



# ABOUT ME

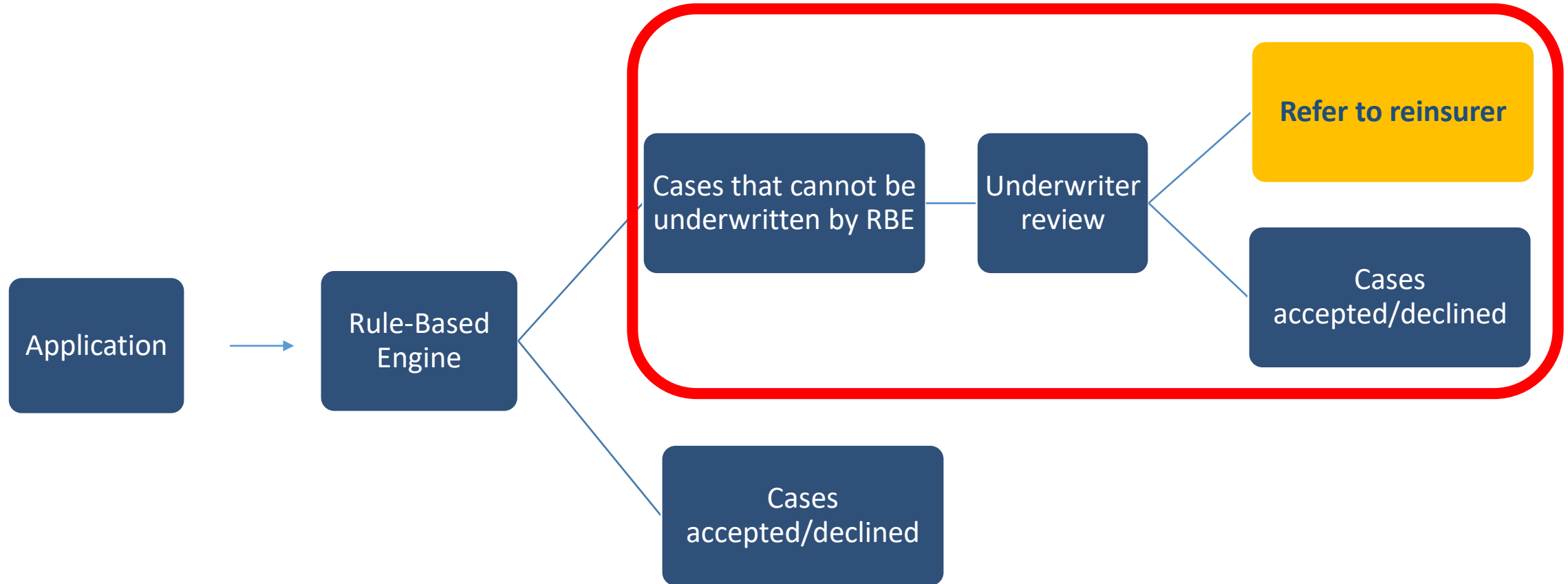
- FIA and FASSA
- MSc Business Analytics Imperial College
- Reserving Actuary at Gen Re
- Gen Re: Berkshire Hathaway

