

# Logistic Regression for ICU Mortality Prediction

Code ▾

Load the libraries and install those we do not have in the system, and Set working directory to source file directory

Hide

```
pacman::p_load(tidyverse, caret, corrplot, caTools, car, ROCR, ggplot2,
               data.table, modelr, broom, relaimpo, ggpubr, moments,
               DMwR, pROC, dlookr, dplyr, ROSE)
```

Read the data files

Hide

```
data = read.csv("data_rescrub1.csv")
```

## Data Preparation and Exploration

Data summary - remove columns not needed (from prior analysis)

Hide

```
data <- subset(data,
               select = -c(ethnicity, icu_admit_source, icu_stay_type, apache_3j_bodysystem,
                           apache_2_bodysystem))
```

Define the factor variable types

Hide

```
data$comor_sum = data$aids + data$cirrhosis + data$diabetes_mellitus + data$hepatic_failure +
  data$immunosuppression + data$leukemia + data$lymphoma + data$solid_tumor_with_metastasis
```

Check the target variable - contrast

Hide

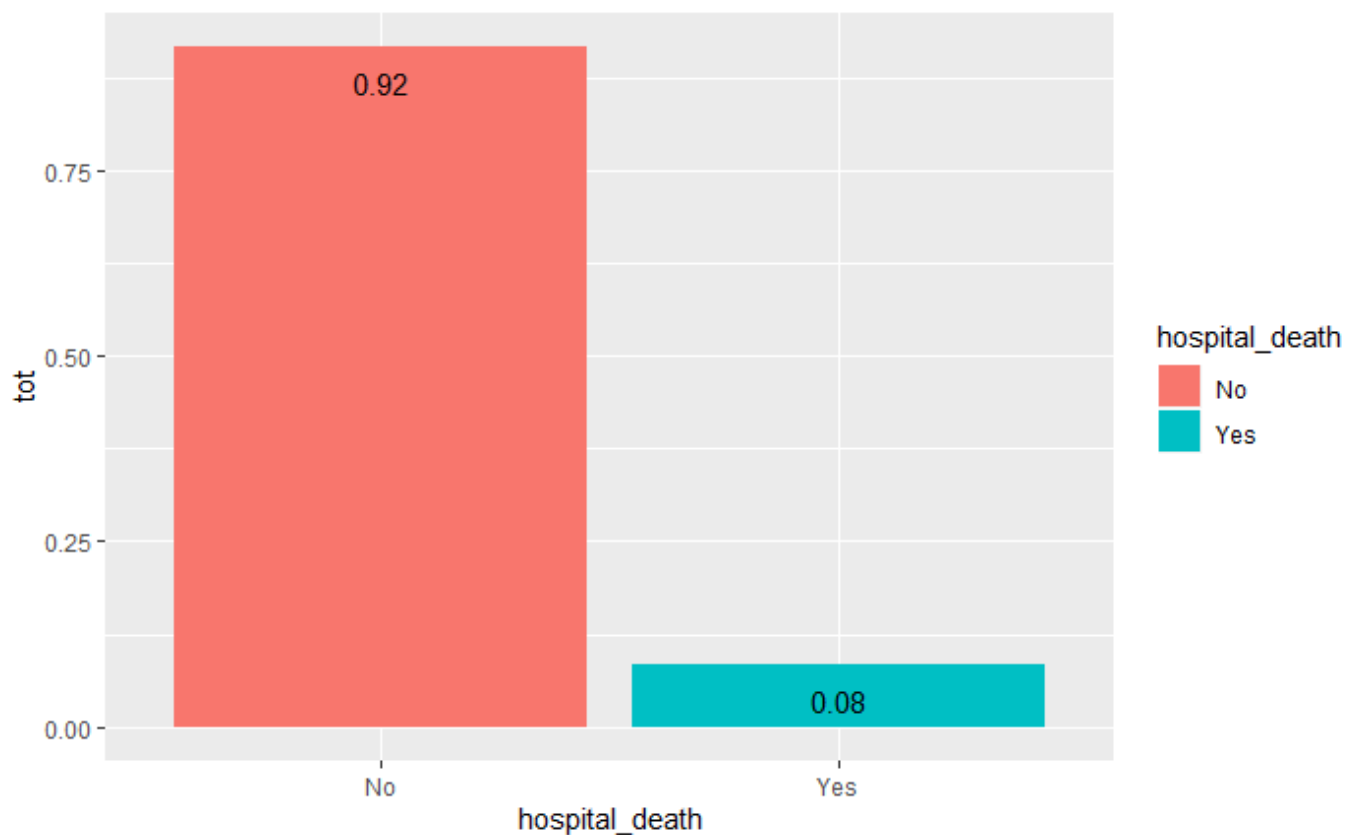
```
contrasts(data$hospital_death)
```

```
      Yes
No      0
Yes     1
```

- histogram

Hide

```
data %>%
  group_by(hospital_death)%>%
  summarise(tot = n()/nrow(data))%>%
  ggplot(aes(x=hospital_death, y=tot, fill = hospital_death)) +
  geom_bar(stat='identity') +
  geom_text(aes(label = round(tot, 2)), vjust = 2)
```



- Count

[Hide](#)

```
table(data$hospital_death)
```

```
   No   Yes
80200  7272
```

Split the data into train and test sets, excluding apache4a predictions

[Hide](#)

```
df <- subset(data,
              select = -c(encounter_id, apache_4a_hospital_death_prob, apache_4a_icu_death_prob))
set.seed(123)
splitData = sample.split(df$hospital_death, SplitRatio = 0.7)
train <- subset(df, splitData==T)
test <- subset(df, splitData==F)
nrow(train)/nrow(df)
```

```
[1] 0.6999954
```

Hide

```
nrow(test)/nrow(df)
```

```
[1] 0.3000046
```

Check the train set

Hide

```
table(train$hospital_death)
```

No	Yes
56140	5090

Hide

```
prop.table(table(train$hospital_death))
```

No	Yes
0.91687081	0.08312919

## Logistic Regression with original dataset

Initial model with iterations to remove aliasing and multi-collinearity between variables

Hide

```
model1 = glm(hospital_death ~ . -gcs_motor -gcs_verbal -gcs_eyes -diabetes_mellitus
             -hr_change -rr_change -temp_change -bun_change -cre_change -glu_change -hco3_cha
nge
             -hto_change -sodium_change -wbc_change -map_change -unk_cnt
             -a2_diag -apache_post_operative -hto_final -bun_final -cre_final -hemoglobin_fin
al
             -map_final -wbc_final,
             data = train, family = binomial)
summary(model1)
```

Call:

```
glm(formula = hospital_death ~ . - gcs_motor - gcs_verbal - gcs_eyes -
    diabetes_mellitus - hr_change - rr_change - temp_change -
    bun_change - cre_change - glu_change - hco3_change - hto_change -
    sodium_change - wbc_change - map_change - unk_cnt - a2_diag -
    apache_post_operative - hto_final - bun_final - cre_final -
    hemoglobin_final - map_final - wbc_final, family = binomial,
    data = train)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-3.1437	-0.3440	-0.2102	-0.1268	3.4384

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	1.338773	0.445261	3.007	0.002641	**
bmi	-0.003291	0.002225	-1.479	0.139043	
a3j_diag1	0.786092	0.272020	2.890	0.003855	**
a3j_diag2	0.725162	0.272143	2.665	0.007707	**
a3j_diag3	0.104310	0.278076	0.375	0.707575	
a3j_diag4	0.291629	0.287406	1.015	0.310251	
arf_apache1	0.307557	0.091543	3.360	0.000780	***
gcs_sum	-0.107036	0.004578	-23.378	< 2e-16	***
intubated_apache1	0.100851	0.046993	2.146	0.031868	*
ventilated_apache1	0.810592	0.046881	17.290	< 2e-16	***
map_risk	0.042481	0.003122	13.606	< 2e-16	***
age	0.034168	0.001295	26.389	< 2e-16	***
elective_surgery1	-1.002745	0.111500	-8.993	< 2e-16	***
genderM	0.029379	0.035377	0.830	0.406292	
icu_typeCCU-CTICU	-0.228014	0.096664	-2.359	0.018332	*
icu_typeCSICU	-0.437704	0.120155	-3.643	0.000270	***
icu_typeCTICU	-0.253354	0.120968	-2.094	0.036225	*
icu_typeMed-Surg ICU	-0.222468	0.075387	-2.951	0.003167	**
icu_typeMICU	-0.061292	0.088405	-0.693	0.488114	
icu_typeNeuro ICU	0.250884	0.095346	2.631	0.008506	**
icu_typeSICU	-0.006081	0.105718	-0.058	0.954128	
pre_icu_los_days	0.047907	0.004951	9.676	< 2e-16	***
aids1	0.676482	0.425296	1.591	0.111696	
cirrhosis1	0.692135	0.139164	4.974	6.57e-07	***
hepatic_failure1	0.460525	0.147641	3.119	0.001813	**
immunosuppression1	0.465394	0.100998	4.608	4.07e-06	***
leukemia1	0.199671	0.172753	1.156	0.247756	
lymphoma1	0.704866	0.202836	3.475	0.000511	***
solid_tumor_with_metastasis1	1.224271	0.102462	11.949	< 2e-16	***
comor_sum	-0.187085	0.045592	-4.103	4.07e-05	***
hr_final1	-1.153043	0.130162	-8.858	< 2e-16	***
hr_final2	-1.052532	0.096478	-10.910	< 2e-16	***
hr_final3	-0.910517	0.101681	-8.955	< 2e-16	***
hr_final4	-0.609943	0.100864	-6.047	1.47e-09	***
hr_final5	-0.479396	0.098162	-4.884	1.04e-06	***
hr_final6	-0.235393	0.109901	-2.142	0.032205	*
hr_final7	-0.055999	0.120234	-0.466	0.641393	
hr_finalUNK	-1.219109	0.578619	-2.107	0.035124	*
rr_final1	-1.681626	0.085256	-19.724	< 2e-16	***
rr_final2	-1.636679	0.088096	-18.578	< 2e-16	***
rr_final3	-1.584425	0.092259	-17.174	< 2e-16	***

rr_final4	-1.294663	0.083243	-15.553	< 2e-16	***
rr_final5	-1.216485	0.091035	-13.363	< 2e-16	***
rr_final6	-1.113173	0.093656	-11.886	< 2e-16	***
rr_final7	-1.347162	0.104539	-12.887	< 2e-16	***
rr_finalUNK	-1.638661	0.379690	-4.316	1.59e-05	***
temp_final1	0.391899	0.189329	2.070	0.038458	*
temp_final2	-0.081987	0.200201	-0.410	0.682156	
temp_final3	-0.480277	0.122479	-3.921	8.81e-05	***
temp_final4	-0.748580	0.098324	-7.613	2.67e-14	***
temp_final5	-1.120700	0.093426	-11.996	< 2e-16	***
temp_finalUNK	-0.598056	0.148164	-4.036	5.43e-05	***
spo2_final	-0.021554	0.001404	-15.357	< 2e-16	***
spo2_change1	0.101521	0.040601	2.500	0.012404	*
spo2_changeUNK	0.956018	0.288676	3.312	0.000927	***
glu_final1	0.033147	0.081552	0.406	0.684411	
glu_final2	-0.090986	0.080221	-1.134	0.256713	
glu_final3	0.356724	0.117238	3.043	0.002344	**
glu_final4	0.508669	0.163560	3.110	0.001871	**
glu_finalUNK	0.240810	0.137057	1.757	0.078916	.
hco3_final3	-0.154845	0.077979	-1.986	0.047065	*
hco3_final4	0.109830	0.083146	1.321	0.186522	
hco3_final5	0.450861	0.095328	4.730	2.25e-06	***
hco3_final6	0.738558	0.100505	7.349	2.00e-13	***
hco3_finalUNK	0.063731	0.125717	0.507	0.612199	
calcium_cat6	0.009747	0.139747	0.070	0.944395	
calcium_cat7	-0.037145	0.085694	-0.433	0.664682	
calcium_cat8	-0.166950	0.068458	-2.439	0.014740	*
calcium_cat9	-0.181317	0.065109	-2.785	0.005356	**
calcium_catUNK	-0.097669	0.122247	-0.799	0.424322	
sodium_final1	-0.026852	0.167176	-0.161	0.872394	
sodium_final2	-0.262835	0.165683	-1.586	0.112656	
sodium_final3	0.303206	0.227038	1.335	0.181717	
sodium_finalUNK	-0.405389	0.216946	-1.869	0.061676	.
platelets_cat1	-0.896230	0.089051	-10.064	< 2e-16	***
platelets_cat2	-1.016270	0.090093	-11.280	< 2e-16	***
platelets_cat3	-0.857904	0.094505	-9.078	< 2e-16	***
platelets_cat4	-0.753381	0.113982	-6.610	3.85e-11	***
platelets_cat5	-0.921248	0.176148	-5.230	1.70e-07	***
platelets_cat6	-0.782645	0.207776	-3.767	0.000165	***
platelets_catUNK	-0.745256	0.117414	-6.347	2.19e-10	***
potassium_cat3	0.067114	0.182934	0.367	0.713711	
potassium_cat4	0.008649	0.182199	0.047	0.962139	
potassium_cat5	0.154435	0.183012	0.844	0.398753	
potassium_cat6	0.368299	0.191495	1.923	0.054444	.
potassium_cat7	0.283380	0.217409	1.303	0.192425	
potassium_catUNK	0.333208	0.233308	1.428	0.153238	
unk_yes	-0.015087	0.088380	-0.171	0.864454	

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 35066 on 61229 degrees of freedom

Residual deviance: 24179 on 61142 degrees of freedom

AIC: 24355

Number of Fisher Scoring iterations: 6

Hide

```
vif(model1)
```

	GVIF	Df	GVIF^(1/(2*Df))
bmi	1.122169	1	1.059325
a3j_diag	4.984529	4	1.222371
arf_apache	1.059287	1	1.029217
gcs_sum	1.619395	1	1.272554
intubated_apache	1.538285	1	1.240276
ventilated_apache	1.787161	1	1.336847
map_risk	1.126432	1	1.061335
age	1.193394	1	1.092426
elective_surgery	3.094493	1	1.759117
gender	1.048950	1	1.024183
icu_type	1.471470	7	1.027974
pre_icu_los_days	1.052304	1	1.025819
aids	1.025019	1	1.012432
cirrhosis	1.703598	1	1.305220
hepatic_failure	1.668597	1	1.291742
immunosuppression	1.516225	1	1.231351
leukemia	1.144460	1	1.069794
lymphoma	1.078163	1	1.038346
solid_tumor_with_metastasis	1.445857	1	1.202438
comor_sum	2.742515	1	1.656054
hr_final	3.612629	8	1.083587
rr_final	2.597115	8	1.061465
temp_final	2.001512	6	1.059530
spo2_final	1.417954	1	1.190779
spo2_change	2.324804	2	1.234800
glu_final	2.995682	5	1.115962
hco3_final	6.974660	5	1.214374
calcium_cat	6.701620	5	1.209534
sodium_final	6.359702	4	1.260171
platelets_cat	3.525885	7	1.094185
potassium_cat	8.517469	6	1.195435
unk_yes	4.991146	1	2.234087

Auto Stepwise Feature Selection - the output shows that there's mild multi-collinearity between bun\_final and sodium\_final (less significant)

Hide

```
model1A = stepAIC(model1, trace = F)
summary(model1A)
```

Call:

```
glm(formula = hospital_death ~ bmi + a3j_diag + arf_apache +
     gcs_sum + intubated_apache + ventilated_apache + map_risk +
     age + elective_surgery + icu_type + pre_icu_los_days + aids +
     cirrhosis + hepatic_failure + immunosuppression + lymphoma +
     solid_tumor_with_metastasis + comor_sum + hr_final + rr_final +
     temp_final + spo2_final + spo2_change + glu_final + hco3_final +
     calcium_cat + sodium_final + platelets_cat + potassium_cat,
     family = binomial, data = train)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-3.1587	-0.3438	-0.2101	-0.1267	3.4341

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	1.377294	0.443892	3.103	0.001917	**
bmi	-0.003507	0.002217	-1.582	0.113758	
a3j_diag1	0.783877	0.271818	2.884	0.003929	**
a3j_diag2	0.722895	0.271934	2.658	0.007853	**
a3j_diag3	0.102349	0.277880	0.368	0.712633	
a3j_diag4	0.291810	0.287251	1.016	0.309691	
arf_apache1	0.305798	0.091494	3.342	0.000831	***
gcs_sum	-0.107089	0.004577	-23.396	< 2e-16	***
intubated_apache1	0.101988	0.046973	2.171	0.029914	*
ventilated_apache1	0.810921	0.046836	17.314	< 2e-16	***
map_risk	0.042381	0.003117	13.597	< 2e-16	***
age	0.034105	0.001291	26.409	< 2e-16	***
elective_surgery1	-1.005308	0.111466	-9.019	< 2e-16	***
icu_typeCCU-CTICU	-0.226933	0.096640	-2.348	0.018863	*
icu_typeCSICU	-0.437742	0.120110	-3.645	0.000268	***
icu_typeCTICU	-0.252506	0.120956	-2.088	0.036835	*
icu_typeMed-Surg ICU	-0.222810	0.075349	-2.957	0.003106	**
icu_typeMICU	-0.061502	0.088365	-0.696	0.486430	
icu_typeNeuro ICU	0.250181	0.095301	2.625	0.008661	**
icu_typeSICU	-0.006786	0.105732	-0.064	0.948827	
pre_icu_los_days	0.048156	0.004948	9.732	< 2e-16	***
aids1	0.660422	0.424862	1.554	0.120079	
cirrhosis1	0.673357	0.138216	4.872	1.11e-06	***
hepatic_failure1	0.444517	0.147095	3.022	0.002511	**
immunosuppression1	0.464443	0.100926	4.602	4.19e-06	***
lymphoma1	0.690082	0.202578	3.406	0.000658	***
solid_tumor_with_metastasis1	1.202852	0.100970	11.913	< 2e-16	***
comor_sum	-0.172069	0.043834	-3.925	8.66e-05	***
hr_final1	-1.152308	0.130136	-8.855	< 2e-16	***
hr_final2	-1.053272	0.096444	-10.921	< 2e-16	***
hr_final3	-0.911014	0.101642	-8.963	< 2e-16	***
hr_final4	-0.611002	0.100831	-6.060	1.36e-09	***
hr_final5	-0.480593	0.098120	-4.898	9.68e-07	***
hr_final6	-0.235154	0.109863	-2.140	0.032320	*
hr_final7	-0.054288	0.120201	-0.452	0.651523	
hr_finalUNK	-1.210131	0.572623	-2.113	0.034574	*
rr_final1	-1.682787	0.085255	-19.738	< 2e-16	***
rr_final2	-1.637973	0.088090	-18.594	< 2e-16	***
rr_final3	-1.585735	0.092260	-17.188	< 2e-16	***
rr_final4	-1.295310	0.083239	-15.561	< 2e-16	***

rr_final5	-1.217054	0.091030	-13.370	< 2e-16	***
rr_final6	-1.114407	0.093649	-11.900	< 2e-16	***
rr_final7	-1.348564	0.104520	-12.902	< 2e-16	***
rr_finalUNK	-1.636799	0.378653	-4.323	1.54e-05	***
temp_final1	0.391711	0.189321	2.069	0.038543	*
temp_final2	-0.085045	0.200213	-0.425	0.671004	
temp_final3	-0.483433	0.122440	-3.948	7.87e-05	***
temp_final4	-0.750758	0.098297	-7.638	2.21e-14	***
temp_final5	-1.122512	0.093399	-12.018	< 2e-16	***
temp_finalUNK	-0.610788	0.134468	-4.542	5.57e-06	***
spo2_final	-0.021568	0.001403	-15.371	< 2e-16	***
spo2_change1	0.102265	0.040588	2.520	0.011750	*
spo2_changeUNK	0.947436	0.285264	3.321	0.000896	***
glu_final1	0.036471	0.081503	0.447	0.654534	
glu_final2	-0.083106	0.079995	-1.039	0.298859	
glu_final3	0.361853	0.117165	3.088	0.002012	**
glu_final4	0.513303	0.163402	3.141	0.001682	**
glu_finalUNK	0.249948	0.136823	1.827	0.067730	.
hco3_final3	-0.153885	0.077969	-1.974	0.048419	*
hco3_final4	0.110551	0.083142	1.330	0.183628	
hco3_final5	0.451036	0.095320	4.732	2.23e-06	***
hco3_final6	0.736269	0.100487	7.327	2.35e-13	***
hco3_finalUNK	0.053788	0.108481	0.496	0.620016	
calcium_cat6	0.012937	0.139714	0.093	0.926225	
calcium_cat7	-0.034356	0.085654	-0.401	0.688344	
calcium_cat8	-0.164249	0.068391	-2.402	0.016323	*
calcium_cat9	-0.179468	0.065072	-2.758	0.005816	**
calcium_catUNK	-0.103640	0.111309	-0.931	0.351801	
sodium_final1	-0.028911	0.167068	-0.173	0.862611	
sodium_final2	-0.265694	0.165567	-1.605	0.108550	
sodium_final3	0.299158	0.226970	1.318	0.187487	
sodium_finalUNK	-0.405117	0.214680	-1.887	0.059151	.
platelets_cat1	-0.906758	0.088451	-10.251	< 2e-16	***
platelets_cat2	-1.030344	0.089338	-11.533	< 2e-16	***
platelets_cat3	-0.875058	0.093635	-9.345	< 2e-16	***
platelets_cat4	-0.771038	0.113232	-6.809	9.80e-12	***
platelets_cat5	-0.938709	0.175716	-5.342	9.18e-08	***
platelets_cat6	-0.795732	0.207434	-3.836	0.000125	***
platelets_catUNK	-0.766862	0.106120	-7.226	4.96e-13	***
potassium_cat3	0.064716	0.182892	0.354	0.723451	
potassium_cat4	0.008554	0.182144	0.047	0.962541	
potassium_cat5	0.155799	0.182937	0.852	0.394405	
potassium_cat6	0.370509	0.191381	1.936	0.052871	.
potassium_cat7	0.286591	0.217297	1.319	0.187206	
potassium_catUNK	0.340840	0.228285	1.493	0.135425	

