DATA SCIENCE

Methods Analytics
AN ALTEN COMPANY

Gaining Greater Insights from Public Consultations with Data Science & NLP

CONTENTS

GAINING GREATER INSIGHTS FROM PUBLIC CONSULTATIONS WITH DATA SCIENCE & NLP

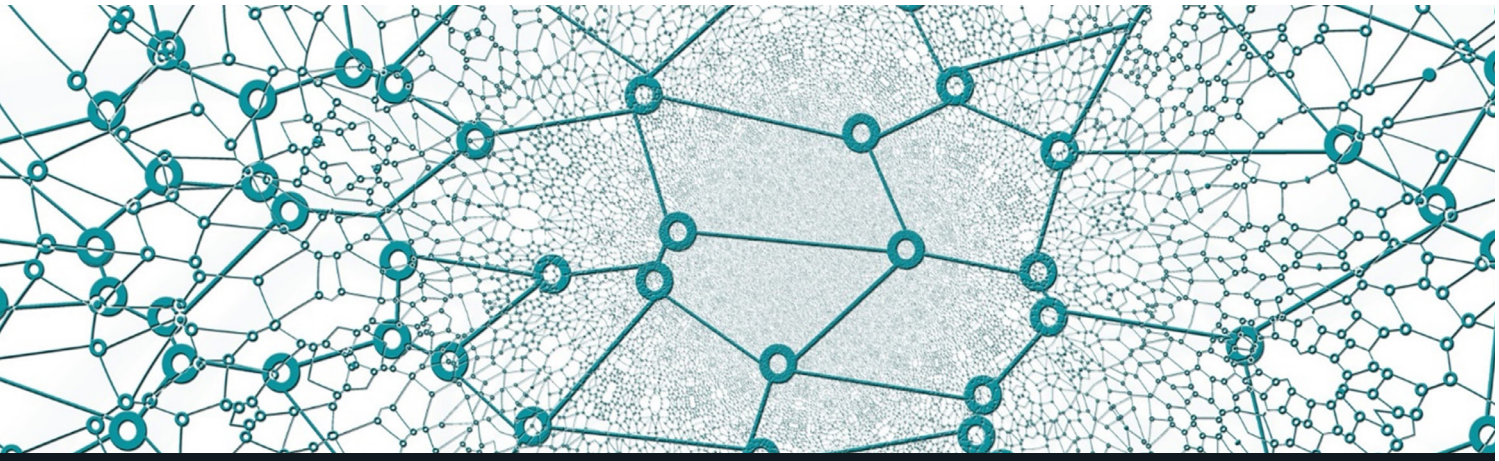
INTRODUCTION

Ongoing dialogue between the government and the public is a vital part of the policy-making process. Sometimes this takes place via roundtables with expert stakeholders, industry leads and lobby groups. Sometimes expert reviews by individuals or organisations can help shape effective policies. And sometimes the government needs to run a formal, open consultation; this provides transparency of emerging policy-thinking, and enables responses from anybody interested in that area.

The scrutiny and feedback provided by an open consultation is important for the healthy democratic process. Ensuring all of this feedback is appropriately reviewed and considered also places an enormous burden on government departments. They will try to steer responses through web forms for ease of analysis, but for accessibility reasons must accept them via any channel (answers on a postcard please...!). The analysis will be largely qualitative in nature, trawling through responses to gather evidence and insight, both to understand the public consensus, and to help build better policy.

> The "**Health and Harmony**" consultation run by Defra in 2018 received a whopping 43,356 responses

There are currently 77 open consultations, with 282 having closed within the last 12 months. The recent Data Reform consultation received 2,924 responses (including ours); the "Health and Harmony" consultation run by Defra in 2018 received a whopping 43,356 responses. Sometimes the volume of responses can be largely predicted and planned for. Sometimes this is harder – when they happen at relatively short notice, or when they're carried out by departments who are also stretched across fast-changing policy or operational area. Teams will have to be stood up quickly to trawl through these thousands of responses, pulling together statistics and pulling out the key content to try to summarise and learn from. Hugely manual, and hugely time-consuming.



