Project Synopsis

**POLYSCAN**

**CDAC DBDA**

**Guided By:**

**Mr. Rahul Parte**

**Submitted By:**

**Group 6**



**Centre for Development of Advanced Computing (CDAC)** Post Graduate Diploma in Big Data Analytics (PG-DBDA) Mumbai Centre – USM VITA

**Academic Year: Feb – Aug 2025**

# Title:

**Polyscan**

# Team Members:

The project team consists of seven dedicated and skilled members:

* Utkarsh Ojha – Team Leader
* Saiteja Gadigoppula– Team Member
* Prathamesh Chaudhari – Team Member
* Nandini Yadav – Team Member
* Yogeshree Chandankhede – Team Member
* Fenil Patel – Team Member
* Anuja Sonawne – Team Member

# Problem Statement:

The U.S. political finance system is complex and often lacks transparency, despite the availability of public data from OpenFEC. The vast and granular nature of this data presents several challenges, including:

* Uncovering donation trends and patterns.
* Detecting potential illicit financing, such as "straw donors" or fraud.
* Assessing donor or industry influence on elections.
* Monitoring changes in donation behavior over time.
* Supporting compliance with campaign finance laws.

# Introduction:

Political donations are an important part of democracy. They help candidates run their campaigns and share their ideas with voters. To keep the system fair and trustworthy, it's important that the flow of money is open and transparent.

In the U.S., the Federal Election Commission (FEC) collects and shares detailed data about who donates to political campaigns. While this data is publicly available, it’s often large, complex, and hard to understand. As a result, it becomes difficult to spot important patterns, trends, or possible signs of illegal or suspicious activity.

# Objective:

The primary objectives of this project are:

* Clean and organize the OpenFEC individual contributions data to make it ready for analysis.
* Find and group important features (like amount, donor info, etc.) that affect donation patterns.
* Spot unusual or suspicious donations using clear, rule-based checks (not machine learning).
* Discover useful insights about who donates, where they come from, and their economic background.
* Show complex donation trends and odd patterns through simple and clear visualizations.
* Build a tool or system that helps analysts, campaign teams, and government bodies better understand and use political donation data.

# Data Dictionary & Description:

This project will primarily utilize the OpenFEC Individual Contributions dataset. This dataset is publicly available from the Federal Election Commission (FEC) and provides detailed information on contributions made by individuals to federal candidates, parties, and political action committees (PACs).

https://[www.kaggle.com/datasets/mukeshdevrath007/indian-5000-cities-weather-data/data](http://www.kaggle.com/datasets/mukeshdevrath007/indian-5000-cities-weather-data/data)

Main dataset:

<https://www.fec.gov/campaign-finance-data/contributions-individuals-file-description/>

<https://www.fec.gov/campaign-finance-data/committee-master-file-description/>

<https://www.fec.gov/campaign-finance-data/candidate-master-file-description/>

**Individuals → Committees: Contributions are made to committees (CMTE\_ID).**

**Committees → Candidates: Committees are linked to candidates (CAND\_ID or CAND\_PCC).**

**Candidates → Elections: Candidate data includes election year, office, and party.**

**Columns Details:**

Column Name Description

CMTE\_ID Committee receiving the contribution

NAME Contributor’s full name

CITY, STATE, ZIP\_CODE Contributor’s location

EMPLOYER, OCCUPATION Contributor’s job details

TRANSACTION\_DT Date of contribution

TRANSACTION\_AMT Amount donated

TRANSACTION\_TP Type of transaction (e.g., direct donation, earmarked)

ENTITY\_TP Entity type (IND = Individual, CAN = Candidate, etc.)

OTHER\_ID FEC ID of contributor if not an individual

SUB\_ID Unique identifier for the contribution

**Data Size:** The OpenFEC individual contributions dataset is substantial, often containing millions of records annually. This volume necessitates efficient data handling and processing techniques.

**Data Characteristics:**

* Granular: Provides highly detailed information at the individual contribution level.
* Categorical and Numerical: Contains a mix of textual (names, occupations), geographical (addresses), temporal (dates), and numerical (amounts) data.
* Longitudinal: Available for multiple election cycles, allowing for time-series analysis and trend identification.
* Publicly Available: Ensures transparency and verifiability of findings.

# Architecture Diagram:

# 

**ACL**

# Expected Outcomes:

* A deployable dashboard for public or organizational use.
* A clean and well-organized dataset ready for analysis.
* Clear understanding of what factors influence political donations.
* Detection of unusual or suspicious donation activities.
* Insights into who is donating — based on location, job, and gender.
* Easy-to-read charts and visuals showing donation trends and red flags.
* A useful tool or system that helps others (like analysts or officials) explore and understand political donation data.

# Use Cases:

* Political Analysts & Researchers: Understand electoral dynamics, identify campaign finance trends, and support academic studies.
* Campaign Managers & Fundraisers: Optimize fundraising, allocate resources effectively, and analyze competitor strategies.
* Regulatory Bodies (e.g., FEC): Enhance compliance monitoring, fraud detection, and inform policy formulation.
* Investigative Journalists: Uncover influence and expose financial irregularities.
* Public & Advocacy Groups: Promote transparency and advocate for campaign finance reform.

| **KPI** | **Description** |
| --- | --- |
| **Total Contributions** | Sum of all donations over a specific period. |
| **Average Donation Size** | Mean contribution amount. |
| **Donor Retention Rate** | Percentage of repeat donors. |
| **Contribution Frequency** | Average number of donations per donor. |
| **Refund Rate** | Percentage of donations refunded. |
| **Earmarked Contribution Ratio** | Proportion of donations earmarked for specific purposes. |
| **Regional Contribution Distribution** | Analysis of donations by geographic location. |
| **Donor Demographics** | Breakdown of donors by occupation, employer, and gender. |

# Conclusion:

This project helps turn complex political donation data into clear and useful information. It allows people like analysts, campaign teams, and officials to see patterns, find unusual donations, and better understand who is funding politics.

By making this information easier to explore, the project supports fairness, transparency, and trust in the election process. It gives the tools needed to keep political financing open and accountable — helping strengthen democracy in the U.S.