**Institution:** University of Essex

**Unit of Assessment:** 10 – Mathematical Sciences

**Title of case study:** Using predictive modelling to improve public service delivery policy and practice of Essex and Suffolk County Councils

**Period when the underpinning research was undertaken:** 2011-2019

**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: |
|---|---|---|
| Aris Perperoglou | Reader in Statistics | Oct 2012 – March 2019 |
| Hongsheng Dai | Reader in Statistics | Jan 2013 – Present |
| Spyridon Vrontos | Senior Lecturer in Actuarial Science | Oct 2014 – Present |
| Berthold Lausen | Professor of Data Science | Jan 2009 – Present |

**Period when the claimed impact occurred:** 2018 - 2020

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

## 1. Summary of the impact

Research at Essex has shaped the policy and practice of Essex and Suffolk County Councils, improving public service delivery. Essex research into novel ensemble tree methods formed the basis of new machine learning platforms and decision support tools to reduce human-unconscious bias in decision making. Adopted by Suffolk County Council the tools support decision making processes and training of social care practitioners, enabling timely intervention and prevention measures for vulnerable children and adults, including children at risk of entering care. The tools developed from this Essex research have enabled the Councils to better understand and improve their use of data, develop staff expertise, and better target resources, improving public service efficiency in an environment of increasing financial pressure and demand on public services.

## 2. Underpinning research

During the last decade there has been a significant growth in the use of predictive statistical modelling in public sector organisations. The most significant challenges that local councils are facing are to identify particular groups which need more support from the government, or to classify people into different categories and provide more specific and appropriate support towards each group. Although there are many algorithms available for classification of such problems, most of them are either not fully data-driven approaches or lack interpretability. For example the specification of the logistic model may not be well suited for the large data sets involved in such projects and model misspecification is likely to happen as more data are collected, while the results of neural networks may be difficult to interpret.

In 2015 Essex Risk Stratification (ERS) researchers (based in the Department of Mathematical Sciences and led by Perperoglou, later Dai and Vrontos, supported by Lausen) were part of the Essex team which received GBP2,200,000 of Catalyst funding from HEFCE for a collaborative project with Essex and Suffolk County Councils (ECC and SCC respectively) [**G1**]. The project

aimed to use Essex research outputs and expertise to create and test novel data analysis methods and machine learning tools that would enable a more preventative approach to delivering public services and to develop a better understanding of the factors affecting outcomes for vulnerable people. To establish a purely data-driven approach, which can reduce human-unconscious bias in decision making and which could be easily adopted by SCC and ECC, the ERS team developed new predictive statistical methods drawing on their research [**R1, R2**]. [**R1**] proposes an ensemble of probability estimation trees, where a proportion of trees are selected on their individual predictive performance. Such an ensemble of trees has better performance than random forest and node harvest methods because only the best trees with highest predictive performances are added into the ensemble (fewer trees are needed without losing predictive accuracy). Extending this approach, [**R2**] evaluated the performance of the ensemble tree method via both unexplained variance and classification error ([**R1**] focused on classification error only). Although [**R2**] was published in 2020, the research in it started in late 2011. Since it uses similar ideas as that in [**R1**], the research method was used in the SCC collaboration from 2018 where ensemble of classification trees were needed in SCC projects. These new approaches improve the predictive ability of random forests and increase the interpretability of the machine learning tool by reduction of the size of the tree ensemble. The research selects the best trees in terms of individual accuracy and diversity from a large ensemble grown by random forest. By doing so, the team demonstrated that better prediction is achieved with even a reduced size of the ensemble. Thus, the computational burden is reduced, and the methodology can be used in the big data problems involved in the Catalyst project.

**Suffolk County Council (SCC)**

Between 2016 and 2019 the algorithms based on [**R1**, **R2**] were developed by the ERS team and embedded in SCC's first data-driven platform, where they reduced unconscious bias in decision-making, allowing human decisions to focus on classification of borderline cases while the remainder are determined automatically by the algorithm. The algorithms employed had a classification accuracy of over 75% in all applications, meaning only a small proportion of cases need to be considered manually. Furthermore, the algorithms provide timely responses and interventions based on the assessment of risks for individuals, by flagging at risk cases earlier than was previously possible. The team further developed the algorithm into user-friendly predictive dashboards that were piloted with social workers in 2019-2020.

**Essex County Council (ECC)**

The ERS team worked with ECC on a pilot project in 2016 on the topic of 'school readiness' examining which factors contribute to children being ready or not ready to start school. Algorithms based on [**R1**] were developed by the ERS the team to make predictions and create a profile comprising low, medium and high risk groups to inform practitioners and enable the successful targeting of resources. The ERS team successfully incorporated their algorithms into ECC's existing PX Risk software platform.

