Deciphering Air Travel Disruptions: A Machine Learning Approach

Authors:

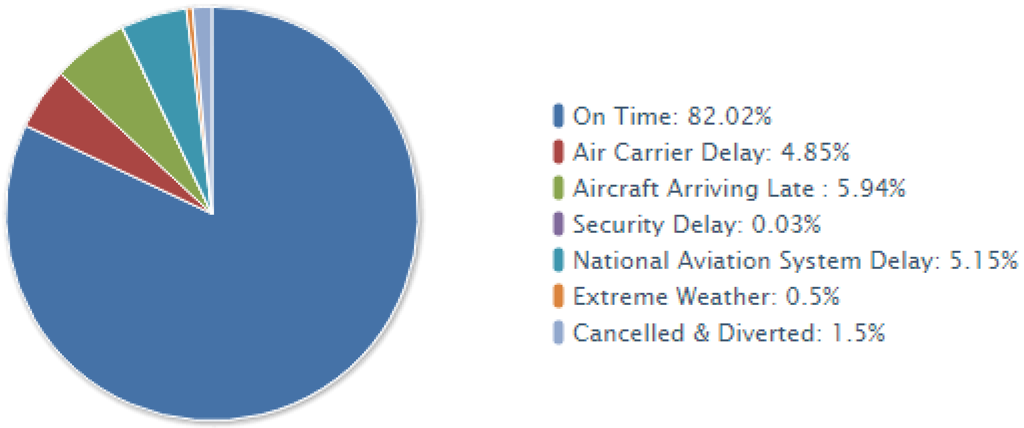
Aravinda Jatavallabha, Aadithya Naresh, Jacob Gerlach

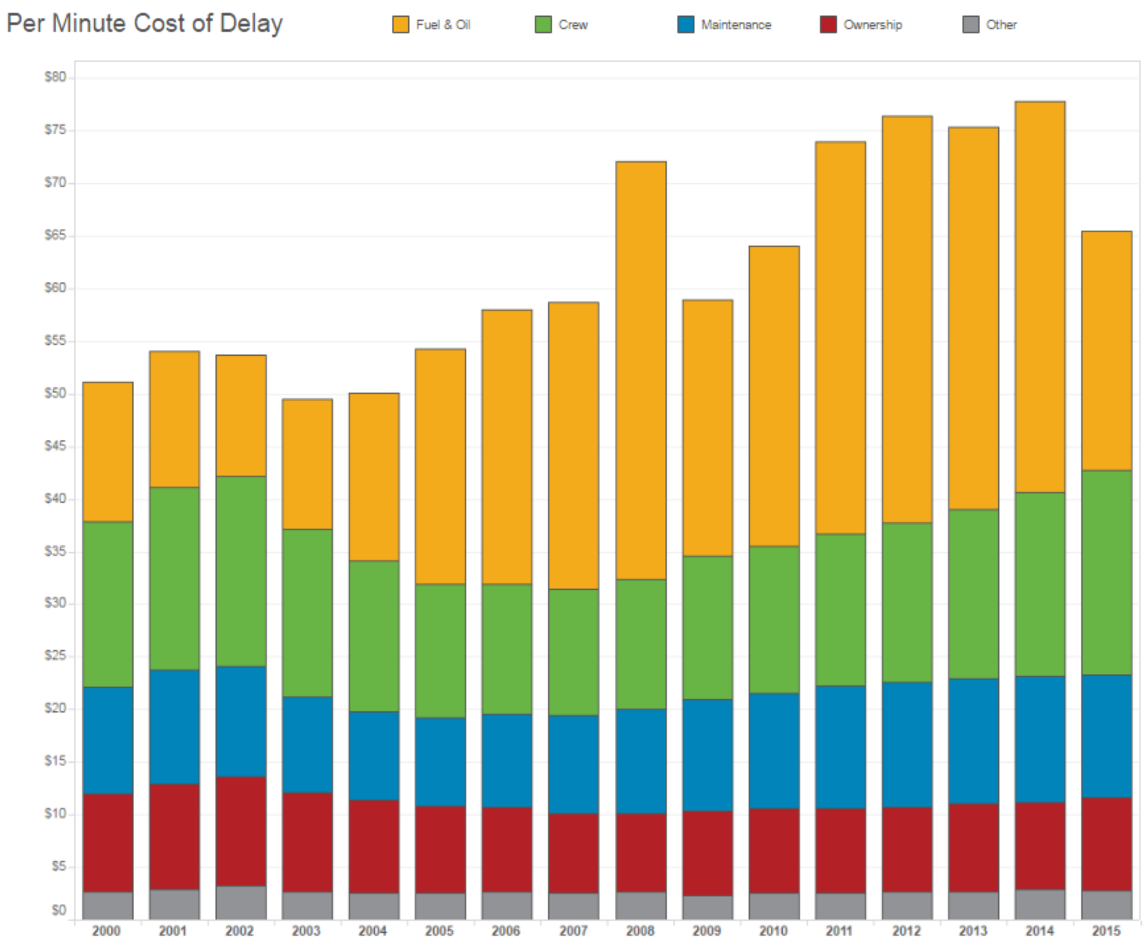
North Carolina State University, Raleigh

Introduction

- Flight punctuality is a critical aspect of airport and airline service quality, but flight delays, both in arrival and departure, pose substantial challenges impacting operational efficiency and customer satisfaction [1].
- The FAA estimated that flight delays cost the aviation industry $33 billion annually in 2019, highlighting the substantial economic impact [2].
- Delays also contribute to environmental concerns through increased fuel emissions.
- Predicting flight delays is essential for proactive planning and resource allocation.
- Considering the intricacies of flight delay factors, it's clear that traditional methods may fall short in accurate predictions.

Understanding Flight Delay Factors and Costs

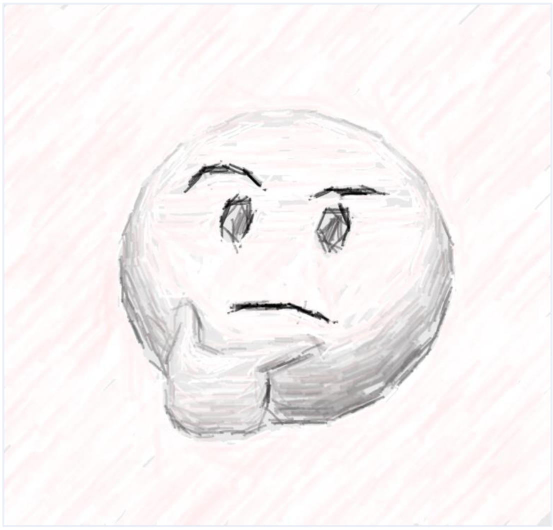




Delay due to various components [3]

Per Minute Cost of Delay [3]

Understanding of the Project



Goal

The goal of this project is to develop machine learning models that can individually predict the contribution of various factors, such as weather conditions, security issues, etc.

Usefulness

- Enable airlines, airports, and stakeholders to address specific issues effectively
- Accurately predict the contribution of each factor to flight delays
- Implement targeted strategies for delay mitigation and operational efficiency improvement

Related works

Authors	Work	Description
Tang, Y	Airline Flight Delay Prediction Using Machine Learning Models [4].	The study evaluated machine learning models using JFK airport data, achieving high accuracy rates of 97.78% with Decision Trees, 92.40% with Random Forest, and 93.34% with Gradient Boosted Trees. Future research aims to refine data imbalance and explore alternative ensemble techniques for improved precision.
Dou, X.	Flight arrival delay prediction and analysis using ensemble learning [5].	Introduced a method using Support Vector Regressor (SVR) to predict flight delays at U.S airports. By organizing and sampling data month by month, they identified 15 key features using cat-boost and employed regression models to predict specific delay times.

Related works

Authors	Work	Description
Esmaeilzadeh, E., & Mokhtarimousavi, S.	Machine Learning Approach for Flight Departure Delay Prediction and Analysis [6].	In this study, authors used Support Vector Machine (SVM) to explore factors influencing air traffic delays at three major New York City airports. They analyzed various explanatory variables to uncover relationships with departure delays, airport operations, and flow management, gaining deeper insights into delay causes.
Li, Q., Guan, X., & Liu, J.	A CNN-LSTM framework for flight delay prediction [7].	Authors proposed a CNN-LSTM deep learning framework to address flight delay prediction complexities. Integrating Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) architectures, their model captured spatial and temporal correlations, achieving 92.39% accuracy on U.S. domestic flights in 2019.

Related works

Authors	Work	Description
Nathalie Kuhn & Navaneeth Jamadagni	Application of Machine Learning Algorithms to Predict Flight Arrival Delays [8].	In this study, authors used Decision Tree, Logistic Regression and Neural Network model to classify the data of U.S. domestic flights in 2017 and predict flight delay. An F1-score of 0.91 was obtained for all the models.
A. Anusha, B. Ratna Kumar, D. Naga Ganesh Reddy, D. Sandeep Kumar, K. Shankar Nayak	Prediction Of Flight Delay using machine learning [9].	The authors have performed a classification task utilizing machine learning algorithms like Random Forest, Multi Layer Perceptron, Naive Bayes Classifier, KNN. The performance of the Random Forest algorithm surpasses other algorithms used achieving an accuracy of 83%.

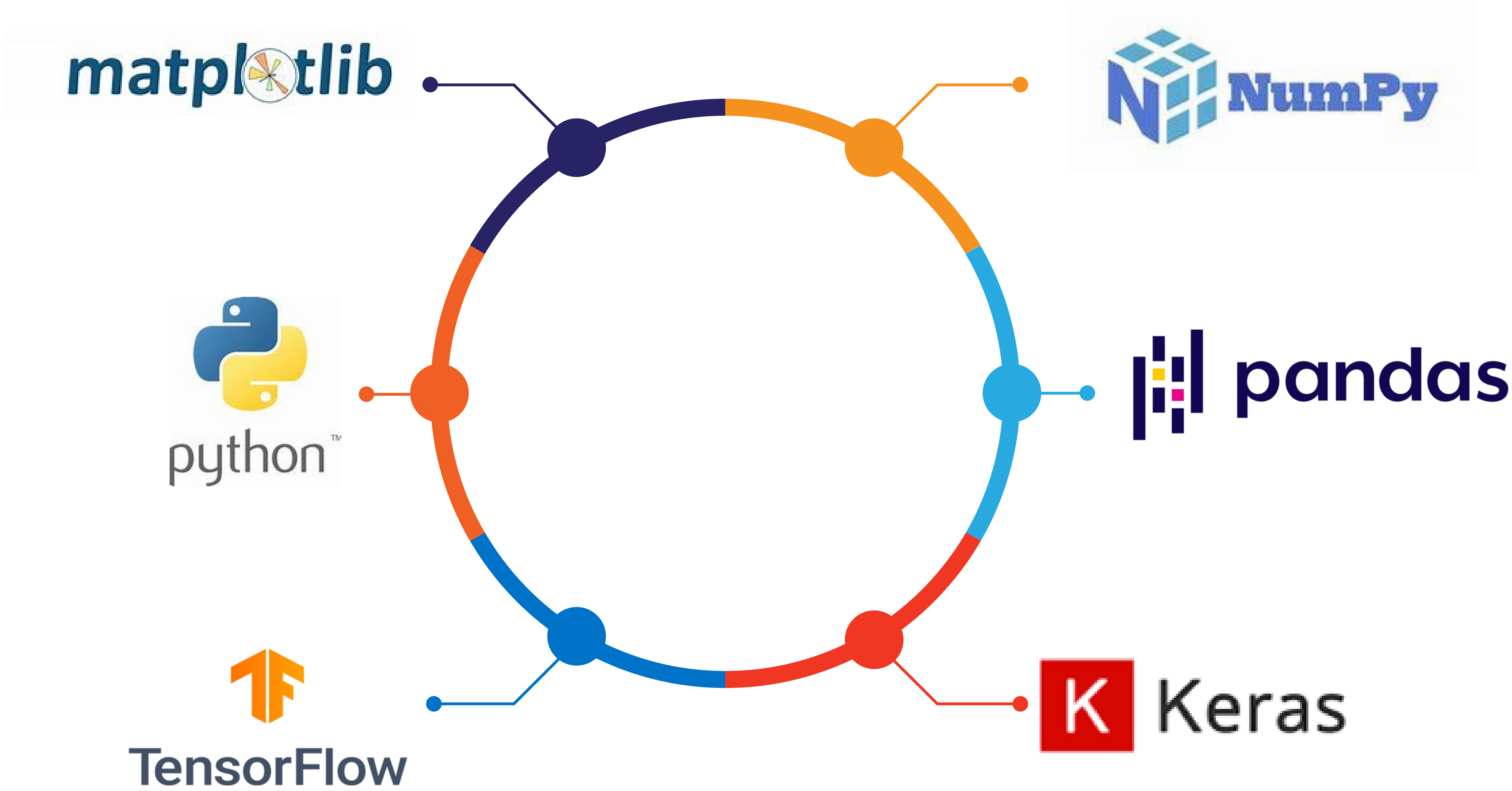
Novelty

- ❑ Regression techniques
- ❑ Individual Delay Components
    - Factors: Security delay, Weather delay, NAS Delay, Carrier Delay, Aircraft Delay.
- ❑ Application of Time-Series Architectures to utilize temporal dependencies in the data.

