# Key User & Business Requirement Identification

**Document Title:** Key User & Business Requirement Identification  
**Activity ID:** M1.2  
Date: [05/04/2025]

**Project Team ID:** TP706 – S1 – Team2  
**Project Team Name:** Team 2

## 1. Description of the Activity

This activity focuses on identifying the key users and gathering detailed business requirements relevant to data analysis. It sets the foundation for building queries, insights, and dashboards that are aligned with user needs and strategic objectives.

## 2. Target Users / Stakeholders

| **User Group/Role** | **Role Description** |
| --- | --- |
| Fraud Analyst | Review alerts, confirm fraud, reduce false positives. |
| Compliance Officer | To meet regulatory requirements (e.g. AML, KYC), and ensure legal compliance. |
| Risk Manager | To proactively reduce financial loss and operational risk from fraudulent activity. |
| Cybersecurity Analyst | To detect digital fraud patterns like phishing, ATO, credential stuffing. |
| Product Manager | To integrate fraud protection into the user experience with minimal friction. |
| Data Scientist / ML Engineer | To build or improve machine learning models that detect evolving fraud patterns. |
| Customer Support Rep  User | Communicates declined transactions or account restrictions to users effectively.  Relies on the platform’s fraud detection to ensure reviews, transactions, and interactions are trustworthy, secure, and authentic. |

## 3. Remarks / Notes

**Assumptions**

1. Labeled Training Data Is Available  
   It is assumed that sufficient labeled data (genuine vs. fake reviews) will be available from public sources (e.g., Amazon, Yelp, Kaggle) or can be labeled manually.
2. Structured Review and Metadata Access  
   Reviews and relevant user metadata (account age, review frequency, timestamps, product categories) will be provided in a structured format such as CSV, JSON, or databases.
3. Historical Data Is Accessible  
   Access to past reviews logs will enable the analysis of long-term user behavior trends.
4. Language Scope:  
   The initial version of the model assumes reviews are in English; multilingual support is out of scope for MVP.
5. Distinctive Behavioral Patterns  
   Fraudulent reviews are expected to exhibit identifiable patterns - such as review bursts and overly positive language - that can be captured through NLP and behavioral analytics.
6. Seamless AI Integration  
   The review platform (or simulated system) allows integration of AI services via APIs or external modules.
7. Continuous Improvement and Validation:  
   Early iterations will incorporate human review and stakeholder feedback to refine detection accuracy and reduce false positives/negatives.
8. User Experience Preservation  
   Flagging or suppressing suspicious reviews is anticipated to maintain a positive user experience without violating platform policies.
9. Adequate Computing Infrastructure  
   Sufficient local or cloud-based computing resources (e.g., GPUs/CPUs) will be available for model training, testing, and deployment.
10. Cross-Dataset Applicability  
    Models developed using one dataset (e.g., Amazon) are expected to generalize to similar datasets with only minor adjustments.

## **Limitations**

1. Limited or Sparse Labeled Data  
   Reliable labels (fake vs. genuine) are often unavailable or hard to verify, which limits supervised learning performance.
2. Class Imbalance  
   Genuine reviews far outnumber fake ones, making it challenging for models to learn the minority class effectively.
3. Short, Vague, or Noisy Reviews  
   Reviews like “Nice product” offer little context for NLP analysis and may lead to false negatives.
4. Evolving Fraud Techniques  
   Fraudsters constantly change writing styles to evade detection, which can update the model quickly.
5. Incomplete or Unavailable Metadata  
   Critical behavioral data like IP address, geolocation, or device ID may not be accessible due to privacy restrictions or API limitations.
6. Cross-Domain Generalization Issues  
   A model trained on one domain (e.g., electronics) may not perform well in others (e.g., restaurants) due to differences in language and review structure.
7. Hard-to-Detect Fraud from Real Accounts  
   Some fake reviews are posted by real or compromised user accounts, making them difficult to distinguish from genuine ones.
8. Uncertain Ground Truth  
   Verifying whether a review is genuinely fake can be subjective or costly, especially without platform-side confirmation or user reports.