The Impact of Covid-19 on Air Traffic: Spatiotemporal Forecasting with Deep Learning

Adelle Driker, Bo Yan, Yuan Hu

Advisor: Professor Rose Yu

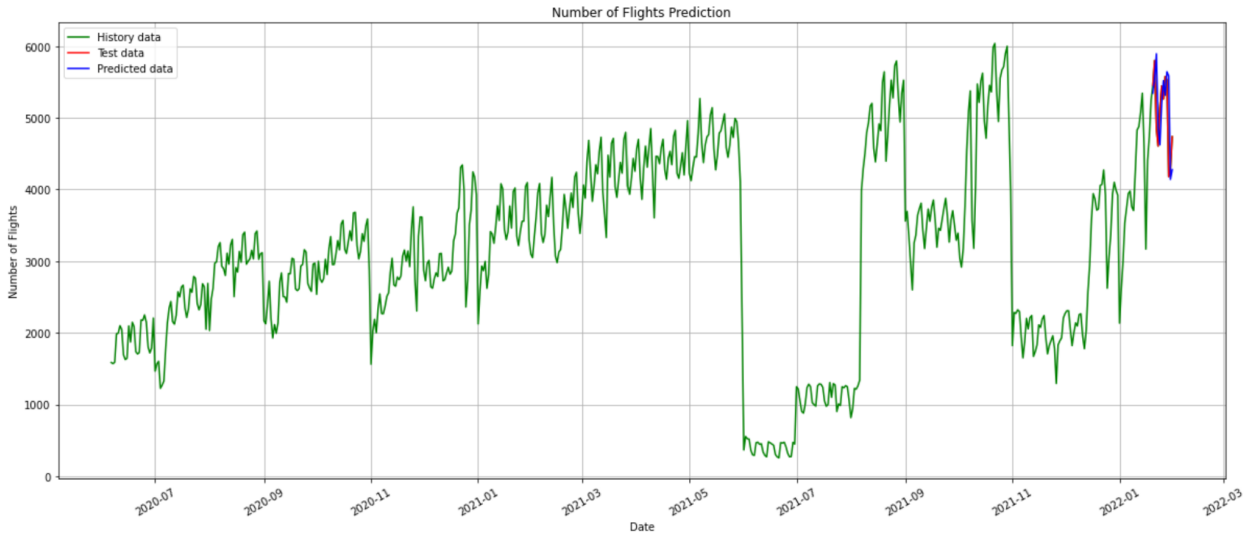
Findings and Reporting

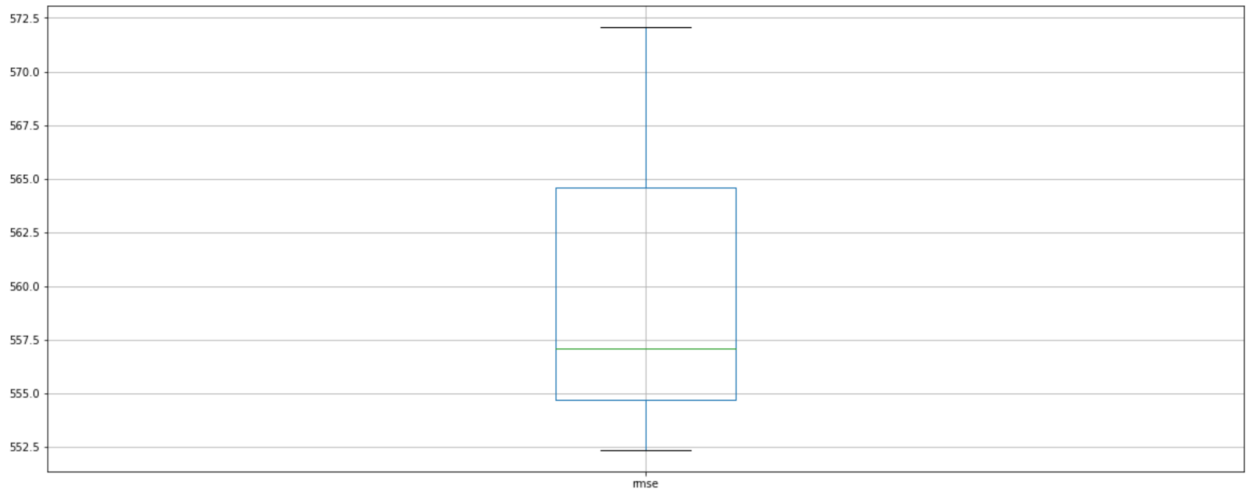
Findings

After applying a deep learning approach to our forecasting product, we found that the deep learning models performed quite well and provided forecast results with high accuracy. We are able to measure this success by comparing the results against classical baseline models - AR and ARIMA. The success was quantitatively measured by taking the R^2, RMSE, and MAE values to confirm that the deep learning models performed as well as and better than the baseline models we chose.

