



**INSTITUTE OF TECHNOLOGY,  
NIRMA UNIVERSITY**

**2CSDE71 (Data Mining)**

**PRACTICAL 1**

**Name : Naman Thaker**

**Roll No. : 20BCE529**

**Semester : 6<sup>th</sup>**

**Aim: Data Domain selection and Identification of Characteristics of selected Dataset of different Formats. Prepare a report(writeup) with following detail**

- **Selection of data domain : Insurance Fraud**
- **Define the data domain : Automobile Insurance fraud detection.**
- **The data source : Data Trained**
- **Objective : The main objective of this dataset is to find out the fraud that is going out in automobile for Insurance company.**
- **Define the selection of fields :**

months\_as\_customer', 'age', 'policy\_number', 'policy\_bind\_date', 'policy\_state', 'policy\_csl', 'policy\_deductable', 'policy\_annual\_premium', 'umbrella\_limit', 'insured\_zip', 'insured\_sex', 'insured\_education\_level', 'insured\_occupation', 'insured\_hobbies', 'insured\_relationship', 'capital-gains', 'capital-loss', 'incident\_date', 'incident\_type', 'collision\_type', 'incident\_severity', 'authorities\_contacted', 'incident\_state', 'incident\_city', 'incident\_location', 'incident\_hour\_of\_the\_day', 'number\_of\_vehicles\_involved', 'property\_damage', 'bodily\_injuries', 'witnesses', 'police\_report\_available', 'total\_claim\_amount', 'injury\_claim', 'property\_claim', 'vehicle\_claim', 'auto\_make', 'auto\_model', 'auto\_year', 'fraud\_reported'

- **Characteristic and behaviours (distribution and inference) of data for each selected field**

|                         |                            |
|-------------------------|----------------------------|
| months_as_customer      | int64 - numerical          |
| age                     | int64 - numerical          |
| policy_number           | int64 - numerical          |
| policy_bind_date        | object - date              |
| policy_state            | object - ordinal           |
| policy_csl              | object - continuous        |
| policy_deductable       | int64 - discrete           |
| policy_annual_premium   | float64 - continuous       |
| umbrella_limit          | int64 - discrete numerical |
| insured_zip             | int64 - discrete numerical |
| insured_sex             | object - ordinal           |
| insured_education_level | object - nominal           |
| insured_occupation      | object - nominal           |
| insured_hobbies         | object - nominal           |
| insured_relationship    | object - ordinal           |
| capital-gains           | int64 - numerical          |
| capital-loss            | int64 - numerical          |
| incident_date           | object - date              |
| incident_type           | object - ordinal           |
| collision_type          | object - ordinal           |
| incident_severity       | object - ordinal           |

|                             |                           |
|-----------------------------|---------------------------|
| authorities_contacted       | object - ordinal          |
| incident_state              | object - nominal          |
| incident_city               | object - nominal          |
| incident_location           | object - address          |
| incident_hour_of_the_day    | int64 - discrete          |
| number_of_vehicles_involved | int64 - discrete          |
| property_damage             | object - Symmetric Binary |
| bodily_injuries             | int64 - discrete          |
| witnesses                   | int64 - discrete          |
| police_report_available     | object - Symmetric Binary |
| total_claim_amount          | int64 - discrete          |
| injury_claim                | int64 - discrete          |
| property_claim              | int64 - discrete          |
| vehicle_claim               | int64 - discrete          |
| auto_make                   | object - nominal          |
| auto_model                  | object - nominal          |
| auto_year                   | int64 - discrete          |
| fraud_reported              | object - Symmetric Binary |

- **Conclusion and Observation :**

Thus we conclude that for detecting fraud activity we need to have lot of features to process on and then after we can predict that Insurance company has to claim the insurance of automobile or not.