**Introduction**

Employee fraud poses a significant threat to organizations worldwide, leading to financial losses, reputational damage, and compromised integrity. Detecting and preventing such fraudulent activities is crucial for maintaining organizational security and trust. In response to this challenge, our project proposes the development of an Employee Fraud Detection System. This system harnesses advanced data analytics, data mining techniques, robust systems development methodologies, and cybersecurity measures to identify suspicious activities and potential instances of fraud within the organization.

**Project Description:** The project aims to develop an employee fraud detection system using Knowledge from what I learned in the following categories:

* Networking and Cybersecurity(CIDM 6356 & CIDM 6340)
* Data Analytics (CIDM 5310)
* Data Management (CIDM 6350 & CIDM 6355)
* System Development (CIDM 6303)

The system will analyze employee behavioral data, financial transactions, and access logs to identify suspicious activities and potential instances of fraud within the organization.

**Data Collection and Integration**

I got the whole idea from CIDS 6356 to detect fraud in a company. The plan involves generating synthetic data, as discussed in CIDM 6356, to simulate financial transactions, access control systems, and HR databases. Here is the [link](https://colab.research.google.com/drive/1dC8_t5cWOVFR99QX-lVKg7piPT5VDL6M#scrollTo=nYab5Wb202Mw) to the code for making fake data. Since each time the result would be different Please find the database [here](https://drive.google.com/file/d/1JKW-ZM8iB3Lc0MUO3thiHeVc6DMuIgfV/view?usp=drive_link).

**Data Management**

I employed the same approach utilized for the project in CIDM 6355 and CIDM 6350 for this section.

**Data Management Problems**

The primary obstacle organizations face in combating employee fraud is the management and analysis of vast amounts of heterogeneous data. Employee activity logs, financial transactions, access control systems, and human resources databases all contain valuable information that, when integrated and analyzed effectively, can reveal patterns indicative of fraudulent behavior. However, the complexity and volume of this data present significant challenges in terms of collection, integration, preprocessing, and analysis.

**Motivation for Our Database Solution**

The motivation behind our database solution stems from the urgent need for organizations to proactively detect and mitigate employee fraud risks. Traditional methods of fraud detection often rely on manual processes and rule-based systems, which are limited in their ability to adapt to evolving fraud schemes and detect subtle anomalies. By leveraging advanced data analytics, machine learning algorithms, and robust system architectures, our solution aims to provide organizations with a proactive and scalable approach to fraud detection.

**Potential Benefits**

The implementation of our Employee Fraud Detection System offers several potential benefits to organizations:

1. **Early Detection:** By continuously analyzing employee behavioral data, financial transactions, and access logs, the system can identify suspicious activities at an early stage, allowing organizations to intervene before significant losses occur.
2. **Actionable Insights:** Through the application of data analytics and mining techniques, the system can extract actionable insights and patterns indicative of fraudulent behavior, enabling organizations to make informed decisions and implement targeted interventions.
3. **Enhanced Security:** By integrating robust networking and cybersecurity measures, the system ensures the secure transmission and storage of sensitive employee data and transaction records, safeguarding organizational integrity.
4. **Operational Efficiency:** The development of user-friendly interfaces, dashboards, and reporting tools streamlines the detection and investigation process, enhancing operational efficiency and reducing response times.

**Potential Users**

The Employee Fraud Detection System is designed to cater to a wide range of users within organizations, including:

1. **Risk Management Professionals:** Responsible for identifying and mitigating fraud risks, risk management professionals can leverage the system to proactively detect and prevent fraudulent activities.
2. **Financial Analysts:** By analyzing financial transactions and uncovering suspicious patterns, financial analysts can use the system to assess fraud risks and inform decision-making processes.
3. **Compliance Officers:** Charged with ensuring regulatory compliance and adherence to internal policies, compliance officers can utilize the system to monitor employee behavior and detect compliance violations.
4. **Executive Leadership:** Senior executives and board members can benefit from the system's insights and analytics to gain a comprehensive understanding of fraud risks and take appropriate strategic actions.

