Machine learning applications for sustainable agriculture supply chain performance

Palak Mittal(IIT2018117), Nehal Singh(IIT2018119), Chinmay Shravanbal Tayade (IIT2018138)

VI Semester, B.Tech, Information Technology

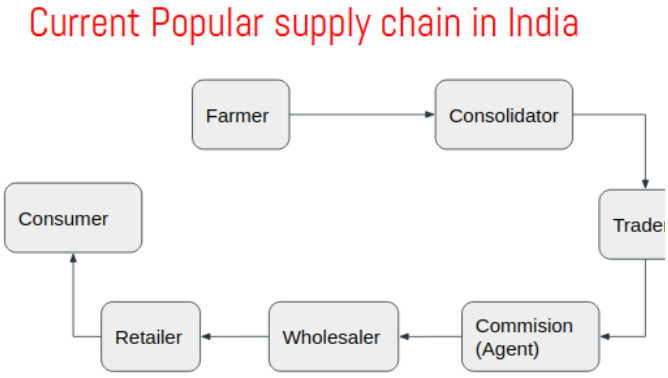
INDIAN INSTITUTE OF INFORMATION TECHNOLOGY, ALLAHABAD, INDIA

Abstract: This paper concerns the improvement in the supply chain and management of supply chain using the machine learning models it focuses on predicting the defect percent in supply chain and recognizing the factors that often show whenever a defect occurs. This will help improve the defect as suppliers can be more focused on the factors thereby leading to a decreased defect percent and better supply chain management.

I. INTRODUCTION

Machine learning , Artificial intelligence and deep learning are evolving concepts in today's era and used in almost every field . Agriculture plays an important role in sustaining human life and activities. To ensure food security for this ever-growing population , we need to sustain agriculture.

Now these activities should not only focus on productivity but also limiting harmful impacts on the environment. Agriculture Supply chain consists of organizations that are responsible for production and distribution of fruits , cereals,vegetables etc. ASCs are facing tremendous pressure nowadays for delivering the right quantities. We would try to develop decision support systems to predict product quality , semi-arid regions , irrigation requirements and improving farming practices.



II. MOTIVATION

In the world of logistics, supply chain management is meant to facilitate the transportation and manufacturing of the product and ease them to reach consumers, supply chain management consists of retailers, distributors , manufacturers and finally consumers. Sustainable supply chain management is outstretched beyond sustainability and going green.

Being sustainable helps to improve productivity and it eventually saves money.

III. BACKGROUND

In the past, few review studies were conducted on Artificial Intelligence and Machine Learning applications for improving the supply chain performance. In this paper we present a systematic literature review of 9 papers on Machine Learning applications in developing sustainable Agriculture Supply chain .

There are some Machine learning Terminologies and Algorithms that needed to be studied before implementing the paper are mentioned in APPENDIX I and APPENDIX II respectively.

IV. PROBLEM STATEMENT

Machine learning applications for sustainable agriculture supply chain performance .Supply chain planning and management (SCP & SCM) using machine learning

V. OBJECTIVES

1. To implement sustainable supply chain management needs to be focused on sustainable goals and targets.
2. To develop strategies leading us to higher levels of productivity and profitability.
3. There are various methods of applying machine learning algorithms to improve supply chain as mentioned in Appendix III.

