

GenAI-Based Data Profiling for Regulatory Reporting



Ravindra Reddy Nandyala
Mahan Kaleswararao
Siva Kumar Bestha
Sai Teja Vavintaparthi
Prashanth

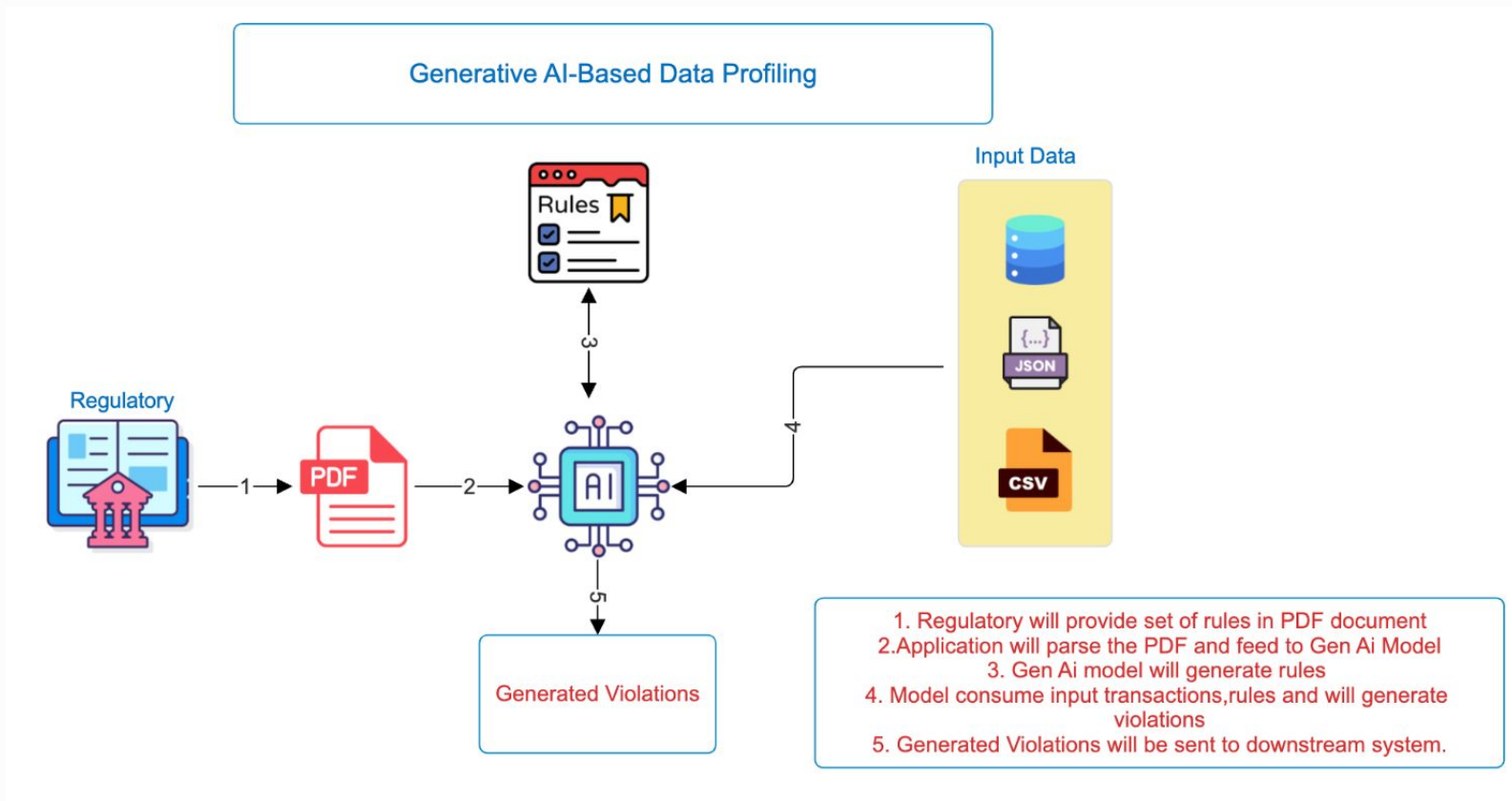
Problem Statement & Challenges

- Regulatory reporting requires accurate data profiling.
- Manual rule definition is time-consuming and error-prone.
- Existing methods lack adaptive risk scoring.
- Compliance teams need automated remediation suggestions.

