EXPLAINABLE AI FOR FRAUD TRANSACTION DETECTION

Shobana Siranjeevilu, Charisma Puli, Kalaiarasi Kaliappan

AGENDA

BUSINESS PROBLEM

SOLUTION PROPOSED

APPROACH

INSIGHTS

CONCLUSION

BUSINESS PROBLEM

In the process of detecting fraudulent claims, companies face the challenge of balancing the costs and risks associated with false negatives (fraudulent claims mistakenly identified as legitimate) and false positives (legitimate claims incorrectly flagged as fraudulent).

This over-sensitivity is primarily due to the model's inability to handle highly imbalanced data (where non-fraudulent claims far outnumber fraudulent ones) and the high cost associated with missing actual fraudulent claims (false negatives).

The company needs a solution that not only improves the accuracy of fraud detection but also provides clear explanations for its decisions, enabling fraud analysts to efficiently review and understand the reasoning behind each flagged claim.

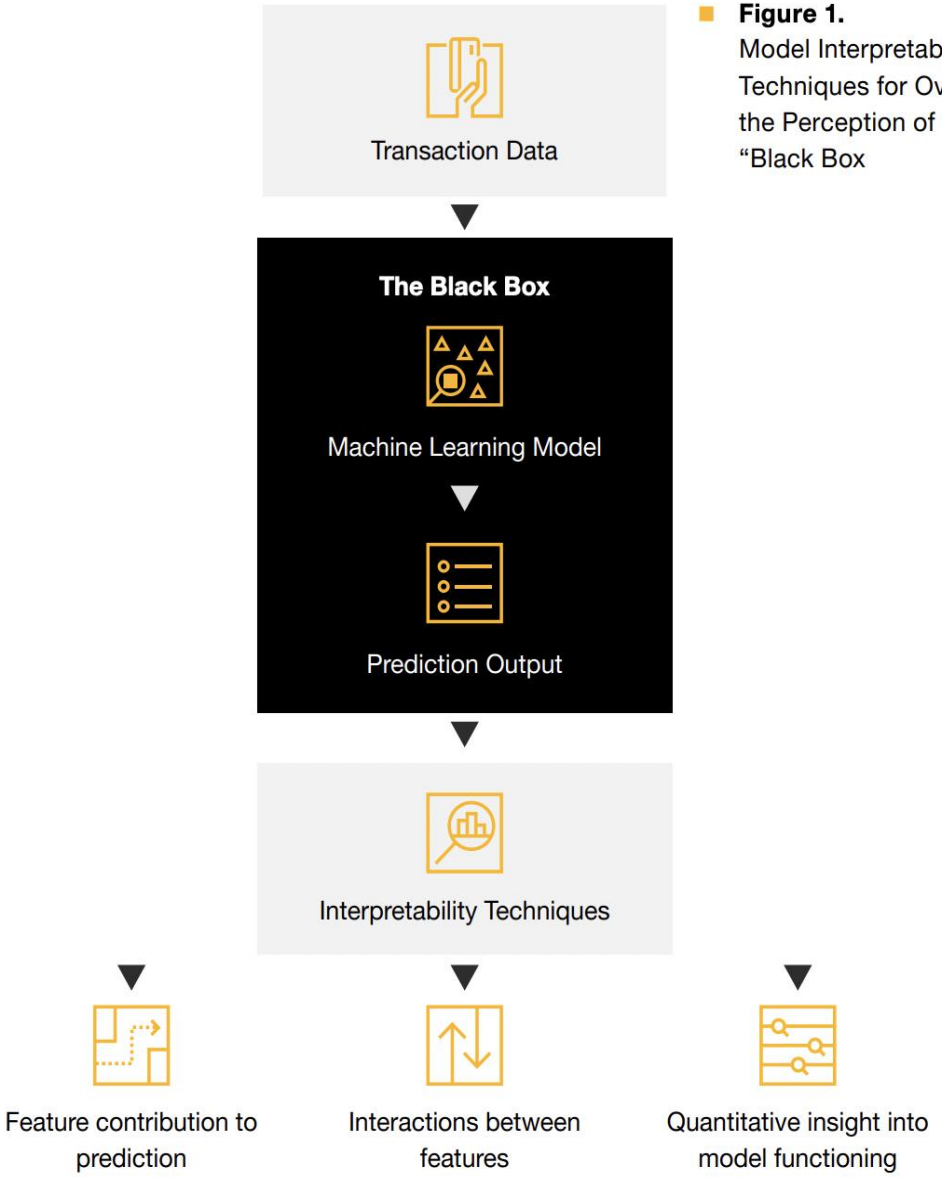


Figure 1.
Model Interpretability Techniques for Overcoming the Perception of the "Black Box

SOLUTION

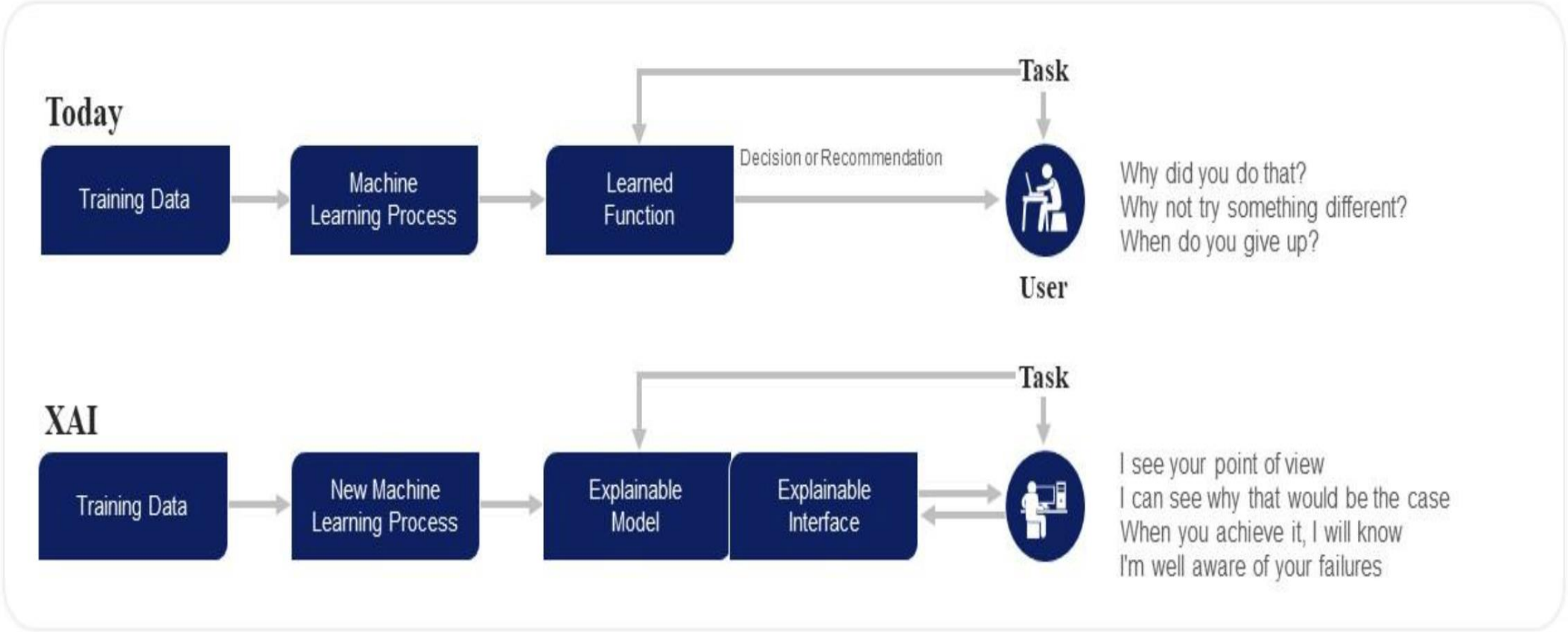
Implementing an Explainable AI (XAI) system could address this specific business problem by providing transparent and interpretable fraud detection models.

XAI would allow the company to understand the factors and patterns the model considers when flagging claims as fraudulent or legitimate.

This enhanced understanding would enable the company to fine-tune the model to reduce false positives (legitimate claims flagged as fraudulent) and false negatives, thus lowering operational costs and improving customer satisfaction.

