Mathematical Statistician and Engineering Applications
ISSN: 2094-0343
2326-9865

Agriculture Crop Recommendation System using Machine-Learning

Kanika Bhatnagar [1], Mamilla Jaahnavi [2], Dr. B. U. Anu Barathi [3]
[1]Student, Department of CSE
Sathyabama Institute of Science and Technology
Durg, India
kbhatnagar75@gmail.com
[2]Student, Department of CSE
Sathyabama Institute of Science and Technology
Durg, India
jaahnavireddy8668@gmail.com
[3]Assistant Professor, Department of CSE
Sathyabama Institute of Science and Technology
Chennai, India
anubarathi.cse@sathyabama.ac.in

Article Info
Page Number: 626 - 637
Publication Issue:
Vol 71 No. 3s2 (2022)

Article History
Article Received*: 28 April 2022*
Revised*: 15 May 2022*
Accepted*: 20 June 2022*
Publication*: 21 July 2022*

Abstract
Machine learning can be a significant method for tracking down valuable and useful arrangements in the rural world. Limit tests will anticipate the reason/result of the prophets' assortment through ensuing investigations. Make an item that utilizes a rundown of factors to change the contribution to the result factors to get the outcomes you really want. Crop yield determining is the determination of a harvest in light of recorded information showing boundaries like the rainfall, temperature, pH, precipitation, and yield name. It provides us with a thought of a very much anticipated plant that will fill in the field as per the logical season. A reasonable comprehension of horticulture is the utilization of new rural strategies. Random Forest and KNN (K Nearest Neighbor), algorithmic law concentrates on machines seldom make such suspicions. The primary plant will be fruitful before it is feasible to make adjustments. The Random Forest techniques are utilized to deliver fascinating harvests.

Keywords: - Crop yield Prediction, precision Agriculture, supervised Learning, Random Forest algorithm, K Nearest Neighbour Classifier.

I. INTRODUCTION

Agriculture could be an important part of the Indian economy further because of the human future. Work is, 1st and foremost, necessary for survival. It additionally generates a major variety of jobs. The demand for production has adult staggeringly as time has passed. Folks are misusing technology so as to make in massive quantities. Every day, new hybrid variants are developed. These sorts, on the opposite hand, don't deliver a similar essential nutrient as naturally adult crops. These artificial strategies degrade the soil. All of this contributes to larger environmental injury. The bulk of those unnatural strategies are used so as to avoid losses. However, once the growers of those crops have corrected info on crop production, the loss is reduced. This project was created

with the goal of finishing it. The knowledge is provided supported previous weather, temperature, and different things.

Machine learning could be a tool that's want to read and analyze information all told doable ways in which. Once analyzing the information, it's won't to forecast for the longer term. It's a good vary of applications. These patterns reveal crop info. Agriculture statistics are utilized in the dataset. This dataset is the muse for the experiment. Following processing, it's divided into two parts: training and testing.

If computer code can be designed to model the interactive impact of climate factors, particularly the impact of maximum events (e.g., heat, rainfalls, and excess water) occurring at completely different growing phases of crops, the information gap between ancient ways of cultivating and new agricultural technologies may be bridged.

II. Literature Survey

[1] Jeevan depicts the utilization of different machine learning methods to speed up. This provides us with a thought of a very much arranged harvest that can be planted in the mountains.

[2] Sujata clarified that past yield data doesn't anticipate gathering and reaping. It likewise gives ranchers the data they need to handle climate data about crops that can be reaped, and to deal with the substance of the harvest at the perfect opportunity.

[3] The conduct of the mentors of Judah, the numerous choice trees, was obvious, and the division or issuance turned into a penance as indicated by the quantity of divisions.

[4] Amrita clarifies the likely advantages of IoT (Internet of Things) casualties to the climate at a decent cost, and clarifies the expected effects of its utilization and use. They utilized a Hadoop sound card. Data, research, and related data don't assist with recognizing bugs and bugs.

