Presenting Authors: - (a)Dinesh Thapak,(b) Tom Olafsson and (c)Vignesh Prajapati
Organization/Institute:-P.hD researcher at R.K University, Rajkot, Gujarat,India
PhD guide:-Dr.Ranjit Goswami, Dean IMT Nagpur
Address: - Residence: - 50/Santoshinagar Bh Ganga Estate Nr.
                         Kubernagar, Sardarnagar, Ahmedabad,India-382340
E-Mail Address:-din.daiict@gmail.com.
Phone: - 09979975221
Country: - **India**

Paper Title: - Cloud RCM (Revenue Cycle Management), Claim denial management with R

Abstract

In this paper we are trying to predict the probability of medical claims to be accepted or denied

This paper will be divided into four parts

  a) Data extraction from database with the help of R
  b) Data cleaning and preparation for analysis with R
  c) Prediction/Forecasting of claim denial with R
  d) Denial management for future claims with R

A clean medical claim is profitable to all stakeholders involving in it and is defined as a "claim that meets the standards set by payer (insurance) for payment on first submission".

With this Denial Management analysis, we are trying to fix those issues which lead to the claim denials and can predict with more than 95% accuracy that whether it be an issues with the claims (internal) or issues with the payers (external).

Final result of this paper will help in building "Clean Claim Rules".
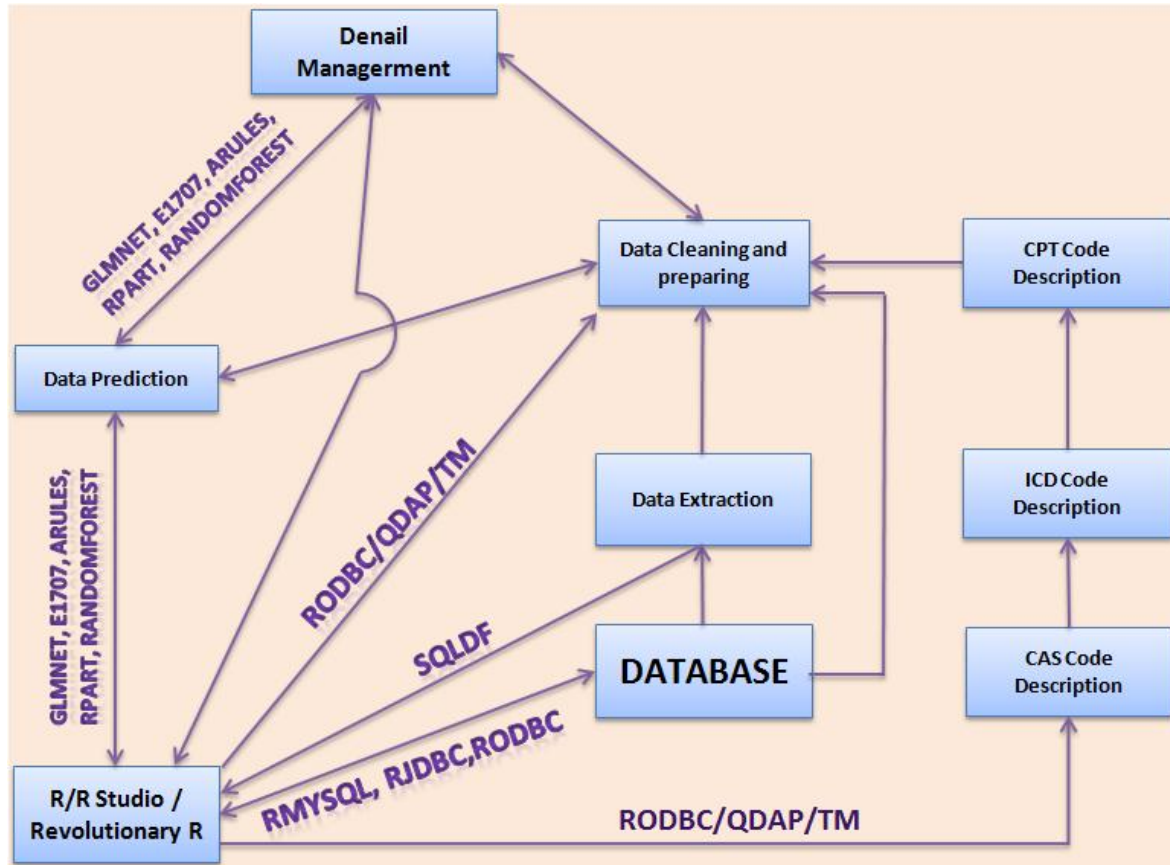
Denial management is a directly associated to accounts receivables and if it is properly monitored/tracked, analysed, trend identification with preventive mechanism then lot of profit and time can be saved.

