**DIABETES CLASSIFICATION**

**1. Summary**

The model for Diabetes Classification is a blending of 8 boosted regression trees (gbm) and 4 random forest models.

Featured creation was done grouping similar diagnostics in a two level categories schema. The medication file was heavily treated for get each drug actives principles (up to four), administration route and dose. More features were calculated as counters (visits, groups of diagnostics, physicians, specialties, actives principles, prescriptions, diagnoses with medication associated…) and as ratios (visits per year, diagnostics per visit, diagnostic per year…)

For model stacking a generalized additive model (gam) with cubic splines was used.

CCS classification (http://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp) helped diagnostics grouping, but not was used literally.

**2. Features Selection / Extraction**

At first step, outliers and possible typos of transcriptsmeasures were cleaned and height median calculated. BMI was recalculated with this constant height for each patient, eliminating so the noise of measures fluctuations.

For weight, height, BMI, systolic blood pressure, diastolic blood pressure, temperature, respiratory rate were calculated median and truncated maximum and minimum.

For 2012 and 2009 years, not complete, a weight was used for calculate features with ratios.

Weight 2012 = 2 \* Total visits 2012 / (Total visits 2010 + Total visits 2011)

Weight 2009 = 2 \* Total visits 2009 / (Total visits 2010 + Total visits 2011)

The anthropometric features and the summarized visits are commented in code.

The diagnostics were grouped at two levels based in clinical etiology or symptoms similarities. 245 level2 and 22 level1 groups were created. For level2 groups not entered in the final models but I considered (intuition, experience or bibliography) they could be correlated with DM, DMSymptom level 1 group were used for create a score: number of different level2 groups (with that level 1) in which the patient had diagnostics. Features were created for number of transcript diagnostics in each group.

The medication was treated at the active principle level. For each NDC code active principle, administration route and dose were extracted. After that, active principles were grouped in families taking into account chemical similarities or common clinical indication. Features were created for maximum dose of an active principle / family (\_dose suffix), number of active principles in the family administered to the patient (\_nap), number of prescriptions (\_npr), and binary flags (\_bin).

Some families treated were ACEI (Angiotensin Converting Enzyme Inhibitor), AIIRA (Angiotensin II Receptor Antagonists), antifungals, benzodiazepines, beta blockers, fibrates, glucocorticoids, L type Calcium channel blockers, statins, thiazides, antilipemics or loop diuretics.

Other features were created for missing data: medication without prescriptions, diagnoses without transcripts, patients without medication, medication with unknown diagnostic, lab panel without lab observation. Some of them were commented in the forum and recognized as data leaks, other could be consequence of a ‘EHR use bias’ (labs are inverse correlated with DM so probably the facilities using lab records are primary care and not specialized medicine).

‘Smoking status’ and ‘Previous smoking situation’ were created. Other possible features like allergy or immunization don’t take account due low number of patients.

For states with more than 450 patients a binary featured was used.

The featured selection was interactive using ‘Relative Influence’ output of gbm models and mainly ‘Increment node purity’ output of random forest models. There is no a methodology here.

**3. Modeling Techniques and Training**

Eight boosted regression trees (gbm) were fitting with different parameters. For cross validation stopping the wrapper ‘gbm.step’ in dismo R package was used, only modified for add n.minobsinnode parameter.

Depth, shrinkage, bag fraction and nodesize are parameters of gbm. Tolerance is parameter of dismo wrapper for controlling cross validation stopping (see packages documentation for detail).

All models trained over ‘dmii’ subset. Models ended in ‘\_ext’ use ‘exten’ subset too.

Four models were fitted with randomForest, all parameters by defect except the detailed in table.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **GBM Model** | **cv error** | **folds** | **trees** | **depth** | **shrinkage** | **bag fr** | **nodesize** | **Tolerance** |
| gbm10\_5\_0.003\_0.80\_30 | 0.31240 | 10 | 7,550 | 5 | 0.003 | 0.80 | 30 | 0.001 |
| gbm10\_5\_0.003\_0.80\_30\_tolhalf\_ext | 0.31153 | 10 | 8,500 | 5 | 0.003 | 0.80 | 30 | 0.0005 |
| gbm10\_5\_0.0025\_0.80\_30 | 0.31319 | 10 | 8,000 | 5 | 0.003 | 0.80 | 30 | 0.001 |
| gbm10\_5\_0.0025\_0.80\_30\_ext | 0.31312 | 10 | 7,750 | 5 | 0.0025 | 0.80 | 30 | 0.001 |
| gbm10\_5\_0.0025\_0.80\_30\_tolhalf | 0.31122 | 10 | 11,000 | 5 | 0.0025 | 0.80 | 30 | 0.0005 |
| gbm10\_5\_0.0025\_0.80\_30\_tolhalf\_ext | 0.31139 | 10 | 10,450 | 5 | 0.0025 | 0.80 | 30 | 0.0005 |
| gbm20\_5\_0.002\_0.80\_10 | 0.31049 | 20 | 13,450 | 5 | 0.002 | 0.80 | 10 | 0.001 |
| gbm20\_5\_0.002\_0.80\_15 | 0.31040 | 20 | 12,950 | 5 | 0.002 | 0.80 | 15 | 0.001 |
| gbm20\_5\_0.0025\_0.80\_20 | 0.30878 | 20 | 12,300 | 5 | 0.0025 | 0.80 | 20 | 0.001 |
| gbm20\_5\_0.0025\_0.80\_40 | 0.31040 | 20 | 10,500 | 5 | 0.0025 | 0.80 | 40 | 0.001 |
| gbm20\_6\_0.002\_0.80\_30 | 0.30931 | 20 | 12,200 | 6 | 0.002 | 0.80 | 30 | 0.001 |

|  |  |  |  |
| --- | --- | --- | --- |
| **RF Model** | **OOB mse** | **trees** | **nodesize** |
| RF1 | 0.10055 | 15,000 | 5 |
| RF5 | 0.10076 | 30,000 | 15 |
| RF2 | 0.10089 | 15,000 | 20 |
| RF3 | 0.10102 | 15,000 | 40 |

**4. Code Description**

The code is split in the following files:

DMfeatureCreation.R

Connect compData.db, clean the data and create features.

The output is file Patient.csv.

DMgbm.R

Read Patient.csv file and fit the boosted trees models. The cross validation stopping is controled with dismo wrapper.

As output the model is saved in RData format, and in csv formatthe predictions of test set and the CV folds predictions of train set for model stacking (the latter with cvest suffix in filename).

DMrandomforest.R

Read Patient.csv file and fit the random forest models.

As output the model is saved in RData format, and in csv format the predictions of test set and the OOB predictionsof train set for model stacking (the latter with cvest suffix in filename).

DMstacking.R

Do the stacking of the models using a generalized additive model (gam) with cubic splines.

