Since the target data for heart disease was not balanced, it was necessary to perform an oversampling technique with default parameters. Thus, the unbalanced data becomes balanced based on the oversampling process that has been carried out. Moreover, the K-Fold testing technique employed K-Fold 10 [19], [20]. The illustration of K-Fold is shown in Figure 5, where the prediction model begins by dividing all data into training data and test data with K-Fold cross-validation, and cross-testing of each - each algorithm. Performance evaluation is carried out on the model with the aim of knowing how well the model is performing using test data.

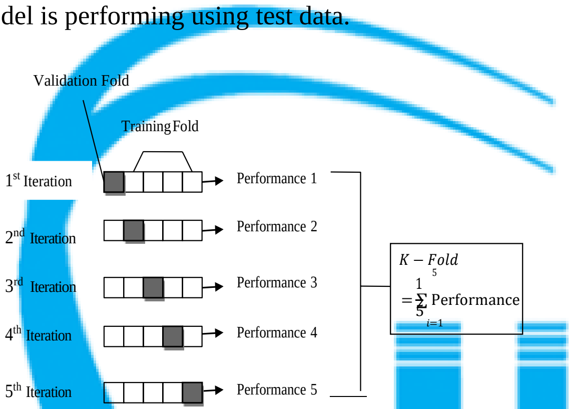

Figure 5. K-Fold Illustration

The evaluation for this study was based on accuracy, precision, sensitivity, and specificity. The evaluation process utilised the test data that has been separated in the previous process and the evaluation results employed the confusion matrix shown in Table2.

Table 1. K-Fold 10

Algorithm	Accuracy K-Fold (%)
C45	86,74
RANDOM FOREST	90,56
C45 + SMOTE	91,77
RF + SMOTE	94,43
C45 + ADASYN	91,71
RF + ADASYN	94,34

The model generated from running Equation 1 to Equation 4 operated the Confusion Matrix on the C45 algorithm which produced an accuracy of 86.74%. On the other hand, the Random Forest Classifier algorithm had succeeded in producing a prediction model with a higher accuracy value than C4.5, which was 90.56%. Table 3 shown the results of the comparison of SMOTE and ADASYN.

Table 2 Comparison of SMOTE and ADASYN

K-Fold Cross Validation		
Algorithm	**SMOTE (%)**	**ADASYN (%)**
Random Forest	94,43	94,34
C45	91,77	91,71

By using the K-Fold calculation based on table 3, in this study the best accuracy value was obtained in the K-Fold 10 calculation for SMOTE applied to Random Forest 94.43% and ADASYN applied to Random Forest 94.34%. Thus, the combined technique of Random Forest with SMOTE and Random Forest with ADSYN had better performance than C45 with ADASYN and C45 with SMOTE. It was proven that the Random Forest algorithm with SMOTE has the best ability to predict class data compared to Random Forest with ADASYN with an accuracy of 94.43%.

4. Conclusion

Based on the results of the research that has been carried out, it was concluded that the SMOTE and ADASYN oversampling techniques had a significant impact on the classification results. It was proven that the increase in accuracy which occurred in the Random Forest and C.45 algorithms was quite significant. However, the highest accuracy was the combination of implementing SMOTE with Random Forest which reached 94.43%. The results of this study can be considered by experts to assist decisions in dealing with heart disease. Moreover, regarding further research, it is suggested to correlate a dashboard and visualization of the relationship between features which affect heart disease.

References

[1] "Kementerian Kesehatan Republik Indonesia." .

[2] BHF, "UK Factsheet," *Br. Hear. Found.*, no. April, pp. 1–21, 2019.

[3] J. J. Pangaribuan, C. Tedja, and S. Wibowo, "PERBANDINGAN METODE ALGORITMA C4.5 DAN EXTREME LEARNING MACHINE UNTUK MENDIAGNOSIS PENYAKIT JANTUNG KORONER," 2019.

[4] A. Rohman and D. M. Rochcham, "MODEL ALGORITMA C4.5 UNTUK PREDIKSI PENYAKIT JANTUNG," 2018.

[5] N. Khasanah, R. Komarudin, N. Afni, Y. I. Maulana, and A. Salim, "Skin Cancer Classification Using Random Forest Algorithm," *Sisfotenika*, vol. 11, no. 2, p. 137, 2021, doi: 10.30700/jst.v11i2.1122.

[6] S. Ath *et al.*, "Jurnal Teknologi Terpadu HYBRID MACHINE LEARNING MODEL UNTUK MEMPREDIKSI PENYAKIT JANTUNG DENGAN METODE LOGISTIC REGRESSION DAN RANDOM," vol. 8, no. 1, pp. 40–46, 2022.

[7] I. M. El-Hasnony, O. M. Elzeki, A. Alshehri, and H. Salem, "Multi-Label Active Learning-Based Machine Learning Model for Heart Disease Prediction," *Sensors*, vol. 22, no. 3, 2022, doi: 10.3390/s22031184.

[8] S. Maldonado, J. López, and C. Vairetti, "An alternative SMOTE oversampling strategy for high-dimensional datasets," *Appl. Soft Comput. J.*, vol. 76, pp. 380–389, 2019, doi: 10.1016/j.asoc.2018.12.024.

[9] D. Elreedy and A. F. Atiya, "A Comprehensive Analysis of Synthetic Minority Oversampling Technique (SMOTE) for handling class imbalance," *Inf. Sci. (Ny).*, vol. 505, pp. 32–64, 2019, doi: 10.1016/j.ins.2019.07.070.

[10] R. . Nurdin, "Pernyataan Keaslian," *Digilib.Uin-Suka.Ac.Id*, no. April 2020, p. 506812, 2021.

[11] "CDC - 2020 BRFSS Survey Data and Documentation." .

[12] R. Siringoringo, "KLASIFIKASI DATA TIDAK SEIMBANG MENGGUNAKAN ALGORITMA SMOTE DAN k-NEAREST NEIGHBOR," 2018.

[13] G. Fico, J. Montalva, A. Medrano, N. Liappas, G. Cea, and M. T. Arredondo, "EMBEC &amp; NBC 2017," *IFMBE Proc.*, vol. 65, pp. 1089–1090, 2018, doi: 10.1007/978-981-10-5122-7.

[14] S. Rahayu, T. Bharata Adji, N. Akhmad Setiawan, and D. Teknik Elektro dan Teknologi Informasi, "Penghitungan k-NN pada Adaptive Synthetic-Nominal (ADASYN-N) dan Adaptive Synthetic-kNN (ADASYN-kNN) untuk Data Nominal-Multi Kategori," *Ktrl.Inst (J.Auto.Ctrl.Inst)*, vol. 9, no. 2, p. 2017.

[15] W. Sullivan, *Machine Learning For Beginners Guide Algorithms*, vol. 4, no. 1. 2017.

[16] A. Cherfi, K. Nouira, and A. Ferchichi, "Very Fast C4.5 Decision Tree Algorithm," *Appl. Artif. Intell.*, vol. 32, no. 2, pp. 119–137, 2018, doi: 10.1080/08839514.2018.1447479.

[17] M. Kretowski, *Evolutionary Decision Trees in Large-Scale Data Mining*. 2019.

[18] A. Primajaya and B. N. Sari, "Random Forest Algorithm for Prediction of Precipitation," *Indones. J. Artif. Intell. Data Min.*, vol. 1, no. 1, p. 27, 2018, doi: 10.24014/ijaidm.v1i1.4903.

[19] T. Djatna, M. K. D. Hardhienata, and A. F. N. Masruriyah, "An intuitionistic fuzzy diagnosis analytics for stroke disease," *J. Big Data*, vol. 5, no. 1, 2018, doi: 10.1186/s40537-018-0142-7.

[20] S. Zitao, "3 min of Machine Learning: Cross Vaildation," *Zitao's Web*, 2020. .