

Analytics with **Alteryx**: (Failing at) Classifying **Diabetes Readmission**



IDS.506 - Robert Minh Duc Bui - mbui7@uic.edu

Executive Summary



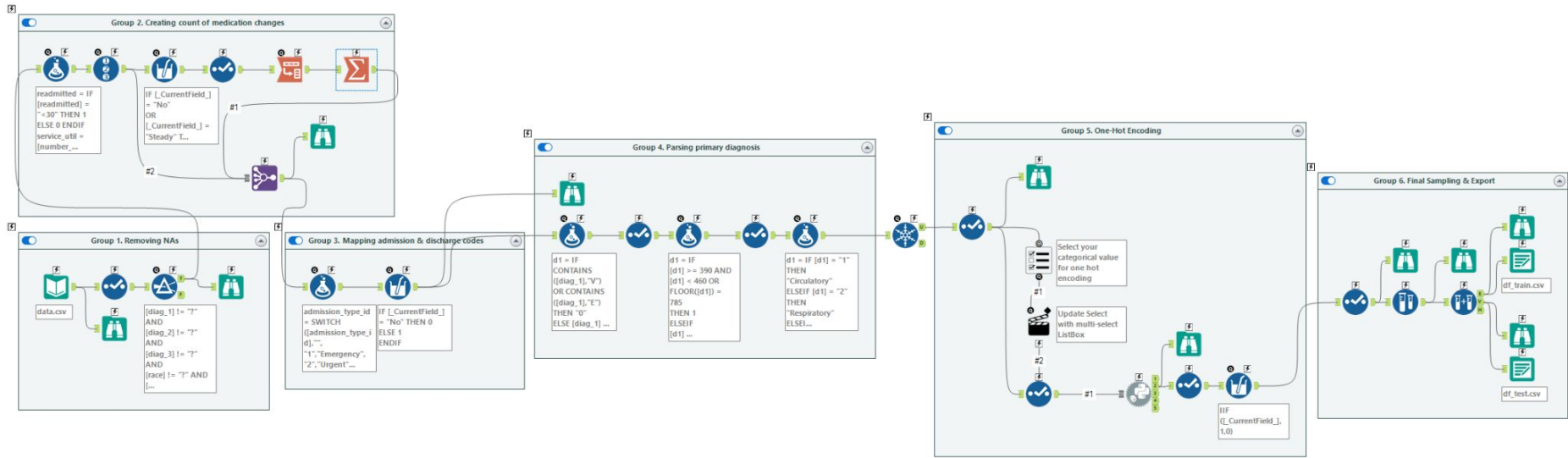
- While can perform ingestion, engineering, and preliminary modeling using **Alteryx** to classify **diabetes** patients' **readmission** under **30** days, due to technical limitations we cannot assess any model performance. Additionally, model training is extremely slow.
- **Candidate** models:
 - **Logistic Regression** (built)
 - **Gradient Boosted Model** (built)
 - **Random Forest** (built)
 - L/M/R/S Support Vector Machine (failure)
 - Single Layer Perceptron (failure)
- Final models at-a-glance:
 - Performance **unassessed**. **Model Comparison Tool** has non-zero exit code.