[5] Pavan discloses to us that "Testament Tree and K Nearest Neighbor" utilized the power calculation. This strategy assists with guaranteeing that the plants in a specific region are typically full grown.

[6] Fatima considered mathematical creation and clarified the various approaches to digging for the advancement of time. Since a great deal of information assortment can be troublesome, K implies that it utilizes registering to deal with a ton of data. Genuine numbers are utilized to choose the collect as a long-lasting issue. Furthermore, they center around authority abilities and forefront.

[7] Yash states the usage of agricultural information with data processing and visual data processing techniques are delineate. Data processing in agriculture is employed for analyzing the varied organic phenomenon and abiotic factors.

[8] Nirupama clarified that it assumes a significant part in data handling in the rural area. They need to decide specific computational abilities, for example, K normal, SVM (Support Vector

Machine), ANN (Artificial Neural Network). Collect is anticipated predominantly founded on the climate, with a genuine score of 95%, C4.5.

[9] Rakesh discusses the factors deciding the crop choice like production rate, value and government policies are mentioned. The projected methodology could improve web yield rate of crops.

[10] Use of ecological variables, emphasis examination, direct relapse calculation, and collected item reviles.

[11] Satish states the wants and designing required for developing a computer code model for exactitude agriculture is mentioned. This approach can end up in the farmer and crop-level support advisories through devices like mobile phones and tablets.

[12] Raorane explains how to boost harvest productivity by applying several data mining tactics. The processes they used for order, like as ANN, SVM, and K implies, among others.

[13] Anshal explains during this paper, varied algorithms are used associated with classification techniques of information mining. A comparative analysis is finished to indicate that classification algorithmic rule is best fitted to predicting the yield with relevance classification techniques.

[14] Aakunuri states the need for crop yield prediction and it facilitate in a very nation's strategic political opinions in exactitude agriculture.

[15] Yash states the usage of agricultural information with data processing and visual data processing techniques are delineate. Raw information/data/information collected from the statistics analysis has helped in crucial the information and by victimization self-organizing and multi-dimensional maps to scale back data. Many reviews are presented in literature by many researchers with respect to ecommerce applications in different domain [21][22][23]. This analysis will surely enable the researchers with the idea of deep learning technique in different applications [24][25][26][27]. Different issues also discussed in machine learning applications [28][29][30].

III. Proposed Architecture

For farmers, crop yield is extremely useful information. Knowing the yield can help you save money by reducing your losses. In the past, experienced farmers predicted crop yields. In a similar way, the proposed system operates. It takes the prior data and applies it to forecast future yields. The weather and fertilizers have the greatest impact on crop productivity. The accuracy of the information provided determines the accuracy of this prediction. As a result, the suggested approach anticipates yield and reduces loss. The expected system takes on the role of a seasoned farmer. However, it is more precise and takes into account a variety of additional parameters. Soil condition, weather forecast, pH, humidity and yield are all factors to consider.

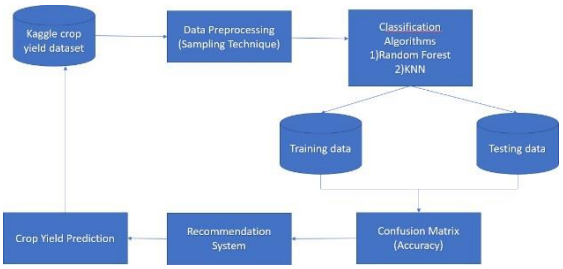

Fig 1: Architecture Diagram

A. Random Forest Classifier

Throughout the coaching section, an outsized range of call trees are created, and therefore the output is separated into two categories: classification and sophistication prediction. Rainfall, perception, temperature, and production are among the variables within the dataset. These variables within the dataset are utilized in the coaching method. Solely simple fraction of the dataset is taken under consideration. The remaining dataset is employed as a work.

