Enhancing Alarm Intelligence through machine learning at OSUM Oil Sands

BUSINESS ANALYTICS CAPSTONE PROJECT – SPRING 2025

VAMSHI KRISHNA KUNCHALA, KARTHIK TULUGU, SAI SURYA KANDELA, SHASHANK BHARADWAJ ASARAM

04-24-2025

Introduction

Objective:

- Analyze alarm data to address alarm management challenges.
- Clean, explore, and model data to understand alarm occurrences.

Dataset Overview:

- Includes alarm tags (Alarm types), chattering behavior (CHB), time-based metrics.
- Contains 20,000 rows and 20 columns.

Data Collection & Preparation

Data Import:

- Imported Data using R's readxl package.
- File: "IM009B-XLS-ENG.xlsx"

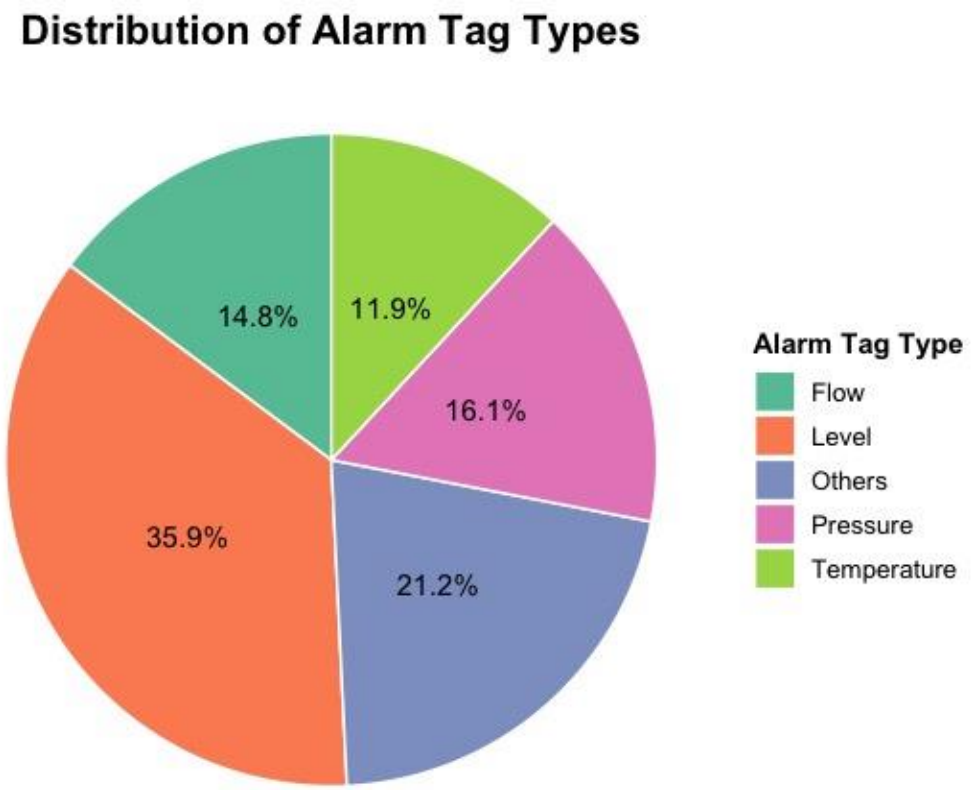
Preparation Steps:

- Converted Variables (e.g., ATD, CHB) to numerical/categorical.
- Removed missing values with drop_na()
- Split data: 70% training, 30% testing.

Exploratory Data Analysis (EDA) - Overview

Key EDA Areas:

- Distribution of Active Time Duration (ATD)
- Frequency of Alarm Tag Types
- Hour-based Alarm Analysis
- CHB (Chattering behavior) Analysis



