Kumpulan jurnaL Ilmu Komputer (KLIK)
Volume 10 No. 01 2023
ISSN: 2406-7857

ANALISIS SPAM KOMENTAR WORDPRESS DIIHAT DARI ALAMAT IP

Imam Suharjo[1], Indah Susilawati[2], Putry Wahyu Setyaningsih[3]
[1,2]Prodi Informatika [3]Prodi Sistem Informasi Universitas Mercu Buana Yogyakarta
Jl Wates Km 10 Yogyakarta
imam@mercubuana-yogya.ac.id, indah@mercubuana-yogya.ac.id, putry@mercubuana-yogya.ac.id

Abstract

Website menjadi sumber rujukan resmi dan formal, meskipun sudah ada media sosial atau media yang lain. Website bersifat terbuka dan bisa diakses secara publik. Spam komentar dapat muncul di berbagai platform situs web, termasuk platform blogging seperti WordPress, platform media sosial seperti Facebook, dan forum online. Namun, platform yang paling sering mengalami masalah spam komentar adalah platform blogging. Spam komen adalah sejenis pesan atau komentar yang tidak diinginkan atau tidak relevan yang dikirimkan ke sebuah forum atau situs web. Biasanya, spam komen dikirimkan dengan tujuan untuk meningkatkan visibilitas suatu situs atau untuk mengirimkan pesan promosi atau iklan kepada orang lain. Penelitian ini bertujuan untuk analisis dan klasifikasi alamat IP yang melakukan spam terhadap website. Daftar alamat IP ini dikumpulkan dari berbagai web dan dirancang bersifat terbuka. Hasil sampel dari 10 alamat IP yang ada di web dan dicek dengan menggunakan data dari cleantalk.org/blacklists, semua alamat IP sampel ini tercatat sebagai alamat IP Spammer. Dilihat dari asal alamat IP spam komen berasal dari luar negeri terutama menggunakan alamat IP dari Rusia dan Latvia, meski sangat dimungkinkan pelaku spammer bersembunnyi dbalik alamat IP ini. Konten spam mengandung beberapa link aktif dengan tak ahref dan juga konten-konten terutama terkait drug dan farmasi. Ada juga konten lain dengan bahasa asing bukan bahasa inggris.
Keywords*: spammer, keamanan web, wordpress, internet*

Abstrak

A website becomes an official and formal reference source, although there are other social media or other media. The website is open and publicly accessible. Spam comments can appear on various website platforms, including blogging platforms such as WordPress, social media platforms such as Facebook, and online forums. However, the platform that most frequently experiences spam comment issues is the blogging platform. Spam comments are unwanted or irrelevant messages or comments sent to a forum or website. Usually, spam comments are sent to increase a site's visibility or send promotional or advertising messages to others. The research was conducted by analyzing and classifying IP addresses that spam websites. This list of IP addresses is collected from various websites and is designed to be open. A sample result of 10 IP addresses on the web and checked using data from cleantalk.org/blacklists, all of these sample IP addresses are recorded as spammer IP addresses. From the origin of spam comment IP addresses, they are mainly from abroad, especially using IP addresses from Russia and Latvia. However, spammer perpetrators are likely hiding behind these IP addresses. Spam content contains several active links with no ahref and also content related to drugs and pharmacies. There is also other content with foreign languages other than English.

Kata kunci*: spammer, web security, WordPress, internet*

1. PENDAHULUAN

Website menjadi sumber rujukan resmi dan formal, meskipun sudah ada media sosial atau media yang lain. Website bersifat terbuka dan bisa diakses secara publik. Website saat ini dinamis, ada interaksi dengan pengunjung baik dalam bentuk formulir (*form*), komentar, buku tamu dan sejenisnya. Formulir terbuka ini menjadikan web lebih dinamis dan ada interaksi 2 arah. Form kementar dibuka sebagai sarana interaksi dengan pembaca untuk memberikan tanggapan atau upan balik. Namun disisi lain dengan formulir yang terbuka ini menjadi sasaran dalam keamanan website.

Spam komentar dapat muncul di berbagai platform situs web, termasuk platform blogging seperti WordPress, platform media sosial seperti Facebook, dan forum online. Namun, platform yang paling sering mengalami masalah spam komentar adalah platform blogging, karena seringkali memiliki fitur yang memungkinkan pengguna untuk mengirimkan komentar pada postingan. Spam komentar dapat menurunkan kualitas konten pada situs dan dapat mengganggu pengalaman pengguna. Spam di internet adalah pesan atau konten yang dikirim secara masal ke orang yang tidak menginginkannya, biasanya melalui email atau pesan singkat [1]. Spam juga dapat muncul di media sosial, forum, komentar blog, dan lainnya. Beberapa contoh lain dari spam di internet adalah email, SMS, media sosial, komentar dan forum.

Menurut Internet Live Stat, jumlah total situs web online telah melonjak lebih dari 1,9 miliar dan terus meningkat. CMS paling populer yang diinstal pada situs web tersebut adalah kerangka kerja berbasis PHP seperti Wordpress dan Joomla, yang akunnya untuk (per Desember 2018) 59% dan 6% dari pangsa pasar. Sebagian besar situs web tersebut milik perusahaan kecil dan individu yang mungkin memiliki sedikit motivasi untuk membayar ribuan dolar untuk layanan pemeliharaan keamanan. Selain itu, karena biaya pengembangan, pemilik – bahkan pengembang yang mereka pekerjakan – mungkin lebih memilih CMS siap pakai untuk mempercepat penerapan [2].

Spam sering berisi tautan-tautan karena spammer mencoba untuk mengarahkan pengguna ke situs web atau layanan yang mereka promosikan. Tautan-tautan tersebut dapat muncul di dalam spam email, pesan spam di media sosial atau komen spam di situs web. *Spammer* menggunakan tautan untuk meningkatkan traffic ke situs web atau layanan yang mereka promosikan [3].

