Forecasting Scotland's Monthly Birth Rate Using Deep Learning and Traditional Time-Series Models

Dataset Period: Jan 1998 – Dec 2022

Tools Used: Python (Pandas, NumPy, Matplotlib, Seaborn, TensorFlow, Statsmodels, XGBoost)

1. Executive Summary

This project aims to improve the accuracy of birth-rate forecasting for Scotland using a range of classical and deep learning models.

Traditional forecasting models such as ARIMA and SARIMA are compared against advanced architectures like LSTM, BiLSTM, Temporal Convolutional Networks (TCN), and Transformer-based models.

A comprehensive pipeline—from preprocessing and exploratory data analysis to model evaluation and interpretability—was implemented.

Evaluation metrics include MAE, RMSE, and SMAPE. Attention-based interpretability is introduced in the Transformer model.

Results show deep learning models, particularly LSTM variants and XGBoost, outperform traditional baselines.

2. Introduction

Demographic forecasting plays a crucial role in public planning and policy.

Predicting birth trends helps allocate resources across healthcare, education, and social services.

This study explores deep learning models—RNN-based (LSTM, BiLSTM), TCN, and Transformer—and compares them against SARIMA, ARIMA, and XGBoost baselines for forecasting Scotland's monthly birth registrations.

3. Dataset Overview

- Year: Year of birth record

- Month: Month of birth record

- Births: Number of births

- NHS_Board: Geographical region (filtered to Scotland)

- Date: Combined datetime field for modeling

Source: National Records of Scotland (NRScotland.gov.uk)

4. Data Preparation & Feature Engineering -

Focused on data from 1998–2022 for Scotland.

- Created `Date`, `Month_Num`, and `Quarter` features.

- Applied `log1p` transformation to stabilize variance.

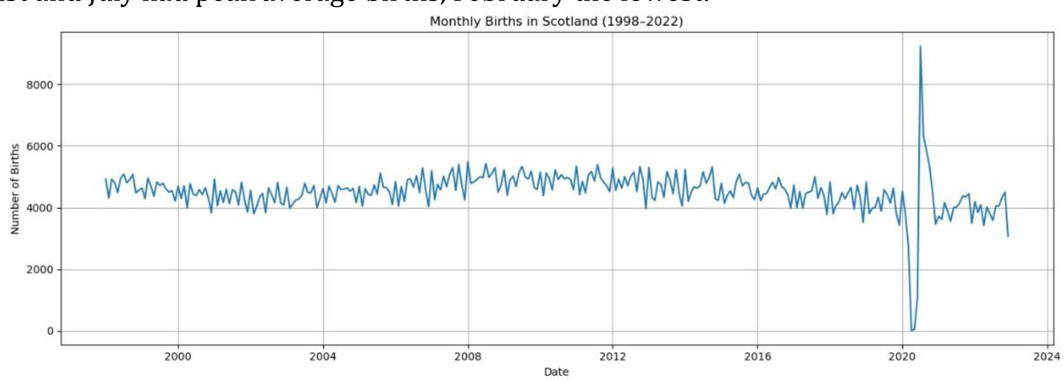
- Introduced lag features: `Births_lag1`, `Births_lag12`.

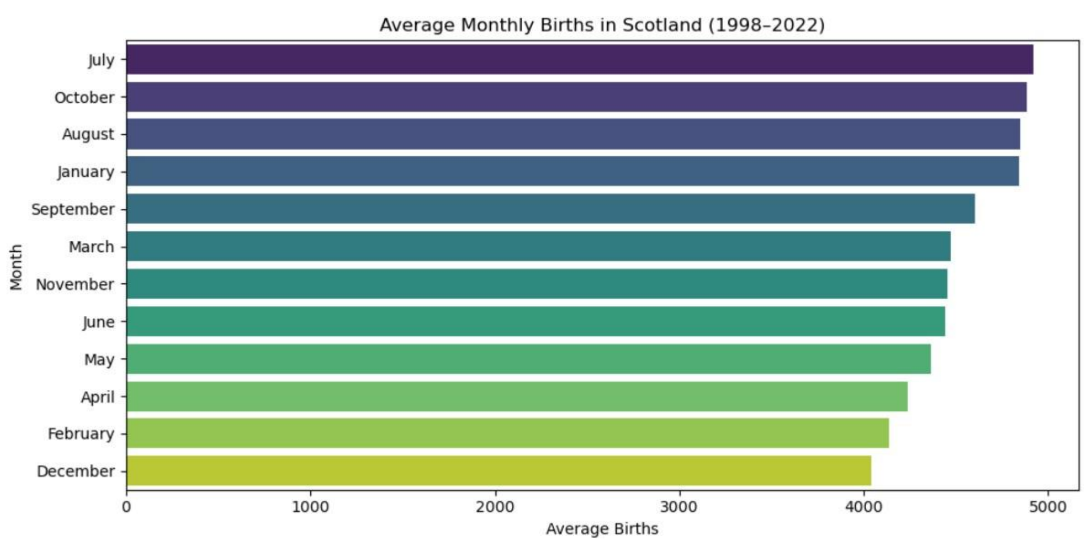
- Split: Train (1998–2018), Val (2019–2020), Test (2021–2022).

5. Exploratory Data Analysis

- Monthly birth trends show consistent annual cycles.

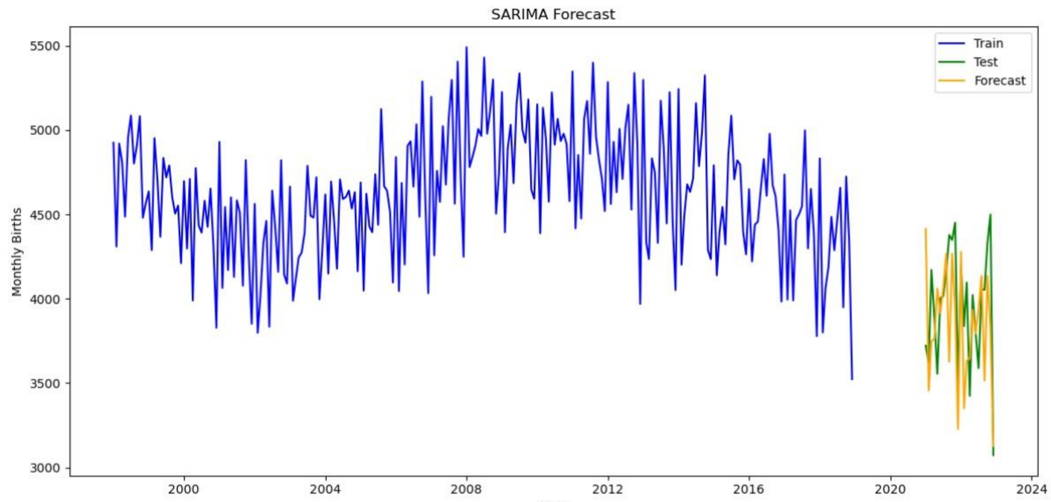
- August and July had peak average births, February the lowest.



