



## Tinjauan Pengesanan Hadis Palsu Menggunakan Pendekatan Pembelajaran Mesin

*[A Survey of Fabricated Hadith Detection Using Machine Learning Approach]*

**Borhan Ab Rahman<sup>1</sup>, \*Mohd Zakree Ahmad Nazri<sup>2</sup>, \*Latifah Abdul Majid<sup>3</sup>**

<sup>1</sup>Pusat Teknologi Kecerdasan Buatan (CAIT), Fakulti Teknologi dan Sains Maklumat, Universiti Kebangsaan Malaysia, [borhan.abrahaman@gmail.com.my](mailto:borhan.abrahaman@gmail.com.my), 0133971376

\* Corresponding Author:

<sup>2</sup>Pusat Teknologi Kecerdasan Buatan (CAIT), Fakulti Teknologi dan Sains Maklumat, Universiti Kebangsaan Malaysia, [zakree@ukm.edu.my](mailto:zakree@ukm.edu.my), 0192672301

<sup>3</sup>Pusat Penyelidikan al-Quran dan al-Sunnah (PQS), Fakulti Pengajian Islam, Universiti Kebangsaan Malaysia, [umilm@ukm.edu.my](mailto:umilm@ukm.edu.my), 0192424799

---

### **Keywords:**

*Fabricated Hadith; Machine Learning; Fabricated Hadith Detection*

---

### **ABSTRACT**

*In Islam, the Prophet's Hadith is the second source of reference after the holy book of the Quran. Everything specified in the Prophet SAW's hadith is a reference for all Muslims. However, not everything cited as a hadith is sahih; it is sourced from the Prophet SAW because there are falsehoods and they should be itemised as fabricated hadiths. Establishing fabricated hadith has become an increasingly significant issue, especially with the widespread use of digital platforms. Although manual detection of fabricated hadith is still a considerable practice, new machine learning approaches can potentially increase the accuracy of authenticating hadith. Various recent studies have explored the use of artificial intelligence methods as an innovative solution to overcome the problem of precision and consistency in hadith authentication. The requirement to accurately understand and verify hadiths has prompted in-depth research into leveraging machine-learning approaches to identify fabricated hadiths. This study provides a detailed review and comparison of previous studies on aspects of machine learning techniques, data sets, and evaluation parameters used to evaluate the potential of the proposed model. This study also discusses the difficulties, challenges, and suggestions that can utilise the potential of machine learning in the domain of hadith knowledge.*

---

### **Kata Kunci:**

Hadis Palsu; Pembelajaran Mesin; Pengesanan Hadis Palsu

---

### **ABSTRAK**

Hadis Rasulullah SAW merupakan sumber rujukan kedua dalam agama Islam selepas kitab suci Al-Quran. Segala yang dinyatakan dalam hadis Rasulullah SAW menjadi rujukan bagi semua umat Islam di seluruh dunia. Namun, bukan semua yang dinyatakan sebagai hadis ialah sahih bersumberkan Rasulullah SAW kerana terdapat pendustaan yang dinamakan sebagai hadis palsu. Pengesanan hadis palsu telah menjadi isu yang semakin penting dalam bidang ilmu hadis, terutama dengan penyebarannya yang luas melalui platform digital. Walaupun pengesanan hadis palsu secara manual masih menjadi amalan, pendekatan baru menggunakan pembelajaran mesin mempunyai potensi untuk meningkatkan

---

keberkesanan dan ketepatan dalam proses pengesahan hadis. Pelbagai kajian terkini telah dijalankan untuk meneroka penggunaan kaedah kecerdasan buatan sebagai satu penyelesaian yang inovatif bagi mengatasi masalah ketepatan dan konsistensi dalam pengesahan hadis. Keperluan untuk memahami dan mengesahkan hadis dengan tepat telah mendorong penyelidikan yang mendalam dalam memanfaatkan pendekatan pembelajaran mesin untuk mengesan hadis palsu. Kajian ini memberikan ulasan terperinci dan perbandingan kajian yang lepas daripada aspek teknik pembelajaran mesin, set data dan parameter penilaian yang digunakan untuk menilai potensi model yang dicadangkan. Kajian ini juga membincangkan kesukaran, cabaran yang dihadapi dan cadangan yang boleh memanfaatkan potensi pembelajaran mesin dalam domain ilmu hadis.

