



**ANALYSIS OF HARUMANIS MANGO  
FLOWERING PREDICTION THROUGH BIOTIC  
AND ABIOTIC FACTORS USING MACHINE  
LEARNING**

by

**ROHANI BINTI S. MOHAMED FAROOK  
(1040210559)**

A thesis submitted in fulfillment of the requirements for the degree of  
Doctor of Philosophy

**School of Computer and Communication Engineering  
UNIVERSITI MALAYSIA PERLIS**

2018

## ACKNOWLEDGEMENT

In the Name of Allah, The Most Beneficent, and The Most Merciful: “My Lord, enable me to be grateful for Your favor which You have bestowed upon me and upon my parents and to do righteousness of which You approve. And admit me by Your mercy into [the ranks of] Your righteous servants. (An-Naml:19)”. Thank You Allah for all the Ni'mah.

First and foremost, I would like to express my gratitude to my Ph.D. supervisor Dato' Prof Dr. Ali Yeon Bin Md Shakaff and Co-supervisor Prof Dr. Mahmad Nor Jaafar for the support and trust they has given to me throughout this journey.

I am very grateful to my friends, for all the joy we had together. Especially to whom always beside me giving me support and coloring my days of finishing this journey, Aznor Hanah, Aini Syuhada, Nazatul Syima, Normaliza and Zulkifli.

A special thanks to my siblings, mother and father who has been there for me in my darkest time. My heartiest thanks to my son Hanif and Naim, my daughter Marsya for all the tolerance and endless love for me.

Last and not least, I would like to say thank you for those who no words can describe his love and support. My beloved one, my late husband Hassan Bin Ali, for all the love, moral and emotional support in my life. Without him, this journey is hard for me to pursue and might never come to end. May Allah bless him with highest rank in Jannah. Indeed we are, Allah willing, soon to follow. May Allah grant well-being for us and for you, my love.

## TABLE OF CONTENT

	<b>PAGE</b>
<b>DECLARATION NOTICE</b>	<b>i</b>
<b>ACKNOWLEDGEMENT</b>	<b>ii</b>
<b>TABLE OF CONTENT</b>	<b>iii</b>
<b>LIST OF TABLES</b>	<b>v</b>
<b>LIST OF FIGURES</b>	<b>viii</b>
<b>LIST OF ABBREVIATIONS</b>	<b>xii</b>
<b>ABSTRAK</b>	<b>xiii</b>
<b>ABSTRACT</b>	<b>xiv</b>
<b>CHAPTER 1</b>	
<b>INTRODUCTION</b>	
1.1 Background	1
1.2 Problem Statement	5
1.3 Objectives	8
1.4. Research Scope	9
1.5. Contribution of the thesis	9
1.6 Layout of Thesis	11
<b>CHAPTER 2</b>	
<b>LITERATURE REVIEW</b>	
2.1 Introduction	12
2.2 Underlying Concepts of Mango Flower Initiation	14
2.3 Mango Phases and Flowering factors	17
2.4 Machine Learning	20
2.4.1 Prediction using Machine Learning	22
2.4.2 Machine Learning in Agriculture Application	28
2.4.3 Machine Learning in Mango Plantation	30
2.5 Summary	32
<b>CHAPTER 3</b>	
<b>MANGO FLOWERING PREDICTION USING BIOTIC FACTORS</b>	<b>34</b>
3.1 Methodology	34
3.1.1 Introduction	34
3.1.2 Biotic Data Accumulation from Harumanis Mango Trees in the Greenhouse	35
3.1.3 Biotic Feature Preliminary Investigation	43

