# AI-Driven Non-Financial Risk Prediction and Control Effectiveness System for Banking

Capstone Project – Post Graduate Diploma in Artificial Intelligence & Machine Learning

Submitted by: GAT

School: Asian Institute of Management (AIM) – Emeritus

# Business and Data Science Problem

## Business Problem

- Banks face thousands of non-financial risk events yearly: system outages, fraud, compliance breaches, vendor failure.
- This could lead to financial losses, customer impact, regulatory sanctions, reputational damage.
- Traditional tools are backward-looking and reactive.
- There is a need for proactive and predictive risk management to prevent losses and to have more effective resource allocation.

## Data Science Problem

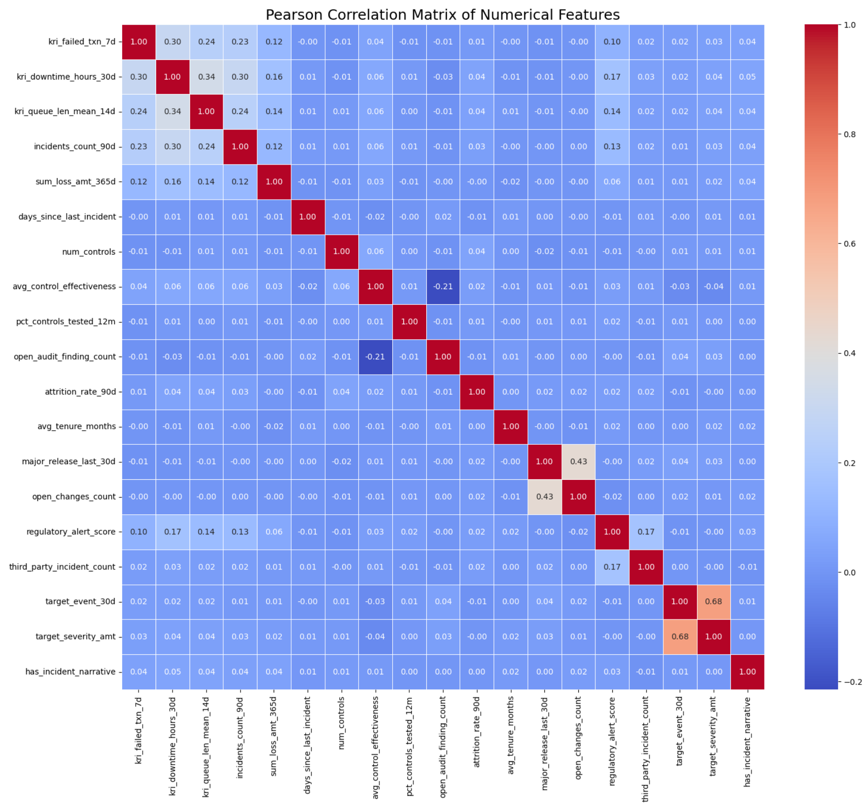
- Predict whether a non-financial risk event will occur within the next 30 days (Event Occurrence, target_event_30d).
- Estimate the potential severity of such events (Severity Amount, target_severity_amt).
- Explains model decisions in a transparent and auditable manner.

# Dataset Preparation and EDA

- Synthetic dataset
- 10,000 records | 120 units | 2-year span
- Structured risk indicators (KRI, RCSA, HR, IT, external signals) + narratives

| Observations | Handling Strategy |
|---|---|
| ‘date’ feature is an ‘object’ | Change to ‘datetime’ |
| ‘incident_narrative’ column has missing values | Create binary indicator (has_incident_narrative)<br>Replace null values with “No narrative’ |
| No duplicates | N/A |
| Mean of the 'target_event_30d' column is approximately 4% | Potential class imbalance but this is the nature of NFR.  Accept and include in interpretations. |
| 'sum_loss_amt_365d' and 'target_severity_amt’ columns are highly skewed | Expected in NFR context. Most days, there are no loss but when it occurs, loss could be significant. Accept and include in interpretations. |
| Outliers are identified for several columns | Outliers represent critical rare events or extreme operational conditions relevant to risk modeling. Keep and include in interpretations. |

