

Comparison of Applying Advanced Spatiotemporal Clustering Algorithms with Machine Learning Methods for Forecasting of Urban Mobility Demand: Case of NYC Taxi

Anna Yasyreva

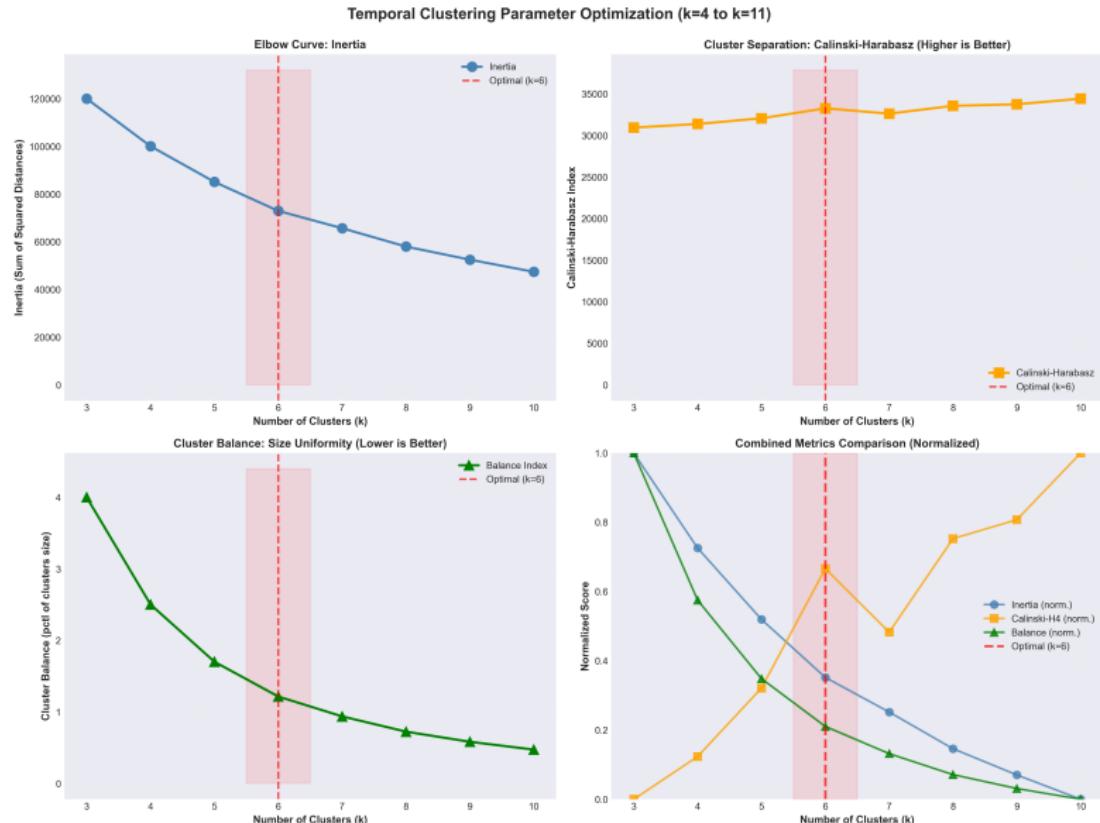
Supervisor: prof. dr. Jurgita Markevičiūtė



