**AI-Based Fraud Detection System for E-Commerce**

Zainab Saeed

08/10/2024

Abstract

**1.0 Introduction**

This project presents an innovative AI-Based Fraud Detection System tailored for e-commerce platforms. With the rising incidence of fraudulent transactions, especially in online retail, there is an urgent need for businesses to implement effective solutions that safeguard their operations and customer trust.

This project not only showcases the practical application of machine learning in the realm of e-commerce but also emphasizes the importance of proactive measures to mitigate fraud risk. The AI-Based Fraud Detection System stands as a valuable tool for small and medium-sized enterprises, enabling them to harness the power of data science to protect their financial interests and maintain customer loyalty.

# **1.1 Problem Statement**

E-commerce businesses, particularly small and medium-sized enterprises (SMEs), face a growing threat from fraudulent transactions, which can lead to significant financial losses, disruption of operations, and erosion of customer trust. As online retail platforms expand, so do the opportunities for cybercriminals to exploit vulnerabilities in payment systems, leading to issues like identity theft, payment fraud, and account takeovers. The losses from such fraudulent activities disproportionately affect SMEs, which often lack the resources to invest in advanced fraud detection systems.

Traditional fraud detection methods, typically based on predefined rules and static thresholds (e.g., flagging transactions above a certain amount), are proving insufficient. These rule-based systems cannot keep pace with the sophisticated and rapidly changing tactics used by fraudsters. Moreover, these systems often result in false positives, where legitimate transactions are flagged as suspicious, causing further disruptions to business operations and frustrating customers.

# **2.0 Market and Customer Needs Assessment**

**2.1 Market Analysis**

The e-commerce sector has experienced unprecedented growth over the past decade, driven by the rapid shift toward online shopping and digital transactions. Along with this growth, fraudulent activities such as identity theft, chargebacks, and account takeovers have also risen sharply. In 2023 alone, global losses from e-commerce fraud amounted to billions of dollars, with estimates suggesting that this trend will continue to rise.

While large corporations have dedicated fraud prevention teams and can afford sophisticated tools, small and medium-sized enterprises (SMEs) are disproportionately affected by these fraud schemes. They often lack the financial and technical resources to deploy advanced fraud detection systems and must rely on outdated or inadequate rule-based models. These models are rigid and unable to adapt to the rapidly changing fraud landscape, leaving SMEs vulnerable to financial losses and reputational damage.

As a result, the demand for scalable, affordable, and AI-driven fraud detection solutions is increasing. There is a significant market gap for fraud prevention systems that are specifically designed for SMEs, offering them an opportunity to protect themselves from financial risk without incurring large costs or requiring extensive technical expertise.

**2.2 Customer Segmentation**

The target customers are primarily small and medium-sized enterprises (SMEs) operating in the e-commerce space. These businesses fall into several distinct categories:

1. **Small Online Retailers**: This group includes independent online stores, boutique retailers, and small marketplace sellers. They are often run by solo entrepreneurs or small teams with limited budgets for IT infrastructure and fraud detection. Despite their size, they face a significant risk of fraudulent transactions, especially with high-value goods.
2. **Local Food Chains and Grocery Delivery Services**: Small food businesses transitioning to online platforms for ordering and delivery face payment fraud risks. These businesses often lack the technical know-how to manage complex fraud detection systems but require real-time protection to maintain customer trust and ensure financial stability.
3. **Service-Based SMEs (e.g., travel, hospitality, and education)**: These businesses frequently handle payments for services rendered online, which makes them vulnerable to chargebacks and fraudulent bookings. Fraud in these sectors can lead to operational disruptions, leading to customer dissatisfaction and revenue loss.
4. **Digital Marketplaces and Subscription-Based Businesses**: SMEs running platforms that facilitate transactions between buyers and sellers (e.g., small-scale digital marketplaces) are at high risk for fraud because they process a large number of transactions daily.

These customers need a fraud detection solution that:

* Scales as their businesses grow,
* Is affordable, with minimal upfront costs,
* Requires little to no technical expertise to implement, and
* Detects fraud in real-time without disrupting legitimate transactions.

# **3.0 Target Specifications and Characterization**

The target customers are primarily small and medium-sized enterprises (SMEs) engaged in e-commerce. Understanding their characteristics and needs is crucial for tailoring the solution effectively. The following specifications and characteristics define this target market:

#### **3.1 Business Size and Structure**

* **Small Enterprises**: Typically have fewer than 50 employees and limited operational budgets. They may operate as sole proprietorships or small partnerships, often relying on a single owner or a small team to manage multiple aspects of the business.
* **Medium Enterprises**: Employ between 50 and 250 employees and may have more structured departments, but often lack dedicated IT security teams. They face similar challenges as small businesses but may have slightly larger budgets for technology investments.

#### **3.2 Industry Segmentation**

* **E-Commerce Retailers**: Businesses selling physical or digital goods online. They often deal with high transaction volumes and are susceptible to various forms of fraud.
* **Service Providers**: SMEs in sectors like travel, hospitality, and education that accept online payments for services. They require solutions that minimize the risk of chargebacks and fraudulent bookings.
* **Food Delivery and Local Restaurants**: With increasing digital orders, they need protection against payment fraud and account takeovers.

**4.0 External Search**

The development of the product necessitates the integration of advanced machine learning techniques that can effectively identify fraudulent activities in real-time. To streamline this process, we can utilise existing machine learning frameworks and tools that offer robust capabilities for fraud detection.

Leveraging pre-trained models from libraries such as TensorFlow, PyTorch, and Scikit-learn can significantly reduce development time. These libraries contain a wealth of algorithms specifically designed for classification, regression, and anomaly detection, which can be adapted to recognize patterns indicative of fraudulent transactions.

Specialized libraries such as CreditCardFraudDetection or PyOD (Python Outlier Detection) offer pre-built algorithms that can be fine-tuned for our specific use case. These libraries are designed to detect anomalies and unusual patterns in transaction data, providing a solid foundation for our system.

The implementation of the fraud detection system relies heavily on the quality of data. Libraries such as Pandas and NumPy will be essential for data preprocessing, enabling us to clean, transform, and analyse transaction data effectively before it is fed into the machine learning models.

**4.1 Benchmarking**

In the market today, there are several fraud detection systems that aim to provide protection against fraudulent activities in e-commerce. These systems range from rule-based models to advanced AI-powered solutions. However, many existing products are either too costly for small and medium-sized enterprises (SMEs) or require complex technical knowledge to operate effectively.

The benchmarking analysis below compares different fraud detection systems based on features such as cost, performance, scalability, ease of integration, and the use of machine learning algorithms. The table highlights how these systems align with the target specifications for an affordable, real-time fraud detection tool tailored to SMEs.

**Benchmarking Table of Products**

| Feature | FraudLabs Pro | Kount | Sift Science | Signifyd |
| --- | --- | --- | --- | --- |
| Size | Cloud-based | Cloud-based | Cloud-based | Cloud-based |
| Weight (Data Handling Capacity) | High | High | Very High | Medium |
| Cost | Low (SME-friendly) | Medium | High (Enterprise Level) | Medium |
| Flexibility | Moderate (Limited customization) | High (Highly customizable) | High | Moderate (Some customization available) |
| Scalability | Moderate | High | Very High | High |
| Ease of Use | High (User-friendly) | Moderate | Moderate | High |
| Machine Learning Integration | No (Rule-based) | Yes | Yes | Yes |
| Support for SMEs | Designed for SMEs | Primarily for large businesses | Primarily for large businesses | SME and large businesses |
| Algorithms Used | Rule-based | Supervised ML | Deep Learning | Supervised ML |

From the benchmarking analysis, it is evident that while systems like FraudLabs Pro offer affordability and ease of use for SMEs, they rely heavily on traditional rule-based systems, which may not be as effective against evolving fraud tactics. On the other hand, more advanced systems like Sift Science and Kount leverage machine learning for fraud detection but are often designed for large enterprises, making them expensive and complex for SMEs to implement.

