attribute *a*. The information gain is based on the entropy function can be calculated according to the following equation:

$$info(T) = -\sum_{j=1}^{Noclass} \frac{freq(C_j,T)}{|T|} \times log_2\left(\frac{freq(C_j,T)}{|T|}\right) \quad (2.4)$$

The function $freq(C_j,T)$ calculates the frequency that case $C$ is classified as $T_j$. C4.5 uses on the information gain ratio to determine the split of data can be calculated according to the following equation:

$$Split(T) = -\sum_{j=1}^{s} \frac{|T_j|}{|T|} \times log_2\left(P\left(\frac{|T_j|}{|T|}\right)\right) \quad (2.5)$$

3) Artificial Neural Network

Copyright by Mahidol University

Artificial Neural Network (ANN) [18] is a set of methods that emulate the way the human brain works. The basic processing unit is a neuron. Multi neurons are

grouped into layers and linked together. In Neural Network, the knowledge is stored in the weight associations of neural networks, most business application are handled using this algorithm [19]. In the general ANN, are suitable for inputs are categorical and numeric and the relationship between inputs and outputs are not linear or the input data are not normally distributed. There are three steps in process of learning in an ANN. First, compute temporary outputs base on inputs and random weights than compute outputs with desired targets and adjust the weights and repeat the process. The general mathematic definition can be calculated according to the following equation:

$$y(x) = g(\sum_{i=1}^{N} u_i x_i) \quad (2.6)$$

Where,

$x$ is a neural with $n$ input dendrites $(x_1, x_2, x_3, \ldots, x_n)$;

$y(x)$ is one output axon;

$u_i$ is weights determining how much the input should be weighted $(u_1, u_2, u_3, \ldots u_n)$.

**2.4.2 Classifier evaluation methods.**

The evaluation process of classification system, this research selects the 10-fold cross validation method [20]. The 10-fold cross-validation is the measure the performance of the model to predict the sample. The base of this technique is re-sampling and takes one-tenth of data from data set as test data and test results from the prediction of the classification model which is constructed by the rest of the data can be calculated according to the following equation:

$$Average\ error = \left(\frac{E_1 + E_2 + E_3 + \cdots + E_k}{k}\right) \quad (2.7)$$

Copyright by Mahidol University

According to the association rule model, this research will focus on and be evaluated by the confidence value and consistency in real usage which will be described in the next subsection.

### 2.4.3 Association Rules Discovery by Apriori Algorithm

Mining association rule is to discover a set of rules that shared large percentage in the data [21]. A goal is to find the relation regulation to shows correlated events or to predict the sequence of events that may occur. Association rule discovery method [21] aims to find an interesting correlation between variable from a large database that be used to analyze shoulder pain patient data. Apriori algorithm is the algorithm was developed in data mining by creating an association rules in the process of finding the frequent item sets by the algorithm often used as an algorithmic basis for comparing results obtained with other algorithm. Among of the association rule technique, Apriori algorithm [17] was used to create the association rule and it is an algorithm for frequent item set mining by generation and test approach over transactional databases. It proceeds frequent item set and extended item set as long as those item sets appear sufficiently often in the database to the next level by creating candidate item sets (n + 1) from joining the frequent N-item set. This will make the cutting if the candidate item set is not frequent item set. To establish association rule are important principles in finding the frequent item sets with the frequency or the support is greater than or equal to the minimum frequency by the rule X => Y holds with support s if s% of transactions in D contain X ∪ Y. Rules that a s greater than a user-specified support said to have minimum support.

### 2.4.4 K-Means clustering method

The cluster is a technique to reduce the size of the data gathered in the group did not know before. The data characteristics can be helped the classification using the meter to tell the same or different data. An advantage of cluster was reducing the scope for a large database consisting of a series of similar groups to come together and the records that were similar to the same group makes it easy to compare and analyze data.

The K-Means algorithm (K stands for the predetermined number of clusters) for partitioning set as X of *n* points where each clusters center is represented by the mean value of the object in the cluster. This process of assigning data point and readjusting centers is repeated until it stabilizes [22]. A formal definition of the algorithm is as follows.

Copyright by Mahidol University

a. Arbitrarily generate *k* random points as initial cluster centers $C_1, C_2, \ldots, C_k$.

b. For each $1 \leq i \leq k$, assign set the cluster centers $C_i$ be the set of points in X that that are closer to $C_i$ than to any $C_j$ with $j \neq i$.

For each $1 \leq i \leq k$, set $C_i = \frac{1}{|C_i|} \sum_{x \in C_i} x$,

c. Recompute the new cluster centers. The partition of X is the set of clusters $C_1, C_2, \ldots, C_k$.

## 2.5 Related works

Tam et al. [23] proposed prediction of the osteoarthritic knee rehabilitation outcome by using the prediction model by the Neural Network. The aims of their researches are to develop a prediction framework and to forecast patient performance for the benefit of decision support for the physical therapist to choose the appropriate treatment techniques and prediction the efficiency of the treatment. Considered the effective treatments for the patient with osteoarthritic knee, and found difference degrees of improvement with the three difference treatment protocols. The Transcutaneous Electrical Nerve Stimulation exercise (TENS) were used with the patient treatment. The patients with the best improvements according to their clinical record are selected and applied the Neural Network method by difference treatment protocols. Found the Spearman's rho was 0.424, which the statistically significant at $p < 0.001$. Their preliminary analysis found that the newly developed prediction protocol can assist the clinical decisions. The result of the research shows TENS, and exercise treatments are directly impacting the results of treatment.

Sukjai [5] proposed analysis of patients with frozen shoulder from the medical record by clustering the range of motion in frozen shoulder patient and association rule for prediction of factors that affected treatment. The study was found that the disease duration, age, range of motion, pain levels, and hands crossed behind levels are the influential factors for the treatment of frozen shoulder patients. There are the satisfactory results with three rules. The first rule was that the range of motion 90-

Copyright by Mahidol University

120 degree should be selected the techniques for treatment following hot pack on neck area, ultrasound in the front of shoulder, ultrasound in infraspinatus muscle and electrical stimulation in the front and back of shoulder. The second rule showed that the range of motion 120- 150 degree should be selected the techniques for treatment following hot pack on neck area, ultrasound in infraspinatus muscle, electrical stimulation in the front and back of shoulder, and electrical stimulation on infraspinatus-rhomboid muscle. The last rule was that the range of motion 150- 180 degree should be selected techniques for treatment following electrical stimulation in the front and back of shoulder. The rules have made a physical approach to deciding techniques to treat patients with frozen shoulder. Anyway, this work is different from the domain of our work in the terms of the types of diseases that occurred with patient. These bodies of research propose how to treat shoulder pain patient in the beginning period before frozen shoulder hazard.

