**ANALISIS TINGKAT AKURASI APLIKASI PENDETEKSI SIMILARITY MENGGUNAKAN ALGORITMA KOMPUTER**



**OLEH :**

**EVI PURNAMASARI, S.SI**

**09042611822002**

**PROGRAM STUDI MAGISTER TEKNIK INFORMATIKA**

**FAKULTAS ILMU KOMPUTER**

 **UNIVERSITAS SRIWIJAYA**

**2018**

**ANALISIS TINGKAT AKURASI APLIKASI PENDETEKSI SIMILARITY MENGGUNAKAN ALGORITMA KOMPUTER**

# **HALAMAN JUDUL**

**PROPOSAL TESIS**

Diajukan Untuk Melengkapi Salah Satu Syarat

Memperoleh Gelar Magister



**OLEH :**

**EVI PURNAMASARI, S.SI**

**09042611822002**

**PROGRAM STUDI MAGISTER TEKNIK INFORMATIKA**

**FAKULTAS ILMU KOMPUTER**

 **UNIVERSITAS SRIWIJAYA**

**2018**

# **LEMBAR PENGESAHAN**

**ANALISIS TINGKAT AKURASI APLIKASI PENDETEKSI SIMILARITY MENGGUNAKAN ALGORITMA KOMPUTER**

Diajukan Untuk Melengkapi Salah Satu Syarat

Memperoleh Gelar Magister

**Oleh :**

**EVI PURNAMASARI, S.SI**

**09042611822002**

Palembang, Mei 2018

****

**DAFTAR ISI**

[**HALAMAN JUDUL** i](#_Toc507588184)

[**LEMBAR PENGESAHAN** ii](#_Toc507588185)

[**DAFTAR ISI** iii](#_Toc507588187)

[**DAFTAR TABEL** iv](#_Toc507588189)

[**BAB I PENDAHULUAN** 5](#_Toc507588190)

[1.1 Latar Belakang 5](#_Toc507588191)

[1.2 Perumusan Masalah 6](#_Toc507588192)

[1.3 Batasan Masalah 6](#_Toc507588193)

[1.4 Tujuan dan Manfaat Penelitian 6](#_Toc507588194)

[1.6 Sistematika Penulisan 6](#_Toc507588196)

[**BAB II TINJAUAN PUSTAKA** 8](#_Toc507588197)

[2.1 Penelitian Terkait 8](#_Toc507588198)

[2.2 Tinjauan Pustaka 10](#_Toc507588199)

[**BAB III METODE PENELITIAN** 11](#_Toc507588201)

[3.1Jadwal Penelitian 11](#_Toc507588204)

[**DAFTAR PUSTAKA** 12](#_Toc507588201)

**DAFTAR TABEL**

**Tabel**[**3.3** *State-of-the-Art* 10](#_Toc507588202)

**Tabel**[**3.1** Rencana Jadwal Penelitian 13](#_Toc507588202)

**BAB I**

**PENDAHULUAN**

**1.1 Latar Belakang Masalah**

Pendidikan merupakan salah satu hal terpenting dalam kehidupan sesorang. Pendidikan lah yang menentukan dan menuntun masa depan dan arah hidup seseorang. Walaupun tidak semua orang berpendapat seperti itu, namun pendidikan tetaplah menjadi kebutuhan manusia nomor wahid. Bakat dan keahlian seseorang akan terbentuk dan terasah melalui pendidikan. Pendidikan juga umumnya dijadikan tolak ukur kualitas setiap orang.

