SciForce

Journal of Artificial Intelligence and Machine Learning

Journal homepage: www.sciforce.org
ISSN : 2995-2336

Open Access Research Article

Forecasting High-Demand Days in Mobile Crisis Response

Juan Quintero, M.S.*

Independent Researcher, United States

Abstract

Mobile Crisis Response (MCR) programs must provide rapid community-based response in spite of fluctuating call volume and limited crisis response staff resources. I tested whether feature engineering of the readily available operational records could be transformed into a useful next-day warning signal for high-demand days without the need to rely on client-level clinical, demographic, or extensive external datasets. Analysis was based on 1,068 full daily observations from March 1, 2023, to Jan. 31, 2026. A day with 10 or more calls was considered to be a high-demand day. Thirty-one predictors were created using the call date and daily call volume, which comprised rolling demand, short-term trend and momentum, prior peak activity, calendar factors and school-calendar factors. An XGBoost model was assessed using a chronological 80/20 split and 5-fold time-series validation, interpreting model contributions using SHAP. The final model had an AUC of 0.849 for a holdout sample. The probability threshold was taken as 0.25, for which recall was 0.90 and accuracy 0.66, with 18 of 20 peak days identified correctly. These results indicate that operational information, such as the volume and variety of calls, can serve as a valuable early warning indicator for behavioral health providers dealing with fluctuating call volumes and could help preparedness efforts by offering practical decision support tools. It is suggested that further research is needed on the usefulness of additional information, other prediction windows and longer-term operational effects.

Introduction

Mobile Crisis Response (MCR) programs are a key part of the behavioral health crisis system, providing immediate, community-based intervention for people with acute behavioral health needs. Despite limited responder numbers, travel demands, documentation requirements, and high daily call volume, these programs must maintain operational capacity[1,2]. One of the major operational challenges is predicting periods of high demand. Different volume of crisis services makes it challenging for the MCR supervisors to be able to anticipate staffing needs and to plan for coverage ahead of time [3]. Predicting demand in mental health helplines and other health services has been shown to be possible using forecasting methods, which can be used to optimize the utilization of resources and workload planning [4 , 5]. But not every community behavioral health provider has a data warehouse and access to a lot of real-time external data [6]. When the conditions are like this, it is not just a question of whether there is any demand that can be predicted, but whether the signals that can be gleaned from the data that an organization can easily access are useful.

Received date: August 25, 2026 Accepted date: October 20, 2026; Published date: October 23, 2026

***Corresponding Author:** Juan Quintero, Independent Researcher, USA Tel.: +1 561-485-2983, Email: juanquint839@gmail.com

Copyright: © 2026 Juan Quintero. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

This study thus focused on next-day high-demand classification with the intentionally restricted information base of daily MCR call counts, historical demand patterns, and calendar information. The final classifier did not contain any client-level information such as identification, diagnosis, demographics, or ZIP. The methodological contribution will not only be the use of XGBoost alone, but the transformation of simple operational data into a specific group of lagged, rolling, momentum, recurrence and calendar features that represent the state of demand over the last few days. The study set forth three goals[1] to identify if routinely available MCR operational data could be employed to differentiate between high demand and non-high demand days [2] to find which engineered temporal and calendar features, if any, were most important for predicting; and [3] to outline a repeatable process for organizations with early-stage analytics infrastructure that would allow them to utilize forecast information as additional support to their preparedness efforts instead of as an automatic

Materials and Methods

Study Design and Data

This study used an operational, retrospective design, with each calendar day as an analytical observation. The study period was from March 1, 2023, to January 31, 2026, and included data from the daily MCR calls. I a*n*alyzed daily MCR calls from March 1, 2023, to January 31, 2026. The first data set consisted of 1,127 daily observations, whereas after filtering and excluding the incomplete observations, the final data set used in the analysis had 1,068 complete daily observations. The study did not include any client-level identifiers, diagnoses, or demographics, nor did it include any information on ZIP codes.

Citation: Juan Quintero. (2026) Forecasting High-Demand Days in Mobile Crisis Response. Journal of Artificial intelligence and Machine Learning, 4(1), 1-6. https://doi.org/10.55124/jaim.v4i1.306

Data preprocessing followed a fixed, chronological pipeline. Dates were first parsed into a date-time format and the series was sorted chronologically to preserve temporal order. Rows at the beginning of the series that could not support complete lagged or rolling-window calculations (for example, a 28-day lag on the 10th day of the series) were removed from the analytical sample as structural exclusions resulting from the feature-construction process, not as missing or incomplete crisis events. Column names and binary calendar indicators, including school-in-session and summer-break flags, were standardized at this stage so that all downstream features were computed from a single, consistently formatted daily table.

