Plant Varieties Selection System

Kitti Koonsanit, Chuleerat Jaruskulchai, Poonsak Miphokasap, and Apisit Eiumnoh

Abstract—In the end of the day, meteorological data and environmental data becomes widely used such as plant varieties selection system. Variety plant selection for planted area is of almost importance for all crops, including varieties of sugarcane. Since sugarcane have many varieties. Variety plant non selection for planting may not be adapted to the climate or soil conditions for planted area. Poor growth, bloom drop, poor fruit, and low price are to be from varieties which were not recommended for those planted area. This paper presents plant varieties selection system for planted areas in Thailand from meteorological data and environmental data by the use of decision tree techniques. With this software developed as an environmental data analysis tool, it can analyze resulting easier and faster. Our software is a front end of WEKA that provides fundamental data mining functions such as classify, clustering, and analysis functions. It also supports pre-processing, analysis, and decision tree output with exporting result. After that, our software can export and display data result to Google maps API in order to display result and plot plant icons effectively.

Keywords—Plant Varieties Selection System, decision tree, Expert Recommendation

I. INTRODUCTION

VARIETY selection is one of the most important decisions a producer makes each year for each plant area. It is the foundation of crop management plan. The variety which agriculturist choose will determine by the yield potential, the area, weather, and feature of soil, so spending time and effort on selecting the right variety for specific conditions is a good investment. Variety plant selection for planted area is of almost importance for all crops, including varieties of sugarcane. Since sugarcane has many varieties, variety plant non selection for planting may not be adapted to the climate or soil conditions for planted area. Poor growth, bloom drop, poor fruit, and low price are to be from varieties which were not recommended for those planted area. Therefore, Variety selection should be driven by environment. No one can recommend certain varieties as a general rule for everyone to plant. In this paper, we present a technique for plant species selection system using WEKA tool and Google maps API so that advanced data processing techniques in WEKA can be visualized by Google maps API. WEKA and Google maps API have become popular libraries in data mining and visualization on a web, respectively, because WEKA is freely

Kitti Koonsanit is with Department of Computer Science, Faculty of Science, Kasetsart University, Bangkok, Thailand e-mail: sc431137@hotmail.com

Chuleerat Jaruskulchai is with Department of Computer Science, Faculty of Science, Kasetsart University, Bangkok, Thailand

Poonsak Miphokasap is with National Center for Genetic Engineering and Biotechnology, Patumthani, Thailand

Apisit Eiumnoh is with National Center for Genetic Engineering and Biotechnology, Patumthani, Thailand

available and contains a large variety of algorithms. The main features of WEKA involve data mining, while the main features of Google Maps API aim toward visualization. Therefore, their integration can make the software more powerful.

II. RELATED WORK

A number of papers and articles have been published regarding the subjects such as recommender system for power plant operation [1] and agricultural selection system. One of the first one, introduced by Goh, K.J [2],[3] proposed fertilizer recommendation systems which are sensitive to the reliability of the input data for precise estimation of fertilizer rates and compromises such as maintaining large field sizes and skipping leaf analysis of some fields. A number of papers and articles have been published regarding the above subjects. However, none of the approach can be employed to incorporate in plant varieties selection system. Therefore, this paper focuses on the development and applications of the plant varieties selection system which was used to recommend the types of plants for plant area, which was used as one of the essential tools for varieties of plants selection in Thailand. This system was developed based on the environmental characteristics of Thailand.

III. WEKA

WEKA[4] is a widely used toolkit for machine learning and data mining originally developed at the University of Waikato in New Zealand. It is a large collection of state-of-the-art machine learning algorithms written in Java. WEKA contains tools for classification, regression, clustering, association rules, visualization, and data pre-processing. WEKA is open source software under the GNU/GPL. It is easily extensible, which allows researchers to contribute new learning algorithms to WEKA, keeping it up-to-date with the latest developments in the field. As a result, WEKA has become very popular with academic and industrial researchers, and is also widely used for teaching purposes. The main focus of WEKA is on classification algorithms. Each data instance is described by its attribute values. For example, clustering which are several different areas based on observations of sky, air temperature, humidity, and wind can be viewed as a clustering task. Each data instance includes values of the observation attributes, such as sunny, warm, humid, strong, and the available classes are {cluster1, cluster2, cluster3, cluster4}. The idea is that a clustering learned on an unlabeled data set can then be used to cluster around unlabeled class. All the above reasons, we choose and use WEKA to help analysis meteorological data.

