

KLASIFIKASI CYBERBULLYING MENGGUNAKAN METODE *LEARNING VECTOR QUANTIZATION (LVQ)* DENGAN EKSTRAKSI FITUR *TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY (TF-IDF)*

SKRIPSI

**Oleh:
NUR AZIZAH MUNIR
NIM. 19650051**



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2024**

KLASIFIKASI *CYBERBULLYING* MENGGUNAKAN METODE *LEARNING VECTOR QUANTIZATION (LVQ)* DENGAN EKSTRAKSI FITUR *TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY (TF-IDF)*

SKRIPSI

Diajukan Kepada :
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk Memenuhi Salah Satu Persyaratan Dalam
Memperoleh Gelar Sarjana Komputer (S.Kom)

Oleh:
NUR AZIZAH MUNIR
NIM. 19650051

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM MALANG
2024**

HALAMAN PERSETUJUAN

KLASIFIKASI *CYBERBULLYING* MENGGUNAKAN METODE *LEARNING VECTOR QUANTIZATION (LVQ)* DENGAN EKSTRAKSI FITUR *TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY (TF-IDF)*

SKRIPSI

Oleh :
NUR AZIZAH MUNIR
NIM. 19650051

Telah Diperiksa dan Disetujui untuk Diuji:
Tanggal: 02 Mei 2024

Pembimbing I,



Okta Oomaruddin Aziz, M.Kom
NIP. 19911019 201903 1 013


Pembimbing II,



Dr. M. Amin Hariyadi, M.T
NIP. 19670018 200501 1 001

Mengetahui,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang




Dr. Fachrul Kurniawan, M.MT, IPM
NIP. 19771020 200912 1 001

HALAMAN PENGESAHAN

KLASIFIKASI *CYBERBULLYING* MENGGUNAKAN METODE *LEARNING VECTOR QUANTIZATION (LVQ)* DENGAN EKSTRAKSI FITUR *TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY (TF-IDF)*

SKRIPSI

Oleh:
NUR AZIZAH MUNIR
NIM. 19650051

Telah dipertahankan
Di depan Dewan Penguji Skripsi dan dinyatakan diterima sebagai salah satu
persyaratan untuk memperoleh gelar Sarjana Komputer (S.Kom)
Tanggal: 12 Juni 2024

Susunan Dewan Penguji

Ketua Penguji	:	<u>Dr. Cahyo Crysdian, MCS</u> NIP. 19740424 200901 1 008	(.....)
Anggota Penguji I	:	<u>Hani Nurhayati, M.T</u> NIP. 19780625 200801 2 006	(.....)
Anggota Penguji II	:	<u>Okta Qomaruddin Aziz, M.Kom</u> NIP. 19911019 201903 1 013	(.....)
Anggota Penguji III	:	<u>Dr. M. Amin Hariyadi, M.T</u> NIP. 19670018 200501 1 001	(.....)

Mengetahui dan Mengesahkan,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang



Dr. Fachrul Kurniawan, M.MT
NIP. 19771020 200912 1 001

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Nur Azizah Munir
NIM : 19650051
Fakultas/Prodi : Sains dan Teknologi / Teknik Informatika
Judul Skripsi : Klasifikasi *Cyberbullying* Menggunakan Metode *Learning Vector Quantization (LVQ)* Dengan Ekstraksi Fitur *Term Frequency-Inverse Document Frequency (TF-IDF)*

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambil alihan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini merupakan hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 20 Mei 2024
Yang membuat pernyataan,



Nur Azizah Munir
NIM. 19650051

HALAMAN MOTTO

**“Tanggung jawab dan selesaikan apa yang sudah
kamu mulai”**

HALAMAN PERSEMBAHAN

Alhamdulillahirobbil'alamiin segala puji bagi Allah subhanahu wa ta'ala, serta shalawat dan salam kepada nabi agung Muhammad SAW. Penulis mempersembahkan skripsi ini untuk orang tua, kakak, serta teman yang telah memberikan banyak support, perhatian, kasih sayang serta arahan sehingga penulis dapat menyelesaikan skripsi ini.

KATA PENGANTAR

Assalamu'alaiakum Wr. Wb.

Bismillahirrahmaanirrahiim, puji syukur kehadiran Allah, dengan itu penulis mengucapkan terima kasih khususnya kepada:

1. Prof. Dr. M. Zainuddin, M.A. selaku Rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Prof. Dr. Sri Harini, M.Si. selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Dr. Fachrul Kurniawan, M.MT. selaku Ketua Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Okta Qomaruddin Aziz, M.Kom. selaku Dosen Pembimbing I dan Dr. M. Amin Hariyadi, M.T. selaku Dosen Pembimbing II, yang membimbing penulis dengan penuh kesabaran dan keikhlasan dalam meluangkan waktunya untuk membimbing sehingga penulis dapat menyelesaikan tugas akhir ini.
5. Seluruh Dosen dan Staf di Program Studi Teknik Informatika, dengan ikhlas memberikan ilmu, bantuan, serta dorongan semangat selama perkuliahan.
6. Kedua orang tua penulis, Bapak Munir, Ibu Khusnul Chotimah dan kakak saya Ria Aisah Munir yang telah memberikan banyak dukungan, doa serta bantuan sehingga penulis mampu menyelesaikan masa studi hingga mencapai gelar sarjana.
7. Keluarga besar Program Studi Teknik Informatika terutama Angkatan 2019 ALIEN “Alliance of Informatics Engineering” yang telah memberikan dukungan untuk saling menyelesaikan skripsi serta satu orang yang berarti bagi saya Deri Fauzi yang senantiasa meluangkan waktu untuk memberikan dorongan dan motivasi dalam masa perkuliahan hingga proses penyusunan skripsi.

8. Diri saya sendiri, yang telah mampu melawan rasa malas dan takut serta bekerja keras untuk menyelesaikan skripsi ini. Terimakasih telah mampu kooperatif dan bertahan dalam menikmati proses pengerjaan skripsi.

Teriring doa, semoga amal baik yang telah diberikan kepada penulis mendapat balasan dari Allah SWT. Laporan Tugas Akhir ini telah ditulis dengan teliti dan sebaik-baiknya, namun saran dan kritikan yang membangun masih penulis harapkan untuk kemudian. Wassalamu'alaikum Wr. Wb.

Malang, 26 April 2023

Penulis

DAFTAR ISI

HALAMAN PENGAJUAN	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN	iv
PERNYATAAN KEASLIAN TULISAN	v
HALAMAN MOTTO	vi
HALAMAN PERSEMBAHAN	vii
KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiii
ABSTRAK	xv
ABSTRACT	xvi
مستخلص البحث	xvii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Pernyataan Masalah	6
1.3 Tujuan Penelitian	7
1.4 Batasan Masalah	7
1.5 Manfaat Penelitian	7
BAB II STUDI LITERATUR	8
2.1 <i>Cyberbullying</i>	8
2.2 <i>Learning vector quantization</i>	9
BAB III METODOLOGI PENELITIAN	13
3.1 Desain Penelitian.....	13
3.2 Pengumpulan Data	14
3.3 Desain Sistem.....	15
3.4 <i>Preprocessing</i>	16
3.4.1 <i>Normalization</i>	17
3.4.2 <i>Tokenizing</i>	19
3.4.3 <i>Stemming</i>	19
3.4.4 <i>Stopword Removal</i>	20
3.4.5 <i>TF-IDF</i>	20
3.4.6 <i>Learning vector quantization</i>	26
BAB IV UJI COBA DAN PEMBAHASAN	35
4.1 Skenario Uji Coba	35
4.1.1 Data Penelitian	39
4.2 Hasil Uji Coba.....	42
4.2.1 Uji Coba Model A	42
4.2.2 Uji Coba Model B	54
4.2.3 Uji Coba Model C	63
4.2.4 Uji Coba K-Fold Cross Validation	73
4.3 Pembahasan.....	77

4.3.1 Perubahan learning rate	78
4.3.2 Pembagian Ratio Data	80
4.3.3 K-fold Cross Validation	81
4.3.4 Integrasi Islam	83
BAB V PENUTUP.....	88
5.1 Kesimpulan	88
5.2 Saran	89
DAFTAR PUSTAKA	
LAMPIRAN	

DAFTAR GAMBAR

Gambar 3. 1 Desain Penelitian.....	13
Gambar 3. 2 Desain Sistem.....	15
Gambar 3. 3 Flowchart Praprocessing	17
Gambar 3. 4 Flowchart Normalization	18
Gambar 3. 5 Flowchart TF-IDF	22
Gambar 3. 6 Arsitektur Jaringan LVQ.....	26
Gambar 3. 7 Flowchart LVQ	28
Gambar 4. 1 Confusion Matrix Multiclass	37
Gambar 4. 2 Visualisasi metode <i>10-fold cross validation</i>	42
Gambar 4. 3 <i>Confusion matrix 60:40 learning rate 0.1</i>	43
Gambar 4. 4 <i>Confusion matrix 60:40 learning rate 0.01</i>	46
Gambar 4. 5 <i>Confusion Matrix 60:40 Learning Rate 0.001</i>	49
Gambar 4. 6 <i>Confusion matrix 60:40 learning rate 0.0001</i>	51
Gambar 4. 7 <i>Confusion matrix 70:30 learning rate 0.1</i>	54
Gambar 4. 8 <i>Confusion matrix 70:30 learning rate 0.01</i>	56
Gambar 4. 9 <i>Confusion matrix 70:30 learning rate 0.001</i>	59
Gambar 4. 10 <i>Confusion matrix 70:30 learning rate 0.0001</i>	61
Gambar 4. 11 <i>Confusion matrix 80:20 learning rate 0.1</i>	63
Gambar 4. 12 <i>Confusion matrix 80:20 learning rate 0.01</i>	65
Gambar 4. 13 <i>Confusion matrix 80:20 learning rate 0.001</i>	68
Gambar 4. 14 <i>Confusion matrix 80:20 learning rate 0.0001</i>	70
Gambar 4. 15 Grafik perubahan learning rate.....	80

DAFTAR TABEL

Tabel 2. 1 Penelitian Terkait	11
Tabel 3. 1 Pengumpulan Data	14
Tabel 3. 2 Contoh Normalization	18
Tabel 3. 3 Contoh <i>Tokenizing</i>	19
Tabel 3. 4 Contoh <i>Stemming</i>	19
Tabel 3. 5 Contoh <i>Stopword removal</i>	20
Tabel 3. 6 Jumlah kata yang muncul dalam kelas.....	22
Tabel 3. 7 Perhitungan Term Frequency.....	23
Tabel 3. 8 Perhitungan IDF.....	24
Tabel 3. 9 Perhitungan TF-IDF.....	24
Tabel 3. 10 Hasil Perhitungan TF-IDF	25
Tabel 3. 11 Data Latih	30
Tabel 3. 12 Data Uji.....	30
Tabel 3. 13 Data bobot akhir.....	33
Tabel 4. 1 Pembagian Dataset.....	35
Tabel 4. 2 Pembagian Learning Rate	36
Tabel 4. 3 Sample Data Penelitian	39
Tabel 4. 4 Jumlah data penelitian.....	40
Tabel 4. 5 Ratio Pembagian Data.....	41
Tabel 4. 6 Tabel Confusion Matrix 60:40 Learning Rate 0.1	43
Tabel 4. 7 Hasil Perhitungan TP, TN, FP, FN (60:40 Learning Rate 0.1).....	44
Tabel 4. 8 Nilai <i>Precision, Recall, F1-score learning rate 0.1</i>	45
Tabel 4. 9 Tabel Confusion Matrix 60:40 Learning Rate 0.01	46
Tabel 4. 10 Hasil Perhitungan TP, TN, FP, FN (60:40 Learning Rate 0.01).....	47
Tabel 4. 11 Nilai <i>Precision, Recall, F1-score learning rate 0.01</i>	48
Tabel 4. 12 Tabel Confusion Matrix 60:40 Learning Rate 0.001	49
Tabel 4. 13 Hasil Perhitungan TP, TN, FP, FN (60:40 Learning Rate 0.001).....	49
Tabel 4. 14 Nilai <i>Precision, Recall, F1-score learning rate 0.001</i>	51
Tabel 4. 15 Tabel Confusion Matrix 60:40 Learning Rate 0.0001	52
Tabel 4. 16 Hasil Perhitungan TP, TN, FP, FN (60:40 Learning Rate 0.0001)....	52
Tabel 4. 17 Nilai <i>Precision, Recall, F1-score learning rate 0.0001</i>	53
Tabel 4. 18 Tabel Confusion Matrix 70:30 Learning rate 0.1	54
Tabel 4. 19 Hasil Perhitungan TP, TN, FP, FN (70:30 Learning rate 0.1).....	55
Tabel 4. 20 Nilai <i>Precision, Recall, F1-score learning rate 0.1</i>	56
Tabel 4. 21 Tabel Confusion Matrix 70:30 Learning rate 0.01	57
Tabel 4. 22 Hasil Perhitungan TP, TN, FP, FN (70:30 Learning rate 0.01).....	57
Tabel 4. 23 Nilai <i>Precision, Recall, F1-score learning rate 0.01</i>	58
Tabel 4. 24 Tabel Confusion Matrix 70:30 Learning rate 0.001	59
Tabel 4. 25 Hasil Perhitungan TP, TN, FP, FN (70:30 Learning rate 0.001).....	59
Tabel 4. 26 Nilai <i>Precision, Recall, F1-score learning rate 0.001</i>	61
Tabel 4. 27 Tabel Confusion Matrix 70:30 Learning rate 0.0001	61
Tabel 4. 28 Hasil Perhitungan TP, TN, FP, FN (70:30 Learning rate 0.0001).....	62

Tabel 4. 29 Nilai <i>Precision, Recall, F1-score</i> learning rate 0.0001	63
Tabel 4. 30 Tabel Confusion Matrix 80:20 Learning rate 0.1	64
Tabel 4. 31 Hasil Perhitungan TP, TN, FP, FN (80:20 Learning rate 0.1)	64
Tabel 4. 32 Nilai <i>Precision, Recall, F1-score</i> learning rate 0.1	65
Tabel 4. 33 Tabel Confusion Matrix 80:20 Learning rate 0.01	66
Tabel 4. 34 Hasil Perhitungan TP, TN, FP, FN (80:20 Learning rate 0.01)	66
Tabel 4. 35 Nilai <i>Precision, Recall, F1-score</i> learning rate 0.01	67
Tabel 4. 36 Tabel Confusion Matrix 80:20 Learning rate 0.001	68
Tabel 4. 37 Hasil Perhitungan TP, TN, FP, FN (80:20 Learning rate 0.001)	68
Tabel 4. 38 Nilai <i>Precision, Recall, F1-score</i> learning rate 0.001	70
Tabel 4. 39 Tabel Confusion Matrix 80:20 Learning rate 0.0001	71
Tabel 4. 40 Hasil Perhitungan TP, TN, FP, FN (80:20 Learning rate 0.0001)	71
Tabel 4. 41 Nilai <i>Precision, Recall, F1-score</i> learning rate 0.0001	72
Tabel 4. 42 Hasil Pengujian Parameter	73
Tabel 4. 43 Hasil akurasi 10-fold cross validation (60:40)	74
Tabel 4. 44 Hasil akurasi 10-fold cross validation (70:30)	75
Tabel 4. 45 Hasil akurasi 10-fold cross validation (80:20)	76
Tabel 4. 46 Perubahan Learning Rate	79
Tabel 4. 47 Pembagian Ratio Data	81

ABSTRAK

Munir, Nur Azizah. 2024. **Klasifikasi Cyberbullying Menggunakan Metode *Learning Vector Quantization (LVQ)* Dengan Ekstraksi Fitur *Term Frequency-Inverse Document Frequency (TF-IDF)***. Skripsi. Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Okta Qomaruddin Aziz, M.Kom (II) Dr. M. Amin Hariyadi, M.T.

Kata Kunci: Klasifikasi *cyberbullying*, *Learning Vector Quantization*, *Cyberbullying*

Seiring meningkatnya penggunaan media sosial, cyberbullying telah mencapai titik puncaknya. Anonimitas di internet membuat cyberbullying sangat merusak, menyebabkan korban merasa tidak ada jalan keluar dari pelecehan tersebut. Setiap individu harus selalu waspada terhadap cyberbullying dan dihimbau untuk melindungi diri sendiri serta orang lain dari hal ini. Penulis membuat model yang secara otomatis menandai tweet yang berpotensi membahayakan dan memecah pola pesan kebencian. Penelitian ini membahas klasifikasi cyberbullying menggunakan metode Learning Vector Quantization (LVQ) dengan ekstraksi fitur Term Frequency-Inverse Document Frequency (TF-IDF). Dataset yang digunakan berisi sekitar 1000 tweet yang telah dilabeli menjadi 5 jenis yaitu religion, gender, age, ethnicity dan not cyberbullying. Data yang telah terlabeli akan dilakukan proses pre-processing dengan melalui empat tahapan yaitu normalization, tokenizing, stemming, dan stopword removal. Kemudian data tersebut akan dilakukan pembobotan kata menggunakan TF-IDF. Skenario pengujian dilakukan dengan menggunakan beberapa model pembagian data dan variasi parameter learning rate. Data training akan dilakukan pemodelan menggunakan metode Learning Vector Quantization. Sehingga dari hasil pemodelan tersebut sistem dapat melakukan klasifikasi. Berdasarkan hasil uji coba yang dilakukan terhadap klasifikasi cyberbullying menggunakan total 12 skenario pengujian mampu menghasilkan hasil akurasi paling optimal sebesar 86.5%. Dengan demikian, hasil penelitian ini menunjukkan bahwa penggunaan metode Learning Vector Quantization dengan ekstraksi fitur TF-IDF serta pemilihan learning rate yang tepat dan pembagian data yang optimal dapat secara efektif mengklasifikasikan cyberbullying pada media sosial twitter.

ABSTRACT

Munir, Nur Azizah. 2024. **Cyberbullying Classification Using the Learning Vector Quantization (LVQ) Method and the Term Frequency-Inverse Document Frequency (TF-IDF) Feature Extraction**. Thesis. Informatics Engineering Department Faculty of Science and Technology Universitas Islam Negeri Maulana Malik Ibrahim Malang. Advisors: (I) Okta Qomaruddin Aziz, M.Kom (II) Dr. M. Amin Hariyadi, M.T.

The increase in social media use has led cyberbullying to achieve its peak point. Anonymity on the internet makes cyberbullying destructive. It causes the victims to feel that they have no way out. Each individual must be wary of cyberbullying and she/he must protect him/herself and other people from this. The researcher has made a model that automatically marks potentially dangerous tweets and breaks hatred message patterns. The research discusses cyberbullying classification using the Learning Vector Quantization (LVQ) method and Term Frequency-Inverse Document Frequency (TF-IDF) feature extraction. The used dataset consists of 1,000 tweets labeled into five types: religion, gender, age, ethnicity, and not cyberbullying. The labeled data will be pre-processed through four steps: normalization, tokenizing, stemming, and stop word removal. Then, the researcher conducted word weighing with the data using TF-IDF. A testing scenario involved some data distribution models and several learning rate parameters. The data training would be modeled using the Learning Vector Quantization method, by which the system could generate classification. Based on the try-out result on cyberbullying classification using 12 testing scenarios, the most optimal accuracy was 86,5%. Therefore, the research result shows that using the Learning Vector Quantization method with TF-IDF feature extraction, adding the correct learning rate, and an optimal data distribution can effectively classify cyberbullying on social media Twitter.

Keywords: Cyberbullying classification, the Learning Vector Quantization, Cyberbullying.

مستخلص البحث

منير، نور عزيزة. 2024. تصنيف التنمر الإلكتروني باستخدام طريقة تعلم تكميم المتجهات (LVQ) مع استخراج ميزة تردد المصطلح-معكوس تردد الوثيقة (TF-IDF). البحث الجامعي. قسم الهندسة المعلوماتية، كلية العلوم والتكنولوجيا بجامعة مولانا مالك إبراهيم الإسلامية الحكومية، مالانج. المشرف الأول: أوكنا قمر الدين عزيز، الماجستير. المشرف الثاني: د. محمد أمين هريادي، الماجستير.

الكلمات الرئيسية: تصنيف التنمر الإلكتروني، تعلم تكميم المتجهات، التنمر الإلكتروني.

مع زيادة استخدام وسائل التواصل الاجتماعي، وصل التنمر الإلكتروني إلى ذروته. إن إخفاء الهوية على الإنترنت يجعل التنمر الإلكتروني ضارا للغاية، مما يجعل الضحايا يشعرون أنه لا توجد طريقة للخروج من المضايقات. يجب أن يكون كل فرد دائما على دراية بالتنمر الإلكتروني ويتم حثه على حماية نفسه والآخرين من ذلك. ابتكرت الباحثة نموذجا يقوم تلقائيا بالإبلاغ عن التغريدات التي يحتمل أن تكون ضارة ويكسر أنماط رسائل الكراهية. ناقش هذا البحث تصنيف التنمر الإلكتروني باستخدام طريقة تعلم تكميم المتجهات (LVQ) من خلال استخراج ميزة تي اف-اي دي دف (TF-IDF). تحتوي مجموعة البيانات المستخدمة على حوالي 1000 تغريدة تم تصنيفها إلى 5 أنواع، وهي الدين والجنس والعمر والعرق وعدم التنمر الإلكتروني. ستم معالجة البيانات المصنفة مسبقا من خلال أربع مراحل، وهي التطبيع، والترميز، والتوقيف، وإزالة الكلمة الموقفة. ثم سيتم ترجيح البيانات باستخدام تي اف-اي دي دف (TF-IDF). تم تنفيذ سيناريو الاختبار باستخدام عدة نماذج مشاركة البيانات والاختلافات في معلمات معدل التعلم. سيتم نمذجة بيانات التدريب باستخدام طريقة تعلم تكميم المتجهات. بحيث من نتائج النمذجة، يمكن للنظام أن يقوم بالتصنيف. استنادا إلى نتائج الاختبارات التي أجريت على تصنيف التنمر الإلكتروني باستخدام 12 سيناريوهات الاختبار، تمكنت من تحقيق أفضل نتيجة دقة بنسبة 86.5%. وبالتالي، أظهرت نتائج هذا البحث أن استخدام طريقة تعلم تكميم المتجهات مع استخراج ميزة تي اف-اي دي دف، بالإضافة إلى اختيار معدل التعلم الصحيح والمشاركة المثلى للبيانات يمكن أن يصنف بشكل فعال التنمر الإلكتروني على وسائل التواصل الاجتماعي "تويتر".

BAB I

PENDAHULUAN

1.1 Latar Belakang

Media sosial telah menjadi alat komunikasi yang dianggap cukup efektif pada saat ini, mengingat jumlah pengguna yang sangat besar. Dengan akses yang lebih mudah ke internet, terutama media sosial, hal ini sangat menguntungkan dunia bisnis, namun tidak sedikit kasus kriminal yang sering timbul dari pengguna media sosial atau menggunakan media sosial sebagai alat (Janisar et al., 2021). Pada saat ini, media sosial tidak hanya berkembang dan menjadi sarana pertukaran informasi kesehatan dan internet saja, akan tetapi juga menjadi tempat untuk saling menghina dan merusak kehidupan satu sama lain. *Cyberbullying* terjadi karena kurangnya pengetahuan yang ada.

Pemahaman mendasar dari bully meliputi pembuli dan korban yang dirundung baik secara fisik maupun perkataan yang dilakukan secara langsung dan atau tidak langsung. *Cyberbullying* atau perundungan daring dipahami sebagai perbuatan yang dengan sengaja serta berulang mengirim pesan maupun komentar yang bersifat *ofensif* atau merendahkan kepada seorang individu dan sebuah kelompok pada media teknologi komunikasi (Margono, 2019). Banyak individu yang menjadi korban dalam konteks perundungan sering mengalami dampak yang menghancurkan, termasuk tingkat kecemasan yang meningkat, perasaan depresi yang mendalam, dan penurunan signifikan dalam rasa percaya diri mereka. Dampak fisik juga seringkali terjadi, seperti sakit kepala, cedera fisik, bahkan mencapai tahap yang tragis seperti kematian. Namun, tidak hanya itu, efek psikologisnya juga

berbahaya dengan depresi menjadi satu di antara beberapa faktor yang dapat memicu seseorang untuk memilih untuk mengakhiri hidupnya. Salah satu bentuk perundungan yang semakin umum terjadi adalah cyberbullying, di mana para pelaku menggunakan berbagai platform online, mulai dari pesan teks hingga media sosial, untuk menyerang dan merendahkan korban. Cyberbullying dapat memanifestasikan dirinya dalam berbagai bentuk, seperti mempermalukan secara fisik (body shaming), merendahkan berdasarkan ras, hobi, orientasi seksual, dan bahkan seksisme. Namun, perlu dicatat bahwa media sosial adalah salah satu arena di mana cyberbullying seringkali mencapai puncaknya. Twitter merupakan salah satu media di mana sering digunakan untuk melakukan aksi *cyberbullying*, karena orang dapat mengunggah tulisan kasar yang memiliki hubungan dengan individu lain dengan tujuan menyudutkan serta merusak kredibilitas korban sehingga korban merasa terluka dan malu, sementara pelaku merasa gembira dan puas dikarenakan tujuannya telah tercapai (Febriana & Budiarto, 2019). Dalam lingkungan Twitter, pesan-pesan yang bernada negatif dapat dengan cepat menyebar luas dan mengakibatkan tekanan emosional yang besar bagi korban. Selain itu, keberadaan anonimitas di balik layar sering kali mendorong perilaku yang lebih agresif dan tanpa belas kasihan. Korban cyberbullying pada Twitter seringkali merasa terasing dan tidak aman dalam berinteraksi online. Penting untuk meningkatkan kesadaran akan dampak negatif cyberbullying di platform seperti Twitter dan mengambil langkah-langkah untuk mencegah dan menanggulangi tindakan tersebut. Sebagai pengguna media sosial, sebenarnya semua orang dapat dengan sengaja atau tidak sengaja menjadi pelaku atau korban *cyberbullying*. *Cyberbullying* di media sosial

memiliki efek yang lebih besar karena dapat dengan mudah menyebar luas dan dapat menjadi pemicu bagi orang lain untuk melakukan hal yang sama. Meningkatnya kasus *cyberbullying* yang menggunakan media sosial, terutama Twitter, seharusnya dihindari dengan mengetahui konten apa pun yang terkait dengan *cyberbullying*. Klasifikasi *cyberbullying* dalam konten media sosial Twitter dianggap perlu agar dapat meminimalkan dampak buruk dari *cyberbullying*.

Islam sebagai agama menyediakan pedoman etika dan moral yang komprehensif untuk membimbing umatnya dalam setiap aspek kehidupan, termasuk dalam dunia maya. Seiring dengan itu, penting untuk menjembatani antara prinsip-prinsip Islam dan perkembangan teknologi modern. Dalam konteks ini, surah Al-Hujurat ayat 13 memberikan pandangan yang mendalam tentang keseragaman penciptaan Allah, yang dapat diaplikasikan dalam pembentukan sistem klasifikasi untuk *cyberbullying*.

يَا أَيُّهَا النَّاسُ إِنَّا خَلَقْنَاكُمْ مِنْ ذَكَرٍ وَأُنْثَىٰ وَجَعَلْنَاكُمْ شُعُوبًا وَقَبَائِلَ لِتَعَارَفُوا ۗ إِنَّ أَكْرَمَكُمْ عِنْدَ اللَّهِ أَتْقَىٰكُمْ ۚ إِنَّ اللَّهَ عَلِيمٌ
حَبِيرٌ

"Tidak ada seekor hewan pun (yang berada) di bumi dan burung-burung yang terbang dengan kedua sayapnya, melainkan semuanya merupakan umat (juga) seperti kamu. Tidak ada sesuatu pun yang Kami luputkan di dalam kitab, kemudian kepada Tuhannya mereka dikumpulkan." (Q.S Al-Hujurat:13).

(Hai manusia, sesungguhnya Kami menciptakan kalian dari seorang laki-laki dan seorang perempuan) yakni dari Adam dan Hawa (dan Kami menjadikan kalian berbangsa-bangsa) lafal Syu'uuban adalah bentuk jamak dari lafal Sya'bun, yang artinya tingkatan nasab keturunan yang paling tinggi (dan bersuku-suku) kedudukan suku berada di bawah bangsa, setelah suku atau kabilah disebut Imarah,

lalu Bathn, sesudah Bathn adalah Fakhdz dan yang paling bawah adalah Fashilah. Contohnya ialah Khuzaimah adalah nama suatu bangsa, Kinanah adalah nama suatu kabilah atau suku, Quraisy adalah nama suatu Imarah, Qushay adalah nama suatu Bathn, Hasyim adalah nama suatu Fakhdz, dan Al-Abbas adalah nama suatu Fashilah (supaya kalian saling kenal-mengenal) lafal Ta'aarafuu asalnya adalah Tata'aarafuu, kemudian salah satu dari kedua huruf Ta dibuang sehingga jadilah Ta'aarafuu; maksudnya supaya sebagian dari kalian saling mengenal sebagian yang lain bukan untuk saling membanggakan ketinggian nasab atau keturunan, karena sesungguhnya kebanggaan itu hanya dinilai dari segi ketakwaan. (Sesungguhnya orang yang paling mulia di antara kalian di sisi Allah ialah orang yang paling bertakwa. Sesungguhnya Allah Maha Mengetahui) tentang kalian (lagi Maha Menegal) apa yang tersimpan di dalam batin kalian.