Outcome Definition

The main outcome was a binary indicator for high demand the following day. The days in which there were ≥10 MCR calls were considered peak days, and days with <10 calls were considered non-peak days. The 10-call level was chosen as a threshold because the number of calls was at a point where routine staffing capacity might become progressively more stretched throughout the study period. The final dataset contained 135 peak days (12.6%) and 933 non-peak days (87.4%).

Table 1. Summary of study design, outcome definition, predictors, and modeling strategy

Component	Description
Study design	Retrospective operational forecasting study
Analytical unit	One calendar day
Study period	March 1, 2023–January 31, 2026
Initial observations	1,127 daily records
Final analytical sample	1,068 complete daily observations
Outcome	Next-day peak-demand classification
Peak-day definition	≥10 MCR calls
Peak days	135 (12.6%)
Non-peak days	933 (87.4%)
Predictor groups	Lagged demand, rolling averages, momentum, peak history, calendar and school-calendar features
Final model	XGBoost classifier
Class-imbalance strategy	Positive-class weight = 6.5
Validation	Chronological 80/20 holdout + five-fold TimeSeriesSplit
Model interpretation	SHAP

Feature Engineering from Routine Operational Data

The feature engineering approach was based on the information that was available prior to the day of prediction. The starting point was a chronologically ordered daily series with a date and total MCR call count.

The analysis in this series generated lagged demand, rolling demand, recent trend and momentum, and prior peak-history variables. Calendar conditions were added since they can be predicted and are not dependent on the client level.

Some rows in the time series that did not have the necessary historical information for lagged or rolling calculations were excluded from the time series as structural missing observations, not as missing crisis events. Historical peak variables had an explicit one-day shift prior to the rolling window, in order to avoid the day being part of its own predictors. The same principle goes for all demand-derived features only data completed before the forecast date should be used. This time separation is necessary for replication and to prevent target leakage.

This framework is not a transportable trained model, but rather a methodological approach; agencies using it should build features based on their own historical data and retrain and validate the trained model locally.

Model Development and Hyperparameter Optimization

The analysis started with the prediction of exact daily call count via linear regression, Lasso regression and XGBoost regression. Given that the operational question was whether the next day would be a more special day than the day before, the final approach was based on binary classification of peak versus non-peak days. Finally, the model for the classification task was an XGBoost classifier that was selected and tuned for the peak-day classification task [7].

Class imbalance was handled by a positive-class weight of 6.5, due to the fact that only 12.6% of the observations were peak days. The final hyperparameters 100 trees, maximum depth 2, learning rate 0.03, subsample 0.65, feature subsample 0.65, gamma 1, and minimum child weight 1 — were selected via randomized search across 40 parameter combinations evaluated over 5 temporal folds, for 200 total candidate model fits. The relatively shallow tree depth and conservative learning rate were selected, among the candidates evaluated, to limit overfitting given the small number of positive (peak-day) examples.

Temporal Validation and Performance Evaluation

The observations were split into an 80/20 holdout split by time, such that the earliest observations were used for the development of the model, and the most recent observations were used for final evaluation. Five-fold TimeSeries Split cross-validation was used for model development, and 200 candidate model fits were evaluated using randomized hyperparameter search.

Area under the receiver operating characteristic curve (AUC), recall, precision and accuracy were used to assess performance. Since the cost of issuing an "unnecessary" preparation alert was deemed to be lower than losing a genuine peak day, probability thresholds for candidates were explored, and a threshold of 0.25 was chosen as the high-recall operating point.

Model Interpretation and Operational Use

Global feature importance and the direction of individual feature contributions8 were explored through model predictions and interpretation using SHapley Additive exPlanations (SHAP). The prediction that was calculated was meant as an "early warning" signal to prepare staff for operations, not as a command to staff. It was intended to be used in conjunction with staff availability, planned coverage, local factors and the judgement of the supervising person.

Citation: Juan Quintero. (2026) Forecasting High-Demand Days in Mobile Crisis Response. Journal of Artificial intelligence and Machine Learning, 4(1), 1-6. https://doi.org/10.55124/jaim.v4i1.306

Table 2. Feature Engineering Reproduction Methods			
Feature family	**Examples**	**Purpose**	**Reproduction rule**
Recent call history	Lags 1–5, 7, 14, and 28 days	Captures short-term persistence and weekly/monthly recurrence	Use the number of calls received 1, 2, 3, 4, 5, 7, 14, or 28 days before the day being predicted.
Rolling demand	7-day and 30-day averages	Represents sustained recent demand relative to a longer baseline	Calculate the average number of calls during the previous 7 or 30 completed days.
Trend and momentum	3-/5-day averages, momentum, call changes, consecutive increases	Represents acceleration or sustained growth in recent demand	Compare recent call volume with earlier days to identify whether demand is increasing, decreasing, or remaining stable. Count consecutive days in which call volume increased.
Prior peak activity	Peak occurrence in prior 7, 14, and 30 days	Captures recent concentration of high-demand days	Mark whether at least one day with 10 or more calls occurred during the previous 7, 14, or 30 days.
Calendar context	Day of week, weekend status, month	Represents recurring calendar structure	Use the weekday, weekend indicator, and month of the day being predicted.
School calendar	School in session, summer break	Represents known community-calendar conditions	Mark whether the prediction date occurs while school is in session or during summer break.

Results

Analytical Sample and Peak-Day Distribution

After data preparation and completeness filtering, the final analytical dataset comprised 1,068 daily observations. Of these, 135 days met the specified peak-demand criterion (10 MCR calls or more): 12.6% were classified as peak demand and 87.4% were classified as non-peak. This distribution shows a significant class imbalance between the two outcome classes and serves as the operational context for assessing the model's accuracy in identifying relatively rare days with high demand.

Baseline Comparison and Cross-Validation Stability

To justify the final modeling choice, the tuned weighted XGBoost classifier was compared against simpler baselines developed earlier in the analysis (Table 3). Exact-count regression approaches (linear, Lasso, and XGBoost regression) produced modest predictive accuracy (R^2 between 0.29 and 0.34) and did not directly answer the operational question of whether a day would cross the peak threshold. Reframing the task as binary classification improved operational relevance: an unweighted XGBoost classifier achieved a test AUC of 0.846 but a recall of only 0.50 at the 0.25 threshold, missing roughly half of all peak days. Applying a positive-class weight of 6.5 to address class imbalance raised test AUC to 0.860 and recall to 0.90 at the same threshold, which is the configuration retained for the final model.

Model stability across time was assessed using five-fold TimeSeriesSplit cross-validation, in which each fold trained on earlier dates and evaluated on later dates. Fold AUC values were 0.646, 0.792, 0.734, 0.800, and 0.829, for a mean cross-validation AUC of 0.760. This mean is lower than the final chronological holdout AUC of 0.849. I interpret this gap as evidence that predictive difficulty varied across different periods of the development data some stretches of the time series were harder to forecast than others rather than as evidence that the holdout result is an overestimate specific to that period. This fold-level variability is itself useful evidence of real, but time-varying, signal, and it argues for ongoing performance monitoring after deployment rather than treating any single evaluation period, including the holdout, as a permanent benchmark of future performance.

Table 3. Comparative evaluation of candidate models, from exact-count regression to the final weighted classifier			
Model	**Outcome**	**Result**	**Role in selection**
Linear regression	Exact daily calls	MAE 1.95; RMSE 2.55; R^2 0.29	Initial parametric benchmark
Lasso regression	Exact daily calls	MAE 1.91; RMSE 2.48; R^2 0.33	Best regularized linear result
XGBoost regressor	Exact daily calls	MAE 1.87; RMSE 2.46; R^2 0.34	Improved nonlinear count prediction
Unweighted XGBoost classifier	Peak10 day	Test AUC 0.846; recall 0.50 at threshold 0.25	Higher specificity, fewer alerts
Weighted XGBoost classifier	Peak10 day	Test AUC 0.860; recall 0.90 at threshold 0.25	Prioritized rare peak detection
Tuned final XGBoost classifier	Peak10 day	Holdout AUC 0.849 at threshold 0.25	Selected, reproducible model

Holdout Model Performance

The final XGBoost classifier was able to predict a meaningful difference between peak and non-peak days, with an AUC value of 0.849 on the holdout evaluation. The model's recall and accuracy were evaluated at a selected probability threshold of 0.25; the model achieved a recall of 90% and accuracy of 66%. It correctly identified 18 of the 20 observed peak days in the holdout period and missed 2 peak days. The model also produced 71 false-positive alerts, reflecting that it was tuned to be as "recall-focused" as possible when trying to detect any potential demand surge.

Citation: Juan Quintero. (2026) Forecasting High-Demand Days in Mobile Crisis Response. Journal of Artificial intelligence and Machine Learning, 4(1), 1-6. https://doi.org/10.55124/jaim.v4i1.306

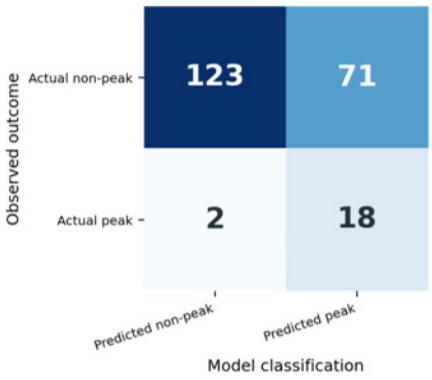

Figure 1: Holdout classification outcomes.

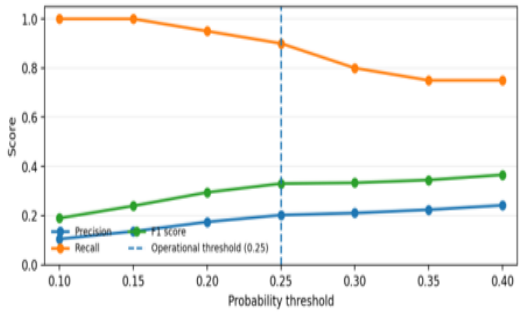

Figure 2: Precision–recall trade-off across candidate probability thresholds.

Probability Threshold Analysis

A clear trade-off was found between recall and the number of days flagged for preparation across candidate probability thresholds (Table 4). Recall was 1.00 at a threshold of 0.10 and predicted 191 peak days an operationally excessive volume of alerts. With the threshold increased to 0.20, recall was 0.95 and 109 peaks were predicted, whereas a threshold of 0.25 gave recall of 0.90 and 89 predicted peaks. Increased thresholds resulted in fewer alerts and more missed peak days: at 0.30, recall fell to 0.80 (four peak days missed), and at 0.40, recall fell to 0.75 (five peak days missed). The threshold of 0.25 was thus kept as the selected high-recall operating point.

Table 4. Threshold trade-off on the chronological holdout period

Threshold	Recall	Precision	Predicted peaks	Operational reading
0.10	1.00	0.105	191	Nearly every day flagged; operationally excessive
0.20	0.95	0.174	109	Maximum sensitivity with very frequent alerts
0.25	0.90	0.202	89	Selected high-recall operating point
0.30	0.80	0.211	76	Fewer alerts; four peak days missed
0.40	0.75	0.242	62	Higher precision; five peak days missed

The 71 false-positive alerts produced at the selected threshold reflect a deliberate, documented operating choice rather than an uncontrolled model weakness. In this operational context, a preparation alert such as a supervisor reviewing on-call coverage, confirming vehicle availability, or checking documentation backlog is a low-cost action compared with an unanticipated surge that catches a program understaffed. Given that asymmetry, prioritizing recall over precision was an intentional design decision, made explicit through the threshold analysis above, rather than an incidental property of the final model. Consistent with this framing, the forecast is intended strictly as an early warning signal to inform supervisory planning and should not be used as an automated staffing command or substitute for professional judgment.

SHAP-Based Model Interpretation

The largest contribution to the model output was the day of the week. Sustained recent demand was also a strong contributor, as generally higher 7- and 30-day rolling averages increased the predicted peak probability, with the 30-day average showing a wide range of impacts. Conditions during school-in-session also helped model predictions, while conditions during the summer break tended to lower the predicted risk. Recent acceleration also brought some warning cues positive momentum from the past 3 days, positive comparison to previous days, increased prior-day volume, and positive multi-day call changes were generally moving predictions up toward the most extreme class. Beyond the rolling averages, weekly and monthly recurrence signals (such as 7-day, 14-day, and 28-day lags and recent peaks in the history signal) played a part. These SHAP values describe how the trained model used the available predictors within this dataset; they do not, on their own, establish that a given calendar condition or prior call pattern causes a surge in demand. Several of the engineered predictors are correlated with one another for example, overlapping lag and rolling-average windows so a feature's apparent importance may be partly shared with, or redistributed across, related variables rather than attributable to that single feature alone.

