

BAB II

LANDASAN TEORI

1.1 Pengertian Pelecehan *Online* (*Online Harassment*)

Pelecehan *online* (*Online Harassment*) disebut sebagai "*cyber harassment*" biasanya berkaitan dengan *e-mail*, pesan instan, atau komentar situs web, yang mengancam atau melecehkan. Hal ini terkadang sering dilakukan berulang kali upaya untuk menargetkan orang tertentu dengan langsung menghubungi mereka, atau secara tidak langsung menggunakan atau menyebarkan informasi pribadi mereka, menyebabkan mereka tertekan, takut, atau marah.

Istilah "*Online Harassment*" telah digunakan untuk menggambarkan berbagai perilaku yang melibatkan ancaman berulang dan atau pelecehan dengan penggunaan *e-mail* atau komunikasi berbasis komputer lainnya yang membuat orang takut atau khawatir akan keselamatan mereka. *Online Harassment* bisa sama mengancamnya seperti mengintai di dunia nyata dan bisa menimbulkan kegelisahan, kesedihan mental, kerusakan fisik, dan bahkan pembunuhan.

Pelecehan *Online* dapat mencakup hal berikut (Privacyright.org, 2017) :

1. Mengirimkan *e-mail* berulang atau pesan instan yang mungkin atau mungkin tidak langsung mengancam penerima;
2. Membanjiri kotak *e-mail* korban dengan surat yang tidak diinginkan, mengirim berkas korban dengan virus;
3. Menggunakan alamat *e-mail* korban untuk berlangganan ke beberapa layanan *online*. Mengirim informasi yang keliru dan pesan yang salah ke forum-forum, atau situs tempat kerja korban;
4. Mencuri identitas *online* seseorang yang digunakan untuk mengirim informasi palsu.
5. Mengirim informasi demografis korban dan mengirim gambar yang berorientasi seksual atau situs pornografi.
6. Mencari dan mengumpulkan berbagai informasi korban yang mungkin telah diposkan oleh situs berita dengan maksud untuk mencari informasi pribadi dan

kemudian menggunakan ini informasi untuk melecehkan, mengancam, dan mengintimidasi korban baik secara *online* maupun di dunia nyata.

1.1.1 Kasus-Kasus Pelecehan Online

Di Indonesia pelecehan *online* kerap menjadi berita-berita media massa antara lain tindak kekerasan berupa berita isu, rumor, dan pelecehan seksual. Meningkatnya penetrasi internet di Indonesia membuat kasus pelecehan *online* di media sosial makin marak terjadi. Di Indonesia, pelecehan *online* tidak hanya menimpa kalangan selebritas, tetapi juga masyarakat biasa. Beberapa kasus bahkan menyebabkan korbannya bunuh diri. Akan tetapi, pelaku merasa tidak bersalah. Contohnya yang dialami artis berinisial ZG akibat melecehkan lambang negara pada Agustus 2015. Berkali-kali pelaku meminta maaf dan berurusan dengan polisi, tidak membuat ZG lepas dari hujatan pada akun media sosialnya.

Salah satu kasus yang berakhir tragis telah dialami oleh YC. Ketua Festival Musik Lockstock 2013 di Yogyakarta ini menabrakkan dirinya ke kereta api setelah akun Twitter-nya dibanjiri oleh hujatan akibat kegagalan festival yang ia tangani. Sebelum bunuh diri, alumnus FISIPOL UGM itu menuliskan salam perpisahan di akun Twitter-nya pada 25 Mei 2013, @effxxx_creatxxx, *“Trimakasih atas segala caci maki @locstockxxx.. Ini gerakan. Gerakan menuju Tuhan.. Salam”*. Tak hanya di Twitter, ia pun menulis status terakhir di Facebook pada 26 Mei 2013 melalui akunnya yang bernama BY, *“Selamat pagi teman2 semua. Mohon maaf atas segala kekurangan dan kelemahan saya, jgn pernah menuntut klg, anak, istri, dan orangtua saya, dan tdk yg harus mrk lakukan.. Berkat Tuhan slalu brsama kalian..”*. Berbagai kasus di atas menunjukkan bahwa perilaku pelecehan di media sosial merupakan masalah serius yang memerlukan penanganan lebih lanjut, baik penanganan yang bersifat kebijakan ataupun kode etik yang disepakati bersama.

Selain itu, ada pula yang menimpa siswi SMA di Medan berinisial SD pada April 2016. SD membentak polisi wanita (polwan) dan mengaku sebagai anak jenderal pada saat ditilang. Hal yang sama dialami oleh selebritas Instagram berinisial AWK yang menjadi bulan-bulanan di media sosial pada pertengahan tahun 2016 karena *posting-an* yang ia unggah dianggap tidak pantas untuk remaja

berusia 19 tahun, seperti berpakaian minim, merokok, minum minuman keras, romantisme bersama kekasih, dan kerap menggunakan kata-kata yang kasar.