Sebagaimana dijelaskan dalam Tafsir Jalalain yang memberikan pandangan tentang upaya membangun sistem klasifikasi cyberbullying yang efektif, kita dapat merujuk pada konsep-konsep yang dijelaskan dalam ayat Al-Qur'an yang menguraikan variasi identitas manusia dan pentingnya saling mengenal. Pertama-tama, seperti yang disebutkan dalam ayat tersebut, manusia berasal dari satu pasangan, Adam dan Hawa, yang mencerminkan asal-usul bersama kita sebagai manusia. Hal ini menggambarkan pentingnya menyatukan pemahaman tentang keberagaman manusia dalam satu kerangka yang inklusif. Pembuatan sistem klasifikasi yang mempertimbangkan lima kelas seperti *religion*, *age*, *gender*, *ethnicity*, dan *not bullying*, mencerminkan kompleksitas identitas manusia yang dijelaskan dalam ayat tersebut. Misalnya, kelas *religion* dapat membantu

mengidentifikasi pesan atau perilaku yang menyerang agama seseorang, sejalan dengan pengertian tentang bangsa dan kabilah. Sementara itu, kelas *age* membantu dalam memetakan pesan-pesan atau perilaku yang menargetkan seseorang berdasarkan usianya, mirip dengan klasifikasi suku dan sub-suku. *Gender*, sebagai kelas dalam sistem klasifikasi cyberbullying, mencakup perilaku atau pesan yang bersifat merendahkan seseorang berdasarkan jenis kelaminnya. Sesuai dengan konsep suku dan sub-suku dalam ayat tersebut adalah dalam memahami variasi identitas manusia secara utuh. *Ethnicity*, mengidentifikasi perilaku yang bersifat rasialis atau diskriminatif terhadap etnis tertentu, sesuai dengan kompleksitas hubungan antara suku dan sub-suku. Terakhir, kelas *not bullying* adalah penanda untuk perilaku yang tidak termasuk dalam kategori cyberbullying, sejalan dengan ajaran ayat tersebut tentang pentingnya saling mengenal dan memahami satu sama lain tanpa menghakimi atau merendahkan. Sistem klasifikasi yang mencerminkan keragaman identitas manusia dapat memperkuat kesadaran nilai kesetaraan dan keadilan, sehingga membantu mencegah dan menangani cyberbullying.

Metode klasifikasi menjadi salah satu pendekatan yang penting dalam upaya penanganan *cyberbullying* (Wang & Potika, 2021). Dengan bantuan teknik-teknik komputasi dan pembelajaran mesin, pendekatan klasifikasi memiliki potensi untuk mengidentifikasi perilaku *cyberbullying* secara otomatis, memungkinkan intervensi yang lebih cepat dan efektif. Salah satu metode yang efektif dalam penelitian ini adalah *Learning vector quantization* (LVQ), yang telah terbukti berhasil dalam berbagai aplikasi klasifikasi. LVQ telah terbukti mampu mengenali pola klasifikasi penting dengan akurasi yang baik (Song et al., 2018). Selain itu,

kemampuan mengatasi multi-kelas menjadi faktor utama dalam pemilihan LVQ. Meskipun banyak penelitian telah dilakukan untuk mengatasi *cyberbullying*, penerapan metode LVQ dalam konteks ini masih relatif terbatas. Kebutuhan untuk mengembangkan pendekatan klasifikasi yang lebih efektif dan tepat dalam mengidentifikasi perilaku *cyberbullying* menjadi semakin mendesak. Oleh karena itu, penelitian ini akan menelaah penerapan metode *Learning vector quantization* (LVQ) dalam klasifikasi *cyberbullying*, dengan tujuan menghasilkan model yang akurat dalam mengidentifikasi perilaku yang merugikan dalam lingkungan digital. Dalam penelitian ini akan membahas bagaimana metode LVQ dapat digunakan untuk mengatasi tantangan dalam mengklasifikasikan perilaku *cyberbullying* dengan akurasi yang lebih tinggi. Penelitian ini juga akan mengeksplorasi dampak penggunaan teknik preprocessing data dalam meningkatkan performa model. Dengan menggabungkan potensi metode LVQ dan pemahaman mendalam mengenai karakteristik perilaku *cyberbullying*, penelitian ini akan memberikan wawasan penting dalam pengembangan pendekatan yang lebih efektif untuk melawan fenomena yang merugikan ini. Penelitian ini diharapkan dapat diterapkan dalam platform digital untuk mendeteksi dan mencegah *cyberbullying* secara real-time serta berkontribusi dalam keamanan digital.

1.2 Pernyataan Masalah

Berapa tingkat akurasi metode *Learning vector quantization* (LVQ) dalam mengklasifikasi *cyberbullying* dengan mempertimbangkan kompleksitas karakteristik bahasa dalam data teks dan kemampuan LVQ dalam mengidentifikasi pola klasifikasi yang relevan.

1.3 Tujuan Penelitian

Mengukur dan mengevaluasi tingkat keakuratan metode *Learning vector quantization* (LVQ) dalam mengklasifikasi *cyberbullying* dengan mempertimbangkan kompleksitas karakteristik bahasa dalam data teks dan kemampuan LVQ dalam mengidentifikasi pola klasifikasi yang relevan.

1.4 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut:

1. Penelitian yang akan dilakukan adalah mengimplementasikan algoritma LVQ untuk klasifikasi *cyberbullying* dengan ekstraksi fitur TF-IDF
2. Dataset yang digunakan dalam penelitian ini diambil dari kaggle berupa teks komentar berbahasa inggris pada sosial media Twitter

1.5 Manfaat Penelitian

Penelitian ini akan membuktikan tingkat akurasi dan hasil klasifikasi dengan algoritma LVQ yang akan memberikan pemahaman mendalam mengenai efektivitas metode ini dalam mengidentifikasi kasus-kasus *cyberbullying*. Sistem ini ditujukan untuk Kementerian Komunikasi dan Informatika (Kominfo) sebagai pihak yang bertanggung jawab atas pengelolaan informasi dan komunikasi di Indonesia, serta Lembaga Perlindungan Anak Indonesia (LPAI) yang berfokus pada perlindungan hak anak, termasuk perlindungan dari *cyberbullying*. Melalui hasil penelitian ini, diharapkan kedua lembaga tersebut dapat mengimplementasikan teknologi klasifikasi berbasis LVQ untuk meningkatkan efektivitas dalam mendeteksi dan menangani kasus-kasus *cyberbullying* di Indonesia.

BAB II

STUDI LITERATUR

2.1 *Cyberbullying*

Cyberbullying merupakan suatu bentuk perilaku yang melibatkan penggunaan teknologi digital seperti internet, media sosial, atau pesan teks untuk secara sengaja dan berulang kali melakukan tindakan merendahkan, mengintimidasi, atau menyakiti individu, yang dapat mengakibatkan dampak psikologis, emosional, dan sosial yang serius pada korban. Menurut Riswanto & Marsinun pada tahun 2020 *Cyberbullying* adalah tindakan negatif yang dilakukan oleh seseorang atau kelompok tertentu dengan cara mengirimkan pesan teks, foto, gambar meme, dan video ke akun media sosial seseorang dengan tujuan untuk menyindir, menghina, melecehkan bahkan mendiskriminasi individu.

Penelitian yang dilakukan Rahman (2021) tentang klasifikasi ujaran kebencian pada media sosial twitter menggunakan SVM. Data yang digunakan pada penelitian ini yaitu 700 data dan data uji sebanyak 300 data. Hasil klasifikasi penelitian ini terdapat 5 kelas diantaranya suku, agama, ras, antar golongan dan netral. Pada penelitian ini penggunaan kernel RBF memperoleh akurasi terbaik yaitu 93% dibanding dengan kernel linear dan sigmoid. Penelitian mengenai klasifikasi *cyberbullying* juga dilakukan Radjavani & Bayu Sasongko (2023) dengan menganalisa perbandingan algoritma CNN dan LSTM. Pada penelitian tersebut menggunakan Word2Vec untuk mengelompokkan kata-kata pada setiap data yang memiliki kemiripan antar satu sama lain. Hasil penelitian tersebut

menunjukkan algoritma LSTM lebih unggul dengan akurasi sebesar 85.28%, sedangkan algoritma CNN menghasilkan akurasi sebesar 77.90%.

Luqyana (2018) melakukan penelitian mengenai Analisis *cyberbullying* pada komentar Instagram dengan menggunakan metode SVM. Pada penelitian ini menggunakan ekstraksi fitur yang sama yaitu TF-IDF. Pengujian dilakukan dengan mengatur nilai degree kernel polynomial menjadi 2, nilai learning rate sebesar 0,0001, dan jumlah iterasi maksimum 200 kali. Dari hasil pengujian tersebut, ditemukan bahwa akurasi tertinggi sebesar 90% tercapai ketika komposisi data latih dan data uji masing-masing 50%. Sebagai perbandingan, penelitian analisis *cyberbullying* yang dilakukan oleh Zaenal (2020) menggunakan metode Naïve Bayes. Dalam penelitian tersebut, terdapat 300 data training dan 100 data testing. Hasil klasifikasi dibagi menjadi dua kategori, yaitu *cyberbullying* dan non *cyberbullying*, dan mendapatkan akurasi sebesar 86%.

2.2 *Learning vector quantization*

Learning vector quantization pertama kali diperkenalkan oleh Kohonen (1990). Algoritma *Learning vector quantization* adalah metode dalam jaringan saraf tiruan yang digunakan untuk mengelompokkan pola-pola sinyal ke dalam berbagai kategori. Konsep utama dari algoritma ini adalah menggunakan codebook vector untuk merepresentasikan setiap kategori yang berbeda dalam inputan dari data observasi. Proses pengambilan keputusan klasifikasi dalam algoritma *Learning vector quantization* didasarkan pada prinsip pemilihan tetangga terdekat. Pemilihan ini dilakukan dengan cara menghitung jarak antara codebook vector dan data observasi menggunakan suatu metrik atau fungsi jarak tertentu.

Learning vector quantization dapat digunakan untuk pengolahan text processing seperti klasifikasi, analisis sentimen, pengolahan citra. Salah satu keunggulan LVQ adalah kemampuannya untuk beradaptasi dengan data yang tidak terstruktur dan heterogen. Akbari (2022) tentang analisis sentimen pada sosial media twitter dengan menggunakan metode *Learning vector quantization* menghasilkan akurasi sebesar 73.15%. Penelitian ini menunjukkan potensi LVQ dalam menangani kompleksitas data teks sosial media.

Penelitian dengan metode klasifikasi dilakukan oleh Lestari (2019) pada penelitiannya tentang klasifikasi video clickbait pada youtube dengan menggunakan *Learning vector quantization* dan lexicon-based features sebagai pembobotan kata. Akurasi sistem yang diperoleh dari hasil klasifikasi penelitian ini sebesar 90.91%. Hal ini menunjukkan bahwa LVQ dapat digabungkan dengan fitur lexicon-based untuk meningkatkan kinerja klasifikasi.

Sudhakar dan Kaliyamurthie (2023) mengenai prediksi berita palsu dengan menggunakan pengukuran jarak hamming. Perbandingan empat algoritma yaitu *Learning vector quantization*, Passive Aggressive Classifier, LS-SVM dan LSTM. Menggunakan tiga fitur ekstraksi antara lain TF-IDF Vectorizer, Count Vectorizer, dan N-Gram. Algoritma LVQ memperoleh nilai akurasi tertinggi sebesar 93.54%. Untuk algoritma lainnya Passive Aggressive Classifier 72.15%, LS-SVM 73.25%, dan LSTM 58.81%. Pada penelitian ini terbukti bahwa metode LVQ mampu mengklasifikasi teks dengan lebih baik. Selain itu, LVQ juga memiliki keunggulan dalam kecepatan konvergensi dan efisiensi komputasi. Hal ini menjadikan LVQ sbagai salah satu algoritma yang fleksibel dan kuat untuk berbagai tugas klasifikasi.

Tabel 2. 1 Penelitian Terkait

No	Nama	Judul Penelitian	Kesamaan	Perbedaan
1.	Rahman et al (2021)	Klasifikasi Ujaran Kebencian Pada Media Sosial Menggunakan SVM	<ul style="list-style-type: none"> - Membahas tentang <i>cyberbullying</i> - Data berasal dari media sosial twitter - Klasifikasi Multiclass 	<ul style="list-style-type: none"> - Algoritma yang digunakan yaitu SVM - Menggunakan tiga perbandingan kernel yaitu kernel linear, sigmoid dan RBF
2.	Alifqi Radjavani, Theopilus Bayu Sasongko (2023)	Analisa Perbandingan Algoritma CNN Dan LSTM untuk Klasifikasi Pesan <i>Cyberbullying</i> pada Twitter	<ul style="list-style-type: none"> - Membahas tentang <i>cyberbullying</i> - Menggunakan dataset yang sama - Data berasal dari media sosial Twitter 	<ul style="list-style-type: none"> - Menggunakan perbandingan metode CNN dan LSTM - Pembobotan kata dilakukan dengan menggunakan <i>Word2Vec</i> - Untuk mengukur kemiripan teks dalam penelitian ini menggunakan cosine similarity
3.	Luqyana et al (2018)	Analisis Sentimen <i>Cyberbullying</i> pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine	<ul style="list-style-type: none"> - Membahas mengenai <i>cyberbullying</i> - Menggunakan ekstraksi fitur TF-IDF 	<ul style="list-style-type: none"> - Data berasal dari media sosial Instagram - Data yang digunakan 400 data. Sedangkan penelitian ini menggunakan data kurang lebih 25000 data - Menggunakan metode SVM
4.	Zaenal et al (2020)	Analisis Sentimen terhadap Komentar Negatif di Media Sosial Facebook dengan Metode Klasifikasi Naïve Bayes	<ul style="list-style-type: none"> - Membahas tentang <i>cyberbullying</i> 	<ul style="list-style-type: none"> - Data berasal dari media sosial Facebook - Menggunakan 300 data training dan 100 data testing - Menggunakan metode Naïve Bayes
5.	Lestari et al (2019)	Klasifikasi Video Clickbait pada YouTube Berdasarkan Analisis Sentimen Komentar Menggunakan <i>Learning vector quantization</i> (LVQ)	<ul style="list-style-type: none"> - Metode <i>Learning vector quantization</i> 	<ul style="list-style-type: none"> - Membahas tentang Video Clickbait pada Youtube - Pembobotan kata dilakukan dengan menggunakan <i>Lexicon Based Features</i>

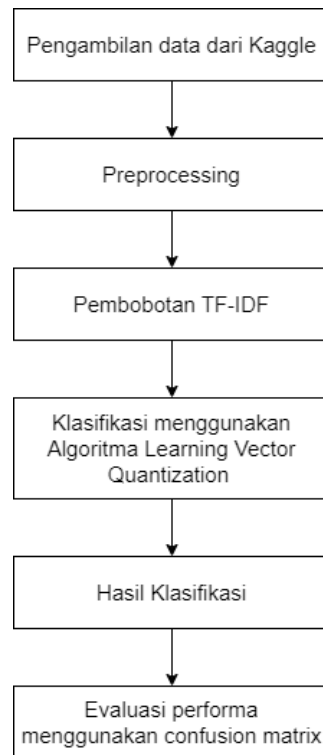
		dan LexiconBased Features		- Data pada penelitian terdahulu lebih sedikit yaitu 300 data sedangkan penelitian ini menggunakan data kurang lebih 25000 data.
6.	Akbari et al (2017)	Analisis Sentimen Menggunakan Metode <i>Learning vector quantization</i>	<ul style="list-style-type: none"> - Data berasal dari media sosial Twitter - Metode <i>Learning vector quantization</i> - Menggunakan ekstraksi fitur TF-IDF 	- Menghasilkan klasifikasi 2 kelas yaitu <i>tweet</i> positif dan <i>tweet</i> negatif
7.	Sudhakar M, K.P. Kaliyamurthie (2023)	Prediksi Berita Palsu dengan Menggunakan Pengukuran Jarak Hamming	<ul style="list-style-type: none"> - Data berasal dari media sosial Twitter - Menggunakan tiga metode salah satunya metode <i>Learning vector quantization</i> - Dilakukan tiga ekstraksi fitur salah satunya TF-IDF 	<ul style="list-style-type: none"> - Menggunakan tambahan ekstraksi fitur Count Vectorizer dan N-Gram - Membandingkan 4 algoritma metode yaitu LVQ, Passive Aggressive Classifier, LS-SVM, dan LSTM - Menghasilkan klasifikasi 2 kelas yaitu berita palsu dan berita akurat

BAB III

METODOLOGI PENELITIAN

3.1 Desain Penelitian

Desain penelitian ini yang akan menjadi pedoman untuk setiap tahap yang dilakukan dalam penelitian ini ditampilkan dalam bentuk diagram pada Gambar 3.1 berikut.



Gambar 3. 1 Desain Penelitian

Proses sistem ini dimulai dengan mengambil data dari Kaggle. Selanjutnya, dataset dibagi menjadi dua bagian, yaitu data pelatihan dan data uji. Data tersebut berisi tweet berbahasa Inggris dan terbagi menjadi 6 kelas diantaranya “not_cyberbullying”, ”gender”, “religion”, “age”, “ethnicity”,

“other_cyberbullying”. Namun pada penelitian ini menggunakan 5 kelas yaitu “gender”, “religion”, “age”, “ethnicity” dan “not_cyberbullying”. Setelah itu, data datih akan melewati tahap preprocessing untuk persiapan analisis. Metode pembobotan TF-IDF diterapkan pada data pelatihan untuk melatih model yang nantinya akan diuji menggunakan data uji. Setelah itu hasilnya akan diproses dengan menggunakan model klasifikasi *Learning vector quantization* untuk melakukan klasifikasi dan dievaluasi dengan menggunakan *confusion matrix*.

3.2 Pengumpulan Data

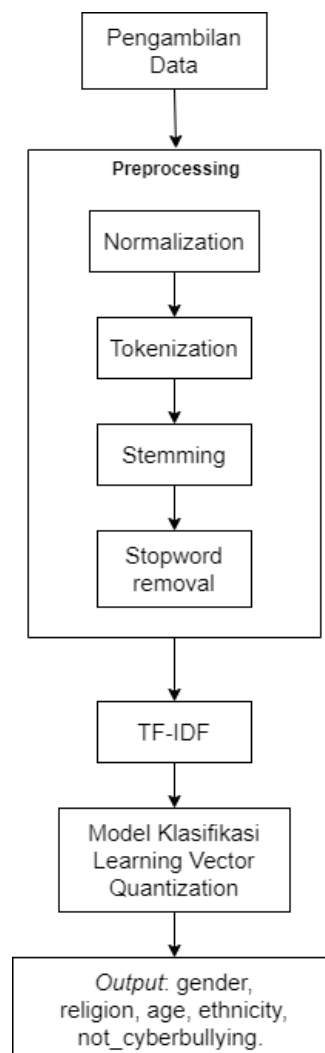
Pada penelitian ini pengumpulan data dan informasi menggunakan data sekunder *cyberbullying* komentar berbahasa inggris yang didapat dari Kaggle dengan dataset yang digunakan *Cyberbullying Classification* yaitu dengan alamat (<https://www.kaggle.com/datasets/andrewmvd/cyberbullying-classification>). Pada penelitian ini hanya menggunakan 1000 data dengan 200 data pada masing-masing kelas. Data tersebut mencakup komentar yang diklasifikasikan ke dalam lima kategori, yaitu age, ethnicity, gender, religion, dan not cyberbullying.

Tabel 3. 1 Pengumpulan Data

Dokumen	Komentar	Kelas
D1	@BelieveTed i feel so fucking bad for you. You dumb fucking nigger im Doxing you and im glad you live in michigan im beating the fuck outa u	ethnicity
D2	Nobody asks to be made fun of but it happens to everyone gay jokes straight jokes rape jokes all of its gonna happen	gender
D3	@biebervalue @greenlinerzjm You fucking moron, slavery is approved right in the Quran. Read it you stupid idiot	religion
D4	Just realized that all the girls who used to bully me in middle school and hs are either single moms, divorced, or dating someone who cheats on them and if thatâ€™s not the best karma ever idk what is	age
D5	That was fun. Now it's over. This is why I can't even consider sites like TechRaptr as anything more than hit piece blogs.	not_cyberbullying

3.3 Desain Sistem

Penelitian ini memiliki sebuah sistem klasifikasi yang dirancang untuk menggambarkan tahap-tahap yang digunakan pada sistem, dimulai dari pengambilan data hingga menghasilkan data yang telah diklasifikasikan. Setelah itu, sistem akan mengevaluasi kinerjanya dengan menggunakan *confusion matrix*. Berikut adalah rancangan desain sistem yang akan diimplementasikan dalam penelitian ini.



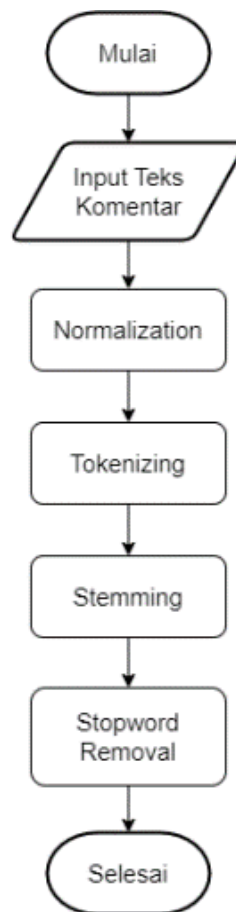
Gambar 3. 2 Desain Sistem

Dalam penelitian ini, sistem klasifikasi yang dirancang memiliki proses awal dengan mengambil data dari sumber Kaggle yang relevan. Setelah data diambil, dataset ini menjalani serangkaian tahapan penting yang dimulai dengan preprocessing data untuk memastikan kebersihan dan kualitasnya, termasuk penghapusan data yang tidak relevan dan normalisasi teks. Selanjutnya, dilakukan pembobotan dengan metode TF-IDF untuk menghasilkan representasi numerik dari teks. Algoritma klasifikasi *Learning vector quantization* digunakan untuk melakukan klasifikasi sentimen pada data tersebut. Hasil klasifikasi akan menjadi 5 kelas yaitu “gender”, “religion”, “age”, “ethnicity” dan “not_cyberbullying”. Untuk mengukur kinerja sistem, sistem evaluasi dilakukan dengan *confusion matrix* untuk mengukur *accuracy*, *precision*, *recall*, dan *F1-score*. Dengan demikian, sistem ini tidak hanya mengklasifikasikan data, tetapi juga melibatkan serangkaian langkah yang komprehensif dalam pengolahan dan evaluasi data untuk menghasilkan hasil yang akurat.

3.4 Preprocessing

Preprocessing merupakan langkah penting dalam mengolah teks untuk membuatnya lebih terstruktur dan lebih mudah dipahami oleh sistem, sehingga memfasilitasi proses klasifikasi. Dalam tahap ini, teks mengalami serangkaian proses untuk mempersiapkannya secara optimal. Proses normalisasi teks, termasuk mengubah huruf menjadi huruf kecil dan menghapus tanda baca, dilakukan untuk memastikan konsistensi data. Kemudian tokenisasi, yang memecah teks menjadi potongan kata yang disebut token. Stemming diterapkan untuk mengubah kata-kata ke bentuk dasarnya, sehingga variasi kata dengan makna yang sama dapat

diidentifikasi dengan benar. Selanjutnya, dilakukan penghapusan stop words, yaitu kata-kata dangkal yang tidak memiliki makna signifikan dalam analisis. Langkah terakhir teks diubah menjadi representasi numerik menggunakan teknik seperti TF-IDF Tahapan pada *preprocessing* dapat dilihat pada Gambar 3.2 dibawah ini.

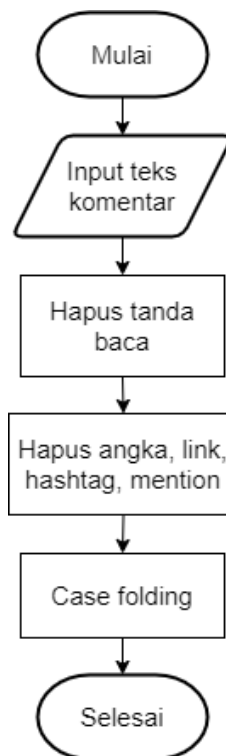


Gambar 3. 3 Flowchart Praprocessing

3.4.1 *Normalization*

Normalization digunakan untuk menghapus perbedaan-perbedaan umum yang terdapat pada dataset teks opini. Tahapan yang dilakukan dalam normalization meliputi penghapusan angka, penghapusan spasi kosong, penghapusan tanda hubung, penghapusan tanda baca, penghapusan *hyperlink*, dan *case folding*, yaitu

mengubah semua huruf besar menjadi huruf kecil. Tujuan dari normalisasi ini adalah untuk menyederhanakan teks dan membuatnya lebih konsisten sehingga memudahkan proses analisis, klasifikasi, dan ekstraksi informasi dari dataset teks opini. Dengan melakukan normalisasi, kita dapat memastikan bahwa data teks yang diolah lebih mudah diproses secara komputasional dan memberikan hasil yang lebih akurat. Diagram alur proses normalization dapat dilihat pada Gambar 3.4 berikut.



Gambar 3. 4 Flowchart Normalization

Tabel 3. 2 Contoh Normalization

Sebelum Normalization	@BelieveTed i feel so fucking bad for you. You dumb fucking nigger im Doxing you and im glad you live in michigan im beating the fuck outa u
Setelah Normalization	i feel so fucking bad for you you dumb fucking nigger im doxing you and im glad you live in michigan im beating the fuck outa u

3.4.2 Tokenizing

Tahap tokenizing adalah tahapan teks atau kalimat komentar diuraikan menjadi kata yang lebih kecil, yang disebut token atau term (Aninditya et al., 2019). Tokenizing ini dapat dianggap sebagai bentuk segmentasi teks. Proses ini penting untuk memudahkan analisis dan pemrosesan lebih lanjut dari teks. Dalam analisis teks, tokenizing berfungsi sebagai langkah awal untuk ekstraksi informasi dan pembuatan model teks.

Tabel 3. 3 Contoh *Tokenizing*

Sebelum <i>Tokenizing</i>	[i feel so fucking bad for you you dumb fucking nigger im doxing you and im glad you live in michigan im beating the fuck outa u]
Setelah <i>Tokenizing</i>	["i", "feel", "so", "fucking", "bad", " for", " you", " you", "dumb", "fucking", "nigger", "im", "doxing", " you", "im", "glad", " you", "live", "in", "michigan", "im", "beating", "the", "fuck", "outa", "u"]

3.4.3 Stemming

Pada tahap stemming, kata diubah menjadi bentuk dasarnya yang disesuaikan dengan struktur bahasa yang dipergunakan (Aninditya et al., 2019). Langkah yang dilakukan pada tahap ini adalah menghapus imbuhan yang terdapat pada kata untuk mendapatkan bentuk kata aslinya. Proses stemming dalam penelitian ini menggunakan library Python yaitu Natural Language Tool Kit (NLTK) yang dapat diakses melalui tautan berikut <https://www.nltk.org/>.

Tabel 3. 4 Contoh *Stemming*

Sebelum <i>Stemming</i>	["i", "feel", "so", "fucking", "bad", " for", " you", " you", "dumb", "fucking", "nigger", "im", "doxing", " you", "im", "glad", " you", "live", "in", "michigan", "im", "beating", "the", "fuck", "outa", "u"]
Setelah <i>Stemming</i>	["i", "feel", "so", "fuck", "bad", " for", " you", " you", "dumb", "fuck", "nigger", "im", "dox", " you", "im", "glad", " you", "live", "in", "michigan", "im", "beat", "the", "fuck", "outa", "u"]

3.4.4 *Stopword Removal*

Stopword removal merupakan sebuah proses dimana kata-kata umum atau yang disebut dengan *stopword* akan dihilangkan dari sebuah teks karena kata-kata tersebut tidak memberikan kontribusi penting dalam analisis teks (Sarica & Luo, 2021). Contohnya, kata-kata seperti kata depan, kata sambung, dan kata-kata umum lainnya. Proses *stopword removal* dilakukan agar proses analisis teks dapat berjalan lebih cepat dan juga dapat mengurangi dimensi data yang tidak perlu, sehingga dapat meningkatkan kualitas dari analisis yang dilakukan. *Stopword removal* merupakan sebuah proses penting dalam pengolahan teks untuk menghilangkan kata-kata umum yang tidak memberikan kontribusi signifikan dalam analisis teks. Hal ini membantu mempercepat proses analisis dan mengurangi dimensi data yang tidak relevan, yang pada gilirannya meningkatkan efisiensi dan akurasi dari analisis teks yang dilakukan

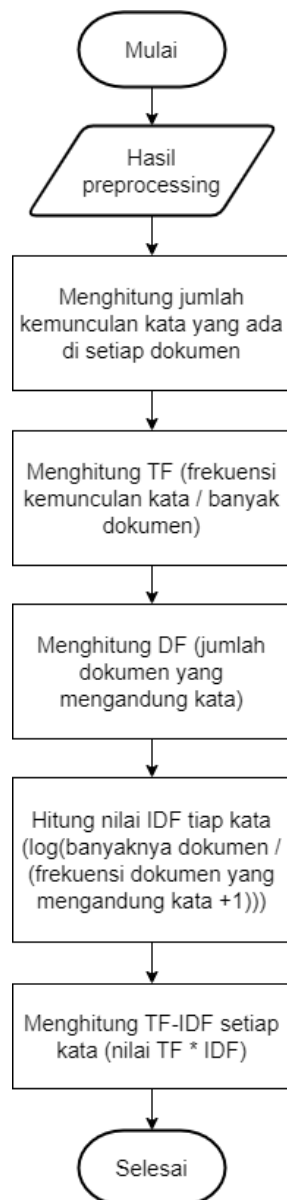
Tabel 3. 5 Contoh *Stopword removal*

Sebelum <i>Stopword removal</i>	["i", "feel", "so", "fuck", "bad", " for", " you", " you", "dumb", "fuck","nigger", "im", "dox", " you", "im", "glad", " you", "live", "in", "michigan", "im", "beat", "the", "fuck", "outa", "u"]
Setelah <i>Stopword removal</i>	["feel", "fuck", "bad", "dumb", "fuck","nigger", "dox", "glad", "live", "michigan", "beat", "fuck"]

3.4.5 *TF-IDF*

Dalam penelitian ini, penulis telah memilih metode Term Frequency-Inverse Document Frequency (TF-IDF) sebagai strategi utama untuk memberikan bobot pada kata-kata dalam dataset yang digunakan. Proses dimulai dengan tahap preprocessing, di mana data disiapkan dan dibersihkan untuk analisis lebih lanjut. Setelah tahap ini selesai, kata-kata dalam dataset menjadi term yang akan menjadi

fokus dalam langkah-langkah berikutnya.. Algoritma ini memberikan bobot yang lebih besar kepada kata-kata yang jarang muncul dalam dataset, sehingga memiliki nilai informatif yang lebih tinggi (Imamah & Rachman, 2020). Term Frequency (TF) merupakan ukuran tingkat keseringan sebuah kata muncul pada sebuah teks. Dengan kata lain, ini adalah ukuran relatif dari frekuensi kemunculan sebuah kata dalam sebuah dokumen. Semakin sering sebuah kata muncul, semakin tinggi pula nilai TF-nya untuk dokumen tersebut. Inverse Document Frequency (IDF) merupakan sebuah algoritma yang dipergunakan dalam menghitung probabilitas kebalikan dari kemunculan sebuah kata dalam seluruh teks. IDF memperhitungkan seberapa umum atau jarang suatu kata dalam kumpulan dokumen. Kata-kata yang sering muncul dalam banyak dokumen akan memiliki nilai IDF yang rendah, sedangkan kata-kata yang jarang muncul dalam dokumen akan memiliki nilai IDF yang tinggi. Hal ini memungkinkan untuk memberikan bobot yang lebih tinggi pada kata-kata yang jarang muncul namun memiliki kepentingan yang besar dalam suatu konteks. Dengan menggabungkan TF dan IDF, metode TF-IDF memberikan bobot yang lebih baik pada kata-kata dalam dataset dengan memperhitungkan kedua faktor tersebut. Hal ini membantu dalam menyoroti kata-kata yang memiliki relevansi dan signifikansi tertentu dalam suatu analisis, seringkali meningkatkan kemampuan model atau algoritma dalam memahami dan mengekstraksi informasi yang relevan dari dataset yang dianalisis. Ilustrasi dari alur pada proses perhitungan TF-IDF dapat dilihat pada Gambar berikut.



