**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/19tvWCtN1lWTZddB8RDBThGfwwkEvzuc7#scrollTo=UgY7rzoTk9_G) 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/1m7Fv_hIkJsBPsNvckV2Rykx0ZvH9uo_w/view?usp=sharing).

**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;

|  |  |  |
| --- | --- | --- |
| EmployeeID | EmployeeName | NumReturns |
| 46255 | Anthony Martinez | 35 |
| 55191 | Sarah Hess | 34 |
| 31836 | Carlos Frazier | 33 |
| 55991 | Janet Howard | 33 |
| 64690 | Neil Liu | 33 |
| 95482 | Larry Hernandez | 33 |
| 14930 | Sheila Rodriguez | 32 |
| 58835 | Katie Delgado | 32 |
| 87030 | Michael Walton | 31 |
| 98257 | Kelsey Barton | 31 |
| 41343 | Billy Tran | 29 |
| 70377 | Jonathan Nicholson | 29 |
| 23266 | Savannah Davis | 28 |
| 41485 | David Oliver | 28 |
| 55718 | Nathan Morris | 28 |
| 70459 | John Gay | 28 |
| 91172 | Lori Jacobson | 28 |
| 92845 | Ms. Julia Becker | 27 |
| 52530 | Mark Bell | 26 |
| 96090 | Judy Kirk | 26 |
| 53250 | Dawn Stevens | 25 |
| 10679 | Julie Harris MD | 24 |
| 57247 | Kerry Ramos | 19 |

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

The analysis reveals that employees exhibit varying frequencies in processing returns without a receipt. Notably, Anthony Martinez stands out with 35 such returns, closely followed by Sarah Hess, Carlos Frazier, and others. With a standard deviation of 3.8, we observe a moderate level of dispersion in the number of returns processed without a receipt across employees. This suggests some variability in performance among employees, albeit within a relatively constrained range. Further investigation into factors contributing to this variance, such as transaction processing protocols and employee training, may offer insights into optimizing efficiency and consistency in handling returns without receipts.

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

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

FROM Return R

JOIN Employee E ON R.EmployeeID = E.EmployeeID

JOIN Customer C ON R.CustomerID = C.CustomerID

WHERE R.IsReceiptPresent = 0 AND C.CreditCardNum IS NOT NULL

GROUP BY E.EmployeeID, E.Name;

|  |  |  |
| --- | --- | --- |
| EmployeeID | EmployeeName | NumReturnsWithoutReceipt |
| 46255 | Anthony Martinez | 36 |
| 55191 | Sarah Hess | 35 |
| 55991 | Janet Howard | 34 |
| 95482 | Larry Hernandez | 34 |
| 14930 | Sheila Rodriguez | 33 |
| 31836 | Carlos Frazier | 33 |
| 58835 | Katie Delgado | 33 |
| 64690 | Neil Liu | 33 |
| 87030 | Michael Walton | 32 |
| 98257 | Kelsey Barton | 31 |
| 41343 | Billy Tran | 30 |
| 23266 | Savannah Davis | 29 |
| 70377 | Jonathan Nicholson | 29 |
| 41485 | David Oliver | 28 |
| 55718 | Nathan Morris | 28 |
| 70459 | John Gay | 28 |
| 91172 | Lori Jacobson | 28 |
| 92845 | Ms. Julia Becker | 27 |
| 96090 | Judy Kirk | 27 |
| 52530 | Mark Bell | 26 |
| 53250 | Dawn Stevens | 25 |
| 10679 | Julie Harris MD | 24 |
| 57247 | Kerry Ramos | 19 |

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

Upon analysis, it's evident that employees handle returns involving debit/credit card refunds and no original receipt with varying frequencies. Notably, Anthony Martinez and Sarah Hess top the list with 36 and 35 such returns, respectively. With a standard deviation of 4.07, we observe a slightly higher level of dispersion compared to the previous query, indicating a broader range of variability in handling these specific types of returns among employees. This variance may stem from factors such as transaction complexity, customer interactions, and adherence to refund policies. Identifying and addressing the underlying causes of this variability can aid in standardizing processes, enhancing employee training, and ultimately improving the efficiency and consistency of handling returns with debit/credit card refunds and no original receipt.

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

SELECT E.EmployeeID, E.Name,

SUM(CASE WHEN R.IsReceiptPresent = 'False' THEN 1 ELSE 0 END) AS ReturnsWithoutReceipts,

SUM(CASE WHEN C.CreditCardNum IS NOT NULL THEN 1 ELSE 0 END) AS CreditCardTransactions

FROM Employee E

LEFT JOIN Return R ON E.EmployeeID = R.EmployeeID

LEFT JOIN Customer C ON R.CustomerID = C.CustomerID

GROUP BY E.EmployeeID, E.Name;

|  |  |  |  |
| --- | --- | --- | --- |
| EmployeeID | Name | ReturnsWithoutReceipts | CreditCardTransactions |
| 55191 | Sarah Hess | 0 | 76 |
| 31836 | Carlos Frazier | 0 | 74 |
| 64690 | Neil Liu | 0 | 70 |
| 55991 | Janet Howard | 0 | 67 |
| 55718 | Nathan Morris | 0 | 63 |
| 14930 | Sheila Rodriguez | 0 | 62 |
| 41343 | Billy Tran | 0 | 61 |
| 23266 | Savannah Davis | 0 | 60 |
| 46255 | Anthony Martinez | 0 | 60 |
| 91172 | Lori Jacobson | 0 | 60 |
| 95482 | Larry Hernandez | 0 | 60 |
| 87030 | Michael Walton | 0 | 58 |
| 70377 | Jonathan Nicholson | 0 | 56 |
| 92845 | Ms. Julia Becker | 0 | 56 |
| 52530 | Mark Bell | 0 | 55 |
| 57247 | Kerry Ramos | 0 | 55 |
| 98257 | Kelsey Barton | 0 | 55 |
| 58835 | Katie Delgado | 0 | 53 |
| 41485 | David Oliver | 0 | 52 |
| 70459 | John Gay | 0 | 45 |
| 96090 | Judy Kirk | 0 | 45 |
| 10679 | Julie Harris MD | 0 | 44 |
| 53250 | Dawn Stevens | 0 | 41 |

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 C.CustomerName, C.Street, C.City, C.State, C.CustomerPhone, COUNT(DISTINCT R.EmployeeID) AS NumEmployees

FROM Return R

JOIN Customer C ON R.CustomerID = C.CustomerID

GROUP BY C.CustomerName, C.Street, C.City, C.State, C.CustomerPhone

HAVING NumEmployees > 1;

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| CustomerName | Street | City | State | CustomerPhone | NumEmployees |
| Caleb Zimmerman | 239 Fuller Glens Suite 371 | Seattle | WA | 212-753-2197 | 5 |
| Jodi Curtis | 4065 Jason Loop Suite 364 | Seattle | WA | 459-479-6854 | 5 |
| Aaron Miller | 0022 Chris Isle | Seattle | WA | 966-370-2811 | 4 |
| Andrew Ward | 4765 Torres Highway | Seattle | WA | (646)901-7069 | 4 |
| Angelica Perez | 6340 Mary Glens Suite 706 | Seattle | WA | 600-623-4207 | 4 |
| Austin Hays | 3491 Amber Shores Suite 307 | Seattle | WA | 8843350445 | 4 |
| Barbara Martinez | 227 Brian Pine Suite 759 | Seattle | WA | 001-613-843- | 4 |
| Connor Armstrong PhD | 76291 Myers Roads Apt. 707 | Seattle | WA | 939.928.0538 | 4 |
| Harry Palmer | 677 Victoria Hill Suite 990 | Seattle | WA | 8799208008 | 4 |
| Joshua Long | 6248 Brennan Island Apt. 128 | Seattle | WA | 567-861-4724 | 4 |
| Matthew Harrison | 04844 Moses Cove | Seattle | WA | (472)809-7551 | 4 |
| Michael Haas | 87958 Ronald Viaduct | Seattle | WA | 973-343-6195 | 4 |
| Mr. John Braun | 1222 Mcfarland Road | Seattle | WA | 503-954-9337 | 4 |
| Nathan Lewis | 378 Tristan Fords Suite 443 | Seattle | WA | 626-581-5835 | 4 |
| Paula Martin | 933 Michael Common | Seattle | WA | (531)269-6633 | 4 |
| Phillip Burton | 029 Patrick Knoll | Seattle | WA | 408-401-7912 | 4 |
| Rebecca Powell | 497 Jennifer Lane Apt. 931 | Seattle | WA | 763-299-1549 | 4 |
| Richard Hernandez | 415 Elizabeth Trace | Seattle | WA | 515-601-7143 | 4 |
| Robert Hale | 401 Powell Pines | Seattle | WA | 658-421-1624 | 4 |
| Sarah Welch | 661 Johnson Motorway | Seattle | WA | 744.373.6969 | 4 |
| Stephanie Williams | 6036 Robertson Divide | Seattle | WA | 532-209-2269 | 4 |
| Thomas Brown | 95958 Young Meadow | Seattle | WA | 785-937-8100 | 4 |
| Tyrone Andersen | 940 Judy Underpass | Seattle | WA | 841.307.2502 | 4 |
| Alan Romero | 7120 Eric Walks Apt. 888 | Seattle | WA | 260.513.4795 | 3 |
| Amy Chan | 93480 Rachel Mountain Suite 538 | Seattle | WA | 934-964-9447 | 3 |
| Andre Hall | 279 Matthew Groves | Seattle | WA | 810-447-7786 | 3 |
| Andrea Casey | 132 Summers Crescent | Seattle | WA | 827-314-6262 | 3 |
| Angela Martin | 3224 Lori Fall Suite 478 | Seattle | WA | 333-913-9721 | 3 |

Table 4:Returns Processed with Different Employee IDs but Same Customer Information

In this analysis, we've identified that out of a total of 364 customers, 276 customers had returns processed by 2 employees, 64 customers by 3 employees, 21 customers by 4 employees, and 2 customers by 5 employees. This indicates a significant number of customers with returns handled by multiple employees, suggesting potential inconsistencies in service provision. To address this issue, it's crucial to identify and rectify the root causes, such as variations in employee training or communication practices.

**Returns Processed with Unusually High Return Price**

SELECT ReturnId, ReturnPrice

FROM Return

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

The query reveals no returns flagged for unusually high prices. 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 Return

GROUP BY EmployeeID

HAVING COUNT(\*) > (SELECT AVG(NumTransactions) \* 2 FROM (SELECT EmployeeID, COUNT(\*) AS NumTransactions FROM Return 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.