# Aplikasi Model Ruang Vektor dan Matriks untuk Mendeteksi Adanya Plagiarisme

Johan Sentosa - 13514026
*Program Studi Informatika*
*Sekolah Teknik Elektro dan Informatika*
*Institut Teknologi Bandung, Jl. Ganesha 10 Bandung 40132, Indonesia*
*Johan_sentosa@students.itb.ac.id*
*Johansentosa17@gmail.com*

**Abstraksi— Konsep dari aljabar linear sangat berguna dan terpakai dalam kehidupan sehari-hari. Salah satu pengembangannya adalah model ruang vektor. Berdasarkan konsep dasar aljabar linear yang digunakan tersebut, bisa diaplikasikan pada aplikasi pendeteksian tindakan plaiarisme. Makalah ini berisi contoh pemodelan sebuah dokumen dengan representasi matriks dan perhitungan sudut antar subruang dari masing-masing dokumen yang akan dibandingkan dan hasilnya dapat bisa dijadikan salah satu pertimbangan untuk menentukan indikasi tindakan plagiarisme.**

**Kata Kunci—aljabar linear, ruang vektor, matriks, plagiarisme**

## I. PENDAHULUAN

Di era globalisasi seperti sekarang ini, perkembangan teknologi informasi sangatlah pesat. Orang akan sangat mudah memperoleh beragam jenis informasi. Informasi kebanyakan dapat diperoleh dari mesin pencarian *(search engine)* seperti Google, Yahoo, Bing, dan sebagainya. Pengguna cukup memasukkan kata kunci yang diinginkan dan kemudian mesin pencari akan menampilkan hasil sesuai dengan kata kunci. Hal ini tentu sangat bermanfaat dan banyak digunakan bagi banyak orang.


***Gambar 1.1 Berbagai Mesin Pencarian (Search engine)***

Source : http://www.entireweb.com/free_submission/ 21 November 2015

Namun seiring dengan perkembangannya, banyak yang menyalahgunakan kemudahan dalam pencarian informasi tersebut. Salah satunya adalah tindakan plagiarisme. Plagiarisme adalah suatu tindakan penyalahgunaan, pencurian, atau pernyataan sebagai milik sendiri sebuah ide, pikiran, tulisan atau ciptaan yang sebenarnya milik orang lain[1].

Plagiarisme ini sering terjadi di kalangan pelajar khususnya mahasiswa. Hal ini dikarenakan kegiatan tulis-menulis sering dilakukan oleh para mahasiswa untuk menyelesaikan tugas-tugas kuliah maupun tugas akhir. Tindak plagiarisme ini semakin didukung oleh tersedianya fasilitas komputer yang mampu melakukan *copy-paste* suatu dokumen. Tindak plagiarisme ini merupakan tindakan yang merugikan. Plagiarisme termasuk tindakan kriminal meniru hak cipta orang lain. Oleh karena itu diperlukan sebuah sistem pendeteksi plagiarism.

Pendeteksian ini menggunakan pemanfaatan konsep aljabar linear yang serupa dengan prinsip kerja mesin pencarian yaitu ruang vektor dan matriks. Metode yang digunakan salah satunya adalah model ruang vektor *(Vector Space Model).* Cara kerjanya adalah dengan mengimplementasikan dokumen sebagai matriks, dan kesamaan antara 2 dokumen atau 2 matriks ini dinyatakan dalam sudut antara 2 vektor. Pertama mencari frekuensi kemunculan kata pada dokumen, Kemudian menghitung kesamaannya dengan dokumen yang dibandingkan dengan metode *cosine similarity.*

## II. DASAR TEORI

2.1 Vektor

Vektor adalah objek geometri yang memiliki besaran dan memiliki arah[4]. Setiap vektor dapat dinyatakan secara geometris sebagai segmen garis berarah pada bidang atau ruang. Vektor jika digambar dilambangkan dengan tanda panah (→). Besar vektor proporsional dengan panjang panah dan arahnya bertepatan dengan arah panah. Vektor dapat melambangkan perpindahan dari titik *A* ke *B*. Vektor sering ditandai sebagai $\overrightarrow{AB}$. Sedangkan unsur vektor tersebut ditulis berurutan atu seperti matriks satu kolom atau memakai notasi vektor satuan $\hat{\imath}$ , $\hat{\jmath}$, *k.*

Panjang sebuah vektor dalam ruang euklidian tiga dimensi dapat didefinisikan sebagai berikut

$$\left\|\vec{a}\right\| = \sqrt{a_1^2 + a_2^2 + a_3^2}$$

yang meruakan konsekuensi dari Teorema Pythagoras karena vektor dasar $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ merupakan vektor-vektor satuan orthogonal[4].

Sebuah vektor yang memiliki panjang satu satuan disebut **vektor satuan** *(unit vector)*. Biasanya vektor satuan digunakan untuk mendefinisikan arah. Untuk membentuk vektor satuan, bagilah vektor tersebut dengan panjang vektor tersebut.

$$\hat{a} = \frac{\vec{a}}{\left\|\vec{a}\right\|}$$

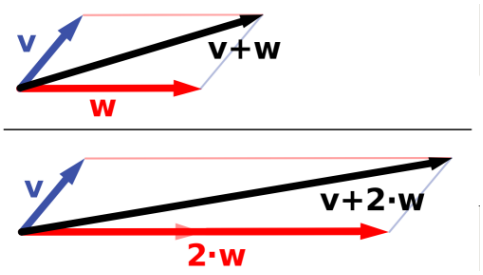
Sedangkan **vektor nol** *(null vector)* adalah suatu vektor yang panjangnya nol.

2.1.1 Operasi Vektor

Vektor pun dapat dikenakan operasi aljabar seperti penjumlahan, pengurangan, dan perkalian.
Sebagai contoh vektor $\mathbf{a} = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}$ dan $\mathbf{b} = b_1\mathbf{i} + b_2\mathbf{j} + b_3\mathbf{k}$.
Hasil dari **a** ditambah **b** adalah:

$$\mathbf{a} + \mathbf{b} = (a_1 + b_1)\mathbf{i} + (a_2 + b_2)\mathbf{j} + (a_3 + b_3)\mathbf{k}$$

Pengurangan vektor juga berlaku dengan cara mengganti tanda + menjadi tanda –

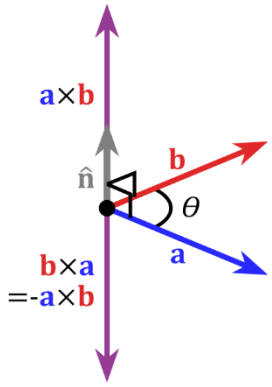



***Gambar 2.1.1 Penjumlahan vektor***

Source : https://en.wikipedia.org/wiki/Vector_space#/media/File:Vector_add_scale.svg Tanggal akses : 21 November 2015

Perkalian vektor hanya dapat dilakukan jika kedua vektor berada pada ruang yang sama. Terdiri dari :

- Hasil kali titik (*dot product*)
  Hasil kali titik akan menghasilkan besaran skalar. misal **a** dan **b** berada pada vector ruang yang sama, maka hasil kali titiknya didefinisikan
  $\bar{a} \bullet \bar{b} = \left\|\bar{a}\right\| \left\|\bar{b}\right\| \cos\alpha$
  Dimana $\left\|\bar{a}\right\|$dan $\left\|\bar{b}\right\|$masing-masing merupakan panjang vektor **a** dan vektor **b.** Dan $\alpha$ adalah sudut yang dibentuk antara dua vektor tersebut
- Hasil kali silang (*cross product)*
  Hasil kali silang merupakan perkalian antara dua vektor yang akan menghasilkan suatu vektor baru.
  $\mathbf{a} \times \mathbf{b} = \|\mathbf{a}\| \, \|\mathbf{b}\| \sin\theta$




***Gambar 2.1.1 Perkalian Silang Vektor***

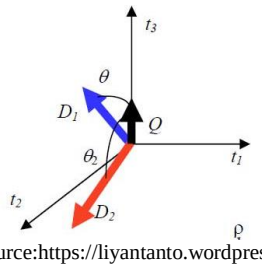
Source : https://en.wikipedia.org/wiki/Cross_product#/media/File:Cross_product_vector.svg

- Perkalian langsung
  Perkalian ini tidak bersifat komutatif

$$\vec{AB} = (a_x\hat{i} + a_y\hat{j} + a_z\hat{k})(b_x\hat{i} + b_y\hat{j} + b_z\hat{k})$$
$$= \hat{i}(a_xb_x)\hat{i} + \hat{i}(a_xb_y)\hat{j} + \hat{i}(a_xb_z)\hat{k}$$
$$+\hat{j}(a_yb_x)\hat{i} + \hat{j}(a_yb_y)\hat{j} + \hat{j}(a_yb_z)\hat{k}$$
$$+\hat{k}(a_zb_x)\hat{i} + \hat{k}(a_zb_y)\hat{j} + \hat{k}(a_zb_z)\hat{k}$$