Read CV and OOB predictions of training set, fit the gam and do a prediction for test set saved as dmpredict.csv file.

gbm.step.R and gbm.utils.R

Used for add n.minobsinnode parameter in the gbm.step wrapper of the dismo package. Dismo use the default for this parameter (n=10) and in small data sets with relatively big number of features increase the overfitting risk.

**5. How To Generate the Solution**

Copy files in c:\dmii\

DMfeatureCreation.R

DMgbm.R

DMrandomforest.R

DMstacking.R

gbm.step.R

gbm.utils.R

Copy in c:\dmii\data

compData.db

drugsap.csv

icd9.csv

Create c:\dmii\models for outputs

Run the source files in this order:

DMfeatureCreation.R

DMgbm.R

DMrandomforest.R

DMstacking.R

**6. Results interpretation**

Boosted regression trees and random forest are powerful in detect deepest interactions in complex model like are frequents in health, but weakness is that they are black box models. It’s hard work extract structured knowledge of them.

The increment of node purity in random forest (average over four models after drop features relates with data leaks and bias selection earlier mentioned), is present.

This is an indicator of relative importance of each feature in the model. But the data interpretation must be done carefully. A small value could be consequence of little influence in the presence of diabetes mellitus or perhaps the high correlation with other features or the presence of confounding variables.

At last, a high value uniquely is interpretable as correlation (comorbidity or simultaneous presence), a relation cause – effect may not be extrapolated. This is a restriction of the transversal character of the data.

**FEATURE INCREMENT NODE PURITY (RANDOM FOREST)**

|  |  |  |  |
| --- | --- | --- | --- |
| Feature | Inc node purity | Feature | Inc node purity |
| L2\_HypertensionEssential | 9.052 | L2\_Anxiety | 0.356 |
| YearOfBirth | 5.755 | L2\_PainJoint | 0.355 |
| BMIMaxT | 5.449 | ACEI\_dose | 0.351 |
| DiastolicBPMedian | 3.542 | PrevSmoker | 0.351 |
| L2\_MixedHyperlipidemia | 3.436 | L2\_GlucoseAbnormal | 0.346 |
| WeightMedian | 2.959 | L2\_SleepApnea | 0.342 |
| active\_principle | 2.585 | L2\_Cough | 0.334 |
| TotDiagPerVisit | 2.497 | L2\_NervousSystem | 0.333 |
| diag\_3digit\_with\_medication | 2.478 | L2\_Osteoporosis | 0.332 |
| SystolicBPMaxT | 2.454 | L2\_Dysrhythmia | 0.332 |
| prescripts | 2.390 | L2\_Gout | 0.327 |
| TotDiagWYear | 2.010 | L2\_Impotence | 0.326 |
| statin\_dose\_adjusted | 2.004 | L1\_NervousSystemOther | 0.324 |
| RangeBMI | 1.956 | STATE\_CA | 0.320 |
| HeightMedian | 1.615 | L2\_Malaise | 0.299 |
| L2\_ChronicRenalFailure | 1.570 | L2\_PeriphNeuropathy | 0.272 |
| TotDiag | 1.432 | atenolol\_dose | 0.270 |
| TemperatureMedian | 1.338 | L2\_Proteinuria | 0.268 |
| VisitPerWYear2Date | 1.334 | inhalation\_npr | 0.267 |
| statin\_nap | 1.327 | L2\_Diarrhea | 0.267 |
| MaxVisitWYear | 1.317 | amlodipine\_dose | 0.267 |
| L2\_HyperlipOther | 1.208 | L2\_HerpesZoster | 0.261 |
| Weighted | 1.205 | L2\_EKGAbnormal | 0.261 |
| Tot3digitICD9 | 1.144 | L2\_Renal | 0.260 |
| DiffLevel2Diag | 1.139 | olmesartan\_dose | 0.246 |
| InternalMedicine | 1.067 | STATE\_NY | 0.244 |
| RespiratoryRateMedian | 1.050 | L2\_MycosisFoot | 0.242 |
| VisitTotal | 1.024 | L2\_Constipation | 0.241 |
| Feature | Inc node purity | Feature | Inc node purity |
| L2\_HypertensionComp | 0.895 | Podiatry | 0.240 |
| MinVisit2Date | 0.894 | L2\_SexDysfunction | 0.230 |
| L2\_DMRelated2 | 0.865 | L2\_IntestinalInfection | 0.229 |
| L2\_AtherosclerosisCoronary | 0.855 | L2\_ImpGlucoseAbnormal | 0.228 |
| L2\_Hypercholesterolemia | 0.820 | L2\_VascularPeripheral | 0.224 |
| fibrate\_npr | 0.773 | L2\_Dermatitis | 0.221 |
| ACEI\_bin | 0.715 | L2\_Ulcer | 0.217 |
| FamilyPractice | 0.678 | antilipid\_bin | 0.214 |
| NumPhysicians | 0.658 | CardiovascularDisease | 0.212 |
| DiffDMOtherRelatedSymptoms | 0.603 | loopdiuretic\_bin | 0.206 |
| L2\_Osteoarthrosis | 0.592 | L2\_VertiginousSyndromes | 0.203 |
| L1\_Back | 0.584 | L1\_Skin | 0.191 |
| L2\_Edema | 0.569 | L2\_Dyspnea | 0.187 |
| GeneralPractice | 0.557 | L2\_VitaminB | 0.180 |
| L2\_Obesity | 0.550 | Gender | 0.174 |
| lisinopril\_dose | 0.548 | L2\_BoneDeformity | 0.166 |
| L2\_DMRelated | 0.506 | L2\_Hearing | 0.165 |
| AIIRA\_npr | 0.488 | L2\_FamilyDM | 0.164 |
| STATE\_TX | 0.475 | L2\_UrinaryIncontinency | 0.162 |
| LtypeCaChB\_npr | 0.468 | L2\_RespiratoryFail | 0.161 |
| L2\_TesticularDysfunction | 0.452 | thiazide\_bin | 0.160 |
| losartan\_telmisartan\_dose | 0.439 | betablocker\_bin | 0.150 |
| L2\_AtherosclerosisPeripheral | 0.438 | L2\_NeurRadiculitis | 0.149 |
| L2\_Esophagus | 0.435 | glucocorticoid\_general\_bin | 0.148 |
| L2\_COPD | 0.426 | L2\_SkinSense | 0.135 |
| L1\_Prostate | 0.420 | L2\_Migraine | 0.132 |
| L1\_RespInfec | 0.408 | L2\_GallBladder | 0.123 |
| L2\_CardiacInsufficiency | 0.399 | benzodiazepine\_bin | 0.123 |
| L2\_Cellulitis | 0.396 | L2\_Hypoglycemia | 0.121 |
| L2\_VitaminD | 0.392 | antiplatelet\_bin | 0.113 |
| L2\_CerebroVascular | 0.387 | L2\_IrritableBowel | 0.109 |
| aspirin\_npr | 0.383 | L2\_Hyperkalemia | 0.084 |
| carvedilol\_dose | 0.366 | gastroparesia\_bin | 0.081 |
| L2\_Hypertriglyceridemia | 0.366 | L2\_MenstruationDisorder | 0.070 |
| L2\_DeficiencyAnemia | 0.363 | osteoporosis\_bin | 0.070 |
| L2\_Vaccine | 0.359 |  |  |

**7. References**

Software used:

R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL

<http://www.R-project.org/>

Liaw and M. Wiener (2002).Classification and Regression by randomForest.R News 2(3), 18--22.

