Journal of Physics: Conference Series

PAPER • OPEN ACCESS

Evaluation performance recall and F2 score of credit card fraud detection unbalanced dataset using SMOTE oversampling technique

To cite this article: B Prasetiyo *et al* 2021 *J. Phys.: Conf. Ser.* **1918** 042002

Evaluation performance recall and F2 score of credit card fraud detection unbalanced dataset using SMOTE oversampling technique

B Prasetiyo[1,*], Alamsyah[1], M A Muslim[1], N Baroroh[2]

[1] Department of Computer Science, Faculty of Mathematics and Natural Sciences, Universitas Negeri Semarang, Indonesia
[2] Department of Accounting, Faculty Economy, Universitas Negeri Semarang, Indonesia

*Corresponding author: bprasetiyo@mail.unnes.ac.id

Abstract. Unbalanced data becomes an interesting research and continues to be studied because of its uniqueness. Unbalanced data requires special treatment prior to making the data balance. In this paper, our study to investigate the performance of unbalanced dataset using diverse oversampling proportion. We use SMOTE to gerentae new syntethic data, then we classify using random forest algorithm. In our experiment we generate new sampling with start 20%, 40%, 60%, 80%, and 100% of majority class, so that the data balancing until 50%: 50%. Each new generated data, we train the data using classification technique. Then, evaluate each algorithm performance. We show that the highest F2 score i.e: 85.34 and 84.93. The new data generated is 60% of majority class, result F2 score 85.34, then the new data generated from 100% of majority class result F2 score 84.93.

1. Introduction

Data mining has been widely used in various fields, ranging from education [1,2], health [3,4], marketing [5], to the economy [6]. Several studies in data mining, including examining how to improve the accuracy of the algorithm, for example in the field of health have been done [7,8]. Meanwhile, implementation in the economy sector, for example [9]. Research studies in the field of economics are quite interesting because, currently, a lot of transaction data or transaction activities are carried out online [10]. Online transactions via e-commerce make it easier in terms of transaction relations between countries. Online transaction activities have broken through time and space boundaries by utilizing internet technology, one of which is increasing export-import activities in China [11,12]. Payment for online transactions is made in non-cash, one of which is a credit card. However, the use of a credit card has risks, for example, when we found a fraud transaction.

Credit card fraud detection has become an interesting study, especially with the use of data mining. Dornadula has conducted a study on the use of machine learning for credit card fraud detection [13,14,15]. The implementation of this problem using on azure ML has been carried out by Shivanna [16]. The credit fraud dataset is the object of much study because of its uniqueness, namely the unbalanced nature of the data. Previously, Pozzolo has conducted research to conduct research on Calibrating Probability on Undersampling for Unbalanced [17]. Unbalanced data is a condition where the number of classes is not balanced. Data imbalances can affect data quality. Data quality can affect classification results. The dataset is unbalanced if the classification categories are not evenly represented [18].

The solution to overcome unbalanced data can be done by undersampling or oversampling. In this paper, we focus on the problem of oversampling, one of which is the Synthetic Minority Over-Sampling Technique (SMOTE). SMOTE simply creates synthetic data based on k nearest neighbors so as to create a balanced class distribution between the majority and minority classes [19]. SMOTE is an oversampling technique that is used to solve the problem of unbalanced datasets by modifying unbalanced datasets and producing a balanced dataset from unbalanced datasets. SMOTE distributes majority and minority class instances evenly thereby increasing the predictive accuracy of the minority class by creating synthetic instances of the minority class without causing overfitting due to synthetic sampling technique [20]. Research for banlancing data on SMOTE includes [21]. In this paper, we conduct a study to determine the classification performance using various class compositions using SMOTE. The data used refers to the credit card fraud detection dataset [17], where the data has an extreme unbalanced class

2. Methods

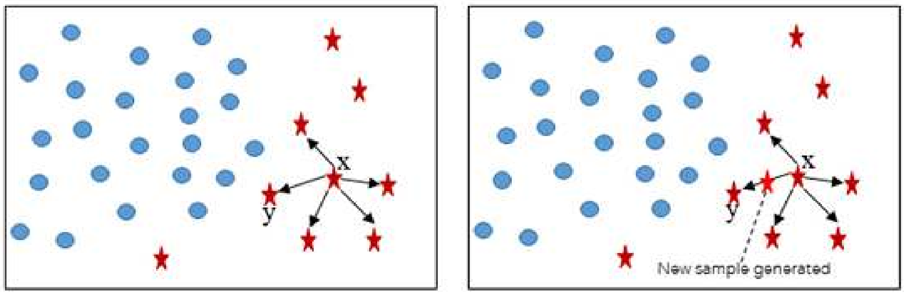
In this study, we performed data preparation: data collecting, data preprocessing. The dataset we use is credit fraud detection [17]. The research data contains transactions made with credit cards in September 2013 by European card holders. This dataset provides transactions that took place in two days. This data contains 492 frauds out of 284,807 transactions, its mean this dataset is extremely unbalanced because the positive class (fraud) has only 0.172% of all transactions. The composition of the class, not fraud compared to fraud was 99.828% : 0.172%. The dataset only contains numeric values, which are the result of PCA transformation, contains 28 attributes, and 1 Class is shown in Table 1.

Table 1. Details of the dataset Features

Features	Features Type
Time	numeric
V1	numeric
V2	numeric
V3	numeric
V4	numeric
V5	numeric
V6	numeric
…	numeric
V27	numeric
V28	numeric
Amount	numeric
Class (label)	categorical

The next step is to perform data preprocessing, which is to normalize the data on the "Amount" attribute using the standard scaller (range -1 to 1), while the other attributes are not. Attribute "Time" is not normalized to see transaction times. Meanwhile, attribute V1 to V8 is the result of PCA transformation. Next, we did the experiment using the SMOTE [18] oversampling technique.

The sample generation of SMOTE illustrate in Figure 1 [22].



(a) (b)

Figure 1. Generating new sample in SMOTE. (a) same K-nearest neighbor samples of the main sample of x (K=5, in this example). (b) New samples generated, see red start symbol.

Next, we conducted experiments with a diverse number of oversampling classes. From low oversampling to balanced oversampling. For each new data from the oversampling results, we conduct training using a classification algorithm. The classification algorithm used is logistic regression. Next, we perform algorithmic performance calculations. The performance in focus is the F2 score and Recall. We chose the F2 score according to the characteristics of the unbalanced data. We use Recall because of the characteristics of the dataset for fraud detection.

3. Results and Discussion

In our experiment dataset [17] show that the credit fraud dataset is extremely unbalanced. We did a visualization to show the proportion of fraud notes compared to fraud shown in Figure 2. In Figure 2 there are two bar charts, the left is Class "not fraud" (284,315 instances), while the right bar chart is "fraud" marked with a red circle (492 instances), with a comparison percentage of 99.828%: 0.172%.

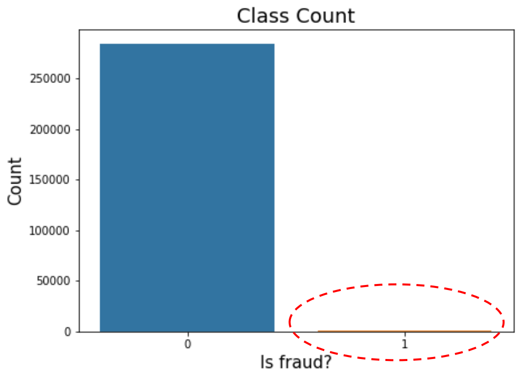


Figure 2. Unbalanced class of no fraud (0) and fraud (1)

We then call the not fraud class the majority class, while the fraud class hereinafter will be called the minority class. In our experiment, we implement diverse oversampling proportional. Then we do classification and matrices the perform. We implement classification in python using sklearn module [23]. In our experiment, we resampled the minority class as x% of majority class as shown in Figure 3.

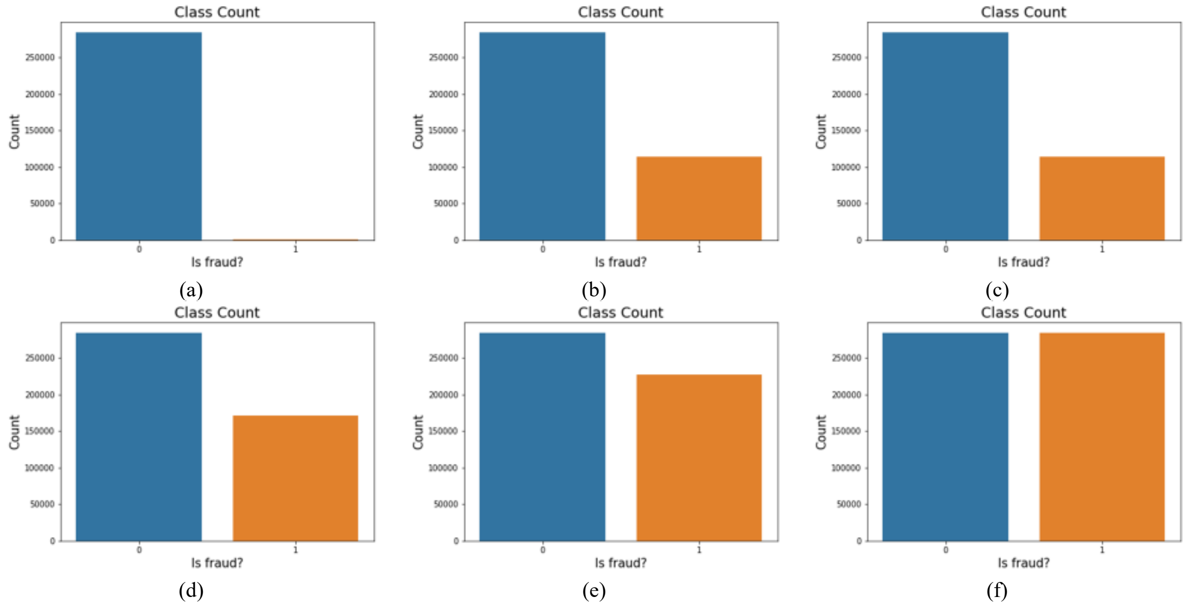


Figure 3. The proportion of new oversampling method using SMOTE. Figure (a) is original data, (b) resampling of 20% from majority class, (c) resampling 40% of majority class, (d) resampling 60% of majority class, (e) resampling 80% of majority clas, (f) resampling 100% of the majority class.

Figure 3 provides 6 figure that visualize diverse resampling proportion. So the percentage proportion of “no fraud : fraud” each figure there are: (a) original data 99.828% : 0.172; (b) 83:17; (c) 71:29; (d) 62:38; (e) 55:44; (f) 50:50. Then each new generated data we classified using Random Forest then metrices the performance provides in Figure 4.

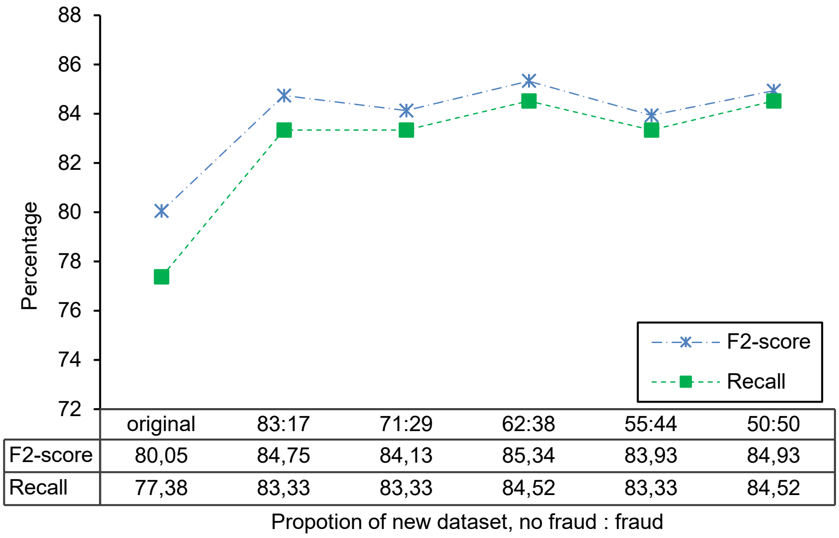


Figure 4. Metric performance using random forest classification

In our experiment, we can show that the recall and f2 performance increase along with the number of classes that are balanced. Figure 4 Provides the graphic that show the performance start from original data until balanced data (50:50). We can show that the recall value is still consistent in the last 3 experiments, amounting to 84.5%; 83.33%; 84.52%. The highest F2 score was in the proportion of data 62:38 amounting to 85.34%, while in the balanced class (50:50), it actually decreased to 84.93%.

4. Conclusion

We have make study to oversampling the extremely unbalanced dataset. We have done the experiment with a diverse proportion of oversampling technique using SMOTE. Then we classified using random forrest classification. In our experiment show that the highest F2 score and Recall in composition 62%: 38%. This oversampling the minority class is to the 80% of the majority class, with a recall value of 84.52 and F2 score of 85.34. Meanwhile, the proportion of 50% : 50% (oversampling as much as 100% of the majority class) was actually lower than 62%:38%, especially in the F2 score class, namely recall of 84.52 and F2 score of 84.93.

References

[1] Baker, R S J D 2010 *Int. encycl. educ.* **7** 112
[2] Koedinger, K R, D'Mello S, McLaughlin E A, Pardos Z A and Rose C P 2015 *Wiley Interdiscip. Rev.: Cogn. Sci.* **6** 333
[3] Yoo I, Alafaireet P, Marinov M, Pena-Hernandez K, Gopidi R, Chang J F and Hua L 2012 *J. med. syst.* **36** 2431
[4] Jothi N and Husain W 2015 *Procedia comput. sci.* **72** 306
[5] Radhakrishnan B, Shineraj G and Muhammed A K M 2013. *Int. J. Comput. Sci. Netw.* **2** 41
[6] Feelders A 2002 *Deal. data flood* 166.
[7] Muslim M A, Herowati A J, Sugiharti E and Prasetiyo B 2018 *In Journal of Physics: Conference Series* **983** 12062
[8] Nurzahputra A, Muslim M A and Prasetiyo B 2019 *J. Phys.: Conf. Ser.* **1321** 32022
[9] Prasetiyo B, Muslim, M A and Baroroh, N 2020 *In Journal of Physics: Conference Series* **1567** 32022
[10] Kim Y H and Kim D J 2005 *Proc. 38th Annu. Hawaii Int. Conf. Syst. Sci.* 170c
[11] Valarezo A, Perez-Amaral T, Garin-Munoz T, Herguera, Garcia L and Lopez R 2018 *Telecommunications Policy* **42** 464
[12] Lin A J, Li E Y and Lee, SY 2018 *J. Electron. Commer. Res.* **19** 36
[13] Dornadula V N and Geetha S 2019 *Procedia Comput. Sci.* 165 p 631
[14] Kumar, A, Anand, K, Jha, S, & Gupta, J 2021 *Data Management, Analytics and Innovation* (Singapore: Springer) p 107
[15] Sadgali I, Sael N and Benabbou F 2019 *Int. Conf. Smart Syst. Data Sci. 2019*
[16] Shivanna A, Ray S, Alshouiliy K and Agrawal D P 2020 *11th IEEE Annu. Ubiquitous Comput. Electron. Mob. Commun. Conf. UEMCON 2020* 268
[17] Pozzolo A D, Caelen O, Johnson R A and Bontempi B 2015 *Symp. Comput. Intell. Data Min. (CIDM), IEEE*
[18] Chawla N V, Bowyer K W, Hall L O and Kegelmeyer W P 2002 *Artif. Intell. Res.* ***16*** 321
[19] He H and Ma Y 2013 *Imbalanced learning: foundations, algorithms, and applications*
[20] Safitri A R and Muslim M A 2020 *J. Soft Comput. Explor.* **1** 70
[21] Jishan S T, Rashu R I, Haque N and Rahman R M 2015 *Decis. Anal.* **2** 1
[22] DuY, Liu Y, Shao Q, Luo L, Dai J, Sheng G and Jiang X 2019 *Trans. Power Deliv.* **34** 1766
[23] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, ... and Duchesnay E 2011 *J. mach. Learn. res.* **12** 2825