2.2 Ruang Vektor

Ruang Vektor adalah struktur matematika yang dibentuk oleh sekumpulan vektor, yaitu objek yang dapat dijumlahkan dan dikalikan dengan suatu bilangan, yang dinamakan skalar[3]. Contoh ruang vektor adalah Vektor Euclides yang sering digunakan untuk melambangkan besaran fisika seperti gaya. Vektor- vektor yang berada di ruang $R^n$ dikenal sebagai vektor Eucides sedangkan ruang vektornya disebuat ruang n–Euclides. Model ruang vektor merupakan teknik dasar dalam perolehan informasi yang dapat digunakan untuk penelitian relevansi dokumen terhadap kata kunci pencarian (*query*) pada mesin pencarian, klasifikasi dokumen, dan pengelompokan dokumen, sistem Temu-balik informasi *(Information Retrieval System)*, dll.




***Gambar 2.2.1 Representasi vektor dalam ruang***

Source:https://liyantanto.wordpress.com/2011/06/28/pencarian-dengan-metode-vektor-space-model-vsm/

2.3 SubRuang Vektor

Subhimpunan W dari sebuah ruang vektor V dinamakan subruang V jika W itu sendiri adalah ruang vektor yang tertutup terhadap operasi penambahan dan perkalian skalar yang didefinisikan pada V[2]. Dengan demikian, syarat agar

W dikatakan sebagai sub ruang V adalah:

- $W \neq \{\}$
- $W \subseteq V$
- Jika $\boldsymbol{u}$ dan $\boldsymbol{v}$ berada pada W, maka $\boldsymbol{u} + \boldsymbol{v}$ juga berada pada W
- Jika $\boldsymbol{u}$ berada di W maka k$\boldsymbol{u}$ juga berada di W, dimana k adalah suatu skalar Riil

2.4 Plagiarisme

Menurut KBBI, Plagiarisme atau sering disebut plagiat adalah penjiplakan atau pengambilan karangan, pendapat, dan sebagainya dari orang lain dan menjadikannya seolah karangan dan pendapat sendiri. Banyak definisi dan klasifikasi yang berbeda-beda tentang plagiarisme. Beberapa contoh yang dianggap sebagai tindakan plagiarisme:

1. *Copy paste* artikel orang lain tanpa mencantumkan referensi.
2. Mengganti nama pemilik karya tulis dengan nama sendiri.
3. Mengambil ide orang lain tanpa mencantumkan sumbernya.
4. Mengubah karya orang lain tanpa seizing pemiliknya

Menurut Sudigdo Sastroasmoro, (2007) dalam tulisannya menyatakan bahwa jenis-jenis plagiarisme yang dapat ditemukan adalah[6]

1. Berdasarkan aspek yang dicuri
    - Plagiarisme ide
    - Plagiarisme isi (data peneletian)
    - Plagiarisme kata, kalimat, paragraph
    - Plagiarisme total
2. Berdasarkan proporsi konten
    - Plagiarisme ringan : < 30%
    - Plagiarisme sedang : 30 – 70 %
    - Plagiarisme berat : > 70%

2.5 *Cosine Similarity*

Cosine Similarity digunakan untuk mengukur kesamaan antara dua buah vektor. *Cosine Similarity* merupakan hasil cosinus dari sudut diantara kedua vektor. Dapat dirumuskan sebagai berikut[5]

$$sim(Q, D_1) = \cos\theta = \frac{\mathrm{Q} \cdot \mathrm{D}}{\|\mathrm{Q}\| \, \|\mathrm{D}\|} = \frac{\sum_{i=1}^{n} Q_i D_i}{\sqrt{\sum_{i=1}^{n} Q_i^2} \sqrt{\sum_{i=1}^{n} B_i^2}}$$

Keterangan:

Q = Query dokumen

$D_1$ = Dokumen uji

## III. PEMBAHASAN

Salah satu penerapan konsep aljabar linear dan vektor adalah model ruang vektor. Model ruang vektor ini awalnya diaplikasikan pada Sitem Temu-balik Informasi *(Information Retrieval System).* Dalam hal ini dokumen yang akan diperiksa dikonversi dahulu menjadi vektor berukuran ***n*** x 1, dimana n adalah banyaknya kata berbeda yang ada dalam dokumen sebagai kamus kata *(vocabulary)* atau indeks kata *(term index).* Kata-kata tersebut akan membentuk ruang vektror berdimensi ***n***. Setiap dokumen maupun query dinyatakan sebagai vektor $\mathbf{w} = (w_1, w_2, w_3, \ldots, w_n)$ di dalam $\mathbf{R}^n$. Nilai $w_i$ menyatakan jumlah kemunculan setiap kata *i* di dalam query atau dokumen *(term frequency)*.[5] Berikut adalah contoh representasi dokumen sebagai vektor