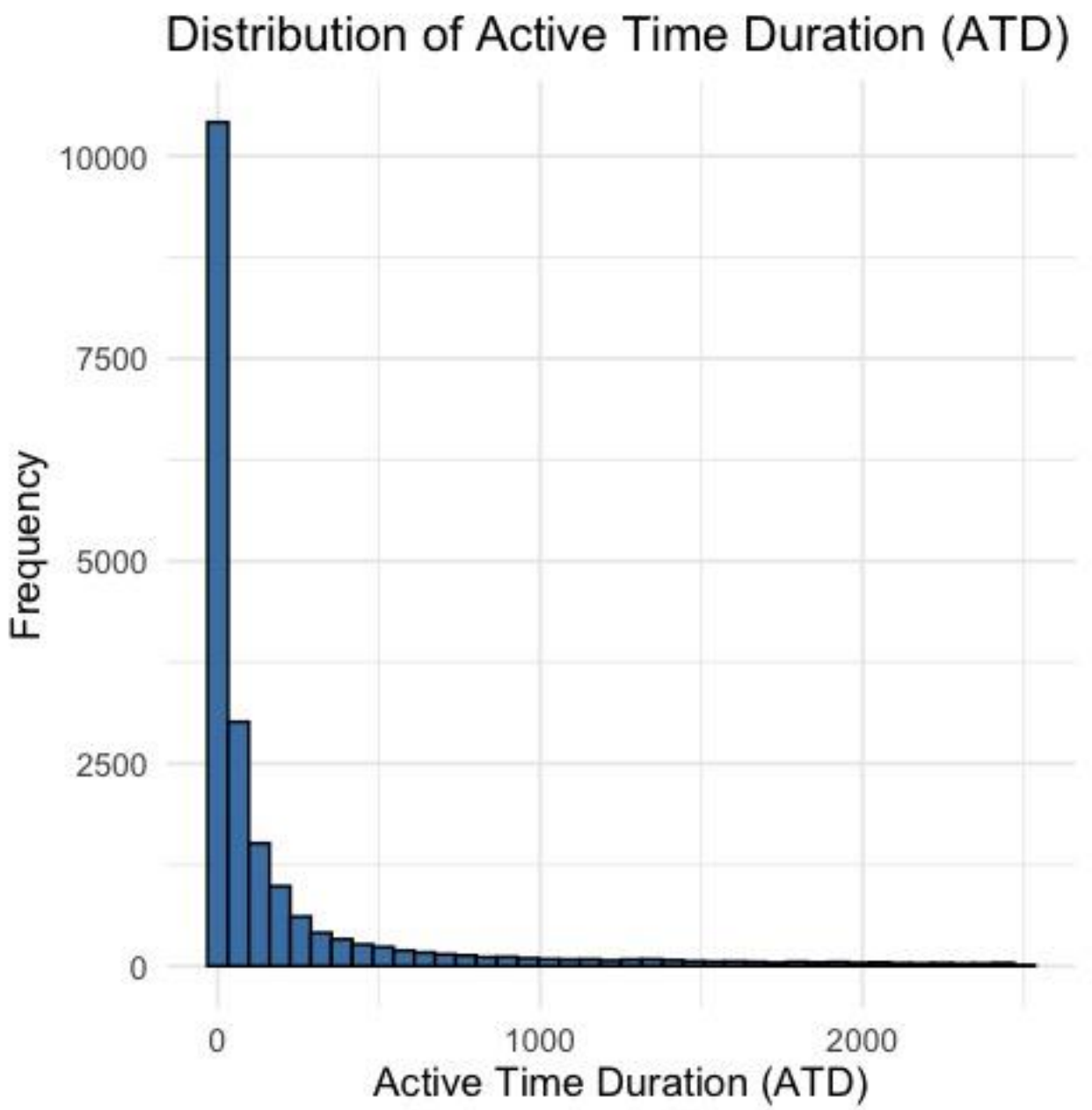
EDA – Distribution of ATD

Key Findings:

- Right-skewed distribution: most alarms resolved quickly

Implication:

- Investigate long-duration alarms to improve processes.



EDA – Alarm Tag Type Frequency

Key Findings:

- Level alarms:

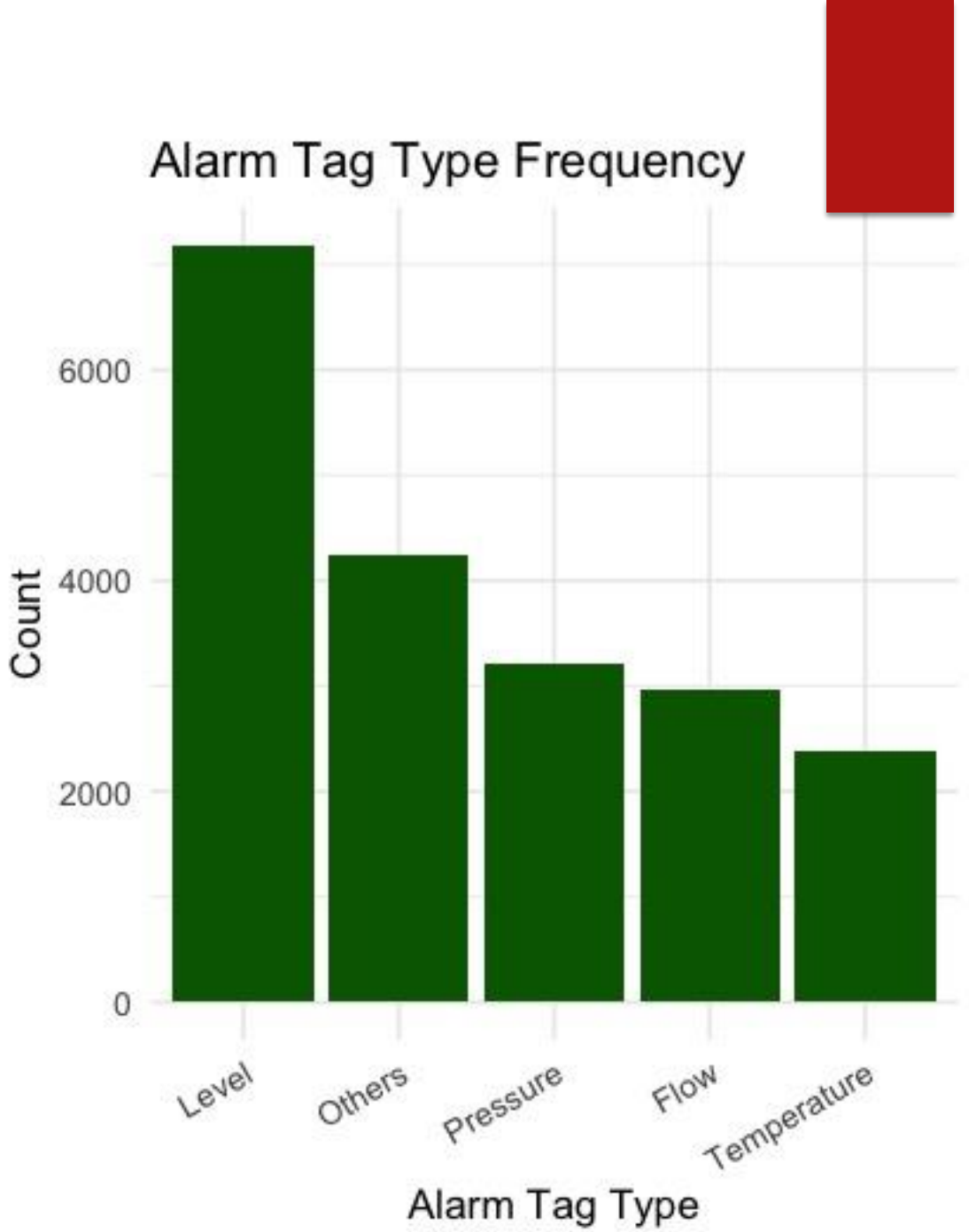
  Highest count (~6,000).