Greg Ridgeway (2012). gbm: Generalized Boosted Regression Models. R package version 1.6-3.2.

<http://CRAN.R-project.org/package=gbm>

Robert J. Hijmans, Steven Phillips, John Leathwick and Jane Elith (2012).dismo: Species distribution modeling. R package version 0.7-17.

<http://CRAN.R-project.org/package=dismo>

Trevor Hastie (2011). gam: Generalized Additive Models. R package version 1.06.2.

<http://CRAN.R-project.org/package=gam>

David A. James (2011). RSQLite: SQLite interface for R. R package version 0.11.1.

<http://CRAN.R-project.org/package=RSQLite>

CCS classification.

<http://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp>

**I. Annex. Feature description.**

|  |  |
| --- | --- |
| Feature | Description |
| PatientGuid | PatientGuid |
| dmIndicator | dmIndicator |
| Gender | Gender |
| YearOfBirth | Year of birth |
| State | State |
| HeightMedian | Height median |
| WeightMedian | Weight median |
| WeightMaxT | Weight truncated max |
| BMIMaxT | BMI truncated max |
| BMIMinT | BMI truncated min |
| BMIMedian | BMI median |
| RangeBMI | Range BMI |
| SystolicBPMaxT | Systolic BP truncated max |
| SystolicBPMinT | Systolic BP truncated min |
| SystolicBPMedian | Systolic BP median |
| DiastolicBPMaxT | Diastolic BP truncated max |
| DiastolicBPMinT | Diastolic BP truncated min |
| DiastolicBPMedian | Diastolic BP median |
| RangeSystolicBP | Range systolic BP |
| RangeDiastolicBP | Range diastolic BP |
| HighLowBP | High Low difference BP |
| RespiratoryRateMaxT | Respiratory rate truncated max |
| RespiratoryRateMedian | Respiratory rate median |
| TemperatureRank2th | Temperature second lowest |
| TemperatureMedian | Temperature median |
| PrevSmoker | Previous smoker |
| Smoker | Smoker status |
| InternalMedicine | Internal medicine |
| CardiovascularDisease | Cardiovascular disease |
| FamilyPractice | Family practice |
| GeneralPractice | General practice |
| Podiatry | Podiatry |
| NumSpecialties | Number of specialties |
| VisitYearBlank | Visits with year blank |
| VisitYear2009 | Visits in year 2009 |
| VisitYear2010 | Visits in year 2010 |
| VisitYear2011 | Visits in year 2011 |
| VisitYear2012 | Visits in year 2012 |
| VisitTotal | Visits total |
| MaxVisitYear | Max visits per year |
| FirstYear | First year with visits |
| Feature | Description |
| LastYear | Last year with visits |
| RangeYear | Range year (last -first) |
| Years2Date | Range year (2012 - first) |
| MaxVisitWYear | Max visits per weighted year |
| MinVisit2Date | Min visits a year (to date) |
| MinVisit2Last | Min visits a year (to last year with visits) |
| NumPhysicians | Number of physicians |
| VisitPerWYear2Date | Visits per weighted year (to date) |
| VisitPerWYear2Last | Visits per weighted year (to last year with visits) |
| Heighted | Visits with not null height |
| Weighted | Visits with not null weight |
| L2\_AbdominalHernia | Abdominal hernia (level 2)(number of transcript diagnostics) |
| L2\_AbdominalPain | Abdominal pain (level 2)(number of diagnostics) |
| L2\_AbuseMonotoring | Abuse monotoring (level 2)(number of diagnostics) |
| L2\_Acne | Acne (level 2)(number of diagnostics) |
| L2\_AcuteBronchitis | Acute bronchitis (level 2)(number of diagnostics) |
| L2\_AcuteCystitis | Acute cystitis (level 2)(number of diagnostics) |
| L2\_Alcohol | Alcohol (level 2)(number of diagnostics) |
| L2\_Allergy | Allergy (level 2)(number of diagnostics) |
| L2\_AMI | AMI (level 2)(number of diagnostics) |
| L2\_AnginaPectoris | Angina pectoris (level 2)(number of diagnostics) |
| L2\_Anxiety | Anxiety (level 2)(number of diagnostics) |
| L2\_Arthropathy | Arthropathy (level 2)(number of diagnostics) |
| L2\_Asthma | Asthma (level 2)(number of diagnostics) |
| L2\_AtherosclerosisCoronary | Atherosclerosis coronary (level 2)(number of diagnostics) |
| L2\_AtherosclerosisPeripheral | Atherosclerosis peripheral (level 2)(number of diagnostics) |
| L2\_BackPain | Back pain (level 2)(number of diagnostics) |
| L2\_Bladder | Bladder (level 2)(number of diagnostics) |
| L2\_BlindDeficiency | Blind deficiency (level 2)(number of diagnostics) |
| L2\_BloodAnormal | Blood anormal (level 2)(number of diagnostics) |
| L2\_BoneCartilage | Bone cartilage (level 2)(number of diagnostics) |
| L2\_BoneDeformity | Bone deformity (level 2)(number of diagnostics) |
| L2\_Calculus | Calculus (level 2)(number of diagnostics) |
| L2\_Candida | Candida (level 2)(number of diagnostics) |
| L2\_CaProstate | Ca prostate (level 2)(number of diagnostics) |
| L2\_CardiacInsufficiency | Cardiac insufficiency (level 2)(number of diagnostics) |
| L2\_CardiacOther | Cardiac other (level 2)(number of diagnostics) |
| L2\_CardiacValve | Cardiac valve (level 2)(number of diagnostics) |
| L2\_CarpalSyndrome | Carpal syndrome (level 2)(number of diagnostics) |
| L2\_CaSkin | Ca skin (level 2)(number of diagnostics) |
| L2\_Cataract | Cataract (level 2)(number of diagnostics) |
| L2\_Cellulitis | Cellulitis (level 2)(number of diagnostics) |
| L2\_CerebralDegeneration | Cerebral degeneration (level 2)(number of diagnostics) |
| Feature | Description |
| L2\_CerebroVascular | Cerebro vascular (level 2)(number of diagnostics) |
| L2\_Cerumen | Earwax (level 2)(number of diagnostics) |
| L2\_Cervical | Cervical (level 2)(number of diagnostics) |
| L2\_ChestPain | Chest pain (level 2)(number of diagnostics) |
| L2\_ChronicCystitis | Chronic cystitis (level 2)(number of diagnostics) |
