INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact factor: 4.295
(Volume 4, Issue 3)
Available online at: www.ijariit.com

Prediction model of crop yield for food crop grown above ground level through big data analytics

Varisha Ashraf
varisha193@gmail.com
School of Engineering and Technology Jain University (SET JU), Bengaluru, Karnataka

Ankit Jain
ankitjaincr@gmail.com
School of Engineering and Technology Jain University (SET JU), Bengaluru, Karnataka

Manjunath C. R
Manjucr123@gmail.com
School of Engineering and Technology Jain University (SET JU), Bengaluru, Karnataka

Sahana Shetty
sahanashe@gmail.com
School of Engineering and Technology Jain University (SET JU), Bengaluru, Karnataka

ABSTRACT

Agriculture is believed to be as the backbone of Indian economic system. For the past few decades, agriculture field has seen lots of technological changes to improve better productivity. Day by day the population is increasing leading to increasing demand for resources but the amount of resources required has been reducing and falling down. Therefore, there has been extensive endeavors to create imaginative and technological advances methodologies for manageable harvest generation. Using prediction methods, farmers can enhance the productivity of crops. These strategies are utilized to find the required number of crops, seeds, moistness, water level and other supplements. Since prediction refers to a statement about an uncertain event, hence modeling the prediction would a good solution to adopt. Predictive modeling uses statistics to predict outcomes. Quantifying the yield is essential to optimize policies to ensure food security. This paper aims at providing a new method to predict the crop yield of food crops grown above the ground level based on big-data analysis technology, which differs with traditional methods in the structure of handling data and in the means of modeling. Firstly, the method can make full use of the existing massive agriculture relevant datasets and can be still utilized with the volume of data growing rapidly, due to big-data friendly processing structure. Secondly, the "nearest neighbors"modeling, which employs results gained from the former data processing structure.

Keywords: *Agriculture, Data processing, Nearest neighbors, Big-data, Relevant datasets.*

1. INTRODUCTION

Agriculture is an important source of survival and has been accepted to be as the backbone of the Indian Financial system. Since years, the agricultural domain has seen tremendous technological and mechanical changes to improve and increase productivity. From the time of cr and developing of people on the planet, it has dependably been a competitiveness to bargain with the asset s to meet the key necessities. The exchange asset is land, water and discuss as said within the Upanishad, Vedas and in ancient Hindu composing and our basic duty is to bargain with this resourcefulness capably and keep them clean. Inside the latest century, there has been an exciting increment inside the quantity of the two individuals and animals. The rapidly extending urbanization and industrialization in recent decades are quickly eating up all the trademark resources and are defiling astonishingly quicker than we had ever envisioned. There has been a enormous hole between the ask and supplying of food and consumable Synonyms/Hypernyms (Ordered by Estimated Frequency) of noun asset and beginning nowadays we do not have true blue sky course of action or instrument to handgrip the crisis since of upshots like climate, temperature and another constituent which are past our restraint .

Our nation has the capacity to achieve uncommon addition within the gather surrender creation with the help of expansion of water framework and mechanical procession in refinement. There are different method acting which can be received to upgrade the phone number and nature of crops. In separate with other made nations completing the essential progression is an extraordinary test in India since of the nonappearance of up keep of the assets on which the time frameworks are subordinate. The productive utilization

of respect developing depends upon diverse components, checking the degree to which conditions Interior a field are known too, overseen, the ampleness of information suggestion and the level of client ascendency. The country India has the capacity to accomplish exceptional increment in the harvest yield creation with the assistance of extension of the water system and mechanical development in horticulture. There are various procedures which can be adopted to enhance the number and nature of crops. In contrast with other created nations accomplishing the fundamental development is an extraordinary challenge in India because of the absence of upkeep of the assets on which the generation frameworks are dependent. The fruitful utilization of value cultivating depends upon different components, checking the degree to which conditions inside a field are known also, managed, the adequacy of data proposition and the level of client control. The extraordinary sorts of value horticulture join a large display of topics including changeability of the dirt resource ground, atmosphere, crop hereditary qualities, trim decent variety, equipment execution and most physical, a substance also, normal data sources used for crop creation. The nature of agribusiness must fit the necessities and capacities of the farmer and must be profitable.

