# Case Study: Subrogation of Insurance Claim

GROUP 13

GAURAV SITAULA, MD REZWAN ISLAM, ALI SHAKIBA

The main purpose of this case study is to identify cases with higher chances of subrogation for an insurance company using text mining analysis. We have around 3000 observations of the insurance data which contains around 35% of the instances when the insurance company sued the third party to recover the compensation. For the insurance company, one of the major challenges is that they need to be sure before subrogation because these litigation processes involve money and time. This analysis is beneficial in two ways. On one hand, if we can extract information from 'adjuster note' and thereby find those cases where our change of wining is much higher, the insurance company might earn substantial money from subrogation. On the other hand, the insurance company can decrease the time, money, man power and effort for this subrogation process which ultimately reduce the cost.

At the beginning, we have explored the data set to detect some certain patterns or clues. The interesting observations are: a) If motor vehicle accident was the cause of injury and neck/head was injured, then the historical chance of subrogation was almost 90% b) If foreign body/object caused the injury, the probability of subrogation was minimal c) If eye/face/hand were injured and the injury was caused by the claimant himself or any object, then the probability of subrogation was very low as well. In the next phase, we have performed the text mining analysis from the corpus of texts along with the typical variables to identify whether the new information extracted from the "adjusted notes" improved our prediction accuracy of not. After analyzing all the models, we have decided the regression model as our final predictive model. Besides these models, we have built models apart from the text mining as well. For this case, we have chosen ROC Index to figure out the best case.
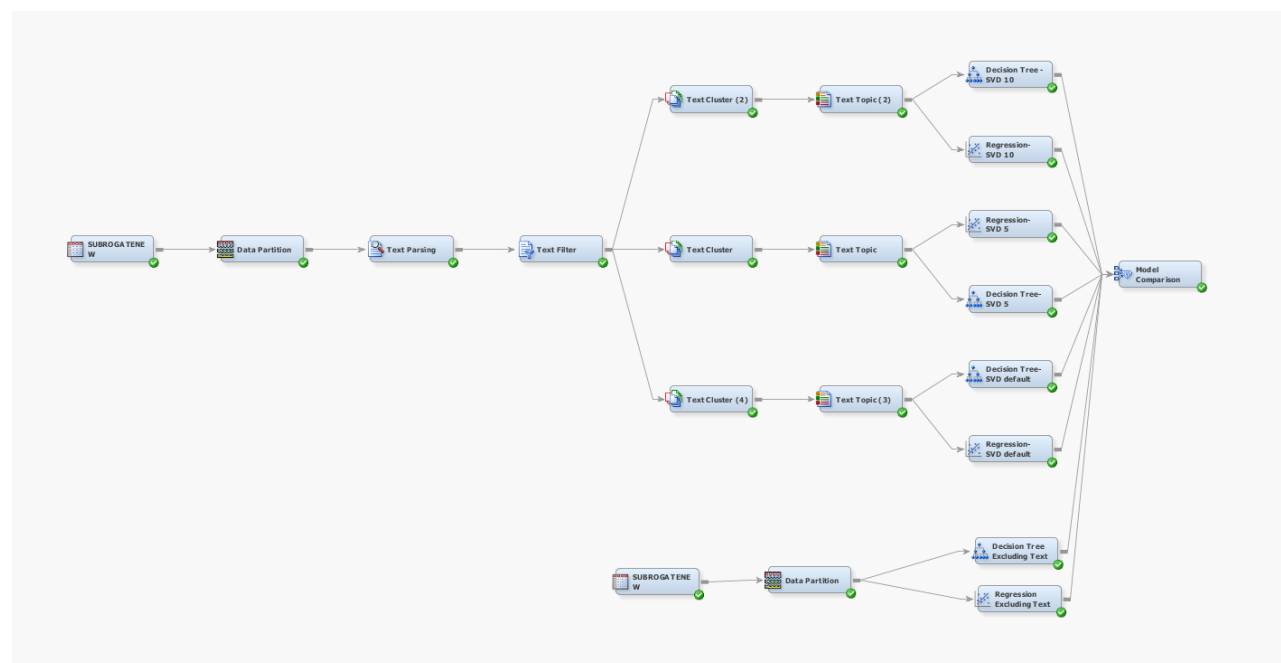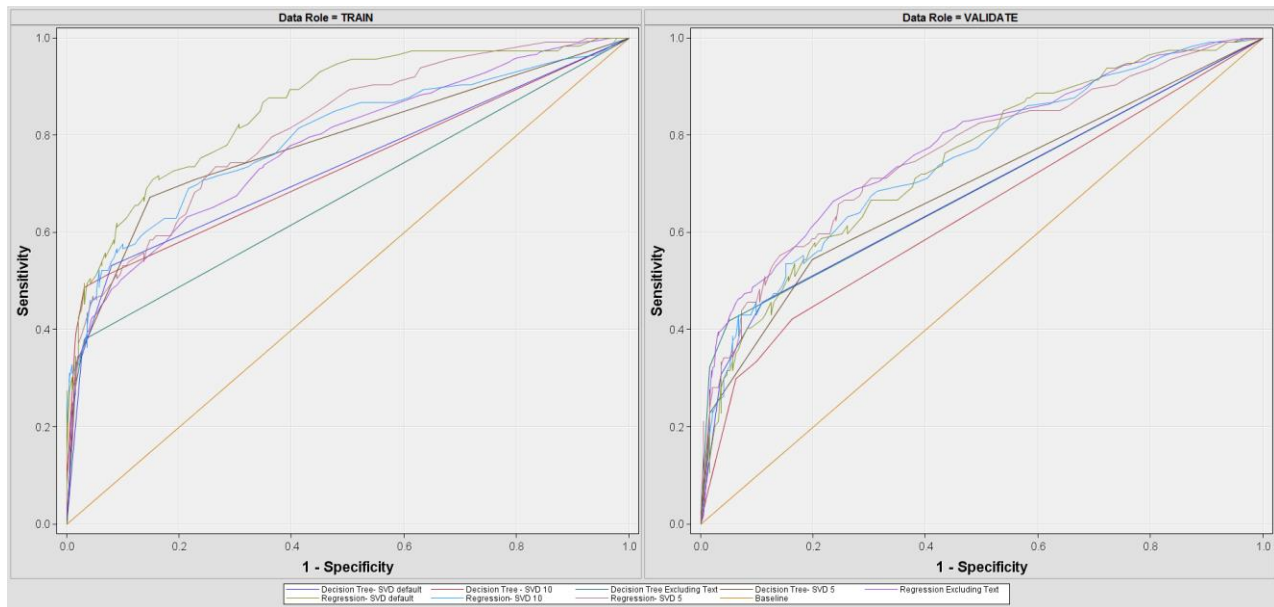
From the text mining analysis, we have figured out three list of words which have obvious trend. Here are the three lists, the first two are related to subrogation and the last one is not related to subrogation:

List 01: Automobile Accident, Auto, Accident, Car Accident, Motor

List 02: Injury, Fall, Neck Pain, Strained Neck, Shoulder

List 03: eye, finger, thumb, hand, ankle, foot, walk, bite, slip, twist etc

Apart from this list, we have found the major causes that have higher odds of subrogation in comparison to unusual body movement. These causes are: MVA, Occupational Illness, Fire, Exposure, Human Conflict, Electrocution. In addition to that, the causes that have low odd ratio in comparison to base case are: Animal Attack, Environmental Cause, Lifting, Pushing/Pulling, Foreign Body.

In conclusion, our analysis tells us that incorporating the text mining ultimately improves the predictive models. For our analysis, we have three specific recommendation:

a) **Subrogate:** If the claim adjuster notes contain anything related to Motor/Auto Vehicle Accident
b) **Subrogate:** If the claim adjuster notes contain anything related to Neck/Shoulder Injury
c) **Do Not Subrogate:** If the claimant causes the damage himself, an object or an animal