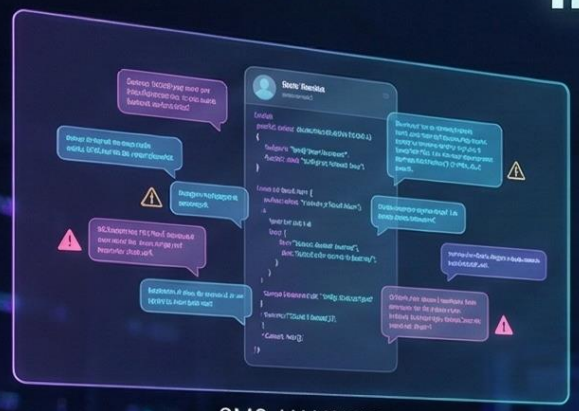
Investigation of Social Engineering Attacks (Call, Email, SMS) Using Machine Learning

Course: IT Incident Investigation

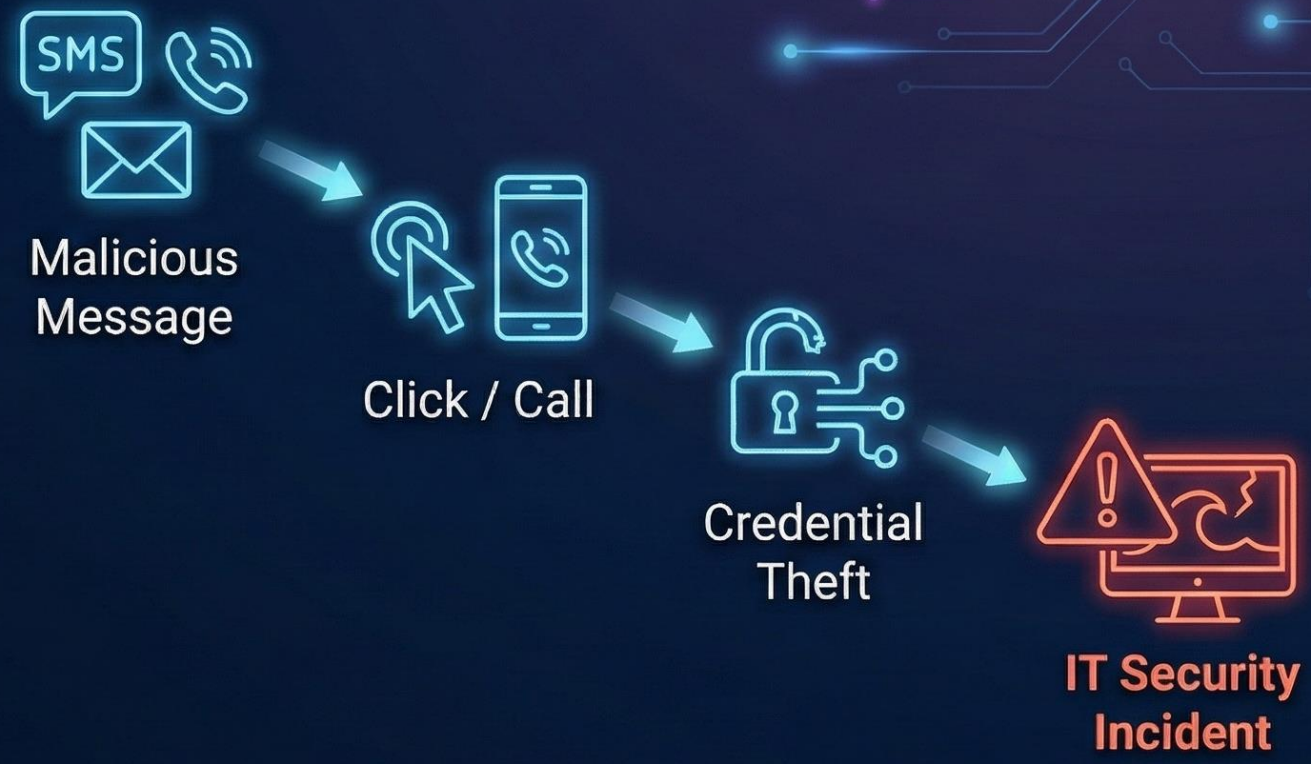


Students: Ali Altynbekov, Timur Kamalov | Group: CTC-241M | Instructor: Professor Dr. Abdul Razaque, PhD & Postdoc