---

Received: July 29, 2024

Accepted: 15 Nov, 2024

Online Published: Dec 31, 2024

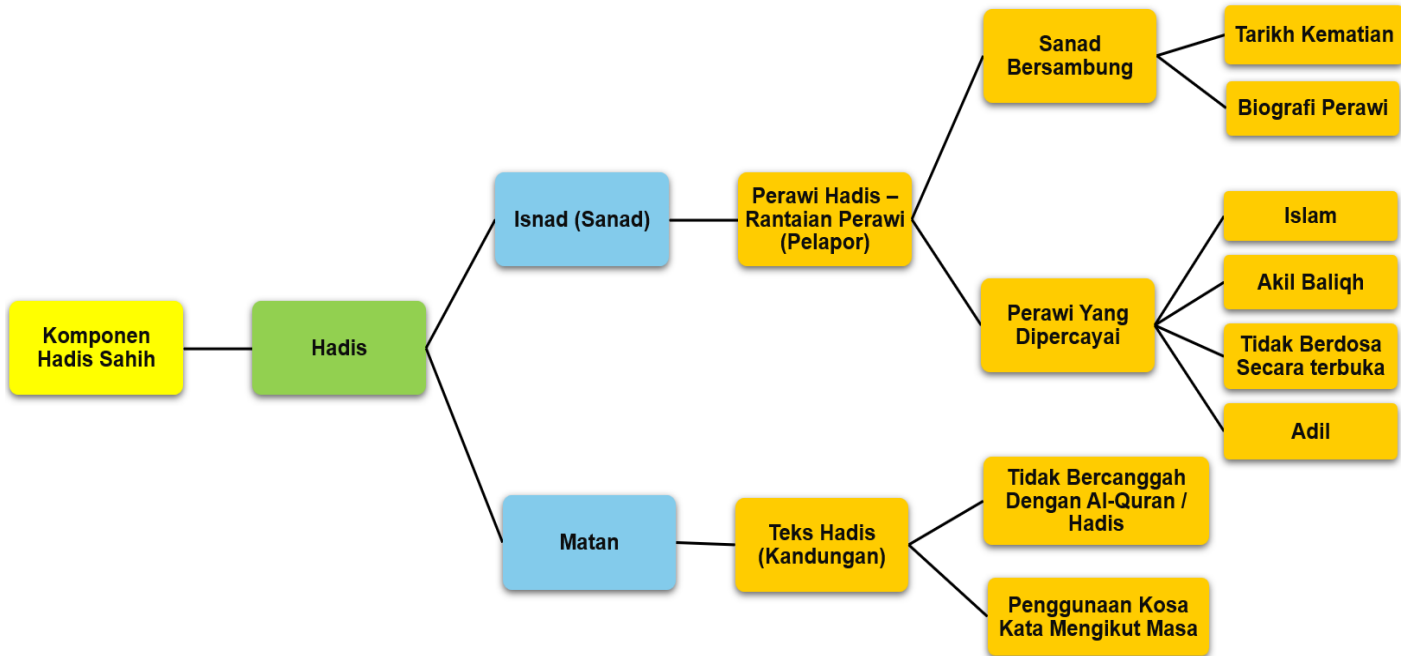
## 1. Latar Belakang

Pengesahan hadis palsu merupakan salah satu cabang penelitian yang penting dalam bidang kajian Islam, memandangkan kesahihan hadis memainkan peranan kritikal dalam menentukan hukum dan amalan dalam agama Islam. Dalam era digital ini, teknologi telah memainkan peranan penting dalam memudahkan proses pengesanan dan penapisan hadis. Secara khususnya, pendekatan pembelajaran mesin telah menunjukkan potensi yang signifikan dalam mengenal pasti hadis palsu dengan lebih efisien dan tepat. Penggunaan teknologi ini bukan sahaja membuka ruang baru dalam kajian hadis yang boleh mempercepatkan proses verifikasi malah dapat juga meningkatkan kebolehpercayaan keputusan yang diperolehi (Binbeshr et al. 2021).

Pembelajaran mesin yang merupakan cabang dari kecerdasan buatan adalah merujuk kepada sistem komputer yang dibina untuk belajar dan membuat ramalan atau keputusan berdasarkan data tanpa diprogramkan secara eksplisit (Belle & Papantonis 2021). Dalam konteks pengesanan hadis, teknik pembelajaran mesin seperti pengklasifikasian teks, pemprosesan bahasa semula jadi, dan analisis sentimen, telah digunakan untuk mengenal pasti ciri-ciri teks hadis yang mungkin menunjukkan tidak sah. Teknik ini memanfaatkan data hadis yang besar dan telah diklasifikasikan oleh para ulama untuk melatih model yang dapat mengenal pasti pola-pola tertentu yang sering muncul dalam hadis palsu (Mohamed & Sarwar 2021).

Untuk mengesahkan keaslian hadis dan memastikan ia benar-benar berasal daripada Nabi Muhammad (SAW), ulama hadis telah menetapkan beberapa kaedah terperinci. Kaedah-kaedah ini termasuk memiliki sanad (rantai perawi) yang bersambung, perawi yang dipercayai dan memiliki kemampuan menghafaz yang cemerlang, serta hadis yang bebas dari penyimpangan yang mencurigakan atau kecacatan tersembunyi (Hakak et al. 2022). Namun menurut Hakak et, al (2022), tiga kaedah utama yang dapat dilaksanakan dengan mudah melalui teknologi ICT ialah hadis yang memiliki sanad bersambung, perawi yang boleh dipercayai dan pendekatan berasaskan kandungan teks (matan) seperti Rajah 1.

**Rajah 1: Komponen Hadis Sahih**



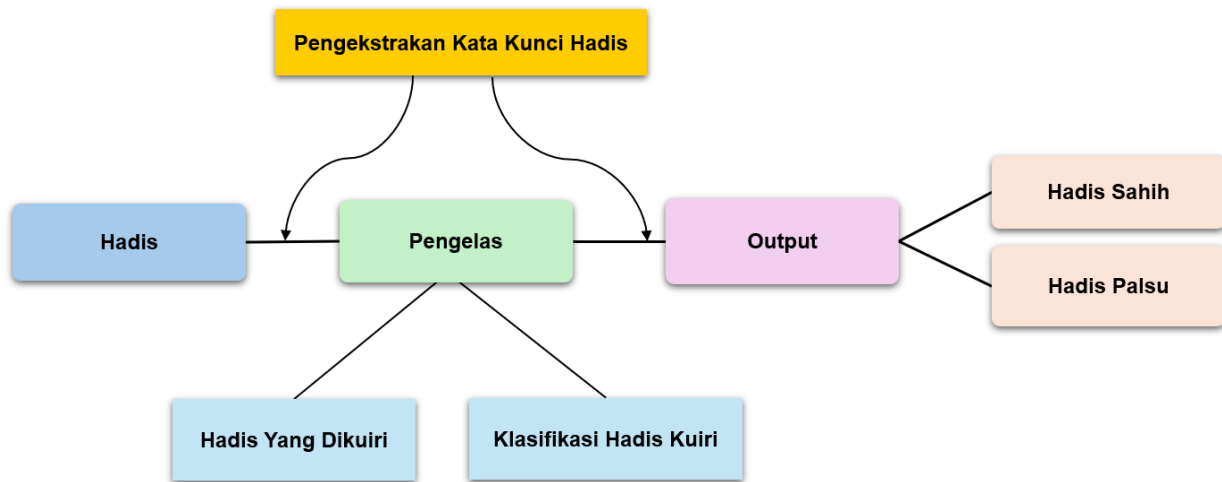
Rajah 1 menggambarkan komponen-komponen hadis sahih yang terbahagi kepada dua bahagian utama iaitu Isnad (Sanad) dan Matan (Teks/Kandungan) (Manas 2022). Hadis sahih merupakan hadis yang memenuhi syarat-syarat tertentu dalam dua komponen utamanya: Isnad (Sanad) dan Matan (Teks/Kandungan). Isnad merujuk kepada rantaian perawi atau pelapor yang menyampaikan hadis daripada Rasulullah SAW kepada generasi seterusnya. Bagi memastikan kesahihan hadis, sanad mesti bersambung, iaitu rantaian perawi tidak terputus di mana setiap perawi mestilah menerima hadis tersebut secara langsung daripada perawi sebelumnya. Keabsahan sanad ini juga disokong oleh rekod sejarah yang sahih seperti tarikh kematian perawi, yang digunakan untuk menentukan sama ada dua perawi hidup pada masa yang sama, dan biografi perawi yang merangkumi maklumat seperti tempat tinggal, guru-guru, dan murid-murid perawi. Selain itu, perawi yang dipercayai juga merupakan elemen penting dalam sanad, di mana setiap perawi mesti mempunyai sifat-sifat seperti Islam, akil baliqh, tidak berdosa secara terang-terangan, dan adil. Kepercayaan kepada perawi ini berdasarkan kepada integriti dan komitmen mereka terhadap ajaran agama.