Introduction

(a)Process flow of Claim
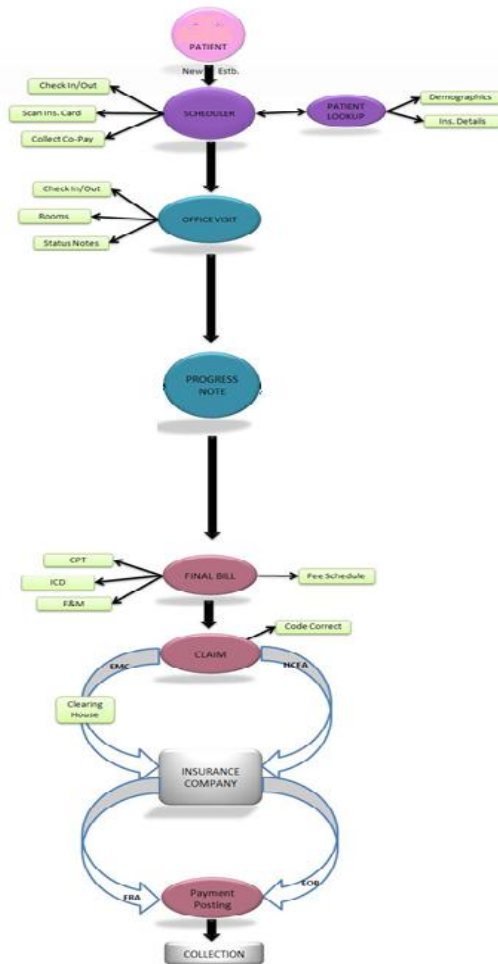
(b)Process flow of paper is as below:-

Keywords

Medical claims, CPT (Current Procedures Terminology), ICD (International Classification of Diseases), CAS (Claim Adjustment Reason) codes, DOS (Date of Service), Payer (Insurance), Modifiers, Demographics.

Methodology

(a)Data to be extracted for following variables

- Claims: Past processed claims with label – Denied/Approved
- CPT: All Procedural codes
- ICD: All Diagnosis codes
- CAS: All group and reason codes
- Patient demographic: Gender, Age
- Payer information

(b)Tools/systems/software's to be used paper:

- R /R Studio/Revelation R
- MySQL/MS SQL – for data dump
- Excel/Open office– for data transformation

(c)Broadly R packages to be used in this paper:-

- Basic packages for loading R
- Rattle-Graphical user interface for data mining in R
- ReporteRs-Microsoft Word, Microsoft Powerpoint and HTML documents generation from R
- reportRx-Tools for automatically generating reproducible clinical report
- spikeslab-Prediction and variable selection using spike and slab regression
- rtf-Rich Text Format (RTF) Output
- jointPm-Risk estimation using the joint probability method
- PCS-Calculate the probability of correct selection (PCS)
- Pmr-Probability Models for Ranking Data
- Batade-HTML reports and so on
- Editrules-R package for parsing, applying, and manipulating data cleaning rules
- Dataview-Human readable data presentation
- Tm-Text Mining Package
- e1071-Misc Functions of the Department of Statistics (for fitting naïve Bayes model)
- glmnet-Lasso and elastic-net regularized generalized linear models – For fitting generalized
- arules-Mining Association Rules and Frequent Itemsets(for identifying the associated rules with the combination of ICD, CPT and patient information)
- arulesSequences-Mining frequent sequences
- arulesViz-Visualizing Association Rules and Frequent Itemsets
- rpart-Recursive Partitioning and Regression Trees
- rpartitions-Code for integer partitioning
- rpart.plot-Plot rpart models
- rpartScore-Classification trees for ordinal responses
- randomForest- Breiman and Cutler's random forests for classification and regression
- randomForestSRC-Random Forests for Survival, Regression and Classification (RF-SRC)
- randomGLM-Random General Linear Model Prediction
- A3-Accurate, Adaptable, and Accessible Error Metrics for Predictive Models

