

Лабораторная работа №4

Загрузка данных

В качестве набора данных будем использовать [набор данных о хоралах Баха](#)

Набор данных содержит следующие колонки:

- chorale_ID - номер хорала
- event_number - id события
- pitch_1 - pitch_12 - наличие в аккорде каждой из 12 нот
- bass - нижняя нота аккорда
- meter - важность аккорда
- chord_label - название аккорда

In [13]:

```
# Импорт библиотек
import numpy as np
import pandas as pd
import seaborn as sns
from io import StringIO
import graphviz
import pydotplus
from IPython.display import Image
import matplotlib.pyplot as plt
%matplotlib inline
sns.set(style="ticks")

from sklearn.preprocessing import LabelEncoder
from sklearn.datasets import load_wine
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.svm import SVC
from sklearn.tree import DecisionTreeClassifier, export_graphviz
from sklearn.metrics import accuracy_score
from sklearn.metrics import confusion_matrix, plot_confusion_matrix

# Загрузка датасета
dataset = pd.read_csv('/content/drive/MyDrive/Colab Notebooks/datasets/bach_choral_set_dataset.csv')
```

In [14]:

```
# Первые 5 строк датасета
dataset.head()
```

Out[14]:

	choral_ID	event_number	pitch_1	pitch_2	pitch_3	pitch_4	pitch_5	pitch_6	pitch_7	pitch_8	pitch_9	pitch_10	pitch_11	pitch_12	bass	meter	chord_label
0	000106b_	1	YES	NO	NO	NO	NO	YES	NO	NO	NO	YES	NO	NO	F	3	F_M
1	000106b_	2	YES	NO	NO	NO	YES	NO	NO	YES	NO	NO	NO	NO	E	5	C_M
2	000106b_	3	YES	NO	NO	NO	YES	NO	NO	YES	NO	NO	NO	NO	E	2	C_M
3	000106b_	4	YES	NO	NO	NO	NO	YES	NO	NO	NO	YES	NO	NO	F	3	F_M
4	000106b_	5	YES	NO	NO	NO	NO	YES	NO	NO	NO	YES	NO	NO	F	2	F_M

In [15]:

```
# Типы колонок
dataset.dtypes
```

Out[15]:

```
choral_ID      object
event_number   int64
pitch_1        object
pitch_2        object
pitch_3        object
pitch_4        object
pitch_5        object
pitch_6        object
pitch_7        object
pitch_8        object
pitch_9        object
pitch_10       object
pitch_11       object
pitch_12       object
bass           object
meter          int64
chord_label    object
dtype: object
```

In [16]:

```
# Количество классов
len(dataset['chord_label'].unique())
```

Out[16]:

102

In [17]:

```
# Будем использовать только те классы, которые составляют более 5%
```

```
freqs = dataset['chord_label'].value_counts(normalize=True)*100
ind = []
for i in range(0, dataset.shape[0]):
    if freqs[dataset.iloc[i]['chord_label']]<5:
        ind.append(i)
dataset.drop(axis=0, index=ind, inplace = True)
dataset.shape
```

Out[17]: (2828, 17)

```
In [18]: # Новое количество классов
len(dataset['chord_label'].unique())
```

Out[18]: 7

```
In [19]: # Кодирование категориальных признаков
le = LabelEncoder()
dataset['choral_ID'] = le.fit_transform(dataset['choral_ID'])
for i in range(1,13):
    col = 'pitch_'+str(i)
    dataset[col] = le.fit_transform(dataset[col])
dataset['bass'] = le.fit_transform(dataset['bass'])
dataset['chord_label'] = le.fit_transform(dataset['chord_label'])
```

```
In [20]: # Разделение данных на целевые и нецелевые колонки
data = pd.DataFrame(dataset[dataset.columns.difference(['chord_label'])])
target = dataset['chord_label'].copy()
```

Разделение на обучающую и тестовую выборки

```
In [33]: # Разделение на обучающую и тестовую выборки
xtrain, xtest, ytrain, ytest = train_test_split(
    data, target, test_size=0.3, random_state=1)

train_acc = dict()
test_acc = dict()
```

