

Data Science for Public Sector

Slava Jankin Mikhaylov
University of Essex

Using Data Science in Policy

A report by the Behavioural Insights Team

- Children's social care
- Given the text of the initial referral and assessment, and structured data relating to the case, could we predict whether the case would be re-referred and escalated if it were closed?
- Machine learning and Natural Language Processing
- <http://bit.ly/2FIL0H3>

Figure 5: The inputs of the machine learning algorithm used to detect escalated closed cases.

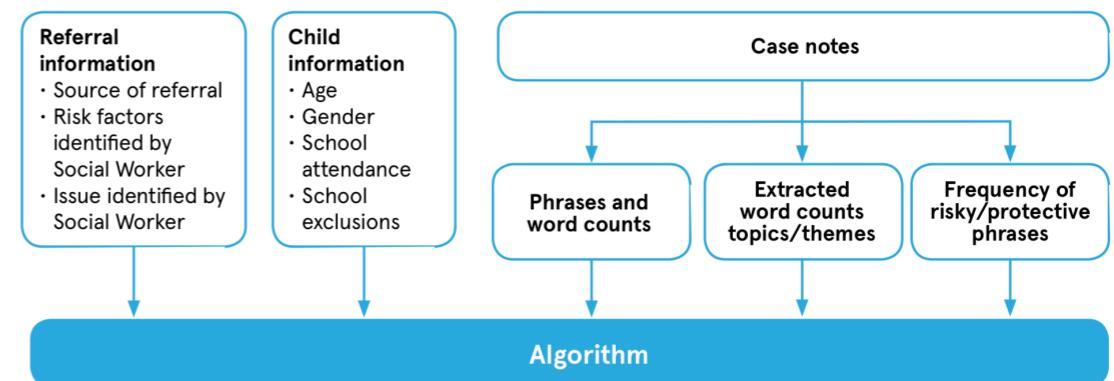
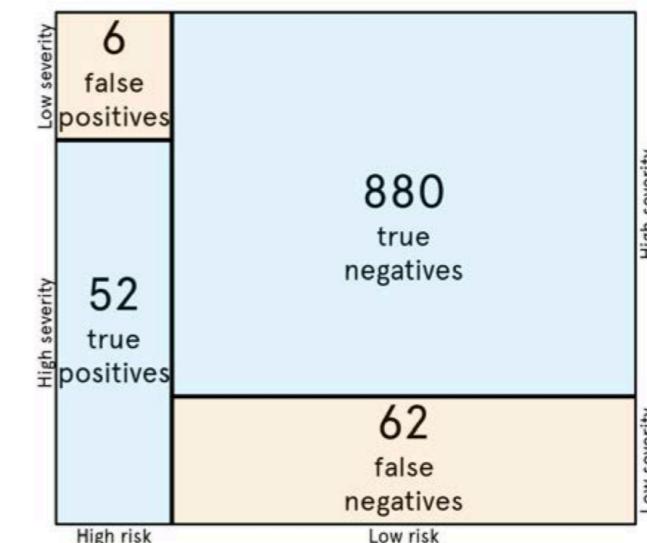


Figure 6: Expected distribution of true and false positives and negatives for 1,000 previously unclassified cases.



'Risk' refers to the model's prediction and 'severity' to the actual outcome. For example, a 'false negative' is a case that escalated but that the model categorised as 'low risk'.

Data Warehouse to Predictive Analytics

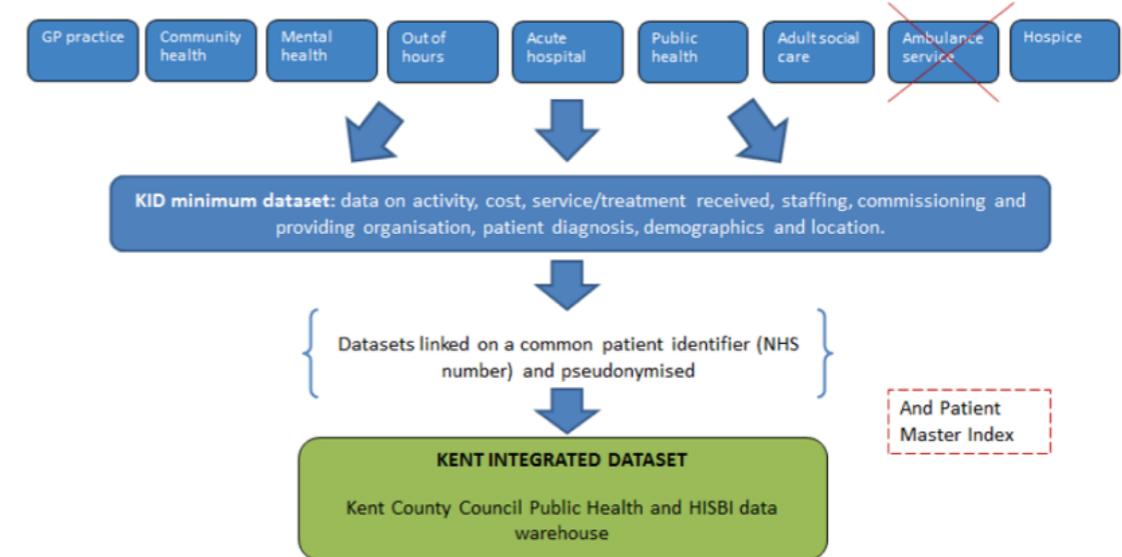
- Successfully integrated 20 core internal datasets to create a data warehouse that feeds predictive analytics projects
- London Borough Newham were behind the nations first borough-wide landlord licensing scheme – the warehouse was used to find unlicensed privately rented properties to target likely offenders.
- Other applications include communication campaigns for eligible benefit recipients, and targeted loft insulation to reduce fuel poverty



Health and Social Care Integrated Datasets

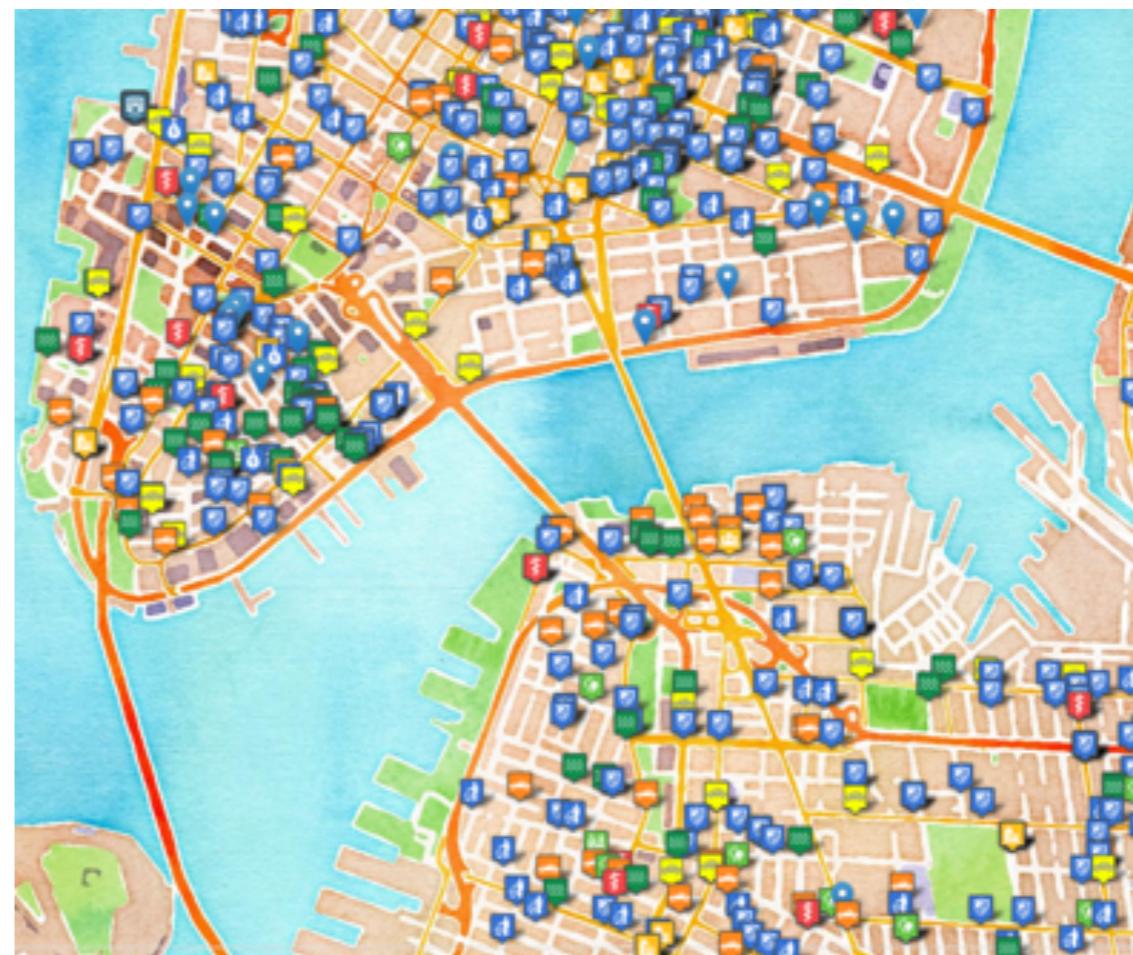
- The purpose of the KID is to provide planners in Kent & Medway with insight into population health and system-level use of services.
- Data from health and social care providers is brought together to allow analysis of the patient journey.
- There have been several preventative uses of the KID including: a localised frailty index; estimating prevalence of rare conditions; comparing risk of non-elective hospitalisation; and many more.

Flow of data into the Kent Integrated Dataset



Building an Open Data Culture: ‘Open Data Values’

- New York’s open data journey began in 2012 with the passing of an open data law. By 2015 1,300 datasets had been fed into their open data platform.
- The open data platform doesn’t just publish datasets, it includes guides for data novices to interpret and analyse.
- The initiative puts New Yorkers as the first consideration for every decision made by the open data team – and abides to six ‘Open Data Values’



Collaborative risk modelling to target individuals

- London Borough of Hackney & Thurrock Council separately worked with tech company Xantura to develop an 'Early Help Predictive System' that used data from multiple partners.
- A risk algorithm gives the respective authorities a list of vulnerable families which is then scrutinised by officers to determine who would benefit most from additional support from their services.
- Brent have worked with IBM to develop an award winning risk model to start identifying children at risk of criminal exploitation.

<https://www.theguardian.com/society/2018/sep/16/councils-use-377000-peoples-data-in-efforts-to-predict-child-abuse>

<https://www.theguardian.com/society/2018/sep/17/data-on-thousands-of-children-used-to-predict-risk-of-gang-exploitation>



Data Science in Criminology

- Historically analytics in the crime sector has been ahead of other public sector counterparts in their innovative use of data science.
- Durham police force have worked with University of Cambridge to create a model that can guide custody decisions.
- Manchester Police used analytics to understand emerging hotspots of robbery, burglary, & theft – leading to a reduction across all 3.
- Kent Police have been using a geo-spatial model to estimate the likelihood of crimes occurring in locations during specific time periods

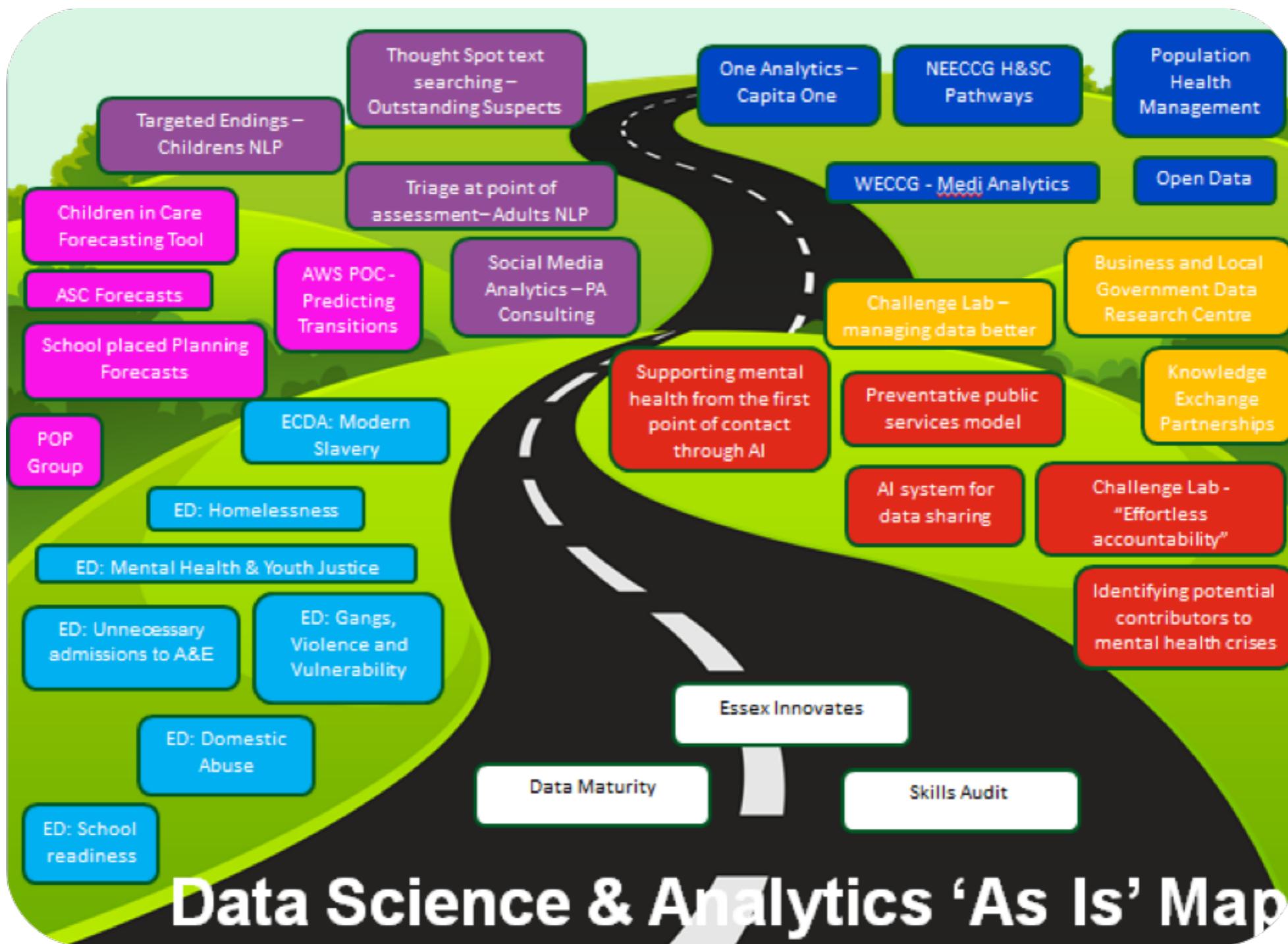


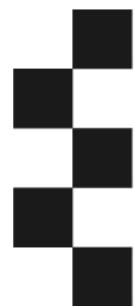
EssexData: A risk profile for school readiness

- Can we predict whether children will be school ready on starting school?
- What factors are most likely to cause this early indicator of “best start in life”?
- 511 households at risk in the pilot area;
- Including 280 previously unknown to public services;
- 74% prediction accuracy rate.



Essex Innovates: Ecosystem





University of Essex

Institute for Analytics and Data Science



ESRC Business and Local Government
Data Research Centre



Administrative Data
Research Network



System approach



Quadruple Helix Innovation

Government, Academia, Industry and Citizens collaborating together to drive structural changes far beyond the scope of any one organization could achieve on its own



“Research in Big Data should be grounded in the quadruple helix model where civil society joins with business, academia, and government sectors to drive changes far beyond the scope of what any organization can do on their own.”