| L2\_ChronicPainSynd | Chronic pain syndrome (level 2)(number of diagnostics) |
| L2\_ChronicRenalFailure | Chronic renal failure (level 2)(number of diagnostics) |
| L2\_Coagulation | Coagulation (level 2)(number of diagnostics) |
| L2\_ColitisNoninfectious | Colitis no infectious (level 2)(number of diagnostics) |
| L2\_Conjunctivitis | Conjunctivitis (level 2)(number of diagnostics) |
| L2\_Constipation | Constipation (level 2)(number of diagnostics) |
| L2\_COPD | COPD (level 2)(number of diagnostics) |
| L2\_Corns | Corns (level 2)(number of diagnostics) |
| L2\_Cough | Cough (level 2)(number of diagnostics) |
| L2\_CRP | CRP (level 2)(number of diagnostics) |
| L2\_DeficiencyAnemia | Deficiency anemia (level 2)(number of diagnostics) |
| L2\_Dementia | Dementia (level 2)(number of diagnostics) |
| L2\_Dermatitis | Dermatitis (level 2)(number of diagnostics) |
| L2\_Diarrhea | Diarrhea (level 2)(number of diagnostics) |
| L2\_DieteticMonotoring | Dietetic monitoring (level 2)(number of diagnostics) |
| L2\_DigestiveBleeding | Digestive bleeding (level 2)(number of diagnostics) |
| L2\_DigestiveUlcer | Digestive ulcer (level 2)(number of diagnostics) |
| L2\_Diverticula | Diverticula (level 2)(number of diagnostics) |
| L2\_DMII | DMII sure related (level 2)(number of diagnostics) |
| L2\_DMRelated | DM medium Related (level 2)(number of diagnostics) |
| L2\_DMRelated2 | DM high Related (level 2)(number of diagnostics) |
| L2\_Drug | Drug (level 2)(number of diagnostics) |
| L2\_DrugOther | Drug other (level 2)(number of diagnostics) |
| L2\_Dysphagia | Dysphagia (level 2)(number of diagnostics) |
| L2\_Dyspnea | Dyspnea (level 2)(number of diagnostics) |
| L2\_Dysrhythmia | Dysrhythmia (level 2)(number of diagnostics) |
| L2\_Dysuria | Dysuria (level 2)(number of diagnostics) |
| L2\_Edema | Edema (level 2)(number of diagnostics) |
| L2\_EKGAbnormal | EKG abnormal (level 2)(number of diagnostics) |
| L2\_Elbow | Elbow (level 2)(number of diagnostics) |
| L2\_EndocrineOther | Endocrine other (level 2)(number of diagnostics) |
| L2\_Enthesopathy | Enthesopathy (level 2)(number of diagnostics) |
| L2\_Epilepsy | Epilepsy (level 2)(number of diagnostics) |
| L2\_Esophagus | Esophagus (level 2)(number of diagnostics) |
| L2\_EssentialTremor | Essential tremor (level 2)(number of diagnostics) |
| L2\_Exanthem | Exanthem (level 2)(number of diagnostics) |
| L2\_ExternalOtitis | External otitis (level 2)(number of diagnostics) |
| L2\_EyeInfection | Eye infection (level 2)(number of diagnostics) |
| L2\_EyeOther | Eye other (level 2)(number of diagnostics) |
| Feature | Description |
| L2\_FamilyCV | Family cardiovascular history (level 2)(number of diagnostics) |
| L2\_FamilyDM | Family diabetes history (level 2)(number of diagnostics) |
| L2\_Fever | Fever (level 2)(number of diagnostics) |
| L2\_Fibromatoses | Fibromatoses (level 2)(number of diagnostics) |
| L2\_Flatulence | Flatulence (level 2)(number of diagnostics) |
| L2\_Flu | Flu (level 2)(number of diagnostics) |
| L2\_FluidDisorder | Fluid disorder (level 2)(number of diagnostics) |
| L2\_Furuncle | Furuncle (level 2)(number of diagnostics) |
| L2\_GallBladder | Gall bladder (level 2)(number of diagnostics) |
| L2\_GastricFunctional | Gastric functional (level 2)(number of diagnostics) |
| L2\_GastritisDuodenitis | Gastritis duodenitis (level 2)(number of diagnostics) |
| L2\_Glaucoma | Glaucoma (level 2)(number of diagnostics) |
| L2\_GlucoseAbnormal | Glucose abnormal (level 2)(number of diagnostics) |
| L2\_Goitier | Goitier (level 2)(number of diagnostics) |
| L2\_Gout | Gout (level 2)(number of diagnostics) |
| L2\_HairDisease | Hair disease (level 2)(number of diagnostics) |
| L2\_Headache | Headache (level 2)(number of diagnostics) |
| L2\_Healthy | Healthy (level 2)(number of diagnostics) |
| L2\_Hearing | Hearing (level 2)(number of diagnostics) |
| L2\_Hematury | Hematury (level 2)(number of diagnostics) |
| L2\_Hemorrhagic | Hemorrhagic (level 2)(number of diagnostics) |
| L2\_HemorrhagicAnemia | Hemorrhagic anemia (level 2)(number of diagnostics) |
| L2\_Hemorrhoids | Hemorrhoids (level 2)(number of diagnostics) |
| L2\_Hepatitis | Hepatitis (level 2)(number of diagnostics) |
| L2\_Hepatobiliary | Hepatobiliary (level 2)(number of diagnostics) |
| L2\_HerpesSimplex | Herpes simplex (level 2)(number of diagnostics) |
| L2\_HerpesZoster | Herpes zoster (level 2)(number of diagnostics) |
| L2\_HighBP | High BP without hypertension (level 2)(number of diagnostics) |
| L2\_Hypercholesterolemia | Hypercholesterolemia (level 2)(number of diagnostics) |
| L2\_Hyperkalemia | Hyperkalemia (level 2)(number of diagnostics) |
| L2\_HyperlipOther | Hyperlipemic other (level 2)(number of diagnostics) |
| L2\_HypertensionComp | Hypertension complicated (level 2)(number of diagnostics) |
| L2\_HypertensionEssential | Hypertension essential (level 2)(number of diagnostics) |
| L2\_Hypertriglyceridemia | Hypertriglyceridemia (level 2)(number of diagnostics) |
| L2\_Hypoglycemia | Hypoglycemia (level 2)(number of diagnostics) |
| L2\_Hypokalemia | Hypokalemia (level 2)(number of diagnostics) |
| L2\_Hyposmolarity | Hyposmolarity (level 2)(number of diagnostics) |
| L2\_Hypotension | Hypotension (level 2)(number of diagnostics) |
| L2\_Hypothyroidism | Hypothyroidism (level 2)(number of diagnostics) |
| L2\_Impetigo | Impetigo (level 2)(number of diagnostics) |