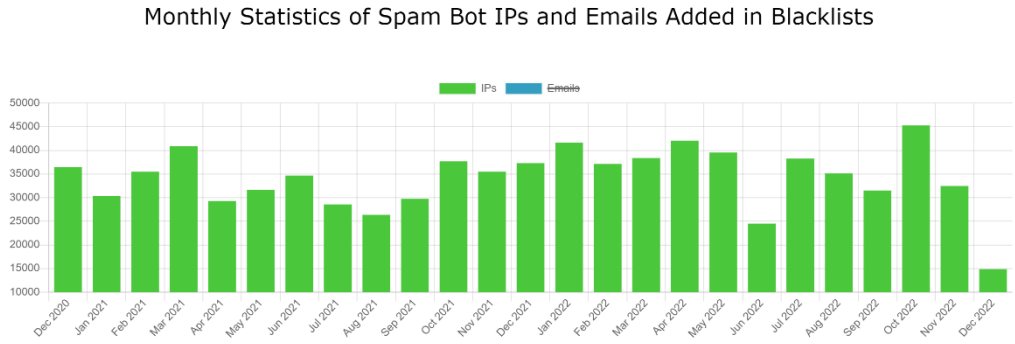
Spammer komen dapat menggunakan cara manual atau otomatis untuk mengirimkan spam komen ke situs web. Spammer dapat menggunakan *software* untuk mengirimkan spam komen ke banyak situs web sekaligus, atau mengatur software tersebut untuk secara teratur mengirimkan spam komen ke situs web tertentu [4]. Beberapa alat yang sering digunakan oleh spammer untuk mengirimkan spam komen antara lain : *software spamming*, *proxies*, layanan hosting dan *form filler* [5].

Keamanan yang terjadi pada formulir yang terbuka bisa berupa percobaan injeksi, serangan yang berulang atau tulisan yang bertujuan untuk melakukan spamming. Meskipun sudah banyak aplikasi atau sistem untuk proteksi namun serangan seperti spam masih saja terjadi. Seperti proteksi spam dengan akismet di

Wordpress, komentar spam masih bisa lolos dan ditandai sebagai spam. Komen spam biasanya diikuti dengan tautan atau link aktif yang bertujuan sebagai optimasi website atau SEO [1].

Website yang besar dengan pengunjung, spam lebih sering terjadi. Penanganan sistem terhadap spam juga akan memperberat kerja server. Untuk sistem yang dikelola sendiri cara yang baik dan memperingan server dengan memblokir alamat IP yang terindikasi sebagai spammer. Teknologi anti spam juga sudah tersedia, namun kadang anti spam seperti *captcha* dirasa membuat tidak nyaman bagi pengunjung.

Antara Oktober 2020 dan September 2021, volume spam harian global mencapai titik tertinggi pada Juli 2021, dengan hampir 283 miliar email spam dari total 336,41 miliar email terkirim. Per Agustus 2021, jumlah ini turun menjadi 65,50 miliar. Menjelang September, rata-rata volume spam kembali meningkat 36 persen, mencapai 88,88 miliar dari total 105,67 miliar email yang dikirim ke seluruh dunia.



Gambar 1. Statistik IP spam setiap per bulan tahun 2022 [4].

Setiap hari CleanTalk mendapatkan informasi tentang ribuan IP/email spam baru dan basis data CleanTalk dari IP aktif spam dan alamat email memiliki informasi tentang lebih dari 20.000.000 alamat IP. Statistik Spam bot IP yang dihimpun dari cleantalk.org (data hingga 15 Desember 2022) mencapai 30.000 hingga 40.000 alamat IP spam ada setiap bulan. Sementara data IP spam setiap hari yang dicatat oleh claertalk bisa mencapai 800-1200 alamat IP spam setiap hari. Data dari Akismet sejak 2005 hingga saat ini Februari 2022, telah menyimpan 527.206.886.474 spam dari web, rata-rata sekitar ada sekitar 7.5juta spam per jam di dunia ini [4], [6].

Keamanan yang terjadi pada formulir yang terbuka bisa berupa percobaan injeksi, serangan yang berulang atau tulisan yang bertujuan untuk melakukan spamming. Meskipun sudah banyak aplikasi atau sistem untuk proteksi namun serangan seperti spam masih saja terjadi. Seperti proteksi spam dengan akismet di Wordpress, komentar spam masih bisa lolos dan ditandai sebagai spam. Komen spam biasanya diikuti dengan tautan atau link aktif yang bertujuan sebagai optimasi website atau SEO [1].

Kehidupan manusia menjadi mudah dengan bantuan komputasi awan. Semua data dan aplikasi sekarang berada di cloud bisa diakses dari mana saja. Namun, masalah keamanan selalu ada di lapisan aplikasi untuk sistem tertentu.

Layanan yang diminta oleh aplikasi akan diberikan berdasarkan kredensialnya. Sistem deteksi intrusi jaringan yang sangat cepat, kuat, dan produktif yang dapat digunakan untuk pencegahan spam formulir atau komentar di web dalam aplikasi yang bebas di internet. Kombinasi kolektif kerangka kerja dan pendekatan kecerdasan buatan dapat sangat membantu untuk menghindari dan mencegah sistem dari ancaman serangan jaringan [7].

Sebagian besar dari hampir dua miliar situs web sekarang di Internet dijalankan oleh individu dan organisasi tanpa keterampilan atau sumber daya untuk mengamankannya. Situs-situs ini menjadi sasaran para penjahat dunia maya yang ingin mengeksploitasi kelemahan untuk 'memonetisasi' mereka. Van-Linh Nguyen, Po-Ching Lin dan Ren-Hung Hwang dari Universitas Nasional Chung Cheng, Taiwan, memeriksa taktik yang digunakan oleh penjahat dan menyajikan kerangka kerja untuk mencegah serangan semacam itu, sambil mengakui bahwa ini adalah pertempuran yang kemungkinan akan berlanjut [8]

Spam Konten atau Spamming Konten adalah jenis spam web pertama dan paling luas karena mengeksploitasi model pencarian informasi berbasis mesin pencari. Model-model ini dibangun dari konten halaman yang selanjutnya memberi peringkat halaman berdasarkan algoritma peringkat halaman. Akibatnya para spammer menganalisis kelemahan model-model yang diterapkan dan mengeksploitasinya. Berbagai jenis konten spamming adalah Judul spam, tubuh spamming, Meta tag spamming, Anchor Text spamming, URL spamming [9].

