FLYR

Mid-term decision making in airline cargo using machine learning

Abhinav Garg
Naman Shukla

AGIFORS 08 May 2024

Abhinav Garg
Senior Applied Scientist

Abhinav is a senior applied scientist at FLYR, where he leads the development of AI algorithms for airline revenue management, with notable achievements in building modern forecasting and pricing systems. He has 7 years of experience building end-to-end ML systems at startups and large corporations. He holds a master's degree in industrial engineering from University of Illinois Urbana-Champaign and bachelor's degree from Indian Institute of Technology Roorkee.

Naman Shukla
Product Manager, Applied Science

Naman is product manager of data science at FLYR, where he leads a team of data scientists and engineers to build products for revenue management on both passenger and cargo domains. Naman has expertise in research and development of end-to-end machine learning solutions that power RM systems for airlines and travel providers.

Extended cargo team

Maarten Wormer
Director, Cargo

Talal Mufti
Manager, Applied Science

Table of contents

Background and context

Mid-term decisions drive a significant share of air cargo revenues…

Relevance of air cargo to airlines

- Air cargo represents 35% (by value) and 1% (by volume) of global trade, due to the higher value of goods transported by air
- Air cargo represents between 12% (2019, 2024F) and 40% (2021 COVID peak) of airline revenues
- The B2B air cargo industry requires materially different decision-making from B2C passenger sales

Mid-term decision-making in air cargo

- Unlike ticket sales, cargo is often-times booked as part of a longer-term customer commitment
- In 2023Q4, 45% of cargo capacity contracts was for a 6+ month duration[1]
- Longer-term contracts are usually negotiated around IATA season changes, to align to the new airline schedules

The importance of making the right decision

- Material revenue is at stake. Airlines trade off longer-term commitments vs shorter-term market opportunities
- Customers may request multiple commitments simultaneously
- Customer relationships are strong and rejections may impact opportunities during future seasons

…and airlines struggle to make confident decisions due to the lack of a **trustworthy mid-term air cargo forecast**

[1] https://www.aircargonews.net/airlines/air-cargo-eyes-a-year-of-uncertainty-and-opportunity

FLYR

From objective to solution

Mid-term forecasting requirements to design from data science perspective

Understanding the Objective

Solution requirements

- Ability to forecast across long horizon i.e > 6 months
- Ability to consider network effects
- Ability to factor in dynamic contexts
- Should be a “useable” forecast i.e. it should be trustworthy and analyst should be able to use it confidently

Principled first approach

Develop a machine learning solution that forecasts demand in terms of gross weight on Origin and Destination (market) level by week.

