

# Anomaly Detection in Real-Time Data with Generative Adversarial Network (GAN)

---

Nathen Byford

Nathen Byford

# Contents

1. Motivation and goals
2. Methods
3. Data

# Motivation and Goals

Nathen Byford

# Motivation

real-time or streaming data analysis is the backbone of modern data-driven decision-making. A key part of utilizing real-time data is having clean data.

Garbage in, garbage out.

In time series processes that are observed in real time it is even more important to be able to detect anomalies so that the cause can be identified.

Anomalies can be caused by:

- Faulty sensor
- Poor data quality
- External factors
- Adversarial Attacks

# Goals

There are many methods being used to detect anomalies in streaming data with data driven methods. Typically clustering algorithms, time series analysis, change point detection, and even statistical tests.

My goal is to utilize Generative Adversarial Network (GAN) models to detect anomalies on the Numanta Anomaly Benchmark (NAB) data set.

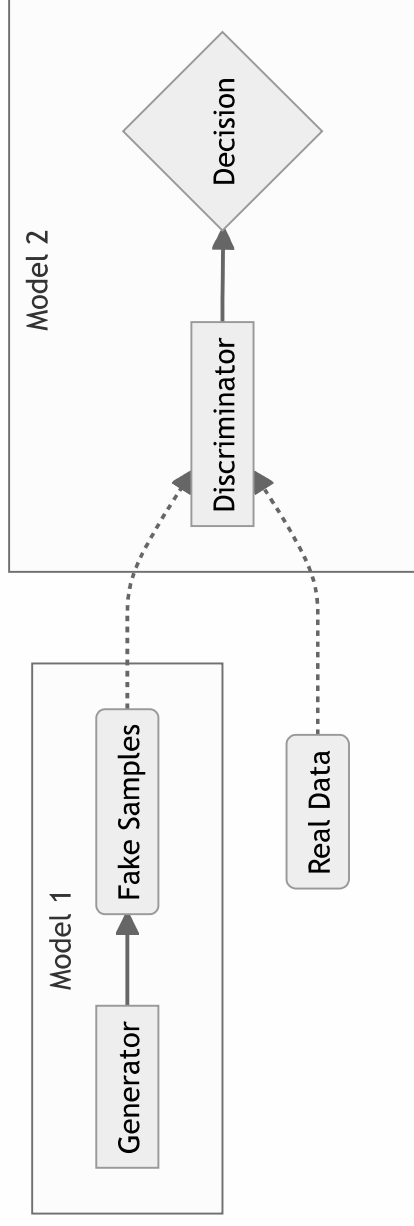
# Methods

Nathen Byford

# Generative Adversarial Network

GAN models, set two AI algorithms against each other, hence “adversarial”.

These two models are:



1. Train the discriminator on real data
2. Use the generator to make a fake data set
3. Have the discriminator decide if the fake data was real or fake
4. Repeat steps 2-3 until the fake data is indistinguishable from the real data.

# For this project

In this project I plan to use these generators from the GAN to predict future values of the time series.

Using these future values of the time series we can determine if the observed values are reasonable based on the prediction of previous data.



# Data

Nathen Byford

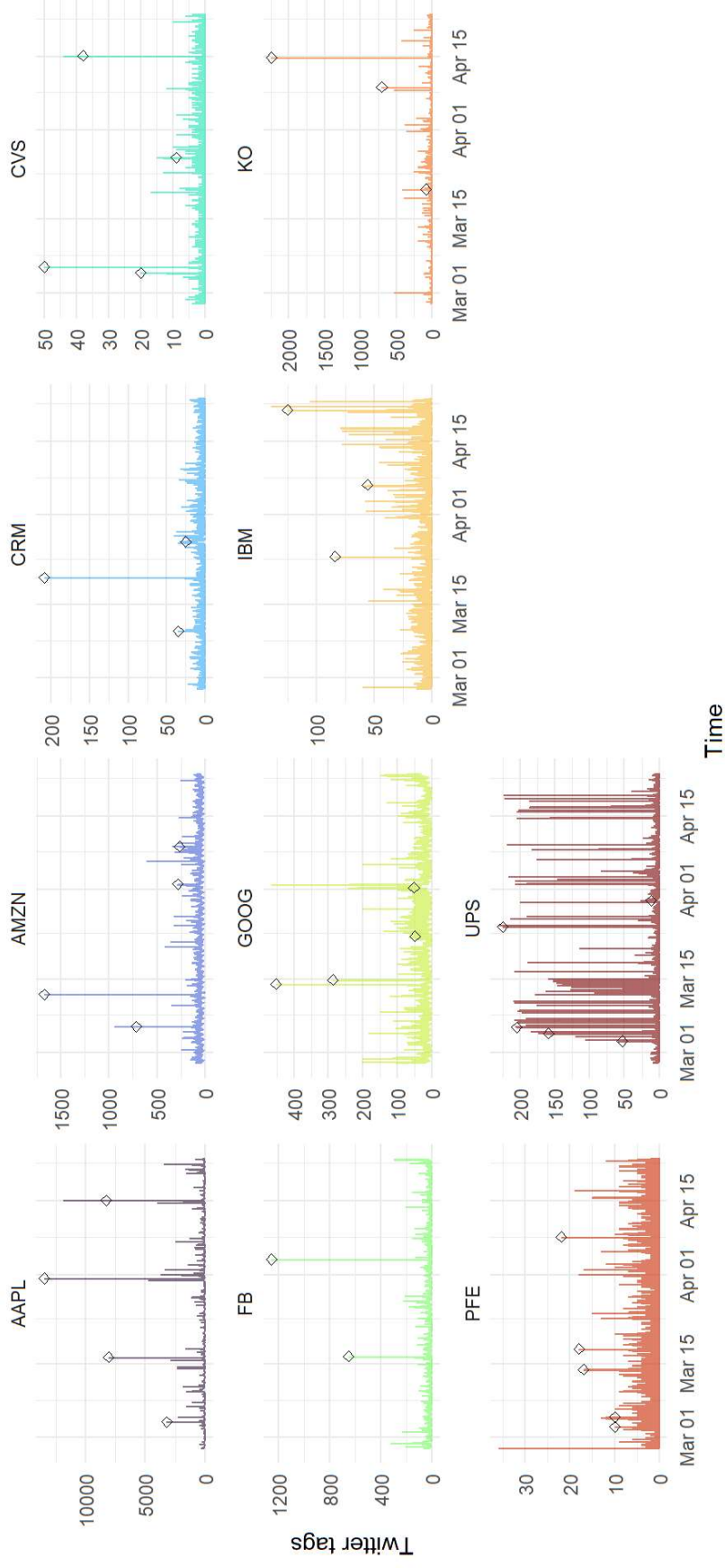
# Numenta Anomaly Benchmark

Numenta, an AI company had open sourced a data set for the purpose of benchmarking anomaly detection algorithms.

- This data set is composed of a couple different data sets, I choose to look at just their twitter tag time series.
- The data set contains the numbers of daily tags for 10 corporate users from February 26 through April 23 2015.
- Four variables: date, number of tags, company, and anomaly.

# EDA

Time plot of Twitter tags with anomalies



CRM is salesforce, and KO is Coca-Cola

Nathen Byford

# Thanks

Nathen Byford

Nathen Byford