Sekarang telah banyak berdirinya Perguruan Tinggi di Indonesia baik Perguruan Tinggi Negeri maupun Swasta. perguruan tinggi pun memiliki kewajiban untuk melakukan penelitian dan pengembangan. Terkait ilmu-ilmu yang diampu di perguruan tinggi tersebut. Sehingga peran perguruan tinggi tidak hanya men-transfer ilmu yang sudah tersedia saja, namun perlu mengembangkannya lagi melalui berbagai kegiatan penelitian.Kewajiban meneliti di perguruan tinggi tidak hanya ditujukan kepada mahasiswanya saja, tapi para dosennya pun memiliki kewajiban yang sama. Tapi bedanya jika mahasiswa melakukannya sebagai syarat kelulusan dengan mengimplementasikan ilmu yang didapat melalui penelitian, sedangkan kalau dosen menjadi prasyarat yang terkait dengan jenjang karir. Namun tujuan utamanya tetap untuk pengembangan ilmu yang ada dan penelitian hal-hal baru.

Beberapa tahun terakhir, dunia pendidikan dikejutkan oleh munculnya pemberitaan tentang kasus yang terkait dengan plagiarisme terhadap sejumlah karya ilmiah. Plagiarisme telah menciptakan iklim yang buruk, khususnya terhadap dunia pendidikan. Tindakan ini dapat membunuh ide dan gagasan serta menurunkan tingkat kreativitas. Tindakan ini juga mengakibatkan terjadinya degradasi moral dan mental kaum intelektual. Terbitnya Permendiknas No. 7 Tahun 2010 diharapkan mampu mencegah dan menanggulangi tindakan tersebut. Namun, tanpa metode yang tepat tindakan ini sulit dicegah.

**1.2 Rumusan Masalah**

Berdasarkan Latar Belakang diatas, maka permasalahannya dapat dirumuskan bagaimana melakukan klasifikasi data menggunakan algoritma komputer dengan text mining.

**1.3 Batasan Masalah**

Adapun batasan masalah yang akan dibahas pada penelitian ini adalah:

1. Data yang digunakan adalah data Universitas Indo Global Mandiri, dari Perpustakaan (Studi Kasus : Fakultas Teknik)
2. Metode yang digunakan dalam melakukan klasifikasi data adalah algoritma komputer dengan text mining.

**1.4 Tujuan Penelitian**

Tujuan dari penelitian ini adalah Menganalisa tingkat terjadinya similarity pada karya Ilmiah

**1.5 Manfaat Penelitian**

Berdasarkan latar belakang masalah yang sudah dijelaskan, maka manfaat yang diperoleh dari penelitian ini adalah :

1. Bagi Universitas

 Membantu Universitas agar mempermudah pengecekan data semilarity

1. Bagi Peneliti

 Manfaat yang didapat bagi peneliti adalah dapat mengimplementasikan ilmu yang sudah didapat dalam pembelajaran maupun dalam penelitian.

**1.6 Sistematika Penulisan**

 Untuk mempermudah dalam penyusunan tesisdan memuat uraian secara garis besar isi dari setiap bab, maka dibuatlah sistematika penulisan:

 1. BAB I - PENDAHULUAN

 Bab ini berisi tentang latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian dan sistematika penulisan.

 2. BAB II - TINJAUAN PUSTAKA

 Bab ini berisi tentang seluruh penjelasan mengenai tinjauan pustaka yag berhubungan dengan permsalahan yang dibahas pada penulisan tesis ini.

 3. BAB III – METODOLOGI PENELITIAN

 Bab ini berisi penjelasan secara bertahap dan terperinci tentang langkah-langkah (metodologi) yang digunakan untuk membuat kerangka berfikir dan kerangka kerja dalam meyelesaikan tesis.

 4. BAB IV – ANALISA DAN PEMBAHASAN

 Bab ini berisi tentang analisa dan pembahasan dari tiap-tiap blok diagram perencanaan rangkaian data-data hasil pengukuran.

 5. BAB V – KESIMPULAN (SEMENTARA)

 Bab ini berisi kesimpulan tentang hasil yang telah diperoleh serta merupakan jawaban dari tujuan yang ingin dicapai pada BAB I (Pendahuluan), akan tetapi masih bersifat sementara.

