Optimal bin number for histogram binning method to calibrate binary probabilities.

Tetyana Honcharenko[1], Olga Solovei[2]

[1] *Kyiv National University of Construction and Architecture, Povitroflots'kyi Ave, 31, Kyiv, 03037, Ukraine.*
[2] *Kyiv National University of Construction and Architecture, Povitroflots'kyi Ave, 31, Kyiv, 03037, Ukraine.*

Abstract

A receiver operating characteristic (ROC) curve analysis is an important instrument while selecting the best for the given dataset a classification algorithm by comparing the area under ROC curve. It is applied in areas: medicine, finance and e-commerce, information retrieval, quality control. However, the accuracy of area under ROC curve imposing on predicted probabilities a threshold >0.5, therefore when predicted probability is calculated with different logic then corresponding area under ROC curve is affected and ROC curve analysis's results are misleading. To guarantee the accuracy of area under ROC curve, predicted probabilities must be calibrated. The subject matter of the article is "fixed-width binning" method which is used to calibrate binary predicted probabilities of machine learning algorithms Naive Bayes Classifier, Random Forest Classifier. In this paper the focus is put on "fixed-width binning" method which algorithm is based on the constant number of bins. The goal of work is to increase the calibration scores by proposing a method to select bin number depending on simple statistics of binary predicted uncalibrated probabilities. To meet the goal in the research were evaluated the feasibility to use two different approaches for the identification of the optimal bins' number: "rule-based" approach, "estimators-based" approach. The results of conducted experiments identified that often used 10 bins with "fixed-width binning" method is not optimal. Our proposal is to identify bin number dynamically according to "estimators-based" approach which algorithm is described in the paper.

Keywords

Histogram binning, bin number, brier score, expected calibration error, calibration curve.

1. Introduction

ROC graphs in machine learning are used to select the best for the given dataset a classification algorithm by comparing the area under ROC curve and to model the classifier predictions depending on the chosen value of false positive rate [1]. ROC curve analysis is applied in a medical field to evaluate a performance of diagnostic tests as it helps in decision-making regarding test accuracy; in finance and e-commerce - to assess the effectiveness of fraud detection systems; in information retrieval systems- to optimize the trade-off between relevant and non-relevant results. ROC graphs measure the ability of a classifier to produce relative instance scores – the numeric values which represents the degree to which an instance is a member of a class. In work [2] is defined that "a classifier need not produce accurate, calibrated probability estimates; it needs only produce relative accurate scores that serve to discriminate positive and negative instances." However, in the same work is underlined that area under ROC' accuracy is imposing a threshold >0.5, so this metric is not appropriate when classifier doesn't produce calibrated scores. The threshold >0.5 is applied only on probabilities which are predicted by Logistic Regression Classifier, others commonly used binary classification algorithms computes probabilities as: Random Forest Classifier (RFC) – predicts probabilities (further "scores") as fractions

Proceedings ITTAP'2023: 3rd International Workshop on Information Technologies: Theoretical and Applied Problems, November 22–24, 2023, Ternopil, Ukraine, Opole, Poland
EMAIL: iust511@ukr.net (A. 1); solovey.ol@knuba.edu.ua (A. 2);
ORCID: https://orcid.org/0000-0003-2577-6916 (A. 1); https://orcid.org/0000-0001-8774-7243 (A. 2);

of samples in a given class within the set of decision trees in the forest; Naive Bayes Classifier (GaussianNB) - computes the probability that a data point belongs to a particular class based on Gaussian distribution of the features; Support Vector Machine Classifier (SVC) – doesn't provide probabilities, instead for each data point in dataset it produces the distance from it to hyperplane. Therefore, it is recommended to calibrate classifier's scores predicted by the mentioned learners in order to execute ROC curve analysis with a goal to select the best binary classifier for the given dataset.