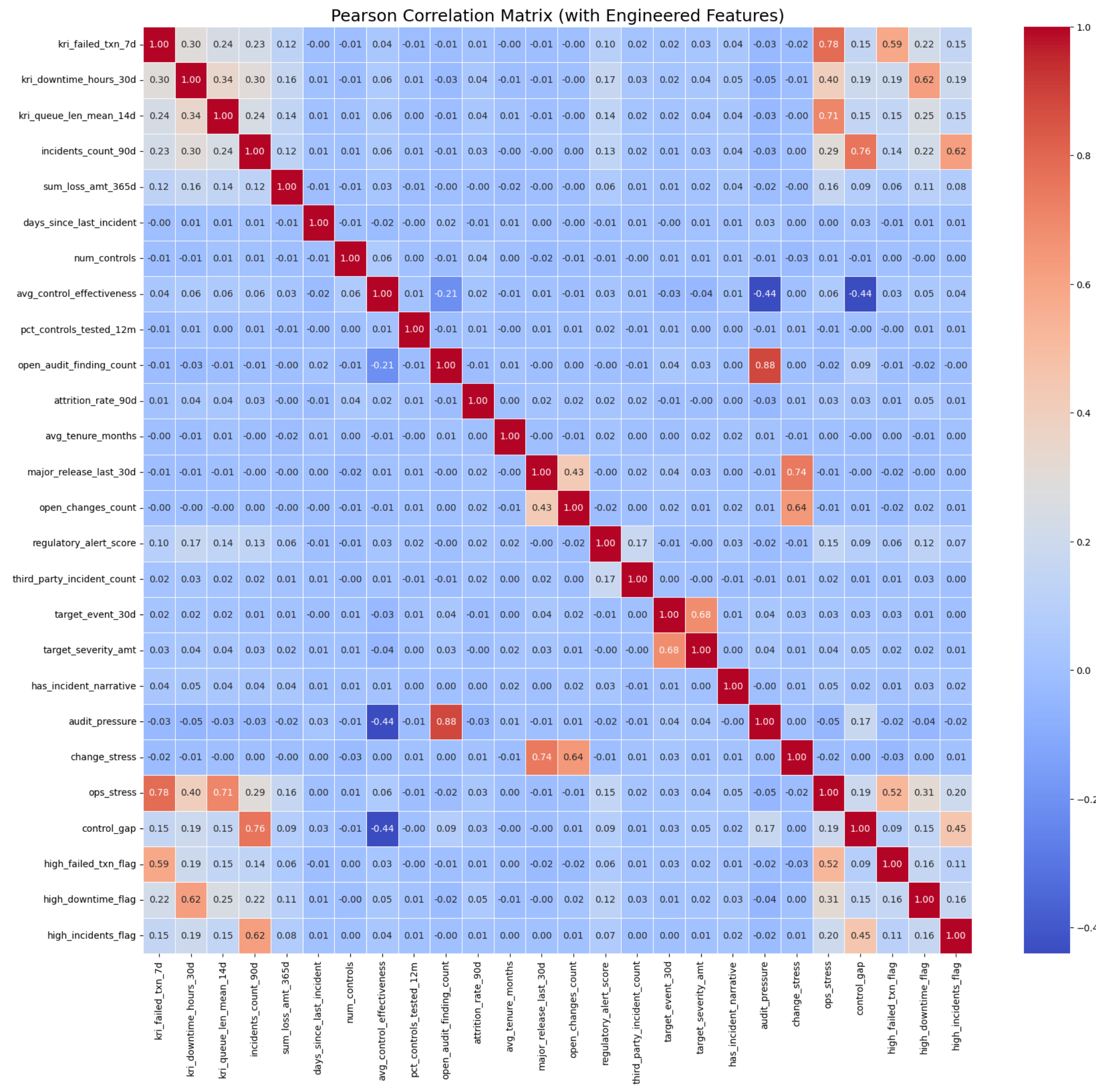
# Feature Engineering




| Observations | Feature Engineering |
|---|---|
| High correlation between the two Target Variables (target_event_30d and target_severity_amt) | Logically the nature of operational events and the dataset. No further feature engineering . |
| Weak correlations of other variables with Target Variables | Feature engineer:<br>• How much unresolved audit pressure exists relative to how strong the controls are?: audit_pressure = open_audit_finding_count / (avg_control_effectiveness + ε)<br>• How stressed the change environment is during a major system release?: change_stress = major_release_last_30d × open_changes_count<br>• How overloaded the operational process is?: ops_stress = kri_failed_txn_7d × kri_queue_len_mean_14d<br>• How often incidents are occurring in an environment with weak controls?: control_gap = (1 − avg_control_effectiveness/5) × incidents_count_90d |
| Heavily skewed distributions | This reflects that most days have no loss, but when losses occur, they can be substantial.<br>Feature engineer:<br>• Is this business unit experiencing an unusually high volume of transaction failures?: high_failed_txn_flag = 1 if kri_failed_txn_7d > 90th percentile<br>• Has the system experienced unusually severe or prolonged downtime recently?: high_downtime_flag = 1 if kri_downtime_hours_30d > 90th percentile<br>• Is this business unit experiencing high incident counts? high_incidents_flag = 1 if incidents_count_90d > 90th percentile |
| Dataset has high dimensionality and complexity | The low variance indicates that operational risk is inherently high-dimensional. As such, dimensionality reduction will not be applied to the final models. |

# After Feature Engineering

- Several of the engineered features show improved or more meaningful correlations with the target variables compared to their individual components.
- This indicates that the feature engineering efforts have been successful in creating stronger signals for risk prediction.



# Modeling (Event Occurrence – Classification)

Objective: Predict whether a non-financial risk event will occur within the next 30 days

| index | ROC-AUC | PR-AUC | Recall | Recall@Top-10% |
|---|---|---|---|---|
| LightGBM | 0.49443526 | 0.04035478 | 0.01265823 | 0.13924051 |
| Logistic Regression | 0.52948425 | 0.04168236 | 0.36708861 | 0.08860759 |
