Evaluating RNN Performance in Unlabeled Time-Series Anomaly Detection

Applications in NYC Taxi Volume Data

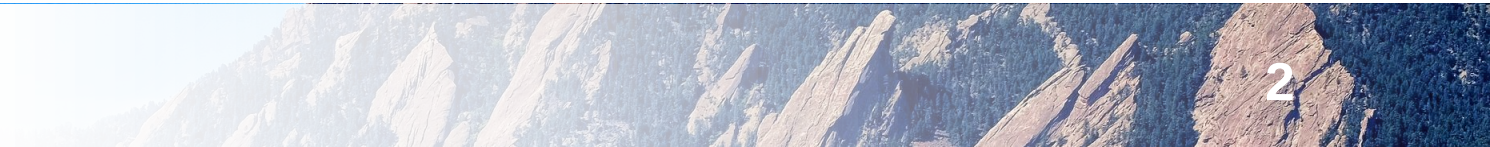
Nick Vastine

DTSA5511 – Intro to Deep Learning
Final Project
April 29, 2025

Data Science
UNIVERSITY OF COLORADO **BOULDER**

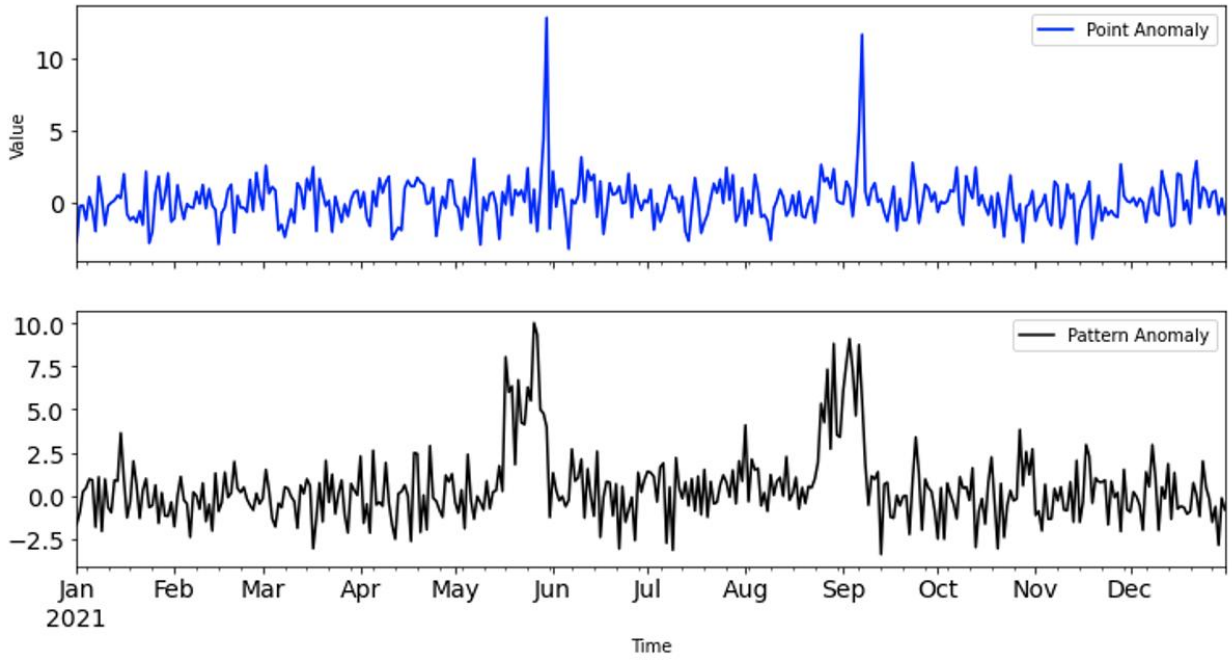
Agenda

- **Problem Statement**
- **Data**
- **Exploratory Data Analysis** – Plotting, Distributions
- **Modeling**
  - Hyperparameter tuning
  - Convolution Layer
  - Feature Engineering
- **Results** & **Analysis**
- **Conclusion** & **Future Work**



Problem Statement

- Use RNN's to detect anomalies in unlabeled, time-series data using prediction error



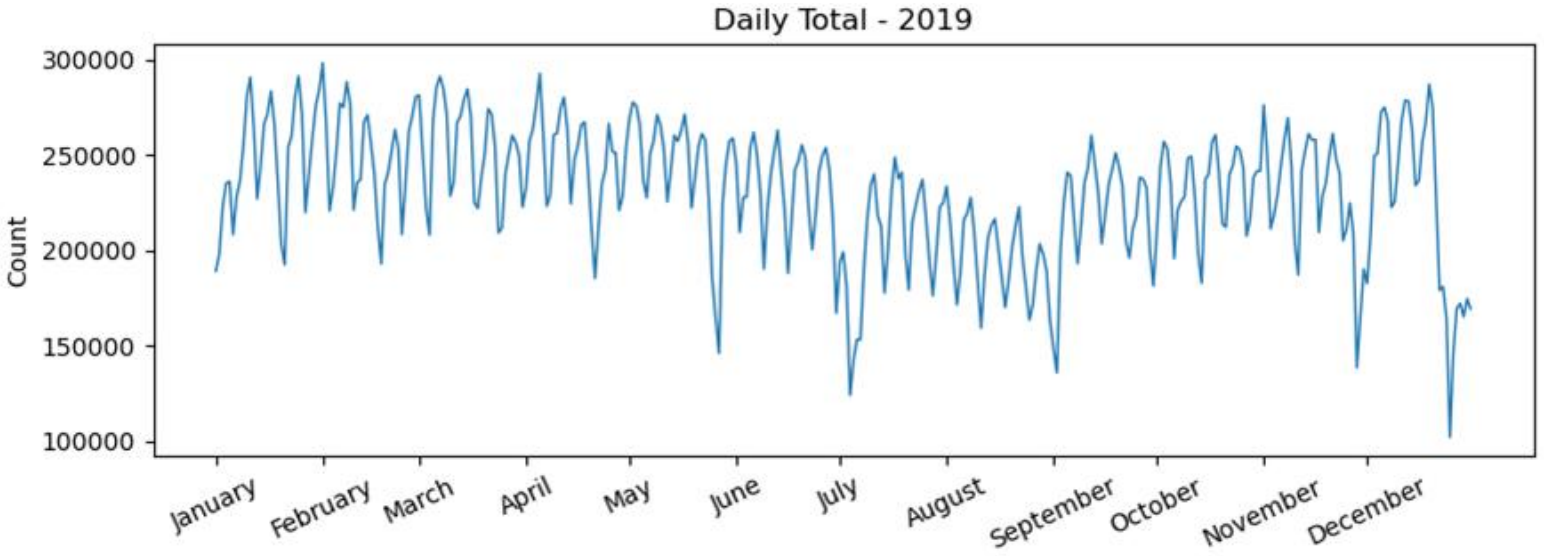
- Evaluate RNN variations
  - Hyperparameter Tuning
    - Sequence Length
    - Number of RNN Units
  - 1-D Convolutional Layer
  - Feature Engineering
    - Rolling Average, Rate of Change
    - Season-Trend Decomposition using LOESS (STL)

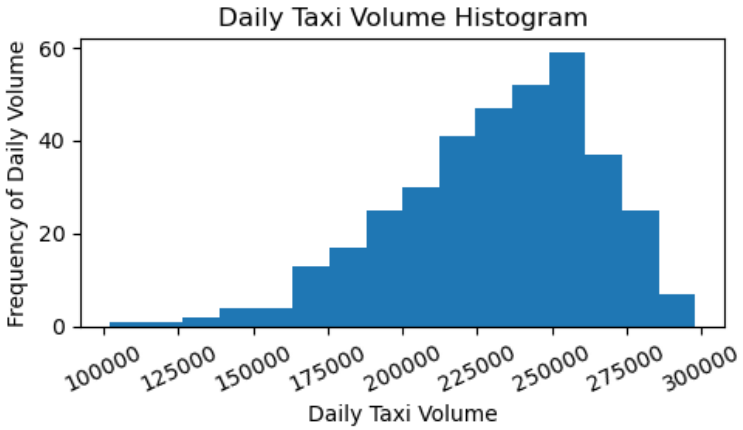
Data

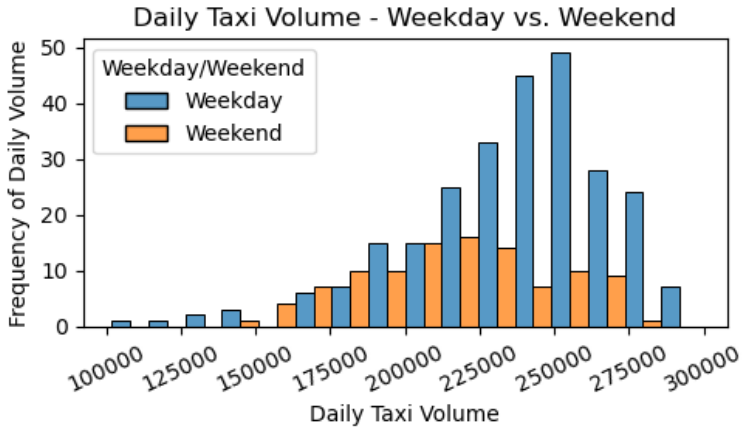
- Data was sourced from NYC Open Data, reporting every individual Yellow Cab taxi trip in 2019.
  - Data should show weekly cycles, long term trends, and anomalies from holidays like Thanksgiving or Christmas.
- Data was aggregated into daily totals using PostgreSQL.

	month	day	count	datetime	weekday
0	1	1	189035	2019-01-01	1
1	1	2	197852	2019-01-02	2
2	1	3	222879	2019-01-03	3
3	1	4	235053	2019-01-04	4
4	1	5	236041	2019-01-05	5
5	1	6	208194	2019-01-06	6
6	1	7	227762	2019-01-07	0