3.2	Result of Machine Learning in Biotic Flowering Prediction	64
3.3	Summary	77
<b>CHAPTER 4</b>		
<b>MANGO FLOWERING PREDICTION USING ABIOTIC FACTORS</b>		<b>78</b>
4.1	Methodology	78
4.1.1	Introduction	78
4.1.2	Abiotic Data Accumulation from Meteorology Department	79
4.1.3	Prediction using Meteorological data	82
4.1.4	Identifying correlation between different factors against the yield on season	88
4.1.5	Experimental Setup and Prediction System in Greenhouse	104
4.2	Results of Machine Learning in Abiotic Flowering Prediction	112
4.3	Summary	118
<b>CHAPTER 5</b>		
<b>CONCLUSION AND FUTURE WORKS</b>		<b>120</b>
5.1	Discussion	120
5.2	Conclusion	121
5.3	Contributions	123
5.4	Future Works	124
<b>REFERENCES</b>		<b>125</b>
<b>LIST OF PUBLICATIONS</b>		<b>140</b>
<b>APPENDIX A</b>		<b>142</b>

## LIST OF TABLES

NO.		PAGE
Table 1.1:	The Export Demand and Supply of Harumanis to Japan	3
Table 3.1:	The Attributes and the Descriptions of Biotic Factors	41
Table 3.2:	The Attributes and the Data Type of the Biotic Factors of Harumanis Branches	62
Table 3.3:	The outcomes of a binary prediction	72
Table 3.4:	Classification Accuracy for different fold cross validation	74
Table 3.5:	Classification Accuracy, Sensitivity and Specificity of the Predictive Models through Biotic Factors	75
Table 3.6:	Accuracy, Sensitivity, Specificity and Precision using different Kernels through Biotic Factors	76
Table 4.1:	The Year, Yield and the yield category for prediction	79
Table 4.2:	The Climatic Data Item and Data Type from Chuping Meteorology	81
Table 4.3:	The Monthly Average Minimum Temperature during Pre-Flowering and Flowering Phase with Pearson r value	86
Table 4.4:	The rainfall during pre-flowering and flowering phase with Pearson r Value	87
Table 4.5:	The Pearson R values of multiple regressions that quantify the relations among the average monthly temperature and the monthly rainfall on yield	87
Table 4.6:	The Pre-Flowering and Flowering Phase Monthly Average Minimum Temperature and Pearson r value	89

Table 4.7:	The Predicted and Actual Yield of Harumanis for Year 1998 – 2004 using January Average Minimum Temperature model (4.1).	90
Table 4.8:	The Predicted and Actual Yield of Harumanis for Year 1998 – 2004 using December Average Minimum Temperature model (4.2).	90
Table 4.9:	The Predicted and Actual Yield of Harumanis for Year 1998 – 2004 using Monthly Average Minimum Temperature by model (4.3).	91
Table 4.10:	The Decade Average Minimum Temperature during and Pearson r value	92
Table 4.11:	The Monthly Average Mean Temperature during Pre-Flowering and Flowering Phase and Pearson r value	94
Table 4.12:	The Decade Average Mean Temperature during and Pearson r value	94
Table 4.13:	The Monthly Rain during Pre-Flowering and Flowering Phase and Pearson r value	96
Table 4.14:	The Rainfall during Pre-Flowering and Flowering Phase and Pearson r value	96
Table 4.15:	The Predicted and Actual Yield of Harumanis for Year 1998 – 2004 using February Rain by model (4.4).	97
Table 4.16:	The Predicted and Actual Yield of Harumanis for Year 1998 – 2004 using December Rain by model (4.5).	97
Table 4.17:	The Predicted and Actual Yield of Harumanis for Year 1998 – 2004 using January Rain by model (4.6).	98

Table 4.18:	The Predicted and Actual Yield of Harumanis for Year 1998 – 2004 using December, January and February Rainfall by model (4.7).	99
Table 4.19:	The Pearson r values of the multiple regressions that quantify the relations among the average monthly temperature and the monthly rainfall on yield.	99
Table 4.20:	The Predicted and Actual Yield of Harumanis for Year 1998 – 2004 using January Average Minimum Temperature and Rainfall resembled by model (4.8).	100
Table 4.21:	The Predicted and Actual Yield of Harumanis for Year 1998 – 2004 using December Average Minimum Temperature and Rainfall resembled by model (4.9).	101
Table 4.22:	The Predicted and Actual Yield of Harumanis for Year 1998 – 2004 using February Average Minimum Temperature and Rainfall resembled by model (4.10).	102
Table 4.23:	The Decade Average Rain during and Pearson r value	103
Table 4.24:	The Abiotic Features accumulated in greenhouse and descriptions	110
Table 4.25:	The Attributes and the Descriptions of Abiotic Factors that accumulated from the greenhouse	111
Table 4.26:	The Abiotic Attributes as input in the model training and testing	112
Table 4.27:	Classification Accuracy, Sensitivity and Specificity of the Predictive Models through Abiotic Factors	118
Table 5.1:	The prediction accuracy of the models through biotic and abiotic factors	121

