

Implementasi Algoritma BERT untuk Question and Answer System Terkait Hadist dalam Bentuk Virtual Youtuber

Moch. Arsyil Albany¹

Ichsan Taufik²

Ichsan Budiman³

^{1,2,3}Teknik Informatika, Universitas Islam Negeri Sunan Gunung Djati Bandung, Jalan AH Nasution No. 105, Cipadung, Kec. Cibiru, Kota Bandung, Jawa Barat, Indonesia

¹moch.arsyil@gmail.com, ²ichsan@uinsgd.ac.id, ³ichsanbudiman@uinsgd.ac.id

*Penulis Korespondensi:

Moch. Arsyil Albany

moch.arsyil@gmail.com

Abstrak

Di tengah pesatnya perkembangan teknologi digital, integrasi ajaran Islam dengan teknologi canggih, khususnya dalam pendidikan, menjadi sangat penting. Salah satu tantangan yang dihadapi dalam pendidikan Islam adalah penyampaian pengetahuan agama, terutama terkait dengan hadits, yang seringkali sulit diakses atau dipahami oleh generasi muda. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan sistem tanya jawab (QnA) Islami berbasis Virtual YouTuber (VTuber) yang menggunakan kecerdasan buatan untuk memudahkan pemahaman hadits. Sistem ini memanfaatkan algoritma IndoBERT-SQuAD untuk pemrosesan bahasa alami dalam menangani pertanyaan-pertanyaan mengenai hadits. Metodologi yang digunakan adalah prototipe, yang meliputi tahapan analisis, desain, implementasi, dan evaluasi. Evaluasi sistem dilakukan dengan menggunakan metrik confidence score dan F1-score untuk mengukur akurasi dan performa jawaban yang diberikan oleh model. Hasil penelitian menunjukkan bahwa model memiliki kinerja yang sangat baik dengan nilai F1-score sebesar 0.96875. Meskipun demikian, sistem masih menghadapi tantangan dalam memberikan jawaban panjang yang lebih akurat dan komprehensif. Penelitian ini berkontribusi pada penerapan teknologi dalam pendidikan Islam, dengan menawarkan solusi praktis dalam mempermudah akses pengetahuan mengenai hadits bagi generasi muda.

Kata Kunci: IndoBERT-SQuAD; Pendidikan Islam; Pemrosesan Bahasa Alami; QnA Hadits; Virtual YouTuber.

Abstract

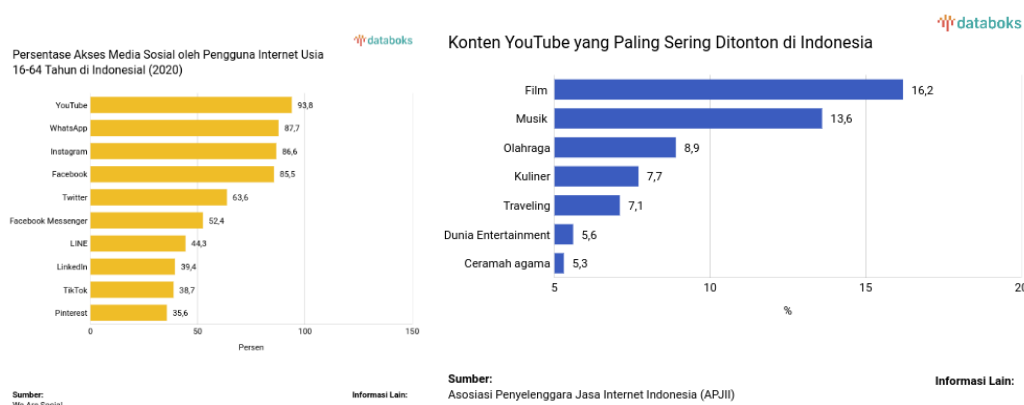
Amid the rapid advancement of digital technology, integrating Islamic teachings with advanced technology, especially in education, has become crucial. One of the challenges in Islamic education is the delivery of religious knowledge, particularly regarding hadith, which is often difficult to access or understand for the younger generation. Therefore, this research aims to develop an Islamic Question and Answer (QnA) system based on Virtual YouTuber (VTuber) using artificial intelligence to facilitate the understanding of hadith. The system utilizes the IndoBERT-SQuAD algorithm for natural language processing in handling questions related to hadith. The methodology used is prototyping, which includes the stages of analysis, design, implementation, and evaluation. The system's evaluation is conducted using confidence score and F1-score metrics to measure the accuracy and performance of the answers provided by the model. The research results show that the model performs excellently with an F1-score of 0.96875. However, the system still faces challenges in providing more accurate and comprehensive long answers. This research contributes to the application of technology in Islamic education, offering a practical solution to make knowledge about hadith more accessible and engaging for the younger generation.

Keywords: IndoBERT-SQuAD; Islamic Education; Natural Language Processing; QnA Hadith; Virtual YouTuber.

1. Pendahuluan

Pada era digital saat ini, akses terhadap informasi telah menjadi semakin mudah dan cepat berkat perkembangan teknologi. Salah satu bentuk perkembangan tersebut adalah dalam bidang pemrosesan bahasa alami (*Natural Language Processing/NLP*), di mana algoritma-algoritma baru terus dikembangkan untuk meningkatkan kemampuan mesin dalam memahami dan merespons teks manusia [1]. Sebagai contoh, pengembangan aplikasi berbasis AR untuk pembelajaran tajwid, seperti "Mahwid," berhasil memberikan media pembelajaran yang inovatif dengan tingkat kepuasan pengguna mencapai 94%. Inovasi ini menunjukkan bahwa teknologi dapat digunakan secara efektif untuk mendukung pembelajaran berbasis agama, khususnya dalam membantu

siswa memahami materi yang kompleks seperti tajwid[2]. Seiring dengan itu, platform seperti YouTube juga semakin dimanfaatkan sebagai media pembelajaran mandiri, termasuk dalam konteks pendidikan agama Islam. Penggunaan YouTube memberikan peluang besar untuk menyampaikan konten Islami secara luas, meskipun dampaknya terhadap sikap remaja Muslim sangat bergantung pada jenis video yang mereka tonton[3]. Di bawah ini merupakan gambar Distribusi Penggunaan Media Sosial dan Konten di Indonesia:



Gambar 1. Distribusi Penggunaan Media Sosial dan Konten di Indonesia (2020)[4].

