Anomaly Detection and Remaining Useful Life Prediction on Industrial Dataset using Deep Learning and Probabilistic Graphical Models

Group Members:

- Karthick Vel Kathirvel (kk37347)
- Pritesh Singh (ps35762)
- Rathil Madihalli (rm63782)
- Sanyam Jain (sj33448)
- Vishal Anand Gupta (vag2478)

Problem Statement:

Predictive maintenance has become a crucial aspect of modern industrial operations, aiming to prevent unexpected equipment failures, enhance operational efficiency, and minimize downtime. The dataset provided captures a range of sensor readings from industrial machinery over time. The goal, using this dataset, is to build models for anomaly detection to identify abnormal behavior and predict the Remaining Useful Life (RUL) of the machinery. Furthermore, understanding causal relationships among the variables and enabling online learning for real-time updates are essential to improve predictive accuracy and the model's adaptability to evolving data patterns.

Dataset source: MetroPT-3 Dataset - UCI Machine Learning Repository

Data Description:

The dataset includes timestamped readings from various sensors, along with features such as temperature, pressure, and motor current among others. The dataset consists of 15,169,480 data points collected at 1Hz from February to August 2020 and is described by 15 features from 7 analogue (1-7) and 8 digital (8-15) sensors. The columns in the dataset are detailed as follows:

- **Timestamp:** The time at which the readings were recorded.
- **TP2:** Measure of the pressure on the compressor.
- **TP3:** Measure of the pressure generated at the pneumatic panel.
- **H1:** Measure of the pressure generated due to pressure drop when the discharge of the cyclonic separator filter occurs.
- **DV_pressure:** Measure of the pressure drop generated when the towers discharge air dryers; a zero reading indicates that the compressor is operating under load.
- **Reservoirs:** Measure of the downstream pressure of the reservoirs, which should be close to the pneumatic panel pressure (TP3).
- **Oil_temperature:** Measure of the oil temperature on the compressor.
- Motor_current: Measure of the current of one phase of the three-phase motor.
- **COMP:** Electrical signal of the air intake valve on the compressor.

- **DV_electric:** Electrical signal that controls the compressor outlet valve.
- **Towers:** Electrical signal that defines the tower responsible for drying the air and the tower responsible for draining the humidity removed from the air.
- **MPG:** Electrical signal responsible for starting the compressor under load by activating the intake valve when the pressure in the air production unit (APU) falls below 8.2 bar.
- **LPS:** Electrical signal that detects and activates when the pressure drops below 7 bars.
- **Pressure_switch:** Electrical signal that detects the discharge in the air-drying towers.
- **Oil_level:** Electrical signal that detects the oil level on the compressor.
- **Caudal_impulses:** Electrical signal that counts the pulse outputs generated by the absolute amount of air flowing from the APU to the reservoirs.

Proposed Approaches:

- **Anomaly Detection:**
  - Employ statistical methods like Z-score or IQR initially to identify basic anomalies.
  - Utilize Long Short-Term Memory (LSTM) and Recurrent Neural Networks (RNN) to capture temporal dependencies and identify anomalies over time.
- **Remaining Useful Life (RUL) Prediction:**
  - Implement LSTM and RNN to forecast the time-series data and predict the RUL.
  - Employ Bayesian Belief Networks (BBN) to incorporate uncertainties inherent in the sensor readings and operational conditions.
- **Causal Relationship Understanding:**
  - Utilize causal inference algorithms to elucidate the causal relationships among the various features.
  - Implement Probabilistic Graphical Models to visualize and compute the causal dependencies.
- **Online Learning:**
  - Adapt models to incorporate online learning algorithms enabling them to learn and update with each new data point.
  - Utilize techniques like Sequential Monte Carlo methods for online learning in Bayesian settings.
- **Exploration of Alternative Models:**
  - Consider other deep learning architectures like Variational Autoencoders (VAE) for anomaly detection.
  - Explore reinforcement learning for decision-making based on the model's insights.
- **Hybrid Approaches:**
  - Combine Bayesian and deep learning approaches for a more robust anomaly detection and RUL prediction.

    - Implement ensemble methods to amalgamate predictions from different models for better accuracy.
- **Evaluation and Validation:**
    - Employ cross-validation and other statistical tests to evaluate the performance of the models.
    - Utilize real-world failure data, if available, to validate the predictive accuracy of the RUL models.

References:

1. *Malhotra, P., et al. (2016) - LSTM-based encoder-decoder for multi-sensor anomaly detection:* This paper introduces an LSTM-based encoder-decoder framework for anomaly detection using multi-sensor data. It highlights the effectiveness of LSTM networks in capturing temporal dependencies, which is crucial for identifying anomalies in time-series data. The methodologies outlined in this paper are directly relevant to the anomaly detection aspect of the problem.
2. *Zhang, W., et al. (2018) - A new remaining useful life prediction algorithm based on nonlinear feature fusion and improved support vector machine:* This work proposes a novel algorithm for Remaining Useful Life (RUL) prediction, emphasizing nonlinear feature fusion and an enhanced support vector machine. The algorithm could provide a fresh perspective on feature engineering and machine learning approaches for RUL prediction, potentially augmenting the predictive accuracy of the models developed for this problem.
3. *Barber, D. (2012) - Bayesian Reasoning and Machine Learning:* This book provides a comprehensive overview of Bayesian reasoning and its application in machine learning. It could serve as a foundational reference for incorporating Bayesian Belief Networks (BBN) in the problem, enabling the assimilation of uncertainties inherent in sensor readings and operational conditions.
4. *Cappe, O., Moulines, E., & Ryden, T. (2005) - Inference in Hidden Markov Models:* This book delves into the realm of Hidden Markov Models (HMMs) and inference methodologies within them. Although HMMs were not explicitly mentioned in the proposed approaches, the book could provide valuable insights if a decision is made to explore HMMs for understanding temporal dynamics and causal relationships among the variables.
5. *Sutton, R.S., & Barto, A.G. (2018) - Reinforcement Learning: An Introduction:* This book is a cornerstone text on reinforcement learning (RL). It could be instrumental in exploring RL for decision-making based on the insights derived from the models. The principles and methodologies discussed in this book could guide the exploration of RL applications in optimizing operational decisions to mitigate the risks associated with equipment failures or other anomalies.