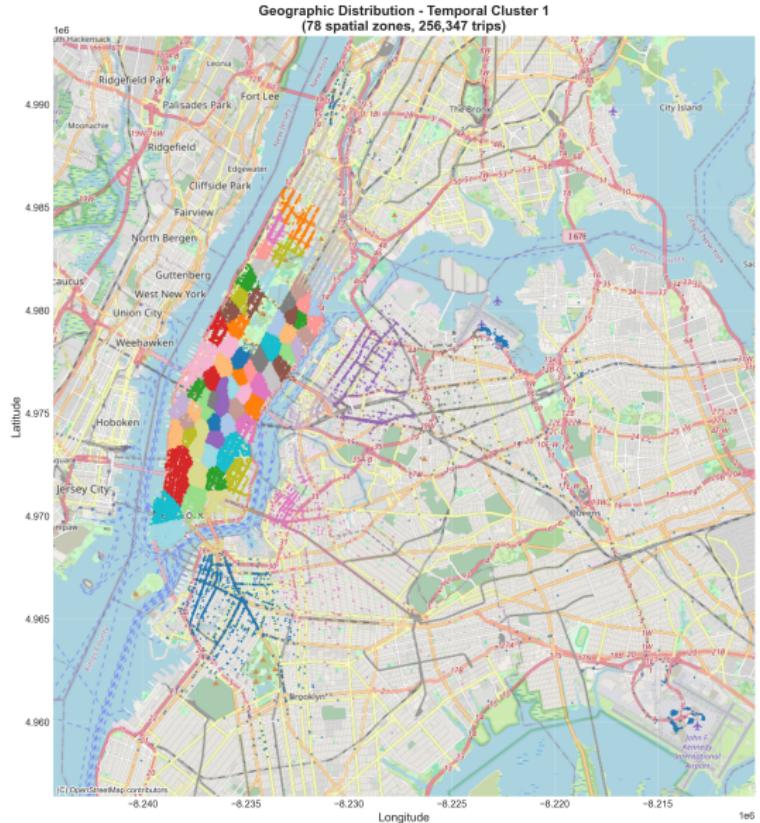
Topic, goal and aims of the work

- Forecasting the traffic demand to help optimize the network
 - To examine whether the application of combined clustering algorithms with different forecasting models can produce actionable insights in urban mobility
 - Applying three different models with different approaches to compare the prediction capabilities - SARIMA, XGBoost and LSTM
 - Additional novelty - to compare the forecasting horizon for each model to evaluate which works best and deteriorates less in which conditions

Clustering algorithms optimization - Temporal



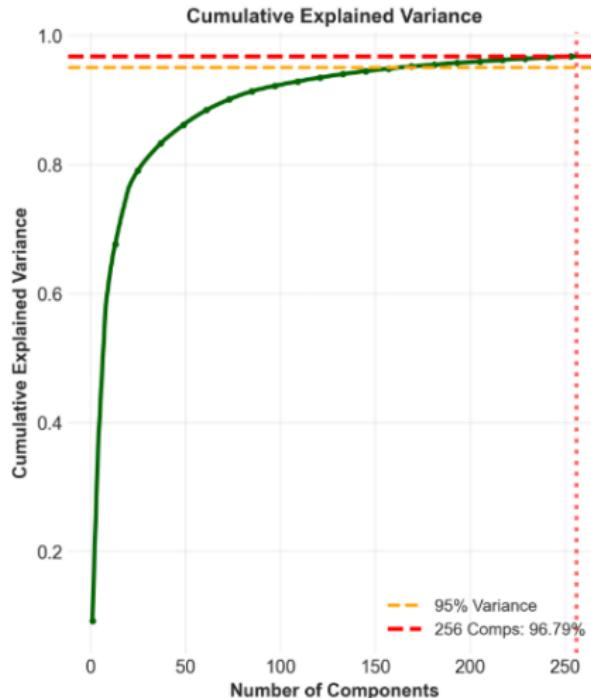
Clustering algorithms optimization - Geographic



Methodology changes

- **Hierarchical Clustering:** Innovative combined clustering - K-means on temporal features; followed by hybrid spatial clustering combining **HDBSCAN** (density-based discovery of natural cluster structures) with K-means refinement - uses HDBSCAN centroids as init points
- **SARIMA** models temporal dependencies within identified clusters, capturing seasonal demand patterns
- **XGBoost** with MultiOutputRegressor simultaneously predicts demand across multiple cluster centroids
- **LSTM** network capture complex non-linear spatiotemporal patterns and multi-step dependencies across clustered regions - required PCA to reduce dimensionality

