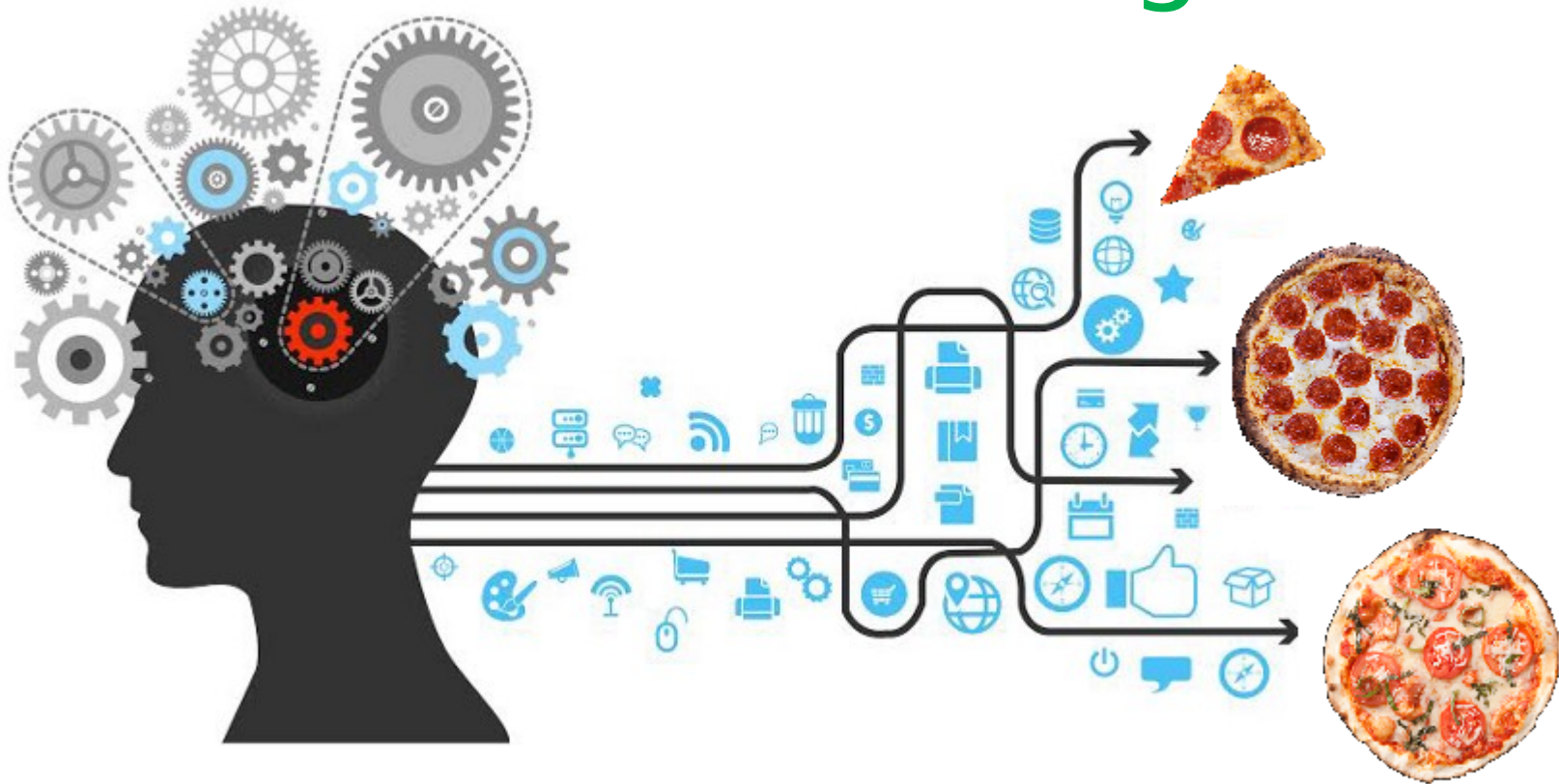
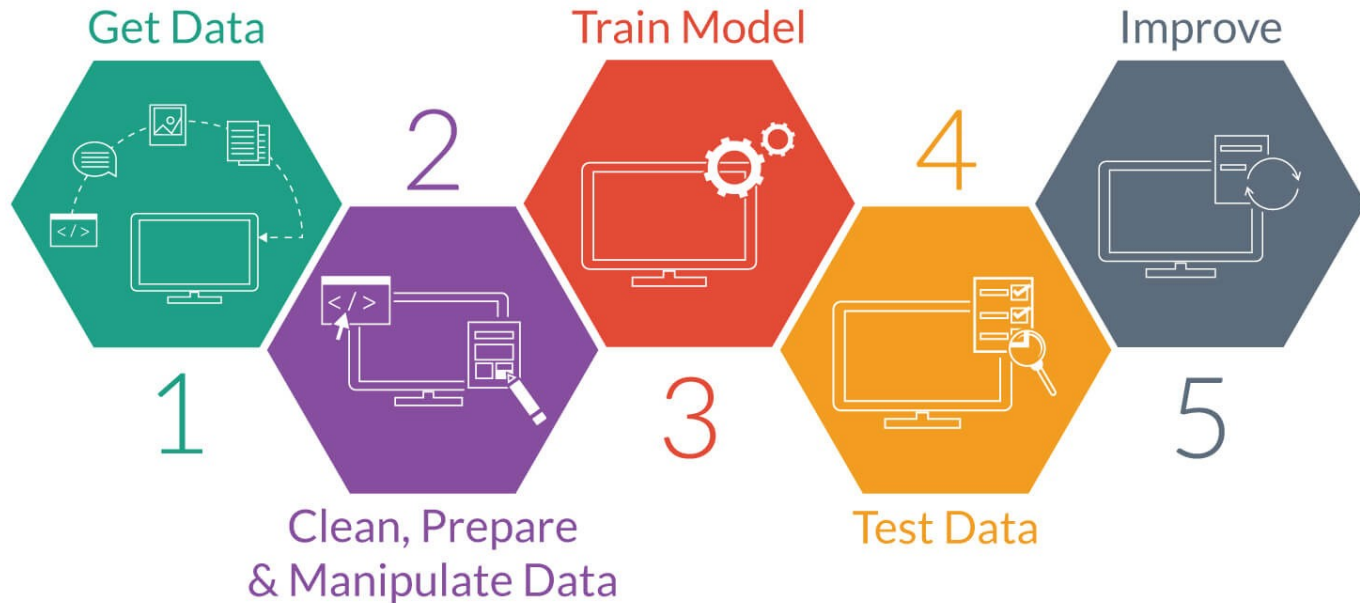


