Interactive Planning Tool for US Flight Delay Prediction

Jake Cuppels, Kevin Polak, Enes Sert,Yanzhe Yin, Amelia Yurkofsky

Motivation

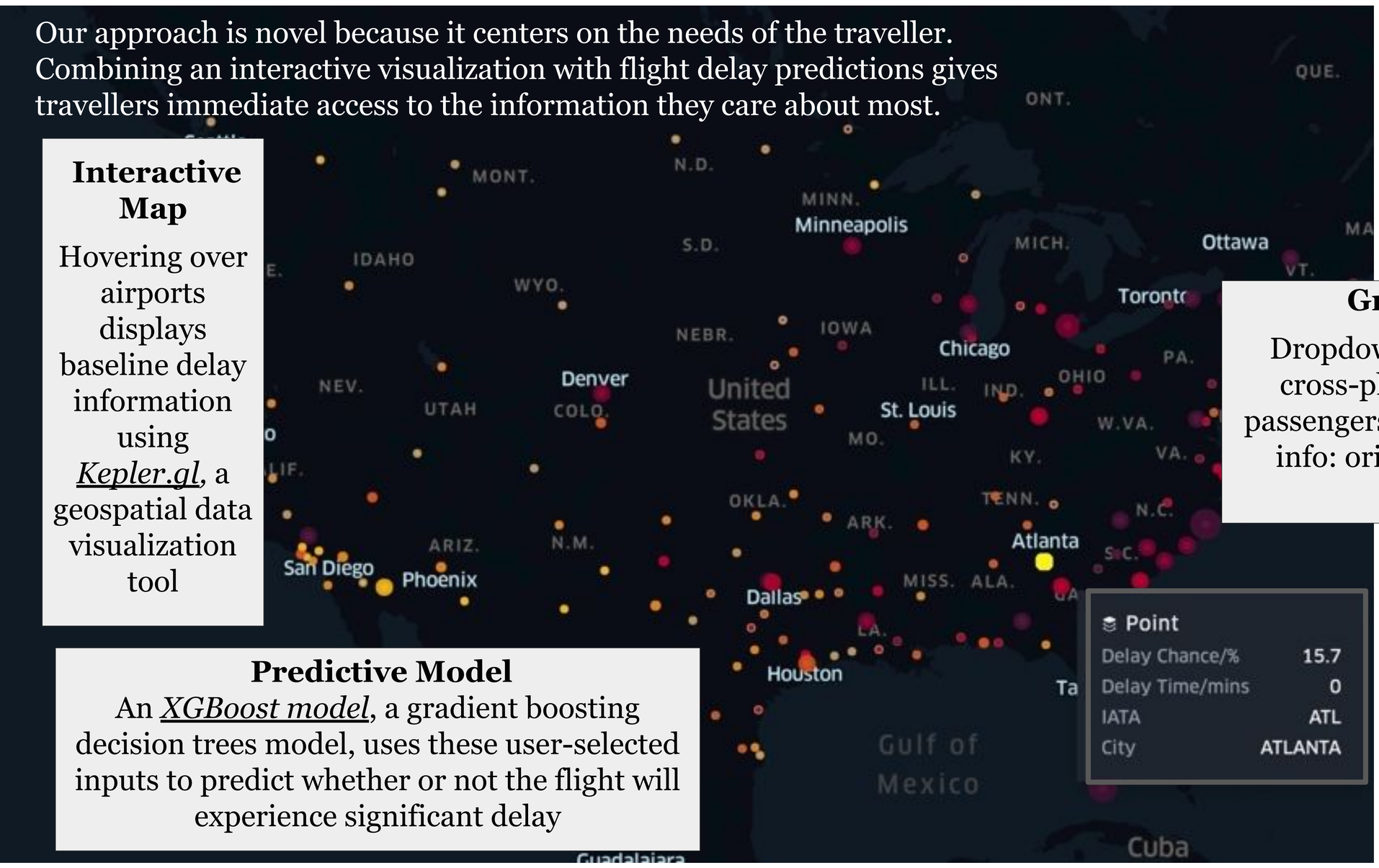
Air transportation dominates other modes for long-distance travel[1], yet this industry experiences high rates of inconvenient delays. In 2021, reporting marketing carriers in the US posted a delay rate of 22.6%[2] with over 10,500 flights are delayed globally each day[3]. Flight delays produce costs for individuals and for the economy at large, including loss of economic output, lowering values of flight tickets, and loss of productive hours[4]. Our project allows travellers to mitigate the adverse effects of delays by providing them accurate, customized predictions of experiencing a **significant delay (> 30 min)** to use in their travel planning.