Why Social Engineering Is an Incident Problem

Social engineering is often the entry point of major IT incidents

- Attacks target human psychology, not systems
- Lead to credential theft, financial fraud, and account takeover
- SMS, Email, and Phone Calls are the most common attack vectors



Social engineering attacks often initiate larger security incidents such as data breaches and financial loss

Investigation Perspective (SOC & Incident Response)

This project supports incident detection and investigation — not full automation.

Investigation Value

- Early detection of suspicious communication (SMS, Email, Calls)
- Triage & prioritization to reduce SOC analyst workload
- Extraction of Indicators of Compromise (IoCs) for evidence support

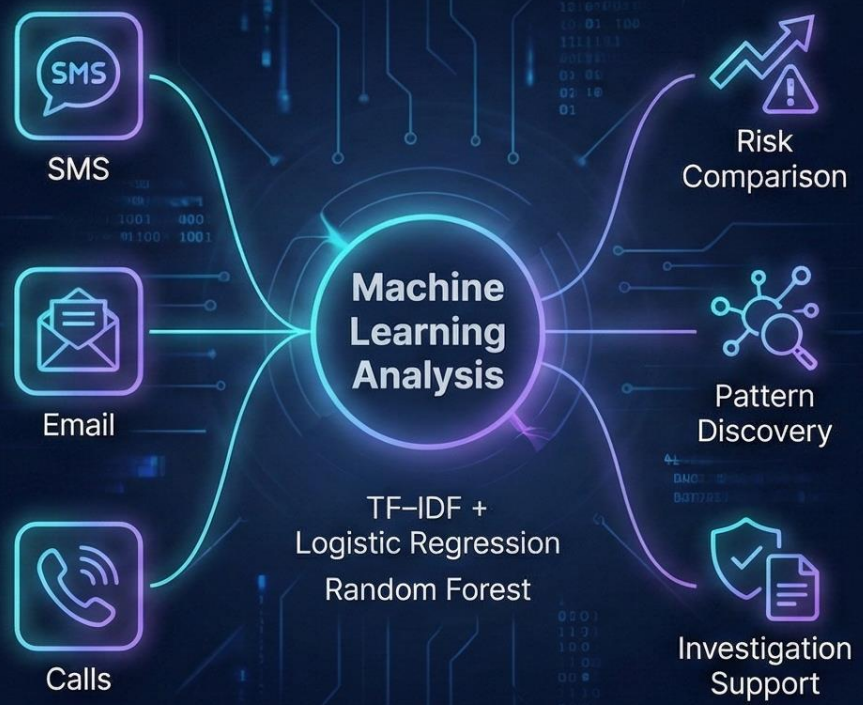
Goal: augment investigator decisions with interpretable signals, not replace analysts.

Project Goal & Research Questions

Project Goals

Detect malicious communication across three channels: SMS, Email, and Phone Calls

Compare machine learning models to support incident detection and investigation workflows



