MS984: Data Analytics in Practice

Case Study 5: Learning Enhancement Using AI (Safeguarding Children)

Group 4

Muneeba Nawaz | Paul Slevin | Ademola Saheed Oladimeji |

Olumuyiwa Olajuwon | Michael Patric | Madhav Thodindala |

Susan Dangana |

1.0   INTRODUCTION

The protection of children and the prevention of domestic abuse are critical concerns that demand thorough analysis and understanding of relevant data. In this report, we embark on an exploratory journey into a dataset extracted from a JSON file, focusing on a section titled "Safeguarding Children Level 2." Through the lens of natural language processing (NLP) techniques, including text summarization, named entity recognition (NER), and word frequency analysis, we aim to uncover meaningful insights that can inform stakeholders and practitioners in the field of child protection and domestic abuse prevention.

2.0   AIM AND OBJECTIVES

The aim of this analysis is to extract valuable insights from the provided dataset using NLP techniques, with a specific focus on understanding safeguarding protocols for children and addressing issues related to domestic abuse. By leveraging advanced text processing methodologies, our goal is to distil complex information into actionable insights that can guide decision-making processes and interventions aimed at protecting vulnerable populations.

The objectives of the analysis are as follows:

- To conduct a comprehensive analysis of the dataset extracted from the JSON file.
- To utilize NLP techniques, including text summarization, NER, and word frequency analysis, to extract meaningful insights.
- To identify and summarize key themes and protocols related to safeguarding children at Level 2.
- To recognize and categorize named entities such as organizations, dates, and cardinal numbers within the text.
- To analyse word frequencies to uncover common themes and topics prevalent in the dataset.
- To derive insights that can inform stakeholders and practitioners involved in child protection and domestic abuse prevention initiatives.

3.0   METHODOLOGY

Data Collection:

The sample data utilized in this study was obtained from a JSON file comprising multiple pages of text. Specific pages pertaining to the section titled "Safeguarding Children Level 2" were selected for detailed analysis.

Text Pre-processing:

Prior to analysis, the collected text underwent pre-processing procedures aimed at enhancing its suitability for computational analysis. This involved the removal of stop words and punctuation, thereby ensuring that only semantically meaningful words remained for subsequent analysis. The pre-processed text served as the basis for all further analyses.

Text Summarization:

Extractive text summarization techniques were employed to distil the essence of the text into a concise summary. This summarization process involved assigning importance scores to individual sentences, which were calculated based on word frequency within each sentence. Consequently, sentences deemed most important were retained to encapsulate the key points of the text.

Named Entity Recognition (NER):

Named Entity Recognition (NER) was applied to identify and categorize named entities present within the text. Entities such as organizations, dates, and cardinal numbers were systematically identified and classified. This facilitated a deeper understanding of the textual content by elucidating the significance of key entities mentioned therein.

Word Frequency Analysis:

A bag-of-words approach was employed to analyse the frequency of occurrence of individual words within the pre-processed text. By quantifying word frequencies, common themes and topics prevalent within the text were discerned. This analysis provided valuable insights into the underlying themes and subjects addressed in the text.

4.0   RESULTS:

Text Summarization Output:

The generated summary provided a condensed version of the original text, highlighting key points such as the impact of certain actions and the services available for support. However, due to the absence of punctuation, the summary may appear less coherent and may require further post-processing for readability. Below is the summary gotten for number of sentences set to 3

'Summary:
Which services may offer support to the family if level two (Additional); they will provide con tact details Identify that Family Solutions criteria is met and discuss how to refer to the Fami ly solutions team if level three (Intensive) or Transfer you to the Initial response team if it is considered that the child is at risk of significant harm (Specialist) Please refer to the Essex Wi ndscreen over the page for explanation of levels and follow the yellow pathway if the family require additional support. If you work with the parents and carers only This list identifies ad ult issues which are known to have an impact on their capacity to parent effectively: Mental Health or Learning disability Alcohol misuse Substance misuse Domestic abuse Family history Financial Problems Chronic Ill health Complex social problems Social isolation Early parentho od Should you concerned about the adults presentation to the hospital, consider: . The impa ct includes: Children may get caught in the 'cross fire' protecting their mother and sustain p hysical injuries Girls often seek to protect siblings during episodes of violence Depression, a nxiety, guilt, fear, insecurity, sleeping and eating disorders, nightmares, shame, self- harming Difficulties with concentration, heightened aggression, violence as 'normal', early sexual and /or drug activity, disobedience/anti-social