Approach

Our approach is novel because it centers on the needs of the traveller. Combining an interactive visualization with flight delay predictions gives travellers immediate access to the information they care about most.

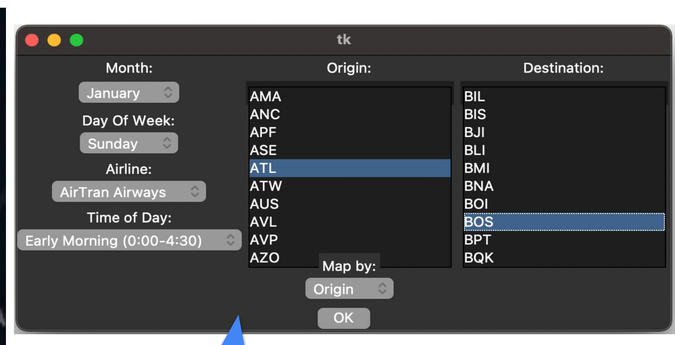
Interactive Map

Hovering over airports displays baseline delay information using *Kepler.gl*, a geospatial data visualization tool



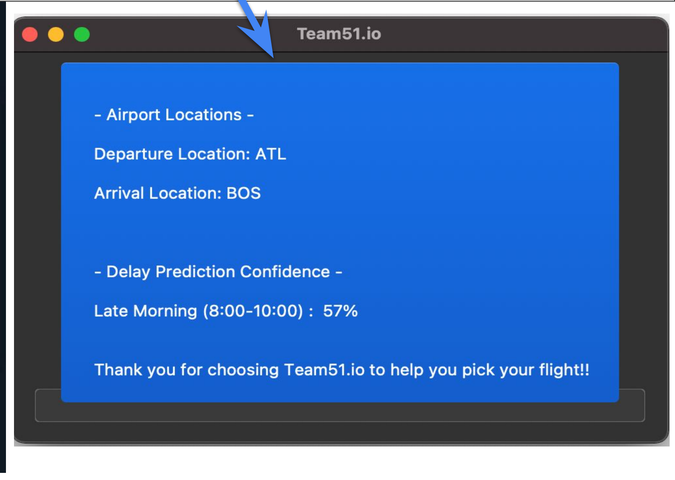
Predictive Model

An *XGBoost model*, a gradient boosting decision trees model, uses these user-selected inputs to predict whether or not the flight will experience significant delay



Graphical User Interface

Dropdown menus created with *tkinter*, a cross-platform package for GUIs, allow passengers to input the most pertinent flight info: origin, destination, date/time, and carrier