Research Questions

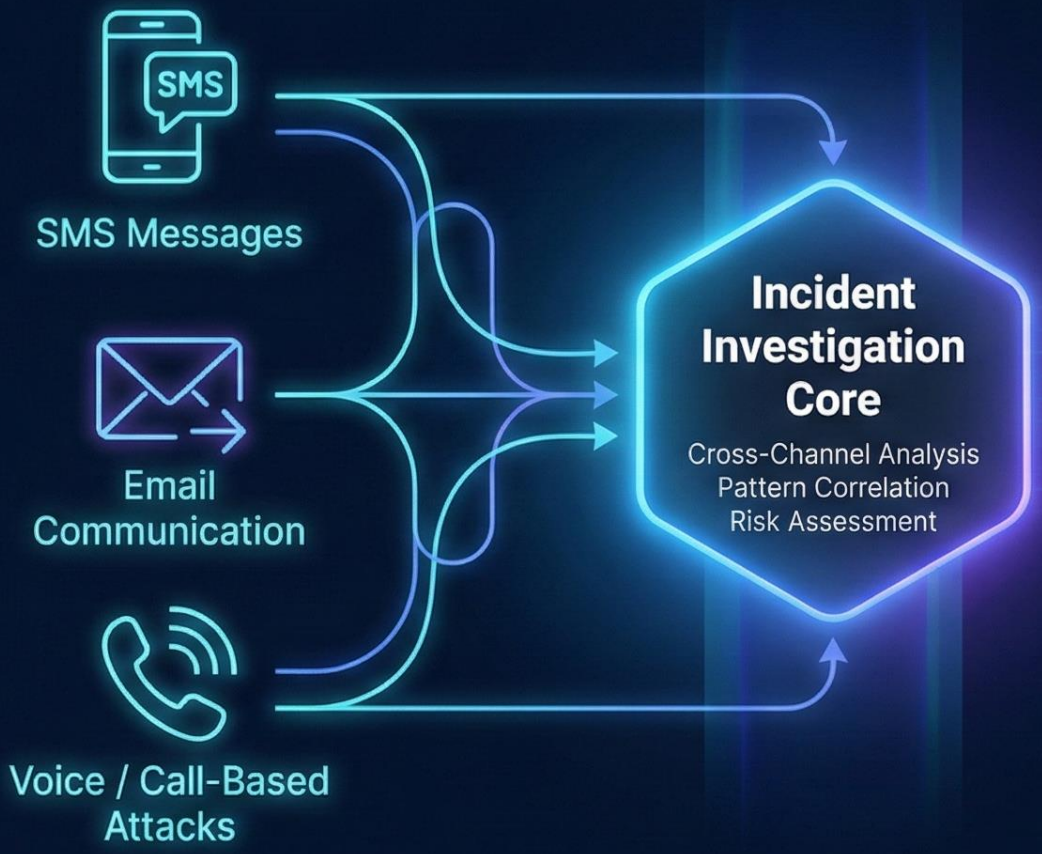
1. Which communication channel presents the highest risk from a social engineering perspective?
2. Which machine learning model performs better for incident investigation support, and why?
3. What linguistic patterns and behavioral indicators are typical of social engineering attacks?

These goals and questions guide the analysis of how machine learning can augment IT incident investigation rather than fully automate decisions.

Multi-Channel Attack Surface

Social engineering attacks rarely operate in isolation

- Attackers reuse the same techniques across multiple communication channels
- The same urgency, authority, and forced-action patterns appear in SMS, Email, and Phone Calls
- Different message formats require different modeling and investigation approaches



Understanding cross-channel attack behavior is critical for effective incident detection and investigation.

Datasets Overview (High-Level)

Three datasets representing different social engineering channels

Channel	Message Characteristics	Investigation Notes
 SMS	Short messages Highly imbalanced High noise level	Common entry point High volume Requires fast triage
 Email	Longer messages Rich textual context Many phishing samples	Strong indicators Links and impersonation Easier pattern detection
 Calls	Very small dataset Synthetic examples Short transcripts	Proof of concept Pipeline extensibility Limited statistical power

Each channel differs significantly in format and risk profile, requiring separate modeling and investigation strategies.

