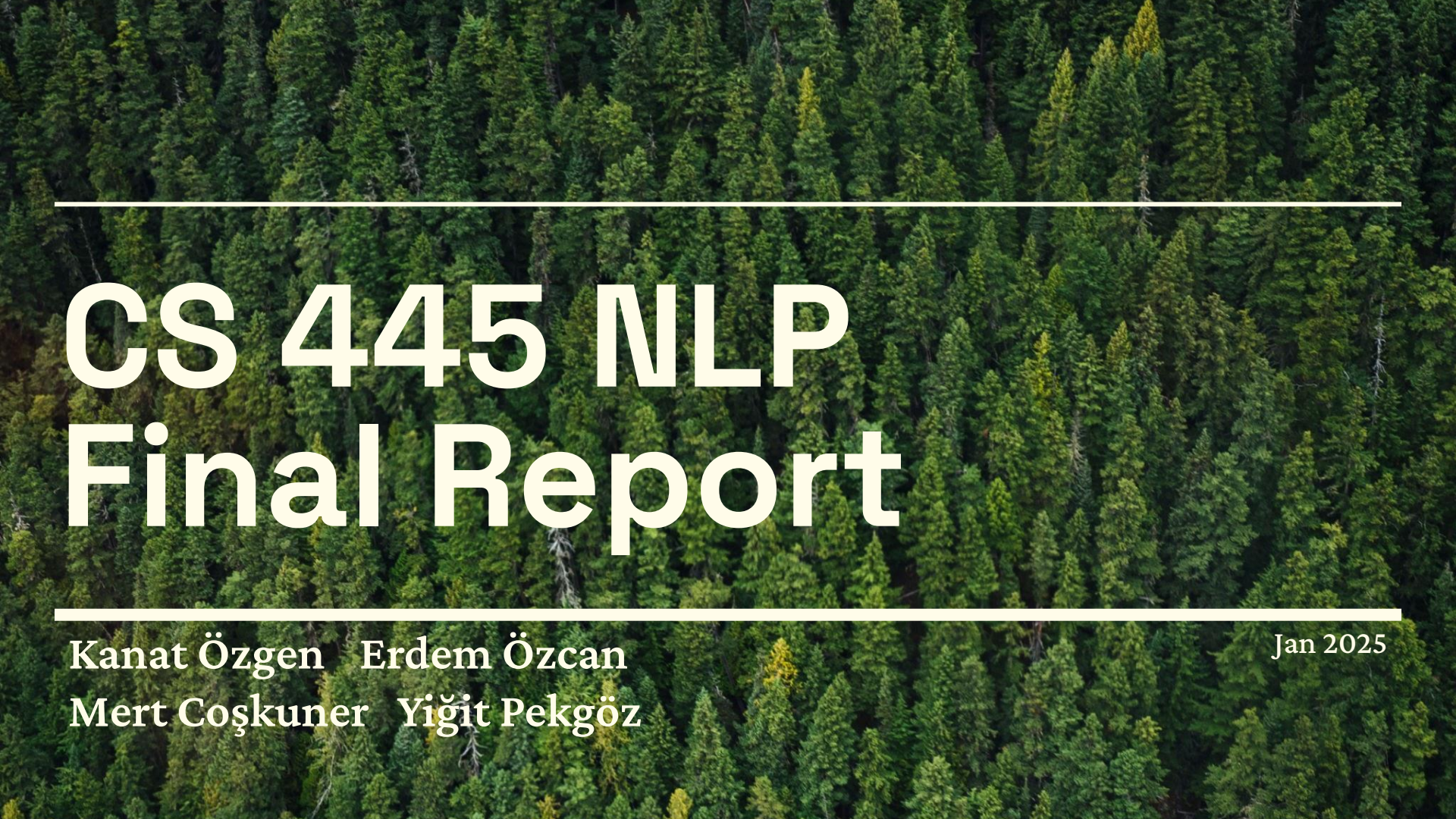


CS 445 NLP Final Report

Kanat Özgen Erdem Özcan
Mert Coşkuner Yiğit Pekgöz

Jan 2025

Section 2

Summary

Here is the summary of the things done in the scope of the NLP project. Steps involve implementing a simple baseline to check the waters and using a paper from NAACL for improving accuracy further by incorporating changes to the existing architecture.

Problem

Increasing the intent accuracy on the NAACL18 paper named Slot-Gated Modeling for Joint Slot Filling and Intent Prediction

- **Step 1**
  Create a baseline for the task using Naive Bayes and TF-IDF embeddings

- **Step 2**
  Using the know-how from the baseline and various resources, improve the accuracy of the paper.

Resolution

- Improved the accuracy from 93.6 to 95.7 with changes to the existing code.

- Used the existing code as a baseline and improved the architecture

Problems Faced

Version Mismatch	The paper that we are implementing was considerably older (from 2018) which required tensorflow version 1.4 and Python 3.5
Data Sparsity	Data is extremely sparse, since the atis_flight label dominates the rest of the classes.
BIO Tagging	BIO tagging helped us to get better attributes, but it became more difficult to work with it.
Lack of Knowledge	We needed to conduct an extensive amount of research to grasp what the paper was actually implementing.

Methodology

Dataset

- **We used the ATIS dataset, which stands for Airline Travel Information System dataset.**
- We chose this particular dataset in order to match up with the existing literature, since there is an abundance of research involving ATIS.

Paper

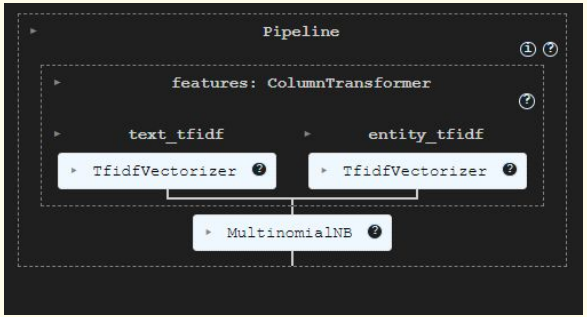
- **We improved the NAACL18 paper called Slot-Gated Modeling for Joint Slot Filling and Intent Prediction.**
- **You can access the paper with the following doi:** 10.18653/v1/N18-2118

Split Ratio

- **We used a train/dev/test split of 75/10/15 to train our model.**
- This was chosen by the paper authors, and we decided to proceed with this.

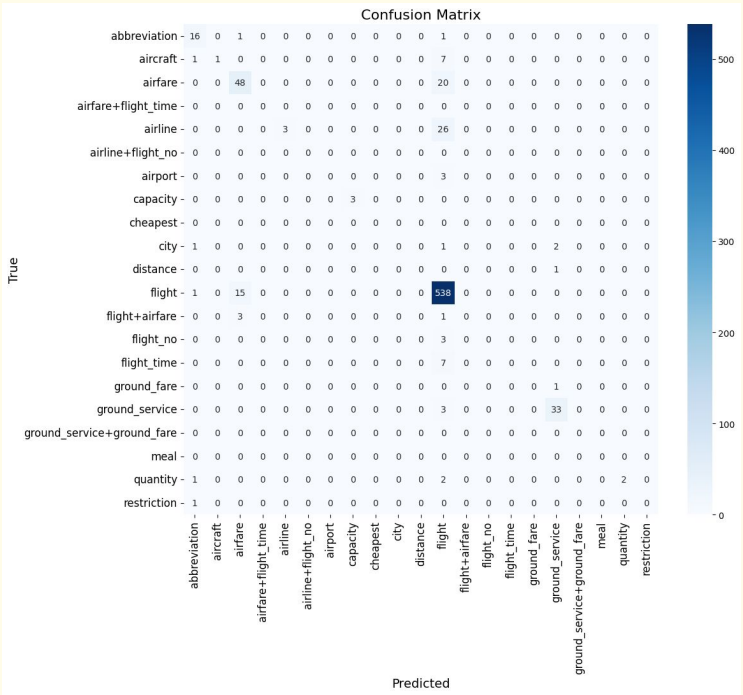
Section 3

The Baseline

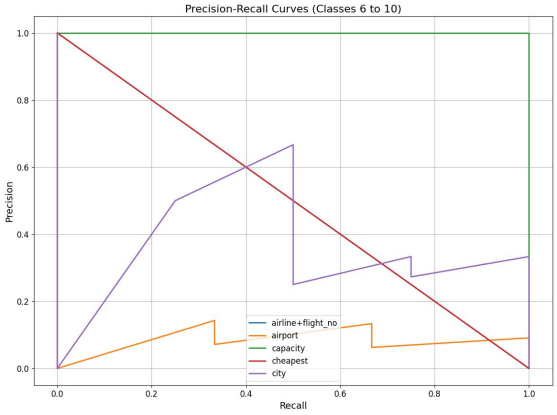


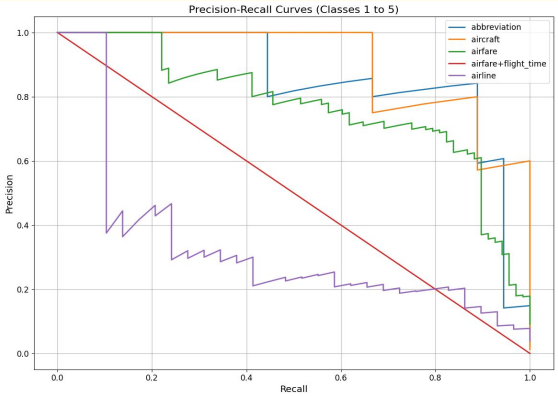
	precision	recall	f1-score	support
abbreviation	0.76	0.89	0.82	18
aircraft	1.00	0.11	0.20	9
aircraft+flight+flight_no	0.00	0.00	0.00	1
airfare	0.72	0.71	0.71	68
airline	1.00	0.10	0.19	29
airport	0.00	0.00	0.00	3
capacity	1.00	1.00	1.00	3
city	0.00	0.00	0.00	4
distance	0.00	0.00	0.00	1
flight	0.88	0.97	0.92	554
flight+airfare	0.00	0.00	0.00	4
flight_no	0.00	0.00	0.00	3
flight_time	0.00	0.00	0.00	7
ground_fare	0.00	0.00	0.00	1
ground_service	0.89	0.92	0.90	36
quantity	1.00	0.40	0.57	5
restriction	0.00	0.00	0.00	1
accuracy			0.86	747
macro avg	0.43	0.30	0.31	747
weighted avg	0.84	0.86	0.83	747

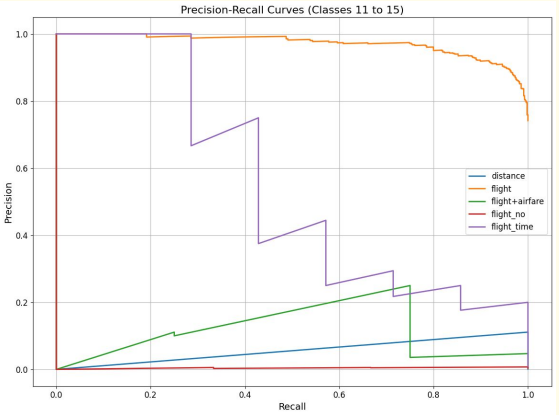
Overall Macro F1 Score: 0.313

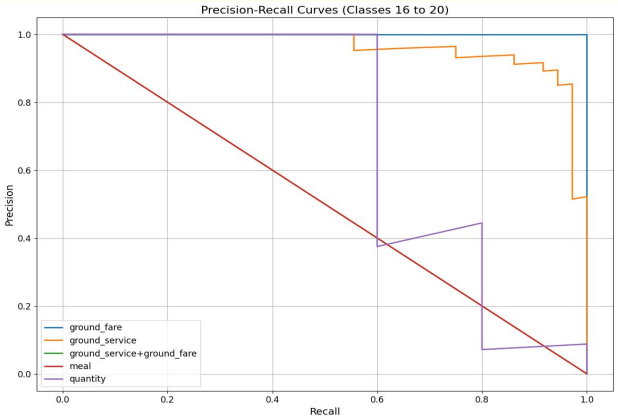


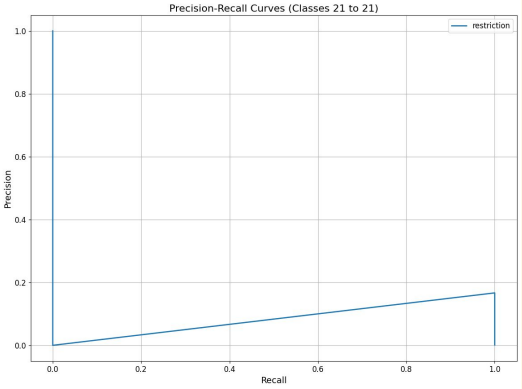
The Baseline contd.





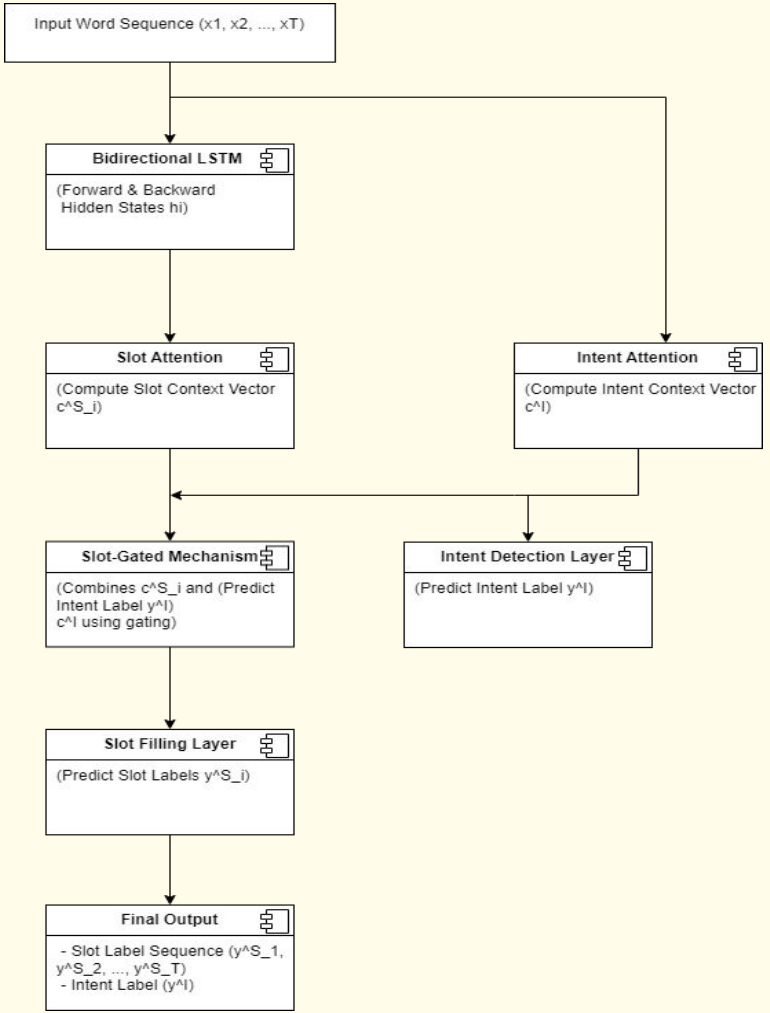






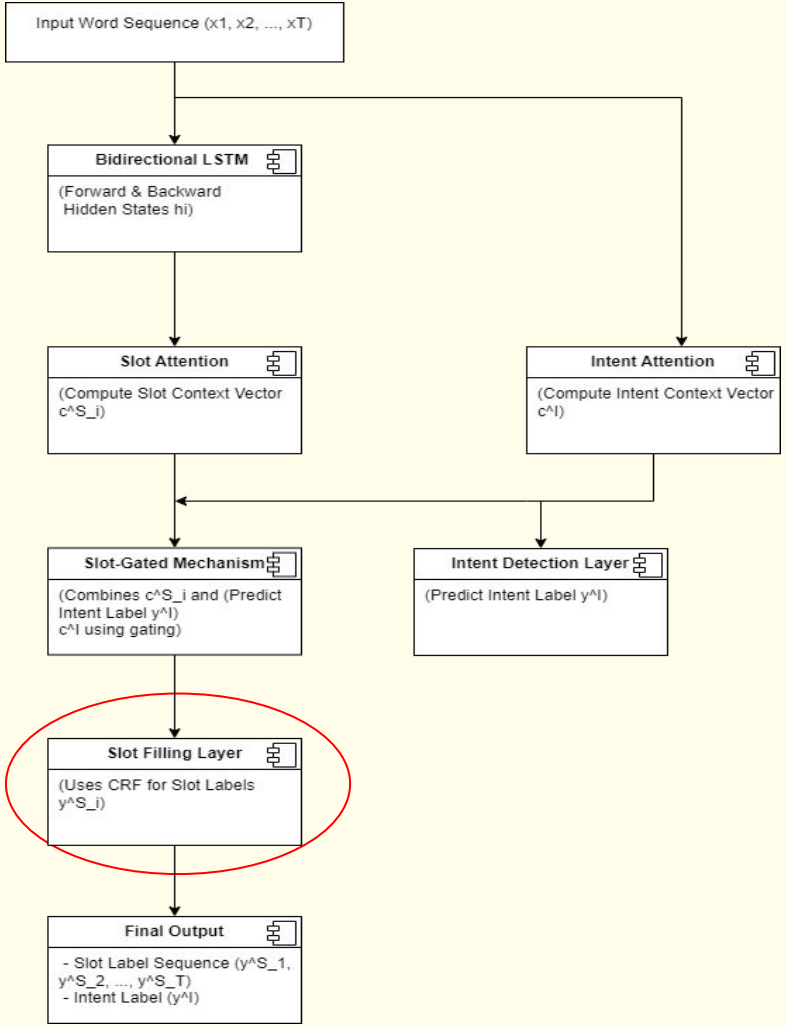
Section 3

Existing Architecture



Section 3

New Architecture



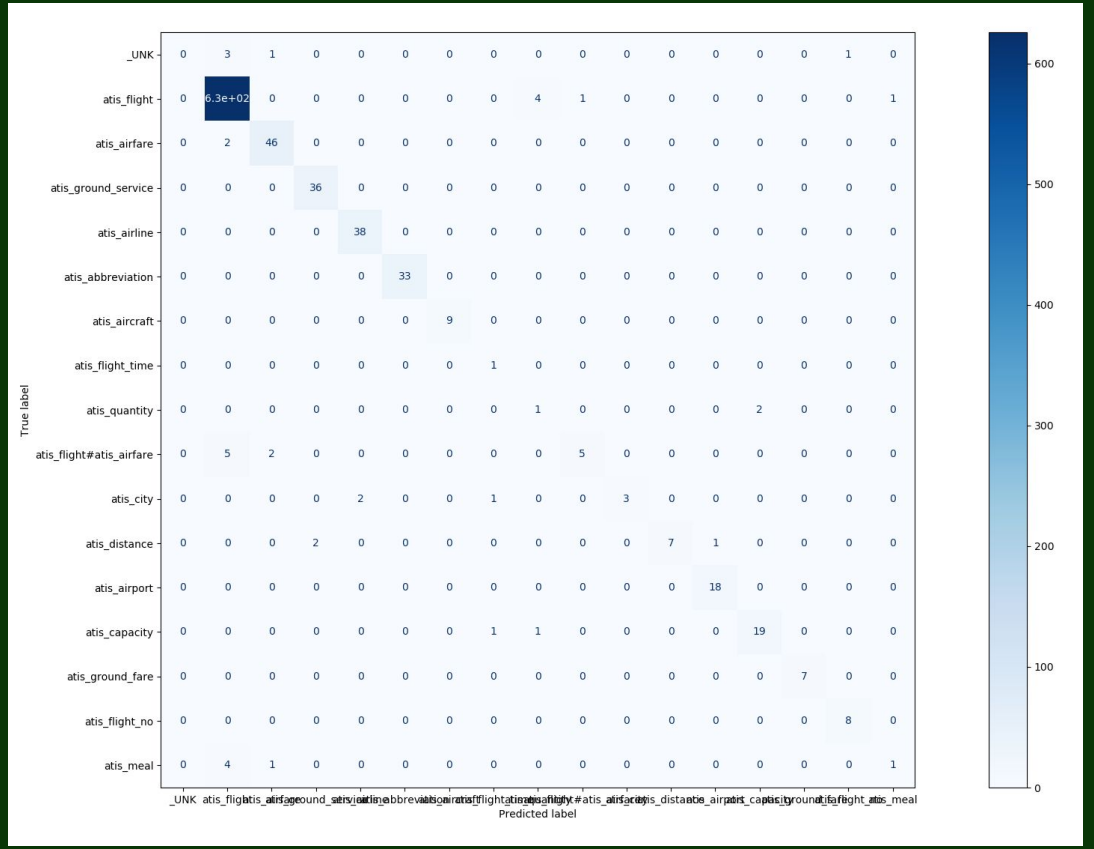
Results

	Accuracy	Macro F1-Measure
Naive Bayes Baseline	0.86	0.313
NAACL18 Paper	0.934	0.639
Our Proposed Architecture	0.957	0.75

Results contd.

	precision	recall	f1-score	support
_UNK	0.00	0.00	0.00	5
atis_flight	0.98	0.99	0.98	632
atis_airfare	0.92	0.96	0.94	48
atis_ground_service	0.95	1.00	0.97	36
atis_airline	0.95	1.00	0.97	38
atis_abbreviation	1.00	1.00	1.00	33
atis_aircraft	1.00	1.00	1.00	9
atis_flight_time	0.33	1.00	0.50	1
atis_quantity	0.17	0.33	0.22	3
atis_flight#atis_airfare	0.83	0.42	0.56	12
atis_city	1.00	0.50	0.67	6
atis_distance	1.00	0.70	0.82	10
atis_airport	0.95	1.00	0.97	18
atis_capacity	0.90	0.90	0.90	21
atis_ground_fare	1.00	1.00	1.00	7
atis_flight_no	0.89	1.00	0.94	8
atis_meal	0.50	0.17	0.25	6
accuracy			0.96	893
macro avg	0.79	0.76	0.75	893
weighted avg	0.96	0.96	0.96	893

```
2024-12-30 05:47:52,851 : INFO : Test intent_acc=96.08
2024-12-30 05:47:55,163 : INFO : Reached max epochs. Stop.
2024-12-30 05:47:55,163 : INFO : Training Completed.
Final Test Intent Accuracy: 96.08
```




Thank
You