**SA Budget Speech predictions using machine learning**

Finance Minister Tito Mboweni is expected to deliver his 2020 Budget Speech on 26th February. No doubt, many South Africans will be interested in what the projected economic figures for the year will be. But could these be predicted using machine learning? "As an experiment, this is exactly what we attempted to do," reports Kimoon Kim, Data Director at Teraflow, a data engineering firm with offices in Johannesburg, Cape Town and London.

### The experiment
With ten years of data, including GDP Growth Rate, Inflation Rate, Interest Rate, Unemployment Rate and the USD/ZAR Exchange Rate, Kim attempted to find relationships that would allow him to project this year's budget figures. It turned out there were no correlations between them strong enough to make an accurate prediction.

However, by scanning a wide range of global data sets, Kim was able to identify similar trends from completely unrelated sources. He found patterns very much like the GDP Growth Rate in US statistics on the number of people killed by hot drinks, food, fats and cooking oils. Similarly, he discovered a close resemblance between the number of people killed by hot water in the US and the SA Inflation Rate.

Although these trends are completely unrelated in a causal sense, their highly correlated changes in value over time means one may be used to gauge the other. Kim's prediction: a GDP Growth Rate of 0.6% and an Inflation Rate of 3.6% for 2020. "Of course, we'll follow the Budget Speech intently to see just how close we are," he says.

### The importance of correlation
Apart from forecasting the country's economic outlook, identifying important correlations, even between seemingly unrelated data, can have a huge impact on business decisions.

For instance, one financing company in China gave its in-house AI system the task of assessing loan applications made through its mobile app. Whereas banks often employ around 10 measurements to rate an applicant's creditworthiness, the AI appraises some 5000 personal attributes based on available data.

For example, it considers how confidently a person types when applying or, more surprisingly, whether their phone battery is low during the process. It seems the AI identified a hidden correlation for a sense of responsibility, by detecting that people with poor repayment habits often did not ensure their battery was properly charged for this important financial transaction. It's something few human loan officers would, or even could, contemplate when evaluating someone's credit risk. This illustrates the power of leveraging correlations correctly.

### Operationalising data for correlation
To identify significant correlations, organisations must operationalise their data by extracting it from siloed enterprise data sources, transforming it to an appropriate

format for analysis, and collecting it into a centralised data repository where it can be easily accessed.

"Only when data is brought together through a repeatable, systematic process can these relationships be effectively exposed and strategically exploited," says Kim. Whether predicting the country's economic trends, understanding behavioural patterns in consumers or making previously unimagined business decisions, the critical role of correlation cannot be overstated.