Exploratory Data Analysis



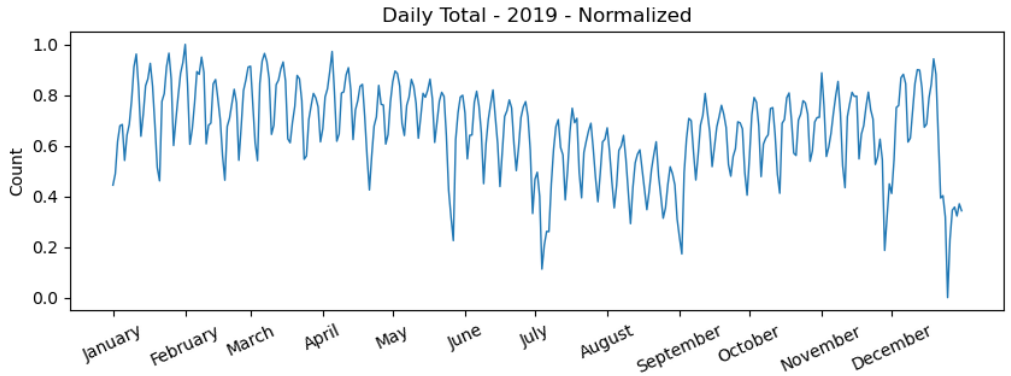


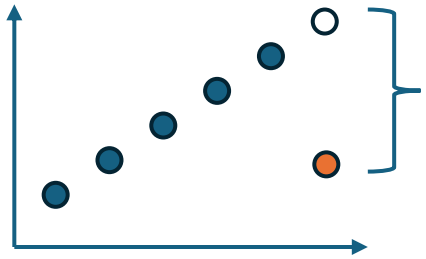


- Daily plot reveals anomalies for model to identify.
- Distribution plots emphasize variation between weekday and weekend.

Modeling – Detecting Anomalies

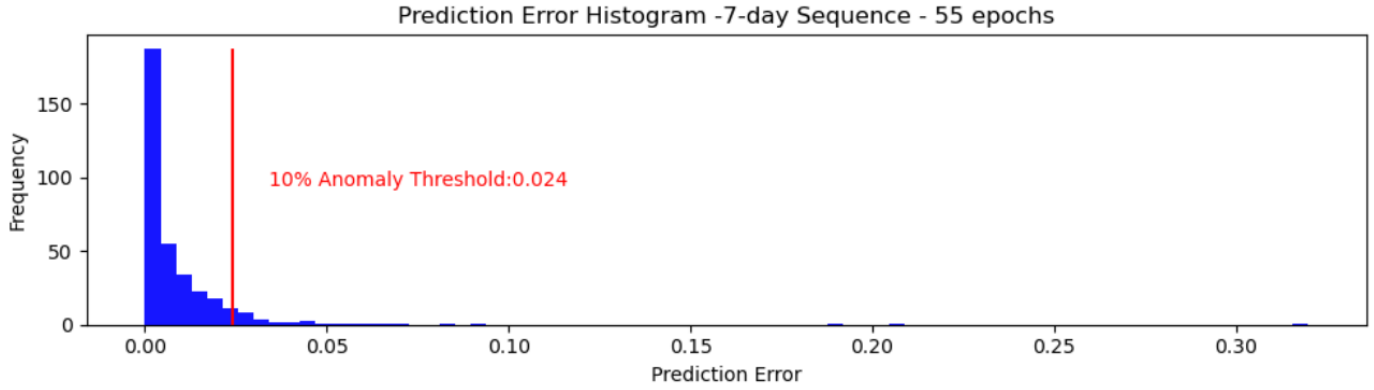
- Data is normalized and sequenced





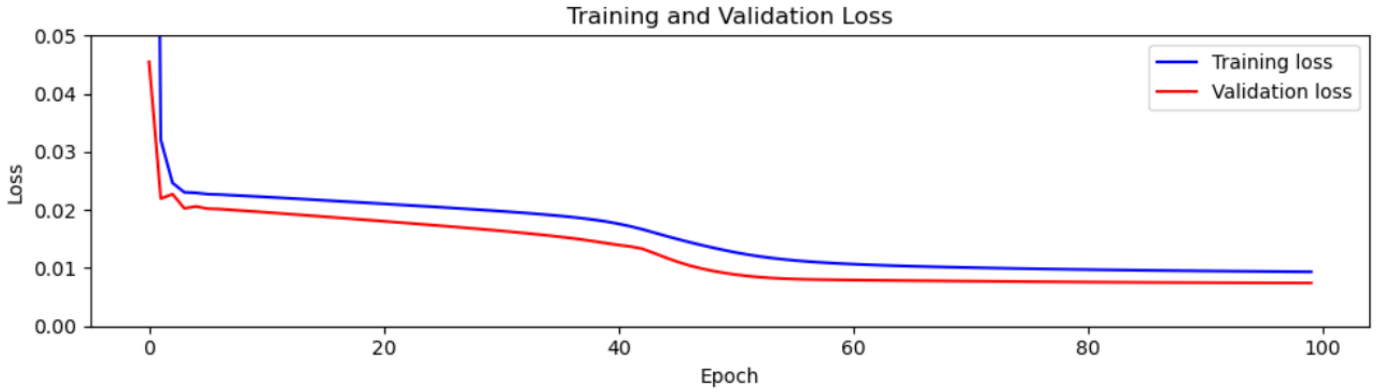
Prediction Error
Larger for anomalies

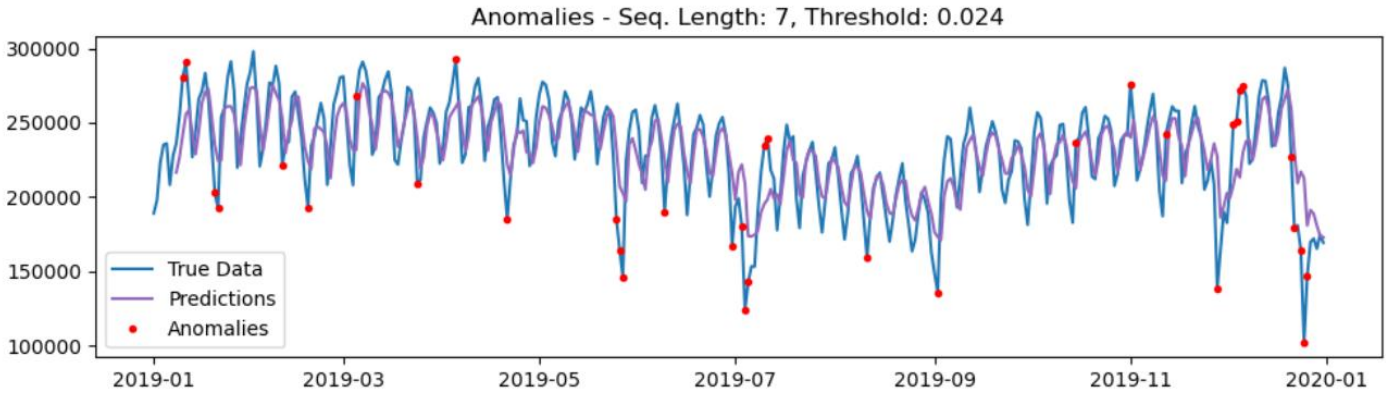
- The distribution of prediction errors can label points as anomalies using an *anomaly threshold.*



