Fraud Stops with Oracle Database 26ai Vector Search + OML

Abi Giles-Haigh, Analytics & Innovation Director, Capgemini

Past & Future TechCasts:

February 5th

Our Favorite Features of OAC: November 2025 & January 2026 Releases

Presented by Tim Vlamis, Branden Pavol, Wayne Van Sluys, Gautam Pisharam & Taiwo Ajayi

March 5th

Fraud Stops with Oracle Database 26ai Vector Search + OML

Presented by Abi Giles-Haigh

April 2nd

Build Practical AI Prototypes with Oracle Analytics (Even with a Small Team)

Presented by Dennis Gray & Adam Hume

TechCast Archive

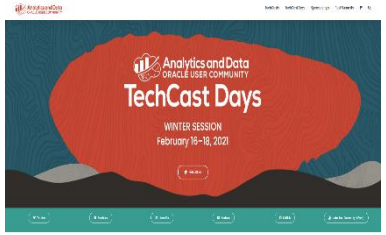
2026 | 2025 | 2024 | 2023 | 2022 | 2021

Date	Title	Presenter(s)	Replay	Download(s)
Feb 5	Our Favorite Features of OAC: November 2025 & January 2026 Releases	Branden Pavol, Gautam Pisharam, & Carter Beaton	Video	Slides
Jan 8	The Oracle AI Microservices Sandbox for RAG Rapid Prototyping	Melli Annamalai & Phani Chilakapati	Video	Slides

Submit a topic to share at **https://andouc.org/techcasts/**

Analytics and Data ORACLE USER COMMUNITY

Let's Connect



Website
http://andouc.org/

Chat with the Experts
https://bit.ly/Join-ANDOUC-Slack

Watch Previous TechCasts
https://bit.ly/3qmGgHN

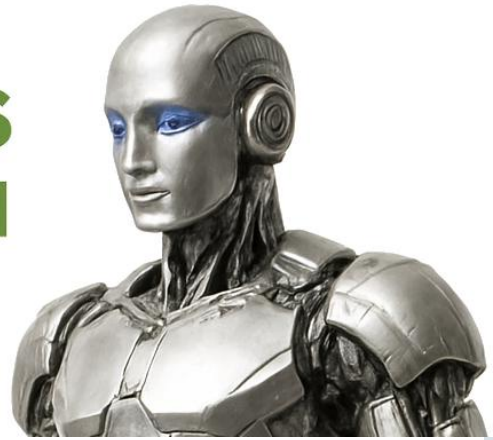
https://www.linkedin.com/company/analytics-and-data-oracle-user-community

Spatial + Graph SIG
bit.ly/Spatial-Graph-LinkedIn

@analyticsndataouc.bsky.social

ANALYTICS AND DATA TechCasts

Analytics and Data SUMMIT 2026



Only 5 weeks to go!
April 14-16, 2026

Scan this for full information:

ANALYTICS AND DATA TechCasts

Past & Future TechCasts:

February 5th

Our Favorite Features of OAC: November 2025 & January 2026 Releases

Presented by Tim Vlamis, Branden Pavol, Wayne Van Sluys, Gautam Pisharam & Taiwo Ajayi

March 5th

Fraud Stops with Oracle Database 26ai Vector Search + OML

Presented by Abi Giles-Haigh

April 2nd

Build Practical AI Prototypes with Oracle Analytics (Even with a Small Team)

Presented by Dennis Gray & Adam Hume

TechCast Archive

2026 | 2025 | 2024 | 2023 | 2022 | 2021

Date	Title	Presenter(s)	Replay	Download(s)
Feb 5	Our Favorite Features of OAC: November 2025 & January 2026 Releases	Branden Pavol, Gautam Pisharam, & Carter Beaton	Video	Slides
Jan 8	The Oracle AI Microservices Sandbox for RAG Rapid Prototyping	Melli Annamalai & Phani Chilakapati	Video	Slides

Submit a topic to share at **https://andouc.org/techcasts/**

FRAUD STOPS HERE

Oracle Database 26ai

Vector Search + OML

for Payment Fraud Detection

DR ABI GILES-HAIGH

Introductions

The Fraud Landscape: Scale & Why Rules Fail

$485B

Global payment fraud losses projected by 2027

73%

Fraud bypasses rule-based detection systems

<200ms

Window to decline a payment transaction

Why Rules Fail

Fraudsters clone legitimate transaction profiles — same MCC, similar amounts, matching time windows