This is where data science – and in particular, Natural Language Processing (NLP) – can be invaluable. Methods Analytics' data scientists have been looking at how this could be used to supplement the manual processing, drawing out insights that might otherwise have been missed, and reducing some of the manual processing that doesn't add any value. This is exciting stuff: and done well, could reduce the cost of consultations, while improving the quality of the outcomes.

GAINING GREATER INSIGHTS FROM PUBLIC CONSULTATIONS WITH DATA SCIENCE & NLP

DATA SCIENCE FOR PUBLIC CONSULTATIONS

Analysing responses to public consultations can be challenging. There is some progress in automating the summarisation of quantitative answers like rating and multiple-choice questions, but a lot of effort is still required to outline free text responses to open-ended questions. Many individual responses are typically read by analysts whose findings are then collected and presented in a report; this is not only time consuming but is often exposed to analysts' subjectivity.

Machine learning, AI, and advanced analytical techniques can be used to solve complex problems and unveil hidden information in a faster, less biased way. These approaches build on traditional mathematical and statistical techniques to learn from the data.

One group of powerful techniques withing data science that can be applied to analyse free text responses in public consultations is Natural Language Processing (NLP). We can leverage these techniques to automate analysis and gain insights into public consultation responses.

NATURAL LANGUAGE PROCESSING

NLP is a branch of data science that enables automated processes to analyse and extract meaningful insights from human language. Methods Analytics has extensive experience in applying NLP to solve problems across the public sector.

Methods Analytics have used its expertise in this field to develop an extensive NLP toolset that can be easily integrated into our clients' environments. This toolset gives expert data scientists the ability to adapt NLP algorithms and gain insight from unstructured text far quicker than conventional approaches. The analysis can then be visualised and presented to produce robust analytics from this unstructured text.

GAINING GREATER INSIGHTS FROM PUBLIC CONSULTATIONS WITH DATA SCIENCE & NLP

NLP FOR PUBLIC CONSULTATIONS

NLP can be used to automate much of the response analysis process and enable new and novel insights which can be presented in intuitive and easy to understand visualisations. It is also possible to ensure that all responses are included in the analysis, reducing bias and making everyone's voice heard. Here are just some of the NLP techniques that can be applied in this context:

Named entity recognition
identifies, extracts and links real-world entities of interest, such as organisations, locations, people & dates. This can provide critical insight into specific entities the public mention in their responses.

Extraction of keywords, phrases & frequently used words that are common across responses. This helps quickly identify any shared key thoughts and feelings in the public responses.

Topic modelling
enables the identification and extraction of core themes, topics and subjects from responses.