Method

- ❑ Selection of advanced time series models: LSTM, Bi-LSTM, and LSTM+CNN chosen for their proficiency in capturing complex temporal patterns.
- ❑ Incorporation of baseline regression models: Multiple Regression, Random Forest Regression, Decision Tree Regression, XGBoost, and Artificial Neural Network for benchmarking.
- ❑ Advantages of LSTM and Bi-LSTM: adeptness in handling vanishing gradient problems and capturing long-term dependencies.
- ❑ Mitigation of vanishing gradient issue: employment of techniques like gradient clipping and batch normalization in LSTM models.
- ❑ Expected superiority of time series models, especially LSTM and its variants, in predicting flight delays due to their capacity to capture temporal dependencies, offering valuable insights into machine learning approaches for flight delay prediction.

Tools and Technologies used



Dataset Description

Data Source:

- We’ve used data obtained from the US Department of Transportation, Bureau of Transportation Statistics from January 2019 – August 2023 [10]

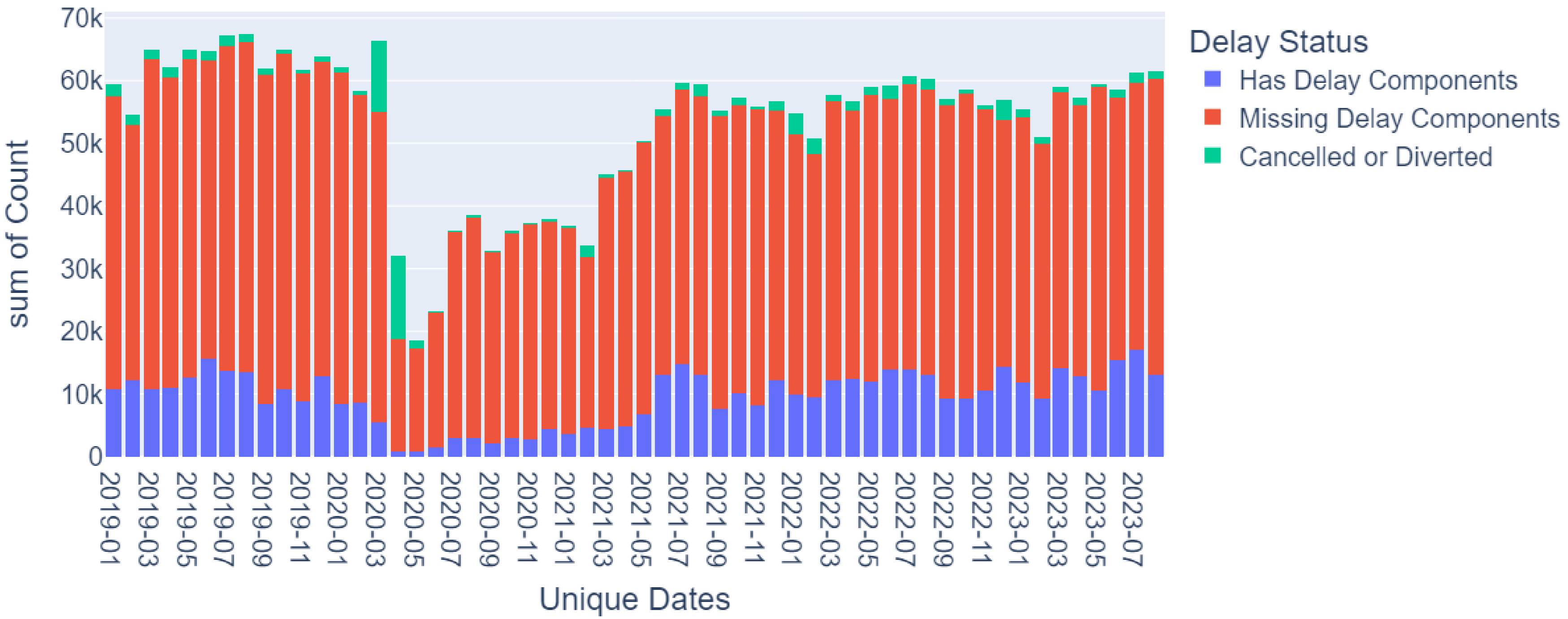
- It contains 32 attributes related to planned flight date-time, airline, planned origin and destination, cancellation and diversion status, overall delay, and delay due to individual components (carrier, weather, NAS, security, late aircraft)

Understanding the Data

FL_DATE	AIRLINE	ORIGIN	DEST	CRS_DEPT_TIME	TAXI_IN	TAXI_OUT	CRS_ARR_TIME	DISTANCE
2019-01-09	United Air Lines Inc.	FLL	EWR	1155	19.0	4.0	1501	1065.0
2022-11-19	Delta Air Lines Inc	MSP	SEA	2120	9.0	38.0	2315	1399.0
2022-07-22	United Air Lines Inc.	DEN	MSP	954	20.0	5.0	1252	680.0
2020-02-23	Spirit Air Lines	MCO	DFW	1840	15.0	14.0	2041	985.0

Pre-Processing: Removing NULLs

The majority of the data was removed for having NULL values for all individual delay components. Leaving us with 533,863 records.



Flight Record Counts

Pre-Processing: Feature Selection

- **Pearson's Correlation** between 6 non-categorical independent attributes: CRS_DEP_TIME, TAXI_OUT, CRS_ARR_TIME, TAXI_IN, CRS_ELASPED_TIME DISTANCE and dependent variable: **ARR_DELAY**
- PCA on normalized data did not help increase correlation

NON-CATEGORICAL ATTRIBUTE	PEARSON'S CORRELATION
CRS_DEP_TIME	0.0704
TAXI_OUT	0.0541
CRS_ARR_TIME	0.0500
TAXI_IN	0.0235
***CRS_ELAPSED_TIME**	**-0.0122**
DISTANCE	-0.0229

***Removed (redundant)**

Pre-Processing: Feature Selection

- **Eliminated redundancies** in categorical attributes. Verified using The **Kruskal-Wallis H-test**

- ORIGIN / DEST had more unique values than ORIGIN_CITY / DEST_CITY

ORIGINAL CATEGORICAL ATTRIBUTE	KEPT CATEGORICAL ATTRIBUTE
AIRLINE	AIRLINE
AIRLINE_CODE	
DOT_CODE	
AIRLINE_DOT	
ORIGIN	ORIGIN
ORIGIN_CITY	
DEST	DEST
DEST_CITY	

