Predicting ICU In-Hospital Mortality with MIMIC-III: A Hands-On ML Tutorial

Tutorial code:

Why This Matters

Clinical motivation: risk stratification, triage, early warning.

Learning goals:

- Merge key MIMIC tables and engineer features
- Train + evaluate ML baselines
- Make the tutorial reproducible for peers

Data Overview (MIMIC-III)

- Tables used: ADMISSIONS, PATIENTS, ICUSTAYS, LABEVENTS
- Target: In-hospital mortality (from DEATHTIME or HOSPITAL_EXPIRE_FLAG)
- Features: AGE, GENDER, FIRST_CAREUNIT, selected LABs (e.g., glucose, creatinine, WBC, hematocrit, lactate)

Data Preparation & Joins

- Merge keys: SUBJECT_ID, HADM_ID
- One row per admission
- Handle datetimes and compute AGE safely (year/month/day method; no timedelta overflow)
- Example code snippet (high level)

```
Merged shape: (58976, 9)
```

	SUBJECT_ID	HADM_ID	ADMITTIME	DISCHTIME	MORTALITY	GENDER	DOB	FIRST_CAREUNIT	AGE
0	22	165315	2196-04-09 12:26:00	2196-04-10 15:54:00	0	F	2131-05-07	MICU	64.0
1	23	152223	2153-09-03 07:15:00	2153-09-08 19:10:00	0	M	2082-07-17	CSRU	71.0
2	23	124321	2157-10-18 19:34:00	2157-10-25 14:00:00	0	M	2082-07-17	SICU	75.0
3	24	161859	2139-06-06 16:14:00	2139-06-09 12:48:00	0	M	2100-05-31	CCU	39.0
4	25	129635	2160-11-02 02:06:00	2160-11-05 14:55:00	0	M	2101-11-21	CCU	58.0

Feature Engineering

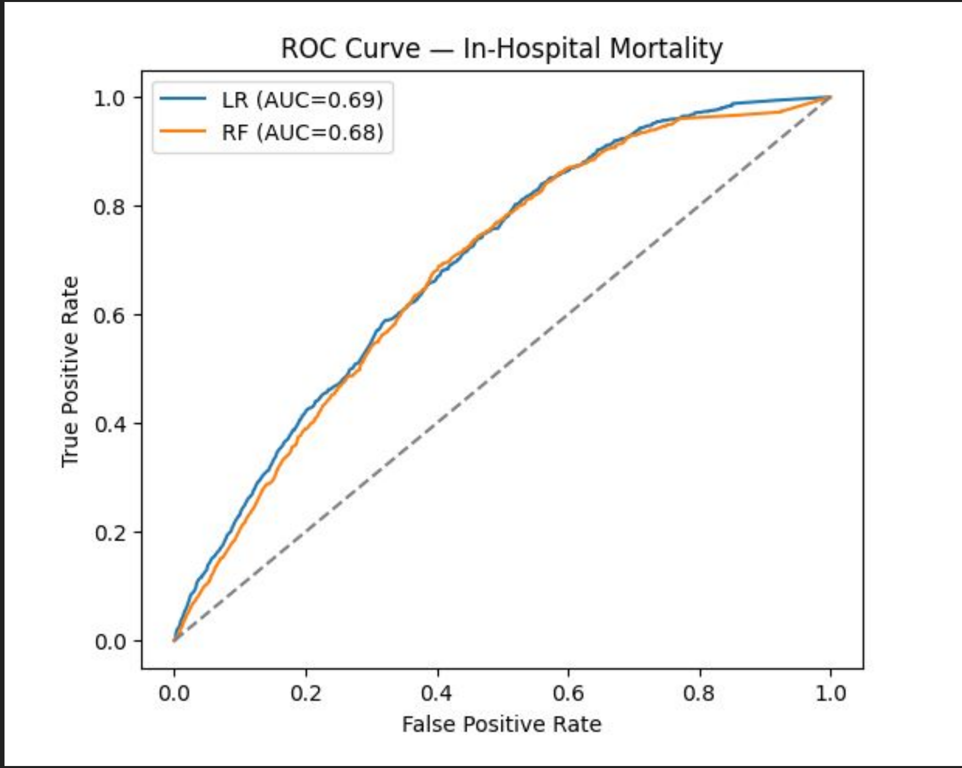
- Categorical → one-hot: GENDER, FIRST_CAREUNIT (with dummy_na=True)
- Labs: select ITEMIDs via D_LABITEMS, pivot per HADM_ID to wide: LAB_<ITEMID>
- Missing values: median imputation
- Class imbalance: show label distribution (~10% positive)

