Int. J. Advance Soft Compu. Appl, Vol. 10, No. 1, March 2018
ISSN 2074-8523

Big Data Analytics using Weight Estimation Algorithm for Oil Palm Plantation Domain

M. Pushparani, and A. Sagaya, S.Ravan

UNITAR International University, Kuala Lumpur, Malaysia
e-mail: ushparani@rcbuminiaga.com.my and sagaya@unitar.my

Abstract

Malaysia and Indonesia are the largest producers and exporters of palm oil and palm oil products in the region and in the world. Both country plays a significant role in contributing to global economy growth. There is always a constant demand for palm oil and its products. In order to cope with demand there are many measure and initiatives taken, managing operation is one of the way to address demands. On the other hand, big data analytics is one of the technology that is assisting many domains. The more data in hand the more accurate information and this would enable plantation stakeholders to make better decisions at any given point of time. This research focus on an approach towards big data analytics using weight estimation algorithm to assist in managing plantation operation effectively. This research emphasis on big data analytics as one of the approach to improve the operational efficiency by ensuring abundance of data in various forms and in various location made available in a single place thus empowering plantation stakeholder to make decisions with the accurate information. The weight estimation algorithm is used in a portable gadget to estimate the weight of the yield at the estate and any discrepancies are addressed with-in the manageable time. With the collected data, analysis is performed at plantation site and headquarters. The weight estimation algorithm is a combination of data collection, estimation of load and ensuring that data is available online on a real-time basis. Data analytics with the captured data will assist in effective decision making that attributes to reduced cost outages within a reduced time. The aim of this research is to create a big data analytics framework using weight estimation algorithm to capture data from various source and making it available on time for better decision making within the plantation estate and the mill.

Keywords: *big data analytics, weight estimation algorithm, artificial intelligence, business intelligence, oil palm plantation*

1 Introduction

Data is valuable resource in today's world, data can transform lives in form of disaster management and prediction, health care or any industry. Data is renewable resource of the information age just like the natural resource data should be managed with care and consideration to make meaningful information [1]. Having said data is a precious resource it is also a challenge in managing enormous amount of available data in various forms. The available data should have integrity and reliability to provide proper business insight that assists organization to make better business decisions from anywhere at any time to keep up with the competitive world. The system or tool that manages that data gets more complex depending the volume of data that has to be managed, data acquired by organization can make it wiser in terms of managing data but it can also disrupt business if data is not being analyzed using the right tools. Data has to be acquired, organized and managed for better insight and careful decision making [2]. Analyzing big data is a challenge as data exists in structured, unstructured and semi-structured formats. Structured data such as fixed fields, spreadsheets or files and unstructured such as image, video, audio files or webpages. Data is collected throughout the organization in various storage location and in various formats, relationship has to be established within data to discover patterns and draw meaningful information from it. [3].

The oil palm is scientifically known as Elaeis guineensis, it comes from the tropical rain forest of West Africa. Processing of oil palm fruits for edible oil has been practiced in Africa for thousands of years. [4]. The oil palm is cultivated in many other countries especially in the tropics, it is an extremely high in demand potential commodity in these regions [5]. The ideal composite of palm fruit bunches is 10 – 40 Kg/ bunch and individual fruit ranging from 6 to 20 grams [4]. The agricultural investments has increased in the Asian region and this has rapidly expanded into the plantation sector as this is an economy driven commodity [5].

The vegetable oil extracted from the palm fruit is in much demand and Malaysia is one of the largest producers of palm oil. Some of the researches conducted in oil palm industry are in areas like image processing on leaves decolorizing, soil quality, to determine the ripeness of FFB and many more. The palm bunches are given more emphasis as this is the raw material for the palm oil products. The plantation operations are manages using plantation management software and the operations at the mill are generally manages using the mill management software. These two systems are used to capture the daily operations at the planation.

2. Problem Definition

Generally, oil palm plantations are located in remote areas with less or no network facilities. Data is collected in various forms through using various system located within the plantation. The usual practice is that the data are captured from various sources and stored in various locations, and consolidated. Decisions are made based on the secondary data, and this consolidated data, which is usually associated with a time lapse. The existing plantation operation has limited real-time data availability, data is stored in various devices and in many locations within the plantation. Data consolidation from these devices has delay because of lack of network infrastructure in remote location. Decisions made using data that are about one week's old are not effective and works performed with outdated data will be redundant [6].

Data entry is performed manually or in a semi-automated manner at the weighbridge. This is time consuming because verification is done at weighbridge by data entry staff manning the weighbridge. The weight of the lorry is measured twice, once with the lorry carrying FFB load and again after the lorry has unloaded FFB load at the mill. Vehicle number is a unique identification number that is used to match the vehicle in order to get the net weight of the FFB load carried by the specific lorry. Since this matching of vehicle number is done manually, there are high chances of mismatching vehicle numbers accidently or incidentally wrong data entry. Example a 3-ton lorry's weight is mapped to a 5-ton lorry weight this could be a loss to one estate and gain to another estate. This mismatched charging may appear to be small for a single entry but when this incident is repeated many times then the accumulated loss would be high.