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 35066 on 61229 degrees of freedom

Residual deviance: 24181 on 61145 degrees of freedom

AIC: 24351

Number of Fisher Scoring iterations: 6

Hide



```
vif(model1A)
```

	GVIF	Df	GVIF^(1/(2*Df))
bmi	1.115883	1	1.056353
a3j_diag	4.972482	4	1.222001
arf_apache	1.058716	1	1.028939
gcs_sum	1.618557	1	1.272225
intubated_apache	1.537277	1	1.239870
ventilated_apache	1.783833	1	1.335602
map_risk	1.122732	1	1.059590
age	1.187842	1	1.089882
elective_surgery	3.092139	1	1.758448
icu_type	1.466312	7	1.027716
pre_icu_los_days	1.050954	1	1.025161
aids	1.023117	1	1.011492
cirrhosis	1.680024	1	1.296157
hepatic_failure	1.655110	1	1.286511
immunosuppression	1.514668	1	1.230719
lymphoma	1.074722	1	1.036688
solid_tumor_with_metastasis	1.403500	1	1.184694
comor_sum	2.531524	1	1.591076
hr_final	3.528728	8	1.081997
rr_final	2.581621	8	1.061068
temp_final	1.463866	6	1.032266
spo2_final	1.417220	1	1.190471
spo2_change	2.267347	2	1.227099
glu_final	2.964384	5	1.114791
hco3_final	4.373912	5	1.159010
calcium_cat	5.200424	5	1.179245
sodium_final	6.079052	4	1.253082
platelets_cat	2.247504	7	1.059550
potassium_cat	7.569788	6	1.183742

### Relative Importance of Features

[Hide](#)

```
imp = as.data.frame(varImp(model1A))
imp = data.frame(overall = imp$Overall, names = rownames(imp))

imp[order(imp$overall,decreasing = T),]
```

	overall	names
	<dbl>	<fctr>
11	26.40935998	age
7	23.39620095	gcs_sum
36	19.73838635	rr_final1
37	18.59424578	rr_final2
9	17.31423440	ventilated_apache1
38	17.18761098	rr_final3

overall		names
	<dbl>	<fctr>
39	15.56141261	rr_final4
50	15.37092577	spo2_final
10	13.59746211	map_risk
40	13.36988712	rr_final5

1-10 of 84 rows

Previous 1 2 3 4 5 6 ... 9 Next

## Results on Train set

Hide

```
trainPredict = predict(model1A, newdata = train, type = 'response')

p_class = ifelse(trainPredict > 0.08, 1, 0)

matrix_table = table(train$hospital_death, p_class)
matrix_table
```

```
      p_class
      0      1
No  44843 11297
Yes   1085  4005
```

Hide

```
# Accuracy
accuracy = sum(diag(matrix_table))/sum(matrix_table)
round(accuracy, 3)
```

```
[1] 0.798
```

## Performance on Train set

Hide

```
pred = prediction(trainPredict, train$hospital_death)

auc.log = performance(pred,"auc"); au_log = as.numeric(auc.log@y.values)
au_log
```

```
[1] 0.8777502
```

Hide

```
acc.perf = performance(pred, measure = "acc")

ind = which.max( slot(acc.perf, "y.values")[[1]] )
acc = slot(acc.perf, "y.values")[[1]][ind]
cutoff = slot(acc.perf, "x.values")[[1]][ind]
print(c(accuracy= acc, cutoff = cutoff))
```

```
accuracy cutoff.34804
0.9283031    0.5647228
```

## Results on Test set

Hide

```
testPredict = predict(model1A, newdata = test, type = 'response')

p_class = ifelse(testPredict > 0.08, 1, 0)

matrix_table = table(test$hospital_death, p_class)
matrix_table
```

```
      p_class
      0      1
No  19226  4834
Yes   491  1691
```

Hide

```
# Accuracy
accuracy = sum(diag(matrix_table))/sum(matrix_table)
round(accuracy, 3)
```

```
[1] 0.797
```

## Performance on Test set

Hide

```
pred = prediction(testPredict, test$hospital_death)

auc.log = performance(pred,"auc"); au_log = as.numeric(auc.log@y.values)
au_log
```

```
[1] 0.8687843
```

Hide

```
acc.perf = performance(pred, measure = "acc")

ind = which.max( slot(acc.perf, "y.values")[[1]] )
acc = slot(acc.perf, "y.values")[[1]][ind]
cutoff = slot(acc.perf, "x.values")[[1]][ind]
print(c(accuracy= acc, cutoff = cutoff))
```

```
accuracy cutoff.52664
0.9293880    0.5858979
```

# More information on the model

- Prediction

```
head(sort(testPredict, decreasing = T),10)
```

5610	34132	56125	29649	57285	31981	11692	38708	36958
42352								
0.9991656	0.9961037	0.9951361	0.9940082	0.9939187	0.9906040	0.9898943	0.9898744	0.9892796
0.9874060								

```
head(testPredict, 10)
```

2	4	5	8	12	17	21
22	25					
0.387643839	0.102317023	0.009850346	0.064857740	0.016015496	0.004408000	0.006015218
0.0533838	0.008155155					
26						
0.642814783						

```
head(test, 10)
```

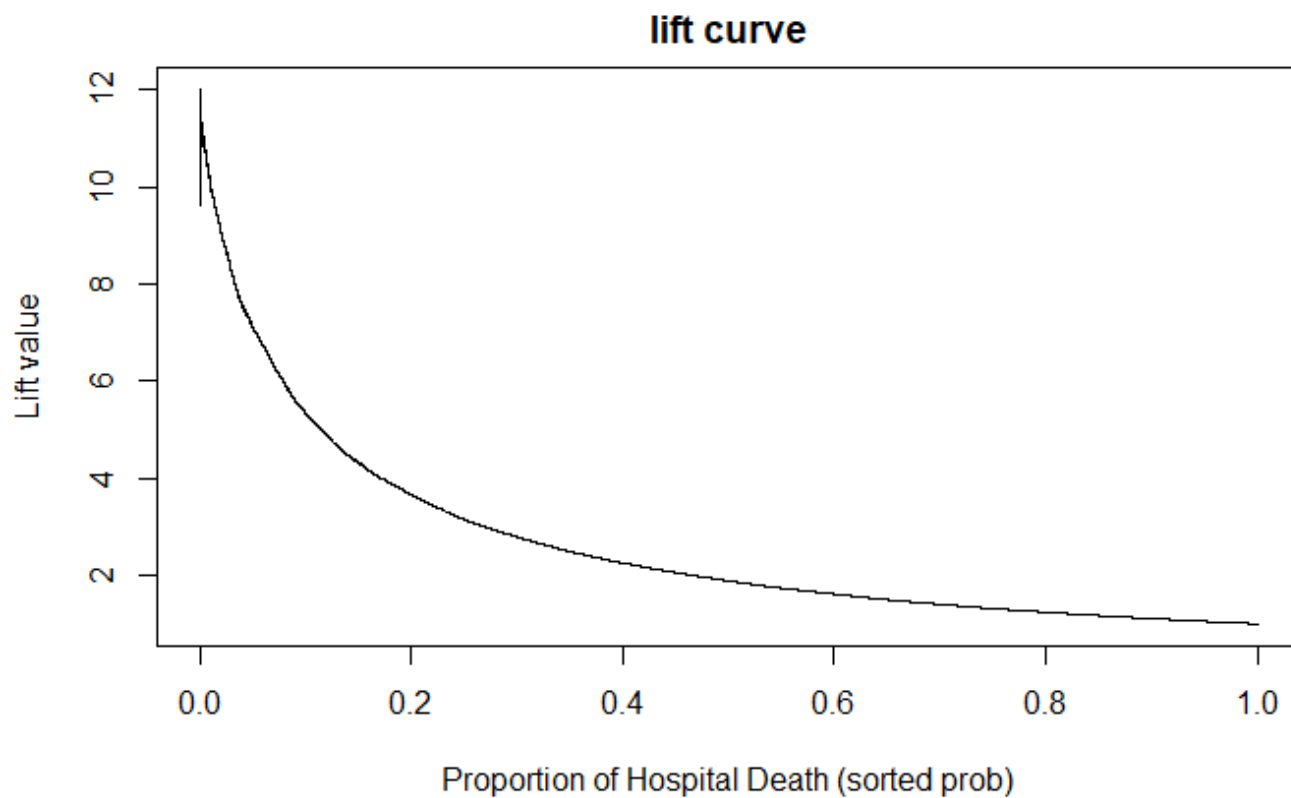
	bmi	a2_diag	a3j_diag	apache_post_operative	arf_apache	gcs_ey...	gcs_mo...	gc
	<dbl>	<fctr>	<fctr>	<fctr>	<fctr>	<fctr>	<fctr>	<fctr>
2	27.42000	1	1	0	0	1	3	1
4	22.64000	2	4	1	0	4	6	5
5	27.94000	1	3	0	0	0	0	0
8	27.94000	1	2	0	0	4	6	5
12	27.38281	2	4	1	0	4	6	4
17	23.38318	3	1	0	0	4	6	5
21	28.37610	3	4	1	0	4	6	5
22	35.65999	1	2	0	0	4	6	5
25	29.50996	1	1	0	0	4	6	5
26	26.01070	1	2	0	0	4	6	3

1-10 of 10 rows | 1-9 of 57 columns

- Lift Chart

```
pred = prediction(trainPredict, train$hospital_death)

perf = performance(pred, "lift", "rpp" )
plot(perf, main="lift curve", xlab = 'Proportion of Hospital Death (sorted prob)')
```



- p-value

[Hide](#)

```
with(model1A, pchisq(null.deviance - deviance,  
                     df.null - df.residual, lower.tail=F))
```

```
[1] 0
```

[Hide](#)

```
confusionMatrix(factor(p_class, levels = c(0,1), labels = c("No", "Yes")), test$hospital_deat  
h, positive = "Yes")
```

## Confusion Matrix and Statistics

	Reference	
Prediction	No	Yes
No	19226	491
Yes	4834	1691

Accuracy : 0.7971

95% CI : (0.7922, 0.8019)

No Information Rate : 0.9169

P-Value [Acc &gt; NIR] : 1

Kappa : 0.3014

McNemar's Test P-Value : &lt;2e-16

Sensitivity : 0.77498

Specificity : 0.79909

Pos Pred Value : 0.25916

Neg Pred Value : 0.97510

Prevalence : 0.08315

Detection Rate : 0.06444

Detection Prevalence : 0.24865

Balanced Accuracy : 0.78703

'Positive' Class : Yes

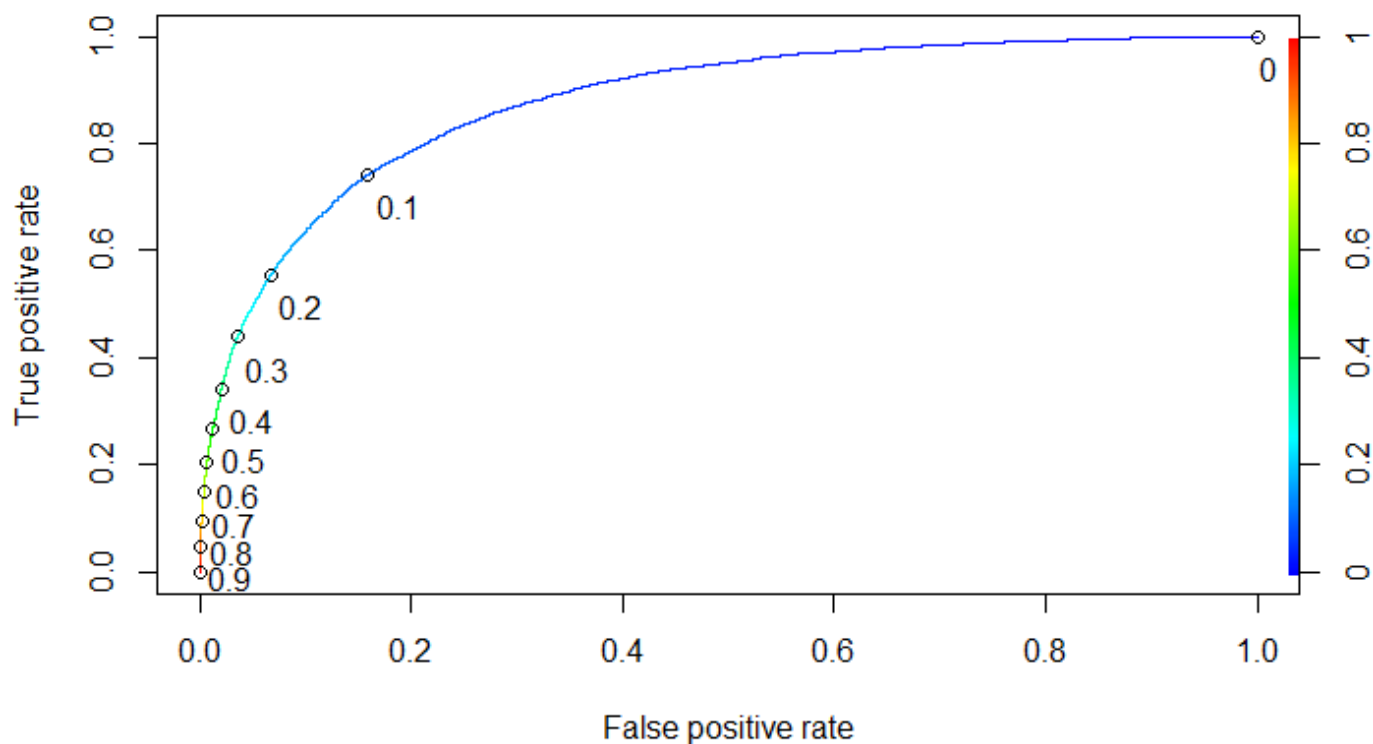
- cut-off

Hide

```
options(repr.plot.width=8, repr.plot.height=6)

perf <- performance(pred, "tpr", "fpr" )

plot( perf, colorize = TRUE,
      print.cutoffs.at = seq(0,1,0.1),
      text.adj = c(-0.2, 1.7))
```



- optimal cut-offs (same or different costs)

[Hide](#)

```
cost.perf = performance(pred, "cost")
pred@cutoffs[[1]][which.min(cost.perf@y.values[[1]])]
```

```
34804
0.5647228
```

[Hide](#)

```
cost.perf = performance(pred, "cost", cost.fp = 4, cost.fn = 1)
pred@cutoffs[[1]][which.min(cost.perf@y.values[[1]])]
```

```
30904
0.8609825
```

[Hide](#)

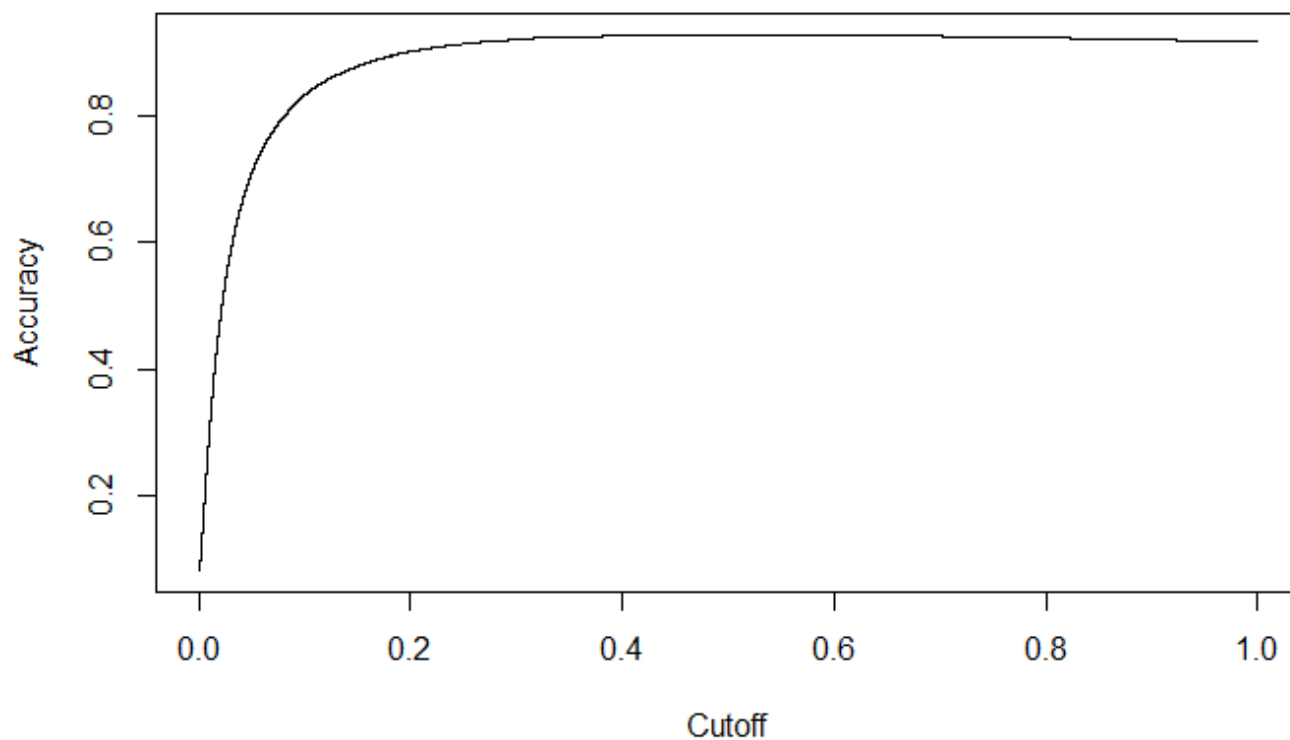
```
auc.log = performance(pred, "auc"); au_log = as.numeric(auc.log@y.values)
au_log
```

```
[1] 0.8777502
```

- optimal accuracy

[Hide](#)

```
acc.perf = performance(pred, measure = "acc")
plot(acc.perf)
```



Hide

```
ind = which.max( slot(acc.perf, "y.values")[[1]] )
acc = slot(acc.perf, "y.values")[[1]][ind]
cutoff = slot(acc.perf, "x.values")[[1]][ind]
print(c(accuracy= acc, cutoff = cutoff))
```

```
accuracy cutoff.34804
0.9283031    0.5647228
```

### Results on Entire dataset

Hide

```
dataPredict = predict(model1A, newdata = data, type = 'response')

p_class = ifelse(dataPredict > 0.08, 1, 0)

matrix_table = table(data$hospital_death, p_class)
matrix_table
```

```
p_class
      0      1
No  64069 16131
Yes   1576  5696
```

Hide



```
# Accuracy
accuracy = sum(diag(matrix_table))/sum(matrix_table)
round(accuracy, 3)
```

```
[1] 0.798
```

## Logistic Regression with balanced dataset

Over-sample target variable =1 and under-sample target variable = 0

[Hide](#)

```
set.seed(789)
data.balance <- SMOTE(hospital_death ~ ., as.data.frame(train), perc.under = 120, perc.over = 500)
table(data.balance$hospital_death)
```

```
      No      Yes
30540  30540
```

Initial model with iterations to remove multi-collinearity

[Hide](#)

```
model2 = glm(hospital_death ~ . -gcs_motor -gcs_verbal -gcs_eyes -diabetes_mellitus
             #-hr_change -rr_change -temp_change -bun_change -cre_change -glu_change -hco3_ch
ange
             #-hto_change -sodium_change -wbc_change -map_change -unk_cnt,
             -a2_diag -apache_post_operative,# -hto_final -bun_final -cre_final -hemoglobin_f
inal
             #-map_final -wbc_final,
             data = data.balance, family = binomial)
summary(model2)
```

