**⚡ General Overview**

This dataset is a simulation/monitoring dataset for the IEEE 6-bus system, containing:

Voltages (VA, VB, VC) per phase at buses.

Currents (IA, IB, IC) per phase on transmission lines.

Fault classification labels (Fault\_Class, Line\_Faulty, Fault\_Location).

**✅ Column Explanation**

🔌 Voltage Measurements

VA1, VB1, VC1: Phase A, B, and C voltages at Bus 1.

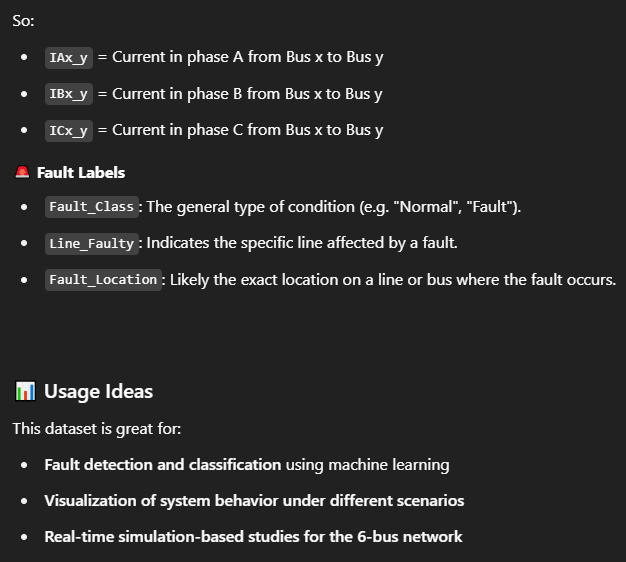
VA2, ..., VC6: Same for Buses 2–6.

⚡ Current Measurements on Lines

IA1-2, IB1-2, IC1-2: Phase currents from Bus 1 to Bus 2.

IA1\_6, IB1\_6, IC1\_6: Currents from Bus 1 to Bus 6.

Similar naming for IA6\_1, IB6\_1, IC6\_1 (Bus 6 to Bus 1), etc.



Target Column (s):  
  
🏁 Target Column: Most Likely Fault\_Class

Based on your dataset, the most likely target column is:

✅ Fault\_Class

This column contains the following labels 'Normal', 'AG', 'BG', 'CG', 'ABG', 'ACG', 'BCG', 'ABC', 'AB', 'AC', and 'BC'.

It's categorical — perfect for classification tasks.

This would be the label you’re trying to predict based on voltage and current features.

🔍 Supporting Columns:

Line\_Faulty: Could be a secondary target if you're doing fault localization (i.e. which line has a fault).

Fault\_Location: Useful for fine-grained fault analysis, like pinpointing the exact segment on a line.

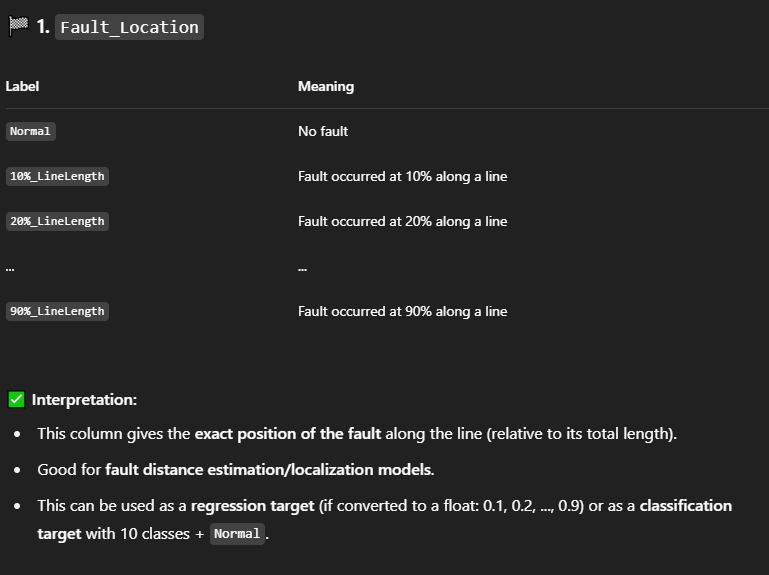
So you could treat:

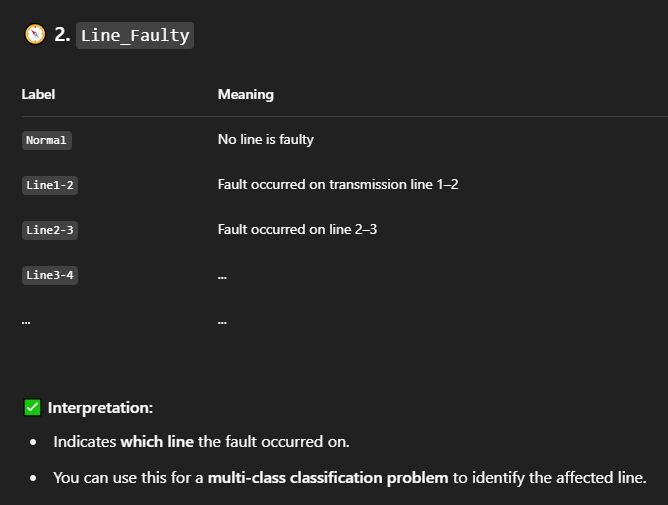
Fault\_Class as the main classification target,

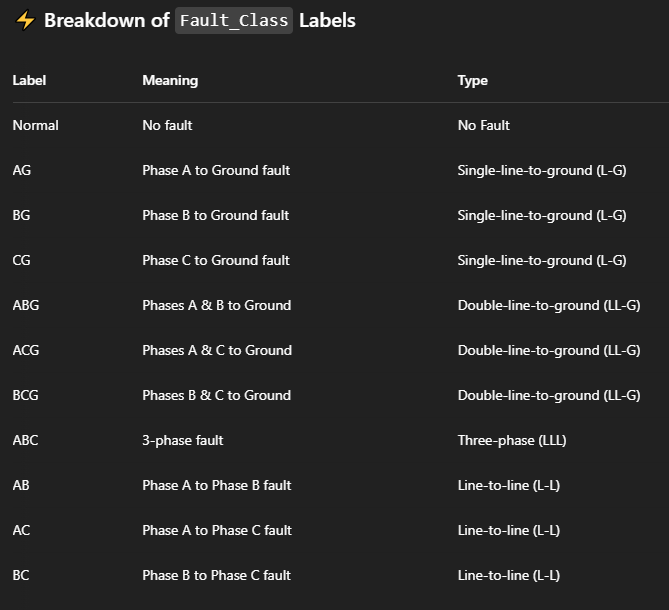
and optionally use Line\_Faulty or Fault\_Location for multi-label or hierarchical classification.

✅ Example Use Cases

| Input Features | Target | ML Type |
| --- | --- | --- |
| Voltages and Currents | Fault\_Class | Classification |
| Voltages and Currents | Line\_Faulty | Classification |
| Voltages, Currents, and Line\_Faulty | Fault\_Location | Classification or Regression (if numeric) |







🧠 What This Means for Modeling

You’ll be training a classifier to recognize these 11 distinct classes based on features like voltage and current readings from various buses and lines.

Example ML Problem Statement:

“Given the voltage and current readings in the IEEE 6-bus system, predict the type of fault that has occurred (or if it's a normal condition).”

Trianing Steps:

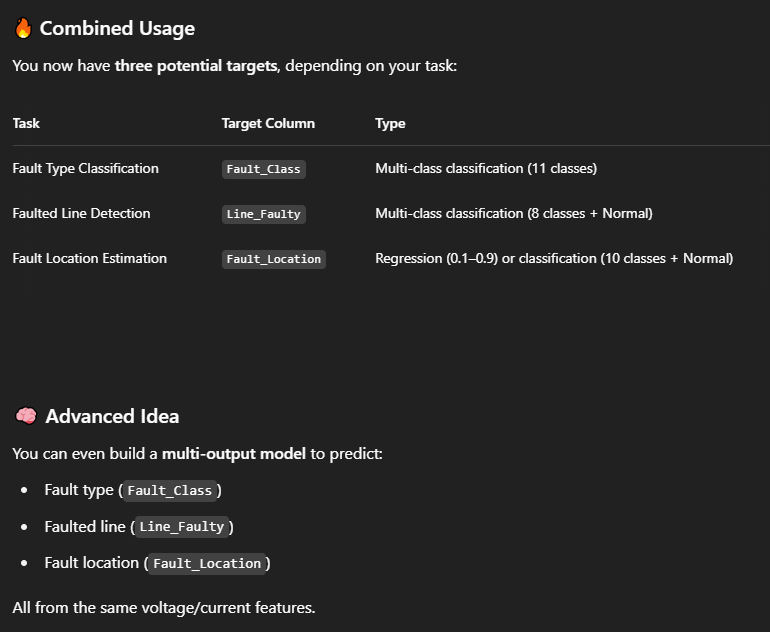
Encode the Fault\_Class column to numeric labels (Label Encoding or One-Hot).

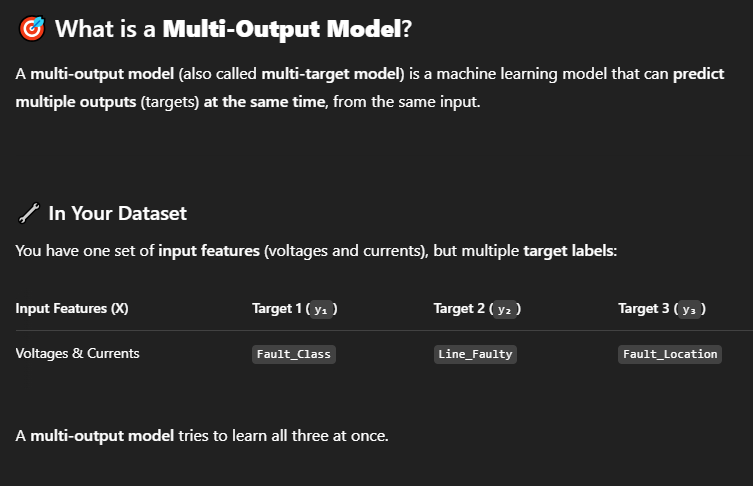
Normalize/standardize the features (voltage/current).

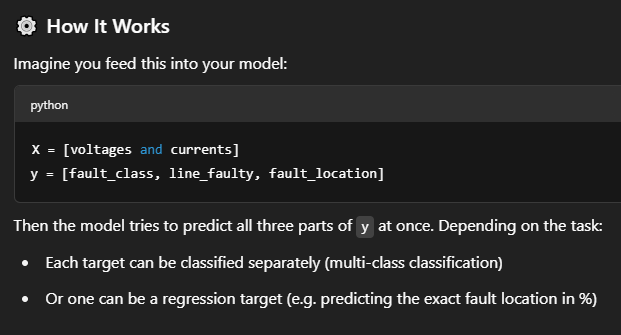
Split into train/test sets.

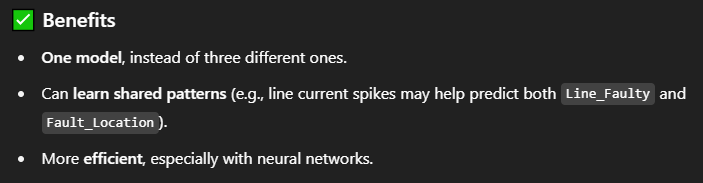
Train a model (e.g. Random Forest, XGBoost, or even deep learning).

Evaluate with confusion matrix, accuracy, F1-score, etc.





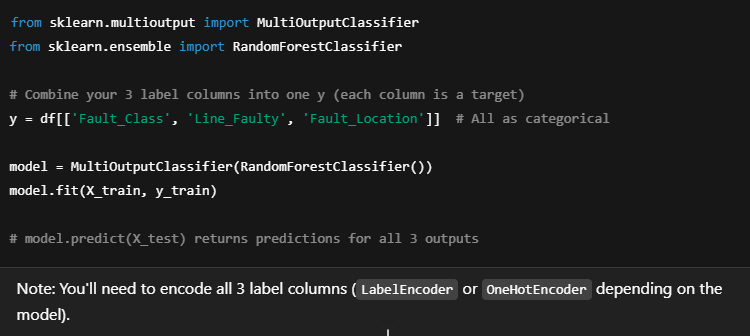




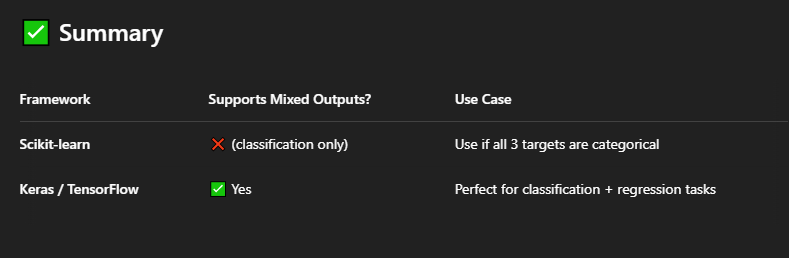
Code Examples

Using Scikit-learn (Treating all targets as classification)

If you treat Fault\_Location as a multi-class classification (e.g. 10 bins like '10%\_LineLength'), then:



One might consider, converting the Fault\_Location to numeric column and using a regression Model…But that won’t be necessary since the data was simulated on simply those 11 categories



With the 3 target columns ; Fault\_Class, Line\_Faulty, and Fault\_Location

We’re tackling a multi-output classification and regression problem in the context of power system fault analysis.

✅ Model Name (Descriptive & Technical)

Here are a few suitable names depending on the level of formality you need:

🔹 Technical Descriptive Names

Multi-Output Fault Diagnosis Model for IEEE 6-Bus System

Integrated Fault Classification, Localization, and Identification Model

Hybrid Fault Analysis Model using Voltage and Current Signatures

End-to-End Intelligent Fault Detection Framework

🔹 Research Paper-style Name

"A Multi-Task Learning Model for Simultaneous Fault Classification, Localization, and Line Identification in Power Systems"

✅ Problem Statement

"This work proposes a multi-output machine learning model that simultaneously classifies the type of fault (Fault\_Class), identifies the faulty transmission line (Line\_Faulty), and estimates the fault location along the line (Fault\_Location) in the IEEE 6-bus system, using voltage and current time-series features. The aim is to enhance the speed and accuracy of automated fault diagnosis and protection in smart grid systems."

Or more simply:

"Develop a model that detects the type, location, and affected line of faults in a 6-bus power system using machine learning."

✅ Problem Type (ML Terminology)

| Target Column | Task Type |
| --- | --- |
| Fault\_Class | Multi-class classification (11 classes) |
| Line\_Faulty | Multi-class classification (8 classes + Normal) |
| Fault\_Location | Either regression (continuous location %) or multi-class classification (10 discrete bins + Normal) |
| Overall | Multi-output classification |

This is technically a multi-output supervised learning problem involving classification.

Final Problem Setup:

| Target Column | Type | Number of Classes |
| --- | --- | --- |
| Fault\_Class | Multi-class classification | 11 |
| Line\_Faulty | Multi-class classification | 8 + 1 ('Normal') |
| Fault\_Location | Multi-class classification | 10 + 1 ('Normal') |

So we're building a multi-output, multi-class classification model.

✅ Suitable Model Types

We can confidently go with:

Scikit-learn: Use MultiOutputClassifier with any base classifier (e.g. RandomForest, XGBoost)

Keras/TensorFlow: Use a model with 3 softmax output layers