Kasus penyimpangan dalam etika berinternet, pelecehan dan perundungan siber di Indonesia cukup tinggi, sebagaimana temuan dari penelitian yang dilakukan oleh Kementerian Komunikasi dan Informatika bersama UNICEF pada tahun 2011 hingga 2012 (Kominfo, 2012). Penelitian tersebut melibatkan 400 anak dan remaja pada rentang usia 10 hingga 19 tahun yang berada di 11 provinsi di Indonesia. Hasil riset menunjukkan bahwa 13% menyatakan mengalami pelecehan *online* dalam bentuk hinaan, ancaman, dan dipermalukan di media sosial dan pesan teks. Tidak hanya itu, 9% menyatakan pernah mengirimkan pesan berupa hinaan dan kemarahan melalui media sosial dan 14% melalui pesan teks. Jumlah ini berarti 13 dari 100 responden merupakan korban perundungan siber, dan 23 dari 100 responden merupakan pelaku perundungan siber.

Perwakilan nasional Pew Research data 2017, survei yang dilakukan terhadap 4.248 orang dewasa A.S. menemukan bahwa 41% orang Amerika telah menjadi sasaran pribadi perilaku pelecehan *online*, dan bahkan lebih besar lagi (66%) telah menyaksikan perilaku pelecehan *online* ini mengarahkan pada orang lain. Dalam beberapa kasus, seperti menyebut nama ras yang menyinggung atau upaya memanggil untuk memermalukan seseorang. Hampir satu dari lima orang Amerika (18%) telah dikenai hukuman terutama dalam bentuk pelecehan yang parah secara *online*, seperti ancaman fisik, pelecehan seksual atau penguntit.

Platform media sosial menjadi tempat yang banyak digunakan untuk pelecehan *online*, seperti Instagram, Facebook, Twitter dll. Sering kali perilaku ini menargetkan pribadi atau karakteristik fisik, 14% orang Amerika mengatakan mereka telah dilecehkan secara *online* secara khusus karena politik mereka, sementara dari satu diantara sepuluh orang telah menjadi target pelecehan karena penampilan fisik mereka (9%), ras atau etnis (8%) dan jenis kelamin (8%).

Pelecehan sering difasilitasi oleh anonimitas yang disediakan internet, pelecehan ini bisa terjadi pada kenalan di *online*, teman atau bahkan anggota keluarga. Bagi mereka yang mengalami pelecehan *online* secara langsung, pelecehan secara langsung dapat memiliki konsekuensi di dunia nyata, mulai dari

tekanan mental atau emosional hingga rusaknya reputasi atau bahkan ketakutan dan terancamnya keamanan pribadi seseorang.

Sedangkan sebaliknya, dampak pelecehan yang dialami secara tidak langsung, meninggalkan dampak bagi korban. Sekitar seperempat orang Amerika (27%) mengatakan bahwa mereka telah memutuskan untuk tidak mem-*posting* sesuatu yang *online* setelah menyaksikan pelecehan orang lain, sementara lebih dari satu per sepuluh (13%) mengatakannya mereka telah berhenti menggunakan layanan *online* setelah menyaksikan pengguna lain terlibat dalam pelecehan perilaku.

Kemudian beberapa pengamat pelecehan *online* yang berperan aktif mengatakan. Bahwa tiga dari sepuluh orang Amerika (30%) mengatakan bahwa mereka telah melakukan intervensi dalam beberapa cara setelah menyaksikannya. Perilaku kasar mengarahkan pada orang lain secara *online*. Namun, bahkan saat pelecehan menembus banyak interaksi pengguna *online*, data publik yang telah dilakukan memberikan pandangan konflik tentang cara terbaik untuk mengatasi masalah ini. Mayoritas orang Amerika (62%) melihat pelecehan *online* sebagai masalah utama, dan hampir delapan dari sepuluh orang Amerika (79%) mengatakan layanan *online* memiliki kewajiban untuk melangkah menangani di saat pelecehan *online* terjadi pada platform mereka. Di sisi lain, mereka (platform media sosial) sangat terbagi atas bagaimana caranya untuk menyeimbangkan kekhawatiran keamanan dengan keinginan untuk mendorong kebebasan berbicara dan terbuka dan juga apakah konten yang melecehkan secara *online* terlalu serius atau diberhentikan terlalu mudah (Maeve Duggan, 2017).

1.1.2 Jenis-Jenis Pelecehan *Online* Pada Twitter

Adapun jenis pelecehan *online* yang terjadi di media sosial menurut survei yang dilakukan oleh Pew Research Center 2017 yaitu:

1. Rasis (*Racist*)

Rasisme adalah suatu gagasan atau teori yang mengatakan bahwa kaitan kausal antara ciri-ciri jasmaniah yang diturunkan dan ciri-ciri tertentu dalam hal kepribadian, intelek, budaya atau gabungan dari semua itu, menimbulkan superioritas dari ras tertentu terhadap yang lain. Rasisme berasal dari dominasi dan

menyediakan dasar pemikiran sosial dan filosofis pembenaran untuk merendahkan dan melakukan kekerasan terhadap orang berdasarkan warna.

