Dairy Production Analysis and Prediction Tool using BIG DATA

Komal Patel¹, Ronak Chudasama², Sagar Dobariya³

¹,²,³Computer Engineering, UCOE.

Abstract—Accurate prefiguring of everyday dairy product production is a fundamental aspect of the dairy industry. During the past decades, although many models using numerous data analytic techniques have been suggested in the literature to address the dairy product production to prefigure the problem, these models have yet to be widely applied in daily operations. Given the increasing amount of milk production information collected every year, trouble also arises for studying big data. To deal with the challenges in dairy supply chains and help dairy producers, specifically for large-scale industry and to make use of data analytics in milk supply decision-making, a targeted effort is developed which is a feasible and cost-effective technique entitled as Dairy Production Analysis and Prediction Tool. For analysis proposed system uses MACD (Moving Average Convergence Divergence) and RSI (Relative Strength Index). The blend of MACD and RSI facilitate the monitoring of real-time progress of a production as well as futuristic analysis. For example, milk producer can use big data analytics to increase new visions about their customers in order to make the decision regarding sales and also to improve the speed of processing big amounts of organized sales data and to deliver data cost efficiently.

Keywords—BIG DATA, MACD, RSI, K-Means, Fuzzy K-Means.

I. INTRODUCTION

Businesses come in different sizes and scales and management of business inventory and supply chain is something that contributes to the business. Currently, ERP and CRM packages deploy a great deal of application frameworks that get the best out of the available market. Product lines like Salesforce, SugarCRM, etc. have shown to contribute to the business trends analysis. But the way these products operate and the cost of operation/licensing these products is very high and can be afforded by major businesses with an operating income of million dollars. The current systems deploy statistical analysis of the product supply and profits per unit. The analysis is supposed to have historic data so as to analyze any risk. So risk mitigation factors are completely out of spectrum in the current systems. The solution to these problems lies within the way the problem is solved. The problem of analysis needs to be done by using specialized algorithms and statistics that provide real-time analysis from historic data. Predictive statistical algorithms fall under this category. The technical indicators used in share market can directly be used here. The MACD (Moving Averages Convergence Divergence) and RSI (Relative Strength Index) are used together to provide a fuzzy function based output. The technical indicators aforementioned when used alongside a fuzzy function yield optimal results as evident from apps like TrackMyShare, Technical Indicators. Business processes can be easily tracked using the MACD and RSI and dairy product strength in the market is shown in the proposed system. The system is intended for analysis and prediction of sales in the dairy products distribution system. The Supply Chain Management (SCM) system includes Manufacturer-Distributor-Retailer-Customer. With this scheme, the business relationship in supply chain management system is digitalizing for the purpose of order placement and delivery. The project aims to reduce the manual workforce required to collect and process the orders from all the various retailers and to cover all the interpersonal communication.
II. EXISTING SYSTEM

In small scale industry, manual work is required for collecting orders from the retailers and then compiling a composite order to be placed at the manufacturer site. They basically maintain a record of orders and sales in a database. Depending on the status of the database they perform analysis manually and make a decision. In large-scale industry, ERP and CRM packages use a great deal of application agendas that get the best out of the available market. Product lines like Salesforce, SugarCRM, etc. have displayed to give to the business trends analysis. But the way these products operate and the cost of operation/licensing these products is very high and can be meet the expense by major businesses with an operating income of million dollars. The current systems deploy statistical analysis of the product supply and profits per unit. The analysis is supposed to have historic data so as to examine any risk. So risk mitigation factors are completely out of range in the current systems.

III. PROPOSED SYSTEM

The system aims to reduce the manual workforces necessary to collect and process the orders from all the various retailers and to cover all the personal communication between retailer-distributor-manufacturer. This approach digitalizes the Manufacturer-Distributor-Retailer business relationship for the purpose of order placement and delivery.

The approach involves
Purchase-Order management, which is a formal order placed by a customer to the retailer, retailer to distributor and distributor to manufacturer.
Inventory Management, at the distributor and manufacturer, end for maintaining the record for checking the products in stock.
Daily updates/notifications regarding order and cancellation of delivery of the product.
Data analysis for the improvement of sales and distribution.

MODULES
The system consists of three main modules which contain an android application, distributor's desktop, and the manufacturer is a desktop website. They are further categorized as follows:

Android application:
Login authentication.
Place an order.
Updates.
Order history.
Send feedback.

Distributor's desktop software:
Login authentication.
Check all orders received and acknowledge (approve/disapprove) an order.
Maintain inventory.
Update current product list.
Send regular notifications/updates to the retailer’s app.
Compile retailer orders and generate order and send to manufacturer.
Generate regular sales analysis reports for evaluation.

Manufacturer’s desktop software:
Region wise login access.
Check all orders received and acknowledge (approve/disapprove) an order.
Maintain inventory.
Update the current product list.
Send regular notifications/updates to the distributor’s software.
Generate regular region wise sales analysis reports for evaluation.
IV. ARCHITECTURE

4.1. SYSTEM ARCHITECTURE

The system consists of three main modules which contain an android application, distributor's desktop, and manufacturer's desktop website. For the placement of order, the retailer is provided with an android application which is easy to use. The retailer will be able to check the list of all the products available with all the detail and can add multiple products with multiple quantities to their shopping cart. The cart will be available to view and edit at any time. During checkout, the retailer has to view the cart and confirm his order. This order is sent to the distributor which is available on the desktop software and is stored in the database server of the distributor. An acknowledgment receipt is sent to the retailer's application as soon as it is received on the desktop software. The distributor checks and processes the order sent by the retailer with respect to his inventory and sends a summary of the final order accepted. The distributor compiles his daily total order with reference to all the orders collected from the various retailers and sends it to the manufacturer. The manufacturer has desktop software for processing the orders and the data is stored in the central database. The manufacturer processes the orders sent by all the distributors and with respect to the availability of the products generate the order to summarize and send them to the respective distributors with a confirmation receipt. The manufacturer generates and updates a central list of products with their details and sends it to the distributor. Further, the distributors generate a list of products with respect to the specific products he distributes and sends it to the retailer. The distributor is able to send daily updates of current rates and offers to the retailer's application. The retailers’ app will be able to send feedback to the distributor which will be received by the manufacturer and stored in the central database. All the data collected is used to provide sufficient analysis to improvise key areas of sales and distribution. Region wise and product wise sales analysis reports will be generated for evaluation of resources to be allotted by the distributor and manufacturer to improve sales, ease out distribution and help in decision making. The technical indicator MACD (Moving Average Convergence/Divergence) is used for current trends of the sale of product and RSI (Relative Strength Index) is used for overall change in trend and to know about the business loss or profit. For prediction, the K-means fuzzy algorithm is used in supply chain management (SCM) to predict the preferable decision for the business. Following figure, Fig.1 explain the system architecture.
4.2. ALGORITHM
In our approach, four algorithms are used. They are RSI (Relative Strength Index) and MACD (Moving Average Convergence Divergence), K-Means and Fuzzy K-Means. They are broadly explained below:

K-Means:
K-Mean Clustering algorithm is a preliminary way of deducing a large dataset into a manageable dataset. In a proposed system large dataset of historical data about sales is used, so it uses a k-means algorithm to break down it into the small dataset. It is an unsupervised learning algorithm. For this algorithm, the system defines k center which is a random integer so the data gets mapped along these k centers. In the end, the objective function is obtained by finding the Euclidean distance. The steps of k-means algorithm are as follows:

**Step 1**: Input Dataset, Clustering Variables and Maximum Number of Clusters (K in Means Clustering)
**Step 2**: Initialize cluster centroid
**Step 3**: Calculate Euclidean Distance. Euclidean is distance measures. Euclidean distance between an observation and initial cluster centroids 1 and 2 is calculated. Based on a Euclidean distance each observation is given to one of the clusters - based on minimum distance.

\[ \text{Euclidean Distance} = \sqrt{(X1 - Y1)^2 + (X2 - Y2)^2} \]

**Step 4**: Move on to next observation and calculate Euclidean Distance
**Step 5**: Calculate Euclidean Distance for the next statement, allocate next observation based on minimum Euclidean distance and update the cluster centroids and continue the steps until all observations are
MACD (Moving Average Convergence Divergence):
MACD is the moving averages convergence divergence method which is used to track the ups and downs in the sale of the product. The dairy sales will be tracked on daily basis. Soon we will find a steady point in our sales of the dairy product. Based on that point we have to find which markets are really helpful in boosting sales and which markets are underperforming.

The following are the steps for the MACD algorithm:
Step 1: Start by tracking the sales for a 12 days period of exponential moving average.
Step 2: Again track exponential moving average of 26 days period.
Step 3: Subtract both the averages.
Step 4: Now calculate the single line and decision line calculate 9 days exponential moving average.
Step 5: Finally subtract the signal line from the MACD line.

Fig. 2
The Figure 2 shown above IBM with its twelve-day EMA (green), twenty six-day EMA (red) and the 12,26,9 MACD in the indicator window. There were eight signal line limits in six months: four up and four down. The yellow area shows a period when the MACD Line flowed above 2 to reach a positive extreme. Even though upward momentum slowed after the flow, upward thrust was still resilient than downside thrust in April-May.

RSI (Relative Strength Index) algorithm:
RSI is a popular drive indicator which determines whether the product is overbought or oversold. A product is said to be overbought when the demand unjustifiably pushes the price upwards. This condition is generally interpreted as a sign that the product is overvalued and the price is likely to go down. A product is said to be oversold when the price goes depressed sharply to a level below its true value. This is a result caused due to panic selling. RSI collections from 0 to 100 and generally when RSI is beyond 70, it may show that the product is overbought and when RSI is below 30, it may indicate the stock is oversold.

The following are the steps for the RSI algorithm:
Step 1: Note down the sales period in intervals of seven days.
Step 2: To calculate the RSI use the following formula,
\[
\text{RSI} = 100 - \frac{100}{(1 + \text{RS})}
\]
Where, \( \text{RS} \) = Relative strength in a given period. (In the initial period or the first relative strength period, we compute the average trades or sales.)

First \( \text{RS} \) = \( \frac{\text{Average gain}}{\text{Average loss}} \)

Step 3: As an error correction we calculate a smooth function for the relative strength.

\[
\text{Smooth RS} = \frac{\{(\text{previous average gain}) \times 13 + \text{current gain}\}/n}{\{(\text{previous average loss}) \times 13 + \text{current loss}\}/n}
\]

Step 4: Generates indices for multiple product lines with similar stake or markets.

Step 5: Compare all the indices and list out weaker indices as a mark of improve.

The DELL example shown above in Figure 3 demonstrates a number of extreme readings as well as a negative divergence. In Oct-99, RSI reached oversold for a short-term period to mark the low around 38. The next extreme reading, in overbought, occurred after a large improvement that peaked in Dec-99. RSI got hold of overbought levels in late Dec-99 and progressed below 50 by the second week of Jan-00. The next oversold reading occurred in Feb. for another short-term period and marked the low around 35.

**Fuzzy clustering**

In the proposed system, the data obtained from the analysis process is highly divergent. The fuzzy K-Mean algorithm is applied on this data to obtain the prediction decision. The steps are as follows:

**Algorithm**

The FCM algorithm shots to partition a finite collection of \( n \) elements \( X = \{x_1, \ldots , x_n\} \) into a collection of \( c \) fuzzy clusters with respect to some given condition.

Given a finite set of data, the algorithm returns a list of \( c \) cluster centers \( C=\{c_1, \ldots , c_2\} \) and a partition matrix.
\[ W = w_{i,j} \in [0,1], \quad i = 1, \ldots, n, \quad j = 1, \ldots, c, \]
where each element, \( W_{ij} \), tells the degree to which element, \( X_i \), belongs to cluster \( C_j \).

The FCM aims to minimize an objective function:
\[
\arg \min_C \sum_{i=1}^{n} \sum_{j=1}^{c} w_{ij}^{m} \| x_i - c_j \|^2,
\]
where:
\[
w_{ij} = \frac{1}{\left( \sum_{k=1}^{c} \left( \frac{\| x_i - c_k \|}{\| x_i - c_j \|} \right)^{2m-2} \right)^{\frac{1}{m-1}}}
\]

V. FUTURE SCOPE
The proposed system involves working only on a single node and if enhanced further than it can be implemented on multiple nodes.

Sentiment Analysis is the process of defining whether provided product facility is positive, negative or neutral. It is also known as opinion mining, deriving the opinion or attitude of a customer. A common intention for this analysis is to discover how people feel about the product and its supplementation. In future, both historical data can analyze along with sentimental analysis.

VI. CONCLUSION
The proposed system minimizes manual work of order and analysis and also helps retailer/distributor to easily predict developments and optimize price for sales and distribution. A Feasible method for managing all responsibilities inside the supply chain management of dairy production system for entering the global markets is also provided. Easy access to information and feedback for convenient communication among all individual in supply chain management, efficient combination, and management of different data that come from different sources.

REFERENCES