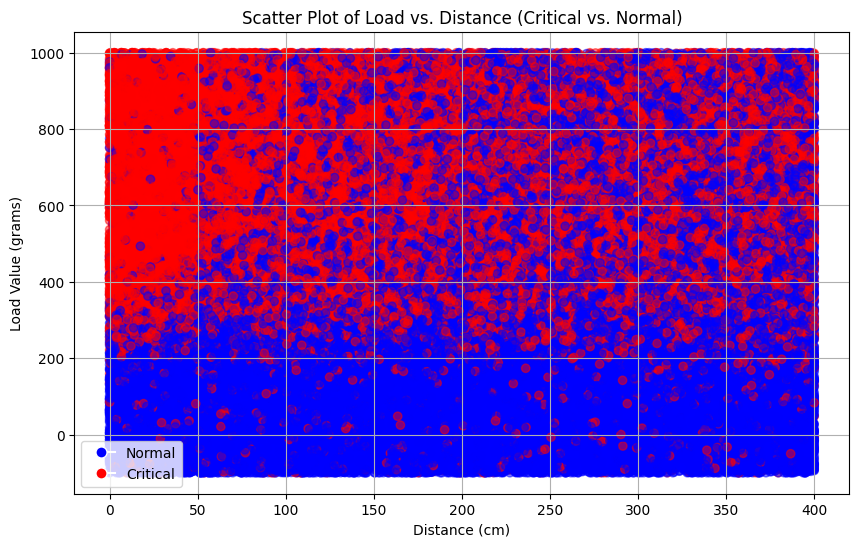
### **Data Analysis and Model Interpretation for Sensor-based Classification**

### **1. Scatter Plot (Load vs. Distance)**



### **What We See**: This scatter plot shows two main features — **distance** (x-axis) and **load value** (y-axis). Points are color-coded by labels: blue for **Normal** and red for **Critical**.

### **Key Observations**:

### **Red (Critical) points** spread across a wide range of load and distance values, indicating that critical conditions can occur in various combinations of these values.

### **Blue (Normal) points** are more concentrated in a lower range of load values and distances.

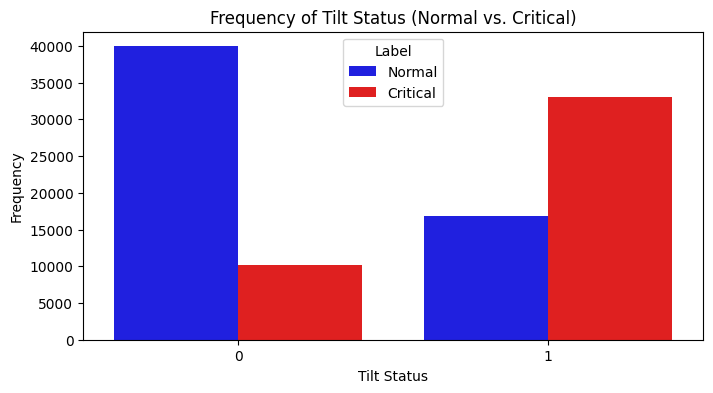
### **Interpretation**:

### **Insight for Model Performance**: This distribution suggests that the model could potentially identify critical conditions based on where data points fall in relation to certain ranges of load and distance. For example, if a point falls within a high load and long distance, it may indicate a higher likelihood of being labeled as critical.

### 

### 

### **2. Bar Plot (Frequency of Tilt Status)**



### **What We See**: This bar plot shows the frequency of each **tilt status** (0 or 1), with each bar divided into **Normal** (blue) and **Critical** (red) segments.

### **Key Observations**:

### **Tilt Status 0** (no tilt) shows a higher number of normal (blue) cases than critical (red) cases.

### **Tilt Status 1** (tilted) has a stronger association with critical (red) cases compared to normal (blue) cases.

### **Interpretation**:

### **Insight for Model Performance**: This pattern suggests that a **tilted status** (1) is a potential **indicator of critical conditions**. If the model sees a tilt status of 1, it could consider this as a significant predictor of a critical label, especially when combined with other factors like load and distance.