VI. LITERATURE SURVEY

S.no	Paper Title	Name of Conference	Purpose	Methodology	Dataset	Results
1.	A Systematic Literature Review on Machine Learning Applications for Sustainable agriculture Chain Performance	Computers & Operations Research, Volume 119, (2020)	The objective was to study the applications of ML algorithms in the Agriculture supply chain.	Steps for methodology- Planning the review Conducting the review Descriptive statistics	The SLR was performed on 93 research articles, which were categorized using different ML algorithms across different ASC phases.	The current study is done on the basis of SLR to determine the current state of research on machine learning (ML) applications in ASC. The study finds that all three ML algorithms, that is, supervised, unsupervised, and reinforcement learning is used to develop relevant ASCs.
2.	A literature review on machine learning in supply chain management	International Conference of Logistics (HICL), Vol. 27 (2019)	In recent years,the use of practical use of logistics using machine learning has increased gradually, mostly in Supply Chain Management (SCM). By using ML methods to the SCM task model, the paper shows current applications in SCM and shows potential research gaps.	Suitable papers with applications of ML in SCM are found on the basis of literature review from 2009-2019 . The used ML methods are linked to the SCM model, forming a reciprocal mapping.	-	With regard to the standard process model, the research focus lies in the modeling phase. Further research emphasizing on and explaining the concepts used in Business, Data Understanding and Evaluation is required. The examined papers cover these phases, but hardly present useful insights and concepts for SCM. The Deployment phase is not treated at all.
3.	Analysis of agriculture data using data mining techniques: application of big data	Journal of Big Data 4, Article : 20 (2017).	The objective of proposed work is to analyse the agriculture data using data mining techniques.	Various data mining techniques are used on the input data to get the accurate performance yielding method.	Dataset in agricultural sector [https://data.gov.in/, http://raitamitra.kar.nic.in/statistics	According to the analyses of clustering quality metrics, DBSCAN gives the better clustering quality than PAM and CLARA, CLARA gives the better clustering quality than the PAM. The proposed work can also be extended to analyse the soil and other factors for the crop and to increase the crop production under the different climatic conditions.
4.	Artificial intelligence-driven innovation for enhancing supply chain resilience and performance under the effect of supply chain dynamism: an empirical investigation	Annals of Operations Research	The purpose of this study was to investigate the effect of AI-based systems in enhancing, directly and indirectly, SCP during the influence of dynamism and uncertainty.	Steps for methodology- Instrument development Sampling design and data collection Data analysis and results Measurement validation	Survey data was collected from 279 firms representing different sizes, operating in various sectors, and countries.	The results support that developing AI-based innovation is positively and significantly related to enhancing both AC and SCC.
5.	Artificial intelligence in supply chain management: theory and applications	International Journal of Logistics Research and Applications [2019]	This paper reviews the past record of success in AI applications to SCM and identifies the most fruitful areas of SCM in which to apply AI.	Classifying and studying the following fields of ML concerning the supply chain management. Artificial neural networks (ANN) and rough set theory Machine learning, expert systems, and GAs Fuzzy logic Agent-based systems	N/A	Since SCM requires the comprehension of complex, interrelated decision-making processes and the creation of intelligent knowledge bases essential for joint problem-solving, SCM has evolved into knowledge management.

6.	Agriculture Analysis Using Data Mining And Machine Learning Techniques	5th International Conference on Advanced Computing & Communication Systems (ICACCS) (2019)	Agriculture is an important application in India. The modern technologies can change the situation of farmers and decision making in agricultural field in a better way	Python is used as a front end for analysing the agricultural data set. Jupyter Notebook is the data mining tool used to predict the crop production. The data mining techniques like K-Means Clustering, KNN, SVM, and Bayesian network algorithm where high accuracy can be achieved.	The parameter includes in the dataset are precipitation, temperature, reference crop, evapotranspiration, area, production and yield for the season from January to December for the years 2000 to 2018. https://data.gov.in/	In this analysis, they used some of the common data mining techniques in the field of agriculture. Some of these techniques, such as the k-means, k-nearest neighbor, SVM, and bayesian network are discussed and an application in agriculture for each of These techniques are presented.
7.	Global Supply Chain Management: A Reinforcement Learning Approach 14 Nov 2010 P. Pontrandolfo,A. Gosavi, O. G. Okogbaa &T. K. Das	Politecnico di Bari - Dipartimento di Progettazione e Produzione Industriale Viale Japigia 182, 70126 Bari, ITALY	This paper is concerned with coordination and integration of MNC's with emphasis on logistics and management of production processes.	The application of Markov decision theory to SCM problems result in a large state-space. reinforcement learning approach to solve the semi-Markov decision problem	Survey data was collected from logistics firms	In particular, we model the GSCM problem as a semi-Markov Decision Problem (SMDP) and solve it using the Reinforcement Learning (RL) algorithm.

