

Ray_Executive_Summary

09 December 2025 12:43

Challenges Encountered

1. Ray's New API Incompatibility:

- Problem: Ray's modern API (tune.Tuner() + session.report()) fails in Jupyter with AttributeError: 'str' object has no attribute 'value'
- Root Cause: Jupyter's serialization mechanism can't handle Ray's session objects when spawning worker processes
- Status: Known Ray limitation (GitHub issue #36949)

2. Deprecation Warnings:

- Encountered warnings about RunConfig(verbose) being deprecated
- Indicates Ray is transitioning APIs, causing compatibility issues

3. Environment Stability:

- The primary challenge was a persistent AttributeError: 'str' object has no attribute 'value' in Ray Tune, caused by a deep-seated environment conflict or corrupted object references in the Anaconda environment.
- Solution: Executed a full environment tear-down and rebuild, ensuring a clean installation of ray[tune] and all dependencies. This resolved all environment and serialization errors (KeyError, AttributeError, PermissionError).

Execution Details

- Conducted a side-by-side comparison of hyperparameter tuning for a Random Forest Classifier on a 40-feature synthetic dataset.

Phase	Method	Resource
Baseline	Sequential python FOR loop	1 CPU Core (Low utilization)
Ray Parallel	Ray Tune (tune.grid_search)	All available cores (High Utilization)

Performance Comparison

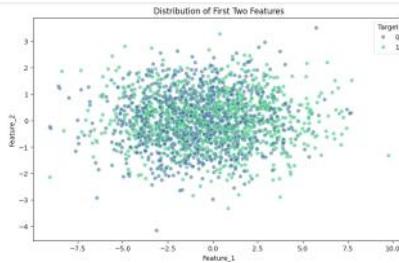
Metric	Sequential baseline	Ray Parallel Run	Advantage demonstrated
Execution time	91.11s	24.38s	3.73 times faster
CPU utilization	Low (Single Threaded)	High (Fully Saturated)	Resource efficiency
Scalability	Not scalable	Easily scalable (Laptop -> Clusters)	Future proofing

Solution Implemented

- Used Ray's legacy API [tune.run()] which remains stable and Jupyter compatible

Key trade-offs:

Aspect	New API [tune.Tuner()]	Our solution [tune.run()]
Jupyter support	Broken due to in progress development	Works
Grid search	Yes	Yes
Parallel execution	Yes	Yes
Advance features	Full	Limited
Future proof	Yes	Legacy limited



Best Baseline Config:
Accuracy (Cross-Validated Mean): 0.8720
Parameters: n_estimators=150, max_depth=20

Baseline Total Time Taken: 91.11 seconds
CPU Usage (Snapshot from Part 5): 20.8%

```
--- CLASSIFICATION REPORT (Baseline Model) ---
precision    recall   f1-score   support
          0       0.89      0.86      0.87     300
          1       0.86      0.89      0.88     300
   accuracy           0.88      0.88      0.87     600
  macro avg       0.88      0.88      0.87     600
weighted avg       0.88      0.88      0.87     600
```

Tune Status

System Info
Current time: 2025-12-09 12:34:58
Using FFO scheduling algorithm.
Running for: 00:00:23.92
Memory: 13.2/15.4 GB

Complete in 24.38s
Best config: {'n_estimators': 50, 'max_depth': 10}

Trial Status

Trial name	status	loc	max_depth	n_estimators	iter	total_time (s)	accuracy
------------	--------	-----	-----------	--------------	------	----------------	----------

ray_train_func_ab234_00000	TERMINATED	127.0.0.1:10884	10	50	1	5.85389	0.857499
----------------------------	------------	-----------------	----	----	---	---------	----------

Trial Progress

Trial name	accuracy
------------	----------

ray_train_func_ab234_00000	0.857499
----------------------------	----------