Oil quality depends on FFB sent to the mill on time, there is a possibility of leaving behind harvested FFB at the field and not sent to mill. These leftover FFB could either be stolen or left to be rotten and possibly unaccounted FFB, this could be due to lack of data availability on time and delay in informing the operational staff to pick the FBB. Although the yield is good but due to lack of real-time data there are many uncounted bunches adding onto the wastage.

2.1. Research Objective

The objective of this research is to create a framework for data analysis using weight estimation algorithm that would assist in collecting consolidated data from primary data source to perform data analysis in order to make effective operational decision reducing or eliminating rework. The better the plantation are equipped with data the more time could be spent on improvising yield thus contributing to the Malaysian economy in oil palm plantation industry. The need for this research is to have data analysis performed with real-time data at the plantation site and also assist in providing data patterns and facts that would assist in effective plantation operation decision making.

This research aims to create a framework for big data analysis using weight estimation algorithm, data analysis is performed using the consolidated data obtained from plantation management system and weighbridge system. There is a huge potential in creating this framework for big data analysis because it can be used for many other crops other than oil palm and this is a significant contribution of my research.

To use weight estimation algorithm in plantation management system and weighbridge system will assist in achieving real-time data. This will be my contribution to the information technology body of knowledge as an element of collecting data from various source in various forms.

3. Related Research

The research area that were covered for this research is big data analytics, business intelligence and machine learning.

3.1. Plantation Management System

There are several existing plantation management systems that manages the daily operation in the plantation however for this research a few of them were studies and the one that was used for this research was iPlantation management system because the modular based development it was easy for accessing the relevant areas to embed the weight estimation algorithm in the system. The iPlantation is an operation management software that records all the daily operation carried at the planation site / estate. There are several modules in this plantation application such as Inventory/ Store/ Procurement keeps track of the received and consumed stock, all stock related information such as stock level, stock valuation and calculation of average price based on the purchases, ageing analysis, tracking of stock items that are issued. The fertilizers and pesticides that are used within and inter estate, tools and the other items that are stored in the inventory. [7]

3.2. Big Data Analytic

BDA is a process of examining huge amount of data used for performing analysis, this process helps to hidden patters and unknown relationship that will assist in making better decision. BDA is used in various domains across many industries and organization to improve their operations through better decision-making abilities.

- BDA is performed in a wide range and is growing exponentially now zeta bytes of data are collected on a daily basis and analysis are drawn to perform necessary action. Data is retrieved from the system and weather the system

has real-time data or delayed information is not specified. The significance of collecting data from the primary source of data is indicated in this research. [8]

- A research on big data stated that although big data has a huge influence on many researches carried out today, it usually focusses on structured data for predictive analysis it has ignored the largest component of unstructured data available in form of video, images and unstructured text. One of the major characteristics is the data size, there should also be on the frequency of data capture to have more accurate data analysis. [9].
- In one of the reports from Gartner he states that information is growing rapidly with a minimum of 59% and rough estimation of 85% of data are unstructured and are not readily available for data analysis as stated in the online magazine [10]. The end users are given a quick breakthrough to have an insight of billions of rows of data using PowerPivot built in Microsoft Excel is an application for power users to perform data visualization.

3.3. Business Intelligence

Business intelligence (BI) is now widely used and described as analytic applications. BI is a process that comprises two main activities such as getting data in referred as warehousing that involves moving data from source systems into an integrated data warehouse (Watson & Wixom, 2007). The value of this data is determined only when users and applications access this data and use it to make decisions. BI usually consists of users and systems accessing the data from the data warehouse to perform enterprise reporting, OLAP, querying, and predictive analytics.

- There was a research in India [11] on oil palm plantation to provide optimized soil condition to maximize yield. The system consisted of many databases, Business Intelligence, Knowledge Management and user interface. The objective of this research was to compute the soil water content change as an important ecological index with respect to time. The integrated information system extract data that was used for agriculture decision making but this research focused mainly on the water retention and no evidence of integration was stated in the research.
- The business intelligent journal by [12] states that the existence of BI was there from the 1988 using the data warehouse architecture and 1993 layered data warehouse architecture replaced it. The business needs for the 21st century was stated as a concept of operational BI and unstructured content analytics. It also described that knowledge density is perceived based on the usage and timeliness and the level in which data is granulized. It also describes that an architecture with three layers such as information, process, and people. The business integrated insight architecture describes the single, consistent and integrated set of information used in an organization for timely operational and strategic decision making.

- In a whitepaper by Devlin[13], the roadmap of the existence of a data warehouse and how it has changed along the years based on the decision-making necessity is described. The study also describes that BI architecture has three main components such as business information resources, business function assembly and personal action domain. The case study was conducted to reduce the current costs of enterprise-wide data infrastructure by diverting investment into a new opening that will enhance information usage and derive better decision-making throughout the organization.
- A study conducted in Indonesia and Malaysia on best management practice [4] to maximize yield, by used BMP to address the inefficiencies during various stages of plantation operations. The FFB production from the BMP implemented location had better results compared with the non-implemented location.

A research in India described the life cycle of BI and various phases that are involved in developing BI systems[14]. It further elaborated that there are many ways to apply tools and technology to better understand the insight of the operation in a business by gathering information from disparate internal and external data sources; the attained data are then interpreted with various data analytics to enable more accurate decision-making. People from various sectors have perceived BI in different forms, but the key is to gather information from disparate systems, and consolidate and perform data analytics to support decision making. Small and medium enterprises are looking out for relevant information and all types of business opportunities. The research highlights that BI is crucial to performing business activities effectively.

