

Start using Machine Learning with Google, it's easier than you

think *By Louis van Schalkwyk, head of Technical Operations at*

Digicloud Africa

Harnessing the power of Machine Learning is now within reach of even the smallest business, thanks to a range of new tools recently brought to market by Google.

Machine Learning (ML) is still in its early stages in South Africa and across Africa, although there is already a great deal of interest in the technology. However, there are some misconceptions around ML; notably its applicability in various environments, and that it is costly and complicated to implement.

Getting started

For some organisations and in some circumstances, ML may not be the right solution, and this is where guidance is needed from the likes of Digicloud's partner network. We also come across customers that would like to embark on an ML journey but are not sure how to get started. Key to ML is having enough and accurate data. Our partners can help architect, build and operate data pipelines to help customers operationalise ML in production.

A common pitfall of ML is creating models that overfit your training data, meaning the model works very well with your training data but not well at all with new, unseen data. This usually happens when the training dataset is too small or the model is too complex. Getting help from an experienced ML team can help customers avoid these kinds of pitfalls.

It is important, therefore, to define what your organisation wants to achieve, and what data you have available, before embarking on an ML journey.

Fast-tracking ML

Contrary to popular belief, getting started with ML is not overly complicated or expensive, thanks to Google's platform to help customers get going much faster.

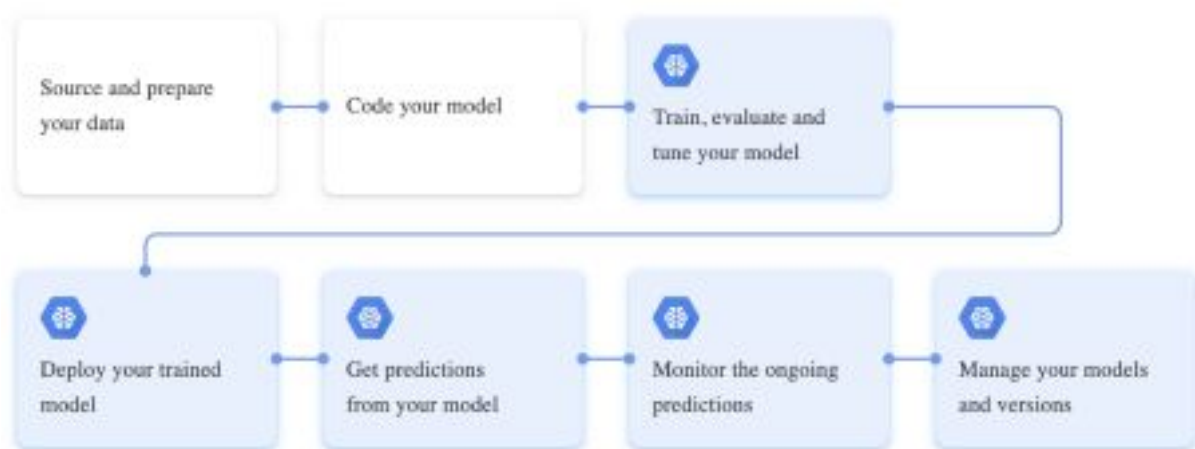
A major component of ML is getting your data ready and cleaned. This can be very time-consuming, so Google offers tools to help customers get their data ready for ML much faster. [Dataprep](#) (by Trifacta) is a serverless platform that allows customers to visually clean and prepare their structured and unstructured data for reporting and ML. It includes rich transformations which are managed through recipes - which can then execute your ETL jobs on a schedule.

Google solutions for all ML requirements

Google's [Cloud AI Platform](#) offers solutions for the entire ML life cycle - prepare, build, validate, deploy & ML Operations, for all customers, no matter where they are in their ML journey.

Companies that want to implement, or even just experiment with ML have a journey to walk. Google understands that different customers are at different stages of their ML journey and offers tools **for any skill level**. Some companies are at the beginning of their journey, meaning they need to design systems to capture data needed for ML, other companies have data but have no idea where to start while some customers are mature in the space and are working on optimising their models while building out new ones.

