Marketing Enhancement via Machine Learning

Description

Creating a winning marketing campaign is intimidating. marketing campaigns are a business' efforts toward promoting a specific company goal. They reach customers through a combination of contact methods and entice customers to purchase.

Leveraging machine learning to identify probable customers would be beneficial for companies to succeed in future campaigns. Indeed, we can utilize the results of previous campaigns to train our machine learning. Then, our machine learning is used to classify the new customers during a new marketing campaign. The result of our classifier can help marketing business planners to focus on customers who are more willing to accept their offers. Therefore, we use data from a bank marketing campaign of a Portuguese banking institution [1] to devise a machine learning pipeline that is able to classify future customers and help marketing campaign planners to increase their campaign success rate.

The information available in the dataset is used to build the machine learning pipeline trained to classify future customers. To evaluate our model, the F1-score is used as the measure of the evaluation, which is the harmonic mean precision and recall. Since, in our case, the cost of false_positive (identifying uninterested customers as potential ones) and false_negative (identifying interested customers as uninterested) is comparatively equal, F1-score is a reasonable choice.

Data Analysis

The data includes information on around 36000 customers along with the result of the bank marketing campaign, which was based on phone calls. Often, more than one contact with the same client was required, in order to access if the product (bank term deposit) would be ('yes') or not ('no') subscribed. The dataset includes information about customers including age, job, marital status, average yearly balance, level of education, the number of contacts performed during the campaign, and the status of the loan.

As the first numerical feature, we consider the number of calls during the campaign through a boxplot. From Figure 1, we can observe that there are some anomalies in the dataset. We remove samples related to more than 30 calls from the dataset.

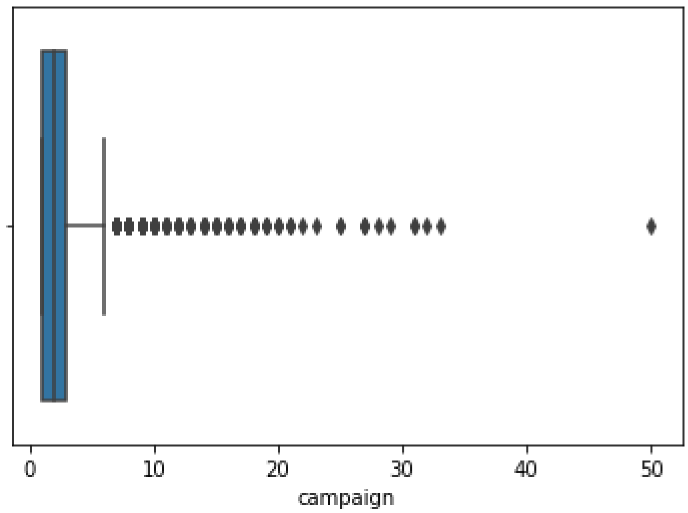


Figure 1: Number of calls

From figure 2, the distribution of age in the dataset is skew normal. The histogram shows that the majority of customers were between 30 to 50 years old. There are some anomalies around 70 to 90 years old. But we keep them because they are not undesired anomalies.

