# The Data Analysis Bureau

*Smarter Data, Better Decisions*

# The Data Analysis Bureau

The Data Analysis Bureau is an end-to-end data science and advanced analytics service provider with the belief that "**smarter data means better decisions**".

From formulating business-led questions to explore your data, or building machine learning models to on-going expertise. We adopt a highly iterative and DataOps driven process to empower you to return the most value from your data.

Our framework helps our clients to navigate the process to discover, build and then run their own data projects. Our combination of technical and delivery expertise helps you quickly understand, scope and mobilise your projects to deliver actionable insights and value to your business.

# Data Accelerator Framework

Our Data Accelerator Framework is an End-to-End Data Service designed to Accelerate Your ability to make Data-driven decisions with your projects, products or people.

## Discover

Define your data requirements with our Data DnA (Discovery & Advisory) Service. Understand your baseline and available solutions to make better data driven decisions.

Understand your data maturity and identify data projects that deliver real business value.

- Data Discovery
- Technical Expertise
- Data Audit & Validation
- Feasibility Assessment
- Project Roadmap

## Build

Analyse your data set and use our Data Launchpad to rapidly drive ideas through to production with an end-to-end project-based, technology agnostic service.

Explore what AI and machine learning can do for your business.

- Data Analysis & Engineering
- Solution Architecture
- Proof of Concept
- Minimum viable products
- Product Deployment

## Run

Manage ongoing data operations with our Data Services, elevate decision making and benefit from continuous delivery of expertise through our DataOps teams.

Maintain optimal performance of your data models and business insights.

- Project Mobilisation & Incubation
- Managed services
- Resource augmentation
- Project Evolution

***Smarter Data, Better Decisions***

# Typical Engagement Model

We scope project work in iterative 2-week sprints and provide expertise on a time & materials basis. This approach helps customers focus on delivering benefits and a return on their investment. Below is an example of how we can engage and collaborate on a project.

## Discovery Workshop

Identify business challenge

Identify data maturity & identify value add data projects

Formulate business-data hypothesis

Outline further analysis.

## Exploratory Analysis

Preliminary data audit, cleaning, data wrangling & munging

Analytical exploration to assess the suitability of your data

Design and fit statistical models to identify correlations within your data of statistical significance

Validate data projects and proposed concepts to define clear roadmap of development to deliver ROI

## Technical Build

Solution design and build

ML model experimentation

Architecture design and build

Assess current data and technology stack to identify potential opportunities and approach.

## PoC Development

PoC Development

Develop analytics and experimental machine learning

Mobilise or migrate data projects

Access trusted technical expertise

Deliver new products to market

*Smarter Data, Better Decisions*

# Machine Learning for Predictive Analytics

## Prediction of spoilage and failure events in the manufacturing chain for a leading packaging manufacturer

### Background

A global manufacturing company was looking to bring predictive analytics to its packaging production line.

In particular, they were keen to understand how machine learning could be applied to reduce machine downtime and spoilage from production errors.

### Solution

T-DAB initially used one years worth of data to use machine learning to firstly mine the dataset for key influential features from an initial list of 64, and then apply machine learning to predict spoilage and tool failure events within future time periods. Included were machine state, output quality, tool life and operational data.

T-DAB first carried out a data audit, cleaning, and wrangling exercise, followed by feature engineering. Machine learning experimentation was carried out in R.

The end result was that a number of ML algorithms were produced able to predict spoilage and tool failure events to a degree of accuracy significant enough (>80%) to have real world impacts on operational processes in reducing spoilage and downtime.

### Benefits

Through the presentation of predictions of spoilage event categories through an easy to understand, interactive UI, machine operators were able to intervene earlier in order to reduce the probability of spoilage. Models not only gave early warning of future spoilage levels, but were also used to return to the user more optimal machine settings than the standard settings, in order to minimise spoilage.

# Machine learning for employee churn

## Time-to-event prediction for sales agent attrition of a major telesales company

### Background

A major telesales company had been suffering attrition rates of it's sales force of over 10%. Managers were caught in 'analysis paralysis', with too many spreadsheets, insufficient visibility, or analytical bandwidth to understand when individual sales agents were at risk of leaving the business..

The brief was several fold. Firstly to develop a machine learning algorithm to predict the number of weeks remaining tenure, as well as a certainty score to assess the accuracy of that prediction compared to past events. Secondly, develop an easy to interpet dashboard to present predictions and accompanying information to a non-technical audience. Finally, to deploy the ML to production on the client systems.

### Solution

Using three years of historical data across 290 initial variables, T-DAB carried out a full ML development workflow from data audit, cleaning, wrangling and feature engineering, to ML experimentation, testing, benchmarking and validation.