- abn-Data Modelling with Additive Bayesian Network
- accrual-Bayesian Accrual Prediction
- ACD-Categorical data analysis with complete or missing responses
- ADM3-An Interpretation of the ADM method - automated detection algorithm
- Approximator-Bayesian prediction of complex computer codes
- AUCRF-Variable Selection with Random Forest and the Area Under the Curve
- BayesLogit-Logistic Regression
- ChainLadder-Statistical methods for the calculation of outstanding claims reserves in general insurance
- DCL-Claims Reserving under the Double Chain Ladder Model
- dbConnect-Provides a graphical user interface to connect with databases that use MySQL
- DBI-R Database Interface
- Mapdata-Extra Map Databases
- ora-Convenient Tools for Working with Oracle Databases
- FAiR-Factor Analysis in R
- Tsfa-Time Series Factor Analysis
- Faoutlier-Influential case detection methods for factor analysis and SEM
- rJava-Low-level R to Java interface
- rJavax-rJava extensions
- RJDBC-Provides access to databases through the JDBC interface
- RMySQL-R interface to the MySQL database
- RODBC-ODBC Database Access
- sqliter-Connection wrapper to SQLite databases
- sqlshare-API for access to SQLShare database
- sqldf-Perform SQL Selects on R Data Frames
- zipcode-U.S. ZIP Code database for geocoding
- caroline-A Collection of Database, Data Structure, Visualization, and Utility Functions for R
- cdb-Reading and Writing Constant DataBases
- lubridate-Make dealing with dates a little easier

(d)Techniques

Descriptive analysis, SEM (Structure equation modelling), FA (Factor analysis), Bayes theorem (Naïve Bayes which can be used for classification), Decision Tree, Random forest, Regression, Association Rules techniques.

(e)Database

Probable database table list:

- claims [Claim Table]
- claimstatus [Claim status description]
- claimstatus_log [Claim status]
- insurance [Insurance Table]
- diagnosis [Diagnosis table]
- paymentdetail [Claim Payment Table]
- adjustments [Claim Payment adjustment Table]
- claimrefund [Claim Refund Table]

(f)Connecting Database with R

(g)Data Cleaning

Steps:
- Data variables with null value should be removed
- Data variables with repeated value to be removed
- Data variables with inappropriate value to be removed

(h)Data Transformation and Preparing

Steps:
- Date of service and Date of Birth need to be transformed to appropriate data format
- Age to be calculated by Date of Birth and Date of service
- All CPT values to be transformed as categorical values
- All modifiers need to be transformed as categorical values
- All the extracted result need to be merged based on the claim no as unique value
- Data need to be splitted/sampled for training and testing datasets

(i)Data Prediction

Steps:
- Prediction model need to be fit over the training dataset.
- Prediction model accuracy to be identified
- Defining the model parameters to improving the model prediction accuracy
- Text mining based Claim filtering for (a) CPT and Claim attributes mapping, (b) ICD and claim attributes mapping, (c) CAS codes and claim attributes mapping (d) mapping of ICD – CPT linking.

(j)Denial Management for future claims

Once the Claim prediction model deployed and integrated with production system. We need to continuously measure the prediction accuracy and improve the model for higher accuracy. Once the model enters in to the production environment, we need to manage or update the model by writing the R script such that it will build the model over new claim data as well. So, again it will improve the model prediction accuracy to very precise level.

(k)Type/source of Data:-Primary/Cognos/Slave Database/ open ERP.

Findings and results

Probability of accepted or denial fresh medical claim and residual error.

Key Findings: - In full paper

Result: - In full paper

Reference:-

Distinguishing the Forest from the TREES:A Comparison of Tree Based Data Mining Methods by Richard Derrig, Ph.D. and Louise Francis, FCAS, MAAA

Rule-based Prediction of Medical Claims' Payments-A Method and Initial Application to Medicaid Data Janusz Wojtusiak
Semantic Data Types in Machine Learning from Healthcare Data by Janusz Wojtusiak

www.cse.buffalo.edu/research/areas/pattern.php

www.mli.gmu.edu/papers/2011/11-4.pdf

www.cs.uiuc.edu/~hanj/pdf/ency99.pdf

www.dais.cs.uiuc.edu/manish/pub/gupta11b_apdsdm.pdf

www.google.com/patents/US20130054259

www.cs.bham.ac.uk/~wbl/biblio/wcci2008.bib

www.casact.org/pubs/forum/06wforum/06w05.pdf

www.cran.r-project.org/

www.humanarc.com/wp-content/uploads/2013/06/HOSPITAL-DENIALS-WHITE-PAPER-by-Holly-Pelaia-2013-05.pdf

www.ats.ucla.edu

www.statmethods.net(Quick-R)

https://www.rstudio.com

www.stackoverflow.com

www.r-tutor.com

www.math.illinoisstate.edu

www.nceas.ucsb.edu

www.programmingr.com

http://r.789695.n4.nabble.com/

www.r-bloggers.com

www.r-statistics.com

tryr.codeschool.com

ww2.coastal.edu

www.eclinicalworks.com

www.poweryourpractice.com/

www.waitingroomsolutions.com

www.ircmhealth.com/

www.revenue360.net/

www.convergentusa.com/

www.wipro.com/Documents/resource-center/library/provider_revenue_cycle_management.pdf?Mobile=1

www.precisionpractice.com/