Pizza Learning 2



Categorização textual

Passos do ML



1 - Dados Brutos

```
1 store_order key,field3,processor_status,send,time,authorization_time,capture_time,customer_name,customer_gender,customer_document_t
2 ype,customer_document_number,customer_dob,customer_phone_1,customer_phone_2,field35,customer_email,field17,field18,country,state,ci
3 ty,field22,address,product_type,field25,other_status,field27,field28,field29,field30,card_flag,card_number,cardholder_name,field33,
4 field34
5 ConnectCar - Texto,FEAEDB96-4AF4-4AFF-872A-E75DA6149B37,243618,Captured,2016-08-10 20:52:19.187,2016-08-10 20:52:19.327,2016-08-10
6 20:55:59.360,DANILLO FERREIRA,M,CPP,77223335058,1988-01-01
7 00:00:00.000,(11)945450456,,,flfo@eonee.com,Personal,Brazil,SP,0,,,"Residencial": País: Brazil, Estado: SP, Cidade: São Paulo,
8 Bairro: Vila São Luís(2008 Oeste), Rua Visconde do Bom Retiro, 654,, 05382-060,"ItemReference":Recarga 50, Nome:
9 ConnectCar - Q,7400,Approved,0,APA,50,0,Visa,2300800***0000,FELIPE DIEGO,FALSO,"Residencial": País: Brazil,,
10 03:58:46.000,ANDUEIA GOMALVES,E,CPP,71870793934,1977-01-13
11 ConnectCar - Texto,B0F8F949-5AAD-4978-8294-F613BF7F8057,958313,Captured,2016-08-25 03:45:41.177,2016-08-25 03:45:41.273,2016-08-25
12 00:00:00.000,(09)222222222,(99)33333333,,71878793934,connectcar.com,Personal,Brazil,PR,Curitiba,,,"Residencial": País: Brazil,
13 Estado: PR, Cidade: Curitiba, Bairro: Jardim Botânico, Avenida Comandador Franco, 415, MAPINE SEGUROS,
14 80215-090,"ItemReference":Recarga 30, Nome: ConnectCar,370.400,Approved,0,APA,30,0,Visa,2300800***0000,ROBERTO
15 SATYNO,FALSO,"Residencial": País: Brazil,,
16 ConnectCar - Texto,CE09CA29-ED3C-4B5A-B5F5-FE955334D7F2,179761,AuthorizedPendingCapture,2016-08-25 05:07:25.137,2016-08-25
17 05:07:25.197,Teste betone,M,CPP,91205398821,1978-02-17 00:00:00.000,(11)345678968,,,gtyee.com,Personal,Brazil,SP,São
18 Paulo,,,"Residencial": País: Brazil, Estado: SP, Cidade: São Paulo, Bairro: Mooca, Rua Dias Leme, 68,,
19 03118-040,"ItemReference":Recarga Online 150, Nome: ConnectCar,343.300,,,150,0,Visa,2300800***0000,FELIPE
20 DIEGO,FALSO,"Residencial": País: Brazil,,
21 ConnectCar - E-commerce,C35548CF-5D5C-418D-96B8-2DA7FFCF8393,45760244318674,Refunded,2016-12-22 11:16:15.477,2016-12-22
22 11:16:16.400,2016-12-22 11:16:16.800,Gabriel,M,CPP,2757984854,1979-06-10
23 00:00:00.000,997146234,,gcalin@uol.com.br,Personal,Brazil,SP,São Carlos,,,"Residencial": País: Brazil, Estado: SP, Cidade: São
24 Carlos, Bairro: Centro, Sete de Setembro, 2656,, 13568101,,,50,1,Mastercard,5162208***7837,GABRIEL CALIN,FALSO,"Residencial":
25 País: Brazil,,
26 ConnectCar - Recarga Cartão Crédito,00B48B5D-74E5-4547-8652-08023697BAA4,107702830006543,Captured,2017-01-07 10:32:24.488,2017-01-07
27 10:32:25.297,2017-01-07 10:35:39.237,aletheia de lima santos correa,F,CPP,31147153892,1983-02-21
28 00:00:00.000,(19)088486813,(19)21212082,,aletheia.correa@uol.com.br,Personal,Brazil,SP,Campinas,,,"Residencial": País: Brazil,
