

Klasifikimi i spam-it

Rudina Gjokokaj
(rudina.gjokokaj@universitetiaab.com)

Abstrakt

Në ditët e sotme, të regjistrohemi ose të identifikohemi në ndonjë uebfaqe, duhet të pajisni me email-in tuaj dhe ndonjëherë edhe numrin tuaj të telefonit. Këto detaje përdoren për të verifikuar përdoruesin. Por, ekziston një shans që këto detaje të mund të keqpërdoren për promovime, mesazhe false, etj. Shembull, nëse shkruani detajet e bankës, numrin e telefonit dhe email-id tuaj për të blerë një produkt nga një faqe interneti në kërkim, disa ditë më vonë ju me siguri do të merrni një postë nga gjysma e rrugës në të gjithë botën duke pretenduar se keni fituar 100 milion dollarë. Shumica prej nesh e dinë që ky mesazh është i rremë dhe kjo email duhet të përfundojë në spam. Ky mashtrim thjesht nuk funksionon më. Email është njëra ndër rrjetet që perdoret më së shumti dhe është me e shpejta. Unë kam rishikuar algoritmin Naïve Bayes.

Fjalët kyçe

Spam, ham, klasifikimi i email, algoritmi i mësimit makinorë, Naïve Bayesian.

1. Hyrje

Spamer(njerzit që dergojne spam) shfrytëzojnë rastet si fatkeqsi natyrore, lojrat sportive për të dërguar spam email. Mesazhet në formë marketingu renditen të parët sa i përket sasisë së spam email. Spam email më së shumti dërgohen nga Kina, SHBA, Russia. Fig.1. [1]

Viteve të fundit, në statistika të bëra gjatë viteve 2012 deri në vitin 2018, Fig.2. shihet qartë se numri i email-ve spam ka pësuar rënie.[2]

Llojet e spam janë:

- Reklama për vende të ndryshme të jashtme, shpesh SNS të tjera
- Mashtrim paraprak duke përfshirë trashëgiminë, llotarinë, vizat dhe mashtrimet e zhdoganimit
- Mashtrime romantike, përfshirë "nusen me porosi" dhe mashtrimet ushtarake
- Reklama për faqet pornografike, kryesisht kamera në internet për të rritur dhe faqet e bisedave të drejtpërdrejta.
- Mbledhja e parave, shpesh nën maskën e vendeve të punës me pagë të lartë "qëndrimi në shtëpi" ose blerjeve "mister"
- Kërkesa për lypje dhe dhurata
- Propozime afariste, etj. [3]

Spam është i pranishëm në SMS, MMS, image files, video files dhe në uebfaqe.

Spam i Uebfaqe-ve është më sfiduesi. Lloji më i zakonshëm në listë është spam-i në internet që shfrytëzon dobësitë dhe boshllëqet në web 2.0 për të injektuar lidhje me përmbajtjen e spamit në përmbajtje dinamike dhe të ndjeshme siç janë bloget, komentet, rishikimet ose wikipages. Shembujt janë komente në bloget e WordPress. Komentet janë të larmishëm. Komenti i parë (figura 3 (a)) duket si një koment i vlefshëm, por fsheh një lidhje në emrin e përdoruesit. Komenti i dytë (figura 3 (b)) përmban lidhje në tekstin gati normal. Komenti i fundit (Figura 3 (c)) është një shembull tipik i

komentit të spam-it në internet me disa lidhje të injektuara pa tekst normal. [4]

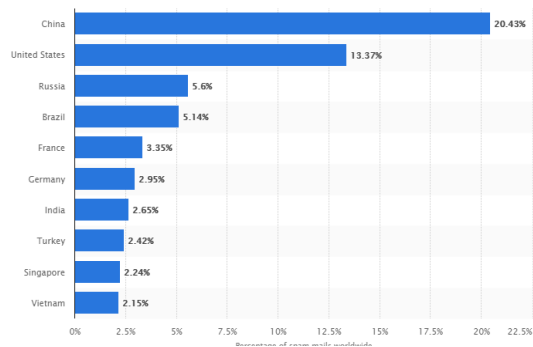


Fig.1. Vendet kryesore të origjinës për e-mail të padëshiruar për spam në tremujorin e 3-të 2019, nga njësia e vëllimit të spamit në të gjithë botën.