| L2\_ImpGlucoseAbnormal | Impaired glucose abnormal (level 2)(number of diagnostics) |
| L2\_Impotence | Impotence (level 2)(number of diagnostics) |
| Feature | Description |
| L2\_InadequatedDiet | Inadequated diet (level 2)(number of diagnostics) |
| L2\_Insomnia | Insomnia (level 2)(number of diagnostics) |
| L2\_InsomnioNO | Insomnio (no organic) (level 2)(number of diagnostics) |
| L2\_IntestinalInfection | Intestinal infection (level 2)(number of diagnostics) |
| L2\_IntestinalMalabsorption | Intestinal malabsorption (level 2)(number of diagnostics) |
| L2\_IronAnemia | Iron anemia (level 2)(number of diagnostics) |
| L2\_IrritableBowel | Irritable bowel (level 2)(number of diagnostics) |
| L2\_Leukocytes | Leukocytes (level 2)(number of diagnostics) |
| L2\_LichenPruritus | Lichen pruritus (level 2)(number of diagnostics) |
| L2\_LimbCramp | Limb cramp (level 2)(number of diagnostics) |
| L2\_LimbPain | Limb pain (level 2)(number of diagnostics) |
| L2\_LiverAbnormal | Liver abnormal (level 2)(number of diagnostics) |
| L2\_LumbarPain | Lumbar pain (level 2)(number of diagnostics) |
| L2\_Lymphadenitis | Lymphadenitis (level 2)(number of diagnostics) |
| L2\_Malaise | Malaise (level 2)(number of diagnostics) |
| L2\_MaleGenitalInf | Male genital infection (level 2)(number of diagnostics) |
| L2\_Malnutrition | Malnutrition (level 2)(number of diagnostics) |
| L2\_MemoryLoss | Memory loss (level 2)(number of diagnostics) |
| L2\_Menopausal | Menopausal (level 2)(number of diagnostics) |
| L2\_MenstruationDisorder | Menstruation disorder (level 2)(number of diagnostics) |
| L2\_MentalMiscelanea | Mental miscelanea (level 2)(number of diagnostics) |
| L2\_MentalRetardation | Mental retardation (level 2)(number of diagnostics) |
| L2\_MetabolicMineral | Metabolic mineral (level 2)(number of diagnostics) |
| L2\_MetabolicOther | Metabolic other (level 2)(number of diagnostics) |
| L2\_Migraine | Migraine (level 2)(number of diagnostics) |
| L2\_MixedHyperlipidemia | Mixed hyperlipidemia (level 2)(number of diagnostics) |
| L2\_Mixuria | Mixuria (level 2)(number of diagnostics) |
| L2\_MoodDisorder | Mood disorder (level 2)(number of diagnostics) |
| L2\_MuscleSpasm | Muscle spasm (level 2)(number of diagnostics) |
| L2\_Myalgia | Myalgia (level 2)(number of diagnostics) |
| L2\_MycosisFoot | Mycosis foot (level 2)(number of diagnostics) |
| L2\_MycosisOther | Mycosis other (level 2)(number of diagnostics) |
| L2\_Myopathy | Myopathy (level 2)(number of diagnostics) |
| L2\_Nails | Nails disease (level 2)(number of diagnostics) |
| L2\_NasalCavity | Nasal cavity (level 2)(number of diagnostics) |
| L2\_NeoSkin | Neo skin (level 2)(number of diagnostics) |
| L2\_NervousSystem | Nervous system (level 2)(number of diagnostics) |
| L2\_NervousSystemOther | Nervous system other (level 2)(number of diagnostics) |
| L2\_NeurRadiculitis | Neuritis radiculitis (level 2)(number of diagnostics) |
| L2\_Nocturia | Nocturia (level 2)(number of diagnostics) |
| L2\_NonFollow | Non follow of treatment (level 2)(number of diagnostics) |
| L2\_Obesity | Obesity (level 2)(number of diagnostics) |
| L2\_OpenWound | Open wound (level 2)(number of diagnostics) |
| L2\_Oral | Oral (level 2)(number of diagnostics) |
| Feature | Description |
| L2\_Osteoarthrosis | Osteoarthrosis (level 2)(number of diagnostics) |
| L2\_Osteoporosis | Osteoporosis (level 2)(number of diagnostics) |
| L2\_Otalgia | Otalgia (level 2)(number of diagnostics) |
| L2\_OtherInfection | Other infection (level 2)(number of diagnostics) |
| L2\_OtitisMedia | Otitis media (level 2)(number of diagnostics) |
| L2\_OvarianCyst | Ovarian cyst (level 2)(number of diagnostics) |
| L2\_Ovaries | Ovaries (level 2)(number of diagnostics) |
| L2\_Overweight | Overweight (level 2)(number of diagnostics) |
| L2\_PainJoint | Pain joint (level 2)(number of diagnostics) |
| L2\_Pancreas | Pancreas (level 2)(number of diagnostics) |
| L2\_ParasiteInfection | Parasite infection (level 2)(number of diagnostics) |
| L2\_Parkinson | Parkinson (level 2)(number of diagnostics) |
| L2\_PerEndMiocard | Pericardic endocardic miocardic (level 2)(number of diagnostics) |
| L2\_PeriphNeuropathy | Peripheral neuropathy (level 2)(number of diagnostics) |
| L2\_Pneumonia | Pneumonia (level 2)(number of diagnostics) |
| L2\_PNS | Peripheral nervous system (level 2)(number of diagnostics) |
| L2\_PNSFace | Nervous system face (level 2)(number of diagnostics) |
| L2\_PNSInflamToxic | PNS Inflamatory or Toxic (level 2)(number of diagnostics) |
| L2\_Polydipsia | Polydipsia (level 2)(number of diagnostics) |
| L2\_Polyuria | Polyuria (level 2)(number of diagnostics) |
| L2\_Proctology | Proctology (level 2)(number of diagnostics) |
| L2\_ProstateBenign | Prostate benign (level 2)(number of diagnostics) |
| L2\_Proteinuria | Proteinuria (level 2)(number of diagnostics) |
| L2\_PSA | PSA (level 2)(number of diagnostics) |
| L2\_Psoriasis | Psoriasis (level 2)(number of diagnostics) |
| L2\_Pyrosis | Pyrosis (level 2)(number of diagnostics) |
| L2\_Renal | Renal (level 2)(number of diagnostics) |
| L2\_RespiratoryFail | Respiratory fail (level 2)(number of diagnostics) |
| L2\_RespUpAcute | Respiratory upper chronic (level 2)(number of diagnostics) |
| L2\_RespUpChronic | Respiratory upper acute (level 2)(number of diagnostics) |
| L2\_RheumaticPain | Rheumatic pain (level 2)(number of diagnostics) |
