

Impact case study (REF3)

Institution: University of Plymouth		
Unit of Assessment: UoA10		
Title of case study: Impact Lab: An Environmental and Big Data Hub for SMEs		
Period when the underpinning research was undertaken: 2015-2020		
Details of staff conducting the underpinning research from the submitting unit:		
Name(s):	Role(s) (e.g. job title):	Period(s) employed by submitting HEI:
Dr Antonio Rago	Associate Professor	2011 - present
Dr Luciana Dalla Valle	Associate Professor of Data Science and Statistics	2011 - present
Dr Matthew Craven	Lecturer in Applied Mathematics	2013 - present
Dr Craig McNeile	Lecturer in Theoretical Physics	2013 - present
Dr Mu Niu	Lecturer in Statistics	Sep 2017 - Aug 2019
Dr Malgorzata Wojtys	Lecturer in Statistics	2013 – present
Period when the claimed impact occurred: 2017-2020		
Is this case study continued from a case study submitted in 2014? N		
<p>1. Summary of the impact (indicative maximum 100 words)</p> <p>The Environmental Futures & Big Data Impact Lab brings together world-class academic research and South West businesses to stimulate the development of new products and services and to see them through to market success. The lab as a whole has already worked with 43 different clients leading to various collaboration schemes and successful Knowledge Transfer Partnerships (KTPs). Our research has led to innovative business solutions with a direct effect on the market positioning and economic prosperity of SMEs. We have reduced costs to insurance companies and contributed to increased sales of the UK's market leading digital printer company. We have become an established provider for developing Data Analytics, Data Science and Big Data skills, which are in short supply, by delivering a large array of workshops and Continuing Professional Development (CPD), with more than 200 attendees having already taken part to our courses.</p>		
<p>2. Underpinning research (indicative maximum 500 words)</p> <p>The Environmental Futures & Big Data Impact Lab is a resource to help businesses solve a key technical challenge in the development of a new product, service or process. It was based on a successful ERDF Priority Axis 1 – Promoting Research & Innovation bid from South West institutions including Exeter (PI) and Plymouth, awarded in 2018 [G1]. The Impact Lab is a partnership of seven world class organisations, and the Centre of Mathematical Sciences (CMS), University of Plymouth (UoP), represents the Big Data pillar of the Plymouth Impact Lab and has been directly involved in 16 of the 43 collaborations with SMEs.</p>		

In the current data revolution, the ability to relate data on weather patterns or climate change to other datasets, to generate predictions about consumer behaviour, transport use, crime, risk of mechanical breakdown, extreme weather events, human health or agricultural production can deliver huge gains both for private sector productivity and profitability and for the efficiency of public sector service delivery. Data analytics is about correlation, probability and prediction (the focus of methods experts), underpinned by an understanding of causation (the focus of domain experts).

The CMS research group behind the Plymouth Impact Lab are world-class experts in data analytics techniques, developed in their main research fields. CMS members Craven, Dalla Valle, McNeile, Niu, Rago and Wojtys, in particular, developed the data analytics and machine learning techniques underpinning the Impact Lab activities. These can be summarised as follows:

1. Machine learning - Evolutionary algorithms Work carried out on the HPC (High Performance Computing) Cluster at the UoP concentrated on AI algorithms (evolutionary, machine learning and automatic design of algorithms) in work programmes of solving cryptographic and financial optimisation problems [3.1]. This was backed up by projects to demystify the 'black box' nature of such algorithms, focusing on the realm of visualisation of algorithmic processes [3.2] and injecting problem understanding into these algorithms.

2. Machine learning - Classifiers: One of the most popular and successful methods of classification are families of classification trees based on the so-called "boosting" algorithms. Wojtys studied regularization of families of classifiers using classification trees in [3.3]. Using an extensive simulation study, she investigated practical aspects of obtaining a regularized classifier that is optimal in the Bayesian sense. In particular, she determined that some theoretical methods of regularizing families of classification trees proposed in the literature suffer from large variance and instability in practical applications and therefore should be approached with caution.

3. Data Science – Risk Modelling: In the research strand focusing on statistics and analytics data science, predictive analytical tools utilizing machine learning techniques are implemented in [3.4] to better exploit and gain more value from large insurance data sets, particularly for commercial property risk assessment. The implemented approach was based on the aggregation of complex insurance databases from various sources, including open-source data, and on the application of machine learning methodologies to assess risks. The research delivers effective predictive risk models by exploiting different pattern recognition techniques, such as neural networks, deep learning, K-nearest neighbour and classification trees. This predictive analytics approach was successfully adopted by the SME Software Solved to obtain an accurate and efficient insurance risk assessment of commercial properties in various locations of the UK and abroad.

4. Big Data – Sentiment Analysis: In the "Big Data" era, social media are amongst the most prolific generators of information on people's opinions and sentiments towards specific topics, products and services. Focusing on specific locations in the UK, we developed a novel methodology to integrate social media and environmental data, based on time series analysis [3.5]. Our approach allowed a machine driven detection of real-world events. In particular, we considered the use of social media for observing natural hazards and we developed a methodology to detect and locate flood events by combining data from different sources.