**Forensic objectives:**

* perform an audit to find any potential fraud in customer returns
* perform an audit to find any improper procedure violations in customer returns. Improper procedure violations are when the employee did not follow the Company Policy Regarding Customer Returns but are not necessarily fraudulent.

**Business Rules Regarding Customer Returns:**

1. Customers may return products for a return of the funds onto their debit/credit card or receive a gift card as follows.
2. For all returns, Employees shall manually record the customers ‘name, street address, city, state, and phone number into the return POS software.
3. If a receipt is present, the employee scans the receipt which causes the computer to record the return date and return price amount automatically.
4. If a receipt is present, the customer may choose to receive the funds as a gift card or to be returned to a credit/debit card.
5. If the customer chooses a gift card, the employee scans the gift card. The POS software records the gift card number and puts the return price onto the gift card automatically.
6. If the customer chooses to have the funds returned to their debit/credit card, the employee swipes the debit/credit card. The computer records the debit/credit card number and expiration date and transfers the return price to the debit/credit card automatically.
7. If the receipt is not present, the employee shall do the following:
8. Returns without receipts must be issued a gift card for store credit and shall not receive funds on a debit/credit card.
9. Employee scans the UPC barcode on the returned item. The POS software records the return price and records today’s date as the return date.
10. The employee scans the gift card. The POS software records the gift card number and puts the return price onto the card automatically.

**Enhanced Entity-Relationship Diagram (EERD)**

A diagram of a data flow

Description automatically generated with medium confidence

**Data Analytics &** **System Development**

I employed the same approach utilized for the project in CIDM 5310 and CIDM 6303 for this section.

The integrity of the database file, **myDB.db**, was meticulously verified using the SHA256 hash algorithm. The calculated hash number is **32912B9F25B591B92454BC4FA2CF146C275715CADEDA24DC85AA7E5A29A3D8FC**

A screenshot of a computer program

Description automatically generated

Image 1: Verification of Hash and File Information

If you're interested in discovering all the query codes within the SQLite file on GitHub, you can do so by following this [link](https://github.com/wtamu-babb/CIDM6395-Spring2024-FarzanehNoroozi.git)

The first query examined returns processed by employees without a receipt, highlighting those who handled more returns in this manner than the average. Notably, employees like Steven Curtis and George Stevens stood out with significantly higher-than-average returns processed without receipts. This suggests a potential need for further investigation into the reasons behind these patterns and to ensure adherence to company return policies.

In the second query, returns processed with credit/debit cards but without original receipts were analyzed. It was found that employees such as Steven Curtis, George Stevens, and Angela Diaz had the highest counts of such returns. This variance in adherence to return policies emphasizes the importance of ongoing training and potential system improvements to encourage better verification of receipts during transactions.

Moving to the third query, it revealed that all employees successfully processed credit card transactions, each with varying numbers of such transactions. Additionally, none of the employees processed returns without receipts, indicating a high level of compliance with company policies regarding return procedures.

The fourth query aimed to identify inconsistencies in return processing where different employee IDs were associated with the same customer information. However, the result showed no such occurrences, indicating a consistent process in associating each customer with a single employee, ensuring accuracy and avoiding redundancy in customer service interactions.

Query five uncovered 135 returns with unusually high prices, suggesting potential outliers in the data. The highest return price was significantly above average, prompting a recommendation to implement policies for auditing such high-value returns to enhance security and accuracy in the return process.

Finally, the sixth query checked for employees processing an unusually high number of transactions, ensuring operational integrity. Fortunately, no instances were found, indicating a stable and controlled environment for return processing across the board. This information provides valuable insights into employee behavior, adherence to policies, potential outliers in return prices, and the overall stability of return processing operations within the company.

**Networking and Cybersecurity Measures**

I employed the same approach utilized for the homework in CIDM 6340 for this section. To ensure the security and integrity of our Employee Fraud Detection System, we could implement the following networking and cybersecurity measures:

**Secure Data Transmission Protocols**: All communication within the system, including data transmission between the frontend interface, backend servers, and database, could be encrypted using HTTPS (HTTP over SSL/TLS) protocols. This ensures that sensitive data exchanged over networks is protected from interception or tampering by unauthorized parties.

