

## DAFTAR TABEL

1.1	Rencana Penelitian . . . . .	5
2.1	Modul NLTK . . . . .	20
3.1	Syarat Penentuan <i>Label Tweet</i> . . . . .	42
4.1	Contoh <i>Tweet</i> Sebelum dan Setelah <i>Preprocessing</i> . . . . .	50
4.2	Pembagian <i>Tweet</i> untuk <i>Labelling</i> . . . . .	52
4.3	Jumlah Data <i>Training Model</i> . . . . .	55
4.4	Hasil Pengujian Model . . . . .	56
4.5	Prediksi pada Kandidat Partai Demokrat . . . . .	58
4.6	Prediksi pada Kandidat Partai Republik . . . . .	58
4.7	Perbandingan Urutan Prediksi dan <i>Polling</i> pada Kandidat Partai De- mokrat . . . . .	59
4.8	Perbandingan Urutan Prediksi dan <i>Polling</i> pada Kandidat Partai Re- publik . . . . .	61
4.9	Jumlah <i>Tweet</i> per Kandidat Setiap <i>n Tweet</i> Pertama . . . . .	64
4.10	Hasil Prediksi <i>Polling</i> 19 Januari dan 9 Februari 2016 . . . . .	65
4.11	Perbandingan Hasil <i>Preprocessing</i> . . . . .	67
4.12	Perbandingan <i>Label</i> pada <i>Tweet</i> dengan Isi yang Sama . . . . .	69
4.13	Sampel <i>Tweet</i> ”Tidak Jelas” yang Disingkirkan . . . . .	70
4.14	Hasil Analisis Diskusi . . . . .	75

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang

Pemilihan presiden merupakan momen penting bagi setiap negara, salah satunya yaitu Amerika Serikat. Pergerakan pasar sekuritas di New York, London, Toronto, dan Tokyo cenderung bergerak ke arah yang sama selama masa Pemilihan Presiden Amerika Serikat (Dobson & Dufrene, 1993). Setiap kebijakan ekonomi Amerika Serikat akan mempengaruhi perekonomian dunia, baik dalam kurun waktu singkat ataupun panjang (Stokes, 2012). Oleh karena itu, prediksi terhadap Presiden Amerika Serikat berikutnya menjadi penting. Prediksi terhadap Presiden Amerika Serikat dapat dilakukan dengan memanfaatkan media sosial seperti *Facebook* dan *Twitter*, yang marak digunakan sebagai media kampanye sejak Pemilihan Presiden Amerika Serikat 2008 (Kreiss, 2014).

Analisis sentimen pada *tweet* mengenai kandidat Presiden Amerika Serikat 2008 telah dilakukan dengan menghitung nilai sentimen berdasarkan jumlah kata positif dan negatif pada setiap *tweet* (O'Connor, Balasubramanyan, Routledge, & Smith, 2010). Satu juta *tweet* dikumpulkan selama 2008-2009, di mana daftar kata positif serta negatif diperoleh dari *OpinionFinder*. Rasio sentimen per kandidat (Obama dan McCain) dihitung dengan *moving average*. Penelitian ini menemu-

kan bahwa perbandingan sentimen pada *tweet* dengan *polling* pemilihan cenderung kompleks karena *polling* memungkinkan masyarakat memilih Obama, McCain, belum memutuskan, tidak berencana memilih, atau kandidat independen.

Penelitian lain melakukan prediksi terhadap hasil pemilihan kandidat presiden dari Partai Republik menggunakan sekitar 300 juta *tweet* yang dikumpulkan melalui *Twitter Streaming API* dan sekitar 10 juta *tweet* yang diperoleh melalui *Twitter API endpoint* selama September 2011 hingga Februari 2012 (Shi, Agarwal, Agrawal, & Garg, 2012). Prediksi dilakukan dengan menggunakan model regresi linear, kemudian dibandingkan dengan *polling* pada *RealClearPolitics.com*. Penelitian ini membuktikan bahwa *polling* dapat dilengkapi, atau bahkan digantikan dengan hasil analisis data dari media sosial.

Prediksi hasil Pemilihan Presiden Amerika Serikat 2012 telah dilakukan dengan memanfaatkan algoritma *Naive Bayes Classifier* pada 32 juta *tweet* yang dikumpulkan menggunakan *Tweepy* pada 29 September hingga 16 November 2012 (Jahanbakhsh & Moon, 2014). *Tweet* dikumpulkan berdasarkan kata kunci untuk masing-masing kandidat, kemudian melalui tahap *preprocessing* meliputi penghapusan URL, *mention*, *RT*, *hashtags*, angka, dan *stop words*; tokenisasi; dan penambahan kata *not\_* pada negasi. Analisis sentimen dilakukan pada 10.000 *tweet* yang hanya mengandung satu nama kandidat. Hasil analisis dibandingkan dengan *polling* pada *Huffington Post*, di mana terbukti bahwa popularitas Barack Obama di *Twitter* sesuai dengan hasil pemilihan. Penelitian ini tidak menggunakan *tweet* yang mengandung lebih dari satu nama kandidat karena membutuhkan teknik pemrosesan

teks yang lebih kompleks.

Pada penelitian lain, analisis sentimen pada *tweet* mengenai Pemilihan Presiden Amerika Serikat 2012 dibuat dan ditampilkan pada sebuah *dashboard* secara *real-time* (Wang et al., 2012). *Tweet* mengenai masing-masing kandidat presiden (Barack Obama serta 9 kandidat Partai Republik) dikumpulkan sejak 12 Oktober 2012 menggunakan *Gnip Power Track*. *Tweet* yang terkumpul dipecah menjadi *token*, kemudian ditandai sentimennya oleh sekitar 800 orang melalui *Amazon Mechanical Turk* (AMT). Sebanyak 17.000 *tweet* (16% positif, 56% negatif, 18% netral, 10% tidak yakin) dilatih menggunakan *Naive Bayes Classifier* dan diperoleh akurasi 59%.

Menjelang Pemilihan Presiden Amerika Serikat 2016, masing-masing kandidat dari Partai Republik dan Demokrat memanfaatkan *Twitter* sebagai media kampanye. Chin et al. (2015) telah menganalisis sentimen pada beberapa kandidat utama berdasarkan *tweet* yang mengandung *emoji* dan berhasil membuat model prediksi menggunakan *Support Vector Machine* (SVM) dengan akurasi 49,22% pada lima kelas: (1) senang, (2) sedih, (3) takut, (4) tertawa, dan (5) marah. Namun, penelitian tersebut terbatas pada *tweet* yang mengandung *emoji* dan jumlahnya hanya sedikit. Oleh karena itu, penelitian ini menganalisis sentimen pada *tweet* seputar kandidat Presiden Amerika Serikat 2016 secara umum. Sentimen akan dikelompokkan menjadi positif dan negatif untuk masing-masing kandidat dengan menggunakan *Naive Bayes Classifier* (Wang et al., 2012). Data dikumpulkan sejak 16 Desember 2015 hingga 29 Februari 2016 dengan kata kunci *#Election2016*.

## 1.2 Rumusan Masalah

Dalam penelitian ini, masalah yang ingin diselesaikan penulis yaitu

1. Bagaimana sentimen pengguna *Twitter* terhadap kandidat Presiden Amerika Serikat 2016?
2. Bagaimana akurasi model *Naive Bayes Classifier* yang dibangun?
3. Bagaimana akurasi prediksi model *Naive Bayes Classifier* jika dibandingkan dengan hasil *polling* pada situs *RealClearPolitics.com*?

## 1.3 Batasan Masalah

Dalam melakukan penelitian ini, penulis memiliki beberapa batasan ruang lingkup penelitian, antara lain

1. Sumber data yang digunakan dalam penelitian merupakan *tweet* dengan kata kunci *#Election2016*.
2. Tahap *preprocessing* yang dilakukan hanya menghilangkan URL dan gambar pada isi *tweet*.
3. Akurasi model dinyatakan baik apabila mencapai 59% (Wang et al., 2012).
4. Hasil prediksi dibandingkan dengan *polling* pada situs *RealClearPolitics.com* (Shi et al., 2012).

## 1.4 Tujuan dan Manfaat Penelitian

### 1.4.1 Tujuan Penelitian

Penelitian ini dilakukan dengan tujuan untuk melakukan prediksi terhadap hasil *polling* kandidat Presiden Amerika Serikat 2016.

### 1.4.2 Manfaat Penelitian

Hasil penelitian ini diharapkan dapat menjadi sumber prediksi Presiden Amerika Serikat 2016. Selain itu, metode penelitian yang digunakan diharapkan dapat menjadi alternatif untuk melakukan prediksi pada penelitian berikutnya.

## 1.5 Rencana Kegiatan

Secara umum, penelitian ini terdiri dari sejumlah kegiatan yang dapat dilihat pada Tabel 1.1.

**Tabel 1.1 Rencana Penelitian**

No	Kegiatan	Minggu Ke-													
		1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	Persiapan Penelitian														
2	Pengumpulan Data														
3	<i>Data Preprocessing</i>														
4	<i>Data Labelling</i>														
5	Pelatihan Model														
6	Pengujian Model														
7	Analisis Hasil														

## 1.6 Sistematika Penulisan

Laporan penelitian ini disusun secara sistematis menjadi lima bagian yang menjelaskan tahap-tahap yang dilakukan penulis, yaitu

1. Bab I: Pendahuluan

Berisi latar belakang penelitian, rumusan dan batasan masalah, serta tujuan dan manfaat yang diperoleh dari penelitian. Bab ini juga mencakup rencana penelitian dan sistematika penulisan laporan.

2. Bab II: Landasan Teori

Berisi teori-teori pendukung yang digunakan dalam penelitian.

3. Bab III: Metodologi Penelitian

Menjelaskan metodologi yang digunakan dalam proses penelitian.

4. Bab IV: Analisis dan Pembahasan

Berisi penjelasan tahap penelitian serta analisis dan pembahasan hasil penelitian yang dilakukan.

5. Bab V: Simpulan dan Saran

Berisi simpulan yang diperoleh dari hasil penelitian serta saran bagi pengembangan berikutnya.

## 1.7 Sumber Dana

Penelitian ini merupakan bagian dari penelitian yang didanai oleh Lembaga Penelitian dan Pengabdian Masyarakat Universitas Multimedia Nusantara (LPPM UMN) dengan nomor 078/LPPM-UMN/III/2016.