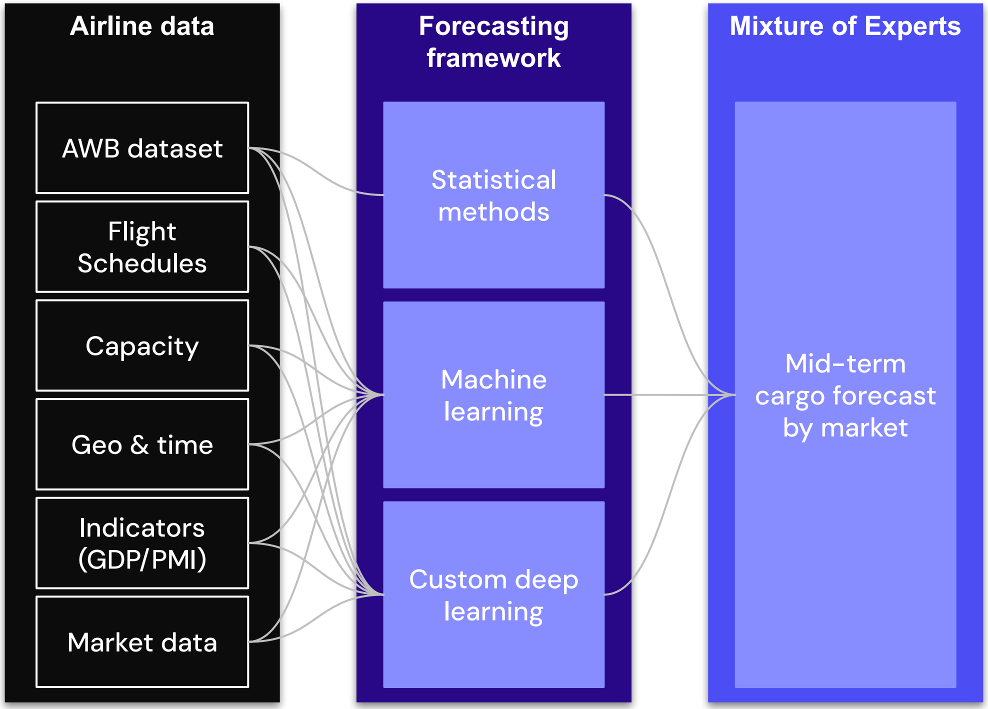
Solution blueprint

- Collect relevant data and context from a cargo carrier
- Leverage deep learning to learn from the context
- Build a robust framework that can be used reliably by cargo RM team
- Evaluate the performance of forecast on key metrics

FLYR

Scientific approach

Mid-term cargo forecasting framework



Mixture of Experts

Approach	Multiple expert networks (learners) are used to divide a problem space into homogeneous regions. Used in most winning models at Kaggle and recent open source LLM's open orca and platypus*
Pros	No "one fits all" approach. It fits for low frequency inference. It works well with a large number of models
Cons	Computationally intensive and difficult to scale. Creates productionisation challenges

* Platypus: Quick, Cheap, and Powerful Refinement of LLMs

FLYR

Statistical methods applied

Croston Classic

Addresses intermittent demand patterns by separately forecasting non-zero demand occurrences and the time intervals between them

Historic Average

Average of all historical observations

Auto ETS

Best ETS (Error, Trend, Seasonality) model (additive/multiplicative) selected using an information criterion

Window Average

Rolling average of past "x" observations

Dynamic Optimized Theta

Decomposes the time-series into 2 lines to learn the short-term and long-term trend

Simple Exponential Smoothing

Exponential rolling average of time-series

Holt Winters

Exponential smoothing of trend and seasonality in the time-series

Seasonal Naive

Forecast the actual value of the same day in the previous season

FLYR

ML & Custom Deep Learning techniques

A Multilayer perceptron

A type of neural network composed of multiple layers of interconnected nodes, used for various machine learning tasks, especially in pattern recognition.

B Long-short term memory (RNN)

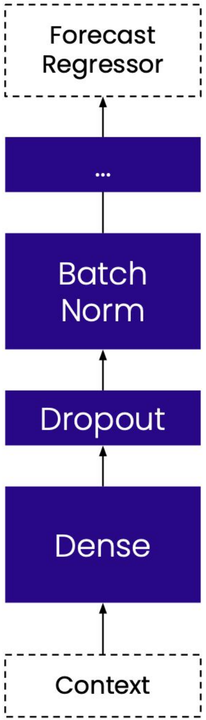
A specialized type of recurrent neural network designed to effectively capture and process sequential data, often used for tasks involving time-series analysis and natural language processing.

C Transformers

Revolutionary architecture for natural language processing and machine translation, using self-attention to enhance parallelized processing and excel at tasks with extensive context.

D Meta learning

Machine learning subfield concerned with developing algorithms that enable models to learn and adapt across different tasks or domains to improve generalization and efficiency