Balgen et al. [24] which proposed the physical therapy and drug therapy techniques in frozen shoulder patient. Each patient had learned pendulum exercise and randomized allocated to one of four treatment types as intra-articular steroids, mobilizations, ice therapy and no treatments. The study consisted of 42 patients with previously untreated frozen shoulder completed this study. Conclusions for inclusion were at least one month of disease duration with sleep disturbance due to night pain. All active and passive of a range of motion acting are restricted. The ranges of age were between 44-74 years old and the duration of the condition at presentation was 1-12 months. The treatments were randomly allocated to one of four treatment group as 1. Steroid group are 11 patients 2. Mobilization group are 11 patients 3. Ice group are 12 patients and 4. Non-treatment group 8 patients. The clinical results shown 39 patients reported excellent general health and were proposed a little effect on long-term advantage in any of treatment regimens, but they found the steroids injections are the most effective pain relief and increasing the range of motion in the early stage of the condition. The result of the research shows the physical therapy had a little effect for the recovery duration. It may be improved if the physical therapy equipment for patient treatment is included for the consideration.

Copyright by Mahidol University

Heijden et al. [25] proposed the treatment with electrotherapy and ultrasound. Their study is controlled trail with two by two factorial design plus an

additional control group in 17 physiotherapy practices. Patients are entered to study with shoulder pain or restricted shoulder mobility, a soft tissue impairment without underlying specific condition were controlled, if the patient had not recovered in six sessions of exercise therapy in two weeks. Additionally, the patient received a maximum 12 sessions of exercise therapy in six weeks. Their outcome measurements are recovery, functional status, chief complaint, pain, clinical status, and range of motion. The result found treatment with interference and ultrasound ineffective in the treatment when combined with exercise therapy in frozen shoulder patient. Interestingly, other previously reviewed researches had not taken a look of the usage of equipment for consideration in any aspects. Therefore, this mentioned paper motivates our work of proposing the data mining applications in shoulder pain patients according to usage of physical therapy equipment in two approaches: recovery duration prediction from equipment usage pattern and association between their amounts.

Interestingly from the research mentioned above, other previously reviewed research had not taken a look of the usage of equipment for consideration in any aspects. Therefore, this mentioned paper motivates our work of proposing the data mining applications in shoulder pain patients according to usage of physical therapy equipment in two approaches: recovery duration prediction from equipment usage pattern and association between their amounts.

Copyright by Mahidol University

# CHAPTER III
# RESEARCH METHODOLOGY

This chapter presents the data mining to study the equipment usage to ensure that patients will receiv the effective treatment from the physical therapist to treating patients with shoulder pain in Department of Rehabilitation, Faculty of Medicine Siriraj Hospital. The guidelines and procedures details are as follows.

3.1 Data preparation and Processing,

3.1.1 Data collection,

3.1.2 Data selection,

3.1.3 Data pre-processing and transformation,

3.2 Preliminary experiment,

3.3 Applying data mining,

3.3.1 Classification in equipment usage pattern,

3.3.2 Association rules discovery,

3.4 Knowledge presentation,

3.5 Research schedule.

## 3.1 Data preparation and Processing

The Data preparation and processing consist of 3 sub-steps which are as follows.