Datasets

Suggestions for plants comprising of hotness, downpour, hotness, and pHq. Data programming downloaded from Kaggle site. The information incorporates 2201 inquiries or information from previous history. These incorporate seven boundaries or choices: nitrogen (N), phosphorus (P), potassium (K), temperature, pH, temperature, and temperature. Rural insights are separated into seven classes: area, locale, year, season, yield, district, and efficiency. Information sources incorporate 2.5 models or information from previous history.

Random forest algorithm

An average vault can be a notable learning calculation comprising of controlled learning methods. It is utilized per milliliter and in reverse. It joins different classifications, supporting the possibility of studying outfit to address shortcomings and further develop execution.

As the name proposes, "Typical memory can be classes that remember the previously mentioned tree goals for various pieces of a given informational index, and utilize a similar bearing to be more successful in anticipating this data." Instead of depending on a solitary choice, thoughts were taken for every unique woodland, upheld by many, and the aftereffects of the last discourse were forecasted.

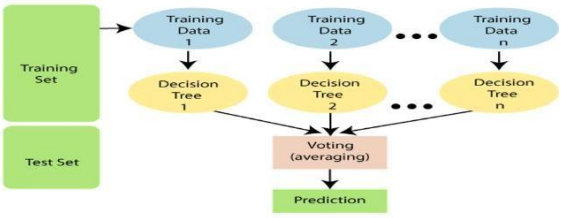

Fig 2: Random Forest

B. K Nearest Neighbour Classifier

KNN (K Nearest Neighbor) Techniques is an instructing based model that exhibits the worth of new data sources when composing pre-example data. Works out the distance between the new information strategy and the model review distance (for example Euclidean, Manhattan, or Minkowski distance) and afterward chooses the proper K (or brief) distance. The weighted normal worth of the K qualities for every one of the neighbors is the worth of the new model worth. Picking the strategy K is a secret; The worth of K is contrasted with the forecast of shock (i.e., better comprehension of data). The lower K shows the huge distinction and the slight contrast, while the long K shows the distinction. The upside of K nearest neighbors is that they don't have to practice or progress admirably. It depends on neighborhood thought and is utilized to take care of issues that are not straight, planned and variable. Since KNN utilizes each datum test to foresee new data, it is existence troublesome. It is an erroneous practice in all parts of machine learning. It expands the size of the vector and creates turmoil. This article analyzes how KNN is utilized in farming.

C. Procedure

The data contained numerous fundamental components like Nitrogen (N), Phosphorus (P), Potassium (K), Heat, Rainfall, Heat, and pH, and was utilized to group and reuse utilizing unlawful ranger service strategies. At first, the choice tree was utilized for model preparing, however it has been observed that the utilization of unlawful memory techniques decreases the seriousness of changes, upgrades, and the utilization of KNN and Random Forest to further develop them.
Information recovered from Kaggle's dataset. To accomplish the best outcomes, we utilized 80% of the information to prepare the example, 20% to test the aftereffects of the set-up information, and afterward to prepare the example utilizing the illicit memory technique. The normal outcomes were then contrasted with the underlying assessments. The reality of the model was determined utilizing a model. Also, numerous strategies have been utilized to precisely anticipate the model. At long last, we have arrived at the resolution that the type of illegal memory makes it significantly more evident. Accordingly, the model was prepared utilizing Random Forest Technique and KNN.

IV. RESULT

Data visualisation is the depiction of interpreting data by displaying it in a graphical environment, allowing for the detection and exposure of patterns, inclinations, and connections.

The most popular charting libraries are:

- Matplotlib: Low-level, gives the user a lot of freedom.
- Pandas Visualization: This border is easy to use. Matplotlib can be used to create it.
- Plotly: Allows you to make an interactive plot for visualisation.

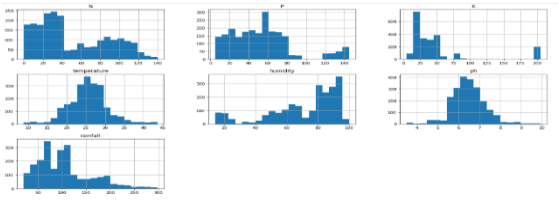

Fig 3: Hist Diagram

Downpour and hotness changes can spread. It is not difficult to see that the hotness dissemination of the property is diminishing fundamentally toward the east. pH dissemination. Gauss or close to Gauss.