Despite that, we also practiced time series prediction with LSTM Recurrent Neural Networks in Python with Keras, below are the prediction results of LSTMs for Univariate Time Series Forecasting and LSTMs for Multi-Step Time Series Forecasting.

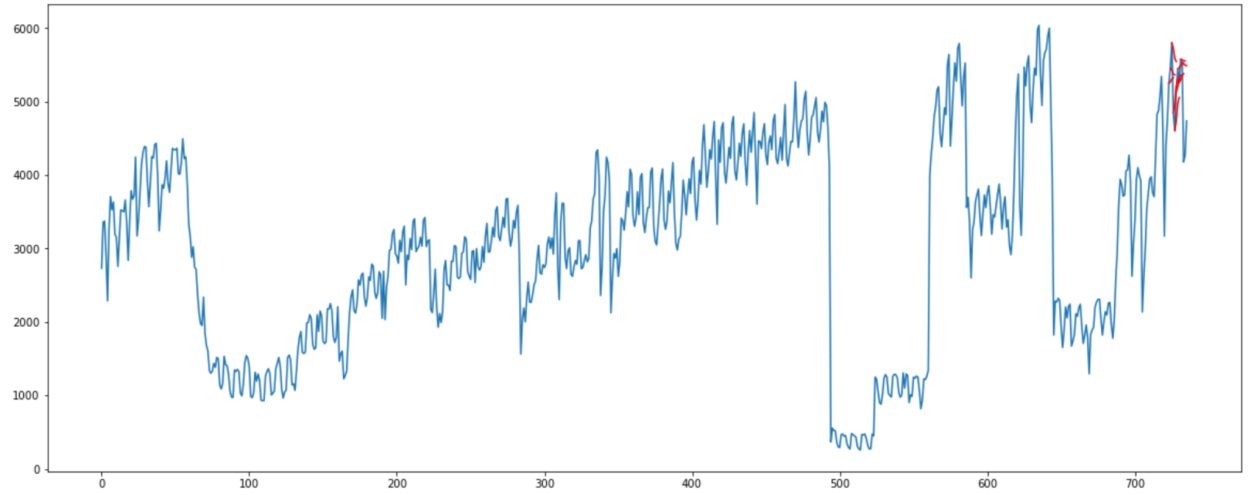
1. Results of LSTMs for Univariate Time Series Forecasting
    - The prediction result of LSTMs for Univariate Time Series Forecasting performs quite well. The box and whisker plot shows the distribution shown which captures the middle of the data as well as the extents and outlier results.
    - The prediction result of LSTMs for Multi-Step Time Series Forecasting performs well too.