Gambar 3. 5 Flowchart TF-IDF

Tabel 3. 6 Jumlah kata yang muncul dalam kelas

Kata	Jumlah Kata				
	D1	D2	D3	D4	D5
fuck	1	0	0	0	0
call	1	0	0	1	0
feel	1	0	0	0	0
bad	1	0	0	0	0
dumb	1	0	0	0	0
nigger	1	0	0	0	0

dox	1	0	0	0	0
glad	1	0	0	0	0
live	1	0	0	0	0
michigan	1	0	0	0	0
beat	1	0	0	0	0
bully	0	1	0	0	0
rude	0	1	0	0	0
school	0	1	0	0	0
destroy	0	0	1	0	0
life	0	0	1	0	0
welcom	0	0	1	0	0
world	0	0	1	0	0
islam	0	0	1	0	0
terror	0	0	1	0	0
female	0	0	0	1	0
basic	0	0	0	1	0
bitch	0	0	0	1	0
start	0	0	0	0	1
thursday	0	0	0	0	1
class	0	0	0	0	1
cancel	0	0	0	0	1
tomorrow	0	0	0	0	1
Total	10	3	6	3	5

Kemudian dihitung menggunakan Algoritma TF-IDF sesuai persamaan berikut. Hal pertama yang dilakukan adalah menghitung *Term Frequency* (TF).

$$TF_{ij} = \frac{f_{ij}}{\max_k f_{kj}} \quad (3.1)$$

Dalam persamaan 3.1, f_{ij} merujuk pada frekuensi istilah i dalam dokumen j , sementara $\max_k f_{kj}$ mengacu pada frekuensi fitur paling umum atau fitur dengan frekuensi tertinggi dalam dokumen tersebut.

Tabel 3.7 Perhitungan Term Frequency

Dokumen	Jumlah Dokumen	Kata	Frekuensi	TF
D1	10	call	1	1/11
D2	3		0	0/3
D3	6		0	0/6
D4	3		1	1/3
D5	5		0	0/5

Selanjutnya dilakukan perhitungan Inverse Document Frequency untuk melakukan keseimbangan antara distribusi nilai. Ini dilakukan dengan rumus berikut:

$$IDF_i = \log\left(\frac{N}{n_i}\right) + 1 \quad (3.2)$$

Dalam persamaan 3.2, N merupakan jumlah total dokumen pelatihan yang digunakan dan n_i adalah jumlah dokumen latih yang mengandung istilah i .

Tabel 3. 8 Perhitungan IDF

Dokumen	Frekuensi	Kata	IDF
D1	1	call	$\log \frac{5}{2} + 1 = 1.39794$
D2	0		
D3	0		
D4	1		
D5	0		

TF-IDF dihitung sesuai dengan persamaan 3.3 seperti berikut :

$$TFIDF_{ij} = TF_{ij} \times IDF_i \quad (3.3)$$

Persamaan 3.3 menggabungkan nilai Term Frequency (Tf_{ij}) dan Inverse Document Frequency (IDF_i) untuk menghasilkan skor bobot untuk istilah i dalam dokumen j . Bobot TF-IDF digunakan menilai seberapa signifikan sebuah kata dalam sebuah dokumen relatif terhadap keseluruhan koleksi dokumen. Nilai TF mengukur seberapa sering sebuah kata muncul dalam sebuah dokumen, sementara IDF mengukur seberapa unik atau jarang kata tersebut muncul dalam keseluruhan korpus dokumen.

Tabel 3. 9 Perhitungan TF-IDF

Dokumen	Kata	TF	IDF	TF-IDF
D1	call	1/11	1.39794	0.12708
D2		0/3		0
D3		0/6		0
D4		1/3		0.46598
D5		0/5		0

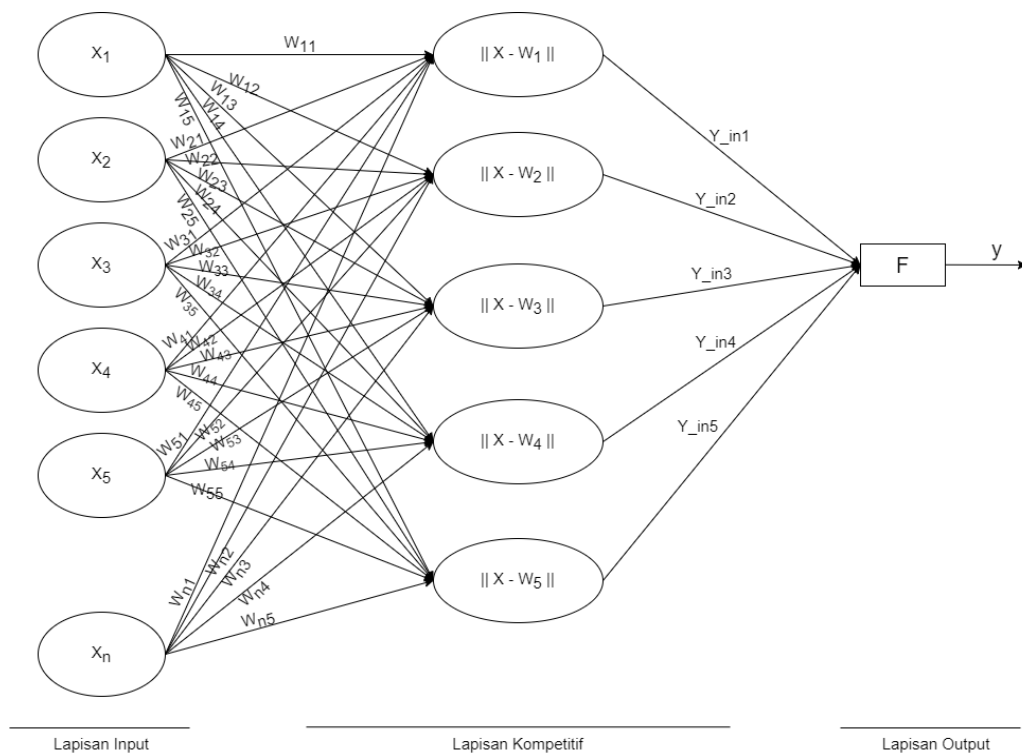
Dengan mengalikan TF dan IDF, kita mendapatkan bobot yang lebih tinggi untuk kata-kata yang muncul secara sering dalam sebuah dokumen tetapi jarang muncul dalam dokumen lainnya, dan sebaliknya. Proses ini membantu melihat kata-kata yang paling relevan dan signifikan dalam sebuah dokumen dalam konteks klasifikasi atau analisis teks. Selanjutnya, nilai TF-IDF ini dapat digunakan sebagai fitur untuk melatih model pembelajaran mesin atau digunakan klasifikasi dokumen. Penggunaan nilai TF-IDF sebagai fitur, model dapat dilatih untuk melakukan klasifikasi dokumen dengan lebih akurat dan efisien. Berikut ini merupakan beberapa hasil perhitungan TF-IDF dalam dokumen.

Tabel 3. 10 Hasil Perhitungan TF-IDF

Kata	Jumlah Kata				
	D1	D2	D3	D4	D5
fuck	0.15444	0	0	0	0
call	0.12708	0	0	0.46598	0
feel	0.15444	0	0	0	0
bad	0.15444	0	0	0	0
dumb	0.15444	0	0	0	0
nigger	0.15444	0	0	0	0
dox	0.15444	0	0	0	0
glad	0.15444	0	0	0	0
live	0.15444	0	0	0	0
michigan	0.15444	0	0	0	0
beat	0.15444	0	0	0	0
bully	0	0.5663	0	0	0
rude	0	0.5663	0	0	0
school	0	0.5663	0	0	0
destroy	0	0	0.28315	0	0
life	0	0	0.28315	0	0
welcom	0	0	0.28315	0	0
world	0	0	0.28315	0	0
islam	0	0	0.28315	0	0
terror	0	0	0.28315	0	0
female	0	0	0	0.42472	0
basic	0	0	0	0.42472	0
bitch	0	0	0	0.42472	0
start	0	0	0	0	0.33979
thursday	0	0	0	0	0.33979
class	0	0	0	0	0.33979
cancel	0	0	0	0	0.33979
tomorrow	0	0	0	0	0.33979

3.4.6 Learning vector quantization

Learning vector quantization yang diperkenalkan Kohonen adalah sebuah jaringan single layer yang terdiri dari lapisan input dan output, dimana terdapat bobot di antara kedua lapisan tersebut (Nugroho et al., 2018). Arsitektur jaringan metode Learning Vector Quantization dalam penelitian ini dapat ditemukan pada Gambar 3.6.

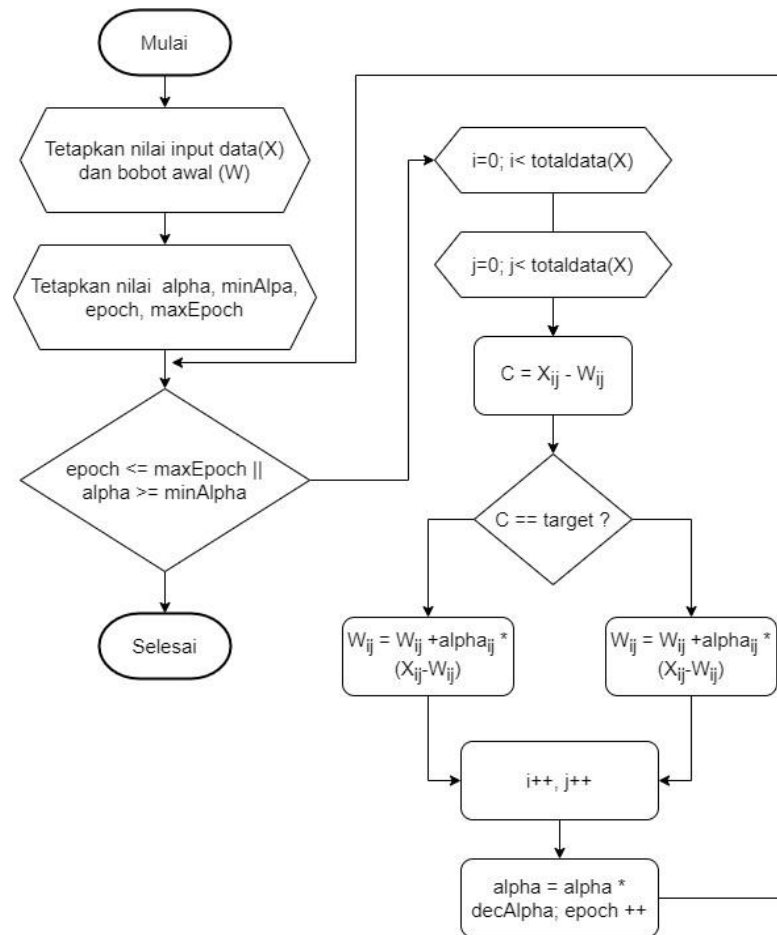


Gambar 3. 6 Arsitektur Jaringan LVQ

Pada Gambar 3.6 terlihat terdapat tiga lapisan, yaitu lapisan input, lapisan kompetitif, dan lapisan output. Lapisan input, di mana X_1 hingga X_n merupakan matriks vektor data dari hasil preprocessing yang mewakili tiap kelasnya masing-masing. Data tersebut kemudian diteruskan ke lapisan kompetitif untuk

dikelompokkan ke dalam kelas-kelas yang sesuai dan disesuaikan vektor bobotnya untuk tiap kelas (W). Pada lapisan ini, akan dibandingkan data input (X) dengan vektor bobot (W) dari setiap kelas, kemudian ditentukan kelas yang paling sesuai dengan data input tersebut. Kemudian $\| X - W_1 \|$ sampai dengan $\| X - W_5 \|$ merupakan perhitungan jarak bobot terkecil sesuai dengan rumus pada Nomor 3.1, dengan W_1 sampai dengan W_5 sebagai nilai vektor bobot kelas data yang diinisialisasi. Dalam proses training, bobot tersebut akan terus diperbarui sesuai dengan rumus Nomor 3.2 dan 3.3. Setelah itu, hasil klasifikasi akan ditentukan di lapisan output (F), yaitu node yang sesuai kelasnya. F diperoleh dari hasil perhitungan jarak yang terkecil. Sedangkan (y) merupakan representasi bobot akhir yang nantinya akan digunakan dalam proses pengujian. Setelah dilakukan proses training, LVQ dapat digunakan untuk pengenalan pola cyberbullying karena metode ini memiliki kemampuan dalam melakukan klasifikasi terhadap data yang kompleks. LVQ juga efektif dalam menangani variasi data yang tinggi, membuatnya cocok untuk aplikasi real-world seperti deteksi cyberbullying. Algoritma LVQ lengkap dapat dilihat pada Gambar 3.7, yang memperinci langkah-langkah dari inisialisasi bobot hingga pembaruan bobot selama training. Dengan pendekatan ini, sistem dapat belajar dari data yang beragam dan meningkatkan akurasi klasifikasi secara bertahap. Proses pembaruan bobot yang berkelanjutan memungkinkan LVQ untuk beradaptasi dengan perubahan pola data seiring waktu. Selain itu, LVQ mampu menangani data yang memiliki dimensi tinggi, yang sering ditemukan dalam analisis teks. Penerapan LVQ dalam deteksi cyberbullying dapat

membantu mengidentifikasi pola perilaku negatif secara lebih cepat dan akurat sehingga lebih efektif dalam pencegahan cyberbullying pada platform online.



Gambar 3. 7 Flowchart LVQ

Berikut ini merupakan Tahapan Algoritma serta contoh perhitungan dari metode

Learning vector quantization:

- Menetapkan data latih (x) dan data serta tetapkan bobot awal (w).
- Selanjutnya, tentukan nilai *alpha* (*learning rate*), *decrement alpha*, *minimum alpha*, dan *maximum epoch* untuk menetapkan batas komputasi.

- c. Proses perhitungan jarak akan dilakukan selama nilai *epoch* kurang dari atau sama dengan *maximum epoch*, atau nilai *alpha* masih lebih besar dari *minimum alpha*.
- d. Menghitung jarak antara nilai input data (x) dengan bobot (w) menggunakan euclidean distance, berikut adalah rumus dari euclidean distance:

$$\text{distance} = \sqrt{\sum_{i=0}^{n-1} (X_{ij} - W_{ij})^2} \quad (3.4)$$

- e. Setelah jarak dihitung, langkah selanjutnya adalah memeriksa target dari X_{ij} (data latih) dan target dari W_{ij} (kelas pemenang). Jika targetnya sama, maka bobot akan diperbarui. Pembaruan bobot dilakukan sesuai dengan rumus yang tercantum pada Nomor 3.2 sebagai berikut:

$$W_{ij} = W_{ij} + (\alpha(ij)) \cdot (X_{ij} - W_{ij}) \quad (3.5)$$

Jika target tidak sama dengan kelas pemenang, maka bobot akan diperbarui sesuai dengan rumus yang tercantum pada Nomor 3.3 sebagai berikut:

$$W_{ij} = W_{ij} - (\alpha(ij)) \cdot (X_{ij} - W_{ij}) \quad (3.6)$$

- f. Kemudian, perbarui nilai α (learning rate) menggunakan rumus yang sama seperti pada Nomor 3.4, dan tambahkan nilai *epoch*. Setelah itu, ulangi proses dari langkah ke-3.

$$\alpha = \alpha * \text{decrement alpha} \quad (3.7)$$

Berikut ini merupakan Tahapan Algoritma serta contoh perhitungan dari metode *Learning Vector Quantization*:

- a. Tetapkan data latih sesuai dengan tabel 3.6 dan data uji sesuai tabel 3.7. Nilai X_1 sampai nilai ke X_n adalah matriks yang menunjukkan jumlah kemunculan

kata-kata unik pada setiap data input yang mewakili satu kelas tertentu. Bobot awal untuk target 0 adalah (0.1, 0.2, 0.7, 0.8) dan bobot awal untuk target 1 adalah (0.3, 0.4, 0.5, 0.6).

Tabel 3. 11 Data Latih

No	1	2	3	4	Target
X ₁	1	0	0	1	0
X ₂	0	1	0	0	0
X ₃	0	0	1	0	1
X ₄	0	0	0	0	1

Tabel 3. 12 Data Uji

No	1	2	3	4
X ₁	0	1	1	0
X ₂	1	1	0	0

Iterasi ke 1:

Data ke 1 {1, 0, 0, 1} dengan target 0, bobot {{0.1, 0.2, 0.5, 0.6}, {0.3, 0.4, 0.7, 0.8}}, kemudian menghitung jarak dengan bobot tiap kelas:

Bobot target 0:

$$\sqrt{(1 - 0.1)^2 + (0 - 0.2)^2 + (0 - 0.5)^2 + (1 - 0.6)^2} = 1.12$$

Bobot target 1:

$$\sqrt{(1 - 0.3)^2 + (0 - 0.4)^2 + (0 - 0.7)^2 + (1 - 0.8)^2} = 1.18$$

Setelah didapatkan jarak kemudian cek target dari X_{ij} (data latih) dan target dari W_{ij} (kelas pemenang) jika targetnya sama maka perbaiki bobot. Karena target sama dengan kelas pemenang maka update bobot dilakukan sesuai dengan rumus pada Nomor 3.2 sebagai berikut:

$$W(1,1) = 0.1 + (0.01 * (1 - 0.1)) = 0.2$$

$$W(1,2) = 0.2 + (0.01 * (0 - 0.2)) = 0.2$$

$$W(1,3) = 0.5 + (0.01 * (0 - 0.5)) = 0.5$$

$$W(1,4) = 0.6 + (0.01 * (1 - 0.6)) = 0.6$$

Iterasi ke 2:

Data ke 2 {0, 1, 0, 0} dengan target 0, bobot {{0.2, 0.2, 0.5, 0.6}, {0.3, 0.4, 0.7, 0.8}}, kemudian menghitung jarak dengan bobot tiap kelas:

Bobot target 0:

$$\sqrt{(0 - 0.2)^2 + (1 - 0.2)^2 + (0 - 0.5)^2 + (0 - 0.6)^2} = 1.13$$

Bobot target 1:

$$\sqrt{(0 - 0.3)^2 + (1 - 0.4)^2 + (0 - 0.7)^2 + (0 - 0.8)^2} = 1.25$$

Berdasarkan data diatas dapat diketahui bahwa kelas pemenang yaitu pada kelas 0, karena target sama dengan kelas pemenang maka update bobot dilakukan sesuai dengan rumus pada Nomor 3.2 sebagai berikut:

$$W(2,1) = 0.2 + (0.01 * (0 - 0.2)) = 0.2$$

$$W(2,2) = 0.2 + (0.01 * (1 - 0.2)) = 0.21$$

$$W(2,3) = 0.5 + (0.01 * (0 - 0.5)) = 0.51$$

$$W(2,4) = 0.6 + (0.01 * (0 - 0.6)) = 0.6$$

Iterasi ke 3:

Data ke 3 {0, 0, 1, 0} dengan target 1, bobot {{0.2, 0.21, 0.51, 0.6}, {0.3, 0.4, 0.7, 0.8}}, kemudian menghitung jarak dengan bobot tiap kelas:

Bobot target 0:

$$\sqrt{(0 - 0.2)^2 + (0 - 0.21)^2 + (1 - 0.51)^2 + (0 - 0.6)^2} = 0.83$$

Bobot target 1:

$$\sqrt{(0 - 0.3)^2 + (0 - 0.4)^2 + (1 - 0.7)^2 + (0 - 0.8)^2} = 0.98$$

Berdasarkan data diatas dapat diketahui bahwa kelas pemenang yaitu pada kelas 0, karena target tidak sama dengan kelas pemenang maka update bobot dilakukan sesuai dengan rumus pada Nomor 3.3 sebagai berikut:

$$W(3,1) = 0.2 - (0.01 * (0 - 0.2)) = 0.2$$

$$W(3,2) = 0.21 - (0.01 * (0 - 0.21)) = 0.21$$

$$W(3,3) = 0.51 - (0.01 * (1 - 0.51)) = 0.51$$

$$W(3,4) = 0.6 - (0.01 * (0 - 0.6)) = 0.61$$

Iterasi ke 4

Data ke 4 {0, 0, 0, 1} dengan target 1, bobot {{0.2, 0.21, 0.51, 0.61}, {0.3, 0.4, 0.7, 0.8}}, kemudian menghitung jarak dengan bobot tiap kelas:

Bobot target 0:

$$\sqrt{(0 - 0.2)^2 + (0 - 0.21)^2 + (0 - 0.51)^2 + (1 - 0.61)^2} = 0.7$$

Bobot target 1:

$$\sqrt{(0 - 0.3)^2 + (0 - 0.4)^2 + (0 - 0.7)^2 + (1 - 0.8)^2} = 0.78$$

Berdasarkan data diatas dapat diketahui bahwa kelas pemenang yaitu pada kelas 0, karena target tidak sama dengan kelas pemenang maka update bobot dilakukan sesuai dengan rumus pada Nomor 3.3 sebagai berikut:

$$W(4,1) = 0.2 - (0.01 * (0 - 0.2)) = 0.2$$

$$W(4,2) = 0.21 - (0.01 * (0 - 0.21)) = 0.21$$

$$W(4,3) = 0.5 - (0.01 * (0 - 0.5)) = 0.52$$

$$W(4,4) = 0.61 - (0.01 * (1 - 0.61)) = 0.61$$

- b. Kemudian, perbarui nilai alpha (learning rate) menggunakan rumus yang sama seperti pada Nomor 3.4, dan tambahkan nilai epoch. Setelah itu, ulangi proses dari langkah ke-3

$$\text{alpha} = \text{alpha} * \text{decrement alpha} \quad (3.8)$$

Setelah semua data telah di hitung kemudian update nilai alpha dengan rumus seperti pada Nomor 3.4 dan nilai epoch di tambah 1

$$\text{alpha} = 0.1 * 0.1 = 0.01$$

$$\text{epoch} = 0 + 1 = 1$$

Dari proses tersebut nilai alpha telah sama dengan nilai minimum alpha dan nilai epoch juga sama dengan nilai maximum epoch, maka didapatkan bobot akhirnya sesuai pada Tabel 3.8

Tabel 3. 13 Data bobot akhir

No	1	2	3	4	Target
X ₁	0.2	0.21	0.52	0.61	0
X ₂	0.3	0.4	0.7	0.8	1

Setelah mendapatkan bobot akhir hasil pelatihan, langkah selanjutnya adalah memproses dengan data uji yang telah disiapkan sesuai dengan Tabel 3.7. Pada proses kali ini setiap data uji dilakukan perhitungan jarak dengan setiap nilai bobot menggunakan rumus *euclidean distance*.

Pengujian data uji ke 1:

Jarak data ke 1 {0, 1, 1, 0} dengan bobot target 1

$$\sqrt{(0 - 0.2)^2 + (1 - 0.21)^2 + (1 - 0.52)^2 + (0 - 0.61)^2} = 1.12$$

Jarak data ke 1 {0, 1, 1, 0} dengan bobot target 1

$$\sqrt{(0 - 0.31)^2 + (1 - 0.41)^2 + (1 - 0.5)^2 + (0 - 0.6)^2} = 1.08$$

Dari hasil uji di atas dapat kita ketahui bahwa data uji ke-1 masuk dalam klasifikasi target 1.

Pengujian data uji ke 2:

Jarak data ke 2 {1, 1, 0, 0} dengan bobot target 0

$$\sqrt{(1 - 0.2)^2 + (1 - 0.21)^2 + (0 - 0.52)^2 + (0 - 0.61)^2} = 1.38$$

Jarak data ke 2 {0, 0, 1, 1} dengan bobot target 1

$$\sqrt{(1 - 0.3)^2 + (1 - 0.4)^2 + (0 - 0.5)^2 + (0 - 0.6)^2} = 1.41$$

Dari hasil uji di atas dapat kita ketahui bahwa data uji ke-2 masuk dalam klasifikasi target 0.

BAB IV

UJI COBA DAN PEMBAHASAN

4.1 Skenario Uji Coba

Pada penelitian ini dilakukan beberapa tahapan uji coba terhadap sistem dengan teratur dan terstruktur. Tahap pertama yang dilakukan dengan membagi dataset menjadi data pelatihan dan pengujian. Kemudian untuk mendapatkan hasil klasifikasi *cyberbullying* menggunakan model LVQ dengan inputan berupa hasil dari TF-IDF. Hasil klasifikasi kemudian dihitung menggunakan *confusion matrix* untuk mendapatkan hasil akurasi, presisi, *recall*, dan *f1-score*. Selain itu, perhitungan juga dilakukan dengan menerapkan beberapa *learning rate* pada ratio split data yang digunakan untuk mengetahui akurasi tertinggi dapat dihasilkan pada *learning rate* beberapa. Untuk ratio data akan dibagi menjadi 3 model yaitu ratio data 60:40 masuk ke dalam uji coba Model A, ratio data 70:30 merupakan uji coba Model B, dan ratio data 80:20 masuk uji coba Model C.

Tabel 4. 1 Pembagian Dataset

Model	Data Latih	Data Uji
A	60%	40%
B	70%	30%
C	80%	20%

Dalam penelitian ini juga dilakukan pembagian learning rate pada masing-masing model rasio data. Learning rate yang digunakan adalah 0.001, 0.003, 0.01, 0.02. Penggunaan berbagai learning rate ini memungkinkan untuk mengeksplorasi bagaimana kinerja model bervariasi dengan perubahan learning rate. Selain itu, penggunaan learning rate yang berbeda juga membantu dalam menyesuaikan

proses pelatihan tergantung pada kompleksitas dataset atau karakteristik spesifik dari data yang digunakan dalam penelitian tersebut.

Tabel 4. 2 Pembagian Learning Rate

Nomor	Learning Rate
1.	0.1
2.	0.01
3.	0.001
4.	0.0001

Tahap pengujian bertujuan untuk mengevaluasi kinerja metode *Learning Vector Quantization* dalam melakukan klasifikasi teks komentar. Proses ini melibatkan perbandingan antara hasil klasifikasi oleh sistem dengan kategori sebenarnya dari data tiga kategori yang berbeda, dengan tujuan untuk menentukan nilai akurasi, presisi, recall, dan f-measure. Pada penelitian ini pengujian dilakukan dengan menggunakan confusion matrix. Confusion matrix memberikan gambaran mengenai hasil klasifikasi, sehingga dapat memahami sejauh mana model LVQ berhasil dalam membedakan antara kasus cyberbullying. Melalui confusion matrix dapat mengetahui berbagai metrik evaluasi yaitu accuracy, precision, recall, dan F1-score yang memberikan wawasan mendalam mengenai performa model. Hasil evaluasi ini dapat digunakan sebagai dasar untuk menyimpulkan bagaimana metode LVQ dengan ekstraksi fitur TF-IDF efektif dalam mengatasi masalah klasifikasi cyberbullying. Dengan pemahaman yang lebih baik mengenai kinerja model, penelitian ini dapat memberikan panduan untuk pengembangan sistem klasifikasi cyberbullying yang lebih efektif dan efisien di masa depan. Dalam penelitian ini menggunakan evaluasi confusion matrix dalam bentuk multiclass sebagai berikut.

		TP				
religion		0	0	FN	0	10
age	2	69	0	0	6	
gender	3	1	54	1	17	
	FP		TN			
ethnicity	2	0	0	74	3	
not bullying	3	0	0	2	61	
		religion	age	gender	ethnicity	not bullying
		Prediction				

Gambar 4. 1 Confusion Matrix Multiclass

- True Positive* (TP) adalah jumlah data yang sebenarnya bernilai positif dan diklasifikasikan sebagai positif.
- False Positive* (FP) adalah jumlah data yang sebenarnya bernilai negatif tetapi diklasifikasikan sebagai positif.
- False Negative* (FN) adalah jumlah data yang sebenarnya bernilai positif tetapi diklasifikasikan sebagai negatif.
- True Negative* (TN) adalah jumlah data yang sebenarnya bernilai negatif dan diklasifikasikan sebagai negatif. Untuk TN dapat dihitung dengan menjumlahkan semua data yang tidak termasuk dalam baris dan kolom kelas yang sedang dievaluasi.