Adapun bentuk penggunaan kata rasis yang belakangan bulan ini sedang *trend* adalah sentimen anti Cina banyak terjadi bahkan dimana saja. Ada yang mengekspresikan sentimen itu dengan terbuka di publik (baik publik "dunia nyata" maupun "dunia maya" seperti media sosial). Ada pula yang melampiaskannya secara diam-diam dan sembunyi-sembunyi, misalnya dengan menggunakan bahasa-bahasa simbol tertentu atau menggunjing mereka di "belakang panggung". Kata "aseng dan pribumi" yang merupakan sebutan peyoratif dan sarkastik untuk Cina dan penduduk asli Indonesia yang kini sedang ramai, ini menjadi bagian dari "bahasa simbol" sekaligus ekspresi dari sentimen anti Cina dan pribumi (Sumanto Al Qurtuby, 2017).

2. Olokan atau Ejekan (*Embarrass*)

Menurut Kamus Besar Bahasa Indonesia (KBBI) mengejek dan mengolok-olok (menertawakan, menyindir) untuk menghinakan (mempermainkan dengan tingkah laku yang ia tidak disukai temannya). Dalam kehidupan berjejaring sosial Twitter tidak hanya digunakan oleh informan sebagai tempat untuk berkomunikasi dengan orang lain, tetapi menjadi tempat untuk melakukan ejekan dan olokan dengan menyampaikan masalah pribadi melalui sebuah *posting*-an yang berisikan pesan ejekan atau olok-olokkan di Twitter. *Posting*-an tersebut merupakan salah satu bentuk pelecehan *online* yang telah diidentifikasi oleh Williard dan Pew Research Global.

Dikutip dari situs (BBC Indonesia, 2017) Bahasa Jawa adalah bahasa yang unik dengan keragaman kosa kata termasuk untuk urusan menyindir atau ejek-mengejek, kata ahli bahasa Adi Deswijaya Dosen Bahasa dan Sastra Jawa dari Universitas Veteran Bangun Nusantara Sukoharjo. "Makian itu banyak sekali (dalam bahasa Jawa) nama orang tua dibawa-bawa, hewan dibawa, buah dibawa, status ekonomi dibawa. Bahasa Jawa itu bahasa unik. Contoh kata ejekan yaitu kambing dan ndeso yang artinya mirip dengan kampung dalam bahasa Indonesia, merujuk pada perilaku yang seperti orang desa. "Ada kesan meremehkan dalam kata ndeso, *kok tingkahmu ndeso, kayak orang ndeso,*" katanya walau menambahkan bahwa

konotasinya bisa baik atau buruk tergantung dengan konteks pemakaian. "Kalau (bilang) "dasar ndeso" itu sudah termasuk ejekan," katanya. "Mengganggap perilaku seperti itu seperti orang desa. Bukan orang desa betulan, tetapi kampungan bahasa Indonesia-nya.

3. Ancaman Fisik (*Physically Threatened*)

Tweet yang mengancam fisik tidak banyak orang tahu kalau pelecehan *online* yang dilakukan dengan kata-kata yang mengancam fisik, tertentu ternyata memiliki efek yang lebih dahsyat dibandingkan dengan pelecehan yang dilakukan dengan kekerasan fisik. Sebab, efeknya memang tidak terlihat tapi cukup membuat emosi korban menjadi sangat-sangat lemah yang menikam sekali ke dalamnya (jiwa). Oleh sebab itu biasanya tingkat bunuh diri paling banyak berasal dari pelecehan yang dengan ancaman fisik.

Contoh ancaman fisik yang terjadi dimedia sosial yang mengandung unsur menyerang dan mengancam fisik dengan tujuan menyakiti (ada unsur ancaman fisik seperti dalam kalimat "*mending diusir saja dari jogja*" dan "*siap-siap nyawamu terancam di jogja*"). Perilaku agresif juga dapat dijumpai biasanya ini merupakan tanggapan atas peristiwa yang terjadi didunia nyata. "usir, tendang, labrak" adalah kata kata yang biasa digunakan dalam mengancam fisik (Apriliawati, 2015).

4. Pelecehan Seksual (*Sexually Harassed*)

Pelecehan seksual adalah perilaku atau perhatian yang bersifat seksual yang tidak diinginkan dan tidak dikehendaki dan berakibat mengganggu diri penerima pelecehan. Intinya pelecehan seksual merupakan suatu perbuatan tidak senonoh yang menyangkut seks, dimana perbuatan ini juga merupakan kendala bagi perkembangan kepribadian baik secara fisik maupun psikis, dan juga bisa mengancam jati diri korban, membuat sulit berkonsentrasi dan tidak percaya diri. Dan sebagian korban pelecehan seksual adalah para generasi bangsa (Triwijati, 2007).

