**Dissertation title**

Machine learning classification and spatial analysis of planning data and aerial imagery in Scotland in support of the monitoring of the aggregate extraction requirements of National Planning Framework 4 and the forthcoming Aggregate Levy Register.

**Introduction**

There is a growing aspiration for the use of machine learning and spatial analysis techniques in the urban planning system but this is not yet deployed on a systematic basis outwith niche areas of innovation. The manageable administrative and spatial scale of the planning system in Scotland offers an opportunity to do so, with the recent publication of the National Planning Framework 4 (NPF4) and an increasing centralisation of locally-applied policies at the regional and national levels, alongside a growth in the availability of development planning and development management data.

One such policy area is that of aggregate extraction, both essential for the construction industry and subject to taxation under new devolved arrangements in Scotland. This encompasses all sand, gravel and rock that has been dug from the ground, dredged from the sea or imported. NPF4 requires local development plans (LDPs) to support a 10 year landbank of aggregates and the forthcoming Aggregates Levy, to be applied in 2025, has the need to create and maintain a site register. There is an inherent conflict between both of these aims, as aggregates policy is to promote reuse of construction and demolition waste material, rather than new extraction, underpinning the importance of policy monitoring.

Machine learning techniques can offer analytical support to both of these policy aims, with the use of natural language processing and unsupervised and supervised classification techniques applied to unstructured and structured data. This will allow the cleaning of data, classification to a common taxonomy, analysis of supply versus demand and the potential for monitoring existing sites and identifying new sites using planning casework data, national minerals mapping and imagery analysis. The application of such techniques to a discrete policy area will provide a useful example for more complex policy areas.

**Literature review**

The literature in this area is disparate, although the research and use of machine learning (ML) and artificial intelligence (AI) in urban planning is growing and converging. From a technical perspective, Tekouabou et al (2021) have considered the application of machine learning methods in urban planning, with a useful taxonomy of methods, models and indicators. However, this does not cover NLP in detail, although it is mentioned in the context of neural networks. Likewise, Chaturvedi and de Vries (2021) provide a good overview of the technical applications, but largely in the context of earth observational data, albeit with a good breakdown of land use indicators, data sources, measurements and applications. Nevertheless, there is a comprehensive overview of ML and AI methods and applications that provides a useful baseline for methodological and model development and a recent overview of the use of machine learning for spatial analyses is also offered by Casali, Aydin and Comes (2022), drawing upon Google Scholar and the Web of Science with a focus upon 2021, identifying the most prominent topics and uses and also knowledge gaps in the sector.

NLP needs to be considered as a detailed subset of ML and AI for a more granular trawl of the literature. Recent research by Mieczko and Desmond (2023) considers the use of NLP in a national zoning and land use database from publicly available administrative data in the United States. Fu, Li and Zhai (2023) have also looked at NLP in the study of resilience plans from the 100 Resilience Cities network, with useful coverage of topic modelling. Brinkley and Stahmer (2021) have looked in more geographical detail at city planning in California. A systematic review of NLP for urban research is also provided by Cai (2021) highlighting a limited use of NLP in urban studies, albeit growing in use in recent years, with a comprehensive survey of literature and geographical areas of study.

Research in the UK is more limited but there are examples from the UK government. Dray (2019) has considered the use of NLP in government and there is a report by the Turing Institute (2021) researching the use of AI and ML in automating planning applications, which does not consider NLP in detail other than in data cleaning, but does look at the application of AI and ML techniques in the context of the English planning development management system, which is essentially the same as in Scotland.

The above research is largely exploratory, demonstrative or occasional, rather than operational. A systemic example of NLP in a production environment is provided by work by NatureScot (2021) reported on the UK Authority website (2022) that illustrate the progress of the InformedDecision platform. This provides for a further case study in terms of potential capability.

For imagery analysis, Evans (2019) has a comprehensive review of the employment of aerial investigation and mapping by Historic England, in the UK context. This does not consider the use of machine learning but highlights practical GIS challenges. Building on the literature relating to the general applicability of machine learning in urban planning, there is a significant weight of research into the classification of aerial imagery. Ouchra et al (2022) and Mehmood (2023) provide an overview of the main methods – manual, unsupervised and supervised learning – related to satellite data and Boguszewski et al (2021) consider the scope for automatic visual mapping at a higher resolution from LandCover.ai aerial imagery using the DeepLab model.

