Time-series forecasting of daily $PM_{2.5}$ air pollution levels using Temporal Fusion Transformer (TFT) model

Stefanela Stevanović
ss51676@student.uni-lj.si
Faculty of Computer and Information Science,
University of Ljubljana
Večna pot 113
SI-1000 Ljubljana, Slovenia

Andrija Stanišić
as86041@student.uni-lj.si
Faculty of Computer and Information Science,
University of Ljubljana
Večna pot 113
SI-1000 Ljubljana, Slovenia

ABSTRACT

Accurate forecasting of $PM_{2.5}$ concentrations is critical for urban air quality management. This study develops a Temporal Fusion Transformer (TFT) model for multi-city $PM_{2.5}$ forecasting, evaluating its performance on daily air quality data from Belgrade and Ljubljana (2022–2025). The model achieves average 5-day mean absolute errors of 10.56 μg/m^3 (Belgrade) and 11.33 μg/m^3 (Ljubljana). Feature importance analysis reveals that mean sea-level pressure dominates historical pattern recognition (encoder phase), while relative humidity and wind and temperature related variables guide current-condition interpretation (decoder phase).

KEYWORDS

$PM_{2.5}$ forecasting, Temporal Fusion Transformer (TFT), Air quality prediction, Multivariate time-series

1 INTRODUCTION

The goal of this project is to develop a forecasting model for daily $PM_{2.5}$ levels using a Temporal Fusion Transformer (TFT), a state-of-the-art deep learning architecture designed for interpretable time-series analysis. High concentrations of $PM_{2.5}$ are linked to respiratory diseases, cardiovascular problems, and reduced life expectancy. Thus, accurate $PM_{2.5}$ forecasting is crucial due to its significant impact on public health and environmental policy.

Traditional forecasting models often struggle to capture complex, nonlinear relationships between $PM_{2.5}$ levels and meteorological factors (e.g., temperature, humidity, wind speed). Additionally, they lack the ability to efficiently handle multi-time-series forecasting (e.g. predicting $PM_{2.5}$ for multiple locations simultaneously within a single model).

In this project, we leverage historical daily $PM_{2.5}$ data and meteorological variables to generate 5-day forecasts for two locations: Ljubljana (Bežigrad) and Belgrade (Stari Grad). We will explain the methodology behind implementing the TFT model for $PM_{2.5}$ time-series forecasting, evaluate its performance, and utilize its interpretability features to analyze the impact of meteorological factors on $PM_{2.5}$ concentrations.

2 METHODOLOGY

2.1 Dataset and Data Preparation

The dataset consists of historical daily $PM_{2.5}$ measurements from two air quality monitoring stations: Ljubljana Bežigrad (https://aqicn.org/city/ljubljana/) and Belgrade Stari Grad (https://aqicn.org/city/serbia/beograd-stari-grad/), covering the period from 1 January 2022 to 20 May 2025. Meteorological variables were obtained from Open-Meteo (https://open-meteo.com/) to supplement the air quality data.

Missing $PM_{2.5}$ values were addressed through forward-filling (propagating the last valid observation forward) and backward-filling for initial NaNs. Wind direction in degrees was transformed into cyclical components using sine and cosine representations.

The final feature set includes:

- Minimum and maximum daily temperature (°C at 2m height)
- Average wind speed (km/h)
- Maximum wind gusts (km/h)
- Average wind direction components ($\sin\theta$, $\cos\theta$)
- Average relative humidity (%)
- Average sea-level pressure (hPa)
- Average cloud cover fraction (0–1)
- Daily precipitation sum (mm)
- Shortwave radiation sum (MJ/m^2)
- Month (categorical: 1–12)
- Weekend indicator (binary: 1=weekend, 0=workday)

Additionally, to capture autoregressive patterns in $PM_{2.5}$ dynamics, all features, including the target variable, were lagged by one timestep (1 day), providing the TFT model with explicit short-term temporal dependencies.

Both datasets were split into training (1,036 days), validation (100 days), and test sets (100 most recent days) to maintain temporal ordering and prevent data leakage. This sequential partitioning ensured no future information contaminated the training process.

2.2 Models

We established a naive baseline predicting each day's $PM_{2.5}$ concentration as the value observed 7 days prior, before developing our primary forecasting system with PyTorch's Temporal Fusion Transformer (TFT). The TFT model was trained to predict $PM_{2.5}$ concentrations over a five-day horizon, as this timeframe represents a reasonable window where meteorological forecasts remain sufficiently accurate to serve as reliable inputs. The encoder analyzed historical patterns using a 60-day look-back period.