**Encryption Methods**: We could employ strong encryption algorithms, such as AES (Advanced Encryption Standard), to encrypt sensitive employee data and transaction records stored in our database (at rest). Additionally, SSL/TLS encryption will be utilized to encrypt data transmitted over networks (in transit), ensuring end-to-end encryption and safeguarding the confidentiality of data.

**Access Control Mechanisms**: Robust access control mechanisms, including role-based access control (RBAC) and attribute-based access control (ABAC), could be implemented to regulate access to sensitive data within the system. Only authorized users with appropriate permissions will be granted access to view or modify sensitive information, thereby preventing unauthorized access.

**Multi-Factor Authentication (MFA)**: To enhance security, we could implement multi-factor authentication (MFA) mechanisms requiring users to provide multiple forms of verification before accessing the system. This includes passwords, biometrics, or one-time passcodes, adding an extra layer of security and mitigating the risk of unauthorized access.

**Intrusion Detection and Prevention Systems (IDPS)**: We could deploy intrusion detection and prevention systems (IDPS) to monitor network traffic in real-time and detect suspicious activities or potential security breaches. Intrusion detection systems (IDS) will be utilized to identify unauthorized access attempts or anomalous behavior within the system, enabling prompt response and mitigation of security threats.

**Regular Security Audits and Penetration Testing**: Regular security audits and penetration testing could be conducted to identify vulnerabilities within the system and address them proactively. Various cyber threats and attacks, including SQL injection, cross-site scripting (XSS), and data breaches, will be simulated to assess the system's resilience and security posture.

**Data Masking and Anonymization**: Personally identifiable information (PII), such as employee names, addresses, and credit card numbers, will be masked or anonymized to prevent unauthorized access or disclosure. This ensures that sensitive information remains protected from identity theft or data leakage.

**Secure Backup and Disaster Recovery**: We could implement secure backup and disaster recovery procedures to ensure the availability and integrity of data in case of system failures, natural disasters, or cyber attacks. Critical data will be regularly backed up and stored in encrypted, geographically redundant locations, facilitating timely recovery and continuity of operations.

By integrating one or all of these networking and cybersecurity measures into our Employee Fraud Detection System, we aim to safeguard sensitive employee data and transaction records, ensuring confidentiality, integrity, and availability throughout the system's lifecycle.

**Appendix:**

**Returns Processed by the Same Employee without a Receipt**

SELECT E.EmployeeID, E.Name AS EmployeeName, COUNT(\*) as NumReturns

FROM Returns R

JOIN Employee E ON R.EmployeeID = E.EmployeeID

WHERE R.IsReceiptPresent = 'False'

GROUP BY R.EmployeeID, E.Name

HAVING NumReturns > 1;

|  |  |  |  |
| --- | --- | --- | --- |
| Num | EmployeeID | EmployeeName | NumReturns |
| 1 | 79023 | Steven Curtis | 46 |
| 2 | 52272 | George Stevens | 40 |
| 3 | 29733 | Angela Diaz | 38 |
| 4 | 27935 | Dr. Thomas Weiss PhD | 36 |
| 5 | 61373 | Chad Dawson | 35 |
| 6 | 81344 | Seth Mckinney | 35 |
| 7 | 41845 | Samuel Obrien | 33 |
| 8 | 92530 | Jessica Short | 33 |
| 9 | 71186 | Katherine Brown | 32 |
| 10 | 54061 | Harry Chen | 31 |
| 11 | 40295 | Anthony Brown | 29 |
| 12 | 57258 | Jonathan Fowler | 29 |
| 13 | 80899 | Timothy Weber | 29 |
| 14 | 84966 | Judy Garcia | 29 |
| 15 | 97986 | David Mullen | 29 |
| 16 | 60657 | Amanda Anderson | 28 |
| 17 | 19694 | Andrew Owens | 26 |
| 18 | 69674 | Kristopher Mendez | 25 |
| 19 | 84034 | Susan Adkins | 25 |
| 20 | 26122 | Lindsey Maxwell | 24 |
| 21 | 26093 | Melissa Garcia | 20 |
| 22 | 28960 | Kaitlyn Matthews | 19 |
| 23 | 94877 | Kyle Brown | 15 |