29 Estado: SP, Cidade: Campinas, Bairro: Cidade Universitária, Rua Professor Achille Bassi, 112,, 13083-530,"ItemReference":Recarga
30 50, Nome: ConnectCar,11.21,Approved,0,APA,200,0,Mastercard,5155900***2257,LUIZ C JUNIOR,FALSO,"Residencial": País: Brazil,,
31 ConnectCar - Recarga Cartão Crédito,077A42BE-F807-4ABA-9266-0007E17A16BA,1667831640244,Captured,2017-03-07 17:01:43.723,2017-03-07
32 17:01:44.803,2017-03-07 17:05:40.993,Luciano Alves Rego,M,CPP,91268737619,1973-03-22
33 00:00:00.000,(63)00041919,,lucianoar19@yahoo.com.br,Personal,Brazil,DF,Brasília,,,"Residencial": País: Brazil, Estado: DF, Cidade:
34 Brasília, Bairro: Asa Norte, Quadra SON 31 Bloco B, 1, Apt 100, 70757-020,"ItemReference":Recarga 100, Nome:
35 ConnectCar,2.97,Approved,0,APA,200,0,Mastercard,55530,,,LUCIANO ALVES REGO,FALSO,"Residencial": País: Brazil,,
36 ConnectCar - Recarga Cartão Crédito,0141756-F801-4187-8269-000A19979209,1570390749268,Captured,2017-02-26 22:52:34.598,2017-02-26
37 22:52:35.520,2017-02-26 22:55:48.157,NOEL DE SOUSA MOTA,M,CPP,5481721880,1964-06-23
38 00:00:00.000,(11)943889931,(11)32471089,,noelmota@uol.com,Personal,Brazil,SP,Tauboão da Serra,,,"Residencial": País: Brazil,
39 Estado: SP, Cidade: Tauboão da Serra, Bairro: Parque Taubão, Estrada São Francisco, 1592, bloco 3 apto. 122,
40 06245-000,"ItemReference":Recarga 50, Nome: ConnectCar,2.0,Approved,0,APA,50,0,Visa,4444444***4040,NOE DE SOUSA
41 MOTA,FALSO,"Residencial": País: Brazil,,
42 ConnectCar - Recarga Cartão Crédito,F18B6BD-F671-4208-8503-000B2E5D61C,35660154450331,Captured,2016-09-12 02:06:51.083,2016-09-12
43 02:06:52.217,2016-09-12 02:10:51.323,Helton Dias Rodrigues,M,CPP,9664507698,1989-12-05
44 00:00:00.000,(11)948957481,,,helton.dias.rodrigues@gmail.com,Personal,Brazil,SP,São Paulo,,,"Residencial": País: Brazil, Estado:
45 SP, Cidade: São Paulo, Bairro: Jardim Aeroporto, Rua Tapas, 375,, 04631-810,"ItemReference":Recarga 50, Nome:
46 ConnectCar,2.94,Approved,0,APA,50,0,Mastercard,5162208***5258,HELTON D RODRIGUES,FALSO,"Residencial": País: Brazil,,
47 ConnectCar - Recarga Cartão Crédito,48A929-8076-4243-88D0-000A5B5A1D43,1627631372988,Captured,2017-03-03 02:05:46.298,2017-03-03
48 02:05:47.017,2017-03-03 02:10:45.127,leandro Henrique da silva,M,CPP,4104569933,1984-01-19
49 00:00:00.000,(43)000490337,(43)03154517,,leandrobella@uol.com,Personal,Brazil,PE,Rolândia,,,"Residencial": País: Brazil,
50 Estado: PE, Cidade: Rolândia, Bairro: Jardim Nobre, Rua Pedro Zorzetti, 190,, 86602-508,"ItemReference":Recarga 50, Nome:
51 ConnectCar,2.94,Approved,0,APA,50,0,Mastercard,5300000***7002,LEANDRO DA SILVA,FALSO,"Residencial": País: Brazil,,
52 ConnectCar - Recarga Cartão Crédito,7187A8AA-797C-4384-A738-000F7C0CE6AD,15070302237973,Captured,2017-02-10 19:46:01.257,2017-02-10
53 19:46:02.403,2017-02-10 19:50:30.933,celso Eduardo,M,CPP,7394919706,1976-04-11
54 00:00:00.000,(21)070154602,,,eduardocelso@hotmail.com,Personal,Brazil,RJ,São Gonçalo,,,"Residencial": País: Brazil, Estado: RJ,
55 Cidade: São Gonçalo, Bairro: Trindade, Fortaleza, 01, Casa 1, 24156-140,"ItemReference":Recarga 50, Nome:
```

