BTP PRESENTATION
Semester 7 Summer'21

Time Series Analysis

Under the guidance of Dr. Gaurav Harit

ADITYA KUMAR
[B18CSE002]

KARTIK VYAS
[B18CSE020]

Overview

- Time : a factor that governs everything, perhaps the most important aspect of our life

- Time series forecasting refers to making scientific predictions based on historical time stamped data

- Anomaly Detection : “An observation which deviates so much from other observations as to arouse suspicions that it was generated by a different mechanism”

Motivation

- Time Series analysis has the most wide range of applications in several research fields and industries.
- Healthcare, Weather, Climate, Economic, Business, Environmental Studies, Finance, Engineering : Some domains
- Make solutions for the real world problems
- Focused mainly on healthcare. Got motivated due to the recent pandemic situation
- Easily be extended to any other problem and dataset.

Goals and Problem Statement

Perform time-series analysis on different datasets using different deep learning architectures

Implement single-step time series forecasting, multistep time series forecasting, and anomaly detection for univariate and multivariate datasets

- Anomaly Detection
  - LSTM Autoencoders
- Single-Step Time Series Prediction
  - LSTM

Goals and Problem Statement

- ❖ Multi-Step Time Series Prediction
  - ➢ LSTM
  - ➢ CNN LSTM Encoder - Decoder
  - ➢ LSTM Encoder - Decoder
- ★ Prediction of number of patients incoming in a hospital
- ★ Prediction of sales of medicines
- ★ Anomaly of patients incoming in hospital

Hint at an outbreak of a disease and would also aid healthcare centers to maintain medicinal inventory and slot management of healthcare officials.

- ★ Also worked on a predictive model for a service outlet

Datasets

We have used three different datasets for time series analysis

- London Bike Sharing Dataset
- Patients Dataset [Generated]
- Pharma Sales Dataset

London Bike Sharing Dataset

Contains data for the number of bike shares on an hourly basis. The dataset spans for a time period of 2 years.

This multivariate dataset has several features in the dataset, some namely:

- Timestamp
- cnt - the count of a new bike shares
- t1 - real temperature in C
- is_holiday - boolean field - 1 holiday / 0 non holiday
- is_weekend - boolean field - 1 if the day is weekend
- season - category field meteorological seasons

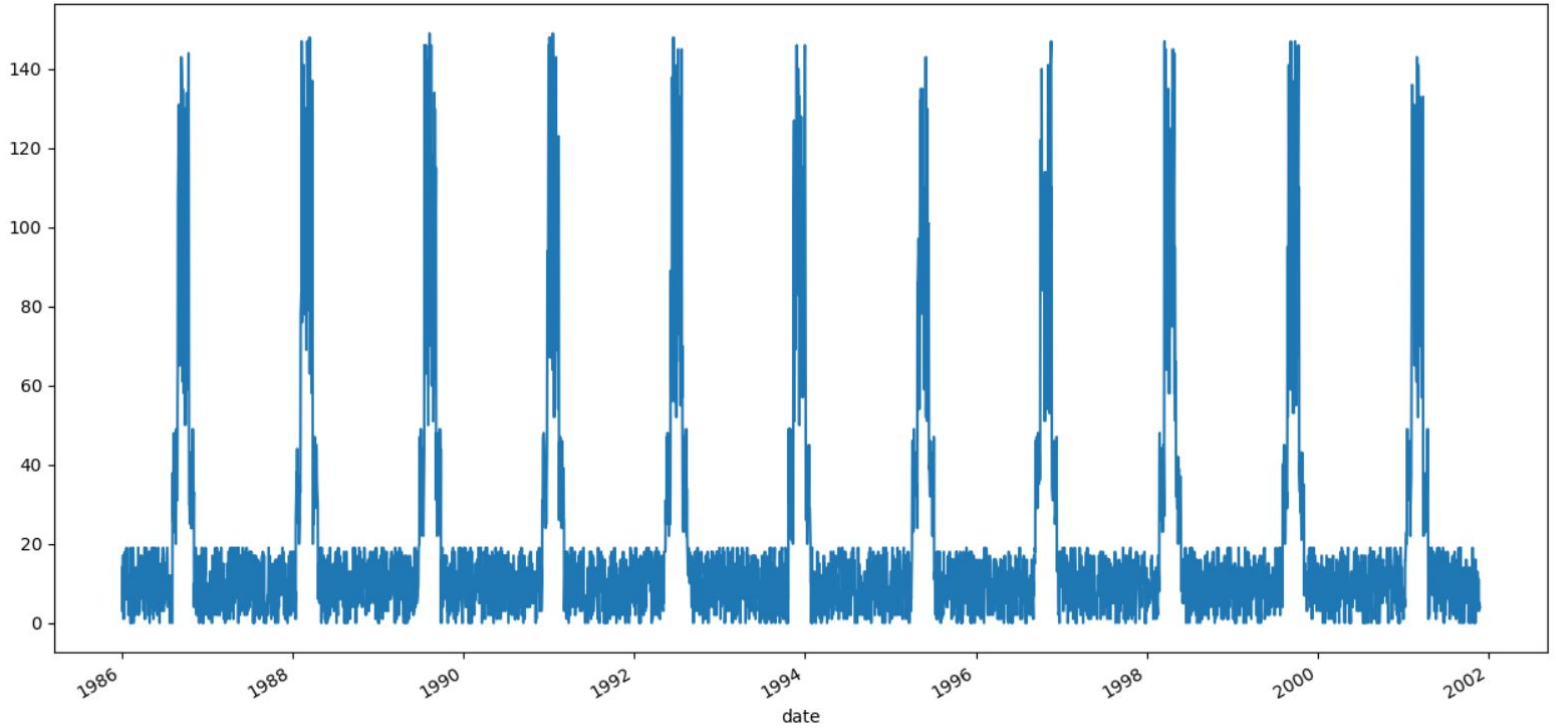
London Bike Sharing Dataset

A	B	C	D	E	F	G	H	I	J
timestamp	cnt	t1	t2	hum	wind_spee	weather_c	is_holiday	is_weeken	season
04-01-2015 00:00	182	3	2	93	6	3	0	1	3
04-01-2015 01:00	138	3	2.5	93	5	1	0	1	3
04-01-2015 02:00	134	2.5	2.5	96.5	0	1	0	1	3
04-01-2015 03:00	72	2	2	100	0	1	0	1	3
04-01-2015 04:00	47	2	0	93	6.5	1	0	1	3
04-01-2015 05:00	46	2	2	93	4	1	0	1	3
04-01-2015 06:00	51	1	-1	100	7	4	0	1	3
04-01-2015 07:00	75	1	-1	100	7	4	0	1	3
04-01-2015 08:00	131	1.5	-1	96.5	8	4	0	1	3
04-01-2015 09:00	301	2	-0.5	100	9	3	0	1	3
04-01-2015 10:00	528	3	-0.5	93	12	3	0	1	3
04-01-2015 11:00	727	2	-1.5	100	12	3	0	1	3

Patients Dataset [Generated]

- Dataset of much relevance could not be found that revolves around number of patients visiting a healthcare center
- Created a dataset using C++ that spans for a period of eleven years
- Made for the purpose of anomaly detection
- In each year, there is a time period of a month or so, where the value is really high, indicating a possible outbreak of a disease
- Dataset resonates with real world datasets

Patients Dataset [Generated]