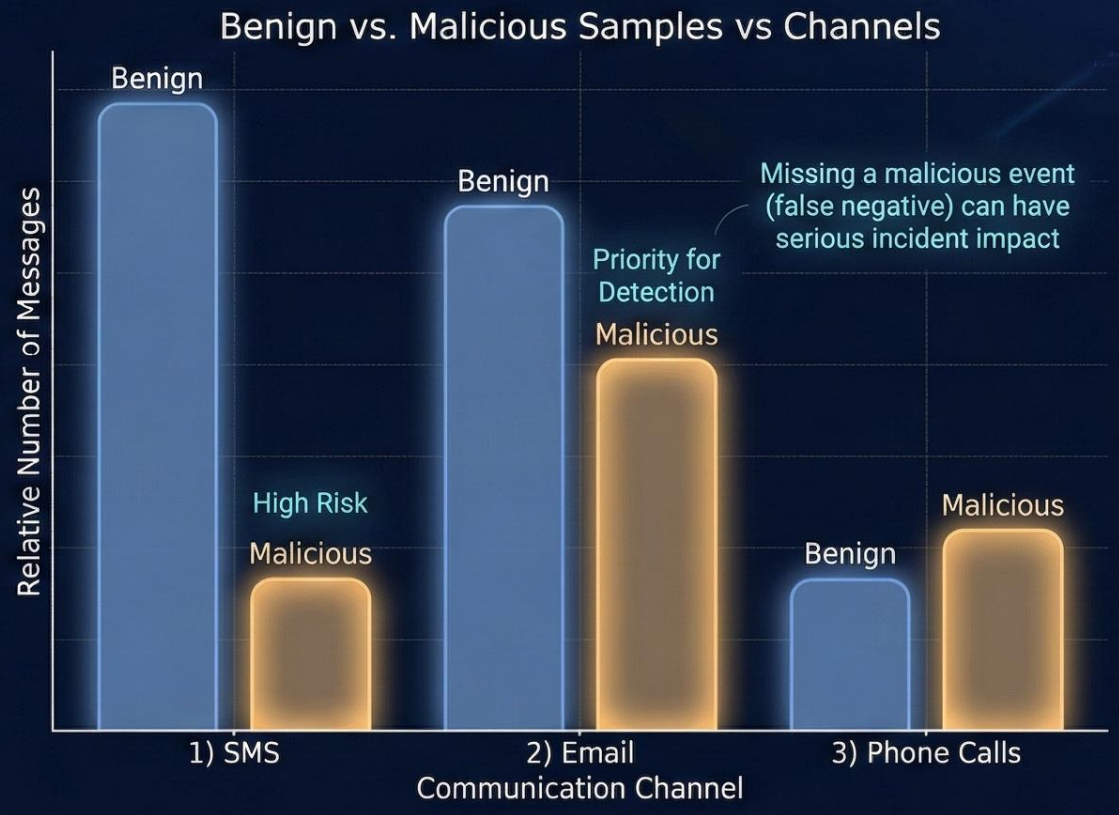
Class Distribution & Why It Matters

Security datasets rarely reflect balanced real-world conditions

Real-world security data is highly imbalanced, with far fewer malicious events than benign ones

Accuracy alone can be misleading when malicious samples are rare

Recall for the malicious class is critical to avoid missing active attacks



In incident investigation, minimizing false negatives is often more important than maximizing overall accuracy.

