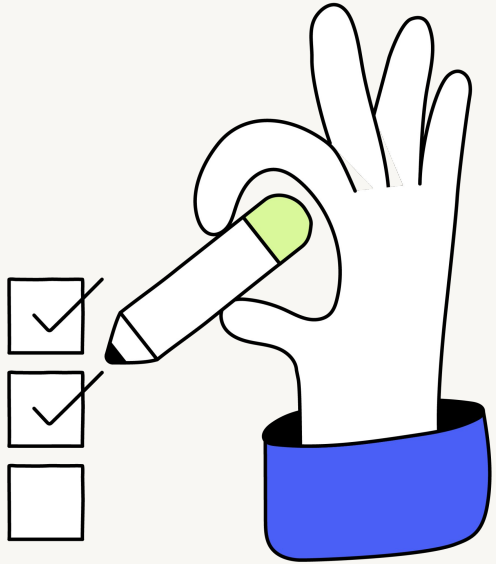
Module 8

Workshop 1: Evaluating Model Performance with Ensemble Methods

Icebreaker

Share one instance where accurate predictive modeling influenced a business decision you've seen or read about.

Drop your answers in the chat.



Module 8

Workshop 1: Evaluating Model Performance with Ensemble Methods

Agenda

- **Review:** Recap key concepts from async units 1 and 2.
- **Demo:** A guided walkthrough of:
  - How to prepare and clean your data
  - Training ensemble classifiers such as Random Forest, Gradient Boosting, XGBoost
  - Applying the hyperparameter optimisation techniques GridSearchCV and RandomizedSearchCV
- **Practice:** Hands-on application of evaluating a models performance with the ensemble methods shown.
- **Closing:** Key takeaways and next steps.

Learning objectives

- **Implement and interpret ensemble methods** (Random Forest, Gradient Boosting, XGBoost) for classification tasks, identifying their relative strengths in improving predictive accuracy.
- **Evaluate ensemble model performance** rigorously using metrics including accuracy, precision, recall, and confusion matrices, effectively communicating results.
- **Apply hyperparameter tuning techniques** (GridSearchCV, RandomizedSearchCV) to optimise ensemble classifiers for specific business contexts.
- **Make data-driven recommendations** based on ensemble model evaluations, effectively translating technical results into actionable business strategies, specifically in customer-retention scenarios.

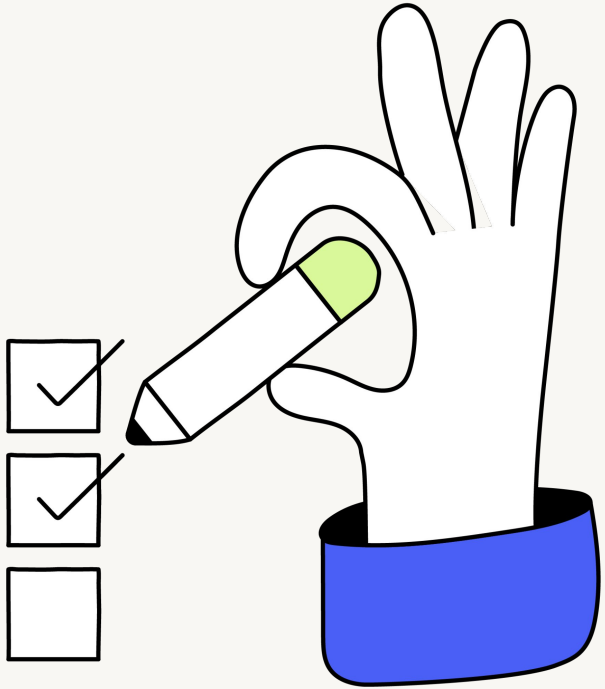
Async review: Units 1, 2, and 3

Go to My Multiverse, Module 8
Workshop 1

Navigate to the **“Async review”** page

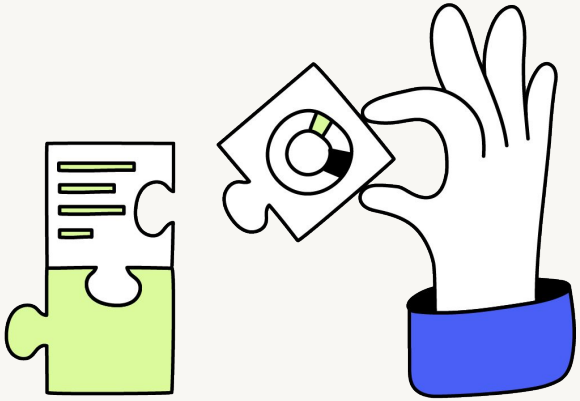
Review the summary of Units 1 and 2, then test your knowledge with the quiz:

Model Evaluation checkpoint



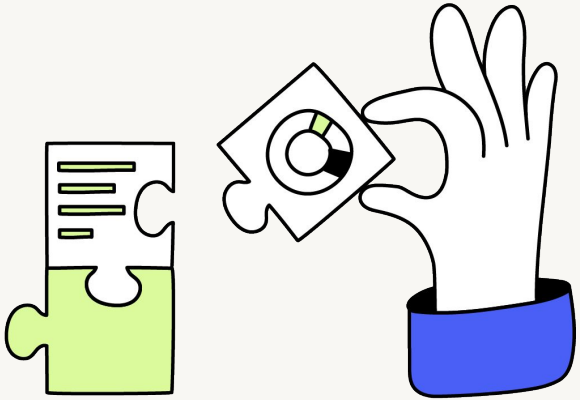
Performance metrics selection

- **Compass vs. Ruler** – Classification (Accuracy, Precision, Recall, F1) steers categorical hits; Regression (MAE, MSE, RMSE, R^2) gauges continuous misses.
- **Single-View Snapshot** – F1-score balances classification; R^2 reveals regression explainability.



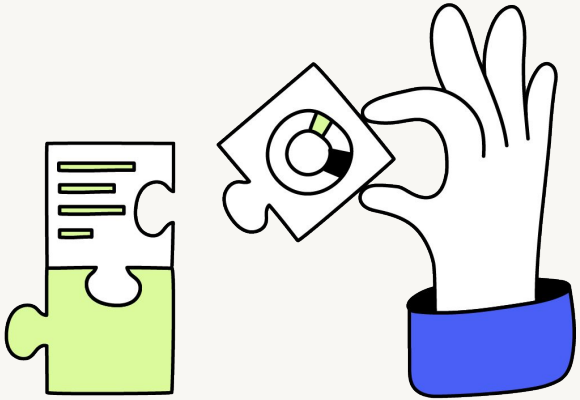
Business alignment in metric selection

- **False-positive phobia** – When too many false alarms erode trust, dial up precision.
- **False-negative fear** – If missing a hit is disaster, crank recall to catch every signal.
- **Balanced battleground** – When both misfires matter, wield the F1-score as your unified gauge.



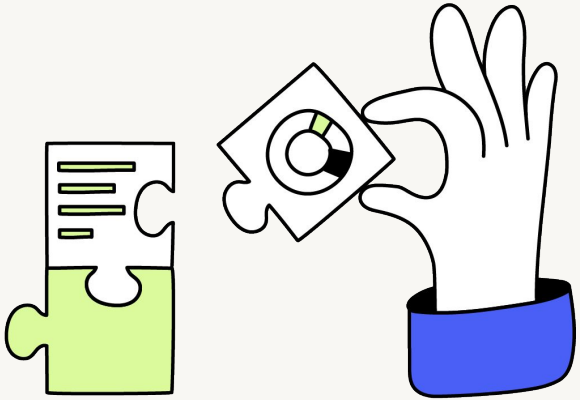
Cross-validation techniques

- Partition data into k folds, training on k–1 and testing on the hold-out fold to average out variance.
- Preserve class proportions in each fold with Stratified Cross-Validation to avoid bias in imbalanced classification.
- Aggregating results across folds delivers more robust, unbiased performance estimates than a single train/test split.
- Use standard K-Fold for fairly uniform data; opt for stratification when class imbalance might skew your evaluation.



Error analysis methods

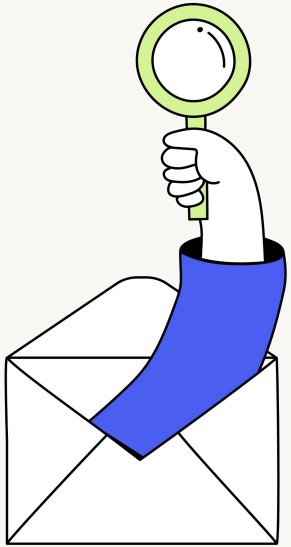
- Confusion matrices decompose predictions into TP, FP, FN, TN to pinpoint exactly where your model errs.
- ROC curves plot true-positive rate against false-positive rate across all thresholds, with AUC summarising overall discrimination.
- Use confusion matrix insights to choose and tune metrics (Precision, Recall) based on specific error costs.
- Leverage ROC/AUC threshold analysis to select the optimal operating point for your business priorities.