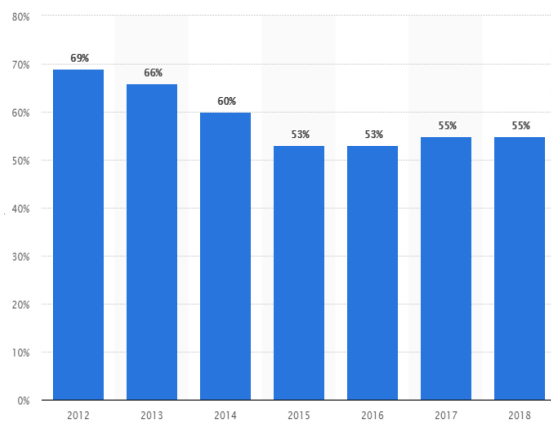


Fig.2. Shkalla globale e spamave të e-mails nga 2012 në 2018

swimwear galore
November 10, 2013 at 11:30 AM

Thanks for one's marvelous posting! I definitely enjoyed reading it, you can be a great author. I will ensure that I bookmark your blog and will often come back someday. I want to encourage continue your great work, have a nice holiday weekend!

(a)

dfcyayrn
November 1, 2013 at 8:24 pm

A more likely explanation is that this lets them get the kinks worked out for a basic set of capabilities, so that its easier to track down the source of the problems that will show up when they extend the capabilities. If Hitler had his "willing executioners" well who am I to argue if America has it's "willing private interests." Of course "hiding in plain sight" makes it hard to call it thievery legal or not. <http://www.klipplust.se/canada-goose/canada-goose-sale> Abyhtw Chrissy decided to rob him for \$32,000 then she ran to Miami to shop and have fun on his dime. [canada goose sverige](http://www.canadagoose.com) [canada goose jacka herr](http://www.canadagoose.com) 1042498304

(b)

lxmauonfz
November 24, 2013 at 11:41 am

Cfcohed Sptposyqx Qrhvoj [Canada Goose Norge](http://www.dit.atl.no/canadagoosejakker.html) Mehfpk Dpqufkqz Jusoko <http://www.dit.atl.no/canadagoosejakker.html> Vcfvfp Fcwltgghi Qrhvoj [Canada Goose Outlet](http://www.dit.atl.no/canadagoosejakker.html) Qtrgebyzd Ndsisctwl Ovgvb <http://www.dit.atl.no/canadagoosejakker.html> Cxdgx Qxdekvez Sprrxdwf [Canada Goose Norge](http://www.olaviken.no/canadagoose.html) Dwtyqp Bcpvaek Eqfkbont <http://www.olaviken.no/canadagoose.html> Rreacmv Qdgsimng Nlxnja [canada goose jakke](http://www.ugress.com/billigcanadagoose.html) Sdzeshj Adtcyfxb Rrelow <http://www.ugress.com/billigcanadagoose.html> <http://coquihilton.com/puerto-rico-politics/#comment-1484> <http://www.ttsports.biz/style/content/CN-04a/underground-mining-equipment-as-an-occasion-to-establish-a-successful-business-at-present/#comment-3894> <http://www.sexy-lingerie-lover.com/blog/index.php/2012/08/how-to-choose-your-own-sexy-underwear/#comment-44404> <http://www.txt-alert.com/2010/08/06/mobile-marketing-is-effective/#comment-14029> <http://prospeccaodeclientes.com.br/blog/index.php/reuniao-de-vendas-reuniao-comercial-367.html/comment-page-1#comment-17546>

(c)

Fig.3. Komentet në blog-et e WordPress

2. Puna e lidhur

Malware dhe Botnet kanë ndihmuar spammers të përhapin spam gjerësisht. Pas marrjes dhe hapjes së një emaili spam, përdoruesit e Internetit janë të ekspozuar ndaj çështjeve të sigurisë pasi që spams transmetohen normalisht për qëllim të keq.

Siponen dhe Stuckye [5] përcaktuan qasje të ndryshme të përdorura për kontrollin e spamit. Ata arritën në përfundimin se filtrimi është mënyra më efektive për kontrollimin e spamit. Wang dhe Cloete [6] debatuan mbi klasifikimin e postës elektronike së bashku me filtrimin e spamit dhe diskutuan teknikat e filtrimit në detaje. Në këtë punim [7] studiuesit aplikuan me sukses teknikat e klasifikimit të email Naïve Bayes. Klasifikuesit e tjerë Bayesian kanë punuar gjithashtu për të filtruar spam [8]. Në këtë punim [9] studiuesi propozoi teknika për të përmirësuar performancën e klasifikuesit Naïve Bayes.

