# Fintech Use of Genetic Algorithms

Tom Pawelek, *tp1333@msstate.edu*

***Abstract*—Artificial Intelligence (AI) systems have been used in the financial applications since their very inception. Most AI/ML models can benefit from co-deployment with a supporting Genetic Algorithm (GA) layer. Those can be mostly used to select features to train or to enable and disable individual features of AI workflows. This paper is a review of a selected literature showing the application of GA in fintech, as well as a summary of future developments in this field. It presents a mix of IEEE and ACM publications from the last 2 decades. For each reviewed article, I highlight the most important findings with the focus on GA application. Most researchers focus on loss prevention such as prediction of insolvency or anti-fraud measures. In many cases, financial ratios are used as the main type of input data. The role of GA is to select the most relevant factors, thus speeding up the training of AI networks and reducing the resource cost of running them. Most articles are rated for their usefulness in real world applications on a scale of 1 (least) to 3 (most useful). In the rating, I rely on my professional experience in commercial use of AI in highly regulated financial applications across the US consumer and B2B markets.**

***Index Terms*—fintech, genetic algorithm, artificial intelligence, neural networks, defi, fraud, insolvency, forecasting**

## I. Introduction

Digital ledgers have been one of the primary use of computers since their very inception. Over decades, we moved from simple bean counting into more advanced algorithmic processes, inevitably entering the realm of Artificial Intelligence (AI). Its use ranges from data science and forecasting to credit underwriting and automated trading.

With ever more powerful hardware and increasing resource capacity, AI focus has recently been on neural networks and deep learning capabilities. In this paper, I want to evaluate if evolutionary algorithms can co-exist with these modern neural nets and how the world of finance could benefit from such symbiosis.

In the following sections, I describe my choice of literature used to conduct my review. Afterwards, I perform analysis of selected works, followed by overall conclusions and comparison to current market trends and sentiments.

## II. Literature Selection

To pick papers for my review, I analyzed relevant keyword searches across IEEE and ACM repositories. In the process, I narrowed down my selection to two particular categories of AI applications:

- Underwriting and risk assessment,
- Fraud prevention

The reasons for my selection were two-fold:

- AI feels like a natural choice for those processes, which traditionally relied heavily on human intuition and expertise rather than clean-cut decision trees.
- My personal experience in both areas allows me to apply real life constrains and conditions to otherwise purely theoretical speculations.

My review begins with the subject of forecasting and risk analysis and then progresses into discovery and investigations.

## III. Literature Review

### *A. An Application of Decision Tree and Genetic Algorithms for Financial Ratios' Dynamic Selection and Financial Distress Prediction*

In [1], the authors propose a genetically trained decision tree model capable of predicting corporate financial distress. Most of the paper is dedicated to theoretical foundation of decision trees and genetic algorithms (GA).

The parameters chosen for their GA are very similar to my own model [2], and ultimately they result in this fairly simple decision tree:

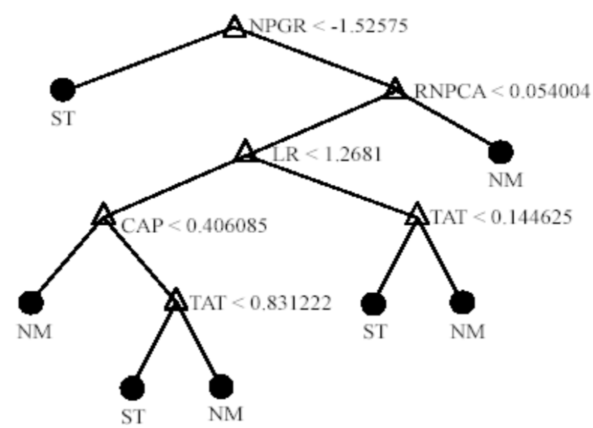


Fig. 1. Resulting Decision Tree

Notably, the paper lacks definitions for the major abbreviations used in the tree, such as NPGR (which stands for net profit growth rate). As far as the result goes, it would be considered a "cookie cutter" of business performance and will not be applicable to many industries, such as distressed lending (also known as Merchant Cash Advances (MCA)) or administration/liquidation.

