UC San Diego

The Impact of Covid-19 on Air Traffic:

Spatiotemporal Forecasting with Deep Learning and Benchmarks

Adelle Driker, Bo Yan, Yuan Hu

Advisor: Professor Rose Yu

Agenda

1. Meet the Team
2. Motivation & Problem Definition
3. Data - {Sources, Pipeline, Preprocessing}
4. ETL & Pipeline
5. EDA & Initial Insights
6. Modeling
7. Scalability & Robustness
8. Product Dashboard Demo
9. Conclusion

MEET THE TEAM

Meet the Team

Advisors

Professor Rose Yu

Professor Ilkay Altintas de Callafon

Meet the Team

About Us

Bo Yan

- Software/ML/DL Engineer
- Record Keeper

Yuan Hu

- Budget Manager
- Data Engineer
- Solution Architect

Adelle Driker

- Project Coordinator/Manager
- Data/Business Analyst

MOTIVATION
&
PROBLEM DEFINITION

Motivation

Crippled Airline Industry to Get $25 Billion Bailout, Part of It as Loans

Airlines will receive billions of dollars in grants and loans to pay flight attendants, pilots and other employees.

The New York Times - April 14, 2020

American Airlines And United Report Nearly $4 Billion In Combined Losses

NPR - April 30, 2020

Air travel down 60 per cent, as airline industry losses top $370 billion: ICAO

UN News - January 15, 2021

Motivation

According to BBC Future (July 9th, 2020):

- As of April 2020, London's Heathrow Airport reported that passenger numbers were down **97%**
- Nearly **30%** of the world's **26,000** commercial aircraft were grounded on tarmacs worldwide

Problem Definition

Creates uncertainty for both passengers and airline companies, especially due to the multiple waves of virus mutations

How should airlines plan future flights? When should passengers schedule their travels?

In other words, given a country's COVID situation, how should an airline/passengers plan ahead?

Target Stakeholders

Airline companies, which would be allowed to:

- Predict future demand and plan to use more efficient aircraft
- Optimize future prices

DATA

{Sources, Pipeline, Preprocessing}

Data

What is Spatiotemporal Data?

SPATIOTEMPORAL

Having to do with location
(x, y, z)

Having to do with time
yyyy-mm-dd
hh:mm:ss

Data Sources

Flight Data
(OpenSky Network)
~10GB

Bansard Airline Code and Country Mapping File

Covid-19 Data
(Johns Hopkins Covid-19 Data Repository)
~2MB

Data Sources

Flight Data

Attributes:

Callsign*
Number
ICAO24
Reg
TypeCode
Origin/Destination Airport
First/Last Seen (DT)
Lat/Long/Alt of Origin/Destination

Data Quality:

About 10% of airports appear as NaN

Negligible number of Lat/Long missing

Data provided as-is

Velocity:

Files updated on a **monthly** basis

Data Sources

COVID-19 DATA

Attributes:

Province/State
Country/Region*
Latitude
Longitude
Dates

Data Quality:

Some Countries have missing counts

A few country names do not match names in Flight data

Format not readily ingestible

Updated for historical accuracy

Velocity:

Files updated on a **daily** basis

Data Preprocessing

1. Missing origin/destination values in Flight dataset were imputed to preserve Data Quality and Size
2. Unnecessary columns were removed to conserve space
3. Country names adjusted to match between Sources

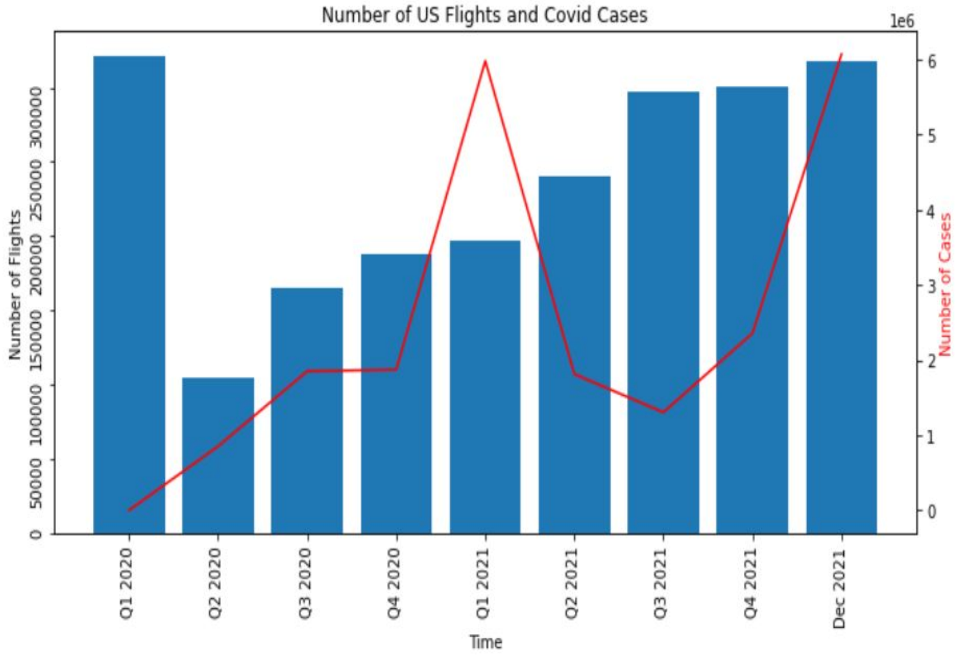
   E.g. "US" → "United States"
4. Timestamps converted to yyyy-mm-dd hh:mm:ss format
5. Covid-19 counts calculated to show daily totals instead of cumulative totals

EDA & INITIAL INSIGHTS

EDA & Initial Insights

Does the number of flights (US) follow the trend of Covid-19 cases (US)?

- Number of flights show a drastic dip in Q2 2020, but continue to recover in the following months
- Pandemic waves are evident and illustrated by red spikes



EDA & Initial Insights

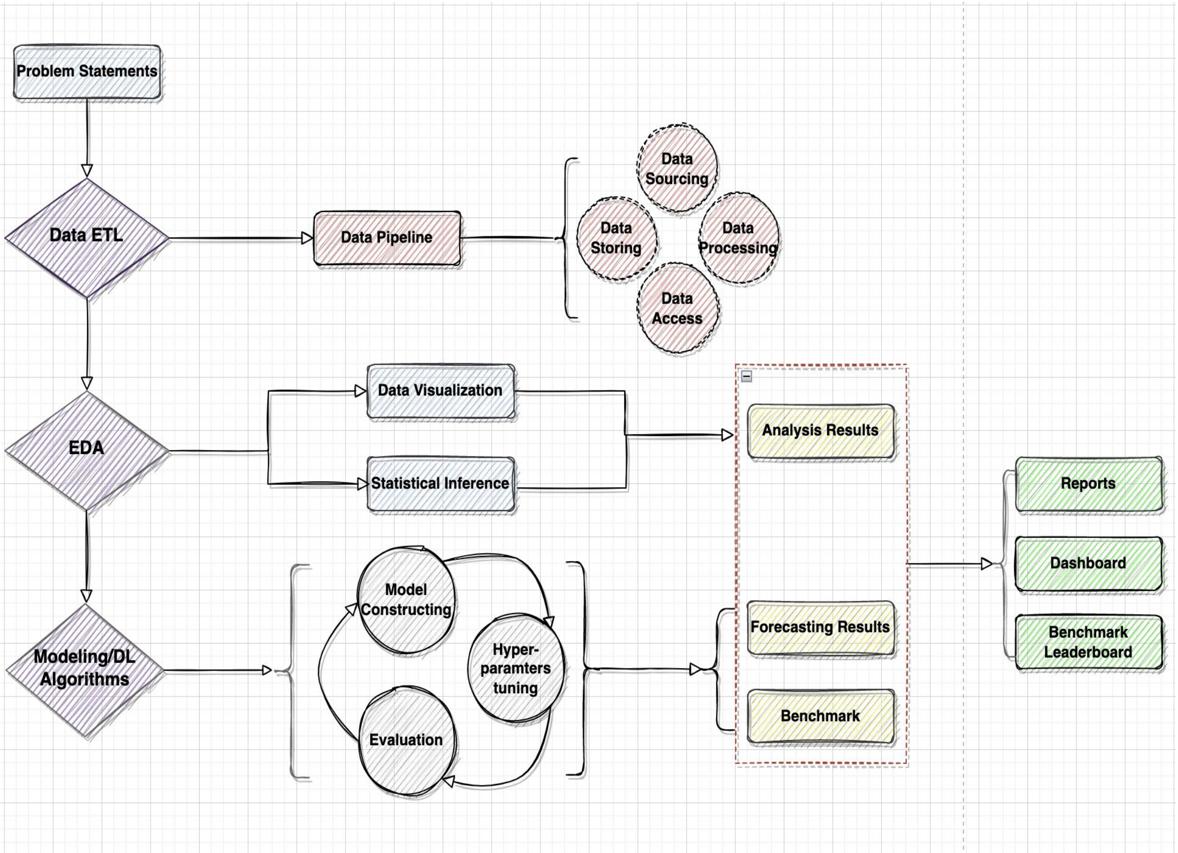
Where are global flights typically concentrated?

- Most flights are concentrated between the US (red) and Europe (yellow)
- China's Air industry activity decreases dramatically due to international travel ban.

Network of Global Flight Trajectory

ETL
&
ARCHITECTURE DESIGN

Product Functional Workflow



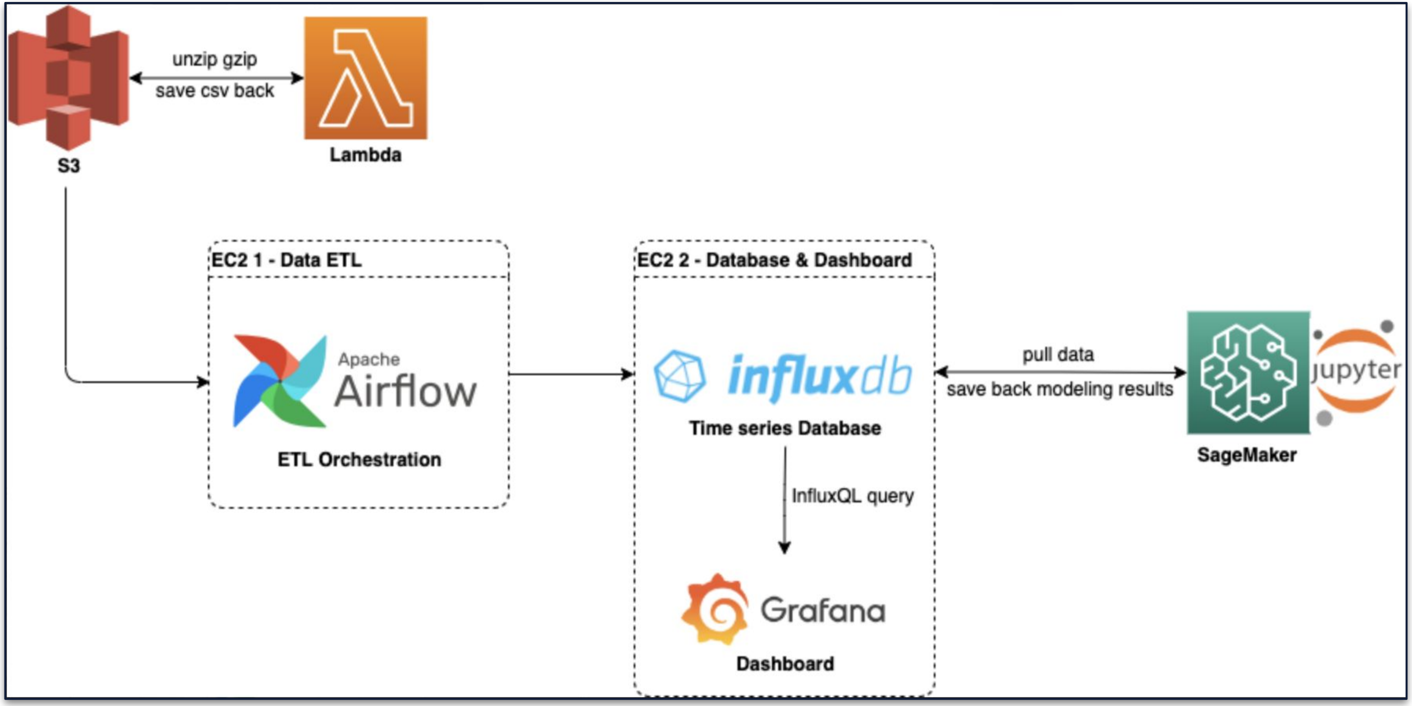
Product Architecture

Cloud Based Architecture

Orchestrated ETL Pipeline

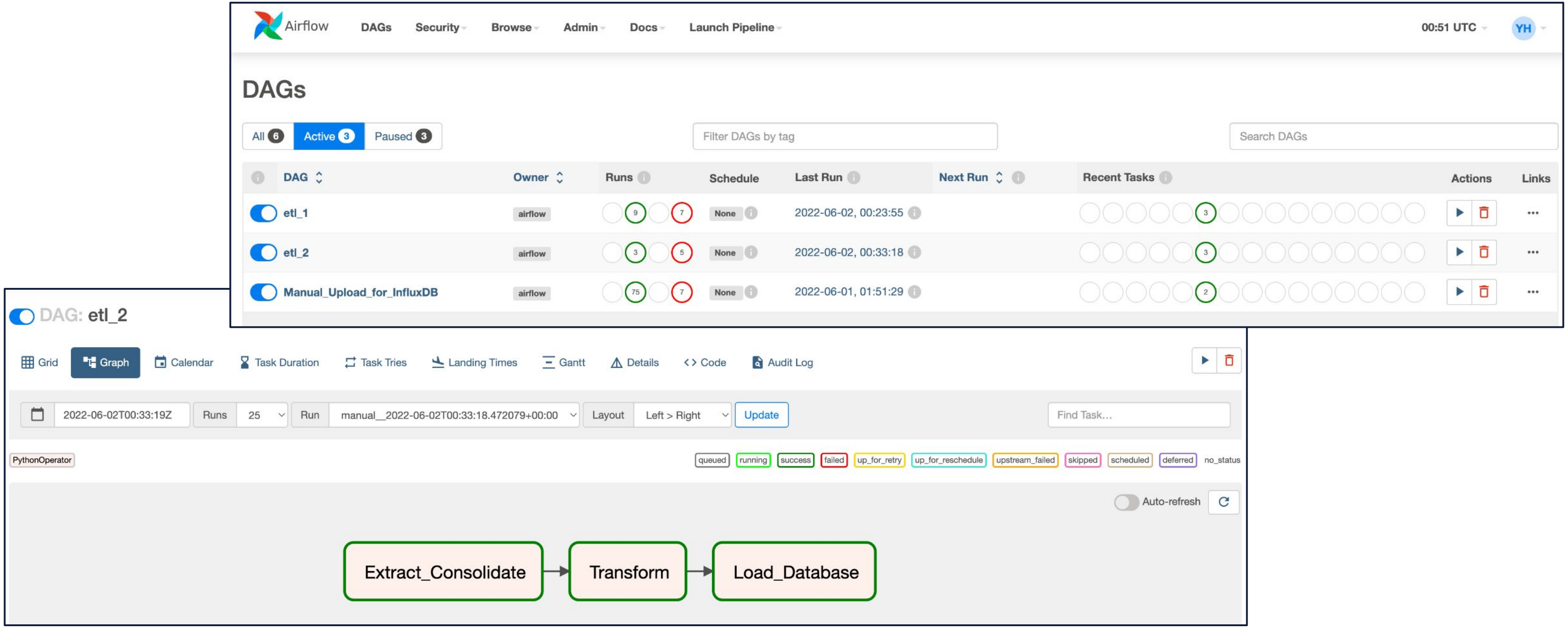
Scalable Components

Cross-Platform Interactive Dashboard

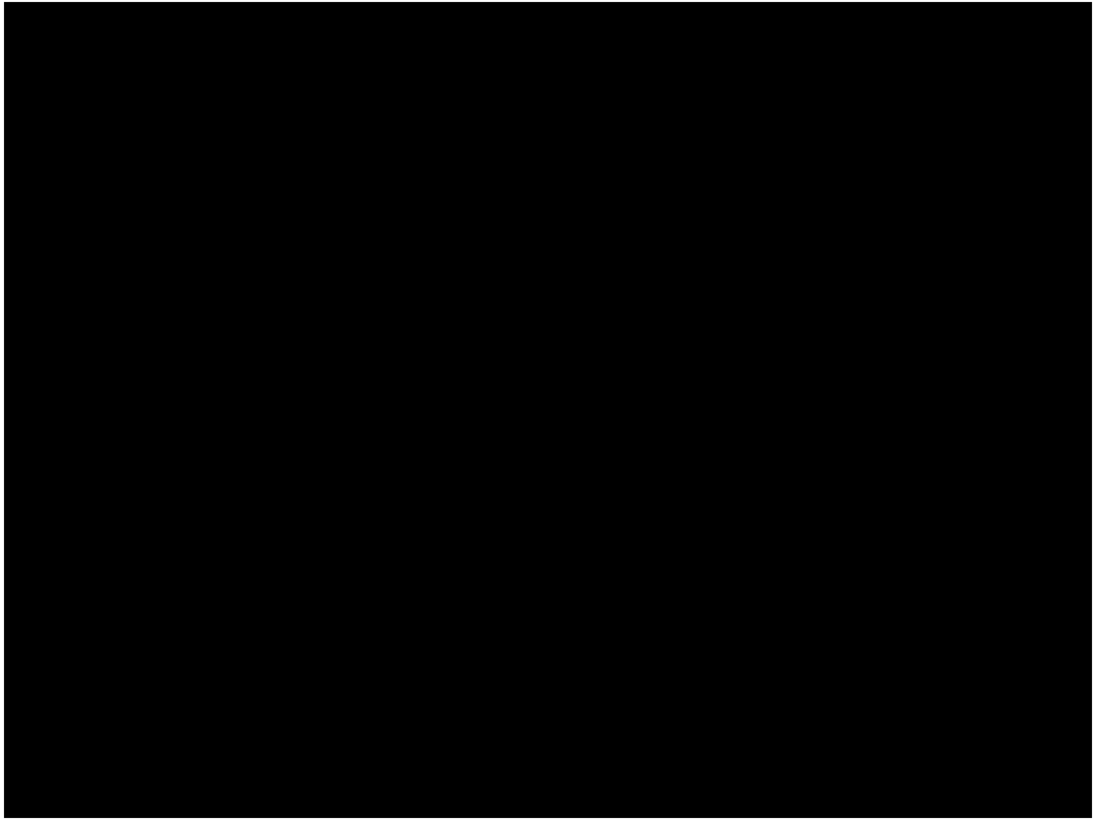


ETL Pipeline

Orchestrated by Apache Airflow Celery Executor and DAGs(Directed Acyclic Graph)



ETL Pipeline Demo



MODELING

Modeling - Approach

Baseline models

- Autoregression (AR) Model
- Autoregressive Integrated Moving Average (ARIMA) Model

Deep learning models

- Long Short Term Memory (LSTM) Model
