

BAB I

PENDAHULUAN

Bab ini menyajikan penjelasan secara sistematis mengenai pokok-pokok yang melatarbelakangi penelitian, dimulai dari identifikasi dan perumusan masalah hingga solusi teknologi yang ditawarkan. Uraian disusun secara sistematis untuk menunjukkan adanya permasalahan. Dimulai dengan menceritakan keresahan dan solusinya di latar belakang masalah, dilanjutkan dengan perumusan masalah, tujuan penelitian, manfaat yang diharapkan, serta ruang lingkup dan batasan penelitian. Melalui pembahasan ini, diharapkan pembaca dapat memperoleh pemahaman yang menyeluruh mengenai konteks permasalahan dan pendekatan yang digunakan dalam penyelesaiannya.

1.1. Latar Belakang

Al-Qur'an merupakan kitab suci umat Islam yang mengandung petunjuk hidup (hudan), pembeda antara yang benar dan salah (furqan), serta penyembuh hati (syifā'). Al-Qur'an sebagai kalamullah yang diturunkan kepada Nabi Muhammad Shallallahu 'alaihi Wassalam yang berfungsi sebagai petunjuk, rahmat dan pedoman hidup sepanjang masa baik dalam hal moral, akhlak dan kepribadian bagi manusia dalam menjalani kehidupannya. Tidak bisa dijadikan sebatas bacaan saja yang penuh dengan keutamaan, Al Qur'an mengandung banyak hikmah, pelajaran yang luar biasa, nilai spiritual, hukum hingga solusi bagi seluruh permasalahan kehidupan manusia bahkan di era modern ini. Oleh sebab itu, interaksi dengan Al Qur'an tidak cukup hanya pada aspek tilawah, tetapi juga harus mencakup aspek tadabbur, pemahaman dan pengamalan [1].

Namun, fenomena sosial hari ini menunjukkan adanya kesenjangan antara keberlimpahan akses terhadap Al-Qur'an dengan implementasi nilai-nilainya dalam kehidupan sehari-hari. Banyak umat Islam yang lebih fokus terhadap bacaan secara ritual, tetapi kurang dalam menginternalisasikan makna dan pesan yang terkandung di dalamnya, terutama menjadikannya sebagai solusi. Padahal dulu Rasulullah dan para sahabatnya itu mencontohkan bagaimana Al-Qur'an itu harus diposisikan sebagai sumber solusi. Hal yang mungkin dianggap remeh, sampai

persoalan yang besar solusinya dicari di dalam Al-Qur'an [1]. Kondisi kesenjangan hari ini ditambah era digital yang semakin modern berdampak pada lemahnya penghayatan spiritual, hilangnya ketenangan jiwa serta meningkatnya berbagai masalah sosial dan psikologi, seperti stress, kecemasan dan krisis identitas [2].

Merujuk sebuah tulisan yang dipublish oleh Kementerian Kesehatan Republik Indonesia pada bulan Oktober tahun 2023, usia 15 tahun keatas dalam kategori remaja yang terjangkit penyakit mental telah mencapai 6,1% dari penduduk Indonesia. Angka gangguan kesehatan ini mengalami peningkatan yang signifikan. Bahkan pada tulisan tersebut ada kutipan pernyataan dari dr. Khamelia Malik yang tergabung dalam PDSKJI (Perhimpunan Dokter Spesialis Kedokteran Jiwa Indonesia), beliau mengungkapkan tingkat kesakitan dan kematian akibat masalah kejiwaan pada kelompok remaja telah mengalami peningkatan hingga 200% dalam beberapa tahun terakhir. Di mana salah satu penyebab ialah ketidakberdayaan dalam mengontrol perilaku maupun menahan emosi. Ada tujuh faktor penyebab seseorang mendapatkan gangguan kesehatan atau penyakit mental menurut artikel yang dirilis Siloam Hospital. Faktor tersebut diantaranya krisis identitas, tekanan dari lingkungan sekitar, konflik dalam keluarga, pengalaman menjadi korban bullying, hubungan yang kurang harmonis dengan teman sebaya, kesulitan dalam aspek sosial ekonomi dan terpengaruh pemahaman atau informasi yang salah dari media sosial. Semua itu baru soalan penyakit mental, belum lagi soal penyimpangan lainnya seperti kerusakan akhlak dan sebagainya yang terjadi di Masyarakat [3].

