

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi internet memudahkan setiap individu untuk berkomunikasi satu sama lain melalui media *online* seperti Instagram, Facebook, dan Twitter. Tentunya hal tersebut memiliki dampak negatif bagi sebagian individu maupun kelompok. Berdasarkan *data* pada PT. Bakrie Telecom, pengguna Twitter di Indonesia adalah sebanyak 19,5 juta dari total 500 juta pengguna global. [1] Dalam satu dekade terakhir, ujaran kebencian yang diungkapkan secara daring menjadi semakin marak dilakukan. [2] Berdasarkan *data Virtual Police* (Dit Tipisiber) Bareskrim Polri pada periode 23 Februari hingga 11 Maret 2021 sebanyak 125 konten diajukan untuk diberikan peringatan *Virtual Police* yang didominasi platform Twitter 79 konten, Facebook 32 konten, Instagram 8 konten, YouTube 5 konten dan Whatsapp 1 konten.

Berkaitan dengan hal tersebut Blankspace bekerja sama dengan CSIS Indonesia melakukan penelitian mengenai “Analisis Sentimen *Hate Speech* di Indonesia”. Dalam melakukan penelitian dibutuhkan *Data Engineer* dan *Data Scientist* sebagai bagian dari penelitian ini. Maka dari itu dibutuhkan, mahasiswa sebagai *Data Engineer Intern* untuk *collect data*, *cleaning data*, dan *preprocessing data*. Pada penelitian ini akan mengambil *data* dari Twitter dengan menggunakan berbagai *keyword* yang mengandung ujaran kebencian berfokus pada 3 kelompok utama yang akan diteliti yaitu Ahmadiyah, Cina, dan Syiah. Dalam penelitian ini akan melalui beberapa proses seperti *scraping*, *cleaning*, dan pembuatan *model machine learning* yang akan dilakukan oleh *Data Engineer* dan *Data Scientist*. Baik *scraping* dan *cleaning data* dilakukan dengan menggunakan Bahasa pemrograman Python. [3] Pada *scraping* menggunakan *library twint*, dikarenakan lebih efisien dan waktu yang diperlukan untuk *scraping* juga lebih cepat. Sedangkan pada proses *cleaning data* menggunakan *library pandas* dan *numpy* untuk membersihkan *data* yang tidak diperlukan dengan menggunakan *dataframe*.

Penelitian yang dilakukan akan langsung dibentuk visualisasi melalui *tools* Power BI dan diberikan kepada CSIS Indonesia untuk ditampilkan pada *website* CSIS Indonesia [4]. Dengan adanya penelitian ini maka terdapat *output* dalam bentuk model yang memprediksi apakah *Tweet* yang ada pada Twitter selanjutnya mengandung ujaran kebencian dan bagaimana peningkatan *hate speech* di Indonesia seiring bertambahnya waktu. Fenomena ujaran kebencian bukanlah hal baru di Indonesia. Sejak konsolidasi demokrasi, berbagai aktor secara berkala menghasut kampanye ujaran kebencian terhadap kelompok rentan, khususnya agama dan etnis minoritas untuk memperkuat kampanye, membangun solidaritas, dan memobilisasi anggota dalam mendukung agenda sosial politik. Ujaran kebencian dan hasutan telah menyebabkan kerusakan fisik dan sistematis yang signifikan terhadap minoritas seperti Ahmadiyah, Syiah, Kristen, dan Tionghoa Indonesia. Pada tahun 2015 hingga 2020, unit *Indonesia's Cyber Patrol* mencatat lebih dari 7.460 kasus yang dilaporkan tentang orang yang menyebarkan konten provokatif (termasuk ujaran kebencian dan hasutan). Arti penting dari ujaran kebencian *online* paling jelas dalam kampanye melawan Gubernur Jakarta, Basuki Tjahaja Purnama, dimana postingan kebencian *online* akhirnya menyebabkan protes massal dan pemenjaraan. Namun, ditengah peningkatan prevalensinya, saat ini tidak ada alat di Indonesia yang cukup untuk secara kuantitatif mengumpulkan dan mengidentifikasi tren ujaran kebencian *online* secara akurat di tingkat nasional. Laporan dan studi yang ada tentang ujaran kebencian di Indonesia sebagian besar berfokus pada kampanye ujaran kebencian individu dan kebanyakan menggunakan metodologi kualitatif daripada kuantitatif.

Dashboard Ujaran Kebencian Nasional dikembangkan untuk mengisi kesenjangan dan memberikan visualisasi kuantitatif atas tren ujaran kebencian *online* di Indonesia. Secara khusus akan memberikan visualisasi pada dua tren penting: volume ujaran kebencian menurut topik menurut waktu, dan pemetaan jaringan aktor yang secara aktif membagikan konten ujaran kebencian. Dengan menyediakan tren tersebut, dasbor ini bertujuan untuk memberikan gambaran yang lebih baik tentang besaran, tren, dan pola isu ujaran kebencian di Indonesia, yang dapat membantu mendorong urgensi isu tersebut kepada pembuat kebijakan dan publik.

1.2 Tujuan Kerja Magang

Kerja magang dijadikan salah satu syarat kelulusan dalam mata kuliah magang memiliki banyak nilai positif bagi mahasiswa salah satunya untuk mendapatkan pengalaman atau ilmu baru yang lebih dalam terkait bidang yang diminati. Dengan adanya kerja magang dapat menerapkan ilmu-ilmu yang dipelajari dalam dunia kerja yang sesungguhnya. Melalui kerja magang ini terdapat beberapa tujuan yang diperoleh perusahaan antara lain mahasiswa magang dapat berkontribusi dalam pengambilan *data* dan pembersihan *data* pada model *machine learning* yang akan dibuat. Pengambilan *data* dan pembersihan *data* yang dilakukan diharapkan dapat membentuk model yang memiliki tingkat akurasi tinggi dan berfungsi untuk prediksi *hate speech* di masa mendatang. *Dashboard* yang dihasilkan dijadikan informasi baru bagi masyarakat Indonesia akan *awareness* ujaran kebencian di Indonesia melalui media sosial.

1.3 Waktu dan Prosedur Pengerjaan Magang

Berikut merupakan penjelasan mengenai waktu dan prosedur pelaksanaan selama kerja magang berlangsung.

1.3.1 Waktu Pelaksanaan Kerja Magang

Magang dilaksanakan pada tanggal 12 Mei 2021 hingga 11 Agustus 2021 dan dilakukan pada jam kerja 9.00 hingga 17.00. Prosedur kerja magang adalah melakukan studi literatur terlebih dahulu untuk mengetahui teknologi apa yang akan digunakan terutama dalam menggunakan *library* Python terkait *scraping* dan *cleaning*. Pada proses pengerjaan proyek ini dilakukan secara *end to end*. Studi literatur dilakukan selama 3 hari dan pengambilan *data* serta proses pembentukan *code* dilakukan kurang lebih 2 minggu. *Timeline* pelaksanaan kerja magang dapat dilihat pada tabel 1.1.

Tabel 1. 1 Timeline Kerja Magang

No	Pekerjaan yang dilakukan	Waktu Pelaksanaan (Per Minggu)							
		1	2	3	4	5	6	7	8
1	Riset jurnal dan informasi terkait teknik pengambilan <i>data</i> melalui Twitter.								
2	Melakukan pengambilan <i>data</i> dengan kata kunci Syiah.								
3	<i>Scraping data</i> Twitter dengan kata kunci Cina dan implementasi regex pada hasil <i>scraping</i> .								
4	<i>Filtering value</i> 1 dan 0. <i>Merge</i> semua <i>data</i> dan menambahkan kolom baru terkait topik Syiah atau Cina.								
5	Melakukan <i>scraping</i> ulang terkait <i>data Tweet</i> Syiah dengan kriteria <i>geo location</i> Jakarta, Indonesia.								
6	<i>Split data</i> untuk diberikan kepada CSIS Indonesia agar dapat di anotasi.								
7	Riset mengenai model <i>machine learning</i> .								
8	<i>Scraping dan Cleaning data</i> Twitter dengan kata kunci “Ahmadiyah” menggunakan <i>geo location</i> Jakarta, Indonesia. <i>Clustering Tweet</i> berdasarkan kata kunci. <i>Finalize code</i> untuk digunakan oleh CSIS.								

1.3.2 Prosedur Pelaksanaan Kerja Magang

Prosedur pelaksanaan kerja magang di Blankspace yang berlaku adalah sebagai berikut:

- 1) Mahasiswa melakukan pelamaran dengan mengirimkan CV melalui email Blankspace.
- 2) Mahasiswa mengajukan pembuatan Surat Pengajuan Magang dari Kampus dengan mengirim *form* KM-01.
- 3) *Team* Blankspace menginformasikan jadwal wawancara dengan *user* yang dilakukan secara *online* melalui Google Meet.
- 4) *Team* Blankspace menginformasikan keputusan bahwa diterimanya mahasiswa magang melalui email.
- 5) Mahasiswa mengisi *form* KM-02 untuk mendapatkan Surat Izin Kerja Magang.
- 6) Pihak Blankspace mengirimkan surat yang berisi informasi kontrak kerja magang yang berisi keterangan jangka waktu magang beserta deskripsi pekerjaan yang dilakukan sebagai *Data Engineer Intern*.
- 7) Mahasiswa menyerahkan surat tersebut kepada sekretaris program studi Sistem Informasi Universitas Multimedia Nusantara.
- 8) Setelah mendapatkan perizinan magang, mahasiswa mulai melanjutkan pengisian *form* KM-03, KM-04, KM-05, KM-06 dan KM-07 yang sudah disediakan pada *website* my.umn.ac.id.