Intel Corp policy position paper on Big Data

Benefits for academia



“In ML, where algorithms get published quickly and state-of-the-art frameworks are open-source, there isn't any first-mover advantage. Rather, competitive edge comes from data accumulation and infrastructure know-how. Which tends to benefit established large companies, rather than nimble upstarts with better tech.”

François Chollet, Deep learning at Google, Author of Keras, @fchollet

Collaboration

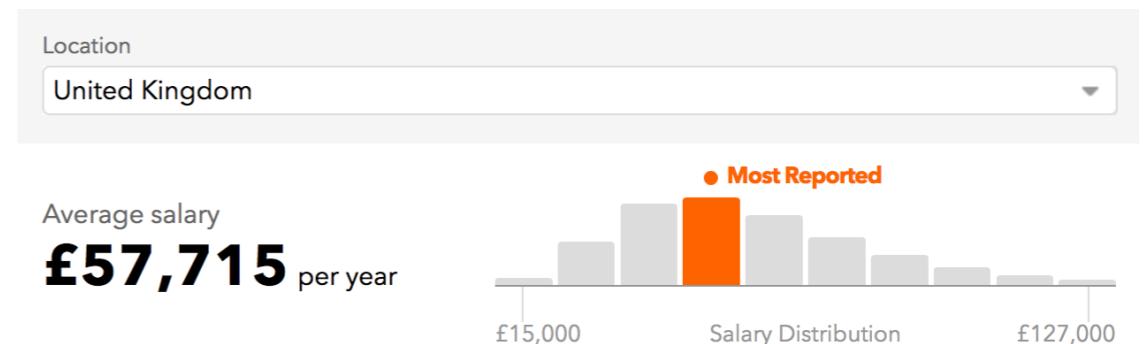
- Cost-Benefit rather than technical issues
- Unclear benefits of sharing data: vague and conceptual rather than tangible and related to business outcomes
- Cost of sharing (perceived privacy, security risks, resource costs) outweigh unclear benefits.



- Knowledge Transfer Partnerships
- Embedding
- Secondments
- Joint appointments

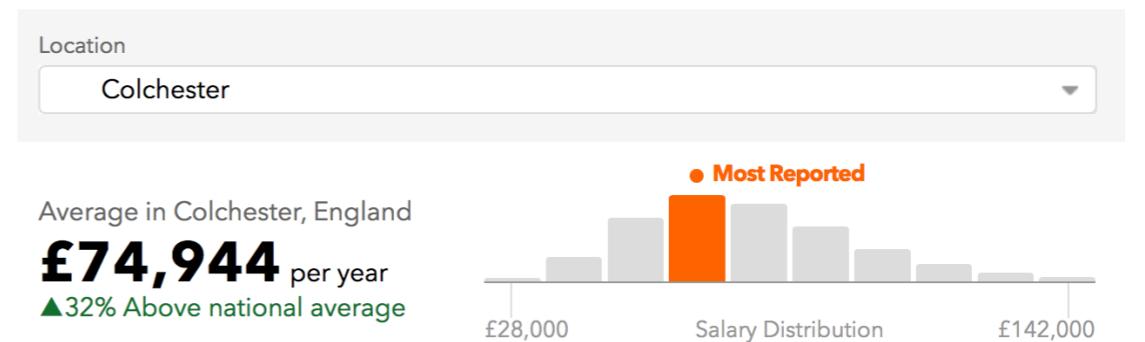
Data Scientist Salaries in the United Kingdom

Salary estimated from 14,248 employees, users, and past and present job advertisements on Indeed in the past 36 months. Last updated: 21 June 2018



Data Scientist Salaries in Colchester, England

Salary estimated from 24 employees, users, and past and present job advertisements on Indeed in the past 36 months. Last updated: 16 March 2018



Thank you!

s.mikhaylov@essex.ac.uk