List of Projects with Consulting and Delivery Contributions

- **Classification of cis-Regulatory Elements in Mouse and Human Genome (2018/19)**

This project was concerned with the prediction of genome-wide distribution of cis-regulatory elements namely enhancers and priming elements in functional genome. The algorithm was applied to a genome wide data of just under a million genomic segments spanning 6 stages of hematopoietic development. For each cell differentiation stage, the data was captured for the presence of transcription factors, histone marks, chromatin openness, mRNA, and footprint data from a large collection of the known motifs. Iterative Random Forests was applied to generate a distribution of enhancers, which was later verified by the web-lab scientists using state of art assays. This verification is still in progress and up to 2k of the predicted enhancers are already verified. This work was done using RDS and cloud technologies. The statistical tool was R and the visualisation was done on Cytoscape.

- **Hypoxia and Cancer Genomics (2018/19)**

This project investigated cancer genomics (Leukaemia) and its association with the differential gene expression pathways during the embryonic growth of hematopoietic stem cells. I worked together with wet-lab scientists using the in-house facilities to grow mouse stem cells which were monitored during the 6 embryonic cell differentiation stages namely ESC, Mesoderm, Hemogenic Endothelium, Hemogenic Progenitors, and Macrophage. My contribution to the project was to build a spatiotemporal model to investigate the dynamics of epigenetic material such as chromatin, transcription factors, histones, and mRNA. It was from the temporal profiles of these models, that provided biological insight on the dynamical properties of the genetic material at a mechanistic level. Such insight was then used to create hypothesis for the next cycle of bioinformatics analysis. This work was done using RDS Cloud-technology on R. The visualisation of the results was done using Cytoscape and MATLAB.

- **Chemical Process Characterisation Work-flow System (2020)**

This project was to develop a work-flow system for a team of 90 process scientists consisting of upstream, downstream, and analytics. The work-flow system was developed from scratch to incorporate statistical methods in DoE work for staff who had limited to no prior experience in the use of statistical machine-learning or other data analytics techniques. The work-flow system incorporated a decision-based framework based on the availability of the number of runs and the number of factors to be characterised. I was the lead data scientists with start-to-end responsibility for the development and implementation of the work-flow system. The system is now a standard of procedure for the process characterisation work within the team.

- **Optimization of Nitrogen Cycle in Agricultural Fertilisers (2020)**

This project aimed to develop a mechanistic model for Nitrogen cycle during the life span of crop plants. NH3-based fertiliser was investigate on a spatiotemporal grid of three large agricultural sites in USA. My contribution to the project was to consult the team on building mechanistic spatiotemporal (reaction-diffusion) model and fit it to the data.

- **Mathematical Modelling of Oxygen Sensors (Clark Electrode) (2014)**

This project was my MSc thesis at the University of Oxford. I used mathematical modelling and adaptive mesh to quantify the electric current generated at the surface of the electrode of a multi-layered Oxygen Sensor (Clark-electrode). This device was invented original by Leland Clark back in 1956 and was the first medical device used for quantitative measurement of Oxygen concentration in patients' blood. I used analytical and numerical techniques to simulation the diffusion process of Oxygen partial pressure and demonstrated that adaptive mesh contributes to the accuracy of electric current arriving at the surface of the electrode in the result of the Redox reaction. I was responsible for both the computational and analytical work of this project. The computational work was conducted using personal PC and MATLAB, which was also used for the time-dependent simulation of the performance of the sensor model.

- **Data Analytics Case Study on the World Happiness Data by Country (2021)**

This was a case study conducted for a private client who required a predictive model for the happiness score of a given country based on a few features that were recorded since 2005 through to 2021, such as GDP per capita, corruption, life-expectancy, social support, and a few more. There were three main parts to this study,

i) Descriptive statistics at regional level, in which the Globe was divided into 10 regions (continental and sub-continental). Time series profile of each region was generated, and a cross comparison was conducted.
ii) Unsupervised clustering algorithm was applied to divide 166 countries of the Globe into 6 disjoint clusters and feature-based similarity of the clusters was further analysed and reported.
iii) Multiple regression model to predict the happiness score of a given country based on the knowledge of these features. Using the model, the happiness score of certain countries was predicted and the general trends of happiness score was compared between the pre and post pandemic times.

This work was done on python using virtual conda environment on Unix platform.

- **UK Power Consumption Predictive Modelling and Source Comparison (2019)**

This project was a data-driven predictive model built on time series data spanning a period of over 11 years in which 22 features were captured with a temporal gap of about 5 minutes between observations from 2008 until 2019. The data was sourced from UK national grid and my contribution to this project was to build a time series predictive model using Fourier transform to predict future demand of energy at national scale and the quantitative comparison between five of the most important sources of energy in the UK, namely coal, solar, nuclear, wind, and biofuel. This work was conducted using Unix environment on Python. The model was developed on Jupyter Notebook and then turned into a pipeline. The visualisation was done using python library matplotlib and seaborn.