Prototyping and experimentation was carried out using standard Python libraries such as Skitlearn and NumPy. A final model, once tested, benchmarked and validated was then pickled and made ready for production. Productionisation was implemented using visual studio, MS SQL Server 2017 combining scheduled tasks of Python pickle files and

A dashboard was developed in Power BI for presentation of the outputs and included interactive graphs, tables, filters and dynamic call out boxes.

### Benefits

Through use of the dashboard, managers of the sales agents are now able to quickly and easily identify early, agents that are at risk of leaving (within 6 – 10 weeks). They are presented with the number of weeks remaining and a certainty score.

In addition, they are presented with easily accessible accompanying information that was previously in many disparate sources.

Because the model highlights sales agents conservatively, managers have plenty of warning to take remedial action, improving the working conditions of sales agents, and minimising losses incurred from attrition to the business.

# Cloud Architecture for Predictive Analytics

Cloud architecture and machine learning to enable predictive analytics for a super-material manufacturer

## Background

A high end innovation, research and manufacturer of advanced super-materials was looking to data science and machine learning as a way to bring efficiencies to their innovation cycle through greater insight and automation.

T-DAB were engaged to design a cloud hosted architecture to host a database and analytical engine. The client required that the solution ingest individual .csv files of data from historical and new materials tests. These were stored through an automated, bespoke ETL process. This data was then analysed and used to train ML algorithms to support a number of innovation challenges.

## Solution

The Data Analysis Bureau designed and built a suitable AWS architecture to batch ingest and database test data from individual .csv files

This consisted of an automated process for file upload to Amazon S3. A scheduled Amazon ECS C# process pulled bucket and data inserted to MS SQL database. For security, this was contained in a private subnet. An amazon EC2 R server instance in a public subnet was connected to MS SQL DB. IAM role and security group restricted access to ECS and EC2 only. An elastic load balancer in a public subnet above the EC2 R Server instance subnet and IP access was restricted using Security Groups. The R server instance was connected to provide an analytics layer. This was used for training ML regression-like algorithms to predict super-material performance.

## Benefits

The architecture carries out automated batch ingestion from the client file system and analytics are regularly carried out using the R Server instance and the MS SQL DB.

This enabled the client to carry out testing in a systematic and holistic manner, as well as visualise and analyse the results of testing as a whole. This has already led to novel insights into material properties and performance.

More importantly, the application of ML to predict material performance has allowed the client to speed up innovation by identifying abnormal tests, as well as make accurate testable predictions of material properties

# Deep Learning for Intelligent Automation

TRIGGER RACING Imperial College London Microsoft

## Deep Reinforcement Learning and Recurrent Neural Nets for Autopilot Control

### Background

High performance ocean racing boats have a wealth of real time IoT data collected by sensors, aggregated and processed. These are used to help the sailor optimise and manage the performance of the boat as they race single handed non-stop around the world. In order to help them do this, sailors actually use an electronic automated autopilot to steer the boat for 98% of the time spent racing. However, analysis shows that these autopilots only perform to 80% of human capability. At the same time, the percentage difference between the last winner of the Vendee Globe, was only 2% time. There is therefore a great opportunity to address this performance gap through machine and deep learning. In addition, there are technical challenges around the remote deployment of models to lower power, low connectivity environments.

### Solution

T-DAB is working alongside a professional sailing team, an academic partner, and technology provider to develop cutting edge deep learning solutions. This involves leveraging the latest deep learning frameworks (TensorFlow, PyTorch), as well as distributed computing and machine learning development technologies (Azure Databricks). Algorithms will, rather than optimise to maintain a constant heading or wind angle, instead optimise for the velocity of the boat to the point it is trying to reach. The solution will account for sea state, boat configuration, and a range of other features not currently accounted for. Multiple ML/DL approaches are in development, along with simulation models for testing and benchmarking. In addition, T-DAB are working with Microsoft to leverage Azure IoT Edge to efficiently ingest data and deploy algorithms remotely via satellite link.

### Benefits

The benefit is a cutting edge advantage. Currently, huge investment is placed by teams in sail, hydrofoil, and composite material development. Route optimisation is also mature. However, little to no progress has been made in marine autopilot control systems despite them being critical to performance.

At it's full potential, the solution can deliver a Vendee Globe winning edge, a challenge that fewer people have survived, completed and won than have walked on the moon.

https://www.vendeeglobe.org/en/presentation

# Key Contacts

If you have any questions or need any more information on The Data Analysis Bureau, please contact the team below:

**George Hancock**
Head of Business Development
George.hancock@t-dab.com
07715 455071

**Dr Eric Topham**
CEO & Data Science Director
Eric.topham@t-dab.com
07728 729208

*Smarter Data, Better Decisions*