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

Following peer-review: [R1] was presented at the second European Conference in Data Analysis (ECDA2014) in Bremen, published by Springer in the Series 'Studies in Classification, Data Analysis, and Knowledge Organization' in 2016 and has been downloaded over 1,900 times; [R2] was published in a leading data science journal (IF1.6).

**R1** Khan Z., Gul A., Mahmoud O., Miftahuddin M. Perperoglou A. Adler W. and Lausen B. 2016. An ensemble of optimal trees for class membership probability estimation. In Analysis of Large and Complex Data (pp. 395-409). (Studies in Classification, Data Analysis, and Knowledge Organization). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-319-25226-1_34

**R2** Khan Z, Gul A., Perperoglou A., Miftahuddin M., Mahmoud O., Adler W., and 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

**G1** University of Essex, Essex and Suffolk CCs, Catalyst Project (HEFCE Catalyst Fund – universities as anchors initiative), Higher Education Funding Council for England (HEFCE), 2015 – 2019, GBP2,200,000.

## 4. Details of the impact

Suffolk and Essex County Councils serve a combined population of 2,500,000 people. Working in partnership with ECC and SCC, the ERS team drew on their research [**R1, R2**] to develop bespoke predictive data analysis tools enabling the Councils to adopt a more preventative approach for delivering public services, better target resources and manage data in an environment of increasing financial pressure. The research informed strategic decisions of SCC and ECC in data use, shaping the Councils' public service delivery and practice. ECC Leader comments '*The collaborative partnership between Essex County Council, the University of Essex […] has put us at the forefront of local authorities using the power of data science and artificial intelligence (AI) to tackle public policy challenges. We are investing in shared resources across our organisations to make Essex a place that is exemplar for integration of data across public bodies and to create the best data science and analytical capabilities in the UK to benefit our people and communities'* [**S1**].

### Changing SCC's data strategy, improving commissioning and public service efficiency

Prior to the Catalyst, SCC had not invested in tools applying machine learning processes to their data. Working with SCC the ERS team *'used their existing research and developed new methods, to produce algorithms for four core SCC support areas'* Assistant Director, Knowledge & Intelligence of SCC [**S2**]. The ERS team *'applied their work on machine learning approaches: Ensemble of Optimal Trees for probability estimation tree [**R1**] and Ensemble of Optimal Trees for classification tree [**R2**], to the identified datasets. They demonstrated both their ensemble of optimal trees and grouped lasso algorithms had an accuracy of over 70%'.This evidenced the algorithms' suitability to support SCC children's services staff in their decision making processes'* [**S2**].

SCC confirm that this work has '*taken SCC forward a huge leap in terms of tools to apply machine learning processes; it was the first time SCC had invested resources in such tools*.' adding that '*their work has also **led to more profound changes in SCC's use of data and supported embedding new approaches**'* [**S2**]. The team's **'***cutting edge statistical analysis methods resulted in more accurate and consistent data entry, as well as improved data processing across several areas in SCC'* adding that the research *'has given SCC new capability to understand key cohorts of service users, and their response to different interventions, from the higher quality datasets their work enables us to obtain. This was not something SCC had achieved before. It proved useful in the design and evaluation of interventions.'* [**S2**]. The ERS team's research enabled SCC to '*understand better what we can achieve with the social care and education data we hold, allowed*

*us to think strategically about our data including the implementation of consistent and accurate data collation processes'. 'We gained the confidence of SCC's senior leadership team to develop this work further because* ***the research provided the evidence they required for the necessary changes to our public service delivery policy and practice****. Taking this forward* ***led to establishment of an ethical framework supporting descriptive and predictive analysis work across SCC***. [**S2**]. The tools developed from Essex research enabled SCC to '*prevent and manage increasing pressures on our public services by identifying community members at highest risk of becoming vulnerable in order to provide timely responses and targeted interventions'* which ultimately '***enabled SCC to improve commissioning and public service efficiency'*** [**S2**].

**Supporting decision making processes of SCC's social care practitioners enabling timely intervention and prevention measures**

The ERS team produced four algorithms based on [**R1, R2**] to aid the decision-making of SCC's social care practitioners in four areas: School readiness, NEET (not in employment, education, training), MASH (Multi-agency Safeguarding Hub), CiC/CiN (children in care/children in need). SCC selected two algorithms for MASH and CiC/CiN, piloted in 2020 as training and decision support tools in practice. *'****The selected algorithms have been deployed to aid the decision-making processes of SCC's social care practitioners****, without overriding their professional judgement.* ***The developed models reduce unconscious bias in decision making, allow replicability of decisions, improve accuracy and enable timely intervention and prevention measures*** *by flagging at risk cases earlier than previously possible'* [**S2**].