Dalam penelitian (Gruber & Fineran, 2008) pelecehan seksual biasanya merupakan fenomena seks lintas jenis yang banyak terjadi, Penelitian yang dilakukan pada remaja menunjukkan bahwa pelecehan penyimpangan seks yang berbahaya. Saat ditanya untuk menggambarkan bagaimana menjengkelkan

pengalaman menjadi korban pelecehan seksual, anak perempuan menyebutkan tentang rumor seksual dan anak laki-laki mengatakan pelecehan dengan kata “alat kelamin” sebagai yang paling menjengkelkan dari semua kata-kata pelecehan.

1.2 Twitter

Twitter merupakan salah satu mikroblog paling populer di Indonesia. Mikroblog ini memungkinkan pengguna untuk mengirim dan membaca pesan yang disebut kicauan berupa teks maksimal 280 karakter yang ditampilkan pada halaman profil pengguna. *Tweet* bisa dilihat secara luar, namun pengirim dapat membatasi pengiriman pesan ke daftar teman-teman mereka saja. Selain itu, Twitter memiliki beberapa fitur andalan yang membuat Twitter menjadi salah satu media sosial yang digemari oleh para penggiat internet. Adapun objek-objek seperti, *tweet*, *re-tweet*, at atau *mention* (@), *hashtag* (#), waktu *tweet* dan *tweet search*. Sosial media yang dibuat oleh Jack Dorsey pada tahun 2006. Pada tahun 2016 Berdasarkan *Country Business Head Twitter Indonesia*, jumlah cuitan orang Indonesia selama Januari hingga Desember 2016 mencapai 4,1 miliar *tweet* (Liputan6.com, 2016).

1.3 Text Mining

Text Mining, yang juga disebut *Text Data Mining* (TDM) atau *knowledge discovery in the text*. Menurut (Berry, M.W. & Kogan, 2010) *Text Mining* merupakan teknik yang digunakan untuk menangani masalah klasifikasi, pengelompokan, ekstraksi informasi, dan pencarian informasi. Untuk data Twitter *text mining* dilakukan pada *unstructure text* untuk mengekstraksi teks yang tidak terstruktur menjadi sebuah informasi yang berguna.

Text Mining mengacu pada proses mengambil informasi berkualitas tinggi dari teks. Informasi berkualitas tinggi biasanya diperoleh melalui peramalan pola dan kecenderungan melalui sarana seperti pembelajaran pola statistik. *Text mining* biasanya melibatkan proses penataan teks input (biasanya *parsing*, bersama dengan penambahan beberapa fitur linguistik turunan dan penghilangan beberapa diantaranya, dan penyisipan *subsequent* ke dalam *database*), menentukan pola dalam data terstruktur, dan akhirnya mengevaluasi dan menginterpretasi *output*

berkualitas tinggi di bidang *text mining* biasanya mengacu ke beberapa kombinasi relevansi, kebaruan, dan *interestingness*. Proses *text mining* yang khas meliputi kategorisasi teks, *text clustering*, ekstraksi konsep/entitas, produksi *taksonomi granular*, *sentiment analysis*, penyimpulan dokumen, dan pemodelan relasi entitas.

1.3.1 *Text Preprocessing*

Tahap *text preprocessing* merupakan tahap awal dari *text mining*. *Text preprocessing* merupakan proses menggali, mengolah dan mengatur informasi dengan cara menganalisis hubungannya dengan aturan-aturan yang ada di data tekstual semi terstruktur atau tidak terstruktur (Luhulima, dan Muflikhah, 2013). Untuk lebih efektif dalam proses *text preprocessing*, dilakukan langkah transformasi data ke dalam suatu format yang memudahkan untuk kebutuhan pemakai. Setelah dalam bentuk yang lebih terstruktur dengan adanya proses di atas, data dapat dijadikan sumber data yang dapat diolah lebih lanjut. Tahapan *text preprocessing*, di antaranya sebagai berikut.

1. *Cleaning*

Adapun kata atau karakter yang akan dihilangkan adalah karakter atau simbol, *link URL*, *taggar (#)*, nama pengguna atau mention (@namapengguna), *emoticon* dan RT (*Retweet*).

2. *Case Folding*

Case Folding adalah mengubah semua karakter huruf menjadi huruf kecil (*lowercase*).

3. *Tokenizing*

Tokenizing yaitu proses penguraian deskripsi yang semula berupa kalimat-kalimat menjadi kata-kata dan menghilangkan delimiter-delimiter seperti tanda titik (.), koma (,), spasi dan karakter angka yang ada pada kata.