3.4. Weighbridge System

The weighbridge system has been used by various industries for load sensing by using both a portable and fixed weighbridge located at a strategic position to monitor a vehicle load. The weighbridge system is used to measure the load of vehicles passing through the weighbridge and stores the data.

- In Tanzania, research was conducted to check on vehicle overload on the trunk road [15], the study showed a significant number of vehicles were carrying overload on these roads. Due to this, the trunk roads were damaged and rework was often called for. To avoid such road damage, the weight of the vehicles was measured, and whenever a weight was detected above the specified range given by the authorities, an alarm was triggered by the control system. The owners of the overweight vehicles were fined. The data obtained were used at that location to raise alerts; however, it was not integrated with any other reporting system, nor was it transferred to other site for further analysis.
- Research was conducted in India on weighing in-motion [16] to capture the

weight of a vehicle while on the move and not stationary. The weighbridge system has the vehicle stationed on the weighbridge scale during the weight capture. But in-motion weighing determined the weight of the vehicle based on the velocity of the vehicle and the number of axles, as it passes through the weighbridge. The record of vehicles passing through the system were stored and later viewed by the authorities. This research captured the aggregate weight of the vehicle and it was a proposed solution at the toll-plazas, since the focus was on number of axles and its approximate weight.

- A research was conducted in Malaysia [17] to automate in-field weighing of FFB using a one-ton trailer. The research conducted linked the FFB with the location and was used to track the FFB movement and generate reports at the plantation. The objective of the study was to automate the in-field weighing of the FFB. Data obtained were used to track and monitor the FFB evacuation and to prevent FFB theft in transit. However, they did not attempt to integrate with any other plantation system.

3.4. Machine Learning Language

The weighbridge system has been used by various industries for load sensing by using both a portable and fixed weighbridge located at a strategic position to monitor a vehicle load. The weighbridge system is used to measure the load of vehicles passing through the weighbridge and stores the data. The importance of learning is classified into animal learning and machine learning stated book about machine learning and the algorithmic perspective. Animal learning is all about remembering, adapting and generalising and animal learning is based on a few factors like the last time when we were in a situation and how was our particular action and did it work. The generalising factor is to recognise similarity between different situations and could be used as knowledge for later on. As computers came along Artificial Intelligence (AI) were introduced as symbolic processing because the computer manipulates symbols that reflects difference environment. In contrast machine learning are sometimes called sub symbolic because no symbol or symbolic manipulation are involved in this [18]. Machine learning is all about making computers modify and adapt to action in order to get more accurate towards the correct once. There are many type of machine learning supervised, unsupervised, reinforcement and evolutionary learning. The machine learning process depends on data collection and preparation, feature selection, algorithm choice, parameters selection, training and finally evaluation.

The area supervision machine learning classification algorithm are;[19] linear classification, that includes logical regression, Naive Bayes, perceptron, support vector machine. Quadratic Classifier, K-means Clustering, Boosting, Decision Tree with Random Forest, Neural Network and Bayesian Network. Some of the algorithm that were reviewed for my research as described as below;

- Naïve Bayes Algorithm – when there is a difficulty in classifying webpage, document and email or any other text based files then this algorithm is used. A classifier is a function that allocates a populations elements value for the available category. This model is particularly used for decease prediction and document classification and more for subjective analysis. [20]
- Artificial Neural Networks – The fundamental processing element of neural network is in neurons. The building block of a human brain encompasses some capabilities, the neurons receives input signals from other sources and combines is a way to perform general nonlinear operations and output the final result.
- The neural network is used in artificial intelligence performs like the nerve model in a brain. The model consists of 3 layers, input layer hidden layer and the output layer. The input layer receives signal in raw data and combines with the weight value and passes it to the hidden layer. The multiplication of the weight and the input value called as simulation function, this hidden layer might consist of several layers within. The output will then be sent to the output layer, the output layer will determine the data accuracy and will be analyzed.[21]

In the year 1947 the early model of artificial neuron was introduced by Warren McCulloch and Walter Pitts. It is called the McCulloch-Pitts neural model and also known as linear threshold gate. It is a set of neuron inputs and one output. The internal state of the weighted sum of inputs and the biased net input is the input minus the threshold and the obtained output is the value achieved by applying based on the activation function to the biased net input. [22].

4. System Design

This research was conducted in plantation domain at Malaysia and Indonesia oil palm plantations. During the initial stage of this research discussions were conducted with the plantation managers, field operators and clerks along with the plantation domain expert. The discussion was carried in the oil palm plantation site and mills, the participants brought out their concerns and challenges they face in carrying out their daily operations. This was then formulated into structured problems statements as stated below;

- same data is capture into various system and this leads to data redundancy
- data entry is performed with a delay and at times data is not accurate leading to unreliable data and data inconsistency
- data analysis is performed using the consolidated data and there is a delay in performing this exercise.
- weight is determined at weighbridge and it is hard to track harvested FFB thus leading to unaccounted or wastage of FFB.

