IJCRT.ORG ISSN : 2320-2882
INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

Classifying Data With Suggestive Causes And Flexible Solution

Ms.Gulla Tusharika*[1],Dr.P.Praveen Kumar*[2]

[1]MCA Student, Department of Master of Computer Applications,
Vignan's Institute of Information Technology(A), Beside VSEZ,Duvvada,Vadlapudi Post,
Gajuwaka, Visakhapatnam-530049.
[2]Associate Professor, Department of Information Technology,
Vignan's Institute of Information Technology(A), Beside VSEZ,Duvvada,Vadlapudi Post,
Gajuwaka, Visakhapatnam-530049.

Abstract:

This project is an effective technique to categorize data on specific requirements and conditions for the majority of marine species. To determine which species are all threatened or endangered, depending on the situation. Not only is the bulk of the data being classified here, but also the causes of and fixes for the classification of the data. The actual process of decision-making and branching based on the qualities of the data is carried out using a set of algorithms and tools. It facilitates the creation of efficient machine learning models that are capable of making precise predictions. Logistic regression is used to estimate discrete values (typically binary values like 0/1) from a collection of independent variables. The logic function is changed to match the data,It helps in estimating how likely an event is to occur. These algorithms function on this application well. Since there are just two outcomes in this project, logistic regression is employed as a binary classifier. It will also explain the causes and a fix for that here.

Keywords: Machine Learning Models,Logistic Regression,Logic Function,Binary Classifier,Precise Prediction.

1. INTRODUCTION

The study of the oceans covers a variety of topics, including as marine life, ocean currents and waves, fluxes, and more. This helped us learn more about how organisms live and explored numerous previously unknown facts and processes that are taking place underwater. This study provides a way for people to comprehend the entire ecosystem and effectively classify the vast majority of marine organisms according to specified needs and conditions. The identification of endangered or threatened species, as well as their causes, are also covered in this study. According to the needs of the clients, this project tries to forecast which oceanic species are endangered or threatened. Researchers must evaluate if the population is dropping and whether the vulnerable or endangered species are in danger in order to decide whether theyThe report's output wouldn't be precise. The difficulty of manually handling a big volume of data with 100% accuracy by people is the cause. They need to use machine learning approaches to overcome this issue. By giving the machine the majority of the data, it can process it and generate an accurate output based on the input. A tool that can duplicate the actions of an intelligent person in a way that is similar to their own is called a machine. Machine learning systems can automate complex activities in a variety of methods that are comparable to how people solve problems in the real world. In the end, based on the results' classification,The report will be created in PDF format according to the specifications and sent to the analyst for analysis on how these species are vulnerable or endangered. As soon as feasible, the final report will be created and provided to the client.

2. LITERATURE SURVEY

The provided text discusses the concept of Heterogeneous Domain Adaptation (HDA) and its foundational requirement of domain alignment, specifically in the context of different feature spaces between the source and target domains. The text introduces two main categories of HDA methods, focusing on the alignment of source and target domains through asymmetrical space transformations. Here's a detailed breakdown:

1. Heterogeneous Domain Adaptation (HDA):

- HDA is described as a more general Domain Adaptation (DA) scenario that relaxes the assumption of homogeneous DA. In HDA, the source and target domains are allowed to come from different feature spaces, indicating a more flexible and versatile adaptation requirement.

2. Foundational Role of Domain Alignment in HDA:

- The text emphasizes that domain alignment, especially in terms of feature space alignment, is foundational to HDA. This implies that successful adaptation between source and target domains in HDA relies on aligning their respective feature spaces.

3. Categories of HDA Methods:

- The text introduces two main categories of HDA methods, both aiming to align source and target domains through asymmetrical space transformations.

a. Asymmetrical Space Transformation:

- The first category involves aligning the source domain to the target domain or vice versa by projecting one domain onto the other.

b. Examples of Asymmetrical Space Transformation Methods:

- The text provides examples of HDA methods falling under the asymmetrical space transformation category:

- Kernel Projection by Kulis et al. [1]:

- This method involves matching the target domain with nonlinear kernel projection to the source domain. Nonlinear kernel projections are used to align the two domains.

- Sparse Feature Conversation by Zhou et al. [4], [5]:

- This method aligns the target domain with sparse feature conversation to the source domain. After alignment, classification is performed for the target instances.

- Distance Modeling by Luo et al. [7]:

- This introduces distance modeling to facilitate DA. The target domain distance space is modeled by initializing it with the source space prior.

- Wasserstein Criterion by Yan et al. [6]:

- This method applies the Wasserstein criterion to measure the domain distance between the source and target domains.

4. Importance of Feature Space Alignment:

- The discussion emphasizes the importance of aligning feature spaces in HDA, indicating that asymmetrical space transformations play a crucial role in achieving this alignment between domains.In summary, the provided text introduces the concept of Heterogeneous Domain Adaptation (HDA) and highlights the significance of domain alignment, particularly in feature space, in this context. It categorizes HDA methods into two groups, with examples of asymmetrical space transformation methods that aim to align source and target domains through various techniques.

