In [ ]:

```
%%html
<marquee style='width: 90%; color: blue;'><b>INSTITUTO INFNET | ESCOLA SUPERIOR DA
TECNOLOGIA DA INFORMAÇÃO</b></marquee>
```

Pós-graduação MIT em Inteligência Artificial, Machine Learning & Deep Learning

Projeto da da disciplina de Mineração de texto com stack ELK.

Bloco: Mineração de Texto e Processamento de Linguagem Natural

Docente: Felipe Fink Grael

Aluno: Winicius Botelho Faquieri

Sumário

Dataset

Este conjunto de dados contém reviews para instrumentos musicais comprados na Amazon - http://jmcauley.ucsd.edu/data/amazon/ (http://jmcauley.ucsd.edu/data/amazon/), no período de 2004 a 2014. Os produtos receberam diversas avaliações de clientes, classificadas em um intervalo de 1 a 5. Mais importante do que a nota atribuída ao produto são os comentários ou reviews, onde os clientes que já adquiriram um produto específico relatam e detalham a sua experiência, desde a chegada do produto no destino até a qualidade do produto em si. O dataset foi obtido da plataforma Kaggle e possui 10.261 registros com as seguintes variáveis:

Colunas	Descrição	N° Registros
reviewerID	ID do revisor	10.261
asin	ID do produto	10.261
reviewerName	Nome do revisor	10.261
helpful	Classificação de utilidade da revisão	10.261
reviewText	Campo textual (resenha /crítica)	10.261
overall	Classificação atribuída (1 a 5)	10.261
summary	Resumo da revisão	10.261
unixReviewTime	hora da revisão (hora unix)	10.261
reviewTime	hora da revisão (bruto)	10.261

Objetivos: No presente trabalho utilizaremos ténicas de NLP para inferir através da análise desses comentários se o sentimento do cliente foi positivo, neutro ou negativo. Mais especificamente, utilizaremos a query More Like This para realizar a predição do sentimento do cliente, simulando o algoritmo KNN (K Nearest Neighbor).

In [254]:

```python
# Bibliotecas:
import pandas as pd
from tqdm.auto import tqdm
from elasticsearch import Elasticsearch
from elasticsearch.helpers import bulk
import urllib3
import json
from wordcloud import WordCloud
import matplotlib.pyplot as plt
from sklearn.metrics import classification_report
```

In [10]:

```
tqdm.pandas()

df = pd.read_json("data/data_raw.json", orient='records', lines=True)

# Visualizando uma amostra aleatória de 10 registros:
df.sample(10)
```

Out[10]:

	reviewerID	asin	reviewerName	helpful	reviewText	overall	summary
3628	A3N27L13CWJ9HY	B0006IQNCK	D. R. Smith	[4, 4]	I love these strings because the sounds are so...	5	Good tones
753	A8AF0VV2LP4LV	B0002D0CA8	Ron1414	[0, 0]	I used to use Elixir Nano string, and a good a...	5	Excellent
5254	A2IZ3ST24HSO4H	B000P5LVSK	David McCarthy	[0, 0]	I play metal so I need a strong pick (the Jazz...	5	one of my fav "Jazz III" guitar pics
659	A19Q4B515ENF9C	B0002D0096	Richard L. Chism	[0, 0]	I bought this amp to replace a Squier SP-10 th...	5	Awesome Effects and a Great Sound
7028	A306YUQBEB2R9T	B001W99HE8	debi	[0, 0]	This microphone pics up instruments quite a bi...	5	GLS Seems Better than Shure!
5176	A10044ECXDUVKS	B000NJ2TIE	Dean J Copely	[0, 0]	Surprised really. For the price I dont think I...	4	These are actually very nice
8927	A2Y2U697SS98E1	B004Z17008	John Wade Long Jr. "LongJnSilver"	[0, 0]	The Snark is great. I have 2 for my 4 ukuleles...	5	A+++
9733	A7IZNVL7JXLM	B007T8OGLK	Carver "Gadget guy"	[4, 6]	I got this from a "direct from china" vendor. ...	3	Not bad for the price
2076	A3QTFLU2ZWH7WS	B0002F7K7Y	Audie	[1, 1]	I love D'Adarrio strings. Great tone and it la...	5	D'Adarrio EXL140-3D Nickel Wound Electric Guit...
7452	AJK15Q9JOEHRH	B002RLLD88	aceofbase	[0, 0]	Nice accessory, hold plenty of sticks, and the...	5	Nice accessory

In [11]:

```
# Algumas informações sobre os dados:
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10261 entries, 0 to 10260
Data columns (total 9 columns):
 #   Column          Non-Null Count  Dtype 
---  ------          --------------  ----- 
 0   reviewerID      10261 non-null  object
 1   asin            10261 non-null  object
 2   reviewerName    10234 non-null  object
 3   helpful         10261 non-null  object
 4   reviewText      10261 non-null  object
 5   overall         10261 non-null  int64 
 6   summary         10261 non-null  object
 7   unixReviewTime  10261 non-null  int64 
 8   reviewTime      10261 non-null  object
dtypes: int64(2), object(7)
memory usage: 721.6+ KB
```

In [12]:

```
# Verificando valores nulos
print('The null values in the dataset:')
df.isnull().sum()
```

```
The null values in the dataset:
```

Out[12]:

```
reviewerID         0
asin               0
reviewerName      27
helpful            0
reviewText         0
overall            0
summary            0
unixReviewTime     0
reviewTime         0
dtype: int64
```

In [13]:

```
# A coluna 'reviewerName' contém o nome do revisor e não é relevante para análise e, portan
# bem como as colunas: helpful e unixReviewTime.
df = df.drop(['reviewerName','unixReviewTime'], axis=1)
```

In [14]:

```
# Ajustando a data:
re_new = df["reviewTime"].str.split(",", n = 1, expand = True)

df["date"] = re_new[0]

df["year"] = re_new[1]

re_new1 = df["date"].str.split(" ", n = 1, expand = True)

df["month"] = re_new1[0]

df["day"] = re_new1[1]

df["date"] = df["year"] + '-' + df["month"] + '-' + df["day"]
```

In [15]:

```
# Excluindo variáveis não relevantes para análise:
df = df.drop(['reviewTime', 'year', 'month', 'day'], axis=1)
df.head()
```

Out[15]:

	reviewerID	asin	helpful	reviewText	overall	summary	date
0	A2IBPI20UZIR0U	1384719342	[0, 0]	Not much to write about here, but it does exac...	5	good	2014-02-28
1	A14VAT5EAX3D9S	1384719342	[13, 14]	The product does exactly as it should and is q...	5	Jake	2013-03-16
2	A195EZSQDW3E21	1384719342	[1, 1]	The primary job of this device is to block the...	5	It Does The Job Well	2013-08-28
3	A2C00NNG1ZQQG2	1384719342	[0, 0]	Nice windscreen protects my MXL mic and preven...	5	GOOD WINDSCREEN FOR THE MONEY	2014-02-14
4	A94QU4C90B1AX	1384719342	[0, 0]	This pop filter is great. It looks and perform...	5	No more pops when I record my vocals.	2014-02-21

In [16]:

```
# Verificando o preenchimento dos dados para cada ano:
df["date"] = df["date"].astype("datetime64")
df["date"].dt.year.value_counts()
```

Out[16]:

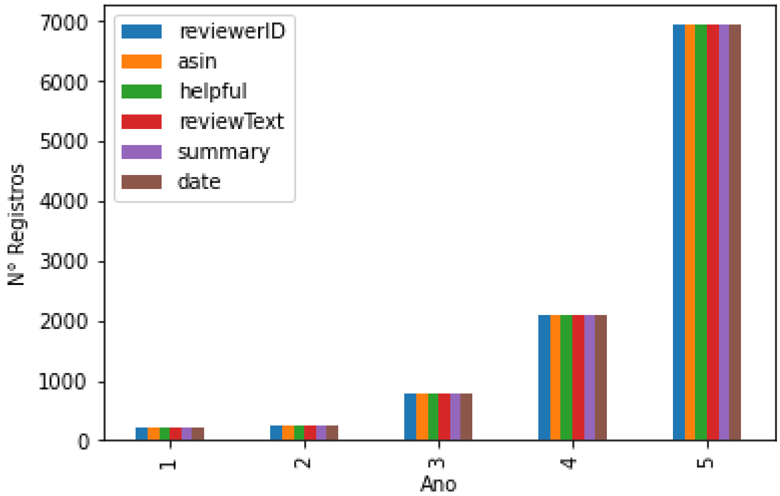
```
2013    4055
2014    2679
2012    1936
2011    1007
2010     350
2009     128
2008      63
2007      22
2006      10
2004       7
2005       4
Name: date, dtype: int64
```

In [17]:

```
# Visualizando a distribuição das notas atribuídas:
df.groupby(df["overall"]).count().plot(kind="bar",xlabel='Ano',ylabel="N° Registros")
```

Out[17]:

```
<AxesSubplot:xlabel='Ano', ylabel='N° Registros'>
```



In [18]:

```
# Combinando as colunas 'reviewText' e 'summary':
df['reviews'] = df['reviewText'] + df['summary']
df = df.drop(['reviewText', 'summary'], axis=1)
df.head()
```

Out[18]:

	reviewerID	asin	helpful	overall	date	reviews
0	A2IBPI20UZIR0U	1384719342	[0, 0]	5	2014-02-28	Not much to write about here, but it does exac...
1	A14VAT5EAX3D9S	1384719342	[13, 14]	5	2013-03-16	The product does exactly as it should and is q...
2	A195EZSQDW3E21	1384719342	[1, 1]	5	2013-08-28	The primary job of this device is to block the...
3	A2C00NNG1ZQQG2	1384719342	[0, 0]	5	2014-02-14	Nice windscreen protects my MXL mic and preven...
4	A94QU4C90B1AX	1384719342	[0, 0]	5	2014-02-21	This pop filter is great. It looks and perform...

In [19]:

```
# Classificando as notas por sentimento:

# Se a pontuação for maior que 3, o sentimento é Positivo
# Se o valor for menor do que 3, o sentimento é Negativo
# Se for igual a 3, o sentimento é Neutro.

def sent(rating):
    if rating['overall'] == 3:
        value = 'Neutral'
    elif rating['overall'] > 3:
        value = 'Positive'
    else:
        value = 'Negative'
    return value

#Applying the function in our new column
df['sentiment'] = df.apply(sent, axis=1)
df.head()
```

Out[19]:

	reviewerID	asin	helpful	overall	date	reviews	sentiment
0	A2IBPI20UZIR0U	1384719342	[0, 0]	5	2014-02-28	Not much to write about here, but it does exac...	Positive
1	A14VAT5EAX3D9S	1384719342	[13, 14]	5	2013-03-16	The product does exactly as it should and is q...	Positive
2	A195EZSQDW3E21	1384719342	[1, 1]	5	2013-08-28	The primary job of this device is to block the...	Positive
3	A2C00NNG1ZQQG2	1384719342	[0, 0]	5	2014-02-14	Nice windscreen protects my MXL mic and preven...	Positive
4	A94QU4C90B1AX	1384719342	[0, 0]	5	2014-02-21	This pop filter is great. It looks and perform...	Positive

In [20]:

```
print('The count of sentiments:')
df['sentiment'].value_counts()
```

```
The count of sentiments:
```

Out[20]:

```
Positive    9022
Neutral      772
Negative     467
Name: sentiment, dtype: int64
```

In [21]:

```
df.groupby([df["date"].dt.year,df["sentiment"]]).size()
```

Out[21]:

```
date  sentiment
2004  Positive        7
2005  Positive        4
2006  Negative        1
      Neutral         1
      Positive        8
2007  Negative        1
      Positive       21
2008  Negative        2
      Neutral         7
      Positive       54
2009  Negative        5
      Neutral         8
      Positive      115
2010  Negative       16
      Neutral        20
      Positive      314
2011  Negative       46
      Neutral        84
      Positive      877
2012  Negative       85
      Neutral       130
      Positive     1721
2013  Negative      170
      Neutral       319
      Positive     3566
2014  Negative      141
      Neutral       203
      Positive     2335
dtype: int64
```

In [22]:

```
# Tornando a coluna 'helpful' mais informativa:

df["helpful"] = df["helpful"].astype(str)

newreview = df["helpful"].str.split(",", n = 1, expand = True)

newreview1 = newreview[0].str.split("[", n = 1, expand = True)
newreview2 = newreview[1].str.split("]", n = 1, expand = True)

#Resetting the index
newreview1.reset_index(drop=True, inplace=True)
newreview2.reset_index(drop=True, inplace=True)

#Dropping empty columns due to splitting
newreview1 = newreview1.drop([0], axis=1)
newreview2 = newreview2.drop([1], axis=1)

#Concatenating the splitted columns
helpfulreview = pd.concat([newreview1, newreview2], axis=1)
helpfulreview
# df.head()

#Converting into integer types
helpfulreview[0] = helpfulreview[0].astype(str).astype(int)
helpfulreview[1] = helpfulreview[1].astype(str).astype(int)

#Dividing the two columns, we have 0 in the second columns when dvided gives error, so I'm
try:
  helpfulreview['result'] = helpfulreview[1]/helpfulreview[0]
except ZeroDivisionError:
  helpfulreview['result']=0

#Filling the NaN values(created due to dividing) with 0
helpfulreview['result'] = helpfulreview['result'].fillna(0)

#Rounding of the results to two decimal places
helpfulreview['result'] = helpfulreview['result'].round(2)

#Attaching the results to a new column of the main dataframe
df['helpful_review'] = helpfulreview['result']

df.head()
```

Out[22]:

	reviewerID	asin	helpful	overall	date	reviews	sentiment	helpful_review
0	A2IBPI20UZIR0U	1384719342	[0, 0]	5	2014-02-28	Not much to write about here, but it does exac...	Positive	0.0
1	A14VAT5EAX3D9S	1384719342	[13, 14]	5	2013-03-16	The product does exactly as it should and is q...	Positive	0.9

	reviewerID	asin	helpful	overall	date	reviews	sentiment	helpful_review
2	A195EZSQDW3E21	1384719342	[1, 1]	5	2013-08-28	The primary job of this device is to block the...	Positive	1.0
3	A2C00NNG1ZQQG2	1384719342	[0, 0]	5	2014-02-14	Nice windscreen protects my MXL mic and preven...	Positive	0.0
4	A94QU4C90B1AX	1384719342	[0, 0]	5	2014-02-21	This pop filter is great. It looks and perform...	Positive	0.0

In [158]:

```
# Exportando...
df.to_json('data/dataset.json', orient='records', lines=True)
```

Análise de Texto com Elasticsearch

In [181]:

```
# Pipeline de pré-processamento: analyzer

# !pip install elasticsearch

ES_URL = 'https://localhost:9200'
ES_USER = 'elastic'
ES_PASS = '08win0991i'

client = Elasticsearch(
    ES_URL,
    basic_auth=(ES_USER, ES_PASS),
    verify_certs=False
)

client
```

```
C:\Users\winicius.faquieri\Anaconda3\lib\site-packages\elasticsearch\_sync\c
lient\__init__.py:395: SecurityWarning: Connecting to 'https://localhost:920
0' using TLS with verify_certs=False is insecure
  _transport = transport_class(
```

Out[181]:

```
<Elasticsearch(['https://localhost:9200'])>
```

In [182]:

```
# Desligando warning para facilitar nossa vida
urllib3.disable_warnings()

dict(client.info())
```

