

## **BAB II**

### **LANDASAN TEORI**

#### **2.1 Tinjauan Teori**

##### **2.1.1 Media Sosial**

Media sosial merupakan istilah yang sering digunakan untuk merujuk sebuah bentuk dari media yang baru yang dimana di dalamnya terdapat partisipasi berupa interaksi antara individu yang satu dengan individu yang lainnya [8]. Media sosial sendiri memiliki fungsi untuk : membangun profil dan informasi pribadi yang terkait sehingga dapat terkoneksi dengan individu lainnya, membagikan atau menikmati konten baik dalam bentuk foto, video, dan teks, media untuk berinteraksi sehingga dapat menciptakan jaringan sosial virtual yang melahirkan kehidupan sosial baru. Dengan adanya media sosial seperti Twitter dan Instagram, suatu individu dapat melakukan media sharing, menyampaikan informasi, dan membangun jejaring sosial dengan individu lainnya.

##### **2.1.2 Kekerasan Seksual**

Kekerasan seksual adalah setiap tindakan baik berupa ucapan ataupun perbuatan yang dilakukan seseorang untuk menguasai atau memanipulasi orang lain serta membuatnya terlibat dalam aktivitas seksual yang tidak dikehendaki [9]. Terdapat beberapa aspek yang menentukan terjadinya kekerasan seksual, yaitu berupa: aspek pemaksaan dan aspek tidak adanya persetujuan dari korban, dan aspek berupa ketidakmampuan korban untuk memberikan persetujuan. Menurut Komnas Perempuan, terdapat 15 bentuk daripada kekerasan seksual yang meliputi : perkosaan, intimidasi seksual, pelecehan seksual, eksploitasi seksual, perdagangan perempuan untuk tujuan seksual, prostitusi paksa, perbudakan seksual, pemaksaan perkawinan, pemaksaan kehamilan,

pemaksaan aborsi, pemaksaan kontrasepsi dan sterilisasi, penyiksaan seksual, penghukuman tidak manusiawi dan bernuansa seksual, praktek tradisi bernuansa seksual yang membahayakan atau mendiskriminasi perempuan, dan kontrol seksual [10].

### **2.1.3 Kekerasan Berbasis Gender Online**

Kekerasan Berbasis Gender Online (KBGO) merupakan pelecehan yang dilakukan dengan menggunakan teknologi digital atau menggunakan ranah online yang dimana didasarkan oleh relasi gender antara korban dan pelaku [11]. Menurut Southeast Asia Freedom of Expression Network, terdapat 8 bentuk kekerasan gender berbasis online, yaitu : pendekatan untuk memperdaya (*cyber grooming*), pelecehan online (*cyber harassment*), peretasan (*hacking*), konten ilegal (*illegal content*), pelanggaran privasi (*infringement of privacy*), ancaman distribusi foto/video pribadi (*malicious distribution*), pencemaran nama baik (*online defamation*), rekrutmen online (*online recruitment*) [12].

### **2.1.4 Kekerasan Seksual Verbal**

Kekerasan seksual verbal merupakan pelecehan seksual yang didasarkan pada kata yang dilontarkan menggunakan nada sindiran, menarik hati, dan menunjuk pada perilaku seksual seseorang di depan umum atau langsung dengan tujuan untuk memermalukan, menghina, dan mengintimidasi seseorang [13]. Bentuk pelecehan seksual verbal sering terjadi di media sosial dikarenakan pelecehan seksual verbal sering dibungkus dengan bentuk candaan yang dimana menyinggung suatu topik terhadap seseorang yang ingin diintimidasi.

### **2.1.5 Knowledge Discovery in Databases**

*Knowledge Discovery in Databases* (KDD) merupakan sebuah metode yang dikembangkan untuk memberikan panduan untuk melakukan data mining

dalam melakukan ekstraksi data yang kemudian dapat diolah menjadi sebuah pengetahuan [14]. Berikut ini merupakan tahapan dari metode KDD:

#### 1. Understanding Goal

*Understanding Goal* adalah tahapan yang digunakan untuk menentukan tujuan dan masalah yang akan diselesaikan

#### 2. Selection

Selection merupakan tahapan untuk menentukan dataset yang akan digunakan

#### 3. Data Cleaning and Preprocessing

*Data Cleaning and Preprocessing* adalah proses untuk melakukan pembersihan terhadap data. Proses pembersihan tersebut akan terdiri dari menghilangkan *outlier*, menangani data yang hilang, dan menghilangkan *noise* yang terdapat pada data

