Institution: University of Essex

Unit of Assessment: 10 – Mathematical Sciences

Title of case study: Creating new predictive data science tools to improve digital marketing success with Profusion Media Ltd

Period when the underpinning research was undertaken: 2009-2020

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:
Berthold Lausen	Professor of Data Science	Jan 2009 - present
Aris Perperoglou	Reader in Statistics	Oct 2012 - Mar 2019
Angel Rolando Medellin-Gasque	KTP Associate	2015 - 2016
Ilan Fridman Rojas	KTP Associate	2017 - 2018
Juan Pablo Equihua	KTP Associate	Nov 2019 – present

Period when the claimed impact occurred: 2015-2020

Is this case study continued from a case study submitted in 2014? N

1. Summary of the impact

Research at Essex has transformed the way Profusion Media Ltd undertakes digital marketing, leading to new revenue streams for Profusion and increased income for commercial clients. Essex data science research in event time models, feature engineering and machine learning was embedded within Profusion's first product, the AI Marketer. This acted as a catalyst for Profusion's transformation from a 'services-only' company to a hybrid 'services and products' company through the creation of a separate division. The application of Essex research to address the business challenges of Profusion's customers has led to significant benefits: a major British home improvement retailer gained over GBP13M in 5 months; a luxury automotive company gained over GBP30M by December 2020.

2. Underpinning research

Successful digital marketing requires an understanding of why customers may choose to purchase from companies at certain times and not others. Improving the estimating of future customer transaction value allows better allocation of budgets and supports informed financial and strategic decision making. Such tasks are referred to as Customer Lifetime Value (CLV) modelling. Marketing efforts are typically focused on retaining customers with a high potential lifetime spend and hence many companies now collect large data sets of customer purchasing behaviour in order to interpret purchasing patterns, make predictions, and target their marketing. To do this more efficiently, new data science methodologies are required. Collaboration between Essex and Profusion Media Ltd (Profusion hereafter) intensified from 2015 onwards resulted in the ESRC and Technology Strategy Board funding a Knowledge Transfer Partnership (KTP) between the University of Essex and Profusion [**G1**].

Medellin-Gasque and Fridman Rojas integrated University of Essex research in specialised software that was utilised in large datasets on the marketing campaigns for clients of Profusion, leading to the development of their first product, the AI Marketer. The scalability was made possible by adapting the penalty optimisation approach [**R1**]. This directly addressed the optimal

frequency that clients should send marketing emails so that individual customers stay engaged with the brand. Timing between campaigns was determined using penalised survival models [**R2**]. Using machine learning methodologies an algorithm was developed to enable a large retail customer of Profusion (a British multinational, the largest home improvement retailer in Europe, with 1,300 stores in ten countries across Europe, and the third largest in the world) to identify customers as being neutral, active, at risk of not interacting again with services within a certain duration (churning), or churned based on their frequency of purchase and the time elapsed since their last purchase. The research led to the development of an R package available for download on the CRAN website [**C1**]. The model had a predictive accuracy of 91% for churned customers and it allowed for the smart targeting of inactive customers.

The AI Marketer was further developed by applying and implementing machine learning approaches and feature engineering developed at Essex [**R3, R4, R5**]. Optimal tree ensembles (OTE) [**R3**] exploit the observation that the predictive performance of a random forest ensemble is highly associated with the strength of individual trees and their diversity. LASSO, random forest, optimal tree ensembles (OTE) [**R3**], ensemble of a subset of k-NN classifiers [**R4**], new feature selection methods [**R5**] and deep neural networks (DNN) were implemented. The response variable represented the value of all transactions in the future year and the feature space was computed from the past transaction data. No other information about the customer's characteristics or the features of the coming year was used. We demonstrated that OTE and random forest outperformed DNN and LASSO modelling within the AI marketer.