Menurut DataBoks, pada tahun 2020, persentase akses media sosial oleh pengguna internet usia 16-64 tahun di Indonesia menunjukkan bahwa YouTube menduduki peringkat pertama dengan 93,8% pengguna aktif. Namun, menurut APJII (Aplikasi Penyedia Jasa Internet Indonesia), konten agama berada di peringkat paling bawah dengan hanya 5,3% dari total konten yang paling sering ditonton di Indonesia.

Di tengah tren ini, artikel ini bertujuan untuk memanfaatkan fenomena tersebut dengan membangun Virtual YouTuber AI (*Artificial Intelligence*) yang berfokus pada konten Islami yang berkaitan dengan hadits. Hal ini dilakukan dengan tujuan menarik minat anak muda terhadap Islam, menyadari bahwa mereka sering terpapar berbagai konten di platform media sosial dan situs streaming[5].

Keputusan untuk menggunakan Virtual YouTuber (*VTuber*) didasarkan pada fakta bahwa Indonesia menempati peringkat pertama sebagai negara dengan komunitas *VTuber* terbesar dan terbanyak di Asia Tenggara, dengan lebih dari 20 juta subscriber[6]. Menurut data, sekitar 66.92% penonton *VTuber* berusia antara 18-25 tahun, sementara 28.72% berusia di bawah 18 tahun[7].

Penggunaan *Vtuber* bukan hanya karena popularitasnya di kalangan audiens muda, tetapi juga karena kemampuan *VTuber* untuk menyediakan pengalaman belajar yang lebih interaktif dan menarik dibandingkan media konvensional. Platform seperti YouTube memungkinkan *VTuber* untuk berinteraksi langsung dengan audiens dalam waktu nyata, menciptakan hubungan yang lebih personal dan lebih mendalam. Hal ini sangat relevan di tengah tingginya persentase pengguna media sosial di Indonesia, terutama di kalangan usia 16-25 tahun, yang merupakan kelompok usia utama penonton *VTuber*. Dibandingkan dengan media tradisional seperti ceramah atau video pembelajaran, *VTuber* menawarkan nuansa yang lebih modern, dengan elemen hiburan dan interaktivitas yang lebih mudah diterima oleh generasi muda yang terbiasa dengan teknologi dan konten digital. Pendekatan ini memberikan kesempatan untuk menyampaikan ajaran Islam melalui karakter yang lebih menarik dan kreatif, membuat pesan-pesan Islami lebih mudah dicerna dan diterima oleh audiens muda.

Dalam konteks keilmuan Islami, hadits merupakan sumber utama dalam memahami ajaran Islam yang bersumber dari ucapan, perbuatan, dan persetujuan Nabi Muhammad SAW[8]. Hadist

memiliki banyak kandungan yang mencakup berbagai aspek kehidupan, mulai dari ibadah, akhlak, hingga hukum-hukum Islam. Namun, keberadaan hadist yang luas dan beragam seringkali membuat pencarian informasi terkait hadist menjadi sulit bagi sebagian individu, terutama bagi mereka yang belum memahami bahasa Arab dengan baik[9].

Untuk mengatasi tantangan tersebut, pengembangan sistem tanya jawab (Question and Answering/QnA) yang dapat mengakomodasi pertanyaan terkait hadist menjadi penting[10]. Salah satu algoritma yang terkenal dalam NLP adalah Bidirectional Encoder Representations from Transformers (BERT), yang dikenal memiliki kemampuan yang baik dalam memahami konteks teks[11]. Implementasi algoritma BERT dalam sistem QnA Islami dapat membantu individu untuk mendapatkan jawaban yang tepat dan relevan terkait pertanyaan mereka yang bersumber dari hadist.

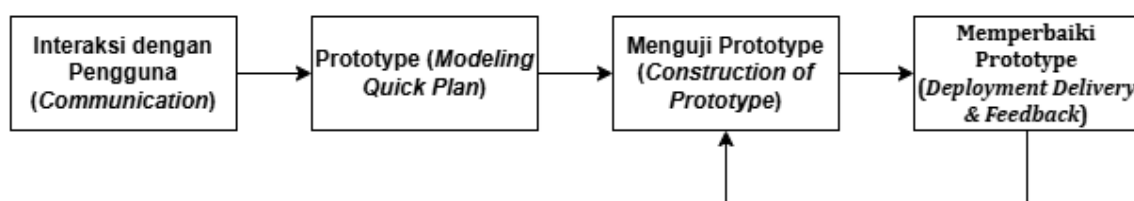
Selain itu, pemilihan algoritma BERT (Bidirectional Encoder Representations from Transformers) didasarkan pada akurasi tinggi yang dimilikinya dalam pemodelan pertanyaan dan jawaban (QnA), dengan kisaran akurasi antara 80,8% hingga 88,5%. Perbandingan dengan algoritma lain menunjukkan bahwa BERT memiliki performa yang lebih baik, dengan akurasi yang jauh lebih tinggi daripada algoritma seperti XGBoost (70%), Random Forest (67%), Support Vector Machine (63%), dan Principal Component Analysis (59%)[12]. Oleh karena itu, BERT dipilih sebagai algoritma utama dalam pengembangan chatbot ini untuk memastikan kualitas dan keakuratan konten yang disampaikan kepada pengguna.

Meskipun BERT telah terbukti efektif dalam berbagai aplikasi NLP, terutama dalam konteks Question and Answering (QnA), penggunaan algoritma ini dalam sistem pendidikan agama, khususnya dalam konteks hadits, masih tergolong minim. Beberapa penelitian yang mengintegrasikan BERT dalam pendidikan agama cenderung terbatas pada analisis teks umum, sementara pemanfaatannya dalam memahami teks agama yang sangat spesifik dan kaya makna, seperti hadits, belum banyak dieksplorasi. Gap ini menunjukkan pentingnya penelitian lebih lanjut mengenai adaptasi dan pengembangan algoritma BERT dalam sistem pendidikan agama Islam, yang dapat membantu menjembatani kesulitan yang dihadapi oleh masyarakat dalam memahami hadits secara lebih mendalam dan kontekstual.

Dengan memanfaatkan teknologi ini, diharapkan akan memudahkan masyarakat umum, baik yang berpengetahuan tinggi maupun yang awam dalam bidang keilmuan Islami, untuk mengakses informasi dan pemahaman yang lebih baik terkait hadist. Implementasi ini juga dapat menjadi langkah awal dalam memperluas aplikasi teknologi dalam konteks agama, menggabungkan kearifan tradisional dengan kemajuan teknologi modern. Selain itu, dengan adanya sistem QnA yang baik, diharapkan dapat membantu meningkatkan pemahaman dan keterlibatan masyarakat terhadap nilai-nilai Islam dalam kehidupan sehari-hari[13].

2. Metode Penelitian

Metode yang akan digunakan dalam penelitian ini adalah *prototype* karena metodologi ini memungkinkan pengembangan sistem QnA hadist dalam bentuk *Vtuber* secara bertahap dan terukur, dengan fokus pada pengujian, evaluasi, dan peningkatan berkelanjutan. Berikut ini merupakan gambar dari alur penelitian:



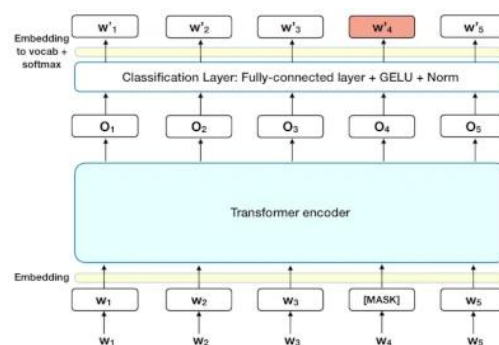
Gambar 2. Alur Penelitian

Interaksi dengan Pengguna (*Communication*)

Pada tahap ini, dilakukan analisis mendalam untuk memahami kebutuhan pengguna terkait sistem QnA hadist berbasis *Vtuber*. Penelitian difokuskan pada jenis-jenis pertanyaan yang kemungkinan diajukan oleh pengguna serta cara mereka berinteraksi dengan sistem. Dataset hadits diambil dari buku *Kuliah Ilmu Hadis 1* [15], yang menyediakan informasi mendalam tentang ilmu hadits, seperti sanad, matan, dan klasifikasi hadits. Buku ini dipilih karena memiliki referensi yang terstandar. Sedangkan dataset untuk pengenalan *VTuber* dibuat berdasarkan hasil penelitian dan observasi langsung terhadap gaya interaksi *VTuber* populer.

Prototype (*Modeling Quick Plan*)

Prototipe awal sistem QnA dikembangkan dengan memanfaatkan algoritma BERT, khususnya model *IndoBERT-SQuAD* yang dirancang untuk bahasa Indonesia. Berikut ini merupakan gambar dari arsitektur algoritma BERT:



Gambar 3. Arsitektur BERT[14].

Implementasi algoritma BERT pada sistem ini mengikuti langkah-langkah utama berikut:

1. Tokenisasi Input Teks

Input berupa pertanyaan dan konteks diubah menjadi token menggunakan tokenizer bawaan dari model pre-trained. Proses ini bertujuan untuk membagi teks menjadi unit-unit terkecil yang dapat dikenali oleh model.

2. Konversi ke ID Numerik

Token yang dihasilkan dikonversi menjadi ID numerik yang sesuai dengan format input model.

3. Pemrosesan oleh Model Pre-trained

Model menerima ID numerik sebagai input dan menghasilkan dua output utama:

- **Start Logits:** Probabilitas posisi awal jawaban dalam konteks.
- **End Logits:** Probabilitas posisi akhir jawaban dalam konteks.

4. Pemilihan Jawaban

Jawaban ditentukan berdasarkan token dengan probabilitas tertinggi untuk posisi awal dan akhir jawaban dalam teks konteks.

5. Decode Token ke Teks

Token yang terpilih didekode kembali menjadi teks asli menggunakan tokenizer, menghasilkan jawaban akhir yang dapat dipahami oleh pengguna.

Langkah-langkah ini memungkinkan model untuk memahami pertanyaan dan konteks secara mendalam, sehingga dapat memberikan jawaban yang relevan dan akurat.

Pengembangan prototipe dilakukan dengan menggunakan *Python* pada platform *Google Colab*, yang mendukung pengolahan data besar dan pemodelan berbasis AI. Untuk mendukung tampilan yang menarik dan interaktif, desain visual avatar *VTuber* dirancang menggunakan *Photoshop*. Hasil desain tersebut kemudian diintegrasikan ke dalam sistem streaming menggunakan aplikasi *OBS* untuk menampilkan avatar *VTuber* secara real-time di YouTube..

Menguji Prototype (*Construction of Prototype*)

Prototipe yang telah dibuat diuji menggunakan *Confident Score* sebagai alat evaluasi utama untuk mengukur kinerja sistem secara kuantitatif. *Confident Score* memberikan penilaian terukur terhadap keandalan sistem dalam memberikan jawaban yang relevan, akurat, dan konsisten. Pada dasarnya, pengujian model BERT Squad biasanya dilakukan menggunakan *confidence score*, yang merepresentasikan keyakinan model terhadap jawaban yang diberikan. Namun, pada penelitian kali ini, saya juga akan menguji kinerja sistem menggunakan *F1-score* untuk mengevaluasi keseimbangan antara *precision* (keakuratan jawaban) dan *recall* (kelengkapan jawaban).

Kombinasi *confidence score* dan *F1-score* memungkinkan evaluasi yang lebih menyeluruh terhadap performa sistem, di mana *confidence score* memberikan wawasan tentang keandalan respons, sementara *F1-score* membantu mengidentifikasi kelemahan sistem dalam hal akurasi atau kelengkapan jawaban. Dengan mengintegrasikan kedua metrik ini, penelitian dapat mengidentifikasi area perbaikan pada model, baik dari segi keyakinan model terhadap jawabannya maupun kemampuan untuk menjawab berbagai jenis pertanyaan secara lengkap dan akurat. Penggunaan metrik evaluasi ini mencerminkan pendekatan yang sistematis dan menyeluruh dalam menilai performa model, yang pada gilirannya dapat mendukung pengembangan sistem yang lebih optimal dan adaptif terhadap kebutuhan pengguna.

Memperbaiki Prototype (*Deployment Delivery & Feedback*)

Berdasarkan hasil evaluasi dan umpan balik yang diperoleh, prototipe diperbaiki untuk mengatasi kelemahan yang teridentifikasi. Proses ini melibatkan penambahan dataset yang lebih luas untuk meningkatkan performa sistem, serta pengujian lanjutan untuk mengukur kemampuan sistem dalam menangani pertanyaan yang lebih kompleks. Setelah diperbaiki, sistem disiapkan untuk disebarkan kepada pengguna atau komunitas yang relevan. Dokumentasi dan panduan penggunaan disediakan untuk memastikan sistem mudah digunakan oleh berbagai kalangan...

3. Hasil

Metode yang akan digunakan dalam penelitian ini adalah prototipe karena metodologi ini memungkinkan pengembangan sistem QnA hadist dalam bentuk *vtuber* secara bertahap dan terukur, dengan fokus pada pengujian, evaluasi, dan peningkatan berkelanjutan.

Hasil Interaksi dengan Pengguna (*Communication*)

Sebagai langkah awal dalam pengembangan sistem QnA hadist berbasis *Vtuber*, proses interaksi dengan pengguna dimulai dengan merancang dataset yang relevan dan sesuai kebutuhan pengguna.

- Dataset untuk hadits: Diambil dari buku Kuliah Ilmu Hadis 1[15], yang menyediakan informasi mendalam terkait ilmu hadis seperti sanad, matan, dan klasifikasi hadis.
- Dataset untuk pengenalan *Vtuber*: Diambil dari hasil research, termasuk menonton dan mempelajari gaya interaksi *Vtuber*, seperti cara mereka menyapa, menjawab pertanyaan ringan. Pertanyaan-pertanyaan ini membantu memperkenalkan *VTuber* kepada audiens dengan cara yang menarik dan interaktif, serta memberikan pengalaman yang lebih personal. Berikut adalah tabel pertanyaan dari dataset;

Tabel 1. Tabel Pertanyaan

No	Pertanyaan
1	Assalamualaikum
2	Siapa nama kamu?,
3	Berapa umur kamu?
4	Siapa penciptamu?
5	Apa itu hadits?
6	Apa peranan hadits dalam islam?
7	Hadits terdiri dari berapa bagian?
8	Apa itu sanad?
9	Apa kegunaan sanad?
10	Apa itu perawi?
11	Apa itu matan?
12	Ada berapa jenis hadits?
13	Ada berapa jenis-jenis hadits berdasarkan isinya?
14	Apa itu hadits qudsi?
15	Apa itu hadits nabawi?
16	Apa itu takhrij?
17	Ada berapa jenis-jenis hadits berdasarkan kualitasnya?
18	Apa itu hadits shahih?
19	Apa ciri-ciri hadits shahih?
20	Apa itu hadits hasan?
21	Apa itu hadits dhoif?
22	Apa itu ilmu mustalaf hadits?
23	Apa itu hadits maudhu?
24	Ada berapa jenis sanad?
25	Apa itu sanad muttaqil?
26	Apa itu sanad muntaqi?
27	Apa itu hadits marfu?
28	Apa itu hadits mursal?
29	Apa itu hadits musnad?
30	Sebutkan kitab-kitab hadits
31	Berikan contoh hadits kitab bukhari!
32	Berikan contoh hadits kitab muslim!

Tabel ini berisi daftar pertanyaan yang digunakan sebagai data input dalam sistem QnA berbasis Vtuber. Pertanyaan-pertanyaan ini mencakup salam, pengenalan diri, dan pertanyaan terkait ilmu hadits, seperti pengertian hadits, bagian-bagiannya, dan jenis-jenis hadits. Tujuan tabel ini adalah untuk memastikan sistem dapat menjawab pertanyaan yang relevan dan umum diajukan oleh pengguna. Berikut ini adalah tabel konteks dari dataset:

Tabel 2. Konteks

Konteks
Assalamualaikum dijawab waalaikumsalam. Nama saya adalah Liysra. Saya berumur 24 tahun. Arash merupakan virtual youtuber yang diciptakan oleh mahasiswa uin bandung bernama moch. Arsyil Albany.
Hadits adalah keterangan-keterangan yang disampaikan secara lisan oleh Rasulullah menyangkut sesuatu masalah duniawi dan agama. Hadits digunakan sebagai sumber hukum kedua dalam agama Islam setelah Al-Qur'an. Hadits memberikan penjelasan, pengajaran, dan tuntunan hidup bagi umat Islam mengenai ajaran-ajaran agama, baik yang berkaitan dengan ibadah, akhlak, maupun tata cara kehidupan sehari-hari. Secara umum, hadits terdiri dari dua bagian yaitu Sanad dan Matan. Sanad adalah orang-orang yang meriwayatkan hadis dari nabi sampai pencatat. Sanad diperlukan untuk mengetahui keaslian hadits. Matan adalah Isi atau teks dari hadits itu sendiri.
Hadits terdapat dua jenis, Berdasarkan kualitasnya, dan berdasarkan isinya. Berdasarkan isinya hadits dibagi menjadi dua hadits qudsi dan hadits nabawi. Hadits qudsi merupakan firman Allah SWT yang disampaikan langsung oleh nabi Muhammad SAW dan tidak tercantum dalam al-quran. Hadits nabawi merupakan ucapan, perbuatan, taqrir, dan sifat-sifat Nabi Muhammad SAW. Proses memverifikasi hadits dikenal sebagai takhrij, di mana ulama melacak sumber dan sanad untuk menilai keabsahannya.
Hadits dapat diklasifikasikan berdasarkan kualitasnya, yaitu Shahih, hasan, dhaif. Hadits shahih merupakan hadits yang memiliki kualitas tinggi dan disepakati para ulama untuk digunakan sebagai sumber hukum. Ciri-ciri hadits shahih yaitu, sanadnya bersambung mulai dari perawi pertama hingga akhir, para perawinya adil dan dabit(kuat daya ingatan), terhindar dari cacat dan janggal, Hadits hasan hampir sama dengan hadits shahih namun hafalan atau ingatan perawi kurang sempurna. Hadits dhoif adalah hadits lemah yang didalamnya tidak terpenuhi syarat-syarat hadits shahih maupun.
Ilmu mustalah hadits mengatur istilah-istilah yang digunakan dalam studi hadits. dan hadits maudhu' adalah istilah untuk hadits palsu yang tidak dapat digunakan sebagai rujukan hukum. Sanad muttasil adalah sanad yang rantai perawi tersambung tanpa putus. sanad munqati adalah sanad yang terputus karena ada perawi yang hilang.. Hadits marfu' adalah hadits yang secara langsung disandarkan kepada Nabi SAW, sedangkan hadits mursal adalah hadits yang sanadnya terputus pada generasi tabi'in. Hadits musnad adalah hadits yang sanadnya bersambung hingga Nabi.
Kitab-kitab Hadits yang terkenal dan paling banyak digunakan sebagai rujukan adalah Shahih Bukhari, Shahih Muslim, Sunan Abu Daud, Sunan Tirmidzi, Sunan Nasa'I, Sunan Ibnu Majah , Musnad Ahmad, Muwatha' Malik, Sunan Darimi.
Salah satu hadits bukhari adalah "Semua perbuatan tergantung niatnya, dan (balasan) bagi tiap-tiap orang (tergantung) apa yang diniatkan: barangsiapa niat hijrahnya karena Allah dan Rasul-Nya, maka hijrahnya adalah kepada Allah dan Rasul-Nya. Barangsiapa niat hijrahnya karena dunia yang ingin digapainya atau karena seorang perempuan yang ingin dinikahnya, maka hijrahnya adalah kepada apa dia diniatkan."(Hadits Riwayat Bukhari no.1).
Salah satu hadits Muslim adalah "Tunaikanlah shalat tepat pada waktunya, bila kalian mampu, tunaikanlah shalat kalian bersama mereka sebagai shalat sunnah."(Hadits Riwayat Muslim no. 1033).

Tabel ini menjelaskan konteks yang digunakan sebagai data pendukung untuk memberikan jawaban dari pertanyaan-pertanyaan pada *Tabel 1*. Konteks ini mencakup informasi dasar seperti salam, pengenalan diri karakter *Vtuber*, serta penjelasan mendalam mengenai hadits, termasuk pengertian, bagian-bagian (sanad dan matan), jenis-jenisnya, serta kitab-kitab hadits terkenal. Tabel ini bertujuan untuk menyediakan data yang cukup bagi model agar dapat memberikan jawaban yang akurat.

Pembersihan Data

Case Folding

Case Folding adalah proses untuk mengubah karakter pada data menjadi huruf kecil[16]. Hal ini akan menyamakan susunan karakter pada data agar lebih mudah untuk dibaca oleh sistem. Berikut adalah gambar codingan dari proses *Case Folding*:

```
[72] hasil_casefolding = str.lower(konteks)
      hasil_casefolding
```

Gambar 4. Case Folding

Stemminng

Stemming data adalah proses untuk mengembalikan suatu kata kedalam bentuk dasarnya[17]. Pada kasus ini akan menghapus kata imbuhan maupun akhiran pada suatu kata. Pada proses ini menerapkan library sastrawi sebagai pendukung natural language processing berbahasa Indonesia. Berikut adalah gambar codingan dari proses Stemming:

```
[73] from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
      factory = StemmerFactory()
      stemmer = factory.create_stemmer()
      input_stemming = str(hasil_casefolding)
      hasil_stemming = stemmer.stem(input_stemming)
      hasil_stemming
```

Gambar 5. *Stemming*

Konstruksi Data

Tokenisasi

Tokenisasi adalah proses untuk memisahkan teks dalam sebuah kalimat, paragraf maupun data lainnya menjadi token atau bagian[18]. Acuan yang digunakan dalam proses ini adalah spasi. Berikut adalah codingan proses tokenisasi:


```
[74] from nltk.tokenize import word_tokenize
      hasil_tokenize = word_tokenize(hasil_stemming)
      hasil_tokenize
```

'adalah',
'kepada',
'allah',
'dan',
'rasul',

Gambar 6. Tokenisasi

Filtering

Filtering adalah penghilangan beberapa fitur yang tidak diinginkan dari data[19]. Filtering yang di terapkan pada penelitian ini adalah stop word removal[20]. Stop word removal adalah proses filtering untuk memproses data dengan menghapus kata yang sering muncul serta tidak memiliki makna seperti “yang”, “dan”, “di”, “dari”. Berikut adalah gambar codingan proses filtering:

```
[75] from nltk.corpus import stopwords
      stop_words = set(stopwords.words('indonesian'))
      hasil_stop = [word for word in hasil_tokenize if not word in stop_words]
      hasil_stop
```

'sunan',
'nasa',
'i',
'sunan',
'ibnu',
'majah',
'musnad',
'ahmad',

Gambar 7. Filtering

Hasil Prototype (Modeling Quick Plan)

Pembangunan Model

Proses pembangunan model menggunakan model BERT ini akan memanfaatkan *library transformer*. Model yang digunakan pada penelitian ini merupakan 'esakrissa/IndoBERT-SQuAD' pemilihan model ini agar model dapat mengolah data berbahasa Indonesia. Sementara SQuAD' adalah model *question and answering* menggunakan BERT.

```
modelQAs = pipeline(
    "question-answering",
    model="esakrissa/IndoBERT-SQuAD",
    tokenizer=tokenizer
)

configjson: 100% 716/716 [00:00<00:00, 10.2kB/s]
pytorch_model.bin: 100% 440M/440M [00:13<00:00, 34.1MB/s]
```

```
cek = modelQAs({
    'context': konteks,
    'question': tanya[0]})
print("Confidence Score = "+str(cek['score']))
print("Jawab = "+cek['answer'])
print("Start = "+str(cek['start']))
```

Confidence Score = 0.25158724188804626
Jawab = arash
Start = 72

Gambar 8. Modeling

4. Pembahasan

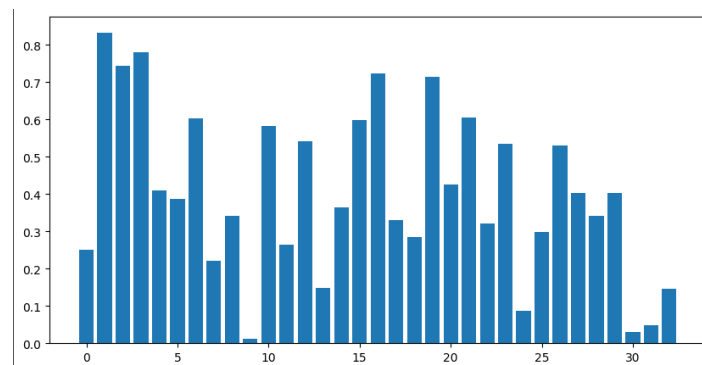
Pembahasan Pengujian Prototype (*Construction of Prototype*)

Pengujian pada model ini dilakukan dengan menggunakan nilai *confidence score* yang diperoleh dari fungsi *pipeline* dalam *library transformers*, berdasarkan jawaban yang diberikan oleh model. *Confidence score* ini merepresentasikan tingkat keyakinan model dalam memberikan jawaban terhadap pertanyaan yang diajukan. Berikut adalah gambar hasil pengujian model:

No	Pertanyaan	Jawaban	Score
1	1	Siapa nama kamu?	Lisyra 0.895842
3	3	Siapa penciptamu?	Moch. Arsyil Albany 0.745701
16	16	Ada berapa jenis-jenis hadits berdasarkan kual...	dua 0.728839
19	19	Apa itu hadits hasan?	dhaif 0.714810
2	2	Berapa umur kamu?	24 0.713980
6	6	Hadits terdiri dari	Sanad dan Matan 0.608345
21	21	Apa itu ilmu mustalaf hadits? mengatur istilah-istilah yang digunakan dalam ...	0.604849

Gambar 9. Hasil Pengujian Model

Pada gambar 9, yang menunjukkan hasil pengujian model, terlihat bahwa secara keseluruhan model menghasilkan *confidence score* yang tinggi. Namun, ketika dilihat lebih mendalam, model ini mengalami kendala dalam menjawab beberapa pertanyaan karena adanya token yang serupa, seperti kata "hadits", yang sering muncul dalam berbagai konteks. Token ini digunakan secara luas dalam banyak pertanyaan dan jawaban, sehingga model kesulitan untuk membedakan konteks spesifik dari setiap pertanyaan yang diajukan. Adapun grafik *confidence score* yang dihasilkan oleh model yaitu:



Gambar 10. Grafik Confidence Score

Setelah mengevaluasi grafik pada gambar 10, model dapat menjawab pertanyaan-pertanyaan dengan baik, namun secara keseluruhan terdapat beberapa pertanyaan yang tidak terjawab dengan baik. Salah satu kendala yang ditemukan selama pengujian adalah jawaban yang terpotong saat sistem menjawab pertanyaan dengan konteks jawaban panjang, seperti "Berikan contoh hadits." Masalah ini terjadi karena keterbatasan panjang keluaran (*output length*) bawaan model, sehingga informasi tidak sepenuhnya disampaikan. Jawaban terpotong ini terutama terjadi pada pertanyaan yang memerlukan penyebutan lengkap sanad dan matan hadits. Model memiliki batas panjang keluaran default yang sering kali terlalu pendek untuk mencakup keseluruhan jawaban, terutama jika dilengkapi dengan penjelasan konteks. Selain itu, struktur hadits yang kompleks, yang terdiri dari sanad dan matan, sering kali menyebabkan model memprioritaskan bagian tertentu yang dianggap penting berdasarkan pelatihan dataset, tetapi tidak selalu relevan untuk menjawab keseluruhan pertanyaan.

Pembahasan Memperbaiki Prototype (*Deployment Delivery & Feedback*)

Dari hasil pengujian sebelumnya, diketahui bahwa banyak *confidence score* yang dihasilkan oleh sistem tidak konsisten. Ketidak konsistenan ini disebabkan oleh kemungkinan model memberikan jawaban yang kurang tepat saat *confidence score* terlalu rendah. Untuk memperbaiki masalah ini, dilakukan pengelompokan konteks berdasarkan kategori tertentu seperti pengenalan, pengertian, jenis isi, jenis kualitas, mustalah, dan kitab, agar sistem dapat memberikan jawaban yang lebih akurat dan maksimal. Berikut adalah contoh gambar model setelah pengelompokan.

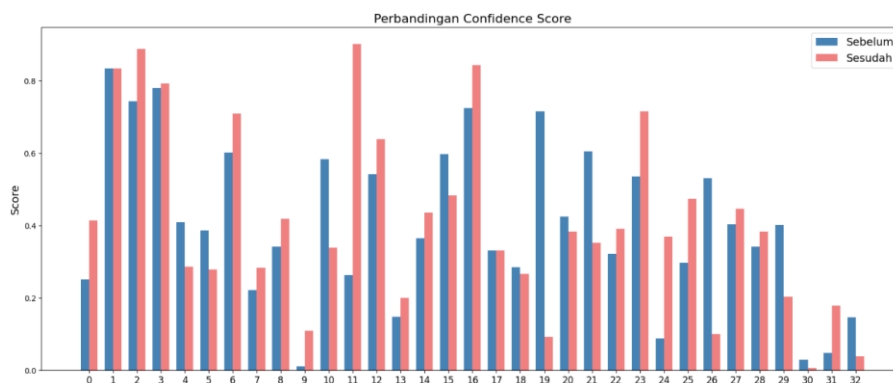
```
[ ] konteksJenisIsi=(
    ' Secara umum Hadits dibagi Berdasarkan kualitasnya dan berdasarkan isinya. Berdasarkan isinya hadits
    )

tanyaJenisIsi=(
    "Ada berapa jenis-jenis hadits berdasarkan isinya?",
    "Apa saja jenis-jenis hadits berdasarkan isinya?",
    "Apa itu hadits qudsi?",
    "Apa itu hadits nabawi?",
    "Apa itu takhrij?"
    )

[ ] cekJenisIsi = modelQAs({
    'context': konteksJenisIsi,
    'question': tanyaJenisIsi[4]
})
print("Confidence Score = "+str(cekJenisIsi['score']))
print("Jawab = "+cekJenisIsi['answer'])
print("Start = "+str(cekJenisIsi['start']))
```

Gambar 11. Model setelah dikelompokkan

Berikut adalah grafik perbandingan *confidence score* sebelum dan sesudah pengelompokan:



Gambar 12. Grafik *confidence score* setelah dan sebelum pengelompokan

Berikut adalah Gambar perbandingan hasil pengujian sebelum dan setelah pengelompokan.

	No	Pertanyaan	Jawaban	Score Sebelum	Score Sesudah	Jawaban Sesudah
1	1	Siapa nama kamu?	Lisya	0.895842	0.914514	Liysra
11	11	Ada berapa jenis-jenis hadits berdasarkan isinya?	dua	0.272454	0.902006	dua
2	2	Berapa umur kamu?	24	0.713980	0.901221	24
16	16	Ada berapa jenis-jenis hadits berdasarkan kual...	dua	0.728839	0.843093	Shahih, hasan, dhaif
3	3	Siapa penciptamu?	Moch. Arsyil Albany	0.745701	0.742050	Moch. Arsyil Albany
23	23	Ada berapa jenis sanad?	dua	0.537309	0.715341	2
6	6	Hadits terdiri dari?	Sanad dan Matan	0.608345	0.709974	Sanad dan Matan
12	12	Apa saja jenis-jenis hadits berdasarkan isinya?	hadits qudsi dan hadits nabawi	0.530268	0.638443	hadits qudsi dan hadits nabawi
15	15	Apa itu takhrij?	Proses memverifikasi hadits	0.600429	0.482557	Proses memverifikasi hadits
25	25	Apa itu sanad muttaqil?	rantai perawi yang tersambung tanpa putus	0.297730	0.474481	berarti rantai perawi yang tersambung tanpa putus

Gambar 13. Hasil pengujian model setelah dan sebelum pengelompokan

Dari gambar di atas, dapat dilihat bahwa setelah dilakukan pengelompokan, *confidence score* model meningkat secara signifikan. Pada pertanyaan "Apa kegunaan Sanad?", model menghasilkan *confidence score* (0.341288) sebelum dan (0.418514) setelah, meskipun jawaban

yang diberikan sama dan benar. Hal ini menunjukkan bahwa meskipun *confidence score* yang lebih rendah menunjukkan tingkat keyakinan model yang lebih kecil, nilai tersebut tidak dapat menjamin bahwa jawaban yang diberikan pasti salah. Oleh karena itu, untuk memperbaiki kinerja model lebih lanjut, dilakukan evaluasi menggunakan *F1-score*, yang memberikan gambaran lebih jelas mengenai keseimbangan antara *precision* (keakuratan jawaban) dan *recall* (kelengkapan jawaban). Berikut merupakan hasil dari pengujian menggunakan metode F1-Score

```
True Positives : 31
True Negatives : 0
False Positives : 2
False Negatives : 0
Accuracy : 0.9393939393939394
Precision : 0.9393939393939394
Recall : 1.0
F1 Score : 0.96875
```

Gambar 14. Hasil *f1-score*

Berdasarkan hasil evaluasi *F1-score* yang diperoleh setelah pengelompokan, sistem menunjukkan performa yang cukup baik dengan *score* 0.96875. Nilai *F1-score* yang tinggi menunjukkan bahwa pengelompokan konteks berhasil membantu model dalam menangani pertanyaan dengan lebih efektif, termasuk pertanyaan yang memiliki token serupa. Meskipun demikian, model tetap mengalami kesulitan dalam menangani pertanyaan yang memerlukan jawaban panjang, di mana jawaban otomatis cenderung terpotong.

Dengan demikian, hasil *F1-score* menunjukkan bahwa meskipun pengelompokan konteks telah meningkatkan kinerja model secara keseluruhan, namun masih ada ruang untuk perbaikan lebih lanjut, terutama dalam menangani jawaban yang lebih kompleks dan panjang.

Perbandingan dengan Model Lain

Sebagai perbandingan, penulis juga menguji model lain, yaitu "malaputri/IndoBERT-SQuAD-ID", yang menghasilkan *F1-score* sebesar (0.4651162). berikut merupakan gambar hasil *F1-score* model "malaputri/IndoBERT-SQuAD-ID"

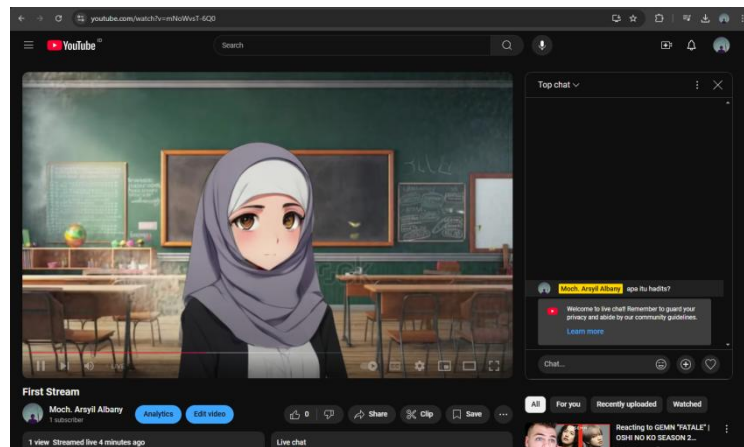
```
True Positives : 10
True Negatives : 0
False Positives : 23
False Negatives : 0
Accuracy : 0.30303030303030304
Precision : 0.30303030303030304
Recall : 1.0
F1 Score : 0.4651162790697675
```

Gambar 15. Hasil *f1-score* model malaputri/IndoBERT-SQuAD-ID

Hasil ini menunjukkan bahwa model 'esakrissa/IndoBERT-SQuAD' yang digunakan dalam penelitian ini memiliki kinerja yang lebih baik dalam menangani pertanyaan berbasis Bahasa Indonesia dibandingkan dengan model "malaputri/IndoBERT-SQuAD-ID". Meskipun kedua model menggunakan BERT untuk tugas question and answering, model 'esakrissa/IndoBERT-SQuAD' menunjukkan kemampuan yang lebih unggul dalam memahami konteks dan memberikan jawaban yang lebih akurat.

Deployment

Berikut ini merupakan gambar dari live streaming youtube:



Gambar 15. *Deployment Virtual Youtuber*

Tahap deployment merupakan langkah akhir dari pengembangan sistem Virtual YouTuber berbasis Question and Answer (QnA) untuk konten hadis. Pada tahap ini, sistem yang telah diuji dan disempurnakan diimplementasikan secara nyata dalam bentuk Live Streaming Youtube yang dapat diakses oleh pengguna.

Implikasi dan Tantangan

Sistem telah menunjukkan potensi besar dalam pendidikan Islam berbasis teknologi. Namun, perbaikan diperlukan untuk menangani jawaban kompleks yang melibatkan teks panjang. Selain itu, sistem perlu dikembangkan untuk memastikan keakuratan yang lebih baik pada berbagai pertanyaan dengan konteks yang berbeda.

5. Penutup

Penelitian ini berhasil mengembangkan sistem tanya jawab Islami berbasis kecerdasan buatan dalam bentuk Virtual YouTuber (VTuber) yang berfokus pada hadits. Dengan memanfaatkan algoritma IndoBERT-SQuAD untuk pemrosesan bahasa alami, sistem menunjukkan kinerja yang baik dalam menjawab pertanyaan terkait hadits. Penerapan metodologi prototipe memungkinkan proses pengembangan yang bertahap dan terukur, mulai dari desain hingga evaluasi.

