


# MACHINE LEARNING BASED ATTACK DETECTION SYSTEM IN CLOUD COMPUTING

CLOUD COMPUTING



**Team-5**

Amruth Mandappa T S  
Ashwini Kumar  
Arjun Sagar NV  
Dhanisht Kumar  
Yugal Deep Singh



# INDEX

- **Services Used and their benefits**
- **Data Set**
- **Implementation**

# Services Used



# SageMaker

- Amazon SageMaker is a machine learning service offered by Amazon Web Services (AWS).
- It provides a platform for building, training, and deploying machine learning models.
- With SageMaker, user can easily create machine learning models without worrying about the underlying infrastructure.
- It offers a wide range of built-in algorithms and frameworks for machine learning.
- SageMaker provides a fully managed environment with automatic scaling, so it can handle large-scale machine learning workloads.
- It also offers features like versioning, security, and integration with other AWS services.
- Its ease of use and scalability make it an attractive option for businesses looking to adopt machine learning.

# Benefits of SageMaker

1. **Easy to use:** SageMaker offers a simple and intuitive interface that allows developers and data scientists to quickly create, train, and deploy machine learning models.
2. **Scalable:** With SageMaker, you can easily scale your machine learning applications up or down based on the demand, without worrying about infrastructure management.
3. **Cost-effective:** SageMaker offers a pay-as-you-go pricing model, which means you only pay for the resources you use, making it a cost-effective option for machine learning projects.
4. **Built-in algorithms:** SageMaker includes a range of pre-built algorithms for common machine learning tasks, such as image classification and text analysis, which can save time and effort in developing custom models.
5. **Secure:** SageMaker offers multiple layers of security and compliance to protect your data and ensure regulatory compliance.

# Dataset

- KDD99 has been widely used as a benchmark dataset for evaluating intrusion detection systems, and many machine learning algorithms have been applied to it for developing effective intrusion detection techniques
- It is a well-known dataset for network intrusion detection research, which was developed by the MIT Lincoln Laboratory in 1999.
- It contains a large number of network traffic data, which were collected from a simulated environment that mimics a typical US Air Force LAN.
- The dataset consists of approximately 4.9 million connection records, with each record containing 41 features such as protocol type, service, source and destination IP addresses, and flags.
- The goal of KDD99 is to classify network traffic as normal or suspicious and identify the type of attack if it is detected.

# Implementation

- We utilized XGBoost to analyze big datasets and discover patterns that are suggestive of intrusion in the context of cloud computing.
- This technique works by building a succession of decision trees that are trained on different subsets of data repeatedly.
- We chose XGBoost because it is both quick and scalable, making it appropriate for huge datasets. It also has great accuracy and can deal with missing values and outliers in the data.

# THANK YOU



Hope You Liked It