IV. GOOGLE MAP API

Google maps API[5] is a web mapping service application and technology provided by Google, free (for non-commercial use), that powers many map-based services, including the Google maps website, and maps embedded on third-party websites via the Google maps API. It offers street maps, a route planner for traveling by foot, car, or public transport and an urban business locator for numerous countries around the world. According to one of its creators (Lars Rasmussen), Google map is a way of organizing the world's information geographically. As Google maps is coded almost entirely in JavaScript and XML, some end users have reverse-engineered the tool and produced client-side scripts and server-side hooks, which allowed a user or website to introduce expanded or customized features into the Google maps interface.

Google maps API allowed developers to integrate Google maps into their websites as shown in Fig. 1. It is a free service.
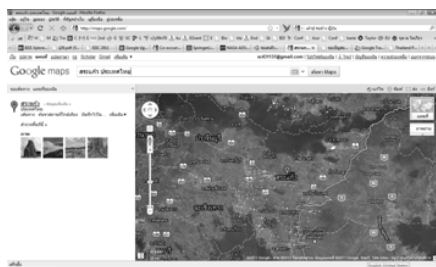

Fig. 1 Google map

V. MASH UP

A web mash-up[6] is a web application that takes information from one or more others sources and presents it in a new way or with a unique layout that combines data, presentation or functionality from two or more others sources to create new services. The mash-up is easy to implement, fast integration, frequently using an open APIs and data sources to produce enriched results. These features provide a unique way of web interface design. Moreover, following the development of visualization as web mash up, Josh Lowensohn[7] proposed a combination of maps and spreadsheets with Google's new mash up tool. This reason above that we choose and use a web mash-up to visualize classification results effectively.

VI. DESIGN

In the GIS (Geographic Information System) world are graphical representations of point data on a map through the use of colors and icons that indicate the representation of some variable such as the plant, plant area. Our software can quickly visualize the representation of locations. Being able to understand the label of point locations makes it much easier to see patterns in result data, especially when using different colors. In this session, we present to show processes for creating maps in our software.

Our software is divided into three distinct functions, including a function for importing meteorological raw dataset,

functions for classification dataset using WEKA library, which is a backend, a function for exporting the XML file, an initialization function for creating the Google map from the XML file, and finally a function to plot the map using the store points read from the XML file on top of the Google Map which is a visualization representation as shown in Fig. 2.
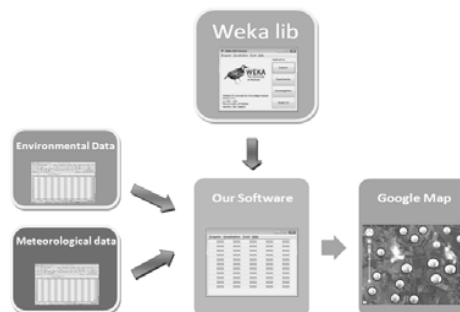

Fig. 2 Overall of System

VII. METHOD

A. Dataset

We used the one hundred sets of sugarcane varieties data that including environmental data and meteorological raw data from different stations[8]. The dataset are obtained from National Center for Genetic Engineering and Biotechnology (BIOTEC), a member of Ministry of Science and Technology of Thailand [9]. We would like to analyze data, which was registered in Thailand and thus try to recommend the approximate type of plant in Thailand.

In this paper, the data used to create our attributes include: day, average temperature, average pressure, soil texture, type of area, soil pH and accumulated precipitations. The next step of our method focuses on transforming these data in order to be used in Weka.

