CED 2026 / RESEARCH PRESENTATION

Cross-Domain Explainable Prediction of Hotel Booking Cancellations

Feature Availability, Calibration, and Transferability under Domain Shift

Oksana Davydova · Vitalii Bezsonnyi · Natalia Balatska

Krystyna Nechepurenko · Larysa Tatar

Simon Kuznets Kharkiv National University of Economics · Ukraine

A model can rank well — and still fail after transfer

A strong ROC-AUC is not sufficient evidence for deployment readiness.

01

DIFFERENT HOTEL

Customer mix, market segments and cancellation prevalence shift between properties.

02

DIFFERENT PERIOD

Demand characteristics change over time even when seasonality is aligned.

03

DIFFERENT FEATURES

Some useful predictors are unavailable, updated or coded differently elsewhere.

Research question

Can cancellation-risk models remain accurate, calibrated and interpretable across these shifts?

119,390 real hotel bookings

Open Hotel Booking Demand datasets · Antonio, de Almeida & Nunes (2019)

H1 · RESORT HOTEL

40,060

bookings · Algarve

27.8% cancelled

H2 · CITY HOTEL

79,330

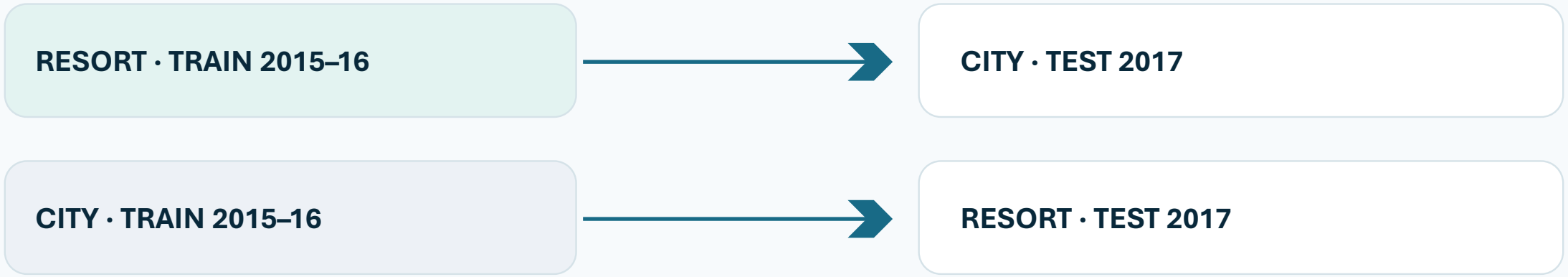
bookings · Lisbon

41.7% cancelled

ARRIVALS 2015–2017 • PRIMARY TRAIN: 2015–2016 • HOLDOUT: 2017 • NEAR-ARRIVAL RISK ASSESSMENT

Two directions · two baselines · one fair comparison

Local and transferred models are evaluated on the same target bookings.



Target-conditioned transfer penalty

$$\Delta AUC(T) = AUC(T \to T) - AUC(S \to T)$$

Primary test: 2017 | Temporal stress test: train through 2015 → Jan–Aug 2016 and Jan–Aug 2017

ANALYTICAL PIPELINE

Three algorithms and three predictor policies

Source-only preprocessing, fixed model settings, no target-specific tuning.

LR **Logistic regression**

Transparent linear baseline

RF **Random forest**

Non-linear ensemble

XGB **XGBoost**

Boosted decision trees

FEATURE-POLICY SENSITIVITY

P20 **20 predictors**

Primary portability-oriented

S18 **18 predictors**

Excludes parking + special requests

N19 **19 predictors**

Excludes Country

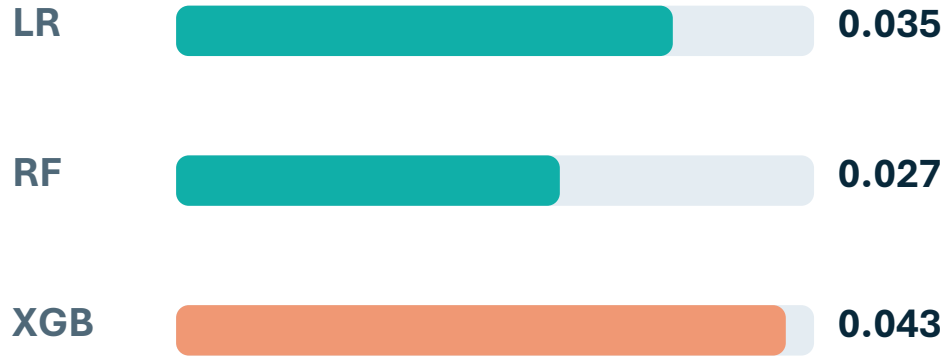
Metrics: ROC-AUC · PR-AUC · Brier · ECE · calibration intercept/slope · SHAP · 95% bootstrap CI