Out[182]:

```
{'name': 'E012366',
 'cluster_name': 'elasticsearch',
 'cluster_uuid': 'Sy1H2ZYYQla4vTUqp3-s6A',
 'version': {'number': '8.2.2',
  'build_flavor': 'default',
  'build_type': 'zip',
  'build_hash': '9876968ef3c745186b94fdabd4483e01499224ef',
  'build_date': '2022-05-25T15:47:06.259735307Z',
  'build_snapshot': False,
  'lucene_version': '9.1.0',
  'minimum_wire_compatibility_version': '7.17.0',
  'minimum_index_compatibility_version': '7.0.0'},
 'tagline': 'You Know, for Search'}
```

In [183]:

```
# Customizando uma lista de stopwords para adicionar ao analyzer:
# Optamos por não utilizar a biblioteca NLTK, pois algumas palavras consideradas stopwords
# importantes na análise

en_stops = ['yourselves', 'between', 'whom', 'itself', 'is', "she's", 'up', 'herself', 'her
             'we', 'he', 'my', "you've", 'having', 'in', 'both', 'for', 'themselves', 'are'
             'and', 'an', 'during', 'their', 'can', 'yourself', 'she', 'until', 'so', 'thes
             'what', 'while', 'have', 're', 'more', 'only', "needn't", 'when', 'just', 'tha
             'very', 'should', 'any', 'y', 'isn', 'who',  'a', 'they', 'to', 'too', "should
             'into', 'yours', "it's", 'do', 'against', 'on',  'now', 'her', 've', 'd', 'by'
             'about', 'further', "that'll", "you'd", 'you', 'as', 'how', 'been', 'the', 'or
             'his', 'himself', 'ourselves',  'was', 'through', 'out', 'below', 'own', 'myse
             'me', 'why', 'once',  'him', 'than', 'be', 'most', "you'll", 'same', 'some', '
             'at', 'after', 'its', 'which', 'there','our', 'this', 'hers', 'being', 'did', 
             'over','again', 'where', 'those', 'then', "you're", 'i', 'because', 'does', 'a
```

Mapping e Configurações de Índice

Para armazenar dados (ou documentos) no Elasticsearch, é necessário fazer isso através de um index. Um index no Elasticsearch é um local para armazenar e organizar documentos relacionados. É comparável a um banco de dados contendo várias tabelas.

In [223]:

```
# Criando um analyzer customizado (rm caracteres especiais, transforma o token para minuscu
# Pré-processamento de textos com index mapping:

INDEX_NAME = 'index_reviews_amazon'

INDEX_MAPPING = {
    "settings": {
        "number_of_shards": 3,
        "analysis": {
            "analyzer": {
                "meu_analisador": {
                    "type": "custom",
                    "tokenizer": "standard",
                    "filter": ["asciifolding","lowercase","elision","stopsss","stop_custom"
                }
            },
            "filter": {
                "stopsss": {
                    "type": "stop",
                    "stopwords": "_english_",
                },
                "stop_custom": {
                    "type": "stop",
                    "stopwords": en_stops,
                },
            }
        }
    },
"mappings": {
        "properties": {
            "sentiment": {
                "type": "text",
                "fields": {
                    "raw": {
                        "type": "keyword"
                    }
                }
            },
            "reviews": {
                "type": "text",
                "analyzer": "meu_analisador",
                "fielddata": True,
                "fielddata_frequency_filter": {
                    # Deve aparecer em pelo menos em 1% dos documentos
                    "min": 0.01,
                    # Definindo um limite inferior para o nº ocorrencias
                    "min_segment_size": 10,
                 }
            }
        }
    }
}
```

In [224]:

```
# Bulk insert
def gen_documents(filename):
    with open('data/dataset.json','r') as fin:
        for line in fin:
            yield json.loads(line)
```

In [225]:

```
# Definindo uma função geradora para transformar um iterável de documentos em um iterável d
# def gen_index_actions(documents):
#     for doc in documents:
#         yield {
#             '_action': 'index',
#             '_index': INDEX_NAME,
#             **doc,
#         }
def gen_index_actions(documents):
    for doc in documents:
        yield {
            '_op_type': 'index',
            '_index': INDEX_NAME,
            '_source': doc
        }
```

In [226]:

```
# Apagar caso o indice exista
if client.indices.exists(index=INDEX_NAME):
    client.indices.delete(index=INDEX_NAME)
client.indices.create(index=INDEX_NAME, **INDEX_MAPPING)
```

Out[226]:

```
ObjectApiResponse({'acknowledged': True, 'shards_acknowledged': True, 'inde
x': 'index_reviews_amazon'})
```

In [227]:

```
# Testando o analyzer:
analyzed = client.indices.analyze(
    index= INDEX_NAME,
    analyzer="meu_analisador",
    text= "I got it to have it if I needed it. I have found that i don't really need it tha
)

print([x["token"] for x in analyzed['tokens']])
```

```
['got', 'needed', 'found', 'really', 'need', 'often', 'rarely', 'use', 'real
ly', 'good', 'see', 'need', 'keyboard', 'organ.it', 'works', 'great', 'hardl
y', 'use']
```

In [228]:

```
%%time
documents = gen_documents('data/dataset.json')
actions = gen_index_actions(documents)

success, errors = bulk(client, actions)

client.indices.refresh(index=INDEX_NAME)
client.indices.flush(index=INDEX_NAME)
```

```
Wall time: 14.1 s
```

Out[228]:

```
ObjectApiResponse({'_shards': {'total': 6, 'successful': 3, 'failed': 0}})
```

In [194]:

```
# Query com Agregação - Realiza operações sobre um conjunto de documentos. Comparável ao GR
QUERY = {
    'match_all': {}
}

AGG = {
    'grupos': {
        'terms': {
            'field': 'sentiment.raw',
            'size': 3,
        }
    }
}
```

In [147]:

```
# Realizando uma busca:
resp = client.search(index=INDEX_NAME, query=QUERY, aggregations=AGG, size=0)
resp
```

Out[147]:

```
ObjectApiResponse({'took': 3, 'timed_out': False, '_shards': {'total': 3, 's
uccessful': 3, 'skipped': 0, 'failed': 0}, 'hits': {'total': {'value': 1000
0, 'relation': 'gte'}, 'max_score': None, 'hits': []}, 'aggregations': {'gru
pos': {'doc_count_error_upper_bound': 0, 'sum_other_doc_count': 0, 'bucket
s': [{'key': 'Positive', 'doc_count': 9022}, {'key': 'Neutral', 'doc_count':
772}, {'key': 'Negative', 'doc_count': 467}]}}})
```

In [148]:

```
resp['hits']
```

Out[148]:

```
{'total': {'value': 10000, 'relation': 'gte'}, 'max_score': None, 'hits':
[]}
```

In [151]:

```
resp['aggregations']
```

Out[151]:

```
{'grupos': {'doc_count_error_upper_bound': 0,
  'sum_other_doc_count': 0,
  'buckets': [{'key': 'Positive', 'doc_count': 9022},
   {'key': 'Neutral', 'doc_count': 772},
   {'key': 'Negative', 'doc_count': 467}]}}
```

In [150]:

```
df_sentiment_agg = pd.DataFrame(resp['aggregations']['grupos']['buckets'])
df_sentiment_agg
```

Out[150]:

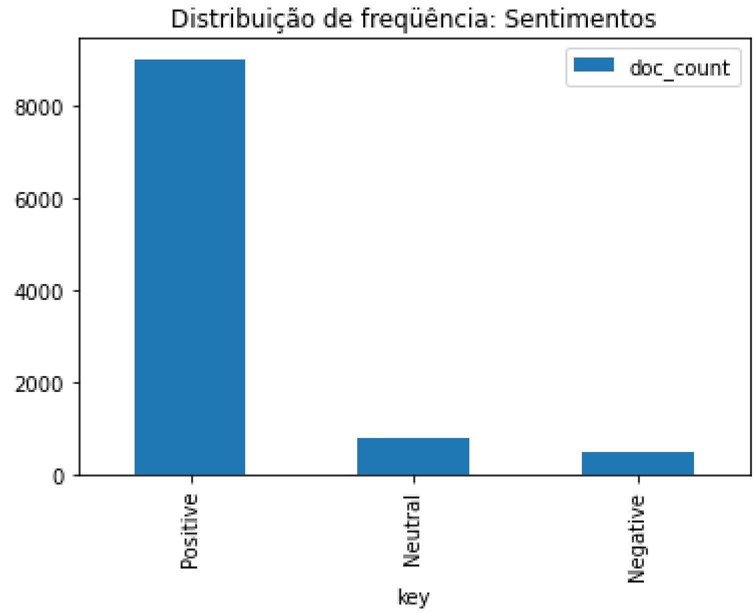
	key	doc_count
0	Positive	9022
1	Neutral	772
2	Negative	467

In [154]:

```
df_sentiment_agg.plot(title = 'Distribuição de freqüência: Sentimentos', x='key', kind='bar
```

Out[154]:

```
<AxesSubplot:title={'center':'Distribuição de freqüência: Sentimentos'}, xla
bel='key'>
```



In [195]:

```python
# Wordcloud para o sentimento 'Positive'
QUERY = {
    'term': {
        'sentiment.raw': 'Positive'
    }
}

AGG = {
    'corpo_textual': {
        'terms': {
            'field': 'reviews',
            'size': 2000,
        }
    }
}


resp = client.search(index=INDEX_NAME, query=QUERY, aggregations=AGG, size=2)
resp
```

Out[195]:

```
ObjectApiResponse({'took': 345, 'timed_out': False, '_shards': {'total':
3, 'successful': 3, 'skipped': 0, 'failed': 0}, 'hits': {'total': {'valu
e': 9022, 'relation': 'eq'}, 'max_score': 0.13400242, 'hits': [{'_index':
'index_reviews_amazon', '_id': '5w-ErIEB5LSonSuUbcnD', '_score': 0.1340024
2, '_source': {'_action': 'index', 'reviewerID': 'A2IBPI20UZIR0U', 'asin':
'1384719342', 'helpful': '[0, 0]', 'overall': 5, 'date': 1393545600000, 'r
eviews': "Not much to write about here, but it does exactly what it's supp
osed to. filters out the pop sounds. now my recordings are much more cris
p. it is one of the lowest prices pop filters on amazon so might as well b
uy it, they honestly work the same despite their pricing,good", 'sentimen
t': 'Positive', 'helpful_review': 0.0}}, {'_index': 'index_reviews_amazo
n', '_id': '6Q-ErIEB5LSonSuUbcnD', '_score': 0.13400242, '_source': {'_act
ion': 'index', 'reviewerID': 'A195EZSQDW3E21', 'asin': '1384719342', 'help
ful': '[1, 1]', 'overall': 5, 'date': 1377648000000, 'reviews': 'The prima
ry job of this device is to block the breath that would otherwise produce
a popping sound, while allowing your voice to pass through with no noticea
ble reduction of volume or high frequencies. The double cloth filter block
s the pops and lets the voice through with no coloration. The metal clamp
```

In [196]:

```
resp['aggregations']['corpo_textual']
```

Out[196]:

```
{'doc_count_error_upper_bound': 0,
 'sum_other_doc_count': 1852,
 'buckets': [{'key': 'great', 'doc_count': 3177},
  {'key': 'guitar', 'doc_count': 2916},
  {'key': 'good', 'doc_count': 2640},
  {'key': 'like', 'doc_count': 2477},
  {'key': 'one', 'doc_count': 2458},
  {'key': 'use', 'doc_count': 2346},
  {'key': 'well', 'doc_count': 2247},
  {'key': 'sound', 'doc_count': 1965},
  {'key': 'price', 'doc_count': 1801},
  {'key': 'get', 'doc_count': 1654},
  {'key': 'works', 'doc_count': 1563},
  {'key': 'really', 'doc_count': 1504},
  {'key': 'little', 'doc_count': 1500},
  {'key': 'would', 'doc_count': 1474},
  {'key': 'nice', 'doc_count': 1410},
  {'key': 'strings', 'doc_count': 1400},
```

In [199]:

```
# Frequência dos tokens mais relevantes para o sentimento positivo:
frequencies = {}

buckets = resp['aggregations']['corpo_textual']['buckets']

for bucket in buckets:
    frequencies[bucket['key']] = bucket['doc_count']

frequencies
```

Out[199]:

```
{'great': 3177,
 'guitar': 2916,
 'good': 2640,
 'like': 2477,
 'one': 2458,
 'use': 2346,
 'well': 2247,
 'sound': 1965,
 'price': 1801,
 'get': 1654,
 'works': 1563,
 'really': 1504,
 'little': 1500,
 'would': 1474,
 'nice': 1410,
 'strings': 1400,
 'much': 1338,
 'quality': 1324,
```

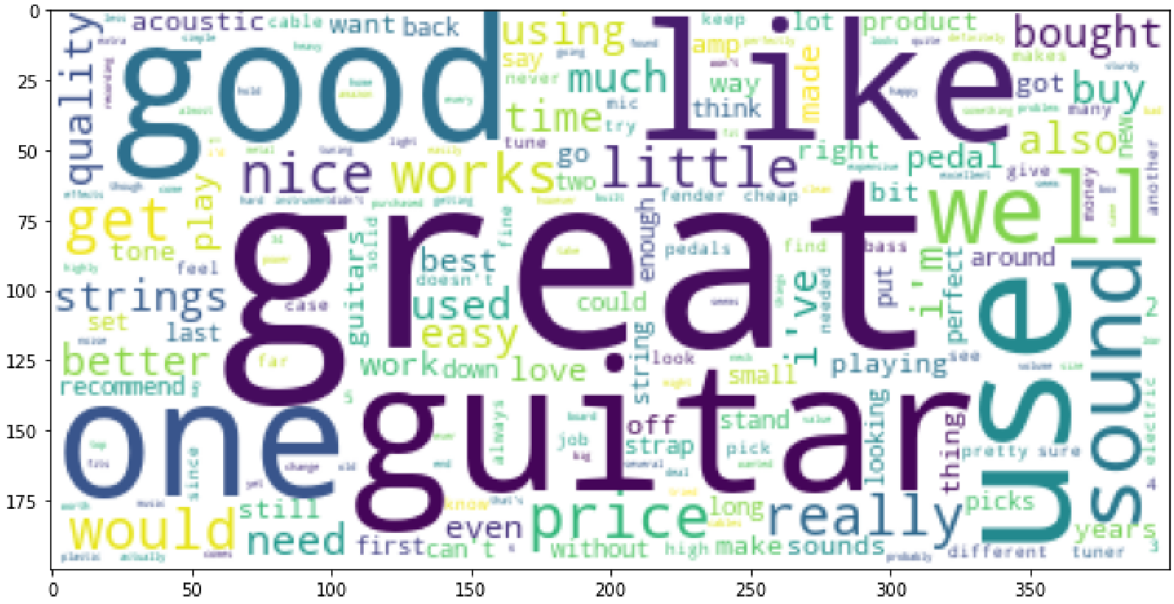
In [201]:

```
# Palavras como great, good, like, nice, well, quality, best definem bem o sentimento posit
cloud = WordCloud(background_color='white')
cloud.generate_from_frequencies(frequencies)

fig, ax = plt.subplots(figsize=(16, 6))
ax.imshow(cloud)
```