3. References to the research (indicative maximum of six references)

G1 Grant Awarded to Exeter University – Col University of Plymouth. ERDF Priority Axis 1 – Promoting Research & Innovation Grant Dates 2018-2021 and subsequently extended to June 2022 [<https://www.plymouth.ac.uk/research/impact-lab>]

- 3.1 M. J. Craven, D. Robertz, A Parallel Evolutionary Approach to Solving Systems of Equations in Polycyclic Group, Groups Complexity Cryptology 8 (2) (2016), 109-125. [<https://pearl.plymouth.ac.uk/handle/10026.1/6507>]
- 3.2 D. J. Walker, M. J. Craven Identifying Good Algorithm Parameters in Evolutionary Multi- and Many-Objective Optimisation: A Visualisation Approach, Applied Soft Computing, Volume 88 (2020) 105902, ISSN 1568-4946, [<https://pearl.plymouth.ac.uk/handle/10026.1/15177>]
- 3.3 M. Wojtys (2019), Families of classifiers, their regularization and Bayesian optimality, *working paper*.
- 3.4 Dalla Valle, L., & Kenett, R. (2018). Social media big data integration: A new approach based on calibration. *Expert Systems with Applications*, 111, 76-90. [<https://pearl.plymouth.ac.uk/handle/10026.1/10453>].
- 3.5 Stander, J., Dalla Valle, L., Eales, J., Baldino, A. and Cortina Borja, M. (2016) The EU referendum: extracting insights from Facebook using R. *Significance*.
- 3.6 Published 19/5/2016, updated 20/6/2016. [<https://pearl.plymouth.ac.uk/handle/10026.1/5418>].

4. Details of the impact (indicative maximum 750 words)

The Impact Lab has been acting as a valuable support for businesses to get in touch with the most advanced techniques in High Performance Computing (HPC), Data Science, Big Data and Mathematical Modelling, which are at the core of our research activities [5.1]. Our research has led to innovative business solutions with a direct effect on the market positioning and economic prosperity of the SMEs, as detailed below. It has also led to SMEs now routinely incorporating a range of cutting-edge techniques in their workforce capabilities. This has been achieved through our Knowledge Transfer Partnerships (KTP) and our Continuing Professional Development (CPD) courses.

Upskilling businesses through CPD courses

The Plymouth Impact Lab has run three CPD courses covering the latest cutting-edge technologies in Data Science and Big Data. There were 65 attendees, 45 of which were from SMEs (subsequent courses postponed due to Covid-19). This has resulted in the upskilling of the participants and the direct engagement of the SMEs with a novel topic [5.2]. Aneeq Ur Rehman, Software Solved said *"It's all highly relevant to the research work I'm doing at Software Solved into the application of machine learning and advanced data analytics for risk modelling and mitigation for the insurance sector... It's technical but truly fascinating and a fundamental part of how organisations will operate, interpret and derive value from the data in the 21st century- the advantages of which are manifold."* [5.3]

Reducing costs and staff resources within insurance companies

The UK insurance sector is the largest in Europe and a leading global centre in a truly global industry. However, the obsolescence of the current risk assessment systems raises a significant need to automate processes and increase system integration, data visualization and predictive analytics. The Plymouth Impact Lab has improved the performance of Software Solved, a South West SME specializing in custom data and software solutions through a KTP collaboration involving Dr Dalla Valle and Dr Ansell (Data Science Research Assistant, Impact Lab). The KTP project [5.4] involved the application of machine learning and computational intelligence to normalize numerous data sources and develop a predictive risk management tool for the insurance sector. The predictive risk assessment algorithm developed through the KTP collaboration has greatly enhanced the speed at which risks can be calculated, has improved accuracy and has resulted in more realistic premiums, making the company more competitive. The research conducted has also reduced costs to the insurance companies by removing the

need for building surveys to assess unusual risk profiles. It has also freed up staff time to deal with other activities such as the support of claimants at a time of need and ensured that corporate businesses are up and running much sooner following a claim. Software Solved has experienced growth and interest in machine learning in other industries that they service, including public sector, logistics and warehouse management. [5.5] *'Software Solved greatly benefitted from the collaboration with Impact Lab, which contributed to the integration of different data sources in the company product offering and has also opened doors for the company to hire interns for small to medium projects from Exeter city futures.'* Jon Stace, Director of Technology. Software Solved. [5.6]

Process innovation resulting in increased sales

The Impact Lab supported the UK's market leading digital printer and photo gift fulfiller, Harrier, leading to increased sales of their digital product. Our research resulted in process innovation that contributed to the development of a new volume-forecasting model. The model creates more accurate and meaningful predictive forecasting. *'A new volume-forecasting model has been devised using external sources of data that significantly influence sales.'* Lewi Griffiths, Director, Harrier [5.7]

5. Sources to corroborate the impact (indicative maximum of 10 references)

5.1 Report from Impact Lab PI In support of the Plymouth group (2020).

Statement of CMS involvement in the Impact Lab activities

5.2 Feedback reports from the CPD courses:

Letter of support from SME Harrier LLC

5.3 Weblink of the SoftwareSolved blog on the CPD

<https://www.softwaresolved.com/blog/no-small-talk-for-big-data-software-solved-spend-time-with-amazon-scientist-at-plymouth-university/>

5.4 General audience publication on the creation of the KTP

collaboration <https://www.devonlive.com/news/business/exeter-firm-leading-pioneering-work-2508131>

5.5 Letter of support from Software Solved (2020). Quantifying the impact of the work developed within the KTP remits.

5.6 Impact of the CPD to the Software Solved policies for data science

5.7 Letter of support from Harrier