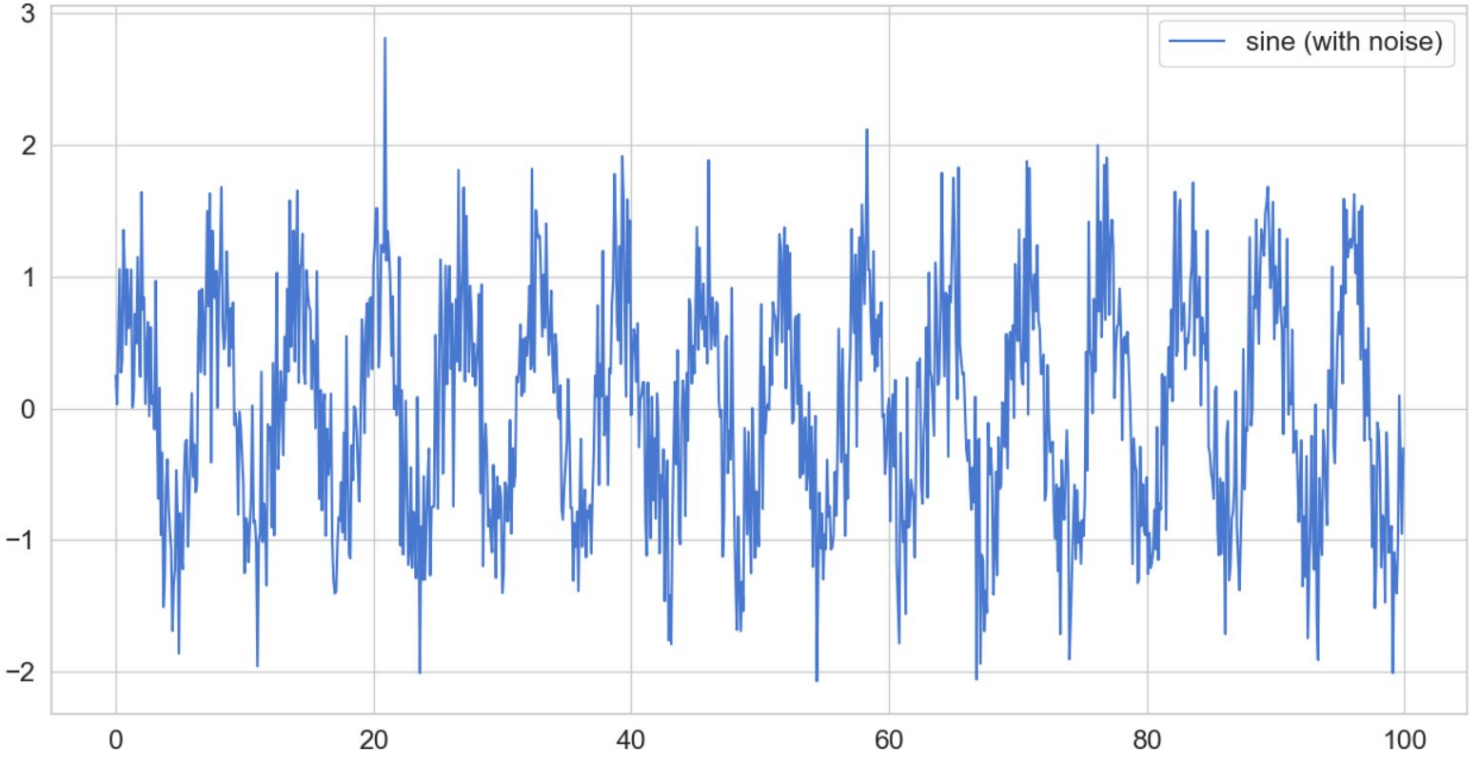


Pharma Sales Dataset

- Built from an initial dataset which included 600000 transactions collected over a period of six years
- Includes the date and time of sale, pharmaceutical drug brand name, quantity sold
- Sales data are resampled to hourly, daily, weekly, and monthly basis
- The selected group of drugs from the dataset (57 drugs) is classified into 8 broad categories : M01AB, M01AE, N02BA, N02BE/B, N05B, N05C, R03, R06

Single Step Univariate Time Series Prediction

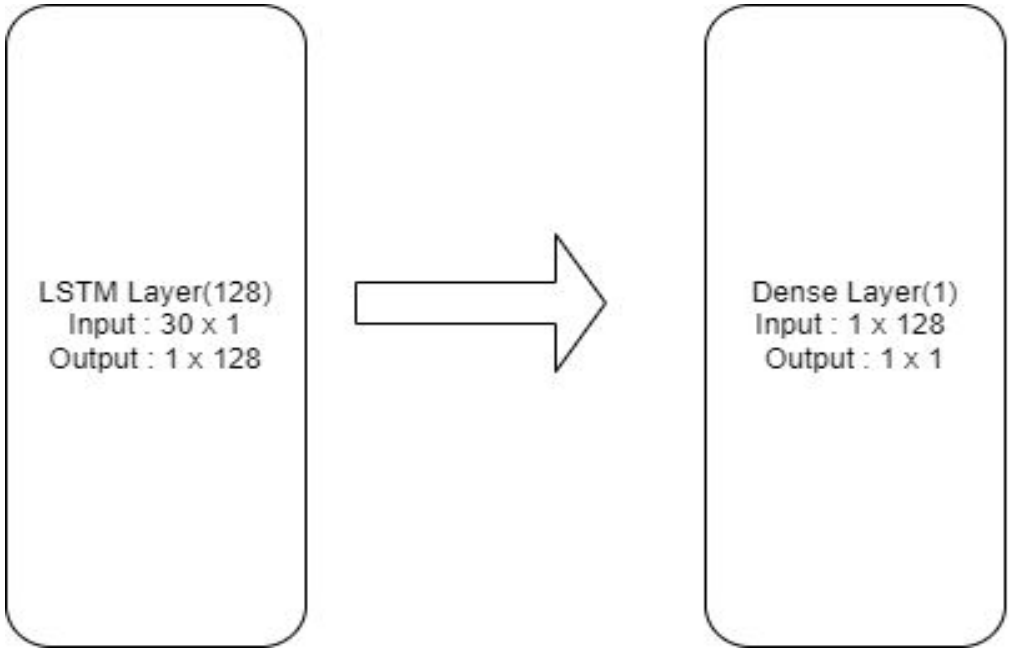
- The dataset has been randomly generated.
- Normal Distribution (mean at 0.5) has been used to generate random values that are added to the sine curve.
- Thousand values are generated
- Data divided in an 80:20 ratio
- The time steps for sequence set at 30