Signifyd provides a middle ground, offering scalability and machine learning integration, but at a cost that may still be too high for some SMEs. This benchmarking highlights the need for a fraud detection system that offers real-time AI-powered detection, scalability, and affordability, specifically tailored to meet the needs of SMEs.

**4.2 Applicable Patents**

**Fraud Detection System Using Machine Learning Algorithms:-** This patent describes an AI-powered fraud detection system that uses a combination of supervised and unsupervised learning algorithms to analyze transaction data and flag potential fraudulent activities. The system employs anomaly detection techniques and can learn from new data over time to improve its fraud detection accuracy.

Link : <https://patents.google.com/patent/US11308497B2/en>

**4.3 Applicable Standards**

**4.3.1 Data Protection and privacy standards**

#### **General Data Protection Regulation (GDPR) - Europe:** GDPR is a comprehensive regulation aimed at protecting the personal data of European citizens. It mandates strict data privacy laws, requiring businesses to manage customer information securely and transparently.

#### **Impact**: As the system will handle personal and transactional data, GDPR compliance is mandatory if the solution operates in or collects data from European users. The system must ensure that:

#### Users provide explicit consent for data collection.

#### Personal and financial data are encrypted and anonymized.

#### Users have the right to request data deletion or modification.

#### **California Consumer Privacy Act (CCPA) - United States:** CCPA is similar to GDPR but applies to California residents. It ensures consumer rights to access their data, opt-out of data sales, and request data deletion.

#### **Impact**: The fraud detection system must provide features allowing customers to view and manage their data. Proper mechanisms must be in place to allow California residents to exercise their rights, such as opting out of data tracking.

#### **Payment Card Industry Data Security Standard (PCI DSS):** PCI DSS is a global standard for businesses that handle credit card transactions. It aims to secure financial data and protect against credit card fraud by enforcing strict security controls.

#### **Impact**: As the system deals with transaction data, PCI DSS compliance is required. The system must ensure encrypted storage, secure payment gateways, and regular security audits to protect cardholder data. Non-compliance can lead to fines and reputational damage.

#### **4.3.2 Cybersecurity Standards:**

#### **National Institute of Standards and Technology (NIST) Cybersecurity Framework - United States:** NIST’s framework provides a structured approach to identifying, managing, and mitigating cybersecurity risks.

**Impact**: The system needs to implement robust cybersecurity measures such as multi-factor authentication (MFA), data encryption, and network security protocols to safeguard e-commerce transactions from fraudsters and hackers.

#### **ISO/IEC 27001 - Information Security Management Systems:** ISO 27001 is an international standard for managing information security, ensuring that an organization keeps sensitive data secure.

**Impact**: The system should follow ISO 27001 guidelines to develop secure data-handling processes, control access to sensitive information, and implement regular security checks to prevent data breaches.

**4.3.3 Artificial Intelligence and Ethical Standards:**

#### **ISO/IEC 23894:2023 - Risk Management in Artificial Intelligence:** This standard outlines the risk management framework specifically designed for AI applications. It covers ethical concerns, risk of bias, transparency, and fairness in decision-making algorithms.

AI algorithms used for fraud detection must be designed with fairness and explainability. The system should offer transparency in decisions to avoid false positives and ensure that flagged transactions can be reviewed by human operators to avoid unfairly penalizing legitimate users.

#### **Ethical Guidelines for Trustworthy AI (EU Commission):** These guidelines ensure that AI systems are developed with respect for fundamental rights, fairness, and non-discrimination.

**Impact**: The fraud detection system must align with these ethical guidelines, particularly regarding the explainability of fraud detection outcomes, ensuring that customers are not unfairly flagged due to biases in the AI model.

**4.3.4 Health, Safety, and Environmental Regulations:**

#### **ISO 14001 -** Environmental Management Systems: ISO 14001 provides standards for managing and reducing environmental impacts in business operations.

**Impact:** If the AI fraud detection system requires physical infrastructure (servers, data centers), the project should minimize its environmental impact by adopting energy-efficient solutions such as cloud computing with low carbon footprints or green data centers.

#### **Occupational Health and Safety (OHSAS 18001):** This standard ensures workplace safety in organizations, especially those handling large-scale data and computing systems.

**Impact:** For businesses managing server infrastructure, it’s essential to follow health and safety regulations to protect staff from hazards related to hardware maintenance or server room operations.

### **4.3.5. Government Policies and Industry Regulations:**

#### **Anti-Money Laundering (AML) Regulations - Global:** E-commerce platforms are often vulnerable to fraudulent activities linked to money laundering schemes. AML regulations require businesses to have systems in place to detect and report suspicious financial activity.

**Impact**: The fraud detection system must comply with AML regulations by incorporating algorithms that can flag transactions potentially linked to money laundering, ensuring compliance with governmental reporting requirements.

#### **Know Your Customer (KYC) Compliance - Financial Sector:** KYC guidelines are used in financial industries to verify the identity of clients to prevent fraud, money laundering, and financing of illegal activities.

**Impact**: The fraud detection system must integrate KYC processes, such as verifying customer identities through official documents or biometric data, ensuring that fraudulent accounts or activities are identified early.

**4.4 Applicable Constraints**

#### **Budget Constraints:** Small and medium-sized enterprises (SMEs) often operate with limited budgets, meaning the fraud detection system must be cost-effective.

**Impact:** The system’s development must prioritize cloud-based, scalable solutions to reduce infrastructure costs. The software should offer modular pricing, allowing SMEs to only pay for features they need.

#### **Technical Expertise and Implementation:** Many SMEs do not have extensive technical expertise to manage complex AI systems.

#### **Impact:** The fraud detection system should be designed to be user-friendly, with minimal technical overhead. Pre-built models and an easy-to-use interface will make it accessible to businesses without a dedicated IT team.

#### **Market and Environmental Constraints:** Variations in market demand, as well as regional privacy laws, may limit the system’s availability.

**Impact:** The system must be adaptable to different markets, incorporating compliance mechanisms for local regulations, such as GDPR in Europe or CCPA in California. Furthermore, the system’s environmental impact should be minimized through sustainable cloud computing practices

**4.5 Business Model**

**1.Subscription-Based Model:**

Basic Plan ($50 - $100/month):  
Targeted at small businesses, this plan provides real-time fraud detection with predefined rule-based algorithms. It includes basic analytics, fraud alerts, and periodic reports on fraud patterns.

Pro Plan ($200 - $500/month):  
Designed for medium-sized enterprises, this plan offers advanced machine learning models, custom fraud detection features, and in-depth analytics on fraudulent activities. Customers will also receive real-time monitoring and reporting with predictive analytics.

Enterprise Plan (starting at $1000/month):  
Tailored for larger businesses or organizations with high transaction volumes, this plan includes fully customized fraud detection models, integration with existing security infrastructure, and dedicated support. Clients in this tier will benefit from bespoke machine learning models and comprehensive AI-driven insights into potential fraud risks.

#### **2. Freemium Model:**

Free Tier:  
This model offers a limited version of the fraud detection system with basic functionality, such as fraud detection alerts and limited transaction monitoring. It allows small businesses to experience the system without upfront costs. They can upgrade to higher tiers for more features, including advanced analytics and machine learning models.

Paid Upgrades:  
Users can unlock additional features like custom reporting, access to more sophisticated algorithms, and real-time monitoring by subscribing to higher-tier plans.

#### **3. White-Label Solutions:**

Licensing to Payment Processors/E-commerce Platforms:  
This model involves partnering with payment processors, e-commerce platforms, or other service providers, allowing them to offer the fraud detection system under their brand. By white-labelling the product, we can license it to these companies as part of their service offering. This will increase market reach while generating recurring revenue through licensing fees or shared revenue from the platform’s customer base.

