MASSEY UNIVERSITY 00 (2023) 000–000

Available online at www.sciencedirect.com

SciVerse ScienceDirect

Journal homepage: www.elsevier.com/locate/rgo

Utilising TabTransformer for Predicting Cross-Selling of Insurance Products

Evie Cheung[a], Shawn He[a], Le Ma[a], Gal Fybish[a]

[a]*Massey University, Auckland 0630, New Zealand*

ARTICLE INFO

Article history:
Received 00 December 00
Received in revised form 00 January 00
Accepted 00 February 00

Keywords:
TabTransformer
Tabular Data
Insurance
Cross-selling

ABSTRACT

Cross-selling of products presents opportunities for insurance companies to increase revenue and improve clients' retention. While enterprises in various industries have successfully implemented algorithm-based recommendation systems for cross-selling, companies in the insurance sector face several unique challenges, such as imbalanced data sets, small sets of potential products, and uncertainty regarding the history of lost opportunities. Previous research in this field used machine learning approaches such as Decision Tree-based approaches, Naïve Bayes Classifiers, Logistic Regression, and Artificial Neural Networks to predict which clients are more inclined to cross-selling. Several transformer-based machine learning approaches have been developed to make predictions on tabular data sets common in many real-world use cases in recent years. Our research applies TabTabular, a transformers-based approach, to solve an insurance cross-selling prediction problem. Using this approach, we show a better result than previous attempts to solve this problem

1. Introduction

As algorithm-based recommendation systems caught the attention of companies across various industries, insurance companies showed particular interest in these systems due to the need to personalise their offerings to customers (Kong et al., 2022).

Insurance companies can use recommendation systems to find new clients or offer additional products or services to existing ones (cross-selling) to improve customer relationships and increase customer potential profit (Altun & Yucekaya, 2021). In addition, cross-selling can reduce the risk of clients leaving (Sidorowicz et al., 2022) and revenue loss (Desirena et al., 2019).

When applying recommendation systems in the insurance domain, particular challenges arise, such as the small set of applicable products, data imbalance, and lack of certainty of past lost opportunities (Sidorowicz et al., 2022).

In addition to the industry's difficulties with data, academic research struggled to find enough good-quality data due to privacy and ownership concerns (Kong et al., 2022).

One of the few publicly available data sets was provided as part of the CoIL Challenge 2000 competition. The dataset was sourced from an insurance company, and the challenge aimed to predict the probability that a customer will purchase caravan insurance (Khalilpour Darzi et al., 2019).

A total of 43 participants submitted a solution to the challenge, with the

* *Corresponding author.* Tel.: +0-000-000-0000 ; fax: +0-000-000-0000.
E-mail address: author@institute.xxx
Peer review under responsibility of xxxxx.

Hosting by Elsevier

http://dx.doi.org/10.1016/j.rgo.2013.10.012

winning submission being able to find 51% (121 out of 238) caravan insurance policyholders in the provided evaluation set, using Naïve Bayes Classifiers (Van der Putten & van Someren, 2004).

Since the original challenge, several attempts to solve this prediction problem have been published in the academic literature. Kim et al. (2005) and Chang and Lai (2021) used Artificial Neural Networks, while Yu et al. (2016) used Logit Trees but failed to achieve a better result than the winner of the original challenge. Darzi et al. (2019) have tried many machine-learning approaches and reported an extraordinary result of finding 237 out of the 238 possible policyholders using Tree Augmented Naive Bayesian with over and under data sampling.

Our research uses the TabTransformer approach, developed by Huang et al. (2020), and set to adopt the successful transformer architecture to the realm of tabular data to try and outperform all previous attempts in solving the CoIL Challenge 2000 problem.

Our research question is as follows: RQ1 – Can using the TabTransformer approach outperform the winning model of the CoIL Challenge 2000 contest?

The remainder of the report is structured as follows. In section 2, we provide relevant literature on insurance products' cross-selling recommendations and the machine learning approaches for the research. Section 3 contains our research methodology, section 4 describes our experiment and its results, and section 5 contains our conclusions.