sine (with noise)
3
2
1
0
−1
−2
0
20
40
60
80
100

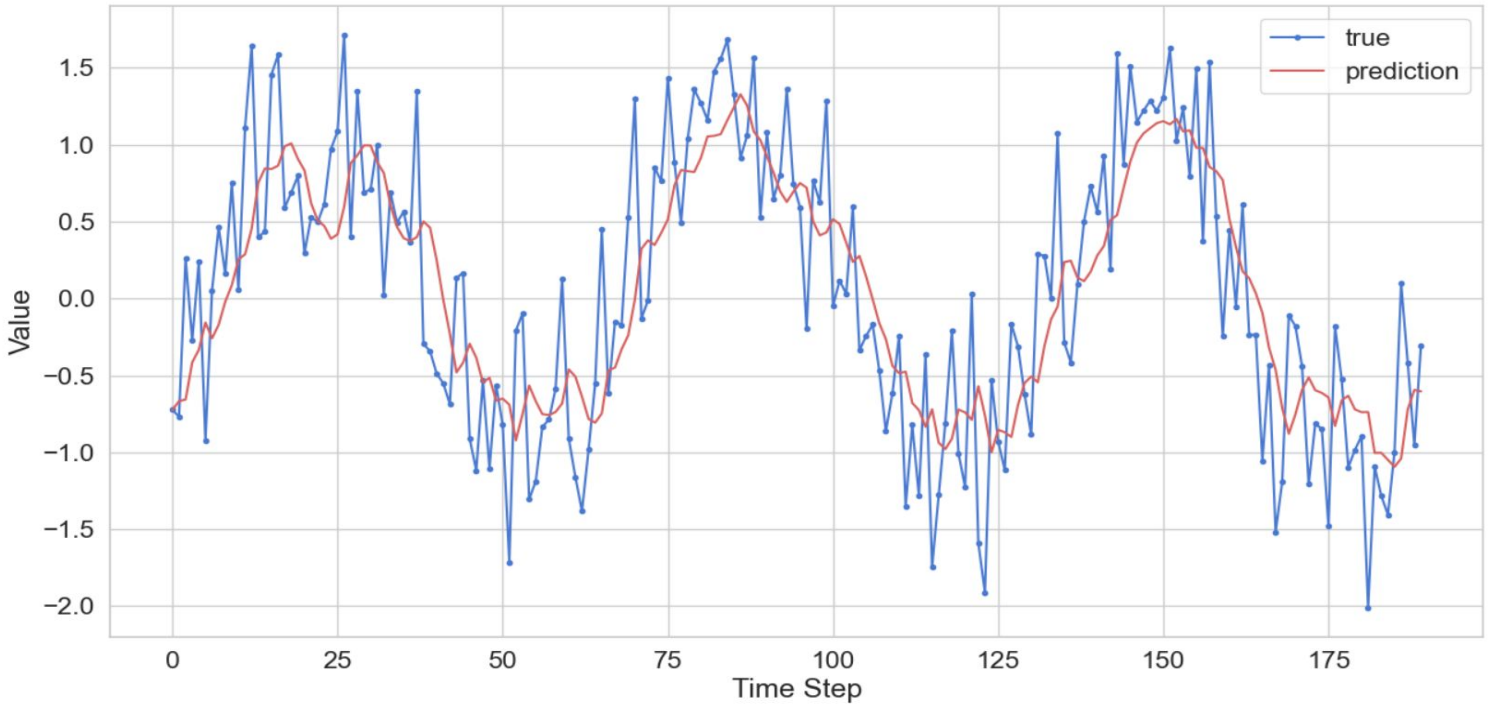
LSTM Model

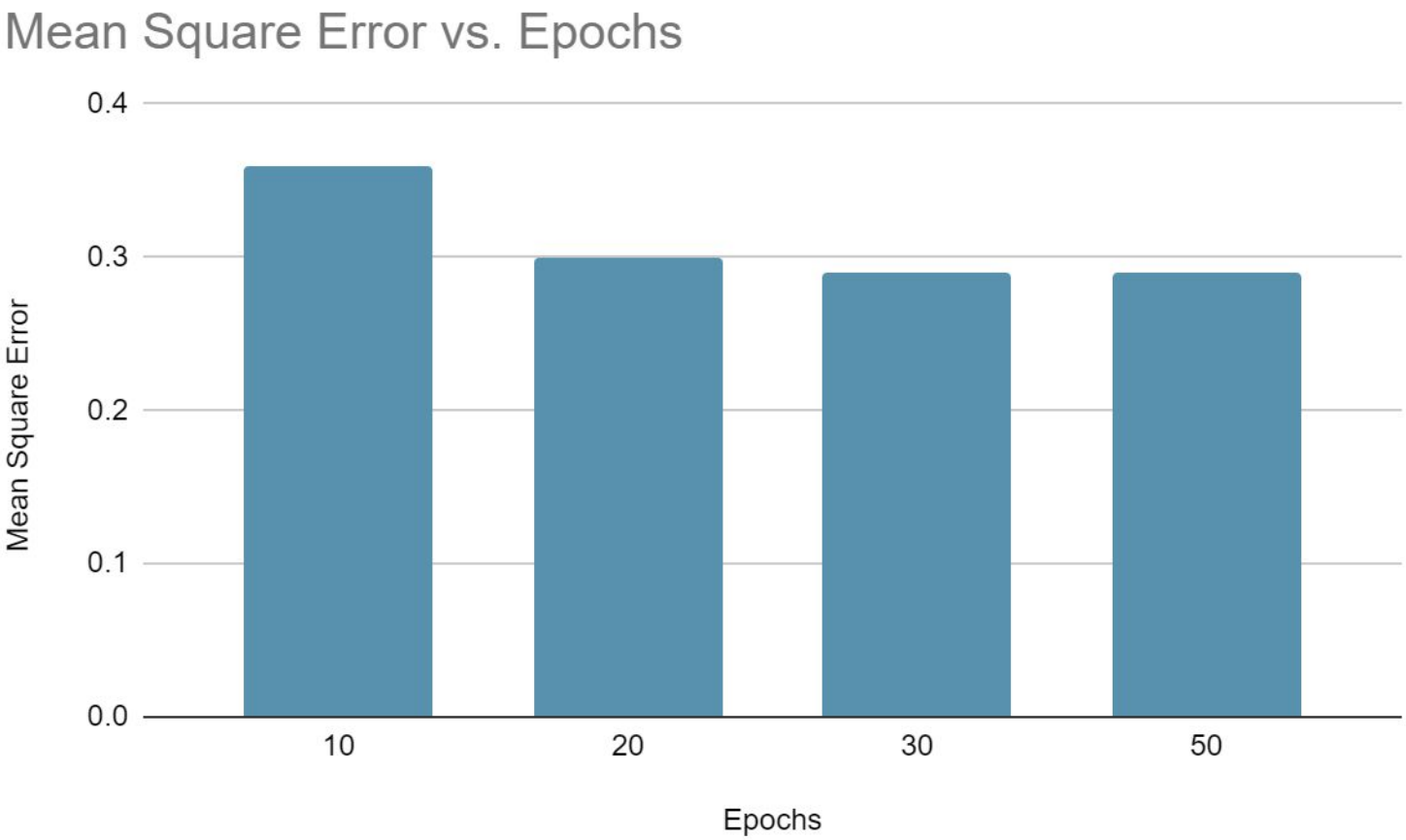
- Need something more robust than the classical RNN models : errors arising due to :
  - Gradient calculations : Vanishing Gradient and Exploding Gradient
  - Memory, the older data points are forgotten
- Used Keras for the ease in implementing the LSTM model
- The intuition behind LSTM develops due to the gates that it has.
- Adam was used as the optimizer.
- Mean Square Error has been used to calculate the error

