Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

Crop and Fertilizer Recommendation System

B. LALITHA RAJESWARI[1], Assistant Professor,Department of CSE,
Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh.

SK. ABDUL MUHEETH[2], **SK. VASEEM NAAZLEEN**[3], **T. PAVAN KUMAR**[4], **V. PHANINDRAAMOULI**[5]
[2,3,4,5] UG Students, Department of CSE,
Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh.
lalitharaji.raji@gmail.com[1], shaikabdulmuheeth@gmail.com[2], shaikvaseemnaazleen@gmail.com[3], thippabathina55@gmail.com[4], v.mouli2002@gmail.com[5]

DOI:10.48047/IJFANS/V11/I12/179

Abstract

Being a source of food, raw resources, and jobs, agriculture is essential to the global economy. However, with a growing population, the demand for food production has increased, making it imperative to improve crop yield and sustainability. Since agriculture is greatly influenced by the surrounding natural conditions, we face many challenges in actual agriculture practices. one of the biggest challenges faced by farmers in determining the right crop and fertilizer to use for optimal yield. Efficient technology can be used to increase yields and reduce possible challenges in this area. One approach is to use machine learning techniques to propose crops and fertilizers to farmers based on their unique needs. In this article, we present a crop and fertilizer recommendation system developed using efficient ML models. We link our model with a web application that allows users to input their data and receive personalized recommendations in multiple regional languages. Our system aims to provide farmers with an easy-to-use tool that can help optimize their crop yield and increase sustainability.

Keywords – LightGBM, GOSS(Gradient One Side Sampling), EFB (Exclusive Feature Bundling)

1. Introduction

Crop and fertilizer recommendation systems may be extremely beneficial to farmers in terms of increasing yields and earnings. In this article, we present a machine-learning model based on LightGBM that can recommend crops and fertilizers based on various input features. Our model was trained on publicly available datasets of crop and fertilizer data, and we evaluated its performance using a range of metrics.

To make our model more accessible to farmers, we also developed a web application that allows users to input their data in various regional languages. The application then returns recommendations for crops and fertilizers based on the input data, making it easier for farmers to make informed decisions about their crops and fertilization strategies.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

2. Literature Review

In [1], the paper aims to build a crop recommendation system. The recommendation system is built by training an ML model on the XGBoost, Random Forest, Logistic Regression and Decision tree. Where the highest accuracy is obtained by the XGBoost algorithm i.e,98%. And, they have developed a web application which takes feature values to predict as input and displays the predicted crop to the users.

In [2], this paper mainly describes two things, Finding the Grade of the Soil and Crop Recommendation. Finding the grade of the soil helps farmers to understand the quality of the soil by considering the various soil nutrients, this is found by the regression algorithm and the Gradient Descent Algorithm minimizes the cost function. Crop recommendation helps to find the suitable crop for the soil details provided, this paper considered the dataset from the ICRISAT Development Center of AP Government, to train the model with the above various supervised ML algorithms like Naïve Bayes, SVM and Random Forest are considered, among all Random Forest Classifier has given the highest accuracy of 72.74%.

In [3], In addition to offering optimal crop recommendations for a particular upazila in Bangladesh, the research attempts to develop a suitable model for categorizing various types of soil series data. It mainly describes two aspects, i.e, soil classification and crop suggestions. The paper's soil dataset was gathered from six upazilas in the Khulna area, Bangladesh. It contains a class label (upazila code) and chemical features of soil (pH, Salinity, Organic matter, P, S, Zn, B, Calcium, Mg, Cu, Fe, and Manganese). The crop dataset is created based on the class label, map unit, upazila code, and crops. Firstly, the given soil is classified and a class label is generated using weighted K-NN, Bagged Tree, and, Gaussian Kernel-based SVM ML techniques. Based on the generated class label, they have shown the crop suggestions. Among all considered ML techniques, SVM gives the highest accuracy of 94.95%. The paper limits itself to a specific region, i.e, Khulna district, Bangladesh. It is difficult to extend a crop dataset to cover all over India. The accuracy can be further increased by using boosting techniques.

