**Project Title: KESS2 PhD Scholarship:** The application of machine learning to optimise propensity-scoring algorithms based on a novel understanding of how customer purchasing behaviour adapts to different product archetypes.

**Project Team:**
Dr Simon Thorne (Academic Lead Supervisor)
Dr Ana Calderon (Academic Supervisor)
Mr Paul Newbury (Lead Company Supervisor)

**Project Outline:**

The problem of market attribution is a non-trivial issue for marketers. Industry standard approaches to market attribution include the 'last touch' and 'single touch' models. These models take a simplistic view of market attribution giving the last click in a user's journey to a purchase the most credit and associated revenue for promotion. These approaches are widely regarded as sub optimal and unrepresentative of the relationship between marketing promotions, user journeys and sales generated. (Dalessandro et al. 2012, Shao and Li 2011, Shina et al 2014, Zhang et al 2014, Geyik et al, 2014, Berman 2018)

Solutions to these problems include enhanced models that examine the effect of multiple promotors on the sale of an item, these models are called 'multi touch data driven' models (Shao and Li 2011, Sina et al 2014, Zhang et al 2014). Indeed data driven models are emerging as the new best solution to the issue of market attribution that allocates credit for sales revenue more fairly (Shao and Li, 2014).

However, these multi touch models still have weaknesses inherit in the methods used to compute the propensity of customer sales. The propensity score is a numerical value that indicates how likely a customer is to purchase an item based on the 'journey' through the website over time. These weaknesses manifest as a 'one size fits all' approach to propensity scoring that assumes uniformity in customer decision-making behaviour regardless of the product type and nature. For instance, the approach taken by a customer to purchase some small consumable such as toothpaste is likely to be radically different to how a customer might behave when buying an expensive piece of furniture or a financial product. These subtleties in customer behaviour offer an opportunity for novel research optimising market attribution through tailoring propensity scoring algorithm to different models of customer behaviour based on product archetypes. A propensity-scoring algorithm that accounts for different customer behaviours would yield a more representative and accurate approach to market attribution that would afford the industrial partner Yard greater advantage and intelligence in attribution and targeting. Such a product would be novel to the market making Yard an unrivalled market leader.

This research therefore aims to develop a machine-learning model that optimises propensity-scoring algorithms based on a novel understanding of dynamic customer behaviour based on product archetypes.

**Application Deadline: 29th April 2019**

Proposed work as detailed below:

**Year 1 Preparation and groundwork phase**

Deliverable 1 literature review on Market Attribution methods, machine learning approaches, customer behaviour models and other relevant topics

Deliverable 2 Identify public and private datasets to be used to model the research

Deliverable 3 Identification of key stakeholders in the development and deployment of the application

Deliverable 4 Identification of key computing infrastructure needed for the development, experimentation and testing of the application

Deliverable 5 Identify candidate features from private and public data collected by Yard

Deliverable 6 Identify candidate methodology for experimental work and other research, including preliminary hypotheses and approaches to testing

**Year 2 Experimentation, testing and reflection phase**

Deliverable 1 A list of system and data requirements drawn from key stakeholders through interview, observation and inspection of processes

Deliverable 2 Collect all relevant data to populate the experimental data set to be used to test hypotheses experimentally

Deliverable 3 A narrative explaining the recursive experimental exploration of the data sets, conclusions drawn, actions taken and hypotheses proved or disproved.

Deliverable 4 A refined candidate data set used in experimentation

**Year 3 Refinement, optimisation and write up phase**

Deliverable 1 A finalised dataset that is cleaned, aggregated and optimised for experimentation

Deliverable 2 A finalised propensity scoring machine learning algorithm and associated measures of success, validation, and testing

Deliverable 3 A finalised infrastructure plan that optimises the maximum efficiency of computing infrastructure resources used for computation of the algorithm

Deliverable 4 Write up of thesis drawing on practical experimental work, data cleaning, aggregation and optimisation and literature

The scholarship offers all eligible participants (student):

- KESS PhD Scholarship holders do not pay fees;
- A monthly stipend in line with RCUK rates: 3 years of funding @ £14,483 per year for PhD
- An additional budget to support your research of £3,000 for travel, conference attendance, minor equipment, consumables and training
- An opportunity to access a suite of higher level skills training & attendance at a residential KESS Grad School
- A minimum of 30 days a year working with the company partner on their allocated project
- Students must complete & submit their thesis within 3.5 years for PhD
- All participants must complete the KESS Postgraduate Skills Development Award

To be eligible for the scholarship, applicants should have UK or EU Citizenship, and be resident in the competitive area of Wales on appointment (or be studying in or working in the competitive area immediately prior to starting the project) and should have the right to work in the region on qualification. Residence confirmed by providing documents such as driver's licence, tenancy agreement or recent statement from bank or building society.

The map for the eligible can be found here. NB: The map also illustrates other assisted areas in the UK, however only the East Wales (areas which are not shaded with any colour) is eligible in this project.

To obtain an application form (*Participant Application Form*) and for further information regarding KESS2 at Cardiff Metropolitan, Please click here

To apply, please submit completed application form to Simon Thorne **(email: SThorne@cardiffmet.ac.uk)**

Interviews for this post will take place in the week beginning the **29/04/2019**, with a proposed commencement date **by the end of May 2019.**

Supported by Knowledge Economy Skills Scholarships 2 (KESS2) which is an All Wales higher level skills initiative led by Bangor University on behalf of the HE Sector in Wales. It is part funded by the Welsh Government's European Social Fund (ESF) competitive area for East Wales.

**References**


- Berman, R., (2018), 'Beyond the Last Touch: Attribution in Online Advertising', *Marketing Science*, (35), 9, Internet, Online, https://pubsonline.informs.org/doi/10.1287/mksc.2018.1104, Available, Accessed 08/10/18
- Dalessandro, B., Perlich, P., Stitelman, O. and Provost, F., (2012), 'Causally Motivated Attribution for Online Advertising', *ADDKDD 2012 Proceedings of the sixth International Workshop on Data Mining for Online Advertising and Internet Economy*, 7, ACM 978-1-4503-1545-6
- Geyik, S., Saxena, A. and Dasdan, A., (2014), 'Multi-touch Attribution Budget Allocation in Online Advertising', ADDKDD 2014 Proceedings of the eighth International Workshop on Data Mining for Online Advertising, 2, 978-1-4503-2999-6
- Shao, X. and Li, L., 2011, 'Data-driven multi-touch attribution models'. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 258-264). ACM.
- Sinha, R., Saini, S. and Anadhavelu, N., 2014, October. Estimating the incremental effects of interactions for marketing attribution. In *Behavior, Economic and Social Computing (BESC), 2014 International Conference on* (pp. 1-6). IEEE.
- Zhang, Y., Wei, Y. and Ren, J., (2014), Multi Touch Attribution in Online Advertising with Survival Theory', ICDM Proceedings of the 2014 IEEE International Conference on Data Mining', pp 687-696, 978-1-4799-4302-9