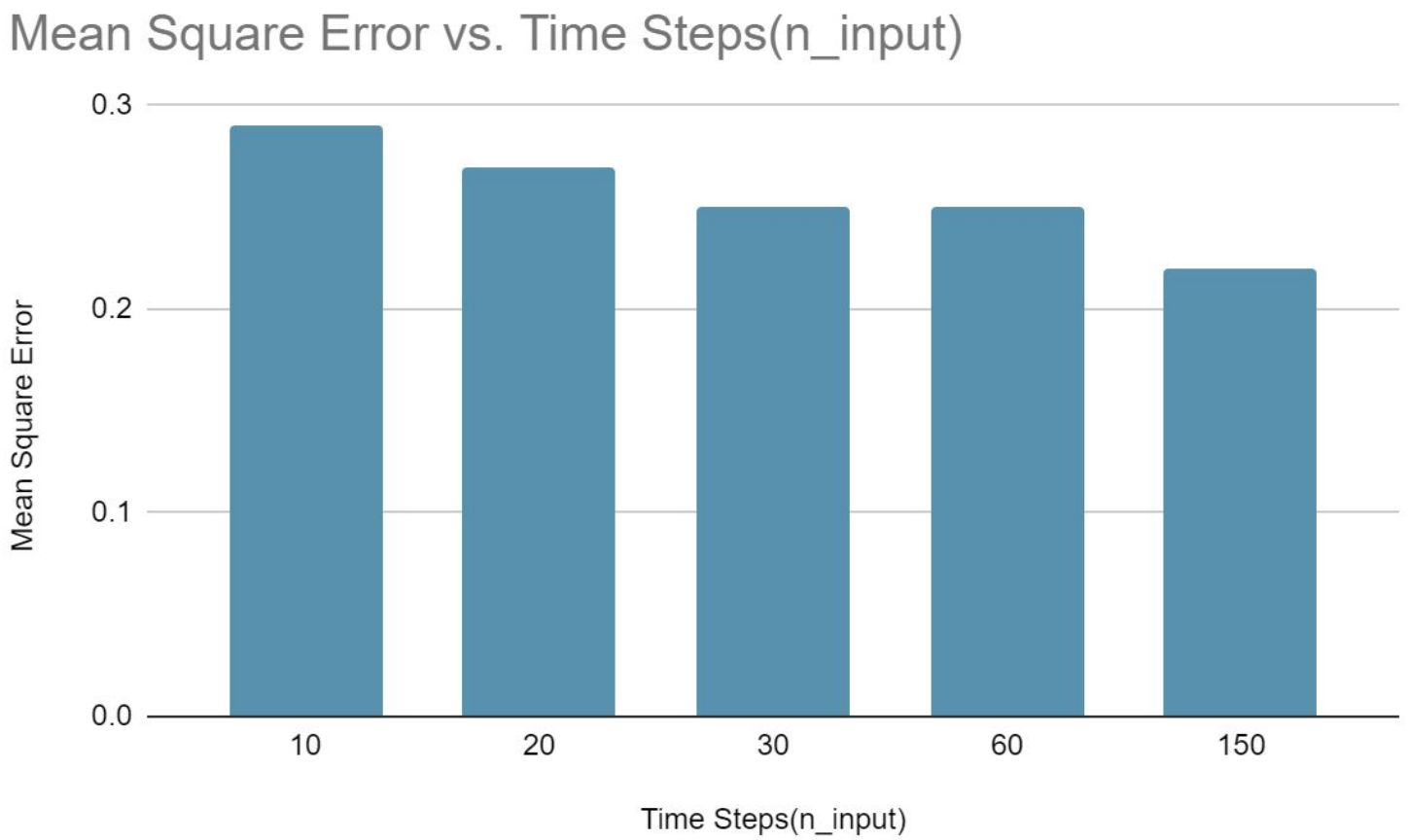
LSTM Model


LSTM Layer(128)
Input : 30 x 1
Output : 1 x 128
Dense Layer(1)
Input : 1 x 128
Output : 1 x 1

Results


true
prediction
Value
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
0
25
50
75
100
125
150
175
Time Step

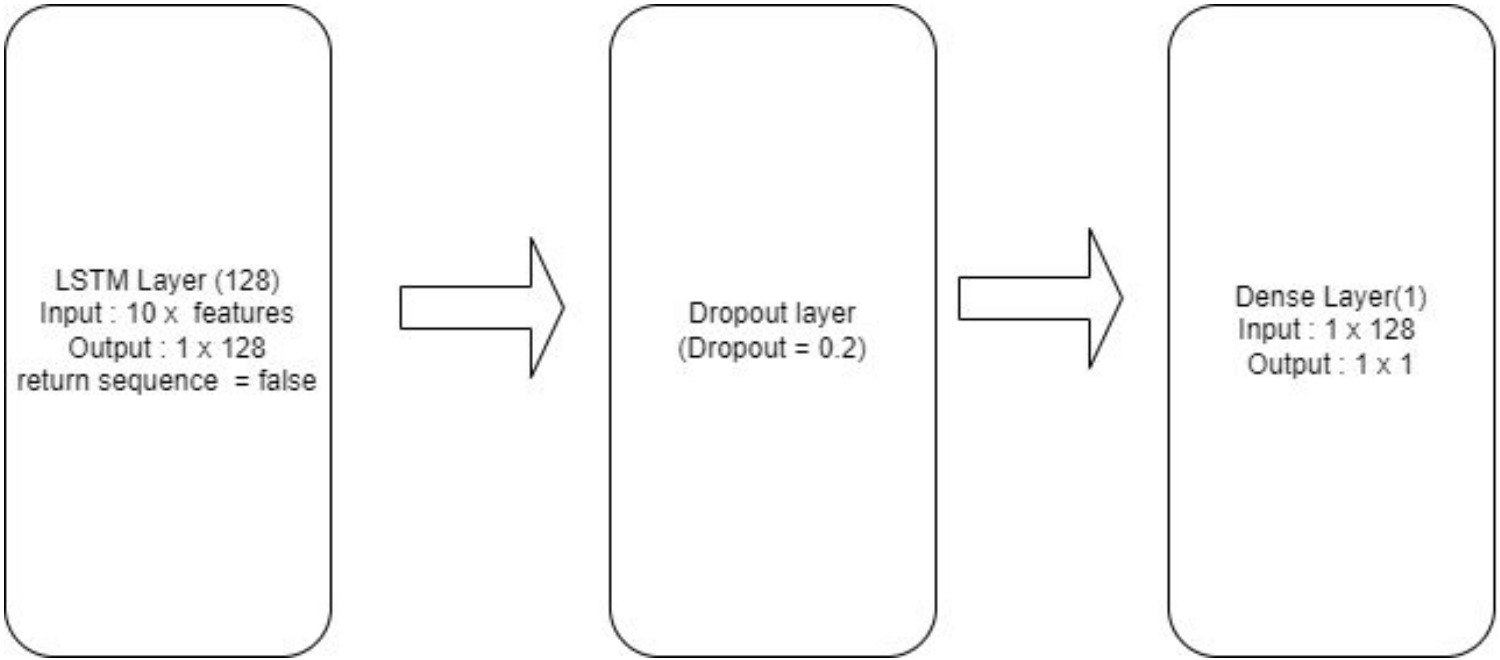

Mean Square Error vs. Epochs
0.4
0.3
0.2
0.1
0.0
Mean Square Error
10
20
30
50
Epochs


Mean Square Error vs. Time Steps(n_input)
0.3
0.2
0.1
0.0
Mean Square Error
10
20
30
60
150
Time Steps(n_input)

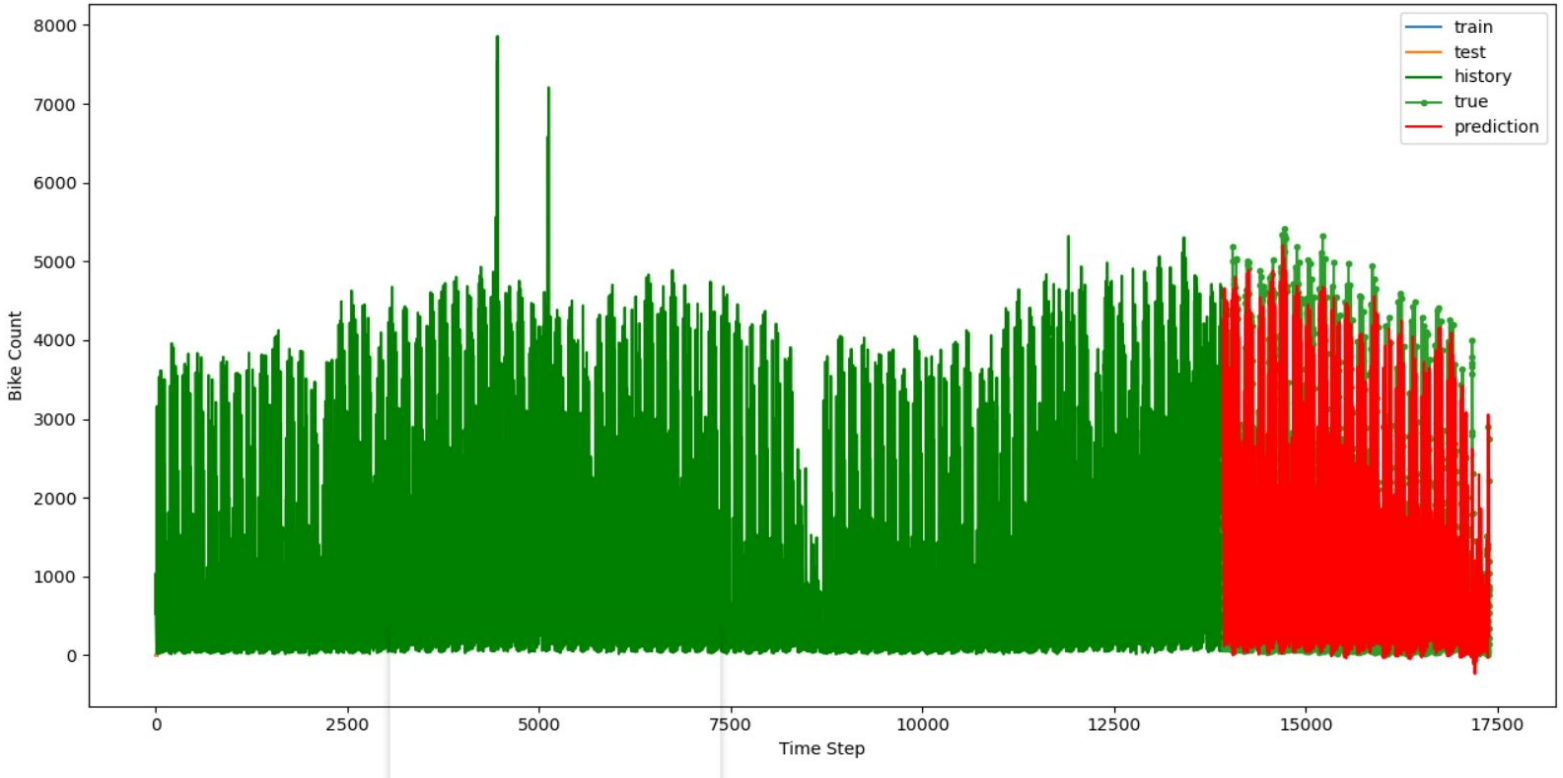
Single Step Multivariate Time Series Prediction