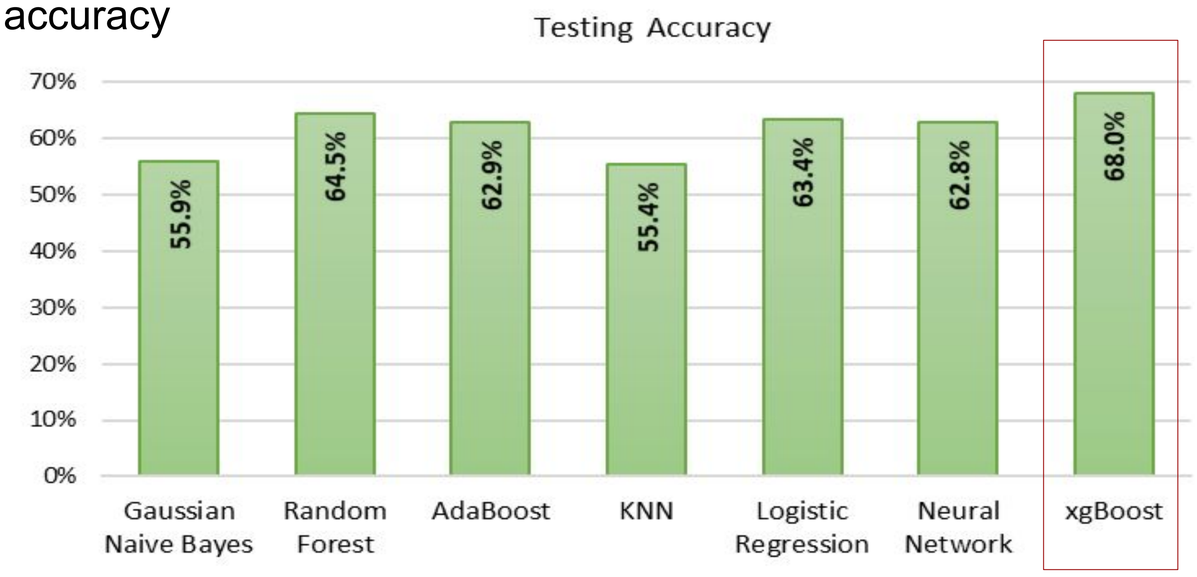


Data

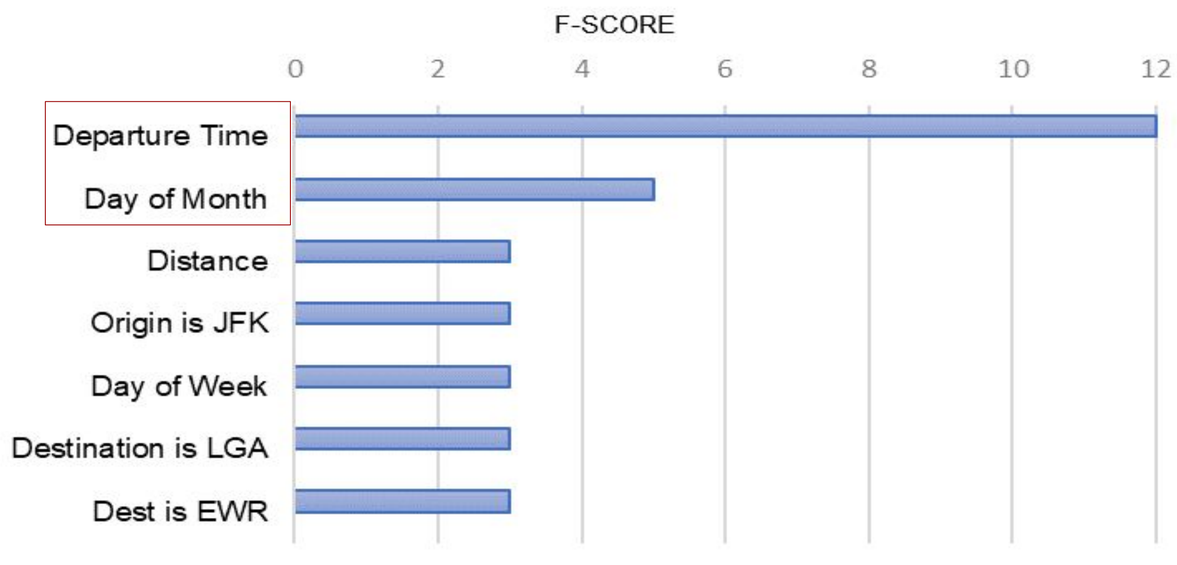
12 GB dataset (over 120 million records), downloaded from Kaggle[5]. Cleaned to include 13 predictor variables including flight date, time, carrier, origin and destination (IATA code, latitude, longitude).

Experiments & Results

Which models predict significant flight delay with the highest accuracy? Methods: Train models, compare testing accuracy



Which features are most predictive of significant delays? Methods: Correlation, entropy, feature importance plot.



[1] Rodrigue, Jean-Paul, and John Bowen. "5.5 Air Transport." The Geography of Transport Systems, 5th ed., Routledge, Taylor & Francis Group, New York, NY, 2020.
[2] "Air Travel Consumer Report: August 2021 Numbers." Air Travel Consumer Report: August 2021 Numbers | Bureau of Transportation Statistics, Bureau of Transportation Statistics, 5 Nov. 2021, https://www.bts.gov/newsroom/air-travel-consumer-report-august-2021-numbers.
[3] "Live Airline Flight Cancellations Info & Statistics." FlightAware, FlightAware, 6 Nov. 2021, https://flightaware.com/live/cancelled/yesterday.
[4] Borsky, Stefan, and Christian Unterberger. "Bad Weather and Flight Delays: The Impact of Sudden and Slow Onset Weather Events." Economics of Transportation, vol. 18, June 2019, pp. 10–26., https://doi.org/10.1016/j.ecotra.2019.02.002.
[5] bulter22. "Airline on-Time Performance Data." Kaggle, Data Expo 2009, 13 Dec. 2020, https://www.kaggle.com/bulter22/airline-data.