Network Intrusion Detection 1 .ipynb

1. **Data Loading from the Google Drive**
2. **Reading and Combining the Dataset**
3. **Displaying the Combined Dataset**
4. **Basic Inspection of the Combined Dataset (Shape, Info, No of Rows and displaying the first 5 rows and statistical analysis)**
5. **Handling the Missing Values**
    a. **Displaying the Count in the heat map**
    b. **Handling the Missing Values (With 0 or Median)**
    c. **Zero tolerance approach**
6. **Oversampling**
    a. **Identifying Minority and Majority Classes**
    b. **Separate Data by Class**
    c. **Using SMOTE for the Oversampling technique**

Network Intrusion Detection 2 . ipynb (Feature Engineering)

1. **Loading the Data using the Apache Spark by Creating the Clusters**
2. **Exploratory Data Analysis with plots which include the Label Distribution and pie charts and Visualized the Important Columns**
3. **Identifying the Numerical and Categorical columns**
4. **Converting the Label from String to the Numerical Values using the StringIndexer**
5. **Converting the Remaining columns also into proper integers(Casting into integers)**
6. **Applying the Standard Scalar to Normalize the Values**
7. **Combining the Features and Labels**
8. **Creating the Train Test Data in the 70-30 Splitting with the Random State 42**
9. **Training the Data using the XGBoost Model and testing the model with the test data and Got the Model Accuracy of the 95 percent**
10. **Saved the Model in the PKL file format for the Future Use of the Model Weights to the Google Drive**

Tasks For the Next Week

Training the Model on Different Algorithms
Models Refining improving the Accuracy and Loss
Saving Model and Preparing the End point using the Fast APi and Integrating it to the Web Application
Building on the React JS
Preparing the Tableau Dashboard