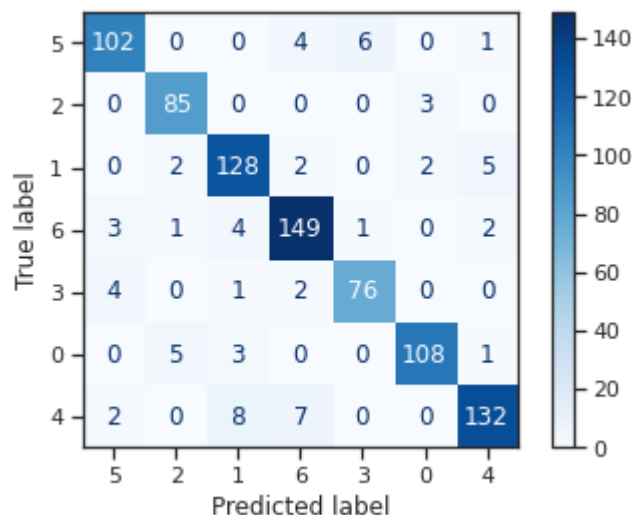
Обучение моделей

```
In [34]: # Логистическая регрессия
lr = LogisticRegression(solver='lbfgs', max_iter=10000)
lr.fit(xtrain, ytrain)
train_acc['Logistic Regression'] = accuracy_score(ytrain, lr.predict(xtrain))
test_acc['Logistic Regression'] = accuracy_score(ytest, lr.predict(xtest))
print('Train accuracy: ', train_acc['Logistic Regression'])
print('Test accuracy: ', test_acc['Logistic Regression'], '\n')
plot_confusion_matrix(lr, xtest, ytest,
                      display_labels=target.unique(), cmap=plt.cm.Blues)
```

Train accuracy: 0.9277412834765033
Test accuracy: 0.9187279151943463

/usr/local/lib/python3.7/dist-packages/sklearn/utils/deprecation.py:87: FutureWarning: Function plot_confusion_matrix is deprecated; Function `plot_confusion_matrix` is deprecated in 1.0 and will be removed in 1.2. Use one of the class methods: ConfusionMatrixDisplay.from_prediction or ConfusionMatrixDisplay.from_estimator.
warnings.warn(msg, category=FutureWarning)

Out[34]: <sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x7fdef38828d0>



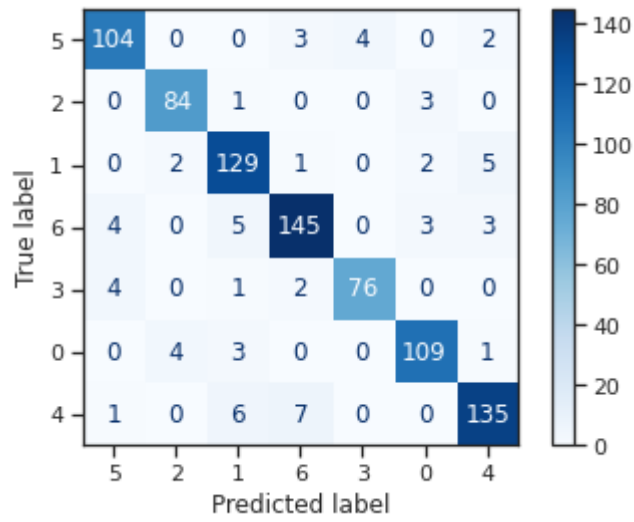
```
In [35]: # SVM
svc = SVC(kernel='linear')
svc.fit(xtrain, ytrain)
train_acc['SVC'] = accuracy_score(ytrain, svc.predict(xtrain))
test_acc['SVC'] = accuracy_score(ytest, svc.predict(xtest))
print('Train accuracy: ', train_acc['SVC'])
print('Test accuracy: ', test_acc['SVC'], '\n')
```

```
plot_confusion_matrix(svc, xtest, ytest,
                      display_labels=target.unique(), cmap=plt.cm.Blues)
```

