BAB I

PENDAHULUAN

1.1. Latar Belakang Penelitian

Lembaga Pendidikan Tenaga Kependidikan (LPTK) merupakan salah satu penyelenggara sektor edukasi yang padat data. Hingga tahun 2019 telah terdapat 19 LPTK rayon di Indonesia. LPTK merupakan salah satu bentuk perguruan tinggi tertua di Indonesia dan sejak berdirinya telah meluluskan banyak mahasiswa.

Salah satu proses bisnis yang dilakukan oleh LPTK adalah pengumpulan data calon mahasiswa. Setelah lulus tahap seleksi, mahasiswa akan diminta melengkapi biodata dan keterangan-keterangan tambahan yang menyangkut identitas pribadinya. Begitu pula setelah kelulusan, identitas tersebut tetap disimpan sebagai arsip data pribadi alumni.

Seiring dengan perkembangan teknologi, kegiatan menyimpan, mengakses, dan mengirimkan data melalui media elektronik, bahkan dari jarak jauh (*remote*) sudah menjadi hal yang umum. Misalnya saja di LPTK, data pribadi dapat disimpan dan diakses secara lebih mudah melalui situs *web* sistem informasi digital yang dikelolanya.

Sejalan dengan kemudahan-kemudahan yang ditawarkan, berbagai resiko baru mulai bermunculan. Berdasarkan penelitian pendahuluan yang telah dilakukan, dengan memanfaatkan mesin pencari serta melakukan penjelajahan

secara langsung di situs *web* sistem informasi alumni dan direktori data dari LPTK terkait, pengakses berpeluang memperoleh data pribadi dengan relatif cepat dan mudah serta tanpa memerlukan hak akses istimewa (*privilege*). Peluang tersebut umumnya muncul karena kurangnya kesadaran dari pemilik data dan pihak pengelola terkait pembatasan yang tegas antara bagian data yang boleh dan tidak boleh disebarkan (Zhao & Dong, 2017).

Data pribadi mahasiswa, alumni, dosen, dan staf di lingkungan LPTK adalah data yang cukup kompleks yang dapat memuat nama, tempat lahir, tanggal lahir, alamat surel, alamat tinggal, alamat bekerja, dan sebagainya. Data tersebut sangat penting untuk dilindungi guna mencegah eksploitasi atau penyalahgunaan oleh pihak yang tidak bertanggung jawab.

Penyalahgunaan data pribadi adalah kasus yang luas dan terus berkembang (Indriyani, et al., 2017) dan (Sautunnida, 2018). Salah satu kasus yang pernah terjadi pada tahun 2016 sehubungan dengan penyalahgunaan data pribadi adalah kasus penipuan oleh sebuah organisasi kriminal yang mencuri lebih dari 100.000 informasi finansial dari situs *web* Internal Revenue Service. Informasi tersebut digunakan untuk mengklaim pengembalian pajak dengan mengatasnamakan pemilik data untuk bantuan pinjaman peserta didik (*student loan*). Pada bulan Januari 2019, kasus kebocoran data pribadi terjadi di Universiti Teknologi MARA dimana sebanyak 1.164.540 *record* data pribadi mahasiswa dan alumni bocor. Data pribadi tersebut mencakup nomor induk mahasiswa, nama lengkap, nomor MyKad, alamat tinggal, alamat surel, nomor telepon, kode perguruan tinggi, dan nama perguruan tinggi. Meskipun belum terindikasi data pribadi tersebut

disalahgunakan oleh pihak yang tidak bertanggung jawab, namun hal tersebut tetap merupakan suatu pelanggaran jika merujuk pada Undang-Undang Malaysia Akta 709 Tahun 2010 tentang Perlindungan Data Peribadi karena data pribadi yang bocor tidak melalui persetujuan pihak pemilik data.

Dengan semakin berkembangnya teknologi, kasus kebocoran (bahkan penyalahgunaan) data pribadi menjadi sangat mungkin terjadi, misalnya saja data pribadi yang tertera pada suatu situs *web* dicocokkan dengan data lain yang diperoleh dari situs *web* media sosial dan sebagainya (misalnya melalui kesamaan nama lengkap, alamat, foto wajah, dan sebagainya), kemudian dijadikan bahan penipuan dengan mengatasnamakan pihak pemilik data seperti kasus penipuan berbasis telepon yang mengatasnamakan Gojek pada tahun 2018 lalu. Kasus serupa dapat saja terjadi dengan memanfaatkan data pribadi sensitif yang diperoleh dari situs *web* sistem informasi di perguruan tinggi. Oleh sebab itu, dalam mencegah terjadinya penyalahgunaan data pribadi maka perlu dilakukan penyesuaian dengan regulasi yang berlaku.

Berkenaan dengan perlindungan data pribadi, berbagai regulasi telah gencar diberlakukan di beberapa negara, misalnya HIPAA di Amerika Serikat, Undang-Undang Akta 709 Tahun 2010 tentang Perlindungan Data Peribadi di Malaysia, Peraturan Menteri Komunikasi dan Informatika Nomor 20 Tahun 2016 tentang Perlindungan Data Pribadi dalam Sistem Elektronik serta Undang-Undang Nomor 23 Tahun 2006 tentang Administrasi Kependudukan di Indonesia, dan GDPR di kawasan regional Uni Eropa. Khusus untuk wilayah Indonesia, di dalam Peraturan Menteri Komunikasi dan Informatika Nomor 20 Tahun 2016 dan

didukung dengan Undang-Undang Nomor 23 Tahun 2006 telah secara tegas dinyatakan mengenai tata cara pengumpulan data pribadi dan perlindungannya, namun dari hasil penelitian pendahuluan yang dilakukan masih ditemukan beberapa LPTK yang belum menerapkan aturan tersebut dengan baik dalam mengelola situs *web* sistem informasinya yang salah satunya dapat dilihat dari masih adanya data pribadi sensitif yang dipublikasikan.

Berdasarkan uraian tersebut, maka pada penelitian ini akan dilakukan uji pencarian data pribadi sensitif, analisis *data mining clustering*, dan analisis secara manual untuk mengidentifikasi jenis data pribadi sensitif yang dipublikasikan beserta sebarannya di situs *web* sistem informasi LPTK di Indonesia. Diharapkan penelitian ini dapat menjadi refleksi bagi LPTK terkait dan secara umum bagi perguruan tinggi di Indonesia untuk dapat secara lebih baik melindungi data pribadi sensitif yang dikelolanya.

Terkait analisis *clustering*, metode yang akan digunakan adalah *Expectation Maximization*. Pada uji awal yang telah penulis lakukan terhadap beberapa sampel yang didapatkan dari penelitian pendahuluan, metode *Expectation Maximization* menghasilkan klaster yang lebih representatif jika dibandingkan dengan metode *clustering* lain, seperti *canopy*, *cobweb*, dan *K-Means*.

Sementara itu terkait subjek penelitian, untuk memfokuskan studi pada ruang lingkup kasus yang lebih spesifik namun dengan tetap mempertahankan generalisasi hasil yang akan diperoleh maka dipilih situs *web* sistem informasi LPTK di Indonesia sebagai subjek. Secara umum LPTK merupakan bentuk

perguruan tinggi yang serupa dengan perguruan tinggi lainnya namun memiliki fungsi tambahan sebagai penyelenggara program Pendidikan dan Latihan Profesi Guru (PLPG). Pemilihan subjek yang lebih spesifik tersebut memungkinkan penalaran dapat lebih disederhanakan dan terarah serta tetap mempertahankan sebaran subjek berdasarkan zona wilayah dan statusnya sehingga pengaruh budaya publikasi data pribadi sensitif berdasarkan zona wilayah maupun status perguruan tinggi dapat diminimalisir.

1.2. Identifikasi Masalah

Berdasarkan latar belakang yang telah dijabarkan sebelumnya, maka dapat diidentifikasi permasalahan-permasalahan yang terjadi sehubungan dengan perlindungan data pribadi sensitif di LPTK di Indonesia, diantaranya:

1) Dengan memanfaatkan bantuan mesin pencari serta penjelajahan secara langsung di situs *web* sistem informasi LPTK terkait, pengakses berpeluang untuk mendapatkan data pribadi sensitif di LPTK tersebut dengan relatif cepat dan mudah serta tanpa memerlukan hak akses istimewa (*privilege*).
2) Pihak pemilik dan pengelola data pribadi sering kali kurang sadar terkait publikasi data pribadi di situs *web* sistem informasi.
3) Masih ada LPTK di Indonesia yang belum menerapkan Peraturan Menteri Komunikasi dan Informatika Nomor 20 Tahun 2016 dan Undang-Undang Nomor 23 Tahun 2006 dengan baik dalam melindungi data pribadi sensitif yang terdapat pada situs *web* sistem informasi yang dikelolanya.

1.3. Batasan Masalah

Berdasarkan identifikasi masalah yang telah diuraikan sebelumnya, maka dilakukan pembatasan terhadap ruang lingkup permasalahan sehingga penelitian menjadi lebih terarah. Berikut merupakan batasan dari permasalahan yang menjadi fokus utama pada penelitian ini:

1) Jenis data pribadi yang diangkat adalah data pribadi sensitif di situs *web* alumni dan direktori data LPTK di Indonesia. Pembatasan pada jenis data pribadi ini dimaksudkan untuk memfokuskan studi pada ruang lingkup kasus yang lebih spesifik sehingga penalaran dapat lebih disederhanakan dan terarah, namun tidak mengurangi generalisasi dari simpulan yang akan dihasilkan.
2) Subdomain yang dipilih adalah situs *web* alumni dan direktori data LPTK karena umumnya publikasi data pribadi di dalam subdomain tersebut bersifat lebih terbuka, tidak seperti subdomain lain (misalnya situs *web* akademik, keuangan, wisuda, dan sebagainya) yang bersifat lebih tertutup.
3) Pencarian tautan (*hyperlinks*) di situs *web* sistem informasi alumni dan direktori data LPTK dilakukan dengan bantuan metode *web crawling*. Penggunaan metode tersebut ditujukan agar proses pencarian tautan dapat dilakukan secara otomatis, sehingga lebih efisien. Namun dalam hal situs *web* sistem informasi tersebut tidak dapat di-*crawl* maka pencarian tautan dilakukan secara manual.
4) Klasterisasi tingkat sensitivitas data pribadi dilakukan dengan bantuan *data mining*. Hasilnya kemudian dikomparasi dengan hasil klasifikasi manual

sehingga dapat dihitung akurasinya. Pengelompokkan hasilnya dibagi ke dalam 3 tingkatan sensitivitas berdasarkan definisi dari National Institute of Standards and Technology (NIST), diantaranya: *Critical* PII (*personally identifiable information* yang bersifat *critical* dari NIST), *Potential* PII (*personally identifiable information* yang memiliki potensi untuk dapat mengidentifikasi seseorang jika dikombinasikan dengan beberapa PII lain), dan nonPII (tidak mengandung *personally identifiable information*).

5) Regulasi yang dijadikan sebagai referensi dalam penyusunan rekomendasi upaya perlindungan data pribadi sensitif adalah Peraturan Menteri Komunikasi dan Informatika Nomor 20 Tahun 2016 tentang Perlindungan Data Pribadi dalam Sistem Elektronik yang di dalamnya mencakup tata cara pemerolehan dan perlindungan data pribadi secara umum.

1.4. Rumusan Masalah

Berdasarkan indentifikasi masalah dan pembatasan masalah yang diuraikan pada pokok bahasan sebelumnya, maka dapat dirumuskan permasalahan-permasalahan yang diangkat pada penelitian ini, diantaranya:

1) Data pribadi sensitif seperti apa yang dipublikasikan oleh LPTK di Indonesia pada situs *web* sistem informasi alumni dan direktori data yang dikelolanya?
2) Seperti apa sebaran data pribadi sensitif yang dipublikasikan pada situs *web* sistem informasi alumni dan direktori data dari LPTK tersebut?
3) Seberapa baik akurasi metode *Expectation Maximization* dalam mengklaster konten-konten yang diperoleh dari situs *web* sistem informasi alumni dan direktori data tersebut?

4) Upaya apa yang dapat dilakukan oleh LPTK di Indonesia dalam melindungi data pribadi sensitif yang dikelolanya jika merujuk pada Peraturan Menteri Komunikasi dan Informatika Nomor 20 Tahun 2016?

1.5. Tujuan Penelitian

Adapun tujuan yang diharapkan dapat dicapai melalui penelitian ini adalah sebagai berikut:

1) Untuk mengetahui data pribadi sensitif apa saja yang dipublikasikan oleh LPTK di Indonesia pada situs *web* sistem informasi alumni dan direktori data yang dikelolanya.
2) Untuk mengetahui sebaran data pribadi sensitif yang dipublikasikan pada situs *web* sistem informasi alumni dan direktori data dari LPTK tersebut.
3) Untuk mengetahui akurasi metode *Expectation Maximization* dalam mengklaster konten-konten yang diperoleh dari situs *web* sistem informasi alumni dan direktori data tersebut.
4) Untuk mengetahui upaya-upaya yang dapat dilakukan oleh LPTK di Indonesia dalam melindungi data pribadi sensitif yang dikelolanya jika merujuk pada Peraturan Menteri Komunikasi dan Informatika Nomor 20 Tahun 2016.

1.6. Manfaat Penelitian

Kebermanfaatan penelitian ini dapat ditinjau dari aspek teoretis dan aspek praktis, yaitu:

1) Secara teoretis, hasil dari penelitian ini diharapkan dapat memperkaya wawasan terkait sebaran data pribadi sensitif pada situs *web* sistem informasi alumni dan direktori data LPTK di Indonesia beserta data apa saja yang dipublikasikan. Selain itu, penelitian ini juga diharapkan dapat dijadikan sebagai bahan acuan untuk kepentingan pengembangan pengetahuan lebih lanjut terhadap hal sejenis ataupun hal lainnya yang belum tercakup pada penelitian ini.
2) Secara praktis, hasil penelitian ini diharapkan dapat dijadikan sebagai refleksi bagi masing-masing LPTK mengenai seberapa baik situs *web* sistem informasinya telah mampu melindungi data pribadi sensitif yang dikelolanya dalam kaitannya dengan implementasi Peraturan Menteri Komunikasi dan Informatika Nomor 20 Tahun 2016 di Indonesia.