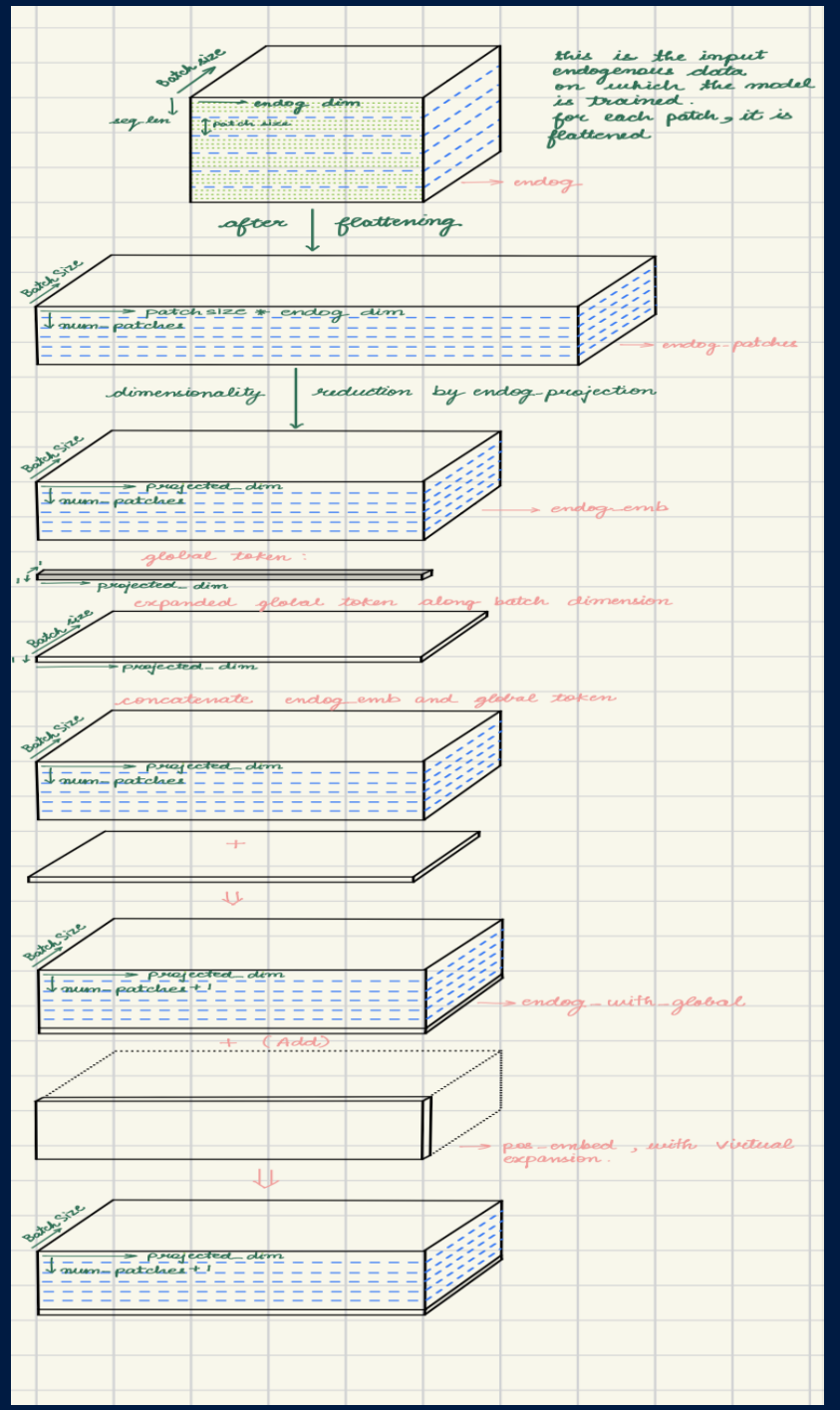
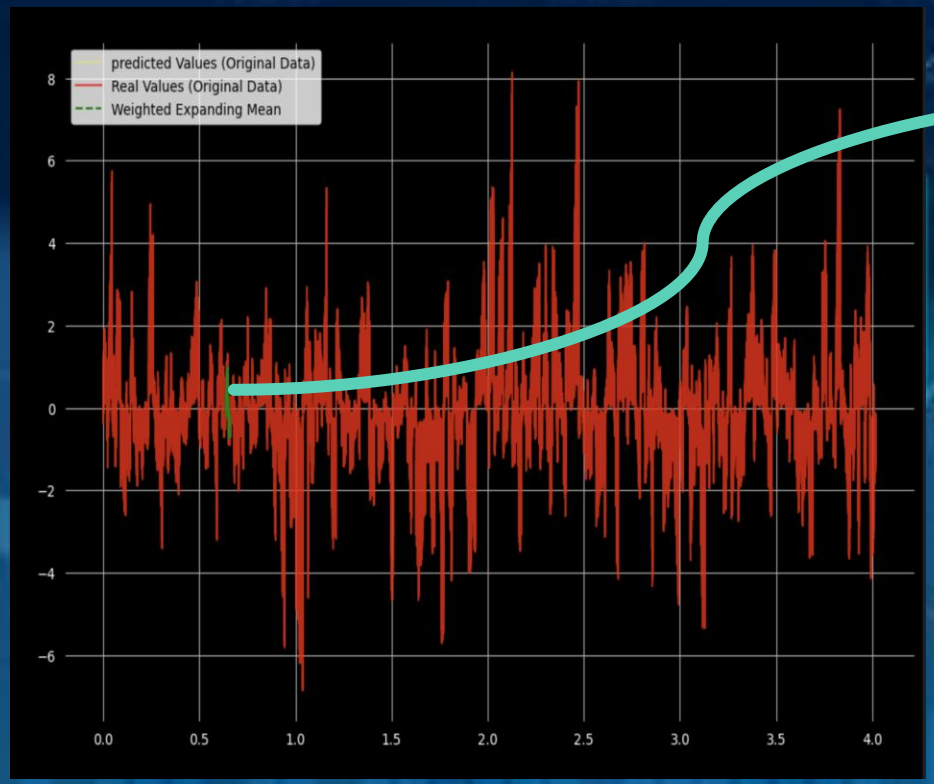