Table 1:Returns Processed by the Same Employee without a Receipt

The average number of returns processed by an employee without a receipt is 29.04, with a standard deviation of 8.33, indicating the variability in this process across employees. To identify outliers, employees whose number of returns without a receipt significantly deviates from the average, considering a deviation of more than one standard deviation, are examined. Notably, Steven Curtis (EmployeeID: 79023) processed 46 returns without a receipt, followed by George Stevens (EmployeeID: 52272) with 40 returns, Angela Diaz (EmployeeID: 29733) with 38 returns, Dr. Thomas Weiss PhD (EmployeeID: 27935) with 36 returns, and Chad Dawson (EmployeeID: 61373) with 35 returns. These individuals demonstrate a significantly higher-than-average volume of returns processed without a receipt, suggesting a need for further investigation to understand the underlying reasons and ensure adherence to company policies regarding return procedures.

**Returns with Debit/Credit Card Refund and No Original Receipt**

SELECT E.EmployeeID, E.Name AS EmployeeName, COUNT(\*) as NumReturnsWithoutReceipt

FROM Returns R

JOIN Employee E ON R.EmployeeID = E.EmployeeID

WHERE R.IsReceiptPresent = 'False' AND R.CreditCardNum IS NOT NULL

GROUP BY E.EmployeeID, E.Name;

|  |  |  |
| --- | --- | --- |
| EmployeeID | EmployeeName | NumReturnsWithoutReceipt |
| 79023 | Steven Curtis | 46 |
| 52272 | George Stevens | 40 |
| 29733 | Angela Diaz | 38 |
| 27935 | Dr. Thomas Weiss PhD | 36 |
| 61373 | Chad Dawson | 35 |
| 81344 | Seth Mckinney | 35 |
| 41845 | Samuel Obrien | 33 |
| 92530 | Jessica Short | 33 |
| 71186 | Katherine Brown | 32 |
| 54061 | Harry Chen | 31 |
| 40295 | Anthony Brown | 29 |
| 57258 | Jonathan Fowler | 29 |
| 80899 | Timothy Weber | 29 |
| 84966 | Judy Garcia | 29 |
| 97986 | David Mullen | 29 |
| 60657 | Amanda Anderson | 28 |
| 19694 | Andrew Owens | 26 |
| 69674 | Kristopher Mendez | 25 |
| 84034 | Susan Adkins | 25 |
| 26122 | Lindsey Maxwell | 24 |
| 26093 | Melissa Garcia | 20 |
| 28960 | Kaitlyn Matthews | 19 |
| 94877 | Kyle Brown | 15 |

Table 2:The count of returns without receipts processed by each employee

The analysis examines returns processed by each employee where credit/debit cards were used for refunds without original receipts. Steven Curtis leads with 46 returns, followed closely by George Stevens with 40 and Angela Diaz with 38. The distribution highlights varying levels of adherence to return policies, prompting the need for further investigation into why original receipts are missing in these transactions. To address this, regular training sessions and system improvements to prompt receipt verification are recommended to ensure compliance and minimize potential errors or fraud in return processing procedures.

**Returns Processed by each employee without Receipts and with Credit Card Transactions**

SELECT e.EmployeeID, e.Name,

COUNT(CASE WHEN r.IsReceiptPresent = FALSE THEN 1 END) AS ReturnsWithoutReceipts,

COUNT(CASE WHEN r.CreditCardNum IS NOT NULL THEN 1 END) AS CreditCardTransactions

FROM employee e

LEFT JOIN returns r ON e.EmployeeID = r.EmployeeId

GROUP BY e.EmployeeID, e.Name;