**BAB II**

**TINJAUAN PUSTAKA**

**2.1 Penelitian Terkait**

 Pada penelitian yang terkait yang melakukan penelitian sebelumnya sebagai acuan referensi untuk melaksanakan tesis ini. Berikut adalah tabel dari beberapa penelitian yang terkait.

**Tabel 2.1 *State-of-the-Art***

|  |  |  |  |
| --- | --- | --- | --- |
| No | Judul Penelitian | Metode | Hasil Penelitian |
| 1. | A Survey of event extraction methods from text for decision support systems | event extraction methods from text | memberikan ringkasan teknik ekstraksi peristiwa untuk data tekstual, yang membedakanantara data-driven, pengetahuan-driven, dan metode hibrida, dan menyajikan evaluasi kualitatif ini. |
| 2. | Research on internet information mining based on agent algorithm | agent algorithm | Algoritma yang diperbaikidapat mengubah berat kelas awal, menghitung bobot dari kata-kata fitur,berat kata karakteristik dihitung, dan sangat meningkatkan eksekusiefisiensi algoritma, sehingga mencegah hasil pengelompokan dari jatuh ke dalamsolusi optimal lokal dan mendapatkan global optimal, dan mendapatkan pengelompokan yang lebih baikefek. Sementara itu, dengan pendekatan eksperimen kita dapat langsung menemukan gugustersangka, itu akan memberikan dasar penting untuk pengambilan keputusan keamanan publikpemimpin dan pengejaran para tersangka sampai akhirnya menemukan para penjahat. |
| 3. | PDLK : Plagiarism detection using linguistic knowledge | linguistic knowledge | Metode ini dapat meningkatkan kinerja deteksi plagiarisme karena mampu menghindari pemilihan sumberkalimat teks yang kemiripannya dengan kalimat teks yang mencurigakan tinggi tetapi maknanya berbeda. |
| 4. | Crime profiling for the Arabic Language using computational linguistic techniques | computational linguistic techniques | Hasil keseluruhan yang diperoleh adalah: untuk presisi 71%, ingat 94% dan F-measure 81%. Juga, untuk mengevaluasi efektivitas penggunaan pendekatan ekstraksi informasi CPS pada teknik pengelompokan Peta Pengorganisasian Diri (SOM), tiga parameter (ukuran data, waktu pembukaan dan kesalahan kuantisasi) digunakan.Hasil presisi dan recall menunjukkan bahwa tata bahasa lokal lokasi dapat mengekstrak 77% entitas lokasi, dengan tingkat presisi 98%. Di sisi lain, menggunakan kamus lokasi telah menyebabkan peningkatan hasil recall menjadi 94%, dengan tingkat presisi 96%. |
| 5. | Applying LSA text mining technique in envisioning social impacts of emerging technologies: The case of drone technology | LSA text mining technique | kegiatan foresight kuantitatif dapat disajikan sebagai cara alternatifmemahami penggambaran masa depan teknologi yang muncul. Kedua,hasilnya bisa menjadi alat penting bagi mereka yang ingin munculteknologi tersebar luas dalam waktu dekat, seperti potensikonsumen dan perusahaan teknologi. |
| 6 | How automated feedback through text mining changes plagiaristic behavior in online assignments | the McNemar‘s Chi-square test | Hasil analisis mengungkapkan perbedaan yang signifikan secara statistik dalam rasio posting plagiarized dan rasio siswa melakukan perilaku plagiaristik sebelum dan sesudah umpan balik. Ditemukan bahwa setelah umpan balik diberikan, sementara rasio posting menjiplak turun 21,07% rata-rata, 83% dari siswa memiliki lebih rendah plagiarized post ratio. Jumlah siswa yang tidak melakukan plagiarisme meningkat 42,37% setelah umpan balik diberikan. |

**2.2 Tinjauan Pustaka**

 **Text Mining** adalah proses ekstraksi pola (informasi dan pengetahuan yang berguna) dari sejumlah besar sumber data tak terstruktur. Penambangan teks memiliki tujuan dan menggunakan proses yang sama dengan penambangan data, namun memiliki masukan yang berbeda.

