A STRATEGIC ANALYSIS OF STORAGE FOR CROP BASED CULTIVATION ON CLIMATIC PARAMETERS USING DECISION TREE COMPARED WITH K-NEAREST NEIGHBOR (KNN) ALGORITHM

Yaragadla Sivaram[1], Aishwarya B[2*]

Article History: Received: 12.12.2022 **Revised:** 29.01.2023 **Accepted:** 15.03.2023

Abstract

Aim: In the study, K-Nearest Neighbor Algorithm (KNN) is to predict climatic parameters of the crop and its performance is tested by comparing it with the k-nearest neighbor algorithm for better accuracy.
Materials and Methods: Support vector machine learning algorithm with sample size n=10 and linear regression algorithm with sample size n= 10 with G-power analysis value of 80%. The crop based cultivation Novel prediction helps to improve the predicted accuracy. The sample size is estimated using G power to be 3101 records in each group with 80% of power and a 0.05 Error rate.
Results: Average accuracy of 92% for linear regression algorithm and 82% for svm to predict customers' sales. Significant difference between Support vector machine and linear regression $p<0.03$ and with a significance value 0.02.
Conclusion: Within the limits of study we found that predicting the novel crop based cultivation yields using natural parameters by using linear regression is better than predicting the climatic parameters by using K-nearest neighbor (KNN).

Keywords: Novel Crop, Climatic Parameters, Machine Learning, Decision Tree, Strategic Analysis, K-Nearest Neighbor.

[1]Research Scholar, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India, Pincode: 602105.
[2*]Project Guide, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India, Pincode: 602105.

1. Introduction

Crop yields and harvest yield conjectures straightforwardly influence the yearly public and worldwide economy and assume a significant part in the food economy (Ma and Ma 2021). Novel crop yields are profoundly subject to water system and environment information. More water systems don't really increase yield, and in this way, improvement of Water Support Vector Machine systems and more proficient water system frameworks are basic ((Ma and Ma 2021; Juma and Beru 2021). Anticipating yield dependent on various sorts of water systems is one method for Point of Sale upgrading the cycle. There are many challenges facing agriculture, such as climate change, environmental impacts, and labor, resources, and land shortages (Lumsden and Schulze, n.d.). The increasing demands on agricultural society to produce more with the same amount of land are also noteworthy (Lumsden and Schulze, n.d.; Utset, Eitzinger, and Alexandrov, n.d.) 2013). The applications involved in the forecasting methodologies of the climatic parameters developed and evaluated by the researchers all over India and the world who will help in the field of cultivation and novel results of agriculture.

Over the past 7 years, nearly 248 articles published in both google scholar and 720 articles in IEEE Xplore and 560 articles in the ScienceDirect related to crop based cultivation on climatic parameters using machine learning (Ahuja and Ma 2020). And the strategical analysis Forecasting of the climatic parameters will affect the growth of the novel crop or the plant in the field (Tomar et al. 2021) support vector machine novel prediction a smart farm is driven by data and information generated by agricultural technology and brings the farmer closer to digital technology and the climatic parameters to get novel results (Lumsden and Schulze, n.d.; Utset, Eitzinger, and Alexandrov, n.d.; Y. Zhang et al. 2015; Wheeler et al., n.d.). This includes the use of sensors and drones and the collection of strategic analysis accurate data such as weather data, soil maps, and other species (Lumsden and Schulze, n.d.; Utset, Eitzinger, and Alexandrov, n.d.; Y. Zhang et al. 2015; Wheeler et al., n.d.; Huntingford et al. 2005). The ability to gain knowledge from these machine learning data and create decision support systems is crucial in order to maximize farms and boost their value, while meeting the food requirements of the population and ensuring the strategic analysis sustainable use of natural resources (Lotze-Campen 2011a). The usage of the natural resources are accurate and easy for the crop based growth.

To help the researchers, increasing the amount of irrigation was not considered as an input to the model (Salisbury and Barbetti 2011). Yield is highly dependent on the amount of irrigation; one change in the amount of water can change a lot of things (Jaradat 2011). Considering irrigation of the novel crop yield scheduling as an input to the model would help to produce better results (Jaradat 2011; Hatfield 2011; Lotze-Campen 2011b). When irrigation scheduling is incorporated into the model, it will help to create an intelligent irrigation schedule that reduces water consumption without Novel results Prediction reducing productivity (Lotze-Campen 2011b). In order to optimize irrigation for optimal productivity, irrigation amounts should be calculated (Lotze-Campen 2011b; Redden 2015).

The main aim of the papers is to increase the novel crop yield of the particular and Agriculture provides nutrients, clothing, shelter, medicine, and recreation to meet the needs of humans and to help shape civilizations (Lotze-Campen 2011b; Redden 2015; Tyack and Dempewolf 2015). Therefore, agriculture is the world's most important Support Vector Machine and machine learning enterprise. It is a productive whole where the free gifts of nature namely food, clothing, shelters, and medicine, are utilized. The gap of this paper is that the primary unit of production problems for humans is agriculture. Secondary units of production are animals and climatic parameters, including livestock, birds, and insects, which consume the primary production and provide concentrated, biodegradable products such as meat, milk, wool, eggs, and honey (Jalota et al. 2018).Our team has extensive knowledge and research experience that has translated into high quality publications(K. Mohan et al. 2022; Vivek et al. 2022; Sathish et al. 2022; Kotteeswaran et al. 2022; Yaashikaa, Keerthana Devi, and Senthil Kumar 2022; Yaashikaa, Senthil Kumar, and Karishma 2022; Saravanan et al. 2022; Jayabal et al. 2022; Krishnan et al. 2022; Jayakodi et al. 2022; H. Mohan et al. 2022)(K. Mohan et al. 2022; Vivek et al. 2022; Sathish et al. 2022; Kotteeswaran et al. 2022; Yaashikaa, Keerthana Devi, and Senthil Kumar 2022; Yaashikaa, Senthil Kumar, and Karishma 2022; Saravanan et al. 2022; Jayabal et al. 2022; Krishnan et al. 2022; Jayakodi et al. 2022; H. Mohan et al. 2022)

The system provides significant value by excluding the need for high-resolution remote control-perception data and by enabling the farmer to take account of adverse climate influences on harvest oscillation and the novel results (Jalota et al. 2018;

Sivakumar and Hansen 2007; Reynolds 2010; Araus and Slafer 2011; Sharannya et al. 2021). In our studies, we forecast the soybean and maize takings for various time periods (Jalota et al. 2018; Sivakumar and Hansen 2007; Reynolds 2010; Araus and Slafer 2011). Researchers found the error metric unit for soybean and maize yield forecasts to be comparable to schemes that provide yield forecast Point of Sale information during the first weeks to calendar months of the crop cycle (Jalota et al. 2018; Sivakumar and Hansen 2007; Reynolds 2010).

2. Materials and Methods

This research study was carried out in the Department of Artificial Intelligence Laboratory belonging to Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (SIMATS), Chennai. This study consists of two sample groups i.e Group 1 k-nearest neighbor and group 2 is .The required 150 samples for this analysis are done using G-power calculation out of the 150 samples. Minimum power of the analysis is fixed as 0.8 and the maximum accepted error is fixed as 0.5 and also significant value is >0.03. The dataset is collected from customer based products through Kaggle, an open source data repository(Marco 2000). These crops will be increasing the fertility of the soil, so that it will have better accuracy. The dependent variables in the given dataset are temperature, humidity, rainfall and pH and the independent variables are type of crop, pesticides and insecticides for the crop (Hooper, Amelon, and Lin 2022). This dataset was collected from cluster, machine learning and not a number values. This dataset is about classification of work with 1000 records and comments from different study sites. This helps in future decision making purposes. Finally, import the dataset into the colab to identify the products and customers apply the coding to perform operations. This data set consists of 3101 samples of the different DT algorithm crop based cultivation samples from different places on the particular districts, temperature, pH, Humidity, including Rainfall falled on the particular districts and types of crops etc, This dataset was collected from cluster, machine learning and different samples (Murakami et al. 2021). This dataset is about classification of work with 3101 records and comments from different study sites. This helps in future decision making purposes. Finally, import the dataset into the colab and spyder to identify the novel crop based cultivation growth and yield based on the climatic parameters and the novel results (Peterson and Phillis 1995).

K-Nearest Neighbor

The k-Nearest neighbors (KNN) algorithm is a simple and interesting machine learning prediction algorithm. Which is used for the prediction of complex problems. And it stores all available data and adds the new data to show similarity using the data. It can't handle the large dataset. It also has Initiative approach problems.

Pseudocode for K-Nearest Neighbor

1. Let knn be the new variable
2. Make use of the imported package
3. KNeighboursClassifier()
4. Let knn.fit be the another variable
5. Using knn.fit train the dataset
6. Like(x_train, y_train)
7. Knn_prediction help to find the result
8. Knn_predictand test the dataset
9. Calculate the formula for the result_4
10. Display result_4

Decision tree

In statistics, the decision tree is a prediction method which is used in supervised machine learning. It is used for making decisions for the multi condition problem such as the restaurant. It will give a multi decision for a single problem. And provides the correct solution for the complex scenarios.

Pseudocode for Decision tree

1. Let dt be the new variable
2. Using dt.fit train the dataset
3. Assume that dt_a=0.8
4. Find the accuracy = accuracy_score(y_test, y_pred)
5. Make use of the result_2 and display result_2

Statistical Analysis

The testing set up for the proposed system to implement with the following system is configuration of hardware Desktop with 64-bit, OS, RAM and Software Window 10, colab SPSS tool used for statistical analysis. In SPSS the dataset is prepared using a sample size of 10 data is analyzed with k-nearest neighbors and decision tree, the statistical analysis is done on the two groups using the train set and test set. A Comparison of means table for k-nearest neighbor and decision tree is shown below. The system architecture of the Point of Sale experimental setup First we import and load data. The classify the data into test and train after the classification applies the algorithm and generates the novel results model. After that apply the generated train model to the test. The dependent variables in the given dataset are no.of crops, name of the crops, growth conditions, pesticides, temperature, humidity,pH value and other parameters.The dependent variables in the given

dataset are temperature, humidity, rainfall and pH and the independent variables are type of crop, pesticides and insecticides for the novel crop.

3. Results

The group statistical analysis on the two groups shows that k nearest Novel Prediction neighbor has more accuracy than the other and its standard error mean is slightly different than the decision tree. In the independent sample test, the significance of both algorithms when the equal variance is 0.311. In Figure 1, the bar chart of accuracies with standard deviation error is plotted for both the algorithms. The k-nearest neighbor algorithm produces an accuracy of 85.82% and the decision tree algorithm has scored 80.06%. Table 2 represents the comparison of decision tree and k nearest neighbor algorithm in terms of mean accuracy and the Accuracy, Precision and Recall are the methods used for measuring the overall performance of data mining. Table 3,Other processes constructed on positive and negative reviews by strategic analysis of calculating accuracy, precision and recall. The Independent samples test categorizes test for equality of variance and T-test for equality of means for mean difference, standard error difference. The confidence interval and level of significance is set to 0.02 ($p<0.03$). The variables constructed on positive and negative reviews calculating accuracy, precision and recall. Group statistics for Mean, Std. Deviation, Std. Error Mean with sample size of 10.

4. Discussion

In this research work the k-nearest neighbor gets higher accuracy though the comparisons with plotting and stabilizing the data in Table 3 and it represents the statistical values obtained by the proposed system k-nearest neighbor have high accuracy value with 85.82% and significance value less with 0.05, when compared decision tree with accuracy value 80.06%.In (Murakami et al. 2021)The Novel predictions of crop based cultivation are done based on knowing the growth condition of crop based cultivation and growth elements. The k-nearest neighbor algorithm is showing high accuracy 85.8% compared to decision tree (80%).This study works on decision tree classifiers for data mining proved with an accuracy of 65% and proposed that decision table is a minimal and most effective algorithm (Murakami et al. 2021; Q. Zhang et al. 2022). The limitations of decision tree algorithms are long training time for large datasets difficult to understand and interpret the final model, variable weights and individual impact. Although it is not possible to predict the cold store for the crop storage, with the help of a k-nearest neighbor algorithm applied to a clean dataset we can find the optimization (Murakami et al. 2021; Q. Zhang et al. 2022; Widener et al. 2021; Kuang et al. 2021). In the future novel results, there is an idea to increase training and testing dataset and to find a variety of accuracy and can deploy as web content for the frameworks (Murakami et al. 2021; Q. Zhang et al. 2022; Widener et al. 2021). The future Scope, there is an idea to increase strategic analysis training and testing dataset and to find a variety of accuracy and can deploy as web content for the frameworks.Although the results of the study showed a better Predictive model performance can be calculated by the algorithms, there are certain limitations. By increasing the number of attributes in the input dataset the performance will be better and accurate. With advanced preprocessing techniques, overfitting can be avoided in the future.

5. Conclusion

Within the limits of this study the proposed k nearest neighbor algorithm shows a significant accuracy than the decision tree. The K-nearest neighbor algorithm will primarily reduce the effort of physically gathering ready data for arrangement. The strategic analysis accuracy has increased by about 7%. The outcome demonstrates that the characterization precision of the support vector machine was moderately low in this examination and k nearest neighbor algorithm has shown a better significant accuracy.

Declarations

Conflict of Interest

No conflict of interest in this manuscript.

Authors Contributions

Author YSR was involved in data collection, data analysis, manuscript writing. Author AB was involved in conceptualization, data validation, and critical review of manuscript.

Acknowledgements

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (Formerly known as Saveetha University) for providing the necessary infrastructure to carry out this work successfully.

Fundings

We thank the following organizations for providing financial support that enabled us to complete the study.

1. Ed-Trix Edu Pvt. Ltd.
2. Saveetha University,
3. Saveetha Institute of Medical and Technical Sciences,
4. Saveetha School of Engineering.

6. References

Ahuja, Lajpat R., and Liwang Ma. 2020. Methods of Introducing System Models into Agricultural Research. John Wiley & Sons.

Araus, José Luis, and Gustavo A. Slafer. 2011. Crop Stress Management and Global Climate Change. CABI.

Hatfield, Jerry L. 2011. "Changing Climate in North America: Implications for Crops." Crop Adaptation to Climate Change. https://doi.org/10.1002/9780470960929.ch5.

Hooper, Sarah E., Sybill K. Amelon, and Chung-Ho Lin. 2022. "Development of an LC-MS/MS Method for Non-Invasive Biomonitoring of Neonicotinoid and Systemic Herbicide Pesticide Residues in Bat Hair." Toxics 10 (2). https://doi.org/10.3390/toxics10020073.

Huntingford, C., F. H. Lambert, J. H. C. Gash, C. M. Taylor, A. J. Challinor, and Process Hydrology. 2005. Aspects of Climate Change Prediction Relevant to Crop Productivity.

Jalota, S. K., B. B. Vashisht, Sandeep Sharma, and Samanpreet Kaur. 2018. Understanding Climate Change Impacts on Crop Productivity and Water Balance. Academic Press.

Jaradat, Abdullah A. 2011. "Energy Crops to Combat Climate Change." Crop Adaptation to Climate Change. https://doi.org/10.1002/9780470960929.ch37.

Jayabal, Ravikumar, Sekar Subramani, Damodharan Dillikannan, Yuvarajan Devarajan, Lakshmanan Thangavelu, Mukilarasan Nedunchezhiyan, Gopal Kaliyaperumal, and Melvin Victor De Poures. 2022. "Multi-Objective Optimization of Performance and Emission Characteristics of a CRDI Diesel Engine Fueled with Sapota Methyl Ester/diesel Blends." Energy. https://doi.org/10.1016/j.energy.2022.123709.

Jayakodi, Santhoshkumar, Rajeshkumar Shanmugam, Bader O. Almutairi, Mikhlid H. Almutairi, Shahid Mahboob, M. R. Kavipriya, Ramesh Gandusekar, Marcello Nicoletti, and Marimuthu Govindarajan. 2022. "Azadirachta Indica-Wrapped Copper Oxide Nanoparticles as a Novel Functional Material in Cardiomyocyte Cells: An Ecotoxicity Assessment on the Embryonic Development of Danio Rerio." Environmental Research 212 (Pt A): 113153.

Juma, Godfrey Shem, and Festus Kelonye Beru. 2021. "Prediction of Crop Yields under a Changing Climate." Agrometeorology. https://doi.org/10.5772/intechopen.94261.

Kotteeswaran, C., Indrajit Patra, Regonda Nagaraju, D. Sungeetha, Bapayya Naidu Kommula, Yousef Methkal Abd Algani, S. Murugavalli, and B. Kiran Bala. 2022. "Autonomous Detection of Malevolent Nodes Using Secure Heterogeneous Cluster Protocol." Computers and Electrical Engineering. https://doi.org/10.1016/j.compeleceng.2022.107902.

Krishnan, Anbarasu, Duraisami Dhamodharan, Thanigaivel Sundaram, Vickram Sundaram, and Hun-Soo Byun. 2022. "Computational Discovery of Novel Human LMTK3 Inhibitors by High Throughput Virtual Screening Using NCI Database." Korean Journal of Chemical Engineering. https://doi.org/10.1007/s11814-022-1120-5.

Kuang, Yun-Hung, Yu-Fu Fang, Shau-Ching Lin, Shin-Fu Tsai, Zhi-Wei Yang, Charng-Pei Li, Shou-Horng Huang, Sherry Lou Hechanova, Kshirod K. Jena, and Wen-Po Chuang. 2021. "The Impact of Climate Change on the Resistance of Rice Near-Isogenic Lines with Resistance Genes Against Brown Planthopper." Rice 14 (1): 64.

Lotze-Campen, Hermann. 2011a. "Climate Change, Population Growth, and Crop Production: An Overview." Crop Adaptation to Climate Change. https://doi.org/10.1002/9780470960929.ch1.

2011b. "Regional Climate Impacts on Agriculture in Europe." Crop Adaptation to Climate Change. https://doi.org/10.1002/9780470960929.ch7.

Lumsden, T. G., and R. E. Schulze. n.d. "Application of Seasonal Climate Forecasts to Predict Regional Scale Crop Yields in South Africa." Climate Prediction and Agriculture. https://doi.org/10.1007/978-3-540-44650-7_21.

Ma, Gang, and Chun-Sen Ma. 2021. "Potential Distribution of Invasive Crop Pests under Climate Change: Incorporating Mitigation Responses of Insects into Prediction Models." Current Opinion in Insect Science 49 (October): 15–21.

Marco, David. 2000. Building and Managing the Meta Data Repository: A Full Lifecycle Guide. Wiley.

Mohan, Harshavardhan, Sethumathavan Vadivel, Se-Won Lee, Jeong-Muk Lim, Nanh Lovanh, Yool-Jin Park, Taeho Shin, Kamala-Kannan Seralathan, and Byung-Taek Oh. 2022.

"Improved Visible-Light-Driven Photocatalytic Removal of Bisphenol A Using V2O5/WO3 Decorated over Zeolite: Degradation Mechanism and Toxicity." Environmental Research. https://doi.org/10.1016/j.envres.2022.113136.

Mohan, Kannan, Abirami Ramu Ganesan, P. N. Ezhilarasi, Kiran Kumar Kondamareddy, Durairaj Karthick Rajan, Palanivel Sathishkumar, Jayakumar Rajarajeswaran, and Lorenza Conterno. 2022. "Green and Eco-Friendly Approaches for the Extraction of Chitin and Chitosan: A Review." Carbohydrate Polymers 287 (July): 119349.

Murakami, Keach, Seiji Shimoda, Yasuhiro Kominami, Manabu Nemoto, and Satoshi Inoue. 2021. "Prediction of Municipality-Level Winter Wheat Yield Based on Meteorological Data Using Machine Learning in Hokkaido, Japan." PloS One 16 (10): e0258677.

Peterson, Patti L., and John W. Phillis. 1995. Novel Therapies for CNS Injuries: Rationales and Results. CRC Press.

Redden, Robert. 2015. "Wild Relatives for the Crop Improvement Challenges of Climate Change: The Adaptation Range of Crops." Crop Wild Relatives and Climate Change. https://doi.org/10.1002/9781118854396.ch4.

Reynolds, Matthew P. 2010. Climate Change and Crop Production. CABI.

Salisbury, Phillip A., and Martin J. Barbetti. 2011. "Breeding OilseedBrassicafor Climate Change." Crop Adaptation to Climate Change. https://doi.org/10.1002/9780470960929.ch31.

Saravanan, A., P. Senthil Kumar, B. Ramesh, and S. Srinivasan. 2022. "Removal of Toxic Heavy Metals Using Genetically Engineered Microbes: Molecular Tools, Risk Assessment and Management Strategies." Chemosphere 298 (July): 134341.

Sathish, T., R. Saravanan, V. Vijayan, and S. Dinesh Kumar. 2022. "Investigations on Influences of MWCNT Composite Membranes in Oil Refineries Waste Water Treatment with Taguchi Route." Chemosphere 298 (July): 134265.

Sharannya, T. M., K. Venkatesh, Amogh Mudbhatkal, M. Dineshkumar, and Amai Mahesha. 2021. "Effects of Land Use and Climate Change on Water Scarcity in Rivers of the Western Ghats of India." Environmental Monitoring and Assessment 193 (12): 820.

Sivakumar, Mannava V. K., and James Hansen. 2007. Climate Prediction and Agriculture: Advances and Challenges. Springer Science & Business Media.

Tomar, Vipin, Daljit Singh, Guriqbal Singh Dhillon, Yong Suk Chung, Jesse Poland, Ravi Prakash Singh, Arun Kumar Joshi, Yogesh Gautam, Budhi Sagar Tiwari, and Uttam Kumar. 2021. "Increased Predictive Accuracy of Multi-Environment Genomic Prediction Model for Yield and Related Traits in Spring Wheat (L.)." Frontiers in Plant Science 12 (October): 720123.

Tyack, Nicholas, and Hannes Dempewolf. 2015. "The Economics of Crop Wild Relatives under Climate Change." Crop Wild Relatives and Climate Change. https://doi.org/10.1002/9781118854396.ch16.

Utset, A., J. Eitzinger, and V. Alexandrov. n.d. "AGRIDEMA: An EU-Funded Effort to Promote the Use of Climate and Crop Simulation Models in Agricultural Decision-Making." Climate Prediction and Agriculture. https://doi.org/10.1007/978-3-540-44650-7_24.

Vivek, J., T. Maridurai, K. Anton Savio Lewise, R. Pandiyarajan, and K. Chandrasekaran. 2022. "Recast Layer Thickness and Residual Stress Analysis for EDD AA8011/h-BN/B4C Composites Using Cryogenically Treated SiC and CFRP Powder-Added Kerosene." Arabian Journal for Science and Engineering. https://doi.org/10.1007/s13369-022-06636-5.

Wheeler, T., A. Challinor, T. Osborne, and J. Slingo. n.d. "Development of a Combined Crop and Climate Forecasting System for Seasonal to Decadal Predictions." Climate Prediction and Agriculture. https://doi.org/10.1007/978-3-540-44650-7_3.

Widener, Sarah, George Graef, Alexander E. Lipka, and Diego Jarquin. 2021. "An Assessment of the Factors Influencing the Prediction Accuracy of Genomic Prediction Models Across Multiple Environments." Frontiers in Genetics 12 (July): 689319.

Yaashikaa, P. R., M. Keerthana Devi, and P. Senthil Kumar. 2022. "Algal Biofuels: Technological Perspective on Cultivation, Fuel Extraction and Engineering Genetic Pathway for Enhancing Productivity." Fuel. https://doi.org/10.1016/j.fuel.2022.123814.

Yaashikaa, P. R., P. Senthil Kumar, and S. Karishma. 2022. "Review on Biopolymers and Composites – Evolving Material as Adsorbents in Removal of Environmental Pollutants." Environmental Research. https://doi.org/10.1016/j.envres.2022.113114.

Zhang, Qiang, Rui Shi, Vijay P. Singh, Chong-Yu Xu, Huiqian Yu, Keke Fan, and Zixuan Wu. 2022. "Droughts across China: Drought Factors, Prediction and Impacts." The Science of the Total Environment 803 (January): 150018.

Zhang, Yi, Yanxia Zhao, Sining Chen, Jianping

Guo, and Enli Wang. 2015. "Prediction of Maize Yield Response to Climate Change with Climate and Crop Model Uncertainties." Journal of Applied Meteorology and Climatology. https://doi.org/10.1175/jamc-d-14-0147.1.

Tables and Figures

Table 1. Descriptive statistics show the output of the descriptive statistics of the dataset. It consists of accuracies of both k-near0est neighbor 85.8270% and decision tree 80.0620%.

No of Experiments	Number of reviews in the training dataset	K-Nearest Neighbor			Decision Tree		
		Accuracy	Precision	Recall	Accuracy	Precision	Recall
1	100	79.91	1.007	1.002	73.43	1.002	1.008
2	200	80.92	1.009	1.004	75.73	0.001	0.007
3	300	81.94	1.005	0.940	76.56	1.007	0.949
4	400	82.93	1.003	1.006	78.58	0.009	0.006
5	500	83.96	0.842	0.762	79.82	0.004	0.003
6	600	86.46	0.861	0.920	81.17	0.036	1.007
7	700	89.39	0.856	1.007	82.49	0.008	0.006
8	800	90.97	1.001	1.009	86.52	0.003	0.005
9	900	95.99	0.956	1.008	85.36	0.952	1.004
10	1000	85.80	0.750	0.945	80.96	0.005	0.003

Table 2. T-Test comparison novel k-nearest neighbor has higher accuracy than decision tree. Descriptive Statistics minimum, maximum, mean and standard deviation of two groups novel k-nearest neighbor and decision tree 10 sample size is taken for both proposed and existing.

	Group	N	Mean	Std.Deviation	Std.Error Means
Accuracy	K-Nearest Neighbor	10	85.8270	5.04326	1.59482
	Decision Tree	10	80.0620	4.14960	1.31222

Table 3. Group statistics T-Test has dependent values for k-nearest neighbor standard error mean and decision tree independent sample t-test is applied for the data set fixing confidence interval as 95% and level of significance as 0.02. There is a significant difference in accuracy(p=0.02).

	Levene's	T-test for equality of means

		test for equality of variances		t	df	sig(2 tailed)	Mean difference	Std error difference	95% confidence interval of the difference	
		f	sig						Lower	Upper
Accuracy	Equal variances assumed	0.311	0.02	2.791	18	0.02	5.76500	2.06527	1.42602	10.10398
	Equal variances not assumed	0.311	0.02	2.791	17.356	0.02	5.76500	2.06527	1.41445	10.11555

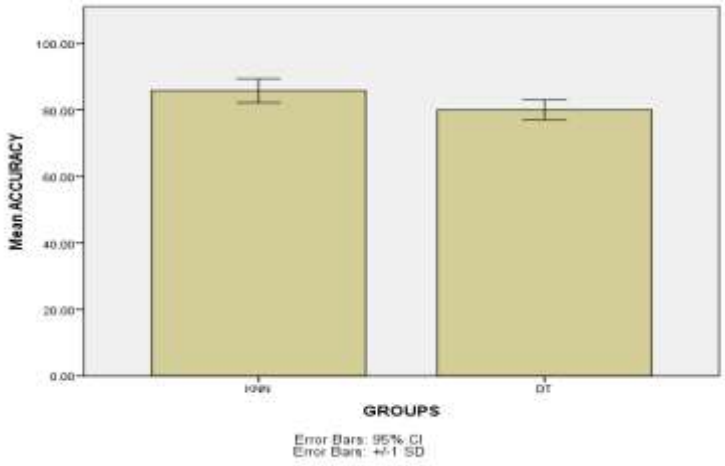

Fig. 1. Bar chart representing the comparison mean accuracy of k-nearest neighbor algorithm and decision tree. The mean accuracy of the novel k-nearest neighbor is better than the decision tree and standard deviation of novel k-nearest neighbor is slightly better than the decision tree. X-axis: k-nearest neighbor algorithm vs decision tree algorithm. Y-Axis: Mean accuracy of detection +/- 1SD.