According to World Trade Organization in the coming decade, feasible harvest creation will be caused by ecological degradation. Therefore, there has been extensive endeavors to create imaginative and technological advances methodologies for manageable harvest generation. sing prediction methods, farmers can enhance the productivity of crops. These strategies are utilized to find the required number of crops, seeds, moistness, water level and other supplements. This may anticipate giving a too high measure of supplements for development, spares cash on pesticides and composts, yet additionally expands yield of product. The real thought of this idea is a higher amount of yield can be acquired by developing required yield at the opportune place and right cost at the right time.

Predictive Analysis Agricultural systems are exceptionally fruitful in expanding profitability and effectiveness of yield creation. However, population grows steadily, while the resource for crop production diminishes day by day. Traditionally agriculture includes planting or harvesting the yield against a foreordained calendar. Precision farming includes gathering real-time information on climate, air quality, soil, crop maturity, hardware, labor costs and availability of existing data.

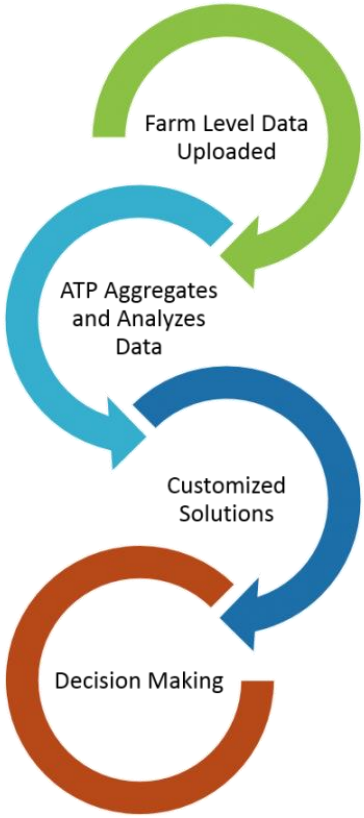


Fig 1.1 Bigdata in Farming

Precision farming takes after perceiving the earlier year yield esteems, understanding the present prerequisite of the condition and misusing data that measures varieties in soil and harvests inside rural fields. Over the past decade, precision horticulture methods play a crucial part in financial framework. Despite the fact that there are various methods accessible, agriculturists find it difficult to deal with the productivity in a stable way. The prediction agriculture idea is in light of the acknowledgment of what occurred in past, examining what is going on in current circumstance and controlling what's to come information require in horticulture. To predict the yield level, past couple of year information is accepted to be as a contribution for the calculation also, the predicted value appears as chart utilizing matplotlib. Here forecast parameters incorporate earlier year crop yield, cultivated territory, water system composes, humidity zone, use of pesticides, manures and labor cost.

The predication model incorporating precision modeling agriculture will then benefit the farmers by providing such benefits that are discussed below. The precise crop variety determination, correct composes and measurements of composts, pesticides, and herbicides, and legitimate irrigation meet the requests of harvests for ideal development and improvement. It leads to increased

yield, particularly in primary territories where uniform product administration methods were customarily drilled. The expectation of cutting-edge advancements, including hardware, apparatuses, and data about information, causes the agriculturists to build the proficiency of work, land and cultivating time. The Prediction of correct amounts of yield at the suitable time diminishes the cost of agrochemical in crop production and furthermore lessens the general cost per unit. Because of expectation, better decision making is done at right time. Also, horticultural apparatuses cause the ranchers to break down and deliver precise data for fitting better decision making in land preparation, seed advances, picking manure, pesticide and herbicide application, water system and seepage, and postproduction.

India is one noteworthy maker of products on the planet. There are different approaches to segment the sorts of products (in view of the area, season, budgetary regard and so forth.). In light of seasons, the yields in India are apportioned into three sorts; Rabi, Kharif, and Zaid.Where Rabi crops incorporate wheat, oats, potato and so forth and are ones sown in October-November and collected in April-May and require the warm atmosphere to edit. Kharif crops incorporate rice, maize, dal and so on, and are sown in June-July and are reaped in September-October.They are monsoon crops that require plenty of water and humid weather conditions to grow. Zaid crops include Cucumber, Pumpkin etc, and are grown in March-June. They are early maturing crops and grown between the Rabi and Kharif crop seasons. Depending on the usage crops can be divided into four major categories. They are

a. Food Crops such as Wheat, Maize, Rice etc.

b. Plantation Crops such as Coffee, Coconut, Rubber etc.

c. Horticulture crops such as Fruits and Vegetables.

d. Cash Crops like Sugarcane, Tobacco, Cotton etc.