8.	Applications of Machine Learning Techniques in Supply Chain Optimization 28 June 2019	Reliability, Quality,Control,Safety, Maintenance and Management ICICCT 2019 – System	The research reviews the cases where Machine Learning Techniques are being used in Supply chain optimization.	Neural networks. Decision Trees & Random Forests. Support Vector Machines. Predictive Analytics for Demand Forecasting: Retail Chain Forecast.	Survey data was collected from e-business firms	This study is an attempt to seek various business applications of Machine Learning(ML) techniques in Supply Chain Management. The research reviews the cases where Machine Learning Techniques are being used in Supply chain optimization.
9.	Predictive big data analytics for supply chain demand forecasting: methods, applications, and research opportunities	Concordia Institute for Information Systems Engineering (CIISE), Concordia University, Montreal H3G 1M8, Canada	Concordia Institute for Information Systems Engineering (CIISE), Concordia University, Montreal H3G 1M8, Canada This survey points to the fact that the literature is particularly lacking on the applications of BDA for demand forecasting in the case of closed-loop supply chains (CLSCs) and accordingly highlights avenues for future research.	The application of BDA and machine learning using Neural networks. Decision Trees & Random Forests. Support Vector Machines. Predictive Analytics for Demand Forecasting: Retail Chain Forecast	Survey data was collected from e-logistics firm	In the survey predictive BDA applications in supply chain demand forecasting propose a classification of these applications, identify the gaps, and provide insights for future research. We classify these algorithms and their applications in sup-ply chain management into time-series forecasting, clustering, K-nearest-neighbors, neural networks, regression analysis, support vector machines, and support vector regression.

VII. APPLICATIONS

The applications of sustainable supply chain management will enable us to have more access to our business and keep the firm ahead of competition, will help to operate the business efficiently and eventually it will enhance customer satisfaction. Through the use of Computerised shipping and tracking systems as well as using RFID for checkout of products will ease the visibility and operations of products.

By using Social media to streamline the supply Chain and finally the Big data

Plays a major role as a catalyst for the Supply chain management systems, helps supply chain systems to grow and proliferate into new markets.

Demand Prediction:

The demand prediction plays a major role in supply chain management. Therefore, it is very necessary to figure out the demand of products accurately in order to place the exact amount that is needed minimizing the wastage and lack of supply in the area. There are many algorithms in machine learning to predict the demand.

VIII. DATASET DESCRIPTION

LINK : DataCo SMART SUPPLY CHAIN FOR BIG DATA ANALYSIS

A DataSet of Supply Chains used by the company DataCo Global was used for the analysis.

We would be using these parameters to predict defect percent in upcoming orders . We will divide the dataset in a 70:30 ratio of training and testing dataset. Out of all the parameters we considered the parameters given below:

X_TRAIN :

Purchase Order Amount	Purchase Order Quant	PO Sent in Advance of Delivery
43103	365	
22884	898	
18671	157	
21403	157	
22479	858	
30451	731	
9452	682	

Y_TRAIN

D
Defect Percent
4.9
8.4
1.2
0.4
0.1
0.5

IX. LANGUAGE AND TOOLS

We will be majorly using Python 3 as the project programming language.

Tools :

A. Pandas → A software library used for data manipulation and analysis.
B. Scikit-learn →A Machine learning library for python.