**5.0 Concept Generation**

As someone who loves shopping online, I’ve noticed a big problem: fraud. Many e-commerce sites face issues like fake transactions and stolen accounts, which can hurt small to medium-sized businesses (SMEs) the most. These businesses often don’t have the money or resources to keep up with advanced fraud detection systems. The old systems they use are not enough to handle the clever tactics of today’s fraudsters, which leads to financial losses and unhappy customers.

Now, imagine a smart fraud detection system built just for these smaller businesses. This system would use artificial intelligence (AI) to check transactions in real time, spotting any suspicious activity while keeping false alarms low. It would be affordable and easy to integrate into existing online stores, helping small businesses protect themselves against fraud without breaking the bank.

In addition to detecting fraud, the system would provide helpful insights so that businesses can understand their unique risks. Features could include customizable alerts, detailed reports, and easy-to-use settings, allowing businesses to take charge of their security. This way, they can focus on growing their business and keeping their customers happy.

This system wouldn’t just solve the immediate problem of fraud; it would also help create a safe shopping experience for everyone. As customers become more aware of fraud risks, knowing that a reliable fraud detection system is in place will boost their confidence in shopping online.

Overall, this idea supports small businesses in staying safe while using modern technology to fight fraud. By making effective security affordable, we can help create a better online shopping environment for everyone.

**6.0 Concept Development**

When first-time users launch the AI-Based Fraud Detection System, they are prompted to create an account, ensuring secure access and personalization of the service. Upon logging in, users are greeted with a friendly and intuitive user interface (UI), designed to be easy to navigate, similar to popular e-commerce platforms.

The home dashboard displays a summary of the user's transaction history, highlighting any recent activities, alerts for suspicious transactions, and an overview of the fraud risk score for their account. Users can quickly access different functionalities, such as:

**Transaction Monitoring:** Users can view real-time alerts for any potentially fraudulent transactions. The system flags suspicious activities based on predefined thresholds and machine learning algorithms, allowing users to take immediate action (e.g., confirming or blocking the transaction).

**Fraud Detection Settings:** Users can customize their fraud detection parameters, such as setting alerts for specific transaction types, choosing their preferred level of risk assessment, and managing notification preferences (e.g., via email or SMS).

**Insights and Analytics:** A dedicated section provides users with visual insights into their transaction patterns, highlighting areas where they may be at risk. This feature will include charts and graphs that illustrate spending habits, frequency of transactions, and any anomalies detected by the system.

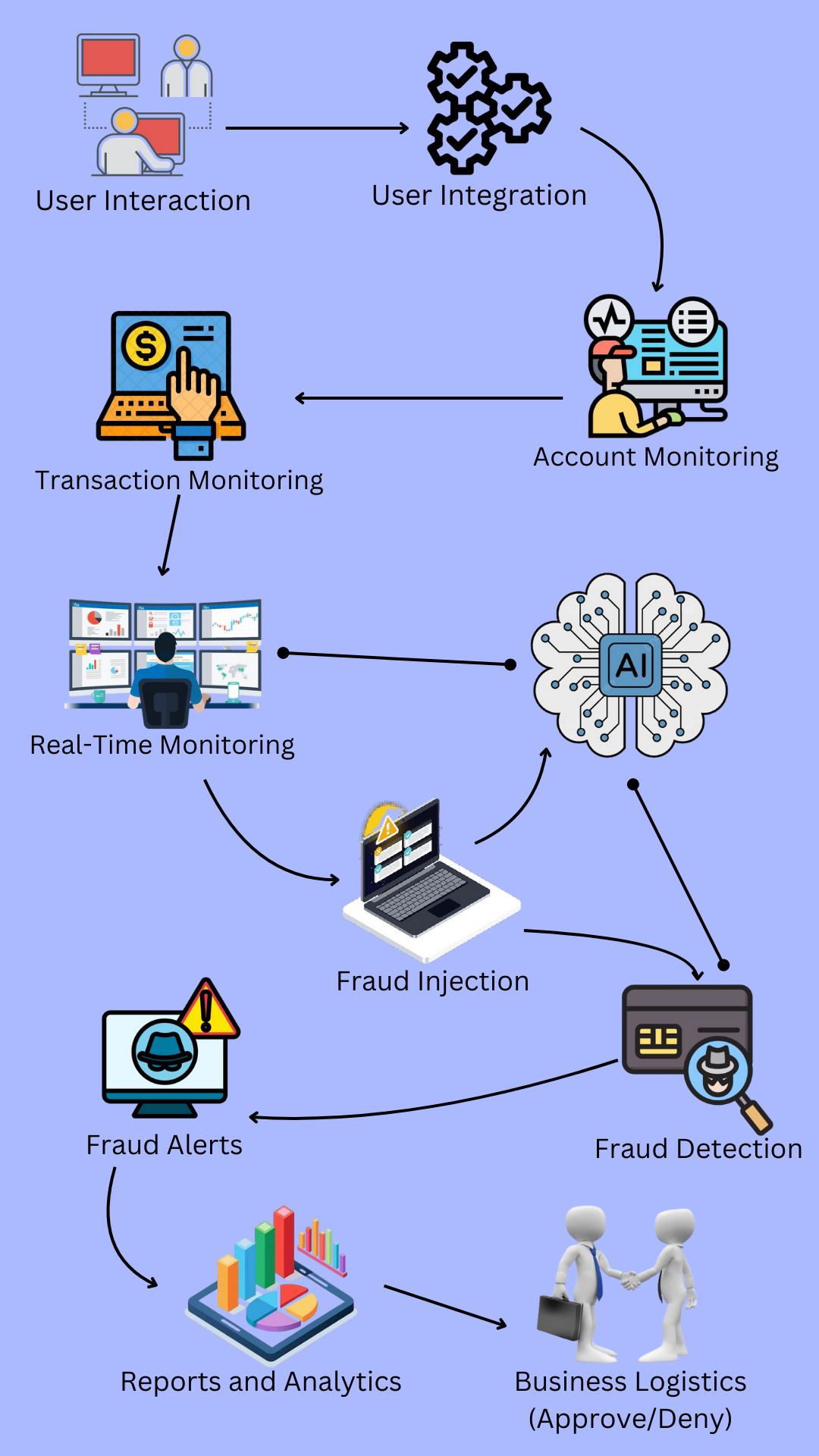
**Support and Resources:** Users can access educational resources on fraud prevention, including tips and best practices, frequently asked questions (FAQs), and direct support options, such as live chat or email support.

For premium users, the system offers advanced features that enhance their experience and provide additional layers of security. These features include:

* **Personalized Fraud Risk Assessment:**
  + Using machine learning algorithms, the system learns from the user’s transaction history to provide personalized risk assessments and tailored alerts based on their unique spending patterns.
* **Customizable Fraud Rules:**
  + Premium users can create custom fraud detection rules, allowing them to tailor the system to their specific needs and preferences, such as excluding certain transaction types from alerts.
* **Regular Reports:**
  + Monthly reports summarising transaction activity and fraud alerts are generated, providing users with insights into their financial behaviour and any potential vulnerabilities.
* **Community Engagement:**
  + Options to join online forums or user groups where they can discuss fraud prevention strategies, share experiences, and receive tips from other users.

**7.0 Final Design**

The final design of the AI-Based Fraud Detection System for E-Commerce integrates an intuitive user interface, a robust AI engine, and a secure database management system to effectively detect and prevent fraudulent transactions in real time, specifically targeting small and medium-sized enterprises (SMEs). The user interface features a streamlined layout for easy navigation, providing real-time notifications and graphical insights into transaction patterns. The AI engine employs machine learning algorithms, such as Random Forest and Gradient Boosting, for anomaly detection and real-time data processing. A secure database management system ensures encrypted storage of sensitive transaction data while complying with data privacy regulations. To enhance the design, a Failure Modes and Effects Analysis (FMEA) was conducted to identify potential failure modes, their impacts, and mitigation strategies, resulting in an emphasis on user trust and financial security. The design decisions prioritize user experience, professional standards, and safety, while a cost-benefit analysis demonstrates the economic viability of the project by highlighting significant cost savings through reduced fraudulent transactions. Mock-ups and simulations were used to validate the design, ensuring the system meets user needs and performs effectively in real-world scenarios.