Bidang tulisan di website sangat penting dalam pencarian informasi, spammer mencoba untuk mengisinya secara berlebihan untuk meningkatkan peringkat halaman dan jenis spamming ini di mana seseorang yang mengisi judul secara berlebihan disebut spamming judul. Dalam body spamming, isi halaman dimodifikasi dan disuntikkan dengan konten atau kueri tertentu yang sering dicari. Meta-tag memainkan peran khusus dalam deskripsi dokumen seperti ketika menggunakan algoritma pencarian dan mesin pencari, hasilnya diambil berdasarkan meta-tag di halaman web. Menempatkan spam di konten ini akan menjadi cara paling efisien untuk mengirim spam ke dokumen. Anchors tag adalah tag yang digunakan untuk menyertakan tautan di situs web sehingga dalam spamming jangkar para spammer membuat tautan dengan teks jangkar yang diinginkan untuk mendapatkan istilah yang tepat untuk halaman target. Dalam spam URL, konten yang akan dicari disuntikkan ke URL itu sendiri [10].

Telah terjadi peningkatan spam SEO dan malware yang bertujuan untuk menghancurkan indeks pencarian aplikasi dengan menambahkan konten, yang tidak sesuai dengan ide aplikasi. Tautan, halaman, atau komentar yang tidak biasa dapat membingungkan pengunjung situs web dan mengarahkan lalu lintas ke situs web yang tidak dikehendaki. Situs web yang sangat aman adalah target yang menarik bagi setiap peretas. Dalam beberapa kasus, hambatan dalam menerapkan setiap jenis perlindungan keamanan hanyalah waktu, tetapi alasan terburuknya adalah pengembang yang kurang informasi atau tidak memenuhi syarat tanpa pengalaman dalam SEO [1]

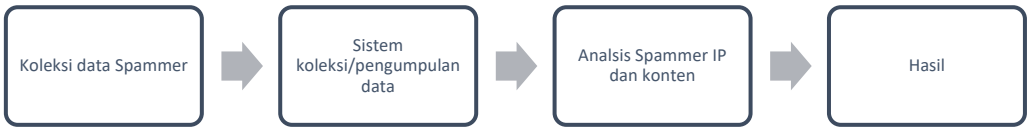
Tujuan utama dari komen spam adalah untuk meningkatkan visibilitas suatu situs web atau mengirimkan pesan promosi atau iklan kepada orang lain. Komen

spam biasanya dikirimkan oleh orang-orang atau perusahaan yang ingin mempromosikan suatu produk atau layanan. Mereka menggunakan spam komen sebagai cara untuk mencapai lebih banyak orang dan meningkatkan traffic ke situs web mereka. Komen spam juga dapat digunakan oleh penjahat *cyber* untuk menyebarkan virus atau *malware* ke komputer pengguna yang menerima spam komen tersebut. Spam komen juga dapat digunakan untuk menipu orang dengan memberikan tautan ke situs web palsu atau mengajukan permintaan informasi pribadi.

2. METODOLOGI PENELITIAN

2.1. Tahapan Penelitian

Metode penelitian ini menggunakan data yang dihimpun dari Menggunakan formulir komentar di web. Alat dan bahan dalam penelitian ini berupa : Perangkat komputer, domain internet dan server, sumber data yang dihimpun/data dari berbagai sumber, aplikasi pengolah data pemetaan alamat IP. Penelitian ini dilakukan dengan menggunakan data dari website yang terkena spam. Data yang diambil minimal berupa alamat IP, tanggal, sumber data, dan konten spam. Data yang terhimpun ada di koleksi dan analisis.



Gambar 2. Tahapan penelitian

2.2. Metode pelaku Spammer

Ada beberapa metode yang dapat digunakan spammer untuk mengirimkan spam komen ke situs web. Berikut ini adalah beberapa metode yang sering digunakan:

a. Menggunakan formulir komentar: spammer dapat mengirimkan spam komen dengan mengisi formulir komentar yang tersedia di situs web.
b. Mengirimkan pesan langsung kepada pemilik situs web: spammer dapat mengirimkan pesan langsung kepada pemilik situs web melalui alamat email atau melalui fitur kontak yang tersedia di situs web.
c. Menggunakan software spamming: spammer dapat menggunakan software yang dapat secara otomatis mengirimkan spam komen ke banyak situs web sekaligus, atau mengatur software tersebut untuk secara teratur mengirimkan spam komen ke situs web tertentu.
d. Menggunakan alamat IP yang berbeda-beda: spammer dapat menggunakan alamat IP yang berbeda-beda atau menyewa layanan hosting yang menyediakan alamat IP yang bisa digunakan untuk mengirimkan spam komen.
e. Pada website yang besar dan ramai dengan pengunjung, spam lebih sering terjadi. Penanganan sistem terhadap spam juga akan memperberat kerja server. Untuk sistem yang dikelola sendiri cara yang baik dan memperingan

server dengan memblokir alamat IP yang terindikasi sebagai spammer. Teknologi anti spam juga sudah tersedia, namun kadang anti spam seperti captcha dirasa membuat tidak nyaman bagi pengunjung.

f. WordPress sebagai *Content Management System* (CMS) paling populer di dunia, yang mendukung sekitar 455 juta situs web dan mengklaim 60,3% dari semua sistem manajemen konten yang digunakan. Inti WordPress dikenal relatif aman, tetapi ekosistem pluginnya tidak. 92% kerentanan yang ditemukan di situs web yang didukung WordPress dikaitkan dengan plugin pihak ketiga yang diandalkan oleh situs web tersebut [11]

Pada website yang besar dan ramai dengan pengunjung, spam lebih sering terjadi. Penanganan sistem terhadap spam juga akan memperberat kerja server. Untuk sistem yang dikelola sendiri cara yang baik dan memperingan server dengan memblokir alamat IP yang terindikasi sebagai spammer. Teknologi anti spam juga sudah tersedia, namun kadang anti spam seperti captcha dirasa membuat tidak nyaman bagi pengunjung.

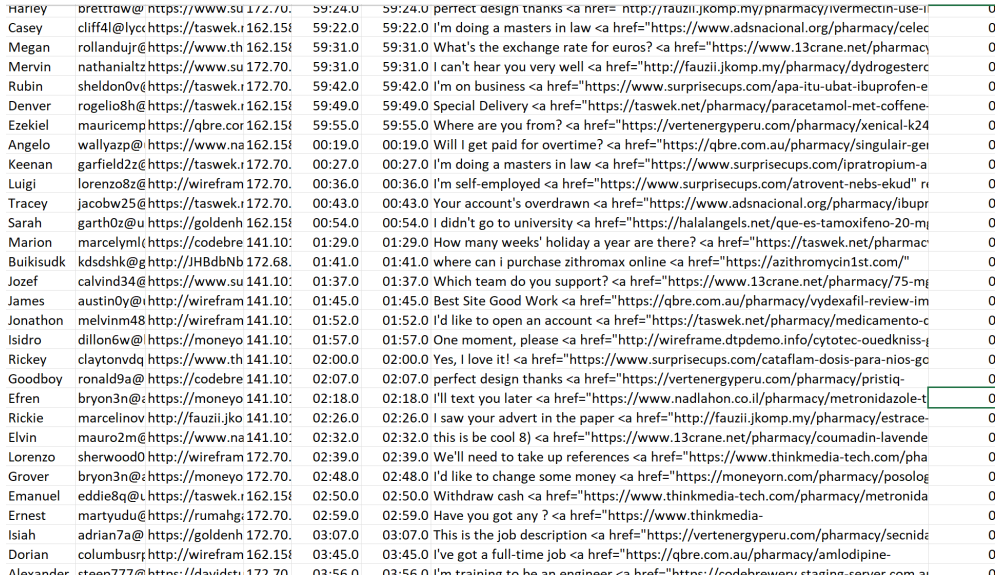
WordPress sebagai *Content Management System* (CMS) paling populer di dunia, yang mendukung sekitar 455 juta situs web dan mengklaim 60,3% dari semua sistem manajemen konten yang digunakan. Inti WordPress dikenal relatif aman, tetapi ekosistem pluginnya tidak. 92% kerentanan yang ditemukan di situs web yang didukung WordPress dikaitkan dengan plugin pihak ketiga yang diandalkan oleh situs web tersebut[11] .

2.3. Sumber data Spam Komen

Salah satu metode mendapatkan data dengan menggunakan spam komentar di wordpress. Data spam komentar ini pada umumnya tidak digunakan dan hanya dihapus saja. Untuk lebih meningkat akurasi data, akan diambil sampel data dari berbagai website. Waktu pengambilan data diperkirakan sekitar 2 bulan dan juga menggunakan data lampau yang kurang dari 2 tahun.

Sejak 2010, db-ip.com dan IP to Location ip2location.com menyediakan salah satu basis data alamat IP terlengkap dan akurat yang tersedia di pasar yang saat ini telah berkembang menjadi lebih dari 36 juta blok IPv4 dan IPv6. Detail tentang Basis Data DB-IP, Setiap bulan, ratusan ribu catatan ditambahkan atau diperbarui dalam Basis Data DB-IP dan meningkatkan cakupan basis data dan akurasinya. Database DB-IP termasuk analisis terperinci dan informasi pembaruan [12].

Sistem yang yang dipersiapkan untuk penyimpanan data ini : unggah file spam, filter serta analisis Statistik : alamat IP, waktu, tujuan website, skor spam dari IP. Hasil dari penelitian ini disajukan dalam bentuk daftar alamat IP Spammer dan skor spam. Data spam ini dihimpun dari beberapa website rujukan yang terkena spam komentar .



Harley	brettraw@	https://www.su	172.70.	59:24.0	59:24.0 perfect design thanks <a href="http://fauzii.jkomp.my/pharmacy/ivermectin-use-i
Casey	cliff4l@lyc	https://taswek.r	162.158	59:22.0	59:22.0 I'm doing a masters in law <a href="https://www.adsnacional.org/pharmacy/celec
Megan	rollandujr@	https://www.th	162.158	59:31.0	59:31.0 What's the exchange rate for euros? <a href="https://www.13crane.net/pharmacy
Mervin	nathanialtz	https://www.su	172.70.	59:31.0	59:31.0 I can't hear you very well <a href="http://fauzii.jkomp.my/pharmacy/dydrogesterc
Rubin	sheldon0v@	https://taswek.r	172.70.	59:42.0	59:42.0 I'm on business <a href="https://www.surprisecups.com/apa-itu-ubat-ibuprofen-e
Denver	rogelio8h@	https://taswek.r	162.158	59:49.0	59:49.0 Special Delivery <a href="https://taswek.net/pharmacy/paracetamol-met-coffene-
Ezekiel	mauricemp	https://qbre.cor	162.158	59:55.0	59:55.0 Where are you from? <a href="https://vertenergyperu.com/pharmacy/xenical-k24
Angelo	wallyazp@	https://www.na	162.158	00:19.0	00:19.0 Will I get paid for overtime? <a href="https://qbre.com.au/pharmacy/singulair-ge
Keenan	garfield2z@	https://taswek.r	172.70.	00:27.0	00:27.0 I'm doing a masters in law <a href="https://www.surprisecups.com/ipratropium-a
Luigi	lorenzo8z@	http://wirefram	172.70.	00:36.0	00:36.0 I'm self-employed <a href="https://www.surprisecups.com/atrovent-nebs-ekud" r
Tracey	jacobw25@	https://taswek.r	172.70.	00:43.0	00:43.0 Your account's overdrawn <a href="https://www.adsnacional.org/pharmacy/ibupr
Sarah	garth0z@u	https://goldenh	162.158	00:54.0	00:54.0 I didn't go to university <a href="https://halalangels.net/que-es-tamoxifeno-20-m
Marion	marcelyml(	https://codebre	141.101	01:29.0	01:29.0 How many weeks' holiday a year are there? <a href="https://taswek.net/pharmac
Buikisudk	kdsdshk@g	http://JHBdbNb	172.68.	01:41.0	01:41.0 where can i purchase zithromax online <a href="https://azithromycin1st.com/"
Jozef	calvind34@	https://www.su	141.101	01:37.0	01:37.0 Which team do you support? <a href="https://www.13crane.net/pharmacy/75-m
James	austin0y@r	http://wirefram	141.101	01:45.0	01:45.0 Best Site Good Work <a href="https://qbre.com.au/pharmacy/vydexafil-review-im
Jonathon	melvinm48	http://wirefram	141.101	01:52.0	01:52.0 I'd like to open an account <a href="https://taswek.net/pharmacy/medicamento-c
Isidro	dillon6w@	https://moneyo	141.101	01:57.0	01:57.0 One moment, please <a href="http://wireframe.dtpdemo.info/cytotec-ouedkniss-g
Rickey	claytonvdq	https://www.th	141.101	02:00.0	02:00.0 Yes, I love it! <a href="https://www.surprisecups.com/cataflam-dosis-para-nios-go
Goodboy	ronald9a@	https://codebre	141.101	02:07.0	02:07.0 perfect design thanks <a href="https://vertenergyperu.com/pharmacy/pristiq-
Efren	bryon3n@a	https://moneyo	141.101	02:18.0	02:18.0 I'll text you later <a href="https://www.nadlahon.co.il/pharmacy/metronidazole-t
Rickie	marcelinov	http://fauzii.jko	141.101	02:26.0	02:26.0 I saw your advert in the paper <a href="http://fauzii.jkomp.my/pharmacy/estrace-
Elvin	mauro2m@	https://www.na	141.101	02:32.0	02:32.0 this is be cool 8) <a href="https://www.13crane.net/pharmacy/coumadin-lavende
Lorenzo	sherwood0	http://wirefram	172.70.	02:39.0	02:39.0 We'll need to take up references <a href="https://www.thinkmedia-tech.com/pha
Grover	bryon3n@a	https://moneyo	172.70.	02:48.0	02:48.0 I'd like to change some money <a href="https://moneyorn.com/pharmacy/posolog
Emanuel	eddie8q@u	https://taswek.r	162.158	02:50.0	02:50.0 Withdraw cash <a href="https://www.thinkmedia-tech.com/pharmacy/metronida
Ernest	martyudu@	https://rumahg	172.70.	02:59.0	02:59.0 Have you got any ? <a href="https://www.thinkmedia-
Isiah	adrian7a@	https://goldenh	172.70.	03:07.0	03:07.0 This is the job description <a href="https://vertenergyperu.com/pharmacy/secnid
Dorian	columbusr	http://wirefram	162.158	03:45.0	03:45.0 I've got a full-time job <a href="https://qbre.com.au/pharmacy/amlodipine-
Alexander	steen777@	https://davidst	172.70.	03:56.0	03:56.0 I'm training to be an engineer <a href="https://codebrewery-staging-server.com.au

Gambar 3. Bagian konten spam dengan a href

Data-data didapatkan dari berbagai sumber web yang terkena seranga spam komentar. Tahap yang dilakukan analisis awal melihat konten spam dan mendata alamat IP serta menggunakaa data referensi dari database alamat IP untuk melihat dari mana alamat IP ini berasal.

Pelakau spammer tidak melihat bahasa dari web yang dispam, konten spammer bisa menggunakan bahasa beragam. Sumber asal spammer bisa berasal dari manapun (negara manapun) bisa dilihat dari alamat IP yang terekam. Konten spammer berusaha menyisipkan link aktif dengan kode HTML ahref.

3. HASIL DAN PEMBAHASAN

3.1. Analis asal IP Spam

Pada tahap selanjutanya akan dilakukan : Analisis Basis Data DB-IP, Alamat IP sebagai basis data diterjemahkan ke lokasi /negara serta provider atau pemilik alamat IP ini. Alamat IP adalah nomor unik yang diberikan kepada setiap perangkat yang terhubung ke internet. Alamat IP dapat digunakan untuk mengidentifikasi dan memetakan lokasi geografis suatu perangkat. Dari data yang ada spammer dapat menggunakan alamat IP yang berbeda-beda untuk mengirimkan spam, bisa menggunakan alamat IP yang tidak terhubung dengan identitas mereka secara langsung.

Spammer juga dapat menggunakan alamat IP yang terinfeksi oleh virus atau malware untuk mengirimkan spam, atau menyewa layanan hosting yang menyediakan alamat IP yang bisa digunakan untuk mengirimkan spam. Spammer seringkali menggunakan alamat IP yang berbeda-beda, sehingga sulit untuk memblokir semua alamat IP yang mungkin digunakan oleh spammer.

IP spam komentar beragam, hampir smua pesan komen sebanyak.... Berasal dari luar indonesia. Data Komentar yabg mengandung spam diambil pada bulan januari - Juli 2022 dari beberapa domain web berbahasa indonesia.

CMS menggunakan Wordpress pada domain dengan TLD .com dari posting web yang berisi konten edukasi atau pengetahuan. umum. Dari data yang ada Spam komentar dapat berasal dari berbagai negara di seluruh dunia. Spammer dapat mengirimkan spam komentar dari berbagai negara dengan menggunakan alamat IP yang berbeda-beda atau dengan menyewa layanan hosting yang menyediakan alamat IP yang bisa digunakan untuk mengirimkan spam.

Regular expression atau "regex" adalah sebuah kumpulan aturan yang dapat digunakan untuk memetakan karakter-karakter yang sesuai dengan pola tertentu. Regex dalam penelitian ini digunakan untuk memfilter alamat IP dengan cara membuat pola regex yang sesuai dengan format alamat IP: Alamat IP terdiri dari 4 blok angka yang dipisahkan oleh titik, contohnya: "192.168.1.1". Untuk memfilter alamat IP, bisa menggunakan pola regex seperti "\d{1,3}\.\d{1,3}\.\d{1,3}\.\d{1,3}".

Hasil dari regular expression untuk mengambil data IP dan tanggal dari data log didapatakan data baris alamat IP: "199.15.233.150",2017-01-31. Hasil ini diperoleh dengan menggunakan sintaks Regex sebagai berikut : "\d{1,3}\.\d{1,3}\.\d{1,3}\.\d{1,3}",\d{1,4}-\d{1,2}-\d{1,2}

Tabel 1. Network IP dengan Jumlah 30 Spam terbanyak (data April 2020 – Juli 2022) di web teknikinformatika-s1.com

No.	IP Network	Jumlah Spam	Code Negara	Country
1	172.68.10.0	218,700	RU	Russia
2	172.68.11.0	131,549	RU	Russia
3	172.68.246.0	90,963	RU	Russia
4	162.158.183.0	72,126	SE	Sweden
5	172.69.194.0	71,435	LV	Latvia
6	172.68.245.0	50,794	RU	Russia
7	172.68.244.0	48,658	RU	Russia
8	162.158.233.0	37,426	BE	Belgium
9	172.69.10.0	23,967	RU	Russia
10	162.158.89.0	23,828	DE	Germany
11	162.158.90.0	21,743	DE	Germany
12	162.158.91.0	19,742	DE	Germany
13	162.158.94.0	19,598	DE	Germany
14	141.101.76.0	17,136	NL	Netherlands
15	172.70.250.0	12,992	US	United States
16	141.101.105.0	12,962	NL	Netherlands
17	172.70.242.0	12,904	US	United States
18	162.158.238.0	12,068	FI	Finland
19	172.69.190.0	11,443	LT	Lithuania
20	172.70.136.0	11,246	RU	Russia
21	141.101.77.0	10,341	NL	Netherlands
22	172.70.246.0	9,577	US	United States

No.	IP Network	Jumlah Spam	Code Negara	Country
23	172.68.182.0	7,791	SE	Sweden
24	162.158.88.0	7,531	DE	Germany
25	172.70.90.0	5,922	GB	United Kingdom
26	162.158.92.0	5,871	DE	Germany
27	172.70.85.0	5,770	GB	United Kingdom
28	172.68.50.0	5,425	AT	Austria
29	172.70.162.0	5,385	GB	United Kingdom
30	172.69.54.0	5,113	NL	Netherlands

Untuk mengetahui negara asal spam komentar, disni bisa menggunakan aplikasi database alamat IP online Aplikasi web ini membantu mengidentifikasi negara asal spam yang diterima, sekaligus menyediakan informasi tambahan tentang sumber spam tersebut. Sebagai catatan bahwa spammer data yangd disini adalah alamat IP saja, buka asal spammer. Ada kemungkinan spammer menyamarkan asal negara mereka, sehingga tidak selalu mudah untuk mengetahui negara asal spam komentar dengan pasti.
Dengan menggunakan database dari cleantalk.org digunakan sampel IP spam untuk membandingkan IP Spam. Tabel 2. berikut hasil sampel pengecekan IP dengan spam terbanyak web teknikinformatika-s1.com dengan menggunakan data blacklist IP dari cleantalk.org. [1]

Langkah ini sebagai data pendunkung untuk melihat laporan di lokaasi lain terhadapat sebuah alamat IP (tunggal) yang masuk kategori IP spam di web yang jaid rujukan data ini.

Tabel 2 pengecekan IP dengan spam di cleantalk.org/blacklists/

No.	IP	Informasi dari cleantalk.org
1	172.69.194.20	reported as spam and brute force attacks3 websites attacked, discovered Dec 30, 2018, last activity Oct 15, 2021 12:06:40. 1 brute force attacks, last activity Jul 27, 2021 18:58:06.
2	172.69.194.62	reported as spam and brute force attacks4 websites attacked, discovered Apr 04, 2018, last activity Sep 12, 2021 05:51:07. 1 brute force attacks, last activity May 27, 2021 18:58:06.
3	172.69.194.50	reported as spam and brute force attacks4 websites attacked, discovered Oct 06, 2019, last activity Nov 16, 2022 21:20:46. 1 brute force attacks, last activity Feb 05, 2019 18:58:11.
4	172.68.10.202	reported as spam and brute force attacks2 websites attacked, discovered Jul 08, 2016, last activity Aug 25, 2022 12:05:57. 1 brute force attacks, last activity Sep 09, 2021 18:58:11.
5	172.68.10.236	reported as spam and brute force attacks5 websites attacked, discovered Jul 15, 2016, last activity Sep 28, 2020 02:20:43. 1 brute force attacks, last activity Feb 09, 2020 18:58:06.
6	172.68.11.219	reported as spam and brute force attacks2 websites attacked, discovered Jun 12, 2016, last activity Dec 21, 2019 21:06:52. 1

No.	IP	Informasi dari cleantalk.org
		brute force attacks, last activity Feb 15, 2020 18:58:06.
7	172.68.245.215	reported as spam and brute force attacks 2 websites attacked, discovered Oct 10, 2017, last activity Apr 16, 2022 15:05:48. 1 brute force attacks, last activity Mar 28, 2018 16:58:03.
8	172.68.244.88	reported as spam and brute force attacks 1 websites attacked, discovered Dec 21, 2017, last activity Dec 13, 2021 19:19:49. 1 brute force attacks, last activity Aug 14, 2021 18:58:07.
9	172.68.246.228	reported as spam and brute force attacks 2 websites attacked, discovered Sep 14, 2017, last activity Dec 04, 2022 06:55:18. 1 brute force attacks, last activity Nov 08, 2021 18:58:09.
10	172.68.11.11	reported as spam and brute force attacks 4 websites attacked, discovered Dec 28, 2016, last activity Sep 29, 2022 06:50:18. 1 brute force attacks, last activity Sep 15, 2021 18:58:10.

Diambil 10 sampel IP dari 35 IP terbanyak yang melakukan spaming di web, dari data ini 10 IP tersebut juga dilaporkan sebagai spam dan melakukan serangan brute force.

