ARTIKEL

PEMODELAN DETEKSI *CYBER BULLYING* PADA JEJARING SOSIAL TWITTER

Oleh:

IKA YUNIDA ANGGRAINI

14.1.03.03.0101

Dibimbing oleh :

1. **SUCIPTO, M.Kom**
2. **RINI INDRIATI, M.Kom**

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS TEKNIK

UNIVERSITAS NUSANTARA PGRI KEDIRI

TAHUN 2018

SURAT PERNYATAAN
ARTIKEL SKRIPSI TAHUN 2018

Yang bertanda tangan di bawah ini:

Nama Lengkap	: Ika Yunida Anggraini
NPM	: 14.1.03.03.0101
Telepun/HP	: 087752955029
Alamat Surel (Email)	: ikayunida123@gmail.com
Judul Artikel	: Pemodelan Deteksi *Cyber Bullying* pada Jejaring Sosial Twitter
Fakultas – Program Studi	: Teknik – Sistem Informasi
Nama Perguruan Tinggi	: Universitas Nusantara PGRI Kediri
Alamat PerguruanTinggi	: JL. K.H. Achmad Dahlan No. 76 Kediri

Dengan ini menyatakan bahwa :

a. artikel yang saya tulis merupakan karya saya pribadi (bersama tim penulis) dan bebas plagiarisme;
b. artikel telah diteliti dan disetujui untuk diterbitkan oleh Dosen Pembimbing I dan II.

Demikian surat pernyataan ini saya buat dengan sesungguhnya. Apabila dikemudian hari ditemukan ketidaksesuaian data dengan pernyataan ini dan atau ada tuntutan dari pihak lain, saya bersedia bertanggung jawab dan diproses sesuai dengan ketentuan yang berlaku.

Mengetahui		Kediri, 22 Juli 2018
Pembimbing I	Pembimbing II	Penulis,
SUCIPTO, M.Kom NIDN. 0721029101	RINI INDRIATI, M.Kom NIDN. 0725057003	IKA YUNIDA ANGGRAINI NPM. 14.1.03.03.0101

PEMODELAN DETEKSI *CYBER BULLYING* PADA JEJARING SOSIAL TWITTER

Ika Yunida Anggraini
14.1.03.03.0101
Fakultas Teknik – Sistem Informasi
ikayunida123@gmail.com
SUCIPTO, M.Kom dan RINI INDRIATI, M.Kom
UNIVERSITAS NUSANTARA PGRI KEDIRI

ABSTRAK

Tindakan kejahatan siber banyak terjadi dalam jejaring sosial. Salah satunya adalah perundungan siber atau *cyber bullying* yang kini marak terjadi terjadi pada situs jejaring sosial Twitter. Dikalangan remaja, tindakan *cyber bullying* dapat mengakibatkan depresi, percobaan pembunuhan bahkan bunuh diri sehingga diperlukan sebuah deteksi terhadap tindak perundungan siber agar tindakan tersebut tidak menimbulkan bahaya bagi korbannya. Untuk mencegah tindakan *cyber bullying* tersebut dapat dilakukan sebuah pemodelan *text mining* dengan mengklasifikasikan *tweets* pada jejaring sosial Twitter ke dalam kelas *bullying* dan kelas bukan *bullying*. Pemodelan dalam penelitian ini dilakukan menggunakan *Naïve Bayes Classifier* dengan lima tahap *preprocessing* yaitu *replace tokens, transform case, tokenization, filter stopwords* dan *n-grams*. Untuk memvalidasi model dilakukan proses *10-Fold Validation* dan untuk menguji kinerja model pada data *real* maka dilakukan proses *testing*. Pemodelan tersebut pada proses validasi menggunakan *10-Fold Validation* memiliki hasil *accuracy* sebesar 82,50% dan *standard deviation* +/-5,12%. Sementara pada proses *testing model* bekerja dengan baik dengan *accuracy* sebesar 90%.