### **7.1 How Does It Work?**

The functioning of the product for E-Commerce is divided into three key sections:

1. **User Experience**

* Account Setup: Business owners create an account and log into the platform to access fraud detection features.
* Transaction Monitoring: Users can monitor transactions in real time, view flagged transactions, and generate fraud analysis reports from the system dashboard.
* Fraud Alerts: Alerts are triggered when transactions are identified as potentially fraudulent based on pre-set criteria.
* Fraud Investigation Tools: Users can deep-dive into flagged transactions, view historical patterns, and make decisions based on the system’s recommendations.
* Customization: Businesses can set fraud detection parameters, thresholds, and adjust model settings to align with their industry and transaction profiles.

1. **Fraud Detection and Business Interaction**

* **Data Integration:** The system pulls data from e-commerce platforms, payment gateways, and customer databases to analyze transactional behavior.
* Anomaly Detection: Machine learning algorithms scan through historical and real-time data to detect deviations from normal transaction patterns, flagging them for review.
* Business Insights: The system generates insights about transaction risks, giving businesses a comprehensive view of their fraud landscape.
* Report Generation: Users can generate custom reports based on fraud detection performance, highlighting risky transactions and fraud trends.

1. **Backend Interaction**

* Transaction Data Analysis: The backend continuously processes new and historical data using machine learning algorithms.
* Model Refinement: The fraud detection models improve over time by learning from flagged and reviewed transactions, leading to more accurate detection.
* Third-Party Integration: The system can be integrated with external fraud detection tools or payment gateways to enhance its capabilities.

### **7.2 Data Sources**

* Transactional Data: Integration with e-commerce platforms and payment gateways for real-time transaction monitoring.
* Customer Data: Access to customer profiles, transaction history, and purchase behavior to identify fraudulent activity.
* External Fraud Databases: Access to industry-standard fraud databases for identifying known fraudulent patterns.
* Machine Learning Models: Data from past flagged transactions is fed into machine learning models to improve detection accuracy over time.

### **7.3 Algorithms, Frameworks, Software**

1. **Algorithms:**

* **Anomaly Detection:** Machine learning algorithms, including decision trees and neural networks, identify abnormal transaction patterns.
* **Clustering Algorithms:** K-Means and DBScan are used to group transactions and customers based on behavior patterns.
* **Natural Language Processing (NLP):** Chatbot for customer support interactions.

1. **Frameworks and Software:**

* **Backend Development:** Django or Flask for managing server logic.
* **Data Processing:** Pandas and NumPy for data analysis, along with TensorFlow for machine learning tasks.
* **Cloud Services:** AWS or Google Cloud for data storage and processing.
* **Database Management:** MySQL for structured transaction data storage and MongoDB for unstructured data.
* **APIs:** REST APIs for integration with payment gateways and e-commerce platforms.

### 

### **7.4 Teams Required**

1. **Product Development:**

* **Data Scientists:** For designing and implementing machine learning models and fraud detection algorithms.
* **Software Engineers:** For developing the platform's backend, APIs, and frontend dashboards.
* **UX/UI Designers:** For designing an intuitive and user-friendly dashboard.

1. **Operations:**

* **Customer Support:** For assisting users with the system and addressing alerts.
* **Data Analysts:** For evaluating fraud detection performance and improving model accuracy.

1. **Business Development:**

* **Partnership Managers:** To establish relationships with payment gateways and e-commerce platforms.
* **Sales and Marketing:** To attract new business customers and promote the product.

1. **Administration:**

* **Legal Advisors:** To ensure compliance with data privacy laws and contracts.
* **HR and Office Management:** To oversee internal operations and staff hiring.

### **7.5 Costs**

1. **Initial Development Costs:**

* **App Development:** Costs associated with developing the platform, including frontend and backend infrastructure.
* **Data Processing:** Expenses for setting up cloud services and databases.
* **Machine Learning Models:** Initial development of fraud detection models.

1. **Ongoing Operational Costs:**

* **Team Salaries:** For the development, data science, and support teams.
* **Cloud and Server Costs:** Ongoing expenses for data processing and storage.
* **Marketing and Sales:** Expenses to promote the system and acquire new customers.

1. **Other Costs:**

* **Legal Fees:** Costs associated with data compliance, insurance, and legal services.
* **Partnership Costs:** Costs incurred through collaborations with e-commerce platforms and payment gateways.