Manakala Matan merujuk kepada teks atau kandungan hadis itu sendiri. Kandungan hadis mestilah tidak bercanggah dengan Al-Quran atau hadis-hadis lain yang sahih. Jika terdapat percanggahan, hadis tersebut dianggap lemah atau palsu. Selain itu, penggunaan kosa kata mengikut masa juga penting; bahasa dan terminologi yang digunakan dalam matan mesti sesuai dengan konteks dan zaman di mana hadis itu diriwayatkan. Ketidaksesuaian dalam penggunaan kosa kata atau logik bahasa yang digunakan boleh menimbulkan keraguan terhadap kesahihan hadis tersebut.

Secara keseluruhannya, hadis dianggap sahih apabila kedua-dua komponen Isnad dan Matan memenuhi syarat-syarat tertentu. Isnad yang bersambung dan perawi yang dipercayai memastikan hadis itu disampaikan dengan tepat dari generasi ke generasi. Sementara itu, Matan yang tidak bercanggah dengan sumber utama Islam dan penggunaan bahasa yang sesuai menjamin keaslian dan relevansi kandungan hadis tersebut. Oleh itu, pemenuhan syarat-syarat ini adalah penting dalam memastikan hadis yang diriwayatkan benar-benar sahih dan dapat dijadikan rujukan dalam kajian ilmu hadis.

Secara ringkas, kajian oleh Fadele et al. (2021) mencadangkan proses klasifikasi hadis seperti yang digambarkan dalam Rajah 2. Proses ini melibatkan penggunaan sistem automatik yang memanfaatkan teknologi pembelajaran mesin untuk menentukan keaslian teks sesuatu hadis iaitu tugas yang sebelum ini biasanya dilakukan oleh pakar hadis melalui analisis manual. Sistem ini dirancang untuk mengautomatiskan proses klasifikasi hadis dengan mengidentifikasikan pola-pola dan ciri-ciri tertentu yang membezakan hadis sahih dari hadis palsu. Dengan cara ini, kaedah pembelajaran mesin yang digunakan dapat menyokong dan mempercepat tugas pengesahan sesuatu hadis.

**Rajah 2: Aliran Proses Klasifikasi Hadis Menggunakan Pembelajaran Mesin**



Aliran proses klasifikasi hadis ini merangkumi beberapa langkah penting yang bertujuan untuk menentukan keaslian sesuatu hadis. Proses ini bermula dengan kemasukan hadis, iaitu pernyataan yang terdiri daripada catatan percakapan, tindakan, dan apa sahaja yang disandarkan kepada Nabi Muhammad SAW. Hadis ini berfungsi sebagai input awal yang akan dianalisis untuk menentukan keasliannya. Langkah kedua adalah pengekstrakan kata kunci hadis. Pengekstrakan ini melibatkan pengenalan kata kunci penting daripada teks hadis menggunakan teknik pemprosesan bahasa semula jadi (NLP). Teknik ini mengenal pasti perkataan atau frasa yang berkaitan yang menunjukkan keaslian atau ketulenan kandungan hadis. Proses pengekstrakan ini penting kerana ia mengenal pasti elemen-elemen yang relevan dalam teks hadis yang boleh membantu dalam proses klasifikasi seterusnya.

Setelah kata kunci diekstrak, ia dimasukkan ke dalam pengelas, iaitu model pembelajaran mesin yang telah dilatih untuk mengkategorikan data berdasarkan ciri-ciri tertentu, termasuk kata kunci yang telah diekstrak daripada hadis. Dalam konteks ini, pengelas berfungsi untuk menentukan kategori hadis berdasarkan ciri-ciri unik yang dikenal pasti dalam teks. Terdapat juga proses pengelasan untuk hadis pertanyaan, di mana teks hadis tertentu dimasukkan ke dalam sistem untuk diklasifikasikan. Pengelas memproses klasifikasi hadis pertanyaan ini berdasarkan kata kunci yang diekstrak dan mengkategorikannya mengikut ciri-ciri tertentu yang telah dipelajari.

Proses klasifikasi ini menghasilkan output yang membahagikan hadis kepada dua kategori utama: Hadis Sahih dan Hadis Palsu. Hadis Sahih adalah hasil daripada pengelas yang menentukan bahawa hadis tersebut berkemungkinan tulen berdasarkan pola yang dipelajari daripada kata kunci. Sebaliknya, Hadis Palsu adalah hasil daripada pengelas yang menentukan bahawa hadis tersebut tidak sejajar dengan pola yang telah dipelajari untuk kesahihan. Keseluruhan aliran proses ini memastikan bahawa hanya hadis-hadis yang sahih dan boleh dipercayai diklasifikasikan sebagai sahih, manakala hadis-hadis yang tidak memenuhi syarat kesahihan dikenal pasti sebagai palsu. Dengan ini, proses ini penting dalam memastikan integriti dan ketepatan maklumat yang disampaikan melalui hadis.

## 2. Kajian Lepas

Sejak beberapa tahun kebelakangan ini terdapat beberapa kajian yang telah dijalankan oleh penyelidik terdahulu yang berkaitan dengan pengesanan hadis palsu menggunakan kaedah pembelajaran mesin. Berdasarkan kepada Bahrudin et al. (2023) mendapati terdapat minat yang ketara dalam kalangan penyelidik terhadap pengesanan hadis, terutamanya dalam aspek pengesanan digital. Dalam era digital ini, kepentingan memastikan keaslian teks agama seperti hadis menjadi semakin penting. Penyalahgunaan dan penyebaran hadis palsu boleh menimbulkan kekeliruan dan salah tafsir dalam komuniti Muslim. Bagi mengamalkan ajaran Islam yang sebenar, umat Islam perlu merujuk kepada hadis-hadis yang telah disahkan kesahihannya oleh para ulama. Jika gagal berpegang kepada hadis-hadis yang sahih, umat Islam berisiko terdedah kepada kekeliruan dan penyelewengan teks hadis, yang boleh membawa kepada kerosakan akidah dan membatalkan syariat yang dituntut ke atas mereka (Abur Hamdi & Mohd Norzi 2022).

