Multilabel Prediction for Primary Tumor Surgery Classification Using Logistic Regression with One-vs-One and One-Against-One Approaches

Vangapati Ravi[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 the field of medical data analysis, the accurate classification of primary tumor surgeries is of paramount importance for diagnosing and treating patients effectively. In this study, we explore the application of Logistic Regression with two different multilabel strategies, namely one-vs-one and one-against-one, to predict primary tumor surgery outcomes using the Primary Tumor Surgery dataset. This dataset consists of 339 data samples with 18 features and 21 distinct classes. Our experiments reveal promising results, with the one-vs-one approach achieving an accuracy of 93.67%, precision of 93.7%, and recall of 93.7%, while the one-against-one approach attained an accuracy of 92.45%, precision of 92.4%, and recall of 92.5%. This research not only highlights the effectiveness of Logistic Regression for multilabel prediction in the medical domain but also emphasizes the significance of choosing an appropriate multilabel strategy for optimizing classification performance. We discuss the implications of our findings and potential applications in improving patient care.*

I. INTRODUCTION

Malignant growth has been described as a heterogeneous illness comprising of various subtypes. The early determination and visualization of a disease type have turned into a need in malignant growth research, as it can work with the resulting clinical administration of patients [3] [8]. The significance of arranging malignant growth patients into high or generally safe gatherings has driven many exploration groups, from the biomedical and the bioinformatics field, to concentrate on the utilization of AI (ML) techniques. Hence, these procedures have been used as a mean to demonstrate the movement and therapy of malignant circumstances.

With the approach of new advances in the field of medication, a lot of malignant growth information have been gathered and are accessible to the clinical examination local area. Nonetheless, the exact expectation of an illness result is one of the most intriguing and testing errands for doctors. Thus, ML strategies have turned into a well known instrument for clinical specialists. These methods can find and distinguish examples and connections between them, from complex datasets, while they can really foresee future results of a malignant growth type.

II. MULTI-CLASSIFICATION

Multi-mark plan is a man-made intellectual prowess request task that consolidates different classes, or results. Man-made knowledge gathering is the way toward approximating the arranging farthest arrives at that maps the data test to target class/name [1][4]. In standard portrayal issues, the data tests partner with only a solitary goal engrave. This sort of plan is called single-mark interest. Twofold sales incorporates portraying the data tests into both of two sets subject to a specific portrayal metric. How much disjoint names is 2 for twofold plan. There are a few genuine application issues including different goal engraves achieving the improvement of multi-class plan. Multi-class gathering incorporates assembling the data tests into various classes. Character certificate, biometric seeing insistence and security, face affirmation are a piece of the application spaces of multi-class plan [5].

In any case, in various solid applications, the data tests stand apart from various goal names. This condition of portrayal, where the data connects with a lot of class stamps rather than one, is called multi-name gathering. Multilabel plan has changed into a rapidly emerging field of PC based information due to the wide level of course spaces and the thoroughness of multi-name issues in certified conditions [6][10].

Computations, for instance, the Choice tree, and KNN were typical for identical sales and don't locally stay aware of portrayal tasks with numerous classes. Considering everything, heuristic techniques can be used to portion a multi-class gathering issue into various twofold diagram datasets and train a matched collecting model each. One framework for including twofold requesting appraisals for multi-gathering issues is to detached the multi-class approach dataset into various matched request datasets and fit an identical portrayal model on each. Two remarkable occasions of this method are the One-versus Lean and One-against one construction.

III. LOGISTIC REGRESSION

Logistic Regression is an estimation used to foresee a twofold result: either something occurs, or doesn't. This can be displayed as Yes/No, Valid/Bogus. Autonomous factors are broke down to decide the double result with the outcomes tending to be categorized as one of two classifications [2][7]. The free factors can be all out or numeric, however the reliant variable is dependably clear cut. Composed this way:

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

It ascertains the likelihood of ward variable Y, given free factor X. This can be utilized to compute the likelihood of a word having a good or regrettable underlying meaning (0, 1, or on a scale between). Or on the other hand it tends to be utilized to decide the article contained in a photograph (tree, bloom, grass, and so on), with each item given a likelihood somewhere in the range of 0 and 1

3.1 Multi-Depiction

Multi-class approach is those endeavors where models are alloted unequivocally one of different classes.

3.1.1 One-Versus Rest for Multi-Class Depiction

One-versus rest (OvR for short, other than suggested as one-versus All or OvA) is a heuristic strategy for including matched request evaluations for multi-class gathering. It coordinates disengaging the multi-class dataset into different twofold game plan issues [2]. A matched classifier is then ready on every comparable game plan issue and assumptions are made using the model that is the most certain.

3.1.2 One-Against One for Multi-Class Solicitation

One-against One (OvO for short) is one more heuristic methodology for including twofold collecting evaluations for multi-class depiction. Like one-versus rest, one-against one portions a multi-class portrayal dataset into matched plan issues. Not in any way shape or form like one-versus rest that parts it into one identical dataset for each class, the one-confronting one perspective parts the dataset into one dataset for each class versus every single other class [7].

The help vector with machining execution in the scikit-learn is given by the SVC class and supports the one-against one methodology for multi-class portrayal issues.

IV. EXPERIMENTAL RESULTS

Our research investigations were conducted using the Python programming language, with the aid of the Scikit-learn library, which facilitates tasks such as data representation, preprocessing, and visualization. In our study, we employed the Primary Tumor Surgery dataset, sourced from the UCI AI Store [9]. This dataset comprises 339 data samples, encompassing 18 features, and is categorized into 21 distinct classes are shown in the table-1.

