Accuracy v/s Explainability: A battle in the financial world

Rishabh Agarwal Jain
Citi Research,
Bits Pilani, Dept. of Computer Science
rishabh.agarwaljain@citi.com

Sushrut Shendre
University of California, Irvine
sshendre@uci.edu

Vikas Sawant
Citi Research
vikas.sawant@citi.com

Jayant Sachdev
Citi Research
jayant.sachdev@citi.com

Abstract

We have come up a long way as far as technology is concerned. The power to extract insights using data mining concepts and the use of machine learning to use historical data to predict optimal performance has led to significant empowerment in making informed decisions. While we have a host of algorithms to facilitate these machine learning approaches and increase accuracies, these often pose a question on the explainable nature of the models built using these very algorithms. While accuracies and prediction rates increase as we delve into more complex algorithms, the explainable nature often decreases with an increase in such complexities. How do we compare these aspects of accuracy and explainability, especially in the finance industry? Do we prefer one over another? How do we maintain a balance between them? These are some of the problems we attempt to address in this paper and introduce the conundrum of explainability in finance domain for researchers and academics to delve upon.

1 Introduction

With the success of deep learning, neural networks, natural language processing, computer vision, and likewise, the concept of machine learning and automated solutions have reached peaks yet still have the potential to give more. The greatest achievement can be seen by the fact that such solutions can be and have been applied to every single field possible. Its use to predict metrics to accelerate business value has been prevalent for a long time. In healthcare, models have been created to detect diseases. In the entertainment and e-commerce industries, the use of recommender systems to drive customer engagement has seen these companies grow manifold. In the automobile industry, the invention of self-driven cars has shown the capabilities of machine learning and AI like never before.

However, with every technological innovation, come some limitations. The interpretability and explainability of machine learning models can be seen as a significant problem in this domain. While we are developing better models and products every day, the question remains as to how explainable these models are. With deep learning and neural networks taking over the standard machine learning models, there has been an aggravation in the explainable nature of these products. This boils down to the fact that such complex models fall under the 'black-box' category.

We know that these algorithms perform very well in the training process, but we do not know what is exactly happening with these algorithms. When coupled with a business frame of mind, the importance of explaining our models can be seen more clearly. In the medical industry, we know that the neural network model can do a great job at detecting health anomalies, but how do we explain

33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.

that to a doctor or a patient. It can even be fearful for a patient to hear that he may have a particular disease because a bunch of mathematical equations, most of which we do not have any idea of, are not in his favor. It is actual human lives that we are talking of here.

2 AI Enigma in Finance

When we look at the finance industry, the problems are similar. It would be worrisome for investors or senior management to know that we wish to approve customers credit/loan because the relations between mathematical entities are in their favor and vice-versa.

This is the major reason why most banks and financial institutions use till date is logistic regression rather than complex deep learning solutions. It is because the simple nature of logistic regression allows the management to understand the importance of variables, relationship between features, and how they collectively interact to come up with a credit score.

Not only do they have to be clear about their strategies, they also have to comply with the law. The Equal Credit Opportunity Act (ECOA), as implemented in Regulation B, and also the Fair Credit Reporting Act (FCA) need the lending institutions to provide reasons as to why a loan application was rejected (Modarres et al. (1); Sarah Ammermann (2)), and having some black box will not help that at all.

2.1 Regulators and Explainibility

When it comes to financial fraud, it is even more worrisome. Not only does the management have to feel comfortable with the understanding of the model and the relation between variables, the regulators too need to have a high level of confidence in order to accept these solutions. While neither the senior management nor the regulators have any problem when the predictions are correct, what happens when there is an error. The frightening fact is that they will not have an answer as to what went wrong. A follow-up question that would essentially come up is, how can we be sure that the errors would not happen every time.

Given the risks associated with anti-money laundering solutions and given that these have repercussions as dangerous as terrorism financing, it is inevitable to be skeptical about these modern complex solutions. It is also the major reason why banks and financial institutions stick with rule engines such as Mantas (3). While nobody would not want better solutions, they surely need solutions that are understandable and explainable.To further our understanding of this topic, we look at a few case studies.

3 Case Studies

The 'black box' which we referred to 1, needs to be unboxed for the sake of its explainability to the regulatory bodies. Feature engineering plays a pivotal role in doing so. In that regard, the top-down modeling approach of creating high-quality features and scoring them in a pre-defined range is much better as compared to the bottom-up modeling approach, which has numerous low-quality features. These features can then be used in traditional, explainable ML models like logistic regression, Decision trees, and even ensemble methods like random forests.

We will have a look at how vital explainability and interpretability are using case studies on Anti-Money Laundering(AML) and Trade surveillance. These case studies bolster the use of the top-down approach as compared to black-box models.

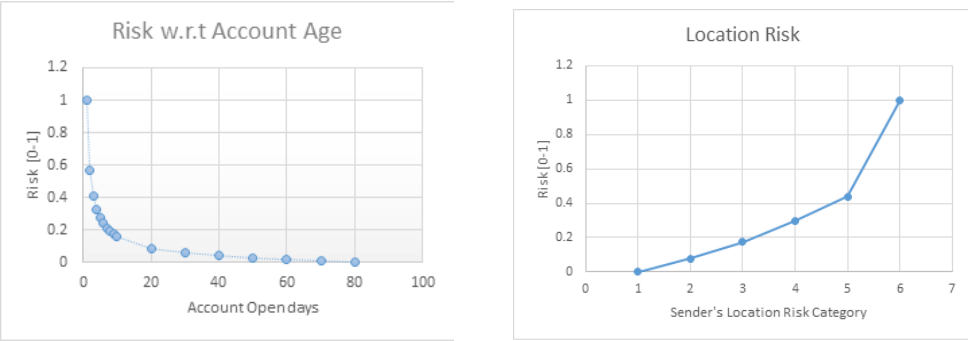
3.1 AML for Money Laundering

Money Laundering is the process of converting black money to legitimate white money and happens via a series of steps which include placement, layering, and integration. The dangers of money laundering can be seen from the fact that not only does it disrupt economies, but it is one of the most common means used for terrorism funding.

Banks have to implement the Anti-money Laundering act by complying to the Financial Action Task Force (FATF (4)). The transaction monitoring applications which banks build, thus require the

machine learning models, to make accurate predictions. The reason is quite simple; one instance of a false negative and it can result in a huge loss not only due to the laundering but also due to the penalty which would be subject to by the regulators. So that is where exactly the catch is. While the models have been accurate, they have to be developed in such a way that they can be easily explainable to the regulators.

To strengthen our ideas and reinforce the need for explainability, we look at a study we conducted. The data corresponds to the transaction patterns of customers belonging to different demographics. A compelling case that came up during detection as a potential fraudulent customer is of a college student. This student was a Mexico resident and moved to the United States for his higher studies and opened a bank account. This student was flagged by an instance of an AML scenario which had a focus on the sender, the receiver, account age, mode of transfer and geography risk which is associated with the risk of the place of residence.



(a) label 1 (b) label 2

Figure 1: Location risk and Age risk features.

Being a Mexico citizen, there was a relatively high geography risk category associated with him (Mexico due to its drug trafficking activities falls under a high-risk zone). As per a rule in the rule-engine, a transaction of amount more than $10,000 is considered a suspicious transaction and involves the proceeds of specified unlawful activity. Moreover, the transfer of his college tuition fee from Mexico to the US of more than this amount meant there was another red flag against him. This was a feature level bias as new account (1) and high risk transfer (1) did not necessary mean that the movement is designed to conceal or disguise the nature, location, source, ownership or control of the proceeds of "specified unlawful activity" or avoid a transaction reporting requirement under federal or state law (see 18 USC Section 1956(a)(2)). (Doyle (5))

It is important to note here that, even though the system flagged a legal transaction, but the use of feature engineering in top-down approach has made it explainable. Also, as discussed above, the AI system in banking has to have a unique trait, i.e., it can flag a false positive, but it should not miss a true positive. It is needless to say that the system has to minimize the false positives in order to be cost-effective as compared to rule-based systems.

3.2 Trade Surveillance

Another scenario in the financial world is trade surveillance, wherein financial institutions are required to monitor the activities of traders to prevent illegal trading practices. Insider information trading, wash trading, and spoofing are typical examples of such malpractices that have happened in the past. Prevention of such events boils down to monitoring the behavior of traders and their behavior. Behavioral analytics needs to be employed by determining features that are precise representations of their trading attributes. A top-down approach of creating high-quality features to carve out suspicious patterns is pivotal in enhancing the explainability of the model.

Consider the law case study of CFTC vs. Wilson ("DRW") where CFTC alleged Wilson ("DRW") of the market manipulation. This case came into notice in 2011 when IDHC initiated an investigation of trading practices being followed by DRW, but the inquiry was dropped without actions. CFTC,

Table 1: Validation accuracy on the test set given the full training dataset or a 30% subset of the training data for the explainable model and black box model

Model	Accuracy (30%)	Accuracy (Full))
Black-Box Model	85.3%	96.6%
Explainable model	81.1%	93.2%

however, accused Wilson of market manipulation but failed to prove the allegations in the court of law since they failed to prove that the prices were artificially inflated. (Sar (6))

This is where the need of an explainable model is realized. If the system deployed were explainable, then the prosecutor could have presented a stronger case in front of the court.

To reinforce our idea of explainability vs. accuracy, we experimented with a trade surveillance scenario called spoofing.

