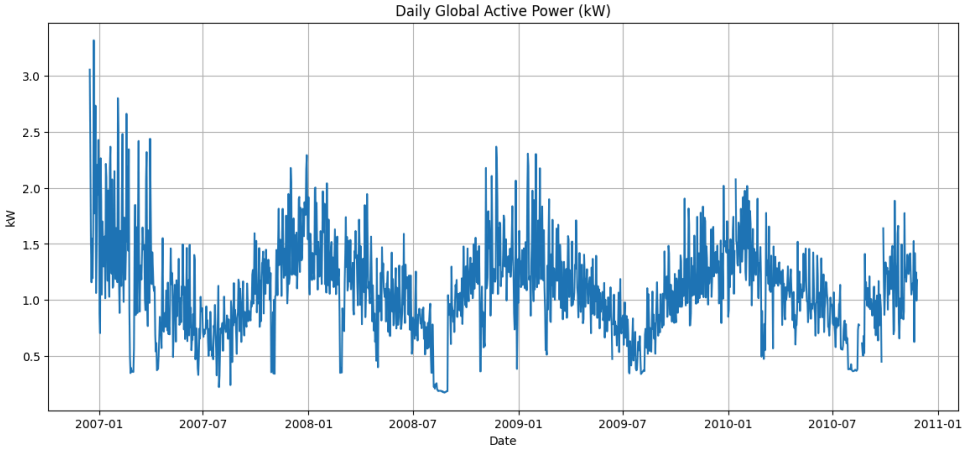
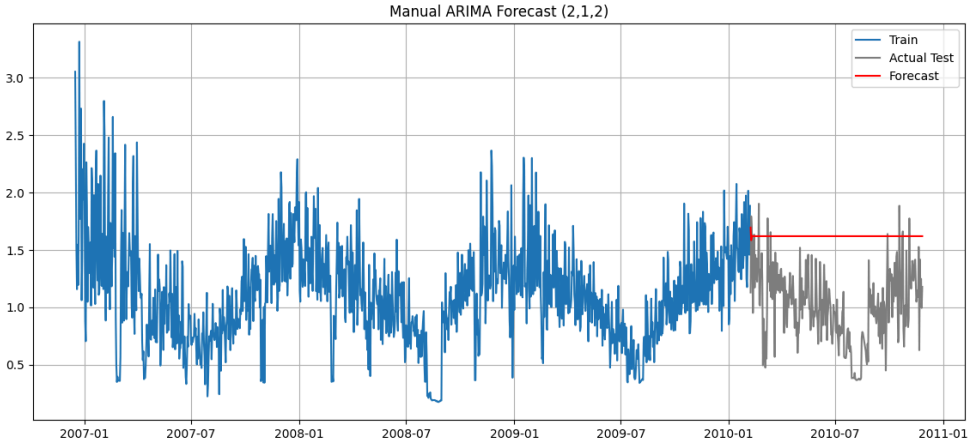
1. Energy consumption prediction
   Analysis using public datasets from:
   https://archive.ics.uci.edu/ml/machine-learning-databases/00235/household_power_consumption.zip
   Dataset preview:

	datetime	Global_active_power	Global_reactive_power	Voltage	Global_intensity	Sub_metering_1	Sub_metering_2	Sub_metering_3
0	2006-12-16 17:24:00	4.216	0.418	234.84	18.4	0.0	1.0	17.0
1	2006-12-16 17:25:00	5.360	0.436	233.63	23.0	0.0	1.0	16.0
2	2006-12-16 17:26:00	5.374	0.498	233.29	23.0	0.0	2.0	17.0
3	2006-12-16 17:27:00	5.388	0.502	233.74	23.0	0.0	1.0	17.0
4	2006-12-16 17:28:00	3.666	0.528	235.68	15.8	0.0	1.0	17.0

   Daily energy consumption plot:

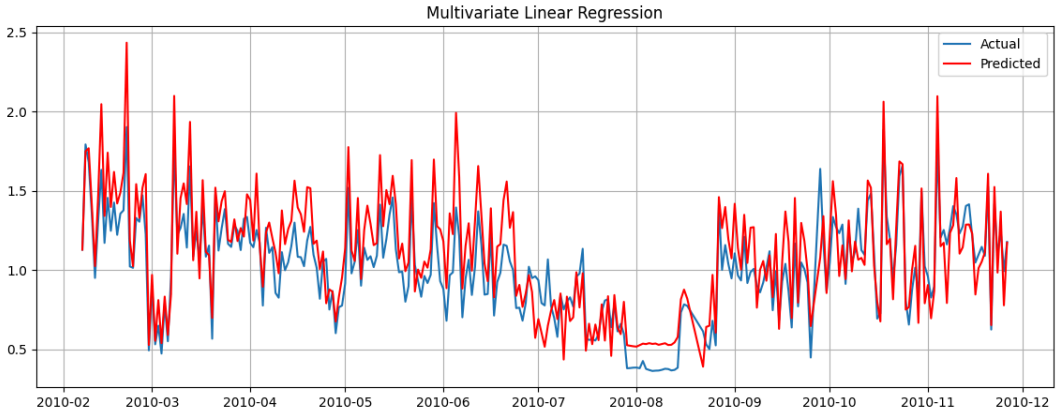


   Manual ARIMA forecast: Resample daily data.



