Mini Project

Time Series Anomaly Detection with LSTM Autoencoders

Submitted By
Amarjeet Kumar CSB15066
Ajay Kumar CSB15057

Abstract

The problem of anomaly detection on time series is to predict whether a newly observed time series normal, to a set of training time series. It is very useful in many monitoring applications such as video surveillance, signal recognition, stock price data etc. Anomaly detection in these applications aims to automatically detect unseen occurrence of events in real time based on historical training instances. Based on LSTM Autoencoders, we propose an instance-based anomaly detection algorithm for S & P 500 daily closing price data and because it is also very efficient for anomaly detection of time series LSTM Autoencoder.

Introduction

Anomalies are patterns in data that do not conform to a well defined notion of normal behavior. Anomaly detection problem for time series is the process of identifying outlier data points relative to some standard or usual signal in data sets. Even though the terminology behind anomaly detection uses the probability theory and some statistics. The purpose of this project is to detect anomalies in S&P 500 daily closing price data .

Need/Importance of Time Series Anomaly Detection

The importance of anomaly detection is due to the fact that anomalies in data translate to significant and actionable information in a wide variety of application domains and that is why Anomaly detection is deployed to check if any data falls outside the control limits and to determine the root cause. For example a retailer must adopt anomaly detection as a way of life because it is effortless, easier, cheaper, better and faster. As Anomaly detection is about identifying outliers in a time series data, correlating it various influencing factors and delivering insights to help business decision makers there are few more example such as:

- Seasonality and trend estimation
- surveillance domain
- social networks

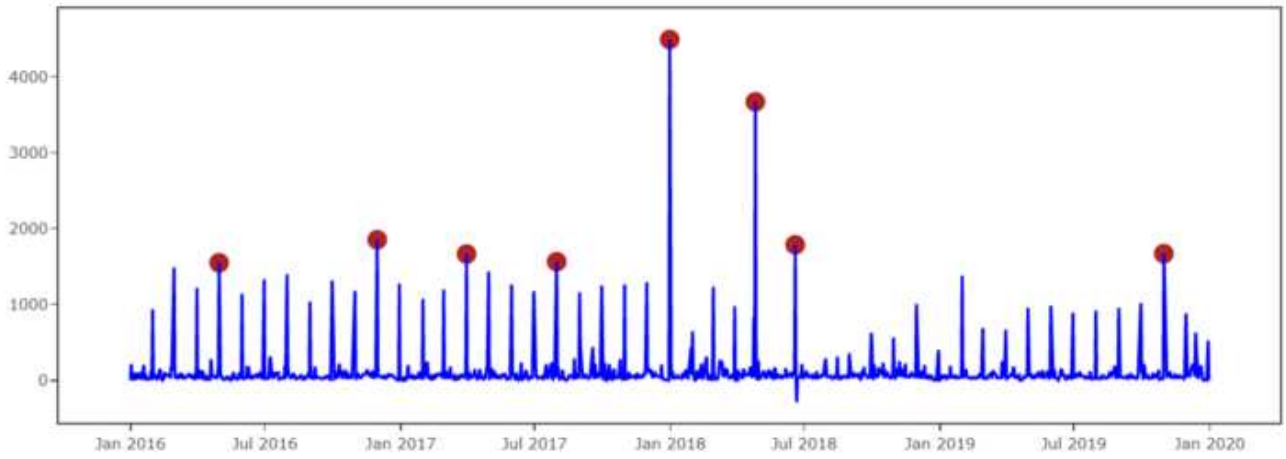
Goals/Objectives

Goal of this project is to detect and analyze anomalies in S&P 500 daily closing price data using LSTM Autoencoders to get that idea of anomaly detection.

Before knowing the solution process we should know some important terminology used to define the project so lets understand them.

Time Series Anomaly Detection:

In time series data, Anomaly detection refers to the task of finding/identifying rare events or data points. For example in a festive season the prize of Goods goes high because of customer demand so if we plot the daily revenue then we can see a hike in revenue at that time of periods as shown below in image-



LSTM:

LSTM is a type of Recurrent Neural Network (RNN) specifically used on sequential or time series data as it known for its ability to extract both long- and short- term effects of pasts event.

Autoencoder:

An autoencoder is a type of artificial neural network for which the input is the same as the output is used to learn efficient data codings in an unsupervised manner.

LSTM Autoencoders:

An LSTM Autoencoder is an implementation of an autoencoder for sequence data using an Encoder-Decoder LSTM architecture.

Anomaly Detection with Autoencoders:

Here are the basic steps to Anomaly Detection using an Autoencoder-

- Train an Autoencoder on normal data (no anomalies)
- Take a new data point and try to reconstruct it using the Autoencoder
- If the error (reconstruction error) for the new data point is above some threshold, we label the example as an anomaly

S&P 500 Index Data:

The data we will be using are daily closing prices for the S&P 500 index from 1986 to 2018.
The S&P 500, or just the S&P, is a stock market index that measures the stock performance of 500 large companies listed on stock exchanges in the United States. It is one of the most commonly followed equity indices, and many consider it to be one of the best representations of the U.S. stock market. link

Steps that we will be using in this project

- Load and Inspect the S&P 500 Index Data
- Data Preprocessing
- Temporalize Data and Create Training and Test Splits
- Build an LSTM Autoencoder
- Train the Autoencoder
- Plot Metrics and Evaluate the Model
- Detect Anomalies in the S&P 500 Index Data

Conclusion

Time Series Anomaly Detection with LSTM Autoencoders can be used to solve many problems such as Cybersecurity breaches, online fraud detection and prevention, predictive maintenance and condition monitoring in various industries including Manufacturing, E-commerce, Banking, Retail, Oil and Gas, Medicine. The value of detecting different anomalies in a regular flow of numerous operations, whether it is about making credit card transactions or eliminating problems in a device work, is hard to overestimate, especially when it goes about the prediction of unexpected anomalies that can be an important factor to influence the enterprise/business income.