Prediction of crop yield using machine learning

Cite as: AIP Conference Proceedings **2418**, 020072 (2022); https://doi.org/10.1063/5.0081726
Published Online: 24 May 2022

Mohammed Ali Shaik, Geetha Manoharan, B. Prashanth, et al.

Prediction of Crop Yield using Machine Learning

Mohammed Ali Shaik[1, a)], Geetha Manoharan[1], B Prashanth[2], NuneAkhil[3], AnumandlaAkash[3] and Thudi Raja Shekhar Reddy[3]

[1]*SR University, Warangal, Telangana, India.*
[2]*Sumathi Reddy Institute of Technology for Women, Warangal, Telangana, India.*
[3]*SR Engineering College, Warangal, Telangana, India*

Corresponding author: [a)]niharali@gmail.com

Abstract. Farming is the major work which is considered as a culture instead like job and farming is the back bone of our economy as farming is the means which carried forth human advancement. India is nation which shows more interest towards farming and also grows all types of crops and its economy generally dependent on harvest profitability. Subsequently we can say that agriculture is major support for all business in our nation. Choosing of each harvest is significant in the choosing as each and every state in India grow various crop and the climate also varies from state to state. The choice of crop will depend on the various factors like, value of the crop, price given by the government, weather conditions and the price given by the private market buyers. Numerous progressions are needed in the field of agriculture to improve the benefit to Indian economy. We can improve agriculture by implementing AI mechanisms which can be same are defficiently on various cultivating areas. With all the advancements in the areas of machines and their improvements we can use them in cultivating the valuable and detailed data concerning various issues in addition to assuming the critical part in it. This paper helps use to getting an idea towards executing all the harvest based strategy with the ambitious techniques that helps in enchanting the maintenance of numerous agriculture and agriculture field issues. This helps the farmers to choose a best crop which helps them getting profit and also helps to increase our nation's economy.

INTRODUCTION

Agriculture established itself as backbone for a country's economy [1]. In a country like India there are over 118.9 million farmers and as the population is raising the food also increases and we need new mechanisms to produce more food products in a less time and agriculture is not a money earning job due to this not many people choose farming as a job [2]. By using advanced mechanisms, we can produce and even help the farmers to earn a good price for their production [3]. For centuries, agriculture is considered as a crucial and superior culture which is practiced in India for centuries [4]. In olden days people use land where they live and choose crop to yield depend on the weather and circumstances in their area, as the greenhouse effect and the change in the climatic conditions the farmers can't expect the conditions when it is going to rain or snow or the water will be available for the land or not etc.,

The Agriculture work carried by the farmers in their daily routine get some benefits from private business companies and also from government help life [5]. Due to the introduction of new advancements and strategies in the field of agriculture is slowly undermining due to these activities, many people are being working around and making up fake things that are hybrid things where they prompt dreadful life. present day's people don't care about all the advantages of the harvests. Because of unwanted methodologies and sudden change in climatic conditions and fundamental assets like soil or water and air that leads to deficiency of food by considering various issues constantly like atmospheric temperature increases and a couple of parts, there are no proper game plan or development to face the situation. In a country like India there are a couple of ways to deal with assemble the reasonable advancement in the agriculture field. There are different ways to deal with these problems by using some of the great innovations in the technology. One of the best and easy technology we can use is AI based Prediction

International Conference on Research in Sciences, Engineering & Technology
AIP Conf. Proc. 2418, 020072-1–020072-8; https://doi.org/10.1063/5.0081726
Published by AIP Publishing. 978-0-7354-4368-6/$30.00

concepts. Data burrowing is like a manner which is important for the prediction and the collection of yield of the crop. Data mining helps the person in researching data from substitute look through and summarize it into specific information and help for the best results.

Data mining is a process which allows the person to break down the data he\she collected depending on various factors and sum up the final data required by the person [6]. Data mining is actually created towards finding connections and examples among the various fields from a huge database [7]. The Data collected by the client or a person can be converted into information about future patterns [8]. For instance, outline data about harvest creation can help the researchers to make the better and safe future [9]. Harvest yield forecast is a significant farming issue [10]. Every single rancher is consistently Interested towards knowing how much yield can he get and contribution towards particular aspect [11]. In the past, yield expectation was determined by the examining rancher's past productions on a specific harvest [12]. The Farming yield is principally relies upon climate conditions, irritations and arranging of collect activity [13]. From the Precise data about history of harvest yield is something critical for settling on choices identified with agriculture danger management. the paper proposes a plan to anticipate the yield of the harvest as the farmer can cross check the produce of harvest according to the section of land, prior to developing onto the field.