Tujuan dari text mining adalah untuk mendapatkan informasi yang berguna dari sekumpulan dokumen. Jadi, sumber data yang digunakan pada text mining adalah kumpulan teks yang memiliki format yang tidak terstruktur atau minimal semi terstruktur. Adapun tugas khusus dari text mining antara lain yaitu pengkategorisasian teks (text categorization) dan pengelompokan teks text clustering).

**BAB III**

**METODOLOGI PENELITIAN**

Pada Metodologi penelitian ini adalah tahapan dari sistematis dan apa yang akan dilakukan oleh penelitian. Adapun langkah yang dilakukan dengan pengumpulan data, pengolahan data, pengujian sistem menggunakan Algoritma Evaluasi dan Validasi Hasil dan Pengujian Algoritma.

**3.1 Jadwal Penelitian**

Penelitian dilakukan mulai semester 1 tahun ajaran 2018/2019 dengan penjadwalan sebagai berikut :

Tabel 3.1 Jadwal Penelitian

|  |  |  |
| --- | --- | --- |
| No | Deskripsi Kegiatan | Tahun 2018 - 2019 |
| Mei | Juni | Juli | Agt | Sep | Okt | Nov | Des | Jan |
| 1 | Kajian Pustaka |  |  |  |  |  |  |  |  |  |
| 2 | Pengumpulan Data |  |  |  |  |  |  |  |  |  |
| 3 | Perancangan Perangkat Lunak |  |  |  |  |  |  |  |  |  |
| 4 | Implementasi  |  |  |  |  |  |  |  |  |  |
| 5 | Testing dan Hasil |  |  |  |  |  |  |  |  |  |
| 6 | Dokumentasi & Laporan Akhir |  |  |  |  |  |  |  |  |  |

**DAFTAR PUSTAKA**

Abdi, A., Idris, N., Alguliyev, R. M., & Aliguliyev, R. M. (2015). PDLK: Plagiarism detection using linguistic knowledge. *Expert Systems with Applications*, *42*(22), 8936–8946.

Akçapinar, G. (2015). How automated feedback through text mining changes plagiaristic behavior in online assignments. *Computers and Education*, *87*, 123–130.

AlSallal, M., Iqbal, R., Palade, V., Amin, S., & Chang, V. (2018). An integrated approach for intrinsic plagiarism detection. *Future Generation Computer Systems*.

Anjali, V., Swapna, T. R., & Jayaraman, B. (2015). Plagiarism detection for java programs without source codes. *Procedia Computer Science*, *46*(Icict 2014), 749–758.

Hogenboom, F., Frasincar, F., Kaymak, U., De Jong, F., & Caron, E. (2016). A Survey of event extraction methods from text for decision support systems. *Decision Support Systems*, *85*, 12–22.

Nokhbeh Zaeem, R., Manoharan, M., Yang, Y., & Barber, K. S. (2017). Modeling and analysis of identity threat behaviors through text mining of identity theft stories. *Computers and Security*, *65*, 50–63.

Oberreuter, G., & Velásquez, J. D. (2013). Text mining applied to plagiarism detection: The use of words for detecting deviations in the writing style. *Expert Systems with Applications*, *40*(9), 3756–3763.

schneider, johannes, Shepherd, D., Damevski, K., vom Brocke, J., & Bernstein, A. (2017). Detecting Plagiarism based on the Creation Process. *IEEE Transactions on Learning Technologies*, *1382*(c), 1–14.

Suarez-Tangil, G., Tapiador, J. E., Peris-Lopez, P., & Blasco, J. (2014). Dendroid: A text mining approach to analyzing and classifying code structures in Android malware families. *Expert Systems with Applications*, *41*(4 PART 1), 1104–1117.