4. *Normalization*

Merupakan perbaikan dan substitusi kata yang salah eja ataupun disingkat dengan bentuk tertentu. Substitusi kata dilakukan untuk menghindari perhitungan dimensi kata yang melebar. Perhitungan dimensi kata akan melebar jika kata disingkat atau salah eja tidak diubah karena kata tersebut

memiliki kontribusi dalam mempresentasikan dokumen tetapi akan dianggap sebagai entitas yang berbeda proses penyusunan matriks.

5. *Stopword Removal*

Stopword removal yaitu proses penghapusan kata-kata yang terdapat pada *stoplist*. *Stoplist* itu sendiri berisi kosakata-kosakata yang bukan merupakan ciri dari suatu dokumen.

6. *Stemming*

Stemming adalah proses pemetaan dan penguraian berbagai bentuk (*variants*) dari suatu kata menjadi bentuk kata dasarnya (*stem*). *Stemming* bertujuan untuk menghilangkan imbuhan imbuhan baik itu berupa *prefiks*, *sufiks*, maupun *konfiks* yang ada pada setiap kata.

1.4 **Algoritma *Enhanced Confix Stripping Stemmer***

Algoritma *enhanced confix stripping stemmer* merupakan peningkatan performa dari algoritma *confix stripping stemmer* yang dikembangkan oleh (Mahendra, 2007), dan merupakan salah satu algoritma yang dapat mengatasi proses *stemming* yang spesifik untuk Bahasa Indonesia. Pada dasarnya, algoritma *enhanced confix stripping stemmer* merupakan modifikasi dari algoritma *confix stripping stemmer* (Asian, 2007) yang dikembangkan dari algoritma *stemming* yang dibuat oleh Nazief dan Adriani (1996) dengan beberapa penambahan aturan tertentu yang telah terbukti mampu meningkatkan kinerja *stemmer* tersebut.

Algoritma *stemming* Nazief dan Adriani dikembangkan berdasarkan pada aturan morfologi Bahasa Indonesia yang mengelompokkan dan mengenkapsulasi imbuhan-imbuhan, termasuk di dalamnya adalah awalan (*prefix*), sisipan(*infix*), akhiran (*suffix*) dan gabungan awalan-akhiran (*confixes*). Algoritma *Enhanced confix stripping stemmer* menambahkan penggunaan kamus kata dasar dan mendukung *recoding*, yakni penyusunan kembali kata-kata yang mengalami proses *stemming* berlebih.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Tabel 2. 1 Kombinasi Awalan Akhiran Yang Dilarang

Awalan (<i>prefix</i>)	Akhiran (<i>suffix</i>) yang tidak diperbolehkan
be-	-i
di-	-an
ke-	-i, -kan
me-	-an
se-	-i, -kan
te-	-an

Algoritma *enhanced confix stripping stemmer* (Mahendra, 2007) adalah sebagai berikut:

1. Cek kombinasi akhiran dan awalan yang dilarang, apabila bernilai benar maka lakukan penghilangan awalan terlebih dahulu. Apabila bernilai salah, maka penghilangan akhiran dilakukan terlebih dahulu.
2. Lakukan *recoding* apabila diperlukan.
3. Lakukan *loop* Pengembalian Akhiran.
4. Cek apakah terdapat tanda hubung ('-') yang menandakan bahwa input kata tersebut adalah kata ulang atau bentuk jamak. Jika ada, maka lakukan *stemming* pada sub-kata di sebelah kiri dan kanan tanda hubung tersebut. Apabila *stemming* memberikan hasil yang sama, maka kata dasar kata ulang tersebut adalah hasil *stemming* yang didapatkan.
5. Jika proses-proses di atas gagal, maka input kata yang di *stemming* dianggap sebagai kata dasar.

Pada setiap langkah, dilakukan proses pengecekan output *stemming* ke kamus. Apabila ditemukan, maka proses ini berhenti. Berikut adalah contoh proses *stemming* pada kata “perpolitikan” dengan menggunakan *enhanced confix stripping stemmer*:

1. Cek kombinasi awalan dan akhiran yang dilarang: salah. Hilangkan akhiran terlebih dahulu. Penghilangan akhiran menyisakan kata “perpoliti”.

2. Penghilangan awalan menyisakan kata “politi”.
3. Kata “politi” tidak ada di kamus, oleh karena itu dilakukan loop Pengembalian Akhiran:
 - a. Awalan-awalan yang telah dihilangkan, dikembalikan lagi. Langkah ini menghasilkan kata “perpoliti”.
 - b. Akhiran-akhirian dikembalikan. Karena akhiran yang sebelumnya dihilangkan adalah “-kan”, maka karakter ‘k’ saja yang dikembalikan terlebih dahulu. Proses ini menghasilkan kata “perpolitik”.
 - c. Karena “perpolitik” tidak ada di kamus, maka proses penghilangan awalan dilakukan. Proses ini menghasilkan kata “politik”.
 - d. Karena “politik” ditemukan dalam kamus, proses loop Pengembalian Akhiran ini berhenti. Kata dasar “perpolitikan” adalah “politik”.

