Journal of
Risk and Financial Management

MDPI

Article

Crypto Exchanges and Credit Risk: Modeling and Forecasting the Probability of Closure

Dean Fantazzini [1,*] and Raffaella Calabrese [2]

[1] Moscow School of Economics, Moscow State University, Leninskie Gory, 1, Building 61, 119992 Moscow, Russia
[2] Business School, University of Edinburgh, 29 Buccleuch Place, Edinburgh EH8 9JS, UK; Raffaella.Calabrese@ed.ac.uk
* Correspondence: fantazzini@mse-msu.ru; Tel.: +7-4955105267; Fax: +7-4955105256

Abstract: While there is increasing interest in crypto assets, the credit risk of these exchanges is still relatively unexplored. To fill this gap, we considered a unique dataset of 144 exchanges, active from the first quarter of 2018 to the first quarter of 2021. We analyzed the determinants surrounding the decision to close an exchange using credit scoring and machine learning techniques. Cybersecurity grades, having a public developer team, the age of the exchange, and the number of available traded cryptocurrencies are the main significant covariates across different model specifications. Both in-sample and out-of-sample analyzes confirm these findings. These results are robust in regard to the inclusion of additional variables, considering the country of registration of these exchanges and whether they are centralized or decentralized.

Keywords: exchange; Bitcoin; crypto assets; cryptocurrencies; credit risk; bankruptcy; default probability

JEL Classification: C21; C35; C51; C53; G23; G32; G33

Citation: Fantazzini, Dean, and Raffaella Calabrese. 2021. Crypto Exchanges and Credit Risk: Modeling and Forecasting the Probability of Closure. *Journal of Risk and Financial Management* 14: 516. https://doi.org/10.3390/jrfm14110516

Academic Editors: Jeffrey Chu, Yuanyuan Zhang, Saralees Nadarajah and Stephen Chan

Received: 27 September 2021
Accepted: 19 October 2021
Published: 27 October 2021

1. Introduction

A cryptocurrency is generally defined as a digital asset designed to work as a medium of exchange, while cryptography is used to protect transactions and to control the creation of additional units of currency[1]. Over the past ten years, since the advent of Bitcoin in 2009, cryptocurrency research has become one of the most relevant topics in the field of finance, see Burniske and Tatar (2018), Fantazzini (2019), and Brummer (2019), Schar and Berentsen (2020) for more details.

Some studies show that cryptocurrencies have been used, not only as an alternative way to carry out transactions, but also as investment assets. According to Glaser et al. (2014), users view their cryptocurrency investments as speculative assets rather than a means of payment. Moreover, Baur et al. (2018) show that the largest cryptocurrency —Bitcoin— is not related to traditional asset classes, such as stocks or bonds, thus indicating the possibility of diversification. Fama et al. (2019) used the empirical strategy originally proposed by Baek and Elbeck (2015), and they found that it is more reasonable to consider Bitcoin as a highly speculative financial asset rather than a peer-to-peer cash system. Furthermore, White et al. (2020) obtained that Bitcoin is diffusing, i.e., it is a technology-based product rather than a currency, so it seems Bitcoin and other cryptocurrencies can be mostly considered as assets rather than currency. However, we should also note that some authors recently derived the fundamental value of Bitcoin as a means of payment, see Schilling and Uhlig (2019), Biais et al. (2020), Giudici et al. (2020), Chen and Vinogradov (2021), and references therein. Therefore, as of writing this paper, a clear distinction between being an asset and a payment mechanism cannot be made.

One of the most popular ways to trade and hold cryptocurrencies is by using crypto exchanges. Moore and Christin (2013) were the first to notice that traders can face the risk of crypto exchange closing down with accounts wiped out. They showed that nearly 45 percent of exchanges that opened before 2013 failed, taking the users' money with them. This result shows the need to develop models that can discriminate between safe and vulnerable exchanges. This goal is important because crypto exchanges are the most popular way to exchange fiat currencies with cryptocurrencies and vice versa, and it is therefore essential to know which exchange to use based on its security and safety profiles. Moreover, the risks of crypto exchanges may significantly contribute to the value of cryptocurrencies as assets, as the famous bankruptcy of the Mt. Gox exchange and the hacks of several exchanges highlighted, see Feder et al. (2017), Gandal et al. (2018), Chen et al. (2019), Twomey and Mann (2020), and Alexander and Heck (2020) for a detailed discussion.

Based on our knowledge, this topic has not been investigated so far. The few studies focused on this topic analyze data before 2015 (at the latest), see Moore and Christin (2013), Moore et al. (2018), and Fantazzini (2019). A quick look at CoinMarketCap[2] highlights that the total cryptocurrency market capitalization in 2021 has grown more than 400 times since 2015, with the total number of listed cryptocurrencies exceeding 10,000. Consequently, there is no doubt that the cryptocurrency market has experienced major changes over the past 6 years.

This paper aims to forecast the probability of a crypto exchange closure using previously identified factors, as well as new ones that have emerged recently. In this regard, recent IT research has suggested that, instead of focusing on specific procedures, it is better to pay attention to the overall security grade of the crypto exchange, as well as to new factors, such as the possibility of sending money to the exchange by wire transfer and/or credit card, the presence of a public developer team, etc., see Votipka et al. (2018) and Hacken Cybersecurity Services (2021) for more details. Therefore, to reach the paper's objective, we first employed a set of models to forecast the probability of closure, using a unique set of covariates (some of which were never used before), including both traditional credit scoring models and more recent machine learning models. The latter are employed because recent literature show their superiority over traditional approaches for credit risk forecasting, see Barboza et al. (2017) and Moscatelli et al. (2020) for more details.

The second contribution of this paper is a forecasting exercise, using a unique set of 144 exchanges that were active from the beginning of 2018 until the end of the first quarter of 2021. Our results show that the cybersecurity grades, having a public developer team, the age of the exchange, and the number of available traded cryptocurrencies are the main factors across several model specifications. Both in-sample and out-of-sample forecasting confirm these findings.

The third contribution of the paper is a set of robustness checks to verify that our results also hold when considering the country of registration of the crypto exchanges and whether they are centralized or decentralized.

The paper is organized as follows: Section 2 briefly reviews the (small amount of) literature devoted to the risks of exchange closure, while the methods proposed to model and forecast the probability of closure are discussed in Section 3. The empirical results are reported in Section 4, while robustness checks are discussed in Section 5. Section 6 briefly concludes.

2. Literature Review

The financial literature dealing with the credit risk involved in crypto exchanges is extremely limited and, as of writing this paper, only three works have examined the main determinants that could lead to the closure of an exchange[3].

Moore and Christin (2013) highlighted that fraudsters can hack the exchanges instead of trying to hack the cryptocurrency system directly, by taking advantage of a specific property of several cryptocurrencies (Bitcoin included): transactions are irrevocable, unlike

most payment mechanisms, such as credit cards and other electronic fund transfers, so that the fraud victims cannot get their money back after revealing the scam; see also Moore et al. (2012) for more details. In this regard, we should note that, when investing in a crypto asset, there are two types of credit risks: the possibility that the asset "dies" and the price goes to zero (or close to zero)[4], and the possibility that the exchange closes, taking most of its users' money with it. The latter is an example of counterparty risk, where the exchange may not fulfill its part of the contractual obligations. In this regard, Moore et al. (2018) examined 80 Bitcoin exchanges established between 2010 and 2015 and found that 38 have since closed: of these 38, 5 fully refunded customers, 5 refunded customers only partially, 6 exchanges did not reimburse anything, while there is no information for the remaining 22 exchanges. These numbers show that closed/bankrupt crypto exchanges imply losses given default (LGD) comparable to subordinated bonds if not public shares; see Shimko (2004) for more details about classical LGDs estimated using the data from Moody's Default Risk Service Database. The best example of the credit risk associated with crypto exchanges is likely represented by the bankruptcy of Mt. Gox in 2014. At that time, this exchange had the most traded volume worldwide (>70%); it dealt with the most important cryptocurrency (Bitcoin), and it was based in a developed country with a sophisticated and advanced legal system (Japan). Moreover, the Bitcoin price increased more than 20 times from the moment the bankruptcy was declared until the moment the available exchange assets were liquidated. Despite these premises, creditors that sued Mt. Gox (not all of them did) will probably be refunded according to the price in April 2014, but it is not clear when, due to competing (and conflicting) legal claims, see the full Reuters and Bloomberg reports by Harney and Stecklow (2017) and Leising (2021), respectively, for more details.

Moore and Christin (2013) first used a Cox proportional hazards model to estimate the time it takes for Bitcoin exchanges to close down, and to discover the main variables that can affect the closure. They found that exchanges that processed more transactions were less likely to shut down, whereas past security breaches and an anti-money laundering indicator were not statistically significant. Secondly, they ran a separate logistic regression to explain the probability that a crypto exchange experienced a security breach, and they found that a higher transaction volume significantly increased this probability, while the age of the exchange was not significant.

Moore et al. (2018) extended the work by Moore and Christin (2013), by considering data between 2010 and March 2015, and up to 80 exchanges. They built quarterly indicators and estimated a panel logit model with an expanded set of explanatory variables. They found that a security breach increases the odds that the exchange will close the same quarter, while an increase in the daily transaction volume significantly decreases the probability that the exchange will shut down that quarter. Interestingly, they found that exchanges that get most of their transaction volume from fiat currencies traded by few other exchanges are 91% less likely to close than other exchanges that trade fiat currencies with higher competition. Moreover, they reported a significant negative time trend decreasing the probability of closure over time, thus implying that the quality of crypto exchanges may be improving. Instead, an anti-money laundering indicator and the two-factor authentication were not significant, similar to what was reported by Moore and Christin (2013).

Fantazzini (2019) showed that crypto exchanges belong to a large 'family' known as small and medium-sized enterprises (SMEs), which represent the vast majority of businesses in most countries. Credit risk management for SMEs is a challenging process due to a lack of data and poor financial reporting; see the report by the European Federation of Accountants (Federation des Experts Comptables Europeens (2005)) for a specific analysis of this problem, the textbooks by Ketz (2003) and Hopwood et al. (2012) for a larger discussion about financial frauds, while Reurink (2018) provides a recent literature review. Given this background and using the dataset by Moore and Christin (2013), Fantazzini (2019) proposed several alternative approaches to forecast the probability of closure of a crypto exchange, ranging from credit scoring models to machine learning methods. However, intensive in-sample and out-of-sample forecasting analyzes were not performed

and the dataset used is now almost ten years old, thus reflecting a completely different market for crypto assets.