Pre-Processing: Data Transformation

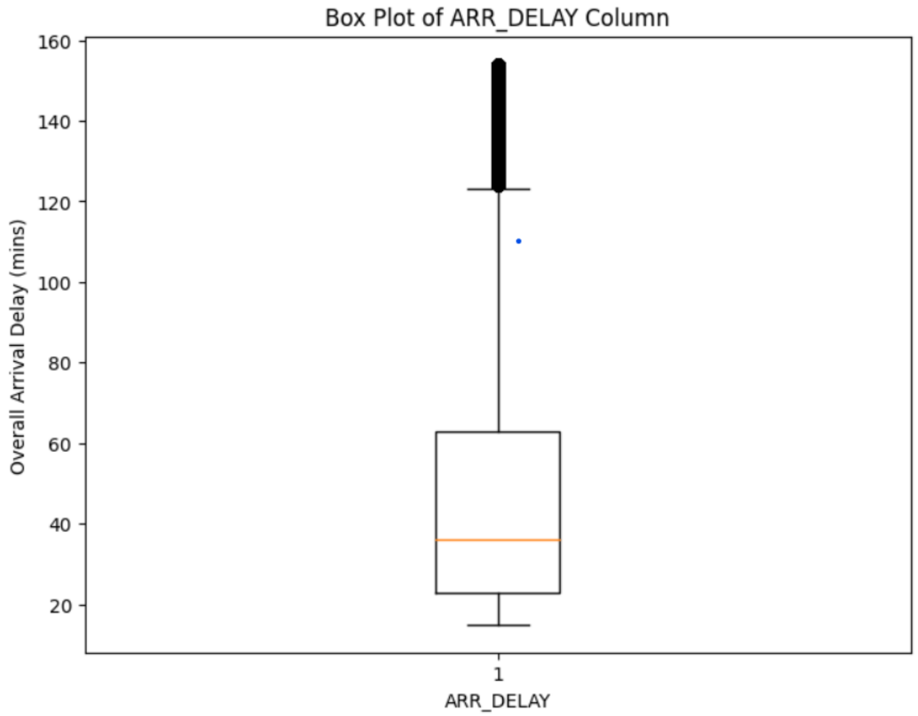
Independent Vars.	Description	Initial Format	Transformed format	Var. Type
CRS_DEP_TIME	Scheduled departure, local time	hhmm	Minutes past midnight	Discrete, Ratio
TAXI_OUT	Taxi out time	Minutes past midnight	Minutes past midnight	Discrete, Ratio
CRS_ARR_TIME	Scheduled arrival, local time	hhmm	Minutes past midnight	Discrete, Ratio
TAXI_IN	Taxi in time	Minutes past midnight	Minutes past midnight	Discrete, Ratio
DISTANCE	Distance between airports	Miles	Miles	Discrete, Ratio
FL_DATE	Flight date	yyyymmdd	3 Vars.: Year (label encoded), Month, Day	3 * Discrete, Ordinal
AIRLINE	Reporting airline	Name of airline	Label encoded	Discrete, Nominal
ORIGIN	Origin airport	3 digit code	Label encoded (with DEST)	Discrete, Nominal
DEST	Destination airport	3 digit code	Label encoded (with ORIGIN)	Discrete, Nominal

Pre-Processing: Outlier Pruning

Lower Bound: Q1 - 1.5*IQR

Upper Bound: Q3 + 1.5*IQR

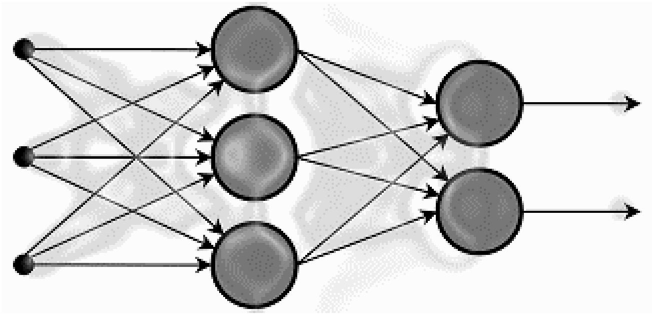
	Pre-Outlier Removal	Post-Outlier Removal
# of Records	533,863	489,828
Mean	67.526	47.828
Standard Dev.	93.909	32.869
Minimum	15	15
Maximum	2934	154



Pre-Processing: Split and Normalization

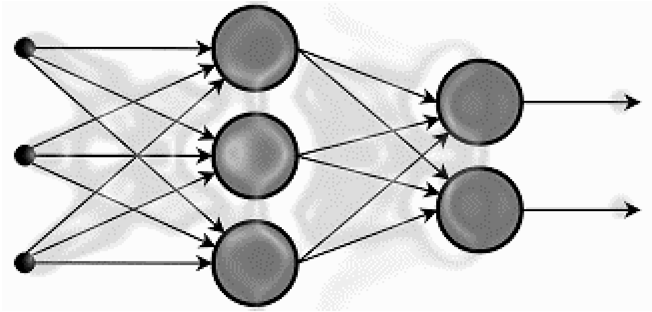
- 75%/25% train/test split
- Recall: Categorical attributes were encoded into integer labels
  - ORIGIN and DEST were encoded together to ensure consistent relationship
- Non-categorical attributes were Z-Score normalized
  - Fitted on train, applied to test

Modeling



- The baseline machine learning algorithms used in this study are:
  - **Random Forest Regressor**
  - **Multiple Regression**
  - **Decision Tree Regression**
  - **XGBoost Regressor**
  - **Artificial Neural Network (ANN)**

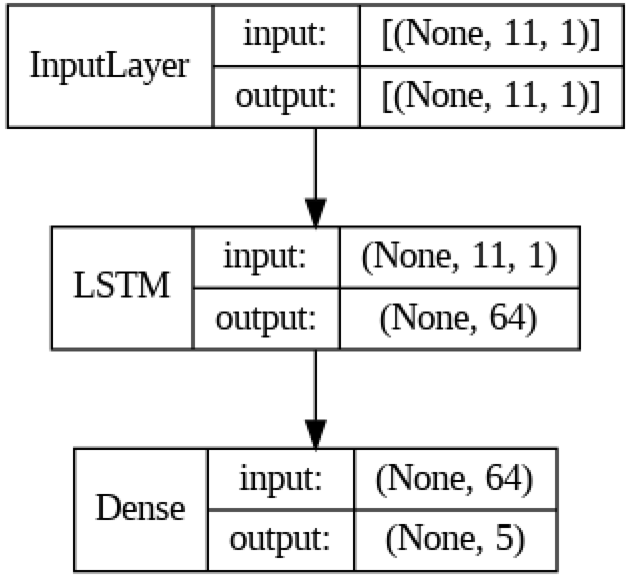
Modeling



- The time-series machine learning algorithms used in this study are:
  - **Long Short-Term Memory (LSTM)**
  - **Bidirectional Long Short-Term Memory (BiLSTM)**
  - **LSTM + CNN Hybrid Model**