KATA KUNCI : Pemodelan, *Cyber Bullying*, Twitter.

I. LATAR BELAKANG

Perkembangan *website* yang sangat pesat menyebabkan pengguna situs jejaring sosial seperti Facebook, Twitter, Instagram dan Youtube meningkat dari tahun ke tahun. Berdasarkan data dari Kementrian Komunikasi dan Informatika Republik Indonesia (Kominfo), pengguna situs jejaring sosial seperti Twitter hampir mencapai 20 juta pengguna aktif dan menempati urutan 5 besar pengguna Twitter terbanyak di dunia.

Generasi millenial banyak menghabiskan waktu mereka di dalam situs jejaring sosial dan seringkali menyebarkan informasi pribadi mereka dengan teman bahkan tak jarang membuat informasi tersebut dapat dilihat secara publik. Hal ini menyebabkan banyak terjadi kejahatan di situs jejaring sosial. Salah satu jenis kejahatan di sosial media yang sering terjadi adalah *cyber bullying* (Sanchez dan Kumar, 2011). Menurut data dari UNICEF (*United Nations Children's Fund*), di Indonesia sendiri pada tahun

2016 sebanyak 41-50 persen remaja dalam rentang usia 13-15 tahun pernah mengalami tindakan *cyber bullying*.

Cyber bullying atau perundungan siber adalah tindakan penyerangan, penghinaan, atau menyakiti orang lain secara sengaja dan berulang-ulang pada sosial media, pesan, atau dengan cara lainnya (Patchin dan Hinduja, 2014). *Cyber bullying* menjadi kekhawatiran publik karena banyak kasus *cyber bullying* sering dikaitkan dengan tindakan bunuh diri. Salah satu organisasi non-profit, Cyber Bullying Research Center mengungkapkan bahwa kebiasaan *bullying* baik secara tradisional maupun *cyber* dikalangan remaja dapat mengakibatkan depresi, tindakan bunuh diri dan percobaan pembunuhan. Dengan bahayanya efek *cyber bullying* tersebut, maka diperlukan sebuah tindakan pencegahan atau deteksi agar tidak menimbulkan bahaya bagi korbannya (Patchin dan Hinduja, 2010).

Untuk mendeteksi tindak *cyber bullying* di Twitter maka dilakukan *text mining* pada *tweets* dengan menggunakan *naïve bayes classifier*. *Naïve bayes classifier* merupakan sebuah algoritma klasifikasi yang dikembangkan oleh Thomas bayes pada awal abad ke-18 dengan menggunakan teorema bayes. Algoritma tersebut bekerja dengan memprediksikan probabilitas keanggotaan suatu *value* dalam kelas-kelas yang telah ditentukan (Han dan Kamber, 2006).

Dalam mengklasifikasikan data pada deteksi *cyber bullying* di situs jejaring sosial Twitter, dalam penelitian ini terdapat dua kelas yaitu kelas "*Bullying*" dan kelas "Bukan *Bullying*". Kelas "*Bullying*" merupakan kelas dari *tweets* yang berisi tindakan *bullying*, sementara kelas "Bukan *Bullying*" merupakan kelas dari *tweets* yang bukan merupakan tindakan *bullyin*g.

II. METODE

1. Text Mining

Text mining yang dikenal juga sebagai *text data mining* merupakan proses penggalian informasi yang berguna dari sekumpulan dokumen dengan menggunakan *tools* analisis yang merupakan komponen dalam *data mining* (Feldman dan Sanger, 2017).

2. Naive Bayes Classifier

Naïve Bayes Classifier merupakan algoritma *data mining* yang menggunakan klasifikator statistik. Algoritma ini dapat memprediksikan probabilitas keanggotaan dalam sebuah kelas.