3.1.1 Data collection,

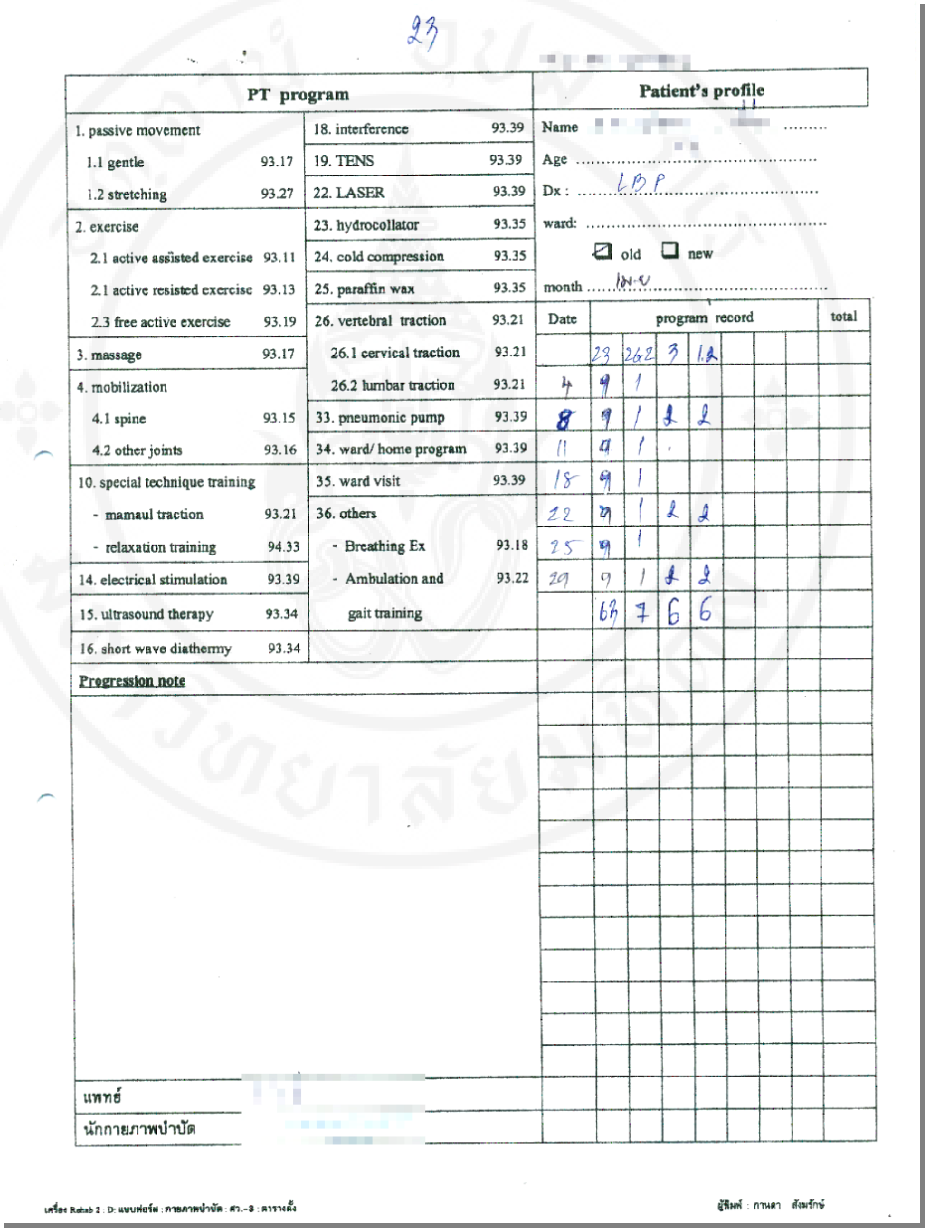
3.1.2 Data selection,

3.1.3 Data pre-processing and transformation.

Copyright by Mahidol University

### 3.1.1 Data collection

Data collections form of the Department of Rehabilitation, Faculty of Medicine Siriraj hospital 105 cases during 2012-2013 from 2 documents consist of the patient medical record form and the physical therapy treatment medical record.



| PT program | | | | Patient's profile |
|---|---|---|---|---|
| 1. passive movement | | 18. interference | 93.39 | Name ..... |
| 1.1 gentle | 93.17 | 19. TENS | 93.39 | Age ..... |
| 1.2 stretching | 93.27 | 22. LASER | 93.39 | Dx : LBP |
| 2. exercise | | 23. hydrocollator | 93.35 | ward: ..... |
| 2.1 active assisted exercise | 93.11 | 24. cold compression | 93.35 | ☑ old ☐ new |
| 2.1 active resisted exercise | 93.13 | 25. paraffin wax | 93.35 | month ..... |
| 2.3 free active exercise | 93.19 | 26. vertebral traction | 93.21 | |
| 3. massage | 93.17 | 26.1 cervical traction | 93.21 | |
| 4. mobilization | | 26.2 lumbar traction | 93.21 | |
| 4.1 spine | 93.15 | 33. pneumonic pump | 93.39 | |
| 4.2 other joints | 93.16 | 34. ward/ home program | 93.39 | |
| 10. special technique training | | 35. ward visit | 93.39 | |
| - mamaul traction | 93.21 | 36. others | | |
| - relaxation training | 94.33 | - Breathing Ex | 93.18 | |
| 14. electrical stimulation | 93.39 | - Ambulation and gait training | 93.22 | |
| 15. ultrasound therapy | 93.34 | | | |
| 16. short wave diathermy | 93.34 | | | |

| Date | program record | | | | | | | | total |
|---|---|---|---|---|---|---|---|---|---|
| | 23 | 262 | 3 | 12 | | | | | |
| 4 | 9 | 1 | | | | | | | |
| 8 | 9 | 1 | 2 | 2 | | | | | |
| 11 | 9 | 1 | | | | | | | |
| 18 | 9 | 1 | | | | | | | |
| 22 | 9 | 1 | 2 | 2 | | | | | |
| 25 | 9 | 1 | | | | | | | |
| 29 | 9 | 1 | 2 | 2 | | | | | |
| | 69 | 7 | 6 | 6 | | | | | |

Progression note

แพทย์

นักกายภาพบำบัด

เครื่อง Rehab 2 : D: แบบฟอร์ม : กายภาพบำบัด : ศว.-3 : ตารางตั้ง ผู้พิมพ์ : กานดา สังฆรักษ์

**Figure 3.1** The patient medical record form.

Copyright by Mahidol University

แบบบันทึกการรักษาทางกายภาพบำบัดของผู้ป่วย ศว. 3 ภาควิชาเวชศาสตร์ฟื้นฟู

ผู้ป่วยชื่อ.....นามสกุล.....อายุ.....ปี

โรค.....HN.....แพทย์เจ้าของไข้.....

ข้อควรระวัง.....

วัดนัดแพทย์.....วันที่.....

โปรแกรมการรักษา.....

กายภาพบำบัดระบบกระดูกและกล้ามเนื้อผู้ใหญ่

ตึกศรีสังวาลย์ ชั้น 3

โทร (02) 419 - 8959

**Figure 3.2** The physical therapy treatment medical record form.

The details of data from patient medical record form are as follows.

Patient details;

1) Name,

2) Age,

3) Sex,

4) Ward (In case of IPD),

5) Admission date,

6) Discharge date,

7) Physician,

8) Physiotherapist,

9) Physical examination information,

Treatment Techniques: There are 20 tools in physical therapy treatment as follows.

1) Gentle passive movement,

2) Stretching passive movement,

3) Active assisted exercise,

4) Active resisted exercise,

Copyright by Mahidol University

5) Free active exercise,

6) Massage,

7) Mobilize spine,

8) Mobilize other joints,

9) Electrical stimulation,

10) Ultrasound,

11) Short wave diathermy,

12) Interference,

13) Transcutaneous electrical nerve stimulation,

14) Hydro collator,

15) Cold compression,

16) Paraffin wax,

17) Cervical traction,

18) Lumbar traction,

19) Pneumonic pump,

20) Other.

**3.1.2 Data selection**

The data selected for analysis are the frequency of physical therapy treatment and the physical therapy equipment.

**3.1.3 Data pre-processing and transformation**

The process that ensures the quality of data to be used in the analysis is correct and checks the accuracy of the patient medical record. Collect the necessary information to model such as removing outlier data if appropriate.

The data transformation for classification technique should to be converted the data into a variable numeric such as Treatment information if the technique is used to represent the actual amount and if no selection is used to represent a ‘0’. An example data structure description as shown in Table 3.1.

