

Klasifikasi Teks Hukum Fiqih Islam menggunakan Multinomial Naïve Bayes

Riyan Wibowo Saputra^a, Benny Sukma Negara^a, Rahmad Kurniawan^{a,b},
Muhammad Irsyad^a, Iis Afrianty^a, Fitra Lestari^b

^aProgram Studi Teknik Informatika, Fakultas Sains dan Teknologi
Universitas Islam Negeri Sultan Syarif Kasim Riau
Jl. HR. Soebrantas, No. 155, Pekanbaru Riau;

^bCenter of Islamic Data Science and Continuous Improvement (CIDSCI),
Universitas Islam Negeri Sultan Syarif Kasim Riau
Jl. HR. Soebrantas, No. 155, Pekanbaru Riau;

e-mail: riyan.wibowo.saputra@students.uin-suska.com, bsnegara@uin-suska.ac.id,
rahmadkurniawan@uin-suska.ac.id, iis.afrianty@uin-suska.ac.id, irsyadtech@uin-suska.ac.id

Abstrak

Indonesia merupakan negara dengan populasi umat muslim terbesar di dunia. Islam tidak hanya mengajarkan tentang ibadah tetapi juga masalah kehidupan seorang muslim yang diatur dalam hukum Islam (Fiqh). Ilmu hukum Islam dapat diketahui dengan cara belajar kepada seorang ulama bidang Fiqih atau membaca buku dari ulama tersebut. Namun, tidak semua orang dapat bertemu ulama dan dapat memahami isi buku dari ulama tersebut. Oleh karena itu, diperlukan sebuah sistem klasifikasi teks yang dapat digunakan umat untuk pertimbangan jawaban dari hukum Islam yang ditanyakan. Algoritme Multinomial Naïve Bayes dipilih sebagai metode untuk menyimpulkan jawaban hukum Islam karena ketepatannya sebagai mesin inferensi. Buku yang ditulis oleh pakar Fiqih Asia Tenggara yaitu Ustaz Abdul Somad yang berjudul "37 Masalah Populer, 77 Tanya Jawab Tentang Shalat dan 33 Tanya Jawab Seputar Kurban" telah digunakan sebagai basis pengetahuan dalam aplikasi berbasis web ini. Berdasarkan hasil pengujian yang dilakukan, didapatkan sistem klasifikasi teks hukum Islam ini menghasilkan akurasi sebesar 75%. Berdasarkan eksperimen dan pengujian tersebut, dapat disimpulkan bahwa sistem klasifikasi teks berdasarkan hukum Islam ini berpotensi digunakan sebagai pertimbangan dalam memahami ilmu Fiqih.

Kata kunci: Fiqih, Hukum Islam, Multinomial Naïve Bayes, Sistem Klasifikasi Teks

Abstract

Indonesia is a country with the most Muslim population in the world. Islam teaches about worship and about the life of a Muslim, which is regulated in Islamic Jurisprudence (Fiqh). The knowledge of Islamic Jurisprudence can be obtained from an Islamic preacher (Fiqh expert) or reading a book from that Islamic preacher. However, not everyone can meet the ulama and can understand the contents of the book. Therefore, it needs a text classification system that the congregation can use to consider the Question and Answer regarding Islamic Jurisprudence. The Multinomial Naïve Bayes algorithm was employed to infer answers to Islamic Jurisprudence because of its accuracy as an inference engine. A Southeast Asian Fiqh expert wrote the book, i.e., Ustaz Abdul Somad, entitled "The 37 Popular Problems, the 77 Questions, and Answers About Prayer and the 33 Questions and Answers About Qurban," which has been used as a knowledge base in this web-based application. Based on the evaluation results, it was obtained that this study produced an accuracy of 75%. Based on these experiments and evaluations, it can be concluded that this study can be used as a consideration in understanding Islamic Jurisprudence (Fiqh).

Keywords: Fiqh, Islamic Jurisprudence, Multinomial Naïve Bayes, Text Classification System

1. Pendahuluan

Indonesia sebagai negara dengan populasi muslim terbesar di dunia, di mana pendidikan Islam bukan hanya sekedar pendidikan dalam bidang akademik saja, tetapi Islam juga mengatur permasalahan kehidupan yang tertuang dalam hukum fiqih [1]. Salah satu cara seorang muslim dapat memahami hukum fiqih adalah dengan bimbingan seorang ulama.

