

Гиперпараметры

ДАТАСЕТ: РАК МОЛОЧНОЙ ЖЕЛЕЗЫ

Ткаченко Елизавета

Параметры RandomForest

Параметры Random Forest

Параметр	Влияние	Описание / Зачем нужен	Тип / Пример
n_estimators	недобучение / переобучение	Количество деревьев. Мало — недобучение, много — лучше качество, но медленнее.	int — 100, 500, 1000
criterion	качество разделения	Функция для измерения качества разделения. "gini" для классификации.	"gini", "entropy"
max_depth	сложность модели	Максимальная глубина дерева. Больше — сложнее модель, риск переобучения.	int — 5, 10, None
min_samples_split	регуляризация	Минимальное количество samples для разделения узла. Больше — более обобщенная модель.	int — 2, 5; float — 0.1
min_samples_leaf	регуляризация	Минимальное количество samples в листе. Больше — более обобщенная модель.	int — 1, 5; float — 0.1
max_features	разнообразие деревьев	Количество features для поиска лучшего разделения. Меньше — более случайные деревья.	"auto", "sqrt", "log2", 0.7
bootstrap	стабилизация	Использование bootstrap samples для построения деревьев.	bool — True, False
oob_score	оценка обобщения	Использование out-of-bag samples для оценки обобщающей способности.	bool — True, False
class_weight	дисбаланс классов	Веса для классов. "balanced" автоматически подстраивает веса по частотам классов.	"balanced", None, {0:1, 1:2}
random_state	воспроизводимость	Фиксирует случайность для воспроизводимости результатов.	int — 42, 123

Сравнения методов подбора параметров

Grid Search

```
Лучшие параметры:  
classifier__criterion -> gini  
classifier__max_depth -> None  
classifier__max_features -> sqrt  
classifier__min_samples_leaf -> 1  
classifier__min_samples_split -> 2
```

Лучшее качество на кросс-валидации: 0.9626373626373628
Точность на тесте: 0.9736842105263158

Random Search

```
Лучшие параметры:  
classifier__criterion -> gini  
classifier__max_depth -> 12  
classifier__max_features -> sqrt  
classifier__min_samples_leaf -> 2  
classifier__min_samples_split -> 4
```

Лучшее качество на кросс-валидации: 0.9626373626373628
Точность на тесте: 0.9736842105263158

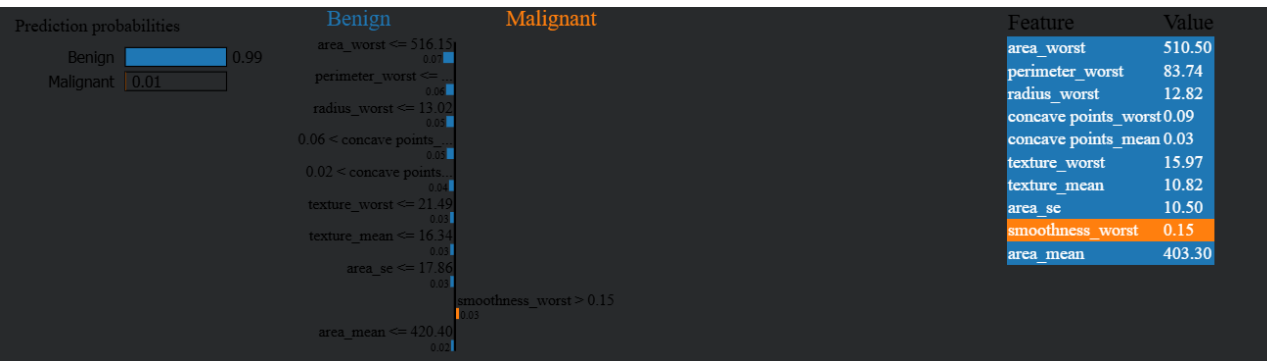
Optuna

```
Лучшие параметры Random Forest:  
classifier__n_estimators -> 592  
classifier__max_depth -> 9  
classifier__min_samples_split -> 14  
classifier__min_samples_leaf -> 1  
classifier__max_features -> sqrt  
classifier__criterion -> gini
```