behaviour (boys) Examples of specialist services in clude statutory children's social care interventions (child in need, child protection or Look Af ter Children Service), statutory youth offending service work, services provided for children a nd young people as a result of statements of special education need and services provided fo r children and young people with complex mental health needs and/or substance misuse de pendency. Degree of threat and coercion It must be acknowledged that Child protection wor k "involves uncertainty we cannot know for sure what is going on in the privacy of family life nor can we predict with certainty what will happen" Munro, 2012 Principles of the Children A ct 1989 and 2004 Children Act 1989 The Children Act 1989 is the main piece of legislation tha t underpins our child protection work.
'

Overall, the summarization process effectively condenses the original text into key points related to service provision, referral pathways, identification of issues impacting families, understanding the impact on children, availability of specialist services, acknowledgment of uncertainties, and adherence to legal frameworks. This condensed information can be valuable for professionals involved in child protection and welfare services, facilitating better understanding and decision-making in safeguarding children.

Named Entity Recognition Output:

The NER output revealed several named entities along with their types, including dates, organizations, and cardinal numbers. This information shed light on important entities referenced in the text, such as legal acts, specific organizations, and significant dates.

Table 1 in the appendix summarizes the named entities and their corresponding types extracted from the pre-processed text.

Organizational Entities: The analysis identified various organizational entities such as "Mid Essex Hospital Services," "NHS Trust," and "Training and Development," highlighting their presence and significance in the document. These organizations likely play crucial roles in the context of the text, possibly related to healthcare, training, or safeguarding initiatives.

Legal Entities and Acts: Named entities related to legal frameworks and acts, including "Children Act 2004," "Data Protection Act," and "Section 11," suggest the document's focus on legal compliance and regulatory requirements, particularly concerning child protection and data privacy.

Temporal References: The presence of temporal references like "March 2014" and "October 2014" indicates specific time periods or events mentioned in the text, providing context and chronological information relevant to the document's content.

Individuals and Professionals: Named entities such as "Victoria Climbie" and "Manas Datta" represent individuals mentioned in the text, potentially referring to notable figures or professionals associated with child welfare, healthcare, or legal domains.

Common Themes Identified:

Through the analysis, we identified common themes mentioned in the text, including references to legal acts, dates, and organizational entities. These recurring themes provide insights into the central topics discussed in the text and help in understanding its primary focus. Table 2 shows a few of the common themes in the text alongside their number of mentions.

VISUAL OVERVIEW OF THE ANALYSIS

The graphical representation in figure 1 displays the most prominent words from the analysed text. The larger the word's font size, the more frequently it appears in the text. This visual succinctly encapsulates the primary focus areas of the report, such as child welfare, legal compliance, and support services, with the most critical topics likely being the most prominent in the cloud.

Figure 2 shows a distribution plot that indicates how the importance scores of sentences or phrases are spread across the document. High-frequency scores correlate with sentences that carry significant weight in the context of safeguarding children, emphasizing the main messages or calls to action within the text.

The heat map above displays the frequency of the top 20 words in a matrix format, using colour intensity to represent the frequency of each word. This visual aids in quickly identifying which terms are most central to the text's discussion about safeguarding children, indicating which topics or concepts are given priority in the literature. As is seen in the heat map child is found at the top with a frequency of 169, which is convenient as it is the primary focus of the text.

5.0   INSIGHTS AND IMPLICATION

Emphasis on Child Welfare: The frequent mention of terms related to child protection laws, healthcare services, and support mechanisms underscores the significance of prioritizing child welfare and safety.

Legal Compliance and Intervention Strategies: References to legal acts and regulations highlight the importance of adhering to established frameworks for safeguarding children and addressing domestic abuse issues. Compliance with these policies is crucial for effective intervention and support.

Collaborative Efforts and Support Services: The presence of terms related to collaborative efforts among healthcare professionals, social services, and law enforcement agencies underscores the importance of coordinated actions in safeguarding vulnerable individuals, particularly children.