Seorang ulama Islam yang berpengaruh dan mempunyai banyak pengikut adalah Ustaz Abdul Somad atau juga dikenal dengan Syaikh Abdul Somad (UAS) [2]. UAS telah menerbitkan beberapa buku yang mengajarkan tentang fiqh dengan beberapa mazhab yang dimuat di dalam buku tersebut. Namun, dalam penelitian Halimah Mohamed [3] menjelaskan bahwa 80% orang mengalami masalah tentang memahami informasi yang ada di buku, internet, ataupun pelatihan. Oleh karena itu sebuah teknik pembelajaran mesin dianggap perlu dalam mengatasi permasalahan ini.

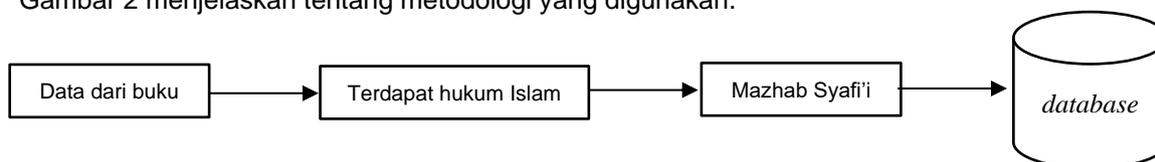
Pembelajaran mesin merupakan sebuah bidang ilmu yang telah banyak diterapkan di berbagai bidang seperti keuangan, militer, medis dan dalam ilmu Islam. Kecerdasan buatan dapat membantu kegiatan manusia lebih efisien [4]. Salah satu bentuk penerapan dari pembelajaran mesin adalah klasifikasi teks. Klasifikasi teks merupakan bagian dari pembelajaran mesin yang mempelajari tentang cara mengklasifikasikan sebuah data tidak terstruktur. Sekarang ini, data sudah mulai beralih dari data yang terstruktur menjadi semi struktur dan tidak terstruktur [2]. Klasifikasi teks telah banyak digunakan termasuk dalam bidang Islam [2].

Beberapa studi tentang klasifikasi teks yang digunakan [5][6][7][8], salah satunya adalah Penelitian yang dilakukan untuk membandingkan algoritme *Naïve Bayes*, *Support Vector Machine*, *K-Nearest Neighbour* pada yang tidak terstruktur, serta *Deep Learning* untuk pengklasifikasian teks arab dengan *Naïve Bayes*. Penelitian tersebut mendapatkan hasil terbaik dengan akurasi mencapai 90% [3], penelitian lainnya dilakukan untuk perbandingan algoritme *Multivariate Bernoulli Naïve Bayes* dan *Multinomial Naïve Bayes* dalam penentuan sentimen terhadap berita. Apakah suatu berita itu bermakna negatif atau positif. Berdasarkan penelitian tersebut, algoritme *Multinomial Naïve Bayes* lebih akurat daripada algoritme *Multivariate Bernoulli Naïve Bayes* [9]. Mengklasifikasikan hukum Islam menggunakan teknik pembelajaran mesin memiliki tantangan yang tinggi karena terdapat empat mazhab dalam Islam, dan setiap mazhab memiliki perbedaan pendapat dalam penentuan suatu hukum [3].

