

ECSI799 Software Engineering Project

Agricultural Decision Support System for Paddy

Literature Review

<i>Module Code :</i>	ECSI799
<i>Supervisor Name:</i>	Mr. Guhanathan Poravi
<i>Student ID:</i>	2014275
<i>Student Name:</i>	Sachitra Dilshan Kaluarachchi

Table of Contents

Overview.....	3
Problem Domain.....	3
Related Work.....	5
Tools and Technologies for Creating a Decision Support System.....	6
SQL Server Analysis Services (SSAS)	6
SQL Server Integration Services (SSIS)	7
Technologies and Approaches for Decision Support System	8
Centralized Data Warehouse	8
Ricardian Analysis	8
Microsoft Time Series Algorithm	9
Conclusion	11
References	12

Overview

This chapter includes the literature review conducted by the author for the project “Agricultural Decision Support System for Paddy”, which is a decision support system to help the industry or the government to take the best decision to understand the market situation and the minimum quantity requirement.

The author has studied some of the prevailing issues that can be solved by implementing a Decision Support System.

Finally the author has discussed about what technologies and approaches should be taken and are most suitable. The design and technologies that will be used in the solution is determined according to the research and discoveries made through this literature survey.

Problem Domain

As the world has become fast moving global village new technology has been applied for the industries to increase the profit and to keep the balance between demand and the supply. Traditional agriculture is known as an industry which does not apply new technology, but follows the same routine which comes from ages. But today the economy human needs and wants are rapidly changing. In order to face that there should be a way to predict or forecast the economical, demand, supply factors. In Sri Lanka there is no way to predict the future aspects.

In Sri Lankan current system there is no proper way to make the farmers aware about future market requirement considering both current and past economic, social and natural factors. There for they face numerous problems once they yield the harvest and try to sell that. There for the government also requires a systems which can forecast and provide accurate predictions considering unique Sri Lankan factors. The existing forecasting models is very primary where they just consider population and prices for the model.

For an example when there is high price for potatoes, all most all of the vegetable farmers cultivate potatoes which will decrease the price of the potatoes when they yielded harvest comes to the market and imbalances the demand and supply where the farmers won't be able to sell the production. Usually the demand varies seasonal wise in Sri Lanka. Therefore the supply or the production should be according to that demand which will keep the price balanced. To do this there should be an accurate prediction system which analyses the past demand, supply, production

and other relevant factors and provide most accurate forecasting for the farmers and to the government. And the effect of the sudden changes of the weather to demand, supply and production should be also forecasted before it imbalances the market. Further impose of tax or decisions on imports and exports of paddy should be verified. But currently there is no proper system to address these problems. Agriculture Decision Support System is the major component which will be the bridge to fulfill the gap between accurate forecasting and the market factors.

As it is mentioned above Agriculture Decision Support System provides factors to marketers, agriculture officers and farmers to make the wise decision to keep the market demand and price stable. The system will enable the area of decisions quantity to import or export, future market price considering present and past factors, future quantity demand, productivity...etc. For an example the government imposes taxes on importing paddy when the market is expected to be fulfilled with the local paddy harvest. If the production is not up to the expected quantity, the market price will be raised and the balance will be shattered. By the proposing system the government can have a prediction about the quantity demand and the supply. So the system will be an innovative advancement to the industry.

Most of time, it is said that the Sri Lankan paddy market runs monopoly and it is controlled by major businessman. This has occurred due to the lack of involvement by the government to help the farmers to take decision. For an example paddy harvest can be stored to sell in the future. But the farmers do not know about the future market aspects. So they sell the harvest to those major business branches even lower than the current expected market price. This is the major reason why the farmers cannot contribute their control to the market, because they don't have a proper accurate decision making support. This has led the paddy market monopoly in the country. Here it is to be mentioned that the corrupted political involvement also mainly affect for ant industry.

The system will fulfill a major requirement by the country to make future decisions based on the development of the agriculture. The system will enable a centralized database structure to create a data warehouse that will be used to do data mining and provide an accurate prediction module. A logical statistical analysis is done considering the econometrics factors that will be affected to the social, economic and environmental factors.

Related Work

There is no such kind of agricultural decision support system, in Sri Lanka which helps to make precise decisions.

“Govi Gnana Seva” project is a SMS based price updating project. This project is based on Dambulla, Meegoda and Narahenpita Dedicated Economic Centers Sri Lanka. Partnership between a not-for-profit company called Govi Gnana Seva (GGS) and Sri Lanka’s largest mobile operator (Dialog Axiata PLC). [1] When the farmer needs to know a price of a certain vegetable he needs to send a SMS with the number specified for that particular vegetable and then this system will reply with the price of that vegetable. This is a real time application, which provides the current prices of vegetables. It doesn’t provide any forecasting and also doesn’t provide rice prices. But this system provides the future prices of rice. [2]

In India there are two systems named “E-Krishi” and “E-Choupal”. Both these systems are created to help farmers. But both of these are not Decision Support systems.

“E-Krishi” is an IT project done in India to support their people in general agriculture. It passes crop market price, weather reports and new technology information via internet. In this project also it provides the current crop prices only, not future prices. It provides growers the new technologies that can be used, to get away from the traditional methods [3]

“E-Choupal” project in India provides relevant and real time information regarding commodity prices, local weather, and news. Consists with 3500 Choupals in 5 states of India covering 21,000 villages, servicing 2.0 million farmers. Sourcing a range of agri commodities like, Oil seeds, Grains, Coffee and Aquaculture. E-Choupal does not provide a forecast of prices, only the real time prices. [4]

Tools and Technologies for Creating a Decision Support System

In this section author discussed some well-known tools used in Creating a Decision Support Systems and there pros and cons.

Author is going to use MSSQL Database to store data. Furthermore, author is going to use SQL Server Analysis Services (SSAS) to build the warehouse and to analyse data. SQL Server Integration Services (SSIS) will be used for data extraction, data transforming and data loading.

SQL Server Analysis Services (SSAS)

SQL Server Analysis Services is an Analytical data engine which can be used in creating Decision Support systems and also in business analytics. SSAS provide the analytical data for business reports and client applications such as Excel, SQL Server Reporting Services (SSRS) reports, and other third-party data visualization tools. [5]

Data Analysing in SSAS can be done by creating workflows.

A typical SSAS workflow includes

- Building an OLAP or tabular data model
- Deploying the model as a database to an Analysis Services instance
- Processing the database to load it with data or metadata
- Assigning permissions to allow data access

By applying the data mining algorithms in Analysis Services to the data, trends can be forecasted, patterns can be identified, rules and recommendations can be created and insights on the data can be gained.

After the analyzing workflow is created, this data model can be accessed by any client application supporting Analysis Services as a data source. Which is beneficial when creating a Decision Support system.

SQL Server Integration Services (SSIS)

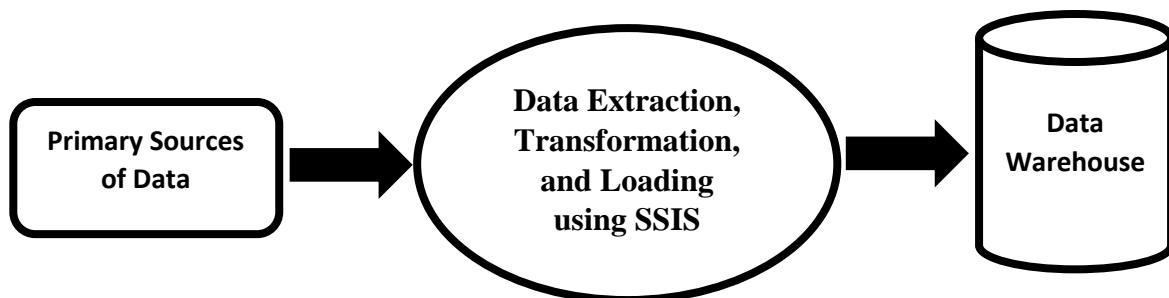
The primary use for SQL Server Integration Services is data warehousing because this is mainly a platform for building enterprise-level data integration and data transformations solutions. [6]

The packages can work alone or in concert with other packages to address complex business needs.

SSIS can extract and transform data from a wide variety of sources such as XML data files, flat files, and relational data sources, and then load the data into one or more destinations. This features a fast and flexible tool for data extraction, transformation, and loading (ETL).

SSIS can be further used to

- Automate maintenance of SQL Server databases
- Updating data warehouses
- Update multidimensional cube data
- Copying or downloading files
- Sending e-mail messages in response to events
- Data mining and data cleaning



Technologies and Approaches for Decision Support System

Centralized Data Warehouse

One major component of the system is the centralized data warehouse, which centralizes all the data for many years past. Data warehouse is a combination of multiple sources of data. Historical data is analysed in different dimensions, using hierarchies. Raw past data is collected based on different factors which need to be analysed. Some factors will not include a continuous data set and missing values has to be filled. Mainly, three steps were carried out in building the data warehouse.

- Extraction
- Transformation
- Loading

Extraction is the method of extracting the data from the data source system for further use and analyze in data warehouse environment. After data has been extracted, it transported physically to the target system for further analysis. Designing and creating the extraction process is often one of the most time-consuming tasks in the ETL process. [7]

Data transformation is the process of converting data or information from one format to another, usually from the format of the source into the required format of the new destination. Data transformation includes data conversions and data scrubbing techniques.

Data loading is a major task in data warehousing. In this research data loading is done using SQL import and export wizard and SSIS packages. The data can be retrieved through Excel sheets, SQL databases, Microsoft Access databases. Visual studio Integration Services are used to design the data flow the system. Therefor SSIS packages have to designed to import excel sheets. To covert the data matching to the data types of the database tables, data conversion tool have to be added in between the source and destination to cater any type of input.

Ricardian Analysis

Ricardian analysis is introduced to estimate the impact of climate change on agriculture.

One of the modules in this system is crop productivity forecast. Crop production depends on climatic, geographical, biological, political and economic factors. Major factor of crop productivity is the weather condition, which is an uncontrollable factor. Therefore it is needed to identify a pattern of the weather changes by analyzing historical weather data. A Ricardian analysis will be performed to check the effect of climate change on paddy.

$$\underset{R}{\text{Max}} \text{NR} = P_i * Q_i(R, E) - C_i(Q_i, R, E)$$

Here P_i and Q_i are respectively the price and quantity of good i . $C_i(\cdot)$ is the relevant cost function; R is a vector of inputs, and E reflects a vector of environmental characteristics of the farmer's land including climate. Given that the farmer chooses inputs R , to maximize NR , one can express the resulting outcome of NR in terms of E alone. [8]

Microsoft Time Series Algorithm

The combination of the source data and the prediction data is called a *series*.

The Microsoft Time Series algorithm provides regression algorithms that can be used for forecasting of continuous values. Unlike other Microsoft algorithms, such as decision trees, the Microsoft Time Series algorithm does not require additional columns of new information as input to predict a trend. A Time Series model can predict trends just only by using the original dataset that is used to create the model. Extra data can also be added according to the requirement. [9]

In this research, this Time Series algorithm can be used for minimum rice requirement.

The sample diagram below shows the prediction that was achieved using the Microsoft Time Series algorithm

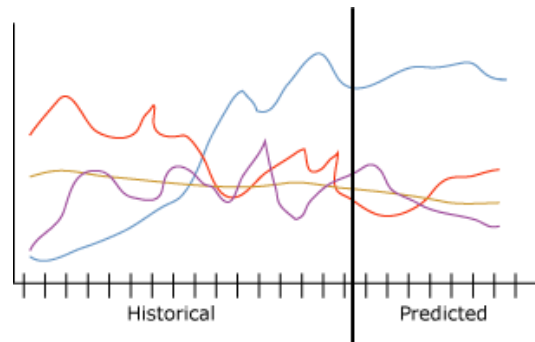


Figure 1: A Sample Time Series prediction

The Microsoft Time Series algorithm includes two separate algorithms for analysing.

- The ARTXP algorithm
- The ARIMA algorithm

The author is going to use ARIMA algorithm for the research project.

The below figure shows a sample ARIMA model parameters for this mining model.

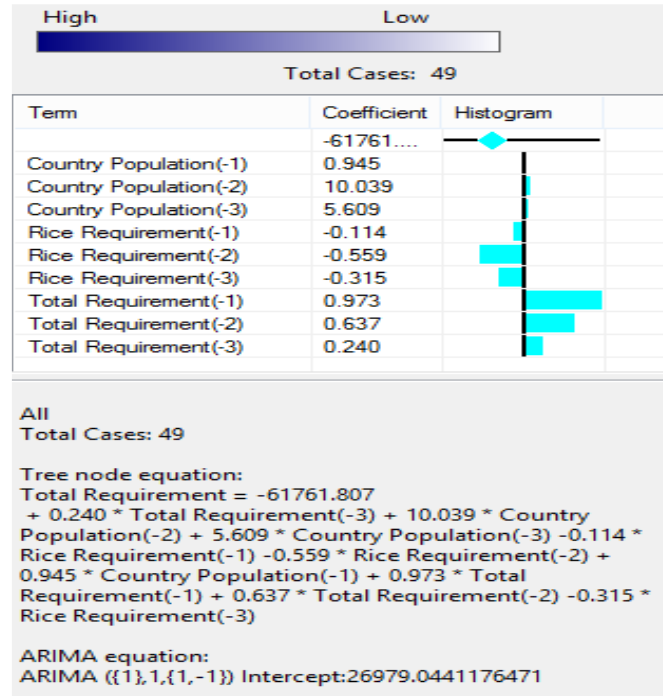


Figure 2 - A sample ARIMA model.

The minimum country requirement is given in two different aspects

- Minimum rice requirement
- Minimum seed requirement for the year.

Total requirement is the summation of minimum rice requirement and the seed requirement.

Conclusion

The system described in this document is totally self-contained and self-independent of any other system. This product is a solution for improper decision making and for the issue of less involvement of technology in agriculture sector.

According to the research conducted by the author there is no known tools which helps in taking decision regarding paddy. The author identified Microsoft SQL Server Analysis Services (SSAS) and Microsoft SQL Server Integration Services (SSIS) as the tools that can be used when creating the Decision Support System. Author also identified Ricardian Analysis and Microsoft Time Series algorithm suits for the algorithm which can be used in predicting values.

When developing such a Decision Support system, it is important to use a language which is widely used and which has enough documentation and support. The author reviewed the most widely used desktop platforms, which are .Net (C#) and Java. The author decides on .Net (C#) due to highly supported IDE and development language fluency.

References

- [1] S. Wijerathna, "Dialog tradenet – Agricultural Commodity Prices via Mobile Phones," December 2009. [Online]. Available: <https://sameeraw.wordpress.com/tag/govi-gnana-seva-ggs/>. [Accessed February 2016].
- [2] H. d. Silva, 2005. [Online]. Available: <http://www.share4dev.info/kb/documents/3260.pdf>. [Accessed February 2016].
- [3] "http://www.inruled.org/," [Online]. Available: <http://www.inruled.org/iERD/Cases/E-KRISHI.pdf>. [Accessed 02 2016].
- [4] S. R. Kuttayan Annamalai, August 2003. [Online]. Available: http://pdf.wri.org/dd_echoupal.pdf. [Accessed February 2016].
- [5] Microsoft, "Analysis Services," Microsoft, [Online]. Available: <https://msdn.microsoft.com/en-us/library/bb522607.aspx>. [Accessed February 2016].
- [6] Microsoft, "SQL Server Integration Services," Microsoft, [Online]. Available: <https://msdn.microsoft.com/en-us/library/ms141026.aspx>. [Accessed February 2016].
- [7] Oracle, "Extraction in Data Warehouses," Oracle, [Online]. Available: https://docs.oracle.com/cd/A97630_01/server.920/a96520/extract.htm. [Accessed February 2016].
- [8] R. M. M. M. Sung-No Niggol Seo, "Climate change and agriculture in Sri Lanka: a Ricardian valuation," [Online]. Available: <http://nersp.nerdc.ufl.edu/~vecy/LitSurvey/Seo.05.pdf>. [Accessed February 2016].
- [9] Microsoft, "Microsoft Time Series Algorithm," Microsoft, [Online]. Available: <https://msdn.microsoft.com/en-us/library/ms174923.aspx>. [Accessed February 2016].