Solution Overview

- Leverage's LLMs to extract data profiling rules from regulatory documents.
- Validates transactions against the extracted rules.
- Detects violations and provides actionable insights.

Architecture Diagram



Key Features

- Scalable and generic.
- Uses fuzzy logic for dynamic column mapping.
- Works with any rule set and dataset.
- No hard coding of rules or column names.

How It Works

- Extract text from regulatory documents (PDFs).
- Parse rules using LLMs and convert them into a structured format.
- Preprocess and pre-aggregate transaction data.
- Map rules to dataset columns using fuzzy matching.
- Validate transactions against the rules.
- Display violations with detailed insights.

Demo - Regulated PDF rules

Rules:

1. Customer ID length must be between 6 to 12 characters.
2. Transaction Amount must be a positive number. The range depends on the Account_Type:
 - If Account_Type is 'Savings', the amount must be between \$100 and \$1,500.
 - For other account types, the amount must be between \$100 and \$1,000,000.
3. Transaction_Date must be in YYYY-MM-DD format.
4. Field 'Capital_Adequacy_Ratio' must be above 8.
5. Field 'Account_Type' must be one of the following: Savings, Checking, or Loan.
6. Field 'Transaction_Type' must be one of the following: Deposit, Withdrawal, or Transfer.
7. Field Permanent Account Number must be 10 characters.
8. Total amount for a customer must not exceed \$1500.



Demo - Input Data

Customer_ID	Transaction_Amt	Transaction_Date	Capital_Adequacy_Ratio	Account_Type	Transaction_Type	Permanent_Account_Number	CURRENCY
12345678	600	2019-10-01	9.5	Savings	Deposit	aeptn4035k	INR
4321	200	2028-10-02	7.2	Checking1	Withdrawal	e235ew24sw	INR
9123456789	-300	2024/11/03	8.8	Loan	Transfer	2wq1234ew	INR
912345678901	900	2024-10-04	6.5	Savings	Deposit		INR
912345678901	700	2024-10-05	6.5	Savings	Deposit		INR
812345664371	1900	2024-11-04	9	Savings	Deposit		INR

Demo - Extracted Rules - 1

[6 rows x 10 columns]

Mapped Rules:

```
[
  {
    "type": "digit_length",
    "field": "Customer_ID",
    "min_value": 6,
    "max_value": 12
  },
  {
    "type": "min_value",
    "field": "Transaction_Amt",
    "value": 100
  },
  {
    "type": "conditional_range",
    "field": "Transaction_Amt",
    "conditions": [
      {
        "condition": "Account_Type == 'Savings'",
        "min_value": 100,
        "max_value": 1500
      },
      {
        "condition": "Account_Type != 'Savings'",
        "min_value": 100,
        "max_value": 1000000
      }
    ]
  }
],
```

```
{
  "type": "date_format",
  "field": "Transaction_Date",
  "format": "YYYY-MM-DD"
},
{
  "type": "min_value",
  "field": "Capital_Adequacy_Ratio",
  "value": 8
},
{
  "type": "allowed_values",
  "field": "Account_Type",
  "values": [
    "Savings",
    "Checking",
    "Loan"
  ]
},
```

Demo - Extracted Rules - 2

```
{
  "type": "allowed_values",
  "field": "Transaction_Type",
  "values": [
    "Deposit",
    "Withdrawal",
    "Transfer"
  ]
},
{
  "type": "digit_length",
  "field": "Permanent_Account_Number",
  "value": 10
},
{
  "type": "aggregated_amount",
  "field": "Total_Amount",
  "threshold": 1500
}
]
```

