**Detailed Baseline Model Inference and Model Selection Report**

**1. Objective**

The goal of baseline model training was to:

* Establish simple benchmark performances.
* Understand if simple models could capture the complexity of cybersecurity incident classification.
* Choose an appropriate model family for advanced modeling.

**2. Baseline Models Trained**

| **Model** | **Notes** |
| --- | --- |
| **Logistic Regression** | Linear model, good for simple and linearly separable datasets. |
| **Decision Tree Classifier** | Non-linear model, handles categorical and complex patterns well. |

**3. Model Performance Summary**

| **Metric** | **Logistic Regression** | **Decision Tree Classifier** |
| --- | --- | --- |
| Accuracy | 55.65% | 92.72% |
| Macro-F1 Score | 0.3517 | 0.8996 |
| Notes | Poor convergence, weak at minority classes | Strong, balanced across all classes |

**4. Detailed Observations**

**Logistic Regression**

* **Convergence Warning**: Did not converge to optimal solution even after 1000 iterations.
* **Poor Class-wise Recall**: Especially bad for minority classes.
* **Problematic Assumptions**:
  + Data was **not scaled** (label-encoded categorical features).
  + Data is **not linearly separable**.
* **Conclusion**: Logistic Regression is **not suitable** for this complex multi-class problem.

**Decision Tree Classifier**

* **High accuracy** (~93%) and **very strong Macro-F1** (~0.9).
* **Balanced precision and recall** across all classes.
* **Handled imbalances and label-encoded categorical data natively**.
* **Conclusion**: Tree-based methods are **naturally a better fit** for this cybersecurity incident classification task.

**5. Model Family Selection**

| **Aspect** | **Decision** |
| --- | --- |
| Best Baseline Model | Decision Tree Classifier |
| Best Model Family | **Tree-based Models** |
| Advanced Models to Try | Random Forest, XGBoost, LightGBM |
| Reason | Handle multi-class, imbalances, categorical data, complex patterns very well |

**6. Final Baseline Conclusion**

**Baseline established**.  
 **Decision Tree gives good results**.  
 **Next move**: Advanced tree-based ensemble models (Random Forest, XGBoost, LightGBM) to further improve generalization and performance.