Amaz... Pizza! x Untitl... Refer... 1.9.Na... fsklear... D TEXTC... x Nalve... x machil... x Steps... x Melhor... x

rocinema.com/filmes/todos/filmes/shotas-espectadores/

er M: Docker & Rails | Swagger Editor | Customizing RS... Mail-Itakuraku | EC2 Managemen... Simility Intelligence

ADOROCINEMA

Ex - O Hábit 2, Crisanta Tios de Circo, Divergente, Cante...

FILMES PROGRAMAÇÃO TRAILERS NOTÍCIAS SÉRIES FILMES ONLINE

MEU ADOROCINEMA

Em cartaz Em breve Críticas Bilheterias Todos os filmes

GO FROM THIS... TO A BELLROY

Incio Filmes em cartaz Melhores filmes em cartaz Filmes mais esperados Top de Filmes

Melhores filmes segundo os espectadores

Melhores filmes Piores filmes Para crianças

Pesquisa avançada Organizar por Notas dos leitores

Todos os gêneros

Ação (254)

Animação (959)

Artes Marcadas (93)

Aventura (1037)

Biografia (977)

Bollywood (3)

Classico (1)

Comédia (515)

Comédia dramática (1823)

Comédia Musical (248)

Crime (255)

1. O Poderoso Chefão

Lançamento 24031972 (chômon)

De Francis Ford Coppola

Com Marlon Brando, Al Pacino

Gênero Policial, Drama

Don Vito Corleone (Marlon Brando) é o chefe de uma "família" de Nova York que está lá, pois Corleone (Al Pacino), sua filha, se casou com Carlo...