Bagi mengatasi masalah ini, kajian terkini telah meneroka penggunaan teknologi pembelajaran mesin sebagai alat untuk mengesan dan pengesanan keaslian hadis. Melalui pendekatan ini, kajian yang telah dijalankan adalah bertujuan untuk mengembangkan model yang mampu membezakan hadis sahih dari yang palsu dengan nilai keakuran yang tinggi dengan menggunakan teks kandungan hadis dan atribut perawi sebagai input utama.

Kajian oleh Gaanoun dan Alsuhaibani (2022) telah memperkenalkan satu set data baru yang dinamakan MAHADDAT, khusus untuk mengesan hadis palsu. Kajian ini merupakan yang pertama seumpamanya dalam bidang pengesanan hadis palsu, di mana set data MAHADDAT dibangunkan dengan menggabungkan set data sedia ada bersama set data hadis palsu tambahan yang dikumpulkan dari pelbagai laman web. Pendekatan ini sangat penting kerana ia menyediakan asas yang kukuh untuk melatih dan menguji model pengesanan hadis palsu yang dicadangkan oleh penyelidik. Dalam kajian ini, penyelidik telah menggunakan model Bahasa Transformer Arab (TLM), terutamanya BERT (Perwakilan Pengekod Dua Arah daripada Transformer), untuk menguji keaslian hadis. Pendekatan ini menekankan pada analisis kandungan dan teks hadis, yang berbeza dengan kaedah pengesanan hadis tradisional yang biasanya memberi fokus pada sanad dan matan tanpa mengoptimalkan teknologi pemrosesan bahasa semula jadi (NLP). Melalui pendekatan baru ini, model mampu menganalisis konteks dan makna yang lebih mendalam dalam teks hadis untuk menentukan kesahihannya.

Untuk latihan dan penilaian model, set data MAHADDAT dibahagikan kepada tiga bahagian: set latihan, set pembangunan, dan set ujian. Pembahagian ini dilakukan dengan kaedah pelabelan yang konsisten untuk memastikan ketepatan dan kebolehpercayaan model dalam mengenal pasti hadis palsu. Keberkesanan model diukur menggunakan skor metrik F1, yang merupakan ukuran yang menggabungkan ketepatan (*precision*) dan penarikan semula (*recall*) untuk memberikan penilaian yang lebih seimbang mengenai prestasi model. Model yang dicadangkan oleh penyelidik ini mencapai ketepatan yang sangat tinggi dengan skor F1 sebanyak 92.47%, menunjukkan keupayaan yang cemerlang dalam mengesan hadis palsu dengan menggunakan pendekatan berasaskan teks dan kandungan. Inovasi ini menunjukkan potensi besar dalam mengaplikasikan model pembelajaran mesin moden untuk kajian hadis, khususnya dalam konteks pengesanan dan pengesanan hadis palsu.

Namun kajian ini mempunyai beberapa keterbatasan. Pertama, kajian ini hanya memfokuskan pada matan (kandungan teks) hadis dan mungkin mengabaikan kepentingan sanad (rantai perawi) yang merupakan elemen penting dalam pengesanan hadis secara tradisional. Integrasi kedua-dua aspek ini mungkin meningkatkan ketepatan model yang digunakan. Kedua, terdapat ketidakseimbangan dalam set data yang digunakan iaitu hadis palsu diwakili dengan jumlah yang lebih rendah. Ketidakseimbangan dalam set data yang digunakan yang mana hadis palsu diwakili dengan jumlah yang lebih rendah, boleh menyebabkan model pembelajaran mesin menjadi bias terhadap kelas yang dominan (hadis sahih) dan kurang efektif dalam mengenal pasti kelas yang kurang diwakili (hadis palsu). Ini boleh mengakibatkan penurunan prestasi model dalam mengesan hadis palsu dan meningkatkan risiko membuat kesimpulan yang salah. Langkah-langkah seperti menyeimbangkan data dan penggunaan metrik penilaian yang sesuai diperlukan untuk mengatasi masalah ini (Torabi Asr & Taboada 2019). Kajian pada masa depan akan mendapat manfaat daripada pengembangan set data untuk memasukkan koleksi hadis yang lebih pelbagai dan seimbang yang boleh memperkuat lagi model tersebut. Selain itu, walaupun model ini dilatih dengan menggunakan teks bahasa Arab dan berfungsi dengan baik dalam konteks ini, keberkesanannya dalam bahasa dan konteks budaya lain belum diuji. Usaha untuk memperluaskan metodologi dengan memasukkan set data berbilang bahasa dan silang budaya boleh meningkatkan kegunaannya secara meluas (Küçük & Can 2021).



Kajian yang dijalankan oleh Fadele et al. (2021) telah memperkenalkan satu taksonomi komprehensif untuk mengklasifikasikan pelbagai teknik yang digunakan dalam pengesanan hadis palsu. Taksonomi ini membahagikan teknik pengesanan hadis kepada tiga kategori utama: teknik berpacuan pengetahuan, teknik berpacuan data dan teknik hibrid. Setiap kategori mempunyai pendekatan dan metodologi tersendiri yang menyumbang kepada peningkatan ketepatan dalam pengesanan hadis palsu. Teknik berpacuan pengetahuan bergantung pada penggunaan peraturan atau leksikon yang spesifik, seperti kamus bahasa atau aturan-aturan tetap yang ditetapkan oleh pakar dalam bidang hadis. Teknik ini menggunakan pengetahuan domain yang dikodkan dalam bentuk peraturan untuk mengenal pasti hadis palsu berdasarkan ciri-ciri tertentu dalam teks. Sebagai contoh, teknik ini menggunakan kamus yang mengandungi kata-kata tertentu yang dianggap mencurigakan atau tidak sesuai dengan bahasa dan struktur hadis yang sah. Teknik ini bergantung kepada input daripada pakar dan memerlukan pengetahuan mendalam tentang bahasa dan konteks agama untuk berfungsi dengan baik.