## LIST OF FIGURES

NO.		PAGE
Figure 1.1 :	Harumanis Mango Tree	2
Figure 1.2:	The Harumanis Mango	2
Figure 2.1:	The Conceptual Flowering model of mango (Thomas L Davenport, 2007).	16
Figure 2.2:	The mango flower induction factors	18
Figure 2.3:	Harumanis Mango Growth and Reproductive Phases	19
Figure 2.4:	Reproductive Phases in Mango	20
Figure 2.5:	CAT node	26
Figure 2.6:	Linear Classifier	27
Figure 2.7:	Non-linear Classifier	28
Figure 2.8:	Machine learning in mango domain	32
Figure 3.1:	The Harumanis mango trees planted in greenhouse at INSAT, UniMAP	36
Figure 3.2:	The flower that blooming in Harumanis Tree in Lysimeter	37
Figure 3.3:	Closer look at the generative flush of a mango tree that has the one-week mango panicles	38
Figure 3.4:	Mango tree in micro lysimeter that has almost two weeks mango panicles	39
Figure 3.5:	The mango terminal stems three whorls.	40
Figure 3.6:	The trees set in the greenhouse	43
Figure 3.7:	The Box Plot for the First Whorl Length	44

Figure 3.8:	The Box Plot for the Second Whorl Length	44
Figure 3.9:	The Box Plot for the Third Whorl Length	45
Figure 3.10:	The Box Plot for the Diameter	46
Figure 3.11:	The boxplot for All the Intercalary Unit	47
Figure 3.12:	The First Whorl Length for different lysimeters	48
Figure 3.13:	The Second Whorl Length for different lysimeters	49
Figure 3.14:	The Third Whorl Length for different lysimeters	50
Figure 3.15:	The Diameter of different lysimeters	51
Figure 3.16:	The First World Length by the Branch State	52
Figure 3.17:	The Second World Length by the Branch State	53
Figure 3.18:	The Third World Length by the Branch State	54
Figure 3.19:	The Diameter of the Branch State	55
Figure 3.20:	The length of the branches for different lysimeter in 2 branch state	56
Figure 3.21:	The length of the branches for two branch state within different lysimeters	57
Figure 3.22:	The Length of the first whorl features for 3 lysimeters in flowering and non-flowering branches	58
Figure 3.23:	The Length of the second whorl features for 3 lysimeters in flowering and non-flowering branches	59
Figure 3.24:	The Length of the third whorl features for 3 lysimeters in flowering and non- flowering branches	60
Figure 3.25:	The diameter of the third whorl features for 3 lysimeters in flowering and non-flowering branches	61

Figure 3.26:	Linear Projection for flowering (1.0) and non-flowering (0.0) branches	63
Figure 3.27:	k-NN Harumanis prediction model through biotic factors	65
Figure 3.28:	The flowering prediction pseudocode using k-NN for through biotic factors	66
Figure 3.29:	Naives Bayes Harumanis prediction model through biotic factors	67
Figure 3.30:	The flowering prediction pseudocode using Naives through biotic factors	68
Figure 3.31:	CART Harumanis prediction model through biotic factors	69
Figure 3.32:	The Harumanis flowering predictive model using SVM	71
Figure 4.1:	Harumanis Yield /Hectar yearly	81
Figure 4.2:	Average Minimum Temperature of High Yield years for the month of November to October	83
Figure 4.3:	Rainfall of High Yield years for the month of November to October	84
Figure 4.4:	Average Minimum Temperature of Low Yield years for the month of November to October	85
Figure 4.5:	Rainfall of Low Yield years for the month of November to October	85
Figure 4.6:	The Harumanis Trees Planted Greenhouse	105
Figure 4.7:	The Temperature sensor and the WiFi Eko Node	106
Figure 4.8:	Eko Node connected with soil moistures	107
Figure 4.9:	The Overview of the WSN system that is developed at greenhouse	108