| Random Forest | 0.47994518 | 0.03739842 | 0 | 0.07594937 |
| Decision Tree | 0.49582232 | 0.0392383 | 0.03797468 | 0.08860759 |
| SVC | 0.46668072 | 0.03611753 | 0.15189873 | 0.06329114 |

ROC-AUC provide limited insight. Note that dataset is imbalanced, with approximately 4% positive events. Accepted in NFR context.

PR-AUC values are low, which is expected given the rarity of events and the weak marginal signal of individual risk indicators.

Recall@Top-10% is the metric that is more aligned with the ability to prioritize risk within constrained investigative capacity.

**Decision:** LightGBM is the best model to use.

**Next Step:** Across all evaluated models, ROC-AUC values are close to 0.5 , use tuning threshold to check if ROC-AUC will improve.

# Modeling (Event Occurrence – Classification, After Tuning)

- Threshold tuning significantly improved the ability of models to detect rare non-financial risk events.
- Logistic Regression achieved perfect recall after threshold tuning but this came at the expense of meaningful risk prioritization.
- **Decision:** LightGBM - provides the best balance between high event detection and effective concentration of true events within the top-risk segment, making it the most suitable model for non-financial risk management.

| index | Optimal Recall Threshold | Optimal Recall | Optimal Recall@Top-10% Threshold | Optimal Recall@Top-10% |
|---|---|---|---|---|
| **LightGBM** | 0.05 | 0.63291139 | 0.05 | 0.13924051 |
| **Logistic Regression** | 0.05 | 1 | 0.05 | 0.08860759 |
| **Random Forest** | 0.05 | 0.27848101 | 0.05 | 0.07594937 |
| **Decision Tree** | 0.05 | 0.03797468 | 0.05 | 0.08860759 |
| **SVC** | 0.05 | 0.01265823 | 0.05 | 0.06329114 |

# Modeling (Severity Amount - Regression)

Objective: Estimate the potential severity of events that will occur within the next 30 days

| index | RMSE | MAE |
|---|---|---|
| **LightGBM** | 12907.0186 | 7765.67504 |
| **Linear Regression** | 12310.5532 | 8256.58213 |
| **Random Forest** | 12797.7757 | 8202.28299 |
| **Decision Tree** | 15250.3297 | 10554.4323 |

Loss severity data in NFR is typically highly skewed and heavy-tailed.

Mean Absolute Error (MAE) - reflects typical prediction error.

Root Mean Squared Error (RMSE) - penalizes large deviations more heavily.

In NFR, where understanding and explaining the drivers of loss magnitude is critical, MAE is the better metric to consider.