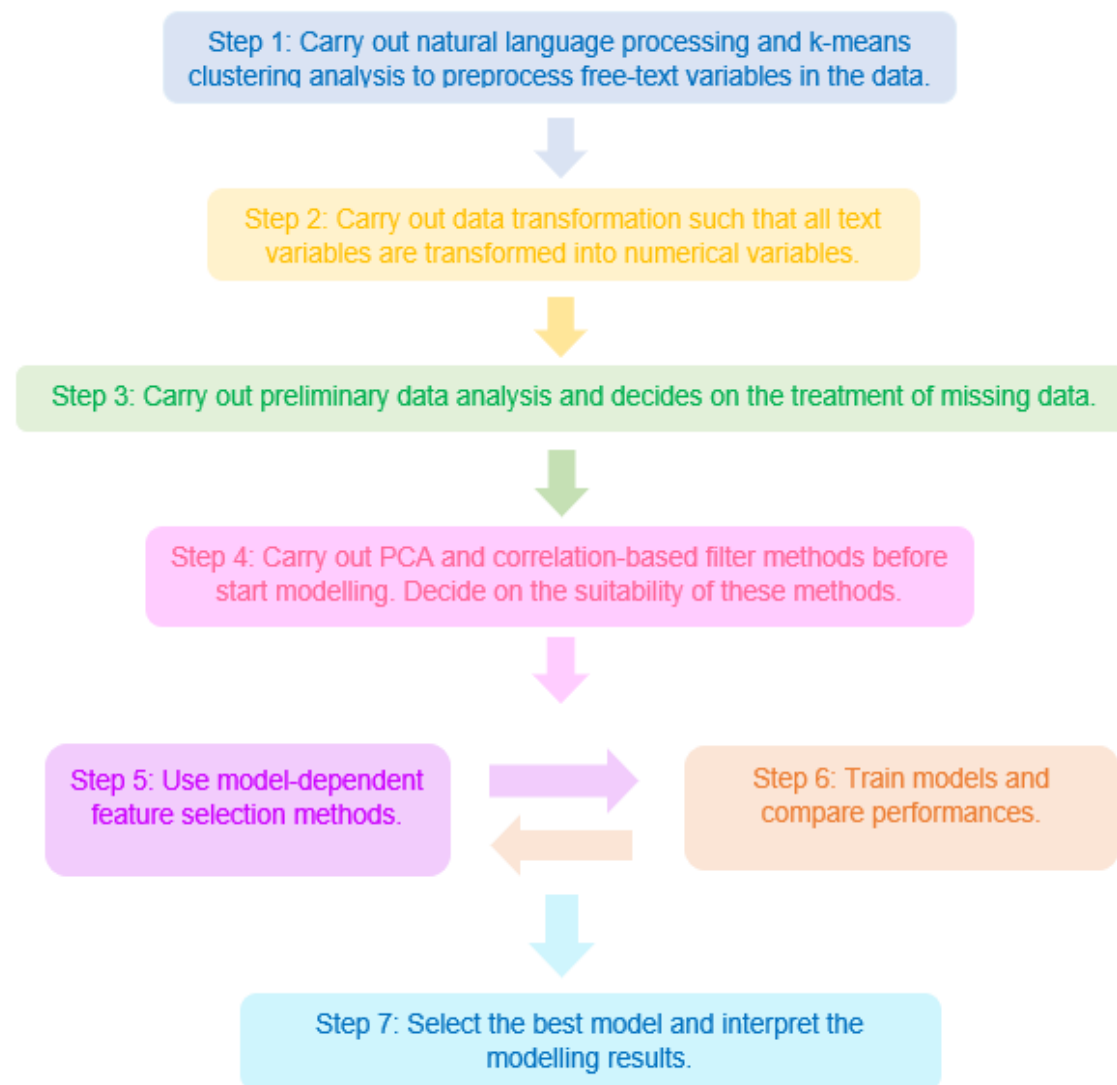




- High prediction accuracies
- Class labels to identify risk groups of interest
- Underwriting insights

# RAW DATA AND UNDERWRITING PROCESS





# DECISION AND CLASS LABELS

Class Label	Description
0	loading is 0, the application is accepted on standard terms
1	loading is $1 * 25\% = 25\%$ , the application is accepted with loading 25%
2	loading is $2 * 25\% = 50\%$ , the application is accepted with loading 50%
3	loading is $3 * 25\% = 75\%$ , the application is accepted with loading 75%
4	loading is $4 * 25\% = 100\%$ , the application is accepted with loading 100%
5	loading is $5 * 25\% = 125\%$ , the application is accepted with loading 125%
6	loading is $6 * 25\% = 150\%$ , the application is accepted with loading 150%
7	loading is $7 * 25\% = 175\%$ , the application is accepted with loading 175%
8	loading is $8 * 25\% = 200\%$ , the application is accepted with loading 200%
9	loading is $9 * 25\% = 225\%$ , the application is accepted with loading 225%
10	loading is $10 * 25\% = 250\%$ , the application is accepted with loading 250%
11	loading is $11 * 25\% = 275\%$ , the application is accepted with loading 275%
12	loading is $12 * 25\% = 300\%$ , the application is accepted with loading 300%
13	loading is $13 * 25\% = 325\%$ , the application is accepted with loading 325%
14	loading is $14 * 25\% = 350\%$ , the application is accepted with loading 350%
15	loading is $15 * 25\% = 375\%$ , the application is accepted with loading 375%
16	loading is $16 * 25\% = 400\%$ , the application is accepted with loading 400%
20	the application is accepted with loading greater than 400%
100	declined applications