3. Metodologjia

Këtu do të përdorë algoritmin Naïve Bayes për klasifikimin e spam e-mails.

3.1. Algoritmi Naïve Bayes

Algoritmi Naive Bayes është një algoritëm mësimor i mbikëqyrur për klasifikimin. Mëson mundësinë e një objekti me veçori të caktuara që i përkasin një grupi të caktuar. Shkurt, është një klasifikues i mundshëm. Algoritmi Naive Bayes quhet "naiv" sepse bën supozimin se ndodhja e një veçorie të caktuar është e pavarur nga shfaqja e tipareve të tjera. Klasifikimi i tekstit përdoret si një metodë e të mësuarit probabilistik për klasifikimin e tekstit. Klasifikuesi Naive Bayes është një nga algoritmet më të suksesshme të njohura kur bëhet fjalë për klasifikimin e dokumenteve të tekstit. Klasifikimi i spam është bërë një mekanizëm popullorë për të dalluar emailin spam nga emaili legjitim. Disa shërbime moderne të email implementojnë filtrimin e spamit Bayesian.

Shumë filtra të postës elektronike të serverit, të tilla si DSPAM, SpamBayes, SpamAssassin, Bogofilter dhe ASSP, përdorin këtë teknikë.

Algoritmi Naïve Bayes në disa statistika renditet si algoritmi më i përdorur për klasifikimin e spam fig.4.[10].

Procesi i filtrimit të spam-it në email bazuar në Algoritmin Naive Bayes është paraqitur në fig.5.[12]

Marcin Luckner, Michał Gad, and Paweł Sobkowiak në hulumtimin e tyre kanë konsideruar dhjetë periudhat mujore. Gjatë periudhave, të dyja komentet e spamit dhe ato ham ishin mbledhur.[11]

Algorithms Used in Spam Filtering

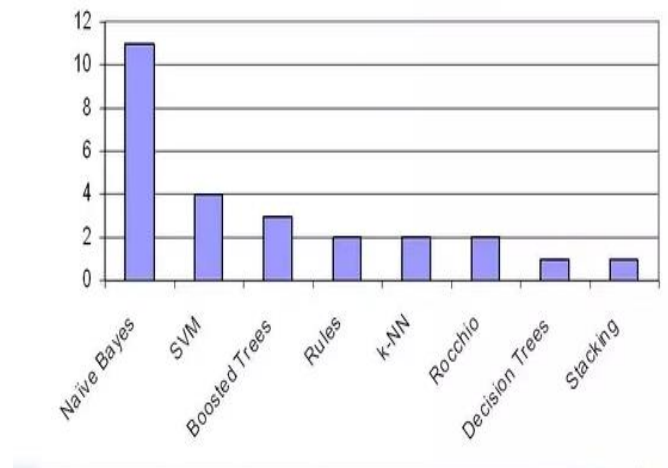


Fig.4. Përdorimi i algoritmeve për klasifikimin e spam-it

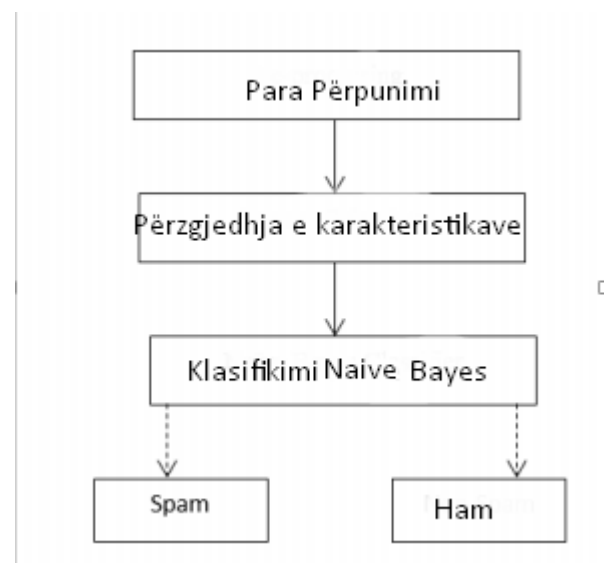


Fig.5.

Ata krahasuan tre strategjitë e mëposhtme të krijimit të grupit mësimor për klasifikimin: dinamike, rritëse dhe statike. Për të gjitha strategjitë, testi është bërë veçmas në nëntë periudha mujore nga T1 në T9.