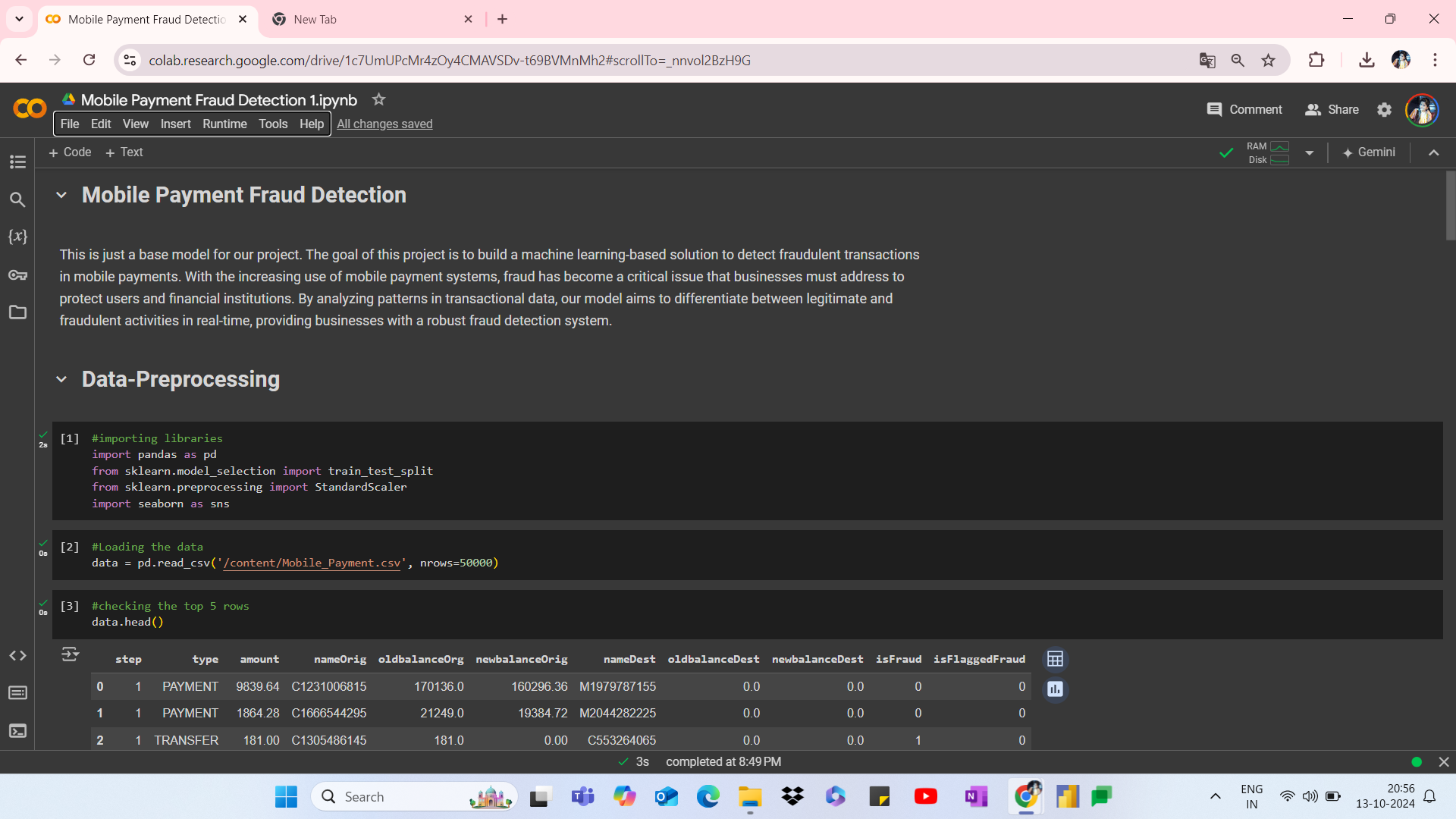
**8.0 Conclusion**

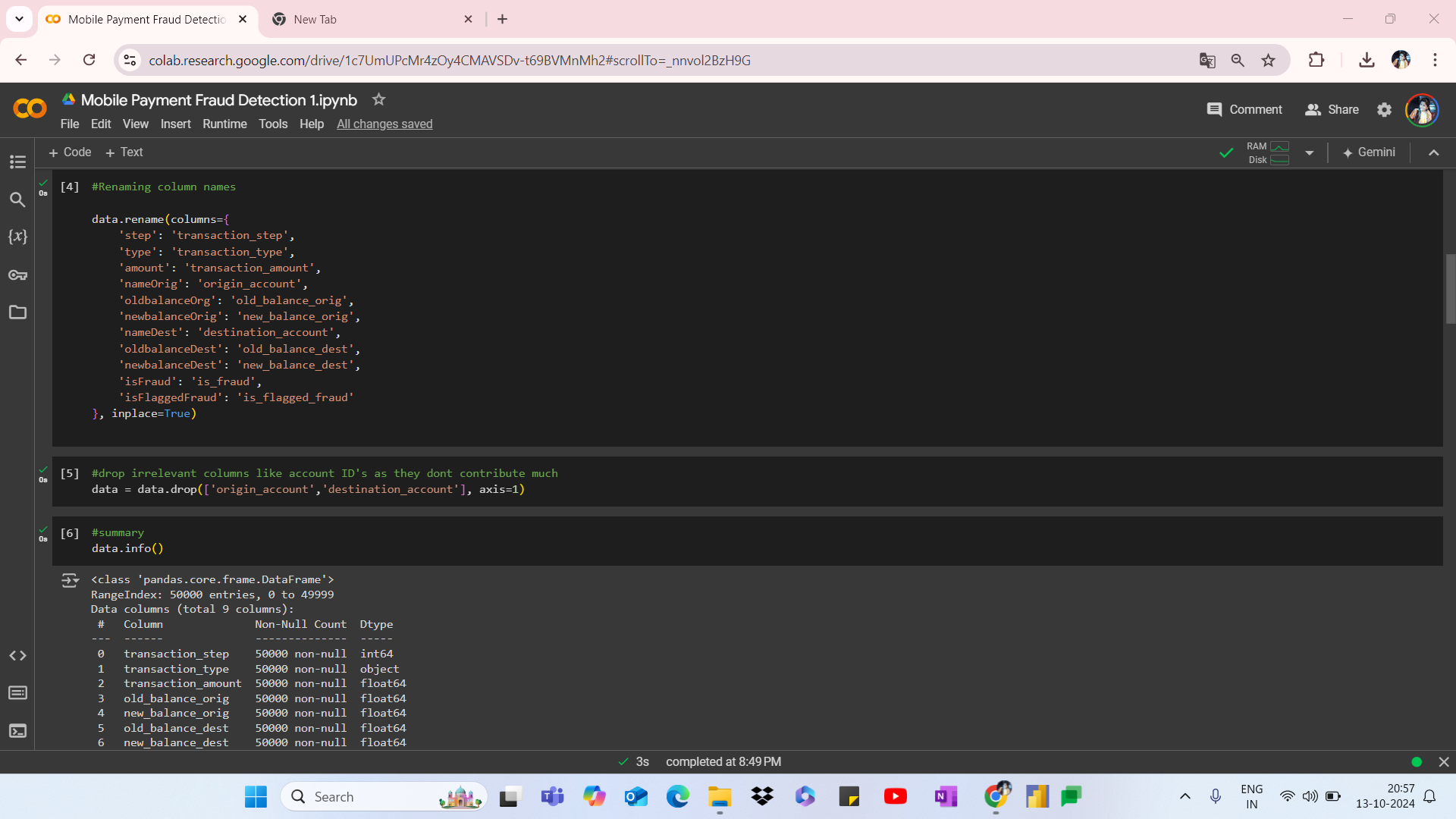
The AI-Based Fraud Detection System for E-Commerce is designed to transform the way small and medium-sized businesses protect themselves against fraudulent transactions. By leveraging machine learning algorithms, the system offers a smart, efficient, and scalable solution to detect suspicious activity in real-time, reducing potential losses.

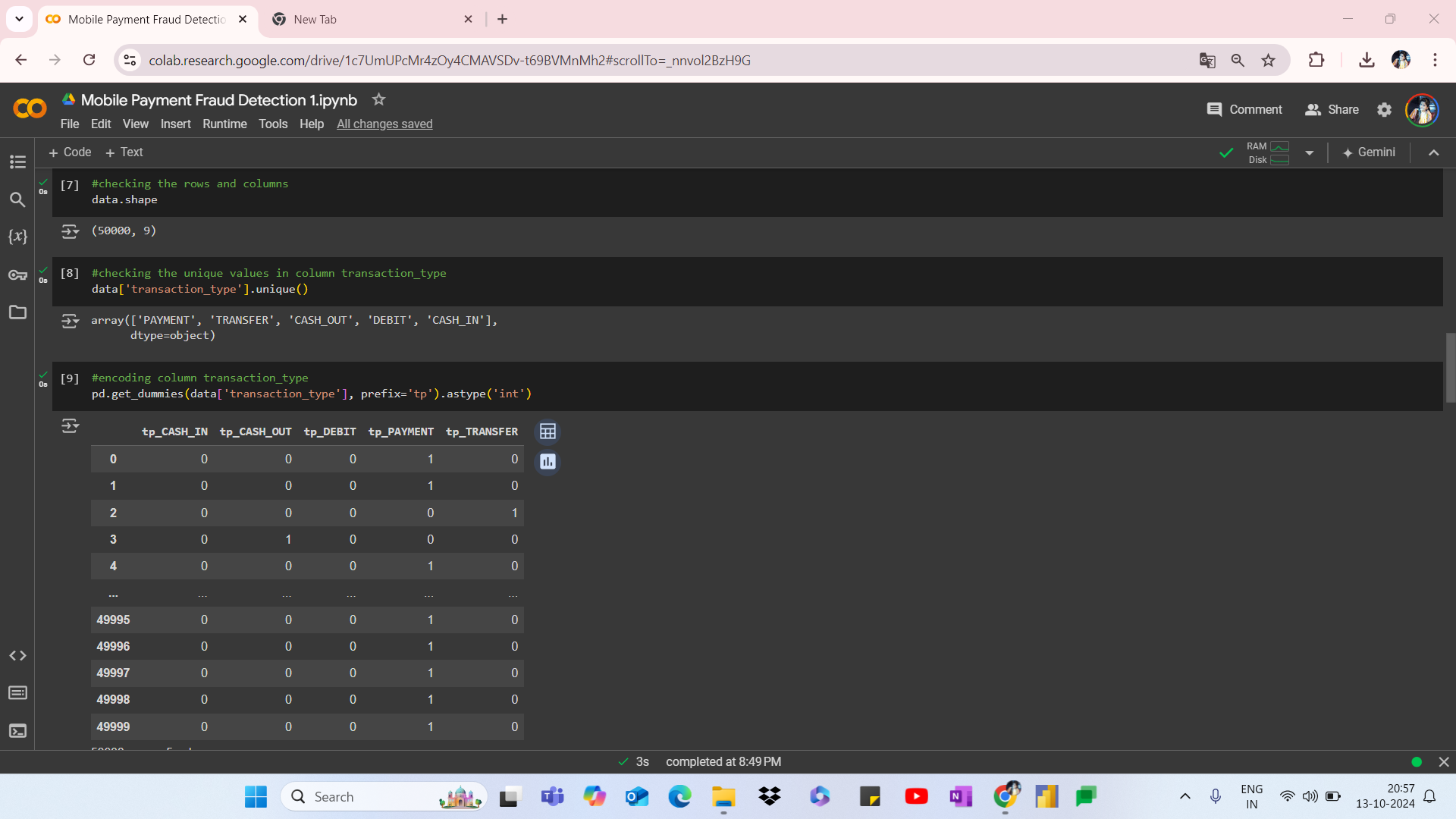
The system’s integration with existing e-commerce platforms and its ability to adapt and learn from new data ensures ongoing effectiveness in fraud prevention. The comprehensive structure—from a user-friendly interface to the backend algorithms and support systems—ensures a seamless experience for businesses. Potential revenue can be generated through tiered subscriptions, customizable features, and partnerships with payment providers.

