

BAB 1. PENDAHULUAN

1.1 Latar Belakang

Politeknik Negeri Jember merupakan salah satu perguruan tinggi di Jember yang menyelenggarakan pendidikan vokasi, yaitu program pendidikan yang mengarah proses belajar mengajar pada tingkat keahlian, keterampilan, dan standar kompetensi yang spesifik. Oleh karena itu, lulusan Politeknik Negeri Jember memiliki keterampilan yang siap kerja pada bidang keahlian masing-masing, seperti pada bidang Teknologi Informasi dimana mahasiswa dituntut untuk mempelajari tentang pengembangan teknologi yang diharapkan mampu bersaing baik di tingkat nasional maupun internasional.

Teknologi Informasi merupakan teknologi yang diciptakan untuk membantu manusia dalam hal membuat, mengubah, menyimpan, mengomunikasikan, dan menyebarkan informasi. Selain itu fungsi dari teknologi informasi untuk memecahkan suatu masalah, membuka kreativitas, meningkatkan efektivitas dan efisiensi dalam aktivitas manusia. Politeknik Negeri Jember memiliki jurusan Teknologi Informasi yang fokus pada pengetahuan tentang teknologi. Salah satu program studi yang ada di jurusan Teknologi Informasi ini yaitu Teknik Informatika.

Teknik Informatika merupakan program studi sarjana terapan di jurusan Teknologi Informasi yang mempelajari beberapa bidang keahlian, seperti analisis sistem, pengembangan sistem informasi, melakukan pengelolaan sumberdaya informasi, dan membuat perangkat lunak baik berbasis *desktop*, *website*, maupun *mobile*. Di program studi ini terdapat mata kuliah yang memberikan tugas untuk melakukan *resume* materi yang telah disampaikan oleh dosen, dalam proses belajar-mengajar yang dinamakan *learning journal* (jurnal belajar). Dengan membuat *learning journal* diharapkan dapat mengingat kembali materi yang telah disampaikan, dan mampu menggambarkan pemahaman mahasiswa secara tertulis.

Learning Journal (jurnal belajar) adalah dokumen yang secara terus-menerus bertambah dan berkembang. Ditulis oleh seorang mahasiswa (pelajar) sebagai bentuk pemahaman terhadap materi yang diberikan oleh dosen (pengajar). Di masa

pandemic Covid'19 ini, mahasiswa belajar secara *daring* (tidak terjadi interaksi secara langsung dengan dosen), oleh karena itu, dosen tidak dapat melihat proses belajar mahasiswa selama masa belajarnya, sehingga dosen memberikan tugas *learning journal* kepada mahasiswa di setiap pertemuan kuliah, baik tugas mata kuliah maupun praktikum. Berdasarkan banyaknya mahasiswa yang mengumpulkan *learning journal*, tidak dapat dipungkiri jika terdapat *plagiarism* pada tugas tersebut.

Plagiarism adalah tindakan menjiplak karya orang lain dan mengaku bahwa karya tersebut adalah karyanya sendiri. Berdasarkan data pada Kemenristek Dikti, sejak tahun 2008 sampai dengan tahun 2017, kasus *plagiarism* mengalami peningkatan dari tahun ke tahun (Sukaesih, 2018). Pemerintah telah memberikan perhatian khusus terkait permasalahan *plagiarism*, seperti dalam (Peraturan Menteri Pendidikan Nasional Republik Indonesia nomor 17 tentang pencegahan dan penanggulangan kasus plagiat di perguruan tinggi 2010) dan (Undang-Undang Republik Indonesia nomor 19 tentang hak cipta 2002). Namun pada kenyataannya, kasus plagiat masih banyak dilakukan di Indonesia, terutama yang terjadi di golongan mahasiswa, dimana dalam mengerjakan tugas yang diberikan oleh bapak/ibu dosen terkait *learning journal* pasti tidak jauh dari kasus plagiat. Salah satu upaya yang dapat dilakukan untuk meminimalisir adanya tindakan *plagiarism* ini adalah dengan melakukan perbandingan *learning journal* antar mahasiswa dengan menghitung tingkat *similaritas*.

Menurut penelitian terdahulu terkait dengan *similaritas* berdasarkan data yang dilakukan oleh (Pratama et al., 2019) dengan judul “Deteksi *Plagiarism* pada Artikel Jurnal menggunakan Metode *Cosine Similarity*” melakukan deteksi *plagiarism* pada artikel jurnal dengan melakukan pengambilan data yang di *upload* di *repository* yang diperoleh dari hasil *grabbing* data DOAJ (*Directory of Open Access Journal*). Hasil penelitian berdasarkan skenario uji coba dalam perhitungan jumlah artikel relevan mendapatkan nilai *recall* sebesar 13%. Sedangkan untuk nilai *precision* mendapatkan hasil 8%.

