Motor Insurance Fraud Detection System

Can a Machine be taught to see through the deception and catch them in the act?

Rohit Devanaboina - 1920181

1. The Facts

1 in 10 Americans would commit fraud if they could get away with it [1]

1 in 4 Americans say it is ok to commit insurance fraud [1]

5.9%

₹ 7,62,000 cr in Total Premium [2]

₹ 45,000 cr. lost to Fraud [3]

The Loss is borne by the Public
Slows down adoption of Insurance

2. Objectives

Create a Machine Learning Model that is able to **partially automate** the Motor Insurance Fraud Detection Process

Two Critical Success Factors -

High Detection Rate - Fraudulent Claims

- i.e. **High Sensitivity**

Moderate Dectection Rate - Legit Claims

- i.e. **Moderate Specificity**

3. Methodology

The **CRISP-DM** Framework for Data Mining has been employed

I. **Business Understanding** - Identifying Fraud **w/ 100% certainty** = Difficult
Identifying **Suspicious Claims** = Relatively Easy

II. **Data Understanding** - **Labeled Insurance Claims** - 52% Fraudulent & 48% Legit
Included **28 features,** each providing info about the claims

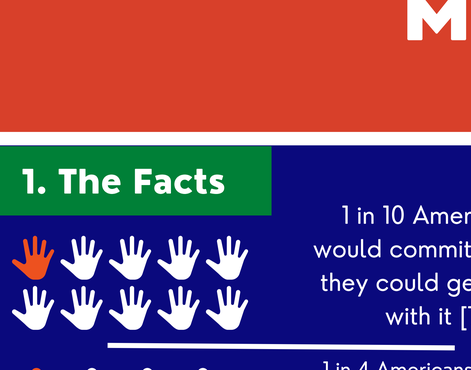
III. **Data Preparation** - **Minimal Preparation** required - the Dataset is well balanced
Uploaded data into RapidMiner and specified data type

IV. **Modeling** - Conducted in Two Phases (see next section)

V. **Evaluation** - Provided at the end

4.1 Modeling - Phase 1: Exploration

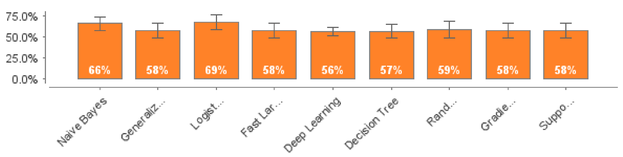
Step 1
Test Models from Reference Article



The 4 Models used in the Reference Article

Step 2
Run Automated Test of Models

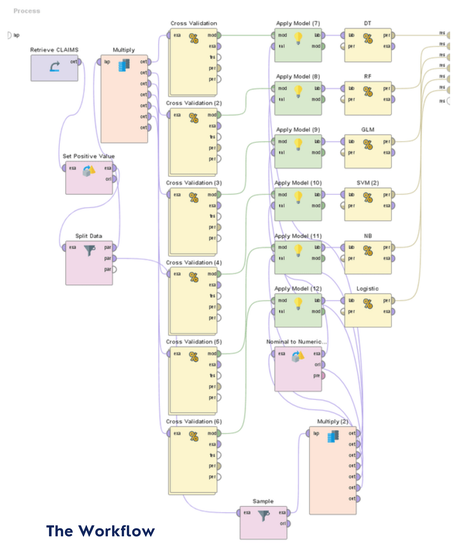
[Above - Sensitivity Rates | Below - Specificity Rates]



The top performers in an Automated Test of 9 Algorithms

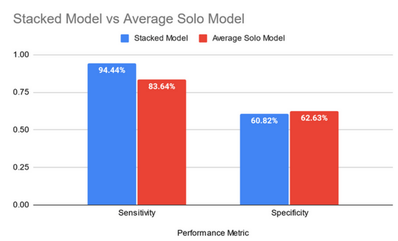
4.2 Modeling - Phase 2 : Optimization

Step 3 - Test Top Models in a Realistic Environment

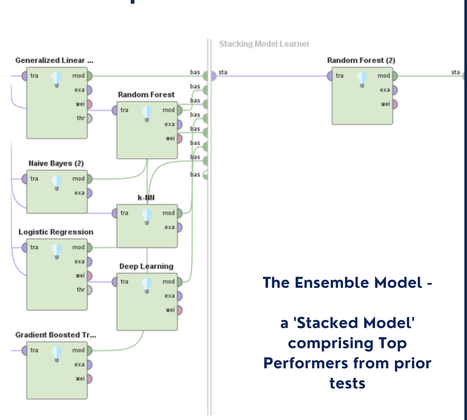


The Workflow

The Results



Step 4- Creating an Ensemble Model of Top Performers



The Ensemble Model -

a 'Stacked Model' comprising Top Performers from prior tests

5. Evaluation (AKA Results!)

The best performing model was **The Stacked Model**

- An **ensemble** of some of the **Best Performing Models** proved to perform the best - **94%** Fraud Detection Rate & elimination of **60%** of claims before human processing takes place
- **Two heads are better than one** - by combining **7 different "heads"** into a single model, we've increased fraud detection by **over 10%!** (compared to individual models)

6. Value Addition

For a Motor Insurance firm that employs this model, the system will -

- Automatically flag 94% of all Motor Fraud Claims
- Reduce claims processing workload by 60% (by eliminating them from the human review process)

Literature Cited and Acknowledgements

RapidMiner and RapidMiner Academy Tutorials

[1] A Two-Step Process for Detecting Fraud using Oracle Machine Learning by Oracle ML

[2] IRDA - Indian Insurance Market

[3] Insurance Frauds Control Act; an urgent need in India - BusinessToday

Key Takeaways

- To make a model on paper is easy with **tools like RapidMiner**
  However, to make a **good model**, learning **Data Mining Principles** is key
- Learning **how to evaluate a model** is the first step in building a good Model
- Learning **how to test a model** is the second step

My early models seemed to perform great, then failed when tested & evaluated properly