3. EXISTING SYSTEM

In the existing system, it is very difficult to identify, which are endangered and which are threatened species from the bulk of data. This data is collected from the oceanographers, which was taken in certain area at certain period of time from the ocean. Over a period of time there is lot of species data that is getting recorded. The count is also massive, the recorded data is compared with the previously recorded data. Finding the endangered species or threatened species is very difficult. At the same time it takes lot time also the accuracy of the manipulated data is questionable. There are 2 million known species are in the ocean which is not an easy job to handle the data manually. This process contains lot of issues like avoidable errors, lot of time consumption, inconsistency, high cost of training, lack of accuracy. The inaccuracy may leads to biggest disaster.

Limitations of the Existing System:

1. The manual process is undertaken by humans who are not infallible in the performance of repetitive tasks.
2. There is no way a man can complete with the machine in term of processing speed.
3. In an environment where consistency of data is crucial to the success of the system.
4. Manual data manipulation service is always a disadvantage as maintaining consistency for humans is a challenging task that must be avoided.
5. Classified information may leak, and sensitive data may develop legs and walk away and thus compromise the entire research.

4. PROPOSED SYSTEM

In this proposed system, a machine learning model is implemented to find the endangered species or threatened species by using a set of algorithms like decision tree, regression models to classify from the bulk data and also finding the causes and provide flexible solution. Machine learning algorithms can achieve much higher accuracy than humans when making predictions or classifying labelled data. The solution is also cost effective which can lead to better business outcomes and increased profits

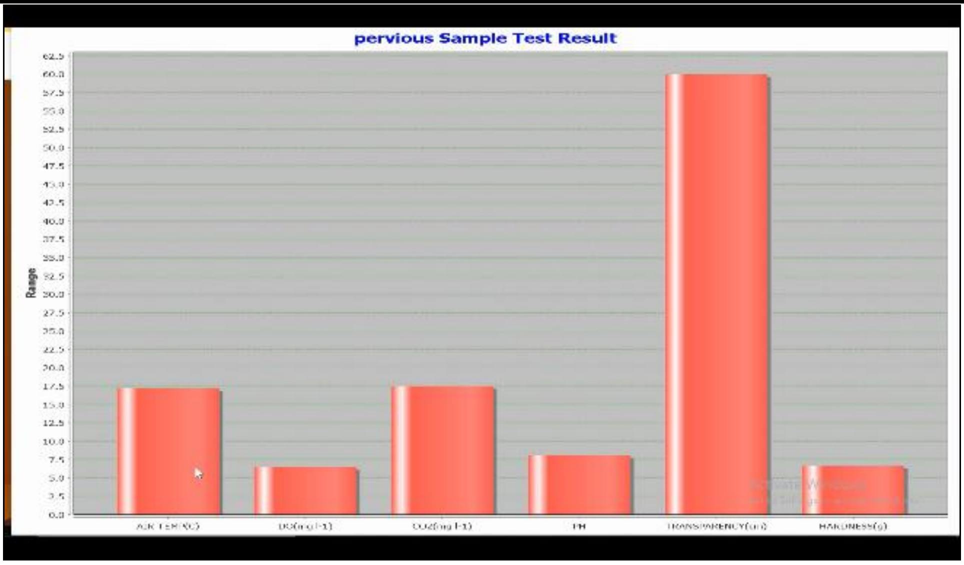
Advantages of the Proposed System:

1. Big data analysts are looking at machine learning as the most effective source for precise data prediction.
2. It will give high accuracy on data manipulation.
3. High accuracy will helps the good research results.
4. This method can handle an unlimited amount of data, assess them and provide a proper analysis for the same.
5. Besides being cost-effective and time-saving, it is also easy to operate.

5. EXPERIMENTAL RESULTS

From the below figures it can be seen that proposed model is more accurate in order to prove our proposed system.

Main Window:



6. CONCLUSION

The decision tree technique and the logistic regression approach have both been used in our suggested version. The two varieties of algorithms operate on the application more successfully. We have to utilize the decision tree method to categorize the majority of the data. We apply the logistic regression algorithm to identify the causes and provide solutions for the issues. Based on the specified criteria, the categorization procedure will determine which species are endangered or threatened. Next, get the oceanographers to give you the results of the lab tests on the water samples. because determining the causes of such species' endangerment or threat. These sampling-related information is gathered over a set amount of time. As a result, the version we've suggested produces excellent results and meets a crucial research need. With this procedure, results will be accurate and timely. In the future, it was improved and put to use while being tested for a situation where it was really needed.

References

[1] B. Kulis, K. Saenko, and T. Darrell, "*What you saw is not what you get: Domain adaptation using asymmetric kernel transforms*," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2011, pp. 1785–1792.