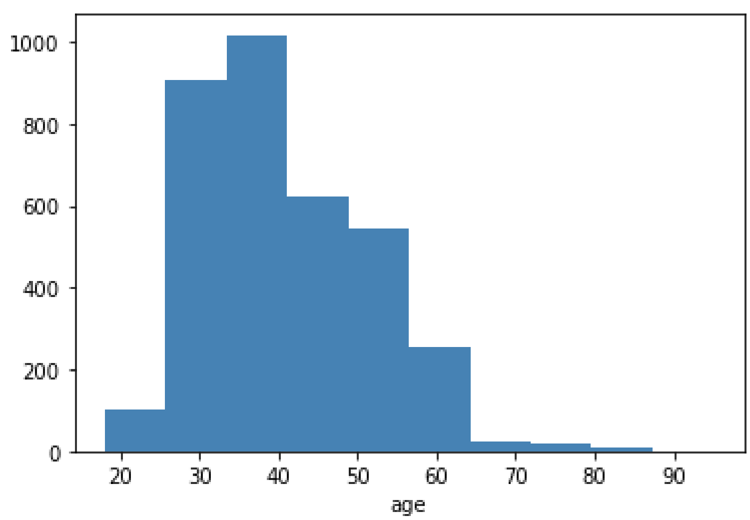
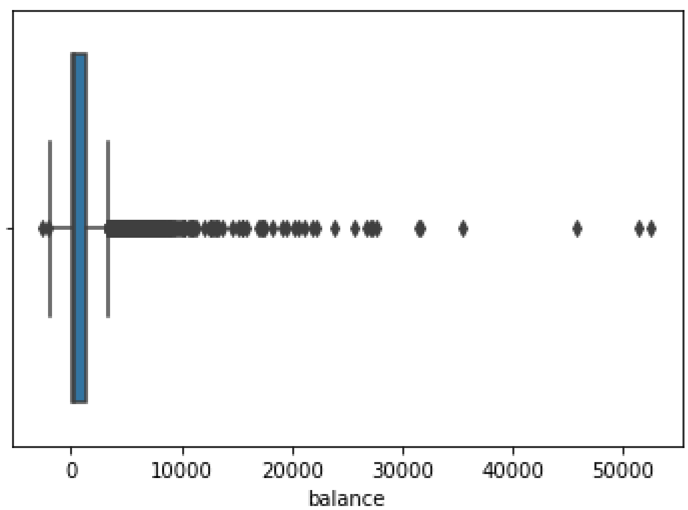


Figure 2: age

Finally, figure 3 represents data with respect to balance. The boxplot suggests that there are some anomalies in the data. To address this issue, we remove samples whose balance is above 35000.



Now, let's explore our data. From figure 1, the majority of clients were married. The next important variable is the job type of clients.

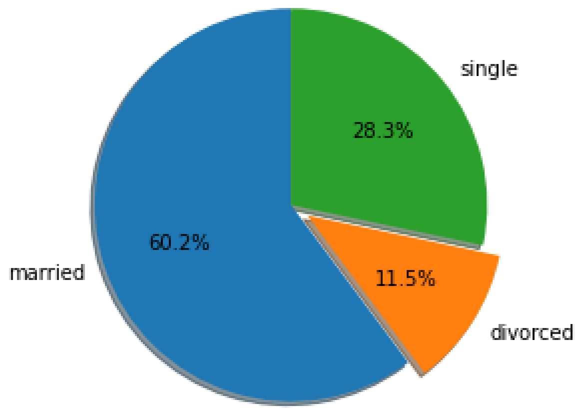


Figure1: Marital status of customers

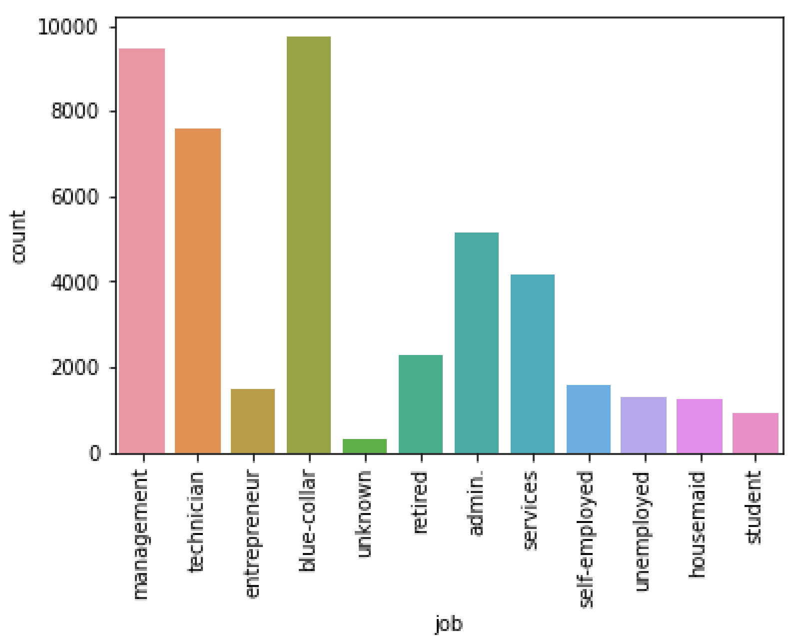


Figure 2: job type of clients

In Figure 2, we can observe that a high number of our clients work in the management sector, or as technicians/workers. Thus, our data is well divided in terms of job type. The last variable, which is worth taking a look at is the level of education.