In [4], the paper aims to recommend a crop to users based in South Indian states like Karnataka, Kerala, Tamil Nadu and Andhra Pradesh. This paper doesn't take any information regarding soil nutrients and quality but rather it takes input about the region i.e, state and district and season. In this paper, the model is trained with a Random Forest Classifier. The web page is built as part of this paper, which has login and registration of users who want to access model prediction. Along with crop recommendation PH, Rainfall and Temperature suitable to grow the crop are provided. Crop Rotation is made Feasible by providing alternate crops which can be grown considering the predicted crop type.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

In [5], the paper aims to come up with a strategy for crop selection based on weather and soil factors to optimize agricultural output. It mainly describes two aspects, i.e, Sessional weather prediction based on that it tries to identify a suitable crop. The data utilized for weather forecasting comes from NRSA Hyd station containing 5 years' worth of meteorological characteristics like temperature, moisture, sun hours, wind direction etc. Recurrent neural network (RNN) has been used to forecast seasonal weather. The crop dataset contains soil features, temperature, humidity etc and crop as the class label. The weather forecast obtained is used to predict the crops. Machine learning models like K-Nearest Neighbour, Decision Tree, and Random forest are used to predict the crops & yield. The accuracy metrics are not mentioned in the research paper but Random Forest most likely gives better accuracy than the rest. The paper can be extended by adding fertilizer recommendations.

In [6], This paper aims to develop a Crop and Fertilizer Recommendation System along with Yield Forecasting and Crop Rotation. Voting-based ensemble classifier containing Naïve Bayes Classifier, Random Forest Classifier, and CHAID Classifier (Chi-square Automatic Interaction Detection) is considered, by following majority voting a crop is recommended to the user. Based on the type of crop recommendation suitable fertilisers will also be recommended to the user. If the area of land is provided as input then the amount of yield will be provided to the user. Crop Rotation is made possible by providing alternate crops which can be grown based on the crop season and recommendation. For Crop Recommendation accuracy of 92% was achieved by following the Voting-based ensemble classifier.

In[7], This research paper attempts to build a model to predict crops based on the input provided, this paper provides a web application which takes phosphorus, nitrogen, potassium, PH, temperature, season etc as input and these inputs are passed to the trained model and it recommends the crop as the output. Along with the type of crop webpage also provides brief details about the type of climate and soil suitable to grow the recommended crop. This paper only recommends 8 types of crops and the model is also trained with the data of these 8 crops only and the model is trained with the KNN algorithm. In this paper Accuracy achieved is not described.

3. Problem Identification

Technological Changes in the agricultural sector tend to be very slow. It is often observed that farmers tend to cultivate the crops which have more demand / high price in the market, without properly considering the suitability of minerals in the soil and other

IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

external factors (rainfall, humidity,...) to grow the crop. Which leads to the cultivation of less yield crops.

By reviewing the considerations and methodologies used in the research papers [10-18], they have tried to develop ML models to recommend crops and fertilizers to produce a good yield. Some considered building a web application to ease the process of accessing the recommendation system. There are certain ways through which we can improve the recommendation system further they are: (i) Rather than asking users to forecast the values of the temperature, humidity, etc. we can make use of the APIs to forecast those values. (ii) Web application built above only supports English. So, expanding language support even further could help reach a wider audience and make the application more accessible to users who do not speak or understand English.

4. Proposed Methodology

We propose the development of a crop and fertilizer recommendation system both distinctly using machine learning techniques. Our methodology involves training a LightGBM model on two datasets - one containing crop recommendations and another containing fertilizer recommendations. The model will be trained on the features present in both datasets, such as soil pH, temperature, humidity, and rainfall.

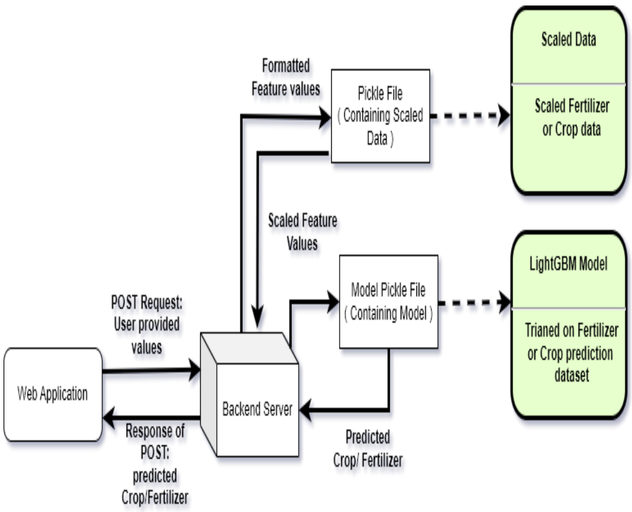


Figure 1:complete working process and architecture of crop and fertilizer recommendation system