Citation: Juan Quintero. (2026) Forecasting High-Demand Days in Mobile Crisis Response. Journal of Artificial intelligence and Machine Learning, 4(1), 1-6. https://doi.org/10.55124/jaim.v4i1.306

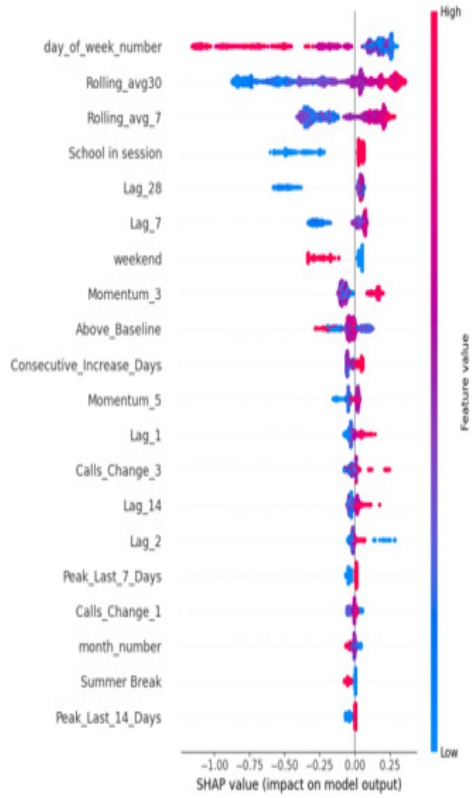

Figure 3: SHAP feature-contribution summary for the final peak-day classifier.

Discussion

Principal Finding and Methodological Contribution

This study shows that, with a relatively small operational information base, next-day high-usage periods in MCR operations can be meaningfully discriminated. The final model achieved a holdout AUC of 0.849 and identified 18 of 20 peak days at the chosen high-recall threshold. These results should not be taken to mean that peak demand is completely predictable. Instead, they show that there is enough information to decrease uncertainty about the next day, arising from recurring temporal structure and recent demand dynamics.

The more general contribution is the feature-engineering framework. The raw data needed for the core approach is fairly straightforward: date, daily call volume, and known calendar conditions. This basic information was enriched using lagged values, rolling summaries, momentum measures, recent peak history and school-calendar features. This is an easy-to-implement step to get started with predictive analytics for providers who do not have a well-developed external-data integration in place, and it does not require clinical, demographic, or geographic variables. The model explanation results further support this interpretation. Important signals were not limited to the day immediately before, but also included sustained rolling demand, weekly and monthly recurrence, calendar structure, school-calendar conditions and recent acceleration. High-demand days are more likely to be perceived as a string of recent pressure and recurring temporal context rather than as one-day events.

Operational Interpretation and Program Context

During the study period, the model was not implemented as an automated staffing system. Rather, the forecasting effort was converted into reporting suggestions and operational readiness recommendations that could be considered alongside staff availability, scheduled coverage, known community events, documentation workload, and supervisory judgment. This is important because the model metrics are retrospective, indicating predictive performance, not the causal effect of deployment.

Forecasting was part of a larger effort to modernize data and improve operations. The program achieved a 78% decrease in failed calls from 2023 to 2026, and the percentage of calls categorized as the fastest response classification (Level 1) rose from around 82% to 94.19%. The patterns presented here are program outcomes and not effects attributable to the forecasting model itself. Implementation of the electronic health record, enhanced data capture, staffing and management, training, call routing, reporting, and other operational changes may all have contributed. A more prospective or interrupted time-series design would be necessary to separate the effect of forecast-informed planning.

Despite this implementation history, a realistic path for adoption is demonstrated. Organizations do not have to start with automated staffing decisions. Forecast outputs can be added to reports and/or supervisory planning as a first step before more formal integration, to assess whether the alerts are understandable, timely and actionable.

Generalizability

The broad feature families used in this study recent call history, rolling demand, trend and momentum, prior peak activity, and calendar structure are plausible candidates for transfer to other mobile crisis or 988-connected providers, since most programs can produce comparable daily call counts and calendar information. However, the specific feature weights, SHAP patterns, and selected threshold reported here are specific to this provider's service area, staffing model, referral network, and school calendar. Transferring this approach to a new setting requires local retraining on that provider's own historical data, a locally appropriate peak-day definition, and independent prospective validation, rather than direct reuse of the trained model reported in this study.

Limitations

There are some limitations to keep in mind when interpreting these results. A primary concern is that only 135 peak days were analyzed, with 20 in the final holdout; recall estimates may therefore differ substantially even with a small change in the number of correctly classified examples. Another consideration is that the chosen high-recall threshold correctly identified 4 out of 5 peak days but resulted in many false positives; this can still work well in cases where taking action when an alert is triggered is not expensive, such as a coverage review, as opposed to hiring staff. This study included only one provider, and applicability may be limited in other settings due to the unique characteristics of local service areas, referral networks, staffing and school calendars.

Further, performance varied across time-based validation periods, meaning that demand relationships are not fixed and should be monitored regularly for change. For some engineered fields included in the model, values were computed in advance,

Citation: Juan Quintero. (2026) Forecasting High-Demand Days in Mobile Crisis Response. Journal of Artificial intelligence and Machine Learning, 4(1), 1-6. https://doi.org/10.55124/jaim.v4i1.306

and a replication should reproduce each of these fields as closely as possible from the raw daily counts. However, there is a chance of differences in calculation logic, which would not invalidate the results, but model accuracy might vary.

Conclusion

This study shows that a community behavioral health (CBH) provider can interpret and transform available call-volume and calendar information routinely into a next-day warning model for high-demand days in the MCR. The final XGBoost classifier had an AUC of 0.849 on the holdout set and identified 90% of peak days at the selected high-recall threshold, though with low precision and a high false-positive alarm rate. The real innovation here, however, is not an automated staffing rule, but a repeatable process of feature engineering and validation to derive meaningful temporal information from a comparatively low-tech operational dataset.

This approach is a good starting point for MCR programs with limited analytical infrastructure: develop dependable daily data, add history and calendar features, maintain temporal validation, understand model behavior, and start with forecasting as a supervisory decision-support tool. Extending this framework to other providers, examining the value of richer external information, and assessing the effects of prospective planning on crisis-response operations are important areas for future study.

Acknowledgments

The author thanks the behavioral health organization for the opportunity to conduct this research and for authorizing the use of anonymized operational data. This work represents an important step toward identifying opportunities to improve service delivery and operational practices across behavioral health providers

Conflicts of Interest

The author declares no conflict of interest.

References

1. Mauri, A. I., Rouhani, S., & Purtle, J. (2025). Characterizing crisis services offered by certified community behavioral health clinics: Results from a national survey. Psychiatric Services, 76(1), 13–21.
2. Peters, R. L., Fine, J., Newton, H., Dalack, G. W., Watson, A., & Lich, K. H. (2025). Mobile crisis effectiveness: A systematic review and associated functions and forms framework. BMC Health Services Research, 26(1), 60.
3. Fix, R. L., Bandara, S., Fallin, M. D., & Barry, C. L. (2023). Creating comprehensive crisis response systems: An opportunity to build on the promise of 988. Community Mental Health Journal, 59(2), 205–208.
4. de Boer, T. R., Mérelle, S., Bhulai, S., Gilissen, R., & van der Mei, R. (2023). Forecasting call and chat volumes at online helplines for mental health. BMC Public Health, 23(1), 984.
5. Vollmer, M. A., Glampson, B., Mellan, T., Mishra, S., Mercuri, L., Costello, C., et al. (2021). A unified machine learning approach to time series forecasting applied to demand at emergency departments. BMC Emergency Medicine, 21(1), 9.
6. Tello, M., Reich, E. S., Puckey, J., Maff, R., Garcia-Arce, A., Bhattacharya, B. S., & Feijoo, F. (2022). Machine learning based forecast for the prediction of inpatient bed demand. BMC Medical Informatics and Decision Making, 22(1), 55.
7. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. https://doi.org/10.1145/2939672.2939785
8. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30.

Citation: Juan Quintero. (2026) Forecasting High-Demand Days in Mobile Crisis Response. Journal of Artificial intelligence and Machine Learning, 4(1), 1-6. https://doi.org/10.55124/jaim.v4i1.306