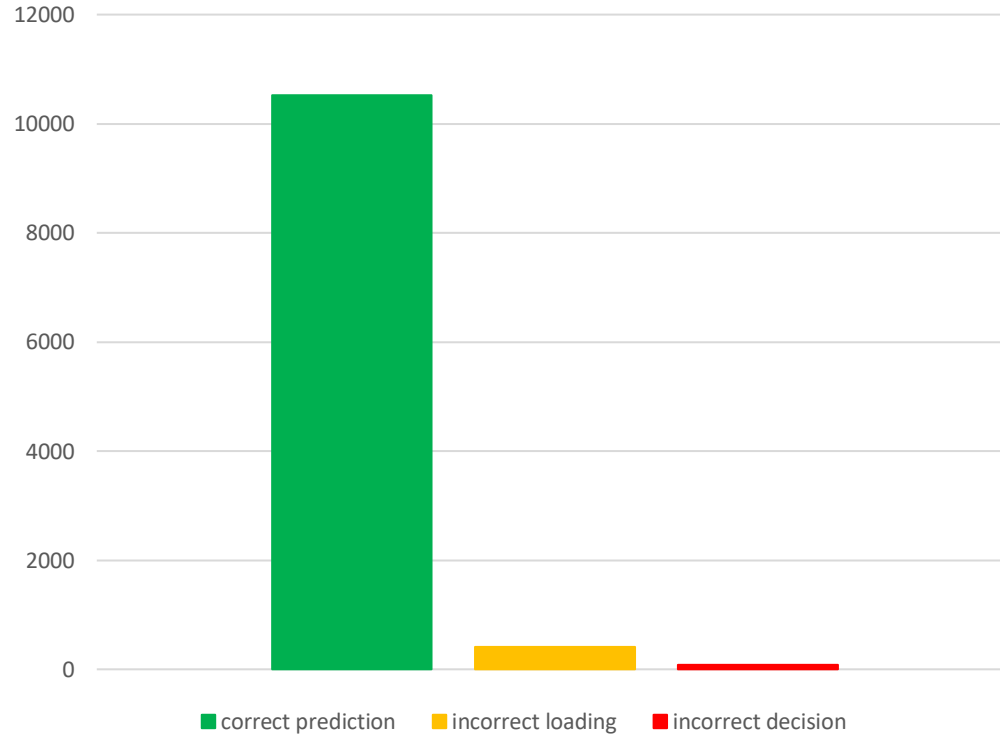
Table 1: Class labels and descriptions

- micro average for precision and recall scores
- XGB best performing model
- Training: testing = 80%:20%

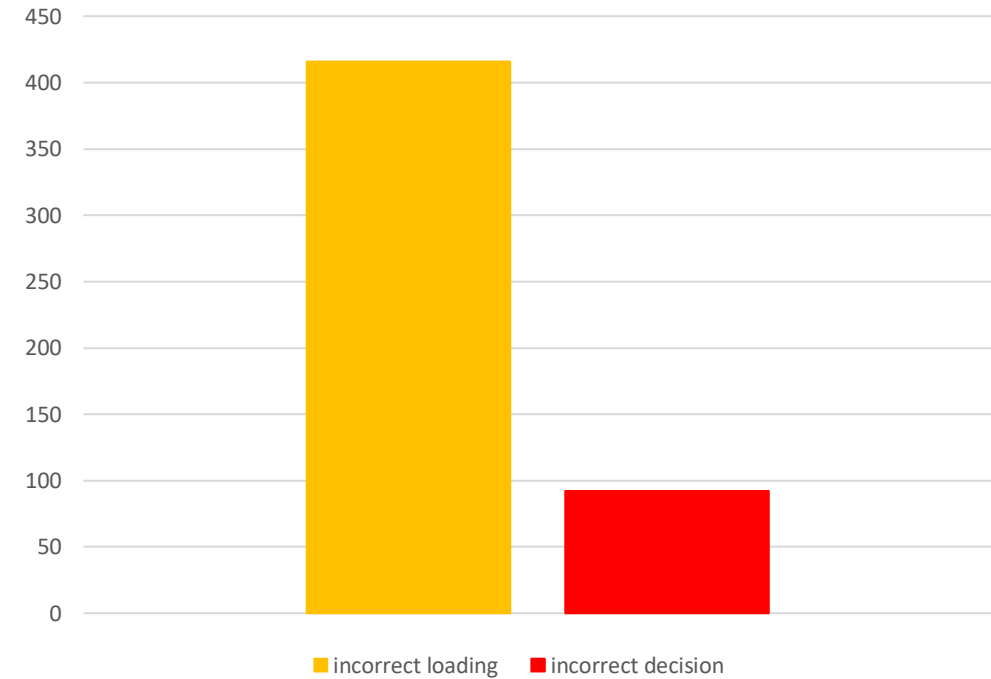
	Model_Name	Precision	Recall	Train_Accuracy	Test_Accuracy	F1_Score
7	XGBClassifier	0.800000	0.800000	1.000000	0.800000	0.800000
1	RandomForestClassifier	0.760000	0.760000	1.000000	0.760000	0.760000
4	BaggingClassifier	0.760000	0.760000	1.000000	0.760000	0.760000
5	GradientBoostingClassifier	0.680000	0.680000	1.000000	0.680000	0.680000
0	DecisionTreeClassifier	0.660000	0.660000	1.000000	0.660000	0.660000
9	KNeighborsClassifier	0.460000	0.460000	0.660000	0.460000	0.460000
3	AdaBoostClassifier	0.420000	0.420000	0.430000	0.420000	0.420000
2	LogisticRegression	0.380000	0.380000	0.390000	0.380000	0.380000
6	SVC	0.380000	0.380000	0.390000	0.380000	0.380000
8	SGDClassifier	0.290000	0.290000	0.290000	0.290000	0.290000



