Visualization and Analysis in Bank Direct Marketing Prediction

Alaa Abu-Srhan[1], Sanaa Al zghoul[3]
Department of Basic Science
The Hashemite University
Zarqa, Jordan.

Bara'a Alhammad[2], Rizik Al-Sayyed[4]
Computer Science Department
University of Jordan
Amman, Jordan.

Abstract—**Gaining the most benefits out of a certain data set is a difficult task because it requires an in-depth investigation into its different features and their corresponding values. This task is usually achieved by presenting data in a visual format to reveal hidden patterns. In this study, several visualization techniques are applied to a bank's direct marketing data set. The data set obtained from the UCI machine learning repository website is imbalanced. Thus, some oversampling methods are used to enhance the accuracy of the prediction of a client's subscription to a term deposit. Visualization efficiency is tested with the oversampling techniques' influence on multiple classifier performance. Results show that the agglomerative hierarchical clustering technique outperforms other oversampling techniques and the Naive Bayes classifier gave the best prediction results.**

Keywords—Bank direct marketing; prediction; visualization; oversampling; Naive Bayes

I. Introduction

Bank direct marketing is an interactive process of building beneficial relationships among stakeholders. Effective multi-channel communication involves the study of customer characteristics and behavior. Apart from profit growth, which may raise customer loyalty and positive responses [1], the goal of bank direct marketing is to increase the response rates of direct promotion campaigns.

Available bank direct marketing analysis datasets have been actively investigated. The purpose of the analysis is to specify target groups of customers who are interested in specific products. A small direct marketing campaign of a Portuguese banking institution dataset [2], for example, was subjected to experiments in the literature. Handling imbalanced datasets requires the usage of resampling approaches. Undersampling and oversampling techniques reverse the negative effects of imbalance [3], these techniques also increase the prediction accuracy of some well-known machine learning classification algorithms.

Data visualization is involved in financial data analysis, data mining, and market analysis. It refers to the use of computer-supported and interactive visual representation to amplify cognition and convey complicated ideas underlying data. This approach is efficiently implemented through charts, graphs, and design elements. Executives and knowledge workers often use these tools to extract information hidden in voluminous data [4] and thereby derive the most appropriate decisions. The usage of data visualization by decision makers and their organizations offers many benefits [2], that

includes absorbing information in new and constructive ways. Visualizing relationships and patterns between operational and business activities can help identify and act on emerging trends. Visualization also enables users to manipulate and interact with data directly and fosters a new business language to tell the most relevant story.

The choice of a proper visualization technique depends on many factors, such as the type of data (numerical or categorical), the nature of the domain of interest, and the final visualization purpose [5], which may involve plotting of the distribution of data points or comparing different attributes over the same data point. Many other factors play a remarkable role in determining the best visualization technique that can detect hidden correlations in text-based data and facilitate recognition by domain experts.

The current research is an attempt to demonstrate the capabilities of different visualization techniques while performing different classification tasks on a direct marketing campaign. The data set, which contains 4521 instances and 17 features that including an output class, originates from a Portuguese banking institution. The goal is to predict whether a client will subscribe to a term deposit. The data set is highly imbalanced. Some oversampling methods are applied as a preprocessing step to enhance prediction accuracy. Random forest, support vector machine (SVM), neural network (NN), Naive Bayes, and k-nearest neighbor (KNN) classifiers are then applied. A comparison is conducted to identify the best results under Gmean and accuracy evaluation metrics.

The rest of the paper is organized as follows. Second section presents a review of the literature and the contributions of the bank direct marketing dataset. The third section provides a brief description of the oversampling techniques used in this research. Fourth section introduces details regarding the data set. Finally, the fifth section discusses the methodology followed in this research and the results obtained from running five different classifiers and their implications on the final prediction.

II. Related Work

From a broad perspective, the work in [6] surveyed the theoretical foundations of marketing analytics, which is a diverse field emerging from operations research, marketing, statistics, and computer science. They stated that predicting customer behavior is one of the challenges in direct marketing analysis. They also discussed big data visualization methods

for the marketing industry, such as multidimensional scaling, correspondence analysis, latent Dirichlet allocation, and customer relationship management (CRM). They debated on geographic visualization as a relative aspect of retail location analysis and tackled the general trade-off between its common practices and art. Additionally, they elaborated on discriminant analysis as a technique for marketing prediction. Discriminant analysis includes methods such as ensemble learning, feature reduction, and extraction. These techniques solve problems such as purchase behavior, review ratings, customer loyalty, customer lifetime value, sales, profit, and brand visibility.