Algorithms :

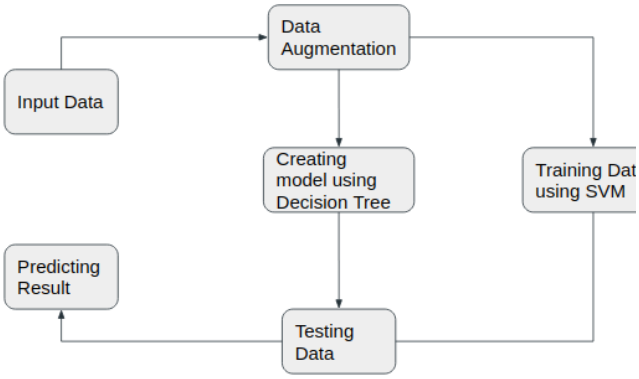
A. Decision Tree
B. Support Vector Machine

X. PROPOSED METHODOLOGY

1) First we will load the necessary set of tools and libraries that includes pandas and scikit.
   a) Pandas is a software library in python which facilities easy reading and writing of CSV files and using its data.
   b) Scikit-learn is a learning library in python which helps in using the ready models of Decision Tree and various other algorithms and helps to predict results out of these models from training and testing dataset.
2) We will load and read the data, both training and testing dataset by reading the excel files of training and testing dataset
3) We will form different lists of the dependent and independent variables taken from the dataset. The following 4 parameters are read from csv and put in Pandas dataframe for further use.
   a) Purchase Order Amount
   b) Purchase Order Quantity
   c) PO Sent in Advance of Delivery
   d) Defect Percent
4) "PO Category" is introduced as a new category . This is done to categorise "PO Amount" value from 0 to 30,000 GBP is classified as PO Category 1, from 30,000 to 60, 000 GBP as PO Category 2 and henceforth.
5) Scale the independent variables of testing and training dataset.
6) Save the scale as "Scale.sav" using pickle.dump().
7) We will train it using the models from the Support Vector Machine and Decision tree as "SVR_TrainedModel.sav" using pickle.
8) Predict the dependent variable i.e. defect per cent values from the trained independent dataset and measure the error/accuracy of the model.

9) From the independent train variable dataset, the prediction is done which is then compared with actual values, and R^2 score for regression estimators is returned by the scoring method.

10) Once predictions are done under error limits, we can go with a testing dataset.



XI. IMPLEMENTATION

1. Import the required modules
2. Read the data source

"Supplier Past Performance.xlsx" → Training Dataset

	Purchase Order Amount	...	Defect Percent
0	33386	...	4.90
1	48477	...	8.40
2	31688	...	1.20
3	27115	...	0.40
4	9028	...	0.10
..	...	...	...
694	44313	...	2.76
695	38823	...	9.87
696	15804	...	5.14
697	47081	...	8.39
698	25606	...	5.41

[699 rows x 4 columns]

"Defect Predict.xlsx" → Testing Dataset

	Purchase Order Amount	...	PO Sent in Advance of Delivery
0	42130	...	1009
1	39952	...	1009
2	18537	...	2494
3	48981	...	570
4	10271	...	2150
..	...	...	...
295	23086	...	2571
296	40294	...	72
297	38860	...	69
298	7457	...	339
299	47283	...	2906

[300 rows x 3 columns]

3. Declaring the independent and dependent train data from the sample.
4. Scaling the independent test and train data
   A. We are using pickle.dump() to save scale as "Scaler.sav"
5. Fit the test data in machine learning model - Support Vector Regressor
   A. The trained model as "SVR_TrainedModel.sav" using pickle is saved.

```
The R-SQUARED VALUE of the Support  Vector model is -0.02
Mean Square Error  of  Vector  model is  8.45
Root Mean Square Error of  Support  Vector  Learning model is  2.91
```

6. Fit the test data in machine learning model - Decision Tree Model

```
The R-SQUARED VALUE of model Decision Tree model is  -0.83
Mean Square Error of Decision Tree model is  15.13
Root Mean Square Error of Decision Tree model is  2.91
```