Call:

```
glm(formula = hospital_death ~ . - gcs_motor - gcs_verbal - gcs_eyes -
     diabetes_mellitus - a2_diag - apache_post_operative, family = binomial,
     data = data.balance)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-4.8266	-0.3647	-0.0107	0.3209	3.3932

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	6.5023524	0.4229962	15.372	< 2e-16	***
bmi	-0.0158943	0.0019149	-8.301	< 2e-16	***
a3j_diag1	0.7559441	0.1525774	4.954	7.25e-07	***
a3j_diag2	0.8186694	0.1527005	5.361	8.26e-08	***
a3j_diag3	0.2818382	0.1582604	1.781	0.074937	.
a3j_diag4	0.7773480	0.1573865	4.939	7.85e-07	***
arf_apache1	0.7563618	0.0678495	11.148	< 2e-16	***
gcs_sum	-0.1864599	0.0039209	-47.556	< 2e-16	***
intubated_apache1	0.1056950	0.0366406	2.885	0.003919	**
ventilated_apache1	0.8093144	0.0329172	24.586	< 2e-16	***
map_risk	0.0349097	0.0029262	11.930	< 2e-16	***
age	0.0387849	0.0011162	34.746	< 2e-16	***
elective_surgery1	-0.3410336	0.0505382	-6.748	1.50e-11	***
genderM	-0.0181535	0.0286055	-0.635	0.525678	
icu_typeCCU-CTICU	-0.1751256	0.0786232	-2.227	0.025920	*
icu_typeCSICU	-0.6361933	0.0970911	-6.553	5.66e-11	***
icu_typeCTICU	-0.4449877	0.0925129	-4.810	1.51e-06	***
icu_typeMed-Surg ICU	-0.2680764	0.0625738	-4.284	1.83e-05	***
icu_typeMICU	-0.0196296	0.0742794	-0.264	0.791574	
icu_typeNeuro ICU	0.3244600	0.0776430	4.179	2.93e-05	***
icu_typeSICU	0.3024003	0.0819043	3.692	0.000222	***
pre_icu_los_days	0.0529001	0.0054871	9.641	< 2e-16	***
aids1	-0.1200392	0.4314694	-0.278	0.780851	
cirrhosis1	0.9186193	0.0894466	10.270	< 2e-16	***
hepatic_failure1	0.5249622	0.0972409	5.399	6.72e-08	***
immunosuppression1	0.1120794	0.0798630	1.403	0.160499	
leukemia1	0.1561066	0.1270052	1.229	0.219021	
lymphoma1	0.9358624	0.1644050	5.692	1.25e-08	***
solid_tumor_with_metastasis1	1.0620681	0.0749723	14.166	< 2e-16	***
comor_sum	0.3766933	0.0305486	12.331	< 2e-16	***
map_change1	-0.3927679	0.0697648	-5.630	1.80e-08	***
map_changeUNK	0.4344950	0.2147769	2.023	0.043073	*
map_final1	-0.8389653	0.0524853	-15.985	< 2e-16	***
map_final2	-1.1031452	0.0605009	-18.234	< 2e-16	***
map_final3	-0.8596885	0.0700648	-12.270	< 2e-16	***
map_final4	-0.5452839	0.1037331	-5.257	1.47e-07	***
map_final5	-1.7142488	0.0962672	-17.807	< 2e-16	***
map_final6	-0.9570719	0.0736204	-13.000	< 2e-16	***
map_final7	-0.9142761	0.0845321	-10.816	< 2e-16	***
map_final8	-1.2849914	0.0829070	-15.499	< 2e-16	***
map_finalUNK	0.0379384	0.2071918	0.183	0.854714	
hr_change1	0.1405356	0.0409467	3.432	0.000599	***
hr_changeUNK	1.1621824	0.2137673	5.437	5.43e-08	***
hr_final1	-1.2959328	0.1093801	-11.848	< 2e-16	***
hr_final2	-1.0233291	0.0866080	-11.816	< 2e-16	***

hr_final3	-1.0558277	0.0877398	-12.034	< 2e-16	***
hr_final4	-0.7369740	0.0878374	-8.390	< 2e-16	***
hr_final5	-0.5484946	0.0863199	-6.354	2.10e-10	***
hr_final6	-0.1774707	0.0977237	-1.816	0.069363	.
hr_final7	-0.2447469	0.1126467	-2.173	0.029803	*
hr_finalUNK	0.2585465	0.2266840	1.141	0.254054	
rr_change1	-0.3545650	0.0468797	-7.563	3.93e-14	***
rr_changeUNK	0.7208311	0.1444928	4.989	6.08e-07	***
rr_final1	-1.7181079	0.0783856	-21.919	< 2e-16	***
rr_final2	-1.5765635	0.0805756	-19.566	< 2e-16	***
rr_final3	-1.3416668	0.0840563	-15.962	< 2e-16	***
rr_final4	-1.1478504	0.0778331	-14.748	< 2e-16	***
rr_final5	-1.1318352	0.0857580	-13.198	< 2e-16	***
rr_final6	-1.1025925	0.0893412	-12.341	< 2e-16	***
rr_final7	-1.3265189	0.0982038	-13.508	< 2e-16	***
rr_finalUNK	-0.3301761	0.1540046	-2.144	0.032038	*
temp_change1	-0.4149226	0.0446301	-9.297	< 2e-16	***
temp_final1	0.6286045	0.1803330	3.486	0.000491	***
temp_final2	-0.2253144	0.1837193	-1.226	0.220046	
temp_final3	-0.4245195	0.1182831	-3.589	0.000332	***
temp_final4	-0.7562448	0.0975604	-7.752	9.08e-15	***
temp_final5	-1.3731564	0.0940073	-14.607	< 2e-16	***
temp_finalUNK	0.1810691	0.1146210	1.580	0.114171	
spo2_final	-0.0288911	0.0014333	-20.158	< 2e-16	***
spo2_change1	0.0711792	0.0304555	2.337	0.019431	*
spo2_changeUNK	0.6798505	0.1570218	4.330	1.49e-05	***
bun_final1	0.4144872	0.0722507	5.737	9.65e-09	***
bun_final2	0.3840277	0.0445158	8.627	< 2e-16	***
bun_final3	0.5385337	0.0565244	9.527	< 2e-16	***
bun_final4	0.7085212	0.0867076	8.171	3.05e-16	***
bun_finalUNK	0.4433193	0.0566181	7.830	4.88e-15	***
cre_final1	-0.2204459	0.0839974	-2.624	0.008679	**
cre_final2	0.0927184	0.0946765	0.979	0.327423	
cre_final3	0.0515654	0.0921113	0.560	0.575605	
cre_finalUNK	0.0482862	0.0926109	0.521	0.602097	
glu_final1	-0.0948551	0.0670500	-1.415	0.157159	
glu_final2	-0.3014825	0.0725833	-4.154	3.27e-05	***
glu_final3	0.1860645	0.1036239	1.796	0.072562	.
glu_final4	0.1457283	0.1553409	0.938	0.348183	
glu_finalUNK	0.0929752	0.0813383	1.143	0.253011	
hco3_final3	-0.2910313	0.0709188	-4.104	4.07e-05	***
hco3_final4	-0.0030785	0.0755070	-0.041	0.967479	
hco3_final5	0.3779072	0.0889392	4.249	2.15e-05	***
hco3_final6	0.4385999	0.0948755	4.623	3.78e-06	***
hco3_finalUNK	0.0694283	0.0782961	0.887	0.375219	
hto_final1	0.2674277	0.0783575	3.413	0.000643	***
hto_final2	0.6049241	0.1653732	3.658	0.000254	***
hto_finalUNK	0.3230969	0.0457204	7.067	1.59e-12	***
calcium_cat6	0.1287038	0.1263555	1.019	0.308400	
calcium_cat7	0.0663433	0.0757927	0.875	0.381396	
calcium_cat8	-0.2432851	0.0604598	-4.024	5.72e-05	***
calcium_cat9	-0.1964232	0.0574202	-3.421	0.000624	***
calcium_catUNK	0.0003067	0.0657400	0.005	0.996277	
hemoglobin_final11	-0.0786463	0.0591531	-1.330	0.183670	
hemoglobin_final12	0.0041614	0.0629717	0.066	0.947311	
hemoglobin_final13	0.0461196	0.0698677	0.660	0.509190	
hemoglobin_final14	0.0708127	0.0900262	0.787	0.431528	
hemoglobin_final15	0.1587362	0.1211769	1.310	0.190211	

hemoglobin_final16	0.3822406	0.1592387	2.400	0.016376	*
hemoglobin_final17	-0.2786793	0.2466639	-1.130	0.258563	
hemoglobin_final15	0.6121655	0.1648126	3.714	0.000204	***
hemoglobin_final6	-0.0245469	0.1153353	-0.213	0.831459	
hemoglobin_final7	-0.1344324	0.0763889	-1.760	0.078434	.
hemoglobin_final8	-0.0306529	0.0646911	-0.474	0.635617	
hemoglobin_final9	0.1167014	0.0584136	1.998	0.045733	*
hemoglobin_finalUNK	0.2279411	0.0580377	3.927	8.58e-05	***
bun_change1	-0.0084790	0.0524893	-0.162	0.871670	
bun_changeUNK	0.0695623	0.0472909	1.471	0.141306	
cre_change1	0.3734156	0.0600194	6.222	4.92e-10	***
cre_changeUNK	0.2347579	0.0467657	5.020	5.17e-07	***
glu_change1	-0.0427411	0.0473525	-0.903	0.366730	
glu_changeUNK	0.2646379	0.0527936	5.013	5.37e-07	***
hco3_change1	-0.3043722	0.0482673	-6.306	2.86e-10	***
hco3_changeUNK	0.0530985	0.0455805	1.165	0.244044	
hto_change1	0.2279715	0.0780960	2.919	0.003510	**
hto_changeUNK	0.3256030	0.0449949	7.236	4.61e-13	***
sodium_change1	-0.0334670	0.0571946	-0.585	0.558452	
sodium_changeUNK	0.0610155	0.0474918	1.285	0.198876	
sodium_final1	0.0735627	0.1645438	0.447	0.654824	
sodium_final2	-0.1846962	0.1669302	-1.106	0.268542	
sodium_final3	0.2898119	0.2195921	1.320	0.186910	
sodium_finalUNK	-0.0222751	0.1700968	-0.131	0.895810	
wbc_change1	-0.0282025	0.0673000	-0.419	0.675175	
wbc_changeUNK	0.1062898	0.0448472	2.370	0.017786	*
wbc_final1	-0.3199342	0.1934563	-1.654	0.098172	.
wbc_final2	-0.6189155	0.1742663	-3.552	0.000383	***
wbc_final3	-0.0926523	0.1756661	-0.527	0.597892	
wbc_final4	-0.2141425	0.1766680	-1.212	0.225467	
wbc_finalUNK	-0.3627462	0.1762613	-2.058	0.039590	*
platelets_cat1	-1.0412780	0.0880849	-11.821	< 2e-16	***
platelets_cat2	-0.9835470	0.0890578	-11.044	< 2e-16	***
platelets_cat3	-0.8536527	0.0928540	-9.193	< 2e-16	***
platelets_cat4	-0.6999962	0.1095146	-6.392	1.64e-10	***
platelets_cat5	-0.7927159	0.1659467	-4.777	1.78e-06	***
platelets_cat6	-0.8846546	0.2015968	-4.388	1.14e-05	***
platelets_catUNK	-0.7209763	0.0924486	-7.799	6.26e-15	***
potassium_cat3	0.0976966	0.1582323	0.617	0.536954	
potassium_cat4	-0.0538703	0.1569931	-0.343	0.731495	
potassium_cat5	-0.0459624	0.1578501	-0.291	0.770916	
potassium_cat6	0.1865967	0.1660295	1.124	0.261065	
potassium_cat7	0.0013531	0.1947008	0.007	0.994455	
potassium_catUNK	0.1973199	0.1599560	1.234	0.217356	
unk_cnt	0.0520577	0.0054169	9.610	< 2e-16	***
unk_yes	-0.6917347	0.0564617	-12.251	< 2e-16	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 84675 on 61079 degrees of freedom

Residual deviance: 33874 on 60931 degrees of freedom

AIC: 34172

Number of Fisher Scoring iterations: 7

vif(model2)

	GVIF	Df	GVIF <sup>1/(2*Df)</sup>
bmi	1.130636	1	1.063314
a3j_diag	2.737078	4	1.134125
arf_apache	1.101088	1	1.049327
gcs_sum	1.337973	1	1.156708
intubated_apache	1.288444	1	1.135097
ventilated_apache	1.404300	1	1.185032
map_risk	1.495987	1	1.223105
age	1.259334	1	1.122201
elective_surgery	1.886542	1	1.373514
gender	1.055550	1	1.027399
icu_type	1.410114	7	1.024852
pre_icu_los_days	1.072578	1	1.035654
aids	1.015273	1	1.007607
cirrhosis	1.326354	1	1.151675
hepatic_failure	1.341800	1	1.158361
immunosuppression	1.235377	1	1.111475
leukemia	1.103309	1	1.050385
lymphoma	1.050793	1	1.025082
solid_tumor_with_metastasis	1.181640	1	1.087033
comor_sum	1.812219	1	1.346187
map_change	1.726297	2	1.146249
map_final	2.793834	9	1.058739
hr_change	2.569470	2	1.266079
hr_final	2.966210	8	1.070317
rr_change	2.692135	2	1.280927
rr_final	2.977116	8	1.070563
temp_change	1.948166	1	1.395767
temp_final	1.317269	6	1.023229
spo2_final	1.246079	1	1.116279
spo2_change	1.470179	2	1.101140
bun_final	4.204014	5	1.154427
cre_final	4.363473	4	1.202206
glu_final	4.211483	5	1.154632
hco3_final	3.579995	5	1.136026
hto_final	5.225633	3	1.317316
calcium_cat	2.587498	5	1.099735
hemoglobin_final	6.257076	13	1.073074
bun_change	2.615526	2	1.271715
cre_change	2.619220	2	1.272164
glu_change	3.906651	2	1.405889
hco3_change	3.158056	2	1.333076
hto_change	2.285784	2	1.229586
sodium_change	3.074862	2	1.324209
sodium_final	3.154683	4	1.154435
wbc_change	3.091936	2	1.326043
wbc_final	3.545399	5	1.134923
platelets_cat	2.660207	7	1.072386
potassium_cat	2.598024	6	1.082813
unk_cnt	5.618706	1	2.370381
unk_yes	3.238390	1	1.799553

Auto Stepwise Feature Selection - the output shows that there's mild multi-collinearity between bun\_final and sodium\_final (less significant)

[Hide](#)

```
timestamp()
```

```
##----- Sat Apr 11 12:26:13 2020 -----##
```

[Hide](#)

```
model2A = stepAIC(model2, trace = F)  
summary(model2A)
```

Call:

```
glm(formula = hospital_death ~ bmi + a3j_diag + arf_apache +
    gcs_sum + intubated_apache + ventilated_apache + map_risk +
    age + elective_surgery + icu_type + pre_icu_los_days + cirrhosis +
    hepatic_failure + lymphoma + solid_tumor_with_metastasis +
    comor_sum + map_change + map_final + hr_change + hr_final +
    rr_change + rr_final + temp_change + temp_final + spo2_final +
    spo2_change + bun_final + cre_final + glu_final + hco3_final +
    hto_final + calcium_cat + hemoglobin_final + cre_change +
    glu_change + hco3_change + hto_change + sodium_final + wbc_change +
    wbc_final + platelets_cat + potassium_cat + unk_cnt + unk_yes,
    family = binomial, data = data.balance)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-4.8338	-0.3651	-0.0107	0.3210	3.3991

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	6.507000	0.418994	15.530	< 2e-16	***
bmi	-0.016045	0.001906	-8.418	< 2e-16	***
a3j_diag1	0.757565	0.152521	4.967	6.80e-07	***
a3j_diag2	0.819623	0.152645	5.369	7.90e-08	***
a3j_diag3	0.281044	0.158179	1.777	0.075610	.
a3j_diag4	0.774927	0.157346	4.925	8.44e-07	***
arf_apache1	0.753906	0.067693	11.137	< 2e-16	***
gcs_sum	-0.186178	0.003915	-47.553	< 2e-16	***
intubated_apache1	0.104727	0.036613	2.860	0.004231	**
ventilated_apache1	0.808828	0.032895	24.588	< 2e-16	***
map_risk	0.034996	0.002924	11.968	< 2e-16	***
age	0.038839	0.001112	34.921	< 2e-16	***
elective_surgery1	-0.340052	0.050541	-6.728	1.72e-11	***
icu_typeCCU-CTICU	-0.170939	0.078569	-2.176	0.029581	*
icu_typeCSICU	-0.636422	0.097062	-6.557	5.49e-11	***
icu_typeCTICU	-0.435014	0.092401	-4.708	2.50e-06	***
icu_typeMed-Surg ICU	-0.265028	0.062572	-4.236	2.28e-05	***
icu_typeMICU	-0.016152	0.074261	-0.218	0.827812	
icu_typeNeuro ICU	0.329174	0.077632	4.240	2.23e-05	***
icu_typeSICU	0.303820	0.081885	3.710	0.000207	***
pre_icu_los_days	0.052827	0.005490	9.622	< 2e-16	***
cirrhosis1	0.900816	0.088653	10.161	< 2e-16	***
hepatic_failure1	0.506818	0.096477	5.253	1.49e-07	***
lymphoma1	0.952150	0.163276	5.832	5.49e-09	***
solid_tumor_with_metastasis1	1.067262	0.074460	14.333	< 2e-16	***
comor_sum	0.396580	0.028175	14.076	< 2e-16	***
map_change1	-0.389873	0.069734	-5.591	2.26e-08	***
map_changeUNK	0.433421	0.214792	2.018	0.043606	*
map_final1	-0.840250	0.052453	-16.019	< 2e-16	***
map_final2	-1.105557	0.060442	-18.291	< 2e-16	***
map_final3	-0.861645	0.069978	-12.313	< 2e-16	***
map_final4	-0.547216	0.103617	-5.281	1.28e-07	***
map_final5	-1.714551	0.096201	-17.823	< 2e-16	***
map_final6	-0.958541	0.073575	-13.028	< 2e-16	***
map_final7	-0.915291	0.084464	-10.836	< 2e-16	***
map_final8	-1.283997	0.082847	-15.498	< 2e-16	***
map_finalUNK	0.037438	0.207324	0.181	0.856700	

hr_change1	0.141323	0.040937	3.452	0.000556	***
hr_changeUNK	1.161420	0.213947	5.429	5.68e-08	***
hr_final1	-1.296776	0.109242	-11.871	< 2e-16	***
hr_final2	-1.023994	0.086545	-11.832	< 2e-16	***
hr_final3	-1.057604	0.087661	-12.065	< 2e-16	***
hr_final4	-0.737841	0.087756	-8.408	< 2e-16	***
hr_final5	-0.548571	0.086254	-6.360	2.02e-10	***
hr_final6	-0.178767	0.097654	-1.831	0.067156	.
hr_final7	-0.242827	0.112590	-2.157	0.031026	*
hr_finalUNK	0.254431	0.226621	1.123	0.261559	
rr_change1	-0.354130	0.046864	-7.556	4.14e-14	***
rr_changeUNK	0.720550	0.144522	4.986	6.17e-07	***
rr_final1	-1.716350	0.078355	-21.905	< 2e-16	***
rr_final2	-1.574527	0.080520	-19.555	< 2e-16	***
rr_final3	-1.338988	0.083977	-15.945	< 2e-16	***
rr_final4	-1.146494	0.077785	-14.739	< 2e-16	***
rr_final5	-1.128973	0.085704	-13.173	< 2e-16	***
rr_final6	-1.099933	0.089285	-12.319	< 2e-16	***
rr_final7	-1.322727	0.098127	-13.480	< 2e-16	***
rr_finalUNK	-0.323821	0.154035	-2.102	0.035531	*
temp_change1	-0.414208	0.044622	-9.283	< 2e-16	***
temp_final1	0.628881	0.180243	3.489	0.000485	***
temp_final2	-0.224463	0.183643	-1.222	0.221602	
temp_final3	-0.423174	0.118229	-3.579	0.000345	***
temp_final4	-0.754108	0.097510	-7.734	1.04e-14	***
temp_final5	-1.370218	0.093951	-14.584	< 2e-16	***
temp_finalUNK	0.187245	0.114557	1.635	0.102151	
spo2_final	-0.028869	0.001432	-20.163	< 2e-16	***
spo2_change1	0.071190	0.030437	2.339	0.019338	*
spo2_changeUNK	0.684020	0.157011	4.357	1.32e-05	***
bun_final1	0.412668	0.070309	5.869	4.37e-09	***
bun_final2	0.383155	0.043922	8.723	< 2e-16	***
bun_final3	0.535529	0.055816	9.595	< 2e-16	***
bun_final4	0.704504	0.085834	8.208	2.25e-16	***
bun_finalUNK	0.457529	0.055717	8.212	< 2e-16	***
cre_final1	-0.226050	0.083797	-2.698	0.006985	**
cre_final2	0.085787	0.094369	0.909	0.363321	
cre_final3	0.044657	0.091785	0.487	0.626582	
cre_finalUNK	0.055209	0.092309	0.598	0.549783	
glu_final1	-0.092076	0.066898	-1.376	0.168706	
glu_final2	-0.297134	0.072404	-4.104	4.06e-05	***
glu_final3	0.191913	0.103522	1.854	0.063762	.
glu_final4	0.147353	0.155216	0.949	0.342446	
glu_finalUNK	0.101472	0.081161	1.250	0.211209	
hco3_final3	-0.291758	0.070887	-4.116	3.86e-05	***
hco3_final4	-0.001267	0.075479	-0.017	0.986604	
hco3_final5	0.379659	0.088896	4.271	1.95e-05	***
hco3_final6	0.438790	0.094781	4.630	3.66e-06	***
hco3_finalUNK	0.079008	0.078174	1.011	0.312173	
hto_final1	0.264340	0.078314	3.375	0.000737	***
hto_final2	0.595503	0.165327	3.602	0.000316	***
hto_finalUNK	0.322895	0.045704	7.065	1.61e-12	***
calcium_cat6	0.125863	0.126216	0.997	0.318667	
calcium_cat7	0.067576	0.075680	0.893	0.371896	
calcium_cat8	-0.244154	0.060377	-4.044	5.26e-05	***
calcium_cat9	-0.197142	0.057370	-3.436	0.000590	***
calcium_catUNK	0.010546	0.065467	0.161	0.872028	
hemoglobin_final11	-0.077518	0.059128	-1.311	0.189850	



hemoglobin_final12	0.003555	0.062952	0.056	0.954962
hemoglobin_final13	0.045287	0.069851	0.648	0.516765
hemoglobin_final14	0.068713	0.089934	0.764	0.444841
hemoglobin_final15	0.159509	0.121093	1.317	0.187756
hemoglobin_final16	0.383104	0.159116	2.408	0.016053 *
hemoglobin_final17	-0.279157	0.246639	-1.132	0.257700
hemoglobin_final5	0.613854	0.164830	3.724	0.000196 ***
hemoglobin_final6	-0.023827	0.115290	-0.207	0.836267
hemoglobin_final7	-0.130577	0.076238	-1.713	0.086757 .
hemoglobin_final8	-0.028763	0.064670	-0.445	0.656483
hemoglobin_final9	0.117177	0.058395	2.007	0.044791 *
hemoglobin_finalUNK	0.227177	0.058008	3.916	8.99e-05 ***
cre_change1	0.368983	0.059398	6.212	5.23e-10 ***
cre_changeUNK	0.248847	0.046275	5.378	7.55e-08 ***
glu_change1	-0.047160	0.047289	-0.997	0.318636
glu_changeUNK	0.267903	0.052795	5.074	3.89e-07 ***
hco3_change1	-0.306994	0.047746	-6.430	1.28e-10 ***
hco3_changeUNK	0.059974	0.045414	1.321	0.186634
hto_change1	0.224720	0.077915	2.884	0.003925 **
hto_changeUNK	0.327033	0.044979	7.271	3.58e-13 ***
sodium_final1	0.088414	0.161396	0.548	0.583826
sodium_final2	-0.161568	0.160243	-1.008	0.313325
sodium_final3	0.285016	0.219205	1.300	0.193524
sodium_finalUNK	0.015788	0.164747	0.096	0.923655
wbc_change1	-0.028526	0.067150	-0.425	0.670978
wbc_changeUNK	0.109468	0.044819	2.442	0.014589 *
wbc_final1	-0.341208	0.193267	-1.765	0.077483 .
wbc_final2	-0.651868	0.173515	-3.757	0.000172 ***
wbc_final3	-0.124508	0.174979	-0.712	0.476740
wbc_final4	-0.242576	0.176130	-1.377	0.168435
wbc_finalUNK	-0.390489	0.175607	-2.224	0.026172 *
platelets_cat1	-1.049315	0.087794	-11.952	< 2e-16 ***
platelets_cat2	-0.992360	0.088773	-11.179	< 2e-16 ***
platelets_cat3	-0.861586	0.092518	-9.313	< 2e-16 ***
platelets_cat4	-0.708950	0.109208	-6.492	8.48e-11 ***
platelets_cat5	-0.801664	0.165730	-4.837	1.32e-06 ***
platelets_cat6	-0.895501	0.201373	-4.447	8.71e-06 ***
platelets_catUNK	-0.728359	0.092188	-7.901	2.77e-15 ***
potassium_cat3	0.089423	0.158185	0.565	0.571865
potassium_cat4	-0.063074	0.156878	-0.402	0.687640
potassium_cat5	-0.054826	0.157724	-0.348	0.728136
potassium_cat6	0.175294	0.165861	1.057	0.290571
potassium_cat7	-0.011544	0.194478	-0.059	0.952667
potassium_catUNK	0.196534	0.159855	1.229	0.218903
unk_cnt	0.054005	0.005321	10.150	< 2e-16 ***
unk_yes	-0.706476	0.056061	-12.602	< 2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 84675 on 61079 degrees of freedom

