Data driven approaches to customer service: What we need to consider
November 2019
Australia 2050

Population: XXm?
Average Age: increase
Retirees to Working age: increase
Urbanisation: increase
Unemployment: mixed
# Megacities: increase
Dominant Industries: Services
Climate: Warmer and Changed

NSW Data Analytics Centre
.. artificial intelligence will become a core technology across many different industries and one of the driving forces of the coming fourth industrial revolution, the standardization community will play a critical role in shaping its future.
Procurement Spend Categorisation

- NSW spends approximately $30bn per annum through procurement.
- Categorisation is currently performed using a rules based system “Spend Cube” with more than 1million rules.
- DAC Natural Language Processing Machine Learning tool was trained and then categorised 10,000,000 transactions from the test set into 274 categories.
- Overall accuracy of >96% and runs in hours (c.f. weeks for spend cube).
- Experiments were run without the “Other” category forcing categorisation of almost 300,000 items.
- Microsoft announced as commercialisation partner for tool.
Predicting school age enrolment

Data Analytics: Deep Learning Based Prediction

Role of the DAC
Multi-agency Data integration
Spatial and temporal visualisations
Deep Learning model evaluations

NSW Data Analytics Centre
Supporting OOHC Reform

- Create “pathways” of all children in OOHC and identify cohorts of particular interest
- Represent OOHC as a sequence of placement events
- Each sequence has a final placement or exit
- At each placement the child accumulates service history
- Millions of records
- Health, FACS, Justice, Education, Industry, Transport

NSW Data Analytics Centre
Data Sharing Strategy

… supporting automated data sharing across government.

Role of the DAC

Working with specialist organizations (including ACS, CSIRO), state and federal governments, industry and privacy advocates to develop frameworks, trial technical solutions, inform policy and help scale automated data sharing.
Data Analytics Strategy for Outcomes Based Budgeting

… Supporting quantification of Outcomes Frameworks

Role of the DAC

Build or enrich the datasets to support quantification of Outcomes Frameworks.
Help agencies identify cohorts and analyse current outcomes.

NSW Data Analytics Centre
Data Analytics Strategy for Outcomes Based Budgeting

... Supporting mapping in data of evidence pathways

Outcomes Framework

Role of the DAC
Build the dataset to support development or enrichment of evidence pathways such as the work done for out-of-home care (OOHC) services.

Help agencies identify cohorts and analyse pathways.
Data Analytics Strategy for Outcomes

services mapped to place, contributing to outcomes frameworks, analysed against evidence pathways to inform policy

Outcomes Framework

NSW Data Analytics Centre
Safe Data: Personal Information Factor

Non personal data

Lowest PIF

Lightly aggregated data

Highly aggregated data

Personal data including Health data

Personal data

N
N
N
N
N
N
N
1

NSW Data Analytics Centre
Context: Different environments for use of data
Privacy Preserving Data Sharing

Quantified “Fives Safes”

Will the results lead to disclosure?

Is there a disclosure risk in the data itself?

Has appropriate and sufficient protection been applied to the data?

Is this use of the data appropriate?

Is the user authorised to access and use the data?
| Safe Organisation | • refers to the systems and processes employed by an organisation to ensure the Safes Framework is applied throughout the project and with the long-term management of data and outputs. Safe organisations may include those which adhere to data protection, quality standards and cyber security standards. Safe Organisation may consider |
| Safe Lifecycle | • refers to the time sensitivity of a dataset or output. Data may be highly sensitive for a specific period and then may be not sensitive at all. For example, a city plan that might involve the mandated acquisition of an individual’s home to enable the construction of a new road may be very sensitive until the home is demolished. At this time there is no remaining value in protecting the data or output. Considering the complete lifecycle of a dataset may add additional insight and tools to help effectively anonymise and protect privacy rights. Safe Lifecycle may consider |
| Safe Outcomes | • refers to the ultimate uses of the project outputs. A variety of “Outcomes Frameworks” have been developed which can be informed by the outputs of individual data linkage and analysis projects. Safe Outcomes may consider |
| Safe Use | • refers to the use of the outputs within the Outcomes framework specifically, how much interpretation or context is required to appropriately use the outputs, including the degree to which a decision or action can be informed, or automated based on this output. Safe Use may consider |
| Safe Response | • refers to the systems and process which need to be in place to address adverse consequences of sharing of data or sharing or actions taken based on outputs. Safe Response may consider |

NSW Data Analytics Centre
Focus Area: Data Sharing and Use Frameworks

Data Holder

- Share Data
- Access Data
- Access Aggregated, Obfuscated or Perturbed Data
- Query Lightly Aggregated, Obfuscated or Perturbed Data
- Query Highly Aggregated, Obfuscated or Perturbed Data
- Aware of Data Scope and Data Dictionary
- Aware of Data
- No Awareness of Data

Level of Access

Decreasing Restrictions

Data User

- Share A copy of Data
- Access Data
- Access Derived Data Product
- Aware of Data Scope and Data Dictionary
- Aware of Data
- No Awareness of Data
Data driven sensitivities:
- Vulnerable groups (e.g. at risk populations)
- Sensitive subjects (e.g. health data)
- Linked data (or linkage is occurring as part of the project)
- Information relating to ethnic or racial origin, political opinions, religious or philosophical beliefs, trade union membership, health or sexual life.

Output driven sensitivities:
- The audience and the use cases for the outputs
- Outputs require high levels of context so as not to be misinterpreted
- Outputs to be used in decision making when the quality of the data/analytics may not support this
- Potential to have unintended consequences (e.g. contradicts government policy, social implications, causes harm to individuals)
- Impacts for individuals, groups or organisations (consider who is impacted and to what extent?)
- Produce unexpected results/insights (e.g. The findings differ from project expectations or other research in the area)
Focus: Safe Use of Outputs (Consequences)

- **Domain Expert Use of Outputs**
  - Post Project Review, Output Informs New Projects
  - Domain-Expert Output Review, Output Informs New Projects
  - Domain-Expert Output Review, Output Shapes New Projects

- **Autonomous Use of Outputs**
  - Automated Use of Output, Automated Checks on System Integrity
  - Exception Handling, Limit Continued Use of Outputs, Outputs Informs Process Modification and Limits
  - Fault Handling, Cease Autonomous Use of Outputs, Outputs Used to Redesign Processes and Limits

**NSW Data Analytics Centre**

**Consequences**

- **Explainable Decision**
  - Reversible Impact
  - Reversible with residual impact
  - Irreversible Impact

- **Low Context Need**
  - Very Low Value
  - 0.01
  - 0.05
  - 0.1
  - 0.5
  - 1.0

- **High Context Need**
  - “Unexpectedness” (Information Content)
What happens when People are Algorithms?

- Will the results lead to disclosure?
- Is the Personal Information Factor less than 1.0?
- Is this use of the data appropriate?
- Is the algorithm transparent and scrutinisable in an ongoing manner?
- Has appropriate and sufficient protection been applied to the data?
Standards Needed

Dr. Ian Oppermann
CEO and Chief Data Scientist, NSW Data Analytics Centre

Data Analytics Centre | Department of Customer Service
e ian.oppermann@customerservice.nsw.gov.au | www.dac.nsw.gov.au
2 – 24 Rawson Place, Sydney NSW 2000