**Macquarie University Hackathon for Navigator Group**

*Introduction*

The intent of the hackathon is to systematically test machine learning and artificial intelligence models to predict *“gross incurred cost”* of workers compensation claims based on psychosocial and demographic data. The end use of these models will be to inform strategies to better manage psychosocial factors that influence these insurance claims.

*Data*

The data set consists of 71 columns of data from 552-cases. Each case consists of onboarding data of psychosocial screening and the claims cost data. The below Table outlines each of the variables and their relevance. The goal is to predict BO-BS, Y, Z, AA-AC.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***Column*** | ***Variable Name*** | ***Data Type*** | ***Description*** | ***Relevance***  *(IV= Independent Variable*  *DV= Dependent Variable)* |
| A | Gender | Qualitative  (Categorical) | 1 = male  2 = female | May be a significant IV. |
| B | State | Qualitative  (Categorical) | State in which claim occurred.  Values: 1-8 | This determines the legislative environment. There are many missing cells. |
| C | Escalated Navigation | Qualitative  (Categorical) | The operative on this case decided to contact the injured person more frequently.  Values: 0,1 | Is an indicator of psychosocial risk. |
| D | Intervention Required | Qualitative  (Categorical) | Indicates a referral to psychological therapy.  ("Yes"= 1, "No"= 2) | Is an indicator of psychosocial risk. |
| E | Claim\_Segment\_Name | Qualitative  (Categorical) | Is an insurance business group specifier.  ("Supported"=1, "Rapid"=2, "Specialised"=3) | May be a severity indicator as different insurance teams handle different claims. |
| F | Date\_of\_Accident | Date | Is the data of occurrence of injury | Relevant for calculating duration of claim. |
| G | Claim\_­How\_­Reported\_Name | Qualitative  (Categorical) | Whether the injured person or employer reported injury.  ("Email"=1, "INI"= 2) | May be a severity indicator. |
| H | Claim­\_Finalised­\_Date | Date | Is the final action on a claim | Is useful to calculate claims duration |
| I | Days\_Open | Quantitative  (Numerical) | Is then number of days the claim ran | Is the duration of the claim |
| J | RTW\_Plan\_Status\_Name | Qualitative  (Categorical) | Is if there is a return-to-work plan in place and the goal of that plan either same duties same job or different duties.  Values: 1-7 | May be a severity indicator. |
| K | Policy\_Issue\_State\_Name | Qualitative  (Categorical) | Is an interaction between state and the insurance policy.  Values: 1-4 | May codify the legislative environment and be a significant variable |
| L | Accident\_type\_code | Qualitative  (Categorical) | Is a coding of how injury occurred | May be useful to predict costs as it infers the cause of injury. |
| M | Nature\_of\_loss\_code | Qualitative  (Categorical) | Is what type of cost was incurred by the insurer | May be useful to identify cost. But is also likely subset of the Independent Variable |
| N | Bodily­\_location\_code | Qualitative  (Categorical) | Is what part of the body was injured | May be useful to identify costs. |
| O | Med\_Cert\_Capacity | Qualitative  (Categorical) | Is the initial work capacity certification as unfit, suitable duties, or fit for work.  Values: 1-6 | May be useful as a severity indicator. |
| P | Med\_cert\_unfit\_restricted\_weekdays. | Quantitative  (Numerical) | Is the number of days a person had reduced capacity for work | May be useful as a severity indicator. |
| Q | Wage\_amount | Quantitative  (Numerical- Continuous) | Is the amount of money paid weekly | May be useful in inferring work laboriousness which may be a factor in recovery |
| R | Interstate\_workers\_flag | Qualitative  (Categorical) | Is if a person lives in a different state from the state in which the injury occurred.  ("Yes"=1, "No"=2) | May be a useful mediating intervention |
| S | Payment\_early\_intervention\_rehab | Quantitative  (Numerical- Continuous) | Is money paid prior to claim being accepted | May be useful indicator of severity. |