| L2\_Rhinitis | Rhinitis (level 2)(number of diagnostics) |
| L2\_RLS | Restless legs syndrome (level 2)(number of diagnostics) |
| L2\_Schizophrenic | Schizophrenic (level 2)(number of diagnostics) |
| L2\_SciaticPain | Sciatic pain (level 2)(number of diagnostics) |
| L2\_Screening | Screening (level 2)(number of diagnostics) |
| L2\_SeborrheicCyst | Seborrheic cyst (level 2)(number of diagnostics) |
| L2\_SeborrheicDermatitis | Seborrheic dermatitis (level 2)(number of diagnostics) |
| L2\_SexDysfunction | Sex dysfunction (level 2)(number of diagnostics) |
| L2\_Shoulder | Shoulder (level 2)(number of diagnostics) |
| L2\_Sinus | Sinus (level 2)(number of diagnostics) |
| L2\_SkinAtrophic | Skin atrophic (level 2)(number of diagnostics) |
| L2\_SkinInflammatory | Skin inflammatory (level 2)(number of diagnostics) |
| Feature | Description |
| L2\_SkinKeratosis | Skin keratosis (level 2)(number of diagnostics) |
| L2\_SkinOther | Skin other (level 2)(number of diagnostics) |
| L2\_SkinSense | Skin sense (level 2)(number of diagnostics) |
| L2\_SkinViric | Skin viric (level 2)(number of diagnostics) |
| L2\_Sleep | Sleep (level 2)(number of diagnostics) |
| L2\_SleepApnea | Sleep apnea (level 2)(number of diagnostics) |
| L2\_SpinalStenosis | Spinal stenosis (level 2)(number of diagnostics) |
| L2\_Spondylosis | Spondylosis (level 2)(number of diagnostics) |
| L2\_Sprains | Sprains (level 2)(number of diagnostics) |
| L2\_Stress | Stress (level 2)(number of diagnostics) |
| L2\_SuperficialInjury | Superficial injury (level 2)(number of diagnostics) |
| L2\_SweatGland | Sweat gland (level 2)(number of diagnostics) |
| L2\_Syncope | Syncope (level 2)(number of diagnostics) |
| L2\_SyndromeX | Syndrome X (level 2)(number of diagnostics) |
| L2\_SynoviumTendon | Synovium and tendon (level 2)(number of diagnostics) |
| L2\_Teeth | Teeth (level 2)(number of diagnostics) |
| L2\_Temporomandibular | Temporomandibular (level 2)(number of diagnostics) |
| L2\_TensionalPsicalgia | Tensional psicalgia (level 2)(number of diagnostics) |
| L2\_TesticularDysfunction | Testicular dysfunction (level 2)(number of diagnostics) |
| L2\_ThyroidOther | Thyroid other (level 2)(number of diagnostics) |
| L2\_Tobacco | Tobacco (level 2)(number of diagnostics) |
| L2\_TonsilsAdenoids | Tonsils and adenoids (level 2)(number of diagnostics) |
| L2\_TriggerFinger | Trigger finger (level 2)(number of diagnostics) |
| L2\_Ulcer | Ulcer (level 2)(number of diagnostics) |
| L2\_UlcerousColitis | Ulcerous colitis (level 2)(number of diagnostics) |
| L2\_UrinaryIncontinency | Urinary incontinency (level 2)(number of diagnostics) |
| L2\_UrinaryUrgency | Urinary urgency (level 2)(number of diagnostics) |
| L2\_Urticaria | Urticaria (level 2)(number of diagnostics) |
| L2\_UTIOther | UTI Other (level 2)(number of diagnostics) |
| L2\_Vaccine | Vaccine (level 2)(number of diagnostics) |
| L2\_Vaginal | Vaginal (level 2)(number of diagnostics) |
| L2\_VarixLeg | Varix leg (level 2)(number of diagnostics) |
| L2\_VascularPeripheral | Vascular peripheral (level 2)(number of diagnostics) |
| L2\_VertiginousSyndromes | Vertiginous syndromes (level 2)(number of diagnostics) |
| L2\_VIH | Vih (level 2)(number of diagnostics) |
| L2\_VitaminB | Vitamin B (level 2)(number of diagnostics) |
| L2\_VitaminD | Vitamin D (level 2)(number of diagnostics) |
| L2\_VomitingNausea | Vomiting and nausea (level 2)(number of diagnostics) |
| L2\_WeightGain | Weight gain (level 2)(number of diagnostics) |
| L2\_WeightLoss | Weight loss (level 2)(number of diagnostics) |
| L1\_Anemia | Anemia (level 1)(number of diagnostics) |
| L1\_Back | Back (level 1)(number of diagnostics) |
| L1\_ColitisMSC | Colitis miscelanea (level 1)(number of diagnostics) |
| L1\_ConnectiveTissue | Connective tissue (level 1)(number of diagnostics) |
| Feature | Description |
| L1\_Dermatitis | Dermatitis (level 1)(number of diagnostics) |
| L1\_Digestive | Digestive (level 1)(number of diagnostics) |
| L1\_DMSymptom | DM Symptom (level 1)(number of diagnostics) |
| L1\_Glucose | Glucose (level 1)(number of diagnostics) |
| L1\_Headache | Headache (level 1)(number of diagnostics) |
| L1\_Herpes | Herpes (level 1)(number of diagnostics) |
| L1\_Hormonal | Hormonal (level 1)(number of diagnostics) |
| L1\_Insomnia | Insomnia (level 1)(number of diagnostics) |
| L1\_IschemicCardiopathy | Ischemic cardiopathy (level 1)(number of diagnostics) |
| L1\_MycosisFoot | Mycosis foot (level 1)(number of diagnostics) |
| L1\_NervousSystemOther | Nervous system other (level 1)(number of diagnostics) |
| L1\_Prostate | Prostate (level 1)(number of diagnostics) |
| L1\_RespInfec | Respinfec (level 1)(number of diagnostics) |
| L1\_Skin | Skin (level 1)(number of diagnostics) |
| L1\_SkinInfection | Skin infection (level 1)(number of diagnostics) |
| L1\_SkinInflammatory | Skin inflammatory (level 1)(number of diagnostics) |
| L1\_SkinOther | Skin other (level 1)(number of diagnostics) |
| StartYear | Oldest start year of all diagnostics |
| PreviousYear | Number of diagnostics with start year < 2009 |
| DiffDMOtherRelatedSymptoms | Number of different level1 with level2=DMSymptom |
| VisitsDMOtherRelatedSymptoms | Visits with diagnostics with level2=DMSymptom |
| DiffLevel2Diag | Different level 2 categories with diagnostics |
| Tot3digitICD9 | Different 3 digit icd9 categories with diagnostics |
