

# Telecom Fraud Visualization Dashboard - Documentation

The Telecom Fraud Visualization Dashboard is a real-time data visualization tool built using Streamlit, Python, and the Reddit API. This documentation provides an overview of the project, including its purpose, features, architecture, setup instructions, and usage guide.

## Table of Contents

Overview

Features

Architecture

Setup

Usage

Contributing

License

## Overview

The Telecom Fraud Visualization Dashboard aims to analyze streaming data from Reddit to identify fraud in the telecommunications industry. It connects to the Reddit API to collect real-time posts related to telecom fraud, processes the data to extract useful information, performs analysis to identify patterns and trends, and presents the results in an interactive dashboard using Streamlit.

## Features

- Real-time data collection: The dashboard connects to the Reddit API to collect real-time posts related to telecom fraud and other keywords.
- Data processing: The collected posts are processed to extract important information such as post text, user name, subreddit, and date/time.
- Analysis and visualization: The data is analyzed to identify patterns and trends related to telecom fraud. The dashboard includes interactive visualizations,

such as bar charts and line charts, to provide insights into fraud mentions by subreddit and over time.

- User-friendly interface: The Streamlit-based dashboard provides a user-friendly interface with interactive controls for exploring the data and gaining insights.

## Architecture

The Telecom Fraud Visualization Dashboard is built using the following technologies:

- Python: The programming language used to write the data collection, processing, and analysis scripts.
- Streamlit: The framework used to create the interactive data visualization dashboard.
- PRAW (Python Reddit API Wrapper): The library used to connect to the Reddit API and collect the real-time posts.
- Pandas: The library used for data manipulation and processing.
- Matplotlib and Seaborn: The libraries used for data visualization.

The architecture of the project consists of the following components:

1. Data Collection: The Python script connects to the Reddit API using PRAW and collects real-time posts related to telecom fraud and other keywords.
2. Data Processing: The collected posts are processed using Pandas to extract important information and store it in a structured format.
3. Data Analysis and Visualization: The processed data is analyzed using Pandas and NumPy for statistical analysis and Matplotlib or Seaborn for creating visualizations. The visualizations are displayed in an interactive dashboard using Streamlit.

## Setup

To set up and run the Telecom Fraud Visualization Dashboard, follow these steps:

1. Clone the project repository:
2. Install the required Python packages:
3. Update the `client_id`, `client_secret`, and `user_agent` variables in the code with your Reddit API credentials.
4. Run the Streamlit app:
5. Access the dashboard by opening the provided local URL in your web browser.

## Usage

- The Telecom Fraud Visualization Dashboard displays real-time information about telecom fraud mentions from Reddit.
- The dashboard includes a bar chart showing the number of fraud mentions by subreddit and a line chart showing the frequency of fraud mentions over time.
- Use the interactive controls provided by Streamlit to explore the data and gain insights into fraud patterns and trends.