Plots are a sort of single direction circuit. It is additionally a method for deciding the appropriation of every trademark in a measurement. Density Plot is a straightforward histogram detail. Thickness addresses an unsteady dissemination measurement. It just acknowledges passages as a numeric rundown.

Density Plot are utilized to track down the dissemination of at least one factors. The main thing to do while recovering new information is to check the dissemination of the factors independently. It gives a great deal of data.

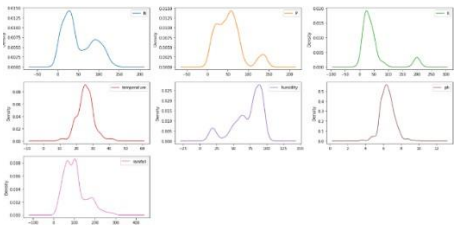

Fig 4: Density Diagram

The degree of conduct joining assists with distinguishing the qualities and shortcomings of the relationship. The positive connection between the factors is communicated as 1. The hazier the number 0, the more terrible the connection.

The framework connection can be set as a gathering variable. Nitrogen (N), Phosphorus (P), Potassium (K), Temperature, Temperature, pH, and Precipitation are the seven fundamental names in this table, and they are typically dispersed by importance.

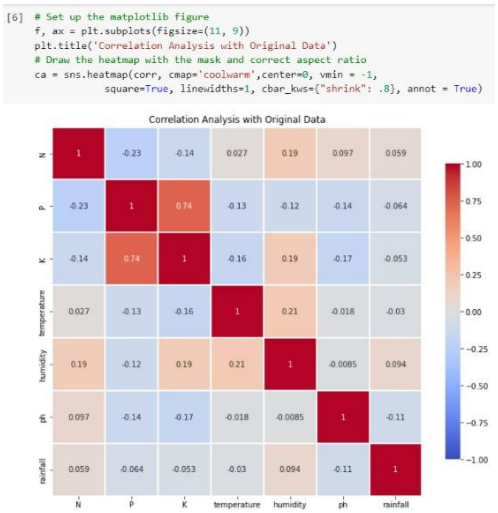

Fig 5: Correlation Diagram

1. Downpour and hotness can spread.
2. It is not difficult to see that the warm properties are disseminated on the left-hand side.
3. Dispersion attributes of pH. Gauss or close to Gauss.

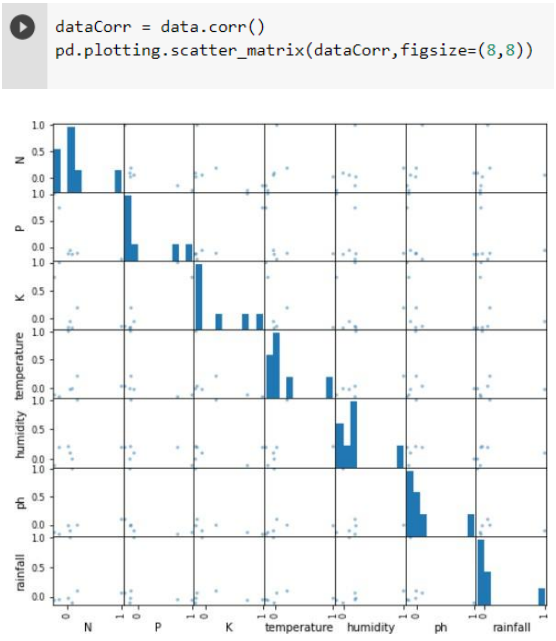

Fig 6: Correlation Matrix Plot Diagram

Crop creation arranging is accessible in all Indian states. In this figure, we will enter the fixings into districts and zones, which will show the plants filling as a cake in light of the info. This item will assist you with deciding the best plants to develop on explicit soils in the Indian state.

