Защищено: Гапанюк Ю.Е.

" " 2022 г.

Демонстрация: Гапанюк Ю.Е.

" " 2022 г.

**Отчет по рубежному контролю № 2 по курсу Технологии машинного обучения**

**ГУИМЦ**

**Тема работы: " Методы построения моделей машинного обучения. "**

16

(количество листов) Вариант № **28**

ИСПОЛНИТЕЛЬ:

студент группы ИУ5Ц-82Б

(подпись)

Дмитриева М.Ю.

" " 2023 г.

Москва, МГТУ - 2022

# Цель лабораторной работы

Изучение линейных моделей, SVM и деревьев решений

# Задание

##### Выберите набор данных (датасет) для решения задачи классификации или регрессии.

1. В случае необходимости проведите удаление или заполнение пропусков и кодирование категориальных признаков.
2. С использованием метода train\_test\_split разделите выборку на обучающую и тестовую.

##### Обучите следующие модели:

одну из линейных моделей (линейную или полиномиальную регрессию при решении задачи регрессии, логистическую регрессию при решении задачи классификации);

##### SVM;

дерево решений.

##### Оцените качество моделей с помощью двух подходящих для задачи метрик. Сравните качество полученных моделей.

1. Постройте график, показывающий важность признаков в дереве решений.
2. Визуализируйте дерево решений или выведите правила дерева решений в текстовом виде.

# Ход лабораторной работы

## Текстовое описание набора данных

В качестве набора данных используется dataset рейтингов университетов мира на основании трёх рейтингов. Датасет доступен по адресу: [https://www.kaggle.com/mylesoneill/world-university-ranking](https://www.kaggle.com/mylesoneill/world-university-rankings)s

Из набора данных будет рассматриваться только файл cwurData.csv . Описание столбцов:

world\_rank - мировой рейтинг университета

institution - название университета

country - страна, в которой расположен университет national\_rank - рейтинг университета в стране его нахождения quality\_of\_education - рейтинг качества образования

quality\_of\_faculty - рейтинг качества профессорско-преподавательского состава

publications - рейтинг публикаций

infuence - рейтинг влияния

citations - количество студентов в университете

##### broad\_impact - рейтинг за широкое влияние (предоставлен только за 2014 и 2015 гг. Остальное - пропуски)

patents - рейтинг за патенты

##### score - общий балл, используемый для определения мирового рейтинга

year - год рейтинга (с 2012 по 2015 год)

## Основные характеристики набора данных

Подключаем все необходимые библиотеки In [1]: **import** numpy **as** np

**import** pandas **as** pd

**import** seaborn **as** sns **import** matplotlib **import** matplotlib\_inline

**import** matplotlib.pyplot **as** plt

**from** sklearn.model\_selection **import** train\_test\_split

**%matplotlib** inline sns**.**set(style**=**"ticks") **from** io **import** StringIO

##### Подключаем Dataset

In [2]: data **=** pd**.**read\_csv('cwurData.csv', sep**=**",")

Размер набора данных In [3]: data**.**shape

Out[3]:(2200, 14)

##### Типы колонок

In [4]: data**.**dtypes

Out[4]:world\_rank int64

institution object

country object

national\_rank int64 quality\_of\_education int64 alumni\_employment int64

quality\_of\_faculty int64

publications int64

influence int64

citations int64

broad\_impact float64

patents int64

score float64

year int64

dtype: object

##### Проверяем, есть ли пропущенные значения

In [5]: data**.**isnull()**.**sum()

|  |  |
| --- | --- |
| Out[5]:world\_rank | 0 |
| institution | 0 |
| country | 0 |
| national\_rank | 0 |
| quality\_of\_education | 0 |
| alumni\_employment | 0 |
| quality\_of\_faculty | 0 |
| publications | 0 |
| influence | 0 |
| citations | 0 |
| broad\_impact | 200 |
| patents | 0 |
| score | 0 |
| year | 0 |
| dtype: int64 |  |

Первые 5 строк датасета In [6]: data**.**head()

Out[6]: **world\_rank institution**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **country** | **national\_rank** | **quality\_of\_education** | **alumni\_employment** | **quality\_of\_faculty** | **publications** | **influence** |
| USA | 1 | 7 | 9 | 1 | 1 | 1 |
| USA | 2 | 9 | 17 | 3 | 12 | 4 |
| USA | 3 | 17 | 11 | 5 | 4 | 2 |

* 1. 1 Harvard University

Massachusetts

* 1. 2 Institute of Technology
  2. 3 Stanford University
  3. 4 University of Cambridge

California

United Kingdom

1 10 24 4 16 16

* 1. 5 Institute of Technology

USA 4 2 29 7 37 22



In [7]: total\_count **=** data**.**shape[0]

print('Всего строк: {}'**.**format(total\_count))

Всего строк: 2200

Процент пропусков в broad\_impact

In [8]: (200 **/** 2200) **\*** 100

Out[8]:9.090909090909092

##### Настройка отображения графиков

In [9]: *# Задание формата графиков для сохранения высокого качества PNG* **from** IPython.display **import** set\_matplotlib\_formats matplotlib\_inline**.**backend\_inline**.**set\_matplotlib\_formats("retina") *# Задание ширины графиков, чтобы они помещались на A4* pd**.**set\_option("display.width", 70)

## Обработка пропусков данных

### Очистка строк

##### Можно очистить строки, содержащие пропуски. При этом останутся данные только за 2014 и 2015 гг (см. описание датасета)

In [10]: *# Удаление строк, содержащих пустые значения* data\_no\_null **=** data**.**dropna(axis**=**0, how**=**'any') (data**.**shape, data\_no\_null**.**shape)

Out[10]:((2200, 14), (2000, 14))

##### Выведем первые 11 строк, чтобы убедиться, что данные в national\_rank числовые (Jupyter Lab в предпросмотре CSV показывает не совсем верно)

In [11]: data\_no\_null**.**head(11)

**institution**

|  |  |  |
| --- | --- | --- |
| Out[11]: | **200** | **world\_rank**  1 |
|  | **201** | 2 |
|  | **202** | 3 |
|  | **203** | 4 |
|  | **204** | 5 |
|  | **205** | 6 |
|  | **206** | 7 |
|  | **207** | 8 |
|  | **208** | 9 |
|  | **209** | 10 |
|  | **210** | 11 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **country**  USA | **national\_rank**  1 | **quality\_of\_education**  1 | **alumni\_employment**  1 | **quality\_of\_faculty**  1 | **publications**  1 | **influenc** |
| USA | 2 | 11 | 2 | 4 | 5 |  |
| USA | 3 | 3 | 11 | 2 | 15 |  |
| United Kingdom | 1 | 2 | 10 | 5 | 10 |  |
| United | 2 | 7 | 12 | 10 | 11 |  |

Harvard University

Stanford University

Massachusetts Institute of Technology

University of Cambridge

University of

Oxford

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Kingdom  USA | 4 | 13 | 8 | 9 | 14 |
| USA | 5 | 4 | 22 | 6 | 7 |
| USA | 6 | 10 | 14 | 8 | 17 |
| USA | 7 | 5 | 16 | 3 | 70 |
| USA | 8 | 9 | 25 | 11 | 18 |
| USA | 9 | 12 | 18 | 19 | 23 |

Columbia University

University of California, Berkeley

University of

Chicago

Princeton University

Yale University

Cornell University



In [12]: total\_count **=** data\_no\_null**.**shape[0]

print('Всего строк: {}'**.**format(total\_count))

Всего строк: 2000

## Кодирование категориальных признаков

##### Преобразуем названия стран, городов, ... в числовые зеачения (label encoding)

In [13]: **from** sklearn.preprocessing **import** LabelEncoder, OneHotEncoder

In [14]: le **=** LabelEncoder()

*# "institution"* le**.**fit(data\_no\_null**.**institution**.**drop\_duplicates()) data\_no\_null**.**institution **=** le**.**transform(data\_no\_null**.**institution)

*# "country"* le**.**fit(data\_no\_null["country"]**.**drop\_duplicates()) data\_no\_null["country"] **=** le**.**transform(data\_no\_null["country"])

/tmp/ipykernel\_143/4210865855.py:4: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame. Try using .loc[row\_indexer,col\_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#re turning-a-view-versus-a-copy

data\_no\_null.institution = le.transform(data\_no\_null.institution)

/tmp/ipykernel\_143/4210865855.py:7: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame. Try using .loc[row\_indexer,col\_indexer] = value instead

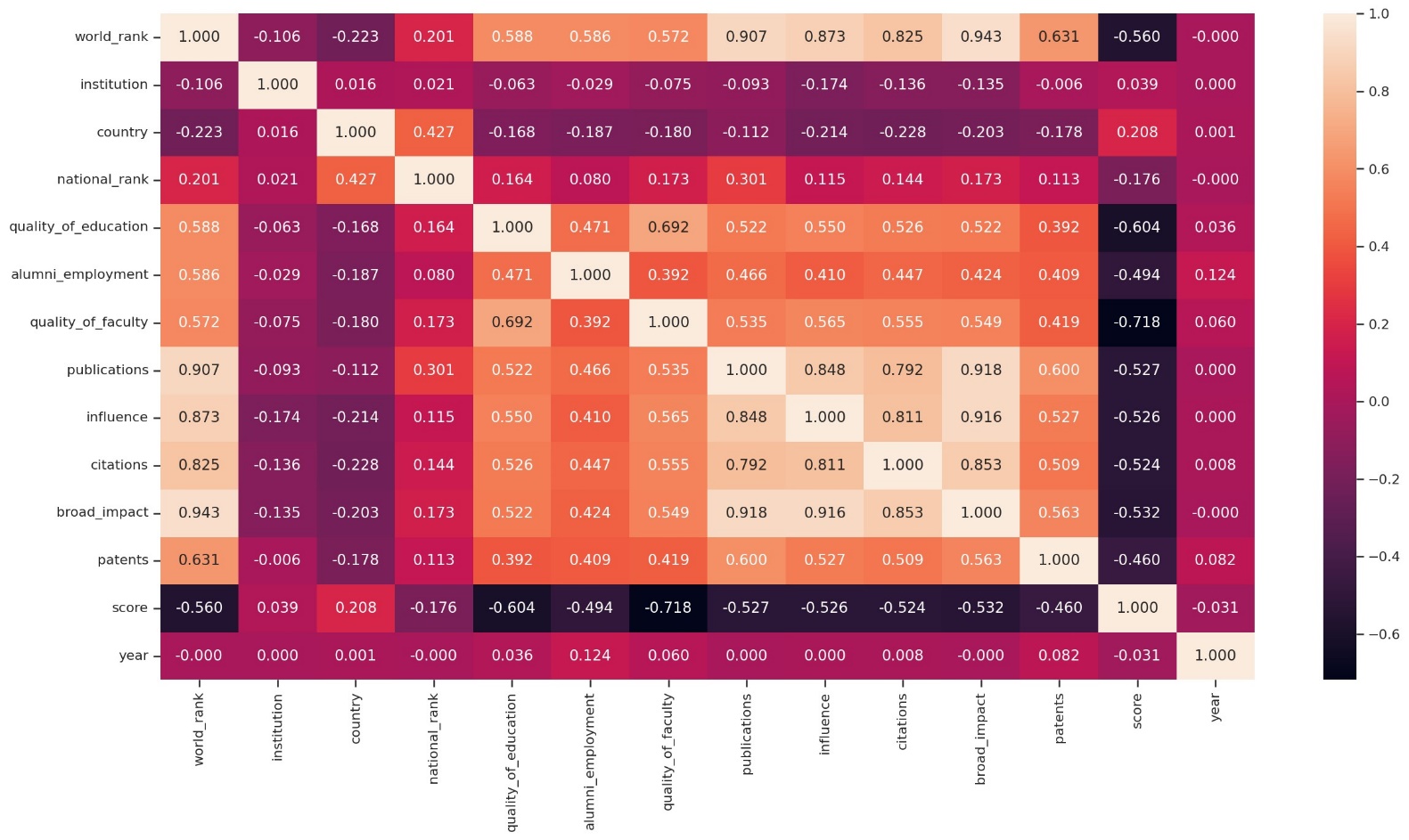
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#re turning-a-view-versus-a-copy

data\_no\_null["country"] = le.transform(data\_no\_null["country"])

##### Построим кореляционную матрицу

In [15]: ig, ax **=** plt**.**subplots(figsize**=**(20,10)) sns**.**heatmap(data\_no\_null**.**corr(method**=**'pearson'), ax**=**ax, annot**=True**, fmt**=**'.3f')

Out[15]:<AxesSubplot:>



|  |
| --- |
|  |
|  |
|  |

## Предсказание целевого признака

Предскажем значение целевого признака world\_rank по broad\_impact и publications , поскольку их значения кореляции ближе всего к 1

### Разбиение выборки на обучающую и тестовую

In [16]: X **=** data\_no\_null[["broad\_impact", "publications"]] Y **=** data\_no\_null["world\_rank"]

print('Входные данные:\n\n', X**.**head(), '\n\nВыходные данные:\n\n', Y**.**head())

Входные данные:

broad\_impact publications

|  |  |  |
| --- | --- | --- |
| 200 | 1.0 | 1 |
| 201 | 4.0 | 5 |
| 202 | 2.0 | 15 |
| 203 | 13.0 | 10 |
| 204 | 12.0 | 11 |

Выходные данные:

200 1

201 2

202 3

203 4

204 5

Name: world\_rank, dtype: int64

##### Разделим выборку на обучающую и тестовую

In [17]: X\_train, X\_test, Y\_train, Y\_test **=** train\_test\_split(X, Y, random\_state **=** 2022, test\_size **=** 0.1)

Входные параметры обучающей выборки In [18]: X\_train**.**head()

|  |  |  |  |
| --- | --- | --- | --- |
| Out[18]: |  | **broad\_impact** | **publications** |
|  | **2164** | 932.0 | 875 |
|  | **1710** | 590.0 | 576 |
|  | **428** | 164.0 | 200 |
|  | **1389** | 164.0 | 233 |
|  | **2089** | 932.0 | 675 |

Входные параметры тестовой выборки In [19]: X\_test**.**head()

|  |  |  |  |
| --- | --- | --- | --- |
| Out[19]: |  | **broad\_impact** | **publications** |
|  | **1218** | 14.0 | 3 |
|  | **1495** | 265.0 | 236 |
|  | **843** | 703.0 | 943 |
|  | **2042** | 850.0 | 803 |
|  | **1869** | 606.0 | 701 |

Выходные параметры обучающей выборки In [20]: Y\_train**.**head()

Out[20]:2164 965

|  |  |
| --- | --- |
| 1710 | 511 |
| 428 | 229 |
| 1389 | 190 |
| 2089 | 890 |

Name: world\_rank, dtype: int64

##### Выходные параметры тестовой выборки

In [21]: Y\_test**.**head()

Out[21]:1218 19

|  |  |
| --- | --- |
| 1495 | 296 |
| 843 | 644 |
| 2042 | 843 |
| 1869 | 670 |

Name: world\_rank, dtype: int64

### Построение линейной регрессии

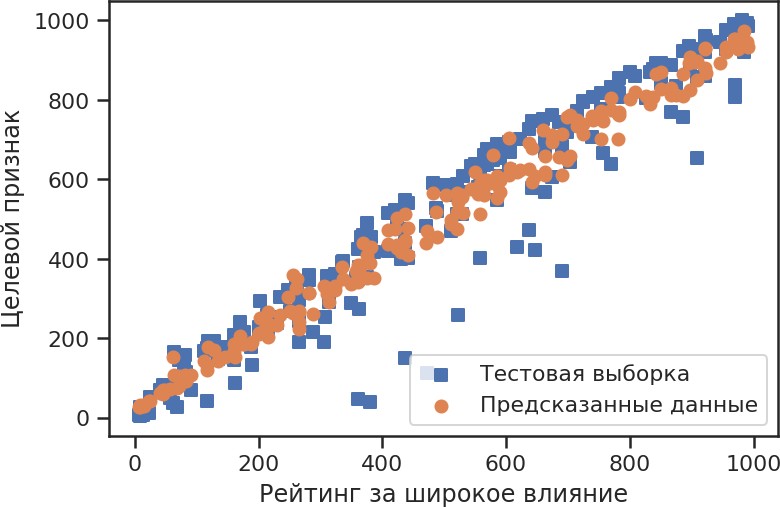
In [22]: **from** sklearn.linear\_model **import** LinearRegression

**from** sklearn.metrics **import** mean\_absolute\_error, mean\_squared\_error, median\_absolute\_error, r2\_score

In [23]: Lin\_Reg **=** LinearRegression()**.**fit(X\_train, Y\_train) lr\_y\_pred **=** Lin\_Reg**.**predict(X\_test)

##### Возьмем тот параметр, чья корреляция ближе всего к единице, т.е. broad\_impact

In [40]: plt**.**scatter(X\_test["broad\_impact"], Y\_test, marker **=** 's', label **=** 'Тестовая выборка') plt**.**scatter(X\_test["broad\_impact"], lr\_y\_pred, marker **=** 'o', label **=** 'Предсказанные данные') plt**.**legend (loc **=** 'lower right')

plt**.**xlabel ('Рейтинг за широкое влияние') plt**.**ylabel ('Целевой признак') plt**.**show()

|  |
| --- |
|  |
|  |
|  |

In [25]: **from** sklearn.metrics **import** mean\_absolute\_error, mean\_squared\_error, median\_absolute\_error, r2\_score

In [26]: print('Средняя абсолютная ошибка:', mean\_absolute\_error(Y\_test, lr\_y\_pred)) print('Средняя квадратичная ошибка:', mean\_squared\_error(Y\_test, lr\_y\_pred)) print('Median absolute error:', median\_absolute\_error(Y\_test, lr\_y\_pred)) print('Коэффициент детерминации:', r2\_score(Y\_test, lr\_y\_pred))

Средняя абсолютная ошибка: 54.70203008487861 Средняя квадратичная ошибка: 6228.270901286782 Median absolute error: 42.05433711920929 Коэффициент детерминации: 0.9234584275958889

### SVM

In [27]: **from** sklearn.svm **import** SVC , LinearSVC

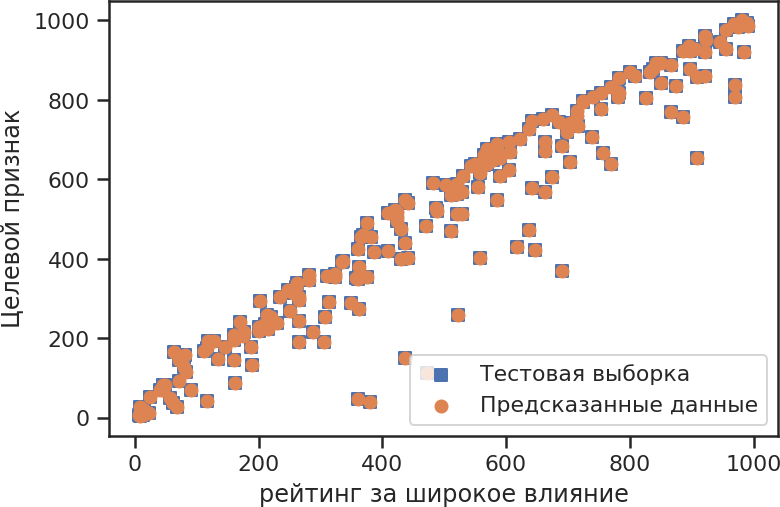
**from** sklearn.datasets **import** make\_blobs