Device fingerprinting alone misses account takeover via legitimate devices

Rule-based systems generate 70-90% false positives, burning analyst capacity

Unstructured intelligence (SARs, fincrime reports) sits siloed from scoring engines

A New Approach

Multi-source signal fusion is essential: structured transactions + unstructured reports

Vector similarity finds semantic near-neighbours in SARs without keyword matching

In-database scoring eliminates ETL latency — score at the point of transaction

Converged database: single engine, single security boundary, no data drift

Fraud Typologies: What We Are Actually Detecting

Mule Networks

8+ accounts sharing 2-3 device fingerprints

Rapid layering: receive → convert → withdraw

Signal: shared device + high velocity + new payees

Structuring / Smurfing

Multiple cash deposits just below £10,000 CTR threshold

Spread across branches, ATMs, and time windows

Signal: amount clustering + multi-location pattern

Account Takeover

Credential change + new payee + immediate transfer

New device / IP from different country than cardholder

Signal: device mismatch + geo anomaly + velocity spike

Trade-Based ML

Invoices for goods/services at prices far above market

Shell companies in low-transparency jurisdictions

Signal: MCC 4829 / SAR similarity to known TBML cases

Standard Data - Transactions

ACCOUNT_ID	ACCOUNT_TYPE	OPEN_DATE	RISK_BAND	COUNTRY_CODE	STATUS	ANNUAL_INCOME	KYC_REFRESH_DA
1	BUSINESS_CURRENT	2024-07-14T00:...	LOW	US	ACTIVE	20-50K	2025-12-07T00:...
2	PERSONAL_CURRENT	2025-11-24T00:...	MEDIUM	US	ACTIVE	20-50K	2025-10-01T00:...
3	BUSINESS_CURRENT	2022-01-23T00:...	MEDIUM	GB	ACTIVE	>250K	2025-06-28T00:...
4	PERSONAL_CURRENT	2024-07-08T00:...	MEDIUM	FR	ACTIVE	<20K	2025-11-08T00:...
5	SAVINGS	2025-03-16T00:...	LOW	HK	ACTIVE	50-100K	2025-12-07T00:...
6	PERSONAL_CURRENT	2024-01-24T00:...	MEDIUM	US	ACTIVE	>250K	2025-09-15T00:...
7	PERSONAL_CURRENT	2023-01-26T00:...	HIGH	GB	ACTIVE	100-250K	2025-12-19T00:...
8	PERSONAL_CURRENT	2021-03-31T00:...	HIGH	NL	ACTIVE	20-50K	2025-12-24T00:...
9	BUSINESS_CURRENT	2021-09-28T00:...	HIGH	GB	ACTIVE	<20K	2025-10-01T00:...
10	SAVINGS	2023-07-15T00:...	MEDIUM	US			
11	BUSINESS_CURRENT	2022-04-27T00:...	LOW	SG			
12	BUSINESS_CURRENT	2024-09-14T00:...	MEDIUM	DE			
13	SAVINGS	2021-05-08T00:...	HIGH	FR			

CARD_ID	ACCOUNT_ID	CARD_TYPE	ISSUED_DATE	EXPIRY_DATE	CARD_STATUS	CONTACTLESS_EI
1	150	DEBIT	2022-11-01T00:0...	2026-12-01T00:...	ACTIVE	1
2	303	DEBIT	2023-03-25T00:...	2026-12-01T00:...	ACTIVE	1
3	78	CREDIT	2025-12-16T00:...	2030-01-01T00:...	ACTIVE	1
4	487	PREPAID	2024-02-12T00:...	2026-11-01T00:0...	ACTIVE	1
5	54	PREPAID	2024-11-04T00:...	2027-07-01T00:...	EXPIRED	1
6	365	CREDIT	2024-06-04T00:...	2026-01-01T00:...	ACTIVE	0
7	359	DEBIT	2023-11-28T00:...	2028-12-01T00:...	ACTIVE	1
8	327	DEBIT	2021-05-25T00:...	2026-02-01T00:...	ACTIVE	0
9	340	CREDIT	2025-04-06T00:...	2028-05-01T00:...	ACTIVE	0
10	359	PREPAID	2021-05-01T00:...	2025-06-01T00:...	ACTIVE	1
11	484	DEBIT	2024-11-09T00:...	2027-10-01T00:...	ACTIVE	1
12	231	DEBIT	2023-05-20T00:...	2027-07-01T00:...	ACTIVE	1

