

```
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  0.814822  0.77558          0.776414        0.81422
```

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

	id	artist	tag	title	views	word_vectors	tag_
0	553931	Adam Mickiewicz	misc	Pan Tadeusz - Inwokacja	1865798	[-0.04651478830724955, 0.0016515845494965713, ...]	
1	3113072	Quebonafide	rap	Half dead	484043	[-0.10396414282670577, -0.021987587568789622, ...]	
2	2077751	(Vitas)	pop	7 элемент The 7th Element	396684	[-0.07940766612927501, -0.016255053190084606, ...]	
3	7158539	Mata	rap	Papuga	335146	[-0.05865863812225622, -0.13251383299773617, ...]	
4	327365	Eripe x Quebonafide	rap	Prowo	304558	[-0.025983956049751948, -0.07024048379267027, ...]	
5	3058392	Quebonafide	rap	Cest la Vie	231780	[0.1404371520233356, -0.10006561728087483, -0....]	
6	5092300	Mata	rap	GOMBAO 33	225523	[-0.0687727609369003, -0.1518336340896418, 0.0...]	
7	2810938	Taco Hemingway	rap	Deszcz na betonie	215613	[-0.1587285369816099, -0.02185077326002476, -0....]	
8	3700075	ReTo	rap	UA	215266	[-0.04667153238768647, -0.08222631633070043, ...]	
9	3394042	Rap Genius Polska	misc	Polskie płyty w 2018 roku. Premiery daty zapow...	214880	[0.012168938323315698, -0.09440418093623877, ...]	

```
In [ ]: vp = ViewsPredictor()

train_mapped, test_mapped = vp.logarithmise_label_and_get_song_count(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  1.693396  0.359783
```

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

Out[10]:	artist	id	tag	title	views	word_vectors	v
0	Adam Mickiewicz	553931	misc	Pan Tadeusz - Inwokacja	1865798	[-0.04651478830724955, 0.0016515845494965713, ...]	1
1	(Vitas)	2077751	pop	7 элемент The 7th Element	396684	[-0.07940766612927501, -0.016255053190084606, ...]	1
2	Rogal DDL	2431545	rap	Ebe Ebe	331225	[-0.03942871812604472, -0.09069790231290455, ...]	1
3	Eripe x Quebonafide	327365	rap	Prowo	304558	[-0.025983956049751948, -0.07024048379267027, ...]	1
4	TACONAFIDE	3620567	rap	8 kobiet Remix	277548	[-0.037840537774856346, -0.02619883215987956, ...]	1
5	RapNews	349009	misc	Polskie płyty w 2014 roku. Premiery daty zapow...	233016	[0.017575087953178005, -0.0794866965136982, -0...]	1
6	Quebonafide	5411320	rap	BUBBLETEA	229619	[-0.03165292132969114, -0.03564107951237666, ...]	1
7	Taco Hemingway	727125	rap	Wszystko jedno	227013	[-0.07648264223999113, -0.04824902051144593, ...]	1
8	Taco Hemingway	3178722	rap	Nostalgia	216403	[-0.07047664582893189, -0.07675312838244001, ...]	1
9	Taco Hemingway	2810938	rap	Deszcz na betonie	215613	[-0.1587285369816099, -0.02185077326002476, ...]	1

In [11]: `spark.stop()`