Modeling – Managing overfitting

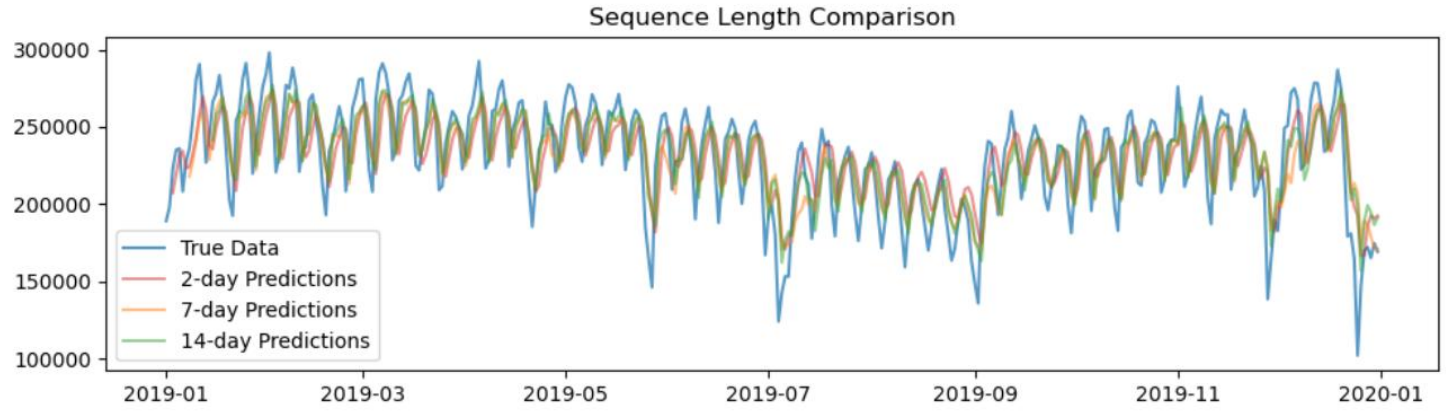
- **Base Model** – 50 RNN units + 1 Dense unit to predict

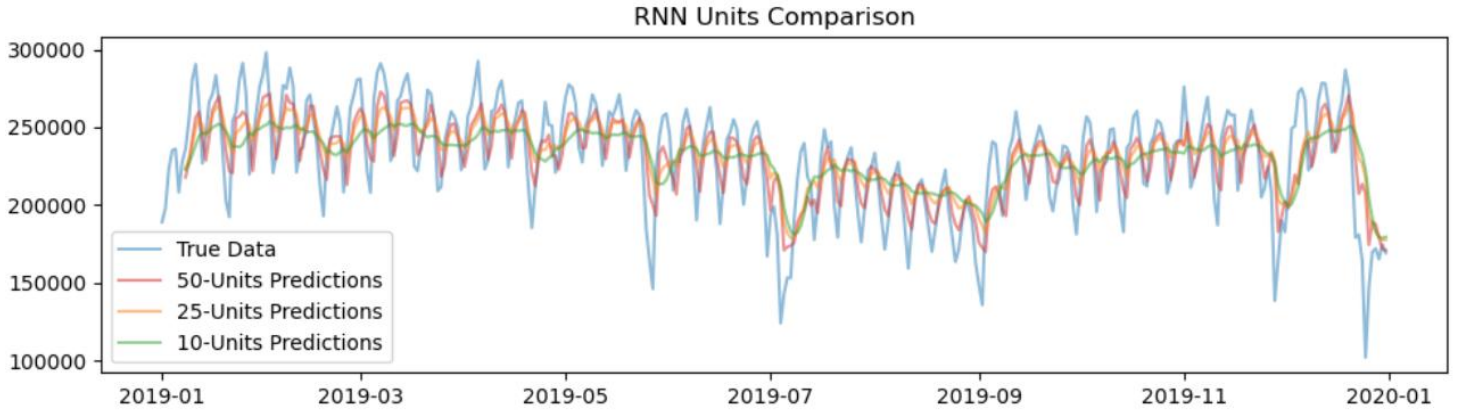




- Avoid overfitting!
  - Overfitting predicts all points precisely, even anomalies.
  - However, variations can improve training rate which is useful for complex problems.
- Models are run for 55 epochs for consistency.