Standard Data - Transactions

TXN_ID	TXN_TIMESTAMP	ACCOUNT_ID	CARD_ID	MERCHANT_ID	TXN_AMOUNT	CURRENCY_CODE	MERCHANT_CATE	CARD_P
37295	2025-12-18T16:2...	222	185	215	133.98	GBP	4814	
37296	2026-01-14T12:5...	242	58	119	54.52	EUR	5411	
37297	2025-09-13T03:...	84	383	178	169.87	EUR	5999	
37298	2025-02-27T18:...	85	112	126	61.92	EUR	7011	
37299	2026-01-19T00:...	113	540	241	166.19	EUR	6011	
37300	2025-06-27T15:...	106	117	131	78.15	EUR	5999	
37301	2025-04-19T21:...	15	591	219	1274.66	SGD	5045	
37302	2025-04-17T10:...	58	469	127	3453.38	USD	5045	
37303	2026-02-03T15:...	3	431	3	59.35	GBP	5541	
37304	2025-03-24T19:...	327	57	214	1749	EUR	6051	
37305	2025-09-04T22:...	395	253	28	60.75	EUR	5812	
37306	2025-10-30T23:...	109	193	249	330.87	HKD	7011	
37307	2025-06-26T05:...	250	316	50	257.85	SGD	5944	
37308	2026-02-11T23:...	88	566	278	261.49	EUR	7011	

TXN_ID	TXN_TIMESTAMP	ACCOUNT_ID	CARD_ID	MERCHANT_ID	TXN_AMOUNT	CURRENCY_CODE	MERCHANT_CATE	CARD_P
37295	2025-12-18T16:2...	222	185	215	133.98	GBP	4814	
37296	2026-01-14T12:5...	242	58	119	54.52	EUR	5411	
37297	2025-09-13T03:...	84	383	178	169.87	EUR	5999	
37298	2025-02-27T18:...	85	112	126	61.92	EUR	7011	
37299	2026-01-19T00:...	113	540	241	166.19	EUR	6011	
37300	2025-06-27T15:...	106	117	131	78.15	EUR	5999	
37301	2025-04-19T21:...	15	591	219	1274.66	SGD	5045	
37302	2025-04-17T10:...	58	469	127	3453.38	USD	5045	
37303	2026-02-03T15:...	3	431	3	59.35	GBP	5541	
37304	2025-03-24T19:...	327	57	214	1749	EUR	6051	
37305	2025-09-04T22:...	395	253	28	60.75	EUR	5812	
37306	2025-10-30T23:...	109	193	249	330.87	HKD	7011	
37307	2025-06-26T05:...	250	316	50	257.85	SGD	5944	
37308	2026-02-11T23:...	88	566	278	261.49	EUR	7011	

Suspicious Activity Report

SAR_ID	SAR_REFERENCE	SAR_EMBEDDING	CREATED_AT	JURISDICTION	ACTIVITY_TYPE	RISK_LEVEL	NARRATIVE	ENTITY_TYPE
1	SAR-2025-100001	[3.64899993E-0...	10/15/2025, 09:0...	SG	STRUCTURING	HIGH	Subject Horizon ...	COMPANY
2	SAR-2025-100002	[7.63599992E-0...	07/14/2025, 09:0...	AE	SANCTIONS_EVA...	HIGH	Ahmed Khalid rec...	COMPANY
3	SAR-2024-100003	[3.63599986E-0...	03/13/2024, 09:0...	LU	ACCOUNT_TAKE...	LOW	Rapid credential ...	COMPANY
4	SAR-2024-100004	[8.99800003E-0...	07/27/2024, 09:0...	DE	SANCTIONS_EVA...	HIGH	Account operate...	COMPANY
5	SAR-2025-100005	[6.0680002E-00...	07/18/2025, 09:0...	NL	SANCTIONS_EVA...	HIGH	Account operate...	INDIVIDUAL
6	SAR-2025-100006	[1.61799997E-00...	02/23/2025, 09:0...	HK	STRUCTURING	MEDIUM	Subject Summit F...	COMPANY
7	SAR-2025-100007	[1.59199998E-0...	10/29/2025, 09:0...	CH	MONEY_LAUNDE...	HIGH	Account held by ...	INDIVIDUAL
8	SAR-2024-100008	[7.61799991E-00...	08/01/2024, 09:0...	MT	SANCTIONS_EVA...	HIGH	Horizon Financial...	COMPANY
9	SAR-2024-100009	[2.6030001E-001...	04/28/2024, 09:...	HK	STRUCTURING	HIGH	Multiple individua...	INDIVIDUAL
10	SAR-2025-100010	[9.00000036E-0...	03/16/2025, 09:0...	SG	MONEY_LAUNDE...	HIGH	Viktor Sorokin ac...	INDIVIDUAL
11	SAR-2024-100011	[1.087E-001,-2.2...	09/20/2024, 09:...	US	MONEY_LAUNDE...	HIGH	Pacific Rim Advis...	INDIVIDUAL
12	SAR-2025-100012	[4.3720001E-001...	10/09/2025, 09:0...	FR	ACCOUNT_TAKE...	MEDIUM	Account belongin...	COMPANY
13	SAR-2025-100013	[-2.34999992E-...	12/20/2025, 09:0...	MT	WIRE_FRAUD	MEDIUM	Multiple wire tran...	COMPANY
14	SAR-2024-100014	[8.30999985E-0...	12/06/2024, 09:0...	DE	MONEY_LAUNDE...	HIGH	Account held by ...	INDIVIDUAL

HIGH

Suspicious Activity Report

Column "NARRATIVE" row 1

Subject Horizon Financial Group conducted 19 cash transactions over 24 days, each below the £10,000 reporting threshold. Transactions occurred at 4 different branches and ATMs, suggesting deliberate structuring to avoid CTR filing. Cumulative total: £368,299. Pattern consistent with smurfing operation.

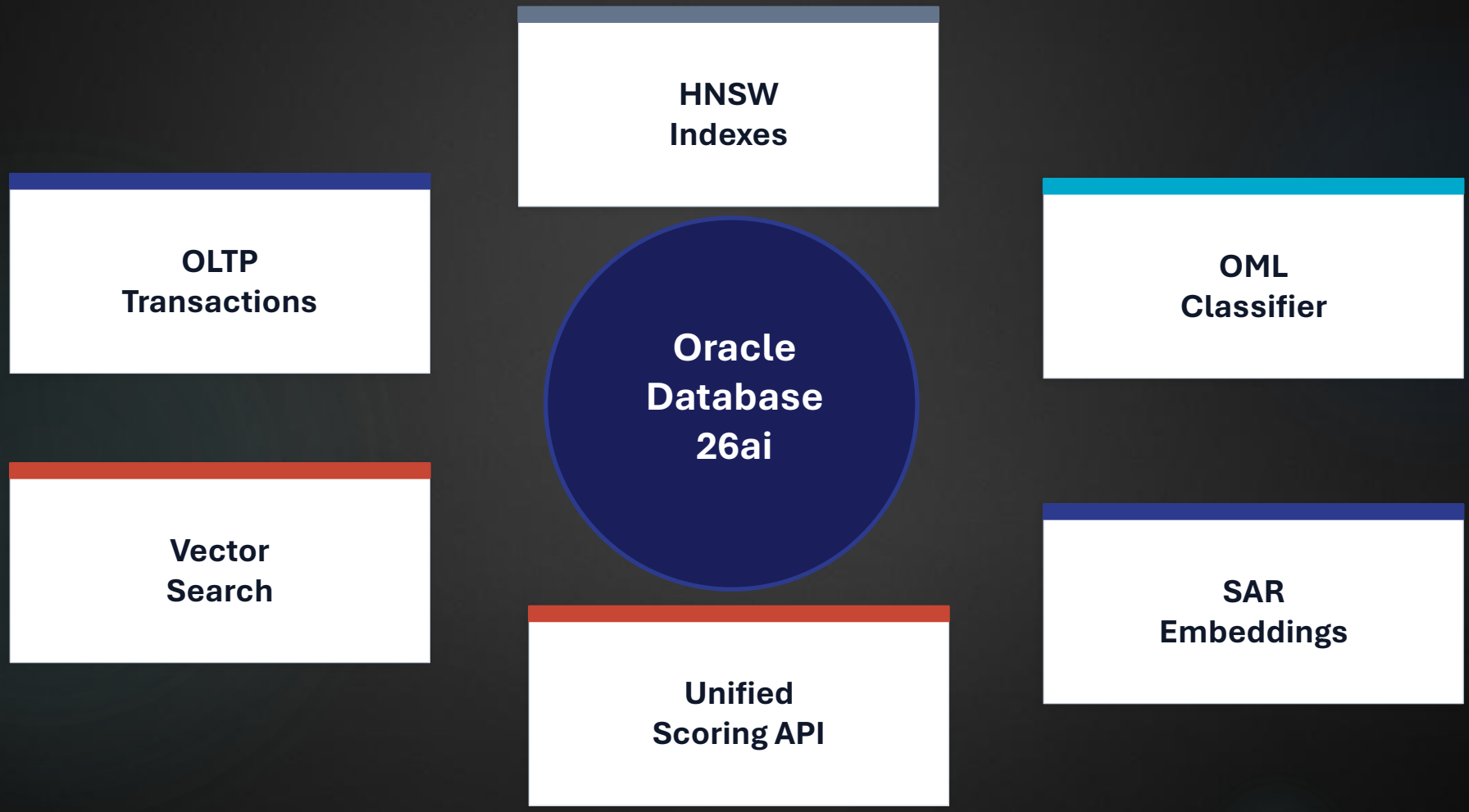
Column "NARRATIVE" row 2

Ahmed Khalid received funds via correspondent chain originating in GB subject to OFAC/HMT restrictions. Intermediary Nexus Advisory Ltd used to obscure nexus to designated entity. Total funds: £1,503,881 across 23 transactions. Nexus to designated individual identified via Adverse Media screen.

Column "NARRATIVE" row 3

Rapid credential changes followed by immediate high-value transfers from Offshore Capital Management account. New payee Vantage Point Consulting added and £2,472,245 transferred within 4 minutes of profile update. Access originated from AE while cardholder is resident in GB. Previous SAR filed on Vantage Point Consulting reference SAR-2021-64938.

Oracle Database 26ai: The Converged Architecture



← eliminates data movement, reduces latency, enforces single security model →

Vector Search: Concepts & Distance Metrics

1

Unstructured Text

SAR narratives, fincrime alerts, investigation notes

2

Embedding Model

Oracle ONNX or third-party model, stored in-DB

3

Vector Store

VECTOR column in Oracle table, HNSW index

4

Similarity Query

VECTOR_DISTANCE() → top-k nearest SARs

Distance Metric Choice

★ Recommended for SAR search

COSINE	Best for text/SAR narratives, measures angle between vectors, ignores magnitude
DOT_PRODUCT	Faster than cosine when embeddings are normalised to unit length
EUCLIDEAN	Useful when magnitude matters, less common for NLP embeddings

Embedding Models: Choosing & Importing into Oracle

Model	Dims	Size	Domain	Rec. Use Case
`all-MiniLM-L6-v2`	384	80MB	General English	Fast baseline; good recall@10 on SARs
`all-mpnet-base-v2`	768	420MB	General English	Higher accuracy; slower; good for investigations
`FinBERT (fine-tuned)`	768	440MB	Financial Text	Best for SAR/AML narrative similarity ★
`text-embedding-3-small`	1536	API	General	OpenAI API; not stored in-DB natively
`e5-large-v2`	1024	1.3GB	Multi-task	Strong cross-lingual SAR matching

I want to vector easily....

```
BEGIN
DBMS_VECTOR.DROP_ONNX_MODEL(
        model_name => 'ALL_MINILM_L12_V2',
        force => true
    );
        EXCEPTION WHEN OTHERS THEN NULL;
END;
/


-- Load directly from Oracle's public URL
BEGIN
DBMS_VECTOR.LOAD_ONNX_MODEL_CLOUD(
        model_name => 'ALL_MINILM_L12_V2',
        credential => NULL,
        uri => 'https://objectstorage.us-ashburn-1.oraclecloud.com/n/adwc4pm/b/OML-
        Resources/o/all_MiniLM_L12_v2.onnx'
    );
END;
/
```

Test the Embedding Model

```
SELECT

VECTOR_EMBEDDING(
    ALL_MINILM_L12_V2 USING 'The quick brown fox' AS DATA
    ) as V_Data
FROM DUAL;
```

V_DATA
[-4.20517325E-002,3.17923352E-002,-7.66987503E-002,-2.84928624E-002,8.20773616E-002,8.69272184E-003,2.37550121E-002,5.82369685E-004,-3.05419974E-0...

Applying to a column

```
ALTER TABLE
suspicious_activity_reports
ADD sar_embedding VECTOR(384,
FLOAT32);

UPDATE
suspicious_activity_reports
SET sar_embedding =
VECTOR_EMBEDDING(
ALL_MINILM_L12_V2 USING
sar_text AS DATA
)
WHERE sar_embedding IS NULL;


COMMIT;
```

```
CREATE VECTOR INDEX
sar_hnsw_idx
ON
suspicious_activity_reports
    (
        sar_embedding
    )
ORGANIZATION INMEMORY NEIGHBOR
GRAPH
DISTANCE COSINE
WITH TARGET ACCURACY 95;
```

OML Feature Engineering: Building the Fraud Signal

Transaction Behaviour

- txn_amount_zscore — amount vs. 90-day account history
- velocity_1h_count / velocity_24h_count
- time_since_last_txn_mins
- card_present_flag (CNP = higher risk)

Entity & Device

- device_fingerprint_hash match flag
- device_shared_account_count (mule signal)
- geo_distance_km from home country
- new_payee_flag (first time beneficiary)

Merchant Context

- merchant_category (MCC code)
- merchant_fraud_rate_90d
- merchant_country_risk_band
- mcc_amount_deviation (vs. MCC avg)

Vector Intelligence ★

- sar_similarity_score — max cosine sim to SAR store
- sar_match_count — SARs above 0.6 threshold
- sar_top_activity_type (encoded)
- sar_entity_name_match_flag

★ The SAR vector features bridge unstructured intelligence into the structured ML pipeline — this is the architectural innovation

Build a Model

```
DECLARE
v_setlst DBMS_DATA_MINING.SETTING_LIST;
BEGIN
    v_setlst('PREP_AUTO') := 'ON';
    v_setlst('ALGO_NAME') := 'ALGO_RANDOM_FOREST';
    DBMS_DATA_MINING.CREATE_MODEL2(

        MODEL_NAME => 'RF_FRAUD_MODEL',
        MINING_FUNCTION => 'CLASSIFICATION',
        DATA_QUERY => 'SELECT * FROM TRAIN_DATA_TXN',
        CASE_ID_COLUMN_NAME => 'TXN_ID',
        SET_LIST => v_setlst,
        TARGET_COLUMN_NAME => 'FRAUD_LABEL'
        );
END;
```

OML Classifier: Training Inside Oracle Database

Training Feature Set

Feature	Type	Importance
txn_amount_zscore	Numeric	High
merchant_category	Categorical	High
device_fingerprint_match	Boolean	Medium
time_since_last_txn_mins	Numeric	Medium
card_present_flag	Boolean	Medium
sar_similarity_score	Numeric	High ★
velocity_24h_count	Numeric	High
geo_distance_km	Numeric	Medium

★ *SAR similarity score bridges vector search → OML*

Algorithm Selection

GLM
Interpretable, regulator-friendly

Random Forest ★ Demo
Best balance: accuracy + explainability

XGBoost
Highest AUC on fraud benchmarks

OML Workflow (In-Database)

1. Create Settings	DBMS_DATA_MINING.CREATE_MODEL
2. Train Model	Uses in-DB parallelism (DOP=8)
3. Batch Score	PREDICTION() SQL function
4. Real-Time Score	REST endpoint via ORDS

Apply the Model

```
BEGIN
    DBMS_DATA_MINING.APPLY
        (

    'RF_FRAUD_MODEL','TEST_DATA_TXN','TXN_ID','APPLY_RESULT_RF’
        );
END;
```

Demo 2 Results: Measuring the Impact of SAR Similarity

Model Performance Comparison (Holdout Batch: 1,000 transactions)

Metric	V2 (OML Only)	V1 (OML + SAR Vector)	Improvement
AUC-ROC	0.871	0.934	+7.2%
KS Statistic	0.512	0.641	+25.2%
Precision@10%FPR	68.4%	84.1%	+15.7pp
False Positive Rate	8.3%	4.1%	−50.6%
False Negative Rate	19.2%	11.4%	−40.6%

Confusion Matrices

V2 — OML Only

	Pred 0	Pred 1
Act 0	**457**	**43**
Act 1	**96**	**404**

V1 — OML + SAR Vector

	Pred 0	Pred 1
Act 0	**479**	**21**
Act 1	**57**	**443**

Combined Scoring: Vector Similarity + OML Classifier

fraud_score = α × PREDICTION_PROBABILITY(oml_model USING *) + β × (1 - min_sar_vector_distance)

Score Range	Action	SLA	Expected FPR
0.0 – 0.4	Approve	<50ms	~0.2%
0.4 – 0.7	Manual Review	<500ms	~3.5%
0.7 – 1.0	Auto-Decline	<50ms	~0.3%

Overview

Fraud is a Multi-Source Problem

No single signal wins. Rule engines, OML classifiers, and vector search on unstructured SARs each add non-overlapping signal ,their combination drives AUC from 0.87 to 0.93+

Converged Database Eliminates Friction

Keeping vectors, transactions, and ML models in one engine eliminates ETL latency, data drift between stores, and security surface area across 5+ systems

Vector Search Changes the Investigation Paradigm

Semantic near-neighbour search finds guilt-by-association in SAR narratives without keyword matching, a new capability that simply does not exist in RDBMS-only architectures

In-Database OML Scales to Production

PREDICTION() SQL function scores millions of rows using parallel query. ORDS wraps scoring in a REST endpoint with no Python microservices or middleware

Governance is Not an Afterthought

PSI drift detection, feature importance tracking, and EXPLAIN PLAN on HNSW indexes give regulators the transparency they require

Dr Abi Giles-Haigh

AbiHaigh@Hotmail.com or Abi.Giles-Haigh@Capgemini.com

@Abi_Giles_Haigh

@ abigail-giles-haigh

Helpful Links –

ORACLE ANALYTICS VIDEOS:
https://www.youtube.com/@OracleAnalytics/videos

OAC NOVEMBER 2025 AND JANUARY 2026 NEW FEATURES VIDEOS BY ORACLE:
https://bit.ly/OACNov25Features
https://bit.ly/OACJan26Features

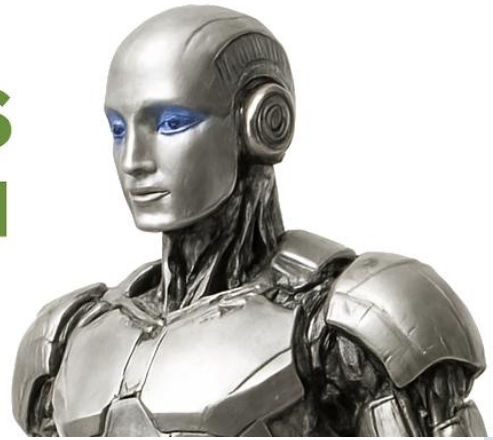
OAC NEW FEATURES DOCUMENTATION BY ORACLE:
https://docs.oracle.com/en/cloud/paas/analytics-cloud/acswn/index.html#GUID-CFF90F44-BCEB-49EE-B40B-8D040F02D476

ORACLE ANALYTICS COMMUNITY:
https://community.oracle.com/products/oracleanalytics

ORACLE ANALYTICS LIBRARY/EXAMPLES:
https://www.oracle.com/business-analytics/data-visualization/examples/

ORACLE ANALYTICS LIVE DEMOS:
https://www.oracle.com/business-analytics/data-visualization/demos/

Analytics and Data SUMMIT 2026



Only 5 weeks to go!
April 14-16, 2026

Scan this for full information:

ANALYTICS AND DATA TechCasts