

Predictive Modelling and Analysis of Biologics Drug Usage in Secondary Healthcare

By Thomas Sellwood (2215661) BSc Computing (Data and Information Systems)
2215661@student.uwtsd.ac.uk



Prifysgol Cymru
Y Drindod Dewi Sant
University of Wales
Trinity Saint David

1. Background

Biologic medicines have transformed the treatment of chronic inflammatory diseases but represent a significant and growing cost to NHS secondary care [2]. Current forecasting approaches are predominantly retrospective, limiting the ability of pharmacy teams to proactively manage stock, procurement and financial planning. Machine learning offers the potential to identify complex patterns in historical medicines usage, enabling more accurate and interpretable forecasting to support data-driven decision making [6][27].

2. Aim and Objectives

Aim:
To analyse the drivers and implications of biologic medicine usage in secondary care healthcare settings and to develop predictive models to forecast future usage.
Objectives:

- Review existing literature on biologics usage forecasting.
- Prepare and explore historical NHS pharmacy issues data.
- Develop and validate XGBoost [27] forecasting models.
- Apply SHAP [29] to explain model predictions.
- Evaluate the framework for medicines management and financial planning.

3. Methodology

Data Collection & Preparation

Historical dermatology biologics issues data were extracted from the Welsh Hospital Pharmacy Stock Management System (WHPSMS). Data were cleaned, transformed and aggregated into monthly observations for forecasting.

Feature Engineering

Lag variables, rolling statistics and temporal features were created to capture historical utilisation patterns across **1-, 3- and 6-month** forecasting horizons.

Model Development

XGBoost [27] was used to develop forecasting models using **TimeSeriesSplit** cross-validation and hyperparameter optimisation.

Model Evaluation

Performance was assessed using **MAE, RMSE, MAPE, WMAPE, SMAPE and R²**, and compared with a naïve forecasting benchmark using the **Diebold–Mariano test [55]**.

Model Explainability

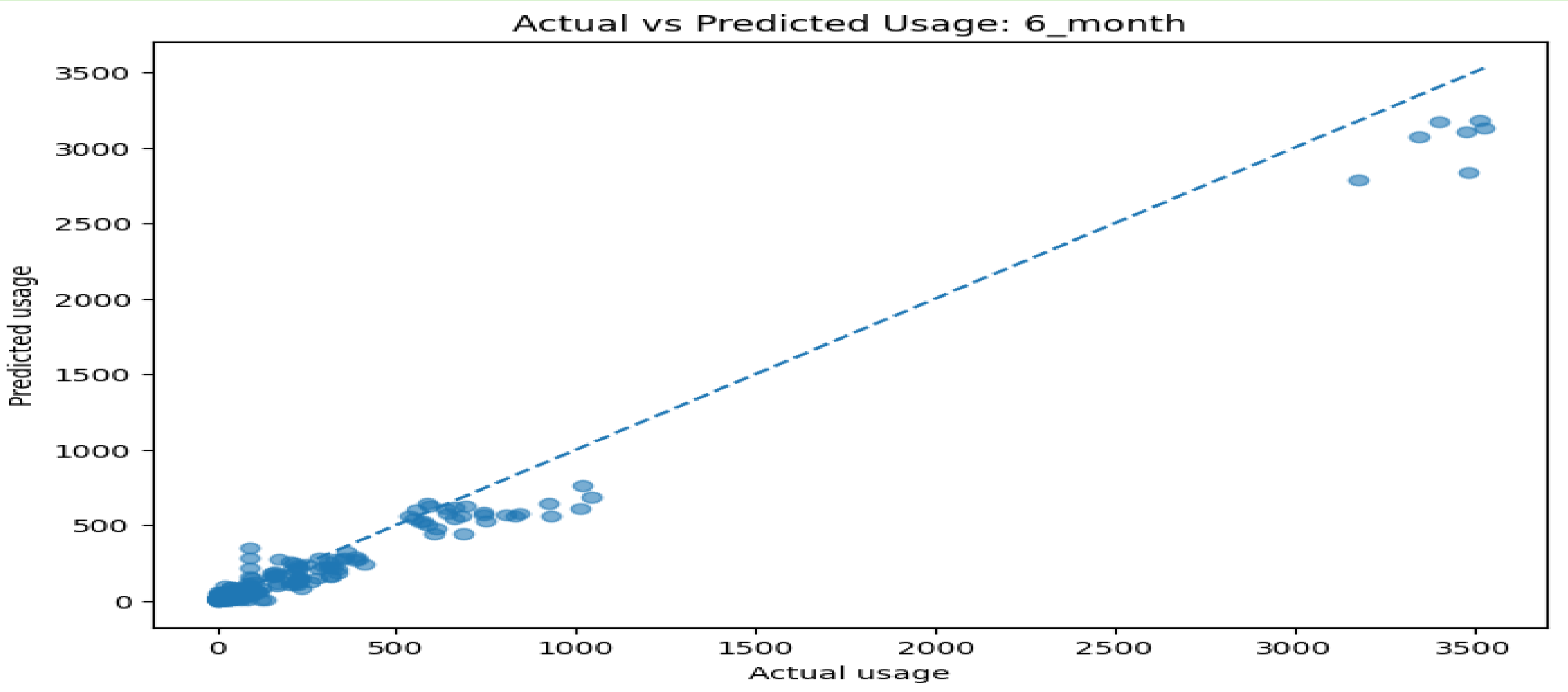
SHAP [29] was applied to explain model predictions and identify the features driving biologics usage forecasts.