Therefore, given the past literature and professional practice, we expect that older exchanges should have a larger experience in terms of system security and a larger user base providing higher transaction fees, which should result in a smaller probability of closure. Similarly, the possibility to send money to the exchange by wire transfer and/or credit card should highlight a higher security level and, thus, a lower probability of default. Moreover, a mature and experienced exchange should be transparent, and the team running it should be composed of accountable individuals with identities publicly available. Furthermore, crypto exchanges with higher overall security grades are expected to show a lower probability of closure, whereas exchanges with a smaller number of tradable assets and a smaller volume of transaction fees may have less funding for the exchange security and thus a higher probability of closure. Finally, a past security breach should increase the probability that the exchange will close or go bankrupt.

3. Materials and Methods

To analyze the determinants behind the decision of closing an exchange, we consider the two main approaches: credit scoring models and machine learning. The literature on credit scoring models is pretty large Baesens and Van Gestel (2009), Joseph (2013). Machine learning techniques have been extensively used in finance; see James et al. (2013), De Prado (2018) and Dixon et al. (2020). Another important contribution of this paper involves comparing the classification accuracy of credit scoring models and machine learning techniques. To do so, we briefly review the models that will be used in the empirical analysis in this section. We remark that our paper employs credit scoring and machine learning models to estimate the probability of closure of crypto exchanges with a cross-sectional dataset. Some of these models could be used for time series forecasting and portfolio management with crypto assets; see Borges and Neves (2020); Sebastião and Godinho (2021), and references therein for more details.

3.1. Credit Scoring Models

Scoring models employ statistical techniques to combine different variables into a quantitative score. Depending on the model, the score can be either interpreted as a probability of default (PD), or used as a classification system. In the former case, a scoring model takes the following form:

$$PD_i = \mathcal{P}(D_i = 1 | D_i = 0; \mathbf{X}_i) = F(\beta' \mathbf{X}_i)$$

where PD_i is the probability of default for the firm i (in our case, a crypto exchange), and $\mathbf{X}$ is a vector of financial ratios or indicators of various kind. If we use a *logit model*, $F(\beta' \mathbf{X}_i)$ is given by the logistic cumulative distribution function,

$$F(\beta' \mathbf{X}_i) = \frac{1}{1 + e^{-(\beta' \mathbf{X}_i)}} \tag{1}$$

The maximum likelihood method is usually used to estimate the parameters vector β in Equation (1), see McCullagh and Nelder (1989) for more details. The logit model is the widely used benchmark for scoring models, because it often shows a good performance in out-of-sample analysis, see Fuertes and Kalotychou (2006), Rodriguez and Rodriguez (2006), Fantazzini and Figini (2008), Fantazzini and Figini (2009), and references therein.

The *linear discriminant analysis* (LDA) proposed by Fisher (1936) uses a set of variables to find a threshold able to separate the reliable firms from insolvent ones. LDA builds a linear combination of these variables for the two populations of firms (alive and defaulted), with the weights chosen to maximize the average distance between the two populations. Once the weights are computed, the observations of the different variables are transformed into a single score for each firm, which is then used to classify the firm based on the distance

of the score from the average scores for the two populations. The variables of the two groups must be distributed as a multivariate normal with the same variance-covariance matrix.

If we have a set of n variables $\mathbf{X}$, the group of alive firms will be separated from the group of defaulted firms based on a discriminating function of this type:

$$Z = \mathbf{a}'\mathbf{X}$$

where Z is the so-called Z-Score, $\mathbf{a}$ is the vector of discriminant coefficients (weights), and the average values for the two groups (defaulted and not defaulted) are $E(\mathbf{a}'\mathbf{X}) = \mathbf{a}'\bar{\mathbf{X}}_\mathbf{1}$ and $E(\mathbf{a}'\mathbf{X}) = \mathbf{a}'\bar{\mathbf{X}}_\mathbf{2}$. The best discriminant function is found by choosing $\mathbf{a}$, so that the squared distance between the sample means of the two groups weighted by the variance/covariance matrix $\mathbf{\Sigma}$ is the maximum:

$$\max_{\mathbf{a}} \quad d = \frac{(\mathbf{a}'\bar{\mathbf{X}}_\mathbf{1} - \mathbf{a}'\bar{\mathbf{X}}_\mathbf{2})^2}{\mathbf{a}'\mathbf{\Sigma}\mathbf{a}}$$

The analytical solution of $\mathbf{a}$ is

$$\mathbf{a} = (\bar{\mathbf{X}}_\mathbf{1} - \bar{\mathbf{X}}_\mathbf{2})'\mathbf{\Sigma}^{-1}$$

while the optimal threshold is given by,

$$\bar{Z}_C = \frac{(\bar{\mathbf{X}}_\mathbf{1} - \bar{\mathbf{X}}_\mathbf{2})'\mathbf{\Sigma}^{-1}\bar{\mathbf{X}}_\mathbf{1} + (\bar{\mathbf{X}}_\mathbf{1} - \bar{\mathbf{X}}_\mathbf{2})'\mathbf{\Sigma}^{-1}\bar{\mathbf{X}}_\mathbf{2}}{2} = \frac{\bar{Z}_1 + \bar{Z}_2}{2}$$

and supposing that $\bar{Z}_1 > \bar{Z}_2$, the discriminant rule is:

$$Z_i \in \begin{cases} \text{Group 1} & if \quad Z_i > \bar{Z}_C \\ \text{Group 2} & if \quad Z_i \leq \bar{Z}_C \end{cases}$$

The Altman (1968) Z-score model is arguably the most well-known classificatory model for credit risk that uses the linear discriminant analysis, and it is still widely used nowadays; see Altman and Sabato (2007) for more details.

3.2. Machine Learning Techniques

Machine learning (ML) is a subfield of artificial intelligence that deals with the development of systems able to recognize complex patterns and make correct choices using a dataset already analyzed. We will consider methods that can be useful for forecasting the probability of closure for a set or crypto exchanges, which is a specific case of supervised learning dealing with a classification problem, where the outputs are discrete and divided into two classes. In general, supervised learning considers all the algorithms where the user provides examples of what the algorithm must learn, containing both the input data and the corresponding output value. The goal is to generate an inference function known as a "classifier" that can be used to predict an output value given a certain input.

The supervised learning algorithm known as *Support Vector Machine* (SVM) was originally developed by V. Vapnik and his team in the 1990s at the Bell AT&T laboratories; see Boser et al. (1992) and Cortes and Vapnik (1995). A SVM interprets the training data as points in space, maps them into one n-dimensional space, and builds a hyperplane to separate these data into different classes. The subsets of points which intersect the separation hyperplane are called support vectors. A classification problem mapped into a vector space can be linearly or not linearly separable. More specifically, the SVM binary classification problems can be formulated as $y = \mathbf{w}'\mathbf{\Phi}(\mathbf{x}) + \mathbf{b}$, where $\mathbf{x}_i \in \mathcal{R}^n$ are the training variables, $y_i \in \{-1, 1\}$ their corresponding labels from two classes, ϕ is the feature-space transformation function, $\mathbf{w}$ is the vector of weights, and $\mathbf{b}$ is the classification bias. The SVM looks for the optimal hyperplane that has a maximum margin between the nearest positive and negative samples, and the search is given by

$$\arg\min_{\mathbf{w},\mathbf{b}} \frac{1}{2}\|\mathbf{w}\|^2, \quad \text{subject to:} \quad y_i(\mathbf{w}'\text{Œ}(\mathbf{x}) + \mathbf{b}) \geq 1$$

If the dataset is large and/or the data are noisy, the usual optimization with the Lagrange multipliers $\boldsymbol{\alpha} = \{\alpha_i\}_{i=1,\ldots,n}$ may become computationally challenging. To deal with this issue, it is possible to introduce control parameters that allow the violation of the previous constraints, using the following dual formulation:

$$\max_{\boldsymbol{\alpha}} D(\boldsymbol{\alpha}) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j k(\mathbf{x}_i, \mathbf{x}_j)$$

$$\text{subject to:} \begin{cases} 0 \leq \alpha_i \leq C & \forall i \\ \sum_{i=1}^{n} y_i \alpha_i = 0 & \forall i \end{cases}$$

where k is the radial kernel $k(\mathbf{x}, \mathbf{y}) = \exp(-\gamma\|\mathbf{x} - \mathbf{y}\|^2)$ with parameter γ, while the parameter C is a regularization term, where small values of C determines a hyperplane with a large-margin separation and several misclassified points, and the opposite is true for large values of C. Other kernel functions can be used, but we chose the radial kernel due to its past success in dealing with non-linear decision boundaries, see Steinwart and Christmann (2008) and Hastie et al. (2009) for more details.

A *classification decision tree* is one of the approaches most commonly used in machine learning. It is similar to a reversed tree diagram that forks each time a choice is made based on the value of a single variable, or a combination of several variables. It consists of two types of nodes: non-terminal nodes, which test the value of a single variable (or a combination of variables) and have two direct branches that represent the outcome of a test; and terminal nodes (or leaves) that do not have further branches and hold a class label. The classification tree performs an exhaustive search at every step among all the possible data splitting, and the best partition is chosen to create branches that are as homogeneous as possible. This procedure continues until a predefined stopping criterion is satisfied that can be, for example, a minimum number of units beyond which a node cannot be further split. This operation is performed by optimizing a cost function, such as the the Gini index: suppose we have a classification outcome taking values $k = 1, 2, \ldots, K$, and $\hat{p}_{mk}$ represents the proportion of class k observations in node m, then the Gini index is given by

$$G = \sum_{k=1}^{K} \hat{p}_{mk}(1 - \hat{p}_{mk})$$

The Gini index is a measure of total variance across the K classes, and it also represents the expected training error if we classify the observations to class k with probability $\hat{p}_{mk}$. When the recursive algorithm ends, it is possible to classify the dependent variable in a specific class using the path determined by the individual tests at each internal node. In our case, the estimated probability of closure for a specific crypto exchange is given by the proportion of closed exchanges in the terminal node where the exchange is included. We refer to Hastie et al. (2009), Maimon and Rokach (2014) and Smith and Koning (2017) for more details about decision trees.

