Месторождения ошибок

Анализ пространства ошибок моделей машинного обучения

Дмитрий Колодезев
ООО Промсофт, Новосибирск
DataConf Barnaul 18.06.2021

План

- Модели и ошибки
- Пространство ошибок
- История вопроса
- Месторождения ошибок
- Источники ошибок
- Инструменты
- Что делать

Ошибки

- Когда модель отдает не то, что мы хотим
- Оценка «ошибочности» модели
  - Как оцениваем
    - Метрики регрессии MAE MSE ...
    - Метрики классификации Accuracy, F-мера, ROC-AU ...
    - Точечные оценки
  - На чем оцениваем
    - Метрики на тесте
    - Метрики на OOF
    - На всем датасете

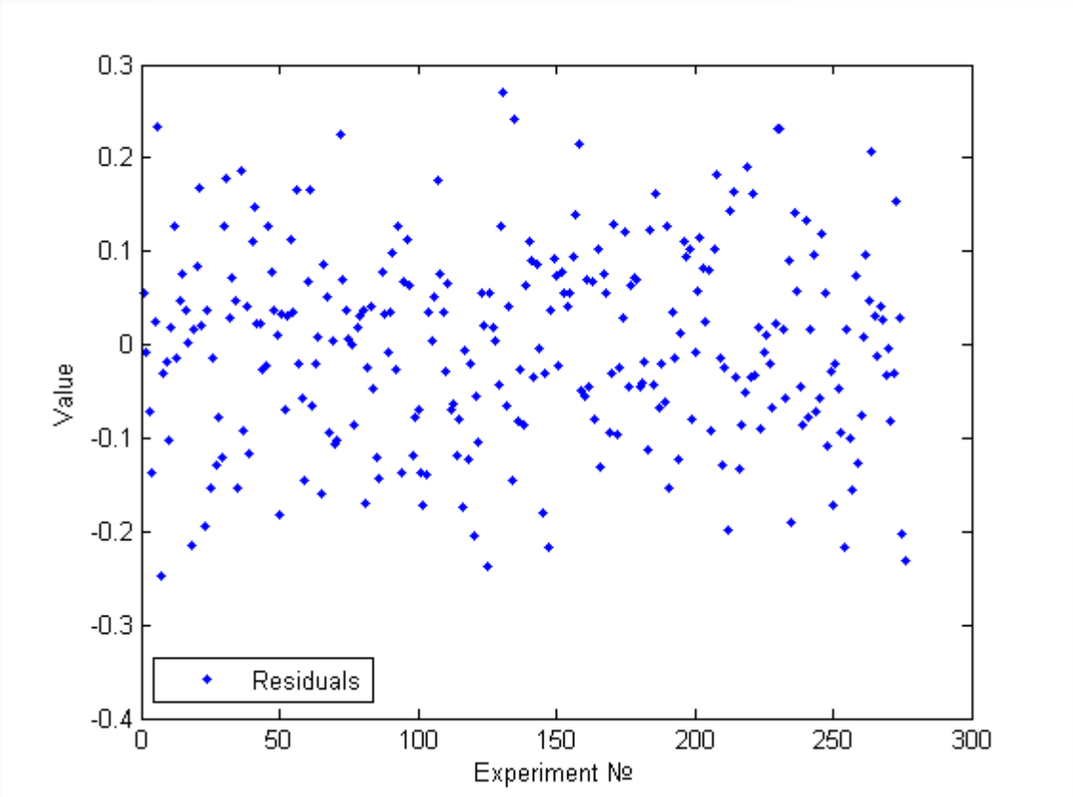
Пространство ошибок

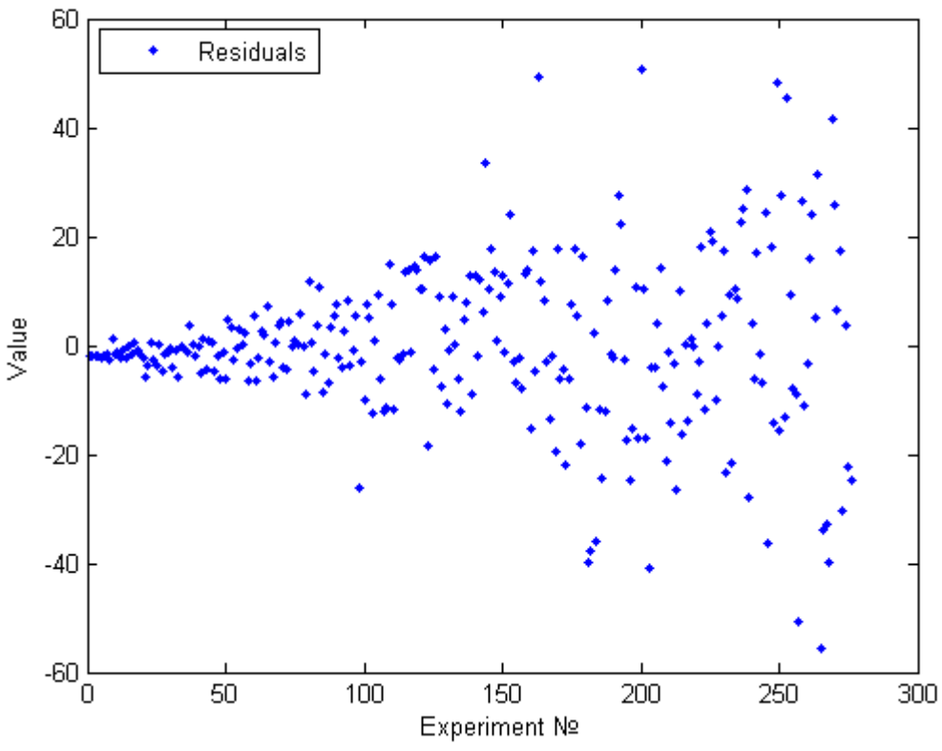
- Берем пространство признаков
- Добавляем точечную оценку ошибки
- Получаем пространство ошибок
- Например, данные медосмотра