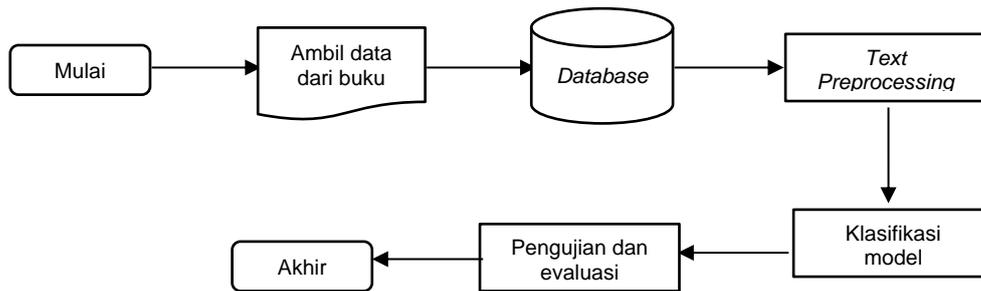
Mazhab merupakan cara pikir, aliran yang menjadi sebuah pedoman dalam menentukan sebuah hukum dalam Islam, di dalam hukum fiqh sendiri terdapat empat mazhab yang populer yaitu mazhab Hanafi, Hambali, Maliki dan Syafi'i. Sedangkan mayoritas umat Islam di Asia tenggara mengikuti mazhab Syafi'i. [10] Oleh karena itu, berdasarkan hasil studi pustaka di atas, maka perlu membangun sebuah sistem yang dapat mengklasifikasikan hukum fiqh Islam menggunakan algoritme *Multinomial Naïve Bayes*.

2. Metode Penelitian

Data latih sebagai basis pengetahuan diambil dari beberapa buku tentang fiqh Islam yang telah ditulis oleh Ustaz Abdul Somad yaitu *37 masalah populer*, *77 tanya jawab tentang shalat* dan *33 tanya jawab seputar qurban*. Dalam pengambilan data, ada dua bagian penting dalam buku tersebut yaitu pertanyaan dan jawaban. Pertanyaan merupakan kumpulan pertanyaan yang berkaitan dengan sebuah topik, jawaban merupakan sebuah jawaban dari sebuah pertanyaan yang terkadang tidak memiliki jawaban secara eksplisit menjelaskan hukum dari pertanyaan tersebut. Berdasarkan teknik pra pemrosesan dalam pengambilan data, maka telah diperoleh 34 topik hukum Islam. Gambar 1 menjelaskan tentang cara menyeleksi dan Gambar 2 menjelaskan tentang metodologi yang digunakan.



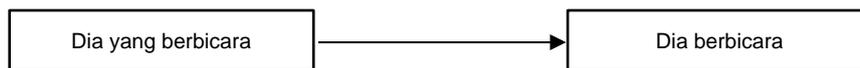
Gambar 1. Seleksi Data Hukum Islam dari Buku Ustadz Abdul Somad



Gambar 2. Metodologi Sistem Klasifikasi Teks Fiqh Hukum Islam

2.1. Pra Pemrosesan Teks

Untuk membuat data latih yang berkualitas dilakukan teknik pra pemrosesan teks yang terdiri dari beberapa tahapan yaitu dimulai dengan mengubah semua huruf menjadi huruf kecil. Tahapan ini dilakukan untuk menghilangkan data yang *redundant*. Langkah kedua yaitu melakukan penghapusan kata tidak penting yang sering muncul pada kalimat atau *stopword* seperti yang dicontohkan pada gambar 3. Langkah ketiga dilakukan pengembalian kata dari kata berimbuhan menjadi kata dasar atau disebut *stemming* yang dicontohkan pada Gambar 4. Langkah terakhir memotong kalimat menjadi satuan kata atau tokenisasi.



Gambar 3. Proses menghilangkan *Stopword*



Gambar 4. Proses *Stemming*

2.2. Klasifikasi Model

Penelitian ini menggunakan algoritme *Multinomial Naïve Bayes* sebagai algoritme dalam melakukan klasifikasi teks. *Multinomial Naïve Bayes* merupakan algoritme yang berdasarkan pembelajaran probabilitas. Kemungkinan kemunculan sebuah dokumen (d) dalam sebuah kelas (c) dapat dituliskan dengan formula:

$$P(c|d) = P(c) \prod_{1 \leq k \leq n_d} P(t_k|c) \quad (1)$$

Di mana $P(t_k|C)$ merupakan kemungkinan munculnya kata dalam dokumen di kelas c . $P(c)$ merupakan kemungkinan munculnya dokumen dalam kelas c . Jika sebuah dokumen tidak bisa memastikan kata tersebut berada dalam kelas yang mana maka akan dibandingkan kata ($t_1, t_2 \dots t_{n_d}$) dalam token yang ada pada setiap dokumen.