The literature review and background study was conducted around the major areas of concerns. The conceptual model was created based on the knowledge acquired through the various process and methodology that were evaluated and studied as discussed in the related studies. The conceptual framework consists of two systems, iPlantation and weighbridge system; data captured at the field is captured and synchronized with the two systems to resolve the data redundancy, inconsistency and delay issues. The aim of this research is to address the data capturing process which will capture data from various location and in various forms, data capture is based on process control. Because of the interrupted network connection, a portable device will be used to capture data and synchronized this data with iPlantation and weighbridge system. This will enable plantation users to have up-to-date information at any given point of time. Data analysis is performed using the primary source of data at the plantation site and decision made using this primary source of data could be reliable compared with the secondary source of data with delayed analysis.

This scope of this research is to focus on oil palm plantation located in oil palm plantation located in Malaysia and Indonesia. The research is used to capture data from various sources in various forms both structure and unstructured. This data will be used to perform data analytics compelling some of big data attributes such as "variety", "varsity" and "velocity". This research focus on getting data from various sources in various form, data in various forms are mapped with the related and relevant data to form meaningful information. Data analysis is performed using this real-time data, data extraction is carried on regular basis using the data modeling and cubes that are built.

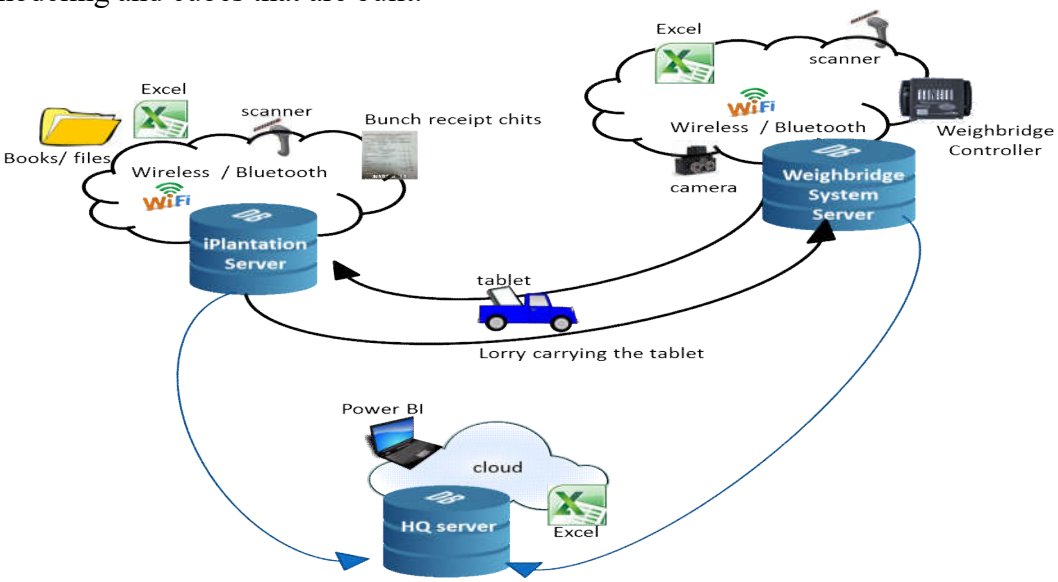

Fig. 1 BDA-WEA System Architecture

The BDA-WEA system will capture information from various control source and connects relevant data before storing it into the system. The weight estimation algorithm is installed in the portable device that communicates with iPlantation and weighbridge system, collected data is then synchronize into both the systems. Data acquired based on this the framework is then used to perform data analysis to provide real-time analysis for better operational decision making.

In order to perform this research there were a few things that were engaged for consideration to name a few are infrastructure, data collection process and data storage as shown in figure 1. That describes the overall scope of this research. Data stored in various forms in terms of files, books, barcode, print, are captured using the portable device and stored in iPlantation system. Hand held or the tablet that has weight estimation algorithm will capture the weight and estimate the weight of the load during the data capturing process. This tablet is carried around in the lorry and when it passes the estate office synchronizes the data using Wifi / Bluetooth connectivity. The lorry with the tablet reaches the weighbridge captured weight data from the weighbridge controller this is then mapped with the data in the tablet and synchronized with the server at the weighbridge. Later this data is transferred to the headquarters to have real-time data, since it has few instance of data entry there is less possibility of data being mismatched or faulty data capture. The data analytics can either be done at the estate/ weighbridge or done at HQ. The research has included both Microsoft excel and also PowerBI tool for data analytics. Data analytics at site will enable the plantation managers to have a better understanding on the current existing situation and better decision can be made instantly for effective operations.

5. BDA-WEA Design

The design comprises of two components, one is Big Data Analytics (BDA) that will be performed based on data captured using Weight Estimation Algorithm (WEA) that is the second component. Data captured from various sources will be brought to a single location without losing data integrity to get real-time information in order to perform data analytics.