4.Results

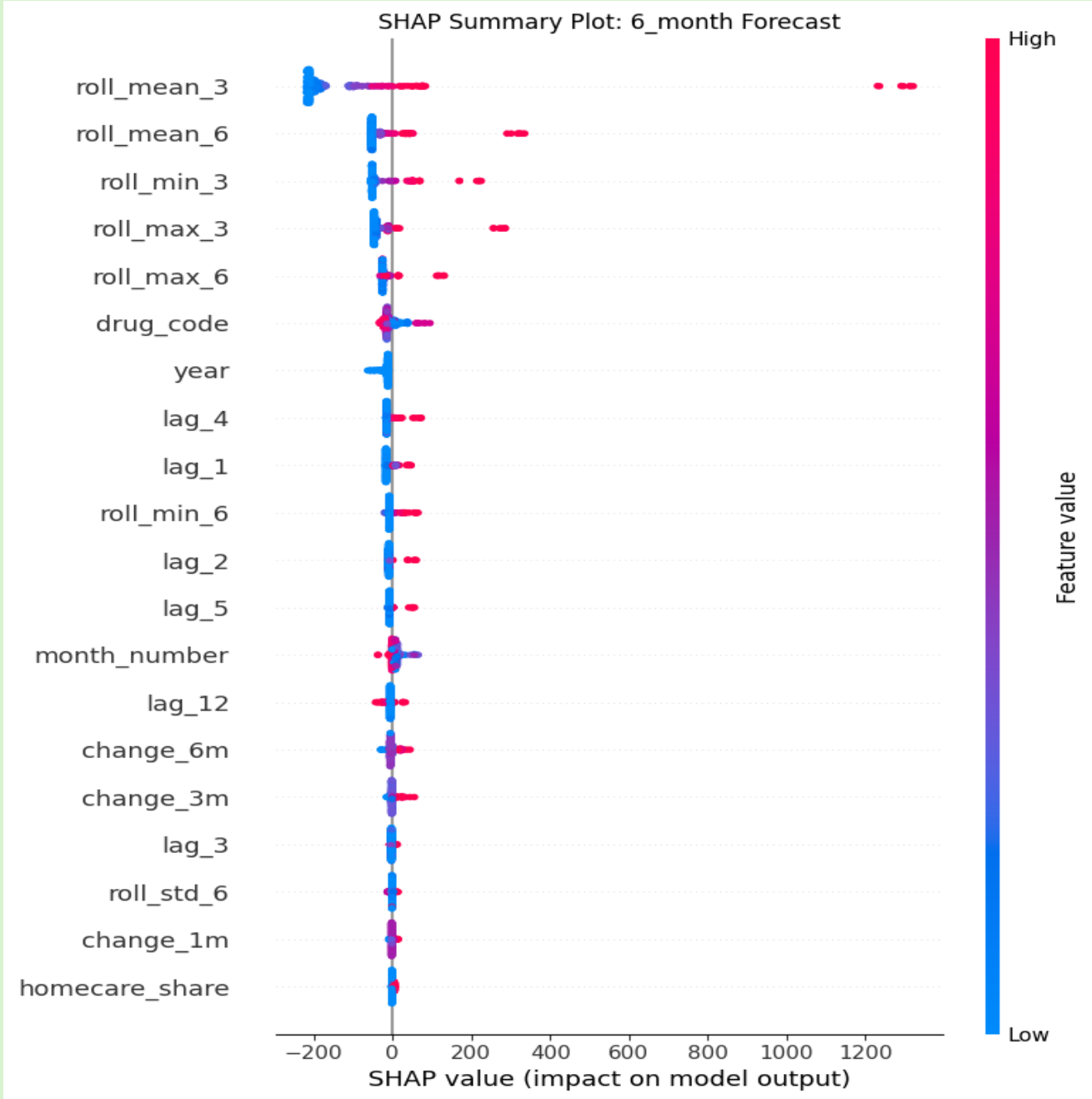
The forecasting framework was evaluated using predictive performance metrics, forecast visualisation, SHAP explainability and Diebold–Mariano statistical testing.