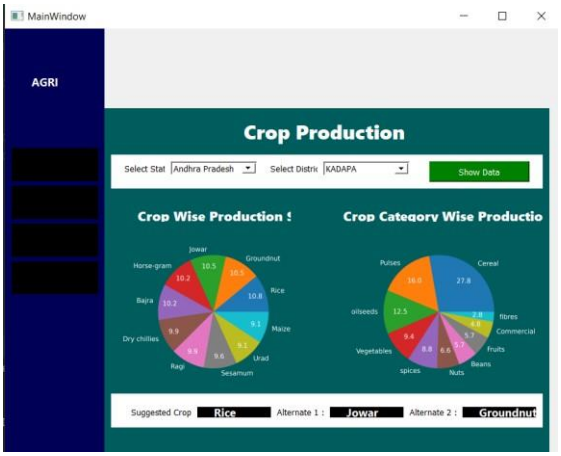

Fig 7: State Prediction

To discover, we will utilize seven unique fixings: nitrogen (N), phosphorus (P), potassium (K), heat, temperature, pH, and precipitation. Presently you want to get a decent gather to get a decent reap.

Fig 8: Crop Prediction

Random Forest were tried by KNN, posting a bunch of test information containing attributes like nitrogen (N), phosphorus (P), potassium (K), temperature, temperature, pH level, and wilderness name.

The outcomes acquired are addressed by a confusion matrix, which shows the connection between the calculation's speculation and the genuine worth when estimating the example subsequent to preparing the informational collection.

1.Random Forest:

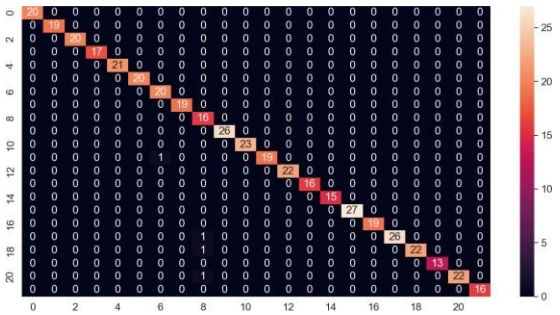

Fig 9: Confusion Matrix for Random Forest

Accuracy for Random Forest: 0.995
Precision for Random Forest: 0.996
Recall for Random Forest: 0.995
F1-score for Random Forest: 0.995

2.KNN:

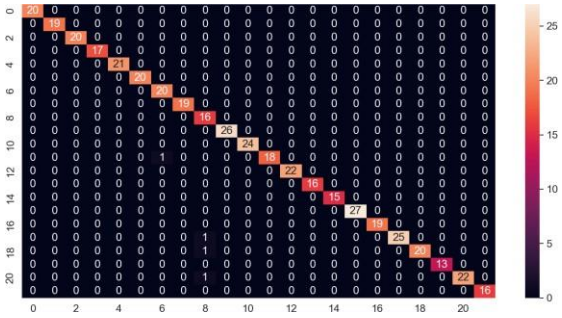

Fig 10: Confusion Matrix for KNN

Accuracy for KNN: 0.988
Precision for KNN: 0.996
Recall for KNN: 0.989
F1-score for KNN: 0.989

Algorithm Comparison:

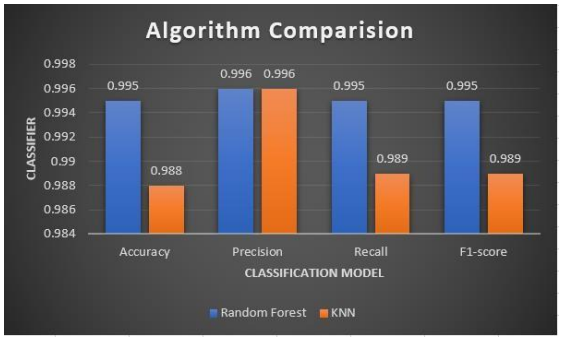

Fig 11: Algorithm Comparison

Contrasted with the past task, we got 9% more on KNN.