Train accuracy: 0.9297625063163214
Test accuracy: 0.9210836277974087

/usr/local/lib/python3.7/dist-packages/sklearn/utils/deprecation.py:87: FutureWarning: Function plot_confusion_matrix is deprecated; Function `plot_confusion_matrix` is deprecated in 1.0 and will be removed in 1.2. Use one of the class methods: ConfusionMatrixDisplay.from_prediction or ConfusionMatrixDisplay.from_estimator.
warnings.warn(msg, category=FutureWarning)

Out[35]: <sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x7fdef3450a90>

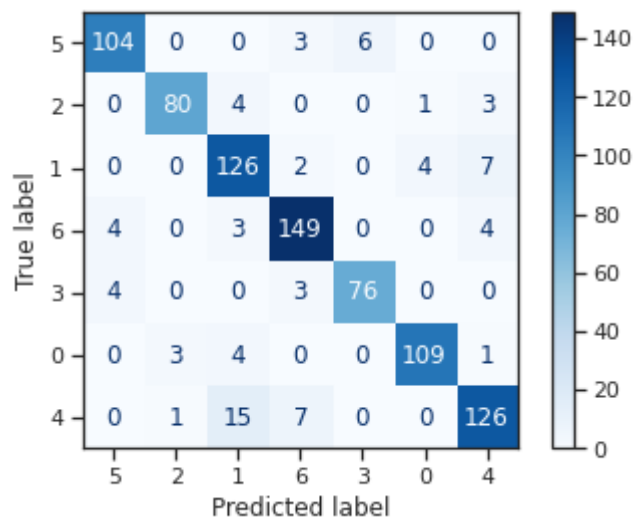


In [36]: *# Решающее дерево*
dtc = DecisionTreeClassifier()
dtc.fit(xtrain, ytrain)
train_acc['Decision Tree'] = accuracy_score(ytrain, dtc.predict(xtrain))
test_acc['Decision Tree'] = accuracy_score(ytest, dtc.predict(xtest))
print('Train accuracy: ', train_acc['Decision Tree'])
print('Test accuracy: ', test_acc['Decision Tree'], '\n')
plot_confusion_matrix(dtc, xtest, ytest,
 display_labels=target.unique(), cmap=plt.cm.Blues)

Train accuracy: 1.0
Test accuracy: 0.9069493521790342

/usr/local/lib/python3.7/dist-packages/sklearn/utils/deprecation.py:87: FutureWarning: Function plot_confusion_matrix is deprecated; Function `plot_confusion_matrix` is deprecated in 1.0 and will be removed in 1.2. Use one of the class methods: ConfusionMatrixDisplay.from_prediction or ConfusionMatrixDisplay.from_estimator.
warnings.warn(msg, category=FutureWarning)