id	age	gender	height	weight	ap_hi	ap_lo	cholesterol	gluc	smoke	alco	active	cardio	bmi
0	18393	2	168	62.0	110	80	1	1	0	0	1	0	21.967120
1	20228	1	156	85.0	140	90	3	1	0	0	1	1	34.927679
2	18857	1	165	64.0	130	70	3	1	0	0	0	1	23.507805
3	17623	2	169	82.0	150	100	1	1	0	0	1	1	28.710479
4	17474	1	156	56.0	100	60	1	1	0	0	0	0	23.011177

```
Age (days)
Height (cm)
Weight (kg)
Gender
Systolic blood pressure
Diastolic blood pressure
Cholesterol
Glucose
Smoking
Alcohol intake
Physical activity
Cardiovascular disease
```

Анализ регрессионных остатков





http://www.machinelearning.ru/wiki/index.php?title=Анализ_регрессионных_остатков

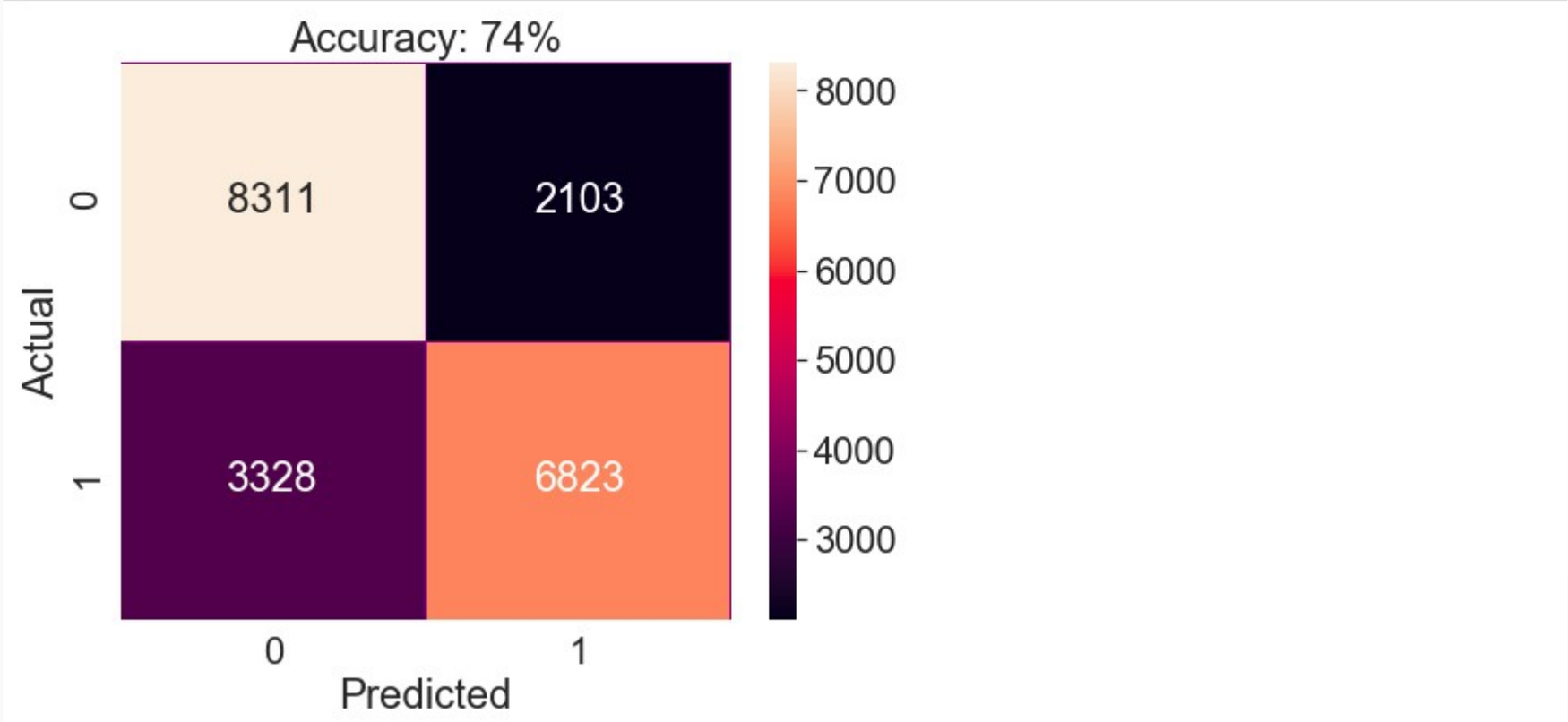
Почистим, поделим, обучим

```
df = pd.read_csv(DATA_FILE, sep=';', index_col='id')
df['bmi'] = df.weight/(df.height/100)**2

idx = df.query('(ap_hi < 50) | (ap_lo < 50) | (ap_lo > 240) | (ap_hi > 300) | (weight < 40) '
               '| (height < 120) | (bmi > 100) | (ap_lo > (ap_hi-10))').index
df_clean = df.drop(idx, axis=0)

X = df_clean.drop('cardio', axis=1)
y = df_clean.cardio
features = list(X.columns)

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3,random_state=SEED)

model = RandomForestClassifier(**model_params)
model.fit(X_train.values, y_train.values)

pred = model.predict(X_test.values)
pred_proba = model.predict_proba(X_test.values)[:, 1]
```