Copyright by Mahidol University

**Table 3.1** The example input data.

| Gentle Passive Movement | Stretching Passive Movement | Active Assisted Exercise | Active Resisted Exercise | Free Active Exercise | Massage | Mobilize Spine | Mobilize Other Joints | Electrical Stimulation | Ultrasound | Short Wave Diathermy | Interference | TENS | Hydro Collator | Cold Compression | Paraffin Wax | Cervical Traction | Lumbar Traction | Pneumonic Pump | Others | Status |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | R |
| 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 24 | 0 | 0 | 0 | 0 | 0 | 0 | S |
| 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 5 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | S |
| 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 2 | 0 | 0 | 8 | 0 | 0 | 0 | 0 | 0 | 0 | S |
| 0 | 2 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 2 | 2 | 0 | 0 | 16 | 0 | 0 | 0 | 0 | 0 | 0 | S |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 6 | 0 | 0 | 0 | 24 | 0 | 0 | 0 | 0 | 0 | 0 | R |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 10 | 0 | 0 | 0 | 40 | 0 | 0 | 0 | 0 | 0 | 0 | S |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 14 | 0 | 0 | 0 | 56 | 0 | 0 | 0 | 0 | 0 | 0 | S |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | S |
| 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | R |
| 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 21 | 0 | 0 | 0 | 0 | 0 | 0 | S |
| 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 5 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | S |
| 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 2 | 0 | 0 | 8 | 0 | 0 | 0 | 0 | 0 | 0 | S |
| 0 | 2 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 2 | 2 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | S |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | R |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 9 | 0 | 0 | 0 | 10 | 0 | 0 | 0 | 0 | 0 | 0 | S |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 0 | 0 | 12 | 0 | 0 | 0 | 0 | 0 | 0 | S |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | S |

To facilitate the data collection and experiment initials are used instead of full name of the physical therapy treatment as shown in Table 3.2:

**Table 3.2** The definitions of equipment.

| No. | Initials | Description |
|---|---|---|
| 1 | GPM | Gentle Passive Movement |
| 2 | SPM | Stretching Passive Movement |
| 3 | ARE | Active Resisted Exercise |
| 4 | FAE | Free Active Exercise |
| 5 | M | Massage |
| 6 | MS | Mobilize Spine |
| 7 | MOJ | Mobilize Other Joints |
| 8 | ES | Emotional Stress |

Copyright by Mahidol University

**Table 3.2** The definitions of equipment. (cont.)

| No. | Initials | Description |
|---|---|---|
| 9 | U | Ultrasound |
| 10 | SWD | Short Wave Diathermy |
| 11 | I | Interference |
| 12 | TENS | Transcutaneous Electrical Nerve Stimulation |
| 13 | H | Hydro Collator |
| 14 | CC | Cold Compression |
| 15 | PW | Paraffin Wax |
| 16 | CT | Cervical Traction |
| 17 | LT | Lumbar Traction |
| 18 | O | Others |

## 3.2 Preliminary experiment

According to the prediction approach, the authors have initially made a preliminary study and found that the amount of equipment used is in direct correlation to the results of treatment. The preliminary study was performed with the Neural Network given by the total amount of tools used for the treatment. The Neural Network structure is as follows Figure 3.4.

Copyright by Mahidol University

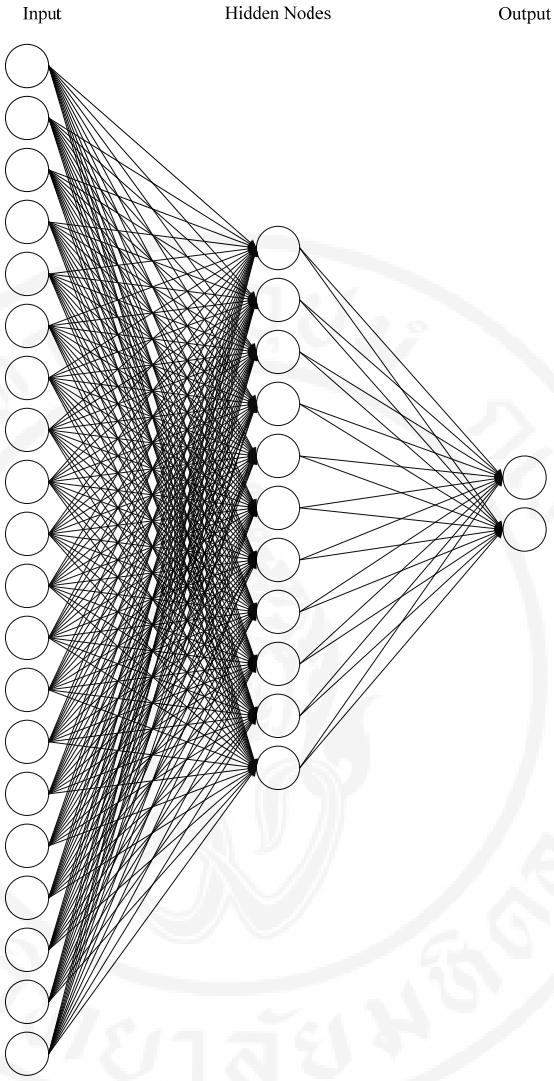



**Figure 3.3** The Neural Network structure.

Unfortunately, the usage amount of each equipment type cannot specify the duration of the treatment of each patient proven by an unsatisfactory root mean squared error of around 45% by number of hidden nodes, learning rate, and momentum are 11, 0.3, and 0.2 respectively. According to the Artificial Neural Network (ANN), our preliminary experiment shown the empirical result had an unsatisfactory root mean squared error of 44.95%. Thus, an extra data transformation procedure is considered. In this research, the study of shoulder pain patient data from the medical records and construct pre-defined patterns of shoulder pain patient. The researcher change the amount of equipment used to create the chart pattern and change the prediction the recovery

Copyright by Mahidol University

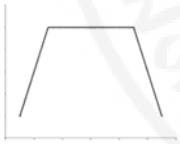
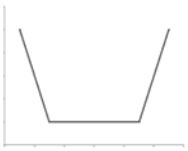
duration to prediction the equipment usage amount. The advantages of predictive equipment could support the decision making in the treatment of shoulder pain patients.

