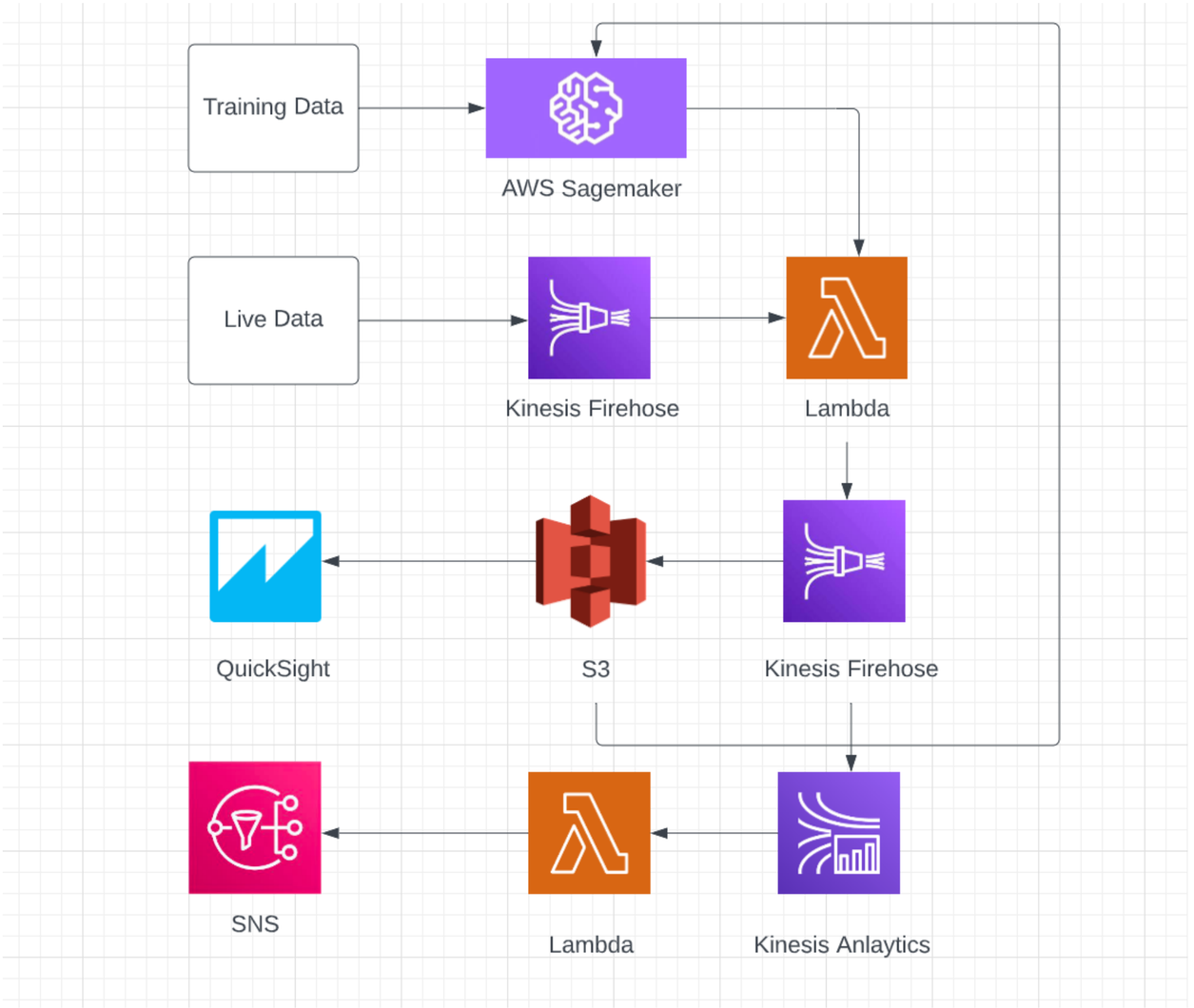
Fraud Detection in Real-Time using AWS

Aditya Pawar
K.J. Somaiya College of Engineering

Introduction

- Anomaly detection is an important aspect of analysing sensitive and critical data
- A technique used to identify rare events or observations which are suspicious in nature
- The following solution leverages AWS for real-time data streaming, fitting it to an appropriate classification model and deploying it as a service endpoint
- The endpoint receives real-time data and classifier tells us whether fraud is detected or not
- An alert will be sent to the client as well as the bank with the necessary details about the same

Architecture



Data to be used

- Currently, I am using the public dataset available on Kaggle for demonstrating the solution
- Link: **https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud**
- The dataset contains 30 variables, among which 28 are principle components obtained with PCA, the only variables which have not been transformed with PCA are '**Time**', '**Amount**' and '**Class**'.
- The dataset is highly unbalanced, the positive class (frauds) account for **0.172 %** of all transactions
- I am also exploring other datasets available publicly

Choosing the right model

- My solution proposes the use of self organising maps, an unsupervised deep learning model
- Based on competitive learning
- The model creates a client segmentation so one segment contains the potential frauds
- The winning neuron i.e. the neuron closest to the specific client will be picked and it's neighbourhood neurons
- With every updation, the neighbourhood radius decreases and we get closer to the fraud

References

- https://aws.amazon.com/solutions/implementations/fraud-detection-using-machine-learning/
- https://www.kaggle.com/
- https://towardsdatascience.com