AKU CINTA BANDUNG DAN AKU ADALAH MAHASISWA INSTITUT TEKNOLOGI BANDUNG YANG CINTA TEKNOLOGI.

| Kata | Frekuensi |
|---|---|
| Aku | 2 |
| Cinta | 2 |
| Bandung | 2 |
| Dan | 1 |
| Adalah | 1 |
| Mahasiswa | 1 |
| Institut | 1 |
| Teknologi | 2 |
| Yang | 1 |

Secara Matematis :

$$Q = \begin{pmatrix} 2 \\ 2 \\ 2 \\ 1 \\ 1 \\ 1 \\ 1 \\ 2 \\ 1 \end{pmatrix}$$

Banyaknya kata berbeda ada 9, artinya, representasi kalimat judul sebagai vektor misal j dengan ukuran ( 9 x 1) merupakan subruang di $R^9$.

Lalu kita memerlukan dokumen uji disini. Kita akan menghitung sudut antara dua dokumen. Ukuran vektor haruslah sama. Pengukuran sudut antar dua buah dokumen diidentikan dengan pengukuran sudut antara dua buah vektor yaitu dengan rumus

$$\cos\theta = \frac{\mathrm{Q} \cdot \mathrm{D}}{\|\mathrm{Q}\| \, \|\mathrm{D}\|}$$

Dimana

- $\theta$ adalah sudut yang diapit oleh kedua vektor
- Q • D adalah hasil perkalian titik dari vektor **Q** dan vektor **D** yang didefinisikan sebagai
  $Q \bullet D = Q_1D_1 + Q_2D_2 + \ldots + Q_nD_n$
- $\|\mathrm{Q}\|$ dan $\|\mathrm{D}\|$ masing-masing adalah panjang

*Euclidean (norm)* dari vektor **Q** dan **D**. Panjang *Euclidean (norm)* diperoleh dari akar penjumlahan kuadrat elemen-elemen vektor tersebut

$\|D\| = \sqrt{a^2 + b^2}$

Lalu kita misalkan ada dokumen dua buah dokumen uji untuk dibandingkan kesamaannya dengan dokumen utama. Untuk saat ini dibatasi kedua dokumen uji harus merepresentasikan vektor yang ukurannya sama.

$$D1 = \begin{pmatrix} 1 \\ 2 \\ 5 \\ 3 \\ 1 \\ 2 \\ 3 \\ 2 \\ 1 \end{pmatrix} \qquad D2 = \begin{pmatrix} 0 \\ 5 \\ 3 \\ 2 \\ 0 \\ 0 \\ 1 \\ 1 \\ 5 \end{pmatrix}$$

D1 = (1,2,5,3,1,2,3,2,1) artinya dalam dokumen pertama terdapat kata "Aku" sebanyak 1 buah, 2 buah kata "Cinta", 5 buah kata "Bandung", dan seterusnya. Begitu juga dengan dokumen kedua tidak ada kata "Aku", ada 5 buah kata "cinta" dan seterusnya.

Kemudian dugaan plagiarisme dilihat dari besar sudut $\theta$ yaitu sudut antara dokumen yang ingin di cek dengan dokumen uji. Jika cos $\theta$ = 1 berarti $\theta$ = 0, vektor Q dan D berhimpit, yang berarti dokumen D sama dengan dokumen Q, atau dengan kata lain, menjiplak[5]. Jika sudut yang dibentuk semakin kecil, maka mengindikasikan telah terjadi plagiarisme. Walaupun begitu, tetap dibutuhkan pemeriksaan lebih lanjut setelah pendugaan ini.

Pada contoh diatas,

Q • D1 = (2)(1) + (2)(2) + (2)(5) + (1)(3) + (1)(1) + (1)(2) + (1)(3) + (2)(2) + (1)(1)
= 2 + 4 + 10 + 3 + 1 + 2 + 3 + 4 + 1
= 30

Q • D2 = (2)(0) + (2)(5) + (2)(3) + (1)(2) + (1)(0) + (1)(0) + (1)(1) + (2)(1) + (1)(5)
= 0 + 10 + 6 + 2 + 0 + 0 + 1 + 2 + 5
= 26

$$\|Q\| = \sqrt{2^2 + 2^2 + 2^2 + 1^2 + 1^2 + 1^2 + 1^2 + 2^2 + 1^2}$$
$$= \sqrt{21}$$

$$\|D1\| = \sqrt{1^2 + 2^2 + 5^2 + 3^2 + 1^2 + 2^2 + 3^2 + 2^2 + 1^2}$$
$$= \sqrt{58}$$

$$\|D2\| = \sqrt{0^2 + 5^2 + 3^2 + 2^2 + 0^2 + 0^2 + 1^2 + 1^2 + 5^2}$$
$$= \sqrt{65}$$

$$sim(Q, D_1) = \cos\theta_1 = \frac{Q \cdot D_1}{\|Q\| \|D_1\|} = \frac{30}{\sqrt{21}\sqrt{58}}$$

$$= \frac{30}{\sqrt{1218}} = 0.8596$$

$$\theta_1 = 30.7283°$$

$$sim(Q, D_2) = \cos\theta_2 = \frac{Q \cdot D_2}{\|Q\| \|D_2\|} = \frac{26}{\sqrt{21}\sqrt{65}}$$

$$= \frac{26}{\sqrt{1365}} = 0.703732$$

$$\theta_2 = 45.2728°$$

Dari dua hasil diatas, dokumen query memiliki beda sudut yang dibentuk dengan dokumen 1 sebesar $30.7283°$, sedangkan dengan dokumen 2 memiliki beda sudut $45.2728°$. Juka sudut yang dibentuk antar Subruang memiliki nilai yang kecil, berarti kedua dokumen memiliki kemiripan yang tinggi. Karena nilai sudut yang terbentuk antara dokumen Query dan dokumen pertama lebih kecil daripada sudut antara dokumen Query dengan dokumen kedua, maka dapat dikatakan dokumen query lebih mirip dengan dokumen pertama.

## IV. HASIL DAN KESIMPULAN

Konsep aljabar linear dapat dijadikan dasar untuk pencocokan dua buah dokumen. Pencocokan ini diaplikasikan pada system temu-balik informasi *(information retrieval system)* dan dapat juga digunakan untuk melakukan pengecekan awal untuk plagiarisme. Pengecekan ditinjau dari jumlah kata yang direpresentasikan sebagai subruang di $\mathbf{R}^n$. Vektor ini nantinya akan diukur sudutnya dengan dokumen uji. Pasangan dokumen yang sudutnya kurang dari sudutt batas, diduga memiliki kecenderungan terjadi plagiarisme. Jika $\theta$ yang dibentuk = 0 maka kedua dokumen dikatakkan sama atau identik. Jika $\theta = \pi/2$, maka dua dokumen dikatakan berbeda. Tetapi hal ini masih belum bisa dikatakan pasti. Masih banyak faktor lain yang tidak diperhitungkan selain yang diatas. Masih banyak hal yang seharusnya diperhitungkan atau bahkan harus dibuang tetapi masih diperhitungkan. Seperti contohnya kita harusnya menggunakan *stopword* yaitu kata-kata yang tidak deskriptif yang dapat dibuang seperti "di", "dan", "yang" dan lainnya, juga teknik *stemming* yaitu teknik yang dilakukan pada perolehan informasi untuk menghilangkan variasi morfologi atau dengan kata lain membuang imbuhan. Selain itu, untuk memastikan terjadinya tindak plagiarisme, kita juga harus melakukan uji empiris, berupa pemeriksaan manual. Diharapkan kedepannya, pemodelan ini akan bisa menghasilkan kinerja yang lebih baik, lebih efisien, lebih akurat dari yang sudah ada.

## REFERENCES


[1] Ridhatillah, Ardini 2003, Dealing with Plagiarism in the Information System Research Community: A Look at F Actors That Drive Plagirism and Ways to Address Them, MIS Quarterly; Vol.27, No.4, p.511-532/December 2003

[2] http://bobo.staff.mipa.uns.ac.id/files/2012/09/BAB-V.pdf.

[3] https://en.wikipedia.org/wiki/Vector_space

[4] https://id.wikipedia.org/wiki/Vektor_%28spasial%29

[5] Munir, Rinaldi. 2015. Aplikasi Aljabar Vektor pada Sistem Temu-balik Informasi. Bahan Kuliah IF2123 Aljabar Geometri. Bandung: Program Studi Teknik Informatika Sekolah Teknik Elektro dan Informatika Institut Teknologi Bandung

[6] Sudigdo, Sastroasmoro. 2007. Beberapa Catatan Tentang Plagiarisme, Majalah Kedokteran Indonesia.


## PERNYATAAN

Dengan ini saya menyatakan bahwa makalah yang saya tulis ini adalah tulisan saya sendiri, bukan saduran, atau terjemahan dari makalah orang lain, dan bukan plagiasi.

Bandung, 30 November 2015

Johan - 13514026