## 3.3 Application mining

### 3.3.1 Classification in equipment usage pattern

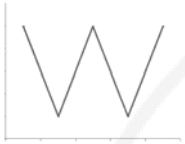
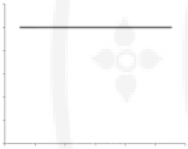
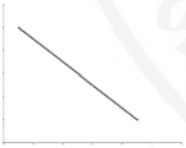
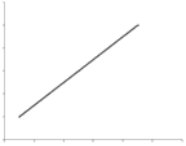
The identification of various shape features from the shoulder pain patient data with seven patterns is shown in Table 3.4. [26]

**Table 3.3** Equation and parameter for pattern simulation.

| Control pattern | Pattern parameters | Pattern equation | Parameter values |
|---|---|---|---|
| A-shape (P1)  | -Mean ($\mu$)<br>-Standard deviation ($\sigma$)<br>-Period ($T$)<br>-Gradient ($g$) | $y_t = \mu + Ti\sigma + ig$<br>$y_t = \mu + Ti\sigma$<br>$y_t = \mu + Ti\sigma + ig$ | $i = [0, a]$<br>$i = [a, b]$<br>$i = [b, n]$ |
| U-shape (P2)  | -Mean ($\mu$)<br>-Standard deviation ($\sigma$)<br>-Period ($T$)<br>-Gradient ($g$) | $y_t = \mu + Ti\sigma + ig$<br>$y_t = \mu + Ti\sigma$<br>$y_t = \mu + Ti\sigma + ig$ | $i = [0, a]$<br>$i = [a, b]$<br>$i = [b, n]$ |
| M-shape (P3) | -Mean ($\mu$)<br>-Standard deviation ($\sigma$)<br>-Period ($T$)<br>-Gradient ($g$)<br>-Amplitude (a)<br>$b - a > \frac{n}{4}$ | $y_t = \mu i + Ti\sigma + a\sin\left(\frac{2\pi i}{T}\right)$<br>$y_t = \mu + Ti\sigma + ig$<br>$y_t = \mu + Ti\sigma + ig$ | $i = [0, n]$<br>$i = [0, a]$<br>$i = [b, n]$ |

Copyright by Mahidol University

**Table 3.3** Equation and parameter for pattern simulation. (cont.)

| Control pattern | Pattern parameters | Pattern equation | Parameter values |
|---|---|---|---|
| W-shape (P4)  | -Mean ($\mu$)<br>-Standard deviation ($\sigma$)<br>-Period ($T$)<br>-Gradient ($g$)<br>-Amplitude (a)<br>$b-a > \frac{n}{4}$ | $y_t = \mu t + T t\sigma + a\sin\left(\frac{2\pi t}{T}\right)$<br>$y_t = \mu + T t\sigma + tg$<br>$y_t = \mu + T t\sigma + tg$ | $t = [0, n]$<br>$t = [0, a]$<br>$t = [b, n]$ |
| Stratification (P5)  | -Mean ($\mu$)<br>-Standard deviation ($\sigma$)<br>-Period ($T$)<br>-Random noise ($\sigma^l$) | $y_t = \mu + T t\sigma^l$ | $\sigma^l = [0.2\sigma, 0.4\sigma]$ |
| Decreasing (P6)  | -Mean ($\mu$)<br>-Standard deviation ($\sigma$)<br>-Period ($T$)<br>-Gradient ($g$) | $y_t = \mu + T t\sigma + tg$ | $g = [-0.1\sigma, -0.$ |
| Increasing (P7)  | -Mean ($\mu$)<br>-Standard deviation ($\sigma$)<br>-Period ($T$)<br>-Gradient ($g$) | $y_t = \mu + T t\sigma + tg$ | $g = [0.05\sigma, 0.1\sigma$ |
| Not used (P0) | - | - | - |

***Note:*** *$i$ = discrete time point at which the pattern is sample ($i$ = 1,2,3,…,$n$), $y_i$= sample value at $i^{th}$ time point.*

Copyright by Mahidol University

The purpose of the classification system is to predict the equipment usage amount by using the physical therapy equipment usage. There are three classification methods, which have been selected, given as first, the Naïve Bayes [27] is a model of

classifier data using probability with strong independence assumptions between the features and it is often used as a baseline because it is convenient to implement. Its assumptions make such efficiency possible but also adversely affect the quality of its results. The second method is C4.5 Decision Tree [16]. It is based on the task domain as one of classification and the decision support tools those decisions possible consequences by discrete value function which the idea behind decision tree is that it recursively divides a training set until each division consists entirely or primary of examples from one class [28]. Another method is Artificial Neural Network (ANN) [28]. It is a set of methods that emulate the way the human brain works. The basic processing unit is a neuron. Multi neurons are grouped into layers and linked together. In Neural Network, the knowledge is stored in the weight associations of neural networks, most business application are handled using this algorithm [18].

To construct the classification system without pattern extraction and dividing of usage range of physical therapy equipment tools was shown the empirical result had an unsatisfactory root, leading us to consider a selecting method to classify attributes with the classification methods of the Naïve Bayes, Decision Tree and ANN with the best tuned parameter. This research is dividing the range with manual method and K-Means clustering method for comparison and use 10-fold cross validation method to evaluation process of classification system. The example of data used in ANN experiment is as follow Table 3.3:

Copyright by Mahidol University

**Table 3.4** The example of data used in ANN experiment.

| GPM | SPM | ARE | FAE | M | MS | MOJ | ES | U | SWD | I | TENS | H | CC | PW | CT | LT | O | Number |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| P0 | P6 | P0 | P0 | P6 | P0 | P0 | P0 | P0 | P0 | P0 | P0 | P0 | P0 | P0 | P0 | P0 | P6 | 3 |
| P0 | P3 | P0 | P0 | P0 | P0 | P0 | P0 | P3 | P0 | P0 | P0 | P3 | P0 | P0 | P0 | P0 | P5 | 4 |
| P0 | P7 | P0 | P0 | P0 | P0 | P0 | P0 | P7 | P0 | P0 | P0 | P0 | P0 | P0 | P0 | P0 | P3 | 3 |
| P0 | P0 | P1 | P0 | P0 | P0 | P0 | P0 | P0 | P0 | P0 | P0 | P1 | P0 | P0 | P0 | P0 | P1 | 3 |
| P0 | P5 | P0 | P0 | P0 | P5 | P5 | P0 | P7 | P3 | P0 | P0 | P7 | P0 | P0 | P0 | P0 | P0 | 6 |
| P0 | P3 | P0 | P0 | P3 | P0 | P0 | P0 | P1 | P0 | P0 | P0 | P3 | P0 | P0 | P0 | P0 | P0 | 4 |
| P0 | P0 | P0 | P0 | P3 | P0 | P0 | P0 | P3 | P0 | P0 | P0 | P1 | P5 | P0 | P0 | P0 | P0 | 4 |
| P5 | P5 | P6 | P0 | P0 | P0 | P0 | P0 | P5 | P0 | P0 | P0 | P7 | P0 | P0 | P0 | P0 | P0 | 5 |
| P0 | P7 | P0 | P0 | P7 | P0 | P0 | P0 | P6 | P0 | P0 | P0 | P6 | P0 | P0 | P0 | P0 | P0 | 4 |
| P0 | P0 | P0 | P0 | P0 | P5 | P0 | P0 | P1 | P0 | P0 | P0 | P3 | P0 | P0 | P0 | P0 | P0 | 3 |
| P5 | P5 | P0 | P0 | P5 | P0 | P5 | P3 | P0 | P0 | P0 | P0 | P0 | P0 | P0 | P0 | P0 | P0 | 5 |

### 3.3.2 Association Rules Discovery

Association rules [28][29] are used to discover relationship between two or more items (or events or concepts. This research studies the relationship of the physical therapy equipment from its usage records to treat the shoulder pain patient by applying the association equipment usage of physical therapy equipment. This

Copyright by Mahidol University

research proposes the association rule with the usage record of physical therapy equipment, and also eliminated missing data from the raw data such the physical equipment tools that were not be used.