Based on the usage of crop the aim is to provide a new prediction model for prediction of crop yield for food crops that are petit in size and grow above the ground level based on bigdata analytics technique, which differs with the various existing traditional techniques within the structure of handling information and within the implies of modelling. Firstly, the strategy can fully utilize the existing gigantic agribusiness important datasets and can still be utilized with the volume of information developing quickly, due to big-data friendly handling structure. Besides the "closest neighbors"modeling, which employs results picked up from the previous information handling structure.

1.1 PROBLEM STATEMENT

Agriculture is the backbone of the Indian economy and one important source of survival. As the days pass by, the population has been increasing but the number of supplements available has been reducing and soon there will be a time where we will be left with not enough supplements. There has been a huge gap between the ask and supply of food and edible assets and beginning nowadays we do not have a genuine course of action or instrument to handle the crisis issues like climate, temperature and other variables which are past our control.

It is of extraordinary incentive to get the crop condition data at beginning periods of cultivation. A similar undertaking is more critical in here and there to gain the correct generation after the harvesting time. Prediction analysis in agribusiness is a crop management idea to increase the environment, economical market and open weights on arable agriculture. To actualize this strategy, past couple of year yield esteem, developed zone, water system strategies, use of manure and pesticides, precipitation level at each season, soil development and climate state of specific region points of interest must be advanced in a viable way. The general idea comprises of following exercises. To start with action is information accumulation from earlier year measurements. Then processing that information using agro algorithm proposed, yield esteem is anticipated

2. LITERATURE SURVEY

There exist a few systems and proposed instrument/models for the forecast of harvest yield with progressive methodologies of inspecting and ordering datasets. Be that as it may, they scarcely ever talk about the issues and techniques for managing expansive and convoluted datasets are along these lines arranging huge datasets remains an extremely intense and dubious test with an extra desire of more noteworthy general execution makes it all the more difficult.
Hsiao et al. (2006) removed coalition designs through social information mining keeping in mind the end goal to enable choice to emotionally supportive network.
Jia (2009) utilized spatial information digging for arrive reviewing. Rajesh (2011) proposed spatial information digging approach for agribusiness.

Kumar et al. (2012) proposed a technique for worldwide auto relationship relating to spatial information mining. Sharma and Mehta (2012) proposed information mining systems to remove business insight from horticultural information. Julia et al. (2012) proposed a procedure for checking horticultural fields through spatiotemporal mining. Yethira (2012) investigated information digging strategies for the field of farming. Mucherino and Rub (2012) examined the most recent improvements in information mining and the field of agribusiness. Rub (2012) investigated spatial information digging for exactness horticulture. Rao et al. (2012) investigated issues and applications in spatiotemporal information mining. Madhuskar and kelkar (2012) made a decent survey of information mining systems.

As of late Cao et al. (2013) proposed a strategy for maize yield forecast in agribusiness. Bhojani (2013) investigated information digging systems for learning revelation in agriculture. Singh et al. (2014) contemplated spatio-fleeting minor departure from earth

surface air. The investigation by H. Guo et.al arranged multi-connection bunch by a few read manifestations while not change inside the first dataset exploitation MRC recipe. the basic arrangement is to utilize cluster methods that progressively utilizes information preparing approach at totally extraordinary levels. At long last Viewed approval recipe was wont to approve the perused however this approach, however, wasn't helpful for horrendously goliath and troublesome relative datasets.

A mode bunching procedure for social databases with the guide of a collection of probabilistic becoming more acquainted with and thinking for direct scaling used to be proposed through Ben Taskar et.al. Tweaking groups and mannequin determination used to be proficient by utilizing training datasets. The reasoning of bunch area know-how is remarkably fundamental for demonstrating, however, transforms into phenomenally difficult to secure and put in the drive in real time if there should arise an occurrence of electronic mannequin development.