APPROACH

PRODUCT OVERVIEW



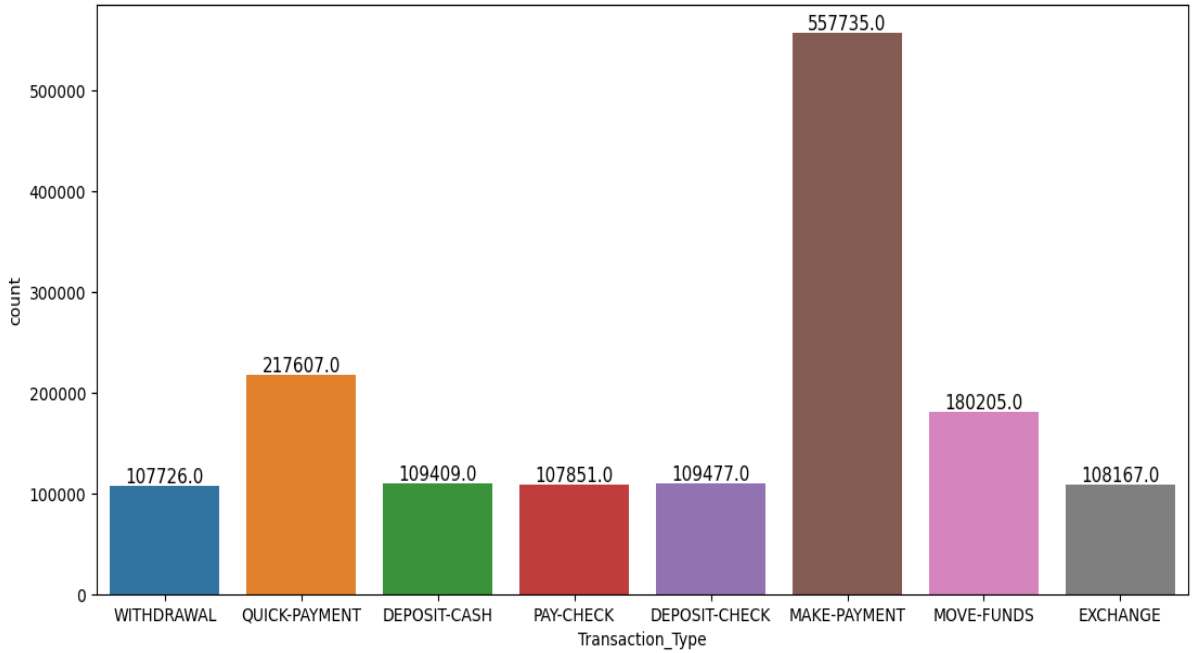
DATA

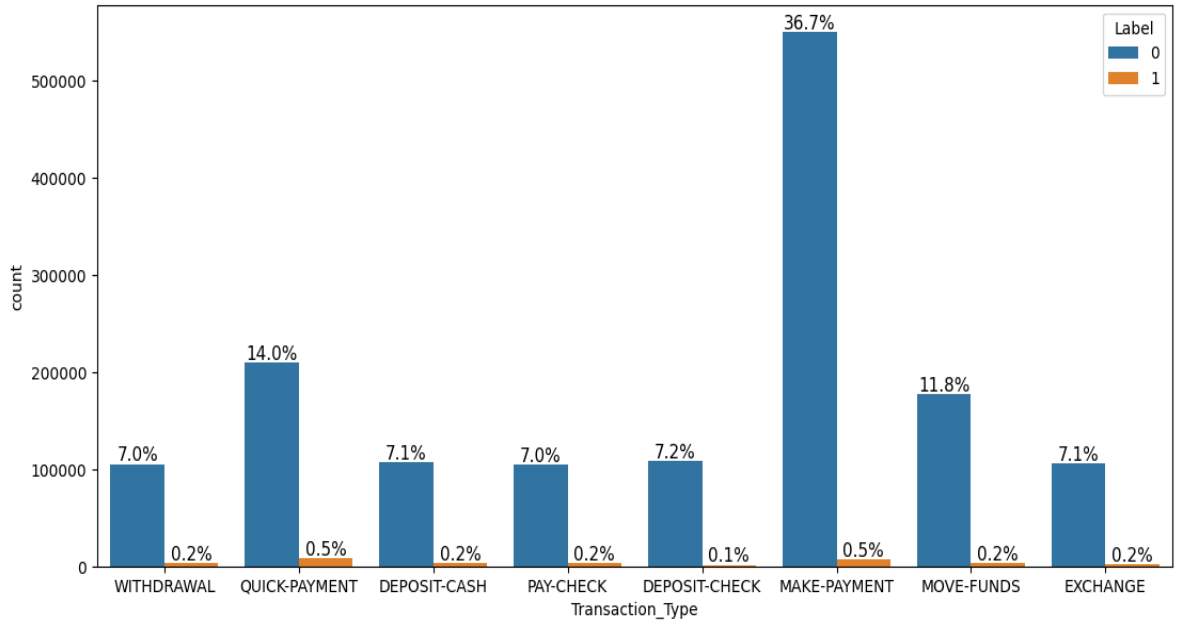
Synthetic Payments Data For Fraud Detection from JPMorgan

Transaction_Id	Sender_Id	Sender_Account	Sender_Country	Sender_Sector	Sender_Iob	Bene_Id	Bene_Account	Bene_Country	USD_Amount	label	Transaction_Type
PAY-BILL-3589	CLIENT-3566	ACCOUNT-3578	USA	21264	CCB	COMPANY-3574	ACCOUNT-3587	GERMANY	492.67	0	MAKE-PAYMENT
WITHDRAWAL-3591	CLIENT-3566	ACCOUNT-3579	USA	18885	CCB				388.92	0	WITHDRAWAL
MOVE-FUNDS-3528	CLIENT-3508	ACCOUNT-3520	USA	4809	CCB	COMPANY-3516	ACCOUNT-3527	GERMANY	280.7	0	MOVE-FUNDS
WITHDRAWAL-3529	CLIENT-3508	ACCOUNT-3519	USA	7455	CCB				118.14	0	WITHDRAWAL
QUICK-DEPOSIT-3471						CLIENT-3442	ACCOUNT-3461	USA	105.16	0	DEPOSIT-CASH
QUICK-DEPOSIT-3473						CLIENT-3442	ACCOUNT-3460	USA	164.97	0	DEPOSIT-CASH
PAY-BILL-3404	CLIENT-3384	ACCOUNT-3395	USA	36316	CCB	COMPANY-3392	ACCOUNT-3401	GERMANY	456.89	0	MAKE-PAYMENT
QUICK-DEPOSIT-3406						CLIENT-3384	ACCOUNT-3396	USA	413.17	0	DEPOSIT-CASH
PAY-CHECK-3347	CLIENT-3330	ACCOUNT-3341	USA	36194	CCB	CLIENT-3333	ACCOUNT-3338	CANADA	377.65	0	PAY-CHECK
PAY-CHECK-3348	CLIENT-3330	ACCOUNT-3340	USA	20626	CCB	CLIENT-3333	ACCOUNT-3338	CANADA	338.03	0	PAY-CHECK
MOVE-FUNDS-3292	CLIENT-3272	ACCOUNT-3284	USA	21568	CCB	CLIENT-3275	ACCOUNT-3291	CANADA	100.85	0	MOVE-FUNDS
MOVE-FUNDS-3294	CLIENT-3272	ACCOUNT-3284	USA	29040	CCB	CLIENT-3273	ACCOUNT-3289	USA	276.66	0	MOVE-FUNDS
PAY-BILL-3232	CLIENT-3203	ACCOUNT-3222	USA	27393	CCB	COMPANY-3210	ACCOUNT-3218	GERMANY	234.88	0	MAKE-PAYMENT
QUICK-DEPOSIT-3234						CLIENT-3203	ACCOUNT-3222	USA	945.22	0	DEPOSIT-CASH
DEPOSIT-CASH-3163						CLIENT-3139	ACCOUNT-3154	USA	655.09	0	DEPOSIT-CASH
PAY-BILL-3162	CLIENT-3139	ACCOUNT-3153	USA	25066	CCB	COMPANY-3147	ACCOUNT-3160	GERMANY	675.37	0	MAKE-PAYMENT
WITHDRAWAL-3100	CLIENT-3075	ACCOUNT-3090	USA	22778	CCB				319.95	0	EXCHANGE
QUICK-PAYMENT-3099	CLIENT-3075	ACCOUNT-3091	USA	39013	CCB	CLIENT-3078	ACCOUNT-3087	TAIWAN	771.54	0	QUICK-PAYMENT
PAY-BILL-3036	CLIENT-3016	ACCOUNT-3028	USA	43951	CCB	COMPANY-3022	ACCOUNT-3033	GERMANY	730.69	0	MAKE-PAYMENT