In [28]: svc **=** SVC(kernel**=**'linear') svc**.**fit(X\_train,Y\_train)

Out[28]:SVC(kernel='linear')

In [29]: pred\_y **=** svc**.**predict(X\_test)

In [41]: plt**.**scatter(X\_test["broad\_impact"], Y\_test, marker **=** 's', label **=** 'Тестовая выборка') plt**.**scatter(X\_test["broad\_impact"], pred\_y, marker **=** 'o', label **=** 'Предсказанные данные') plt**.**legend (loc **=** 'lower right')

plt**.**xlabel ('рейтинг за широкое влияние') plt**.**ylabel ('Целевой признак') plt**.**show()

|  |
| --- |
|  |
|  |
|  |

In [31]: print('Средняя абсолютная ошибка:', mean\_absolute\_error(Y\_test, pred\_y)) print('Средняя квадратичная ошибка:', mean\_squared\_error(Y\_test, pred\_y)) print('Median absolute error:', median\_absolute\_error(Y\_test, pred\_y)) print('Коэффициент детерминации:', r2\_score(Y\_test, pred\_y))

Средняя абсолютная ошибка: 57.19 Средняя квадратичная ошибка: 9379.4 Median absolute error: 27.0

Коэффициент детерминации: 0.8847330124868531

### Дерево (Tree)

In [32]: **from** sklearn.tree **import** DecisionTreeClassifier, DecisionTreeRegressor, export\_graphviz

**from** sklearn.tree **import** export\_graphviz

**from** sklearn **import** tree

**import** re

**from** IPython.core.display **import** HTML **from** sklearn.tree **import** export\_text **import** graphviz

**from** IPython.display **import** Image

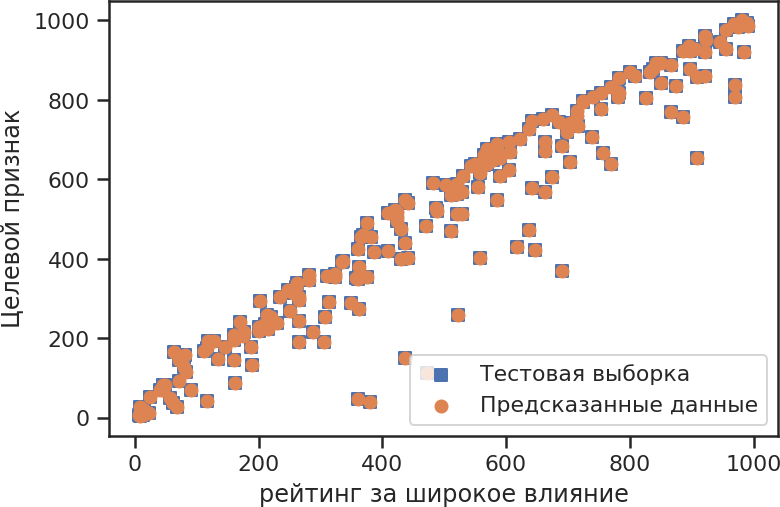
**import** pydotplus

##### Обучим дерево на всех признаках

In [33]: reg **=** tree**.**DecisionTreeRegressor() reg **=** reg**.**fit(X\_test, Y\_test)

In [42]: pred\_y **=** reg**.**predict(X\_test)

plt**.**scatter(X\_test["broad\_impact"], Y\_test, marker **=** 's', label **=** 'Тестовая выборка') plt**.**scatter(X\_test["broad\_impact"], pred\_y, marker **=** 'o', label **=** 'Предсказанные данные') plt**.**legend (loc **=** 'lower right')

plt**.**xlabel ('рейтинг за широкое влияние') plt**.**ylabel ('Целевой признак') plt**.**show()

|  |
| --- |
|  |
|  |
|  |

**Дерево в текстовом виде**

In [35]: tree\_rules **=** export\_text(reg, feature\_names**=**list(X**.**columns)) HTML('<pre>' **+** tree\_rules **+** '</pre>')

Out[35]:|--- broad\_impact <= 477.00

| |--- broad\_impact <= 201.50

| | |--- publications <= 105.00

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| | | | |--- publications <= 24.50 | | | | | | | |
| | | | | | | | | |--- publications <= 23.50 | | | |
| | | | | | | | | | |--- publications <= 5.00 | | | |
| | | | | | | | | | | |--- value: [19.00] | | | |
| | | | | | | | | | |--- publications > 5.00 | | | |
| | | | | | | | | | | |--- broad\_impact <= 18.00 | | | |
| | | | | | | | | | | | |--- broad\_impact <= 8.00 | | | |
| |  | | |  | | |  | | |  | | |  | | |  | | |  | | | |--- value: [7.00]  |--- broad\_impact > 8.00 |
| | | | | | | | | | | | | | | | |--- publications <= 10.50 |
| | | | | | | | | | | | | | | | | |--- value: [5.00] |
| | | | | | | | | | | | | | | | |--- publications > 10.50 |
| | | | | | | | | | | | | | | | | |--- value: [6.00] |
| | | | | | | | | | | | | |--- broad\_impact > 18.00 | |
| | | | | | | | | | | | | | |--- value: [11.00] | |
| | | | | | | | | |--- publications > 23.50 | | | |
| | | | | | | | | | |--- value: [27.00] | | | |
| | | | | | | |--- publications > 24.50 | | | | |
| | | | | | | | |--- broad\_impact <= 69.50 | | | | |
| | | | | | | | | |--- broad\_impact <= 52.00 | | | | |
| | | | | | | | | | |--- publications <= 30.00 | | | | |
| | | | | | | | | | | | | | | |--- value: [53.00] |
| | | | | | | | | | | | | |--- publications > 30.00 | |
| | | | | | | | | | | | | | |--- broad\_impact <= 43.50 | |
| | | | | | | | | | | | | | | |--- value: [69.00] | |
| | | | | | | | | | | | | | |--- broad\_impact > 43.50 | |
| | | | | | | | | | | | | | | |--- publications <= 65.50 | |

| |--- value: [82.00]

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |

|--- publications > 65.50

| |--- value: [70.00]

| | | | | |--- broad\_impact > 52.00

|--- publications <= 46.50

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |

| |--- value: [28.00]

|--- publications > 46.50

| |--- value: [51.00]

|--- broad\_impact > 69.50

|  |  |  |  |
| --- | --- | --- | --- |
| | | | | | | | |
| | | | | | | | |

| |--- publications <= 73.00

|--- broad\_impact <= 75.00

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |

| |--- value: [93.00]

|--- broad\_impact > 75.00

| |--- publications <= 66.00

| | |--- value: [116.00]

| |--- publications > 66.00

| | |--- value: [127.00]

|--- publications > 73.00

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| | | | | | | | | | |
| | | | | | | | | | |

| |--- broad\_impact <= 139.00

|--- broad\_impact <= 103.50

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |

| |--- value: [70.00]

|--- broad\_impact > 103.50

| |--- value: [43.00]

| | | | | | |--- broad\_impact > 139.00

|--- publications <= 82.50

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |

| |--- value: [87.00]

|--- publications > 82.50

| |--- value: [145.00]

| | |--- publications > 105.00

|--- broad\_impact <= 63.00

|  |  |  |
| --- | --- | --- |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |

| |--- value: [37.00]

|--- broad\_impact > 63.00

| |--- publications <= 155.50

|--- publications <= 138.50

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| | | | | | | | | | |
| | | | | | | | | | |

| |--- publications <= 120.50

|--- publications <= 115.00

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |

| |--- value: [147.00]

|--- publications > 115.00

| |--- value: [157.00]

|--- publications > 120.50

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | |
| | | | | | | | | | | | |

| |--- publications <= 133.00

|--- broad\_impact <= 130.50

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |

| |--- value: [175.00]

|--- broad\_impact > 130.50

| |--- publications <= 127.00

| | |--- value: [178.00]

| |--- publications > 127.00

| | |--- value: [179.00]

|--- publications > 133.00

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |

| |--- value: [208.00]

| | | | | |--- publications > 138.50

|--- publications <= 145.00

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |

| |--- value: [146.00]

|--- publications > 145.00

| |--- value: [133.00]

| | | | |--- publications > 155.50

|--- broad\_impact <= 165.50

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| | | | | | | | | | |
| | | | | | | | | | |

| |--- publications <= 190.50

|--- publications <= 173.00

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |

| |--- value: [165.00]

|--- publications > 173.00

| |--- value: [168.00]

| | | | | | |--- publications > 190.50

|--- publications <= 220.50

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |

| |--- value: [197.00]

|--- publications > 220.50

| |--- publications <= 262.50

| | |--- value: [194.00]

| |--- publications > 262.50

| | |--- value: [193.00]

|--- broad\_impact > 165.50

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| | | | | | | | | | |
| | | | | | | | | | |

| |--- publications <= 202.50

|--- broad\_impact <= 175.00

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |

| |--- value: [205.00]

|--- broad\_impact > 175.00

| |--- broad\_impact <= 188.00

| | |--- value: [216.00]

|--- broad\_impact > 188.00

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |

| |--- value: [219.00]

| | | | | | |--- publications > 202.50

|--- broad\_impact <= 185.00

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |

| |--- value: [240.00]

|--- broad\_impact > 185.00

| |--- value: [229.00]

| |--- broad\_impact > 201.50

| | |--- broad\_impact <= 380.50

| | | |--- publications <= 399.00

| | | | |--- publications <= 382.00

|--- publications <= 277.00

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| | | | | | | | | | |
| | | | | | | | | | |

| |--- broad\_impact <= 368.00

|--- broad\_impact <= 355.00

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |

| |--- publications <= 197.50

|--- publications <= 100.00

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |

| |--- value: [191.00]

|--- publications > 100.00

| |--- broad\_impact <= 276.50

| | |--- truncated branch of depth 3

| |--- broad\_impact > 276.50

| | |--- value: [217.00]

|--- publications > 197.50

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |

| |--- broad\_impact <= 253.50

|--- broad\_impact <= 212.00

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | |

| |--- value: [237.00]

|--- broad\_impact > 212.00

| |--- truncated branch of depth 3

| | | | | | | | | |--- broad\_impact > 253.50

|--- broad\_impact <= 260.50

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | |

| |--- value: [313.00]

|--- broad\_impact > 260.50

| |--- truncated branch of depth 3

| | |--- broad\_impact > 355.00

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |

| | | |--- value: [48.00]

| |--- broad\_impact > 368.00

| | |--- value: [354.00]

|--- publications > 277.00

| |--- broad\_impact <= 309.50

| | |--- broad\_impact <= 293.50

| | | |--- broad\_impact <= 257.50

|--- publications <= 350.50

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |

| |--- broad\_impact <= 218.50

| | |--- value: [293.00]

| |--- broad\_impact > 218.50

| | |--- value: [303.00]

|--- publications > 350.50

| |--- value: [234.00]

| | | | | | | | |--- broad\_impact > 257.50

|--- publications <= 368.50

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |

| |--- value: [359.00]

|--- publications > 368.50

| |--- value: [346.00]

| | | | | | | |--- broad\_impact > 293.50

|--- publications <= 363.50

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |

| |--- value: [253.00]

|--- publications > 363.50

| |--- value: [190.00]

| | | | | | |--- broad\_impact > 309.50

|--- publications <= 364.50

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |

| |--- publications <= 328.50

|--- broad\_impact <= 342.50

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |

| |--- publications <= 303.50

| | |--- truncated branch of depth 2

| |--- publications > 303.50

| | |--- value: [355.00]

|--- broad\_impact > 342.50

| |--- value: [349.00]

| | | | | | | | |--- publications > 328.50

|--- publications <= 346.50

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |

| |--- value: [380.00]

|--- publications > 346.50

| |--- value: [395.00]

| | | | | | | |--- publications > 364.50

|--- publications <= 371.50

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |

| |--- value: [275.00]

|--- publications > 371.50

| | | | | |--- value: [353.00]

|  |  |  |  |
| --- | --- | --- | --- |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |

|--- publications > 382.00

| |--- value: [40.00]

| | | |--- publications > 399.00

| | | | |--- broad\_impact <= 348.00

|--- broad\_impact <= 298.50

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| | | | | | | | | | |
| | | | | | | | | | |

| |--- broad\_impact <= 258.00

|--- broad\_impact <= 251.50

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |

| |--- value: [321.00]

|--- broad\_impact > 251.50

| |--- value: [317.00]

| | | | | | |--- broad\_impact > 258.00

|--- publications <= 516.00

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |

| |--- value: [329.00]

|--- publications > 516.00

| |--- value: [340.00]

| |--- broad\_impact > 298.50

|  |  |  |  |
| --- | --- | --- | --- |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |

| | |--- value: [392.00]

|--- broad\_impact > 348.00

| |--- broad\_impact <= 363.50

| | |--- value: [425.00]

