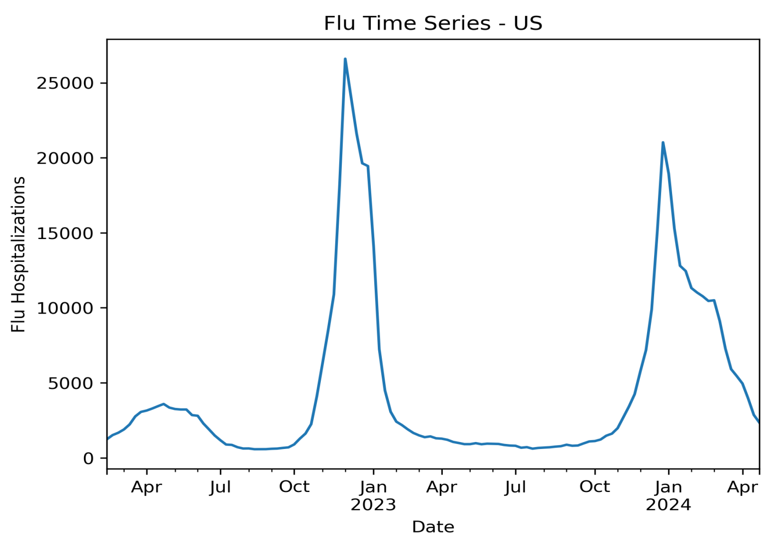
Analysis of Continuously Retrained Models in Time Series Forecasting

Student: Arya Palla
Mentor: Dr. Aniruddha Adiga

Background

- It is traditional machine learning practice to use a single train-test split for a model, essentially training only once before testing the performance on remaining data
- Forecasting may benefit from continuously retrained models that stay up to date
- Data was provided by the CDC's FluSight initiative
- Data is limited and lacks strong seasonality



Purpose

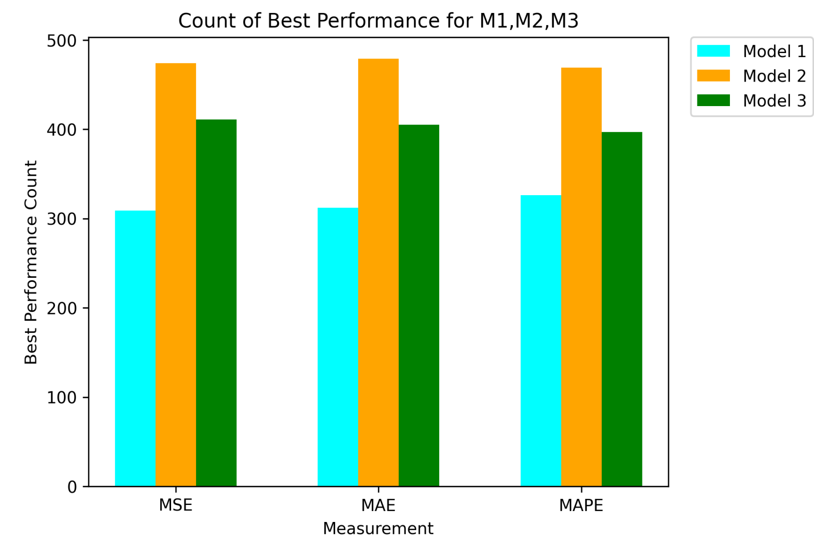
- Explore the variance in retrained vs static machine learning models in flu forecasting
- Understand possible variance in model efficacy across different states and time periods
- Build robust forecasting models relevant for influenza research

Experiments

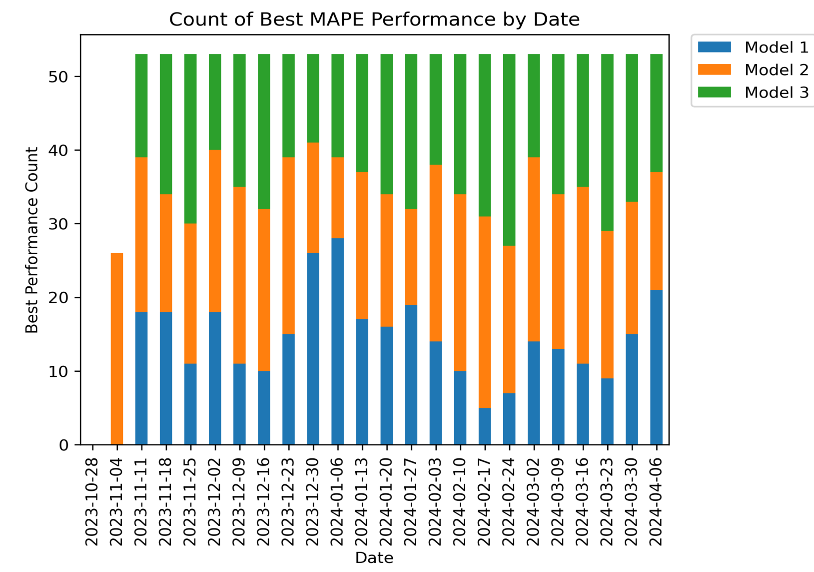
- 3 different models, all LSTM, were set up to forecast flu patterns 4 weeks ahead at every date from the start of flu season
  - Nov 4 - Nov 25, Nov 11 - Dec 2, Nov 18 - Dec 9, etc.
- M1
  - Static model trained only up until October 28, 2023
- M2
  - Continuous model updated weekly, essentially always forecasting the next 4 weeks from the data that the model was trained on
- M3
  - A delayed version of M2, where it is still updated continuously but trained up until a week before the data M2 is trained on
  - When M2 is trained up until Nov 11, M3 is trained only until Nov 4, but both will still forecast from Nov 11 – Dec 2

Results

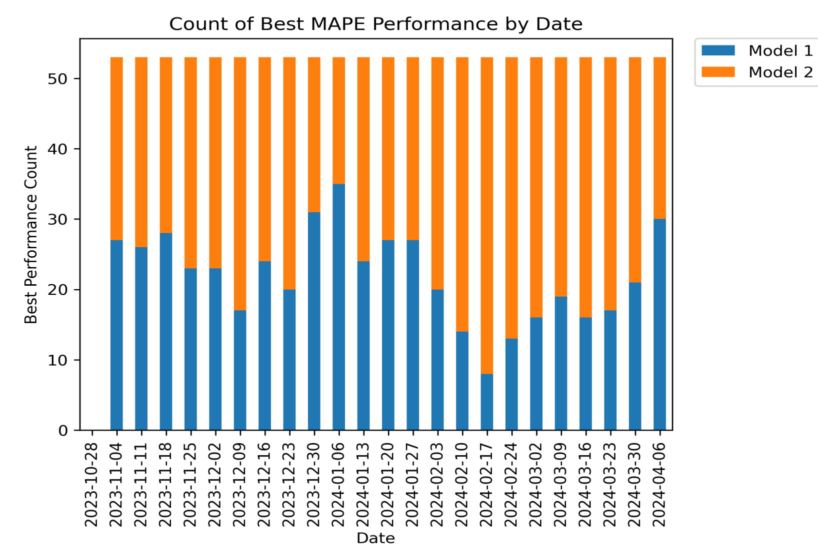
- M2 & M3 outperformed M1 in comparisons across all three of measurements
  - Mean Squared Error (MSE)
  - Mean Absolute Error (MAE)
  - Mean Absolute Percentage Error (MAPE)



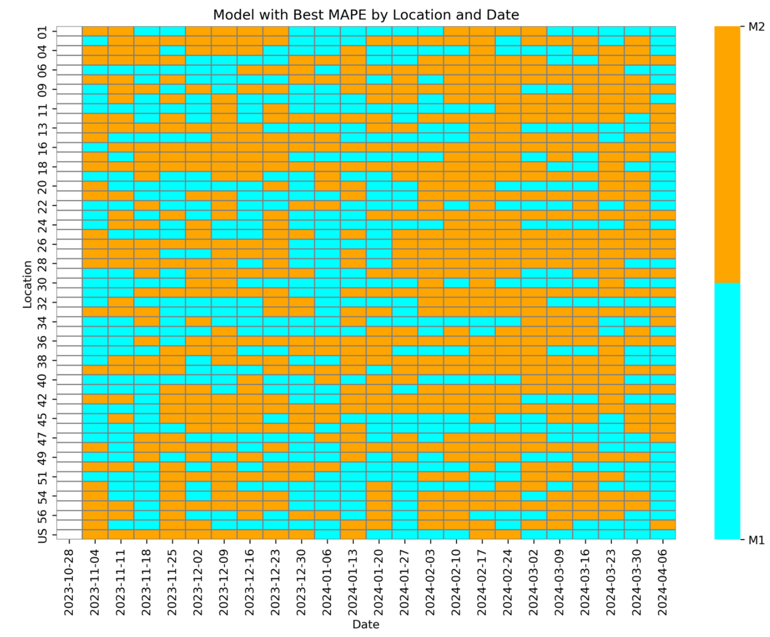
- As season progressed, retrained models outperformed the static model, except during periods of decline or stagnation in hospitalizations where the models performed similarly



- When comparing just M1 & M2, the pattern remains consistent as M2 performs better the longer the season goes on except moments of sharp decline, where again, the models are relatively equal



- States/regions within the US were numerated alphabetically, and when displayed across a heatmap, variance shows in model performance for certain states and certain dates
- For example, M1 outperforms M2 in states like Tennessee or New Mexico, but M2 appears to have a clear advantage in most states from January onwards



Future Work

- Conducting similar experiments on different forecasting models like ARIMA may strengthen the argument for retraining
- Exploring and understanding variance in particular states/dates can yield a more thorough understanding of influenza patterns and future forecasting