LITERATURE SURVEY

Machine Learning manages issues where the connection among information and yield factors isn't clear or difficult to understand [14, 8]. The term "learning" means the programmed procurement of primary depictions from occurrences of what is being described [15]. Unlike the customary measurable techniques, Machine Learning doesn't make suppositions regarding the right organization of information model as it clearly depicts the information whose trademark is helpful to perform modelling the complex and non-straight practices, for example, a capacity for crop yield expectation [16, 19]. Most of the Machine Learning techniques are effectively applied to Crop Yield Prediction. with the help of supervised learning [17, 20], we use this arrangement of the available factors [18], we will create a capacity which will guide the contributions towards the wanted yield [19]. The preparation measurements will proceed until the model accomplished an ideal degree precision on the preparation information. In the development of this project we used different methods which are regression [20, 33], Decision tree [21, 34], Random Forest [22, 35], KNN [23, 36], logistic Regression [24, 37] and so on. There are many similar projects in the market what makes our project different from the existing is that we are integrating python with machine learning so the we are reduce the lines of code and production cost too and we can get accurate result with this project.

METHODOLOGY

Random Forest Classifier

Random forest classifier [25] is a method which is a well-known methodology with the ML calculation fit for performing both the arrangements, relapse undertaking, then it works by building large number of choice trees with the given data at preparing time which yields the class or the method of classes that groups with the mean expectation value which is called as the relapse of a singular tree [26]. The power of the forecast depends on the number of trees in the forest [27]. Random forest right for choice trees propensity for over fitting to their preparation set. The informational collections are considered as precipitation, discernment, creation, temperature that generates random forest, an assortment of choice trees by allowing two thirds of the records in the dataset [28]. The excess records are filled with the choice trees for exact arrangement. The resultant preparation set is applied on the test information for the right forecast of the harvest yield dependent on the information credited. Random Forest calculation was used to consider for the execution of the methodology on the dataset. There are various benefits of random forest calculations in random forest that overfits and is less of an issue and not at all like a choice trees of AI/ML calculations as in Random forest there is no need of pruning. Random Forest AI/ML calculations can be filled in equal.

The Random forest algorithm can run efficiently on large databases and also it has the maximum classification accuracy. The three parameters that are available in the random forest algorithm are [29]:

- ntree- in this the name suggests that the parameter n defines we can grow ‘n’ trees. The larger the tree the more it costs to build models [30].

- mtry–in this it refers the number of variables to be selected at a specific node that splitsbased on the default value attained for performing regression that is $^{p}/_{3}$ and the sqrt(p) is utilized for performing classification which always try to utilize smaller values for 'mtry' that avoids the process of overfitting [31].
- nodesize– denotes the number of observations in terminals that we require along with the terminal nodes that comprises of distinct parameters which are clearly related with the depth of a tree where the lower is the tree depth considered to be the higher the number is and which may also consider that the tree might even fail to consider the signals generated from the data if the tree that comprises of minimal depth [32].

Tools Used

In order to predict the data, we have used the Different data sets and we have compared the data using Train Split Concepts, Random Forest Algorithm and Linear Regression for desired output[46-48]. We made different kinds of data comparisons and different data visualizations to make standard output. We made model (Random Forest Algorithm and Linear Regression) comparisons to make best suitable model in order to develop the output. In total process we came with Random forest has high r2score when compare to other model. So, we have used these models and Split concepts for the development of the project.

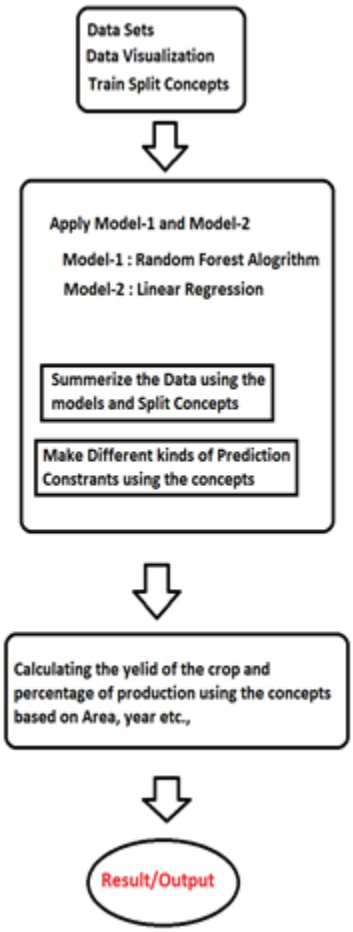


FIGURE 1. This Flow chart represent the complete detail process and execution

Dataset Used

All the data used in this project are gathered from the straightforwardly open records of the Indian Government and some third-party websites. The data we collected consists of records from the year 1997 to 2014 for distinct

seasons like Kharif and Rabi with respect to rice creation process. From the really big beginning data, simply a predetermined number of huge factors which have the most raised impact on cultivating yield were picked for the current assessment and the limits decided for the current project recorded under[43-45].

- Precipitation (mm): The total amount of drizzle for the yield or crop
- Kharif is the Harvesttime, Entire year and Rabi time of each season of every zone.
- perception: Insight data for every year in each picked region of Andhra Pradesh was considered for careful yield.
- crop prediction (in Tons): The yield created zone in Hectares and creation in tons for Kharif and Rabi seasons for every year in each picked region of Andhra Pradesh state was considered for the present research.

Decision Tree

The Decision tree classifiers utilizes various methodologies thus a characteristic pick from the starting step which can't be utilized any longer which can give a better arrangement whenever used in later advances. Additionally, itisoverfit for the preparation information which can give us helpless outcomes from inconspicuous information. Thus, to beat this impediment group model is used and the troupe model outcomes from various models are joined. The outcome acquired from the outfit model is typically in a way that is better than the outcome for any of the individual models.

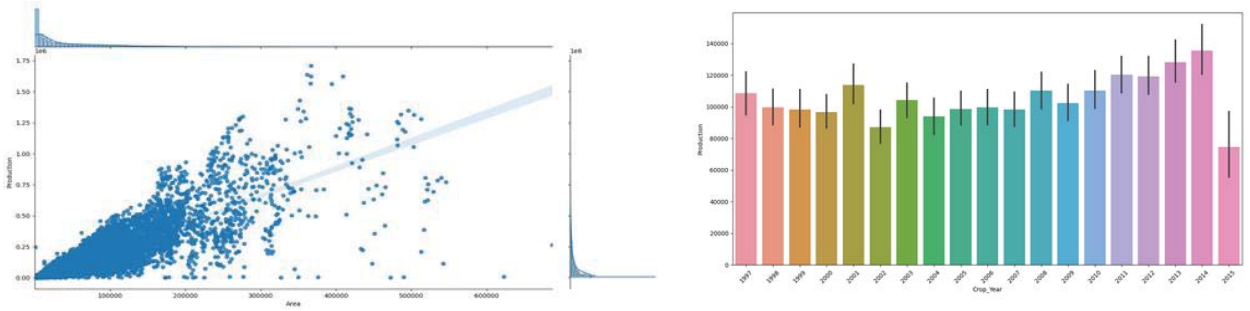
Arbitrary Woodlands is a group classifier which utilizes numerous decision trees models to anticipate the outcome. An alternate subset of preparing information is chosen, with substitution to prepare each tree. An assortment tree is a woodland, and the trees are being prepared on subsets which are being chosen aimlessly, consequently irregular timberlands which can be used towards the order and relapse issues in a class that comprises of tasks that are made by the quantity of votes from all the trees and for relapse the normal of the outcomes is utilized.

The Straight relapse endeavors tends to establish the connectivity that is merely attainable in between distinct factors that are fittedwith a specific or predefined condition which is also the noticed information over distinct variables that are considered as one of the logical variables and is further viewed as the needy variable as it directly relates to the distinct loads of people who tend to utilize the straight relapse model [12].

Procedure

As per the algorithm, it converts the gathered informational collections into csv document configuration &burden the informational collections. Split the stacked informational collections into two sets, for example, preparing information & test the information in the split proportion of either 67 rates or 33 rates that it is 0.67 or 0.33. To Isolate the collected preparation information by classifying esteems with the goal that the quality guide to a reasonable quality and put away in fitting rundown. At that point ascertain Mean and the Standard Deviation for required tuples and afterward sum up collected informational indexes. Analyze the summed-up information list and the first informational collections ascertain the likelihood. In light of the outcome the biggest likelihood created is taken for forecast[38-42]. The exactness can be anticipated by contrasting the subsequent class esteem and the test informational collection. The precision can go from0% to 100%.

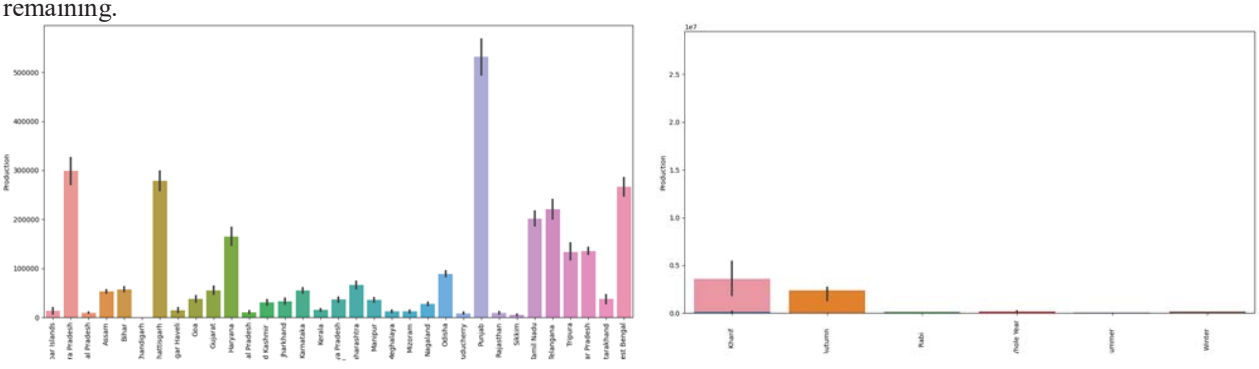
EXPERIMENTAL RESULT



(a) (b)

FIGURE 2.(a) Production of Crop (b) Crop_Yearvs Production

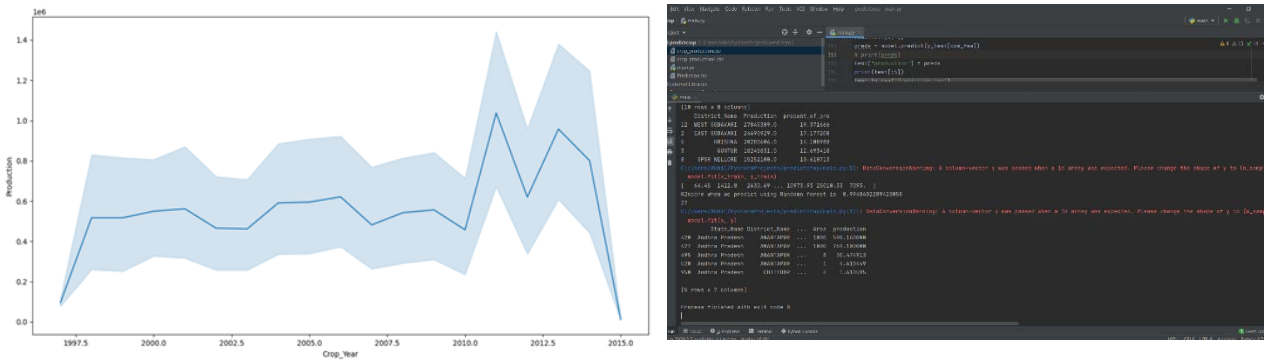
Figure 2 (a) denotes the year wise production attained by taken dataset and (b) denotes the Crop_Year versus production for example the in the year 2000 the production was in between 80000 to 100000 similarly to the remaining.



(a) (b)

FIGURE 3. (a) State vs Production (b) Season vs Production

Figure 3(a) denotes the state wise production. We can see the different level of production in particular state and (b) denotes the production in different seasons as the production will be different in different seasons we can see the production is high when we compare kharif seasons production with autumn season because of the demand of the crop and the price for the particular crop.



(a) (b)

FIGURE 4. (a) Crop_Yearvs Production (b) Results attained or output attained

Figure 4(a) denotesthe rise and fall of the production in difference years. By this we can see the status of the production, the growth and profit of the production in year wise and (b) denotes the overall output of the project. We can see the production percentage of the crop over the years and the usage of the area for the cultivation and all other factors/parameters.

CONCLUSION

By using the random forest method algorithm by which we can attain a precise crop yield that predicts from the result we can understand we can get a better prediction using random forest method by using the Random forest algorithm we can achieve maximum amount of crop yield that models with respect to the least model as the random forest method is suitable for performing huge crop yield prediction process in the phase of agricultural planning. Through this we can help the farmer to take right decision for crop such that the farmer can get more money and also develop the agricultural sector.

The paper describes the crop yield prediction using the random forest algorithm and in future we want to identify the efficiency of algorithm that is based on their accurateness of metrics that tends to help in choosing the efficiency of algorithm for performing the crop yield prediction. In further we will develop the project more efficiently and make the product into the market with best outputs.

REFERENCES

1. Awan, A. M., & Sap, M. N. M. (2006), An intelligent system based on kernel methods for crop yield prediction. In Pacific-Asia Conference on Knowledge Discovery and Data Mining (pp. 841-846). Springer, Berlin, Heidelberg
2. Bhosale, S. V., Thombare, R. A., Dhemey, P. G., &Chaudhari, A. N. (2018), Crop Yield Prediction Using Data Analytics and Hybrid Approach. In 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA) (pp. 1-5). IEEE
3. Gandhi, N., Armstrong, L. J., &Petkar, O. (2016), Proposed decision support system (DSS) for Indian rice crop yield prediction. In 2016 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR) (pp. 13-18). IEEE.
4. Mohammed Ali Shaik and DhanrajVerma, (2020), Deep learning time series to forecast COVID-19 active cases in INDIA: a comparative study, 2020 IOP Conf. Ser.: Mater. Sci. Eng. 981 022041 https://doi.org/10.1088/1757-899X/981/2/022041
5. Ahamed, A. M. S., Mahmood, N. T., Hossain, N., Kabir, M. T., Das, K., Rahman, F., &Rahman, R. M. (2015). Applying data mining techniques to predict annual yield of major crops and recommend planting different crops in different districts in Bangladesh. In 2015 IEEE/ACIS 16th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD) (pp. 1-6). IEEE.
6. Suresh, A., Kumar, P. G., &Ramalatha, M. (2018). Prediction of major crop yields of Tamilnadu using K-means and Modified KNN. In 2018 3rd International Conference on Communication and Electronics Syst ems (ICCES) (pp. 88-93). IEEE.
7. Mohammed Ali Shaik and DhanrajVerma, (2020), Enhanced ANN training model to smooth and time series forecast, 2020 IOP Conf. Ser.: Mater. Sci. Eng. 981 022038 https://doi.org/10.1088/1757-899X/981/2/022038
8. Veenadhari, S., Misra, B., & Singh, C. D. (2014). Machine learning approach for forecasting crop yield based on climatic parameters. In 2014 International Conference on Computer Communication and Informatics (pp. 1-5).
9. Mohammed Ali Shaik, DhanrajVerma, P Praveen, K Ranganath and BonthalaPrabhanjanYadav, (2020), RNN based prediction of spatiotemporal data mining, 2020 IOP Conf. Ser.: Mater. Sci. Eng. 981 022027 https://doi.org/10.1088/1757-899X/981/2/022027
10. Mohammed Ali Shaik, “Time Series Forecasting using Vector quantization”, International Journal of Advanced Science and Technology (IJAST), ISSN: 2005-4238, Volume-29, Issue-4 (2020), Pp. 169-175.
11. Mohammed Ali Shaik, (2019), “A Survey on Text Classification methods through Machine Learning Methods”, International Journal of Control and Automation (IJCA), ISSN: 2005-4297, Volume-12, Issue-6 (2019), Pp. 390-396.

12. T. SampathKumar,B. Manjula, D. Srinivas,(2017), A New Technique to Secure Data Over Cloud Jour of Adv Research in Dynamical & Control Systems vol 11 pp 145-149
13. T. Sampath Kumar, B. Manjula, Mohammed Ali Shaik, Dr. P. Praveen, (2019), A Comprehensive Study on Single Sign on Technique International Journal of Advanced Science and Technology (IJAST) Vol-127 pp 156-162
14. Mohammed Ali Shaik, T. Sampath Kumar, P. Praveen, R. Vijayaprakash, (2019), "Research on Multi-Agent Experiment in Clustering", International Journal of Recent Technology and Engineering (IJRTE), ISSN: 2277-3878, Volume-8, Issue-1S4, June 2019. Pp. 1126-1129.
15. R. Ravi Kumar, M. Babu Reddy and P. Praveen,(2017), A review of feature subset selection on unsupervised learning Third International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB), pp163-167
16. P.Praveen, B.Rama,(2019), An Efficient Smart Search Using R Tree on Spatial Data Journal of Advanced Research in Dynamical and Control Systems vol 4 pp1943-1949.
17. P. Praveen, B. Rama and T. Sampath Kumar,(2017), An efficient clustering algorithm of minimum Spanning Tree Third International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB pp 131-135
18. Naresh Kumar S, Pramod Kumar P, Sandeep C, Thirupathi V, Shwetha S,(2019), A study on deep Q-learning and single stream Q-network architecture Int J AdvSciTechnolvol 28(20) pp 586-592
19. Kumar RR, Reddy MB, Praveen P. (2019), Text classification performance analysis on machine learning Int J AdvSciTechnolvol 28(20) pp 691-697
20. Ravi Kumar R, Babu Reddy M, Praveen P,(2019), An evaluation of feature selection algorithms in machine learning Int J SciTechnol Res vol 8(12) pp 2071-2074
21. Jamalpur, Bhavana,(2020), Data Exploration As A Process Of Knowledge Finding And The Role Of Mining Data Towards Information Security Journal Of Mechanics Of Continua And Mathematical Sciences vol 15(6) pp 2020-2028
22. JamalpurBhavana, KomuravellySudheer Kumar,(2019), Implementation of Bovw Model Towards Obtaining Discriminative Features of the Images International Journal of Advanced Science and Technology vol 28(17) pp 205-213
23. J.Bhavana, KomuravellySudheer Kumar,(2019), A Study on the Enhanced Approach of Data Mining Towards Providing Security for Cloud Computing Indian Journal of Public Health Research & Development vol 9(11) pp 225-232
24. Kothandaraman D, Sheshikala M, SeenaNaik K, Chanti Y, Vijaykumar B. (2019), Design of an optimized multicast routing algorithm for internet of things Int J Recent TechnolEngvol 8(2) pp 4048-4053
25. Dr.SeenaNaik,(2016), An Effective use of Data Mining Techniques to Creation International Journal Of Advancement In Engineering vol 3(10) pp 157-163
26. D.Ramesh, Syed Nawaz Pasha, G.Roopa,(2017), A Comparative Analysis of Classification Algorithms on Weather Dataset Using Data Mining Tool Oriental Journal of Computer Science and Technology vol 10(4) pp1-5
27. Chythanya KR, Kumar KS, Rajesh M, Tharun Reddy S,(2020), Sensor Cloud: A Breakdown information on the Utilization of Wireless Sensor Network by Means of Cloud Computing Test Eng Manage vol 82 pp13945-13954
28. P. Praveen and B. Rama,(2016), An empirical comparison of Clustering using hierarchical methods and K-means International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB) pp.445-449
29. SallauddinMohmmad, G. Sunil, RanganathKanakam,(2017), A Survey On New Approaches Of Internet Of Things Data Mining International Journal of Advanced Research in Computer Science vol 8(8) pp 123-135
30. SallauddinMohmmad, Dr.M.Sheshikala, Shabana,(2018), Software Defined Security (SDSec): Reliable centralized security system to decentralized applications in SDN and their challenges Jour of Adv Research in Dynamical & Control Systems Vol 10(10) pp 147-152
31. Mohammed Ali Shaik and DhanrajVerma 2020 Enhanced ANN training model to smooth and time series forecast IOP ConfSer: Mater SciEngVol (981) 022038 https://doiorg/101088/1757-899X/981/2/022038
32. SallauddinMohmmad, Shabana, RanganathKanakam,(2017), Provisioning Elasticity On IoT's Data In Shared-Nothing Nodes International Journal of Pure and Applied Mathematics vol 117(7) pp 165-173.

33. Pramod Kumar P, Naresh Kumar S., Thirupathi V. and Sandeep C,(2019), QOS and security problems in 4G networks and QOS mechanisms offered by 4G. International Journal of Advanced Science and Technology 28(20) 600-606.
34. Mohammed Ali Shaik and DhanrajVerma 2020 Deep learning time series to forecast COVID-19 active cases in INDIA: a comparative study IOP ConfSer: Mater SciEngVol (981) 022041 https://doiorg/101088/1757-899X/981/2/022041
35. Pramod Kumar P and Sagar K,(2019), Vertical Handover Decision Algorithm Based On Several Specifications in Heterogeneous Wireless Networks International Journal of Innovative Technology and Exploring Engineering, 8(9) 972-82.
36. YerrollaChanti, BandiBhaskar, NagendarYamsani, "Li-Fi Technology Utilized In Leveraged To Power In Aviation System Entertainment Through Wireless Communication", J. Mech. Cont.& Math. Sci., Vol.-15, No.-6, June (2020) pp 405-412.
37. Chanti, Y., Korra, S.N., Bhaskar, B., Harshavardhan, A., Srinivas, V., 2020. "Sturdy goals coverage for power harvesting Wi-Fi detector coterie". IOP Conf. Ser.: Mater. Sci. Eng. 981, 022070.
38. Yadav, B.P., Prasad, C.S.S., Padmaja, C., Korra, S.N. and Sudarshan, E., 2020, December. A Coherent and Privacy-Protecting Biometric Authentication Strategy in Cloud Computing. In IOP Conference Series: Materials Science and Engineering (Vol. 981, No. 2, p. 022043). IOP Publishing.
39. Yadav, B.P., Sheshikala, M., Swathi, N., Chythanya, K.R. and Sudarshan, E., 2020, December. Women Wellbeing Assessment in Indian Metropolises Using Machine Learning models. In IOP Conference Series: Materials Science and Engineering (Vol. 981, No. 2, p. 022042). IOP Publishing.
40. Shabana, Mohmmad, S., Shaik, M.A., Mahender, K., Kanakam, R., Yadav, B.P.Average Response Time (ART):Real-Time Traffic Management in VFC Enabled Smart Cities(2020) IOP Conference Series: Materials Science and Engineering, 981 (2), art. no. 022054.
41. Sravanthi, T., Hema, V., Tharun Reddy, S., Mahender, K., Venkateshwarlu, S.Detection of Mentally Distressed Social Media Profiles Using Machine Learning Techniques(2020) IOP Conference Series: Materials Science and Engineering, 981 (2), art. no. 022056.
42. Akarapu, M., Martha, S., Donthamala, K.R., Prashanth, B., Sunil, G., Mahender, K.Checking for Identity-Based Remote Data Integrity Cloud Storage with Perfect Data Privacy(2020) IOP Conference Series: Materials Science and Engineering, 981 (2), art. no. 022034.
43. Prasad, C.H.S.S., Yadav, B.P., Mohmmad, S., Gopal, M., Mahender, K.Study of threats associated with cloud infrastructure systems(2020) IOP Conference Series: Materials Science and Engineering, 981 (2), art. no. 022055.
44. 20. Sudarshan, E., Korra, S.N., Rajasekharaiah, K.P., Venkatesulu, S. and Harshavardhan, A., 2020, December. IoT Based Smart Solar Atmospheric Water Harvesting System. In IOP Conference Series: Materials Science and Engineering (Vol. 981, No. 4, p. 042004). IOP Publishing.
45. 21. Venkatesulu, S., Sudarshan, E., Korra, S.N., Kumari, D.R., Yadav, B.P. and Mahender, K., 2020, December. Real Time Fitness Analysis of Bitumen Road and Vehicle through Their Acoustic Signals. In IOP Conference Series: Materials Science and Engineering (Vol. 981, No. 3, p. 032004). IOP Publishing.
46. 22. Korra, S.N., Sudarshan, E., Venkatesulu, S., Kumar, P.P. and Yadav, B.P., 2020, December. Portable manure dispenser machine. In IOP Conference Series: Materials Science and Engineering (Vol. 981, No. 2, p. 022026). IOP Publishing.
47. 23. Mannanuddin, K., Aluvala, S., Sneha, Y., Kumaraswamy, E., Sudarshan, E. and Mahender, K., 2020, December. Confluence of Machine Learning with Edge Computing for IoT Accession. In IOP Conference Series: Materials Science and Engineering (Vol. 981, No. 4, p. 042003). IOP Publishing.
48. 24. Yadav, B.P., Ghate, S., Harshavardhan, A., Jhansi, G., Kumar, K.S. and Sudarshan, E., 2020, December. Text categorization Performance examination Using Machine Learning Algorithms. In IOP Conference Series: Materials Science and Engineering (Vol. 981, No. 2, p. 022044). IOP Publishing.