Abstract:
A large number of problems in data mining are related to fraud detection. Fraud is a common problem in auto insurance claims, health insurance claims, credit card transactions, financial transaction and so on. The data in this particular case comes from an actual auto insurance company. Each record represents an insurance claim. The last column in the table tells you whether the claim was fraudulent or not. A number of people have used this dataset and here are some observations from them:
• "This is an interesting data because the rules that most tools are coming up with do not make any intuitive sense. I think a lot of the tools are overfitting the data set."
• "The other systems are producing low error rates but the rules generated make no sense."
• "It is OK to have a higher overall error rate with simple human understandable rules for a business use case like this."

There are two datasets (Excel Files) –
1. Insurance Fraud – TRAIN-3000, and
2. Insurance Fraud – TEST-12900.

Attribute Information:
Input variables:
1. **MONTH**: Jan through Dec.
2. **WEEKOFMONTH**: Continuous – 1 through 5.
3. **DAYOFWEEK**: Monday through Sunday.
4. MAKE: Acura, BMW, Chevrolet, Dodge, Ford, Toyota, VW, Nissan, etc.
5. **ACCIDENTAREA**: Urban, Rural.
6. **DAYOFWEEKCLAIMED**: Monday through Friday.
7. **MONTHCLAIMED**: Jan through Dec.
8. **WEEKOFMONTHCLAIMED**: Continuous – 1 through 5.
9. **SEX**: Male/Female.
10. **MARITALSTATUS**: Married, Single, Divorced, Widow.
11. **AGE**: continuous – 0 through 80.
12. **FAULT**: Policy_Holder, Third_Party.
13. **POLICYTYPE**: Sport-Collision, Sedan-All_Perils, Sedan-Collision, Sedan-Liability etc.
14. **VEHICLECATEGORY**: Sport, Sedan, Utility, etc.
15. **VEHICLEPRICE**: 20000_to_29000,30000_to_39000, 40000_to_59000 etc.
16. **REPNUMBER**: Continuous – 1 through 16
17. **DEDUCTIBLE**: Continuous – 300 through 700.
18. **DRIVERRATING**: Continuous – 1 through 4.

19. **DAYS_POLICY_ACCIDENT**: none, 1_to_7, 8_to_15, 15_to_30, more_than_30, etc.
20. **DAYS_POLICY_CLAIM**: none, 1_to_7, 8_to_15, 15_to_30, more_than_30, etc.
21. **PASTNUMBEROFCLAIMS**: none, 2_to_4, more_than_4, etc.
22. **AGEOFVEHICLE**: new, 3_years, 4_years, 5_years, 6_years, 7_years, more_than_7, etc.
23. **AGEOFPOLICYHOLDER**: 16_to_17, 21_to_25, 31_to_35, etc.
24. **POLICEREPORTFILED**: Yes/No.
25. **WITNESSPRESENT**: Yes/No.
26. **AGENTTYPE**: Internal/External.
27. **NUMBEROFSUPPLIMENTS**: none, 1_to_2, 3_to_5, more_than_5.
28. **ADDRESSCHANGE_CLAIM**: no_change, under_6_months, 1_year, 2_to_3_years, etc.
29. **NUMBEROFCARS**: 1_vehicle, 2_vehicles, 3_to_4 etc.
30. **YEAR**: Continuous – 1994 through 1996.
31. **BASEPOLICY**: Collison, All_Perils, Liability etc.

Output variable (desired target):
32. **FRAUDFOUND**: Yes/No.