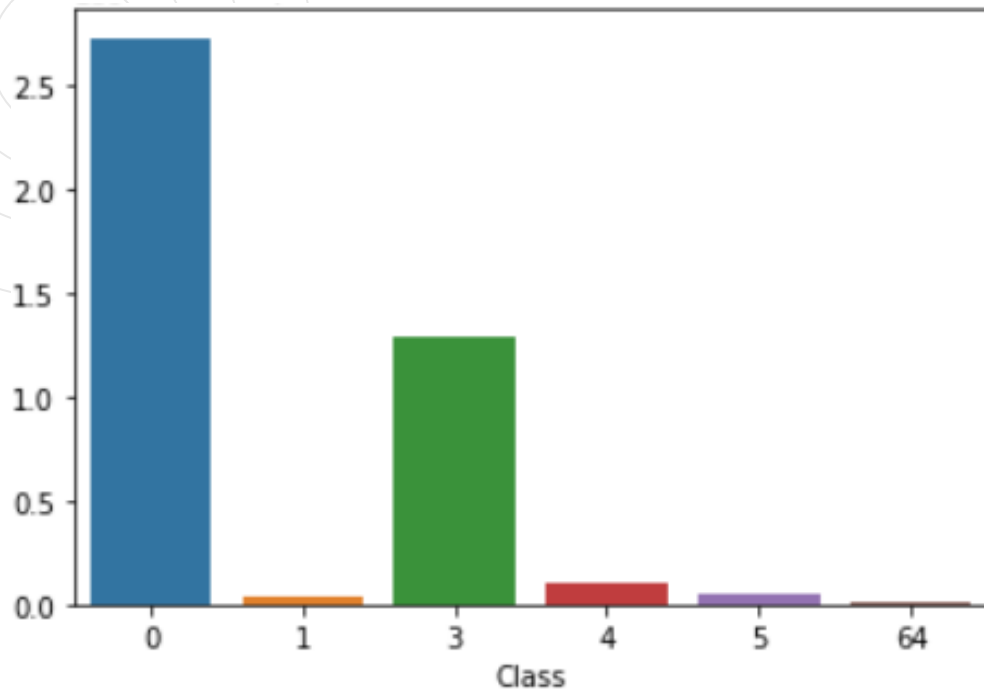


Классификация объектов точек Для железнодорожного транспорта

Хафизова Дарья

6 классов точек лазерных отражений



0 – земля

1 – опоры

3 – растительность

4 – точки рельсов

5 – контактная сеть

64 – шумы

Шумы – самый **редкий класс**



1

SMOTE

для дисбаланса
классов

SMOTE

Synthetic Minority Oversampling Technique

```
weights_smote = ml_data[target].value_counts().to_dict()  
weights_smote
```

Исходный баланс классов в обучающей выборке

{0: 2731040, 3: 1287816, 4: 103453, 5: 55985, 1: 39384, 64: 4416}

Целевой баланс классов для SMOTE

{0: 2731040, 3: 1287816, 4: 103453, 5: 103453, 1: 103453, 64: 103453}

Автоматически **синтезировали** новые точки
для **`minority`** классов :

5, 1, 64 - контактная сеть, опоры, шумы

#Imblearn
библиотека



2

Новые

Параметры точек

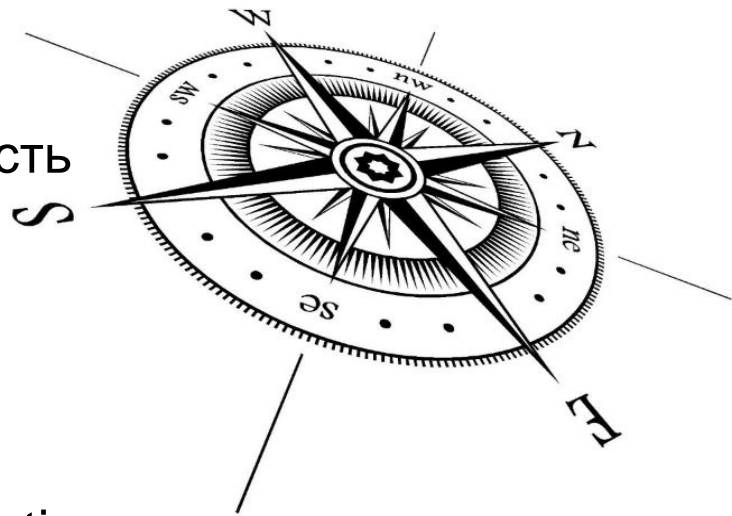
Параметры точек лазерных отражений

1. *Northing* – направление на север
2. *Easting* – направление на восток
3. *Height* – высота (м)
4. *Reflectance* – отражательная способность

