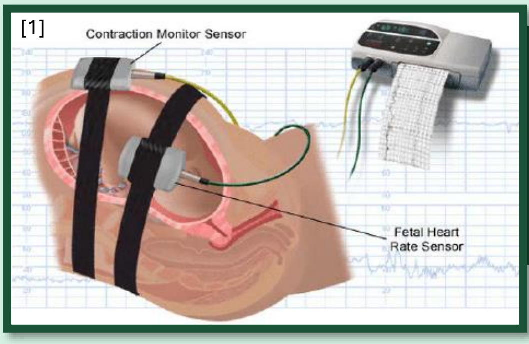
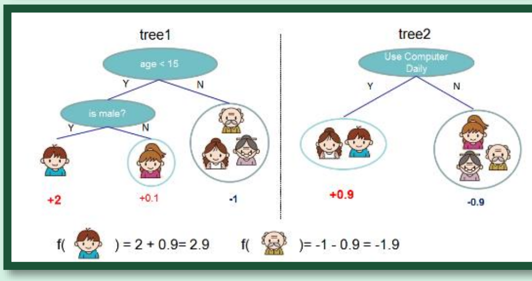
Alyssa Grogorenz-de Oliveira

DEVELOPING MACHINE LEARNING METHODS TO IDENTIFY HIGH RISK PREGNANCIES

[1]





[7]

To help achieve the SDG goal 3.2 [3], cardiotography (CTG) data [4] monitoring foetal health was used to train a classification algorithm. The dataset consisted of 21 features and one target variable, classifying each instance as **normal**, **suspect**, or **pathological**. This multiclass classification problem required rebalancing of the dataset to allow classifiers to detect suspect and pathological instances, which are rarer but more crucial to detect [5]. Repeated stratified k-fold cross validation reduced the likelihood of overfitting [6].

TARGET 3·2

5 YRS

END ALL PREVENTABLE DEATHS UNDER 5 YEARS OF AGE [2]

XGBoost (extreme gradient boosting) [7] is an ensemble of multiple Decision Trees. It calculates the sum of predictions from each tree and uses a quantile sketch function to handle weighted data and efficiently calculate the split of each tree based on the distribution of features. A regularisation parameter, shrinkage, and subsampling prevent overfitting.

Feature Selection (FS)

Classifier performance can be enhanced and overfitting avoided by removing irrelevant features [8], [9].

To compare performance, XGBoost was either trained with all **21** features or **7** relevant features, chosen based on their correlations [10] and mutual information [11].

Resampling

SMOTE: Synthetic Minority Over-Sampling Technique. Increases the number of instances in minority classes by generating samples between each instance and its k-nearest neighbour [12].

RUS: (Repeated) Random Under-Sampling. Reduces the number of instances in majority classes by (repeatedly) selecting random instances.

Assessment: Performance Metrics

Recall: How many relevant instances of a class are detected [13].

Precision: How well a model correctly predicts a given instance of a class [13].

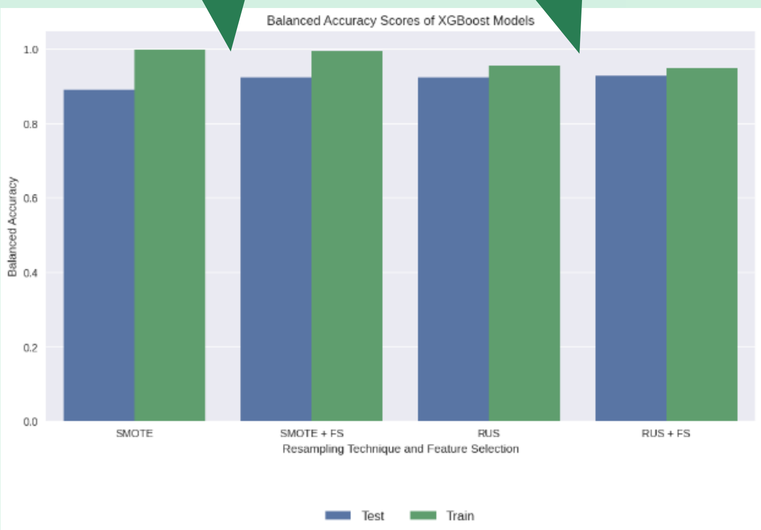
F_β: Summary statistic of a class, β = 2 values Recall twice as much as Precision [14]. Important in medicine to create a less-discriminant classifier [13].

Balanced Accuracy: Accuracy of a classifier that accounts for class imbalance [15].

SMOTE: higher Train Accuracy; the models might have overfit on replicated samples

RUS: smaller difference between Train and Test accuracy; models are more generalisable.

Model	Precision			Recall			F_β		
	N	S	P	N	**S**	**P**	N	S	**P**
SMOTE	0.96	0.87	0.91	0.98	0.78	***0.91***	0.98	0.79	0.91
SMOTE+FS	0.98	0.78	0.97	0.96	0.86	***0.94***	**0.97**	**0.84**	**0.95**
RUS	1.00	**0.61**	0.80	0.88	***0.95***	***0.94***	0.90	0.85	0.91
RUS+FS	1.00	**0.62**	0.77	0.87	***0.97***	***0.94***	**0.89**	**0.87**	**0.90**



FS increased Test Accuracy and lowered the difference between Test and Train Accuracy by reducing the likelihood of overfitting; models with FS perform better.

In Summary: Prioritising ***high Recall of S and P*** is important in a medical context to lower the False Negative rate and save lives [13]. However, **low Precision** should still be avoided where possible to reduce the False Positive rate and prevent unnecessary medical interventions. $F_{\beta=2}$ strikes a balance between these two metrics. SMOTE + FS had a **higher F_β for P and N**, whilst RUS + FS a higher **F_β for S**. However, given the similar values between the SMOTE + FS and RUS + FS models overall, the latter should be favoured due to the lower likelihood of overfitting to the train data.

[1] Y. Salini et al., ‘Cardiotocography Data Analysis for Fetal Health Classification Using Machine Learning Models’, *IEEE Access*, vol. 12, pp. 26005–26022, 2024, doi: 10.1109/ACCESS.2024.3364755.
[2] International Centre for the Study of the Preservation and Restoration of Cultural Property, *SDG 3.2: End All Preventable Deaths Under 5 Years of Age*. [Online]. Available: https://ocm.iccrom.org/sdgs/sdg-3-good-health-and-well-being/sdg-32-end-all-preventable-deaths-under-5-years-age
[3] World Health Organization, ‘SDG Target 3.2: End preventable deaths of newborns and children under 5 years of age’, The Global Health Observatory. Accessed: Mar. 23, 2024. [Online]. Available: https://www.who.int/data/gho/data/themes/topics/sdg-target-3_2-newborn-and-child-mortality
[4] D. Ayres-de-Campos and J. Bernardes, ‘Cardiotocography’. UC Irvine Machine Learning Repository, 2010. doi: 10.24432/C51S4N.
[5] M. S. Shelke et al., ‘A Review on Imbalanced Data Handling Using Undersampling and Oversampling Technique’, *IJRTER*, vol. 3, no. 4, pp. 444–449, May 2017, doi: 10.23883/IJRTER.2017.3168.0UWXM.
[6] scikit-learn developers, ‘Cross Validation’, Scikit-learn. Accessed: Apr. 24, 2024. [Online]. Available: https://scikit-learn.org/stable/modules/cross_validation.html
[7] T. Chen and C. Guestrin, ‘XGBoost: A Scalable Tree Boosting System’, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco California USA: ACM, Aug. 2016, pp. 785–794. doi: 10.1145/2939672.2939785.
[8] M. Dash and P. W. Koot, ‘Feature Selection for Clustering’, in *Encyclopedia of Database Systems*, L. Liu and M. T. Özsu, Eds., Boston, MA: Springer US, 2009, pp. 1119–1125. doi: 10.1007/978-0-387-39940-9_613.
[9] X. Ying, ‘An Overview of Overfitting and its Solutions’, *J. Phys.: Conf. Ser.*, vol. 1168, p. 022022, Feb. 2019, doi: 10.1088/1742-6596/1168/2/022022.
[10] A. Bravais, *Analyse mathématique sur les probabilités des erreurs de situation d’un point.* Impr. Royale, 1844.
[11] scikit-learn developers, ‘Mutual Information Score’, scikit-learn. Accessed: Apr. 23, 2024. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.mutual_info_score.html
[12] N. V. Chawla et al., ‘SMOTE: Synthetic Minority Over-sampling Technique’, *Journal of Artifical Intelligence Research*, vol. 16, pp. 321–357, Jun. 2002.
[13] D. M. W. Powers, ‘Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation’. arXiv, Oct. 10, 2020. Accessed: Apr. 24, 2024. [Online]. Available: http://arxiv.org/abs/2010.16061
[14] C. J. Van Rijsbergen, *Information Retrieval*, 2nd ed. Butterworth-Heinemann, 1979.
[15] K. H. Brodersen et al., ‘The Balanced Accuracy and Its Posterior Distribution’, in *2010 20th International Conference on Pattern Recognition*: IEEE, Aug. 2010, pp. 3121–3124. doi: 10.1109/ICPR.2010.764.