  - Simple LSTM model
  - LSTM with Mean Interval Score Regression (1 Confidence Interval)
  - LSTM with Mean Interval Score Regression (2 Confidence Intervals)
  - LSTM with Quantile Regression
  - LSTM with One-step Univariate Time Series Forecasting
  - LSTM with Multi-step Time Series Forecasting
- Other models such as Seq2Seq, DCRNN

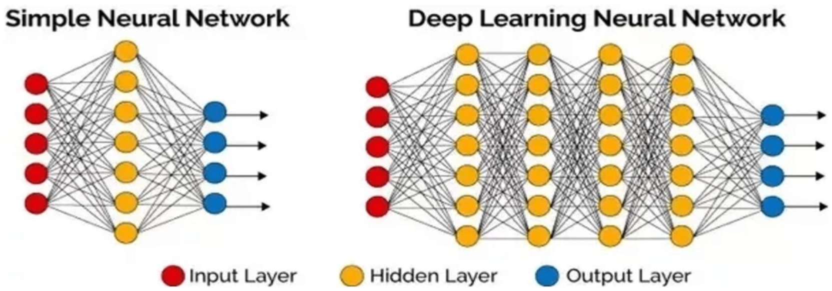
Modeling - Approach

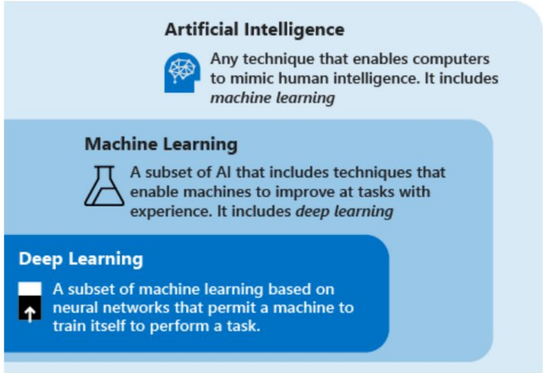
What are Baseline Models (AR, ARIMA)?

- A simple model that acts as a reference in a machine learning project.
- Main function: contextualize the results of trained models.
- Based on the statistical concept of serial correlation, where past data points influence future data points.

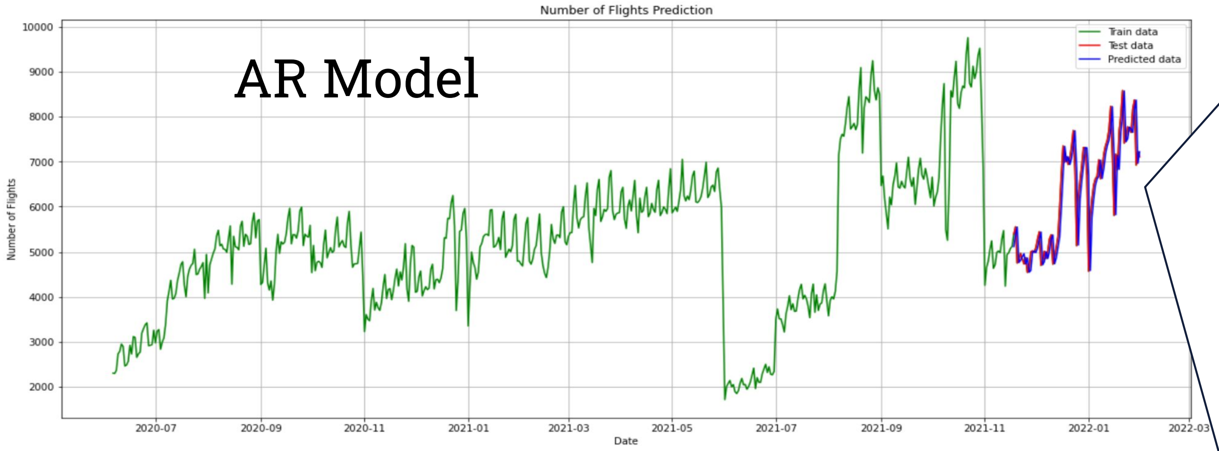
What are Deep Learning models (LSTM)?