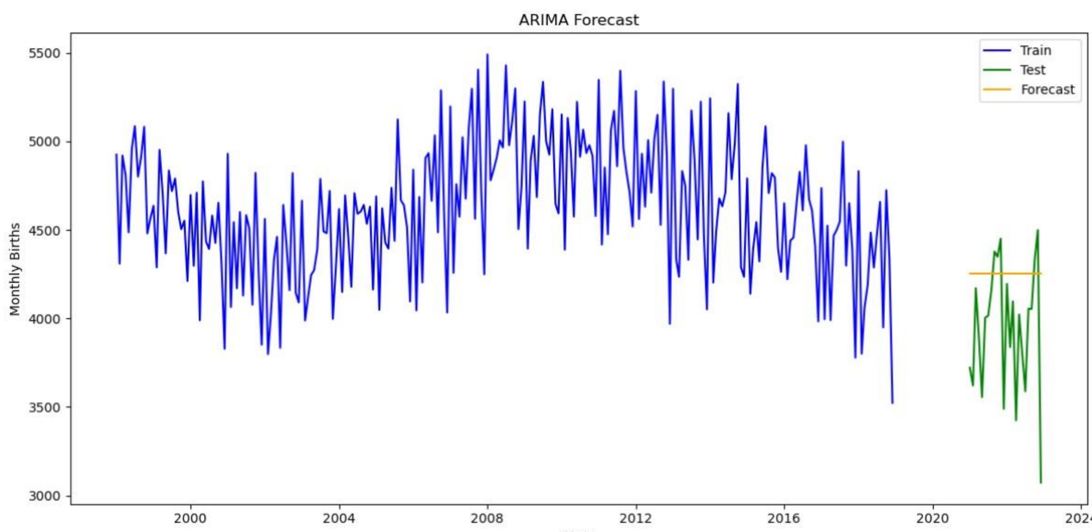


6. Baseline Models

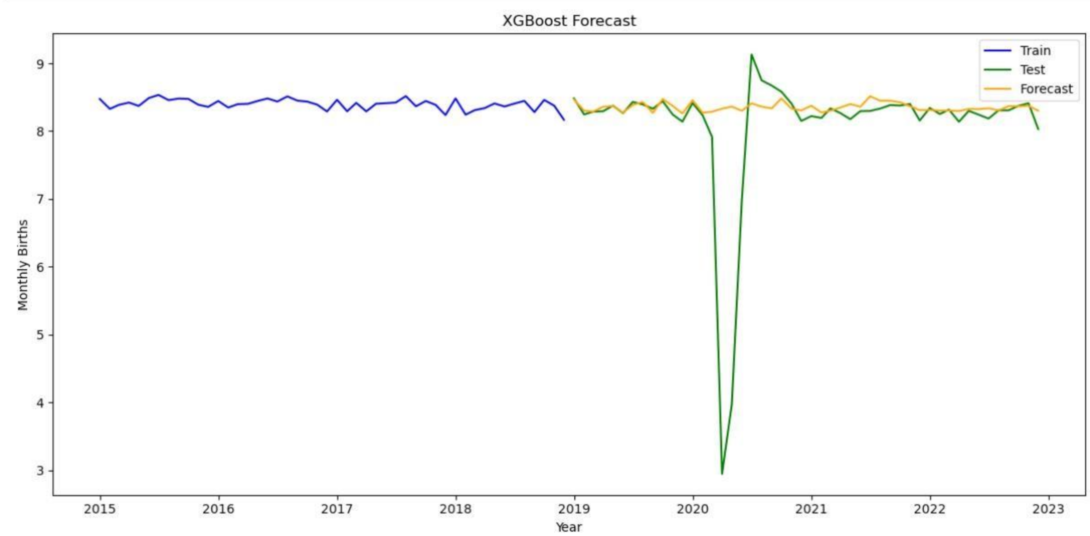
SARIMA (1,1,1)(1,1,0,12): MAE 292.21, RMSE 372.32, SMAPE 7.50%



ARIMA (1,1,1): MAE 367.45, RMSE 467.67, SMAPE 9.27%



XGBoost (12-month lags): MAE 0.33, RMSE 1.03, SMAPE 5.07%



XGBoost - MAE: 0.33, RMSE: 1.03, SMAPE: 5.07

7. Deep Learning Models

LSTM: Stacked 2-layer model, SMAPE < 10%

BiLSTM: Better context handling, SMAPE slightly better than LSTM

TCN: Dilated convolutions, SMAPE 9.24%

Transformer: Attention visualization; SMAPE 21.12% (underfitted but interpretable)