| TotDiagWYear | Total diagnostics per weighted year |
| TotDiag | Total diagnostics |
| TotDiagPerVisit | Total diagnostics per visit |
| TotDiagWOTranscript | Total diagnostics without transcripts |
| acetaminophen\_bin | Acetaminophen (binary) |
| acetaminophen\_dose | Acetaminophen (max dose) |
| acetaminophen\_npr | Acetaminophen (prescriptions) |
| ACEI\_bin | ACEI (binary) |
| benazepril\_bin | Benazepril (binary) |
| benazepril\_dose | Benazepril (max dose) |
| enalapril\_bin | Enalapril (binary) |
| enalapril\_dose | Enalapril (max dose) |
| ramipril\_bin | Ramipril (binary) |
| ramipril\_dose | Ramipril (max dose) |
| lisinopril\_bin | Lisinopril (binary) |
| lisinopril\_dose | Lisinopril (max dose) |
| ACEI\_dose | ACEI (max dose) |
| ACEI\_npr | ACEI (prescriptions) |
| aspirin\_bin | Aspirin (binary) |
| aspirin\_dose | Aspirin (max dose) |
| aspirin\_npr | Aspirin (prescriptions) |
| Feature | Description |
| losartan\_bin | Losartan (binary) |
| olmesartan\_bin | Olmesartan (binary) |
| telmisartan\_bin | Telmisartan (binary) |
| valsartan\_bin | Valsartan (binary) |
| losartan\_npr | Losartan (prescriptions) |
| olmesartan\_npr | Olmesartan (prescriptions) |
| telmisartan\_npr | Telmisartan (prescriptions) |
| valsartan\_npr | Valsartan (prescriptions) |
| losartan\_dose | Losartan (max dose) |
| olmesartan\_dose | Olmesartan (max dose) |
| telmisartan\_dose | Telmisartan (max dose) |
| valsartan\_dose | Valsartan (max dose) |
| AIIRA\_nap | AIIRA (different active principles) |
| AIIRA\_bin | AIIRA (binary) |
| losartan\_telmisartan\_dose | Losartan (max dose) |
| AIIRA\_npr | AIIRA (prescriptions) |
| antifungical\_general\_npr | Antifungal (prescriptions) |
| antifungical\_general\_nap | Antifungal (different active principles) |
| antifungical\_general\_bin | Antifungal (binary) |
| antifungical\_topical\_npr | Antifungal (prescriptions) |
| antifungical\_topical\_nap | Antifungal (different active principles) |
| antifungical\_topical\_bin | Antifungal (binary) |
| antiplatelet\_bin | Antiplatelet (binary) |
| antiplatelet\_dose | Antiplatelet (max dose) |
| antiplatelet\_npr | Antiplatelet (prescriptions) |
| alprazolam\_bin | Alprazolam (binary) |
| alprazolam\_dose | Alprazolam (max dose) |
| clonazepam\_bin | Clonazepam (binary) |
| clonazepam\_dose | Clonazepam (max dose) |
| diazepam\_bin | Diazepam (binary) |
| diazepam\_dose | Diazepam (max dose) |
| lorazepam\_bin | Lorazepam (binary) |
| lorazepam\_dose | Lorazepam (max dose) |
| temazepam\_bin | Temazepam (binary) |
| temazepam\_dose | Temazepam (max dose) |
| benzodiazepine\_nap | Benzodiazepine (different active principles) |
| benzodiazepine\_bin | Benzodiazepine (binary) |
| benzodiazepine\_npr | Benzodiazepine (prescriptions) |
| atenolol\_bin | Atenolol (binary) |
| atenolol\_npr | Atenolol (prescriptions) |
| atenolol\_dose | Atenolol (max dose) |
| carvedilol\_bin | Carvedilol (binary) |
| carvedilol\_npr | Carvedilol (prescriptions) |
| carvedilol\_dose | Carvedilol (max dose) |
| Feature | Description |
| metoprolol\_bin | Metoprolol (binary) |
| metoprolol\_dose | Metoprolol (max dose) |
| nebivolol\_bin | Nebivolol (binary) |
| nebivolol\_dose | Nebivolol (max dose) |
| propranolol\_bin | Propranolol (binary) |
| propranolol\_dose | Propranolol (max dose) |
| betablocker\_nap | Betablocker (different active principles) |
| betablocker\_bin | Betablocker (binary) |
| betablocker\_npr | Betablocker (prescriptions) |
| fenofibrate\_dose | Fenofibrate (max dose) |
| fenofibric\_acid\_dose | Fenofibric (max dose) |
| gemfibrozil\_dose | Gemfibrozil (max dose) |
| fenofibrate\_bin | Fenofibrate (binary) |
| fenofibric\_acid\_bin | Fenofibric (binary) |
| gemfibrozil\_bin | Gemfibrozil (binary) |
| fibrate\_npr | Fibrate (prescriptions) |
| fibrate\_bin | Fibrate (binary) |
| fibrate\_nap | Fibrate (different active principles) |
| gabapentin\_dose | Gabapentin (max dose) |
| pregabalin\_dose | Pregabalin (max dose) |
| gabapentin\_bin | Gabapentin (binary) |
| pregabalin\_bin | Pregabalin (binary) |
| GABA\_npr | GABA (prescriptions) |
| GABA\_bin | GABA (binary) |
| gastroparesia\_bin | Gastroparesia drugs (binary) |
| gastroparesia\_dose | Gastroparesia drugs (max dose) |
| gastroparesia\_npr | Gastroparesia drugs (prescriptions) |
| glucocorticoid\_injec\_bin | Glucocorticoid injectable (binary) |
| glucocorticoid\_injec\_npr | Glucocorticoid injectable (prescriptions) |
| glucocorticoid\_local\_npr | Glucocorticoid local (prescriptions) |
| glucocorticoid\_local\_nap | Glucocorticoid local (different active principles) |
| glucocorticoid\_local\_bin | Glucocorticoid local (binary) |
| betamethasone\_local\_npr | Betamethasone (prescriptions) |
| betamethasone\_local\_bin | Betamethasone (binary) |
| hydrocortisone\_local\_npr | Hydrocortisone (prescriptions) |
| hydrocortisone\_local\_bin | Hydrocortisone (binary) |
| triamcinolone\_local\_npr | Triamcinolone (prescriptions) |
| triamcinolone\_local\_bin | Triamcinolone (binary) |
| glucocorticoid\_general\_npr | Glucocorticoid (prescriptions) |
| glucocorticoid\_general\_nap | Glucocorticoid (different active principles) |
| glucocorticoid\_general\_bin | Glucocorticoid (binary) |
| hydrocortisone\_general\_dose | Hydrocortisone (max dose) |
| hydrocortisone\_general\_bin | Hydrocortisone (binary) |
| hydrocortisone\_general\_npr | Hydrocortisone (prescriptions) |
| Feature | Description |
| methylprednisolone\_general\_dose | Methylprednisolone (max dose) |
| methylprednisolone\_general\_bin | Methylprednisolone (binary) |
| methylprednisolone\_general\_npr | Methylprednisolone (prescriptions) |
| prednisone\_general\_dose | Prednisone (max dose) |
| prednisone\_general\_bin | Prednisone (binary) |
| prednisone\_general\_npr | Prednisone (prescriptions) |
| LtypeCaChB\_nap | L type calcium channel blocker (different active principles) |
| LtypeCaChB\_bin | L type calcium channel blocker (binary) |
| LtypeCaChB\_npr | L type calcium channel blocker (prescriptions) |
| amlodipine\_dose | Amlodipine (max dose) |
| amlodipine\_bin | Amlodipine (binary) |
| protonpumpinhibitor\_nap | Proton pump inhibitor (different active principles) |
| protonpumpinhibitor\_bin | Proton pump inhibitor (binary) |
| protonpumpinhibitor\_npr | Proton pump inhibitor (prescriptions) |
| statin\_nap | Statin (different active principles) |
| statin\_npr | Statin (prescriptions) |
| statin\_bin | Statin (binary) |
| atorvastatin\_dose | Atorvastatin (max dose) |
| lovastatin\_dose | Lovastatin (max dose) |
| pravastatin\_dose | Pravastatin (max dose) |
| rosuvastatin\_dose | Rosuvastatin (max dose) |
| simvastatin\_dose | Simvastatin (max dose) |
| atorvastatin\_bin | Atorvastatin (binary) |
| lovastatin\_bin | Lovastatin (binary) |
| pravastatin\_bin | Pravastatin (binary) |
| rosuvastatin\_bin | Rosuvastatin (binary) |
| simvastatin\_bin | Simvastatin (binary) |
| statin\_dose\_adjusted | Statin |
| thiazide\_nap | Thiazide (different active principles) |
| thiazide\_bin | Thiazide (binary) |
| thiazide\_npr | Thiazide (prescriptions) |
| hydrochlorothiazide\_dose | Hydrochlorothiazide (max dose) |
| niacin\_npr | Niacin (prescriptions) |
| niacin\_bin | Niacin (binary) |
| ezetimibe\_npr | Ezetimibe (prescriptions) |
| ezetimibe\_bin | Ezetimibe (binary) |
| antilipid\_npr | Antilipemics (prescriptions) |
| antilipid\_bin | Antilipemics (binary) |
| nasal\_npr | Nasal route (prescriptions) |
| inhalation\_npr | Inhalation route (prescriptions) |
| otic\_nmed | Otic route (number of different drugs) |
| otic\_nap | Otic route (different active principles) |
| otic\_bin | Otic route (binary) |
| otic\_npr | Otic route (prescriptions) |
| Feature | Description |
| allergy\_inflam\_opht\_nap | Ophtalmic allergy inflamatory (different active principles) |
| allergy\_inflam\_opht\_bin | Ophtalmic allergy inflamatory (binary) |
| allergy\_inflam\_opht\_npr | Ophtalmic allergy inflamatory (prescriptions) |
| antibiotic\_opht\_nap | Ophtalmic antibiotic (different active principles) |
| antibiotic\_opht\_bin | Ophtalmic antibiotic (binary) |
| antibiotic\_opht\_npr | Ophtalmic antibiotic (prescriptions) |
| glaucoma\_opht\_nap | Ophtalmic glaucoma (different active principles) |
| glaucoma\_opht\_bin | Ophtalmic glaucoma (binary) |
| glaucoma\_opht\_npr | Ophtalmic glaucoma (prescriptions) |
| oral\_contracep\_npr | Oral contraceptive (prescriptions) |
| oral\_contracep\_bin | Oral contraceptive (binary) |
| no\_oral\_contracep\_npr | No oral contraceptive (prescriptions) |
| no\_oral\_contracep\_bin | No oral contraceptive (binary) |
| alfa1blocker\_npr | Alfa1 blocker (prescriptions) |
| calcium\_npr | Calcium (prescriptions) |
| calcium\_bin | Calcium (binary) |
| coxib\_npr | Coxib (prescriptions) |
| digoxin\_npr | Digoxin (prescriptions) |
| laxatives\_npr | Laxatives (prescriptions) |
| laxatives\_bin | Laxatives (binary) |
| fiveareductase\_npr | 5a-reductase (prescriptions) |
| iron\_npr | Iron (prescriptions) |
| loopdiuretic\_bin | Loop diuretic (binary) |
| Kdiuretic\_npr | K-diuteric (prescriptions) |
| Kdiuretic\_bin | K-diuretic (binary) |
| isosorbide\_npr | Isosorbide (prescriptions) |
| omega\_3\_npr | Omega-3 (prescriptions) |
| potassium\_chloride\_bin | Potassium chloride (binary) |
| isopropyl\_npr | Isopropyl (prescriptions) |
| null\_npr | Null medication description (prescriptions) |
| osteoporosis\_bin | Osteoporosis drugs (binary) |
| alkylamine\_bin | Alkylamine (binary) |
| amphetamine\_npr | Amphetamine (prescriptions) |
| BPOther\_npr | Blood preassure other drugs (prescriptions) |
| BPOther\_bin | Blood preassure other drugs (binary) |
| triptan\_npr | Triptan (prescriptions) |
| corticosteroids\_npr | Corticosteroids (prescriptions) |
| gout\_npr | Gout drugs (prescriptions) |
| gout\_bin | Gout drugs (binary) |
| vitamin\_bin | Vitamin (binary) |
| active\_principle | Total active principles |
| prescripts | Total prescriptions |
| medication\_wo\_prescript | Medication without prescription |
| nomedication | No medication at all |
| Feature | Description |
| diag\_3digit\_with\_medication | Number of 3 digit groups of diagnostics with medication |
| medication\_diag\_zero | Medication with diagnostic "0000" |
| medication\_diag\_NA | Medication with diagnostic NA |
| dm\_npr | Miglitol or acarbose (prescriptions) |
| HasLabWithNA | Has lab panel without lab observation |
| HasLab | Has lab panel |
| STATE\_CA | State CA (binary) |
| STATE\_TX | State TX (binary) |
| STATE\_FL | State FL (binary) |
| STATE\_MO | State MO (binary) |
| STATE\_NJ | State NJ (binary) |
| STATE\_NY | State NY (binary) |
| STATE\_OH | State OH (binary) |
| STATE\_NV | State NV (binary) |
| STATE\_VA | State VA (binary) |

**Refers to files drugsap.csv, icd9.csv and DMfeatureCreation.R files for more details.**