|  |  |  |  |
| --- | --- | --- | --- |
| EmployeeID | Name | ReturnsWithoutReceipts | CreditCardTransactions |
| 29733 | Angela Diaz | 0 | 69 |
| 52272 | George Stevens | 0 | 69 |
| 27935 | Dr. Thomas Weiss PhD | 0 | 67 |
| 92530 | Jessica Short | 0 | 67 |
| 54061 | Harry Chen | 0 | 66 |
| 79023 | Steven Curtis | 0 | 65 |
| 57258 | Jonathan Fowler | 0 | 61 |
| 41845 | Samuel Obrien | 0 | 60 |
| 84034 | Susan Adkins | 0 | 60 |
| 84966 | Judy Garcia | 0 | 60 |
| 69674 | Kristopher Mendez | 0 | 57 |
| 81344 | Seth Mckinney | 0 | 57 |
| 97986 | David Mullen | 0 | 57 |
| 61373 | Chad Dawson | 0 | 55 |
| 71186 | Katherine Brown | 0 | 54 |
| 19694 | Andrew Owens | 0 | 52 |
| 40295 | Anthony Brown | 0 | 52 |
| 60657 | Amanda Anderson | 0 | 51 |
| 80899 | Timothy Weber | 0 | 49 |
| 26093 | Melissa Garcia | 0 | 46 |
| 28960 | Kaitlyn Matthews | 0 | 45 |
| 26122 | Lindsey Maxwell | 0 | 44 |
| 94877 | Kyle Brown | 0 | 43 |

Table 3: Employee without Receipts and with Credit Card Transactions

we can observe that each employee has successfully processed credit card transactions, with varying numbers of such transactions. Additionally, none of the employees have processed returns without receipts. This indicates a high level of adherence to company policies regarding return procedures and suggests efficient handling of credit card transactions across the board.

This data could be used to assess employee performance in terms of adherence to company policies regarding returns and credit card transactions. It also provides insights into areas where further training or guidance may be needed, such as ensuring consistent compliance with return policies or optimizing credit card transaction processes.

**Returns Processed with Different Employee IDs but Same Customer Information**

SELECT CustomerName, Street, City, State, CustomerPhone, COUNT(DISTINCT EmployeeID) AS NumEmployees

FROM Returns

GROUP BY CustomerName, Street, City, State, CustomerPhone

HAVING NumEmployees > 1;

The query identifies instances where returns are processed with different employee IDs but share the same customer information. By grouping returns based on customer details and counting distinct employee IDs, it detects potential inconsistencies in handling returns for the same customer. However, the result reveals no such occurrences, indicating that the return process maintains consistency in associating each customer with a single employee, ensuring accuracy and preventing redundancy in customer service interactions.

**Returns Processed with Unusually High Return Price**

SELECT ReturnId, ReturnPrice

FROM Returns

WHERE ReturnPrice > (SELECT AVG(ReturnPrice) \* 2 FROM Returns);

|  |  |
| --- | --- |
| ReturnId | ReturnPrice |
| 6272 | 1290690993 |
| 4141 | 1155423713 |
| 6651 | 998571806 |
| 8518 | 992847328 |
| 4772 | 987716552 |
| 6993 | 974698248 |
| 2306 | 961414129 |
| 2133 | 935456677 |
| 182 | 931165902 |
| 1482 | 927153189 |
| 9696 | 919549129 |
| 1336 | 914887550 |
| 9104 | 913195879 |
| 6297 | 909979615 |
| 7632 | 905438471 |
| 6944 | 895201776 |
| 5544 | 891588257 |
| 4986 | 879235101 |
| 1284 | 878234576 |

Table 4: Returns Processed with Unusually High Return Price

The query reveals 135 returns flagged for unusually high prices. The highest return price, at 1290690993, indicates a significant outlier. Even the lowest return price, 131529198, remains notably high. To enhance security, implementing a rule in company policies to audit such high-value returns could be beneficial.

**Returns Processed with Unusually High Number of Transactions per Employee**

SELECT EmployeeID, COUNT(\*) AS NumTransactions

FROM Returns

GROUP BY EmployeeID

HAVING COUNT(\*) > (SELECT AVG(NumTransactions) \* 2 FROM (SELECT EmployeeID, COUNT(\*) AS NumTransactions FROM Returns GROUP BY EmployeeID) AS T);

The query confirms that there are no occurrences of returns processed with an unusually high number of transactions per employee. By analyzing each employee's transaction count and comparing it to double the average count, it ensures consistency across the board. This finding reflects a stable and controlled environment for return processing, where no employee demonstrates an abnormal surge in transactions, ensuring operational integrity.