Research **Mission**

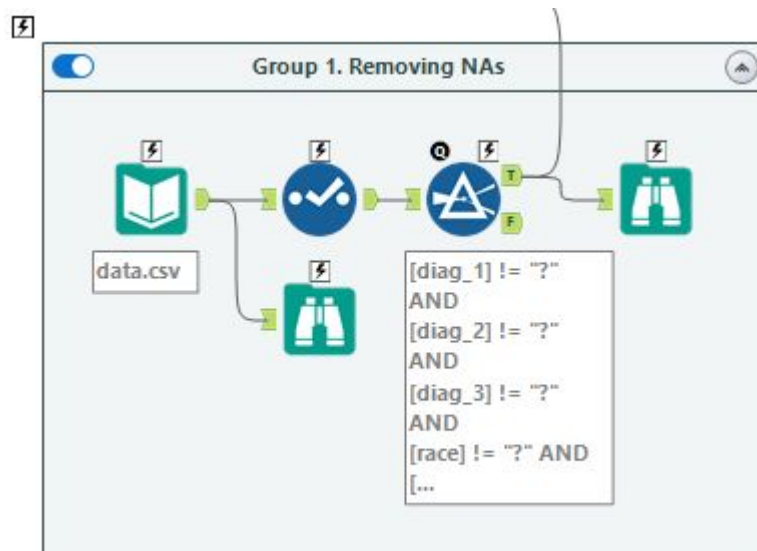


- To develop a model that:
 - Uses aggregated claims data from medical providers to:
 - **Classify** patients at risk of readmission under 30 days of discharge.

Data **Processing** Workflow

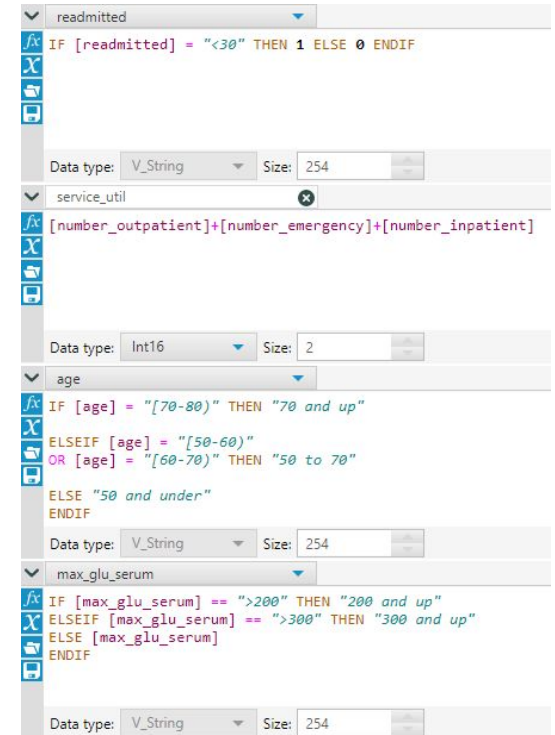
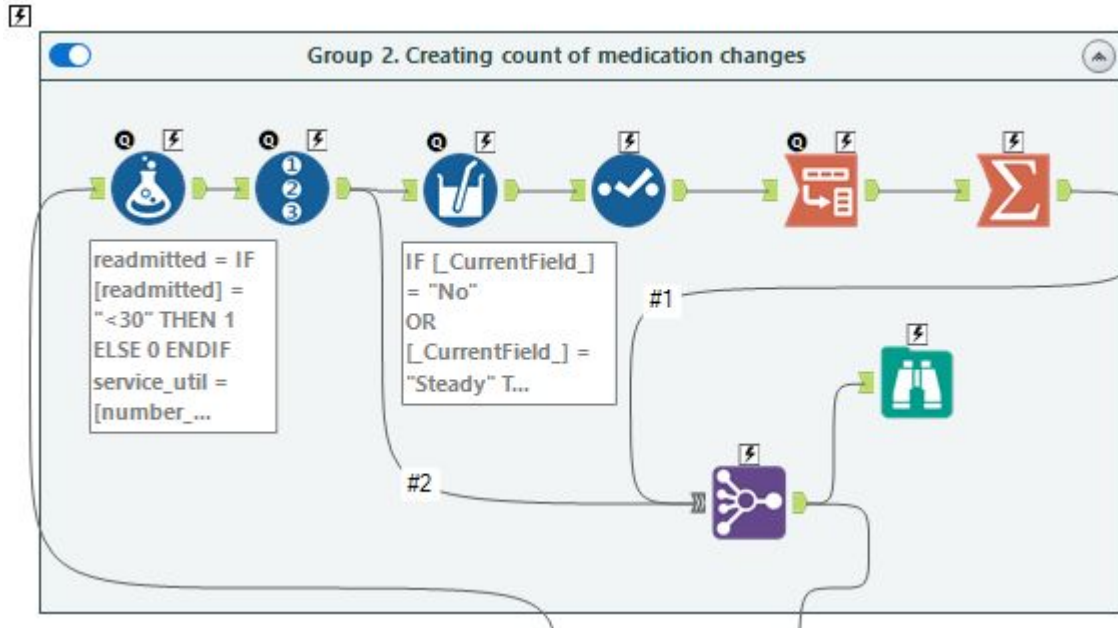