**Model Choice:** LightGBM - achieves the lowest MAE.

# Addressing Limitations and Ethical AI

## Addressing Limitations

- Class Imbalance
  - NFR events are rare (~4%), which can cause under detection.
  - Addressed through: stratified sampling, class weighting, and probability threshold tuning.
  - Ensure material risk events are prioritized within limited review capacity.
- Target Leakage
  - Preserve predictive integrity.
  - Addressed through: severity amount feature was excluded from event prediction, event occurrence was excluded from severity amount modeling.
  - Ensure all predictions are forward-looking and decision-relevant.
- Overfitting
  - Prevent the models from learning noise or historical artefacts.
  - Addressed through: independent train–test evaluation, model regularization, and the use of ensemble tree-based methods. Conservative learning behavior was applied.
  - Ensure stable performance on unseen data.

## Ethical AI

- Current Scope
  - Synthetic data contains no direct sensitive attributes, so demographic bias testing is not applicable.
- Potential Bias Risks
  - Business units, HR indicators, historical patterns, and incident narratives may act as indirect proxies in real-world data.
- Fairness Governance
  - Monitor prediction parity, error rates, and threshold impacts across identified proxy groups.
- Mitigation Controls
  - Apply debiasing techniques, fairness-aware validation, careful feature design, and ongoing monitoring.
- Risk Committee Assurance
  - The model framework is designed to support fair, explainable, and regulator-ready deployment.

# Business Impact and Results

**Priority Business Units**

- UNIT_116
- UNIT_112
- UNIT_045
- UNIT_098
- UNIT_070
- UNIT_051
- UNIT_090
- UNIT_017
- UNIT_032
- UNIT_103

- The AI-Driven NFR Prediction Model forecasted that NFR events will occur in the next 30 days.
- These are the top 10 business units that should be prioritized because of highest predicted probability.
- Focus on proactive measures, mitigation strategies, and effective control design and implementation.

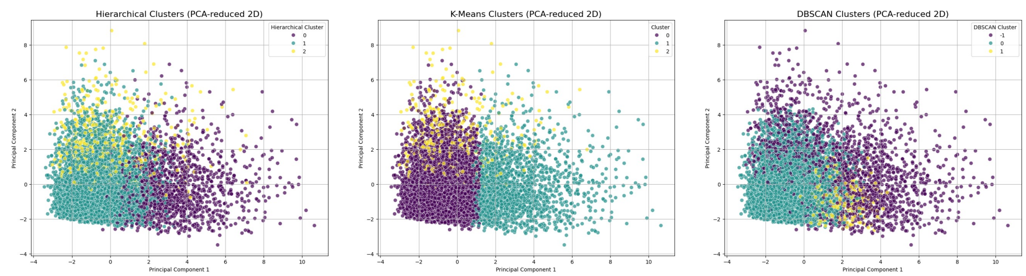
**Potential Loss Avoided**

**$7.6 M**

This figure represents the projected financial loss should risk events materialize within the top 10 prioritized business units.

# Experimentation: Unsupervised Learning

- If we must choose: Hierarchical Clustering
  - Higher Silhouette Score and balanced cluster structure indicate better internal cohesion and separation, which is desirable for exploratory segmentation of risk profiles.
- However:
  - Silhouette Scores across all methods highlight that NFR data does not naturally form strongly separable clusters.
  - This reflects the complex, overlapping, and multi-factor nature of operational and non-financial risks.
- Conclusion:
  - Use unsupervised clustering as an exploratory and complementary analysis, not a core predictive mechanism.




| index | Sil Score |
|---|---|
| K-Means | 0.214 |
| DBSCAN | 0.182 |
| Hierarchical Clustering | 0.239 |

# Experimentation: Recommender System and Deep Learning

## Recommender System

- Suitability for NFR - content-based recommendation approach (instead collaboration)
- Data Preparation and Risk Profiling
  - 24 relevant numerical risk indicators  (excludes identifiers and target variable to prevent leakage)
  - 120 distinct business unit profiles
  - Scaled to ensure comparability
- Similarity computation – Cosine similarity used for comparing high-dimensional feature vectors
- Implementation and Validation - Testing the system using UNIT_001 successfully returned a ranked list of five peer units with closely aligned risk profile.
- For further study: Incorporate textual data, dynamic profiles, hybrid similarity metrics

## Deep Learning

- Deep learning models were not performed due to limited data volume, tabular feature structure, and the need for transparent, auditable explanations in a non-financial risk context.
- Area for Further Study for NFR: Management: Deep learning techniques are most appropriate for non-financial risk applications involving unstructured or sequential data, such as incident narratives, regulatory communications, and time-series escalation patterns.