Figure 4.10: Sensors deployment in mango greenhouse	109
Figure 4.11: Sensor Placement for Micro-Lysimeter	111
Figure 4.12: Predictive model flowchart using abiotic factors K-NN	113
Figure 4.13: Predictive model flowchart using abiotic factors through Naives Bayes	114
Figure 4.14: Predictive model flowchart using abiotic factors through CART	115
Figure 4.15: The Harumanis predictive model using abiotic factors through SVM	117

©This item is protected by original copyright

## LIST OF ABBREVIATIONS

AI	Artificial Intelligent
CAT	Classification Trees
DOA	Department of Agriculture
DW	Diameter of the whorl
FN	False Negative
FP	False Positive
GHPS	Greenhouse Production System Technology
ICT	Information and Communication Technology
INSAT	Institute of Sustainable Agrotechnology
k-NN	k-Nearest neighbors
LFW	Length of the first whorl
LSW	Length of the second whorl
LTW	Length of the third whorl
ML	Machine Learning
RBF	Radial Basis Function
RF	Random Forests
SVM	Support Vector Machine
TN	True Negative
TP	True Positive
UNIMAP	University Malaysia Perlis

## Analisis Ramalan Pembungaan Mangga Harumanis melalui Faktor Biotik dan Abiotik menggunakan Pembelajaran Mesin

### ABSTRAK

Mangga Harumanis (*Mangifera indica*) dikenali sebagai salah satu jenis buah tropikal yang terbaik, disebabkan aroma dan kemanisannya. Penanaman mangga Harumanis dimasukkan di dalam agenda nasional sebagai buah mangga khas Perlis, Malaysia untuk dunia. Walaupun permintaan dari dalaman Malaysia dan antarabangsa begitu banyak, penawaran buah ini tidak pernah memenuhi permintaan. Ramalan bunga mangga adalah sebuah faktor yang penting bagi meramal keluaran mangga bagi membolehkan pelaksanaan pemasaran ke hadapan yang efektif. Pemasaran Kehadapan ialah sebuah kontrak yang ditandatangani antara pemborong barangan dan pembeli yang mempersetujui bilangan penghantaran dan harga barangan di masa hadapan yang ditetapkan berdasarkan keluaran ramalan. Mangga Harumanis ialah spesies yang berbuah hanya sekali setahun. Faktor-faktor biotik dan persekitaran dilaporkan dalam kertas-kertas kajian sebagai faktor yang mempengaruhi pokok mangga berbunga dan berbuah. Analisis yang dijalankan menunjukkan bahawa faktor biotik dan abiotik mempunyai hubungan tak linear dengan hasil mangga. Adalah penting untuk membina, melatih dan menguji model peramalan pembungaan pokok mangga Harumanis melalui pendekatan pembelajaran mesin seperti *K-Nearest Neighbors (k-NN)*, *Naives Bayes*, *Support Vector Machine (SVM)*, *Classification Trees (CAT)* dan *Random Forests (RF)*. Model ramalan pembungaan pokok mangga menggunakan faktor biotik dan abiotik dibina, dilatih dan diuji menggunakan data yang dikumpulkan dari rumah hijau pokok Harumanis. Faktor-faktor biotik ialah lysimeter, *Length of First Whorl (LFW)*, *Length of Second Whorl (LSW)*, *Length of Third Whorl (LTW)* dan *Diameter of the Whorl (DW)*. Ramalan pembungaan Harumanis menggunakan faktor biotik menunjukkan bahawa ketepatan ramalan teknik SVM adalah 79.9% berbanding dengan *k-NN*, *Naives Bayes*, *CAT* dan *RF* pada 66.5%, 74%, 71.3% and 72% masing-masing. Model ramalan SVM diuji bagi beberapa kernel iaitu linear, polinomial, radial basis dan sigmoid. Kernel radial basis meramal dengan ketepatan 79.9% berbanding dengan linear, polinomial dan sigmoid pada 65%, 65.7% dan 59.8% masing-masing. Data persekitaran daripada Jabatan Meteorology Perlis dan hasil keluaran dari Ladang Bukit Bintang di proses dan di analisis bagi mengenalpasti kepentingan faktor abiotik dalam meramal hasil keluaran Harumanis. Seterusnya, faktor abiotik daripada rumah hijau yang ditanam mangga Harumanis iaitu purata suhu minimum dan purata kebasahan tanah bagi setiap 10 hari digunakan bagi membina model ramalan pembungaan Harumanis. Ramalan pembungaan pokok mangga Harumanis melalui faktor abiotik menunjukkan bahawa kadar ketepatan teknik SVM adalah 90.6% berbanding kadar *k-NN*, *Naives Bayes*, *CAT* dan *RF* pada 76.6%, 75%, 82.1% dan 72% masing-masing. Kesimpulannya, model ramalan menggunakan SVM dalam kernel radial basis melalui faktor biotik dan abiotik menunjukkan kadar ketepatan yang tertinggi pada 79.9% dan 90.6% masing-masing. Model ramalan SVM dalam kernel radial basis dapat melaksanakan ramalan walaupun menggunakan data yang terhad yang disebabkan oleh ciri-ciri domain pertanian dimana pengumpulan data dan pemerhatian memerlukan masa yang panjang.

