

# Analysis of Results for Intent Classification Model

## Assignment # 2

### 1. Training Progress:

- The model was trained over **20 epochs**, achieving consistent improvement in both **accuracy** and **loss** metrics.
- Starting with an accuracy of **72.52%** and a loss of **1.0968** in the first epoch, the model rapidly converged, reaching an accuracy of **99.55%** and a loss of **0.0205** by the final epoch.
- Validation accuracy followed a similar trend, starting at **86.02%** and stabilizing around **98.11%**, indicating good generalization.

### 2. Test Set Performance:

- The model achieved a **test accuracy of 98.31%**, which confirms that it generalizes well to unseen data.
- This performance highlights the effectiveness of the architecture and training process in capturing the intent-specific patterns of the ATIS dataset.

### 3. Classification Report Highlights:

- **Precision, Recall, and F1-Score:**
  - Most classes, such as `atis_flight`, `atis_airfare`, and `atis_ground_service`, show **precision, recall, and F1-scores** close to **1.00**, indicating highly accurate predictions for these intents.
  - The `atis_flight` class, having the highest support (865 samples), achieved perfect metrics, reflecting the model's robustness on dominant classes.
  - Minor underperformance is observed in smaller classes, such as `atis_aircraft` and `atis_flight_time`, which have lower support (15 and 9 samples, respectively). These classes exhibit reduced metrics (e.g., `atis_flight_time` has an F1-score of 0.80), likely due to data imbalance.

### 4. Macro vs Weighted Averages:

- The **macro average** F1-score is **0.91**, slightly lower than the **weighted average** of **0.98**.

- This discrepancy suggests that smaller classes are contributing less to the overall performance due to their limited representation.

## 5. Model Generalization:

- The **low validation loss** (0.1039) and the close match between training and validation accuracies confirm that the model is not overfitting, a key indicator of a well-trained neural network.

## 6. Recommendations for Further Improvement:

- **Error Analysis:**
  - Perform detailed error analysis on misclassified samples to understand the confusion patterns between intents and refine the model accordingly.

---

## Conclusion

The model demonstrates excellent performance on the ATIS dataset, with a weighted F1-score of **0.98**, making it highly *reliable* for *intent classification tasks*. Addressing the minor *class imbalance issues* can further optimize the model's performance across all intents.