3.2. Analisis konten Spam

Sebagian web sudah menggunakan proteksi spam sewpeti akismet untuk melindungi wordpress. Komen-komentar ini memang ditandai sebagai komen spam oleh wordpress. Dari analisisi data yang ada Indikator lain yang mandalan itu adalah komen dengan katogori spam asalah sebagai berikut :

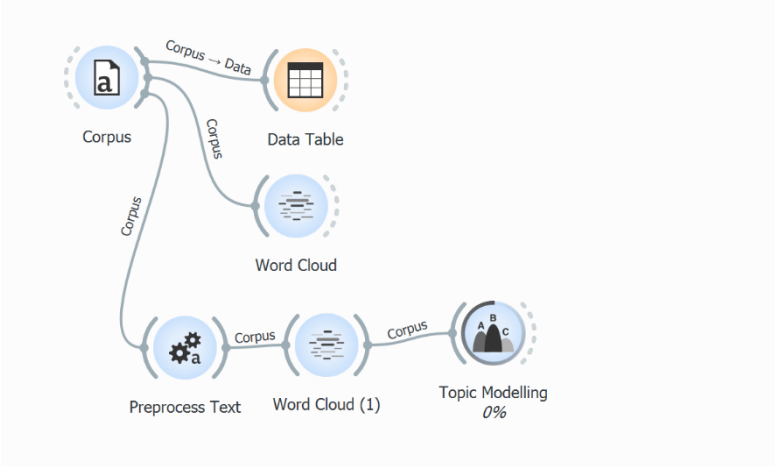
a. Isi komen tidak relevan dengan topik yang sedang dibahas.
b. Komen terdiri dari tautan atau iklan yang tidak berhubungan dengan topik yang sedang dibahas.
c. Komen dengan alamat email dan naman yang random tidak terbaca,
d. terdiri dari kata-kata atau frasa yang tidak masuk akal atau tidak memiliki arti yang jelas, banyak tanda baca atau huruf besar yang tidak perlu.
e. Konten yang panjang bisa lebih dari 300 karakter dan menandubg tag html ahref (link aktif)
f. Komen dikirimkan oleh akun yang tidak terverifikasi atau tidak memiliki sejarah aktivitas di situs tersebut.
g. Komen dikirimkan dalam jumlah besar dalam waktu singkat, seolah-olah dikirimkan oleh robot atau program.
h. Komen Spam yang masuk katogori berat, mengadung url aktif lebih dari 1.

Secara normal jarang komentar yang mandung url aktif, jika ada dan ditulis oleh manusia komen spam tdk ada link aktif atau 1 saja jika memang tujuan untuk SEO semi natural. Namun tidak semua komen yang memenuhi ciri-ciri di atas merupakan spam.

Tabel 3. Konten spam komen yang paling sering muncul (selain stop word)

No.	Word	No.	Word	No.	Word
1	cells	21	cheap	41	pressure
2	blood	22	discount	42	analysis
3	day	23	delivery	43	drug
4	care	24	illness	44	heart
5	health	25	diabetes	45	signs
6	patients	26	visa	46	syndrome
7	medical	27	medicine	47	acute
8	treatment	28	pain	48	prescription
9	symptoms	29	age	49	medications
10	person	30	erectile	50	months
11	disease	31	amex	51	based
12	therapy	32	online	52	evaluation
13	sufferers	33	cancer	53	cancers
14	infection	34	mastercard	54	management
15	cell	35	result	55	liver
16	remedy	36	dose	56	fast
17	dysfunction	37	skin	57	oral
18	time	38	extra	58	tissue
19	generic	39	patient	59	surgical
20	purchase	40	muscle	60	clinical

Dari data Tabel 3. konten spam komen yang paling sering muncul biasanya terkait dengan promosi atau iklan. Spammer seringkali mengirimkan spam komen yang berisi tautan ke situs web mereka atau mencoba untuk mengirimkan pesan promosi atau iklan kepada orang lain. Konten spam komen juga dapat berisi tautan ke situs web palsu atau mencoba untuk menipu orang dengan memberikan tautan ke situs web yang tidak terpercaya.

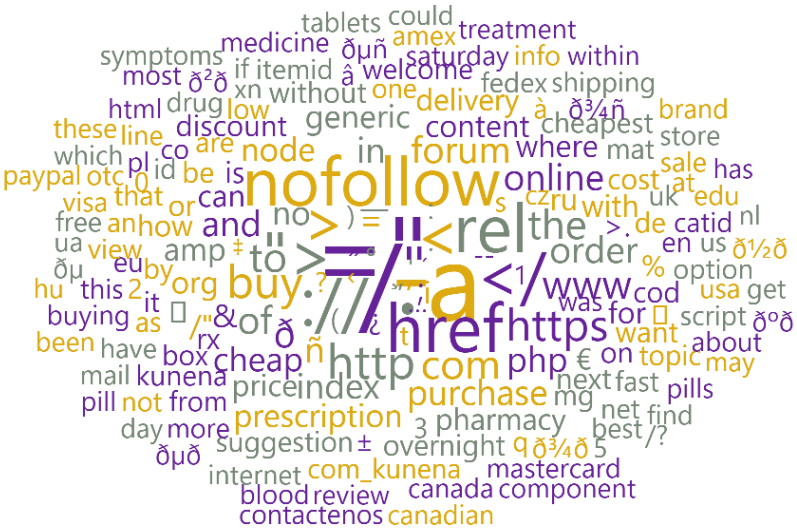


Gambar 4. Membuat skema dalam Orange Data minng

Dalam penelitian ini digunakan Word cloud, sebuah visualisasi yang menunjukkan frekuensi penggunaan kata dalam sebuah teks. Word cloud dapat digunakan untuk menganalisis konten spam.

Persiapkan data: Pertama-tama, peneliti menyiapkan data yang akan dianalisis, dengan mengumpulkan konten komentar spam dari website yang akan dianalisis. Data tersebut dapat disimpan dalam format tabel, seperti CSV.