V. Conclusion

Carry out an arrangement to conjecture generally upheld rural creation. Plant creation is gotten ready for data handling strategies. Today, random forest is utilized to characterize a fascinating yield as a genuine harvest. In agribusiness, it is for the most part consistent with anticipate yields. The higher the yield, the higher the yield. The proposed procedures assist ranchers with being careful about various harvest necessities and costs. This assists ranchers with picking what harvests to plant. This work is frequently used to distinguish extra plants that can be gathered monetarily and proficiently. This innovation can grow a wide assortment of yields. Indian ranchers might enjoy the benefit of precisely foreseeing yields in various pieces of India.

VI. Future Scope

India may be a country wherever agriculture is extremely vital. The prosperity of the farmers ends up in the prosperity of the state. Thus, our work would assist farmers in sowing the acceptable seed supported soil necessities so as to extend productivity and exploit such a way. As a result, farmers will plant the acceptable crop, increasing their yield and therefore the nation's overall

productivity. Our future work can concentrate on associate degree improved knowledge set with an oversized variety of attributes, in addition as yield prediction.

REFERENCES

[1] Aruvansh Nigam, Saksham Garg, Archit Agrawal "Crop Yield Prediction using ML Algorithms ",2019
[2] LeoBrieman, "Random Forests", 2019
[3] Priya, P., Muthaiah, U., Balamurugan, M." Predicting Yield of the Crop Using Machine Learning Algorithm",2020
[4] Mishra, S., Mishra, D., Santra, G. H.,"Applications of machine learning techniques in agricultural crop production",2020
[5] Dr.Y Jeevan Kumar,"Supervised Learning Approach for Crop Production",2020
[6] RameshMedar,Vijay S, Shweta, "Crop Yield Prediction using Machine Learning Techniques", 2018
[7] Ranjini B Guruprasad, Kumar Saurav, Sukanya Randhawa,"Machine Learning Methodologies for Paddy Yield Estimation in India: A CASE STUDY", 2019
[8] https://en.wikipedia.org/wiki/Crop_yield
[9] https://en.wikipedia.org/wiki/Random_forest
[10] https://en.wikipedia.org/wiki/Knearest_neighbors_algorithm
[11] N. Suresh, Nvk Ramesh, Syed Inthiyaz, P. Poorna Priya, KurraNagasowmika,Kota V. N. Harish Kumar, Mashkoor Shaik, B. Naresh Kumar Reddy,"CropYieldPrediction Using Random Forest Algorithm",2021
[12] Mamunur Rashid, Bifta Sama Bari, YusriYusup, Mohamad Anuar, NuzhaKhan, "A Comprehensive Review of Crop Yield Prediction Using Machine Learning Approaches with Special Emphasis on Palm Oil Yield Prediction",2021
[13] DhivyaElavarasan, P. M. Durairaj Vincent**,** "Crop Yield Prediction Using Deep Reinforcement Learning Model for Sustainable Agrarian Applications",2020
[14] Roberto Luciani, Giovanni Laneve, MunzerJahjah**, "**Agriculture Monitoring, an Automatic Procedure for Crop Mapping and Yield Estimation: The Great Rift Valley of Kenya Case",2019
[15] Nida Rasheed**,** Shoab Ahmed Khan, Ali Hassan**,** Saria Safdar, "A Decision Support Framework for National Crop Production Planning",2021
[16] Li He, Craig A, Zhi-Jie Wang, Wei Feng, Tian-Cai Guo, "Reduced Prediction Saturation and View Effects for Estimating the Leaf Area Index of Winter Wheat",2019
[17] Alex F. Haufler, John H. Booske, Susan C. Hagness**,** "Microwave Sensing for Estimating Cranberry Crop Yield: A Pilot Study Using Simualted Canopies and Field Measurement Testbeds",2022
[18] Laura Martínez-Ferrer, Maria Piles, Gustau Camps-Valls, "Crop Yield Estimation and Interpretability with Gaussian Processes",2021
[19] Abhinav S**,** Arpit Jain, Prateek Gupta, Vinay C, "Machine Learning Applications for Precision Agriculture: A Comprehensive Review",2021

[20] Emily Myers, John Kerekes, Craig Daughtry, Andrew Russ, “Effects of Satellite Revisit Rate and Time-Series Smoothing Method on Throughout-Season Maize Yield Correlation Accuracy”,2021.