After training the model, we will link it to a web application that asks users to enter values of the features present in the datasets. The application will be designed to display text in multiple regional languages for better accessibility. Based on the user inputs, the model will provide recommendations for the most suitable crop or fertilizer for the given conditions.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

We are trying to improve the following aspects through our proposed system: (i) improving the accuracy, speed of processing and reduced memory usage (ii) We will be extending our web applicationto support different regional languages like Telugu, Hindi, Tamil, … (iii) Making use of weather APIs to forecast the values of features like temperature, humidity, etc.

5. Implementation

Implementation of the project can be described through the following steps:

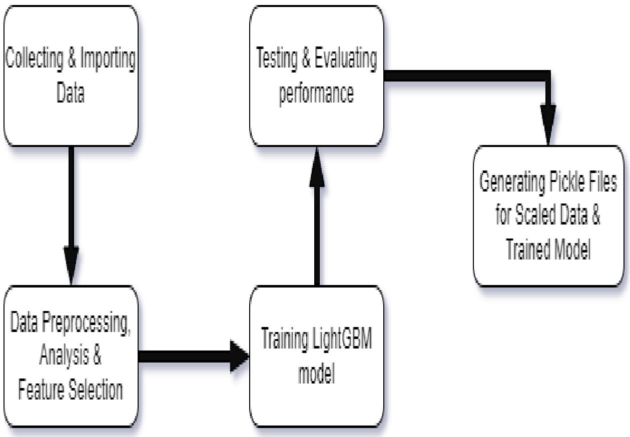


Figure 2: Working process of LightGBM model

1. Data Collection: We gathered the crop dataset from Kaggle, which contains information on various factors such as soil type, rainfall, temperature, humidity, and pH levels, to predict the most suitable crop for a given set of environmental conditions. Similarly, we also collected the fertilizer dataset from Kaggle, which contains information on various fertilizers and their suitability for different crops.
2. Data Preprocessing and Data Analysis: Once we collected the datasets, we performed data cleaning and preprocessing to ensure that the data was suitable for training the machine learning model. Data sets have no missing or incomplete data, no outliers, and encoding. After that, we tried to identify different patterns and relations between the features by finding correlations
3. Model Training and Evaluation: We trained the LightGBM model using the preprocessed data and evaluated its performance using various metrics such as accuracy, precision, recall, and F1 score.
4. Integration with Web Application: With the trained and optimized LightGBM model in hand, we integrated it with a web application to provide a user-friendly interface for users to input their environmental conditions and receive recommendations for

ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

the most suitable crop or fertilizer. We used Flask as the web application framework and deployed it on a cloud-based platform to ensure scalability and availability.

5. Testing and Deployment: Finally, we tested the crop and fertilizer recommendation system thoroughly to ensure that it performs well in a real-world scenario.

5.1 Algorithms

LIGHTGBM ALGORITHM:

LightGBM is a gradient-boosting framework that employs a tree-based learning algorithm. It is designed to be efficient and scalable and can handle large-scale datasets with millions of examples and features. One of the key features of LightGBM is its ability to handle categorical features directly, without requiring one-hot encoding or other preprocessing steps. This makes it particularly well-suited for datasets with mixed data types, such as the crop and fertilizer datasets used in this study. Another advantage of LightGBM is its ability to handle imbalanced datasets. Overall, LightGBM has become a popular choice for machine learning practitioners due to its speed, accuracy, and flexibility by providing a few other features like EFB (Exclusive Feature Bundling) and GOSS (Gradient One Side Sampling).