Out[36]: <sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x7fdef32b5d50>



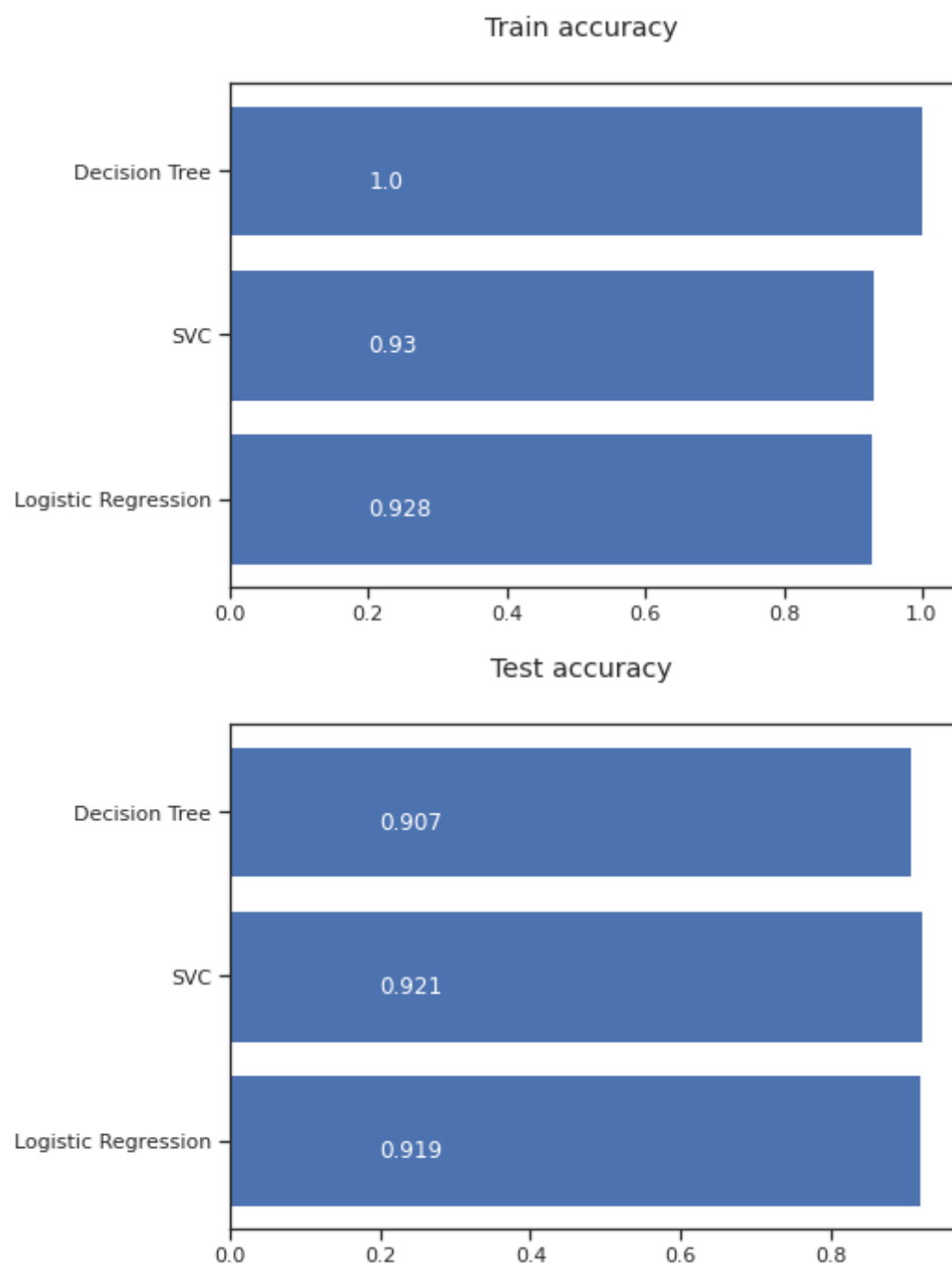
Сравнение качества моделей

In [37]:

```
def plot_acc(metric, subtitle):
    fig, ax = plt.subplots(figsize=(7,5))
    fig.suptitle(subtitle)
    ind = np.arange(len(metric))
    plt.barh(ind, metric.values())
    plt.yticks(ind, metric)
    # Вывод значений
    for a,b in zip(ind, metric.values()):
        plt.text(0.2, a-0.1, str(round(b,3)), color='white')
    plt.show()
    return
```

In [38]:

```
plot_acc(train_acc, 'Train accuracy')
plot_acc(test_acc, 'Test accuracy')
```