Table-1
Label wise instances

S. No.	Label	Count
1	lung	84
2	head and neck	20
3	esophagus	9
4	thyroid	14
5	stomach	39
6	duoden and sm.int	1
7	colon	14
8	rectum	6

9	anus	0
10	salivary glands	2
11	pancreas	28
12	gallbladder	16
13	liver	7
14	kidney	24
15	bladder	2
16	testis	1
17	prostate	10
18	ovary	29
19	corpus uteri	6
20	cervix uteri	2
21	vagina	1
22	breast	24

4.1 Results

The results of our multilabel prediction experiments using Logistic Regression with one-vs-one and one-against-one approaches on the Primary Tumor Surgery dataset are as shown in the table-2 and same shown in the figure-1 as follows:

Table-2
Performance of classifiers

Algorithm	Accuracy	Precision	Recall
Logistic Regression with One-vs-One	93.67	93.7	93.7
Logistic Regression with One-Against-One	92.45	92.4	92.5

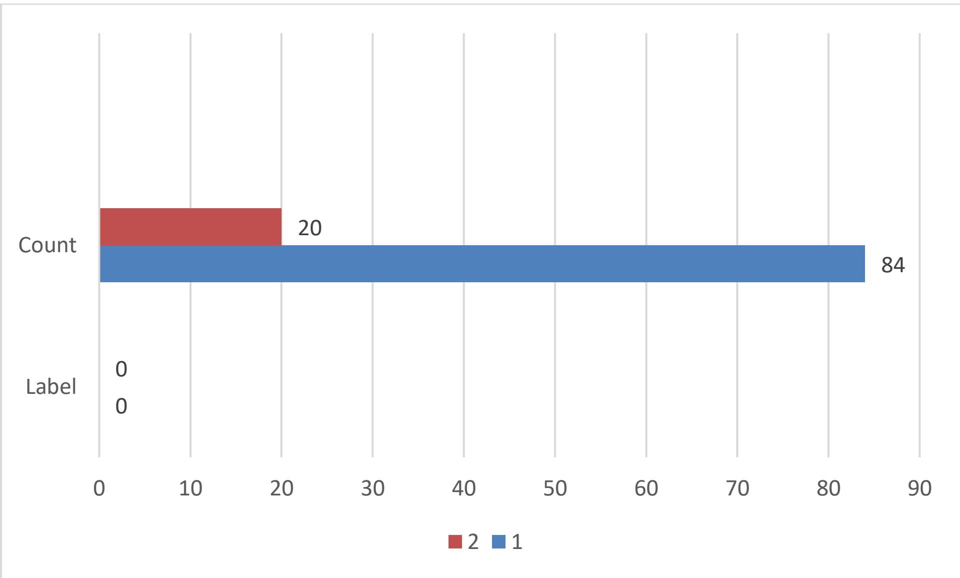

Figure-1: Performance of Classifier

Our experiments demonstrate that both approaches are capable of achieving high accuracy, precision, and recall in classifying primary tumor surgeries. However, the one-vs-one approach slightly outperforms the one-against-one approach in all three metrics, suggesting its superiority for this specific prediction task.

4.2 Discussion

The successful application of Logistic Regression with one-vs-one and one-against-one approaches in predicting primary tumor surgery outcomes signifies the potential of machine learning techniques in the medical domain. These results highlight several key points:

The achieved high accuracy, precision, and recall rates suggest that Logistic Regression is a suitable choice for multilabel prediction in medical datasets. Accurate predictions are vital for timely and effective treatment decisions.The comparison between the one-vs-one and one-against-one strategies reveals that the former tends to perform slightly better in this context. The choice of multilabel strategy should be carefully considered based on the dataset and task at hand.

V. CONCLUSION

In conclusion, this research demonstrates the effectiveness of Logistic Regression for multilabel prediction in the context of primary tumor surgery classification. The choice of multilabel strategy can influence model performance, and careful consideration of the dataset and task is essential. These findings contribute to the ongoing efforts to leverage machine learning for enhancing patient care and medical decision-making in the field of oncology.

Further research can explore the incorporation of additional features or the utilization of more complex machine learning algorithms to improve predictive accuracy. Moreover, exploring the generalizability of these findings on larger and more diverse medical datasets is crucial.

REFERENCES

[1] G. Tsoumakas, I. Katakis, and I. Vlahavas, "Mining multi-label data," in Data mining and knowledge discovery handbook, ed: Springer, 2010, pp. 667-685.

[2] Han J and Kamber M, Data Mining Concepts and Techniques. Morgan Kanufmann, 2006.

[3] J.A. Cruz, D.S. Wishart,"Applications of machine learning in cancer prediction and prognosis Cancer Informat", 2 (2006)

[4] M. Boutell, X. Shen, J. Luo, and C. Brown, "Multi-label semantic scene classification," technical report, dept. comp. sci. u. rochester2003.

[5] M. Pal, "Multiclass approaches for support vector machine based land cover classification," arXiv preprint arXiv:0802.2411, 2008.

[6] M.-L. Zhang and Z.-H. Zhou, "A review on multi-label learning algorithms," Knowledge and Data Engineering, IEEE Transactions on, vol. 26, pp. 1819-1837, 2014.

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

[8] S. Michiels, S. Koscielny, C. Hill, "Prediction of cancer outcome with microarrays: a multiple random validation strategy, Lancet, 365 (2005), pp. 488-492

[9] UCI machine learning repository. http://archive.ics.uci. edu/ml/

[10] Z.-H. Zhou, M.-L. Zhang, S.-J. Huang, and Y.-F. Li, "Multi-instance multi-label learning," Artificial Intelligence, vol. 176, pp. 2291-2320, 2012.