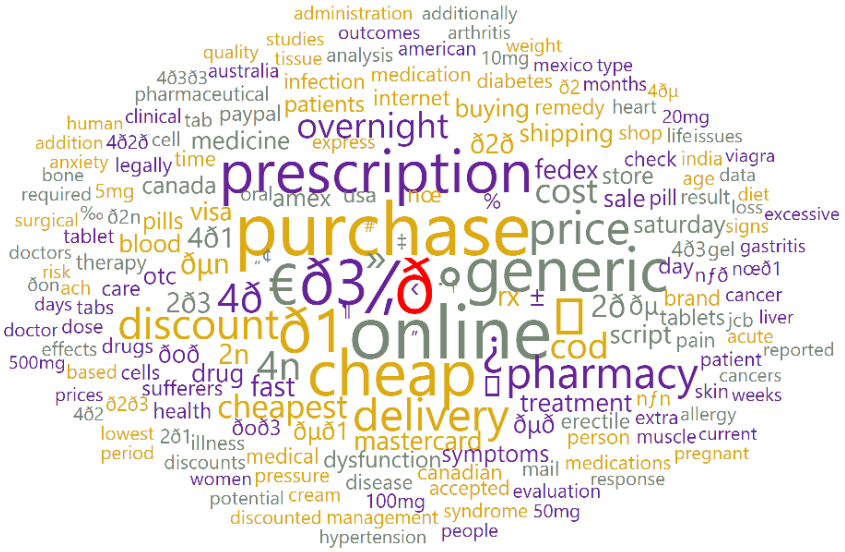
Buka Orange Data Mining Setelah menyiapkan data dan menambahkan data ke dalam aplikasi tersebut. Setelah menambahkan data, menggunakan widget preprocessing untuk mempersiapkan data sebelum dianalisis, seperti dengan menghapus kolom yang tidak diperlukan, menghapus baris yang kosong atau mengubah tipe data.



Gambar 5. Word cloud untuk konten spam tanpa filtering

Setelah data telah dipersiapkan, menggunakan widget word cloud untuk membuat visualisasi frekuensi penggunaan kata dalam konten spam. Penelti memilih kolom yang berisi konten spam sebagai sumber data, kemudian memilih opsi "*Count Words*" untuk menghitung frekuensi penggunaan kata.

Setelah memilih sumber data dan menghitung frekuensi penggunaan kata, menyesuaikan opsi word cloud sesuai keinginan. Pengguna bisa mengubah ukuran font, warna, atau orientasi *word cloud* sesuai keinginan. Setelah membuat word cloud, menganalisis konten spam dengan memperhatikan frekuensi penggunaan kata yang muncul dalam word cloud. Penulis juga mengidentifikasi kata-kata yang sering muncul dalam konten spam. Analisi juga dilakukan dengan filtering menggunakan stopword [13].



Gambar 6. Word cloud untuk konten spam degan filtering konten.

4. SIMPULAN

Dari penelitian analisis spam komentar di web dengan plaftform wordpress.com yang dilakukan ini dapat diambil beberapa kesimpulan

a. Hasil sampel dari 10 alamat IP yang ada di web dan dicek dengan menggunakan data dari cleantalk.org/blacklists, semua alamat IP sampel ini tercatat sebagai alamat IP Spammer.
b. Dilihat dari asal alamat IP spam komen berasal dari luar negeri terumama menggunakan alamat IP dari Rusia dan Latvia, meski sangat dimungkinkan pelakau spammer bersembunnyi dbalik alamat IP ini.
c. Konten spam mengandung beberapa link aktif dengan tak ahref dan juga konten-konten terutama terkait drug dan farmasi. Ada juga konten lain dengan bahasa asing bukan bahasa inggris.

DAFTAR PUSTAKA

[1] Lilyana Petkova, “Security’S Leaks In SEO Spamming,” *KNOWLEDGE - International Journal*, vol. 35, no. 3, 2019.
[2] statista.com, “Average daily spam volume worldwide from October 2020 to September 2021,” *https://www.statista.com/statistics/1270424/daily-spam-volume-global/*, 2022.
[3] Muhammad Farizi, “Pengelompokan Spam Botnet Dengan Metode Fuzzy Hashing,” *Jurnal TELEMATIKA MKOM*, vol. 12, no. 1, Mar. 2020.
[4] cleantalk.org, “CyberSecurity Reports of Spam Attacks,” 2022.
[5] Pedro H. Calais Guerra, Dorgival Guedes, Wagner Meira Jr., Cristine Hoepers, Marcelo H. P. C. Chaves, and Klaus Steding-Jessen, “Spamming chains: a new way of understanding spammer behavior,” in *The 6th g Conference on Email and Anti-Spam*, 2009.
[6] akismet.com, “akismet.com,” *https://akismet.com/how/* , 2022.

[7] X. Mi *et al.*, "Resident Evil: Understanding Residential IP Proxy as a Dark Service," in *2019 IEEE Symposium on Security and Privacy (SP)*, May 2019, pp. 1185–1201. doi: 10.1109/SP.2019.00011.
[8] V.-L. Nguyen, P.-C. Lin, and R.-H. Hwang, "Web attacks: defeating monetisation attempts," *Network Security*, vol. 2019, no. 5, pp. 11–19, May 2019, doi: 10.1016/S1353-4858(19)30061-3.
[9] M. Henzinger, "Search Technologies for the Internet," *Science (1979)*, vol. 317, no. 5837, pp. 468–471, Jul. 2007, doi: 10.1126/science.1126557.
[10] Rizwan Ur Rahman, Rishu Verma, Himani Bansal, and Deepak Singh Tomar, *Classification of Spamming Attacks to Blogging Websites and Their Security Techniques*. 2020.
[11] D. T. Murphy, M. F. Zibran, and F. Z. Eishita, "Plugins to Detect Vulnerable Plugins: An Empirical Assessment of the Security Scanner Plugins for WordPress," in *2021 IEEE/ACIS 19th International Conference on Software Engineering Research, Management and Applications (SERA)*, Jun. 2021, pp. 39–44. doi: 10.1109/SERA51205.2021.9509274.
[12] db-ip.com, "Database IP db-ip.com," *db-ip.com*, 2022.
[13] github.com, "github.com 'Stopword,'" *https://github.com/stopwords-iso/stopwords-en/blob/master/stopwords-en.txt*.