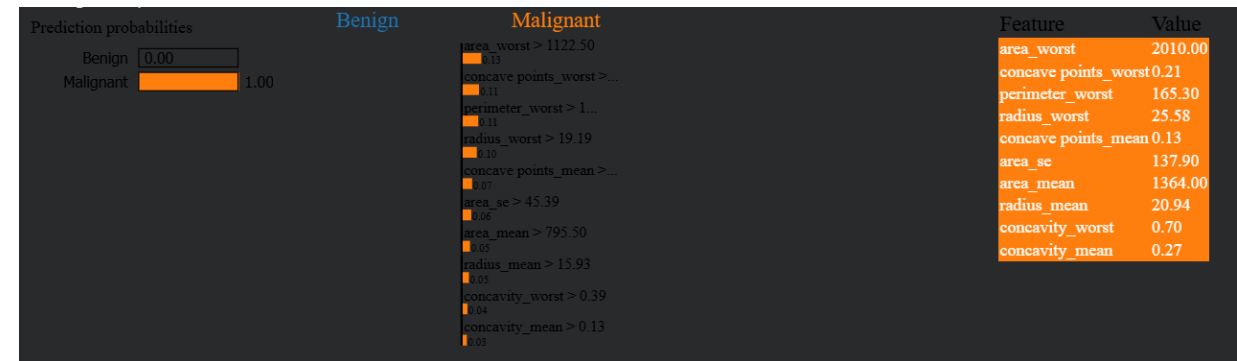
Лучшее качество на кросс-валидации: 0.9582417582417584
Точность Random Forest на тесте: 0.9649122807017544

LIME

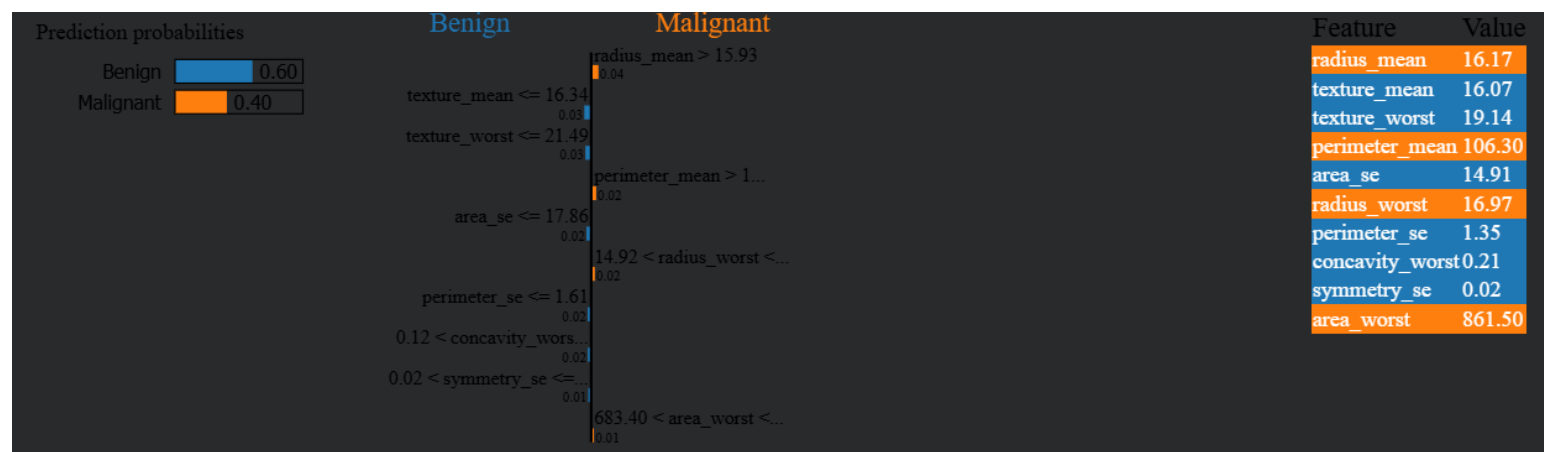
Экземпляр 1



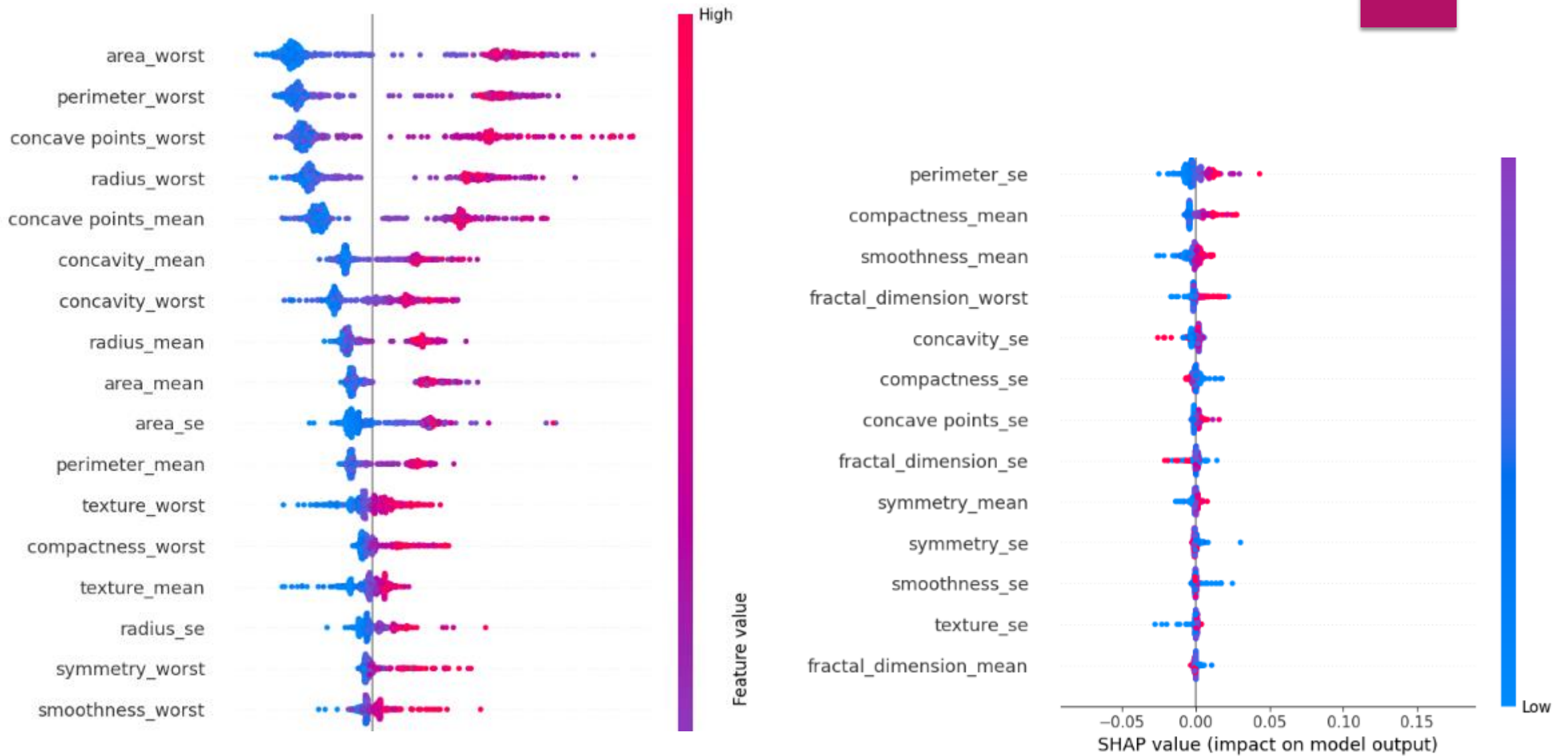
Экземпляр 2



Экземпляр 3



SHAP глобальное



SHAP глобальное