## 3.4 Knowledge presentation

The best results of classification system and association rules can be applied to help the decision support for physical therapists in the treatment of shoulder pain patients. The ANN structure enables to predict the amount of equipment in the physical therapy with a high average accuracy. Additionally, the rules extracted by the Association Rule by Apriori algorithm to study the association equipment usage of physical therapy equipment and could be supported the decision making of specialist in different aspect, an association of equipment usage. For example the association rules method, if the physical therapist selecting the Stretching, Passive movement and Massage tend to select Hydro collator to treat the patient. This research may be further considered for the inclusion of the classification system and the association rule altogether. This may provide more effective results in the prediction of treatment, because the classification can predict the recovery status of patients from the usage of the pattern of physical therapy equipment, while the association rules discovery method studies the association amount of the physical therapy equipment used.

Copyright by Mahidol University

## 3.5 Research Schedule

**Table 3.5** Thesis research schedule.

| **Implementation process** | **Period (2014-2015)** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | **Oct.** | **Nov.** | **Dec.** | **Jan.** | **Feb.** | **Mar.** | **Apr.** | **May.** | **Jun.** |
| **1) Educational process data** | ↔ | | | | | | | | |
| **2) Certifies ethical research in humans** | | ↔ | | | | | | | |
| **3) Collect data and Prepare Data** | | | ↔ | | | | | | |
| **4) Process for treatment factors.** | | | | ↔ | | | | | |
| **5) Process for Data Mining** | | | | | ↔ | | | | |
| **6) Process for evaluation** | | | | | | ↔ | | | |
| **7) Process for documentation** | | | | | | | ↔ | ↔ | ↔ |

Copyright by Mahidol University

# CHAPTER IV
# EXPERIMENTAL RESULTS AND DISCUSSION

This chapter presents the results of the experiments in three parts which have the results as follows.

4.1 Experimental results of classification system,

4.2 Experimental results of association rule,

4.3 Discussion.

## 4.1 Experimental results of classification system.

The classification system is to predict the number of equipments usages by using the physical therapy equipment usage.

From Table 4.1, it shows the comparison of the results obtained from the experiments for the accuracy of the proposition. It has been found that the interesting result of ANN method provided the most satisfactory accuracy is in the range 84.62%-91.35%. The accuracy is improved by more than 1-3% on average when the network is applied up to 92.30% of accuracy in case of experimental 5. The experiment shows the experimental 5 with K-Means [27] cluster the result is the most satisfactory.

To provide more details in the experiments, for the comparison of the accuracies in each classification method, it has found that the accuracy of 91.35% of neural network was most satisfactory. After tuning the configuration of ANN in experimental scenario 5, the performance was further improved to 92.30% accuracy. The 83.65% accuracy was the most satisfactory obtained from the Naïve Bayes method. The 81.73% accuracy was the most satisfactory obtained from the decision tree method. Furthermore, the 83.65% accuracy was the most satisfactory obtained after processing Naïve Bayes method. The 81.73% accuracy was the most satisfactory obtained after processed decision tree method, which are as follows in Table 4.1.

Copyright by Mahidol University

**Table 4.1** Comparison of the results obtained from the experiments.

| Experimental | Accuracy | | | | | |
|---|---|---|---|---|---|---|
| | | | Neural Network | | | |
| | Naïve Bayes | Decision Tree | Structure (Input-Hidden node-Output) | Learning Rate | Momentum | Accuracy |
| **Experiment 1 (3 manual cluster) A =1-2 B = 3-4 C = 5-7** | 69.23% | 66.35% | 18-17-3 | 0.5 | 0.5 | 87.50% |
| **Experiment 2 (3 manual cluster) A = 1-3 B = 4-5 C = 6-7** | 65.38% | 72.12% | 18-1-3 | 0.1 | 0.8 | 83.65% |
| **Experiment 3 (3 manual cluster) A = 1-2 B = 3-5 C = 6-7** | 75.00% | 76.92% | 18-19-3 | 0.1 | 0.5 | 85.57% |
| **Experiment 4 (2 manual cluster) A = 1-4 B = 5-7** | 83.65% | 79.81% | 18-13-2 | 0.3 | 0.1 | 90.38% |
| **Experiment 5 (2 K-Means cluster) A = 1-3 B = 4-7** | 77.88% | 81.73% | 18-1-2 | 0.1 | 0.2 | 92.30% |
| **Experiment 6 (3 K-Means cluster) A = 1-3 B = 4 C = 5-7** | 60.58% | 45.19% | 18-13-3 | 0.9 | 0.1 | 79.80% |

Copyright by Mahidol University

From The Table 4.1, it displays the result obtained from the experiment no. 5 which is around to 92.30% accuracy proves the ANN is an effective model for equipment usage amount prediction from equipment usage pattern since its parameters are tunable i.e. number of hidden nodes, learning rate, and momentum. These parameters are appropriately tuned and found their optimums to 1, 0.1, and 0.2 respectively.

Although the 92.30% accuracy of an ANN after tuning the parameter but this research wants to gain the knowledge and can be interpreted. Therefore, the researcher has to propose the decision tree. As the experimental no. 5, the 81.73% accuracy was the most satisfactory obtained from the decision tree method, as shown in Figure 4.1.