Out[201]:

```
<matplotlib.image.AxesImage at 0x1fecf3c4490>
```



Imitando k-NN com a query More Like This

In [229]:

```
# document
doc = client.search(index=INDEX_NAME, q='great', size=150)
doc
```

Out[229]:

```
ObjectApiResponse({'took': 262, 'timed_out': False, '_shards': {'total':
3, 'successful': 3, 'skipped': 0, 'failed': 0}, 'hits': {'total': {'valu
e': 3390, 'relation': 'eq'}, 'max_score': 2.2631245, 'hits': [{'_index':
'index_reviews_amazon', '_id': 'xQ-SrIEB5LSonSuUrPch', '_score': 2.263124
5, '_source': {'reviewerID': 'A35JPJHGOSPZX9', 'asin': 'B0002E1O2M', 'help
ful': '[0, 0]', 'overall': 5, 'date': 1385510400000, 'reviews': 'Great gre
at great great great great great great great.  Been uning them for years.
Love the sound and playability and durability.  matin coated are good too.
Great', 'sentiment': 'Positive', 'helpful_review': 0.0}}, {'_index': 'inde
x_reviews_amazon', '_id': 'ORCSrIEB5LSonSuU0A_W', '_score': 2.13108, '_sou
rce': {'reviewerID': 'AB7WPLA1CN5I1', 'asin': 'B002VA464S', 'helpful':
'[0, 0]', 'overall': 5, 'date': 1397347200000, 'reviews': 'This was an exc
ellent purchase.  Beside having a great range of audio pick-up, the mic is
so versatile.  Great product!Great Mic!', 'sentiment': 'Positive', 'helpfu
l_review': 0.0}}, {'_index': 'index_reviews_amazon', '_id': 'pRCSrIEB5LSon
SuUzglF', '_score': 2.12974, '_source': {'reviewerID': 'A2JQWSA0GPVSQN',
'asin': 'B0013PU75Y', 'helpful': '[0, 0]', 'overall': 5, 'date': 139760640
0000, 'reviews': "These have it all. Sound great, play great, feel great,
```

In [233]:

```
doc = doc['hits']['hits'][-1]
doc
```

Out[233]:

```
{'_index': 'index_reviews_amazon',
 '_id': '7BCSrIEB5LSonSuU1RWp',
 '_score': 1.8714406,
 '_source': {'reviewerID': 'A3G5BIFX6VS1OP',
  'asin': 'B005FKF3L6',
  'helpful': '[0, 0]',
  'overall': 5,
  'date': 1388793600000,
  'reviews': 'These strap pins are fantastic. They securely hold the strap o
n your guitar very nicely. The oval shape is perfect. There is no need for p
roprietary or expensive "strap locks". They are easy to install. These ellip
tical pins are great. Terrific yet simple concept. They also look great! I h
ighly recommend these to anyone who has a guitar. I will never use strap loc
ks again after trying these. Great value.Fantastic product',
  'sentiment': 'Positive',
  'helpful_review': 0.0}}
```

In [234]:

```
print(doc['_source']['sentiment'])
```

```
Positive
```

In [237]:

```
QUERY = {
    'more_like_this': {
        'fields': ['reviews'],
        'like': [
            {
                '_index': INDEX_NAME,
                '_id': 'xQ-SrIEB5LSonSuUrPch',
            }
        ],
        'min_term_freq': 1,
        'max_query_terms': 12,
    }
}
resp = client.search(index=INDEX_NAME, query=QUERY, size=10)
resp
```