Residual deviance: 33883 on 60939 degrees of freedom

AIC: 34165

Number of Fisher Scoring iterations: 7

Hide

vif(model2A)

	GVIF	Df	GVIF <sup>1/(2*Df)</sup>
bmi	1.120276	1	1.058431
a3j_diag	2.730478	4	1.133783
arf_apache	1.097514	1	1.047623
gcs_sum	1.334885	1	1.155372
intubated_apache	1.286790	1	1.134368
ventilated_apache	1.402837	1	1.184414
map_risk	1.494281	1	1.222408
age	1.250486	1	1.118252
elective_surgery	1.886079	1	1.373346
icu_type	1.396681	7	1.024151
pre_icu_los_days	1.071441	1	1.035105
cirrhosis	1.306486	1	1.143016
hepatic_failure	1.324797	1	1.150998
lymphoma	1.042123	1	1.020844
solid_tumor_with_metastasis	1.164491	1	1.079116
comor_sum	1.540108	1	1.241011
map_change	1.726790	2	1.146331
map_final	2.780874	9	1.058466
hr_change	2.569228	2	1.266049
hr_final	2.952607	8	1.070010
rr_change	2.691797	2	1.280886
rr_final	2.963163	8	1.070249
temp_change	1.948185	1	1.395774
temp_final	1.313668	6	1.022996
spo2_final	1.244997	1	1.115794
spo2_change	1.468187	2	1.100767
bun_final	3.830941	5	1.143749
cre_final	4.234214	4	1.197696
glu_final	4.178717	5	1.153730
hco3_final	3.527314	5	1.134343
hto_final	5.215004	3	1.316869
calcium_cat	2.529848	5	1.097260
hemoglobin_final	6.185854	13	1.072602
cre_change	2.510128	2	1.258705
glu_change	3.891996	2	1.404569
hco3_change	3.075648	2	1.324293
hto_change	2.276621	2	1.228352
sodium_final	2.333432	4	1.111730
wbc_change	3.076341	2	1.324368
wbc_final	3.482913	5	1.132907
platelets_cat	2.619465	7	1.071204
potassium_cat	2.540878	6	1.080808
unk_cnt	5.411819	1	2.326332
unk_yes	3.192566	1	1.786775

Hide

timestamp()

##----- Sat Apr 11 13:41:19 2020 -----##

Relative Importance of Features

Hide

```
imp = as.data.frame(varImp(model2A))
imp = data.frame(overall = imp$Overall, names = rownames(imp))

imp[order(imp$overall,decreasing = T),]
```

	overall	names
	<dbl>	<fctr>
7	47.55266095	gcs_sum
11	34.92115521	age
9	24.58793834	ventilated_apache1
49	21.90487966	rr_final1
64	20.16334042	spo2_final
50	19.55453016	rr_final2
29	18.29118519	map_final2
32	17.82256010	map_final5
28	16.01913618	map_final1
51	15.94468516	rr_final3

1-10 of 140 rows

Previous123456...14Next

Results on Train set

Hide

```
trainPredict = predict(model2A, newdata = data.balance, type = 'response')

p_class = ifelse(trainPredict > 0.5, 1, 0)

matrix_table = table(data.balance$hospital_death, p_class)
matrix_table
```

	p_class	
	0	1
No	27215	3325
Yes	3696	26844

Hide

```
# Accuracy
accuracy = sum(diag(matrix_table))/sum(matrix_table)
round(accuracy, 3)
```

```
[1] 0.885
```

Performance on Train set

Hide

```

pred = prediction(trainPredict, data.balance$hospital_death)

auc.log = performance(pred,"auc"); au_log = as.numeric(auc.log@y.values)
au_log

```

```
[1] 0.9531957
```

Hide

```

acc.perf = performance(pred, measure = "acc")

ind = which.max( slot(acc.perf, "y.values")[[1]] )
acc = slot(acc.perf, "y.values")[[1]][ind]
cutoff = slot(acc.perf, "x.values")[[1]][ind]
print(c(accuracy= acc, cutoff = cutoff))

```

```

accuracy cutoff.2762
0.8851506    0.4994677

```

### Results on Test set

Hide

```

write.csv(dataPredict, "Data Predict.csv")
write.csv(data, "Data Original.csv")

```

### Performance on Test set (overfit train data)

Hide

```

pred = prediction(testPredict, test$hospital_death)

auc.log = performance(pred,"auc"); au_log = as.numeric(auc.log@y.values)
au_log

```

```
[1] 0.8382478
```

Hide

```

acc.perf = performance(pred, measure = "acc")

ind = which.max( slot(acc.perf, "y.values")[[1]] )
acc = slot(acc.perf, "y.values")[[1]][ind]
cutoff = slot(acc.perf, "x.values")[[1]][ind]
print(c(accuracy= acc, cutoff = cutoff))

```

```

accuracy cutoff.66804
0.9261870    0.9536216

```

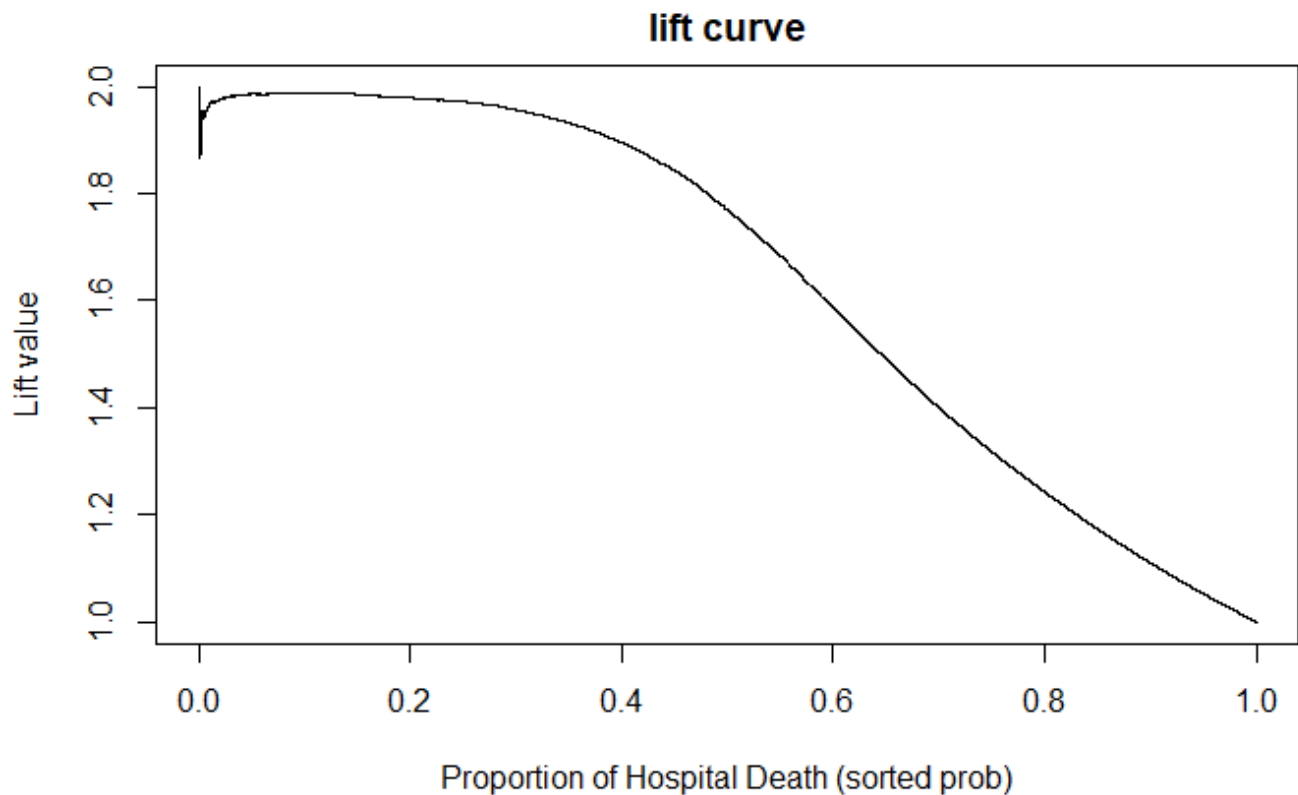
## More information on the model

- Lift Chart

Hide

```
pred = prediction(trainPredict, data.balance$hospital_death)

perf = performance(pred, "lift", "rpp" )
plot(perf, main="lift curve", xlab = 'Proportion of Hospital Death (sorted prob)')
```



- p-value

Hide

```
with(model2A, pchisq(null.deviance - deviance,
                     df.null - df.residual, lower.tail=F))
```

```
[1] 0
```

Hide

```
confusionMatrix(factor(p_class, levels = c(0,1), labels = c("No", "Yes")), test$hospital_deat  
h, positive = "Yes")
```

## Confusion Matrix and Statistics

	Reference	
Prediction	No	Yes
No	21778	1012
Yes	2282	1170

Accuracy : 0.8745

95% CI : (0.8704, 0.8785)

No Information Rate : 0.9169

P-Value [Acc &gt; NIR] : 1

Kappa : 0.349

McNemar's Test P-Value : &lt;2e-16

Sensitivity : 0.53621

Specificity : 0.90515

Pos Pred Value : 0.33893

Neg Pred Value : 0.95559

Prevalence : 0.08315

Detection Rate : 0.04459

Detection Prevalence : 0.13154

Balanced Accuracy : 0.72068

'Positive' Class : Yes

- cut-off

Hide

```
#trainPredict = predict(glmFit, newdata = data.balance)#, type = 'response')
trainPredict = glmFit$finalModel$fitted.values

p_class = ifelse(trainPredict > 0.5, 1, 0)

matrix_table = table(data.balance$hospital_death, p_class)
#matrix_table = table(data.balance$hospital_death, trainPredict)
matrix_table
```

	p_class	
	0	1
No	27274	3266
Yes	3592	26948

Hide

```
# Accuracy
accuracy = sum(diag(matrix_table))/sum(matrix_table)
round(accuracy, 3)
```

[1] 0.888

- optimal cut-offs (same or different costs)

Hide

```
cost.perf = performance(pred, "cost")  
pred@cutoffs[[1]][which.min(cost.perf@y.values[[1]])]
```

```
2762  
0.4994677
```

Hide

```
cost.perf = performance(pred, "cost", cost.fp = 4, cost.fn = 1)  
pred@cutoffs[[1]][which.min(cost.perf@y.values[[1]])]
```

```
226641  
0.8082511
```

Hide

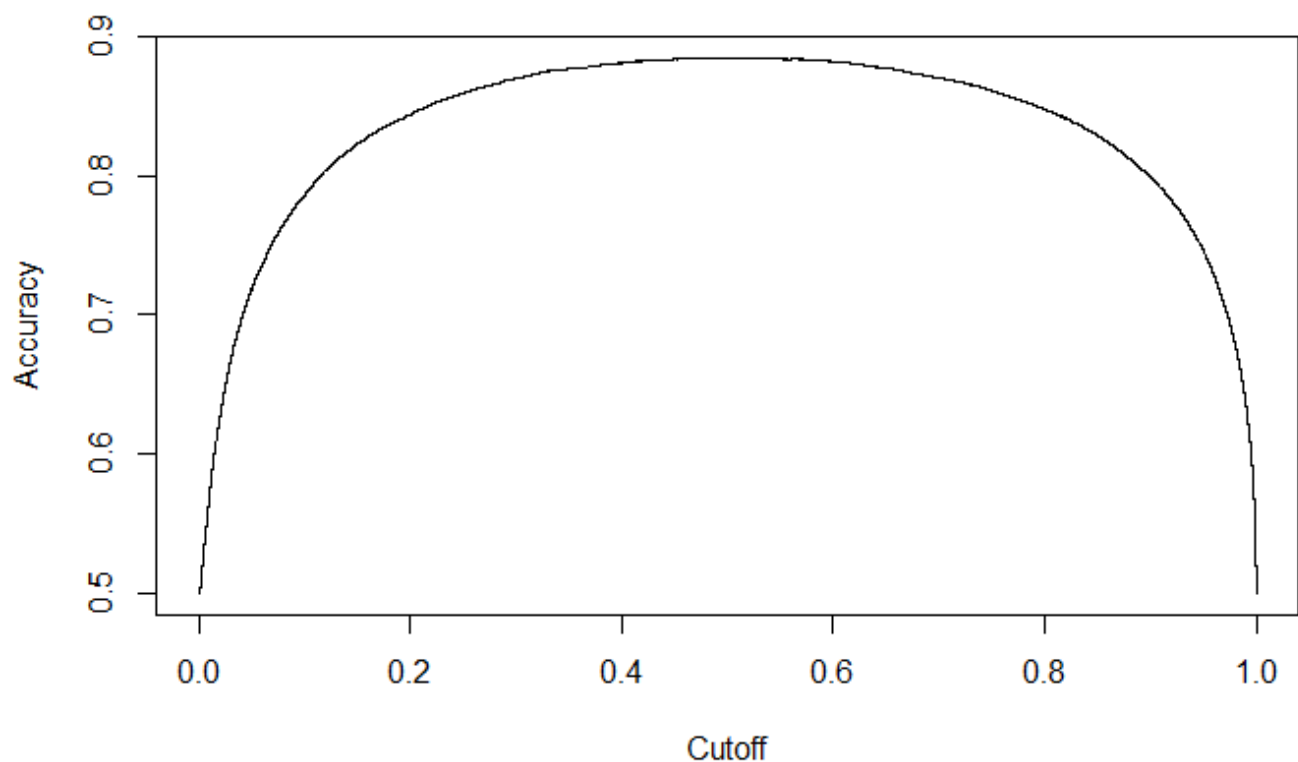
```
auc.log = performance(pred, "auc"); au_log = as.numeric(auc.log@y.values)  
au_log
```

```
[1] 0.9531957
```

- optimal accuracy

Hide

```
acc.perf = performance(pred, measure = "acc")  
plot(acc.perf)
```



Hide

```
ind = which.max( slot(acc.perf, "y.values")[[1]] )
acc = slot(acc.perf, "y.values")[[1]][ind]
cutoff = slot(acc.perf, "x.values")[[1]][ind]
print(c(accuracy= acc, cutoff = cutoff))
```

```
accuracy cutoff.2762
0.8851506  0.4994677
```

## Results on Entire dataset

Hide

```
dataPredict = predict(model2A, newdata = data, type = 'response')

p_class = ifelse(dataPredict > 0.577, 1, 0)

matrix_table = table(data$hospital_death, p_class)
matrix_table
```

```
p_class
      0      1
No  73202  6998
Yes   3417  3855
```

Hide

```
# Accuracy
accuracy = sum(diag(matrix_table))/sum(matrix_table)
round(accuracy, 3)
```

```
[1] 0.881
```

# Logistic Regression using datasets from Other Sampling Methods

## Train-Over dataset

Hide

```
train = read.csv("train_over.csv")
test = read.csv("test.csv")
```

## Data Preparation

Hide



```

train$X = NULL
test$X = NULL

train$comor_sum = train$aids + train$cirrhosis + train$diabetes_mellitus + train$hepatic_failure +
  train$immunosuppression + train$leukemia + train$lymphoma + train$solid_tumor_with_metastasis
test$comor_sum = test$aids + test$cirrhosis + test$diabetes_mellitus + test$hepatic_failure +
  test$immunosuppression + test$leukemia + test$lymphoma + test$solid_tumor_with_metastasis

factor.list = c('a2_diag', 'a3j_diag', 'apache_post_operative', 'arf_apache', 'intubated_apache',
  'ventilated_apache', 'elective_surgery', 'aids', 'cirrhosis', 'diabetes_mellitus',
  'hepatic_failure', 'immunosuppression', 'leukemia', 'lymphoma', 'solid_tumor_with_metastasis',
  'gcs_eyes', 'gcs_motor', 'gcs_verbal')
train[, factor.list] = lapply(train[, factor.list], as.factor)
train$hospital_death = factor(train$hospital_death, levels = c(0,1), labels = c("No", "Yes"))
test[, factor.list] = lapply(test[, factor.list], as.factor)
test$hospital_death = factor(test$hospital_death, levels = c(0,1), labels = c("No", "Yes"))

```

Check the target variable

Hide

```
contrasts(train$hospital_death)
```

```

      Yes
No      0
Yes     1

```

Hide

```
table(train$hospital_death)
```

```

  No   Yes
51350 60930

```

Hide

```
prop.table(table(train$hospital_death))
```

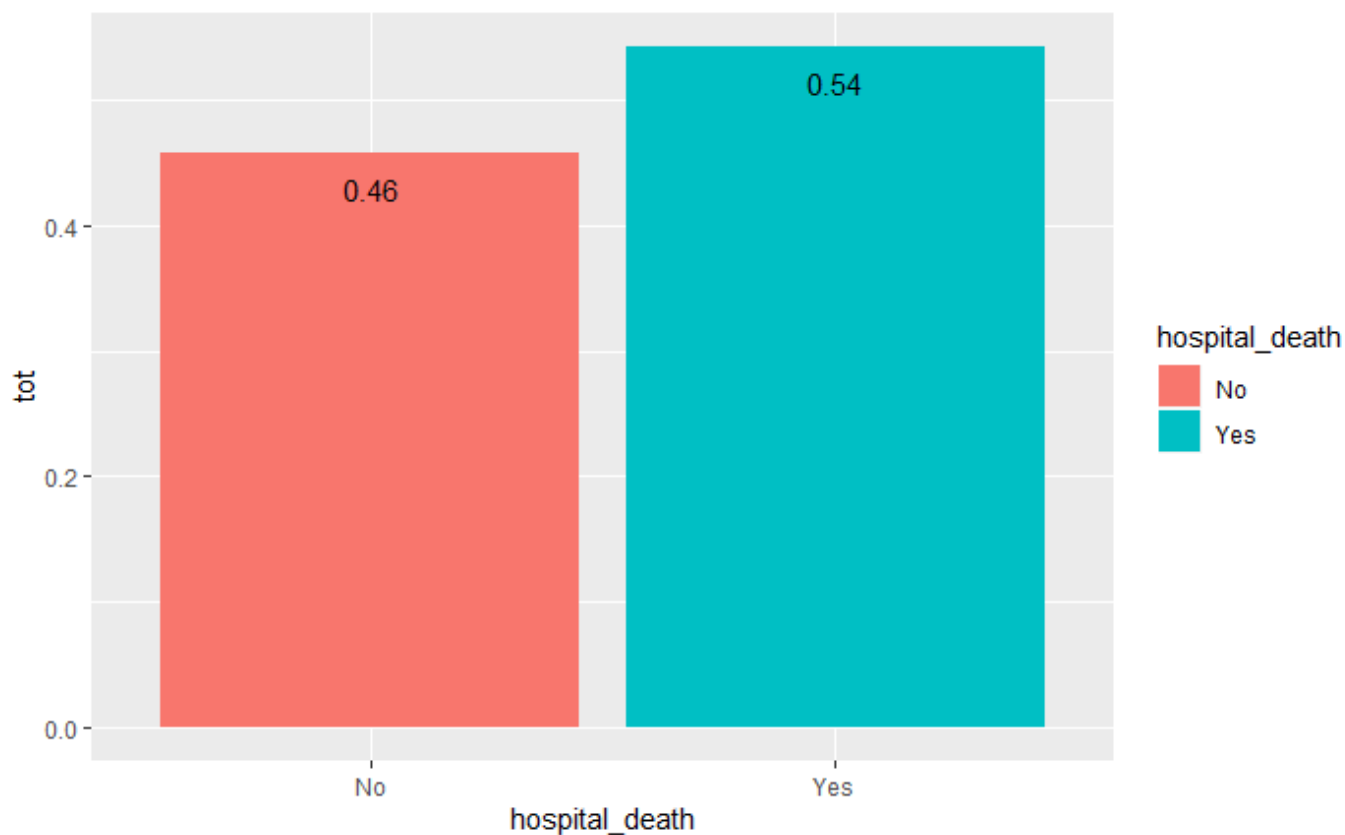
```

      No      Yes
0.4573388 0.5426612

```

Hide

```
train %>%  
  group_by(hospital_death)%>%  
  summarise(tot = n()/nrow(train))%>%  
  ggplot(aes(x=hospital_death, y=tot, fill = hospital_death)) +  
  geom_bar(stat='identity') +  
  geom_text(aes(label = round(tot, 2)), vjust = 2)
```



## Initial Model

[Hide](#)

```
model3 = glm(hospital_death ~ . -gcs_motor -gcs_verbal -gcs_eyes  
  -apache_2_bodysystem -apache_3j_bodysystem -a2_diag  
  -diabetes_mellitus -hospital_admit_source  
  -apache_post_operative -hto_final -cre_final -potassium_cat -sodium_final,  
  #-ethnicity -wbc_final,  
  data = train, family = binomial)  
summary(model3)
```