Authors of [7] analyzed customer behavior patterns through CRM. They applied the Naive Bayes, J48, and multi-layer perceptron NNs on the same data set used in the current work. They also assessed the performance of their model using sensitivity, accuracy, and specificity measures. Their methodology involves understanding the domain and the data, building the model for evaluation, and finally visualizing the outputs. The visualization of their results showed that the J48 classifier outperformed the others with an accuracy of 89.40.

Moreover, [8] employed the same data set for other customer profiling purposes. Naive Bayes, random forests, and decision trees were used on the extended version of the data set examined in the current work. Preprocessing and normalization were conducted before evaluating the classifiers. RapidMiner tool was used for conducting the experiments and evaluation processes. They illustrated the parameter's adjustments of each classifier using a normalization technique applied previously. Furthermore, they showed the impact of these parameter values on accuracy, precision, and recall. Their results showed that decision trees are the best classifier for customer profiling and behavior prediction.

By contrast, [9] used the extended data set to create a logistic regression model for customer behavior prediction. This model is built on top of specific feature selection algorithms. Mutual information (MI) and data-based sensitivity analysis (DSA) are used to improve the performance over false-positive hits. They reduced the number of feature sets influencing the success of this marketing sector. They found that DSA is superior in the case of low false-positive ratio with nine selected features. MI is slightly better when false-positive values are marginally high with 13 selected features among a wide range of different features.

Additionally, a framework of three feature selection strategies was introduced by [10] to reveal novel features that directly affect data quality, which, in turn, exerts a significant impact on decision making. The strategies include identification of contextual features and evaluation of historical features. A problem is divided and conquered into sub-problems to reduce the complexity of the feature selection search space. Their framework tested the extended version of the data set used in the current work. Their goal was to target the best customers in marketing campaigns. The candidacy of the highest correlated hidden features was determined using DSA. The process involved designing new features of past occurrences aided by a domain expert. The last strategy split the original data upon the highest relevant set of features. The experiments confirmed the enrichment of data for better decision-making processes.

From visualization aspects, [11] explained several types of visualization techniques, such as radial, hierarchical, graph, and bar chart visualization, and presented the impact of human–computer interaction knowledge on opinion visualization systems. Prior domain knowledge yielded high understandability, user-friendliness, usefulness, and informativeness. Age factor affected the usability metrics of other systems, such as visual appeal, comprehensiveness, and intuitiveness. These findings were projected to the visualization of the direct marketing industry because it is mainly aided by end users and customers.

III. Oversampling Techniques

Oversampling is a concept that relates to the handling of imbalanced datasets. This method is performed by replicating or synthesizing minority class instances. Common approaches include randomly choosing instances (ROS) or choosing special instances on the basis of predefined conditions. Although oversampling methods are information-sensitive [12], they often lead to the overfitting problem, which may cause misleading classifications. This problem can be overcome by combining oversampling techniques with an ensemble of classifiers at an algorithmic level to attain the best performance.

An overview of the oversampling techniques used in the preprocessing phase of the current research, along with a brief description of the exploited classification algorithm, is introduced.

A. Synthetic Minority Oversampling Technique

Synthetic minority oversampling (SMOTE) generates synthetic instances on the basis of existing minority observations and calculates the k-nearest neighbors for each one [13]. The amount of oversampling needed determines the number of synthetic k-nearest neighbors created randomly on the link line.

B. Adaptive Synthetic Sampling Technique

Adaptive synthetics minority (ADASYN) is an improved version of SMOTE. After creating the random samples along the link line, ADASYN adds up small values to produce scattered and realistic data points [14], which are of reasonable variance built upon a weighted distribution. This approach is implemented according to the level of difficulty in learning while emphasizing the minority classes that are difficult to learn.

C. Random Over Sampling Technique

ROS is a non-heuristic technique. It is less computational than other oversampling methods and is competitive relative to complex ones [15]. A large number of positive minority instances are likely to produce meaningful results under this technique.

D. Adjusting the Direction of the Synthetic Minority Class Examples Technique

Adjusting the direction of the synthetic minority class (ADOMS) examples is another common oversampling technique [15] that relies on the principal component analysis

of the local data distribution in a feature space using the Euclidean distance between each minority class example and a random number of its k-nearest neighbors aided by projection and scaling parameters.

E. Selective Preprocessing of Imbalanced Data Technique

The selective preprocessing of imbalanced data (SPIDER) technique introduces a new concept of oversampling [16] and comprises two phases. The first phase is identifying the type of each minority class example by flagging them as safe or noisy using the nearest neighbor rule. The second phase is processing each example on the basis of one of three strategies; weak amplification, weak amplification with relabeling, and strong amplification.