B. Preprocessing

According to our experience, we used evaluate on training data test mode for J48 classifier algorithm[4], that can only work with nominal variables, so we have to adapt the data to Weka format by a discretize filter, which is an instance filter that discretizes a range of numeric attributes in the dataset into nominal attributes. After choosing a discretize filter (weka.filters.unsupervised.attribute.Discretize), it can be applied to the initial dataset. The result of this transformation is shown in Table I- Table V.

TABLE I
TEMPERATURE INTERVALS

Temperature intervals	Represented by [a]
-inf-23	1
24-29	2
30-33	3
34-38	4
39-inf	5

[a] celsius

The average temperature values will be divided in equal intervals, each ones having attached a number: 1 – (-inf, 23), 2 – (24, 29), 3 – (30, 33), 4 – (34, 38) and 5 – (39, inf).

TABLE II
AIR PRESSURE INTERVALS

Pressure intervals	Represented by
-inf -1004	1
1005-1008	2
1009-1015	3
1016- 1019	4
1020-1027	5
1028- inf	6

The average air pressure values will be divided in equal intervals, each ones having attached a number: 1 – (-inf, 1004), 2 – (1005, 1008), 3 – (1009, 1015), 4 – (1016, 1019), 5 – (1020, 1027) and 6 – (1028, inf).

TABLE III
SOIL PH INTERVALS

Soil pH intervals	Represented by
-inf-1.0	1
1.1- 2.3	2
2.4-3.5	3
3.6- 4.2	4
4.3- 7.5	5
7.6-9.1	6
9.2-inf	7

The same codification is used for the soil pH values: 1 – (-inf, 1.0), 2 – (1.1, 2.3), 3 – (2.4, 3.5), 4 – (3.6, 4.2), 5 – (4.3, 7.5), 6 – (7.6, 9.1) and 7 – (9.2, inf).

TABLE IV
ACCUMULATED PRECIPITATION INTERVALS

Accumulated precipitation intervals	Represented by
-inf-88	1
89-185	2
186-279	3
280-393	4
394-531	5
532-657	6
658-inf	7

The accumulated precipitation is measured in mm. and used 1 for the (-inf, 88), 2 – (89, 185), 3 – (186, 279), 4 – (280, 393), 5 – (394, 531), 6 – (532, 657) and 7 – (658, inf).

TABLE V
EXAMPLE DATASET

Station	LO	LA	AT	AP	pH	ACP	. . . .	RP
1	13.45N	100.35 E	2	1	1	1	. . . .	A
2	18.47N	98.59E	3	1	2	1	. . . .	A
3	19.52N	99.50E	2	1	1	1	. . . .	A
4	12.40N	101.20 E	3	1	1	1	. . . .	C
5	14.48N	100.37 E	3	1	2	1	. . . .	C

where,

LO is longitude,

LA is latitude,

AT is average temperature,

AP is average air pressure,[10]

pH is soil pH,

ST is soil texture,

TA is type of area,

ACP is accumulated precipitation,

and RP= recommended varieties plant. [11]

C. Algorithm

After the input dataset is transformed in the format that is suitable for the machine learning scheme, we use J48 to build a classification tree to predict the recommended plant for areas in Thailand because the J48 model format as tree or rules is easy to understand. To predict the type of plant, we need to build decision tree based on the training data set. The results of decision tree as shown in the Fig. 3 and Table III.
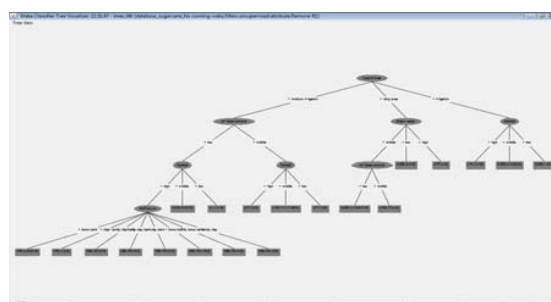

Fig. 3 Example decision tree

TABLE III
EXAMPLE DETAILED ACCURACY BY VARITIES

icon	variety of sugarcane	F-measure
	K76-4	0.8
	UT1	0.923
	CN-2	0.667
	KU50	0.8
	KU60-2	0.75

Furthermore, with the coding rules used for the database, these can be explained: In this example, out of a total of 100 instances, only 74 have been correctly classified meaning 74.00 %.