Call:

```
glm(formula = hospital_death ~ . - gcs_motor - gcs_verbal - gcs_eyes -
     apache_2_bodysystem - apache_3j_bodysystem - a2_diag - diabetes_mellitus -
     hospital_admit_source - apache_post_operative - hto_final -
     cre_final - potassium_cat - sodium_final, family = binomial,
     data = train)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-3.8925	-0.6490	0.1372	0.6564	2.6303

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	5.741e+00	2.290e-01	25.068	< 2e-16 ***
a3j_diag1	-1.690e+00	1.470e-01	-11.496	< 2e-16 ***
a3j_diag2	-1.685e+00	1.468e-01	-11.478	< 2e-16 ***
a3j_diag3	-2.203e+00	1.475e-01	-14.935	< 2e-16 ***
a3j_diag4	-1.670e+00	1.565e-01	-10.672	< 2e-16 ***
arf_apache1	4.478e-01	4.405e-02	10.166	< 2e-16 ***
gcs_sum	-8.782e-02	2.629e-03	-33.402	< 2e-16 ***
intubated_apache1	-8.551e-02	2.539e-02	-3.368	0.000757 ***
ventilated_apache1	7.411e-01	2.194e-02	33.774	< 2e-16 ***
map_risk	2.277e-02	1.558e-03	14.620	< 2e-16 ***
age	2.912e-02	6.177e-04	47.145	< 2e-16 ***
elective_surgery1	-6.149e-01	5.232e-02	-11.754	< 2e-16 ***
ethnicityAsian	3.626e-04	7.920e-02	0.005	0.996347
ethnicityCaucasian	6.627e-02	2.834e-02	2.338	0.019378 *
ethnicityHispanic	2.474e-01	4.791e-02	5.165	2.41e-07 ***
ethnicityNative American	8.684e-02	9.829e-02	0.883	0.376977
ethnicityOther/Unknown	-4.215e-02	4.420e-02	-0.954	0.340297
genderM	5.620e-02	1.671e-02	3.363	0.000770 ***
icu_admit_sourceFloor	2.147e-01	2.284e-02	9.398	< 2e-16 ***
icu_admit_sourceOperating Room / Recovery	-4.481e-01	5.381e-02	-8.327	< 2e-16 ***
icu_admit_sourceOther Hospital	6.412e-01	4.493e-02	14.269	< 2e-16 ***
icu_admit_sourceOther ICU	2.068e+00	1.018e-01	20.307	< 2e-16 ***
icu_admit_sourceUnknown	1.051e+00	3.201e-01	3.282	0.001031 **
icu_stay_typereadmit	5.256e-02	9.939e-02	0.529	0.596918
icu_stay_typetransfer	-4.147e-01	3.586e-02	-11.565	< 2e-16 ***
icu_typeCCU-CTICU	3.317e-03	4.459e-02	0.074	0.940692
icu_typeCSICU	-2.085e-01	5.554e-02	-3.754	0.000174 ***
icu_typeCTICU	-2.928e-01	5.513e-02	-5.311	1.09e-07 ***
icu_typeMed-Surg ICU	-7.686e-02	3.629e-02	-2.118	0.034156 *
icu_typeMICU	-4.141e-03	4.298e-02	-0.096	0.923259
icu_typeNeuro ICU	4.066e-01	4.558e-02	8.920	< 2e-16 ***
icu_typeSICU	7.348e-02	4.923e-02	1.493	0.135559
pre_icu_los_days	3.683e-02	3.520e-03	10.461	< 2e-16 ***
apache_4a_death_prob	1.823e+00	5.643e-02	32.307	< 2e-16 ***
aids1	-2.573e-01	2.812e-01	-0.915	0.360169
cirrhosis1	5.351e-01	7.172e-02	7.461	8.61e-14 ***
hepatic_failure1	4.508e-01	7.653e-02	5.890	3.86e-09 ***
immunosuppression1	5.231e-01	4.993e-02	10.477	< 2e-16 ***
leukemia1	1.550e-01	8.694e-02	1.783	0.074589 .
lymphoma1	6.867e-01	1.114e-01	6.162	7.16e-10 ***
solid_tumor_with_metastasis1	1.025e+00	5.486e-02	18.687	< 2e-16 ***
mbp_final	-4.186e-02	4.093e-03	-10.226	< 2e-16 ***
hr_final1	-9.047e-01	6.557e-02	-13.797	< 2e-16 ***

hr_final2	-8.815e-01	5.270e-02	-16.727	< 2e-16	***
hr_final3	-5.806e-01	5.452e-02	-10.649	< 2e-16	***
hr_final4	-4.244e-01	5.490e-02	-7.729	1.08e-14	***
hr_final5	-2.364e-01	5.417e-02	-4.364	1.28e-05	***
hr_final6	-1.203e-01	6.135e-02	-1.960	0.049944	*
hr_final7	2.926e-02	6.766e-02	0.433	0.665375	
hr_finalunknown	-1.098e+01	4.075e+01	-0.269	0.787598	
rr_final1	-1.518e+00	4.956e-02	-30.631	< 2e-16	***
rr_final2	-1.473e+00	5.057e-02	-29.132	< 2e-16	***
rr_final3	-1.486e+00	5.228e-02	-28.417	< 2e-16	***
rr_final4	-1.122e+00	4.963e-02	-22.606	< 2e-16	***
rr_final5	-9.821e-01	5.295e-02	-18.546	< 2e-16	***
rr_final6	-9.320e-01	5.442e-02	-17.125	< 2e-16	***
rr_final7	-1.001e+00	5.826e-02	-17.186	< 2e-16	***
rr_finalunknown	6.790e-01	1.744e-01	3.892	9.93e-05	***
temp_final1	2.947e-01	1.408e-01	2.093	0.036356	*
temp_final2	-2.215e-01	1.318e-01	-1.681	0.092857	.
temp_final3	-4.249e-01	7.835e-02	-5.424	5.84e-08	***
temp_final4	-6.538e-01	6.644e-02	-9.840	< 2e-16	***
temp_final5	-9.849e-01	6.451e-02	-15.267	< 2e-16	***
temp_finalunknown	-7.567e-01	8.149e-02	-9.286	< 2e-16	***
spo2_final	-1.674e-02	7.262e-04	-23.057	< 2e-16	***
bun_final1	1.367e-01	3.740e-02	3.654	0.000258	***
bun_final2	5.066e-01	2.196e-02	23.067	< 2e-16	***
bun_final3	7.248e-01	2.733e-02	26.517	< 2e-16	***
bun_final4	8.104e-01	4.457e-02	18.184	< 2e-16	***
bun_finalunknown	4.498e-01	6.604e-02	6.810	9.74e-12	***
glu_final1	1.517e-01	4.206e-02	3.606	0.000311	***
glu_final2	6.177e-02	4.091e-02	1.510	0.131089	
glu_final3	3.401e-01	6.294e-02	5.403	6.55e-08	***
glu_final4	4.284e-01	1.006e-01	4.257	2.07e-05	***
glu_finalunknown	1.930e-01	6.390e-02	3.021	0.002523	**
hco3_final3	-2.382e-01	3.627e-02	-6.568	5.10e-11	***
hco3_final4	-3.417e-03	3.932e-02	-0.087	0.930747	
hco3_final5	1.953e-01	4.787e-02	4.079	4.52e-05	***
hco3_final6	3.965e-01	5.176e-02	7.660	1.86e-14	***
hco3_finalunknown	-2.313e-01	5.196e-02	-4.451	8.53e-06	***
calcium_cat5	-3.190e-01	1.110e-01	-2.873	0.004060	**
calcium_cat6	-4.606e-01	8.265e-02	-5.573	2.50e-08	***
calcium_cat7	-6.561e-01	7.660e-02	-8.566	< 2e-16	***
calcium_cat8	-6.865e-01	7.553e-02	-9.089	< 2e-16	***
calcium_cat9	-6.556e-01	7.820e-02	-8.383	< 2e-16	***
calcium_catunknown	-6.098e-01	8.587e-02	-7.102	1.23e-12	***
hemaglobin_cat2	4.394e-02	3.284e-02	1.338	0.180929	
hemaglobin_cat3	-1.056e-01	3.345e-02	-3.158	0.001591	**
hemaglobin_cat4	-1.598e-01	3.741e-02	-4.270	1.95e-05	***
hemaglobin_cat5	1.923e-01	5.595e-02	3.437	0.000589	***
hemaglobin_catunknown	-2.389e-02	6.689e-02	-0.357	0.721023	
wbc_final1	-1.159e-01	1.131e-01	-1.025	0.305323	
wbc_final2	-3.541e-01	1.002e-01	-3.534	0.000410	***
wbc_final3	-1.682e-02	1.044e-01	-0.161	0.872039	
wbc_final4	4.419e-02	1.046e-01	0.423	0.672577	
wbc_finalunknown	-3.212e-01	1.178e-01	-2.726	0.006412	**
platelets_cat1	-9.633e-01	5.224e-02	-18.441	< 2e-16	***
platelets_cat2	-1.046e+00	5.290e-02	-19.775	< 2e-16	***
platelets_cat3	-9.022e-01	5.478e-02	-16.469	< 2e-16	***
platelets_cat4	-7.616e-01	6.327e-02	-12.038	< 2e-16	***
platelets_cat5	-1.012e+00	8.756e-02	-11.554	< 2e-16	***

```

platelets_cat6                -7.817e-01  1.062e-01  -7.364 1.79e-13 ***
platelets_catunknown          -7.743e-01  8.737e-02  -8.862 < 2e-16 ***
comor_sum                     -2.623e-01  2.126e-02 -12.338 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 154835 on 112279 degrees of freedom  
 Residual deviance: 95264 on 112176 degrees of freedom  
 AIC: 95472

Number of Fisher Scoring iterations: 10

[Hide](#)

```
vif(model3)
```

	GVIF	Df	GVIF^(1/(2*Df))
a3j_diag	12.305576	4	1.368557
arf_apache	1.090547	1	1.044292
gcs_sum	1.871876	1	1.368165
intubated_apache	1.608246	1	1.268167
ventilated_apache	1.838753	1	1.356006
map_risk	1.408894	1	1.186968
age	1.347933	1	1.161005
elective_surgery	4.536317	1	2.129863
ethnicity	1.205983	5	1.018906
gender	1.068791	1	1.033824
icu_admit_source	11.739456	5	1.279278
icu_stay_type	1.093182	2	1.022523
icu_type	1.656259	7	1.036697
pre_icu_los_days	1.319881	1	1.148861
apache_4a_death_prob	2.770581	1	1.664506
aids	1.018336	1	1.009126
cirrhosis	1.584770	1	1.258877
hepatic_failure	1.568997	1	1.252596
immunosuppression	1.424242	1	1.193416
leukemia	1.133439	1	1.064631
lymphoma	1.076187	1	1.037394
solid_tumor_with_metastasis	1.320739	1	1.149234
mbp_final	1.413162	1	1.188765
hr_final	1.324686	8	1.017729
rr_final	1.594407	8	1.029586
temp_final	1.396352	6	1.028213
spo2_final	1.066357	1	1.032646
bun_final	8.541300	5	1.239231
glu_final	2.822681	5	1.109344
hco3_final	4.603339	5	1.164950
calcium_cat	5.150988	6	1.146368
hemaglobin_cat	8.811160	5	1.243092
wbc_final	9.181506	5	1.248221
platelets_cat	14.497278	7	1.210456
comor_sum	2.357700	1	1.535480

# Auto Stepwise Feature Selection

[Hide](#)

```
model3A = stepAIC(model3, trace = F)
summary(model3A)
```

Call:

```
glm(formula = hospital_death ~ a3j_diag + arf_apache + gcs_sum +
    intubated_apache + ventilated_apache + map_risk + age + elective_surgery +
    ethnicity + gender + icu_admit_source + icu_stay_type + icu_type +
    pre_icu_los_days + apache_4a_death_prob + cirrhosis + hepatic_failure +
    immunosuppression + leukemia + lymphoma + solid_tumor_with_metastasis +
    mbp_final + hr_final + rr_final + temp_final + spo2_final +
    bun_final + glu_final + hco3_final + calcium_cat + hemaglobin_cat +
    wbc_final + platelets_cat + comor_sum, family = binomial,
    data = train)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-3.8922	-0.6489	0.1371	0.6566	2.6305

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	5.739e+00	2.290e-01	25.063	< 2e-16 ***
a3j_diag1	-1.689e+00	1.470e-01	-11.489	< 2e-16 ***
a3j_diag2	-1.684e+00	1.468e-01	-11.473	< 2e-16 ***
a3j_diag3	-2.202e+00	1.475e-01	-14.928	< 2e-16 ***
a3j_diag4	-1.669e+00	1.565e-01	-10.665	< 2e-16 ***
arf_apache1	4.484e-01	4.404e-02	10.181	< 2e-16 ***
gcs_sum	-8.780e-02	2.629e-03	-33.397	< 2e-16 ***
intubated_apache1	-8.567e-02	2.539e-02	-3.374	0.000741 ***
ventilated_apache1	7.411e-01	2.194e-02	33.777	< 2e-16 ***
map_risk	2.277e-02	1.558e-03	14.617	< 2e-16 ***
age	2.914e-02	6.174e-04	47.201	< 2e-16 ***
elective_surgery1	-6.152e-01	5.231e-02	-11.759	< 2e-16 ***
ethnicityAsian	8.020e-04	7.919e-02	0.010	0.991920
ethnicityCaucasian	6.654e-02	2.834e-02	2.348	0.018883 *
ethnicityHispanic	2.468e-01	4.791e-02	5.152	2.58e-07 ***
ethnicityNative American	8.773e-02	9.828e-02	0.893	0.372054
ethnicityOther/Unknown	-4.183e-02	4.420e-02	-0.946	0.344013
genderM	5.611e-02	1.671e-02	3.358	0.000785 ***
icu_admit_sourceFloor	2.146e-01	2.284e-02	9.398	< 2e-16 ***
icu_admit_sourceOperating Room / Recovery	-4.481e-01	5.381e-02	-8.328	< 2e-16 ***
icu_admit_sourceOther Hospital	6.402e-01	4.493e-02	14.249	< 2e-16 ***
icu_admit_sourceOther ICU	2.067e+00	1.018e-01	20.301	< 2e-16 ***
icu_admit_sourceUnknown	1.050e+00	3.201e-01	3.279	0.001043 **
icu_stay_typereadmit	5.195e-02	9.937e-02	0.523	0.601103
icu_stay_typetransfer	-4.145e-01	3.586e-02	-11.561	< 2e-16 ***
icu_typeCCU-CTICU	3.650e-03	4.459e-02	0.082	0.934754
icu_typeCSICU	-2.079e-01	5.553e-02	-3.744	0.000181 ***
icu_typeCTICU	-2.924e-01	5.513e-02	-5.303	1.14e-07 ***
icu_typeMed-Surg ICU	-7.648e-02	3.629e-02	-2.108	0.035069 *
icu_typeMICU	-4.035e-03	4.299e-02	-0.094	0.925209
icu_typeNeuro ICU	4.070e-01	4.558e-02	8.929	< 2e-16 ***
icu_typeSICU	7.406e-02	4.923e-02	1.504	0.132464
pre_icu_los_days	3.683e-02	3.520e-03	10.462	< 2e-16 ***
apache_4a_death_prob	1.822e+00	5.642e-02	32.298	< 2e-16 ***
cirrhosis1	5.363e-01	7.170e-02	7.480	7.46e-14 ***
hepatic_failure1	4.525e-01	7.651e-02	5.914	3.33e-09 ***
immunosuppression1	5.242e-01	4.991e-02	10.503	< 2e-16 ***
leukemia1	1.570e-01	8.691e-02	1.807	0.070789 .
lymphoma1	6.845e-01	1.114e-01	6.143	8.12e-10 ***

solid_tumor_with_metastasis1	1.027e+00	5.484e-02	18.724	< 2e-16	***
mbp_final	-4.185e-02	4.093e-03	-10.225	< 2e-16	***
hr_final1	-9.043e-01	6.557e-02	-13.790	< 2e-16	***
hr_final2	-8.811e-01	5.271e-02	-16.718	< 2e-16	***
hr_final3	-5.800e-01	5.452e-02	-10.638	< 2e-16	***
hr_final4	-4.238e-01	5.490e-02	-7.719	1.17e-14	***
hr_final5	-2.359e-01	5.417e-02	-4.354	1.33e-05	***
hr_final6	-1.202e-01	6.136e-02	-1.959	0.050114	.
hr_final7	2.999e-02	6.766e-02	0.443	0.657623	
hr_finalunknown	-1.098e+01	4.075e+01	-0.269	0.787609	
rr_final1	-1.518e+00	4.956e-02	-30.634	< 2e-16	***
rr_final2	-1.473e+00	5.057e-02	-29.138	< 2e-16	***
rr_final3	-1.486e+00	5.228e-02	-28.421	< 2e-16	***
rr_final4	-1.122e+00	4.963e-02	-22.617	< 2e-16	***
rr_final5	-9.824e-01	5.295e-02	-18.552	< 2e-16	***
rr_final6	-9.321e-01	5.442e-02	-17.129	< 2e-16	***
rr_final7	-1.002e+00	5.826e-02	-17.196	< 2e-16	***
rr_finalunknown	6.778e-01	1.744e-01	3.885	0.000102	***
temp_final1	2.944e-01	1.408e-01	2.091	0.036564	*
temp_final2	-2.219e-01	1.318e-01	-1.683	0.092283	.
temp_final3	-4.257e-01	7.834e-02	-5.433	5.53e-08	***
temp_final4	-6.549e-01	6.643e-02	-9.858	< 2e-16	***
temp_final5	-9.857e-01	6.451e-02	-15.281	< 2e-16	***
temp_finalunknown	-7.572e-01	8.148e-02	-9.293	< 2e-16	***
spo2_final	-1.675e-02	7.262e-04	-23.060	< 2e-16	***
bun_final1	1.367e-01	3.740e-02	3.654	0.000258	***
bun_final2	5.066e-01	2.196e-02	23.068	< 2e-16	***
bun_final3	7.248e-01	2.733e-02	26.516	< 2e-16	***
bun_final4	8.097e-01	4.456e-02	18.172	< 2e-16	***
bun_finalunknown	4.488e-01	6.603e-02	6.796	1.07e-11	***
glu_final1	1.512e-01	4.205e-02	3.595	0.000325	***
glu_final2	6.083e-02	4.090e-02	1.487	0.136904	
glu_final3	3.397e-01	6.294e-02	5.398	6.75e-08	***
glu_final4	4.283e-01	1.006e-01	4.256	2.08e-05	***
glu_finalunknown	1.928e-01	6.391e-02	3.016	0.002558	**
hco3_final3	-2.382e-01	3.627e-02	-6.568	5.10e-11	***
hco3_final4	-3.289e-03	3.932e-02	-0.084	0.933321	
hco3_final5	1.947e-01	4.786e-02	4.068	4.74e-05	***
hco3_final6	3.963e-01	5.177e-02	7.655	1.93e-14	***
hco3_finalunknown	-2.312e-01	5.195e-02	-4.450	8.60e-06	***
calcium_cat5	-3.190e-01	1.110e-01	-2.873	0.004060	**
calcium_cat6	-4.615e-01	8.265e-02	-5.584	2.35e-08	***
calcium_cat7	-6.561e-01	7.660e-02	-8.566	< 2e-16	***
calcium_cat8	-6.865e-01	7.552e-02	-9.089	< 2e-16	***
calcium_cat9	-6.555e-01	7.820e-02	-8.382	< 2e-16	***
calcium_catunknown	-6.095e-01	8.587e-02	-7.099	1.26e-12	***
hemaglobin_cat2	4.377e-02	3.284e-02	1.333	0.182625	
hemaglobin_cat3	-1.058e-01	3.345e-02	-3.162	0.001566	**
hemaglobin_cat4	-1.598e-01	3.741e-02	-4.272	1.93e-05	***
hemaglobin_cat5	1.923e-01	5.595e-02	3.437	0.000588	***
hemaglobin_catunknown	-2.245e-02	6.689e-02	-0.336	0.737175	
wbc_final1	-1.170e-01	1.131e-01	-1.035	0.300553	
wbc_final2	-3.541e-01	1.002e-01	-3.534	0.000410	***
wbc_final3	-1.665e-02	1.044e-01	-0.159	0.873291	
wbc_final4	4.379e-02	1.046e-01	0.419	0.675435	
wbc_finalunknown	-3.199e-01	1.179e-01	-2.715	0.006636	**
platelets_cat1	-9.629e-01	5.224e-02	-18.433	< 2e-16	***
platelets_cat2	-1.046e+00	5.290e-02	-19.768	< 2e-16	***



```

platelets_cat3          -9.019e-01  5.478e-02 -16.464 < 2e-16 ***
platelets_cat4          -7.611e-01  6.326e-02 -12.030 < 2e-16 ***
platelets_cat5          -1.011e+00  8.756e-02 -11.548 < 2e-16 ***
platelets_cat6          -7.810e-01  1.062e-01  -7.358 1.87e-13 ***
platelets_catunknown    -7.770e-01  8.737e-02  -8.894 < 2e-16 ***
comor_sum               -2.635e-01  2.121e-02 -12.419 < 2e-16 ***

```

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 154835 on 112279 degrees of freedom  
 Residual deviance: 95265 on 112177 degrees of freedom  
 AIC: 95471

Number of Fisher Scoring iterations: 10

Hide

vif(model3A)

	GVIF	Df	GVIF^(1/(2*Df))
a3j_diag	12.300059	4	1.368480
arf_apache	1.090311	1	1.044180
gcs_sum	1.871537	1	1.368041
intubated_apache	1.608259	1	1.268171
ventilated_apache	1.838707	1	1.355989
map_risk	1.408844	1	1.186947
age	1.346398	1	1.160344
elective_surgery	4.536043	1	2.129799
ethnicity	1.205122	5	1.018833
gender	1.068784	1	1.033820
icu_admit_source	11.728271	5	1.279156
icu_stay_type	1.093119	2	1.022508
icu_type	1.655555	7	1.036666
pre_icu_los_days	1.319834	1	1.148840
apache_4a_death_prob	2.769451	1	1.664167
cirrhosis	1.584136	1	1.258625
hepatic_failure	1.567946	1	1.252177
immunosuppression	1.423548	1	1.193125
leukemia	1.132740	1	1.064302
lymphoma	1.075576	1	1.037100
solid_tumor_with_metastasis	1.319601	1	1.148739
mbp_final	1.413089	1	1.188734
hr_final	1.324099	8	1.017701
rr_final	1.593608	8	1.029553
temp_final	1.395704	6	1.028173
spo2_final	1.066374	1	1.032654
bun_final	8.534940	5	1.239139
glu_final	2.820209	5	1.109247
hco3_final	4.600186	5	1.164870
calcium_cat	5.145007	6	1.146257
hemaglobin_cat	8.807143	5	1.243035
wbc_final	9.173001	5	1.248105
platelets_cat	14.478248	7	1.210343
comor_sum	2.348589	1	1.532511

Relative Importance of Features

Hide