Data **Processing** Workflow #1

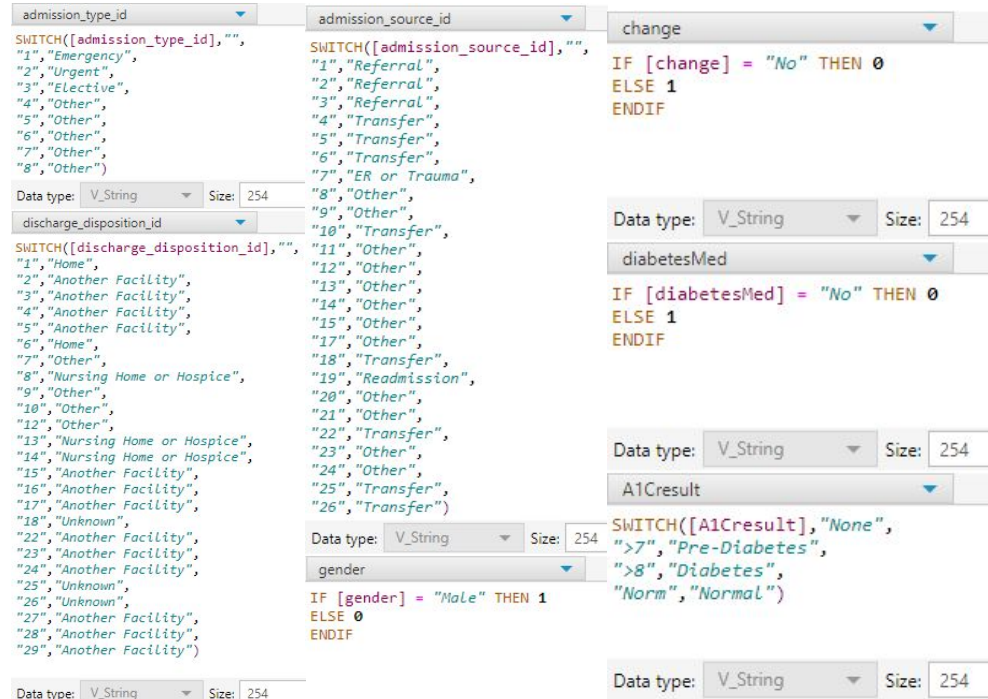
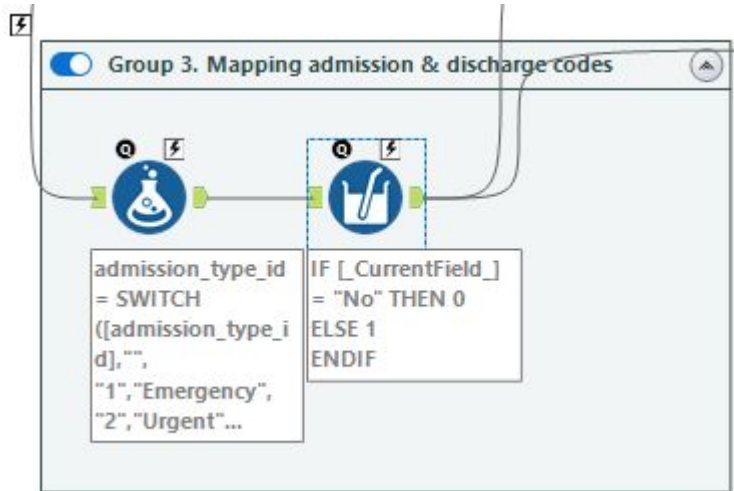


```
[diag_1] != "?" AND  
[diag_2] != "?" AND  
[diag_3] != "?" AND  
[race] != "?" AND  
[gender] != "Unknown/Invalid" AND  
[discharge_disposition_id] != "11" AND  
[discharge_disposition_id] != "19" AND  
[discharge_disposition_id] != "20" AND  
[discharge_disposition_id] != "21"
```