Sebaliknya, teknik berpacuan data menggunakan data sedia ada untuk melatih model pembelajaran mesin dan membuat ramalan baru. Dalam kajian ini, beberapa algoritma pembelajaran mesin seperti Pokok Keputusan (*Decision Tree*, DT), *Naïve Bayes* (NB) dan Mesin Vektor Sokongan (*Support Vector Machine*, SVM) telah digunakan untuk membangunkan dan menguji model-model pengesanan hadis palsu. Teknik ini memanfaatkan data berlabel yang besar untuk melatih model dalam mengenal pasti pola-pola dalam teks hadis yang menunjukkan ketulenan atau kepalsuan. Walau bagaimanapun, teknik berpacuan data ini memerlukan sejumlah besar data berlabel yang berkualiti tinggi, yang mungkin sukar diperoleh dan memerlukan kos yang tinggi untuk dikumpulkan dan diproses (Belle & Papantonis 2021). Teknik hibrid pula menggabungkan kedua-dua pendekatan berdasarkan peraturan dan pembelajaran mesin (ML), dengan tujuan untuk mewujudkan sistem pengesanan yang lebih efektif. Teknik hibrid ini memanfaatkan kelebihan pemahaman kontekstual yang diperoleh daripada pendekatan berpacuan pengetahuan bersama dengan keupayaan ramalan data yang diperoleh daripada model pembelajaran mesin. Ini bermakna, teknik hibrid tidak hanya bergantung pada pengetahuan statik tetapi juga dapat belajar dan menyesuaikan diri dengan data baru, menjadikannya lebih dinamik dan fleksibel dalam mengesan hadis palsu (Najeeb 2022).

Kajian oleh Fadele et al. (2021) ini juga membincangkan beberapa cabaran yang dihadapi oleh setiap teknik. Misalnya, teknik berasaskan kamus dan peraturan memerlukan input pakar yang intensif dan terhad kepada domain tertentu, manakala teknik berpacuan data menghadapi cabaran dalam mendapatkan data berlabel yang berkualiti tinggi, khususnya dalam bahasa Arab yang memiliki struktur linguistik yang kompleks. Kajian ini memberi panduan kepada penyelidikan masa depan dalam pengesanan hadis, seperti keperluan untuk mengatasi masalah data yang tidak seimbang dan meningkatkan pemahaman tentang bahasa Arab dalam konteks pembelajaran mesin. Meskipun model pembelajaran mesin yang dicadangkan memiliki potensi besar dalam pengesanan hadis palsu, mereka juga menghadapi batasan untuk pengembangan lanjut. Oleh itu, kajian lebih mendalam diperlukan untuk mengembangkan sistem hibrid yang lebih maju yang dapat menggabungkan kekuatan kedua-dua pendekatan ini secara efektif dan meningkatkan ketepatan serta keberkesanan dalam pengesanan hadis palsu.

Kajian yang dijalankan oleh Abdelaal & Youness (2019) bertujuan untuk mengkaji pengelasan hadis secara automatik ke dalam empat kategori utama: Sahih, Hasan, Da'if, dan Maudu'. Pengelasan ini berdasarkan kepercayaan dan ingatan perawi dalam menyampaikan hadis. Kajian ini menggunakan teknik Perlombongan Data dan algoritma pembelajaran mesin seperti Pokok Keputusan (*Decision Tree*, DT) dan *Naïve Bayes* (NB) untuk membangunkan model yang mampu membezakan antara kategori-kategori hadis tersebut. Pemilihan Pokok Keputusan adalah kerana kemampuannya untuk menilai dan memanfaatkan maklumat secara efisien dalam membuat keputusan, manakala *Naïve Bayes* dipilih kerana kecepatan dan ketepatannya dalam pengklasifikasian teks berdasarkan kebarangkalian bersyarat.

Kajian ini juga memberi penekanan pada bagaimana bahagian-bahagian hadis, seperti sanad (rantai perawi) dan matan (teks hadis), mempengaruhi proses pengelasan. Model yang dibangunkan dalam kajian ini diuji dengan menggunakan teknik pengesahan silang dan pembahagian peratusan untuk mengukur keberkesannya serta mengurangkan isu bias dalam pengklasifikasian. Teknik pengesahan silang ini membolehkan penilaian yang lebih kukuh terhadap model dengan menggunakan data yang berlainan untuk latihan dan pengujian, memastikan model lebih umum dan boleh diaplikasikan pada set data yang berbeza. Dalam kajian ini, penggunaan *Naïve Bayes*

menunjukkan ketepatan yang tinggi dengan mencapai skor F1 sebanyak 93.75%, menunjukkan keberkesannya dalam mengklasifikasikan hadis berdasarkan kategori yang telah ditetapkan. Skor F1 yang tinggi ini mencerminkan kemampuan model untuk menggabungkan ketepatan (*precision*) dan penarikan semula (*recall*) secara seimbang, yang menunjukkan bahawa model tersebut berprestasi baik dalam membezakan hadis yang sahih dan palsu serta kategori lain berdasarkan kepercayaan perawi.

Walau bagaimanapun, kajian ini juga mempunyai beberapa keterbatasan. Pertama, fokus utama kajian hanya pada kebolehppercayaan perawi dan tidak melibatkan analisis mendalam terhadap matan hadis. Ini boleh menyebabkan pemahaman yang tidak lengkap mengenai keaslian hadis kerana tidak semua aspek kandungan teks dianalisis. Misalnya, kesahihan sesuatu hadis bukan sahaja ditentukan oleh sanad, tetapi juga oleh keserasian kandungan matan dengan prinsip-prinsip asas dalam Islam. Kedua, pendekatan yang digunakan dalam kajian ini adalah berdasarkan kelas-kelas yang telah ditetapkan sebelumnya yang mungkin tidak mencakupi keseluruhan spektrum pengkategorian hadis secara menyeluruh. Ini bererti terdapat kemungkinan bahawa beberapa jenis hadis tidak sesuai dimasukkan dalam kategori yang telah ditentukan, menyebabkan beberapa hadis mungkin diklasifikasikan secara tidak tepat.

Keberkesanan model pengklasifikasian ini juga sangat bergantung kepada kualiti set data yang digunakan. Jika set data yang digunakan mengandungi isu bias atau kesalahan, ia boleh mempengaruhi ketepatan model secara signifikan (Gasparetto et al. 2022; Wang et al. 2020). Misalnya, jika terdapat kesilapan dalam pelabelan data atau jika data tidak seimbang antara kategori hadis yang berbeza, model mungkin cenderung untuk mempelajari pola yang tidak diinginkan atau salah. Walaupun teknik pengesanan silang dapat membantu menilai keberkesanan model pada data baru, kejayaan sebenar model tersebut bergantung pada kemampuannya untuk berfungsi dengan baik pada pelbagai jenis teks hadis yang tidak termasuk dalam set data awal. Oleh itu, lebih banyak kajian dan pengujian diperlukan untuk memastikan model tersebut berfungsi dengan baik dalam situasi dunia sebenar dan untuk memperbaiki pendekatan pengelasan yang digunakan agar lebih holistik dan tepat.

