



Innovation
Procurement
Empowerment
Centre

A Connected Places
Catapult Initiative

Extracting Insights from Procurement Data - A Text Mining Approach for Social Housing Retrofits in the West Midlands

Annum Rafique
City-REDI. University of Birmingham

Pei-Yu Yuan
University of Manchester

Research brief no. 10

Publication Date: 02/2025



UNIVERSITY OF
BIRMINGHAM

MANCHESTER
1824

The University of Manchester

Extracting Insights from Procurement Data - A Text Mining Approach for Social Housing Retrofits in the West Midlands¹

Executive Summary

To ensure public procurement aligns with strategic objectives and supports regional economic growth, data-driven methods are essential for analysing spending patterns and supplier engagement. A thorough examination of procurement contracts provides policymakers and stakeholders with valuable insights to assess procurement efficiency, optimise resource allocation, and monitor progress toward the goals.

This research brief explores how the data-driven methodological approach of text mining can be applied to procurement contract analysis, particularly in the context of social housing retrofitting and net-zero initiatives. The text mining methodology is based on keyword-based filtering and data extraction and offers a more straightforward alternative, relying on structured keyword identification and multi-stage filtering using spreadsheet-based tools.

By applying the methodology to procurement contracts from the [Tussell Database](#), this research brief examines the procurement landscape for social housing retrofitting in the West Midlands, uncovering key trends in 2022. The analysis highlights the disproportionate awarding of contracts to suppliers outside the region, the significant involvement of SMEs (65%), and the prevalence of short-term contracts, which may pose challenges for long-term retrofitting initiatives.

Future research directions include the use of artificial intelligence (AI) for contract selection, longitudinal analysis of procurement trends, supply chain mapping to understand contractor networks, and workforce development assessments to address skills gaps in retrofitting projects. The proposed methodology offers a scalable and adaptable framework for contract data analysis, with potential applications in broader procurement research and sustainability policy evaluations.

¹The West Midlands is one of nine official regions of England consisting of the counties of Herefordshire, Shropshire, Staffordshire, Warwickshire, West Midlands and Worcestershire.

1. Introduction

Understanding how government procurement practices contribute to advancing net-zero objectives is vital. By examining procurement strategies, we can gain insights into the adoption of innovative technologies and processes aimed at accelerating progress towards net-zero emissions. By understanding the procurement process and contracts, we can identify innovation capabilities at both firm and regional levels and assess the impact of procurement on transitioning to new sectors.

With the global imperative to combat climate change, governments worldwide are setting ambitious targets to reduce greenhouse gas (GHG) emissions. Taking the lead in this effort, the United Kingdom has pledged to reach net-zero carbon emissions by 2050, marking a pioneering step as the first major economy to make this commitment. This commitment, based on scientific evidence and backed by public support, underscores the urgency of addressing climate change.

To translate this ambitious goal into action, the UK government is leveraging its procurement processes as a strategic tool for driving innovation and sustainability across sectors. A recent policy measure mandates suppliers for major government contracts to demonstrate their commitment to achieving net-zero emissions by 2050 and publish comprehensive carbon reduction plans (Cabinet Office, 2021).

In light of this policy context, we aim to examine the text-mining methodological approach to analyse and derive insights from procurement contract data. We aim to effectively extract and structure relevant data from procurement contracts by employing text mining tools, data analysis techniques, and domain-specific knowledge. Through a multi-stage filtering process and identifying pertinent keywords, the proposed methodology

facilitates extracting and organising information related to the desired topics, such as net-zero initiatives and social housing retrofitting activities.

In this research brief, we examine and highlight the value of a text-mining approach adopted to analyse social housing retrofit procurement contracts. The data analysis and visualisation components will enable a comprehensive assessment of various aspects, including contract values, supplier dynamics, geographical distribution, and other relevant factors. Additionally, the methodology incorporates expert knowledge to ensure the research aligns with the unique context and objectives of the study.