Correct vs Incorrect Loading vs Incorrect Decision

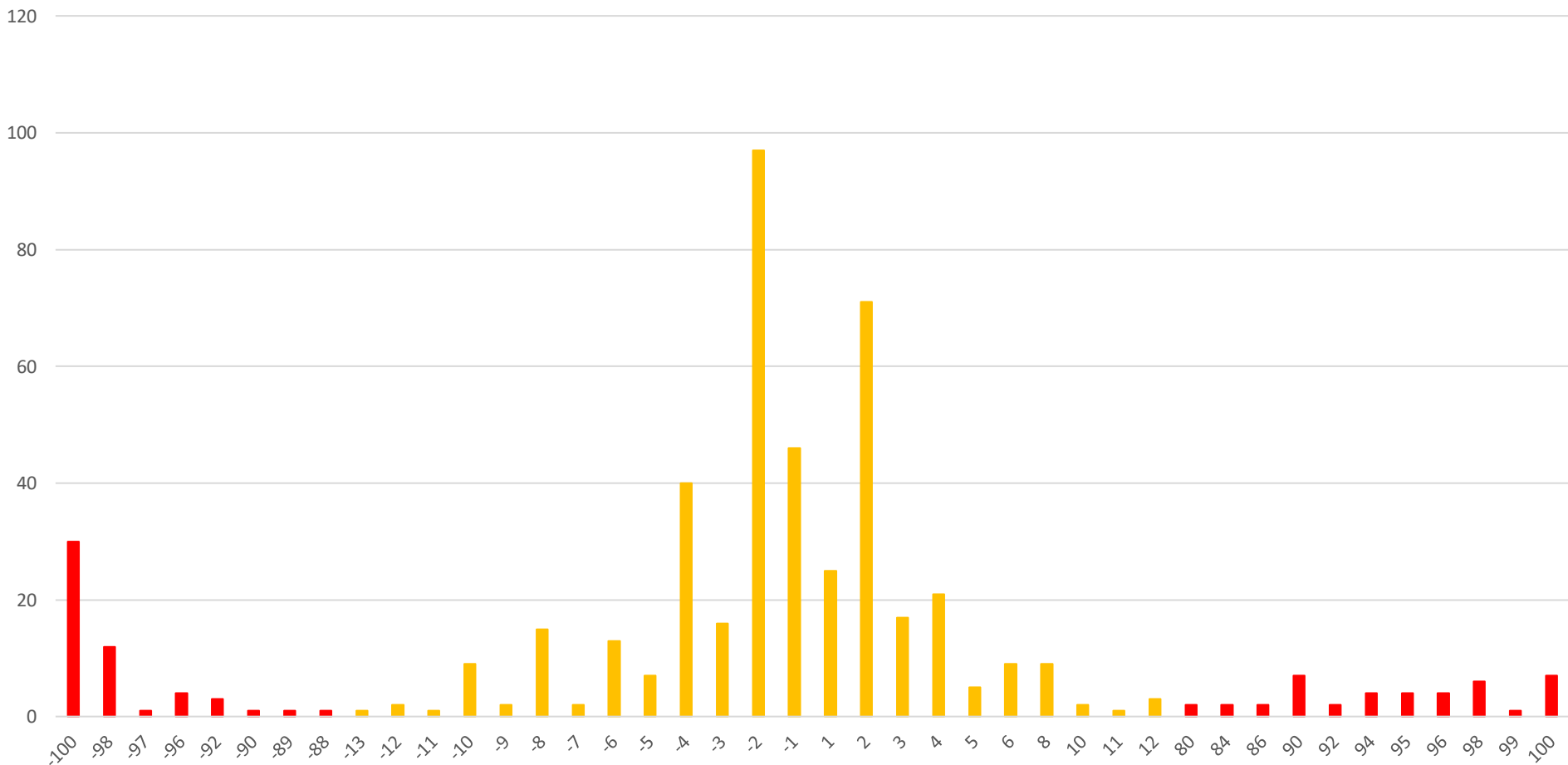


Incorrect Loading vs Incorrect Decision





Failed Predictions Grouped by Difference



- High accuracy for standard class, large loadings (400%+), as well as a few other classes/loadings
- Declined class has fairly high accuracy
- A few classes with low accuracies have a small number of data points, e.g. loading 350%