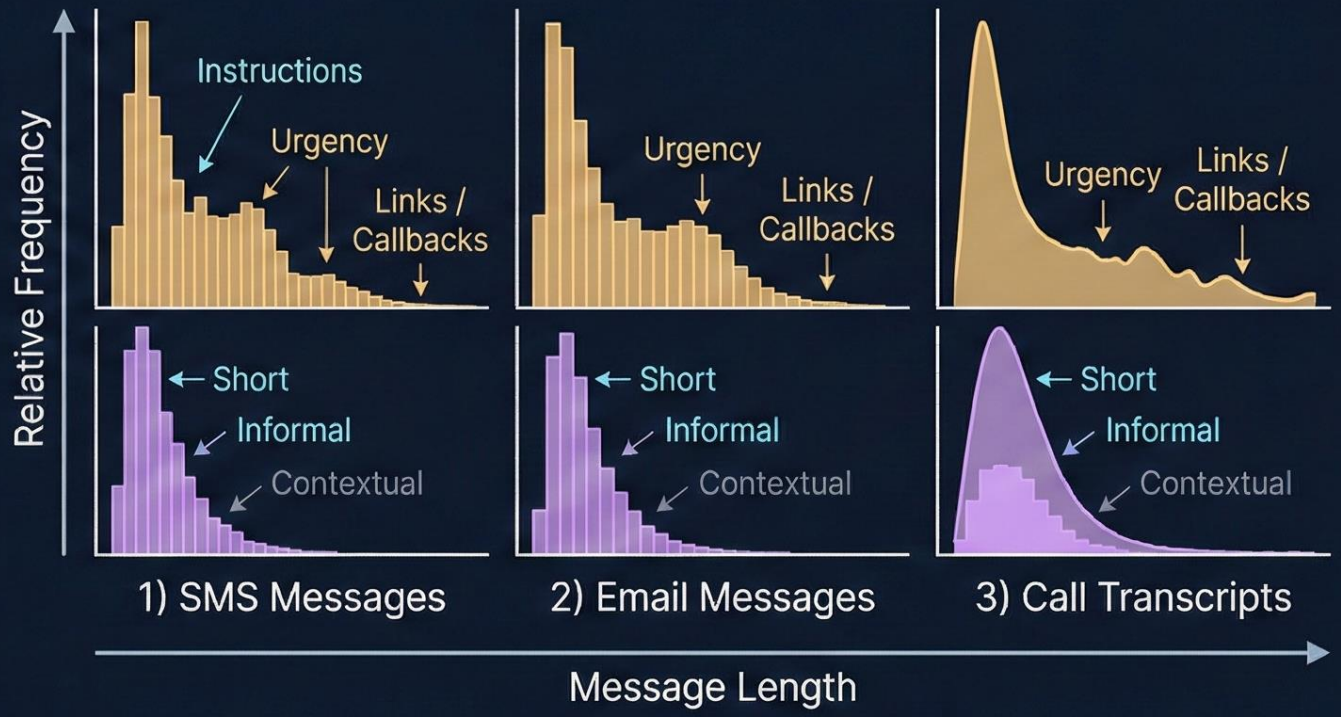
Message Length & Attacker Behavior

Message length and structure often reveal social engineering intent

Spam and phishing messages are often longer than legitimate communication

They frequently include step-by-step instructions, links, phone numbers, or follow-up actions

Attackers use pressure language to force quick decisions and reduce critical thinking



Longer, instruction-heavy messages are a strong behavioral indicator of social engineering across channels.

Preprocessing & Feature Engineering Engineering

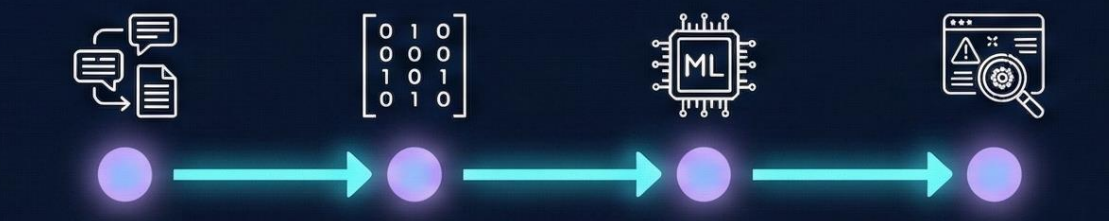
Text Preprocessing

Text normalization
- Lowercasing all text
- Unified text field per channel

Noise handling
- English stopwords removed
- Punctuation handled by vectorization

Channel-specific preparation
- SMS, Email, and Calls modeled separately
- Different vocabulary and structure per channel

Transforming raw communication data into structured features for investigation



Raw Text
SMS • Email • • Call Transcripts

Feature Extraction
TF–IDF Features (1–2 n-grams, max 10k)

Classification Model
Logistic Regression Random Forest

Investigation Output
Risk Scores & IoCs for Analyst Review

Feature Engineering

TF–IDF Vectorization
- Unigrams + Bigrams (n-gram range: 1–2)
- Captures short phrases like 'verify account', 'urgent action', 'call now'

Vocabulary Control
- Maximum of 10,000 features per channel
- Prevents noise and overfitting

Pipeline-based modeling
- TF–IDF → Classifier
- Logistic Regression and Random Forest

Why This Works

Social engineering attacks reuse recurring phrases and psychological triggers, which are effectively captured by TF–IDF n-gram features.

Why These Models?