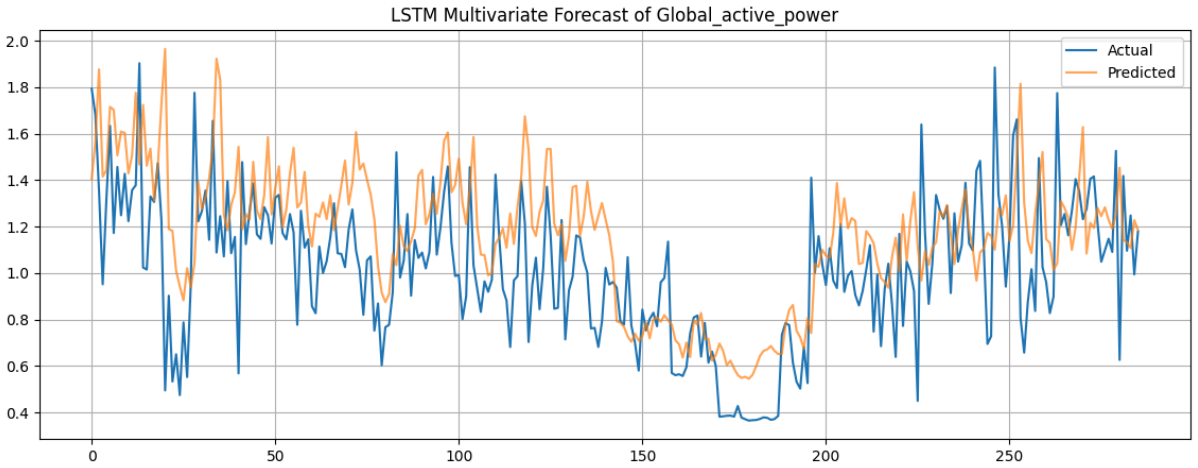
   ARIMA predictions are very flat. ARIMA follows the final level of the training data. The model fails to capture the variability of the test set.

Multivariate model: adds additional features such as Voltage and Sub_metering.



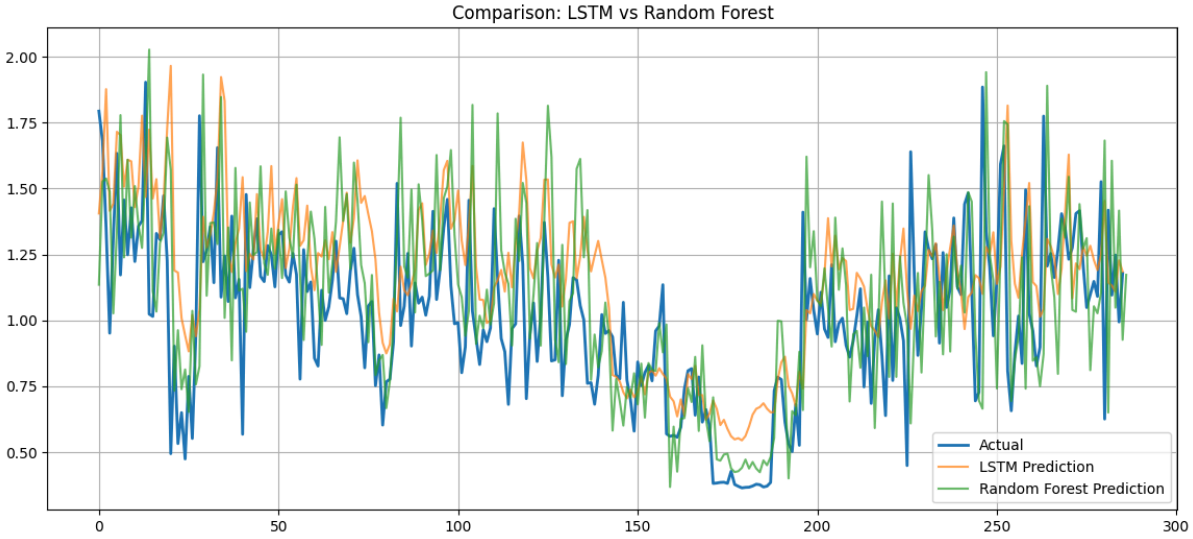
The model is able to follow general trends. Overfitting to extreme values. Systematic errors in certain periods. Lack of temporal context.

Multivariate LSTM to predict Global_active_power based on: Voltage, Global_reactive_power, Sub_metering_1, Sub_metering_2, Sub_metering_3.