Penelitian lain yang telah dilakukan dengan menggunakan *cosine similarity* dalam melakukan deteksi kemiripan pada *content* tugas akhir mahasiswa dengan

judul “*Implementation of Clustering and Similarity Analysis for Detecting Content Similarity in Student Final Projects*” yang dilakukan oleh (Ilham et al., 2020). Penelitian ini mendapatkan hasil bahwa sistem dapat mengelompokkan dokumen dan mendeteksi kemiripan *content* dengan algoritma TF-IDF dan *Cosine Similarity*.

Berdasarkan beberapa penelitian yang telah dijelaskan di atas, analisis *similaritas* menggunakan algoritma TF-IDF (*Term Frequency-Inverse Document Frequency*) sebagai pembobotan kata dan *Cosine Similarity* sebagai penentuan tingkat *similaritas* yang terjadi antar dua dokumen atau lebih memiliki persentase ketepatan yang lebih baik. Berdasarkan tingkat ketepatan algoritma tersebut dapat disimpulkan bahwa penelitian ini menggunakan algoritma TF-IDF (*Term Frequency-Inverse Document Frequency*) dan *Cosine Similarity* untuk mendeteksi tingkat persentase *plagiarism* atau *similaritas* yang terjadi antar *learning journal* mahasiswa.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, dapat dirumuskan beberapa rumusan masalah diantaranya yaitu:

1. Bagaimana merancang sistem deteksi *plagiarism* dengan mengukur tingkat persentase *similaritas* pada *learning journal* mahasiswa D4-Teknik Informatika?
2. Bagaimana hasil implementasi dari algoritma TF-IDF (*Term Frequency-Inverse Document Frequency*) dan *Cosine Similarity* dalam mendeteksi *plagiarism*?
3. Bagaimana menguji hasil akurasi tingkat persentase *similaritas* pada sistem deteksi *plagiarism* menggunakan algoritma TF-IDF (*Term Frequency-Inverse Document Frequency*) dan *Cosine Similarity*?

1.3 Tujuan

Dari rumusan masalah di atas maka, penelitian ini memiliki beberapa tujuan diantaranya yaitu:

1. Dapat menghasilkan sistem deteksi *plagiarism* dengan mengukur tingkat persentase *similaritas* pada *learning journal* mahasiswa D4-Teknik Informatika.
2. Dapat mengimplementasikan algoritma TF-IDF (*Term Frequency-Inverse Document Frequency*) dan *Cosine Similarity* untuk sistem deteksi *plagiarism*.
3. Dapat mengetahui hasil akurasi tingkat persentase *similaritas* pada sistem deteksi *plagiarism* menggunakan algoritma TF-IDF (*Term Frequency-Inverse Document Frequency*) dan *Cosine Similarity*.

1.4 Manfaat

Manfaat yang diharapkan dari penelitian dengan judul Analisis *Similaritas Learning Journal* pada D4-Teknik Informatika adalah sebagai berikut:

1. Memaparkan hasil penelitian untuk dijadikan media pendeteksi *plagiarism*.
2. Mengetahui tingkat persentase *similaritas* pada *learning journal* antar mahasiswa di D4-Teknik Informatika.
3. Mengurangi kasus *plagiarism* pada mahasiswa di D4-Teknik Informatika.

1.5 Batasan Masalah

Dalam penelitian ini terdapat batasan-batasan masalah yang dapat dijadikan fokus utama pada penelitian, sehingga mampu mendapatkan hasil yang lebih baik. Berikut adalah batasan-batasan masalah pada penelitian ini, antara lain:

1. Penelitian ini berfokus pada tugas *learning journal* (*resume* materi) mahasiswa D4-Teknik Informatika angkatan tahun 2018 semester 6 mata kuliah Organisasi dan Arsitektur Komputer yang dikumpulkan di link *google form* yang disediakan oleh bapak/ibu dosen.
2. Dilakukan perbandingan *learning journal* antar mahasiswa berdasarkan pengelompokan data yang sesuai dan menampilkan persentase tingkat *similaritas* $> 20\%$.
3. Dilakukan deteksi *plagiarism* pada *learning journal* pada teks Bahasa Indonesia.