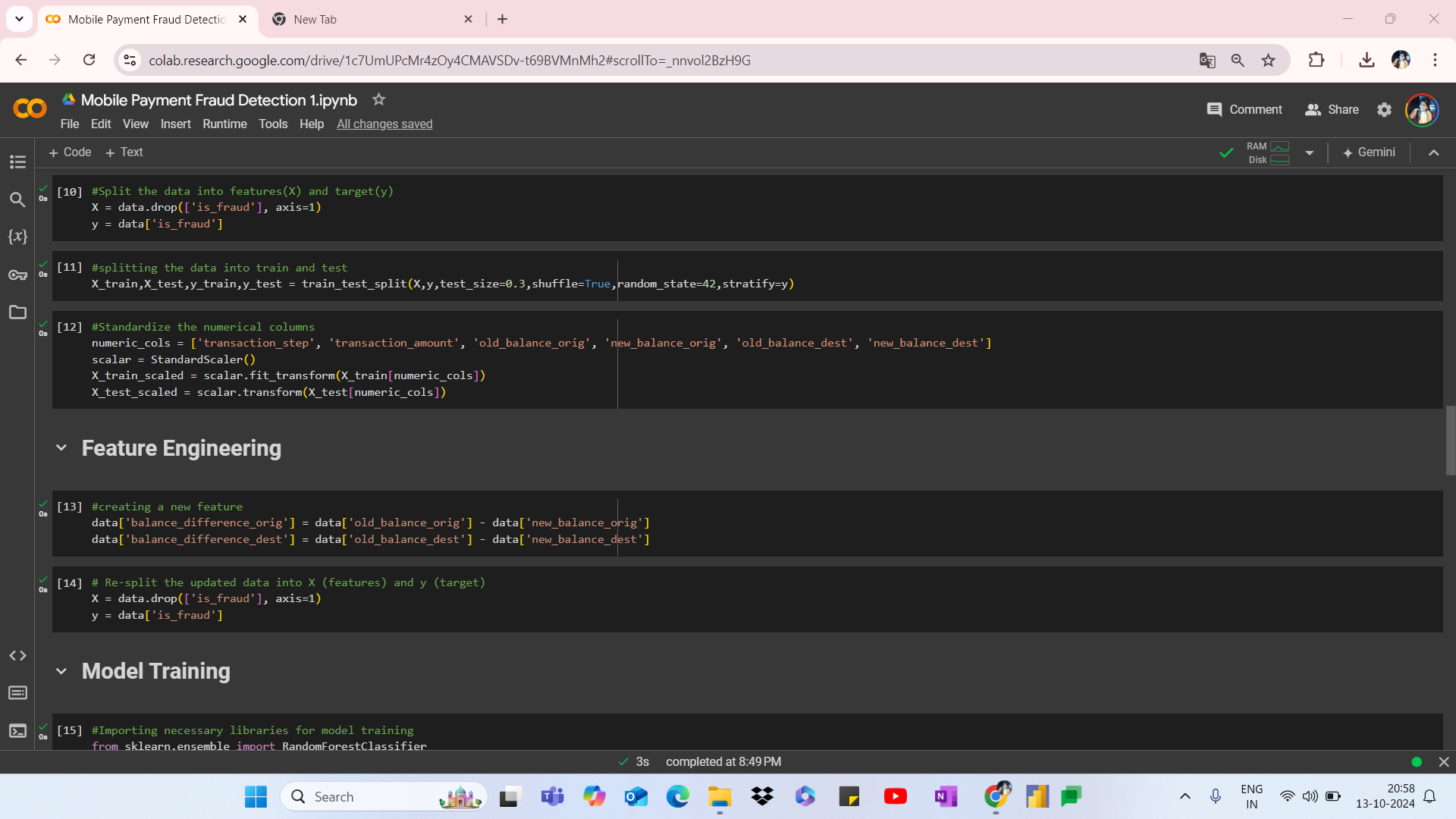
With the combined efforts of a dedicated development team, strong operational support, and ongoing updates, this solution is poised to meet the growing demand for secure, efficient fraud detection, helping businesses operate safely in the digital marketplace.

