1. gettingDataset

October 18, 2017

```
In [6]: from pattern.web import Twitter, plaintext
        from pattern.en import tag
        import re

        twitter = Twitter()
        file = open('datasetG.txt', 'w+')

        for tweet in twitter.search("Machine Learning", cached=False, count=100):
            s = plaintext(tweet.text.lower())
            print("Tweet: " + s + "\n")

            # Remove the incomplet words of the end.
            s = re.sub(ur'[^ ]*\u2026','', s)

            # Remove the line breacks.
            s = re.sub(r'\n', r' ', s)

            # Remove the URLs
            s = re.sub(r'https:\/\/t.co\/([a-z]|[0-9]){10}', r'', s)

            # Remove user names
            s = re.sub(r'@[^ ]*', r'', s)

            # Remove "rt"
            s = re.sub(r'rt', r'', s)

            s = re.sub(r'#', r'', s)

            # Replace successive spaces with a single space
            s = re.sub(r' +', r' ', s)

            # Remove all the characters that aren from ASCII code.
            s = re.sub(r'[^\x00-\x7f]',r'', s)

            print("Tweet filtrado: " + s + "\n")
            file.write(s + "\n")

Tweet: east africa: machine learning, cybersecurity take centre stage at #oow17 - https://t.co/
```

1

Tweet filtrado: east africa: machine learning, cybersecurity take centre stage at oow17 - afri

Tweet: oracles new database uses machine learning to automate administration https://t.co/ytaq

Tweet filtrado: oracles new database uses machine learning to automate administration

Tweet: rt @acutusllp: want to know more about #machinelearning and #ai? by @drnatalie https://t

Tweet filtrado: want to know more about machinelearning and ai? by via bigdata

Tweet: @gsohn ich lese gerade bücher über machine learning. kann mann auf alles anwenden, was

Tweet filtrado: ich lese gerade bcher ber machine learning. kann mann auf alles anwenden, was

Tweet: #fenergo cto niall twomey discussing how automation & machine learning can improve the

Tweet filtrado: fenergo cto niall twomey discussing how automation & machine learning can impro

Tweet: building a $60 billion data model to stop us healthcare fraud (part 1) https://t.co/ce7

Tweet filtrado: building a $60 billion data model to stop us healthcare fraud (pa 1)

Tweet: rt @bobehayes: 10 best #artificialintelligence (#ai) and #machinelearning tools (ml) fo

Tweet filtrado: 10 best aificialintelligence (ai) and machinelearning tools (ml) for you

Tweet: rt @bernardmarr: 5 #myths about #artificial #intelligence (#ai) you must stop believing

https://t.co/5wema6hin3 #transform #machine #learn

Tweet filtrado: 5 myths about aificial intelligence (ai) you must stop believing transform mac

Tweet: @dd_jessica_ how to derive actionable insights from #rawdata and then use ai to optimize

Tweet filtrado: how to derive actionable insights from rawdata and then use ai to optimize yo

Tweet: google, mit researchers creat
https://t.co/sf2cwkywdn

Tweet filtrado: google, mit researchers creat

Tweet: google, mit researchers creat
https://t.co/povukqxw7z

Tweet filtrado: google, mit researchers creat

Tweet: rt @bobehayes: 27 great articles about #machinelearning #algorithms https://t.co/umfgtp

Tweet filtrado: 27 great aicles about machinelearning algorithms

Tweet: rt @cvcnd: 10 best artificial intelligence (ai) and machine learning tools (ml) for you

Tweet filtrado: 10 best aificial intelligence (ai) and machine learning tools (ml) for you

Tweet: machine learning and ai marketing trends to watch for 2018 | marketing insider group htt

Tweet filtrado: machine learning and ai marketing trends to watch for 2018 | marketing insider

Tweet: rt @daverubal: how to build and deploy scalable #machinelearning with #apachekafka | #bi

Tweet filtrado: how to build and deploy scalable machinelearning with apachekafka | bigdata ml

Tweet: machine learning engineer https://t.co/2jypw5a3ro

Tweet filtrado: machine learning engineer

Tweet: rt @ronald_vanloon: the mathematics of machine learning | #datascience #machinelearning

Tweet filtrado: the mathematics of machine learning | datascience machinelearning

Tweet: 3 machine learning algorithms you need to know - dzone ai https://t.co/selctzfeds

Tweet filtrado: 3 machine learning algorithms you need to know - dzone ai

Tweet: rt @cambwireless: chairman of @prowler_io & prof of machine learning @cambridge_uni carl