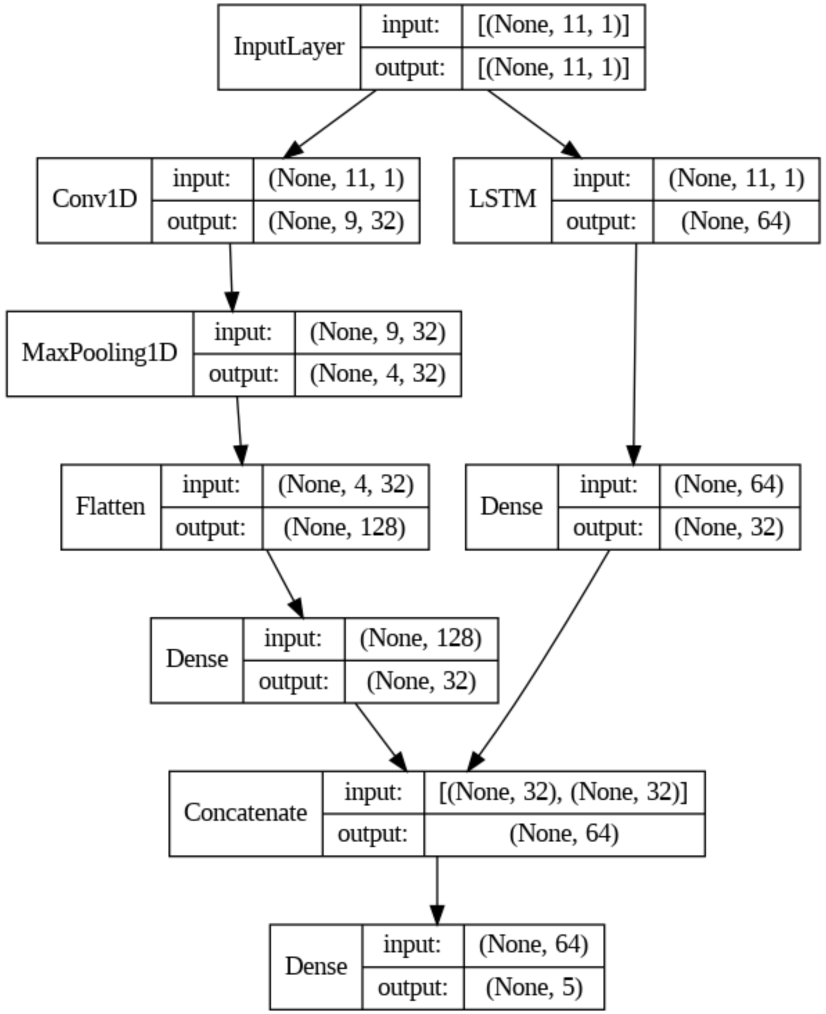
Time-Series Architecture

LSTM



Hand tuned batch size (265) and LSTM hidden nodes (64)

LSTM/CNN Hybrid



Training and Evaluation Metrics

Mean Absolute Error (MAE)

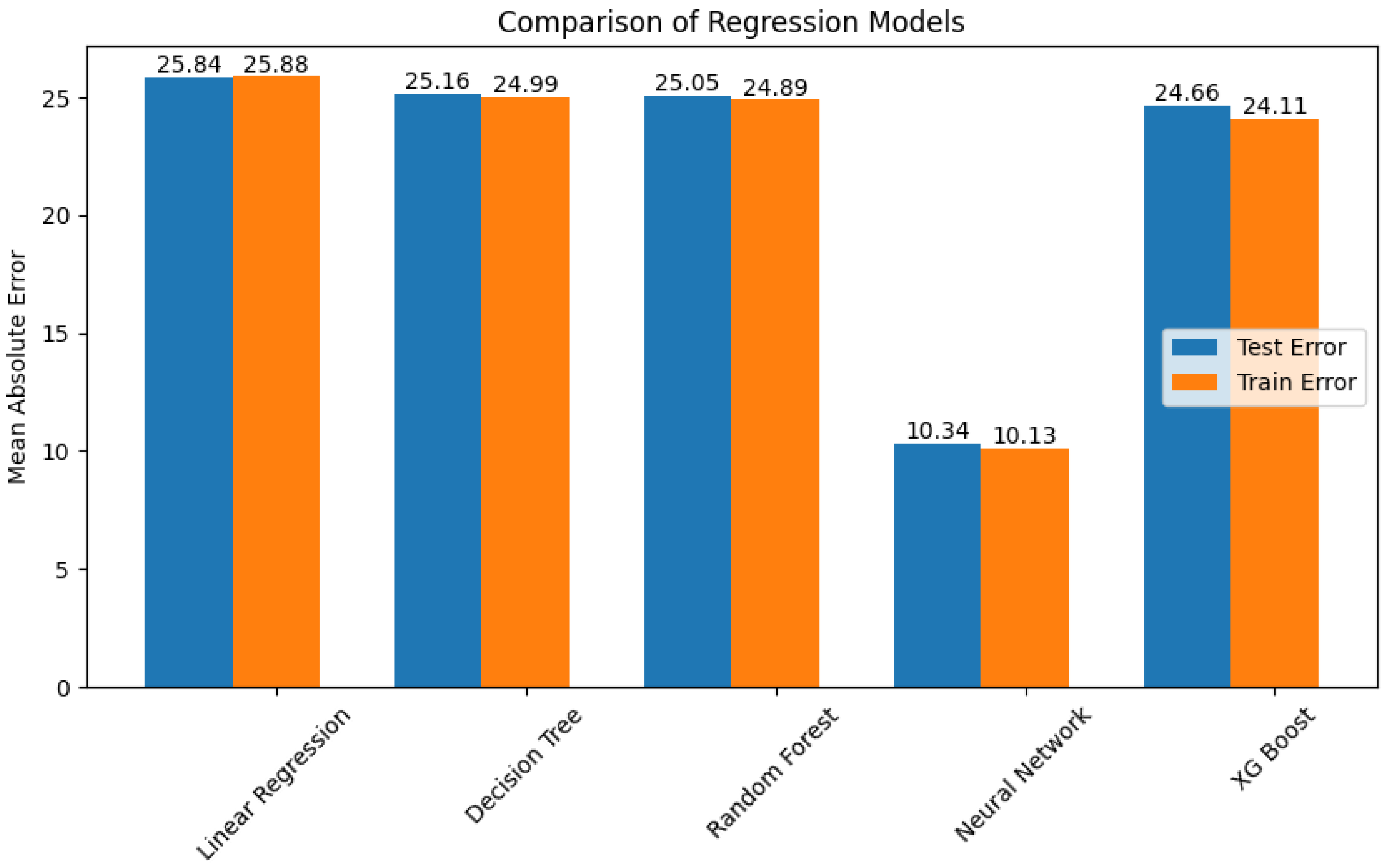
$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_{\text{pred},i} - y_{\text{true},i}|$$