## **Analysis of Harumanis Mango Flowering Prediction through Biotic and Abiotic Factors using Machine Learning**

### **ABSTRACT**

Harumanis Mango (*Mangifera indica*) is known as one of the best table tropical fruit, due to its aroma and sweetness. Harumanis mango cultivar is included in the national agenda as a specialty fruit from Perlis, Malaysia for the world. Despite its overwhelming local demand in Malaysia and also internationally, the fruit supply never meets the demand. Mango flowering prediction is important as one of the factors to predict mango yield in order to implement effective forward marketing. Forward marketing is a contract that is signed between supplier and client based on the amount of delivery and the price of delivery in future, based on the predicted yield. Harumanis mango is a species that only bear fruit once a year. The biotic and environmental factors are reported in the literature as the factors that influence the mango trees flowering and fruit-bearing. The pre-processing and analysis done shows that the biotic and abiotic factors have non-linear relation with the yield. It is essential to develop, train and test the Harumanis mango tree flowering prediction model through machine learning approaches such as K-Nearest Neighbors (k-NN), Naive Bayes, Support Vector Machine (SVM), Classification Trees (CAT) and Random Forests (RF). Harumanis flowering predictive model on biotic and abiotic factors developed, trained and tested through the data accumulated from Harumanis trees in the greenhouse. The biotic factors are lysimeter, Length of First Whorl (LFW), Length of Second Whorl (LSW), Length of Third Whorl (LTW) and Diameter of the Whorl (DW). The Harumanis flowering prediction on biotic factors indicates that SVM technique prediction accuracy is at 79.9% as compared to k-NN, Naive Bayes, CAT and RF at 66.5%, 74%, 71.3% and 72%, respectively. The SVM predictive model further tested on several kernels which are linear, polynomial, radial basis and sigmoid. The radial basis kernel accuracy is at 79.9% compared to linear, polynomial, and sigmoid at 65%, 65.7% and 59.8% respectively. The environmental data from Perlis Meteorology Department and the yield from Bukit Bintang Orchard were analyzed to identify the significant abiotic factors in predicting the Harumanis mango yield. Later, the abiotic factors which are average minimum temperature and average soil moisture of the 10 days from Harumanis greenhouse are calculated and utilized in developing the Harumanis flowering predictive model. The Harumanis mango tree flowering prediction on abiotic factors shows that SVM technique is at 90.6% accuracy rate compared to k-NN, Naive Bayes, CAT and RF at 76.6%, 75%, 82.1% and 72% respectively. Concluded that the prediction model using SVM on radial basis kernel through biotic and abiotic factors displays the highest prediction accuracy at 79.9% and 90.6% accordingly. The SVM with radial basis kernel model able to perform flowering prediction although using a limited data due to the nature of the agricultural domain where the data collection and observation require a longer period of time.