Decision trees have several well-known drawbacks: their performance is poor in the case of too many classes and/or relatively small datasets. They can be computationally intensive, particularly if a "pruning" procedure is required to make its structure interpretable and to avoid overfitting. Moreover, the pruning procedure may suffer from a certain degree of subjectivity and does not fully solve the problem of overfitting. Furthermore, decision trees can be highly unstable, with small changes to the dataset resulting in completely different trees. *Random forests* solve the problem of instability and overfitting of a single tree by aggregating several decision trees into a so-called "forest", where each tree is obtained by introducing a random component in their construction. More specifically, each decision tree in a forest is built using a bootstrap sample from the original data, where 2/3 of these

data are used to build a tree, while the remaining 1/3 is used as a control set, which is known as out-of-bag (OOB) data. m variables out of the original n variables are randomly selected at each node of the tree, and the best split based on these m variables is used to split the node. The random selection of variables at each node decreases the correlation among the trees in the forest so that the algorithm can deal with redundant variables and avoid model overfitting. Moreover, each tree is grown up to its maximum size and not pruned to maximize its instability, which is neutralized by the high number of trees created to have the "forest". Note that, for a given i-th exchange in the OOB control set, the forecasts are computed using a majority vote: in simple terms, the probability of closure is given by the proportion of trees voting for the closure of exchange i. This procedure is repeated for all observations in the control set, which leads to the computation of the overall OOB classification error. The main drawback of random forests is interpretability, which is not immediate as it is for decision trees. See Hastie et al. (2009) and Smith and Koning (2017) for more details about random forests.

Finally, we will also consider the random forest with conditional inference trees proposed by Strobl et al. (2007), Strobl et al. (2008), and Strobl et al. (2009), which perform better than the original random forests in case of variables of different type (both discrete and continuous). Fantazzini (2019) showed that this approach was the best among the machine learning methods used to forecast the probability of closure with the dataset collected by Moore and Christin (2013).

3.3. Model Evaluation

Several evaluation metrics can be used to compare a set of forecasting models for binary variables. These metrics usually employ a dataset different from the one used for estimation and they can be applied to all the models considered, even if they belong to different classes, see Section 5 in Giudici and Figini (2009) for a review. Given the size of our dataset, after in-sample forecasting, we will also consider the *Leave One Out Cross Validation* (LOOCV): one observation is left out for forecasting purposes, while the model is estimated using all other observations in the training dataset. This process is then repeated for all observations in the dataset. Once the predicted values for the validation dataset are computed, we can check the forecasting performance of a model using the confusion matrix by Provost and Kohavi (1998), see Table 1:

Table 1. Theoretical confusion matrix. Number of: a true positive, b false positive, c false negative, d true negative.

Observed/Predicted	Closed Exchange	Alive
Closed Exchange	a	b
Alive	c	d

In our case, the entries in the confusion matrix have the following meaning: a is the number of correct predictions that an exchange is closed/bankrupt, b is the number of incorrect predictions that an exchange is closed/bankrupt, c is the number of incorrect predictions that an exchange is open/solvent, while d is the number of correct predictions that an exchange is open/solvent. The confusion matrix is then used to compute the area under the receiver operating characteristic curve (AUC or AUROC) proposed by Metz (1978), Metz and Kronman (1980), and Hanley and McNeil (1982) for all forecasting models. The ROC curve is computed by plotting, for any probability cut-off value between 0 and 1, the proportion of correctly predicted closed/bankrupt exchanges $a/(a+b)$ on the y-axis, also known as sensitivity or hit rate, and the proportion of open/solvent exchanges predicted as closed/bankrupt exchanges $c/(c+d)$ on the x-axis, also known as false positive rate or as 1—specificity, where the latter is $d/(d+c)$. The AUC lies between zero and one and the closer it is to one the more accurate the forecasting model is, see Sammut and Webb (2011), pp. 869–875, and references therein for more details.

It is possible to show that the area under an empirical ROC curve, when calculated by the trapezoidal rule, is equal to the Mann–Whitney U-statistic for comparing distributions of values from the two samples, see Bamber (1975). DeLong et al. (1988) used this nonparametric statistic to test the equality of two or more ROC areas, and we used this test in our analysis. This method has become popular because it does not make the strong normality assumptions required in alternative approaches, such as those proposed by Metz (1978) and McClish (1989).

Even though the AUC is one of the most common measures to evaluate the discriminative power of a predictive model for binary data, it has also some drawbacks, as discussed in detail by Krzanowski and Hand (2009), p. 108. Therefore, we also computed the model confidence set (MCS) proposed by Hansen et al. (2011) and extended by Fantazzini and Maggi (2015) to binary models, to select the best forecasting models among a set of competing models with a specified confidence level. The MCS procedure selects the best forecasting model and computes the probability that the other models are indistinguishable from the best one using an evaluation rule based on a loss function that, in the case of binary models, is given by the Brier (1950) score. More specifically, the MCS approach tests at each iteration that all models in the set of forecasting models $M = M_0$ have an equal forecasting accuracy using the following null hypothesis for a given confidence level $1 - \beta$,

$$H_{0,M} = E(d_{ij}) = 0, \quad \forall i, j \in M, \quad \text{vs.} \quad H_{A,M} = E(d_{ij}) \neq 0$$

where $d_{ij} = L_i - L_j$ is the sample loss differential between forecasting models i and j and L_i stands for the loss function of model i (in our case, the Brier score). If the null hypothesis cannot be rejected, then $\widehat{M}^*_{1-\beta} = M$. If the null hypothesis is rejected, an elimination rule is used to remove the worst forecasting models from the set M. The procedure is repeated until the null hypothesis cannot be rejected, and the final set of models define the so-called model confidence set $\widehat{M}^*_{1-\beta}$.

Among the different equivalence tests proposed by Hansen et al. (2011), we briefly discuss the T-max statistic that will be used in the empirical analysis. First, the following t-statistics are computed, $t_{i\cdot} = \bar{d}_{i\cdot} / \widehat{var}(\bar{d}_{i\cdot})$, for $i \in M$, where $\bar{d}_{i\cdot} = m^{-1} \sum_{j \in M} \bar{d}_{ij}$ is the simple loss of the i-th model relative to the average losses across models in the set M, and $\bar{d}_{ij} = H^{-1} \sum_{h=1}^{H} d_{ij,h}$ measures the sample loss differential between model i and j, and H is the number of forecasts. The T-max statistic is then calculated as $T_{max} = \max_{i \in M}(t_{i\cdot})$. This statistic has a non-standard distribution that is estimated using bootstrapping methods with 2000 replications, see Hansen et al. (2011) for details. If the null hypothesis is rejected, one model is eliminated using the following elimination rule: $e_{max,M} = \arg\max_{i \in M} \left(\bar{d}_{i\cdot} / \widehat{var}(\bar{d}_{i\cdot}) \right)$.

4. Results

4.1. Data

The dataset examined in this paper was collected using four sources of information:

- CoinGecko[5]: it is a platform that aggregates information from different crypto exchanges and has a free application programming interface (API) with access to its database;
- Cybersecurity Ranking and Certification platform[6]: it is an organization performing security reviews and assessments of crypto exchanges;
- Cryptowisser[7]: it is a site specialized in comparison of different crypto exchanges, including those closed and bankrupt;
- Mozilla Observatory[8]: it is a service allowing users to test the security of a particular website.

The dataset consisted of 144 cryptocurrencies that were alive or closed between the beginning of 2018 and the first quarter of 2021. We discarded earlier data because the cryptocurrency market has changed dramatically since 2015, see also Section 4.1 in

Fantazzini and Kolodin (2020) and references therein for a discussion about structural changes in Bitcoin markets.

Safety is essential for crypto exchanges because it builds trust among users. The more customers are sure that their money is safe on a specific crypto exchange, the more they will use that crypto exchange, and this explains why several crypto exchanges try to improve their security. Moreover, in case of a security breach, a crypto exchange may be obliged to compensate users for the lost money. Consequently, security grades can affect the probability that a crypto exchange will close. Past studies focused on the presence of some peculiar security procedures, such as the two-step authentication process or a security audit, but most of these variables turned out to be not statistically significant. Therefore, following the latest professional IT research (see Hacken Cybersecurity Services (2021)), we decided to use aggregated overall grades of the exchange's cybersecurity in place of single testing procedures.

The Cybersecurity Ranking and Certification platform developed a methodology that allows assessing the overall cybersecurity grade of different exchanges. This grade depends on the results of testing procedures performed in six different categories:

- *Server security*. This category consists of testing cryptographic protocols, such as the Transport Security Layer (TLS), the Secure Sockets Layer (SSL), the Web Application Firewall (WAF) in combination with a Content Delivery Network (CDN), the Domain Name System Security Extensions (DNSSEC), Sender Policy Framework (SPF), and many others.
- *User security*. This category assesses the implementation of security measures related to the user experience, such as the two-factor authentication, CAPTCHA, password requirements, device management, anti-phishing code, withdrawal whitelist, and previous hack cases.
- *Penetration test (or ethical hacking test)*. This kind of test looks for vulnerabilities of the exchange security and how fraudsters may use them.
- *Bug bounty program*. The program aims at stimulating hackers and cybersecurity specialists to find bugs or errors in the crypto exchange software in exchange for a reward.
- *ISO 27001*. The test verifies compliance with the standard published by the International Organization of Standardization (ISO) and the International Electrotechnical Commission (IEC) that regulates information security management systems.
- *Fund insurance*. It verifies that the crypto exchange has identifiable wallets and minimum funding.

The final cybersecurity grade takes all the previous security factors into account and assigns an aggregated score between 0 and 10. It is important to note that these cybersecurity grades changed over time for most crypto exchanges, particularly for the exchanges that closed. Therefore, in the case of closed crypto exchanges, we considered the cybersecurity grades published in the periods before the closure using cache versions of the certification platform.

We also considered a second variable to measure the security of a crypto exchange using data collected from the so-called Mozilla Observatory. The Mozilla Observatory developed a grading system that allows a user to check a website's security level, with grades ranging from A+ to F. Moreover, it is possible to transform these grades into numerical variables. The grades for the crypto exchanges that are alive refer to the first quarter of 2021, while the grades for the closed crypto exchanges refer to the last quarter when they worked. Possible grades and the corresponding numerical grades are reported in Table 2[9].