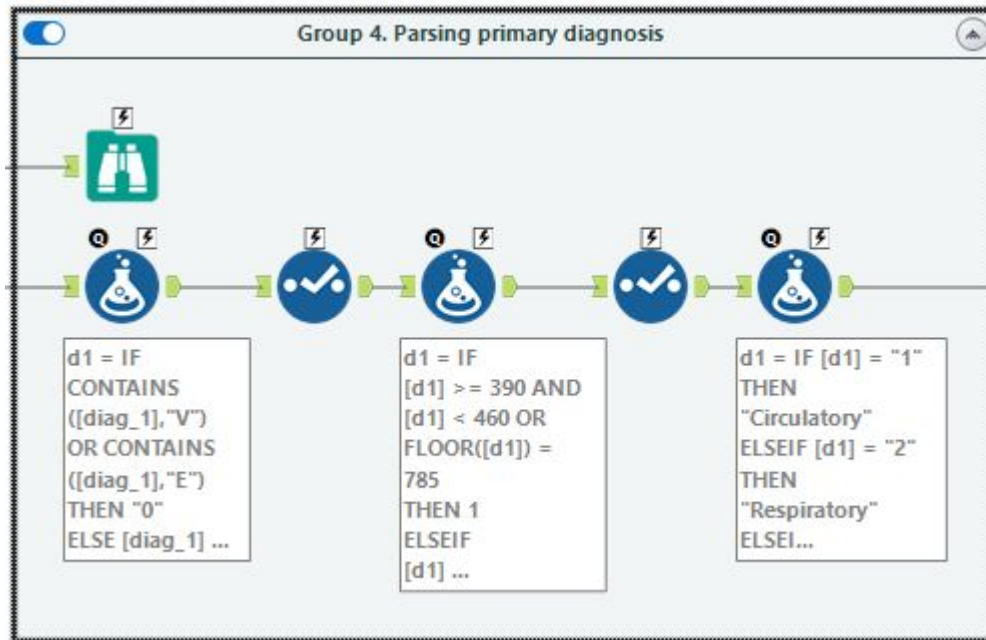
Data Processing Workflow #2



Data Processing Workflow #3



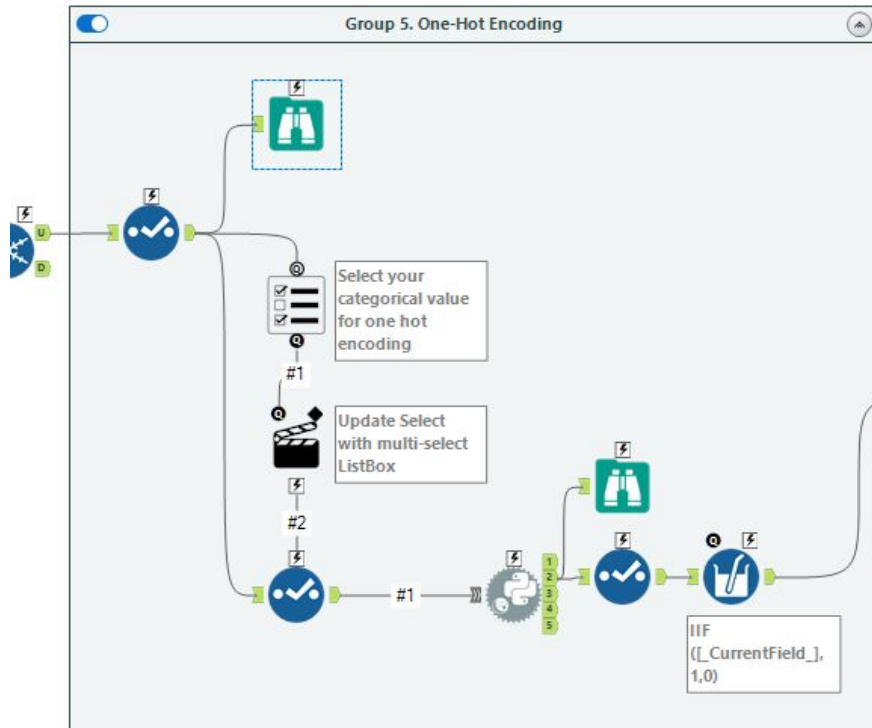
Data Processing Workflow #4



d1

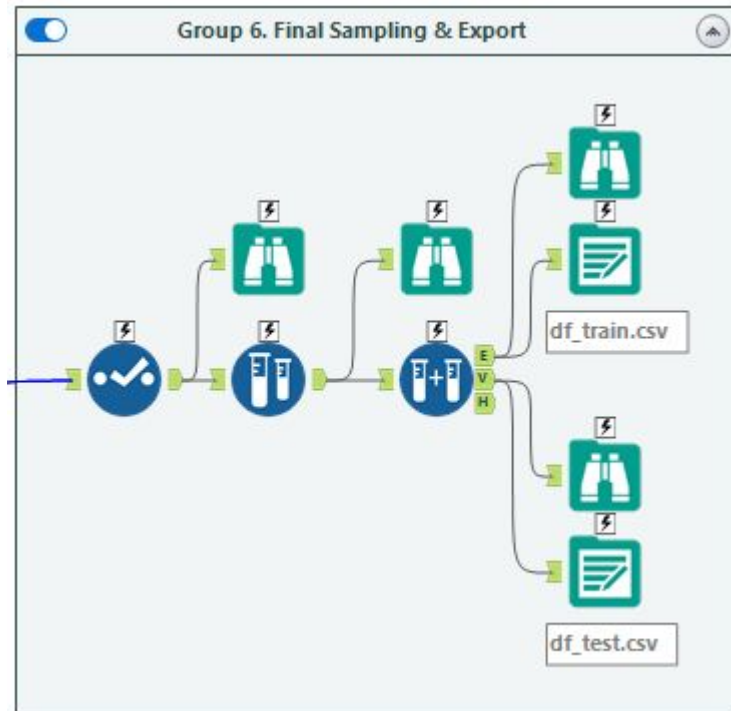
```
IF [d1] = "1" THEN "Circulatory"  
ELSEIF [d1] = "2" THEN "Respiratory"  
ELSEIF [d1] = "3" THEN "Digestive"  
ELSEIF [d1] = "4" THEN "Diabetes"  
ELSEIF [d1] = "5" THEN "Injury"  
ELSEIF [d1] = "6" THEN "Musculoskeletal"  
ELSEIF [d1] = "7" THEN "Genitourinary"  
ELSEIF [d1] = "8" THEN "Neoplasms"  
ELSE "Other"  
ENDIF
```


Data Processing Workflow #5



- ☒ race_AfricanAmerican
- ☒ race_Asian
- ☒ race_Hispanic
- ☒ race_Other
- ☒ age_50 and under
- ☒ age_50 to 70
- ☒ age_70 and up
- ☒ admission_type_id_Emergency
- ☒ admission_type_id_Other
- ☒ admission_type_id_Urgent
- ☒ discharge_disposition_id_Another Facility
- ☒ discharge_disposition_id_Nursing Home or Hospice
- ☒ discharge_disposition_id_Other
- ☒ discharge_disposition_id_Unknown
- ☒ admission_source_id_ER or Trauma
- ☒ admission_source_id_Other
- ☒ admission_source_id_Transfer
- ☒ d1_Diabetes
- ☒ d1_Digestive
- ☒ d1_Genitourinary
- ☒ d1_Injury
- ☒ d1_Musculoskeletal
- ☒ d1_Neoplasms
- ☒ d1_Other
- ☒ d1_Respiratory
- ☒ max_glu_serum_200 and up
- ☒ max_glu_serum_300 and up
- ☒ max_glu_serum_None
- ☒ A1Cresult_Diabetes
- ☒ A1Cresult_None
- ☒ A1Cresult_Pre-Diabetes
- ☐ Dynamic or Unknown Fields

Data **Processing** Workflow #6



Record allocation

Estimation sample percent

80

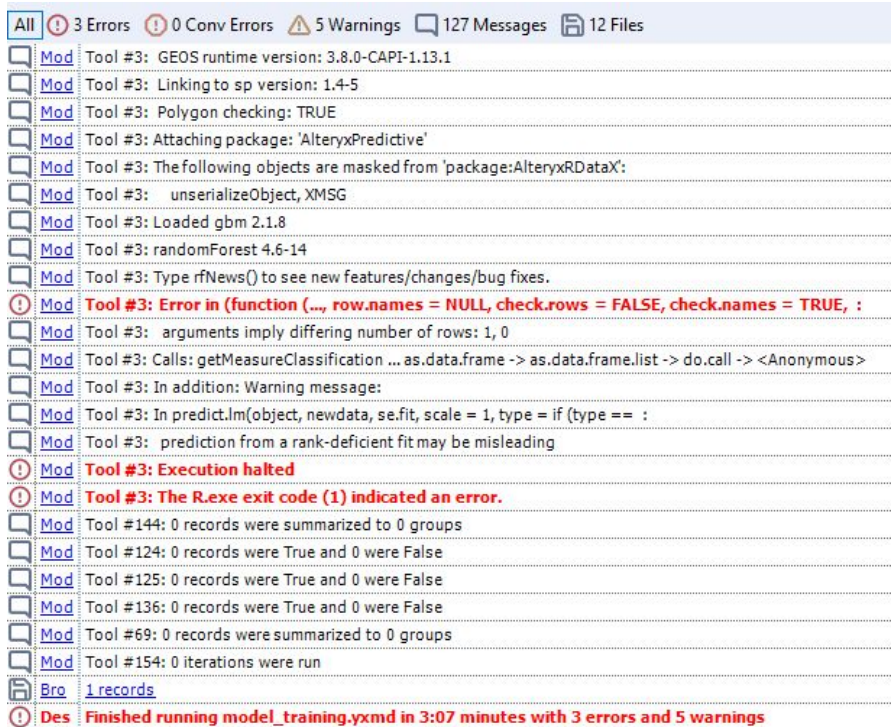
Validation sample percent

20

The total of the estimation and validation percentages should be less than or equal to 100. If the sum is less than 100, then the residual percentage is placed in the holdout sample. Using the default settings, 34% of the records are in the estimation sample while the validation and holdout sample will have 33% of the data records each.

Random seed

1



 Des Finished running model training.yxmd in 3:07 minutes with 3 errors and 5 warnings

(Unsuccessful) Remedial Steps Taken

The screenshot shows the 'Runtime' tab of the Alteryx configuration window. The following settings are highlighted with red boxes:

- Memory Limit:** ☒ Use Specific Amount: 65536 Megabytes
- Temporary files:** ☒ Use Global Default
- Conversion Errors:** ☐ Limit Conversion Errors (Maximum Errors per Location: 10), ☐ Stop Processing When Limit is Reached
- Predictive Tools Code Page:** Western European (CP 1252)
- Record Limit for All Inputs:** No Limit
- ☐ Cancel Running Workflow on Error
- ☐ Disable All Browse Tools
- ☒ Show All Macro Messages
- ☐ Disable All Tools that Write Output
- ☐ Enable Performance Profiling
- ☒ Use AMP Engine
- ☐ Engine compatibility mode

- Problem: **Unclear Stacktrace**
 - Ticked “**Show All Macro Messages**” to expand stack trace
 - **Result:** unclear what the error is, and cannot troubleshoot.
- Problem: **Very Long Execution Time**
 - Extended **RAM** allocation to maximum possible w/o pagefile.sys
 - Enabled **AMP multithreading**
 - **Unparked** all cores (set alteryx.exe and services to **Priority 0** in Task Manager)
 - Elevating alteryx.exe to **Administrator** access
 - **Result:** sped up execution by x8.

LogReg coefs

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.928927	0.235625	-3.94240	8e-05 ***
gender	0.062626	0.047208	1.32659	0.18465
race_AfricanAmerican	-0.103997	0.062184	-1.67242	0.09444 .
race_Asian	0.117805	0.277625	0.42433	0.67132
race_Hispanic	-0.095965	0.163489	-0.58698	0.55722
race_Other	-0.506582	0.195630	-2.58949	0.00961 **
age_50.and.under	-0.191759	0.060323	-3.17889	0.00148 ***
age_50.to.70	-0.128265	0.059126	-2.16935	0.03006 *
admission_type_id_Emergency	-0.067026	0.097094	-0.69032	0.48999
admission_type_id_Other	0.130843	0.124486	1.05107	0.29323
admission_type_id_Urgent	0.128103	0.084083	1.52354	0.12762
admission_source_id_ER.or.Trauma	0.125788	0.082055	1.53298	0.12528
admission_source_id_Other	-0.047455	0.130770	-0.36289	0.71669
admission_source_id_Transfer	-0.188312	0.105680	-1.78191	0.07476 .
discharge_disposition_id_Other	0.438975	0.263177	1.66798	0.09532 .
discharge_disposition_id_Unknown	0.134033	0.115769	1.15777	0.24696
discharge_disposition_id_Another.Facility	0.654225	0.057881	11.30295	< 2.2e-16 ***
discharge_disposition_id_Nursing.Home.or.Hospice	-0.844400	0.288316	-2.92873	0.0034 **
d1_Diabetes	0.023127	0.098273	0.23533	0.81395
d1_Digestive	-0.276585	0.088621	-3.12100	0.0018 **
d1_Genitourinary	-0.064941	0.112817	-0.57563	0.56486
d1_Injury	0.006948	0.096092	0.07230	0.94236
d1_Musculoskeletal	-0.242063	0.115397	-2.09765	0.03594 *
d1_Neoplasms	-0.182996	0.128684	-1.42206	0.15501
d1_Other	-0.137587	0.072464	-1.89868	0.05761 .

d1_Respiratory	-0.364451	0.080128	-4.54833	1e-05 ***
max_glu_serum_None	-0.088815	0.166775	-0.53254	0.59435
max_glu_serum_200.and.up	-0.086042	0.246019	-0.34974	0.72653
max_glu_serum_300.and.up	-0.119280	0.283829	-0.42025	0.6743
A1Cresult_Diabetes	0.006556	0.132391	0.04952	0.96051
A1Cresult_None	0.186636	0.106892	1.74602	0.08081 .
A1Cresult_Pre.Diabetes	0.122825	0.153270	0.80137	0.42292
time_in_hospital	0.035970	0.009461	3.80204	0.00014 ***
num_lab_procedures	0.003221	0.001389	2.31818	0.02044 *
num_procedures	-0.022367	0.016071	-1.39181	0.16398
num_medications	0.003598	0.003771	0.95407	0.34005
number_diagnoses	0.024341	0.014072	1.72979	0.08367 .
change	-0.086353	0.082217	-1.05031	0.29358
diabetesMed	0.175243	0.084627	2.07078	0.03838 *
service_util	0.201408	0.015122	13.31893	< 2.2e-16 ***
drug_change	-0.006650	0.064092	-0.10375	0.91737
metformin	-0.151448	0.068865	-2.19919	0.02786 *
glimepiride	-0.018273	0.109713	-0.16656	0.86772
glipizide	0.108829	0.078706	1.38273	0.16675
glyburide	-0.032787	0.086439	-0.37930	0.70446
pioglitazone	-0.112811	0.095058	-1.18676	0.23532
rosiglitazone	-0.021812	0.101619	-0.21464	0.83005
insulin	0.115178	0.075794	1.51961	0.12861
age_70.and.up	NA	NA	NA	NA

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1