Akurasi adalah seberapa dekat nilai prediksi dengan nilai sebenarnya. Ini adalah hasil dari membandingkan nilai uji dengan nilai aktual. Tingkat akurasi

biasanya dinyatakan dalam bentuk persentase, yaitu jumlah prediksi yang benar dibagi dengan total percobaan yang dilakukan. Tingkat akurasi dapat dihitung menggunakan persamaan 4.1

$$Akurasi_{kelas} = \frac{TP}{Jumlah\ Dokumen} 100\% \quad (4.1)$$

Precision adalah seberapa tepat informasi yang diminta oleh pengguna dengan hasil yang diberikan oleh sistem. Ini merupakan hasil persentase dari perbandingan nilai True Positif (TP) dengan total hasil yang diprediksi positif. Tingkat presisi dapat dihitung menggunakan persamaan 4.2.

$$Precision_{kelas} = \frac{TP}{TP + FP} 100\% \quad (4.2)$$

Recall adalah seberapa baik sistem dalam menemukan kembali informasi. Ini mengukur persentase dari perbandingan nilai True Positif (TP) dengan total data yang sebenarnya positif. Pengukuran recall dapat dihitung menggunakan persamaan 4.3.

$$Recall_{kelas} = \frac{TP}{TP + FN} 100\% \quad (4.3)$$

F1-Score adalah hasil dari evaluasi yang menggabungkan nilai recall dan presisi. Nilai F1-Score digunakan untuk menilai tingkat efektivitas suatu pengujian. Pengukuran F1-Score dapat dihitung menggunakan persamaan 4.4.

$$F1 - Score_{kelas} = 2 \frac{Precision \times Recall}{Precision + Recall} 100\% \quad (4.4)$$

Langkah selanjutnya adalah melakukan pengujian menggunakan metode k-fold cross validation. Teknik ini digunakan untuk mengukur seberapa akurat suatu model. Pelatihan dan pengujian dilakukan sebanyak K kali, di mana pada setiap iterasi, dataset ke-i digunakan sebagai testing set sedangkan dataset lainnya

digunakan sebagai training set. Setelah itu, rata-rata hasil dari setiap fold dihitung menggunakan persamaan 4.5, dan standar deviasi dihitung menggunakan persamaan 4.6.

$$\bar{x} = \frac{\sum x_i}{N} \quad (4.5)$$

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{N-1}} \quad (4.6)$$

4.1.1 Data Penelitian

Dataset yang digunakan pada peneliitian ini terdiri dari lima kelas yaitu Religion, Age, Gender, Ethnicity, dan Not Bullying. Masing-masing kelas terdiri dari 200 data. Berikut ini merupakan contoh dataset yang diambil dari Kaggle, dimana terdapat beberapa jenis *cyberbullying* yang berbeda, seperti yang terlihat pada Tabel 4.1

Tabel 4. 3 Sample Data Penelitian

No.	Komentar	Label
1.	@Raja5aab @Quickieleaks Yes, the test of god is that good or bad or indifferent or weird or whatever, it all proves gods existence.	not_cyberbullying
2.	I think @bxokrissy third period teacher doesn't like me he always tells me to go to class when I walk u to class	not_cyberbullying
3.	@GlennF @Spacekatgal it is rather funny how all of us are so completely different, yet complement each other's skill sets so well.	not_cyberbullying
4.	@TURBOCUNT i have no idea what red bean tastes like. i'm assuming it tastes like beans? i'm not really down with beans.	not_cyberbullying
5.	@doofsmack @wadhwa YUP. incidentally, I know the editor. She's a woman, a gamer, and in tech. It's all factual info.	not_cyberbullying
6.	Cant call her a bitch, cause a bitch is a female dog, and female dogs are loyal, which these hoes are not.	gender
7.	Men are empowered to call female politicians a "fucking bitch" when women speak of them hatefully and other women do nothing.	gender
8.	@ManhattaKnight I mean he's gay, but he uses gendered slurs and makes rape jokes	gender
9.	Idgaf if you are gay, lesbian, bisexual, or whatever the fuck you are, you don't fuck make a rape joke and think it's fucken ok	gender
10.	So I call you female Iâ€™m basically calling you a bitch ?	gender
11.	@officialuarh @sunniqaziboard you can see the mentality of idiots I want to see him jail rascal basterd.he is offering to kill muslim	religion
12.	U have destroyed your life. Welcome to the world of Islamic Terrorism.	religion

13.	@KellyTurner99 @buellerishere The Quran sanctifies sexual slavery. I only wish that you could have been one of these. http://t.co/MGAD8ICWYx	religion
14.	U SOLD OUT PRESS FOR WAHABI ISLAMIC TERRORISM OPENLY FUNDED BY QATAR WHICH UR NGO SHARES	religion
15.	@Sheikh_Times3 Nice of ISIS to get them together in one place so we can take out the trash.	religion
16.	This kid who called me a pig In high school referring to my weight and bullied this trans girl and like everyone else at the school too just dried my dads car at the carwash and I told dad who it was and he pulled \$5 out of his tip and offered to run him over aw #daddysgirl	age
17.	Oh haha you made a joke about your homophobic high school bully being gay how funny lol	age
18.	Here at home. Neighbors pick on my family and I. Mind you my son is autistic. It feels like high school. They call us names attack us for no reason and bully us all the time. Can't step on my front porch without them doing something to us	age
19.	i did go to international primary school but i never thought I am lucky because ameriken kids used to bully me by asking me to strip in the toilet because i was smol	age
20.	All the mean girls that bullied me in high school need to pay me reparations.	age
21.	Fuck u bitch RT @tayyoung_: FUCK OBAMA, dumb ass nigger	ethnicity
22.	That nigger food in the cafe today was disgusting	ethnicity
23.	@CJ_Jacobs oh look it's that dumb niggers ghost acct! Fuck you you dumb motherfucker!	ethnicity
24.	i wear white socks because there not nigger colored. #Racism #FuckStahlbaum	ethnicity
25.	@StevieJohnson13 Nice drop you nigger, guess you didn't know they play two halves you dumb fuck	ethnicity

Data yang telah dikumpulkan tersebut kemudian dilakukan proses preprocessing, pembobotan TF-IDF, pemodelan LVQ, dan pengujian kinerja model menggunakan *Confusion matrix*. Data tersebut digunakan sebagai data pelatihan yang kemudian dilakukan pemisahan data menjadi beberapa ratio data, yaitu 60:40, 70:30, dan 80:20. Data pelatihan yang digunakan lebih besar daripada data pengujian. Dataset yang digunakan dapat dilihat pada Tabel 4.2.

Tabel 4. 4 Jumlah data penelitian

Kelas	Jumlah
Religion	200 data
Age	200 data
Gender	200 data
Ethnicity	200 data
Not Bullying	200 data

Pembagian data pelatihan dan pengujian dilakukan dengan ratio 0.6, 0.7, dan 0.8 dengan cara melakukan inisialisasi variabel ratio. Pembagian dataset ini dilakukan dengan membagi antara training dan validation berdasarkan ratio. Variasi pembagian data yang berbeda pada penelitian ini dapat memberikan hasil akurasi yang optimal dengan ratio pembagian terbaik yang diperoleh. Pembagian data ratio training dan validation untuk masing-masing kelas dapat dilihat pada Tabel sebagai berikut.

Tabel 4. 5 Ratio Pembagian Data

Ratio	Training	Testing
0.6	600	400
0.7	700	300
0.8	800	200

Dalam penelitian ini, pendekatan pembagian dataset menjadi data pelatihan dan pengujian dilakukan dengan cermat. Rasio pembagian data yang digunakan adalah 60% untuk data pelatihan dan 40% untuk data validasi, 70% untuk data pelatihan dan 30% untuk data validasi, serta 80% untuk data pelatihan dan 20% untuk data uji. Tujuan dari variasi ini adalah untuk mengevaluasi performa model dengan berbagai proporsi data latih dan validasi. Selanjutnya, dalam skenario uji coba ini, fokus diberikan pada perubahan learning rate untuk menilai pengaruhnya terhadap akurasi model yang dihasilkan. Learning rate yang diujikan adalah 0.1, 0.01, 0.001, dan 0.0001. Seleksi parameter ini dilakukan berdasarkan serangkaian percobaan dan referensi jurnal yang diperoleh untuk mendapatkan parameter terbaik yang dapat memberikan hasil optimal. Selain itu, epoch yang digunakan dalam penelitian ini sebesar 100. Pada skenario uji coba ini juga dilakukan perhitungan menggunakan confusion matrix. Selanjutnya, pengujian dilakukan

menggunakan metode k-fold cross validation, yang digunakan untuk mengukur seberapa akurat sebuah model. Dalam konteks ini, metode tersebut digunakan untuk mengevaluasi model klasifikasi *cyberbullying* dengan menggunakan *Learning Vector Quantization*, dengan menggunakan dataset berjumlah 1000 dokumen untuk pelatihan dan pengujian. Pengujian dilakukan dengan K sebanyak 4, 5, dan 10-fold. Ilustrasi alur kerja *k-fold cross validation* dapat dilihat pada Gambar 4.1.

Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Fold-6	Fold-7	Fold-8	Fold-9	Fold-10
testing	training	training	Training	Training	Training	Training	Training	Training	Training
training	testing	training	Training	Training	Training	Training	Training	Training	Training
training	training	testing	Training	Training	Training	Training	Training	Training	Training
training	training	training	testing	Training	Training	Training	Training	Training	Training
training	training	training	training	testing	Training	Training	Training	Training	Training
training	training	training	training	Training	testing	Training	Training	Training	Training
training	training	training	training	Training	Training	testing	Training	Training	Training
training	training	training	training	Training	Training	training	testing	Training	Training
training	training	training	training	Training	training	training	Training	testing	Training
training	training	training	training	Training	Training	training	Training	Training	testing

Gambar 4. 2 Visualisasi metode 10-fold cross validation

4.2 Hasil Uji Coba

4.2.1 Uji Coba Model A

Hasil uji coba pada skenario Model A ini dilakukan dengan menerapkan ratio data 60:40. Dengan menggunakan variasi 4 parameter learning rate.



Gambar 4. 3 Confusion matrix 60:40 learning rate 0.1

Uji coba pertama dengan menggunakan learning rate sebesar 0.1 dan nilai epoch sebanyak 100. Sehingga diperoleh hasil dari pengujian tersebut menggunakan confusion matrix seperti pada Gambar 4.3. Hasil ini menunjukkan seberapa efektif model dalam mengklasifikasikan data dengan parameter yang telah ditentukan.

Tabel 4. 6 Tabel Confusion Matrix 60:40 Learning Rate 0.1

Prediksi \ Aktual	Religion	Age	Gender	Ethnicity	Not bullying
Religion	80	0	0	0	0
Age	17	62	1	0	0
Gender	28	0	43	9	0
Ethnicity	17	0	0	63	0
Not bullying	75	0	0	3	2

Dalam hasil confusion matrix, nilai True Positive (TP), True Negative (TN), False Negative (FN), dan False Positive (FP) diketahui melalui skema confusion matrix. Berikut perhitungan nilai TP, TN, FN, dan FP untuk kelas Religion:

$$TP = 80$$

$$TN = 62+43+63+2 = 170$$

$$FN = 0+0+0+0 = 0$$

$$FP = 17+28+17+75 = 137$$

Tabel 4. 7 Hasil Perhitungan TP, TN, FP, FN (60:40 Learning Rate 0.1)

Learning Rate	Kelas	TP	TN	FP	FN
0.1	Religion	80	170	137	0
	Age	62	188	0	18
	Gender	43	207	1	37
	Ethnicity	63	187	12	17
	Not bullying	2	248	2	80

Dari hasil prediksi, langkah selanjutnya adalah melakukan perhitungan untuk mendapatkan nilai akurasi, presisi, recall, dan f-measure untuk setiap kelas.

- Evaluasi Learning rate 0.1

$$Akurasi_{kelas} = \frac{80 + 62 + 43 + 63 + 2}{400} 100\% = 62\%$$

$$Precision_{(Religion)} = \frac{80}{80 + 137} 100\% = 37\%$$

$$Precision_{(Age)} = \frac{62}{62 + 0} 100\% = 100\%$$

$$Precision_{(Gender)} = \frac{43}{43 + 1} 100\% = 98\%$$

$$Precision_{(Ethnicity)} = \frac{63}{63 + 12} 100\% = 84\%$$

$$Precision_{(Not bullying)} = \frac{2}{2 + 2} 100\% = 100\%$$

$$Recall_{(Religion)} = \frac{80}{80 + 0} 100\% = 100\%$$

$$Recall_{(Age)} = \frac{62}{62 + 18} 100\% = 78\%$$

$$Recall_{(Gender)} = \frac{43}{43 + 37} 100\% = 54\%$$

$$Recall_{(Ethnicity)} = \frac{63}{63 + 17} 100\% = 79\%$$

$$Recall_{(Not bullying)} = \frac{2}{2 + 80} 100\% = 3\%$$

$$F1 - Score_{(Religion)} = 2 \frac{37 \times 100}{37 \times 100} 100\% = 54\%$$

$$F1 - Score_{(Age)} = 2 \frac{100 \times 78}{100 \times 78} 100\% = 87\%$$

$$F1 - Score_{(Gender)} = 2 \frac{98 \times 54}{98 \times 54} 100\% = 69\%$$

$$F1 - Score_{(Ethnicity)} = 2 \frac{84 \times 79}{84 \times 79} 100\% = 81\%$$

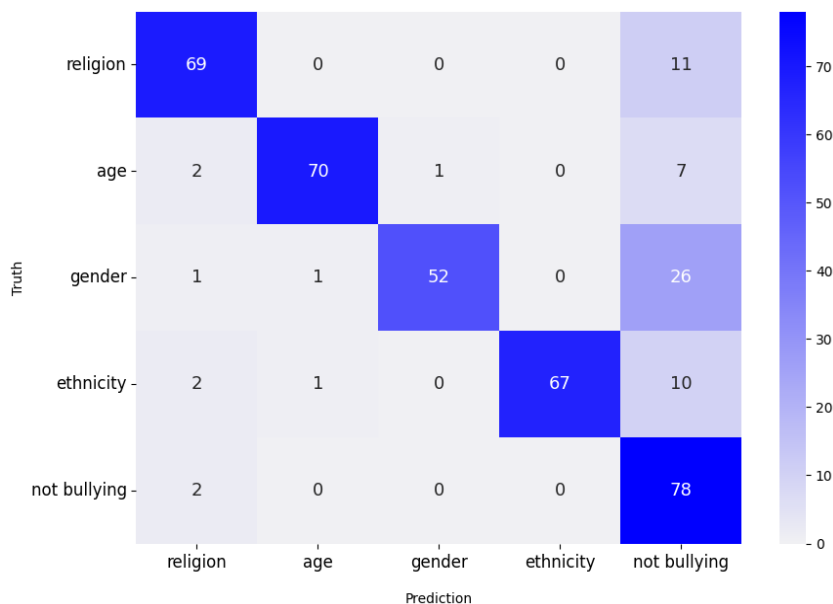
$$F1 - Score_{(Not bullying)} = 2 \frac{100 \times 3}{100 \times 3} 100\% = 5\%$$

Berdasarkan perhitungan pada Tabel 4.7 mampu menghasilkan akurasi sebesar 0.62 dan diperoleh nilai *precision*, *recall*, dan *fi-score* jika menggunakan parameter *learning rate* sebesar 0.1 yang dapat ditunjukkan pada Tabel 4.8. Dengan akurasi sebesar 0.62, model menunjukkan tingkat kesesuaian yang moderat dalam mengklasifikasikan data. Parameter *learning rate* yang digunakan mempengaruhi performa model, dengan nilai *precision*, *recall*, dan *f1-score* memberikan gambaran lebih lanjut tentang kemampuan model dalam mengidentifikasi kelas-kelas tertentu.

Tabel 4. 8 Nilai *Precision*, *Recall*, *F1-score learning rate* 0.1

Kelas	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>
Religion	0.37	1.00	0.54
Age	1.00	0.78	0.87
Gender	0.98	0.54	0.69
Ethnicity	0.84	0.79	0.81
Not bullying	1.00	0.03	0.05

Parameter pengujian selanjutnya yang digunakan adalah *learning rate* 0.01 dengan ratio 60:40 menghasilkan nilai *confusion matrix* seperti Gambar 4.4 berikut.



Gambar 4. 4 *Confusion matrix* 60:40 *learning rate* 0.01

Tabel 4. 9 Tabel Confusion Matrix 60:40 Learning Rate 0.01

Prediksi \ Aktual	Religion	Age	Gender	Ethnicity	Not bullying
Religion	69	0	0	0	11
Age	2	70	1	0	7
Gender	1	1	52	0	26
Ethnicity	2	1	0	67	10
Not bullying	2	0	0	0	78

Pada hasil tabel confusion matrix di atas dapat diketahui nilai aktual dan prediksi dari masing-masing kelas dengan menggunakan skema confusion matrix. Berikut merupakan hasil perhitungan nilai TP, TN, FN dan FP pada ratio 60:40 dan learning rate 0.01. Informasi ini memberikan pemahaman yang lebih mendalam tentang

kinerja model dalam melakukan klasifikasi pada skenario 60:40 dengan learning rate 0.01.

Tabel 4. 10 Hasil Perhitungan TP, TN, FP, FN (60:40 Learning Rate 0.01)

Learning Rate	Kelas	TP	TN	FP	FN
0.01	Religion	69	267	7	11
	Age	70	266	2	10
	Gender	52	284	1	28
	Ethnicity	67	269	0	13
	Not bullying	78	258	54	2

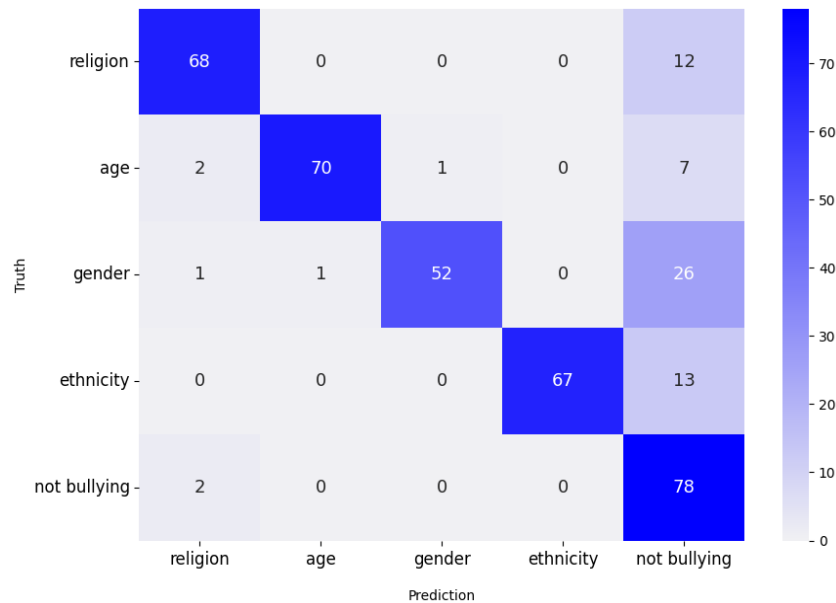
Berdasarkan tabel tersebut terdapat 69 TP yang menunjukkan jumlah prediksi yang benar untuk kelas Religion. Sebanyak 267 TN menunjukkan jumlah prediksi benar untuk data yang bukan kelas Religion. Sedangkan, 7 FP menunjukkan jumlah prediksi yang salah untuk kelas Religion, dan 11 FN menunjukkan jumlah prediksi yang salah untuk data yang bukan kelas Religion. 70 TP, menandakan jumlah prediksi yang benar untuk kelas Age. Terdapat 266 TN, yang menunjukkan jumlah prediksi benar untuk data yang bukan kelas Age. Hanya ada 2 FP, menunjukkan jumlah prediksi yang salah untuk kelas Age, dan 10 FN yang menunjukkan jumlah prediksi yang salah untuk data yang bukan kelas Age. Ditemukan 52 TP, yang menunjukkan jumlah prediksi yang benar untuk kelas Gender. Terdapat 284 TN, menandakan jumlah prediksi benar untuk data yang bukan kelas Gender. Hanya ada 1 FP, yang menunjukkan jumlah prediksi yang salah untuk kelas Gender, dan 28 FN yang menunjukkan jumlah prediksi yang salah untuk data yang bukan kelas Gender. Ditemukan 67 TP, yang menunjukkan jumlah prediksi yang benar untuk kelas Ethnicity. Terdapat 269 TN, menandakan jumlah prediksi benar untuk data yang bukan kelas Ethnicity. Tidak ada FP, menunjukkan tidak ada prediksi yang salah untuk kelas Ethnicity, dan 13 FN yang menunjukkan jumlah prediksi yang

salah untuk data yang bukan kelas Ethnicity. erdapat 78 TP, yang menunjukkan jumlah prediksi yang benar untuk kelas Not bullying. Sebanyak 258 TN, menandakan jumlah prediksi benar untuk data yang bukan kelas Not bullying. Ada 54 FP, menunjukkan jumlah prediksi yang salah untuk kelas Not bullying, dan hanya 2 FN yang menunjukkan jumlah prediksi yang salah untuk data yang bukan kelas Not bullying. Kelas Age menunjukkan keseimbangan baik antara TP, TN, FP, dan FN, sedangkan kelas Ethnicity menunjukkan tidak adanya FP namun FN yang signifikan, mengindikasikan perlu peningkatan dalam mendeteksi bullying berbasis etnis. Nilai akurasi yang diperoleh menggunakan parameter ini sebesar 0.84 dimana secara keseluruhan model dengan skenario pembagian data 60:40 dengan learning rate 0.01 telah cukup baik dalam mengklasifikasikan cyberbullying. Berdasarkan hasil perhitungan TP, TN, FP, FN pada Tabel 4.12 diperoleh nilai *precision*, *recall*, dan *fi-score* yang dapat ditunjukkan pada Tabel 4.11 berikut.

Tabel 4. 11 Nilai *Precision*, *Recall*, *F1-score* learning rate 0.01

Kelas	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>
Religion	0.91	0.86	0.88
Age	0.97	0.88	0.92
Gender	0.98	0.65	0.78
Ethnicity	1.00	0.84	0.91
Not bullying	0.59	0.97	0.74

Untuk langkah pengujian berikutnya, parameter yang digunakan adalah learning rate sebesar 0.001 dengan pembagian data 60:40. Hal ini bertujuan untuk mengeksplorasi dampak dari pengaturan learning rate yang lebih rendah terhadap kinerja model sehingga menghasilkan nilai *confusion matrix* seperti pada Gambar 4.5 berikut.



Gambar 4. 5 Confusion Matrix 60:40 Learning Rate 0.001

Tabel 4. 12 Tabel Confusion Matrix 60:40 Learning Rate 0.001

Prediksi \ Aktual	Religion	Age	Gender	Ethnicity	Not bullying
Religion	68	0	0	0	12
Age	2	70	1	0	7
Gender	1	1	52	0	26
Ethnicity	0	0	0	67	13
Not bullying	2	0	0	0	78

Pada hasil tabel confusion matrix di atas dapat diketahui nilai aktual dan prediksi dari masing-masing kelas. Berikut hasil perhitungan nilai TP, TN, FN dan FP pada ratio 60:40 dan learning rate 0.001 berdasarkan tabel confusion matrix sebelumnya.

Tabel 4. 13 Hasil Perhitungan TP, TN, FP, FN (60:40 Learning Rate 0.001)

Learning Rate	Kelas	TP	TN	FP	FN
0.001	Religion	68	267	5	12
	Age	70	265	1	10
	Gender	52	283	1	28
	Ethnicity	67	268	0	13
	Not bullying	78	257	58	2

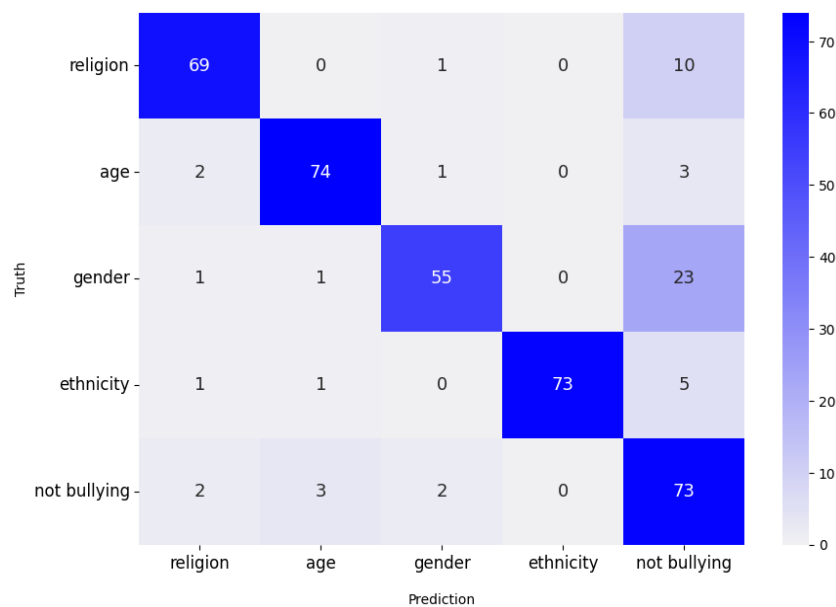
Terdapat 68 TP, menunjukkan jumlah prediksi yang benar untuk kelas Religion. Sebanyak 267 TN, menunjukkan jumlah prediksi benar untuk data yang bukan kelas Religion. Ada 5 FP, menunjukkan jumlah prediksi yang salah untuk kelas Religion, dan 12 FN, menunjukkan jumlah prediksi yang salah untuk data yang bukan kelas Religion. Ditemukan 70 TP, menandakan jumlah prediksi yang benar untuk kelas Age. Terdapat 265 TN, menunjukkan jumlah prediksi benar untuk data yang bukan kelas Age. Hanya ada 1 FP, menunjukkan jumlah prediksi yang salah untuk kelas Age, dan 10 FN, menunjukkan jumlah prediksi yang salah untuk data yang bukan kelas Age. Terdapat 52 TP, yang menunjukkan jumlah prediksi yang benar untuk kelas Gender. Terdapat 283 TN, menunjukkan jumlah prediksi benar untuk data yang bukan kelas Gender. Ada 1 FP, menunjukkan jumlah prediksi yang salah untuk kelas Gender, dan 28 FN, menunjukkan jumlah prediksi yang salah untuk data yang bukan kelas Gender. Ditemukan 67 TP, menandakan jumlah prediksi yang benar untuk kelas Ethnicity. Terdapat 268 TN, menunjukkan jumlah prediksi benar untuk data yang bukan kelas Ethnicity. Tidak ada FP, menunjukkan tidak ada prediksi yang salah untuk kelas Ethnicity, dan 13 FN, menunjukkan jumlah prediksi yang salah untuk data yang bukan kelas Ethnicity. Terdapat 78 TP, yang menunjukkan jumlah prediksi yang benar untuk kelas Not bullying. Sebanyak 257 TN, menunjukkan jumlah prediksi benar untuk data yang bukan kelas Not bullying. Ada 58 FP, menunjukkan jumlah prediksi yang salah untuk kelas Not bullying, dan hanya 2 FN, menunjukkan jumlah prediksi yang salah untuk data yang bukan kelas Not bullying. Nilai akurasi yang diperoleh menggunakan parameter ini

sebesar 0.838. Kemudian diperoleh nilai *precision*, *recall*, dan *fi-score* yang dapat ditunjukkan pada Tabel 4.14 berikut.

Tabel 4. 14 Nilai *Precision*, *Recall*, *F1-score* learning rate 0.001

Kelas	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>
Religion	0.93	0.85	0.89
Age	0.99	0.88	0.93
Gender	0.98	0.65	0.78
Ethnicity	1.00	0.84	0.91
Not bullying	0.57	0.97	0.72

Hasil klasifikasi cyberbullying dengan menggunakan parameter ratio data 60:40 dan learning rate sebesar 0.0001 terdokumentasi dalam bagian 4.6. Dalam pengujian ini, data dibagi dalam rasio 60:40 antara data latih dan data uji, dan learning rate yang digunakan adalah 0.0001. Hasil dari klasifikasi ini memberikan gambaran tentang kemampuan model dalam mengenali kasus cyberbullying.



Gambar 4. 6 *Confusion matrix* 60:40 learning rate 0.0001

Tabel 4. 15 Tabel Confusion Matrix 60:40 Learning Rate 0.0001

Prediksi Aktual	Religion	Age	Gender	Ethnicity	Not bullying
Religion	69	0	1	0	10
Age	2	74	1	0	3
Gender	1	1	55	0	23
Ethnicity	1	1	0	73	5
Not bullying	2	3	2	0	73

Dengan menggunakan skema confusion matrix, nilai True Positive (TP), True Negative (TN), False Negative (FN), dan False Positive (FP) dapat dihitung dari hasil tabel confusion matrix. Berikut hasil perhitungan TP, TN, FN dan FP pada pengujian dengan rasio data 60:40 dan learning rate sebesar 0.0001.

Tabel 4. 16 Hasil Perhitungan TP, TN, FP, FN (60:40 Learning Rate 0.0001)

Learning Rate	Kelas	TP	TN	FP	FN
0.0001	Religion	69	275	6	11
	Age	74	270	5	6
	Gender	55	289	4	25
	Ethnicity	73	271	0	7
	Not bullying	73	271	41	7

Tabel menunjukkan 69 TP, menunjukkan jumlah prediksi yang benar untuk kelas Religion. Sebanyak 275 TN menunjukkan jumlah prediksi benar untuk data yang bukan kelas Religion. Ada 6 FP, menunjukkan jumlah prediksi yang salah untuk kelas Religion, dan 11 FN, menunjukkan jumlah prediksi yang salah untuk data yang bukan kelas Religion. Ditemukan 74 TP, menandakan jumlah prediksi yang benar untuk kelas Age. Terdapat 270 TN, menunjukkan jumlah prediksi benar untuk data yang bukan kelas Age. Hanya ada 5 FP, menunjukkan jumlah prediksi yang salah untuk kelas Age, dan 6 FN, menunjukkan jumlah prediksi yang salah untuk data yang bukan kelas Age. Terdapat 55 TP, yang menunjukkan jumlah prediksi yang benar untuk kelas Gender. Terdapat 289 TN, menunjukkan jumlah

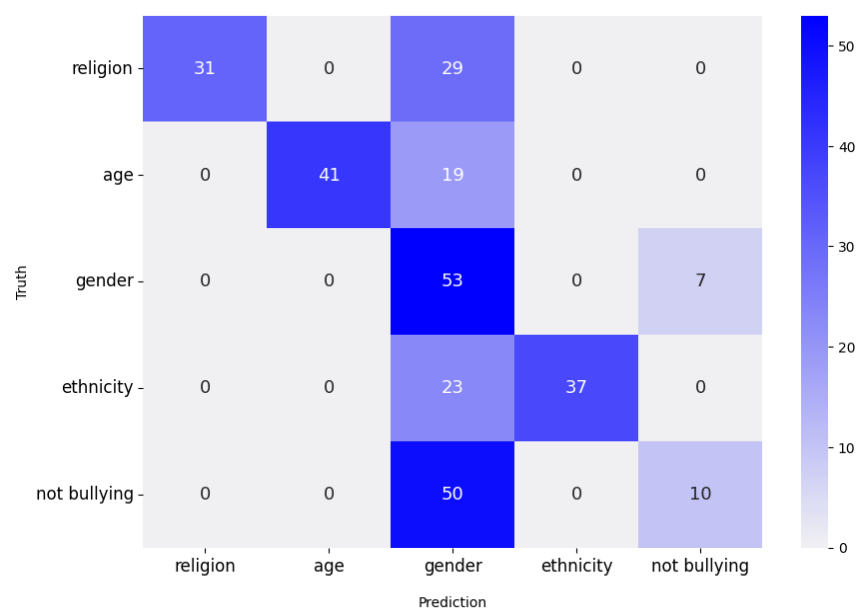
prediksi benar untuk data yang bukan kelas Gender. Ada 4 FP, menunjukkan jumlah prediksi yang salah untuk kelas Gender, dan 25 FN, menunjukkan jumlah prediksi yang salah untuk data yang bukan kelas Gender. Ditemukan 73 TP, menandakan jumlah prediksi yang benar untuk kelas Ethnicity. Terdapat 271 TN, menunjukkan jumlah prediksi benar untuk data yang bukan kelas Ethnicity. Tidak ada FP, menunjukkan tidak ada prediksi yang salah untuk kelas Ethnicity, dan 7 FN, menunjukkan jumlah prediksi yang salah untuk data yang bukan kelas Ethnicity. Terdapat 73 TP, yang menunjukkan jumlah prediksi yang benar untuk kelas Not bullying. Sebanyak 271 TN, menunjukkan jumlah prediksi benar untuk data yang bukan kelas Not bullying. Ada 41 FP, menunjukkan jumlah prediksi yang salah untuk kelas Not bullying, dan hanya 7 FN, menunjukkan jumlah prediksi yang salah untuk data yang bukan kelas Not bullying. Tabel menunjukkan hasil dari confusion matrix dengan detail TP, TN, FP, dan FN untuk masing-masing kelas, serta akurasi keseluruhan sebesar 0.86 menggunakan learning rate 0.0001. Hasil perhitungan untuk precision, recall, dan F1-score dapat ditemukan pada Tabel 4.17 untuk memberikan gambaran lebih lanjut tentang kinerja model dalam melakukan klasifikasi berdasarkan masing-masing metrik evaluasi tersebut.