[21] Kanyadara Saakshara, Kandula Pranathi, R.M. Gomathi, A. Sivasangari, P. Ajitha, T. Anandhi, "Speaker Recognition System using Gaussian Mixture Model", 2020 International Conference on Communication and Signal Processing (ICCSP), pp.1041-1044, July 28 - 30, 2020.

[22] R. M. Gomathi, P. Ajitha, G. H. S. Krishna and I. H. Pranay, "Restaurant Recommendation System for User Preference and Services Based on Rating and Amenities," 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), 2019, pp. 1-6, doi: 10.1109/ICCIDS.2019.8862048.

[23] Subhashini R, Milani V, "IMPLEMENTING GEOGRAPHICAL INFORMATION SYSTEM TO PROVIDE EVIDENT SUPPORRT FOR CRIME ANALYSIS", Procedia Computer Science, 2015, 48(C), pp. 537–540

[24] Harish P, Subhashini R, Priya K, "Intruder detection by extracting semantic content from surveillance videos", Proceeding of the IEEE International Conference on Green Computing, Communication and Electrical Engineering, ICGCCEE 2014, 2014, 6922469

[25] Sivasangari, A., Krishna Reddy, B.J., Kiran, A., Ajitha, P.(2020), “ Diagnosis of liver disease using machine learning models”, ISMAC 2020, 2020, pp. 627–630, 9243375

[26] Sivasangari, A., Nivetha, S., Pavithra,, Ajitha, P., Gomathi, R.M. (2020),” Indian Traffic Sign Board Recognition and Driver Alert System Using CNN”, 4th International Conference on Computer, Communication and Signal Processing, ICCCSP 2020, 2020, 9315260

[27] Ajitha, P., Lavanya Chowdary, J., Joshika, K., Sivasangari, A., Gomathi, R.M., "Third Vision for Women Using Deep Learning Techniques", 4th International Conference on Computer, Communication and Signal Processing, ICCCSP 2020, 2020, 9315196

[28] Ajitha, P.Sivasangari, A.Gomathi, R.M.Indira, K."Prediction of customer plan using churn analysis for telecom industry",Recent Advances in Computer Science and Communications,Volume 13, Issue 5, 2020, Pages 926-929.

[29] Gowri, S. and Divya, G., 2015, February. Automation of garden tools monitored using mobile application. In International Confernce on Innovation Information in Computing Technologies (pp. 1-6). IEEE.

[30] Gowri, S., and J. Jabez. "Novel Methodology of Data Management in Ad Hoc Network Formulated Using Nanosensors for Detection of Industrial Pollutants." In International Conference on Computational Intelligence, Communications, and Business Analytics, pp. 206-216. Springer, Singapore, 2017.

Bibliography

Master Kanika Bhatnagar graduating from the beloved university Sathyabama Institute of Science and Technology in the stream of Computer Science and Engineering has excelled in this stream. Her research interest is Machine Learning, Neural Network and Cyber Security.

Master Mamilla Jaahnavi graduating from the beloved university Sathyabama Institute of Science and Technology in the stream of Computer Science and Engineering has excelled in this stream. Her research interest is Machine Learning, Neural Network and Web Development.

Dr.B.U. Anu Barathi, Assistant professor in the Department of Computer Science and Engineering at Sathyabama Institute of Science and Technology in Chennai. Her area of specialization is Database Management Systems, Data Mining, Machine Learning, and Data Analytics. She has authored in various reputed conferences and in international journals.