Table 2. Mozilla grading chart.

Scoring Range	Grade
100+	A+
90–99	A
85–89	A−
80–84	B+
70–79	B
65–69	B−
60–64	C+
50–59	C
45–49	C−
40–44	D+
30–39	D
25–29	D−
0–24	F

Moore et al. (2018) found that a negative time trend significantly affected the probability of a crypto exchange closure. As a consequence, we included in the analysis a variable named "age" to measure the operational longevity of exchanges: in the case of alive exchanges, this variable is equal to the number of years from their foundation until the first quarter of 2021, while for closed exchanges to the number of years between their launch and their closure[10].

Moore et al. (2018) also discovered that a security breach increased the odds of an exchange closing in the same quarter. Therefore, we added a binary variable to model the case of whether the crypto exchange was hacked or not[11].

Crypto exchanges give the possibility to trade different cryptocurrencies: a higher number of available assets to trade may result in higher transaction volumes and higher incomes from fees. Thus, the number of traded cryptocurrencies may potentially decrease the probability of closure, so we added this variable in our analysis[12].

Finally, recent professional research has suggested studying whether the exchange's developer team is public or anonymous because this information can be a potential harbinger of future scams, see Digiconomist (2016), Reiff (2020), Sze (2020) for more details. A mature and experienced exchange should be transparent, and the team running it should be composed of accountable individuals. Unfortunately, it is common for scammers to create fake identities and biographies for their projects, so that is important to check whether the members of the development team and their qualifications are real. Therefore, we also added a binary variable, which is 1 if the team information is public and 0 otherwise[13]. For similar reasons, we also considered two dummy variables that are equal to 1 if the exchange supports credit card/wire transfers, respectively, and zero otherwise.

The final dataset consisted of 144 exchanges[14] active from the beginning of 2018 until the first quarter of 2021 (but they could start working before 2018): 51 exchanges closed, while 93 were still active. A brief description of the variables used in the empirical analysis is reported in Table 3.

Table 3. Description of the explanatory variables used in the analysis.

Variable	Description	Source
Closed (dep. variable)	Binary variable that is 1 if the exchange is closed and zero otherwise	CoinGecko/Cryptowisser
Wire transfer	Binary variable that is 1 if the exchange supports wire transfers and zero otherwise	Data from exchanges
Credit card	Binary variable that is 1 if the exchange supports credit card transfers and zero otherwise	Data from exchanges
Age	Age of the exchange in years	CoinGecko/Cryptowisser
Number of tradable assets	Number of cryptocurrencies traded on the exchange	Cryptowisser
Public team	Binary variable that is 1 if the exchange's developer team is public and zero otherwise	CoinGecko
CER Cyber security grade	Security grade of the exchange assigned by the CER platform. It ranges between 0 and 10	Cybersecurity Ranking and CERtification Platform
Mozilla security grade	Security grade of the exchange assigned by the Mozilla Observatory. It ranges between 0 and 100	Mozilla Observatory
Hacked	Binary variable that is 1 if the exchange experienced a security breach and zero otherwise	Data collected manually from websites, blogs, and official Twitter accounts of the exchanges

The variance inflation factors of the regressors that are reported in Table A2 and their correlation matrix in Table A3 (both of them in the Appendix A) show that collinearity is not a problem in our dataset[15]. Their box plots are reported in Figure 1.

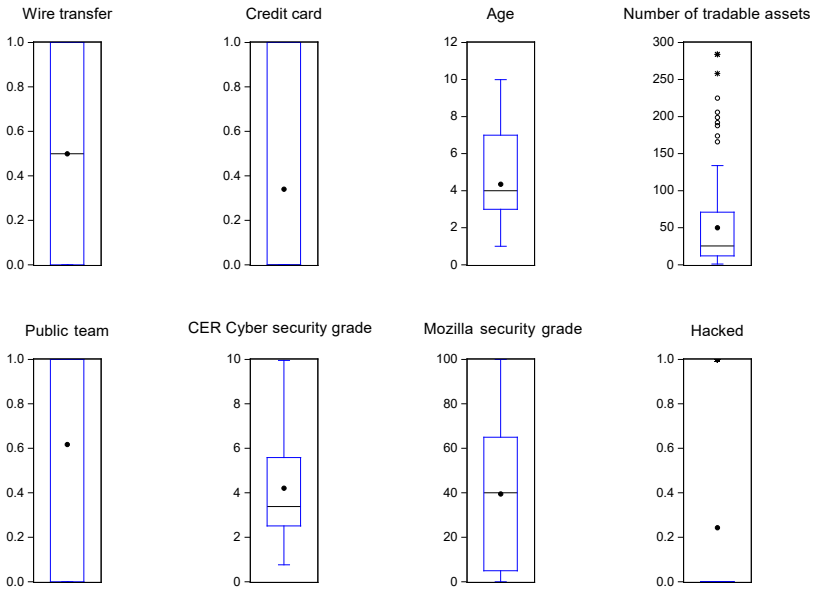

Figure 1. Box plots of the regressors.

4.2. In-Sample Analysis

Table 4 reports the results for the logit model, together with its traditional diagnostics and goodness-of-fit tests, such as the McFadden (1974) pseudo R^2, the Hosmer and Lemesbow (1980) test, the Osius and Rojek (1992) test, and the Stukel (1988) test, where the latter

two tests are robust variants of the original Hosmer and Lemesbow (1980) test, see Bilder and Loughin (2014) Section 5 for a detailed discussion at the textbook level.

Table 4. Logit model estimation results.

Variable	Estimate	Std. Error	z-Statistic	Pr(>|z|)
(Intercept)	3.51	0.82	4.30	0.00
Wire transfer	−0.98	0.54	−1.83	0.07
Credit card	−0.56	0.54	−1.03	0.30
Age	−0.22	0.13	−1.63	0.10
Number of tradable assets	−0.01	0.01	−1.32	0.19
Public team	−1.79	0.52	−3.48	0.00
CER Cyber security grade	−0.37	0.16	−2.34	0.02
Mozilla security grade	−0.00	0.01	−0.36	0.72
Hacked	0.97	0.59	1.65	0.10
McFadden R-squared:	0.38			
Hosmer-Lemeshow statistic	*p*-value:	0.14		
Osius-Rojek statistic	*p*-value:	0.01		
Stukel statistic	*p*-value:	0.17		

The logit diagnostics show a pretty good fit and the lack of major misspecification problems, while the signs of all coefficients correspond to what we expected. Interestingly, only the presence of a public team and the CER security grade are strongly significant at the 5% probability level, while the possibility of a wire transfer, the exchange age, and the presence of a security breach are only weakly significant at the 10% level. All other regressors are not statistically significant.

The estimated coefficients of the linear discriminant function that is used to classify the two response classes are reported in Table 5: the signs and sizes of the coefficients are rather similar to the coefficients of the logit model.

Table 5. LDA: Coefficients of linear discriminants.

Variable	Coefficients
Wire transfer	−0.72
Credit card	−0.30
Age	−0.11
Number of tradable assets	−0.00
Public team	−1.37
CER cyber security grade	−0.20
Mozilla security grade	−0.00
Hacked	0.51

Figure 2 reports a stacked histogram of the values of the discriminant function separately for each group (alive and closed exchanges in our case), which is a common way to display the results of a LDA: positive values are generally associated with closed exchanges, while negative values with alive exchanges.

The estimated decision tree with our dataset is reported in Figure 3.

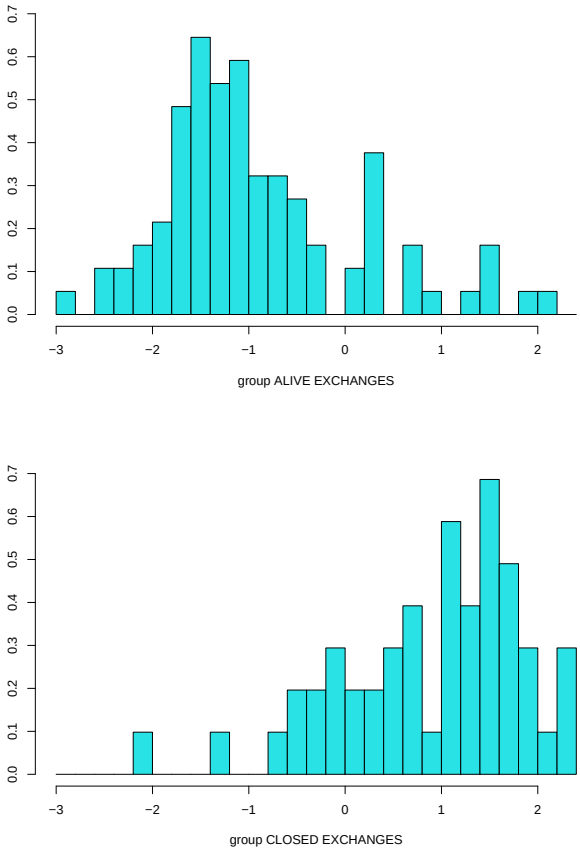

Figure 2. Stacked histogram of the scores of the discriminant function separately for each group.

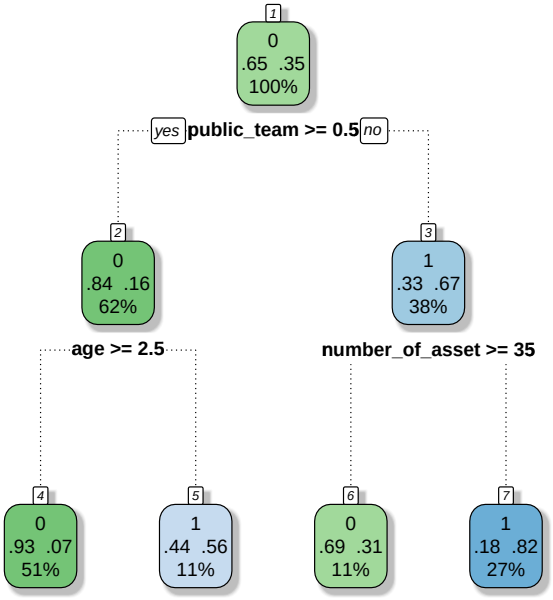

Figure 3. Estimated decision tree.

The meaning of the plot is the following: 51 exchanges closed (∼35% of the total sample), while 93 exchanges remained alive (∼65% of the total sample). In the dataset, there were 89 exchanges (∼62% of the total sample) that had a public developer team: out of these 89, 14 exchanges closed (∼16% of 89 exchanges), while 75 remained alive (∼84% of 89 exchanges). Out of the 55 exchanges (∼38% of the total sample) that did not have a

public team, 37 exchanges closed (∼67% of 55 exchanges), while 18 remained alive (∼33% of 55 exchanges). In the last row:

- 51% of exchanges (=73 exchanges) had a public team and an age bigger than 2.5 years (68 remained alive and 5 closed, 93% and 7%, respectively);
- 11% of exchanges (=16 exchanges) had a public team and an age smaller than 2.5 years (7 remained alive and 9 closed, 44% and 56%, respectively);
- 11% of exchanges (=16 exchanges) did not have a public team and they had a number of tradable assets bigger than 35 (11 remained alive and 5 closed, 69% and 31%, respectively);
- 27% of exchanges (=39 exchanges) did not have a public team and they had a number of tradable assets smaller than 35 (7 remained alive and 32 closed, 18% and 82%, respectively).

Summarizing: an exchange that has a public team, which has operated for more than 2.5 years, and which has a number of tradable assets bigger than 35 has a high probability to survive and to keep working.

Support vector machines, random forests, and conditional random forests do not have straight interpretations. To compare these models with the previous ones, we followed Fantazzini and Figini (2008) and Moscatelli et al. (2020) and we first report in Table 6 the models' AUCs together with their 95% confidence intervals for the in-sample forecasting performance, their Brier scores, and whether the models were included in the MCS or not. Table 7 reports the joint test for the equality of the AUCs estimated for all models using the test statistic proposed by DeLong et al. (1988). Finally, Table 8 reports the difference (in %) between the models' AUCs (with all variables included) and the AUCs of the same models with a specific variable excluded: this approach was proposed in Moscatelli et al. (2020) as a measure of variable importance across different models.

Table 6. AUC and 95% confidence intervals for each model, Brier scores, and model inclusion in the MCS.

Model	AUC	[AUC 95% Conf. Interval]		Brier Score	MCS
LOGIT	0.89	0.83	0.95	0.12	not included
LDA	0.89	0.83	0.94	0.13	not included
Decision Tree	0.87	0.81	0.93	0.12	not included
Random Forest	0.99	0.98	1.00	0.02	**included**
Conditional R.F.	0.95	0.92	0.98	0.11	not included
SVM	0.97	0.94	0.99	0.07	not included

Table 7. Joint test of equality for the AUCs of the six models.

H_0: *AUC(LOGIT) = AUC(LDA) = AUC(Decision Tree) = = AUC(Random Forest) = AUC(Conditional R.F.) = AUC(SVM)*	
Test statistics ($\chi^2(5)$)	25.73
p-value	0.00

Table 8. Difference (in %) between the baseline AUCs and the AUCs of the same models without a specific variable.

Excluded Variable	LOGIT	LDA	Decision Tree	Random Forest	Conditional R.F.	SVM
Wire transfer	−0.90%	−1.26%	0.00%	0.00%	−0.45%	−2.26%
Credit card	−0.40%	−0.34%	0.00%	0.00%	−0.65%	−0.61%
Age	−0.85%	−0.45%	−2.35%	−0.06%	−0.60%	−1.81%
Number of tradable assets	−0.64%	−0.24%	2.17%	−0.04%	−0.54%	−2.68%
Public team	−3.25%	−3.43%	−0.79%	0.00%	−0.63%	−2.42%
CER Cyber security grade	−1.66%	−0.98%	0.00%	0.00%	−0.67%	−1.48%
Mozilla security grade	−0.27%	−0.08%	0.00%	0.00%	−0.83%	−1.00%
Hacked	−0.79%	−0.62%	0.00%	0.00%	−0.69%	−1.79%

The random forest is the best model (but conditional R.F. and SVM are close), while the age of the exchange, the number of tradable assets, and a public developer team seem to be the most important variables to model the probability of closure. The reported high values of the AUCs were expected, given that we did in-sample forecasting with a small dataset, so that out-of-sample forecasting should give better insights about the real forecasting capabilities of the models.

4.3. Out-of-Sample Analysis

After in-sample forecasting, we implemented the leave one out cross validation (LOOCV), where one observation is left out for forecasting purposes, while the model is estimated using all other observations in the dataset. This process is then repeated for all observations in the dataset.

Table 9 reports the models' AUCs together with their 95% confidence intervals for the LOOCV forecasting performance, their Brier scores, and whether the models were included in the MCS or not. Table 10 reports the joint test for the equality of the AUCs estimated for all models using the test statistic proposed by DeLong et al. (1988), while Table 11 reports the difference (in %) between the models' AUCs (with all variables included) and the AUCs of the same models with a specific variable excluded.

Table 9. AUC and 95% confidence intervals for each model, Brier scores, and model inclusion in the MCS.

Model	AUC	[AUC 95% Conf. Interval]		Brier Score	MCS
LOGIT	0.85	0.78	0.92	0.14	not included
LDA	0.85	0.78	0.92	0.15	not included
Decision Tree	0.67	0.54	0.79	0.18	not included
Random Forest	0.90	0.85	0.95	0.12	**included**
Conditional R.F.	0.90	0.85	0.95	0.14	not included
SVM	0.89	0.84	0.94	0.13	**included**

Table 10. Joint test of equality for the AUCs of the six models.

H_0: AUC(LOGIT) = AUC(LDA) = AUC(Decision Tree) = = AUC(Random Forest) = AUC(Conditional R.F.) = AUC(SVM)	
Test statistics ($\chi^2(5)$)	21.75
p-value	0.00

Table 11. Difference (in %) between the baseline AUCs and the AUCs of the same models without a specific variable.

Excluded Variable	LOGIT	LDA	Decision Tree	Random Forest	Conditional R.F.	SVM
Wire transfer	−0.47%	−0.89%	0.00%	−0.46%	−1.62%	−2.32%
Credit card	0.05%	0.22%	0.00%	−0.36%	−0.35%	0.88%
Age	0.02%	−0.65%	0.00%	−3.71%	−2.72%	0.50%
Number of tradable assets	−0.57%	0.00%	2.27%	−2.37%	−1.67%	−4.93%
Public team	−3.89%	−4.36%	−17.73%	−5.83%	−4.93%	−4.98%
CER Cyber security grade	−2.16%	−1.66%	5.88%	−0.80%	−0.70%	−1.52%
Mozilla security grade	0.77%	0.44%	0.00%	0.49%	0.66%	0.95%
Hacked	0.32%	0.12%	0.00%	−0.35%	−0.33%	−0.97%

The performance criteria highlight that there is not a clear model that strongly outperforms the others, since they all show a similar AUC close to 85%–90%. An exception is the decision tree model that had the worst performance; thus, confirming well-known problems of model instability with small changes to the dataset. However, the MCS shows that the random forest and the SVM have significantly better forecasts than the competing models, according to the Brier score.

This empirical evidence seems to partially confirm past evidence and the theoretical discussion reported by Hand (2006), who showed that "*the marginal gain from complicated*

models is typically small compared to the predictive power of the simple models", and that "*simple methods typically yield performance almost as good as more sophisticated methods, to the extent that the difference in performance may be swamped by other sources of uncertainty that generally are not considered in the classical supervised classification paradigm*". Moreover, simple classification models may be preferred thanks to their interpretability, which may be a legal requirement in some cases (like credit scoring).

As for the main determinants of the decision of closing an exchange, a public developer team is the most important variable across all models, followed by the number of tradable crypto assets, the age of the exchange, and the CER cybersecurity grade. The evidence that a public developer team is by far the most important determinant did not come as a surprise: scammers and fraudsters alike always try to hide their identity to avoid being discovered (and prosecuted).

5. Robustness Checks

We wanted to verify that our previous results also hold with different model specifications. Therefore, we performed a series of robustness checks considering the additional information of whether the exchanges are centralized or decentralized, as well as their country of registration.

5.1. Centralized or Decentralized Exchanges: Does It Matter?

Decentralized exchanges allow for direct peer-to-peer cryptocurrency transactions without the need for an intermediary, thus reducing the risk of theft from hacking that can take place in centralized exchanges. Moreover, they can prevent price manipulation or faked trading volume through wash trading[16], and they are more anonymous than centralized exchanges that require "know your customer" (KYC) procedures[17]. However, they have also some drawbacks, such as slippage and front running; see Lin et al. (2019), Daian et al. (2020), Johnson (2021), and Alkurd (2021) for more details.

The number of decentralized exchanges in our dataset is less than 5%, so their influence on the probability of closure can be minor at best. Nevertheless, we added a binary variable to our dataset that is 1 if the exchange is decentralized and zero otherwise, and we redid our analysis due to the increasing interest in these exchanges[18]. Table 12 reports the models' AUCs together with their 95% confidence intervals for the LOOCV forecasting performance, their Brier scores, and whether the models were included in the MCS or not. Table 13 reports the joint test for the equality of the AUCs estimated for all models using the test statistic proposed by DeLong et al. (1988), while Table 14 reports the difference (in %) between the models' AUCs (with all variables included) and the AUCs of the same models with a specific variable excluded.

Table 12. AUC and 95% confidence intervals for each model, Brier scores, and model inclusion in the MCS.

Model	AUC	[AUC 95% Conf. Interval]		Brier Score	MCS
LOGIT	0.85	0.78	0.92	0.15	**included**
LDA	0.85	0.78	0.92	0.15	**included**
Decision Tree	0.67	0.54	0.79	0.18	not included
Random Forest	0.90	0.85	0.95	0.13	**included**
Conditional R.F.	0.90	0.85	0.95	0.14	**included**
SVM	0.88	0.82	0.94	0.14	**included**

Table 13. Joint test of equality for the AUCs of the six models.

H_0: *AUC(LOGIT) = AUC(LDA) = AUC(Decision Tree) = = AUC(Random Forest) = AUC(Conditional R.F.) = AUC(SVM)*	
Test statistics ($\chi^2(5)$)	20.05
p-value	0.00

Table 14. Difference (in %) between the baseline AUCs and the AUCs of the same models without a specific variable.

