Unit 2 - Literature Review Outline

1. What is the focus and aim of your review? Who is your audience?

The review investigates how machine learning (ML) techniques can be applied to predict health insurance fraud, focusing on the comparative strengths and weaknesses of supervised, unsupervised, and ensemble learning methods. The primary aim is to provide health insurance companies and fraud prevention professionals with a comprehensive understanding of ML's potential in combating fraud. The audience includes academics, data scientists, and practitioners in fraud detection.

2. Why is there a need for your review? Why is it significant?

Health insurance fraud is an escalating global problem, costing billions annually despite detection efforts (Insurance Europe, 2024; AARP, 2023). As technological advancements in ML offer new tools for fraud prevention, this review is significant in guiding stakeholders toward more efficient and effective fraud detection solutions (Guo, 2024; Sharma, 2024).

3. What is the context of the topic or issue? What perspective do you take? What framework do you use to synthesise the literature?

The review is situated within the growing context of applying advanced ML techniques to financial fraud detection. It adopts a practical and critical

perspective, evaluating methods for real-world applicability. The framework synthesises literature by categorising approaches into supervised, unsupervised, and ensemble methods to highlight common themes and gaps.

4. How did you locate and select sources for inclusion in the review?

Sources were selected through broad Google Scholar searches, backward searches from key reviews, and inclusion criteria focusing on studies published from 2016 onwards, with a preference for research from the past five years.

5. How is your review structured?

The review is structured into an introduction outlining the problem and research aim, followed by a critical review of literature categorised into supervised, unsupervised, and ensemble methods, concluding with a synthesis of findings and research gaps.

6. What are the main findings in the literature on this topic?

Supervised models show strong performance but require large datasets and are prone to overfitting (Hancock and Khoshgoftaar, 2021; Nabrawi and Alanazi, 2023). Unsupervised methods excel at anomaly detection but struggle with high-dimensional data (Massi et al., 2020; Zhang et al., 2020). Ensemble methods

improve accuracy by combining models but increase computational complexity and reduce interpretability (Kunickaite et al., 2020; Chaurasiya and Jain, 2025).

7. What are the main strengths and limitations of this literature?

Strengths include the diversity of approaches and promising performance across techniques. Limitations involve class imbalances, inconsistent definitions of fraud, lack of standardised benchmarks, and limited research on real-world deployment and explainability (Karangara et al., 2024; Seshagiri and Prema, 2025).

8. Are there any discrepancies in this literature?

The literature reveals discrepancies, particularly in the comparative effectiveness of algorithms. Some studies report conflicting findings on computational efficiency (Gupta et al., 2021; Hancock and Khoshgoftaar, 2021) and model performance depending on the dataset used.

9. What conclusions do your draw from the review? What do you argue needs to be done as an outcome of the review?

While ML holds promise, systematic comparative studies using uniform datasets are needed. Future research should prioritize real-world deployment, enhance

explainability, and explore ensemble methods further to bridge the gap between academic insights and practical applications (Wang et al., 2025).

List of References:

American Association of Retired Persons (2023) Medicare Fraud. Available at: https://www.aarp.org/money/scams-fraud/medicare/ (Accessed 22 February 2025)

Chaurasiya, R. and Jain, K. (2025) 'Healthcare Fraud Detection Using Machine Learning Ensemble Methods', *South Eastern European Journal of Public Health* 26(1), pp. 4789-4795. Available at: https://seejph.com/index.php/seejph/article/view/4988

Guo, Y. (2024) 'Application of Machine Learning in Insurance Fraud Detection: Achievements and Future Prospects', *Proceedings of the 2024 International Conference on Artificial Intelligence and Communication*. Available at: https://www.atlantis-press.com/proceedings/icaic-24/126003412

Gupta, R. Y., Mudigonda, S. S. and Baruah, P. K. (2021) 'A Comparative Study of Using Various Machine Learning and Deep Learning-Based Fraud Detection Models For Universal Health Coverage', *International Journal of Engineering*

Trends and Technology 69(3), pp. 96-102. Available at: https://ijettjournal.org/archive/ijett-v69i3p216

Hancock, J. T. and Khoshgoftaar, T. M. (2021) 'Gradient Boosted Decision Tree Algorithms for Medicare Fraud Detection', *SN Computer Science* 2. Available at: https://link.springer.com/article/10.1007/s42979-021-00655-z

Insurance Europe (2024) 'Insurance Europe Annual Report 2023 - 2024'. Available from: https://insurancefraud.org/fraud-stats/ (Accessed 22 February 2025)

Karangara, R., Devineni, S. K. and Challa, N. (2024) 'Enhancing Explainability in AI Fraud Detection', *International Journal of Computer Techniques* 11(1). Available at: https://www.researchgate.net/publication/377329151_Enhancing_Explainability_in_AI_Fraud_Detection

Kunickaite, R., Zdanaviciute, M. and Krilavičius, T. (2020) 'Fraud Detection in Health Insurance Using Ensemble Learning Methods', *International Conference on Information Technology*. Available at: https://www.semanticscholar.org/paper/Fraud-Detection-in-Health-Insurance-

Using-Ensemble-Kunickaite-Zdanaviciute/ac028d54c9f5bedf8d22dc1e2cb41d4d4095ec06

Massi, M. C., Ieva, F. and Lettieri, E. (2020) 'Data mining application to healthcare fraud detection: a two-step unsupervised clustering method for outlier detection with administrative databases', *BMC Medical Informatics and Decision Making* 20. Available at: https://doi.org/10.1186/s12911-020-01143-9

Nabrawi, E. and Alanazi, A. (2023) 'Fraud Detection in Healthcare Insurance Claims Using Machine Learning', *Risks* 11(9). Available at: https://doi.org/10.3390/risks11090160

Seshagiri, S. and Prema, K. V. (2025) 'Efficient Handling of Data Imbalance in Health Insurance Fraud Detection Using Meta-Reinforcement Learning', *IEEE Access* 13, pp. 23482-23497. Available at: https://ieeexplore.ieee.org/document/10858064

Sharma, A. (2024) 'Predictive Accuracy of Machine Learning Models in Fraud Detection for Health Insurance in India', *American Journal of Statistics and Actuarial Science* 5(2), pp. 1-12. Available at: https://ajpojournals.org/journals/index.php/AJSAS/article/view/2253

Wang, Z., Chen, X., Wu, Y., Jiang, L., Lin, S. and Qiu, G. (2025) ‘A robust and interpretable ensemble machine learning model for predicting healthcare insurance fraud’, *Scientific Reports* 15(1). Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC11697271/

Zhang, C., Xiao, X. and Wu, C. (2020) ‘Medical Fraud and Abuse Detection System Based on Machine Learning’, *International Journal of Environmental Research and Public Health* 17(19). Available at: https://doi.org/10.3390/ijerph17197265