This is the first of a four-part series of articles from DBS Group Audit, documenting how they have leveraged on automation and innovative machine learning predictive auditing technique to become more Productive, Proactive, Predictive & Preventive in their journey towards the Future of Auditing.

Using Data Analytics for Predictive Auditing

By DBS Group Audit

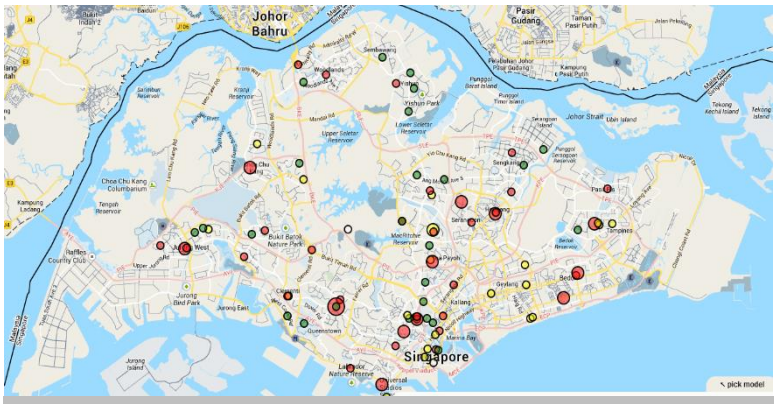
"Numbers have a convincing story to tell.
They rely on you to give them a clear and convincing voice."
- *Stephen Few, a leading expert in data visualisation*

The age of data analytics, machine learning and artificial intelligence is here - we've already seen IBM Watson beat humans at Jeopardy. So can data analytics help banks identify potential risks better than humans? We embarked on a data analytics quest to validate this hypothesis.

Annually, the DBS Group Audit team will do a risk profiling exercise to select branches that will be audited that year. This exercise is traditionally based on a certain set of criteria, such as [X1, X2, …Xn. These criteria are chosen based on our collective years of experience and understanding of how they can impact the riskiness of the branch. Each criterion is then assigned weights to determine the risk rating of the branches. While this approach has been successful the last eight years, it begs a few questions: 1) are these criteria good predictors of branch riskiness and 2) are the weights assigned reflective of the sensitivity of each criterion? Can we review and enhance our approach by letting the data speak for itself – can we use data analytics to develop a new risk model that more accurately predicts whether a branch needs to be audited?

With these questions in mind, the Audit Data Analytics team together with a handful of auditors embarked on a collaboration with A*STAR's I^{2}R research lab to work on this predictive model. The approach involved an ensemble model that incorporates logistic regression, random forest and gradient boosting – technical jargon referring to the analysis of associations and attributions between data variables. In layman terms, this involved taking measurements of multiple attributes of our branches, and studying correlations between these attributes and the likelihood of risk events within these branches. After a process of data crunching, these correlations are distilled into a data model that profiles the risk score of each branch, which in turn facilitates the selection of branches for audit.

Comparing the new data-driven model with the traditional experience-based approach, the data-driven model successfully correlated more than 130 attributes with the risk score, compared to a mere seven used in the traditional approach. And here's the kicker – the new data-driven model is **50% more accurate** at predicting whether a branch needed to be audited compared to the traditional approach! With this achievement, it means that we can now more accurately focus our attention and energies on the correct



Info at our fingertips – at a glance, we can now predict where the high risk branches are, improving our ability to spot which branches to focus on.

branches – resulting in significant productivity improvements! Plotting the risk scores against the branches map locations, we can also predict where the higher risk branches are, and even study changes in risk scores over time.

The insights don't end there. Among other observations, one interesting story that the data told us was that the central branches seemed to have fewer cash discrepancies and risk events than their peers, despite having a large number of transactions. Good management? Perhaps. Worth studying further? Definitely! We could use this information as part of our continuous auditing and computer-assisted audit techniques, to interrogate the data further before we go down to the branches for the actual audits.

Naturally, DBS Group Audit has shared the completion of this milestone with senior management and various stakeholders. The response has been unanimously characterised by excitement in the results achieved, and interest to collaborate to refine this model further. For example, one suggestion raised was to rank the risk events, so that the model can sound louder alarm bells when it anticipates a severe event is about to occur.

Furthermore, why stop at analysing the *current* risk profile of a branch? Why don't we use the data further, to predict risk events even *before* they happen? With the success of the initial model, this is precisely the next problem statement that the team is working on in the next phase of this project.

Welcome to the exciting future of auditing!