# .5-78020,91

# Conformal LSTM Classifier for Time-Series Anomaly Detection

Presented By: Anish Rao

**Professor:** Devesh Jawla

```
WILL 113 Inesseries
       ad clais Progle (o : 27 Irtuarful Viedils ell)
       Mesites: = Inidoer : Prado_Sereal>
Straterient, Detioweny Rubal Notaced Timpertar
Mettiensel: 1513: 6558) +4318_6090)
personal: canturdaylild = castefine air popless fistt, Pacacles for 1850
Orrughd, lastal: Twited Geraserietl>
  aly pertonrce = Latgle Wicmo>
  crtal: Wrwdll.lesits:
    wrices fraptint coplechencissictes, 16 reduer antumal notar action land
           had! for carlee this liwiced Neuriel New Mill
           ( 0: for tarflytions al tiolacy uthacting Its: Rest = Deci
     tarfuloren.re: (ary_invlice = godlsfetil, utin bet, round at sales and
       155Tons(pall, = rust; = juli lames, = chtt the opd on the to a
   curiles titer inrales)
cetwies involourthes dat):
         tnarical onoite seraplares; tomi sslattoreli
   cirles lest to for latstutes,
         tree: franciser hat assing blesti):
         (() no: netsrise engoleses; = towile (rep No.)
                 vecaess the selents (( setable
```

# Project Objective & Scope



# Main Objective

Use Julia to Implement a conformal LSTM classifier to detect anomalies

?

# Scope of the Project

# Why LSTM?

- A type of model that helps understand data over time
- It remembers
   patterns of the past
   to make better
   future decisions

# Why Conformal **Prediction?**

- Helps us know how sure the model is about its decision
- Instead of guessing a number to detect anomalies, it calculates a limit from the data itself.

### **About NAB**

- A large collection of time-based data
- Contains real and fake data
- It's commonly used to test how good anomaly detectors are.

# NAB Leaderboard: What Top Models Do

The top-performing models on the NAB leaderboard use a **diverse mix of strategies**, mostly **unsupervised**, to detect anomalies in time series:

# HTM (Hierarchical Temporal Memory):

Inspired by the human brain, it learns patterns over time and detects anything that breaks the flow

# **Bayesian Models:**

Use probability to guess how likely a new value is — if it seems very unlikely, it's marked as an anomaly.

# KNN-CAD (Nearest Neighbour):

Compares current patterns to past ones — if they look very different, something's probably wrong.

# Forecasting Models (like ARIMA):

Try to predict what's coming next based on past trends — big surprises are flagged as anomalies.

# **Ensemble Models:**

Combine the opinions of multiple small detectors to make better overall decisions.

**Common Goal:** 

Minimise false alarms while catching real, subtle anomalies — especially in **noisy, unlabelled, real-world data**.

# How It Works

# **Unsupervised (Forecasting)**

- Data: Only values
- Training: predict the next value and minimise the MSE between predicted and actual values.
- Conformal Prediction: use residuals
   ( | predicted actual | ) to calculate
   threshold
- Evaluation: Predict next values,
   flag if residual > threshold

# **01**. Data Preparation

Normalise Data → Split into train/ calibration/ test→ Create windows/sequences

# 02. Training

Create model → Train model

# 03. Conformal Prediction

Create residuals → Set threshold (Confidence level)

### 04. Evaluation

Predict anomalies → Compare to true labels for metrics

# **Supervised (Classification)**

- Data: Both values and labels
- Training: Classify sequences as normal or anomaly using Sigmoid + Binary Cross-Entropy Loss.
- Conformal Prediction: use residuals
   ( | probability actual | ) to calculate
   threshold
- Evaluation: Predict probabilities, flag
   if residual > threshold

# **Key Results**

Unsupervised model outperformed the supervised

one in 4 out of 6 real-world categories

Highest precision

88.89

Supervised

61.9

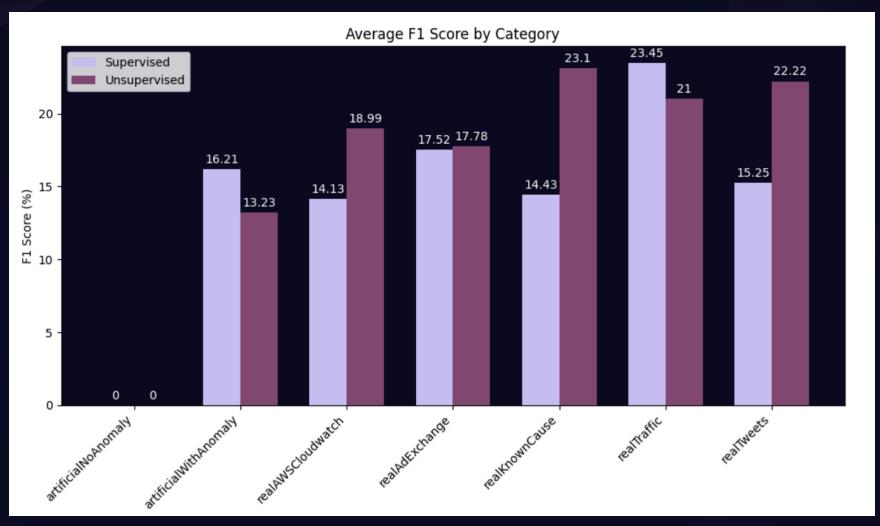
Unsupervised

Highest recall

81.05
Supervised
Unsupervised

Unsupervised caught more anomalies, but also had more false alarms.

Supervised was more precise, but missed several real anomalies.



Model	TP (Correct Detections)	FP (False Alarms)	FN (Missed Anomalies)
Supervised	8780	37991	37362
Unsupervised	16585	76082	34569

# Conclusion

The unsupervised model shows higher overall recall and slightly better F1-score, highlighting its strength in detecting anomalies broadly, while the supervised model achieves slightly higher precision, reducing false alarms.

Unsupervised model performed betteron noisy, real-world data like servermetrics and social media.

Supervised model excelled on clean, synthetic data with clearer anomaly patterns.