**MASH,** is responsible for safeguarding vulnerable children and adults and triages notifications of need from agencies with safeguarding responsibilities. The ERS team's algorithms support MASH's current practices in assessing the level of risk for individuals to provide timely responses and interventions**,** decrease the number of false negatives and support MASH decision making, **helping more vulnerable people in need at a faster rate**: '*The MASH team interrogate, extract and use data related to vulnerable children in a more efficient way as a result of the Risk Stratification team's input'* [**S2**]. The MASH model is now used in training and decision making for new and current SCC staff [**S2**].

**CiC/CiN** The number of children entering care in England has increased in recent years but funding to tackle this has reduced. The ERS team's predictive model uses machine learning methods [based on **R1, R2**] to assign a risk level to individual cases supporting the decision making of case workers, helping them to prioritise workload. SCC confirms '***the CiC/CiN model has helped identify those factors that can, in combination, indicate a greater or lesser risk of entry into care***; *it quantifies the risk effects of circumstances such as family dis-function, neglect, age of the first entry into the system along with other contributing indicators of this outcome and provides a dynamic estimation for these risks based on new data collected by SCC.* This model '*has enabled us to look in far more depth at particular cases whose characteristics are suggested […] as potentially leading to an outcome of CiC*' which has '*speeded up the decision making process'* [**S2**]. This work aids the timely identification of children in need who are at high risk of going into care allowing preventative interventions to avoid the complexities and costs involved with entering care. According to SCC '*the spend on services for Children in Care in SCC alone is more than £40m per year with over 300 children per year entering care. It is a challenge to identify children that are on a potential trajectory to care.* ***Using the data to inform work that reduces 1% of children entering care would have a value of £400,000 per year. The human value of enabling a child to sustainably remain safely living with their birth family is priceless***' [**S2**].

SCC concludes that the ERC team's research has enabled them '***to explore and understand our data in new ways, and make interventions with some of the most vulnerable in our society***, *children and young people who may end up with an outcome of being in care; and contacts made in relation to safeguarding.' '****These initiatives could not have achieved the success they had without the tools developed from the predictive analytics research undertaken by the Risk Stratification Team****'* [**S2**].

**Integrating data analytics into ECC's decision making processes informing service delivery**

The ERS team applied advanced machine learning techniques based on [**R1**] to help ECC predict levels of school readiness in children to supplement existing expertise and facilitate targeted community interventions. The work supported the integration of machine learning techniques in the Council's decision making processes '***supporting the cultural evolution of ECC relating to the introduction and adoption of data analytics*** *e.g. predictive analytics, data visualization and machine learning'* Head of Corporate Strategy, ECC [**S4**]. ECC confirm that this work has '*enabled us to optimise our data science approaches'* [**S3**] and '***helped ECC to embed techniques and models in their processes so that analytics can inform their service delivery***.' [**S4**]. The ERS team's work has '*added value to the work being undertaken around the delivery of key Council strategies within the Essex Vision to support vulnerable people […] and in the democratization, integration and management of their data, to derive intelligence that has* ***informed the design and implementation of service provision***. *The team has played a key role in drawing on its academic assets to work alongside ECC to unlock barriers and tackle challenges aligned to ECC priorities'* [**S2**].

**Embedding expertise in the Councils through staff training and professional development**

SCC data analysts underwent '*an intensive training programme in the use and generation of the Risk Stratification academics' algorithms* [**S2**]*'*. Specifically, the ERS team worked closely with SCC's data analysts providing training in decision support tools for MASH and CiC/CiN '*embedding expertise in the techniques born out of their research'* [**S2**]. This was the first time that the council staff were trained and had used such tools. SCC adds *'It has allowed upskilling of our in-house data analysts [...] SCC's data analysts now apply the project's machine learning approaches in their own work with increased confidence in visualising, linking and analysing data; this reduced report preparation time and improves the quality of reporting to senior management and decision makers.'* [**S2**]. Working with staff at ECC, the team also provided *'opportunities for upskilling ECC staff via training materials and one to one interaction'* [**S4**].

**5. Sources to corroborate the impact**

**S1** Quote from Leader of Essex County Council
https://www.essex.ac.uk/research-projects/catalyst
**S2** Testimonial, Assistant Director, Knowledge & Intelligence, Suffolk County Council
**S3** Quote from Data Science Fellow, Essex County Council
https://www.essex.ac.uk/research-projects/catalyst/risk
**S4** Testimonial, Head of Corporate Strategy, Essex County Council