D. Display

To display the type of recommended plant on the Google maps, the result of recommended plant prediction is stored in XML form. We plot picture markers for interesting areas. Our

XML file, called Landmarks.xml, contains several landmarks in a format as seen in the Fig 4.

```
<markers>

<!-- E-tracking : Loading the data from an XML file -->

  <marker lat="43.65654" lng="-79.90138" label="Variety-A"/>
<marker lat="42.31892" lng="-78.89231" label="Variety-B "/>
<marker lat="42.42589" lng="-79.10040" label="Variety-A "/>
<marker lat="43.44654" lng="-79.90138" label="Variety -C "/>
<marker lat="43.78892" lng="-78.89231" label="Variety -A "/>
<marker lat="43.11589" lng="-79.10040" label="Variety -B "/>
<marker lat="42.39892" lng="-78.89231" label="Variety -B "/>
<marker lat="41.44589" lng="-79.10040" label="Variety -A "/>

......

</markers>
```

Fig. 4 Example XML file contains several dozen landmarks

Finally, this file contains elements for latitude and longitude which we will use to plot the marker along with one attribute elements, including the varieties of plant such as A- Variety, B-Variety and C-Variety. We are going to do in this simple application is read this file, plot out a new instance of GMarker using the <Longitude> and <Latitude> elements for each site, and displayed in Google maps as markers.

VIII. RESULTS

A. Experimental Result

This rule from our experiments can be used in recommender system. Thus, if we know environmental data such as average temperature, average air pressure, soil texture, type of area, soil pH value and accumulated precipitations and other data. Using this decision tree, we can estimate the future probability for the plant varieties selection.

Finally, we are displaying and exporting an auto-updating XML file that is clustered around results containing 100 points provided by plant varieties selection system. This system integrates WEKA and Google maps to deliver interesting area and locations to plant areas around Thailand. It is built on a web mash up, a web mapping interface that displays interesting locations to publish users as seen in the Fig 5.
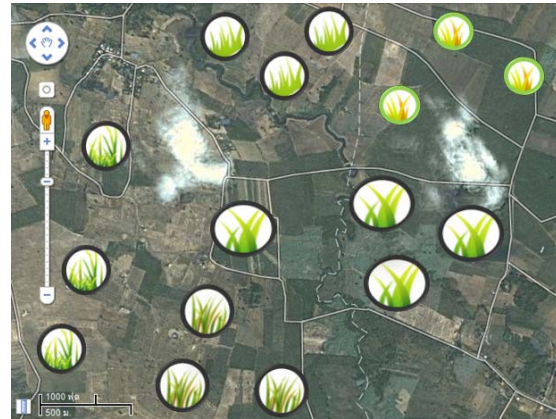

Fig. 5 Recommended Result image

IX. CONCLUSION

This paper presents a plant varieties selection system for planted areas in Thailand from meteorological data and environmental data by the use of decision tree techniques. With this software developed as an environmental data analysis tool, it can analyze resulting easier and faster. Our software is a front end of WEKA that provides fundamental data mining functions such as classify, clustering, and analysis functions. It also supports pre-processing, analysis, and selected output with exporting result. After that, our software can export and display data result to Google maps API in order to display result and plot labels and visualization in different plant icon on Google maps. We created our software using java ,developed on Netbeans IDE version 6.8 and tested the proposed process on meteorological data and environmental data in Thailand, and the experimental results show that our proposed process can export data result to Google maps and display classification results in order to help visualize its effectively. The outcome of this research will be used in further steps for visualization tools that help visualize meteorological data such as e-monitoring, e-tracking[12], natural resources, agricultural, environment and climate change.

