Feature Audit & Manual Validation

สรุปการตรวจ clean_dataset.csv เทียบกับ engineered/model features และ manual return-rate check รายลูกค้า

Manual Return Rate Sample

customer_id	orders	returns	formula	rate	percent
C0324	19	10	10/19	0.5263	52.63%
C0413	13	5	5/13	0.3846	38.46%
C0043	11	4	4/11	0.3636	36.36%
C0309	11	2	2/11	0.1818	18.18%
C0270	10	5	5/10	0.5000	50.00%
C0086	10	1	1/10	0.1000	10.00%
C0334	9	1	1/9	0.1111	11.11%
C0094	8	5	5/8	0.6250	62.50%
C0163	7	3	3/7	0.4286	42.86%
C0052	6	4	4/6	0.6667	66.67%

Feature Summary by Version

version	features	used	created	leakage	post-delivery	note
clean_dataset	65		0	item_condition, refund_amount, return_date, return_id, return_reason, return_scenario, return_status, risk_score, risk_tier, score_id, scored_at, shap_values	delay_days, delivery_date, delivery_days	ฐานข้อมูล clean หลังรวม order/return/customer/product แต่ยังมี post-event/leakage fields ต้องตัดก่อน train
v1_engineered	51	136	34			มี feature engineering จำนวนมากและ one-hot/encoding ทำให้ final feature count สูง
v1_model_used	136	136	124			ใช้ encoded features 136 ตัว เหมาะเป็น baseline experiment
v2_engineered	39	38	6		delay_days, delivery_days	ลด feature ให้ดีความง่ายขึ้นและใช้feature history/customer/product/logistics
v2_model_used	38	38	6		delay_days, delivery_days	V2 baseline 38 features แต่ยังมี delivery_days/delay_days จึงไม่ order-time safe
v2_xgboost_safe_plus_rolling	60	60	30			ตัด post-delivery fields และเพิ่ม rolling history 30/60/90/180/365d เหมาะกับ Feature Store

version	features	used	created	leakage	post-delivery	note
v3_stacking	38	38	6		delay_days, delivery_days	ไม่ได้สร้าง feature ใหม่เอง ใช้ V2 feature set แล้วเปลี่ยน model เป็น stacking ensemble
v4_generated_engineered	123	180	123			ใช้ synthetic/generated data สำหรับทดลอง imbalance/SMOTE/Optuna
v4_generated_model_used	180	180	160			feature set จาก generated pipeline ไม่ควรเทียบ production โดยตรงกับข้อมูลจริง

Model Summary

version	model	accuracy	recall	f1	auc	cost	rating
V1	XGBClassifier	70.80%	26.80%	34.82%	68.82%	35,900	D
V2 baseline	V2 tuned model	67.87%	65.60%	54.27%	73.66%	19,550	B
V2 XGBoost safe rolling	XGBClassifier	71.07%	56.88%	53.33%	71.47%	20,250	B
V3 stacking	V3 Stacking (XGB+LGB+CAT)	66.67%	63.76%	52.65%	71.90%	20,400	B
V4 generated	XGBoost_SMOTE_Optuna	83.45%	46.39%	45.69%	85.38%	31,650	C

Key Findings

V2 baseline เป็น version ที่ครบที่สุดบนข้อมูลจริง ทั้ง feature engineering, evaluation, SHAP, threshold และ prediction output

V2 XGBoost safe rolling เป็น candidate ที่เหมาะกว่าเมื่อจะใช้จริงแบบ order-time เพราะตัด `delivery_days` และ `delay_days` แล้ว

V3 มี model process แล้ว แต่ feature engineering ไม่ได้แยกเอง ใช้ feature/split จาก V2

V4 ครบ end-to-end มากที่สุดในเชิง pipeline แต่เป็น synthetic/generated data จึงควรระบุให้ชัดเวลานำเสนอ

Feature ที่เป็น identifier เช่น `customer_id`, `order_id` ควรใช้สำหรับ query/audit ไม่ควรส่งเข้า model ให้จำ identity โดยตรง

Feature หลังเหตุการณ์ เช่น return/refund/risk_score/shap_values ต้อง drop ก่อน train เพื่อกัน leakage