Berdasarkan hasil penelitian oleh Badan Pusat Statistik yang berkolaborasi dengan Asosiasi Penyelenggara Jasa Internet Indonesia (APJII), tercatat bahwa jumlah pengguna internet di Indonesia terus meningkat secara signifikan sejak tahun 2013, yaitu sebesar 71,19%, dan naik menjadi 83,7% pada tahun 2014. Salah satu kelompok pengguna terbesar adalah remaja, yang cenderung menghabiskan banyak waktu dengan perangkat digital. Sebuah studi pada tahun 2016 mengungkapkan bahwa sekitar 42,4% remaja mengalami kecanduan internet, dan lebih dari 70% dari mereka mengakses konten negatif seperti game daring berlebihan, pornografi siber, hingga aktivitas kejahatan digital (*cybercrime*).

Penggunaan gadget yang berlebihan ini berpotensi memperburuk kondisi kesehatan mental di kalangan remaja [3].

Padahal di tahun 2020, sebuah pusat penelitian ternama yakni *Pew Research Center* pernah melakukan survei "The Global God Divide" pada 34 negara. Di mana hasilnya, Indonesia menempati posisi pertama sebagai negara paling religius. Di Indonesia, Islam adalah agama yang paling banyak pengikutnya. Menjadi mayoritas dengan jumlah muslim mencapai 87,08% atau sejumlah 245,93 juta jiwa berdasarkan data Kementerian Dalam Negeri (Kemendagri) per-Juni 2024. Tentu ini menjadi hal yang sangat serius, ada konsekuensi sebagai seorang muslim untuk memiliki arah hidup yang jelas karena Al-Qur'an adalah pedoman hidup yang memberikan petunjuk [4]. Kalau umat Islam menekuni tadabbur Al-Qur'an, niscaya akan mendapatkan solusi hidup yang kokoh dan ketenangan batin yang sejati. Maka upaya memahami dan mengakses ayat-ayat Al-Qur'an secara relevan dengan kebutuhan personal menjadi semakin penting, terutama di tengah derasnya arus informasi dan kompleksitas permasalahan hidup saat ini [5].

Seiring berkembangnya teknologi informasi, masyarakat dari level anak – anak, remaja hingga dewasa sebenarnya memiliki akses yang sangat luas terhadap Al-Qur'an. Baik itu melalui aplikasi digital, *website*, *platform* media sosial, terutama *gadget* yang hari ini bahkan menjadi kebutuhan *primer* karena pola hidup digital. Kemudahan akses tersebut memunculkan permasalahan baru, yaitu kesulitan dalam menemukan ayat-ayat yang relevan dengan kebutuhan atau topik tertentu. Al-Qur'an terdiri dari ribuan ayat dengan ragam tema yang saling berkaitan, sehingga pencarian ayat secara manual atau hanya berdasarkan ingatan pengguna menjadi tidak efisien, khususnya bagi pengguna yang tidak memiliki latar belakang keilmuan agama yang memadai. Pendekatan pencarian konvensional berbasis kata kunci sederhana juga sering kali menghasilkan ayat yang kurang relevan karena tidak mempertimbangkan kesamaan teks secara menyeluruh. Tingginya intensitas penggunaan perangkat digital di masyarakat mendorong kebutuhan akan sistem cerdas yang mampu membantu pengguna menyaring informasi secara efektif. Dalam konteks pencarian teks Al-Qur'an, diperlukan suatu mekanisme yang tidak hanya menampilkan ayat berdasarkan kecocokan kata secara literal, tetapi juga mampu mengelompokkan ayat-ayat berdasarkan kemiripan

konten sehingga proses pencarian menjadi lebih terarah. Disinilah muncul urgensi untuk menghadirkan sebuah sistem cerdas yang mengkombinasikan antara pendekatan agama dengan teknologi, yakni sebuah sistem yang merekomendasikan ayat-ayat Al-Qur'an secara tepat sesuai kebutuhan pengguna [6].

Sistem rekomendasi yang selama ini banyak dimanfaatkan dalam bidang *e-commerce* dan hiburan digital, kini mulai diterapkan dalam konteks keislaman untuk membantu pengguna menemukan konten yang sesuai. Prinsip utamanya adalah membantu pengguna menemukan informasi yang relevan dari sekian banyak data yang tersedia. Dalam konteks Al-Qur'an, sistem rekomendasi berpotensi menghadirkan ayat-ayat yang mirip, relevan secara leksikal-tematik bahkan diharapkan sampai bisa kontekstual berdasarkan kata kunci atau preferensi pengguna. Hal ini tidak hanya memudahkan eksplorasi Al-Qur'an secara relevan, tetapi juga meningkatkan kedekatan spiritual dan kebermanfaatan praktis bagi kehidupan sehari-hari. Sampai menghadirkan kembali budaya untuk menjadikan Al-Qur'an sebagai solusi mulai dari membaca dan mentadabburinya.