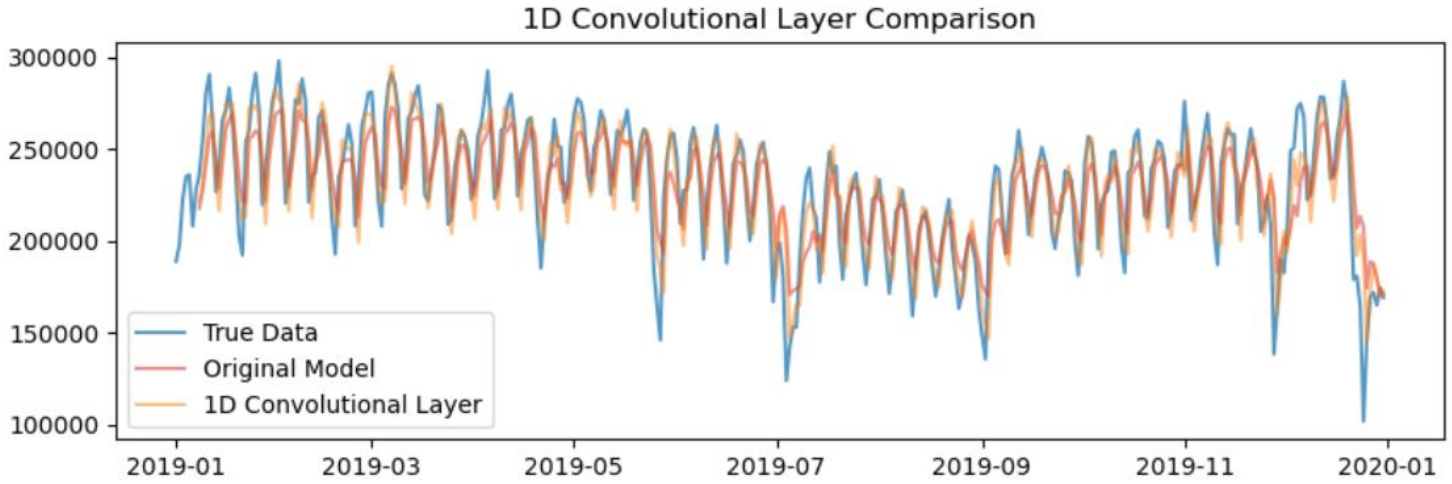
Modeling - Hyperparameters

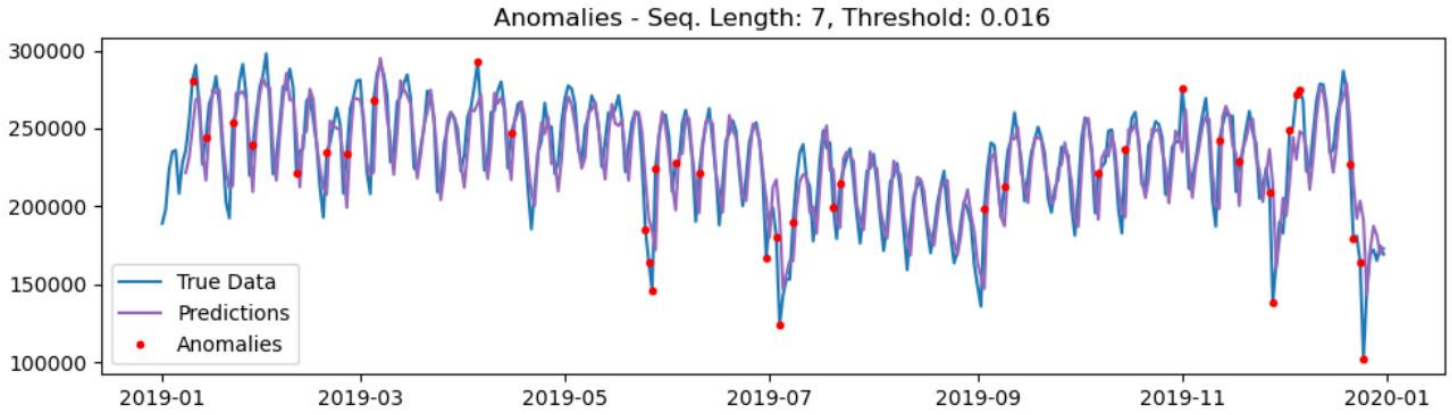




- Varying sequence length controls the points leading up to the prediction
  - Longer better, but barely so.
- Decreasing RNN reduces performance after decreasing past minimum.