The model's hyperparameters were optimized using Optuna, a Bayesian optimization framework that efficiently explores parameter spaces. After 15 trials, the optimization process yielded the following optimal parameters:

- Batch size: 63

- Dropout rate: 0.2
- Hidden layer size: 32
- Hidden continuous size: 4
- Attention head size: 4
- Learning rate: 0.016

The learning rate was dynamically adjusted during training using a plateau-based reduction scheduler to improve convergence. Training incorporated an early stopping mechanism to prevent overfitting, halting optimization if the validation loss failed to improve by at least 1×10^{-4} over eight consecutive epochs. Model performance was evaluated using a sliding-window validation approach on the held-out validation set, where predictions were made for five-day sequences at daily intervals. The optimization process used quantile loss as the objective function. The model trained for 25 epochs before early stopping.

2.3 Evaluation

The final evaluation was performed on non-overlapping 5-day windows across the test set (totaling 20 windows). For each window, we computed the Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). These metrics were averaged across all windows to assess model consistency over time. We also computed the global RMSE and MAE by aggregating the predictions and errors across the entire test set.

The same procedure was applied to the baseline model. Both 5-day window-averaged and global metrics were calculated to ensure consistent comparison.

3 RESULTS

3.1 Model Performance

The model performance was evaluated through two complementary metrics: (1) windowed error statistics calculated across twenty 5-day rolling forecasts, and (2) global error measures aggregated over the entire test period. Tables 1 and 2 present these results for Ljubljana and Belgrade respectively, comparing the TFT model against the naive baseline.

Table 1: Model Performance for Ljubljana

	Baseline	TFT
Avg. Window RMSE	30.49 ± 21.13	13.68 ± 8.96
Global RMSE	36.80	16.26
Avg. Window MAE	26.63 ± 19.46	11.33 ± 7.76
Global MAE	26.63	11.33

Table 2: Model Performance for Belgrade

	Baseline	TFT
Avg. Window RMSE	27.51 ± 16.05	12.38 ± 4.83
Global RMSE	31.65	13.24
Avg. Window MAE	22.50 ± 13.69	10.56 ± 4.05
Global MAE	22.50	10.56

Test set $PM_{2.5}$ concentrations averaged $79.69\mu g/m^3$ in Belgrade and $57.48\mu g/m^3$ in Ljubljana, with similar variability ($\sigma \approx 35\ \mu g/m^3$) in both cities. The model achieved lower errors in Belgrade despite its higher pollution levels, suggesting more systematic patterns in its $PM_{2.5}$ dynamics. Performance remained consistent across both rolling windows and global metrics, with marginally better stability in Belgrade.

3.2 Feature Importance

The feature importance patterns, presented in Figures 1-2 on the final page, demonstrate how the TFT model allocated attention across different variables during processing.

During the encoder phase, mean sea-level pressure stood out as the most influential variable in both cities. This aligns with meteorology principles, as pressure systems govern atmospheric stability and pollution dispersion. The model also consistently relied on historical $PM_{2.5}$ levels, suggesting strong temporal persistence in particulate matter concentrations.

During the decoder phase, relative humidity played a consistently significant role in both locations. This likely reflects how moisture influences aerosol formation and particle growth, particularly during winter inversion episodes and spring transitional periods.

An interesting observation was Belgrade's comparatively stronger weighting of wind characteristics during decoding, possibly indicating subtle differences in how local topography or emission patterns interact with atmospheric transport. However, the overall similarity in feature importance between the two cities suggests that $PM_{2.5}$ forecasting in urban environments may rely on universal meteorological drivers, with only secondary variations based on local geography.

4 CONCLUSION

In this project, we demonstrated that the Temporal Fusion Transformer (TFT) model achieves promising results in forecasting $PM_{2.5}$ levels five days in advance, with a 5-day MAE of 11.33 for Ljubljana and 10.56 for Belgrade on the test set. However, it is important to note that these results may represent an optimistic estimation, as the model relies on meteorological variables assumed to be known in advance. In practice, inaccuracies in weather forecasts would also impact $PM_{2.5}$ predictions.

Interestingly, the model exhibits a lower error in Belgrade despite the city generally having higher $PM_{2.5}$ concentrations, with a similar standard deviation to Ljubljana in the test set. This suggests that the TFT may better capture the pollution patterns in Belgrade.

Furthermore, feature importance analysis revealed that mean sea level pressure was the most significant encoder feature, followed by the target $PM_{2.5}$ value. For decoder features, the key predictors varied between cities, though relative humidity, daily temperature, and wind patterns consistently played important roles. Notably, wind-related variables (such as maximum wind gusts and wind direction) ranked higher in Belgrade than in Ljubljana, indicating their stronger influence on $PM_{2.5}$ levels in the Serbian capital.

