BAB I
PENDAHULUAN

1.1 Latar Belakang

Kemajuan teknologi informasi dalam segala bidang membuat kebutuhan manusia terhadap informasi semakin mudah didapatkan. Hal tersebut dapat dilihat dari mudahnya manusia untuk mencari informasi melalui internet. Internet merupakan singkatan dari *Interconected Networking*, yaitu suatu jaringan komputer yang terhubung dengan luas. Jaringan tersebut berawal dari ARPANET yang dibentuk oleh Departemen Pertahanan Amerika Serikat. Kemudian dikembangkan sampai sekarang dan menjadi tulang punggung global sumber daya informasi yang disebut internet (Hadi, 1999).

Internet tidak hanya dibutuhkan oleh perorangan, tapi juga dibutuhkan oleh organisasi atau instansi yang melakukan bisnis online, dan organisasi ini biasanya menawarkan perdagangan online (Jiming Liu, 2001). Dengan kemudahan tersebut, pengguna internet dapat mencari informasi dan dapat melakukan transaksi online secara praktis. Namun, pengguna internet dapat rentan terhadap beberapa ancaman penggunaan *website* yang dapat menyebabkan kerugian finansial, pencurian identitas, kehilangan informasi pribadi, kerugian nama baik dan kehilangan kepercayaan pelanggan bagi *e-commerce* dan *e-banking*. Salah satu bentuk pencurian informasi yang dilakukan di internet adalah *phishing*.

Phishing merupakan mekanisme kejahatan yang dilakukan dengan teknik *social engineering* dan teknik *subterfuge* untuk mencuri identitas pengguna dan *financial account credentials*. Skema teknik *social engineering* adalah menggunakan *instant message, chat,* atau *e-mail* palsu yang mengaku sebagai instansi atau lembaga yang sah, dan mengarahkan pengguna ke *website* palsu untuk membocorkan data pribadinya seperti *username* dan *password*. Skema teknik *subterfuge* dilakukan dengan menanamkan *crimeware* ke perangkat target untuk mencuri informasi secara langsung, menghadang akses *login* pengguna dan

memberikan *form* palsu untuk mendapatkan *username* dan *password* (Greg Aaron, 2016).

Phishing websites telah menjadi sebuah permasalahan yang serius, tidak hanya karena meningkatnya jumlah dari *websites* ini, tapi juga strategi yang semakin pintar untuk meniru dari *websites* yang sebenarnya. Pertumbuhan *phishing websites* pada kuartal akhir tahun 2015 mengalami peningkatan yang tinggi hingga lebih dari 21.000 *websites* antara bulan November dan Desember, dan untuk keseluruhan selama bulan Oktober sampai bulan Desember telah ditemukan 158.574 *phishing websites* yang terdeteksi (Greg Aaron, 2016).

Secara teknis, ada dua pendekatan yang dapat dilakukan untuk mengidentifikasi *phishing websites*. Pertama adalah berdasarkan *blacklists* (Nuttapong Sanglerdsinlapachai, 2010), URL yang dicurigai dibandingkan dengan URL yang ada dalam *blacklists*. Kelemahan dari pendekatan ini adalah *phishing website* baru yang tidak ada dalam *blacklists* tidak bisa dideteksi. Salah satu penyedia layanan *anti-phishing* dengan pendekatan ini adalah Netcraft anti-phishing toolbar. Netcraft toolbar menentukan kemungkinan *phishing* dari *website* yang dikunjungi dengan menentukan berapa lama domain telah terdaftar. Hal itu dilakukan menggunakan *database websites* yang dikelola oleh perusahaan. Kelemahannya adalah bahwa pendekatan ini tidak dapat mengenali *phishing websites* baru yang tidak ada dalam *database* (V Pavankumar, 2011). Netcraft toolbar bergantung pada *blacklists* untuk memblokir *phishing websites* dari pengguna, *blacklists* memiliki kerentanan antara *phishing website* yang baru muncul dan penambahan *website* ke dalam *blacklists* yang membuat *phishing websites* tidak diblokir (P. Likarish, 2008).

Pendekatan kedua adalah *heuristic-based* (Sophie Gastellier-Prevost, 2011), dimana beberapa parameter dikumpulkan dari *website* untuk mengklasifikasikan *website* tersebut ke dalam *phishing websites* atau *website* yang sah. Kelebihannya adalah *heuristic-based* dapat mengenali *phishing websites* yang baru dibuat secara *real time*. Akurasi dari *heuristic-based* tergantung pada keunikan parameter yang didapatkan dari *website*. SpoofGuard merupakan salah satu *tool* untuk membantu memprediksi *phishing website* yang menggunakan pendekatan *heuristic-based*.

SpoofGuard adalah sebuah *toolbar* untuk Internet Explorer, yang menentukan sebuah *website* merupakan *phishing* atau tidak dengan memeriksa URL, gambar, *link, passwords* dan usia dari domain (P. Likarish, 2008).

Penelitian sebelumnya mengenai identifikasi *phishing websites* dengan pendekatan *heuristic-based* telah dilakukan oleh Rami M. Mohammad dkk (2014) dengan judul "*Intelligent Rule-Based Phishing Websites Classification"*. Penelitian ini menggunakan beberapa metode dalam *data mining* seperti C4.5, RIPPER, PRISM dan CBA. Sebanyak 19 parameter digunakan dalam penelitian untuk melakukan pengujian fitur seleksi dan tingkat akurasi. Penelitian yang dilakukan menghasilkan sembilan parameter yang paling efektif dalam memprediksi *phishing websites*, yaitu *IP Address, URL Length, Prefix or Suffix, Sub Domain, SSL, Request URL, URL Anchor, Age of Domain* dan *Website Traffic,* dengan metode CBA terbukti sebagai metode terbaik yang memiliki *error-rate* paling kecil yaitu 4.75%.

Penelitian selanjutnya dilakukan oleh Slamet Widodo dengan judul "Klasifikasi Situs *Phishing* dengan Menggunakan Neural Network dan K-Nearest Network". Penelitian ini menggunakan *backpropagation* sebagai metode dari *neural network* dan *k-nearest neighbor* sebagai metode dari *data mining*. Penelitian ini membuktikan bahwa *backpropagation* mampu memberikan hasil akurasi klasifikasi sebesar 91.21%, lebih besar dibanding menggunakan metode *k-nn* dengan hasil sebesar 90.33% (Widodo, 2017).

Berdasarkan perbandingan dua penelitian sebelumnya, *data mining* merupakan algoritma yang banyak digunakan untuk melakukan pengujian klasifikasi *phishing websites* dan menghasilkan tingkat akurasi yang baik. Pada penelitian pertama, algoritma *data mining* menunjukkan tingkat akurasi yang tinggi dan menghasilkan parameter yang efektif melalui pengujian fitur seleksi yang dilakukan. Pada penelitian kedua, algoritma *data mining* yaitu *K-Nearest Neighbor* memiliki tingkat akurasi yang sedikit lebih rendah dari pada algoritma lawannya yaitu *backpropagation*. Penelitian tersebut menggunakan 17 parameter.

Penelitian mengenai *data mining* lainnya telah dilakukan sebelumnya oleh Parvin dengan judul "*A Modification on K-Nearest Neighbor Classifier"*. Penelitian mengusulkan sebuah metode modifikasi dari *K-Nearest Neighbor* yang disebut

Modified K-Nearest Neighbor. Penelitian ini melakukan pengujian dan evaluasi terhadap sembilan dataset berbeda yang diperoleh dari UCI *datasets*. Evaluasi yang dilakukan menunjukkan bahwa MKNN memberikan peningkatan hasil akurasi yang sangat baik dari pada KNN dengan nilai K yang digunakan adalah 3, 5 dan 7.

Penelitian menggunakan MKNN lainnya adalah "Implementasi Algoritma *Modified K-Nearest Neighbor* (MKNN) untuk Klasifikasi Penyakit Demam" yang dilakukan oleh Fakihatin. Penelitian ini melakukan klasifikasi terhadap penyakit demam berdasarkan 15 gejala penyakit. Hasil rata-rata akurasi dari penelitian ini adalah 96.35%. Kemudian, penelitian dengan judul "Implementasi *Modified K-Nearest Neighbor* Dengan Otomatisasi Nilai K Pada Pengklasifikasian Penyakit Tanaman Kedelai" yang dilakukan oleh Tri Halomoan Simanjuntak. Penelitian ini menunjukkan bahwa parameter nilai K sangat berpengaruh terhadap hasil klasifikasi, rata-rata akurasi cenderung menurun dengan bertambahnya nilai K yang digunakan. Rata-rata akurasi pada penelitian ini adalah 98.83%.

Berdasarkan latar belakang yang telah disebutkan, maka penulis tertarik melakukan penelitian untuk mengklasifikasikan *phishing websites* dengan pendekatan *heuristic-based* menggunakan metode *Modified K-Nearest Neighbor* (MKNN) dengan harapan hasil akurasi klasifikasi akan lebih baik dari pada penelitian sebelumnya. Penelitian ini menggunakan sembilan parameter pada penelitian sebelumnya yang telah terbukti efektif dalam mengklasifikasikan *phishing websites*.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah disebutkan, maka dibuatlah rumusan masalah seperti berikut:

1. Bagaimana menerapkan metode *Modified K-Nearest Neighbor* (MKNN) untuk mengklasifikasikan *phising websites*.
2. Mengukur tingkat akurasi pengklasifikasian *phising websites* menggunakan metode *Modified K-Nearest Neighbor* (MKNN).

1.3 Batasan Masalah

Beberapa hal yang menjadi batasan masalah dalam penelitian yang dilakukan adalah sebagai berikut:

1. Penelitian ini menggunakan *phishing websites datasets* dari UCI Machine Learning Repository.
2. Parameter yang digunakan adalah: *IP Address, URL Length, Prefix or Suffix, Sub Domain, SSL, Request URL, URL Anchor, Age of Domain, Website Traffic*.
3. Kelas yang digunakan dalam klasifikasi *phishing websites* adalah: *Legitimate* dan *Phishing.*
4. Nilai parameter k yang digunakan adalah bilangan ganjil 1 sampai 11.
5. Perbandingan data latih dan data uji yang diterapkan adalah 70:30, 80:20 dan 90:10.

1.4 Tujuan

Tujuan dalam penelitian ini adalah mengetahui akurasi dari klasifikasi *phishing websites* menggunakan metode *Modified K-Nearest Neighbor* (MKNN) dan mengetahui pengaruh nilai K terhadap hasil klasifikasi.

1.5 Sistematika Penulisan

Sistematika penulisan pada laporan tugas akhir ini adalah sebagai beriku:

BAB I PENDAHULUAN

Bab ini memberikan gambaran penelitian secara keseluruhan yang meliputi latar belakang, rumusan masalah, batasan masalah, tujuan dan sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini menjelaskan teori dasar dan teori pendukung yang digunakan dalam penelitian. Teori tersebut berupa definisi, konsep dan rumus-rumus yang digunakan untuk menyelesaikan masalah dalam penelitian.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan langkah-langkah yang dilakukan dalam proses penelitian.

BAB IV ANALISA DAN PERANCANGAN

Bab ini berisi pembahasan tentang analisa sistem yang akan dibangun, meliputi analisa kebutuhan data, algoritma, diagram, *relational database*, dan *user interface*.

BAB V IMPLEMENTASI DAN PENGUJIAN

Bab ini berisi tentang implementasi sistem dan pengujian dalam penelitian.

BAB VI PENUTUP

Bab ini berisi tentang kesimpulan dari hasil penelitian yang telah dilakukan dan saran terhadap penelitian berikutnya.