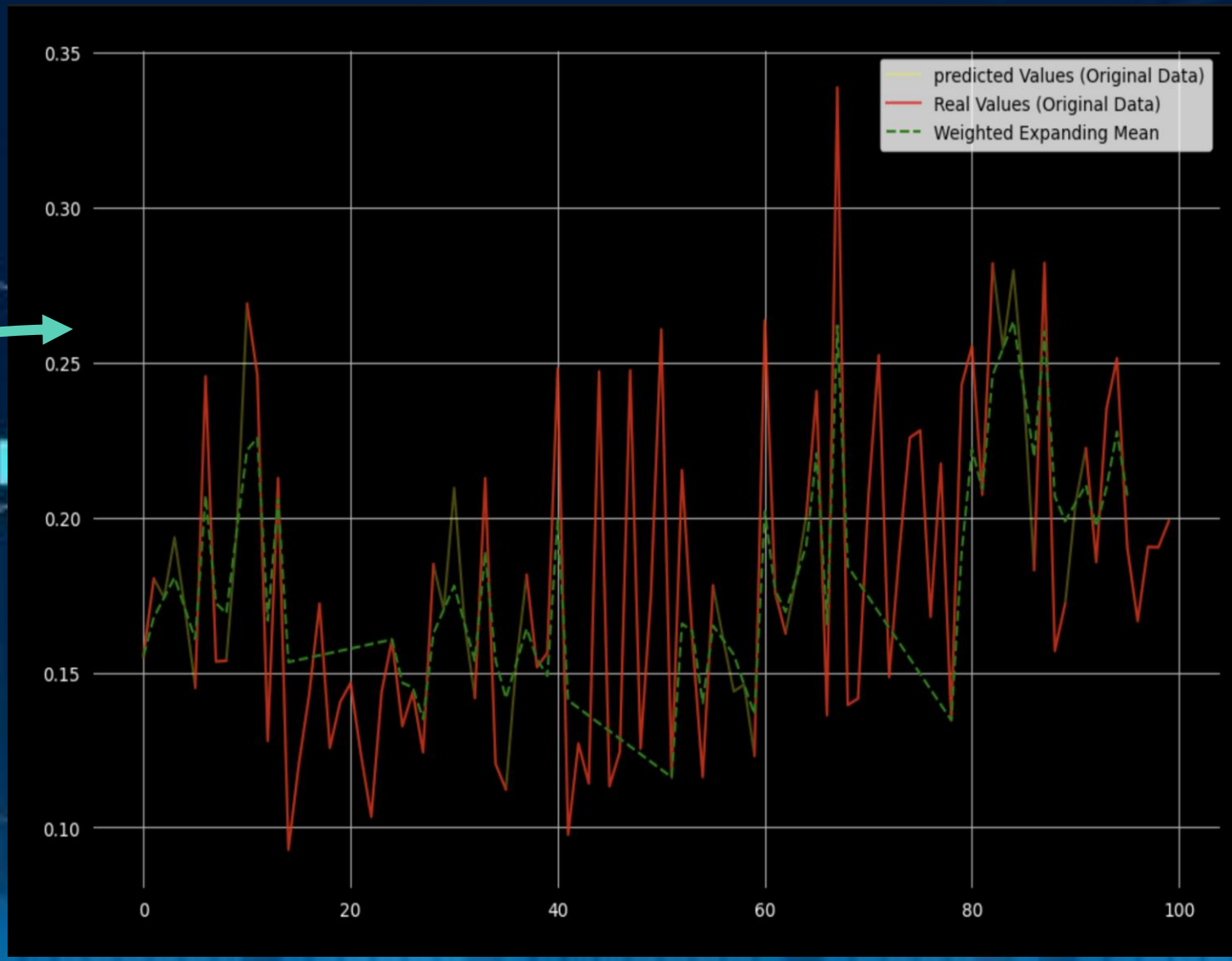
Data Pipeline