Untuk klasifikasi teks menggunakan *multinomial naïve bayes* lebih berfokus pada kemungkinan terbesar teks itu muncul di suatu dokumen atau sering disebut *maximum a posteriori (MAP)* dengan formula:

$$C_{map} = \arg \max_{c \in C} [\log P(c) + \sum_{1 \leq k \leq n_d} \log P(t_k|c)] \quad (2)$$

Dan untuk menghindari terjadinya kemungkinan kata yang tidak muncul dalam dokumen yang menyebabkan hasilnya 0 maka digunakan *laplace smoothing* pada perhitungan probabilitas perhitungan kata dengan formula.

$$P(t|c) = \frac{t_{ct} + 1}{(\sum_{t \in V} t_{ct}) + B} \quad (3)$$

Di mana $B = |V|$ merupakan jumlah kata yang ada pada sebuah dokumen tersebut. Untuk melakukan pelatihan data pada *Multinomial Naïve Bayes* dapat menggunakan sebuah *Pseudocode* di bawah ini:

```

TrainMultinomialNB(C, D)
1. V ← getJumlahKata(D)
2. N ← HitungDokumen(D)
3. for each c ∈ C
4. do Nc ← HitungDokumenDikelas(D, c)
5.   prior[c] ← Nc/N
6.   textc ← GabungTeksSemuaKelas(D, c)
7.   for each t ∈ V
8.     do Tct ← hitungToken(textc, t)
9.     for each t ∈ V
10.    do prob[t][c] ←  $\frac{t_{ct}+1}{(\sum_{t \in V} t_{ct})+B}$ 
11. return V, prior, prob
    
```

Sedangkan untuk melakukan prediksi dapat dilakukan dengan menggunakan *Pseudocode* di bawah ini:

```

PredictMultinomialNB(C, V, prior, prob, d)
1. W ← ambilTokenDariDokumen(V, d)
2. for each c ∈ C
3. do score[c] ← log prior[c]
4.   for each t ∈ W
5.     do score[c] += log prob[t][c]
6. return arg max(score[c])
    
```

2.3 Pengujian dan Evaluasi

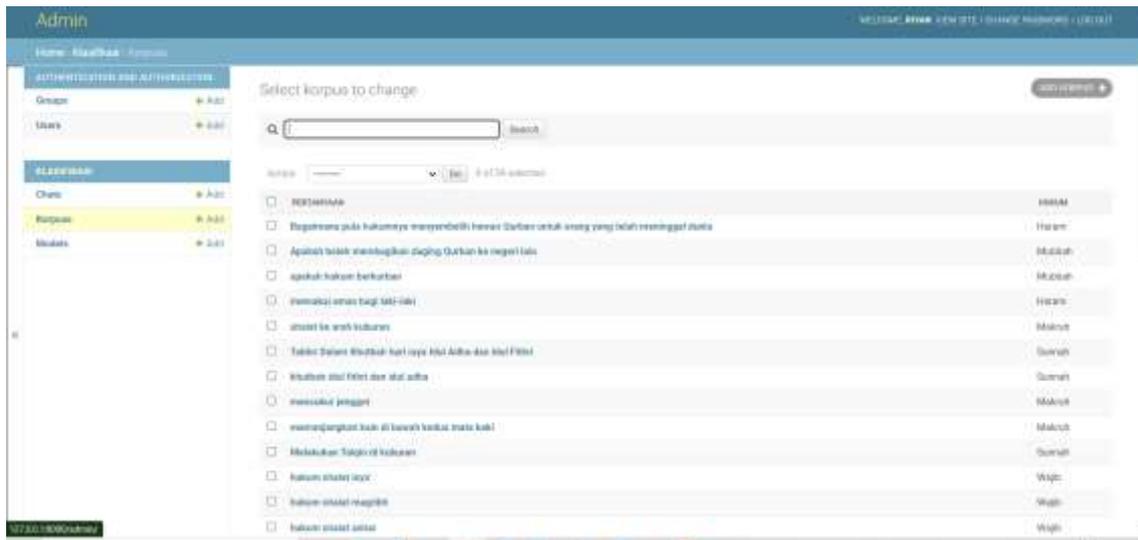
Dalam pengujian dilakukan proses percobaan prediksi dan melakukan proses akurasi dengan *confusion matrix*.