Table 1. Examples of Public Agricultural Big Data (*Adapted from* [1])

Type of Public AgBD	Provider
Satellite imagery and meteorological information	**Cloud-computer based** (e.g., Earth on Amazon Web Services [28], Google Earth Engine [29], and NASA Earth Exchange [30]) and others (e.g., National Oceanic and Atmospheric Administration (NOAA) [25], National Aeronautics and Space Administration (NASA) [26] and U.S. Bureau of Labor Statistics [27])
Survey data	National Agricultural Statistics Service (NASS)* [31]
Financial data	Economic Research Service (ERS)* [32] National Water Economy Database (NWED) [33]
Scientific data	Agricultural Research Service (ARS - U)* [34]
Soil, water, and geospatial data	Natural Resources Conservation Service (NRCS)* [35]
Price and sales data	Agricultural Marketing Service (AMS)* [36]
Commodity and market data	World Agricultural Outlook Board (WAOB)* [37]
Generic data	Global Open Data for Agriculture and Nutrition (GODAN)* [24] VegScape* [38]
Animal disease incidence data	World Animal Health Information System (WAHIS) [39] EMPRES Global Animal Disease Information System (Empres-i2) [40]
Citizen data	Social media platforms (e.g., Twitter posts about food)

* Service provided by U.S. Department of Agriculture (USDA)

Different endeavor is created to survey the utilization of knowledge mining way of life within the subject of agriculture. Soft computing and advanced technical ways are applied within the field of agriculture like artificial neural networks, the k nearest neighbor, the k-means, support vector machines and ID3 algorithm. the appliance of knowledge mining techniques in farming could be a moderately new approach and provides new insight into animal direction, program line and within the prediction of agricultural craw.

So as to forecast enhanced nature of the quality of crop one should know have a total idea regarding the harvest in terms of cultivation, plant science, and other important zones. Requires being taken an interest in crop plantation. The following basic steps are some of the guidelines required in quality farming:

A. Assessing Variability
B. Managing Variability
C. Evaluate

3. BACKGROUND WORK

3.1 EXISTING SYSTEM

The present framework depends on Constrained K-implies Algorithm. Here the compelled based and dividing techniques are converged in the calculation beneath.

1) K-mean Algorithm
   a. In this algorithmic normal info like kernel matrix, a variety of clusters, weight for every function, stopping criteria, a penalty term argument area unit other.
   b. For better performance, good timbre clusters are used for huge information sets.
2) Expectation Maximization (EM)
   Algorithm: The parameters are refined based on prospect and maximization senses of steps.

In the present system, we get the facts from the Distributed File System (DFS) which is a customer/server-based application that permits customers to access and method facts put away on the server as although it had been own PC and study it line with the aid of line and operate normalization for the data via taking statistical common imply of data.
Here normalization is a procedure of rebuilding a social database as per a progression of supposed typical structures with a specific end goal to reduce data redundancy and improve data integrity.

There are numerous issues adopting quality farming in all the international locations based on farming across the globe. These problems differ in terms of areas but some troubles are particular to the Indian conditions as cited below:

- Institutional constraints, land possession, and infrastructure
- Observation.
- Technical Gaps
- Small farm size is smaller.
- Lack of Achiever stories.
- Local mechanical expertise is very few.
- Market failures and Heterogeneity of harvesting systems.
- Quality, costs, and data availability.
- It lacks the modern artificial intelligence and machine learning techniques that would still manage to operate the existing system in a more efficient manner.

