

BAB I

PENDAHULUAN

1.1. Latar Belakang

Ledakan pertumbuhan dari berbagai jenis informasi yang tersedia di internet menimbulkan dampak berupa manusia harus memilah banyaknya informasi dan memilih mana yang sesuai dengan kebutuhan mereka. Salah satu yang relevan dengan masalah ini adalah artikel berita. Tidak jarang dengan banyaknya banyaknya opsi yang diberikan justru dapat membuat manusia kewalahan dan mengarah ke pemilihan yang kurang sesuai dengan apa yang sebenarnya diinginkan [1].

Untuk membantu menemukan informasi yang relevan diantara banyaknya informasi yang ada, dikembangkanlah Sistem Rekomendasi [2], [3]. Tujuan dari sistem rekomendasi adalah untuk memberikan konten (*item*) yang efektif dan berarti kepada pengguna yang aktif didalam *platform* [4]. Sistem Rekomendasi adalah seperangkat peralatan dan teknik yang memberikan saran kepada pengguna tentang item spesifik yang kemungkinan mereka minati. Sebuah sistem rekomendasi mencatat profil pengguna dan berdasarkan minat mereka, menyarankan sebuah produk atau layanan [5]. Saran yang diberikan bisa dalam domain apapun mulai dari saran layanan web apa yang digunakan hingga artikel berita untuk dibaca [6]–[8].

Terdapat dua tipe sistem rekomendasi, yang dipersonalisasi dan yang tidak dipersonalisasi. Sistem rekomendasi yang dipersonalisasi adalah sistem yang mana setiap pengguna mendapatkan saran yang berbeda, sedangkan sistem rekomendasi

yang tidak dipersonalisasi semua pengguna mendapatkan saran yang sama [9]. Yang perlu dilakukan oleh sistem rekomendasi yang dipersonalisasi adalah dua hal, yaitu (i) mengestimasi nilai prediksi untuk sebuah item (ii) memberikan *ranking item-item* tersebut berdasarkan nilai prediksinya [10].

Terdapat beberapa pengelompokan dalam sistem rekomendasi. Beberapa yang populer antara lain *Content-based Filtering*, *Collaborative Filtering*, dan kombinasi antara keduanya yang disebut dengan *Hybrid*. Penelitian ini menggunakan *Content-based Filtering* karena dalam proses rekomendasinya independen tidak memerlukan data pengguna lain [11], tidak memiliki masalah *cold-start* sehingga dapat langsung merekomendasikan *item* baru [12], dan karena berbasis konten dan atribut *item* [10], yang mana bersinergi dengan data artikel.

Dalam *Content-based Filtering*, *item* dideskripsikan dengan bantuan kata kunci. Algoritma yang dipakai dalam *Content-based Filtering* memprediksi *item* berdasarkan apa yang disukai pengguna diwaktu lampau. Dalam *Content-based Filtering*, sistem membandingkan profil dari pengguna dengan profil dari *item* lalu mencari *item* yang mempunyai kemiripan kemudian menyarankannya ke pengguna [13].

Artikel berita merupakan salah satu bentuk data yang tidak terstruktur. Terdapat berbagai teknik yang dapat dilakukan untuk membuat teks menjadi lebih terstruktur [14]. Salah satu teknik adalah *stemming* yang bertujuan untuk mengubah kata ke bentuk akarnya [15].

Artikel direpresentasikan kedalam *Vector Space Model (VSM)*. *Term Frequency – Inverse Document Frequency (TF-IDF)* merupakan algoritma berbasis

VSM yang efektif dan umum digunakan dalam *Content-based Filtering* [16], [17]. TF-IDF memperhitungkan perbedaan frekuensi kata dalam semua dokumen, mampu membedakan dokumen, dan algoritma ini juga digunakan secara luas dalam komputasi kesamaan (*similarity computing*) dokumen [16]. Namun konversi ini cenderung menghilangkan konteks dari kata. Sebagai contoh kata “kecerdasan buatan” kehilangan konteksnya saat diubah oleh TF-IDF menjadi “kecerdasan” dan “buatan” yang jika dipisah masing-masing bermakna berbeda. Model N-Gram dapat membantu menyelesaikan masalah ini dengan memprediksi kata yang mempunyai keterkaitan berdasarkan frekuensi kemunculannya [18].

Salah satu algoritma yang populer untuk bekerja dengan VSM untuk menghitung kemiripan *item* adalah *Cosine Similarity* [19]. Algoritma ini memiliki kelebihan yaitu tidak bergantung pada ukuran dokumen, selain itu algoritma ini juga mudah untuk diimplementasikan [20].

Untuk menjawab permasalahan diawal, penelitian ini menawarkan solusi sistem rekomendasi *Content-based Filtering* dalam bentuk *package* atau pustaka pada bahasa pemrograman Python yang dapat dipergunakan untuk membantu mempercepat pembuatan prototipe sistem rekomendasi.

Package diuji menggunakan data artikel dari portal berita Kabar Informatika Universitas Amikom Yogyakarta yang terdapat di *platform* Medium dan tersedia secara publik. Medium tidak menyediakan data riwayat interaksi, oleh sebab itu penelitian ini menggunakan data interaksi buatan yang dihasilkan secara acak menggunakan kode program.

1.2. Rumusan Masalah

Berdasarkan latar belakang penelitian tentang sistem rekomendasi *Content-Based Filtering* diatas maka dapat dirumuskan beberapa masalah berikut:

- a. Bagaimana perancangan dalam pembuatan sistem rekomendasi ini?
- b. Bagaimana penerapan sistem rekomendasi ini kedalam sebuah *package*?

1.3. Batasan Masalah

Penelitian ini menggunakan batasan sebagai berikut:

- a. Sumber data uji berasal dari portal berita Kabar Informatika Universitas Amikom Yogyakarta
- b. Objektif penelitian adalah uji penerapan sistem rekomendasi *Content-based Filtering* kedalam *package* pada bahasa pemrograman Python.
- c. Evaluasi penelitian dilakukan secara luring (*offline*)
- d. Hasil penelitian hanya sebatas *Minimum Viable Product*

1.4. Maksud dan Tujuan Penelitian

- a. Mengetahui bagaimana perancangan dalam pembuatan sistem rekomendasi *Content-based Filtering*
- b. Mengetahui penerapan *package* sistem rekomendasi *Content-based Filtering*.

1.5. Manfaat Penelitian

1.5.1. Manfaat Teoritis

Penelitian ini memberikan gambaran tentang upaya penerapan teknik sistem rekomendasi *Content-based Filtering* dalam sebuah *package* pada bahasa

pemrograman Python menggunakan data dari portal berita Kabar Informatika Universitas Amikom Yogyakarta sebagai data uji.

1.5.2. Manfaat Praktis

Penelitian dapat digunakan sebagai sumber referensi dalam perancangan sistem rekomendasi Kabar Informatika di lapangan.

1.5.3. Bagi Peneliti

Memberikan pengetahuan tentang bagaimana penerapan sistem rekomendasi dengan teknik *Content-based Filtering* dalam sebuah *package* pada bahasa pemrograman Python. Serta bisa menjadi bahan referensi dan masukan bagi peneliti lain yang sedang meneliti hal yang sama.

1.6. Metode Penelitian

Data yang digunakan dalam penelitian berasal dari portal berita Kabar Informatika yang terdapat di *platform* Medium. Data dikumpulkan menggunakan metode *consume API* yang disediakan oleh *platform*. API ini memberikan respon data *RSS Feed* dalam format XML. Data yang didapatkan dikonversi kedalam format JSON dan digunakan dalam penelitian.

1.7. Sistematika Penulisan

1.7.1. Bab I Pendahuluan

Bab I berisi mengenai latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, bagaimana metode-metode yang digunakan dalam penelitian dan yang terakhir adalah sistematika penulisan laporan.

1.7.2. Bab II Landasan Teori

Bab II berisi mengenai kajian pustaka dan dasar teori yang digunakan sebagai landasan dalam melaksanakan penelitian ini.

1.7.3. Bab III Metode Penelitian

Bab III berisi mengenai alat bahan yang dibutuhkan dan bagaimana alur penelitian dilakukan.

1.7.4. Bab IV Hasil dan Pembahasan

Bab IV berisi mengenai penjabaran hasil penelitian yang telah dilakukan.

1.7.5. Bab V Penutup

Bab V berisi mengenai kesimpulan dan saran yang dapat dilakukan bagi peneliti lain yang ingin meneliti hal yang sama kedepannya.