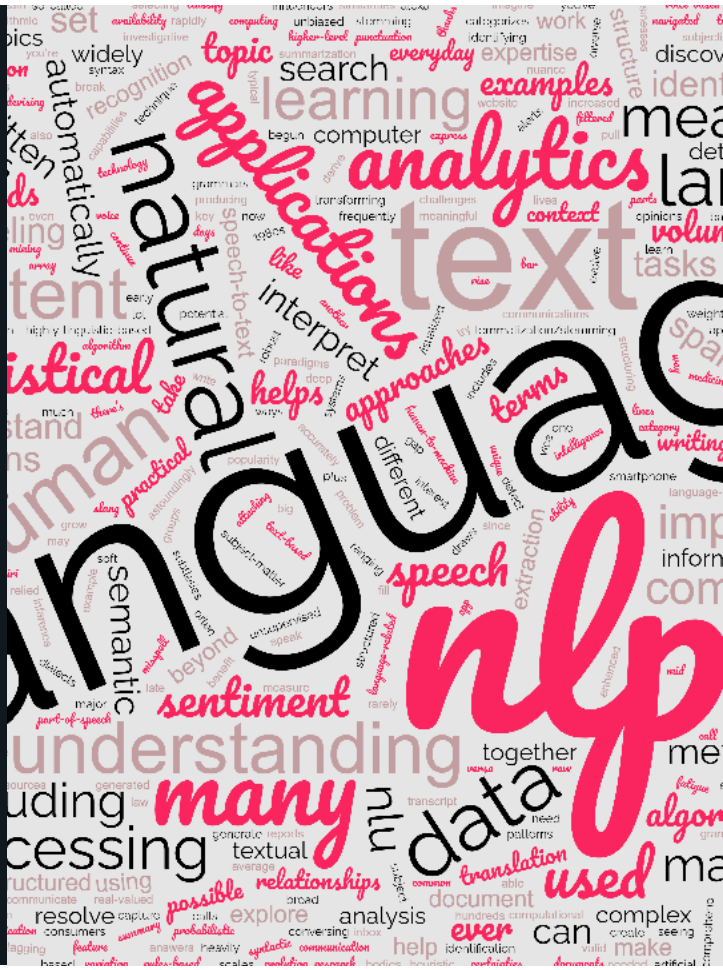
Sentiment analysis
can be used to see if the text of a survey response is positive, negative, or neutral. These can be aggregated across respondents to automatically extract respondents' perception about the policy under consultation.

Named entity linking
allows us to link these entities to a knowledge graph to acquire additional information such as definitions, aliases and conceptual categories. This allows crucial insight beyond just extracting individual entities, as it gives them context by creating connections and associations. This network of connections and associations can then be visualised in an intuitive, dynamic and easy to understand way.

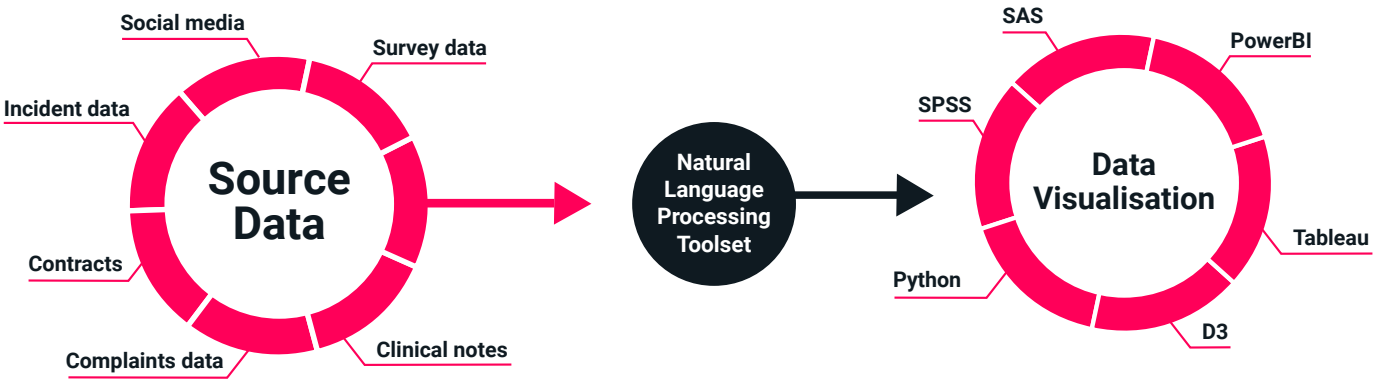
OTHER APPLICATIONS

Aside from NLP there are other areas of data science and analytics expertise can be used to advantageous effect. We've noticed that many public consultation reports mention there are regional differences in survey responses, but these differences are often not quantified and analysed rigorously, leaving decision makers in the dark. Along with geographic data, demographic data is also often collected but not utilised to see how different groups' opinions differ on certain questions: these differences could be valuable for policy makers but are often not examined in detail.

In current public consultation analysis, it is often difficult, and requires painstaking work, to link closed answers, such as multiple choice, with open free-text answers and responses. A particularly exciting outcome from using data science techniques is the ability to combine the more traditional statistical and data analysis approach with NLP insights, approaching this problem in a novel and rigorous way. This has the great advantage of reducing the resources needed, reducing the time spent on analysis and bringing new and previously undiscovered insights and connections between different types of responses.