```
imp = as.data.frame(varImp(model3A))
imp = data.frame(overall = imp$Overall, names = rownames(imp))

imp[order(imp$overall,decreasing = T),]
```

	overall	names
	<dbl>	<fctr>
10	47.20112449	age
8	33.77696915	ventilated_apache1
6	33.39679210	gcs_sum
33	32.29830340	apache_4a_death_prob
49	30.63433339	rr_final1
50	29.13819311	rr_final2
51	28.42083967	rr_final3
66	26.51616534	bun_final3
65	23.06764636	bun_final2
63	23.05961842	spo2_final

1-10 of 102 rows

Previous123456...11Next

Results on Train set

Hide

```
#trainPredict = predict(glmFit, newdata = data.balance)#, type = 'response')
trainPredict = predict(model3A, newdata = train, type = 'response')

p_class = ifelse(trainPredict > 0.54, 1, 0)

matrix_table = table(train$hospital_death, p_class)
matrix_table
```

	p_class	
	0	1
No	42032	9318
Yes	12806	48124

Hide

```
# Accuracy
accuracy = sum(diag(matrix_table))/sum(matrix_table)
round(accuracy, 3)
```

```
[1] 0.803
```

## Performance on Train set

Hide

```
pred = prediction(trainPredict, train$hospital_death)

auc.log = performance(pred,"auc"); au_log = as.numeric(auc.log@y.values)
au_log
```

```
[1] 0.8863967
```

Hide

```
acc.perf = performance(pred, measure = "acc")

ind = which.max( slot(acc.perf, "y.values")[[1]] )
acc = slot(acc.perf, "y.values")[[1]][ind]
cutoff = slot(acc.perf, "x.values")[[1]][ind]
print(c(accuracy= acc, cutoff = cutoff))
```

```
accuracy cutoff.106304
0.8087104      0.4769985
```

## Results on Test set

Hide

```
testPredict = predict(model3A, newdata = test, type = 'response')

p_class = ifelse(testPredict > 0.54 , 1, 0)

matrix_table = table(test$hospital_death, p_class)
matrix_table
```

```
p_class
      0      1
No 18079 3928
Yes  432 1602
```

Hide

```
# Accuracy
accuracy = sum(diag(matrix_table))/sum(matrix_table)
round(accuracy, 3)
```

```
[1] 0.819
```

## Performance on Test set

Hide

```
pred = prediction(testPredict, test$hospital_death)

auc.log = performance(pred,"auc"); au_log = as.numeric(auc.log@y.values)
au_log
```

```
[1] 0.8840048
```

Hide

```
acc.perf = performance(pred, measure = "acc")

ind = which.max( slot(acc.perf, "y.values")[[1]] )
acc = slot(acc.perf, "y.values")[[1]][ind]
cutoff = slot(acc.perf, "x.values")[[1]][ind]
print(c(accuracy= acc, cutoff = cutoff))
```

```
accuracy cutoff.21502
0.9287467    0.9421910
```

## Train-Rose dataset

Hide

```
train = read.csv("train_rose.csv")
```

### Data Preparation

Hide

```
train$X = NULL

train$comor_sum = train$aids + train$cirrhosis + train$diabetes_mellitus + train$hepatic_fail
ure +
  train$immunosuppression + train$leukemia + train$lymphoma + train$solid_tumor_with_metastas
is

factor.list = c('a2_diag', 'a3j_diag', 'apache_post_operative', 'arf_apache', 'intubated_apac
he',
               'ventilated_apache', 'elective_surgery', 'aids', 'cirrhosis', 'diabetes_melli
tus',
               'hepatic_failure', 'immunosuppression', 'leukemia', 'lymphoma', 'solid_tumor_
with_metastasis',
               'gcs_eyes', 'gcs_motor', 'gcs_verbal')
train[, factor.list] = lapply(train[, factor.list], as.factor)
train$hospital_death = factor(train$hospital_death, levels = c(0,1), labels = c("No", "Yes"))
```

### Check the target variable

Hide

```
contrasts(train$hospital_death)
```

```
Yes
No    0
Yes   1
```

Hide

```
table(train$hospital_death)
```

```
  No   Yes
28009 28087
```

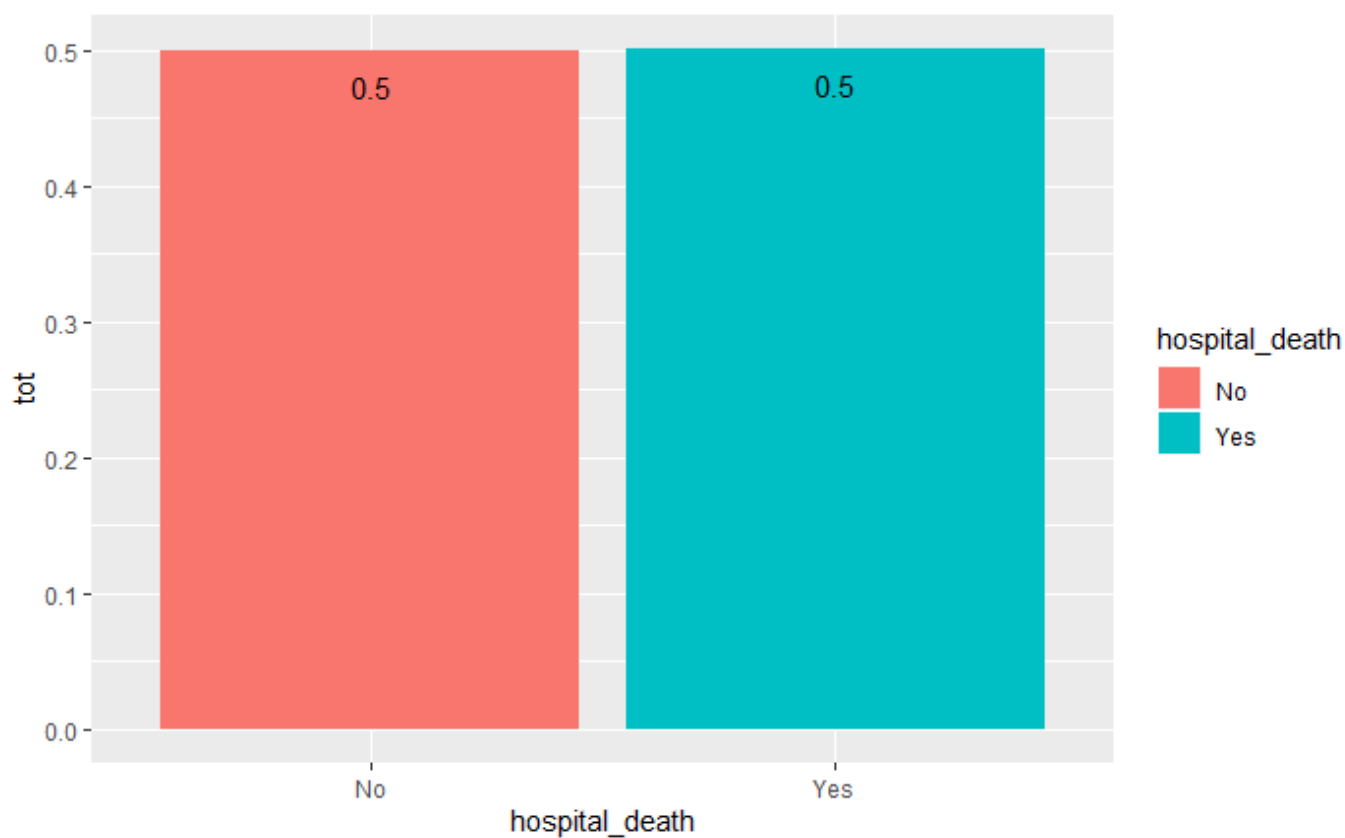
Hide

```
prop.table(table(train$hospital_death))
```

```
      No      Yes
0.4993048 0.5006952
```

Hide

```
train %>%
  group_by(hospital_death)%>%
  summarise(tot = n()/nrow(train))%>%
  ggplot(aes(x=hospital_death, y=tot, fill = hospital_death)) +
  geom_bar(stat='identity') +
  geom_text(aes(label = round(tot, 2)), vjust = 2)
```



# Initial Model

[Hide](#)

```
model4 = glm(hospital_death ~ . -gcs_motor -gcs_verbal #-gcs_eyes
             -apache_2_bodysystem -apache_3j_bodysystem -a2_diag
             -diabetes_mellitus -hospital_admit_source
             -apache_post_operative -hto_final -cre_final -potassium_cat -sodium_final,
             data = train, family = binomial)
summary(model4)
```

Call:

```
glm(formula = hospital_death ~ . - gcs_motor - gcs_verbal - apache_2_bodysystem -
  apache_3j_bodysystem - a2_diag - diabetes_mellitus - hospital_admit_source -
  apache_post_operative - hto_final - cre_final - potassium_cat -
  sodium_final, family = binomial, data = train)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-3.8377	-0.6462	0.0313	0.6455	2.8666

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	4.946e+00	3.118e-01	15.866	< 2e-16	***
a3j_diag1	-1.168e+00	1.838e-01	-6.356	2.08e-10	***
a3j_diag2	-1.180e+00	1.836e-01	-6.427	1.30e-10	***
a3j_diag3	-1.797e+00	1.854e-01	-9.697	< 2e-16	***
a3j_diag4	-1.212e+00	1.994e-01	-6.078	1.22e-09	***
arf_apache1	3.739e-01	6.303e-02	5.932	3.00e-09	***
gcs_eyes1	2.692e-01	9.733e-02	2.765	0.005684	**
gcs_eyes2	-2.771e-01	1.064e-01	-2.605	0.009191	**
gcs_eyes3	-4.811e-01	1.093e-01	-4.402	1.07e-05	***
gcs_eyes4	-5.502e-01	1.154e-01	-4.770	1.85e-06	***
gcs_sum	-3.810e-02	5.182e-03	-7.352	1.95e-13	***
intubated_apache1	-6.926e-02	3.560e-02	-1.945	0.051730	.
ventilated_apache1	7.438e-01	3.071e-02	24.219	< 2e-16	***
map_risk	2.299e-02	1.995e-03	11.524	< 2e-16	***
age	2.641e-02	7.930e-04	33.305	< 2e-16	***
elective_surgery1	-4.480e-01	7.511e-02	-5.964	2.46e-09	***
ethnicityAsian	5.882e-02	1.137e-01	0.517	0.604887	
ethnicityCaucasian	7.187e-02	4.014e-02	1.790	0.073381	.
ethnicityHispanic	2.472e-01	6.712e-02	3.683	0.000231	***
ethnicityNative American	-2.483e-02	1.346e-01	-0.184	0.853662	
ethnicityOther/Unknown	4.415e-02	6.220e-02	0.710	0.477807	
genderM	1.060e-02	2.370e-02	0.447	0.654849	
icu_admit_sourceFloor	2.821e-01	3.155e-02	8.939	< 2e-16	***
icu_admit_sourceOperating Room / Recovery	-5.704e-01	7.658e-02	-7.448	9.47e-14	***
icu_admit_sourceOther Hospital	7.439e-01	6.431e-02	11.567	< 2e-16	***
icu_admit_sourceOther ICU	1.445e+00	1.361e-01	10.615	< 2e-16	***
icu_admit_sourceUnknown	1.200e-02	4.647e-01	0.026	0.979402	
icu_stay_typereadmit	6.352e-02	1.373e-01	0.463	0.643683	
icu_stay_typetransfer	-4.070e-01	5.139e-02	-7.919	2.39e-15	***
icu_typeCCU-CTICU	-1.340e-01	6.330e-02	-2.117	0.034271	*
icu_typeCSICU	-2.883e-01	7.915e-02	-3.643	0.000270	***
icu_typeCTICU	-3.028e-01	7.747e-02	-3.908	9.30e-05	***
icu_typeMed-Surg ICU	-1.250e-01	5.085e-02	-2.458	0.013983	*
icu_typeMICU	1.637e-02	6.050e-02	0.271	0.786659	
icu_typeNeuro ICU	4.536e-01	6.400e-02	7.087	1.37e-12	***
icu_typeSICU	7.158e-02	6.889e-02	1.039	0.298737	
pre_icu_los_days	3.490e-02	4.193e-03	8.322	< 2e-16	***
apache_4a_death_prob	1.236e+00	6.363e-02	19.420	< 2e-16	***
aids1	-5.657e-01	4.087e-01	-1.384	0.166327	
cirrhosis1	7.938e-01	1.004e-01	7.903	2.72e-15	***
hepatic_failure1	2.429e-01	1.087e-01	2.235	0.025393	*
immunosuppression1	7.438e-01	6.924e-02	10.743	< 2e-16	***
leukemia1	6.005e-03	1.171e-01	0.051	0.959105	
lymphoma1	8.726e-01	1.542e-01	5.658	1.53e-08	***

solid_tumor_with_metastasis1	1.017e+00	7.727e-02	13.158	< 2e-16	***
mbp_final	-3.869e-02	5.244e-03	-7.378	1.61e-13	***
hr_final1	-9.180e-01	9.312e-02	-9.858	< 2e-16	***
hr_final2	-9.246e-01	7.478e-02	-12.364	< 2e-16	***
hr_final3	-7.591e-01	7.734e-02	-9.815	< 2e-16	***
hr_final4	-4.306e-01	7.781e-02	-5.534	3.14e-08	***
hr_final5	-3.101e-01	7.674e-02	-4.041	5.32e-05	***
hr_final6	-8.544e-02	8.720e-02	-0.980	0.327168	
hr_final7	-1.672e-02	9.570e-02	-0.175	0.861280	
hr_finalunknown	-1.108e+01	5.450e+01	-0.203	0.838904	
rr_final1	-1.495e+00	6.804e-02	-21.978	< 2e-16	***
rr_final2	-1.534e+00	6.963e-02	-22.032	< 2e-16	***
rr_final3	-1.472e+00	7.197e-02	-20.448	< 2e-16	***
rr_final4	-1.118e+00	6.785e-02	-16.474	< 2e-16	***
rr_final5	-9.764e-01	7.263e-02	-13.443	< 2e-16	***
rr_final6	-8.688e-01	7.501e-02	-11.583	< 2e-16	***
rr_final7	-9.070e-01	8.059e-02	-11.255	< 2e-16	***
rr_finalunknown	-1.075e-01	2.703e-01	-0.398	0.690899	
temp_final1	1.851e-01	1.920e-01	0.964	0.335108	
temp_final2	-4.057e-02	1.936e-01	-0.210	0.834023	
temp_final3	-5.163e-01	1.089e-01	-4.741	2.13e-06	***
temp_final4	-7.722e-01	9.216e-02	-8.379	< 2e-16	***
temp_final5	-1.099e+00	8.930e-02	-12.310	< 2e-16	***
temp_finalunknown	-9.907e-01	1.131e-01	-8.756	< 2e-16	***
spo2_final	-1.511e-02	8.865e-04	-17.041	< 2e-16	***
bun_final1	2.844e-01	5.300e-02	5.367	8.00e-08	***
bun_final2	5.930e-01	3.118e-02	19.018	< 2e-16	***
bun_final3	8.779e-01	3.869e-02	22.692	< 2e-16	***
bun_final4	9.532e-01	6.165e-02	15.462	< 2e-16	***
bun_finalunknown	5.369e-01	9.266e-02	5.795	6.85e-09	***
glu_final1	1.797e-01	5.924e-02	3.033	0.002423	**
glu_final2	4.535e-02	5.754e-02	0.788	0.430622	
glu_final3	3.917e-01	8.729e-02	4.488	7.19e-06	***
glu_final4	7.537e-01	1.428e-01	5.278	1.31e-07	***
glu_finalunknown	8.240e-02	9.016e-02	0.914	0.360732	
hco3_final3	-2.047e-01	5.149e-02	-3.976	7.01e-05	***
hco3_final4	3.705e-02	5.593e-02	0.662	0.507677	
hco3_final5	3.154e-01	6.795e-02	4.642	3.45e-06	***
hco3_final6	4.012e-01	7.297e-02	5.499	3.83e-08	***
hco3_finalunknown	-6.499e-02	7.330e-02	-0.887	0.375253	
calcium_cat5	-4.949e-02	1.578e-01	-0.314	0.753801	
calcium_cat6	-2.650e-01	1.154e-01	-2.296	0.021670	*
calcium_cat7	-4.919e-01	1.073e-01	-4.585	4.53e-06	***
calcium_cat8	-4.764e-01	1.057e-01	-4.506	6.59e-06	***
calcium_cat9	-3.971e-01	1.095e-01	-3.626	0.000288	***
calcium_catunknown	-4.571e-01	1.215e-01	-3.761	0.000169	***
hemaglobin_cat2	9.744e-02	4.613e-02	2.112	0.034659	*
hemaglobin_cat3	-1.198e-01	4.705e-02	-2.547	0.010879	*
hemaglobin_cat4	-1.561e-01	5.290e-02	-2.951	0.003164	**
hemaglobin_cat5	2.319e-02	7.974e-02	0.291	0.771217	
hemaglobin_catunknown	-2.237e-02	9.373e-02	-0.239	0.811339	
wbc_final1	-9.213e-02	1.560e-01	-0.591	0.554779	
wbc_final2	-3.299e-01	1.379e-01	-2.392	0.016750	*
wbc_final3	9.228e-02	1.439e-01	0.641	0.521211	
wbc_final4	1.954e-01	1.441e-01	1.356	0.174973	
wbc_finalunknown	-2.256e-01	1.632e-01	-1.382	0.166863	
platelets_cat1	-1.076e+00	7.261e-02	-14.820	< 2e-16	***
platelets_cat2	-1.160e+00	7.351e-02	-15.787	< 2e-16	***



```
platelets_cat3          -1.015e+00  7.623e-02 -13.317 < 2e-16 ***
platelets_cat4          -7.970e-01  8.842e-02 -9.014 < 2e-16 ***
platelets_cat5          -9.060e-01  1.227e-01 -7.383 1.54e-13 ***
platelets_cat6          -1.077e+00  1.497e-01 -7.191 6.43e-13 ***
platelets_catunknown    -1.013e+00  1.230e-01 -8.236 < 2e-16 ***
comor_sum               -3.196e-01  3.014e-02 -10.603 < 2e-16 ***
```

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 77765 on 56095 degrees of freedom

Residual deviance: 47619 on 55988 degrees of freedom

AIC: 47835

Number of Fisher Scoring iterations: 10

[Hide](#)

```
vif(model4)
```

	GVIF	Df	GVIF^(1/(2*Df))
a3j_diag	11.399937	4	1.355541
arf_apache	1.087972	1	1.043059
gcs_eyes	5.257205	4	1.230536
gcs_sum	4.350700	1	2.085833
intubated_apache	1.583414	1	1.258338
ventilated_apache	1.800916	1	1.341982
map_risk	1.305862	1	1.142743
age	1.273428	1	1.128463
elective_surgery	4.660995	1	2.158934
ethnicity	1.216070	5	1.019755
gender	1.073023	1	1.035868
icu_admit_source	9.643538	5	1.254364
icu_stay_type	1.101123	2	1.024375
icu_type	1.680664	7	1.037781
pre_icu_los_days	1.235640	1	1.111593
apache_4a_death_prob	2.149834	1	1.466231
aids	1.018023	1	1.008971
cirrhosis	1.589279	1	1.260666
hepatic_failure	1.576500	1	1.255588
immunosuppression	1.444520	1	1.201882
leukemia	1.155779	1	1.075071
lymphoma	1.074917	1	1.036782
solid_tumor_with_metastasis	1.330371	1	1.153417
mbp_final	1.311622	1	1.145261
hr_final	1.302053	8	1.016633
rr_final	1.486655	8	1.025093
temp_final	1.390316	6	1.027841
spo2_final	1.056102	1	1.027668
bun_final	8.425162	5	1.237536
glu_final	2.806245	5	1.108696
hco3_final	4.716364	5	1.167779
calcium_cat	5.533507	6	1.153232
hemaglobin_cat	8.638903	5	1.240640
wbc_final	9.092343	5	1.247003
platelets_cat	14.731702	7	1.211844
comor_sum	2.382514	1	1.543539

## Auto Stepwise Feature Selection

[Hide](#)

```
model4A = stepAIC(model4, trace = F)
summary(model4A)
```

Call:

```
glm(formula = hospital_death ~ a3j_diag + arf_apache + gcs_eyes +
    gcs_sum + intubated_apache + ventilated_apache + map_risk +
    age + elective_surgery + ethnicity + icu_admit_source + icu_stay_type +
    icu_type + pre_icu_los_days + apache_4a_death_prob + cirrhosis +
    hepatic_failure + immunosuppression + lymphoma + solid_tumor_with_metastasis +
    mbp_final + hr_final + rr_final + temp_final + spo2_final +
    bun_final + glu_final + hco3_final + calcium_cat + hemaglobin_cat +
    wbc_final + platelets_cat + comor_sum, family = binomial,
    data = train)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-3.8393	-0.6463	0.0314	0.6452	2.8644

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	4.948e+00	3.113e-01	15.893	< 2e-16 ***
a3j_diag1	-1.167e+00	1.837e-01	-6.352	2.13e-10 ***
a3j_diag2	-1.180e+00	1.836e-01	-6.425	1.32e-10 ***
a3j_diag3	-1.796e+00	1.853e-01	-9.693	< 2e-16 ***
a3j_diag4	-1.210e+00	1.993e-01	-6.070	1.28e-09 ***
arf_apache1	3.748e-01	6.300e-02	5.949	2.71e-09 ***
gcs_eyes1	2.695e-01	9.732e-02	2.770	0.005610 **
gcs_eyes2	-2.780e-01	1.064e-01	-2.613	0.008972 **
gcs_eyes3	-4.809e-01	1.093e-01	-4.400	1.08e-05 ***
gcs_eyes4	-5.501e-01	1.153e-01	-4.769	1.85e-06 ***
gcs_sum	-3.805e-02	5.182e-03	-7.343	2.09e-13 ***
intubated_apache1	-6.891e-02	3.560e-02	-1.936	0.052909 .
ventilated_apache1	7.436e-01	3.070e-02	24.221	< 2e-16 ***
map_risk	2.297e-02	1.993e-03	11.525	< 2e-16 ***
age	2.642e-02	7.912e-04	33.397	< 2e-16 ***
elective_surgery1	-4.488e-01	7.509e-02	-5.977	2.28e-09 ***
ethnicityAsian	6.221e-02	1.136e-01	0.547	0.584043
ethnicityCaucasian	7.417e-02	4.009e-02	1.850	0.064335 .
ethnicityHispanic	2.488e-01	6.712e-02	3.707	0.000210 ***
ethnicityNative American	-2.172e-02	1.346e-01	-0.161	0.871771
ethnicityOther/Unknown	4.645e-02	6.215e-02	0.747	0.454789
icu_admit_sourceFloor	2.821e-01	3.155e-02	8.943	< 2e-16 ***
icu_admit_sourceOperating Room / Recovery	-5.696e-01	7.656e-02	-7.440	1.00e-13 ***
icu_admit_sourceOther Hospital	7.438e-01	6.432e-02	11.564	< 2e-16 ***
icu_admit_sourceOther ICU	1.444e+00	1.361e-01	10.614	< 2e-16 ***
icu_admit_sourceUnknown	8.034e-03	4.651e-01	0.017	0.986219
icu_stay_typereadmit	6.230e-02	1.372e-01	0.454	0.649867
icu_stay_typetransfer	-4.066e-01	5.138e-02	-7.914	2.49e-15 ***
icu_typeCCU-CTICU	-1.320e-01	6.329e-02	-2.085	0.037061 *
icu_typeCSICU	-2.858e-01	7.914e-02	-3.612	0.000304 ***
icu_typeCTICU	-2.998e-01	7.744e-02	-3.872	0.000108 ***
icu_typeMed-Surg ICU	-1.232e-01	5.085e-02	-2.422	0.015445 *
icu_typeMICU	1.844e-02	6.049e-02	0.305	0.760501
icu_typeNeuro ICU	4.553e-01	6.400e-02	7.113	1.13e-12 ***
icu_typeSICU	7.413e-02	6.887e-02	1.076	0.281765
pre_icu_los_days	3.490e-02	4.193e-03	8.325	< 2e-16 ***
apache_4a_death_prob	1.235e+00	6.356e-02	19.424	< 2e-16 ***
cirrhosis1	7.968e-01	1.001e-01	7.962	1.70e-15 ***
hepatic_failure1	2.460e-01	1.083e-01	2.272	0.023075 *

immunosuppression1	7.467e-01	6.919e-02	10.791	< 2e-16	***
lymphoma1	8.626e-01	1.541e-01	5.598	2.17e-08	***
solid_tumor_with_metastasis1	1.018e+00	7.631e-02	13.341	< 2e-16	***
mbp_final	-3.860e-02	5.242e-03	-7.364	1.78e-13	***
hr_final1	-9.168e-01	9.312e-02	-9.845	< 2e-16	***
hr_final2	-9.241e-01	7.480e-02	-12.355	< 2e-16	***
hr_final3	-7.582e-01	7.736e-02	-9.801	< 2e-16	***
hr_final4	-4.301e-01	7.782e-02	-5.527	3.25e-08	***
hr_final5	-3.096e-01	7.675e-02	-4.034	5.49e-05	***
hr_final6	-8.556e-02	8.719e-02	-0.981	0.326470	
hr_final7	-1.549e-02	9.571e-02	-0.162	0.871467	
hr_finalunknown	-1.108e+01	5.449e+01	-0.203	0.838882	
rr_final1	-1.496e+00	6.804e-02	-21.988	< 2e-16	***
rr_final2	-1.535e+00	6.963e-02	-22.044	< 2e-16	***
rr_final3	-1.472e+00	7.197e-02	-20.456	< 2e-16	***
rr_final4	-1.119e+00	6.785e-02	-16.486	< 2e-16	***
rr_final5	-9.768e-01	7.262e-02	-13.450	< 2e-16	***
rr_final6	-8.694e-01	7.500e-02	-11.593	< 2e-16	***
rr_final7	-9.096e-01	8.056e-02	-11.291	< 2e-16	***
rr_finalunknown	-1.107e-01	2.703e-01	-0.410	0.682145	
temp_final1	1.852e-01	1.920e-01	0.965	0.334717	
temp_final2	-4.023e-02	1.936e-01	-0.208	0.835376	
temp_final3	-5.173e-01	1.089e-01	-4.751	2.03e-06	***
temp_final4	-7.738e-01	9.215e-02	-8.398	< 2e-16	***
temp_final5	-1.101e+00	8.929e-02	-12.324	< 2e-16	***
temp_finalunknown	-9.905e-01	1.131e-01	-8.756	< 2e-16	***
spo2_final	-1.511e-02	8.864e-04	-17.052	< 2e-16	***
bun_final1	2.852e-01	5.297e-02	5.383	7.31e-08	***
bun_final2	5.943e-01	3.106e-02	19.136	< 2e-16	***
bun_final3	8.792e-01	3.839e-02	22.901	< 2e-16	***
bun_final4	9.559e-01	6.130e-02	15.594	< 2e-16	***
bun_finalunknown	5.370e-01	9.265e-02	5.797	6.77e-09	***
glu_final1	1.786e-01	5.922e-02	3.016	0.002564	**
glu_final2	4.418e-02	5.741e-02	0.770	0.441540	
glu_final3	3.913e-01	8.725e-02	4.485	7.30e-06	***
glu_final4	7.551e-01	1.427e-01	5.291	1.22e-07	***
glu_finalunknown	8.213e-02	9.003e-02	0.912	0.361652	
hco3_final3	-2.040e-01	5.146e-02	-3.964	7.37e-05	***
hco3_final4	3.746e-02	5.592e-02	0.670	0.502910	
hco3_final5	3.150e-01	6.794e-02	4.636	3.55e-06	***
hco3_final6	3.989e-01	7.296e-02	5.468	4.55e-08	***
hco3_finalunknown	-6.496e-02	7.329e-02	-0.886	0.375416	
calcium_cat5	-4.943e-02	1.578e-01	-0.313	0.754055	
calcium_cat6	-2.664e-01	1.154e-01	-2.308	0.021006	*
calcium_cat7	-4.919e-01	1.073e-01	-4.585	4.53e-06	***
calcium_cat8	-4.763e-01	1.057e-01	-4.505	6.62e-06	***
calcium_cat9	-3.972e-01	1.095e-01	-3.627	0.000286	***
calcium_catunknown	-4.568e-01	1.215e-01	-3.759	0.000171	***
hemaglobin_cat2	9.794e-02	4.612e-02	2.124	0.033690	*
hemaglobin_cat3	-1.192e-01	4.704e-02	-2.535	0.011249	*
hemaglobin_cat4	-1.538e-01	5.276e-02	-2.916	0.003544	**
hemaglobin_cat5	2.721e-02	7.940e-02	0.343	0.731835	
hemaglobin_catunknown	-1.920e-02	9.367e-02	-0.205	0.837586	
wbc_final1	-9.423e-02	1.559e-01	-0.604	0.545660	
wbc_final2	-3.308e-01	1.378e-01	-2.400	0.016380	*
wbc_final3	9.197e-02	1.438e-01	0.640	0.522477	
wbc_final4	1.952e-01	1.440e-01	1.355	0.175319	
wbc_finalunknown	-2.255e-01	1.631e-01	-1.383	0.166756	

platelets_cat1	-1.075e+00	7.199e-02	-14.939	< 2e-16	***
platelets_cat2	-1.161e+00	7.286e-02	-15.929	< 2e-16	***
platelets_cat3	-1.017e+00	7.555e-02	-13.456	< 2e-16	***
platelets_cat4	-7.977e-01	8.770e-02	-9.096	< 2e-16	***
platelets_cat5	-9.067e-01	1.223e-01	-7.416	1.21e-13	***
platelets_cat6	-1.076e+00	1.495e-01	-7.201	5.99e-13	***
platelets_catunknown	-1.016e+00	1.226e-01	-8.287	< 2e-16	***
comor_sum	-3.211e-01	2.910e-02	-11.035	< 2e-16	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 77765 on 56095 degrees of freedom

Residual deviance: 47621 on 55991 degrees of freedom

AIC: 47831

Number of Fisher Scoring iterations: 10

[Hide](#)

vif(model4A)

	GVIF	Df	GVIF^(1/(2*Df))
a3j_diag	11.380471	4	1.355252
arf_apache	1.087484	1	1.042825
gcs_eyes	5.249631	4	1.230314
gcs_sum	4.350457	1	2.085775
intubated_apache	1.583511	1	1.258376
ventilated_apache	1.799962	1	1.341626
map_risk	1.303184	1	1.141571
age	1.267618	1	1.125886
elective_surgery	4.660061	1	2.158717
ethnicity	1.211877	5	1.019403
icu_admit_source	9.626762	5	1.254146
icu_stay_type	1.100809	2	1.024302
icu_type	1.675526	7	1.037554
pre_icu_los_days	1.235474	1	1.111519
apache_4a_death_prob	2.145073	1	1.464607
cirrhosis	1.577923	1	1.256154
hepatic_failure	1.565218	1	1.251087
immunosuppression	1.443305	1	1.201376
lymphoma	1.071999	1	1.035374
solid_tumor_with_metastasis	1.298172	1	1.139374
mbp_final	1.310951	1	1.144968
hr_final	1.298912	8	1.016480
rr_final	1.482607	8	1.024918
temp_final	1.388059	6	1.027702
spo2_final	1.055913	1	1.027576
bun_final	8.244349	5	1.234854
glu_final	2.773961	5	1.107414
hco3_final	4.689412	5	1.167110
calcium_cat	5.515591	6	1.152920
hemaglobin_cat	8.474680	5	1.238261
wbc_final	9.000560	5	1.245739
platelets_cat	14.288469	7	1.209202
comor_sum	2.222166	1	1.490693

## Relative Importance of Features

[Hide](#)

```
imp = as.data.frame(varImp(model4A))
imp = data.frame(overall = imp$Overall, names = rownames(imp))

imp[order(imp$overall,decreasing = T),]
```

### overall names

<dbl> <fctr>

14	33.39668589	age
12	24.22056037	ventilated_apache1
68	22.90096366	bun_final3
52	22.04398275	rr_final2
51	21.98754913	rr_final1
53	20.45589608	rr_final3

overall names		
	<dbl>	<fctr>
36	19.42426096	apache_4a_death_prob
67	19.13646286	bun_final2
65	17.05179197	spo2_final
54	16.48622861	rr_final4

1-10 of 104 rows

Previous123456...11Next

Results on Train set

Hide

```
#trainPredict = predict(glmFit, newdata = data.balance)#, type = 'response')
trainPredict = predict(model4A, newdata = train, type = 'response')

p_class = ifelse(trainPredict > 0.5, 1, 0)

matrix_table = table(train$hospital_death, p_class)
matrix_table
```

p_class		
	0	1
No	22935	5074
Yes	5889	22198

Hide

```
# Accuracy
accuracy = sum(diag(matrix_table))/sum(matrix_table)
round(accuracy, 3)

[1] 0.805
```

Performance on Train set

Hide

```
pred = prediction(trainPredict, train$hospital_death)

auc.log = performance(pred,"auc"); au_log = as.numeric(auc.log@y.values)
au_log

[1] 0.8872267
```

Hide

```
acc.perf = performance(pred, measure = "acc")

ind = which.max( slot(acc.perf, "y.values")[[1]] )
acc = slot(acc.perf, "y.values")[[1]][ind]
cutoff = slot(acc.perf, "x.values")[[1]][ind]
print(c(accuracy= acc, cutoff = cutoff))
```

```
accuracy cutoff.43645
0.8054407    0.4918746
```

## Results on Test set

Hide

```
testPredict = predict(model4A, newdata = test, type = 'response')

p_class = ifelse(testPredict > 0.5 , 1, 0)

matrix_table = table(test$hospital_death, p_class)
matrix_table
```

```
p_class
      0      1
No 18050 3957
Yes  445 1589
```

Hide

```
# Accuracy
accuracy = sum(diag(matrix_table))/sum(matrix_table)
round(accuracy, 3)
```

```
[1] 0.817
```

## Performance on Test set

Hide

```
pred = prediction(testPredict, test$hospital_death)

auc.log = performance(pred,"auc"); au_log = as.numeric(auc.log@y.values)
au_log
```

```
[1] 0.8817173
```

Hide

```
acc.perf = performance(pred, measure = "acc")

ind = which.max( slot(acc.perf, "y.values")[[1]] )
acc = slot(acc.perf, "y.values")[[1]][ind]
cutoff = slot(acc.perf, "x.values")[[1]][ind]
print(c(accuracy= acc, cutoff = cutoff))
```



```
accuracy cutoff.7933
0.9282892    0.9298824
```

## Train-Both dataset

Hide

```
train = read.csv("train_both.csv")
```

### Data Preparation

Hide

```
train$X = NULL

train$comor_sum = train$aids + train$cirrhosis + train$diabetes_mellitus + train$hepatic_failure +
  train$immunosuppression + train$leukemia + train$lymphoma + train$solid_tumor_with_metastasis

factor.list = c('a2_diag', 'a3j_diag', 'apache_post_operative', 'arf_apache', 'intubated_apache',
               'ventilated_apache', 'elective_surgery', 'aids', 'cirrhosis', 'diabetes_mellitus',
               'hepatic_failure', 'immunosuppression', 'leukemia', 'lymphoma', 'solid_tumor_with_metastasis',
               'gcs_eyes', 'gcs_motor', 'gcs_verbal')
train[, factor.list] = lapply(train[, factor.list], as.factor)
train$hospital_death = factor(train$hospital_death, levels = c(0,1), labels = c("No", "Yes"))
```

### Check the target variable

Hide

```
contrasts(train$hospital_death)
```

```
      Yes
No      0
Yes     1
```

Hide

```
table(train$hospital_death)
```

```
   No   Yes
28156 27940
```

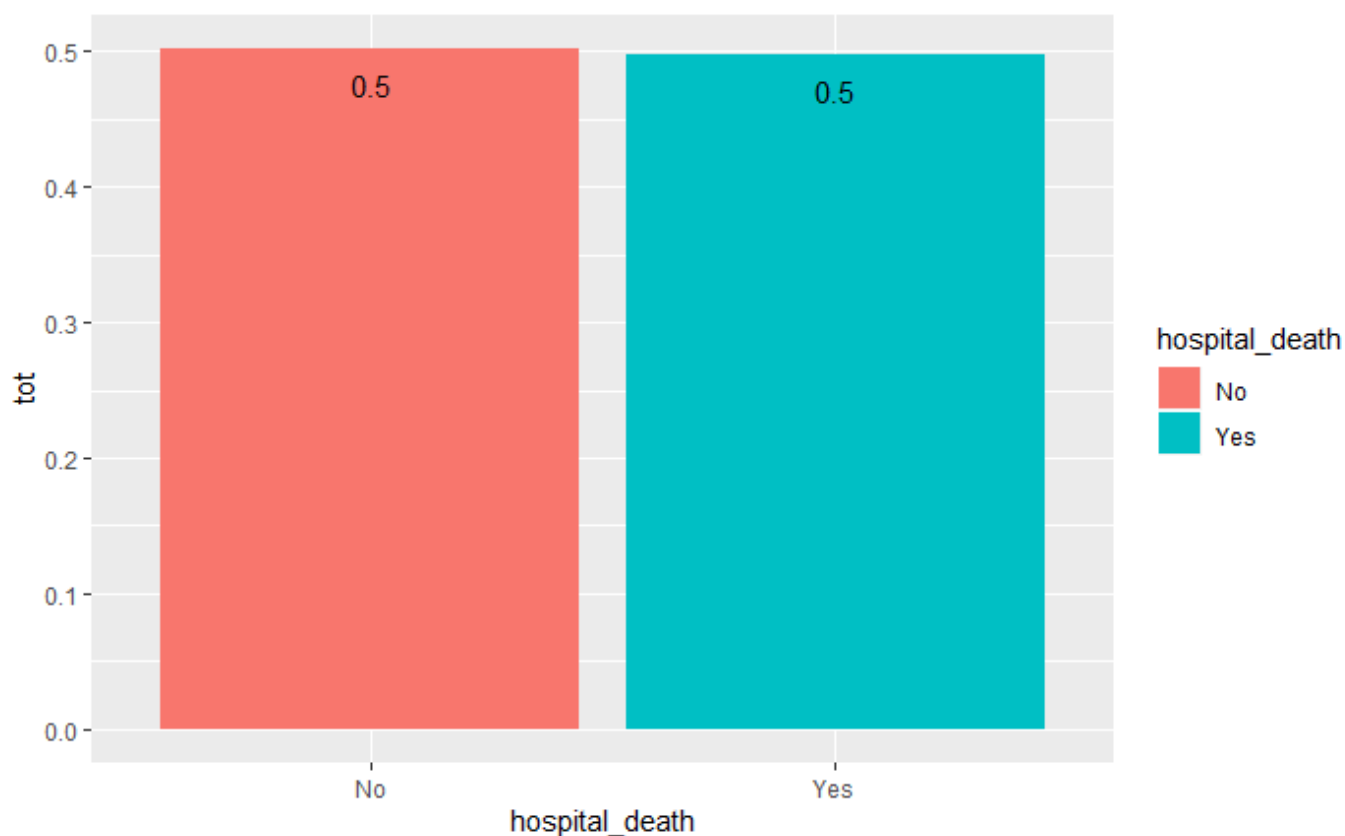
Hide

```
prop.table(table(train$hospital_death))
```

No	Yes
0.5019253	0.4980747

[Hide](#)

```
train %>%
  group_by(hospital_death)%>%
  summarise(tot = n()/nrow(train))%>%
  ggplot(aes(x=hospital_death, y=tot, fill = hospital_death)) +
  geom_bar(stat='identity') +
  geom_text(aes(label = round(tot, 2)), vjust = 2)
```



## Initial Model

[Hide](#)

```
model5 = glm(hospital_death ~ . -gcs_motor -gcs_verbal -gcs_eyes
              -apache_2_bodysystem -apache_3j_bodysystem -a2_diag
              -diabetes_mellitus -hospital_admit_source
              -apache_post_operative -hto_final -cre_final -potassium_cat -sodium_final,
              data = train, family = binomial)
summary(model5)
```

Call:

```
glm(formula = hospital_death ~ . - gcs_motor - gcs_verbal - gcs_eyes -
     apache_2_bodysystem - apache_3j_bodysystem - a2_diag - diabetes_mellitus -
     hospital_admit_source - apache_post_operative - hto_final -
     cre_final - potassium_cat - sodium_final, family = binomial,
     data = train)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-3.6880	-0.6426	-0.1519	0.6385	2.6587