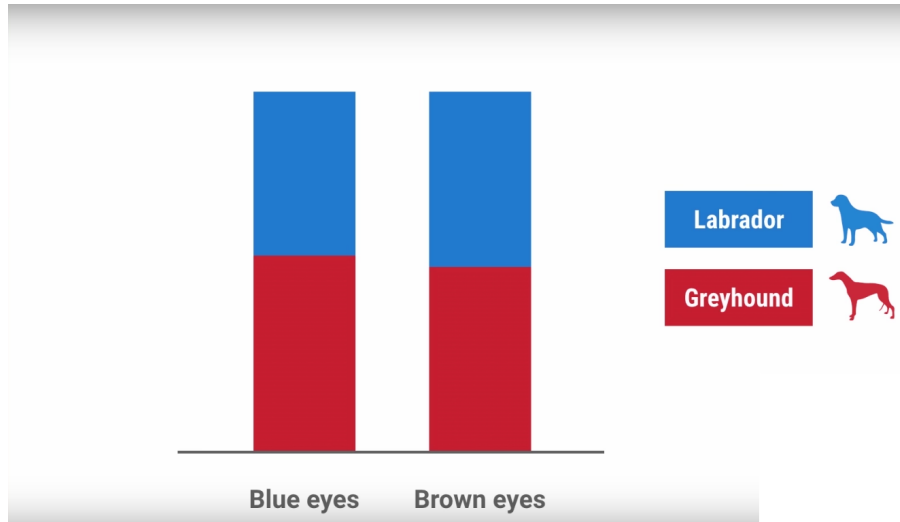
AdoroCinema

Imprensa

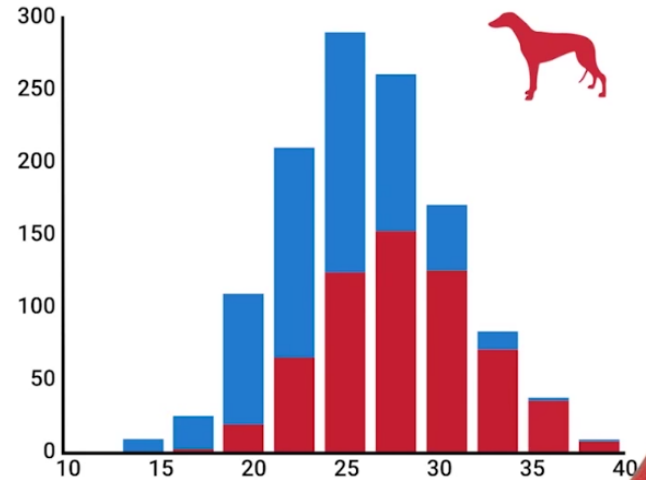
a partir de **R\$ 84**



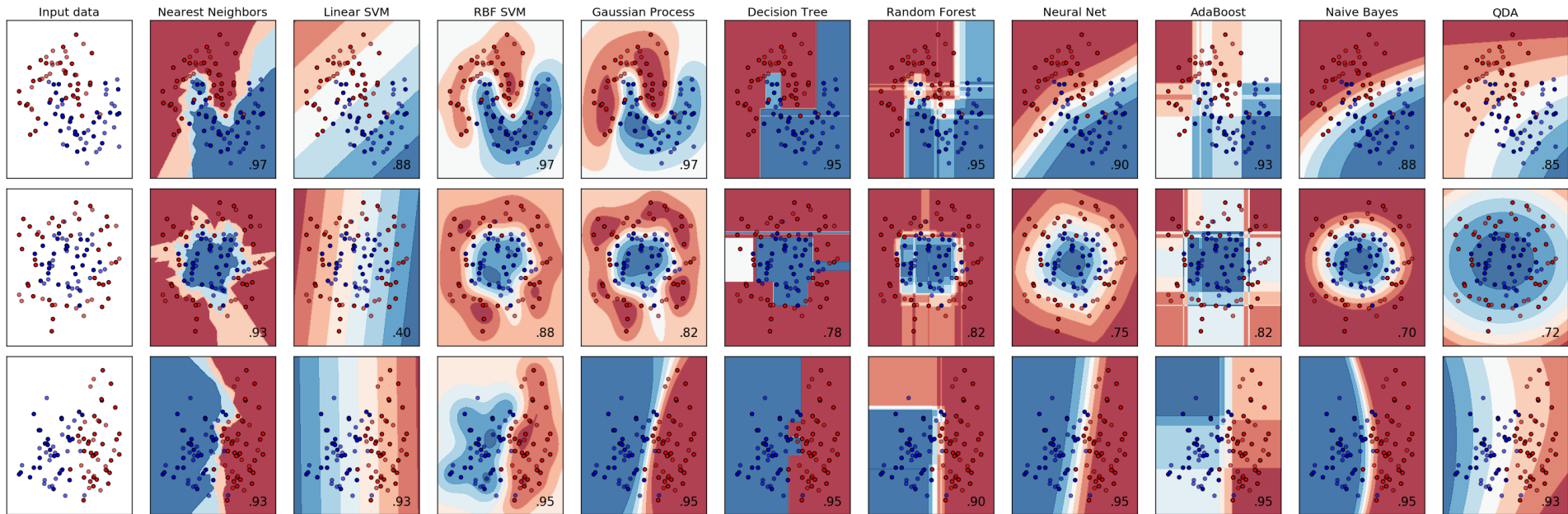
2 - Análise dos Dados



About 95% of dogs at this height are greyhounds.



