# Anomaly Detection in Real-time Ad Performance

**Cassie Huang**
Center for Data Science
New York University
qh2044@nyu.edu

**Helen Zhang**
Center for Data Science
New York University
qz2086@nyu.edu

**Chenxin Gu**
Center for Data Science
New York University
cg3423@nyu.edu

## 1 Question to Answer

Our intended research topic is **anomaly detection in real-time Ad performance**. The key question to answer is how can various anomaly detection algorithms effectively identify anomalies in real-time ad performance metrics such as cost-per-click (CPC) and cost-per-thousand-impressions (CPM), and what model provides the most accurate early detection? **Our study will evaluate multiple models to identify Ad performance anomalies on the NAB dataset.**

## 2 Dataset

We will utilize the **realAdExchange** dataset, which provides insights into online advertisement performance through metrics such as CPC and CPM. CPC represents the amount paid by advertisers each time a user clicks on their ad, while CPM reflects the cost incurred for every one thousand impressions of an advertisement. Each dataset is organized as a time series, where each entry corresponds to a specific time interval (e.g., every hour). This structure allows for the analysis of trends, seasonality, and anomalies over time. The datasets include instances of known anomalies, such as unusual spikes or drops in CPC and CPM values, which will serve as benchmarks for evaluating the effectiveness of the selected anomaly detection methods.

## 3 Methods

**Time Series Models**. We will employ various statistical approaches, including **ARMA** (AutoRegressive Moving Average) models to capture patterns in the time series data, **HMM** (Hidden Markov Models) to account for state transitions, **EWMA** (Exponentially Weighted Moving Average) for smoothing and detecting gradual shifts, and **CUSUM** (Cumulative Sum Control Chart) for identifying abrupt changes in the series.

**Machine Learning Techniques**. On the other hand, we will implement advanced machine learning models such as **LSTM** (Long Short-Term Memory networks) that are designed for time series data. Additionally, we will explore **one-class SVM** (Support Vector Machine) to model the normal behavior of the advertisement metrics and identify deviations as anomalies.

## 4 Evaluation Metric

The Numenta Anomaly Benchmark (NAB) Score is the evaluation metric used in this project, designed for real-time anomaly detection from [1]. It rewards early and accurate detections while penalizing late detections and false alarms using a weighted scoring system. This flexibility allows for prioritization of minimizing detection delay or reducing false positives, making it ideal for time-sensitive metrics like CPC and CPM.

## 5 Allocation of Work

Allocation of work of this project is as follows: Helen will train the CUSUM and EWMA models, Cassie will handle ARMA and SVM, while Chenxin will focus on LSTM and HMM. The team will then collaborate to evaluate all models, compare their performance, and finalize the presentation together.

## References

[1] Alexander Lavin and Subutai Ahmad. Evaluating real-time anomaly detection algorithms–the numenta anomaly benchmark. In *2015 IEEE 14th international conference on machine learning and applications (ICMLA)*, pages 38–44. IEEE, 2015.