2. Methodology

We use the data-driven text mining methodology to analyse procurement contract data. The methodology is more manual and relies heavily on creating good keyword lists and carefully filtering the data compared to complex methodologies involving more intricate steps like sentence labelling and language model training (Yıldız, 2019). The text mining methodology can be implemented using relatively simple tools in MS Excel or another spreadsheet program, making it more accessible for those without expertise in advanced data analysis, machine learning or natural language processing (NLP) techniques.

The text mining approach has been widely used to analyse contracts in various industries. Marzouk and Enaba (2019) used this approach to understand the nature of communication in construction project contracts. They highlighted the importance of this approach in the architecture, engineering and construction sectors where the contract documents are complex and unstructured. Shamshiri, Ryu and Park (2024) also

highlighted the growing need for text analytics in construction management, where over 80% of data is unstructured, spanning emails, drawings, contracts, and reports. They suggested that with the increasing volume of construction projects, the adoption of big data analytics and automation tools has become essential to enhance efficiency and reduce workload.

In the context of green and sustainable procurement, El Haddadi *et al.*(2021) used the approach to analyse and evaluate tenders for purchasing IT equipment in Morocco. Their study findings indicated that procurement tenders rarely included environmental considerations, suggesting a low infiltration rate of sustainability-related criteria. Fantoni *et al.* (2021) developed a text-mining tool to assist engineers in the railway sector during the most repetitive tender phases. Their findings indicate that automating tender analysis through text mining saves time and provides a strategic advantage to companies by improving efficiency in bid evaluation.

The strength of this approach lies in its simplicity and ease of implementation. The downside is that it may not be as accurate or comprehensive as some of the complex methodologies driven by artificial intelligence (AI), which uses machine learning techniques. However, for basic analysis of procurement contracts, the text mining methodology can still provide valuable insights, especially when dealing with large volumes of text data that would be impractical to review in its entirety manually.

Step 1 - Data Extraction: Extract the required data from the relevant database.

Step 2 - Keyword Identification: Use expert knowledge or domain-specific literature to develop a comprehensive list of:

- **Inclusionary keywords** - specific terms used to identify relevant procurement contracts that align with the research focus. These keywords are selected based on expert knowledge, domain-specific literature, and policy objectives to ensure that the dataset includes contracts related to the relevant context.
- **Exclusionary keywords** - terms or phrases used to remove irrelevant contracts from the dataset. These keywords help eliminate contracts unrelated to the study's scope, preventing misclassification and ensuring data accuracy.

Step 3 - Multi-stage Filtering: Implement a multi-stage filtering process to refine the results:

- **Stage 1:** Filter using the inclusionary keywords related to the topic of interest.
- **Stage 2:** Filter out irrelevant contracts using exclusionary keywords.

Step 4: Manual review: Manually review the filtered contracts from Step 3.

Step 5 - Data Extraction: Extract relevant information from the filtered contracts, such as contract details (e.g., contract ID, contracting authority, supplier, contract value, start and end dates, duration) and project descriptions. Structure the extracted data in a consistent format (e.g., database, spreadsheet) for further analysis.

2.1. Methodological Consideration

When analysing procurement contracts, a few important considerations must be made.

- **Multiple keyword occurrences:** The contract documents may contain more than one relevant keyword. To avoid double-counting or inflating the results, the analysis must ensure that each document is counted only once, even if it contains multiple relevant keywords.
- **Multiple suppliers per contract:** It is possible for a single procurement contract to involve more than one supplier or contractor. This could happen if the project is divided into different components or phases, each awarded to a separate supplier. The analysis should account for this possibility and correctly attribute the contract to all the relevant suppliers.
- **Keyword ambiguity:** Some keywords used may appear in contexts other than what is required. A two-stage approach may be beneficial to ensure accurate results.

