HITECH NOTEBOOK SUMMARIES

# HITECH\_cleaned.ipynb Summary

Objective: Clean and preprocess the HITECH dataset for further analysis.

* Handled missing values, cleaned columns.
* Parsed dates, created severity buckets, encoded categories.

Dataset ready for modeling.

# HITECH\_EDA.ipynb Summary

Objective: Perform exploratory data analysis on HITECH breach dataset.

* Breaches per year/month, types, severity visualized.
* Correlation and insights analysis.

Clear understanding of breach trends.

# HITECH\_FeatureEngineering\_SeverityModel.ipynb Summary

Objective: Feature engineering and severity modeling with LightGBM.

* Extracted, encoded features.
* Built and evaluated LightGBM model.

Identified key features influencing severity.

# HITECH\_LSTM\_Anamoly.ipynb Summary

Objective: Detect anomalies in monthly breaches with LSTM autoencoder.

* Aggregated, normalized monthly data.
* Trained LSTM to detect anomalies.

Identified unusual breach activities.

# HITECH\_CNN\_SeverityModel.ipynb Summary

Objective: CNN-based severity classification from text descriptions.

* Tokenized, padded data.
* Built CNN model with early stopping.

Classifier differentiated severity effectively.

# HITECH\_Topic\_Classifier.ipynb Summary

Objective: Topic modeling (LDA) and severity classification (CNN).

* Text preprocessing, TF-IDF, LDA.
* CNN classifier for severity prediction.

Gained insights and accurate classification.