| T | Payment­\_medicolegal | Quantitative  (Numerical- Continuous) | Is money paid for a review by a specialist doctor for an opinion | May be an indicator of claim complexity. **Is a subset of total cost** |
| U | Payment\_Rehab | Quantitative  (Numerical- Continuous) | Is the money paid to rehabilitation providers to assist in return to work | May be an indicator of claim severity and complexity |
| V | Payment\_travel\_accomodation | Quantitative  (Numerical- Continuous) | Is money paid for the injured person to travel and stay | May be a variable to partial out of costs |
| W | Payment\_weekly\_compensation | Quantitative  (Numerical- Continuous) | Is money paid because the injured person cannot work | May be a severity indicator Is part of total costs. |
| X | Hours\_worked\_per\_week | Quantitative  (Numerical- Continuous) | Is the number of hours of work per week | Is a covariate to wages cost. |
| Y | Fitness\_week2 | Qualitative  (Categorical) | Is fitness for work at week 2.  Values: 1-3 | Is a variable to be predicted (DV) |
| Z | Fitness\_week6 | Qualitative  (Categorical) | Is fitness for work at week 6.  Values: 1-3 | Is a variable to be predicted (DV) |
| AA | Fitness\_week12 | Qualitative  (Categorical) | Is fitness for work at week 12.  Values: 1-3 | Is a variable to be predicted (DV) |
| AB | Fitness\_week26 | Qualitative  (Categorical) | Is fitness for work at week 26.  Values: 1-3 | Is a variable to be predicted (DV) |
| AC | Fitness\_week52 | Qualitative  (Categorical) | Is fitness for work 1-year after injury.  Values: 1-3 | Is a variable to be predicted (DV) |
| AD | Work Status at Referral | Qualitative  (Categorical) | Is the work status at the time the injured person was reviewed by Navigator.  Values: 1-3 | Is a recovery indicator |
| AE | Priority | Qualitative  (Categorical) | Is the urgency of review of the case.  Values: 1-3 | Is a severity indicator |
| AF | Navigation Status | Qualitative  (Categorical) | Is the need for support identified by the Navigator.  Values: 1-5 | Is a severity indicator |
| AG | All calls are recorded, can we continue? | Qualitative  (Categorical) | Permission to record call.  Values: 0,1 | May indicate degree of trust |
| AH | Are you getting better? | Qualitative  (Categorical) | Self-perception of recovery.  Values: 0,1 | May be a useful indicator of optimism and attitude to injury |
| AI | How do you find your doctor? | Qualitative  (Categorical) | Relationship to doctor and perception of trust in treatment.  Values: 1-3 | May be a useful indicator of optimism about recovery |
| AJ | How do you find your case manager | Qualitative  (Categorical) | Relationship to insurer.  Values: 1-3 | May be a useful indicator of trust and attitude to claim |
| AK | Spoken to any of your co-workers | Qualitative  (Categorical) | Relationship with employer.  Values: 0,1 | May be a useful indicator of trust and attitude to work |
| AL | Are you seeing a therapist? | Qualitative  (Categorical) | Indicates psychological treatment.  (“Yes”=1, “No”=0) | May be a useful predictor |
| AM | Is therapy helpful for you? | Qualitative  (Categorical) | Indicates perception of treatment as helpful.  (“Yes”=1, “No”=0) | May be a useful predictor. |
| AN | How long have you had your pain problem? | Qualitative  (Categorical) | Indicator of duration of pain.  Values: 0-10 | May be a useful predictor. |
| AO | Rate the pain that you have had? | Qualitative  (Categorical) | Indicator of self-perception of pain severity.  Values: 0-10 | May be a useful predictor. |
| AP | I can do light work for 1-hour | Qualitative  (Categorical) | Indicator of self-efficacy.  Values: 0-10 | May be a useful predictor |
| AQ | I can sleep at night? | Qualitative  (Categorical) | Indicator of goodness of sleep.  Values: 0-10 | May be a useful predictor |
| AR | How tense or anxious have you felt? | Qualitative  (Categorical) | Indication of presence of anxiety.  Values: 0-10 | May be a useful predictor |
| AS | Have been bothered by feeling depressed? | Qualitative  (Categorical) | Indication of presence of low mood.  Values: 0-10 | May be a useful predictor |