F. Agglomerative Hierarchical Clustering Technique

Agglomerative hierarchical clustering (AHC) [16] starts with clusters of every minority class example. In each iteration, AHC merges the closest pair of clusters by satisfying some similarity criteria to maintain the synthetic instances within the boundaries of a given class. The process is repeated until all the data are in one cluster. Clusters with different sizes in the tree can be valuable for discovery.

IV. CLASSIFICATION TECHNIQUES

In this section, a concise description of the classification algorithms used in this research is presented.

A. Random Forests

Random forest [17] is a supervised learning algorithm used for classification and regression tasks. It is distinguished from decision trees by the randomized process of finding root nodes to split features. Random forest is efficient in handling missing values. Unless a sufficient number of trees is generated to enhance prediction accuracy, the overfitting problem is a possible drawback of this algorithm.

B. Support Vector Machines

SVM is a learning algorithm used in regression tasks. However, SVM [18] is preferable in classification tasks. This algorithm is based on the following idea: if a classifier is effective in separating convergent non-linearly separable data points, then it should perform well on dispersed ones. SVM finds the best separating line that maximizes the distance between the hyperplanes of decision boundaries.

C. Artificial Neural Networks

ANN is an approximation of some unknown function and is performed by having layers of "neurons" work on one another's outputs. Neurons from a layer close to the output use the sum of the answers from those of the previous layers. Neurons are usually functions whose outputs do not linearly depend on their inputs. ANN [19] uses initial random weights to determine the attention provided to specific neurons. These weights are iteratively adjusted in the back-propagation algorithm to reach a good approximation of the desired output.

D. Naive Bayes

Naive Bayes [20] is a direct and powerful classifier that uses the Bayes theorem. It predicts the probability that a given record or data point belongs to a particular class. The class with the highest probability is considered to be the most likely class. This algorithm assumes that all features are independent and unrelated. The Naive Bayes model is simple and easy to build and particularly useful for large data sets. This model is known to outperform even highly sophisticated classification methods.

E. K-Nearest Neighbor

The KNN learning algorithm is a simple classification algorithm that works on the basis of the smallest distance from the query instance to the training sample [21] to determine the simple majority of KNN as a prediction of the query. KNN is used due to its predictive power and low calculation time, and it usually produces highly competitive results.

V. VISUALIZATION METHODS

Data visualization exerts considerable impact on user software experience. The decision-making process benefits from the details obtained from large data volumes [22], which are usually built in a coherent and compact manner.

The purpose of the established model is to emphasize the importance of data visualization methods. It helps in conducting a perceptual analysis of a given situation. Visualization methods are used to illustrate hidden patterns inside data sets. This section introduces the characteristics of the visualization methods used in this research.

A. Scatter Plot

Scatter plots are a graphical display of data dispersion in Cartesian coordinates [22] that shows the strength of relationships between variables and determines their outliers. Variations include scatter plots with trend line. It is used to reveal the patterns in the normal distribution of the data points.

B. Bar Charts

Bar charts are used to represent discrete single data series [23]. The length usually represents corresponding values. Variations include multi-bar charts, floating bar charts, and candlestick charts.

C. Pie Charts

A synonym of a circle graph [23] is divided into a number of sectors to describe the size of a data wedge. Sectors are compared by using a labeled percentage. Variations include doughnut, exploding, and multi-level pie charts for hierarchical data.

D. Line Charts

Line charts visualization technique [23] is used to display the trend of data as connected points on a straight line over a specific interval of time. Variations include step and symbolic line charts, vertical–horizontal segments, and curve line charts.

VI. DataSet and Methodology

A. DataSet

The data set is obtained from the UCI machine learning repository. This data set is related to the direct marketing campaigns of a Portuguese banking institution. The current research uses the small version of the raw data set, which contains 4521 instances and 17 features (16 features and output). The classification goal is to predict whether the client will subscribe (yes/no) to a term deposit (variable y). Table I shows the dataset description, and Fig. 1 illustrates the distribution of each feature.

Fig. 2 shows the relationship between each attribute and the output. The attribute with unique values (distribution) of less than 6 is selected.

