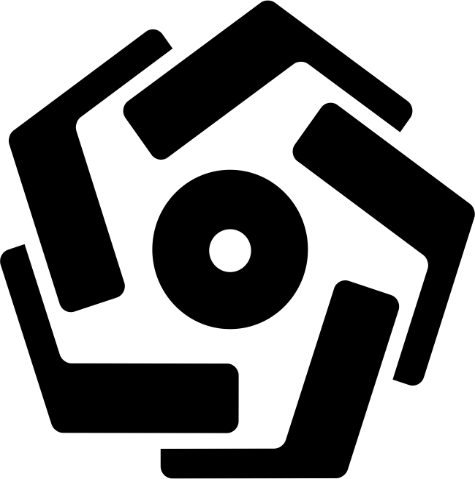
**SELEKSI FITUR BERBASIS INFORMATION GAIN PADA KINERJA ALGORITMA KLASIFIKASI PADA DETEKSI WEB PHISHING**

**SKRIPSI**



disusun oleh

Anggun Wahyu Andriyanto

17.11.1732

**PROGRAM SARJANA**

**PROGRAM STUDIINFORMATIKA**

**UNIVERSITAS AMIKOM YOGYAKARTA**

**YOGYAKARTA**

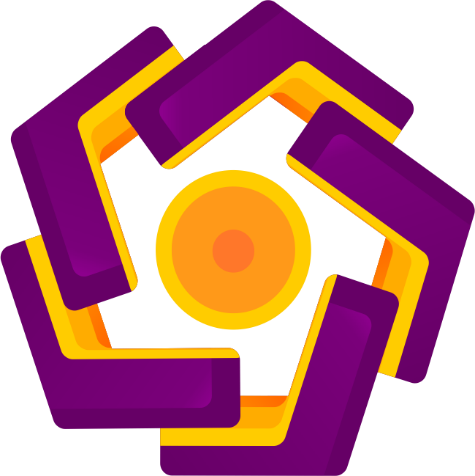
**2020**

**SELEKSI FITUR BERBASIS INFORMATION GAIN PADA KINERJA ALGORITMA KLASIFIKASI PADA DETEKSI WEB PHISHING**

**SKRIPSI**

untukmemenuhisebagianpersyaratan  
mencapaigelarSarjana

padaProgram StudiSistemInformasi



disusun oleh

Anggun Wahyu Andriyanto

17.11.1732

**PROGRAM SARJANA**

**PROGRAM STUDISISTEM INFORMASI**

**SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER**

**AMIKOM YOGYAKARTA**

**YOGYAKARTA**

**2020**

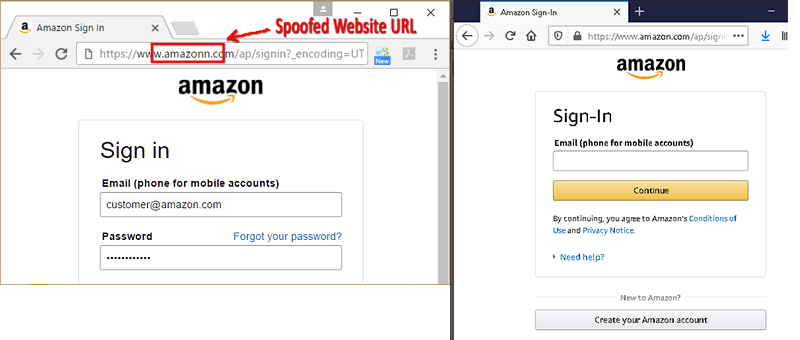
**BAB I**  
Pendahuluan

1. 1. **LatarBelakang**

Perkembangan internet daritahunketahunmembuatfasilitas yang ditawarkannyamenjadilebihbanyak. Dari segikemudahanakseshinggabanyaknyafitur yang ditawarkanpastinyaakanmudahmembuat orang-orang tertarikuntukmenggunakannya, sehinggatidakdipungkirijikapenggunanyatidaklahsedikit. Sebagaicontohnyaterdapat media sosial, internet atau mobile banking. Duahaltersebuttelahmemberikankemudahanbagi orang-orang untuksalingberhubunganmeskidipisahkanolehjarakdanmembantu proses transaksi, yang mana bisadiaksesdandilakukandari mana sajasecara onlinedan*realtime*. Namundibalikkemudahantersebut, bisamenjadicelahkeamananbagipengguna internet dalambertransaksisecara online maupunmengakses media sosialseperti data pribadi, kata sandi, e-mail, informasisensitif internet banking atau mobile banking. Hal tersebutbiasanyadilakukanolehpelakudenganbantuanmenggunakan situs phising.

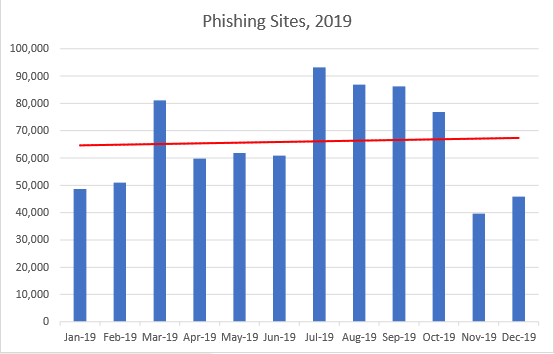
Phisingadalahsuatumetodeuntukmelakukanpenipuandenganmengelabui target denganmaksuduntukmencuriakun target. Istilahiniberasaldari kata “fishing” = “memancing” korban untukterperangkapdijebakannya.Sedangkan situs phisingadalah situs yang dibuatsemiripmungkindengan situs aslinyadandigunakanuntukmengelabuilalumenjebak korban yang mengaksesnya[1]. Situs phisingmempunyaitampilan yang dibuatsemiripmungkindengan situs aslinyasehingga korban akanterkelabuidanyakinbahwa situs tersebutadalah situs yang sebenarnya. Jika korban mengakses situs tersebutdengan*username*atau e-mail dan kata sandiakun korban makainformasitersebutakandicuriolehpembuat situs phising, bahkanbila korban terpancinguntukmemasukkaninformasisensitiftentang banking onlinenyamakapelakudapatdenganmudahmendapatkanakseskeakunkeuangan korban.

Phising juga terjadipada media sosialseperti Facebook, Gmail, dan Twitter. Dimana situs phising yang dibuatakanmemberikaninformasilayaknya situs aslinyakemudianketikainformasipenggunadimasukkanpadakolom*login,* situs tersebutakanmengarahke situs phisingdaripadake situs aslinya[2]. Dibawahiniadalahperbandingan situs phisingdengan situs non-phisingyaitu situs jualbeli online Amazon, pada situs phisingberalamatkan di [www.amazonn.com](http://www.amazonn.com) :



Gambar 1.1 Perbandingan situs phising (kiri) dengan non-phising (kanan)

Terlihatsekalibahwa situs phishing tersebutsangatmiripdengan situs aslinya, jikatidaktelitimakaakansulitmembedakannya. Dari laporan APWG (*Anti-Phishing Working Group*)[3], tahun 2019 menjadinaikturunnyaperbuatan phishing, meskipunbegitutetapsajaterjadipeningkatanpesatpadabulan-bulantertentu. Dari laporan APWG kuartalkeempat 2019 (Gambar 1.2),padabulanOktober – Desemberada 162.555 situs web yang terdeteksioleh APWG. Hasiltersebutturundari 266.387 padabulanJuli – September.



Gambar 1.2 Grafik*phishing activity* padatahun 2019

Hal tersebutbisamenimbulkanketidakpercayaanpengguna internet pada situs media sosialmaupun situs transaksi. Olehsebabitudibutuhansistem yang mampumendeteksi situs phishing untukmencegahkriminalitas phishing sehinggamengurangidanbisamenghindarikerugian situs phishing terhadappengguna internet. Dari beberapapenelitiansebelumnyadigunakanmetodeklasifikasiuntukdeteksisitus phishing dan non-phishing sehinggadapatdilakukanpencegahanterkenaserangan phishing. Metodeklasifikasimenggunakansemuafitur yang terdapatpadasebuah data untukmembangun model,namunmenurut[4]tidaksemuafiturtersebutrelevanterhadaphasilklasifikasi. Apabilahaltersebutpada data yang memilikiukuranbesar, makakinerjaklasifikasiakanmenjaditidakefisiendanefektif, misalkanbeban yang dilakukanolehkomputeruntukmelakukankomputasimenjadilebihberat, waktupemrosesanmenjadilebih lama.

Olehkarenaitupadapenelitianiniakanditerapkanseleksiatribut yang berbasis Information Gain terhadapkinerjaalgoritmaklasifikasiyang popular yaitu Decision Tree, Naïve Bayes dan K-NN. Dengandemikiankinerjaklasifikasi yang dihasilkandapatterukursecarasistematissetelahmelaluiseleksiatributatauseleksifitur, sehinggahasilpenelitianinidapatdigunakanuntukmenunjangpenelitian lain yang sejenisdikemudianhari. Tentunyaapabila model klasifikasidaripenelitianinidiimplementasikan, makadapatmengurangiresikopengguna internet terkenaserangan phishing.

* 1. **RumusanMasalah**

Berdasarkanlatarbelakang yang telahdisampaikan, di bawahiniadalahrumusanmasalahatausesuatu yang menjadipertanyaan-pertanyaan yang harusdiselesaikanolehpenilitipadapenelitian yang akandibuat :

1. Apakahseleksifiturmempengaruhikinerjaalgoritmaklasifikasi?
2. Fitur-fiturapasaja yang relevanterhadaphasilklasifikasi?
3. Adakahpeningkatankinerjaklasifikasisetelahdilakukanseleksiatribut?
4. Adakahpenurunankinerjaklasifikasisetelahdilakukanseleksiatribut?
   1. **BatasanMasalah**

Untukmembatasipermasalahan yang lebihluaspadakonteksobjek, makaruanglingkupdalampenelitianiniakandibatasisesuaikebutuhandankemampuan. Batasanmasalahdalampenelitianiniantara lain:

1. Data situs phishing didapatkandarirepository UCI Machine Learning : Phishing Websites Data SetdanPhishThank.
2. Dalampenelitianiniakandigunakan Information Gain untukseleksiatributataufiturpada dataset kemudiandilakukanklasifikasimenggunakanalgoritmaklasifikasi. Denganbatashasilakurasisetelahseleksiatribut.
3. Metodeklasifikasi yang digunakandalampenelitianiniadalah Decision Tree, Naïve Bayes dan K-NN. Denganbatashasilakurasidanhasilpenerapan
   1. **MaksuddanTujuanPenelitian**

Di bawahiniadalahtujuandaripenelitianini, yaitu:

1. Tujuandaripenelitianiniadalahmenerapkanseleksifiturberbasis Information Gain pada dataset latih, kemudiandilakukanklasifikasimenggunakanalgoritmaklasifikasi.
2. Tujuandaripenelitianiniadalahuntukmengetahuikinerjaklasifikasisetelahditerapkanseleksiatribut yang relevanmenggunakan Information Gain, apakahmeningkatataumenurun.
3. PemilihanalgoritmaDecision Tre, Naïve bayes, K-NNdilakukanuntukpembandingmanakah yang hasilnya paling akurat.
   1. **ManfaatPenelitian**

Adapunmanfaat yang diharapkandaripenelitaniniadalah:

1. Pemrosesanklasifikasipadadeteksi web phishing menjadilebihefektifdanefisiensetelahdilakukanseleksiatribut.
2. Mengetahuiatributataufiturapasaja yang relevanpada data set.
3. Mengetahuipengaruhpenerapanseleksifiturpadakinerjaklasifikasi.
   1. **MetodePenelitian**

Dalampenelitianini, penelitimenggunakan data yang berasaldari repository UCI Machine Learning bagian Phishing Websites Data Set. Data tersebut yang nantinyaakandigunakanuntukmengujiketigametodeklasifikasi yang sudahdilakukanseleksiatribut. Sehinggahasilujitersebutbisadibandingkan mana yang memilikihasilakurasitertinggidan juga berapapeningkatan yang terjadipadaketigaklasifikasitersebut.

1. **SistematikaPenulisan**

Padabagianinidituliskanurut-urutandansistematikapenulisan yang dilakukan. Berikanringkasanmengenaiisimasing-masingbab.

* **BAB I PENDAHULUAN**

Padababinidijelaskanmengenailatarbelakang, rumusanmasalah, batasanmasalah, maksuddantujuanpenelitian, manfaatpenelitian, metodepenelitan, dansistematikapenulisandalampenelitan.

* **BAB II LANDASAN TEORI**

Padababiniakandijelaskanmengenailandasanteori-teoridankajianpustakadariberbagaipenelitian yang memilikiketerkaitandenganpenelitianini. Haltersebutbergunauntukmemperkuatdasar, analisa, penulisandanalasandilakukannyapenelitianini. Sumberdarilandasarteoriini juga berasaldaribuku, jurnal, blog, maupunlaporan yang berasaldari internet.

* **BAB III METODOLOGI PENELITIAN**

Padababiniakandijelaskanmengenailangkah-langkahpenelitianbesertametode yang digunakan.

* **BAB IV HASIL DAN PEMBAHASAN**

Padababiniakandilakukanperancangansistemsertapembahasannya. Kemudian di babini juga hasilpenelitianakandidapatkandandilakukanpembahasannya.

* **BAB V KESIMPULAN**

Padababiniberisikesimpulandaripenelitianinidan juga saran bagipenelitianmendatang yang berasaldarikekurangandaripenelitian ini.

Daftarpustaka

1. Y. Ping, G. Yuxiang, Z. Futai, Y. Yao, W. Wei, Z. Ting. “Web Phishing Detection Using a Deep Learning Framework”, Wireless Communication and Mobile Computing, Vol. 2018, Article ID 4678746, 9 pages, 2018. https://doi.org/10.1155/2018/4678746
2. G. Varshney, M. Misra, P. K. Atrey, “A Survey and Classification of Web Phishing Detection Schemes”, Wiley Online Library., Vol. 9, No. 18, hal. 6266-6284, Okt 2016.
3. APWG, “Phishing Activity Trends Report, 4th Quarter 2019.” 2020.
4. H. N. Firqiani, A. Kustiyo, E. P. Giri“SeleksiFiturMenggunakan Fast Correlation Based Filter PadaAlgoritma Voting Feature Interval 5”, InstitutPertanian Bogor, 2008.
5. N. Abdelhamid, A. Ayesh, F. Thabtah “Phishing Detection Based Associative Classification Data Mining.”, ELSEVIER, Vol. 41, Issue 13, hal 5498-5959, Okt 2014.