5.1. Field Experiment

The weighbridge system has been Experiment were conducted in the plantation site; the weight of the palm fruit bunches is determined at the collection area in the estate or fields. There are many variables that determine the weight of bunches, the most significant one is the age of a palm tree, soil type and weather condition to determine the wetness in the fruit. However, for this research the age of the palm tree and the weather is taken into consideration. In order to estimate the weight of the palm bunches there are some additional parameters that has to be

considered. When the lorry is loaded with the palm bunches the weight of the loaded palm bunches has to be determined. The algorithm has to determine the estimated weight. The mathematical model of a neuron by [18] is used for this research. With this small change things are made a lot easier and we can now define precisely what a McCulloch-Pitts neuron is:

- a cell which can output a 0 or a 1
- a number of excitatory/ causing inputs
- a single inhibitory / hindered input
- a threshold values

The McCulloch and Pitts mathematical model of a neuron as shown in figure 2, the input x_i are multiplied with the weight w_i and the neurons sum their values. If the sum is greater than the threshold θ then the neurons fire, otherwise it does not.

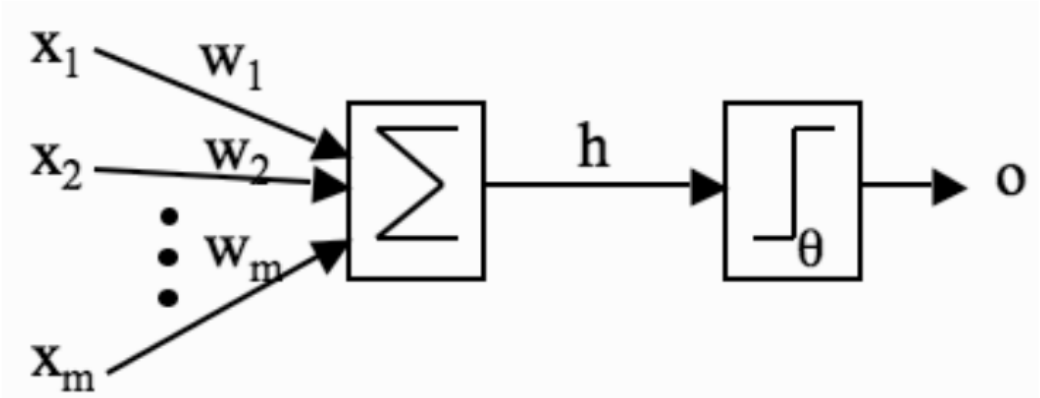

Fig. 2 McCulloch and Pitts Model of a Neuron

The x_i are the input nodes each of these input nodes will have its respective weightage w_i and will be multiplied. The output from these neurons are summed, it is then checked against the set threshold, if it is > 0 then the neurons fire, otherwise there is no action.[18]. The neural network has 3 layers, input layer, hidden layer and output layer. The situation of neuron in each layer will only effect neurons in the next layer. If the output layer does not provide the desired result then the neuron will be fired.

5.2. Algorithm Design

The algorithm is designed based on the preliminary research and literature review carried out with relevant area of research, the components taken into consideration are from various parameters that is used to capture data from various location. This weight estimation algorithm is installed in the portable device that can estimate the weight of the FFB bunches as it is loaded into the lorry. This weight

is later compared with the actual weight to calculate the accuracy and alert the right authority for better decision making. The algorithm is designed based on the problem statement that was derived during the initial stage of this research, the weight of the FFB bunches in several plantation management system are not derived until the lorry reaches mill. [6].

The oil palm estates are generally divided by division and block or also called as fields. The best management practice followed in the plantation domain is the naming conviction of the field that determine the year of planting, division and block of a plantation. The iPlantation system is a system that assist in capturing data on daily operation at planation estates. All the master data will be unique data are created and maintained at the headquarters for control purpose. This master data is synchronized at the plantation estate to store the daily operational data, this data is then used as references during the weight determination process.

This research aims to create an algorithm to estimate the weight at site based on the standards and compare with the actuals once the FFB is sent to the mill. Each harvester will record details such as the number of bunches, they field from where the harvesting is done, which team the harvester belong to. The information regarding the field will determine the YOP which will in turn determine the average weight of the FFB bunch. This data will be collected at the collection point and stored using the portable device, then when the lorry reaches the mill the actual weight is measured using the weighbridge. The algorithm will specify if there is a weight discrepancy, if there is a weight difference ± 5% then it is within the tolerance level, should the weight range is out of the tolerance level then alert is generated based on this action can be taken by the authorized person at the plantation.

The plantation standard practice is the plant the entire field during the same period of the year, this is one of the best management practice that is engaged for many year in the oil palm plantation domain for easy tracking [4]. The average bunch weight is 23 to 27 kg. Each collection point will have more than one harvester's details all these data are collected and scanned into the system and a print out with the barcode is printed and left at the collection point as proof. Approximate weight will be determined based on the data provided by the harvester's chit. The harvester's data is used for many purposes, one to calculate the salary and also to distribute the operational cost in terms of lorry, fuel and rest of the operational cost.

The standard reference table for weight is retrieved from the historical data, this data is obtained from the average bunch weight based on the actuals and are updated on the regular basis in both weighbridge and iPlantation systems.

5.2. Modelling Process

The model design process used for this research is to estimate the weight of the load, the weight of the palm bunch is the dependent variable W_i and the number of bunches, the average weight based and on the block where the bunches are harvested are independent variable and is described as a_i and b_i respectively as shown in equation [1].

$$W_{i=} a_i, X\ b_i \quad [1]$$

The weight of the palm bunches is determined based on the number of palm bunches times the average bunch weight that is determined based on the year of planting. Each collection point referred as 'harvesters' will have more than one harvester harvested bunches, the number of bunches loaded in the lorry will have to have information of a number of bunches and the collection point that determine the block details to extract the age of the tree.

Collection Point (1) = number of bunches (1) X bunch weight based on the year of planting (1)

Collection Point (2) = number of bunches (2) X bunch weight based on the year of planting (2)

Collection Point (n) = number of bunches (n) X bunch weight based on the year of planting (n)

Number of collection point is 'h', number of bunches is 'b' and average bunch weight is 'a', the equation [2] gives the total of each individual harvester's weight.

$h_1=(h_1a_1*h_1b_1)+(h_1a_2*h_1b_2)+\ldots..+(h_1a_x*h_1a_x)$

$h_2=(h_2a_1*h_2b_2)+(h_2a_2*h_2b_2)+\ldots\ldots+(h_2a_x*h_2b_x)$

….

$h_n=(h_na_1*h_nb_1)+(h_na_2*h_nb_2)+\ldots\ldots+(h_na_x*h_nb_x)$ [2]

$$h_x = \sum_{i=0}^{n} (b_i a_i) \quad [3]$$

The individual harvesters total bunch weight is calculated using the equation [3], this describes the total number of blocks/ fields harvested using the variable 'i', for every block the total number of bunches 'b' and the bunch weight is 'a'. The equation used to estimate the weight of the load is described in equation [4]. The weight W_e = to the number of harvester's individual weight, the total number of

harvesters at that collection point 'j' and the individual harvester is 'h_x' the summation of all the harvester weight at that collection point loaded into the vehicle will give an estimate of the FFB loaded into the vehicle.

$$W_e = \sum_{j=0}^{n} (h_x) \quad [4]$$

The actual weight of the load is determined from the weight bridge controller 'Wa', this is then compared with the estimated weight 'We'. The difference 'd' is calculated using the estimated weight and the actual weight as shown in equation [5].

$$d = W_e - W_a \quad [5]$$

The linear threshold gate model as shown in the below figure 3, the neurons is a set of inputs in $\boldsymbol{h_1, h_2 \ldots h_j}$ and one output $\boldsymbol{y}$, the threshold gate classifies the set of inputs into two different classes. Thus, the output is binary and this equation is mathematically represented as in equation [6]

$$sum = \sum_{j=0}^{n} (h_j w_j) \quad [6]$$

The weight 'h', rain 'w' is the rainfall as a factor, if there is rain then the weightage is added. The threshold 't' is calculated using equation [7] based on the tolerance level and if the threshold is '0' then there is no action and if the threshold is '1' then action taken to send alert.

$$sum = \sum_{j=0}^{n} (h_j w_j) = \begin{cases} 1 & \text{If } t=1 \\ 0 & \text{if } t=0 \end{cases} \quad [7]$$

The total weight estimated algorithm is designed to capture the data, synchronize and also alert based on the difference in weight, if the difference is within the range or it is out of range.

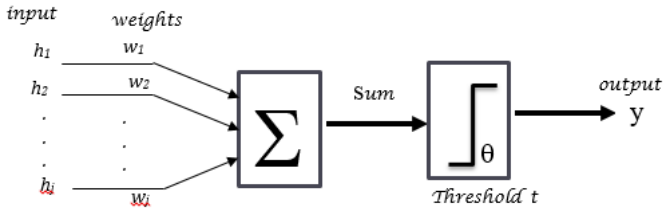

Fig. 3 Topological structure of a simple three-layer feed

6. Big Data Analytics

The big data analytics design comprised of structured and unstructured data from various source within the estate and the mill. This data then goes through the ETL process and stored in data marts. The cubes are created, using star schema then is then pushed into the presentation layer for data analytics. Online Analysis processing (OLAP) analysis is used to perform analysis. These analysis that are on a real-time basis will enable better decision making as described in Fig 4.

- Multi sources connection – data from various source were acquired from the estate office, field and mill
- Data Cleansing and Data Transformation – the data was combined without losing the data integrity and transformed into meaningful data when synchronized with the iPlantation and Weighbridge system
- Data modelling – cubes and dimension were built using the captured data
- Visualization (Interactive Reports and Dashboards) – Dashboards and BI reports were generated for plantation users and managers to make decision.

ETL is a process in data warehousing, this process is used for extracting or attracting data out of the source systems and placing it into a data warehouse.

The data mart is a subset of the data warehouse and is usually specific to a certain line or topic, data warehouses collection of data marts and have a wide depth of data reserve, the information in data marts pertains to the domain.

The star schema in data analytics is a style of data mart representation, this shows the relationship and connectivity of data. The schema is used to develop dimensional data marts in data warehouse. The star schema consists of one or more fact tables referencing any number of dimension tables.

OLAP is the technology used in many business intelligence (BI) applications. this technology is responsible for data discovery, report viewing and complex analytical calculation

The data captured from the weighbridge controller and the portable device that carries information from the estate this synchronized data is used for data analytics to have better decision making.

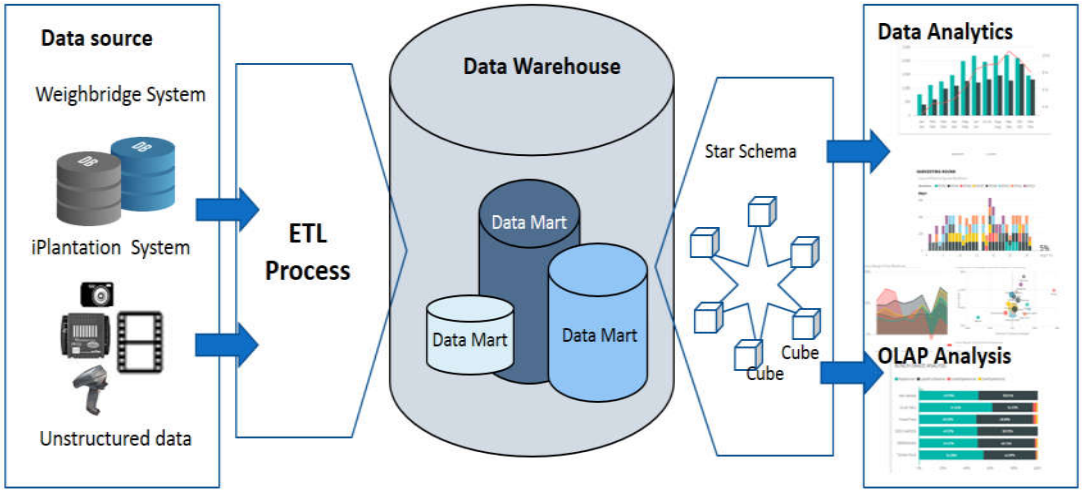

Fig. 4 Architecture for Data Analytics

7. Evaluation

The pilot study was conducted at the estate, preliminary discussion and interview were conducted with the data entry staff, the operation working staff, supervisors and estate managers. Based on the interview and initial discussion questionnaires were designed and given to the same group. There was set of questionnaires used for developing the prototype, the group of people involved during the initial discussion were used as respondents for these questionnaires. The group consisted of 20 respondent comprising data entry staff, managers and weighbridge operators. All the respondents stated that they have experience is using both plantation management and weighbridge systems. The group has more than three years of experience in using these systems.

Data Captured at the mill and the estate that are used for data analysis the online data analytics can also be performed if there is internet connection for the top management to look at the information at any point of time and make decision based on the data analytics. One of the dashboard from the on-line business tool as shown is figure 5. The tool that was used for this research purpose is Microsoft PowerBI and Microsoft Excel PowerPivot.

The prototype was designed and developed and the implemented at one of the estates at Indonesia. The sample data was extracted from the data capture for 2 months. The sample data is as shown in table below.

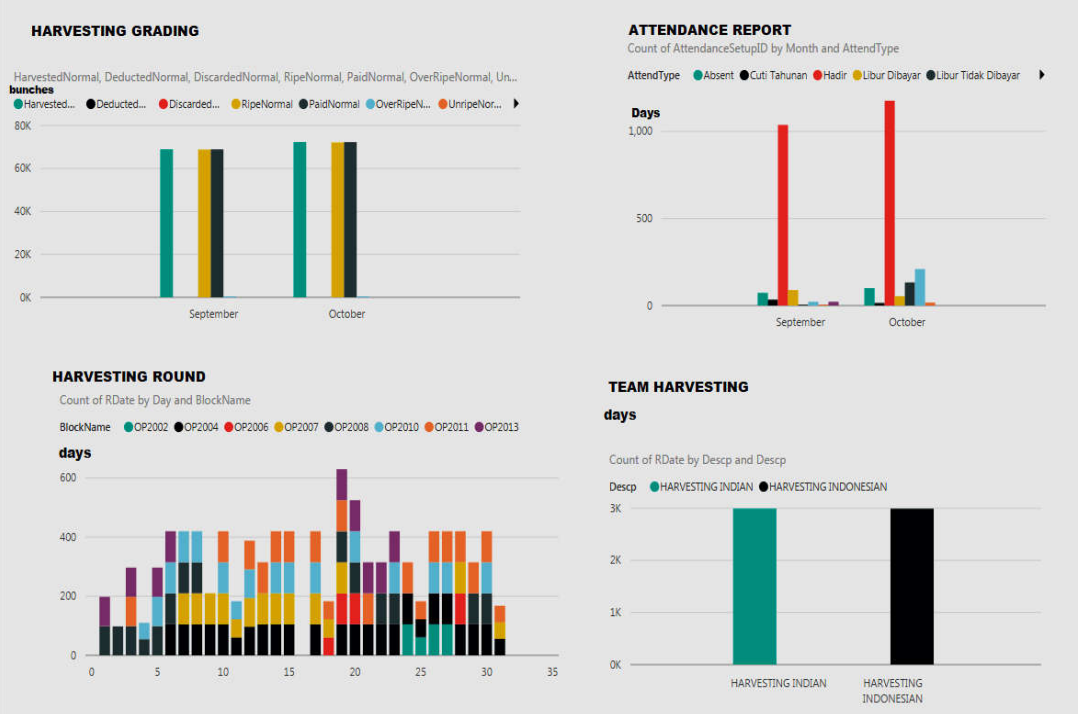

Figure 5: Data Analytics using on-line BI tool

8. Evaluation

The aim of this research is to provide a framework for big data analytics using the weight estimation algorithm. As stated earlier there is a lot of potential in creating data analytics as these analyses will provide real time business information at the estate for effective decision making. Palm oil being the commodity industry instant decision making based on the real-time available information will contribute to the economy by reducing the cost outages. The primary data for performing the data analytics will be attained by the weight estimation algorithm from the iPlantation and Weighbridge. The portable device with the weight estimation algorithm, estimate the weight and compare it with the actual weight of the FFB being carried by the lorry. This approach will remove the general latent way of data capturing done using various legacy system with some integration, this leads to data inconsistency and ambiguity. In order to get timely information, data has to be transferred to data presentation state at any instance without causing delays. The big data analytics will provide accurate and real-time data for better decision making, with data gathering at the weighbridge system and the planation

management system, users such as mill managers, estate managers, mill operators and operation assistant will be able to get reports that will assist in addressing the current issues whilst waiting for the top management to assist them with countermeasures. Although the parameters used in the weight estimation algorithm could produce the desired result future research can be conducted to add in more parameters such as soil pH values and it can be combined with RGB color combination of the fruits. This research could also be used for other industries other than the oil palm plantation

Acknowledgment

The first author would like to thank RC Buminiaga Sdn Bhd for supporting this research.

References

[1] M. Advisory, L. Version, W. House, and F. Sept, "Data : a resource more valuable than gold ?," *The international Research Data Alliance convenes in Washington, D.C., Sept. 16-18.*, pp. 6–8, 2017.
[2] Wired, "Is Big Data the New Black Gold?," pp. 1–3, 2017.
[3] A. Sathi, *Big Data Analytics*. MC Press Online, 2012.
[4] K. Poku, "Small-scale palm oil processing in Africa," *Fao Agric. Serv. Bull. 148*, pp. 1–56, 2002.
[5] M. Gatto, "Land-use dynamics, economic development , and institutional change in rural communities – Evidence from the Indonesian oil palm sector," 2015.
[6] A. S. M. Pushparani, "Towards a Framework for Embedded Weight Comparison Algorithm with Business Intelligence in the Plantation Domain," *World Acad. Sci. Eng. Technol. Int. J. Comput. Electr. Autom. Control Inf. Eng. Vol10, No7, 2016*, vol. 10, no. 7, pp. 1243–1255, 2016.
[7] RCBN, "iPlantation," *Product Details*, 2014. [Online]. Available: www.rcbuminiaga.com.my.
[8] T. K. Das and P. M. Kumar, "BIG Data Analytics: A Framework for Unstructured Data Analysis," *Int. J. Eng. Sci. Technol.*, vol. 5, no. 1, pp. 153–156, 2013.
[9] A. Gandomi and M. Haider, "Beyond the hype: Big data concepts, methods, and analytics," *Int. J. Inf. Manage.*, vol. 35, no. 2, pp. 137–144, 2015.
[10] Pressreader, "Microsoft SQL Server 2012 to empower Asian enterprises," *The Freeman*, 2012. [Online]. Available: http://www.pressreader.com/philippines/the-freeman/20120312/282583079953103.
[11] K. Babaria and M. G. Jacob, "Survey on Predictive analysis for formulating real time data in precision agriculture .," *IEEE Spons. Int. Conf. Innov. information, Embed. Commun. Syst.*, pp. 1–8, 2015.

[12] H. J. Watson and B. H. Wixom, “The Current State of Business Intelligence,” vol. IEEE Compu. 2007.
[13] B. Devlin, “Beyond business intelligence,” *Bus. Intell. J.*, vol. 15, no. 2, pp. 7–16, 2010.
[14] G. R. Gangadharan and S. N. Swami, “Business intelligence systems: design and implementation strategies,” *Inf. Technol. Interfaces, 2004. 26th Int. Conf.*, p. 139–144 Vol.1, 2004.
[15] A. Kumar, “A Review of Road Sector Reforms in Tanzania,” 2002.
[16] H. Kumawat and P. Jain, “In-Motion Weighing with Vehicle Data Collection System,” *Int. J. Eng. Res. Technol. ISSN*, vol. 3, no. 7, pp. 1237–1241, 2014.
[17] B. Liang, “Research on System of Process Information Integrating and Monitoring for Production Process of Metallurgical Industry Based on Web,” vol. 2, no. 2, pp. 2–4, 2016.
[18] H. Crc and S. Marsland, *Machine Learning: An Algorithmic Perspective*, Second Edi. Taylors & Francis Group, 2015.
[19] T. O. Ayodele, “Types of Machine Learning Algorithms,” *New Adv. Mach. Learn.*, pp. 19–49, 2010.
[20] Dezyre.com, “Top 10 Machine Learning Algorithms List of Common Machine Learning Algorithms,” 2017. .
[21] S. Abdullah and E. C. Tiong, “Prediction of Palm Oil Properties using Artificial Neural Network,” *J. Comput. Sci.*, vol. 8, no. 8, pp. 101–106, 2008.
[22] M. Hooker and M. Chua, “An Introduction to the Theory and Practice of Light Microscopy,” 2006.