

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Keterbukaan akses informasi berbagai karya ilmiah yang di *publish* pada internet dengan tujuan awal untuk memberikan referensi dan sumber ilmu pengetahuan membuat pihak-pihak tertentu menyalahgunakan keterbukaan akses tersebut dalam bentuk tindakan plagiarisme (Wahyuni 2018). Berbagai bentuk tindakan plagiarisme telah banyak dilakukan oleh kalangan akademisi seperti dosen dan mahasiswa yang pada praktiknya akan mencoreng integritas dan nama baik bagi diri sendiri maupun institusi terkait (Aji 2015).

Untuk mengetahui apakah dokumen tersebut merupakan hasil plagiasi atau tidak, dapat dilakukan dengan melakukan cek kemiripan. Ada beberapa algoritma dan metode untuk mengukur tingkat kesamaan suatu naskah dengan naskah yang lain salah satunya algoritma Rabin Karp dan metode Vector Space Model (VSM). Algoritma Rabin Karp menggunakan hashing untuk menemukan sebuah substring dalam sebuah teks. Hashing adalah metode yang menggunakan fungsi hash untuk mengubah suatu jenis data menjadi beberapa bilangan bulat sederhana. Dokumen dalam **Vector Space Model (VSM)** berupa matriks yang berisi bobot seluruh kata pada tiap dokumen (Wiguna and Rizqa 2017). Bobot tersebut menyatakan kepentingan atau kontribusi kata terhadap suatu dokumen dan kumpulan dokumen. Kepentingan suatu kata dalam dokumen dapat dilihat dari frekuensi kemunculannya terhadap dokumen. Vector Space Model (VSM) memiliki 8 rumus varian pengukuran jarak diantaranya dalam (Raharjo 2014) yaitu (1)

Cosine, (2) Jaccard, (3) Dice, (4) Euclidean, (5) Manhattan, (6) Minkowski, (7) Mahalanobis, (8) sehingga dirasa perlu untuk melakukan pengukuran nilai jarak atau nilai kesamaan antar metode untuk mendapatkan hasil terbaik. Penerapan pengukuran jarak menggunakan metode Vector Space Model (VSM) juga dapat dilakukan untuk mendeteksi naskah karya ilmiah bilingual dengan menerjemahkan terlebih dahulu ke dalam bahasa Indonesia.

Suatu karya ilmiah seharusnya dibuat berdasarkan pada orisinalitas baik dari segi judul maupun isi. Lemahnya kontrol dan tidak adanya sanksi dari institusi terkait plagiasi membuat kegiatan tersebut menjadi hal yang biasa untuk dilakukan. Pada dasarnya terdapat toleransi untuk mengutip dan memparafrase kalimat pada suatu karya ilmiah yang menjadi bahan rujukan namun harus disertakan sumber identitas pemilik karya agar tidak termasuk dalam tindakan plagiat. Namun tetap saja berbagai cara untuk melakukan kecurangan tetap terjadi misalnya dengan melakukan duplikasi sebagian atau keseluruhan artikel berbahasa Inggris yang diterjemahkan ke bahasa Indonesia kemudian diakui menjadi karya milik sendiri.

Berdasarkan penjelasan di atas, penulis akan melakukan evaluasi pengukuran jarak menggunakan metode Vector Space Model (VSM) dengan beberapa varian rumus perhitungan jarak diantaranya adalah (1) Cosine, (2) Jaccard, (3) Dice, (4) Euclidean, (5) Manhattan, (6) Minkowski, (7) Mahalanobis, (8) Weighted pada algoritma Rabin Karp untuk mendeteksi plagiasi bilingual. Dengan dilakukannya penelitian ini maka diharapkan nantinya dapat mengetahui metode mana yang memiliki tingkat akurasi yang lebih baik dalam pendeteksian

dokumen bilingual sehingga meminimalisir terjadinya kecurangan dalam pembuatan karya ilmiah yang bersumber dari artikel nasional maupun internasional sehingga keorisinalitasan penelitian akan lebih terjamin. Hal tersebut dapat meningkatkan integritas dan kejujuran bagi diri peneliti dalam melakukan kajian ilmiah dan memajukan penelitian akademik di Indonesia.

### **1.2. Rumusan Masalah**

Berdasarkan latar belakang permasalahan yang telah dikemukakan, maka rumusan masalah pada penelitian ini yaitu bagaimana mengevaluasi pengukuran jarak pada algoritma Rabin Karp untuk mendeteksi plagiasi bilingual?

### **1.3. Tujuan**

Tujuan dari penelitian ini yaitu untuk mengetahui hasil evaluasi pengukuran jarak pada algoritma Rabin Karp untuk mendeteksi plagiasi bilingual.

### **1.4. Manfaat**

Adapun manfaat dari penelitian ini yaitu untuk mengetahui perbandingan tingkat akurasi dari metode Cosine, Jaccard, Dice, Euclidean, Manhattan, Minkowski, Mahalanobis, Weighted pada algoritma Rabin Karp sehingga nantinya akan ditemukan metode mana yang lebih baik dalam mendeteksi plagiasi pada dokumen bilingual.



### 1.5. Batasan Masalah

Batasan masalah pada penelitian ini yaitu:

- a. Menggunakan algoritma Rabin Karp dengan membandingkan 8 metode yaitu Cosine, Jaccard, Dice, Euclidean, Manhattan, Minkowski, Mahalanobis, Weighted
- b. Pengujian tingkat akurasi pada hasil plagiasi dilakukan dengan mendeteksi dokumen di tiap metode pada sistem yang akan dibangun
- c. Sistem pendeteksi plagiasi pada algoritma Rabin Karp akan dibangun menggunakan bahasa pemrograman Python dan database MySQL
- d. Sumber dokumen yang akan dibandingkan dan dijadikan acuan dalam pencocokan data dengan dokumen lain untuk menentukan nilai plagiasi merupakan dokumen berupa jurnal ilmiah yang terdapat dalam database sistem.