3 - Treinar o Modelo



4 - Testar o Modelo

```
# preprocess dataset, split into training and test part  
X, y = ds  
X = StandardScaler().fit_transform(X)  
X_train, X_test, y_train, y_test = \  
    train_test_split(X, y, test_size=.4, random_state=42)
```

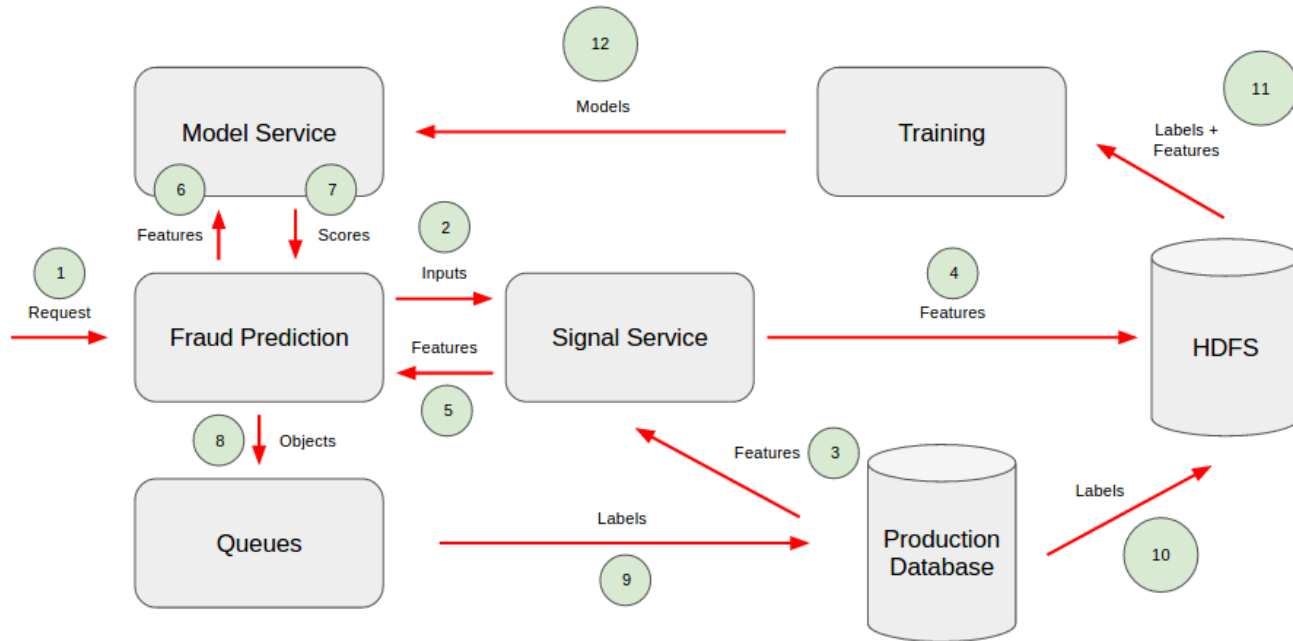
Accuracy: (número de classificações certas) / (total de amostras)

Recall: (positivos) / (positivos + falsos negativos)

Accuracy pode não ser o melhor indicador de performance.

Vide classificação de fraude

5 - Produção



Problema

De acordo com a descrição de um determinado filme, quero classificá-lo em dentro de um set de gêneros pré-definidos utilizando machine learning

Bayes

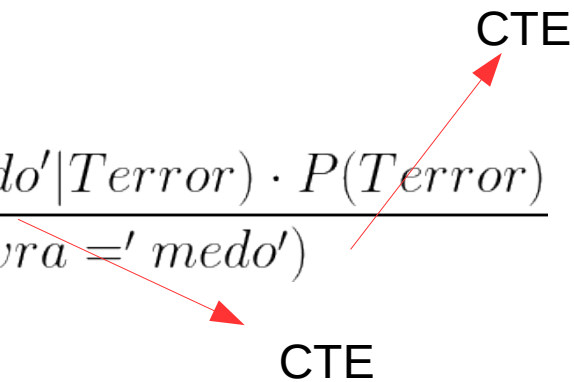
$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$