Model choice impacts both detection performance and investigation outcomes

Logistic Regression (Baseline)

Strong baseline for text classification
- Achieved near-perfect results on Email (Precision ≈ 0.99, Recall ≈ 0.99)

Interpretable model
- Feature weights highlight key phrases
- Supports evidence-based investigation

Stable on high-dimensional sparse TF–IDF data
- Consistent performance across SMS and Email
- SMS Recall ≈ 0.75–0.77 with Precision ≈ 1.00

Random Forest (Non-Linear Model)

Captures non-linear feature interactions
- Useful for complex patterns

Precision–Recall trade-off observed
- SMS: Precision ≈ 1.00, Recall ≈ 0.75
- Calls: Precision ≈ 1.00, Recall ≈ 0.50

Less stable on sparse text features
- Higher variance on small datasets
- Limited interpretability for investigators

Investigation Perspective

For incident investigation, missing a malicious message (false negative) is often more costly than reviewing a false positive.

Results: SMS Channel (Spam vs Ham)

High precision with moderate recall on SMS spam detection

Malicious Class Performance (Spam = 1)

Model	Precision (Spam)	Recall (Spam)
Logistic Regression	1.000	0.765
Random Forest	1.000	0.805

Typical SMS Spam Indicators (IoCs)

free | win | prize | call now | claim reward

Key Observations

- SMS spam is relatively easy to detect using TF-IDF text features
- Both models achieve perfect precision (1.00), meaning almost no false positives
- Recall is lower, especially for shorter or ambiguous spam, highlighting the challenge of missed attacks

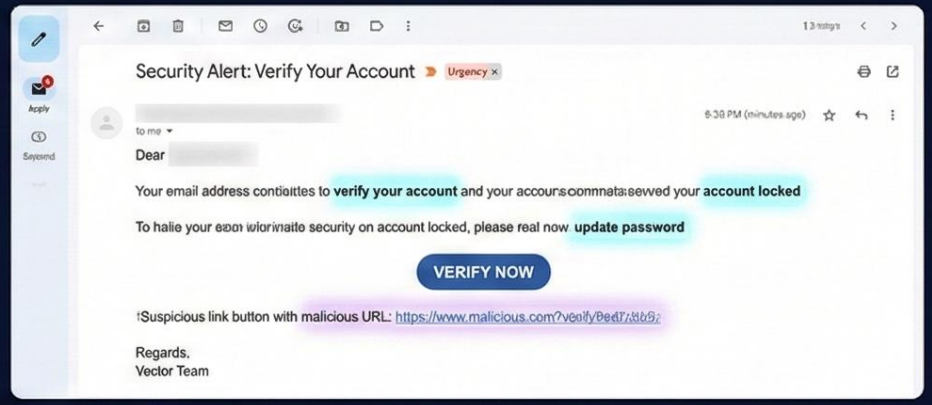
High precision makes SMS models suitable for automated triage, but moderate recall means some spam messages may still evade detection.

Results: Email Channel (Phishing vs Legitimate)

Near-perfect detection performance on phishing emails

Malicious Class Performance (Phishing = 1)

Model	Precision (Phishing)	Recall (Phishing)
Logistic Regression	0.995	0.995
Random Forest	0.997	0.997



Why Email Performs Best

- Phishing emails combine formal language with strong urgency cues
- Repeated phishing templates and security-related keywords make patterns easier to detect
- Rich text content provides more signals than short SMS messages

Although the email dataset is more complex than SMS, the richness of text and repeated templates enable near-perfect detection performance.

Results: Call Channel (Scam vs Normal)

Proof of Concept — Not a Production System

Malicious Class Performance (Scam = 1)

	Model	Precision (Scam)	Recall (Scam)
1	Logistic Regression	0.667	1.000
2	Random Forest	1.000	0.500

Typical Scam Call Patterns

- This is your bank... your account will be locked
- Urgent action required
- Call immediately or make a payment

Key Limitations

- Very small synthetic dataset (call_log.csv)
- Metrics are unstable and sensitive to individual samples
- High risk of overfitting and poor generalization

The call channel demonstrates pipeline extensibility. Real-world voice scam detection would require audio-based features (e.g., MFCCs, prosody), not just short text transcripts.