Sistem rekomendasi adalah suatu mekanisme yang dirancang untuk membantu pengguna dalam menyaring informasi yang berlimpah dengan memberikan saran atau rekomendasi yang sesuai secara spesifik. Tujuan utamanya adalah untuk menjawab kebutuhan dan preferensi pengguna secara lebih tepat. Terdapat beberapa pendekatan dalam sistem rekomendasi, antara lain metode berbasis konten (*Content-Based*), penyaringan kolaboratif (*Collaborative Filtering*), dan metode gabungan (*Hybrid Based*) [3].

Berdasarkan penelitian Putriando dan Sutanto (2024) mengembangkan sebuah sistem rekomendasi ayat Al-Qur'an atas masalah yang dibagikan masyarakat di media sosial. Di mana dalam penelitian ini menggunakan metode *Content-Based Filtering* yang digabungkan *Cosine Similarity* sebagai metode perhitungannya. Evaluasi model yang dicocokkan dengan penilaian para ahli mendapatkan hasil yang cukup baik di mana dari sepuluh percobaan yang dilakukan mendapatkan hasil yang sesuai sebesar 60% [3].

Penelitian dari Nurulhikmah dan Abdi (2024) tentang pengklasifikasi makanan berdasarkan kandungan nutrisinya menggunakan *K-Means* dan *DBSCAN*. Di mana kualitas model *clustering*-nya dievaluasi menggunakan *Silhouette Score*

dan *Davies-Bouldin Index (DBI)*. Hasil penelitian menunjukkan *K-Means* lebih unggul dan efektif dibanding *DBSCAN* dengan peforma model yang optimal atas evaluasi multiparameternya [7].

Teknik pencarian berbasis kata kunci masih memiliki keterbatasan dalam menemukan ayat yang relevan dalam Al-Qur'an, sehingga diperlukan pendekatan berbasis leksikal untuk meningkatkan relevansi hasil rekomendasi. Pada penelitian ini pendekatannya dengan mengombinasikan keunggulan masing-masing metode, baik dalam menangkap struktur global data maupun dalam mengukur kesamaan antar ayat secara presisi [8]. Maka digunakanlah *K-Means Clustering* dan *Cosine Similarity* yakni dua pendekatan penting dalam pengembangan sistem rekomendasi ini. *K-Means Clustering* digunakan untuk mengelompokkan ayat-ayat Al-Qur'an berdasarkan kemiripan konten sehingga ayat dengan tema yang sejenis dapat berada dalam satu kelompok. Representasi teks dilakukan menggunakan metode *Term Frequency–Inverse Document Frequency (TF-IDF)*, yang mampu merepresentasikan bobot kata secara numerik berdasarkan tingkat kepentingannya dalam dokumen.

Model *K-Means Clustering* merupakan pendekatan yang terbukti efektif dalam pengelompokan dokumen berbasis teks. Untuk memastikan keandalan model, maka digunakan juga skenario pengujian dengan *tuning hyperparameter* dan evaluasi model oleh *Silhouette Score* dan *Davies-Bouldin Index (DBI)*. Keduanya diterapkan guna mengukur kualitas pengelompokan dan performa model secara menyeluruh. Evaluasi hasil *Clustering* merupakan tahap terpenting dalam proses analisis data pada algoritma atau model tipe *Unsupervised Learning*. Kedua metrik tersebut memberikan informasi yang berharga tentang kualitas dan kehomogenan *Cluster* yang dihasilkan oleh algoritma *K-Means Clustering* [9].