- Others category:

  High volume, needs refinement

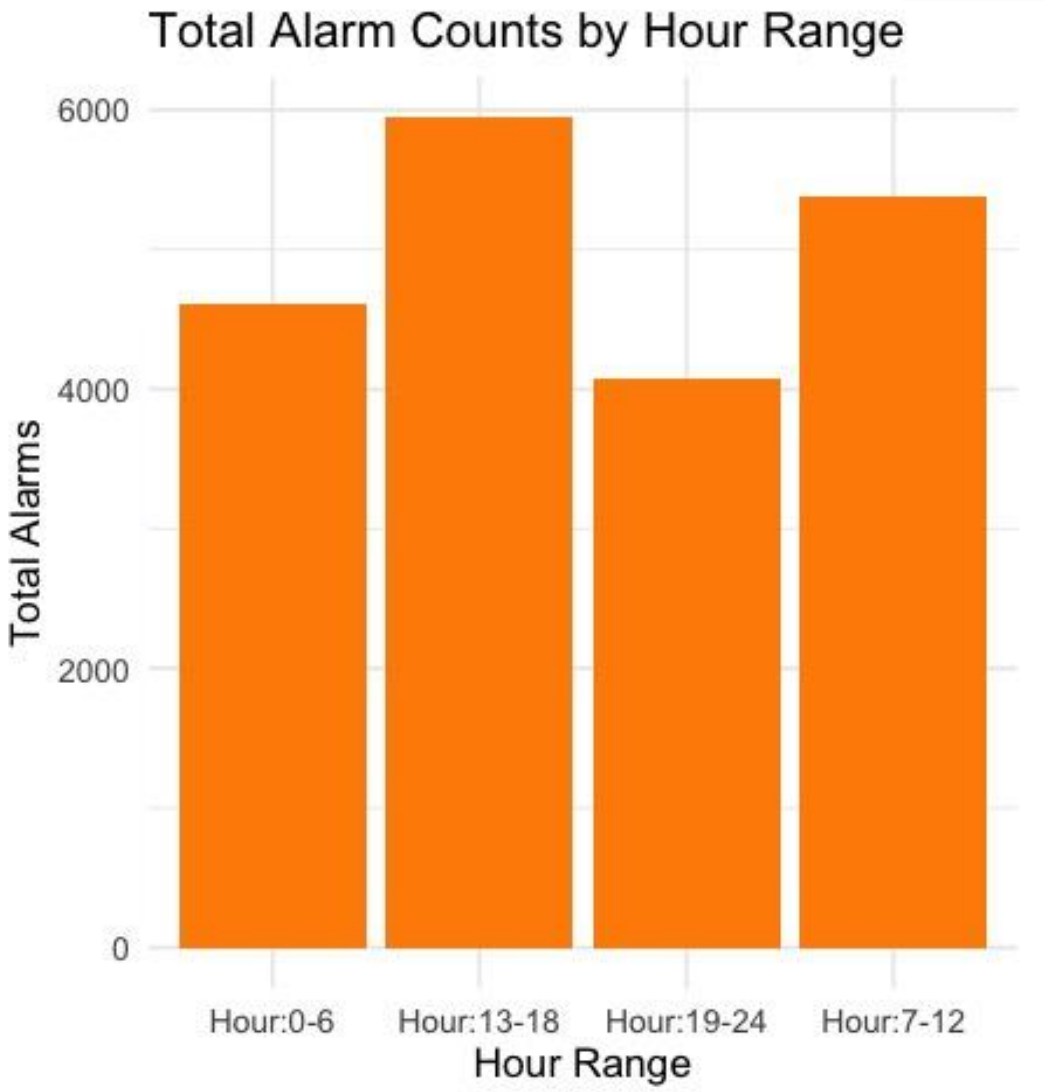
- Implication:

  Reclassify unclear alarm types



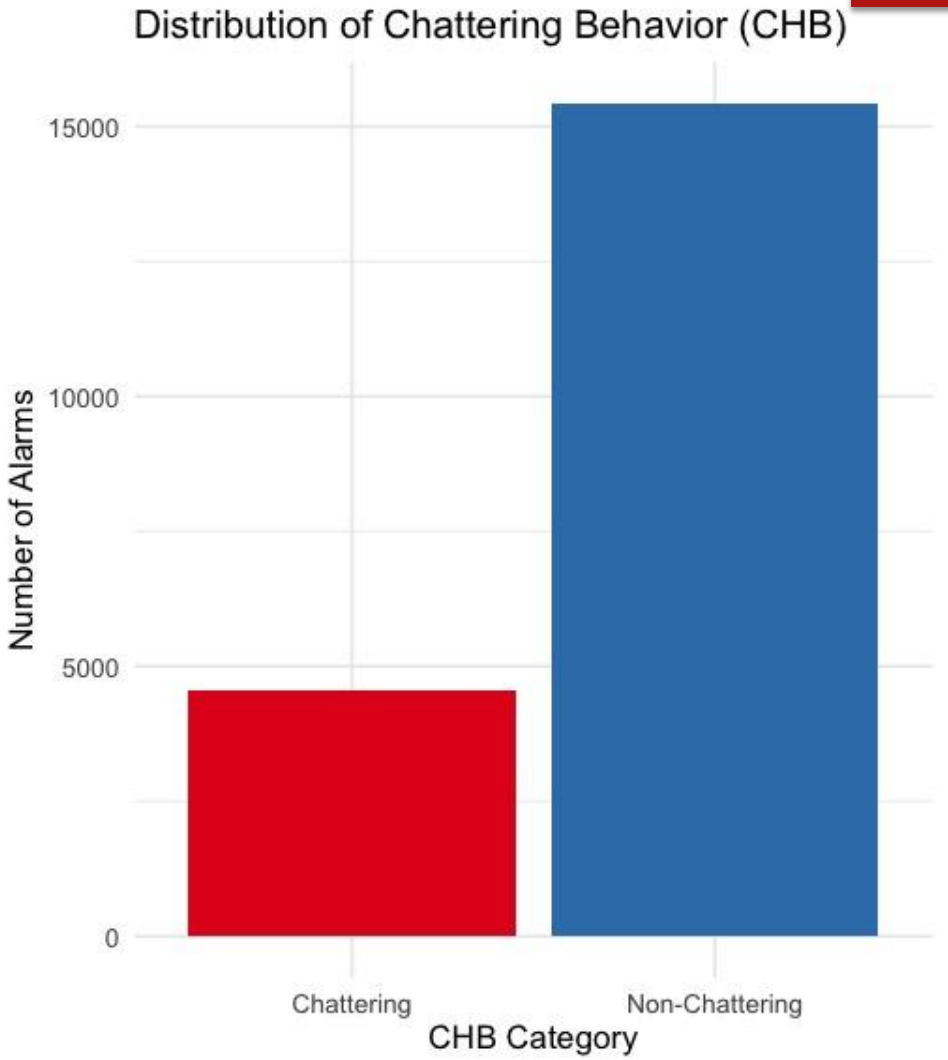
EDA – Hour-Based Alarm Analysis

- **Highest activity:** 13:00 – 18:00
- **Moderate:** 0:00–6:00 and 19:00–24:00
- **Implication:** Afternoon shifts need more staffing and monitoring



