

```
In [1]: from repository.cassandra_provider import CassandraProvider
from ml.genre_classifier import GenreClassifier
from ml.views_predictor import ViewsPredictor
```

```
In [2]: from spark.spark_config import create_spark_session
spark = create_spark_session()
```

```
In [3]: cassandra_provider = CassandraProvider()
df = cassandra_provider.load(spark)
```

```
In [4]: train, test = df.randomSplit(weights=[0.8, 0.2], seed=200)
```

```
In [5]: gr = GenreClassifier()

gr_results = gr.fit(train).predict(test)
gr_evaluation = gr.evaluate_all(gr_results)
```

```
In [6]: print(gr_evaluation)

accuracy      f1  weightedPrecision  weightedRecall
0  0.818106  0.787121          0.776236        0.816539
```

```
In [7]: gr_results.sort("views", ascending=False).limit(10).toPandas()
```

Out [7]:

	id	artist	tag	title	views	word_vectors
0	553931	Adam Mickiewicz	misc	Pan Tadeusz - Inwokacja	1865798	[0.0565991740503038 -0.03586697340399648 0.0..
1	2029397	Taco Hemingway	rap	6 zer	344414	[-0.05327897261055868 -0.10824612087663515, ..
2	5092219	Mata	rap	Patointeligencja	310719	[-0.05081893871531631 -0.06867769900850296 0..
3	327365	Eripe x Quebonafide	rap	Prowo	304558	[-0.028140135611031663 -0.014489627847644059,..
4	5188731	Paris Platynov	rap	HOTELOWY FREESTYLE	289676	[-0.08874175558935518 -0.0488890867041038, ..
5	3058392	Quebonafide	rap	Cest la Vie	231780	[0.1330781336308869 0.06976153864219907 0.22..
6	2810938	Taco Hemingway	rap	Deszcz na betonie	215613	[0.01652138757002403 -0.11031870971406427, ..
7	3376425	Bedoes & Kubi Producent	rap	05:05	211734	[-0.0283162603013669 -0.08440886317949856 0... ..
8	3065829	Quebonafide	rap	Zorza	195184	[0.0034886569649513277 -0.07163304948068645, ..
9	2470767	PRO8L3M	rap	Molly	193069	[-0.02478589512311559 -0.051998302226468014 ..

In []:

```
vp = ViewsPredictor()

train_mapped, test_mapped = vp.logarithmise_label_and_get_avg_views(train)

vp_results = vp.fit(train_mapped).predict(test_mapped)
vp_evaluation = vp.evaluate_all(vp_results)
```

In [9]:

```
print(vp_evaluation)

      rmse        r2
0  1.13534  0.722831
```

In [10]:

```
vp_results.sort("views", ascending=False).limit(10).toPandas()
```

	Out [10]:	artist	id	tag	title	views	word_vec
0		Adam Mickiewicz	553931	misc	Pan Tadeusz - Inwokacja	1865798	[0.0565991740503 -0.03586697340399
1		Quebonafide	5240143	rap	ROMANTICPSYCHO	239558	[-0.05345424741233 -0.08227868943623
2		Quebonafide	5411320	rap	BUBBLETEA	229619	[0.016639173674985 -0.08069375495417
3		Mata	5092300	rap	GOMBAO 33	225523	[0.01500892147444 -0.2691422674934507
4		Taco Hemingway	3178722	rap	Nostalgia	216403	[-0.004972895757796 -0.1174805215479307
5		ReTo	3700075	rap	UA	215266	[-0.07767287366817 -0.08698837589133
6		TACONAFIDE	3585011	rap	Art-B	213712	[0.006288606781350 0.016805721306247
7		Quebonafide	3105552	rap	Kawa i xanax	204835	[-0.04548617502884 -0.1311005343999517
8		Taco Hemingway	4703918	rap	KABRIOLETY	203205	[-0.0283391281988 -0.07430131285350
9		Malik Montana, DaChoyce & SRNO	7644235	rap	Jetlag	198135	[-0.043330651176259 0.00295313327649691
	In [11]:	spark.stop()					
	In []:						