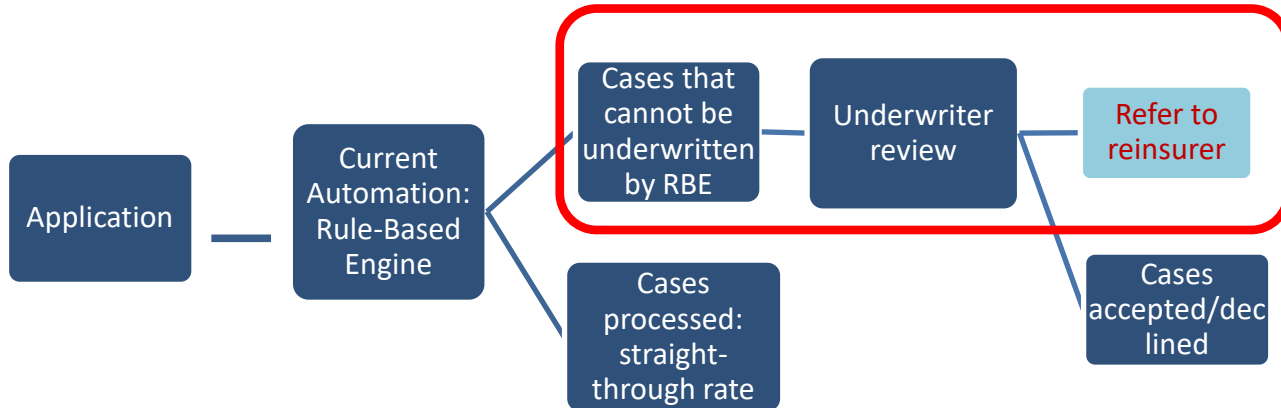# Machine Learning and XAI To Predict Underwriting Decision of Life and Health Insurance

Yafei (Patricia) Wang, Lloyd's of London

Existing Approach

- ML model trained on cases that RBE fails to process
- Predictive ML vs deterministic RBE
- XAI highlight data issues, errors, inconsistencies..

Methodology

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

- Multiclass classification
- Loading: an indication of riskiness, not calculated from a formula
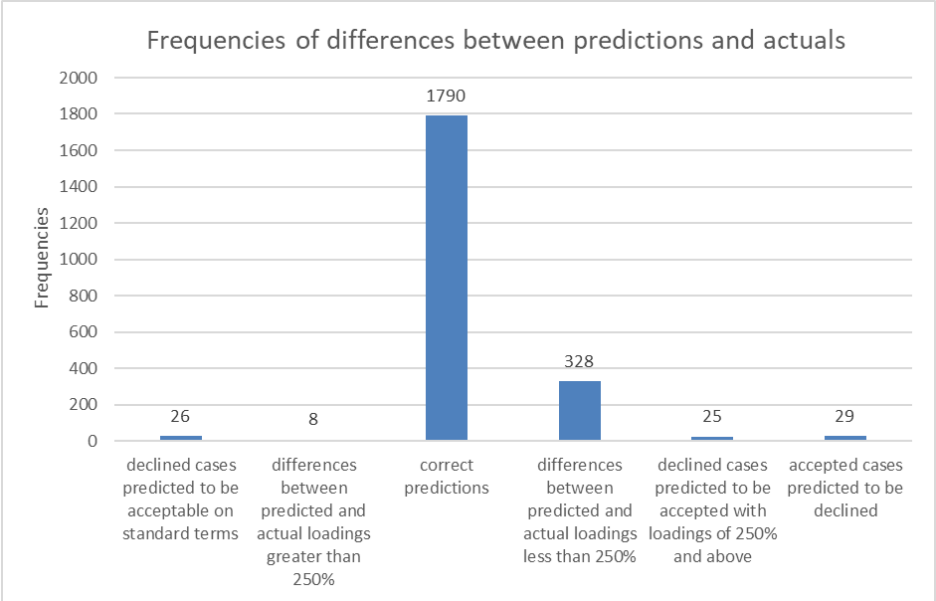- >400%: grouped together in one class