2. Background and Literature Review

2.1. Using machine learning models in products' cross-selling

Financial institutions often use customer demographics and product purchase history to build models that identify potential customers and efficiently enhance cross-selling opportunities (Boustani et al., 2023).

In their paper, Boustani et al. (2023) used several machine learning approaches, including Gradient Boosting, Decision Trees and Random Forests, to model demographic, product ownership, and transactional data to improve the accuracy of predicting cross-selling of consumer loans.

Decision Trees, which are tree-based machine learning algorithms, offer significant interpretability. Each data point traverses from the tree's root to a leaf node during the prediction process, following a transparent and easily interpretable path (Costa & Pedreira, 2023). This characteristic makes Decision Trees invaluable for exploratory analysis, mainly when comprehending the fundamental drivers of cross-selling.

Random Forests improve model generalisation by combining the results of multiple decision trees. It has the inherent ability to balance errors, ensuring that no single feature dominates the prediction. This property facilitates automatic feature selection. Notably, Random Forests excel in classification problems with large imbalanced datasets (Lin et al., 2017), a common challenge in the insurance sector where data imbalance is often prevalent.

In the study by Boustani et al. (2023), Gradient Boosters outperformed other models when used individually. Gradient boosting is an ensemble approach that sequentially combines weak learners to create a more robust model (Lesage et al., 2020). This technique has given rise to advanced algorithms like XGBoost, LightGBM, and CatBoost, which are renowned for their speed and accuracy (Bentéjac et al., 2021). In their comparative study, Bentéjac et al. (2021) found CatBoost to excel in generalisation accuracy and AUC among various gradient boosting and random forest algorithms. This superiority can be attributed to CatBoost's capability to handle categorical features through ordered boosting (Prokhorenkova et al., 2018), a significant advantage given the prevalence of categorical features in cross-selling prediction datasets.

Azzone et al. (2022) consider Logistic Regression as a traditional machine learning technique regularly used to solve data classification challenges, while Leiria et al. (2022) argue that logistic regression can be used to calculate the probability of a binary target outcome and has been used in insurance-related studies to predict customers' behaviour. In addition, Logistic Regression has the advantage of highly interpretable parameters (Hanafy & Ruixing, 2022).

2.2. Cross-selling recommendations for insurance products

Although recommendation systems have been thoroughly researched, the literature is scarce on applying them in the context of insurance products (Sidorowicz et al., 2022).

In recent years, a few published works have applied advanced analytical techniques to solve the problem of insurance product recommendations.

Desirena et al. (2019) investigated using Neural Networks to recommend insurance products to improve customers' Customer Lifetime Value (CLV), while Bruun (2021) employed Neural Networks to estimate the probability of a customer purchasing an insurance product based on online interactions.

Other researchers used Bayesian Networks to predict the relevance of products to an insurance company's clients (Qazi et al., 2020); meanwhile, Lesage et al. (2020) used an XGBoost model to recommend which clients will purchase additional car insurance coverage and Sidorowicz et al. (2022) used Positive Unlabelled Learning and advanced feature engineering to suggest prospective clients for cross-selling insurance products.

In contrast to the other works, which focused on using a single approach, Kong et al. (2022) compared the effectiveness of using Collaborative Filtering models versus Logistic Regression and XGBoost models to recommend insurance products. They concluded that the Collaborative Filtering models are effective when applied to existing customers, whereas the other models performed better when recommending products to new clients.

2.3. Machine learning approaches and tabular data

Data in a tabular form is frequently used in many predictive analytical use cases (Kong et al., 2023; Silva et al., 2022). In contrast to imagery or audio data, tabular data structure does not offer any positional information (Liu et al., 2023). It includes numerical columns, which may be of several different formats, and categorical columns, which may be unbalanced, thus adding complexity to the prediction process (Kong et al., 2023; Silva et al., 2022).

Tree-based ensemble approaches such as gradient-boosted decision trees are state-of-the-art approaches for analysing tabular data (Gorishniy et al., 2021; Huang et al., 2020; Sethi et al., 2023). The tree-based approaches perform better than deep learning (Shwartz-Ziv & Armon, 2022; Zheng et al., 2023) and offer high levels of interpretability (Silva et al., 2022). However, the tree-based approaches are lacking in handling noisy or incomplete data (Huang et al., 2020), which requires extensive feature engineering before they can be used (Sethi et al., 2023),

To counter these deficiencies, researchers proposed to use the multi-layer perceptron (MLP) and various deep learning approaches, which have end-to-end learning capabilities (Silva et al., 2022). However, they could not match the performance of the tree-based ensemble approaches (Huang et al., 2020). In addition, Huang et al. (2020) argue that MLP models are not interpretable and underperform in cases where the data is noisy or missing values.

2.4. TabTransformer

To overcome the performance issues of MLPs and other deep learning approaches in the tabular data domain, Huang et al. (2020) introduced, TabTransformer, an approach built on the transformer's architecture introduced by Vaswani et al. (2017).

Since their introduction, transformers have been successfully used in computer vision applications and Natural Language Processing (NLP) use cases (Xu & Zheng, 2021). Due to the successful implementation of transformers in these domains, researchers sought to adapt the transformers technology to handle tabular data (Alzahrani et al., 2022; Gorishniy et al., 2021; Sethi et al., 2023; Zheng et al., 2023).

The TabTransformer implementation of transformers uses a series of multi-head transformer layers to convert categorical features to contextual embeddings. These embeddings are, in turn, combined with normalised continuous features and then fed into a Multi-layer Perceptron to predict the output (Alzahrani et al., 2022; Huang et al., 2020; Zheng et al., 2023).

TabTransformer has been showed to achieve performance levels comparable to tree-based ensemble approaches (Xu & Zheng, 2021) while being more robust to issues stemming from missing or noisy data (Sethi et al., 2023) and having higher accuracy (Xu & Zheng, 2021).

2.5. Imbalanced datasets

Longadge & Dongre (2013) argue that dataset imbalances occur when one label has significantly more instances than another, and they are common in various real-world applications (Singh et al., 2022; Thabtah et al., 2020). Conducting traditional machine learning approaches on an imbalanced dataset may lead to undesirable results due to the emphasis on the majority class (Kumar et al., 2021; Thabtah et al., 2020). Resampling and algorithmic techniques are the most used approaches to deal with imbalanced data sets.

Data resampling uses adjustments to the dataset's class ratio by oversampling the minority class or under-sampling the majority class (Khoshgoftaar et al., 2016; Shamsuddin et al., 2023). Oversampling techniques suitable to solve imbalance issues include synthetic sampling, focused oversampling, and advanced heuristic techniques such as SMOTE (synthetic minority oversampling technique). In contrast, under-sampling includes condensed or edited nearest neighbours' techniques (Kumar et al., 2021).

In algorithmic techniques, the machine learning model is adjusted to focus on the minority class (Thabtah et al., 2020). For this purpose, hybrid approaches such as balance cascade and easy ensemble and classical approaches such as thresholding and cost-sensitive learning can be used (Buda et al., 2018).

3. Materials and Methodologies

3.1. Materials

The dataset used in this research was provided to the CoIL Challenge 2000 competition by an insurance company and is accessible to the public on UC Irvin's machine learning library under the CC BY 4.0 license (Putten, 2000).

The dataset includes information on 9,822 insurance company clients divided into training and evaluation sets (5,822 and 4,000, respectively). The data consists of 85 attributes, and the 86th column, which serves as the target feature, contains a binary value denoting whether a client opted for a caravan policy. The training set is imbalanced, with only 6% of the instances having a positive value for the target feature.

The first 43 attributes in the data set are socio-demographic features derived from the customers' area of residence and are not individual characteristics. Thus, all the customers from the exact location will have the same socio-demographic values.

The rest of the attributes include details on insurance policies owned by the customers.

3.2. Methodologies

3.2.1 Algorithm description

Our research presents a machine learning workflow leveraging the TabTransformer approach tailored for tabular data.

We preprocess the CoIL Challenge 2000 dataset, where categorical variables are encoded numerically. To mitigate class imbalance, we use the SMOTE technique, ensuring a fair representation of all classes. The TabTransformer model is configured with hyperparameters, including the number of unique categorical values, the count of continuous features, embedding dimensions, and the structure of the transformer itself, such as comprising depth, attention heads, and dropout rates. The model then undergoes training over eight epochs, using the Adam optimiser and Binary Cross-Entropy with Logits loss function. Each epoch bifurcates into a training phase (optimising the model's weights) and a validation phase (evaluating performance metrics without further weight adjustments).

3.2.2 Evaluation metrics

To evaluate the effectiveness of machine learning approaches, metrics of precision, accuracy, recall, and F1-Score are usually used. However, the accuracy rate is unsuitable for assessing a model trained on an imbalanced dataset since the overweight given to the majority class creates misleading accuracy results (Bekkar et al., 2013). Therefore, we use recall, precision, and F1-Score to measure the quality and performance of our model. Following is an explanation of the evaluation metrics, as adapted from (Tharwat, 2021), with Tables 1 and 2 presenting the detailed measures.

- **Precision**: The share of predicted positive cases that are positive. Higher precision suggests fewer false positives, thus better results.
- **Recall (Sensitivity)**: The ratio of actual positive cases correctly identified as positives by the model to the total number of positive cases.
- **F1-Score**: Provides a balance between the recall and precision metrics.

Table 1 – Two-class confusion matrix.

	Actual Positive +	**Actual Negative -**
Predicted Positive +	True Positive (TP)	False Positive (FP)
Predicted Negative -	False Negative (FN)	True Negative (TN)

Table 2 – Evaluation metrics.

Measure	**Formula**
Precision	TP/ (TP + FP)
Recall	TP/ (TP + FN)
F1-score	2*Precision*Recall/ (Precision + Recall)

4. Experiment

To run the experiment, we use tab-transformer-pytorch version 0.2.6, a TabTransfomrmer implementation, and Google's colab platform with a Python 3 runtime environment.

The training and validation follow the process described in section

3.2.1. To deal with the imbalance class, we test several approaches, such as over-sampling and classical approaches, such as cost-sensitive learning. In our experiment, because of the limitation of dataset size, SMOTE can generate synthetic examples that are similar, but not identical, to the existing instances in the minority class, which proves to be the best strategy in our experiment.

The training process provides a log of training and validation losses, augmented by validation accuracy, AUC-ROC scores, and ROC curves (see Fig. 1) after each epoch. We observe the training and validation losses decrease and AUC-ROC scores increase to ensure the model is learning and generalising rather than overfitting. In our experiment, training and validation losses and AUC-ROC scores show that the training loop should be stopped at eight epochs. This process ensures a robust evaluation of the model's predictive capabilities in the CoIL Challenge 2000 dataset.

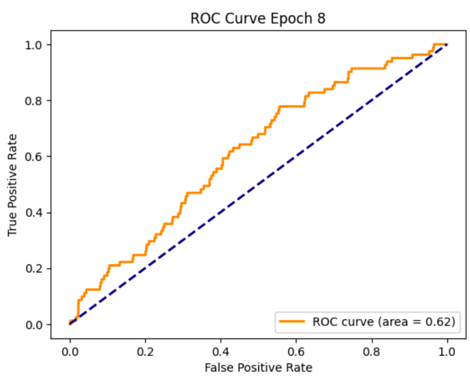

Fig. 1 – ROC curve in epoch 8.

After the training and validation, the CoIL Challenge 2000 test data set undergoes the same data preprocessing. The test set is a separate dataset with 3,762 negative and 238 positive samples.

Table 3 presents the two-class confusion matrix created by the prediction run on our test set, with 142 instances of the target variable's positive class predicted correctly. Table 4 presents the classification report with our pre-defined evaluation metrics.

Table 3 –Confusion matrix for CoIL Challenge 2000 test set.

		Predicted	
		0	**1**
Actual	**0**	2,279	1,483
	1	96	142

Table 4 –Classification report for CoIL Challenge 2000 test set.

	Precision	Recall	F1-score	Support
0	0.96	0.61	0.74	3,762
1	0.09	0.60	0.15	238
accuracy			0.61	4,000
macro avg	0.52	0.60	0.45	4,000
weighted avg	0.91	0.61	0.71	4,000

5. Conclusions

This research aims to outperform the winning model in the CoIL Challenge 2000 competition by using the TabTransformer approach, known for its effectiveness with tabular data sets, to predict which customers will purchase caravan insurance. Our experiment correctly found 142 true positives out of the possible 238, compared to the winner's 121 in the CoIL Challenge 2000. This result demonstrates the fulfilment of our research goal with a positive answer for RQ1.

From the classification report (Table 4), we observe that the precision score of the positive class is low. Future work can include using different data preprocessing such as a more sophisticated encoding technique and hyperparameter tuning techniques to improve the precision score while keeping a high level of correct prediction of the positive class.

References

Altun, Y., & Yucekaya, A. (2021). A probabilistic approach to maximize cross-selling revenues of financial products. *American Scientific Research Journal for Engineering, Technology, and Sciences*, *79*(1), 1–14.

Alzahrani, A. I. A., Al-Rasheed, A., Ksibi, A., Ayadi, M., Asiri, M. M., & Zakariah, M. (2022). Anomaly detection in fog computing architectures using custom Tab Transformer for Internet of Things. *Electronics (Switzerland)*, *11*(23). https://doi.org/10.3390/electronics11234017

Azzone, M., Barucci, E., Giuffra Moncayo, G., & Marazzina, D. (2022). A machine learning model for lapse prediction in life insurance contracts. *Expert Systems With Applications*, *191*. ScienceDirect.

Bekkar, M., Djema, H., & Alitouche, T. A. (2013). Evaluation measures for models assessment over imbalanced data sets. *Journal of Information Engineering and Applications*, *3*, 27–38.

Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, *54*(3), 1937–1967.

Boustani, N., Emrouznejad, A., Gholami, R., Despic, O., & Ioannou, A. (2023). Improving the predictive accuracy of the cross-selling of consumer loans using deep learning networks. *Annals of Operations Research*, 1–18. https://doi.org/10.1007/s10479-023-05209-5

Bruun, S. B. (2021). Learning dynamic insurance recommendations from users' click sessions. *RecSys 2021 - 15th ACM Conference on Recommender Systems*, *RecSys 2021-15th ACM Conference on Recommender Systems*, 860-863–863. https://doi.org/10.1145/3460231.3473900

Buda, M., Maki, A., & Mazurowski, M. A. (2018). A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, *106*, 249–259. https://doi.org/10.1016/j.neunet.2018.07.011

Chang, T. W., & Lai, H. K. (2021). A neural network-based approach in predicting consumers' intentions of purchasing insurance policies. *Acta Informatica Pragensia*, *10*(2), 138–154.

Costa, V. G., & Pedreira, C. E. (2023). Recent advances in decision trees: An updated survey. *Artificial Intelligence Review: An International Science and Engineering Journal*, *56*(5), 4765–4800. Springer Nature Journals. https://doi.org/10.1007/s10462-022-10275-5

Darzi, M. R. K., Niaki, S. T. A., & Khedmati, M. (2019). Binary classification of imbalanced datasets: The case of CoIL challenge 2000. *Expert Systems with Applications*, *128*, 169–186. https://doi.org/10.1016/j.eswa.2019.03.024

Desirena, G., Diaz, A., Desirena, J., Moreno, I., & Garcia, D. (2019). Maximizing customer lifetime value using stacked neural networks: An insurance industry application. *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, 541–544. https://doi.org/10.1109/ICMLA.2019.00101

Gorishniy, Y., Rubachev, I., Khrulkov, V., & Babenko, A. (2021). *Revisiting Deep Learning Models for Tabular Data*. arXiv.

Hanafy, M., & Ruixing, M. (2022). Classification of the insureds using integrated machine learning algorithms: A comparative study. *Applied Artificial Intelligence*, *36*(1), 1–32. Business Source Complete. https://doi.org/10.1080/08839514.2021.2020489

Huang, X., Khetan, A., Cvitkovic, M., & Karnin, Z. (2020). *TabTransformer: Tabular data modeling using contextual embeddings*. arXiv.

Khalilpour Darzi, M. R., Niaki, S. T. A., & Khedmati, M. (2019). Binary classification of imbalanced datasets: The case of CoIL challenge 2000. *Expert Systems with Applications*, *128*, 169–186. https://doi.org/10.1016/j.eswa.2019.03.024

Khoshgoftaar, T. M., Fazelpour, A., Dittman, D. J., & Napolitano, A. (2016). Ensemble vs. Data sampling: Which option is best suited to improve classification performance of imbalanced bioinformatics data? *Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI*, *2016-January*, 705–712. https://doi.org/10.1109/ICTAI.2015.106

Kim, Y., Street, W. N., Russell, G. J., & Menczer, F. (2005). Customer targeting: A neural network approach guided by genetic algorithms. *Management Science*, *51*(2), 264–276. Health Business Elite.

Kong, Kim, J.-H., Moon, I.-C., Kim, W. C., Yun, W., Joo, W., & Kim, K.-K. (2022). Constructing a personalised recommender system for life insurance products with machine-learning techniques. *Intelligent Systems in Accounting, Finance and Management*, *29*(4), 242-253–253. https://doi.org/10.1002/isaf.1523

Kong, S., Cheng, W., Shen, Y., & Huang, L. (2023). AutoSrh: An embedding dimensionality search framework for tabular data prediction. *IEEE Transactions on Knowledge and Data Engineering, Knowledge and Data Engineering, IEEE Transactions on, IEEE Trans. Knowl. Data Eng.*, *35*(7), 6673–6686. IEEE Xplore Digital Library. https://doi.org/10.1109/TKDE.2022.3186387

Kumar, P., Bhatnagar, R., Gaur, K., & Bhatnagar, A. (2021). Classification of imbalanced data: Review of methods and applications. *IOP Conference Series: Materials Science and Engineering*, *1099*(1), 012077. https://doi.org/10.1088/1757-899X/1099/1/012077

Leiria, M., Rebelo, E., & deMatos, N. (2022). Measuring the effectiveness of intermediary loyalty programmes in the motor insurance industry: Loyal versus non-loyal customers. *European Journal of Management and Business Economics*, *31*(3), 305–324. Emerald Insight. https://doi.org/10.1108/EJMBE-05-2020-0103

Lesage, L., Deaconu, M., Lejay, A., Meira, J. A., Nichil, G., & State, R. (2020). A recommendation system for car insurance. *European Actuarial Journal*, *10*(2), 377–398. https://doi.org/10.1007/s13385-020-00236-z

Lin, W., Wu, Z., Lin, L., Wen, A., & Li, J. (2017). An ensemble random forest algorithm for insurance big data analysis. *IEEE Access, Access, IEEE*, *5*, 16568–16575. IEEE Xplore Digital Library. https://doi.org/10.1109/ACCESS.2017.2738069

Liu, J., Tian, T., Liu, Y., Hu, S., & Li, M. (2023). iTabNet: An improved neural network for tabular data and its application to predict socioeconomic and environmental attributes. *Neural Computing and Applications*, *35*(15), 11389–11402. Springer Nature Journals. https://doi.org/10.1007/s00521-023-08304-7

Longadge, R., & Dongre, S. (2013). *Class imbalance problem in data mining review*. arXiv. https://arxiv.org/abs/1305.1707

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). Catboost: Unbiased boosting with categorical features. *Advances in Neural Information Processing Systems*, *December 2018*, 6638–6648.

Putten, P. (2000). *Insurnace Company Benchmark (COIL 2020)* [dataset]. UCI Machine Learning Repository. https://doi.org/10.24432/C5630S

Sethi, A., S. Gupta, A. Agarwal, N. Agrawal, & S. Asthana. (2023). Auto-TabTransformer: Hierarchical transformers for self and semi supervised learning in tabular data. *2023 International Joint Conference on Neural Networks (IJCNN)*, 1–8. https://doi.org/10.1109/IJCNN54540.2023.10192021

Shamsuddin, S. N., Ismail, N., & Nur-Firyal, R. (2023). Life insurance prediction and its sustainability using machine learning approach. *Sustainability (Switzerland)*, *15*(13). https://doi.org/10.3390/su151310737

Shwartz-Ziv, R., & Armon, A. (2022). Tabular data: Deep learning is not all you need. *Information Fusion*, *81*, 84–90. ScienceDirect.

Sidorowicz, T., Peres, P., & Li, Y. (2022, July 18). A novel approach for cross-selling insurance products using positive unlabelled learning. *2022 International Joint Conference on Neural Networks (IJCNN), Neural Networks (IJCNN), 2022 International Joint Conference On.* https://doi.org/10.1109/IJCNN55064.2022.9892762

Silva, A. A., Xavier, A. S., Macedo, D., Zanchettin, C., & Oliveira, A. L. I. (2022). An adapted GRASP approach for hyperparameter search on deep networks applied to tabular data. *2022 International Joint Conference on Neural Networks (IJCNN), Neural Networks (IJCNN), 2022 International Joint Conference On*, 1–8. IEEE Xplore Digital Library. https://doi.org/10.1109/IJCNN55064.2022.9892288

Singh, A., Ranjan, R. K., & Tiwari, A. (2022). Credit card fraud detection under extreme imbalanced data: A comparative study of data-level algorithms. *Journal of Experimental and Theoretical Artificial Intelligence*, *34*(4), 571–598. https://doi.org/10.1080/0952813X.2021.1907795

Thabtah, F., Hammoud, S., Kamalov, F., & Gonsalves, A. (2020). Data imbalance in classification: Experimental evaluation. *Information Sciences*, *513*, 429–441. ScienceDirect.

Tharwat, A. (2021). Classification assessment methods. *Applied Computing and Informatics*, *17*(1), 168–192. Directory of Open Access Journals. https://doi.org/10.1016/j.aci.2018.08.003

van der Putten, P., & van Someren, M. (2004). A Bias-Variance analysis of a real world learning problem: The CoIL Challenge 2000. *Machine Learning*, *57*(1–2), 177–195. Springer Nature Journals. https://doi.org/10.1023/b:mach.0000035476.95130.99

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (Vol. 30). Curran Associates, Inc.

Wang, P. (2023). *Tab-transformer-pytorch* (0.2.6) [Computer software]. https://github.com/lucidrains/tab-transformer-pytorch/

Xu, X., & Zheng, X. (2021). Hybrid model for network anomaly detection with gradient boosting decision trees and tabtransformer. *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 8538–8542. https://doi.org/10.1109/ICASSP39728.2021.9414766

Yu, P. L. H., Lee, P. H., Cheung, S. F., Lau, E. Y. Y., Mok, D. S. Y., & Hui, H. C. (2016). Logit tree models for discrete choice data with application to advice-seeking preferences among Chinese Christians. *Computational Statistics*, *31*(2), 799–827. https://doi.org/10.1007/s00180-015-0588-4

Zheng, Q., Peng, Z., Dang, Z., Zhu, L., Liu, Z., Zhang, Z., & Zhou, J. (2023). Deep tabular data modeling with dual-route structure-adaptive graph networks. *IEEE Transactions on Knowledge and Data Engineering, Knowledge and Data Engineering, IEEE Transactions on, IEEE Trans. Knowl. Data Eng.*, *35*(9), 9715–9727. IEEE Xplore Digital Library. https://doi.org/10.1109/TKDE.2023.3249186