Models & Training

- Baselines:
    - Logistic Regression (Impute → Scale → LR)
    - Random Forest (Impute → RF)
- Train/Test split (stratified 80/20)
- Freeze training feature names to avoid misalignment

Results — ROC Curves

```
Columns with NaNs: [] … total: 0

==== Logistic Regression ====
              precision    recall  f1-score   support

           0      0.901     1.000     0.948     10625
           1      0.000     0.000     0.000      1171

    accuracy                          0.901     11796
   macro avg      0.450     0.500     0.474     11796
weighted avg      0.811     0.901     0.854     11796

ROC-AUC: 0.693

==== Random Forest ====
              precision    recall  f1-score   support

           0      0.901     1.000     0.948     10625
           1      0.000     0.000     0.000      1171

    accuracy                          0.900     11796
   macro avg      0.450     0.500     0.474     11796
weighted avg      0.811     0.900     0.854     11796

ROC-AUC: 0.682
```



ROC with LABS

==== LogReg + LABS ====

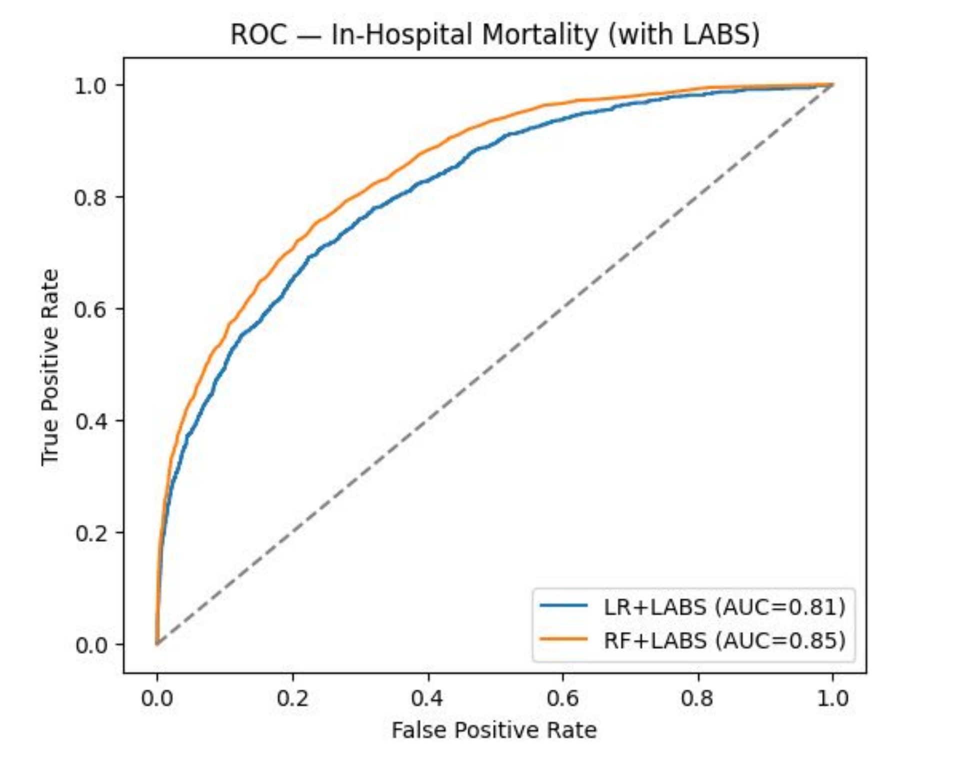
	precision	recall	f1-score	support
0	0.917	0.992	0.953	10625
1	0.725	0.183	0.292	1171
accuracy			0.912	11796
macro avg	0.821	0.588	0.623	11796
weighted avg	0.898	0.912	0.887	11796