- London Bike Sharing Dataset
- Multivariate Data
- Data divided in 80:20 ratio
- Robust Scaler is being used for scaling : outliers removed. Removes median and scales data in between 1st quartile and 3rd quartile.
- The time step for input sequence set at 10
- Adam was used as the optimizer and MSE as the loss function.
- After fine tuning the model, the least MSE value was 0.02

LSTM Model


LSTM Layer (128)
Input : 10 x features
Output : 1 x 128
return sequence = false
Dropout layer
(Dropout = 0.2)
Dense Layer(1)
Input : 1 x 128
Output : 1 x 1

Result


train
test
history
true
prediction
Bike Count
Time Step
0
1000
2000
3000
4000
5000
6000
7000
8000
0
2500
5000
7500
10000
12500
15000
17500

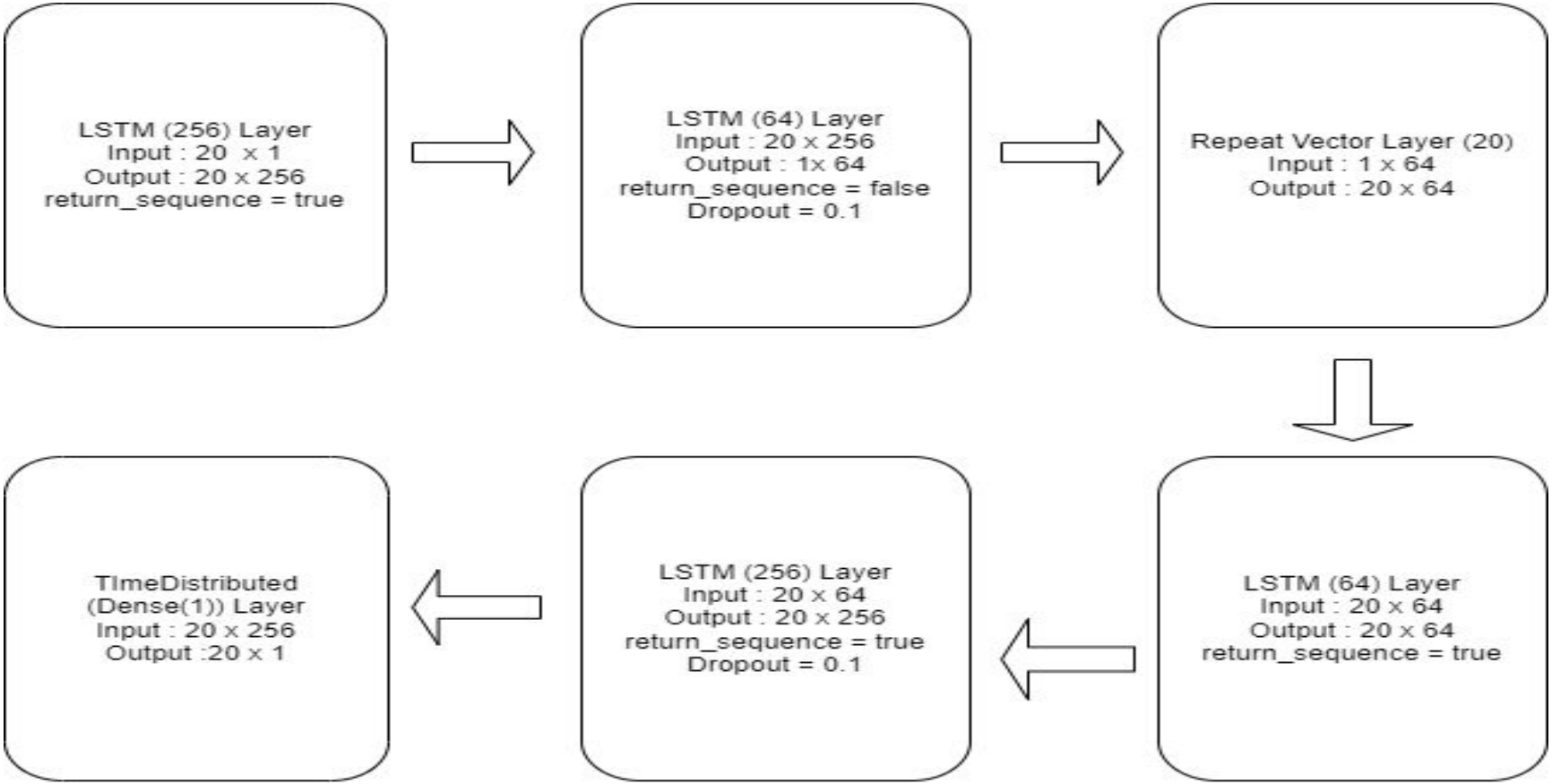
Anomaly Detection

- The task of identifying rare events in data
- Bank fraud detection, tumor detection in medical imaging
- Detecting abnormal events in the hospital patients dataset
- Highlight the cases of anomalies for any suspicious events
- Data divided in 90:10 ratio
- Standard scaler makes mean equal to zero and scales data to unit variance because it follows standard normal distribution
- Time step for the sequence set at 20

LSTM Autoencoder Model

- Autoencoder architecture is divided into two modules, encoder and decoder, both consists of LSTM layers
- Two Dimensional matrix is fed as input into the LSTM network
- Repeat vector layer acts as a bridge between the encoder and decoder part.
- The task of the decoder layer is to unfold the decoding thus the layers of decoder are stacked in the reverse order to that of encoder.
- Time distributed layer is added to get the final desired output