Model Performances

Model_Name	Precision Train	Recall Train	F1 Score Train	Train Accuracy	Precision Test	Recall Test	F1 Score Test	Test Accuracy
XGBClassifier	0.997	0.993	0.995	0.995	0.799	0.72	0.754	0.811
BaggingClassifier	0.996	0.994	0.995	0.995	0.703	0.612	0.647	0.775
RandomForestClassifier	0.996	0.994	0.995	0.995	0.703	0.615	0.65	0.774
GradientBoostingClassifier	0.997	0.994	0.995	0.995	0.662	0.636	0.645	0.757
DecisionTreeClassifier	0.995	0.995	0.995	0.995	0.55	0.556	0.55	0.691
KNeighborsClassifier	0.581	0.501	0.534	0.66	0.375	0.293	0.312	0.471
AdaBoostClassifier	0.072	0.09	0.072	0.442	0.066	0.089	0.071	0.438
LogisticRegression	0.021	0.053	0.03	0.39	0.02	0.053	0.029	0.388
SVC	0.073	0.053	0.03	0.39	0.073	0.053	0.03	0.388
SGDClassifier	0.003	0.053	0.006	0.056	0.003	0.053	0.006	0.061

Table 2: Performances of all models

XGB: Model Performances By Class

Class label	Class Description	Precision Test	Recall Test	Accuracy Test	Count Test	Precision Train	Recall Train	Accuracy Train	Count Train
0	Standard	0.853	0.938	0.938	855	0.997	0.996	0.996	3436
1	Loaded 25	0.821	0.696	0.696	46	0.996	0.996	0.996	235
2	Loaded 50	0.713	0.771	0.771	345	0.988	0.996	0.996	1370
3	Loaded 75	0.852	0.687	0.687	134	0.992	0.986	0.986	496
4	Loaded 100	0.741	0.76	0.76	200	0.997	0.994	0.994	870
5	Loaded 125	0.953	0.788	0.788	52	0.995	0.986	0.986	216
6	Loaded 150	0.814	0.731	0.731	108	0.993	0.998	0.998	448
7	Loaded 175	0.833	0.667	0.667	15	1	0.987	0.987	78
8	Loaded 200	0.794	0.617	0.617	81	1	0.996	0.996	279
9	Loaded 225	1	0.75	0.75	8	1	1	1	26
10	Loaded 250	0.821	0.667	0.667	69	1	0.988	0.988	252
11	Loaded 275	1	0.909	0.909	11	1	0.964	0.964	28
12	Loaded 300	0.711	0.73	0.73	37	0.994	0.989	0.989	177
13	Loaded 325	0	0	0	2	1	1	1	3
14	Loaded 350	0.818	0.6	0.6	15	1	1	1	54
15	Loaded 375	1	1	1	1	1	1	1	1
16	Loaded 400	0.864	0.76	0.76	25	1	1	1	93
20	Loaded 400+	0.773	0.895	0.895	19	1	1	1	122
100	Decline	0.819	0.716	0.716	183	0.989	0.995	0.995	637
Overall	All Classes	0.799	0.72	0.811	2206	0.997	0.993	0.995	8821

XGB Accuracy

Frequencies of differences between predictions and actuals

LIME

SHAP

XGB(C)

XGB(B,C)

No feature

XGB(B)

XGB(A,C)

XGB(A,B,C)

XGB(A)