ROC-AUC: 0.813

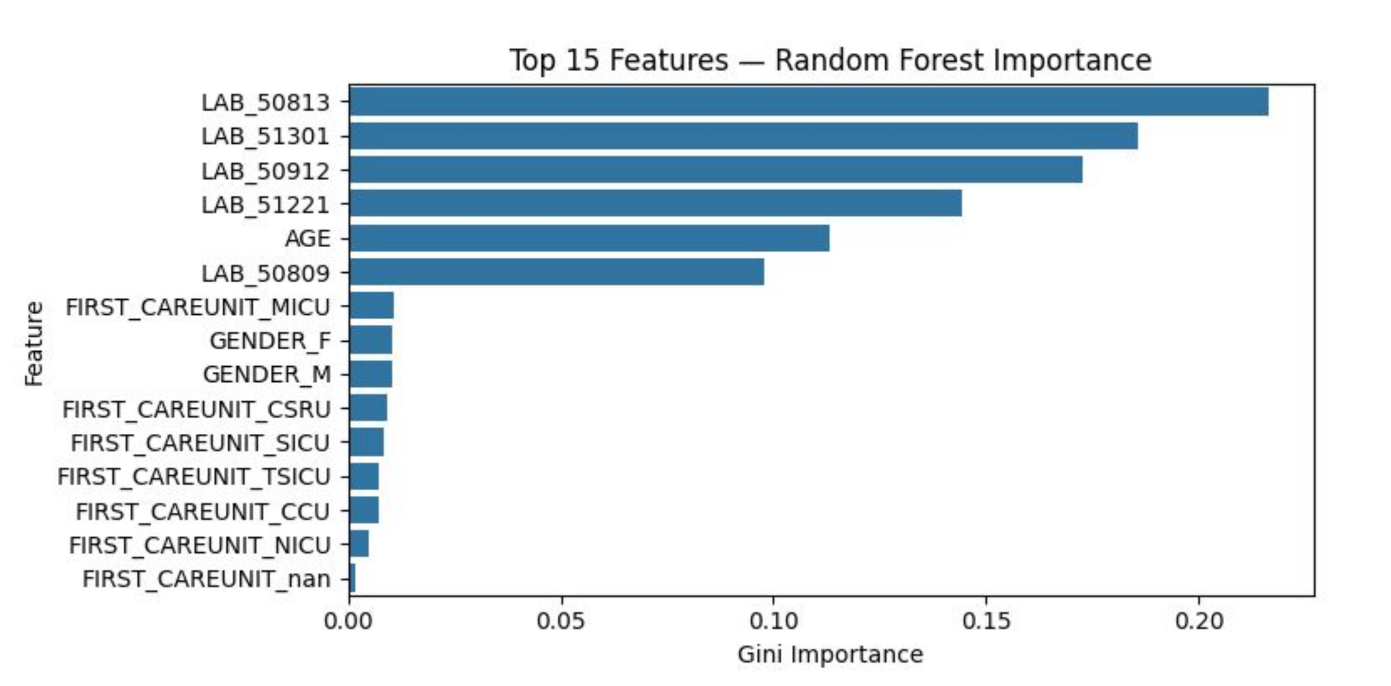
==== RF + LABS ====

	precision	recall	f1-score	support
0	0.923	0.988	0.955	10625
1	0.705	0.257	0.377	1171
accuracy			0.916	11796
macro avg	0.814	0.623	0.666	11796
weighted avg	0.902	0.916	0.897	11796