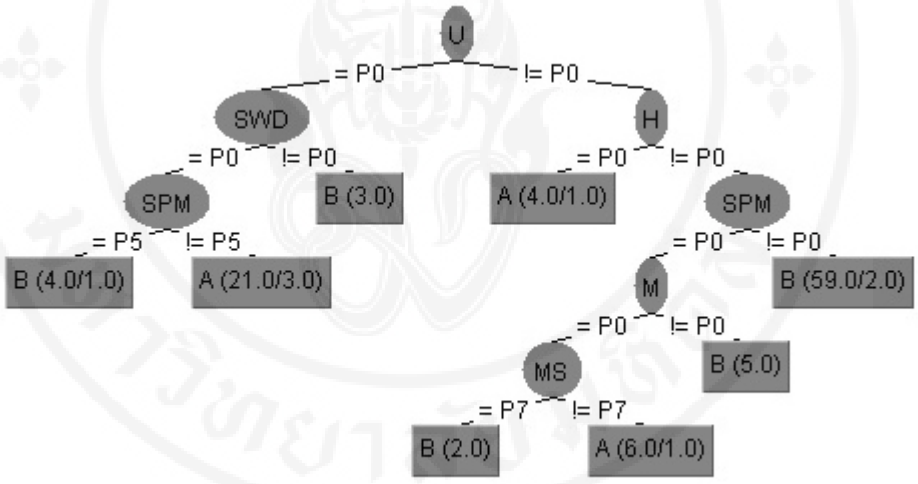



**Figure 4.1** Tree visualizer of experiment 5.

According to the result of the experiment no. 5, the decision tree was constructed by J48. Referring to the tree in Figure 4.1, If Ultrasound was not used, must be used of Short Wave Diathermy (SWD), which if used the Short Wave Diathermy (SWD) will require using other 4-7 types of equipment treatment. Otherwise, if the Short Wave Diathermy (SWD) was not used in the physical therapy treatment, should to be used Stretching Passive Movement (SPM), if the Stretching Passive Movement (SPM) which used equipment amount treating is unstable in each time will require another 1-3 type of equipment treatment, and if used the Stretching

Copyright by Mahidol University

Passive Movement (SPM) which is used amount of treating is stable in in each time, will require another 4-7 type of equipment treatment.

## 4.2 Experimental results of Association Rule

Association Rule method was applied to study the association equipment usage of physical therapy equipment. The result was satisfactory as shown in Table 4.2.

**Table 4.2** The results of Association rules.

| Rules | Confidence |
|---|---|
| 1. Stretching Passive movement and Massage ==> Hydro collator. | 97% |
| 2. Massage ==> Hydro collator. | 96% |
| 3. Ultrasound ==> Hydro collator. | 93% |
| 4. Stretching Passive Movement and Ultrasound ==> Hydro collator. | 93% |
| 5. Massage, Hydro collator ==> Stretching Passive Movement | 92% |
| 6. Stretching Passive Movement ==> Hydro collator. | 92% |
| 7. Massage ==> Stretching Passive Movement | 91% |

In addition, we present the association rule method to study the association equipment usage of the physical therapy equipment. We found the top five interesting rules to make a decision to support the physical therapist to select the treatment equipment treat the patient with shoulder pain are as follows.

1) If the physical therapist selected Stretching Passive Movement and Massage then tend to choose Hydro collator.

2) If the physical therapist selected Massage then tend to choose Hydro collator.

3) If the physical therapist selected Ultrasound then tend to choose Hydro collator.

4) If the physical therapist selected Stretching Passive Movement and Ultrasound then tend to choose Hydro collator.

Copyright by Mahidol University

5) If the physical therapist selected Massage and Hydro collator then tend to choose Stretching Passive Movement.

## 4.3 Knowledge Presentation

After proposing the classification system to discover and predict the equipment usage amount from physical therapy equipment usage patterns and the association rule method to establish selection rules of physical therapy techniques. The Advantages of this research is the both data mining applications of this research could support the decision making in the treatment of shoulder pain patients. The limitations is the amount of each equipment type are used can not specify the duration of the treatment in each patient.

According to the experimental results proposed of the Classification and Association Rules methods are consistent this makes the research results are reliable and can be combined as a rule to make the decision to support the treatment of shoulder pain patients . There are 18 instruments used in the treatment of patients with shoulder pain, the results of both experiments showed 4-5 instruments are used in the treatment of patients with shoulder pain as Ultrasound, Short Wave Diathermy, Stretching Passive Movement, Massage, and Hydro Collator. The Policy can be created to make recommendations for selecting the physical therapy equipment is as follows:

1) If the Hydro Collator was used then tend to choose Stretching Passive Movement.

2) If the Ultrasound and Hydro Collator was used then tend to choose Stretching Passive Moment.

3) If the Ultrasound was used then tend to choose Hydro Collator.

Copyright by Mahidol University

# CHAPTER V
# CONCLUSION

This thesis research proposed analyzing and studied the factors that affect the treatment of shoulder pain patients can be summarized are as follows.

5.1 Conclusion,

5.2 Limitation,

5.3 Future works.

## 5.1 Conclusion

This thesis is to apply the data mining techniques to gain the knowledge from the pattern usage in the physical therapy equipment and establish selection rules physical therapy techniques to help make a decision to support the physical therapists in the treatment of shoulder pain patients. Technically, the classification methods of Naïve Bayes, Decision Tree, and Artificial Neural Network (ANN) were considered in this research. This thesis also presented the association rule method to study the association equipment usage of the physical therapy equipment.

According to classification system, our classifier clearly proved that the system is effective for considering the equipment usage pattern to predict the equipment usage amount. As illustrated in an experimental result, the traceability of the ANN structure is able to predict the equipment usage amount of shoulder pain patient with a high average accuracy. Additionally, the rules extracted by the Association Rule discovery method could support the decision making of specialist in different aspect, an association of equipment usage.

A total of 105 shoulder pain patient cases, limited to only the shoulder pain patient without an accident. In the experiment, the result of 92.30% accuracy proves the ANN is an effective model for predict the equipment usage amount from equipment usage pattern since its parameters are tunable i.e. number of hidden nodes,

Copyright by Mahidol University

learning rate, and momentum. These parameters are appropriately tuned and found their optimums to 1, 0.1, and 0.2 respectively.

According the experimental results proposed of the Classification and Association Rules methods are consistent this makes the research results are reliable and can be combined as a rule to make the decision to support the treatment of shoulder pain patients. The Policy can be created to make recommendations for selecting the physical therapy equipment is as mentioned in Chapter IV.

## 5.2 Limitation

According to technical aspect, this research may be further considered for the inclusion of the classification system and the association rule. This may provide more effective results in the prediction of treatment, because the classification can predict the equipment usage amount of patients from usage of the pattern of physical therapy equipment while the association rules discovery method studies the association amount of physical therapy equipment used.