GAINING GREATER INSIGHTS FROM PUBLIC CONSULTATIONS WITH DATA SCIENCE & NLP



BENEFITS

There are several tangible benefits from approaching public consultation analysis with data science and NLP. Traditionally the process for analysing open questions is difficult, complex and requires many person hours to unearth valuable information. With this novel approach we can dramatically reduce the investment needed to reach these answers, freeing up staff to carry out higher value analysis. This of course, also means that the process can be done in considerably less time. As we've already highlighted sometimes consultations can happen at short notice - using these techniques we would be able to get comparable results between these and consultations that departments were better prepared for.

The second advantage is coverage. With these methods we can ensure that every open question and any piece of free text that is sent by a member of the public can be considered, and their thoughts and feelings incorporated into the analysis. Whilst AI is not free of bias, in this context and when used with care, it can be utilised to gather information in an objective way. Traditionally, different analysts will look at different answers, with each bringing their own experiences and perspectives to their interpretation, using these techniques we can ensure that each answer is being viewed in the same light.

The final, and perhaps most exciting, group of benefits is the ability to generate new insights. We can bring out common themes that might have otherwise been missed, identify keywords, organisations, and people the public are citing, gauge the public sentiment in responses to open questions and see how these responses change between different demographic, economic and geographic groups. We can also start to connect insights between the more structured closed multiple choice questions and open questions and responses sent directly by members of the public, again helping to ensure more voices are heard generating a fairer and more in-depth results than has previously been possible using traditional technologies and techniques.

GAINING GREATER INSIGHTS FROM PUBLIC CONSULTATIONS WITH DATA SCIENCE & NLP

ETHICAL CONSIDERATIONS

Often organisations are sceptical about the use of AI to process sensitive documents, mainly concerned about ethical issues. Methods Analytics recommend using AI in this context to augment, rather than replace, the human review of responses, allowing for greater scrutiny of feedback, rather than less. AI should be used as a tool to decrease bias in analysing responses, highlighting the main themes and the links they have with the responders.

EXPERIMENTAL APPROACH

A possible way to increase the trust in AI and tune the models according to specific needs, is to use an experimental approach. A fraction of the consultation responses can be analysed using NLP and validated by humans independently. At this point the NLP techniques can be tailored to optimise time-consuming tasks, feeding summary dashboards that can be verified by subject matter experts. This iterative process should increase the trust in AI, allowing analysts to use these techniques as a tool, speeding up processes and saving time and money

SUMMARY AND CONCLUSIONS

Open public consultations are an important part of a healthy democratic process. Ensuring feedback is reviewed and analysed as fairly as possible places an enormous burden on government resources, especially when demands on a department may already be high, or consultations happen in a short time frame. We believe that novel data science & NLP techniques present an excellent opportunity to help relieve some of these burdens.

These techniques can be used to extract greater insights at a faster pace, whilst also increasing the coverage by using NLP to tap into open responses and reducing bias by looking at all responses equally. We can also make new and interesting connections between the open and closed questions, like multiple choice answers, and dive into the different perspectives from different geographical, demographic, and economic groups. This combined approach will give analysts and policy makers a much more comprehensive view of the public's attitude, emotions, and perspectives on the issues in any given public consultation.

GAINING GREATER INSIGHTS FROM PUBLIC CONSULTATIONS WITH DATA SCIENCE & NLP

Methods Analytics
AN ALTEN COMPANY

Contributors

Andrew Knott
Senior Data Scientists
andrew.knott@methods.co.uk

Felicia Ziparo
Lead Data Scientist
felicia.ziparo@methods.co.uk

Martine Clarke
Head of Government Sector
martine.clark@methods.co.uk

Registered in England and Wales Company No. 2485577
VAT No. GB 859 1403 14