BAB II

LANDASAN TEORI

2.1 *Internet of Things*

Internet of Things atau dikenal juga dengan singkatan IoT, merupakan perkembangan keilmuan yang sangat menjanjikan untuk mengoptimalkan kehidupan berdasarkan sensor cerdas dan peralatan pintar yang terhubung melalui jaringan internet (Keoh et al., 2014).

Selain itu, *Internet of Things* juga merupakan sebuah konsep yang bertujuan untuk memperluas manfaat dari konektivitas internet yang tersambung secara terus-menerus. Dengan semakin berkembangnya infrastruktur internet, bukan hanya *smartphone* atau komputer saja yang dapat terkoneksi dengan internet. Namun berbagai macam benda nyata akan terkoneksi dengan internet. Sebagai contohnya dapat berupa : mesin produksi, mobil, peralatan elektronik, peralatan yang dapat dikenakan manusia (*wearables*), dan termasuk benda nyata apa saja yang semuanya tersambung ke jaringan lokal dan global menggunakan sensor dan atau aktuator yang tertanam (Arafat et al., 2016). Sensor mampu mengkonversi data fisik mentah menjadi sinyal digital dan mengirimkan ke pusat control sehingga perubahan dapat dikontrol dengan jarak jauh melalui internet (Aini et al., n.d., 2018).

Data yang dikirimkan melalui internet tersebut dapat menjadi jalur rentan untuk disusupi oleh *malware* yang bisa merusak perangkat serta mencuri data yang sangat penting, tetapi serangan *malware* yang menyerang sistem jaringan IoT tersebut dapat dipantau *network traffic* nya, sehingga ketika terdapat serangan

malware atau anomali data *network traffic* terhadap sistem jaringan IoT bisa diprediksi karakteristik data *network traffic* nya yang termasuk kedalam *benign* atau *malicious*.

2.2 *Malware*

Malware merupakan perangkat lunak yang secara eksplisit didesain untuk melakukan aktifitas berbahaya atau perusak perangkat lunak lainnya seperti *Trojan*, *Virus*, *Spyware* dan *Exploit* (Kramer & Bradfield, 2010). *Malware* diciptakan dengan maksud tertentu yaitu melakukan aktifitas berbahaya yang berdampak sangat merugikan bagi para korbannya, antara lain seperti penyadapan serta pencurian informasi pribadi, hingga kasus perusakan sistem yang dilakukan oleh penyusup (*Intruder*) terhadap perangkat korban dengan berbagai alasan (Cahyanto et al., 2017).

2.3 *Machine Learning*

Machine learning mempelajari bagaimana sebuah mesin atau komputer dapat belajar dari pengalaman atau bagaimana cara memprogram mesin untuk dapat belajar. *Machine learning* membutuhkan data untuk belajar sehingga biasa juga diistilahkan dengan *learn from data*.

Ada beberapa teknik *machine learning* tetapi ada dua teknik belajar dasar seperti *supervised learning* dan *unsupervised learning*, *supervised learning* merupakan teknik yang bisa diterapkan pada pembelajaran mesin yang bisa menerima informasi yang sudah memiliki label sehingga ketika di *training* akan

mempelajari berdasarkan data yang telah ada, pola tersebut akan digunakan sebagai acuan untuk data-data berikutnya. Sedangakan *unsupervised learning* bersifat deskriptif, yang bisa digunakan untuk mengelompokkan atau mengkategorikan data (Santoso et al., 2021).

2.4 *Data Mining*

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik-teknik, metode-metode, atau algoritma dalam *data mining* sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan untuk memperoleh pengetahuan dari *database* atau dataset yang ada (Mardi Y, 2013).

2.5 Algoritma *K-Nearest Neighbour* (K-NN)

Algoritma *K-Nearest Neighbor (K-NN)* adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Data pembelajaran diproyeksikan ke ruang berdimensi banyak, dimana masing-masing dimensi merepresentasikan fitur dari data. Ruang ini dibagi menjadi bagian-bagian berdasarkan klasifikasi data pembelajaran. Sebuah titik pada ruang ini ditandai kelas c jika kelas c merupakan klasifikasi yang paling banyak ditemui pada k buah tetangga terdekat titk tersebut. Dekat atau jauhnya tetangga biasanya dihitung berdasarkan jarak *Euclidean* dengan rumus seperti pada persaman dibawah :

$$distance = \sqrt{\sum_{i=1}^{n} (X^{i}_{training} - X_{testing})^2}$$

Dengan

$X^{i}_{training}$: data training ke-*i*,

$X_{testing}$: data testing,

i : record (baris) ke-*i* dari tabel,

n : jumlah data training.

Pada fase pembelajaran, algoritma ini hanya melakukan penyimpanan vektor-vektor fitur dan klasifikasi dari data pembelajaran. Pada fase klasifikasi, fitur-fitur yang sama dihitung untuk *data test* (yang klasifikasinya tidak diketahui). Jarak dari vektor yang baru ini terhadap seluruh vektor data pembelajaran dihitung, dan sejumlah K buah yang paling dekat diambil. Titik yang baru klasifikasinya diprediksikan termasuk pada klasifikasi terbanyak dari titik-titik tersebut. Nilai K yang terbaik untuk algoritma ini tergantung pada data. Secara umum, nilai K yang tinggi akan mengurangi efek *noise* pada klasifikasi, tetapi membuat batasan antara setiap klasifikasi menjadi lebih kabur (Yustanti, 2012).

2.6 *Dataset*

Himpunan data (*dataset*) merupakan kumpulan dari objek dan atributnya (Hermawati, 2013). Dalam *dataset*, jenis data dapat dibagi menjadi 2 bagian yaitu:

1. *Data Training*

Data training adalah data yang digunakan untuk perhitungan probabilitas dari data berdasarkan data pembelajaran yang dilakukan (Pratiwi et al., 2016). *Data training* merupakan data yang sebelumnya sudah ada sesuai dengan fakta.

2. *Data Testing*

   Data testing merupakan data yang akan atau sedang terjadi dan dipergunakan sebagai bahan uji yang sebelumnya sudah didapatkan pada *data training* (Pratiwi et al., 2016). Data ini digunakan untuk mengukur sejauh mana klasifikasi berhasil melakukan klasifikasi dengan benar.

2.7 Penelitian Terdekat

Tabel 2.1 Tabel Penelitian Terdekat

No.	Peneliti/Tahun	Judul	Masalah Penelitian	Algoritma	State Of The Art
1	Green Arther Sandag, Jonathan Leopold, Vinky Fransiscus Ong [2018]	*Klasifikasi Malicious Websites Menggunakan Algoritma K-NN Berdasarkan Application Layers dan Network Characteristics*	1. Modus kejahatan di dunia *cyber* saat ini sangat beragam 2. Banyak ancaman terhadap keamanan Web 3. *Malware* dapat menyebar dengan cepat di jaringan tanpa campur tangan dari pengguna	Metode: *10-fold cross validation* Algoritma: *K-Nearest Neighbour* (K-NN)	1. Mengklasifikasikan *malicious website*. 2. *K-NN* menjadi algoritma dengan performa yang baik ketika dibandingkan dengan algoritma yang lain seperti : *Decision Tree, Logistic Regression,* dan *Random Forest*. 3. Melakukan proses data *cleaning* dan data *reduction*.
2	Ari Sandriana [2022]	*klasifikasi serangan malware terhadap lalu lintas jaringan IoT menggunakan algoritma K-Nearest Neighbour (K-NN)*	1. Pada tahun 2019 serangan siber menggunakan *malware* di Indonesia merupakan yang tertinggi di Asia Pasifik 2. 1,6 miliar lebih anomali *traffic* diwilayah Indonesia disebabkan oleh serangan *malware* 3. Ancaman keamanan pada perangkat IoT meningkat	Algoritma: *K-Nearest Neighbour* (K-NN)	1. Mengklasifikasikan serangan *malware* yang bersifat *benign* atau *malicious* 2. Memprediksi data *network traffic* kedalam *benign* atau *malicious*

Tabel 2.1 Menerangkan penelitian yang dijadikan acuan terkait Klasifikasi serangan *malware* menggunakan algoritma *K-Nearest Neighbour* (K-NN) berdasarkan *anomaly detection* pada lalu lintas jaringan IoT. Penelitian berjudul *"Klasifikasi Malicious Websites Menggunakan Algoritma K-NN Berdasarkan Application Layers dan Network Characteristics"* yang dilakukan oleh Green Arther Sandag, Jonathan Leopold, Vinky Fransiscus Ong (2018). Penelitian ini melakukan proses klasifikasi *malicious website* dan menjadikan *K-Nearest Neighbour* (K-NN) sebagai algoritman, karena algoritma *K-NN* mempunyai performa yang paling baik dibandingkan dengan algoritma lain seperti : *Decision Tree, Logistic Regression,* dan *Random Forest*.

Berdasarkan penelitian yang telah disebutkan maka di usulkan penelitian yang berjudul *"Klasifikasi serangan malware terhadap lalu lintas jaringan IoT menggunakan algoritma K-Nearest Neighbour (K-NN)"*.