Bayes

$$P(Terror|palavra = 'medo') = \frac{P(palavra = 'medo'|Terror) \cdot P(Terror)}{P(palavra = 'medo')}$$

Bayes

$$P(Terror|palavra = 'medo') = \frac{P(palavra = 'medo'|Terror) \cdot P(Terror)}{P(palavra = 'medo')}$$


CTE

CTE

Naive Bayes

$$P(Terror|palavra = 'medo') = P('medo'|Terror) \cdot P('susto'|Terror)$$

Dados do Problema

ADOROCINEMA

FILMES

PROGRAMAÇÃO

TRAILERS

NOTÍCIAS

SÉRIES

FILMES ONLINE

Em cartaz

Em breve

Críticas

Bilheterias

Todos os filmes

Todos os gêneros >

Ação (2524)

Animação (959)

Artes Marciais (93)

Aventura (1907)

Biografia (977)

Bollywood (3)

Clássico (1)

Comédia (5915)

Comédia dramática (1823)

Comédia Musical (348)

Crime (295)

Desenho Animado (12)

Documentário (3956)

Doramas (2)

Drama (11612)

Épico (30)

Erótico (131)

Espetáculo (2)

Espionagem (125)

Esporte (142)

Experimental (132)

Família (1097)

Fantasia (1309)


Faroeste (471)

Ficção científica (1004)

Guerra (605)

Histórico (759)

Movie night (2)



1. O Poderoso Chefão

Lançamento 24/03/1972 (2h55min)

De [Francis Ford Coppola](#)

Com [Marlon Brando](#), [Al Pacino](#)

Gênero [Policial](#), [Drama](#)

Don Vito Corleone (Marlon Brando) é o chefe de uma "família" de Nova York que está feliz, pois Connie (Talia Shire), sua filha, se casou com Carlo...

AdoroCinema


Imprensa

[Leitores](#)

4,8

Ver o trailer

Sessões



2. A Lista de Schindler

Lançamento 31/12/1993 (3h15min)

De [Steven Spielberg](#)

Com [Liam Neeson](#), [Ben Kingsley](#)

Gênero [Histórico](#), [Drama](#), [Guerra](#), [Biografia](#)

A inusitada história de Oskar Schindler (Liam Neeson), um sujeito oportunista, sedutor, "armador", simpático, comerciante no mercado negro, mas,...

AdoroCinema

Imprensa

[Leitores](#)

4,8

Ver o trailer

Sessões

Dados do Problema

```
movie_extractor.py × predict.py × link_extractor.py × movies_links × movies_test.csv × movies.csv ×
1 http://www.adorocinema.com/filmes/filme-136634/
2 http://www.adorocinema.com/filmes/filme-59124/
3 http://www.adorocinema.com/filmes/filme-4616/
4 http://www.adorocinema.com/filmes/filme-189555/
5 http://www.adorocinema.com/filmes/filme-2436/
6 http://www.adorocinema.com/filmes/filme-43039/
7 http://www.adorocinema.com/filmes/filme-42379/
8 http://www.adorocinema.com/filmes/filme-210032/
9 http://www.adorocinema.com/filmes/filme-45563/
10 http://www.adorocinema.com/filmes/filme-28363/
11 http://www.adorocinema.com/filmes/filme-25861/
12 http://www.adorocinema.com/filmes/filme-57673/
13 http://www.adorocinema.com/filmes/filme-231662/
14 http://www.adorocinema.com/filmes/filme-702/
15 http://www.adorocinema.com/filmes/filme-27638/
16 http://www.adorocinema.com/filmes/filme-127661/
17 http://www.adorocinema.com/filmes/filme-89485/
18 http://www.adorocinema.com/filmes/filme-45359/
19 http://www.adorocinema.com/filmes/filme-43019/
20 http://www.adorocinema.com/filmes/filme-25478/
21 http://www.adorocinema.com/filmes/filme-1422/
22 http://www.adorocinema.com/filmes/filme-28919/
23 http://www.adorocinema.com/filmes/filme-26038/
24 http://www.adorocinema.com/filmes/filme-3996/
25 http://www.adorocinema.com/filmes/filme-198903/
26 http://www.adorocinema.com/filmes/filme-229155/
27 http://www.adorocinema.com/filmes/filme-61640/
28 http://www.adorocinema.com/filmes/filme-28777/
29 http://www.adorocinema.com/filmes/filme-130209/
30 http://www.adorocinema.com/filmes/filme-1810/
31 http://www.adorocinema.com/filmes/filme-229453/
32 http://www.adorocinema.com/filmes/filme-53667/
33 http://www.adorocinema.com/filmes/filme-171259/
34 http://www.adorocinema.com/filmes/filme-52501/
35 http://www.adorocinema.com/filmes/filme-110941/
```