1.5 Naïve Bayes Classifier

Algoritma *Naïve Bayes Classifier* merupakan algoritma yang digunakan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasi data uji pada kategori yang paling tepat (Feldman, R & Sanger, 2007). Dalam penelitian ini yang menjadi data uji adalah dokumen *tweet*. Ada dua tahap pada klasifikasi dokumen. Tahap pertama adalah pelatihan terhadap dokumen yang sudah diketahui kategorinya. Sedangkan tahap kedua adalah proses klasifikasi dokumen yang belum diketahui kategorinya.

Bayes merupakan teknik prediksi berbasis probabilistik sederhana yang berdasar pada penerapan teorema *Bayes* (atau aturan *Bayes*) dengan asumsi independensi (ketidaktergantungan) yang kuat (naif). Dengan kata lain, *Naïve Bayes*, model yang digunakan adalah “model fitur independen”. Dalam *Bayes* (terutama *Naïve Bayes*), maksud independensi yang kuat pada fitur adalah bahwa sebuah fitur pada sebuah data tidak berkaitan dengan ada atau tidaknya fitur lain

dalam data yang sama. Prediksi *Bayes* didasarkan pada teorema *Bayes* dengan formula umum sebagai berikut

$$P(H|E) = \frac{P(E|H) \times P(H)}{P(E)} \quad (2.1)$$

Penjelasan dari formula (2.1) tersebut adalah sebagai berikut:

Parameter	Keterangan
$P(H E)$	Probabilitas akhir bersyarat (<i>conditional probability</i>) suatu hipotesis H terjadi jika diberikan bukti (<i>evidence</i>) E terjadi.
$P(E H)$	Probabilitas sebuah bukti E terjadi akan memengaruhi hipotesis H.
$P(H)$	Probabilitas awal (priori) hipotesis H terjadi tanpa memandang bukti apapun.
$P(E)$	Probabilitas awal (priori) bukti E terjadi tanpa memandang hipotesis/bukti yang lain.

Ide dasar dari aturan *Bayes* adalah bahwa hasil dari hipotesis atau peristiwa (H) dapat diperkirakan berdasarkan pada beberapa bukti (E) yang diamati. Ada beberapa hal penting dari aturan *Bayes* tersebut, yaitu (Feldman, R & Sanger, 2007):

1. Sebuah probabilitas awal/prior H atau $P(H)$ adalah probabilitas dari suatu hipotesis sebelum bukti diamati.
2. Sebuah probabilitas akhir H atau $P(H|E)$ adalah probabilitas dari suatu hipotesis setelah bukti diamati.

1.5.1 Model Probabilistik *Naïve Bayes*

Model probabilitas untuk klasifikasi adalah model kondisional terhadap variabel kelas dependen C dengan sejumlah kecil hasil atau kelas, tergantung pada beberapa variabel fitur F_1 sampai F_N . Masalahnya adalah bahwa jika jumlah fitur n besar atau bila fitur bisa mengambil sejumlah besar nilai, maka membuat sebuah model pada tabel probabilitas adalah tidak mungkin. Oleh karena itu di reformulasi model untuk membuatnya lebih fleksibel.

$$p(C|F_1, \dots, F_N) \quad (2.2)$$

1.5.2 Naïve Bayes Untuk Klasifikasi Dokumen

Berikut ini adalah sebuah contoh dari metode *Naïve Bayes* untuk permasalahan klasifikasi dokumen. Masalah mengklasifikasikan dokumen adalah berdasarkan konten, misalnya *spam* dan *non-spam e-mail*. Bayangkan bahwa dokumen yang diambil dari beberapa kelas dokumen yang dapat dimodelkan sebagai set kata-kata dimana probabilitas (independen) bahwa kata ke-*i* dari suatu dokumen tertentu terjadi dalam dokumen dari kelas *C* dapat ditulis sebagai

$$p(w_i|C) \tag{2.3}$$

mengasumsikan bahwa kata-kata secara acak terdistribusi dalam dokumen, yaitu: kata-kata tidak tergantung pada panjang dokumen, posisi dalam dokumen, dengan hubungannya dengan kata lain, atau dokumen-konteks yang lain.

Dalam algoritma *Naïve Bayes Classifier* setiap dokumen direpresentasikan dengan pasangan atribut “ $x_1, x_2, x_3, \dots, x_n$ ” dimana x_1 adalah kata pertama, x_2 adalah kata kedua dan seterusnya. Sedangkan *V* adalah himpunan kategori *tweet*. Pada saat klasifikasi algoritma akan mencari probabilitas tertinggi dari semua kategori dokumen yang diujikan (V_{MAP}), dimana persamaannya adalah sebagai berikut :

$$V_{MAP} = \underset{V_j \in V}{\operatorname{argmax}} \frac{P(x_1, x_2 | x_3, \dots, x_n | V_j) P(V_j)}{P(x_1, x_2, x_3, \dots, x_n)} \tag{2.4}$$