3. Example: An approach for social housing retrofits

We apply the text mining approach to analyse procurement contracts related to social housing retrofitting projects in the West Midlands. Retrofitting housing stock with energy-efficient technologies and sustainable practices is essential for achieving net-zero goals. It reduces carbon emissions, creates jobs across the supply chain, and delivers broader socioeconomic benefits.

To gain a deeper understanding of the procurement landscape in the West Midlands Combined Authority (WMCA) area, a subregion within the broader West Midlands,

refer to the IPEC publication: [“Retrofitting Social Housing in the West Midlands Combined Authority \(WMCA\) Area²”](#). This report by Rafique (2025) provides valuable insights into procurement challenges, supply chain dynamics, and policy considerations specific to the region.

By examining the procurement contracts for house retrofitting, we can help policymakers, housing authorities, and stakeholders understand the scale, scope, and regional distribution of retrofitting efforts, informing better planning and execution of projects. The insights gained can also support researchers and advocacy groups working towards sustainable and affordable housing solutions.

3.1. Data and Methodology

To identify the social housing retrofitting procurement contracts using the **Tussell Database**, a platform established in 2015 to enhance transparency, competition, and accountability in government contracting. Tussell provides comprehensive data on government contracts and spending, supporting effective collaboration between the public and private sectors in delivering high-quality public services.

We extracted the following data fields from the Tussell database:

- Procurement notice
- Contracting Authority
- Supplier
- Total award value
- Award value per supplier

² The West Midlands Combined Authority is county in the West Midlands region which includes Wolverhampton, Dudley, Walsall, Sandwell, Birmingham, Solihull, Coventry and Worcestershire. The WMCA area is home to almost 50% of the West Midlands region population.

- Supplier award count
- Contract start date
- Contract end date
- Contract duration (months)
- Contract Type
- Contract Title
- Contract description

Step 1: Data Extraction from Tussell Dashboard:

As a starting point, we conducted a single-year analysis focusing on **2022**. During this period, **4,499 procurement notices** were issued by various contracting and local authorities in the **West Midlands**, resulting in **5,315 contracts awarded** to the suppliers.

The higher number of contracts than procurement notices suggests that multiple suppliers often fulfil a single procurement notice— a common practice in large-scale projects to ensure competition, diversify the supplier base, and efficiently cover different project components.

Step 2: Keyword Identification: Using expert knowledge and domain-specific literature, we created a list of inclusionary and exclusionary keywords required to identify social housing procurement contracts.

- Total inclusionary keywords = 24 (Table 1)
- Total exclusionary keywords = 16. (Table 2)

Step 3: Multi-stage Filtering: We perform text mining on the pre-processed data using the identified keywords to extract relevant information.

- **Stage 1:** Filtered data using inclusionary keywords related to the topic of interest, e.g., social housing retrofitting, net-zero, as

given in Table 1. This process identified 1,294 keyword occurrences across 1,099 contracts, indicating that some contracts contained multiple relevant keywords (Table 1).

- **Stage 2:** Filtered out irrelevant contracts using exclusion keywords, e.g., schools, hospitals, and private sector. This step filtered down the dataset from 1,099 contracts to 269 contracts that met the selection criteria.

Table 1: Inclusionary keywords - Net-zero social housing retrofitting keywords REQUIRED in procurement documents

Inclusionary keywords	Number of contracts
solar	10
heat pumps	2
insulation	12
glazing	4
boilers	12
ventilation	8
biomass	2
heating control	0
thermostat	0
draught proofing	2
energy-efficient doors/windows	2
retrofit	21
social	159
housing	116
SAP	13
EPC	9
sustainability	12
green	49
homes	693
carbon	35
zero	18
flat	31
SHDF	2
light	82
Total	1,294

Table 2: Exclusionary Keywords - Keywords NOT REQUIRED in Procurement Documents.

