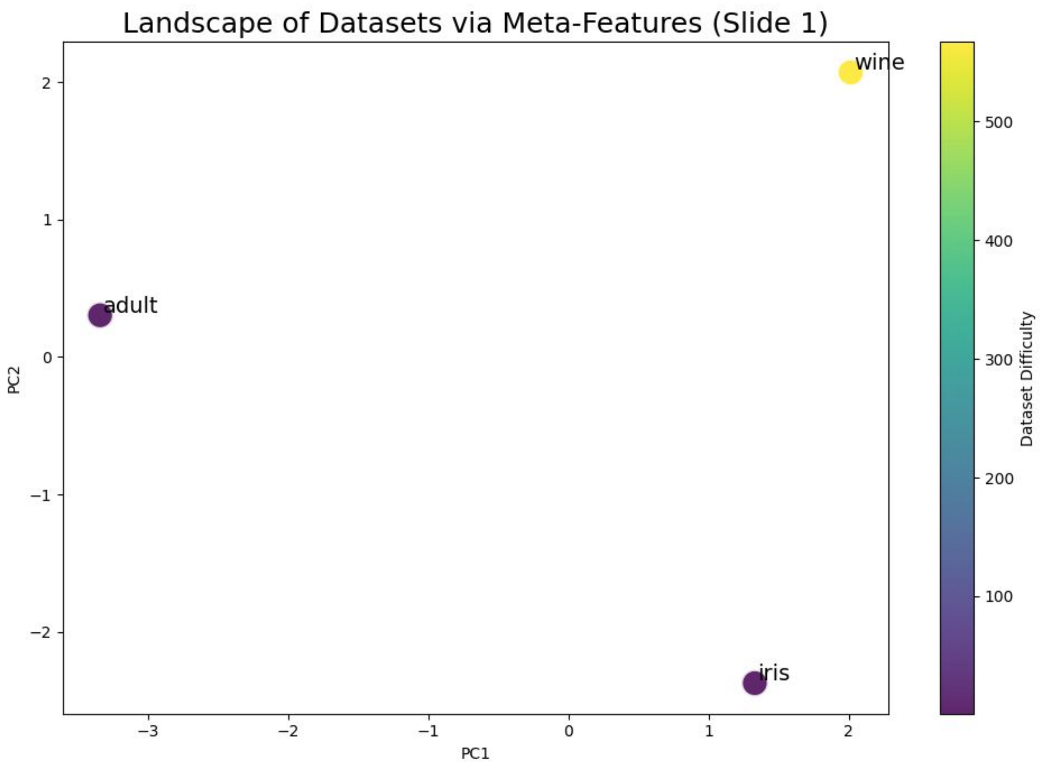
Lightweight AutoML for Efficient Model Selection

- Manual model selection is slow, subjective, and error-prone.
- Existing AutoML systems are powerful but computationally expensive.
- Need for efficient, explainable, hardware-friendly AutoML



Goal:

- Automate preprocessing, model selection, hyperparameter tuning.
- Support ensembles, warm starts, TinyNAS.
- Fast, reproducible, hardware-efficient system.

Methodology Overview

- 1. Preprocessing Module:

  - Schema inference, missing values, scaling, encoding.

- 2. Model Zoo:

  - Logistic, Ridge, Trees, RF, GBM, MLP, TinyNAS.

- 3. Hyperparameter Search:

  - Grid, Random, Bayesian Optimization.

- 4. Meta-Learning Warm Starts:

  - Meta-features, KNN similarity, warm priors.

- 5. Ensembles:

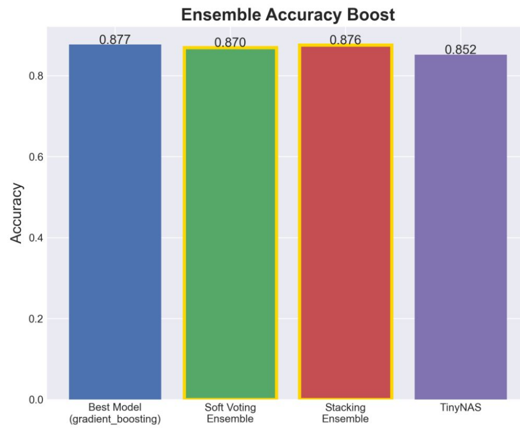
  - Soft Voting, Stacking.