Untuk $P(x_1, x_2, x_3, \dots, x_n)$ nilainya konstan untuk semua kategori (V_j) sehingga persamaan dapat ditulis sebagai berikut :

$$V_{MAP} = \underset{V_j \in V}{\operatorname{argmax}} P(x_1, x_2, x_3, \dots, x_n | V_j) P(V_j) \tag{2.5}$$

Persamaan diatas dapat disederhanakan menjadi sebagai berikut :

$$V_{MAP} = \underset{V_j \in V}{\operatorname{argmax}} \prod_{i=1}^n P(x_i | V_j) P(V_j) \tag{2.6}$$

Keterangan:

V_j = Kategori *tweet*, $j = 1, 2, 3, \dots, n$. Dimana dalam penelitian ini,

J_1 = Kategori *tweet racist*, J_2 = Kategori *tweet Embarrass*,

J_3 = Kategori *tweet Physically threatened*, J_4 =

Kategori *tweet Sexual Harassment*, dan

J_5 = Kategori *tweet Neutral*.

$P(X_i|V_j)$ = Peluang kemunculan X_i pada kategori V_j .

$P(V_j)$ = Peluang kemunculan dokumen yang memiliki kategori j .

Selanjutnya, melakukan perhitungan probabilitas setiap kelas j , dengan cara:

$$P(V_j) = \frac{|\text{docs}_j|}{|\text{contoh}|} \quad (2.7)$$

Keterangan:

$P(V_j)$ = Peluang kemunculan dokumen yang memiliki kategori j .

$|\text{docs}_j|$ = Jumlah dokumen setiap kategori j .

$|\text{contoh}|$ = Jumlah dokumen dari semua kategori.

Dan terakhir melakukan perhitungan probabilitas setiap kata kunci (kata selain kata kunci diabaikan) pada data *testing* terhadap data uji pada setiap kelas j , dengan cara:

$$P(X_i|V_j) = \frac{n_k+1}{n+|\text{kosakata}|} \quad (2.8)$$

Keterangan:

$P(X_i|V_j)$ = Peluang kemunculan X_i pada kategori V_j .

n_k = Jumlah frekuensi kemunculan setiap kata.

n = Jumlah frekuensi kemunculan kata dari setiap kategori.

$|\text{kosakata}|$ = Jumlah semua kata dari semua kategori.

1.5.3 Karakteristik *Naïve Bayes*

Klasifikasi dengan *Naïve Bayes* bekerja berdasarkan teori probabilitas yang memandang semua fitur dari data sebagai bukti dalam probabilitas. Hal ini memberikan karakteristik *Naïve Bayes* sebagai berikut:

1. Metode *Naïve Bayes* bekerja teguh (*robust*) terhadap data-data yang terisolasi yang biasanya merupakan data dengan karakteristik berbeda (*outliner*). *Naïve Bayes* juga bisa menangani nilai atribut yang salah dengan mengabaikan data latih selama proses pembangunan model dan prediksi.
2. Tangguh menghadapi atribut yang tidak relevan.
3. Atribut yang mempunyai korelasi bisa mendegradasi kinerja klasifikasi *Naïve Bayes* karena asumsi independensi atribut tersebut sudah tidak ada.

1.6 Confusion Matrix

Confusion Matrix adalah sebuah tabel yang menyatakan jumlah data uji yang benar diidentifikasi dan jumlah data uji yang salah diidentifikasi (Indriani, 2014). Contoh *Confusion Matrix* klasifikasi biner ditunjukkan pada Tabel 2.2.

Tabel 2. 2 Confusion Matrix

		Kelas Prediksi	
		1	0
Kelas Sebenarnya	1	TP	FN
	0	FP	TN

Perhitungan akurasi dinyatakan dalam persamaan berikut:

$$Akurasi = \frac{TP+TN}{TP+FN+FP+TN} \times 100\% \quad (2. 10)$$

Keterangan untuk Tabel 2.1 dinyatakan sebagai berikut (Indriani, 2014):

True Positive (TP), yaitu jumlah dokumen dari kelas 1 yang benar dan diklasifikasikan sebagai kelas 1.

True Negative (TN), yaitu jumlah dokumen dari kelas 0 yang benar diklasifikasikan sebagai kelas 0.

False Positive (FP), yaitu jumlah dokumen dari kelas 0 yang salah diklasifikasikan sebagai kelas 1.

False Negative (FN), yaitu jumlah dokumen dari kelas 1 yang salah diklasifikasikan sebagai kelas 0.

1.7 Pengujian White box

Pengujian *White Box* adalah pengujian untuk memperlihatkan cara kerja dari produk secara rinci sesuai dengan spesifikasinya. *White Box Testing* adalah salah satu cara untuk menguji suatu aplikasi atau *software* dengan cara melihat modul untuk dapat meneliti dan menganalisa kode dari program yang di buat ada yang salah atau tidak. Kalau modul yang telah dan sudah di hasilkan berupa *output* yang tidak sesuai dengan yang diharapkan maka akan dikompilasi ulang dan di cek

kembali kode-kode tersebut hingga sesuai dengan yang diharapkan.