Tweet filtrado: chairman of & prof of machine learning carl edward rasmussen is here talking a

Tweet: when #machine learning benefits #employees and #customers alike https://t.co/rcmkr9nuyg

Tweet filtrado: when machine learning benefits employees and customers alike via

Tweet: rt @daverubal: how to build and deploy scalable #machinelearning with #apachekafka | #bi

Tweet filtrado: how to build and deploy scalable machinelearning with apachekafka | bigdata ml

Tweet: splunk execs: partners taking machine learning into new iot, industrial applications htt

Tweet filtrado: splunk execs: paners taking machine learning into new iot, industrial applicati

Tweet: @tomroud t'as pas vraiment le choix que d'aller faire du machine learning. avec des pers

Tweet filtrado: t'as pas vraiment le choix que d'aller faire du machine learning. avec des pe

Tweet: rt @markthoma: the usefulnessand possible dangersof machine learning - regulatory reviev

3

Tweet filtrado: the usefulnessand possible dangersof machine learning - regulatory review

Tweet: rt @dd_wen_: introduction to machine learning with python: a guide for data scientists

Tweet filtrado: introduction to machine learning with python: a guide for data scientists data

Tweet: rt @lorelai66: how machine learning and big data are helping chemists search the vast c

Tweet filtrado: how machine learning and big data are helping chemists search the vast chemica

Tweet: https://t.co/ojnvknkijw

Tweet filtrado:

Tweet: rt @cambwireless: chairman of @prowler_io & prof of machine learning @cambridge_uni carl

Tweet filtrado: chairman of & prof of machine learning carl edward rasmussen is here talking a

Tweet: rt @petioteric: [#tech]
10 best #artificialintelligence and #machinelearning tools for you
https://t.co/3t4j9nmwzl

#ai #ml #cloud
#bigda

Tweet filtrado: [tech] 10 best aificialintelligence and machinelearning tools for you ai ml

Tweet: rt @bernardmarr: 5 #myths about #artificial #intelligence (#ai) you must stop believing

https://t.co/5wema6hin3 #transform #machine #learn

Tweet filtrado: 5 myths about aificial intelligence (ai) you must stop believing transform mac

Tweet: rt @genusllc: #ai is changing #devops enterprise architecture. here is a basic explanati

Tweet filtrado: ai is changing devops enterprise architecture. here is a basic explanation of

Tweet: software engineer, computer vision and machine learning - osmo - palo alto, ca - https:/

Tweet filtrado: software engineer, computer vision and machine learning - osmo - palo alto, ca

Tweet: researchers just trained a robot to use nunchucks using trial and error: researchers fr

Tweet filtrado: researchers just trained a robot to use nunchucks using trial and error: resea

Tweet: researchers just trained a robot to use nunchucks using trial and error: researchers fr

Tweet filtrado: researchers just trained a robot to use nunchucks using trial and error: resea

Tweet: rt @ronald_vanloon: a machine learning workflow | #machinelearning #artificialintellige

Tweet filtrado: a machine learning workflow | machinelearning aificialintelligence

Tweet: shipping forecast: cloudy with a chance of machine learning - cbr - https://t.co/vvg9mt

Tweet filtrado: shipping forecast: cloudy with a chance of machine learning - cbr -

Tweet: rt @creative_punch: artificial dataset for machine learning with #python and #numpy / #
https://t.co/y94orwsvgb

#programming #mac

Tweet filtrado: aificial dataset for machine learning with python and numpy / theano --- prog

Tweet: rt @daverubal: how to build and deploy scalable #machinelearning with #apachekafka | #b

Tweet filtrado: how to build and deploy scalable machinelearning with apachekafka | bigdata m

Tweet: rt @miketamir: cartoon: future #machinelearning class https://t.co/wjgw46bkle #ai #deep

Tweet filtrado: caoon: future machinelearning class ai deeplearning datascience

Tweet: take a course ft. latest in tech, machine learning, ai at #treehousetx member @divergen

Tweet filtrado: take a course ft. latest in tech, machine learning, ai at treehousetx member -

Tweet: how machine learning and artificial #intelligence can transform your #sales department.

Tweet filtrado: how machine learning and aificial intelligence can transform your sales depame

Tweet: how machine learning and artificial #intelligence can transform your #sales department.