LSTM Autoencoder Model



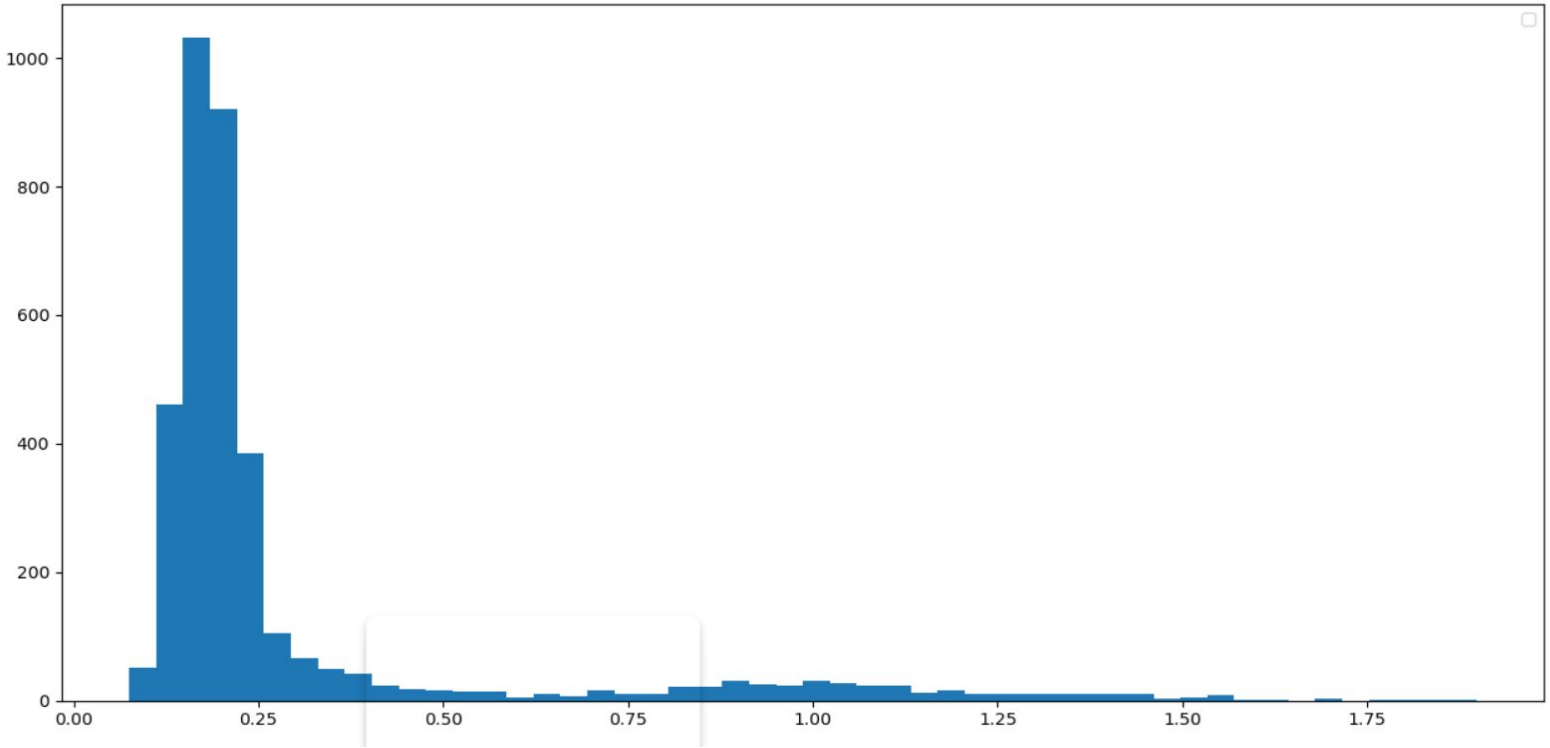
LSTM Autoencoder Model

- Adam was used as the optimizer while training the model.
- Mean absolute error was used as the loss function while training the model.
- Rectified Linear Unit(ReLU) is being used as the activation function

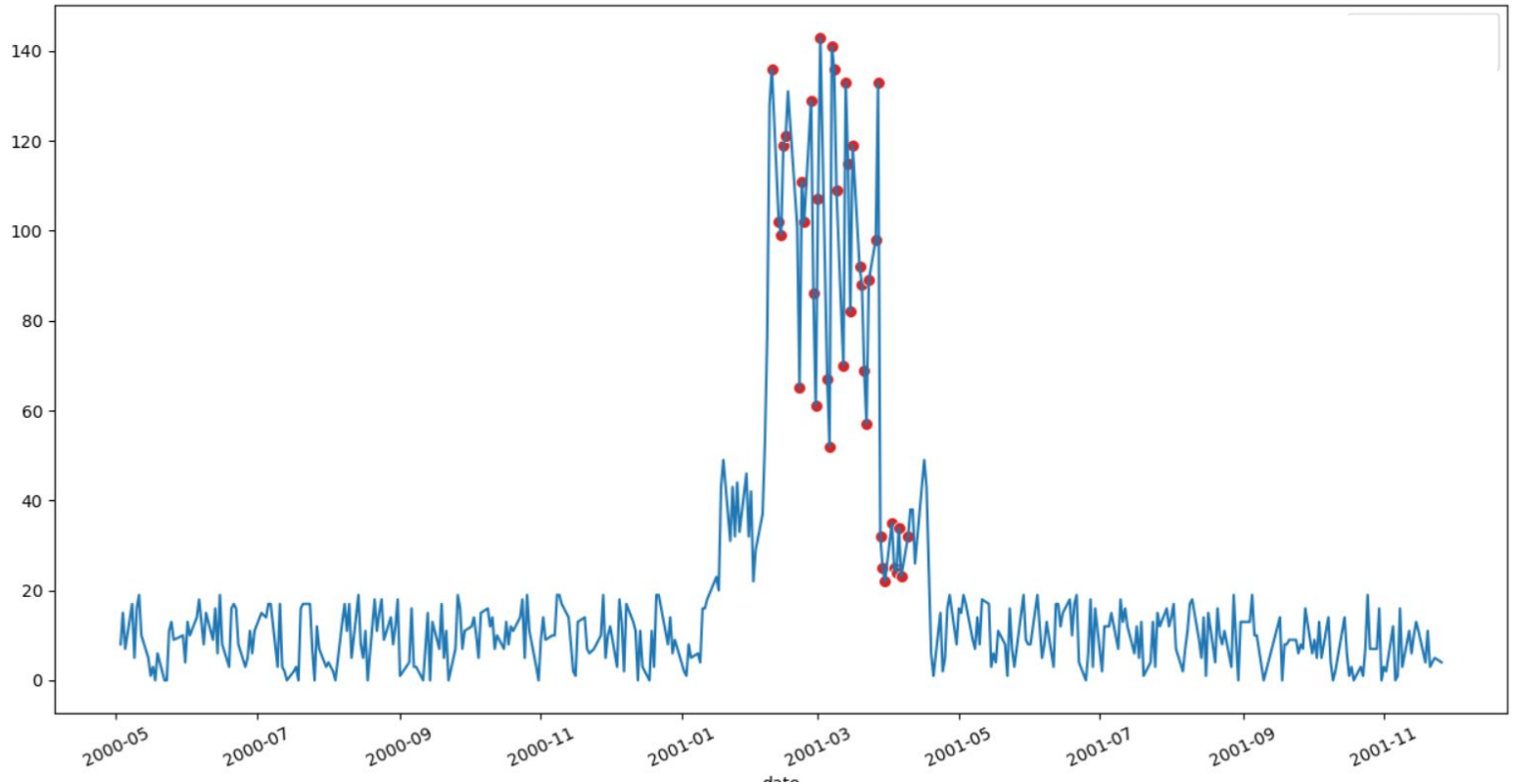
Evaluation :

- Calculate MAE on the training data and plotting the spread of error
- Making predictions on test data and calculate MAE.
- If MAE is more than threshold value, then data point is declared as anomaly.