As the benchmarking method to calibrate binary scores is used "fixed-width binning" method proposed for algorithms RFC and GaussianNB in research [3]. The method described as a "fixed-width binning" where interval [0,1] is partitioned into bins and a number of bins is recommended to be 10. The computational simplicity and ability to measure a calibration error with "fixed-width binning" made it often used [4-6].

However, in study [7] is concluded that a binning method is effective with properly selected bin's width, as depending on dataset characteristics predicted scores are differently distributed through bins and too small or big number of bins could result calibrated scores are either "over-detailed" or "over-smoothed".

The current research objectives are to empirically study whether the simple statistics of uncalibrated predicted binary scores can be used to choose optimal number of bins for the "fixed-width binning" method and to propose the solution approach. To achieve the research's objectives will be considered the simple statistics of the predicted scores: a range, a standard deviation and an interquartile range and bin size estimators: David W. Scott rule, Freedman-Diaconis rule as their results are directly proportional to a standard deviation and an interquartile range correspondingly.

2. Study Research

2.1. Related literature review

To improve "fixed-width binning" method results in work [8] was proposed a scaling binning method – the algorithm divides data into two subsets - the 1st subset is calibrated by the other continuous calibration method such as "Platt calibration"; the 2nd subset is used to choose the bins so that an equal number of points are landed in each bin. The scaling binning method addresses two issues:

- Reduces a calibration error.
- Calculates bin width which is adopted to already calibrated scores.

However, "Platt calibration" method has a few problems [9]:

- It is the most efficient when distortion of predicted scores is sigmoid-shaped.
- It is computationally intensive as is solving a convex optimization problem to find sigmoid function parameters.

In research [10] to identify bin size to construct a histogram for dataset's distribution is proposed to use Freedman-Diaconis rule. However, in statistic theory depending on actual data distribution it is recommended for the identification of bin size and bin numbers the estimators: David W. Scott, Freedman-Diaconis, Sturges rule, Doane formula, Rice Rule and others. For this reason, applying only Freedman-Diaconis rule may not result that the optimal bin size is found.

The related works [11-12] do not recommend a logic to define the number of bins for "fixed-width binning" method to calibrate predicted probabilities so the problem is actual to study.

2.2. Methodology

To achieve the study's goals we will evaluate the feasibility to use two different approaches for the identification of the optimal bins' number: 1) "rule-based"; 2) "estimators-based".

The "rule-based" approach suggests having a set of rules which depending on simple statistics of predicted uncalibrated scores to propose an optimal bins' number to be used with "fixed-width binning" method.

The "estimators-based" approach suggests identifying bins' number by using different estimators and selection the best bins' number to be used with "fixed-width binning" method as a result of evaluation of calibration error.

Feasibility study for "rule-based" approach includes steps: 1) set rules which input parameters are simple statistics of predicted uncalibrated scores; 2) specify the expected results; 3) execute the rules and record the actual results; 3) compare actual and expected results: if actual and expected results are the same then recommend a "rule-based" approach.

Feasibility study for "estimators-based" approach include steps: 1) calculate bins' number using selected estimators; 2) calibrate predicted scores with "fixed-width binning" method and bins' number received from step 1; 3) compare actual and expected results: if the minimum calibration error does not correspond to bin's number equal to 10 then propose a "estimators-based" approach.

2.3. Materials

The following rules to be included in feasibility study for "rule-based" approach:

1. The 1st rule's clause: given predicted scores have a low standard deviation (less than 0.1) and scores' variability is small (<0.5) and the 2nd rule's clause: given predicted scores have a low standard deviation (less than 0.1) and scores' variability shows a degree of dispersion (from 0.5 to 0.7). The expected results: when David W. Scott rule calculates the required bins number for the 1st rule's clause and the 2nd rule's clause then the results are closed as David W. Scott rule is not expected to care about scores' range.

2. Given predicted scores have a low standard deviation and a low interquartile range (less than 0.1) and scores' variability shows a degree of dispersion (from 0.5 to 0.7). The expected results: when David W. Scott rule calculates the required bin number and Freedman-Diaconis rule calculates the required bin number then the results are closed as both estimators are not expected to care about scores' range.

In case, the actual results from execution of the rules 1-2 look the same as the expected then our recommendation will be to develop "rule-base" approach as stable rules can be specified based on simple statistic of the predicted binary scores.

The following formulas and algorithms to be included in feasibility study for "estimators-based" approach.

The predicted binary scores are considered as well calibrated when its values are closed or equal to actual value of target class $y \in \{0,1\}$. Brier score estimates the calibration's error as the mean squared error of the actual target class y_i and predicted binary score s_i (1) [13]. The lower value of Brier score the better calibration results.

$$BS = \frac{1}{n}\sum_{i=0}^{n-1}(y_i - s_i)^2 \tag{1}$$

where n – is the number of observations in dataset.

The lower Brier score does not always mean a better calibration, the reason for it is the bias-variance decomposition of the mean squared error [14]. Other approach to measure calibrations is to calculate expected calibration error (ECE) (2) [15].

$$ECE = \frac{1}{n}\sum_{i=1}^{B} h_i \cdot |\overline{y_i} - \overline{s_i}| \tag{2}$$

where $\overline{y_i} = \frac{1}{|y|}\sum_{j \in h_i}^{B} y_j$ - the mean of the actual target class values which are belonging to a single bin with a width is h_i; $\overline{s_i} = \frac{1}{|s|}\sum_{j \in h_i}^{B} s_j$ - the mean of calibrated scores which are belonging to a single bin with a width is h_i; B – number of bins.

A Calibration curve plots the calibration results as the relationship between the mean predicted binary scores s_i in each bin, placed on x-axis and fraction of actual values of target class in each bin – placed on y-axis. The closer a calibration curve to diagonal line the better calibration [16].

When bin's width (h) is identify using David W. Scott (further Scott) estimator (3) then it makes bin's width to be proportional to dataset standard deviation (σ) and inversely proportional to the number of observations in dataset (n) [17].

$$h = \frac{3.49 \cdot \sigma}{\sqrt[3]{n}} \tag{3}$$

When bin's width (h) is identify using Freedman-Diaconis estimator (4) then it makes bin's width to be proportional to interquartile rate (IQR) and inversely proportional to the number of observations in dataset (n) so the calculated width is optimal when dataset is normally distributed but contains outliers [18].

$$h = \frac{2 \cdot IQR}{\sqrt[3]{n}} \tag{4}$$

The required bin's number (b) for the calculated bin's width to be derived as a fraction between scores' range and bin's width (h) according to equation (5).

$$b = \frac{\max(scores) - \min(scores)}{h} \tag{5}$$

Algorithm 1 specifies a "fixed-width binning" method to calibrate predicted binary scores for a given constant number of bins, noted as "n_bins". In lines 6-8 it calculates bins' edges. In lines 9-14 for each uncalibrated score received as input parameter is calculated its index of bin, denoted as "score_inx". In lines 16-18: the algorithm iterates through bins indices, finds scores' which indices coincide with current bin's index, those score's indices are denoted as "mask" and calculates calibrated score as an average of scores inside the bin.

Algorithm 1. "fixed-width binning" method to calibrate binary scores

```
Input: n_bins, scores
Output: calibrated_scores
1.      calibrated_scores=zero array of size score
2.      score_inx=zero array of size score
3.      start =0
4.      stop = 1
5.      h=(stop-start)/n_bins
6.      For i in range(|n_bins|)
7.          bin_edges=start+i·h
8.      EndFor
9.      For i in range(|scores|)
10.       For j in range (n_bins)
11.         If bin_edges [j]<=scores[i]< bin_edges [j+1]
12.           score_inx[i]=j
13.         break
14.       EndFor
15.     EndFor
16.     For  i in range(|n_bins|)
17.       mask= score_inx==i
          calibrated_scores[mask]=sum(calibrated_scores[mask])/len(mask)
18.     EndFor
```

Algorithm 2 specifies logic to identify an optimal bin's number to be used as input parameter for "fixed-width binning" method to calibrate predicted binary scores. In line 6 bin's width is calculated by estimator and in line 8 predicted scores are calibrated by algorithm 1 which is called with bin's number calculated in line 7. In lines 9-10 the calibration results are evaluated by Brier score and expected calibration error and marks are saved in arrays. The lines 6-10 are repeated for each estimator. In lines 12-16 the optimal bin number is selected where the lower values of metrics' marks is identifier. If metrics' marks disagree, then the default bins' number equal to 10 is returned.

In case, the minimum calibrations error, which is received from the execution of algorithm 2 is for calibration with bins' numbers different from 10 bins, then our recommendation will be to develop "estimators-base" approach.

Algorithm 2. "Estimators-based" approach to identify optimal bin numbers for "fixed-width binning" to calibrate binary scores

```
Input: scores, estimators
Output: bin number
1.      bin_number=10;
2.      n_bins = ece = brier = zero array of size estimators.
3.      σ=standard deviation (scores)
4.      IQR = interquartile rate (scores)
5.      For i in range (|estimators|)-1
6.         h=call of estimator[i](σ, IQR)
7.         n_bins[i]= formula 5
8.         calibrates scores = algorithm1(n_bins [i], scores)
9.         brier[i]=formula 1
10.        ece[i]=formula 2
11.     EndFor
12.     For i in range (|estimators|)-1
13.        If Brier[i]<= Brier[i+1] and ece[i]<= ece[i+1] Then
14.          bin number= n_bins[i]
15.          Break
16.     EndFor
```

2.4. Experiments

The execution of feasibility study for "rule-based" approach consists of:

Step1. Calculate bin's width using estimator (3-4) and bins' number according to (5) with input parameters:

Rule 1: $\sigma \leq 0.1$ and n = 250 and scores range ≤ 0.1.

Rule 2: $\sigma \leq 0.1$ and n = 250 and 0.5< scores range ≤ 0.7.

Rule 3: IQR ≤ 0.1 and n = 250 and 0.5< scores range ≤ 0.7.

Step2. Compare actual results and expected results (specified in sec. "Materials"), make the recommendations.

The execution of feasibility study for "estimators-based" approach consists of three steps:

Step1. Generate two synthetic datasets for classification problem: the 1st dataset is from skewed Gaussian distribution; the 2nd – from normal distribution. The size of datasets is: two features and 1000 observations; a target class values are 1 and 0 for positive and negative class correspondingly. The machining learning algorithms which are included in the experiments are RandomForestClassifier and GaussianNB to be used with default values for hyperparameters.

Step2. Split dataset on a train and test subset in proportion 80/20; train a learner and receive predicted binary scores for test subset.

Step3. Execute algorithm 2, compare actual results and expected results (specified in sec. "Materials"), make the recommendations.

2.5. Results and discussions

The actual results from the execution of the rules 1-3 to study the feasibility to use "rule-based" approach are:

1. when $\sigma \leq 1$ and n = 250 and scores range ≤ 0.5 then estimator (3) will make bins' width less than 0.0349 and number of bins is 29;
2. when $\sigma \leq 1$ and $n = 250$ and 0.5<scores range ≤ 0.7 then estimator (3) will make bins' width less than 0.08 and number of bins is 5;

3. when $IQR \leq 1$ and $n = 250$ and 0.5< scores range ≤ 0.7 then estimator (4) will make bins' width less than 0.01 and number of bins is 60.

The actual results which are obtained from the execution of the rules 1-2 show that small changes in scores' variability may impact David W. Scott's rule's result so that calculated numbers of bin differ appx. by a factor of 6 which is not expected as standard deviation and number of observations are similar low in rules 1-2. The actual results which are obtained from the execution of the rules 2-3 show that both estimators (3-4) calculate numbers of bin which differ by a factor of 12 which is not expected as we kept low standard deviation and IQR, so expecting the similar results from both estimators.

To summarize the results, we won't recommend "rule-based" approach as stable rules cannot be defined based on the selected simple statistic of the predicted binary scores.

The results of the execution of feasibility study for "estimators-based" approach is recorded in tables 1-6 and illustrated on figures 1-2.

Table 1 records simple statistics for predicted by GausianNB and RandomForestClassifier uncalibrated binary scores in lines 1 and 2 correspondingly. The learning algorithm had been trained on the skewed dataset from Gaussian distribution. Line 1 recodes standard deviation of uncalibrated predicted scores and IQR are less than 0.1, as specified in scenario 1, in Table 2 is visible that the number of bins tends to be bigger than 10 – it is 60 and 14 bins. Line 2 records increased spreads and in Table 3 is visible that the number of bins tends to be smaller than 10 – it is 9 and 8.

Calibration results for scores which statistics are described in Table1 is presented in Table 2-3 in columns: "Brier score", "ECE" and visualized with calibration curves on Figure1. For calibrated GausianNB scores the lower values of metrics are captured for 60 bins and calibration curves on picture (b) from Figure 1 and line 1, shows that half from total curve's points are very closed to diagonal. On picture (c) from Figure 1 and line 1 can also be seen that half of the point are closed to diagonal and Brier score is almost the same, however according to ECE metric – the better calibration is achieved with 60 bins. For calibrated RandomForestClassifier's scores the lower values of Brier score is captured for binning with 8 bins. The difference in ECE metric between 8 bins and 9 bins binning is less than 10^{-4}, however, calibration curve on picture (c) from Figure 1, line 2 shows more curve's points are closed to diagonal line compared to curve on picture (b) which indicate the better calibration with 8 bins.

Table 1

Simple statistics of predicted scores when learning algorithm is trained on dataset from skewed Gaussian distribution.

Estimator	Predicted scores counts per bin	Scores range	σ	IQR
GausianNB	2, 2, 4, 7, 18, 53, 132, 19, 5, 8	0.66	0.086	0.035
RandomForestClassifier	8, 12, 32, 28, 43, 40, 30, 26, 22, 9	0.89	0.2	0.3

Table2.

Bin numbers to be used with Algorithm 1 to calibrate scores predicted by GaussianNB trained on dataset from skewed Gaussian distribution.

Estimator	Bin width	Bin number	Brier score	ECE
Not applied	0.1	10	0.2619	$6*10^{-4}$
Freedman-Diaconis	0.011	60	**0.2614**	**$3*10^{-4}$**
Scott	0.047	14	0.2615	$5*10^{-4}$

Table3.
Bin numbers to be used with Algorithm 1 to calibrate scores predicted by RandomForestClassifer trained on dataset from skewed Gaussian distribution.

Estimator	Bin width	Number	Brier score	ECE
Not applied	0.1	10	0.2853	$6*10^{-4}$
Freedman-Diaconis	0.09	9	0.2812	$5*10^{-4}$
Scott	0.113	8	**0.2795**	$\mathbf{5*10^{-4}}$

Table 4 records simple statistics for predicted by GausianNB and RandomForestClassifier uncalibrated binary scores in lines 1 and 2 correspondingly. The learning algorithms had been trained on the dataset from normal distribution. Calibration results for scores from Table 4 are presented in Table 5-6.

As in line 1 and line 2 from Table 1 is seen increased compared to Table 1 in the spread of uncalibrated predicted scores from mean and median, so the result in Table 5 record the number of bins is closed to 10 – it is 12 and 8. In Table 6 due to more increased spread we see 6 bins are needed.

For calibrated GausianNB scores the lower values of Brier score is captured for default 10 bins, however the difference in ECE metric is less than 10^{-4}, and calibration curves on the picture (c) from Figure 2, line 1 shows 5 from 8 curve points are on diagonal line so 8 bins could be considered as optimal as well. For calibrated RandomForestClassifier scores the lower values of metrics are captured for 6 bins and calibration curves on pictures (b)-(c) from Figure 2, line 2 indicate better calibration compared to curve with 10 bins on picture (a) from Figure 2, line 2.

Table 4.
Simple statistics of predicted scores when learning algorithm is trained on dataset from normal distribution.

Estimator	Predicted scores counts per bin	Scores range	σ	IQR
GausianNB	115, 49, 20, 18, 13, 9, 10, 3, 8, 5.	0.65	0.15	0.18
RandomForestClassifier	25, 30, 20, 32, 26, 18, 19, 19, 20, 41.	0.99	0.3	0.55

Table5.
Bin numbers to be used with Algorithm 1 to calibrate scores predicted by GaussianNB trained on dataset from normal distribution.

Estimator	Bin width	Number	Brier score	ECE
Not applied	0.1	10	**0.2409**	$\mathbf{1*}10^{-4}$
Freedman-Diaconis	0.05	12	0.2424	$2*10^{-4}$
Scott	0.081	8	0.245	$1*10^{-4}$

Table6.
Bin numbers to be used with Algorithm 1 to calibrate scores predicted by RandomForestClassifer trained on dataset from normal distribution.

Estimator	Bin width	Number	Brier score	ECE
Not applied	0.1	10	0.2008	$5*10^{-4}$
Freedman-Diaconis	0.165	6	**0.1961**	$\mathbf{3*}10^{-4}$
Scott	0.165	6	**0.1961**	$\mathbf{3*}10^{-4}$

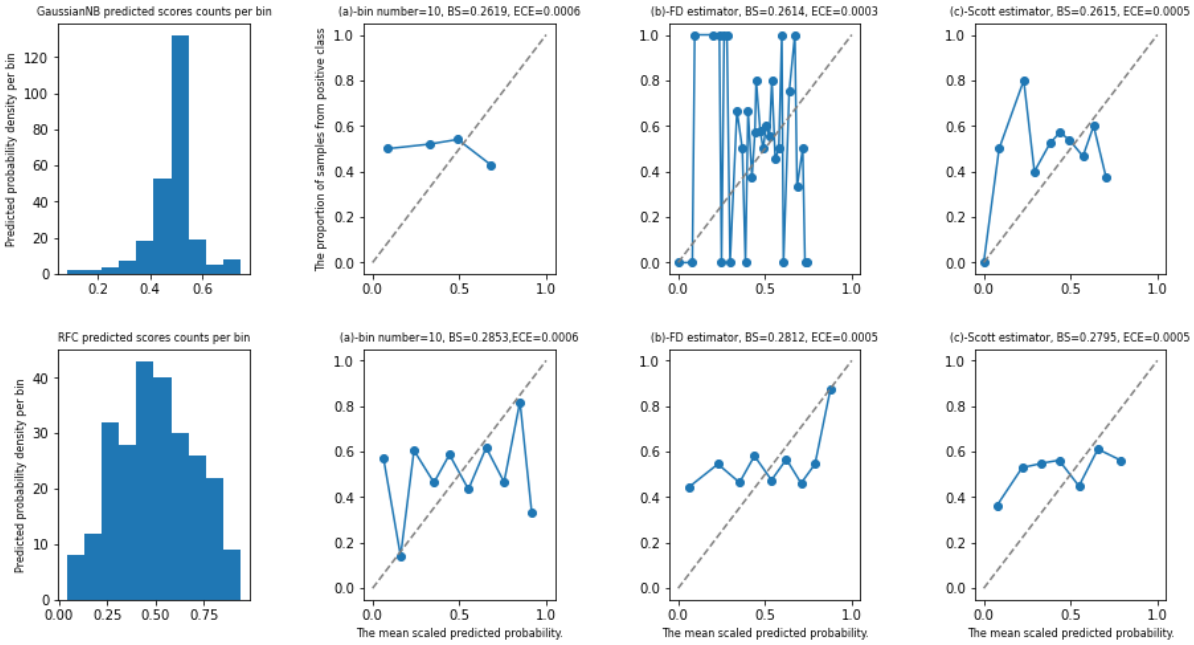

Figure 1: Calibration curves of the scaled scores predicted by GaussianNB and RandomForestClassifier for dataset from skewed Gaussian distribution.

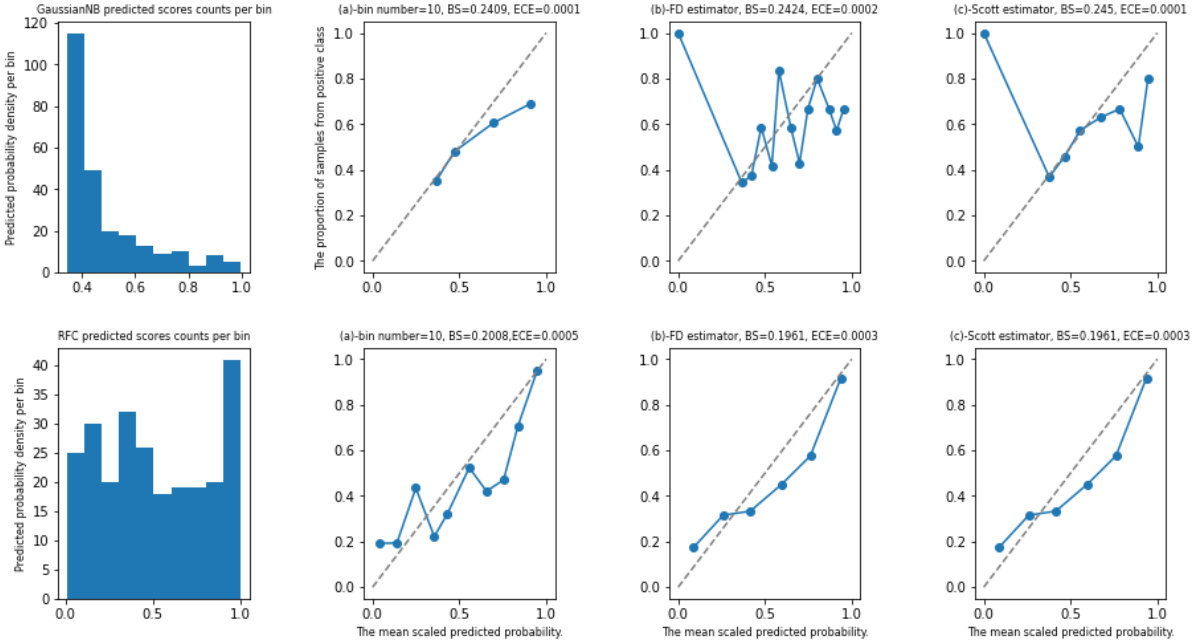

Figure 2: Calibration curves of the scaled scores predicted by GaussianNB and RandomForestClassifier for dataset from normal distribution.

3. Conclusions

In current study had been considered two different approaches for identification of the optimal bin number to be used with "fixed-width binning" method. The "rule-based" approach – according to which a set of rules depending on the values of uncalibrated scores' range, standard deviation and an interquartile range will propose an optimal bin number had not been recommended to further development as the actual results from the rules execution compared to expected results showed that the small changes in scores' variability may impact David W. Scott's rule results and David W. Scott's rule and Freedman-Diaconis rule calculated the numbers of bin which differ by a factor of 12 which was not expected as we fixed to be low standard deviation and an interquartile range. Our conclusion

for “rule-based” approach is that stable rules can’t be defined based on the selected simple statistic of the predicted binary scores.

The effectiveness of “estimators-based” approach - according to which bins’ number is calculated by different estimators and the optimal bin number is selected as a result of the evaluation of a calibration error revealed the following: 10 bins for “fixed-width binning” method to calibrated predicted probabilities is not optional for all datasets. When uncalibrated score’s range, standard deviation and interquartile range are low then 60 bins can be optimal according to Freedman-Diaconis rule, at the same when scores’ spread is low, however a standard deviation and an interquartile range are increasing then 8 bins can be optimal according to David W. Scott’s rule. Further increase of spread will cause optimal bin number is decreased to 6 bins according to both estimators.

Our proposal is to identify bin number dynamically according to “estimators-based” approach which is described per algorithm 2. The proposed approach will improve calibrations of binary predicted probabilities based on ECE and Brier score metrics as visible from calibration curves on Figure 1 and Figure 2 so that the accuracy of area under ROC curve is good to conduct ROC curve analysis.

Further work will be to extend the proposed “estimators-based” approach to calculate optimal bin number with estimators: Sturgers’ formula, Rice rule, Doane’s formula as those estimators considers bin numbers based on the range of the data.

4. References

[1] T. Fawcett, An introduction to ROC analysis. Pattern recognition letters. 2006 Jun 1;27(8):861-74.

[2] T. Fawcett, ROC graphs: Notes and practical considerations for researchers. Machine learning. 2004 Mar 16;31(1):1-38.

[3] B. Zadrozny, C. Elkan, Obtaining calibrated probability estimates from decision trees and naive bayesian classifiers. InIcml 2001 Jun 28 (Vol. 1, pp. 609-616).

[4] M. Naeini, G. Cooper, M. Hauskrecht. Obtaining well calibrated probabilities using Bayesian binning. In AAAI Conference on Artificial Intelligence, 20

[5] C. Guo, G. Pleiss, Y. Sun, KQ. Weinberger. On calibration of modern neural networks. In International conference on machine learning 2017 Jul 17 (pp. 1321-1330).

[6] R. Roelofs, N. Cain, J. Shlens, M. Mozer. Mitigating bias in calibration error estimation. arXiv preprint arXiv:2012.08668, 2020

[7] C. Gupta, A. Ramdas, Distribution-free calibration guarantees for histogram binning without sample splitting. In International Conference on Machine Learning 2021 Jul 1 (pp. 3942-3952). PMLR.

[8] A. Kumar, P. Liang, M. Tengyu, Verified uncertainty calibration. Advances in Neural Information Processing Systems. 2019;32.

[9] J. Platt, Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. Advances in large margin classifiers. 1999 Mar 26;10(3):61-74.

[10] Jiang, T., Luo, S., Wang, D., Li, Y., Wu, Y., He, L. and Zhang, G., 2023. A new bin size index method for statistical analysis of multimodal datasets from materials characterization. Scientific Reports, 13(1), p.10915.

[11] KH. Knuth, Optimal data-based binning for histograms and histogram-based probability density models. Digital Signal Processing. 2019 Dec 1;95:102581.

[12] WK. Leow, R. Li, The analysis and applications of adaptive-binning color histograms. Computer Vision and Image Understanding. 2004 Apr 1;94(1-3):67-91.

[13] G. Brier, Verification of forecasts expressed in terms of probability. Monthly weather review 78 (1950) 1–3.

[14] A. Bella, C. Ferri, J. Hernández-Orallo, MJ. Ramírez-Quintana, Calibration of machine learning models. InHandbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques 2010 (pp. 128-146). IGI Global.

[15] M.P. Naeini, G. Cooper, and M. Hauskrecht, Obtaining well calibrated probabilities using bayesian binning. In: Proceedings of the AAAI Conference on Artificial Intelligence, 2015.

[16] DS, Wilks, On the combination of forecast probabilities for consecutive precipitation periods. Weather and forecasting. 1990 Dec;5(4):640-50.

[17] DW. Scott, Sturges' rule. Wiley Interdisciplinary Reviews: Computational Statistics. 2009 Nov;1(3):303-6.
[18] D. Freedman D, P. Diaconis, On the histogram as a density estimator: L 2 theory. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete. 1981 Dec;57(4):453-76.