The minerals industry has a wealth of literature considering the use of big data and machine learning techniques, but this can be complex and specific to the geology sector. Xihong et al (2018) provide an example of the use of deep learning to map iron mineralisation for exploitation purposes. A useful overview is provided by Zuo (2020) – a contributor to the preceding study – on geodata science-based techniques for minerals mapping, employing algorithms, that notes the need for text mining..

From the perspective of planning in Scotland, there is no literature pertaining to the use of AI and ML in the planning system as it stands. Information is available on planning performance statistics, collected regularly and published quarterly and annually by the Scottish Government. The most recent statistics (2021/22) provide a categorisation of application and development types which, aligned with the topic modelling and UK-specific research mentioned earlier, provides the elements of a taxonomy for NLP and categorisation. However, for minerals, this is generic and does not consider the types of extraction, ie for aggregates.

The recently-published NPF4 (2023) is the overarching land use policy document for Scotland. Among the policy requirements is that LDPs should support a landbank of construction aggregates of at least 10-years at all times.

In terms of the aggregate levy, this was introduced for devolution from the UK government (dating from 2002) by the Scotland Act 2016 and the legislative process is continuing with the completion of initial legal discussions in 2019, research published in 2020, and the commitment to aggregates levy legislation made in the Programme for Government 2022-23. Comprehensive research published in 2020 by the Scottish Government contains quantitative information and also coverage of market issues, such as the balance between recycling and import and export, with the consideration of recycling of construction and demolition waste as a positive outcome , although notes a lack of recent Scottish Aggregates Survey data since 2012. As well as the market review, the spatial coverage of aggregates operations highlights West Central Scotland, Tayside and Fife, and also Highland and Moray as the primary extraction areas.

**Methodology**

This research employs Python and various libraries throughout the process. This is summarised below in **Figure 1**.

**Figure 1. Methodology**

A screenshot of a computer

Description automatically generated

There are three main stages:

**1. Data Identification, Gathering and Analysis**

Two main datasets are used for this analysis - Planning Applications and Development Plan Policies – alongside minerals mapping from British Geological Survey and aggregates industry information and also aerial imagery. These are detailed below and summarised in **Figure 2**.

**Planning Applications:** A historic point and polygon dataset of up to 1million records dating back to 2009. This will provide, at various scales, both an overview and a detailed record of historical land use development activities. There is an element of language consistency within this dataset. The fields include the general description of development, some coarse categorisation by land use and application type, planning decision (approve or refuse) and casework dates.

**Development Plan Policies:** The overarching national plan is NPF4, with subordinate local development plans (LDPs) which will be required to define minerals subject policies, including aggregate extraction. As NPF4 was published in 2023, there will be no LDPs prepared under the overarching requirements, although aggregate supply policy is unlikely to have changed.

**Minerals Mapping:** BGS has comprehensive coverage of geological maps at 1:10,000, 1:25,000 and 1:50,000 scales. BGS has also published minerals mapping covering the Central Belt of Scotland. This includes crushed igneous rock aggregates and sand and gravel. The BRITPITS database also contains details of active and historic quarrying and minerals workings.

**Aerial Imagery:** A range of resolutions and years are available within 1km squares. The likely resolution is 25cm, although 12.5cm will be investigated, subject to processing feasibility and study areas.

**Figure 2. Datasets and Uses**

|  |  |  |
| --- | --- | --- |
| **Datasets** | **NPF4 Requirements** | **Aggregate Levy Requirements** |
| Ten year supply of aggregate land | National Aggregates Register |
| Local Development Plans | Land supply measurement | Search areas for site identification |
| BGS minerals map[[1]](#footnote-1) | Search areas for additional supply potential | Search areas for site identification |
| BRITPITS[[2]](#footnote-2) | Historic and current activity | Historic and current activity |
| Planning Applications | Activity monitoring | Site identification |
| Aerial imagery | Activity monitoring | Site identification |
| Industry data | Capacity forecasting |  |

The definition of study areas will be undertaken at this stage and is anticipated to include some of the following planning authority areas, possibly at the regional level:

* Aberdeenshire
* Angus
* Argyll and Bute
* Clackmannanshire
* Dumfries and Galloway
* East Ayrshire
* East Dunbartonshire
* East Lothian
* East Renfrewshire
* Falkirk
* Fife
* Highland
* Midlothian
* Moray
* North Ayrshire
* North Lanarkshire
* Renfrewshire
* South Ayrshire
* South Lanarkshire
* West Dunbartonshire
* West Lothian

This stage includes exploratory data analysis, to allow more precise definition of the analytical process, and preliminary visualisations. However, a common taxonomy, vocabulary and categorisation system will be required, which is considered in **Stage 2** below.

**2. Text Mining and Classification**

In order to allow meaningful analysis, entities require to be categorised in the same way. Also, some datasets have essential information that is locked up in text fields. This is true for both the NPF and also for planning applications, which have a narrative description of developments that includes quantitative information.

NLP techniques are used to extract tokens and also quantities and units of development, to provide both a keyword/token list for categorisation, quantitative values otherwise missing, and also to label data that can be used to train a classifier. At this stage, these two elements are considered as an output required to progress to **Stage 3**, but also as an assessment of the success of NLP techniques as applied to this research requirement.

In this case, NLP will be used to identify and quantify planning applications for aggregate-related activity and also to eliminate ancillary uses.

**3. Data Modelling, Analysis and Visualisation**

Appropriately-categorised datasets can be used to identify sites for the Aggregates Levy Register and to assess the delivery of the policies required by NPF4. This will be augmented by an analysis of demand based upon industry data and historic planning applications.

These datasets will also be used to identify clusters of aggregates activity to support the definition of minerals search areas and to train imagery analysis models to identify active aggregate sites.

The outcome of this stage will be a series of methods, models and visualisations that allow the classification and identification of aggregate extraction sites from planning casework data, development plans and imagery analysis, with an analysis of supply versus demand at various scales. Recommendations of appropriate techniques for development plan monitoring and aggregates register population and monitoring will follow from this stage, with observations as to the scope for wider application in the planning system.

**Qualitative Analysis**

In addition to the quantitative analysis, subject matter experts in the aggregates and planning sectors will be consulted.

**References:**

***Papers***

Bide, T., Evans, E., Idoine, N. E, & Mankelow, J. (2023) *United Kingdom Minerals Yearbook 2022.* Nottingham, UK, British Geological Survey, 70pp. (OR/23/001) Online] Available from: <https://nora.nerc.ac.uk/id/eprint/534312/>

Boguszewski, A., Batorski, D., Ziemba-Jankowska, N., Dziedzic, T. & Zambrzycka, A. (2021)*. LandCover. ai: Dataset for automatic mapping of buildings, woodlands, water and roads from aerial imagery.* Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 1102-1110). [Online] Available from: <https://openaccess.thecvf.com/content/CVPR2021W/EarthVision/papers/Boguszewski_LandCover.ai_Dataset_for_Automatic_Mapping_of_Buildings_Woodlands_Water_and_CVPRW_2021_paper.pdf>

Brinkley, C., & Stahmer, C. (2021). *What Is in a Plan? Using Natural Language Processing to Read 461 California City General Plans*. Journal of Planning Education and Research, 0(0).  [Online] Available from: <https://doi.org/10.1177/0739456X21995890>

Cai, M. (2021). *Natural language processing for urban research: A systematic review.* Heliyon. 2021 Mar 8;7(3):e06322. [Online] Available from: <https://doi.org/10.1016/j.heliyon.2021.e06322>

Casali, Y., Aydin, N. Y., & Comes, T. (2022). *Machine learning for spatial analyses in urban areas: a scoping review.* Sustainable Cities and Society, Volume 85, 2022, 104050, ISSN 2210-6707. [Online] Available from: <https://doi.org/10.1016/j.scs.2022.104050>.

Chaturvedi, V. & de Vries, W. (2021). *Machine Learning Algorithms for Urban Land Use Planning: A Review.* Urban Sci.2021*,* 5*,* 68.  [Online] Available from: <https://doi.org/10.3390/urbansci5030068>

Evans, S. (2019). *Historic England Aerial Investigation and Mapping Standards Technical Review.* Historic England Research Report Series 46-2019.

Fu, X., Li, C. & Zhai, W. (2023). *Using Natural Language Processing to Read Plans.* Journal of the American Planning Association, 89:1, 107-119. [Online] Available from: <https://doi.org/10.1080/01944363.2022.2038659>

Mehmood, M., Shahzad, A., Zafar, B., Shabbir, A. & Ali, N. (2022). *Remote Sensing Image Classification: A Comprehensive Review and Applications. M*athematical Problems in Engineering vol 2022. [Online] Available from: <https://doi.org/10.1155/2022/5880959>

Mleczko, M., & Desmond, M. (2023). *Using natural language processing to construct a National Zoning and Land Use Database*. Urban Studies, 0(0). [Online] Available from: <https://doi.org/10.1177/00420980231156352>

Ouchra, H., Belangour, A. & Erraissi, A. (2022). *Machine learning for satellite image classification: A comprehensive review*. International Conference on Data Analytics for Business and Industry (ICDABI), Sakhir, Bahrain, 2022, pp. 1-5. [Online] Available from: <https://ieeexplore.ieee.org/document/10041606/>

Tekouabou, S. C. K., Diop, E.B., Azmi, R., Jaligot, R., Chenal, J. (2021). *Reviewing the application of machine learning methods to model urban form indicators in planning decision support systems: Potential, issues and challenges.* Journal of King Saud University - Computer and Information Sciences, Volume 34, Issue 8, Part B. pp 5943-5967, ISSN 1319-1578. [Online] Available from: <https://doi.org/10.1016/j.jksuci.2021.08.007>

Xiong, Y., Zuo, R. & Carranza, E. J. M. (2018). *Mapping mineral prospectivity through big data analytics and a deep learning algorithm*. Ore Geology Reviews, Volume 102, 2018, Pages 811-817, ISSN 0169-1368. [Online] Available from: <https://doi.org/10.1016/j.oregeorev.2018.10.006>.

***Other References***

British Geological Survey & Scottish Government (2008). *A Guide to Minerals Information in the Central Belt of Scotland.* [Online] Available from: <https://www2.bgs.ac.uk/mineralsuk/download/scotland/mins_info_scotland_guide.pdf>

British Geological Survey (2020). *Directory of Mines and Quarries.* [Online] Available from: <https://www2.bgs.ac.uk/mineralsuk/download/dmq/Directory_of_Mines_and_Quarries_2020.pdf>

Data Study Group team. (2021). *Data Study Group Final Report: Agile Datum*. Turing Institute. [Online] Available from: <https://doi.org/10.5281/zenodo.5729594>

Dray (2019). *Natural Language Processing in Government.* [Online] Available from: <https://dataingovernment.blog.gov.uk/2019/06/14/natural-language-processing-in-government/>

NatureScot (2021). *Digital Revolution For Nature Sites.* [Online] Available from:  <https://www.nature.scot/digital-revolution-nature-sites>

Scottish Government (2022). *Policies and Taxes.* [Online] Available from: <https://www.gov.scot/policies/taxes/aggregates-levy/>

Scottish Government (2020). *Evidence Review and Illustrative Policy Options for a Scottish Aggregates Levy.* [Online]. Available from: <https://www.gov.scot/binaries/content/documents/govscot/publications/research-and-analysis/2020/08/evidence-review-illustrative-policy-options-scottish-aggregates-levy/documents/evidence-review-illustrative-policy-options-scottish-aggregates-levy-final-report/evidence-review-illustrative-policy-options-scottish-aggregates-levy-final-report/govscot%3Adocument/evidence-review-illustrative-policy-options-scottish-aggregates-levy-final-report.pdf>

Scottish Government (2022). *Planning Performance Statistics 2021-22*. [Online] Available from: <https://www.gov.scot/binaries/content/documents/govscot/publications/statistics/2022/08/planning-performance-statistics-2021-2022-annual-and-quarterly-october-2021-to-march-2022/documents/annual-planning-performance-statistics-2021-22/annual-planning-performance-statistics-2021-22/govscot%3Adocument/annual-planning-performance-statistics-2021-22.pdf>

Scottish Government (2023). *National Planning Framework 4 (Revised Draft).* [Online] Available from: <https://www.gov.scot/binaries/content/documents/govscot/publications/strategy-plan/2023/02/national-planning-framework-4/documents/national-planning-framework-4-revised-draft/national-planning-framework-4-revised-draft/govscot%3Adocument/national-planning-framework-4.pdf>

Scottish Government (2023). *National Planning Framework 4 – Draft Housing Land Requirement Explanatory Report.* [Online] Available from: <https://www.gov.scot/publications/scotland-2045-fourth-national-planning-framework-draft-housing-land-requirement-explanatory-report/>

UK Authority (2022). *NatureScot To Use AI In Land Management And Planning.* <https://www.ukauthority.com/articles/naturescot-to-use-ai-in-land-management-and-planning/>

1. <https://www2.bgs.ac.uk/mineralsuk/planning/resource.html#scotland> [↑](#footnote-ref-1)
2. <https://www.bgs.ac.uk/datasets/britpits/> [↑](#footnote-ref-2)