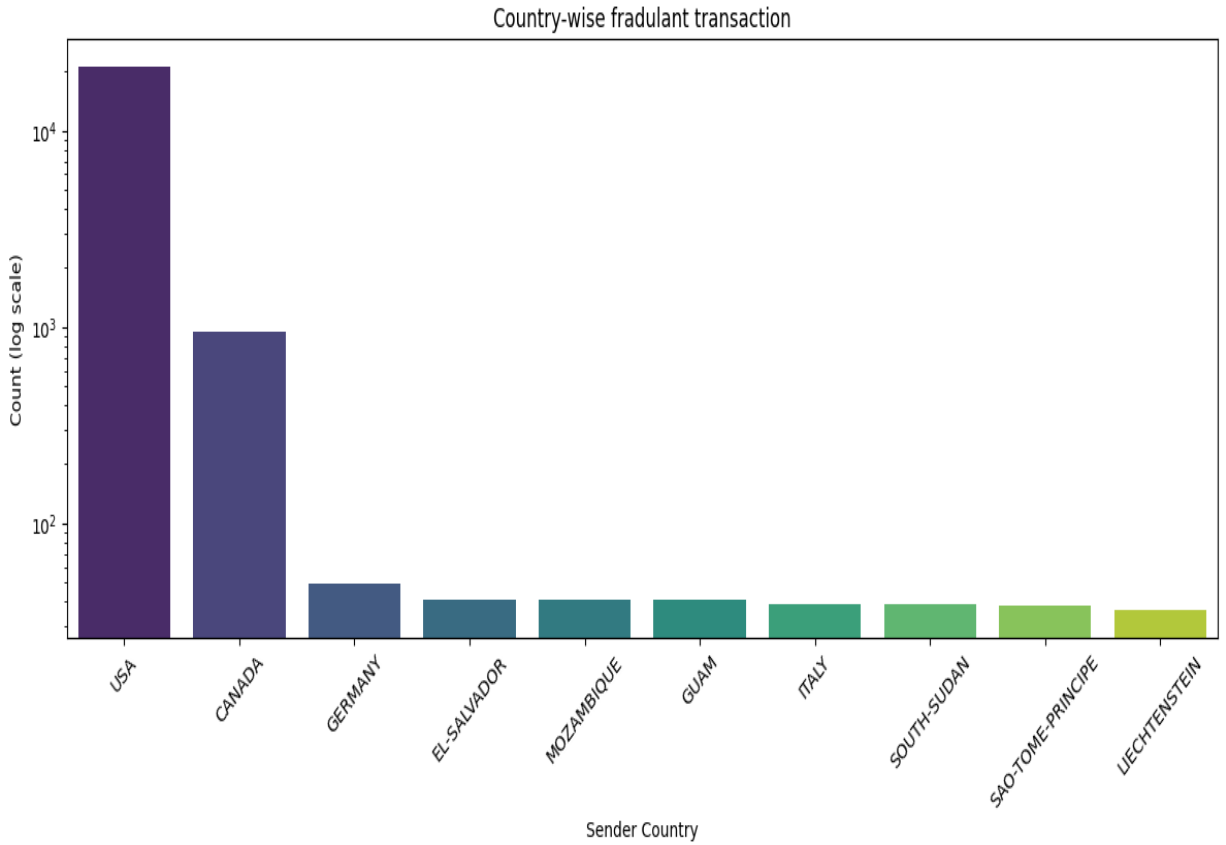
EXPLORATORY DATA ANALYSIS

INSIGHTS INTO TRANSACTION TYPE

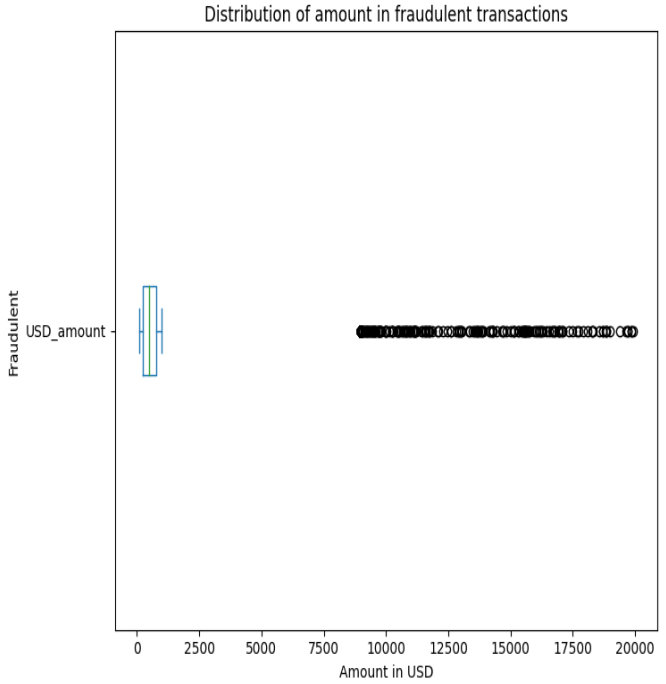


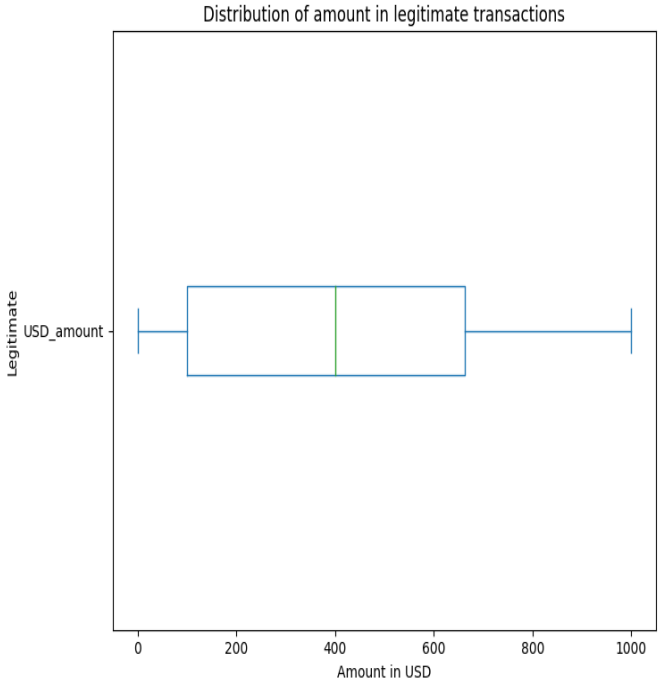


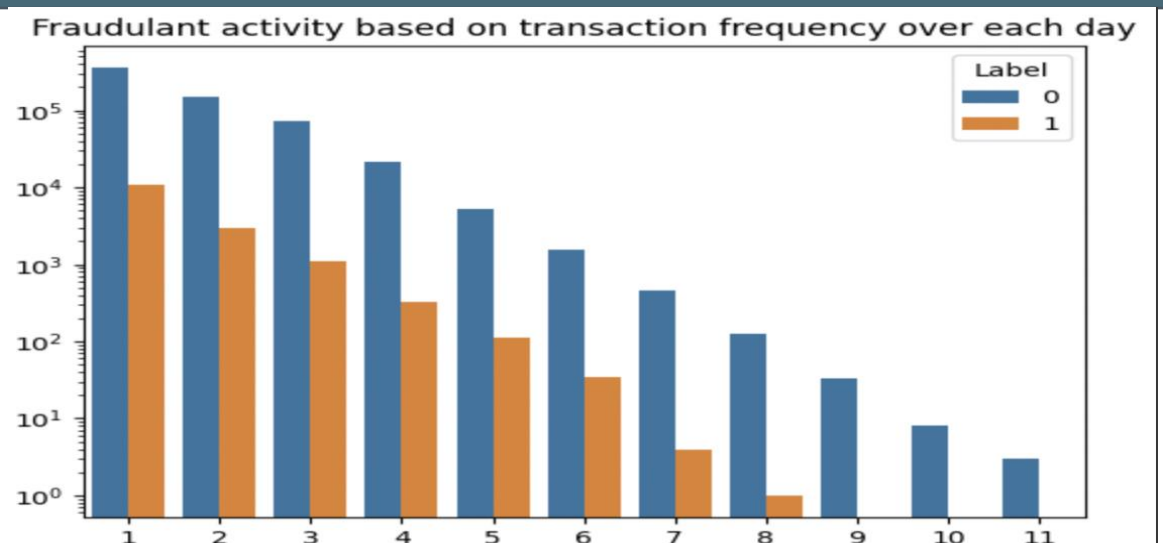
COUNTRY-WISE FRAUDULENT HISTORY

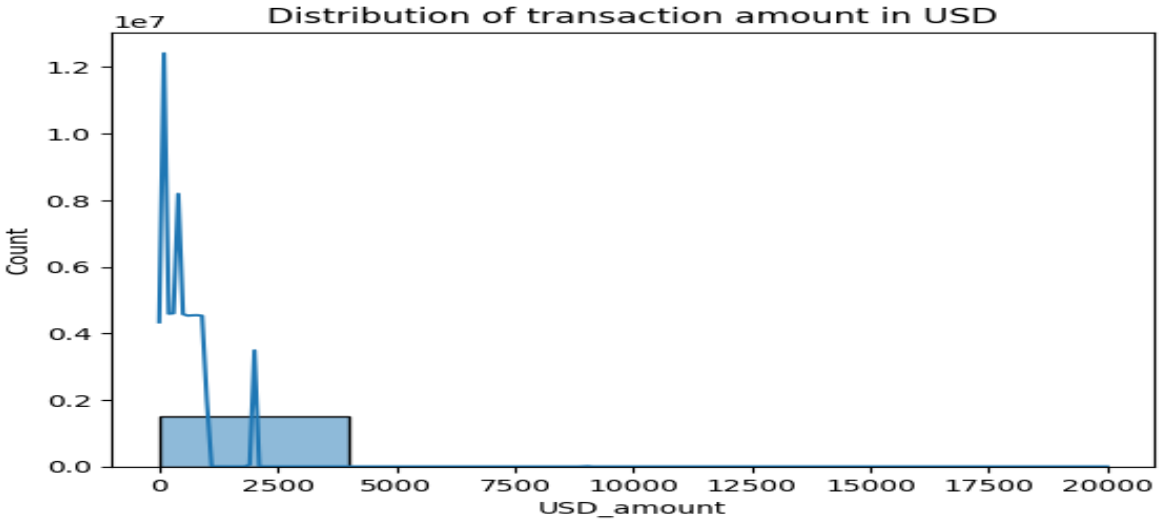


DISTRIBUTION OF FRAUDULENT TRANSACTION









FEATURE ENGINEERING

From EDA we derived these columns to train the model

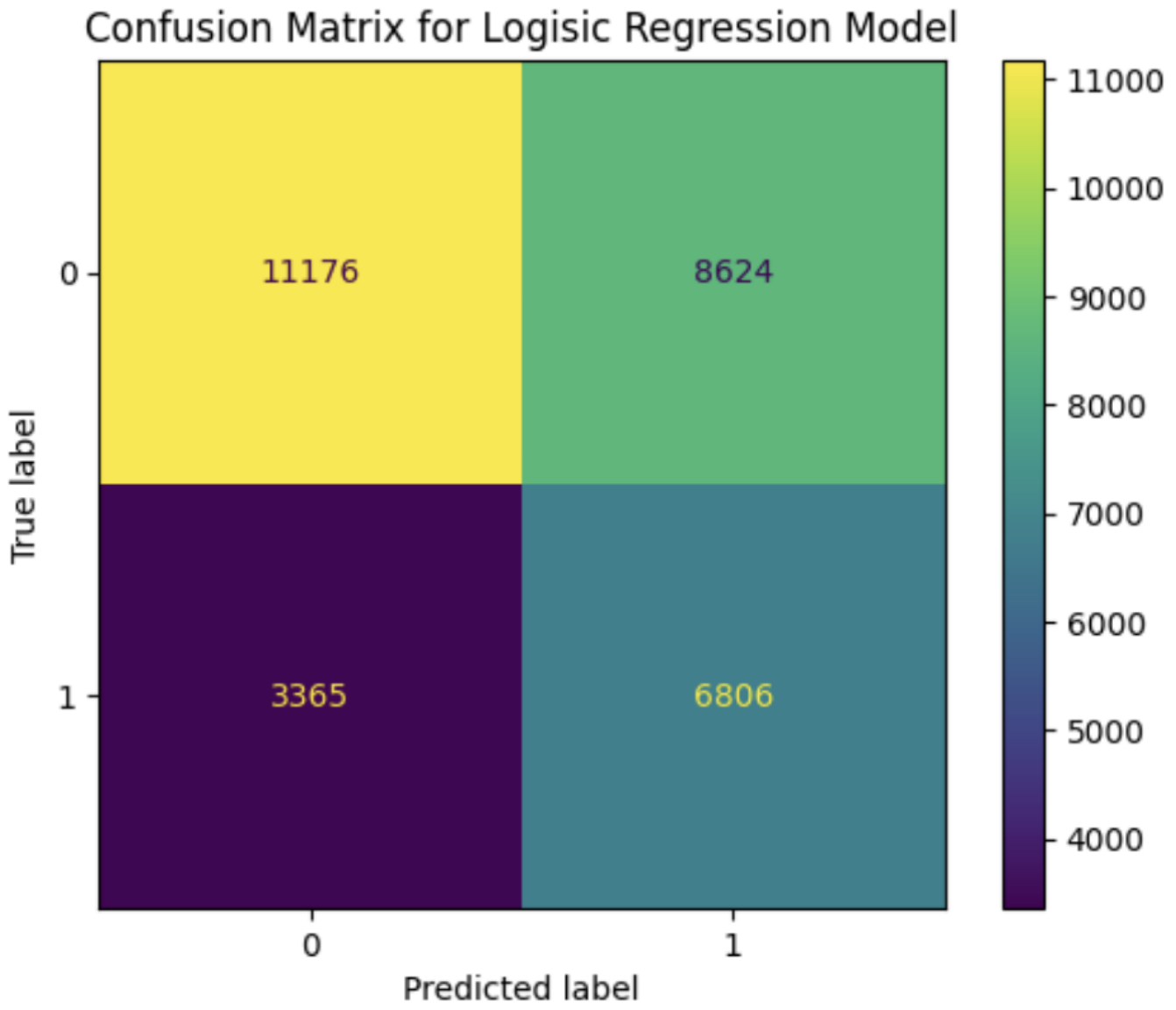
	is_international	hour_of_transaction	time_since_last_transaction_sender	time_since_last_transaction_bene	new_bene_for_sender	sender_bene_unique_count
0	1	6	693.833333	0.0	1	7
1	0	20	243.833333	0.0	1	3
2	1	3	0.000000	0.0	1	5
3	1	13	36.666667	849.5	0	5
4	0	17	109.500000	0.0	1	5
...	...	...	...	...	...	...