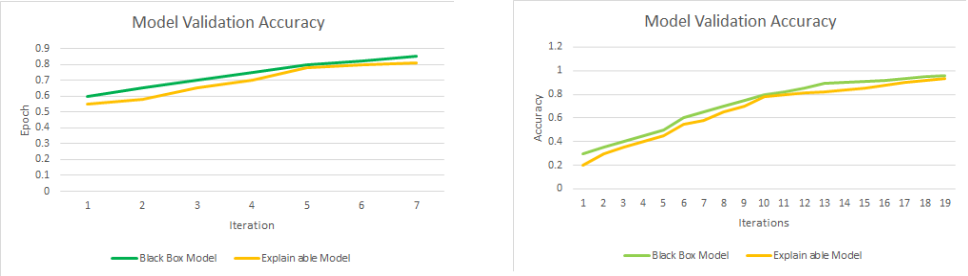
3.2.1 Experiments

To explain spoofing, consider the scenario where a trader, to increase the price of an underlying asset class places a buy order and waits for its successful execution. Once attaining an initial position for the particular asset class, the trader starts manipulating the market by placing other buy orders but without intent for the successful execution of the order.

The trader deliberately cancels out the order after placing it with the intent of generating an open interest in the market. As a result of these orders, there is a positive impact on the price of the underlying asset class. Once the targeted price for booking profit is achieved, the trader exits from its position by booking the desired levels of profits. This behavior amounts to spoofing and is illegal under the rule of law.

In our experiment, we built two different models; first, a deep learning model(black-boxed) using native trade data as an input while the second model is a tree-based structure and uses the top-down approach of feature engineering.

We report results from the experiment on Spoofing trade using the full dataset and using a subset of 30%. 1



(a) label 1

(b) label 2

Figure 2: Validation accuracy for spoofing on 30% data (left) and full dataset(right)

To simulate a smaller dataset, we had randomly selected 30% of the training set and used it instead of the entire training set. 2 shows the validation accuracy (on the entire validation set, not a subset) over time. A similar trend is observed for the partial dataset. The gap in final accuracies between the deep learning-based model and the tree-based model increases slightly (Table 11, “30%” column); the final accuracy for the explainable model drops only slightly.

The validation accuracy of these models over time is shown in Fig. 2. The best results are obtained when epochs are high; the explainable tree-based model has underperformed in all dataset(s) as compared to neural network-based models but to an acceptable range. In the regulatory requirement of explaining the models, the two model accuracy is roughly equally well (Table 1) 1, and the tree-based model prevails, further highlighting the advantage of an explainable model.

4 Conclusions

The finance industry still experiences a dire need for the underlying models and strategies to be clear and easy to use. The complexity curve in the algorithmic and modeling paradigm is increasing with each passing day, but being a stringent industry, and rightly so, finance has taken a backseat incorporating these complex methodologies, also, because of the inability of these complex methodologies to explain themselves on a ground level. From our analysis of the case studies, we observed two vital points. First, whenever it comes to decision making, we need to understand the trade-off between the accuracy and explainability of models. If the difference between accuracies is minimal, we might as well use the simpler model just because of its simplicity and an explainable nature. Secondly, we must continuously try to explore how complex algorithms can be incorporated such that there is no compromise in their explainable and interpretable nature. As financial analysts and modelers, we need to take care of these situations and adhere to the expectations that the business has. With this comprehensive research survey, coupled with illustrations of two case studies, we believe this paper can prove as a guide and point of reference for leaders, managers, and researchers in the financial data science world.

Disclaimer

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of their respective employers/institutions .

References

[1] Ceena Modarres, Mark Ibrahim, Melissa Louie, and John Paisley. Towards Explainable Deep Learning for Credit Lending: A Case Study. pages 1–8, 2018. URL http://arxiv.org/abs/1811.06471.

[2] Sarah Ammermann. Adverse Action Notice Requirements Under the ECOA and the FCRA - Consumer Compliance Outlook: Second Quarter 2013 - Philadelphia Fed, 2013. URL https://consumercomplianceoutlook.org/2013/second-quarter/adverse-action-notice-requirements-under-ecoa-fcra/.

[3] Mantas. A product by oracle financial services for anti money laundering. URL https://www.oracle.com/industries/financial-services/banking/products/anti-money-laundering/.

[4] FATF. Money laundering. URL https://www.fatf-gafi.org/faq/moneylaundering/.

[5] Charles Doyle. Money laundering: An overview of 18 u.s.c. § 1956 and related federal criminal law. URL https://fas.org/sgp/crs/misc/RL33315.pdf.

[6] Meric Sar. Dodd-Frank and the Spoofing Prohibition in Commodities Markets. *Fordham Journal of Corporate & Financial Law*, 22(3):383, 2017. ISSN 1532-303X. URL https://news.law.fordham.edu/jcfl/wp-content/uploads/sites/5/2017/10/MSar-Note.pdf.