Excluded Variable	LOGIT	LDA	Decision Tree	Random Forest	Conditional R.F.	SVM
Wire transfer	−0.50%	−0.84%	0.00%	−1.32%	−1.72%	−1.99%
Credit card	0.15%	0.17%	0.00%	−0.15%	0.05%	1.05%
Age	−0.12%	−0.62%	0.00%	−4.11%	−2.40%	−0.74%
Number of tradable assets	−0.42%	−0.20%	2.27%	−2.01%	−0.96%	−5.13%
Public team	−4.20%	−4.51%	−13.52%	−5.40%	−5.18%	−5.25%
CER Cyber security grade	−2.39%	−1.79%	5.88%	−0.67%	−0.24%	−1.53%
Mozilla security grade	0.72%	0.42%	0.00%	0.59%	1.01%	0.91%
Hacked	0.47%	0.30%	0.00%	−0.05%	0.16%	−0.60%
Decentralized	0.17%	0.15%	0.00%	0.20%	0.24%	1.61%

The models' performances are very close, if not identical, to the baseline out-of-sample forecasting case. The only small difference is the Brier scores that are now slightly higher, so the MCS includes all models except for the decision tree model. The noise introduced by an additional insignificant regressor worsened the model performances just enough to make them no more statistically different from each other, and the MCS was unable to separate good and bad models. This outcome was expected due to the small sample size involved and the small number of decentralized exchanges present in the dataset.

5.2. Country of Registration: Does It Matter?

To verify the effect of the country of registration of crypto exchanges on their probability of closure, we followed Moore and Christin (2013) and Moore et al. (2018), and we used an index computed by World Bank economists (Yepes (2011)) to identify each country's compliance with "Anti-Money Laundering and Combating the Financing of Terrorism" (AML-CFT) regulations; see Yepes (2011) for more details.

Table 15 reports the models' AUCs together with their 95% confidence intervals for the LOOCV forecasting performance, their Brier scores, and whether the models were included in the MCS or not. Table 16 reports the joint test for the equality of the AUCs estimated for all models using the test statistic proposed by DeLong et al. (1988), while Table 17 reports the difference (in %) between the models' AUCs (with all variables included) and the AUCs of the same models with a specific variable excluded.

Table 15. AUC and 95% confidence intervals for each model, Brier scores, and model inclusion in the MCS.

Model	AUC	[AUC 95% Conf. Interval]		Brier Score	MCS
LOGIT	0.85	0.78	0.92	0.15	not included
LDA	0.85	0.78	0.92	0.15	not included
Decision Tree	0.67	0.54	0.79	0.18	not included
Random Forest	0.90	0.85	0.95	0.12	**included**
Conditional R.F.	0.90	0.84	0.95	0.14	not included
SVM	0.89	0.83	0.94	0.13	**included**

Table 16. Joint test of equality for the AUCs of the six models.

H_0: *AUC(LOGIT) = AUC(LDA) = AUC(Decision Tree) = = AUC(Random Forest) = AUC(Conditional R.F.) = AUC(SVM)*	
Test statistics ($\chi^2(5)$)	21.95
p-value	0.00

Table 17. Difference (in %) between the baseline AUCs and the AUCs of the same models without a specific variable.

Excluded Variable	LOGIT	LDA	Decision Tree	Random Forest	Conditional R.F.	SVM
Wire transfer	−0.35%	−1.11%	0.00%	−1.36%	−1.20%	−2.04%
Credit card	0.57%	0.17%	0.00%	−0.37%	0.35%	0.50%
Age	−0.07%	−0.84%	0.00%	−2.79%	−3.04%	−0.40%
Number of tradable assets	−0.65%	−0.25%	2.27%	−1.22%	−1.08%	−4.58%
Public team	−4.03%	−4.80%	−13.52%	−5.47%	−4.66%	−5.55%
CER Cyber security grade	−2.12%	−1.63%	5.88%	−1.76%	−0.73%	−1.61%
Mozilla security grade	0.67%	0.35%	0.00%	0.64%	1.01%	−0.19%
Hacked	0.10%	0.07%	0.00%	−0.21%	0.00%	0.69%
AML−CFT	0.37%	0.02%	0.00%	0.20%	0.28%	0.00%

The models' performances and the tests statistics are almost identical to the baseline out-of-sample forecasting case, thus confirming that the AML-CFT index is not a statistically significant variable as reported by Moore and Christin (2013) and Moore et al. (2018).

6. Conclusions

This paper investigated the determinants surrounding the decision to close an exchange, using a set of variables consisting of previously identified factors, and new ones that emerged from the latest professional IT research.

To reach this objective, we first proposed a set of models to forecast the probability of closure of a crypto exchange, including both traditional credit scoring models and more recent machine learning models. Secondly, we performed a forecasting exercise using a unique set of 144 exchanges that were active from the beginning of 2018 until the end of the first quarter of 2021. We found that having a public developer team is by far the most important determinant, followed by the CER cybersecurity grade, the age of the exchange, and the number of traded cryptocurrencies available on the exchange. Both in-sample and out-of-sample forecasting confirm these findings. The fact that having a public developer team is the most important factor is probably a confirmation that cryptocurrencies' returns merely depend on financial conventions and that these assets have become part of the traditional financial system, as discussed in Fama et al. (2019).

The general recommendation for investors that emerged from our analysis is to choose an exchange with a public developer team (scammers and fraudsters always try to hide), with a high CER cybersecurity grade, preferably with a working experience of several years, and with a high number of available tradable assets, which can guarantee a large volume of transaction fees and, thus, better funding for exchange security.

Finally, we performed a set of robustness checks to verify that our results also hold when considering whether the exchanges are centralized or decentralized, and when considering their country of registration by using an index to identify the country's compliance with the AML–CFT regulations. We found that the models' performances and the tests statistics were almost identical to the baseline out-of-sample forecasting case; thus, showing that the exchange being decentralized or not, and the AML–CFT index, are not statistically significant variables.

We should note that the number of exchanges that we used is rather low compared to traditional studies dealing with credit risk for SMEs, despite our analysis being the largest so far in this field of research. We are aware that this limitation may make our models suffer from a certain degree of selection bias. For example, some small exchanges were discarded from our dataset because we were unable to collect all the regressors required for our analysis: it was not possible to find information about their public team, past hacks, age, methods of money transfers, etc. However, we are confident that the addition of these exchanges, mainly small and no more working, would strengthen our results instead of weakening, because they would likely confirm the need to choose exchanges with a public team, without past hacks, and with several years of experience. The retrieval and the analysis of additional exchanges data are left as an avenue for future research.

Another possibility of future work will be to check how the credit risk for crypto exchanges will change when the number of decentralized exchanges and their trading volume increase to a more sizable level. The recent crackdown in China, where both crypto mining and transactions involving crypto assets are now fully prohibited, may stimulate the growth of decentralized exchanges. Their development may spread a form of "fully denationalized financial money" from which only a few social groups will benefit with increasing social inequalities, but it may also stimulate financial circuits that can enable a more equitable distribution of the wealth created by social cooperation, as recently discussed by Fama et al. (2019). This is why this phenomenon will have to be monitored.

Author Contributions: Conceptualization, D.F.; Methodology, D.F., R.C.; Software, D.F.; Validation, D.F., R.C.; Formal Analysis, D.F.; Investigation, D.F.; Resources, D.F.; Data Curation, D.F.; Writing—Original Draft, D.F., R.C.; Writing—Review and Editing, D.F., R.C.; Visualization, D.F., R.C.; Supervision, D.F., R.C.; Project Administration, D.F.; Funding Acquisition, D.F. All authors have read and agreed to the published version of the manuscript.

Funding: The first-named author gratefully acknowledges financial support from the grant of the Russian Science Foundation n. 20-68-47030.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: The first-named author gratefully acknowledges financial support from the grant of the Russian Science Foundation n. 20-68-47030.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Exchanges Names

Table A1. Exchanges names.

3xbit	BTSE	FTX	NLexch
6X	Bybit	Gate.io	Oceanex
Aax	C-Cex	Femini	Okcoin
Alterdice	Chainrift	Gopax	Okex
Altilly	Chilebit	HB.top	Otcbtc
Altsbit	Cobinhood	Hbtc	Paribu
Ascend	Coinegg	HBUS	Phemex
B2Bx	Coinbene	Hitbtc	Poloniex
Bancor	Coinchangex	Hoo	Poloni dex
Bankera	Coincheck	Hotbit	Probit
Bibox	Coindeal	Huobi	Purcow
Bigone	Coinex	Ice3x	Shortex
Biki	Coinfalcon	ICOCryptex	Sistemkoin
Bilaxy	Coinfloor	Indodax	Sparkdex
Binance	Coinhub	Instant Bitex	Stex
Bitbank	Coinlim	IQFinex	Stormgain
Bitbay	Coinmetro	Itbit	The PIT
Bitbox	Coinnest	Koineks	TheRockTrading
Bitfinex	Coinone	Korbit	Tidex
Bitflyer	Coinrate	Kraken	TokensNet
Bitforex	Coinsbit	Kucoin	TopBTC
Bitget	Coinsuper	Kuna	Trade Satoshi
Bithumb	Cointiger	Lakebtc	Tux exchange
Bitkub	CPDAX	Latoken	Unichange
Bitlish	Credox	Lbank	Upbit
Bitmart	Crypto Bridge	LEOxChange	Vbitex
Bitmesh	Crypto Dao	Liquid	VeBitcoin

Table A1. *Cont.*

Bitmex	CryTrEx	Livecoin	VirWox
Bitopro	Dcoin	Lukki	Wazirx
Bitpanda	Deribit	luno	Whitebit
Bitso	Dex-trade	Max Maicoin	XT
Bitstamp	Dflow	Mercado Bitcoin	Yobit
Bittrex	Digifinex	Mercatox	Zaif
Bleutrade	Exmo	Narkasa	ZB.com
BTCbear	Fcoin	Neraex	ZBG
BTCturk	Fisco	Nicehash	ZG.top

Table A2. Variance inflation factors (VIFs) of the regressors.

Wire transfer	1.36
Credit card	1.08
Age	1.27
Number of tradable assets	1.24
Public team	1.42
CER cyber security grade	1.46
Mozilla security grade	1.26
Hacked	1.09

Table A3. Correlation matrix of the regressors.

	Wire Transfer	Credit Card	Age	Number of Tradable Assets	Public Team	CER Cyber Security Grade	Mozilla Security Grade	Hacked
Wire transfer	1	0.22	0.38	−0.14	0.27	0.09	0.18	−0.15
Credit card	0.22	1	0.19	0.05	0.14	0.02	0.12	0.04
Age	0.38	0.19	1	0.10	0.26	0.03	0.13	−0.03
Number of tradable assets	−0.14	0.05	0.10	1	0.10	0.31	0.24	0.14
Public team	0.27	0.14	0.26	0.10	1	0.41	0.30	0.11
CER cyber security grade	0.09	0.02	0.03	0.31	0.41	1	0.37	−0.04
Mozilla security grade	0.18	0.12	0.13	0.24	0.30	0.37	1	0.04
Hacked	−0.15	0.04	−0.03	0.14	0.11	−0.04	0.04	1

Notes

1 This is a general definition of cryptocurrency that is based on the current practices among both financial and IT professionals, see, for example, the official technical report by the Association of Chartered Certified Accountants (ACCA (2021)), as well as the formal definition of cryptocurrency proposed by Lansky (2018), which is considered the most precise by IT specialists, and which was later adopted by Fantazzini and Zimin (2020) to formally define credit risk for cryptocurrencies. Antonopoulos (2014) and Narayanan et al. (2016) to provide a larger discussion at the textbook level.

2 https://coinmarketcap.com/charts/ (accessed on 1 August 2021). CoinMarketCap is the main aggregator of cryptocurrency market data, and it has been owned by the crypto exchange Binance since April 2020, see https://crypto.marketswiki.com/index.php?title=CoinMarketCap (accessed on 1 August 2021) for more details. Website accessed on June 15, 2021.

3 We will use the terms 'probability of closure' and 'probability of default' interchangeably.

4 This type of risk was originally defined by Fantazzini and Zimin (2020), pp. 24–26, as "*the gains and losses on the value of a position of a cryptocurrency that is abandoned and considered dead according to professional and/or academic criteria, but which can be potentially revived and revamped*".

5 https://www.coingecko.com (accessed on 1 August 2021).

6 https://cer.live (accessed on 1 August 2021).

7 https://www.cryptowisser.com (accessed on 1 August 2021).

8 https://observatory.mozilla.org (accessed on 1 August 2021).

9 https://github.com/mozilla/http-observatory/blob/master/httpobs/docs/scoring.md (accessed on 1 August 2021).

10 The dates of crypto exchange foundations were taken from CoinGecko, while the dates of closure (if any) from Cryptowisser.

11 The information about security breaches was collected manually from websites, blogs, and official Twitter accounts of the exchanges.

12 Cryptowisser reports how many cryptocurrencies are traded on each exchange.

[13] Information about the exchanges' developer team is available at CoinGecko.

[14] The names of these exchanges are reported in Table A1 in the Appendix A.

[15] The variance inflation factors (VIF) are used to measure the degree of collinearity among the regressors in an equation. They can be computed by dividing the variance of a coefficient estimate with all the other regressors included by the variance of the same coefficient estimated from an equation with only that regressor and a constant. Classical "rules of thumbs" to get rid of collinearity are to eliminate those variables with a VIF higher than 10 or to eliminate one of the two variables with a correlation higher than 0.7–0.8 (in absolute value).

[16] Wash trading is a process whereby a trader buys and sells an asset to feed misleading information to the market. It is illegal in most regulated markets, see James Chen (2021) and references therein for more details. However, there is recent evidence that up to 30% of all traded tokens on two of the first popular decentralized exchanges on the Ethereum blockchain (IDEX and EtherDelta) were subject to wash trading activity, see Victor and Weintraud (2021) for more details.

[17] The "know your customer" or "know your client" check is the process of identifying and verifying the client's identity when opening a financial account, see https://en.wikipedia.org/wiki/Know_your_customer (accessed on 1 August 2021) and references therein for more details.

[18] https://trends.google.ru/trends/explore?date=all&q=decentralized%20exchanges (accessed on 1 August 2021).

References

ACCA. 2021. *Accounting for Cryptocurrencies*. London: Association of Chartered Certified Accountants.

Alexander, Carol, and Daniel F. Heck. 2020. Price discovery in Bitcoin: The impact of unregulated markets. *Journal of Financial Stability* 50: 100776. [CrossRef]

Alkurd, Ibrahim. 2021. *The Rise of Decentralized Cryptocurrency Exchanges*. Forbes. Available online: https://www.forbes.com/sites/theyec/2020/12/01/the-rise-of-decentralized-cryptocurrency-exchanges/ (accessed on 1 August 2021)

Altman, Edward. 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance* 23: 589–609. [CrossRef]

Altman, Edward, and Gabriele Sabato. 2007. Modelling credit risk for SMEs: Evidence from the US market. *Abacus* 43: 332–57. [CrossRef]

Antonopoulos, Andreas. 2014. *Mastering Bitcoin: Unlocking Digital Cryptocurrencies*. Sevastopol: O'Reilly Media, Inc.

Baek, Chung, and Matt Elbeck. 2015. Bitcoins as an investment or speculative vehicle? a first look. *Applied Economics Letters* 22: 30–34. [CrossRef]

Baesens, Bart, and Tony Van Gestel. 2009. *Credit Risk Management: Basic Concepts*. Oxford: Oxford University Press.

Bamber, Donald. 1975. The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *Journal of Mathematical Psychology* 12: 387–415. [CrossRef]

Barboza, Flavio, Herbert Kimura, and Edward Altman. 2017. Machine learning models and bankruptcy prediction. *Expert Systems with Applications* 83: 405–17. [CrossRef]

Baur, Dirk G., Kihoon Hong, and Adrian D. Lee. 2018. Bitcoin: Medium of exchange or speculative assets? *Journal of International Financial Markets, Institutions and Money* 54: 177–89. [CrossRef]

Biais, Bruno, Christophe Bisiere, Matthieu Bouvard, Catherine Casamatta, and Albert J. Menkveld. 2020. Equilibrium Bitcoin Pricing. Available online https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3261063 (accessed on 1 August 2021).

Bilder, Christopher, and Thomas Loughin. 2014. *Analysis of Categorical Data with R*. Boca Raton: CRC Press.

Borges, Tome Almeida, and Rui Ferreira Neves. 2020. Ensemble of machine learning algorithms for cryptocurrency investment with different data resampling methods. *Applied Soft Computing* 90: 106187. [CrossRef]

Boser, Bernhard E., Isabelle M. Guyon, and Vladimir N. Vapnik. 1992. A training algorithm for optimal margin classifiers. Paper presented at the Fifth Annual Workshop on Computational Learning Theory, Pittsburgh, PA, USA, July 27–29. pp. 144–52.

Brier, Glenn. 1950. Verification of forecasts expressed in terms of probability. *Monthly Weather Review* 78: 1–3. [CrossRef]

Brummer, Chris. 2019. *Cryptoassets: Legal, Regulatory, and Monetary Perspectives*. Oxford: Oxford University Press.

Burniske, Chris, and Jack Tatar. 2018. *Cryptoassets: The Innovative Investor's Guide to Bitcoin and Beyond*. New York: McGraw-Hill.

Chen, Weili, Jun Wu, Zibin Zheng, Chuan Chen, and Yuren Zhou. 2019. Market manipulation of Bitcoin: Evidence from mining the Mt. Gox transaction network. Paper presented at the IEEE INFOCOM 2019—IEEE Conference on Computer Communications, Paris, France, April 29–May 2, pp. 964–972.

Chen, Yi-Hsuan, and Dmitri Vinogradov. 2021. *Coins with Benefits: On Existence, Pricing Kernel and Risk Premium of Cryptocurrencies*. Technical Report. Berlin: Humboldt University of Berlin, International Research Training Group 1792. Discussion Paper No. 2021-006.

Cortes, Corinna, and Vladimir Vapnik. 1995. Support-vector networks. *Machine Learning* 20: 273–97. [CrossRef]

Daian, Philip, Steven Goldfeder, Tyler Kell, Yunqi Li, Xueyuan Zhao, Iddo Bentov, Lorenz Breidenbach, and Ari Juels. 2020. Flash boys 2.0: Frontrunning in decentralized exchanges, miner extractable value, and consensus instability. Paper presented at the 2020 IEEE Symposium on Security and Privacy (SP), Francisco, CA, USA, May 17–21. pp. 910–927.

De Prado, Marcos Lopez. 2018. *Advances in Financial Machine Learning*. New York: John Wiley & Sons.

DeLong, Elizabeth, David DeLong, and Daniel Clarke-Pearson. 1988. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics* 44: 837–45. [CrossRef] [PubMed]
Digiconomist. 2016. *Introducing the Fraud Assessment Tool*. Available online: https://digiconomist.net/introducing-fat (accessed on 1 August 2021).
Dixon, Matthew F., Igor Halperin, and Paul Bilokon. 2020. *Machine Learning in Finance*. New York: Springer.
Federation des Experts Comptables Europeens. 2005. *How SMEs Can Reduce the Risk of Fraud*. Bruxelles: European Federation of Accountants (FEE) .
Fama, Marco, Andrea Fumagalli, and Stefano Lucarelli. 2019. Cryptocurrencies, monetary policy, and new forms of monetary sovereignty. *International Journal of Political Economy* 48: 174–94. [CrossRef]
Fantazzini, Dean. 2019. *Quantitative Finance with R and Cryptocurrencies*. Seattle: Amazon KDP. ISBN-13: 978-1090685315.
Fantazzini, Dean, and Silvia Figini. 2008. Default forecasting for small-medium enterprises: Does heterogeneity matter? *International Journal of Risk Assessment and Management* 11: 138–63. [CrossRef]
Fantazzini, Dean, and Silvia Figini. 2009. Random survival forests models for sme credit risk measurement. *Methodology and Computing in Applied Probability* 11: 29–45. [CrossRef]
Fantazzini, Dean, and Nikita Kolodin. 2020. Does the hashrate affect the Bitcoin price? *Journal of Risk and Financial Management* 13: 263. [CrossRef]
Fantazzini, Dean, and Mario Maggi. 2015. Proposed coal power plants and coal-to-liquids plants in the us: Which ones survive and why? *Energy Strategy Reviews* 7: 9–17. [CrossRef]
Fantazzini, Dean, and Stephan Zimin. 2020. A multivariate approach for the simultaneous modelling of market risk and credit risk for cryptocurrencies. *Journal of Industrial and Business Economics* 47: 19–69. [CrossRef]
Feder, Amir, Neil Gandal, J. T. Hamrick, and Tyler Moore. 2017. The impact of DDoS and other security shocks on Bitcoin currency exchanges: Evidence from Mt. Gox. *Journal of Cybersecurity* 3: 137–44. [CrossRef]
Fisher, Ronald A. 1936. The use of multiple measurements in taxonomic problems. *Annals of Eugenics* 7: 179–88. [CrossRef]
Fuertes, Ana-Maria, and Elena Kalotychou. 2006. Early warning systems for sovereign debt crises: The role of heterogeneity. *Computational Statistics and Data Analysis* 51: 1420–41. [CrossRef]
Gandal, Neil, J. T. Hamrick, Tyler Moore, and Tali Oberman. 2018. Price manipulation in the Bitcoin ecosystem. *Journal of Monetary Economics* 95: 86–96. [CrossRef]
Giudici, Giancarlo, Alistair Milne, and Dmitri Vinogradov. 2020. Cryptocurrencies: Market analysis and perspectives. *Journal of Industrial and Business Economics* 47: 1–18. [CrossRef]
Giudici, Paolo, and Silvia Figini. 2009. *Applied Data Mining for Business and INDUSTRY*. New York: Wiley Online Library.
Glaser, Florian, Kai Zimmermann, Martin Haferkorn, Moritz Christian Weber, and Michael Siering. 2014. Bitcoin-asset or currency? revealing users' hidden intentions. *Revealing Users' Hidden Intentions (15 April 2014). ECIS*. Available online: https://ssrn.com/abstract=2425247 (accessed on 1 August 2021).
Hacken Cybersecurity Services. 2021. *Cryptocurrency Exchange Security Assessment Methodology*. Available online: https://hacken.io/researches-and-investigations/cryptocurrency-exchange-security-assessment-methodology/ (accessed on 1 August 2021).
Hand, David J. 2006. Classifier technology and the illusion of progress. *Statistical Science* 21: 1–14. [CrossRef]
Hanley, James, and Barbara McNeil. 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143: 29–36. [CrossRef] [PubMed]
Hansen, Peter, Asger Lunde, and James Nason. 2011. The model confidence set. *Econometrica* 79: 453–97. [CrossRef]
Harney, Alexandra, and Steve Stecklow. 2017. Twice burned—How Mt. Gox's Bitcoin customers could lose again. *Reuters*, November 30.
Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer.
Hopwood, William, Jay Leiner, and George Young. 2012. *Forensic Accounting and Fraud Examination*. New York: McGraw-Hill.
Hosmer, David, and Stanley Lemesbow. 1980. Goodness of fit tests for the multiple logistic regression model. *Communications in Statistics-Theory and Methods* 9: 1043–69. [CrossRef]
James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. *An Introduction to Statistical Learning*. New York: Springer, Volume 112.
James Chen. 2021. *Wash TRADING*. Investopedia. Available online: https://www.investopedia.com/terms/w/washtrading.asp (accessed on 1 August 2021)
Johnson, Kristin N. 2021. Decentralized finance: Regulating cryptocurrency exchanges. *William & Mary Law Review* 62: 1911.
Joseph, Ciby. 2013. *Advanced Credit Risk Analysis and Management*. New York: John Wiley & Sons.
Ketz, Edward. 2003. *Hidden Financial Risk: Understanding Off-Balance Sheet Accounting*. New York: John Wiley & Sons.
Krzanowski, Wojtek, and David Hand. 2009. *ROC Curves for Continuous Data*. Boca Raton: CRC Press.
Lansky, Jan. 2018. Possible state approaches to cryptocurrencies. *Journal of Systems Integration* 9: 19–31. [CrossRef]
Leising, Matthew. 2021. CoinLab Cuts Deal With Mt. Gox Trustee Over Bitcoin Claims. *Bloomberg*, January 15.
Lin, Lindsay X., Eric Budish, Lin William Cong, Zhiguo He, Jonatan H. Bergquist, Mohit Singh Panesir, Jack Kelly, Michelle Lauer, Ryan Prinster, Stephenie Zhang, and et al. 2019. Deconstructing decentralized exchanges. *Stanford Journal of Blockchain Law & Policy* 2: 58–77.

Maimon, Oded, and Lior Rokach. 2014. *Data Mining with Decision Trees: Theory and Applications*. London: World Scientific, Volume 81.
McClish, Donna. 1989. Analyzing a portion of the roc curve. *Medical Decision Making* 9: 190–5. [CrossRef] [PubMed]
McCullagh, Peter, and John A. Nelder. 1989. *Generalized Linear Model*. London: Chapman Hall.
McFadden, Daniel. 1974. Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics* 105–42.
Metz, Charles. 1978. Basic principles of ROC analysis. In *Seminars in Nuclear Medicine*. Amsterdam: Elsevier, Volume 8, pp. 283–98.
Metz, Charles, and Helen Kronman. 1980. Statistical significance tests for binormal ROC curves. *Journal of Mathematical Psychology* 22: 218–43. [CrossRef]
Moore, Tyler, and Nicolas Christin. 2013. Beware the middleman: Empirical analysis of Bitcoin-exchange risk. In *International Conference on Financial Cryptography and Data Security*. New York: Springer, pp. 25–33.
Moore, Tyler, Nicolas Christin, and Janos Szurdi. 2018. Revisiting the risks of Bitcoin currency exchange closure. *ACM Transactions on Internet Technology 18*: 1–18. [CrossRef]
Moore, Tyler, Jie Han, and Richard Clayton. 2012. The postmodern ponzi scheme: Empirical analysis of high-yield investment programs. In *International Conference on Financial Cryptography and Data Security*. New York: Springer, pp. 41–56.
Moscatelli, Mirko, Fabio Parlapiano, Simone Narizzano, and Gianluca Viggiano. 2020. Corporate default forecasting with machine learning. *Expert Systems with Applications* 161: 113567. [CrossRef]
Narayanan, Arvind, Joseph Bonneau, Edward Felten, Andrew Miller, and Steven Goldfeder. 2016. *Bitcoin and Cryptocurrency Technologies: A Comprehensive Introduction*. Princeton: Princeton University Press.
Osius, Gerhard, and Dieter Rojek. 1992. Normal goodness-of-fit tests for multinomial models with large degrees of freedom. *Journal of the American Statistical Association* 87: 1145–52. [CrossRef]
Provost, Foster, and R. Kohavi. 1998. Glossary of terms. *Journal of Machine Learning* 30: 271–74. [CrossRef]
Reiff, Nathan. 2020. *How to Identify Cryptocurrency and ICO Scams*. Investopedia. Available online: https://www.investopedia.com/tech/how-identify-cryptocurrency-and-ico-scams/ (accessed on 1 August 2021)
Reurink, Arjan. 2018. Financial fraud: A literature review. *Journal of Economic Surveys* 32: 1292–325. [CrossRef]
Rodriguez, Arnulfo, and Pedro N. Rodriguez. 2006. Understanding and predicting sovereign debt rescheduling: A comparison of the areas under receiver operating characteristic curves. *Journal of Forecasting* 25: 459–79. [CrossRef]
Sammut, Claude, and Geoffrey Webb. 2011. *Encyclopedia of Machine Learning*. New York: Springer.
Schar, Fabian, and Aleksander Berentsen. 2020. *Bitcoin, Blockchain, and Cryptoassets: A Comprehensive Introduction*. Cambridge: MIT Press.
Schilling, Linda, and Harald Uhlig. 2019. Some simple Bitcoin economics. *Journal of Monetary Economics* 106: 16–26. [CrossRef]
Sebastião, Helder, and Pedro Godinho. 2021. Forecasting and trading cryptocurrencies with machine learning under changing market conditions. *Financial Innovation* 7: 1–30. [CrossRef]
Shimko, David. 2004. *Credit Risk Models and Management*. London: Risk Books.
Smith, Chris, and Mark Koning. 2017. *Decision Trees and Random Forests: A Visual Introduction for Beginners*. Blue Windmill Media.
Steinwart, Ingo, and Andreas Christmann. 2008. *Support Vector Machines*. New York: Springer Science & Business Media.
Strobl, Carolin, Anne-Laure Boulesteix, Thomas Kneib, Thomas Augustin, and Achim Zeileis. 2008. Conditional variable importance for random forests. *BMC Bioinformatics* 9: 307. [CrossRef] [PubMed]
Strobl, Carolin, Anne-Laure Boulesteix, Achim Zeileis, and Torsten Hothorn. 2007. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics* 8: 25. [CrossRef] [PubMed]
Strobl, Carolin, Torsten Hothorn, and Achim Zeileis. 2009. Party on! A new, conditional variable importance measure available in the party package. *The R Journal* 2: 14–17. [CrossRef]
Stukel, Thérèse. 1988. Generalized logistic models. *Journal of the American Statistical Association* 83: 426–31. [CrossRef]
Sze, Jin. 2020. *Coingecko Trust Score Explained*. Coingecko. Available online: https://blog.coingecko.com/trust-score-team-presence-incidents-update (accessed on 1 August 2021) .
Twomey, David, and Andrew Mann. 2020. Fraud and manipulation within cryptocurrency markets. In *Corruption and Fraud in Financial Markets: Malpractice, Misconduct and Manipulation*. New York: Wiley, pp. 205–50.
Victor, Friedhelm, and Andrea Marie Weintraud. 2021. Detecting and quantifying wash trading on decentralized cryptocurrency exchanges. In *Proceedings of the Web Conference 2021*. Ljubljana: International World Wide Web Conference Committee, pp. 23–32.
Votipka, Daniel, Rock Stevens, Elissa Redmiles, Jeremy Hu, and Michelle Mazurek. 2018. Hackers vs. testers: A comparison of software vulnerability discovery processes. Paper presented at 2018 IEEE Symposium on Security and Privacy, San Francisco, CA, USA, May 21–23. pp. 374–391.
White, Reilly, Yorgos Marinakis, Nazrul Islam, and Steven Walsh. 2020. Is Bitcoin a currency, a technology-based product, or something else? *Technological Forecasting and Social Change* 151: 119877. [CrossRef]
Yepes, Concepcion Verdugo. 2011. *Compliance with the AML/CFT International Standard: Lessons from a Cross-Country Analysis*. Technical Report. Washington: International Monetary Fund.