Mean Squared Error (MSE)

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n} (y_i - \tilde{y}_i)^2$$

- MAE quantifies the average magnitude of errors between predicted and actual values
  - Useful for interpreting our results because it is measured in **minutes**
- MSE quantifies the average squared magnitude of errors between predicted and actual values
  - Useful for training models because it gives higher weighting to lower frequency values, and **we have a lot of zeros**

Results for Baseline Models



Time-Series Results

Total Model Error

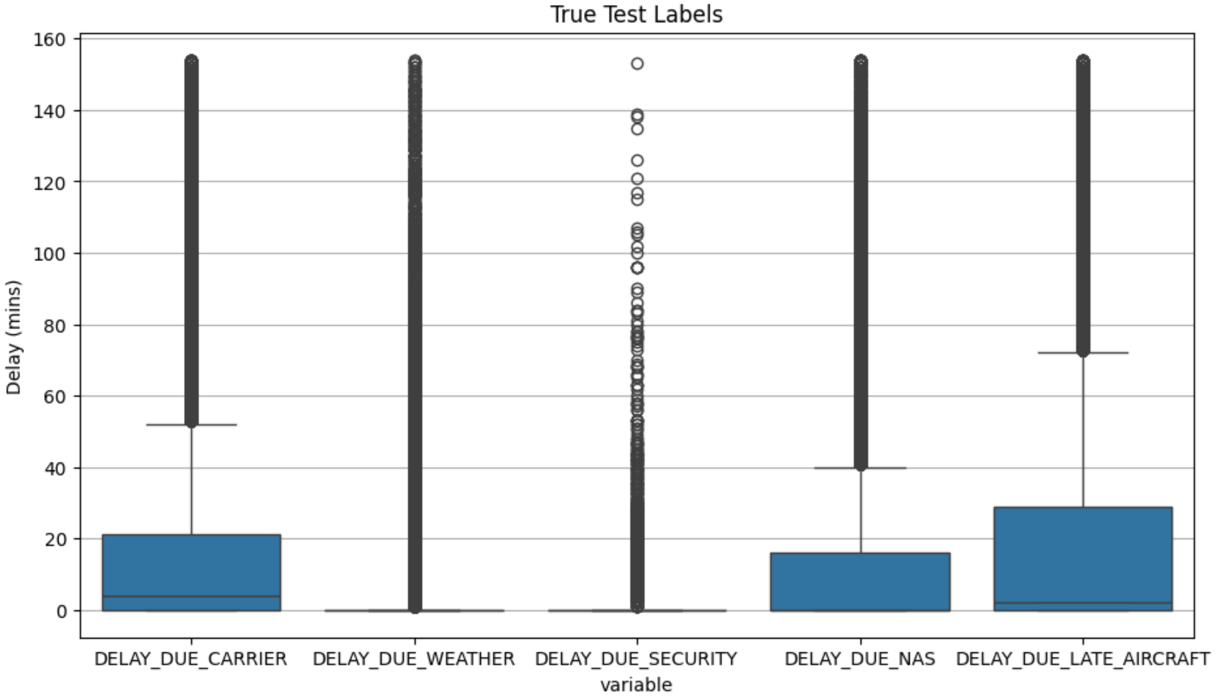
	Mean Squared Error	Mean Absolute Error
LSTM	361.833	10.022
Bidirectional LSTM	365.645	10.296
LSTM + CNN Hybrid	367.868	9.684
*Neural Network	371.4737	10.340

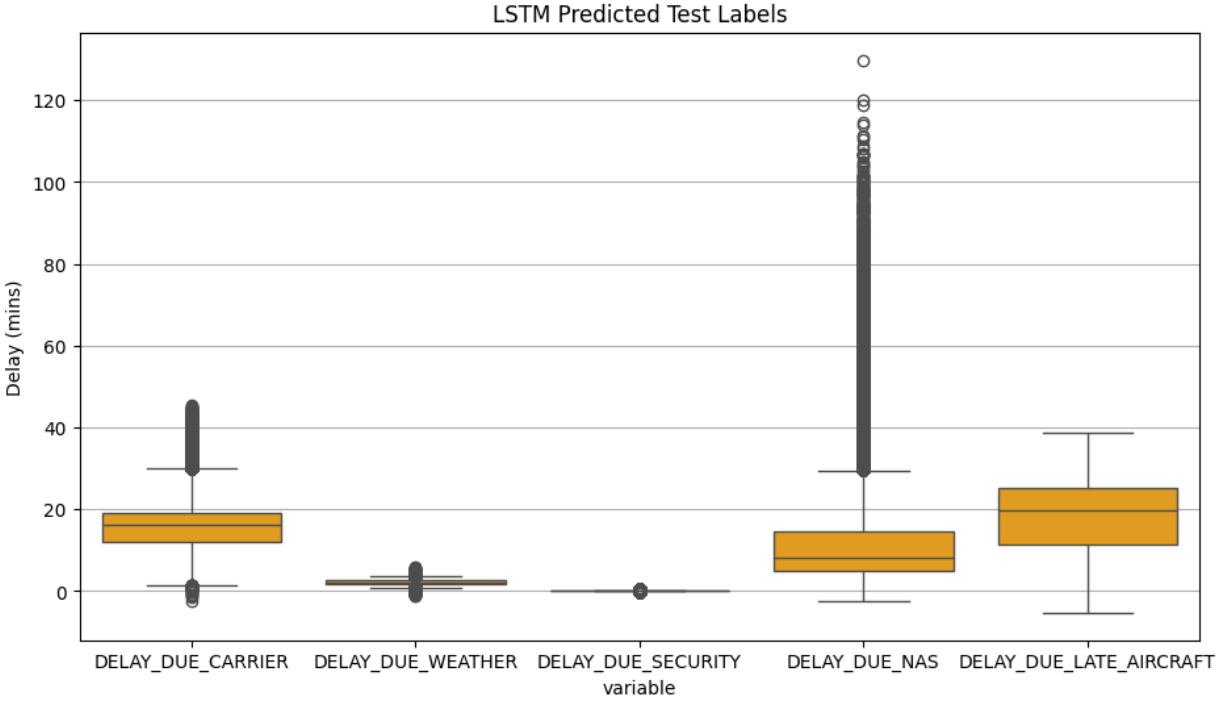
*Best non-time-series model

LSTM Test Results

Delay Component	True Mean	Mean of Predictions	Mean Absolute Error
Carrier	15.877	16.288	17.050
Weather	1.978	2.246	3.979
Security	0.130	0.141	0.270
NAS	10.914	11.512	9.475
Late Aircraft	19.374	18.031	19.338