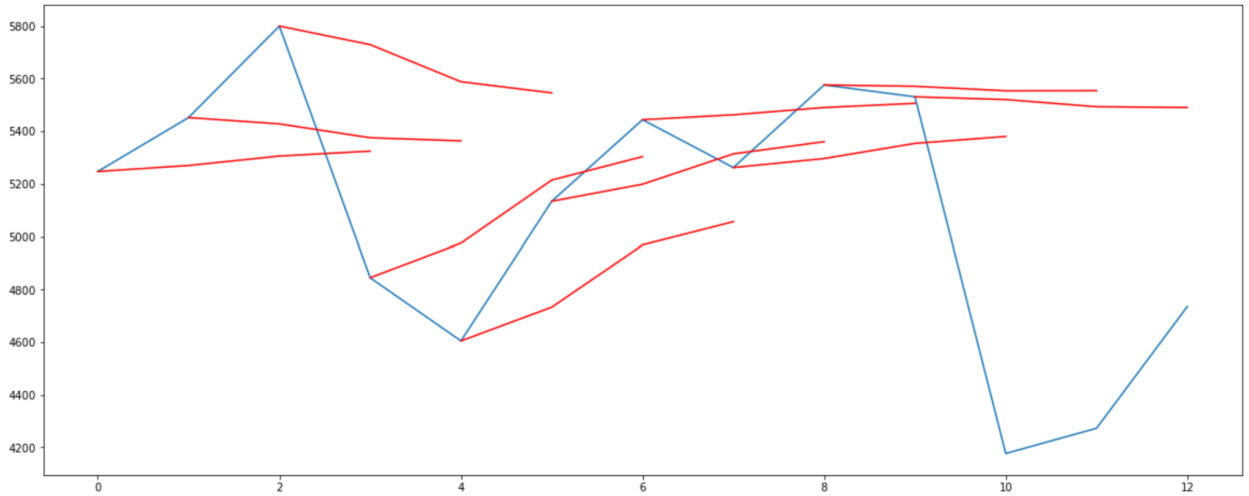




2. Results of LSTMs for Multi-Step Time Series Forecasting

- As we see, the lines connect to the appropriate input value for each forecast. It shows that although the skill of the model is better, some of the forecasts are not very good and that there is plenty of room for improvement.





Audience for Findings

The results from our findings are mainly geared towards analyst and management teams at airline companies so that they may accurately predict future demand, optimize future prices, and adjust their fleets of planes as needed.

Determining What to Present in Reporting

In our reporting, we will present our findings mainly focusing on the following points.

- The architecture of our system design.
- The models we choose to create and performance results.
- The dashboard we created to show our visualizations.
- The findings of our capstone project and how these findings benefit our stakeholders.
- Future works.

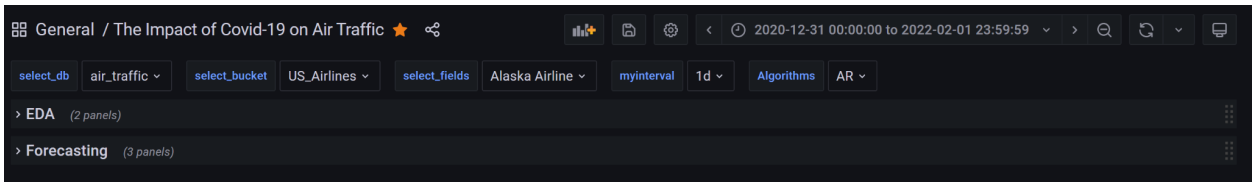
Our reports that are featured on the dashboard have been compiled in such a way that our target audience will be able to confidently make important business decisions. Although some models that we worked with, such as Seq2Seq, did not bring anticipated results, we plan to continue to improve those results and eventually incorporate them into the dashboard as well.

Techniques and Tools for Communication of Results

In order to create the dashboard, we used a tool called Grafana, which is an open-source web application suited for analytics and interactive visualizations. Grafana allows us to both link to our InfluxDB database which store our raw data and create EDA-like visualizations as well as upload csv files with our forecast modeling results to display our final product. To access and view the dashboard, a user will need to have an account to log in, thus adding a layer of security as well.

Visualizations and Other Reportable Products

Our dashboard consists of several visualizations, each of which can be controlled by a set of filters at the top. A user may select to view a set of visualizations of their choice - EDA or Forecasting.



Each plot can be filtered by date both by using the filters on top and by dragging the blue highlighted window as shown below. Additionally, a user may filter by airline type or hover over a point on the plot to reveal the quantitative information behind the data point such as the date, time and number of flights around the world.