Визуализация решающего дерева

```
In [39]: # График важности признаков
from operator import itemgetter
# Сортировка значений важности признаков по убыванию
list_to_sort = list(zip(pd.DataFrame(data=data, columns=data.columns).columns.values, dtc.feature_importances_))
sorted_list = sorted(list_to_sort, key=itemgetter(1), reverse = True)
# Названия признаков
labels = [x for x, _ in sorted_list]
# Важности признаков
dat = [x for _, x in sorted_list]
# Вывод графика
fig, ax = plt.subplots(figsize=(15,7))
ind = np.arange(len(labels))
plt.bar(ind, dat)
plt.xticks(ind, labels, rotation='vertical')
# Вывод значений
for a,b in zip(ind, dat):
    plt.text(a-0.05, b+0.01, str(round(b,3)))
plt.show()
```



```

|--- bass > 13.00
|--- event_number <= 111.00
|--- class: 6
|--- event_number > 111.00
|--- class: 3
|--- pitch_11 > 0.50
|--- choral_ID <= 54.50
|--- pitch_8 <= 0.50
|--- class: 1
|--- pitch_8 > 0.50
|--- choral_ID <= 42.50
|--- class: 1
|--- choral_ID > 42.50
|--- meter <= 3.50
|--- class: 6
|--- meter > 3.50
|--- class: 1
|--- choral_ID > 54.50
|--- meter <= 3.50
|--- class: 5
|--- meter > 3.50
|--- class: 1
|--- pitch_12 > 0.50
|--- pitch_8 <= 0.50
|--- pitch_9 <= 0.50
|--- bass <= 9.50
|--- pitch_3 <= 0.50
|--- class: 5
|--- pitch_3 > 0.50
|--- choral_ID <= 10.50
|--- class: 0
|--- choral_ID > 10.50
|--- meter <= 4.00
|--- event_number <= 102.00
|--- choral_ID <= 34.50
|--- class: 6
|--- choral_ID > 34.50
|--- truncated branch of depth 2
|--- event_number > 102.00
|--- meter <= 2.50
|--- class: 5
|--- meter > 2.50
|--- class: 6
|--- meter > 4.00
|--- class: 0
|--- bass > 9.50
|--- class: 5
|--- pitch_9 > 0.50
|--- pitch_6 <= 0.50
|--- pitch_2 <= 0.50
|--- class: 4
|--- pitch_2 > 0.50
|--- class: 0
|--- pitch_6 > 0.50
|--- class: 3
|--- pitch_8 > 0.50
|--- meter <= 2.50
|--- choral_ID <= 3.00
|--- class: 5
|--- choral_ID > 3.00
|--- pitch_3 <= 0.50
|--- pitch_6 <= 0.50
|--- class: 6
|--- pitch_6 > 0.50
|--- class: 5
|--- pitch_3 > 0.50
|--- event_number <= 48.50
|--- bass <= 4.00
|--- class: 2
|--- bass > 4.00
|--- event_number <= 47.50
|--- choral_ID <= 21.50
|--- truncated branch of depth 5
|--- choral_ID > 21.50
|--- class: 6
|--- event_number > 47.50
|--- class: 2
|--- event_number > 48.50
|--- choral_ID <= 30.00
|--- class: 6
|--- choral_ID > 30.00
|--- choral_ID <= 33.50
|--- event_number <= 70.00
|--- class: 6

```

```

|--- event_number > 70.00
|   |--- class: 3
|       |--- choral_ID > 33.50
|           |--- choral_ID <= 37.50
|               |--- truncated branch of depth 3
|                   |--- choral_ID > 37.50
|                       |--- class: 6
|--- meter > 2.50
|   |--- pitch_10 <= 0.50
|       |--- choral_ID <= 43.50
|           |--- class: 6
|               |--- choral_ID > 43.50
|                   |--- event_number <= 8.50
|                       |--- choral_ID <= 49.50
|                           |--- class: 6
|                               |--- choral_ID > 49.50
|                                   |--- event_number <= 4.50
|                                       |--- class: 6
|                                           |--- event_number > 4.50
|                                               |--- class: 2
|--- event_number > 8.50
|   |--- choral_ID <= 45.50
|       |--- event_number <= 47.00
|           |--- class: 6
|               |--- event_number > 47.00
|                   |--- event_number <= 63.50
|                       |--- class: 2
|                           |--- event_number > 63.50
|                               |--- class: 6
|--- choral_ID > 45.50
|   |--- class: 6
|--- pitch_10 > 0.50
|   |--- event_number <= 21.50
|       |--- event_number <= 13.50
|           |--- class: 6
|               |--- event_number > 13.50
|                   |--- class: 3
|--- event_number > 21.50
|   |--- class: 6
|--- pitch_5 > 0.50
|   |--- pitch_1 <= 0.50
|       |--- pitch_10 <= 0.50
|           |--- pitch_9 <= 0.50
|               |--- bass <= 12.00
|                   |--- pitch_8 <= 0.50
|                       |--- bass <= 6.50
|                           |--- class: 0
|                               |--- bass > 6.50
|                                   |--- pitch_3 <= 0.50
|                                       |--- class: 4
|                                           |--- pitch_3 > 0.50
|                                               |--- class: 6
|--- pitch_8 > 0.50
|   |--- choral_ID <= 4.00
|       |--- class: 1
|           |--- choral_ID > 4.00
|               |--- event_number <= 99.00
|                   |--- choral_ID <= 8.00
|                       |--- class: 6
|                           |--- choral_ID > 8.00
|                               |--- meter <= 1.50
|                                   |--- class: 6
|                                       |--- meter > 1.50
|                                           |--- truncated branch of depth 7
|--- event_number > 99.00
|   |--- bass <= 6.50
|       |--- class: 6
|           |--- bass > 6.50
|               |--- class: 1
|--- bass > 12.00
|   |--- choral_ID <= 53.50
|       |--- meter <= 4.50
|           |--- pitch_3 <= 0.50
|               |--- class: 6
|                   |--- pitch_3 > 0.50
|                       |--- choral_ID <= 33.50
|                           |--- event_number <= 127.00
|                               |--- class: 6
|                                   |--- event_number > 127.00
|                                       |--- class: 3
|                                           |--- choral_ID > 33.50
|                                               |--- class: 3
|--- meter > 4.50
|   |--- class: 3