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	5.439e+00	3.138e-01	17.333	< 2e-16 ***
a3j_diag1	-1.707e+00	2.026e-01	-8.427	< 2e-16 ***
a3j_diag2	-1.707e+00	2.023e-01	-8.439	< 2e-16 ***
a3j_diag3	-2.245e+00	2.035e-01	-11.029	< 2e-16 ***
a3j_diag4	-1.818e+00	2.162e-01	-8.405	< 2e-16 ***
arf_apache1	5.183e-01	6.214e-02	8.341	< 2e-16 ***
gcs_sum	-9.022e-02	3.700e-03	-24.384	< 2e-16 ***
intubated_apache1	-4.276e-02	3.566e-02	-1.199	0.230496
ventilated_apache1	7.224e-01	3.108e-02	23.245	< 2e-16 ***
map_risk	1.939e-02	2.219e-03	8.741	< 2e-16 ***
age	2.727e-02	8.725e-04	31.257	< 2e-16 ***
elective_surgery1	-5.596e-01	7.443e-02	-7.519	5.54e-14 ***
ethnicityAsian	1.672e-01	1.119e-01	1.495	0.134915
ethnicityCaucasian	7.202e-02	4.016e-02	1.793	0.072932 .
ethnicityHispanic	2.042e-01	6.789e-02	3.008	0.002629 **
ethnicityNative American	-1.583e-02	1.359e-01	-0.117	0.907223
ethnicityOther/Unknown	8.893e-02	6.214e-02	1.431	0.152399
genderM	6.266e-02	2.366e-02	2.648	0.008101 **
icu_admit_sourceFloor	2.179e-01	3.223e-02	6.759	1.39e-11 ***
icu_admit_sourceOperating Room / Recovery	-3.862e-01	7.538e-02	-5.123	3.01e-07 ***
icu_admit_sourceOther Hospital	5.148e-01	6.315e-02	8.152	3.59e-16 ***
icu_admit_sourceOther ICU	1.946e+00	1.436e-01	13.548	< 2e-16 ***
icu_admit_sourceUnknown	1.043e+00	4.216e-01	2.474	0.013350 *
icu_stay_typereadmit	-1.761e-01	1.346e-01	-1.308	0.190911
icu_stay_typetransfer	-3.205e-01	5.106e-02	-6.278	3.44e-10 ***
icu_typeCCU-CTICU	-1.035e-02	6.381e-02	-0.162	0.871167
icu_typeCSICU	-3.084e-01	8.037e-02	-3.837	0.000125 ***
icu_typeCTICU	-1.386e-01	7.827e-02	-1.771	0.076631 .
icu_typeMed-Surg ICU	-1.073e-01	5.176e-02	-2.072	0.038235 *
icu_typeMICU	-6.281e-03	6.132e-02	-0.102	0.918409
icu_typeNeuro ICU	4.553e-01	6.502e-02	7.003	2.51e-12 ***
icu_typeSICU	1.060e-01	7.034e-02	1.507	0.131703
pre_icu_los_days	3.548e-02	4.917e-03	7.216	5.37e-13 ***
apache_4a_death_prob	1.781e+00	7.920e-02	22.483	< 2e-16 ***
aids1	6.265e-02	4.485e-01	0.140	0.888921
cirrhosis1	5.766e-01	9.942e-02	5.799	6.66e-09 ***
hepatic_failure1	4.957e-01	1.049e-01	4.726	2.28e-06 ***
immunosuppression1	5.146e-01	7.047e-02	7.301	2.85e-13 ***
leukemia1	2.686e-01	1.239e-01	2.168	0.030158 *
lymphoma1	7.769e-01	1.538e-01	5.052	4.38e-07 ***
solid_tumor_with_metastasis1	1.078e+00	7.941e-02	13.577	< 2e-16 ***
mbp_final	-5.667e-02	5.838e-03	-9.708	< 2e-16 ***
hr_final1	-8.256e-01	9.096e-02	-9.077	< 2e-16 ***

hr_final2	-7.944e-01	7.277e-02	-10.917	< 2e-16	***
hr_final3	-4.980e-01	7.550e-02	-6.595	4.25e-11	***
hr_final4	-3.480e-01	7.592e-02	-4.584	4.56e-06	***
hr_final5	-1.159e-01	7.494e-02	-1.546	0.122098	
hr_final6	-1.991e-02	8.523e-02	-0.234	0.815244	
hr_final7	1.519e-01	9.498e-02	1.600	0.109683	
hr_finalunknown	-1.122e+01	5.451e+01	-0.206	0.836932	
rr_final1	-1.417e+00	6.853e-02	-20.682	< 2e-16	***
rr_final2	-1.412e+00	7.001e-02	-20.169	< 2e-16	***
rr_final3	-1.388e+00	7.245e-02	-19.153	< 2e-16	***
rr_final4	-1.036e+00	6.855e-02	-15.116	< 2e-16	***
rr_final5	-9.136e-01	7.327e-02	-12.468	< 2e-16	***
rr_final6	-8.438e-01	7.561e-02	-11.160	< 2e-16	***
rr_final7	-9.521e-01	8.027e-02	-11.861	< 2e-16	***
rr_finalunknown	5.208e-01	2.674e-01	1.947	0.051486	.
temp_final1	1.942e-01	1.953e-01	0.994	0.320072	
temp_final2	-9.196e-02	1.890e-01	-0.486	0.626625	
temp_final3	-4.723e-01	1.085e-01	-4.354	1.33e-05	***
temp_final4	-6.513e-01	9.175e-02	-7.098	1.26e-12	***
temp_final5	-1.013e+00	8.876e-02	-11.410	< 2e-16	***
temp_finalunknown	-8.118e-01	1.124e-01	-7.222	5.12e-13	***
spo2_final	-1.638e-02	1.006e-03	-16.280	< 2e-16	***
bun_final1	2.056e-01	5.271e-02	3.901	9.59e-05	***
bun_final2	5.363e-01	3.128e-02	17.146	< 2e-16	***
bun_final3	7.229e-01	3.869e-02	18.687	< 2e-16	***
bun_final4	9.283e-01	6.314e-02	14.703	< 2e-16	***
bun_finalunknown	2.507e-01	9.281e-02	2.701	0.006914	**
glu_final1	2.788e-01	5.918e-02	4.711	2.47e-06	***
glu_final2	1.877e-01	5.754e-02	3.262	0.001107	**
glu_final3	5.084e-01	8.781e-02	5.790	7.02e-09	***
glu_final4	5.645e-01	1.358e-01	4.156	3.24e-05	***
glu_finalunknown	3.539e-01	9.112e-02	3.884	0.000103	***
hco3_final3	-2.539e-01	5.160e-02	-4.920	8.64e-07	***
hco3_final4	2.403e-02	5.585e-02	0.430	0.667066	
hco3_final5	2.978e-01	6.749e-02	4.412	1.02e-05	***
hco3_final6	4.690e-01	7.294e-02	6.431	1.27e-10	***
hco3_finalunknown	-1.636e-01	7.355e-02	-2.224	0.026142	*
calcium_cat5	-4.391e-01	1.549e-01	-2.835	0.004581	**
calcium_cat6	-3.401e-01	1.144e-01	-2.974	0.002936	**
calcium_cat7	-4.906e-01	1.060e-01	-4.628	3.69e-06	***
calcium_cat8	-5.348e-01	1.046e-01	-5.114	3.15e-07	***
calcium_cat9	-4.760e-01	1.086e-01	-4.384	1.17e-05	***
calcium_catunknown	-3.453e-01	1.193e-01	-2.894	0.003803	**
hemaglobin_cat2	1.363e-02	4.623e-02	0.295	0.768154	
hemaglobin_cat3	-1.303e-01	4.719e-02	-2.761	0.005757	**
hemaglobin_cat4	-1.018e-01	5.289e-02	-1.924	0.054296	.
hemaglobin_cat5	2.128e-01	7.977e-02	2.668	0.007634	**
hemaglobin_catunknown	7.996e-02	9.455e-02	0.846	0.397722	
wbc_final1	-3.016e-01	1.550e-01	-1.946	0.051672	.
wbc_final2	-4.209e-01	1.363e-01	-3.088	0.002018	**
wbc_final3	-3.240e-02	1.426e-01	-0.227	0.820317	
wbc_final4	-1.158e-01	1.426e-01	-0.812	0.416802	
wbc_finalunknown	-3.536e-01	1.624e-01	-2.178	0.029415	*
platelets_cat1	-1.012e+00	7.240e-02	-13.973	< 2e-16	***
platelets_cat2	-1.056e+00	7.329e-02	-14.411	< 2e-16	***
platelets_cat3	-9.020e-01	7.606e-02	-11.860	< 2e-16	***
platelets_cat4	-7.264e-01	8.796e-02	-8.258	< 2e-16	***
platelets_cat5	-1.051e+00	1.233e-01	-8.525	< 2e-16	***

```

platelets_cat6                -7.066e-01  1.469e-01  -4.811 1.51e-06 ***
platelets_catunknown          -9.200e-01  1.233e-01  -7.460 8.65e-14 ***
comor_sum                     -3.042e-01  3.026e-02 -10.052 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 77765 on 56095 degrees of freedom  
 Residual deviance: 47530 on 55992 degrees of freedom  
 AIC: 47738

Number of Fisher Scoring iterations: 10

[Hide](#)

```
vif(model5)
```

	GVIF	Df	GVIF^(1/(2*Df))
a3j_diag	12.207628	4	1.367190
arf_apache	1.086497	1	1.042352
gcs_sum	1.888607	1	1.374266
intubated_apache	1.609445	1	1.268639
ventilated_apache	1.840359	1	1.356598
map_risk	1.419545	1	1.191446
age	1.336645	1	1.156134
elective_surgery	4.469284	1	2.114068
ethnicity	1.219072	5	1.020006
gender	1.066226	1	1.032582
icu_admit_source	11.131880	5	1.272497
icu_stay_type	1.094080	2	1.022733
icu_type	1.680388	7	1.037769
pre_icu_los_days	1.316290	1	1.147297
apache_4a_death_prob	2.769002	1	1.664032
aids	1.021023	1	1.010457
cirrhosis	1.607286	1	1.267788
hepatic_failure	1.601195	1	1.265383
immunosuppression	1.442855	1	1.201189
leukemia	1.141677	1	1.068493
lymphoma	1.091012	1	1.044515
solid_tumor_with_metastasis	1.302858	1	1.141428
mbp_final	1.419396	1	1.191384
hr_final	1.324103	8	1.017701
rr_final	1.549711	8	1.027758
temp_final	1.386802	6	1.027625
spo2_final	1.073526	1	1.036111
bun_final	8.329099	5	1.236118
glu_final	2.828435	5	1.109570
hco3_final	4.565386	5	1.163986
calcium_cat	5.137607	6	1.146120
hemaglobin_cat	8.926977	5	1.244716
wbc_final	9.564048	5	1.253326
platelets_cat	14.813560	7	1.212324
comor_sum	2.411335	1	1.552847

# Auto Stepwise Feature Selection

[Hide](#)

```
model15A = stepAIC(model15, trace = F)
summary(model15A)
```

Call:

```
glm(formula = hospital_death ~ a3j_diag + arf_apache + gcs_sum +
    ventilated_apache + map_risk + age + elective_surgery + ethnicity +
    gender + icu_admit_source + icu_stay_type + icu_type + pre_icu_los_days +
    apache_4a_death_prob + cirrhosis + hepatic_failure + immunosuppression +
    leukemia + lymphoma + solid_tumor_with_metastasis + mbp_final +
    hr_final + rr_final + temp_final + spo2_final + bun_final +
    glu_final + hco3_final + calcium_cat + hemaglobin_cat + wbc_final +
    platelets_cat + comor_sum, family = binomial, data = train)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-3.6829	-0.6427	-0.1516	0.6377	2.6597

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	5.422e+00	3.134e-01	17.300	< 2e-16 ***
a3j_diag1	-1.694e+00	2.022e-01	-8.376	< 2e-16 ***
a3j_diag2	-1.693e+00	2.019e-01	-8.386	< 2e-16 ***
a3j_diag3	-2.231e+00	2.031e-01	-10.982	< 2e-16 ***
a3j_diag4	-1.805e+00	2.159e-01	-8.360	< 2e-16 ***
arf_apache1	5.191e-01	6.213e-02	8.354	< 2e-16 ***
gcs_sum	-8.994e-02	3.692e-03	-24.361	< 2e-16 ***
ventilated_apache1	7.079e-01	2.863e-02	24.722	< 2e-16 ***
map_risk	1.932e-02	2.218e-03	8.710	< 2e-16 ***
age	2.733e-02	8.711e-04	31.369	< 2e-16 ***
elective_surgery1	-5.621e-01	7.440e-02	-7.556	4.17e-14 ***
ethnicityAsian	1.660e-01	1.119e-01	1.484	0.137838
ethnicityCaucasian	7.251e-02	4.016e-02	1.806	0.070953 .
ethnicityHispanic	2.015e-01	6.785e-02	2.970	0.002974 **
ethnicityNative American	-1.879e-02	1.359e-01	-0.138	0.890046
ethnicityOther/Unknown	8.945e-02	6.213e-02	1.440	0.149983
genderM	6.232e-02	2.366e-02	2.634	0.008446 **
icu_admit_sourceFloor	2.183e-01	3.223e-02	6.773	1.26e-11 ***
icu_admit_sourceOperating Room / Recovery	-3.871e-01	7.540e-02	-5.134	2.84e-07 ***
icu_admit_sourceOther Hospital	5.133e-01	6.311e-02	8.133	4.18e-16 ***
icu_admit_sourceOther ICU	1.926e+00	1.426e-01	13.502	< 2e-16 ***
icu_admit_sourceUnknown	1.023e+00	4.218e-01	2.426	0.015273 *
icu_stay_typereadmit	-1.781e-01	1.346e-01	-1.323	0.185843
icu_stay_typetransfer	-3.167e-01	5.093e-02	-6.218	5.02e-10 ***
icu_typeCCU-CTICU	-1.155e-02	6.381e-02	-0.181	0.856398
icu_typeCSICU	-3.104e-01	8.037e-02	-3.863	0.000112 ***
icu_typeCTICU	-1.393e-01	7.828e-02	-1.780	0.075107 .
icu_typeMed-Surg ICU	-1.060e-01	5.175e-02	-2.048	0.040563 *
icu_typeMICU	-5.499e-03	6.132e-02	-0.090	0.928544
icu_typeNeuro ICU	4.565e-01	6.502e-02	7.022	2.19e-12 ***
icu_typeSICU	1.056e-01	7.034e-02	1.501	0.133236
pre_icu_los_days	3.566e-02	4.917e-03	7.253	4.07e-13 ***
apache_4a_death_prob	1.766e+00	7.824e-02	22.576	< 2e-16 ***
cirrhosis1	5.781e-01	9.939e-02	5.816	6.02e-09 ***
hepatic_failure1	4.941e-01	1.048e-01	4.713	2.44e-06 ***
immunosuppression1	5.161e-01	7.042e-02	7.329	2.33e-13 ***
leukemia1	2.653e-01	1.238e-01	2.143	0.032144 *
lymphoma1	7.768e-01	1.537e-01	5.055	4.31e-07 ***
solid_tumor_with_metastasis1	1.077e+00	7.936e-02	13.568	< 2e-16 ***
mbp_final	-5.657e-02	5.837e-03	-9.692	< 2e-16 ***

hr_final1	-8.264e-01	9.092e-02	-9.089	< 2e-16	***
hr_final2	-7.943e-01	7.272e-02	-10.922	< 2e-16	***
hr_final3	-4.974e-01	7.545e-02	-6.592	4.35e-11	***
hr_final4	-3.474e-01	7.586e-02	-4.579	4.66e-06	***
hr_final5	-1.155e-01	7.489e-02	-1.543	0.122905	
hr_final6	-2.013e-02	8.518e-02	-0.236	0.813182	
hr_final7	1.509e-01	9.493e-02	1.589	0.112015	
hr_finalunknown	-1.122e+01	5.460e+01	-0.205	0.837232	
rr_final1	-1.418e+00	6.852e-02	-20.698	< 2e-16	***
rr_final2	-1.413e+00	7.000e-02	-20.186	< 2e-16	***
rr_final3	-1.388e+00	7.244e-02	-19.160	< 2e-16	***
rr_final4	-1.036e+00	6.854e-02	-15.110	< 2e-16	***
rr_final5	-9.128e-01	7.326e-02	-12.460	< 2e-16	***
rr_final6	-8.420e-01	7.559e-02	-11.139	< 2e-16	***
rr_final7	-9.516e-01	8.026e-02	-11.857	< 2e-16	***
rr_finalunknown	5.056e-01	2.669e-01	1.894	0.058184	.
temp_final1	1.949e-01	1.952e-01	0.998	0.318184	
temp_final2	-9.534e-02	1.890e-01	-0.505	0.613905	
temp_final3	-4.717e-01	1.084e-01	-4.350	1.36e-05	***
temp_final4	-6.509e-01	9.172e-02	-7.097	1.28e-12	***
temp_final5	-1.012e+00	8.873e-02	-11.408	< 2e-16	***
temp_finalunknown	-8.108e-01	1.124e-01	-7.215	5.38e-13	***
spo2_final	-1.642e-02	1.006e-03	-16.327	< 2e-16	***
bun_final1	2.052e-01	5.271e-02	3.892	9.93e-05	***
bun_final2	5.360e-01	3.128e-02	17.137	< 2e-16	***
bun_final3	7.240e-01	3.868e-02	18.718	< 2e-16	***
bun_final4	9.306e-01	6.310e-02	14.748	< 2e-16	***
bun_finalunknown	2.485e-01	9.275e-02	2.679	0.007387	**
glu_final1	2.791e-01	5.918e-02	4.715	2.41e-06	***
glu_final2	1.890e-01	5.752e-02	3.287	0.001014	**
glu_final3	5.099e-01	8.779e-02	5.809	6.29e-09	***
glu_final4	5.648e-01	1.359e-01	4.157	3.22e-05	***
glu_finalunknown	3.574e-01	9.102e-02	3.926	8.62e-05	***
hco3_final3	-2.554e-01	5.158e-02	-4.952	7.35e-07	***
hco3_final4	2.265e-02	5.584e-02	0.406	0.684966	
hco3_final5	2.955e-01	6.745e-02	4.381	1.18e-05	***
hco3_final6	4.660e-01	7.289e-02	6.393	1.63e-10	***
hco3_finalunknown	-1.637e-01	7.353e-02	-2.226	0.025997	*
calcium_cat5	-4.410e-01	1.549e-01	-2.847	0.004413	**
calcium_cat6	-3.403e-01	1.143e-01	-2.977	0.002910	**
calcium_cat7	-4.909e-01	1.060e-01	-4.633	3.61e-06	***
calcium_cat8	-5.338e-01	1.045e-01	-5.107	3.27e-07	***
calcium_cat9	-4.746e-01	1.085e-01	-4.372	1.23e-05	***
calcium_catunknown	-3.441e-01	1.193e-01	-2.885	0.003920	**
hemaglobin_cat2	1.342e-02	4.622e-02	0.290	0.771606	
hemaglobin_cat3	-1.311e-01	4.718e-02	-2.778	0.005468	**
hemaglobin_cat4	-1.028e-01	5.288e-02	-1.944	0.051930	.
hemaglobin_cat5	2.115e-01	7.977e-02	2.651	0.008020	**
hemaglobin_catunknown	7.981e-02	9.447e-02	0.845	0.398199	
wbc_final1	-3.015e-01	1.549e-01	-1.946	0.051655	.
wbc_final2	-4.220e-01	1.363e-01	-3.097	0.001955	**
wbc_final3	-3.454e-02	1.426e-01	-0.242	0.808584	
wbc_final4	-1.168e-01	1.425e-01	-0.820	0.412340	
wbc_finalunknown	-3.551e-01	1.623e-01	-2.188	0.028659	*
platelets_cat1	-1.011e+00	7.240e-02	-13.970	< 2e-16	***
platelets_cat2	-1.055e+00	7.329e-02	-14.402	< 2e-16	***
platelets_cat3	-9.016e-01	7.606e-02	-11.853	< 2e-16	***
platelets_cat4	-7.255e-01	8.796e-02	-8.248	< 2e-16	***



```

platelets_cat5                -1.052e+00  1.233e-01  -8.527  < 2e-16 ***
platelets_cat6                -7.037e-01  1.468e-01  -4.794  1.64e-06 ***
platelets_catunknown          -9.177e-01  1.231e-01  -7.454  9.06e-14 ***
comor_sum                     -3.033e-01  3.021e-02 -10.043  < 2e-16 ***
---

```

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 77765 on 56095 degrees of freedom

Residual deviance: 47532 on 55994 degrees of freedom

AIC: 47736

Number of Fisher Scoring iterations: 10

[Hide](#)

vif(model5A)

	GVIF	Df	GVIF^(1/(2*Df))
a3j_diag	12.166321	4	1.366611
arf_apache	1.086164	1	1.042192
gcs_sum	1.880827	1	1.371432
ventilated_apache	1.562516	1	1.250007
map_risk	1.418449	1	1.190987
age	1.332173	1	1.154198
elective_surgery	4.465946	1	2.113278
ethnicity	1.213621	5	1.019549
gender	1.066003	1	1.032474
icu_admit_source	10.944307	5	1.270337
icu_stay_type	1.089427	2	1.021644
icu_type	1.672799	7	1.037434
pre_icu_los_days	1.315173	1	1.146810
apache_4a_death_prob	2.700027	1	1.643176
cirrhosis	1.606420	1	1.267446
hepatic_failure	1.599719	1	1.264800
immunosuppression	1.441311	1	1.200546
leukemia	1.140614	1	1.067995
lymphoma	1.090784	1	1.044406
solid_tumor_with_metastasis	1.301827	1	1.140977
mbp_final	1.419139	1	1.191276
hr_final	1.322329	8	1.017616
rr_final	1.540769	8	1.027386
temp_final	1.385795	6	1.027563
spo2_final	1.072311	1	1.035525
bun_final	8.309224	5	1.235822
glu_final	2.821193	5	1.109285
hco3_final	4.554477	5	1.163708
calcium_cat	5.121098	6	1.145813
hemaglobin_cat	8.909243	5	1.244469
wbc_final	9.545736	5	1.253086
platelets_cat	14.751582	7	1.211961
comor_sum	2.402944	1	1.550143

## Relative Importance of Features

```
imp = as.data.frame(varImp(model15A))
imp = data.frame(overall = imp$Overall, names = rownames(imp))

imp[order(imp$overall,decreasing = T),]
```

overall		names
	<dbl>	<fctr>
9	31.3693715	age
7	24.7224384	ventilated_apache1
6	24.3605662	gcs_sum
32	22.5759753	apache_4a_death_prob
48	20.6975768	rr_final1
49	20.1856278	rr_final2
50	19.1596908	rr_final3
65	18.7180612	bun_final3
64	17.1371838	bun_final2
62	16.3267386	spo2_final
1-10 of 101 rows		
Previous		1 2 3 4 5 6 ... 11 Next

Results on Train set

```
#trainPredict = predict(glmFit, newdata = data.balance)#, type = 'response')
trainPredict = predict(model15A, newdata = train, type = 'response')

p_class = ifelse(trainPredict > 0.5, 1, 0)

matrix_table = table(train$hospital_death, p_class)
matrix_table
```

p_class	
	0 1
No	23183 4973
Yes	5998 21942

```
# Accuracy
accuracy = sum(diag(matrix_table))/sum(matrix_table)
round(accuracy, 3)
```

```
[1] 0.804
```

Performance on Train set

Hide

```
pred = prediction(trainPredict, train$hospital_death)

auc.log = performance(pred,"auc"); au_log = as.numeric(auc.log@y.values)
au_log
```

```
[1] 0.8881072
```

Hide

```
acc.perf = performance(pred, measure = "acc")

ind = which.max( slot(acc.perf, "y.values")[[1]] )
acc = slot(acc.perf, "y.values")[[1]][ind]
cutoff = slot(acc.perf, "x.values")[[1]][ind]
print(c(accuracy= acc, cutoff = cutoff))
```

```
accuracy cutoff.52089
0.8071699    0.4288148
```

## Results on Test set

Hide

```
testPredict = predict(model5A, newdata = test, type = 'response')

p_class = ifelse(testPredict > 0.5 , 1, 0)

matrix_table = table(test$hospital_death, p_class)
matrix_table
```

```
p_class
      0      1
No 18136 3871
Yes  450 1584
```

Hide

```
# Accuracy
accuracy = sum(diag(matrix_table))/sum(matrix_table)
round(accuracy, 3)
```

```
[1] 0.82
```

## Performance on Test set

Hide

```
pred = prediction(testPredict, test$hospital_death)

auc.log = performance(pred,"auc"); au_log = as.numeric(auc.log@y.values)
au_log
```

```
[1] 0.8837864
```

[Hide](#)

```
acc.perf = performance(pred, measure = "acc")

ind = which.max( slot(acc.perf, "y.values")[[1]] )
acc = slot(acc.perf, "y.values")[[1]][ind]
cutoff = slot(acc.perf, "x.values")[[1]][ind]
print(c(accuracy= acc, cutoff = cutoff))
```

```
accuracy cutoff.9882
0.9285387  0.9330996
```