5.1.1 EFB (Exclusive Feature Bundling): LightGBM safely identifies features which are mutually exclusive i.e, they never take zero value simultaneously and bundle them into a single feature to reduce complexity. LightGBM considers the following properties to find out the mutually exclusive features in the EFB:

- Correlation: LightGBM bundles the features which are not highly correlated with each other. This is because features that are highly correlated with each other are likely to contain similar information and therefore should not be bundled together to avoid loss of information.
- Information Gain: Bundles the features that have similar information gain, which helps to reduce the overfitting & dimensionality.
- Split Gain: Features that have high split gain are likely to be more important and should not be bundled with other features. Split gain measures how much the target variable's variability is reduced by splitting the data based on a particular feature.
- Similarity: Features that are similar in terms of data distribution, mean that they are providing similar information to the model. By bundling these features together, LightGBM can reduce the number of splits needed to capture the information they provide, which can help prevent overfitting.

EFB can reduce the number of features in the dataset and improve the model's performance by reducing the risk of overfitting.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

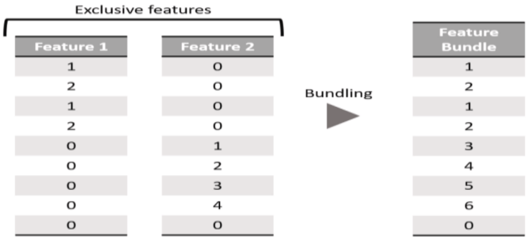


Figure 3: Exclusive Feature Bundling of two features into one[8]

5.1.2 GOSS (Gradient based One Side Sampling): It is a technique used in the LightGBM algorithm to reduce the computational cost of gradient boosting. GOSS aims to sample more examples that have large gradients (i.e., the data points that are more informative for the model) while reducing the number of examples that have small gradients.

The GOSS approach is a two-step process. In the first step, GOSS samples a subset of the training examples based on the gradients of the loss function. Specifically, GOSS calculates the gradient of each example in the training set and sorts them in descending order based on their magnitude. Then, it selects the top 20% of examples with large gradients and 80% of the examples with small gradients will be sampled. In the second step, GOSS trains the LightGBM model using the sampled training examples. This allows the model to focus on informative examples while reducing the number of computations required.

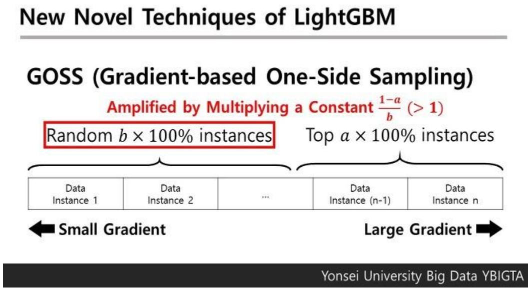
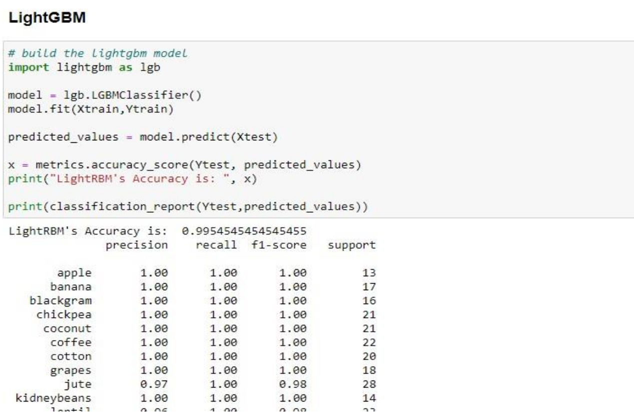


Figure 4: Selection of instances using GOSS[9]

6. Results & Conclusion



LightGBM

```
# build the lightgbm model
import lightgbm as lgb

model = lgb.LGBMClassifier()
model.fit(Xtrain,Ytrain)

predicted_values = model.predict(Xtest)

x = metrics.accuracy_score(Ytest, predicted_values)
print("LightRBM's Accuracy is: ", x)

print(classification_report(Ytest,predicted_values))
```

```
LightRBM's Accuracy is:  0.9954545454545455
              precision    recall  f1-score   support

       apple       1.00      1.00      1.00        13
      banana       1.00      1.00      1.00        17
   blackgram       1.00      1.00      1.00        16
    chickpea       1.00      1.00      1.00        21
     coconut       1.00      1.00      1.00        21
      coffee       1.00      1.00      1.00        22
      cotton       1.00      1.00      1.00        20
      grapes       1.00      1.00      1.00        18
        jute       0.97      1.00      0.98        28
 kidneybeans       1.00      1.00      1.00        14
```

