Project Methods Reflection – Interpretable and Reliable Machine Learning for Epidemic Forecasting

In developing my project on interpretable and reliable machine learning for epidemic forecasting, it is essential to align the research methods and data collection approach with both the technical and practical aspects of the topic. The nature of the work combines computational modelling with the need for transparency and usability in public health settings.

Research Methodology Choice

Based on this week's reading, I believe a mixed methods approach is most appropriate. As defined by Creswell (2022), mixed methods combine quantitative and qualitative strategies to offer a more comprehensive understanding of complex research problems. My project will involve:

- Developing and evaluating forecasting models using historical epidemic data (quantitative).
- Collecting expert feedback on the interpretability and usefulness of the model outputs (qualitative).

This combination suits the dual aims of building a technical artefact and ensuring its real-world usability.

Suitable Research Methods

Drawing from Dawson (2015) and the British Research Methodology (2018), the following methods seem best aligned with my aims:

- Design Science Research (DSR): Appropriate for developing new artefacts, such as interpretable machine learning models. It emphasises iterative refinement, evaluation, and solving real-world problems (Dawson, 2015).
- Case Study Elements: Although I will not use a full case study design, I may include a small, applied example to demonstrate the model's use in a specific context (e.g., influenza in the UK). As noted by Saunders, Lewis and Thornhill (2012), this can help ground abstract methods in a realistic scenario.

- CRISP-DM Framework: The data mining lifecycle outlined by IBM (2021) provides a useful structure, particularly for projects that involve large datasets and model building.

Data Collection Methods

I plan to use both secondary quantitative data and primary qualitative data.

- Secondary quantitative data will include historical epidemic datasets, such as COVID-19 or influenza case numbers and hospital admissions. These will come from national and international health sources such as the NHS or WHO. This data will be used for model training, validation and performance testing.
- Primary qualitative data will be collected through short interviews or structured questionnaires with public health professionals or data scientists. Their feedback will help assess whether the model outputs are interpretable, relevant and practical in real-world decision-making.

Combining both types of data will allow me to evaluate not only the accuracy of the forecasting model but also its trustworthiness and usability, which are key to developing responsible AI (Wu, Zhao and Fils-Aimé, 2022).

Skills Required

To complete this project, I will need to strengthen several technical and research-related skills.

- Experience with machine learning and time series forecasting, particularly using Python and libraries such as Scikit-learn, XGBoost and SHAP to ensure explainability.
- Strong data wrangling and evaluation skills, including the ability to clean data, apply validation metrics, use cross-validation and compare model performance.
- Ability to design and apply mixed methods, construct ethical and unbiased questionnaires, and analyse qualitative feedback (Hurst, 2023)
- Knowledge of interpretability frameworks such as feature attribution techniques and rule-based explanations will be essential for ensuring that the models can be understood and trusted by public health users.

Conclusion

In summary, this project will benefit from a structured research design grounded in design science and supported by the CRISP-DM lifecycle. A mixed methods approach allows for both technical depth and real-world evaluation, supporting the dual goals of accuracy and interpretability. The readings from this week have helped clarify how to combine methods effectively and highlighted the need for strong practical and research skills to carry out a project that is both rigorous and relevant to public health contexts.

References:

British Research Methodology (2018) *BRM Module 3: Research Design*. Available at https://www.scribd.com/document/128347797/BRM-Module-3 (Accessed 16 May 2025)

Creswell, J. W. (2022) *What is Mixed Methods Research?* University of Michigan. Video. Available at https://www.youtube.com/watch?v=I5fUxGX2LJk&t=17s (Accessed 16 May 2025)

Dawson, C.W. (2015). *Projects in Computing and Information Systems: A Student's Guide*. 2nd edn. Harlow: Pearson Education. Chapter 2, Section 2.3, pp. 26–27.

Hurst, A. (2023) *Introduction to Qualitative Research Methods*. Oregon State University. Chapter 15, 'Mixed Methods. Available at https://open.oregonstate.education/qualresearchmethods/front-matter/preface/ (Accessed 16 May 2025)

IBM (2021) *CRISP-DM Help Overview*. Available at: https://www.ibm.com/docs/en/spss-modeler/saas?topic=dm-crisp-help-overview (Accessed 16 May 2025)

Saunders, M., Lewis, P. and Thornhill, A. (2012) *Research Methods for Business Students*. 6th edn. Harlow: Pearson Education Limited. Chapter 5, pp. 172–220.

Wu, M.-J., Zhao, K. and Fils-Aimé, F. (2022) *Response rates of online surveys in published research: a meta-analysis*, *Heliyon*, 8(8), e10409. Available at https://www.sciencedirect.com/science/article/pii/S2451958822000409 (Accessed 16 May 2025)