Tweet filtrado: how machine learning and aificial intelligence can transform your sales depame

Tweet: #confmlearn conférence #orsys : tout savoir sur le machine learning ça se passe dans #l

Tweet filtrado: confmlearn confrence orsys : tout savoir sur le machine learning a se passe da

Tweet: the world is changing rapidly and so do tax and revenue agencies. in tax we use machine

Tweet filtrado: the world is changing rapidly and so do tax and revenue agencies. in tax we us

Tweet: rt @deep_in_depth: jpmorgan's massiv guide to machine learning jobs in finance https://

Tweet filtrado: jpmorgan's massiv guide to machine learning jobs in finance deeplearning mach

Tweet: rt @deep_in_depth: jpmorgan's massiv guide to machine learning jobs in finance https://t

Tweet filtrado: jpmorgan's massiv guide to machine learning jobs in finance deeplearning mach

Tweet: rt @deep_in_depth: jpmorgan's massiv guide to machine learning jobs in finance https://t

Tweet filtrado: jpmorgan's massiv guide to machine learning jobs in finance deeplearning mach

Tweet: googlelanguager analysing language through the google cloud machine learning apis https

Tweet filtrado: googlelanguager analysing language through the google cloud machine learning a

Tweet: rt @spirosmargaris: essential #cheatsheets for #machinelearning & #deeplearning enginee

https://t.co/knxbgxud4d #fintech @deeplearn007

Tweet filtrado: essential cheatsheets for machinelearning & deeplearning engineers fintech

Tweet: rt @spirosmargaris: essential #cheatsheets for #machinelearning & #deeplearning enginee

https://t.co/knxbgxud4d #fintech @deeplearn007

Tweet filtrado: essential cheatsheets for machinelearning & deeplearning engineers fintech

Tweet: rt @v_vashishta: #opensource is leading #machinelearning https://t.co/uljbxs14jl

Tweet filtrado: opensource is leading machinelearning

Tweet: @harryaevans @connectedchris yes we used machine learning in health insurance data and

Tweet filtrado: yes we used machine learning in health insurance data and determine the predi

Tweet: researchers just trained a robot to use nunchucks using trial and error https://t.co/x3

Tweet filtrado: researchers just trained a robot to use nunchucks using trial and error via

Tweet: rt @homeaiinfo: data science machine learning – data science within machine le https:/

Tweet filtrado: data science machine learning – data science within machine le machine-learn

Tweet: how can machine learning affect your organizational data strategy? – dataversity https:

Tweet filtrado: how can machine learning affect your organizational data strategy? – dataversit

Tweet: researchers just trained a robot to use nunchucks using trial and error #machinelearning

Tweet filtrado: researchers just trained a robot to use nunchucks using trial and error machine

Tweet: rt @code_likeagirl: machine learning for pms by @dozenrose on #codelikeagirl https://t.c

Tweet filtrado: machine learning for pms by on codelikeagirl machinelearning ai tensorflow

Tweet: rt @denodo: rt @kirkdborne diving deep into whats new with #azure #machinelearning: http

Tweet filtrado: diving deep into whats new with azure machinelearning: bigdata datascience ai

Tweet: rt @denodo: rt @kirkdborne diving deep into whats new with #azure #machinelearning: http

Tweet filtrado: diving deep into whats new with azure machinelearning: bigdata datascience ai

Tweet: rt @burnout72: want to know more about #machinelearning and #ai? https://t.co/wrhvspruu

Tweet filtrado: want to know more about machinelearning and ai? aificialintelligence neuralnet

Tweet: rt @kirkdborne diving deep into whats new with #azure #machinelearning: https://t.co/vzu

Tweet filtrado: diving deep into whats new with azure machinelearning:

Tweet: rt @v_vashishta: #opensource is leading #machinelearning https://t.co/uljbxs14jl

Tweet filtrado: opensource is leading machinelearning

Tweet: rt @homeaiinfo: data science machine learning - data science within machine le https:/

Tweet filtrado: data science machine learning - data science within machine le machine-learni

Tweet: steve @steve_db2 during @idugdb2 in lisbon talking about machine learning & ai in a pack

Tweet filtrado: steve during in lisbon talking about machine learning & ai in a packed room!!!

