Machine Learning in DeFi: Credit Risk Assessment and Liquidation Prediction

Abstract—**This paper investigates the application of Machine Learning for credit risk assessment in Multichain Decentralized Finance (DeFi). With DeFi expanding its scope, the need for effective credit risk evaluation becomes paramount. Our study utilizes a diverse dataset gathered from multiple blockchains, including Ethereum, and employs rigorous data preprocessing techniques. DeFi-specific features are extracted, capturing transaction-related statistics. Machine learning models, such as Logistic Regression, Random Forest, XGBoost, CatBoost, LightGBM and a CNN, are deployed to predict wallet liquidations. Evaluation metrics, including accuracy, ROC curve and Area Under the Curve, demonstrate the efficacy of DeFi-related features in credit risk assessment. Furthermore, we analyze feature importance and inter-feature correlations, providing insights into critical risk factors within the DeFi ecosystem. This research contributes valuable insights to the DeFi landscape, offering data-driven approaches to credit risk management and investment strategies. Our findings hold significance for DeFi stakeholders seeking to navigate the evolving financial frontier while mitigating credit risk effectively.**

Index Terms—**Supervised Machine Learning, Credit Score, Multichain, Ethereum, Decentralized Finance, Blockchain.**

I. Introduction

Since the emergence of Blockchain as the accounting method for Bitcoin [1], technologies and ideas evolved, and with the development of the Ethereum project new solutions are offered for various application domains [2] by enabling the execution of smart contracts [3]. One of the most promising applications is DeFi [4], and in this ecosystem, any entity can deploy a financial protocol, by implementing the respective smart contracts and deploying it on the Ethereum blockchain network [5]. Users can interact with lending pools such as Aave [6] and Maker [7], Automated Market Maker exchanges, stablecoins, derivatives, and asset management platforms. Smart contracts are now also available on a variety of other blockchains, which we have included in this research.

In DeFi blockchain-based lending, lenders supply assets to a lending smart contract, and borrowers provide collateral to secure cryptocurrency loans. These loans are typically over-collateralized, meaning the collateral's value exceeds the loan amount, to mitigate default risks. This over-collateralization allows borrowers to leverage their assets. When the collateral value in a lending system falls below a specified threshold, the corresponding debt position becomes eligible for liquidation. During a liquidation event, a liquidator settles the outstanding debts and is compensated with a part of the position's collateral as a reward [5]. Additionally, attacks including many transactions have been detected [8], including Pump Attacks and Arbitrage, Oracle Manipulation Attacks, often facilitated by Flash Loans (instantaneous access to massive capitals), yielding an ROI beyond 500% [9].

In the DeFi space, liquidations are a common practice. Studies have shown that throughout 2020 and 2021, liquidators accrued financial gains exceeding 800 million from conducting liquidations, making creditworthiness a number one issue to be resolved [10]. Furthermore, prior research has shown that Machine Learning (ML) methods, such as Boosting models and the random forest algorithm, outperform traditional credit evaluation techniques in terms of accuracy [11].

This work aims in developing a new approach to calculate the borrowers' trustworthiness based on their possibility of being liquidated by taking into account their historic transactional data and leveraging ML algorithms for liquidation prediction.

Contribution: Overall our paper makes the following contributions:

1) **DeFi credit risk assessment model:** We propose a system that leverages AI methods to develop an intelligent credit rating system for DeFi users and protocols by analyzing users' activity and recognizing patterns, leveraging digital, decentralized transaction data, giving emphasis on a user's behavior in the DeFi ecosystem, while traditional credit scoring methods (e.g. FICO) rely on centralized financial transactions and behaviors.
2) **Machine Learning Benchmarking for DeFi:** Our work benchmarks various ML models in the context of DeFi, establishing a foundational understanding for future related studies and applications in this domain.

The rest of this paper is structured as follows: In Section II we present the research baseline including related work. Section III provides an overview of the methodology followed for data collection, pre-processing, and feature extraction as well as the ML algorithms used in our study. Section IV offers an evaluation of our model's effectiveness in DeFi credit scoring, while Section V concludes with brief remarks on potential avenues for future research.

II. Related Work

The evolution of credit scoring methodologies has been significantly influenced by advancements in ML, yielding more precise and predictive capabilities.

Mariam Ossama Kotb [13] proposes an innovative credit scoring system combining ML and blockchain technology. Addressing the limitations of traditional credit scoring methods, this study employs ML algorithms like Logistic Regression (LR), XGBoost, and Random Forest (RF), integrated with

TABLE I: Description and details on the DeFi events

Event	Description	Events Amount
Borrow	Taking a loan from a DeFi protocol. This process involves overcollateralization since users are typically required to provide as a collateral assets worth more the loan.	8,5M
Deposit	Depositing assets into a DeFi protocol for various purposes such as earning interest or as a collateral for borrowing [12]	27,6M
Flashloan	Due to the atomicity of blockchain transactions, lenders can offer flash loans, i.e., loans that are only valid within one transaction and must be repaid by the end of that transaction. This concept has lead to a number of attack possibilities and arbitrage [9]	82,4K
Liquidation	In the context of lending protocols, a borrower defaults on a loan when the value of the locked collateral drops below some fixed liquidation threshold. The liquidation thresholds vary between asset markets across different protocols. In an event of default, the lending protocol seizes and liquidates the locked collateral at a discount to cover the underlying debt. Additionally, a penalty fee is charged against the debt, prior to paying out the remaining collateral to the borrower. [4] [5]	225K
Repay	Paying back borrowed funds to a DeFi protocol	4,7M
Withdraw	Retrieving assets that were deposited in a DeFi protocol	13,3M

blockchain data from Aave's smart contracts. It demonstrates that the RF model excels in predicting credit scores, highlighting the potential for more transparent, secure, and fair credit assessments in the banking sector. In the same direction, Chopra et al. discovered that RF algorithm surpasses LR and K-Nearest Neighbors (KNN) in predicting loan defaulters. Additionally, RF outperformed traditional models like Support Vector Machines (SVM) in identifying the most reliable borrowers in peer-to-peer lending scenarios [14]. Regarding ensemble methods, Lenka et al. [15] conducted studies affirming the superiority of ensemble classifiers over single models in credit scoring. These findings align with Gedela's research, which specifically pointed to the efficacy of ensemble techniques such as AdaBoost and XGBoost, with decision trees as base classifiers [16].

Our method, is distinguished by its specific focus on the DeFi ecosystem and the use of a multichain dataset. This approach is more nuanced than traditional credit scoring methods, as it takes into account the unique features of blockchain transactions. Our study's comprehensive use of various ML models, combined with a detailed analysis of feature importance and correlations, provides a deep insight into DeFi credit risk factors. In contrast, the other studies either focus on traditional credit scoring enhanced by blockchain data, peer-to-peer lending scenarios, or the effectiveness of specific types of ML models (such as ensemble methods) in credit scoring. In summary, our study's unique contributions lie in its multichain approach, DeFi-specific feature extraction, comprehensive use of diverse ML models, detailed feature analysis with focus on practical implications for DeFi stakeholders.

III. Methodology

A. Data Collection

Data used for this study originate from Ethereum as well as publicly available datasets published by existing research studies and complementary publicly available web resources. Python and NodeJS, scripts were used for data retrieval and analysis and learning models are applied using the scikit-learn package [17]. For initial experiments, we filtered for the relevant data from the Google BigQuery's public Ethereum dataset [18]. As described below, the main data on the DeFi events were provided by Covalent [19].

We have also set up an Ethereum full archive node (8-core Intel i7-11700 CPU, 4.8GHz, 32GB RAM, 10TB SSD), in order to have a dedicated server for the purposes of this research, and to have a resource for running all the conducted experiments and extracting results. By utilizing the node, we have unlimited access to all the transactions that took place since the first Ethereum block in July 2015, also called the Genesis block and we were able to search for transactions including specific actions that characterize the applications and protocols of Decentralized Finance.

We collected transactions occurring between blocks 5771740 (May 5, 2019 1:10:38 PM UTC) and 36843000 (March 12, 2023 4:46:11 PM UTC), inclusive, for 23 protocols including Aave, Compound, Curve, Lido, and Yearn. Each protocol has upgraded the initial versions of their smart contracts, and in some cases multiple versions of the protocol may exist at one time. Therefore, we collected data from the set of smart contracts associated to the different protocol versions.

1) DeFi transactions dataset: We identified the smart contracts belonging to DeFi protocols. Following the data collection process described above, in this querying process for DeFi activities, we focused on addresses interacting with these DeFi protocols and the respective transactions. To this direction, we gathered all the transaction records through these contracts assembling them into the "DeFi transactions dataset", gathering more than 54,000,000 transactions for 550,000 unique entities (addresses).

2) Covalent dataset on decoded DeFi events: Covalent [19] is a new blockchain company that offers one of the most complete dataset on DeFi protocols through APIs to their clients. They have indexed more than 100B transactions across 60 blockchains and 200k smart contracts. On top, they clean and normalize the data, to make it more accessible for users. Covalent has graciously offered us to use their dataset for the research presented in this paper. The underlying dataset presented is hence one of the most advanced in the industry.

B. Data Pre-processing

Selecting and filtering: Wallet addresses that have at least one borrow event and at least ten transactions were selected, to assure that the training data is relevant and that the only wallets taken into account, would be the ones that could potentially be liquidated.

- **Temporal Aggregation:** To gain insights over time and identify patterns in the data, temporal aggregation on a monthly basis was performed. The data is hence a numerical operation of all transactions that took place in a given month.
- **Feature calculation:** For the feature determination, we aggregated on the six events: borrow, deposit, flashloan, liquidation, repay, and withdraw. For each event, we count the number of transactions per month, and then calculate the sum, mean, and standard deviation, of the

Features	Details
6	**1 calculation on 6 events:** Count of all transactions per month
18	**3 calculations on 6 events:** Sum, mean, and standard deviation of the inflowing amount in USD per month
18	**3 calculations on 6 events:** Sum, mean, and standard deviation of the outflowing amount in USD per month
42	Total number of DeFi features

TABLE II: Introducing DeFi-related features

amount in and out in USD, respectively. This results in 42 features, leading to the events depicted in Table I. In the future, also other features could be considered, such as token names, min, max, frequencies, higher granularity than monthly, or other secondary derived features.

- **Training and prediction engineering:** In order to evaluate the historical information of previous transactions of the same address, we select the first 40 months in the dataset as training data. The following 6 months are then used as the prediction targets. The predictor is a liquidation as a boolean, i.e., whether the address was liquidated or not within this 6 months timeframe.
- **Pivot Transformation:** In order to preprocess the dataset in a way it could be fed into our models, we pivot the temporal dimension of the dataset into columns, i.e., instead of 40+1 rows per wallet, we only have 1 row, but with 40x + 1 as many features due to the time dimension. This is a necessary transformation, in order to experiment with non-recurrent or non-attention models for a timeseries prediction.
- **Data Oversampling:** To address the class imbalance in the dataset, we employ an oversampling method. To mitigate this imbalance, our approach focuses on employing data-level re-sampling techniques, specifically the Synthetic Minority Over-sampling Technique (SMOTE) [20]. This technique is utilized as a preprocessing step to equilibrate the dataset prior to classification. It enhances the training set by creating synthetic samples of the minority class (liquidation events), selected based on the k-nearest neighbors, thus avoiding information loss inherent in under-sampling and the overfitting risk of simple over-sampling.

C. Feature Extraction

Table II explains the extracted features based on the interactions of entities with DeFi protocols. Moreover, feature correlation was explored as it is depicted in Figure 1. Features correlated with a correlation coefficient greater than 0.85 [21], were removed in order to avoid non-informative distribution and to decrease the model's complexity.

D. Machine Learning Algorithms

For comparison reasons, six different ML algorithms were employed for the prediction experiment: Logistic Regression (LR), Random Forest (RF), LightGBM, Catboost, XGBoost and a Convolutional Neural Network (CNN).

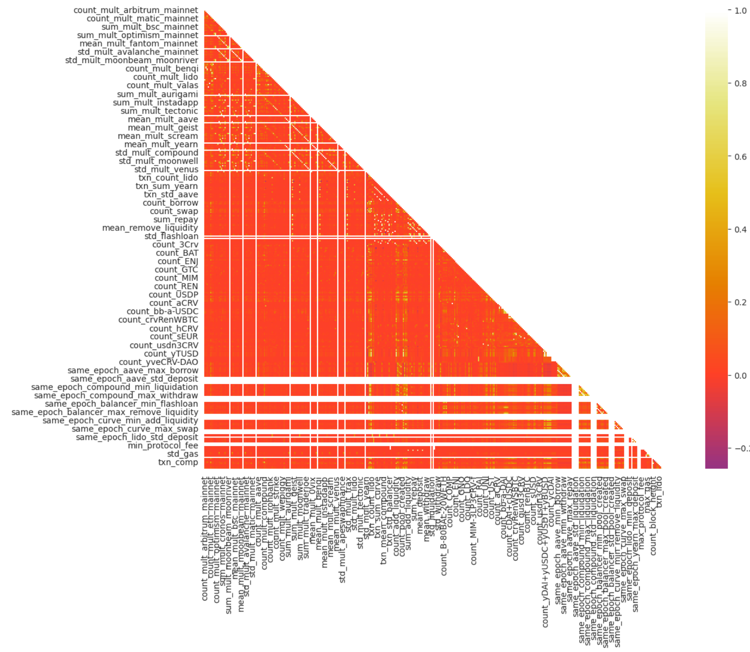

Fig. 1: Correlation

The XGBoost algorithm, known for its efficiency and effectiveness in various ML challenges [22], was selected due to its sophisticated handling of both underfitting and overfitting, which is critical in our scenario involving an imbalanced dataset. While XGBoost often demonstrates superior performance compared to other gradient boosting frameworks, it is important to consider alternatives such as LightGBM and CatBoost for their specific strengths.

LightGBM, a gradient boosting framework that is optimized for speed and efficiency [23], was chosen for its faster training times and lower memory usage compared to XGBoost. This makes LightGBM particularly suitable for our study where the dataset features a high dimensionality. Moreover, CatBoost, given its exceptional handling of categorical features [24], provides an interesting comparison point. It offers an advanced approach to processing categorical data, which is prevalent in blockchain transactions.

The inclusion of an ensemble model, namely RF, adds depth to our analysis. RF, with its approach of averaging the results of numerous decision trees [25], mitigates the risk of overfitting, a common concern in complex datasets. The ensemble nature of RF makes it robust against the noise and allows it to capture more complex patterns in the data.

Furthermore, LR serves as an essential baseline model [26]. Its simplicity and the utilization of regularization techniques (L1 or L2) [27] provide a valuable contrast to more complex models in terms of handling overfitting. The comparison with LR helps in evaluating the incremental benefits brought in by more sophisticated algorithms.

Finally, our decision to include a Convolutional Neural Network (CNN), though unconventional for tabular data, is driven by the CNN's proven capability in extracting spatial and temporal features in high-dimensional data [28]. CNNs have shown remarkable success in various domains [29], and their application to complex transaction data of DeFi could potentially unearth novel insights that other models might overlook.

- **Logistic Regression and Random Forest:** Standard implementation with sklearn.
- **XGBoost, CatBoost, LightGBM:** Standard XGBoost and LightGBM classifier from the 'xgboost' and 'lightgbm' package respectively, tuned to optimize both speed and accuracy. CatBoost classifier with its default settings, including the 'silent' mode to streamline the training process.
- **CNN:** The model has multiple layers: an input layer with 1000 neurons (ReLU activation), followed by dense layers with 400, 20, and 10 neurons. The final layer is a dense one with a single neuron (sigmoid activation). Each layer includes a 0.2 dropout rate. Training involves 100 epochs, a 256 batch size, using Adam optimizer and Mean Squared Error (MSE) for loss calculation.

IV. Experimental Results

A. Model Verification Techniques

To ascertain that our models are capable of generalizing effectively to new, unseen data, we implement the method of stratified k-fold cross-validation. This approach involves creating folds in a manner that maintains the original proportion of classes, thereby ensuring a more representative and balanced validation process.

Furthermore, to evaluate the performance of our models, two assessment methods are utilized. Firstly, we calculate the average Accuracy of the models on the dataset. Secondly, to address the class imbalance and prioritize the minority class (liquidation prediction), the ROC curve and the Area Under the Curve (AUC) [30] are being calculated. ROC curve a two-dimensional representation of classification performance, plots the true positive rate against the false positive rate. Its interpretability lies in its ability to evaluate classifier performance across various thresholds, with the AUC being a crucial measure. The AUC can be calculated discretely using the following formula:

$$\text{AUC} = \frac{1}{mn}\sum_{i=1}^{n_+}\sum_{j=1}^{n_-} 1_{f(x_i^+)>f(x_j^-)} \tag{1}$$

where $f(\cdot)$ is the scoring function, x^+ and x^- denote positive and negative samples, n_+ and n_- are the counts of positive and negative samples, and 1 is the indicator function. Those methods combined, ensure a balanced consideration of all classes in the evaluation.

B. Evaluating the performance of the classification algorithms

In this section, the effectiveness of the preprocessing, feature extraction and prediction is evaluated, with the results demonstrated in Table III. In terms of accuracy metrics for credit scoring problems, CatBoost emerges as the superior prediction algorithm. However, when considering deployment in real-world scenarios, where factors like training speed and ease of parameter tuning are crucial, alternatives like LightGBM are also viable. Despite being only marginally less effective than CatBoost (by a mere -0.01% in AUC score), LightGBM offers advantages in these areas. On the other hand, while XGBoost shows commendable performance in average accuracy, its trade-off between precision and recall (as reflected in the AUC score) is less optimal compared to CatBoost, RF, and LightGBM. Surprisingly, the performance of LR and CNN rivals that of the more sophisticated models, despite LR being primarily used as a baseline model and CNN not being the ideal fit for tabular data.

TABLE III: Average Accuracy and AUC Metrics for the Six Models

Model	Accuracy	AUC
XGBoost	0.976	0.815
CatBoost	0.975	0.847
LightGBM	0.971	0.837
RF	0.974	0.831
LR	0.942	0.777
CNN	0.926	0.756

V. Conclusion and Future Work

This paper investigates the use of ML techniques for identifying entities interacting with DeFi protocols, as marked in specialized Ethereum datasets. Our approach, focused on credit risk assessment via liquidation prediction, compiling transaction datasets from major DeFi protocols. After the DeFi-based data preprocessing and feature extraction, six machine learning models were employed for the prediction task. In the context of credit risk assessment, CatBoost distinguishes itself with its exceptional overall accuracy, especially when evaluated using the AUC metric, a critical measure in this field. However, it is important to note that other models like LightGBM and RF, also demonstrate commendable performance, closely rivaling that of CatBoost.

Nonetheless, for future work, creating more datasets for DeFi credit scoring assessments is recommended. This will allow for model refinement using advanced sequential machine learning techniques, particularly attention-based models like LSTMs and Large Language Models (LLMs). Additionally, our ML framework will be enhanced with advanced preprocessing and strategies for oversampling and undersampling.

References

[1] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," *Decentralized business review*, p. 21260, 2008.

[2] G. Palaiokrassas, P. Skoufis *et al.*, "Combining blockchains, smart contracts, and complex sensors management platform for hyper-connected smartcities: An iot data marketplace use case," *Computers*, vol. 10, no. 10, p. 133, 2021.

[3] T.-T. Davarakis, G. Palaiokrassas, A. Litke, and T. Varvarigou, "Reinforcement learning with smart contracts on blockchains," *Future Generation Computer Systems*, vol. 148, pp. 550–563, 2023.

[4] L. Gudgeon, S. Werner, D. Perez, and W. J. Knottenbelt, "Defi protocols for loanable funds: Interest rates, liquidity and market efficiency," in *Proceedings of the 2nd ACM Conference on Advances in Financial Technologies*, 2020, pp. 92–112.

[5] K. Qin, L. Zhou, P. Gamito, P. Jovanovic, and A. Gervais, "An empirical study of DeFi liquidations: incentives, risks, and instabilities," in *Proceedings of the 21st ACM Internet Measurement Conference*, ser. IMC '21. New York, NY, USA: Association for Computing Machinery, Nov. 2021, pp. 336–350. [Online]. Available: https://doi.org/10.1145/3487552.3487811

[6] Aave, "Aave – open source liquidity protocol," https://aave.com/, 2023, accessed on: December 18, 2023.

[7] MakerDAO, "MakerDAO | an unbiased global financial system, https://makerdao.com/," 2023. [Online]. Available: https://makerdao.com/en/

[8] G. Palaiokrassas, S. Scherrers, I. Ofeidis, and L. Tassiulas, "Leveraging machine learning for multichain defi fraud detection," *arXiv preprint arXiv:2306.07972*, 2023.

[9] K. Qin, L. Zhou, B. Livshits, and A. Gervais, "Attacking the defi ecosystem with flash loans for fun and profit," in *International conference on financial cryptography and data security*. Springer, 2021, pp. 3–32.

[10] D. Perez, S. M. Werner, J. Xu, and B. Livshits, "Liquidations: Defi on a knife-edge," in *Financial Cryptography and Data Security: 25th International Conference, FC 2021, Virtual Event, March 1–5, 2021, Revised Selected Papers, Part II 25*. Springer, 2021, pp. 457–476.

[11] M. Schmitt, "Deep learning vs. gradient boosting: Benchmarking state-of-the-art machine learning algorithms for credit scoring," *arXiv preprint arXiv:2205.10535*, 2022.

[12] M. Darlin, G. Palaiokrassas, and L. Tassiulas, "Debt-financed collateral and stability risks in the defi ecosystem," in *2022 4th Conference on Blockchain Research & Applications for Innovative Networks and Services (BRAINS)*. IEEE, 2022.

[13] M. O. Kotb, "Credit scoring using machine learning algorithms and blockchain technology," in *2023 Intelligent Methods, Systems, and Applications (IMSA)*. IEEE, 2023, pp. 381–386.

[14] A. Chopra and P. Bhilare, "Application of ensemble models in credit scoring models," *Business Perspectives and Research*, vol. 6, no. 2, pp. 129–141, 2018.

[15] S. R. Lenka *et al.*, "Empirical analysis of ensemble learning for imbalanced credit scoring datasets: A systematic review," 2022.

[16] B. Gedela and P. R. Karthikeyan, "Credit card fraud detection using adaboost algorithm in comparison with various machine learning algorithms to measure accuracy, sensitivity, specificity, precision and f-score," 2022.

[17] scikit learn, "scikit-learn: machine learning in Python — scikit-learn 1.2.2 documentation," 2023. [Online]. Available: https://scikit-learn.org/stable/

[18] G. BigQuery, "BigQuery Enterprise Data Warehouse," 2023. [Online]. Available: https://cloud.google.com/bigquery

[19] Covalent-API, "Covalent - One unified API. One billion possibilities." 2023. [Online]. Available: https://www.covalenthq.com/

[20] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.

[21] I. Alarab and S. Prakoonwit, "Effect of data resampling on feature importance in imbalanced blockchain data: Comparison studies of resampling techniques," *Data Science and Management*, vol. 5, no. 2, pp. 66–76, 2022.

[22] 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*. ACM, 2016, pp. 785–794.

[23] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "Lightgbm: A highly efficient gradient boosting decision tree," in *Advances in Neural Information Processing Systems*, vol. 30. Curran Associates, Inc., 2017.

[24] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "Catboost: unbiased boosting with categorical features," *Advances in Neural Information Processing Systems*, vol. 31, 2018.

[25] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[26] D. W. Hosmer Jr, S. Lemeshow, and R. X. Sturdivant, *Applied logistic regression*. John Wiley & Sons, 2013.

[27] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

[28] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[29] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[30] T. Fawcett, "An introduction to roc analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006.