

Data Collection and Preprocessing Phase

Date	18 June 2024
Team ID	739990
Project Title	Auto Insurance Fraud Detection
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification Report:

Elevate your data strategy with the Data Collection plan and the Raw Data Sources report, ensuring meticulous data curation and integrity for informed decision-making in every analysis and decision-making endeavor.

Data Collection Plan:

Section	Description
Project Overview	Auto insurance fraud detection involves using advanced technologies and methodologies to identify and prevent fraudulent activities in insurance claims. This includes comprehensive data collection from diverse sources, predictive modeling, machine learning algorithms, and rule-based systems to detect patterns and anomalies indicative of fraud.
Data Collection Plan	<ol style="list-style-type: none"> 1. Claims History: <ul style="list-style-type: none"> ○ Collect detailed historical claims data, including claim amounts, dates, and types. 2. Customer Information: <ul style="list-style-type: none"> ○ Gather personal and demographic details of policyholders, such as age, address, occupation, and driving history. 3. Vehicle Details: <ul style="list-style-type: none"> ○ Acquire comprehensive information about insured vehicles, including make, model, year, and maintenance records. 4. Telematics Data: <ul style="list-style-type: none"> ○ Integrate data from telematics devices, such as GPS location, driving behavior, and vehicle usage patterns.

5. **External Databases:**

- Access external sources like criminal records, vehicle history reports, and social media profiles.

6. **Repair and Medical Records:**

- Collect records from repair shops and medical providers involved in claims.

7. **Industry Collaboration:**

- Share data with other insurance companies and industry bodies to enhance detection efforts.

8. **Public Records:**

- Utilize public databases, including property records and legal documents, for additional verification.

9. **Incident Reports:**

- Obtain police reports and other incident documentation related to claims.

10. **Customer Interactions:**

- Record and analyze customer communications, including phone calls, emails, and chat logs.

11. **Financial Transactions:**

- Monitor payment records and transaction histories associated with claims.

12. **Document Verification:**

- Validate submitted documents, such as driver's licenses and vehicle registration, for authenticity.

Raw Data Sources Report:

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle Dataset	This dataset typically includes comprehensive information such as claim IDs, policy details, incident dates, types of incidents (like collisions or theft), claim amounts, and indicators of whether claims were found to be fraudulent or not.	C:\Users\narra\Downloads\insurance claims (1).csv	CSV	15 kB	Public