3. Hasil dan Pembahasan

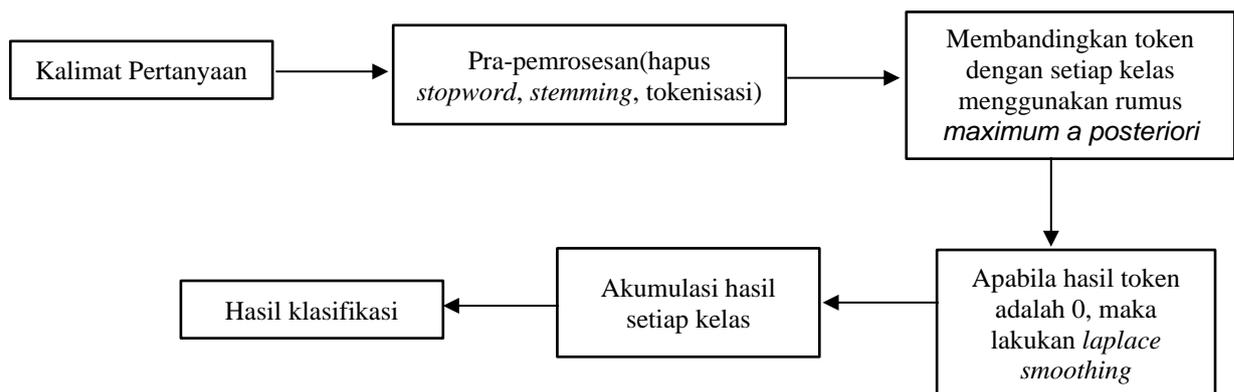
Penulis telah berhasil melakukan implementasi terhadap penelitian tentang menerapkan algoritme *Multinomial Naïve Bayes* untuk teks klasifikasi hukum Fiqih.

3.1. Hasil Implementasi Sistem

Implementasi sistem menggunakan *Python* dengan *framework Django*, sistem terdiri dari beberapa bagian menu, di antaranya menu *chat* yang berfungsi menyimpan data data teks prediksi yang telah dilakukan oleh pengguna. Pada menu korpus merupakan menu untuk menyimpan data latihan dan menu model yang menyimpan model *Multinomial Naïve Bayes* yang dapat disimpan dan dipakai di sistem yang lain bila diperlukan. Untuk klasifikasi model pada sistem dijelaskan pada gambar 6.



Gambar 5. Sistem Klasifikasi Teks Hukum Islam Berbasiskan Web



Gambar 6. Proses Klasifikasi dalam Sistem

3.2. Analisa Pengujian Prediksi

Pada tahapan ini merupakan hasil pengujian untuk melihat bagaimana sistem dapat memprediksi teks yang diajukan oleh jemaah. Pertanyaan yang didapatkan dari Jemaah merupakan pertanyaan yang telah diajukan sebelumnya melalui sebuah *chatbot* bernama Azka Hasub, pertanyaan dibatasi kepada tiga tema yaitu tentang salat, qurban, dan tata cara berbusana. Pertanyaan yang didapatkan umumnya merupakan pertanyaan yang tidak terstruktur polanya. Penulis mengambil 100 pertanyaan untuk percobaan kali ini.



Gambar 7. Hasil Prediksi Sistem mengenai Pertanyaan yang diajukan Jemaah

Tabel 1 menunjukkan contoh pertanyaan yang diajukan ke sistem klasifikasi teks tersebut dengan menggunakan 100 pertanyaan. Berdasarkan evaluasi, sistem klasifikasi teks mendapatkan hasil 75 jawaban benar dan 25 jawaban salah. Sehingga akurasi sistem klasifikasi teks ini sebesar 75%. Tabel 2 menunjukkan *Confusion Matrix*.

