Visual Analytics Group 11

Prompt:
Developing an Advanced Fraud Detection and Prevention System for Financial Services

1. **Summary:**

   In an era where financial crimes continue to evolve, the "Fraud Detection and Prevention" project aims to develop an advanced model for detecting fraudulent activities within the financial services.. Leveraging data analytics and Tableau, this project provides a comprehensive solution to identify and visualize potentially fraudulent transactions and activities.

2. **Objective:**

   The primary objectives of this project are as follows:

- Fraud Detection Model: Create an advanced machine learning model that can effectively detect and classify fraudulent activities from legitimate transactions. This model should continuously learn and adapt to emerging fraud patterns.
- Visualization with Tableau: Use Tableau to build an interactive and user-friendly dashboard that provides insights into fraud trends and incidents. This will empower organizations to monitor and prevent fraudulent activities.
- Alerts and Reporting: Implement an alerting system within the Tableau dashboard to immediately notify relevant personnel when suspicious activities are detected. Additionally, generate detailed reports for fraud investigations.
- Adaptability: Develop a solution that evolves to address new types of fraud and remains up-to-date with emerging trends in financial crime.

3. **Target Audience:**

   Our project caters to professionals and stakeholders within the financial services and insurance sectors, particularly:

- Chief Risk Officers (CROs)
- Fraud Analysts and Investigators
- Data Scientists and Analysts
- Compliance Officers
- C-Suite Executives

   Persona - "Alexandra Smith" (Chief Risk Officer):

   Name: Alexandra Smith
   Age: 48

Gender: Female
Location: New York City, New York

Description of the User:
Alexandra Smith is a seasoned Chief Risk Officer (CRO) with extensive experience in the financial services sector. She is responsible for overseeing all risk management activities within her financial institution, which includes proactively identifying and addressing financial fraud. Alexandra recognizes the critical importance of staying ahead of evolving fraud schemes and is committed to leveraging cutting-edge technology to protect her organization.

Key Goals:

- Ensuring the financial institution is safeguarded against financial fraud.
- Implementing innovative solutions for fraud detection and prevention.
- Keeping up with the latest industry trends in financial crime prevention.
- Collaborating with the fraud analysis team to investigate and mitigate fraud.
- Effectively communicating fraud-related insights to the organization's senior management and stakeholders.

This project will cater to Alexandra's needs by providing her with an advanced fraud detection model and Tableau dashboard that offers real-time insights into potential fraudulent activities, thereby enhancing the security and reputation of her financial institution.

4. Data Source

a. Reasons for choosing this dataset (2-4 lines)

This dataset was chosen due to its relevance in the context of financial crime prevention, which is a critical concern for organizations in the financial services and insurance industries. Leveraging Tableau for visualization, we aim to demonstrate how data analytics can be used to detect and prevent fraudulent activities.

b. Construct a description of the dataset.

The dataset contains 2 sets of data i.e., fraudtrain.csv with 22 columns and 1048574 rows and fraudtest.csv with 22 columns and 555718 rows, which is ideal for our objective of building a fraud detection model. We will not be limiting ourselves to this dataset and are ambitious about finding more relevant datasets that will cater to our project objectives.

Additionally, transaction-related information makes it suitable for creating and testing fraud detection models. The dataset's attributes, such as transaction date and time, merchant details, and transaction amount, align with the project's objectives.

A table describing each variable in a few words.

Sl no	Variable name	Description
1	trans_date_trans_time	This variable is discrete in nature and is a date/time data type.It represents the date and time of the transaction.
2	cc_num	Unique Identifier This variable is discrete in nature and is an integer data type. It represents the Credit card number used in making purchases
3	merchant	This variable is discrete in nature and is a string data type. It represents the name of the merchants that are getting paid
4	category	This variable is discrete in nature and is a string data type. It represents the name of the category the merchant deal in
5	amt	This variable is continuous in nature and is a numeric data type. It represents the amount of money spent by the customer in USD
6	first	This variable is discrete in nature and is a string data type. It represents the first name of the card holder
7	last	This variable is discrete in nature and is a string data type. It represents the last name of the card holder
8	gender	This variable is discrete in nature and is a string data type. It represents the gender of the card holder. F = Female, M = Male
9	street	This variable is discrete in nature and is a string data type. It represents the street of the card holder’s residence
10	city	This variable is discrete in nature and is a string data type. It represents the city of the card holder’s residence
11	state	This variable is discrete in nature and is a string data type. It represents the state of the card holder’s residence
12	zip	This variable is discrete in nature and is an integer data type. It represents the zipcode of the card holder’s residence
13	lat	This variable is discrete in nature and is a Geospatial data type. It represents the Latitude of the card holder’s residence
14	long	This variable is discrete in nature and is a Geospatial data type. It represents the Longitude of the card holder’s residence
15	city_pop	This variable is continuous in nature and is an integer data type. It

		represents the population of the city the card holder's reside in
16	job	This variable is discrete in nature and is a string data type. It represents the job of the card holder
17	dob	This variable is discrete in nature and is a date data type. It represents the date of birth of the card holder
18	trans_num	Unique Identifier This variable is discrete in nature and is a string data type. It represents the transaction id of the card holder's transactions
19	unix_time	This variable is discrete in nature and is a date/time data type. It represents the timestamp that counts the number of seconds that have elapsed since the Unix epoch (January 1, 1970, at 00:00:00 UTC)
20	merch_lat	This variable is discrete in nature and is a Geospatial data type. It represents the latitude of the merchant
21	merch_long	This variable is discrete in nature and is a Geospatial data type. It represents the longitude of the merchant
22	is_fraud	This variable is discrete in nature and is a boolean data type. 0 - False indicating that the transaction is not fraudulent, 1 - True indicating that the transaction is fraudulent

c. Link to the original data source.

https://www.kaggle.com/datasets/dermisfit/fraud-transactions-dataset?select=fraudTest.csv

5. Approach:

In examining the dataset, we plan to:

- Perform exploratory data analysis (EDA) to understand the distribution and characteristics of the data.
- Create an advanced machine learning model for fraud detection, utilizing attributes like transaction date and time, transaction amount, merchant details, and geographical information.
- Implement an alerting system within Tableau that provides real-time insights into fraud trends and incidents.
- Develop a user-friendly dashboard in Tableau that can be accessed by Chief Risk Officers (CROs), fraud analysts, data scientists, and other stakeholders.

- Use Tableau to generate detailed reports for fraud investigations, offering insights into transaction patterns.
- Ensure that the solution is adaptable, capable of addressing new types of fraud by continuously learning and adapting to emerging fraud patterns.
- Draw insights from the dataset to showcase how data analytics and machine learning can empower organizations to monitor and prevent fraudulent activities in real-time.

This project aims to provide a holistic solution for fraud detection and prevention by integrating advanced analytics, machine learning, and data visualization techniques while using the provided dataset. The findings will be presented through interactive Tableau dashboards and detailed reports that cater to the needs of the target audience, including CROs, fraud analysts, data scientists, compliance officers, and C-suite executives. The project's objective is to showcase the practical application of these technologies in safeguarding financial institutions against financial fraud.

Appendix

https://legacy.acfe.com/report-to-the-nations/2020/
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6798528/
https://www.tableau.com/solutions/finance-risk-analytics#reveal-172133
https://www.managementsolutions.com/en/microsites/whitepapers/financial-crime
https://www2.deloitte.com/content/dam/Deloitte/uy/Documents/strategy/gx-fsi-evolving-role-of-chief-data-officer.pdf
https://www.sas.com/en/whitepapers/managing-fraud-risk-digital-age-109846.html?utm_source=google&utm_medium=cpc&utm_campaign=industry-banking-fraud-us&utm_content=GMS-187726-fsi-pyf-us&dclid=&gclid=CjwKCAjw7c2pBhAZEiwA88pOFxcK_xxPPDb0CRTqS7sVWUN89ghpGKjUShBzIkaIIoGP1sDbk5l6vhoCWe0QAvD_BwE