

**KLASIFIKASI KECELAKAAN KERJA PADA BERITA *ONLINE*
MENGUNAKAN *MULTINOMIAL NAÏVE BAYES***

SKRIPSI

Oleh:
QORINA SETYANINGRUM
NIM. 200605110120



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

**KLASIFIKASI KECELAKAAN KERJA PADA BERITA *ONLINE*
MENGUNAKAN *MULTINOMIAL NAÏVE BAYES***

SKRIPSI

Diajukan kepada:
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Komputer (S.Kom)

Oleh:
QORINA SETYANINGRUM
NIM. 200605110120

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

HALAMAN PERSETUJUAN

**KLASIFIKASI KECELAKAAN KERJA PADA BERITA *ONLINE*
MENGUNAKAN *MULTINOMIAL NAÏVE BAYES***

SKRIPSI

Oleh:

OORINA SETYANINGRUM

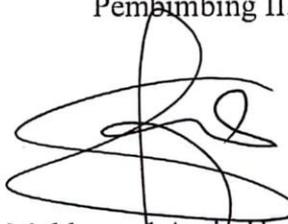
NIM. 200605110120

Telah Diperiksa dan Disetujui untuk Diuji:
Tanggal: 5 Juni 2024

Pembimbing I,


Dr. Cahyo Crys dian, MCS
NIP. 19740424 200901 1 008

Pembimbing II,


Dr. Mokhamad Amin Hariyadi, M.T
NIP. 19670118 200501 1 001

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




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

HALAMAN PENGESAHAN

**KLASIFIKASI KECELAKAAN KERJA PADA BERITA *ONLINE*
MENGUNAKAN *MULTINOMIAL NAÏVE BAYES***

SKRIPSI

Oleh:

OORINA SETYANINGRUM

NIM. 200605110120

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Tanggal: 19 Juni 2024

Susunan Dewan Penguji

Ketua Penguji : Dr. Irwan Budi Santoso, M.Kom
NIP. 19770103 201101 1 004

Anggota Penguji I : Roro Inda Melani, M.T, M.Sc
NIP. 19780925 200501 2 008

Anggota Penguji II : Dr. Cahyo Crysdian, MCS
NIP. 19740424 200901 1 008

Anggota Penguji III : Dr. Mokhamad Amin Hariyadi, M.T
NIP. 19670118 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, IPM
NIP. 19771020 200912 1 001

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Qorina Setyaningrum
NIM : 200605110120
Fakultas / Program Studi : Sains dan Teknologi / Teknik Informatika
Judul Skripsi : Klasifikasi Kecelakaan Kerja pada Berita *Online*
Menggunakan *Multinomial Naïve Bayes*

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, 24 Juni 2024

Yang membuat pernyataan,



Qorina Setyaningrum

NIM. 200605110120

MOTTO

“Mintalah Semua Ketidakmungkinan itu pada Allah SWT.”

HALAMAN PERSEMBAHAN

Puji Syukur atas kehadiran Allah Subhanahu wa ta'ala, karena berkat rahmat dan petunjuk-Nya penulis dapat menyelesaikan skripsi ini. Shalawat serta salam kepada Rasulullah Shallallahu 'alaihi wasallam, yang telah membawa kita dari zaman *jahiliyah* menuju *addinul Islam*.

Penulis mempersembahkan karya ini kepada kedua orang tua, kerabat, teman, kakak tingkat, dan sahabat yang telah menemani perkuliahan daring maupun luring. Tanpa bantuan mereka penulis tidak akan bisa sampai di titik ini.

KATA PENGANTAR

Assalamualaikum wr wb.

Puji syukur penulis panjatkan kepada kehadiran Allah Swt yang telah melimpahkan nikmat serta karunia-Nya sehingga penulis mampu menyelesaikan penulisan Skripsi yang berjudul “Klasifikasi Kecelakaan Kerja pada Berita *Online* Menggunakan *Multinomial Naïve Bayes*” dengan baik dan tepat waktu.

Dalam penulisan skripsi ini, penulis menyadari banyak pihak yang terlibat baik dalam proses membimbing penulisan dan juga memberikan semangat dan dukungan moril atau materiil. Untuk itu, penulis ingin menyampaikan terima kasih kepada:

1. Prof. Dr. H.M. Zainuddin, MA., selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Prof. Dr. Sri Hariani, M.Si., selaku dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Dr. Fachrul Kurniawan, M.MT., IPM, selaku Ketua Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Dr. Cahyo Crysdiyan, MCS dan Dr. Mokhamad Amin Hariyadi, M.T selaku Dosen Pembimbing I dan Dosen Pembimbing II yang telah memberikan banyak dukungan dan bimbingan dalam penulisan skripsi ini.
5. Dr. Irwan Budi Santoso, M.Kom dan Roro Inda Melani, S.Kom., M.Sc selaku Dosen Penguji I dan Dosen penguji II yang telah memberikan banyak saran untuk menyelesaikan skripsi ini.

6. Segenap Dosen, Admin, Laboran dan jajaran pada Program Studi Teknik Informatika yang telah memberikan bimbingan dan bantuan selama studi.
7. Bapak Iwan Setiawan, Ibu Nanik Rokhaniyah, Kak Wiya, serta adik Ifa dan Yazid yang selalu memberikan dukungan dan rangkaian doa terbaik kepada penulis untuk menyelesaikan skripsi ini dengan lancar dan tepat waktu.
8. Zul Iflah Al Juhaeda selaku teman, sahabat, dan orang istimewa yang selalu menemani, mendukung, serta bangga terhadap pencapaian penulis.
9. Sahabat-sahabatku “vbt family” dan “cari pitis” yaitu Anisa, Zulfa, Zahro, Jihan, Izzul, Viola, Mukhlis, sekumpulan manusia terbaik yang senantiasa membantu penulis melalui segala bentuk opini dan obrolan-obrolan seru.
10. Teman-teman satu bimbingan dengan Dr. Cahyo Crysdiyan, MCS yaitu Zulfa, Rizha, Fahrendra, Yoga, dan Galan yang telah membantu melalui segala bentuk teori dan opini yang didiskusikan bersama.
11. Teman-teman Angkatan 2020 Teknik Informatika “INTEGER” yang telah memberikan banyak bantuan baik material maupun dukungan intelektual, semangat, serta motivasi dalam menyelesaikan penyusunan skripsi ini.

Akhir kata, penulis mengakui bahwa penulisan pada skripsi ini masih banyak kekurangan.

Wassalamualaikum Warahmatullahi Wabarakatuh.

Malang, 24 Juni 2024

Penulis

DAFTAR ISI

HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN.....	iv
PERNYATAAN KEASLIAN TULISAN	v
MOTTO	vi
HALAMAN PERSEMBAHAN	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	x
DAFTAR TABEL	xi
DAFTAR GAMBAR.....	xii
ABSTRAK	xiii
ABSTRACT	xiv
مستخلص البحث.....	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Pernyataan Masalah	6
1.3 Tujuan Penelitian	6
1.4 Batasan Penelitian	6
1.5 Manfaat Penelitian	7
BAB II KAJIAN PUSTAKA	8
2.1 Klasifikasi Kecelakaan Kerja	8
2.2 <i>Multinomial Naïve Bayes</i>	13
BAB III DESAIN DAN IMPLEMENTASI	18
3.1 Pengumpulan Data	18
3.2 Desain Sistem.....	19
3.2.1 <i>Preprocessing</i>	21
3.2.2 <i>Term Frequency-Inverse Document Frequency</i>	27
3.2.3 Data Karakteristik	30
3.2.4 <i>Multinomial Naïve Bayes</i>	32
BAB IV UJI COBA DAN PEMBAHASAN.....	37
4.1 Langkah-Langkah Uji Coba.....	37
4.1.1 Data Pengujian	37
4.1.2 Mengukur Kinerja Sistem	38
4.2 Hasil Uji Coba.....	41
4.3 Pembahasan.....	47
BAB V KESIMPULAN DAN SARAN	66
5.1 Kesimpulan	66
5.2 Saran.....	67
DAFTAR PUSTAKA	
LAMPIRAN	

DAFTAR TABEL

Tabel 4.1 Data Pengujian	37
Tabel 4.2 Rasio Pembagian Data	38
Tabel 4.3 Hasil Uji Coba 70:30	41
Tabel 4.4 Hasil Perhitungan Presisi, <i>Recall</i> , <i>F-1 Score</i> pada rasio 70:30.....	42
Tabel 4.5 Hasil Uji Coba 80:20	43
Tabel 4.6 Hasil Perhitungan Presisi, <i>Recall</i> , <i>F-1 Score</i> pada rasio 80:20.....	44
Tabel 4.7 Hasil Uji Coba 90:10	45
Tabel 4.8 Hasil Perhitungan Presisi, <i>Recall</i> , <i>F-1 Score</i> pada rasio 90:10.....	46
Tabel 4.9 Rata-Rata Akurasi, Presisi, <i>Recall</i> , <i>F-1 Score</i> pada 3 Rasio Berbeda ..	46
Tabel 4.10 Hasil <i>10-Fold Cross Validation</i>	47
Tabel 4.11 Ranking Bobot TF-IDF Kelas Terjatuh	54
Tabel 4.12 Ranking Bobot TF-IDF Kelas Tertimpa Benda-Benda Jatuh.....	55
Tabel 4.13 Ranking Bobot TF-IDF Kelas Menginjak, Menabrak, atau Tertimpa Benda, kecuali Benda yang Jatuh.....	56
Tabel 4.14 Ranking Bobot TF-IDF Kelas Terperangkap di dalam atau diantara Benda-Benda	57
Tabel 4.15 Ranking Bobot TF-IDF Kelas Melakukan Gerakan Berat.....	58
Tabel 4.16 Ranking Bobot TF-IDF Kelas Paparan dengan Suhu Ekstrem.....	60
Tabel 4.17 Ranking Bobot TF-IDF Kelas Kontak dengan Arus Listrik	61
Tabel 4.18 Ranking Bobot TF-IDF Kelas Paparan/Kontak dengan Zat atau Radiasi Berbahaya.....	62

DAFTAR GAMBAR

Gambar 3.1 Blok Diagram Desain Sistem	21
Gambar 3.2 Blok Diagram Tahap <i>Preprocessing</i>	21
Gambar 3.3 <i>Flowchart</i> Tahap <i>Cleaning</i>	22
Gambar 3.4 <i>Flowchart</i> Tahap <i>Case Folding</i>	23
Gambar 3.5 <i>Flowchart</i> Tahap <i>Tokenizing</i>	24
Gambar 3.6 <i>Flowchart</i> Tahap <i>Stopword Removal</i>	25
Gambar 3.7 <i>Flowchart</i> Tahap <i>Stemming</i>	26
Gambar 3.8 <i>Flowchart</i> Pembobotan TF-IDF.....	28
Gambar 3.9 Distribusi Fitur Kata dalam Setiap Dokumen	31
Gambar 3.10 Hasil Uji Normalitas <i>Kolmogorov Smirnov</i> dan <i>Shapiro Wilk</i>	32
Gambar 3. 11 <i>Flowchart</i> Algoritma MNB.....	33
Gambar 4.1 Skenario Pengujian <i>10-Fold Cross Validation</i>	40
Gambar 4.2 Visualisasi Prediksi Data Rasio 70:30	42
Gambar 4.3 Visualisasi Prediksi Data Rasio 80:20	44
Gambar 4.4 Visualisasi Prediksi Data Rasio 90:10	46
Gambar 4.5 <i>Wordcloud</i> Kelas Terjatuh	53
Gambar 4.6 <i>Wordcloud</i> Kelas Tertimpa Benda-Benda Jatuh	55
Gambar 4.7 <i>Wordcloud</i> Kelas Menginjak, Menabrak, atau Tertimpa Benda, kecuali Benda yang Jatuh.....	56
Gambar 4.8 <i>Wordcloud</i> Kelas Terperangkap di dalam atau diantara Benda-Benda	57
Gambar 4.9 <i>Wordcloud</i> Kelas Melakukan Gerakan Berat.....	58
Gambar 4.10 <i>Wordcloud</i> Kelas Paparan/Kontak dengan Suhu Ekstrem	59
Gambar 4.11 <i>Wordcloud</i> Kelas Kontak dengan Arus Listrik	61
Gambar 4.12 <i>Wordcloud</i> Kelas Paparan/Kontak dengan Zat atau Radiasi Berbahaya.....	62

ABSTRAK

Setyaningrum, Qorina. 2024. **Klasifikasi Kecelakaan Kerja pada Berita *Online* Menggunakan *Multinomial Naïve Bayes***. Skripsi. Program Studi Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Dr. Cahyo Crysdiyan, MCS (II) Dr. Mokhamad Amin Hariyadi, MT.

Kata kunci: Kecelakaan Kerja, Klasifikasi Berita, *Multinomial Naïve Bayes*

Portal berita *online* berperan penting dalam menyediakan informasi aktual dan faktual kepada masyarakat mengenai berbagai isu, salah satunya kecelakaan kerja yang merupakan permasalahan serius di lingkup industri karena berdampak pada keselamatan dan produktivitas pekerja. Dengan meningkatnya kasus kecelakaan kerja secara global maupun nasional, diperlukan tindakan pencegahan dengan mengklasifikasikan jenis kecelakaan kerja di berbagai sektor melalui portal berita *online*. Namun, klasifikasi manual saat ini memakan waktu dan rentan kesalahan. Penelitian ini bertujuan untuk mengukur kinerja *Multinomial Naïve Bayes* (MNB) dalam mengklasifikasikan berita kecelakaan kerja dan menganalisis faktor-faktor yang mempengaruhi kinerja tersebut. Dataset diolah melalui tahap *preprocessing*, yaitu *cleaning*, *case folding*, *tokenizing*, *stopwords removal*, dan *stemming*. Selanjutnya, data hasil *preprocessing* dihitung tiap kemunculan kata pada setiap dokumen melalui ekstraksi fitur TF-IDF yang dikonversi menjadi data diskrit dengan dikategorikan dalam 20 rentang. Hasil inilah yang digunakan sebagai input pada metode MNB. Hasil penelitian menunjukkan bahwa nilai akurasi tertinggi sebesar 88% pada rasio 90:10. Pengujian *k-fold cross validation* dengan nilai $k=10$ menunjukkan rata-rata akurasi sebesar 84% dengan akurasi tertinggi pada iterasi ke-9 sebesar 89.58%. Selain itu didapatkan hasil bahwa faktor-faktor yang mempengaruhi kinerja MNB dalam proses klasifikasi yaitu penggunaan rasio data pelatihan dan data pengujian, pemilihan fitur, serta karakteristik dataset.

ABSTRACT

Setyaningrum, Qorina. 2024. **Classification of Workplace Accidents in Online News Using Multinomial Naïve Bayes**. Undergraduate Thesis. Department of Informatics Engineering Faculty of Science and Technology Maulana Malik Ibrahim State Islamic University Malang. Supervisor: (I) Dr. Cahyo Crysdiyan, MCS (II) Dr. Mokhamad Amin Hariyadi, MT.

Online news portals play an important role in providing actual and factual information to the public on various issues, one of which is work accidents which are a serious problem in the industrial sphere because they have an impact on worker safety and productivity. With the increasing cases of work accidents globally and nationally, preventive measures are needed by classifying the types of work accidents in various sectors through online news portals. However, current manual classification is time-consuming and error-prone. This study aims to measure the performance of Multinomial Naïve Bayes (MNB) in classifying work accident news and analyze the factors that affect the performance. The dataset is processed through the preprocessing stage, namely cleaning, case folding, tokenizing, stopwords removal, and stemming. Furthermore, the preprocessing data is calculated for each word occurrence in each document through TF-IDF feature extraction which is converted into discrete data by being categorized in 20 ranges. These results are used as input to the MNB method. The results showed that the highest accuracy value was 88% at a ratio of 90:10. Testing k-fold cross validation with a value of $k = 10$ shows an average accuracy of 84% with the highest accuracy at the 9th iteration of 89.58%. In addition, it was found that the factors that affect the performance of MNB in the classification process are the use of the ratio of training data and test data, feature selection, and dataset characteristics.

Keywords: Multinomial Naïve Bayes, News Classification, Occupational Accidents

مستخلص البحث

سيتيانجروم، قورينا. 2024. تصنيف حوادث العمل في الأخبار على الإنترنت باستخدام الخوارزميات الساذجة متعددة الحدود. الأطروحة. برنامج دراسة هندسة المعلوماتية، كلية العلوم والتكنولوجيا، الجامعة الإسلامية الحكومية، مولانا مالك إبراهيم مالانج. المشرف: (الأول) د. كاهيو كريسديان، MCS (الثاني) د. محمد أمين هريادي، MT.

الكلمات المفتاحية: حوادث العمل، تصنيف الأخبار، تصنيف الأخبار، بايز ساذج متعدد الحدود

تلعب البوابات الإخبارية على الإنترنت دورًا مهمًا في توفير المعلومات الفعلية والواقعية للجمهور حول مختلف القضايا، ومن بينها حوادث العمل التي تعد مشكلة خطيرة في المجال الصناعي لما لها من تأثير على سلامة العمال وإنتاجيتهم. مع تزايد حالات حوادث العمل على الصعيدين العالمي والوطني، هناك حاجة إلى اتخاذ تدابير وقائية من خلال تصنيف أنواع حوادث العمل في مختلف القطاعات من خلال البوابات الإخبارية على الإنترنت. ومع ذلك، فإن التصنيف اليدوي الحالي يستغرق وقتًا طويلاً وعرضة للأخطاء. تهدف هذه الدراسة إلى قياس أداء برنامج "تعدد الحدود الساذجة" (MNB) في تصنيف أخبار حوادث العمل وتحليل العوامل التي تؤثر على الأداء. تتم معالجة مجموعة البيانات من خلال مرحلة المعالجة المسبقة، وهي التنظيف، وطيّ الحالات، والتمييز، وإزالة الكلمات المتوقفة، والوقف. علاوةً على ذلك، يتم حساب بيانات المعالجة المسبقة لكل كلمة واردة في كل مستند من خلال استخراج ميزة TF-IDF التي يتم تحويلها إلى بيانات منفصلة من خلال تصنيفها في 20 نطاقًا. تُستخدم هذه النتائج كمدخلات لطريقة MNB. أظهرت النتائج أن أعلى قيمة دقة كانت 88% بنسبة 90:10. يُظهر اختبار التحقق التبادلي المضاعف k بقيمة k = 10 متوسط دقة 84% مع أعلى دقة عند التكرار التاسع بنسبة 89.58%. بالإضافة إلى ذلك، وُجد أن العوامل التي تؤثر على أداء MNB في عملية التصنيف هي استخدام نسبة بيانات التدريب وبيانات الاختبار، واختيار الميزات، وخصائص مجموعة البيانات.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Semakin berkembangnya zaman, informasi menjadi hal yang krusial di kalangan masyarakat sebagai pemahaman terbaru mengenai peristiwa yang terjadi di seluruh dunia. Sebagian besar sumber informasi didapatkan melalui media *online* berupa portal berita yang mudah diakses kapanpun melalui *smartphone*. Dalam hal ini, informasi berita dapat tersedia di berbagai portal berita *online*, khususnya detik.com sebagai salah satu situs web ternama di Indonesia yang menyediakan banyak informasi berita aktual dan faktual. Portal berita tersebut berisi suatu keadaan, peristiwa, dan pendapat-pendapat berupa tanggapan terkait persoalan aktual di kalangan masyarakat. Hal ini dapat mempengaruhi pola kritik masyarakat terkait berita yang tersebar sebagai perbaikan di masa depan dari permasalahan yang sudah terjadi. Oleh karena itu, portal berita berperan krusial dalam pemahaman masyarakat mengenai berbagai isu, salah satunya kecelakaan kerja.

Kecelakaan kerja merupakan salah satu permasalahan serius dalam lingkup industri yang berdampak pada keselamatan dan produktivitas kinerja pekerja. Berdasarkan data *International Labour Organization*, terjadinya kecelakaan kerja di dunia mencapai rata-rata per tahun sebanyak 430 juta kasus. Sedangkan menurut BPJS Ketenagakerjaan, terjadinya kecelakaan kerja di Indonesia sejak tahun 2019 hingga 2021 meningkat mencapai 234.370 kasus (Kementerian Ketenagakerjaan RI, 2022). Berdasarkan jumlah tersebut, sekitar 70% kecelakaan kerja berakibat fatal menyebabkan cacat seumur hidup bahkan kematian. Sebagian besar

kecelakaan kerja terjadi karena kelalaian pekerja dalam penggunaan perangkat mesin berteknologi dan perilaku penggunaan APD saat bekerja. Teknologi di bidang industri menjadikan adanya peningkatan dalam penggunaan perangkat mesin dan bahan kimia agar menghasilkan produk berkualitas baik. Namun, perkembangan teknologi tersebut dapat menyebabkan peningkatan potensi bahaya bagi keselamatan dan kesehatan pekerja di suatu industri. Maka dari itu, diperlukan adanya pencegahan dengan melakukan tindakan-tindakan perbaikan melalui peristiwa yang sudah terjadi sebelumnya sebagai upaya mengurangi jumlah kecelakaan kerja di Indonesia.

Salah satu pencegahan terjadinya kecelakaan kerja dapat dilakukan dengan mengklasifikasikan jenis kecelakaan kerja di berbagai sektor industri melalui portal berita *online*. Hal tersebut dilakukan sebagai bentuk pencegahan dini keselamatan kerja agar pihak industri berfokus pada aspek tertentu yang mungkin perlu diperbaiki, sehingga jenis kecelakaan kerja yang sama tidak terjadi lagi dikemudian hari. Dalam hal ini, portal berita yang beredar menjadi sumber informasi yang menunjukkan kejadian aktual di lapangan kerja. Namun, klasifikasi berita kecelakaan kerja secara manual rentan terhadap kesalahan. Selain itu, klasifikasi secara manual memerlukan waktu cukup lama dan sukar dalam mengatasi volume berita yang mengalami peningkatan setiap harinya. Perlu adanya metode pengklasifikasian data guna merancang model yang dapat mengenali faktor penyebab kecelakaan kerja.

Dalam Islam, Allah SWT. menunjukkan adanya klasifikasi atau pengelompokan atas penciptaan-Nya pada qur'an surah An-Nahl ayat 13 sebagai berikut.

وَمَا ذَرَأْنَاكُمْ فِي الْأَرْضِ مُخْتَلِفًا أَلْوَنًا، إِنَّ فِي ذَلِكَ لَآيَةً لِّقَوْمٍ يَدَّبُّونَ

“dan Dia (menundukkan pula) apa yang Dia ciptakan untuk kamu di bumi ini dengan berlain-lainan macamnya. Sesungguhnya pada yang demikian itu benar-benar terdapat tanda (kekuasaan Allah) bagi kaum yang mengambil pelajaran. (Q.S An-Nahl:13)

Menurut tafsir Kementrian Agama RI, ayat tersebut menjabarkan mengenai kekuasaan Allah SWT. dalam penciptaan segala sesuatu di bumi, beserta hewan dan tumbuhan yang dikelompokkan beragam jenis, bentuk, serta warna agar manusia dapat memanfaatkan sesuai tujuan penciptaan-Nya. Seperti halnya dalam penelitian tentang klasifikasi kecelakaan kerja pada berita *online* menggunakan metode MNB yang dibuat bertujuan untuk mengelompokkan jenis kecelakaan kerja di berbagai sektor, sehingga dapat dimanfaatkan untuk tujuan yang bermanfaat yaitu mengenali potensi resiko kecelakaan kerja secara dini dengan mengambil tindakan pencegahan yang tepat.

Selain itu, klasifikasi atas penciptaan-Nya juga dinyatakan dalam qur'an surah Al-Hujurat ayat 13 sebagai berikut.

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

“Hai manusia, sesungguhnya Kami menciptakan kamu dari seorang laki-laki dan seorang perempuan dan menjadikan kamu berbangsa-bangsa dan bersuku-suku supaya kamu saling kenal-mengenal. Sesungguhnya orang yang paling mulia diantara kamu disisi Allah ialah orang yang paling takwa diantara kamu.

Sesungguhnya Allah Maha Mengetahui lagi Maha Mengetahui.” (QS. Al-Hujurat: 13)

Berdasarkan tafsir Tahlili dalam Departemen Keagamaan Republik Indonesia (KEMENAG), ayat tersebut menjelaskan mengenai penciptaan manusia yang dikelompokkan berdasarkan suku bangsa, jenis kelamin, dan warna kulit untuk saling mengenal dan berbuat tolong menolong. Allah melarang tindakan manusia yang menunjukkan kesombongannya dengan tingkat pangkat, keturunan, dan kekayaan. Bagi-Nya, umat mulia hanyalah orang-orang yang bertakwa pada sang pencipta alam. Ibnu Hibban dan at-Tirmidhi dari Ibnu Umar meriwayatkan manusia terdiri dari dua lagam yaitu umat yang bertakwa di sisi Allah dengan berbuat kebaikan dan umat yang hina di sisi Allah dengan berbuat celaka dan durhaka. Beliau memohon ampun pada Allah untuk dirinya dan seluruh manusia karena sesungguhnya Allah maha menerima taubat dan maha mengetahui hal-hal yang ada di pikiran serta jiwa manusia.

Sesuai dengan tafsir QS. Al-Hujurat: 13, umat manusia dianjurkan untuk saling mengenal dan tolong menolong dibalik keberagaman suku bangsa, jenis kelamin, dan warna kulit karena kedudukan manusia di hadapan Allah SWT. adalah setara, kecuali bagi orang-orang yang bertakwa di sisi-Nya. Seperti halnya membantu pekerja dalam memberi perlindungan tanpa saling membedakan, salah satunya dengan melakukan pencegahan dini terjadinya kecelakaan kerja melalui klasifikasi berita *online*. Hal tersebut bertujuan agar pihak keselamatan kerja dapat menetapkan resiko dan kebutuhan perlindungan pekerja tanpa adanya diskriminasi atas pangkat atau kekayaan dalam suatu industri. Seluruh pekerja akan memperoleh

perlindungan yang sama guna mewujudkan lingkungan kerja adil dimana hak serta kewajiban dihargai tanpa adanya pengecualian.

Metode pengklasifikasian data dikembangkan dengan perpaduan kecerdasan buatan dan pemrosesan bahasa alami dimana sistem akan melakukan pengklasifikasian berita berdasarkan pola yang diberikan. Dalam membangun algoritma *machine learning*, banyak metode yang diaplikasikan dalam pengklasifikasian teks, salah satunya adalah *Multinomial Naïve Bayes* (MNB) yang menerapkan distribusi multinomial dalam setiap fiturnya. MNB dikembangkan berdasarkan dasar metode probabilistik klasifikasi berita menurut frekuensi kemunculan kata. Adapun alasan mengapa MNB banyak diterapkan pada pengklasifikasian dokumen teks karena kemampuannya yang dapat menyelesaikan perhitungan secara cepat dengan tingkat akurasi cukup tinggi (Hamzah, 2012). Penelitian oleh Rahman *et al.* (2017) tentang pengklasifikasian berita bahasa Indonesia dengan menerapkan metode *Multinomial Naïve Bayes*. Proses seleksi fitur mengimplementasikan TF-IDF guna mengurangi jumlah dimensi data yang menghasilkan akurasi tertinggi sebesar 86,62%. Setianingrum *et al.* (2017) melakukan penelitian mengenai pengklasifikasian surat keluar untuk penentuan penomoran surat secara otomatis dengan mengimplementasikan *Multinomial Naïve Bayes*. Pada proses klasifikasi menggunakan *confix-stripping stemmer* guna mendapatkan kata dasar. Hasil dari penelitian ini memiliki nilai *accuracy* 89,58%, *precision* 79,17%, *recall* 78,72%, dan *f-measure* 77,05% pada uji orisinal.

1.2 Pernyataan Masalah

Berdasarkan latar belakang yang telah dipaparkan, maka pernyataan masalah yang dapat dirumuskan dalam penelitian ini yaitu:

1. Bagaimana kinerja *Multinomial Naïve Bayes* dalam pengklasifikasian kecelakaan kerja berdasarkan berita *online*?
2. Faktor-faktor apa yang dapat mempengaruhi kinerja *Multinomial Naïve Bayes* pada klasifikasi kecelakaan kerja berdasarkan berita *online*?

1.3 Tujuan Penelitian

Adapun tujuan penelitian ini yaitu:

1. Mengukur kinerja algoritma *Multinomial Naïve Bayes* dalam pengklasifikasian kecelakaan kerja berdasarkan berita *online*.
2. Menganalisis faktor-faktor yang dapat mempengaruhi kinerja *Multinomial Naïve Bayes* pada klasifikasi kecelakaan kerja berdasarkan berita *online*.

1.4 Batasan Penelitian

Mengingat luasnya cakupan masalah yang akan diteliti, maka penulis membatasi penelitian ini yaitu:

1. Data penelitian berasal dari portal berita *online* Detik berbahasa Indonesia sebanyak 960 data.
2. Data penelitian yang dikumpulkan hanya terkait berita kecelakaan kerja dengan 8 kategori jenis kecelakaan kerja berdasarkan *International Labour Organization*.

1.5 Manfaat Penelitian

Dengan dilakukannya penelitian ini, diharapkan dapat memberi beberapa manfaat yaitu:

1. Untuk bidang kesehatan dan keselamatan kerja, memudahkan dalam menganalisis jenis kecelakaan kerja di berbagai sektor industri, sehingga dapat mengenali pola kejadian dan potensi resiko kecelakaan kerja secara dini dengan mengambil tindakan pencegahan yang tepat.
2. Untuk pekerja sektor industri, membantu meningkatkan edukasi dan pelatihan dalam hal keselamatan kerja.

BAB II

KAJIAN PUSTAKA

2.1 Klasifikasi Kecelakaan Kerja

Klasifikasi berasal dari bahasa latin yang artinya *classis* atau pengelompokan data ke dalam kelas berbeda menurut atribut tertentu yang ada pada data tersebut guna mengidentifikasi pola data. Klasifikasi merupakan proses pembagian objek dalam kelas, sub kelas, dan sub-sub kelas menurut kesamaan pada objek secara umumnya agar dapat dibedakan secara logika (Habsyi, 2012). Berdasarkan Kamus Besar Bahasa Indonesia (KBBI), klasifikasi merupakan pengelompokan sistem berlandaskan standar yang sudah ditentukan. Tujuan dari adanya pengklasifikasian yaitu memprediksi kelas data dengan mengolah nilai pada data acuan menggunakan algoritma klasifikasi agar menghasilkan data yang akurat.

Kecelakaan kerja merupakan peristiwa tidak direncanakan bahkan dinantikan dan terjadi di lingkup industri yang menyebabkan kerugian materil dan immateril. Berlandaskan Peraturan Menteri Tenaga Kerja No. 3, kecelakaan kerja merupakan peristiwa yang terjadi di tempat kerja pada saat melakukan tugasnya serta dalam perjalanan berangkat maupun pulang yang mengakibatkan korban jiwa, cedera, hingga kematian tergantung pada jenis kecelakaan kerja yang menimpanya. Berdasarkan *International Labour Organization* (ILO), kecelakaan kerja diklasifikasikan menurut penyebab, sifat luka, letak luka, dan jenisnya (ILO, 1996). Klasifikasi menurut penyebabnya yaitu perangkat mesin, peralatan angkut, bahan-bahan kimia, peralatan listrik, dan instalasi pendingin di ruang lingkup kerja. Klasifikasi menurut sifat lukanya yaitu patah tulang, dislokasi, regang otot, memar,

pengaruh bahan kimia, luka bakar, remuk, dan amputasi. Klasifikasi menurut letak lukanya yaitu kepala, leher, badan, anggota atas, dan bawah. Sedangkan klasifikasi menurut jenis kecelakaan kerja sesuai ketentuan ILO dalam (ILO, 1996), diantaranya:

1. Terjatuh
 - a) Jatuhnya orang dari ketinggian (pohon, bangunan, perancah, tangga, mesin, kendaraan) pada kedalaman (sumur, parit, galian, lubang di tanah)
 - b) Terjatuh di permukaan datar
2. Tertimpa benda-benda yang jatuh
 - a) Longsor dan masuk ke dalam gua (batu, tanah, salju)
 - b) Runtuh (bangunan, perancah, tumpukan barang)
 - c) Tertimpa benda jatuh saat memindahkan atau mengangkat barang
3. Menginjak, menabrak, atau tertimpa benda, kecuali benda yang jatuh
 - a) Menginjak benda
 - b) Membentur benda yang tidak bergerak, kecuali apabila benturan tersebut terjadi akibat jatuh sebelumnya
 - c) Membentur benda bergerak
 - d) Terkena benda bergerak (termasuk serpihan dan partikel yang beterbangan), namun tidak termasuk benda yang jatuh
4. Terperangkap di dalam atau diantara benda-benda
 - a) Terperangkap di dalam sebuah benda
 - b) Terjepit di antara benda yang tidak bergerak dan benda yang bergerak

- c) Terperangkap diantara benda-benda yang bergerak, kecuali benda yang terbang atau jatuh
5. Terlalu banyak beraktivitas atau melakukan gerakan berat
 - a) Terlalu banyak tenaga dalam mengangkat benda
 - b) Terlalu banyak tenaga dalam mendorong atau menarik benda
 - c) Terlalu banyak bergerak saat memegang atau melempar benda
 - d) Gerakan Berat
 6. Paparan atau kontak dengan suhu ekstrem
 - a) Paparan panas (atmosfer atau lingkungan)
 - b) Paparan terhadap suhu dingin (atmosfer atau lingkungan)
 - c) Kontak dengan zat atau benda panas
 - d) Kontak dengan zat atau benda yang sangat dingin
 7. Terpapar atau kontak dengan arus listrik
 8. Paparan atau kontak dengan zat atau radiasi berbahaya
 - a) Kontak dengan penghirupan, konsumsi, atau penyerapan zat berbahaya
 - b) Paparan radiasi pengion
 - c) Paparan terhadap radiasi, selain radiasi pengion

Klasifikasi jenis kecelakaan kerja pada berbagai sektor dilakukan untuk mengidentifikasi pola jenis kecelakaan kerja di berbagai sektor industri sebagai bentuk pencegahan dini untuk perbaikan aspek-aspek tertentu yang diperlukan, sehingga peristiwa yang sama tidak terjadi kedua kalinya. Tujuan dari pengklasifikasian jenis kecelakaan kerja yaitu memberikan label pada sebuah berita dengan jenis kecelakaan kerja yang sesuai. Nisak & Nugraha (2018) melakukan

penelitian mengenai pengklasifikasian penanganan kecelakaan kerja menggunakan algoritma *Decision Tree*. Peneliti menggunakan model log linier untuk memahami pola hubungan antar variabel menurut faktor status, klasifikasi, dan tempat terjadinya kecelakaan kerja di PT. Pertamina Cepu. Pada penelitian tersebut, peneliti mendapat nilai akurasi sebesar 68.9% menggunakan algoritma C4.5 dan nilai ICI sebesar 31% dimana terdapat 3 faktor yang mempengaruhi pola penanganan kecelakaan kerja yaitu pengamatan lapangan, tingkat resiko, dan lama penanganan.

Abdullah *et al.* (2023) mengerjakan penelitian mengenai pengklasifikasian tingkat keparahan korban kecelakaan lalu lintas dengan mengimplementasikan metode *Naïve Bayes*. Peneliti menggunakan *data mining* untuk pengolahan data kecelakaan di kota Pekanbaru dari tahun 2015 hingga 2021. Penelitian ini menggunakan 5 kategori tingkat keparahan korban kecelakaan yaitu luka ringan, luka berat, dan meninggal dunia. Penelitian dilakukan dengan proses *cleaning*, transformasi, dan *feature selection* melalui variabel yang dominan dalam penentuan tingkat keparahan kecelakaan yaitu waktu, kendaraan, usia, dan tingkat parahnya kecelakaan tersebut. Pengujian dilakukan dengan menerapkan *10-fold cross validation* guna mendapatkan nilai akurasi tertinggi saat data terbagi menjadi data *testing* dan data *training*. Peneliti mendapat nilai akurasi sebesar 57%.

Utami & Irsyadi (2018) melakukan penelitian mengenai pengklasifikasian kecelakaan lalu lintas di kota Boyolali dengan menerapkan metode *Naïve Bayes*. Penelitian ini dilakukan untuk menganalisa data kecelakaan yang didapat dari Satlantas Polres Boyolali tahun 2016 sebanyak 742 kasus. Klasifikasi kecelakaan

ini terbagi menjadi 2 kategori yaitu sering dan tidak sering terjadinya kecelakaan lalu lintas tersebut. Peneliti melakukan observasi dan wawancara terkait penyebab umum terjadinya kecelakaan di wilayah tersebut, kemudian data diolah melalui tahap *cleaning* dengan data *training* sebanyak 742 dan data *testing* sebanyak 25. Peneliti melaksanakan uji akurasi menggunakan *black box* sebanyak 5 kali dengan pengambilan data *test* secara acak dan mendapatkan hasil akurasi tertinggi 68,5% dengan rata-rata akurasi sebesar 41,67%.

Wisdayani *et al.* (2019) melakukan penelitian mengenai pengimplementasian *K-Nearest Neighbor* dalam mengklasifikasikan tingkat keparahan korban kecelakaan lalu lintas di Kabupaten Jawa Tengah dengan menerapkan teknik *data mining*. Dataset pada penelitian ini berasal dari Satlantas Kabupaten Pati tahun 2017 mengenai jumlah kecelakaan lalu lintas di daerah tersebut yang diolah dengan variabel label dan atribut. Penelitian ini menggunakan variabel label berupa tingkat keparahan korban, kemudian untuk variabel atributnya berupa jenis kelamin, tanggal peristiwa, jenis pekerjaan, jenis kendaraan, faktor dan peran korban, alat keselamatan, dan jenis kecelakaan. Nilai Parameter K yang diimplementasikan pada penelitian ini yaitu 3 dengan jarak *Euclidean* 1,0,2,3,4,0,1,0,2,1,3 dimana dilakukan pengujian tingkat akurasi dari 241 data latih. Penelitian ini menghasilkan nilai prediksi yang tinggi dibandingkan nilai yang tidak sinkron dengan tingkat akurasi 88%, presisi 64%, *recall* 60%, dan *f-measure* berturut-turut 62%.

Penelitian yang dilakukan oleh Franseda *et al.* (2020) mengenai pengklasifikasian kecelakaan lalu lintas menggunakan *Decision Tree* dan *Synthetic*

Minority Over-sampling Technique (SMOTE) sebagai solusi dalam mengatasi ketidakseimbangan kelas dengan tingkat akurasi tinggi. Pada penelitian ini menggunakan data sekunder pada situs web catatan keselamatan pemerintah Australia Selatan tahun 2018 dengan total 13.599 data. Kategori pada penelitian ini terdiri dari kerusakan barang berharga (*property damage only*) dan korban jiwa (*casualties*). Pada dataset tersebut dilakukan *preprocessing* guna memperlancar keberhasilan proses *data mining* melalui transformasi data. Selanjutnya data diuji menurut jumlah keseluruhan data pada setiap rasio dengan membandingkan antara *Decision Tree* dan SMOTE + *Decision Tree* dimana ketika diuji dengan SMOTE hasil akurasi meningkat yaitu 71,12% pada rasio 0.1 yang sebelumnya tingkat akurasinya sebesar 65,97%. Penelitian ini juga menghasilkan *precision* sebesar 89% pada rasio 3:7 beserta *Area Under Curve* (AUC) 0,77% pada rasio 1:9 yang dinilai sebagai kategori klasifikasi cukup berhasil.

2.2 *Multinomial Naïve Bayes*

Multinomial Naïve Bayes (MNB) adalah algoritma klasifikasi turunan dari *Naïve Bayes* yang dapat diterapkan dalam mengklasifikasikan teks karena efisien dalam dataset bervolume tinggi dengan menyertakan berbagai fitur. MNB memproses pengklasifikasian data yang berfungsi untuk menentukan kelas dengan menghitung frekuensi setiap kata beserta jumlahnya pada data tersebut. MNB memperkirakan bahwa seluruh variabel saling berkaitan satu sama lain berdasarkan kelasnya dengan melupakan seluruh dependensi antar variabel (Bunga *et al.*, 2018). Dalam algoritma MNB, data teks dimodelkan sebagai distribusi multinomial dimana setiap data direpresentasikan dalam bentuk vektor yang didalamnya berisi

frekuensi kemunculan kata dalam data tersebut. MNB akan mengklasifikasikan data ke dalam kelas dengan probabilitas posterior paling tinggi dan nantinya digunakan sebagai prediksi klasifikasi.

Dewi (2021) melakukan penelitian mengenai pengklasifikasian berita *online* Kompas dan Tempo dengan menerapkan metode *multinomial naïve bayes* dan *text mining* guna membentuk data dan pemodelan klasifikasi. Dataset pada penelitian ini akan dikategorikan menjadi 7 jenis, yaitu kategori bisnis, polhukam, hiburan, olahraga, otomotif, kesehatan, dan teknologi dengan jumlah data 10.500 dan masing-masing 1.500 data. Pemrosesan data menerapkan *scraping* secara otomatis menggunakan *web scrape* dengan mengambil informasi berita *online* yang dilabeli kategori secara manual. Tahap *preprocessing* melalui *tokenization*, *stopword removal*, *case folding*, *stemming* kemudian dilakukan pembobotan TF-IDF. Pengujian pada penelitian ini menerapkan model *pipeline* kemudian membagi dataset menjadi rasio 90:10, didapat hasil nilai yang sama yaitu *akurasi*, *presisi*, *recall*, dan *f1-score* berturut-turut sebesar 96%.

Penelitian oleh Zhafira *et al.* (2021) mengenai analisis sentimen kebijakan kampus merdeka menurut komentar *youtube* dengan menerapkan metode *Naïve Bayes Classifier* (NBC). Penelitian ini melalui beberapa tahap, diantaranya tahap pemberian label secara manual, *preprocessing*, ekstraksi fitur TF-IDF, serta pengujian *k-fold cross validation*, kemudian tahap klasifikasi NBC. Pemrosesan data dilakukan menggunakan *scraping* dengan mengambil data API *youtube* sebanyak 1000 komentar dengan data *balance* antara sentimen negatif dan positif. Pelabelan data dilakukan secara manual oleh bidang ilmu psikologi, ilmu komputer,

dan ilmu bahasa. Hasil pengujian *10-fold cross validation* dengan nilai presisi 90%, *recall* 93%, dan *f-1 score* 91% dengan nilai setiap iterasi sebesar 91% dan nilai akurasi tertinggi pada iterasi urutan ke-4 sebesar 97%.

Hamzah (2012) melakukan penelitian mengenai pengklasifikasian teks berita dan abstrak akademis menggunakan algoritma *Naïve Bayes Classifier* (NBC) dengan dataset berjumlah 1000 artikel berita dan 450 abstrak akademis. Penelitian ini ditujukan untuk mengetahui seberapa tinggi kinerja algoritma NBC dalam pengklasifikasian teks berita dan akademis melalui abstrak artikel dari berbagai sumber. Pada penelitian ini menerapkan filter frekuensi teks dalam kamus guna membedakan kategori dokumen satu dengan yang lain dengan rasio pembagian 90:10. Penelitian ini berfokus pada pemilihan *feature* kata untuk merepresentasikan teks dengan meneliti berapa banyak kata dalam kamus yang berpengaruh terhadap kinerja algoritma NBC. Hasil penelitian ini mencapai akurasi 91% pada dokumen berita dan 82% pada dokumen akademik dengan uji seleksi kata minimal 4 hingga 5 dokumen dimana terjadi penurunan tingkat akurasi dikarenakan fitur kata terpilih pada dokumen akademik bersifat ambigu.

Penelitian oleh Rahman *et al.* (2017) tentang pengklasifikasian berita *online* menggunakan algoritma *Multinomial Naïve Bayes* (MNB) dengan melakukan perhitungan frekuensi tiap kata yang muncul pada data. Dataset pada penelitian ini berasal dari data berita lembaga pengolahan dan penyedia informasi, kementerian komunikasi dan informatika, serta dirjen informasi publik berjumlah 1011 yang dikategorikan pada 15 kelas. Tahapan pada penelitian ini diantaranya *preprocessing*, proses seleksi fitur atau TF-IDF guna menghitung dokumen dengan

kata tertentu, kemudian menentukan *threshold* data untuk pengurangan dimensi data yang tinggi, lalu dilakukan klasifikasi MNB. Pada proses pengujian, nilai akurasi dinilai rendah dikarenakan fitur kata tidak muncul, sehingga dilakukan perbandingan tingkat akurasi antara MNB, TFIDF-MNB, dan *DF-Threshold* dengan nilai akurasi masing-masing sebesar 93%, 94%, dan 92%.

Randhika *et al.* (2021) melakukan penelitian mengenai pengklasifikasian kategori berita online dengan menggabungkan algoritma *Multinomial dan Complement Naïve Bayes* dari subset data berita pada PT. Merah Putih Media yang berjumlah 42.155 artikel. Ada 3 kategori dalam penelitian ini diantaranya hiburan dan gaya hidup, olahraga, dan Indonesiaku yang diberi label secara manual oleh redaktur berita. Tahapan pada penelitian ini dimulai dari tahap *preprocessing*, pelatihan dan evaluasi data dengan menginisialisasikan model klasifikasi (MNB, CNB, dan kombinasi kedua algoritma) serta pembobotan TF-IDF. Pada pengujian setiap model dilakukan evaluasi menggunakan *f1-score* dengan tingkat akurasi tinggi sebesar 90% dengan kesalahan prediksi sebesar 5% dari kategori olahraga dan kesalahan prediksi benar sebesar 20% pada kategori Indonesiaku.

Penelitian yang dilakukan oleh Singh *et al.* (2019) mengenai pengklasifikasi teks dengan membandingkan algoritma *Multinomial dan Bernoulli Naïve Bayes*. Penelitian ini berfokus pada analisis sentimen artikel berita dengan kelas positif serta negatif menggunakan dua pendekatan *Multivariate Bernoulli Naïve Bayes Classification* dan *Multinomial Naïve Bayes Classification*. Dataset dalam penelitian ini berupa 312 record berita tahun 2018 di India terdiri dari kolom 'news' (artikel berita) dan 'polarity' (kecenderungan artikel berita

dalam hal positif dan negatif). Tahapan pada penelitian ini diantaranya *preprocessing* (*lowercasing, tokenization, punctuation removal, dan stopwords removal*), lalu dilakukan klasifikasi dengan membandingkan kedua metode yaitu *Multinomial* dan *Multivariate Bernoulli Naïve Bayes*. Pada proses pengujian, nilai akurasi menggunakan *Multinomial* memiliki performa lebih baik daripada *Multivariate Bernoulli* dengan hasil akurasi berturut-turut adalah 73.4% dan 69.15%. Kemudian dibuktikan dengan grafik batang akurasi dan kurva ROC yang menyatakan bahwa kinerja *Multinomial* lebih baik dari kinerja *Bernoulli* dengan tingkat positif sebenarnya lebih besar.

BAB III

DESAIN DAN IMPLEMENTASI

3.1 Pengumpulan Data

Penelitian ini menggunakan data berita kecelakaan kerja yang diambil dari portal detik.com dengan teknik *web scrape*. Data tersebut diberi label secara manual yang diambil berdasarkan khalayak ramai melalui kuesioner. Data dilabeli sesuai ketentuan *International Labour Organization* (ILO) dengan label terjatuh, tertimpa benda-benda jatuh, menginjak, menabrak, atau tertimpa benda, kecuali benda yang jatuh, terperangkap di dalam atau diantara benda-benda, paparan/kontak dengan suhu ekstrem, kontak dengan arus listrik, paparan/kontak dengan zat berbahaya atau radiasi, dan melakukan gerakan berat (ILO, 1996). Data disimpan dalam file berekstensi *.csv* yang berisi kolom teks berita dan kelas jenis kecelakaan kerja dengan pengolahan data menggunakan bahasa pemrograman python. Contoh dataset dalam penelitian ini ditunjukkan pada Tabel 3.1.

Tabel 3.1 Contoh Data Penelitian

No	Tanggal	Link	Berita	Label
1.	6/6/2023 22:39	https://news.detik.com/berita/d-6758894/pekerja-tewas-terjatuh-dari-lantai-7-di-jakpus-2-orang-lain-luka	Pekerja Tewas Terjatuh dari Lantai 7 di Jakpus, 2 Orang Lain Luka Seorang pekerja bangunan berinisial HA (30) tewas setelah terjatuh dari lantai 7 sebuah bangunan di Gondangdia, Jakarta Pusat (Jakpus). Korban terjatuh diduga karena tali gondola putus.	Terjatuh
2.	9/14/2023 16:48	https://www.detik.com/jabar/berita/d-6931306/nasib-tragis-wawan-di-pabrik-penggilingan-gabah-tasikmalaya	Nasib Tragis Wawan di Pabrik Penggilingan Gabah Kecelakaan kerja yang berakibat fatal terjadi di pabrik penggilingan gabah di Kampung Perintis, Kelurahan Sukamanah, Kecamatan Cipedes, Kota	Terperangkap di dalam atau diantara benda-benda

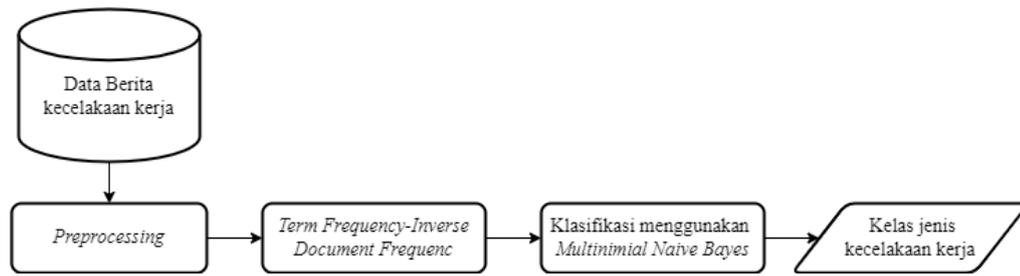
Tabel 3.1 Contoh Data Penelitian

No	Tanggal	Link	Berita	Label
			Tasikmalaya, Kamis (14/9/2023) siang. Seorang pekerja tewas terlilit putaran mesin saat sedang mengoperasikan mesin diesel berkapasitas besar itu.	
3.	1/10/2023 7:45	https://www.detik.com/sulsel/berita/d-6507100/nahas-2-pekerja-telkom-di-konawe-sultra-tewas-tersengat-listrik	Nahas 2 Pekerja Telkom di Konawe Sultra Tewas Tersengat Listrik Dua pekerja Telkom di Kabupaten Konawe, Sulawesi Tenggara (Sultra) tewas tersengat listrik. Kejadian nahas itu juga menyebabkan 3 pekerja lainnya turut menjadi korban.	Kontak dengan arus listrik
4.	7/6/2020 15:34	https://news.detik.com/berita-jawa-barat/d-5082059/diduga-gas-bocor-rumah-produksi-krupuk-di-cirebon-ludes-terbakar	Diduga Gas Bocor, Rumah Produksi Krupuk di Cirebon Ludes Terbakar Home industri atau rumah usaha produksi krupuk di Desa Karang Asem, Kecamatan Plumbon, Kabupaten Cirebon, ludes terbakar. Kebakaran diduga akibat kebocoran gas. Kasie Tanggap Darurat Dinas Pemadam Kebakaran (Damkar) Kabupaten Cirebon Eno Sujana mengatakan kebakaran di rumah usaha produksi krupuk tersebut terjadi sekitar pukul 12.00.	Paparan/Kontak dengan Zat atau Radiasi Berbahaya
5.	2/16/2024 19:46	https://www.detik.com/jabar/berita/d-7197093/kelelahan-seorang-pengawas-pemilu-desa-di-cianjur-meninggal-dunia	Kelelahan, Seorang Pengawas Pemilu Desa di Cianjur Meninggal Dunia Beny Karyanto (55), anggota Panwas Kelurahan Desa (PKD) Desa Cibadak Kecamatan Cibeder, Kabupaten Cianjur meninggal dunia usai menjalani perawatan intensif di rumah sakit. Diduga penyelenggara pemilu di tingkat desa itu kelelahan saat bekerja menyiapkan pelaksanaan pemungutan suara.	Melakukan gerakan berat

3.2 Desain Sistem

Alur desain sistem pada penelitian ini ditunjukkan pada Gambar 3.1 melalui tahapan-tahapan yaitu input data berupa berita kecelakaan kerja, *preprocessing*,

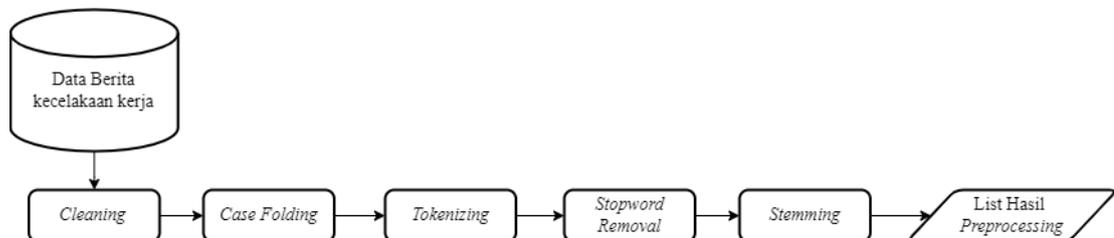
term frequency, dan klasifikasi menggunakan *Multinomial Naïve Bayes*. Tahap tersebut menghasilkan *output* berupa kelas jenis kecelakaan kerja. Tahap pertama dimulai dengan pengambilan dataset menggunakan teknik *web scrape* pada portal berita detik.com mengenai artikel berita kecelakaan kerja. Kemudian dilakukan pelabelan secara manual sesuai ketentuan dari ILO yang diambil dari khalayak ramai. Tahap kedua dilakukan persiapan data melalui *preprocessing* teks, yaitu *cleaning, case folding, tokenizing, stopword removal, dan stemming*. Tahap ketiga, data diperhitungkan dengan pembobotan TF-IDF guna melatih model yang diimplementasikan dalam data uji. Pembobotan TF-IDF digunakan untuk menilai pentingnya suatu kata dalam dokumen. Kata-kata yang sering ditemukan dalam teks, namun jarang muncul pada teks lain bernilai bobot lebih tinggi. Pada penelitian ini, hasil TF-IDF awalnya bersifat kontinu dikonversi menjadi data diskrit yang dikategorikan ke dalam 20 rentang atau bin. Rentang ini dimulai dari 0 hingga 1 dengan interval 0.05 untuk setiap bin. Bin 0 mencakup nilai dari 0 hingga kurang dari 0.05, bin 1 mencakup nilai dari 0.05 hingga kurang dari 0.1, dan seterusnya. Proses ini berlanjut dengan interval yang sama hingga bin 19 yang mencakup nilai dari 0.95 hingga 1. Konversi ini dilakukan untuk memastikan bahwa hasil TF-IDF sesuai dengan persyaratan model klasifikasi MNB yang membutuhkan data dalam bentuk diskrit. Selanjutnya pada tahap terakhir, hasil dari perhitungan bobot TF-IDF digunakan sebagai input dalam model klasifikasi MNB. Model MNB dilatih menggunakan data yang telah diproses dan dibobotkan untuk mengklasifikasikan jenis kecelakaan kerja berdasarkan teks berita.



Gambar 3.1 Blok Diagram Desain Sistem

3.2.1 Preprocessing

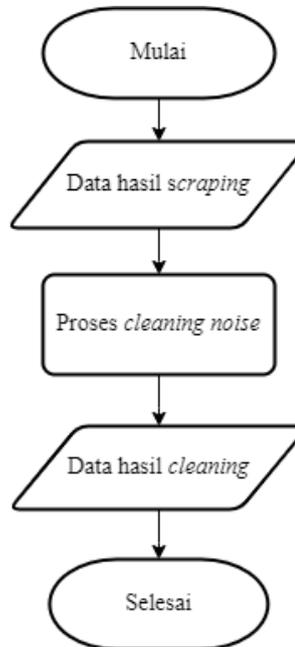
Pada tahap ini, data berita yang didapat dari *web scrape* diproses dalam bentuk file excel. Selanjutnya, dilakukan pengolahan melalui beberapa tahapan agar menghasilkan data yang terstruktur untuk pemodelan klasifikasi. Tahap *preprocessing* melalui *cleaning*, *case folding*, *tokenizing*, *stopword removal*, dan *stemming*. Hasil dari tahap ini di-list dalam bentuk data untuk diproses pada tahap pembobotan TF-IDF. Alur tahap *preprocessing* ditunjukkan pada Gambar 3.2.

Gambar 3.2 Blok Diagram Tahap *Preprocessing*

3.2.2.1 Cleaning

Tahap ini dilakukan dengan membersihkan *noise* yang tidak dibutuhkan guna memudahkan model memproses tahap klasifikasi. *Noise* yang dimaksud adalah angka, simbol, emotikon, tagar, tanda baca, tautan, dan karakter khusus lainnya. Tahap ini bertujuan untuk menghasilkan teks berita yang terstruktur, sehingga dapat meningkatkan akurasi model klasifikasi. Pada penelitian ini

menggunakan library *nltk* (*Natural Language Toolkit*) dan *sastrawi* untuk tahap *cleaning*. Alur tahap *cleaning* data ditunjukkan pada Gambar 3.3.



Gambar 3.3 *Flowchart* Tahap *Cleaning*

Pada tahap ini, input data yang digunakan adalah berita kecelakaan kerja. Apabila dalam kalimat tersebut terdapat *noise*, maka akan dihapuskan. Dapat dilihat pada Tabel 3.2 mengenai perbedaan hasil dataset sebelum dan sesudah tahap *cleaning*.

Tabel 3.2 Dataset sebelum dan sesudah tahap *Cleaning*

Sebelum <i>Cleaning</i>	Sesudah <i>Cleaning</i>
Pekerja Hotel di Medan Terjatuh dari Kanopi Setinggi 10 Meter	Pekerja Hotel di Medan Terjatuh dari Kanopi Setinggi Meter
Detik-detik Crane di Kemayoran Roboh Timpa Rumah dan Lukai 3 Orang	Detik detik Crane di Kemayoran Roboh Timpa Rumah dan Lukai Orang
Kecelakaan Kerja, 3 Jam Heri Terjepit Alat Berat Penyusun Kontainer di JICT	Kecelakaan Kerja Jam Heri Terjepit Alat Berat Penyusun Kontainer di JICT
6 Orang Pekerja Tersengat Listrik di Cikarang Utara, 2 Orang Tewas	Orang Pekerja Tersengat Listrik di Cikarang Utara Orang Tewas

3.2.2.2 Case Folding

Tahap *case folding* dilakukan dengan mengubah seluruh teks dengan huruf besar menjadi bentuk *lowercase* atau huruf kecil guna menyelaraskan seluruh huruf pada kalimat menggunakan library *nlk* dan *sastrawi*. Hal ini bertujuan agar model dapat mengidentifikasi kata setiap kalimat tanpa adanya ambiguitas teks. Alur tahap *case folding* ditunjukkan pada Gambar 3.4.



Gambar 3.4 Flowchart Tahap Case Folding

Pada tahap ini, input data adalah hasil dari tahap *cleaning*. Kalimat tersebut diubah menjadi bentuk *lowercase* atau huruf kecil. Ditunjukkan pada Tabel 3.3 mengenai perbedaan hasil dataset sebelum dan sesudah tahap *case folding*.

Tabel 3.3 Dataset sebelum dan sesudah tahap *Case Folding*

Sebelum Case Folding	Sesudah Case Folding
Pekerja Hotel di Medan Terjatuh dari Kanopi Setinggi Meter	pekerja hotel di medan terjatuh dari kanopi setinggi meter
Detik detik Crane di Kemayoran Roboh Timpa Rumah dan Lukai Orang	detik detik crane di kemayoran roboh timpa rumah dan lukai orang
Kecelakaan Kerja Jam Heri Terjepit Alat Berat Penyusun Kontainer di JICT	kecelakaan kerja jam heri terjepit alat berat penyusun kontainer di jict
Orang Pekerja Tersengat Listrik di Cikarang Utara Orang Tewas	orang pekerja tersengat listrik di cikarang utara orang tewas

3.2.2.3 Tokenizing

Tahap *tokenizing* dilakukan dengan memecah teks dalam suatu kalimat menurut spasi atau tanda baca menjadi *token* menggunakan library *nlk* dan *sastrawi*. Hal ini bertujuan agar model dapat memahami struktur teks dalam entitas terpisah. Alur tahap *tokenizing* ditunjukkan pada Gambar 3.5.



Gambar 3.5 Flowchart Tahap Tokenizing

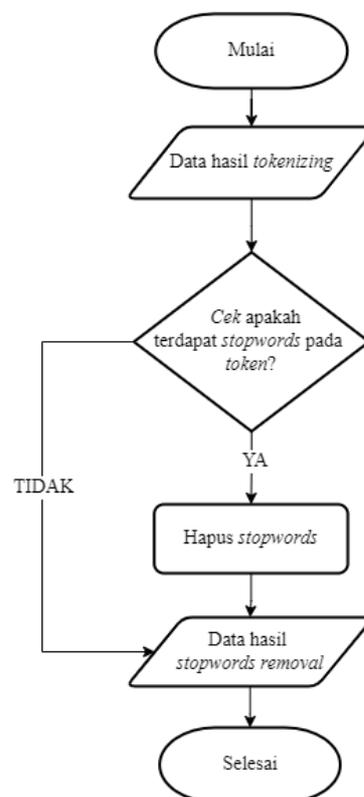
Input data yang digunakan adalah hasil tahap *case folding*. Kalimat tersebut dipecah menjadi *token* kata. Dapat dilihat pada Tabel 3.4 mengenai perbedaan hasil dataset sebelum dan sesudah tahap *Tokenizing*.

Tabel 3.4 Dataset sebelum dan sesudah tahap *Tokenizing*

Sebelum Tokenizing	Sesudah Tokenizing
pekerja hotel di medan terjatuh dari kanopi setinggi meter	[pekerja, hotel, di, medan, terjatuh, dari, kanopi, setinggi, meter]
detik detik crane di kemayoran roboh timpa rumah dan lukai orang	[detik, detik, crane, di, kemayoran, roboh, timpa, rumah, dan, lukai, orang]
kecelakaan kerja jam heri terjepit alat berat penyusun kontainer di jict	[kecelakaan, kerja, jam, heri, terjepit, alat, berat, penyusun, kontainer, di, jict]
orang pekerja tersengat listrik di cikarang utara orang tewas	[orang, pekerja, tersengat, listrik, di, cikarang, utara, orang, tewas]

3.2.2.4 Stopword Removal

Tahap ini dilakukan dengan menghapus *stopwords* atau kata-kata umum yang tidak bermakna signifikan dalam suatu kalimat, misalnya “atau”, “dan”, “dari”, “yang” menggunakan jenis *stopwords* bahasa Indonesia dari *library* nltk. Hal tersebut bertujuan agar sejumlah kata yang diproses oleh model lebih instruktif, sehingga dapat mengurangi adanya *noise*. Alur tahap *stopword removal* ditunjukkan pada Gambar 3.6.



Gambar 3.6 Flowchart Tahap Stopword Removal

Input data yang digunakan adalah hasil tahap *tokenizing*. Kalimat tersebut dipecah menjadi *token* kata dengan menghilangkan kata-kata umum yang tidak berpengaruh. Perbedaan hasil dataset sebelum dan sesudah tahap *stopword removal* ditunjukkan pada Tabel 3.5.

Tabel 3.5 Dataset sebelum dan sesudah tahap *Stopword Removal*

Sebelum <i>stopword removal</i>	Sesudah <i>stopword removal</i>
[pekerja, hotel, di, medan, terjatuh, dari, kanopi, setinggi, meter]	[pekerja, hotel, medan, terjatuh, kanopi, meter]
[detik, detik, crane, di, kemayoran, roboh, timpa, rumah, dan, lukai, orang]	[detik, detik, crane, kemayoran, roboh, timpa, rumah, lukai, orang]
[kecelakaan, kerja, jam, heri, terjepit, alat, berat, penyusun, kontainer, di, jict]	[kecelakaan, kerja, jam, heri, terjepit, alat, berat, penyusun, kontainer, jict]
[orang, pekerja, tersengat, listrik, di, cikarang, utara, orang, tewas]	[orang, pekerja, tersengat, listrik, cikarang, utara, orang, tewas]

3.2.2.5 *Stemming*

Tahap ini dilakukan dengan mengubah kata dalam suatu kalimat menjadi bentuk kata dasar melalui penghilangan imbuhan menggunakan *library* sastrawi. Hal ini bertujuan untuk mengidentifikasi kata setiap kalimat tanpa adanya ambiguitas teks dan mengurangi adanya kesalahan dalam pembobotan TF-IDF. Alur tahap *stemming* ditunjukkan pada Gambar 3.7.

Gambar 3.7 *Flowchart Tahap Stemming*

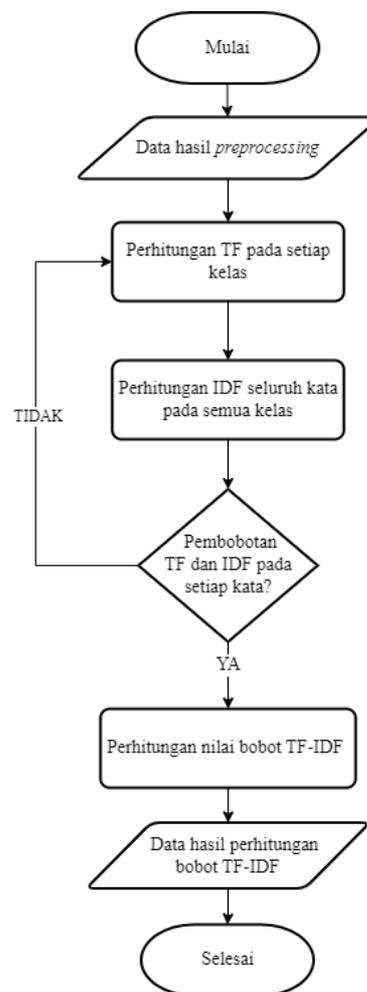
Input data yang digunakan adalah hasil tahap *stopword removal*, kemudian kalimat tersebut dipecah menjadi *token* kata dengan mengubah setiap kata menjadi bentuk kata dasarnya. Dapat dilihat pada Tabel 3.6 mengenai perbedaan hasil dataset sebelum dan sesudah tahap *stemming*.

Tabel 3.6 Dataset sebelum dan sesudah tahap *Stemming*

Sebelum <i>stemming</i>	Sesudah <i>stemming</i>
[pekerja, hotel, medan, terjatuh, kanopi, meter]	[kerja, hotel, medan, jatuh, kanopi, meter]
[detik, detik, crane, kemayoran, roboh, timpa, rumah, lukai, orang]	[detik, detik, crane, mayor, roboh, timpa, rumah, luka, orang]
[kecelakaan, kerja, jam, heri, terjepit, alat, berat, penyusun, kontainer, jict]	[celaka, kerja, jam, heri, jepit, alat, berat, susun, kontainer, jict]
[orang, pekerja, tersengat, listrik, cikarang, utara, orang, tewas]	[orang, kerja, sengat, listrik, cikarang, utara, orang, tewas]

3.2.2 *Term Frequency-Inverse Document Frequency*

Setelah tahap *preprocessing* dilakukan, maka perlu ditentukan pembobotan tiap kata dengan menghitung TF-IDF (*Term Frequency-Inverse Document Frequency*). Dalam melakukan perhitungan frekuensi kata, perlu dihitung setiap kata yang mencakup dua langkah, diantaranya TF (pengukuran kemunculan jumlah kata yang ada pada data) dan IDF (pengukuran sejauh mana kata itu muncul di semua data). Pembobotan TF-IDF bertujuan untuk mengevaluasi kepentingan kata dalam data dengan pertimbangan kata yang sering muncul serta keunikan kata tersebut. Kata-kata yang sering ditemukan dalam data, namun jarang muncul pada data lain bernilai bobot lebih tinggi. Sebaliknya, kata-kata yang muncul secara umum di banyak dokumen akan memiliki bobot yang lebih rendah karena dianggap kurang informatif. Dengan menggunakan TF-IDF, kita dapat menyaring kata-kata yang kurang relevan dan lebih fokus pada kata-kata yang signifikan dalam analisis teks. Alur pembobotan TF-IDF ditunjukkan pada Gambar 3.8.



Gambar 3.8 Flowchart Pembobotan TF-IDF

Pembobotan TF-IDF pada suatu data dilakukan setelah tahap *preprocessing* melalui beberapa tahap. Tahap pertama yaitu menghitung frekuensi kata atau TF di dalam suatu data dengan perhitungan rumus pada Persamaan 3.1 yang merupakan rasio jumlah kemunculan kata terhadap total jumlah kata dalam dokumen tersebut. Selanjutnya tahap kedua yaitu melakukan perhitungan IDF dengan rumus Persamaan 3.2 dengan membagi jumlah keseluruhan dokumen dengan jumlah dokumen berisi kata tersebut, kemudian mengambil logaritma dari hasil pembagian tersebut, maka akan didapatkan nilai TF dan IDF. Kemudian langkah ketiga yaitu

melakukan pembobotan TF-IDF dengan rumus Persamaan 3.3. Berikut rumus persamaan dalam perhitungan nilai bobot TF-IDF.

$$tf_{t,d} = \frac{n_{i,j}}{\sum k n_{i,j}} \quad (3.1)$$

$$Idf_t = \log\left(\frac{N}{df_t}\right) \quad (3.2)$$

$$W_{t,d} = tf_{t,d} \times Idf_t \quad (3.3)$$

Keterangan:

$tf_{t,d}$ = Frekuensi Kata

$n_{i,j}$ = Jumlah suatu kata yang muncul dalam dokumen

$\sum k n_{i,j}$ = Total seluruh kata yang ada pada dokumen

Idf_t = Frekuensi kemunculan kata t pada seluruh dokumen

df_t = Dokumen yang mengandung t

$W_{t,d}$ = Bobot kata dalam suatu data

Tahap pembobotan TF-IDF menggunakan input hasil *preprocessing* pada tahap terakhir yaitu *stemming*. Model TF-IDF digunakan untuk mengubah teks artikel berita menjadi bentuk vektor TF-IDF berupa angka-angka yang merepresentasikan bobot setiap kata berdasarkan frekuensi kemunculan kata dan kepentingannya dalam seluruh dokumen. Pada penelitian ini, hasil ekstraksi fitur TF-IDF awalnya bersifat kontinu dikonversi menjadi data diskrit yang dikategorikan ke dalam 20 rentang atau bin. Rentang ini dimulai dari 0 hingga 1 dengan interval 0.05 untuk setiap bin. Bin 0 mencakup nilai dari 0 hingga kurang dari 0.05, bin 1 mencakup nilai dari 0.05 hingga kurang dari 0.1, dan seterusnya. Proses ini berlanjut dengan interval yang sama hingga bin 19 yang mencakup nilai dari 0.95 hingga 1. Konversi ini dilakukan untuk memastikan bahwa hasil TF-IDF sesuai dengan persyaratan model klasifikasi MNB yang membutuhkan data dalam bentuk diskrit. Hasil TF-IDF dalam bentuk diskrit selanjutnya menjadi input dalam

model klasifikasi MNB. Hasil TF-IDF pada masing-masing teks artikel berita ditunjukkan pada Tabel 3.7 berikut.

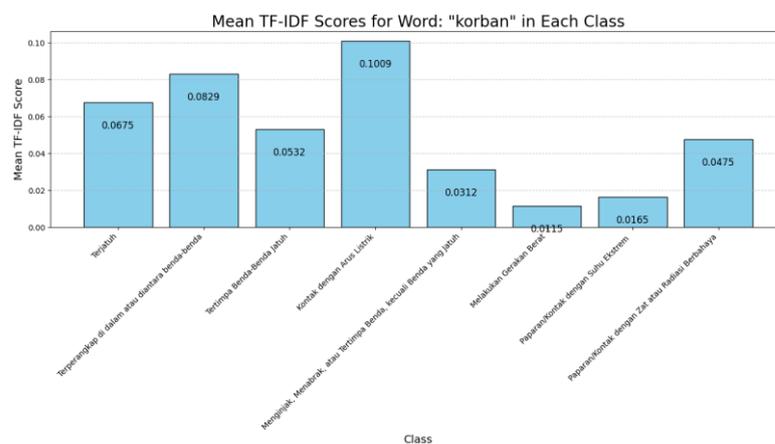
Tabel 3.7 Dataset sesudah tahap TF-IDF

<i>List Hasil Preprocessing</i>	Vektor TF-IDF	TF-IDF dengan Data Diskrit
kerja hotel medan jatuh kanopi meter	[0.17540781, 0.50021173, 0.43055563, ..., 0, 0, 0]	[3, 10, 8, ..., 0, 0, 0]
detik detik crane mayor roboh timpa rumah luka orang	[0.36122969, 0.70008565, 0.26693940, ..., 0, 0, 0]	[7, 14, 5, ..., 0, 0, 0]
celaka kerja jam heri jepit alat berat susun kontainer jict	[0.23613833, 0.43311769, 0.41615901, ..., 0, 0, 0]	[4, 8, 8, ..., 0, 0, 0]
orang kerja sengat listrik cikarang utara	[0.70841457, 0.27928636, 0.19077223, ..., 0, 0, 0]	[14, 5, 3, ..., 0, 0, 0]

3.2.3 Data Karakteristik

Karakteristik data pada penelitian ini menggunakan distribusi multinomial yang dianalisis berdasarkan nilai TF-IDF yang dikonversi menjadi data diskrit. Distribusi multinomial cocok untuk data diskrit dimana setiap fitur dapat memiliki beberapa kategori dengan probabilitas tertentu. Ketika data didistribusikan, rata-rata nilai antar kelas cenderung berdekatan bahkan sama, sehingga tidak tampak adanya perubahan kelas yang signifikan. Hasil histogram distribusi fitur-fitur setiap dokumen dapat dilihat pada Gambar 3.9 dan Lampiran II dimana sumbu x ditampilkan skor tf-idf pada fitur kata tersebut. Sedangkan pada sumbu y ditampilkan jumlah kemunculan kata pada setiap dokumen. Histogram tersebut menunjukkan bahwa rata-rata nilai antar kelas cenderung berdekatan. Hal tersebut berkaitan dengan karakteristik dari data diskrit menggunakan distribusi multinomial dimana data cenderung tidak menunjukkan variasi yang besar diantara setiap kelasnya dengan setiap kategori memiliki nilai probabilitas tetap. Nilai-nilai yang mendekati rata-rata tersebut bercampur diantara berbagai kelas, akibatnya sulit dalam mengelompokkan data secara jelas, sehingga perlu mendefinisikan data

ini sebagai data kategorikal. Data kategorikal adalah data yang dibagi ke dalam kelompok atau kategori yang berbeda berdasarkan atribut tertentu. Pada penelitian ini, data dibagi ke dalam kelompok berdasarkan kelas. Misalnya, kata “bangunan”, “roboh” yang menunjukkan kelas tertimpa benda-benda jatuh. Data kategorikal memiliki salah satu ciri yaitu fitur kata dalam setiap kelas berbeda dimana setiap kelas memiliki atribut unik.



Gambar 3.9 Distribusi Fitur Kata dalam Setiap Dokumen

Selanjutnya, pengujian distribusi multinomial dilakukan uji normalitas untuk memverifikasi apakah fitur-fitur yang diaplikasikan oleh model berdistribusi normal atau tidak. Data yang mengikuti distribusi normal memiliki kecenderungan untuk menghasilkan nilai residual yang lebih konsisten dan tidak bias. Pada penelitian ini menggunakan pengujian normalitas *Kolmogorov-Smirnov* guna menentukan apakah data sampel berasal dari distribusi normal atau tidak. Dalam membuktikan distribusi tersebut peneliti menggunakan software SPSS dengan menganalisa variabel dependent dan independent sesuai data penelitian. Setelah itu, software akan menampilkan hasil uji normalitas yang berisi mengenai nilai statistik *Kolmogorov-smirnov* yaitu nilai *degrees of freedom* (df) dan nilai signifikansi (Sig).

Nilai signifikansi merupakan kunci dalam menentukan data berdistribusi normal atau tidak dimana apabila nilai $\text{Sig.} > 0.05$, maka dianggap gagal menolak hipotesis nol yang artinya data dianggap memiliki distribusi normal. Sebaliknya, apabila nilai $\text{Sig.} \leq 0.05$, maka dianggap menolak hipotesis nol yang artinya data tidak berdistribusi normal. Hasil pengujian normalitas pada fitur-fitur kata dapat dilihat pada Gambar 3.10 yang menunjukkan bahwa fitur tersebut memiliki distribusi tidak normal dengan nilai signifikansi < 0.05 yang membuktikan bahwa menolak hipotesis nol bahwa distribusi fitur-fitur tersebut berdistribusi normal.

	Tests of Normality					
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
aa	.517	960	<.001	.028	960	<.001
aantadi	.512	960	<.001	.012	960	<.001
aanimbau	.512	960	<.001	.012	960	<.001
membesarkantor	.512	960	<.001	.012	960	<.001
tambun	.512	960	<.001	.012	960	<.001
strip	.516	960	<.001	.022	960	<.001
stroke	.526	960	<.001	.056	960	<.001
membesarsecurity	.512	960	<.001	.012	960	<.001
aamin	.512	960	<.001	.012	960	<.001
adeipekerja	.512	960	<.001	.012	960	<.001
apisaksi	.512	960	<.001	.012	960	<.001
apisalah	.512	960	<.001	.012	960	<.001
apisimak	.516	960	<.001	.021	960	<.001
apisehingga	.516	960	<.001	.021	960	<.001
antusias	.512	960	<.001	.012	960	<.001
aktifitas	.521	960	<.001	.035	960	<.001
aamiin	.512	960	<.001	.012	960	<.001
aamenyenggol	.512	960	<.001	.012	960	<.001

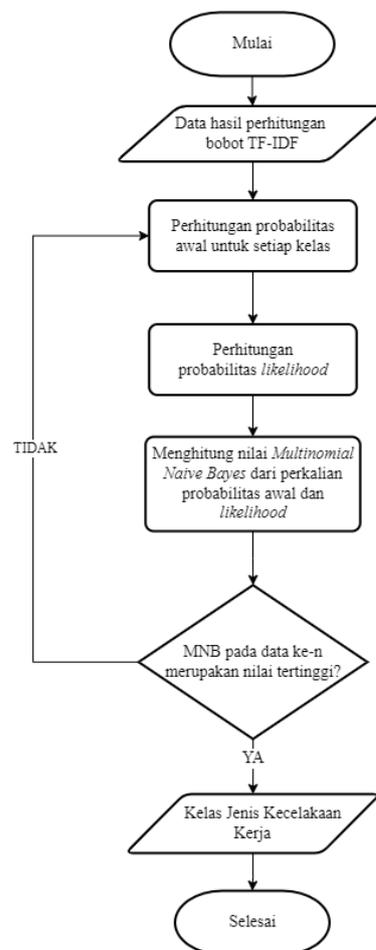
Gambar 3.10 Hasil Uji Normalitas *Kolmogorov Smirnov* dan *Shapiro Wilk*

3.2.4 *Multinomial Naïve Bayes*

MNB adalah salah satu model turunan dari algoritma *naïve bayes* yang cocok sebagai model klasifikasi teks dalam lingkup data bervolume besar. MNB memperkirakan bahwa seluruh variabel saling berkaitan satu sama lain berdasarkan kelasnya dengan melupakan seluruh dependensi antar variabel (Bunga *et al.*, 2018). Cara kerja metode ini dengan melakukan perhitungan probabilitas *token* dalam setiap kelas beserta jumlahnya pada data tersebut dimana data teks dimodelkan

sebagai distribusi multinomial. Kemudian dari hasil perhitungan tersebut dipilih kelas dengan nilai probabilitas posterior paling tinggi sebagai prediksi klasifikasi.

Alur proses klasifikasi MNB ditunjukkan pada Gambar 3.9.



Gambar 3. 11 *Flowchart* Algoritma MNB

Distribusi multinomial adalah generalisasi dari distribusi binomial. Dalam konteks klasifikasi teks, dokumen d yang telah diberi label selanjutnya direpresentasikan $d = \{x_1, x_2, \dots, x_n, c\}$ dengan variabel x yang merepresentasikan suatu kata pada dokumen d , kemudian c merupakan label kelas dari dokumen d . Dalam asumsi *Naïve Bayes* menyatakan bahwa kemunculan kata dan jumlah kemunculannya pada suatu dokumen independen secara kondisional. Rumus

persamaan tersebut dapat dilihat pada Persamaan 3.5 berikut (Yance Nanlohy *et al.*, 2020):

$$P(c|d) = P(c) \prod_{i=1}^n P(X_n|c) \quad (3.4)$$

Keterangan:

$P(c|d)$ = Probabilitas dokumen d berada pada kelas c

$P(c)$ = Probabilitas awal di kelas c

$P(X_n|c)$ = Probabilitas bersyarat kata n muncul pada kelas c (*likelihood*)

Hasil klasifikasi dokumen yaitu memperoleh kelas terbaik dengan nilai perhitungan MNB tertinggi dari beberapa kelas. Hal tersebut dilakukan dengan melakukan perhitungan nilai *maximum a posterior* (MAP) kelas C_{map} pada seluruh kelas yang dapat dilihat pada Persamaan 3.5 (Yance Nanlohy *et al.*, 2020). Untuk P ditulis \hat{P} dikarenakan nilai $P(c|d)$ dan $P(X_n|c)$ yang sebenarnya belum diketahui, sehingga nantinya akan dihitung pada proses *training*.

$$C_{map} = arg \max_{c \in C} \hat{P}(c) \prod_{k=1}^n \hat{P}(X_n|c) \quad (3.5)$$

Pada Persamaan 3.5, apabila dilakukan banyak perkalian probabilitas bersyarat pada suatu dokumen akan menyebabkan terjadinya *floating point underflow*, sehingga diperlukan adanya penjumlahan logaritma dari probabilitas, yaitu $\log(xy) = \log(x) + \log(y)$. Dengan menerapkan persamaan tersebut, maka akan diperoleh kelas hasil klasifikasi pada nilai probabilitas tertinggi. Rumus Persamaan 3.5 diubah menjadi Persamaan 3.6 berikut (Yance Nanlohy *et al.*, 2020):

$$C_{map} = arg \max_{c \in C} [\log \hat{P}(c) + \sum_{1 \leq k \leq n} \log \hat{P}(X_n|c)] \quad (3.6)$$

Untuk menghitung nilai probabilitas *maximum likelihood* dilakukan dengan mendapatkan nilai $\hat{P}(c)$ dan $\hat{P}(X_n|c)$. Pada Persamaan 3.4, variabel c

merepresentasikan kelas, sedangkan variabel d merepresentasikan dokumen, sehingga dapat diinterpretasikan bahwa $(X_n|c)$ dapat mengukur seberapa banyak kemunculan sebuah kata n dalam dokumen yang memberikan kontribusi bahwa c adalah kelas yang tepat. Rumus perhitungan probabilitas *prior* $P(c)$ ditunjukkan pada Persamaan 3.7 berikut.

$$P(c) = \frac{N_{ci}}{N} \quad (3.7)$$

Keterangan:

N_{ci} = Jumlah dokumen pada kelas c index i

N = Jumlah seluruh dokumen

Untuk menghitung nilai probabilitas dari kata n (*likelihood*), dilakukan pendekatan dengan menggunakan distribusi multinomial yang ditunjukkan pada Persamaan 3.8 berikut. Misalkan $x_c = \{x_{c_1}, x_{c_2}, \dots, x_{c_n}\}$ merupakan variabel acak diskrit yang merupakan vektor untuk menghitung jumlah kemunculan kata X yang muncul dengan total N dalam dokumen.

$$\begin{aligned} P(X_n|c) &= \frac{N!}{x_1!x_2!\dots x_i!} \theta_{c_1}^{x_1} \theta_{c_2}^{x_2} \dots \theta_{c_i}^{x_i} \\ &= \frac{N!}{\prod_{i=1}^n x_i!} \prod_{i=1}^V \theta_{c_i}^{x_i} \end{aligned} \quad (3.8)$$

Keterangan:

x_i = Jumlah kemunculan kata ke- n dalam dokumen d

N = Jumlah seluruh kata dalam dokumen kelas c

θ_{c_i} = Probabilitas kemunculan kata ke- n pada kelas c

Estimasi parameter dalam MNB dapat menggunakan *Maximum Likelihood Estimate* (MLE). Metode MLE dapat digunakan apabila didasarkan pada suatu distribusi tertentu, dimana pada MNB didasarkan pada distribusi multinomial. Dalam mencari nilai estimasi parameter menggunakan MLE, perlu menguraikan

fungsi *likelihood* dari *probability mass function* distribusi multinomial, sehingga menghasilkan nilai estimasi parameter θ_{c_i} yang ditunjukkan pada Persamaan 3.9 berikut.

$$\theta_{c_i} = \frac{x_{c_i}}{N} \quad (3.9)$$

Dalam mengatasi probabilitas nol (*zero probability*), dilakukan cara sederhana yaitu menggunakan teknik *Laplace smoothing*. *Laplace smoothing* adalah metode yang digunakan mengatasi nilai peluang nol dengan menambahkan nilai $\alpha = 1$ pada pembilang dan V yang merepresentasikan jumlah kata unik dalam data teks pada penyebut (Harjito *et al.*, 2020).

$$\hat{P}(X_n|c) = \frac{x_{c_i} + \alpha}{N(c) + V} \quad (3.10)$$

Keterangan:

x_{c_i} = Jumlah kemunculan kata ke- n pada kelas c

$N(c)$ = Jumlah seluruh kata yang muncul pada kelas yang diketahui

α = *Add-one smoothing* untuk setiap kata pada kelas c

V = Total kata unik pada seluruh dokumen

BAB IV

UJI COBA DAN PEMBAHASAN

4.1 Langkah-Langkah Uji Coba

Pengujian sistem dalam penelitian ini dilakukan secara sistematis mengikuti serangkaian langkah-langkah pengujian yang telah dirancang. Hal ini bertujuan untuk memastikan bahwa pengujian sistem yang dikembangkan dapat dilakukan secara efisien. Skenario pengujian yang digunakan dalam penelitian ini adalah sebagai berikut.

4.1.1 Data Pengujian

Penelitian ini menggunakan data uji berjumlah 960 data dengan 8 kelas sesuai ketentuan ILO dimana setiap masing-masing kelas terdiri dari 120 teks berita. Dataset bersumber dari portal berita detik.com yang diambil dengan teknik *web scrape*. Dataset diolah melalui tahapan *preprocessing* dan ekstraksi fitur TF-IDF. Selanjutnya, dilakukan pembagian data training dan testing dengan perbandingan rasio 70:30, 80:20, dan 90:10. Data pengujian ditunjukkan pada Tabel 4.1 berikut.

Tabel 4.1 Data Pengujian

Kelas	Jumlah
Terjatuh	120 data
Tertimpa Benda-Benda Jatuh	120 data
Menginjak, Menabrak, atau Tertimpa Benda, kecuali Benda yang Jatuh	120 data
Terperangkap di Dalam atau Diantara Benda-Benda	120 data
Melakukan Gerakan Berat	120 data
Paparan atau Kontak dengan Suhu Ekstrem	120 data
Kontak dengan Arus Listrik	120 data
Paparan atau Kontak dengan Zat Berbahaya	120 data

Pembagian dataset dalam beberapa variasi dilakukan untuk mendapatkan rasio pembagian data terbaik agar mengoptimalkan hasil akurasi. Untuk pengujian awal, persiapannya menggunakan 960 artikel berita yang dibagi dengan rasio 70:30, menghasilkan 672 artikel berita untuk data latih dan 288 artikel berita untuk data uji. Pembagian data tiap kelas memiliki 84 artikel untuk data latih dan 36 artikel untuk data uji. Pada tahap pengujian selanjutnya, untuk rasio 80:20 menghasilkan 768 artikel berita untuk data latih dan 192 artikel berita untuk data uji dengan pembagian data tiap kelas memiliki 96 data latih dan 24 data uji. Kemudian untuk rasio 90:10 menghasilkan 864 artikel berita untuk data latih dan 96 artikel berita untuk data uji. Pembagian data tiap kelasnya yaitu 108 data latih dan 12 data uji. Pembagian dataset dalam beberapa variasi untuk menghasilkan rasio terbaik sehingga dapat memaksimalkan hasil akurasi. Selanjutnya dilakukan penginisialisasian nilai *random seed* untuk mengacak urutan pembagian dataset. Jumlah pembagian data latih dan data uji pada variasi rasio ditunjukkan pada Tabel 4.2 berikut.

Tabel 4.2 Rasio Pembagian Data

Rasio	Jumlah Data <i>Training</i>	Jumlah Data <i>Testing</i>	<i>Random Seed</i>
0.7	672	288	42
0.8	768	192	42
0.9	864	96	42

4.1.2 Mengukur Kinerja Sistem

Pada penelitian ini menggunakan tahap pengukuran kinerja sistem dengan konsep *confusion matrix* untuk mengukur tingkat akurasi dari model klasifikasi. Tahap ini dilakukan dengan skenario pembagian data rasio 70:30, 80:20, dan 90:10. Konsep *confusion matrix* ini menunjukkan hasil perbandingan antara kelas prediksi

dan aktual dengan menentukan nilai *True Positive* (TP), *False Positive* (FP), *True Negative* (TN), dan *False Negative* (FN). *True Positive* (TP) merupakan hasil nilai positif yang diklasifikasikan dengan benar oleh model sebagai nilai positif. *True Negative* (TN) merupakan hasil nilai negatif yang diklasifikasikan dengan benar oleh model sebagai nilai negatif. *False Positive* (FP) merupakan hasil nilai negatif yang diklasifikasikan salah oleh model sebagai nilai positif. *False Negative* (FN) merupakan hasil nilai positif yang diklasifikasikan salah oleh model sebagai nilai negatif. Dari *Confusion Matrix* tersebut dapat dilakukan perhitungan nilai akurasi, presisi, *recall*, dan *f-1 score*.

Akurasi merupakan nilai keakuratan dari hasil perbandingan nilai aktual dengan nilai prediksi untuk mengukur tingkat keberhasilan model klasifikasi. Nilai akurasi dapat dirumuskan dengan Persamaan 4.1.

$$Akurasi = \frac{TP+TN}{TP+FN+FP+TN} \quad (4.1)$$

Presisi merupakan metrik yang diperoleh berdasarkan nilai aktual untuk mengukur tingkat akurat model dalam mengidentifikasi kelas positif. Nilai presisi dapat dirumuskan dengan Persamaan 4.2.

$$Presisi = \frac{TP}{TP+FP} \quad (4.2)$$

Recall merupakan metrik yang digunakan untuk mengukur kemampuan model dalam mendeteksi seluruh sampel positif yang benar dalam sebuah data. Nilai *recall* dapat dirumuskan dengan Persamaan 4.3.

$$Recall = \frac{TP}{TP+FN} \quad (4.3)$$

F-1 Score merupakan nilai bobot rata-rata harmonik antara nilai presisi dan *recall* untuk mengukur tingkat keseimbangan antar kelas positif dan negatif. Nilai *f-1 score* dapat dirumuskan dengan Persamaan 4.4 berikut:

$$F - 1 \text{ Score} = 2 \times \frac{\text{Recall} \times \text{Presisi}}{\text{Recall} + \text{Presisi}} \quad (4.4)$$

Selanjutnya, dilakukan pengujian *k-fold cross validation* dengan pembagian data 90:10 karena menunjukkan hasil akurasi tertinggi melalui rasio perbandingan tersebut. Pertama, dataset dibagi menjadi *k* subset atau *fold* yang berukuran sama. Kemudian, model dilatih menggunakan *k-1 fold* sebagai data pelatihan dan diuji pada *fold* yang tersisa sebagai data uji. Proses ini diulangi 10 kali, setiap kali menggunakan *fold* yang berbeda sebagai data uji dan *fold* yang tersisa sebagai data pelatihan. Setelah semua *k*-iterasi selesai, hasil evaluasi dari setiap iterasi dirata-rata untuk memberikan gambaran keseluruhan tentang kinerja model. Selanjutnya, melakukan evaluasi menggunakan *confusion matrix* melalui perhitungan akurasi, presisi, *recall*, dan *f-1 score* dengan hasil yang ditunjukkan pada Gambar 4.1.

		K-Fold Cross Validation									
Iterasi	iterasi-1	testing	training								
	iterasi-2	training	testing	training							
	iterasi-3	training	training	testing	training						
	iterasi-4	training	training	training	testing	training	training	training	training	training	training
	iterasi-5	training	training	training	training	testing	training	training	training	training	training
	iterasi-6	training	training	training	training	training	testing	training	training	training	training
	iterasi-7	training	training	training	training	training	training	testing	training	training	training
	iterasi-8	training	training	training	training	training	training	training	testing	training	training
	iterasi-9	training	training	training	training	training	training	training	training	testing	training
	iterasi-10	training	training	training	training	training	training	training	training	training	testing
		Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Fold-6	Fold-7	Fold-8	Fold-9	Fold-10
		Fold									

Gambar 4.1 Skenario Pengujian *10-Fold Cross Validation*

4.2 Hasil Uji Coba

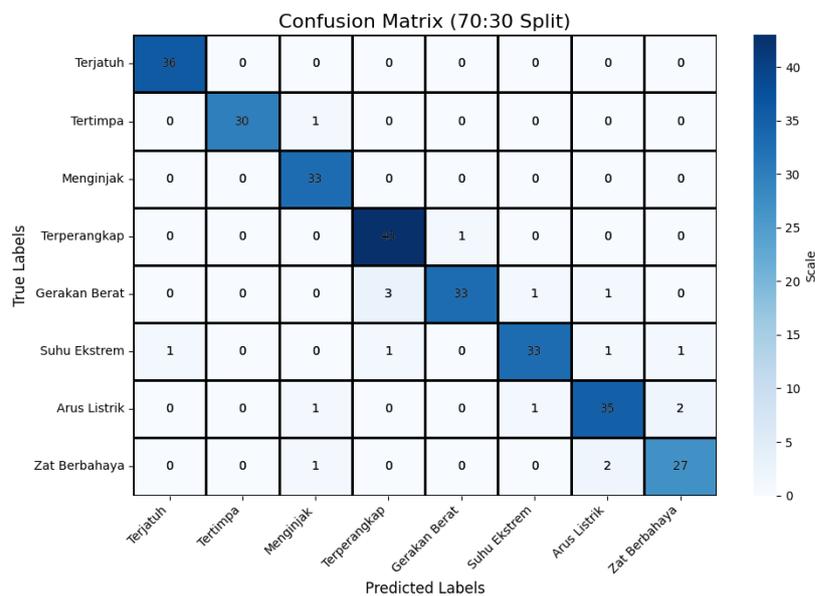
Berdasarkan skenario pengujian yang telah dijabarkan, hasil uji coba pada pembagian data latih dan data uji dengan perbandingan rasio 70:30, 80:20, dan 90:10 untuk menguji kinerja model pada pembagian data dengan rasio berbeda. Hasil model klasifikasi MNB dengan rasio pembagian data 70:30 dapat dilihat pada Tabel 4.3, sedangkan nilai *confusion matrix* ditunjukkan pada Gambar 4.2 berikut.

Tabel 4.3 Hasil Uji Coba 70:30

NO.	Index Berita	True Label	Predicted Label	TP	TN	FP	FN
1	836	Tertimpa Benda-Benda Jatuh	Tertimpa Benda-Benda Jatuh	1	0	0	0
2	477	Paparan/Kontak dengan Zat atau Radiasi Berbahaya	Paparan/Kontak dengan Zat atau Radiasi Berbahaya	1	0	0	0
3	350	Melakukan Gerakan Berat	Melakukan Gerakan Berat	1	0	0	0
4	893	Menginjak, Menabrak, atau Tertimpa Benda, kecuali Benda yang Jatuh	Menginjak, Menabrak, atau Tertimpa Benda, kecuali Benda yang Jatuh	1	0	0	0
5	923	Menginjak, Menabrak, atau Tertimpa Benda, kecuali Benda yang Jatuh	Menginjak, Menabrak, atau Tertimpa Benda, kecuali Benda yang Jatuh	1	0	0	0
6	261	Terperangkap di dalam atau diantara benda-benda	Tertimpa Benda-Benda Jatuh	0	0	0	1
7	215	Menginjak, Menabrak, atau Tertimpa Benda, kecuali Benda yang Jatuh	Menginjak, Menabrak, atau Tertimpa Benda, kecuali Benda yang Jatuh	1	0	0	0
8	334	Melakukan Gerakan Berat	Melakukan Gerakan Berat	1	0	0	0
9	86	Terjatuh	Terjatuh	1	0	0	0
10	878	Paparan/Kontak dengan Suhu Ekstrem	Paparan/Kontak dengan Suhu Ekstrem	1	0	0	0
...
282	895	Menginjak, Menabrak, atau Tertimpa Benda, kecuali Benda yang Jatuh	Menginjak, Menabrak, atau Tertimpa Benda, kecuali Benda yang Jatuh	1	0	0	0
283	668	Paparan/Kontak dengan Zat atau Radiasi Berbahaya	Paparan/Kontak dengan Zat atau Radiasi Berbahaya	1	0	0	0
284	501	Paparan/Kontak dengan Suhu Ekstrem	Paparan/Kontak dengan Suhu Ekstrem	1	0	0	0

Tabel 4.3 Hasil Uji Coba 70:30

285	796	Terperangkap di dalam atau diantara benda-benda	Terperangkap di dalam atau diantara benda-benda	1	0	0	0
286	634	Paparan/Kontak dengan Zat atau Radiasi Berbahaya	Paparan/Kontak dengan Zat atau Radiasi Berbahaya	1	0	0	0
287	405	Melakukan Gerakan Berat	Melakukan Gerakan Berat	1	0	0	0
288	741	Kontak dengan Arus Listrik	Kontak dengan Arus Listrik	1	0	0	0



Gambar 4.2 Visualisasi Prediksi Data Rasio 70:30

Berdasarkan Gambar 4.2 didapatkan hasil perhitungan *confusion matrix* untuk presisi, *recall*, dan *f-1 score* ditunjukkan pada Tabel 4.4 sebagai berikut.

Tabel 4.4 Hasil Perhitungan Presisi, *Recall*, *F-1 Score* pada rasio 70:30

Kelas	Confusion Matrix		
	Presisi	Recall	F-1 Score
Terjatuh	90%	73%	81%
Tertimpa Benda-Benda Jatuh	57%	83%	68%
Menginjak, Menabrak, atau Tertimpa Benda, kecuali Benda yang Jatuh	72%	79%	75%
Terperangkap di Dalam atau Diantara Benda-Benda	81%	74%	77%
Melakukan Gerakan Berat	85%	90%	88%
Paparan atau Kontak dengan Suhu Ekstrem	90%	86%	88%
Kontak dengan Arus Listrik	90%	100%	95%
Paparan atau Kontak dengan Zat Berbahaya	86%	79%	82%

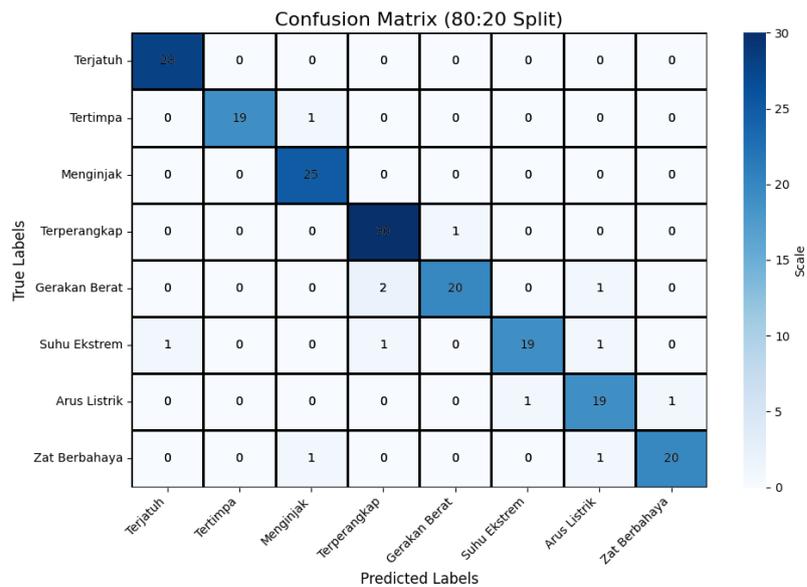
Selanjutnya, hasil klasifikasi kecelakaan kerja pada pembagian data latihan 80% dan data uji 20% ditunjukkan pada Tabel 4.5, sedangkan nilai *confusion matrix* dapat dilihat pada Gambar 4.3 berikut.

Tabel 4.5 Hasil Uji Coba 80:20

NO.	Index Berita	True Label	Predicted Label	TP	TN	FP	FN
1	836	Tertimpa Benda-Benda Jatuh	Tertimpa Benda-Benda Jatuh	1	0	0	0
2	477	Paparan/Kontak dengan Zat atau Radiasi Berbahaya	Paparan/Kontak dengan Zat atau Radiasi Berbahaya	1	0	0	0
3	350	Melakukan Gerakan Berat	Melakukan Gerakan Berat	1	0	0	0
4	893	Menginjak, Menabrak, atau Tertimpa Benda, kecuali Benda yang Jatuh	Menginjak, Menabrak, atau Tertimpa Benda, kecuali Benda yang Jatuh	1	0	0	0
5	923	Menginjak, Menabrak, atau Tertimpa Benda, kecuali Benda yang Jatuh	Menginjak, Menabrak, atau Tertimpa Benda, kecuali Benda yang Jatuh	1	0	0	0
6	261	Terperangkap di dalam atau diantara benda-benda	Terperangkap di dalam atau diantara benda-benda	1	0	0	0
7	215	Menginjak, Menabrak, atau Tertimpa Benda, kecuali Benda yang Jatuh	Menginjak, Menabrak, atau Tertimpa Benda, kecuali Benda yang Jatuh	1	0	0	0
8	334	Melakukan Gerakan Berat	Melakukan Gerakan Berat	1	0	0	0
9	86	Terjatuh	Terjatuh	1	0	0	0
10	878	Paparan/Kontak dengan Suhu Ekstrem	Paparan/Kontak dengan Suhu Ekstrem	1	0	0	0
...
183	739	Kontak dengan Arus Listrik	Kontak dengan Arus Listrik	1	0	0	0
184	866	Paparan/Kontak dengan Zat atau Radiasi Berbahaya	Paparan/Kontak dengan Zat atau Radiasi Berbahaya	1	0	0	0
185	554	Paparan/Kontak dengan Suhu Ekstrem	Paparan/Kontak dengan Suhu Ekstrem	1	0	0	0
186	312	Menginjak, Menabrak, atau Tertimpa Benda, kecuali Benda yang Jatuh	Menginjak, Menabrak, atau Tertimpa Benda, kecuali Benda yang Jatuh	1	0	0	0
187	954	Kontak dengan Arus Listrik	Kontak dengan Arus Listrik	1	0	0	0
188	462	Paparan/Kontak dengan Zat atau Radiasi Berbahaya	Paparan/Kontak dengan Zat atau Radiasi Berbahaya	1	0	0	0

Tabel 4.5 Hasil Uji Coba 80:20

189	673	Paparan/Kontak dengan Zat atau Radiasi Berbahaya	Tertimpa Benda-Benda Jatuh	0	0	0	1
190	33	Terjatuh	Terjatuh	1	0	0	0
191	31	Terjatuh	Terjatuh	1	0	0	0
192	714	Kontak dengan Arus Listrik	Kontak dengan Arus Listrik	1	0	0	0



Gambar 4.3 Visualisasi Prediksi Data Rasio 80:20

Berdasarkan Gambar 4.3 didapatkan hasil *confusion matrix* untuk perhitungan presisi, *recall*, dan *f-1 score* dapat dilihat pada Tabel 4.6 sebagai berikut.

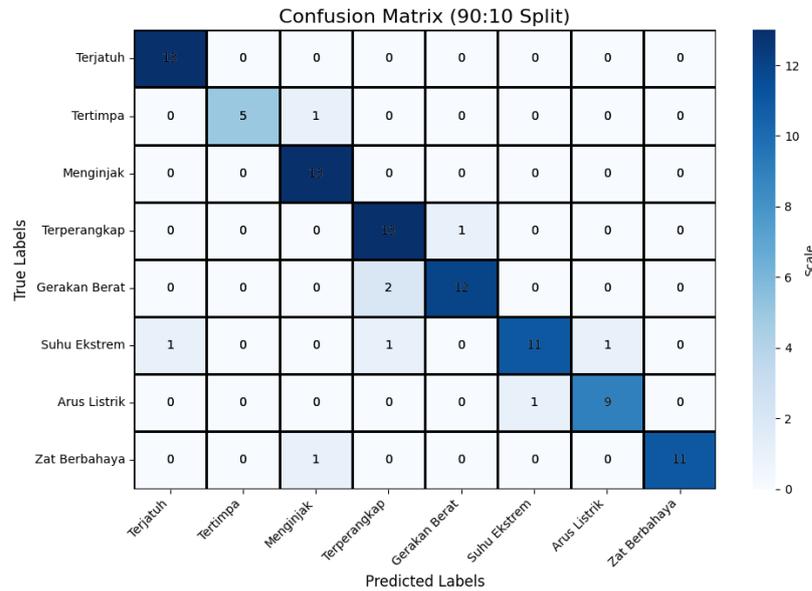
Tabel 4.6 Hasil Perhitungan Presisi, *Recall*, *F-1 Score* pada rasio 80:20

Kelas	Confusion Matrix		
	Presisi	Recall	F-1 Score
Terjatuh	79%	68%	73%
Tertimpa Benda-Benda Jatuh	73%	86%	79%
Menginjak, Menabrak, atau Tertimpa Benda, kecuali Benda yang Jatuh	88%	88%	88%
Terperangkap di Dalam atau Diantara Benda-Benda	80%	76%	78%
Melakukan Gerakan Berat	95%	95%	95%
Paparan atau Kontak dengan Suhu Ekstrem	93%	84%	88%
Kontak dengan Arus Listrik	100%	100%	100%
Paparan atau Kontak dengan Zat Berbahaya	78%	78%	78%

Selanjutnya, Hasil model klasifikasi MNB dengan rasio pembagian data *training* 90% dan data *testing* 10% dapat dilihat pada Tabel 4.7, sedangkan nilai *confusion matrix* ditunjukkan pada Gambar 4.4 berikut.

Tabel 4.7 Hasil Uji Coba 90:10

NO.	Index Berita	True Label	Predicted Label	TP	TN	FP	FN
1	836	Tertimpa Benda-Benda Jatuh	Tertimpa Benda-Benda Jatuh	1	0	0	0
2	477	Paparan/Kontak dengan Zat atau Radiasi Berbahaya	Paparan/Kontak dengan Zat atau Radiasi Berbahaya	1	0	0	0
3	350	Melakukan Gerakan Berat	Melakukan Gerakan Berat	1	0	0	0
4	893	Menginjak, Menabrak, atau Tertimpa Benda, kecuali Benda yang Jatuh	Menginjak, Menabrak, atau Tertimpa Benda, kecuali Benda yang Jatuh	1	0	0	0
5	923	Menginjak, Menabrak, atau Tertimpa Benda, kecuali Benda yang Jatuh	Menginjak, Menabrak, atau Tertimpa Benda, kecuali Benda yang Jatuh	1	0	0	0
6	261	Terperangkap di dalam atau diantara benda-benda	Terperangkap di dalam atau diantara benda-benda	1	0	0	0
7	215	Menginjak, Menabrak, atau Tertimpa Benda, kecuali Benda yang Jatuh	Menginjak, Menabrak, atau Tertimpa Benda, kecuali Benda yang Jatuh	1	0	0	0
8	334	Melakukan Gerakan Berat	Melakukan Gerakan Berat	1	0	0	0
9	86	Terjatuh	Terjatuh	1	0	0	0
10	836	Tertimpa Benda-Benda Jatuh	Tertimpa Benda-Benda Jatuh	1	0	0	0
...
90	605	Paparan/Kontak dengan Suhu Ekstrem	Paparan/Kontak dengan Suhu Ekstrem	1	0	0	0
91	423	Melakukan Gerakan Berat	Melakukan Gerakan Berat	1	0	0	0
92	814	Terjatuh	Terjatuh	1	0	0	0
93	518	Paparan/Kontak dengan Suhu Ekstrem	Paparan/Kontak dengan Suhu Ekstrem	1	0	0	0
94	770	Kontak dengan Arus Listrik	Kontak dengan Arus Listrik	1	0	0	0
95	286	Tertimpa Benda-Benda Jatuh	Tertimpa Benda-Benda Jatuh	1	0	0	0
96	597	Paparan/Kontak dengan Suhu Ekstrem	Paparan/Kontak dengan Suhu Ekstrem	1	0	0	0



Gambar 4.4 Visualisasi Prediksi Data Rasio 90:10

Berdasarkan Gambar 4.4 didapatkan hasil perhitungan *confusion matrix* untuk presisi, *recall*, dan *f-1 score* ditunjukkan pada Tabel 4.8 sebagai berikut.

Tabel 4.8 Hasil Perhitungan Presisi, *Recall*, *F-1 Score* pada rasio 90:10

Kelas	Confusion Matrix		
	Presisi	Recall	F-1 Score
Terjatuh	92%	79%	85%
Tertimpa Benda-Benda Jatuh	100%	92%	96%
Menginjak, Menabrak, atau Tertimpa Benda, kecuali Benda yang Jatuh	75%	92%	83%
Terperangkap di Dalam atau Diantara Benda-Benda	91%	100%	95%
Melakukan Gerakan Berat	100%	100%	100%
Paparan atau Kontak dengan Suhu Ekstrem	87%	93%	90%
Kontak dengan Arus Listrik	93%	100%	96%
Paparan atau Kontak dengan Zat Berbahaya	85%	79%	81%

Hasil uji coba dengan perbandingan 3 rasio yang berbeda-beda melalui pembagian data *training* dan data *testing* menggunakan MNB ditunjukkan pada Tabel 4.9 berikut.

Tabel 4.9 Rata-Rata Akurasi, Presisi, *Recall*, *F-1 Score* pada 3 Rasio Berbeda

Rasio	Akurasi	Presisi	Recall	F-1 Score
70:30	83%	83%	83%	83%
80:20	85%	85%	85%	85%
90:10	88%	88%	88%	87%

Berdasarkan hasil uji coba perbandingan rasio pembagian data tersebut diperoleh hasil tertinggi pada rasio pembagian data 90:10. Hal tersebut menunjukkan bahwa sistem mampu melakukan klasifikasi kecelakaan kerja secara maksimal pada rasio 90% untuk data latih dan 10% untuk data uji. Pada saat uji coba dengan rasio 70:30, nilai akurasi mengalami penurunan, kemudian pada rasio 80:20 nilai akurasi mengalami kenaikan. Hal tersebut menunjukkan bahwa penambahan data *training* mempengaruhi hasil kinerja *Multinomial Naïve Bayes* pada klasifikasi kecelakaan kerja.

Tahap selanjutnya adalah hasil uji coba *10-fold cross validation* dengan 10 iterasi untuk pembagian data latih dan data uji. Hasil pengujian *10-fold cross validation* dapat dilihat pada Tabel 4.10 yang menghasilkan nilai rata-rata akurasi sebesar 84%. Hasil akurasi tertinggi terdapat pada iterasi ke-9 dengan nilai akurasi, presisi, *recall*, dan *f-1 score* berturut-turut adalah 89.58%, 90.92%, 89.58%, dan 89.72%.

Tabel 4.10 Hasil *10-Fold Cross Validation*

Iterasi ke-	Akurasi	Presisi	<i>Recall</i>	<i>F-1 Score</i>
1	87.50%	88.26%	87.50%	87.23%
2	81.25%	83.12%	81.25%	81.70%
3	84.38%	85.40%	84.38%	84.37%
4	81.25%	82.52%	81.25%	81.17%
5	86.46%	89.08%	86.46%	87.00%
6	84.38%	85.75%	84.38%	84.35%
7	81.25%	82.56%	81.25%	81.18%
8	84.38%	84.85%	84.38%	84.09%
9	89.58%	90.92%	89.58%	89.72%
10	80.21%	82.28%	80.21%	80.48%
Rata-Rata	84.06%	85.47%	84.06%	84.13%

4.3 Pembahasan

Hasil uji coba pembagian variasi rasio data diperoleh hasil tertinggi pada rasio 90:10 dengan nilai akurasi, presisi, *recall*, dan *f-1 score* berturut-turut adalah

88%, 88%, 88%, dan 87%. Akurasi tinggi menunjukkan bahwa kinerja model sangat baik dalam mengklasifikasikan data. Pada saat uji coba dengan rasio 80:20, akurasi mengalami penurunan, namun tingkat akurasinya cukup dekat dengan rasio 90:10 yaitu sebesar 85%. Kemudian pada rasio 70:30 terjadi penurunan akurasi yang menyebabkan kinerja model lebih rendah dibanding pengujian rasio lainnya. Penurunan kinerja model klasifikasi tersebut dapat disebabkan oleh kurangnya melatih model dengan baik menggunakan dataset yang mencakup variabel-variabel penting, sehingga model kurang mampu dalam menangkap informasi penting selama tahap pengujian. Sedangkan hasil kenaikan kinerja model tertinggi disebabkan oleh jumlah data *training* untuk melatih model lebih besar dibandingkan pada rasio sebelumnya, sehingga menjadikan model lebih memahami pola data serta memperoleh informasi yang lebih beragam. Model dapat memprediksi klasifikasi kecelakaan kerja dengan baik dan memperoleh nilai akurasi tinggi pada rasio 90:10. Hal tersebut membuktikan bahwa penggunaan rasio pembagian data merupakan aspek yang mempengaruhi kinerja dalam proses klasifikasi kecelakaan kerja pada berita *online* menggunakan *Multinomial Naïve Bayes*.

Hasil evaluasi kinerja model untuk pembagian data rasio 70:30 ditunjukkan pada Tabel 4.4. Pada kelas “Melakukan Gerakan Berat”, “Paparasi atau Kontak dengan Suhu Ekstrem”, dan “Kontak dengan Arus Listrik” memiliki kinerja sangat baik dalam mengenali jenis kecelakaan kerja dilihat dari nilai *f-1 score* tertinggi. Sedangkan untuk kelas “Terjatuh” memiliki kinerja kurang dalam mengidentifikasi beberapa jenis kecelakaan kerja tertentu yang dilihat dari nilai *recall* rendah. Pada

kelas “Menginjak, Menabrak, atau Tertimpa Benda, kecuali Benda Jatuh”, model cukup baik dalam mengklasifikasikan jenis kecelakaan kerja pada kelas tersebut, namun masih terdapat beberapa prediksi yang salah. Selanjutnya, pada kelas “Tertimpa Benda-Benda Jatuh” memiliki kinerja cukup rendah dengan presisi 57% dan *recall* 83% yang membuktikan bahwa meskipun model dapat memprediksi banyak kasus dari kelas ini, namun model sering salah dalam mengklasifikasikan data sebagai kelas “Tertimpa Benda-Benda Jatuh”. Hal ini membuktikan bahwa karakteristik distribusi kelas dan pemilihan fitur yang cukup informatif sangat mempengaruhi kinerja klasifikasi kecelakaan kerja pada berita *online* menggunakan *Multinomial Naïve Bayes*.

Selanjutnya, hasil evaluasi kinerja model pada pembagian data rasio 80:20 ditunjukkan pada Tabel 4.6 mengalami kenaikan signifikan dibanding rasio sebelumnya. Pada kelas “Melakukan Gerakan Berat”, “Paparasi atau Kontak dengan Suhu Ekstrem”, dan “Kontak dengan Arus Listrik” memiliki kinerja sangat baik dalam mengenali jenis kecelakaan kerja dilihat dari nilai *f-1 score* yang tinggi. Namun, pada kelas “Tertimpa Benda-Benda Jatuh” memiliki nilai presisi rendah dengan nilai *recall* tinggi yang berarti model dapat mengenali seluruh kasus pada kelas tersebut, namun terdapat masalah dalam ketepatan prediksi. Selanjutnya pada kelas “Terjatuh” dan “Terperangkap di Dalam atau Diantara Benda-Benda” menunjukkan bahwa model sering gagal dalam mengenali kejadian sebenarnya dari kelas tersebut yang dilihat dari nilai *recall* rendah. Kemudian, untuk rasio data 90:10 yang ditunjukkan pada Tabel 4.8 terjadi peningkatan kinerja dibanding rasio sebelumnya. Pada kelas “Tertimpa Benda-Benda Jatuh”, “Melakukan Gerakan

Berat”, “Terperangkap di Dalam atau Diantara Benda-Benda”, dan “Kontak dengan Arus Listrik” menunjukkan kinerja model sangat baik dalam mengklasifikasikan jenis kecelakaan kerja yang dilihat dari peningkatan nilai presisi, *recall*, dan *f1-Score*. Namun, pada kelas “Terjatuh” dan “Paparan/Kontak dengan Zat Berbahaya” memiliki nilai *recall* terendah yang menunjukkan bahwa model sering gagal dalam mengklasifikasikan jenis kecelakaan kerja pada kelas tersebut. Hal ini membuktikan bahwa perubahan rasio pembagian data *training* dan data *testing* berpengaruh terhadap kinerja klasifikasi kecelakaan kerja pada berita *online* menggunakan *Multinomial Naïve Bayes*. Jumlah data *training* yang lebih banyak menjadikan model klasifikasi lebih optimal dalam mengidentifikasi dan mempelajari pola yang berkaitan dengan kelas tersebut.

Tahap pengujian selanjutnya yaitu *k-fold cross validation* yang digunakan untuk mengevaluasi kinerja model dan memastikan bahwa kinerja model konsisten pada data acak. Pengujian ini melibatkan data uji dan data latih secara acak untuk memastikan model klasifikasi memahami seluruh pola data, sehingga persebaran datanya menjadi lebih representatif dan membantu mengoptimalkan kinerja klasifikasi. Pengujian ini menggunakan 10 *fold* yang menyesuaikan rasio 90:10 pada hasil pengujian sebelumnya dengan nilai akurasi tertinggi. Pada Tabel 4.10 ditunjukkan hasil uji coba *10-fold cross validation* dimana hasil akurasi tertinggi terdapat pada iterasi ke-1 dengan nilai akurasi 89.58%, presisi 90.92%, *recall* 89.58%, dan *f-1 score* 89.72%. Hal ini membuktikan bahwa pengujian *k-fold cross validation* berpengaruh pada kinerja klasifikasi kecelakaan kerja menggunakan *Multinomial Naïve Bayes*.

Berdasarkan hasil pengujian, penggunaan skenario pengujian dengan ketiga rasio berbeda menghasilkan nilai akurasi yang cukup konsisten. Jika dilihat dari hasil uji coba, terdapat perbedaan dalam kinerja MNB kategorisasi pada perbandingan tiga skenario yaitu 70:30, 80:20, dan 90:10. Model klasifikasi MNB memiliki kinerja relatif lebih optimal pada skenario 90:10 dibanding rasio sebelumnya. Peningkatan ukuran data *training* dapat memberikan lebih banyak informasi kepada model untuk mengidentifikasi dan mempelajari pola dalam setiap kelas. Namun, terjadi penurunan kinerja model pada skenario 70:30 dan 80:20 yang dapat dikaitkan dengan kurangnya model dalam mewakili variabel-variabel penting pada dataset, sehingga model kurang mampu dalam menangkap informasi selama tahap pengujian. Secara keseluruhan, perbandingan dari skenario pengujian tersebut menekankan bahwa ukuran data *training*, pemilihan fitur, dan karakteristik dataset berpengaruh signifikan terhadap kinerja model klasifikasi menggunakan *Multinomial Naïve Bayes* dalam mengenali pola dalam kumpulan data.

Penelitian serupa yang dilakukan oleh Dewi (2021) mengenai klasifikasi berita *online* Kompas dan Tempo dalam 7 jenis, yaitu kategori bisnis, politik, hiburan, olahraga, otomotif, kesehatan, dan teknologi. Penelitian tersebut menggunakan metode yang sama yaitu *Multinomial Naïve Bayes*, tetapi memiliki perbedaan pada dataset. Dalam penelitiannya menggunakan pembobotan TF-IDF dengan hasil akurasi sebesar 96% pada rasio 90:10 sebagai rasio terbaik. Selanjutnya, penelitian yang dilakukan oleh Zhafira *et al.* (2021) mengenai analisis sentimen kebijakan kampus merdeka menurut komentar *youtube* dengan menggunakan metode yang sama yaitu *Naïve Bayes*, tetapi memiliki perbedaan

pada topik penelitian dan dataset yang digunakan. Hasil penelitian tersebut mencapai akurasi sebesar 97% pada rasio pembagian 90:10 sebagai rasio terbaik. Walaupun kedua penelitian mencapai akurasi yang lebih tinggi, penelitian yang dilakukan menonjol pada jumlah kategori artikel berita yang lebih banyak dan kesederhanaan implementasi yang memberikan respon akurat.

Kemudian juga terdapat penelitian oleh Singh *et al.* (2019) mengenai pengklasifikasi teks pada analisis sentimen artikel berita dengan kelas positif dan negatif. Penelitian tersebut menggunakan metode yang sama yaitu *Multinomial Naïve Bayes*, tetapi memiliki perbedaan pada dataset dan tahap *preprocessing*. Dalam penelitiannya menggunakan tahap *preprocessing* berupa *lowercasing*, *tokenization*, *punctuation removal*, dan *stopwords removal* dengan hasil akurasi sebesar 73.4%. Hal ini menunjukkan bahwa penelitian yang dilakukan lebih baik karena menggunakan tahap *preprocessing* yang lebih optimal yaitu *cleaning*, *case folding*, *tokenizing*, *stemming*, dan *stopwords removal* dengan hasil akurasi 88% di rasio pembagian data 90:10.

Pada penelitian ini, tahap *preprocessing* menggunakan TF-IDF dimana nilai tersebut dapat digunakan untuk menganalisis kata dalam dokumen. Nilai TF-IDF dapat digunakan untuk menilai seberapa penting suatu kata dalam data dan mengidentifikasi kata-kata yang paling mendominasi atau mewakili dalam sebuah kelas. *Output* dari nilai TF-IDF yaitu memberikan bobot pada setiap kata dalam dokumen yang menunjukkan kepentingan suatu kata dalam setiap kelas. Analisis kata ini diterapkan untuk memahami preferensi dan evaluasi pola jenis kecelakaan kerja.

Representasi visual pada kelas terjatuh dalam bentuk *wordcloud* ditunjukkan pada Gambar 4.5, kemudian bobot TF-IDF setiap kata sesuai ranking ditunjukkan pada Tabel 4.11. Kata-kata yang mendominasi pada kelas ini adalah “kerja”, “jatuh”, “korban”, “lift”, dan “tewas”. Kata-kata tersebut menunjukkan tingginya frekuensi kemunculan kata dalam konteks ini. Adanya kata “proyek” dan “bangun” menegaskan aktivitas dan lokasi berisiko di lingkungan kerja. Kata-kata seperti “jatuh”, “tewas”, “celaka”, dan “korban” mengindikasikan insiden kecelakaan kerja dan dampaknya terhadap tenaga kerja. Selanjutnya, kata-kata “lift” dan “lantai” menunjukkan aspek spesifik di lingkungan kerja yang perlu diawasi dan diperbaiki untuk mencegah terjadinya jenis kecelakaan kerja terjatuh. Hal ini membuktikan bahwa aspek-aspek yang perlu ditingkatkan pada jenis kecelakaan kerja tersebut adalah memastikan pengawasan dan pemeliharaan rutin terhadap infrastruktur yang berisiko seperti lift di lingkungan kerja. Pemeliharaan ini harus mencakup pemeriksaan berkala dan perbaikan segera jika ditemukan kerusakan atau potensi bahaya. Selain itu, menerapkan SOP ketat dan memberikan pelatihan keselamatan kerja terkait jatuh dari ketinggian, terutama di proyek-proyek konstruksi dan lingkungan kerja yang melibatkan bangunan bertingkat.

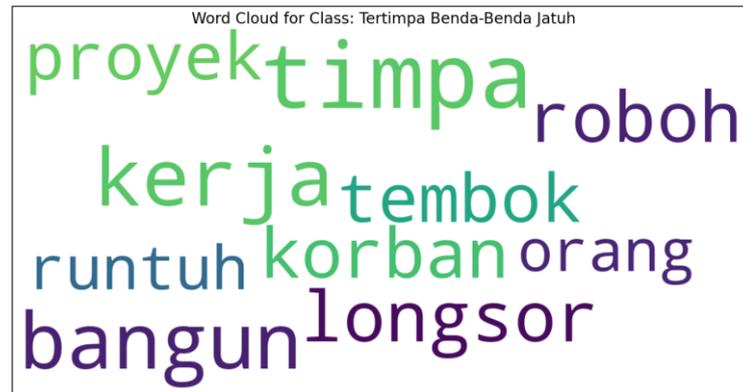


Gambar 4.5 *Wordcloud* Kelas Terjatuh

Tabel 4.11 Ranking Bobot TF-IDF Kelas Terjatuh

No	Kata	Nilai Bobot TF-IDF	Nilai TF-IDF dalam Data Diskrit
1	kerja	0.999821	19
2	jatuh	0.997183	19
3	korban	0.730221	14
4	lift	0.536544	10
5	tewas	0.452075	9
6	proyek	0.362592	7
7	bangun	0.360709	7
8	lantai	0.347914	6
9	celaka	0.336366	6
10	orang	0.318584	6

Selanjutnya, representasi visual pada kelas Tertimpa Benda-Benda Jatuh dalam bentuk *wordcloud* ditunjukkan pada Gambar 4.6, kemudian bobot TF-IDF setiap kata sesuai ranking ditunjukkan pada Tabel 4.12. Kata-kata yang mendominasi pada kelas ini adalah “timpa”, “kerja”, “bangun”, “roboh”, dan “longsor”. Kata-kata ini memiliki nilai TF-IDF tinggi pada kelas tertimpa benda-benda jatuh yang menunjukkan tingginya frekuensi kemunculan kata dalam konteks ini. Adanya kata “timpa”, “tembok”, dan “roboh” menunjukkan bahwa lingkungan kerja dan potensi risiko kecelakaan kerja berasal dari struktur bangunan yang tidak stabil perlu diperhatikan dalam evaluasi keselamatan kerja. Kata “longsor” dan “runtuh” menegaskan pentingnya aspek kondisi tanah dan struktur bangunan di lingkungan kerja. Hal ini membuktikan bahwa aspek-aspek yang perlu ditingkatkan pada jenis kecelakaan kerja tersebut adalah memastikan stabilitas struktur bangunan dan kondisi tanah di lingkungan kerja untuk mencegah risiko runtuhnya bangunan. Penerapan SOP yang ketat terkait konstruksi dan penggalian untuk meminimalkan risiko tertimpa benda-benda jatuh. Selain itu, memberikan pelatihan keselamatan kerja mengenai cara mengidentifikasi dan menghindari area berisiko di lingkungan kerja.

Gambar 4.6 *Wordcloud* Kelas Tertimpa Benda-Benda Jatuh

Tabel 4.12 Ranking Bobot TF-IDF Kelas Tertimpa Benda-Benda Jatuh

No	Kata	Nilai Bobot TF-IDF	Nilai TF-IDF dalam Data Diskrit
1	timpa	0.782619	15
2	kerja	0.761972	15
3	bangun	0.723449	14
4	roboh	0.580337	11
5	longsor	0.579405	11
6	korban	0.563072	11
7	tembok	0.538252	10
8	proyek	0.523178	10
9	runtuh	0.441522	8
10	orang	0.416966	8

Representasi visual pada kelas Menginjak, Menabrak, atau Tertimpa Benda, kecuali Benda yang Jatuh dalam bentuk *wordcloud* ditunjukkan pada Gambar 4.7, kemudian bobot TF-IDF setiap kata sesuai ranking ditunjukkan pada Tabel 4.13. Kata-kata yang mendominasi pada kelas ini adalah “tabrak”, “mobil”, “truk”, dan “kendaraan”. Kata-kata tersebut memiliki bobot TF-IDF tinggi pada kelas ini yang menunjukkan tingginya frekuensi kemunculan kata dalam konteks ini. Adanya kata “tabrak”, “kendara” dan “jalan” menunjukkan jenis kecelakaan kerja yang melibatkan kendaraan di jalan raya. Kata “mobil” dan “truk” mengindikasikan jenis kendaraan yang sering terlibat dalam kecelakaan kerja pada kelas ini. Hal ini membuktikan bahwa aspek-aspek yang perlu ditingkatkan pada jenis kecelakaan

kerja tersebut adalah memastikan pemeliharaan dan kondisi kendaraan, seperti mobil dan truk untuk mencegah risiko kecelakaan kerja terkait kendaraan di jalan raya.



Gambar 4.7 Wordcloud Kelas Menginjak, Menabrak, atau Tertimpa Benda, kecuali Benda yang Jatuh

Tabel 4.13 Ranking Bobot TF-IDF Kelas Menginjak, Menabrak, atau Tertimpa Benda, kecuali Benda yang Jatuh

No	Kata	Nilai Bobot TF-IDF	Nilai TF-IDF dalam Data Diskrit
1	tabrak	0.636937	12
2	mobil	0.608929	12
3	truk	0.585585	11
4	kendaraan	0.558081	11
5	celaka	0.490865	9
6	sopir	0.445904	8
7	jalan	0.384708	7
8	orang	0.362991	7
9	kemudi	0.340807	6
10	kereta	0.297083	5

Selanjutnya, representasi visual pada kelas Terperangkap di dalam atau diantara Benda-Benda dalam bentuk *wordcloud* ditunjukkan pada Gambar 4.8, kemudian bobot TF-IDF setiap kata sesuai ranking ditunjukkan pada Tabel 4.14. Kata-kata yang mendominasi pada kelas ini adalah “korban”, “mesin”, “jepit”, dan “longsor”. Kata-kata tersebut memiliki bobot TF-IDF tinggi pada kelas terjatuh yang menunjukkan tingginya frekuensi kemunculan kata dalam konteks ini. Adanya kata “mesin”, “giling”, dan “lift” menunjukkan aspek spesifik di

lingkungan kerja yang berpotensi sebagai penyebab jenis kecelakaan ini. Kata “jepit” menunjukkan risiko terkait penggunaan benda-benda berat yang menyebabkan terjepit. Selain itu, kata “longsor” menunjukkan risiko bencana alam yang dapat menyebabkan terperangkap. Hal ini membuktikan bahwa aspek-aspek yang perlu ditingkatkan pada jenis kecelakaan kerja tersebut adalah memastikan pemeliharaan rutin pada mesin-mesin yang digunakan di lingkungan kerja serta menerapkan SOP dalam pengoperasian mesin dan alat berat. Langkah-langkah ini bertujuan untuk mencegah adanya malfungsi dan kerusakan teknis yang dapat menyebabkan kecelakaan kerja. Selain itu, perlu dilakukan pelatihan bagi pekerja tentang prosedur keselamatan dan tindakan pencegahan.

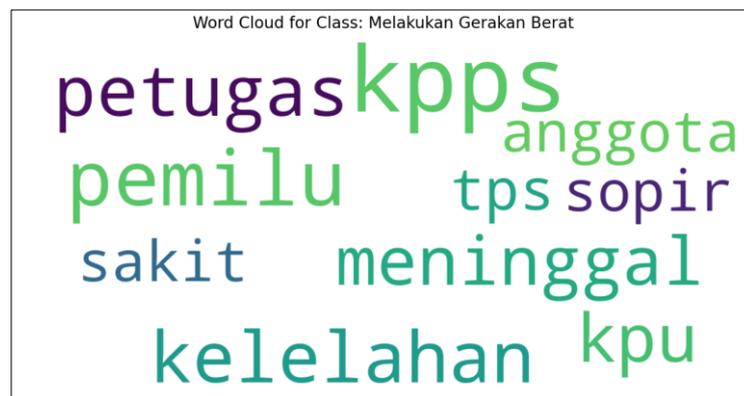


Gambar 4.8 Wordcloud Kelas Terperangkap di dalam atau diantara Benda-Benda

Tabel 4.14 Ranking Bobot TF-IDF Kelas Terperangkap di dalam atau diantara Benda-Benda

No	Kata	Nilai Bobot TF-IDF	Nilai TF-IDF dalam Data Diskrit
1	korban	0.829391	16
2	mesin	0.848335	16
3	jepit	0.796577	15
4	kerja	0.738566	14
5	longsor	0.578303	11
6	evakuasi	0.536515	10
7	giling	0.461246	9
8	tewas	0.438803	8
9	lift	0.435964	8
10	orang	0.355497	7

Representasi visual pada kelas Melakukan Gerakan Berat dalam bentuk *wordcloud* ditunjukkan pada Gambar 4.9, kemudian bobot TF-IDF setiap kata sesuai ranking ditunjukkan pada Tabel 4.15. Kata-kata yang mendominasi pada kelas ini adalah “kpps”, “pemilu”, “kelelahan”, “petugas”, dan “meninggal”. Kata-kata tersebut memiliki bobot TF-IDF tinggi pada kelas melakukan gerakan berat yang menunjukkan tingginya frekuensi kemunculan kata dalam konteks ini. Adanya kata “kpu” dan “tps” menunjukkan bahwa organisasi dan lokasi tempat pemungutan suara menjadi faktor penting yang diperhatikan dalam evaluasi beban kerja. Hal ini membuktikan bahwa aspek-aspek yang perlu ditingkatkan pada jenis kecelakaan kerja tersebut adalah memastikan adanya dukungan yang memadai untuk mengurangi risiko kelelahan dan kesehatan akibat beban kerja yang berat.

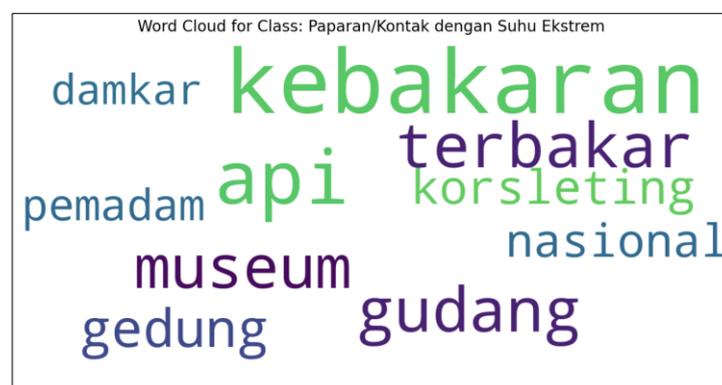


Gambar 4.9 *Wordcloud* Kelas Melakukan Gerakan Berat

Tabel 4.15 Ranking Bobot TF-IDF Kelas Melakukan Gerakan Berat

No	Kata	Nilai Bobot TF-IDF	Nilai TF-IDF dalam Data Diskrit
1	kpps	0.874356	17
2	pemilu	0.743560	14
3	kelelahan	0.672877	13
4	petugas	0.631425	12
5	meninggal	0.515625	10
6	kpu	0.509814	10
7	tps	0.416446	8
8	anggota	0.367916	7
9	sakit	0.357239	7
10	sopir	0.351636	7

Representasi visual pada kelas Paparan/Kontak dengan Suhu Ekstrem dalam bentuk *wordcloud* ditunjukkan pada Gambar 4.10, kemudian bobot TF-IDF setiap kata sesuai ranking ditunjukkan pada Tabel 4.16. Kata-kata yang mendominasi pada kelas ini adalah “kebakaran”, “api”, “gudang”, “terbakar”, dan “museum”. Kata-kata tersebut memiliki bobot TF-IDF tinggi pada kelas ini yang menunjukkan tingginya frekuensi kemunculan kata dalam konteks ini. Adanya kata “damkar” dan “pemadam” menunjukkan bahwa layanan pemadam kebakaran dan respons terhadap insiden kebakaran menjadi hal yang diperhatikan dalam evaluasi keselamatan kerja. Kata “korsleting” dan menegaskan pentingnya perhatian terhadap penyebab kebakaran di lingkungan kerja. Selain itu, kata “gedung” menunjukkan risiko kebakaran terkait dengan bangunan dan infrastruktur. Hal ini membuktikan bahwa aspek-aspek yang perlu ditingkatkan pada jenis kecelakaan kerja tersebut adalah pencegahan kebakaran, penanganan insiden kebakaran, dan memastikan keselamatan risiko paparan suhu ekstrem di berbagai lingkungan kerja dengan dilengkapi sistem deteksi kebakaran yang efektif, seperti alarm asap. Selain itu, perlu dilakukan pemeriksaan rutin terhadap instalasi listrik untuk mencegah korsleting yang dapat memicu kebakaran.



Gambar 4.10 *Wordcloud* Kelas Paparan/Kontak dengan Suhu Ekstrem

Tabel 4.16 Ranking Bobot TF-IDF Kelas Paparan dengan Suhu Ekstrem

No	Kata	Nilai Bobot TF-IDF	Nilai TF-IDF dalam Data Diskrit
1	kebakaran	0.897521	17
2	api	0.539033525	10
3	gudang	0.356674818	7
4	terbakar	0.331549842	6
5	museum	0.240194452	4
6	gedung	0.192401265	3
7	korsleting	0.185575509	3
8	nasional	0.185096858	3
9	pemadam	0.163611085	3
10	damkar	0.152856905	3

Representasi visual pada kelas Kontak dengan Arus Listrik dalam bentuk *wordcloud* ditunjukkan pada Gambar 4.11, kemudian bobot TF-IDF setiap kata sesuai ranking ditunjukkan pada Tabel 4.17. Kata-kata yang mendominasi pada kelas ini adalah sebagai berikut: “listrik”, “korban”, “tersengat”, “tersetrum”, dan “tewas”. Kata-kata tersebut memiliki bobot TF-IDF tinggi pada kelas kontak dengan arus listrik yang menunjukkan risiko terkait arus listrik dan dampaknya terhadap pekerja. Adanya kata “kabel” dan “tiang” menunjukkan bahwa infrastruktur listrik, seperti kabel dan tiang listrik menjadi hal yang penting untuk diperhatikan dalam evaluasi keselamatan kerja. Hal ini membuktikan bahwa aspek-aspek yang perlu ditingkatkan pada jenis kecelakaan kerja tersebut adalah pencegahan insiden tersengat listrik dan memastikan keselamatan pekerja dari risiko kontak dengan arus listrik di berbagai lingkungan kerja dengan pemeliharaan dan pemeriksaan rutin terkait instalasi listrik. Selain itu, menyediakan APD yang sesuai dilengkapi perangkat pengaman, seperti pemutus arus listrik otomatis (*circuit breaker*) di lingkungan kerja yang rentan terhadap risiko listrik.



Gambar 4.11 Wordcloud Kelas Kontak dengan Arus Listrik

Tabel 4.17 Ranking Bobot TF-IDF Kelas Kontak dengan Arus Listrik

No	Kata	Nilai Bobot TF-IDF	Nilai TF-IDF dalam Data Diskrit
1	listrik	0.891262	17
2	korban	0.887644	17
3	tersetrum	0.854364	17
4	tersengat	0.585062	11
5	kabel	0.472317	9
6	tewas	0.416810	8
7	tiang	0.385530	7
8	jebakan	0.382463	7
9	rumah	0.346975	6
10	warga	0.337644	6

Representasi visual pada kelas Paparan/Kontak dengan Zat atau Radiasi Berbahaya dalam bentuk *wordcloud* ditunjukkan pada Gambar 4.12, kemudian bobot TF-IDF setiap kata sesuai ranking ditunjukkan pada Tabel 4.18. Kata-kata yang mendominasi pada kelas ini adalah “gas”, “kebocoran”, “ledakan”, dan “kebakaran”. Kata-kata tersebut memiliki bobot TF-IDF tinggi pada kelas paparan/kontak dengan zat atau radiasi berbahaya yang menandakan bahwa risiko terkait kebocoran gas dan insiden ledakan menjadi hal yang perlu diperhatikan. Adanya kata “pipa” dan “sumur” menunjukkan bahwa infrastruktur seperti pipa gas dan sumur minyak/gas menjadi faktor penting yang diperhatikan dalam evaluasi keselamatan kerja. Kata “pabrik” dan “pt” menegaskan kecelakaan kerja ini sering

terjadi di lingkungan industri dan perusahaan. Hal ini membuktikan bahwa aspek-aspek yang perlu ditingkatkan pada jenis kecelakaan kerja tersebut adalah pencegahan kebocoran dan ledakan gas, penanganan insiden kebakaran yang mungkin terjadi, serta memastikan keselamatan pekerja dari risiko paparan zat berbahaya di lingkungan industri. Selain itu, menyediakan APD yang sesuai dilengkapi pemasangan sistem deteksi gas di lingkungan kerja yang rentan terhadap kebocoran gas.



Gambar 4.12 Wordcloud Kelas Paparan/Kontak dengan Zat atau Radiasi Berbahaya

Tabel 4.18 Ranking Bobot TF-IDF Kelas Paparan/Kontak dengan Zat atau Radiasi Berbahaya

No	Kata	Nilai Bobot TF-IDF	Nilai TF-IDF dalam Data Diskrit
1	gas	0.850712	17
2	ledakan	0.837677	16
3	pt	0.720138	14
4	kebocoran	0.700861	14
5	korban	0.503235	10
6	sumur	0.494911	9
7	kebakaran	0.472669	9
8	pipa	0.432887	8
9	bakar	0.424495	8
10	pabrik	0.398259	7

Hasil analisis kata menggunakan nilai TF-IDF memberikan pemahaman lebih mendalam terhadap aspek-aspek penting yang berhubungan dengan keselamatan kerja. Fitur-fitur (kata) dengan nilai TF-IDF tinggi menunjukkan jenis

kecelakaan kerja yang sering terjadi dan area yang perlu diperhatikan untuk mengurangi risiko kecelakaan kerja. Dengan analisis ini, pengelola proyek atau pelaku industri dapat memanfaatkannya untuk mengidentifikasi aspek-aspek yang perlu ditingkatkan dalam operasi keselamatan kerja dengan merancang strategi pencegahan yang lebih efektif.

Sistem ini dirancang dengan harapan mampu melakukan klasifikasi kecelakaan kerja dari berita *online* secara akurat sehingga dapat memudahkan bidang kesehatan dan keselamatan kerja dalam menganalisis jenis kecelakaan kerja di berbagai sektor industri. Perhitungan kinerja sistem dilakukan dengan mengukur tingkat akurasi agar menghasilkan model yang akurat. Dalam Islam juga dijelaskan mengenai ketepatan pengukuran sebagaimana tercantum dalam qur'an surah Al-Qamar ayat 49 sebagai berikut.

إِنَّا كُلَّ شَيْءٍ خَلَقْنَاهُ بِقَدَرٍ

“*Sesungguhnya Kami menciptakan segala sesuatu menurut ukuran.*” (Q.S Al-Qamar:49)

Berdasarkan tafsir Al-Muyassar, ayat tersebut menjelaskan bahwa Allah SWT. telah menciptakan segala sesuatu sesuai ukuran dan takaran yang ditetapkan yaitu tidak kurang maupun lebih. Hal tersebut telah ditulis dalam Lauhul Mahfudz bahwa segala ciptaan-Nya bersifat mutlak dan sudah ditentukan. Seperti halnya dalam penelitian mengenai klasifikasi kecelakaan kerja pada berita *online* menggunakan *Multinomial Naïve Bayes* yang dibuat dengan memperhatikan konsep ketepatan pengukuran. Dalam penelitian ini, pengukuran kinerja model klasifikasi dilakukan menggunakan *confusion matrix* dengan membandingkan nilai

aktual dan prediksi untuk mengukur tingkat ketepatan model klasifikasi. Hal ini selaras dengan qur'an surah Al-Hijr ayat 19 yang berbunyi:

وَالْأَرْضَ مَدَدْنَاهَا وَأَلْقَيْنَا فِيهَا رُؤُوسَ رُؤُوسِي وَأَنْبَتْنَا فِيهَا مِنْ كُلِّ شَيْءٍ مَوْزُونٍ

“Dan Kami telah menghamparkan bumi dan menjadikan padanya gunung-gunung dan Kami tumbuhkan padanya segala sesuatu menurut ukuran.” (Q.S Al-Hijr:19)

Menurut tafsir Al-Madinah Al-Munawwarah, ayat tersebut menjelaskan bahwa Allah SWT. menciptakan segala yang ada di bumi dengan teratur, detail, dan terukur. Dalam konteks penelitian mengenai klasifikasi kecelakaan kerja pada berita *online* menggunakan *Multinomial Naïve Bayes* memiliki pendekatan pada konsep pengukuran yang teratur dan terukur sesuai dengan surah Al-Hijr:19. Pada penelitian ini, perhitungan kinerja sistem dilakukan secara detail menggunakan *confusion matrix* untuk mengukur seberapa baik model bekerja, sehingga menghasilkan sistem yang akurat. Sistem juga dirancang untuk mengklasifikasikan jenis kecelakaan kerja secara sistematis, sehingga memberikan pemahaman mendalam mengenai pola kejadian dan potensi resiko kecelakaan kerja agar dapat mengambil tindakan pencegahan yang tepat.

Dari Jabir Ibnu Abdullah Radiyallahu ‘anhu berkata:

وَعَنْ جَابِرِ بْنِ عَبْدِ اللَّهِ رَضِيَ اللَّهُ عَنْهُ قَالَ: رَسُولُ اللَّهِ صَلَّى اللَّهُ عَلَيْهِ وَسَلَّمَ نَهَى عَنْ بَيْعِ الْعَرْرِ وَهُوَ بَيْعُ السَّمَكِ فِي الْمَاءِ الَّذِي لَمْ يُصَدِّدْ وَبَيْعِ الطَّيْرِ فِي السَّمَاءِ

“Rasulullah SAW melarang jual beli setumpuk kurma yang tidak diketahui takarannya dengan kurma yang diketahui takarannya.” (H.R Muslim).

Hadits tersebut mengajarkan umat-Nya untuk menggunakan takaran (ukuran) yang jelas dan tepat. Hal ini selaras dengan prinsip ketepatan pengukuran

guna memastikan keakuratan perhitungan. Dalam penelitian ini, dilakukan perhitungan kinerja model dengan menghitung akurasi, presisi, *recall*, dan *f-1 score* untuk mengukur tingkat keakuratan sistem. Diharapkan, hasil implementasi dalam penelitian ini dapat memberikan manfaat bagi bidang kesehatan dan keselamatan kerja untuk memudahkan dalam menganalisis jenis kecelakaan kerja di berbagai sektor industri, sehingga dapat mengenali pola kejadian dan potensi resiko kecelakaan kerja secara dini dengan mengambil tindakan pencegahan yang tepat.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Penelitian ini menggunakan dataset yang didapatkan melalui teknik *scraping* pada portal berita *online* detik.com yang menghasilkan 960 data untuk diproses pada 8 kelas dengan masing-masing jumlah datanya sebanyak 120 artikel berita kecelakaan kerja. Pengujian pertama dilakukan dengan membandingkan rasio data *training* dan data *testing* dimana diperoleh rasio 90:10 menjadi rasio terbaik dengan hasil akurasi 88%, presisi 88%, *recall* 88%, dan *f-1 score* 87%. Pada saat uji coba dengan rasio 80:20, akurasi mengalami penurunan dengan tingkat akurasi sebesar 85%. Kemudian pada rasio 70:30 terjadi penurunan akurasi yang menyebabkan kinerja model lebih rendah dibanding pengujian rasio lainnya. Penurunan kinerja model klasifikasi tersebut disebabkan oleh kurangnya melatih model dengan baik menggunakan dataset yang mencakup variabel-variabel penting, sehingga model kurang mampu dalam menangkap informasi penting selama tahap pengujian. Sedangkan hasil kenaikan kinerja model tertinggi disebabkan oleh jumlah data *training* untuk melatih model lebih besar dibandingkan rasio sebelumnya. Hal ini menjadikan model lebih memahami pola data lebih banyak serta memperoleh informasi yang beragam. Pengujian kedua dilakukan dengan *10-fold cross validation* dengan hasil rata-rata akurasi sebesar 84% dan akurasi tertinggi terletak pada iterasi ke-9 sebesar 89.58%. Selanjutnya, hasil analisis kata menggunakan TF-IDF memberikan pemahaman lebih mendalam terkait aspek-aspek yang perlu ditingkatkan oleh pekerja dan pengelola

keselamatan kerja untuk mengenali pola jenis kecelakaan kerja yang telah terjadi sebelumnya, sehingga dapat mengambil tindakan pencegahan yang tepat. Selain itu, diperoleh bahwa faktor-faktor yang mempengaruhi kinerja *Multinomial Naïve Bayes* dalam proses klasifikasi adalah penggunaan rasio data pelatihan dan data pengujian, pemilihan fitur, serta karakteristik dataset.

5.2 Saran

Setelah melakukan uji coba, peneliti menyadari bahwa masih memiliki beberapa kekurangan. Adapun saran yang diberikan untuk mengembangkan sistem yang telah dibuat agar dapat digunakan dalam penelitian selanjutnya adalah

1. Melakukan penambahan jumlah dataset dengan menggunakan data dari berbagai portal berita *online* selain detik.com, seperti kompas, cnnindonesia, dan tribunnews. Hal tersebut bertujuan untuk menjadikan data lebih bervariasi dan menghasilkan model klasifikasi yang akurat.
2. Mengoptimalkan berbagai fungsi pada tahap *preprocessing* untuk memastikan dataset yang digunakan lebih bersih, karena hal tersebut mempengaruhi hasil kinerja sistem.
3. Menggunakan ekstraksi fitur pembobotan kata yang berbeda dan metode selain *Multinomial Naïve Bayes* untuk membandingkan hasil kinerja terbaik diantara berbagai metode.

DAFTAR PUSTAKA

- Abdullah, S. N., Gusti, S. K., Wulandari, F., & Syafria, F. (2023). Klasifikasi Tingkat Keparahan Korban Kecelakaan Lalu Lintas Menggunakan Naive Bayes. *Jurnal Ilmiah Komputer*, 19(2), 858–866.
- Bunga, M. T. ., S, B., Djahi, & Nabuasa, Y. Y. (2018). Multinomial Naive Bayes Untuk Klasifikasi Status Kredit Mitra Binaan Di Pt . Angkasa Pura I Program Kemitraan. *J-Icon*, 6(2), 30–34. <https://media.neliti.com/media/publications/292504-multinomial-naive-bayes-untuk-klasifikas-2536567f.pdf>
- Dewi, F. K. S. (2021). Klasifikasi Berita Menggunakan Metode Multinomial Naïve Bayes. *SCAN - Jurnal Teknologi Informasi Dan Komunikasi*, 16(3), 1–8. <https://doi.org/10.33005/scan.v16i3.2870>
- Franseda, A., Kurniawan, W., Anggraeni, S., & Gata, W. (2020). Integrasi Metode Decision Tree dan SMOTE untuk Klasifikasi Data Kecelakaan Lalu Lintas. *Jurnal Sistem Dan Teknologi Informasi (Justin)*, 8(3), 282–290. <https://doi.org/10.26418/justin.v8i3.40982>
- Habsyi, S. H. P. (2012). *Pengantar Tajuk Subjek dan Klasifikasi*. Alauddin University Press.
- Hamzah, A. (2012). Klasifikasi Teks dengan Naïve Bayes Classifier (NBC) untuk Pengelompokan Teks Berita dan Abstract Akademis. *Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST) Periode III*, 269–277.
- Harjito, B., Wijayanto, A., Aini, K. N., & Murtiyas, B. (2020). Comparison of Multinomial Naïve Bayes with K-Nearest Neighbors, Support Vector Machine and Random Forest for Classification of “Network Attacks” Document. *Proceedings of 2020 4th International Conference on Informatics and Computing, ICIC 2020*. <https://doi.org/10.1109/ICIC47613.2019.8985919>
- ILO. (1996a). *Recording and Notification of Occupational Accidents and Diseases* (Vol. 148). https://www.ilo.org/wcmsp5/groups/public/---ed_protect/---protrav/---safework/documents/normativeinstrument/wcms_107800.pdf
- ILO. (1996b). *Recording and Notification of Occupational Accidents and Diseases*. International Labour Organization.
- Kementerian Ketenagakerjaan RI. (2022). *Profil Keselamatan dan Kesehatan Kerja Nasional Indonesia Tahun 2022*. KEMENTERIAN KETENAGAKERJAAN REPUBLIK INDONESIA.
- Nisak, S. K., & Nugraha, J. (2018). Penerapan Klasifikasi Decision Tree dan Model Log Linear dalam Penanganan Kecelakaan Kerja. *PRISMA, Prosiding Seminar Nasional ...*, 1, 445–455.

<https://journal.unnes.ac.id/sju/index.php/prisma/article/view/19691>

- Rahman, A., Wiranto, & Doewes, A. (2017). Online News Classification Using Multinomial Naive Bayes. *ITSMART: Jurnal Ilmiah Teknologi Dan Informasi*, 6(1), 32–38. www.kompas.com
- Randhika, M. N., Young, J. C., Suryadibrata, A., & Mandala, H. (2021). Implementasi Algoritma Complement dan Multinomial Naïve Bayes Classifier Pada Klasifikasi Kategori Berita Media Online. *Ultimatics : Jurnal Teknik Informatika*, 13(1), 19–25. <https://doi.org/10.31937/ti.v13i1.1921>
- Setianingrum, A. H., Kalokasari, D. H., & Shofi, I. M. (2017). Implementasi Algoritma Multinomial Naive Bayes Classifier. *Jurnal Teknik Informatika*, 10(2), 109–118. <https://doi.org/10.15408/jti.v10i2.6822>
- Singh, G., Kumar, B., Gaur, L., & Tyagi, A. (2019). Comparison between Multinomial and Bernoulli Naive Bayes for Text Classification. *International Conference on Automation, Computational and Technology Management, ICACTM*, 593–596. <https://doi.org/10.1109/ICACTM.2019.8776800>
- Utami, S. B., & Irsyadi, Y. Al. (2018). Sistem Klasifikasi Kecelakaan Lalu Lintas Jalan Raya Di Kota Boyolali Menggunakan Metode Naive Bayes. *Jurnal Mitra Manajemen*, 2(3), 190–203. <https://doi.org/10.52160/ejmm.v2i3.91>
- Wisdayani, D. S., Manfaati Nur, I., & Wasono, R. (2019). Penerapan Algoritma K-Nearest Neighbor dalam Klasifikasi Tingkat Keparahan Korban Kecelakaan Lalu Lintas di Kabupaten Jawa Tengah. *Prosiding Mahasiswa Seminar Nasional Unimus*, 2, 373–380. <http://prosiding.unimus.ac.id>
- Yance Nanlohy, L., Mulyanto Yuniarno, E., & Mardi Susiki Nugroho, S. (2020). Classification of public complaint data in sms complaint using naive bayes multinomial method. *4th International Conference on Vocational Education and Training, ICOVET*, 241–246. <https://doi.org/10.1109/ICOVET50258.2020.9229941>
- Zhafira, D. F., Rahayudi, B., & Indriati, I. (2021). Analisis Sentimen Kebijakan Kampus Merdeka Menggunakan Naive Bayes dan Pembobotan TF-IDF Berdasarkan Komentar pada Youtube. *Jurnal Sistem Informasi, Teknologi Informasi, Dan Edukasi Sistem Informasi*, 2(1), 55–63. <https://doi.org/10.25126/justsi.v2i1.24>

LAMPIRAN

LAMPIRAN

Lampiran I Dataset

No.	Tanggal	Link	Berita	Label
1	6/6/2023 22:39	https://news.detik.com/berita/d-6758894/pekerja-tewas-terjatuh-dari-lantai-7-di-jakpus-2-orang-lain-luka	<p>Pekerja Tewas Terjatuh dari Lantai 7 di Jakpus, 2 Orang Lain Luka</p> <p>Seorang pekerja bangunan berinisial HA (30) tewas setelah terjatuh dari lantai 7 sebuah bangunan di Gondangdia, Jakarta Pusat (Jakpus). Korban terjatuh diduga karena tali gondola putus."(Penyebab jatuh) Masih didalami. Kalau terlihat di TKP memang terlihat ada gondola yang talinya putus itu saat ini masih kita lakukan pendalaman," kata Kapolres Metro Jakarta Pusat Kombes Komarudin kepada wartawan, Selasa (6/6/2023).Dari kejadian tersebut, ada dua korban lain berinisial P dan S. Korban terjatuh pada Senin (5/6), sekitar pukul 17.30.</p>	Terjatuh
2	11/3/2020 20:07	https://news.detik.com/berita-jawa-timur/d-5240290/polisi-selidiki-penyebab-lift-anjlok-dalam-proyek-pembangunan-di-surabaya	<p>Polisi Selidiki Penyebab Lift Anjlok dalam Proyek Pembangunan di Surabaya</p> <p>Salah seorang kuli bangunan, Wawan, warga Surabaya mengatakan, peristiwa lift anjlok terjadi sekitar pukul 15.55. Ia melihat langsung kejadian tersebut."Tadi saya turun duluan. Tadi (melihat) langsung anjlok ke bawah. Tadi ada enam orang di dalam lift," kata Wawan kepada wartawan di lokasi kejadian, Selasa (3/11/2020)."Iya saya lihat langsung kejadiannya. Waktu jatuh dan mental ke atas terdengar suara teriakan.</p>	Terjatuh
3	9/17/2020 11:30	https://news.detik.com/berita-jawa-	<p>Kecelakaan Kerja Tewaskan 4 Orang, Polisi Panggil Yayasan Unisma-Kontraktor</p>	Terjatuh

Lampiran I Dataset

No.	Tanggal	Link	Berita	Label
		timur/d-5176598/kecelakaan-kerja-tewaskan-4-orang-polisi-panggil-yayasan-unisma-kontraktor	Polisi memanggil yayasan dan kontraktor terkait kecelakaan kerja di proyek RSI Unisma. Sebab, peristiwa lift jatuh itu menewaskan 4 pekerja."Kami akan memanggil pihak yayasan Unisma maupun kontraktor. Kemungkinan dalam minggu ini," ujar Kasat Reskrim Polresta Malang Kota AKP Azi Pratasa Guspitu, Kamis (17/9/2020).Satreskrim Polresta Malang Kota telah mengantongi identitas pelaksana proyek, yaitu PT Dwi Ponggo Seto, yang disebut beralamat di Ponorogo.	
4	13 Agu 2020 04:17	https://news.detik.com/berita/d-5131024/cerita-saksi-soal-tewasnya-4-pekerja-tower-saat-bongkar-sutet-di-sumedang	Cerita Saksi soal Tewasnya 4 Pekerja Tower Saat Bongkar SUTET di Sumedang Tiang penyangga SUTET listrik di Kabupaten Sumedang, Jawa Barat roboh saat sedang dibongkar yang mengakibatkan empat pekerja tewas. Salah seorang warga yang berjaga di proyek pengerjaan SUTET, Deni Abdullah (38), mengatakan keempat bekerja itu jatuh saat sedang membongkar SUTET."Keempatnya lagi kerja di atas (tower SUTET), nggak tau seling yang di bawah copot yang di sawah.	Terjatuh
5	6/2/2020 22:50	https://news.detik.com/berita/d-5038102/1-tewas-4-luka-luka-saat-perbaiki-jembatan-bekas-banjir-bandang-di-sumbar	1 Tewas-4 Luka-luka Saat Perbaiki Jembatan Bekas Banjir Bandang di Sumbar Satu orang tewas dan sejumlah pekerja lainnya terluka akibat ambruknya sisa jembatan yang akan diperbaiki di Sikabu Lubuk Alung, Kabupaten Padang Pariaman. Tiga tahun lalu, jembatan itu ambruk akibat banjir bandang dan akan diperbaiki serta	Terjatuh

Lampiran I Dataset

No.	Tanggal	Link	Berita	Label
			diganti dengan jembatan baru. Kepala Badan Pelaksana BPBD Kabupaten Padang Pariaman Budi Mulya mengatakan peristiwa ini terjadi saat para pekerja sedang berusaha membongkar sisa jembatan. Proses evakuasi sempat mengalami kendala karena korban korban terakhir yang meninggal dunia, jatuh ke dalam sungai dan terimpit besi jembatan.	
6	4/16/2020 17:33	https://news.detik.com/berita-jawa-timur/d-4979809/apd-3-pekerja-pabrik-bioetanol-mojokerto-yang-tewas-ditemukan-di-kolam-pengendapan	APD 3 Pekerja Pabrik Bioetanol Mojokerto yang Tewas Ditemukan di Kolam Pengendapan Kasat Reskrim Polres Mojokerto Kota AKP Sodik Efendi mengatakan, pihaknya telah memeriksa sejumlah saksi terkait kecelakaan kerja ini. Berdasarkan keterangan dua korban selamat, lanjut Sodik, tiga korban tewas memakai APD saat membersihkan lumpur campuran spentwash dan yeast (ragi) di dalam kolam pengendapan. Terdiri dari masker respirator, helm dan sepatu bot. "Yang jelas hasil lidik kami, APD korban ditemukan di TKP. Kondisinya tercebur di dalam kolam pengendapan," terangnya	Terjatuh
7	4/14/2020 16:15	https://news.detik.com/berita-jawa-timur/d-4976719/ini-penyebab-tewasnya-3-pekerja-pabrik-bioetanol-di-mojokerto	Ini Penyebab Tewasnya 3 Pekerja Pabrik Bioetanol di Mojokerto Kasat Reskrim Polres Mojokerto Kota AKP Sodik Efendi mengatakan, ketiga korban tewas karena tenggelam di dalam lumpur campuran spentwash dan yeast (ragi). Lumpur tersebut berada di kolam pengendapan yang	Terjatuh

Lampiran I Dataset

No.	Tanggal	Link	Berita	Label
			akan dibersihkan oleh para korban.Hal itu berdasarkan keterangan sementara dari tim kedokteran Forensik yang mengautopsi jenazah ketiga korban di RS Bhayangkara, Surabaya, Minggu (12/4). Menurut dia, kolam pengendapan di PT Enero mempunyai luas sekitar 4x10 meter persegi. Kedalaman lumpur di dasar kolam sekitar 60-80 cm."Untuk sementara hasil autopsi yang disampaikan secara lisan ke kami, ketiga korban meninggal karena tenggelam di dalam lumpur.	
8	4/12/2020 13:53	https://news.detik.com/berita-jawa-timur/d-4974123/3-pekerja-pabrik-bioetanol-mojokerto-tewas-saat-bersihkan-kolam-pengendapan	3 Pekerja Pabrik Bioetanol Mojokerto Tewas Saat Bersihkan Kolam Pengendapan Insiden tewasnya 3 pekerja pabrik bioetanol PT Energi Agro Nusantara (Enero) di Desa Gempolkrep, Kecamatan Gedeg, Mojokerto, mulai terkuak. Kecelakaan kerja itu terjadi saat para korban membersihkan kolam presettling atau kolam pengendapan.Keterangan perihal insiden yang merenggut nyawa 3 pekerja PT Enero salah satunya dari Kasat Reskrim Polres Mojokerto Kota AKP Sodik Efendi. Dia mengatakan, kecelakaan terjadi saat tiga pekerja yang tewas sedang membersihkan lumpur sisa etanol di dalam kolam pada Jumat (11/4) pagi.	Terjatuh
9	9/20/2022 12:53	https://www.detik.com/jateng/berita/d-6301778/aksi-damkar-evakuasi-	Aksi Damkar Evakuasi Pekerja Terjepit Mesin Cetak di Semarang Insiden kecelakaan kerja terjadi di sebuah percetakan di Jalan Sunan Kalijaga,	Terperangkap di dalam atau diantara benda-benda

Lampiran I Dataset

No.	Tanggal	Link	Berita	Label
		pekerja-terjepit-mesin-cetak-di-semarang	Pedurangan, Kota Semarang, Selasa (20/9/2022). Seorang pekerja terjepit tangannya di sebuah mesin cetak. Petugas pemadam kebakaran mendapatkan laporan pukul 10.42. Laporan itu menyebutkan ada warga yang tangan kanannya terjepit mesin cetak ketika proses produksi. Saat petugas damkar tiba di lokasi, pergelangan tangan kanan korban terjepit di antara dua rol cetak.	
10	9/4/2023 21:30	https://www.detik.com/bali/berita/d-6913417/imb- as-lift-maut-kemenaker-cek-keselamatan-kerja-ayuterra-resort	Imbas Lift Maut, Kemenaker Cek Keselamatan Kerja Ayuterra Resort Tragedi lift putus yang merenggut nyawa lima karyawan Ayuterra Resort di Desa Kedewatan, Ubud, Gianyar, Bali, Jumat (4/9/2023), menjadi perhatian Kementerian Ketenagakerjaan (Kemenaker). Rombongan Kemenaker mendatangi lokasi kejadian, Senin siang (4/9/2023). "Karena ini tentang lift yang jatuh mengakibatkan semua penumpangnya meninggal ya jadi kami selidiki apakah sebelumnya ada uji atau tidak untuk memastikan keamanan dan keselamatan penumpangnya. Baik proses instalasinya, hingga penerapan penggunaannya seperti apa," beber Dinar.	Terjatuh
11	7/6/2023 14:42	https://www.detik.com/su mbagsel/berita/d-6809073/4-korban-tragedi-lift-anjlok-di-	4 Korban Tragedi Lift Anjlok di Sekolah Az-Zahra Lampung Dimakamkan Empat korban tewas dalam tragedi lift barang terjun bebas di Sekolah Az-Zahra dimakamkan. Empat korban di antaranya merupakan warga	Terjatuh

Lampiran I Dataset

No.	Tanggal	Link	Berita	Label
		sekolah-az-zahra-lampung-dimakamkan	Bandar Lampung. Dari penelusuran detikSumbagsel, empat dari tujuh korban tewas ini tinggal di Kelurahan Olok Gading, Kecamatan Teluk Betung Barat, Bandar Lampung. Mereka secara bersama dimakamkan di Tempat Pemakaman Umum (TPU) Tanjung Jati. Pemakaman sendiri berlangsung Kamis (6/7/2023), pukul 10.00.	
12	11/3/2020 21:36	https://news.detik.com/berita-jawa-timur/d-5240388/ini-dugaan-penyebab-lift-anjlok-dalam-proyek-pembangunan-di-surabaya	Ini Dugaan Penyebab Lift Anjlok dalam Proyek Pembangunan di Surabaya Polisi masih menyelidiki penyebab lift anjlok dalam proyek pembangunan showroom mobil di Surabaya. Di lokasi, garis polisi sudah dipasang. Proyek pembangunan showroom ini berada di Jalan Kartini No 35, Surabaya. Lift anjlok sekitar pukul 15.55, dari ketinggian 20 meter. Akibatnya, 6 pekerja bangunan yang menumpang lift tersebut mengalami luka-luka. Semuanya masih menjalani perawatan di rumah sakit.	Terjatuh
13	9/14/2023 16:48	https://www.detik.com/jabar/berita/d-6931306/nasib-tragis-wawan-di-pabrik-penggilingan-gabah-tasikmalaya	Nasib Tragis Wawan di Pabrik Penggilingan Gabah Tasikmalaya Kecelakaan kerja yang berakibat fatal terjadi di pabrik penggilingan gabah di Kampung Perintis, Kelurahan Sukamanah, Kecamatan Cipedes, Kota Tasikmalaya, Kamis (14/9/2023) siang. Seorang pekerja tewas terlilit putaran mesin saat sedang mengoperasikan mesin diesel berkapasitas besar itu.	Terperangkap di dalam atau diantara benda-benda

Lampiran I Dataset

No.	Tanggal	Link	Berita	Label
			Tubuh pria yang diketahui bernama Wawan Saefulloh (58) warga Cisayong Kabupaten Tasikmalaya tersebut ditemukan terjepit atau terlilit di pulley (katrol) berukuran besar.	
...
952	6/1/2020 17:55	https://news.detik.com/berita/d-372978/tabrakan-ka-di-lampung-14-pegawai-pt-ka-diperiksa-di-bandung	Tabrakan KA di Lampung, 14 Pegawai PT KA Diperiksa di Bandung Palembang - Tentunya masih ingat dengan peristiwa tabrakan kereta api (KA) di Bandar Lampung, Lampung, yang menewaskan empat orang beberapa waktu lalu. Buntut dari peristiwa itu, sebanyak 14 pegawai PT KA Sub Divisi Regional III.2 Tanjungkarang, Lampung, diperiksa Tim CO (sidang intern) di kantor Pusat PT KA, Jalan Perintis Kemerdekaan No.1 Bandung.	Menginjak, Menabrak, atau Tertimpa Benda, kecuali Benda yang Jatuh
953	1/29/2024 21:17	https://www.detik.com/jogja/berita/d-7167028/pekerja-ditemukan-tewas-kesetrum-di-atap-gapura-perumahan-di-bantul	Pekerja Ditemukan Tewas Kesetrum di Atap Gapura Perumahan di Bantul Seorang pekerja bangunan ditemukan meninggal dunia di atap gapura masuk salah satu perumahan di Jalan Pleret KM 2.5, Jambidan, Banguntapan, Bantul. Pekerja itu tersengat listrik saat mengecek instalasi air di atap bangunan tersebut. Kasi Humas Polres Bantul, AKP I Nengah Jeffry menjelaskan, kejadian bermula saat korban yakni Aji Budianto (39), warga Sindet, Wukirsari, Imogiri, Bantul mengerjakan renovasi ruko di Jambidan pukul 16.00 tadi.	Kontak dengan Arus Listrik

Lampiran I Dataset

No.	Tanggal	Link	Berita	Label
954	01 Des 2023 01:00	https://www.detik.com/sumbagsel/berita/d-7065679/tersangkut-tali-penambat-kapal-di-bakauheni-rusdian-alami-patah-kaki	Tersangkut Tali Penambat Kapal di Bakauheni, Rusdian Alami Patah Kaki Nahas dialami Rusdian Fadli, pekerja di dermaga Pelabuhan Bakaheuni, Lampung Selatan mengalami patah di bagian kaki kiri usai tersangkut di tali penambat kapal. Dari foto yang diterima detikSumbagsel, tampak bagian kaki kiri korban bengkok karena patah tulang pada bagian betisnya.	Terperangkap di dalam atau diantara benda-benda
955	11/16/2023 14:12	https://www.detik.com/jogja/berita/d-7040179/tersetrum-kuli-bangunan-di-bantul-luka-bakar-di-sekujur-tubuh	Tersetrum, Kuli Bangunan di Bantul Luka Bakar di Sekujur Tubuh Seorang kuli bangunan bernama Antok tersengat aliran listrik saat bekerja di wilayah Kalurahan Sabdodadi, Kapanewon Bantul, Bantul, pagi tadi. Akibatnya, korban mengalami luka bakar di sekujur tubuhnya. Kasi Humas Polres Bantul, Iptu I Nengah Jeffry mengatakan kejadian ini bermula saat korban yang beralamat di Wukirsari, Imogiri, Bantul itu sedang bekerja menyelesaikan pembangunan rumah di wilayah Sabdodadi.	Kontak dengan Arus Listrik
956	07 Des 2023 23:00	https://www.detik.com/sumut/berita/d-7078132/2-pekerja-sumur-bor-di-deli-serdang-tewas-tersengat-listrik	2 Pekerja Sumur Bor di Deli Serdang Tewas Tersengat Listrik Tiga pekerja sumur bor tersengat listrik di Jalan Delitua, Kecamatan Delitua, Deli Serdang. Akibatnya dua pekerja tewas. Satu lainnya dirawat di rumah sakit. Surya Andi Putra selaku warga sekitar mengatakan kejadian itu berlangsung, Kamis (7/12/2023) sekitar pukul 09.30 ."Awalnya mereka lagi	Kontak dengan Arus Listrik

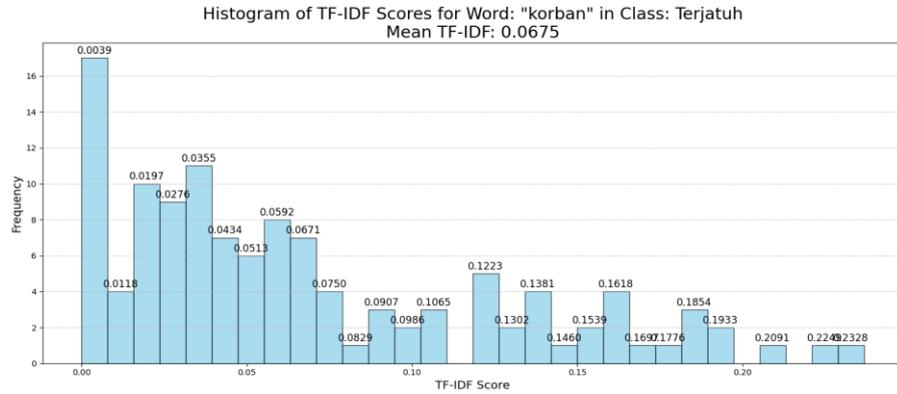
Lampiran I Dataset

No.	Tanggal	Link	Berita	Label
			mengebor sumur air untuk buat instalasi apa gitu. Terus kesetrum karena kena ke kabel listrik. Korban tiga orang, yang meninggal ada dua orang," kata Surya.	
957	11 Des 2023 05:19	https://news.detik.com/internasional/d-7082546/9-orang-tewas-tersengat-listrik-di-sebuah-kamp-di-brasil	<p>9 Orang Tewas Tersengat Listrik di Sebuah Kamp di Brasil</p> <p>Sembilan orang tewas tewas tersengat listrik di sebuah kamp di Brasil utara. Insiden itu terjadi ketika para pekerja yang memasang antena internet secara tidak sengaja menyentuh kabel listrik bertegangan tinggi sehingga menyebabkan tiga karyawan dan enam warga tersengat listrik. Dilansir AFP, Senin (11/12/2023), kecelakaan yang terjadi pada hari Sabtu (9/12) itu juga memicu kebakaran besar di kamp yang didirikan oleh Gerakan Pekerja Tak Bertanah (MST) Brasil di luar kota Parauapebas.</p>	Kontak dengan Arus Listrik
958	20 Des 2023 14:20	https://www.detik.com/suamut/berita/d-7099457/4-pekerja-kesetrum-saat-pasang-baliho-caleg-di-padang-1-tewas	<p>4 Pekerja Kesetrum saat Pasang Baliho Caleg di Padang, 1 Tewas</p> <p>Empat orang pekerja di Kelurahan Pasie Nan Tigo, Kecamatan Koto Tangah, Padang, Sumatera Barat (Sumbar), tersengat arus listrik saat sedang memasang baliho calon legislatif (caleg) DPR RI dari salah satu partai. Akibatnya, satu di antara mereka meninggal dunia. Kasi Humas Polresta Padang Ipda Yanti Devina membenarkan kejadian tersebut. Dia menyebut peristiwa itu terjadi pada Selasa (19/12/2023) sore.</p>	Kontak dengan Arus Listrik

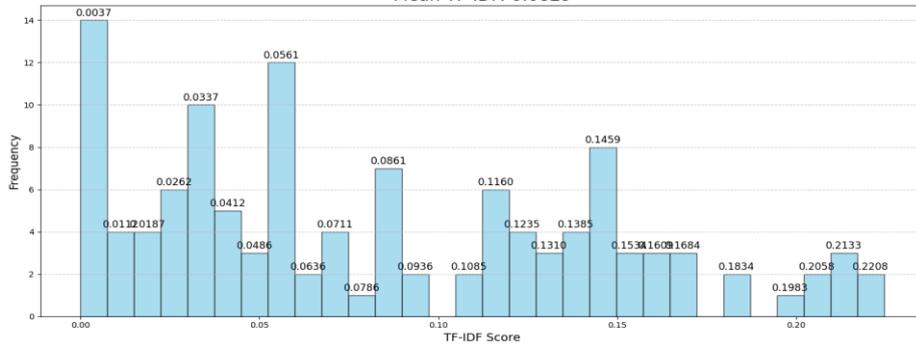
Lampiran I Dataset

No.	Tanggal	Link	Berita	Label
959	22 Des 2023 13:30	https://www.detik.com/su/mbagsel/berita/d-7103200/pekerja-bangunan-di-gandus-tersengat-listrik-tegangan-tinggi	Pekerja Bangunan di Gandus Tersengat Listrik Tegangan Tinggi Perry (55) warga Lubuk Lancang Banyuasin tersengat listrik tegangan tinggi saat sedang mengerjakan bangunan gapura di Gandus, Palembang. Ia langsung dibawa ke rumah sakit Gandus dan kini sudah dirujuk ke RSMH Palembang. Camat Gandus Jufriansyah mengatakan pekerja bangunan tersebut saat kejadian sedang membangun gapura di perumahan Pemkot Gandus Palembang. Seketika korban Perry tersengat listrik bertegangan tinggi.	Kontak dengan Arus Listrik
960	16 Mei 2022 14:12	https://www.detik.com/jabar/berita/d-6079912/ngebut-dan-oleng-detik-detik-elf-maut-tewaskan-7-orang-di-karawang	Ngebut dan Oleng! Detik-detik Elf Maut Tewaskan 7 Orang di Karawang Sebuah rekaman CCTV menampilkan sebuah elf yang memicu terjadinya kecelakaan maut di Jalan Tamelang, Purwasari, Kabupaten Karawang. Elf itu menabrak sejumlah kendaraan dan menyeruduk warga. Insiden tersebut mengakibatkan tujuh orang tewas dan sembilan lainnya terluka. Sekadar diketahui, kecelakaan maut di Purwasari Karawang ini terjadi Minggu (15/5).	Menginjak, Menabrak, atau Tertimpa Benda, kecuali Benda yang Jatuh

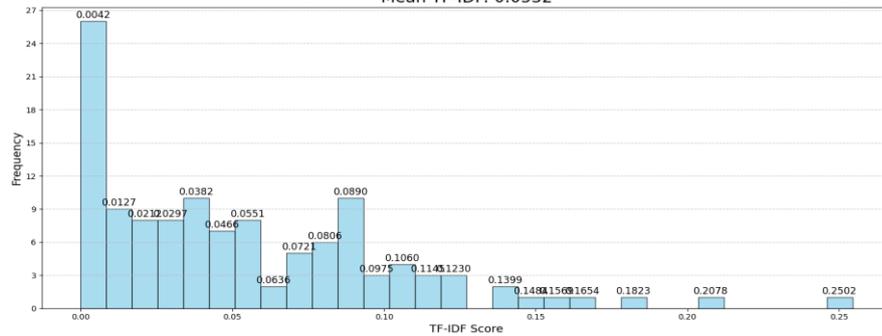
Lampiran II Histogram distribusi fitur kata setiap dokumen



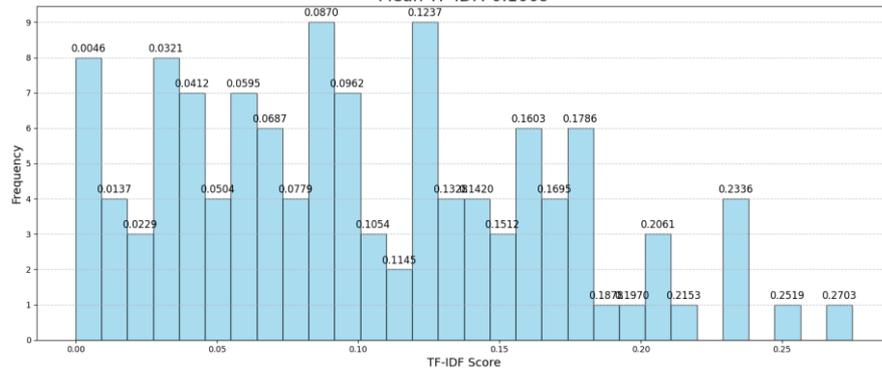
Histogram of TF-IDF Scores for Word: "korban" in Class: Terperangkap di dalam atau diantara benda-benda
Mean TF-IDF: 0.0829



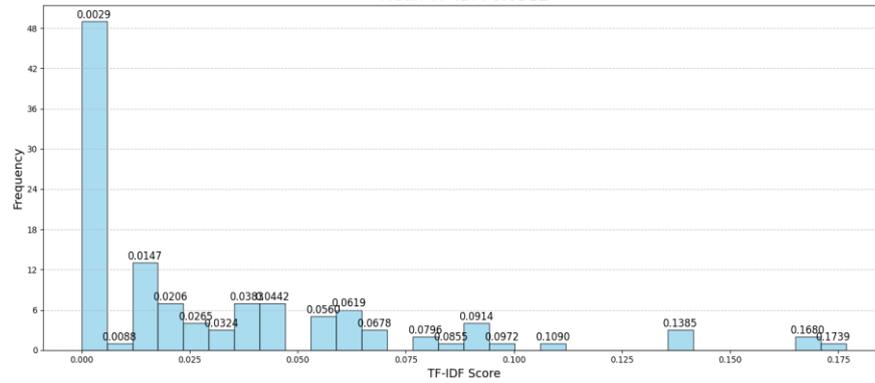
Histogram of TF-IDF Scores for Word: "korban" in Class: Tertimpa Benda-Benda Jatuh
Mean TF-IDF: 0.0532



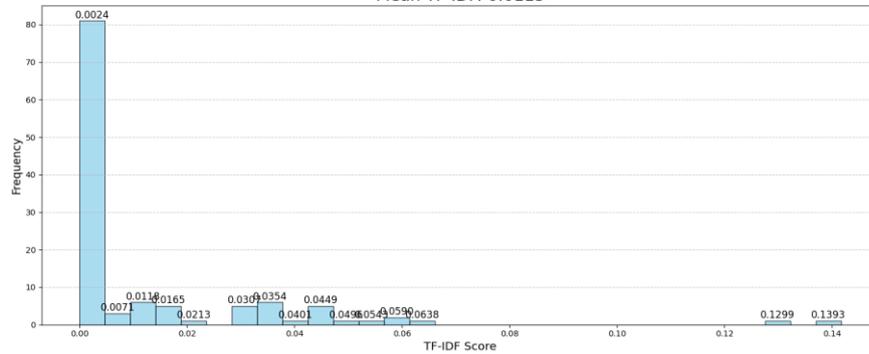
Histogram of TF-IDF Scores for Word: "korban" in Class: Kontak dengan Arus Listrik
Mean TF-IDF: 0.1009



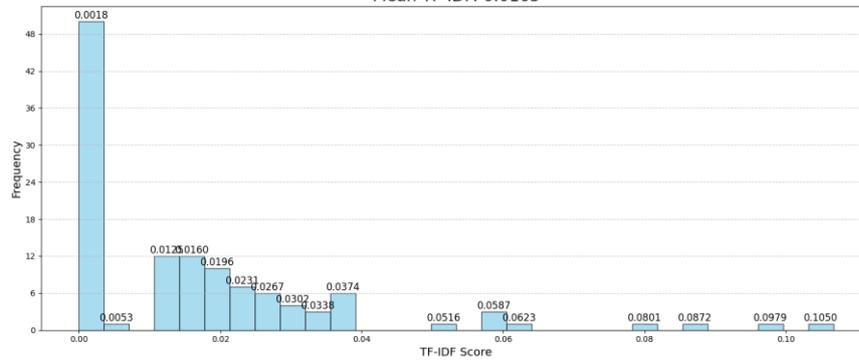
Histogram of TF-IDF Scores for Word: "korban" in Class: Menginjak, Menabrak, atau Tertimpa Benda, kecuali Benda yang Jatuh
 Mean TF-IDF: 0.0312



Histogram of TF-IDF Scores for Word: "korban" in Class: Melakukan Gerakan Berat
 Mean TF-IDF: 0.0115



Histogram of TF-IDF Scores for Word: "korban" in Class: Paparan/Kontak dengan Suhu Ekstrem
 Mean TF-IDF: 0.0165



Histogram of TF-IDF Scores for Word: "korban" in Class: Paparan/Kontak dengan Zat atau Radiasi Berbahaya
 Mean TF-IDF: 0.0475