Датасет https://www.kaggle.com/sulianova/cardiovascular-disease-dataset

Оценим

```
df_cm = pd.DataFrame(confusion_matrix(y_test, pred), columns=np.unique(y_test), index = np.unique(y_test))
df_cm.index.name = 'Actual'
df_cm.columns.name = 'Predicted'
plt.figure(figsize=(8, 6))
score = accuracy_score(y_test.values, pred)
plt.title('Accuracy: {:.0%}'.format(score))
sns.heatmap(df_cm, fmt=".0f", annot=True,linewidths=0.2, linecolor="purple", );
```

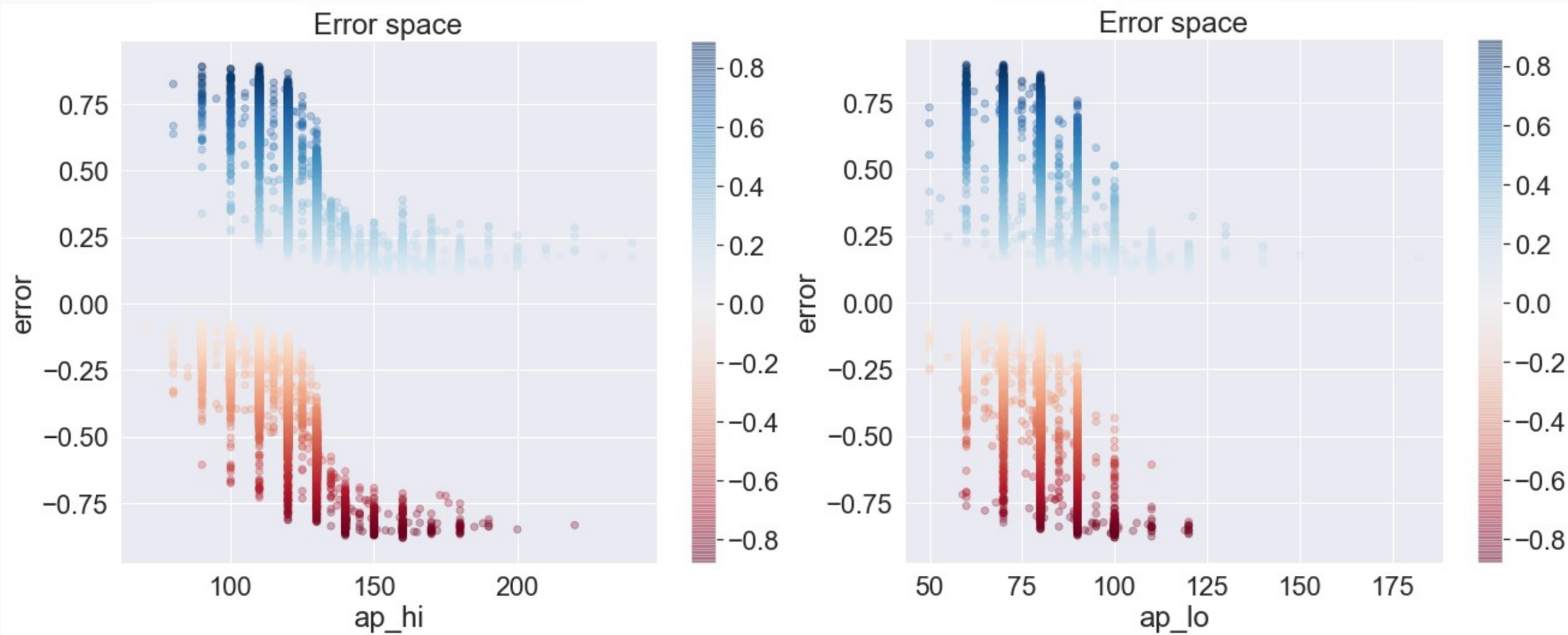


Соберем ошибки

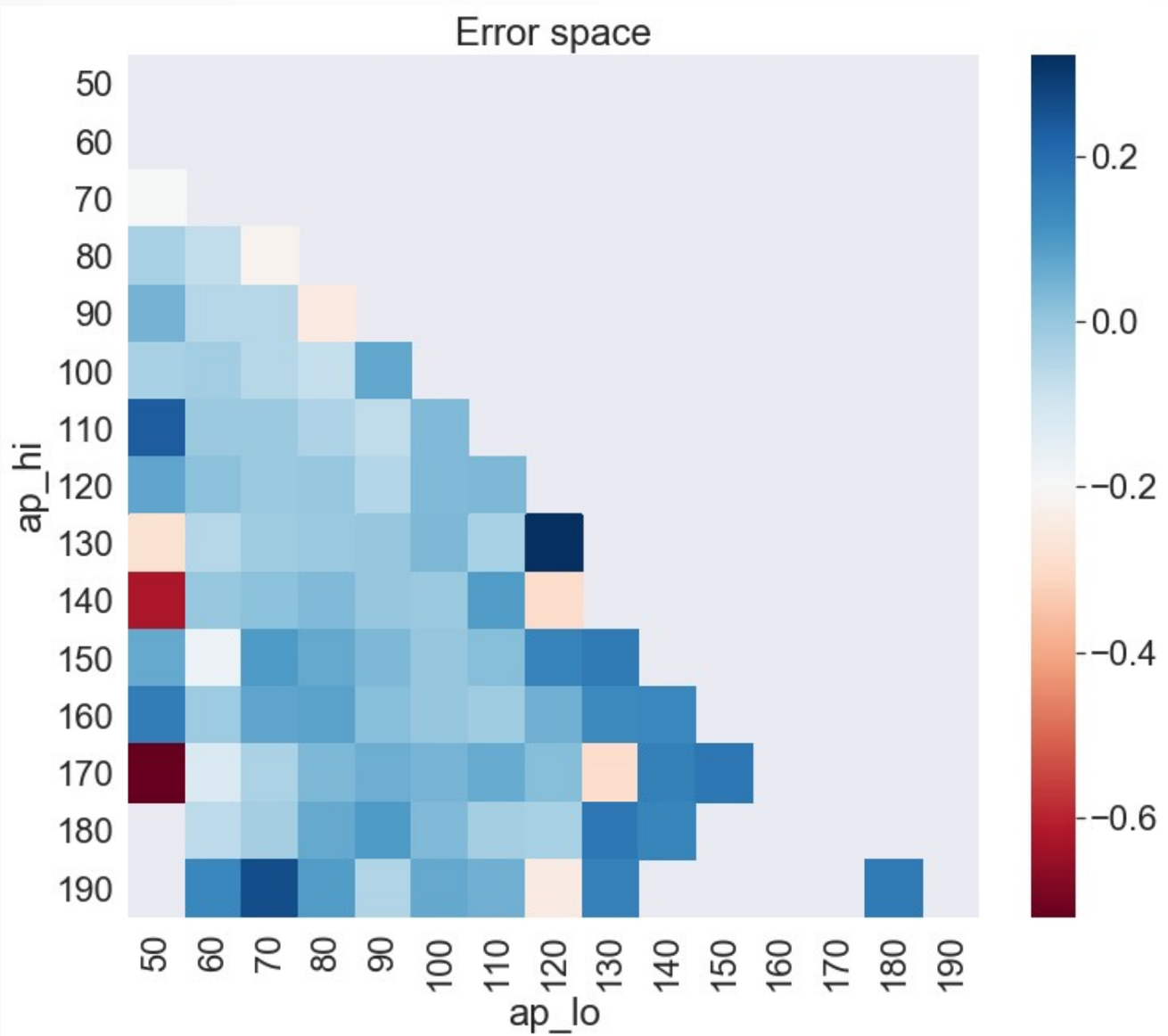
```
df_error = X_test.copy()
df_error['error'] = (y_test - pred_proba)
df_error['abs_error'] = (y_test - pred_proba).abs()
df_error.head()
```

id	age	gender	height	weight	ap_hi	ap_lo	cholesterol	gluc	smoke	alco	active	bmi	error	abs_error
10289	23636	1	160	63.0	120	80	1	1	0	0	0	24.609375	-0.528284	0.528284
17843	19676	1	169	75.0	120	80	1	1	0	0	1	26.259585	-0.276148	0.276148
9901	21032	1	168	67.0	120	70	1	1	0	0	1	23.738662	0.629541	0.629541
93721	20214	2	165	90.0	130	80	1	1	1	0	0	33.057851	-0.558215	0.558215
37683	20438	1	160	64.0	100	60	1	1	0	0	1	25.000000	-0.281052	0.281052

Ошибка зависит



Взгляд сверху



Microsoft raiWidget

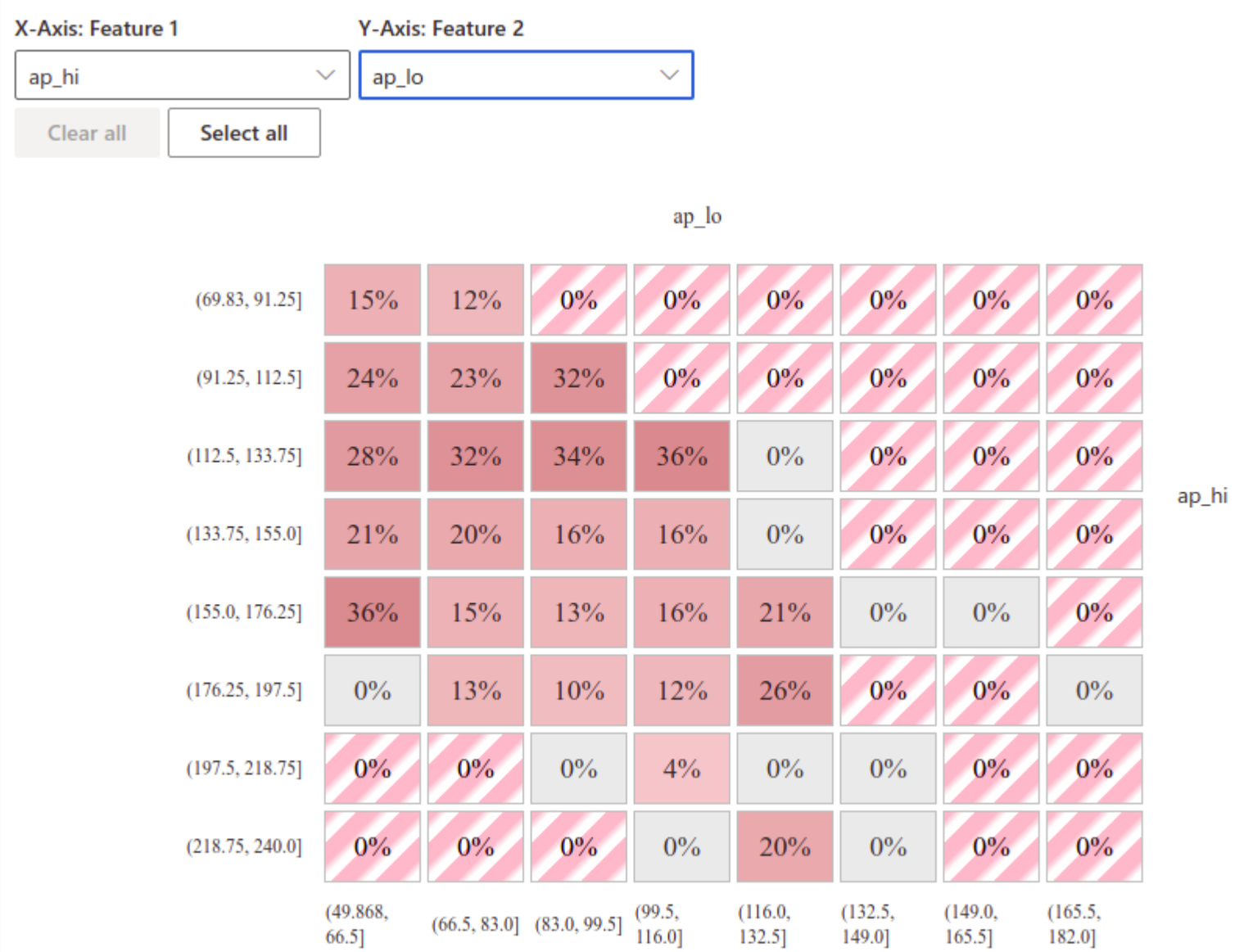
- https://github.com/microsoft/responsible-ai-widgets
- https://github.com/interpretml/interpret-community
- https://github.com/fairlearn/fairlearn

```
from raiwidgets import ErrorAnalysisDashboard
```

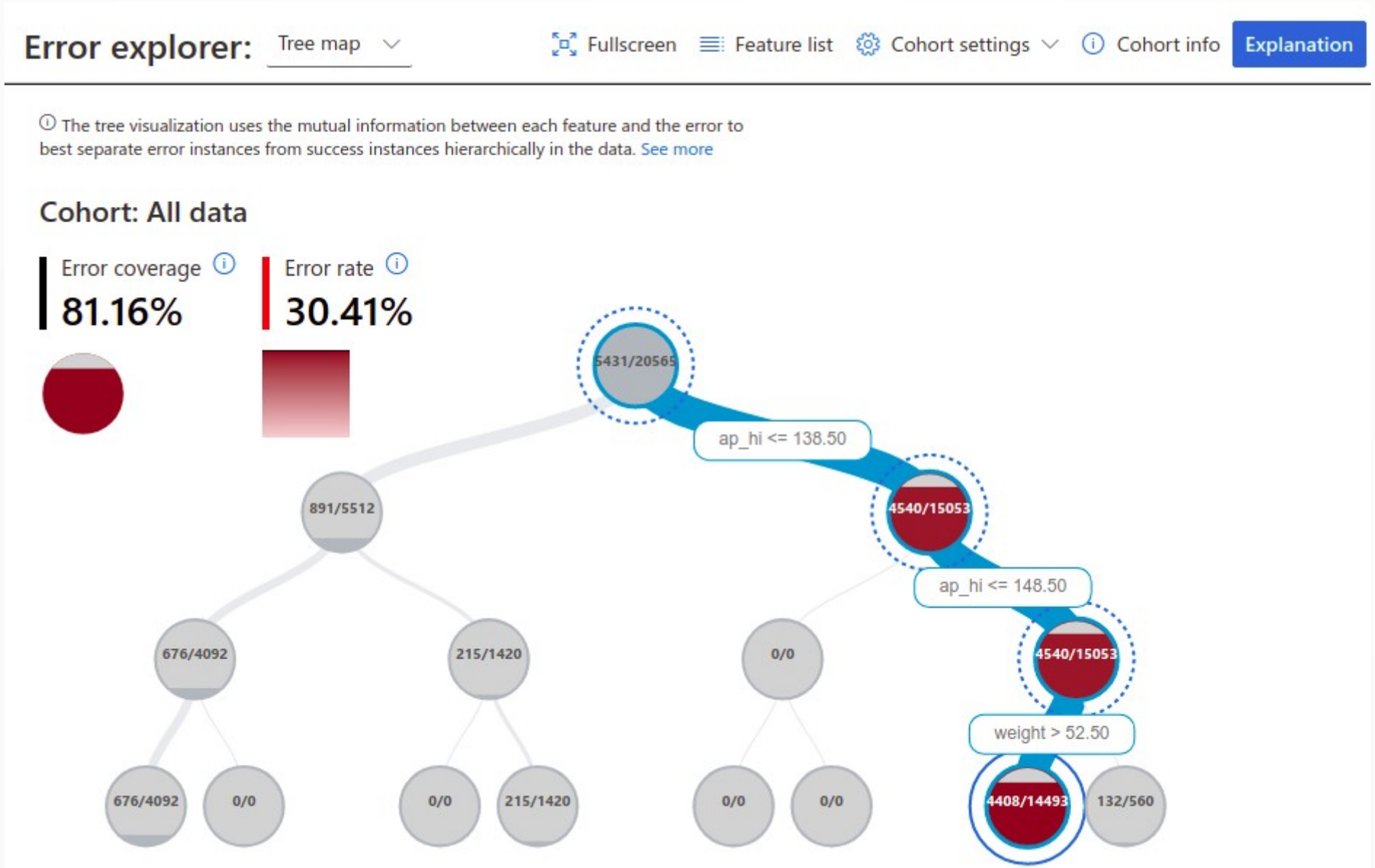
```
ErrorAnalysisDashboard(dataset=X_test.values, true_y=y_test.values, features=features, pred_y=pred);
```

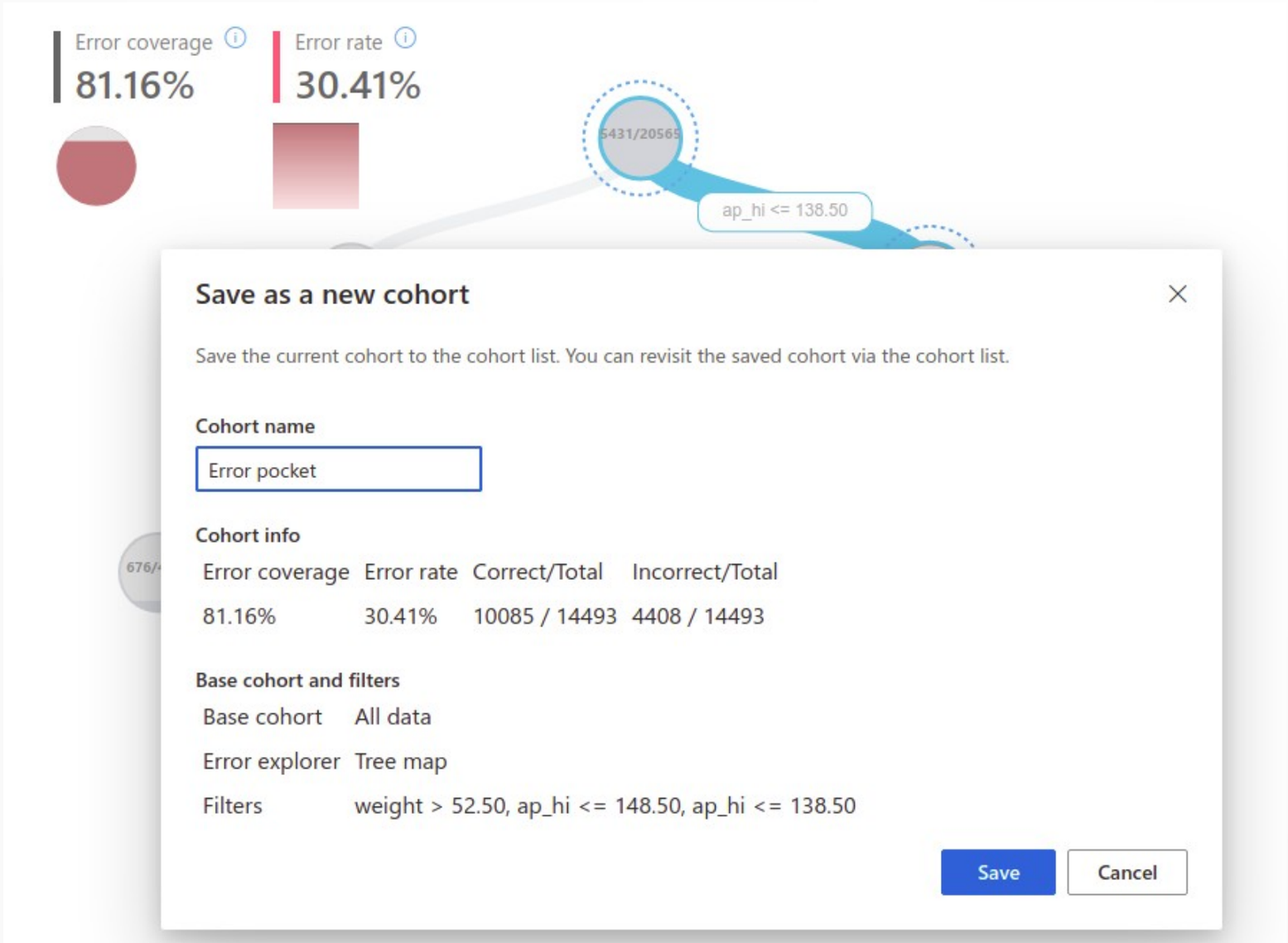
```
ErrorAnalysis started at http://localhost:5001
```

Карта ошибок

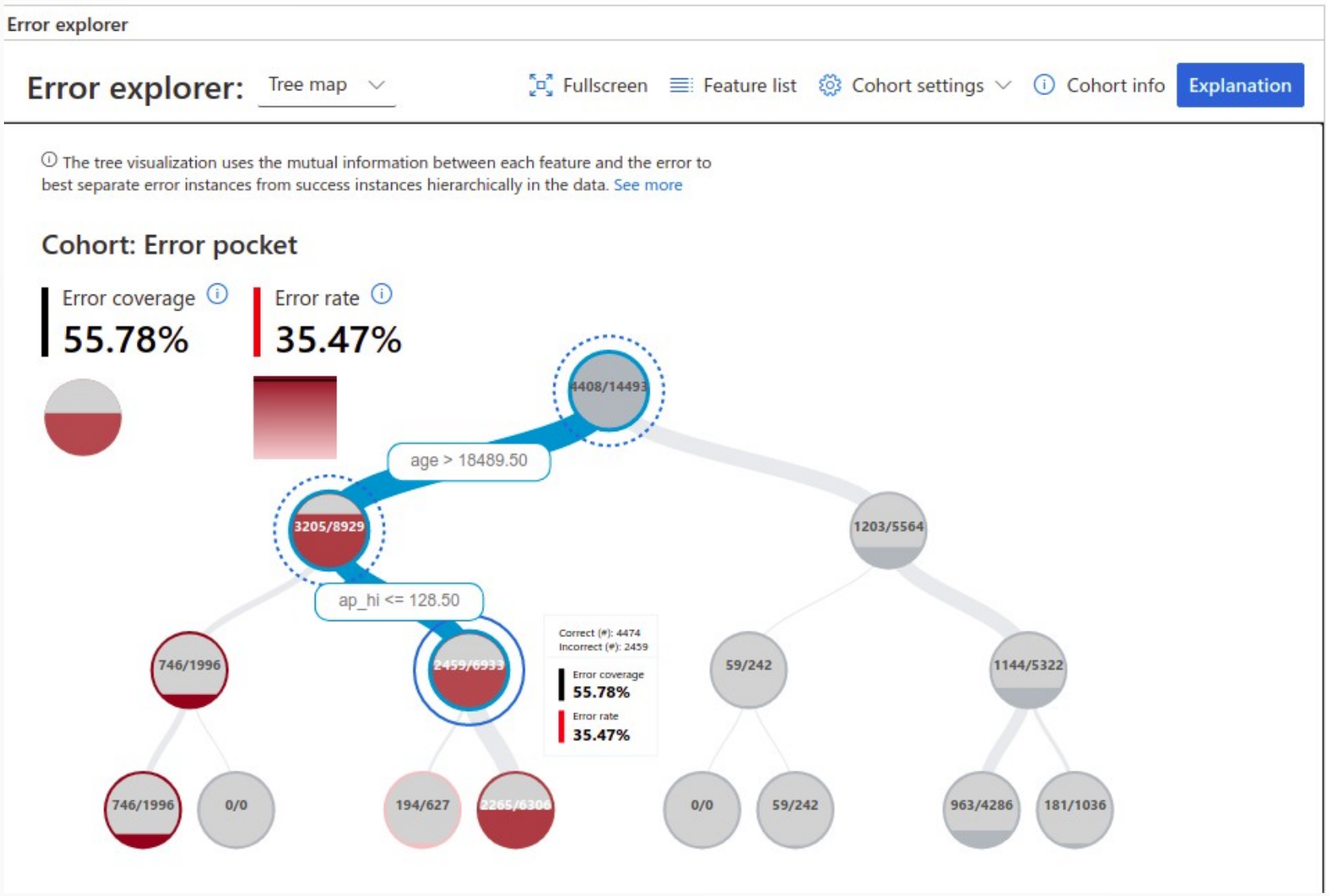


Дерево ошибок




Error coverage
81.16%
Error rate
30.41%
5431/20565
ap_hi <= 138.50
Save as a new cohort
Save the current cohort to the cohort list. You can revisit the saved cohort via the cohort list.
Cohort name
Error pocket
Cohort info
Error coverage
Error rate
Correct/Total
Incorrect/Total
81.16%
30.41%
10085 / 14493
4408 / 14493
Base cohort and filters
Base cohort
All data
Error explorer
Tree map
Filters
weight > 52.50, ap_hi <= 148.50, ap_hi <= 138.50
Save
Cancel

Работа с когортами



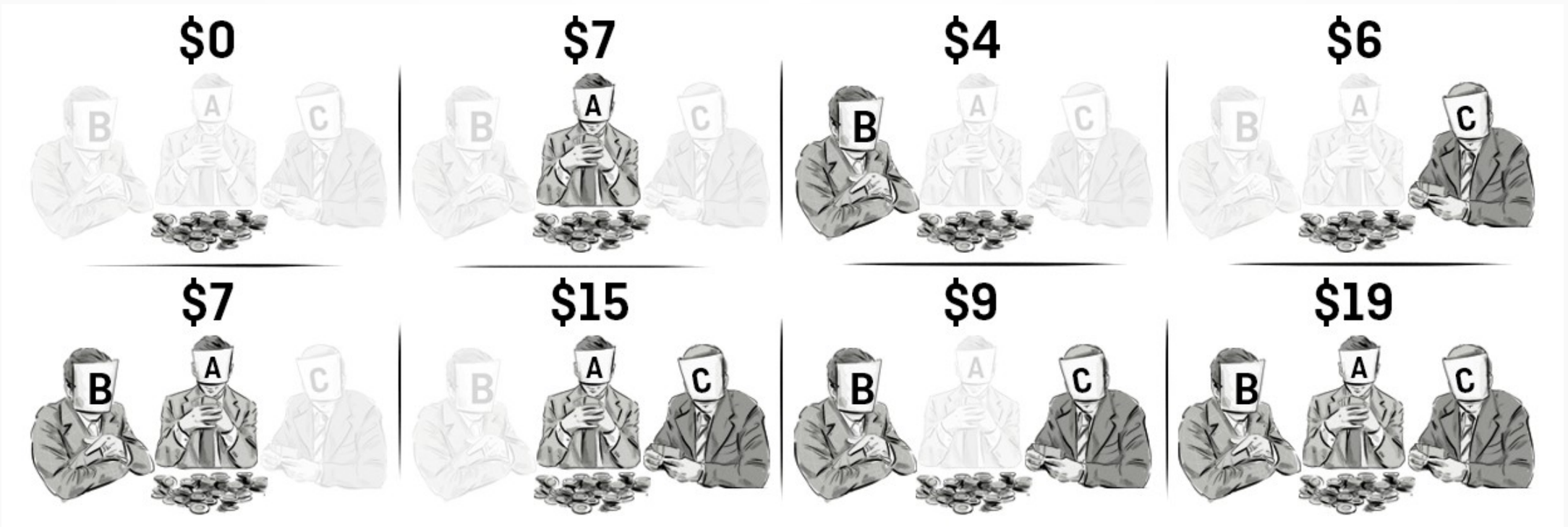
Источники ошибок

- Исходные данные
- Генерация признаков
- Генерация признаков внутри модели
- Особенности моделей
- Параметры моделей

Что делать

- Идентифицировать и рассмотреть месторождения ошибок
- Сравнить месторождения для разных типов моделей
- Добавить новые признаки
- Не верить ответу модели в этих случаях
- Стекать!

Shapley Values



https://clearcode.cc/blog/game-theory-attribution/

А еще

- Обычно мы с помощью Shapley Values оцениваем вклад каждого признака в предсказание, для объяснения решений модели
- Можно строить Shapley Values-объяснения для ошибок. Например,
  Explaining the Loss of a Tree Model и
  Use SHAP loss values to debug your model

Вопросы

Слайды тут

dkolodezev

promsoft

dkolodezev

d_key

dmitry_kolodezev

https://kolodezev.ru/download/slides-dataconf-2021.pdf