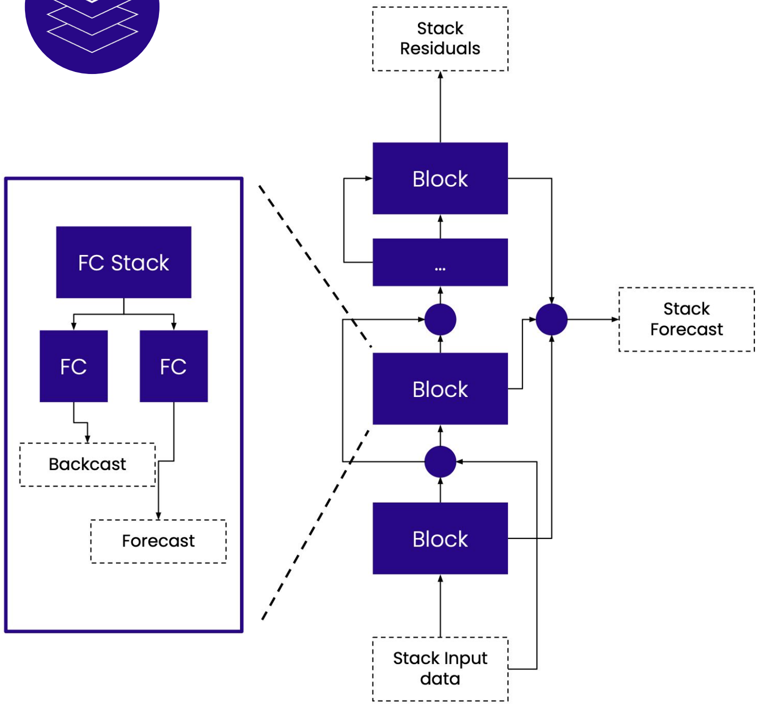


Multilayer Perceptron

- This architecture is based out of fully connected layers with the idea building a model based on i.i.d (independent and identically distributed) data based on provided context.
- Primary hypothesis for implementing multilayer perceptron was to **capture the trend based on context** not necessarily driven by time series.

Key characteristics

- Address skewed data using suitable outlier removal techniques like the interquartile range (IQR) method.
- Explored the potential benefits of incorporating lagged variables, but acknowledge the limitation of insufficient data for creating meaningful weekly or monthly lags.
- Note that capping features, despite intention, detrimentally impacts a basic neural network model; investigate reasons for this outcome.

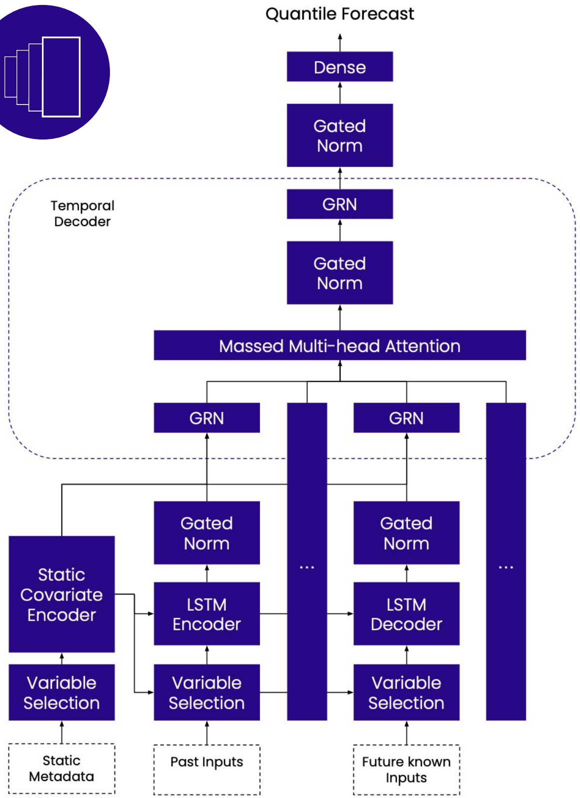


Long-Short Term Memory (RNN)

- This architecture design methodology relies on a few key principles.
  a. First, the base architecture should be **simple and generic**, yet expressive (deep).
  b. Second, the architecture should not rely on timeseries-specific feature engineering or input scaling. These prerequisites let us explore the potential of pure DL architecture in TS forecasting.
  c. Finally, as a prerequisite to explore interpretability, the architecture should be extendable towards making its outputs human interpretable.

Key characteristics

- Performs better when patterns are seasonal and somewhat context-driven
- Struggles when data is sparse
- OD context in the same model help learn patterns across ODs

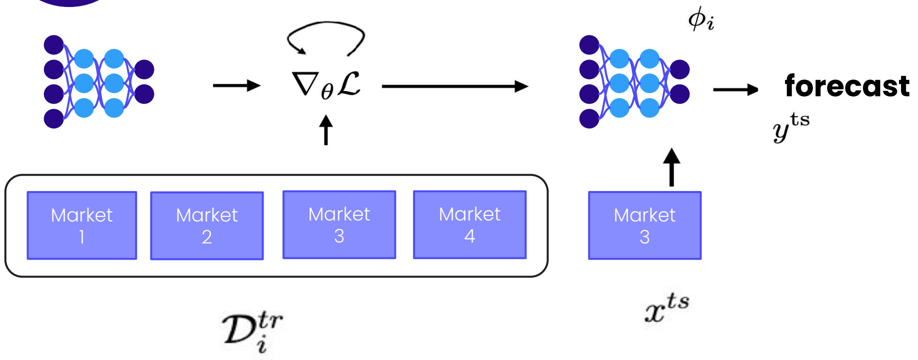


Transformers

- This architecture inputs static metadata, time-varying past inputs and time varying a priori known future inputs
- Time-dependent processing is based on LSTMs for local processing, and multi-head attention for integrating information from any time step
- Gated Residual Network blocks enable **efficient information flow with skip connections** and gating layers