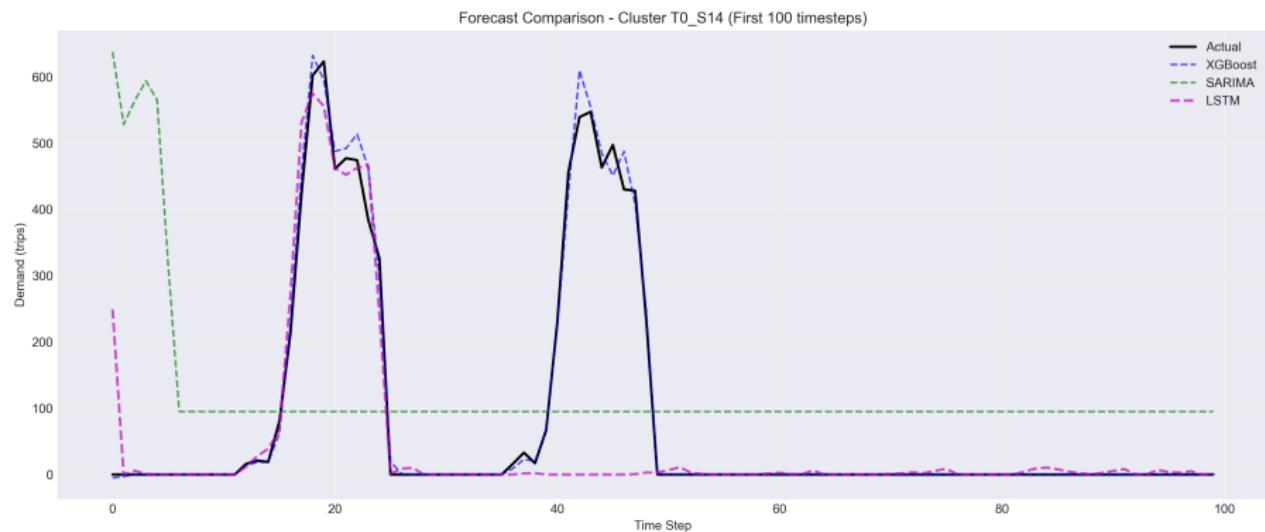
PCA



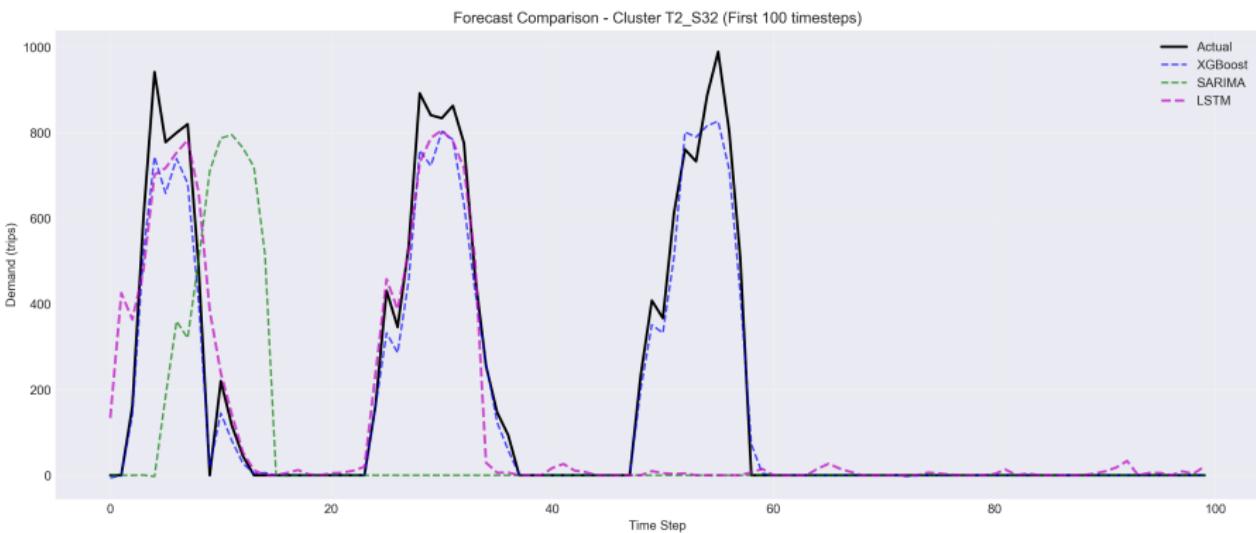
Results

- SARIMA model was not able to model and forecast properly without external regressors
- VAR model, even able to predict up to 10 clusters' demand, is not suitable for such high amount count of clusters due to the curse of dimensionality
- Both XGBoost and LSTM models, when tuned and optimized, proved to be robust and showed good performance in forecasting