Secara umum, kajian-kajian lepas mengenai pengesanan dan pengelasan hadis palsu telah menunjukkan kemajuan yang signifikan dalam menggunakan pendekatan berasaskan pembelajaran mesin. Kajian-kajian ini telah memperkenalkan model-model baru serta taksonomi teknik pengesanan yang lebih komprehensif, yang merangkumi teknik berpacuan pengetahuan, berpacuan data, dan teknik hibrid. Dengan adanya set data khusus dan penggunaan model bahasa yang canggih seperti BERT, pendekatan baru ini menunjukkan potensi yang besar dalam meningkatkan ketepatan dan kecekapan pengesanan hadis (Luthfi et al. 2022). Walaupun terdapat beberapa cabaran seperti keperluan untuk set data yang lebih seimbang dan kesukaran dalam memproses bahasa Arab, kajian-kajian ini telah membuka laluan untuk penyelidikan masa depan dalam pengesanan hadis yang lebih holistik dan menyeluruh. Secara keseluruhannya, pendekatan yang lebih interaktif dan inovatif diperlukan untuk terus memperbaiki teknik pengesanan dan memastikan keaslian kandungan hadis dapat dipelihara dengan lebih baik.

### 3. Perbincangan

Bidang pengesanan hadis secara digital masih berada pada peringkat awal perkembangan, dengan jumlah kajian yang terhad dan banyak ruang untuk diteroka (Luthfi et al. 2022; Mohamed & Sarwar 2021). Kajian-kajian yang ada kebanyakannya menawarkan kerangka teori untuk mengklasifikasikan hadis sebagai sahih atau palsu, namun pelaksanaan praktikal bagi kerangka ini masih terhad dan tidak menyeluruh. Fokus utama kajian biasanya terarah kepada pengesanan hadis melalui analisis rantaian perawi dengan menitikberatkan aspek kepercayaan dan kesinambungan para perawi. Walau bagaimanapun, aspek lain seperti kepercayaan perawi berdasarkan kriteria peribadi atau analisis kandungan matan (teks hadis) sering kali diabaikan atau tidak diberikan penekanan yang sewajarnya. Walaupun terdapat usaha untuk menerapkan peraturan berasaskan analisis matan, kajian-kajian ini masih kurang mendalam dan terbatas dalam skopnya.

Bagi menghadapi pelbagai cabaran dalam pengesanan hadis dalam bentuk digital, para penyelidik telah mula mengaplikasikan teknik pembelajaran mesin untuk membantu dalam pengelasan hadis namun kaedah ini juga mempunyai keterbatasan dari segi pemahaman kontekstual dan nuansa bahasa Arab yang kompleks (Abdelkader et al. 2019; Nassif et al. 2022). Oleh itu, untuk menangani cabaran dalam pengesanan hadis digital,



pendekatan hibrid yang menggabungkan teknik pembelajaran mesin dengan analisis manual atau separa manual boleh digunakan. Gabungan kaedah pembelajaran mesin dan kaedah manual dalam pengesahan hadis dapat mengintegrasikan teknologi automatik dan analisis pakar untuk mencapai ketepatan yang lebih tinggi dalam sistem pengesahan. Model pembelajaran mesin dapat berfungsi dalam pra-pemrosesan dan klasifikasi hadis dan menggunakan keupayaannya untuk mengenal pasti corak dalam jumlah data yang besar. Bagaimanapun, pengesahan secara manual oleh pakar masih diperlukan bagi memastikan ketepatan konteks dan nuansa hadis, terutama dalam kes kompleks yang mungkin tidak dapat dikendalikan sepenuhnya oleh model automatik sahaja. Dengan menggabungkan keupayaan pengecaman pola tertentu dalam pembelajaran mesin dengan sensitiviti dan pengetahuan mendalam pakar manusia, sistem ini boleh memastikan hasil yang lebih tepat dan komprehensif. Model automatik dapat mengendalikan tugas rutin dan yang mudah serta dapat mengurangkan beban kerja manual. Manakala pakar hadis pula dapat menumpukan pada kes yang memerlukan penilaian yang lebih mendalam. Pendekatan ini dijangka dapat meningkatkan ketepatan dan kebolehpercayaan hasil pengesahan dengan memanfaatkan kelebihan kedua-dua pendekatan teknologi dan tradisional.

Selain daripada kaedah pembelajaran mesin, kaedah pembelajaran mendalam, yang merupakan cabang kecerdasan buatan, turut menawarkan potensi besar dalam analisis dan pengesahan hadis. Pembelajaran mendalam berfungsi dengan melatih algoritma untuk mengenal pasti dan mengesahkan kesahihan hadis berdasarkan data yang telah dilatih sebelumnya (Hakak et al. 2022; Najeeb 2021). Ini dilakukan dengan menggunakan rangkaian neural yang kompleks untuk memproses maklumat dan membuat ramalan atau keputusan mengenai data yang baru. Sebagai contoh, algoritma pembelajaran mendalam boleh dilatih untuk mengenal pasti ciri-ciri khusus dalam teks hadis yang menunjukkan sama ada ia sahih atau tidak, berdasarkan pola yang telah dipelajari daripada set data latihan (Najeeb 2021). Proses ini membolehkan pengesahan yang automatik dan berkesan terhadap kesahihan hadis tanpa memerlukan campur tangan manusia secara langsung pada setiap langkah.

Di samping itu, pengkomputeran pinggir muncul sebagai pendekatan pengkomputeran masa depan yang lebih efisien dan responsif berbanding pengkomputeran awan. Berbanding pengkomputeran awan yang bergantung pada pusat data jauh untuk memproses maklumat, pengkomputeran pinggir memproses data lebih dekat kepada pengguna atau sumber data. Ini mengurangkan kelewatan dalam pemindahan dan pemrosesan data kerana ia mengurangkan jarak fizikal dan masa yang diperlukan untuk data bergerak antara pengguna dan pusat data (Marozzo et al. 2022). Dalam konteks pengesahan hadis, banyak kandungan hadis kini tersedia melalui aplikasi mudah alih dan laman web yang dapat diakses melalui Internet (Ahmad et al. 2019). Pengkomputeran pinggir memanfaatkan kelebihan ini dengan membolehkan pengesahan kandungan hadis dilakukan secara masa nyata. Dengan pemrosesan data segera dan alat analisis data yang terletak berhampiran dengan pengguna, pengkomputeran pinggir dapat meningkatkan kelajuan dan ketepatan dalam proses pengesahan, menjadikannya lebih responsif kepada perubahan dan keperluan pengguna (Marozzo et al. 2022). Ini memastikan bahawa verifikasi kandungan hadis dapat dilakukan dengan cepat dan berkesan, meningkatkan kebolehpercayaan sistem dalam menyediakan maklumat yang sahih.