| |--- broad\_impact > 363.50

|--- broad\_impact <= 372.00

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |

| |--- publications <= 509.50

| | |--- value: [456.00]

| |--- publications > 509.50

| | |--- value: [461.00]

|--- broad\_impact > 372.00

| |--- value: [491.00]

| | |--- broad\_impact > 380.50

| | | |--- broad\_impact <= 471.50

| | | | |--- publications <= 551.00

|--- publications <= 455.50

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| | | | | | | | | | |
| | | | | | | | | | |

| |--- publications <= 321.00

|--- publications <= 265.00

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |

| |--- value: [418.00]

|--- publications > 265.00

| |--- value: [402.00]

|--- publications > 321.00

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | |
| | | | | | | | | | | | |

| |--- broad\_impact <= 427.00

|--- publications <= 436.00

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |

| |--- value: [495.00]

|--- publications > 436.00

| |--- value: [505.00]

|--- broad\_impact > 427.00

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |

| |--- publications <= 386.00

|--- publications <= 351.50

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |

| |--- value: [484.00]

|--- publications > 351.50

| |--- value: [474.00]

| | | | | | | | |--- publications > 386.00

|--- broad\_impact <= 433.50

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |

| |--- value: [400.00]

|--- broad\_impact > 433.50

| |--- value: [441.00]

| | | | | |--- publications > 455.50

|--- publications <= 477.50

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |

| |--- value: [150.00]

|--- publications > 477.50

| |--- publications <= 518.50

| | |--- value: [419.00]

| |--- publications > 518.50

| | |--- value: [454.00]

|--- publications > 551.00

|  |  |  |  |
| --- | --- | --- | --- |
| | | | | | | | |
| | | | | | | | |

| |--- broad\_impact <= 430.50

|--- publications <= 621.00

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |

| |--- value: [524.00]

|--- publications > 621.00

| |--- broad\_impact <= 416.50

| | |--- value: [516.00]

| |--- broad\_impact > 416.50

| | |--- value: [519.00]

| | | | | |--- broad\_impact > 430.50

|--- broad\_impact <= 439.50

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |

| |--- value: [547.00]

|--- broad\_impact > 439.50

| |--- value: [541.00]

| | | |--- broad\_impact > 471.50

| | | | |--- value: [114.00]

|--- broad\_impact > 477.00

| |--- broad\_impact <= 709.00

| | |--- publications <= 626.50

| | | |--- broad\_impact <= 694.50

|--- publications <= 592.50

|  |  |  |  |
| --- | --- | --- | --- |
| | | | | | | | |
| | | | | | | | |

| |--- publications <= 458.00

|--- broad\_impact <= 539.00

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | |
| | | | | | | | | | | | |

| |--- publications <= 434.50

|--- broad\_impact <= 499.50

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |

| |--- value: [520.00]

|--- broad\_impact > 499.50

| |--- broad\_impact <= 515.50

| | |--- value: [562.00]

| |--- broad\_impact > 515.50

| | |--- value: [563.00]

|--- publications > 434.50

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |

| |--- value: [469.00]

| | | | | | |--- broad\_impact > 539.00

|--- publications <= 390.50

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |

| |--- value: [401.00]

|--- publications > 390.50

| |--- value: [370.00]

|--- publications > 458.00

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| | | | | | | | | | |
| | | | | | | | | | |

| |--- publications <= 510.00

|--- publications <= 505.50

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |

| |--- broad\_impact <= 652.50

| | |--- broad\_impact <= 558.00

|--- broad\_impact <= 525.50

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | |

| |--- value: [589.00]

|--- broad\_impact > 525.50

| |--- value: [608.00]

| | | | | | | | | |--- broad\_impact > 558.00

|--- publications <= 463.50

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | |

| |--- value: [548.00]

|--- publications > 463.50

| |--- value: [579.00]

| | | | | | | | |--- broad\_impact > 652.50

|--- publications <= 488.00

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |

| |--- value: [672.00]

|--- publications > 488.00

| |--- value: [568.00]

| |--- publications > 505.50

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |

| | |--- value: [422.00]

|--- publications > 510.00

| |--- broad\_impact <= 638.00

|--- broad\_impact <= 559.50

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |

| |--- value: [581.00]

|--- broad\_impact > 559.50

| |--- publications <= 567.50

| | |--- publications <= 528.50

| | | |--- value: [654.00]

| | |--- publications > 528.50

| | | |--- value: [637.00]

| |--- publications > 567.50

| | |--- value: [678.00]

|--- broad\_impact > 638.00

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |

| |--- value: [745.00]

|--- publications > 592.50

|  |  |  |  |
| --- | --- | --- | --- |
| | | | | | | | |
| | | | | | | | |

| |--- publications <= 596.50

|--- broad\_impact <= 504.50

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |

| |--- value: [529.00]

|--- broad\_impact > 504.50

| |--- value: [258.00]

| | | | | |--- publications > 596.50

|--- broad\_impact <= 573.50

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |

| |--- value: [568.00]

|--- broad\_impact > 573.50

| |--- broad\_impact <= 627.50

| | |--- value: [429.00]

| |--- broad\_impact > 627.50

| | |--- value: [473.00]

| | | |--- broad\_impact > 694.50

|--- broad\_impact <= 701.00

|  |  |  |  |
| --- | --- | --- | --- |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |

| |--- value: [729.00]

|--- broad\_impact > 701.00

| | | | | |--- value: [741.00]

| | |--- publications > 626.50

| | | |--- broad\_impact <= 536.00

| | | | |--- publications <= 699.00

|--- publications <= 664.00

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |

| |--- value: [512.00]

|--- publications > 664.00

| |--- value: [513.00]

| | | | |--- publications > 699.00

|--- publications <= 754.50

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |

| |--- value: [585.00]

|--- publications > 754.50

| |--- value: [590.00]

| | | |--- broad\_impact > 536.00

| | | | |--- broad\_impact <= 614.00

|--- publications <= 628.00

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |

| |--- value: [608.00]

|--- publications > 628.00

| |--- broad\_impact <= 561.50

|--- broad\_impact <= 553.50

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |

| |--- publications <= 740.00

| | |--- value: [633.00]

| |--- publications > 740.00

| | |--- value: [639.00]

|--- broad\_impact > 553.50

| |--- value: [616.00]

| | | | | | |--- broad\_impact > 561.50

|--- publications <= 943.00

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |

| |--- broad\_impact <= 582.00

|--- publications <= 662.50

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |

| |--- value: [638.00]