	MAE	RMSE	MAPE	WMAPE	SMAPE	R2
1_month	11.263416	29.872100	68.103716	39.043255	120.363141	0.871214
3_month	20.190313	51.039271	58.469541	22.710281	98.674976	0.957831
6_month	41.336035	87.083867	57.803144	23.829332	94.218482	0.969298

XGBoost [27] achieved strong predictive performance across all forecasting horizons, with the 6-month model demonstrating the lowest forecast error.



Predicted usage closely followed observed medicines utilisation, demonstrating strong agreement between forecasts and actual values.



SHAP [29] analysis identified lag variables and rolling statistics as the dominant contributors to model predictions.

5. Conclusion

This research demonstrated that machine learning can be successfully applied to forecast biologics medicines usage within NHS secondary healthcare. XGBoost [27] produced accurate forecasts across the 1-, 3- and 6-month prediction horizons, while SHAP [29] improved model transparency by identifying the historical utilisation features that most influenced predictions. The forecasting framework provides an interpretable and data-driven approach that has the potential to support medicines management, procurement planning and financial decision-making within the NHS.

6. Future Work

- Validate the framework using data from multiple NHS Health Boards.
- Investigate additional machine learning algorithms (e.g. LightGBM and CatBoost).
- Incorporate external variables such as patient demographic data, seasonality and medicine shortages.
- Develop an automated forecasting pipeline with Power BI integration.
- Evaluate implementation within routine NHS pharmacy practice.

7. References

[2] A. Angelis, J. Lomas, B. Woods and H. Naci, "Promoting Population Health through Pharmaceutical Policy," 2023. [Online]. Available: <https://www.lse.ac.uk/lse-health/assets/documents/Reports/23-0275-Pharma-Report-V10.pdf>

[6] K. P. Fourkiotis and A. Tsadiras, "Applying Machine Learning and Statistical Forecasting Methods for Enhancing Pharmaceutical Sales Predictions," vol. 6, no. 1, pp.186, 2024, doi: 10.3390/forecast6010010.

[27] C. Tianqi and Carols Guestrin, "XGBoost | Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining," 13/08 2026, [Online]. Available: <https://dl.acm.org/doi/10.1145/2939672.2939785>

[29] S. M. Lundberg and S. Lee, "A unified approach to interpreting model predictions", in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, California, USA, pp. 4768–4777.

[55] F. X. Diebold, "Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold-Mariano Tests," vol. 33, no. 1, pp.1, 2015, doi: 10.1080/07350015.2014.983236.