Out[237]:

```
ObjectApiResponse({'took': 249, 'timed_out': False, '_shards': {'total':
3, 'successful': 3, 'skipped': 0, 'failed': 0}, 'hits': {'total': {'valu
e': 2794, 'relation': 'eq'}, 'max_score': 14.878317, 'hits': [{'_index':
'index_reviews_amazon', '_id': '9BCSrIEB5LSonSuU1xl6', '_score': 14.87831
7, '_source': {'reviewerID': 'A3KZEGBTPH6MMF', 'asin': 'B00IZCSW3M', 'help
ful': '[0, 0]', 'overall': 5, 'date': 1402704000000, 'reviews': "My father
is a full-time gigging musician primarily on Steel Guitar and Telecaster.
He has been playing over 50 years and it takes a lot to get him interested
in new strings.  We both use D'Addario strings almost exclusively (myself
on bass guitar, and him on the electric).  These are *outstanding* strings
with great durability, playability and exceptional sound.  They remain in
tune for long gigs or recording sessions, they generate less fret noise th
an comparable strings and have a bright/crisp sound that really stands ou
t. Well worth the extra investment.  Highly recommended!Great for pickin'
on the Tele or almost anything else!", 'sentiment': 'Positive', 'helpful_r
eview': 0.0}}, {'_index': 'index_reviews_amazon', '_id': 'BA-SrIEB5LSonSuU
rvkU', '_score': 13.420241, '_source': {'reviewerID': 'ABC68JUCPTVOE', 'as
in': 'B0002E3CHC', 'helpful': '[0, 0]', 'overall': 5, 'date': 136503360000
0, 'reviews': "I'm impressed with these EXP16 guitar strings.* What I Expe
ctedA more dead sound than the non-coated strings they replaced. Coated st
rings usually lack the bright high-end brilliance that uncoated light stri
ngs have.* What I GotA very bright, yet not to twangy sound.  When I put L
ight strings on my acoustic, I'm going for easy playability and, most of a
ll, a very bright and airy sound.  These deliver on both counts.* Durabili
tyThe coating on these strings are supposed to prolong the time between ch
anges.  So far they are holding their tone and brightness.  I will report
back after three months.  My last set started to sound dull, and exhibit m
ore tone variation after six months (when viewed with a very sensitive tun
er, it seems that my dirtier strings detune sharp a bit more on attack, th
e harder you strum them.  This is confirmed by the fact that they just sou
nd out of tune to me when they get dirty).* ConclusionI will be ordering m
ore of these.  I think they perform very well for the price. Love the brig
ht light gauge sound and playability.Sound like uncoated strings", 'sentim
ent': 'Positive', 'helpful_review': 0.0}}, {'_index': 'index_reviews_amazo
n', '_id': '4hCSrIEB5LSonSuU1xl6', '_score': 12.940178, '_source': {'revie
werID': 'A29B4PAIOL7HYG', 'asin': 'B00HFRXACG', 'helpful': '[1, 1]', 'over
all': 5, 'date': 1402444800000, 'reviews': "I have to admit this up front
- I'm a terrible bass player. I play guitar and keep a bass in the music r
oom for times when I have friend over to jam. I do dabble with the bass, b
ut I'm not proficient with it. That being said; I really appreciate the di
```

```
fference these strings make in sound, playability and durability. These st
rings sound better and feel great to play for long periods of time. They h
ave &#34;punch&#34; and &#34;snap&#34; which are good things to have in a
bass. I haven't had them on for a real long time, but in the time I've had
them they remain consistent and show absolutely no sound or metal degradat
ion in the time they've been on my Fender MIM Jazz bass. When I need anoth
er set, I'll buy these!Great strings", 'sentiment': 'Positive', 'helpful_r
eview': 1.0}}, {'_index': 'index_reviews_amazon', '_id': 'BQ-SrIEB5LSonSuU
rvkU', '_score': 12.078281, '_source': {'reviewerID': 'A22Z554ZQ8NFPC', 'a
sin': 'B0002E3CHC', 'helpful': '[2, 2]', 'overall': 5, 'date': 12990240000
00, 'reviews': "I've used D'Addario strings for many, many years without p
roblems - both acoustic & electric.  I've tried these coated strings in th
e past, and I actually prefer them to Elixir because they feel more like r
egular strings - the wrap is coated, rather than a sheathe being placed ov
er the entire string, like on Elixirs.Do they last longer?  Probably.  Eve
n when I'm in situations where I'm gigging regularly I rarely change strin
gs more than every few months, just because I don't see the need.  I think
the main benefit to coated strings is if you have guitars that you can't p
lay regularly the strings won't corrode.  But these sound great and feel f
ine, and the're the least expensive coated strings I've found.  I've alway
s found D'Addario quality to be great.Good strings", 'sentiment': 'Positiv
e', 'helpful_review': 1.0}}, {'_index': 'index_reviews_amazon', '_id': 'TB
CSrIEB5LSonSuUxQVC', '_score': 11.74545, '_source': {'reviewerID': 'A24Z0T
LWP8FXGV', 'asin': 'B000KIRT74', 'helpful': '[0, 0]', 'overall': 5, 'dat
e': 1395619200000, 'reviews': 'I love this pedal. It gives just the right
amount of crunch when needed. Typical Behringer pedal with plastic case bu
t a really good sound. I use this in my home studio so durability is not a
concern. Nice and quiet when not in use and can scream when you need it t
o. Highly recommended.Great little pedal', 'sentiment': 'Positive', 'helpf
ul_review': 0.0}}, {'_index': 'index_reviews_amazon', '_id': 'CBCSrIEB5LSo
nSuU1xp6', '_score': 11.617835, '_source': {'reviewerID': 'A1SD1C8XK3Z3V
1', 'asin': 'B00JBIVXGC', 'helpful': '[0, 0]', 'overall': 5, 'date': 14040
86400000, 'reviews': "I've used Elixirs for about five years now. This set
offers the same great sound for my acoustic as their electric strings prov
ide for my electric. Some people don't like the polymer (nanoweb) that coa
ts the string, but I find that it cuts back on string noise. Sliding your
fingers around doesn't cause too much noise, and with acoustic strings I f
ind that very important, especially if you're using a piezo pickup in your
instrument. Phosphor Bronze has always been a great alloy for acoustic gui
tars IMHO. These are great strings, worth the money, and sound great. They
also last 3 times longer than regular, non-coated strings in my experience
(been playing for 24yrs). Great product.Elixirs just sound good", 'sentime
nt': 'Positive', 'helpful_review': 0.0}}, {'_index': 'index_reviews_amazo
n', '_id': '8Q-SrIEB5LSonSuUsfz_', '_score': 11.4297695, '_source': {'revi
ewerID': 'A9P4Q7KX35RS2', 'asin': 'B0002H04NE', 'helpful': '[0, 0]', 'over
all': 5, 'date': 1390694400000, 'reviews': "I usually use the Pro Steels
(by D'Addario)  but these sound and fit nicely with my Squire Fender P Bas
s. Amazon seems to have great prices and it's nice to have them delivered
too.Great Pratice and Live Strings", 'sentiment': 'Positive', 'helpful_rev
iew': 0.0}}, {'_index': 'index_reviews_amazon', '_id': 'AQ-SrIEB5LSonSuUsf
3_', '_score': 11.4031105, '_source': {'reviewerID': 'A398X9POBHK69N', 'as
in': 'B0002H0A3S', 'helpful': '[0, 0]', 'overall': 5, 'date': 138317760000
0, 'reviews': "These were the same strings originally strung on my Godin S
eagull, and I absolutely love their sound, feel, and durability. The one r
eason I actually needed them because my G string broke while I was tuning
too fast. The sound is almost identical to what D'Addario describes, these
strings have a great, uplifting sound to them, hard to soft strums. Playin
g feels great with them, nothing out of the ordinary either strumming or t
he fretboard feel, clean and smooth. When replacing, I ran into no problem
s, quick and easy. I love the strings and their price, definitely will sti
ck with this brand as well.Sound Amazing On My Seagull, Top Pick For Strin
```

```
gs", 'sentiment': 'Positive', 'helpful_review': 0.0}}, {'_index': 'index_r
eviews_amazon', '_id': 'Cw-SrIEB5LSonSuUrvkU', '_score': 11.313062, '_sour
ce': {'reviewerID': 'A1LHMSY3Q46PJS', 'asin': 'B0002E3CHC', 'helpful':
'[5, 6]', 'overall': 5, 'date': 1197158400000, 'reviews': "I've been using
these strings for about eight years, and I also use them on my two Taylo
r's (814-CE & I use the mediums on my 710) I agree with the other reviewer
in saying they are the best sounding coated string out there. I still use
Martin strings once in awhile, but these strings last and improve in sound
as they get a few hours of play on them.Actually I own three other acousti
cs other than my two Taylor's and right now they're all strung with EXP's.
I know people want different sounds from their strings, but I find these t
o be bright with a great bottom on them, unlike the Elixirs which I find t
o be unbearably bright with a weak bottom end. These strings sound great w
ith great dynamics on them. Even on my Laguna Grand Symphony (slightly lar
ger than a Grand Concert) they pump out the lows and sparkle on the highs.
These on my Grand Auditorium do the same thing, they produce the sound I d
emand and make a real positive improvement over the other coated strings
I've used.The Mediums on my 710 have an endless bottom end and still have
great range on the mid's and treble.As I said I still use Martin Phosphor
Bronze strings once in awhile, but honestly these strings hold up extremel
y well and sound great. Darn near as near to an uncoated string that you'r
e going to find.Great sounding set of coated strings.", 'sentiment': 'Posi
tive', 'helpful_review': 0.83}}, {'_index': 'index_reviews_amazon', '_id':
'TQ-SrIEB5LSonSuUsf3_', '_score': 11.042872, '_source': {'reviewerID': 'A2
SKZZZBV1XYWP', 'asin': 'B0002H0A3S', 'helpful': '[8, 10]', 'overall': 5,
'date': 1268524800000, 'reviews': "Not much to say here. Strings are a ver
y personal thing to a guitar player. You need to try different gauges and
brands till you find the ones that work best for you and your style. These
are the string of choice for my acoustic and have been for many many year
s. I replace my strings about once a month with casual play and proper mai
ntenance so why run to the store or pay for shipping 10 times a year. The
strings are great, priced great and was on sale which made it even better.
There is less waste with the packaging which help the planet just a bit so
that's good too.Great Strings for Great Price", 'sentiment': 'Positive',
'helpful_review': 0.8}}]}})
```

In [238]:

```
resp_df = pd.DataFrame({'_id': x['_id'], '_score': x['_score'], **x['_source']} for x in re
resp_df
```

Out[238]:

	_id	_score	reviewerID	asin	helpful	overall	
0	9BCSrIEB5LSonSuU1xl6	14.878317	A3KZEGBTPH6MMF	B00IZCSW3M	[0, 0]	5	1
1	BA-SrIEB5LSonSuUrvkU	13.420241	ABC68JUCPTVOE	B0002E3CHC	[0, 0]	5	1
2	4hCSrIEB5LSonSuU1xl6	12.940178	A29B4PAIOL7HYG	B00HFRXACG	[1, 1]	5	1
3	BQ-SrIEB5LSonSuUrvkU	12.078281	A22Z554ZQ8NFPC	B0002E3CHC	[2, 2]	5	1
4	TBCSrIEB5LSonSuUxQVC	11.745450	A24Z0TLWP8FXGV	B000KIRT74	[0, 0]	5	1
5	CBCSrIEB5LSonSuU1xp6	11.617835	A1SD1C8XK3Z3V1	B00JBIVXGC	[0, 0]	5	1
6	8Q-SrIEB5LSonSuUsfz_	11.429770	A9P4Q7KX35RS2	B0002H04NE	[0, 0]	5	1
7	AQ-SrIEB5LSonSuUsf3_	11.403111	A398X9POBHK69N	B0002H0A3S	[0, 0]	5	1
8	Cw-SrIEB5LSonSuUrvkU	11.313062	A1LHMSY3Q46PJS	B0002E3CHC	[5, 6]	5	1