|--- publications > 662.50

| |--- publications <= 686.50

| | |--- value: [676.00]

| |--- publications > 686.50

| | |--- truncated branch of depth 2

| | | | | | | | |--- broad\_impact > 582.00

|--- publications <= 630.50

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |

| |--- value: [660.00]

|--- publications > 630.50

| |--- broad\_impact <= 605.50

| | |--- truncated branch of depth 2

| |--- broad\_impact > 605.50

| | |--- value: [670.00]

|--- publications > 943.00

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |

| |--- value: [623.00]

|--- broad\_impact > 614.00

|  |  |  |  |
| --- | --- | --- | --- |
| | | | | | | | |
| | | | | | | | |

| |--- broad\_impact <= 661.00

|--- publications <= 715.00

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |

| |--- value: [701.00]

|--- publications > 715.00

|--- broad\_impact <= 639.50

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |

| |--- value: [728.00]

|--- broad\_impact > 639.50

| |--- publications <= 856.00

| | |--- value: [748.00]

| |--- publications > 856.00

| | |--- value: [752.00]

|--- broad\_impact > 661.00

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| | | | | | | | | | |
| | | | | | | | | | |

| |--- publications <= 780.00

|--- broad\_impact <= 669.00

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |

| |--- value: [693.00]

|--- broad\_impact > 669.00

| |--- value: [762.00]

| | | | | | |--- publications > 780.00

|--- broad\_impact <= 682.50

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |

| |--- value: [607.00]

|--- broad\_impact > 682.50

|--- publications <= 938.50

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |

| |--- publications <= 867.00

| | |--- value: [684.00]

| |--- publications > 867.00

| | |--- value: [719.00]

|--- publications > 938.50

| |--- value: [644.00]

| |--- broad\_impact > 709.00

| | |--- broad\_impact <= 915.00

| | | |--- broad\_impact <= 769.50

| | | | |--- broad\_impact <= 755.00

|--- broad\_impact <= 740.00

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| | | | | | | | | | |
| | | | | | | | | | |

| |--- publications <= 803.50

|--- publications <= 753.50

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |

| |--- value: [797.00]

|--- publications > 753.50

| |--- value: [795.00]

|--- publications > 803.50

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | |
| | | | | | | | | | | | |

| |--- broad\_impact <= 715.50

|--- publications <= 838.50

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |

| |--- value: [773.00]

|--- publications > 838.50

| |--- value: [759.00]

| | | | | | | |--- broad\_impact > 715.50

|--- broad\_impact <= 727.50

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |

| |--- value: [735.00]

|--- broad\_impact > 727.50

| |--- value: [707.00]

| | | | | |--- broad\_impact > 740.00

|--- publications <= 686.50

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |

| |--- value: [776.00]

|--- publications > 686.50

| |--- broad\_impact <= 747.50

| | |--- value: [808.00]

| |--- broad\_impact > 747.50

| | |--- value: [818.00]

| | | | |--- broad\_impact > 755.00

|--- broad\_impact <= 762.50

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |

| |--- value: [667.00]

|--- broad\_impact > 762.50

| |--- value: [639.00]

| | | |--- broad\_impact > 769.50

| | | | |--- publications <= 824.00

|--- broad\_impact <= 903.00

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| | | | | | | | | | |
| | | | | | | | | | |

| |--- publications <= 651.00

|--- broad\_impact <= 833.50

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |

| |--- value: [807.00]

|--- broad\_impact > 833.50

| |--- value: [756.00]

| | | | | | |--- publications > 651.00

|--- publications <= 760.50

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |

| |--- broad\_impact <= 807.50

|--- publications <= 750.00

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |

| |--- publications <= 732.50

| | |--- value: [818.00]

| |--- publications > 732.50

| | |--- value: [819.00]

|--- publications > 750.00

| |--- value: [855.00]

|--- broad\_impact > 807.50

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |

| |--- broad\_impact <= 870.50

|--- publications <= 699.00

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | |

| |--- value: [889.00]

|--- publications > 699.00

| |--- truncated branch of depth 2

| | | | | | | | | |--- broad\_impact > 870.50

|--- publications <= 670.50

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | |

| |--- value: [878.00]

|--- publications > 670.50

| |--- value: [834.00]

|--- publications > 760.50

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |

| |--- publications <= 799.50

|--- broad\_impact <= 846.50

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |

| |--- value: [804.00]

|--- broad\_impact > 846.50

| |--- value: [770.00]

| | | | | | | | |--- publications > 799.50

|--- publications <= 806.50

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |

| |--- value: [843.00]

|--- publications > 806.50

| |--- value: [833.00]

|--- broad\_impact > 903.00

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| | | | | | | | | | |
| | | | | | | | | | |

| |--- value: [653.00]

| | | | |--- publications > 824.00

|--- broad\_impact <= 868.00

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| | | | | | | | | | |
| | | | | | | | | | |

| |--- publications <= 919.50

|--- broad\_impact <= 804.50

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |

| |--- value: [871.00]

|--- broad\_impact > 804.50

| |--- value: [861.00]

| |--- publications > 919.50

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |

| | |--- value: [892.00]

|--- broad\_impact > 868.00

| |--- broad\_impact <= 907.50

|--- broad\_impact <= 891.00

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |

| |--- value: [923.00]

|--- broad\_impact > 891.00

|--- broad\_impact <= 896.50

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |

| |--- value: [935.00]

|--- broad\_impact > 896.50

| |--- publications <= 946.50

| | |--- value: [929.00]

| |--- publications > 946.50

| | |--- value: [924.00]

|--- broad\_impact > 907.50

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | |
| | | | | | | | | | | | |

| |--- value: [858.00]

| | |--- broad\_impact > 915.00

| | | |--- broad\_impact <= 972.50

|--- broad\_impact <= 969.50

|  |  |  |  |
| --- | --- | --- | --- |
| | | | | | | | |
| | | | | | | | |

| |--- broad\_impact <= 922.00

|--- publications <= 993.50

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |

| |--- publications <= 898.00

| | |--- value: [920.00]

| |--- publications > 898.00

| | |--- value: [860.00]

|--- publications > 993.50

| |--- value: [962.00]

|--- broad\_impact > 922.00

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| | | | | | | | | | |
| | | | | | | | | | |

| |--- publications <= 894.00

|--- broad\_impact <= 951.00

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |

| |--- broad\_impact <= 934.50

| | |--- value: [950.00]

| |--- broad\_impact > 934.50

| | |--- value: [946.00]

|--- broad\_impact > 951.00

| |--- value: [928.00]

| | | | | | |--- publications > 894.00

|--- broad\_impact <= 962.50

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |

| |--- value: [975.00]