Pengelompokan konteks terbukti meningkatkan performa sistem, seperti yang ditunjukkan oleh peningkatan nilai F1-score menjadi 0.96875. Meskipun demikian, sistem masih menghadapi kendala dalam menjawab pertanyaan dengan jawaban panjang, di mana hasil jawaban cenderung terpotong. Hal ini menunjukkan bahwa meskipun model telah optimal untuk pertanyaan pendek hingga menengah, perbaikan lebih lanjut diperlukan untuk menangani jawaban yang lebih panjang.

Kesimpulannya, sistem ini tidak hanya menawarkan kemudahan akses terhadap informasi hadits tetapi juga memberikan pendekatan interaktif dan modern dalam mendekatkan generasi muda kepada ajaran Islam. Sebagai rekomendasi untuk pengembangan model di masa mendatang, integrasi dengan chatbot berbasis audio dapat dipertimbangkan. Pendekatan ini memungkinkan pengguna untuk berinteraksi dengan sistem melalui percakapan lisan, yang dapat meningkatkan aksesibilitas dan kenyamanan.

6. Referensi

- [1] P. Chowdhary, "Natural Language Processing," 2020, pp. 603–649. doi: 10.1007/978-81-322-3972-7_19.
- [2] F. R. Pradhana, T. Harmini, and H. R. A. S, "Implementasi Teknologi Augmented Reality Dalam Pembelajaran Tajwid Kelas 5 Pada Hukum Bacaan Mim Sukun Dan Tanwin Berbasis MDA Framework," *Smatika Jurnal*, vol. 13, no. 02, pp. 350–360, 2023, doi: 10.32664/smatika.v13i02.1002.
- [3] E. DANINGRUM, "Dampak Penggunaan Media Sosial Youtube Melalui Film Pendek Islami Terhadap Perilaku Remaja Muslim," 2019.
- [4] Andrea Lidwina, "94% Orang Indonesia Akses YouTube dalam Satu Bulan Terakhir." Accessed: Nov. 25, 2024. [Online]. Available: <https://databoks.katadata.co.id/teknologi-telekomunikasi/statistik/9246278d026597c/94-orang-indonesia-akses-youtube-dalam-satu-bulan-terakhir>
- [5] A. R. B. Zaman and M. M. Assarwani, "Habib Husein Jafar Al-Hadar's Da'wa Content Commodification on Youtube," vol. 15, pp. 1–11, 2021, doi: 10.24090/KOMUNIKA.V15I1.3986.
- [6] H. Kurniawan, "Indonesia Rangking Pertama dengan Komunitas Vtuber Terbesar di Asia Tenggara." Accessed: Nov. 25, 2024. [Online]. Available: <https://tekno.sindonews.com/read/1058393/207/indonesia-rangking-pertama-dengan-komunitas-vtuber-terbesar-di-asia-tenggara-1680008621>
- [7] Y. Tan, "More Attached, Less Stressed: Viewers' Parasocial Attachment to Virtual Youtubers and Its Influence on the Stress of Viewers During the COVID-19 Pandemic," *SHS Web of Conferences*, vol. 155, p. 03012, 2023, doi: 10.1051/shsconf/202315503012.
- [8] M. Jayadi, "Kedudukan Dan Fungsi Hadis Dalam Islam," *Jurnal Adabiyah*, vol. XI, no. 2, pp. 242–255, 2011.
- [9] F. Alaba and O. C. Kolade, "A Novel Hadith Authentication Mobile System for Android and Ios Phones With Arabic to English Language Translation," *Journal of Media & Management*, 2023, doi: 10.47363/jmm/2023(5)160.
- [10] K. Gaanoun and M. Alsuhaibani, "Fabricated Hadith Detection: A Novel Matn-Based Approach With Transformer Language Models," *IEEE Access*, vol. 10, pp. 113330–113342, 2022, doi: 10.1109/ACCESS.2022.3217457.
- [11] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. Salakhutdinov, and Q. V. Le, "XLNet: Generalized autoregressive pretraining for language understanding," *Advances in Neural Information Processing Systems*, vol. 32, no. NeurIPS, pp. 1–18, 2019.
- [12] Y. Chen and F. Zulkernine, *BIRD-QA: A BERT-based Information Retrieval Approach to Domain Specific Question Answering*. 2021. doi: 10.1109/BigData52589.2021.9671523.
- [13] S. Rosyad and M. Alif, "Hadis di Era Digital: Tantangan dan Peluang Penggunaan Teknologi dalam Studi Hadis," *Jurnal Ilmu Agama: Mengkaji Doktrin, Pemikiran, dan Fenomena Agama*, vol. 24, no. 2, pp. 185–197, 2023, doi: 10.19109/jia.v24i2.18979.
- [14] D. P. Dika, "Sekilas Mekanisme BERT: Algoritma di Balik Model Bahasa yang Canggih," 07 Mar 2024. Accessed: Nov. 25, 2024. [Online]. Available: <https://dikakaryatech.com/software/development/2024/03/07/sekilas-mekanisme-BERT-algoritma-di-balik-model-bahasa-yang-canggih.html>
- [15] M. A. PROF DR. ZIKRI DARUSSAMIN, *Kuliah Ilmu Hadis 1*. 2020.
- [16] Y. HaCohen-Kerner, D. Miller, and Y. Yigal, "The influence of preprocessing on text classification using a bag-of-words representation," *PLoS ONE*, vol. 15, no. 5, pp. 1–22, 2020, doi: 10.1371/journal.pone.0232525.
- [17] J. Singh and V. Gupta, "Text Stemming: Approaches, Applications, and Challenges," *ACM Comput. Surv.*, vol. 49, no. 3, Sep. 2016, doi: 10.1145/2975608.
- [18] L. Mullen, K. Benoit, O. Keyes, D. Selivanov, and J. Arnold, "Fast, Consistent Tokenization of Natural Language Text," *J. Open Source Softw.*, vol. 3, p. 655, 2018, doi: 10.21105/jOSS.00655.

- [19] C. Schiffman *et al.*, "Filtering procedures for untargeted LC-MS metabolomics data," *BMC Bioinformatics*, vol. 20, 2019, doi: 10.1186/s12859-019-2871-9.
- [20] J. Raulji and J. Saini, "Stop-Word Removal Algorithm and its Implementation for Sanskrit Language," *International Journal of Computer Applications*, vol. 150, pp. 15–17, 2016, doi: 10.5120/IJCA2016911462.