Overall, the paper did not strike me as particularly useful in the financial field. It felt like an academic introduction to GAs more than a practical application.

**Useful Rate: 1**

### *B. Using Genetic Algorithms to Predict Financial Performance - Evidence from China*

The following publication [3] expands on the decision tree idea by introducing a wider range (50) of accounting variables used for their analysis. In addition, it provides an interesting view focused solely on Chinese economy.

Another improvement is the introduction of probabilistic neural networks (PNNs). Through their experiments, authors compare the performance of GA versus Decision Trees and PNN. Their results are shown below:

| | GA | PNN | DT |
|---|---|---|---|
| **Train Data** | 63.81 | 59.80 | 95.52 |
| **Test Data** | 61.50 | 55.23 | 55.60 |

TABLE I
PERFORMANCE COMPARISON OF GA, PNN AND DT

Clearly, decision trees are the most prone to overfitting. GAs show the smallest performance loss between Training and Test data, as well as the best overall prediction rate.

I wish the paper provided more insight into the internal structure of GAs and decision trees, but even without those elements, it is a very interesting approach to evaluating Chinese companies.

**Useful Rate: 2**

### *C. A Novel Financial Risk Evaluation Model Based on Adaptive Genetic Algorithm*

An improvement of GA is proposed in [4], based on a concept called Adaptive Genetic Algorithm (AGA). In this publication, focused on electric power enterprises, AGA is used to evaluate the following types of risk:

- operating,
- financing,
- investing,
- accounting,
- funds recovery,
- other (such as tax).

Compared to a traditional genetic algorithm, AGA introduces more sophisticated adaptive exchange and mutation operation, by dynamically controlling those phases in respect of a global solution domain. Through this improvement, AGA is guaranteed to converge to the optimal solution.

In a real life problem of credit risk, a sample AGA is shown to outperform traditional backpropagation neural networks. It performs much faster and is immune to local minima, which financial neural networks tend to struggle with.

I like the simple yet detailed description of adaptive improvements applied to GAs. It seems to be a good candidate for other applications related to risk evaluation.

**Useful Rate: 3**

### *D. Variable selection for financial distress classification using a genetic algorithm*

Continuing the subject of risk evaluation, [5] focuses on bankruptcy predictions with the use of GAs and financial ratios. It analyzes 60 UK companies which went through a period of distress around the year 2000.

Using ratios to predict future performance dates back to 1960, and the role of a chromosome in this particular study is to select the ratios which are being evaluated (rather than to tune their thresholds). In other words, a chromosome bit value of 1 means that a given ratio should be considered valid, while 0 negates its use.

This publication stands out as one of the best I evaluated. One reason is the simplicity and clarity of GA application, which results in a very concise output: only 3 selected ratios outperformed 4 used in a conventional application, and their formula (weights) are explicitly given.

The second reason has to do with how GA is being used: rather than competing with classification or regression networks, it's used as a higher level modeling tool - it selects which layers (here: ratios) should be used. This corelates with my personal experimentation and commercial use of GAs.

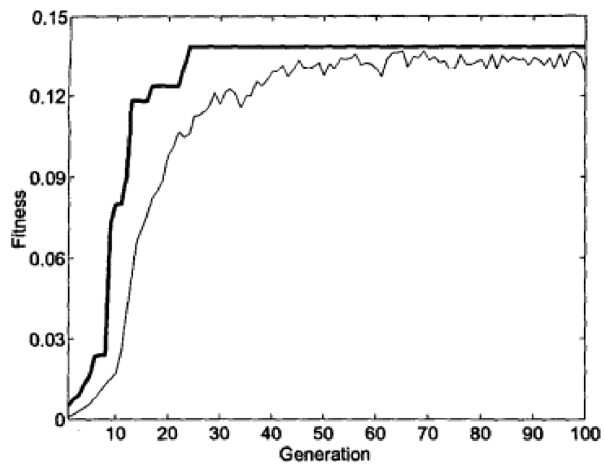


Fig. 2. Financial Distress fitness function

Interestingly, fitness function also seems to plateau around generation 30-40, which also matches my own GA applications.