```



```

|--- choral_ID > 53.50
|--- class: 0
|--- pitch_9 > 0.50
|--- pitch_2 <= 0.50
|--- pitch_12 <= 0.50
|--- meter <= 2.50
|--- class: 3
|--- meter > 2.50
|--- class: 4
|--- pitch_12 > 0.50
|--- choral_ID <= 44.50
|--- choral_ID <= 9.50
|--- event_number <= 37.00
|--- event_number <= 31.00
|--- class: 4
|--- event_number > 31.00
|--- class: 0
|--- event_number > 37.00
|--- class: 4
|--- choral_ID > 9.50
|--- class: 4
|--- choral_ID > 44.50
|--- meter <= 3.50
|--- event_number <= 32.50
|--- class: 4
|--- event_number > 32.50
|--- event_number <= 33.50
|--- class: 0
|--- event_number > 33.50
|--- truncated branch of depth 5
|--- meter > 3.50
|--- class: 4
|--- pitch_2 > 0.50
|--- pitch_3 <= 0.50
|--- bass <= 12.50
|--- event_number <= 82.50
|--- class: 4
|--- event_number > 82.50
|--- class: 0
|--- bass > 12.50
|--- class: 0
|--- pitch_3 > 0.50
|--- class: 3
|--- pitch_10 > 0.50
|--- pitch_2 <= 0.50
|--- bass <= 5.00
|--- pitch_9 <= 0.50
|--- class: 0
|--- pitch_9 > 0.50
|--- event_number <= 75.50
|--- class: 4
|--- event_number > 75.50
|--- class: 0
|--- bass > 5.00
|--- pitch_3 <= 0.50
|--- event_number <= 83.00
|--- class: 4
|--- event_number > 83.00
|--- event_number <= 134.00
|--- class: 0
|--- event_number > 134.00
|--- class: 3
|--- pitch_3 > 0.50
|--- event_number <= 64.00
|--- class: 3
|--- event_number > 64.00
|--- event_number <= 140.50
|--- class: 0
|--- event_number > 140.50
|--- class: 3
|--- pitch_2 > 0.50
|--- bass <= 9.00
|--- pitch_9 <= 0.50
|--- pitch_12 <= 0.50
|--- pitch_8 <= 0.50
|--- bass <= 4.50
|--- class: 0
|--- bass > 4.50
|--- choral_ID <= 33.00
|--- truncated branch of depth 3
|--- choral_ID > 33.00
|--- class: 0
|--- pitch_8 > 0.50
|--- event_number <= 77.50