LSTM: Test True vs. Predicted





	Carrier	Weather	Security	NAS	Late Aircraft
True Std.	25.567	11.406	2.423	19.660	29.453
Predictions Std.	6.164	0.618	0.060	10.536	8.720

Discussion and Future Scope

Although all of our time-series approaches performed better than our baseline regression models, they do not provide meaningful predictions.

Possible issues

- Effect of COVID-19 on data
- Outlier values

Prior work involved classification on overall delay, our work involved regression on delay components

Future

- Handling COVID-19 data
- Different Outlier removal methods
- Delay components may be too complicated of a task, maybe consider total delay

References

[1] Carvalho, A. Sternberg, L. Maia Gonçalves, A. Beatriz Cruz, J.A. Soares, D. Brandao, D. Carvalho, e E. Ogasawara, 2020, On the relevance of data science for flight delay research: a systematic review, Transport Reviews

[2] F. Wieland. Limits to growth: results from the detailed policy assessment tool [air traffic congestion]. In 16th DASC. AIAA/IEEE Digital Avionics Systems Conference. Reflections to the Future. Proceedings, volume 2, pages 9.2–1–9.2–8 vol.2, Oct. 1997.

[3] Team, T. (2016, August 31). What is the impact of flight delays? Trefis. https://www.trefis.com/stock/dal/articles/375013/what-is-the-impact-of-flight-delays/2016-08-31

[4] Tang, Y. (2022). Airline Flight Delay Prediction Using Machine Learning Models. In Proceedings of the 2021 5th International Conference on E-Business and Internet (ICEBI '21) (pp. 151–154). Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3497701.3497725

[5] Dou, X. (2020). Flight arrival delay prediction and analysis using ensemble learning. In 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC) (pp. 1-4). New York: IEEE. https://doi.org/10.1109/ITNEC48623.2020.9084929

References

[6] Esmaeilzadeh, E., & Mokhtarimousavi, S. (2020). Machine Learning Approach for Flight Departure Delay Prediction and Analysis. Transportation Research
Record, 2674(8), 145-159. https://doi.org/10.1177/0361198120930014

[7] Li, Q., Guan, X., & Liu, J. (2023). A CNN-LSTM framework for flight delay prediction. Expert Systems with Applications, 227, 120287. https://doi.org/10.1016/j.eswa.2023.120287

[8] Nathalie Kuhn, Navaneeth Jamadagni (2017). Application of Machine Learning Algorithms to Predict Flight Arrival Delays, Stanford University.

[9]A. Anusha,B. Ratna Kumar,D, et al., (2022). Prediction of Flight Delay Using Machine Learning. IRJMETS e-ISSN: 2582-5208.

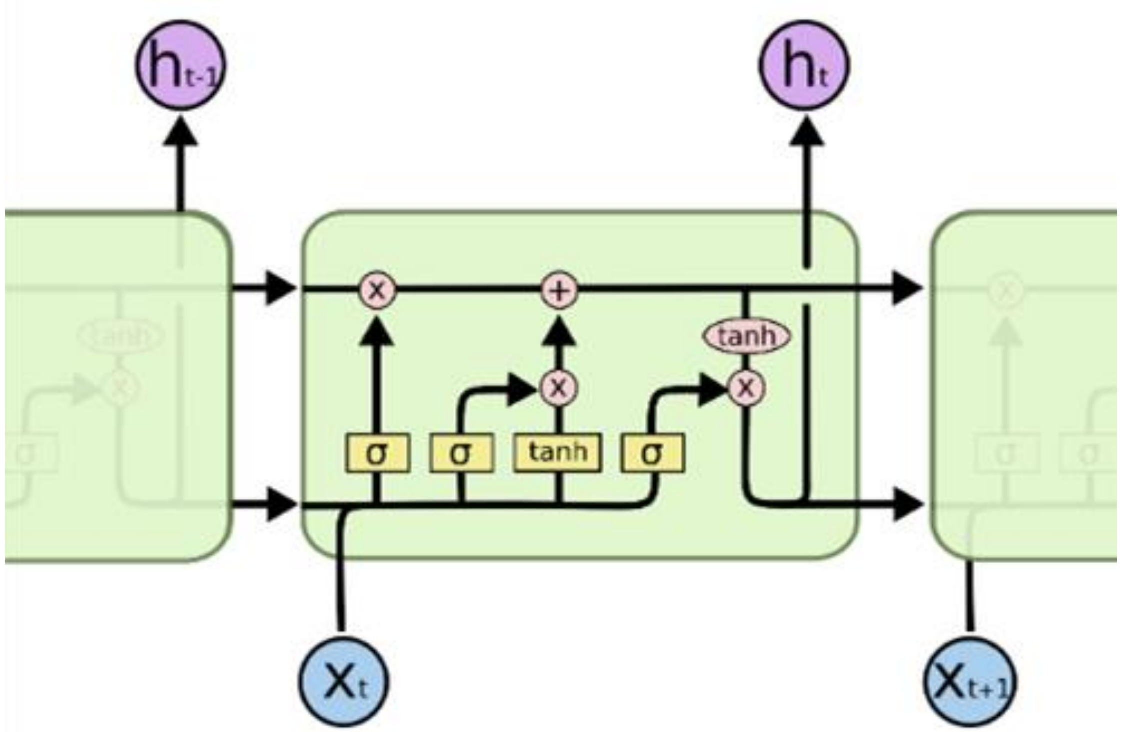
[10] https://www.kaggle.com/datasets/patrickzel/flight-delay-and-cancellationdataset-2019-2023?select=flights_sample_3m.csv