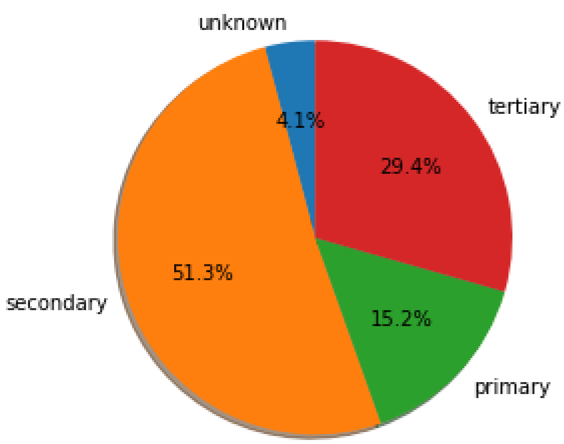


Figure 3: Level of education of customers

Figure 3 shows that only 30% of our clients have tertiary education.

Data Preprocessing:

To make the dataset ready to use by our machine learning model, some data munging steps should be performed. First, predictor variables job and education contain a few null values. To address this issue, we remove these null values from the dataset because they are less than 1% of the whole dataset. Second, as we discussed in the previous section, there are some *anomalies in campaign and salary predictor variables. We remove these anomalies from the dataset since they can deteriorate the performance of our machine learning model. Next, we transform the categorical features into numeric ones by using the one-hot encoding technique [2]. Finally, since generally, machine algorithms except decision trees need to have normalized data, we perform normalization on our data.*

Methodology

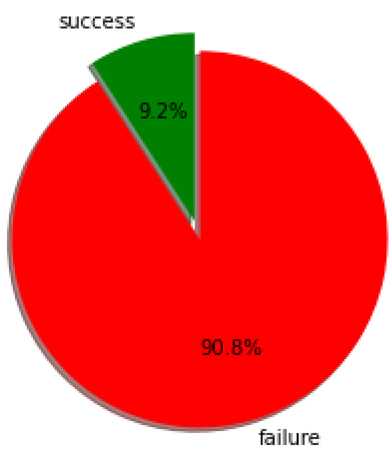


Figure 5: the result of the campaign

In this section, we present our machine learning model. According to the above figure, only 9% of the customers said yes to the campaign. Therefore, the problem at hand is an imbalanced classification problem.

Since the task at hand is classification, we utilize some conventional classification methods such as KNN, Logistic Regression, bagging, and random forest trees. To enhance the performance of our decision tree classifiers, we use the dynamic ensemble selection technique [3]. In this method, we train ensemble decision tree classifiers. Then, during the test phase, we consider the trees that make at least one correct prediction on the neighborhood of the test sample. Then, the test sample is classified based on the majority vote of classifiers. Finally, we are unaware of the nature of the system that produces the dataset. Thus, utilizing deep learning models that can draw meaningful inferences from highly complex structure data would be reasonable. To this end, we first use three stacked layers of 1D convolutional neural networks, then the extracted features are fed to a feed-forward neural network to classify the data. We hope that this architecture can capture the complexity of our data.