3. References to the research [can be supplied by HEI on request]

R1 Perperoglou A (2014), Cox models with dynamic ridge penalties on time-varying effects of the covariates, *Statistics in Medicine* 33, 1, 170-180. https://doi.org/10.1002/sim.5921
R2 Zuliana SU, Perperoglou A (2016). The weight of penalty optimization for ridge regression. In: Wilhelm A, Kestler H (eds), Analysis of Large and Complex Data, series: Studies in Classification, Data Analysis, and Knowledge Organization, Springer, 231-239. https://doi.org/10.1007/978-3-319-25226-1_20
R3 Khan Z, Gul A, Perperoglou A, Miftahuddin M, Mahmoud O, Adler W, Lausen B (2020), Ensemble of Optimal Trees, Random Forest and Random Projection Ensemble Classification, *Advances in Data Analysis and Classification*, 14, 97-116. https://doi.org/10.1007/s11634-019-00364-9
R4 Gul A, Perperoglou A, Khan Z, Mahmoud O, Miftahuddin M, Adler W, Lausen B (2018), Ensemble of a subset of kNN classifiers, *Advances in Data Analysis and Classification* 12, 827–840. https://doi.org/10.1007/s11634-015-0227-5
R5 Mahmoud O, Harrison A, Perperoglou A, Gul A, Khan Z, Metodiev MV, Lausen B (2014). A feature selection method for classification within functional genomics experiments based on the proportional overlapping score. BMC Bioinformatics. 15 (1), 274. https://doi.org/10.1186/1471-2105-15-274

Software:

C1 Stoyanov K, Nordmark H, Perperoglou A, Medellin R, Fridman Rojas I, Lausen B (2018), hkclustering: Ensemble clustering using k-means and hierarchical clustering, R package version 1.0.1. https://CRAN.R-project.org/package=hkclustering

Grants:

G1 Lausen, B., Perperoglou, A., Citi, L., Knowledge Transfer Partnership (KTP) with Data Science and Marketing Services Company Profusion, London, GBP215,000 (2014-2018) Economic &

Social Research Council (20%), Innovate UK (80%);
https://info.ktponline.org.uk/action/details/partnership.aspx?id=9713
G2 Lausen, B., Scherp, A., Russel-Rose, T., Ali, M., Knowledge Transfer Partnership (KTP) with Data Science and Marketing Services Company Profusion, London, GBP264,000 (2019-2022); Innovate UK (100%). https://info.ktponline.org.uk/action/details/partnership.aspx?id=11896

4. Details of the impact

Profusion is a UK-based data science and marketing services consultancy company, which specialises in helping retailers, FMCGs, SMEs and financial services businesses to grow their businesses using the power of data science. Its clients include Kingfisher, Topps Tiles Screwfix, HSBC, and First Direct [**S1**]. From 2014 Profusion worked together with Essex researchers '*towards applying state of the art statistical techniques and machine learning methods to solve business problems'* CEO/Director of Science, Data & Innovation, Profusion [**S2**]. Profusion goes on to note: '*In retrospect, this was a very significant moment in Profusion's history as a company. It marked the beginning of a transformation for Profusion from a company mainly focused on digital marketing services with a strong emphasis on email delivery towards the company that it is today in which data science, data engineering, data visualisation and data strategy are Profusion's core offerings to the market*' [**S2**].

Improved marketing and business performance for Profusion clients

Profusion applied Essex research to address the real world business challenges faced by its customers [**S3**]. The CEO/Director of Science, Data & Innovation at Profusion testifies that Essex research was '*very beneficial to the successful delivery of projects for Profusion's clients*' [**S2**]. Profusion worked with a leading DIY retailer on a project which addressed a need to predict customer churn, i.e. when a customer might stop making any further purchase [**S2, S3**]. Profusion notes: *'Essex's mathematical sciences research [**R1**] was applied to develop a sophisticated customer churn model with a predictive accuracy of 91% for churned customers. This enabled the firm to predict which of their customers were most at risk of churning and design a win back email campaign to retain as many customers as possible.* ***This generated approximately £4 million from previously inactive customers who were at risk of becoming permanently churned in the first month of using the algorithm****. Subsequently, this number* ***rose to an uplift of £13 million in the first 5 months and which was reflected by 76,264 inactive customers shifting to being active****.'* [**S2, S3**].
One of Profusion's customers, a leading bank, wanted to determine the optimal frequency of sending their marketing emails to increase customer engagement [**S3**]. Profusion state that this was *'accomplished in collaboration with researchers at Essex by adapting the penalty optimisation approach to enable the scalability of Penalised Regression with Individual Deviance Effects as proposed in [**R2**]. The approach had* ***good results improving the open rates and dwell time of the email marketing communications****'* [**S2**]. This work led to **a member of Profusion's Data Science team winning the Data Scientist/Big Data Leader category at the data IQ Talent Awards 2016** for the innovative application of these mathematical methods in a new business context [**S2, S4**]**.**