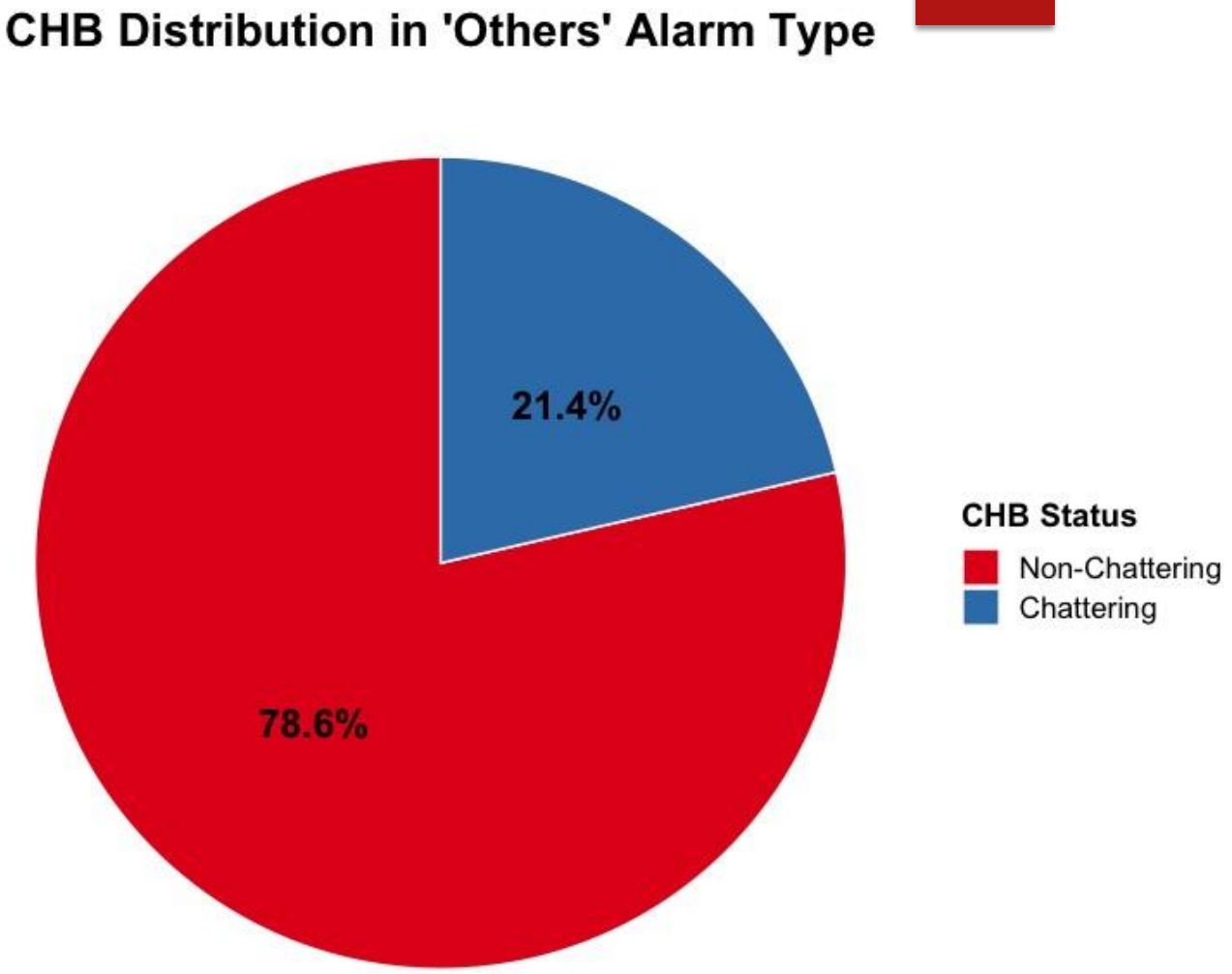
EDA - CHB – Chattering Behavior Analysis

- **CHB =1 (Chattering):** ~25%
- **CHB=0 (non-chattering):** ~75%
- **Implication:** Class imbalance → emphasize recall in models



Exploring Alarm Noise in Miscellaneous Tags ('Others')

- ~21% of "Others" are chattering
- Many alarms poorly labeled (~79%)
- **Implication:** Clean/reclassify to reduce noise



CHB Classification – Modeling Approach

- **Problem Type:** Binary Classification

Target: CHB (0 = No, 1 = Yes)

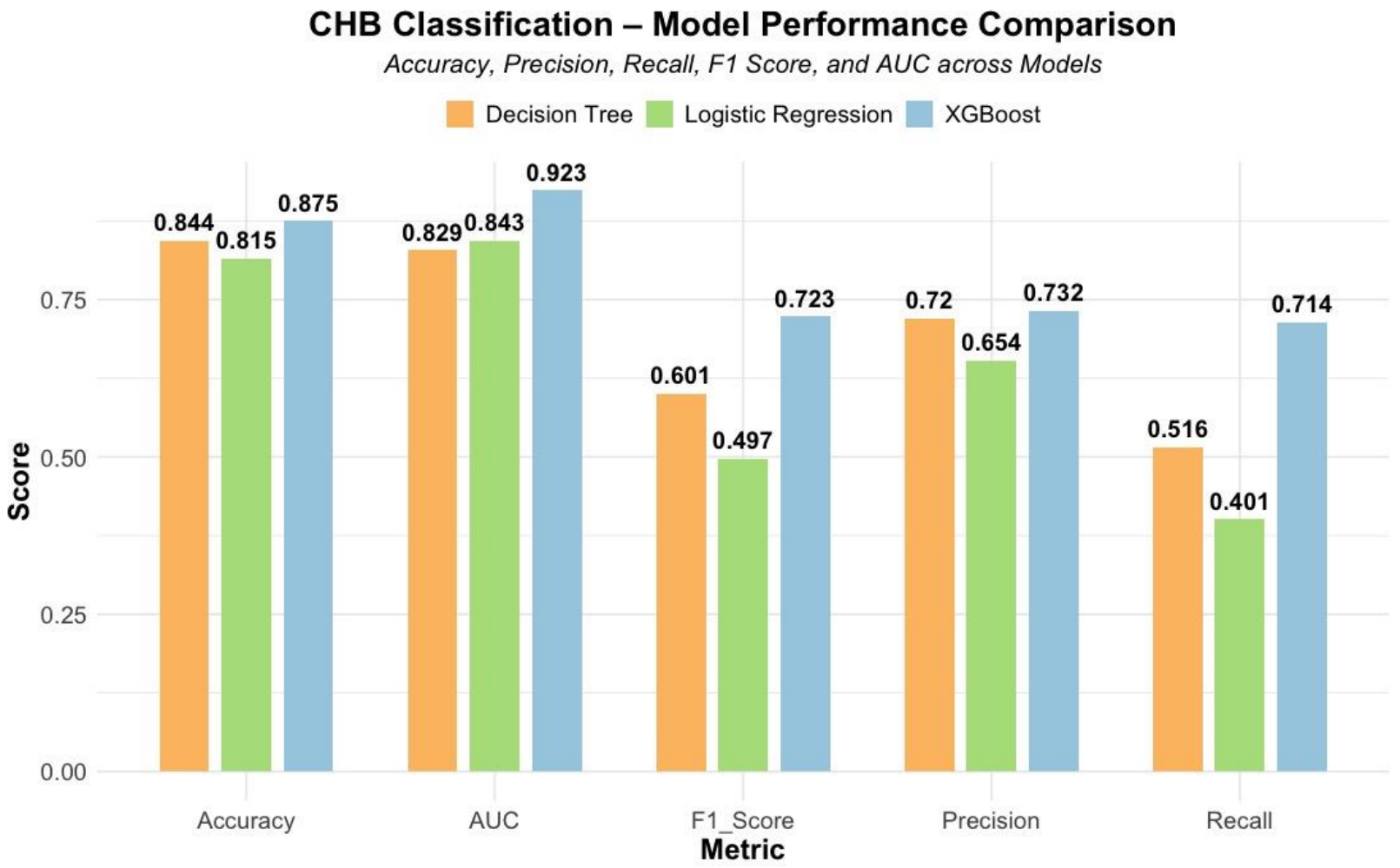
Models Used:

- Logistic Regression
- Decision Tree
- XGBoost

CHB Classification – Model Evaluation

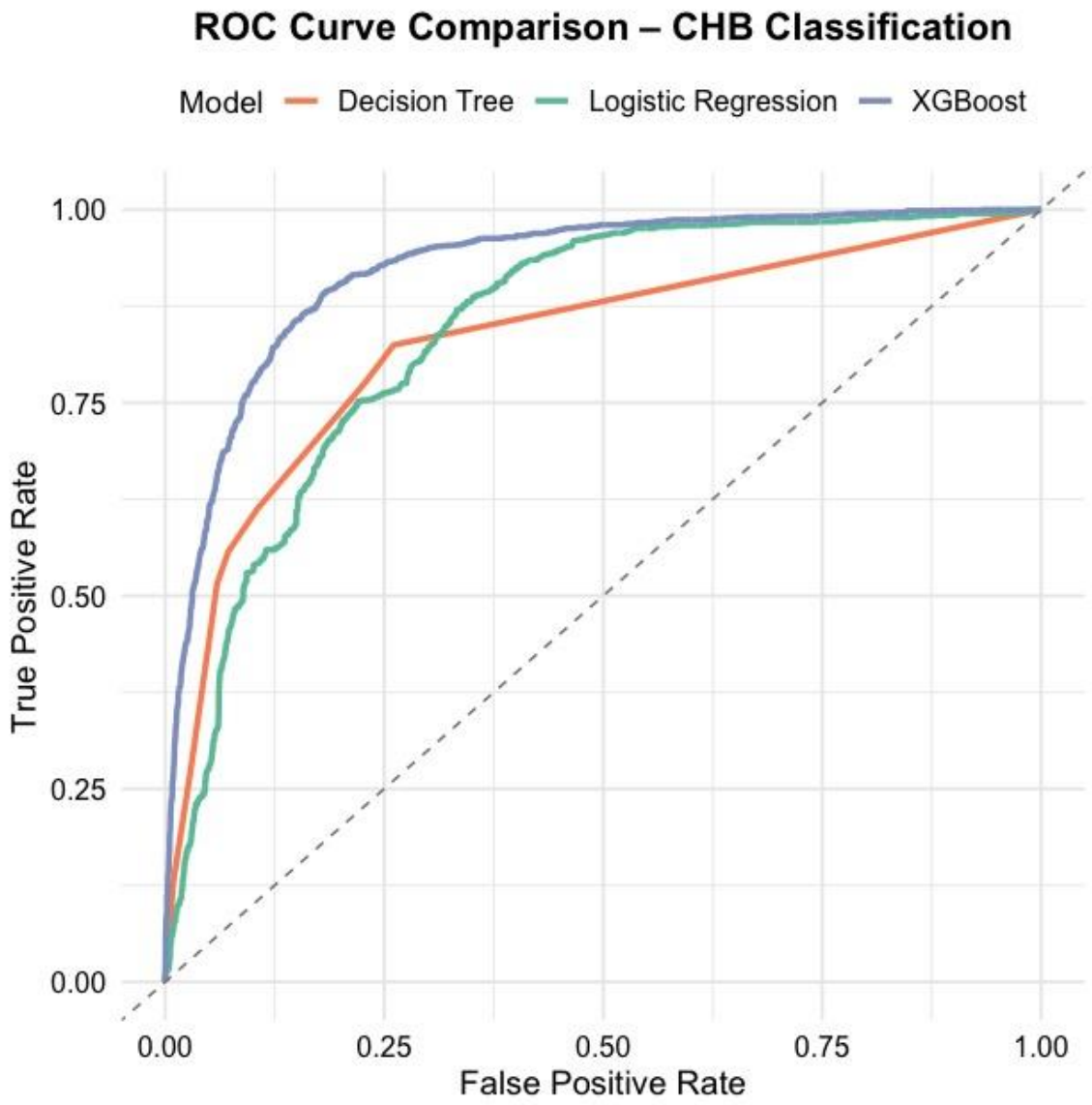
XGBoost Highlights:

- Best performance across all metrics
- **Recall:** 66% (crucial for chattering)
- **AUC:** 0.91
- **Conclusion:** Ideal for real-time deployment



CHB ROC Curve Evaluation

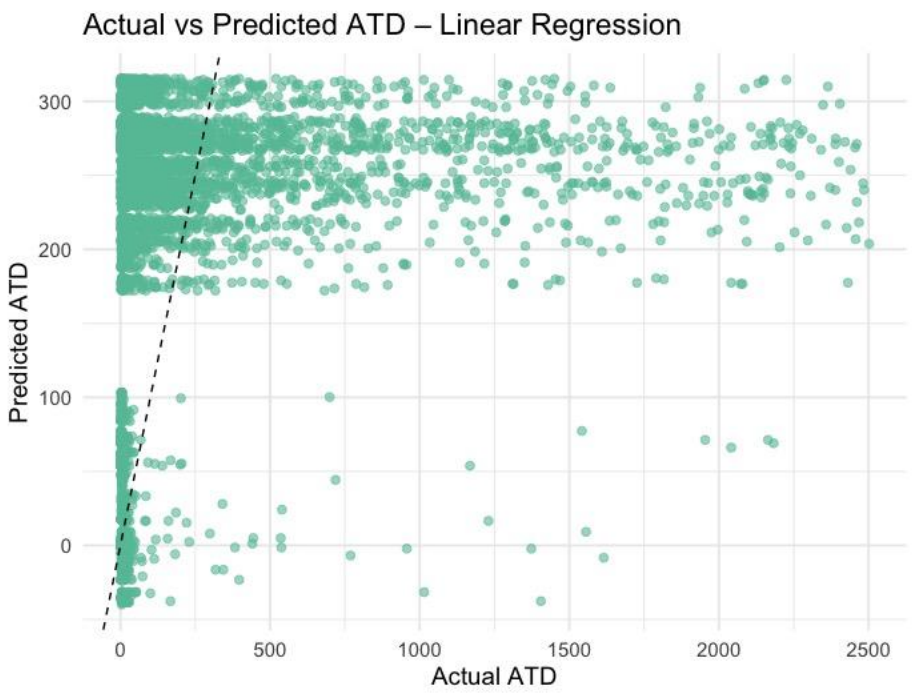
- **XGBoost:** AUC 0.923
- **Decision Tree:** AUC 0.875
- **Logistic Regression:** AUC 0.829
- **Takeaway:** AUC reflects alarm ranking capability → higher = better filtering

