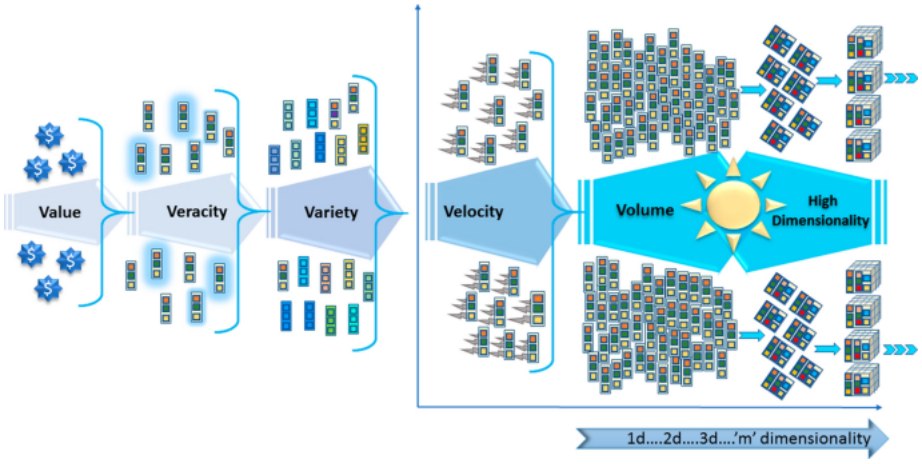
AI2002 Artificial Intelligence
Real-Time Anomaly Detection in Big Data Streams using Machine Learning

Sharif Ali
P20-0130

Use a lot of diagrams to describe the problem and your proposed solution.

1 Problem Statement

The problem to be addressed is the difficulty of detecting anomalies in big data streams in real time, which is crucial for many applications such as fraud detection and network intrusion detection. Anomalies can occur due to various reasons, such as sensor errors, malicious attacks, or equipment malfunctions, and can have severe consequences, such as system downtime or financial losses. Therefore, there is a need for an efficient and accurate anomaly detection system that can process data streams in real-time.



2 Motivation

Anomaly detection is an important and challenging problem in big data analytics, and the ability to detect anomalies in real-time can lead to better decision-making and improved security.

2.1 Example

In today's world, most people prefer to use credit cards for online shopping or other transactions. However, fraudulent transactions can cause a significant loss to both the credit card provider and the customer. Detecting these transactions in real-time can help prevent the fraud from occurring or minimize the loss.

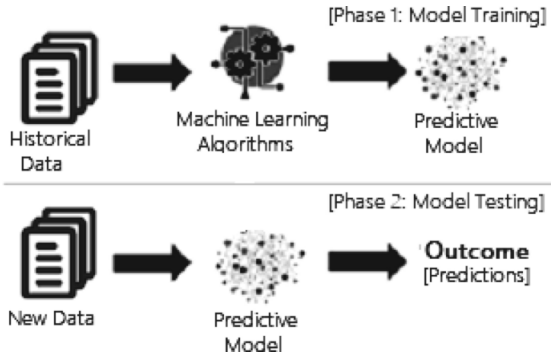
2.2 Solution for examples problem

The challenge with detecting fraudulent transactions is that it requires processing large amounts of data in real-time, which can be time-consuming and computationally expensive. Therefore, the need for efficient and accurate

real-time anomaly detection algorithms that can handle big data streams has become increasingly important.

3 Background

To understand the problem, readers need to know about the challenges of processing big data streams, such as the need for real-time processing, high volume, and variability of data. Readers also need to be familiar with machine learning techniques, such as clustering and classification, commonly used in anomaly detection.

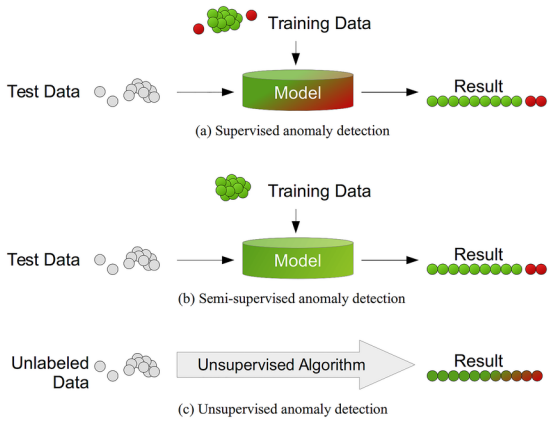


4 Related Work

Several studies have been conducted to address the challenge of real-time anomaly detection in big data streams. For example, [1] proposed an unsupervised anomaly detection method that used clustering techniques to identify anomalies in data streams. [2] developed a framework for scalable and automated anomaly detection using machine learning techniques. These studies provide a foundation for our proposed work.

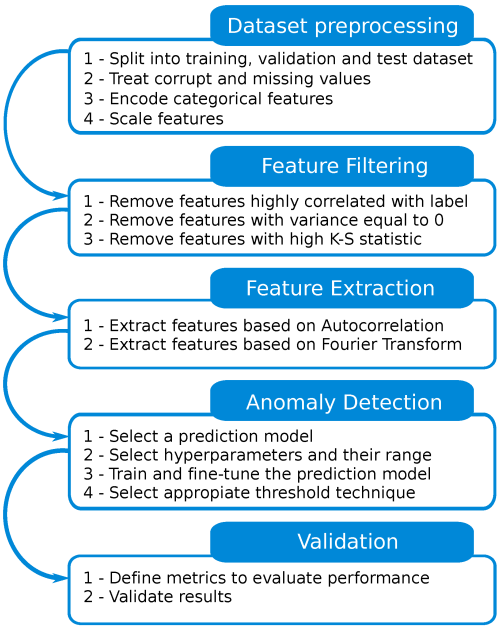
5 Proposed Work

The proposed work will involve the development and implementation of a machine learning-based anomaly detection system for big data streams. A diagram of the proposed system can be included, along with a description of the feature extraction and selection process, and the machine learning algorithm used for anomaly detection.



6 Evaluation Methodology

We plan to evaluate our implementation using real-world datasets and compare its performance with existing anomaly detection methods. We will evaluate the system's accuracy, precision, and recall to assess its effectiveness in detecting anomalies. We will also measure the system's scalability and efficiency to evaluate its suitability for large-scale applications.



7 Hypothesis

We hypothesize that our proposed real-time anomaly detection system will provide efficient and accurate detection of anomalies in big data streams, outperforming existing methods in terms of accuracy and scalability.

8 Proposed Timeline

The tentative weekly timeline giving concrete milestones would be as follows. Modify the following timeline for your project.

- **February 20**: This proposal.
- **March 20**:Familiarization
- **March 30**: Implementation of step1 ...
- **April 6**:step2
- **April 14**:step3
- **April 20**: Evaluation in progress.
- **April 27** : Complete experiments.
- **April 28**: Writing report.
- **May 4**: Presentation and Final report due.

9 References

9.1 Reference

[1] Chauhan, K., & Deshpande, B. (2019). Unsupervised Anomaly Detection in Streaming Data. In Proceedings of the 10th ACM International Conference on Data and Application Security and Privacy (pp. 157-167).

9.2 Reference

[2] Gupta, D., & Gupta, A. (2019). Scalable Automated Anomaly Detection using Machine Learning for Big Data Applications. Journal of Big Data, 6(1), 1-23.