Добавлены **2 признака** направлений:

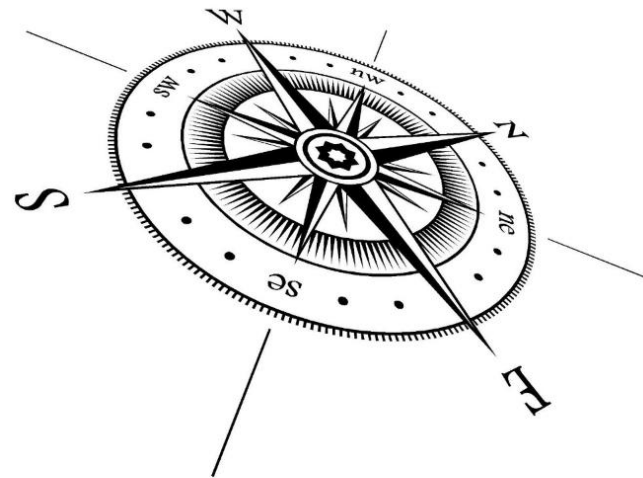
combine_north_east = Northing + Easting

height_north_east = Height + Northing + Easting



Важность признаков для предсказаний

	importance
Height	0.382397
combine_north_east	0.207798
Reflectance	0.175139
height_north_east	0.173542
Easting	0.031708
Northing	0.029416



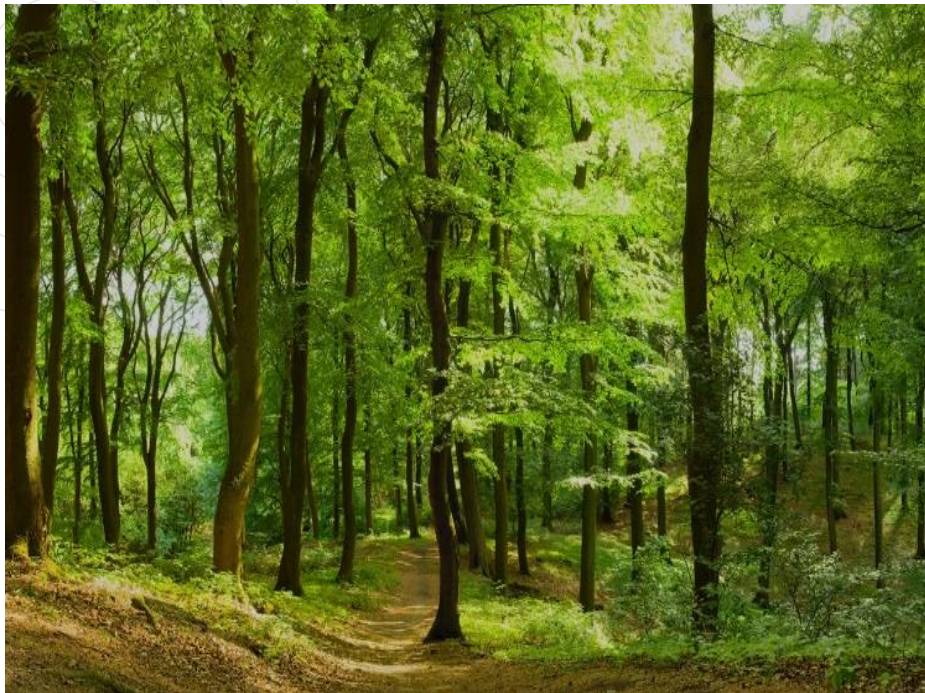


3

Random Forest

финальная модель

RandomForest для предсказаний




```
model = BalancedRandomForestClassifier(  
    n_estimators=100,  
    bootstrap=False,  
    max_features=3)
```

```
clf = imblearn.pipeline.Pipeline([  
    ('SMOTE', SMOTE(random_state=12,  
                    sampling_strategy=weights_smote,  
                    k_neighbors=13)),  
    ('model', model)  
])
```

BalancedRF – использует **UnderSampling**

Пример результатов на **валидации**

recall




0	0.97
1	0.99
3	0.97
4	1.00
5	0.99
64	0.89

Выборка для обучения ~ 2 000 000 точек

Лучше всего модель определяет **3 класса**:

- (1) Точки опор
- (4) Точки рельсов
- (5) Точки контактной сети

* Полный датасет – 4 222 094 точек

The background is black. A large white circle is centered on the slide. To the left of the white circle, there are two overlapping circles: a dark gray one in front of a lighter gray one. To the right of the white circle, there are several concentric white circles of varying sizes.

**Спасибо
за внимание!**