XII. RESULTS

a. SVM

```
The prediction by SVM model is  [5.05348482 4.95700187 5.77890377 4.
 5.62233194 5.35991581 5.47839017 5.23974272 5.38218093 5.32273432
 5.42106986 5.22165591 4.91371383 5.42289676 5.4459813  4.96670967
 5.69325422 5.05656883 5.69261619 5.53886166 5.30478037 5.87252239
 5.17523037 5.03411014 5.68035191 5.35023246 5.21156652 4.73058861
 5.26771173 4.9386791  5.32191455 5.76322284 5.78500915 5.20384568
 5.46861008 5.31131451 5.28630659 5.13705173 5.50184101 5.22266727
 5.72168516 5.17153693 5.142541   5.51562398 5.43979658 4.93727165
 5.03369931 5.02829304 5.51372195 5.58982092 5.43665735 5.73757409
 5.24904742 5.14955689 4.8320443  5.14766602 5.12690701 5.29750048
 5.27393569 5.08171016 5.79919509 5.20749262 5.07568145 5.34282941
 5.35100328 5.18465257 5.54293917 5.44040517 5.628021   5.64030268
 5.51872643 5.7724452  5.27335267 5.44358469 5.07429904 5.81619352
 5.43801216 5.51475284 5.62363185 5.10166039 5.45536964 5.85537976
 5.01603037 5.01978641 5.43420741 5.29930581 5.52648989 5.42065469
 5.01982647 5.68630594 4.98754378 5.52154526 5.02566379 5.51356026
 5.54822562 4.94752567 5.05197362 5.17403924 5.61343082 5.84867
 5.13667534 5.05589668 4.88660847 5.43383636 5.32690111 4.98260819
 5.11147499 5.03381402 5.57278184 5.01690434 5.66336779 5.19448775
```

b. Decision Tree model

```
The prediction by Decision Treemodel is  [1.76 6.81 6.62 0.37 6.27 4.41
 8.07 8.43 2.12 0.87 6.27 1.76 6.62 5.24 0.87 6.36 8.5  7.67 0.94 4.23
 4.9  2.14 2.49 1.67 4.23 9.39 9.39 0.4  3.69 2.04 5.55 4.33 6.62 0.4
 5.55 0.4  3.31 5.55 2.33 1.76 0.87 9.31 6.47 1.22 9.39 7.72 7.89 8.88
 6.81 4.41 7.67 0.37 0.4  4.41 3.57 0.4  5.14 1.76 5.24 1.22 4.44 9.39
 8.43 7.36 1.22 6.62 7.34 1.76 0.4  6.62 8.61 9.2  2.12 7.67 6.62 5.93
 3.31 0.12 8.43 8.5  6.62 3.69 1.67 3.65 7.67 8.43 8.61 0.4  5.03 4.9
 4.9  1.22 5.66 9.39 0.87 8.66 4.9  9.76 3.69 1.76 0.4  7.76 6.93 4.9
 5.66 0.4  6.62 9.31 7.76 6.27 2.12 7.72 4.23 4.9  7.72 2.12 2.49 5.03
 3.69 8.43 5.93 6.81 5.14 4.44 0.4  5.14 2.27 5.24 5.24 3.69 2.49 7.72
 6.62 0.94 1.22 1.76 2.76 2.33 5.24 6.36 0.4  1.22 2.12 6.62 6.62 0.87
 5.03 8.43 1.76 3.33 2.33 3.65 8.97 4.23 2.27 0.4  1.76 0.87 8.97 6.81
 1.22 1.22 1.22 7.67 5.24 3.65 7.34 7.34 6.41 2.36 4.23 6.27 7.72 6.36
 5.24 0.4  7.67 2.12 1.22 6.81 1.16 6.54 8.4  3.69 6.93 8.97 7.34 7.67
 0.12 0.4  0.87 8.43 9.76 7.76 3.57 2.33 8.97 9.83 5.24 9.76 6.36 6.62
 9.39 6.36 1.76 0.4  0.87 2.67 5.66 4.44 1.22 8.5  8.43 2.36 3.69 8.61
 4.23 4.23 7.43 0.84 0.87 3.65 2.49 9.39 8.43 7.67 8.43 6.36 2.76 4.23
 0.4  2.36 8.61 0.87 7.72 6.62 1.76 7.67 7.34 5.14 8.61 1.92 5.41 7.34
 7.89 6.62 5.46 7.72 2.12 9.76 1.22 3.65 4.41 0.93 0.93 0.1  0.93 8.97
 2.33 2.49 4.9  1.76 7.72 7.67 2.04 4.9  7.72 2.61 7.36 8.5  7.72 5.24
 0.67 2.49 4.41 5.03 4.41 2.36 1.76 8.97 8.97 7.72 2.36 9.39 1.22 3.65
 2.49 4.23 9.88 5.77 7.67 5.24]
```