Modeling – Convolutional Layer

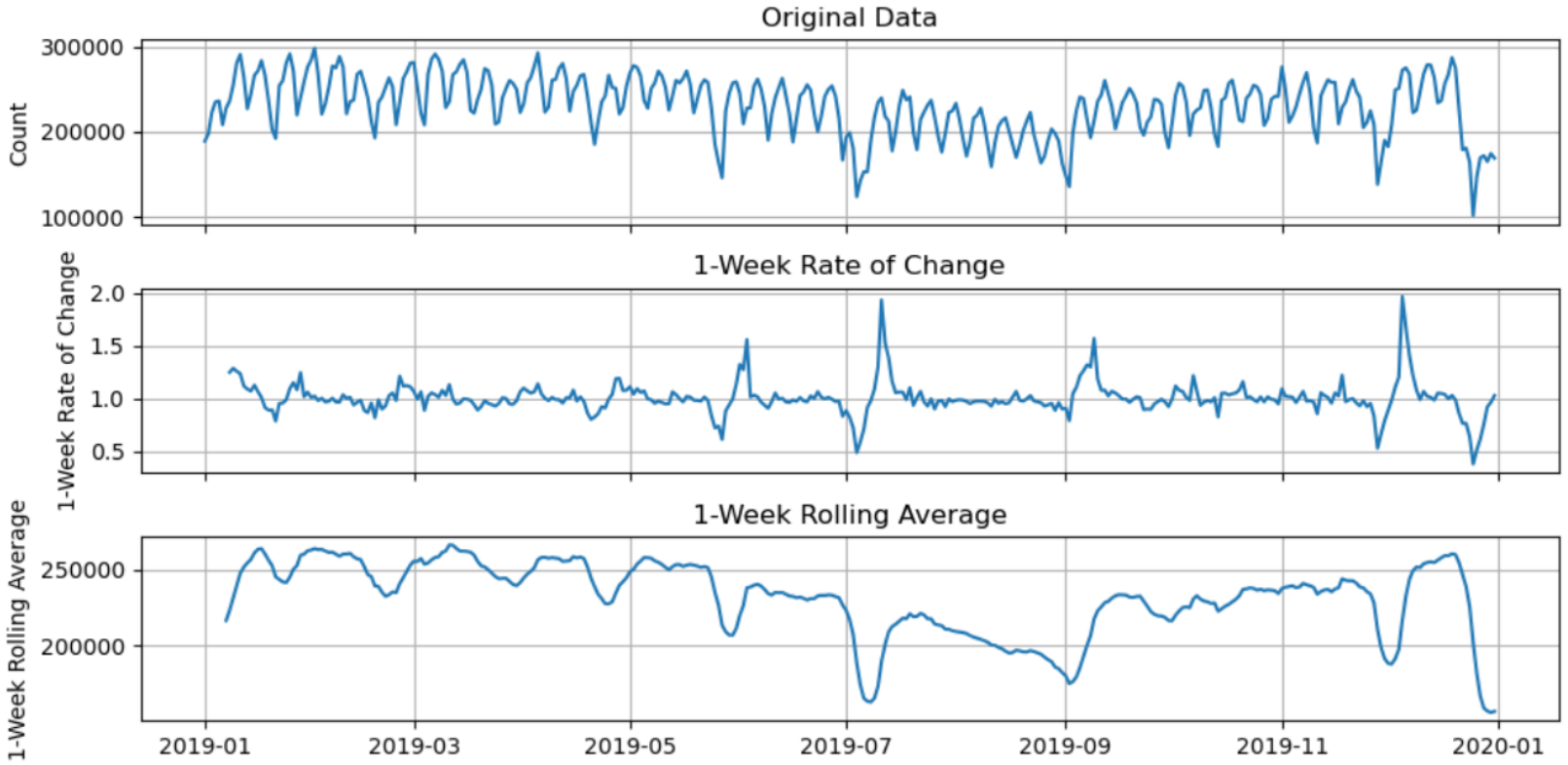




- Adding the convolution layer drastically improved training time!
- …which results in overfitting. 😔

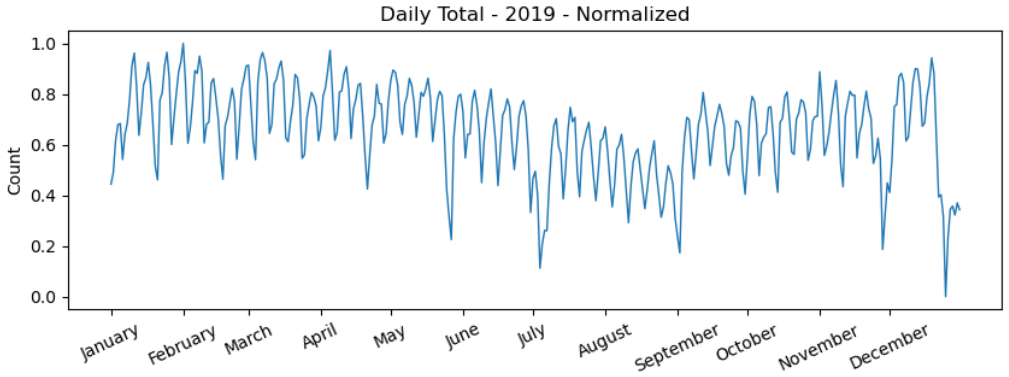
Modeling – Feature Engineering

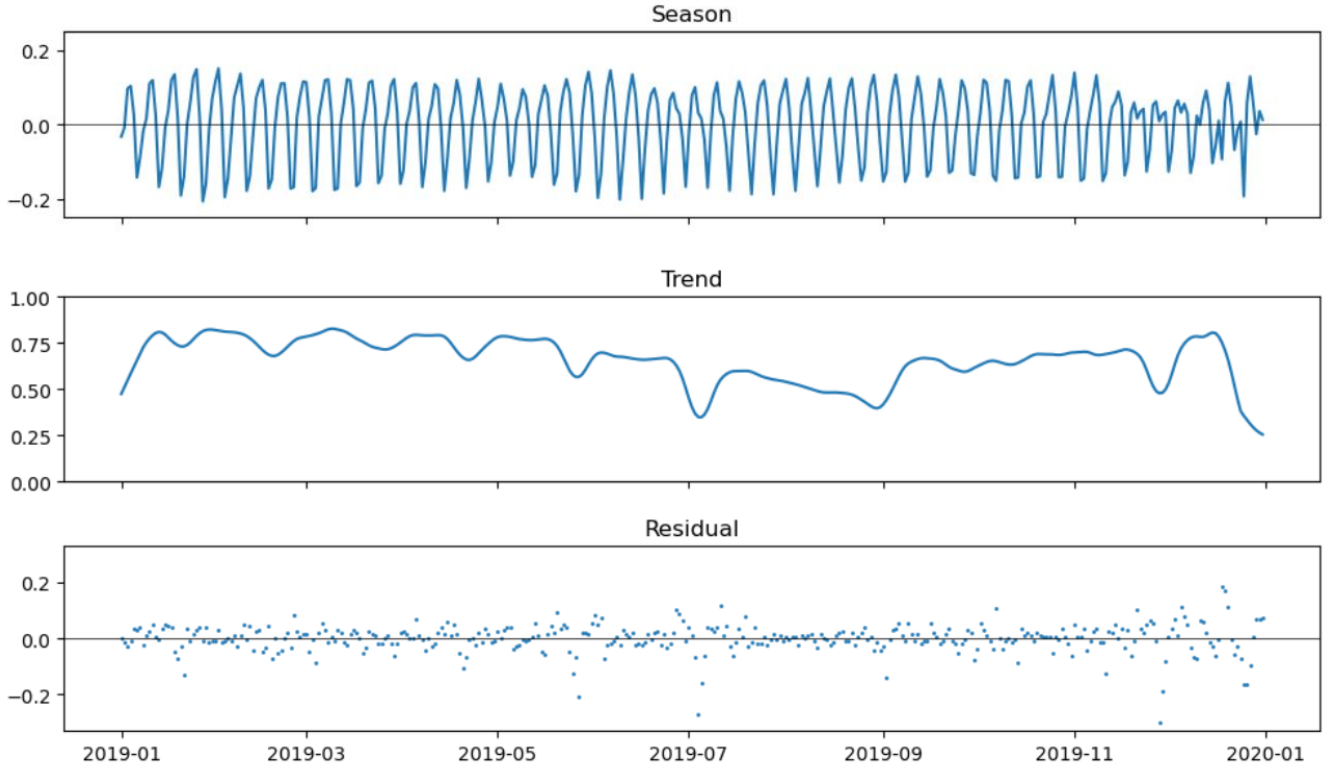
- Additional features improved modeling (more information)



Modeling – STL Decomposition

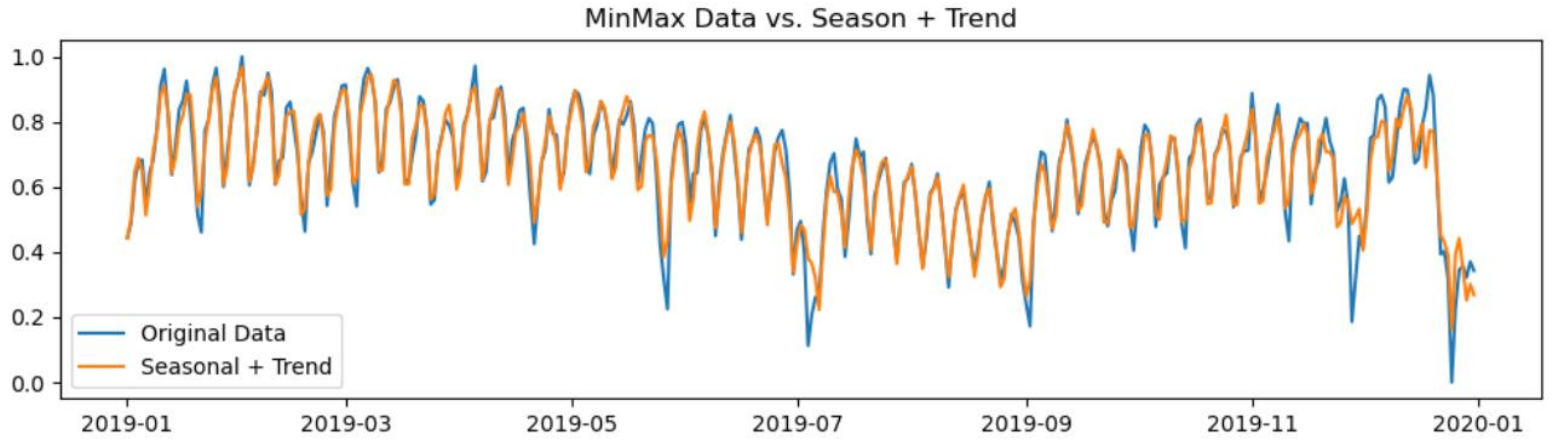
- Season-Trend Decomposition using LOESS (STL) is a way to process data into three component parts.





Modeling – STL Decomposition (2)

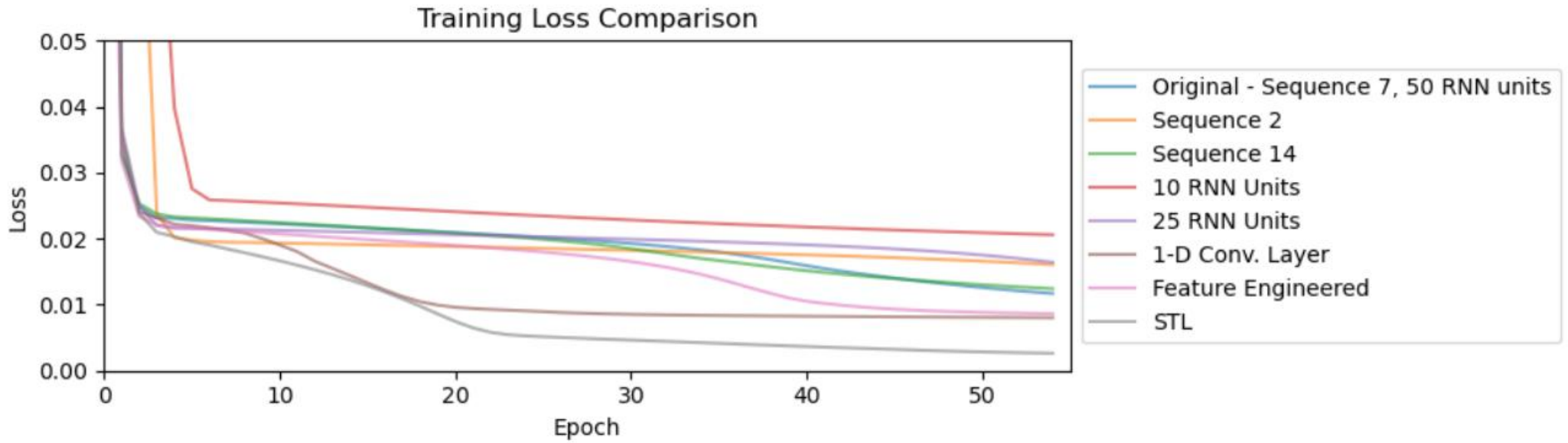
- The residual component is an alternative approach to anomaly detection, independent from an RNN implementation.



- STL improved training speed most of the variations tested.

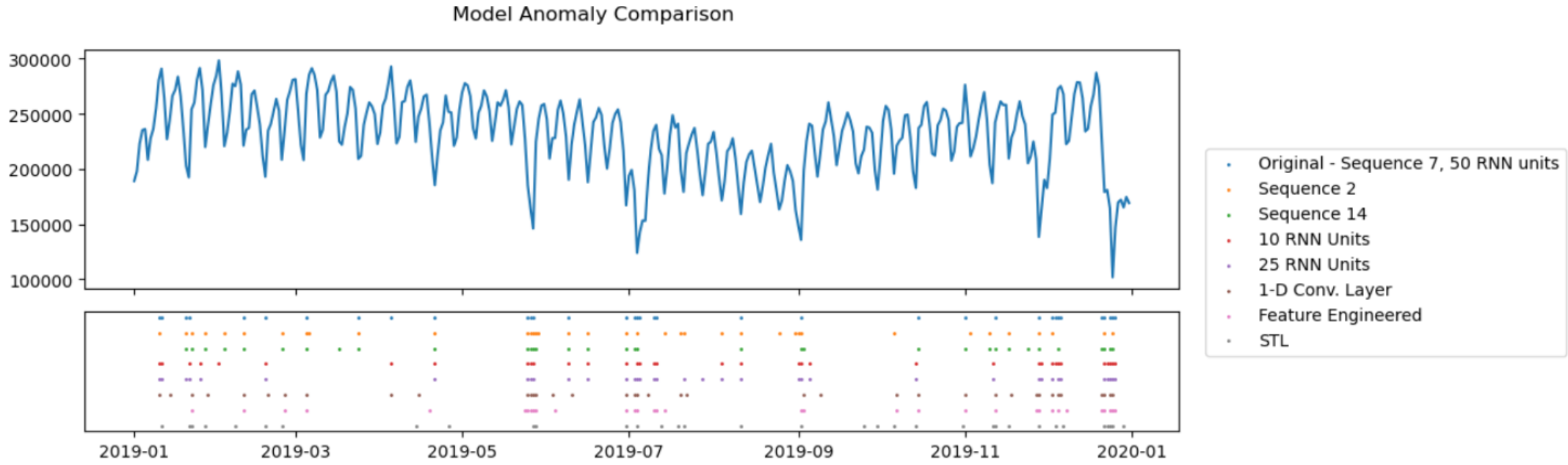
Results & Analysis

- Rate of Learning Comparison
  - More units and longer sequences learn faster (to an extent)
  - 1-D Convolutional layer 2nd best improvement (also adds parameters)
  - STL performed best, but more features better



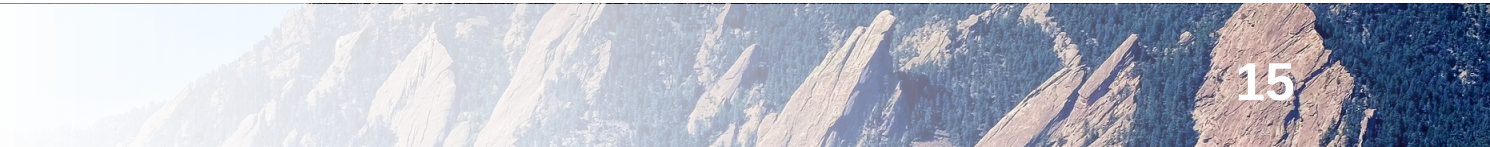
Results & **Analysis** (2)

- Model Anomaly Comparison
  - Overfitting adds harder to interpret anomalies, while underfitting misses some critical anomalies
  - Note variation in each, in part due to randomness of model training



Conclusion & Future Work

- RNN is suitable for anomaly detection via prediction.
- Model training may be improved, in order, by:
  - Pre-processing with STL
  - Implementing a 1-D Convolutional Layer
  - Engineering Features
  - increasing epochs, RNN units, sequence length… but only marginally
- Concern with anomaly detection is overfitting, which then labels non-anomalous points as anomalies.
- Future work can consider other regularization techniques, various architectures on complex data, and non-neural network approaches such as one class classification.



Thank you for listening!

Questions, clarifications, concerns

Nick Vastine

DTSA5511 – Intro to Deep Learning
Final Project
April 29, 2025

Data Science
UNIVERSITY OF COLORADO **BOULDER**