Problem Statement

As a member of the Gallagher Bassett Analytics team you have been tasked with developing a model to predict whether a workers’ compensation claim will involve prescription opioids based on available information about the claimant, the accident that produced the claim, and his/her medical procedures subsequent to the accident.

Assignment

Build a model to predict Opioids Used, True (1) or False (0), from the data provided.

You will be asked to present your development process to other members of the team. Be prepared to explain:

* Any data preprocessing steps taken
* Why a particular model was selected
* How the performance of the model was evaluated
* What are the important features for predicting the target variable
* The steps to calculate a prediction for a claim not included in the provided data set
* In general terms that could be understood by a non-technical user, how the model generates its predictions

Prior to the meeting, please provide a brief summary of your model, any code that you used to build the model, as well as any exploratory analysis done. Also provide a saved version (e.g. a pickle file) of your model, and any additional details (e.g. particular software versions) that would be needed to run your code and validate your model.

Data Definitions

A CSV file (GB Analytics Modeling Data) has been provided with the following fields:

|  |  |
| --- | --- |
| **Field Name** | **Definition** |
| ClaimID | Unique identifier for a claim |
| Accident DateID\* | DateID when the accident occurred |
| Claim Setup DateID\* | DateID when Resolution Manager sets up the claim |
| Report To GB DateID\* | DateID when employer (client) notifies Gallagher Bassett of a claim |
| Employer Notification DateID\* | DateID when claimant notifies employer (client) of an injury |
| Benefits State | The jurisdiction whose benefits are applied to a claim. |
| Accident State | State in which the accident occurred |
| Industry ID | Broad industry classification categories |
| Claimant Age | Age of the injured worker (claimant) |
| Claimant Sex | Sex of the injured worker (claimant) |
| Claimant State | State in which the claimant resides |
| Claimant Marital Status | Marital status of the injured worker (claimant) |
| Number Dependents | Number of dependents the claimant has |
| Weekly Wage | An average of the claimant’s weekly wages as of the injury date. |
| Employment Status Flag | A code representing the claimant’s employment status at the time the claim was reported.  Valid employment status codes are:  F – Regular full-time employee  P – Part-time employee  U – Unemployed  S – On strike  D – Disabled  R – Retired  O – Other  L – Seasonal worker  V – Volunteer worker  A – Apprenticeship full-time  B – Apprenticeship part-time  C – Piece worker |
| RTW Restriction Flag | A Y/N flag, used to indicate whether the employee`s responsibilities upon returning to work were limited as a result of his/her illness or injury. |
| Max Medical Improvement DateID | DateID of Maximum Medical Improvement, after which further recovery from or lasting improvements to an injury or disease can no longer be anticipated based on reasonable medical probability. |
| Percent Impairment | Indicates the percentage of anatomic or functional abnormality or loss, for the body as a whole, which resulted from the injury and exists after the date of maximum medical improvement |
| Post Injury Weekly Wage | The weekly wage of the claimant after returning to work, post-injury, and/or the claim is closed |
| NCCI Job Code | A code that is established to identify and categorize jobs for workers’ compensation. |
| Surgery Flag | Indicates if the claimant’s injury will require or did require surgery |
| Disability Status | Type of disability benefit claim is classified as  Possible Values: Temporary Total Disability (TTD) Temporary Partial Disability (TPD) Permanent Partial Disability (PPD) Permanent Total Disability (PTD) |
| SIC Group | Standard Industry Classification (SIC) group for the client |
| NCCI BINatureOfLossDescription | Description of the end result of the bodily injury loss occurrence |
| Accident Source Code | A code identifying the object or source which inflicted the injury or damage. |
| Accident Type Group | A code identifying the general action which occurred resulting in the loss |
| Neurology Payment Flag | Indicates if there were any payments made for diagnosis and treatment of disorders of the nervous system without surgical intervention |
| Neurosurgery Payment Flag | Indicates if there were any payments made for services by physicians specializing in the diagnosis and treatment of disorders of the nervous system, including surgical intervention if needed |
| Dentist Payment Flag | Indicates if there were any payments made for prevention, diagnosis, and treatment of diseases of the teeth and gums |
| Orthopedic Surgery Payment Flag | Indicates if there were any payments made for surgery dealing with the skeletal system and preservation and restoration of its articulations and structures. |
| Psychiatry Payment Flag | Indicates if there were any payments made for treatment of mental, emotional, or behavioral disorders. |
| Hand Surgery Payment Flag | Indicates if there were any payments made for surgery only addressing one or both hands. |
| Optometrist Payment Flag | Indicates if there were any payments made to specialists who examine the eye for defects and faults of refraction and prescribe correctional lenses or exercises but not drugs or surgery |
| Podiatry Payment Flag | Indicates if there were any payments made for services from a specialist concerned with the care of the foot, including its anatomy, medical and surgical treatment, and its diseases. |
| HCPCS A Codes - HCPCS Z Codes | Count of the number of HCPCS (Healthcare Common Procedure Coding System) codes that appear on the claim within each respective code group |
| ICD Group 1 - ICD Group 21 | Count of the number of ICD (International Classification of Diseases) codes that appear on the claim within each respective code group |
| CPT Category - Anesthesia | Count of the number of CPT (Current Protocol Terminology) codes on the claim within each respective code category |
| CPT Category - Eval\_Mgmt |
| CPT Category - Medicine |
| CPT Category - Path\_Lab |
| CPT Category - Radiology |
| CPT Category - Surgery |
| NDC Class - Benzo | Count of the number of NDC (National Drug Code) codes on the claim within each respective code class |
| NDC Class - Misc (Zolpidem) |
| NDC Class - Muscle Relaxants |
| NDC Class - Stimulants |
| Opioids Used | A True (1) or False (0) indicator for whether or not the claim included opioid prescriptions |

\*DateID fields represent the number of days from an arbitrary day in the past. All DateID fields have been calculated from the same arbitrary day to allow for determining the time between any two DateID fields.