Using the above two classifiers we are now predicting the defect percent of orders in the testing dataset.

XIII. ACTIVITY SCHEDULE

Steps	Time Required	Predicted Date & Time
Requirement Verify	Done	5th February 2021
Project Planning	Done	15th February 2021
System & Detail Design	Done	25th February 2021
Coding	Done	20th March 2021
Debugging and coding	Done	5th April 2021
Testing	Done	10th April 2021
Documentation and Final	Done	15th April 2021

XIV. CONCLUSION

Here we proposed the methodology to predict the defect percent in the supply chain and recognizing the variables that lead to the increase in defect percent by using the two algorithms SVM and Decision Tree.

Predicting the results will lead to a significant decrease in the defect percent as executioners in the supply chain management will be fully aware of the factors causing the defect in the supply chain.

As the population increases and the world deflects more towards finding solutions using ML, AI.

Further, we can develop strategies to avoid environmental risks, and increase productivity considering health and pure environment by reducing hazardous by-products like greenhouse gases ,emission of chemicals and by-products directly into water streams, solid waste etc.

REFERENCES

[1]. Rohit Sharma, Sachin S. Kamble, Angappa Gunasekaran, Vikas Kumar, Anil Kumar, A systematic literature review on machine learning applications for sustainable agriculture supply chain performance, Computers & Operations Research, Volume 119, 2020, 104926,ISSN 0305-0548,
DOI: https://doi.org/10.1016/j.cor.2020.104926.

[2]. Wenzel, Hannah; Smit, Daniel; Sardesai, Saskia (2019) : A literature review on machine learning in supply chain management, In: Kersten, Wolfgang Blecker, Thorsten Ringle, Christian M. (Ed.): Artificial Intelligence and Digital Transformation in Supply Chain Management: Innovative Approaches for Supply Chains. Proceedings of the Hamburg International Conference of Logistics (HICL), Vol. 27, ISBN 978-3-7502-4947-9, epubli GmbH, Berlin, pp. 413-441
DOI: http://dx.doi.org/10.15480/882.2478

[3]. Majumdar, J., Naraseeyappa, S. & Ankalagi, S. Analysis of agriculture data using data mining techniques: application of big data. J Big Data 4, 20 (2017).
DOI: https://doi.org/10.1186/s40537-017-0077-4

[4]. Belhadi, A., Mani, V., Kamble, S.S. et al. Artificial intelligence-driven innovation for enhancing supply chain resilience and performance under the effect of supply chain dynamism: an empirical investigation. Ann Oper Res (2021).
DOI: https://doi.org/10.1007/s10479-021-03956-x

[5]. Min, Hokey. (2010). Artificial intelligence in supply chain management: Theory and applications. International Journal of Logistics-research and Applications - INT J LOGIST-RES APPL. 13. 13-39.
DOI: https://doi.org/10.1080/13675560902736537

[6] C. N. Vanitha, N. Archana and R. Sowmiya, "Agriculture Analysis Using Data Mining And Machine Learning Techniques," 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS), Coimbatore, India, 2019, pp. 984-990,
DOI: 10.1109/ICACCS.2019.8728382.