## CHAPTER 1

### INTRODUCTION

Tremendous advances have been made in realising the precision agriculture in achieving the goals to increase productivity and yield. There are multi perspective that are being investigated at various domain such as agricultural domain, processes and technologies. Enabling the precision agriculture requires the knowledge on the factors that involved in particular cultivar that impact the yield directly or indirectly.

#### 1.1 Background

Harumanis mango has been a vital commodity of economic to the state of Perlis and Malaysia as a whole for its value in the agricultural field. Perlis state has always known as a state that offers Harumanis mango with excellent taste. The taste and the aroma of this particular mango cultivar have captured wide interest in import value from various countries (Nadzmi, 2009). Figure 1.1 is a Harumanis Mango Tree planted in one of the famous Harumanis orchards at Bukit Bintang Perlis, Malaysia. The trees are planted, managed and monitored by the orchard staffs and the trees are producing revenue each year. The trees are being pruned and fertilized accordingly. Since the Harumanis mango trees are yearly bearer trees the pruning and fertilizing can be scheduled according to the growth and harvest plan yearly.



Figure 1.1 : Harumanis Mango Tree

The look and the color of this variety of mango are beautiful and captivating. The rich orange color of the mango is an attractive nature and resembles healthy mango as in Figure 1.2.



Figure 1.2: The Harumanis Mango

Every year the demand increment is very high and the production needed to be planned and scrutinized according to the export standard. This particular mango had been included in national agenda as a specialty fruit from Perlis for the world (Malaysia, 2015). The export of this particular mango is in very high demand especially to Japan (Nadzmi, 2010). It always has been a challenge to Malaysia to fulfill the supply to achieve the export demands. The demand from Japan since the year 2010 as shown in Table 1.1.

Table 1.1 displays, Harumanis export demand from Japan is 20 metric ton each year from 2010 to 2013 and the actual export supply never achieve the demand. Starting from 2010 supplied only 0.56 metric ton which is 2.8% of the demand, the year 2011 supplied 2.24 metric ton at 11.2% of the demand, the year 2012 supplied 0.617 metric ton at 3.1% of the demand and year 2013 supplied 2.12 metric ton at 10.6% of the demand (Bahri, 2013).

Table 1.1: The Export Demand and Supply of Harumanis to Japan. (Bahri, 2013)

Year	Demand(MT)	Supply (MT)	%
2010	20.0	0.56	2.8
2011	20.0	2.24	11.2
2012	20.0	0.617	3.1
2013	20.0	2.12	10.6

Due to this, Department of Agriculture (DOA) has draft several missions in increasing the Harumanis mango plantation such as Harumanis Mango Plantation Cluster Development Project that could undertake the increment in the plantation output of this particular mango synchronize with the increment of the productivity through the technology implementation and good plantation practice (Malaysia, 2015). University Malaysia Perlis (UniMAP) as one of research entity in Perlis vows to engage in this plan for Perlis and launches the plantation and researchs of Harumanis in greenhouses at Institute of Sustainable Agrotechnology (INSAT), UniMAP. There are 100 greenhouses

build in UniMAP with a total of 200 trees in each greenhouse. Total of 20,000 trees planted in greenhouses in Greenhouse Production System Technology (GHPS) that involves the internet, remote sensing and automation (Illi Shazwani, 2014).

The export process also requires that the supplier country to sign a consignment called forward marketing contract that includes the quantity of the demand, the quantity of the supply and the costing of the export that will take place in the date that agreed together. Knowing the yield information is essential in signing the consignments. There will be one or several consignments per year signed and agreed upon per season accordingly.

Perlis state in Malaysia claimed that has convenient agro-climatic conditions for this breed of mango but the Harumanis mango production never been an easy path. It had been real challenges for the producers in increasing the supply since it is a yearly bearing fruit tree. This increasing demand locally and internationally requires an approach to increase the yield per year of Harumanis. The journey in increasing the yield has been slow since the limited of knowledge on the factors that could increase the yield. The relationship between the factors and the yield that could be looked into the flowering phase is complex and highly nonlinear. Therefore, there is an essential need in developing a prediction model using the factors and parameters which able to accurately predict the flowering of the Harumanis mango trees.