	Class	Precision	Recall	Accuracy	Count
0	Standard	0.848	0.923	0.923	831.0
1	Loading 25%	0.941	0.653	0.653	49.0
2	Loading 50%	0.695	0.772	0.772	360.0
3	Loading 75%	0.743	0.677	0.677	124.0
4	Loading 100%	0.773	0.736	0.736	227.0
5	Loading 125%	0.846	0.767	0.767	43.0
6	Loading 150%	0.875	0.798	0.798	114.0
7	Loading 175%	1.000	0.800	0.800	15.0
8	Loading 200%	0.673	0.507	0.507	69.0
9	Loading 225%	0.800	0.500	0.500	8.0
10	Loading 250%	0.797	0.638	0.638	80.0
11	Loading 275%	0.857	0.857	0.857	7.0
12	Loading 300%	0.806	0.694	0.694	36.0
13	Loading 325%	NaN	NaN	NaN	NaN
14	Loading 350%	0.700	0.467	0.467	15.0
15	Loading 375%	NaN	NaN	NaN	NaN
16	Loading 400%	0.957	0.710	0.710	31.0
17	Loading 400%+	0.630	0.895	0.895	19.0
18	Decline	0.787	0.725	0.725	178.0

# ***SAMPLE FAILED PREDICTION***

## Policy 1

Actual decision: **standard**

Predicted decision: **decline**

TA with SA < 1 mil

Age: 72

BMI: 26 (overweight)

Occupational Class: 0

Company X

## Policy 2

Actual decision: **decline**

Predicted decision: **standard**

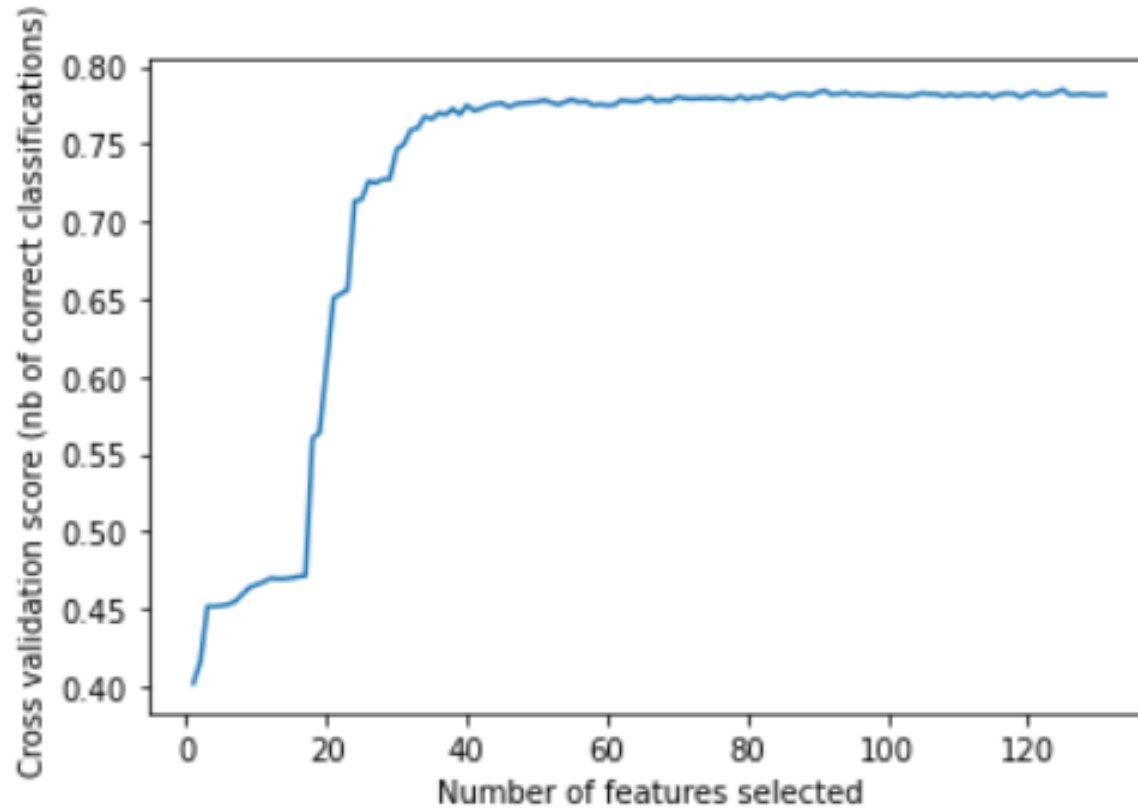
TA with SA < 1 mil

Age 33

BMI: 20 (healthy)