4. DESIGN METHODOLOGY

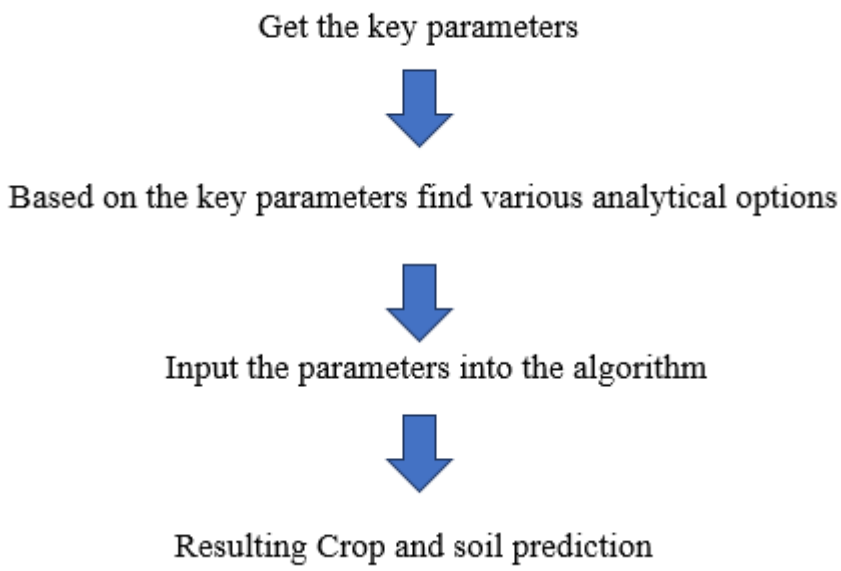


Fig 4.1 Architecture of Proposing Prediction Model

Numerous analysts contributed towards crop yield forecast. The examination concentrated on the expectation of various harvests. The techniques utilized were similarly extraordinary. We believed that consideration of all possible parameters in crop yield prediction can improve the accuracy of prediction.

We then proposed an architecture in which we initially get normalized data from multiple sources, sets and systems, then analyze the collected resource data. Soon after the analysis, we perform the classification procedure using different classifiers needed such as Bayes classifier.

Bayes theorem:

$$P(A | B) = P(B | A)*P(A)/P(B)$$

This equation applies to any probability distribution over events and, it has a particularly nice interpretation in the case where A represents hypothesis and B represents some observed evidence.

Expected Output:

This further uses a weighted mean Poisson distribution for the prediction which is given by

$$p(x) = \begin{cases} \dfrac{e^{-\alpha}\alpha^{x}}{x!}, & x = 0,1,\ldots \\ 0, & \text{otherwise} \end{cases}$$

Where alpha is mean.

Through the classification and distribution further, the final prediction for the soil and crop to be planted is gained.

Features:

With the assistance of Big Data performance tools, investigate the data in light of the need of zone and sort of yield suggestion can be passed on to farmers through this structure. The structure can contain features as indicated:

- Firstly, System farmers come to consider the different diseases of the considerable number of harvests and get the solution for it from the specialists.
- New Innovations to keep pace with the cutting-edge world by providing up-to-date distinctive cultivation techniques for ground-level crops.
- Using the advancing system of rural things with picture will help the farmers to offer their agricultural items as they want from any side of the country.
- The measure of cultivation field can be provided as well the model can forecast an estimated measure of the cultivated items in any season.

Fig 4.2 Important factors affecting our system

5. PROPOSED SYSTEM

Our program system is aimed at creating Agriculture Department well organized by predicting and so enhancing the crop yield by victimization soil data. We tend to propose a brand new Agro algorithmic rule that is utilized to foresee the nature of a yield for a chose soil kind and improves the nature of agrarian generation.

This likewise encourages the agriculturists to choose a specific yield to sow contingent upon the climatic condition and gives fundamental data to pick the best climate to do quality cultivating.

Algorithm:

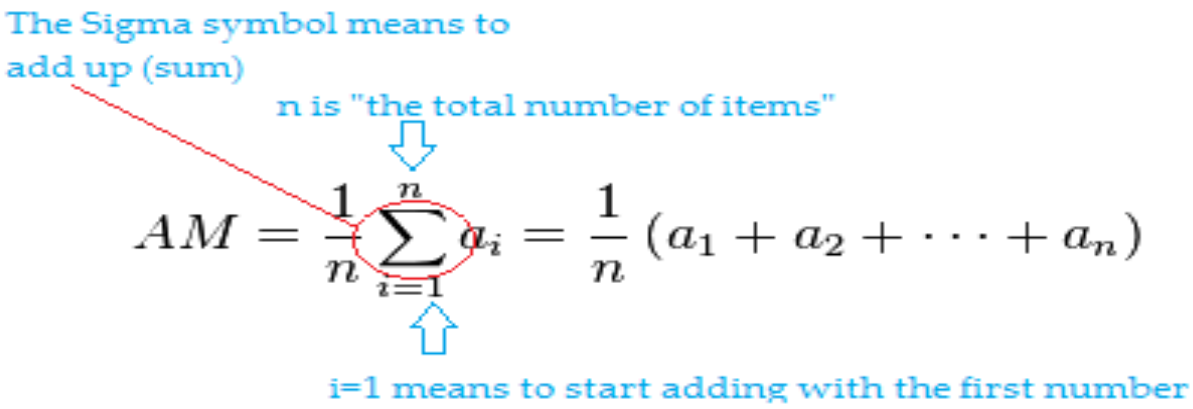
Step1:
Collection and refined supported commonality uses like soil sort, temperature, humidity, water level, spacing, depth, soil ph , season, chemical and months of the data set are done. This knowledge sets square measure entered into the information victimization my SQL queries. From these parameters name of the crop and web yield rate of the crop may be foreseen.
Step 2:
In view of a scope of examinations, the parameters soil compose and temperature are taken as entering and forecast have been attempted. The property soil kind determines the sort of soil in a particular zone, for example, Alluvial, Loamy, Black soil, Clay and Red and the characteristic temperature indicates the water content reachable in the dirt.
Step 3:
By the utilization of Bayesian calculation, the one of a kind crop is analyzed and anticipated by utilizing different parameters into an account such as soil kind and temperature.
How to predict and analyze?

1. Identify the tuples with attributes in the datasets.

2. Summarize the properties in the dataset and calculate the probabilities and make predictions. After that, use the summarize dataset to generate a single prediction.

3. The classifier will predict the particular attributes needed with the conditional probability theorem.

1. If the needed attributes are not available in the data sets then the conditional probability theorem predicts the suitable output and will be analyzed.



The classification algorithm is simple, fast learning and training statistical-based algorithm that requires small runtime memory. Using the classifier, the data based on disease is classified and prediction regarding the soil and particular type of crop are made.

5.1 Limitations of Our Current Work

- Inability to control the weather
- Inability to choose the right fertilizer
- Limited data storage and preservation: The increasing volume, variety, and velocity of agriculture big data sets demands excessive computing power and computational resources to manage and analyze.
- Insufficient data documentation: To support big data analytical methods in agriculture, such as data mining, it is increasingly common for satellite imagery to be supplemented with more and more field data.
- Data sharing barriers: Data privacy is a major concern and private owners may be reluctant to share their data.

6. FUTURE SCOPE

There exist distinctive sorts of sicknesses influencing a specific sort of harvest and it is often hard to anticipate which sort of ailment may happen its recurrence, the reason of cause and speculative period of time of the event. In Future, the examination will consolidate arrangement of diverse sorts of malady for a specific yield and help in identifying expectations with respect to the region which would upgrade crop generation and additionally wipe out wastage.

7. CONCLUSION

As a farming country, we have a tendency to need a solid advancement and information base structure to stay pace with exhibit day factory farm of the planet which can create our country an especially developed one. This paper helps to create a new well-organized

way of performing agriculture and improve the crop yield. A new Agro-algorithm is introduced which helps predict the suitability of a crop for a particularly unique type of soil and enhances the agricultural production. This results in helping the farmers to increase the crop yield relying on the climatic condition and offers indispensable information to select the first-class weather to do the best farming.