	_id	_score	reviewerID	asin	helpful	overall	
9	TQ-SrIEB5LSonSuUsf3_	11.042872	A2SKZZZBV1XYWP	B0002H0A3S	[8, 10]	5	1

In [239]:

```
# Classificação baseado somente nos 10 mais próximos
resp_df['sentiment'].value_counts()
```

Out[239]:

```
Positive    10
Name: sentiment, dtype: int64
```

In [241]:

```
# Definindo um classificador:
def classify_with_score(like, size=10):

    query = {
        'more_like_this': {
            'fields': ['reviews'],
            'like': like,
            'min_term_freq': 1,
            'max_query_terms': 12,
        }
    }

    resp = client.search(index=INDEX_NAME, query=query, size=size)

    resp_df = pd.DataFrame(
        {
            '_id': x['_id'],
            '_score': x['_score'],
            **x['_source']
        } for x in resp['hits']['hits']
    )
    return resp_df.groupby('sentiment').sum().index[0]
```

In [242]:

```
resp_df.groupby('sentiment').sum().index[0]
```

Out[242]:

```
'Positive'
```

In [243]:

```
resp_df['_id'].unique()
```

Out[243]:

```
array(['9BCSrIEB5LSonSuU1xl6', 'BA-SrIEB5LSonSuUrvkU',
       '4hCSrIEB5LSonSuU1xl6', 'BQ-SrIEB5LSonSuUrvkU',
       'TBCSrIEB5LSonSuUxQVC', 'CBCSrIEB5LSonSuU1xp6',
       '8Q-SrIEB5LSonSuUsfz_', 'AQ-SrIEB5LSonSuUsf3_',
       'Cw-SrIEB5LSonSuUrvkU', 'TQ-SrIEB5LSonSuUsf3_'], dtype=object)
```

In [244]:

```
resp
```

Out[244]:

```
ObjectApiResponse({'took': 249, 'timed_out': False, '_shards': {'total':
3, 'successful': 3, 'skipped': 0, 'failed': 0}, 'hits': {'total': {'valu
e': 2794, 'relation': 'eq'}, 'max_score': 14.878317, 'hits': [{'_index':
'index_reviews_amazon', '_id': '9BCSrIEB5LSonSuU1xl6', '_score': 14.87831
7, '_source': {'reviewerID': 'A3KZEGBTPH6MMF', 'asin': 'B00IZCSW3M', 'help
ful': '[0, 0]', 'overall': 5, 'date': 1402704000000, 'reviews': "My father
is a full-time gigging musician primarily on Steel Guitar and Telecaster.
He has been playing over 50 years and it takes a lot to get him interested
in new strings.  We both use D'Addario strings almost exclusively (myself
on bass guitar, and him on the electric).  These are *outstanding* strings
with great durability, playability and exceptional sound.  They remain in
tune for long gigs or recording sessions, they generate less fret noise th
an comparable strings and have a bright/crisp sound that really stands ou
t. Well worth the extra investment.  Highly recommended!Great for pickin'
on the Tele or almost anything else!", 'sentiment': 'Positive', 'helpful_r
eview': 0.0}}, {'_index': 'index_reviews_amazon', '_id': 'BA-SrIEB5LSonSuU
rvkU', '_score': 13.420241, '_source': {'reviewerID': 'ABC68JUCPTVOE', 'as
in': 'B0002E3CHC', 'helpful': '[0, 0]', 'overall': 5, 'date': 136503360000
0, 'reviews': "I'm impressed with these EXP16 guitar strings.* What I Expe
ctedA more dead sound than the non-coated strings they replaced. Coated st
rings usually lack the bright high-end brilliance that uncoated light stri
ngs have.* What I GotA very bright, yet not to twangy sound.  When I put L
ight strings on my acoustic, I'm going for easy playability and, most of a
ll, a very bright and airy sound.  These deliver on both counts.* Durabili
tyThe coating on these strings are supposed to prolong the time between ch
anges.  So far they are holding their tone and brightness.  I will report
back after three months.  My last set started to sound dull, and exhibit m
ore tone variation after six months (when viewed with a very sensitive tun
er, it seems that my dirtier strings detune sharp a bit more on attack, th
e harder you strum them.  This is confirmed by the fact that they just sou
nd out of tune to me when they get dirty).* ConclusionI will be ordering m
ore of these.  I think they perform very well for the price. Love the brig
ht light gauge sound and playability.Sound like uncoated strings", 'sentim
ent': 'Positive', 'helpful_review': 0.0}}, {'_index': 'index_reviews_amazo
n', '_id': '4hCSrIEB5LSonSuU1xl6', '_score': 12.940178, '_source': {'revie
werID': 'A29B4PAIOL7HYG', 'asin': 'B00HFRXACG', 'helpful': '[1, 1]', 'over
all': 5, 'date': 1402444800000, 'reviews': "I have to admit this up front
- I'm a terrible bass player. I play guitar and keep a bass in the music r
oom for times when I have friend over to jam. I do dabble with the bass, b
ut I'm not proficient with it. That being said; I really appreciate the di
fference these strings make in sound, playability and durability. These st
rings sound better and feel great to play for long periods of time. They h
ave &#34;punch&#34; and &#34;snap&#34; which are good things to have in a
bass. I haven't had them on for a real long time, but in the time I've had
them they remain consistent and show absolutely no sound or metal degradat
ion in the time they've been on my Fender MIM Jazz bass. When I need anoth
er set, I'll buy these!Great strings", 'sentiment': 'Positive', 'helpful_r
eview': 1.0}}, {'_index': 'index_reviews_amazon', '_id': 'BQ-SrIEB5LSonSuU
rvkU', '_score': 12.078281, '_source': {'reviewerID': 'A22Z554ZQ8NFPC', 'a
sin': 'B0002E3CHC', 'helpful': '[2, 2]', 'overall': 5, 'date': 12990240000
00, 'reviews': "I've used D'Addario strings for many, many years without p
roblems - both acoustic & electric.  I've tried these coated strings in th
e past, and I actually prefer them to Elixir because they feel more like r
egular strings - the wrap is coated, rather than a sheathe being placed ov
er the entire string, like on Elixirs.Do they last longer?  Probably.  Eve
```

```
n when I'm in situations where I'm gigging regularly I rarely change strin
gs more than every few months, just because I don't see the need.  I think
the main benefit to coated strings is if you have guitars that you can't p
lay regularly the strings won't corrode.  But these sound great and feel f
ine, and the're the least expensive coated strings I've found.  I've alway
s found D'Addario quality to be great.Good strings", 'sentiment': 'Positiv
e', 'helpful_review': 1.0}}, {'_index': 'index_reviews_amazon', '_id': 'TB
CSrIEB5LSonSuUxQVC', '_score': 11.74545, '_source': {'reviewerID': 'A24Z0T
LWP8FXGV', 'asin': 'B000KIRT74', 'helpful': '[0, 0]', 'overall': 5, 'dat
e': 1395619200000, 'reviews': 'I love this pedal. It gives just the right
amount of crunch when needed. Typical Behringer pedal with plastic case bu
t a really good sound. I use this in my home studio so durability is not a
concern. Nice and quiet when not in use and can scream when you need it t
o. Highly recommended.Great little pedal', 'sentiment': 'Positive', 'helpf
ul_review': 0.0}}, {'_index': 'index_reviews_amazon', '_id': 'CBCSrIEB5LSo
nSuU1xp6', '_score': 11.617835, '_source': {'reviewerID': 'A1SD1C8XK3Z3V
1', 'asin': 'B00JBIVXGC', 'helpful': '[0, 0]', 'overall': 5, 'date': 14040
86400000, 'reviews': "I've used Elixirs for about five years now. This set
offers the same great sound for my acoustic as their electric strings prov
ide for my electric. Some people don't like the polymer (nanoweb) that coa
ts the string, but I find that it cuts back on string noise. Sliding your
fingers around doesn't cause too much noise, and with acoustic strings I f
ind that very important, especially if you're using a piezo pickup in your
instrument. Phosphor Bronze has always been a great alloy for acoustic gui
tars IMHO. These are great strings, worth the money, and sound great. They
also last 3 times longer than regular, non-coated strings in my experience
(been playing for 24yrs). Great product.Elixirs just sound good", 'sentime
nt': 'Positive', 'helpful_review': 0.0}}, {'_index': 'index_reviews_amazo
n', '_id': '8Q-SrIEB5LSonSuUsfz_', '_score': 11.4297695, '_source': {'revi
ewerID': 'A9P4Q7KX35RS2', 'asin': 'B0002H04NE', 'helpful': '[0, 0]', 'over
all': 5, 'date': 1390694400000, 'reviews': "I usually use the Pro Steels
(by D'Addario)  but these sound and fit nicely with my Squire Fender P Bas
s. Amazon seems to have great prices and it's nice to have them delivered
too.Great Pratice and Live Strings", 'sentiment': 'Positive', 'helpful_rev
iew': 0.0}}, {'_index': 'index_reviews_amazon', '_id': 'AQ-SrIEB5LSonSuUsf
3_', '_score': 11.4031105, '_source': {'reviewerID': 'A398X9POBHK69N', 'as
in': 'B0002H0A3S', 'helpful': '[0, 0]', 'overall': 5, 'date': 138317760000
0, 'reviews': "These were the same strings originally strung on my Godin S
eagull, and I absolutely love their sound, feel, and durability. The one r
eason I actually needed them because my G string broke while I was tuning
too fast. The sound is almost identical to what D'Addario describes, these
strings have a great, uplifting sound to them, hard to soft strums. Playin
g feels great with them, nothing out of the ordinary either strumming or t
he fretboard feel, clean and smooth. When replacing, I ran into no problem
s, quick and easy. I love the strings and their price, definitely will sti
ck with this brand as well.Sound Amazing On My Seagull, Top Pick For Strin
gs", 'sentiment': 'Positive', 'helpful_review': 0.0}}, {'_index': 'index_r
eviews_amazon', '_id': 'Cw-SrIEB5LSonSuUrvkU', '_score': 11.313062, '_sour
ce': {'reviewerID': 'A1LHMSY3Q46PJS', 'asin': 'B0002E3CHC', 'helpful':
'[5, 6]', 'overall': 5, 'date': 1197158400000, 'reviews': "I've been using
these strings for about eight years, and I also use them on my two Taylo
r's (814-CE & I use the mediums on my 710) I agree with the other reviewer
in saying they are the best sounding coated string out there. I still use
Martin strings once in awhile, but these strings last and improve in sound
as they get a few hours of play on them.Actually I own three other acousti
cs other than my two Taylor's and right now they're all strung with EXP's.
I know people want different sounds from their strings, but I find these t
o be bright with a great bottom on them, unlike the Elixirs which I find t
o be unbearably bright with a weak bottom end. These strings sound great w
ith great dynamics on them. Even on my Laguna Grand Symphony (slightly lar
ger than a Grand Concert) they pump out the lows and sparkle on the highs.
```

```
These on my Grand Auditorium do the same thing, they produce the sound I d
emand and make a real positive improvement over the other coated strings
I've used.The Mediums on my 710 have an endless bottom end and still have
great range on the mid's and treble.As I said I still use Martin Phosphor
Bronze strings once in awhile, but honestly these strings hold up extremel
y well and sound great. Darn near as near to an uncoated string that you'r
e going to find.Great sounding set of coated strings.", 'sentiment': 'Posi
tive', 'helpful_review': 0.83}}, {'_index': 'index_reviews_amazon', '_id':
'TQ-SrIEB5LSonSuUsf3_', '_score': 11.042872, '_source': {'reviewerID': 'A2
SKZZZBV1XYWP', 'asin': 'B0002H0A3S', 'helpful': '[8, 10]', 'overall': 5,
'date': 1268524800000, 'reviews': "Not much to say here. Strings are a ver
y personal thing to a guitar player. You need to try different gauges and
brands till you find the ones that work best for you and your style. These
are the string of choice for my acoustic and have been for many many year
s. I replace my strings about once a month with casual play and proper mai
ntenance so why run to the store or pay for shipping 10 times a year. The
strings are great, priced great and was on sale which made it even better.
There is less waste with the packaging which help the planet just a bit so
that's good too.Great Strings for Great Price", 'sentiment': 'Positive',
'helpful_review': 0.8}}]}})
```