[2] Y. Wang, X. Chen, S. Chen, and H. Xue, "*Adaptive teacher-and-student model for heterogeneous domain adaptation*," in Proc. IEEE Int. Conf. Data Mining, 2019, pp. 1342–1347.

[3] J. Hoffman, E. Rodner, J. Donahue, B. Kulis, and K. Saenko, "*Asymmetric and category invariant feature transformations for domain adaptation*," Int. J. Comput. Vis., vol. 109, nos. 1–2, pp. 28–41, 2014.

[4] J. T. Zhou, I. W. Tsang, S. J. Pan, and M. Tan, "*Heterogeneous domain adaptation for multiple classes*," in Proc. Artif. Intell. Stat., 2014, pp. 1095–1103.

[5] J. T. Zhou, I. W. Tsang, S. J. Pan, and M. Tan, "*Multi-class heterogeneous domain adaptation*," J. Mach. Learn. Res., vol. 20, no. 57, pp. 1–31, 2019.

[6] Y. Yan, W. Li, H. Wu, H. Min, M. Tan, and Q. Wu, "*Semi-supervised optimal transport for heterogeneous domain adaptation*," in Proc. Int. Joint Conf. Artif. Intell., 2018, pp. 2969–2975.

[7] Y. Luo, Y. Wen, T. Liu, and D. Tao, "Transferring knowledge fragments for learning distance metric from a heterogeneous *domain*," IEEE Trans. Pattern Anal. Mach. Intell., vol. 41, no. 4, pp. 1013–1026, Apr. 2019.

[8] H.-J. Ye, X.-R. Sheng, D.-C. Zhan, and P. He, "*Distance metric facilitated transportation between heterogeneous domains*," in Proc. Int. Joint Conf. Artif. Intell., 2018, pp. 3012–3018.

[9] S. Sukhija, N. C. Krishnan, and G. Singh, "*Supervised heterogeneous domain adaptation via random forests*," in Proc. Int. Joint Conf. Artif. Intell., 2016, pp. 2039–2045.

[10] C. Wang and S. Mahadevan, "*Heterogeneous domain adaptation using manifold alignment*," in Proc. 22nd Int. Joint Conf. Artif. Intell., 2011, pp. 1541–1546.

[11] M. Xiao and Y. Guo, "*Semi-supervised subspace co-projection for multiclass heterogeneous domain adaptation*," in Proc. Joint Eur. Conf. Mach. Learn. Knowl. Disc. Databases, 2015, pp. 525–540.

[12] S. M. Kakade and D. P. Foster, *"Multi-view regression via canonical correlation analysis*," in Proc. Int. Conf. Computing. Learn. Theory, 2007, pp. 82–96.

[13] H. Zuo, J. Lu, G. Zhang, and W. Pedrycz, "*Fuzzy rule-based domain adaptation in homogeneous and heterogeneous spaces*," IEEE Trans. Fuzzy Syst., vol. 27, no. 2, pp. 348–361, Feb. 2019.

[14] J. Li, K. Lu, Z. Huang, L. Zhu, and H. T. Shen, "*Transfer independently together: A generalized framework for domain adaptation*," IEEE Trans. Cybernet., vol. 49, no. 6, pp. 2144–2155, Jun. 2019.
[15] H. Hotelling, **"***Relations between two sets of variates*," in Breakthroughs in Statistics. New York, NY, USA: Springer, 1992, pp. 162–190.

[16] D. R. Hardoon, S. Szedmak, and J. Shawe-Taylor, "*Canonical correlation analysis: An overview with application to learning methods*," Neural Computing., vol. 16, no. 12, pp. 2639–2664, 2004.

[17] W.-Y. Chen, T.-M. H. Hsu, Y.-H. H. Tsai, M.-S. Chen, and Y.-C. F. Wang, "*Transfer neural trees: Semi-supervised heterogeneous domain adaptation and beyond*," IEEE Trans. Image Process., vol. 28, no. 9, pp. 4620–4633, Sep. 2019.

[18] H. Li, S. J. Pan, R. Wan, and A. C. Kot, "*Heterogeneous transfer learning via deep matrix completion with adversarial kernel embedding*," in Proc. AAAI Conf. Artif. Intell., vol. 33, 2019, pp. 8602–8609.

[19] Y.-H. H. Tsai, Y.-R. Yeh, and Y.-C. F. Wang, "*Learning cross-domain landmarks for heterogeneous domain adaptation*," in Proc. IEEE Conf. Computing. Vis. Pattern Recognition., 2016, pp. 5081–5090

[20] Y. Yao, Y. Zhang, X. Li, and Y. Ye, "*Discriminative distribution alignment: A unified framework for heterogeneous domain adaptation*," Pattern Recognition., vol. 101, May 2020, Art. no. 107165.