- **Input:** Multivariate Time-series of shape → [batch_size, num_features, seq_len]
- **Flattening:** Endogenous features are flattened for each patch → [batch_size, seq_len x num_features]
- **Exogenous join:** Append learned embeddings (e.g., 8D) for each feature → richer context per feature
- **Linear projection:** Combined input is projected to a shared projected_dim using a feedforward layer
- **Global token extraction:** One token is cropped from each sequence to represent global temporal context
- **Token fusion:** Global token is concatenated back with the sequence for joint learning
- **Positional embedding:** Fixed sinusoidal positions are added to the input sequence
- **Transformer encoder:** 2 stacked blocks with:
  - Multi-head self-attention + cross-attention
  - Feed-forward MLP + layernorm + dropout
- **Forecast head:**
  sequence is flattened → MLP head predicts future values (e.G., Next 1 step)
- **Output:**
  final shape → [batch_size, output_dim], trained with **weighted huber loss**

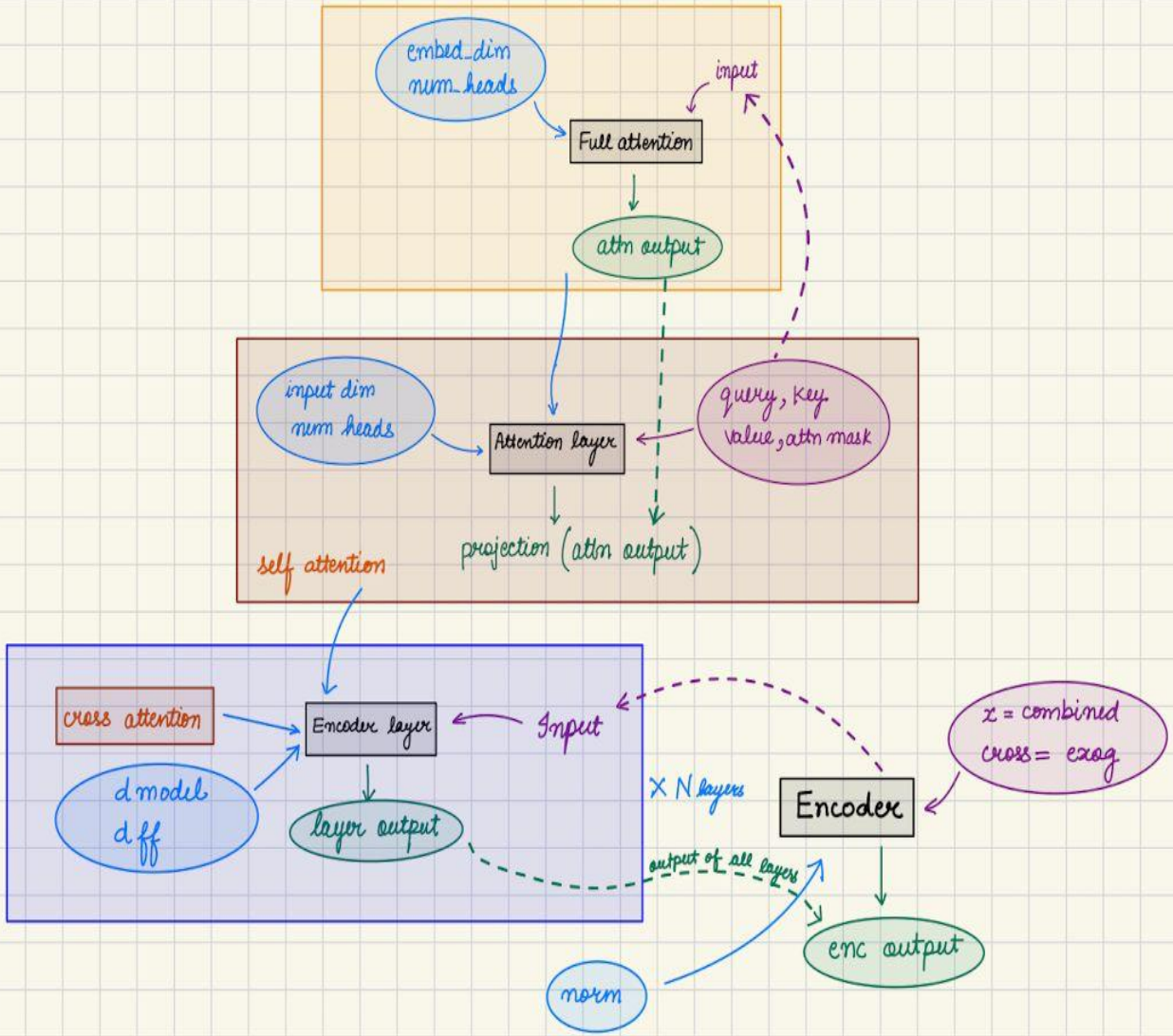


Data Pre-Processing





Encoder



Version 2.7 : Few fixes in 2.6

Code Flow :

FORECAST ILLUSTRATION