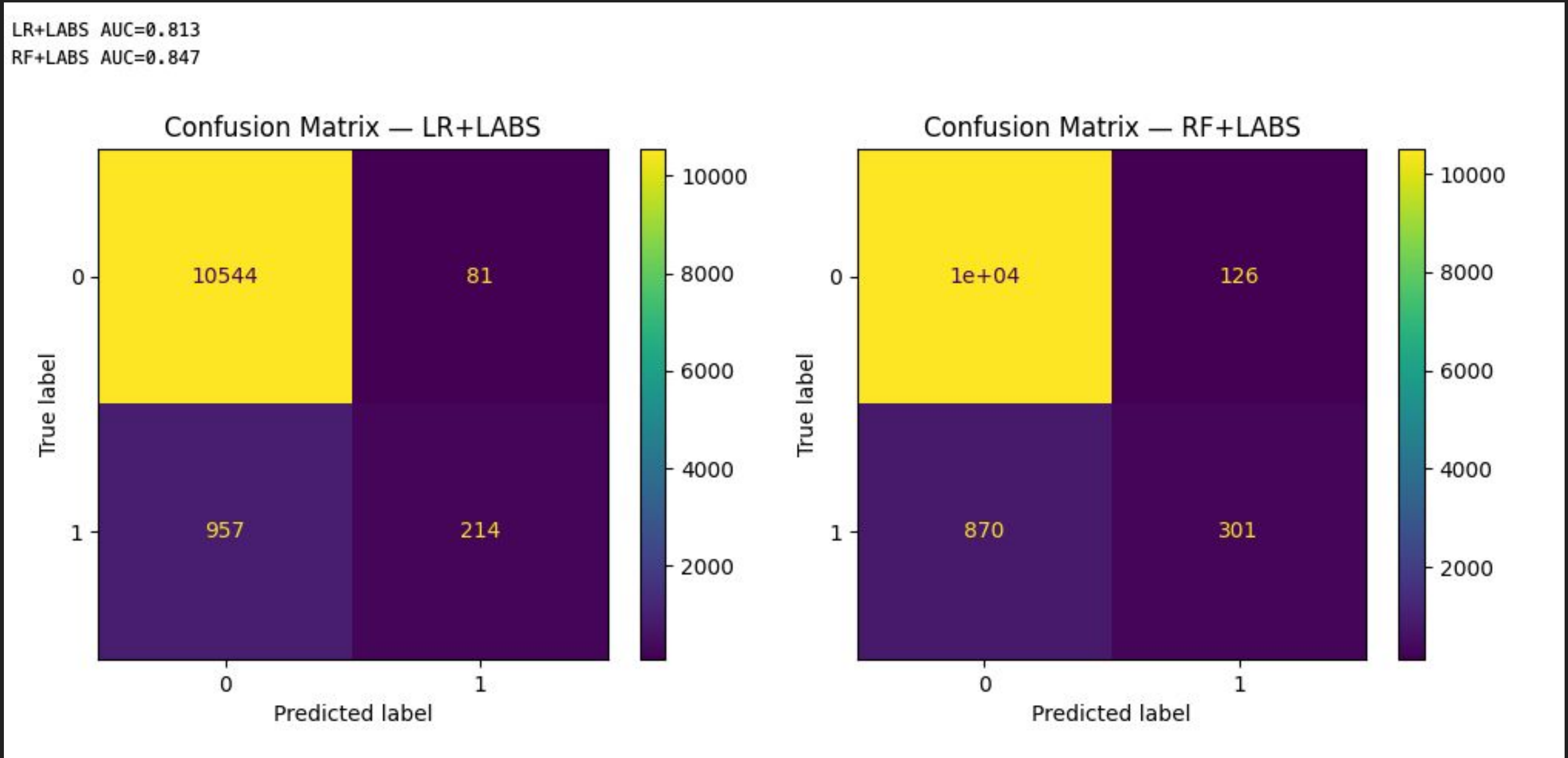
ROC-AUC: 0.847



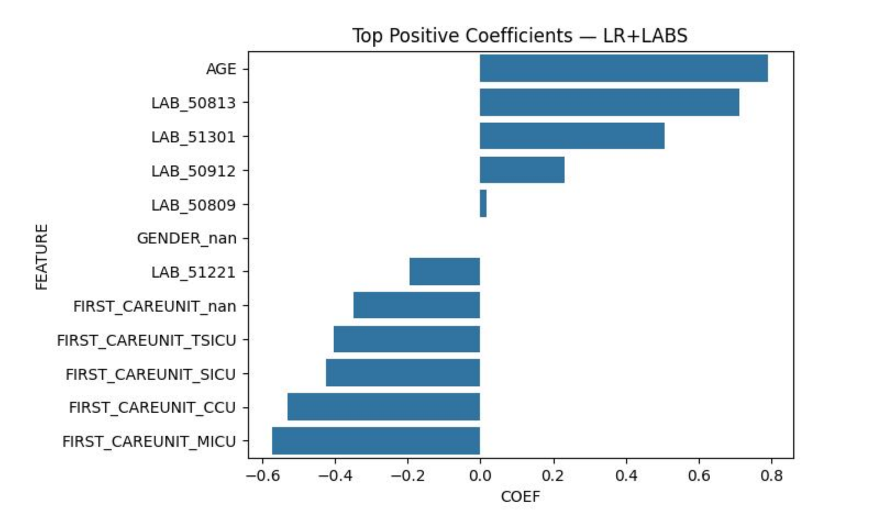
Random Forest Importance

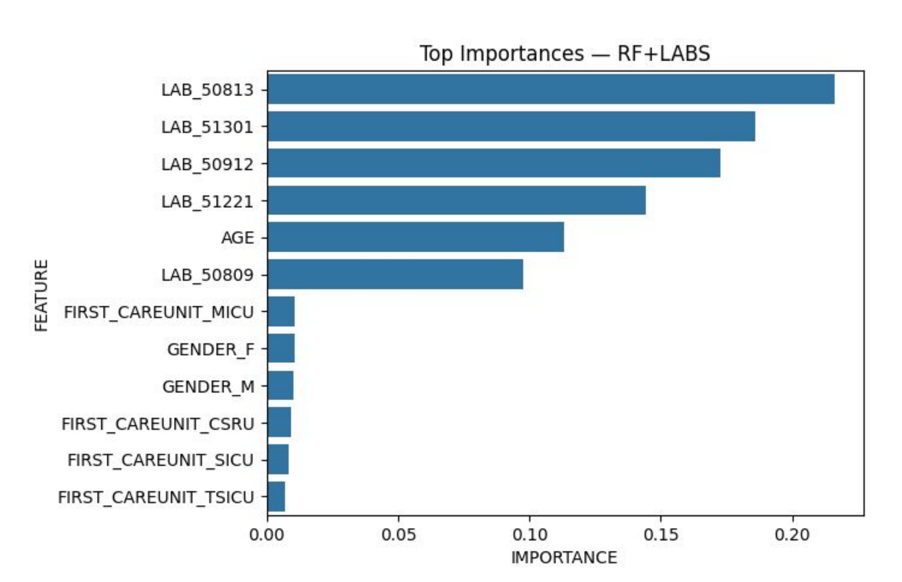


Results — Confusion Matrices & Thresholds



Model Interpretation — LR Coefficients and RF Importances





Limitations & Next Steps

- Single-timepoint features (not time-series); limited lab set
- Missingness patterns; potential selection bias
- Future work:
  - Add vitals/time-series (e.g., LSTMs/Transformers)
  - Calibration & PR curves
  - SHAP for local explanations
  - Cost-sensitive learning or focal loss