The boxes in blue below show where Google's AI Platform provides managed services and APIs:



When we look at big data, there's the concept of the "Five Vs of Big Data" - Volume, Variety, Velocity, Veracity & Value. Google offers tools for customers efficiently and effectively handle these -

- Volume - Google's BigQuery essentially consists of 2 decoupled services - compute for SQL execution and storage for your data. Google Cloud provides near unlimited storage and you only pay for what you use.
- Variety - Structured and unstructured data can be stored and analysed in Google Cloud.
- Velocity - Google's messaging queues and ETL pipeline tools can handle data ingestion at global scale, BigQuery for example allows 100,000 rows to be inserted per second into a table, and even more if using [Bigtable](#).
- Veracity - the accuracy or truthfulness of data, which becomes more apparent as you combine data from multiple sources - Google's data preparation tools help you identify possible discrepancies in data.
- Value - actually getting business value from your data, Google provides ongoing tools for running and retraining your data to ensure continued value is derived from it.

Google Cloud AI has prepackaged solutions to help customers solve some of their most important challenges and realise their investment in AI sooner. One such solution is [Contact Centre AI](#), a solution that integrates various Google AI building blocks like Speech to Text, Text to Speech, [Natural Language Processing](#) and [Dialogflow](#) to allow your 'virtual agent' to converse naturally with your customers and provide expert advice to your human agents on complex cases in the background.

Another challenge customers face is the cost of running ML, from the skills needed to the tools required. Google Cloud brings ML to customers and meets them at the skill level and budget constraints they have.

Google offers ML models which are pre-trained and ready to use, with functions such as [analysing photos](#) to understand what's in them, identifying text, objects, animals, people, etc. A useful [Speech to Text](#) API allows customers to translate voice to text (and [Text to Speech API](#) from written text to human like spoken speech).

But let's take that one step further: should a customer want to train a model using their own data, they can use [Cloud AutoML](#), a service that allows you to train models using your own data. Once the model has been trained, it is published as a rest API on Google, ready to be used by your own application. This can all be done without writing complex models; using a point and click UI, it allows even inexperienced users to create and operationalise ML models with ease. Cloud AutoML can work with structured data such as tables, as well as with images and audio.

For companies with SQL skills; Google's data warehouse platform, of which BigQuery is a major component, offers ML through standard SQL syntax. Using [BigQuery ML](#) customers can create and execute ML models using SQL queries - it currently supports around 10 different ML models, depending on your use case.

On the other end of the spectrum, you'll find customers with their own custom built ML models, training these models with lots and lots of data can be very time and compute intensive. Google offers one of the best places to train your models: they have developed hardware specifically designed for training Tensorflow ML models. This hardware - called [TPUs](#) (Tensor Processing Units) can train models 27 times faster compared to using GPUs, at 38% lower cost. Significant business benefit can be achieved by training your model in hours instead of days.

Powering ahead

A common objection to ML is that it's not scientific or academic in nature, it simply learns how to answer a question given a set of input data, it doesn't explain how it came to an answer, it doesn't do research into a topic, nor does it take into account data points that it doesn't have access to. Machine learning is exactly that - answering a question given a set of data.

Google recently announced Explainable AI (currently in Beta) that aims to help customers understand and interpret the predictions made by their ML models. This will help customers improve their models, understand its behaviour and even visually investigate how the model will behave using a what-if tool. Using this tool helps customers to detect and resolve biases in their data, creating more inclusive AI.

Get expert guidance

At Digicloud - Google's enablement and distribution partner for Africa - we work with our network of partners across Africa to bring Google technology to end customers. Our partners have deep, local knowledge and expertise to help customers implement Google AI technologies. Whether you're just starting out with ML or you're already running your own models; our partners are able to assist and guide you through the journey as needed.