

# Model Card for Milan Mobile Traffic Prediction

Version 1.0 | Date: June 13, 2025

## Model Details

- Model Name: Milan Mobile Traffic Prediction
- Version: 1.0
- Date: June 13, 2025
- Primary Use: Forecast short-term mobile network demand in 10-minute intervals for specific areas within Milan
- Dataset Source: Barlacchi, G., De Nadai, M., et al. (2015). A multi-source dataset of urban life in Milan and the Province of Trentino

## Model Performance

Target Variable	R <sup>2</sup> Score (Higher is Better)	RMSE (Lower is Better)
SMS Activity	0.9210	0.1420
Call Activity	0.9626	0.0738
Internet Traffic	0.9470	0.0923

The model demonstrates high accuracy across all targets, explaining over 92% of the variance. It is most effective at predicting call activity, which has the lowest prediction error.

## Data Preparation

### Initial Processing

- Timestamps were converted to a standard datetime format localized to 'Europe/Rome'.
- The dataset was filtered to focus on the central 2500 grid squares of Milan.
- Directional data (e.g., 'SMS-in'/'SMS-out') was merged into single activity features like 'SMS activity'.

### Feature Engineering

- Temporal Features: Created cyclical features for daily/weekly patterns and an is weekend flag.

- **Spatial Features:** Derived grid coordinates and a grid density feature to represent the activity level of each location.
- **Target Transformation:** Applied log transformations to target variables to normalize their distributions and improve model performance.

## Data Splitting

A time-based splitting strategy was used to ensure the model is evaluated on data it has not seen before, which mimics a real-world forecasting scenario.

- **Training & Testing Period (Nov 1 - Dec 13):** The data from the beginning of the dataset until December 13th was used for model training and hyperparameter tuning, following an 80/20 split for training and testing respectively.
- **Validation Period (Dec 14 - Dec 31):** The data from December 14th to December 31st was held out as a final validation set to evaluate the performance of the chosen model on completely unseen future data.

## About the Model

The selected model is designed for a specific purpose. Here are its core details and intended uses.

### Intended Use

- **Primary Users:** Network operators for dynamic resource allocation and urban planners studying city dynamics.
- **Scope:** Forecasting short-term mobile network demand in Milan.

### Out-of-Scope Uses

- Predicting traffic for individual users (privacy-sensitive).
- Making long-term (e.g., year-ahead) forecasts.
- Forecasting for cities other than Milan without retraining on relevant local data.

## Ethical Considerations & Limitations

### Temporal Bias

The model is trained on data from Nov/Dec 2013 and may not generalize to other seasons, years, or periods with significantly different mobility patterns.

## Geographic Bias

The model is specific to Milan. Its spatial features and learned patterns will not be applicable to other cities without complete retraining.

## Data Representation

The dataset only captures activity on the Telecom Italia network and does not include traffic from other carriers or non-cellular activity (e.g., Wi-Fi).

## Fairness Considerations

The model predicts aggregate traffic and does not use demographic data. However, socio-economic factors may implicitly influence network usage patterns across different grid squares.