A result of this research was developed to provide users process to visual classification result on Google maps. Many data mining techniques in WEKA libraries can be included within our software in order to make the software more useful and valuable. We hope that it can be used as visualization tool and help develop research in agricultural selection system.

X. FUTURE WORK

Future work, we will include the extension of the database with other important marketing parameters or weather parameters like sugarcane price, wind speed, wind direction or radiation. Beside this aspect, we can enlarge our database with records from other years not only from this year. Having all this improvements in mind, we think that it can increase the precision in building the decision tree and the recommended result based on it.

REFERENCES

[1] Reyes, A., Sucar, L.E., Morales, E. ",AsistO: A qualitative MDP-based recommender system for power plant operation," *Computación y Sistemas*, 13(1), 2009, pp. 5-20.
[2] ANUAR, A.R., GOH, K.J., TEE, B.H. and AHMED, O.H. "Transforming spatial-temporal yield maps to classified management zone maps for efficient fertilizer management of oil palm,". Am. J. Applied Sciences 5(10), 2008, pp. 1392-1396.
[3] Xu, L.; Liang, N.; Gao, Q.; "An Integrated Approach for Agricultural Ecosystem Management," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, pp 590-599
[4] Remco R. Bouckaert, "WEKA Manual," Retrieved: Sep 2, 2010, WAIKATO, Available from University, pp.1-303, January 2010.
[5] "Google Map API," Retrieved: Sep 2, 2010, Available to: http://code.Google.com/apis/maps/index.html
[6] LIN Hui and YE Lei, "A Mash-up of the Tracking System and Real-time Monitoring System Based on MAS," *IJCSNS International Journal of Computer Science and Network Security 2006*, vol.6 No.12, December 2006
[7] Josh Lowensohn, "*Combine maps and spreadsheets with Google's new mashup tool*," Retrieved: Sep 2, 2010, Available to: http://news.cnet.com /8301-17939_109-9704486-2.html
[8] "*Meteorological raw data*," Retrieved: Sep 2, 2010, Available to: http://www.tmd.go.th/climate/climate.php
[9] "*Image map 3 hrs of rainfall, temperature, temperature, relative humidity and MSL air pressure*," Retrieved: Sep 2, 2010, Available from http://www.arcims.tmd.go.th/dailydata/Map3hrs.php
[10] "*Accumulated Precipitation Data,*" Retrieved: Sep 2, 2010, Available from http://hydromet.tmd.go.th/Monitor/Forecast.aspx
[11] "*Sugarcane List Data,*" Retrieved: April 12, 2011, Available from http://sugarcanecenter.rdi.kps.ku.ac.th/
[12] K. Koonsanit, Chuleerat Jaruskulchai, "The Integration of WEKA and Google Maps in Visualization with a Web Mash-up of Deluge Risks Tracking in E-Tracking System," *In Proceeding of 9th international conference on e-business 2010 (iNCEB)*,2010, pp. 48-53

Kitti Koonsanit is a phd student studying a Ph.D. 's degree program in Ph.D. (Computer Science) at Kasetsart University, Bangkok, Thailand (now in the second years). He was born in Prachinburi, Thailand in May, 1983. In 1990, he studied a primary school in Sumretvittaya School, Aranyaprathet, Sakaew and 1996, high-school in Benchama ratcharangsarit School, Chachoengsao. In 2004, he received a bachelor's degree in B.Sc. (Computer Science) from Burapha University, Bang Sean, Chon Buri, Thailand. In year (2004), his senior project was about Java IDE. He received M.S. degrees in Computer Science from Kasetsart University in 2008. Mr.Kitti interests in computer science very much, his interested fields and research areas in Human Computer Interface, Image Processing, Computer Graphics, Internet Technology, Web Technology, Future Computer Technology, Assistive Technology and Computer Software.