BAB I
PENDAHULUAN

1.1. Latar Belakang Penelitian

Seiring dengan kemajuan dan pesatnya penggunaan sistem informasi berbasis situs *web (websites)* saat ini, menawarkan berbagai kemudahan dalam proses menyimpan, mengakses, mempublikasikan dan mengirimkan data karena dapat dilakukan secara *online* dan dikontrol dari jarak jauh. Dalam situs *web* penggunaan berbagai jenis data sudah merupakan hal yang umum, akan tetapi penggunaan data yang dipublikasikan sering kali tidak memperhatikan aturan terkait perlindungan data. Mengingat banyak situs *web* yang ada menampilkan data dan informasi yang tidak layak untuk dipublikasikan seperti data pribadi seseorang yang bersifat sensitif. Tanpa disadari hal tersebut memberikan peluang besar kepada pihak luar untuk melakukan tindakan kejahatan (Norman & Yasin , 2010).

Berbagai macam kasus kriminal yang pernah terjadi dengan memanfaatkan data pribadi, pada tahun 2016 di Amerika terjadi kasus pencurian lebih dari 100.000 informasi finansial masyarakat pada situs *web Internal Revenue Service* yang dilakukan oleh organisasi kriminal. Tujuannya adalah untuk melakukan penipuan dengan mengklaim pengembalian pajak dengan mengatasnamakan pemilik data untuk bantuan pinjaman peserta didik. Pada bulan Nopember tahun 2017 terjadi kebocoran data pelanggan operator seluler di Malaysia secara *online* sebesar 46 juta data. Motif dari kasus ini adalah untuk

menjual data pelanggan kepada pihak ketiga (Maulida, 2017). Pada bulan Juli tahun 2019 di Indonesia terjadi kasus penipuan

yang dilakukan oleh perusahaan pinjaman *online*, dimana mereka menggunakan data pribadi seseorang yang didapat dari aksi pencurian data pada aplikasi GOJEK dan Tokopedia (Ngazis, 2019).

Beberapa negara di dunia sudah memberlakukan aturan khusus dan pasti terkait dengan perlindungan data pribadi seperti: *Health Insurance Protability Auditor Accountability* (HIPAA) di Amerika Serikat, *General Data Protection Regulation* (GDPR) di kawasan regional Uni Eropa dan Undang-Undang Akta 709 Tahun 2010 tentang Perlindungan Data Peribadi di Malaysia. Sementara untuk di Indonesia sendiri aturan khusus yang mengatur tentang perlindungan data pribadi masih dalam bentuk Rancangan Undang-Undang (RUU) yang belum disahkan. Akan tetapi aturan umum terkait perlindungan data pribadi sudah ada dalam UU No. 23 Tahun 2006 tentang Administrasi Kependudukan dan Peraturan Menteri Komunikasi Dan Informatika Republik Indonesia Nomor 20 Tahun 2016 Tentang Perlindungan Data Pribadi Dalam Sistem Elektronik.

Berkaitan dengan perlindungan data pribadi di Indonesia, banyak sekali ditemukan data pribadi masyarakat sipil yang dipublikasikan melalui media elektronik (*online*) tanpa mempertimbangkan aspek penting terkait dengan perlindungan data pribadi (Sautunnida, 2018). Data pribadi yang disebar banyak terdapat pada situs *web* instansi pemerintahan atau swasta. Berdasarkan pengamatan, data masyarakat yang sering disebarkan mencakup: nama,

tempat/tanggal lahir, NIK (nomor induk kependudukan), NKK (nomor kartu keluarga), nama orang tua, alamat, dan foto wajah tanpa mempertimbangkan aturan yang ada. Hal ini dapat dilihat pada saat pelaksanaan PEMILU (Pemilihan Umum) yang ada di Indonesia, dimana seluruh data masyarakat yang mempunyai hak pilih dipublikasikan dalam situs resmi Komisi Pemilihan Umum (KPU) dan juga ditempel di setiap desa atau kelurahan yang ada tanpa mempertimbangkan perlindungan akan data pribadi. Hal ini berpotensi membahayakan keamanan data penduduk Indonesia mengingat celah ini akan digunakan oleh pihak-pihak yang berkepentingan untuk melakukan hal-hal yang tidak baik.

Dalam memperoleh data pribadi pada situs *web* dapat dilakukan dengan berbagai cara seperti: dengan memanfaatkan mesin pencari, menerapkan metode *crawling*, dan langsung melakukan pencarian pada situs *web* yang bersangkutan. Penggunaan metode *crawling* membuat seseorang dengan mudah dan cepat untuk memperoleh suatu konten dalam sebuah situs *web* seperti data pribadi masyarakat (Sreeja & Chaudhari, 2014).

Data pribadi penduduk yang harus dilindungi memuat: NKK, NIK, tanggal/bulan/tahun lahir, keterangan tentang kecacatan fisik dan/atau mental, NIK ibu kandung, NIK ayah, dan beberapa isi catatan peristiwa penting (Indonesia, 2006). Oleh karena itu, pelaksanaan berbagai macam hal yang menggunakan data pribadi penduduk harus mematuhi aturan yang ada terkait pelindungan data pribadi. Pada kenyataannya di Indonesia pelaksanaan berbagai hal yang menggunakan data pribadi tidak sesuai seperti yang disampaikan sebelumnya. Mengingat di Indonesia undang-undang khusus yang mengatur

tentang masalah pelindungan data pribadi sampai saat ini belum disahkan dan masih berbentuk RUU.

Berdasarkan uraian tersebut, maka pada penelitian ini dilakukan pengkajian data pribadi masyarakat Indonesia yang dipublikasikan oleh KPU untuk mengetahui data apa saja yang dipublikasikan dan tingkat publikasi data pribadi sensitif masyarakat Indonesia. Dalam penelitan ini menggunakan bantuan metode *web crawling*, praproses (*preprocessing)* data dengan *Text Mining*, pembobotan kata dan analisis *data mining* klasifikasi. Melalui penelitian ini diharapkan KPU lebih protektif dan teliti dalam mempublikasikan data pribadi masyarakat dengan mempertimbangkan aturan yang ada.

Berkaitan dengan pembobotan kata dan analisis *data mining* klasifikasi, pada penelitian ini digunakan metode pembobotan *Term Frequency Invers Document Frequency* (TF-IDF) dan metode klasifikasi Naïve Bayes. Penggunaan metode pembobotan TF-IDF dikarenakan metode ini paling umum dan sudah teruji dalam pembobotan kata, dimana metode TF-IDF memiliki nilai skor yang lebih baik dibandingkan dengan metode BLAU (Nugroho, 2014) dan (Saputro, dkk, 2017). Sementara untuk penggunaan metode klasifikasi Naïve Bayes dikarenakan metode Naïve Bayes memiliki akurasi yang lebih baik dan stabil dibandingkan dengan metode K-Nearest Neighbor, C.45 dan ID3 (Marcos & Utomo, 2015), (Riri , 2018) & (Astuti & Rifai, 2018). Penggunaan metode Naïve Bayes juga berdasarkan hasil uji pendahuluan yang telah dilakukan, dimana metode Naïve Bayes memiliki akurasi yang paling baik terhadap *dataset* pada

penelitian ini dibandingkan metode klasifikasi lainnya seperti: C.45 dan K-Nearest Neighbor.

1.2. Identifikasi Masalah

Berdasarkan latar belakang yang telah dijabarkan sebelumnya maka dapat diidentifikasi permasalahan yang ada, antara lain:

1) Dalam hal publikasi data pribadi masyarakat, masih ada pemilik data atau pengelola situs resmi KPU di daerah tingkat provinsi yang kurang menyadari akan pentingnya menjaga dan melindungi data pribadi masyarakat yang dipublikasikan di internet. Mana data yang bisa dipublikasi dan mana yang tidak.
2) Mesin pencari, *web crawling* maupun pencarian langsung ke situs resmi KPU, memberikan kesempatan yang besar kepada seseorang untuk memperoleh data pribadi masyarakat dengan cepat dan mudah. Hal ini disebabkan karena tidak adanya kesepakatan dan keseragaman dalam melakukan publikasi data pribadi masyarakat antara KPU Pusat dengan daerah, sehingga memberikan celah yang besar bagi pihak tertentu dalam memperoleh data pribadi masyarakat untuk kepentingan tertentu.
3) Belum adanya keseragaman dan kepatuhan penerapan aturan yang ada berkaitan dengan publikasi data pribadi yang dilakukan oleh KPU. Hal ini dilihat dari masih adanya situs-situs resmi KPU Provinsi yang masih mempublikasikan data pribadi masyarakat tanpa mempertimbangkan aturan yang ada seperti UU No. 23 Tahun 2006 Tentang Administrasi Kependudukan maupun Peraturan Menteri Komunikasi dan Informatika Nomor 20 Tahun 2016 tentang Perlindungan Data Pribadi dalam Sistem Elektronik.

1.3. Batasan Masalah

Berdasarkan indentifikasi masalah sebelumnya, adapun batasan masalahnya antara lain:

1) Penelitian ini menguji situs resmi KPU Pusat dan Provinsi dimana untuk situs provinsi telah dikelompokkan menjadi 3 wilayah yaitu timur, tengah dan barat.
2) Data yang digunakan dalam penelitian ini adalah data pribadi masyarakat sipil, mencakup: Nama, NIK, NKK, NIP (Nomor Induk Pegawai), Tempat/Tanggal Lahir, Alamat, Nomor Telepon/HP, Nomor Rekening, Nomor Pokok Wajib Pajak (NPWP), Jabatan/ Pekerjaan, Jenis Kelamin, Agama, Alamat *Email*, Foto Wajah.
3) Pencarian data dilakukan dengan menggunakan bantuan metode *web crawling* untuk memperoleh *hyperlink* dari masing-masing situs yang diteliti. Hal ini bertujuan agar data yang didapat menjadi lebih cepat dan mudah. Jika terdapat situs yang tidak menginjinkan *hyperlink*-nya diperoleh menggunakan metode *web crawling*, maka dilakukan pencarian menggunakan cara manual.
4) Data dari *hyperlink* didapat dengan cara mengunduh konten yang ada didalamnya menggunakan bantuan *Google Sheets*. Data tersebut dibersihkan terlebih dahulu untuk selanjutnya dilakukan pembobotan serta pengklasifikasian data menggunakan bantuan *data mining*.

5) Data diklasifikasikan berdasarkan 2 kelas yaitu: kelas ya dan kelas tidak. Untuk mengetahui apakah konten yang ada mengandung data pribadi sensitif atau tidak.

6) Dalam memperoleh rekomendasi terkait upaya yang harus dilakukan KPU dalam publikasi data pribadi masyarakat berpedoman pada Undang-undang No. 23 Tahun 2006 tentang Administrasi Kependudukan dan Peraturan Menteri Komunikasi dan Informatika Nomor 20 Tahun 2016 tentang Perlindungan Data Pribadi dalam Sistem Elektronik.

1.4. Rumusan Masalah

Berdasarkan identifikasi masalah dan batasan masalah sebelumnya adapun permasalahan yang ingin diselesaikan dalam penelitian ini antara lain:

1) Informasi dan data pribadi sensitif masyarakat apa saja yang dipublikasikan pada situs resmi KPU Pusat dan Provinsi berdasarkan kelompok wilayahnya?

2) Seberapa tinggi tingkat publikasi data pribadi masyarakat Indonesia yang dipublikasikan pada situs resmi KPU Pusat dan Provinsi berdasarkan kelompok wilayahnya?

3) Seberapa baik tingkat akurasi metode Naïve Bayes dalam mengklasifikasikan data yang diperoleh dari pembobotan TF-IDF?

4) Bagaimana seharusnya KPU dalam melakukan publikasi data pribadi masyarakat sehingga data pribadi masyarakat Indonesia dapat terlindungi jika dikaitkan dengan Undang-undang No. 23 Tahun 2006 tentang

Administrasi Kependudukan Pasal 84 ayat (1) dan Peraturan Menteri Komunikasi dan Informatika Nomor 20 Tahun 2016 tentang Perlindungan Data Pribadi dalam Sistem Elektronik?

1.5. Tujuan Penelitian

Adapun tujuan dari penelitian ini antara lain:

1) Mendeskripsikan informasi dan data pribadi masyarakat yang terpublikasi apakah sudah sesuai dengan aturan yang ada dalam Undang-Undang No. 23 Tahun 2006 Tentang Administrasi Kependudukan dan Peraturan Menteri Komunikasi dan Informatika Nomor 20 Tahun 2016 tentang Perlindungan Data Pribadi dalam Sistem Elektronik.
2) Mendeskripsikan sebaran dan tingkat publikasi data pribadi masyarakat yang dipublikasikan oleh situs resmi KPU Pusat dan Provinsi berdasarkan kelompok wilayahnya.
3) Mendeskripsikan tingkat akurasi dari metode Naïve Bayes dalam mengklasifikasikan data yang telah dibobotkan dengan metode TF-IDF.
4) Merumuskan rekomendasi terkait publikasi data pribadi masyarakat Indonesia yang dilakukan oleh KPU sesuai dengan Undang-Undang No. 23 Tahun 2006 Tentang Administrasi Kependudukan dan Peraturan Menteri Komnunikasi dan Informatika Nomor 20 Tahun 2016 tentang Perlindungan Data Pribadi dalam Sistem Elektronik, sehingga keamanan data pribadi masyarakat dapat terjamin kerahasiaannya.

1.6. Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah untuk memperoleh suatu rekomendasi tentang bagaimana seharusnya pelaksanaan hal yang berkaitan dengan publikasi data pribadi masyarakat dengan memperhatikan aturan terkait perlindungan data pribadi. Adapun manfaat penelitian ini dapat dilihat dari 2 aspek yaitu aspek teoritis dan praktis.

1) **Aspek Teoretis**

Dilihat dari aspek teoretis, penelitian ini diharapkan memberikan wawasan untuk memperkaya pengetahuan tentang bagaimana seharusnya melakasanakan hal yang berkaitan dengan publikasi data pribadi masyarakat dengan memperhatikan aturan yang ada terkait perlindungan data pribadi.

2) Aspek Praktis

Dilihat dari aspek praktis, penelitian ini diharapkan memberikan dasar pertimbangan kepada lembaga-lembaga yang ada khususnya KPU Pusat maupun Provinsi dalam melakukan publikasi data pribadi masyarakat sehingga tidak melanggar aturan terkait perlindungan data pribadi.