

Evaluasi Pengaruh Varian Daftar *Stopword* terhadap Kinerja Klasifikasi Teks Al-Qur'an dengan *Support Vector Machine* dan *Backpropagation Neural Network*

Afit Ajis Solihin^{*1}, Fandy Setyo Utomo², Azhari Shouni Barkah³

^{1,2}Program Magister Ilmu Komputer, Fakultas Ilmu Komputer, Universitas Amikom Purwokerto, Indonesia

³Program Studi Teknologi Informasi, Fakultas Ilmu Komputer, Universitas Amikom Purwokerto, Indonesia

Email: ¹afitajissolihin@gmail.com, ²fandy_setyo_utomo@amikompurwokerto.ac.id,
³azhari@amikompurwokerto.ac.id

Abstrak

Latar belakang penelitian ini adalah tantangan dalam mengklasifikasikan teks Al-Qur'an, yang disebabkan oleh kompleksitas struktur bahasa Arab dan perbedaan antara bahasa Arab klasik dan modern. Penggunaan teknik Natural Language Processing (NLP), khususnya *stopword removal*, menjadi penting dalam meningkatkan akurasi klasifikasi teks. Namun, pengaruh penggunaan varian *stopword* terhadap performa model klasifikasi teks Al-Qur'an belum banyak dieksplorasi. Tujuannya adalah untuk mengevaluasi pengaruh penerapan varian daftar *stopword* yang berbeda terhadap kinerja dua algoritma klasifikasi teks, yaitu *Support Vector Machine* (SVM) dan *Backpropagation Neural Network* (BPNN), dalam mengklasifikasikan ayat-ayat Al-Qur'an. Penelitian ini juga bertujuan untuk menganalisis bagaimana teknik seleksi fitur *Chi-Square* dan representasi *TF-IDF* dapat mempengaruhi efektivitas model. Metode yang digunakan meliputi pengumpulan dataset ayat-ayat Al-Qur'an dalam Bahasa Indonesia yang melalui preprocessing seperti tokenisasi, normalisasi, dan penghapusan *stopword* menggunakan tiga varian *stopword list*: *Sastrawi*, *NLTK*, dan kombinasi keduanya. Model klasifikasi diterapkan dengan algoritma SVM dan BPNN, serta dievaluasi menggunakan metrik akurasi, presisi, recall, dan *F1-score*. Hasil penelitian menunjukkan bahwa algoritma SVM memberikan performa yang lebih baik dan konsisten dibandingkan BPNN. Penggunaan *stopword NLTK* memberikan hasil terbaik dengan akurasi tertinggi sebesar 0,5849 dan *F1-score* 0,5438 pada SVM. BPNN menunjukkan hasil yang kurang optimal dengan akurasi tertinggi hanya 0,4292 dan *F1-score* yang lebih rendah dari 0,3 pada semua varian *stopword*. Kontribusi penelitian ini adalah menegaskan pentingnya pemilihan daftar *stopword* yang tepat untuk meningkatkan kinerja klasifikasi teks Al-Qur'an serta memberikan wawasan berharga dalam pengembangan sistem klasifikasi teks keagamaan yang lebih akurat menggunakan algoritma pembelajaran mesin.

Kata kunci: *Backpropagation Neural Network, Klasifikasi Teks Al-Qur'an, Penghapusan Stopword, Seleksi Fitur, Support Vector Machine, TF-IDF.*

Evaluation of the Impact of Stopword List Variants on Quranic Text Classification Performance Using Support Vector Machine and Backpropagation Neural Network

Abstract

*This research addresses the challenges of classifying Quranic texts due to the complex structure of Arabic and the differences between classical and modern Arabic. The use of Natural Language Processing (NLP) techniques, particularly *stopword removal*, is crucial for improving classification accuracy. However, the impact of different *stopword variants* on Quranic text classification models has not been widely explored. The aim of this study is to evaluate the effect of various *stopword lists* on the performance of two classification algorithms, *Support Vector Machine* (SVM) and *Backpropagation Neural Network* (BPNN), in classifying Quranic verses. Additionally, this research analyzes how feature selection techniques like *Chi-Square* and *TF-IDF* representation influence model effectiveness. The method involves collecting a dataset of Quranic verses in Indonesian, followed by preprocessing steps such as tokenization, normalization, and *stopword removal* using three *stopword lists*: *Sastrawi*, *NLTK*, and their combination. Classification models are applied using SVM and BPNN, evaluated with accuracy, precision, recall, and *F1-score* metrics. Results show that SVM performs better and more consistently than BPNN. The *NLTK* *stopword list* yields the best results, with the highest accuracy of 0.5849 and *F1-score* of 0.5438 on SVM. BPNN shows suboptimal performance with a maximum accuracy of 0.4292 and an *F1-score* below 0.3 across all*

stopword variants. This study emphasizes the importance of selecting the appropriate stopword list to improve Quranic text classification and offers valuable insights for developing more accurate religious text classification systems using machine learning.

Keywords: *Backpropagation Neural Network, Feature Selection, Qur'anic Text Classification, Stopword Removal, Support Vector Machine, TF-IDF.*

1. PENDAHULUAN

Teks Al-Qur'an memiliki peranan yang sangat penting dalam kehidupan umat Islam. Sebagai sumber utama ajaran Islam, Al-Qur'an mengandung prinsip-prinsip spiritual, hukum, moral, dan etika yang membimbing umat Islam dalam menjalani kehidupan sehari-hari. Penelitian menunjukkan bahwa pemahaman yang tepat terhadap teks Al-Qur'an dapat membantu dalam membentuk karakter individu yang berkeadaban dan meningkatkan kualitas kehidupan sosial masyarakat. Al-Qur'an juga berfungsi sebagai pedoman dalam membangun masyarakat yang adil, sejahtera, dan berperadaban [1], [2]. Oleh karena itu, memahami teks Al-Qur'an dengan tepat adalah suatu keharusan bagi setiap umat Islam. Namun, pemahaman teks Al-Qur'an tidaklah mudah. Salah satu tantangan utama yang dihadapi adalah perbedaan konteks sosial dan budaya antara zaman klasik dan modern. Banyak tafsir atau penafsiran yang terjadi dalam sejarah Al-Qur'an, yang bisa jadi tidak lagi relevan atau sulit diterapkan dalam konteks saat ini. Oleh karena itu, pendekatan yang menggabungkan metode hermeneutik—yakni pendekatan interdisipliner dengan ilmu sosial dan humaniora—menjadi sangat penting untuk memperkaya pemahaman Al-Qur'an dalam menghadapi dinamika zaman modern [3], [4]. Fenomena Living Qur'an, yang merujuk pada penerapan ajaran Al-Qur'an dalam kehidupan sehari-hari, memerlukan kajian yang mendalam untuk memastikan ajaran tersebut tetap relevan dengan kondisi zaman tanpa menyimpang dari makna asli yang terkandung dalam teks [5], [6]. Selain itu, tantangan yang lebih besar muncul dengan pengaruh globalisasi dan liberalisme yang memengaruhi cara pandang terhadap agama, termasuk dalam hal pemahaman ajaran Islam. Pendekatan moderasi (wasatiyah), yang menekankan pada jalan tengah, menjadi penting untuk menjaga keseimbangan antara interpretasi agama yang ekstrim dan inklusif [7]. Oleh karena itu, keberadaan ulama dan cendekiawan yang mampu memberikan penafsiran yang kontekstual sangat dibutuhkan untuk menjaga agar ajaran Al-Qur'an tetap relevan dengan perkembangan zaman.

Bahasa Arab, sebagai bahasa asli Al-Qur'an, memiliki struktur gramatikal yang kompleks dan kaya akan nuansa makna. Bahasa ini memiliki sistem morfologi yang memungkinkan pembentukan kata dari akar kata melalui penambahan afiks yang menciptakan ribuan kata dengan makna yang berbeda, bergantung pada konteks. Hal ini menjadikan bahasa Arab sangat khas dan memiliki kedalaman dalam menyampaikan makna yang tidak mudah diterjemahkan ke dalam bahasa lain [8], [9]. Struktur nahu dalam bahasa Arab memungkinkan penggunaan teknik retorik seperti balaghah yang memberikan keindahan serta kedalaman pada teks Al-Qur'an [10], [11]. Kehadiran metafora, perumpamaan, dan gaya bahasa yang penuh nuansa dalam Al-Qur'an memberikan lapisan-lapisan makna yang hanya bisa dipahami dengan pemahaman yang mendalam tentang bahasa Arab itu sendiri [12]. Oleh karena itu, untuk benar-benar memahami Al-Qur'an, dibutuhkan pemahaman tentang bahasa Arab, yang bukan hanya berfungsi sebagai alat komunikasi, tetapi juga sarana untuk menyampaikan nilai-nilai ilahi. Kesalahan dalam interpretasi teks Al-Qur'an bisa berakibat pada pemahaman yang keliru tentang prinsip-prinsip hukum Islam, hak asasi manusia, dan hubungan antaragama. Misalnya, kesalahan dalam menafsirkan ayat-ayat terkait hukum warisan dapat menyebabkan ketidakadilan gender, yang sering kali berdampak pada perempuan yang kurang mendapatkan hak warisan yang seharusnya mereka terima [13]. Interpretasi yang tidak tepat ini juga berpotensi memperburuk pandangan yang salah terhadap hak asasi manusia dalam Islam. Pemahaman yang keliru bisa menciptakan pandangan yang seolah-olah Islam mendukung diskriminasi atau pelanggaran hak, yang tentu saja bertentangan dengan ajaran sebenarnya [5]. Selain itu, kesalahan tafsir juga dapat memperburuk hubungan antaragama, mengingat kesalahpahaman terhadap ajaran toleransi dan kerukunan dapat memperburuk hubungan antara komunitas Muslim dan non-Muslim. Kesalahpahaman ini berpotensi menguatkan stereotip negatif dan menghalangi dialog antarbudaya yang konstruktif [14]. Oleh karena itu, penting untuk memiliki pemahaman yang tepat terhadap teks Al-Qur'an agar dapat menjaga keharmonisan dalam masyarakat yang multikultural.

Perkembangan teknologi, terutama dalam bidang Natural Language Processing (NLP), telah membuka peluang besar untuk menganalisis teks Al-Qur'an dan meningkatkan pemahaman terhadapnya. NLP memungkinkan penggunaan algoritma canggih untuk mengolah teks, termasuk stopword removal dan feature selection, yang dapat meningkatkan akurasi klasifikasi teks. Teknologi ini telah banyak digunakan dalam analisis sentimen, klasifikasi teks, dan pencarian informasi, yang dapat diaplikasikan untuk memahami lebih dalam makna dari teks-teks suci seperti Al-Qur'an [15], [16]. Di bidang pendidikan, NLP telah terbukti efektif dalam membantu siswa mempelajari bahasa, dan dalam konteks Al-Qur'an, dapat membantu umat Islam untuk memahami makna

setiap ayat dengan lebih mendalam. Penelitian menunjukkan bahwa aplikasi berbasis NLP dapat memberikan umpan balik yang berguna bagi pembelajaran bahasa, termasuk bahasa Arab [17]. Selain itu, NLP juga digunakan dalam menganalisis media sosial dan data tidak terstruktur lainnya untuk memberikan wawasan yang lebih objektif tentang pandangan publik [18], [19]. Namun, meskipun NLP menawarkan banyak manfaat, tantangan tetap ada, seperti potensi bias dalam model yang dapat memengaruhi akurasi dan interpretasi teks. Oleh karena itu, penting untuk terus melakukan penelitian untuk mengembangkan model NLP yang lebih transparan dan akurat agar penerapannya lebih efektif dan bebas dari bias [20].

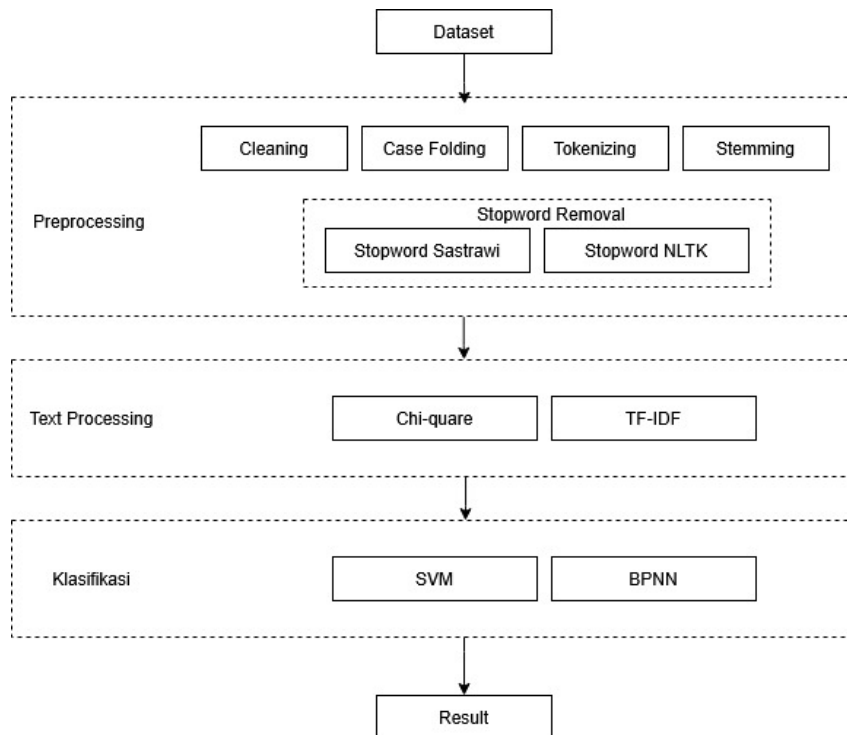
Klasifikasi teks Al-Qur'an menghadapi tantangan besar karena perbedaan antara bahasa Arab klasik yang digunakan dalam teks asli dan bahasa Arab modern, serta konteks sejarah yang melingkupi setiap ayat. Bahasa Arab klasik memiliki struktur gramatikal yang lebih kompleks dan kaya nuansa makna, yang membutuhkan pemahaman mendalam untuk mencegah kesalahan interpretasi terhadap ajaran Islam [21]. Masalah ini semakin kompleks dengan adanya variasi dalam pembacaan Al-Qur'an dan perbedaan dalam pengenalan stopword dalam bahasa Arab klasik dan modern [22]. Meskipun banyak penelitian menggunakan Natural Language Processing (NLP) dan stopword removal, penelitian yang menguji pengaruh varian stopword terhadap klasifikasi teks Al-Qur'an masih terbatas. Beberapa penelitian menunjukkan bahwa penghapusan stopword dapat meningkatkan akurasi klasifikasi teks umum, tetapi dalam konteks Al-Qur'an, hal ini tidak selalu tepat karena nuansa yang terkandung dalam kata-kata yang dianggap stopword [23]. Penelitian lebih lanjut dibutuhkan untuk mengeksplorasi lebih dalam pengaruh varian stopword dalam klasifikasi teks Al-Qur'an, dengan mempertimbangkan aspek linguistik Arab klasik dan konteks sejarah [24]. Sistem klasifikasi yang lebih akurat diperlukan untuk mendukung pemahaman yang lebih baik terhadap teks-teks agama, yang dapat mempercepat proses pembelajaran dan memastikan pemahaman yang benar [25].

Penggunaan Natural Language Processing (NLP) dalam teks keagamaan, khususnya dalam penghapusan stopword, telah diterapkan dalam berbagai konteks, seperti analisis Hadis dan terjemahan Al-Qur'an. Stopword removal dapat menyederhanakan teks untuk analisis lebih lanjut, namun dalam teks keagamaan, kata-kata yang tampaknya tidak bermakna mungkin memiliki makna yang mendalam, yang dapat mengubah pemahaman [26]. Dalam analisis Hadis, penggunaan teknik ini dapat membantu dalam mengklasifikasikan hadis berdasarkan tema, meskipun harus ada keseimbangan agar tidak mengurangi makna yang terkandung dalam struktur kalimat [27]. Penghapusan stopword juga menimbulkan tantangan dalam terjemahan Al-Qur'an karena menghilangkan nuansa penting yang terdapat dalam teks asli [28]. Oleh karena itu, metode berbasis konteks diperlukan untuk memilih stopword yang tepat dan mempertahankan makna teks [29]. Metode yang digunakan dalam penelitian NLP pada teks keagamaan termasuk teknik klasifikasi berbasis pembelajaran mesin seperti SVM dan BPNN, serta metode pengolahan teks seperti TF-IDF. Misalnya, [30] menerapkan teknik pemilihan fitur menggunakan TF-IDF untuk mengidentifikasi kata-kata yang relevan sebelum mengaplikasikannya dalam model klasifikasi, yang menunjukkan pentingnya pertimbangan terhadap nuansa makna dalam teks keagamaan. Penggunaan teknik ini memberikan hasil yang lebih baik dalam klasifikasi teks yang lebih kompleks. Beberapa penelitian relevan menunjukkan bahwa stopword removal dapat memberikan manfaat dalam analisis teks keagamaan, meskipun hasilnya bervariasi tergantung konteks. Riset [31] mengembangkan daftar stopword komprehensif untuk teks Al-Qur'an, yang membantu meningkatkan efisiensi alat analisis teks. Riset [32] menekankan pentingnya teknik text mining dalam Al-Qur'an, di mana stopword removal digunakan untuk membersihkan data dan meningkatkan proses pengolahan algoritmik. Secara keseluruhan, penelitian-penelitian ini menyoroti pentingnya penggunaan stopword removal dengan memperhatikan konteks linguistik dan budaya dalam analisis teks keagamaan untuk mencapai pemahaman yang lebih akurat.

Penelitian ini bertujuan untuk mengevaluasi pengaruh varian daftar stopword terhadap kinerja klasifikasi teks ayat Al-Qur'an menggunakan dua algoritma pembelajaran mesin, yaitu SVM dan BPNN. Selain itu, penelitian ini juga bertujuan untuk menganalisis pengaruh seleksi fitur menggunakan teknik Chi-Square serta representasi TF-IDF terhadap kinerja model dalam klasifikasi teks Al-Qur'an. Dengan pendekatan ini, penelitian berharap dapat memberikan wawasan lebih dalam tentang pentingnya pemilihan stopword yang tepat dalam meningkatkan akurasi model klasifikasi dalam konteks teks keagamaan. Penelitian ini diharapkan memberikan kontribusi signifikan terhadap pengembangan sistem klasifikasi teks keagamaan, khususnya Al-Qur'an, dengan memanfaatkan teknologi Natural Language Processing (NLP) dan algoritma pembelajaran mesin. Hasil penelitian ini akan membantu memperbaiki pemahaman dan aplikasi teks Al-Qur'an melalui sistem klasifikasi yang lebih akurat. Implikasi dari penelitian ini juga dapat diterapkan pada bidang serupa, seperti pengklasifikasian teks agama lainnya atau teks dengan konteks linguistik yang kompleks, serta dapat meningkatkan efektivitas dan akurasi dalam analisis teks keagamaan yang lebih luas.

2. METODE PENELITIAN

Tahapan yang dilakukan dalam penelitian ini mengacu pada kerangka penelitian seperti pada Gambar 1 berikut.



Gambar 1. Kerangka Penelitian

2.1 Dataset

Dataset adalah kumpulan informasi yang diatur dan digunakan untuk berbagai tujuan, seperti analisis, pelatihan demonstrasi pembelajaran mesin, atau penelitian. Dataset dapat berisi berbagai jenis informasi, seperti teks, gambar, angka, atau kategori informasi, bergantung pada tujuan penggunaannya[33]. Penelitian ini menggunakan dataset Terjemahan Alquran dalam Bahasa Indonesia, Tafsir Quraish Shihab, dan Tafsir Quran dari Kementerian Agama RI, Topik tematik dari Alquran Cordoba dijadikan sebagai kelas target untuk eksperimen, seperti dicontohkan pada Tabel 1. Studi ini menggunakan tiga kategori yang sudah dikelompokkan sebelumnya, yaitu Akhlak, Al-Quran, dan Bangsa-bangsa terdahulu (Tabel 2) [34].

Tabel 1. Contoh Dataset

| Surah id | Surah name | Surah meaning | Verse text arabic | Verse text Indonesian translation | Thematic domain name |
|----------|------------|---------------|--|---|----------------------|
| 2 | Al-Baqarah | Sapi Betina | بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ | Alif laam miim. | Al-Quran |
| 2 | Al-Baqarah | Sapi Betina | ذَلِكَ الْكِتَابُ لَا رَيْبَ فِيهِ هُدًى لِّلْمُتَّقِينَ | Kitab (Al Quran) ini tidak ada keraguan padanya; petunjuk bagi mereka yang bertakwa, | |
| 2 | Al-Baqarah | Sapi Betina | الَّذِينَ يُؤْمِنُونَ بِالْغَيْبِ وَيُقِيمُونَ الصَّلَاةَ وَمِمَّا رَزَقْنَاهُمْ يُنْفِقُونَ | (yaitu) mereka yang beriman kepada yang ghaib, yang mendirikan shalat, dan menafkahkan sebahagian rezeki yang Kami anugerahkan kepada mereka. | Akhlak |

Tabel 2. Daftar Topik Tematik

| No | Label | Jumlah |
|--------------|-------------------------|------------|
| 1 | Akhlak | 218 |
| 2 | Al-Quran | 183 |
| 3 | Bangsa-bangsa Terdahulu | 127 |
| Total | | 528 |

Selain dataset Terjemahan Al-quran penelitian ini juga menggunakan dua buah daftar stopwords yang digunakan dalam eksperimen yaitu Daftar Stopwords Indonesia dari Library Sastrawi dan Stopwords Indonesia dari Library NLTK.

2.2 Pra-pemrosesan

Pra-pemrosesan digunakan untuk mengubah data yang sebelumnya tidak terstruktur menjadi data terstruktur[35]. Proses ini bertujuan untuk membersihkan dan menyiapkan data sebelum memasukkan model. Penelitian ini menggunakan pra-pemrosesan cleaning, case folding, tokenizing, stemming dan stopwords removals[36].

a) *Cleaning*

Cleaning merupakan proses pembersihan data, proses pembersihan data meliputi menghilangkan angka, simbol, dan karakter non-huruf yang bertujuan untuk meningkatkan kualitas data dan akurasi model pembelajaran mesin[7]. Proses cleaning ditunjukkan pada Tabel 3 dan semua karakter yang tidak relevan akan dihilangkan.

Tabel 3. Contoh hasil *Cleaning*

| Data Input | Data Output |
|---|--|
| Kitab (Al Quran) ini tidak ada keraguan padanya; petunjuk bagi mereka yang bertakwa, (yaitu) mereka yang beriman kepada yang ghaib, yang mendirikan shalat, dan menafkahkan sebahagian rezeki yang Kami anugerahkan kepada mereka. Dan mereka yang beriman kepada Kitab (Al Quran) yang telah diturunkan kepadamu dan Kitab-kitab yang telah diturunkan sebelumnya, serta mereka yakin akan adanya (kehidupan) akhirat. | Kitab Al Quran ini tidak ada keraguan padanya petunjuk bagi mereka yang bertakwa yaitu mereka yang beriman kepada yang ghaib yang mendirikan shalat dan menafkahkan sebahagian rezeki yang Kami anugerahkan kepada mereka dan mereka yang beriman kepada Kitab Al Quran yang telah diturunkan kepadamu dan Kitab-kitab yang telah diturunkan sebelumnya serta mereka yakin akan adanya kehidupan akhirat |

b) *Case folding*

Untuk mendukung Pra-pemrosesan harus dilakukan penyamaan format teks yang diolah, penyamaan format teks dapat dilakukan dengan teknik *case folding*. *Case folding* merupakan proses mengubah seluruh teks menjadi huruf kecil untuk konsistensi [37]. Tabel 4 menunjukkan contoh hasil *case folding*.

Tabel 4. Contoh hasil *case folding*

| Data Input | Data Output |
|--|--|
| Kitab Al Quran ini tidak ada keraguan padanya petunjuk bagi mereka yang bertakwa yaitu mereka yang beriman kepada yang ghaib yang mendirikan shalat dan menafkahkan sebahagian rezeki yang Kami anugerahkan kepada mereka dan mereka yang beriman kepada Kitab Al Quran yang telah diturunkan kepadamu dan Kitab-kitab yang telah diturunkan sebelumnya serta mereka yakin akan adanya kehidupan akhirat | kitab al quran ini tidak ada keraguan padanya petunjuk bagi mereka yang bertakwa yaitu mereka yang beriman kepada yang ghaib yang mendirikan shalat dan menafkahkan sebahagian rezeki yang kami anugerahkan kepada mereka mereka yang beriman kepada kitab al quran yang telah diturunkan kepadamu dan kitab-kitab yang telah diturunkan sebelumnya serta mereka yakin akan adanya kehidupan akhirat |

c) *Stopwords removal*

Salah satu tahapan penting dalam preprocessing adalah penghapusan kata-kata umum atau stopwords, yaitu kata-kata yang sering muncul namun tidak memiliki makna penting dalam klasifikasi. Tujuan penghapusan ini adalah untuk mengurangi noise dan meningkatkan efisiensi serta akurasi model klasifikasi.

Dalam penelitian ini, digunakan tiga varian stopwords list :

1. *Stopword List Sastrawi*

Daftar ini diambil dari pustaka Sastrawi, yang umum digunakan dalam pemrosesan teks Bahasa Indonesia, dicontohkan pada Tabel 5.

Tabel 5. Daftar Stopwords list Sastrawi

| Daftar Stopwords |
|--|
| yang, untuk, pada, dan, ke, dari, di, adalah, itu, dalam, tidak, ini, dengan, sebagai, oleh, atau, juga, tapi, karena, maka, bahwa, yaitu, telah, bisa, hanya, sudah, bila, setiap, semua, sangat, namun, saat, agar, akan, harus, menjadi, dimana, tanpa, apakah, antara, sesudah, sebelum, ketika, tentang, kepada, sehingga, mereka, kami, saya, aku, dia, ia, kita, kamu, engkau |

Penghapusan kata umum (*stopword removal*) menggunakan daftar stopwords dari library Sastrawi. Tabel 6 menampilkan contoh hasil proses penghapusan stopwords dari salah satu ayat dalam dataset:

Tabel 6. Contoh Hasil Stopwords Removal dengan Sastrawi

| Data Input | Data Output |
|---|--|
| kitab al quran ini tidak ada keraguan padanya petunjuk bagi mereka yang bertakwa yaitu mereka yang beriman kepada yang ghaib yang mendirikan shalat dan menafkahkan sebahagian rezeki yang kami anugerahkan kepada mereka dan mereka yang beriman kepada kitab al quran yang telah diturunkan kepadamu dan kitabkitab yang telah diturunkan sebelumnya serta mereka yakin akan adanya kehidupan akhirat | kitab al quran keraguan padanya petunjuk bertakwa beriman ghaib mendirikan shalat menafkahkan sebahagian rezeki anugerahkan beriman kitab al quran diturunkan kepadamu kitabkitab diturunkan sebelumnya yakin adanya kehidupan akhirat |

2. *Stopword List NLTK (Bahasa Indonesia)*

Daftar ini diambil dari pustaka NLTK Bahasa Indonesia, dicontohkan pada Tabel 7.

Tabel 7. Daftar Stopwords list NLTK

| Daftar Stopwords |
|--|
| ada, adalah, adakah, akankah, aku, anda, apakah, atau, bagaimana, bahwa, banyak, beberapa, belum, boleh, bisa, bukan, dalam, dan, dengan, dia, dahulu, dulu, dimana, hingga, ia, ini, itu, jadi, jika, juga, kami, karena, kata, ke, kembali, ketika, kita, lagi, lalu, lebih, maka, melalui, mereka, meskipun, harus, misalnya, mungkin, tidak, untuk, walaupun, yang |

Penghapusan kata umum (*stopword removal*) menggunakan daftar stopwords dari library NLTK. Tabel 8 menampilkan contoh hasil proses penghapusan stopwords dari salah satu ayat dalam dataset.

Tabel 8. Contoh Hasil Stopwords Removal dengan NLTK

| Data Input | Data Output |
|---|---|
| kitab al quran ini tidak ada keraguan padanya petunjuk bagi mereka yang bertakwa yaitu mereka yang beriman kepada yang ghaib yang mendirikan shalat dan menafkahkan sebahagian rezeki yang kami anugerahkan kepada mereka dan mereka yang beriman kepada kitab al quran yang telah diturunkan kepadamu dan kitabkitab yang telah diturunkan sebelumnya serta mereka yakin akan adanya kehidupan akhirat | Kitab al quran keraguan petunjuk bertakwa beriman ghaib mendirikan shalat menafkahkan sebahagian rezeki anugerahkan beriman kitab al quran diturunkan kepadamu kitabkitab diturunkan sebelumnya kehidupan akhirat |

3. *Stopword Kombinasi (Sastrawi dan NLTK)*

Daftar ini merupakan gabungan dari stopwords Sastrawi dan NLTK, yang kemudian disaring untuk menghapus duplikasi, ditampilkan pada Tabel 9.

Tabel 9. Daftar Stopwords list Kombinasi

| Daftar Stopwords |
|---|
| ada, adalah, adakah, akan, akankah, aku, anda, apakah, atau, bagaimana, bahwa, banyak, beberapa, belum, boleh, bisa, bukan, dalam, dan, dengan, dia, dahulu, dulu, dimana, hingga, ia, ini, itu, jadi, jika, juga, kami, karena, kata, ke, kembali, ketika, kita, lagi, lalu, lebih, maka, melalui, mereka, meskipun, harus, misalnya, mungkin, tidak, untuk, walaupun, yang, pada, sebagai, oleh, tapi, telah, hanya, sudah, bila, setiap, semua, sangat, namun, saat, agar, menjadi, tanpa, sesudah, sebelum, tentang, kepada, saya, engkau, kamu |

Penghapusan kata umum (*stopword removal*) menggunakan daftar stopwords gabungan. Tabel 10 menampilkan contoh hasil proses penghapusan stopwords dari salah satu ayat dalam dataset.

Tabel 10. Contoh Hasil Stopwords Removal dengan Sastrawi & NLTK

| Data Input | Data Output |
|---|---|
| kitab al quran ini tidak ada keraguan padanya petunjuk bagi mereka yang bertakwa yaitu mereka yang beriman kepada yang ghaib yang mendirikan shalat dan menafkahkan sebahagian rezeki yang kami anugerahkan kepada mereka dan mereka yang beriman kepada kitab al quran yang telah diturunkan kepadamu dan kitabkitab yang telah diturunkan sebelumnya serta mereka yakin akan adanya kehidupan akhirat | kitab al quran keraguan petunjuk bertakwa beriman ghaib mendirikan shalat menafkahkan sebahagian rezeki anugerahkan beriman kitab al quran diturunkan kepadamu kitabkitab diturunkan sebelumnya kehidupan akhirat |

d) *Tokenisasi*

Tokenisasi memecah teks menjadi unit-unit kecil, seperti kata atau karakter, yang memudahkan analisis dan pelatihan model. Pada tabel 11 menampilkan contoh hasil tokenisasi.

Tabel 11. Contoh Hasil *Tokenisasi*

| Data Input | Data Output |
|---|--|
| kitab al quran keraguan petunjuk bertakwa beriman ghaib mendirikan shalat menafkahkan sebahagian rezeki anugerahkan beriman kitab al quran diturunkan kepadamu kitabkitab diturunkan sebelumnya kehidupan akhirat | [kitab, al, quran, keraguan, petunjuk, bertakwa [beriman, ghaib, mendirikan, shalat, menafkahkan, sebahagian, rezeki, anugerahkan] [beriman, kitab, al, quran, diturunkan, kepadamu, kitabkitab, diturunkan, sebelumnya, kehidupan, akhirat] |

e) *Stemming*

Stemming mengubah kata menjadi bentuk dasarnya dengan menghilangkan imbuhan, seperti "berlari" dan "pelari" menjadi "lari", untuk menyederhanakan variasi kata dengan makna serupa, dicontohkan pada Tabel 12. Semua langkah ini penting untuk meningkatkan kualitas data dan akurasi dalam aplikasi seperti klasifikasi teks dan analisis sentimen [38].

Tabel 12. Contoh Hasil *Stemming*

| Data Input | Data Output |
|---|---|
| kitab al quran keraguan petunjuk bertakwa beriman ghaib mendirikan shalat menafkahkan sebahagian rezeki anugerahkan beriman kitab al quran diturunkan kepadamu kitabkitab diturunkan sebelumnya kehidupan akhirat | kitab al quran ragu tunjuk takwa iman ghaib diri shalat nafkah sebahagian rezeki anugerah iman kitab al quran turun kepada kitabkitab turun belum hidup akhirat |

2.3 Pemrosesan Awal Teks

Pemrosesan awal teks dilakukan dengan menggunakan metode Chi-Square dan TF-IDF untuk mengubah teks menjadi representasi numerik yang dapat diproses lebih lanjut dalam model pembelajaran mesin. TF-IDF (Term Frequency-Inverse Document Frequency) digunakan untuk menilai seberapa penting suatu kata dalam suatu dokumen dibandingkan dengan kumpulan dokumen (korpus). Dengan menghitung frekuensi kata dalam dokumen relatif terhadap frekuensi kemunculannya di seluruh korpus, TF-IDF membantu dalam menilai relevansi kata dalam konteks tertentu, sehingga meningkatkan akurasi model klasifikasi teks. Penggunaan TF-IDF dipilih karena

kemampuannya untuk mengidentifikasi kata yang paling signifikan dalam dokumen, yang memungkinkan model untuk fokus pada informasi yang relevan dan mengurangi pengaruh kata yang tidak terlalu informatif [39], [40].

Untuk mengurangi dimensi data dan meningkatkan efisiensi model, Pemilihan Fitur Chi-Square digunakan untuk memilih 1000 fitur terbaik yang paling relevan dengan label klasifikasi. Chi-Square dipilih karena metode ini efektif dalam menyeleksi fitur yang signifikan dengan menghitung statistik Chi-Square, yang mengukur hubungan statistik antara fitur dan label dalam data. Beberapa penelitian menunjukkan bahwa penggunaan Chi-Square dalam seleksi fitur dapat meningkatkan akurasi model klasifikasi seperti SVM dan Naive Bayes, dengan peningkatan akurasi rata-rata sebesar 2,45% [41]. Selain itu, seleksi fitur ini juga mempercepat proses komputasi model, yang sangat penting dalam pengolahan data besar dan pemodelan algoritma yang lebih efisien [42], [43]. Dengan demikian, kombinasi teknik Chi-Square dan TF-IDF terbukti mampu menghasilkan model klasifikasi yang lebih akurat dan efisien dalam analisis teks, meningkatkan performa model berbasis NLP [44].

2.4 Klasifikasi

Pemilihan algoritma SVM dan BPNN dalam penelitian ini didasarkan pada keunggulannya yang telah terbukti dalam berbagai studi terdahulu dalam pengolahan teks, khususnya untuk klasifikasi teks berbasis Natural Language Processing (NLP). SVM dikenal efektif dalam menangani data berdimensi tinggi, memiliki kemampuan untuk menangani data linear maupun non-linear dengan menggunakan kernel, dan terbukti lebih unggul dibandingkan algoritma klasifikasi lain dalam banyak penelitian. Sebagai contoh, penelitian oleh [45] menunjukkan bahwa SVM mencapai akurasi tinggi dalam analisis sentimen dengan hasil terbaik mencapai 95.56%. Di sisi lain, BPNN, sebagai jaringan syaraf tiruan yang menggunakan teknik backpropagation untuk memperbarui bobot, juga memberikan performa yang sangat baik dalam tugas klasifikasi berbasis data kompleks. Penelitian yang dilakukan oleh [46] menunjukkan bahwa BPNN dapat mencapai akurasi yang hampir setara dengan SVM, yaitu 99.84%, dalam tugas klasifikasi tertentu.

Pada penelitian ini, SVM dan BPNN diterapkan untuk mengklasifikasikan teks Al-Qur'an dengan parameter yang disesuaikan untuk memperoleh hasil yang optimal. Untuk SVM, parameter yang digunakan meliputi C (penalti atau regularisasi) dan kernel. Nilai C mengontrol trade-off antara margin dan kesalahan klasifikasi, dengan nilai yang lebih tinggi mengurangi toleransi terhadap kesalahan dan berisiko menyebabkan overfitting. Pada penelitian ini, kernel yang digunakan adalah Radial Basis Function (RBF), yang efektif dalam menangani data non-linear. Model SVM dilatih menggunakan 80% data pelatihan dan diuji dengan 20% data uji, kemudian dievaluasi dengan metrik Accuracy, Precision, dan Recall. Sedangkan pada BPNN, model jaringan syaraf terdiri dari lapisan input, lapisan tersembunyi dengan 128 neuron pada lapisan pertama dan 64 neuron pada lapisan kedua, serta lapisan keluaran yang jumlah neuronnya disesuaikan dengan jumlah kelas yang ada. BPNN dilatih selama 10 epoch dengan ukuran batch 32. Salah satu parameter utama yang digunakan dalam pelatihan BPNN adalah learning rate, yang mengontrol kecepatan pembaruan bobot.

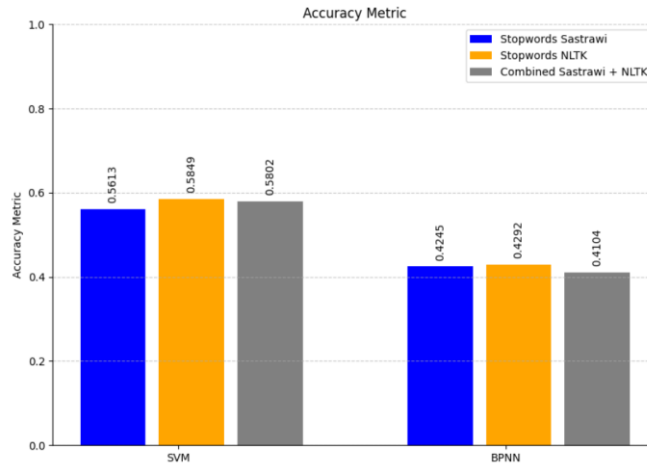
2.5 Evaluasi

Setelah model klasifikasi dibangun menggunakan SVM dan BPNN, langkah selanjutnya adalah mengevaluasi kinerja model. Metrik evaluasi yang umum dan standar yang digunakan dalam klasifikasi meliputi akurasi, presisi, recall, dan F1-score [47]. Evaluasi akurasi, presisi, dan recall dihitung untuk membandingkan kinerja SVM dan BPNN. Kemudian akan dibuat visualisasi grafik batang untuk menampilkan perbandingan metrik (akurasi, presisi, dan recall) dari kedua metode [48].

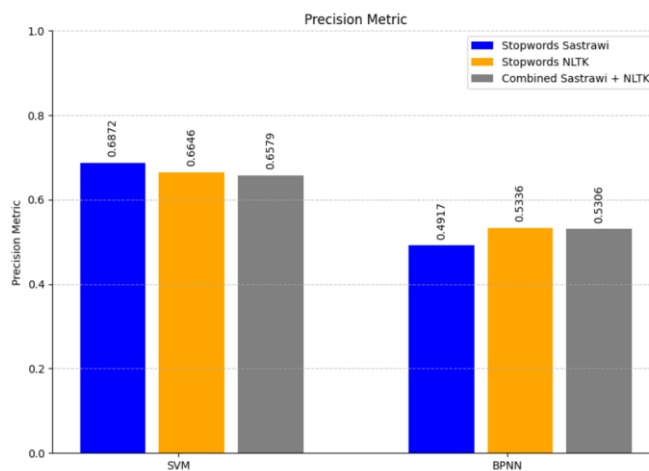
3. HASIL

Bagian ini menyajikan hasil penelitian yang diperoleh dari serangkaian eksperimen yang dilakukan untuk mengevaluasi pengaruh penggunaan varian daftar stopwords terhadap performa klasifikasi teks Al-Qur'an menggunakan algoritma SVM dan BPNN. Penelitian ini bertujuan untuk menganalisis sejauh mana strategi preprocessing, khususnya dalam pemilihan dan penerapan daftar stopwords yang berbeda, dapat memengaruhi efektivitas model klasifikasi. Eksperimen dilakukan dengan menggunakan dataset ayat-ayat Al-Qur'an yang telah diberi label berdasarkan klasifikasi tertentu, dan melalui tahapan preprocessing standar seperti tokenisasi, normalisasi, serta pembersihan teks dengan stopwords list. Eksperimen dalam penelitian ini dibagi menjadi dua bagian utama, yaitu pengujian algoritma SVM dan BPNN, masing-masing dengan beberapa varian daftar stopwords. Setiap varian stopwords list diterapkan dalam proses preprocessing untuk melihat pengaruhnya terhadap performa klasifikasi. Daftar stopwords yang digunakan dalam penelitian ini adalah Daftar Stopwords Indonesia dari Library Sastrawi dan Stopwords Indonesia dari Library NLTK.

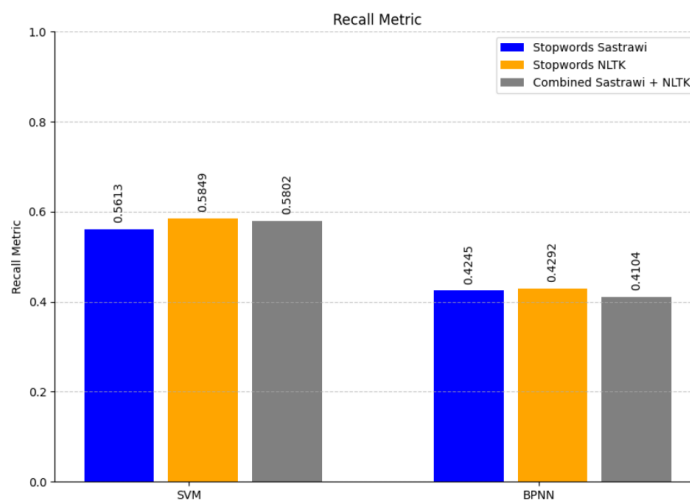
Evaluasi dilakukan dengan menggunakan empat metrik utama, yaitu akurasi, presisi, recall, dan F1-score. Untuk mempermudah pemahaman terhadap hasil evaluasi, Gambar 2, 3, 4 dan 5 menampilkan grafik perbandingan metrik performa antara algoritma dan varian stopwords, dan Tabel 13 menampilkan ringkasan hasil evaluasi.



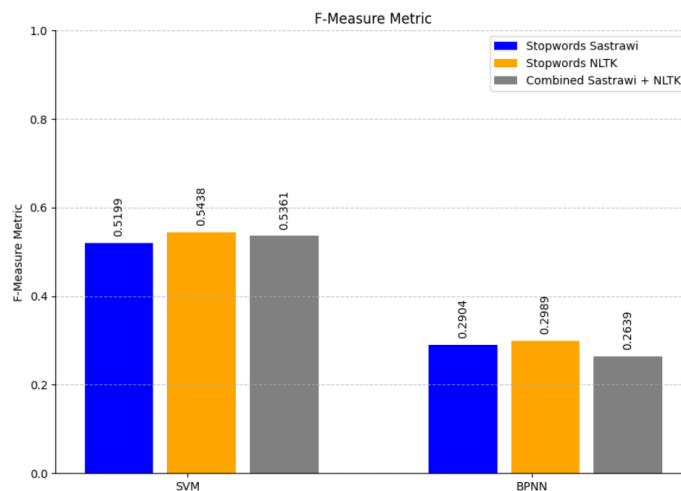
Gambar 2. Hasil evaluasi Accuracy



Gambar 3. Hasil evaluasi Precision



Gambar 4. Hasil evaluasi Recall



Gambar 5. Hasil evaluasi *F-Measure*

Tabel 13. Hasil Evaluasi

| Varian Stopword | Akurasi | | Presisi | | Recall | | F1-score | |
|-----------------------------|---------|--------|---------|--------|--------|--------|----------|--------|
| | SVM | BPNN | SVM | BPNN | SVM | BPNN | SVM | BPNN |
| Sastrawi | 0,5613 | 0,4245 | 0,6872 | 0,4917 | 0,5613 | 0,4245 | 0,5199 | 0,2904 |
| NLTK | 0,5849 | 0,4292 | 0,6646 | 0,5336 | 0,5849 | 0,4292 | 0,5438 | 0,2989 |
| kombinasi sastrawi dan nltk | 0,5802 | 0,4104 | 0,6579 | 0,5306 | 0,5802 | 0,4104 | 0,5361 | 0,2639 |

Hasil evaluasi menunjukkan bahwa pemilihan varian stopwords list memberikan pengaruh signifikan terhadap kinerja klasifikasi, serta perbedaan mencolok antara performa algoritma SVM dan BPNN. Akurasi tertinggi dicapai oleh SVM dengan stopwords NLTK 0,5849, menandakan bahwa NLTK cukup efektif menghapus kata-kata tidak penting tanpa menghilangkan informasi bermakna. Sementara BPNN memiliki akurasi tertinggi hanya 0,4292 dengan NLTK, menunjukkan performa yang masih jauh dari optimal. Presisi tertinggi ditemukan pada SVM dengan stopwords Sastrawi 0,6872, meskipun recall-nya lebih rendah. Hal ini mengindikasikan bahwa model lebih berhati-hati dalam mengklasifikasi dan cenderung menghindari false positive. Di sisi lain, BPNN cenderung memiliki presisi yang lebih rendah dan tidak konsisten. Recall tertinggi juga ditemukan pada SVM dengan stopwords NLTK 0,5849, artinya model mampu menangkap lebih banyak instance yang relevan. Ini penting untuk konteks klasifikasi ayat yang beragam secara tematik dan semantik. F1-score tertinggi diperoleh dari SVM dengan stopwords NLTK 0,5438, menandakan kombinasi presisi dan recall yang paling seimbang. Sementara BPNN memiliki F1-score yang jauh lebih rendah, bahkan di bawah 0,3 pada semua varian, yang menunjukkan ketidakseimbangan antara presisi dan recall-nya.

4. PEMBAHASAN

Hasil penelitian ini menunjukkan bahwa pemilihan varian stopwords list memiliki pengaruh yang signifikan terhadap performa klasifikasi teks Al-Qur'an, dan terdapat perbedaan mencolok antara dua algoritma yang digunakan, yaitu SVM dan BPNN. Pada umumnya, SVM menunjukkan kinerja yang lebih baik dibandingkan dengan BPNN, baik dalam hal akurasi, presisi, recall, maupun F1-score. Hal ini mengindikasikan bahwa SVM lebih mampu menangani data yang memiliki dimensi tinggi dan kompleksitas linguistik seperti yang ada pada teks Al-Qur'an. Hasil terbaik pada SVM tercatat dengan penggunaan stopwords NLTK, dengan akurasi mencapai 0,5849, presisi 0,6646, recall 0,5849, dan F1-score 0,5438. Ini menunjukkan bahwa stopwords NLTK efektif dalam menghapus kata-kata yang tidak bermakna tanpa mengurangi informasi penting yang terkandung dalam teks, sehingga membantu meningkatkan kinerja klasifikasi.

Di sisi lain, BPNN menunjukkan performa yang lebih rendah, dengan akurasinya hanya mencapai 0,4292 untuk stopwords NLTK, yang jauh lebih rendah dibandingkan dengan SVM. Penurunan performa BPNN ini dapat disebabkan oleh beberapa faktor, termasuk kesulitan model dalam menangani kompleksitas data teks yang kaya akan makna dan semantik dalam Al-Qur'an. BPNN mungkin mengalami kesulitan dalam memahami hubungan antar kata yang lebih dalam, dan kurangnya penyesuaian parameter, seperti jumlah lapisan dan ukuran batch, bisa menyebabkan performa yang tidak optimal [46]. Hal ini menunjukkan bahwa meskipun BPNN memiliki potensi dalam klasifikasi data kompleks, perlu ada peningkatan dalam penyesuaian model untuk mencapai hasil yang

optimal. Selain itu, meskipun BPNN mampu menangkap pola dalam data, SVM cenderung lebih stabil dan unggul dalam tugas klasifikasi teks seperti yang terlihat pada penelitian sebelumnya oleh [49]. Dalam konteks klasifikasi ayat-ayat Al-Qur'an, metrik yang digunakan, seperti presisi, recall, dan F1-score, sangat relevan karena mengukur seberapa baik model dapat mengenali dan menangkap ayat-ayat yang relevan serta menghindari kesalahan klasifikasi yang berpotensi merugikan pemahaman makna ayat tersebut.

Perbandingan dengan penelitian terdahulu yang mengaplikasikan stopwords removal dalam analisis teks keagamaan menunjukkan konsistensi hasil yang serupa. Misalnya, penelitian oleh [31] juga menunjukkan bahwa penghapusan stopwords dapat meningkatkan efisiensi analisis teks, namun tantangan utama tetap pada kebutuhan untuk mempertimbangkan konteks bahasa yang lebih mendalam. Riset [50] menyarankan agar meskipun stopwords removal dapat berguna untuk menyederhanakan analisis, pendekatan semantik yang lebih mendalam perlu dipertimbangkan untuk memastikan bahwa makna yang lebih dalam tidak hilang selama proses tersebut. Selain itu, studi oleh [32] mengindikasikan bahwa teknik seperti text mining yang menggabungkan stopwords removal dapat membantu mengidentifikasi tema utama dalam Al-Qur'an, tetapi juga menekankan bahwa perbaikan teknik dalam proses preprocessing sangat penting untuk menjaga keakuratan klasifikasi, sebagaimana yang ditemukan dalam penelitian ini.

Dalam konteks klasifikasi ayat-ayat Al-Qur'an, metrik yang digunakan memiliki makna yang sangat penting untuk mengevaluasi performa model. Akurasi mengukur seberapa sering model melakukan prediksi yang benar, namun tidak memberikan gambaran yang lengkap jika terdapat ketidakseimbangan dalam data kelas. Presisi mengukur seberapa banyak prediksi positif yang benar dibandingkan dengan seluruh prediksi positif, penting untuk memastikan bahwa model tidak menghasilkan banyak false positives. Recall mengukur seberapa banyak instance positif yang berhasil ditangkap oleh model dari seluruh data positif yang ada, yang sangat penting dalam konteks klasifikasi ayat-ayat Al-Qur'an, di mana kita ingin menangkap sebanyak mungkin ayat yang relevan dengan tema tertentu. F1-score, yang merupakan rata-rata harmonis dari presisi dan recall, memberikan gambaran yang lebih seimbang tentang kinerja model, dan dalam penelitian ini, SVM dengan stopwords NLTK memberikan nilai F1-score tertinggi (0,5438), yang menunjukkan keseimbangan yang baik antara presisi dan recall. SVM terbukti lebih unggul dalam hal klasifikasi teks Al-Qur'an, dengan penggunaan stopwords NLTK memberikan hasil yang paling seimbang antara presisi dan recall, sehingga mampu memberikan model klasifikasi yang lebih akurat dan efisien. Di sisi lain, BPNN masih membutuhkan peningkatan dalam hal penanganan data dan pengaturan parameter agar dapat mencapai kinerja yang optimal dalam klasifikasi teks keagamaan yang kompleks seperti Al-Qur'an. Penelitian ini juga menunjukkan pentingnya pemilihan daftar stopwords yang tepat, karena pengaruhnya terhadap performa model klasifikasi sangat signifikan, terutama dalam konteks teks yang memiliki kedalaman makna dan kompleksitas semantik seperti Al-Qur'an.

5. KESIMPULAN

Penelitian ini mengevaluasi pengaruh penggunaan varian daftar stopwords terhadap performa klasifikasi teks Al-Qur'an dengan menggunakan dua algoritma pembelajaran mesin, SVM dan BPNN. Hasil penelitian menunjukkan bahwa SVM dengan stopwords NLTK memberikan kinerja terbaik dengan mencapai akurasi 0,5849, presisi 0,6646, recall 0,5849, dan F1-score 0,5438. Sementara itu, BPNN menunjukkan performa yang lebih rendah di seluruh metrik yang digunakan, dengan akurasi tertinggi hanya 0,4292 menggunakan stopwords NLTK. SVM terbukti lebih efektif dalam menangani teks berdimensi tinggi dan kompleksitas semantik dalam teks Al-Qur'an, sementara BPNN masih membutuhkan perbaikan lebih lanjut dalam hal parameter dan representasi data untuk dapat mencapai performa optimal. Keterbatasan utama dalam penelitian ini adalah penggunaan daftar stopwords yang terbatas dan penerapan hanya dua algoritma klasifikasi, yaitu SVM dan BPNN. Penelitian selanjutnya disarankan untuk memperluas variasi daftar stopwords, termasuk varian yang lebih kontekstual dan relevan dengan teks keagamaan seperti stopwords dalam bahasa Arab, atau bahkan mengembangkan daftar stopwords yang lebih spesifik untuk teks-teks agama. Selain itu, penting untuk mengeksplorasi penggunaan algoritma lain, seperti Random Forest atau XGBoost, yang mungkin memiliki keunggulan dalam mengatasi data yang tidak seimbang dan memiliki lebih banyak fitur. Selain itu, penelitian mendatang bisa mengintegrasikan pendekatan deep learning lebih lanjut, termasuk pre-trained models seperti BERT atau RoBERTa, yang telah terbukti efektif dalam menangani konteks semantik yang kompleks dalam teks keagamaan. Penggunaan teknik transfer learning juga dapat menjadi arah yang baik untuk meningkatkan akurasi klasifikasi, mengingat model tersebut telah dilatih pada data teks besar dan dapat menangkap konteks yang lebih luas.

DAFTAR PUSTAKA

- [1] M. R. Ramadhan dan N. Nasrulloh, "Pengaruh Konsep Keadilan Dalam Al Qur'an Dan Relevansinya Dalam Hukum Manusia," *Jimr*, vol. 2, no. 11, hlm. 132–139, 2024, doi: 10.62504/jimr972.

-
- [2] D. F. A. Sakhi, P. A. I. Amin, dan K. Kurniati, "Etika Politik Islam Dalam Masyarakat Kontemporer: Perspektif Al-Mawardi," *J. Ilm. Falsafah J. Kaji. Filsafat Teol. Dan Hum.*, vol. 10, no. 2, hlm. 95–106, 2024, doi: 10.37567/jif.v10i2.3052.
- [3] H. Maburur dan S. Abas, "Hermeneutik Sebagai Tawaran Metodologis Dalam Menafsirkan Al-Quran Yang Diperdebatkan," *Setyaki J. Studi Keagamaan Islam*, vol. 1, no. 1, hlm. 78–89, 2023, doi: 10.59966/setyaki.v1i1.251.
- [4] Abd. Muid N. Amiril Ahmad, "Pendekatan Konstruktivis-Interpretis (Hermeneutik) Sebagai Metode Penafsiran," *Al Dhikra J. Studi Qur Dan Hadis*, vol. 2, no. 2, hlm. 169–182, 2022, doi: 10.57217/aladhikra.v2i2.779.
- [5] A. Ghoni dan G. Saloom, "Idealisasi Metode Living Qur'an," *Himmah J. Kaji. Islam Kontemporer*, vol. 5, no. 2, hlm. 413, 2021, doi: 10.47313/jkik.v5i2.1510.
- [6] G. Muradlo, A. K. Khotimah, D. Alawiyah, E. Elviana, Y. C. Nugroho, dan Z. Ayuni, "Mendalami Living Qur'an: Analisis Pendidikan Dalam Memahami Dan Menghidupkan Al-Qur'an," *Pandu*, vol. 1, no. 2, hlm. 112–118, 2023, doi: 10.59966/pandu.v1i2.206.
- [7] I. Irawan, "Al-Tawassut Waal-I'tidal: Menjawab Tantangan Liberalisme Dan Konservatisme Islam," *Afkaruna*, vol. 14, no. 1, 2018, doi: 10.18196/aiijis.2018.0080.49-74.
- [8] A. Fauzan, D. Muriyatmoko, dan S. N. Utama, "Penerapan Teknologi Augmented Reality Pada Media Pembelajaran Bahasa Arab: Durus Al-Lughah Jilid 1," *Elem. Sch. Educ. J. J. Pendidik. Dan Pembelajaran Sekol. Dasar*, vol. 4, no. 1, hlm. 63, 2020, doi: 10.30651/else.v4i1.4379.
- [9] A. Salida dan Z. Zulpina, "Keistimewaan Bahasa Arab Sebagai Bahasa Al-Quran Dan Ijtihadiyyah," *Sathar*, vol. 1, no. 1, hlm. 23–33, 2023, doi: 10.59548/js.v1i1.40.
- [10] A. Atabik, "Teori Makna Dalam Struktur Linguistik Arab Perspektif Mufasir Masa Klasik," *J. Theol.*, vol. 31, no. 1, hlm. 65–86, 2020, doi: 10.21580/teo.2020.31.1.5631.
- [11] I. R. A. S dan Z. Zuraidah, "Optimasi Media Pembelajaran Untuk Meningkatkan Keterampilan Menulis Huruf Hijaiyah Pada Taman Pendidikan Al-Qur'an (TPQ)," *Kontribusi J. Penelit. Dan Pengabd. Kpd. Masy.*, vol. 2, no. 1, hlm. 46–55, 2021, doi: 10.53624/kontribusi.v2i1.62.
- [12] E. Y. Manurung dan S. Nasution, "Penggunaan Kitab Al-Qira'atur Ar-Rasyidah Dalam Melatih Kemampuan Penerjemahan Siswa Kelas XII Agama MAS Tafhizil Qur'an Islamic Center Medan," *J. Educ. Res.*, vol. 4, no. 3, hlm. 1212–1221, 2023, doi: 10.37985/jer.v4i3.377.
- [13] R. Muhammad dkk., "Literasi Hukum: Pembagian Warisan Berdasarkan Kaidah Hukum Islam," *Joehr*, vol. 1, no. 1, hlm. 28–36, 2024, doi: 10.34304/joehr.v1i1.207.
- [14] M. Ridwan, "Membuka Wawasan Keislaman: Kebermaknaan Bahasa Arab Dalam Pemahaman Islam," *Jazirah J. Perad. Dan Kebud.*, vol. 4, no. 2, hlm. 102–115, 2023, doi: 10.51190/jazirah.v4i2.100.
- [15] R. Huang, "Design and Implementation of English Writing Aids Based on Natural Language Processing," *Jes*, vol. 20, no. 6s, hlm. 2178–2187, 2024, doi: 10.52783/jes.3132.
- [16] A. H. Nasution dan A. Onan, "ChatGPT Label: Comparing the Quality of Human-Generated and LLM-Generated Annotations in Low-Resource Language NLP Tasks," *Ieee Access*, vol. 12, hlm. 71876–71900, 2024, doi: 10.1109/access.2024.3402809.
- [17] J. Peng, "A Comprehensive Review of the Application of NLP Technology in Language Learning," *Appl. Comput. Eng.*, vol. 92, no. 1, hlm. 163–168, 2024, doi: 10.54254/2755-2721/92/20241735.
- [18] Y.-X. Yu, Y. Guan, dan Y. Hu, "Natural Language Processing Applications in Social Network Analysis: A Data Mining Approach," *J. Phys. Conf. Ser.*, vol. 2813, no. 1, hlm. 012009, 2024, doi: 10.1088/1742-6596/2813/1/012009.
- [19] A. M. Wahid, T. Turino, K. A. Nugroho, T. S. Maharani, D. Darmono, dan F. S. Utomo, "Optimasi Logistic Regression dan Random Forest untuk Deteksi Berita Hoax Berbasis TF-IDF," *J. Pendidik. Dan Teknol. Indones.*, vol. 4, no. 8, Art. no. 8, 2024, doi: 10.52436/1.jpti.602.
- [20] S. Khan, I. Qasim, W. Khan, K. Aurangzeb, J. A. Khan, dan M. S. Anwar, "A Novel Transformer Attention-based Approach for Sarcasm Detection," *Expert Syst.*, vol. 42, no. 1, 2024, doi: 10.1111/exsy.13686.
- [21] M. Fajri, "Dynamics of the Study of the Quran in Indonesia: Language and Paradigm," *Islam Transform. J. Islam. Stud.*, vol. 5, no. 1, hlm. 59, 2021, doi: 10.30983/it.v5i1.4130.

- [22] D. AbuZeina dan T. M. Abdalbaset, "Exploring the Performance of Tagging for the Classical and the Modern Standard Arabic," *Adv. Fuzzy Syst.*, vol. 2019, hlm. 1–10, 2019, doi: 10.1155/2019/6254649.
- [23] H. M. Abdelaal, A. M. Ahmed, W. Ghribi, dan H. Youness, "Knowledge Discovery in the Hadith According to the Reliability and Memory of the Reporters Using Machine Learning Techniques," *Ieee Access*, vol. 7, hlm. 157741–157755, 2019, doi: 10.1109/access.2019.2944118.
- [24] N. J. Ibrahim, M. Y. I. Idris, Z. M. Yusoff, N. N. A. Rahman, dan M. I. Dien, "Robust Feature Extraction Based on Spectral and Prosodic Features for Classical Arabic Accents Recognition," *Malays. J. Comput. Sci.*, hlm. 46–72, 2019, doi: 10.22452/mjcs.sp2019no3.4.
- [25] A. M. Alashqar, "A Classification of Quran Verses Using Deep Learning," *Int. J. Comput. Digit. Syst.*, vol. 15, no. 1, hlm. 1041–1053, 2024, doi: 10.12785/ijcds/160176.
- [26] W. Darmalaksana, "Studi Penggunaan Analisis Pendekatan Ilmu-Ilmu Sosial Dalam Penelitian Hadis Metode Syarah," *Khazanah Sos.*, vol. 2, no. 3, hlm. 155–166, 2020, doi: 10.15575/ks.v2i3.9599.
- [27] F. Fitriani, "Fungsi, Kategori, Dan Peran Sintaksis Bahasa Arab: Perspektif Linguistik Modern," *El-Fakhr*, vol. 1, no. 1, hlm. 180–212, 2023, doi: 10.46870/iceil.v1i1.473.
- [28] A. Hadiyanto, C. Samitri, dan S. M. Ulfah, "Model Pembelajaran Bahasa Arab Multiliterasi Berbasis Kearifan Lokal Dan Moderasi Islam Di Perguruan Tinggi Negeri," *Hayula Indones. J. Multidiscip. Islam. Stud.*, vol. 4, no. 1, hlm. 117–140, 2020, doi: 10.21009/004.01.07.
- [29] S. Syandri, E. Halid, dan A. R. Sakka, "K.H. Lanre Said View in Takbir Zawaid (A Criticical Hadith)," *Al-Risal.*, vol. 14, no. 1, hlm. 194–210, 2023, doi: 10.34005/alrisalah.v14i1.2437.
- [30] A. Abdullahi, N. Samsudin, A. Mustapha, dan S. Khalid, "Automating Quranic Verses Labeling Using Machine Learning Approach," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 16, no. 2, hlm. 925, 2019, doi: 10.11591/ijeecs.v16.i2.pp925-931.
- [31] T. M. T. Sembok dan B. M. Abuata, "Arabic Stop Words for Information Retrieval Systems," *Int. J. Relig.*, vol. 6, no. 1, hlm. 121–127, 2025, doi: 10.61707/k7rnm813.
- [32] A. A. Badawy, "Topic Discovery in the Digital Quran: A Text Mining Approach," *J. Inf. Syst. Eng. Manag.*, vol. 10, no. 18s, hlm. 642–649, 2025, doi: 10.52783/jisem.v10i18s.2976.
- [33] I. Budiman, M. R. Faisal, dan D. T. Nugrahadi, "Studi Ekstraksi Fitur Berbasis Vektor Word2Vec pada Pembentukan Fitur Berdimensi Rendah," *J. Komputasi*, vol. 8, no. 1, Art. no. 1, Apr 2020, doi: 10.23960/komputasi.v8i1.2517.
- [34] Y. Purwati, F. S. Utomo, N. Trinarsih, dan H. Hidayatulloh, "Feature Selection Technique to Improve the Instances Classification Framework Performance for Quran Ontology," *Int. J. Inform. Vis.*, vol. 7, no. 2, hlm. 615–620, 2023, doi: 10.30630/joiv.7.2.1195.
- [35] M. Hasanudin, S. Dwiasnati, dan W. Gunawan, "Pelatihan Datascience pada Pra-Pemrosesan Data untuk Siswa SMK Media Informatika - Jakarta:," *J. Pengabd. Pada Masy.*, vol. 9, no. 4, Art. no. 4, Nov 2024, doi: 10.30653/jppm.v9i4.921.
- [36] P.-O. Côté, A. Nikanjam, N. Ahmed, D. Humeniuk, dan F. Khomh, "Data cleaning and machine learning: a systematic literature review," *Autom. Softw. Eng.*, vol. 31, no. 2, hlm. 54, 2024, doi: 10.1007/s10515-024-00453-w.
- [37] F. Deny Elfianita, O. Virgantara Putra, T. Harmini, dan A. Trisnani, "Klasifikasi Teks Terjemahan Ayat Al-Qur'an dalam Bahasa Indonesia dengan Machine Learning Classification of Translated Texts of Al-Qur'an Verses in Indonesian with Machine Learning," *Semin. Nas. Has. Penelit. Pengabd. Masy. Bid. Ilmu Komput.*, hlm. 49–57, 2022.
- [38] O. Kononova, T. He, H. Huo, A. Trewartha, E. A. Olivetti, dan G. Ceder, "Opportunities and challenges of text mining in aterials research," *iScience*, vol. 24, no. 3, hlm. 102155, 2021, doi: 10.1016/j.isci.2021.102155.
- [39] C. Chairunnisa, I. Ernawati, dan M. M. Santoni, "Klasifikasi Sentimen Ulasan Pengguna Aplikasi PeduliLindungi Di Google Play Menggunakan Algoritma Support Vector Machine Dengan Seleksi Fitur Chi-Square," *Inform. J. Ilmu Komput.*, vol. 18, no. 1, hlm. 69, 2022, doi: 10.52958/iftk.v17i4.4594.
- [40] M. Tsani, A. P. G. Rupaka, L. Asmoro, dan B. Pradana, "Analisis Sentimen Review Transportasi Menggunakan Algoritma Support Vector Machine Berbasis Chi Square," *Smart Comp Jurnalnya Orang Pint. Komput.*, vol. 9, no. 1, hlm. 35–39, 2020, doi: 10.30591/smartcomp.v9i1.1817.

-
- [41] T. A. Y. Siswa, "Komparasi Optimasi Chi-Square, CFS, Information Gain Dan ANOVA Dalam Evaluasi Peningkatan Akurasi Algoritma Klasifikasi Data Performa Akademik Mahasiswa," *Inform. Mulawarman J. Ilm. Ilmu Komput.*, vol. 18, no. 1, hlm. 62, 2023, doi: 10.30872/jim.v18i1.11330.
- [42] A. Rahmadeyan dan M. Mustakim, "Seleksi Fitur Pada Supervised Learning: Klasifikasi Prestasi Belajar Mahasiswa Saat Dan Pasca Pandemi COVID-19," *J. Nas. Teknol. Dan Sist. Inf.*, vol. 9, no. 1, hlm. 21–32, 2023, doi: 10.25077/teknosi.v9i1.2023.21-32.
- [43] J. Suntoro dan C. N. Indah, "Average Weight Information Gain Untuk Menangani Data Berdimensi Tinggi Menggunakan Algoritma C4.5," *J. Buana Inform.*, vol. 8, no. 3, 2017, doi: 10.24002/jbi.v8i3.1315.
- [44] C. F. Suharno, M. A. Fauzi, dan R. S. Perdana, "Klasifikasi Teks Bahasa Indonesia Pada Dokumen Pengaduan Sambat Online Menggunakan Metode K-Nearest Neighbors Dan Chi-Square," *Syst. Inf. Syst. Inform. J.*, vol. 3, no. 1, hlm. 25–32, 2017, doi: 10.29080/systemic.v3i1.191.
- [45] S. Mutmainah, "Kemungkinan Depresi Dari Postingan Pada Sosial Media," *Snati*, vol. 1, no. 2, 2022, doi: 10.20885/snati.v1i2.11.
- [46] I. F. Yuliati, S. Wulandary, dan P. R. Sihombing, "Penerapan Metode SVM Dan BPNN Dalam Pengklasifikasian PUS Di Jawa Barat," *J. Stat. Dan Apl.*, vol. 4, no. 1, hlm. 23–34, 2020, doi: 10.21009/jsa.04103.
- [47] T. Sugihartono dan R. R. C. Putra, "Penerapan Metode Support Vector Machine Dalam Classifikasi Ulasan Pengguna Aplikasi Mobile Jkn," *SKANIKA Sist. Komput. Dan Tek. Inform.*, vol. 7, no. 2, hlm. 144–153, 2024, doi: 10.36080/skanika.v7i2.3193.
- [48] S. F. Hussain, H. Z.-U.-D. Babar, A. Khalil, R. M. Jillani, M. Hanif, dan K. Khurshid, "A Fast Non-Redundant Feature Selection Technique for Text Data," *IEEE Access*, vol. 8, hlm. 181763–181781, 2020, doi: 10.1109/ACCESS.2020.3028469.
- [49] R. K. Putri dan M. Athoillah, "Support Vector Machine Untuk Identifikasi Berita Hoax Terkait Virus Corona (Covid-19)," *J. Inform. J. Pengemb. It*, vol. 6, no. 3, hlm. 162–167, 2021, doi: 10.30591/jpit.v6i3.2489.
- [50] S. Chukhanov dan N. Kairbekov, "The importance of a semantic approach in understanding the texts of the Holy Quran and Sunnah," *Pharos J. Theol.*, vol. 105, no. 3, Jun 2024, doi: 10.46222/pharosjot.105.36.