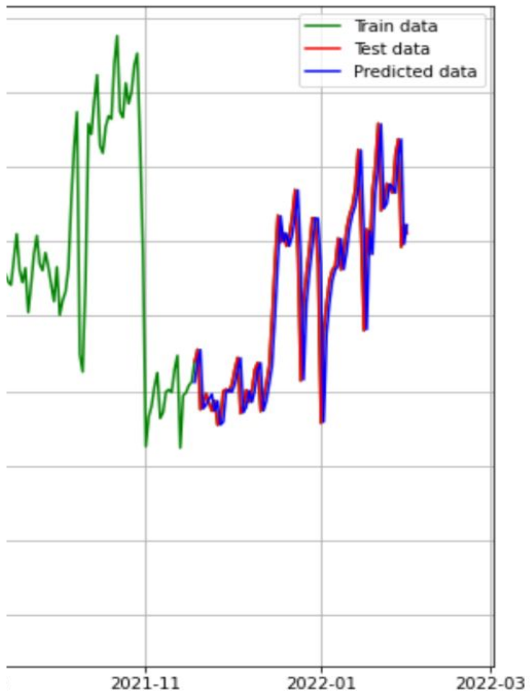
- Type of RNN that handles sequential data very well.
- Used for handwriting and speech recognition, machine translation, etc
- Developed to solve the Vanishing Gradient Problem seen in traditional RNNs.
- Forgets and remembers things selectively!





Modeling - Baseline Model 1

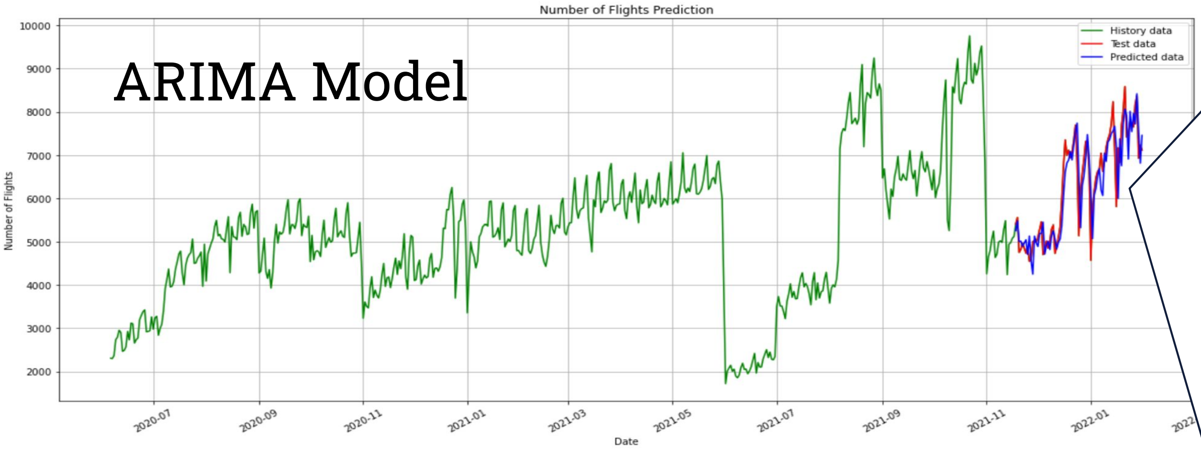


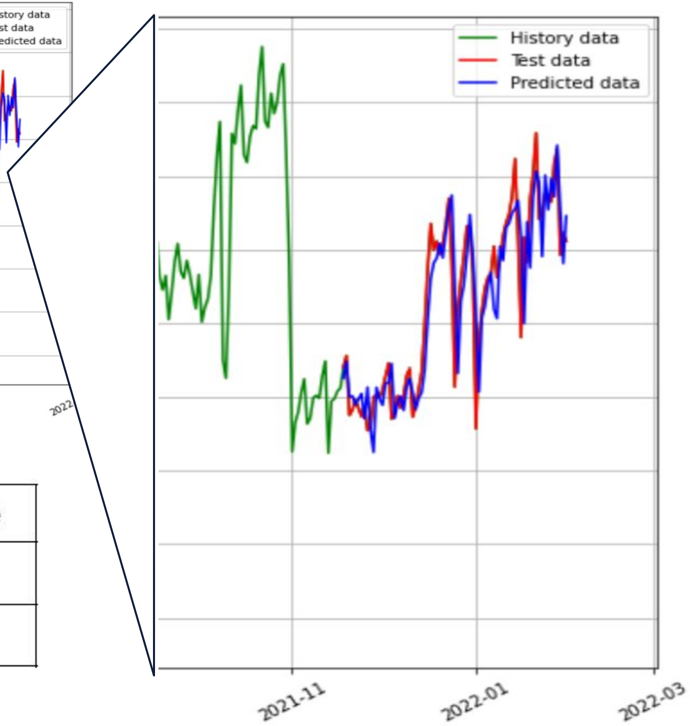


Models	ME	MSE	MAE	RMSE	R^2 score
AR	1,530.095	259,351.994	381.051	509.266	0.808

1. r2: 81% of the data fit the AR model (better fit)
2. Strong effect size, 81% of the variance of the dependent variable can be explained by the variance of the independent variable.
3. The lower value of MAE, and RMSE implies higher accuracy of AR model.

Modeling - Baseline Model 2



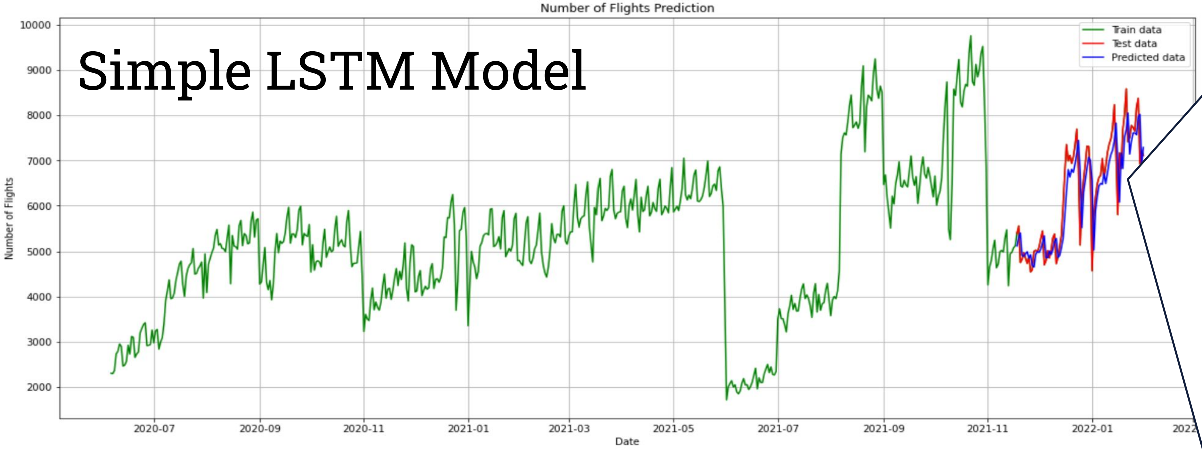


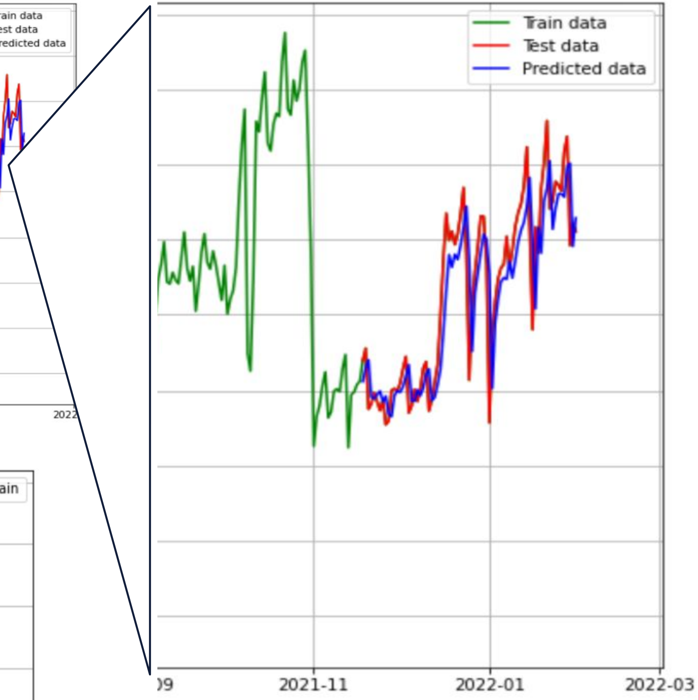
Models	ME	MSE	MAE	RMSE	R^2 score
AR	1,530.095	259,351.994	381.051	509.266	0.808
ARIMA	1,782.427	240,363.422	339.022	490.269	0.845

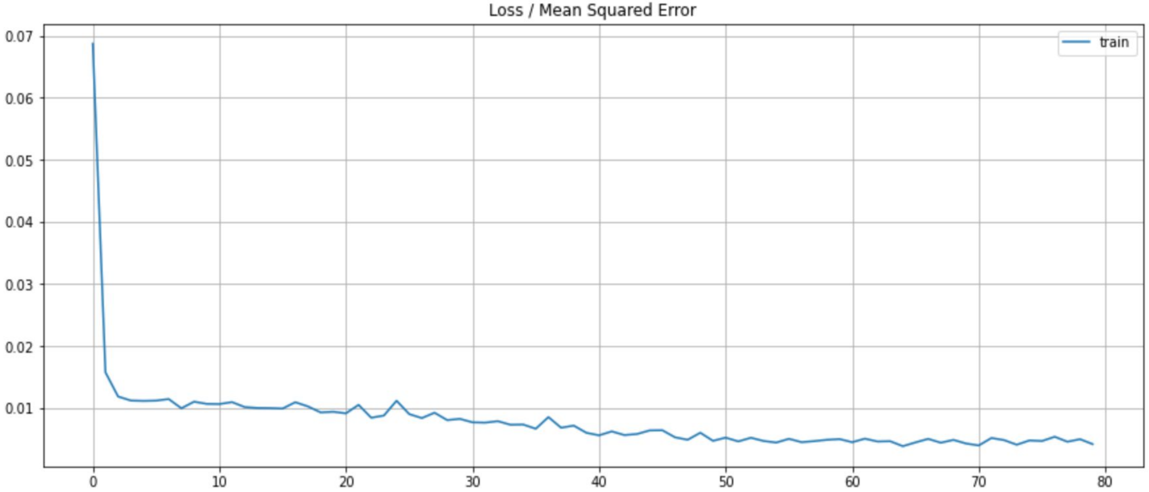
1. r2: 85% of the data fit the ARIMA model (better fit)
2. Strong effect size, 85% of the variance of the dependent variable can be explained by the variance of the independent variable.
3. The lower value of MAE, and RMSE implies higher accuracy of ARIMA model.

Modeling - Deep Learning Model 1

Simple LSTM Model

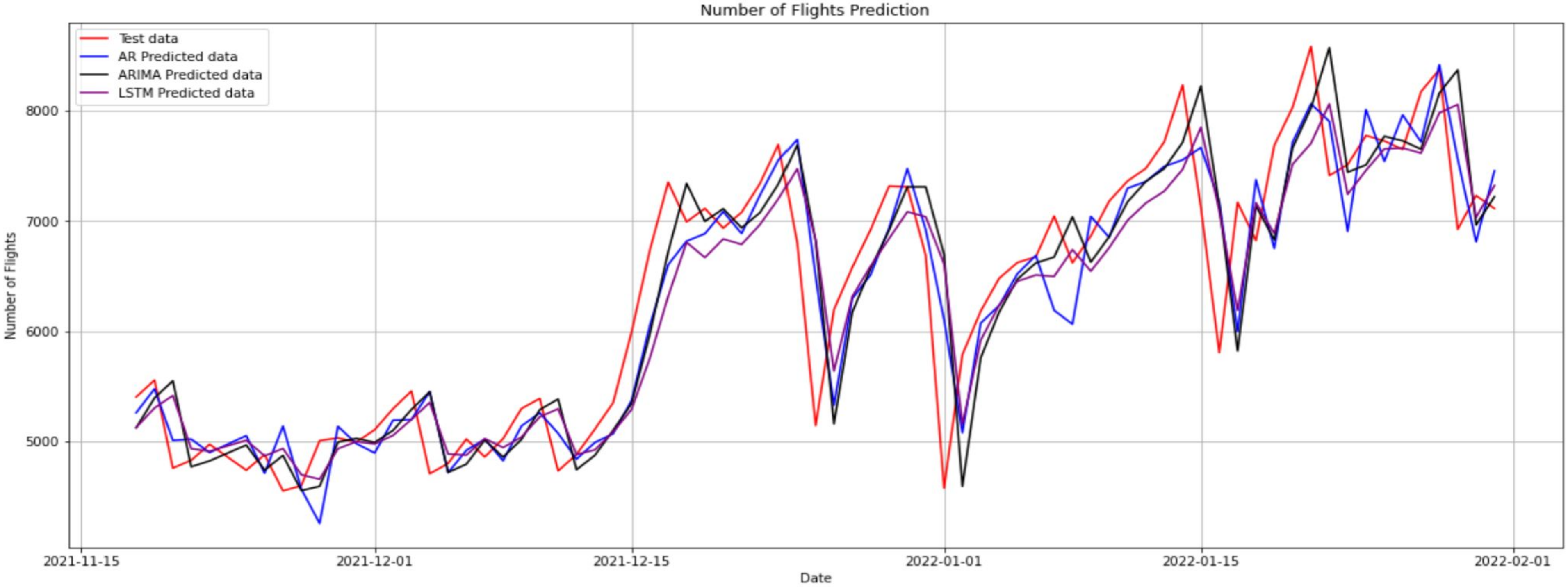






1. Reflect the trend information
2. Best epoch is 60
3. Performs well

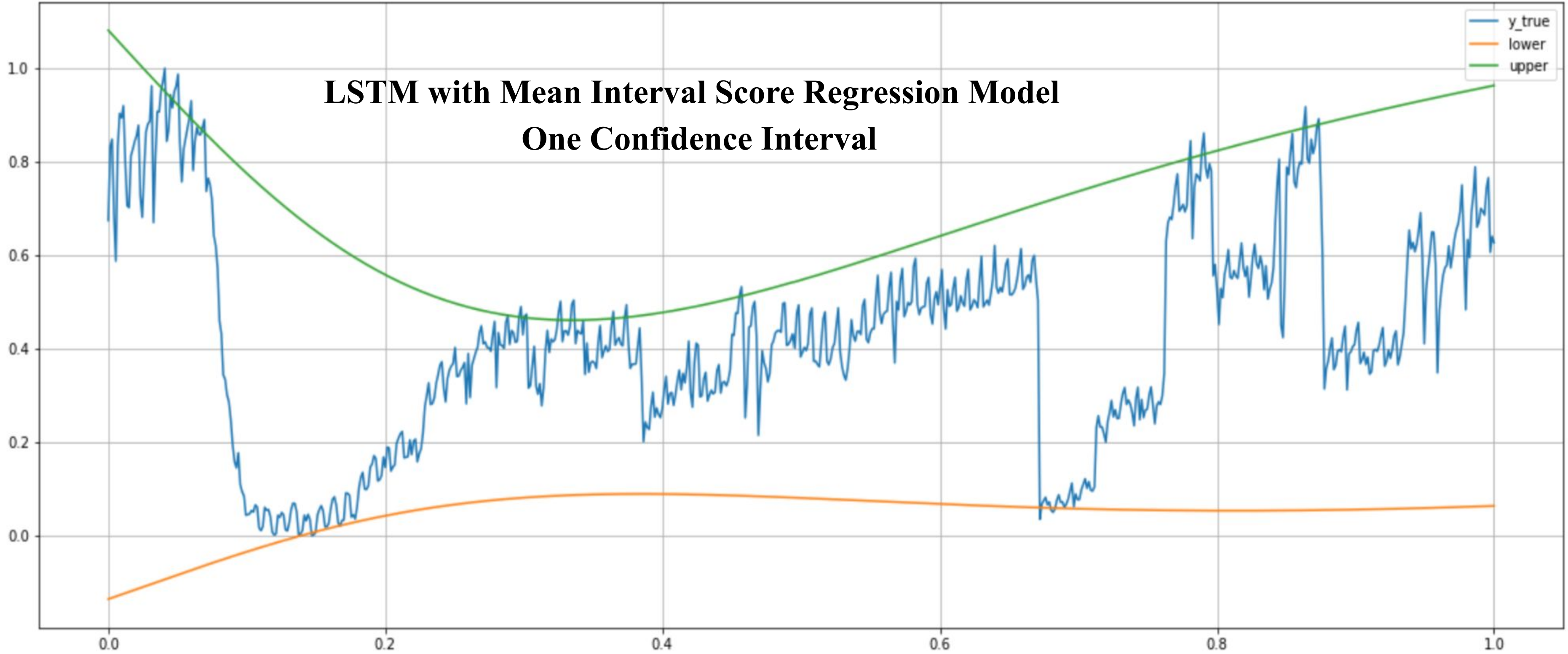
Modeling - Summary of Three Core Models



Models	ME	MSE	MAE	RMSE	R^2 score
AR	1,530.095	259,351.994	381.051	509.266	0.808
ARIMA	1,782.427	240,363.422	339.022	490.269	0.845
LSTM	1,982.785	313,407.582	427.684	559.828	0.768

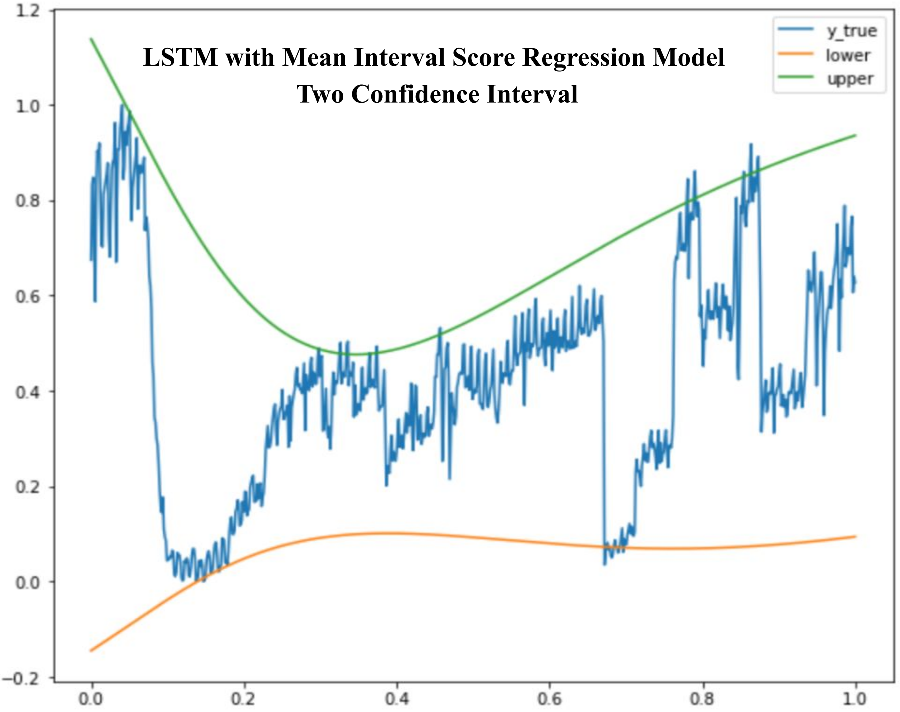
1. All models: high accuracy.
2. LSTM: performed as well as baseline models.
3. All models: major deviation throughout Dec 2021 and Jan 2022, which signifies the peak of the Omicron wave

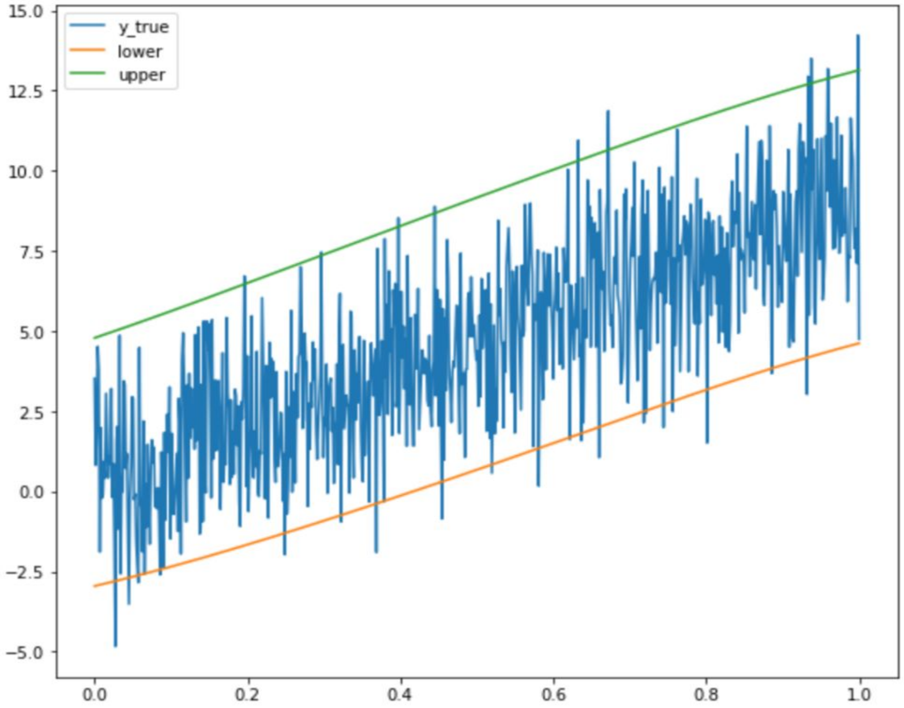
Modeling - Deep Learning Model 2-1



1. Use a mean interval score loss function with a LSTM network
2. Tend to stay within the estimation corridor
3. A larger sample will reduce the sampling error, give more precise estimates and thus smaller intervals.