Evaluation



Result


140
120
100
80
60
40
20
0
2000-05
2000-07
2000-09
2000-11
2001-01
2001-03
2001-05
2001-07
2001-09
2001-11
date

Multi Step Time Series Prediction

- Predict the sales of drugs at a healthcare center
- Predict the values of how many medicines would be sold in the duration of the next seven days
- Pharma sales dataset
- Data divided in 85:15 ratio
- Time step for input sequence set at fourteen
- Multiple ways for multi-step time series prediction

Models

- Direct Multi-step Forecast Strategy: Developing a separate model for each forecast time step
- Recursive Multi-step Forecast: One-step model multiple times where the prediction for the prior time step is used as an input for the future predictions
- Direct-Recursive Hybrid Strategies: This type of model can be called a hybrid model of the above two. A different model is constructed for each time step to be predicted, and each model uses the predictions made by models for earlier values

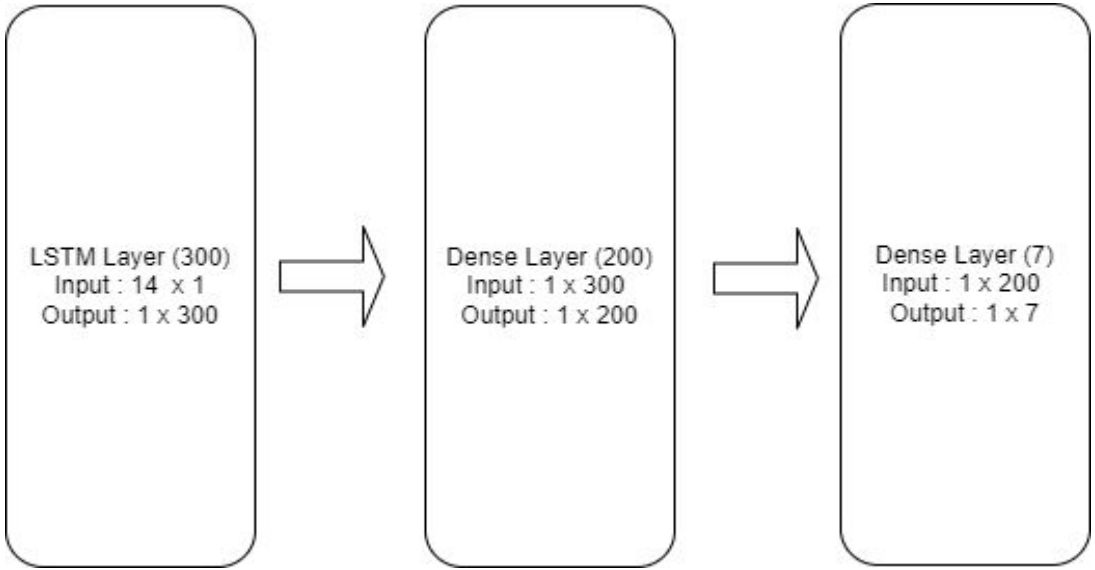
Models

- ❖ Multiple Output Strategy: Involves developing one model that is capable of predicting the entire forecast sequence
- ★ Chosen the last approach, that is, Multiple Output Strategy
- ★ A single model saving upon the computation and predicted values not used for further prediction, thus there are no aggregating errors.

We have worked on three models : LSTM, CNN LSTM Encoder Decoder, LSTM Encoder Decoder

LSTM Model

- ReLU, Mean Square Error and Adam optimizer used

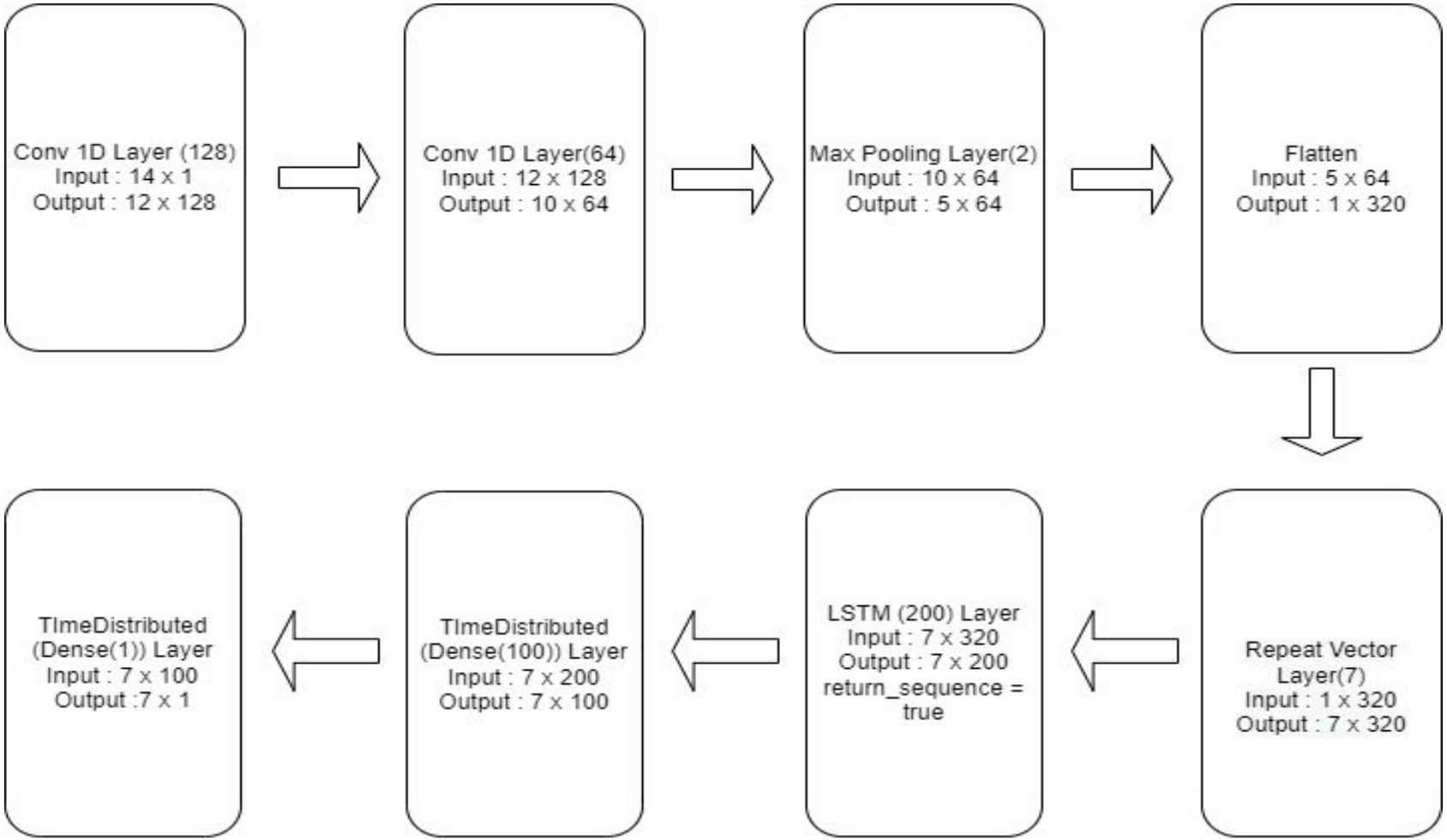


CNN LSTM Encoder Decoder Model

- Model consists of two modules, encoder and decoder
- Encoder consists of two convolution layers and a max pooling layer
- Convolution layer reads the input and projects it into feature maps
- The second convolution layer further tries to amplify any salient feature
- Input sequences in the convolutional layer are read with a kernel size of three time steps.