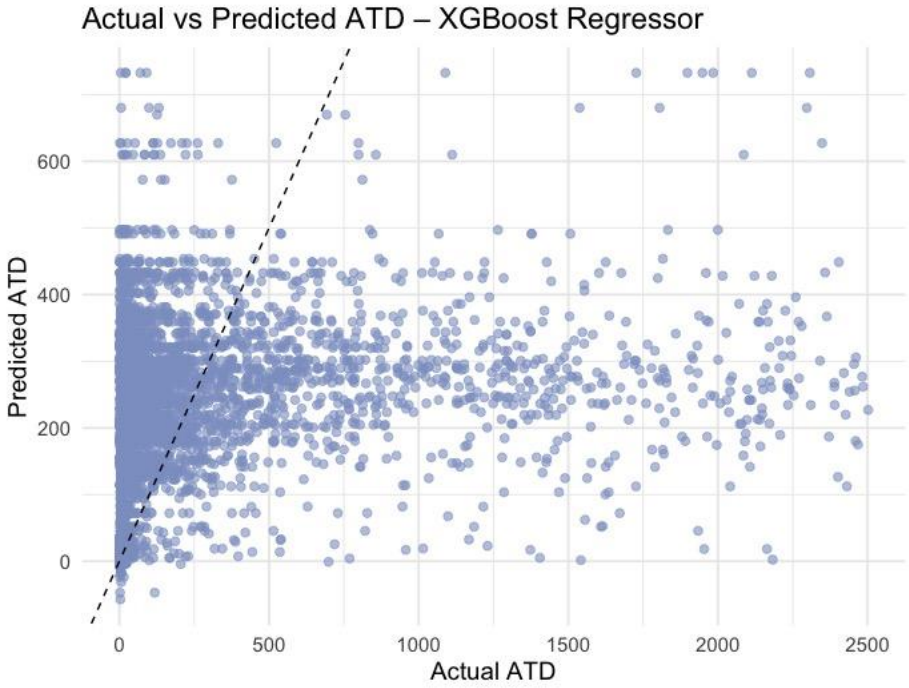


ATD Prediction – Model Comparison

- **Models:** Linear Regression vs XGBoost Regressor

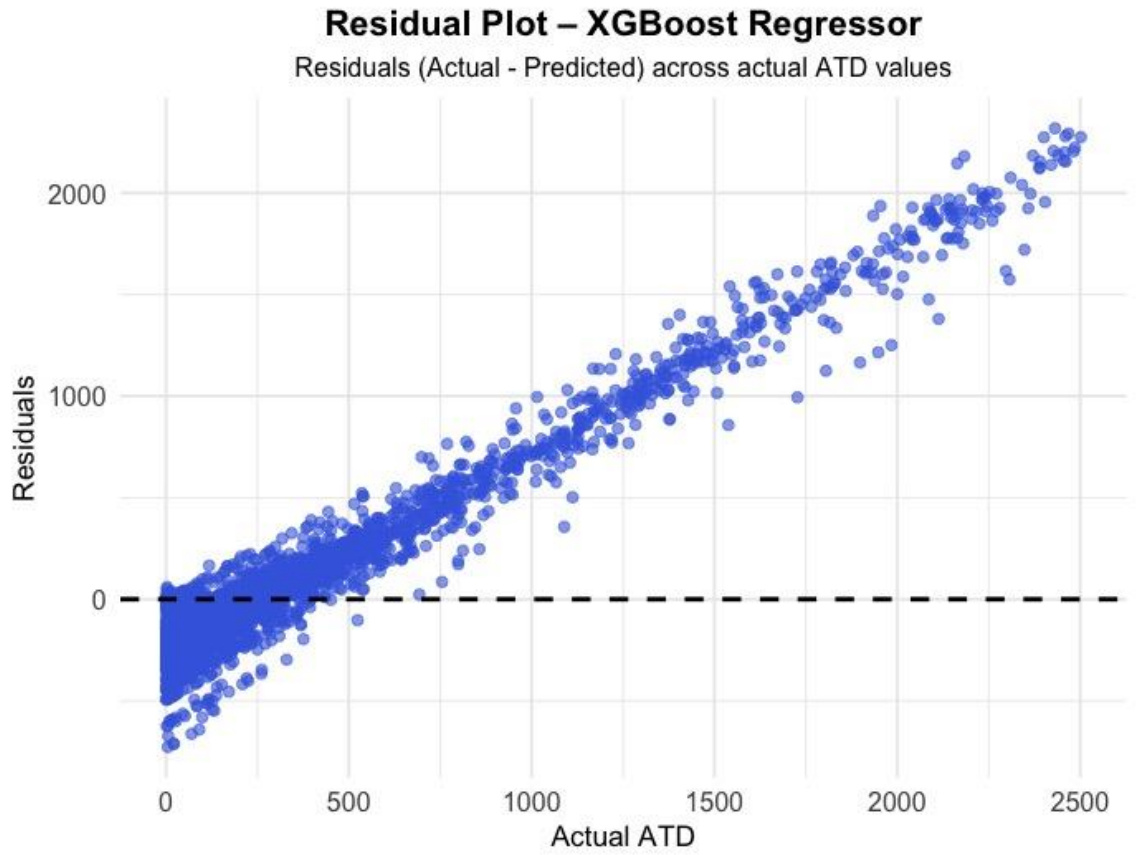
- **Key Results:**
- XGBoost had lower RMSE (404 vs. 407.5) and MAE (231.8 vs. 240)
- Captures complex patterns better





Residual Analysis – XGBoost ATD Prediction

- **Residuals mostly centered around zero**
- Slight overprediction at high durations
- **Conclusion:** Stable predictions → confidence in long-duration alarm triage



Insights Recap:

25% ALARMS ARE CHATTERING: SOURCE OF OPERATOR OVERLOAD

"OTHERS" CATEGORY: POOR CLARITY; RECLASSIFICATION NEEDED

XGBOOST CLASSIFIER: TOP PERFORMER (AUC = 0.91)

XGBOOST REGRESSOR: BEST FOR ATD PREDICTION

RESIDUAL PLOTS CONFIRM STABLE PREDICTION PERFORMANCE EVEN ON LONG ALARMS

Recommendations & Next Steps

- Deploy XGBoost Classifier in real-time systems
- Use ATD predictions to prioritize alarms
- Clean and reclassify “Others” category
- Develop a Shiny Dashboard
- Continuously refine models with live data

Conclusion

Summary

- CHB classification and ATD prediction enhance alarm management
- XGBoost models provide best performance
- Results enable proactive alarm filtering, improving decision-making

Final Thought

- Data-driven alarm intelligence reduces fatigue and boosts safety

Thank you!