Asian Journal of Electrical Sciences
ISSN: 2249-6297 (P) Vol.9 No.2, 2020, pp.13-16

DOI: https://doi.org/10.51983/ajes-2020.9.2.2551

Visualization of Net Effects for Image Hiding Using Gain/Lift

R.Ganesh[1], S.Thabasu Kannan[2] and S.Selvam[3]
[1]Research Scholar, Periyar University, Salem, Tamil Nadu, India.
[2]Pannai College of Engg & Tech, Sivagangai, Tamil Nadu, India
[3]Head, Dept. of Computer Applications, N.M.S.S. Vellaichamy Nadar College, Madurai, Tamil Nadu, India
E-mail: yourmrganesh@gmail.com, thabasukannan@gmail.com, s.selvammscmphil@gmail.com.
(Received 5 July 2020; Revised 24 July 2020; Accepted 20 August 2020; Available online 26 August 2020)

***Abstract* - At present most of the larger companies depends heavily on their data science capabilities for taking decisions. On the basis of complexity and diversity of analysis, the big data units are transformed into larger and more technologies. Internet technologies are now playing a vital role in our day to day life. It has the advantages along with the disadvantages also, which in term generates the requirements of image hiding technology for maintaining the secrecy of the secret information. The interpretability of findings plays a major role for the success of delivering data science solutions into business reality. Even if the existing method provides outstanding accuracy, they may be neglected if they do not hide the image or text in an exact manner for various cases. When evaluating ML/DL [1] models there is an excess of possible metrics to assess performance. There are things like accuracy, precision-recall, ROC and so on. All of them can be useful, but they can also be misleading or don't answer the question at hand very well. The ROC AUC score is not informative enough for taking decisions since it is abstract for non-technical managers.**

Hence two more informative and meaningful metrics that every analyst should take into consideration when illustrating the results of their binary classification models: Cumulative Gains and Lift charts. Both the metrics are extremely useful to validate the predictive model (binary outcome) quality.

Gain and Lift charts [2] are used to update the performance of binary classification model. They measure how much better one can expect to do with the predictive model. It also helps to find the best predictive model among multiple challenger models. The main intention behind this paper is to assess the performance of the binary classification model and compares the results with the random pick. It shows the percentage of gains reached when considering a certain percentage of the data set with the highest probability to be target according to the classifier. This paper proposes a broad look at the ideas of cumulative gains chart and lift chart to develop a binary classifier model quality which can be used theoretically to evaluate the quality of a wide range of classifiers in a standardized fashion. This paper proposes a hybrid solution of image hiding binary classifier using vicinity value based image hiding classification model as main complimented by gain calculation to increase image hiding classification accuracy.
The study has shown that implementing the image hiding binary classification using Gain and Lift is feasible. Experiment of the study has confirmed that the image hiding binary classification model can be improved.
***Keywords*: ROC, binary matrix, AUC**

I. INTRODUCTION

Confusion matrix [2,3] is a specific table layout that allows visualization of the performance of our system, typically a supervised learning one. Each row of the matrix represents the instances in a predicted class while each column represents the instances in an actual class. This matrix makes it easy to see if the system is confusing two classes. To test our strategy, we need to compare the output of the model to the actual results in the real world. This is done by comparing the results and creating a contingency table of misclassification errors. It is a very important decision making tool when we evaluate the quality of the model.

The Gains and the Lift curve [1,3] are visualizations showing overall performance of the models. The shape of the curves will tell us a lot about the behavior of the model. It clearly shows how much our taken three cases are better than case assigning categories randomly and how far we are from the optimal model which is in practice unachievable. These curves can help in setting the final cut-off point for deciding which percentage of data set reaches the gain value. The cut-off point will define who should be targeted by the campaign and who should not be. It is very important to mention that ROC, Gains or Lift charts are connected only by one predicted category. There are also analogical Gains and ROC charts that represent the negative response as well. If the main goal for prediction was finding the images to hide with negative response, the criterion for quality of the model would be rather Gains or ROC curve for negative response category.

II. APPRAISAL

	Low	High	
Low	4	1	5
High	7	21	28
	11	22	33

Fig 1(a) Confusion matrix for Less_White

	Low	High	
Low	6	4	10
High	0	3	3
	6	7	13

Fig 1(b) Confusion matrix for More_White

	Low	High	
Low	8	5	13
High	0	10	10
	8	15	23

Fig 1(c) Confusion matrix for No_White

Fig.1Confusion matrix for three cases.

The fig.1(a) deals with the confusion matrix for Less White case. Here the system judged that out of 11 positive valued relevant 7 were high valued relevant. And also, out of 22 negative valued relevant 1 is low valued relevant. The precision for this case is 0.8 and the recall is 0.4. the accuracy for this case is 0.76. The classifier for this case is 0.76 can be calculated using diagonal. This classifier is used to visually inspect for prediction errors, as they will be represented by values outside the diagonal.

The fig.1(b) deals with the confusion matrix for More White case. Here the system judged that all the 8 positive valued relevant were fully positive. And also, out of 15 negative valued relevant 5 is low valued relevant.

The precision for this case is 0.6 and the recall is 0.1, the accuracy for this case is 0.78. The classifier for this case is 0.78 can be calculated using diagonal.

This classifier is used to visually inspect for prediction errors, as they will be represented by values outside the diagonal.

In any binary classification, the ML/DL based methods return a probability, then converted to a prediction. To convert this probability into the actual prediction, a certain threshold has to be chosen.

If the probability >= threshold, the object is hided in an exact manner else the object is not hide in an exact manner. If the obtained probability is greater than 0.5, the prediction is a 1 i.e the image is hided in good manner, otherwise it's a 0 i.e the image is hided in not good manner.

III. GAIN CHART

This Gain chart [1,3] is used to study the probability of a positive and negative reaction. Here we can sort the population according to the probability of a positive reaction to the execution and run the execution only for a percentage of items with highest probability of image hiding. The below chart and table clearly demonstrates the Gains chart for three cases, namely, Less White image hiding, More_White image hiding and Nonwhite image hiding.

The choice of the percentage to be targeted in the campaign depends on the concrete costs for the campaign and gain from the expected positive responses.

The Gains chart is a display of the expected results based on the choice of the percentage targeted. Our final strategy therefore consists of the model and the targeted percentage.

In the above figure 1(a) clearly depicts the gain chart for less white case, the first 10% of the data set gained 10% of observations with the highest probability are located*;* 20% of the dataset gained 40% of observations with the highest probability are located; 30% of the dataset gained 70% of observations with the highest probability are located; 40% of the dataset gained 80% of observations with the highest probability are located*;* 60% of the dataset gained 90% of observations with the highest probability are located*;* 80% *of* the dataset gained 100% *of* observations with the highest probability are located*;*

The above figure 1(b) clearly illustrates the More_White image hiding, which means, the first 10% of the data set gained 20% of observations with the highest probability are located; 20% of the dataset gained 30% of observations with the highest probability are located; 30% of the dataset gained 50% of observations with the highest probability are located*;* 50% of the dataset gained 80% of *observations* with the highest probability are located; 60% of the dataset gained 100% of observations with the highest probability are located*;*

TABLE I-VI GAIN FOR 3 CASES OF IMAGE HIDING

#	C	Score
1	P	0.9
2	N	0.95
3	P	0.8
4	P	0.85
5	P	0.8
6	P	0.78
7	P	0.76
8	P	0.74
9	P	0.7
10	N	0.68
11	N	0.64
12	N	0.61
13	P	0.59
14	N	0.56
15	N	0.53
16	N	0.51
17	N	0.49
18	N	0.47
19	P	0.46
20	N	0.45
21	N	0.44
22	N	0.43
23	N	0.42
24	N	0.39
25	N	0.31
26	P	0.23
27	N	0.22
28	N	0.19
29	N	0.15
30	N	0.12
31	N	0.11
32	N	0.04
33	N	0.01

Table I: P,N values with threshold - LW

x	y
0.0	0.0
0.0	0.1
0.1	0.1
0.1	0.2
0.1	0.3
0.2	0.4
0.2	0.5
0.2	0.5
0.2	0.6
0.3	0.7
0.3	0.7
0.3	0.7
0.4	0.7
0.4	0.8
0.4	0.8
0.5	0.8
0.5	0.8
0.5	0.8
0.5	0.8
0.6	0.9
0.6	0.9
0.6	0.9
0.7	0.9
0.7	0.9
0.7	0.9
0.8	0.9
0.8	1.0
0.8	1.0
0.8	1.0
0.9	1.0
0.9	1.0
0.9	1.0
1.0	1.0
1.0	1.0

Table II: Coordinates for Table1

#	C	Score
1	P	0.9
2	P	0.95
3	N	0.8
4	P	0.85
5	N	0.8
6	P	0.78
7	P	0.76
8	P	0.74
9	N	0.7
10	N	0.68
11	N	0.64
12	N	0.61
13	N	0.59

Table III: P,N values with threshold - MW

x	y
0.0	0.0
0.1	0.2
0.2	0.3
0.2	0.3
0.3	0.5
0.4	0.5
0.5	0.7
0.5	0.8
0.6	1.0
0.7	1.0
0.8	1.0
0.8	1.0
0.9	1.0
1.0	1.0

Table IV: Coordinates for TableIII

#	C	Score
1	P	0.9
2	N	0.95
3	N	0.8
4	P	0.85
5	P	0.8
6	P	0.78
7	P	0.76
8	P	0.74
9	P	0.7
10	P	0.68
11	N	0.64
12	N	0.61
13	N	0.59
14	N	0.56
15	N	0.53
16	N	0.51
17	N	0.49
18	N	0.47
19	N	0.46
20	N	0.45
21	N	0.44
22	N	0.43
23	N	0.42

Table V: P,N values with threshold - NW

x	y
0.0	2.9
0.1	1.4
0.1	1.0
0.2	1.4
0.2	1.7
0.3	1.9
0.3	2.1
0.3	2.2
0.4	2.2
0.4	2.3
0.5	2.1
0.5	1.9
0.6	1.8
0.6	1.6
0.7	1.5
0.7	1.4
0.7	1.4
0.8	1.3
0.8	1.2
0.9	1.2
0.9	1.1
1.0	1.0
1.0	1.0

Table VI: Coordinates for TableV

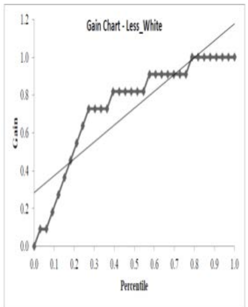

Fig 2(a) Gain chart for Less_White image hide

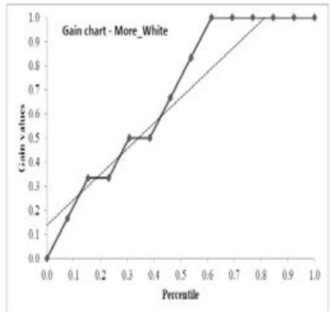

Fig 2(b) Gain chart for More_White image hide

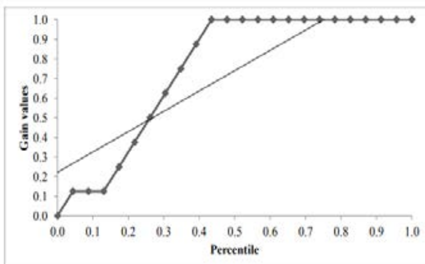

Fig 2(c) Gain chart for No_White image hide

Fig.2 Gain chart for three cases of image hiding

The above figure *1(C)* clearly demonstrates the No_White image hiding, which means, the first 10% of the data set gained 10% of observations with the highest probability are located; 20% of the dataset gained 50% of observations with the highest probability are located*;* 30% *of the dataset* gained 80% *of* observations with the highest probability are located*;* 40% of the dataset gained 100% of observations with the highest probability are located*;*

From the above evaluation, the gain for Nonwhite case reached the higher gain of 100% of the total dataset at the beginning i.e 40%, the More_white case reached the higher gain at 60% and the No_white case reached the higher gain at 80%. Hence the No_white case is a good model wrt gain ratio.

IV. LIFT CHART

It is used to assess the probability of how the three cases is performing, and how well they are identifying the positive (1s or hiding is visible) *or negative (*0s or hiding is perfect*)* instances of our Dataset. The Lift is easily calculated as the ratio of 1s on a certain sample point, divided by the ratio of 1s on the whole dataset, which can also be viewed as the predictions that a random algorithm would be making. The lift curve we can make a cut in a certain population size that is the most likely to react positively. Building a Lift Curve [5] is very easy. First we must sort out the predictions of our cases from highest (closest to 1) to smallest (closest to zero). In this way we have our total dataset ranked by how likely the images are hided. The Maximum Lift is a metric that can be quickly used to compare three cases: the one with the highest Maximum Lift is generally better.

In the figure 2(a), the first point is (0.5, 2.6), second point is (0.35, 2.0), third point is (0.8, 1.3), and the fourth point is (0.95, 0.0). Here the point A (0.5, 2.6) is called the *Maximum Lift point.* If this point is higher, the performance is better, as there is a lot of TP in a proportion of our population which has a very high probability of being positive. The next Point B (0.35, 2.0)**,** we have chosen a sample including 35% of our data set, where we can see that we have more than 3 times more of positive labels than the average. It showed that 35% of the data set had hided in an exact manner and 65% did not. The Point C (0.8, 1.3) **i**ncludes the 80% of our data set with the highest chance of having the image hiding. For this sample the proportion of image hiding is 1.3 times as much as the mean. The Point D (0.95, 0.0) includes the 95% of our data set with the highest chance of having the image hiding. For this sample the proportion of image hiding is 1 time as much as the mean.

In the figure 2(b), the first point is (0.1, 1.8), second point is (0.5, 1.5), third point is (0.85, 1.0), and the fourth point is (1.0, 1.0). Here the Point A (0.1, 1.8) is a lot of TP in a proportion of our population which has a very high probability of being positive. The Point B (0.5, 1.5)**,** have chosen a sample including 35% of our population, where we can see that we have more than 3 times more of positive labels than the average.

It showed that 35% of the data set had hided in an exact manner and 65% did not. **The** Point C(0.85, 1.2)includes the 85% of our population with the highest chance of having the image hiding. For this sample the proportion of image hiding is 1.3 times as much as the mean. The last **Point D** (1.0. 1.0) includes the 100% of our population with the highest chance of having the image hiding. For this sample the proportion of image hiding is 1 time as much as the mean.

In the figure 2(C), the first point is (0.05, 2.1), second point is (0.25, 1.9), third point is (0.65, 1.5), and the fourth point is (1.0, 1.0).Here the Point A (0.05, 1.8) is a lot of TP in a proportion of our population which has a very high probability of being positive. The Point B (0.25, 1.9)**,** have chosen a sample including 25% of our population, where we can see that we have more than 4 times more of positive labels than the average.

It showed that 25% of the data set had hided in an exact manner and 75% did not. **The** Point C(0.65, 1.5)includes the 65% of our population with the highest chance of having the image hiding. For this sample the proportion of image hiding is 1.5 times as much as the mean. The Point D (1.0. 1.0) includes the 100% of our population with the highest chance of having the image hiding. For this sample the proportion of image hiding is 1 time as much as the mean.

x	y
0.0	3.0
0.1	1.5
0.1	2.0
0.1	2.3
0.2	2.4
0.2	2.5
0.2	2.6
0.2	2.6
0.3	2.7
0.3	2.4
0.3	2.2
0.4	2.0
0.4	2.1
0.4	1.9
0.5	1.8
0.5	1.7
0.5	1.6
0.5	1.5
0.6	1.6
0.6	1.5
0.6	1.4
0.7	1.4
0.7	1.3
0.7	1.3
0.8	1.2
0.8	1.3
0.8	1.2
0.8	1.2
0.9	1.1
0.9	1.1
0.9	1.1
1.0	1.0
1.0	1.0

Table VII: Coordinates for Table I

x	y
0.1	2.2
0.2	2.2
0.2	1.4
0.3	1.6
0.4	1.3
0.5	1.4
0.5	1.5
0.6	1.6
0.7	1.4
0.8	1.3
0.8	1.2
0.9	1.1
1.0	1.0

Table VIII: Coordinates for Table III

x	y
0.0	2.9
0.1	1.4
0.1	1.0
0.2	1.4
0.2	1.7
0.3	1.9
0.3	2.1
0.3	2.2
0.4	2.2
0.4	2.3
0.5	2.1
0.5	1.9
0.6	1.8
0.6	1.6
0.7	1.5
0.7	1.4
0.7	1.4
0.8	1.3
0.8	1.2
0.9	1.2
0.9	1.1
1.0	1.0
1.0	1.0

Table IX: Coordinates for Table V

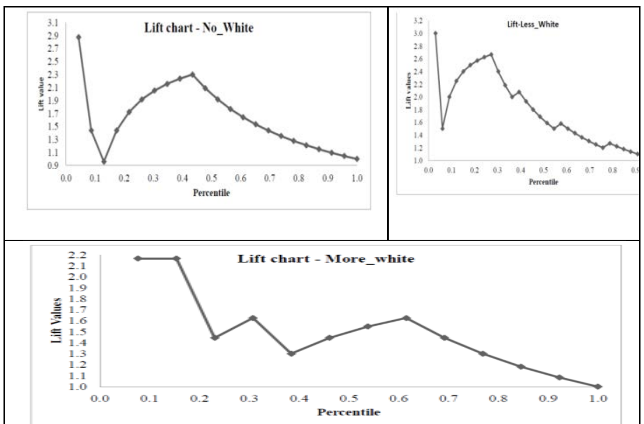

Fig. 3 (a,b,c) Lift curve for Less _White, More_White and No_White image hide

Fig.3 Lift curve for three cases of image hiding

In the above graph there is no ideal lift for Less_White hiding and Nonwhite hiding, because, the higher lift starts from 2 for first case and 2.1 for third case. But for second case, there is an ideal lift since lift starts from two points. In our Less White case, the class distribution is 11 for positive, and 22 for negative; so: 33,3% (11.33) is positive, for More White and Nonwhite case the positive rate is 0.46 (6/13) and 0.53 (8/15) respectively. Here, the first best case is Nonwhite, second rank for More_Whie and lest for Less White case. Here the best cut point value **is** not a unique and relies on the case needs, and is constrained by the rate of FP and FN.

V. CONCLUSION

So, we were able to observe that the Cumulative Gain and Lift charts are useful tools for decision making and binary classification evaluation. They are relatively easy to interpret not only by professionals with a technical background but by others as well. Eventually, both the charts can be applied in various fields, which are market segment targeting, financial budgeting, human resource evaluation, etc.Just like every other evaluation metric lift charts aren't an one-off solution. But they help to get a better picture of the overall performance of our classification model. We can quickly spot flaws, if the slope of the lift chart is not monotonic. Additionally, it helps to set a threshold, which users are worth targeting. Further it can be extended into the model learning process.

REFERENCES

[1] C. Ferri, P. A. Flach, and J. Hernández-Orallo, "Learning decision trees using the area under the ROC curve," in Proceedings of the 19th International Conference on Machine Learning (ICML '19), pp. 139–146, Sydney, Australia, July 2019.

[2] Dr.S.Thabasu Kannan and Ganesh, "Analysis on Enhancing the Effectiveness and Efficiency of Multiresolution Watermarking,"*Intnl Journal of Applied Engineering Research*,Vol.10,No.55, 2015.

[3] E. Marcade, "Evaluating modeling techniques," Technical Whitepaper, KXEN, Inc., San Francisco, Calif, USA, 2018.

[4] Dr.S.ThabasuKannan and Ganesh, " A Capable System for Hiding Image in an Image," *Asian Journal of Engineering and Applied Technology* , Vol.6,No.2, pp.14-17, 2017.

[5] G. Piatetsky-Shapiro and S. Steingold, "Measuring lift quality in database marketing," ACM SIGKDD Explorations Newsletter, Vol. 2, No. 2, pp. 76–80, 2020.

[6] Dr.S.ThabasuKannan and Ganesh, "Design of Efficient Method for Multiresolution Watermarking Algorithm to improve the Robustness," *Intnl Research Journal*, Vol.5, No.1, 2015.

[7] Dr.S.ThabasuKannan and Ganesh, "A Dynamic and Interactive System for Text hiding Using Visibility," *Journal of Analysis and Computation*, Vol.6, No.12, 2020.

[8] S.Selvam and Dr.S.Thabasu Kannan,"Image Retrieval Optimization With Genitic Algorithm", published IJAER ,*International Journal of Applied Engineering Research as Special issue*,Vol.10, No. 55(2015) and IJAER is indexed by SCOPUS, And also listed in Anna University Chennai Annexure II-2014(Sl.No.8565).

[9] F.Long,H.Zhang and H.Dagan,"Fundamentals of content based image retrieval," in D.Feng,W.Siu,H.Zhang(Eds.): "Color Image Retrieval and Management. Technological Fundamentals and Applications,"*Multimedia Signal Processing Book*, Chapter1, Springer-Verlag, BerlinHeidelberg NewYork,2013, pp.1-26.

[10] S.Selvam and Dr.S.Thabasu Kannan," A New Technique for Color-Based Image Retrieval System Using Histogram", published IJAER ,*International Journal of Applied Engineering Research as Special issue*, Vol. 10, No. 82(2015), ISSN 0973-4562 and IJAER is indexed by SCOPUS.

[11] S.Selvam and Dr.S.Thabasu Kannan,"An Empirical Review on Enhancing the Robustness of Multi resolution Water Marking", published IJAER ,*International Journal of Applied Engineering Research as Special issue*, Vol.10, No. 82(2015), ISSN 0973-4562 and IJAER is indexed by SCOPUS.