#### 4. Transformation

*Transformation* adalah proses untuk menemukan fitur yang dapat mewakili data. Proses tersebut terdiri dari reduksi dan transformasi pada data

#### 5. Data Mining Task

*Data Mining Task* adalah tahapan untuk menentukan metode yang akan digunakan dalam proses *Data Mining*

#### 6. Data Mining Algorithm Selection

*Data Mining Algorithm Selection* adalah tahapan untuk menentukan algoritma yang akan dipakai untuk melakukan *Data Mining*

#### 7. Data Mining

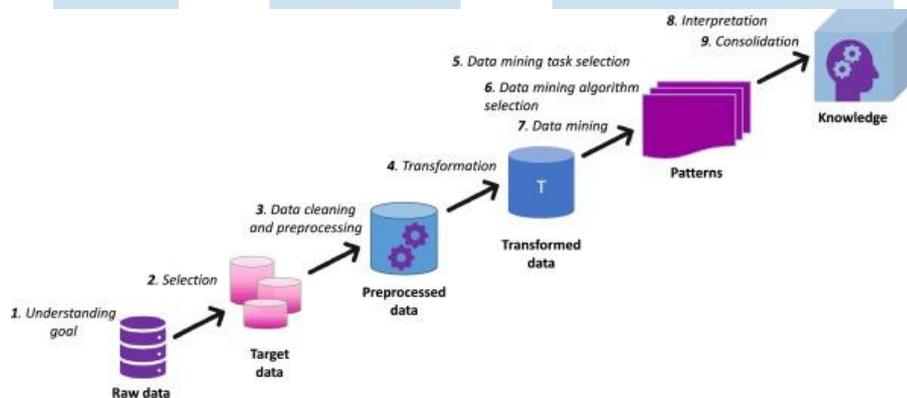
*Data Mining* adalah sebuah proses untuk mencari sebuah pola yang didasarkan pada metode yang telah ditentukan

## 8. Interpretation

*Interpretation* adalah tahapan untuk membuat visualisasi dari permodelan yang dibuat. Visualisasi yang dibuat akan digunakan untuk memperoleh pengetahuan

## 9. Consolidation

*Consolidation* adalah tahapan untuk melakukan dokumentasi terhadap pengetahuan yang telah ditemukan



**Gambar 2.1. Arsitektur Knowledge Discovery in Databases [14]**

### 2.1.6 Text Preprocessing

Text Preprocessing dapat diartikan sebagai sebuah metode untuk membersihkan data yang berisi noise, emoji dan tanda baca yang kemudian akan dimasukkan ke dalam permodelan machine learning [15]. Berikut ini merupakan tahapan umum dari proses Text Preprocessing, yaitu :

#### 1. Sentence Segmentation

*Sentence Segmentation* adalah proses pemisahan suatu kalimat dari suatu paragraph

#### 2. Tokenization

*Tokenization* adalah tahapan pemisahan teks menjadi potongan-potongan kecil yang disebut sebagai token

### 3. Filtering

*Filtering* adalah proses pengambilan kata-kata penting dari hasil token. Proses pengambilan kata tersebut akan disesuaikan dengan kondisi yang diinginkan

### 4. Stemming and lemmatization

*Stemming and lemmatization* merupakan proses normalisasi kata yang sama akan tetapi berbeda imbuhan. *Stemming* merupakan proses menghilangkan imbuhan pada awal atau akhir dari suatu kata sehingga kembali ke bentuk dasarnya, sedangkan *lemmatization* merupakan proses perubahan kata menjadi bentuk dasar dengan mempertimbangkan konteks dari suatu kalimat