90814	0	9	73.500000	0.0	1	5
90815	0	3	217.666667	0.0	1	6
90816	0	1	0.000000	0.0	1	2
90817	0	20	97.500000	0.0	1	8
90818	0	8	0.166667	0.0	1	7

90819 rows × 12 columns

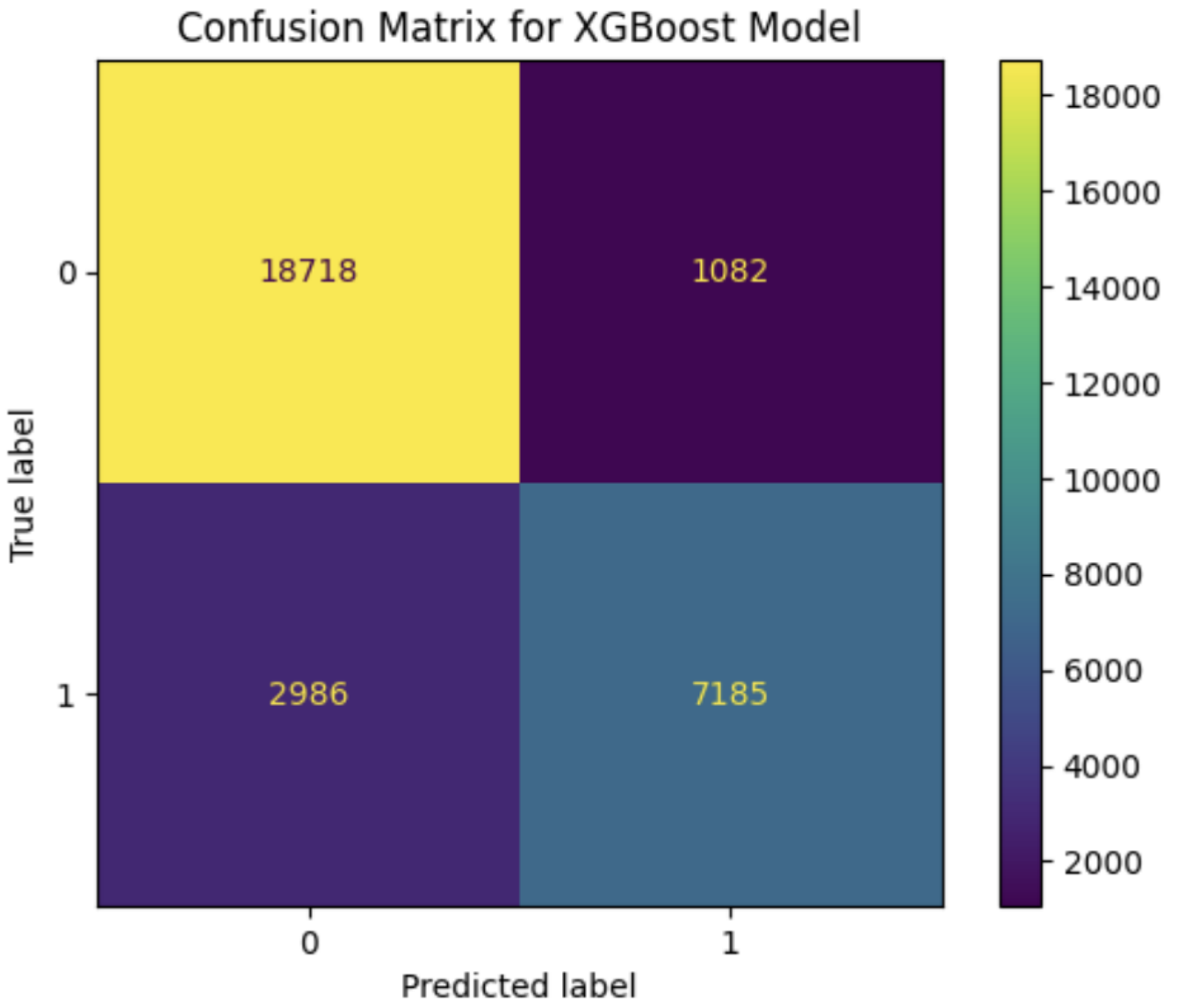
transaction_amount_deviation_sender	transaction_amount_deviation_bene	is_round_amount	average_sender_amount	average_bene_amount	Label
-423.915000	0.000	0	428.085000	4.1700	0
-533.002500	0.000	0	535.562500	439.3200	0
-330.028571	-90.675	0	363.348571	123.9950	0
-148.678571	90.675	0	363.348571	123.9950	0
-264.198571	0.000	0	363.348571	477.7625	0
...	...	...	...	...	...
78.800000	0.000	0	616.690000	416.4775	0
77.363333	0.000	0	320.236667	570.4700	0
258.320000	0.000	0	673.550000	931.8700	0
-305.923333	0.000	0	470.753333	668.4860	0
477.655000	0.000	0	428.085000	566.9700	1

