**Case Study: Electrical Faults Detection and Classification using Machine Learning**

**1. Introduction**

Electrical faults in power systems can cause significant damage if not identified and addressed promptly. In this case study, we explore the use of machine learning models to detect and classify faults in electrical power systems. The dataset provided contains measurements from various sensors, including current and voltage values, under different fault conditions. Our objective is to train and evaluate machine learning models to classify the type of fault present in the system.

**2. Dataset Overview**

The dataset consists of several features related to the power system's operation, including:

* **Ia, Ib, Ic**: Current values for phases A, B, and C.
* **Va, Vb, Vc**: Voltage values for phases A, B, and C.
* **Fault\_Type**: A categorical label indicating the fault condition (e.g., 'No Fault', 'Line A to Ground Fault', etc.).

We combined multiple columns to form a single column (Fault\_Type), representing different types of faults in the power system.

**3. Data Preprocessing**

We performed the following steps to preprocess the data:

* **Combining Fault Types**: The Fault\_Type was created by combining the individual fault columns (A, B, C, Ground).
* **Mapping Fault Types**: Fault types were mapped to more descriptive labels (e.g., 'Line A to Ground Fault').
* **Label Encoding**: The Fault\_Type column was label-encoded into numerical values for machine learning models.

**4. Exploratory Data Analysis (EDA)**

**Fault Type Distribution**

The following two visualizations show the distribution of fault types in the dataset:

1. **Count Plot**: This plot displays the frequency of each fault type in the dataset. It helps us understand how common each fault type is.
2. **Pie Chart**: This pie chart shows the proportional distribution of fault types. It highlights which fault types are most common and which are least frequent.

**Current vs. Voltage Plots for Different Fault Types**

We also visualized the current (Ia, Ib, Ic) and voltage (Va, Vb, Vc) measurements for various fault conditions:

1. **No Fault (Healthy System)**: The current and voltage waveforms for a healthy system (No Fault) are expected to be stable and periodic.
2. **Line A to Ground Fault**: When a fault occurs, the current and voltage waveforms show abrupt changes. For example, in a Line A to Ground Fault, the waveform of the current in Line A will show a sudden rise.
3. **Line A Line B to Ground Fault**: Similar to the previous example, but involving multiple phases.
4. **Line B to Line C Fault**: This fault causes a significant change in the waveforms of currents and voltages between these two lines.

**5. Model Development**

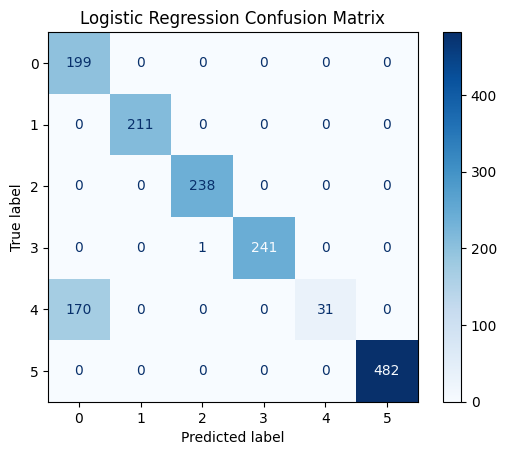
We applied multiple machine learning models to classify the fault types. These models include:

* **Logistic Regression**
* **Decision Tree Classifier**
* **Random Forest Classifier**
* **XGBoost Classifier**
* **Support Vector Machine (SVM)**

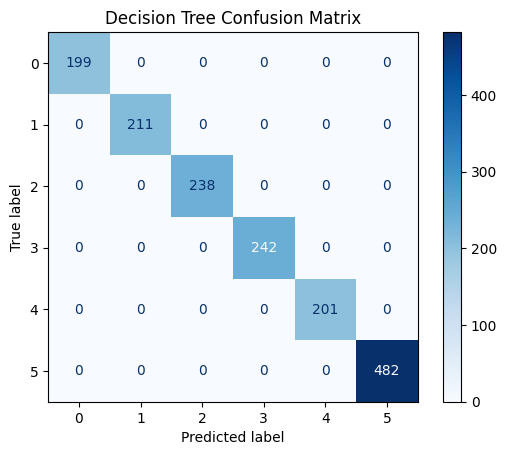
**Model Evaluation Metrics**

The following confusion matrices were plotted to evaluate each model's performance:

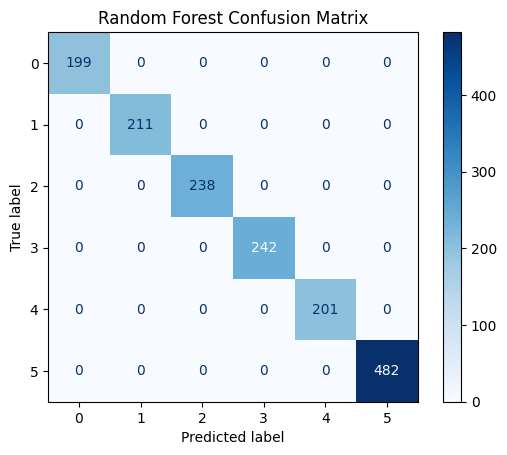
1. **Logistic Regression** Confusion Matrix:



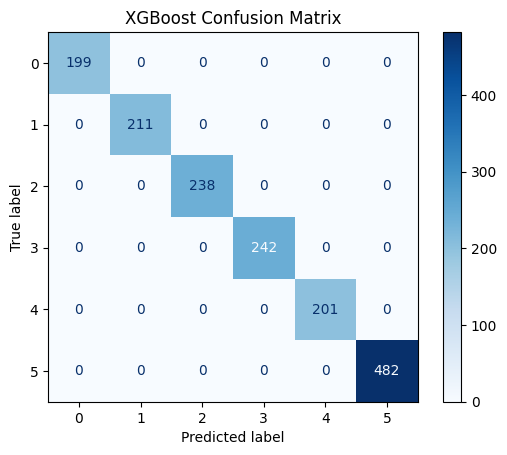
1. **Decision Tree Confusion Matrix**:



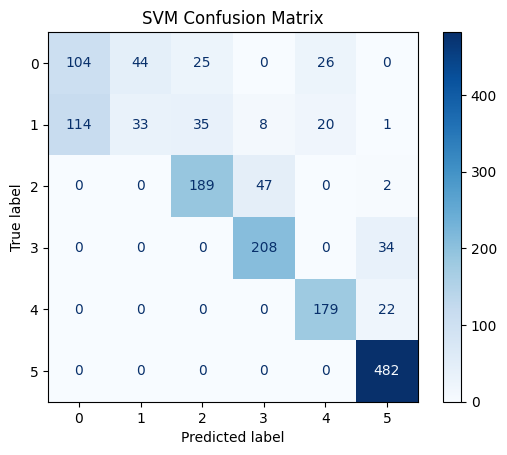
1. **Random Forest Confusion Matrix**:



1. **XGBoost Confusion Matrix**:



1. **Support Vector Machine Confusion Matrix**:



**Comparison of Models**

The table below compares the performance of each model based on **Training Accuracy** and **Model Accuracy**:

|  |  |  |
| --- | --- | --- |
| **Model** | **Training Accuracy (%)** | **Model Accuracy Score (%)** |
| **Random Forest** | 100 | 100 |
| **Decision Tree** | 100 | 100 |
| **Logistic Regression** | 100 | 100 |
| **XGBoost** | 89.19 | 89.13 |
| **Support Vector Machine** | 76.07 | 75.97 |

**6. Model Deployment & Application**

**These trained models can be used in real-time power systems to predict faults based on real-time measurements. Applications include:**

* **Fault Diagnosis: Automatically classifying faults based on real-time data to initiate corrective actions.**
* **Maintenance Planning: Based on fault predictions, scheduling maintenance for affected parts to prevent downtime.**
* **System Monitoring: Identifying faulty conditions early, preventing system-wide failures, and enhancing grid reliability.**

**7. Conclusion**

**The machine learning models applied to classify electrical faults performed impressively, with Random Forest being the most accurate, followed closely by Decision Tree. Both models achieved near-perfect performance on the training dataset and excellent generalization to the test data.**

**By leveraging machine learning for fault detection, utilities can enhance their response times, improve system stability, and reduce maintenance costs. Further research could focus on optimizing model hyperparameters, incorporating additional data features, and exploring real-time deployment strategies.**