**Useful Rate: 3**

### *E. Research on detecting technique of financial statement fraud based on Fuzzy Genetic Algorithms BPN*

The following paper [6] continues the study of GAs, but pivots their application towards detection of fraudulent financial statements of public companies.

In their study, the authors analyze 97 public Chinese companies which were all accused by the Chinese SEC of fraud. Each statement year is treated as a separate unit, resulting in total sample count of 190.

As in the previous article, financial ratios remain the dominant group of factors, representing 30 out of 57 total variables. Combining fuzzy logic with GAs, authors build a model capable of recognizing 86.3% fraudulent statements.

The combination of fuzzy logic and GA is the exact structure of my professionally used model. The only difference between this paper and my approach was a hybrid structure of chromosome (both selection and numerical values) that I ended up using for maximum performance.

**Useful Rate: 3**

*F. Knowledge-guided genetic algorithm for financial forecasting*

To summarize current state of GA applications in the world of finance, authors of [7] published a more theoretical piece, describing in detail how regression models and evolutionary algorithms can work together with knowledge based systems.

In the article, KGAF is defined as a forecasting-capable GA capable of tapping into the knowledge resources. Its domain knowledge covers both business and marked performance indicators, focusing mostly on balance sheets and income statements.

As in previous documents, GA acts as a selector of 46 parameters, which are used in the linear regression model. Domain knowledge is then applied directly during the mutation phase, guiding the changes towards optimal results.

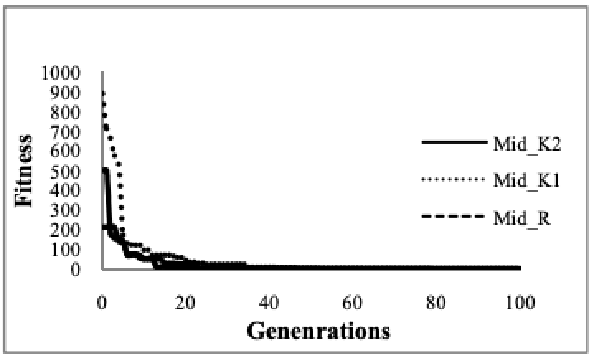


Fig. 3. KGAF fitness function

This improvement approach results in a model which shows better overall results ( 20% better fitness outcome) and reaches its full potential faster, albeit still within the expected 30-40 generations.

I find this solution more elegant than random mutations, but at the same time I think we need to be careful no to over-guide the evolution, which ultimately does depend on random changes to discover new optima. This over-guiding would be an evolutionary equivalence of over-fitting in conventional regression models.

**Useful Rate: 2**

*G. Using Genetic Algorithms for Feature Selection in Predicting Financial Distresses with Support Vector Machines*

GAs are not restricted to working with decision trees and linear regressions. In [8], GAs are applied to Support Vector Machines (SVM) in the pursuit of financial distress prediction.

Here, authors use GA to replace gain ratio feature selection. As previously, chromosomes are pure selectors, rather than thresholds. The procedure is shown below:

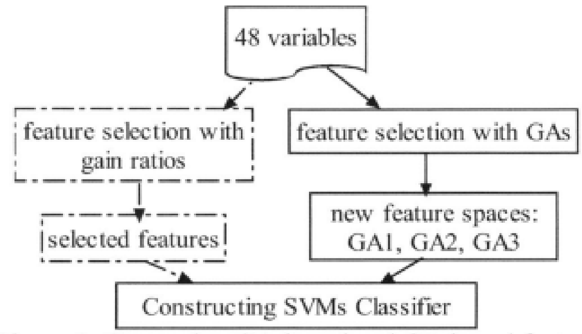


Fig. 4. GR vs GA feature selection

There are 48 variables to choose from. Same as in previous papers, most of them are financial ratios. What differs this publication from others, though, is use of two different selection methods:

- roulette
- tournament

Using evolution, authors were able to predict distress on average 2-3 years prior to it taking place. While not proposing any revolutionary measures, this publication serves as a solid foundation for further research into GA assisted SVMs.