Exclusionary Keywords	Number of contracts
LAD	22
Sustainable Warmth Competition	4
school	262
hospital	100
university	535
private	25
owner	20
children	155
office	103
adult	161
care	760
water	123
nuclear	6
waste	71
company	127
food	63
Total	2,537

Step 4: Manual review: A detailed manual review of the 269 filtered contracts was conducted to ensure relevance and accuracy in contract selection.

Step 5: Contract and Supplier Identification: We extracted the data from filtered social housing retrofitting procurement contracts:

- We also identified 13 contracting authorities and 20 individual suppliers, implying multiple contracts with a single supplier in a year. The list of contracting authorities and suppliers is given in Table 3.
- We identified 19 procurement notices with 23 contracts through the two-stage filtration process, showing multiple suppliers for a single procurement notice.

A flowchart of the text mining methodology is presented in Figure 1.

Table 3: List of contracting authorities and suppliers

Contracting Authority	Supplier
Birmingham City Council	Arcadis Consulting (Uk) Limited
Bromsgrove District Council	Bidconnector Limited
Citizen Housing Group Ltd	Cocuun Ltd
Communities & Housing Investment Consortium (CHIC)	Communities & Housing Investment Consortium (Chic)
Connexus Housing Ltd	Complete Certification Ltd
Coventry City Council	Eco Approach Ltd
NATIONAL GAS TRANSMISSION PLC	Equans Regeneration (Fhm) Limited
Platform Housing Group	Evolve Home Energy Solutions Ltd
Platform Plus Limited	Jones Lang Lasalle Services Limited
Sandwell Metropolitan Borough Council	Morgan Lambert Limited
Shropshire Council	My Energy Solutions (Wm) Limited
Staffordshire County Council	Novus Property Solutions Limited
Worcester City Council	Parcor Ltd
	Pennington Choices Limited
	Property Tectonics Limited
	Reed Talent Solutions Limited
	Sustainable Building Services (Uk) Limited
	Synergize Ltd
	Yesss (B) Electrical Ltd
	1 unknown

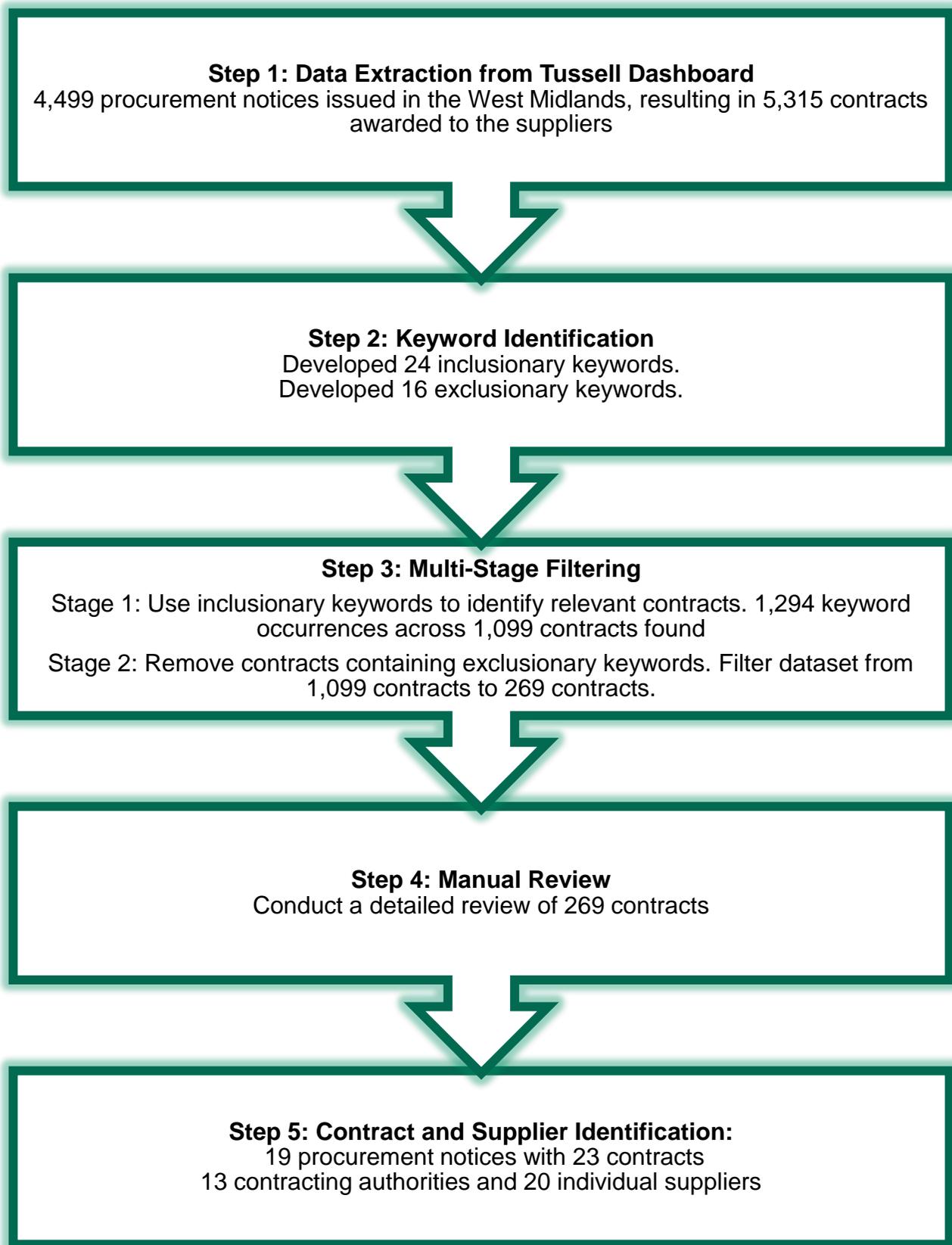


Figure 1: Flowchart of text mining methodology for analysing social housing procurement contracts

3.2. Data Analysis and Findings

Contract Value: Table 4 summarises the total, minimum, maximum, and average contract values across the identified social housing retrofitting contracts.

The average contract value is overestimated due to an outlier contract (CHIC) with a significantly higher value than the others, likely due to its longer. The distribution of awarded contracts is shown in Figure 2. 65% of the contracts were valued below £500,000, with seven contracts falling under £100,000 and eight contracts ranging between £100,000 and £500,000.

Table 4: Contract value summary (£)

	Total	Minimum	Maximum	Average
Total Contract Value – High (£)	315,022,735	4,480	280,000,000	14,319,215
Total Award Value (£)	314,801,908	4,480	280,000,000	14,309,178
Award Value per Supplier (£)	311,069,130	4,480	280,000,000	14,139,506

Suppliers per contract: Table 5 shows the distribution of procurement contracts based on the number of suppliers involved.

Table 5: Number of suppliers under one procurement award

Number of suppliers in a contract	Number of contracts
1	16
2	2
3	1
unknown	1

It indicates that **most contracts involved a single supplier**, while a few involved multiple suppliers.

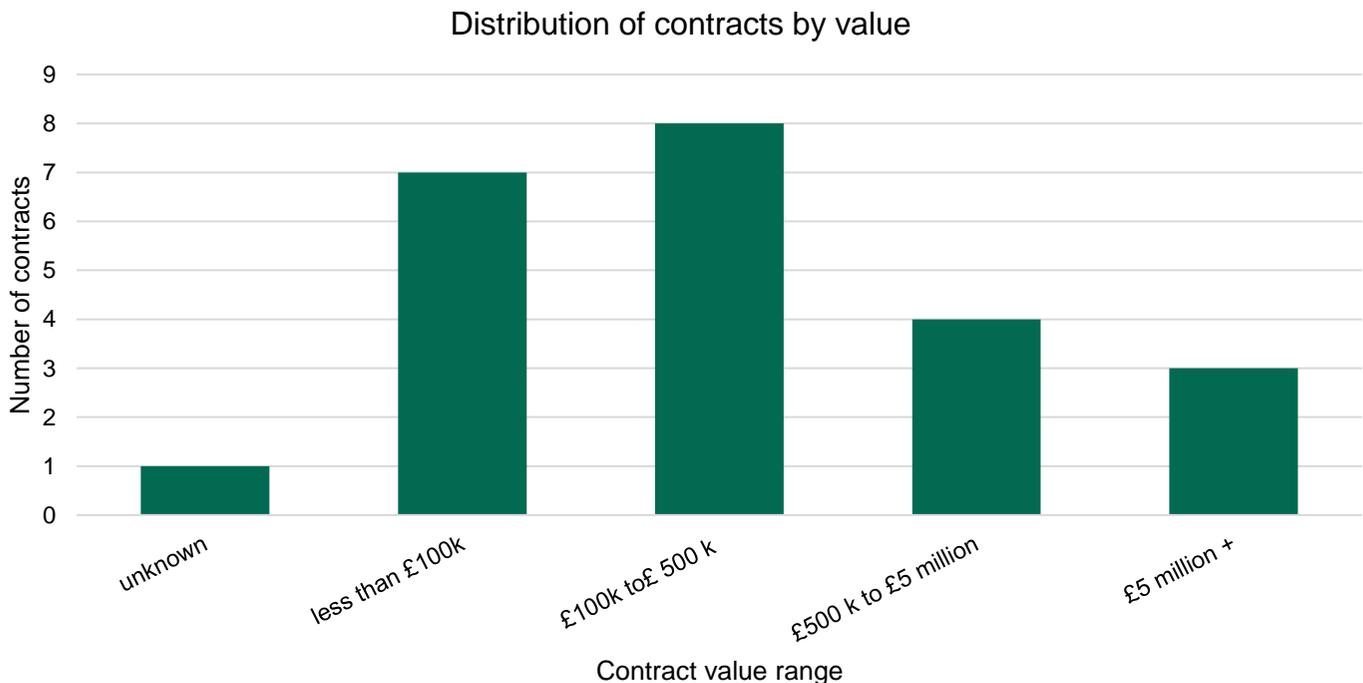


Figure 2: Distribution of contracts by value

Contract Duration: The time duration of procurement notices is given in Table 6.

Table 6: Time duration of procurement notices.

Time duration	Number of procurement notices
unknown	2
5 months	3
6 months	3
8 months	2
12 months	2
18 months	1
24 months	1
36 months	2
48 months	2
85 months	1

Half of the procurement notices issued in the year were for contracts of a year’s duration or less. **Most contracts are short-term, an issue as retrofitting activities typically require a longer timeframe.**

Location of Suppliers: The locations of the suppliers for the contracts located in the West Midlands regions are presented in Table 7.

Table 7: Location of the suppliers for the West Midlands region

	Number of suppliers
Yorkshire And The Humber	4
East Midlands	1
North West	4
West Midlands	3
South East	1
Wales	1
London	4
Unknown	1
North East	1
Total	20

The supplier distribution shows that the highest number of suppliers are located in Yorkshire and the Humber, London, and the North West (4 suppliers each).

The West Midlands, the selected region, has only three suppliers, which suggests that **many contracts are awarded to suppliers outside the immediate local areas of the contracting authorities.**

SME Status of Suppliers: Table 8 presents the classification of suppliers based on their SME status.

Table 8: Are the Suppliers SMEs?

	Are the Suppliers SMEs?
Yes	13
No	7

The SME participation is significant, with 13 out of 20 suppliers (65%) being SMEs. This suggests **a positive trend towards engaging smaller firms** in social housing retrofitting projects, which can enhance local economic development and competition.

4. Potential Future Research Directions

- **Use of AI in analysis and contract selection:**

The AI-driven methodology leverages NLP and machine learning, involving manual sentence labelling and large language model (LLM) training to enhance accuracy and automation in contract analysis. While this approach requires significant initial setup, it enables more precise and scalable text analysis, making it valuable for large datasets.

This approach combines human knowledge with AI. By leveraging a pre-trained language model, the analysis of extensive textual data can be streamlined and automated. However, AI driven analysis requires specialised training and knowledge, making it less accessible to those without a background in machine learning or NLP (Fantoni *et al.*, 2021). This presents a barrier for policymakers, researchers, and procurement professionals who may lack the technical skills to use the AI model effectively.

The methodology described below is presented as a general framework, and the specific implementation details, such as the choice of language model, labelling guidelines, and evaluation metrics, would need to be determined based on the specific requirements and constraints of the project.

Step 1 - Identifying Keywords: Identify relevant keywords related to the topic of interest.

Step 2 - Ranking Keywords: Establish a ranking system for these keywords to prioritise their importance.

Step 3 - Filtering Sentences: Once we have identified and ranked the keywords, we'll need to filter sentences based on these keywords to extract relevant information.

Step 4 - Labelling Sentences: This step involves labelling sentences through a systematic process to outline guidelines and rules for training models. This will likely be an iterative process until we complete 500-1000 sentences, depending on the scale and complexity of the data we aim to process.

Step 5 - Using an LLM to perform the labelling task: we designed a framework to support AI annotators with prompt instructions for the AI language model, such as ChatGPT

or BERT, to understand and classify text accordingly.

- **Analysing longitudinal data**

While the proposed methodological approach of text mining provides a framework for analysing procurement contract data related to social housing retrofitting and net-zero initiatives, it is important to note that access to data spanning multiple years would be necessary to draw more informed and comprehensive conclusions. Having **a longitudinal view of the data** would enable researchers and policymakers to effectively track trends, evaluate the impact of policies and initiatives over time, and make data-driven decisions for long-term planning and resource allocation strategies.

- **Analysing the Retrofitting Supply Chain**

We can list the key suppliers and contractors involved in social housing retrofitting projects in the region. This information can help **map out the supply chain network** and **understand the major players** in the retrofitting industry. The distribution of contract values among different suppliers will reveal information about their market share and **capacity to take on large-scale projects**. This can then help **identify potential supply chain clusters** or regions with a higher concentration of retrofitting activity. It will also help identify contracts where smaller, local suppliers are involved in retrofitting projects. This information can be used to evaluate the participation of **local businesses in the supply chain and the potential barriers** they may face. Moreover, we can assess the resilience and potential risks within the retrofitting supply chain by combining supplier information with contract values, geographical distribution, and project timelines.

By mapping the retrofit supply chain through procurement contracts, we can identify dependencies on specific suppliers or regions and potential disruptions or capacity constraints.

• **Workforce Development and Skills Gap Analysis**

By analysing contract descriptions and project details extracted through the methodologies, we can **identify specific retrofitting measures** being implemented, such as insulation, heating system upgrades, energy-efficient window replacements, or renewable energy installations. Each measure requires a unique set of workforce skills; for instance, insulation projects necessitate expertise in installing various materials, adhering to codes, and ensuring proper air sealing, while renewable energy installations demand knowledge of solar photovoltaic systems, wind turbines, or biomass boilers. Mapping these measures to required skills reveals the **workforce capabilities** needed for implementation.

Furthermore, the analysis provides insights into retrofitting contracts' geographical distribution and volume, enabling estimation of potential **demand for specific skills across regions**. Comparing identified skill requirements with existing workforce availability can **highlight skill gaps** within the retrofitting industry. These insights inform workforce development strategies, guiding the **creation of training programs, apprenticeships, or upskilling initiatives** tailored to the skills required for various retrofitting measures. Policymakers and stakeholders can leverage this information to **develop targeted incentives or support programs**, addressing skill gaps and ensuring a sufficient supply of skilled workers for the growing retrofitting industry.