Adapun untuk meningkatkan efisiensi dan kualitas pengelompokan data teks berdimensi tinggi, penelitian ini menerapkan algoritma *K-Means Clustering* yang dikombinasikan dengan *Principle Component Analysis (PCA)* sebagai teknik reduksi dimensi untuk mendukung proses *clustering*. *PCA* digunakan sebagai alat atau teknik reduksi dimensi untuk menyederhanakan ruang fitur hasil vektorisasi *TF-IDF* tanpa menghilangkan informasi utama yang terkandung di dalam data, sehingga proses komputasi menjadi lebih efisien. Selain itu juga *PCA* memang

cocok untuk algoritma *Unsupervised Learning* seperti *K-Means*. *PCA* pada penelitian ini tidak diposisikan sebagai algoritma optimasi, melainkan sebagai teknik reduksi dimensi untuk mendukung kinerja *K-Means* [10]. Proses *clustering* bertujuan membentuk struktur global pada ayat Al-Qur'an, sehingga pencarian rekomendasi dapat difokuskan pada kelompok ayat yang paling relevan. Selanjutnya, *Cosine Similarity* digunakan sebagai metode pengukuran tingkat kemiripan antara vektor kata kunci pengguna dan vektor ayat Al-Qur'an dalam klaster terpilih. Dengan kombinasi tersebut, sistem diharapkan mampu memberikan rekomendasi ayat yang relevan secara leksikal-tematik, serta lebih efisien dibandingkan pencarian langsung pada seluruh data [11].

Berdasarkan dari uraian latar belakang tersebut, pengembangan "Sistem rekomendasi Ayat Al-Qur'an berbasis *K-Means Clustering* dan *Cosine Similarity* dengan reduksi dimensi *PCA*" menjadi sangat penting. Penelitian ini tidak hanya berkontribusi pada bidang teknologi informasi, khususnya sistem rekomendasi berbasis teks, tetapi juga memberikan nilai tambah dalam konteks keagamaan dengan menghadirkan solusi cerdas yang memudahkan umat Islam dalam menemukan ayat-ayat Al-Qur'an yang relevan dengan kebutuhan hidup. Dengan adanya sistem ini, diharapkan umat Islam dapat lebih terdorong untuk mentadabbur Al-Qur'an yang bisa memperkuat kesehatan mental, sampai bisa menjadikan Al-Qur'an sebagai pedoman utama dalam menghadapi tantangan zaman modern. Diharapkan juga inovasi ini tidak hanya memperluas cakrawala integrasi teknologi dalam keilmuan Islam, tetapi juga membuka ruang baru bagi pengembangan dakwah digital berbasis data dan kecerdasan buatan.

1.2. Rumusan Permasalahan

Berikut beberapa rumusan pemasalahan yang akan terjawab dan diselesaikan melalui penelitian ini diantaranya :

1. Bagaimana membuat sistem rekomendasi ayat Al-Qur'an berbasis *K-Means Clustering* dan *Cosine Similarity*?
2. Bagaimana penerapan *Principle Component Analysis (PCA)* sebagai teknik reduksi dimensi serta pengaruhnya terhadap kualitas klaster *K-Means Clustering*?

3. Bagaimana mekanisme pemanfaatan sistem rekomendasi yang telah dirancang agar dengan mudah dapat digunakan oleh pengguna?
4. Bagaimana hasil sistem rekomendasi yang dikembangkan dalam memberikan ayat-ayat yang relevan terhadap kata kunci pengguna?

1.3. Tujuan Penelitian

Atas dasar dari rumusan permasalahan tersebut maka berikut beberapa tujuan utama penelitian yang akan dicapai diantaranya :

1. Mengembangkan sebuah sistem rekomendasi ayat Al-Qur'an menggunakan *K-Means Clustering* dan *Cosine Similarity*.
2. Mengetahui pengaruh dari *Principle Component Analysis (PCA)* sebagai teknik reduksi dimensi atas *K-Means Clustering* melalui hasil evaluasi kualitas klaster dengan metriks *Silhouette Score* dan *Davies-Bouldin Index*.
3. Mengimplementasikan sistem rekomendasi ayat Al-Qur'an berbasis website yang mudah digunakan.
4. Pengguna mendapatkan ayat-ayat yang sesuai dengan kata kunci atau preferensi sebagaimana yang dibutuhkan.

1.4. Manfaat Penelitian

Penelitian ini sangat diharapkan memberikan manfaat baik untuk peneliti sendiri maupun orang lain yang dijabarkan sebagaimana berikut :

1. Menambah literatur dan wawasan dalam bidang sistem rekomendasi berbasis teks, khususnya yang menggabungkan model *K-Means* dan *Cosine-Similarity* ditambah pengaruh *Principle Component Analysis (PCA)* sebagai teknik reduksi dimensi dalam konteks pengolahan teks Al-Qur'an.
2. Memberikan kemudahan bagi pengguna, khususnya umat Islam dalam menemukan dan mendabbur ayat-ayat Al-Qur'an yang relevan dengan permasalahan atau topik tertentu yang dialaminya, sehingga Al-Qur'an mudah diakses dan dijadikan pedoman hidup.
3. Membantu dalam proses pencarian ayat-ayat yang mirip secara kontekstual untuk kebutuhan, baik untuk kepentingan pembelajaran, kajian, penelitian, maupun menyusun materi dakwah.

4. Sebagai referensi dan dasar pengembangan bagi penelitian lanjutan atau sistem serupa yang lebih baik dari segi fungsionalitas, akurasi, maupun kemanfaatan secara luas.

1.5. Ruang Lingkup

Adapun pada penelitian ini agar tidak terlalu melebar dan membuat kebingungan, maka akan dibatasi ruang lingkupnya dibeberapa hal berikut :

1. Dataset yang digunakan diambil dari sumber terbuka (*Kaggle*) yang telah tersedia berisi data Al-Qur'an Kemenag.
2. Pada tahap pemodelan menggunakan *library* agar hasil lebih baik dan sesuai yang diharapkan karena banyak fungsi yang tersedia. Dalam hal *K-Means* semisal menggunakan "*sklearn*", begitu juga yang lainnya.
3. Menetapkan skenario pengujian dengan *hyperparameter tuning* (*n_cluster*, *max_iter* dan *init_method*) dan evaluasi model oleh *Silhouette Score* dan *Davies-Bouldin Index*.
4. *Principle Component Analysis (PCA)* digunakan sebagai teknik reduksi dimensi dari vektorisasi data yang dihasilkan *TF-IDF*.
5. Penelitian ini tidak hanya mengevaluasi kualitas klaster, tetapi juga menjelaskan hubungan antara proses *clustering* dan mekanisme rekomendasi.
6. *Cluster* digunakan sebagai penyempit ruang pencarian (*candidate generation*), sehingga rekomendasi ayat diambil dari cluster dengan karakteristik teks yang paling mendekati kata kunci pengguna.
7. Proses analisis pada penelitian ini hanya berdasarkan teks terjemahan (*translation*) berbahasa Indonesia.
8. Sistem akan menampilkan 5 ayat teratas dengan kemiripan tertinggi terhadap kata kunci yang dicari. Di mana hasilnya menjadi referensi yang diperuntukkan untuk tadabbur, bukan tafsir.
9. Hasil rekomendasi yang ditampilkan di website mencakup: nomor ayat, nama surah, lafadz Arab, teks terjemahan dalam Bahasa Indonesia dan tafsir beserta persentase kemiripannya (*similarity score*).
10. Website diakses dengan *localhost* yang dirunnig di dalam terminal *visual studio code*.

1.6. Kontribusi Penelitian

Terakhir penelitian ini memberikan beberapa kontribusi penting baik dari sisi metodologis, konseptual, maupun aplikatif. Di mana diharapkan dari penelitian ini bisa diambil banyak pelajaran untuk kedapannya dalam penelitian lanjutan. Kontribusi dari penelitian ini diuraikan dalam beberapa poin dan bagian berikut :

1. Kontribusi metodologis, yakni mengintegrasikan *K-Means Clustering + Cosine Similarity + PCA* dalam satu arsitektur sistem rekomendasi ayat Al-Qur'an. *PCA* tidak hanya untuk visualisasi, tapi diuji dampaknya secara kuantitatif terhadap kualitas klaster (*SS & DBI*).
2. Kontribusi konseptual, yakni memperkenalkan konsep *cluster* sebagai *candidate generation*, bukan sekadar analisis eksploratif. Artinya, *K-Means* dipakai untuk menyempitkan ruang pencarian sebelum *cosine similarity*.
3. Kontribusi empiris, yakni menyediakan evaluasi nyata atas dampak *PCA* terhadap klaster. Bersifat koseptual karena diuji dengan matriks *Precision* untuk sistem rekomendasi dengan lima ayat teratas (*top-N*) dan ini jarang ada di penelitian Al-Qur'an yang biasanya hanya demo fungsionalitas sistem atau aplikasi yang dibangun.
4. Kontribusi aplikatif, yakni membangun *website* sistem rekomendasi nyata, bukan hanya model di notebook. Bisa langsung digunakan untuk tadabbur atas ayat yang direkomendasikan.
5. Kontribusi domain keislaman, yakni menjadi salah satu penelitian yang memakai *clustering* sebagai struktur global ayat. Bukan sekadar pencarian kata kunci literal.

(Halaman ini sengaja dikosongkan)