Data di Internet berkembang pesat setiap hari, termasuk kandungan hadis yang banyak dan perlu diubah ke bentuk digital serta disahkan. Sebahagian besar kandungan hadis tersedia dalam bahasa Arab, dan setelah di digital, jumlah data yang terhasil boleh mencapai unit Gigabait (GB) atau Terabait (TB). Walaupun terdapat alat analisis data besar yang sedia ada, terdapat cabaran khusus dalam pemrosesan bahasa semula jadi (NLP) berkaitan teks Arab yang perlu diatasi untuk memastikan ketepatan dan keberkesanan analisis (Abdelkader et al. 2019; Bhardwaj et al. 2022; Nassif et al. 2022). Cabaran utama termasuk penggunaan diakritik, gaya penulisan, tafsiran yang berbeza dan proses digital (Abdelkader et al. 2019). Diakritik yang merupakan tanda khas dalam bahasa Arab untuk menunjukkan sebutan vokal, memainkan peranan penting dalam menentukan makna perkataan. Ketiadaan atau ketidaktepatan diakritik boleh mengakibatkan kekeliruan dalam analisis NLP, memerlukan algoritma yang dilatih untuk memproses teks dengan dan tanpa diakritik .

Gaya penulisan dalam teks hadis mungkin berbeza-beza, mempengaruhi keupayaan model NLP untuk memahami dan menganalisis teks dengan tepat. Variasi dalam dialek dan struktur ayat memerlukan model yang dilatih pada korpus yang pelbagai untuk menangani gaya penulisan yang berbeza (Haque et al. 2020). Tafsiran hadis yang berbeza berdasarkan konteks sejarah, budaya, dan metodologi juga mempengaruhi analisis NLP, memerlukan model untuk memahami konteks dan nuansa yang berbeza (Nassif et al. 2022). Proses mengubah teks hadis dari bentuk fizikal ke digital melibatkan penggunaan OCR, yang mungkin menghadapi kesukaran



dengan huruf dan tulisan tangan. Penggunaan perisian OCR khusus untuk bahasa Arab dan memastikan kualiti imbasan yang tinggi adalah penting untuk mengurangkan kesalahan (Varshney & Vishwakarma 2023). Secara keseluruhan, pengendalian data hadis dalam bahasa Arab memerlukan pendekatan yang menyeluruh untuk mengatasi cabaran dalam pemprosesan bahasa semula jadi dan memastikan analisis yang tepat dan berkesan.

#### **4. Kesimpulan**

Kesahihan hadis ialah isu penting yang perlu segera diselesaikan dalam masyarakat Islam. Penelitian dalam mengesahkan hadis masih terbatas kerana isu ini sangat rumit. Secara keseluruhan, bidang pengesahan hadis secara digital berada pada fasa awal perkembangan dengan banyak ruang untuk penambahbaikan dan eksplorasi lanjut. Kajian-kajian lepas menunjukkan bahawa pendekatan sedia ada, termasuk penggunaan teknik pembelajaran mesin, telah memberikan kerangka teori yang berguna untuk mengklasifikasikan hadis sebagai sahih atau palsu, namun aplikasi praktikalnya masih terhad. Kajian-kajian ini umumnya memfokuskan kepada analisis rantaian perawi, dengan menitikberatkan aspek kepercayaan dan kesinambungan para perawi, tetapi sering kali mengabaikan aspek lain seperti kepercayaan perawi berdasarkan kriteria peribadi dan analisis matan.

Teknik pembelajaran mesin telah menunjukkan potensi dalam menguruskan jumlah data hadis yang besar dan dalam mengklasifikasikan hadis secara automatik. Walau bagaimanapun, kajian-kajian lepas juga menunjukkan bahawa teknik ini menghadapi cabaran dalam memahami nuansa bahasa Arab dan kontekstual yang kompleks. Ini termasuk penggunaan diakritik, variasi gaya penulisan, dan tafsiran yang berbeza, yang sering kali menyebabkan kekeliruan dalam analisis. Kajian terdahulu menunjukkan bahawa walaupun pembelajaran mesin boleh memproses data dalam skala besar, ia memerlukan pemahaman mendalam yang sering kali tidak dapat dicapai tanpa campur tangan manusia.

Dalam usaha untuk mengatasi cabaran ini, terdapat juga kajian memberikan perhatian kepada penggunaan pembelajaran mendalam yang menawarkan potensi besar dengan kemampuan untuk melatih algoritma yang boleh mengenal pasti dan mengesahkan kesahihan hadis berdasarkan ciri-ciri khusus dalam teks. Walau bagaimanapun, pembelajaran mendalam juga memerlukan latihan yang teliti dan penggunaan data yang luas untuk mencapai ketepatan yang tinggi. Pengkomputeran pinggir, sebagai teknologi masa depan, juga muncul sebagai penyelesaian yang menjanjikan untuk meningkatkan kecekapan dan ketepatan pengesahan hadis dengan memproses data lebih dekat kepada sumbernya dan mengurangkan kelewatan. Ini adalah relevan dalam konteks di mana banyak kandungan hadis boleh diakses melalui aplikasi mudah alih dan laman web.

Oleh itu, pendekatan hibrid yang menggabungkan keupayaan teknologi automatik dengan analisis manual atau separa manual kelihatan sebagai jalan penyelesaian yang paling berpotensi. Model pembelajaran mesin dapat mengendalikan tugas-tugas rutin dan pengelasan awal, sementara pakar manusia dapat memberikan penilaian mendalam untuk kes-kes yang kompleks dan memerlukan konteks yang lebih terperinci. Gabungan ini dapat meningkatkan ketepatan dan kebolehpercayaan sistem pengesahan dengan memanfaatkan kekuatan kedua-dua pendekatan teknologi dan tradisional.

Secara keseluruhan, kejayaan dalam pengesahan hadis digital memerlukan pendekatan yang menyeluruh dan inovatif. Mengintegrasikan teknologi canggih dengan kepakaran manusia dan mengatasi cabaran dalam pemprosesan bahasa semula jadi adalah kunci untuk mencapai sistem pengesahan hadis yang lebih tepat, boleh dipercayai dan juga berkesan.

#### **5. Penghargaan**

Kertas ini adalah hasil penyelidikan yang disokong dan dibiayai oleh Skim Geran Penyelidikan Fundamental (FRGS) Kementerian Pengajian Tinggi Malaysia (Kod: FRGS/1/2022/ICT02/UKM/02/3).