- **Bayesian Parameter Estimation for PDEs (BPEPDE) (2018)**

This project was a short-term contract during the final few months of my PhD at Sussex. I was tasked to develop a user-guide documentation for a python-based machine learning package, which was prospected to become commercial. The algorithm integrated Bayesian statistics with MCMC methods and iteratively fitted the parameters of mechanistic models with data. The algorithm later delivered provable results in 3-D cell-tracking biotechnologies in Switzerland and Germany. This work was conducted using personal PC with sufficient Ram and Storage. The python library matplotlib was used to visualise the results for the test runs.

- **Modelling Stingray Spot Evolution by Reaction-Diffusion Systems (2018)**

This project was a part of my PhD thesis and somewhat intellectually motivated case study. My specialisation is the temporal dynamics of pattern formation models of Turing type. One of the geometries on which I was exploring the dynamics of patterns formation was disc-shape surface. The domain-dependent analysis of the pattern formation models was developed on the disc-shape, and it was computational demonstrated that spots do undergo temporal changes, which agreed with the seasonal changes of spot formation on a certain family of stingrays. This work generated the second publication toward my PhD and the article was featured in the International Journal of Bifurcation and Chaos. The Journal selected the computer simulations for the cover page of the volume in which the paper was published. The news was also echoed by the Sussex University Research media at the time. The computational simulation of this project was conducted on computer cluster provided by Sussex University. The finite element solver was written from scratch using MATLAB, which was also used for the visualisation of the results.

- **Modelling Honeycomb Panels for Soundproof in Construction and Aviation (2015)**

I worked on this project for a client and the aim was to quantify the acoustic characteristics of honeycomb panels using Sound Pressure Level (SPL) and Sound Transmission Loss (STL) combined with a detailed study of extracting natural frequency of sound, whilst treating air as the medium of sound propagation. My contribution helped to construct from scratch a method by STL and SPL was analytically developed and implemented using ANSYS. To preserve the freedom of choice in the size of honeycomb and material characteristics, the simulations are executed using APDL rather than GUI. My contribution to the project was also in the implementation of the finite element on adaptive 2-D mesh grid and in the analytical derivation of SPL and STL equations. I was responsible mainly for the analytical side of this project and other members of the teams conducted the computational work using ANSYS.

- **Wear Analysis of Aircraft Wheel by Modelling Temperature at Touch Down Friction (2016)**

This project was to investigate factors contributing to the wear of aircraft wheel. It had two components namely optimize the pre-touch down velocity of aircraft and developing a model to analyse the touch-down related friction induced temperature diffusion into the wheel material. My contribution in the project was to develop from scratch

a heat-diffusion model, that was implemented using MATLAB on a moving wheel. The second part of my contribution was related the diffusion-model with various pre-touch down velocities and find the optimal velocity that minimized the temperature generation at and post touch down. I was mainly responsible for the analytical development of a heat-diffusion model from the surface of touch-down whilst the wheel is spinning. The computational work of this project was done using ANSYS.

- **Analysis and Predictive Modelling of Car Price (2021)**

This project was delivered for a private client who needed a car price predictive model using several qualitative and quantitative features. The descriptive and inferential statistics were conducted to analyse the significance of difference in certain features of cars based on their manufacturing characteristics. This dataset was analysed using excel and the data-analysis add on within was employed to develop a multiple regression model for price prediction.

- **Positioning Index (2020)**

This project was to conduct competitor analysis for higher education institutions based on the fluctuation and analytics of the annual report post to the publication of the same. We created performance-based metrics for each of the client and then applied unsupervised clustering algorithm to relate the annual report performance differences to the space of admission and assessment policies. This project was half way paused by the first wave of the pandemic. My contribution to the project until it was going was to create the workflow plan from the hypothesis generation to the insight extraction.

- **Leaky Hosepipe (2020)**

This project was concerned with preventing the loss of applicants for higher education institution by delivering data analytics and descriptive statistics solutions. We investigated the collective journey of the applicants from the time of submission all the way through to the first day of the academic year. By data analytics we identified and provided data-driven consultation to the institution on what policies to devise in order to maintain the initial number of the applicant and thereby prevent leakage/loss of applicants. My contribution in the project was to deliver stage-based descriptive statistics and extract data-driven insight for action.

- **Learner Personas (2020)**

This was unsupervised machine-learning pipeline that classified the personas of clients based on the characteristics within the data. K-mean clustering was applied to develop the pipeline, and this was a commercial data analytics product for clients from higher education institutions in UK. My contribution to this project was to develop modify the relevancy score algorithm and provide as predictive feature for the classification of the clusters.

- **Offer Rate Calculator (2020)**

This project was to develop a supervised machine-learning product for predicting the prospect of success for securing a place at a given university within the UK when the A-level qualifications and grades of the user are known. The logistic regression for the initial proof of concept was applied and the product was deployed to the website for public use. By replacing the logistic regression model by Random Forests, the model was significantly improved in terms of its accuracy and ROC measures. I led this project and my specific portion to the workflow was to develop a probability-based algorithm for relevancy score of A-Level qualification with any given degree program.