MACHINE LEARNING

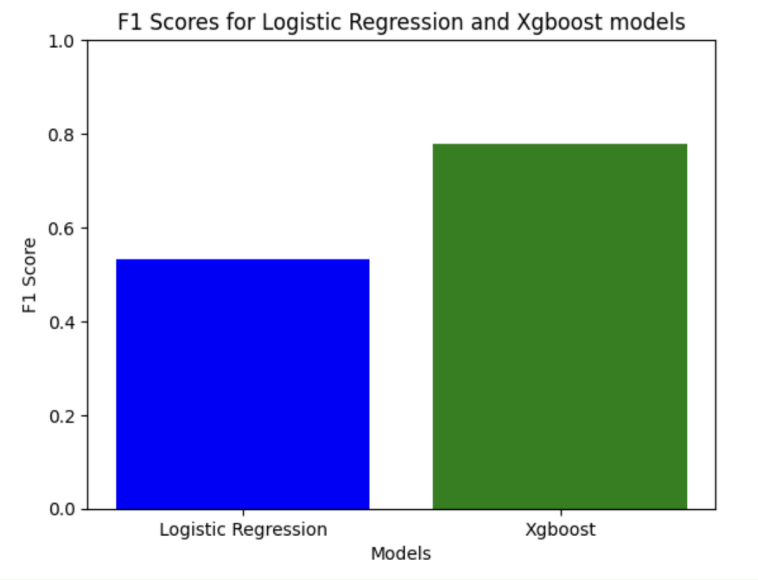
LOGISTIC REGRESSION (BASE)

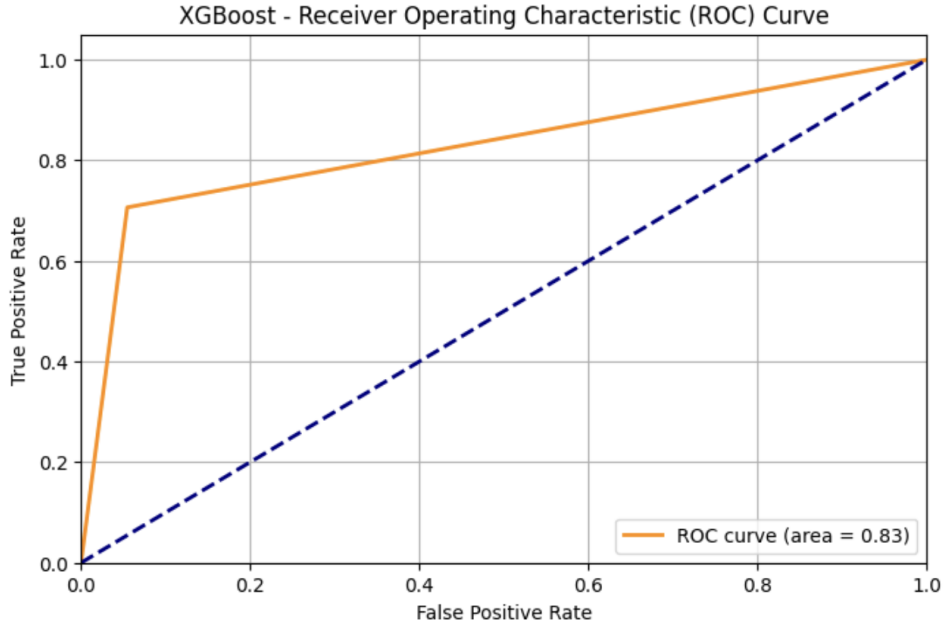


XGBOOST


Confusion Matrix for XGBoost Model
True label
Predicted label
0
1
18718
1082
2986
7185
18000
16000
14000
12000
10000
8000
6000
4000
2000