Tabel 1. Contoh Pertanyaan yang diajukan ke dalam Sistem

Pertanyaan	Jawaban	Ketepatan
hukum shalat jumat	Wajib	Benar
hukum berpuasa	Wajib	Benar
hukum meninggalkan shalat jumat	Wajib	Salah
menggunakan sayyidina pada nama nabi	Sunah	Benar
...
hukum berdoa qunut pada shalat subuh	Sunah	Benar

Tabel 2. Akurasi *Confusion Matrix*

Hasil Sebenarnya	Prediksi Wajib	Prediksi Sunah	Prediksi Mubah	Prediksi Makruh	Prediksi Haram
Wajib	30	2	2	0	1
Sunah	4	17	3	1	0
Mubah	0	1	11	0	0
Makruh	2	4	0	11	0
Haram	3	2	0	0	6

$$\text{Akurasi} = \frac{30+17+11+11+6}{30+2+2+0+1+4+17+3+1+0+0+1+11+0+0+2+4+0+11+0+3+2+0+0+6} = \frac{75}{100} = 0.75 * 100\% = 75\%$$

Pada tabel 2 menunjukkan bahwasanya sistem dapat memprediksi masalah hukum fiqih Islam yang benar yaitu, 30 hukum wajib, 17 hukum sunah, 11 hukum mubah, 11 hukum makruh dan 6 hukum haram dari 100 pertanyaan yang diajukan.

4. Kesimpulan

Sebuah sistem prediksi hukum fiqih Islam dengan *Multinomial Naïve Bayes* telah berhasil dibangun. Berdasarkan hasil pengujian dengan basis pengetahuan dari tiga buku karya Ustadz Abdul Somad, maka sistem klasifikasi teks ini dapat memprediksi hukum fiqih Islam dengan ketepatan 75%. Pada penelitian ini hanya menggunakan sumber mazhab Syafi'i saja. Untuk penelitian selanjutnya diharapkan dapat menggunakan mazhab yang lain sebagai basis pengetahuan dan menggunakan fungsi *tagging* untuk melakukan pelatihan berbasis multi kelas.

Daftar Pustaka

- [1] R. Kurniawan, Akbarizan, K. Jamal, A. Nur, M. Z. Ahmad, and D. Kholilah, "Advise-giving expert systems based on Islamic jurisprudence for treating drugs and substance abuse," *J. Theor. Appl. Inf. Technol.*, 2018.
- [2] K. Jamal, R. Kurniawan, A. S. Batubara, Z. A. Nazri, F. Lestari, and P. Papilo, "Text Classification Based On Islamic Jurisprudence Using Machine Learning Techniques," 2019.
- [3] H. Mohamed, S. Sulaiman, and M. Sabudin, "A hybrid of rule and frame based approach in solving hajj complex problems," *SoCPaR 2009 - Soft Comput. Pattern Recognit.*, pp. 466–471, 2009, doi: 10.1109/SoCPaR.2009.96.
- [4] Akbarizan, R. Kurniawan, M. Z. A. Nazri, S. N. H. S. Abdullah, S. Murhayati, and Nurcahaya, "Using Bayesian Network for Determining The Recipient of Zakat in BAZNAS Pekanbaru," in *2018 2nd International Conference on Electrical Engineering and Informatics (ICon EEI)*, Oct. 2018, pp. 12–17, doi: 10.1109/ICon-EEI.2018.8784142.
- [5] X. Lin, Z. Li, Y. Zhou, and Y. Xue, "Text classification algorithm study based on rough set theory," 2010, doi: 10.1109/IFITA.2010.203.
- [6] F. Miao, P. Zhang, L. Jin, and H. Wu, "Chinese News Text Classification Based on Machine Learning Algorithm," 2018, doi: 10.1109/IHMSC.2018.10117.
- [7] T. Yao, Z. Zhai, and B. Gao, "Text Classification Model Based on fastText," 2020, doi: 10.1109/ICAIS49377.2020.9194939.
- [8] F. Zhou, F. Zhang, B. Yang, and X. Yu, "Research on short text classification algorithm based on statistics and rules," 2010, doi: 10.1109/ISECS.2010.9.
- [9] A. McCallum and K. Nigam, "A Comparison of Event Models for Naive Bayes Text Classification," *AAAI/ICML-98 Work. Learn. Text Categ.*, 1998, doi: 10.1.1.46.1529.
- [10] M. Ikhsan, "Sejarah Mazhab Fiqih di Asia Tenggara," *J. Bid. Keislam.*, vol. 4, no. 2, pp. 187–203, 2018.