- Different models performance differs on 1day, 3days and 7days forecast horizons.

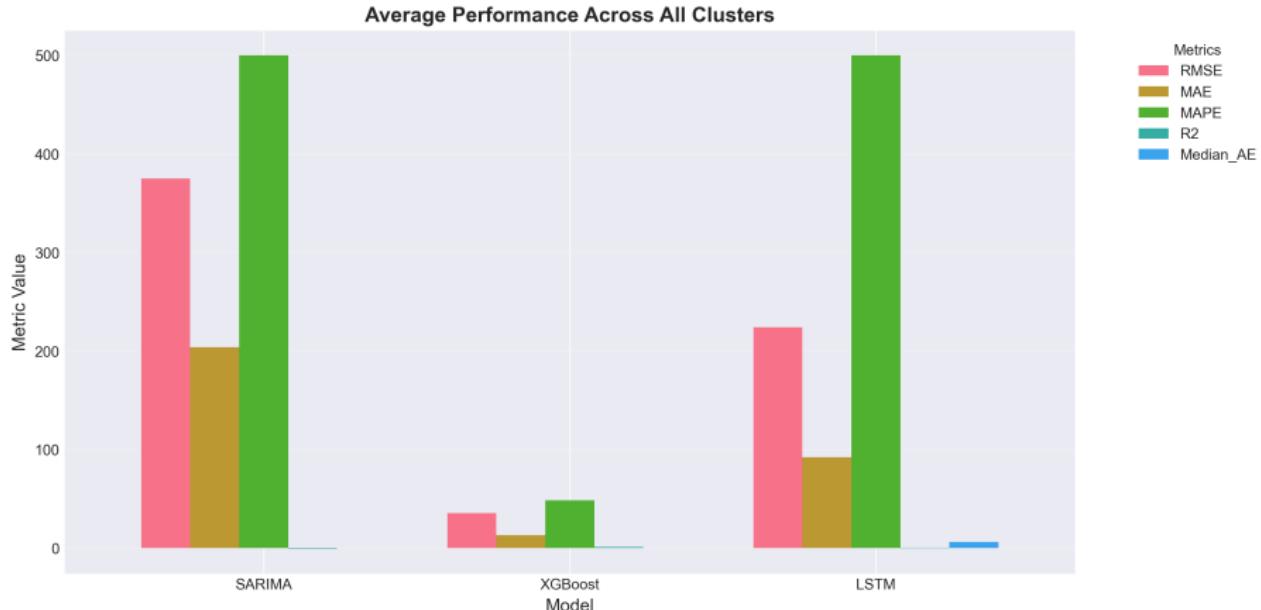
Results - Forecasting



Results - Forecasting



Results - Evaluation



What's left

- Polish Literature review section
- Finish writing Methodology section
- Finish describing Experimental Setup part
- Refine Results and Conclusions section

Thank you for your attention!