Klasifikasi bayes menerapkan teorema bayes. Teorema bayes ditemukan oleh Thomas bayes pada abad ke-18. Teorema bayes dirumuskan sebagai berikut :

$$\mathbf{P(H|X)} = \frac{\mathbf{P(X|H)P(H)}}{\mathbf{P(X)}}$$

Di mana X adalah data dengan kelas yang belum diketahui. H adalah hipotesis kelas dari data X. P(H|X) adalah probabilitas H berdasarkan kondisi X (probabilitas posterior). P(H) adalah probabilitas H (probabilitas prior). P(X|H) adalah probabilitas X berdasarkan kondisi H. Dan P(X) adalah probabilitas X.

Dalam penelitian ini pendekatan *Naïve Bayes* yang digunakan adalah Gaussian *Naïve Bayes Classifier* yang menggunakan distribusi Gaussian atau distribusi normal. Ukuran probabilitas dalam distribusi normal adalah :

$$\boldsymbol{f(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}}$$

Di mana μ adalah mean dari distribusi, σ adalah *standard deviation* dan σ^2adalah varian (Han dan Kamber, 2006).

3. Confusion Matrix

Dalam model klasifikasi, evaluasi performa dihitung berdasarkan banyaknya objek *testing* yang diprediksikan secara benar dan salah. Salah satu metode evaluasi dalam klasifikasi adalah *confusion matrix*. *Confusion matrix* atau yang disebut juga *error matrix* merupakan tabel yang menggambarkan performa dari sebuah algoritma. Setiap kolom dari *matrix* merepresentasikan *predicted class* dan *actual class*. Evaluasi menggunakan *confusion matrix* pada penelitian ini menggunakan *precision, recall dan accuracy* (Gorunescu, 2011)

4. Preprocessing

Sebelum data dimasukkan ke dalam proses *mining*, terlebih dahulu dilakukan *preprocessing*. *Preprocessing* adalah tahapan di mana data disiapkan untuk proses *mining* dan disesuaikan dengan lingkungan *data mining*. *Preprocessing* penting untuk dilakukan karena data pada dunia nyata cenderung tidak komplit, *noisy* dan tidak konsisten (Han dan Kamber, 2006).

Dalam penelitian ini, ada beberapa teknik *preprocessing* yang digunakan, yaitu *replace tokens, transform case, tokenization, filter stopwords* dan *n-grams*.

5. K-Fold Validation

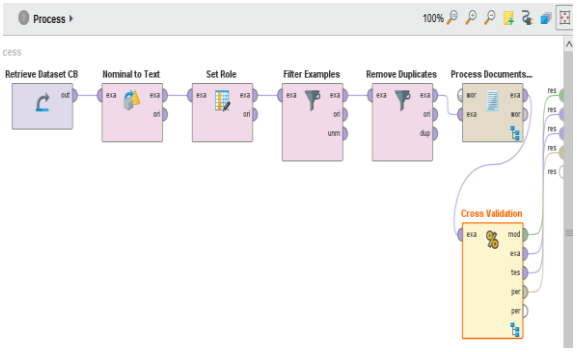
Pada *cross validation* atau *k-fold validation*, data secara acak dipartisi menjadi *k subsets* atau *"folds"*, D1, D2, . . . , Dk, yang berukuran sama. Keseluruhan *folds* tersebut kemudian akan melalui proses *training* dan *testing* yang dilakukan sebanyak *k*-kali. Dalam setiap iterasi *i*, partisi D_i digunakan sebagai *test set*, dan sisanya akan digunakan untuk me-*train model*. Secara umum, *10-fold cross validation* direkomendasikan untuk mengestimasi *accuracy* karena relatif berbias rendah dan beragam (Han dan Kamber, 2006).

III. HASIL DAN KESIMPULAN

1. Proses Validasi Pemodelan

Proses validasi merupakan proses untuk penilaian terhadap sebuah model. Dalam penelitian ini proses validasi dilakukan menggunakan operator *Cross Validation (K-Fold Validation)*.

Cross Validation yang digunakan dalam penelitian ini adalah *10-Fold Validation*. Dalam *10-Fold Validation*, data dibagi menjadi 10 *fold* berukuran sama dan dalam setiap *fold* akan dilakukan pengujian dengan 9 *fold* sebagai *training dataset* dan 1 *fold* sebagai *validation dataset*. Proses *cross validation* pada RapidMiner seperti pada gambar berikut :



Gambar 1. Proses *Cross Validation*

Dalam proses *cross validation* terdapat dua sub-proses, yaitu sub-proses *training* dan sub-proses *testing*. Sub-proses *training* digunakan untuk melakukan proses *training* model menggunakan operator *Naïve Bayes*. *Training model* tersebut kemudian diaplikasikan pada sub-proses *testing* menggunakan operator *Apply Model*. Selanjutnya performa dari model tersebut dinilai dengan operator *Performance (Binominal Classification)* pada sub-proses *testing*.

Pada sub-proses *training* dalam *10-Fold Validation* digunakan operator *Naïve Bayes*. Selain itu pada pilihan paramater digunakan parameter *Laplace's correction*. Hal tersebut dilakukan untuk menghindari probabilitas nol pada nilai atribut yang belum pernah muncul sebelumnya.

Gambar 2 menunjukkan hasil keseluruhan performa dari pemodelan pada penelitian ini menggunakan *10-Fold Validation* pada *software* RapidMiner.

accuracy: 82.50% +/- 5.12% (mikro: 82.50%)

	true Bullying	true Bukan Bullying	class precision
pred. Bullying	99	29	77.34%
pred. Bukan Bullying	6	66	91.67%
class recall	94.29%	69.47%	

Gambar 2. Performa *10-Fold Validation*

Precision yang menunjukkan tingkat ketepatan informasi yang diminta dan jawaban yang diberikan model dari hasil pemodelan adalah 77,34% untuk kelas *Bullying* dan 91,67% untuk kelas Bukan *Bullying*. Sedangkan *recall* yang menunjukkan model mempunyai keberhasilan cukup baik dalam menemukan kembali sebuah informasi dari hasil pemodelan di atas adalah 94,29% untuk kelas *Bullying* dan 69,47% untuk kelas Bukan *Bullying*.

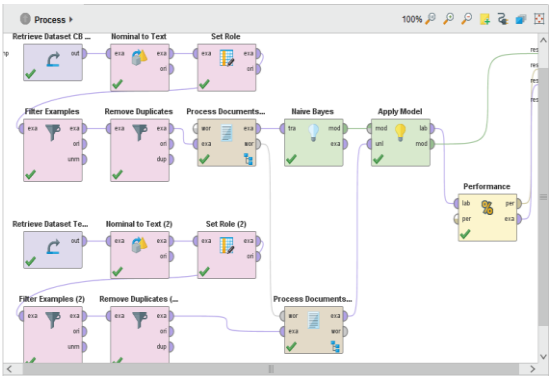
Accuracy keseluruhan yang menunjukkan tingkat kedekatan antara nilai prediksi dengan hasil aktual dari semua *fold* pada *10-Fold Validation* adalah 82,50%, yang menunjukkan persentase semua prediksi yang benar dari keseluruhan data. Sementara nilai deviasi standar dari kesepuluh *accurcy* dalam *10-Fold Validation* adalah +/-5,12. Nilai deviasi standar yang rendah menunjukkan stabilitas sebuah model. Sebuah model dinilai stabil karena memiliki nilai *accuracy* yang cukup baik dan memiliki *range* kecil dari kasus terbaik dan terburuk yang menunjukkan kualitas dari prediksi.

Setelah melihat hasil *cross validation* yang dirasa cukup baik, selanjutnya dilakukan proses *testing* untuk melihat performa pemodelan pada *unseen data* atau data yang belum pernah dilihat sebelumnya.

2. Proses *Testing*

Proses *testing* merupakan proses yang dilakukan untuk melihat performa pemodelan terhadap *unseen real-world data*. Proses *testing* juga dilakukan untuk mengevaluasi performa dari *classifier* yang telah di-*training*.

Dalam penelitian ini 200 *row data training* yang telah di-*training* pada proses *cross validation* serta *unseen dataset* sebanyak 60 *row* digunakan sebagai *data testing*. *Data testing* tersebut kemudian melalui proses *cleaning* serta *preprocessing* yang sama dengan *data training*. Selanjutnya *data training* yang telah dimodel akan diuji dengan *data testing* untuk mengetahui performanya. Proses *testing* dalam penelitian ini pada RapidMiner seperti pada gambar berikut.



Gambar 3. Proses *Testing* dalam RapidMiner

Proses *testing* tersebut kemudian akan menghasilkan *output* performa pemodelan dalam *confusion matrix* seperti pada gambar 4. Dari output pada gambar 5, diketahui bahwa *precision* dari kelas *Bullying* adalah 92,86% dan *precision* dari kelas Bukan *Bullying* adalah 83,33%. Selanjutnya *recall* dari kelas *Bullying* adalah 92,86% dan *recall* pada kelas Bukan *Bullying* adalah 83,33. Kemudian *accuracy* pada keseluruhan proses *testing* adalah 90%.

accuracy: 90.00%

	true Bullying	true Bukan Bullying	class precision
pred. Bullying	39	3	92.86%
pred. Bukan Bullying	3	15	83.33%
class recall	92.86%	83.33%	

Gambar 4. Hasil Akurasi Proses *Testing*

3. Kesimpulan

Berdasarkan hasil penelitian yang dilakukan peneliti, maka dapat ditarik kesimpulan bahwa deteksi *cyber bullying* pada jejaring sosial Twitter dapat dilakukan dengan beberapa teknik. Pertama, data dikumpulkan melalui proses *data crawling*. Kedua, dilakukan proses seleksi data, *cleaning* data dan *preprocessing* untuk mempersiapakan data pada proses *mining*. Ketiga, dilakukan klasifikasi menggunakan *Naïve Bayes Classifier*. Hasil dari proses pemodelan deteksi *cyber bullying* pada jejaring sosial Twitter tersebut mempunyai *accuracy* sebesar 82,50% dan *standard deviation* sebesar +/-5,12. Sementara pada proses *testing* terhadap data *real*, pemodelan *cyber bullying* bekerja dengan baik dengan *accuracy* yang dihasilkan model tersebut sebesar 90%.

IV. DAFTAR PUSTAKA

Feldman, R., & Sanger, J. (2017). *The Text Mining Handbook : Advanced Approaches in Analyzing Unstructured Data*. New York, USA : Cambridge University Press.

Gorunescu, F. (2011). *Data Mining : Concept, Models and Techniques*. Berlin : Springer.

Han, J., & Kamber, M. (2006). *Data Mining : Concept and Techniques*. San Francisco : Morgan Kaufmann Publishers.

Patchin, J. W., & Hinduja, S. (2010). *Cyberbullying Research Summary*. USA : The Cyberbullying Research Center

Patchin, J. W., & Hinduja, S. (2014). *Words Wound : Delete Cyberbullying and Make Kindness Go Viral*. Minneapolis, USA : Free Spirit Publishing Inc.

Kasus Bullying Anak Meningkat Pada Tahun 2017. (2017). Dipetik Desember 17, 2017, dari VIVA.co.id:https://www.viva.co.id/berita/nasional/938446/kasus-bullying-anak-meningkat pada-2017

Kominfo : Pengguna Internet di Indonesia 63 Juta Orang. (2017) Dipetik Desember 17, 2017, dari Kementrian Komunikasi dan Informatika: https://www.kominfo.go.id/content/detail/3415/kominfo-pengguna -internet-di-indonesia-63-juta -orang/0/berita_satker