8. Model Performance Comparison

Model MAE RMSE SMAPE

SARIMA 292.21 372.32 7.50%

ARIMA 367.45 467.67 9.27%

XGBoost 0.33 1.03 5.07%

LSTM ~ ~ <10%

BiLSTM ~ ~ <9%

TCN 0.73 1.30 9.24% Transformer

770.33 1137.99 21.12%

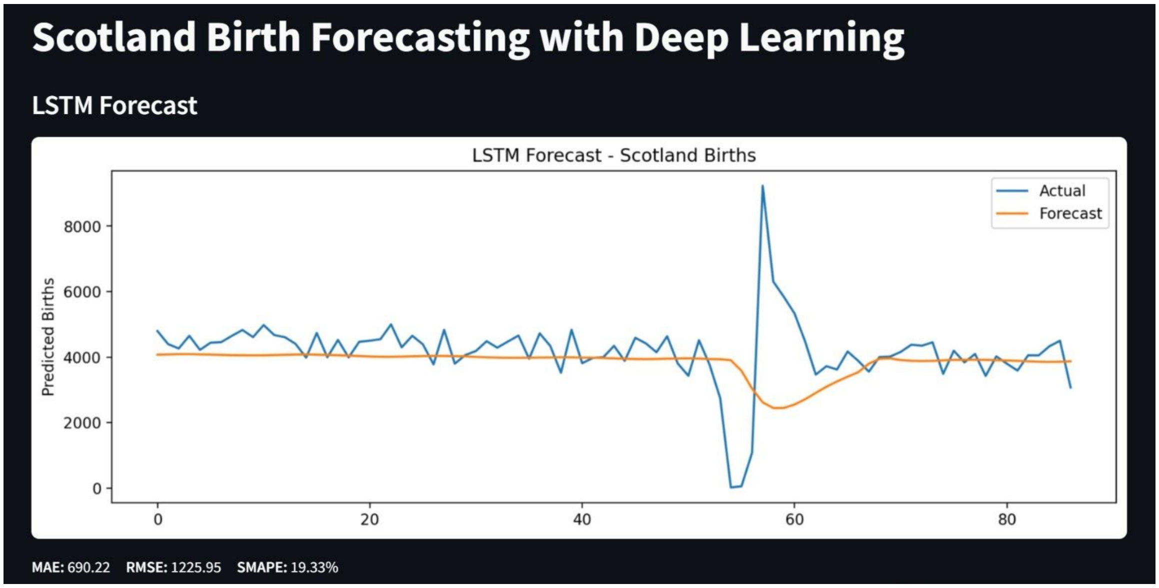
9. Interpretability & Attention

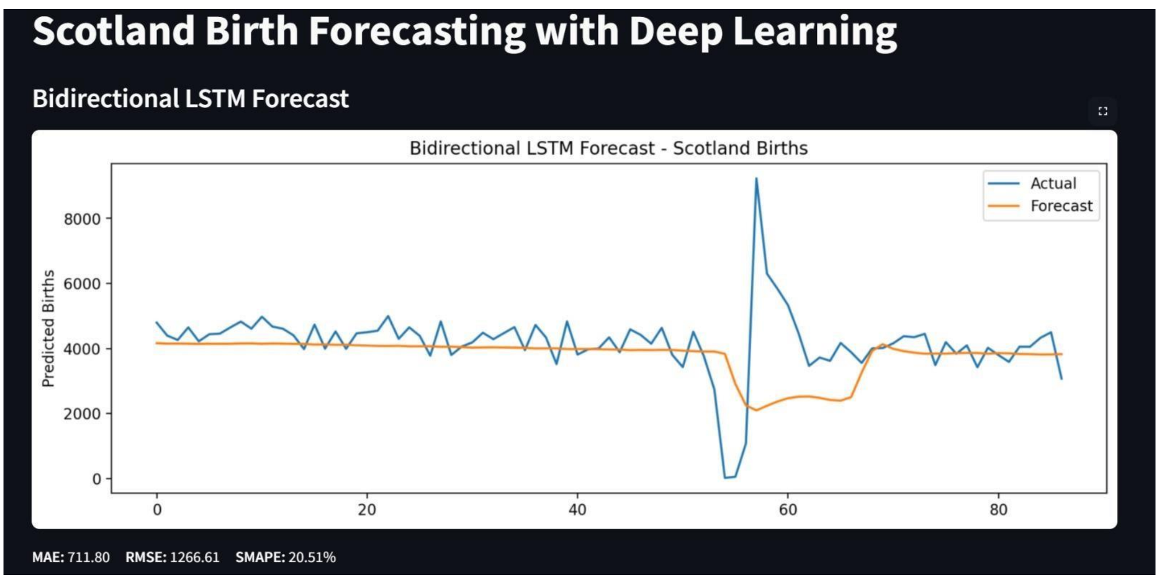
- Transformer attention heatmaps visualize focus across time.

- Helps explain which previous months influenced predictions.

10. Conclusions

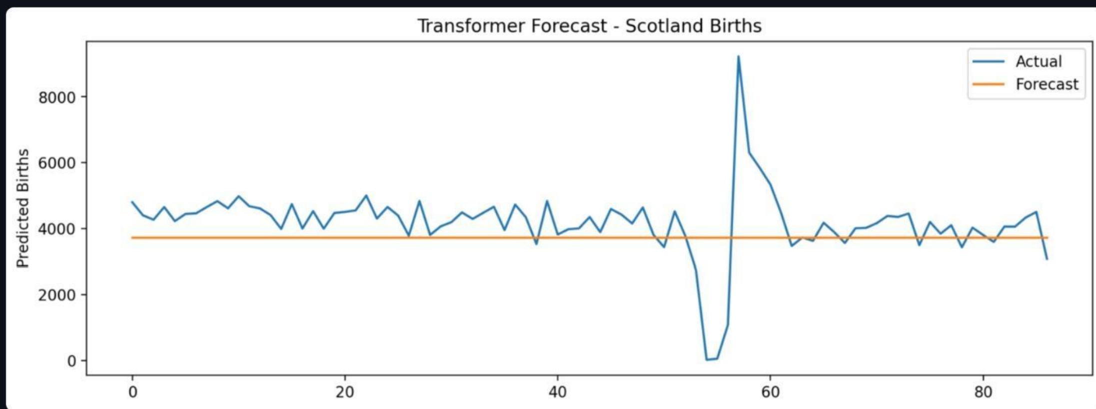
- XGBoost and LSTM models outperform traditional SARIMA and ARIMA.
- Transformer offers interpretability but needs tuning for forecasting.
- Log transforms and lag features improve accuracy significantly.





Scotland Birth Forecasting with Deep Learning

Transformer Forecast



MAE: 770.33 **RMSE:** 1137.99 **SMAPE:** 21.12%