```

```

|--- class: 0
|--- event_number > 77.50
|--- event_number <= 78.50
|--- class: 3
|--- event_number > 78.50
|--- class: 0
|--- pitch_12 > 0.50
|--- event_number <= 52.00
|--- class: 0
|--- event_number > 52.00
|--- bass <= 3.00
|--- class: 4
|--- bass > 3.00
|--- class: 0
|--- pitch_9 > 0.50
|--- bass <= 3.00
|--- class: 0
|--- bass > 3.00
|--- class: 3
|--- bass > 9.00
|--- bass <= 12.00
|--- pitch_9 <= 0.50
|--- event_number <= 151.00
|--- event_number <= 24.50
|--- choral_ID <= 8.00
|--- class: 0
|--- choral_ID > 8.00
|--- class: 4
|--- event_number > 24.50
|--- choral_ID <= 16.00
|--- class: 0
|--- choral_ID > 16.00
|--- truncated branch of depth 3
|--- event_number > 151.00
|--- class: 3
|--- pitch_9 > 0.50
|--- class: 4
|--- bass > 12.00
|--- class: 0
|--- pitch_1 > 0.50
|--- pitch_10 <= 0.50
|--- pitch_9 <= 0.50
|--- pitch_12 <= 0.50
|--- pitch_11 <= 0.50
|--- pitch_6 <= 0.50
|--- class: 2
|--- pitch_6 > 0.50
|--- event_number <= 20.00
|--- class: 5
|--- event_number > 20.00
|--- class: 2
|--- pitch_11 > 0.50
|--- bass <= 12.00
|--- pitch_3 <= 0.50
|--- class: 2
|--- pitch_3 > 0.50
|--- bass <= 6.00
|--- class: 1
|--- bass > 6.00
|--- class: 2
|--- bass > 12.00
|--- class: 5
|--- pitch_12 > 0.50
|--- event_number <= 67.50
|--- class: 2
|--- event_number > 67.50
|--- choral_ID <= 43.00
|--- event_number <= 90.00
|--- class: 6
|--- event_number > 90.00
|--- class: 2
|--- choral_ID > 43.00
|--- class: 3
|--- pitch_9 > 0.50
|--- choral_ID <= 32.00
|--- class: 4
|--- choral_ID > 32.00
|--- class: 2
|--- pitch_10 > 0.50
|--- choral_ID <= 38.00
|--- bass <= 1.50
|--- choral_ID <= 24.00
|--- class: 5
|--- choral_ID > 24.00

```



```

| | | |--- class: 3
| | | |--- bass > 1.50
| | | |--- class: 4
| | | |--- pitch_8 > 0.50
| | | |--- class: 2
|--- pitch_3 > 0.50
| |--- pitch_12 <= 0.50
| | |--- pitch_2 <= 0.50
| | | |--- bass <= 6.00
| | | | |--- meter <= 4.50
| | | | |--- class: 3
| | | |--- meter > 4.50
| | | | |--- pitch_1 <= 0.50
| | | | |--- class: 3
| | | | |--- pitch_1 > 0.50
| | | | |--- class: 6
| | | |--- bass > 6.00
| | | |--- class: 3
|--- pitch_2 > 0.50
| |--- bass <= 3.00
| | |--- class: 0
| |--- bass > 3.00
| | |--- choral_ID <= 4.00
| | |--- class: 0
| | |--- choral_ID > 4.00
| | |--- class: 3
|--- pitch_12 > 0.50
| |--- pitch_8 <= 0.50
| | |--- event_number <= 62.50
| | | |--- event_number <= 7.50
| | | |--- class: 6
| | |--- event_number > 7.50
| | | |--- meter <= 2.50
| | | | |--- pitch_10 <= 0.50
| | | | |--- event_number <= 30.00
| | | | |--- class: 3
| | | | |--- event_number > 30.00
| | | | |--- event_number <= 43.50
| | | | |--- class: 6
| | | | |--- event_number > 43.50
| | | | |--- class: 3
| | | |--- pitch_10 > 0.50
| | | |--- event_number <= 48.00
| | | | |--- class: 6
| | | |--- event_number > 48.00
| | | | |--- class: 3
| | | |--- meter > 2.50
| | | |--- class: 3
| |--- event_number > 62.50
| | |--- event_number <= 121.50
| | |--- class: 6
| | |--- event_number > 121.50
| | |--- event_number <= 155.00
| | | |--- class: 3
| | |--- event_number > 155.00
| | | |--- class: 4
|--- pitch_8 > 0.50
| |--- class: 6

```