Fraud Detection - Machine learning to the rescue
Enterprise value chains lead us from product design and development through to customer engagement and service, with various support functions playing an enabling function. While focusing on the core business value delivery, it is easy to lose sight of functions like risk management that have an uncanny way of striking when least expected.

While quite rare in incidence, the impact of process failure and fraud cannot be underestimated. Given that the underlying causes are not always easy to establish, examining fraud arising from inadvertent process failure can provide methods to mitigate the adverse impact.

The impact of fraud is enormous. The ACFE (Association of Certified Fraud Examiners) report estimates that organizations lose up to 5% of their annual revenues to fraud. We have also seen organizations go bankrupt or lose significant market capitalisation as a result of fraud. These outcomes can often be attributed to failure in oversight of low-priority but potentially high impact processes.

A practical challenge in fraud detection is often posed by the very business rules and algorithms intended to prevent fraud. These set off false positives with the result that the benefit of loss prevention is offset by the cost of investigating the false positives, customer dissatisfaction and reputational damage.
Traditional audit approaches to detecting frauds and identifying control failure are built using heuristics around fraud indicators such as rules, thresholds, checklists, scorecards and profiles using:

- Data that is scientifically sampled to be representative of the population
- Quality datasets that enable analysis leveraging traditional tools
- Periodic examination to ensure that processes haven’t degraded over time - with rules refined from previous learnings

The key challenges with this approach are the following:

- Traditional audit approaches are challenged by the inability to react quickly before evidence is lost, the inability to handle large scale data (especially log data) and correlate between feeds
- Given the low incidence rate and the fact that new fraud patterns emerge continuously, a retrospective audit is unlikely to unearth process failures
- Manually recalibrating rules has limitations and only a partial set of available signals are factored

The fundamental point of difference between Machine Learning (ML) approaches and traditional approaches is this: traditional approaches use prior data patterns to define normal behavior and view outliers from that perspective. This means rules for genuine exceptions need to be created, data errors need to be systemically addressed and abnormalities require investigation. Given the data-centric orientation of the exercise (as opposed to domain expertise in traditional approaches), experimentation and willingness to recalibrate is of paramount importance.

ML models are developed and tested through a variety of algorithmic techniques as the underlying data patterns could behave differently. Based on field investigations, the learnings in terms of false positives and cases that were missed and detected through other sources are incorporated into the model. The model is periodically re-configured and improves in predictive accuracy with the availability of new data.
The above philosophy for anomaly detection has been leveraged in our implementations and has delivered significantly better actionable results. Using Apollo, our analytics platform, the scale of data that can be dealt with tends to be 100 to 1000 times more than traditional approaches, enabling leverage of more historic data and also log data.

The process of examining this data is experimental. Rather than predefining attributes of relevance, a variety of attribute combinations are tried to evaluate the combinations that deliver the best results.

Likewise, thresholds and parameters are varied to achieve the best balance that meets the required need. In terms of measurement, the optimal approach is to achieve a balance between precision (fraction of actual incidents in the identified shortlist) and recall (fraction of actual incidents identified).

Based on working with a variety of algorithms, the following have been the key learnings:

- Invariably more data and better quality data leads to improved predictive outcomes. This seems to be a bigger driver for prediction quality compared to the choice of algorithms.

- We have observed improved outcomes from ensemble models (aggregation of simple models) compared to complex models. We can compare this to the “wisdom of crowds” where polling a jury of commoners, each operating with partial information, delivers better results than a single expert operating a sophisticated process.

- In multiple datasets we have seen a phenomenon of uncertain outcome from investigation. We were able to improve predictive accuracy by up to 10% across multiple datasets by leveraging embedded intelligence through smart classification of grey data points.

While our focus is on improving the quality of algorithms for improved predictive accuracy, it would be remiss to ignore the critical role of stakeholder management through the process. Experience shows that success of a new paradigm requires taking critical business owners along in addition to educating and empowering audit and compliance teams to leverage this capability along with their domain insight.
About Wipro

Wipro Ltd. (NYSE:WIT) is a leading information technology, consulting and business process services company that delivers solutions to enable its clients do business better. Wipro delivers winning business outcomes through its deep industry experience and a 360 degree view of “Business through Technology.” By combining digital strategy, customer centric design, advanced analytics and product engineering approach, Wipro helps its clients create successful and adaptive businesses. A company recognized globally for its comprehensive portfolio of services, strong commitment to sustainability and good corporate citizenship, Wipro has a dedicated workforce of over 160,000, serving clients in 175+ cities across 6 continents. For more information, please visit wipro.com or write to us at info@wipro.com

About the author

R. Guha heads Corporate Business Development at Wipro and is charged with driving certain key initiatives of strategic value. He oversees the development of the Apollo platform and its deployment for various customers.