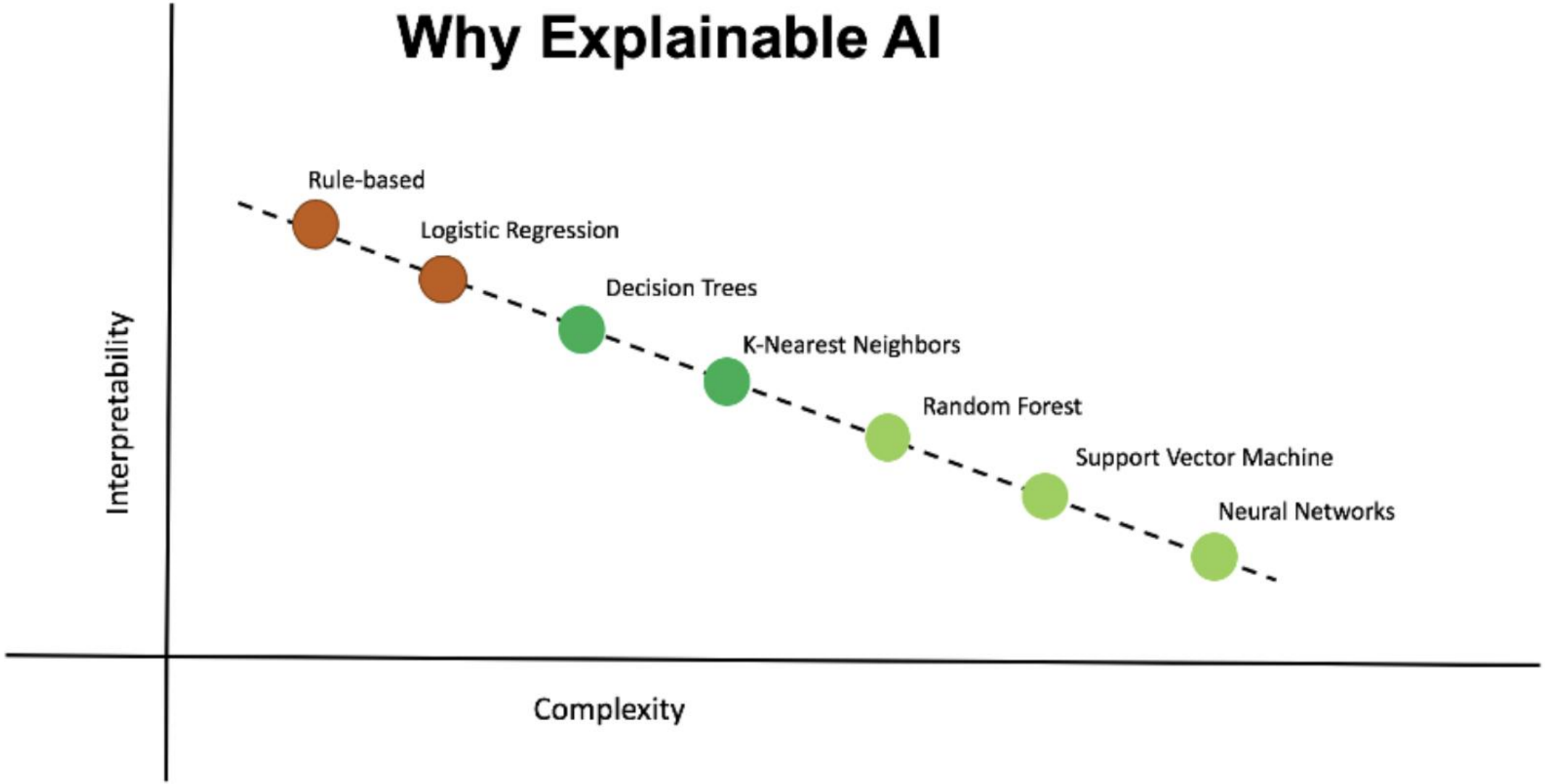
COMPARING MODELS



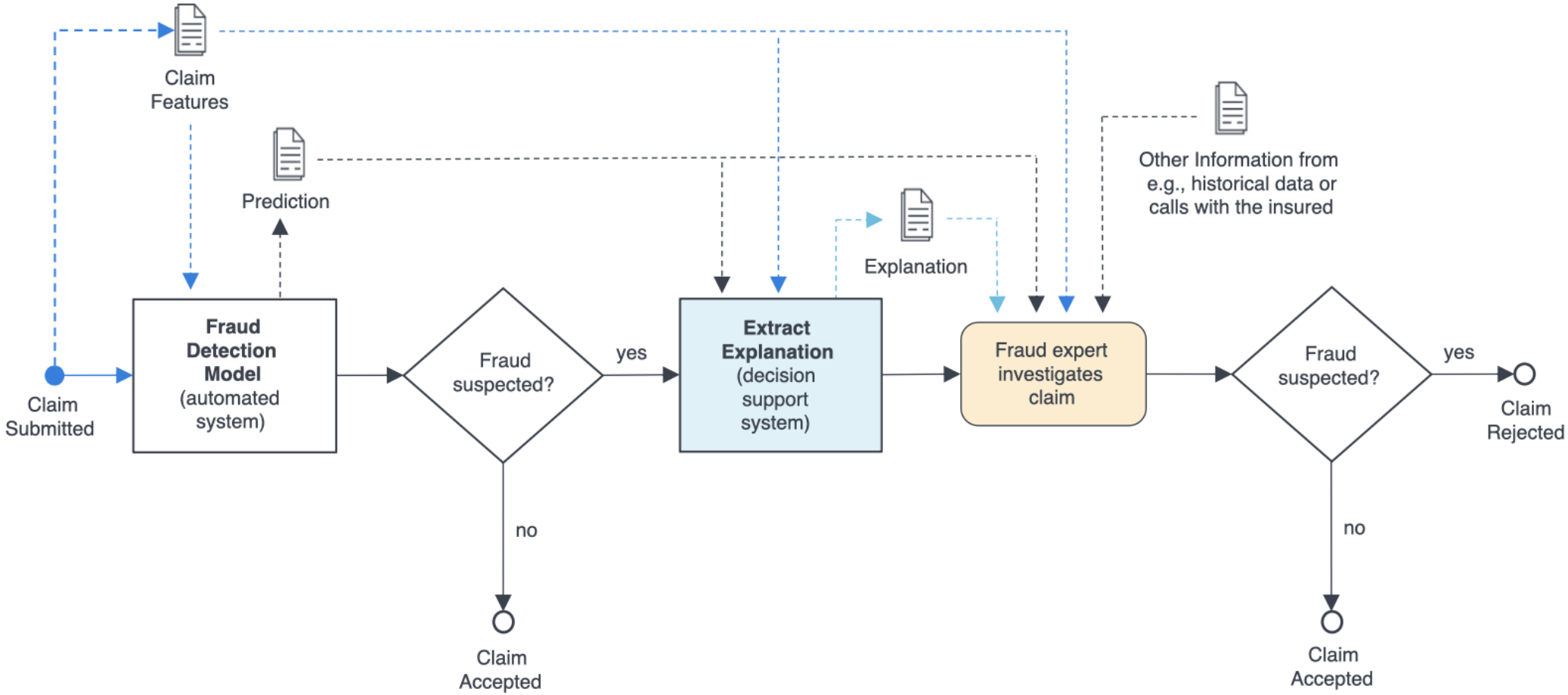


EXPLAINABLE AI(XAI)