Skills application: Model evaluation demo

Skills application introduction

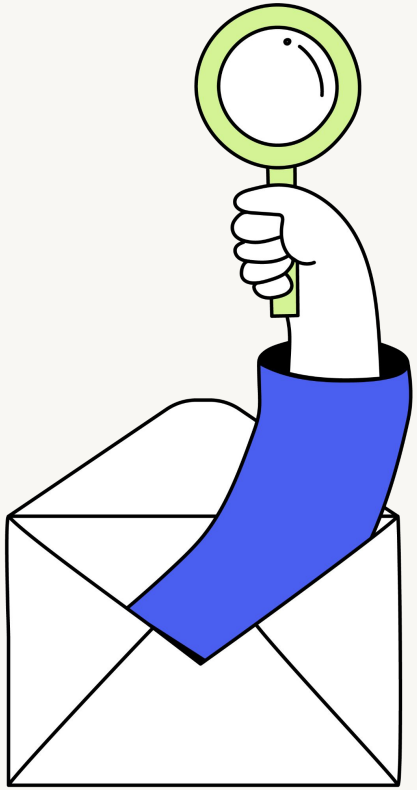
- **Objective:** Apply an end-to-end ensemble classification pipeline to predict London house price categories, reinforcing practical skills in data preparation, model training, tuning, and evaluation for robust, generalisable results.
- **We'll show you how to:**
  - Load, explore, clean, and impute the London House Price dataset.
  - Train Random Forest, Gradient Boosting, and XGBoost models with GridSearchCV and RandomizedSearchCV.
  - Evaluate via confusion matrices, accuracy, precision, and recall; interpret key trade-offs.
  - Reinforce strategies for model robustness and generalisation.



Skills application: Model evaluation

Skills application

- Work in breakout rooms in small groups.
- Use the Jupyter Notebook and provided dataset.
- The challenge is designed to be solved individually, but collaboration is encouraged.
- Find all instructions and resources on Ariel.

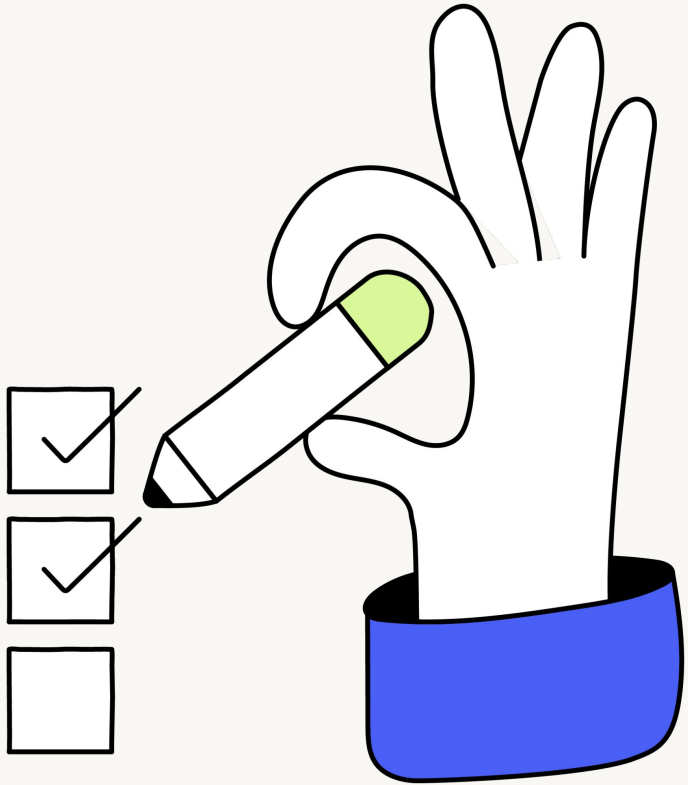


Access the skills application

Go to My Multiverse, Module 8
Workshop 1

Go to the Skills application page and follow the instructions provided

Be prepared to discuss your learnings when we come back together



Skills application debrief

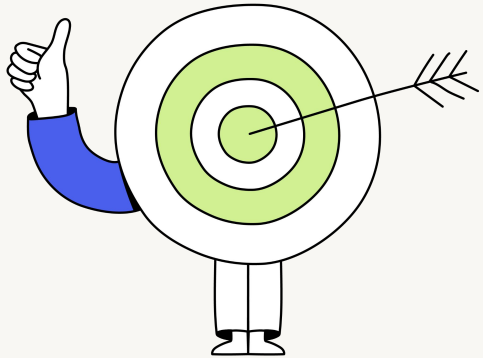
Skills application debrief

What techniques did you apply, and how did they impact the model’s performance?