XGB(A,B)
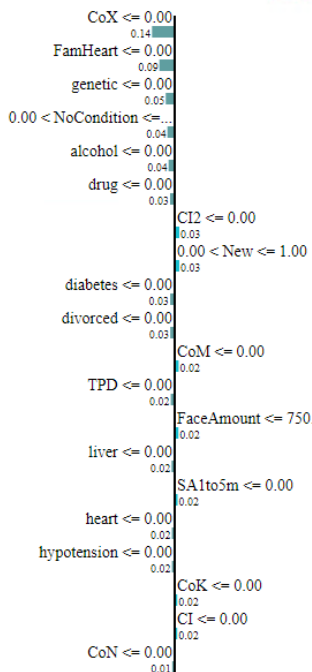
- **SH**apley **A**dditive ex**P**lanations
- Game theory
- Add features sequentially
- Consider all combinations
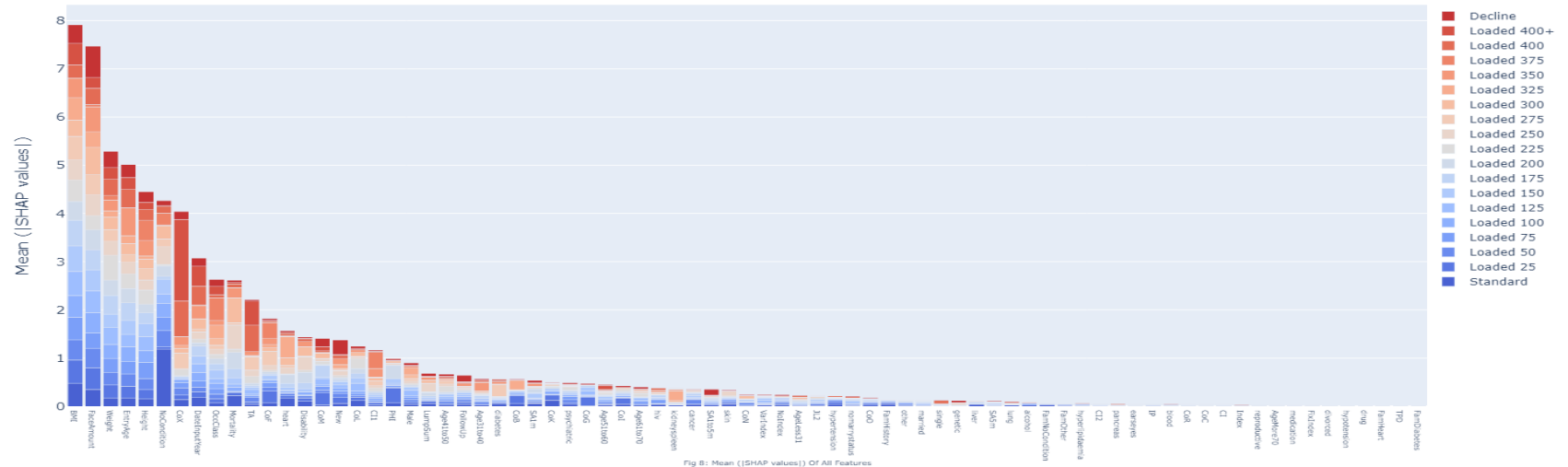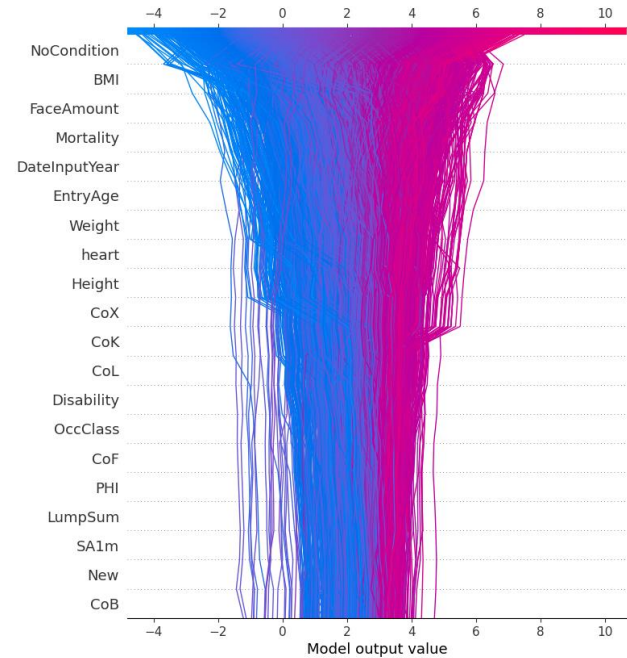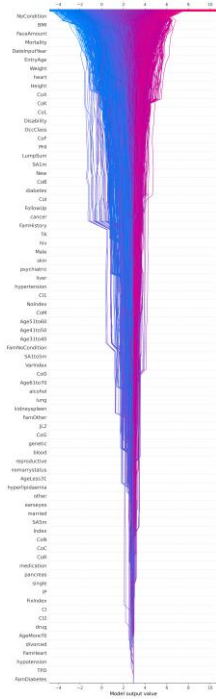- SHAP value = weighted average
- Model agonistic
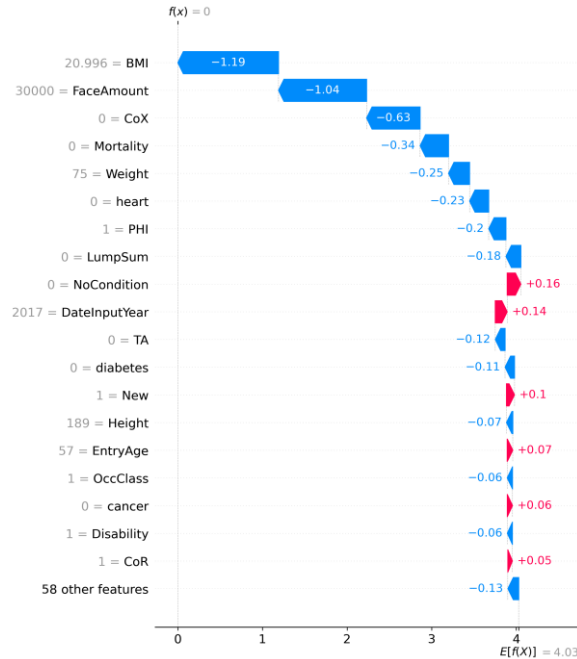- More stable than model-dependent feature ranking
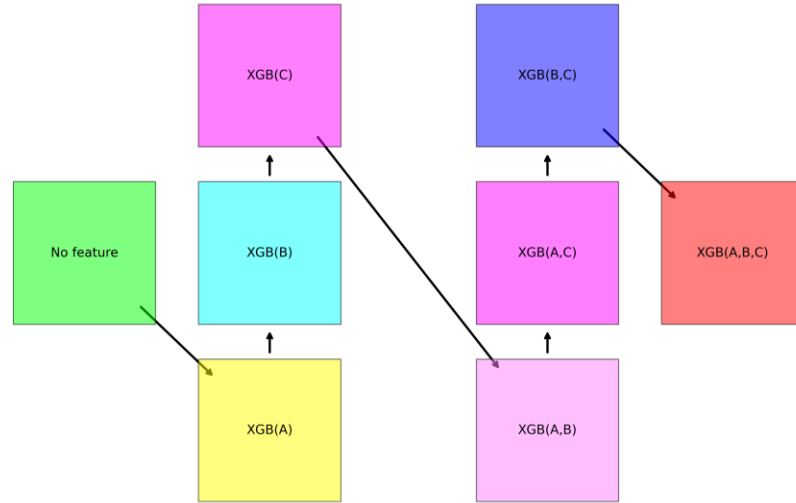
Figure 1: Methodology flowchart

SHAP Feature Ranking

Fig 8: Mean (|SHAP values|) Of All Features

SHAP Feature Ranking

Fig 9: Mean (|SHAP|) for Standard vs Decline Class
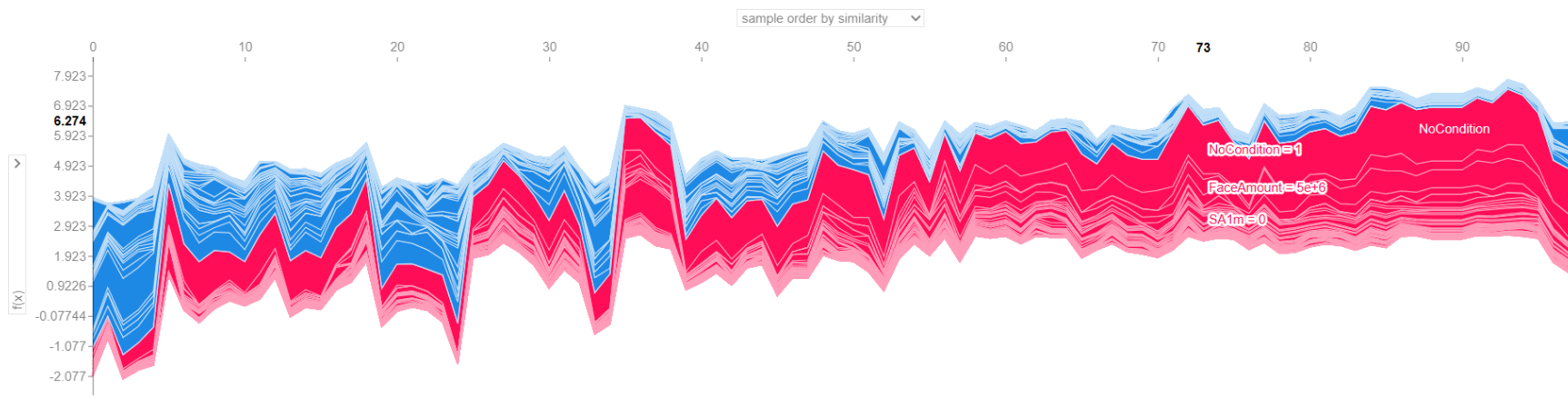
SHAP: Decision Plot
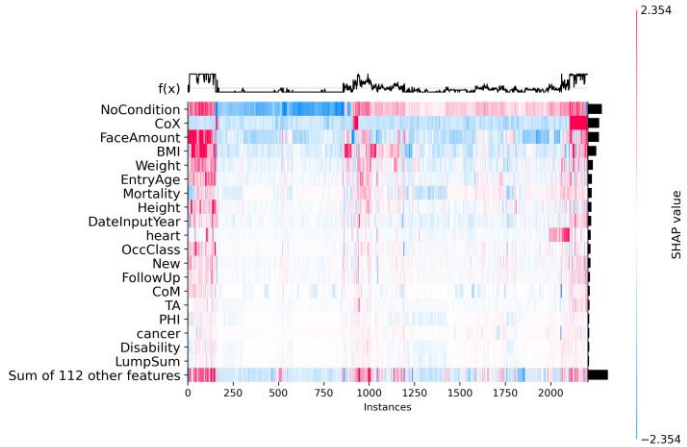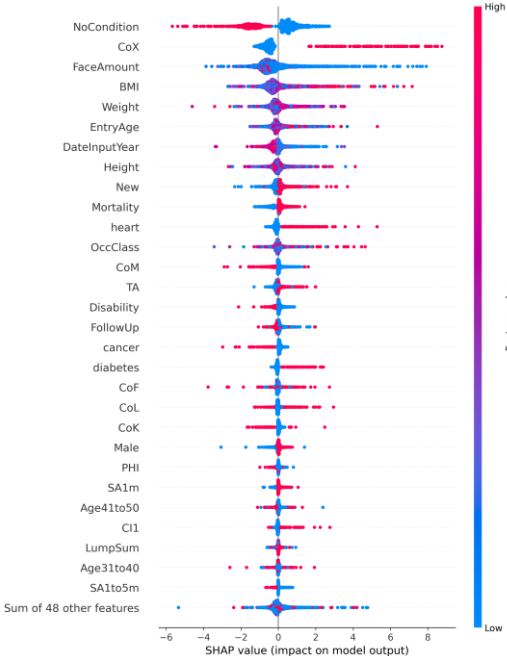
SHAP: Single Instance

- Ranked using absolute values, but show negative and positive sign
- Re-rank features for instance of interest
- Show the values for all the features of this instance
- Easy to explain to applicant or broker
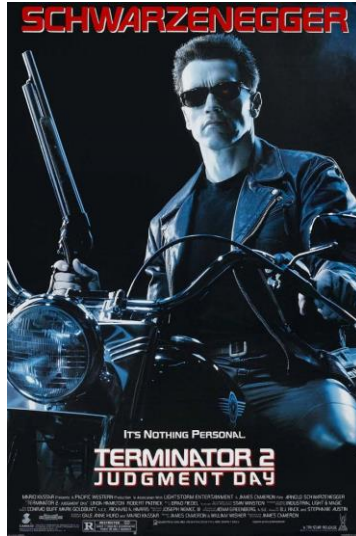
SHAP: Cohort

SHAP: Other plots

Conclusion

- Accuracy: 99.5% on training set and 81.1% on test set

- Accuracy by each class indicates the model is rather accuracy at predicting standard class, i.e. accuracy 93.8% on test set

- Majority of incorrect predictions are small differences in predicted loadings as opposed incorrect bucket (55 cases, 2.5% of testing dataset)

- XAI explain model outputs

- XAI provide underwriting insights

- XAI highlights some data issues, human errors and inconsistencies

- Post-modelling analysis, data collection and improvement, discussion with underwriters could further improve model performances

Thank You

TO HELP, NOT TO KILL