Submission

# **Project Proposal: Optimizing EMT Station Placement on Pennsylvania Highways**

### **Decision Question**

The primary goal of this project is to determine the optimal locations for temporary Emergency Medical Technician (EMT) stations along Pennsylvania highways for each quarter of the year. The central decision question is: **Where should temporary EMT stations be located each quarter to minimize the weighted response distance to predicted crash locations, while adhering to budget and resource capacity constraints?** This analysis will also address how many stations to activate, the specific highway sections each should cover, and how to adapt placements to account for seasonal variations in crash patterns.

### **Data Sources and Preparation**

This analysis will be supported by a robust set of publicly available data. The primary dataset is the **PennDOT Crash Data (2019-2024)**, which provides historical records including the location, time, and severity of accidents. This will be combined with **Pennsylvania Highway Network Data** to define road segments and calculate distances, **NOAA Weather Data** to identify seasonal patterns, and **PennDOT Traffic Volume Data (AADT)** to correlate traffic flow with accident frequency. To prepare the data for analysis, highways will be segmented into discrete sections, and historical crash data will be aggregated by section and quarter. A distance matrix will then be computed between all potential station locations and each highway section.

### **Proposed Analytical Methods**

Two analytical methods will be developed and compared to address the decision question.

* **Predict-then-Optimize Framework** This two-stage approach first uses a machine learning model (e.g., Gradient Boosting/XGBoost) to predict the number and severity of crashes for each highway segment in a given quarter. The features for this model will include historical crash trends, weather patterns, and traffic volumes. In the second stage, the crash predictions will be used as inputs for a Mixed-Integer Programming (MIP) optimization model. This model will identify the set of station locations that minimizes the total predicted response distance, subject to budget, capacity, and maximum response time constraints. We will also explore a Smart Predict-then-Optimize (SPO) integration to train the prediction model with awareness of the downstream optimization goal.
* **Reinforcement Learning (RL) Approach** This method frames the problem as a sequential decision-making process. The state space will include the current quarter, active station locations, and remaining budget. The action space will consist of decisions to open or close stations at various locations. The model will be trained to maximize a reward function based on minimizing response distances, with penalties for exceeding the budget. A Deep Q-Network (DQN) will be trained on several years of historical data to learn a policy that dynamically adjusts station locations each quarter in response to changing conditions.

### **Expected Impact**

The successful implementation of this project is expected to provide PennDOT and emergency services with a data-driven framework for resource allocation. By optimizing station placements, the models could significantly reduce emergency response times, potentially leading to better patient outcomes and lives saved. The developed framework would also be scalable and adaptable for other emergency services and regions.

Tab 1

## **Project Proposal**

### **1. Decision Question**

**Primary Decision Question**: Where should temporary EMT stations be located along Pennsylvania highways each quarter to minimize weighted response distance to predicted crash locations, given budget and capacity constraints?

**Sub-questions**:

* How many stations should be activated each quarter?
* Which highway sections should each station cover?
* How should station locations adapt to seasonal crash patterns?

### **2. Data Requirements & Sources**

**Available Data**:

* **PennDOT Crash Data** (2019-2024): Historical crash records with location, time, severity, weather conditions
* **Highway Network Data**: Pennsylvania highway segments with entry/exit points, distances between nodes
* **Temporal Features**: Quarter, month, day of week, holidays, special events
* **Weather Data**: Historical weather patterns by quarter from NOAA
* **Traffic Volume Data**: AADT (Annual Average Daily Traffic) from PennDOT

**Data Processing Plan**:

* Segment highways into ~500-1000 sections based on mile markers
* Aggregate crash data by section and quarter
* Calculate distance matrix between potential EMT locations and highway sections

### **3. Analytical Methods from Course**

Based on the course topics, here's how to frame your two methods:

## **Method 1: Predict-then-Optimize Framework (Week 4 Topic)**

### **Prediction Stage**

* **Features**: Historical crashes, weather patterns, traffic volume, temporal indicators
* **Model**: Use gradient boosting (XGBoost) or deep learning to predict p\_j^t
* **Output**: Crash predictions for each section j in quarter t

### **Optimization Stage (MIP with Gurobi - Week 1)**

****Minimize: Σ\_t w\_t × Σ\_j Σ\_i p\_j^t × d\_ij × x\_ij^t

Subject to:

- Coverage: Σ\_i x\_ij^t ≤ 1, ∀j,t

- Activation: x\_ij^t ≤ y\_i^t, ∀i,j,t

- Capacity: Σ\_i y\_i^t ≤ K, ∀t

- Budget: Σ\_i Σ\_t C\_i × y\_i^t ≤ B

- Response time: x\_ij^t × d\_ij ≤ R\_max

### **Smart Predict-then-Optimize Integration (Week 4)**

* Use decision-focused learning to train prediction model aware of downstream optimization
* Implement SPO+ loss function that accounts for optimization objective during training

## **Method 2: Reinforcement Learning Approach (Week 3 Topic)**

Frame as sequential decision problem using course's RL framework:

### **State Space**

* Current quarter t
* Active station locations
* Historical crash patterns
* Budget remaining

### **Action Space**

* Open/close station at location i
* Assign sections to stations

### **Reward Function**

* Negative weighted response distance
* Penalty for budget overrun
* Bonus for high-risk section coverage

### **Solution Approach**

* **Q-Learning** for smaller problem instances
* **Deep Q-Network (DQN)** for full-scale problem
* Train on 3 years of data, validate on 1 year

### **Dynamic Programming Baseline (Week 3)**

* Use backward induction for comparison
* Solve finite-horizon problem quarter by quarter

## **Implementation Plan for Proposal**

### **For 1-Page Document (8%):**

**Paragraph 1 - Problem Statement** (3-4 sentences): Define the emergency response optimization problem, emphasizing the quarterly temporal component and the critical nature of minimizing response times to save lives.

**Paragraph 2 - Data Description** (3-4 sentences): Specify PennDOT crash database (2019-2024), highway network structure, and supplementary weather/traffic data. Mention data preprocessing to create section-quarter aggregations.

**Paragraph 3 - Method 1: Predict-then-Optimize** (4-5 sentences): Describe ML prediction model for crashes, followed by MIP optimization using Gurobi. Reference course topics from Weeks 1 and 4, including SPO+ integration.

**Paragraph 4 - Method 2: Reinforcement Learning** (4-5 sentences): Frame as MDP, describe state/action/reward design, and specify Q-learning approach from Week 3.

**Paragraph 5 - Expected Impact** (2-3 sentences): Quantify potential reduction in response times and lives saved. Mention scalability to other emergency services.

### **For 1-Page Summary Slide (2%):**

Create a visual layout with:

* **Title**: "Quarterly EMT Station Optimization for PA Highways"
* **Problem**: Visual of PA highway network with crash hotspots
* **Data Sources**: Icons/logos for PennDOT, NOAA, traffic data
* **Methods**: Two-column comparison of approaches with key equations
* **Timeline**: Gantt chart showing project phases
* **Expected Outcomes**: Metrics for success

## **What-If Analysis Opportunities (using OptiGuide)**

As mentioned in the syllabus, consider these scenarios:

1. Budget sensitivity analysis
2. Impact of climate change on seasonal patterns
3. Trade-offs between coverage and response time
4. Station capacity constraints
5. Multi-objective optimization (cost vs. coverage)