**Useful Rate: 2**

*H. Application of regression analysis based on genetic particle swarm algorithm in financial analysis*

Continuing the research of various GA applications, I encountered this 2010 paper about the Particle Swarm Algorithm (PSA) [9].

The underlying model is a GA-assisted regression analysis. It is then improved through the introduction of particle swarm algorithms. It enhances the evolution of following generations, making sure that local optima don't affect overall fitness.

While the idea is very interesting and supported by somewhat convincing results (faster training, less resources), the paper is unfortunately poorly organized and reads as if it was machine translated. I am skeptical of the findings presented by its authors. Even the input data is limited to only a few financial indicators spread over 6 years for one company.

I think the use of PSA warrants further investigation, but it needs to be performed in a controlled and verifiable environment.

**Useful Rate: 1**

*I. Integration of Genetic Algorithm and Neural Network for Financial Early Warning System: An Example of Taiwanese Banking Industry*

Keeping my focus on Asian markets, the next article [10] I reviewed describes an attempt to combine GA and neural networks for the purpose of early warning in a Taiwanese banking network.

Here, GA is used to actually modify the weights of underlying neural networks, rather than purely select its layers or input features. Notably, authors use a mutation rate of 0.3 which is a very big value.

In their experiments, the GA-improved networks score a better result than their traditional counterparts (topping up at 85.71% hit ratio). While clearly outperforming other models, the use of such a high mutation rate makes me question the academic value of this approach.

I suspect that almost purely randomized neural weights adjustments were ran long enough to achieve favorable results, effectively choosing an over-fitted network for this particular scenario. I am very much doubtful it would perform equally well presented with different data sets.

**Useful Rate: 1**

### *J. Extracting earnings information from financial statements via genetic algorithms*

The next publication I reviewed [11] is one of the oldest documents I read, dating back to 1999. It does, however, present an interesting use of GAs: rather than forecast future performance, it assists financial analysts in data extraction from financial statements.

It stands out as a critic of subjectivity in the financial markets, pointing out the sheer number of factors (input variables) makes it virtually impossible to agree on automated methodology that would aggregate all that data into a single output.

The article is written in a very educational manner, aiming to teach its reader the principles of evolutionary algorithms. Chromosomes are 67-bit long, and are pure selectors of features, which are then present (or not) in the LOGIT estimation of whether the coming up earning should flip its sign.

While resulting in an accuracy of only 65% - notably lower than other documents in this review - I still consider this article to be a must read for any GA beginner looking for real life application of evolutionary coding.

**Useful Rate: 3**

### *K. Optimization of financial network stability by genetic algorithm*

Growing the scope from one entity to a multi-party environment, the authors of [12] propose a GA designed to improve the stability of a financial network of banks.

When we consider the entirety of a financial ecosystem (such as a network of trading banks), its stability can be defined as resilience to disturbance. The goal is to prevent the capital value of any bank to fall below certain threshold, which would lead to its insolvency.

In this directed network graph, we want to maximize the minimum value of capital in banks (nodes), after a fixed time of money flowing through its edges. Utilizing GAs, the authors manage to design a model which significantly improves this network stability.

As in previous papers, the fitness network plateaus around generation 40-60.

This paper shows GA usefulness in directed networks with flow dynamics, increasing the complexity of its application, but keeping the somewhat stable resource cost.

**Useful Rate: 3**

### *L. Comparing ML Algorithms on Financial Fraud Detection*

Stepping into the realm of financial fraud, I wanted to evaluate how the GAs can be used in anti-fraud applications. The 2019 evaluation [13] of ML algorithms used for fraud detection provides a summary view of various techniques used in this particular application.

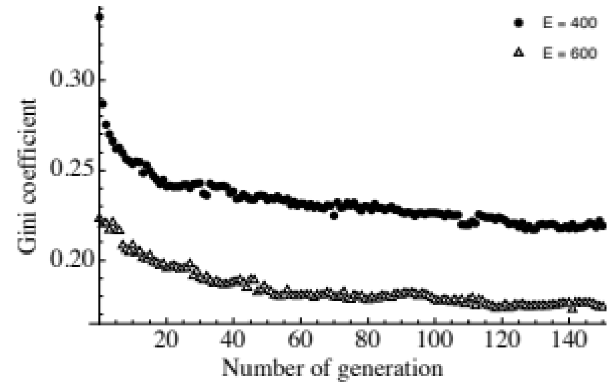


Fig. 5. Network Stability fitness function

Its authors used a public credit card fraud dataset from 2013 and measure the efficiency of the most popular ML tools in its evaluation. The algorithms analyzed comprise:

- K-Nearest Neighbor
- Logistic Regression
- Naive Bayes
- Decision Tree
- Random Forest
- AdaBoost
- Neural Networks

Ignoring the training speed and tuning cost, the highest predictive accuracy can be attributed to Neural Networks, Random forest and Adaboost. I have already shown that Neural Networks are perfect candidates for working with GA.

As to AdaBoost, it has already been coupled with GA, albeit in the appliaction of nanocomposite materials. Similarly, Random forest tuned with GA has found itself into the realm of engineering.

It presents a potential research opportunity to apply those two combinations in the realm of financial anti-fraud.

**Useful Rate: 3**

### *M. Research on Prediction of Anti-Fraud in Automobile Finance Based on XGBoost Machine Learning Algorithm*

I want to conclude this review with real-life applications of AI as fraud prevention and evaluate the possibility of GA improvements. In [14], Chinese researchers show the use of XGBoost to prevent car loan fraud.

There are 18 input variables available in the dataset, most of which would be available in US consumer markets. The logical regression with XGBoost duo achieves an AUC of 83-87%.

The easiest application of GA would be to use an 18-bit chromosome to selectively filter on features used in the regression. Previous papers shown this to be a highly effective strategy.

### N. AI-powered Fraud Detection in Decentralized Finance: A Project Life Cycle Perspective

Finally, in [15], we are presented with a very comprehensive overview of current AI applications to detect fraud in decentralized finance (DeFi) environments, most notably - cryptocurrency blockchains.

With the number of fraud incidents on a steady incline, prevention becomes a top priority for all parties involved: from the exchange administrators to the very end-users of crypto.

Most common types of DeFi fraud include:

- Ponzi schemes,
- fake token offerings,
- rug pulls.

Due to the public nature of blockchains, AI systems are perfect candidates to assist in the prevention and early detection of fraud. The two models noted by the authors are Random Forest and Graph Neural Networks.

Both of these structures were already discussed in this paper, and both seem to be ideal candidates for further improvements via evolutionary algorithms.

## IV. Conclusion

As shown in my review, the two most dominant applications of GA-assisted AI in fintech are:

- distress prediction,
- fraud prevention.

This is not surprising. Although automated traders and crypto startups might be behind the AI hype, loss prevention remains the most important goal of any financial institution. This remains true for both conventional (banks) and modern (DeFi wallets) providers.

In all but one paper I reviewed, chromosome bits were used to select features, rather than fine-tune numerical values. This becomes obvious once we realize that most datasets mentioned contained around 50 variables. If we assume that - even with down-scaling - we would want to encode 256 values for each, that would result in 400-bit chromosomes. The number of their possible combinations ($10^{120}$) is much greater than the number of atoms in an observable universe, which would make it somewhat challenging to evolve them in a meaningful way.

Finally, I observed many performance similarities between GAs, even when applied to vastly different environments. For example:

- Most fit functions plateaued around **generation 30-40**,
- Most models resulted in prediction rate of **around 85%**

Those numbers match my own results from my own developments in the fintech GA field.

## References


[1] J. Sun and X.-f. Hui, "An application of decision tree and genetic algorithms for financial ratios' dynamic selection and financial distress prediction," in *2006 International Conference on Machine Learning and Cybernetics*, 2006, pp. 2413–2418.

[2] R. O. A. Tom Pawelek, "Genetic evolution of expert ai systems," *Mississippi State University*, 2024.

[3] Y. Jiang, D. Ke, Y. Wang, and L. Xu, "Using genetic algorithms to predict financial performance – evidence from china," in *2007 IEEE International Conference on Systems, Man and Cybernetics*, 2007, pp. 3225–3229.

[4] J. Jiang, L. Zhang, and Z. Liu, "A novel financial risk evaluation model based on adaptive genetic algorithm," in *2009 Second International Symposium on Electronic Commerce and Security*, vol. 2, 2009, pp. 396–399.

[5] R. Galveo, V. Becerra, and M. Abou-Seada, "Variable selection for financial distress classification using a genetic algorithm," in *Proceedings of the 2002 Congress on Evolutionary Computation. CEC'02 (Cat. No.02TH8600)*, vol. 2, 2002, pp. 2000–2005 vol.2.

[6] J. Liang and W. Lv, "Research on detecting technique of financial statement fraud based on fuzzy genetic algorithms bpn," in *2009 International Conference on Management Science and Engineering*, 2009, pp. 1462–1468.

[7] J. Du and R. Rada, "Knowledge-guided genetic algorithm for financial forecasting," in *2012 IEEE Conference on Computational Intelligence for Financial Engineering  Economics (CIFEr)*, 2012, pp. 1–8.

[8] P.-W. Huang and C.-L. Liu, "Using genetic algorithms for feature selection in predicting financial distresses with support vector machines," in *2006 IEEE International Conference on Systems, Man and Cybernetics*, vol. 6, 2006, pp. 4892–4897.

[9] X. Cheng, L. Sun, and P. Liu, "Application of regression analysis based on genetic particle swarm algorithm in financial analysis," in *2010 International Conference On Computer Design and Applications*, vol. 4, 2010, pp. V4–335–V4–338.

[10] J.-C. Hsieh, P.-C. Chang, and S.-H. Chen, "Integration of genetic algorithm and neural network for financial early warning system: An example of taiwanese banking industry," in *First International Conference on Innovative Computing, Information and Control - Volume I (ICICIC'06)*, vol. 1, 2006, pp. 562–565.

[11] J. Trigueros, "Extracting earnings information from financial statements via genetic algorithms," in *Proceedings of the IEEE/IAFE 1999 Conference on Computational Intelligence for Financial Engineering (CIFEr) (IEEE Cat. No.99TH8408)*, 1999, pp. 281–296.

[12] J. Wang, W. Guo, and K. Y. Szeto, "Optimization of financial network stability by genetic algorithm," in *Proceedings of the International Conference on Web Intelligence*, ser. WI '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 517–524. [Online]. Available: https://doi.org/10.1145/3106426.3106524

[13] C. M. Tae and P. D. Hung, "Comparing ml algorithms on financial fraud detection," in *Proceedings of the 2019 2nd International Conference on Data Science and Information Technology*, ser. DSIT 2019. New York, NY, USA: Association for Computing Machinery, 2019, p. 25–29. [Online]. Available: https://doi.org/10.1145/3352411.3352416

[14] Y. Bao and H. Wen, "Research on prediction of anti-fraud in automobile finance based on xgboost machine learning algorithm," in *Proceedings of the International Conference on Digital Economy, Blockchain and Artificial Intelligence*, ser. DEBAI '24. New York, NY, USA: Association for Computing Machinery, 2024, p. 367–375. [Online]. Available: https://doi.org/10.1145/3700058.3700116

[15] B. Luo, Z. Zhang, Q. Wang, A. Ke, S. Lu, and B. He, "Ai-powered fraud detection in decentralized finance: A project life cycle perspective," *ACM Comput. Surv.*, vol. 57, no. 4, Dec. 2024. [Online]. Available: https://doi.org/10.1145/3705296