| AT | Risk current pain may become persistent? | Qualitative  (Categorical) | Indication of beliefs about recovery.  Values: 0-10 | May be a useful predictor |
| AU | Chance you’ll be working normal duties? | Qualitative  (Categorical) | Indication of beliefs about recovery.  Values: 0-10 | May be a useful predictor |
| AV | Increase in pain indicates I should stop? | Qualitative  (Categorical) | Indications about the meaning of pain.  Values: 0-10 | May be a useful predictor |
| AW | I should not do my normal work with pain | Qualitative  (Categorical) | Indication about the meaning of pain.  Values: 0-10 | May be a useful predictor |
| AX | Unable to control the important things? | Qualitative  (Categorical) | Beliefs about self-efficacy.  Values: 0-4 | May be a useful predictor |
| AY | Your ability to handle personal Problems | Qualitative  (Categorical) | Beliefs about self-efficacy.  Values: 0-4 | May be a useful predictor |
| AZ | You felt that things were going your way | Qualitative  (Categorical) | Beliefs about self-efficacy.  Values: 0-4 | May be a useful predictor |
| BA | Felt Difficulties were piling up high? | Qualitative  (Categorical) | Beliefs about self-efficacy.  Values: 0-4 | May be a useful predictor |
| BB | How are you going financially? | Qualitative  (Categorical) | Indicator of financial duress.  Values: 0-10 | May be a useful predictor |
| BC | Perceived Stress Scale Short Form Total | Quantitative  (Numerical- Continuous) | Total score | May be a useful predictor |
| BD | Z-score | Quantitative  (Numerical- Continuous) | Z-score for Perceived Stress Scale Short Form | May be a useful predictor |
| BE | Significant Stress | Qualitative  (Categorical) | Indicates clinically significant stress.  Values: 1,2 | May be a useful predictor |
| BF | First Orebro Score | Quantitative  (Numerical- Continuous) | Is the total score for the Orebro Musculoskeletal Pain Questionnaire | May be a useful Predictor |
| BG | Significant Pain Complications | Qualitative  (Categorical) | Classified as having significant pain.  Values: 1,2 | May be a useful predictor |
| BH | Orebro Musculoskeletal Pain Total | Quantitative  (Numerical- Continuous) | Total score for the Orebro | May be a useful predictor |
| BI | Claims­\_Risk\_Assessment | Qualitative  (Ordinal) | Is the insurers concluded risk rating of the claim.  ("Low"=1, "Medium"= 2, "High"=3) | Useful to determine how good the insurers current risk rating is relative to the models calculated at the hackathon |
| BJ | Initial\_estimate | Quantitative  (Numerical- Continuous) | Is the amount of money set aside for the cost of claim. | This can be used to determine how much more accurate our models are |
| BK | Payment\_legal\_defendant | Quantitative  (Numerical- Continuous) | Money paid for legal involvement | Subset of claims cost, but an indicator of litigation. Could possibly be predicted (Bonus DV) |
| BL | Payment\_legal\_plaintiff | Quantitative  (Numerical- Continuous) | Money paid to injured persons law firm if a lawsuit is lost | Subset of claims cost, but indicator of litigation. Could possibly be predicted (Bonus DV) |
| BM | Payment\_Impairment | Quantitative  (Numerical- Continuous) | Is money paid for permanent injury | Could possibly be predicted (Bonus DV) |
| BN | Payment\_investigation\_surveillance | Quantitative  (Numerical- Continuous) | Money paid for fact finding and surveillance | Could possibly be predicted (Bonus DV) |
| BO | Net\_total\_incurred | Quantitative  (Numerical- Continuous) | Is the amount spent on the claim | Is one variable to be predicted (DV) |
| BP | Total\_Paid | Quantitative  (Numerical- Continuous) | Is the total amount paid on a claim | Is another variable to be predicted (DV) |
| BQ | Other\_Paid | Quantitative  (Numerical- Continuous) | Is additional claims cost | Is another variable to be predicted (DV) |
| BR | Payment\_medical | Quantitative  (Numerical- Continuous) | Is the amount of money paid for treatment | Is a Dependent Variable (DV) |
| BS | Other\_paid\_risk | Quantitative  (Numerical- Continuous) | Is money paid on a claim | Is a variable to be predicted (DV) |