Referral Pathways and Early Intervention: Insights derived from the analysis shed light on the referral pathways and early intervention strategies outlined in the text. Clear protocols for identifying and responding to cases of child abuse or neglect are essential for timely and appropriate intervention.

Temporal References: Dates and temporal references such as "October 2014" and "March 2014" indicate specific time periods relevant to the context of the text. These temporal markers may correspond to events, regulations, or initiatives mentioned in the text.

5.1 Additional Insights

Further analysis could explore the relationships between identified entities and themes, providing deeper insights into the text's content and implications.

Integration of sentiment analysis techniques could help assess the emotional tone and implications of the text, enhancing understanding and interpretation.

Overall, the analysis offers valuable insights into safeguarding practices and highlights the importance of proactive measures and collaborative approaches in addressing child welfare concerns. The findings provide a foundation for informed decision-making and policy development in the field of child protection and domestic violence prevention.

5.2 Limitations

A limitation of our model can be seen from the common themes which highlights October 2014 despite that being less important than its rank suggests due to it appearing as footer on each page. Additionally, FII and Munchausen can be seen from the text to be the same illness but are classified separately by the model. Also, there are multiple references that refer to the Childrens Acts in part or entirely which are again classified separately and therefore underestimate the importance of the acts in the text.

5.3 Recommendations

Train the language model with explicit data to become domain-specific. It's a way of giving the language model information regarding specific vocabularies, syntaxes, and semantics of the area, hence helping in summarization or text generation.

Sentiment Analysis Incorporated: Sentiment analysis techniques, therefore, shall assist in trying to understand the emotional tone and the implications in what the text means to portray. This was, therefore, able to expand on the understanding and interpretation of the educational resources, which allowed the artificial intelligence tools to react better keeping in mind the nuances of language expressing the different sentiments, something that is very key in the form of child welfare and child protection.

6.0 CONCLUSION

In conclusion, the analysis of the sample text using NLP techniques provided valuable insights into common themes, key entities, and temporal references. By categorizing named entities into predefined types, the analysis facilitates a better understanding of the document's content and its implications in various domains such as healthcare, legal compliance, and child welfare and safeguarding. By leveraging NLP tools and techniques, stakeholders can gain a deeper understanding of text data, leading to more informed decision-making and action.

REFERENCE

Mid Essex Hospital Services NHS. Your Mandatory Training Booklet October 2014 First Edition

Appendix

Table 1: Named Entity Recognition Output

Entity Name	Entity Type
Mid Essex Hospital Services	ORG
NHS Trust	ORG
Training and Development	ORG
March 2014	DATE
Safeguarding Children and	ORG
Introduction Mid Essex Hospital Services NHS Trust	ORG
MEHT	ORG
Safeguarding Children and Young People	ORG
Working Together to Safeguard Children	ORG
Section 11	LAW
Children Act 2004	LAW
The Children Act	LAW
October 2014	DATE
First Edition	ORG
Munro	WORK_OF_ART
Victoria Climbie	PERSON
Data Protection Act	LAW
England	GPE
Wales	GPE
Northern Ireland	GPE
Foster	PERSON
MEHT Consent Policy	ORG
Slavery Servitude	ORG

English	LANGUAGE
Shared Family Assessment and Lead Professional	ORG
The Common Assessment Framework	LAW
Specialist Specialist	ORG
Specialist Midwife	ORG
Manas Datta	PERSON

Table 2: Common Themes

Common themes	Number of mentions
October 2014	11
1989	4
Section 47	3
Children Act	2
Age 16 18	2
UK	2
Family Solutions	2
Essex	2
IRT	2
999	2
CQC	1
Foster	1
NHS Trust	1
Training Development	1
Intercollegiate Document	1
Introduction Mid Essex Hospital Services	1
Safeguarding Children Young People	1
Section 11 Children Act	1

Data Protection Act	1
England Wales Northern Ireland	1
Foster	1
MEHT Consent Policy Children Act	1
Victoria Climbie Act	1
families/ household involvement local authority care	1
Training Booklet	1
FII	1
Munchausen	1
Imposition	1
Home	1

Fig. 1: Word Cloud

Fig. 2: Distribution of Importance Scores

Fig. 3: Heatmap of Top 20 Word Frequencies