Occupational Class: 3

Company M

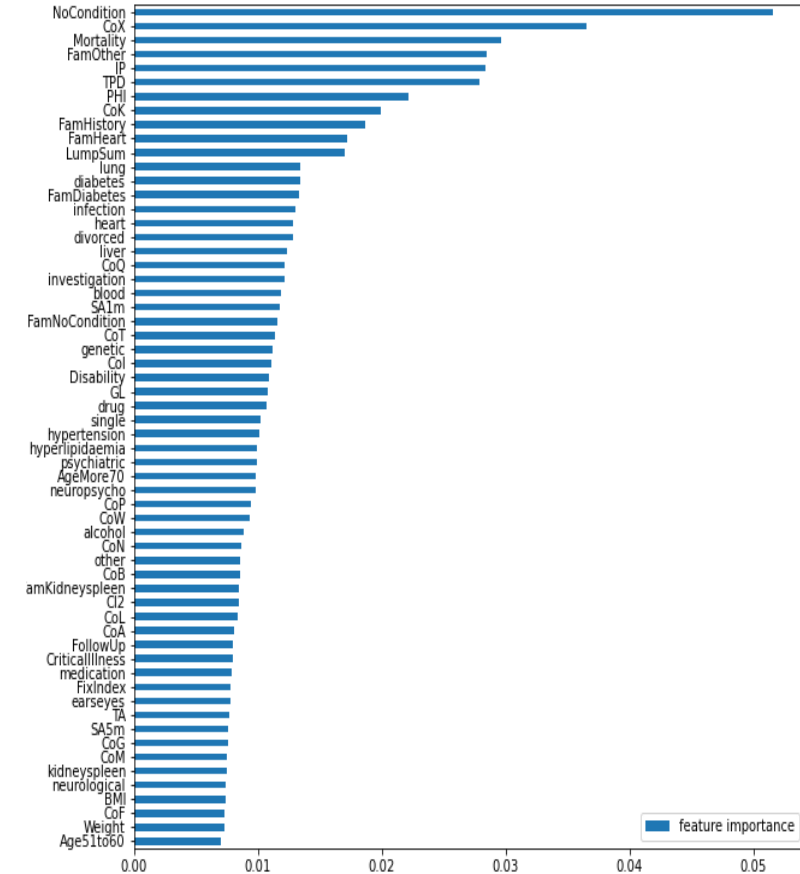
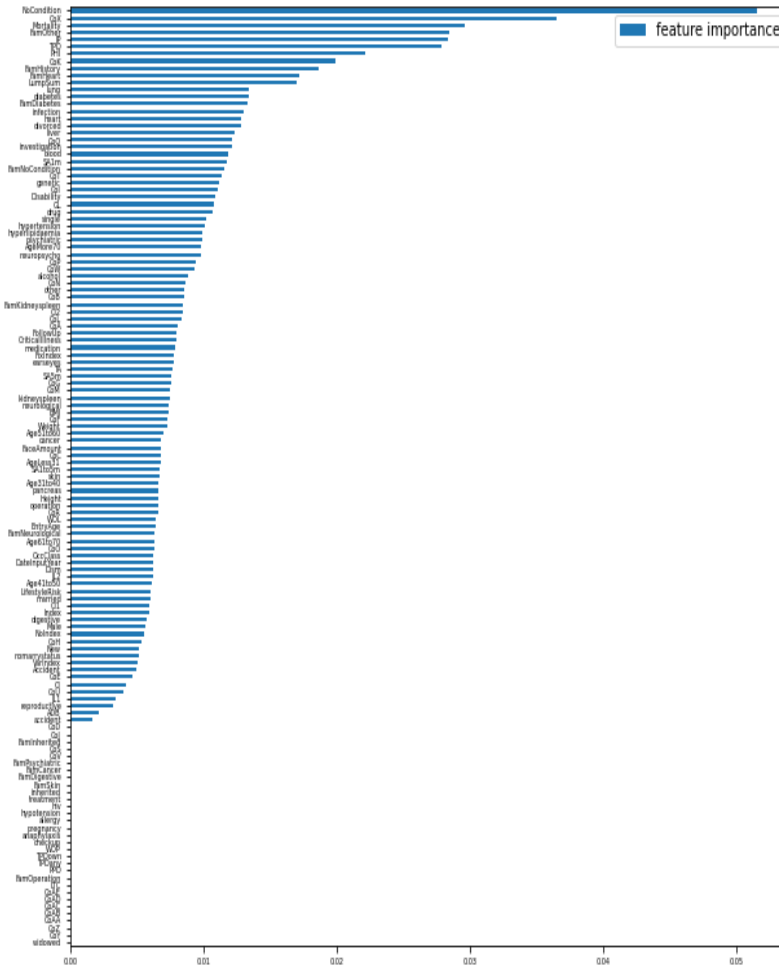


Optimal number of features required: 125

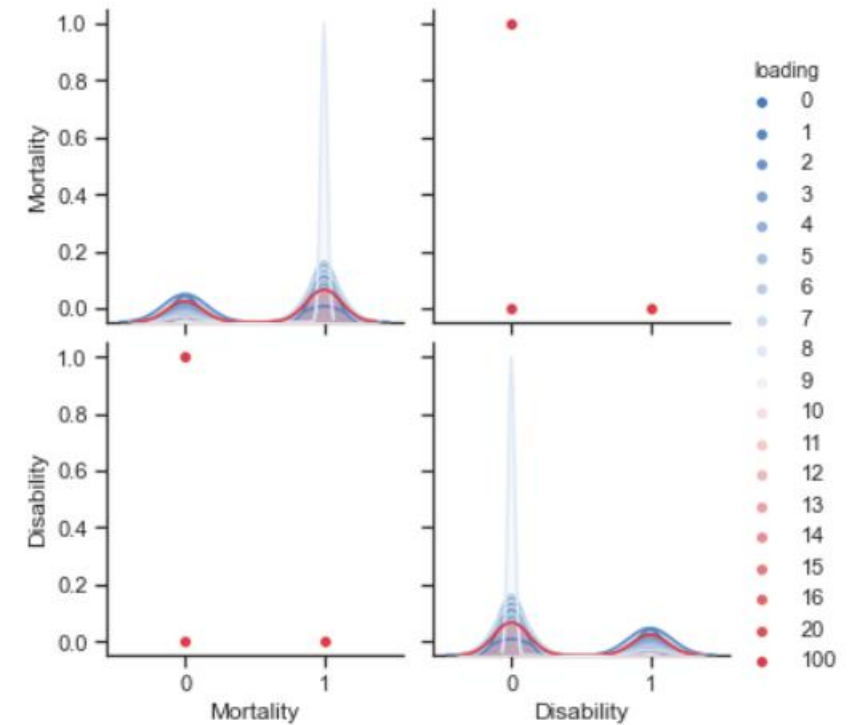
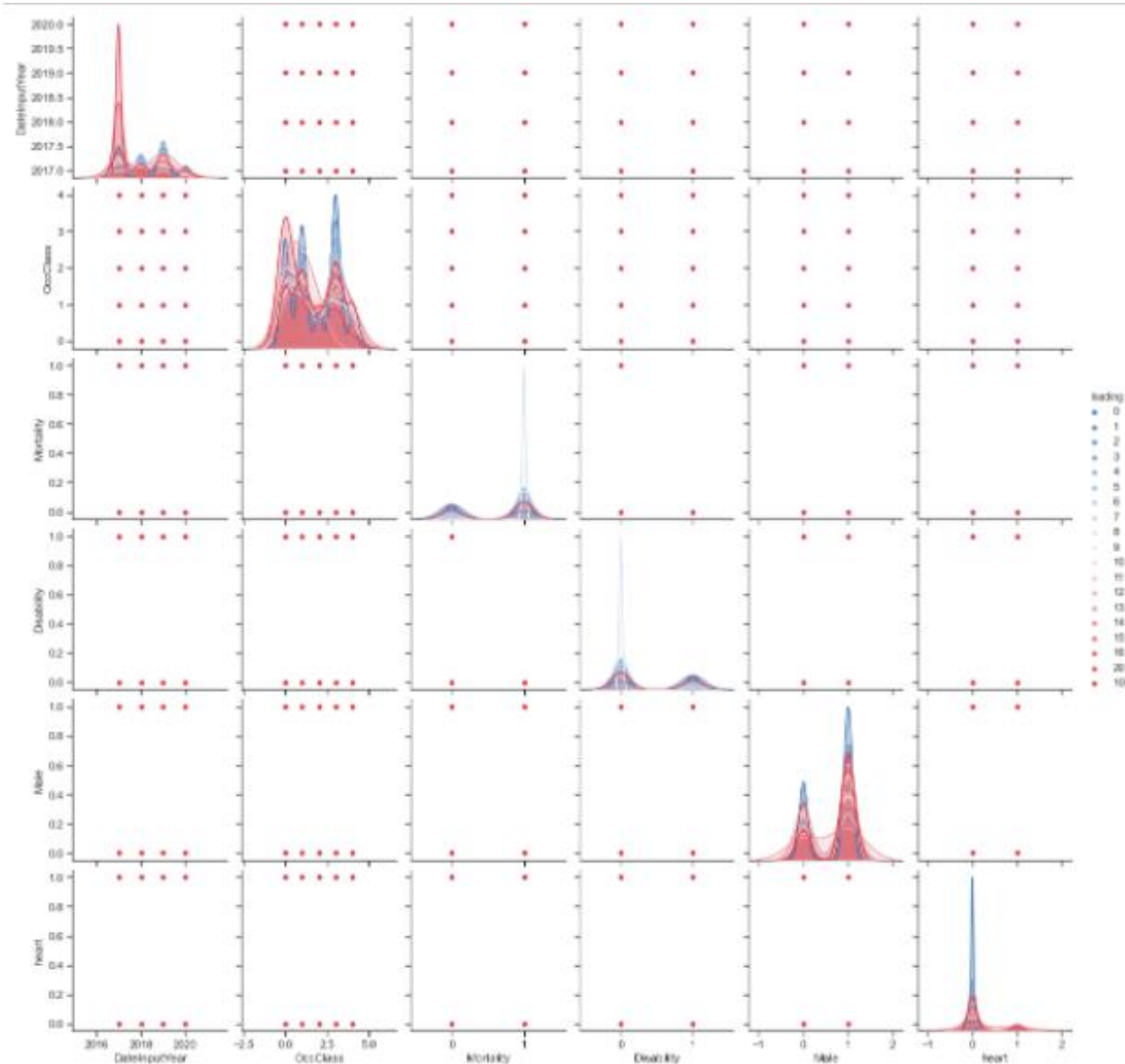
Performance flattens after about 40 features

Sharp increase in performance between 20 to 40 features

# FEATURE RANKING



# SCATTERPLOT MATRIX



## 19-class classification

	Model_Name	Precision	Recall	Train_Accuracy	Test_Accuracy	F1_Score
7	XGBClassifier	0.807	0.807	1.000	0.807	0.807
1	RandomForestClassifier	0.787	0.787	1.000	0.787	0.787
4	BaggingClassifier	0.786	0.786	1.000	0.786	0.786
0	DecisionTreeClassifier	0.688	0.688	1.000	0.688	0.688
5	GradientBoostingClassifier	0.684	0.684	1.000	0.684	0.684
9	KNeighborsClassifier	0.421	0.421	0.640	0.421	0.421
3	AdaBoostClassifier	0.369	0.369	0.349	0.369	0.369
6	SVC	0.342	0.342	0.320	0.342	0.342
2	LogisticRegression	0.337	0.337	0.314	0.337	0.337
8	SGDClassifier	0.116	0.116	0.118	0.116	0.116

## 3-class classification

	Model_Name	Precision	Recall	Train_Accuracy	Test_Accuracy	F1_Score
7	XGBClassifier	0.911	0.911	1.000	0.911	0.911
1	RandomForestClassifier	0.904	0.904	1.000	0.904	0.904
4	BaggingClassifier	0.904	0.904	1.000	0.904	0.904
5	GradientBoostingClassifier	0.878	0.878	1.000	0.878	0.878
0	DecisionTreeClassifier	0.864	0.864	1.000	0.864	0.864
3	AdaBoostClassifier	0.802	0.802	0.784	0.802	0.802
9	KNeighborsClassifier	0.680	0.680	0.790	0.680	0.680
6	SVC	0.640	0.640	0.658	0.640	0.640
2	LogisticRegression	0.589	0.589	0.609	0.589	0.589
8	SGDClassifier	0.589	0.589	0.609	0.589	0.589



- Accuracy: 99.5% on training set and 80% on test set
- Accuracy by each class indicates the model is rather accurate at predicting standard class, i.e. accuracy 92% on test set
- Majority of incorrect predictions are small differences in predicted loadings as opposed incorrect bucket (0.8%)
- Feature ranking provide underwriting insights
- Data improvement, Post-modelling analysis and discussion with underwriters could further improve results





# EAA e-Conference on Data Science & Data Ethics

12 May 2022

# Thank you

## Contact

# Yafei (Patricia) Wang

*Email: [patriciawangyafei@gmail.com](mailto:patriciawangyafei@gmail.com)*

LinkedIn: <https://www.linkedin.com/in/patricia-wang-1446894a/>