

# ANALISIS SENTIMEN PENGURUSAN BENCANA BANJIR MENGUNAKAN TWITTER

Liyanna Sarah Wong binti Abdullah<sup>1</sup>, Wandeep Kaur a/p Ratan Singh<sup>2</sup>

<sup>1,2</sup>*Fakulti Teknologi & Sains Maklumat, Universiti Kebangsaan Malaysia*

## ABSTRAK

Populariti penggunaan media sosial telah meningkat dengan begitu ketara sejak beberapa tahun belakngan ini di mana orang ramai menyatakan pendapat serta berkongsi pengalaman melalui pelbagai pelantar media sosial. Namun, media sosial masih kurang digunakan dalam strategi pengurusan bencana kerana ia tidak dilihat sebagai alat memperoleh maklumat masa nyata semasa berlaku bencana alam. Memanfaatkan maklumat sumber khalayak memberikan potensi besar bagi komuniti setempat mendapatkan maklumat masa nyata untuk respons kecemasan tepat pada masanya. Dalam kertas kerja ini, pendekatan model pembelajaran terselia yang digunakan ialah *Logistic Regression* dan *Naïve Bayes* dan model pembelajaran mendalam yang digunakan ialah *Multilayer Perceptron* untuk pengelasan teks. Teks dikelaskan kepada enam kategori iaitu sumber diperlukan, sumber tersedia, ucapan terima kasih, aduan, amaran dan lain-lain. MNB-TFIDF menunjukkan keputusan yang terbaik menggunakan set data asal dengan mencapai *accuracy* sebanyak 89.53% manakala MLP-TFIDF menunjukkan keputusan yang terbaik menggunakan set data campur dengan mencapai *accuracy* sebanyak 95.49%. Penemuan projek ini mencadangkan bahawa cadangan fitur tambahan meningkatkan prestasi model klasifikasi secara keseluruhannya. Fitur “amaran” diperkenalkan sebagai fitur baharu dalam klasifikasi tweet bencana untuk menambahbaik pengurusan bencana semasa.

**Kata kunci:** Pengurusan bencana banjir, analisis sentimen

## 1. PENGENALAN

Banjir merupakan kejadian bencana alam yang paling kerap berlaku di mana tanah yang kering ditenggelamkan oleh air yang berlebihan (Shreevastav et al. 2022). Hujan yang lebat dan berpanjangan meningkatkan paras air dalam tempoh yang singkat dan menyebabkan limpahan air anak sungai, saluran dan sungai (Siegel 2020). Banjir boleh menyebabkan kerosakan yang meluas termasuk kemusnahan terhadap kemudahan awam yang penting serta kerosakan harta benda peribadi dan kematian (Menne et al. 2013). Oleh itu, pengurusan bencana yang melibatkan organisasi, perancangan dan pelan tindakan untuk persediaan, tindak balas dan pemulihan daripada bencana sangat penting semasa kejadian (Coppola 2006). Orang ramai menggunakan Twitter untuk menciap maklumat semasa tentang keadaan mereka, meminta bantuan dan mendapatkan maklumat berguna mengenai makanan, pusat pemindahan dan pengangkutan (Maharani 2020).

### 1.1 Peranan Media Sosial

Kemunculan media sosial sebagai alat kritikal untuk pengurangan risiko bencana termasuk penyediaan, tindak balas dan pemulihan, adalah signifikan (Ogie et al. 2019). Semasa berlakunya krisis dan malapetaka, pelantar media sosial termasuk Twitter, Facebook, WeChat dan WeiBo sering digunakan oleh komuniti untuk terus berhubung, berkongsi pengalaman dan mendapatkan akses kepada maklumat dan sumber penting untuk membantu tindak balas dan pemulihan bencana. (Jamali et al. 2019). Sebagai contoh, semasa Taufan Meranti mendarat, orang ramai menggunakan media sosial untuk memeriksa keselamatan satu sama lain dan memberi bantuan (Boas et al. 2020). Agensi bantuan yang ingin berjaya bertindak balas terhadap malapetaka mesti mendapatkan maklumat yang tepat tentang kejadian semasa dengan pantas (De Bruijn et al. 2018). Banyak kajian mencadangkan tiga pendekatan untuk menggunakan media sosial selepas bencana alam (Lin et al. 2020):

1. Bersedia menghadapi bencana alam – media sosial boleh membantu individu dalam membuat persediaan yang lebih baik untuk menghadapi malapetaka dan memahami kumpulan mana yang akan membantu kawasan mereka.

2. Bertindak balas semasa dan selepas bencana alam – semasa krisis, media sosial membolehkan pengguna berhubung terus dengan saudara-mara mereka, pemberita dan kumpulan sukarelawan serta penduduk tempatan yang lain, serta memberikan maklumat dengan segera. Selain itu, ia menghadkan penyebaran berita palsu dengan menjadikannya lebih mudah bagi organisasi untuk mengesahkan fakta.
3. Pulih daripada bencana alam – media sosial membantu menyatukan masyarakat bersama untuk membincangkan malapetaka, bertukar maklumat, mengatur aktiviti pemulihan dan mendapatkan maklumat bantuan.

Data media sosial dalam talian (OSM) merujuk kepada cerapan mentah dan maklumat yang diperoleh daripada akaun media sosial seseorang (Behl et al. 2021). Data tidak berstruktur ini terus berkembang disebabkan pertumbuhan eksponen populariti media sosial dalam talian (Jamali et al. 2019). Terdapat keperluan untuk penyelesaian yang menggunakan data OSM untuk menghubungkan usaha bantuan, mangsa dan mereka yang ingin membantu dengan ubat-ubatan, tempat perlindungan dan sumber lain. Ia memudahkan pengumpulan nombor, peratusan dan data daripada ciapan mangsa.

## **1.2 Komponen Pengurusan Bencana Yang Komprehensif**

Pengurusan bencana yang komprehensif adalah berdasarkan empat komponen berbeza iaitu mitigasi, kesediaan, tindak balas dan pemulihan (Coppola 2006). Mitigasi melibatkan mengurangkan kemungkinan atau kesan bencana, atau kedua-duanya. Kesediaan melibatkan melengkapkan orang yang mungkin terjejas akibat bencana atau yang mungkin dapat membantu mereka yang terjejas dengan alatan untuk meningkatkan peluang mereka untuk terus hidup dan meminimumkan kerugian kewangan dan lain-lain. Tindak balas melibatkan mengambil tindakan untuk mengurangkan atau menghapuskan kesan bencana yang telah berlaku atau sedang berlaku, untuk mengelakkan penderitaan selanjutnya, kerugian kewangan, atau gabungan kedua-duanya. Pemulihan melibatkan mengembalikan kehidupan mangsa kepada keadaan normal berikutan kesan akibat bencana. Walau bagaimanapun, kerjasama pelbagai agensi menghadapi banyak cabaran dalam pengurusan bencana

termasuk maklumat, komunikasi, alam sekitar, sosial, politik, antara organisasi dan cabaran intra-organisasi (Abdeen et al. 2021).

### **1.3 Pernyataan Masalah**

Apabila sesuatu krisis mula timbul, orang ramai menggunakan pelantar media sosial seperti Facebook dan Twitter untuk menyampaikan situasi semasa, mencari maklumat berguna dan meminta bantuan (Jamali et al. 2019). Maklumat peringkat bawah yang dikumpul daripada pelantar media sosial diperoleh berdasarkan pengalaman langsung (Maharani 2020), justeru, data Twitter mempunyai peluang yang besar dalam membantu organisasi, mangsa dan orang sekeliling mereka bertindak dalam menghadapi kecemasan. Walau bagaimanapun, kajian sedia ada tidak memberikan gambaran yang jelas tentang bagaimana orang ramai mendapat manfaat daripada komunikasi di media sosial dan bagaimana aktiviti ini boleh membantu dalam proses pemulihan berikutan bencana alam (Bukar et al. 2022). Oleh itu, terdapat keperluan untuk mengkaji pemboleh ubah yang boleh membantu dalam pengurusan bencana semasa.

#### **1.3.1 Mengklasifikasi Ciapan Berkaitan Bencana Menggunakan Pembelajaran Mesin Untuk Respons Kecemasan**

Projek ini bertujuan untuk mengklasifikasikan ciapan berkaitan bencana sebagai usaha awal untuk respons kecemasan dan meningkatkan kesedaran situasi. Kajian lepas telah menggunakan model pembelajaran mesin seperti *Support Vector Machine* (Delimayanti et al. 2020), (Firlia et al. 2021), *Naïve Bayes* (Ridwan et al. 2022), *Multilayer Perceptron* (Behl et al. 2021), BERT (Maharani 2020) dan LDA (Rahmadan et al. 2020). Kelemahan kaedah ini ialah ketepatan pengelasan sangat bergantung pada data berlabel (Ridwan et al. 2022) dan sangat dipengaruhi oleh teknik pengekstrakan fitur (Ahuja et al. 2019). Oleh itu, dalam projek ini, penambahbaikan kaedah pembelajaran mesin

akan dicadangkan untuk mengklasifikasikan ciapan berkaitan bencana dan memfokuskan pada analisis fitur.

### **1.3.2 Mengenal Pasti Fitur Untuk Respons Kecemasan**

Ketika berlakunya kecemasan dan bencana, keperluan asas, rawatan perubatan, menghubungi keluarga dan rakan serta mencari maklumat berkaitan adalah penting (Tyshchuk & Wallace 2018). Walau bagaimanapun, penyelidikan semasa sedia ada mengenai analisis sentimen berkaitan bencana tidak termasuk fitur yang mencukupi yang digunakan untuk membina model klasifikasi. Sebagai contoh, mengkategorikan ciapan kepada sumber diperlukan, ketersediaan sumber dan lain-lain (Behl et al. 2021) serta mengklasifikasikan ciapan mengikut menyatakan rasa terima kasih, kelegaan, aduan, memerlukan bantuan dan lain-lain (Nair et al. 2017).

Berdasarkan kajian susastera lepas, terdapat kekurangan fitur untuk mengenal pasti amaran bencana seperti banjir. Oleh itu, projek ini memfokuskan untuk mengenal pasti fitur tambahan yang boleh digabungkan untuk meningkatkan prestasi model klasifikasi yang dicadangkan.

### **1.4 Sorotan Susastera**

Tinjauan susastera yang dilakukan memfokuskan kepada fitur serta model yang digunakan pada kajian sedia ada. Perbandingan dilakukan bagi mengenal pasti jurang pada kajian sedia ada untuk ditambah baik pada projek ini. Jadual 1.1 dan Jadual 1.2 di bawah menunjukkan tinjauan susastera.

| No | Rujukan            | Set data   | Sumbangan   | Dapatan Kajian  |
|----|--------------------|--|---|---|
| 1  | Behl et al. 2021   | Gempa bumi Nepal 2015 (51846 ciapan), gempa bumi Italy 2016 (70897 ciapan), data asal Covid-19 (2274 ciapan) | <ol style="list-style-type: none"> <li>1. Set data gempa bumi Nepal dan set data gempa bumi Itali digunakan untuk melatih model, set data asal COVID-19 Twitter digunakan untuk menguji kebolegunaan semula model terlatih.</li> <li>2. Menggunakan MLP, CNN dan <i>Logistic Regression</i> untuk pengelasan teks.</li> <li>3. Mengklasifikasikan teks kepada 3 kategori iaitu sumber diperlukan, sumber tersedia dan lain-lain.</li> </ol> | MLP dengan <i>pre-trained Word2Vec embedding</i> mencapai <i>accuracy</i> tertinggi 83% pada data COVID-19.   |
| 2  | Nair et al. 2017   | #chennaiflood 2015 (1125 ciapan)   | <ol style="list-style-type: none"> <li>1. Menggunakan <i>Decision trees</i>, <i>Random Forests</i> dan <i>Naïve Bayes</i> untuk pengelasan teks.</li> <li>2. Mengklasifikasikan teks kepada 5 kategori iaitu ucapan terima kasih, sumber tersedia, aduan, sumber diperlukan dan lain-lain.</li> </ol>   | <i>Random Forest</i> mencapai ketepatan, dapatan semula dan skor-F tertinggi iaitu 99.7%.   |
| 3  | Ridwan et al. 2022 | ciapan "prayforkalsel"   | Menggunakan kaedah pengelas <i>Naïve Bayes</i> dengan fitur <i>N-Gram</i> .   | Tokenisasi menggunakan <i>bigram</i> menghasilkan <i>accuracy</i> yang lebih tinggi (71.05%) berbanding <i>unigram</i> (63.16%) dan <i>trigram</i> (52.63%)   |
| 4  | Ahuja et al. 2019  | 4242 set data Kekuatan Sentimen Twitter  | <ol style="list-style-type: none"> <li>1. Mengekstrak fitur menggunakan teknik <i>N-grams</i> dan TF-IDF. \</li> <li>2. Menggunakan <i>Random Forest</i>, <i>Decision Tree</i>, <i>Naive Bayes</i>, <i>Support Vector Machine</i>, <i>Logistic Regression</i>, dan <i>K-Nearest Neighbour</i> untuk pengelasan teks.</li> </ol>   | <ol style="list-style-type: none"> <li>1. Fitur TF-IDF memberikan hasil yang lebih baik (3-4%) berbanding fitur <i>N-Gram</i>.</li> <li>2. <i>Logistic regression</i> dengan TF-IDF mencapai <i>accuracy</i> dan <i>precision</i> tertinggi (57%), <i>recall</i> dan skor-F tertinggi (50%).</li> </ol> |

Jadual 1.1: Tinjauan Susastera

| No | Rujukan                 | Set data  | Sumbangan   | Dapatan Kajian   |
|----|-------------------------|---|---|--|
| 5  | Rahmadan et al. 2020    | #banjirjakarta, #banjir, 'banjir jakarta', 'bencana banjir jakarta', 64754 ciapan | <ol style="list-style-type: none"> <li>1. Pemodelan topik menggunakan LDA.</li> <li>2. Analisis sentimen berasaskan leksikon menggunakan kamus InSet Lexicon.</li> </ol>  | <ol style="list-style-type: none"> <li>1. Data ciapan dikelompokkan kepada 9 topik dan setiap topik diwakili oleh 10 perkataan paling dominan.</li> <li>2. Kebanyakan pendapat yang disampaikan melalui media sosial Twitter menunjukkan sentimen negatif iaitu sebanyak 79%.</li> </ol> |
| 6  | Maharani 2020           | 10806 ciapan  | Menggunakan <i>BERT base Multilingual Cased model</i> dengan parameter pra-latihan untuk penalaan halus.  | Model ini mencapai <i>accuracy</i> 90% pada set data latihan dan 79% pada set data ujian.  |
| 7  | Delimayanti et al. 2020 | pan   | <ol style="list-style-type: none"> <li>1. 2 jenis kamus kata henti (stopword dictionaries) digunakan dalam penyingkiran kata henti.</li> <li>2. <i>Support Vector Machine</i> dengan menimbang perkataan menggunakan TF-IDF digunakan untuk mengklasifikasikan mesej bencana banjir.</li> </ol> | <i>Accuracy</i> tertinggi yang diperoleh menggunakan kamus kata henti 1 ialah 75.13% manakala <i>accuracy</i> tertinggi yang diperoleh menggunakan kamus kata henti 2 ialah 77.87%.  |

Jadual 1.2: Tinjauan Susastera (sambungan daripada jadual sebelumnya)

Behl et al. (2021) menggunakan set data gempa bumi Nepal dan set data gempa bumi Itali untuk melatih model, set data asal COVID-19 Twitter untuk menguji kebolehgunaan semula model terlatih. MLP, CNN dan *Logistic Regression* digunakan untuk pengelasan teks. MLP dengan *pre-trained Word2Vec embedding* mencapai *accuracy* tertinggi 83% pada data COVID-19. Nair et al. (2017) menggunakan *Decision trees*, *Random Forests* dan *Naïve Bayes* untuk pengelasan teks serta mengklasifikasikan teks kepada 5 kategori iaitu ucapan terima kasih, sumber tersedia, aduan, sumber diperlukan dan lain-lain. *Random Forest* mencapai ketepatan, dapatan semula dan skor-F tertinggi iaitu 99.7%.

Ridwan et al. (2022) menggunakan kaedah pengelas *Naïve Bayes* dengan fitur *N-Gram*. Tokenisasi menggunakan *bigram* menghasilkan *accuracy* yang lebih tinggi (71.05%) berbanding *unigram* (63.16%) dan *trigram* (52.63%). Ahuja et al. (2019) mengekstrak fitur menggunakan teknik *N-grams* dan TF-IDF, *Random Forest*, *Decision Tree*, *Naive Bayes*, *Support Vector Machine*, *Logistic Regression*, dan *K-Nearest Neighbour* untuk pengelasan teks. Fitur TF-IDF memberikan hasil yang lebih baik (3-4%) berbanding fitur *N-Gram*. *Logistic regression* dengan TF-IDF mencapai *accuracy* dan *precision* tertinggi (57%), *recall* dan skor-F tertinggi (50%).

Rahmadan et al. (2020) menggunakan LDA untuk pemodelan topik. Analisis sentimen berasaskan leksikon menggunakan kamus InSet Lexicon. Data ciapan dikelompokkan kepada 9 topik dan setiap topik diwakili oleh 10 perkataan paling dominan. Kebanyakan pendapat yang disampaikan melalui media sosial Twitter menunjukkan sentimen negatif iaitu sebanyak 79%. Maharani (2020) menggunakan *BERT base Multilingual Cased model* dengan parameter pra-latihan untuk penalaan halus. Model ini mencapai *accuracy* 90% pada set data latihan dan 79% pada set data ujian. Delimayanti et al. (2020) menggunakan 2 jenis kamus kata henti (stopword dictionaries) untuk penyingkiran kata henti. *Support Vector Machine* dengan menimbang perkataan menggunakan TF-IDF digunakan untuk mengklasifikasikan mesej bencana banjir. *Accuracy* tertinggi yang diperoleh menggunakan kamus kata henti 1 ialah 75.13% manakala *accuracy* tertinggi yang diperoleh menggunakan kamus kata henti 2 ialah 77.87%.

## 1.5 Cadangan Penyelesaian

Kajian ini akan dijalankan dengan menggabungkan fitur tambahan sebagai input ke dalam algoritma pembelajaran mesin. Ia dilakukan dengan mengasingkan data ciapan kepada enam kategori iaitu sumber diperlukan, sumber tersedia, ucapan terima kasih, aduan, amaran dan lain-lain. Analisis



perbandingan ketepatan menggunakan metrik penilaian akan dilaksanakan pada algoritma pembelajaran mesin.

## 1.6 Objektif Kajian

Objektif kajian ini adalah seperti berikut:

1. Untuk mengenal pasti fitur-fitur relevan yang diekstrak daripada Twitter berkaitan respons kecemasan semasa pengurusan bencana banjir.
2. Untuk mencadangkan algoritma pembelajaran mesin yang mampu menganalisis ciapan untuk pengklasifikasian sentimen semasa pengurusan bencana banjir.
3. Untuk menilai algoritma pengelasan yang dicadang menggunakan metrik penilaian

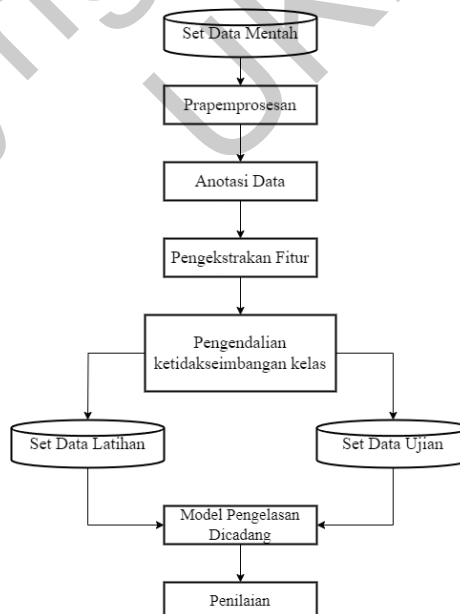
## 1.7 Skop dan Kekangan

Projek ini memberi tumpuan kepada analisis fitur untuk menambah baik langkah-langkah mitigasi pengurusan banjir semasa melalui klasifikasi ciapan berkaitan bencana menggunakan algoritma pembelajaran mesin. Ini dilakukan dengan memanfaatkan data sumber khalayak sukarelawan yang tersedia di Twitter mengenai banjir. Ciapan dalam bahasa Inggeris mengenai banjir terbaharu di Malaysia dengan tanda pagar seperti #DaruratBanjir, #banjirdarurat, #banjir, #RakyatJagaRakyat dan #kitajagakita antara 18 Disember 2021 hingga 21 Disember 2021 dan banjir di Pakistan dengan tanda pagar seperti #FloodsInPakistan, #Floods20Pakistan #pakistanflood2022, #pakistanfloods2022 sejak 14 Jun 2022 dipertimbangkan. Pendekatan pembelajaran mesin disesuaikan untuk mengklasifikasikan ciapan untuk analisis sentimen mengenai pengurusan bencana banjir. Terdapat banyak kekangan dalam kajian ini. Memandangkan kajian ini tertumpu terutamanya pada ciapan di Twitter, data yang disiarkan pada pelantar media sosial yang lain tidak dapat dianalisis. Mungkin terdapat siaran yang lebih bernas pada pelantar media sosial lain kerana demografi Twitter tidak sama dengan pelantar media sosial lain. Kumpulan umur terbesar untuk Twitter berkisar antara 18-29 manakala Facebook dan Instagram berkisar antara 25-34 sebagai kumpulan umur terbesar, TikTok berjulat antara 10-19

sebagai kumpulan umur terbesar (Barnhart 2021). Selain itu, hanya ciapan yang diakses terbuka boleh diekstrak, ini bermakna kami tidak akan dapat menganalisis ciapan yang diciap oleh akaun peribadi. Hanya ciapan bahasa Inggeris diekstrak yang juga menjadikan kekangan dalam kajian ini.

Dalam laporan ini mengandungi 4 bahagian di mana Bahagian 1. Pengenalan menerangkan latar belakang projek, membincangkan pernyataan masalah, melakukan sorotan susastera. Bahagian ini juga menyatakan cadangan penyelesaian, objektif projek serta skop dan kekangan projek. Bahagian 2. Metodologi membincangkan algoritma klasifikasi sentimen. Bahagian 3. Keputusan dan Perbincangan membincangkan keputusan klasifikasi sentimen menggunakan set data asal, set data campur dan membuat perbandingan keputusan projek ini dengan kajian lepas. Bahagian 4. Kesimpulan membuat rumusan projek ini, serta menyatakan kekangan dan cadangan penambahbaikan pada masa akan datang.

## 2. METODOLOGI CADANGAN ALGORITMA KLASIFIKASI SENTIMEN



Rajah 2.1: Metodologi Cadangan Algoritma Klasifikasi Analisis Sentimen Pengurusan Bencana Banjir Menggunakan Twitter

Rajah 2.1 di atas menunjukkan metodologi cadangan algoritma klasifikasi analisis sentimen pengurusan bencana banjir menggunakan Twitter.

## 2.1 Pengumpulan Data

Semasa fasa pengumpulan data, ciapan yang berkaitan dengan banjir diekstrak menggunakan API Twint. Twint ialah alat mengikis (scraping tool) Twitter berasaskan Python yang berkuasa yang mengikis (scrapes) Tweet daripada akaun Twitter tanpa menggunakan API Twitter. Data yang diekstrak disimpan ke dalam fail CSV.

## 2.2 Prapemprosesan Data

Set data menjalani teknik prapemprosesan untuk meminimumkan hingar data (data noise). Data yang dikumpul adalah berkaitan banjir antara 1 Disember 2022 hingga 31 Disember 2022 menggunakan kata kunci *flood*. Pada mulanya, data yang diperolehi mempunyai 50001 rekod dengan banyak lajur. Hanya lajur tweet yang tinggal manakala lajur lain telah dipadamkan kerana ia tidak mempunyai nilai tambahan untuk projek.

Projek ini memberi tumpuan kepada kandungan bahasa Inggeris, kandungan bukan bahasa Inggeris ditapis. Tweet berulang telah dialih keluar. URL, sebutan (@sebutan), tanda pagar (#), aksara seperti kurungan, 'RT' dan aksara bukan ASCII lain seperti &, elips dan aksara Unicode yang sepadan dengan emoji dialih keluar menggunakan ungkapan biasa (regular expression) dan perpustakaan *cleantext*<sup>1</sup>. Tweet dinormalkan dengan mengalih keluar ruang putih tambahan dan ditukarkan kepada huruf kecil.

Kata henti seperti a, an, diekstrak menggunakan perpustakaan NLTK (Ramachandran & Parvathi 2019). Tokenisasi dilakukan untuk memisahkan teks kepada token menggunakan perpustakaan NLTK. Tweet yang mempunyai kurang daripada 4 token dialih keluar kerana ia tidak

---

<sup>1</sup> <https://pypi.org/project/cleantext/>

menyampaikan apa-apa makna dan sukar untuk dianotasi. Lemmatisasi dilakukan menggunakan WordNetLemmatizer untuk menormalkan perkataan kepada bentuk asalnya (Qorib et al. 2023).

### 2.3 Penglabelan Data

Tiga pencatat yang terdiri daripada penyelidik khusus dalam kejuruteraan awam, pakar dalam linguistik dan orang awam yang terlibat dalam penyediaan data standard emas yang mengandungi tweet yang berkaitan dengan sumber diperlukan, sumber tersedia, ucapan terima kasih, aduan, amaran atau lain-lain. Label “amaran” dikenal pasti sebagai ciri baharu dalam projek ini.

Setiap pencatat diminta untuk melabelkan tweet secara bebas tanpa berbincang dengan pencatat lain. Walau bagaimanapun, dalam kes ketidaksetujuan pencatat, sebagai contoh, pencatat pertama menyatakan “amaran”, pencatat kedua menyatakan “aduan” dan pencatat ketiga mengatakan “lain-lain”, ujian kebolehpercayaan antara penilai alfa Krippendorff digunakan untuk mengukur tahap persetujuan dalam kalangan pencatat (Krippendorff 2018). Jika pekali alfa ialah  $> 0.8$ , ia menunjukkan kebolehpercayaan antara penilai yang kuat. Jika pekali alfa ialah 0.67 hingga 0.8, ia menunjukkan kebolehpercayaan yang rendah. Pekali alfa kurang daripada 0.67 menunjukkan kebolehpercayaan antara penilai yang sangat rendah. Apabila pekali alfa menunjukkan kebolehpercayaan antara penilai yang sangat rendah, perbincangan dengan pencatat telah dianjurkan untuk menyemak dan memahami alasan mereka untuk kategori yang dipilih. Kategori anotasi yang paling sesuai ditentukan sepanjang perbincangan. Dalam kes ini, pekali alfa yang diperolehi ialah 0.92 iaitu melebihi 0.8 dan dianggap baik.

Dalam situasi kedua di mana pencatat pertama menyatakan “amaran”, pencatat kedua menyatakan “amaran” manakala pencatat ketiga menyatakan “lain-lain”, kategori catatan akhir akan menjadi “amaran” kerana ia dipilih oleh majoriti pencatat.

Pencatat-pencatat fasih berbahasa Inggeris dan pengguna Twitter yang kerap, tetapi tiada seorang pun daripada mereka yang terlibat dengan projek ini. 1200 data yang dipilih secara rawak

telah dihantar untuk anotasi dalam kalangan pencatat pada 23 Mac 2023. Walau bagaimanapun, hanya sejumlah 627 data telah diterima daripada pakar dan orang awam pada 8 Mei 2023. Projek ini menggunakan set data asal dan set data campur untuk penilaian model pengelasan. Set data campur ialah gabungan set data asal dan set data penanda aras Banjir Sri Lanka 2017 yang berkaitan dengan bantuan kemanusiaan semasa bencana alam (Alam et al. 2021).

## 2.4 Pengekstrakan Ciri

Pengekstrakan ciri mengubah data menjadi ciri berangka yang boleh difahami oleh model pembelajaran mesin. Pembinaan CrisisNLP Word2Vec<sup>2</sup> khusus bencana digunakan kerana perkataan diwakili sebagai vektor dimensi rendah yang mewakili kedua-dua semantik dan sintaks perkataan (Behl et al. 2021). Pembinaan Word2Vec telah dilatih menggunakan 52 juta tweet dianotasi manusia berkaitan dengan bencana yang berlaku antara 2013 dan 2015. Setiap perkataan diwakili oleh 300 dimensi menggunakan seni bina beg perkataan berterusan dalam model Word2Vec yang digunakan di sini. *Term Frequency-Inverse Document Frequency* (TFIDF) juga digunakan untuk memetakan perkataan kepada vektor nombor nyata bagi proses pengekstrakan ciri.

## 2.5 Pengendalian Ketidakeimbangan Kelas

Teknik pengendalian ketidakseimbangan kelas, *Synthetic Minority Over-Sampling Technique* (SMOTE) digunakan untuk menangani masalah ketidakseimbangan kelas. Jadual 2.1 dan Jadual 2.2 berikut menunjukkan taburan kelas sebelum SMOTE dan selepas SMOTE bagi kedua-dua set data asal dan set data campur.

---

<sup>2</sup> <https://crisisnlp.qcri.org/lrec2016/lrec2016.html>

Jadual 2.1: Perbandingan Sebelum dan Selepas SMOTE Set Data Asal

| Kategori            | Sebelum SMOTE | Selepas SMOTE |
|---------------------|---------------|---------------|
| Sumber diperlukan   | 85            | 172           |
| Sumber tersedia     | 172           | 172           |
| Ucapan terima kasih | 72            | 172           |
| Aduan               | 68            | 172           |
| Amaran              | 158           | 172           |
| Lain-lain           | 72            | 172           |

Jadual 2.2: Perbandingan Sebelum dan Selepas SMOTE Set Data Campur

| Kategori            | Sebelum SMOTE | Selepas SMOTE |
|---------------------|---------------|---------------|
| Sumber diperlukan   | 115           | 473           |
| Sumber tersedia     | 473           | 473           |
| Ucapan terima kasih | 109           | 473           |
| Aduan               | 68            | 473           |
| Amaran              | 224           | 473           |
| Lain-lain           | 72            | 473           |

## 2.6 Model Pengelasan

Model pengelasan yang digunakan dalam projek ini ialah *Logistic Regression*, *Multinomial Naïve Bayes* bagi pembelajaran mesin diselia dan *Multilayer Perceptron* bagi pembelajaran mendalam. Setiap model digunakan menjalani penalaan halus hiperparameter untuk menghasilkan keputusan pengelasan yang terbaik.

## 2.7 Penilaian

Data dibahagikan kepada 75% untuk set data latihan dan 25% untuk set data ujian. Metrik penilaian yang digunakan untuk menilai prestasi model ialah kejituan, *logarithmic loss*, ketepatan, dapatan semula, skor-F, *confusion matrix* dan AUC/ROC.

### 3. KEPUTUSAN DAN PERBINCANGAN

#### 3.1 Keputusan Set Data Asal

Jadual 3.1: Ringkasan Keputusan Set Data Asal

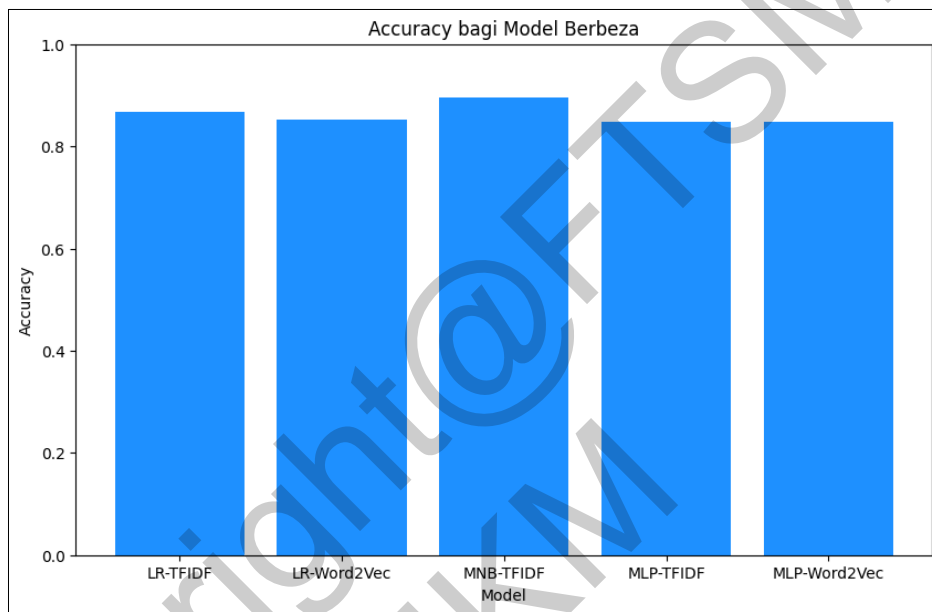
| Model        | Metrik          |               |                  |               |                 |
|--------------|-----------------|---------------|------------------|---------------|-----------------|
|              | <i>Accuracy</i> | <i>Loss</i>   | <i>Precision</i> | <i>Recall</i> | <i>F1-Score</i> |
| LR-TFIDF     | 0.8682          | 0.4301        | 0.8720           | 0.8698        | 0.8682          |
| LR-Word2Vec  | 0.8527          | 0.5381        | 0.8507           | 0.8564        | 0.8572          |
| MNB-TFIDF    | <b>0.8953</b>   | <b>0.4060</b> | <b>0.8967</b>    | <b>0.8939</b> | <b>0.8953</b>   |
| MLP-TFIDF    | 0.8488          | 0.4883        | 0.8590           | 0.8564        | 0.8566          |
| MLP-Word2Vec | 0.8488          | 0.6059        | 0.8469           | 0.8512        | 0.8488          |

Daripada Jadual 3.1 di atas, MNB-TFIDF mencapai *accuracy* tertinggi berbanding model lain iaitu 89.53% manakala MLP-TFIDF dan MLP-Word2Vec mencapai *accuracy* terendah berbanding model lain iaitu 84.88% bagi kedua-dua model pembelajaran mendalam. MNB-TFIDF mempunyai nilai *loss* yang paling rendah berbanding model lain iaitu 40.06% manakala MLP-Word2Vec mempunyai nilai *loss* yang paling tinggi berbanding model lain iaitu 60.59%. Nilai *loss* tertinggi dan terendah mempunyai julat yang agak ketara berbanding metrik penilaian yang lain iaitu sebanyak 19.99%.

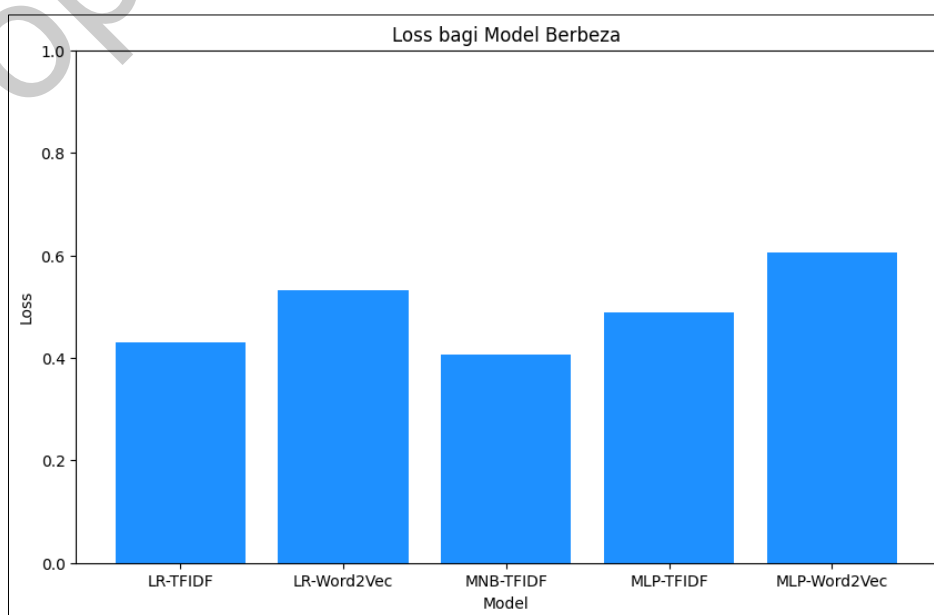
Model klasifikasi yang menggunakan Word2Vec sebagai pengekstrakan fitur diperhatikan mempunyai nilai *loss* yang agak tinggi iaitu melebihi 50%. MNB-TFIDF mempunyai nilai tertinggi bagi *precision*, *recall*, *F1-score* iaitu 89.67%, 89.39% dan 89.53%. Nilai tertinggi *precision* dan *recall* menghasilkan nilai tertinggi *F1-score*. MLP-Word2Vec mempunyai nilai terendah bagi *precision*, *recall*, *F1-score* iaitu 84.69%, 85.12% dan 84.88%

Pendekatan pembelajaran diselia menunjukkan prestasi yang lebih baik berbanding menggunakan pendekatan pembelajaran mendalam dalam menggunakan set data asal. Hal ini kerana set data asal merupakan set data yang kecil. Seni bina yang kompleks dalam pendekatan mendalam kurang sesuai digunakan pada set data kecil. Penggunaan Word2Vec khusus bencana sebagai kaedah pengekstrakan fitur menunjukkan nilai *loss* yang agak tinggi. Model Word2Vec yang digunakan dalam

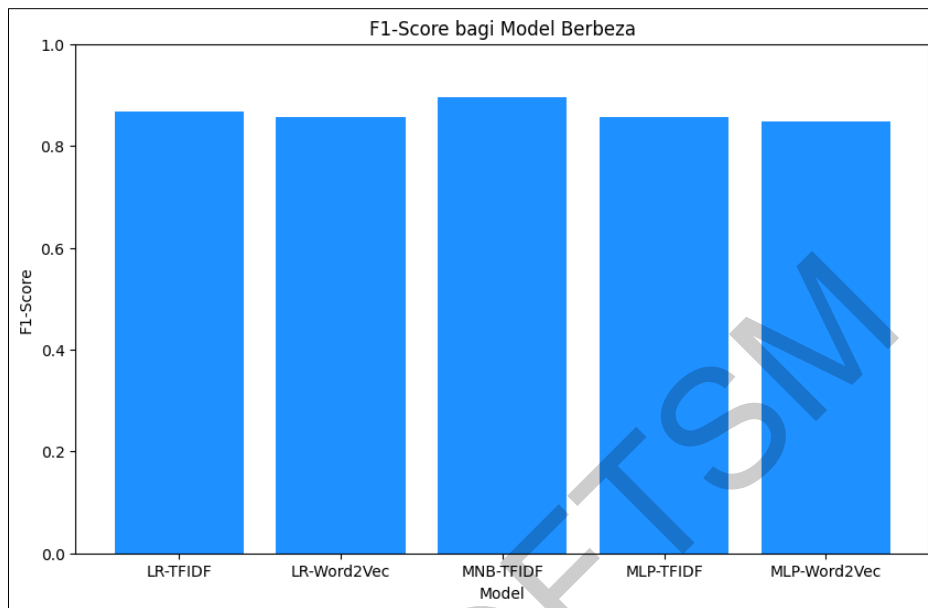
ujian projek ini dilatih pada set data dengan perbendaharaan kata yang berbeza berbanding set data yang digunakan dalam projek ini. Hal ini berkemungkinan terdapat perkataan di luar perbendaharaan kata (*out-of-vocabulary*) (OOV) dalam dataset projek ini yang tidak mempunyai pembenaman kata. Model tidak dapat mengendalikan perkataan OOV mengakibatkan prestasi menurun. Perbandingan *accuracy*, *loss*, *F1-Score* bagi model berbeza menggunakan set data asal ditunjukkan dalam Rajah 3.1, Rajah 3.2, Rajah 3.3 di bawah.



Rajah 3.1: Perbandingan *Accuracy* bagi Model Berbeza dalam Set Data Asal





Rajah 3.2: Perbandingan *Loss* bagi Model Berbeza dalam Set Data AsalRajah 3.3: Perbandingan *F1-score* bagi Model Berbeza dalam Set Data Asal

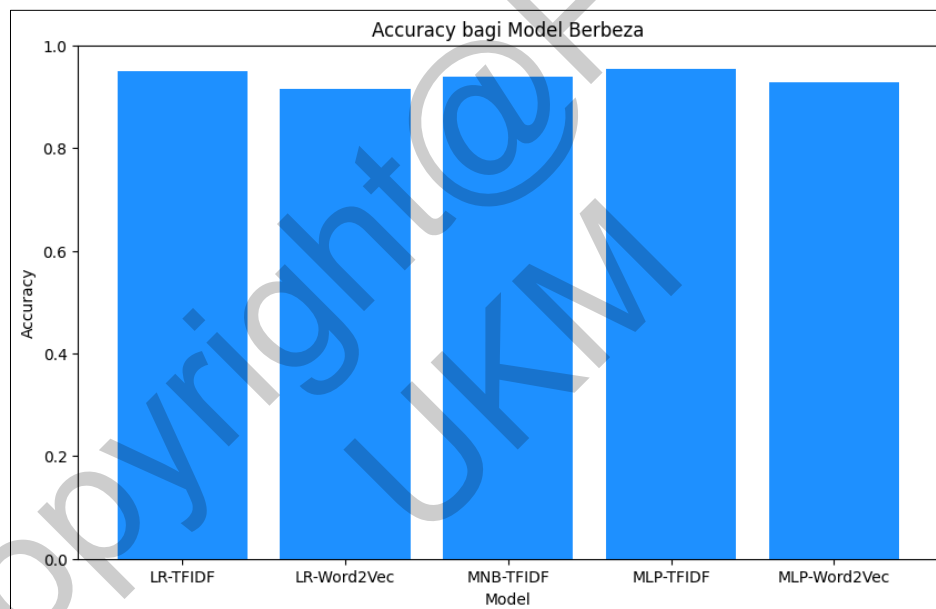
### 3.2 Keputusan Set Data Campur

Jadual 3.2: Ringkasan Keputusan Set Data Campur

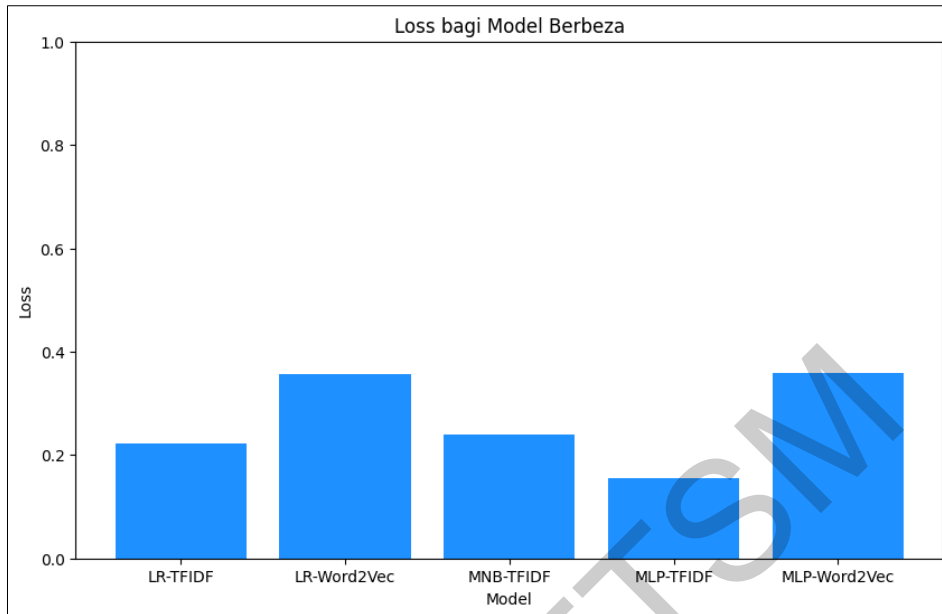
| Model        | Metrik          |               |                  |               |                 |
|--------------|-----------------|---------------|------------------|---------------|-----------------|
|              | <i>Accuracy</i> | <i>Loss</i>   | <i>Precision</i> | <i>Recall</i> | <i>F1-Score</i> |
| LR-TFIDF     | 0.9493          | 0.2222        | 0.9493           | 0.9502        | 0.9493          |
| LR-Word2Vec  | 0.9141          | 0.3575        | 0.9129           | 0.9157        | 0.9141          |
| MNB-TFIDF    | 0.9380          | 0.2386        | 0.9385           | 0.9389        | 0.9380          |
| MLP-TFIDF    | <b>0.9549</b>   | <b>0.1546</b> | <b>0.9558</b>    | <b>0.9572</b> | <b>0.9549</b>   |
| MLP-Word2Vec | 0.9282          | 0.3582        | 0.9303           | 0.9315        | 0.9282          |

Daripada Jadual 3.2 di atas, MLP-TFIDF mencapai *accuracy* tertinggi berbanding model lain iaitu 95.49% manakala LR-Word2Vec mencapai *accuracy* terendah berbanding model lain iaitu 91.41%. MLP-TFIDF mempunyai nilai *loss* yang paling rendah berbanding model lain iaitu 15.46% manakala MLP-Word2Vec mempunyai nilai *loss* yang paling tinggi berbanding model lain iaitu 35.82%. Nilai *loss* tertinggi dan terendah mempunyai julat yang agak ketara berbanding metrik penilaian yang lain iaitu sebanyak 20.36%.

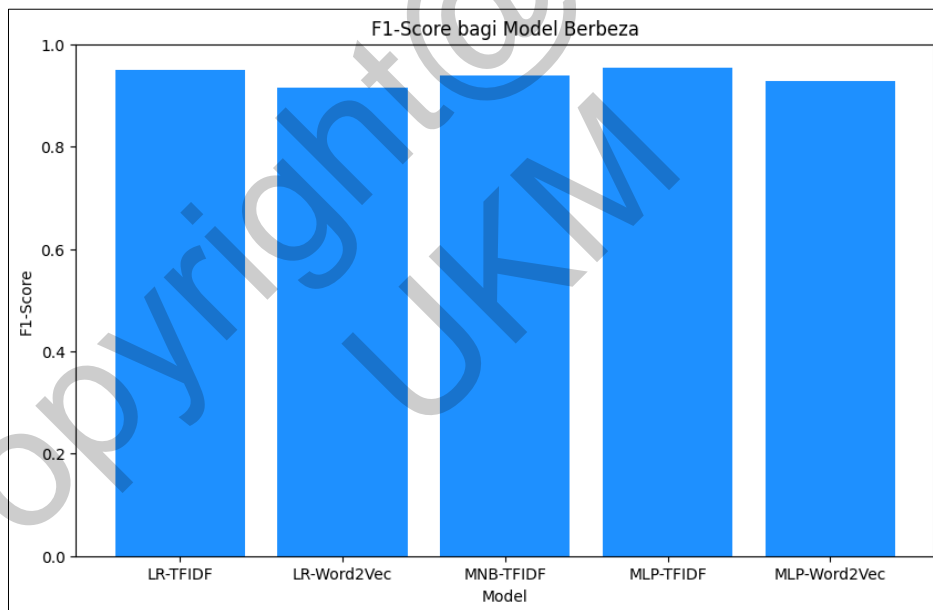
MLP-TFIDF mempunyai nilai tertinggi bagi *precision*, *recall*, *F1-score* iaitu 95.58%, 95.72% dan 95.49%. Nilai tertinggi *precision* dan *recall* menghasilkan nilai tertinggi *F1-score*. LR-Word2Vec mempunyai nilai terendah bagi *precision*, *recall*, *F1-score* iaitu 91.29%, 91.57% dan 91.41%. Kekuatan model MLP-TFIDF terletak pada keupayaannya untuk menangkap corak kompleks dalam data melalui seni bina mendalam dan fungsi pengaktifan bukan linear. Keupayaan model untuk mengendalikan data yang lebih meluas dan pelbagai menghasilkan prestasi yang lebih baik dalam menggunakan set data campur. Jadi, pendekatan mendalam lebih sesuai dalam menggunakan set data yang bersaiz lebih besar. Perbandingan *accuracy*, *loss*, *F1-Score* bagi model berbeza menggunakan set data campur ditunjukkan dalam Rajah 3.4, Rajah 3.5, Rajah 3.6 di bawah.



Rajah 3.4: Perbandingan *Accuracy* bagi Model Berbeza dalam Set Data Campur



Rajah 3.5: Perbandingan *Loss* bagi Model Berbeza dalam Set Data Campur



Rajah 3.6: Perbandingan *F1-score* bagi Model Berbeza dalam Set Data Campur

### 3.3 Perbandingan Keputusan dengan Kajian Lepas

Jadual 3.3: Perbandingan *Accuracy* Model dengan Kajian Lepas

|              | Behl et al. (2021) | Set Data Asal | Set Data Campur |
|--------------|--------------------|---------------|-----------------|
| LR-TFIDF     | 0.84               | 0.87          | 0.95            |
| MLP-TFIDF    | 0.86               | 0.85          | 0.95            |
| MLP-Word2Vec | 0.82               | 0.86          | 0.93            |

Daripada Jadual 3.3 di atas, secara keseluruhannya *accuracy* dalam projek ini menunjukkan peningkatan berbanding dengan kajian lepas. LR-TFIDF menunjukkan peningkatan sebanyak 3% daripada 84% *accuracy* dalam kajian lepas kepada 87% dalam projek ini menggunakan set data asal. Namun, MLP-TFIDF menunjukkan sedikit penurunan iaitu sebanyak 1% daripada 86% dalam kajian lepas kepada 85% dalam projek ini menggunakan set data asal. MLP-Word2Vec menunjukkan peningkatan sebanyak 4% daripada 82% *accuracy* dalam kajian lepas kepada 86% dalam projek ini menggunakan set data asal. LR-TFIDF dan MLP-TFIDF menggunakan set data campur menunjukkan *accuracy* tertinggi iaitu sebanyak 95%.

Dalam kajian lepas, klasifikasi teks hanya menggunakan tiga fitur iaitu sumber diperlukan, sumber tersedia dan lain-lain manakala dalam projek ini, klasifikasi teks menggunakan 6 fitur iaitu, sumber diperlukan, sumber tersedia, ucapan terima kasih, aduan, amaran dan lain-lain. Fitur “amaran” diperkenalkan sebagai fitur baharu dalam klasifikasi tweet bencana. Cadangan fitur tambahan menyumbang kepada peningkatan sebanyak 3% daripada *accuracy* kajian lepas dalam menggunakan LR-TFIDF serta peningkatan sebanyak 4% daripada *accuracy* kajian lepas dalam menggunakan MLP-Word-2Vec. Penemuan projek ini mencadangkan bahawa cadangan fitur tambahan meningkatkan prestasi model klasifikasi.

#### 4. KESIMPULAN

Projek ini telah mencapai objektif yang dinyatakan seperti berikut:

1. Mengenal pasti fitur-fitur relevan berkaitan respons kecemasan dalam pengurusan bencana banjir.

Fitur “amaran” diperkenalkan sebagai fitur baharu dalam klasifikasi tweet bencana untuk menambahbaik pengurusan bencana banjir semasa. Penemuan projek ini mencadangkan bahawa integrasi fitur tambahan meningkatkan prestasi model klasifikasi secara keseluruhannya.

2. Mencadangkan algoritma pembelajaran mesin yang mampu menganalisis ciapan untuk pengklasifikasian sentimen semasa pengurusan bencana banjir.

MNB-TFIDF menunjukkan keputusan yang terbaik menggunakan set data lebih kecil manakala MLP-TFIDF menunjukkan keputusan yang terbaik menggunakan set data lebih besar.

3. Menilai algoritma pengelasan yang dicadangkan menggunakan metrik penilaian.

MNB-TFIDF mencapai *accuracy* sebanyak 89.53%, *loss* sebanyak 40.60% dan *F1-score* sebanyak 89.53% menggunakan set data asal. MLP-TFIDF mencapai *accuracy* sebanyak 95.49%, *loss* sebanyak 15.46% dan *F1-score* sebanyak 95.49% menggunakan set data campur.

##### 4.1 Kekangan dan Cadangan Penambahbaikan

Data beranotasi manusia yang terhad menimbulkan cabaran besar dalam projek pengelasan teks disebabkan kos anotasi manusia yang tinggi. Ini menjadikannya tidak praktikal untuk mendapatkan set data berlabel yang komprehensif, menjadikannya sukar untuk melatih model yang mantap dan tepat. Ketidakseimbangan kelas ialah satu lagi isu dalam tugas pengelasan teks pengurusan bencana,

di mana kelas tertentu mungkin kurang diwakili disebabkan oleh kejadian yang jarang berlaku. Pembinaan CrisisNLP Word2Vec tidak boleh dilatih lebih lanjut mengenai data baharu kerana seni bina tetapnya, yang mungkin termasuk perkataan di luar perbendaharaan kata (OOV), menyebabkan prestasi menurun.

Memindahkan teknik pembelajaran boleh membantu mengatasi data beranotasi terhadap dengan memanfaatkan pengetahuan daripada model pra-latihan pada set data yang besar. Pendekatan ini membolehkan model menyesuaikan diri dengan konteks pengurusan bencana sambil mendapat manfaat daripada perwakilan yang dipelajari daripada set data yang lebih besar. Teknik pensampelan berlebihan dan pensampelan rendah boleh menangani ketidakseimbangan kelas dengan menghasilkan sampel sintetik untuk kelas minoriti dan mengalih keluar acara secara rawak daripada kelas majoriti. Set data seimbang ini meningkatkan keupayaan model untuk belajar daripada semua kelas dan meramalkan kejadian yang jarang berlaku. Untuk mengatasi penghadan benam tetap dalam CrisisNLPWord2Vec, terokai benam perkataan pra-latihan lain dan perwakilan perkataan kontekstual seperti BERT atau FastText. Teknik ini boleh diperhalusi untuk mengendalikan perkataan yang tiada perbendaharaan kata dan menyesuaikan diri dengan set data sasaran.

Kesimpulannya, integrasi fitur tambahan meningkatkan prestasi model klasifikasi secara keseluruhannya. Pendekatan kaedah mesin pembelajaran diselia menunjukkan keputusan terbaik menggunakan set data kecil manakala pendekatan kaedah pembelajaran mendalam menunjukkan keputusan terbaik menggunakan set data yang lebih besar.

## **PENGHARGAAN**

Sekalung penghargaan dan jutaan terima kasih saya ucapkan kepada penyelia saya Dr. Wandeep Kaur a/p Ratan Singh atas dedikasi dan kesabaran beliau dalam memberikan bimbingan dan nasihat sepanjang tempoh saya menyiapkan projek ini.

## RUJUKAN

- Abdeen, F. N., Fernando, T., Kulatunga, U., Hettige, S. & Ranasinghe, K. D. A. 2021. Challenges in multi-agency collaboration in disaster management: A Sri Lankan perspective. *International Journal of Disaster Risk Reduction* 62: 102399.
- Ahuja, R., Chug, A., Kohli, S., Gupta, S. & Ahuja, P. 2019. The impact of features extraction on the sentiment analysis. *Procedia computer science* 152: 341-348.
- Alam, F., Qazi, U., Imran, M. & Ofli, F. 2021. Humaid: Human-annotated disaster incidents data from twitter with deep learning benchmarks. *Proceedings of the International AAAI Conference on Web and social media*, pp.933-942.
- Barnhart, B. 2021. Social media demographics to inform your brand's strategy in 2021. *Sprout Social*.
- Behl, S., Rao, A., Aggarwal, S., Chadha, S. & Pannu, H. 2021. Twitter for disaster relief through sentiment analysis for COVID-19 and natural hazard crises. *International Journal of Disaster Risk Reduction* 55: 102101.
- Boas, I., Chen, C., Wiegel, H. & He, G. 2020. The role of social media-led and governmental information in China's urban disaster risk response: The case of Xiamen. *International Journal of Disaster Risk Reduction* 51: 101905.
- Bukar, U. A., Sidi, F., Jabar, M. A., Nor, R. N. H. B., Abdullah, S. & Ishak, I. 2022. A Multi-Stage Analysis of Predicting Public Resilience of Impactful Social Media Crisis Communication in Flooding Emergencies. *IEEE Access*.
- Coppola, D. P. 2006. *Introduction to international disaster management*. Elsevier.
- De Bruijn, J. A., De Moel, H., Jongman, B., Wagemaker, J. & Aerts, J. C. 2018. TAGGS: Grouping tweets to improve global geoparsing for disaster response. *Journal of Geovisualization and Spatial Analysis* 2(1): 1-14.
- Delimayanti, M. K., Sari, R., Laya, M., Faisal, M. R. & Naryanto, R. F. 2020. The effect of pre-processing on the classification of twitter's flood disaster messages using support vector machine algorithm. *2020 3rd International Conference on Applied Engineering (ICAE)*, pp.1-6.
- Firlia, K. Y., Faisal, M. R., Kartini, D., Nugroho, R. A. & Abadi, F. 2021. Analysis of New Features on the Performance of the Support Vector Machine Algorithm in Classification of Natural Disaster Messages. *2021 4th International Conference of Computer and Informatics Engineering (IC2IE)*, pp.317-322.
- Jamali, M., Nejat, A., Ghosh, S., Jin, F. & Cao, G. 2019. Social media data and post-disaster recovery. *International Journal of Information Management* 44: 25-37.
- Krippendorff, K. 2018. *Content analysis: An introduction to its methodology*. Sage publications.
- Lin, A., Wu, H., Liang, G., Cardenas-Tristan, A., Wu, X., Zhao, C. & Li, D. 2020. A big data-driven dynamic estimation model of relief supplies demand in urban flood disaster. *International Journal of Disaster Risk Reduction* 49: 101682.
- Maharani, W. 2020. Sentiment analysis during Jakarta flood for emergency responses and situational awareness in disaster management using BERT. *2020 8th International Conference on Information and Communication Technology (ICoICT)*, pp.1-5.
- Menne, B., Murray, V. & Organization, W. H. 2013. *Floods in the WHO European Region: health effects and their prevention*. World Health Organization. Regional Office for Europe.
- Nair, M. R., Ramya, G. & Sivakumar, P. B. 2017. Usage and analysis of Twitter during 2015 Chennai flood towards disaster management. *Procedia computer science* 115: 350-358.

- Ogie, R. I., Clarke, R. J., Forehead, H. & Perez, P. 2019. Crowdsourced social media data for disaster management: Lessons from the PetaJakarta.org project. *Computers, Environment and Urban Systems* 73: 108-117.
- Qorib, M., Oladunni, T., Denis, M., Ososanya, E. & Cotae, P. 2023. COVID-19 vaccine hesitancy: Text mining, sentiment analysis and machine learning on COVID-19 vaccination twitter dataset. *Expert Systems with Applications* 212: 118715.
- Rahmadan, M. C., Hidayanto, A. N., Ekasari, D. S. & Purwandari, B. 2020. Sentiment Analysis and Topic Modelling Using the LDA Method related to the Flood Disaster in Jakarta on Twitter. *2020 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*, pp.126-130.
- Ramachandran, D. & Parvathi, R. 2019. Analysis of twitter specific preprocessing technique for tweets. *Procedia computer science* 165: 245-251.
- Ridwan, A., Nuha, H. H. & Dharayani, R. 2022. Sentiment Analysis of Floods on Twitter Social Media Using the Naive Bayes Classifier Method with the N-Gram Feature. *2022 International Conference on Data Science and Its Applications (ICoDSA)*, pp.114-118.
- Shreevastav, B. B., Tiwari, K. R., Mandal, R. A. & Singh, B. 2022. "Flood risk modeling in southern Bagmati corridor, Nepal"(a study from Sarlahi and Rautahat, Nepal). *Progress in Disaster Science*: 100260.
- Siegel, F. R. 2020. Coastal city flooding. In (ed.). *Adaptations of Coastal Cities to Global Warming, Sea Level Rise, Climate Change and Endemic Hazards*, pp.27-34. Springer.
- Tyshchuk, Y. & Wallace, W. A. 2018. Modeling human behavior on social media in response to significant events. *IEEE Transactions on Computational Social Systems* 5(2): 444-457.

Liyanna Sarah Wong binti Abdullah (A179939)  
Dr. Wandeep Kaur a/p Ratan Singh  
Fakulti Teknologi & Sains Maklumat,  
Universiti Kebangsaan Malaysia.