Machine Learning is an approach that consists of statistical and structural technique in performing prediction, classification and decision support. There are many domains that employ machine learning in predictions of various tasks. The ability of machine learning techniques to predict in highly non-linear problems brings credibility to the technique.

## 1.2 Problem Statement

Harumanis mango cultivar is highly demanded fruit at local and international level. The particular cultivar bears fruit for only once a year and this is a challenge for Malaysia since the export demand keeps increasing each year, not to mention the demand in local. The Perlis State has come up with a few mission in increasing the production of the particular fruit including increasing the yearly bearing plant to twice bearing plant in order to fulfill the local and export demand. The research is important to the National Agenda to export the mango as the demand increases every year. The research is also important to both, the Harumanis orchards management and the farmers, who concern about the mango yield and interest in performing early reproductive shoot detection.

Currently, only the highly experienced farmers are aware of the influencing factors of the Harumanis mango yield. Majority of the farmers have little knowledge of the factors that influence the yield and the impact of biotic factors and climatic conditions on yield that is quite prominent especially to mango trees. Biotic factors are the features of the physiology of the tree that could be manipulated to initiate flowering, while abiotic features are the environmental effect that provides great impact to flowering initiation in mango trees. Their flowering process for the fruit bearing trees are almost similar depending on the climate and the type of fruits (Fumie Nishikawa, 2013). The conceptual flowering model has been utilised for mango, lychee and other citrus trees in discussing the flowering (T L Davenport, 2000). The specific crop such as Harumanis mango has a specific sensitivity to different conditions on the plant physiology or the environmental conditions. Mango yield producing process is a complex decision-making process at the different level of the crop phases especially at pre-flowering and flowering phases. According to the conceptual model, pruning, defoliation, application of nitrogen fertilizer

and irrigation during dry conditions are a few factors that could stimulate the shoot growth initiation.

Currently, there is lack of method to identify the Harumanis mango trees that would flower. In this thesis, the Harumanis mango flowering prediction model developed through machine learning approach by looking into mango pre-flowering and flowering phase. Currently, there is lack of machine learning model for predicting the flowering of the Harumanis mango. The conceptual flowering model that exist now are being utilised in mango, lychee and citrus fruits in agricultural researches (T L Davenport, 2000). The other similar project is where the mango flowering captured from the trees that are blooming with flowers using image processing and the flowers being counted (Wang et al., 2016).

There are a number of mechanistic models and mathematical models have been developed in the crop growth and agricultural field. Most of them are specific for different crops. Drummond et al (2003) claimed that these models are expensive and take more time to be developed. There are also data-driven models have been developed using classical statistics (Dixon et al., 1994; Sudduth et al., 1996) and machine learning methods (Drummond et al., 2003; Roel & Plant, 2004; Irmak et al., 2006) but none of Harumanis mango has been explored.

A statistical method such as regression has been widely used in the agricultural area to explore the model building. There are researchs that focus on mango trees but different cultivars according to the countries. The statistical method requires planned and repeated experiments in collecting data and building the models. The amount of time needed for a particular experiment could be very long and take decades especially for mango cultivars that only flowers and bear fruit one a year (J. W. Jones et al., 2017a).

Other countries have been progressing in research for a different cultivar of Mango but not specifically for Harumanis mango cultivar that we have in Malaysia. Harumanis mango in Malaysia especially in Perlis state, experiencing very different climate and background that for sure varies from other countries. Statistical approach is usually the main technique utilized in the agricultural domain in looking through the researches (J. W. Jones et al., 2017b). In this research, preliminarily, regression analysis is used as explorative data analysis to look through the nature of the data. Machine learning approaches are utilized to cater the non-linearity of the features relation with the flowering.