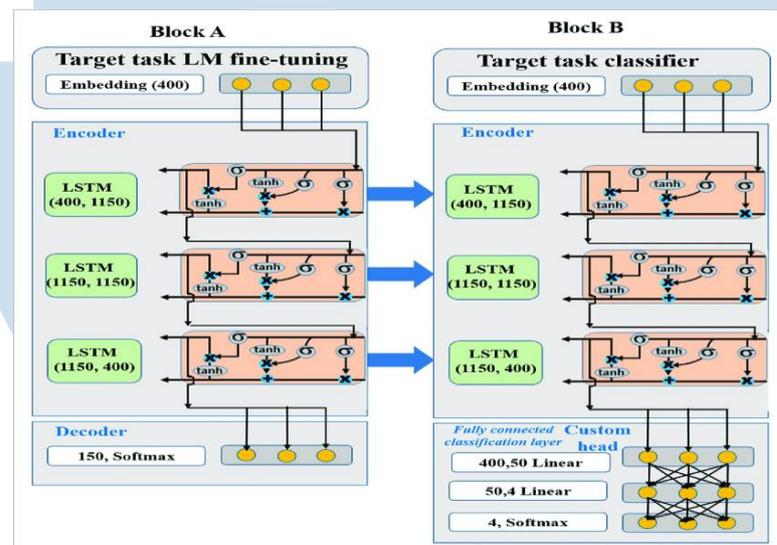
### 5. Word Embedding / Text Vector

*Word Embedding / Text Vector* merupakan proses perubahan kata menjadi suatu *vector*. *Vector* tersebut akan merepresentasikan syntax dan makna dari suatu kata.

#### **2.1.7 Average Stochastic Gradient Descent Weight-Dropped LSTM (AWD-LSTM)**

*Average Stochastic Gradient Descent Weight-Dropped LSTM (AWD-LSTM)* merupakan sebuah arsitektur yang berasal dari modifikasi algoritma *Long Short-Term Memory (LSTM)* yang menambahkan fitur berupa *DropConnect*, *Average Random Gradient Descent Method*, dan beberapa regulasi ke dalam algoritma *LSTM*. Modifikasi tersebut dilakukan karena kondisi loop dari permodelan *LSTM* yang dimana seringkali menyebabkan permodelan yang dibuat menjadi *overfitting*. Oleh karena hal tersebut, ditambahkan sebuah metode berupa *WeightDropout* yang juga dapat disebut sebagai *DropConnect*. Metode *DropConnect* tersebut akan diterapkan ke dalam matriks yang berada pada *hidden states* yang akan menyebabkan *Traditional Dropout* menjadi nol untuk tiap subset aktivasi yang dipilih secara acak di setiap

lapisan dari permodelan yang akan dibuat. Kombinasi dari *DropConnect*, *Average Random Gradient Descent Method*, dan beberapa regulasi yang ditambahkan ke permodelan akan membuang sebagian informasi dari *weight matrix* yang berada di *hidden states* yang dimana akan mengurangi masalah *overfitting* yang biasanya dihasilkan oleh algoritma *LSTM* tanpa modifikasi [16].



**Gambar 2.2. Arsitektur AWD-LSTM [17]**

### 2.1.8 Flask

*Flask* merupakan sebuah *micro framework* yang dapat digunakan untuk mengembangkan sebuah website dengan basis bahasa pemrograman *Python*. *Flask* dapat dideskripsikan sebagai sebuah *micro framework* karena tidak memerlukan sebuah *tools* atau *library* untuk menjalankannya. Dalam penerapannya, *Flask* melakukan integrasi terhadap *micro services* lainnya, seperti *Werkzeug* dan *Jinja*. *Werkzeug* merupakan sebuah alat yang menyediakan *Web Server Gateway Interface* (WSGI) yang dimana memungkinkan program yang berbasis *Python* untuk dapat berinteraksi dengan web server, sedangkan *Jinja* merupakan sebuah *template* yang dapat digunakan untuk memungkinkan interaksi antara program berbasis *Python* dengan halaman web [18]

## 2.2 Penelitian Terdahulu

Berikut merupakan tabel perbandingan antara penelitian terdahulu yang berkaitan dengan kekerasan seksual dan metode yang akan dilakukan dalam penelitian.

**Tabel 2.1. Penelitian Terdahulu**

Author	Judul Artikel	Nama Jurnal	Metodologi	Hasil
Samuel Bruno da S. Sousa, Evangelos Milios, Norbert Zeh, dan Lilian Berton (2019)	Categorizing Online Harassment on Twitter	ECML PKDD 2019: Machine Learning and Knowledge Discovery in Databases	<ol style="list-style-type: none"> <li>1. Menggunakan TF-IDF Vectorizer</li> <li>2. Menggunakan Word2Vec untuk mencari asosiasi kata</li> <li>3. Menggunakan algoritma klasifikasi berupa : Logistic Regression, Gaussian Naïve Bayes, Decision Trees, Random Forest, Linear SVM, Gaussian SVM, Polynomial SVM, Multi-Layer Perceptron, dan AdaBoost</li> </ol>	Permodelan yang paling baik dihasilkan oleh algoritma Linear SVM, TF-IDF Vectorizer dan Word2Vec yang dimana menghasilkan akurasi sebesar 92%. Hasil tersebut dapat terjadi karena metode Linear SVM dapat menemukan hyperplane yang dapat membantu pemisahan kelas yang dipakai untuk pengkategorian online harassment
Mohammad reza Rezvan, Saeedeh Shekarpour, Krishnaprasad Thirunarayanan, Valerie L. Shalin, dan	Analyzing and learning the language for different types of harassment	-	<ol style="list-style-type: none"> <li>1. Menggunakan TF-IDF Vectorizer</li> <li>2. Menggunakan Word2Vec untuk mencari asosiasi kata</li> <li>3. Menggunakan algoritma</li> </ol>	Permodelan yang paling baik dihasilkan oleh algoritma Gradient Boosting Machine yang dimana menghasilkan nilai precision,

Author	Judul Artikel	Nama Jurnal	Metodologi	Hasil
Amit Sheth (2018)			klasifikasi berupa : Support Vector Machine, K-Nearest Neighbors, Gradient Boosting Machine, Naive Bayes	recall, dan f1-score sebesar : 66%, 69%, dan 68%
Ivan, Yuita Arum Sari, Putra Pandu Adikara (2019)	Klasifikasi Hate Speech Berbahasa Indonesia di Twitter Menggunakan Naive Bayes dan Seleksi Fitur Information Gain dengan Normalisasi Kata	-	1. Melakukan normalisasi pada kata 2. Menggunakan seleksi fitur berupa Information Gain 2. Menggunakan Naive Bayes untuk mengklasifikasi Hate Speech	Permodelan yang paling baik dihasilkan oleh metode Naive Bayes, seleksi fitur Information Gain dengan normalisasi kata yang dimana menghasilkan akurasi sebesar 98%. Hasil tersebut merupakan hasil yang lebih baik apabila dibandingkan dengan melakukan permodelan dengan menggunakan metode Naive Bayes yang dimana tidak menggunakan fitur Information Gain atau normalisasi kata.
Arijit Ghosh Chowdhury, Ramit Sawhney, Rajiv Ratn Shah,	YouToo? Detection of Personal Recollections of Sexual Harassment	Proceedings of the 57th Annual Meeting of the Association for	Menggunakan metode klasifikasi berupa AWD-LSTM yang dibandingkan	Metode klasifikasi AWD-LSTM merupakan metode yang paling baik

<b>Author</b>	<b>Judul Artikel</b>	<b>Nama Jurnal</b>	<b>Metodologi</b>	<b>Hasil</b>
Debanjan Mahata (2019)	on Social Media	Computational Linguistics	dengan metode CNN, LSTM, dan RNN	apabila dibandingkan dengan metode yang lain dikarenakan nilai recall yang tinggi dan memiliki kemampuan untuk menangkap ketergantungan kata yang lebih kompleks. Metode ini menghasilkan nilai akurasi precision, recall, dan f1-score sebesar : 96%, 95%, 97%, dan 96%
Margarita Bugueno dan Marcelo Mendoza (2020)	Learning to detect online harassment on Twitter with the transformer	-	Menggunakan algoritma deep learning yang bernama BERT dalam mengkategorikan online harassment	Menghasilkan macro-averaged F1-score yang memiliki nilai sebesar 48,1%. Hal tersebut merupakan hasil yang cukup baik apabila dibandingkan dengan penelitian sebelumnya
Kholiq Budiman, Nurul Zaatsiyah, Ulfatun Niswah, Fikry Muhanna Nur Faizi (2020)	Analysis of Sexual Harassment Tweet Sentiment on Twitter in Indonesia using Naïve Bayes Method through National	-	1. Melakukan normalisasi pada kata 2. Menggunakan algoritma Naïve Bayes untuk mengklasifikasi sexual harassment antar dokumen	Permodelan Naïve Bayes menghasilkan nilai akurasi, precision, recall, dan F1-Score sebesar : 83%, 57%, 25%, dan 35%

Author	Judul Artikel	Nama Jurnal	Metodologi	Hasil
	Institute of Standard and Technology Digital Forensic Acquisition Approach			
Rezi Iwardani Saputri (2021)	Perbandingan Metode Naïve Bayes Classifier dan Support Vector Machine untuk Klasifikasi Cyber Harassment pada Twitter	-	<ol style="list-style-type: none"> <li>1. Menggunakan feature selection</li> <li>2. Menggunakan TF-IDF</li> <li>3. Menggunakan Naïve Bayes Classifier dan Support Vector Machine</li> </ol>	Hasil kinerja masing – masing metode dengan pembagian data training dan data testing yaitu 80% : 20% menunjukkan metode Naïve Bayes Classifier dengan accuracy 86.30%, precision 84.51%, recall 87.21% dan dan f1 score 85.84%. dan Support Vector Machine dengan accuracy 89.56%, precision 83.62%, recall 94.5% dan f1 score 88.73%
Shailvi Sharma, Dharmveer Singh (2020)	Cyber-Bullying Detection using Naive Bayes and N-Gram	-	<ol style="list-style-type: none"> <li>1. Menggunakan Case Folding, Stemming, Lemmatization untuk tahapan preprocessing, dan Bog of Words untuk feature extraction.</li> <li>2. Menggunakan algoritma Naïve Bayes untuk</li> </ol>	Naïve Bayes dengan Uni-Gram mencapai akurasi 66,77%, Naïve Bayes dengan Bi-Gram mencapai akurasi 67,29%, Naïve Bayes dengan Tri-Gram mencapai 57,86% akurasi dan Naïve Bayes

Author	Judul Artikel	Nama Jurnal	Metodologi	Hasil
			mengklasifikasi cyber bullying di dalam dokumen	dengan Ni-Gram mencapai 65,09% ketepatan. Oleh karena itu, akurasi rata-rata adalah sekitar 64,46%
Aditya Gaydhani, Vikrant Doma, Shrikant Kendre and Laxmi Bhagwat (2018)	Detecting Hate Speech and Offensive Language on Twitter using Machine Learning: An N-gram and TFIDF based Approach	-	1. Menggunakan fitur n-gram 2. Menggunakan TF-IDF Vectorizer 3. Menggunakan algoritma klasifikasi berupa : Logistic Regression, Naive Bayes and Support Vector Machine	Permodelan yang paling baik dihasilkan oleh algoritma Logistic Regression dengan nilai akurasi sebesar 95.6 %
Irfan Dwi Wijaya, Aji Gautama Putrada, Dita Oktaria (2021)	Penggunaan Metode K-Fold untuk Data Imbalance pada Klasifikasi HWE dan QPQ dalam Kejahatan Tweet Pelecehan Seksual	-	1. Menggunakan tahapan preprocessing berupa metode K-Fold 2. Menggunakan 2 metode klasifikasi berupa Naïve Bayes dan KNN	Nilai akurasi pada model klasifikasi Naïve bayes dengan menggunakan K-Fold cross Validation sebesar 90.7%, sedangkan model klasifikasi KNN yang menggunakan metode yang sama memiliki nilai akurasi mencapai 87.9%.
Firman Sriyono, Kusriani, Asro Nasiri (2022)	Detecting Hate Speech In Twitter Using Long Short-Term Memory and	-	1. Menggunakan Case Folding, Stemming, Lemmatization untuk tahapan preprocessing,	Metode Long Short-Term Memory memiliki akurasi 84,39% yang lebih baik

Author	Judul Artikel	Nama Jurnal	Metodologi	Hasil
	Naïve Bayes Method		dan Bog of Words untuk feature extraction. 2. Menggunakan metode Long Short-Term Memory dibandingkan dengan metode Naïve Bayes.	apabila dibandingkan dengan metode Naïve Bayes yang memiliki akurasi 55,18%. Metode tersebut memiliki akurasi terbaik untuk parameter dengan epoch 50 dan fungsi aktivasi sigmoid
Jawaid Ahmed Siddiqui, Siti Sophiyati Yuhaniz, Zulfiqar Ali Memon, Yumna Amin (2021)	Improving Hate Speech Detection Using Machine and Deep Learning Techniques: A Preliminary Study	-	1. Menggunakan TF-IDF Vectorizer 2. Menggunakan Long Short-Term Memory, Linear SVC, Logistic Regression, Random Forest dan Naïve Bayes	Logistic Regression menghasilkan model yang memiliki akurasi sebesar 0,74, diikuti oleh Random Forest 0,72, dan Linear SVC yang menghasilkan model terbaik dengan memberikan akurasi 0,85. Untuk hasil subkategorisasi, model Random Forest memberikan hasil kategorisasi yang sedikit lebih baik dibandingkan yang lain. Random Forest menghasilkan model terbaik dengan memberikan akurasi sebesar 0,95.

Author	Judul Artikel	Nama Jurnal	Metodologi	Hasil
Muhammad Okky Ibrohim, Erryan Sazany, Indra Budi (2019)	Identify Abusive and Offensive Language in Indonesian Twitter using Deep Learning Approach	-	1. Menggunakan TF-IDF Vectorizer 2. Menggunakan FastText dan Word2Vec 2. Menggunakan metode klasifikasi berupa Long-Short Term Memory	LSTM menghasilkan akurasi sebesar 83,68% F1-Score, meningkat sekitar 19,44% dari penelitian sebelumnya yang hanya mencapai 70,06%. Di antara dua feature selection yang dipakai, FastText memberikan klasifikasi yang sedikit lebih baik dibandingkan Word2Vec, khususnya saat membedakan bahasa kasar yang menyerang atau tidak menyerang suatu pihak
Junanda Patihullah, Edi Winarko (2019)	Hate Speech Detection for Indonesia Tweets Using Word Embedding And Gated Recurrent Unit	-	1. Menggunakan TF-IDF Vectorizer. 2. Menggunakan Word2Vec 3. Menggunakan metode Gated Recurrent Unit (GRU), Support Vector Machine, Naive Bayes, Random Forest dan Regresi logistik	Hasil terbaik dihasilkan dari metode GRU dengan akurasi sebesar 92,96% yang lebih besar dari model lain dengan hasil berupa : akurasi dari SVM sebesar 83,65%, akurasi dari NB sebesar 79,80%. akurasi dari BLR sebesar 78,84%. akurasi dari

Author	Judul Artikel	Nama Jurnal	Metodologi	Hasil
				RFDT sebesar 82,69%
Iqbal Zulfikar Muhammad, Muhammad Nasrun, Casi Setianingsih (2020)	Hate Speech Detection using Global Vector and Deep Belief Network Algorithm	-	1. Menggunakan feature selection 2. Menggunakan algoritma berupa Deep Belief Network	Deep Belief Network menghasilkan nilai akurasi sebesar 86,00%, presisi sebesar 82,00%, recall sebesar 89,13% dan F1-Score sebesar 85,42%
M. Ali Fauzi, Anny Yuniarti (2018)	Ensemble Method for Indonesian Twitter Hate Speech Detection	-	1. Menggunakan Bag of Words dengan TF-IDF Vectorizer Menggunakan Naïve Bayes, K-Nearest Neighbors, Maximum Entropy, Random Forest, Support Vector Machines, dan dua ensemble methodes, yaitu hard voting dan soft voting	Hasil terbaik dihasilkan oleh model Naïve Bayes dengan nilai F1-Skor sebesar 78,3% apabila menggunakan dataset yang seimbang. Hasil terbaik dengan menggunakan dataset yang tidak seimbang dihasilkan oleh model Maximum Entropy dengan nilai F1-skor sebesar 84.1%
Angela Marpaung, Rita Rismala, Hani Nurrahmi (2021)	Hate Speech Detection in Indonesian Twitter Texts using Bidirectional Gated Recurrent Unit	-	1. Menggunakan SVM dan RFDT 2. Menggunakan BiGRU dan IndoBERT yang telah dilatih dengan BiGRU. 3. Menggunakan Word2vec dan FastText	Metode terbaik dihasilkan oleh IndoBERT-GRU yang mendapatkan hasil akurasi sebesar 84.77%. BiGRU mendapatkan hasil terbaik karena kemampuannya dalam

Author	Judul Artikel	Nama Jurnal	Metodologi	Hasil
				menyimpan data, terutama kemampuan penyimpanan teks
Fauzi Ihsan, Iwan Iskandar, Nazruddin Safaat Harahap, Surya Agustian (2021)	Algoritme decision tree untuk mendeteksi ujaran kebencian dan bahasa kasar multilabel pada Twitter berbahasa Indonesia	-	1. Menggunakan fitur Word Embedding 2. Menggunakan algoritma Decision Tree untuk klasifikasi.	Fitur lexicon di dalam model klasifikasi Decision Tree menghasilkan akurasi tertinggi untuk mendeteksi ketiga kelas, yaitu kelas ujaran kebencian, kata-kata kasar dan level ujaran kebencian, . Rata-rata akurasi dari ketiga kelas tersebut mengalami peningkatan dari 69,77 % menjadi 70,48 % untuk komposisi data train-test 90:10, dan dari 69,35 % menjadi 69,54 % untuk komposisi train-test sebesar 80:20
Luh Putu Ary Sri Tjahyanti (2020)	Pendeteksian Bahasa Kasar (Abusive Language) Dan Ujaran Kebencian (Hate Speech) Dari	-	1. Menggunakan fitur ekstraksi berupa fitur frekuensi seperti kata n-gram dan huruf n-gram 2. Menggunakan algoritma :	Pada skenario pertama model Naïve Bayes dengan fitur kata unigram + bigrams memberikan hasil terbaik

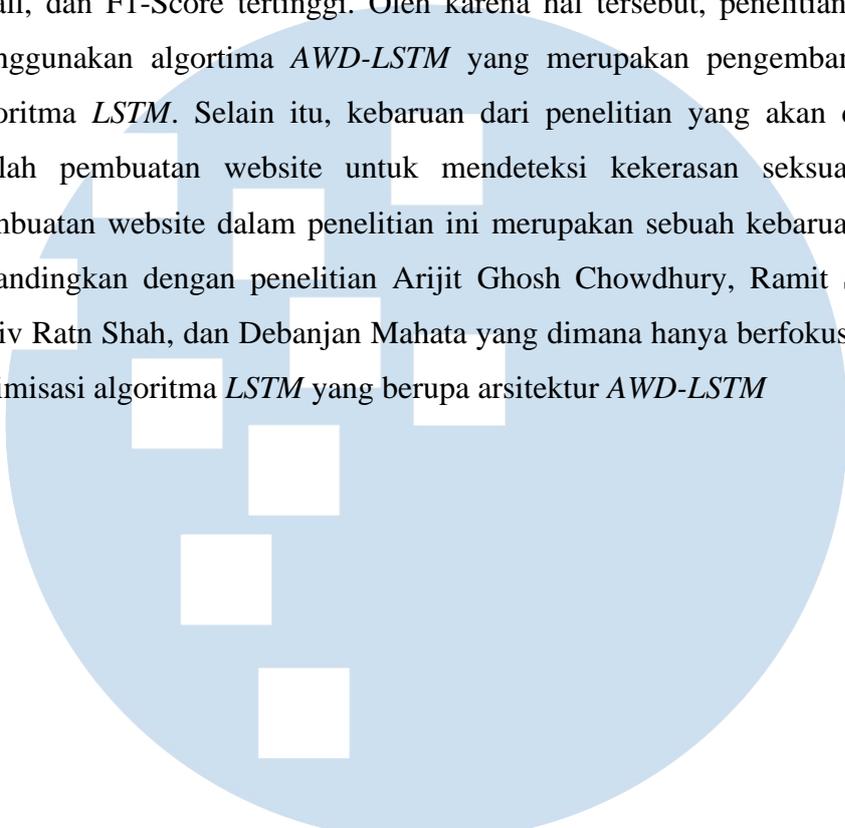
Author	Judul Artikel	Nama Jurnal	Metodologi	Hasil
	Komentar Di Jejaring Sosial		Naïve Bayes (NB), Support Vector Machine (SVM), dan Random Forest Decision Tree (RFDT) 3. Menggunakan 2 skenario dengan jenis label yang berbeda untuk menemukan akurasi tertinggi	dengan F1-Score sebesar 71,15%, sedangkan untuk skenario kedua, Naïve Bayes dengan kata unigram memberikan hasil terbaik dengan F1-Score sebesar 87,26%
Nassharikh Abdulloh, Ahmad Fathan Hidayatullah (2020)	Deteksi Cyberbullying pada Cuitan Media Sosial Twitter	-	1. Menggunakan TF-IDF Vectorizer. 2. Menggunakan algoritma berupa: K-Nearest Neighbor (KNN), Support Vector Machine dengan linear kernel (SVM), Logistic Regression, dan Multinomial Naïve Bayes	Model yang menghasilkan akurasi terbaik adalah model SVM dengan akurasi 99.75%, diikuti oleh Logistic Regression dengan akurasi sebesar 99.5%, Multinomial Naïve Bayes sebesar 96.2%, dan K-Nearest Neighbor (KNN) sebesar 92%
Ahmad Fathan Hidayatullah, Afa Aulia Fadila Yusuf, Kiki Purnama Juwairi, Royan Abida N. Nayoan (2019)	Identifikasi Konten Kasar pada Tweet Bahasa Indonesia	-	1. Menggunakan TF-IDF Vectorizer. 2. Menggunakan Multinomial Naïve Bayes dan Support Vector Machine dengan linear kernel.	Hasil akurasi algoritma SVM mencapai 0.9928, sedangkan Multinomial Naïve Bayes menghasilkan akurasi sebesar 0.9834. Nilai presisi dari SVM adalah 0.9914 dan Multinomial Naive Bayes

Author	Judul Artikel	Nama Jurnal	Metodologi	Hasil
				0.9912. Nilai recall untuk SVM adalah 0.9946, sedangkan nilai recall untuk Multinomial Naïve Bayes yaitu sebesar 0.9762. Nilai FI dari SVM adalah 0.9930 dan Multinomial Naive Bayes adalah 0.9836.
Hadi Kurniawan Sidiq, Dana Sulistyono, Kusumo, Indra Lukmana Sardi (2019)	Mendeteksi Cyberhate pada Twitter Menggunakan Text Classification dan Crowdsourced Labeling	-	1. Menggunakan TF-IDF Vectorizer. 2. Menggunakan SVM dan Crowdsourced Labelling.	Hasil dari kernel linear SVM adalah 96%. Apabila dikombinasikan dengan metode crowdsourced labelling hasil akurasi yang dihasilkan adalah 99%.
Rashi Shah, Srushti Aparajit, Riddhi Chopdekar, Rupali Patil (2020)	Machine Learning based Approach for Detection of Cyberbullying Tweets	-	1. Menggunakan TF-IDF Vectorizer 2. Menggunakan Support Vector Classifier, Logistic Regression, Naïve Bayes, Random Forest Classifier	Logistic Regression adalah model yang menghasilkan hasil yang paling akurat di antara yang model lainnya. Model ini memiliki nilai presisi sebesar 91%, recall sebesar 96%, F1 skor sebesar 93%, dan akurasi sebesar 93%.

Author	Judul Artikel	Nama Jurnal	Metodologi	Hasil
Monirah Abdullah Al-Ajlan, Mourad Ykhlef (2018)	Deep Learning Algorithm for Cyberbullying Detection	-	1. Menggunakan TF-IDF dan LDA 2. Menggunakan algoritma berupa : Naïve Bayes, SVM, J 48 dan CNN-CB	Model yang paling baik dihasilkan oleh algoritma CNN-CB. Model tersebut menghasilkan model dengan nilai akurasi sebesar 95%
Aaminah Ali, Adeel M. Syed (2022)	Cyberbullying Detection Using Machine Learning	-	Menggunakan algoritma berupa : Support Vector Classifier, Logistic Regression, Naïve Bayes, Random Forest Classifier dan Ensemble	Hasil menunjukkan bahwa SVM dan Ensemble memiliki performa yang lebih baik dibandingkan classifier lainnya dengan akurasi rata-rata 79%. Hasil kedua adalah model Regresi Logistik dengan akurasi sebesar 78%, kemudian Random Forest dengan sebesar 76,7%, dan Naïve Bayes dengan akurasi sebesar 76%

Berdasarkan tabel 2.1, penelitian ini akan mengambil referensi dari penelitian Arijit Ghosh Chowdhury, Ramit Sawhney, Rajiv Ratn Shah, dan Debanjan Mahata yang dimana menggunakan metode klasifikasi berupa *Average Stochastic Gradient Descent Weight-Dropped LSTM (AWD-LSTM)* sebagai salah satu algoritma pembandingan. Metode tersebut dipilih dikarenakan kemampuan dari algoritma *AWD-LSTM* yang memiliki kemampuan baik untuk menangkap ketergantungan kata yang lebih kompleks dibandingkan dengan

algoritma yang lain yang hanya berfokus dalam mencapai nilai akurasi, presisi, recall, dan F1-Score tertinggi. Oleh karena hal tersebut, penelitian ini akan menggunakan algoritma *AWD-LSTM* yang merupakan pengembangan dari algoritma *LSTM*. Selain itu, kebaruan dari penelitian yang akan dilakukan adalah pembuatan website untuk mendeteksi kekerasan seksual verbal. Pembuatan website dalam penelitian ini merupakan sebuah kebaruan apabila dibandingkan dengan penelitian Arijit Ghosh Chowdhury, Ramit Sawhney, Rajiv Ratn Shah, dan Debanjan Mahata yang dimana hanya berfokus terhadap optimisasi algoritma *LSTM* yang berupa arsitektur *AWD-LSTM*

A large, light blue watermark logo of Universitas Multimedia Nusantara (UMMN) is centered on the page. It features a stylized 'U' and 'M' inside a circle, with a grid of squares below it.

# UMMN

U N I V E R S I T A S  
M U L T I M E D I A  
N U S A N T A R A