Tweet: rt @msexcel: create forecasting models using excel and #machinelearning. learn how: http

Tweet filtrado: create forecasting models using excel and machinelearning. learn how:

Tweet: rt @homeaiinfo: data science machine learning - data science within machine le https:/

Tweet filtrado: data science machine learning - data science within machine le machine-learni

Tweet: rt @ronald_vanloon: data science, machine learning, bi explained in a amazing few pictu

Tweet filtrado: data science, machine learning, bi explained in a amazing few pictures | datas

Tweet: rt @userkoded: hey machine learning hub thanks for following me. id love to help with yo

Tweet filtrado: hey machine learning hub thanks for following me. id love to help with your re

Tweet: software-based #marketingautomation is a key to the search marketing of tomorrow read t

Tweet filtrado: software-based marketingautomation is a key to the search marketing of tomorrow

Tweet: rt @userkoded: hey machine learning hub thanks for following me. id love to help with yo

Tweet filtrado: hey machine learning hub thanks for following me. id love to help with your re

Tweet: databricks 'lays foundations' for machine learning (sorry) - open source insider https:/

Tweet filtrado: databricks 'lays foundations' for machine learning (sorry) - open source inside

Tweet: markov models: definitive guide to understanding unsupervised machine learning in py htt

Tweet filtrado: markov models: definitive guide to understanding unsupervised machine learning

Tweet: an introduction to statistical learning with applications in r https://t.co/huxtsqvi84

Tweet filtrado: an introduction to statistical learning with applications in r machine learning

Tweet: rt @webcodegeeks: machine learning for the modern web developer - web dev https://t.co/s

Tweet filtrado: machine learning for the modern web developer - web dev

Tweet: rt @dflazaro: machine learning, #iot bring big changes to #data management systems https

Tweet filtrado: machine learning, iot bring big changes to data management systems ai

Tweet: rt @ardentisys: external article
#machine learning, #iot bring big changes to #datamanagement systems
https://t.co/gcrbe93k4q

Tweet filtrado: external aicle machine learning, iot bring big changes to datamanagement syste

Tweet: machine learning in cyber security domain #bgamakale
https://t.co/xnfi5hwdij

Tweet filtrado: machine learning in cyber security domain bgamakale

Tweet: rt @lorelai66: how machine learning and big data are helping chemists search the vast ch

Tweet filtrado: how machine learning and big data are helping chemists search the vast chemica

Tweet: rt @floydram: strata 2017 postmortem: more virtual data lake, more operational machine l

Tweet filtrado: strata 2017 postmoem: more viual data lake, more operational machine learning

Tweet: rt @floydram: strata 2017 postmortem: more virtual data lake, more operational machine l

Tweet filtrado: strata 2017 postmoem: more viual data lake, more operational machine learning

Tweet: rt @lorelai66: how machine learning and big data are helping chemists search the vast ch

Tweet filtrado: how machine learning and big data are helping chemists search the vast chemica

Tweet: tech giants including baidu and google spent between $20b to $30b on ai in 2016 #ai #ma

Tweet filtrado: tech giants including baidu and google spent between $20b to $30b on ai in 201

Tweet: strata 2017 postmortem: more virtual data lake, more operational machine learning | zdne

Tweet filtrado: strata 2017 postmoem: more viual data lake, more operational machine learning

Tweet: rt @corey_c: 3 reasons why its crucial for brands to introduce #machinelearning into the

Tweet filtrado: 3 reasons why its crucial for brands to introduce machinelearning into their r

Tweet: how unity is tackling the crossroads between machine learning and gaming

https://t.co/ldzqlp9n1h

Tweet filtrado: how unity is tackling the crossroads between machine learning and gaming

Tweet: rt @v_vashishta: #opensource is leading #machinelearning https://t.co/uljbxs14jl

Tweet filtrado: opensource is leading machinelearning

Tweet: rt @deep_in_depth: jpmorgan's massiv guide to machine learning jobs in finance https://t

Tweet filtrado: jpmorgan's massiv guide to machine learning jobs in finance deeplearning machi

Tweet: this morning has been progress bars in machine learning jupyter notebooks because machi

Tweet filtrado: this morning has been progress bars in machine learning jupyter notebooks becau

Tweet: rt @deep_in_depth: jpmorgan's massiv guide to machine learning jobs in finance https://t

Tweet filtrado: jpmorgan's massiv guide to machine learning jobs in finance deeplearning machi

Tweet: rt @v_vashishta: #opensource is leading #machinelearning https://t.co/uljbxs14jl

Tweet filtrado: opensource is leading machinelearning

Tweet: researchers just trained a robot to use nunchucks using trial and error - business insi

Tweet filtrado: researchers just trained a robot to use nunchucks using trial and error - busi

Tweet: rt @v_vashishta: #opensource is leading #machinelearning https://t.co/uljbxs14jl

Tweet filtrado: opensource is leading machinelearning

Tweet: what's the difference between #ai, #machinelearning, and #deeplearning? | nvidia blog h

Tweet filtrado: what's the difference between ai, machinelearning, and deeplearning? | nvidia l

In [ ]:

In [ ]: