## Project Report: Vibration-Based Gear Fault Diagnosis

### 1. Project Objective

The primary goal of this project was to develop a highly accurate and reliable model for the predictive maintenance of industrial gears. The objective was to diagnose a "broken tooth" fault using vibration sensor data. A key component of this analysis was to compare the effectiveness of a traditional Machine Learning (ML) approach (**Random Forest**) against a Deep Learning (DL) approach (**1D-Convolutional Neural Network**) to determine the optimal solution for this type of time-series classification problem.

### 2. The Dataset

The analysis was performed on a comprehensive dataset designed to simulate real-world industrial conditions.

* **Data Source:** Vibration data collected from a gear test rig.
* **Classes:** The data was categorized into two distinct conditions:
  1. **Healthy:** Gears operating under normal conditions.
  2. **Faulty:** Gears with a deliberately broken tooth.
* **Sensors:** Four vibration sensors (a1, a2, a3, a4) were used to capture the system's dynamics.
* **Operating Conditions:** To ensure robustness, data was recorded under a range of ten different load conditions, from 0% to 90%.
* **Dataset Size:** The consolidated dataset contained over **2 million (2000k+) time-series data points**, providing a rich source for training and validation.

### 3. Data Preprocessing & Preparation

Before model training, the raw data was meticulously cleaned and structured to create a suitable feature set.

1. **Data Consolidation:** Individual CSV files for each load condition and fault type were merged into a single, master Pandas DataFrame.
2. **Feature Engineering:** Two crucial columns were added to provide context to the sensor readings:
   * load: The specific load percentage under which the data was recorded.
   * fault: The class label ('H' for Healthy, 'F' for Faulty).
3. **Data Standardization:** The four sensor reading columns were scaled using Scikit-Learn's StandardScaler. This step normalizes the data to have a mean of 0 and a standard deviation of 1, which is essential for the optimal performance of both ML and DL models.

### 4. Model Development & Comparative Analysis

Two distinct modeling approaches were implemented and evaluated on the same dataset.

#### 4.1. The Failed Model: Random Forest (Machine Learning)

A Random Forest Classifier was initially trained to distinguish between the healthy and faulty states.

* **Problem Encountered:** The model exhibited a classic and severe case of **overfitting**. It perfectly memorized the patterns in the training data but was unable to generalize this knowledge to new, unseen data.
* **Performance:**
  + **Training Accuracy:** 100%
  + **Test Accuracy:** **59.04%**
* **Evidence of Failure:** The confusion matrix for the test set showed that the model was only slightly better than random guessing, misclassifying nearly half of the instances. This rendered it useless for any practical diagnostic application.

#### 4.2. The Successful Model: 1D-Convolutional Neural Network (Deep Learning)

A 1D-CNN was designed using TensorFlow and Keras, as this architecture is exceptionally well-suited for finding patterns in time-series data.

* **How it Worked:** The CNN automatically learned the relevant, discriminative features from the raw vibration signals, identifying the unique signatures of a broken tooth fault.
* **Performance:**
  + **Training Process:** The model's accuracy on the validation (test) set rapidly improved with each epoch, reaching a perfect score.
  + **Final Test Accuracy:** **99.98%**
* **Evidence of Success:** To visualize the features learned by the CNN, a **t-SNE plot** was generated. The plot showed two perfectly distinct and widely separated clusters for the 'Healthy' and 'Faulty' classes, providing clear visual proof that the model had learned to effectively differentiate between the two conditions.

### 5. Final Results & Conclusion

The direct comparison of the two models on the unseen test data reveals a clear winner.

| Metric | Random Forest (ML) | 1D-CNN (DL) |

| Test Accuracy | 59.04% | 99.98% |

| Outcome | Failed (Overfit) | Successful & Reliable |

Conclusion:

This project successfully demonstrates that for complex, high-frequency sensor data, a Deep Learning approach is significantly more effective than traditional Machine Learning. The 1D-CNN was able to overcome the challenge of overfitting and build a generalized, highly accurate model capable of reliably diagnosing a broken tooth fault in gears. The final model serves as a robust and effective tool for predictive maintenance.

### 6. Technical Summary

* **Programming Languages:** Python
* **Scientific Packages:** NumPy | Pandas | Seaborn | Matplotlib | Scikit-Learn | TensorFlow | Keras
* **Key Skills:** Data Preprocessing, Feature Engineering, Supervised Learning, Deep Learning (CNNs), Model Evaluation, Overfitting Diagnosis, Dimensionality Reduction (t-SNE).