Key characteristics

- Able to selectively distinguish between important and noisy features
- Performs better on sparse data and context-driven markets
- Computationally expensive to train
- High-latency during inference



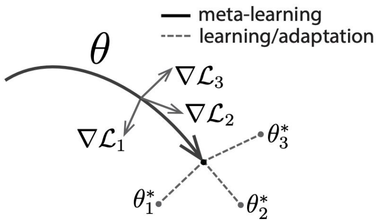
Meta-Learning

- Multi-task based machine learning approach to jointly minimize each individual market's forecast error and network-wide global error
- Enhanced model training routine to generate a forecast with the same model architecture

Key characteristics

- Generalizes well with newer tasks
- Able to provide zero-shot forecasts
- Computationally expensive to train

Optimization-Based Adaptation



pre-trained parameters

Fine-tuning [test-time] $\phi \leftarrow \theta - \alpha \nabla_\theta \mathcal{L}(\theta, \mathcal{D}^{\text{tr}})$

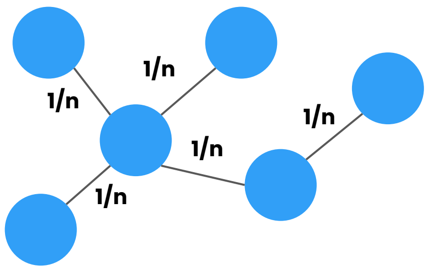
training data for new task

Meta-learning $\min\limits_{\theta} \sum\limits_{\text{task } i} \mathcal{L}(\theta - \alpha \nabla_\theta \mathcal{L}(\theta, \mathcal{D}_i^{\text{tr}}), \mathcal{D}_i^{\text{ts}})$

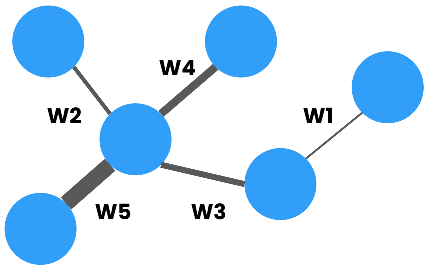
FLYR

Key metrics for accuracy evaluation

Metrics	Formulation
Root mean square error (RMSE)	biased towards markets with high values
Normalised RMSE (nRMSE)	remove the bias on market level by first normalising with average actuals and then taking the overall average [Figure (a)]
Weighted nRMSE	take the weighted average instead of overall to measure performance with proportionate market relevance on network [Figure (b)]
Mean Bias Error (MBE)	mean error to measure average deviation