#### **6. Rujukan**



- Abdelkader, A., Najeeb, M., Alnamari, M. & Malik, H. 2019. How Existing Nlp Tools of Arabic Language Can Serve Hadith Processing. *International Journal of Computer Engineering and Technology* 10(6). doi:10.34218/ijcet.10.6.2019.003
- Abur Hamdi, U. & Mohd Norzi, N. 2022. Penyebaran Hadis Palsu dalam Penerbitan di Malaysia dan Kesannya kepada Masyarakat. *Journal Hadis* 12(23): 34–50. doi:<https://doi.org/10.53840/hadis.v12i23.182>
- Ahmad, K., Suliaman, I., Ahmad Shah, F. & Mohd Rokhibi, I. A. 2019. Kriteria Utama dalam Pembangunan Laman Web Berkaitan Hadith: Satu Analisis dan Cadangan. *Ma'ālim al-Qur'ān wa al-Sunnah* 15(2): 28–53. doi:10.33102/jmq.s.v15i2.176
- Bahrudin, M., Suliaman, I. & Mohamed Adnan, M. A. 2023. Systematic Literature Review on Hadith Authentication Between Year 2000 To 2021. *Journal Hadis* 13(25): 55–69. doi:10.53840/hadis.v13i25.218
- Belle, V. & Papantonis, I. 2021. Principles and Practice of Explainable Machine Learning. *Frontiers in Big Data* 4(July): 1–25. doi:10.3389/fdata.2021.688969
- Bhardwaj, R., Vaidya, T. & Poria, S. 2022. Towards solving NLP tasks with optimal transport loss. *Journal of King Saud University - Computer and Information Sciences* 34(10): 10434–10443. doi:10.1016/j.jksuci.2022.10.031
- Binbeshr, F., Kamsin, A. & Mohammed, M. 2021. A Systematic Review on Hadith Authentication and Classification Methods. *ACM Transactions on Asian and Low-Resource Language Information Processing* 20(2). doi:10.1145/3434236
- Fadele, A. A., Kamsin, A., Ahmad, K. & Hamid, H. 2021. A novel classification to categorise original hadith detection techniques. *International Journal of Information Technology*. doi:10.1007/s41870-021-00649-3
- Gasparetto, A., Marcuzzo, M., Zangari, A. & Albarelli, A. 2022. A Survey on Text Classification Algorithms: From Text to Predictions. *Information (Switzerland)* 13(2): 1–39. doi:10.3390/info13020083
- Hakak, S., Kamsin, A., Zada Khan, W., Zakari, A., Imran, M., bin Ahmad, K. & Amin Gilkar, G. 2022. Digital Hadith authentication: Recent advances, open challenges, and future directions. *Transactions on Emerging Telecommunications Technologies* 33(6): 1–14. doi:10.1002/ett.3977
- Haque, F., Orthy, A. H. & Siddique, S. 2020. Hadith Authenticity Prediction using Sentiment Analysis and Machine Learning. *14th IEEE International Conference on Application of Information and Communication Technologies, AICT 2020 - Proceedings* (March 2022). doi:10.1109/AICT50176.2020.9368569
- Küçük, D. & Can, F. 2021. Stance Detection: Concepts, Approaches, Resources, and Outstanding Issues. *SIGIR 2021 - Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval* 2673–2676. doi:10.1145/3404835.3462815
- Luthfi, E. T., Yusoh, Z. I. M. & Aboobaidar, B. M. 2022. BERT based Named Entity Recognition for Automated Hadith Narrator Identification. *International Journal of Advanced Computer Science and Applications* 13(1): 604–611. doi:10.14569/IJACSA.2022.0130173
- Manas, S. A. A. 2022. Asas Kenali Hadis, hlm. 4th Edisi . Kuala Lumpur: MALAHA PLT.
- Marozzo, F., Orsino, A., Talia, D. & Trunfio, P. 2022. Edge Computing Solutions for Distributed Machine Learning. *Proceedings of the 2022 IEEE International Conference on Dependable, Autonomic and Secure Computing, International Conference on Pervasive Intelligence and Computing, International Conference on Cloud and Big Data Computing, International Conference on Cy* 1–8. doi:10.1109/DASC/PiCom/CBDCoM/Cy55231.2022.9927824
- Mohamed, E. & Sarwar, R. 2021. Linguistic features evaluation for hadith authenticity through automatic machine learning. *Digital Scholarship in the Humanities*. doi:10.1093/lc/fqab092
- Najeeb, M. M. A. 2021. Towards a Deep Learning-based Approach for Hadith Classification. *European Journal of Engineering and Technology Research* 6(3): 9–15. doi:10.24018/ejers.2021.6.3.2378
- Najeeb, M. M. A. 2022. A Hidden Markov Model-Based Tagging Approach for Arabic Isnads of Hadiths. *Mathematical Problems in Engineering* 2022: 1–12. doi:10.1155/2022/7160509
- Nassif, A. B., Elnagar, A., Elgendy, O. & Afadar, Y. 2022. Arabic Fake News Detection Based On Deep Contextualized Embedding Models. *Neural computing & applications* 4: 1–14. doi:10.1007/s00521-022-07206-4
- Torabi Asr, F. & Taboada, M. 2019. Big Data and quality data for fake news and misinformation detection. *Big Data and Society* 6(1): 1–14. doi:10.1177/2053951719843310
- Varshney, D. & Vishwakarma, D. K. 2023. Framework for detection of probable clues to predict misleading information proliferated during COVID-19 outbreak. *Neural Computing and Applications* 35(8): 5999–6013. doi:10.1007/s00521-022-07938-3
- Wang, W., Wang, S., Fan, W., Liu, Z. & Tang, J. 2020. Global-and-local aware data generation for the class



Published biannually by:  
**HADITH AND AQIDAH RESEARCH INSTITUTE (INHAD),  
Selangor International Islamic University College (KUIS)**  
Email: [jurnalhadis@kuis.edu.my](mailto:jurnalhadis@kuis.edu.my)  
Web: [www.jurnal.kuis.edu.my/hadis/](http://www.jurnal.kuis.edu.my/hadis/)  
Bandar Seri Putra, 43600, Bangi, Selangor (Darul Ehsan) Malaysia.  
Tel: 03-8911 7000 Ext: 6129/6130, Fax: 03-8926 6279  
**Vol. 14, No. 28 (December 2024)**

imbalance problem. *Proceedings of the 2020 SIAM International Conference on Data Mining, SDM 2020* 307–315.