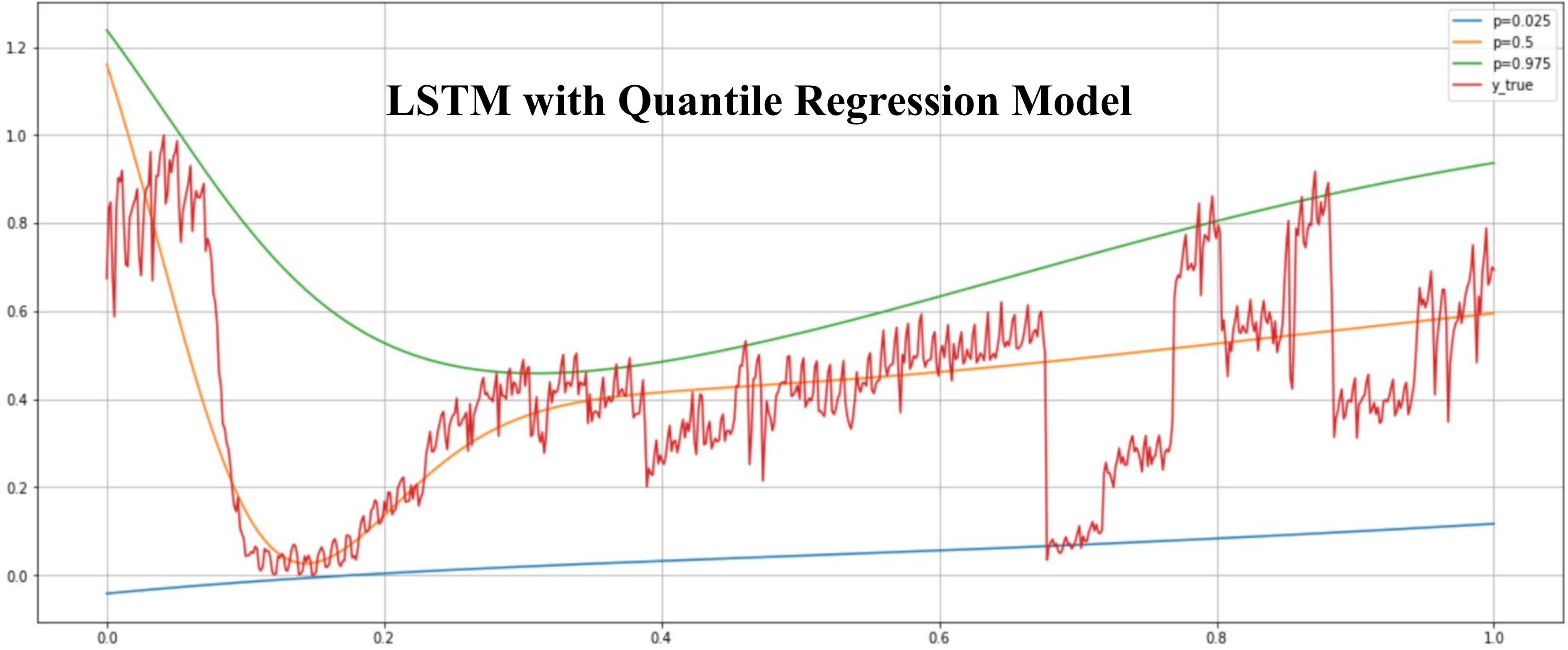
Modeling - Deep Learning Model 2-2





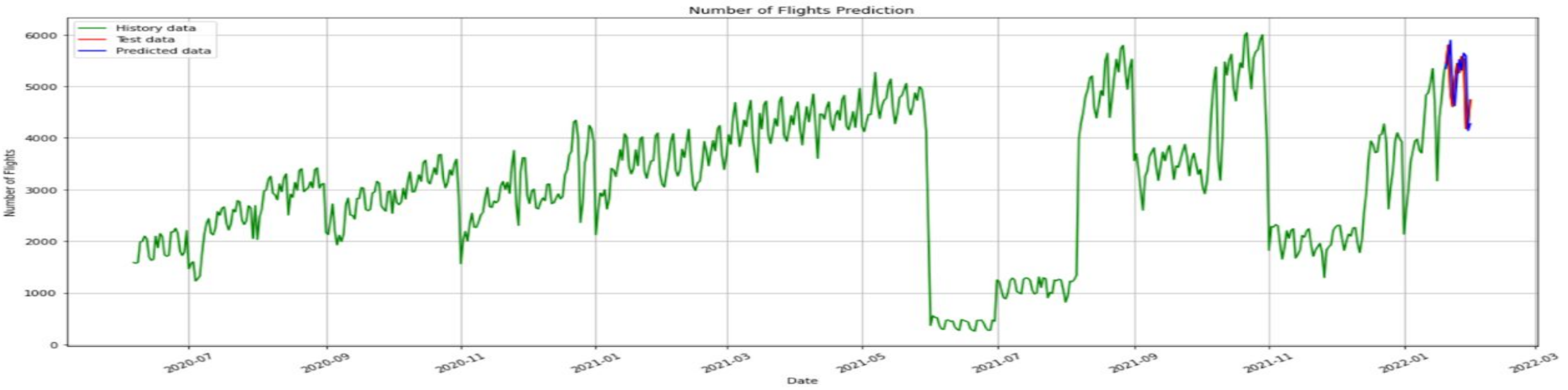
1. Use a mean interval score loss function with a LSTM network.
2. Consistently tend to stay within the estimation corridor, and wherever the plot travels beyond the confidence intervals shown in blue and green indicates that the confidence interval becomes narrower.
3. A larger sample will reduce the sampling error, give more precise estimates and thus smaller intervals.

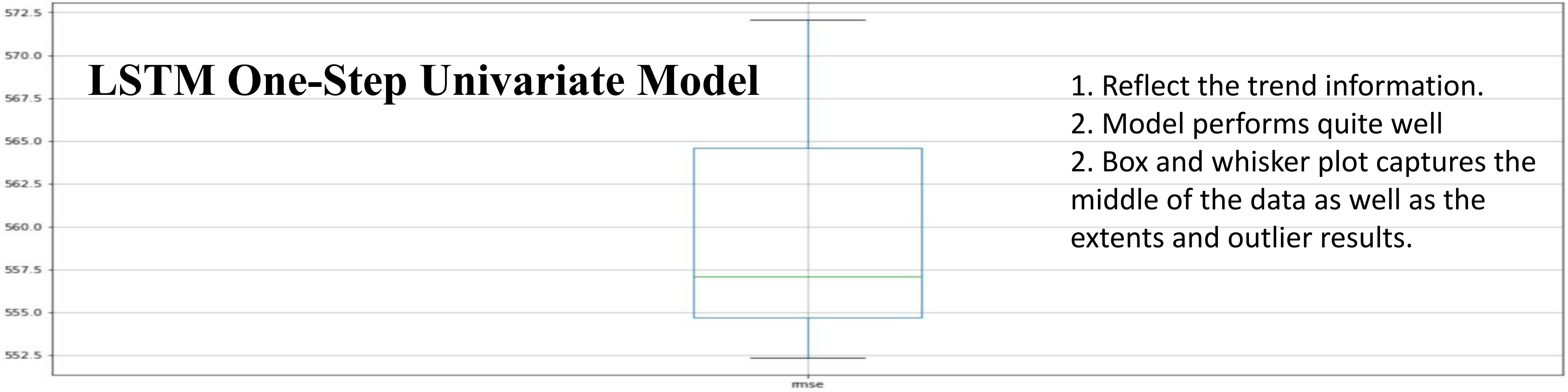
Modeling - Deep Learning Model 3



1. Use a quantile loss function with a LSTM network.
2. Required quantiles, 0.025, 0.5 and 0.975, have 3 output nodes, with each node having a different loss function.
3. This ensures that the structure of the data is shared in the first few layers.

