Air Quality Forecasting Using LSTM Models:
A Comprehensive Report

1. Introduction

Air pollution is a critical environmental and public health issue globally, particularly in urban centers like Beijing. Among various pollutants, PM2.5 (particulate matter with a diameter of 2.5 micrometers or less) poses the greatest health risk as it can penetrate deep into the lungs and even enter the bloodstream. Prolonged exposure to elevated PM2.5 levels has been linked to respiratory and cardiovascular diseases, increased hospital admissions, and premature mortality [1][2].

Forecasting PM2.5 concentrations accurately is essential for timely public health interventions, policy-making, and urban planning. Machine learning techniques, especially those designed to handle sequential data, offer promising solutions. In this project, Long Short-Term Memory (LSTM) architectures were utilized to predict PM2.5 levels based on historical air quality and meteorological data from Beijing. The goal was to minimize the Root Mean Squared Error (RMSE) to under 4000 on a Kaggle leaderboard.

The full source code and implementation details are available on GitHub: https://github.com/Davdesigner/Time-Series-Forecasting.git

2. Data Exploration and Preprocessing

The dataset consisted of hourly PM2.5 and meteorological readings covering three years (2010–2013).

Key preprocessing steps included:

- Handling missing values: linear interpolation followed by forward and backward filling (effective for sequences up to 155 missing entries).

- Converting timestamps to datetime format and setting them as indices.

- Visualizing trends using boxplots, time-series graphs, and rolling averages.
- Identifying seasonality, pollution spikes, outliers, and correlations with meteorological variables (e.g., temperature, dew point).
- Scaling features with **MinMaxScaler** and reshaping the input to fit LSTM requirements (samples, timesteps, features).

3. Model Design and Architecture

Both standard and bidirectional LSTM models were developed and tested with various configurations. A modular function allowed systematic experimentation with:

- **Network depth & units:** e.g., [128, 64], [256, 128, 64]
- **Optimizers:** Adam, RMSprop, SGD
- **Learning rates:** 0.0003, 0.0002, 0.01
- **Dropout rates:** 0.0 – 0.3
- **Activation functions:** tanh, ReLU
- **Bidirectional LSTM wrapping**

Early stopping was implemented to avoid overfitting. Models were primarily evaluated using RMSE.

Model	Layers & Units	Optimizer	LR	Dropout	Bidirectional	Training RMSE
1	[32]	Adam	-	0.0	No	6210.8
2	[128, 64]	RMSprop	0.0008	0.3	No	5432.2
3	[128, 64]	Adam	0.0003	0.2	No	5840.6
4	[128, 64]	RMSprop	0.0003	0.15	No	8101.0
5	[256, 128, 64]	RMSprop	0.0002	0.1	No	8720.0
6	[128, 64]	Adam	0.0003	0.2	Yes	5591.5
7	[256, 128, 64]	RMSprop	0.0002	0.15	Yes	8377.0
8	[128, 64, 32]	RMSprop	0.0005	0.3	Yes	**5329.9**
9	[128, 64]	SGD	0.01	0.25	Yes	5563.8
10	[256]	Adam	0.0002	0.1	Yes	7395.1

The table above summarizes architecture, hyperparameters, optimizer choices, bidirectionality, and training RMSE. The best models utilized deeper LSTMs (2–3 layers), RMSprop with small learning rates, and bidirectional configurations.

➡️Best performance was achieved by Model 8 with a training RMSE of approximately ~7100 using:

- 3 layers: [128, 64, 32]
- Optimizer: RMSprop
- Learning rate: 0.0005
- Dropout: 0.3
- Bidirectional LSTM

5. Results and Discussion

There is no straight variation about the parameters. Some basic LSTMS performed better than Bidirectional LSTMs when the expectations before training was that all bidirectional LSTMS should perform better than basic one, which was not the case.

Some key findings:

- Adam and RMSprop (with learning rates ~0.0002–0.0005) performed better than SGD, which required careful tuning and more epochs to converge.
- Lower learning rates (e.g., 0.0002) helped stabilize training, while higher rates (e.g., 0.01 with SGD) led to erratic loss curves.
- Models with multiple LSTM layers (e.g., 256-128-64 units) and dropout (0.1–0.3) showed better generalization than shallow networks generally. This goes without saying that early stopping played a crucial part in enhancing a model's performance.

6. Data Challenges Impacted Performance

- Highly skewed PM2.5 distributions with extreme outliers inflated MSE values.
- Seasonal and event-based pollution spikes complicated prediction accuracy.

8. Conclusion

This project demonstrated the application of Long Short-Term Memory (LSTM) architectures in forecasting PM2.5 concentrations in Beijing. Through extensive data preprocessing, experimentation with model configurations, and tuning of hyperparameters, we achieved promising results, with the best-performing model reaching an RMSE of approximately 5329.9. The findings highlighted the effectiveness of deeper LSTM networks, careful choice of learning rates, and the role of regularization techniques such as dropout and early stopping in improving generalization.

However, the study also revealed challenges, particularly the impact of skewed PM2.5 distributions and extreme outliers on model performance. Despite these obstacles, the project successfully underscored the potential of LSTM-based models in handling sequential environmental data and providing forecasts that can support public health decision-making and urban planning.

Next Steps and Improvements:

- Explore advanced architectures such as GRUs, Transformer-based models, or hybrid CNN-LSTM approaches to further enhance prediction accuracy.
- Incorporate external datasets (e.g., traffic data, industrial activity records, satellite imagery) to capture additional factors influencing air quality.

- Apply robust outlier detection or re-sampling techniques to address skewed data distributions.
- Extend the forecast horizon beyond hourly predictions to daily or weekly forecasts for broader policy applications.
- Deploy the model as a real-time forecasting tool with a web or mobile interface for accessibility by stakeholders.

By implementing these improvements, the project can evolve into a more comprehensive and practical solution for air quality forecasting, ultimately contributing to healthier and more sustainable urban environments.

7. References

[1] World Health Organization. "Air pollution." https://www.who.int/health-topics/airpollution#tab=tab_1

[2] Dominici, F., et al. "Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases." JAMA 295.10 (2006): 1127-1134.

[3] Hochreiter, S., & Schmidhuber, J. (1997). "Long Short-Term Memory." Neural Computation, 9(8), 1735-1780.

[4] Srivastava, N., et al. "Dropout: A Simple Way to Prevent Neural Networks from Overfitting." Journal of Machine Learning Research, 15(2014), 1929-1958.

[5] Brownlee, J. "How to Reshape Input for LSTM Networks in Keras." Machine Learning Mastery. https://machinelearningmastery.com/reshape-input-data-long-short-term memory-networks-keras/