High-Risk Messages & Indicators of Compromise (IoCs)

Model-extracted examples that provide direct investigation value

SMS — High-Risk Spam Examples

URGENT! You have won a FREE membership. Claim your prize now. Short link: https://shOwUE

Predicted risk ≈ 0.92 – 0.96

Get your FREE ringtone now! Reply with TONE — limited offer

Predicted risk ≈ 0.92 – 0.96

You have WON a guaranteed £1000 cash. Call now to claim your prize

Predicted risk ≈ 0.92 – 0.96

Email — High-Risk Phishing Examples

Security Alert: Verify your account immediately https://www.securityia.com/

Predicted risk ≈ 0.99

Your account has been locked. Update your password to restore access https://www.uourdameaocument.com/

Predicted risk ≈ 0.99

Unusual activity detected. Login now to avoid suspension https://www.clsoin.com/

Predicted risk ≈ 0.99

Phone Calls — High-Risk Scam Examples

This is your bank security department. Your account will be locked in 2 hours.

Predicted risk ≈ 0.66 – 0.68

Unusual activity detected. Press 1 now to avoid suspension.

Predicted risk ≈ 0.66 – 0.68

Provide your card number and PIN to restore access immediately.

Predicted risk ≈ 0.66 – 0.68

Common Indicators of Compromise Across Channels

- Urgency and time pressure
- Authority impersonation (bank, security, support)
- Forced actions (click, call, press, reply)
- Requests for credentials or payment

These high-risk examples can be used as:

- Analyst training material
- IoC patterns for detection rules
- Inputs to SIEM / SOC playbooks

Mapping to the Incident Response Lifecycle

How machine learning outputs support real-world incident investigation

Project Contribution	
Detection	→ ML classification & risk scoring
Triage	→ Prioritize high-risk messages
Analysis	→ IoC extraction from text
Containment	→ Support blocking & response
Lessons Learned	→ Pattern reuse & awareness

The system augments investigator decision-making by providing **early detection, prioritization**, and **investigation-ready evidence** — not automation.

Strengths & Limitations of the Approach

Balancing practical value with academic and operational constraints

Strengths

- Unified machine learning pipeline for SMS, Email, and Phone Call analysis
- Simple and interpretable models (Logistic Regression & Random Forest) with strong real-world performance
- High precision and recall on text-based channels, especially Email phishing (≈ 0.99 precision/recall)
- Practical for SOC use: supports triage, prioritization, and automated filtering

Limitations

- Synthetic and very small call dataset leads to unstable metrics for voice scams
- Text-only analysis: no audio features, email headers, or user behavior context
- Imbalanced datasets across channels, especially for SMS and call data
- Limited deep semantic understanding; models rely on recurring phrases
- Regular retraining required as attackers change wording and tactics

This balance highlights why the system is well-suited for investigation support and SOC triage, but not as a fully autonomous detection solution.

Future Work & Final Conclusion

From proof of concept to investigation-ready systems

Future Work

Explainability & Operations
- Feature importance and SHAP explanations
- Integration with SIEM / SOAR workflows

Multi-Channel Correlation
- Link SMS, Email, and Calls
- Detect coordinated attack campaigns

Advanced Language Models
- Transformer-based models (e.g., BERT)
- Better understanding of subtle phishing language

Richer Call Analysis
- Larger and realistic call datasets
- Audio-based features (voice, prosody, caller behavior)

Machine learning can augment incident investigation — not replace analysts

Our results show that simple, interpretable models can effectively support social engineering detection across SMS, Email, and Calls, while keeping human investigators at the center of decision-making.

This project demonstrates how machine learning can transform raw communication logs into actionable evidence for IT incident investigation.

Thank You

Questions & Discussion

References

Almeida, T. A., Hidalgo, J. M. G., & Yamakami, A. (2011). SMS Spam Collection Data Set. UCI / Kaggle, SMS Spam Collection.
https://www.kaggle.com/code/dejavu23/sms-spam-or-ham-beginner/notebook

Naser Abdullah Alam. CEAS 2008 Phishing Email Dataset. Kaggle.
https://www.kaggle.com/datasets/naserabdullahalam/phishing-email-dataset/data

call_log.csv — Synthetic phone scam dataset created as part of this project.

All datasets are used for educational and research purposes.