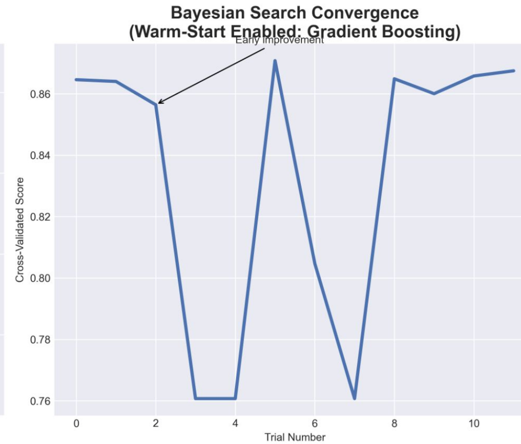
Results & Key Findings

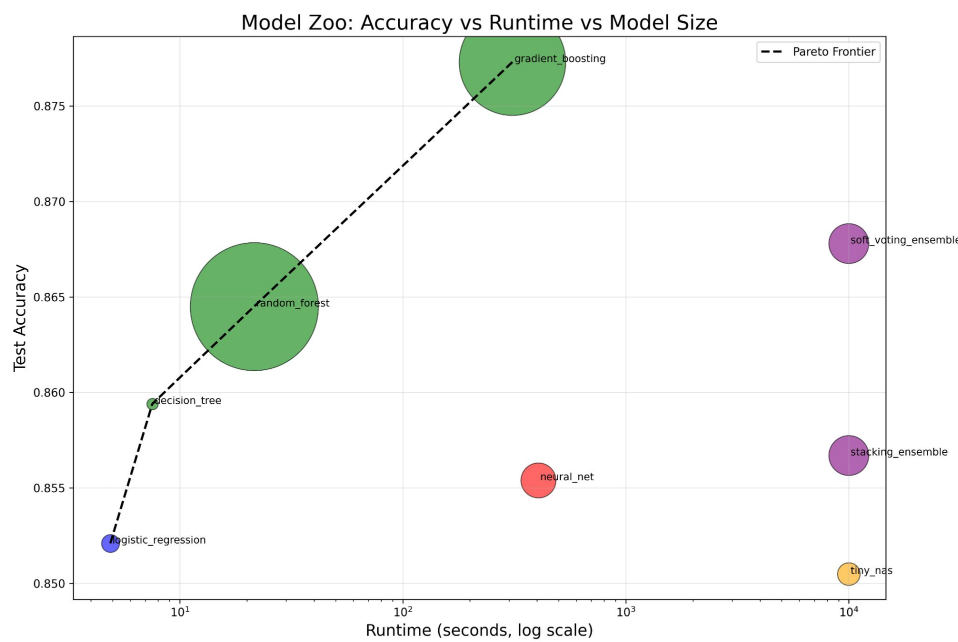
- Search Method Comparison:
  - Grid ~0.852
  - Random ~0.863
  - Bayesian 0.877

- Ensemble Accuracy:
  - Soft Voting: 0.867 - 0.870
  - Stacking: 0.874 - 0.876

- Warm Start Benefits:
  - 20–30% faster convergence.
  - More stable early exploration.







Discussion & Conclusion

- Key Takeaways:

  - Full modular AutoML system.

  - Integrated warm starts, ensembles, TinyNAS.

  - Strong performance & efficiency.

- Limitations:

  - Bayesian optimization slower.

  - Sizes of our datasets: NAS & Meta-learning dependon dataset size.

- Future Work:

  - Expanding evaluation metrics

  - Expand meta-learning DB.

  - Cost-aware optimization.

  - Faster NAS searches.

  - Cloud/distributed AutoML.