Demo - Violations Generated By Model

```
]
Violations Found:

Record ID: 1
Record Data: {'Customer_ID': 4321, 'Transaction_Amt': 200, 'Transaction_Date': '2028-10-02', 'Capital_Adequacy_Ratio': 7.2, 'Account_Type': 'Checking1', 'Transaction_Type': 'Withdrawal', 'Permanent_Account_Number': 'e235ew24sw', 'CURRENCY': 'INR', 'Total_Amount': 200, 'Transaction_Count': 1}
Violated Rules:
- Customer_ID length is less than the minimum required length of 6 digits.
- Account_Type value 'Checking1' is not allowed. Allowed values are 'Savings', 'Checking', 'Loan'.
- Transaction_Date is not in the required format 'YYYY-MM-DD'.
- Capital_Adequacy_Ratio is less than the minimum required value of 8.

Record ID: 2
Record Data: {'Customer_ID': 9123456789, 'Transaction_Amt': -300, 'Transaction_Date': '2024/11/03', 'Capital_Adequacy_Ratio': 8.8, 'Account_Type': 'Loan', 'Transaction_Type': 'Transfer', 'Permanent_Account_Number': '2wq1234ew', 'CURRENCY': 'INR', 'Total_Amount': -300, 'Transaction_Count': 1}
Violated Rules:
- Customer_ID length is more than the maximum allowed length of 12 digits.
- Transaction_Amt is less than the minimum required value of 100.
- Transaction_Date is not in the required format 'YYYY-MM-DD'.
- Permanent_Account_Number length is not equal to the required length of 10 digits.

Record ID: 3
Record Data: {'Customer_ID': 912345678901, 'Transaction_Amt': 900, 'Transaction_Date': '2024-10-04', 'Capital_Adequacy_Ratio': 6.5, 'Account_Type': 'Savings', 'Transaction_Type': 'Deposit', 'Permanent_Account_Number': nan, 'CURRENCY': 'INR', 'Total_Amount': 1600, 'Transaction_Count': 2}
Violated Rules:
- Customer_ID length is more than the maximum allowed length of 12 digits.
- Capital_Adequacy_Ratio is less than the minimum required value of 8.
- Permanent_Account_Number is missing.
- Total_Amount exceeds the threshold of 1500.
- Aggregated amount 1600 exceeds threshold 1500

Record ID: 4
Record Data: {'Customer_ID': 912345678901, 'Transaction_Amt': 700, 'Transaction_Date': '2024-10-05', 'Capital_Adequacy_Ratio': 6.5, 'Account_Type': 'Savings', 'Transaction_Type': 'Deposit', 'Permanent_Account_Number': nan, 'CURRENCY': 'INR', 'Total_Amount': 1600, 'Transaction_Count': 2}
Violated Rules:
- Customer_ID length is more than the maximum allowed length of 12 digits.
- Capital_Adequacy_Ratio is less than the minimum required value of 8.
- Permanent_Account_Number is missing.
- Total_Amount exceeds the threshold of 1500.
- Aggregated amount 1600 exceeds threshold 1500

Record ID: 5
Record Data: {'Customer_ID': 812345664371, 'Transaction_Amt': 1900, 'Transaction_Date': '2024-11-04', 'Capital_Adequacy_Ratio': 9.0, 'Account_Type': 'Savings', 'Transaction_Type': 'Deposit', 'Permanent_Account_Number': nan, 'CURRENCY': 'INR', 'Total_Amount': 1900, 'Transaction_Count': 1}
Violated Rules:
- Customer_ID length is more than the maximum allowed length of 12 digits.
- Permanent_Account_Number is missing.
- Total_Amount exceeds the threshold of 1500.
- Aggregated amount 1900 exceeds threshold 1500
(ravi_venv)
```

Conclusion

- The GenAI-powered regulatory compliance system provides a scalable, intelligent, and adaptable solution for automating compliance validation.
- It utilizes Gen AI models for rule extraction and fuzzy logic for dynamic column mapping to enhance accuracy and efficiency.
- The system's flexible architecture enables seamless extension to new use cases and ensuring future scalability.