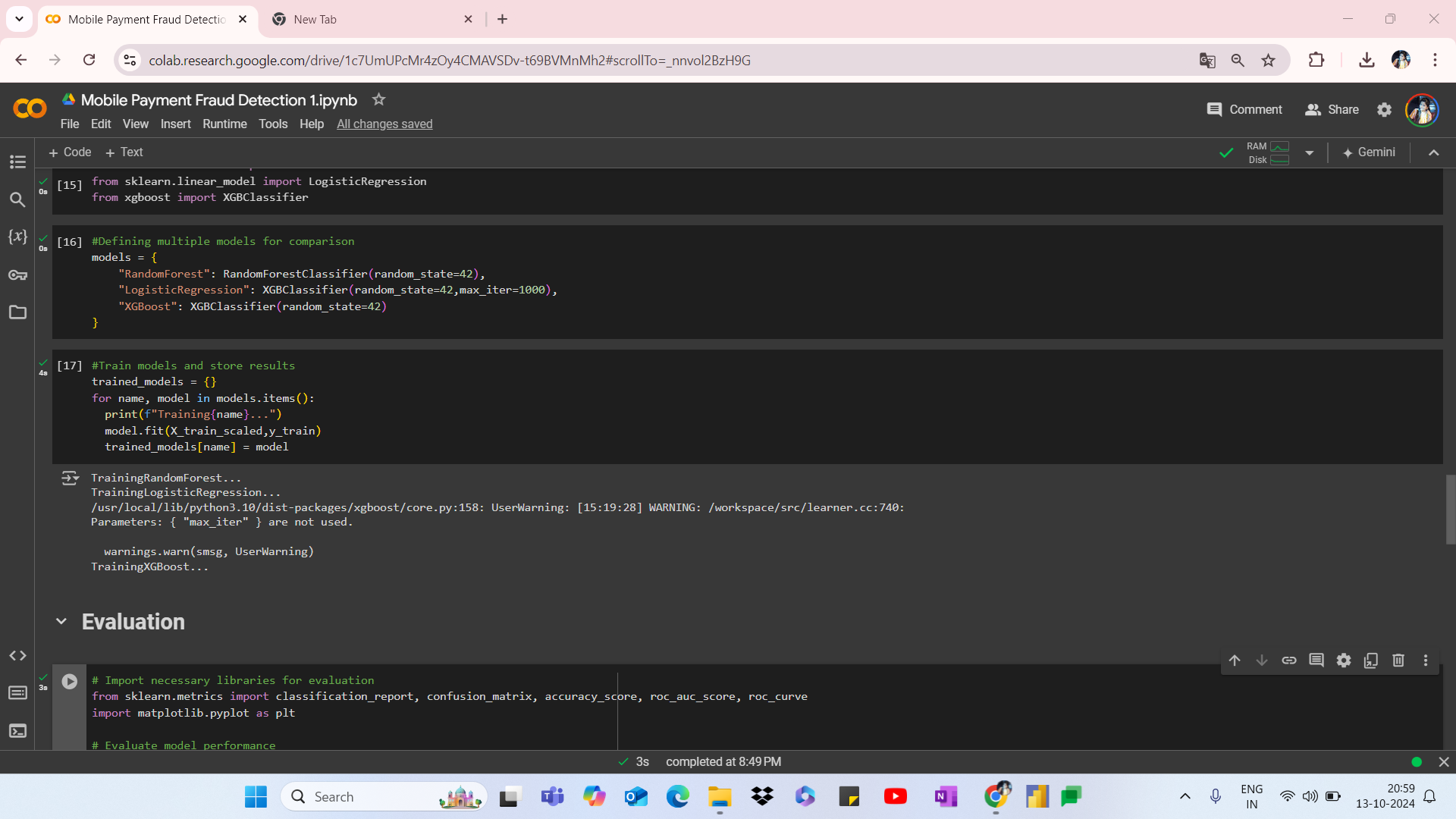
**Implementation**

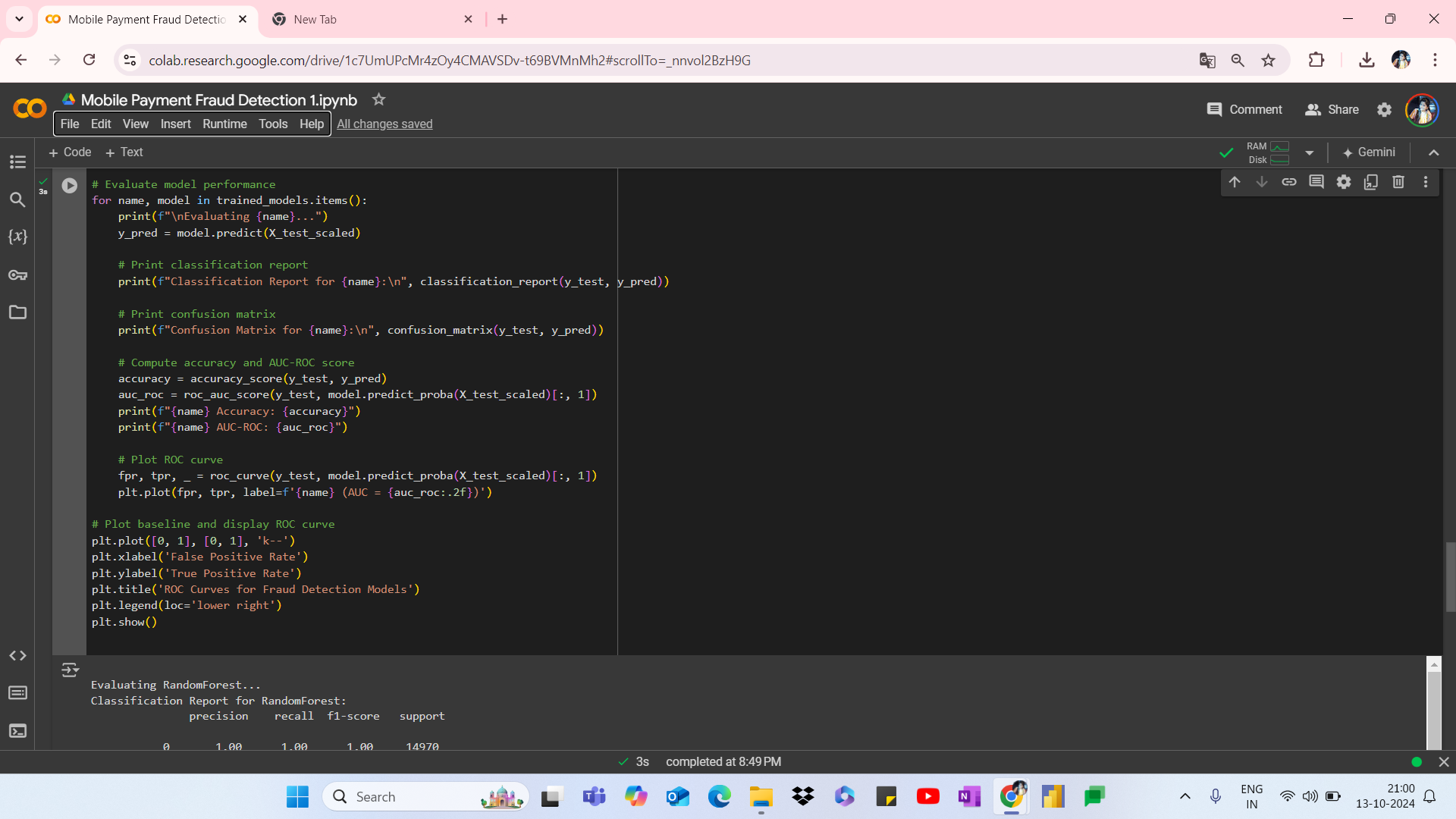
****

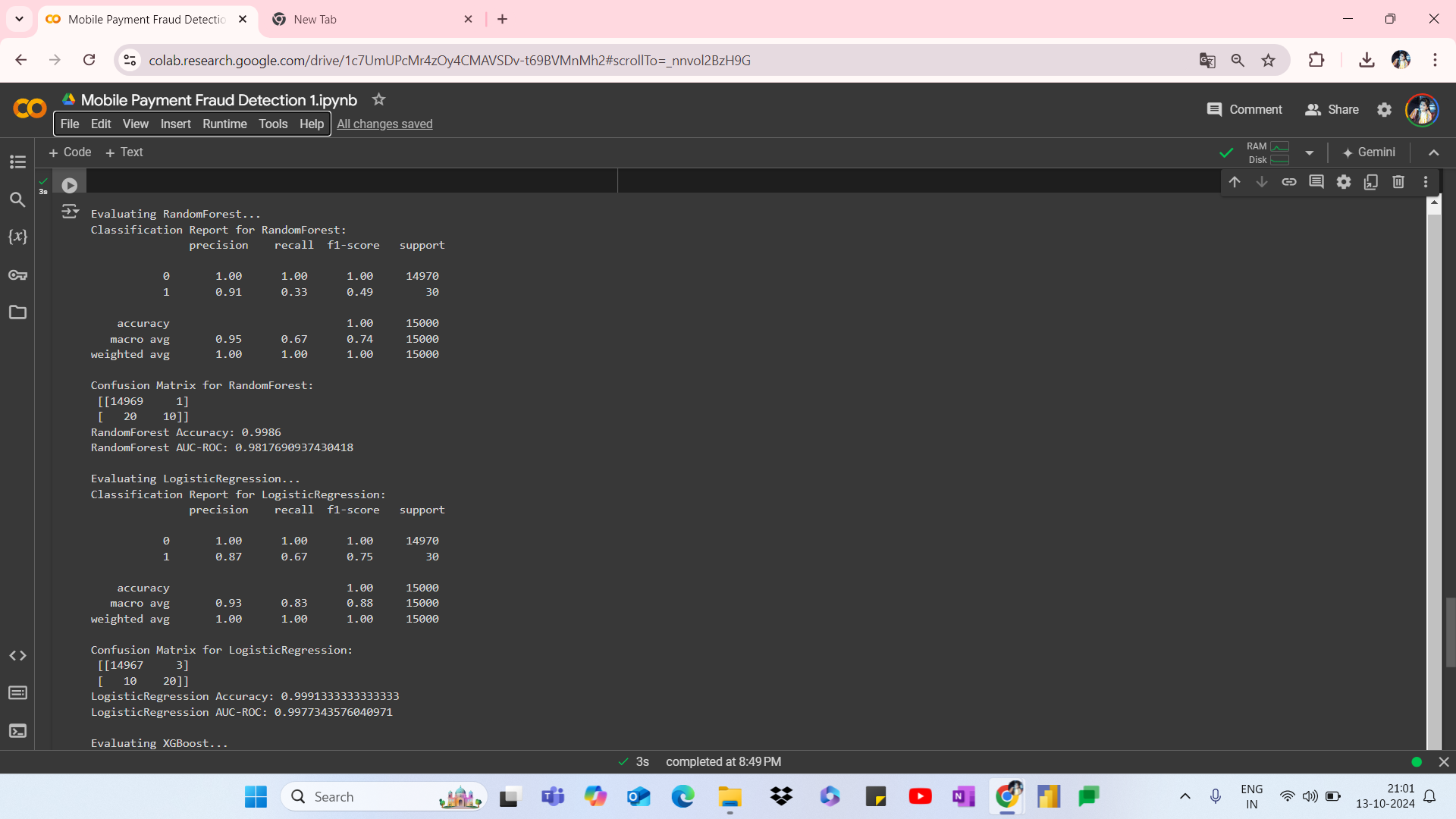
****

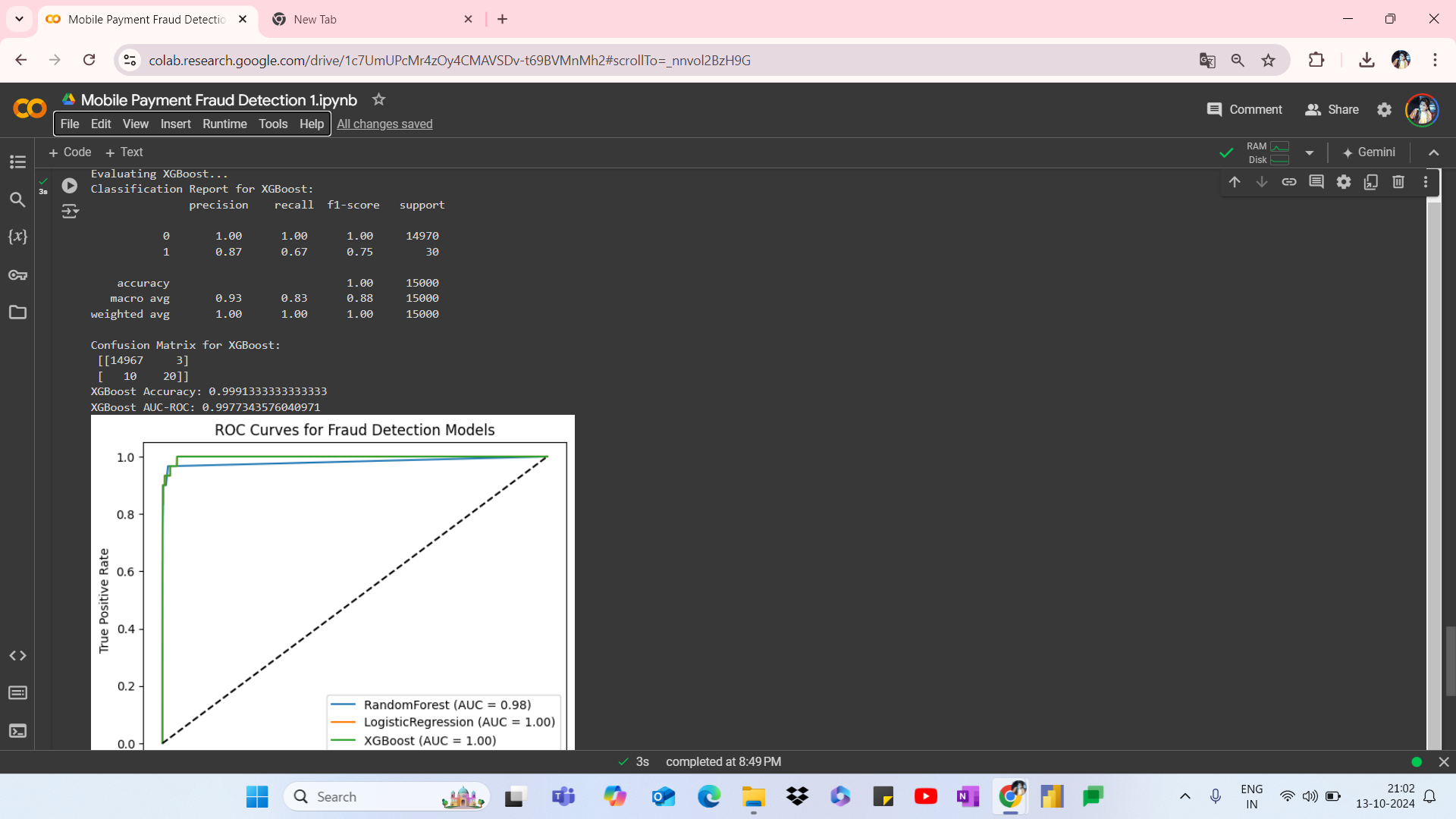
****

****

****

****

****

****

**Data Dictionary**

| **Column Name** | **Description** |
| --- | --- |
| **transaction\_step** | Represents the unit of time for the transaction (could be in hours or other time measures). It indicates when the transaction occurred in the sequence. |
| **transaction\_type** | The type of transaction performed. Common values include: PAYMENT, TRANSFER, CASH\_OUT, etc. Each type defines the action taken in the transaction. |
| **transaction\_amount** | The monetary amount involved in the transaction. |
| **origin\_account** | The identifier for the account that initiated the transaction. This could be an anonymized account number. |
| **old\_balance\_orig** | The balance of the originating account before the transaction took place. |
| **new\_balance\_orig** | The balance of the originating account after the transaction took place. |
| **destination\_account** | The identifier for the account that is the recipient of the transaction. This could be an anonymized account number. |
| **old\_balance\_dest** | The balance of the destination account before the transaction took place (in cases like TRANSFER or CASH\_OUT). |