TABLE I. Attribute Information

| Attribute	Description	Unique Values
age	Age	67
job	Type of job	12
marital	Marital status	3
education	Education	4
default	Has credit in default	2
balance	Credit balance	2353
housing	Has housing loan	2
loan	Has personal loan?	2
contact	Contact communication type	3
day	Last contact day of the week	31
month	Last contact month of Year	12
duration	Last contact duration, in seconds	875
campaign	Number of contacts performed during this campaign and for this client	32
pdays	Number of days that passed by after the client was last contacted from a previous campaign	292
previous	Number of contacts performed before this campaign and for this client	24
poutcome	Outcome of the previous marketing campaign	4
out	has the client subscribed a term deposit	2
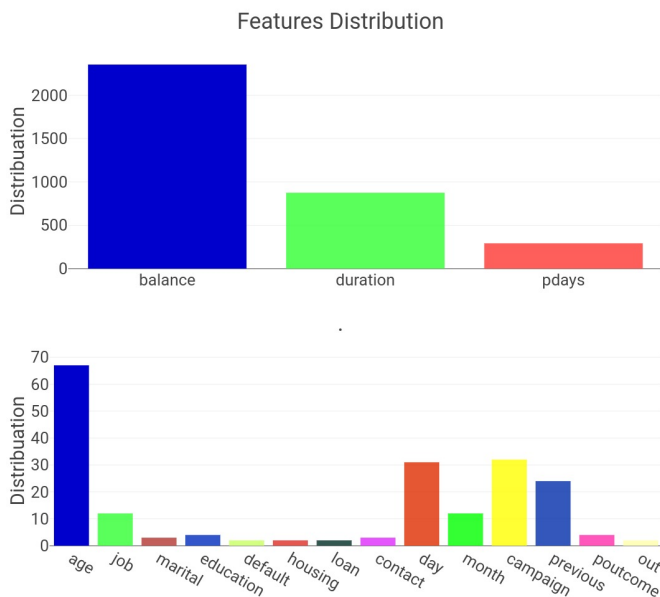
Fig. 1. Features Distribution
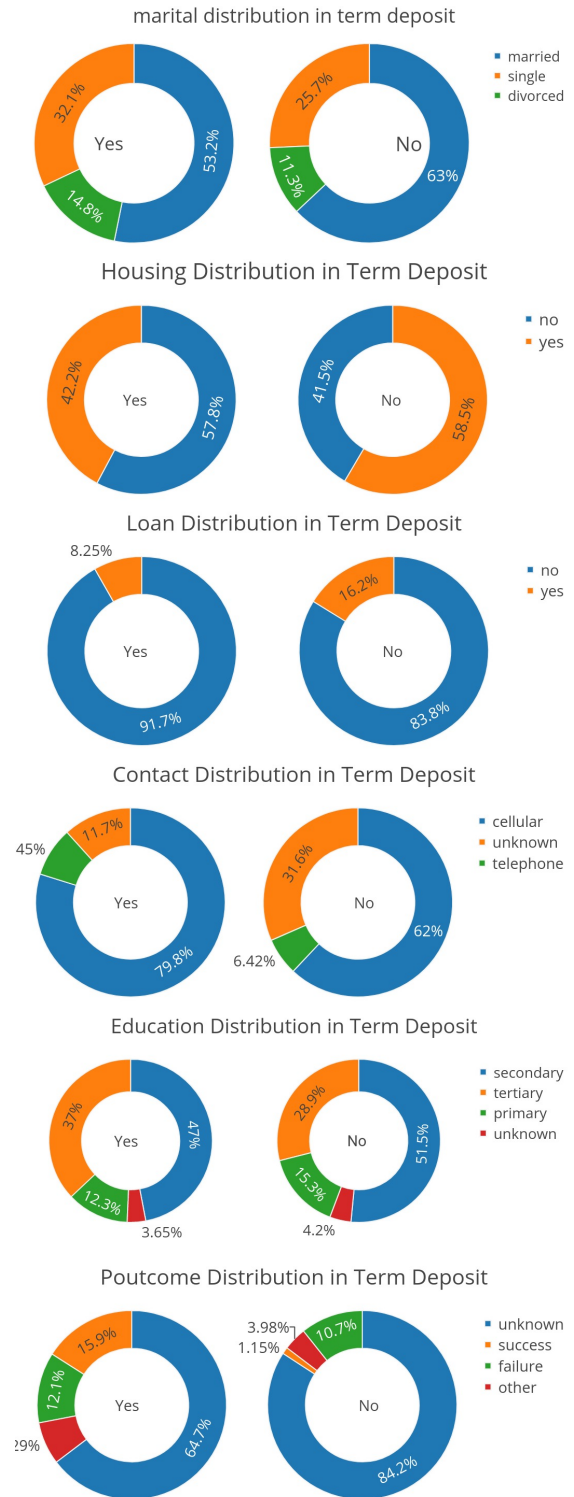
Fig. 2. Pie Chart Visualization Method for Features Distribution

Fig. 2 shows the relationship between each attribute and the output. The attribute that has unique values (distribution) less than 6 is selected.