Vlerësimi i rezultateve të marra në grupet e testimit është bërë nga masat e mëposhtme: TP (true positive) numri i regjistrimeve të spamëve të njohura saktë. TN (true negative) numri i hyrjeve ham të njohura saktë. FP (false pozitive) numri i hyrjeve spam të njohura gabimisht. FN (false negative) numri i hyrjeve ham të njohura gabimisht. Në këtë mënyrë kanë punuar edhe Antonio Lupher, Cliff Engle, Reynold Xin[13].

Evaluation Measure	Evaluation Function
Accuracy	$Acc = \frac{TN+TP}{TP+FN+FP+TN}$
Recall	$r = \frac{TP}{TP+FN}$
Precision	$P = \frac{TP}{TP+FP}$
F-measure	$F = \frac{2pr}{p+r}$

Përqindja e mesazheve të identifikuara në mënyrë korrekte dhe jo spam:

$$acc = \frac{TN+TP}{TP+FN+FP+TN}$$

Përqindja e mesazhit spam arrin të bllokojë:

$$R = \frac{TP}{TP+FN}$$

Përqindja e mesazhit të saktë për spam email:

$$P = \frac{TP}{TP+FP}$$

Mesatarja e ponderuar e e R dhe P:

$$F = \frac{2PR}{P+R}$$

Në secilën tabelë, rezultatet e theksuara janë rezultatet më të mira të marra për një masë të caktuar midis tre strategjive të të mësuarit. Strategjia dinamike dhe strategjia rritëse japin rezultate të mira të klasifikimit.

TABLE I
DATASETS

	T_0	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9
Spam trap	2588	7673	7371	4176	1783	7746	17419	15323	14112	2065
Quora	45	40	41	53	59	47	54	53	73	49
Reddit	105	362	51	555	319	40	24	867	626	343
Stack Overflow	1104	951	972	1028	1057	1007	849	717	884	689

TABLE II
STATISTICS FOR QUORA OBTAINED BY RANDOM FOREST

	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9
Dynamic									
ACC	1.0000	0.9981	0.9991	0.9973	0.9995	0.9994	0.9997	0.9997	0.9991
TPR	1.0000	1.0000	0.9993	0.9983	0.9999	0.9996	0.9997	0.9997	0.9990
SPC	1.0000	0.7455	0.9804	0.9655	0.9388	0.9216	1.0000	1.0000	1.0000
FI	1.0000	0.9990	0.9995	0.9986	0.9997	0.9997	0.9998	0.9999	0.9995
Incremental									
ACC	1.0000	0.9943	0.9991	0.9843	0.9850	0.9997	0.9997	0.9998	0.9991
TPR	1.0000	1.0000	1.0000	0.9983	0.9997	0.9998	0.9997	0.9998	0.9990
SPC	1.0000	0.4940	0.9298	0.6829	0.2812	0.9804	1.0000	1.0000	1.0000
FI	1.0000	0.9971	0.9995	0.9918	0.9924	0.9999	0.9999	0.9999	0.9995
Static									
ACC	1.0000	0.9950	0.9790	0.8333	0.6687	0.4434	0.5244	0.7724	0.8723
TPR	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
SPC	1.0000	0.5256	0.3732	0.1612	0.0179	0.0055	0.0072	0.0221	0.1536
FI	1.0000	0.9975	0.9892	0.9058	0.8000	0.6128	0.6866	0.8708	0.9301

TABLE III
STATISTICS FOR STACK OVERFLOW OBTAINED BY RANDOM FOREST

Stat	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9
Dynamic									
ACC	0.9686	0.9942	0.9773	0.9299	0.9721	0.9931	0.9948	0.9945	0.9851
TPR	1.0000	0.9997	0.9995	0.9981	0.9992	0.9970	0.9953	0.9953	0.9805
SPC	0.7782	0.9547	0.8984	0.8432	0.8079	0.9150	0.9832	0.9808	1.0000
FI	0.9820	0.9967	0.9857	0.9410	0.9840	0.9964	0.9973	0.9971	0.9902
Incremental									
ACC	0.9667	0.9942	0.9710	0.8908	0.9481	0.9681	0.9981	0.9977	0.9938
TPR	1.0000	0.9999	1.0000	0.9993	0.9999	0.9994	0.9990	0.9989	0.9918
SPC	0.7682	0.9538	0.8719	0.7736	0.6895	0.5942	0.9804	0.9786	1.0000
FI	0.9809	0.9967	0.9816	0.9048	0.9698	0.9830	0.9990	0.9988	0.9959
Static									
ACC	0.9681	0.8632	0.8100	0.6067	0.3728	0.2069	0.1436	0.2279	0.3635
TPR	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
SPC	0.7757	0.4600	0.5097	0.4862	0.1550	0.0554	0.0496	0.0709	0.2821
FI	0.9818	0.9161	0.8657	0.5439	0.4511	0.2880	0.1877	0.3045	0.2625

Fig.6. Statistikat në nëntë periudha

4. Përfundimi dhe puna në të ardhmen

4.1. Përfundimi

Klasifikuesi Naive Bayes kishte një saktësi prej 82% kur ishte trajnuar në fjalët më të mira. Saktësia e saj ra në 77% kur u shtuan tipare ekspertësh. Rënia e saktësisë pas shtimit të veçorive të ekspertëve ka shumë të ngjarë të shpjegohet me supozimin "naiv" në këtë algoritëm se tiparet janë të pavarura. Disa nga karakteristikat që kemi përdorur, p.sh., gjatë jetës së llogarisë dhe nëse një përdorues ka miq ose foto, në të vërtetë ka të ngjarë të jenë të varura, duke çuar në një mbi peshë të këtyre karakteristikave gjatë klasifikimit.[13]

Deshtimi i strategjisë Statike:[11]

Arsyeja kryesore pse strategjia statike nuk funksionon është një koncept. Përkufizimi statistikor i ndryshimeve spam dhe jo-spam me kalimin e kohës në mënyrë të paparashikuar. Kjo shkakton probleme sepse parashikimet bëhen më pak të sakta me kalimin e kohës. Për të vërtetuar këtë arsyetim ata kanë testuar ndryshime të rëndësishme të tipareve në procesin e klasifikimit.

Strategjia dinamike është më e madhe se sa për strategjia e rritjes në shumicën e provave dhe më e vogël në disa teste me vetëm një sasi të vogël. Ata krahasuan të dy strategjitë në të 27 kombinimet e të dhënave dhe muajt. Për 16 çifte acc që është llogaritur për strategjinë dinamike ishte më e madhe. Një situatë e kundërt ka ndodhur në 8 raste. Në pjesën tjetër të rasteve, rezultatet ishin të njëjta për të dy strategjitë.

4.2. Puna në të ardhmen

Për shkak se koha ishte e shkurtër dhe marrja e informative dhe e hulumtimeve ishte më e vogël, unë shpreh interesimin tim për të vazhduar këtë hulumtim. Kësaj here dëshiroj që të hulumtoj rreth klasifikimit të spam me algoritmin SVM(Support Vector Machine).

5. Referencat

[1] <https://www.statista.com/statistics/263086/countries-of-origin-of-spam/>

[2] <https://www.statista.com/statistics/270899/global-e-mail-spam-rate/>

[3] Feature Selection and Classification of Spam on Social Networking Sites(Antonio Lupher, Cliff Engle, Reynold Xin) fq.4

[4] Antyscam—Practical Web Spam Classifier Marcin Luckner, Michał Gad, and Paweł Sobkowiak

[5] M. Siponen and C. Stucke, "Effective anti-spam strategies in companies", An international study, In Proceedings of HICSS 2006, vol. 6, 2006.

[6] X. Wang and I. Cloete, "Learning to classify email: a survey", In Proceedings of the 2005 International Conference on Machine Learning and Cybernetics, ICMLC 2005, pp. 5716-5719, 2005.

[7] D. D. Lewis, and M. Ringuette, "Comparison of two learning algorithms for text categorization", In Proceedings of SDAIR, fq. 81–93, 1994.

[8] D. Koller, and M. Sahami, "Hierarchically classifying documents using very few words", In Machine Learning: Proceedings of the Fourteenth International Conference, fq. 170–178, 1997.

[9] J. D. Rennie, L. Shih, J. Teevan, and D. R. Karger, "Tackling the poor assumptions of naïve bayes text classifiers". In Proceedings of ICML, fq. 616–623, 2003.

[10] <https://www.quora.com/Which-algorithms-are-best-to-use-for-spam-filtering-How-should-they-be-implemented>

[11] Antyscam—Practical Web Spam Classifier Marcin Luckner, Michał Gad, and Paweł Sobkowiak, fq. 725

[12] "Analysis of Naïve Bayes Algorithm for Email Spam Filtering across Multiple Datasets", Nurul Fitriah Rusland, Norfaradilla Wahid, Shahreen Kasim, Hanayanti Hafit, fq.4-5.

[13] Feature Selection and Classification of Spam on Social Networking Sites Antonio Lupher, Cliff Engle, Reynold Xin, fq.8