Comparative Analysis of Machine Learning Classifier Algorithms for Agricultural Multi-Class Prediction: A Case Study on the Fish Catch Dataset

Metta Vaishnavi[1], G V Ramesh Babu[2]
[1]PG Scholar, Dept. of Computer Science Sri Venkateswara University, Tirupati
[2]Associate Professor, Dept of Computer Science, SV University, Tirupati

Abstract— *In this research paper, we present a comprehensive analysis of machine learning classifier algorithms applied to agricultural multi-class prediction, using the Fish Catch dataset sourced from the UCI Machine Learning Repository. We assess the performance of two machine learning classifier algorithms, K-Nearest Neighbors (KNN) and Logistic Regression, using key performance metrics such as Accuracy, Precision, and Recall. The comparison between these algorithms illustrates the potential of machine learning in agricultural applications, particularly in multi-class prediction scenarios. Logistic Regression's effectiveness in achieving high accuracy and precision underscores its suitability for such tasks. Experimental results demonstrate that Logistic Regression outperforms KNN in terms of accuracy and precision. This exploratory study sheds light on the suitability and effectiveness of classification algorithms for fish catch prediction.*

I. INTRODUCTION

Information mining is the key stage in the information revelation process. The primary assignments of Information mining are for the most part isolated into two classes: Prescient and Clear. The target of the prescient undertakings is to anticipate the worth of a specific quality in light of the upsides of different traits, while for the unmistakable ones, the goal is to remove beforehand obscure and helpful data like examples, affiliations, changes, peculiarities and huge designs, from enormous data sets [2].

Information mining is an interdisciplinary field, drawing work from regions including man-made consciousness, AI, brain organizations, insights, design acknowledgment, data set innovation, data recovery, and information perception. In this examination we considered Arrangement Rule Digging for information disclosure and produced the standards by applying our created approach on fish get dataset.

II. CLASSIFICATION

Grouping is the most common way of finding a model or a capability that portrays and recognizes information classes and ideas, to utilize the model to foresee the classes of items whose class mark isn't known. Information characterization can be seen as a two stage process: learning step in which a classifier is fabricated depicting a foreordained arrangement of classes or ideas by breaking down the preparation set comprised of data set tuples and their related marks [3][6]. In the subsequent step model is utilized for grouping by first assessing the prescient exactness of classifier worked during the initial step. It is finished utilizing the test information. The precision of classifier on a given test set tuples is level of tuples that are accurately characterized by the classifier. On the off chance that the precision is over some OK level, the classifier can be utilized to anticipate future tuples whose class name isn't known.

III. METHODOLOGY

A great many kinds of request methodology have been proposed recorded as a hard copy that integrates Choice Trees, Credulous Bayesian strategies, Brain Organizations, Calculated Relapse, SVM and KNN, etc. In this paper, we evaluate the presentation of the Strategic Relapse and of K-Closest Neighbors (KNN) for anticipating the Fish Catch and diagrams the goals of this near study.

3.1 Logistic Regression

Key backslide is an assessment used to predict a twofold outcome: either something happens, or doesn't. This can be shown as Yes/No, Substantial/False. Independent variables are bankrupt down to determine the twofold outcome with the end results having a tendency to be ordered as one of two groupings [3][4]. The free factors can be hard and fast or numeric, but the dependent variable is constantly obvious. Made along these lines:

P(Y=1|X) or P(Y=0|X)

It determines the probability of ward variable Y, given free element X. This can be used to process the probability of a word having a decent or deplorable basic significance (0, 1, or on a scale between). Then again it will in general be used to conclude the article contained in a photo (tree, blossom, grass, etc), with every thing given a probability some place in the scope of 0 and 1.

3.2 K-Nearest Neighbor (KNN)

KNN request portrays events considering their closeness. A thing is requested by a larger piece of its neighbors. K is reliably a positive number. The neighbors are browsed a lot of things for which the right gathering is known [3]. The planning tests are depicted by n layered numeric qualities.

Every model tends to a point in a n-layered space. Thusly, the planning tests are undeniably taken care of in a n-layered plan space. Exactly when given a dark model, a k-nearest neighbor classifier glance through the model space for the k planning tests that are closest to the dark model. "Closeness" is portrayed concerning Euclidean distance [4][5]. The dark model is allotted the most generally perceived class among its k nearest neighbors. When k=1, the dark model is given out the class of the readiness test that is closest to it in plan space.

IV. EXPERIMENTAL RESULTS

The objective of this section is to evaluate two machine learning algorithms in terms of learning accuracy for Fish Catch dataset sourced from the UCI Machine Learning Repository [7]. We have used the Python Language to experiment our proposed algorithms. The dataset consists of 252 cases with 15 recorded features and 7 class labels, including 252 cases in the Dynamic class and 7 cases in the Dormant class. To conduct the evaluation, the dataset was divided into a training set (70%) and a testing set (30%).

The performance of two ml classifier algorithms, KNN and Logistic Regression, was assessed using various performance metrics including Accuracy, Precision, and Recall. The experimental results are summarized in Figure 1.

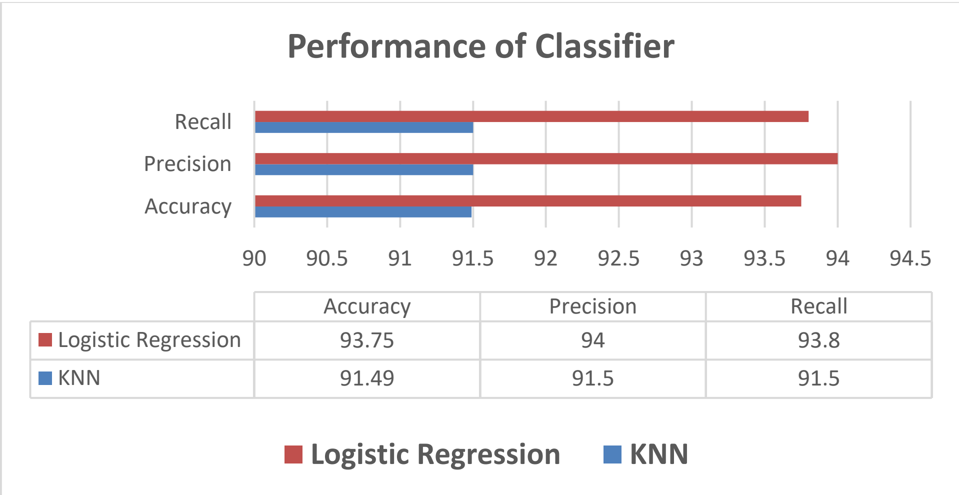

Figure-1: Performance of classifiers

Figure-1 presents the performance of the KNN and Logistic Regression classifier algorithms for agricultural multi-class prediction using the Fish Catch dataset. The evaluation is based on crucial performance metrics:

V. DISCUSSION

The experimental results highlight the performance of machine learning classifier algorithms in predicting agricultural multi-class outcomes using the Fish Catch dataset. Both KNN and Logistic Regression algorithms demonstrate strong predictive capabilities, with Logistic Regression exhibiting a slight advantage.

Logistic Regression achieves the highest accuracy of 93.75%, indicating its proficiency in accurately classifying agricultural multi-class cases. Furthermore, its precision and recall scores of 94% and 93.8%, respectively, emphasize its ability to minimize both false positives and false negatives, critical factors in agricultural decision-making.

KNN, while slightly behind Logistic Regression in terms of accuracy and precision, still performs commendably, with an accuracy score of 91.49% and precision and recall scores of 91.5%. These results demonstrate KNN's competence in agricultural multi-class prediction.

The comparison between these algorithms illustrates the potential of machine learning in agricultural applications, particularly in multi-class prediction scenarios. Logistic Regression's effectiveness in achieving high accuracy and precision underscores its suitability for such tasks.

VI. CONCLUSION

In conclusion, this research underscores the efficacy of machine learning classifier algorithms, specifically KNN and Logistic Regression, in agricultural multi-class prediction using the Fish Catch dataset. Logistic Regression exhibits superior performance in terms of accuracy, precision, and recall, making it a valuable tool for agricultural decision support.

These findings have significant implications for agriculture, where accurate prediction and decision-making are crucial for crop management, yield optimization, and resource allocation. The use of machine learning algorithms in agriculture holds promise for enhancing productivity and sustainability.

Future research can explore the integration of additional features, advanced algorithms, and real-world validation to further improve predictive accuracy and the generalizability of agricultural multi-class prediction models.

REFERENCES

[1] G. Ravi Kumar, K. Venkata Sheshanna, S. Rahamat Basha, and P. Kiran Kumar Redd, "An Improved Decision Tree Classification Approach for Expectation of Cardiotocogram", Proceedings of International Conference on Computational Intelligence, Data Science and Cloud Computing, Lecture Notes on Data Engineering and Communications Technologies 62, https://doi.org/10.1007/978-981-33-4968-1_26

[2] Ian H. Witten and Eibe Frank. Data Mining: Practical machine learning tools and techniques. 2nd ed. San Francisco: Morgan Kaufmann, 2005.

[3] J.Han and M.Kamber,"Data Mining concepts and Techniques", the Morgan Kaufmann series in Data Management Systems, 2nd ed.San Mateo, CA; Morgan Kaufmann, 2006.

[4] N.Michael, "Artificial Intelligence – A Guide to Intelligent Systems", 2nd Edition, Addison Wesley 2005

[5] P.N.Tan, M.Steinbach and V.Kumar "Introduction to Data Mining", A: Addision-Wesley, 2005.

[6] M. V. Lakshmaiah, G. Ravi Kumar and G. Pakardin, "Frame work for Finding Association Rules in Bid Data by using Hadoop Map/Reduce Tool", International Journal of Advance and Innovative Research, Volume 2, Issue1(1), PP:6-9,2015, ISSN: 2394-7780.

[7] UCI Machine Learning Repository. https://archive.ics.uci.edu/ml/.