KEY RESULT 1

Cross-domain portability is asymmetric

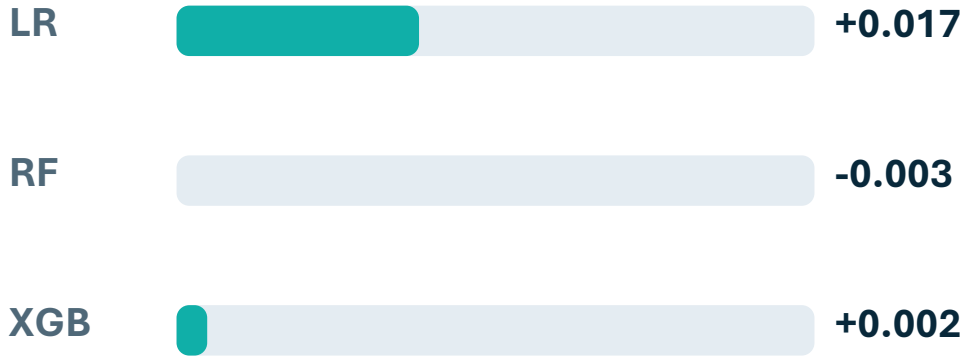
Target-conditioned ROC-AUC penalty (local target benchmark minus transferred model).

CITY TARGET • R → C



XGBoost: 0.043 AUC penalty on city bookings

RESORT TARGET • C → R



RF / XGBoost: near-zero resort penalty

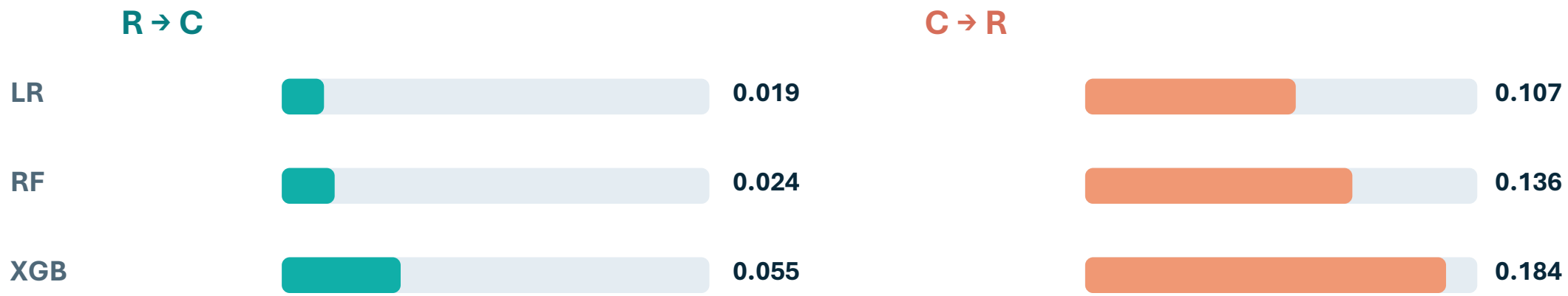
The finding describes two individual hotels; hotel type and property identity are confounded.

KEY RESULT 2

Transfer performance changes over time

Prospective stress test — trained on source data available through December 2015.

Change in cross-domain ROC-AUC: Jan–Aug 2016 → Jan–Aug 2017



Both directions deteriorate; these drops are diagnostic, not the primary model estimates.

Calendar-matched windows control months, but the model uses only partial 2015 training history.

KEY RESULT 3

Good discrimination ≠ reliable probabilities

Resort-trained XGBoost deployed on the 2017 city-hotel target.