Dados do Problema

```
ies_test.csv /
movie_extractor.py x predict.py x link_extractor.py x movies_links x movies_test.csv x movies.csv x
The file size (2901982 bytes) exceeds configured limit (2560000 bytes). Code insight features are not available.
1 "category_1","category_2","category_3","description"
2 "13","7","0","Don Vito Corleone (Marlon Brando) é o chefe de uma 'família' de Nova York que está feliz, pois Co
3 "7","12","4","A inusitada história de Oskar Schindler (Liam Neeson), um sujeito oportunista, sedutor, 'armador
4 "10","3","1","Sauron planeja um grande ataque a Minas Tirith, capital de Gondor, o que faz com que Gandalf (Ia
5 "7","0","0","Em 1946, Andy Dufresne (Tim Robbins), um jovem e bem sucedido banqueiro, tem a sua vida radicalme
6 "2","3","9","Mufasa (voz de James Earl Jones), o Rei Leão, e a rainha Sarabi (voz de Madge Sinclair) apresentar
7 "13","7","0","Início do século XX. Após a máfia local matar sua família, o jovem Vito (Robert De Niro) foge da
8 "7","5","12","Durante a Segunda Guerra Mundial na Itália, o judeu Guido (Roberto Benigni) e seu filho Giosué s
9 "6","14","0","Quarenta anos da história dos Estados Unidos, vistos pelos olhos de Forrest Gump (Tom Hanks), um
10 "1","0","0","Após os acontecimentos em Londres, Dom (Vin Diesel), Brian (Paul Walker), Letty (Michelle Rodrigue
11 "10","13","0","1935, no corredor da morte de uma prisão sulista. Paul Edgecomb (Tom Hanks) é o chefe de guarda
12 "10","3","0","Harry Potter (Daniel Radcliffe) e seus amigos Rony Weasley (Rupert Grint) e Hermione Granger (Em
13 "1","7","0","Após dois anos desde o surgimento do Batman (Christian Bale), os criminosos de Gotham City têm mu
14 "10","3","1","Numa terra fantástica e única, chamada Terra-Média, um hobbit (seres de estatura entre 80 cm e 1
15 "11","3","5","Um jovem (Michael J. Fox) aciona acidentalmente uma máquina do tempo construída por um cientista
16 "10","3","1","Após a captura de Merry (Dominic Monaghan) e Pippa (Billy Boyd) pelos orcs, a Sociedade do Anel
17 "5","0","0","Philippe (François Cluzet) é um aristocrata rico que, após sofrer um grave acidente, fica tetrapl
18 "2","3","9","Cinco anos após convencer os habitantes de seu vilarejo que os dragões não devem ser combatidos,
19 "11","3","10","As forças imperais comandadas por Darth Vader (David Prowse) lançam um ataque contra os membros
20 "7","0","0","Jackson, pequena cidade no estado do Mississippi, anos 60. Skeeter (Emma Stone) é uma garota da soc
21 "2","5","9","Riley é uma garota divertida de 11 anos de idade, que deve enfrentar mudanças importantes em sua v
22 "3","1","11","Décadas após a queda de Darth Vader e do Império, surge uma nova ameaça: a Primeira Ordem, uma o
23 "7","0","0","Parker Wilson (Richard Gere) é um professor universitário que, ao retornar do trabalho, encontra
24 "14","5","0","Inglaterra, 1797. As cinco irmãs Bennet - Elizabeth (Keira Knightley), Jane (Rosamund Pike), Lyd
25 "0","0","0","Django (Jamie Foxx) é um escravo libertado cujo passado brutal com seus antigos proprietários leva-
26 "7","0","0","Carl Brashear (Cuba Gooding Jr.) veio de uma humilde família negra que vivia em uma área rural e
```


Categorias

```
13
14
15 class Category(Enum):
16     Outros = 0
17     Acao = 1
18     Animacao = 2
19     Aventura = 3
20     Biografia = 4
21     Comedia = 5
22     Comedia_dramatica = 6
23     Drama = 7
24     Documentario = 8
25     Familia = 9
26     Fantasia = 10
27     Ficcao_cientifica = 11
28     Guerra = 12
29     Policial = 13
30     Romance = 14
31     Terror = 15
32
```

Demo