According to the distribution of the target class shown in Fig. 3 and 4, the data set is imbalanced. That is, the percentage of class yes is 11.5 (500 records out of 4521), whereas that of class no is 88.5 (4021 records out of 4521).
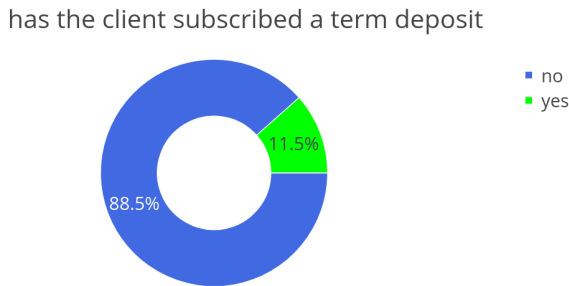

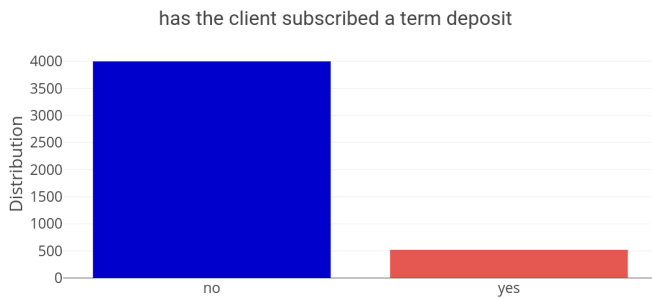Fig. 3. Pie Chart Visualization Method for Output Distribution

Fig. 4. Bar Chart Visualization Method for Output Distribution

B. Methodology

A preprocessing phase is first implemented to balance the data distribution by applying oversampling techniques, such as SMOTE with varying percentages. The next step is to determine which among ADASYN, ROS, SPIDER, ADOMS, and AHC is superior. The selection is aided by proper visualization methods. Then, random forest, SVM, ANN, Naive Bayes, and KNN classification algorithms are applied, and their assessment is conducted using Gmean as an evaluation metric. Other essential measurements, such as accuracy and recall, are also used.

VII. RESULTS AND DISCUSSION

In the preprocessing step, the SMOTE percentage is set to 100. Gmean and accuracy are calculated for different classifiers. Table II shows the results.

Accordingly, the Naive Bayes technique has the highest Gmean among all techniques. The accuracy and Gmeans of all five classifiers in the original dataset are shown in Table III.

Then, Naive Bayes with different SMOTE percentages is applied to select the most appropriate one that throws the best results from this data set. Table IV shows the results of Naive Bayes on the data set with different SMOTE percentages ranging from 100 to 800.

TABLE II. GMEAN AND ACCURACY OF DIFFERENT CLASSIFICATION TECHNIQUES AFTER USING SMOTE OVERSMPLING TECHNIQUE WITH PERCENTAGE 100

Method	Gmean	Accuracy
Random forest	0.620	89.98
SVM	0.658	89.38
NN	0.667	87.635
naive bayes	0.710	85.11
K nearest neighbor	0.547	85.02

TABLE III. RESULTS OF DIFFERENT CLASSIFICATION TECHNIQUES ON THE ORIGINAL DATASET (GMEAN,ACCURACY)

Method	Gmean	Accuracy
Random forest	0.52	90.02
SVM	0.61	89.29
NN	0.59	88.18
Naive Bayes	0.68	86.88
K nearest neighbor	0.49	86.5

The best results with Naive Bayes are those with a SMOTE value set to 400, as shown in Fig. 5. The figure also shows the Gmeans of different SMOTE percentages using the Naive Bayes classifier. The results in Table IV are shown as a chart to present a clear reading of the best Gmean.
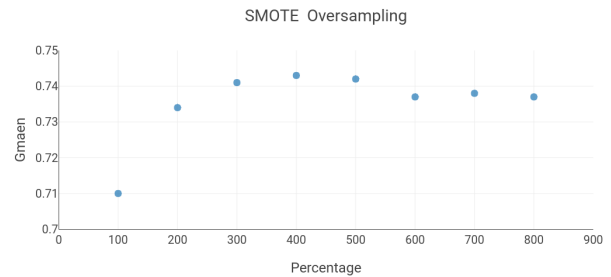

Fig. 5. Gmean of Naive Bayes with Different SMOTE Percentages Ranging from 100 to 800

Another experiment is conducted to determine the second best classifier under SMOTE set to 400. Table V shows the results. As shown in Fig. 6, the best classifier is the one with the highest Gmean, that is, Naive Bayes and SVM with a Gmean value of 0.74.

SVM and Naive Bayes are good candidates for further analysis. Naive Bayes is preferred over SVM because it is fast and easy to install. Fig. 6 shows the Gmean value before and after applying SMOTE. SMOTE enhances the Gmeans for all applied classification techniques.

The final step is to compare oversampling techniques with

TABLE IV. GMEAN AND ACCURACY OF NAIVE BAYES WITH DIFFERENT SMOTE PERCENTAGES RANGING FROM 100 TO 800

Percentage	Gmean	Accuracy
100	0.710	85.11
200	0.734	84.03
300	0.741	82.48
400	0.743	80.8
500	0.742	79.51
600	0.737	78.23
700	0.738	77.17
800	0.737	75.98

TABLE V. ACCURACY AND GMEAN OF DIFFERENT TECHNIQUES USING SMOTE WITH PERCENTAGE 400

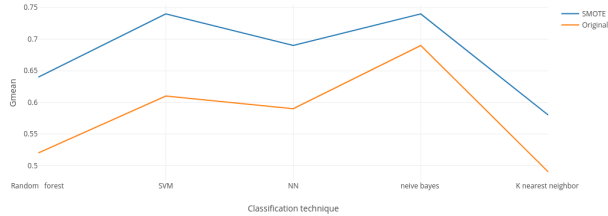| Method	Gmean	Accuracy
Random forest	0.64	88.8
SVM	0.74	87.25
NN	0.69	86.92
naive bayes	0.74	80.80
K nearest neighbor	0.58	83.98

Fig. 6. Gmean of different techniques using SMOTE with percentage 400 compared with original

the best SMOTE percentage. Comparing these results with those of the original dataset is also important. Naive Bayes is used with all oversampling techniques. Accuracy and Gmean are calculated by different percentages of SMOTE. Table VI shows the results.

The scatter plot method is used to visualize the results of the six oversampling techniques, as illustrated in Fig. 7.

TABLE VI. COMPARISON BETWEEN DIFFERENT OVERSAMPLING TECHNIQUES (PERCENTAGES RANGING FROM 100 TO 800)

	Gmean					
	SMOTE	ADASYN	ROS	ADOMS	SPIDER	AHC
100	0.71	0.680	0.68	0.68	0.59	0.67
200	0.73	0.690	0.68	0.68	0.61	0.78
300	0.741	0.741	0.69	0.70	0.68	0.78
400	0.743	0.740	0.68	0.70	0.71	0.78
500	0.742	0.720	0.68	0.71	0.72	0.79
600	0.737	0.682	0.69	0.72	0.72	0.79
700	0.738	0.651	0.68	0.72	0.72	0.732
800	0.737	0.671	0.68	0.67	0.60	0.72
	Accuracy					
	SMOTE	ADASYN	ROS	ADOMS	SPIDER	AHC
100	85.11	81.43	80.90	84.83	82.51	84.43
200	84.03	78.07	79.81	83.12	82.45	84.43
300	82.48	77.28	77.03	81.89	79.13	83.71
400	80.8	74.22	76.88	81.80	80.12	82.27
500	79.51	68.52	77.08	80.11	79.15	80.96
600	78.23	66.88	76.81	80.20	78.22	77.18
700	77.17	66.23	76.90	80.27	80.11	77.63
800	75.98	66.00	76.97	80.21	79.15	76.15

Table VII shows the best percentage for all used oversampling techniques. Apart from the Gmean and accuracy of the best percentage of each oversampling technique used, Fig. 8 shows the final results. Oversampling techniques are found to enhance performance. AHC with Naive Bayes has the best Gmean among all the oversampling techniques.

VIII. CONCLUSION

This research aimed to provide a visualization mechanism for simple classification tasks. Experiments were conducted on an imbalanced data set for a direct marketing campaign of a
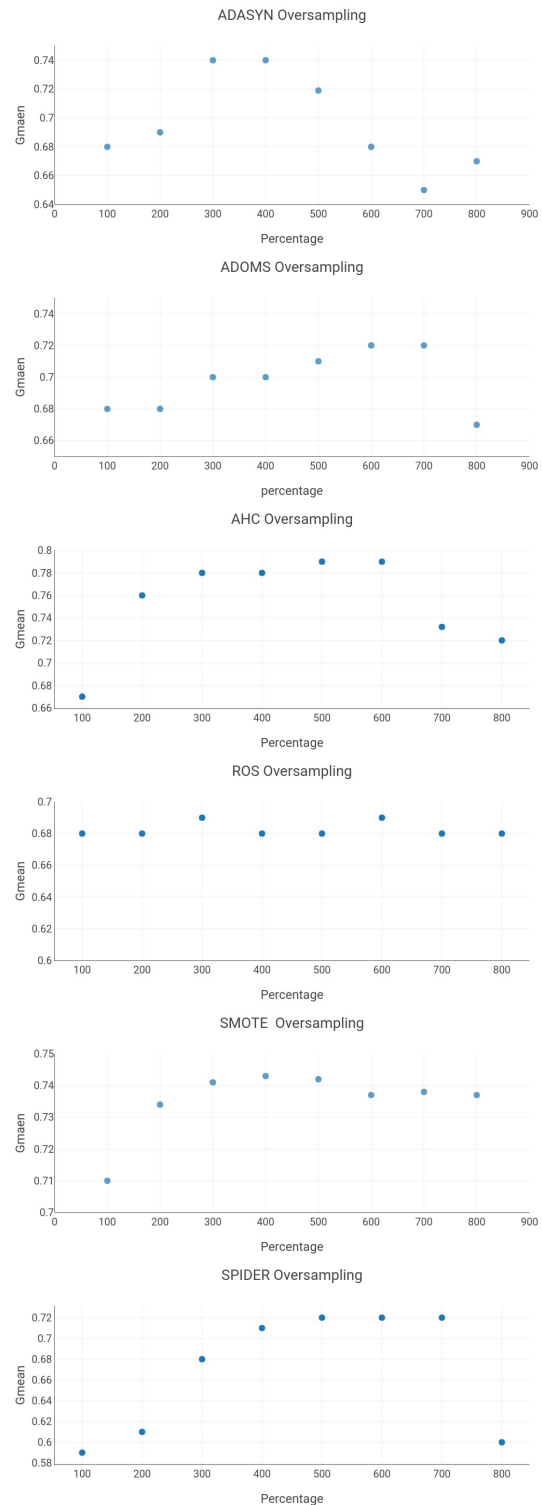

Fig. 7. Gmean of Different Oversampling Techniques with Different Percentages Ranging from 100 to 400

TABLE VII. COMPARISON BETWEEN DIFFERENT OVERSAMPLING
TECHNIQUES WITH BEST GMEAN PERCENTAGE

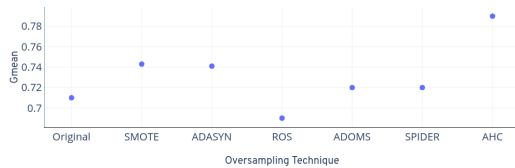| Technique	Best Percentage	Naive Bayes Gmean	Naive Bayes Accuracy
SMOTE	400	0.743	80.8
ADASYN	300	0.741	77.278
ROS	300	0.69	77.030
ADOMS	700	0.72	80.272
SPIDER	700	0.72	80.110
AHC	500	0.79	80.960

Fig. 8. Gmean of Different Oversampling Techniques Compared with
Original

Portuguese bank institution. The goal was to predict whether a customer will subscribe to a term deposit. The experiments were conducted using different oversampling techniques, along with five selected common classifiers from the literature. The results showed that AHC with SMOTE of 500 outperformed the other oversampling techniques. The Naive Bayes and SVM classifiers provided the best prediction accuracy and Gmean values. Naive Bayes was preferred for its simplicity and maintainability. This research is limited to the applying of the most used oversampling techniques and it only performed on the small version of the direct marketing campaign of a Portuguese banking institution. Also, not all known classifiers involved in conducting the experiments, nor the visualization techniques and the results can be further analyzed by other performance measurements to draw more precise conclusions.

REFERENCES

[1] V. L. Miguéis, A. S. Camanho, and J. Borges, "Predicting direct marketing response in banking: comparison of class imbalance methods," *Service Business*, vol. 11, no. 4, pp. 831–849, 2017.

[2] S. Ghosh, A. Hazra, B. Choudhury, P. Biswas, and A. Nag, "A comparative study to the bank market prediction," in *International Conference on Machine Learning and Data Mining in Pattern Recognition*. Springer, 2018, pp. 259–268.

[3] G. Marinakos and S. Daskalaki, "Imbalanced customer classification for bank direct marketing," *Journal of Marketing Analytics*, vol. 5, no. 1, pp. 14–30, 2017.

[4] J. Kokina, D. Pachamanova, and A. Corbett, "The role of data visualization and analytics in performance management: Guiding entrepreneurial growth decisions," *Journal of Accounting Education*, vol. 38, pp. 50–62, 2017.

[5] G. Costagliola, M. De Rosa, V. Fuccella, and S. Perna, "Visual languages: A graphical review," *Information Visualization*, vol. 17, no. 4, pp. 335–350, 2018.

[6] S. L. France and S. Ghose, "Marketing analytics: Methods, practice, implementation, and links to other fields," *Expert Systems with Applications*, 2018.

[7] S. S. Raju and P. Dhandayudam, "Prediction of customer behaviour analysis using classification algorithms," in *AIP Conference Proceedings*, vol. 1952, no. 1. AIP Publishing, 2018, p. 020098.

[8] S. Palaniappan, A. Mustapha, C. F. M. Foozy, and R. Atan, "Customer profiling using classification approach for bank telemarketing," *JOIV: International Journal on Informatics Visualization*, vol. 1, no. 4-2, pp. 214–217, 2017.

[9] N. Barraza, S. Moro, M. Ferreyra, and A. de la Peña, "Mutual information and sensitivity analysis for feature selection in customer targeting: A comparative study," *Journal of Information Science*, p. 0165551518770967, 2018.

[10] S. Moro, P. Cortez, and P. Rita, "A framework for increasing the value of predictive data-driven models by enriching problem domain characterization with novel features," *Neural Computing and Applications*, vol. 28, no. 6, pp. 1515–1523, 2017.

[11] K. Sagar and A. Saha, "A systematic review of software usability studies," *International Journal of Information Technology*, pp. 1–24, 2017.

[12] F. Shakeel, A. S. Sabhitha, and S. Sharma, "Exploratory review on class imbalance problem: An overview," in *Computing, Communication and Networking Technologies (ICCCNT), 2017 8th International Conference on*. IEEE, 2017, pp. 1–8.

[13] D. Elreedy and A. F. Atiya, "A novel distribution analysis for smote oversampling method in handling class imbalance," in *International Conference on Computational Science*. Springer, 2019, pp. 236–248.

[14] Y. E. Kurniawati, A. E. Permanasari, and S. Fauziati, "Adaptive synthetic-nominal (adasyn-n) and adaptive synthetic-knn (adasyn-knn) for multiclass imbalance learning on laboratory test data," in *2018 4th International Conference on Science and Technology (ICST)*. IEEE, 2018, pp. 1–6.

[15] M. S. Santos, J. P. Soares, P. H. Abreu, H. Araujo, and J. Santos, "Cross-validation for imbalanced datasets: Avoiding overoptimistic and overfitting approaches [research frontier]," *ieee ComputatioNal iNtelligeNCe magaziNe*, vol. 13, no. 4, pp. 59–76, 2018.

[16] S. Fotouhi, S. Asadi, and M. W. Kattan, "A comprehensive data level analysis for cancer diagnosis on imbalanced data," *Journal of biomedical informatics*, 2019.

[17] Z.-H. Zhou and J. Feng, "Deep forest," *arXiv preprint arXiv:1702.08835*, 2017.

[18] G. Haixiang, L. Yijing, J. Shang, G. Mingyun, H. Yuanyue, and G. Bing, "Learning from class-imbalanced data: Review of methods and applications," *Expert Systems with Applications*, vol. 73, pp. 220–239, 2017.

[19] W. Cao, X. Wang, Z. Ming, and J. Gao, "A review on neural networks with random weights," *Neurocomputing*, vol. 275, pp. 278–287, 2018.

[20] J. Korst, V. Pronk, M. Barbieri, and S. Consoli, "Introduction to classification algorithms and their performance analysis using medical examples," in *Data Science for Healthcare*. Springer, 2019, pp. 39–73.

[21] J. Gou, H. Ma, W. Ou, S. Zeng, Y. Rao, and H. Yang, "A generalized mean distance-based k-nearest neighbor classifier," *Expert Systems with Applications*, vol. 115, pp. 356–372, 2019.

[22] M. Behrisch, M. Blumenschein, N. W. Kim, L. Shao, M. El-Assady, J. Fuchs, D. Seebacher, A. Diehl, U. Brandes, H. Pfister *et al.*, "Quality metrics for information visualization," in *Computer Graphics Forum*, vol. 37, no. 3. Wiley Online Library, 2018, pp. 625–662.

[23] K. Santhi and R. M. Reddy, "Critical analysis of big visual analytics: A survey," in *2018 IADS International Conference on Computing, Communications & Data Engineering (CCODE)*, 2018.