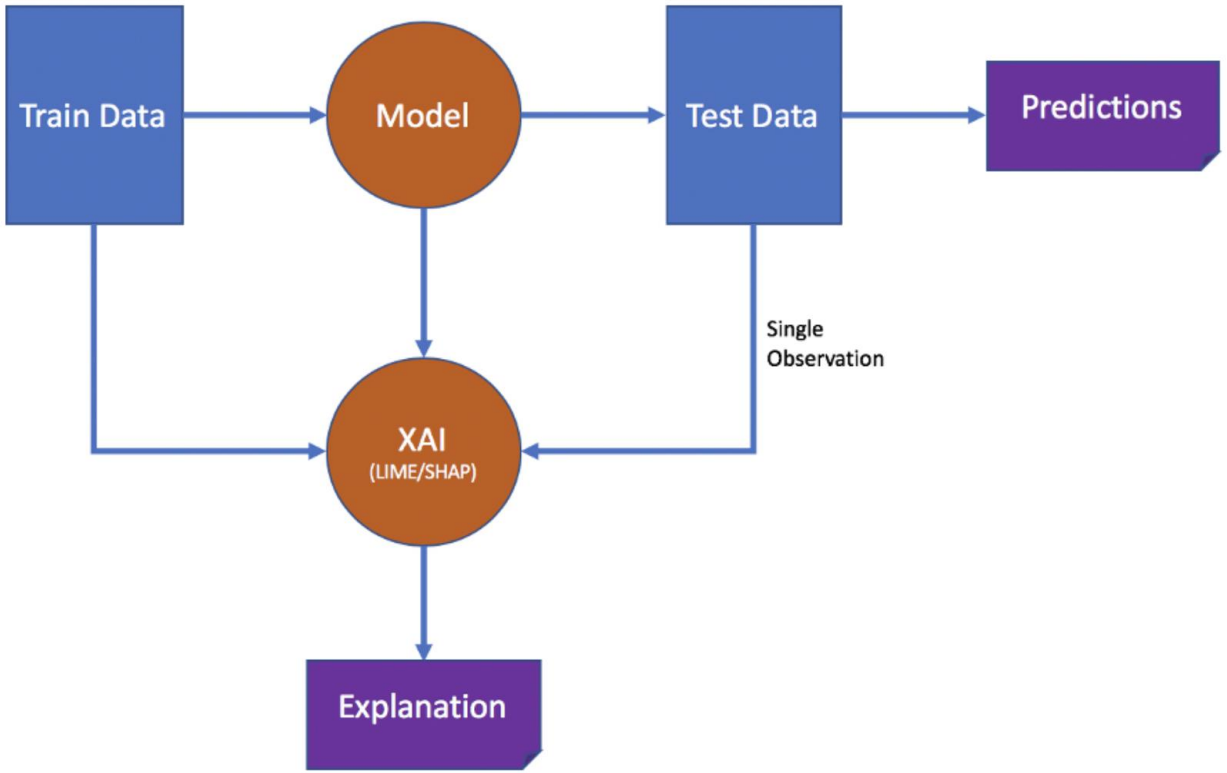
Why Explainable AI



FLOW CHART

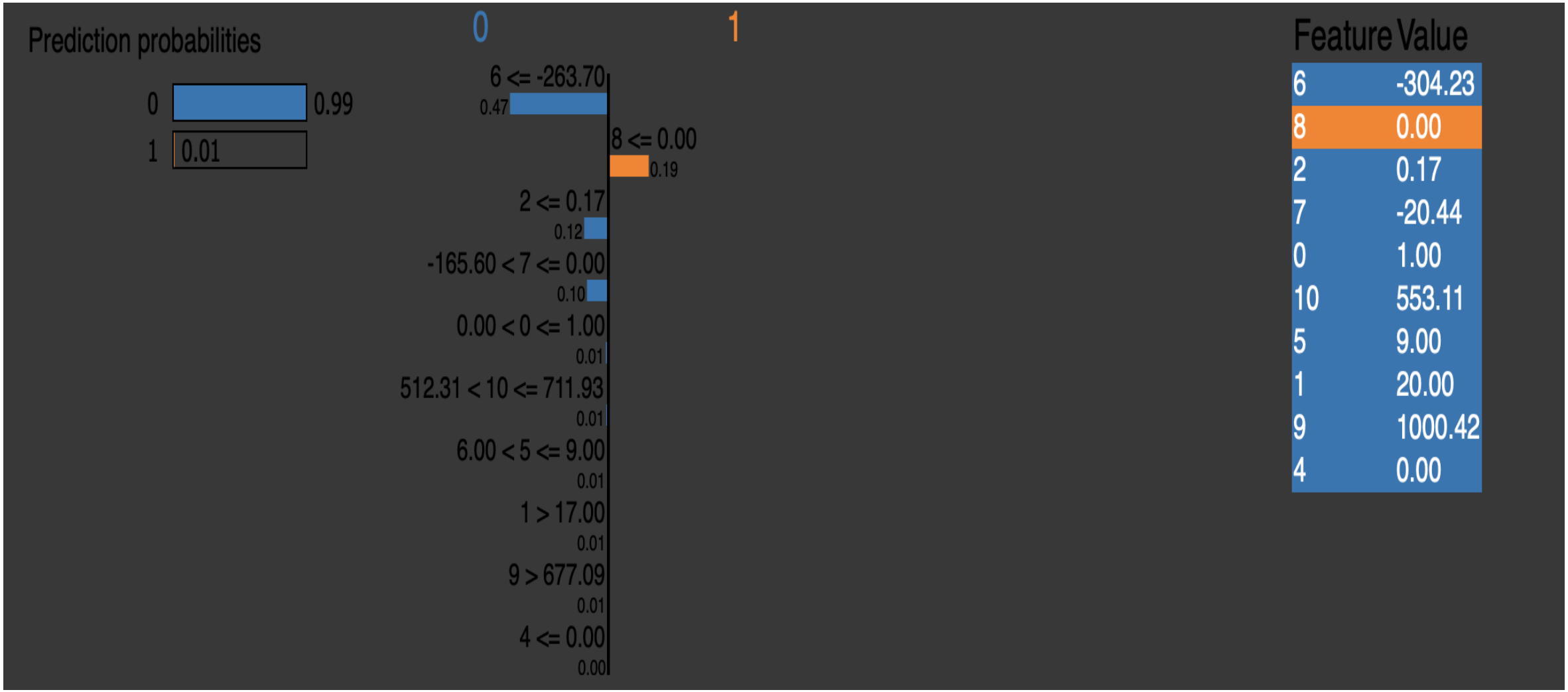

Claim Features
Prediction
Other Information from e.g., historical data or calls with the insured
Explanation
Claim Submitted
Fraud Detection Model (automated system)
Fraud suspected?
yes
Extract Explanation (decision support system)
Fraud expert investigates claim
Fraud suspected?
yes
Claim Rejected
no
Claim Accepted
no
Claim Accepted

METHODOLOGY

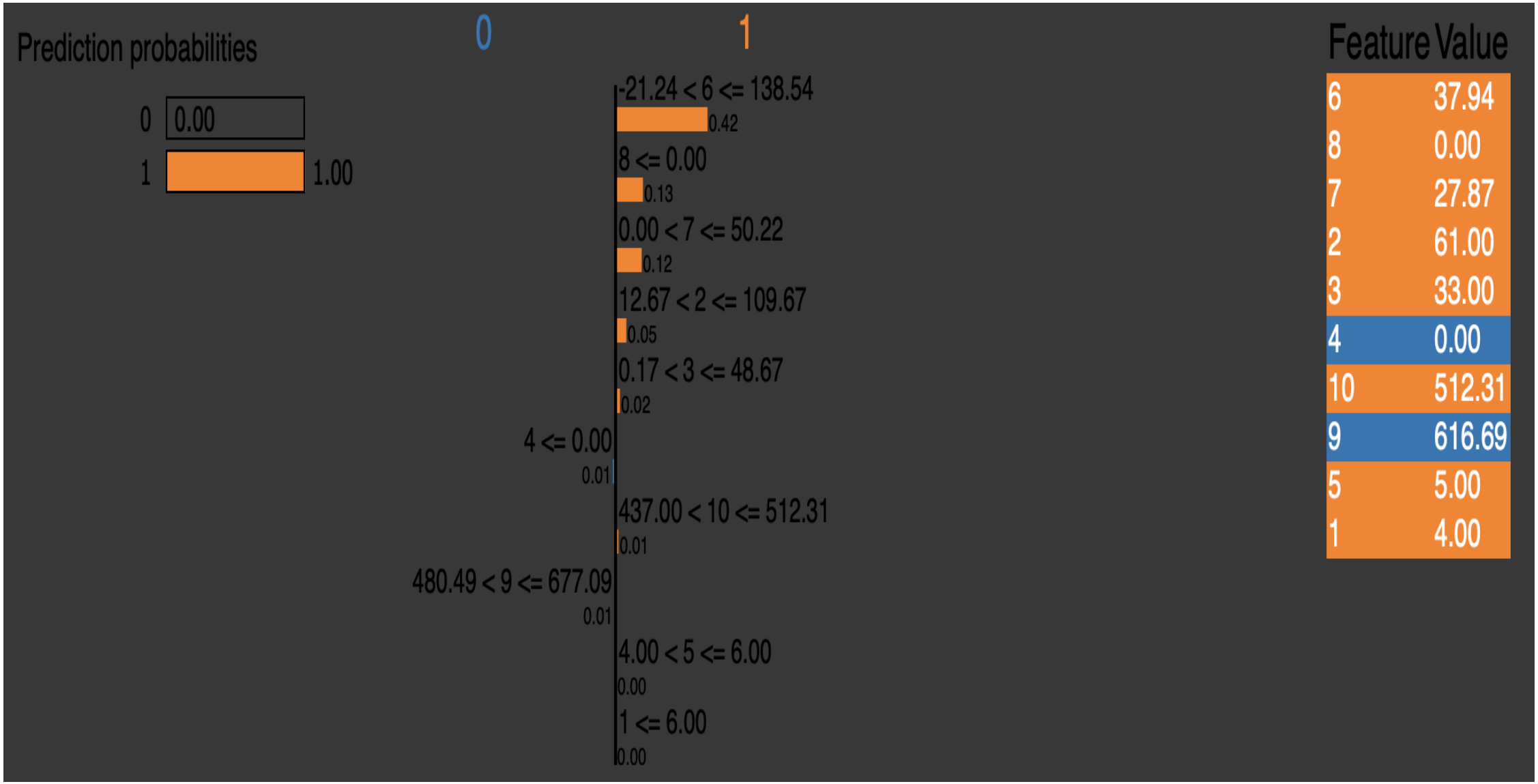


Interpret ML
SHAP
DLIME
Anchors

KEY RESULTS WITH INTERPRETATION



KEY RESULTS WITH INTERPRETATION



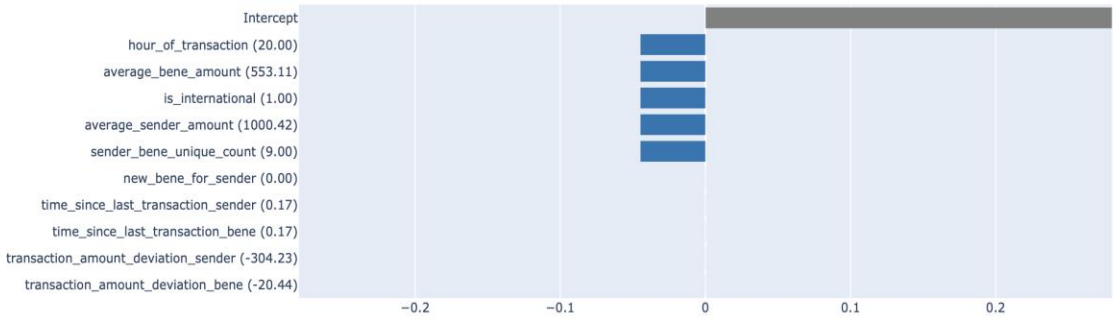
KEY RESULTS WITH INTERPRETATION

Select Component to Graph

0 : Actual (1) | Predicted (0.009999999776482582) | Resid (0.99)

LIME

Actual: 1 | Predicted: 0.0104

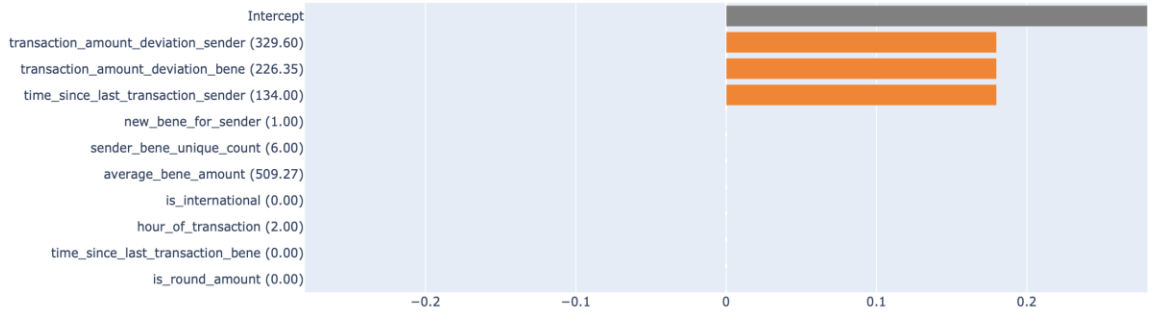


Select Component to Graph

1 : Actual (0) | Predicted (1) | Resid (-1)

LIME

Actual: 0 | Predicted: 1

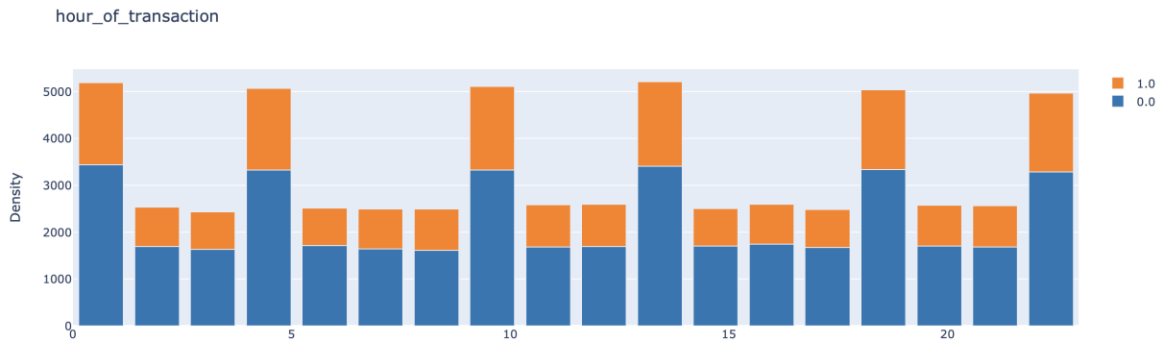


KEY RESULTS WITH INTERPRETATION

Select Component to Graph

1 : Name (hour_of_transaction) | Type (continuous) | # Unique (24)

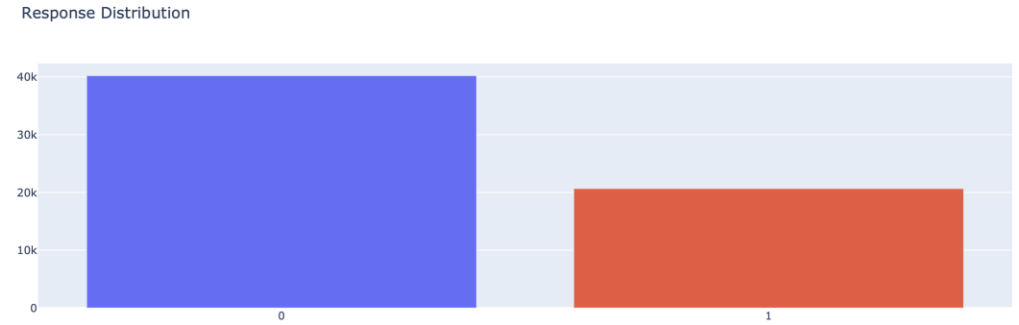
Train Data



Select Component to Graph

Summary

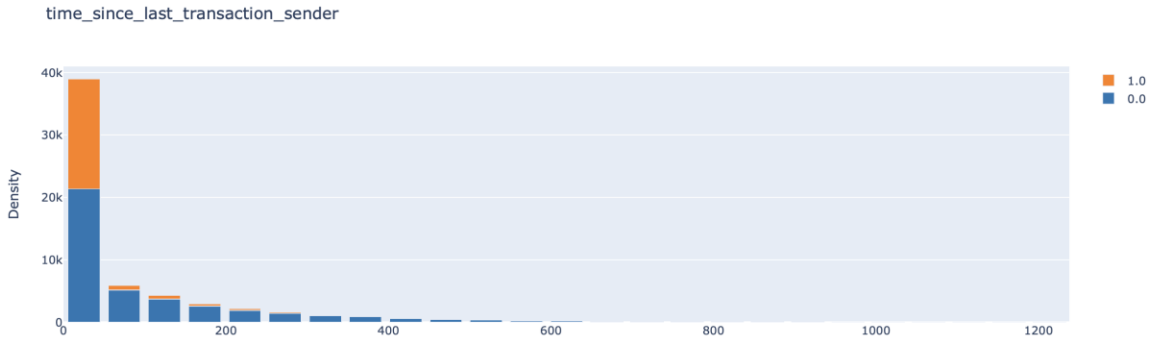
Train Data



Select Component to Graph

2 : Name (time_since_last_transaction_sender) | Type (continuous) | # Unique (2673)

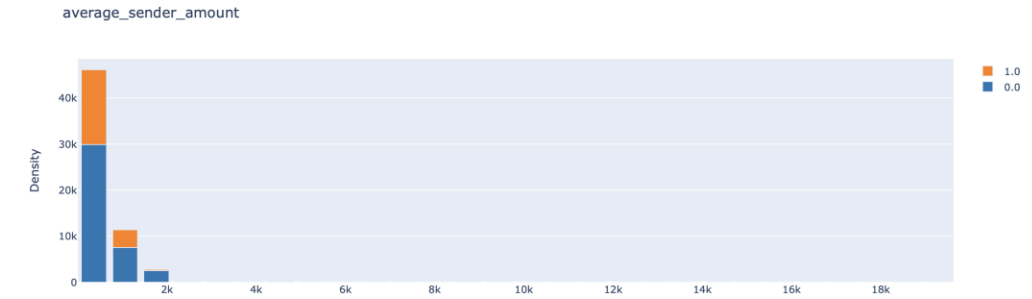
Train Data

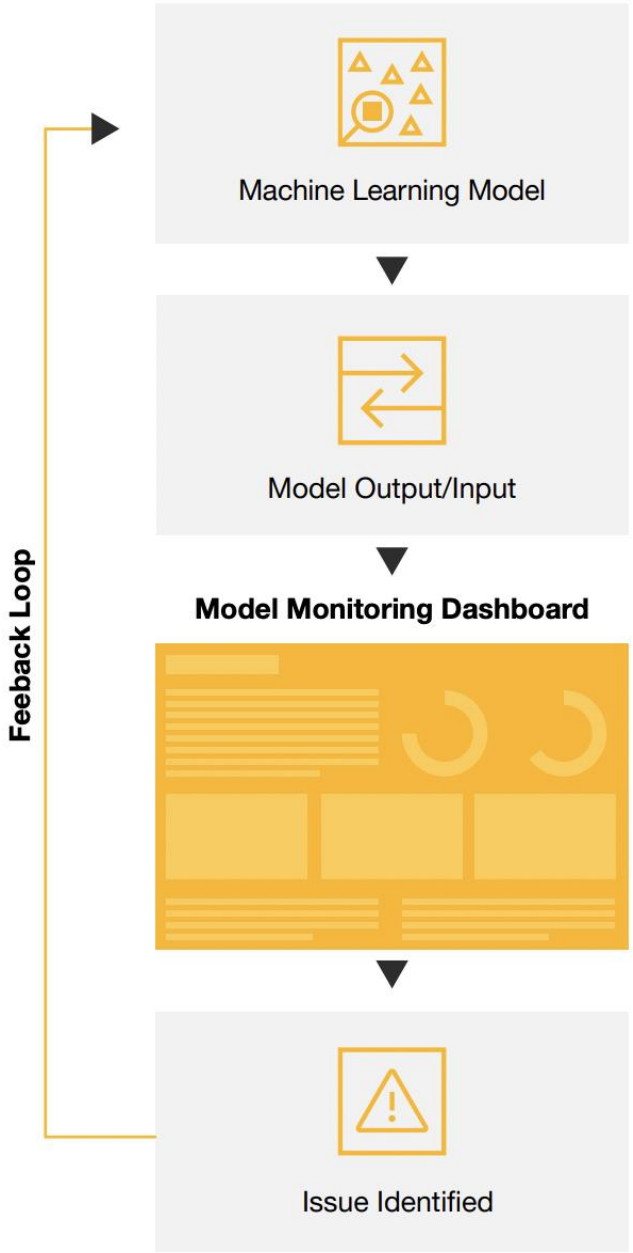


Select Component to Graph

9 : Name (average_sender_amount) | Type (continuous) | # Unique (13624)

Train Data





CONCLUSION

- **Incorporation of Explainable Artificial Intelligence (XAI) establishes a systematic feedback loop between the Machine Learning Model and the Model Monitoring.**
- **Swift detection of inconsistencies or errors in model outputs.**
- **Continuous feedback mechanism can lead to a significant reduction in false alarms.**
- **Model becomes progressively more accurate in predictions and decisions.**


QUESTIONS ??


THANK YOU