### 

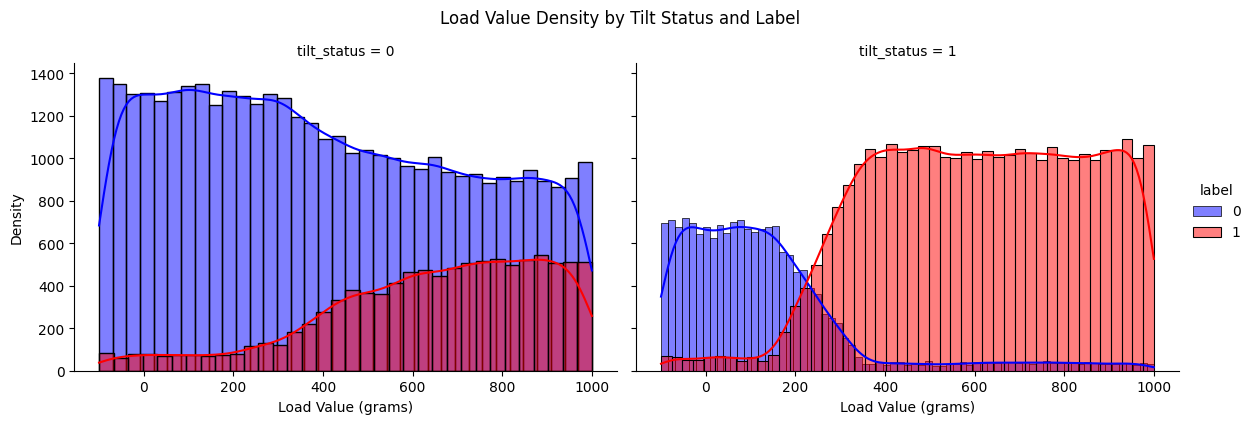
### 

### 

### 

### 

### **3. Conditional Density Plot (Load Value Density by Tilt Status and Label)**



* **Observation**: In the density plots, for tilt\_status = 0, the normal (blue) label shows a wide distribution across load values, while the critical (red) label appears more concentrated at higher load values, gradually increasing from around 200 grams onward. For tilt\_status = 1, the critical label dominates the distribution, especially after 200 grams, while the normal label is minimal.
* **Conclusion**: The load value is a strong indicator of critical status, especially when the tilt status is 1. This suggests that high load values combined with an active tilt status are likely predictors of critical events. This conditional dependence can inform model adjustments to emphasize these features for predicting critical conditions.

### 

### 

### 

### 

### 

### 

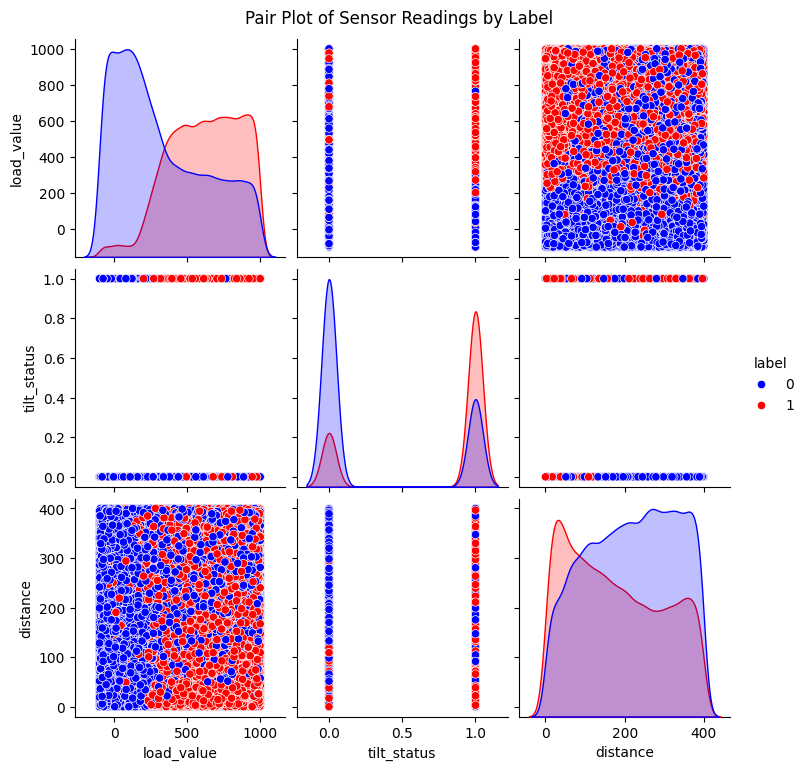
### 

### 

### 

### 

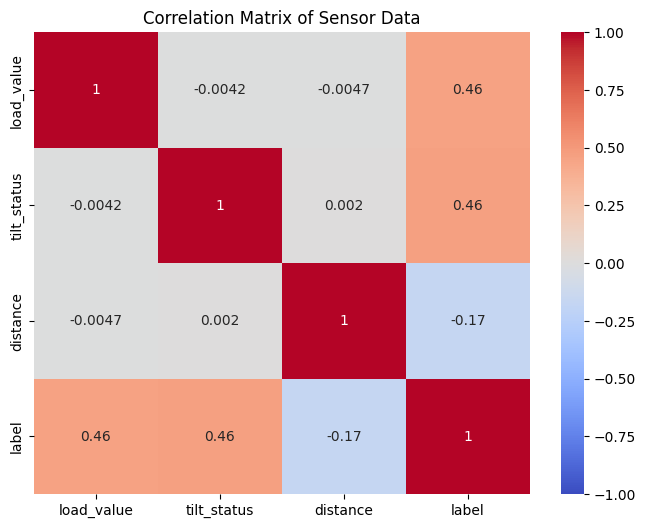
### **4. Pair Plot of Sensor Readings by Label**



* **Observation**: The pair plot reveals distinct distributions and potential separations for the labels across combinations of load\_value, tilt\_status, and distance. For instance, load\_value appears to show different density trends between the labels, and tilt\_status provides clear separation between critical and normal labels.
* **Conclusion**: The relationships between load\_value and tilt\_status are significant, indicating that they may be key predictors of the critical label. The visualization suggests that load\_value and tilt\_status interact in a way that can help distinguish between normal and critical labels, while distance has less visible separation and may contribute less.