Kasus yang sering menggunakan *white box testing* akan di uji dengan beberapa tahapan yaitu:

1. Pengujian seluruh keputusan yang menggunakan logikal.
2. Pengujian keseluruhan *loop* yang ada sesuai batasan-batasannya.
3. Pengujian pada struktur data yang sifatnya internal dan yang terjamin validitasnya.

Kelebihan *White Box Testing* antara lain :

1. Kesalahan Logika

Menggunakan sintax '*if*' dan *sintax* pengulangan. Langkah selanjutnya metode *white box testing* ini akan mencari dan mendeteksi segala kondisi yang di percaya tidak sesuai dan mencari kapan suatu proses perulangan di akhiri.

2. Ketidaksesuaian Asumsi

Menampilkan dan memonitor beberapa asumsi yang diyakini tidak sesuai dengan yang diharapkan atau yang akan diwujudkan, untuk selanjutnya akan dianalisa kembali dan kemudian diperbaiki.

3. Kesalahan Pengetikan

Mendeteksi dan mencaribahasa-bahasa pemograman yang di anggap bersifat *match case*.

Kelemahan *White Box Testing* adalah pada perangkat lunak yang jenisnya besar, metode *white box testing* ini dianggap boros karena melibatkan banyak sumber daya untuk melakukannya (Nidhra dan Dondeti, 2012).

1.8 Penelitian Terkait

Penelitian yang pernah dilakukan tentang klasifikasi akun pelecehan *online* dengan menggunakan atau menerapkan *Naïve Bayes*. Penelitian terkait dapat dilihat pada Tabel 2.6.

Tabel 2. 3 Penelitian Terkait

1. Penelitian Terkait Kasus

No	Peneliti, Judul	Metode	Tahun	Keterangan
1	(Dawei Yin , dkk) <i>Detection of Harassment on Web 2.0</i>	SVM	2009	Menggunakan metode SVM. Dilengkapi (TFIDF), sentimen dan atribut fitur kontekstual Sedangkan pada penelitian penulis menggunakan <i>dataset</i> berbahasa indonesia dan metode <i>Naïve Bayes</i> .
2	(A.Saravananaraj,J.I. Sheeba Assistant dan S. Pradeep Devaneyan) <i>Automatic Detection Of Cyberbullying From Twitter</i>	<i>Naïve Bayes</i>	2016	Menggunakan metode <i>stemming syntatyc</i> dan semantik dan pengklasifikasi dengan algoritma NVC dan <i>Random Forest Classifier</i> sedangkan penelitian penulis menggunakan Algoritma <i>Stemming Enhanced Confix Stripping</i> dan klasifikasi dengan metode NVC.

2. Penelitian Terkait Metode

No	Peneliti	Metode	Tahun	Keterangan
1	(Dhita) Perbandingan Klasifikasi Tugas Akhir Mahasiswa Jurusan Teknik Informatika	<i>Metode Naïve Bayes Classifier dan K-Nearest Neighbor</i>	2016	Perbandingan Klasifikasi Tugas Akhir Mahasiswa Jurusan Teknik Informatika Menggunakan Metode <i>Naïve Bayes Classifier</i> dan <i>K-Nearest Neighbor</i> , didapatkan hasil dengan metode NBC memberikan hasil dengan akurasi yang lebih baik hingga 87%. Sedangkan pengujian pada metode <i>K-Nearest Neighbor</i> menghasilkan nilai akurasi 84%
2	(Bambang Saputro),	<i>Naïve Bayes Classifier</i>	2016	Klasifikasi Dan Pemetaan Posdaya Tematik Berbasis Masjid. Ada tiga

<p>Hak Cipta Dilindungi Undang-Undang</p>	<p>Klasifikasi Dan Pemetaan Posdaya Tematik Berbasis Masjid</p>			<p>klasifikasi yaitu baik, cukup baik dan kurang baik. Sedangkan penelitian penulis menggunakan kasus jenis pelecehan <i>online</i> kategori dengan lima klasifikasi <i>tweet</i>.</p>
<p>3</p>	<p>(Rodiansyah dan Winarko) Klasifikasi <i>Posting</i> Twitter Kemacetan Lalu Lintas Kota Bandung</p>	<p><i>Naïve Bayes Classifier</i></p>	<p>2016</p>	<p>Dari hasil uji coba aplikasi, menghasilkan akurasi tertinggi 93,58% pada pengujian deta sampel sebanyak 13106. dengan data <i>tweet</i> yang mengandung informasi lalu lintas di kota bandung dengan hasil klasifikasi ditampilkan dalam visualisai di peta kota Bandung.</p>

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.