From Incident Pipelines to Predictive Traffic Operations

By: Yancy Castro & Niki Nadezard

Date: June 16, 2026

Introduction

Traffic incident management depends on timing. When an incident happens on the freeway network, agencies need to understand what is happening, where it is happening, how long it may take to clear, and whether the pattern looks normal or unusual for that location and time.

The TIMS data pipeline work is a foundation for that kind of analysis. CHP, PeMS, and SWITRS each describe a different part of the incident picture. CHP gives a recent live-feed view, PeMS gives a historical incident-dispatch view, and SWITRS gives a structured collision-reporting view. Once these datasets are cleaned, geoprocessed, and stored in a repeatable way, they can support more than dashboard reporting.

The next opportunity is machine learning. The goal is not to replace operators or analysts. The goal is to give them better signals earlier: estimated clearance times, likely safety hotspots, and alerts when incident patterns do not look normal.

Motivation

Today, incident and collision analysis is mostly retrospective. The dashboard can show what happened yesterday, last month, or over the past five years. That is valuable, but the same data can also help answer forward-looking questions.

For example:

- If a crash is reported now, how long is it likely to take to clear?
- Which freeway segments are showing repeated safety problems?
- Is the current number or type of incidents unusual for this corridor, day, or hour?

These questions are hard to answer manually because the patterns are spread across multiple datasets and time windows. A machine learning layer could help summarize those patterns and surface them sooner.

The important first step is data readiness. Machine learning only becomes useful when the underlying data is consistent. The TIMS pipeline work moves the project in that direction by preserving raw inputs, standardizing cleaned outputs, assigning corridor and freeway segment geography, and creating repeatable datasets for analysis.

Proposed Machine Learning Direction

The recommendation is to explore three machine learning use cases first:

1. Clearance-time prediction.
2. Safety-hotspot detection.
3. Unusual-pattern alerts.

These use cases are practical because they build directly on fields already produced or documented in the TIMS pipeline work: incident type, category, severity, duration, goal time, corridor flags, freeway segment category, time of day, day of week, truck and hazmat flags, and collision attributes from SWITRS.

Clearance-Time Prediction

Clearance time is one of the most important operational metrics in the TIMS dashboard. A model could estimate how long an active or recently reported incident may take to clear based on similar historical incidents.

Useful inputs could include:

- Incident type and category.
- Location, corridor, and freeway segment.
- Time of day and day of week.
- Truck or hazmat involvement.
- Historical median and 90th percentile clearance time for similar incidents.
- Whether the incident falls in a peak period.

The first version does not need to be complicated. A baseline model could predict a range, such as short, medium, or long clearance, or estimate the expected duration in minutes. The output should be presented as a decision-support signal, not as a guarantee.

Safety-Hotspot Detection

The PeMS and SWITRS datasets both support location-based analysis. PeMS captures historical incident activity, while SWITRS captures reported collisions with richer collision details. Together, they can help identify places where risk is consistently higher.

A safety-hotspot workflow could score freeway segments based on:

- Total incidents or collisions over time.
- Crash and hazard counts.
- Severe or fatal collisions.
- Truck, hazmat, wrong-way, or peak-period patterns.
- Clearance-time performance.
- Changes compared with prior months or quarters.

The goal would be to highlight segments that may need closer review. This could help analysts prioritize where to look first instead of scanning every corridor manually.

Unusual-Pattern Alerts

The CHP live feed makes it possible to compare current activity against historical expectations. Once recent incidents are cleaned and geoprocessed, a model or statistical alerting process could flag unusual patterns.

Examples include:

- More incidents than expected on a corridor during a specific time window.
- A sudden increase in long-duration incidents.
- Repeated crashes near the same segment over a short period.
- A pattern of truck or hazmat-related incidents that is higher than normal.

This does not have to start as a complex model. A strong first version could use rolling averages, thresholds, and simple anomaly detection. More advanced models can come later after the team validates which alerts are actually useful.

Methodology

The recommended approach is to start small and build trust in the outputs.

First, create a training dataset from the curated TIMS outputs. For PeMS, this would use `incidents_cleaned` and `incidents_cleaned_geoprocessed`. For CHP, it would use the cleaned Recent Incidents outputs. For SWITRS, it would use `geoprocessed_incidents.csv` or a future curated equivalent.

Second, define clear target variables. For clearance-time prediction, the target could be `duration_minutes` or a duration category. For hotspot detection, the target could be a segment-level risk score. For unusual-pattern alerts, the target could be whether a corridor-hour combination is outside normal historical behavior.

Third, build baseline models before moving to more complex methods. Simple regression, classification, clustering, or threshold-based anomaly detection may be enough for the first phase. The goal should be usefulness and explainability, not model complexity.

Fourth, validate results with staff. A model that performs well statistically is not enough. The outputs should make sense to the people who understand incident operations, corridor context, and data limitations.

Responsible Use

Machine learning should be treated as an operational aid, not an automated decision-maker. The data has known limits: sources update on different schedules, live feeds can change, and collision records may be corrected retroactively.

Because of that, the first ML outputs should include confidence ranges, clear labels, and plain-language explanations. Staff should be able to see why a location or incident was flagged and what data contributed to the result.

The system should also track false positives and missed events. That feedback is what makes the model better over time.

Next Steps

The recommended next step is a small proof of concept:

1. Build a clean modeling table from PeMS, CHP, and SWITRS curated outputs.
2. Start with one clearance-time prediction model using PeMS historical incidents.
3. Create one segment-level hotspot score using PeMS and SWITRS geography.
4. Test simple unusual-pattern alerts using CHP Recent Incidents.
5. Review results with MTC staff before expanding the model.

The TIMS pipelines already create the structure needed for this work. The next phase is to turn clean, repeatable incident data into predictive signals that help staff focus attention earlier and make better use of the information they already collect.