For each of the classifiers, a range of hyperparameters is considered to better evaluate them. For Logistic regression, different class weighting methods are utilized as the task at hand is imbalanced classification. Besides, we perform the VIF test [5], and remove predictor variables that are highly correlated to each other, i.e. predictor variables whose VIF score is above 2.6. For the KNN classifier, we consider a different number of neighbors. For decision trees and the ensemble classifier, we consider a different number *of trees. Finally, for the CNN classifier, a different number of layers is considered for hyperparameter tuning.*

Results

To evaluate our classifiers, k-fold cross-validation is used, which is a powerful method to flag problems like overfitting or selection bias [4]. As we discussed earlier, the F1 score is used as a measure evaluation. Moreover, as has been discussed in the previous section, hyperparameter tuning is performed, and the best model, which achieved the highest score is reported.

Figure 5 presents, the F1 scores achieved by the proposed classifiers via boxplots. From the figure, we can observe that CNN and Ensemble learning algorithms achieve a higher F1 score compared to the other classifiers. The result suggests that our CNN could capture some complexity of data and extract meaningful features. However, since we ensemble bagging and random forest via the ensemble method, it could get a good result, which is comparable to a more complex method like the CNN approach.

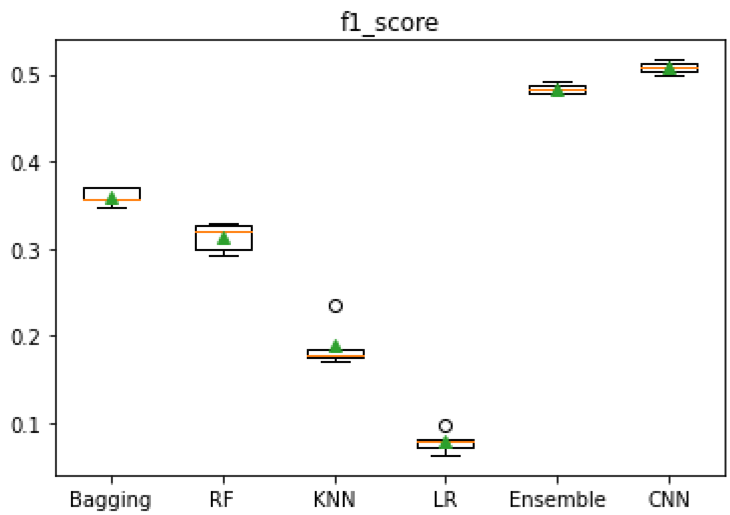


Fig 5: F1-score of classifiers

Here, we present the result of hyperparameter tuning for each classifier.

F1-score Logistic regression	
class weight: balanced	0.085
class weights: None	0.071

F1-score Logistic regression classifier

number of neighbors: 3	0.175
number of neighbors: 5	0.191
number of neighbors: 10	0.185

F1-score KNN classifier

number of trees	Random Forest	Ensemble	Bagging
20	0.305	0.475	0.356
50	0.309	0.48	0.358
100	0.313	0.484	0.361
200	0.313	0.484	0.361

F1-score decision tree-based approaches

number of layers	F1-score
1	0.495
2	0.502
3	0.508

F1-score CNN

Conclusion

In conclusion, the result implies that the proposed machine learning pipeline can be considered by the marketing campaign planners for future campaigns. It can be used as a tool to better customer identification. To improve the model, we can report the probability of class membership. This approach summarizes the uncertainty of an example belonging to each class label, and it is more nuanced and can be interpreted by a human operator in decision-making. Furthermore, a wider range of hyperparameters *can be tuned, or a broader range of classifiers can be leveraged such as different neural networks, and LDA models.*

References

[1]. archive.ics.uci.edu/ml/datasets/bank+marketing.

[2].https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.OneHotEncoder.html.

[3].https://www.amazon.com/Ensemble-Methods-Foundations-Algorithms-Recognition/dp/1439830037/ref=as_li_ss_tl?dchild=1&keywords=Ensemble+Methods&qid=1595369184&s=books&sr=1-2&linkCode=sl1&tag=inspiredalgor-20&linkId=50dfc0e5b2cdde98a1303c4de913741d&language=en_US.

[4]. https://en.wikipedia.org/wiki/Cross-validation_(statistics).

[5].https://www.statisticshowto.com/variance-inflation-factor/#:~:text=In%20general%2C%20a%20VIF%20above,like%20x%20and%20x2.