**Predictive Maintenance Using Machine Learning**

A Project Report for the DEPI Scholarship

**Team Members**

- Ahmed Nasser Sayed  
- Ahmed Ashraf Mohammed  
- Mohammed Sobhy Lotfy  
- Youssef Salah Ahmed  
- Youssef Mohsen Poules

**Table of Contents**

1 Introduction

2. Objective

3. Dataset Description

4. Methodology

5. Exploratory Data Analysis (EDA)

6. Model Development

7. Results and Evaluation

8. Conclusion

9. Future Work

## Introduction

Predictive maintenance is a crucial application of machine learning in industrial settings, enabling the prediction of potential machine failures before they occur. This project leverages historical operational data to build models that predict equipment failures, reducing downtime and improving overall efficiency.  
  
This report documents the development, evaluation, and findings of the Predictive Maintenance project conducted as part of the DEPI Scholarship initiative.

## Objective

The primary objective of this project is to develop a predictive maintenance model that:  
- Accurately forecasts machine failures.  
- Provides actionable insights for maintenance scheduling.  
- Reduces unplanned downtime and maintenance costs.

## Dataset Description

- \*\*Dataset Name:\*\* Predictive Maintenance Dataset  
- \*\*Source:\*\* Provided for the DEPI Scholarship project.  
- \*\*Description:\*\* The dataset contains operational data with features including:  
 - `Air temperature [K]`  
 - `Process temperature [K]`  
 - `Rotational speed [rpm]`  
 - `Torque [Nm]`  
 - `Tool wear [min]`  
 - `Type` (Categorical: L, M, H)  
 - `Target` (Indicates failure or non-failure of the machine)  
- \*\*Download Link:\*\* [Predictive Maintenance Dataset](#)

## Methodology

### Data Preprocessing:  
1. Handled missing values and checked for data imbalances.  
2. One-hot encoded categorical data (`Type` column with values L, M, H).  
3. Applied SMOTE (Synthetic Minority Over-sampling Technique) to address target class imbalance.  
  
### Exploratory Data Analysis (EDA):  
1. Visualized correlations among features using heatmaps.  
2. Analyzed target variable distribution.  
3. Explored failure distribution across machine types.  
  
### Model Development:  
1. Experimented with multiple machine learning models:  
 - Logistic Regression  
 - Decision Tree Classifier  
 - Random Forest Classifier  
 - Support Vector Machine (SVM)  
2. Evaluated model performance using metrics such as accuracy and classification reports.  
  
### Hyperparameter Tuning:  
- Optimized the Random Forest Classifier by adjusting parameters like:  
 - Number of estimators: 500  
 - Maximum depth: 10

## Exploratory Data Analysis (EDA)

### Correlation Heatmap:  
- Visualized feature correlations to identify significant relationships.  
  
### Target Variable Distribution:  
- Balanced the `Target` variable using SMOTE.  
  
### Machine Type Failures:  
- Visualized failure rates across machine types (L, M, H).

## Model Development

### Machine Learning Models:  
- \*\*Logistic Regression\*\*  
 - Simple baseline model with moderate performance.  
- \*\*Decision Tree Classifier\*\*  
 - Captured feature importance but prone to overfitting.  
- \*\*Random Forest Classifier\*\*  
 - Achieved the best results due to its ensemble approach.  
- \*\*Support Vector Machine (SVM)\*\*  
 - Effective for complex patterns but computationally expensive.  
  
### Model Comparison:  
| Model | Accuracy |  
|-------------------------|----------|  
| Logistic Regression | 0.845 |  
| Decision Tree Classifier| 0.880 |  
| Random Forest Classifier| \*\*0.910\*\*|  
| Support Vector Machine | 0.875 |  
  
### Final Model:  
- \*\*Tuned Random Forest Classifier\*\*  
 - \*\*Accuracy:\*\* 91.0%  
 - \*\*Hyperparameters:\*\*  
 - Number of estimators: 500  
 - Maximum depth: 10

## Results and Evaluation

### Key Metrics:  
- \*\*Precision, Recall, F1-Score:\*\* Detailed performance metrics are included in the classification reports.  
- \*\*Model Robustness:\*\* Random Forest proved to be robust with consistent results on test data.  
  
### Model Insights:  
- `Tool wear [min]` and `Torque [Nm]` were significant predictors of failure.  
- Machine Type (`L`, `M`, `H`) influenced failure patterns.

## Conclusion

This project successfully developed a machine learning pipeline for predictive maintenance, achieving high accuracy with the Random Forest Classifier. The model provides actionable insights for scheduling maintenance and preventing machine failures.

## Future Work

1. Integrate real-time data for live predictions.  
2. Explore advanced algorithms like Gradient Boosting and Neural Networks.  
3. Deploy the model as a web service for operational use.