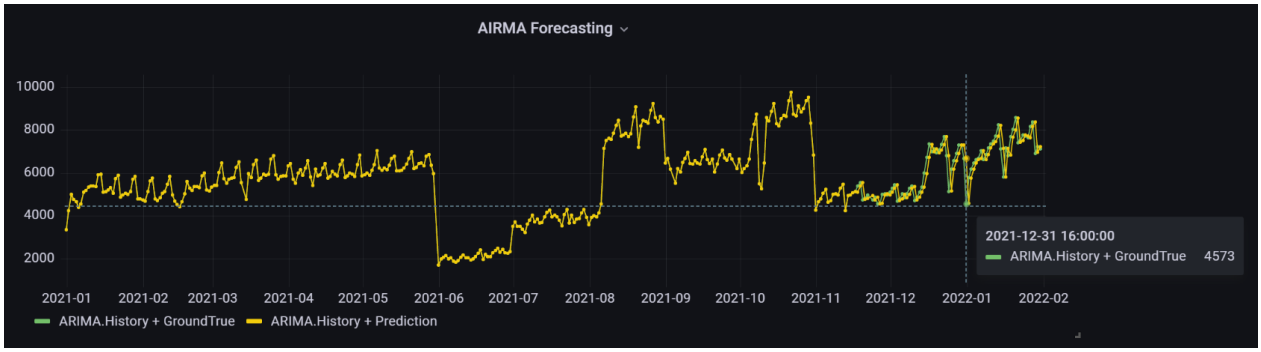

General / The Impact of Covid-19 on Air Traffic
2020-12-31 00:00:00 to 2022-02-01 23:59:59
select_db
Enter variable value
select_bucket
US_Airlines
select_fields
American Airline
myinterval
1d
Algorithms
AR
EDA
_internal
air_traffic
covid
models
Departing Flights
Number of Departing Flights
2.75 K
2.50 K
2.25 K
2 K
1.75 K
1.50 K
1.25 K
1 K
750
2021-02
2021-04
2021-06
2021-08
2021-10
2021-12
2022-02
US_Airlines.max_American Airline Min: 985 Max: 3.06 K


General / The Impact of Covid-19 on Air Traffic
2020-12-31 00:00:00 to 2022-02-01 23:59:59
select_db
air_traffic
select_bucket
US_Airlines
select_fields
Enter variable value
myinterval
1d
Algorithms
AR
EDA
Selected (1)
All
Alaska Airline
American Airline
Delta Airline
JetBlue Airline
SouthWest Airline
United Airline
Number of Departing Flights
3.25 K
3 K
2.75 K
2.50 K
2.25 K
2 K
1.75 K
1.50 K
1.25 K
1 K
750
2021-02
2021-04
2021-06
2021-08
2021-10
2021-12
2022-02
US_Airlines.max_American Airline Min: 985 Max: 3.06 K


General / The Impact of Covid-19 on Air Traffic
2020-12-31 00:00:00 to 2022-02-01 23:59:59
select_db
air_traffic
select_bucket
US_Airlines
select_fields
American Airline
myinterval
Enter variable value
Algorithms
AR
EDA
auto
1d
7d
14d
30d
Departing Flights
Number of Departing Flights
3.25 K
3 K
2.75 K
2.50 K
2.25 K
2 K
1.75 K
1.50 K
1.25 K
1 K
750
2021-02
2021-04
2021-06
2021-08
2021-10
2021-12
2022-02
US_Airlines.max_American Airline Min: 985 Max: 3.06 K

COVID-19 data at the country level is also available in parallel in order for the user to track the number of covid cases in both origin and destination of travel. A user may click and drag the map to change location as well as zoom in and out to see more granular data. The amount of Covid cases are also encoded with a color scale where red is the more severe, while green is very low.

Once the user selects Forecasting, they will see the following figures, which contain the forecasting results from the different models. Each plot can be filtered by date both by using the filters on top and by dragging the blue highlighted window as shown below. Additionally, a user may hover over a point on the plot to reveal the quantitative information behind the data point such as the date, time and number of flights around the world.



Major Updates to Steps 1-8

As we work to finalize our final project, we are focusing on several items to ensure that our product is ready for production: streamlining the architecture to help automate the pipeline, updating the final dashboard to include more insightful visualizations, and applying final tuning and optimization to our models. We have added two additional models to our collection - LSTMs for Univariate Time Series Forecasting and LSTMs for Multi-Step Time Series Forecasting.

Team Member Contributions

Bo:

- Wrote and debug scripts for LSTMs-Univariate model and LSTMs-Multi-Step model, tuned performance for both of these two models
- Contributed to “Findings” and “Determining What to Present in Reporting”
- Maintained Capstone Project Planning spreadsheet, GitHub, and Rose’s Documentation

Yuan:

- Continued to add to the Grafana dashboard to include new visualizations and user interaction
- Worked on streamlining the data pipeline to ensure proper data flow

Adelle:

- Contributed to “Findings and Reporting” subsections and “Major Updates Section”
- Coordinated with teammates to devise a final plan for the project