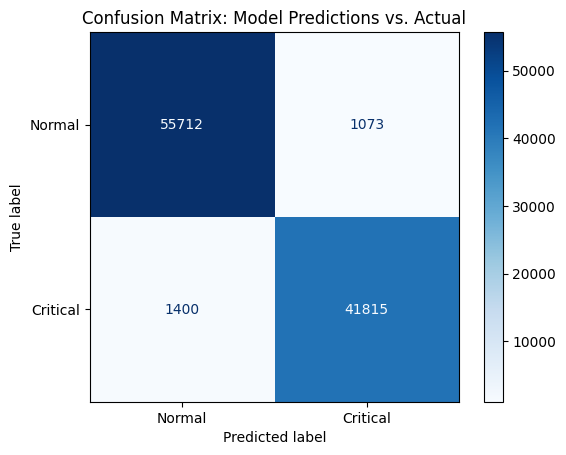
### 

### **5. Correlation Matrix of Sensor Data**



* **Observation**: The correlation matrix shows that both load\_value and tilt\_status have a moderate positive correlation (0.46) with the label. In contrast, distance has a weak negative correlation (-0.17) with label.
* **Conclusion**: Load\_value and tilt\_status are influential features with respect to the critical label, reinforcing the importance of these features in model training. The weak correlation of distance with the label suggests it might play a smaller role in classification, possibly acting as a supplementary feature rather than a primary predictor.

**6. Model Confusion Matrix**

****

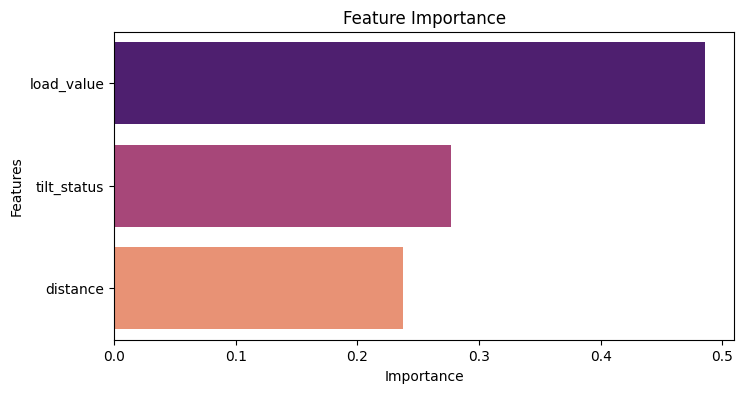
This confusion matrix provides insights into the performance of a binary classification model that predicts "Normal" and "Critical" conditions. Here’s a breakdown of each element:

* **True Positives (Bottom-right, 41815):** Cases that were correctly predicted as "Critical" (actual label is Critical, and the model predicted Critical).
* **True Negatives (Top-left, 55712):** Cases that were correctly predicted as "Normal" (actual label is Normal, and the model predicted Normal).
* **False Positives (Top-right, 1073):** Cases that were incorrectly predicted as "Critical" when they were actually "Normal" (model false alarms).
* **False Negatives (Bottom-left, 1400):** Cases that were incorrectly predicted as "Normal" when they were actually "Critical" (missed detections of critical conditions).

### **Interpretation:**

* **High True Positive and True Negative values** indicate that the model is accurately classifying a large portion of both "Normal" and "Critical" cases.
* **Low False Positive and False Negative values** show that the model is making relatively few errors, but there are still some critical cases missed (1400).

**7. Feature Importance**

****

**Feature Importance**: Each feature's importance is a measure of its contribution to the model’s predictive power. In Random Forests, this is typically computed by evaluating how much each feature reduces the error (or increases accuracy) across all the trees in the ensemble.

**Features**:

* **load\_value**: This feature has the highest importance score, indicating it contributes the most to the model’s predictions.
* **tilt\_status**: This feature is moderately important, meaning it has a significant impact but is less influential than load\_value.
* **distance**: This feature has the lowest importance among the three, but it still contributes to the model's predictions.

**Interpretation**:

* **load\_value** likely has the most predictive power, suggesting that variations in load\_value are strongly associated with changes in the target variable (label in your dataset).
* **tilt\_status** has a moderate impact, which may mean that the status (or angle) of tilt influences the target but not as heavily as load\_value.
* **distance** has the least impact, meaning it contributes the least information for predicting the target variable compared to the other two features.