The paper "Predicting Time to Pushback of Flights in U.S. Airports" by Daniil Filienko et al. presents a machine learning model for predicting the pushback time of flights. The model, which uses a gradient boosting decision tree approach, is trained on an extensive dataset that includes weather data, airport activity, airline, and aircraft characteristics. The study emphasizes the importance of accurate pushback time predictions for improving overall flight scheduling and management efficiency. The authors participated in the 2023 "Pushback to the Future" competition hosted by NASA and achieved significant results.

The researchers collected data from 10 U.S. airports, comprising over 3.8 million flights, and developed a feature set that captures various relevant factors. They trained both global models, which use combined data from all airports, and local models, tailored to each specific airport. Their findings indicate that local models are more memory efficient and perform comparably to global models in terms of mean absolute error (MAE). The study reveals that integrating multiple sources of information and leveraging feature engineering can significantly enhance prediction accuracy.

In their methodology, the authors employed gradient boosting decision trees, utilizing the LightGBM library for training the models. They compared the performance of their approach against baseline methods and neural network models, finding that their LightGBM models consistently yielded lower MAEs. The study also highlighted the computational efficiency of local models and provided insights into feature importance, with the estimated time of departure and the flight's Globally Unique Flight Identifier being among the most critical features. The paper concludes with suggestions for future work, including the potential use of federated learning to incorporate sensitive airline data while preserving privacy.