



Project Report

Title: Multivariate Forecasting of Scotland's Monthly Birth Rates

Company: ITSOLERA PRIVATE LIMITED

Team Members: Muhammad Shoaib, Maryam, Umair

1. Introduction

This report presents the final deliverables and key learnings from our project titled *"Multivariate Forecasting of Scotland's Monthly Birth Rates."* The project was executed under the umbrella of ITSOLERA PRIVATE LIMITED by team members Muhammad Shoaib, Maryam, and Umair. The goal was to build a robust and accurate forecasting framework that would enable policymakers to make data-driven decisions related to public health, education, and social planning.

2. Project Objectives

- Develop a multivariate forecasting pipeline using economic, health, and educational indicators.
 - Improve accuracy over traditional univariate birth forecasting methods.
 - Create an explainable forecasting system for use by stakeholders and decision-makers.
-

3. Achievements

Phase 1: Data Integration & Preprocessing

- Collected monthly data from Jan 1998 to Dec 2022 for:
 - Birth registrations (NRScotland.gov.uk)
 - Unemployment rate (ONS)
 - Inflation (CPI - ONS)
 - Maternal health indicators (Public Health Scotland)

- Female tertiary enrollment (HESA)
 - UK Government holidays
- Cleaned and merged datasets on a common monthly Date index.
- Interpolated annual series (e.g., education data) to monthly granularity.
- Imputed missing values and handled outliers.
- Performed EDA including correlation heatmaps, seasonal decomposition, and trend analysis.

Phase 2: Feature Engineering

- Generated lagged features (1-, 3-, 6-month lags) for all predictors.
- Computed rolling aggregates (3-month moving averages).
- Created calendar effect encodings such as month-of-year dummies and holiday flags.

Phase 3: Model Development

- Baseline model: Vector Auto Regression (VAR).
- Machine Learning models:
 - XGBoost with lag features.
 - Random Forest with Recursive Feature Elimination (RFE).
- Deep Learning models:
 - Long Short-Term Memory (LSTM) networks.
 - Temporal Convolutional Networks (TCNs).

Phase 4: Training & Validation

- Used rolling-window approach:
 - Training: 1998–2018
 - Validation: 2019–2020
 - Testing: 2021–2022
- Evaluation Metrics:
 - MAE, RMSE, SMAPE
- Performed hyperparameter tuning using Optuna.

Phase 5: Explainability & Interpretability

- Applied SHAP values to tree-based models to quantify feature importance.
- Used Partial Dependence Plots (PDPs) for Random Forest models.
- Identified shifting drivers of birth rates over time.

Phase 6: Deployment

- Developed an interactive dashboard using Streamlit.
 - Features included real-time forecasts, what-if scenarios, and variable importance analysis.
 - Automated data pipeline for monthly updates and retraining.
-

4. Tools & Technologies Used

- **Programming:** Python (Pandas, NumPy, Scikit-learn, Statsmodels)
 - **Visualization:** Seaborn, Matplotlib, Plotly
 - **ML/DL Frameworks:** XGBoost, LightGBM, TensorFlow, PyTorch
 - **Dashboarding:** Streamlit
 - **Version Control:** GitHub
 - **Environment:** Google Colab, Local Machine
-

5. Final Outcomes

- Demonstrated significant improvement in forecasting accuracy compared to univariate methods.
 - Delivered a live, interactive tool for policy and public-health planners.
 - Provided interpretable models highlighting socio-economic drivers of fertility behavior.
-

6. Conclusion

Through this project, we successfully demonstrated how integrating diverse predictors can improve the accuracy and reliability of monthly birth rate forecasts in Scotland. The combination of statistical, machine learning, and deep learning models enabled us to build a future-ready system that can be used by policymakers to explore various demographic scenarios and plan services more efficiently.

This project stands as a valuable contribution toward data-driven decision-making in public policy and healthcare resource planning.

Team: Muhammad Shoaib, Maryam, Umair

Company: ITSOLERA PRIVATE LIMITED