Tabel 4. 17 Nilai *Precision*, *Recall*, *F1-score* learning rate 0.0001

Kelas	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>
Religion	0.92	0.86	0.89
Age	0.94	0.93	0.93
Gender	0.93	0.69	0.79
Ethnicity	1.00	0.91	0.95
Not bullying	0.64	0.91	0.75

4.2.2 Uji Coba Model B

Skenario uji coba Model B kedua menggunakan ratio data 70:30 dengan *epochs* 100. Dilakukan dengan variasi 4 learning rate yaitu 0.1, 0.01, 0.001, 0.0001. Pengujian pertama menerapkan parameter *learning rate* 0.1 ditunjukkan pada Gambar 4.7.



Gambar 4. 7 Confusion matrix 70:30 learning rate 0.1

Tabel 4. 18 Tabel Confusion Matrix 70:30 Learning rate 0.1

Prediksi \ Aktual	Religion	Age	Gender	Ethnicity	Not bullying
Religion	31	0	29	0	0
Age	0	41	19	0	0
Gender	0	0	53	0	7
Ethnicity	0	0	23	37	0
Not bullying	0	0	50	0	10

Pada hasil tabel confusion matrix di atas dapat diketahui nilai-nilai aktual dan prediksi dari masing-masing kelas dengan menggunakan skema confusion matrix.

Berikut merupakan hasil perhitungan nilai TP, TN, FN dan FP pada ratio 70:30 dan learning rate 0.1.

Tabel 4. 19 Hasil Perhitungan TP, TN, FP, FN (70:30 Learning rate 0.1)

Learning Rate	Kelas	TP	TN	FP	FN
0.1	Religion	31	141	0	29
	Age	41	131	0	19
	Gender	53	119	121	7
	Ethnicity	37	135	0	23
	Not bullying	10	162	7	50

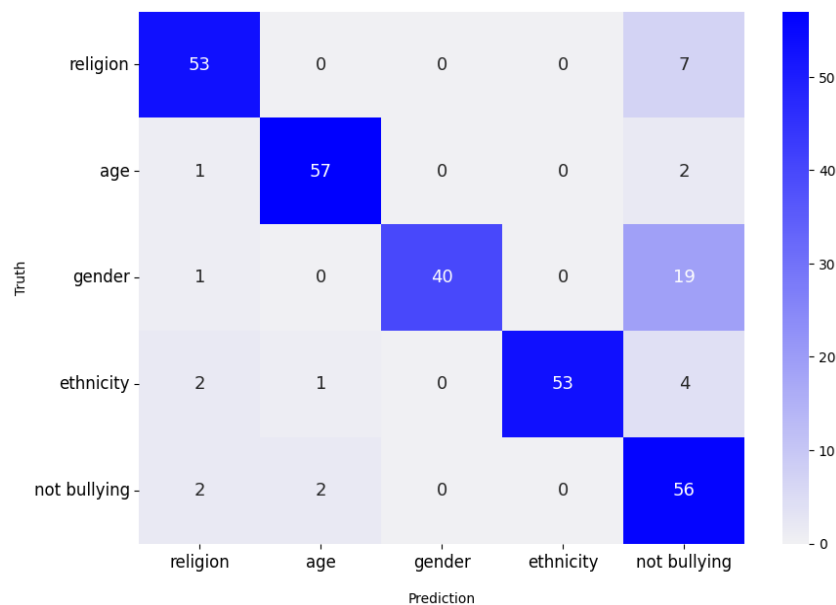
Pada tabel terdapat 31 TP, menunjukkan jumlah prediksi yang benar untuk kelas Religion. Sebanyak 141 TN menunjukkan jumlah prediksi benar untuk data yang bukan kelas Religion. Tidak ada FP, menunjukkan tidak ada prediksi yang salah untuk kelas Religion, dan 29 FN, menunjukkan jumlah prediksi yang salah untuk data yang bukan kelas Religion. Ditemukan 41 TP, menandakan jumlah prediksi yang benar untuk kelas Age. Terdapat 131 TN, menunjukkan jumlah prediksi benar untuk data yang bukan kelas Age. Tidak ada FP, menunjukkan tidak ada prediksi yang salah untuk kelas Age, dan 19 FN, menunjukkan jumlah prediksi yang salah untuk data yang bukan kelas Age. Terdapat 53 TP, yang menunjukkan jumlah prediksi yang benar untuk kelas Gender. Sebanyak 119 TN, menunjukkan jumlah prediksi benar untuk data yang bukan kelas Gender. Ada 121 FP, menunjukkan jumlah prediksi yang salah untuk kelas Gender, dan 7 FN, menunjukkan prediksi yang salah untuk data yang bukan kelas Gender. Ditemukan 37 TP, menandakan prediksi yang benar untuk kelas Ethnicity. Terdapat 135 TN, menunjukkan prediksi benar untuk data yang bukan kelas Ethnicity. Tidak ada FP, menunjukkan tidak ada prediksi yang salah untuk kelas Ethnicity, dan 23 FN, menunjukkan prediksi yang salah untuk data yang bukan kelas Ethnicity. Terdapat 10 TP, yang menunjukkan

prediksi yang benar untuk kelas Not bullying. Sebanyak 162 TN, menunjukkan prediksi benar untuk data yang bukan kelas Not bullying. Ada 7 FP, menunjukkan prediksi yang salah untuk kelas Not bullying, dan 50 FN, menunjukkan prediksi yang salah untuk data yang bukan kelas Not bullying. Hasil perhitungan TP, TN, FP dan FN yang diperoleh dari *learning rate* 0.1 yang digunakan menghasilkan tingkat akurasi sebesar 0.573. Hasil perhitungan untuk *precision*, *recall*, dan *f1-score* berdasarkan hasil tersebut ditunjukkan pada Tabel 4.20.

Tabel 4. 20 Nilai *Precision*, *Recall*, *F1-score* *learning rate* 0.1

Kelas	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>
Religion	1.00	0.52	0.68
Age	1.00	0.68	0.81
Gender	0.30	0.88	0.45
Ethnicity	1.00	0.62	0.76
Not bullying	0.59	0.17	0.26

Parameter *learning rate* selanjutnya pada 0.01, digunakan untuk pengujian pada ratio 70:30. Hasil dari pengujian ini direpresentasikan pada Gambar 4.8.



Gambar 4. 8 *Confusion matrix* 70:30 *learning rate* 0.01

Tabel 4. 21 Tabel Confusion Matrix 70:30 Learning rate 0.01

Prediksi Aktual	Religion	Age	Gender	Ethnicity	Not bullying
Religion	53	0	0	0	7
Age	1	57	0	0	2
Gender	1	0	40	0	19
Ethnicity	2	1	0	53	4
Not bullying	2	2	0	0	56

Pada hasil tabel confusion matrix di atas dapat diketahui nilai-nilai aktual dan prediksi dari masing-masing kelas dengan menggunakan skema confusion matrix. Berikut merupakan hasil perhitungan nilai TP, TN, FN dan FP pada ratio 70:30 dan learning rate 0.01.

Tabel 4. 22 Hasil Perhitungan TP, TN, FP, FN (70:30 Learning rate 0.01)

Learning Rate	Kelas	TP	TN	FP	FN
0.01	Religion	53	206	6	7
	Age	57	202	3	3
	Gender	40	219	0	20
	Ethnicity	53	206	0	7
	Not bullying	56	203	32	4

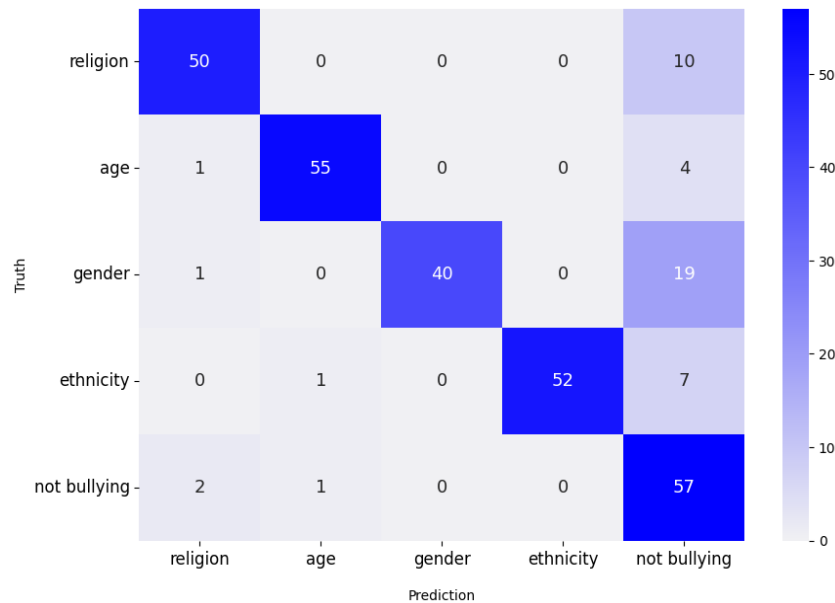
Dari Tabel 4.22 Terdapat 53 TP, menunjukkan jumlah prediksi yang benar untuk kelas Religion. Sebanyak 206 TN menunjukkan jumlah prediksi benar untuk data yang bukan kelas Religion. Ada 6 FP, menunjukkan jumlah prediksi yang salah untuk kelas Religion, dan 7 FN, menunjukkan jumlah prediksi yang salah untuk data yang bukan kelas Religion. Ditemukan 57 TP, menandakan jumlah prediksi yang benar untuk kelas Age. Terdapat 202 TN, menunjukkan jumlah prediksi benar untuk data yang bukan kelas Age. Hanya ada 3 FP, menunjukkan jumlah prediksi yang salah untuk kelas Age, dan 3 FN, menunjukkan jumlah prediksi yang salah untuk data yang bukan kelas Age. Terdapat 40 TP, yang menunjukkan jumlah prediksi yang benar untuk kelas Gender. Terdapat 219 TN, menunjukkan jumlah

prediksi benar untuk data yang bukan kelas Gender. Tidak ada FP, menunjukkan tidak ada prediksi yang salah untuk kelas Gender, dan 20 FN, menunjukkan jumlah prediksi yang salah untuk data yang bukan kelas Gender. Ditemukan 53 TP, menandakan jumlah prediksi yang benar untuk kelas Ethnicity. Terdapat 206 TN, menunjukkan jumlah prediksi benar untuk data yang bukan kelas Ethnicity. Tidak ada FP, menunjukkan tidak ada prediksi yang salah untuk kelas Ethnicity, dan 7 FN, menunjukkan jumlah prediksi yang salah untuk data yang bukan kelas Ethnicity. Terdapat 56 TP, yang menunjukkan jumlah prediksi yang benar untuk kelas Not bullying. Sebanyak 203 TN, menunjukkan jumlah prediksi benar untuk data yang bukan kelas Not bullying. Ada 32 FP, menunjukkan jumlah prediksi yang salah untuk kelas Not bullying, dan 4 FN, menunjukkan jumlah prediksi yang salah untuk data yang bukan kelas Not bullying. Berdasarkan Tabel 4.22 diperoleh hasil akurasi sebesar 0.863. Selain itu, berikut ini merupakan nilai *precision*, *recall*, dan *fi-score* diperoleh berdasarkan hasil perhitungan nilai TP, TN, FP dan FN diatas yang dapat ditunjukkan pada Tabel 4.23.

Tabel 4. 23 Nilai *Precision*, *Recall*, *F1-score* learning rate 0.01

Kelas	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>
Religion	0.90	0.88	0.89
Age	0.95	0.95	0.95
Gender	1.00	0.67	0.80
Ethnicity	1.00	0.88	0.94
Not bullying	0.64	0.93	0.76

Pengujian selanjutnya dilakukan dengan menggunakan parameter learning rate sebesar 0.001, dengan rasio pembagian data sebesar 70:30 antara data latih dan data uji. Hasil dari pengujian ini direpresentasikan dalam confusion matrix yang terdokumentasi pada Gambar 4.9.



Gambar 4. 9 Confusion matrix 70:30 learning rate 0.001

Tabel 4. 24 Tabel Confusion Matrix 70:30 Learning rate 0.001

Prediksi \ Aktual	Religion	Age	Gender	Ethnicity	Not bullying
Religion	50	0	0	0	10
Age	1	55	0	0	4
Gender	1	0	40	0	19
Ethnicity	0	1	0	52	7
Not bullying	2	1	0	0	57

Pada hasil tabel confusion matrix di atas dapat diketahui nilai nilai nilai aktual dan prediksi dari masing-masing kelas dengan menggunakan skema confusion matrix. Berikut merupakan hasil perhitungan nilai TP, TN, FN dan FP pada ratio 70:30 dan menggunakan learning rate 0.001.

Tabel 4. 25 Hasil Perhitungan TP, TN, FP, FN (70:30 Learning rate 0.001)

Learning Rate	Kelas	TP	TN	FP	FN
0.001	Religion	50	204	4	10
	Age	55	199	2	5
	Gender	40	214	0	20
	Ethnicity	52	202	0	8
	Not bullying	57	197	40	3

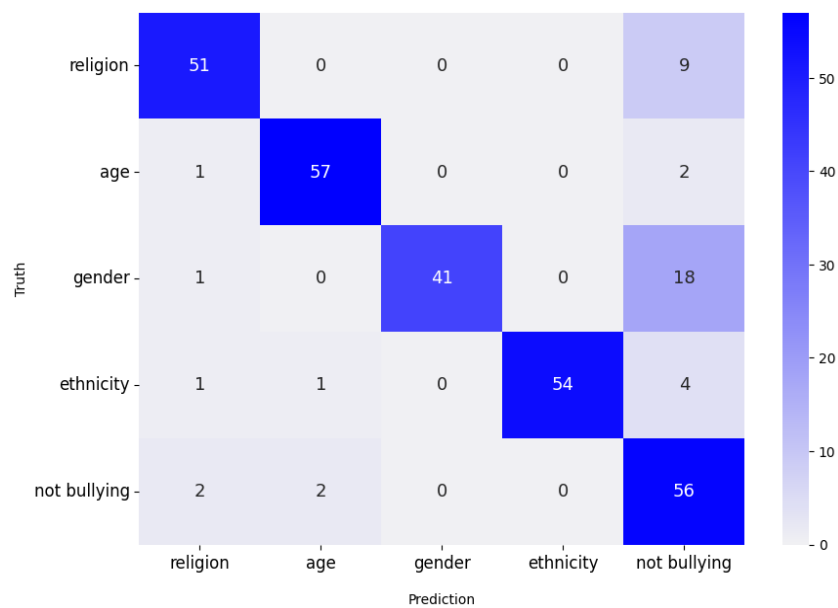
Berdasarkan Tabel 4.25, nilai 50 TP, menunjukkan jumlah prediksi yang benar untuk kelas Religion. Sebanyak 204 TN menunjukkan jumlah prediksi benar untuk data yang bukan kelas Religion. Ada 4 FP, menunjukkan jumlah prediksi yang salah untuk kelas Religion, dan 10 FN, menunjukkan jumlah prediksi yang salah untuk data yang bukan kelas Religion. 55 TP, menandakan jumlah prediksi yang benar untuk kelas Age. Terdapat 199 TN, menunjukkan jumlah prediksi benar untuk data yang bukan kelas Age. Hanya ada 2 FP, menunjukkan jumlah prediksi yang salah untuk kelas Age, dan 5 FN, menunjukkan jumlah prediksi yang salah untuk data yang bukan kelas Age. 40 TP, yang menunjukkan jumlah prediksi yang benar untuk kelas Gender. Terdapat 214 TN, menunjukkan jumlah prediksi benar untuk data yang bukan kelas Gender. Tidak ada FP, menunjukkan tidak ada prediksi yang salah untuk kelas Gender, dan 20 FN, menunjukkan jumlah prediksi yang salah untuk data yang bukan kelas Gender. 52 TP, menandakan jumlah prediksi yang benar untuk kelas Ethnicity. Terdapat 202 TN, menunjukkan jumlah prediksi benar untuk data yang bukan kelas Ethnicity. Tidak ada FP, menunjukkan tidak ada prediksi yang salah untuk kelas Ethnicity, dan 8 FN, menunjukkan jumlah prediksi yang salah untuk data yang bukan kelas Ethnicity. Terdapat 57 TP, yang menunjukkan jumlah prediksi yang benar untuk kelas Not bullying. Sebanyak 197 TN, menunjukkan jumlah prediksi benar untuk data yang bukan kelas Not bullying. Ada 40 FP, menunjukkan jumlah prediksi yang salah untuk kelas Not bullying, dan 3 FN, menunjukkan jumlah prediksi yang salah untuk data yang bukan kelas Not bullying. Hasil perhitungan yang ditunjukkan oleh Tabel 4.25 diperoleh hasil

akurasi sebesar 0.847 dengan nilai *precision*, *recall*, dan *fi-score* untuk masing-masing kelas seperti ditunjukkan pada Tabel 4.26.

Tabel 4. 26 Nilai *Precision*, *Recall*, *F1-score* learning rate 0.001

Kelas	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>
Religion	0.93	0.83	0.88
Age	0.96	0.92	0.94
Gender	1.00	0.67	0.80
Ethnicity	1.00	0.87	0.93
Not bullying	0.59	0.95	0.73

Selanjutnya pengujian dilakukan dengan mengubah parameter *learning rate* menjadi 0.0001 untuk ratio data 70:30 yang ditunjukkan pada Gambar 4.10 berikut.



Gambar 4. 10 *Confusion matrix* 70:30 learning rate 0.0001

Tabel 4. 27 Tabel *Confusion Matrix* 70:30 Learning rate 0.0001

Prediksi \ Aktual	Religion	Age	Gender	Ethnicity	Not bullying
Religion	51	0	0	0	9
Age	1	57	0	0	2
Gender	1	0	41	0	18
Ethnicity	1	1	0	54	4
Not bullying	2	2	0	0	56

Pada hasil tabel confusion matrix di atas dapat diketahui nilai-nilai aktual dan prediksi dari masing-masing kelas dengan menggunakan skema confusion matrix.

Berikut merupakan hasil perhitungan nilai TP, TN, FN dan FP kasus tersebut.

Tabel 4. 28 Hasil Perhitungan TP, TN, FP, FN (70:30 Learning rate 0.0001)

Learning Rate	Kelas	TP	TN	FP	FN
0.0001	Religion	51	208	5	9
	Age	57	202	3	3
	Gender	41	218	0	19
	Ethnicity	54	205	0	6
	Not bullying	56	203	33	4

Berdasarkan Tabel 4.28 Terdapat 51 TP, menunjukkan jumlah prediksi yang benar untuk kelas Religion. Sebanyak 208 TN menunjukkan jumlah prediksi benar untuk data yang bukan kelas Religion. Ada 5 FP, menunjukkan jumlah prediksi yang salah untuk kelas Religion, dan 9 FN, menunjukkan jumlah prediksi yang salah untuk data yang bukan kelas Religion. Ditemukan 57 TP, menandakan jumlah prediksi yang benar untuk kelas Age. Terdapat 202 TN, menunjukkan jumlah prediksi benar untuk data yang bukan kelas Age. Hanya ada 3 FP, menunjukkan jumlah prediksi yang salah untuk kelas Age, dan 3 FN, menunjukkan jumlah prediksi yang salah untuk data yang bukan kelas Age. Terdapat 41 TP, yang menunjukkan jumlah prediksi yang benar untuk kelas Gender. Terdapat 218 TN, menunjukkan jumlah prediksi benar untuk data yang bukan kelas Gender. Tidak ada FP, menunjukkan tidak ada prediksi yang salah untuk kelas Gender, dan 19 FN, menunjukkan jumlah prediksi yang salah untuk data yang bukan kelas Gender. Ditemukan 54 TP, menandakan jumlah prediksi yang benar untuk kelas Ethnicity. Tidak ada FP, menunjukkan tidak ada prediksi yang salah untuk kelas Ethnicity, dan 6 FN, menunjukkan jumlah prediksi yang salah untuk data yang bukan kelas Ethnicity.

Terdapat 56 TP, yang menunjukkan jumlah prediksi yang benar untuk kelas Not bullying. Ada 33 FP, menunjukkan jumlah prediksi yang salah untuk kelas Not bullying, dan 4 FN, menunjukkan jumlah prediksi yang salah untuk data yang bukan kelas Not bullying. Hasil perhitungan TP, TN, FP dan FN yang ditunjukkan oleh Tabel 4.30 diperoleh hasil akurasi sebesar 0.86 dengan nilai *precision*, *recall*, dan *fi-score* seperti ditunjukkan pada Tabel 4.29.

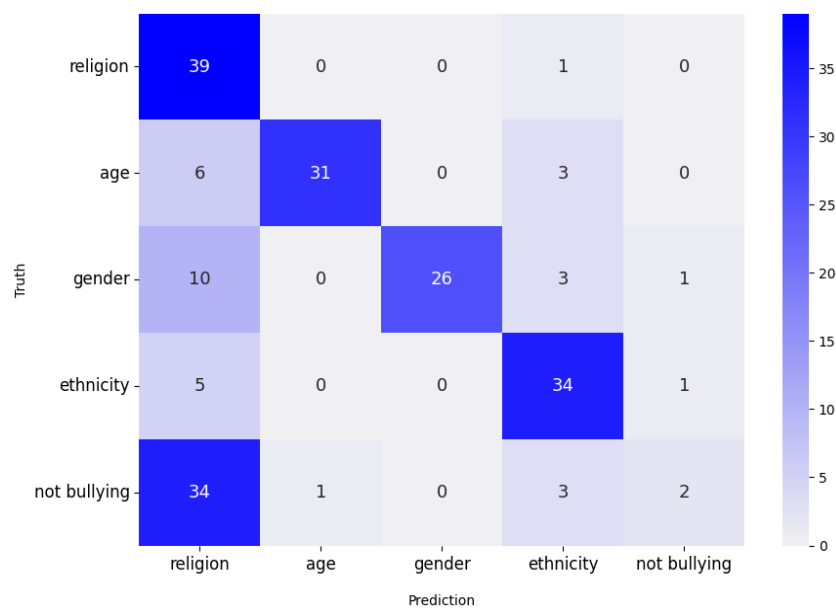
Tabel 4. 29 Nilai *Precision*, *Recall*, *F1-score* learning rate 0.0001

Kelas	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>
Religion	0.91	0.85	0.88
Age	0.95	0.95	0.95
Gender	1.00	0.68	0.81
Ethnicity	1.00	0.90	0.95
Not bullying	0.63	0.93	0.75

4.2.3 Uji Coba Model C

Uji coba skenario Model C dilakukan dengan menerapkan ratio data 80:20.

Pengujian pertama menggunakan parameter *learning rate* 0.1.



Gambar 4. 11 *Confusion matrix* 80:20 learning rate 0.1

Tabel 4. 30 Tabel Confusion Matrix 80:20 Learning rate 0.1

Prediksi Aktual	Religion	Age	Gender	Ethnicity	Not bullying
Religion	39	0	0	1	0
Age	6	31	0	3	0
Gender	10	0	26	3	1
Ethnicity	5	0	0	34	1
Not bullying	34	1	0	3	2

Pada hasil tabel confusion matrix di atas dapat diketahui nilai-nilai aktual dan prediksi dari masing-masing kelas. Berikut merupakan hasil perhitungan nilai TP, TN, FN dan FP pada ratio 80:20 dan learning rate 0.1.

Tabel 4. 31 Hasil Perhitungan TP, TN, FP, FN (80:20 Learning rate 0.1)

Learning Rate	Kelas	TP	TN	FP	FN
0.1	Religion	39	93	55	1
	Age	31	101	1	9
	Gender	26	106	0	14
	Ethnicity	34	98	10	6
	Not bullying	2	130	2	38

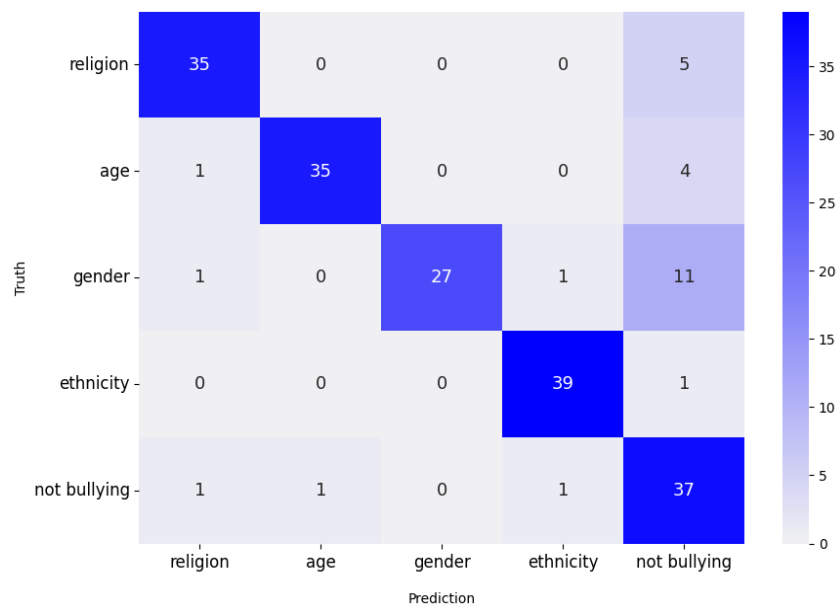
Pada tabel tersebut nilai 39 TP, menunjukkan jumlah prediksi yang benar untuk kelas Religion. Sebanyak 93 TN menunjukkan jumlah prediksi benar untuk data yang bukan kelas Religion. Ada 55 FP, menunjukkan jumlah prediksi yang salah untuk kelas Religion, dan 1 FN, menunjukkan jumlah prediksi yang salah untuk data yang bukan kelas Religion. Nilai 31 TP, menandakan jumlah prediksi yang benar untuk kelas Age. Hanya ada 1 FP, menunjukkan jumlah prediksi yang salah untuk kelas Age, dan 9 FN, menunjukkan jumlah prediksi yang salah untuk data yang bukan kelas Age. Untuk 26 TP, yang menunjukkan jumlah prediksi yang benar untuk kelas Gender. Terdapat 106 TN, menunjukkan jumlah prediksi benar untuk data yang bukan kelas Gender. Tidak ada FP, menunjukkan tidak ada prediksi yang salah untuk kelas Gender, dan 14 FN, menunjukkan jumlah prediksi yang

salah untuk data yang bukan kelas Gender. Terdapat 2 TP, yang menunjukkan jumlah prediksi yang benar untuk kelas Not bullying. Sebanyak 130 TN, menunjukkan jumlah prediksi benar untuk data yang bukan kelas Not bullying. Ada 2 FP, menunjukkan jumlah prediksi yang salah untuk kelas Not bullying, dan 38 FN, menunjukkan jumlah prediksi yang salah untuk data yang bukan kelas Not bullying. Hasil perhitungan TP, TN, FP dan FN menghasilkan akurasi 0.72, diperoleh nilai untuk *precision*, *recall*, dan *fi-score* seperti pada Tabel 4.32.

Tabel 4. 32 Nilai *Precision*, *Recall*, *F1-score* learning rate 0.1

Kelas	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>
Religion	0.41	0.97	0.58
Age	0.97	0.78	0.86
Gender	1.00	0.65	0.79
Ethnicity	0.77	0.85	0.81
Not bullying	0.50	0.05	0.09

Pengujian berikutnya dengan mengubah parameter *learning rate* menjadi 0.01. Hasil pengujian ditunjukkan pada gambar 4.12 sebagai berikut.



Gambar 4. 12 *Confusion matrix* 80:20 learning rate 0.01

Tabel 4. 33 Tabel Confusion Matrix 80:20 Learning rate 0.01

Prediksi Aktual	Religion	Age	Gender	Ethnicity	Not bullying
Religion	35	0	0	0	5
Age	1	35	0	0	4
Gender	1	0	27	1	11
Ethnicity	0	0	0	39	1
Not bullying	1	1	0	1	37

Pada hasil tabel confusion matrix di atas dapat diketahui nilai-nilai aktual dan prediksi dari masing-masing kelas dengan menggunakan skema confusion matrix. Berikut merupakan hasil perhitungan nilai TP, TN, FN dan FP pada ratio 80:20 dan learning rate 0.01.

Tabel 4. 34 Hasil Perhitungan TP, TN, FP, FN (80:20 Learning rate 0.01)

Learning Rate	Kelas	TP	TN	FP	FN
0.01	Religion	35	138	3	5
	Age	35	138	1	5
	Gender	27	146	0	13
	Ethnicity	39	134	2	1
	Not bullying	37	136	21	3

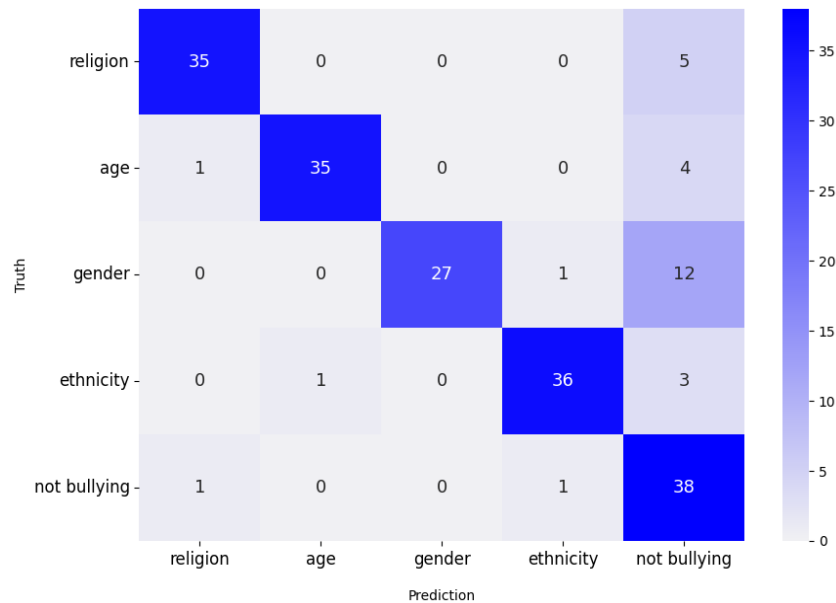
Pada tabel terdapat 35 TP, menunjukkan jumlah prediksi yang benar untuk kelas Religion. Sebanyak 138 TN menunjukkan jumlah prediksi benar untuk data yang bukan kelas Religion. Ada 3 FP, menunjukkan jumlah prediksi yang salah untuk kelas Religion, dan 5 FN, menunjukkan jumlah prediksi yang salah untuk data yang bukan kelas Religion. Ditemukan 35 TP, menandakan jumlah prediksi yang benar untuk kelas Age. Terdapat 138 TN, menunjukkan jumlah prediksi benar untuk data yang bukan kelas Age. Hanya ada 1 FP, menunjukkan jumlah prediksi yang salah untuk kelas Age, dan 5 FN, menunjukkan jumlah prediksi yang salah untuk data yang bukan kelas Age. Terdapat 27 TP, yang menunjukkan jumlah prediksi yang benar untuk kelas Gender. Terdapat 146 TN, menunjukkan jumlah prediksi benar

untuk data yang bukan kelas Gender. Tidak ada FP, menunjukkan tidak ada prediksi yang salah untuk kelas Gender, dan 13 FN, menunjukkan jumlah prediksi yang salah untuk data yang bukan kelas Gender. Ditemukan 39 TP, menandakan jumlah prediksi yang benar untuk kelas Ethnicity. Terdapat 134 TN, menunjukkan jumlah prediksi benar untuk data yang bukan kelas Ethnicity. Ada 2 FP, menunjukkan jumlah prediksi yang salah untuk kelas Ethnicity, dan 1 FN, menunjukkan jumlah prediksi yang salah untuk data yang bukan kelas Ethnicity. Terdapat 37 TP, yang menunjukkan jumlah prediksi yang benar untuk kelas Not bullying. Sebanyak 136 TN, menunjukkan jumlah prediksi benar untuk data yang bukan kelas Not bullying. Ada 21 FP, menunjukkan jumlah prediksi yang salah untuk kelas Not bullying, dan 3 FN, menunjukkan jumlah prediksi yang salah untuk data yang bukan kelas Not bullying. Tabel 4.35 berikut menunjukkan nilai *precision*, *recall*, dan *f1-score* yang diperoleh berdasarkan jumlah data yang mampu diklasifikasi pada Gambar 4.35 yang dapat menghasilkan akurasi sebesar 0.865.

Tabel 4. 35 Nilai *Precision*, *Recall*, *F1-score* learning rate 0.01

Kelas	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>
Religion	0.92	0.88	0.90
Age	0.97	0.88	0.92
Gender	1.00	0.68	0.81
Ethnicity	0.95	0.97	0.96
Not bullying	0.64	0.93	0.76

Pengujian berikutnya dengan mengubah parameter *learning rate* dari 0.01 menjadi 0.001 untuk mengetahui tingkat akurasi yang dihasilkan apakah dapat lebih baik daripada parameter *learning rate* sebelumnya. Jumlah data yang mampu diklasifikasi oleh *confusion matrix* ditunjukkan pada Gambar 4.13.

Gambar 4. 13 *Confusion matrix 80:20 learning rate 0.001*

Tabel 4. 36 Tabel Confusion Matrix 80:20 Learning rate 0.001

Prediksi \ Aktual	Religion	Age	Gender	Ethnicity	Not bullying
Religion	35	0	0	0	5
Age	1	35	0	0	4
Gender	0	0	27	1	12
Ethnicity	0	1	0	36	3
Not bullying	1	0	0	1	38

Pada hasil tabel confusion matrix di atas dapat diketahui nilai-nilai aktual dan prediksi dari masing-masing kelas dengan menggunakan skema confusion matrix. Berikut merupakan hasil perhitungan nilai TP, TN, FN dan FP pada ratio 80:20 dan learning rate 0.001.

Tabel 4. 37 Hasil Perhitungan TP, TN, FP, FN (80:20 Learning rate 0.001)

Learning Rate	Kelas	TP	TN	FP	FN
0.001	Religion	35	136	2	5
	Age	35	136	1	5
	Gender	27	144	0	13
	Ethnicity	36	135	2	4
	Not bullying	38	133	24	2

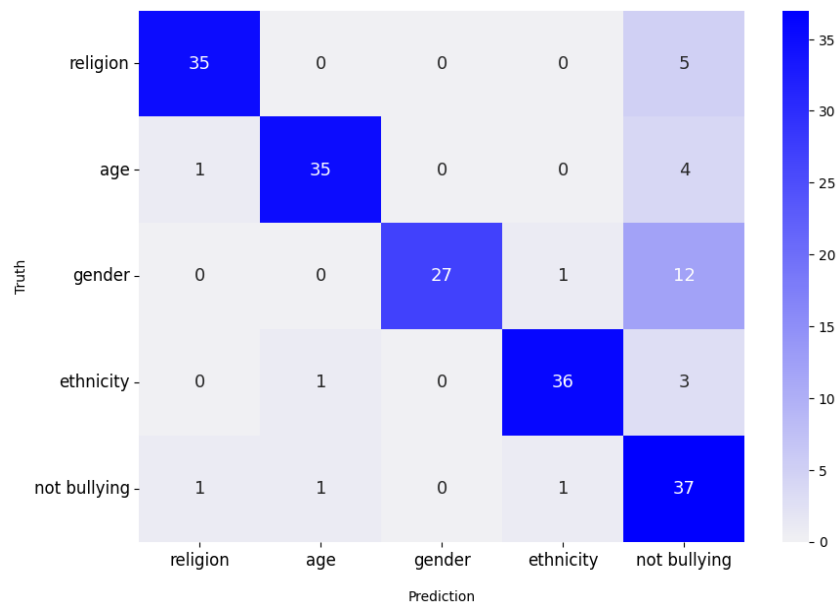
Berdasarkan tabel tersebut terdapat 35 TP, menunjukkan jumlah prediksi yang benar untuk kelas Religion. Sebanyak 136 TN menunjukkan jumlah prediksi benar untuk data yang bukan kelas Religion. Ada 2 FP, menunjukkan jumlah prediksi yang salah untuk kelas Religion, dan 5 FN, menunjukkan jumlah prediksi yang salah untuk data yang bukan kelas Religion. Ditemukan 35 TP, menandakan jumlah prediksi yang benar untuk kelas Age. Terdapat 136 TN, menunjukkan jumlah prediksi benar untuk data yang bukan kelas Age. Hanya ada 1 FP, menunjukkan jumlah prediksi yang salah untuk kelas Age, dan 5 FN, menunjukkan jumlah prediksi yang salah untuk data yang bukan kelas Age. Terdapat 27 TP, yang menunjukkan jumlah prediksi yang benar untuk kelas Gender. Terdapat 144 TN, menunjukkan jumlah prediksi benar untuk data yang bukan kelas Gender. Tidak ada FP, menunjukkan tidak ada prediksi yang salah untuk kelas Gender, dan 13 FN, menunjukkan jumlah prediksi yang salah untuk data yang bukan kelas Gender. Ditemukan 36 TP, menandakan jumlah prediksi yang benar untuk kelas Ethnicity. Terdapat 135 TN, menunjukkan jumlah prediksi benar untuk data yang bukan kelas Ethnicity. Ada 2 FP, menunjukkan jumlah prediksi yang salah untuk kelas Ethnicity, dan 4 FN, menunjukkan jumlah prediksi yang salah untuk data yang bukan kelas Ethnicity. Terdapat 38 TP, yang menunjukkan jumlah prediksi yang benar untuk kelas Not bullying. Sebanyak 133 TN, menunjukkan jumlah prediksi benar untuk data yang bukan kelas Not bullying. Ada 24 FP, menunjukkan jumlah prediksi yang salah untuk kelas Not bullying, dan 2 FN, menunjukkan jumlah prediksi yang salah untuk data yang bukan kelas Not bullying. Berdasarkan Tabel 4.37 yang menunjukkan hasil perhitungan TP, TN, FP dan FN yang diperoleh

dengan menerapkan *learning rate* 0.001 terhadap ratio 80:20 menghasilkan tingkat akurasi sebesar 0.85. Selain itu diperoleh nilai *precision*, *recall*, dan *fi-score* seperti pada Tabel 4.38.

Tabel 4. 38 Nilai *Precision*, *Recall*, *F1-score* *learning rate* 0.001

Kelas	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>
Religion	0.95	0.88	0.91
Age	0.97	0.88	0.92
Gender	1.00	0.68	0.81
Ethnicity	0.95	0.90	0.92
Not bullying	0.61	0.95	0.75

Pengujian selanjutnya dilakukan dengan mengubah parameter *learning rate* menjadi 0.0001. Hasil yang mampu diklasifikasi oleh sistem menggunakan *learning rate* ini ditunjukkan pada *confusion matrix* pada Gambar 4.14. Dengan *learning rate* yang lebih rendah ini, diharapkan model dapat melakukan pembaruan bobot dengan lebih halus dan meningkatkan akurasi klasifikasi.



Gambar 4. 14 *Confusion matrix* 80:20 *learning rate* 0.0001

Tabel 4. 39 Tabel Confusion Matrix 80:20 Learning rate 0. 0001

Prediksi Aktual	Religion	Age	Gender	Ethnicity	Not bullying
Religion	35	0	0	0	5
Age	1	35	0	0	4
Gender	0	0	27	1	12
Ethnicity	0	1	0	36	3
Not bullying	1	1	0	1	37

Pada hasil tabel confusion matrix di atas dapat diketahui nilai-nilai aktual dan prediksi dari masing-masing kelas dengan menggunakan skema confusion matrix. Berikut merupakan hasil perhitungan nilai TP, TN, FN dan FP pada ratio 80:20 dan learning rate 0.0001.

Tabel 4. 40 Hasil Perhitungan TP, TN, FP, FN (80:20 Learning rate 0.0001)

Learning Rate	Kelas	TP	TN	FP	FN
0.0001	Religion	35	135	2	5
	Age	35	135	2	5
	Gender	27	143	0	13
	Ethnicity	36	134	2	4
	Not bullying	37	133	24	3

Pada tabel menunjukkan terdapat 35 TP, menunjukkan jumlah prediksi yang benar untuk kelas Religion. Sebanyak 135 TN menunjukkan jumlah prediksi benar untuk data yang bukan kelas Religion. Ada 2 FP, menunjukkan jumlah prediksi yang salah untuk kelas Religion, dan 5 FN, menunjukkan jumlah prediksi yang salah untuk data yang bukan kelas Religion. Ditemukan 35 TP, menandakan jumlah prediksi yang benar untuk kelas Age. Terdapat 135 TN, menunjukkan jumlah prediksi benar untuk data yang bukan kelas Age. Ada 2 FP, menunjukkan jumlah prediksi yang salah untuk kelas Age, dan 5 FN, menunjukkan jumlah prediksi yang salah untuk data yang bukan kelas Age. Terdapat 27 TP, yang menunjukkan jumlah prediksi yang benar untuk kelas Gender. Terdapat 143 TN, menunjukkan jumlah prediksi

benar untuk data yang bukan kelas Gender. Tidak ada FP, menunjukkan tidak ada prediksi yang salah untuk kelas Gender, dan 13 FN, menunjukkan jumlah prediksi yang salah untuk data yang bukan kelas Gender. Ditemukan 36 TP, menandakan jumlah prediksi yang benar untuk kelas Ethnicity. Terdapat 134 TN, menunjukkan jumlah prediksi benar untuk data yang bukan kelas Ethnicity. Ada 2 FP, menunjukkan jumlah prediksi yang salah untuk kelas Ethnicity, dan 4 FN, menunjukkan jumlah prediksi yang salah untuk data yang bukan kelas Ethnicity. Terdapat 37 TP, yang menunjukkan jumlah prediksi yang benar untuk kelas Not bullying. Sebanyak 133 TN, menunjukkan jumlah prediksi benar untuk data yang bukan kelas Not bullying. Ada 24 FP, menunjukkan jumlah prediksi yang salah untuk kelas Not bullying, dan 3 FN, menunjukkan jumlah prediksi yang salah untuk data yang bukan kelas Not bullying. Kemudian diperoleh tingkat akurasi sebesar 0.85 dengan nilai *precision*, *recall*, dan *fi-score* yang ditunjukkan pada Tabel 4.41.

Tabel 4. 41 Nilai *Precision*, *Recall*, *F1-score* learning rate 0.0001

Kelas	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>
Religion	0.95	0.88	0.91
Age	0.95	0.88	0.91
Gender	1.00	0.68	0.81
Ethnicity	0.95	0.90	0.92
Not bullying	0.61	0.93	0.73

Hasil uji coba menggunakan ketiga ratio data yaitu 60:40, 70:30, dan 80:20 serta perubahan parameter *learning rate* 0.1, 0.01, 0.001, dan 0.0001 pada *epoch* yang digunakan yaitu 100 mampu menghasilkan tingkat akurasi seperti yang tertera pada Tabel 4.44 berikut ini. Penelitian ini mengeksplorasi pengaruh kombinasi rasio data dan *learning rate* terhadap performa model, dengan tujuan untuk mengidentifikasi konfigurasi optimal yang menghasilkan akurasi prediksi yang tinggi dan konsisten.

Tabel 4. 42 Hasil Pengujian Parameter

Ratio Data	Learning rate	Akurasi
60:40	0.1	62%
60:40	0.01	84%
60:40	0.001	83.8%
60:40	0.0001	86%
70:30	0.1	57.3%
70:30	0.01	86.3%
70:30	0.001	84.7%
70:30	0.0001	86%
80:20	0.1	72%
80:20	0.01	86.5%
80:20	0.001	85%
80:20	0.0001	85%

Berdasarkan Tabel 4.42 diatas, akurasi terbaik yang mampu dihasilkan ketika menggunakan ratio data 80:20 dengan parameter *learning rate* sebesar 0.01 yaitu 86.5%. Pada pengujian ratio data 60:40 mampu menghasilkan akurasi rata-rata sebesar 78.9%, ratio data 70:30 juga mampu menghasilkan akurasi rata-rata sebesar 78.5%. Sedangkan pada pengujian ratio data 80:20, diperoleh akurasi rata-rata sebesar 82.1%. Pada Tabel 4.42 dapat dilihat bahwa penerapan *learning rate* yang berbeda dapat mempengaruhi hasil akurasi yang dihasilkan. Adapun jumlah *epoch* yang diterapkan pada pengujian parameter ini sama untuk semua ratio data dan *learning rate*, yaitu sebesar 100.

4.2.4 Uji Coba K-Fold Cross Validation

Hasil terbaik yang diperoleh dari sejumlah pengujian sebelumnya adalah ketika menggunakan rasio perbandingan data 60:40 dan learning rate sebesar 0.0001, yang menghasilkan tingkat akurasi sebesar 86%. Proses selanjutnya adalah dilakukan pengujian. Hasil pengujian tersebut di validasi dengan teknik *k-fold cross validation*, dengan pemilihan nilai k sebanyak 10. Berikut hasil dari validasi menggunakan *10-fold cross validation* dapat dilihat pada Tabel 4.43 berikut.

Tabel 4. 43 Hasil akurasi 10-fold cross validation (60:40)

<i>fold ke-</i>	Hasil Akurasi
1	86.6%
2	80%
3	91.6%
4	86.6%
5	83.3%
6	90%
7	83.3%
8	93.3%
9	85%
10	91.6%
Rata-rata	87.1%
Standar Deviasi	0.0414

Tabel tersebut menggambarkan sejauh mana model mampu mempertahankan performa yang konsisten ketika dihadapkan pada variasi data yang luas. Berdasarkan pengujian menggunakan teknik *k-fold cross validation* dengan pemilihan nilai k sebanyak 10 kali. Hasilnya menunjukkan bahwa akurasi terbaik yang diperoleh terjadi pada iterasi ke-8 dengan akurasi sebesar 93.3%. Rata-rata akurasi yang tinggi ini menunjukkan bahwa model memiliki performa yang sangat baik secara keseluruhan. Standar deviasi yang lebih besar mengindikasikan bahwa hasil akurasi antar fold bervariasi lebih signifikan. Meskipun demikian, performa model secara keseluruhan tetap sangat baik. Tingginya rata-rata akurasi ini menunjukkan bahwa model dapat menangani berbagai subset data dengan efektif. Namun, variasi yang lebih besar dalam hasil akurasi antar fold menunjukkan kurangnya konsistensi. Ini berarti bahwa dalam beberapa subset data, model mungkin tidak tampil sebaik di subset lainnya.

Hasil terbaik yang diperoleh dari sejumlah pengujian sebelumnya adalah ketika menggunakan rasio perbandingan data 70:30 dan learning rate sebesar 0.01, yang menghasilkan tingkat akurasi sebesar 86.3%. Proses selanjutnya adalah dilakukan pengujian. Hasil pengujian tersebut di validasi dengan teknik *k-fold cross validation*, dengan pemilihan nilai k sebanyak 10. Berikut hasil dari validasi menggunakan *10-fold cross validation* dapat dilihat pada Tabel 4.44 berikut.

Tabel 4. 44 Hasil akurasi *10-fold cross validation* (70:30)

<i>fold</i> ke-	Hasil Akurasi
1	88.5%
2	87.1%
3	85.7%
4	88.5%
5	82.8%
6	91.4%
7	78.5%
8	84.2%
9	90%
10	84.2%
Rata-rata	86.1%
Standar Deviasi	0.0363

Tabel tersebut menggambarkan sejauh mana model mampu mempertahankan performa yang konsisten ketika dihadapkan pada variasi data yang luas. Berdasarkan pengujian menggunakan teknik *k-fold cross validation* dengan pemilihan nilai k sebanyak 10 kali. Hasilnya menunjukkan bahwa akurasi terbaik yang diperoleh terjadi pada iterasi ke-6 dengan akurasi sebesar 91.4%. Rata-rata akurasi yang tinggi ini menunjukkan bahwa model memiliki performa yang sangat baik secara keseluruhan. Standar deviasi yang lebih besar mengindikasikan bahwa hasil akurasi antar fold bervariasi lebih signifikan. Meskipun demikian,

performa model secara keseluruhan tetap sangat baik. Tingginya rata-rata akurasi ini menunjukkan bahwa model dapat menangani berbagai subset data dengan efektif. Namun, variasi yang lebih besar dalam hasil akurasi antar fold menunjukkan kurangnya konsistensi. Ini berarti bahwa dalam beberapa subset data, model mungkin tidak tampil sebaik di subset lainnya.

Hasil terbaik yang diperoleh dari sejumlah pengujian sebelumnya adalah ketika menggunakan rasio perbandingan data 80:20 dan learning rate sebesar 0.01, yang menghasilkan tingkat akurasi sebesar 86.5%. Proses selanjutnya adalah dilakukan pengujian. Hasil pengujian tersebut di validasi dengan teknik *k-fold cross validation*, dengan pemilihan nilai k sebanyak 10. Berikut hasil dari validasi menggunakan *4-fold cross validation* dapat dilihat pada Tabel 4.45 berikut.

Tabel 4. 45 Hasil akurasi 10-*fold cross validation* (80:20)

<i>fold ke-</i>	Hasil Akurasi
1	85%
2	83.7%
3	85%
4	85%
5	92.5%
6	95%
7	86.2%
8	88.7%
9	81.2%
10	78.7%
Rata-rata	86.1%
Standar Deviasi	0.0463

Tabel tersebut menggambarkan sejauh mana model mampu mempertahankan performa yang konsisten ketika dihadapkan pada variasi data yang luas. Berdasarkan pengujian menggunakan teknik *k-fold cross validation*

dengan pemilihan nilai k sebanyak 10 kali. Hasilnya menunjukkan bahwa akurasi terbaik yang diperoleh terjadi pada iterasi ke-6 dengan akurasi sebesar 95%. Rata-rata akurasi yang tinggi ini menunjukkan bahwa model memiliki performa yang sangat baik secara keseluruhan. Standar deviasi yang lebih besar mengindikasikan bahwa hasil akurasi antar fold bervariasi lebih signifikan. Meskipun demikian, performa model secara keseluruhan tetap sangat baik. Tingginya rata-rata akurasi ini menunjukkan bahwa model dapat menangani berbagai subset data dengan efektif. Namun, variasi yang lebih besar dalam hasil akurasi antar fold menunjukkan kurangnya konsistensi. Ini berarti bahwa dalam beberapa subset data, model mungkin tidak tampil sebaik di subset lainnya.

4.3 Pembahasan

Berdasarkan hasil uji coba yang dilakukan dengan menggunakan 3 jenis pembagian ratio data dan juga 4 jenis *learning rate* yang berbeda mampu menghasilkan akurasi terbaik pada ratio data 80:20 dengan menggunakan *learning rate* yaitu 0.01. Jumlah data yang digunakan juga berpengaruh terhadap hasil akurasi yang mampu dihasilkan oleh model yang digunakan sebagai penelitian. Pada penelitian ini, model yang digunakan adalah *Learning vector quantization*. Penggunaan data dalam jumlah besar untuk pelatihan menggunakan metode ini dapat mempengaruhi beberapa aspek sehingga kemungkinan hasil yang diperoleh juga tidak terlalu optimal. Semakin besar data yang digunakan akan semakin lama juga waktu yang digunakan untuk pelatihan dan bisa memunculkan kemungkinan *overfitting*. Dataset yang digunakan pada penelitian ini memiliki jumlah yang sama pada masing-masing kelas. Sehingga menghindari terjadinya *oversampling* dan

undersampling karena jumlah data yang tidak seimbang untuk pelatihannya. Penentuan penggunaan parameter untuk model ini memiliki pertimbangan yang bertujuan untuk membandingkan parameter yang paling optimal yang mampu diterapkan pada model LVQ dengan menggunakan *epoch* yang sudah diterapkan dalam jumlah data sekian. Kemudian dilanjutkan untuk pengujian dengan menggunakan teknik *K-fold Cross validation*, dengan pemilihan nilai k sebesar 4, 5 dan 10 fold.

4.3.1 Perubahan learning rate

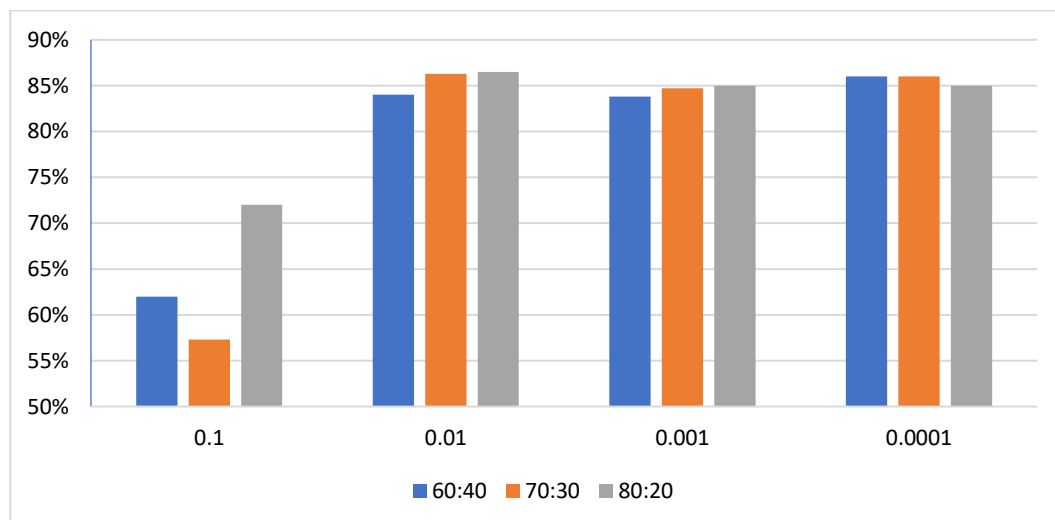
Salah satu faktor yang dapat mempengaruhi hasil akurasi yang berbeda pada penelitian ini adalah variasi terhadap *learning rate* yang diterapkan pada model. *Learning rate* yang terlalu kecil dapat menyebabkan konvergensi memakan waktu yang lama sehingga menyebabkan proses pelatihan berjalan dengan sangat lambat. Namun jika *learning rate* yang digunakan terlalu besar juga dapat menyebabkan minimum yang tidak baik karena konvergensi yang berlangsung terjadi secara cepat. Pada penelitian ini, *learning rate* yang digunakan terdapat 4 variasi yaitu 0.1, 0.01, 0.001, 0.0001. Seperti yang terlihat pada Tabel 4.46 dimana hasil akurasi mencapai hasil paling optimal pada *learning rate* 0.01 dengan rasio data 80:20 dimana menghasilkan akurasi sebesar 86.5%. Untuk learning rate 0.1 menghasilkan rata-rata akurasi sebesar 63.7%. Learning rate 0.01 menghasilkan akurasi dengan rata-rata 85.6%. pada learning rate 0.001 menghasilkan rata-rata akurasi kurang lebih 84.5%. sedangkan untuk learning rate 0.0001 mendapatkan rata-rata akurasi sebesar 85.6% dimana dari keempat learning rate, learning rate 0.1 menghasilkan rata-rata akurasi terendah. Hal ini menunjukkan bahwa memilih learning rate yang

tepat sangat penting untuk mencapai kinerja model yang optimal. Terlalu besar atau terlalu kecilnya learning rate dapat secara signifikan mempengaruhi hasil akhir model. Dengan menggunakan learning rate 0.01, model dapat mencapai keseimbangan yang baik antara kecepatan konvergensi dan stabilitas. Oleh karena itu, pemilihan parameter yang tepat dalam pelatihan model sangat krusial untuk memperoleh hasil yang optimal.

Tabel 4. 46 Perubahan Learning Rate

Learning Rate	Ratio Data	Akurasi	Rata-Rata
0.1	60:40	62%	63.7%
	70:30	57.3%	
	80:20	72%	
0.01	60:40	84%	85.6%
	70:30	86.3%	
	80:20	86.5%	
0.001	60:40	83.8%	84.5%
	70:30	84.7%	
	80:20	85%	
0.0001	60:40	86%	85.6%
	70:30	86%	
	80:20	85%	

Pada Gambar 4.15 merepresentasikan grafik dari perubahan learning rate. Dalam grafik tersebut setiap nilai learning rate diletakkan di sumbu x, sementara tinggi batang mewakili akurasi. Setiap ratio pembagian data (60:40, 70:30, 80:20) memiliki beberapa batang yang mewakili nilai akurasi untuk masing-masing ratio pembagian data. Grafik ini memberikan representasi visual mengenai bagaimana nilai α dan pembagian data mempengaruhi akurasi dalam model atau percobaan tertentu.



Gambar 4. 15 Grafik perubahan learning rate

4.3.2 Pembagian Ratio Data

Skenario pengujian yang digunakan dengan melakukan pembagian rasio data menjadi beberapa variasi yaitu 60:40, 70:30, dan 80:20. Tujuan dilakukannya variasi pembagian rasio data adalah untuk melakukan evaluasi terhadap performa model dan melakukan pengujian untuk mengetahui sampai sejauh mana model mampu diterapkan ke dalam data yang belum pernah dilihat sebelumnya oleh model. Pada Tabel 4.47 dapat dilihat bahwa rasio data yang berbeda memberikan hasil akurasi yang berbeda. Pada tabel tersebut akurasi paling optimal ditunjukkan oleh rasio data 80:20 pada variasi *learning rate* 0.01.

Pada rasio data 80:20 dengan learning rate 0.01 mampu menghasilkan *precision*, *recall*, dan *f1-score* sebesar 89.8%, 86.8%, dan 87%. Dalam hal ini, *precision* sebesar 89.8% menunjukkan bahwa dari semua prediksi yang diberi label positif oleh model, sekitar 89.8% di antaranya adalah benar-benar positif. Ini menunjukkan bahwa model memiliki kemampuan untuk mengurangi jumlah positif

palsu. Recall sebesar 86.8% menunjukkan bahwa model dapat menangkap sekitar 86.8% dari semua instance positif yang sebenarnya (aktual). Ini menunjukkan bahwa model cukup baik dalam menangani positif sejati. F1-score 87% menunjukkan bahwa model mencapai keseimbangan yang baik antara ketepatan dan kelengkapan dalam prediksinya. Jika dilihat pada tabel 4.47 rata-rata akurasi yang diperoleh pada ratio data 60:40 dengan menggunakan 4 learning rate menghasilkan akurasi kurang lebih 78.9%. Untuk ratio data 70:30 menghasilkan rata-rata akurasi 78.5%. Sedangkan pada ratio data 80:20 menghasilkan rata-rata akurasi terbesar yaitu 82.1%.

Tabel 4. 47 Pembagian Ratio Data

Ratio Data	Learning Rate	Akurasi	Rata-Rata
60:40	0.1	62%	78.9%
	0.01	84%	
	0.001	83.8%	
	0.0001	86%	
70:30	0.1	57.3%	78.5%
	0.01	86.3%	
	0.001	84.7%	
	0.0001	86%	
80:20	0.1	72%	82.1%
	0.01	86.5%	
	0.001	85%	
	0.0001	85%	

4.3.3 K-fold Cross Validation

Pengujian k-fold cross validation dengan berbagai nilai k dan rasio pembagian data menunjukkan variasi performa model yang signifikan. Dengan rasio pembagian data 60:40 dan learning rate sebesar 0.0001, pengujian menggunakan 10-fold cross validation menghasilkan akurasi rata-rata sebesar 87.1%, tetapi standar deviasi juga meningkat, menunjukkan adanya variasi hasil

yang lebih besar antar fold. Pengujian ini menunjukkan bahwa dengan nilai k yang lebih tinggi, meskipun akurasi rata-rata bisa sedikit lebih tinggi, variasi antar fold juga meningkat, yang mengindikasikan kurangnya konsistensi dalam performa model pada beberapa subset data. Pada rasio data 70:30 dan 80:20 dengan learning rate sebesar 0.01, pola yang serupa juga terlihat. Untuk rasio 70:30, akurasi rata-rata 86.1% pada 10-fold, dengan standar deviasi yang juga meningkat. Demikian pula, pada rasio 80:20, akurasi rata-rata 86.1% pada 10-fold, dengan standar deviasi yang cukup tinggi. Berdasarkan hasil pengujian k -fold cross validation dengan berbagai rasio pembagian data, terlihat bahwa penggunaan rasio 60:40 dengan learning rate 0.0001 menunjukkan performa yang lebih konvergen dibandingkan rasio lainnya. Tanpa menggunakan k -fold, rasio ini mencapai akurasi 86.3%, sementara dengan 10-fold cross validation, rata-rata akurasi mencapai 87.1% dengan standar deviasi 0.0414. Meskipun akurasi rata-rata lebih tinggi dari hasil non k -fold, k -fold cross validation memberikan gambaran yang lebih lengkap tentang performa model dengan mengurangi kemungkinan overfitting dan memastikan bahwa model diuji pada berbagai subset data. Secara keseluruhan, pengujian menunjukkan bahwa penggunaan k yang lebih tinggi dapat memberikan gambaran yang lebih rinci tentang performa model, namun juga dapat mengungkapkan variabilitas yang lebih besar dalam hasil prediksi. Model tetap menunjukkan performa yang baik secara umum, tetapi peningkatan variasi hasil antar fold menunjukkan perlunya perhatian lebih terhadap stabilitas model pada berbagai subset data.

4.3.4 Integrasi Islam

Tujuan penelitian ini dilakukan yaitu untuk mengukur dan mengevaluasi tingkat keakuratan metode Learning vector quantization (LVQ) dalam mengklasifikasi cyberbullying pada media sosial Twitter berdasarkan lima kelas yaitu *Religion, Age, Gender, Ethnicity, Not Bullying*. Penelitian ini memberikan kontribusi penting dalam mengklasifikasi dan mencegah *cyberbullying* lebih dini sehingga dapat mengidentifikasi dengan cepat perilaku yang termasuk dalam kategori *cyberbullying* dan memungkinkan tindakan pencegahan yang lebih efektif. Dalam penelitian ini, penting untuk menggunakan proporsi data *training* yang tepat untuk meningkatkan kinerja model sehingga dapat mengidentifikasi dengan cepat perilaku yang termasuk dalam kategori *cyberbullying* dan memungkinkan tindakan pencegahan yang lebih efektif. Dalam hal ini sesuai dengan Q.S Qamar ayat 49 yang memberikan wawasan tentang ketelitian dan kebijaksanaan Allah dalam penciptaan-Nya:

إِنَّا كُلَّ شَيْءٍ خَلَقْنَاهُ بِقَدَرٍ

“*Sesungguhnya Kami menciptakan segala sesuatu sesuai dengan ukuran..*” (Q.S Qamar:49)

Menurut tafsir Jalalain mengenai ayat ini yaitu (Sesungguhnya segala sesuatu itu Kami) dinashabkan oleh Fi'il yang terdapat pada firman selanjutnya yang berfungsi menafsirkannya (ciptakan menurut ukuran) masing-masing. Menurut suatu qiraat lafal Kulla dibaca Kullu dan dianggap sebagai Muftada, sedangkan Khabarnya adalah lafal Khalaqnaahu. Tafsir tersebut menggambarkan bahwa segala sesuatu yang ada di dunia ini diciptakan oleh Allah dengan penuh

kebijaksanaan dan ketelitian sesuai dengan ukuran yang telah ditentukan-Nya. Ini mencerminkan prinsip kesempurnaan dan keakuratan dalam penciptaan-Nya. Sistem klasifikasi yang benar-benar akurat harus mampu mengidentifikasi dan memetakan perilaku online sesuai dengan standar yang jelas dan tepat. Sebagaimana Allah menciptakan setiap entitas sesuai dengan ukuran-Nya, demikian juga sistem klasifikasi harus dapat mengukur dan mengklasifikasikan perilaku secara akurat berdasarkan norma etika dan nilai-nilai kebaikan dalam Islam. Sebagaimana firman tersebut menegaskan bahwa Allah menciptakan segala sesuatu sesuai dengan ukuran, sistem klasifikasi yang akurat dan tepat akan mencerminkan tindakan yang mengedepankan kerjasama, kebaikan, dan saling menghormati dalam dunia maya. Oleh karena itu, seperti Allah menciptakan segala sesuatu dengan keakuratan yang luar biasa, sistem klasifikasi juga perlu diatur dengan akurat untuk menciptakan lingkungan online yang seimbang dan mematuhi prinsip-prinsip moral serta etika Islam. Kemudian dalam surah Al-A'la ayat 1-3.

سَبِّحْ اسْمَ رَبِّكَ الْأَعْلَى الَّذِي خَلَقَ فَسَوَّى وَالَّذِي قَدَّرَ فَهَدَى

"Sucikanlah nama Tuhanmu Yang Mahatinggi, yang menciptakan, lalu menyempurnakan (ciptaan-Nya), yang menentukan kadar (masing-masing) dan memberi petunjuk," (QS. Al-A'la:1-3)

Dalam tafsir Jalalain, disebutkan bahwa perintah "Sucikanlah nama Rabbmu" mengandung makna untuk membersihkan nama Allah dari sifat-sifat yang tidak pantas bagi-Nya. Kata "Ismu" di sini dianggap sebagai kata tambahan (Za'id) yang menggambarkan sifat-Nya yang Maha Tinggi, yaitu Al-A'la, yang berperan sebagai kata sifat bagi kata "Rabbika". Ayat berlanjut dengan menjelaskan

bahwa Allah menciptakan segala sesuatu dengan sempurna, menjadikan seluruh bagian makhluk-Nya seimbang dan tidak ada yang cacat atau tidak seimbang. Allah pula yang menentukan segala sesuatu sesuai kehendak-Nya, dan memberi petunjuk kepada manusia tentang apa yang telah ditetapkan-Nya, baik dalam hal perbuatan baik maupun buruk. Dalam ayat tersebut mengajarkan bahwa Allah menciptakan segala sesuatu dengan seimbang dan sempurna, tanpa kurang suatu apapun. Hal ini mencakup semua aspek kehidupan, baik fisik maupun spiritual, yang memberikan nilai penting pada setiap bagian dari ciptaan-Nya. Ayat ini juga menegaskan bahwa Allah menentukan segala sesuatu dengan arahan yang terbaik. Dalam konteks penelitian tentang sistem klasifikasi cyberbullying, tujuannya adalah untuk mencari kebenaran dan arahan yang tepat dalam mengoptimalkan lingkungan digital agar lebih aman dan beradab bagi semua pengguna.

Sistem Klasifikasi cyberbullying membantu kita mengidentifikasi tindakan-tindakan yang melanggar prinsip-prinsip yang diajarkan dalam Surah Al-Hujurat ayat 11. Dengan memahami jenis-jenis perilaku yang termasuk dalam kategori cyberbullying, kita dapat lebih efektif dalam mengatasi dan mencegahnya. Misalnya, komentar-komentar yang menghina, mengancam, atau menyebarkan informasi pribadi orang lain secara tidak sah dapat dianggap sebagai bentuk cyberbullying yang harus ditangani. Prinsip-prinsip keadilan dan pengampunan harus menjadi bagian dari pendekatan dalam menangani kasus-kasus cyberbullying, sesuai dengan ajaran Islam yang mengajarkan untuk memberikan kesempatan kepada orang lain untuk bertaubat dan memperbaiki diri. Dengan memahami dan mengimplementasikan nilai-nilai yang terkandung dalam Surah Al-

Hujurat ayat 11 dalam konteks klasifikasi cyberbullying, dapat membangun lingkungan online yang lebih aman, lebih menghormati, dan lebih bermartabat bagi semua orang.

يَا أَيُّهَا الَّذِينَ آمَنُوا لَا يَسْخَرْ قَوْمٌ مِّن قَوْمٍ عَسَىٰ أَن يَكُونُوا خَيْرًا مِّنْهُمْ وَلَا نِسَاءٌ مِّن نِّسَاءٍ عَسَىٰ أَن يَكُنَّ خَيْرًا مِّنْهُنَّ
وَلَا تَلْمِزُوا أَنفُسَكُمْ وَلَا تَنَابَرُوا بِالْأَلْقَابِ بِئْسَ الْإِسْمُ الْفُسُوقُ بَعْدَ الْإِيمَانِ وَمَنْ لَّمْ يَتُبْ فَأُولَٰئِكَ هُمُ الظَّالِمُونَ

“Wahai orang-orang yang beriman, janganlah suatu kaum mengolok-olok kaum yang lain (karena) boleh jadi mereka (yang diolok-olokkan itu) lebih baik daripada mereka (yang mengolok-olok) dan jangan pula perempuan-perempuan (mengolok-olok) perempuan lain (karena) boleh jadi perempuan (yang diolok-olok itu) lebih baik daripada perempuan (yang mengolok-olok). Janganlah kamu saling mencela dan saling memanggil dengan julukan yang buruk. Seburuk-buruk panggilan adalah (panggilan) fasik setelah beriman. Siapa yang tidak bertobat, mereka itulah orang-orang zalim.” (QS. Al-Hujurat:11)

Dalam tafsir Jalalain ayat ini diturunkan berkenaan dengan delegasi dari Bani Tamim sewaktu mereka mengejek orang-orang muslim yang miskin, seperti Ammar bin Yasir dan Shuhaib Ar-Rumi. As-Sukhriyah memiliki arti merendahkan dan menghina (suatu kaum) yaitu sebagian di antara kalian (kepada kaum yang lain karena boleh jadi mereka yang diolok-olokkan lebih baik dari pada mereka yang mengolok-olokkan) di sisi Allah (dan jangan pula wanita-wanita) di antara kalian mengolok-olokkan (wanita-wanita lain karena boleh jadi wanita-wanita yang diperolok-olokkan lebih baik dari wanita-wanita yang mengolok-olokkan dan janganlah kalian mencela diri kalian sendiri) sebagaimana artinya, janganlah kalian mencela, maka karenanya kalian akan dicela. Makna yang dimaksud ialah, janganlah sebagian dari kalian mencela sebagian yang lain (dan janganlah kalian panggil-memanggil dengan gelar-gelar yang buruk) janganlah sebagian di antara kalian memanggil sebagian yang lain dengan nama julukan yang tidak disukainya,

antara lain seperti, hai orang fasik, atau hai orang kafir. (Seburuk-buruk nama) panggilan yang telah disebutkan di atas, yakni memperolok-olokkan orang lain mencela dan memanggil dengan nama julukan yang buruk (ialah nama yang buruk sesudah iman) lafal Al-Fusuq adalah Badal dari lafal Al-Ismu, dikarenakan nama panggilan yang dimaksud memberikan pengertian fasik dan juga karena nama panggilan itu biasanya diulang-ulang (dan barang siapa yang tidak bertobat) dari perbuatan tersebut (maka mereka itulah orang-orang yang dzalim). Perbuatan dzalim ini sangat dikecam dalam Islam karena merusak hubungan sosial dan mencederai persaudaraan di antara sesama muslim. Dengan demikian, ayat ini memberikan pedoman moral yang jelas bagi umat Islam untuk menjaga keharmonisan dan kedamaian dalam masyarakat dengan menghindari tindakan-tindakan yang merendahkan dan menyakiti orang lain.

BAB V

PENUTUP

5.1 Kesimpulan

Setelah dilakukan penelitian terhadap klasifikasi cyberbullying menggunakan metode Learning Vector Quantization dengan ekstraksi fitur Term Frequency-Inverse Document Frequency pada penggunaan ratio 80:20 dan learning rate 0.01 yang menghasilkan nilai akurasi terbaik yaitu 86.5%, presisi 89,6%, recall 86,8% dan f1-score 87%. Sedangkan untuk setiap kelas, pada kelas Religion, didapatkan nilai presisi 92%, recall 88% dan f1-score 90%. Pada kelas Age, nilai presisi 97%, recall 88% dan f1-score 92%. Kelas Gender, nilai presisi 100%, recall 68% dan f1-score 81%. Kelas Ethnicity nilai presisi 95%, recall 97% dan f1-score 96%. Pada kelas Not Bullying nilai presisi 64%, recall 93% dan f1-score 76%. Berdasarkan skenario ratio pembagian data dan learning rate pada skenario pengujian dapat diketahui bahwa pembagian data dan learning rate mempengaruhi hasil klasifikasi sistem dengan metode Learning Vector Quantization. Learning Rate yang terkecil yaitu 0.0001 menghasilkan akurasi rata-rata 85.6%, sedangkan learning rate yang terbesar yaitu 0.1 menghasilkan akurasi rata-rata 63.7%. Untuk ratio pembagian data 60:40 menghasilkan akurasi dengan rata-rata 78.9%, sedangkan untuk ratio pembagian data 80:20 menghasilkan akurasi dengan rata-rata 82.1%. Sehingga dapat disimpulkan bahwa nilai learning rate yang terlalu kecil atau terlalu besar dan pembagian data yang tidak optimal akan berpengaruh pada hasil klasifikasi.

5.2 Saran

Berdasarkan hasil percobaan pada penelitian ini, diharapkan penelitian selanjutnya dapat meningkatkan hasil klasifikasi yang lebih akurat. Oleh karena itu, penulis memiliki saran untuk pengembangan penelitian di masa yang akan datang, sebagai berikut:

1. Menambahkan dataset yang digunakan dari platform media sosial lain seperti Instagram atau Facebook agar data yang digunakan menjadi lebih bervariasi.
2. Melakukan percobaan menggunakan pendekatan pengolahan teks lainnya seperti metode klasifikasi berbasis deep learning.
3. Melakukan percobaan dengan menggunakan metode seleksi fitur yang berbeda seperti *Chi-Square*, *Information Gain*, *Mutual Information*, *Recursive Feature Elimination* dan sebagainya.

DAFTAR PUSTAKA

- Akbari, M. I. H. A. D., Noviantv, A., Setianingsih, C., Kusuma, D. A., Eryanto, F. M., Pratama, M. F. A., & Hanafia, S. R. (2022). Sentiment Analysis using *Learning vector quantization* Method. *2022 5th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, 277–282. <https://doi.org/10.1109/ISRITI56927.2022.10053040>
- Aninditya, A., Hasibuan, M. A., & Sutoyo, E. (2019). Text mining approach using TF-IDF and naive Bayes for classification of exam questions based on cognitive level of bloom's taxonomy. *2019 IEEE International Conference on Internet of Things and Intelligence System (IoT&IS)*, 112–117.
- Fahmy Amin, M. (2022). *Confusion matrix* in Binary Classification Problems: A Step-by-Step Tutorial. *Journal of Engineering Research*, 6(5), 0–0. <https://doi.org/10.21608/erjeng.2022.274526>
- Febriana, T., & Budiarto, A. (2019). Twitter Dataset for Hate Speech and *Cyberbullying* Detection in Indonesian Language. *2019 International Conference on Information Management and Technology (ICIMTech)*, 379–382. <https://doi.org/10.1109/ICIMTech.2019.8843722>
- Firmansyah, D., & Suryana, A. (2022). Konsep Pendidikan Akhlak: Kajian Tafsir Surat Al Hujurat Ayat 11-13. *Al-Mutharahah: Jurnal Penelitian dan Kajian Sosial Keagamaan*, 19(2), 58–82. <https://doi.org/10.46781/al-mutharahah.v19i2.538>
- Imamah, & Rachman, F. H. (2020). Twitter Sentiment Analysis of Covid-19 Using Term Weighting TF-IDF And Logistic Regresion. *2020 6th Information Technology International Seminar (ITIS)*, 238–242. <https://doi.org/10.1109/ITIS50118.2020.9320958>
- Janisar, A. A., Afzal, H., & Kumar, G. (2021). Identification of HATE speech tweets in Pashto language using Machine Learning techniques. *International Journal*, 10(3).
- Kohonen, T. (1990). Improved versions of *learning vector quantization*. *1990 IJCNN International Joint Conference on Neural Networks*, 545–550 vol.1. <https://doi.org/10.1109/IJCNN.1990.137622>
- Lestari, D. W. P., Perdana, R. S., & Adikara, P. P. (n.d.). *Klasifikasi Video Clickbait pada YouTube Berdasarkan Analisis Sentimen Komentar Menggunakan Learning vector quantization (LVQ) dan Lexicon-Based Features*.

- Luqyana, W. A., Cholissodin, I., & Perdana, R. S. (n.d.). *Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine*.
- M, S., & Kaliyamurthie, K. P. (2023). Effective prediction of fake news using a *learning vector quantization* with hamming distance measure. *Measurement: Sensors*, 25, 100601. <https://doi.org/10.1016/j.measen.2022.100601>
- Margono, H. (n.d.). *Analysis of the Indonesian Cyberbullying through Data Mining*:
- Nugroho, W. H., Handoyo, S., & Akri, Y. J. (2018). An Influence of Measurement Scale of Predictor Variable on Logistic Regression Modeling and Learning Vector Quntization Modeling for Object Classification. *International Journal of Electrical and Computer Engineering (IJECE)*, 8(1), 333. <https://doi.org/10.11591/ijece.v8i1.pp333-343>
- Oryza Habibie Rahman, Gunawan Abdillah, & Agus Komarudin. (2021). Klasifikasi Ujaran Kebencian pada Media Sosial Twitter Menggunakan Support Vector Machine. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 5(1), 17–23. <https://doi.org/10.29207/resti.v5i1.2700>
- Radjavani, A., & Bayu Sasongko, T. (2023). Analisa Perbandingan Algoritma CNN dan LSTM untuk Klasifikasi Pesan *Cyberbullying* pada Twitter. *Indonesian Journal of Computer Science*, 12(4). <https://doi.org/10.33022/ijcs.v12i4.3287>
- Riswanto, D., & Marsinun, R. (2020). Perilaku *Cyberbullying* Remaja di Media Sosial. *Analitika*, 12(2), 98–111. <https://doi.org/10.31289/analitika.v12i2.3704>
- Sarica, S., & Luo, J. (2021). Stopwords in technical language processing. *PLOS ONE*, 16(8), e0254937. <https://doi.org/10.1371/journal.pone.0254937>
- Song, K., Li, F., Hu, X., He, L., Niu, W., Lu, S., & Zhang, T. (2018). Multi-mode energy management strategy for fuel cell electric vehicles based on driving pattern identification using *learning vector quantization* neural network algorithm. *Journal of Power Sources*, 389, 230–239. <https://doi.org/10.1016/j.jpowsour.2018.04.024>
- Wang, A., & Potika, K. (2021). *Cyberbullying* Classification based on Social Network Analysis. *2021 IEEE Seventh International Conference on Big Data Computing Service and Applications (BigDataService)*. 2021 IEEE Seventh International Conference on Big Data Computing Service and Applications (BigDataService), Oxford, United Kingdom.

Zaenal, Z., Salim, Y., & Ilmawan, L. B. (2020). Analisis Sentimen terhadap Komentar Negatif di Media Sosial Facebook dengan Metode Klasifikasi Naïve Bayes. *Buletin Sistem Informasi dan Teknologi Islam*, 1(4), 259–265. <https://doi.org/10.33096/busiti.v1i4.666>

LAMPIRAN

Lampiran I Tabel Hasil klasifikasi cyberbullying ratio 60:40

No.	Learning Rate							
	0.1		0.01		0.001		0.0001	
	Aktual	Prediksi	Aktual	Prediksi	Aktual	Prediksi	Aktual	Prediksi
1	4	0	4	4	4	4	4	4
2	4	0	4	4	4	4	4	4
3	4	0	4	0	4	0	4	0
4	4	0	4	4	4	4	4	4
5	4	0	4	4	4	4	4	4
6	4	0	4	4	4	4	4	1
7	4	0	4	4	4	4	4	4
8	4	0	4	4	4	4	4	4
9	4	0	4	4	4	4	4	4
10	4	0	4	4	4	4	4	4
11	4	0	4	4	4	4	4	4
12	4	0	4	4	4	4	4	4
13	4	0	4	4	4	4	4	4
14	4	0	4	4	4	4	4	4
15	4	0	4	4	4	4	4	4
16	4	0	4	4	4	4	4	4
17	4	0	4	4	4	4	4	4
18	4	0	4	4	4	4	4	2
19	4	0	4	4	4	4	4	4
20	4	0	4	4	4	4	4	1
21	4	0	4	4	4	4	4	4
22	4	0	4	4	4	4	4	4
23	4	0	4	4	4	4	4	4
24	4	0	4	4	4	4	4	4
25	4	0	4	4	4	4	4	4
26	4	0	4	4	4	4	4	4
27	4	0	4	4	4	4	4	4
28	4	3	4	4	4	4	4	4
29	4	0	4	4	4	4	4	4
30	4	0	4	4	4	4	4	4
31	4	3	4	4	4	4	4	4
32	4	3	4	4	4	4	4	4
33	4	0	4	4	4	4	4	4
34	4	0	4	4	4	4	4	4
35	4	0	4	4	4	4	4	4
36	4	0	4	4	4	4	4	4
37	4	0	4	4	4	4	4	4
38	4	0	4	4	4	4	4	4
39	4	0	4	4	4	4	4	4
40	4	0	4	4	4	4	4	4
41	4	0	4	4	4	4	4	4
42	4	0	4	4	4	4	4	4
43	4	0	4	4	4	4	4	4
44	4	0	4	4	4	4	4	4
45	4	4	4	4	4	4	4	4
46	4	0	4	4	4	4	4	4
47	4	0	4	4	4	4	4	4
48	4	4	4	4	4	4	4	4
49	4	0	4	4	4	4	4	4

50	4	0	4	4	4	4	4	4
51	4	0	4	4	4	4	4	4
52	4	0	4	4	4	4	4	4
53	4	0	4	4	4	4	4	4
54	4	0	4	4	4	4	4	4
55	4	0	4	4	4	4	4	4
56	4	0	4	4	4	4	4	2
57	4	0	4	4	4	4	4	4
58	4	0	4	4	4	4	4	4
59	4	0	4	4	4	4	4	4
60	4	0	4	4	4	4	4	4
61	4	0	4	4	4	4	4	4
62	4	0	4	4	4	4	4	4
63	4	0	4	4	4	4	4	4
64	4	0	4	4	4	4	4	4
65	4	0	4	4	4	4	4	4
66	4	0	4	4	4	4	4	4
67	4	0	4	4	4	4	4	4
68	4	0	4	4	4	4	4	4
69	4	0	4	4	4	4	4	4
70	4	0	4	4	4	4	4	4
71	4	0	4	4	4	4	4	4
72	4	0	4	4	4	4	4	4
73	4	0	4	4	4	4	4	4
74	4	0	4	4	4	4	4	4
75	4	0	4	4	4	4	4	4
76	4	0	4	4	4	4	4	4
77	4	0	4	0	4	0	4	0
78	4	0	4	4	4	4	4	1
79	4	0	4	4	4	4	4	4
80	4	0	4	4	4	4	4	4
81	2	2	2	2	2	2	2	2
82	2	2	2	2	2	2	2	2
83	2	0	2	4	2	4	2	4
84	2	0	2	2	2	2	2	2
85	2	3	2	4	2	4	2	2
86	2	0	2	0	2	0	2	0
87	2	0	2	2	2	2	2	2
88	2	2	2	2	2	2	2	2
89	2	2	2	2	2	2	2	2
90	2	0	2	1	2	1	2	1
91	2	2	2	2	2	2	2	2
92	2	0	2	4	2	4	2	4
93	2	0	2	4	2	4	2	4
94	2	2	2	2	2	2	2	2
95	2	2	2	2	2	2	2	2
96	2	0	2	2	2	2	2	2
97	2	0	2	4	2	4	2	4
98	2	0	2	2	2	2	2	2
99	2	0	2	2	2	2	2	2
100	2	3	2	4	2	4	2	4
101	2	2	2	2	2	2	2	2
102	2	3	2	2	2	2	2	2

103	2	0	2	4	2	4	2	4
104	2	3	2	2	2	2	2	2
105	2	2	2	2	2	2	2	2
106	2	0	2	4	2	4	2	4
107	2	3	2	4	2	4	2	4
108	2	2	2	2	2	2	2	2
109	2	2	2	2	2	2	2	2
110	2	3	2	2	2	2	2	2
111	2	2	2	2	2	2	2	2
112	2	0	2	4	2	4	2	4
113	2	2	2	2	2	2	2	2
114	2	0	2	4	2	4	2	4
115	2	2	2	2	2	2	2	2
116	2	0	2	4	2	4	2	4
117	2	2	2	2	2	2	2	2
118	2	0	2	4	2	4	2	4
119	2	2	2	2	2	2	2	2
120	2	2	2	2	2	2	2	2
121	2	3	2	4	2	4	2	4
122	2	0	2	2	2	2	2	2
123	2	2	2	2	2	2	2	2
124	2	0	2	4	2	4	2	2
125	2	2	2	2	2	2	2	2
126	2	0	2	4	2	4	2	2
127	2	2	2	2	2	2	2	2
128	2	2	2	2	2	2	2	2
129	2	2	2	2	2	2	2	2
130	2	3	2	4	2	4	2	4
131	2	2	2	2	2	2	2	2
132	2	3	2	4	2	4	2	4
133	2	2	2	2	2	2	2	2
134	2	2	2	2	2	2	2	2
135	2	0	2	4	2	4	2	4
136	2	2	2	2	2	2	2	2
137	2	0	2	4	2	4	2	4
138	2	2	2	2	2	2	2	2
139	2	2	2	2	2	2	2	2
140	2	0	2	4	2	4	2	4
141	2	0	2	4	2	4	2	4
142	2	2	2	2	2	2	2	2
143	2	2	2	2	2	2	2	2
144	2	2	2	2	2	2	2	2
145	2	2	2	2	2	2	2	2
146	2	2	2	2	2	2	2	2
147	2	2	2	2	2	2	2	2
148	2	0	2	4	2	4	2	4
149	2	0	2	4	2	4	2	4
150	2	2	2	2	2	2	2	2
151	2	2	2	2	2	2	2	2
152	2	2	2	2	2	2	2	2
153	2	0	2	4	2	4	2	4
154	2	2	2	2	2	2	2	2
155	2	2	2	2	2	2	2	2

156	2	0	2	4	2	4	2	4
157	2	2	2	2	2	2	2	2
158	2	2	2	2	2	2	2	2
159	2	2	2	2	2	2	2	2
160	2	2	2	2	2	2	2	2
161	0	0	0	0	0	0	0	0
162	0	0	0	0	0	0	0	0
163	0	0	0	0	0	0	0	0
164	0	0	0	0	0	0	0	0
165	0	0	0	4	0	4	0	4
166	0	0	0	0	0	0	0	0
167	0	0	0	0	0	0	0	0
168	0	0	0	4	0	4	0	4
169	0	0	0	0	0	0	0	0
170	0	0	0	0	0	0	0	0
171	0	0	0	0	0	0	0	0
172	0	0	0	4	0	4	0	4
173	0	0	0	0	0	0	0	0
174	0	0	0	0	0	0	0	0
175	0	0	0	0	0	0	0	0
176	0	0	0	4	0	4	0	4
177	0	0	0	0	0	0	0	0
178	0	0	0	0	0	0	0	0
179	0	0	0	0	0	0	0	0
180	0	0	0	0	0	0	0	0
181	0	0	0	0	0	0	0	0
182	0	0	0	0	0	0	0	0
183	0	0	0	0	0	0	0	0
184	0	0	0	0	0	0	0	0
185	0	0	0	0	0	0	0	0
186	0	0	0	0	0	0	0	0
187	0	0	0	4	0	4	0	4
188	0	0	0	0	0	0	0	0
189	0	0	0	0	0	0	0	0
190	0	0	0	0	0	0	0	0
191	0	0	0	0	0	0	0	0
192	0	0	0	0	0	0	0	0
193	0	0	0	0	0	0	0	0
194	0	0	0	4	0	4	0	0
195	0	0	0	0	0	0	0	0
196	0	0	0	0	0	0	0	0
197	0	0	0	0	0	0	0	0
198	0	0	0	4	0	4	0	4
199	0	0	0	0	0	0	0	0
200	0	0	0	0	0	0	0	0
201	0	0	0	0	0	0	0	0
202	0	0	0	0	0	0	0	0
203	0	0	0	0	0	0	0	0
204	0	0	0	0	0	0	0	0
205	0	0	0	0	0	0	0	0
206	0	0	0	0	0	0	0	0
207	0	0	0	0	0	0	0	0
208	0	0	0	0	0	0	0	0

209	0	0	0	0	0	0	0	0
210	0	0	0	0	0	0	0	0
211	0	0	0	4	0	4	0	4
212	0	0	0	0	0	0	0	0
213	0	0	0	0	0	0	0	0
214	0	0	0	0	0	0	0	0
215	0	0	0	4	0	4	0	2
216	0	0	0	0	0	0	0	0
217	0	0	0	0	0	0	0	0
218	0	0	0	0	0	0	0	0
219	0	0	0	0	0	0	0	0
220	0	0	0	0	0	0	0	0
221	0	0	0	0	0	0	0	0
222	0	0	0	0	0	0	0	0
223	0	0	0	0	0	0	0	0
224	0	0	0	0	0	0	0	0
225	0	0	0	0	0	0	0	0
226	0	0	0	0	0	0	0	0
227	0	0	0	0	0	0	0	0
228	0	0	0	0	0	0	0	0
229	0	0	0	0	0	0	0	0
230	0	0	0	0	0	0	0	0
231	0	0	0	0	0	0	0	0
232	0	0	0	4	0	4	0	4
233	0	0	0	0	0	0	0	0
234	0	0	0	0	0	0	0	0
235	0	0	0	0	0	4	0	4
236	0	0	0	0	0	0	0	0
237	0	0	0	0	0	0	0	0
238	0	0	0	0	0	0	0	0
239	0	0	0	4	0	4	0	4
240	0	0	0	0	0	0	0	0
241	1	1	1	1	1	1	1	1
242	1	1	1	1	1	1	1	1
243	1	0	1	1	1	1	1	1
244	1	1	1	1	1	1	1	1
245	1	1	1	4	1	4	1	1
246	1	1	1	1	1	1	1	1
247	1	1	1	1	1	1	1	1
248	1	0	1	0	1	0	1	0
249	1	1	1	1	1	1	1	1
250	1	1	1	1	1	1	1	1
251	1	0	1	4	1	4	1	1
252	1	1	1	1	1	1	1	1
253	1	1	1	1	1	1	1	1
254	1	1	1	1	1	1	1	1
255	1	1	1	1	1	1	1	1
256	1	1	1	1	1	1	1	1
257	1	1	1	1	1	1	1	1
258	1	1	1	1	1	1	1	1
259	1	1	1	1	1	1	1	1
260	1	1	1	1	1	1	1	1
261	1	0	1	1	1	1	1	1

262	1	1	1	1	1	1	1	1
263	1	1	1	1	1	1	1	1
264	1	1	1	1	1	1	1	1
265	1	0	1	4	1	4	1	1
266	1	2	1	2	1	2	1	2
267	1	1	1	1	1	1	1	1
268	1	1	1	1	1	1	1	1
269	1	0	1	4	1	4	1	1
270	1	1	1	1	1	1	1	1
271	1	1	1	1	1	1	1	1
272	1	1	1	1	1	1	1	1
273	1	1	1	1	1	1	1	1
274	1	1	1	1	1	1	1	1
275	1	0	1	1	1	1	1	1
276	1	0	1	1	1	1	1	1
277	1	1	1	1	1	1	1	1
278	1	0	1	1	1	1	1	1
279	1	1	1	1	1	1	1	1
280	1	1	1	1	1	1	1	1
281	1	0	1	4	1	4	1	4
282	1	1	1	1	1	1	1	1
283	1	1	1	1	1	1	1	1
284	1	1	1	1	1	1	1	1
285	1	1	1	1	1	1	1	1
286	1	1	1	1	1	1	1	1
287	1	1	1	1	1	1	1	1
288	1	1	1	1	1	1	1	1
289	1	1	1	1	1	1	1	1
290	1	1	1	1	1	1	1	1
291	1	1	1	1	1	1	1	1
292	1	1	1	1	1	1	1	1
293	1	1	1	1	1	1	1	1
294	1	1	1	1	1	1	1	1
295	1	1	1	1	1	1	1	1
296	1	0	1	4	1	4	1	4
297	1	1	1	1	1	1	1	1
298	1	1	1	1	1	1	1	1
299	1	1	1	1	1	1	1	1
300	1	1	1	1	1	1	1	1
301	1	1	1	1	1	1	1	1
302	1	1	1	1	1	1	1	1
303	1	1	1	1	1	1	1	1
304	1	0	1	1	1	1	1	1
305	1	1	1	1	1	1	1	1
306	1	1	1	1	1	1	1	1
307	1	1	1	1	1	1	1	1
308	1	0	1	4	1	4	1	4
309	1	0	1	1	1	1	1	1
310	1	1	1	1	1	1	1	1
311	1	1	1	1	1	1	1	1
312	1	1	1	1	1	1	1	1
313	1	1	1	1	1	1	1	1
314	1	1	1	1	1	1	1	1

315	1	1	1	1	1	1	1	1
316	1	0	1	1	1	1	1	1
317	1	1	1	1	1	1	1	1
318	1	0	1	1	1	1	1	1
319	1	1	1	1	1	1	1	1
320	1	0	1	0	1	0	1	0
321	3	3	3	3	3	3	3	3
322	3	0	3	3	3	3	3	3
323	3	3	3	4	3	4	3	4
324	3	3	3	3	3	3	3	3
325	3	3	3	3	3	3	3	3
326	3	0	3	3	3	3	3	3
327	3	3	3	3	3	3	3	3
328	3	3	3	3	3	3	3	3
329	3	3	3	3	3	3	3	3
330	3	3	3	3	3	3	3	3
331	3	3	3	3	3	3	3	3
332	3	3	3	3	3	3	3	3
333	3	3	3	3	3	3	3	3
334	3	3	3	3	3	3	3	3
335	3	3	3	3	3	3	3	3
336	3	3	3	3	3	3	3	3
337	3	3	3	3	3	3	3	3
338	3	3	3	4	3	4	3	3
339	3	3	3	3	3	3	3	3
340	3	3	3	3	3	3	3	3
341	3	3	3	3	3	3	3	3
342	3	3	3	3	3	3	3	3
343	3	3	3	3	3	3	3	3
344	3	3	3	3	3	3	3	3
345	3	3	3	3	3	3	3	3
346	3	3	3	3	3	3	3	3
347	3	3	3	3	3	3	3	3
348	3	3	3	3	3	3	3	3
349	3	3	3	3	3	3	3	3
350	3	3	3	3	3	3	3	3
351	3	3	3	3	3	3	3	3
352	3	0	3	3	3	4	3	3
353	3	0	3	4	3	4	3	4
354	3	3	3	3	3	3	3	3
355	3	3	3	3	3	3	3	3
356	3	3	3	3	3	3	3	3
357	3	3	3	3	3	3	3	3
358	3	0	3	3	3	3	3	3
359	3	3	3	3	3	3	3	3
360	3	3	3	3	3	3	3	3
361	3	3	3	3	3	3	3	3
362	3	3	3	3	3	3	3	3
363	3	3	3	3	3	3	3	3
364	3	3	3	3	3	3	3	3
365	3	3	3	3	3	3	3	3
366	3	0	3	4	3	4	3	4
367	3	3	3	3	3	3	3	3

368	3	0	3	1	3	4	3	1
369	3	0	3	3	3	3	3	3
370	3	3	3	3	3	3	3	3
371	3	3	3	3	3	3	3	3
372	3	0	3	3	3	3	3	3
373	3	0	3	0	3	3	3	3
374	3	3	3	3	3	3	3	3
375	3	3	3	3	3	3	3	3
376	3	3	3	3	3	3	3	3
377	3	0	3	4	3	4	3	4
378	3	3	3	3	3	3	3	3
379	3	3	3	3	3	3	3	3
380	3	3	3	3	3	3	3	3
381	3	3	3	4	3	4	3	3
382	3	3	3	3	3	3	3	3
383	3	3	3	3	3	3	3	3
384	3	3	3	3	3	3	3	3
385	3	3	3	3	3	3	3	3
386	3	0	3	0	3	4	3	0
387	3	3	3	3	3	3	3	3
388	3	0	3	4	3	4	3	4
389	3	3	3	3	3	3	3	3
390	3	3	3	3	3	3	3	3
391	3	0	3	4	3	4	3	3
392	3	3	3	3	3	3	3	3
393	3	0	3	4	3	4	3	3
394	3	0	3	3	3	3	3	3
395	3	3	3	3	3	3	3	3
396	3	0	3	4	3	4	3	3
397	3	3	3	3	3	3	3	3
398	3	3	3	3	3	3	3	3
399	3	3	3	3	3	3	3	3
400	3	3	3	3	3	3	3	3
388	3	3	3	3	3	3	3	3
389	3	0	3	1	3	4	3	1
390	3	0	3	3	3	3	3	3
391	3	3	3	3	3	3	3	3
392	3	3	3	3	3	3	3	3
393	3	0	3	3	3	3	3	3
394	3	0	3	0	3	3	3	3
395	3	3	3	3	3	3	3	3
396	3	3	3	3	3	3	3	3
397	3	3	3	3	3	3	3	3
398	3	0	3	4	3	4	3	4
399	3	3	3	3	3	3	3	3
400	3	3	3	3	3	3	3	3

Lampiran II Tabel Hasil klasifikasi cyberbullying ratio 70:30

No.	Learning Rate							
	0.1		0.01		0.001		0.0001	
	Aktual	Prediksi	Aktual	Prediksi	Aktual	Prediksi	Aktual	Prediksi
1	4	4	4	4	4	4	4	4
2	4	2	4	4	4	4	4	4
3	4	2	4	0	4	0	4	0
4	4	2	4	4	4	4	4	4
5	4	4	4	4	4	4	4	4
6	4	2	4	4	4	4	4	4
7	4	2	4	4	4	4	4	4
8	4	2	4	4	4	4	4	4
9	4	2	4	4	4	4	4	4
10	4	2	4	4	4	4	4	4
11	4	2	4	4	4	4	4	4
12	4	2	4	4	4	4	4	4
13	4	2	4	4	4	4	4	4
14	4	2	4	4	4	4	4	4
15	4	2	4	1	4	1	4	1
16	4	2	4	4	4	4	4	4
17	4	2	4	4	4	4	4	4
18	4	2	4	4	4	4	4	4
19	4	2	4	4	4	4	4	4
20	4	2	4	4	4	4	4	4
21	4	4	4	4	4	4	4	4
22	4	2	4	4	4	4	4	4
23	4	2	4	4	4	4	4	4
24	4	2	4	4	4	4	4	4
25	4	2	4	4	4	4	4	4
26	4	2	4	4	4	4	4	4
27	4	4	4	4	4	4	4	4
28	4	2	4	4	4	4	4	4
29	4	4	4	4	4	4	4	4
30	4	2	4	4	4	4	4	4
31	4	2	4	4	4	4	4	4
32	4	2	4	4	4	4	4	4
33	4	4	4	4	4	4	4	4
34	4	4	4	4	4	4	4	4
35	4	2	4	4	4	4	4	4
36	4	4	4	4	4	4	4	4
37	4	2	4	4	4	4	4	4
38	4	2	4	4	4	4	4	4
39	4	2	4	4	4	4	4	4
40	4	2	4	4	4	4	4	4
41	4	2	4	4	4	4	4	4
42	4	4	4	4	4	4	4	4
43	4	2	4	4	4	4	4	4
44	4	2	4	4	4	4	4	4
45	4	2	4	4	4	4	4	4
46	4	2	4	4	4	4	4	4
47	4	2	4	4	4	4	4	4
48	4	2	4	4	4	4	4	4
49	4	2	4	4	4	4	4	4

50	4	2	4	4	4	4	4	4
51	4	2	4	4	4	4	4	4
52	4	4	4	4	4	4	4	4
53	4	2	4	4	4	4	4	4
54	4	2	4	4	4	4	4	4
55	4	2	4	4	4	4	4	4
56	4	2	4	4	4	4	4	4
57	4	2	4	0	4	0	4	0
58	4	2	4	1	4	4	4	1
59	4	2	4	4	4	4	4	4
60	4	2	4	4	4	4	4	4
61	2	2	2	2	2	2	2	2
62	2	2	2	2	2	2	2	2
63	2	2	2	0	2	0	2	0
64	2	2	2	2	2	2	2	2
65	2	2	2	2	2	2	2	2
66	2	2	2	2	2	2	2	2
67	2	4	2	4	2	4	2	4
68	2	2	2	4	2	4	2	4
69	2	2	2	4	2	4	2	4
70	2	2	2	2	2	2	2	2
71	2	2	2	2	2	2	2	2
72	2	4	2	4	2	4	2	4
73	2	2	2	2	2	2	2	2
74	2	2	2	4	2	4	2	4
75	2	2	2	2	2	2	2	2
76	2	2	2	2	2	2	2	2
77	2	2	2	2	2	2	2	2
78	2	2	2	2	2	2	2	2
79	2	2	2	4	2	4	2	4
80	2	2	2	2	2	2	2	2
81	2	2	2	2	2	2	2	2
82	2	2	2	4	2	4	2	4
83	2	2	2	2	2	2	2	2
84	2	2	2	4	2	4	2	4
85	2	2	2	2	2	2	2	2
86	2	4	2	4	2	4	2	4
87	2	2	2	2	2	2	2	2
88	2	4	2	4	2	4	2	4
89	2	2	2	2	2	2	2	2
90	2	2	2	2	2	2	2	2
91	2	2	2	4	2	4	2	4
92	2	2	2	2	2	2	2	2
93	2	2	2	2	2	2	2	2
94	2	2	2	4	2	4	2	4
95	2	2	2	2	2	2	2	2
96	2	2	2	4	2	4	2	2
97	2	2	2	2	2	2	2	2
98	2	2	2	2	2	2	2	2
99	2	2	2	2	2	2	2	2
100	2	4	2	4	2	4	2	4
101	2	2	2	4	2	4	2	4
102	2	2	2	2	2	2	2	2

103	2	2	2	2	2	2	2	2
104	2	2	2	2	2	2	2	2
105	2	2	2	2	2	2	2	2
106	2	4	2	4	2	4	2	4
107	2	2	2	2	2	2	2	2
108	2	2	2	2	2	2	2	2
109	2	2	2	2	2	2	2	2
110	2	2	2	2	2	2	2	2
111	2	2	2	2	2	2	2	2
112	2	2	2	4	2	4	2	4
113	2	2	2	2	2	2	2	2
114	2	2	2	2	2	2	2	2
115	2	2	2	4	2	4	2	4
116	2	2	2	2	2	2	2	2
117	2	2	2	2	2	2	2	2
118	2	4	2	4	2	4	2	4
119	2	2	2	2	2	2	2	2
120	2	2	2	2	2	2	2	2
121	0	2	0	0	0	0	0	4
122	0	2	0	0	0	0	0	0
123	0	2	0	0	0	0	0	0
124	0	2	0	4	0	4	0	4
125	0	0	0	0	0	0	0	0
126	0	0	0	0	0	0	0	0
127	0	2	0	4	0	4	0	4
128	0	0	0	0	0	0	0	0
129	0	2	0	4	0	4	0	4
130	0	0	0	0	0	0	0	0
131	0	2	0	0	0	0	0	0
132	0	0	0	0	0	0	0	0
133	0	2	0	0	0	0	0	0
134	0	0	0	0	0	0	0	0
135	0	2	0	0	0	0	0	0
136	0	0	0	0	0	0	0	0
137	0	2	0	0	0	0	0	0
138	0	0	0	0	0	0	0	0
139	0	2	0	4	0	4	0	4
140	0	0	0	0	0	0	0	0
141	0	0	0	0	0	0	0	0
142	0	0	0	0	0	0	0	0
143	0	2	0	0	0	0	0	0
144	0	0	0	0	0	0	0	0
145	0	2	0	0	0	0	0	0
146	0	2	0	0	0	0	0	0
147	0	0	0	0	0	0	0	0
148	0	0	0	0	0	0	0	0
149	0	2	0	4	0	4	0	4
150	0	0	0	0	0	0	0	0
151	0	0	0	0	0	0	0	0
152	0	0	0	0	0	0	0	0
153	0	0	0	0	0	0	0	0
154	0	2	0	0	0	0	0	0
155	0	0	0	0	0	0	0	0

156	0	0	0	0	0	0	0	0
157	0	0	0	0	0	0	0	0
158	0	0	0	0	0	0	0	0
159	0	2	0	4	0	4	0	4
160	0	0	0	0	0	0	0	0
161	0	2	0	0	0	0	0	0
162	0	0	0	0	0	0	0	0
163	0	0	0	0	0	0	0	0
164	0	0	0	0	0	0	0	0
165	0	0	0	0	0	0	0	0
166	0	0	0	0	0	0	0	0
167	0	2	0	0	0	0	0	0
168	0	2	0	0	0	0	0	0
169	0	2	0	0	0	0	0	0
170	0	2	0	0	0	0	0	0
171	0	2	0	0	0	0	0	0
172	0	2	0	0	0	4	0	0
173	0	2	0	0	0	4	0	0
174	0	2	0	0	0	0	0	0
175	0	2	0	0	0	4	0	4
176	0	0	0	0	0	0	0	0
177	0	0	0	0	0	0	0	0
178	0	2	0	0	0	0	0	0
179	0	2	0	4	0	4	0	4
180	0	0	0	0	0	0	0	0
181	1	1	1	1	1	1	1	1
182	1	1	1	1	1	1	1	1
183	1	2	1	1	1	1	1	1
184	1	1	1	1	1	1	1	1
185	1	1	1	1	1	1	1	1
186	1	1	1	1	1	1	1	1
187	1	1	1	1	1	1	1	1
188	1	1	1	1	1	1	1	1
189	1	2	1	1	1	4	1	1
190	1	1	1	1	1	1	1	1
191	1	1	1	1	1	1	1	1
192	1	2	1	1	1	1	1	1
193	1	1	1	1	1	1	1	1
194	1	2	1	1	1	1	1	1
195	1	1	1	1	1	1	1	1
196	1	1	1	1	1	1	1	1
197	1	2	1	1	1	1	1	1
198	1	1	1	1	1	1	1	1
199	1	2	1	1	1	4	1	1
200	1	1	1	1	1	1	1	1
201	1	2	1	1	1	1	1	1
202	1	2	1	1	1	1	1	1
203	1	1	1	1	1	1	1	1
204	1	1	1	1	1	1	1	1
205	1	1	1	1	1	1	1	1
206	1	1	1	1	1	1	1	1
207	1	2	1	1	1	1	1	1
208	1	1	1	1	1	1	1	1

209	1	2	1	1	1	1	1	1
210	1	1	1	1	1	1	1	1
211	1	1	1	1	1	1	1	1
212	1	2	1	4	1	4	1	4
213	1	1	1	1	1	1	1	1
214	1	1	1	1	1	1	1	1
215	1	1	1	1	1	1	1	1
216	1	1	1	1	1	1	1	1
217	1	1	1	1	1	1	1	1
218	1	1	1	1	1	1	1	1
219	1	2	1	1	1	1	1	1
220	1	1	1	1	1	1	1	1
221	1	1	1	1	1	1	1	1
222	1	1	1	1	1	1	1	1
223	1	1	1	1	1	1	1	1
224	1	1	1	1	1	1	1	1
225	1	1	1	1	1	1	1	1
226	1	1	1	1	1	1	1	1
227	1	2	1	1	1	1	1	1
228	1	1	1	1	1	1	1	1
229	1	2	1	1	1	1	1	1
230	1	1	1	1	1	1	1	1
231	1	2	1	1	1	1	1	1
232	1	1	1	1	1	1	1	1
233	1	1	1	1	1	1	1	1
234	1	2	1	4	1	4	1	4
235	1	1	1	1	1	1	1	1
236	1	2	1	1	1	1	1	1
237	1	1	1	1	1	1	1	1
238	1	2	1	1	1	1	1	1
239	1	1	1	1	1	1	1	1
240	1	2	1	0	1	0	1	0
241	3	3	3	3	3	3	3	3
242	3	2	3	3	3	3	3	3
243	3	3	3	3	3	3	3	3
244	3	3	3	3	3	3	3	3
245	3	3	3	3	3	3	3	3
246	3	3	3	3	3	3	3	3
247	3	3	3	3	3	3	3	3
248	3	3	3	3	3	3	3	3
249	3	3	3	3	3	3	3	3
250	3	3	3	3	3	3	3	3
251	3	2	3	3	3	3	3	3
252	3	3	3	3	3	3	3	3
253	3	3	3	3	3	3	3	3
254	3	3	3	3	3	3	3	3
255	3	3	3	3	3	3	3	3
256	3	3	3	3	3	3	3	3
257	3	2	3	3	3	3	3	3
258	3	3	3	3	3	3	3	3
259	3	3	3	3	3	3	3	3
260	3	3	3	3	3	3	3	3
261	3	3	3	3	3	3	3	3

262	3	2	3	4	3	4	3	4
263	3	2	3	3	3	3	3	3
264	3	3	3	3	3	3	3	3
265	3	2	3	3	3	3	3	3
266	3	3	3	3	3	3	3	3
267	3	3	3	3	3	3	3	3
268	3	3	3	3	3	3	3	3
269	3	2	3	4	3	4	3	4
270	3	3	3	3	3	3	3	3
271	3	2	3	1	3	1	3	1
272	3	2	3	3	3	3	3	3
273	3	3	3	3	3	3	3	3
274	3	3	3	3	3	3	3	3
275	3	2	3	3	3	3	3	3
276	3	2	3	3	3	3	3	3
277	3	3	3	3	3	3	3	3
278	3	2	3	3	3	3	3	3
279	3	2	3	3	3	3	3	3
280	3	2	3	4	3	4	3	4
281	3	3	3	3	3	3	3	3
282	3	3	3	3	3	3	3	3
283	3	3	3	3	3	3	3	3
284	3	2	3	3	3	3	3	3
285	3	3	3	3	3	3	3	3
286	3	3	3	3	3	3	3	3
287	3	3	3	3	3	3	3	3
288	3	3	3	3	3	3	3	3
289	3	2	3	0	3	4	3	0
290	3	2	3	4	3	4	3	4
291	3	2	3	3	3	3	3	3
292	3	2	3	0	3	4	3	3
293	3	3	3	3	3	3	3	3
294	3	2	3	3	3	4	3	3
295	3	2	3	3	3	3	3	3
296	3	2	3	3	3	3	3	3
297	3	2	3	3	3	3	3	3
298	3	3	3	3	3	3	3	3
299	3	3	3	3	3	3	3	3
300	3	3	3	3	3	3	3	3

Lampiran III Tabel Hasil klasifikasi cyberbullying ratio 80:20

No.	Learning Rate							
	0.1		0.01		0.001		0.0001	
	Aktual	Prediksi	Aktual	Prediksi	Aktual	Prediksi	Aktual	Prediksi
1	4	0	4	4	4	4	4	4
2	4	0	4	0	4	0	4	0
3	4	0	4	4	4	4	4	4
4	4	0	4	4	4	4	4	4
5	4	0	4	4	4	4	4	4
6	4	3	4	4	4	4	4	4
7	4	3	4	4	4	4	4	4
8	4	0	4	4	4	4	4	4

9	4	0	4	4	4	4	4	4
10	4	0	4	4	4	4	4	4
11	4	0	4	4	4	4	4	4
12	4	0	4	4	4	4	4	4
13	4	0	4	4	4	4	4	4
14	4	4	4	4	4	4	4	4
15	4	0	4	4	4	4	4	4
16	4	0	4	4	4	4	4	4
17	4	0	4	4	4	4	4	4
18	4	0	4	4	4	4	4	4
19	4	0	4	4	4	4	4	4
20	4	0	4	4	4	4	4	4
21	4	0	4	4	4	4	4	4
22	4	1	4	4	4	4	4	4
23	4	0	4	4	4	4	4	4
24	4	0	4	4	4	4	4	4
25	4	0	4	4	4	4	4	4
26	4	4	4	4	4	4	4	4
27	4	0	4	4	4	4	4	4
28	4	0	4	4	4	4	4	4
29	4	0	4	4	4	4	4	4
30	4	0	4	4	4	4	4	4
31	4	0	4	4	4	4	4	4
32	4	3	4	4	4	4	4	4
33	4	0	4	4	4	4	4	4
34	4	0	4	4	4	4	4	4
35	4	0	4	4	4	4	4	4
36	4	0	4	4	4	4	4	4
37	4	0	4	4	4	4	4	4
38	4	0	4	3	4	3	4	3
39	4	0	4	1	4	4	4	1
40	4	0	4	4	4	4	4	4
41	2	2	2	2	2	2	2	2
42	2	0	2	3	2	3	2	3
43	2	2	2	2	2	2	2	2
44	2	2	2	2	2	2	2	2
45	2	2	2	2	2	2	2	2
46	2	0	2	4	2	4	2	4
47	2	2	2	2	2	2	2	2
48	2	2	2	2	2	2	2	2
49	2	3	2	4	2	4	2	4
50	2	2	2	2	2	2	2	2
51	2	3	2	2	2	2	2	2
52	2	2	2	2	2	2	2	2
53	2	2	2	2	2	2	2	2
54	2	0	2	4	2	4	2	4
55	2	2	2	2	2	2	2	2
56	2	0	2	4	2	4	2	4
57	2	2	2	2	2	2	2	2
58	2	3	2	4	2	4	2	4
59	2	2	2	2	2	2	2	2
60	2	2	2	2	2	2	2	2
61	2	2	2	2	2	2	2	2

62	2	0	2	4	2	4	2	4
63	2	2	2	2	2	2	2	2
64	2	0	2	0	2	4	2	4
65	2	2	2	2	2	2	2	2
66	2	2	2	2	2	2	2	2
67	2	2	2	2	2	2	2	2
68	2	4	2	4	2	4	2	4
69	2	0	2	4	2	4	2	4
70	2	2	2	2	2	2	2	2
71	2	2	2	2	2	2	2	2
72	2	2	2	2	2	2	2	2
73	2	2	2	2	2	2	2	2
74	2	0	2	4	2	4	2	4
75	2	2	2	2	2	2	2	2
76	2	0	2	4	2	4	2	4
77	2	2	2	2	2	2	2	2
78	2	2	2	2	2	2	2	2
79	2	2	2	2	2	2	2	2
80	2	0	2	4	2	4	2	4
81	0	0	0	0	0	0	0	0
82	0	0	0	0	0	0	0	0
83	0	0	0	4	0	4	0	4
84	0	0	0	0	0	0	0	0
85	0	0	0	0	0	0	0	0
86	0	0	0	0	0	0	0	0
87	0	0	0	0	0	0	0	0
88	0	0	0	0	0	0	0	0
89	0	0	0	0	0	0	0	0
90	0	0	0	0	0	0	0	0
91	0	0	0	0	0	0	0	0
92	0	0	0	4	0	4	0	4
93	0	0	0	0	0	0	0	0
94	0	0	0	0	0	0	0	0
95	0	0	0	0	0	0	0	0
96	0	0	0	0	0	0	0	0
97	0	0	0	4	0	4	0	4
98	0	0	0	0	0	0	0	0
99	0	0	0	0	0	0	0	0
100	0	0	0	0	0	0	0	0
101	0	0	0	0	0	0	0	0
102	0	0	0	0	0	0	0	0
103	0	0	0	0	0	0	0	0
104	0	0	0	0	0	0	0	0
105	0	0	0	0	0	0	0	0
106	0	0	0	4	0	4	0	4
107	0	0	0	0	0	0	0	0
108	0	0	0	0	0	0	0	0
109	0	0	0	0	0	0	0	0
110	0	0	0	0	0	0	0	0
111	0	0	0	0	0	0	0	0
112	0	0	0	0	0	0	0	0
113	0	0	0	0	0	0	0	0
114	0	0	0	0	0	0	0	0

115	0	0	0	0	0	0	0	0
116	0	0	0	0	0	0	0	0
117	0	0	0	0	0	0	0	0
118	0	3	0	0	0	0	0	0
119	0	0	0	4	0	4	0	4
120	0	0	0	0	0	0	0	0
121	1	1	1	1	1	1	1	1
122	1	1	1	1	1	1	1	1
123	1	1	1	1	1	1	1	1
124	1	1	1	1	1	1	1	1
125	1	0	1	4	1	4	1	4
126	1	1	1	1	1	1	1	1
127	1	1	1	1	1	1	1	1
128	1	1	1	1	1	1	1	1
129	1	1	1	1	1	1	1	1
130	1	1	1	1	1	1	1	1
131	1	0	1	4	1	4	1	4
132	1	1	1	1	1	1	1	1
133	1	1	1	1	1	1	1	1
134	1	0	1	1	1	1	1	1
135	1	1	1	1	1	1	1	1
136	1	1	1	1	1	1	1	1
137	1	1	1	1	1	1	1	1
138	1	1	1	1	1	1	1	1
139	1	1	1	1	1	1	1	1
140	1	0	1	4	1	4	1	4
141	1	1	1	1	1	1	1	1
142	1	1	1	1	1	1	1	1
143	1	1	1	1	1	1	1	1
144	1	1	1	1	1	1	1	1
145	1	1	1	1	1	1	1	1
146	1	1	1	1	1	1	1	1
147	1	1	1	1	1	1	1	1
148	1	1	1	1	1	1	1	1
149	1	0	1	1	1	1	1	1
150	1	1	1	1	1	1	1	1
151	1	1	1	1	1	1	1	1
152	1	0	1	1	1	1	1	1
153	1	3	1	1	1	1	1	1
154	1	1	1	1	1	1	1	1
155	1	3	1	4	1	4	1	4
156	1	1	1	1	1	1	1	1
157	1	1	1	1	1	1	1	1
158	1	1	1	1	1	1	1	1
159	1	1	1	1	1	1	1	1
160	1	3	1	0	1	0	1	0
161	3	3	3	3	3	3	3	3
162	3	3	3	3	3	3	3	3
163	3	3	3	3	3	3	3	3
164	3	3	3	3	3	3	3	3
165	3	3	3	3	3	3	3	3
166	3	3	3	3	3	3	3	3
167	3	3	3	3	3	3	3	3

168	3	3	3	3	3	3	3	3
169	3	3	3	3	3	3	3	3
170	3	3	3	3	3	3	3	3
171	3	3	3	3	3	3	3	3
172	3	3	3	3	3	3	3	3
173	3	3	3	3	3	3	3	3
174	3	3	3	3	3	3	3	3
175	3	3	3	3	3	3	3	3
176	3	0	3	3	3	3	3	3
177	3	3	3	3	3	3	3	3
178	3	3	3	3	3	3	3	3
179	3	0	3	3	3	4	3	4
180	3	3	3	3	3	3	3	3
181	3	0	3	3	3	1	3	1
182	3	0	3	3	3	3	3	3
183	3	3	3	3	3	3	3	3
184	3	3	3	3	3	3	3	3
185	3	3	3	3	3	3	3	3
186	3	3	3	3	3	3	3	3
187	3	0	3	3	3	4	3	4
188	3	3	3	3	3	3	3	3
189	3	3	3	3	3	3	3	3
190	3	3	3	3	3	3	3	3
191	3	3	3	3	3	3	3	3
192	3	3	3	3	3	3	3	3
193	3	3	3	3	3	3	3	3
194	3	4	3	4	3	4	3	4
195	3	3	3	3	3	3	3	3
196	3	3	3	3	3	3	3	3
197	3	3	3	3	3	3	3	3
198	3	3	3	3	3	3	3	3
199	3	3	3	3	3	3	3	3
200	3	3	3	3	3	3	3	3