Predicting customer behaviour can give a company the edge in a fierce and competitive market and Profusion worked with a luxury automotive company to design a model that could predict when its customers were likely to buy a new car. Research from [**R3**] was '*applied successfully to predict a window of opportunity for when different customers are most likely to be receptive to make a purchase. These insights were given to sales agents in different locations worldwide.* ***This has***

created an uplift of £30 million to date' [**S2, S3**].

Profusion worked with a global airline, to address their need to identify customers who were most likely to purchase a flight within the next three months, a challenging problem because of the complex and very large dataset. Profusion testifies that research from [**R5**] '*was used to discover the most important features to focus on to predict which customers were likely to fly'*. They add '*This allowed the* ***project to be delivered well beyond the performance contractually agreed of 80% precision and recall, exceeding their expectations with 93% precision and 97% recall in identifying customers*** *that would go on to purchase a flight'* and conclude that '*this project* ***provided Profusion with an initial contract worth £53k***' [**S2**].

AI Marketer and transformation of Profusion business

In addition to the examples above, Essex research led Profusion to develop a new product, the AI Marketer. According to the Profusion CEO/Director of Science, Data & Innovation, this online platform '*has within it some of the most successful statistical and machine learning methodologies we had learned to apply, including many methodologies directly inspired by research done at Essex such as [**R3, R4**] to make customer lifetime value predictions or [**R1**] for propensity to purchase*' [**S2**]. The AI Marketer was launched in 2020 and the University of Essex and Profusion were awarded a second KTP in July 2019 [**G2**], to further develop the product [**S5**]. The development of the AI Marketer product is an entirely new revenue stream for Profusion, who previously only focused on consultancy (services rather than products) [**S2**]. [text removed for publication] The AI marketer enabled Profusion for example to win an important new retail client at a time when many retailers were suffering due to the pandemic. This **new contract is worth GBP70,000 of work initially, followed by annual recurring revenue of GBP20,000** for the continued use of the AI Marketer [**S2**].

Expansion of Profusion's data science capability and skills

[text removed for publication] The development of a product function in the business with the AI Marketer also led to two direct new jobs being created, further jobs created as a result of partial redeployment of existing resources to support this [**S2**] [text removed for publication]. Profusion CEO/Director of Science, Data & Innovation state: '***Lausen and Perperoglou's research increased the depth and breadth of Profusion's Data Science team's skills and knowledge directly*** *and* ***its application enabled the recruitment of new members****. Moreover, if it was not for the success of selling data science services, which benefited heavily from knowledge sharing and collaboration with the university, the Data Science team would not have grown to the extent which it has with* ***12 FTE jobs created since 2014****.'* [**S2**].

5. Sources to corroborate the impact

S1 Profusion's Website screenshot: showing its customers
S2 Endorsement Letter Profusion Media Ltd (2021 Letter from CEO @ Profusion and Director of Science, Data & Innovation @ Profusion).
S3 Profusion customer case studies: https://profusion.com/case-studies/
S4 Data 50 Awards: 10 UK-based data entrepreneurs you should know about in 2017, Guy Marson (Co-Founder and Managing Director, Profusion) http://www.information-age.com/10-uk-based-data-entrepreneurs-know-2017-123466017/
S5 KTP11896: https://info.ktponline.org.uk/action/details/partnership.aspx?id=11896
S6 [text removed for publication]