

UNIVERSITI TEKNOLOGI MARA

**AGARWOOD OIL QUALITY
CLASSIFICATION USING
K-NEAREST NEIGHBORS AND
CORRELATION-BASED FEATURE
SELECTION**

MOHAMAD AQIB HAQMI BIN ABAS

Thesis submitted in fulfilment
of the requirements for the degree of
Master of Science

Faculty of Electrical Engineering

June 2018

ABSTRACT

Agarwood oil is an essential oil which is a concentrated volatile aromatic compound that is produced by agarwood plant. It is widely being used as incense and fragrance in religious prayers and traditional ceremonies. The market price of agarwood oil being traded depends on its quality. It has been shown from literatures that agarwood oil quality can be classified into high or low quality. The latest method combined lab based GC-MS, z-score technique and ANN for agarwood oil quality classification, however the classification accuracy score achieved was only between range of 81-86% and does not achieve 100%. This thesis describes a representation of k -nearest neighbors to classify the agarwood oil sample quality to high quality and low quality. In this study, the chemical compound of agarwood oil samples is obtained from GC-MS analysis. Different type of feature scaling and data splitting technique were used in this experiment to analyze the effect on training the classifier model on the agarwood oil quality sample dataset. Using correlation-based feature selection, it was found that out of seven chemical compounds abundances only five were predictive. The k -nearest neighbors model was analyzed to get a comparison between different number of neighbors used to the overall classification accuracy and performance measure score of the model. The best scaling method found from the experiment is min-max scaling. From the results of the experiment, it shows that using stratified k -fold cross validation splitting technique have a much better performance and stable model scores when compared to using hold-out test set technique. The k -NN classifier was built with number of neighbours ranging from 1 to 20 to obtain the best range of number of neighbours to be used to get the highest classification accuracy. The results from the experiment shows that the best range for number of neighbours used for k -NN classifier are 1 to 8 and best data splitting technique is stratified k -fold cross validation as the combination that have the highest classification accuracy to classify the quality of agarwood oil which is at 100%. The best number of neighbors parameters for the model is five.

ACKNOWLEDGEMENT

I would like to thank all ASPRG members for sharing their experience, knowledge and assistance. May Allah reward them all. I extend my appreciation to Forest Research Institute Malaysia (FRIM), Malaysia and Universiti Malaysia Pahang (UMP), Malaysia for providing the agarwood oil dataset and to my Faculty of Electrical Engineering, Universiti Teknologi MARA, Malaysia.

Last but not least, I would like to express my love and gratitude especially to *Ibu, Abah*, my family and colleagues. It would not be possible for me to complete this research without their moral support to complete the study.

Mohamad Aqib Haqmi bin Abas
June 2018

TABLE OF CONTENTS

	Page
CONFIRMATION BY PANEL OF EXAMINERS	ii
AUTHOR'S DECLARATION	iii
ABSTRACT	iv
ACKNOWLEDGEMENT	v
TABLE OF CONTENTS	vi
LIST OF TABLES	ix
LIST OF FIGURES	x
LIST OF ABBREVIATIONS	xi
CHAPTER ONE: INTRODUCTION	1
1.1 Research Background	1
1.2 Problem Statement	2
1.3 Objectives of The Study	3
1.4 Scope of The Study	3
1.5 Thesis Organization	4
CHAPTER TWO: LITERATURE REVIEW	6
2.1 Introduction	6
2.2 Agarwood	6
2.2.1 Agarwood Oil	8
2.3 Machine Learning	9
2.3.1 Machine Learning Optimization	10
2.3.2 Machine Learning Evaluation	11
2.3.3 Machine Learning in Agriculture	11
2.3.4 K-Nearest Neighbor (<i>k</i> -NN)	12
2.4 Feature Selection	14
2.4.1 Pearson Correlation Coefficient	15
2.4.2 Curse of Dimensionality	15

CHAPTER ONE

INTRODUCTION

1.1 RESEARCH BACKGROUND

Agarwood or also known as Gaharu is a resinous impregnated heartwood of *Aquilaria* species plant [1]–[4]. It can be found in countries such as India, Malaysia, Indonesia, Thailand, Burma, Laos and Vietnam [5]. It is widely known by names such as aloeswood, eagle wood, agarwood, oud, kalambak, jinkoh and chen xiang [2], [5], [6]. Agarwood oil is an essential oil which is a concentrated volatile aromatic compound that is produced by agarwood plant [2], [7]–[9]. Agarwood oil is widely traded and has a very high demand internationally because of its special usage; as incense, perfumes and medicine [1], [2], [7]. Agarwood oil is also more expensive than perfume and fragrance oil because it contains true essence of the plant that it derived from [10].

Agarwood oil has been widely used in religious ceremony, traditional medicine and perfumery industry [2], [11]. It also represents the symbol of wealth in some Middle East countries [2]. Agarwood oil is being traded internationally and it has a very high market demand, especially from countries in the Middle East such as United Arab Emirates and Saudi Arabia; China and Japan [12]. The demand for agarwood oil has increase every year especially in Middle East countries, Malaysia, Japan and Taiwan [9]. The market price of agarwood oil being traded depends on its quality, purity and concentration. The grading method used to classify the quality of agarwood oil is different for each country.

Common grading method that was found from previous literatures and practices are by classifying the oil quality as low quality and high quality. High quality oil have thicker color compared to low quality; and it has a longer lasting odour. The common process of grading the agarwood oil is basically by using the agarwood oil physical appearances like it's color, odor and consumer perception [6], [10], [13]. For this reason, traditionally, an expert is hired to classify the agarwood oil quality manually by the agarwood oil physical appearances such as its color and odor. Studies shows that the grading of agarwood oil can be done by using its chemical profiles [6], [12], [14]. In the past few years, there has been