Key takeaways

1. **Clean & slice for fairness:** Derive and bin demographic proxies, impute missing values, and encode categoricals to prep six demographic slices for disaggregated analysis.

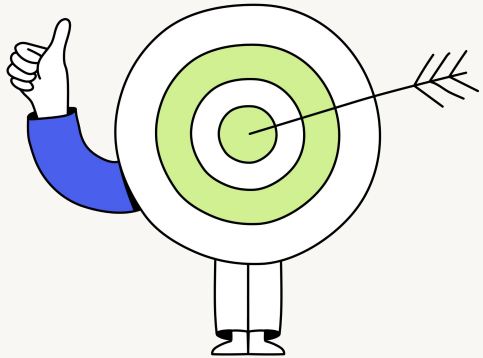
2. **Train & calibrate:** Fit your CNN baseline with class-weighted loss and regularisation, then apply post-hoc temperature scaling and instance re-weighting to mitigate bias.



Key takeaways

3 **Disaggregate & diagnose:** Use confusion matrices alongside per-slice FNR and max-gap metrics—visualised in histograms and bar plots—to pinpoint and interpret disparate errors.

4 **Balance & govern:** Leverage these insights to fine-tune fairness–accuracy trade-offs, ensuring your model meets both performance targets and governance KPIs.



Applying key takeaways to your organisation

Applying key takeaways

1. Where does messy or incomplete data crop up in your projects? How are you currently exploring, imputing, and encoding it to build more reliable models?

2. Which classification challenges in your team could benefit from ensemble methods? How might hyperparameter tuning with GridSearchCV or RandomizedSearchCV unlock performance gains?

3. How are you interpreting model outcomes today? In what ways could confusion matrices, precision, recall, and accuracy sharpen your insights and drive better business decisions?

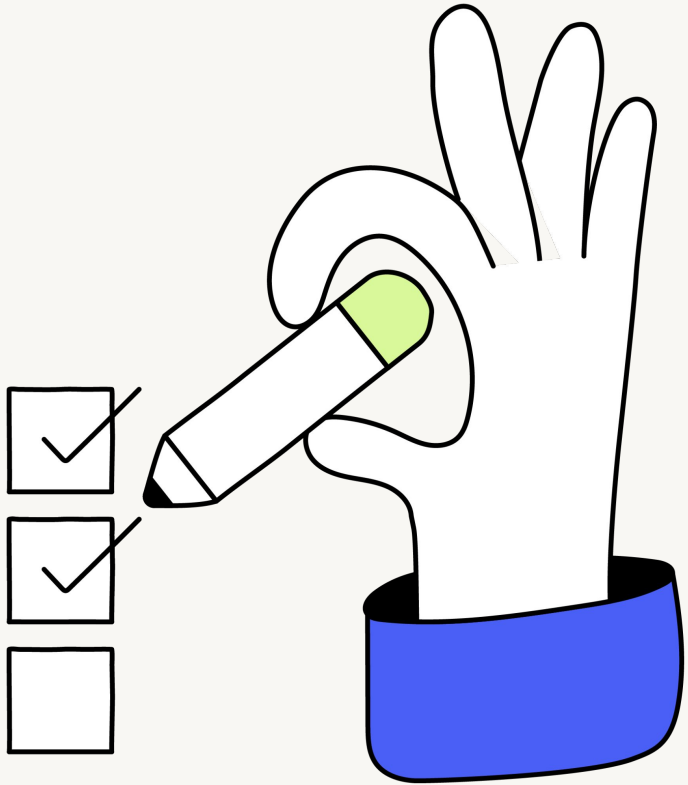
Share your plan to apply new skills

Go to My Multiverse, Module 8
Workshop 2

On the Apply new skills to your role page, find the discussion activity

Take 2 minutes to share how you will apply this skill in your role.

Take time this week to read what others have written and get new ideas!



Closing out

- **Next workshop:** Module 8 Workshop 2
- **Have questions?**
  - Chat with Atlas
  - Reach out to your Cohort Coach
  - Sign up for Tutoring
- **Log your OTJ hours**
- **Take the Workshop Survey** on My Multiverse

multiverse