[7]Global Supply Chain Management: A Reinforcement Learning Approach by Abhijit Gosavi ,O. Geoffrey Okogbaa ,Tapas K. Das.
DOI: 10.1080/00207540110118640

[8] Applications of Machine Learning Techniques in Supply Chain Optimization ,ICICCT 2019 – System Reliability, Quality Control, Safety, Maintenance and Management.
DOI - 10.1007/978-981-13-8461-5_98

[9]Predictive big data analytics for supply chain demand forecasting: methods, applications,and research opportunities, Seyedan and Mafakheri J Big Data.
DOI- 10.1186/s40537-020-00329-2.

[10]. Constante, Fabian; Silva, Fernando; Pereira, António (2019), "DataCo SMART SUPPLY CHAIN FOR BIG DATA ANALYSIS", Mendeley Data, V5, DOI: 10.17632/8gx2fvg2k6.

APPENDIX

APPENDIX I:

MACHINE LEARNING ALGORITHMS

ALGORITHM	DESCRIPTION
BAYESIAN NETWORK	A probabilistic graphical model that represents a set of variables and the class conditional probability and prior probability.
DECISION TREE	A decision support tool that creates and uses a tree-like model of decision and the outcomes . Also This algorithm classifies the data into smaller subsets where each subset contains responses of one class i.e either "yes" or "no".
SUPPORT VECTOR MACHINE(SVM)	A boundary detection algorithm to find a hyperplane in an N dimensional space that classified data points. It works by separating data points to their respective classes
CLUSTERING	Algorithms such as k-means clustering find k centroids by dividing the data into k clusters
DEEP LEARNING	Subset of ML that has networks that can learn unsupervised from the data that is unstructured or unlabeled. Most common model is CNN.
ARTIFICIAL NEURAL NETWORK(ANN)	It is a computational and mathematical model that is inspired by the biological nervous system. The weights in the network learn to reduce the error between actual and prediction

APPENDIX II:

MACHINE LEARNING TERMINOLOGIES

Example	Instances of data used for learning.
Test Sample	Examples used for evaluating ML algorithm performance .
Training Sample	Examples used for training Ml algorithm .
Validation Sample	Examples used for tuning the parameters of a learning sample .
Features	Set of attributes associated with an example.
Hyperparatmeters	Parameter used as input in ML algorithms.
Label	Values assigned to Examples

APPENDIX III

METHODS	MACHINE LEARNING ALGORITHMS
CROP YIELD PRODUCTION	ANN, Bayesian network, Regression, SVM, Deep Learning, Decision tree
PREDICTING SOIL PROPERTIES	ANN, SVM, Clustering, Deep Learning, Regression
IRRIGATION MANAGEMENT	ANN, SVM, Clustering, Regression, Ensemble Learning
WEATHER PRODUCTION	ANN, Decision Tree, Clustering, Regression
CROP PROTECTION	ANN, Decision Tree, Clustering, Regression, SVM
WEED DETECTION	ANN, SVM, Clustering, Deep Learning, Regression
LIVESTOCK MANAGEMENT	ANN, Bayesian network , Clustering, Deep Learning, Decision Tree, Regression
SITE-SPECIFIC NUTRIENT MANAGEMENT	ANN, Clustering, Deep Learning, Decision Tree, Ensemble-learning, Regression, Instance-based learning
DEMAND PREDICTION	ANN, Genetic Algorithm
PRODUCTION PLANNING	Bayesian Algorithm, Clustering, Genetic Algorithm
TRANSPORTATION	Clustering, Genetic Algorithm, Regression
CONSUMER ANALYTICS	ANN, Bayesian Algorithm, Clustering, Decision Tree, Genetic Algorithm
INVENTORY MANAGEMENT	Genetic Algorithm