THANK YOU

APPENDIX

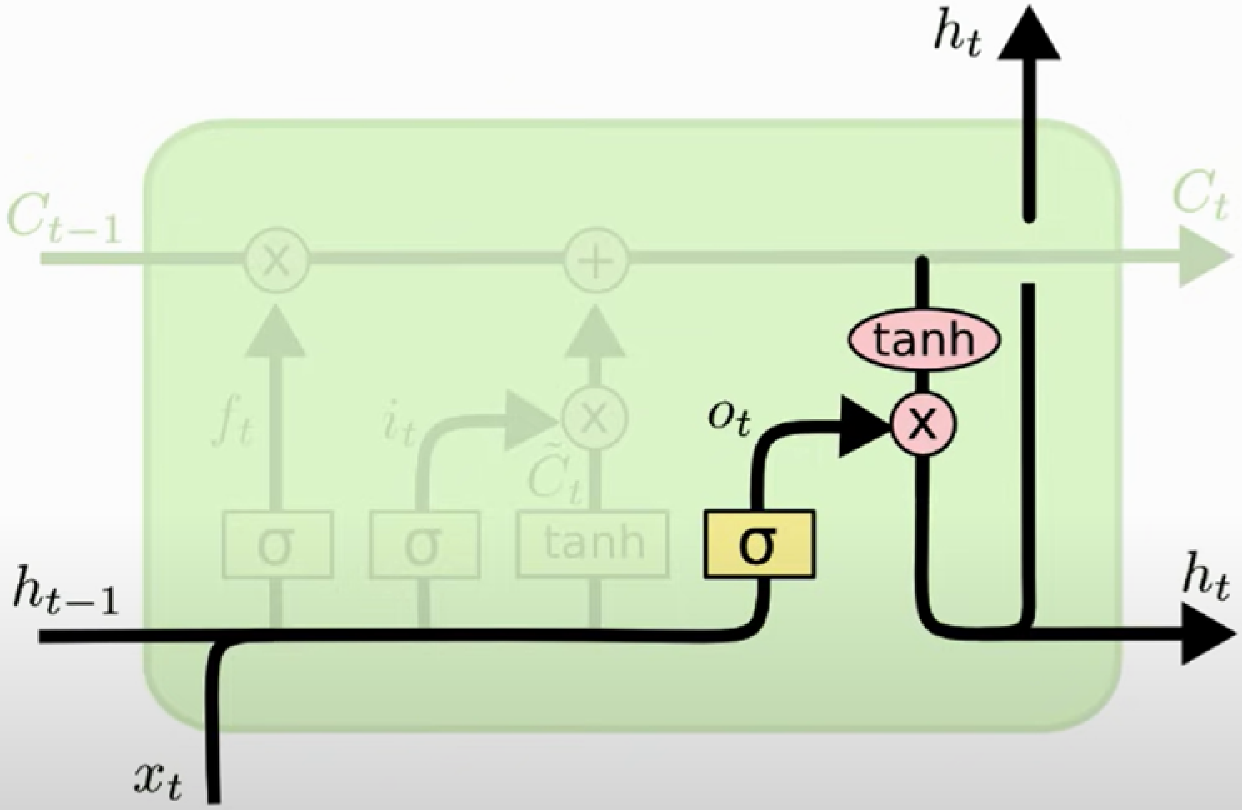
LSTM Model

- Long Short-Term Memory (LSTM) models provide a powerful framework for modeling sequential data, such as time series prediction tasks like flight delay prediction.
- LSTMs are a type of recurrent neural network (RNN) architecture designed to address the vanishing gradient problem and capture long-term dependencies in sequential data effectively.



The repeating module in an LSTM contains four interacting layers. [7]

LSTM



Three Gates of LSTM Cell:

- Input Gate — Is Cell Updated?

$$i^{(t)} = \sigma\left(W^i\left[h^{(t-1)}, x^{(t)}\right] + b^i\right)$$

- Forget Gate — Is memory set to 0?

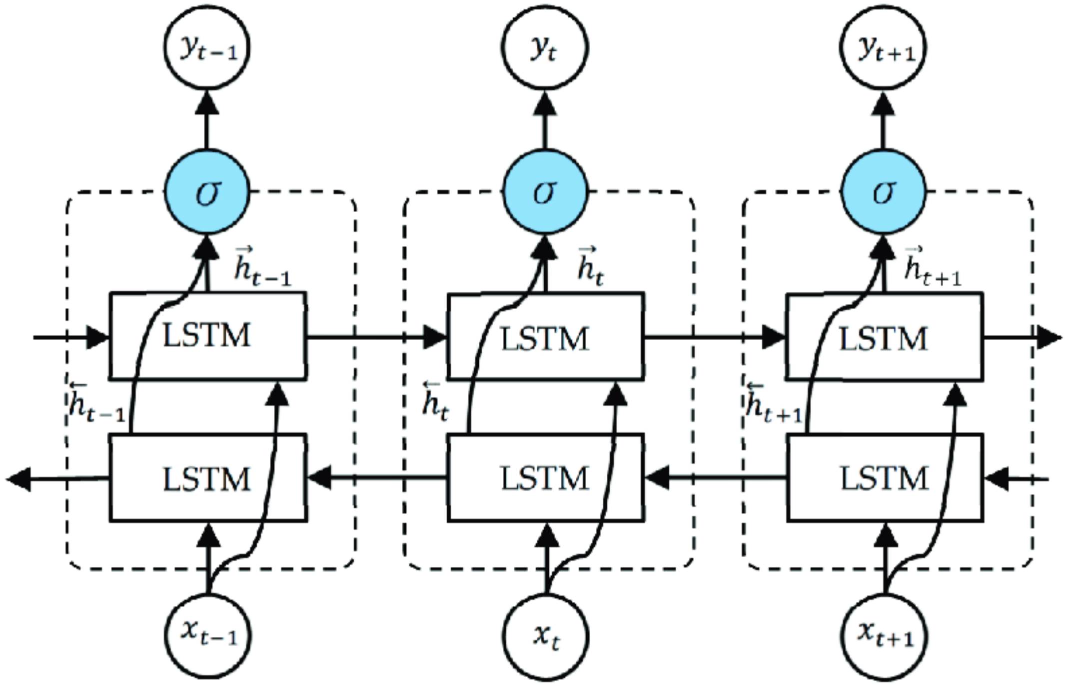
$$f^{(t)} = \sigma\left(W^f\left[h^{(t-1)}, x^{(t)}\right] + b^f\right)$$

- Output Gate — Is current info visible?

$$o^{(t)} = \sigma\left(W^o\left[h^{(t-1)}, x^{(t)}\right] + b^o\right)$$

BiLSTM Model

- Bidirectional Long Short-Term Memory (BiLSTM) models extend the capabilities of LSTM by processing input sequences in both forward and backward directions.
- This allows the model to capture contextual information from both past and future time steps, making it particularly effective for tasks requiring a comprehensive understanding of sequence data.



The unfolded architecture of Bidirectional LSTM (BiLSTM) with three consecutive steps.