R → C · XGBOOST

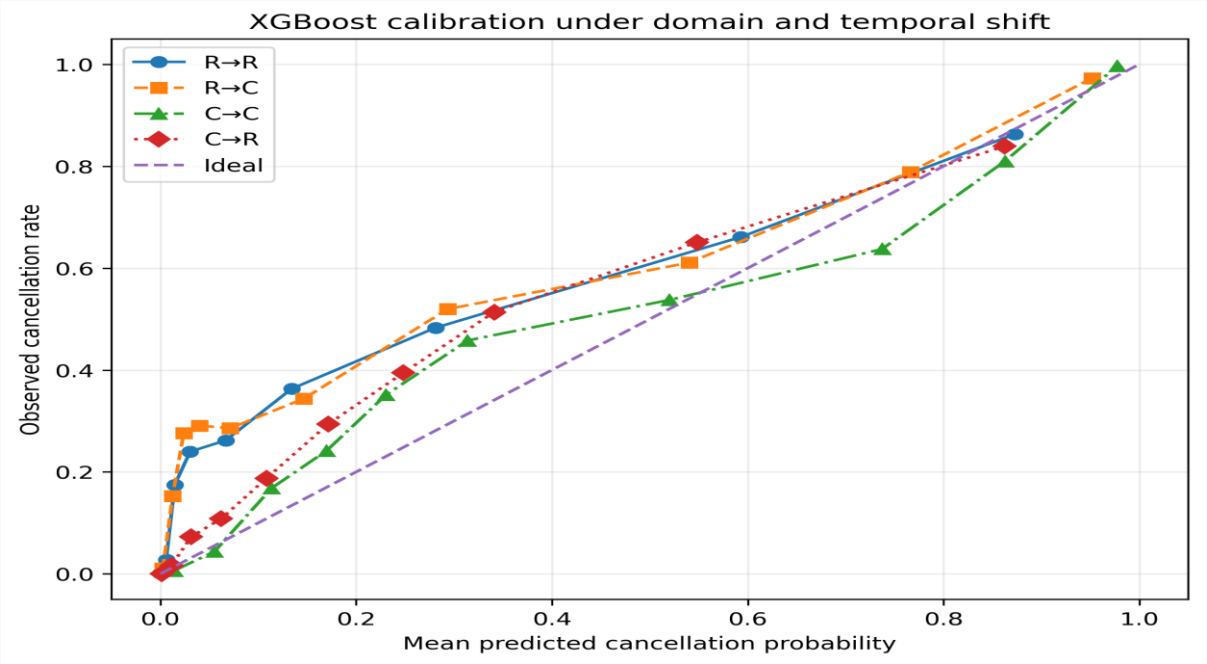
28.5%

Mean predicted cancellation risk

42.5%

Observed cancellation frequency

AUC 0.821 • ECE 0.143



Calibration intercept +0.662 | slope 0.577

KEY RESULT 4

Feature availability changes the transfer story

XGBoost target-conditioned ROC-AUC penalty; positive values indicate performance loss.

Predictor policy		City target · R → C	Resort target · C → R
P20	20 predictors	+0.043	+0.002
S18	minus 2 request fields	+0.067	-0.010
N19	without Country	+0.037	+0.063

Removing Country increases the city → resort XGBoost penalty from 0.002 to 0.063.

Country is a contextual market-origin marker, not a causal or stand-alone customer-treatment rule.

Stable feature rankings can hide changed effects

SHAP summary statistics for XGBoost across source and deployment domains.

RESORT-TRAINED

0.941

Spearman rank correlation

1.000	Top-10 Jaccard
0.121	TVD of normalized SHAP

CITY-TRAINED

0.964

Spearman rank correlation

1.000	Top-10 Jaccard
0.131	TVD of normalized SHAP

High rank stability does not imply unchanged contribution magnitude in the new hotel.

Four gates before reusing a hotel model

Operational reliability is a deployment assessment, not one leaderboard score.

01

DISCRIMINATION

Does risk ranking hold locally?

02

CALIBRATION

Are probabilities reliable for thresholds?

03

FEATURE COMPATIBILITY

Are inputs meaningful and available?

04

EXPLANATION DRIFT

Have contribution profiles changed?

Use local validation and, where justified, recalibration before threshold-based decisions.

Transfer requires more than predictive accuracy

Evidence from two Portuguese hotels with 119,390 reservations.

01	**Asymmetric transfer**	Target-local comparisons reveal a penalty mainly for resort → city.
02	**Fragile probabilities**	Good ROC-AUC can coexist with serious underprediction.
03	**Conditional robustness**	Feature policy, temporal shift and SHAP effects matter.

Reproducibility package · doi.org/10.5281/zenodo.23187750