Theoretical Questions

1) Turbit primarily deals with time-series data generated by sensors for temperature, pressure, and other quantities. Which modern time-series forecasting approach would you choose to model such data? Describe its advantages and disadvantages. How would you implement it using TensorFlow?

   For time-series forecasting using sensor data such as Temperature and Pressure, I would choose to implement the Long Short-Term Memory networks (LSTMs). The following are the advantages and disadvantages of this model:

Advantages	Disadvantages
Captures temporal dependencies	Computationally expensive
Can identify patterns and trends in the time-series data	Prone to overfitting
Flexible framework that can be customized	Not as interpretable as traditional models

   To implement LSTM using TensorFlow, the following steps will be applied:

   1. Import necessary libraries such as TensorFlow, Pandas and any other required module.
   2. Load and Preprocess data. Scaling the data and splitting it into training and testing sets is a must.
   3. Reshape data as LSTM in TensorFlow expects input data to have a 3D shape: (number of samples, time steps, number of features).
   4. Build LSTM Model and Train it.
   5. Evaluate trained model on test set.
   6. Use trained LSTM to make predictions on unseen data.
   7. Finetune and optimize.

2) We often deal with data gaps. What approaches can you use to deal with missing data in the context of time-series? What are the advantages and disadvantages?

a) Drop data with missing value.

Adv.: Simple and easy to implement.

Avoids errors such as distorting the distribution of the dataset.

Dis.: Reduces the dataset size

Can result in the loss of valuable information.

b) Interpolation

Adv.: Preserves trends and seasonality

Dis.: Sensitive to outliers

c) Forward filling (LOCF)

Adv.: Preserves the temporal order of the data.

Simple to implement.

Dis.: Assumes that neighboring values are similar which could cause an error if the missing value occurred in a different condition.

d) Backward filling (NOCB)

Adv.: Preserves the temporal order of the data.

Simple to implement.

Dis.: Can cause leakage when implementing forecasting models

3) When reporting turbine anomalies, customers usually want to know the underlying reason for the unusual data points. Imagine that a thermometer in the gearbox of a turbine suddenly started reporting temperatures ten degrees higher than expected. List possible causes for such a pattern and how Turbit could distinguish between them?

a) Possible causes

1. Sensor Malfunction
2. Environmental changes
3. Mechanical issues
4. Algorithm/model error

b) Distinguishing between causes

1. Comparative Analysis: To determine whether there is a sensor malfunction or a more significant environmental change, compare the anomalous temperature data with those from the other sensors.

2. Historical Data Analysis: Look for patterns in historical data. A sensor error could be indicated by sudden spikes, while gradual variations could indicate mechanical or environmental issues.

3. Expert Consultation: To gain insight into possible causes, consult domain experts with experience in turbine operations.