In [246]:

```
like = [
    {
        '_index': INDEX_NAME,
        '_id': 'Cw-SrIEB5LSonSuUrvkU',
    }
]
classify_with_score(like)
```

Out[246]:

```
'Positive'
```

In [247]:

```
def classify_document(doc_id, size=10):

    return classify_with_score(
        [
            {
                '_index': INDEX_NAME,
                '_id': doc_id
            }
        ]
    )
```

In [248]:

```
classify_document('Cw-SrIEB5LSonSuUrvkU')
```

Out[248]:

```
'Positive'
```

Avaliando o "classificador"

In [251]:

```
test_docs_resp = client.search(index=INDEX_NAME, size=100, _source=['sentiment'])

test_docs = pd.DataFrame(
        {
            '_id': x['_id'],
            **x['_source']
        } for x in test_docs_resp['hits']['hits']
    )
test_docs
```

Out[251]:

	_id	sentiment
0	_g-SrIEB5LSonSuUp_Gk	Positive
1	_w-SrIEB5LSonSuUp_Gk	Positive
2	AA-SrIEB5LSonSuUp_Kk	Positive
3	Ag-SrIEB5LSonSuUp_Kk	Positive
4	BA-SrIEB5LSonSuUp_Kk	Positive
...	...	...
95	KQ-SrIEB5LSonSuUp_Ok	Positive
96	Lg-SrIEB5LSonSuUp_Ok	Neutral
97	Mg-SrIEB5LSonSuUp_Ok	Positive
98	Ng-SrIEB5LSonSuUp_Ok	Positive
99	Nw-SrIEB5LSonSuUp_Ok	Positive

100 rows × 2 columns

In [253]:

```
test_docs['predicted'] = test_docs['_id'].apply(classify_document)

test_docs
```

Out[253]:

	_id	sentiment	predicted
0	_g-SrIEB5LSonSuUp_Gk	Positive	Negative
1	_w-SrIEB5LSonSuUp_Gk	Positive	Negative
2	AA-SrIEB5LSonSuUp_Kk	Positive	Positive
3	Ag-SrIEB5LSonSuUp_Kk	Positive	Positive
4	BA-SrIEB5LSonSuUp_Kk	Positive	Positive
...	...	...	...
95	KQ-SrIEB5LSonSuUp_Ok	Positive	Neutral
96	Lg-SrIEB5LSonSuUp_Ok	Neutral	Negative
97	Mg-SrIEB5LSonSuUp_Ok	Positive	Positive
98	Ng-SrIEB5LSonSuUp_Ok	Positive	Negative
99	Nw-SrIEB5LSonSuUp_Ok	Positive	Neutral

100 rows × 3 columns

In [256]:

```
print(
    classification_report(
        test_docs['sentiment'],
        test_docs['predicted']
    )
)
```

```
              precision    recall  f1-score   support

    Negative       0.03      0.50      0.05         2
     Neutral       0.00      0.00      0.00         8
    Positive       0.89      0.38      0.53        90

    accuracy                           0.35       100
   macro avg       0.31      0.29      0.19       100
weighted avg       0.81      0.35      0.48       100
```