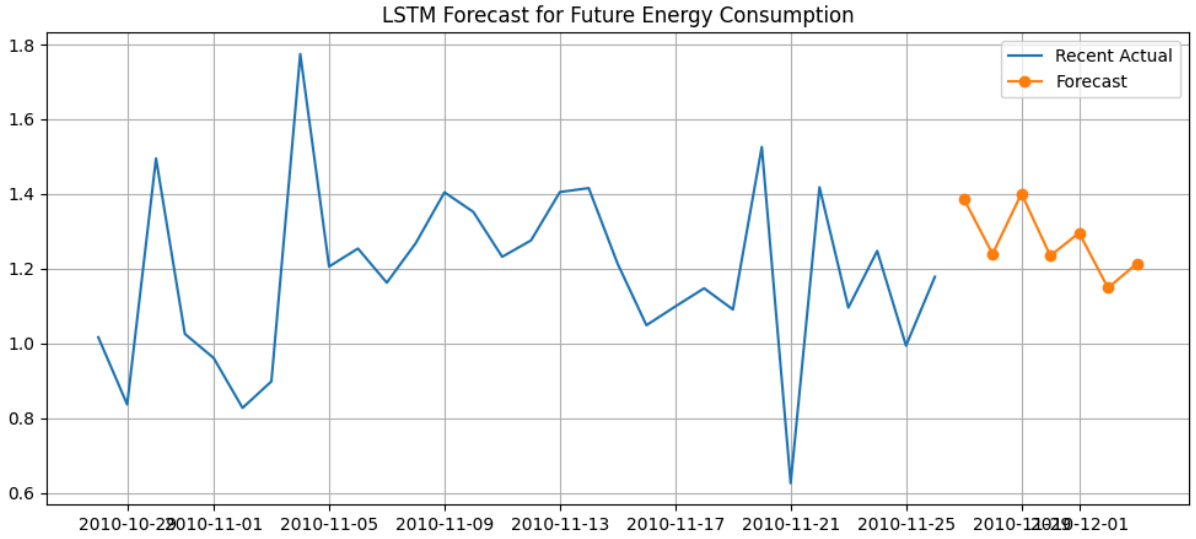
LSTM is able to follow the general pattern of data trends. It struggles to capture extreme spikes. Underpredictions occur at the bottom (low consumption). The model shows stability, but is somewhat conservative.

Comparison of LSTM and Random Forest performance for predicting Global_active_power with the same feature inputs: Voltage, Global_reactive_power, Sub_metering_1, Sub_metering_2, Sub_metering_3.



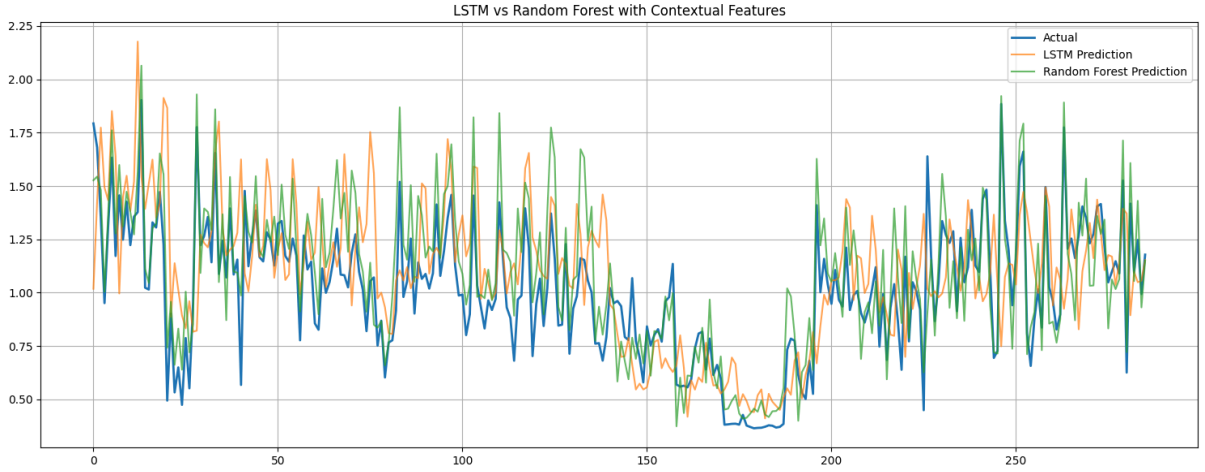
The LSTM's medium-term trend is good, while the Random Forest's tends to be unstable. The LSTM's response to spikes is slow, while the Random Forest's is responsive. The LSTM's predictions are very smooth, while the Random Forest's are sharp and sometimes overfit. The LSTM's fit to the time series is excellent, while the Random Forest's is poor.

Predict future values for the next few days (extrapolation) based on recent data with a trained LSTM model.



Prediction patterns follow historical trends. Fluctuations are more stable than actual data. Autoregression-based predictions. Potential for over/underestimation.

Comparison of Random Forest & LSTM using additional contextual features: Voltage, Global_reactive_power, Sub_metering_1/2/3, day_of_week, is_weekend, is_holiday, temp_avg.



Both models followed the general trend quite well. Random Forest was more volatile (responsive) but noisy. LSTM was more stable but somewhat slower to respond to drastic changes. Both still had gaps with the actual data.