(a) nRMSE: Equal weights for each market



(b) Weighted nRMSE: Weighted edges for each market

Results and quality review

Forecast performance results

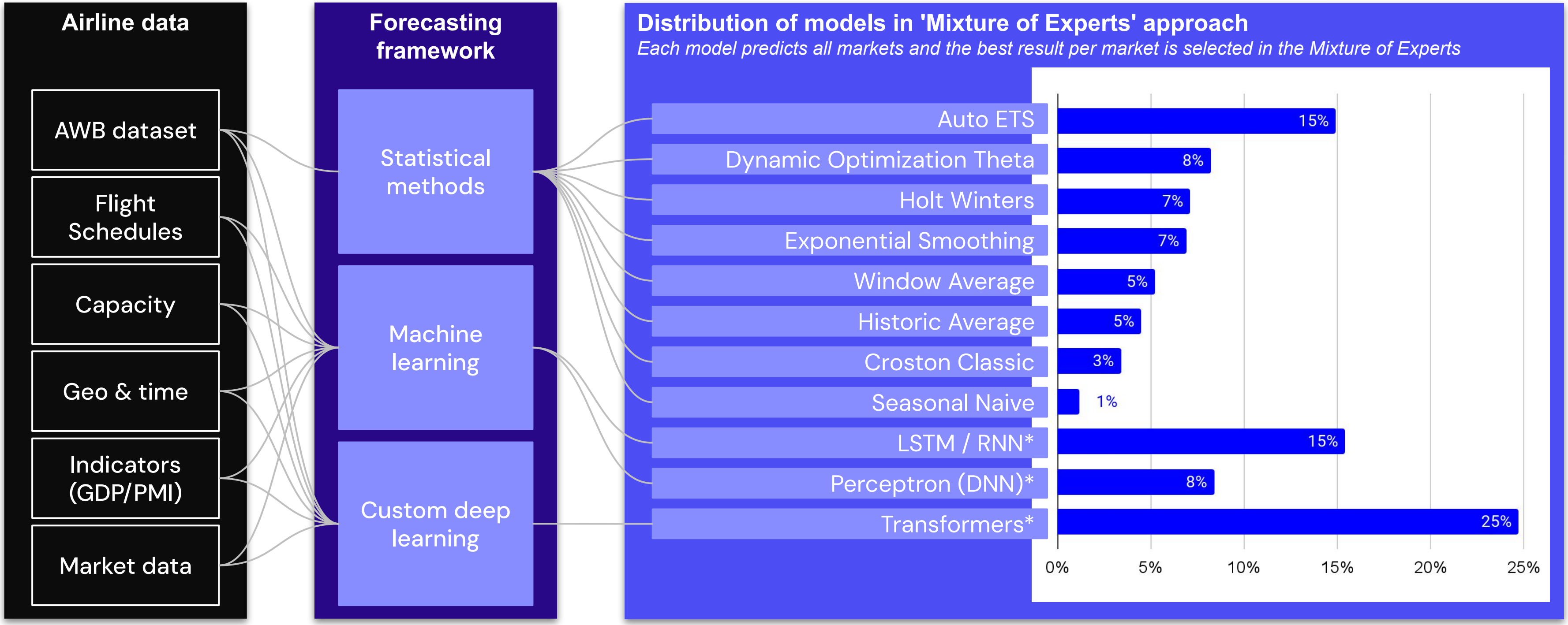
Market bracket	nRMSE (by week)	weighted nRMSE (by week)	% ML won	ML vs benchmark improvement	Overall RMSE
Significant markets	1.04	1 0.52	3 49%	4 13.2%	7,565
Top 100 markets	0.34	0.30	42%	8.7%	23,840
Market 101-500	0.56	2 0.51	49%	5.6%	10,099
Market 501-1000	0.75	0.69	46%	4.5%	5,141
Market 1001-end	1.32	0.99	50%	18.3%	2,688

Overall forecast quality review

1. Full IATA Summer 2023 season forecast on Significant markets (90% of revenue) at nRMSE=0.52 accuracy (weighted)
2. Performance on top 500 markets (with 2/3 of cargo revenues) significantly better than network average
3. ML-models provide best results for 49% of markets (increases as we go down the markets bracket)
4. When ML-models "win", they outperform statistical models by 13% on average

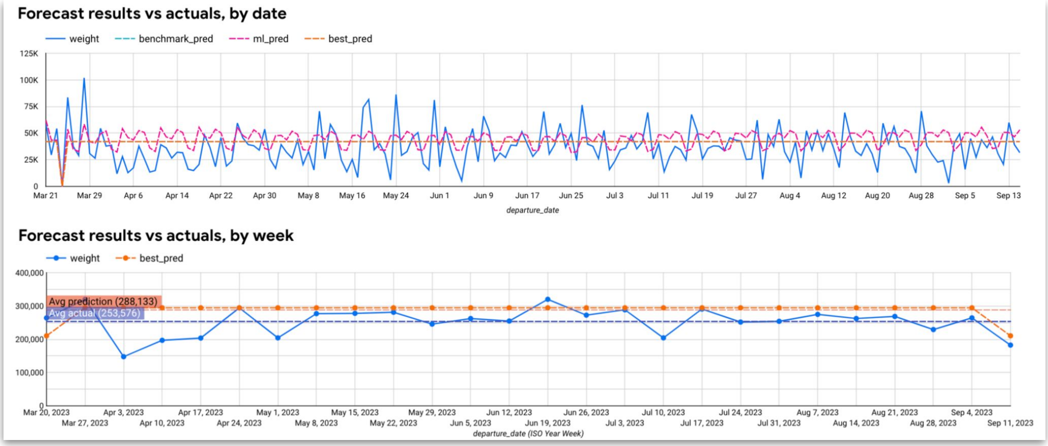
FLYR

Mid-term cargo forecasting framework



* These models are trained using meta learning approach

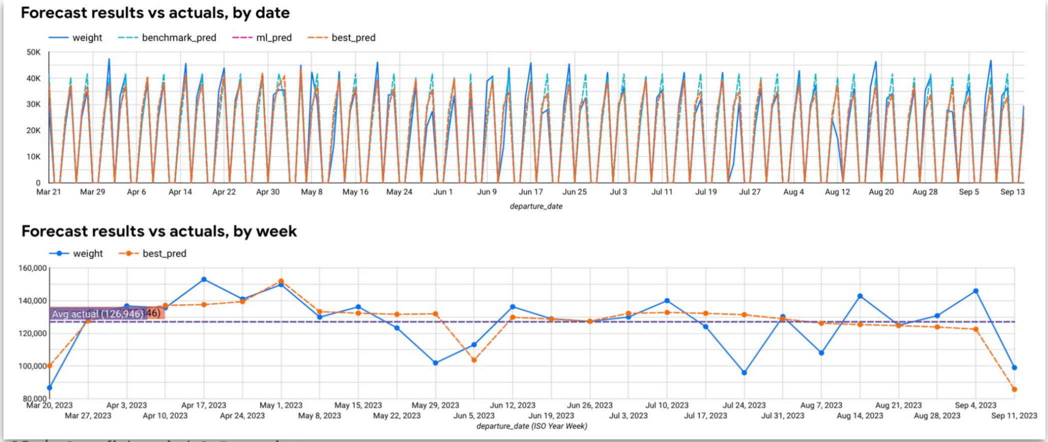
FLYR



#1 Market

nRMSE = 0.20

MBE = 36,890



#23 Market

nRMSE = 0.10

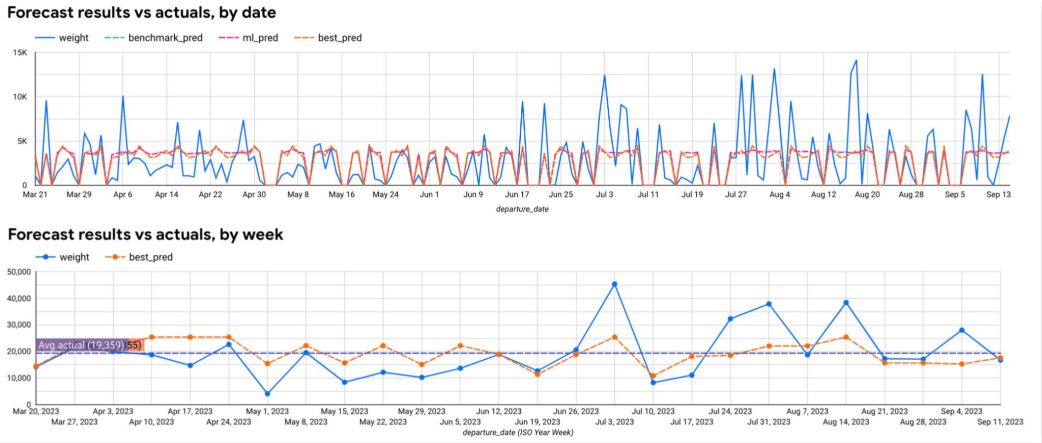
MBE = 509

Validation approach

For each market,

- Day-level forecasts generated from each model
- Model with minimum val RMSE selected as best performing
- Best performing model forecasts aggregated to week
- Metric calculation for comparison across markets

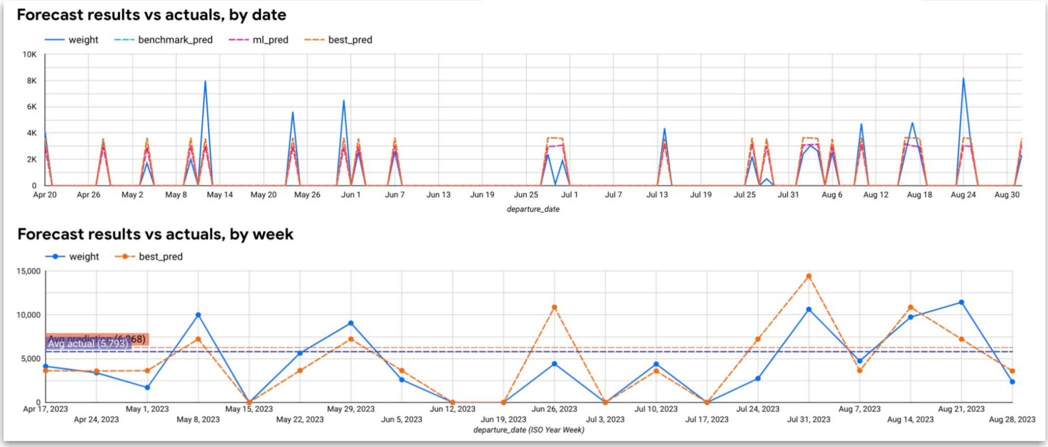
FLYR



#109 Market

nRMSE = 0.43

MBE = -179



#651 Market

nRMSE = 0.47

MBE = 475

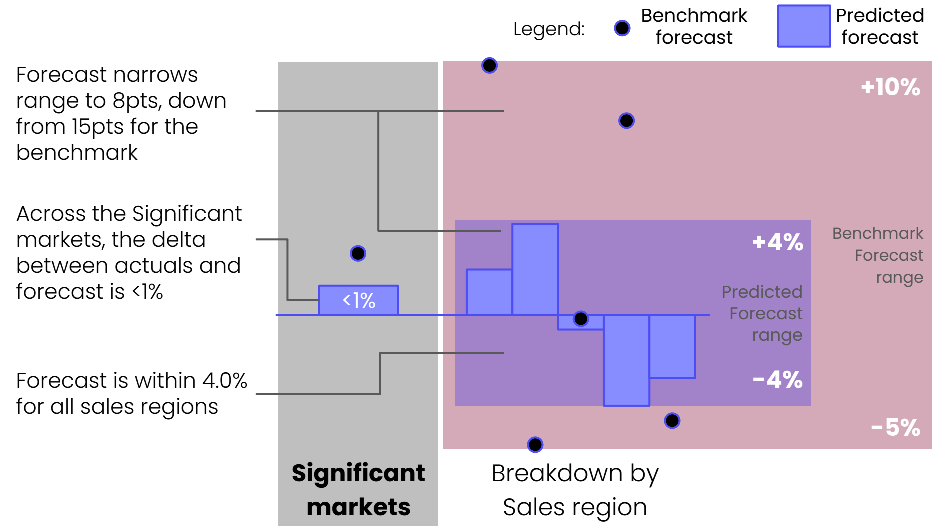
Key observations

- Mixture of experts approach is necessary to capture the diverse behavior across a wide-range of markets
- nRMSE and MBE provide useful tools to gauge forecast quality

Conclusion and implications

Accurate expectations for market-specific & aggregated performance

Forecasted cargo performance (April-May 2023)
% difference between forecasted tonnes and actuals



Relevance to mid-term decision-making

- Aggregated results (by country, region, network, month, quarter) drive sales and budgetary expectations (and targets)
- Market-specific forecasts support allotment vs free sales decisions and allotment acceptance guidance
- Recreating forecasts for different forward-looking schedules inform attractiveness of future flight (change)s for network planning

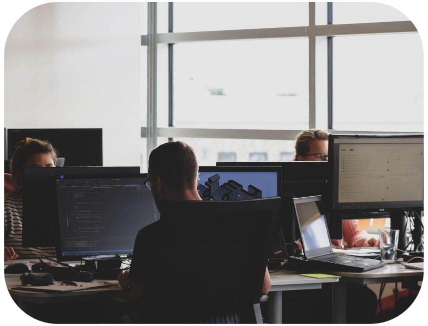
Conclusions & implications

Value to mid-term cargo planning

Forecast results are foundational for sales & budgeting and mid-term decision support including allotment acceptance

Framework is suitable for continuous forecast refreshes and context expansion

Scientific approach



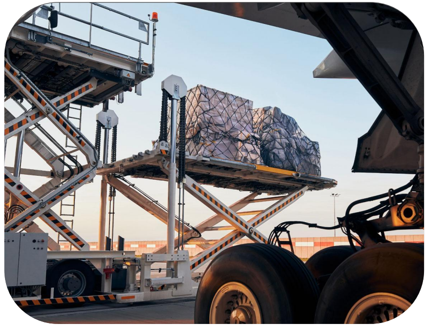
Accurate forecast results can be achieved with data sources relevant to air cargo, typically readily available at an airline, and the application of a 'Mixture of Experts' approach

Results & quality review

Quality is consistent across geographies and across the five-month forecast horizon

ML-models outperform on smaller markets, statistical best on larger/stable markets

Tangible next steps

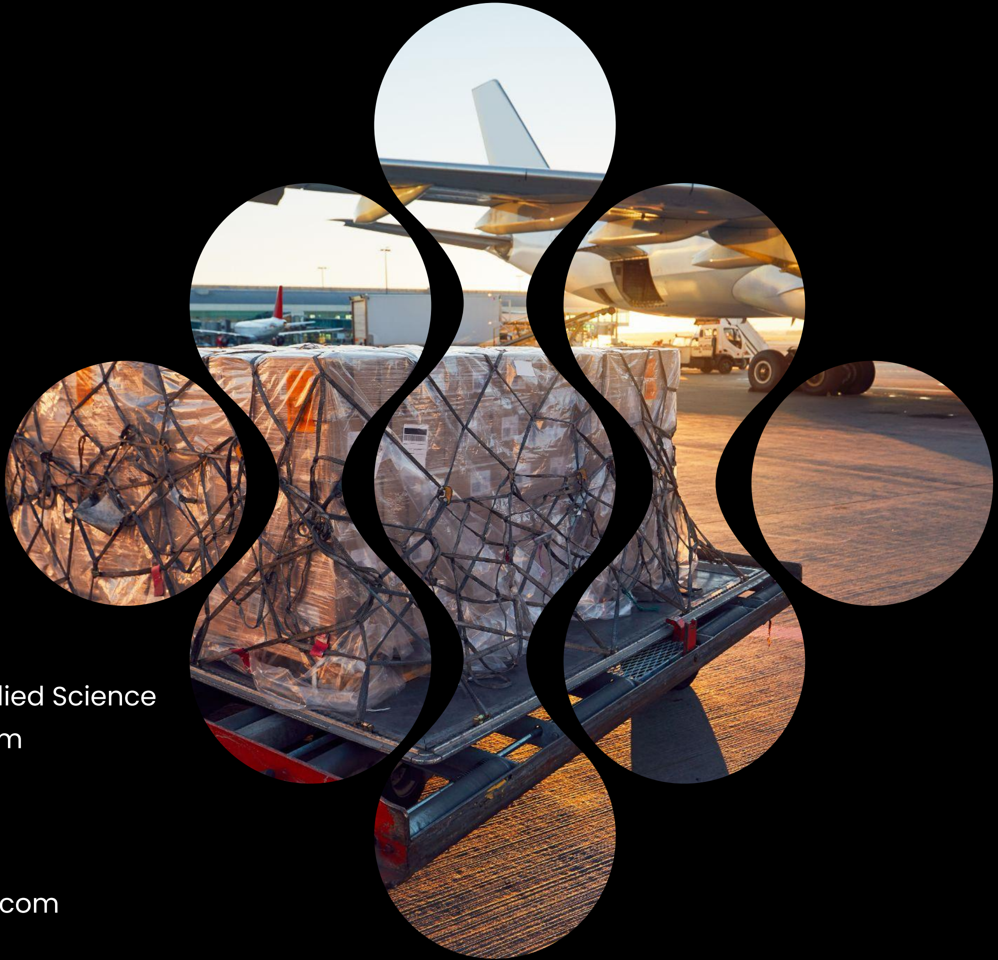


Expansion into revenue and chargeable weight outputs

Further refinement of model features from current context and targeted infusion of further model context

FLYR

FLYR



Thank you

Abhinav Garg
Senior Applied Scientist
abhinav.garg@flyr.com
+1 217-904-2848

Naman Shukla
Product Manager, Applied Science
naman.shukla@flyr.com
+1 646-267-9093

Talal Mufti
Manager, Applied Science
talal.mufti@flyr.com
+971 50-595-5736

Maarten Wormer
Director, Cargo
maarten.wormer@flyr.com
+31 638 455 876