

Bab 1. Pendahuluan

Pada bab ini, akan dibicarakan mengenai segala hal yang menjadi awal mula dilakukannya penelitian yang meliputi latar belakang, permasalahan, tujuan penelitian, metodologi yang digunakan dalam penelitian serta sistematika penulisan.

1.1 Latar Belakang

Twitter adalah salah satu media sosial yang populer di kalangan masyarakat. Terhitung 21 Maret 2016, media sosial ini secara global memiliki sekitar 332 juta pengguna bulanan, dengan 500 juta *tweet* dikirim setiap hari dan 200 miliar *tweet* dalam setahun[1]. Kota Bandung menjadi kota di Indonesia yang menempati urutan ke enam di dunia pada kategori kota paling aktif nge-*tweet*[2].

Sangat mudah berbagi informasi di media sosial *Twitter*, dengan hanya mem-*follow* suatu akun pada *Twitter*, otomatis kita akan mendapatkan setiap *tweet* yang di *post* oleh akun tersebut, begitu pun sebaliknya informasi yang kita *post* melalui *tweet* juga otomatis akan diterima oleh setiap akun yang mem-*follow* kita. Hal ini menunjukkan betapa mudahnya kita berbagi informasi di *Twitter*.

Namun sekarang ini orang-orang memanfaatkan kemudahan tersebut untuk tujuan promosi. Mereka dapat dengan mudah mempromosikan suatu produk kepada orang lain melalui *tweet* yang di *post*-nya[3]. Meskipun *Twitter* hanya membatasi jumlah karakter dalam setiap *tweet* nya, biasanya mereka para pengguna yang ingin mempromosikan produknya melampirkan *link* ke dalam *tweet* promosinya untuk memberikan keterangan lebih lanjut tentang produknya. Hal tersebut sangat berguna bagi mereka para pengguna yang memang menginginkan iklan tersebut, tapi sebaliknya *tweet* tersebut dapat dianggap sebagai *spam* apabila isi dalam *tweet* tersebut tidak relevan dengan *link* yang dilampirkannya, tentunya hal ini akan merugikan mereka para pengguna yang menginginkan iklan tersebut[4]. Maka dari itu dibutuhkan suatu sistem yang dapat mengklasifikasi *tweet* tersebut kedalam dua kategori yaitu *spam* dan *non spam*.

Hye-Jin Kim dan Yong-Hyuk Kim telah melakukan klasifikasi *tweet spam* berisi *link* dengan metode *Naïve Bayes* dan *URL Analysis* dan menghasilkan akurasi yang cukup baik yaitu 74% dan 82%[5], namun pada penelitian tersebut tidak dilakukan *preprocessing* terhadap dataset yang digunakan. Menurut Datta H.Deshmukh, Tushar Ghorpade, dan Puja Padiya pada penelitiannya, *preprocessing* yang dilakukan sebelum pengklasifikasian dapat meningkatkan akurasi *classifier*[6].

Pada penelitian ini *Naive Bayes* dan *URL Analysis* akan diujikan kembali pada dataset berbahasa Indonesia dengan melakukan *preprocessing* terhadap dataset untuk mengetahui performansi setiap metode tersebut dalam melakukan klasifikasi data yang berupa *tweet* berbahasa Indonesia yang terdapat *link* di dalamnya. Selanjutnya hasil dari penelitian pada tugas akhir ini diharapkan dapat dimanfaatkan untuk penelitian selanjutnya mengenai analisis *spam tweet* pada *Twitter*.

1.2 Perumusan Masalah

Berdasarkan pada latar belakang diatas permasalahan yang akan diuji dan diteliti adalah :

1. Bagaimana cara mengimplementasikan metode *Naïve Bayes* dan *URL Analysis* dalam mengklasifikasikan data *tweets* berbahasa Indonesia?

2. Bagaimana performansi sistem klasifikasi *spam* dan *non-spam* pada *tweets* berbahasa Indonesia dengan metode *Naïve Bayes* dan *URL Analysis*?

Dalam penyelesaian masalah yang diangkat dalam penelitian, terdapat beberapa lingkup yang menjadi batasan dari penelitian ini. Adapun batasan masalah tersebut adalah :

1. *Dataset* yang akan digunakan adalah *tweet* berbahasa Indonesia yang dimana terdapat *URL* didalamnya.
2. Data yang diambil hanya berupa teks, tidak termasuk suara, gambar, dan *video*.
3. Sistem melakukan pembelajaran secara *offline*.