| **new\_balance\_dest** | The balance of the destination account after the transaction took place. |
| **is\_fraud** | A binary indicator (1 for fraud, 0 for non-fraud) that marks whether the transaction was identified as fraudulent. |
| **is\_flagged\_fraud** | A binary indicator (1 for flagged fraud, 0 for not flagged). It marks whether the transaction was flagged by the system as potentially fraudulent, but might not necessarily have been classified as fraud. |

**9.0 References**

DataSet source:- GABRIEL ATKIN, "MOBILE fraud detection”. Available at:[<https://www.kaggle.com/code/gcdatkin/mobile-payment-fraud-detection/input> ]

**Machine Learning Libraries Used:-** Pedregosa, F., Varoquaux, G., Gramfort, A., et al., "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, 12, pp. 2825–2830, 2011.

McKinney, W., "Pandas: A Foundational Python Library for Data Analysis and Statistics," Python for High-Performance and Scientific Computing, 2010.

Van der Walt, S., Colbert, S.C., Varoquaux, G., "The NumPy Array: A Structure for Efficient Numerical Computation," Computing in Science & Engineering, 13, 22–30, 2011.

**Fraud Detection research Paper:-** Phua, C., Lee, V., Smith, K., and Gayler, R., "A Comprehensive Survey of Data Mining-based Fraud Detection Research," arXiv preprint arXiv:1009.6119, 2010.