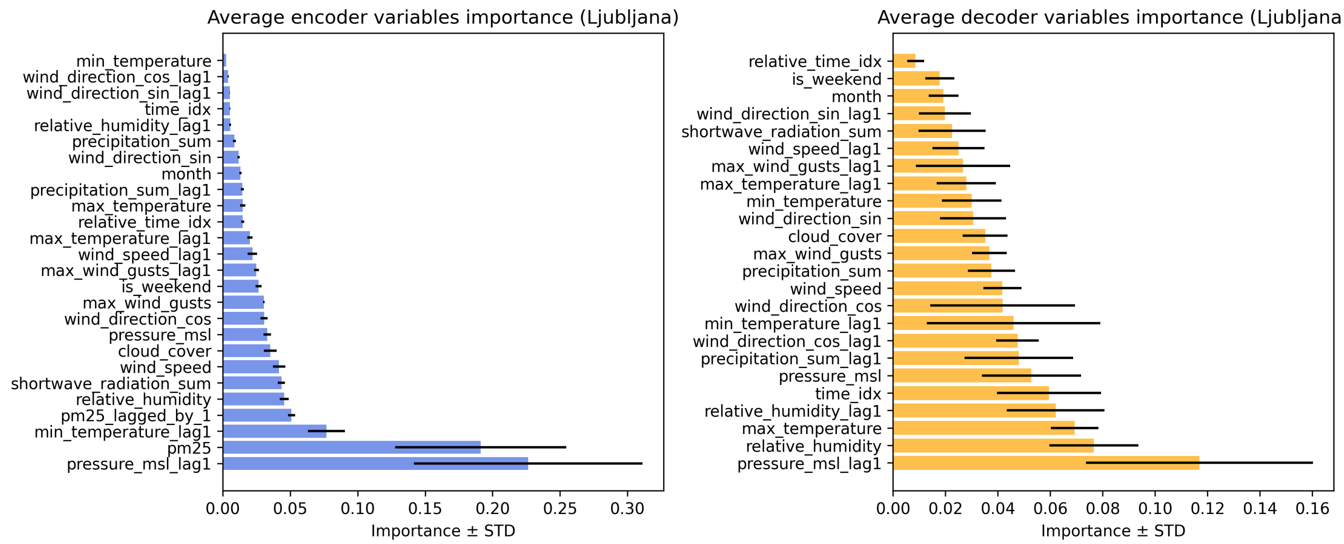

Figure 1: Variable importance in Ljubljana forecasts showing (left) encoding-phase attention to historical patterns, and (right) decoding-phase reliance on current conditions.

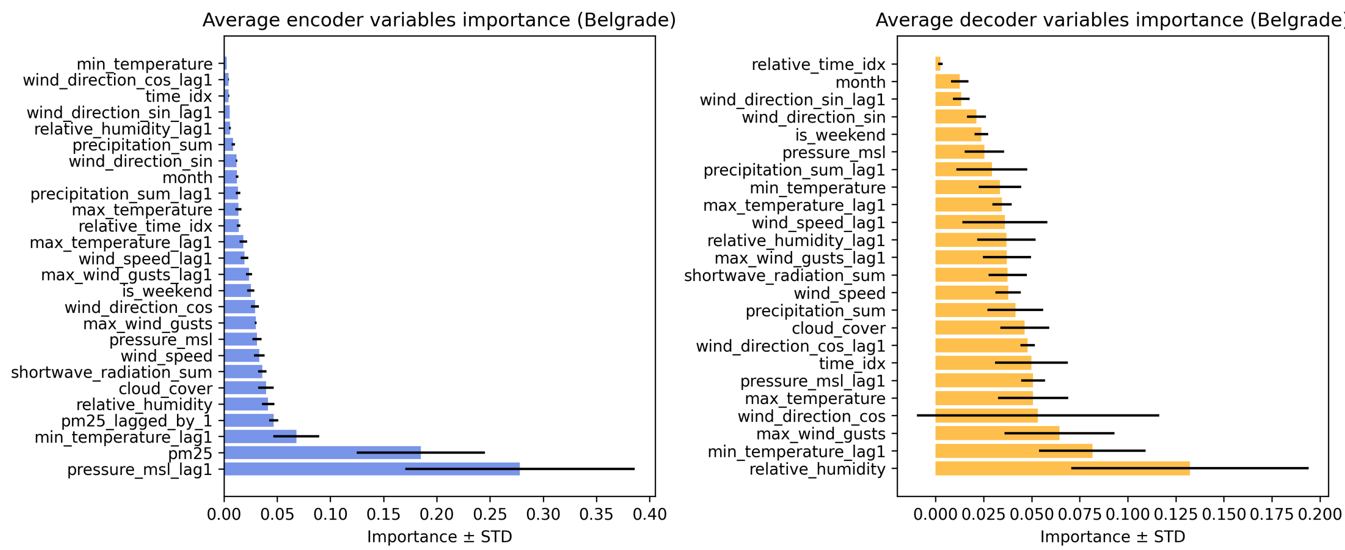

Figure 2: Variable importance in Belgrade forecasts showing (left) encoding-phase attention to historical patterns, and (right) decoding-phase reliance on current conditions.