5. Potential Users

The proposed approach would be of value to several user groups, including:

- Government agencies and policymakers responsible for housing, energy, and environmental policies.
- Local authorities and housing associations involved in social housing management and retrofitting initiatives.
- Researchers and academics studying energy efficiency, sustainable housing, and climate change mitigation.
- Industry stakeholders, such as contractors, suppliers, and retrofitting service providers, to understand market dynamics and opportunities.
- Environmental organisations and advocacy groups interested in monitoring progress towards net-zero goals.

6. Conclusion - Ways forward

As illustrated above, with the case of social housing retrofit, the proposed methodological tool offers numerous potential benefits, including mapping out the supply chain by identifying various retrofitting programs and engaging stakeholders to address procurement barriers and challenges. It also facilitates the analysis of supplier relationships by examining consistency across contracts, thereby uncovering any instances of preferential treatment or longstanding relationships between local councils and contractors. Additionally, the tool allows for the refinement of contract categories, enabling clearer insights into fund allocation based on the specific nature of retrofitting projects. Moreover, it enhances understanding by contextualising contract types within the desired domain, providing a deeper comprehension of the work being carried out and associated risks, ultimately improving the overall analysis and decision-making process.

This methodological approach provides a systematic and structured way to leverage text mining and data analysis techniques to gain valuable insights from procurement contract data. It can inform policy decisions, resource allocation strategies, and supply chain management, ultimately contributing to achieving net-zero goals and sustainable housing initiatives.

The methodological approach can be adapted and applied to other domains and studies that analyse and rely on extensive textual data, such as contracts, reports, or project documentation for **infrastructure projects, renewable energy initiatives, or environmental conservation efforts**. For instance, by modifying the keyword lists and

filtering criteria, the approach can be tailored to extract information from contracts, enabling targeted analysis of procurement trends, policy compliance, and sustainability commitments across different industries.

References

- Cabinet Office. (2021). Firms must commit to net zero to win major government contracts. Lord Agnew.
- El Haddadi, T., El Haddadi, O., Mourabit, T., El Allaoui, A. and Ben Ahmed, M. (2021) 'Automatic analysis of the sustainability of public procurement based on Text Mining: The case of the Moroccan ICT markets', *Cleaner and Responsible Consumption*, 3, pp. 100037.
- Fantoni, G., Coli, E., Chiarello, F., Apreda, R., Dell'Orletta, F. and Pratelli, G. (2021) 'Text mining tool for translating terms of contract into technical specifications: Development and application in the railway sector', *Computers in Industry*, 124, pp. 103357.
- Marzouk, M. and Enaba, M. (2019) 'Text analytics to analyse and monitor construction project contract and correspondence', *Automation in Construction*, 98, pp. 265-274.
- Rafique, A. (2025) *Retrofitting Social Housing in the West Midlands Combined Authority (WMCA) Area: Overcoming Procurement Barriers for a Net-Zero Future: Innovation Procurement Empowerment Centre.*
- Shamshiri, A., Ryu, K. R. and Park, J. Y. (2024) 'Text mining and natural language processing in construction', *Automation in Construction*, 158, pp. 105200.
- Yıldız, T. (2019) 'Examining The Concept of Industry 4.0 Studies Using Text Mining and Scientific Mapping Method', *Procedia Computer Science*, 158, pp. 498-507.



**Innovation
Procurement
Empowerment
Centre**



**UNIVERSITY OF
BIRMINGHAM**



The University of Manchester

Author Details

Annum Rafique

City-REDI

University of Birmingham

a.rafique@bham.ac.uk

Pei-Yu Yuan

University of Manchester

pei-yu.yuan@manchester.ac.uk

IPEC website: <https://www.ipec.org.uk/>

IPEC LinkedIn: <https://www.linkedin.com/company/93121184>

IPEC email: contact@ipec.org.uk