## 5.3 Future works

However, there are some challenging issues for future works in both physical therapy and technical aspects. According to physical therapy aspect, it may be extended to gain the knowledge in the discovery of the knowledge management in herniated intervertebral disk patients since it is a disease with high level pain and difficulty in moving. The knowledge acquired for analyzing the appropriate treatment technique will be the key to a successful treatment plan for herniated intervertebral disk patients.

Copyright by Mahidol University

# REFERENCES

1 Department of Health Education [Internet]. Bangkok: Pranangklao Hospital; [Cited 2013 September 20]. Available from:
http://www.pranangklao.go.th/information1-2.php

2 Surasak Srisuk LP, Nuananong Chaipiyaporn. Shoulder Pain. 7, editor. Bangkok 2007.

3 Prajane P. Shoulder. Academic Service Centre Khon Kaen Universiry. 2003;11(2).

4 Sirichartvapee W. Frozen Shoulder. Sri Nakharin Medical Journal. 2001;16:84.

5 Vuttigoso S. Usage of Data Mining in Frozen Shoulder Patient of Pranangklao Hospital [dissertation]. Nakhon Si Thammarat: Walailak University; 2008.

6 Reeves B. The natural history of the frozen shoulder syndrome. Scandinavian journal of rheumatology. 1975(4):193-6.

7 Jan K, Richardson Z, Iglarsh A. Clinical Orthopaedic Physical Therapy. USA: W.B. Saunders; 1994.

8 Suteerawattananon M, RV. The Prevalence Study of Therapeutic Techniques Used in Physical Therapy. 1996;18(1):38-47.

9 Vongsirinavarat M. Physical Therapy in Common Shoulder Dysfunction. Bangkok: Mahidol University; 2004.

10 Rakesh Agrawal, Heikki Mannila, Ramakrishnan Srikant, Hannu Toivonen, Verkamo AI. Advances in Knowledge Discovery and Data Mining. IBM Almaden Research Center. 1996;12(1): 307-28.

11 De Martin M. Bertone, A. Albertoni, R. Hauska, H. Demsar, U. Dunkars, M. Technical Report of Data Mining. No. INVISIP project IST-2000-29640. Technical Report for the European Commission, 2002.

12 Rennie JD, Shih L, Teevan J, Karger DR, editors. Tackling the poor assumptions of naive bayes text classifiers. ICML; 2003: Washington DC).

13 Mitchell TM. The discipline of machine learning: Carnegie Mellon University, School of Computer Science, Machine Learning Department; 2006.

Copyright by Mahidol University

14 Irina Rish.An Empirical Study of the Naïve Bayes Classifier.IBM Research Report, 2001 Nov 2. New York. 2001.

15 Quinlan JR. Simplifying decision trees. International journal of man-machine studies. 1987;27(3):221-34.

16 Turban E, RS, D. Support and Business Intelligence Systems. 9, editor. Pearson2011.

17 Thammaboosadee S, Watanapa B, SILPARCHA U. A Two-Stage Classifier That Identifies Charge and Punishment under Criminal Law of Civil Law System. IEICE TRANSACTIONS on Information and Systems. 2014;97(4):864-75.

18 Kröse B, Krose B, van der Smagt P, Smagt P. An introduction to neural networks. 1993.

19 Kohavi R, editor A study of cross-validation and bootstrap for accuracy estimation and model selection. Ijcai; 1995.

20 Zaki MJ. Scalable algorithms for association mining. Knowledge and Data Engineering, IEEE Transactions on. 2000;12(3):372-90.

21 Agrawal R, Imieliński T, Swami A, editors. Mining association rules between sets of items in large databases. ACM SIGMOD Record; 1993: ACM.

22 Vattani A. K-means requires exponentially many iterations even in the plane. Discrete & Computational Geometry. 2011;45(4):596-616.

23 Tam S-F, Cheing GL, Hui-Chan CW. Predicting osteoarthritic knee rehabilitation outcome by using a prediction model developed by data mining techniques. International Journal of Rehabilitation Research. 2004;27(1):65-9.

24 Bulgen D, Binder A, Hazleman B, Dutton J, Roberts S. Frozen shoulder: prospective clinical study with an evaluation of three treatment regimens. Annals of the rheumatic diseases. 1984;43(3):353-60.

25 Van der Heijden GJ, Leffers P, Wolters PJ, Verheijden JJ, van Mameren H, Houben JP, et al. No effect of bipolar interferential electrotherapy and pulsed ultrasound for soft tissue shoulder disorders: a randomised controlled trial. Annals of the rheumatic diseases. 1999;58(9):530-40.

Copyright by Mahidol University

26 Bag M, Gauri SK, Chakraborty S. Recognition of Control Chart Patterns using Discriminant Analysis of Shape Features. 2010.

27 Rennie JD, Shih L, Teevan J, Karger DR, editors. Tackling the poor assumptions of naive bayes text classifiers. ICML; 2003: Washington DC).

28 Turban E, Watkins PR. Integrating expert systems and decision support systems. MIS quarterly. 1986:121-36.

29 Lai K, Cerpa N, editors. Support vs. confidence in association rule algorithms. Proceedings of the OPTIMA Conference, Curicó; 2001.

Copyright by Mahidol University

# BIOGRAPHY

| | |
|---|---|
| **NAME** | Mr. Kittisak Kaewbooddee |
| **DATE OF BIRTH** | 10 June 1984 |
| **PLACE OF BIRTH** | Surin, Thailand |
| **INSTITUTIONS ATTENDED** | King Mongkut's University of Technology Thonburi, 2006-2008<br>Bachelor of Science<br>(Information Technology)<br>Mahidol University, 2011-2015<br>Master of Science<br>(Technology of Information System Management) |
| **HOME ADDRESS** | 4/501 Moo 5 Sakaengam Villa<br>Sakaengam 7, Rama II Rd.<br>Samaedum, Bangkhunthien<br>Bangkok 10150<br>Tel. 094-405-3525<br>E-mail: kittisak.kae@mahidol.ac.th |
| **PUBLICATION / PRESENTATION** | Kittisak Kaewbooddee, The Data Mining Applications of Shoulder Pain Patients Treatment: Physical Therapy Equipment Usage Approaches, 2nd Int. Symposium on Future Information and Communication Technologies for Ubiquitous HealthCare - 2015, 28-30 May 2015, BUPT Hotel, Beijing, China. |

Copyright by Mahidol University