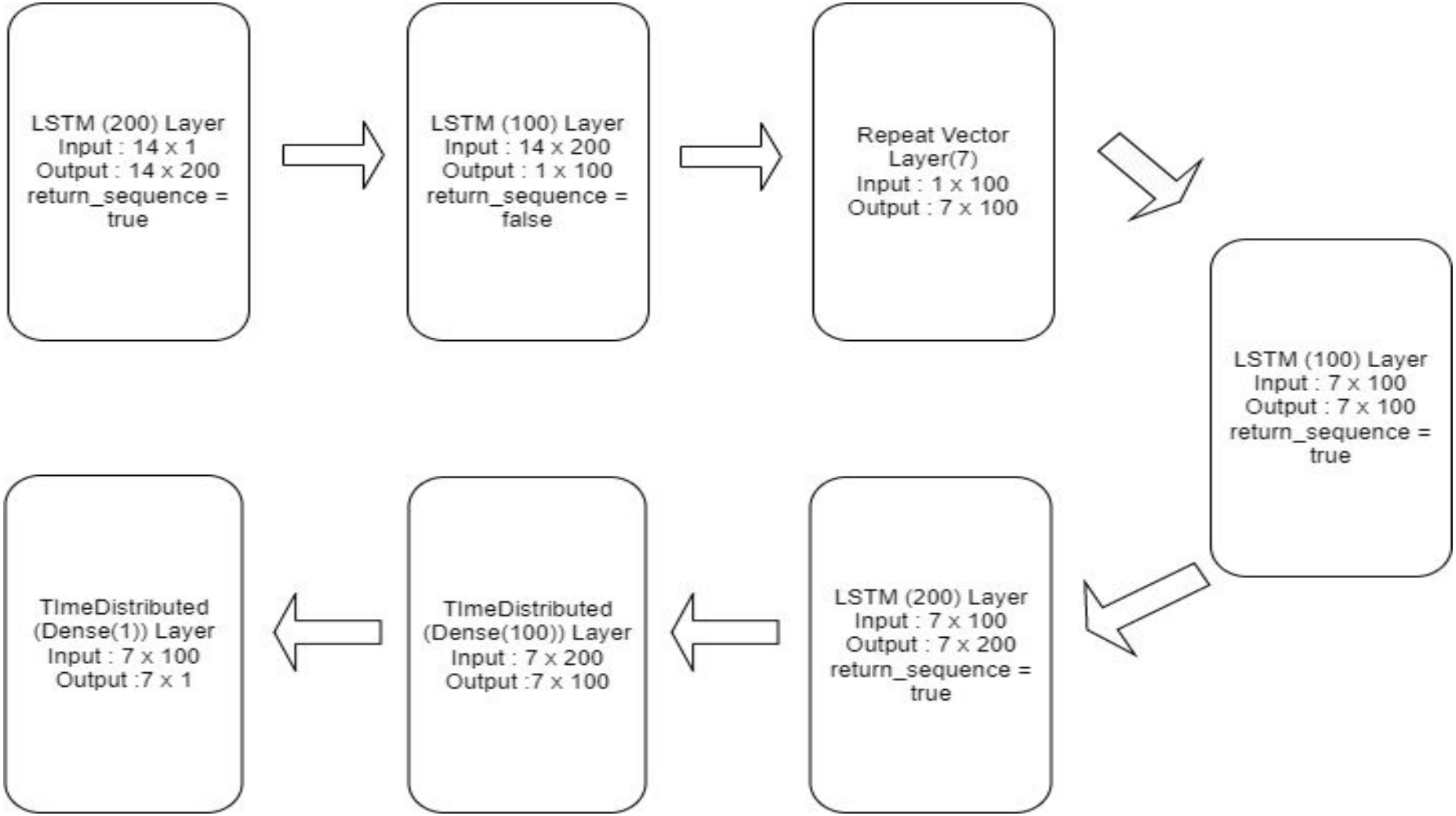
CNN LSTM Encoder Decoder Model

- Feature maps simplified by pooling layer by keeping half of the values from max signals
- Distilled feature map is flattened into vectors and fed into a repeat vector layer
- Vector is fed into the LSTM layer of the decoder
- Dense layer is being used to interpret each time step
- Output layer wrapped in Time Distributed Wrapper
- ReLU, Mean Square Error and Adam optimizer used


Conv 1D Layer (128)
Input : 14 x 1
Output : 12 x 128
Conv 1D Layer(64)
Input : 12 x 128
Output : 10 x 64
Max Pooling Layer(2)
Input : 10 x 64
Output : 5 x 64
Flatten
Input : 5 x 64
Output : 1 x 320
Repeat Vector Layer(7)
Input : 1 x 320
Output : 7 x 320
LSTM (200) Layer
Input : 7 x 320
Output : 7 x 200
return_sequence = true
TImeDistributed (Dense(100)) Layer
Input : 7 x 200
Output : 7 x 100
TImeDistributed (Dense(1)) Layer
Input : 7 x 100
Output :7 x 1

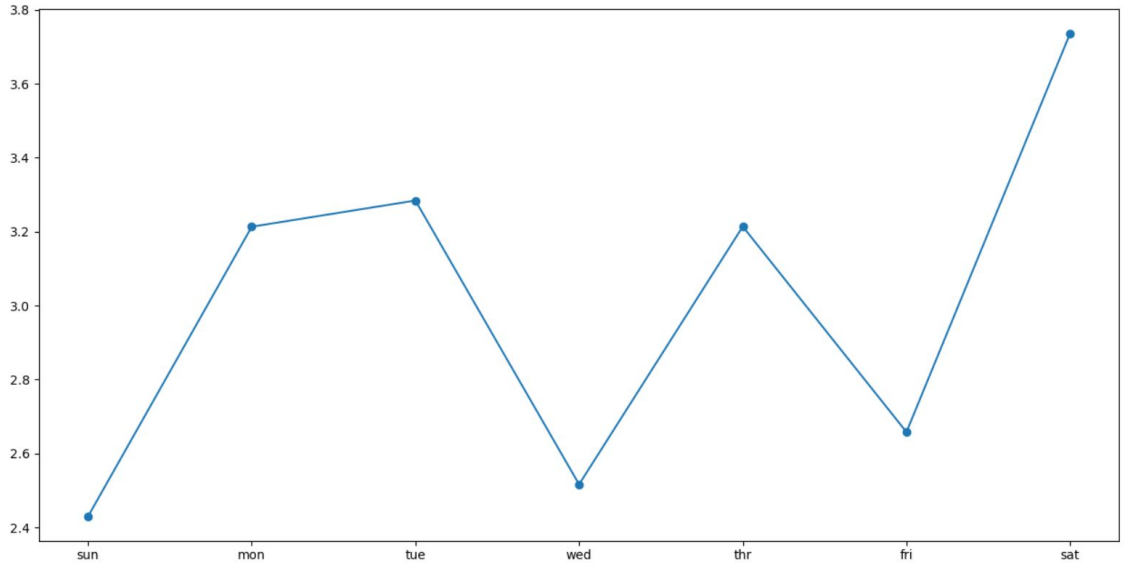
LSTM Encoder Decoder Model

- Both encoder and decoder modules are composed of LSTM layers
- Repeat vector layer acts as a bridge
- LSTM in the decoder beneficial because it knows what was predicted for a prior day in the sequence and accumulates internal states while outputting the sequence
- Mean Squared Error used as loss function
- Adam used as the optimizer
- ReLU used as Activation Function


LSTM (200) Layer
Input : 14 x 1
Output : 14 x 200
return_sequence = true
LSTM (100) Layer
Input : 14 x 200
Output : 1 x 100
return_sequence = false
Repeat Vector Layer(7)
Input : 1 x 100
Output : 7 x 100
LSTM (100) Layer
Input : 7 x 100
Output : 7 x 100
return_sequence = true
LSTM (200) Layer
Input : 7 x 100
Output : 7 x 200
return_sequence = true
TImeDistributed (Dense(100)) Layer
Input : 7 x 200
Output :7 x 100
TImeDistributed (Dense(1)) Layer
Input : 7 x 100
Output :7 x 1

Evaluation

- Evaluated on the basis of MSE calculated for each forecast day as well as overall MSE calculated
- MSE used because it is more punishing for larger mistakes



Evaluation

- Duly experimented with the models on the basis of fine tuning several parameters
- Chosen epochs as 20 and number of timesteps as 14

Epochs	n_input	LSTM	CNN LSTM Encoder-Decoder	LSTM Encoder-Decoder
20	7	2.880	2.897	2.864
20	14	2.907	2.877	2.863

Future Prospects

- The foundation has already been laid for multivariate multi step time series prediction, they can be explored in a much better depth
- GNNs can be experimented upon after due research on the topic

Tech Stack

- Programming languages : Python, C++
- DL libraries : Tensorflow, Keras

Conclusion

- Different types of time series including univariate and multivariate time series have been analyzed using different techniques
- Anomaly detection performed using LSTM autoencoder
- Prediction of sales using multivariate London bike service data
- Multistep time series forecasting performed on the pharma sales dataset using three different models : LSTM, CNN-LSTM Encoder - Decoder and LSTM Encoder - Decoder
- All the models duly fine-tuned and made in such a way they can be easily extrapolated to new problems.

Acknowledgement

We are heartily thankful to our instructor, Dr. Gaurav Harit, for providing us the necessary guidance, the needed constant support, and helping us throughout the course of the project via continuous interaction and evaluation of the course at regular intervals.

Thank You