Modeling - Deep Learning Model 4

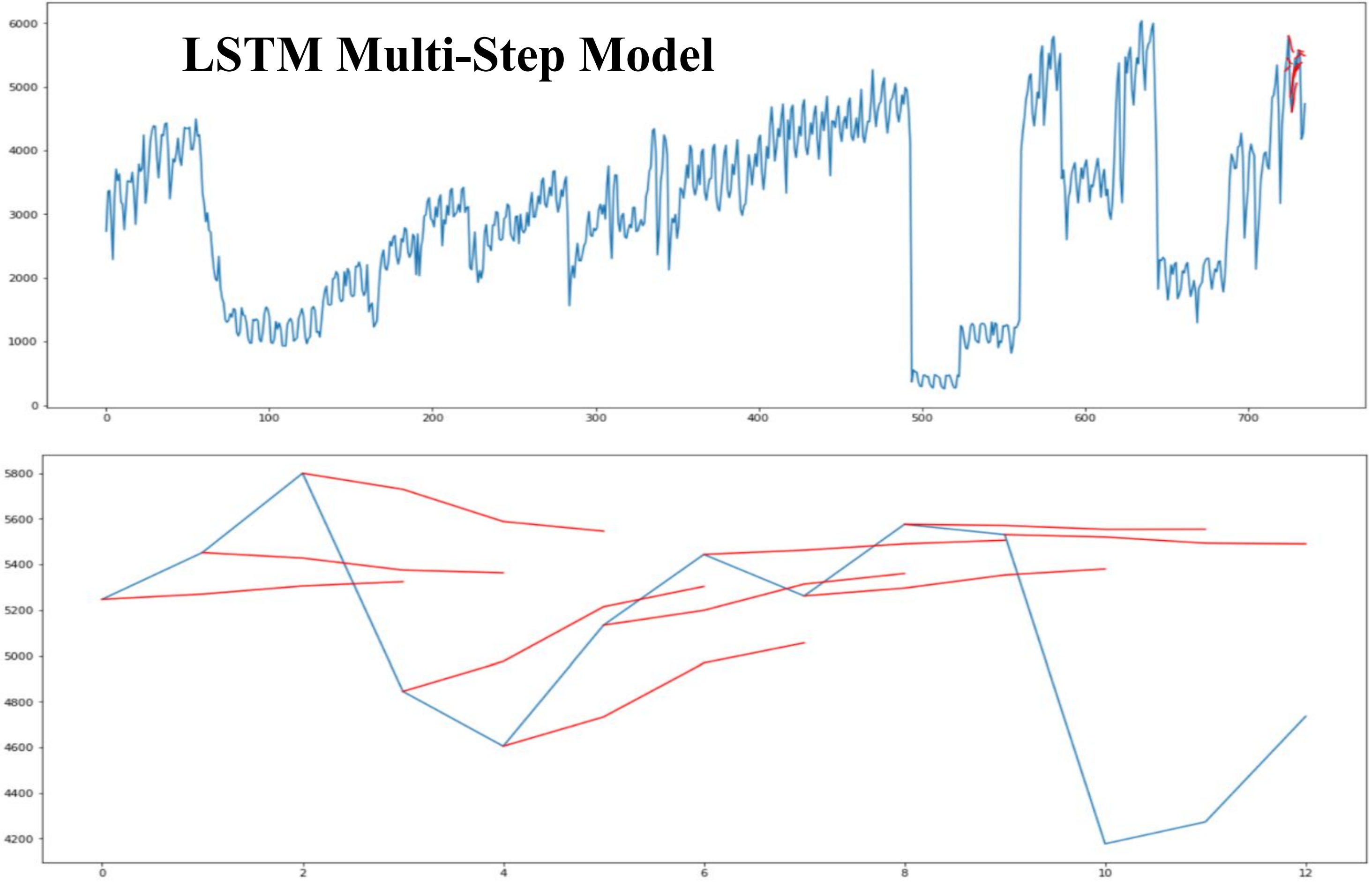




LSTM One-Step Univariate Model

1. Reflect the trend information.
2. Model performs quite well
2. Box and whisker plot captures the middle of the data as well as the extents and outlier results.

Modeling - Deep Learning Model 5



1. t+2 appears easier to forecast
2. Results at each forecasted time step are better, in some cases much better
3. Skill of the model is better, some of the forecasts are not very good
4. Plenty of room for improvement

Scalability & Robustness

Data Scalability:

- Model performance evaluated by incorporating different subsamples of the whole datasets, 10%, 20%, and 100% to analyze data scalability.

Model Scalability:

- Model scalability is tested by assessing the compute time for different model network architecture.

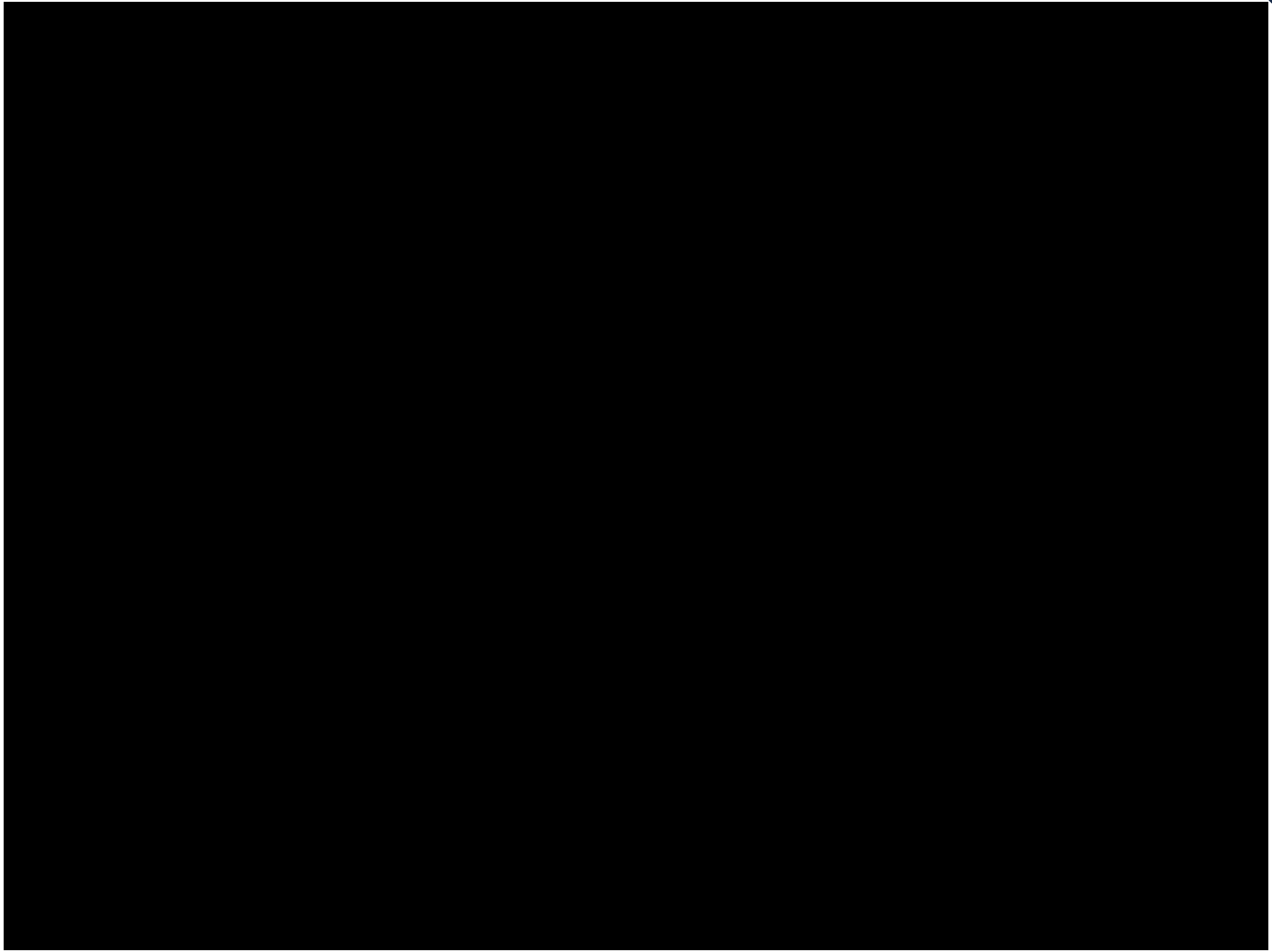
Compute Scalability:

- Compute scalability was achieved by wrapping deep learning models around PyTorch Lightning, AWS S3 and Celery executor to elastic scalability.

System Robustness:

- **Amazon s3**: Easily manage data at any scale with robust access controls & Backup and restore critical data with robust application feature
- **Amazon EC2**: Load balancers: distributes network or application traffic

PRODUCT DASHBOARD DEMO



CONCLUSIONS

CONCLUSIONS

- Deep learning models performed quite well and provided forecast results with high accuracy
- By comparing deep learning modeling results against classical baseline modeling results, we confirmed that the deep learning models performed as well as and better than the baseline models
- Our deep learning results are also visualized to create benchmarks so that they can aid in easily identifying the best options to help users accurately and confidently make decisions

THANK YOU!
Q&A