1.3 Tujuan

Tujuan dari penyusunan Tugas Akhir ini adalah :

1. Mengimplementasikan metode *Naive Bayes* dan *URL Analysis* untuk klasifikasi *spam-non spam tweets* berbahasa Indonesia.
2. Membandingkan hasil klasifikasi dari kedua metode yang digunakan yaitu *Naïve Bayes* dan *URL Analysis*.

1.4 Metodologi Penyelesaian Masalah

Metodologi untuk menyelesaikan masalah dalam penelitian ini adalah sebagai berikut :

1. Studi Literatur
Studi literatur yang digunakan dalam penelitian ini berdasarkan paper baik nasional maupun internasional, buku, informasi dari ahli bidang keilmuan yang bersangkutan, dan website. Penelusuran yang dilakukan adalah mengenai *Twitter*, *Naive Bayes*, *URL Analysis*, *preprocessing* pada data teks dan evaluasi sistem yang nantinya akan digunakan sebagai dasar untuk melakukan penelitian.
2. Pengumpulan Data
Data didapatkan dengan cara memanfaatkan *API* yang sudah disediakan oleh *Twitter* dan di ambil dengan menggunakan *Tweepy*.
3. Perancangan Sistem
Perancangan sistem dapat dilihat pada *flowchart* yang dijelaskan pada bab III buku tugas akhir ini.
4. *Preprocessing*
Preprocessing dilakukan untuk merapikan *dataset*. *Preprocessing* yang dilakukan adalah dengan *case holding*, *tokenization*, *stopword removal*, *stemming* dan menghapus *blank space* dalam kalimat.
5. Implementasi
Data yang telah dikumpulkan akan di klasifikasikan menggunakan metode *Naïve Bayes* dan *URL Analysis*.
6. Pengujian
Pada tahap ini adalah melakukan evaluasi terhadap impelementasi yang dilakukan. Tujuannya adalah untuk mengetahui akurasi terhadap data yang diperoleh dengan menggunakan metode *Naive Bayes* dan *URL Analysis*.
7. Analisis
Tahap pengujian dari sistem yang sudah dibangun apakah sudah sesuai dengan yang diharapkan dan membandingkan hasil klasifikasi dari kedua metode yang digunakan.
8. Pembuatan Laporan
Seluruh rangkaian proses pengujian dan analisis yang telah dilakukan akan ditarik kesimpulan yang sesuai kemudian dituangkan dalam bentuk laporan.

1.5 Sistematika Penulisan

Sistematika penulisan yang digunakan dalam menyusun laporan Tugas Akhir ini adalah :

BAB 1 : Pendahuluan

Bab pendahuluan, berisikan penjelasan singkat mengenai latar belakang, rumusan masalah, tujuan, batasan masalah, metodologi penyelesaian masalah, dan sistematika penulisan laporan. Pada bab ini, menjelaskan mengenai judul yang telah dipilih, yang diantaranya adalah apa penelitian yang akan dilakukan, mengapa melakukan penelitian tersebut, apa tujuan dari penelitian, serta penjelasan singkat mengenai tahapan-tahapan yang akan dilakukan.

BAB 2 : Kajian Pustaka

Pada bab ini, dijelaskan mengenai landasan atau dasar teori yang digunakan dan berkaitan dalam pengerjaan Tugas Akhir. Teori yang tertulis dalam bab ini merupakan rujukan dan acuan yang digunakan selama penelitian.

BAB 3 : Perancangan Sistem

Pada bab ini, dijelaskan mengenai tahap-tahap perancangan dan pengerjaan sistem yang akan dibangun dalam Tugas Akhir ini. Tahapan-tahapn tersebut diantaranya adalah pengumpulan data, *preprocessing*, implementasi metode *Naive Bayes* terhadap data, implementasi *URL Analysis* terhadap data .

BAB 4 : Pengujian dan Analisis

Pada bab ini berisi hasil pengujian yang dilakukan terhadap sistem dan hasil analisis pengujiannya.

BAB 5 : Kesimpulan dan Saran

Pada bab ini, berisikan kesimpulan yang ditarik dari hasil analisis yang telah dilakukan serta saran pengembangan dan perbaikan sistem selanjutnya.