IMPACT: International Journal of Research in Humanities, Arts and Literature (IMPACT: IJRHAL)
ISSN (P): 2347–4564; ISSN (E): 2321–8878
Vol. 7, Issue 1, Jan 2019, 575–584

CROP YIELD PREDICTION USING DATA MINING

N. Manoharan[1] & C. Sudha[2]

[1]Head, Department of Computer Science, PA College, Chengalpattu, Tamilnadu, India

[2]Assistant Professor, Department of Computer Science, PA College, Chengalpattu, Tamilnadu, India

Received: 24 Jan 2019 **Accepted: 29 Jan 2019** **Published: 31 Jan 2019**

ABSTRACT

The crop yield prediction can be defined as a supervised problem, such that for any given pre-defined forecast horizon, the model should predict the future yielders. For any given data associated with each consumer or subscriber of the system, the input includes data on past history of crop transaction for each subscriber, together with all personal and business information that is maintained by the business firm. In addition, the system needs to incorporate the training phase, where labels are provided in the form of a list of yielders together with the corresponding yield issues. This paper outlines the crop issues and reasons for their yield with emphasis on the challenges the business firm faces over the product selling attributes. The research work focuses on need for adaptive and efficient computational learning algorithms which can work on large databases to achieve reasonably good prediction performance through cross-validation for the entire crop knowledge base[1]. Based on detailed work and analysis of existing research approaches it can be understood that the literature focuses primarily only on mechanisms to predict crop yield and do not support on decision making for agricultural commodity. Much research work has been invested towards approach of yield prediction and commodity selling in Agri-market. A significant advance has been made in this research work, but however few major drawbacks to be reported are:

- *Current research methods available do not provide adequate time for companies to identify and retain the predict yielders. There is a lack of time element in commodity yield prediction.*
- *Fails to acknowledge the expensive problem of misclassifying non-yielders as yielders.*
- *The complexity theory of identifying the chances of"partial yielders" from a business firm is the major issue.*

Even though crop growth is considered as a major factor being identified behind crop yield, the challenges and issues behind achieving consistent yield is debatable behind research phenomenon. Major research analysis adopts using crop demographic and crop growth data along with farm governing policies, regional issues and environmental aspects. It is proposed in this research work that the analysis adopts data gathered from crop yield supports in resolving conflicts and provides suggestions towards suggesting a suitable alternative for crop yield analysis and improving the predictive measures.

KEYWORDS: *Crop Yield, Data Mining*

INTRODUCTION

The primary aim of the review work is fivefold, comprising of [a] Analysis of Crop Yield and issues behind crop yield [b] Methodologies adopted to control cost of crop commodity profit, [c] Schemes and polices adopted by farming support activities [d] Mechanisms to identify the commodity price pattern [e] Schemes adopted to understand market crop demand and its outcome .The purchasing capacity of crop along with their behavioral characteristics plays increasingly vital for developing a competitive and mature business firm which can engage in selling house hold materials or articles. Need for a mechanism to propose on commodity yield rate over household materials is considered as a major research challenge since the yield rate is found to be highly volatile in this domain. The survey and analysis show that farmers can easily swap their accounts and balances from one business firm to another over a short period of time. Hence developing a mechanism as a valuable and adaptable tool to identify the chances behind yield as well the challenges of sequence of commodity yield in a large business firm such as retail stores where commodity sales is optimality is required. To reduce additional investments cost and maximize effectiveness, yield prediction has to be as accurate as possible to ensure that only the crop who are planning to switch their service providers are being targeted for retention.

COMMODITY AND CROP RELATIONSHIP INDICATORS (CRI)

Crop Yield and Analysis has been applied in consumer-based marketing field due to its major application of manipulation of crops based on product usage and manufacturing firm's interest. Research studies show that any product sold in market has dynamic association metrics with type of crop and their behavioral characteristics. Methodological aspects of data mining play major role in benefiting the understanding the crop relationship with product usage [2]. The studies reveal that data mining approaches such as classification, segmentation of crop characteristics helps to formulate efficient and effective marketing strategies on survey it could be suggested that few of crop relationship indicators belongs to organizational behavior of business firm as well product usage along with crop demographic parameters such as location of livelihood, salary earned per month, functional behavior among others.

RESEARCH METHODOLOGIES EMPLOYED

Crop commodity and yield occurs when crop switch vendors or cancel service altogether and can be categorized as: unavoidable yield, involuntary yield and voluntary yield "Unavoidable yield" is defined as a sequence of event when the provider changes their operating area or when the crop moves completely out of the provider's operating area. "Involuntary yield" can be considered as an occurrence when a user fails to pay for service and the provider terminates the service as a result. Decreasing the yield rate is advantageous and should be considered as the primary goal of a business firm, since the cost involved in retaining current crop s is much less than the cost of obtaining new one. The work suggested by argues that a 5% increase in crop yielding produces reduction of operating costs by 19%[3]. Hence, if a crop indicates a "partial yield", which indicates an intention to leave (yield), then the business firm should follow an anti-yield strategy as being suggested. It is analyzed and identified that consumer satisfaction and crop purchasing behavior has a positive relation with an individual business growth, which leads to increased profit. An effective crop yield management plays an important role in enhancing the quality of crop relationships[4. The traditional literature deals with crop satisfaction by regular questionnaire-based investigation procedure which increases the working difficulties because both indicators are qualitative characteristics, using computational procedures especially adopting fitting measures.

CROP YIELD ANALYSIS AND DISCUSSION

Crop yield analysis benefits any business firm managers in identifying right decision making over crops using right marketing strategy to retain their crop[5]. It had been predicted by business analysts that the next turn of manufacturing industry needs to focus on developing malls and crop item chain industry which focuses on selling regular house hold items and goods.

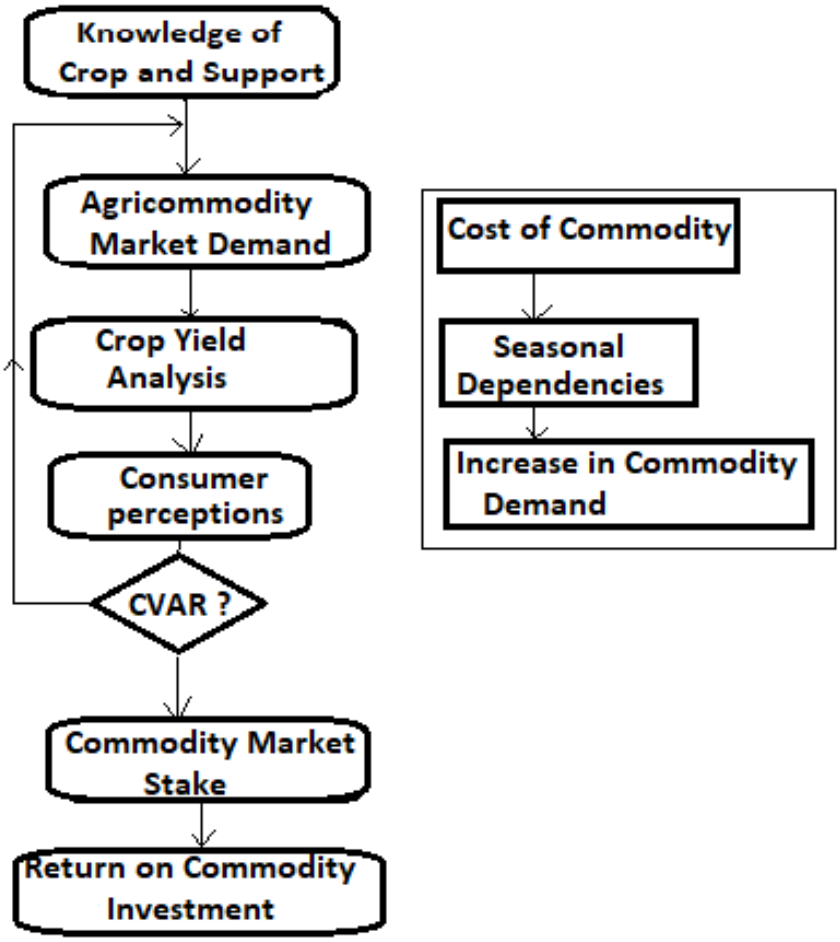

Figure 1: Crop Yield and Marketing Process.

Hence, the major objective of this work is to develop and adopt crop yield prediction model to identify consistent crop yields over period[6]. The model Fig aims to predict the characteristics of consumers so decision-makers can understand the rules of crop yield and use multiple indicators to formulate efficient and effective strategies for marketing methods to reducing yield.With associative rule approach being applied on new data sets for crop yield prediction the need for computational algorithms is required which needs to be heuristic and adaptive for dynamic and variable data set. Hence the data set gets updated with frequent crop value and notions of their pattern regarding product and its metric values. To analyze the feasibility of adaptation of decision rules in this work, various computationally functional methods need to be executed on test-run[7]. The results suggest that the crop hit rates are 95.2% for voluntary yield class 0, 88.7% for the involuntary yield class 1 and 92.3% for partial yield class 2, which invokes major outcome to prove that the new objects should also fit into the existing decision classes.Several interesting patterns are found from mined data: (a) 64% of males crop usually fall into the involuntary class 1, while 43% of married crop s usually fall into the involuntary class 1, and

23% of single crop susually fall into the voluntary yield class 0 (b) 34.92% of crops aged between 30–39 are more likely to be in the voluntary yield class[8] (c) 44.82% of crops do accept that the annual returns or benefit is higher for long term crop s than for yield crops (d) the greater the average purchase amount, the greater the chance that the crop will get higher benefit[9] (e) involuntary yield crop s usually do not have consumption up time during the last 6 months and (f) partial yield crop s and voluntary yield crop s normally do not attempt for automatic debit transfers.

COMMODITY – CROP VALUE

Crop value analysis is highly critical for any good business firm or retail market involved with marketing and maintaining a crop relationship management strategyrepresented in Figure Crop retention[10] rate is considered as a major strategy which has a major impact on crop lifetimevalue and understanding the true value of a possible crop yield will help the company in its crop relationship management.

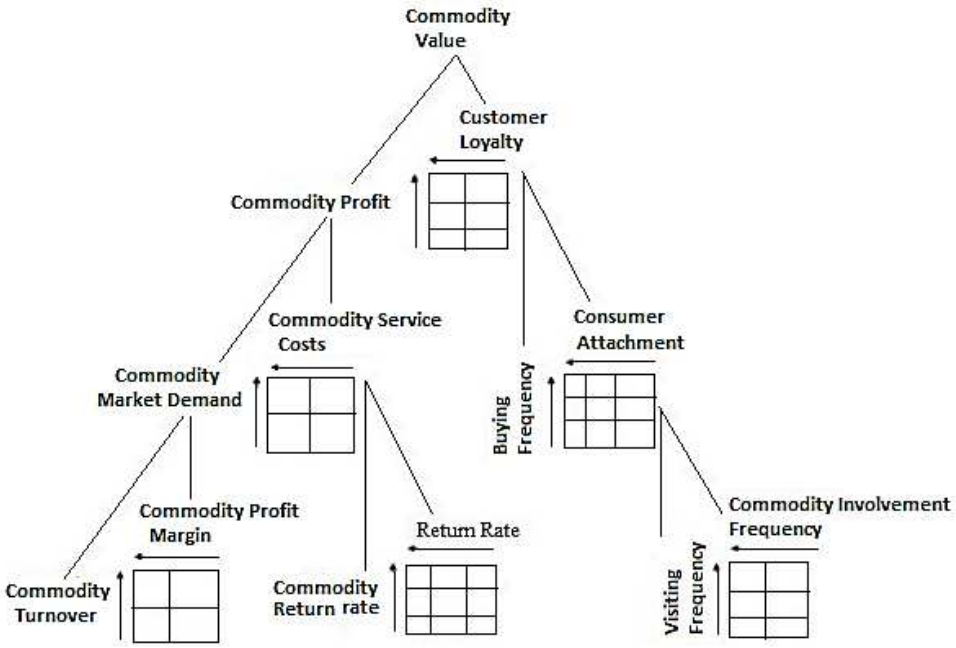

Figure 2: Commodity - Crop Value.

Several previous works have researched on the prediction of yield through using machine learning and analytical tools. A recent work adopts a decision tree classifier when applied to the geographical region of the crop which also plays an important role is the yield prediction. Proposes a dual-step model building approach, where the clustering approach is adopted to differentiate the types of crop s and then learns the classification models for each cluster separately. This approach also evaluated cost-sensitive learning technology for yield prediction[11]. The proposed yield prediction research issue is highly challengeable, since the method predicts on need to identify the “likely yielders” with good accuracy rate and within “tight time frame” and may not create a loss for the business firm. Based on this requirement and efficiency there could likely be some fixed patterns which signals that a crop would yield soon in their purchasing efforts. These challenges invoke mechanisms to develop suitable feature computation strategy and to extract the relevant features from crop data for predicting the near-future yielders and possible yield backs[12]. In predictive modeling approaches imbalance correction was performed through sampling techniques with an aim to improve the precision of prediction for the various class of crops.

[b] Research on Consumer Loyalty. As a mechanism to measure the consumer loyalty, synthesized consumer loyalty and criticized the traditional behavioral measures of crop yield which focuses on propensity for repeated purchase with additional attention to attitudinal factors behind the purchase decisions, which is labeled as "spurious loyalty"[13]. This approach indicatespurchase which is not being guided by concomitant or strong attitude but merely by situational exigencies.[c] Crop buying pattern analysis

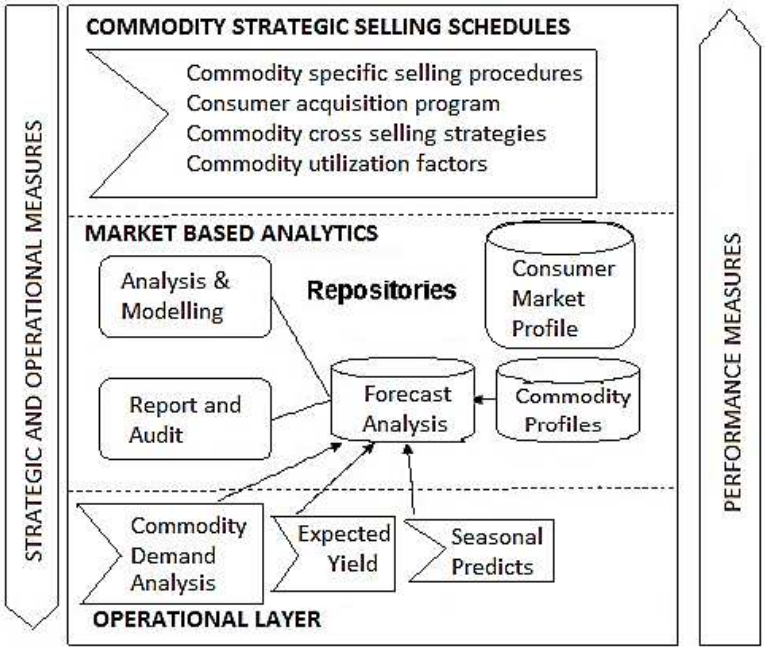

Figure 3: Architecture of Commodity Buying Pattern Analysis.

This figure shows the strategic architecture of buying pattern analysis of crop s in a business firm. The architecture suggests on three layers such as operational layer, which focuses on the business firms marketing process, product buying patterns and selling strategies[14]. The second analytical layer maintains the crop profiles, their market consistent demand issues, and commodity as product buying patterns along with analysis of crop yield over product buying approaches. The third layer called as strategic planning layer focuses on approaches to be scheduled over selling / buying of commodities, creation of marketing strategies and yield issues as well approaches to support farming community. These commodities selling profile and product buying profiles maintained as repositories help in improving the performance of business firm based on its predefined strategic and operational objectives which may be modified as per the new strategies to be evolved.

SOFT COMPUTING METHODOLOGIES EMPLOYED

Crop yield deals primarily with handling large number of data gathered among various sources of information and types of commodities to be sold. Computational approaches play major role in predicting crop yield along with data mining approaches to identify and predict crop yield effectively within minimal time period.Fuzzy classification approaches have been proposed in the marketing literature which occupies higher role among other computational approaches. Proposed a segmentation of crop using fuzzy clustering methods. A cluster wise regression model for simultaneous fuzzy market structuring was proposed by Hsu's Fuzzy Group Positioning Model allows an understanding of the relationship between consumer consumption patterns, and the business firm's competitive situation with strategic positioning. The modeling of

fuzzy data in qualitative marketing research was suggested by suggested a fuzzy Classification Query Language for crop relationship management. Fuzzy control is being applied to classical marketing issues in major cited research literature. However, fuzziness has not yet been adapted for e-business, e-commerce mechanisms or e-government. Electronic shop is a specific model where the power of fuzzy classification model is adopted. In this approach online crop s are not assigned to classical crop segments but bended to fuzzy classes. This approach leads to differentiated online business concepts as well helps to improve the crop equity of shoppers on web as well minimizes the shopping challenges among users or buyers. An important difference between a fuzzy classification and normal data mining approach is the fact that a crop can be clustered based on fuzzier defined mapping. In conventional computational methods the groups or segments of crop s are typically constituted by a small number of qualifying attributes. If the transaction data value corresponds to two or more crop s in similarity, then the same membership functions can be applied. The data gathered and analysis of online crop s compared to traditional approach has the advantage that more information related to the crop s' behavior is automatically logged in the system.Some factors which establish service quality as one of the major factors with a significant impact on crop satisfaction appeared to be product utilization quality, value-added services, and product support[15. This suggests that, while product service providers (PSP) have an improved product quality over historical past scenarios through massive equipment investments or the scalability of product usage, clarity and coverage, according to crop s' perceptions, still retain their importance. In addition, the business firms or carriers should also concentrate their efforts on developing value-added services to increase product utilization and extend their convenience. In the area of crop support, carriers should consider much effort as strive to minimize crop s' inconvenience by speedily processing crop s' complaints through a variety of systems and channels.

DATA SET

Various data mining techniques can be applied to large data set to analyze crop properties attributes. The data set has been assembled from various business firm surveys conducted by Net Research Labs at household item selling marketing firms in and around Bengaluru city. The research has also utilized an existing data collected from literature survey analysis to identify the common set of patterns and correlations between crop s and marketing firm properties.

The sample data set is selected from differentareas at Bengaluru with 1,207,000 consumers or crops. The data set consists of primarily the crop 's product transaction data, product return data, time taken for servicing, crop 's personal information between three different time periods as first quarterly of 01.01.2011 and 31.04.2011, second quarterly of 01.06.2011 and 31.12.2011, third quarterly of 01.01.2012 and 31.04.2012 and finally between 01.01.2010 and 31.04.2011 as four varying time intervals and 10 various data sets collected from various farmers in TamilNadu, Andhra Pradesh and Maharastra. India.The data set was collected from 2500out of which 1532 members volunteered to participate all the four surveys at various period.

- The proposed crop yield model being surveyed links the probability of yielding to the level of crop yield satisfaction with crop growth and not to farmers subsidies or benefits attached to the growth of crop.
- The crops yield permitted to change the crop growth or another alternate crop growth or similar crop yield analysis as switching from one crop to another without termination or yield with the existing crop yield.
- The latter point highlights the profusion of product usage and subsidies before June 2000, which implies that the consumer dissatisfaction with inadequacy of subsidies would have been rare. The data are collected from 260

farmers. From this method by applying ANN the total energy produced between the input and output of potato production was 83,723 and 83,059 MJ. From the analysis, it's found that electricity, seed and chemical fertilizers are the most influential factors for energy consumption for potato production. In this work an expert system is designed for tomato crop analysis aimed for efficient plantation of tomato. This system integrates eminent experts, agriculture scientists, programmers and designers to develop the tomato analysis expert system. The proposed system is divided into two subdivisions namely is tomato crop expert system and a tomato analysis system. The proposed system is developed by ID3 algorithm and website is developed by Java programming as a front end and SQL for storing data. The crop expert system shows the varieties of different tomato, with support for tomato yield analysis and their selling aspects based on change in market demand is studied.This work reviews many strategic applications of Machine Learning in maize breeding[16]. Quantitative attribute loci mapping, heterotic cluster assignment and thewell-liked genome-wide choicearea number of the key areas presentlyself-addressed by the literature. Results are encouraging and the proposed Machine Learning algorithms are valuable alternative to ancientstatistical techniques applied in maize, even a lot of recently introduced linear mixed models.This work analysis the totally different applications of a simulation tool for winter wheat crop management previously developed has made public the difficulties encountered by crop production plan engineers in planning ways that diverge from those presently used. The author presenteda groundwork Support System that tend to present the approaches primarily based on reinforcement learning and genetic rule technics. It defines attention-grabbing ways in a progressive manner and seems to be a promising approach to help users to explore a wider variety of solutions.This work proposes the rice crop observation system developed with the variations as applied to a neural network classification. The rice production areas are systematically illustrated for wet and dry season andwas able to extract data on rice cultivation as a perform of various planting dates. A minimum mapping accuracy of 96 % was achieved for each season. This data was then utilized in a neural network-based yield model to predict rice yield on a regional basis for the wet season. Once the yields expected by the neural network were compared with government statistics and achieved 94% than others[17]. This work represents the forecastingbehaviourof the international business models of rice based on supply and demand with unpredictable factors like environmental, biological, metrological factors and subsidiaries etc. The artificial neural networks are applied to evaluate the forecasting behaviour for Thailand rice exporters and the performance measure is compared with ARIMA and exponential smoothing model. The evaluation measures considered in this work are MAPE,MAE, RMSE and MSE. From this work ANN performed relatively higher compared to the other existing models.Data mining applications in horticulture is a generally new approach for anticipating/foreseeing of farming crop/animal administration. In the present investigation an endeavour has been made to think about the impact of climatic parameters on soybean efficiency utilizing choice tree acceptance strategy. The discoveries of Decision tree were confined into various rules for better comprehension by the end clients. The investigation discoveries will help the specialists, approach creators and ranchers in anticipating/determining the harvest yield ahead of time for advertise progression.In agriculture area where agriculturists and agribusinesses need to settle on incalculable choices consistently and complicated complexities includes the different variables affecting them. A fundamental issue for farming goal is the exact yield estimation for the various harvests associated with the planning. Data mining methods are vital approach for achieving functional and successful answers for this issue. Agribusiness has been a conspicuous focus for huge information about natural conditions, changeability in soil, input levels,

blends and product costs have made everything the more important for agriculturists to utilize data and inspire help to settle on basic cultivating choices. This work centers around the examination of the agriculture information and finding ideal parameters to boost the yield creation utilizing data mining strategies like PAM, CLARA, DBSCAN and Multiple Linear Regression. Mining the expansive measure of existing yield, soil and climatic information, and breaking down new, non-test information streamlines the creation and makes agribusiness stronger to climatic change.

With the effect of environmental change in India, the share of the agricultural products are as a rule seriously influenced in-terms of their execution over a time of most recent two decades. Foreseeing the product yield well in front of it reap would help the strategy creators and ranchers for taking fitting measures for promoting and capacity. Such forecasts will likewise help the related enterprises for arranging the coordination of their business. A few strategies for anticipating and demonstrating crop yields have been created in the past with shifting rateof achievement, as these don't consider characteristics of the weather, and are mostly observational. In the present investigation a product apparatus named `Crop Advisor' has been created as an easy-to-use website page for foreseeing the impact of climatic parameters on the harvest yields. C4.5 calculation is utilized to discover the most affecting climatic parameter on the product yields of chosen trims in chosen regions of Madhya Pradesh. This product gives a sign of relative impact of various climatic parameters on the harvest yield, other agro-input parameters in charge of harvest yield are not considered in this instrument, since utilization of these information parameters fluctuates with singular fields in space and time.In the work light of late advances in spectral imaging innovation, exceedingly adaptable demonstrating strategies must be created to appraise different soil and yield parameters for exactness cultivating from airborne hyperspectral symbolism. The capability of Artificial neural networks (ANNs) for the advancement of in-season yield mapping and determining frameworks was analyzed. Hyperspectral pictures of corn plots in eastern Canada, subjected to various preparation rates and different weed administration conventions, were obtained by a minimal airborne ghastly imager. Factual and ANN approaches alongside different vegetation files were utilized to create yield forecast models. Primary segment was utilized to lessen the quantity of information factors and more prominent expectation precision was gotten with an ANN show than with both three traditional observational models in view of standardized contrast vegetation file, basic proportion, or photochemical reflectance record. No unmistakable distinction was seen amongst ANNs and stepwise numerous straight relapse models. Although the high potential helpfulness of ANNs was affirmed, especially in the formation of yield maps, facilitate examinations are required before their application at the field scale.

RESULT

Status of commodity and yield is being analysed using SPSS in comparison with past two-year historical dataset. The yield metrics are discussed in Section 3.0 which is primarily considered as major parameters for analysis. Statistical approaches (Chi-square, T-test) are applied on crop data and their monthly transactions data, it was identified that for a crop whose Monthly Salary mean lies between 200 and 700 per person then their monthly transactions of purchase were more than 20% - 25% of their salary. The list of proxy indicators used in business firms which have the need to implement crop yield analysis. The following indicators propose methodologies adopted by business firm experts to analyse the phenomenon of crop yield implemented in a region. On detailed survey and analysis of research work, the proposed work aims to develop a crop profiling approach to effectively predict and control crop yield in advance, while keeping the misclassification of yield levels to a minimum. This methodology also incorporates time element in the prediction of crop yield for maximizing

future yield capture by identifying a potential loss of crop at an earlier business rate. Two case studies are identified and carried out for validating the proposed methodology of Bee-Hive approach, Fuzzy Multi-criteria decision approach and Taguchi's optimization approach using repairs and complaints data. Finally, the results from each proposed methods are analyzed and compared against existing popular yield prediction techniques reported in literature. The proposed research demonstrates that crops can be placed into one of several profiles' clusters according to their interactions with the service provider.

REFERENCES

1. *A.M. Khorasani, M. R. S. Yazdi, and M. S. Safiza-deh, "Tool Life Prediction in Face Milling Machining of 7075 Al by Using Artificial Neural Networks (ANN) and Taguchi Design of Experiment (DOE)", International Journal of Engineering and Technology 3, no. 1, 2011*
2. *Aída Jiménez, Fernando Berzal ,Juan-Carlos Cubero "Using trees to mine multirelational databases" Data Mining and Knowledge Discovery Issn: 1384-5810 pp-1-39*
3. *Alder JR and Hostetler SW (2015) Web based visualization of large climate data sets, Environmental Modelling &Software 68, 175-180*
4. *Amol D. Vibhute, K. V. Kale, Rajesh K. Dhumal and S. C. Mehrotra "Soil Type Classification and Mapping using Hyperspectral Remote Sensing Data" 2015 International Conference on Man and Machine Interfacing (MAMI).*
5. *Ankalaki, S., Chandra, N., Majumdar, J., "Applying Data Mining Approach and Regression Model to Forecast Annual Yield of Major Crops in Different District of Karnataka", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 5, Special Issue 2, pp.25-29, 2016.*
6. *Anup K. Prasad, Lim Chai, Ramesh P. Singh, MenasKafato, Crop yield estimation model for Iowa using remote sensing and surface parameters, International Journal of Applied Earth Observation and Geoinformation Vol 8 (2006) 26–33*
7. *Arun Kumar Sangaiah and Arun Kumar Thangavelu: Applications Of Soft Computional Approaches In Bioinformatics. IIOABJ, 2015; Vol. 6 (4): 01-02*
8. *AsefaTaa, Douglas Tanner, Alan T.P. Bennie, Effects of stubble management, tillage and cropping sequence on wheat production in the south-eastern highlands of Ethiopia Soil & Tillage Research 76 (2004) 69–82*
9. *B. Taskar, E. Segal, and D. Koller. "Probabilistic clustering in relational data" In IJCAI-01 pages 870--876, 2001.*
10. *Bart Kosko, Fuzzy Cognitive Maps, International Journal of Man Machine studies, 24, 65-75, 1986*
11. *Box G, Jenkins G. Time series analysis: forecasting and control. 5th ed. Hoboken Wiley 2015.*
12. *C.J.C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," Data Mining and Knowledge Discovery, vol. 2, pp. 121-168, 1998.*

13. *Ching-JuiKeng, Tseng-Lung Huang, Li-Jie Zheng, Maxwell K. Hsu, (2007) "Modeling service encounters and customer experiential value in retailing: An empirical investigation of shopping mall customers in Taiwan", International Journal of Service Industry Management, Vol. 18 Issue: 4, pp.349-367*

14. *Chipanshi, Y. Zhang, L. Kouadio, N. Newlands, A. Davidson, H. Hill et al. "Evaluation of the Integrated Canadian Crop Yield Forecaster (ICCYF) model for in-season prediction of crop yield across the Canadian agri-cultural landscape." Agricultural and Forest Meteorology, pp: 137-150, 2015.*

15. *Christine Preisach, Lars Schmidt-Thieme "Ensembles of Relational Classifiers" Springer Knowldge Information System vol.14, Issue-3, pg.no-249–272.*

16. *Colaco,M.Castro Rego.F.,Meirs.P& Santos(2005 May). What are the opinions of foresters in portugal regarding fire?.In proceedings of the 2ndInternationsl conference on prevention strategies of Fire in southern Europe ", Barcelona (Spain) (PP 9-11).*

17. *Cortez, P.,&Morais, .AD.J.R.(2007). A data mining approach to predict cotton yields using meteorological data.*