Figure 5: Crop Prediction using LightGBM

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

```
# Set up the grid search with cross-validation
grid_search = GridSearchCV(lgbm, params, scoring="accuracy", cv=5)

# Fit the grid search to the training data
grid_search.fit(X_train, y_train)

# Print the best hyperparameters
print("Best hyperparameters:", grid_search.best_params_)

# Use the best model to make predictions on the testing data
y_pred = grid_search.predict(X_test)

# Calculate the accuracy score
accuracy = accuracy_score(y_test, y_pred)
temp = cross_val_score(grid_search, X_test,y_test,cv=5)
class_report = classification_report(y_test,y_pred)

print("Accuracy: ",accuracy)
print("\nClassification Report: \n\n",class_report)
```

```
[LightGBM] [Warning] lambda_l1 is set=0.1, reg_alpha=0.0 will be ignored. Current value: lambda_l1=0.1
[LightGBM] [Warning] lambda_l2 is set=0.1, reg_lambda=0.0 will be ignored. Current value: lambda_l2=0.1
Accuracy:  1.0

Classification Report:

              precision    recall  f1-score   support

           0       1.00      1.00      1.00         2
           1       1.00      1.00      1.00         1
           3       1.00      1.00      1.00         1
           4       1.00      1.00      1.00         5
           5       1.00      1.00      1.00         5
           6       1.00      1.00      1.00         6

    accuracy                           1.00        20
   macro avg       1.00      1.00      1.00        20
weighted avg       1.00      1.00      1.00        20
```

Figure 6: Fertilizer Prediction using LightGBM

Figure7: Web Application Home Page

Figure 8: Crop Recommender Page

ISSN PRINT 2319 1775 Online 2320 7876

Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

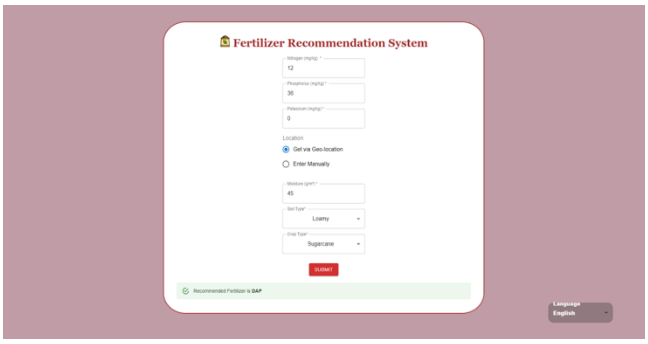


Figure 9: Fertilizer Recommender Page

Model	Recommendation System	ACCURACY
LightGBM Model	Crop	99.54%
	Fertilizer	99.99 %

Table 1: Obtained model Accuracies for both Crop and Fertilizer Recommendation System

In conclusion, the crop and fertilizer recommendation system developed using the advanced gradient boosting algorithm LightGBM has shown promising results in accurately predicting the best crop and fertilizer based on input parameters.

7. Limitations and Future Scope

It can be further extended by integrating pesticide recommendation system. Moreover, we can also consider developing a mobile application to make it easier for farmers to access the recommendation system on their smartphones. Overall, this article provides a promising solution to the problem of crop and fertilizer recommendation, and there are numerous opportunities for further research and development in this field.

References

[1] [1] Atharva Jadhav, Nihar Riswadkar, Pranay Jadhav, Yash Gogawale, (April 2022) “Crop Recommendation System Using Machine Learning Algorithms”. International Research Journal of Engineering and Technology (IRJET), Volume-9 Issue-4, e-ISSN: 2395-0056, Pg.1227-1230.

[2] [2] Keerthan Kumar T G, Shubha C, Sushma S A (Nov 2019) “Random Forest Algorithm for Soil Fertility Prediction and Grading Using Machine Learning”, International Journal of Innovative Technology and Exploring Engineering (IJITEE), Volume-9 Issue-1, ISSN: 2278-3075 (Online), Pg.1301-1304.

Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

[3] [3] Sk Al Zaminur Rahman, Kaushik Chandra Mitra, S.M. Mohidul Islam (Dec 2018) "Soil Classification using Machine Learning Methods and Crop Suggestion Based on Soil Series", International Conference of Computer and Information Technology (ICCIT)

[4] [4] Chaitanya Bn, Dr Kiran B Malagi, Gangothri N V, Lokesh P, Divya Shree (Aug 2021) "Features Identification For Growth Of Certain Crops In Indian Agriculture", International Journal of Creative Research Thoughts (IJCRT), Volume 9, Issue 8, ISSN: 2320-2882, pg: a735-739.

[5] [5] G Ravi Kumar, Geetesh Challur, Pradhan Babu Tammana (Nov 2022) "Machine Learning Techniques For Crop Yield Prediction In Indian Agriculture", International Journal of Scientific Development and Research (IJSDR), Volume 7 Issue 11, ISSN: 2455-2631, Pg. 475 - 482.

[6] [6] K. Archana, Dr K.G.Saranya (May 2020) "Crop Yield Prediction, Forecasting and Fertilizer Recommendation using Voting Based Ensemble Classifier", SSRG International Journal of Computer Science and Engineering (SSRG-IJCSE), Volume 7 Issue 5, ISSN: 2348 – 8387.

[7] [7] Prakash U.M, Sristika Bora, Abhishek Gautam, K.R. Gokul Anand (April 2020) "Knn based Crop and Fertilizer Prediction", International Journal of Engineering and Advanced Technology (IJEAT), ISSN: 2249 – 8958 (Online), Volume-9 Issue-4, Pg:1453-1456.

[8] [8] Hirochika Yamashiro (2021) Estimation of processing time using machine learning and real factory data for optimization of parallel machine scheduling problem, DOI: https://doi.org/10.1016/j.orp.2021.100196

[9] [9] Yonsei University of Big data YBIGTA, (July 2018) from LightGBM: A Highly Efficient Gradient Boosting Decision Tree.

[10] Sri Hari Nallamala, et al., "A Literature Survey on Data Mining Approach to Effectively Handle Cancer Treatment", (IJET) (UAE), ISSN: 2227 – 524X, Vol. 7, No 2.7, SI 7, Page No: 729 – 732, March 2018.

[11] Sri Hari Nallamala, et.al., "An Appraisal on Recurrent Pattern Analysis Algorithm from the Net Monitor Records", (IJET) (UAE), ISSN: 2227 – 524X, Vol. 7, No 2.7, SI 7, Page No: 542 – 545, March 2018.

[12] Sri Hari Nallamala, et.al, "Qualitative Metrics on Breast Cancer Diagnosis with Neuro Fuzzy Inference Systems", International Journal of Advanced Trends in Computer Science and Engineering, (IJATCSE), ISSN (ONLINE): 2278 – 3091, Vol. 8 No. 2, Page No: 259 – 264, March / April 2019.

[13] Sri Hari Nallamala, et.al, "Breast Cancer Detection using Machine Learning Way", International Journal of Recent Technology and Engineering (IJRTE), ISSN: 2277-3878, Volume-8, Issue-2S3, Page No: 1402 – 1405, July 2019.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper **© 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022**

[14] Sri Hari Nallamala, et.al, "Pedagogy and Reduction of K-nn Algorithm for Filtering Samples in the Breast Cancer Treatment", International Journal of Scientific and Technology Research, (IJSTR), ISSN: 2277-8616, Vol. 8, Issue 11, Page No: 2168 – 2173, November 2019.

[15] Kolla Bhanu Prakash, Sri Hari Nallamala, et al., "Accurate Hand Gesture Recognition using CNN and RNN Approaches" International Journal of Advanced Trends in Computer Science and Engineering, 9(3), May – June 2020, 3216 – 3222.

[16] Sri Hari Nallamala, et al., "A Review on 'Applications, Early Successes & Challenges of Big Data in Modern Healthcare Management'", Vol.83, May - June 2020 ISSN: 0193-4120 Page No. 11117 – 11121.

[17] Nallamala, S.H., et al., "A Brief Analysis of Collaborative and Content Based Filtering Algorithms used in Recommender Systems", IOP Conference Series: Materials Science and Engineering, 2020, 981(2), 022008.

[18] Nallamala, S.H., Mishra, P., Koneru, S.V., "Breast cancer detection using machine learning approaches", International Journal of Recent Technology and Engineering, 2019, 7(5), pp. 478–481.