

Next-Gen Forecasting Pipeline

Andy Andikko

Introduction

Forecasting pipeline

Artifacts for further downstream applications



5 - ARTIFACTS

Making predictions from stacked model and using conformal quantile regression for guaranteed confidence intervals



4 - PREDICTION

Selecting and tuning of meta-learners on top of tuned base-learners



3 - STACKING

Leveraging Optuna and Dask



2 - HYPERPARAMETER OPTIMISATION

Preparing data from data engineering pipeline for modeling



1 - DATA PREPARATION

Introduction

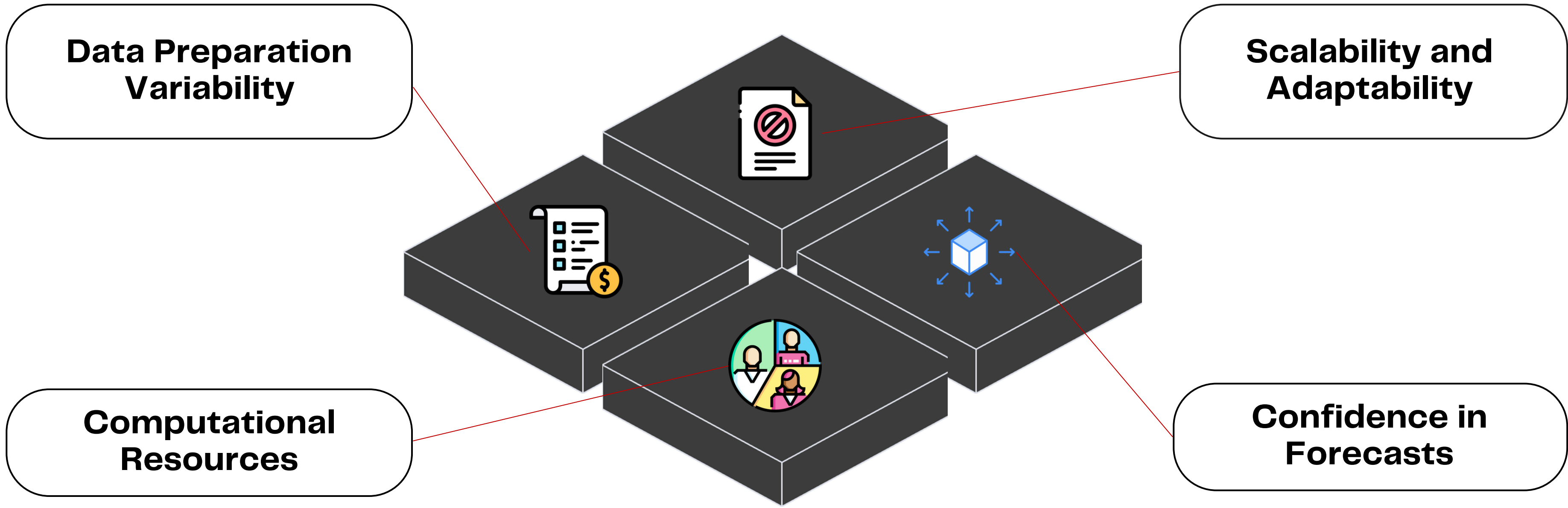
Challenges

Solutions

Conclusion

The Challenge

A next-gen forecasting pipeline must be optimized for parallelization and cloud computing



Introduction	Challenges	Solutions	Conclusion
--------------	------------	-----------	------------

The Solution

A unified interface built on the Darts framework



Support across different ML packages

- Develop a standardized library of functions for Darts to unify data preparation, training, optimization, and prediction across multiple machine learning packages.
- This standardization enhances computational efficiency and facilitates parallelization.

Integration with Optuna and Dask

- Integrate Dask for parallel execution of validation and computationally intensive tasks within the unified interface.
- Custom Optuna functions to streamline hyperparameter optimization across models leveraging the unified interface, enhancing overall efficiency.

Simplified stacking and HPO of Meta learner

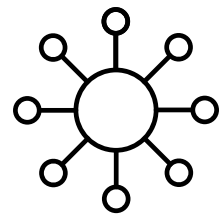
- Reusable custom functions utilised on the meta learners and stacking processes.
- Employing tuned meta learners with automated best selection markedly boosts prediction accuracy and reduced confidence intervals

Conformalised Quantile Regression

- Utilize MAPIE, a model-agnostic Python package, to create prediction intervals for any regression model, enhancing prediction certainty.
- MAPIE's flexibility reduces CQR complexity.
- Employ Conformalized Quantile Regression (CQR) for statistically guaranteed coverage, bolstering confidence in business decision-making.

Introduction	Challenges	Solutions	Conclusion
--------------	------------	-----------	------------

Progress



Custom functions and libraries simplify the prediction workflow, organizing and facilitating the integration of current and future models. This approach streamlines scaling to cloud environments and boosts computational efficiency.



Introducing stacking and CQR develops more accurate and reliable forecasting models

Next Steps



Deployment to AWS Sagemaker (GPU) + Fargate (CPU)



Using Metaflow for parallel sagemaker instances