Accurately knowing the mango tree flowering status can assist to analyze the growth status and the flowering induction status for early flowering detection. Currently acquiring the mango flowering status is not viable. It is costly and subjective, especially in large scale in a field environment since the flowering required to be observed manually and will be too late for performing the forward contract with suppliers. Developing such a model using machine learning technique enable the predicting the flowering using the existing biotic and abiotic data accumulated from the mango area. The model could be used in planning the yearly flowering into twice a year flowering depending on the trees. The model could assist to plan in increasing the mango yield without increasing the plantation land area in the near future. Could the model predict the flowering with plausible accuracy and is it possible to implement the models in predicting the flowering. Several machine learning models are developed and performed applied to the flowering stem prediction. In summary, the following research questions were addressed:

- How can the biotic and abiotic features extracted to predict the Harumanis mango flowering

- How biotic and biotic data collected from the greenhouse planted mango trees could be the flowering predictor for the model generated using one or more of the following techniques: KNN, Naive Bayes, SVM, Random Forest and Classification Trees
- How machine learning technique accurately predict Harumanis mango flowering

This research is focused on the predictive ability of machine learning technique in agricultural domain specifically in predicting Harumanis mango flowering. The yield prediction features are essential for future crop planning and harvest increment in fulfilling the high export demand in future.

### 1.3 Objectives

The general aim of this research is to design and develop mango flowering predictive model using machine learning. The objectives are :

- 1) To identify the appropriate factors that have influence in Harumanis mango flowering
- 2) To develop a Harumanis mango flowering predictive model using biotic factors and validate the accuracy of the model in predicting the Harumanis mango flowering
- 3) To develop a Harumanis mango flowering predictive model using abiotic factors and validate the accuracy of the model in predicting the Harumanis mango flowering

#### **1.4. Research Scope**

Mango tree generally has two stages which are growth and reproductive stages. The reproductive stage is an important stage and directly connected to the flowering and fruiting. This research only focuses on the reproductive phase since the aim to conduct research on the ability to predict the flowering stems.

The data utilized in this research were collected from various sources according to the function and aims. The climatic data from meteorology department of Chuping Perlis were collected from 2005 to 2011 to develop the relational model with the yield data that were accumulated from Bukit Bintang Orchard. The data from Bukit Bintang Orchard are from the year 2006 to 2011 that reported the yield of Harumanis mango trees in a metric ton.

The data utilized to develop and test the machine learning models are accumulated from the Harumanis mango trees in the greenhouse at INSAT from 2011 to 2014. The data consist of the mango trees attributes called biotic factors and also the environmental data called abiotic factors related to the research objectives. The reproductive phase span from pre-flowering, flowering, fruit bearing and harvest phases. In this research only focus on pre-flowering and flowering phases since the goal of the research is in interest of predict flowering in the mango trees.

The effectiveness of the predictive models developed is validated in terms of prediction accuracy and also displayed using confusion matrix.

#### **1.5. Contribution of the thesis**

The thesis reports the main contributions of the work which are as listed below:

- Harumanis mango predictive models through biotic and abiotic features using SVM – RBF kernel is identified and displays better prediction compared to other approaches. The Harumanis mango predictive model highly contribute in Harumanis mango planted greenhouses to plan, manage and control the microclimate of the greenhouse to obtain the optimum surrounding to flowering as planned to fulfill the forward marketing contracts in future.
- Predictive models using ML that applied in agricultural areas specifically Harumanis mango plantation
- Usefulness of the plant and environmental parameters through application of machine learning techniques and the knowledge of a domain expert
- Exploratory data analysis of the flowering stem of the Harumanis mango trees real data
- Develop the Harumanis mango flowering predictive model that could be utilized in developing monitoring and yield increasing system in greenhouses.
- The models enable the mango flowering triggering using the automated system in greenhouses and support the management of mass production in greenhouses as such a project in UniMAP with 50 greenhouses that consist of 200 trees each greenhouse.
- Wireless Sensor Network-based system implementation for greenhouse microclimate data accumulation
- A new data set of Harumanis mango trees and the environment through feature extraction of biotic and abiotic factors that enable the prediction of Harumanis mango flowering prediction. The factors are First Whorl Length (LFW), Second Whorl Length (LSW), Whorl Diameter (DW) and lysimeter for biotic attributes