

Logs Data/monitor

Server /cloud infrastructure

[Logs]

App Request

Security Service & Model

**With TensorFlow Keras and labeled data**

-Tokenize text to sequence of token indices

-Pad sequences to desired sentence length

-Compile the model

-Train the model

-Run predictions

-Adjust model for better results

Classification was chosen over clustering because it provided better results for text classification.

Test model

Determine likelihood of SQLi accuracy

[set acceptable boundary]

[update] Model in production [CD]

CheckIfSQLi(request\_text\_from\_webApp)

Automated Response

**Benign** login attempt looks like:

SELECT \* FROM users WHERE email = ’xxx@gmail.com’ and password=’Mypwd1$’

**Malicious** request may look like:

SELECT \* FROM users WHERE email = ’xxx@gmail.com’ and Password = **”or ‘1’=’1’**

Preprocessing data for model

Remove unnecessary columns from data source. Add columns if needed. [enrich with other DS]

Resulted in 2 columns: request and label [1,0]

Try to understand training and test data. Remove any unnecessary words and characters. This step was minimal because every character, word e.g. AND, OR, IN, or special character is important to detect sqli. Use Word count, word cloud, graphs to visualize data.

[SQL Injection Detection Using Machine Learning (sjsu.edu)](https://scholarworks.sjsu.edu/cgi/viewcontent.cgi?article=1727&context=etd_projects)

[Detecting Web Attacks Using Multi-stage Log Analysis (sc.edu)](https://cse.sc.edu/~huangct/CSCE813F16/07544930.pdf)

[NLP with Python: Text Clustering - Sanjaya’s Blog (sanjayasubedi.com.np)](https://sanjayasubedi.com.np/nlp/nlp-with-python-document-clustering/)