Areas	AgBD Opportunities and Research Needs
Workforce Development	Training for farmers and AgBD companies with coursework on big data methods at land-grant universities and beyond in collaboration with department of education.
CyberInfrastructure	NIFA-supported storage‡ for valuable AgBD sets along the lines of NIFA support for genome databases (e.g., MaizeGDB) and seed banks (e.g., the USDA Plant Introduction System). Improve rural broadband infrastructure to support AgBD collection in rural areas.
Private data sharing and compilation	Models for sharing private AgBD. For example, administrative data may provide behavioral and societal information that are not well studied. Standards for sharing of private AgBD (e.g., data format, statistical guidance). Methods for compiling public and private AgBD.
Novel Data Collection	New data collection methods for model validation, combined with funding the development and testing of algorithms to fill in data gaps using predictions from existing information (e.g., remotely sensed data, market data). Public data on food movement and food consumption. New approaches to improve data transfer capacity between farms and data center (e.g., use of TV white space or other less frequently used channels).
Spatiotemporal Machine Learning	Leveraging of new high-resolution (e.g., daily, 1 meter) satellite data to monitor crops on a large scale. Spatiotemporal hotspot detection to identify risks in supply-chain (e.g., heavily localized plants that are subject to climate change). Spatial optimization for land-use and land-cover allocation, and identification of potential production improvements through changes in management. Disease risk forecasting for livestock, including environmental, epidemiological, and weather-related data.
Mechanistic Models	Combining empirical models and mechanistic models to link observation with theory.
Citizen Engagement	Social Media, Apps, Easy-to-use Decision Support for growers and ranchers. Downstream behavioral change through apps (e.g., reduce food waste). Cognitive and behavioral science applied to enhance feedback for technology improvement, scientific advancement and innovation.
Data analytics in animal agriculture	Application of data science methods to detect aberrations in AgBD data streams, which may indicate changing or emerging threats to animal health. New approaches to optimize animal health and production through linking processes occurring at multiple spatial and temporal scales. Development of data pipelines to promote analysis of data in near real-time.

8. REFERENCES

[1] Anu Peisker and Soumya Dalai, "Data Analytics for Rural Development Indian Journal of Science and Technology", Vol 8(S4), 50-60, February 2015

[2] Purnima Shah, Deepak Hiremath and Sanjay Chaudhary, "Big Data Analytics Architecture for Agro Advisory System", 2016 IEEE 23rd International Conference on High-Performance Computing Workshops

[3] Er. Rupinder Kaur, Raghu Garg, Dr. Himanshu Aggarwal, "Big Data Analytics Framework to Identify Crop Disease and Recommendation a Solution", 2016

[4] T. Rajasekaran , P.Jayasheelan and K.R. Sri Preethaa3," Predictive Analysis in Agriculture to Improve the Crop Productivity using ZeroR algorithm" International Journal of Computer Science and Engineering Communications, Volume.4, Issue.2 (2016)

[5] SPATIAL-TEMPORAL DATA MINING FOR CROP YIELD PREDICTION, International Journal of Advanced Research(IJAR) 23rd October 2016, Aakunuri Manjula and Dr. G. Narsimha -Research Scholar and Associate Professor, CSE, Affiliated to JNTUH Hyderabad, Telangana State, India
[6] Development of Smart Crop Production System using Big Data- International Journal of Research and Innovation in Applied Science (IJRIAS)|Volume II, Issue V, May 2017 -Shikha Ujjainia1, Pratima Gautam2, S. Veenadhari3
[7] Analysis of agriculture data using data mining techniques: application of big data (2017) Jharna Majumdar*, Sneha Naraseeyappa and Shilpa Ankalaki
[8] Big Data in Smart Farming – A review, Elsevier (2017)- SjaakWolfert, LanGeaCorVerdouwabMarc-JeroenBogaardta
[9] Ch. Chandra Sekhar, Ch Sekhar, "Productivity improvement in agriculture sector using big data tools",2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDACI)
[10] E. Manjula and S. Djodiltachoumy "A Model for Prediction of Crop Yield" International Journal of Computational Intelligence and Informatics, Vol. 6: No. 4, March 2017
[11] Aditya Saxena, Surendra Pratap Singh, Manju Rani, Manoj Kumar, "A Study of Crop Yield pattern with Climate Change based on Physical Parameters: Temperature and Rainfall in Western Uttar Pradesh to make future predictions for better Crop Management and Yield".
[12] Yiqing Cai, Kristen Moore, Adam Pellegrini, Aymn Elhaddad, Jerrod Lessel, Christianna Townsend, Hayley Solak, Nemo Semret "Crop yield predictions - high-resolution statistical model for intra-season forecasts applied to corn in the US" Gro Intelligence, Inc.