|--- broad\_impact > 962.50

| |--- value: [992.00]

| | | | |--- broad\_impact > 969.50

|--- publications <= 936.00

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |

| |--- value: [837.00]

|--- publications > 936.00

| |--- value: [807.00]

| | | |--- broad\_impact > 972.50

|--- publications <= 945.00

|  |  |  |  |
| --- | --- | --- | --- |
| | | | | | | | |
| | | | | | | | |

| |--- broad\_impact <= 978.00

|--- publications <= 873.50

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |

| |--- value: [986.00]

|--- publications > 873.50

| |--- value: [983.00]

| | | | | |--- broad\_impact > 978.00

|--- publications <= 840.00

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |

| |--- value: [985.00]

|--- publications > 840.00

| |--- broad\_impact <= 985.00

| | |--- value: [1000.00]

| |--- broad\_impact > 985.00

| | |--- value: [994.00]

|--- publications > 945.00

|  |  |  |  |
| --- | --- | --- | --- |
| | | | | | | | |
| | | | | | | | |

| |--- value: [920.00]

#### Визуализация дерева

In [36]: *# Визуализация дерева*

**def** get\_png\_tree(tree\_model\_param, feature\_names\_param): dot\_data **=** StringIO()

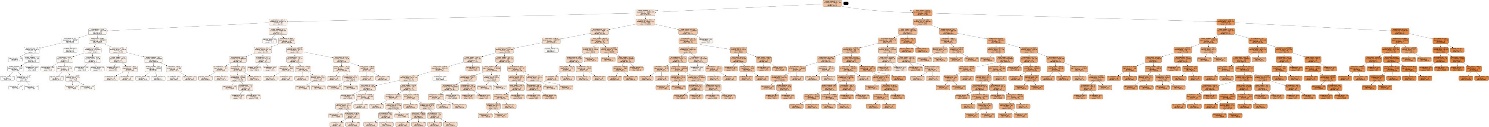
export\_graphviz(tree\_model\_param, out\_file**=**dot\_data, feature\_names**=**feature\_names\_param, filled**=True**, rounded**=True**, special\_characters**=True**)

graph **=** pydotplus**.**graph\_from\_dot\_data(dot\_data**.**getvalue())

**return** graph**.**create\_png()

In [37]: Image(get\_png\_tree(reg, X**.**columns), height**=**'100%')

Out[37]:



|  |
| --- |
|  |
|  |
|  |

#### Важность признаков в дереве

In [38]: **from** operator **import** itemgetter

**def** draw\_feature\_importances(tree\_model, X\_dataset, figsize**=**(18,5)): """

Вывод важности признаков в виде графика """

*# Сортировка значений важности признаков по убыванию*

list\_to\_sort **=** list(zip(X\_dataset**.**columns**.**values, tree\_model**.**feature\_importances\_)) sorted\_list **=** sorted(list\_to\_sort, key**=**itemgetter(1), reverse **= True**)

*# Названия признаков*

labels **=** [x **for** x,\_ **in** sorted\_list]

*# Важности признаков*

data **=** [x **for** \_,x **in** sorted\_list]

*# Вывод графика*

fig, ax **=** plt**.**subplots(figsize**=**figsize) ind **=** np**.**arange(len(labels)) plt**.**bar(ind, data)

plt**.**xticks(ind, labels, rotation**=**'vertical')

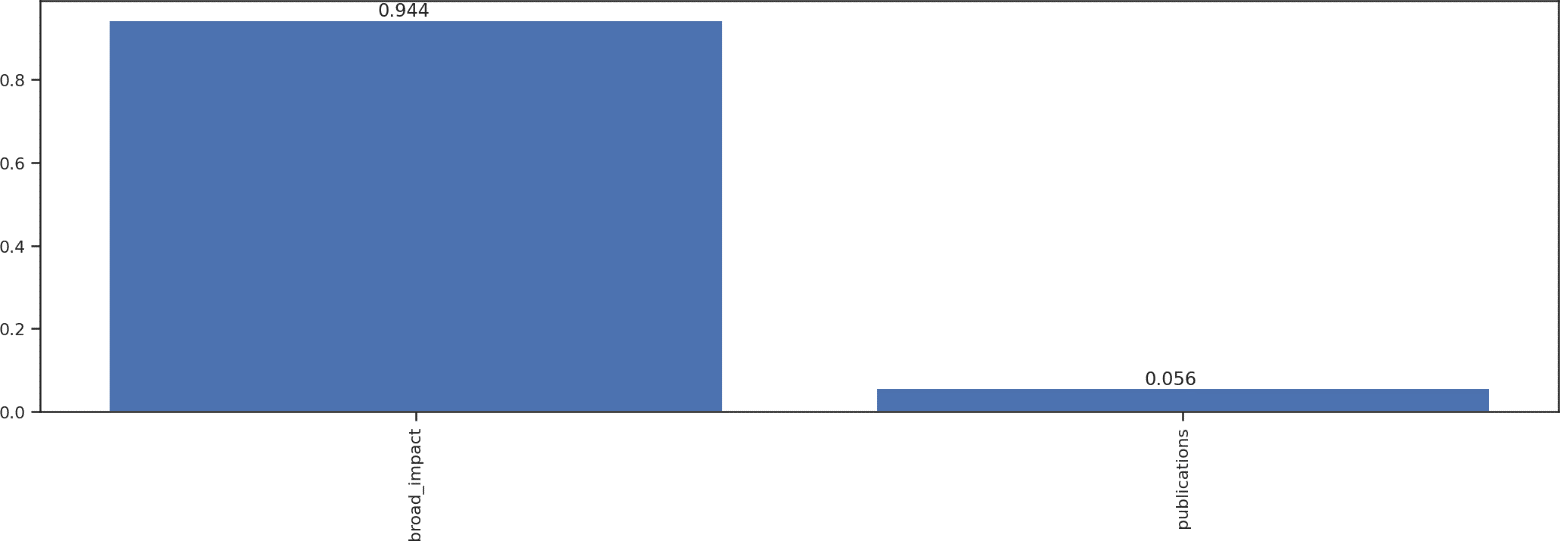
*# Вывод значений*

**for** a,b **in** zip(ind, data):

plt**.**text(a**-**0.05, b**+**0.01, str(round(b,3))) plt**.**show()

**return** labels, data

In [39]: boston\_tree\_regr\_fl, boston\_tree\_regr\_fd **=** draw\_feature\_importances(reg, X)



|  |
| --- |
|  |
|  |
|  |

##### In [ ]: