

BAB 1. PENDAHULUAN

1.1 Latar Belakang

Tri Dharma Perguruan Tinggi merupakan landasan utama dalam penyelenggaraan pendidikan tinggi di Indonesia yang mencakup pendidikan dan pengajaran, penelitian, serta pengabdian kepada masyarakat (Kementerian Pendidikan dan Kebudayaan RI, 2012). Sesuai dengan Undang-Undang Nomor 12 Tahun 2012 tentang Pendidikan Tinggi Pasal 1 Ayat 9, perguruan tinggi wajib melaksanakan ketiga aspek tersebut guna mendorong perkembangan ilmu pengetahuan dan kesejahteraan masyarakat. Dosen, sebagai bagian dari perguruan tinggi, memiliki kewajiban tidak hanya dalam mendidik mahasiswa tetapi juga dalam menghasilkan luaran akademik dari penelitian dan pengabdian, seperti publikasi ilmiah dan Hak Kekayaan Intelektual (HKI), yang menjadi indikator produktivitas akademik serta berkontribusi terhadap pengembangan ilmu dan penerapannya di masyarakat (Suryaningsum *et al.*, 2020). Kewajiban ini diperkuat dalam Pasal 12 Ayat (2) dan (3) UU No. 12 Tahun 2012, yang menyatakan bahwa dosen memiliki tugas mengembangkan ilmu pengetahuan dan teknologi melalui penelitian serta menyebarkanluaskannya dalam bentuk buku ajar atau publikasi ilmiah sebagai bagian dari pengembangan budaya akademik.

Publikasi ilmiah berperan penting dalam penyebaran hasil penelitian, validasi metode ilmiah melalui tinjauan sejawat (*peer review*), serta perluasan jaringan akademik (Ridwan Institute, 2024). Di tingkat global, publikasi ilmiah menjadi faktor strategis dalam peningkatan reputasi perguruan tinggi, sementara di Indonesia, publikasi ilmiah juga berperan dalam evaluasi kinerja penelitian dan kinerja dosen, pemeringkatan institusi, serta akreditasi perguruan tinggi. Selain itu, publikasi ilmiah berkontribusi terhadap syarat kelulusan mahasiswa, kenaikan pangkat akademik dosen, serta meningkatkan daya saing perguruan tinggi (Ridwan Institute, 2024). Seiring berkembangnya teknologi, publikasi ilmiah kini semakin terdigitalisasi, memungkinkan akses lebih luas serta kolaborasi akademik yang lebih efektif (Suryaningsum *et al.*, 2020). Sebagai bentuk evaluasi kinerja

penelitian, publikasi ilmiah dosen perlu terdokumentasi dalam sistem yang dapat mengukur dampak dan kualitasnya secara terstruktur.

Hasil luaran penelitian dan pengabdian dicatat dan direkap dalam berbagai platform atau sistem indeksasi ilmiah guna memantau serta menilai produktivitas akademik dosen. Salah satu platform utama adalah *Google Scholar*, layanan mesin pencari yang dikembangkan oleh *Google* untuk mencari literatur akademik dalam berbagai format publikasi, seperti jurnal ilmiah, buku, prosiding konferensi, dan laporan penelitian (Publikasi Indonesia, 2024). Diluncurkan pada tahun 2004, layanan ini dirancang untuk mempermudah akses terhadap referensi ilmiah dengan menampilkan hasil pencarian berdasarkan kelengkapan teks, penulis, jumlah sitasi, dan sumber publikasi. Selain *Google Scholar*, pemerintah Indonesia mengembangkan *Science and Technology Index (SINTA)* sebagai sistem indeksasi publikasi ilmiah yang dikelola oleh Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi (Kemendikbudristek) (Publikasi Indonesia, 2024). *SINTA* berfungsi sebagai katalog jurnal nasional terakreditasi yang digunakan untuk mengukur dan menilai kinerja penelitian akademisi di Indonesia, dengan jurnal yang dikategorikan ke dalam enam tingkatan (*SINTA 1* hingga *SINTA 6*) berdasarkan kualitas editorial, keterbukaan akses, dan dampak penelitian.

Meski kaya informasi, *Google Scholar* dan *SINTA* masih mengharuskan banyak institusi, termasuk Politeknik Negeri Jember, untuk mengumpulkan data secara manual. Proses ini memakan waktu, tidak efisien, dan tidak didukung oleh fitur ekspor data dalam jumlah besar, sehingga pencatatan masih harus dilakukan secara individu. Setiap tahun atau pada saat akreditasi, institusi pendidikan tinggi wajib melakukan evaluasi publikasi dosen sebagai bagian dari pemantauan kinerja akademik dan peningkatan kualitas institusi. Namun, keterbatasan dalam aksesibilitas dan pengelolaan data publikasi sering kali menjadi kendala dalam proses evaluasi ini. Dalam mengatasi tantangan tersebut, salah satu teknik yang dapat diterapkan adalah *web scraping*.

Web scraping memungkinkan pengambilan data secara otomatis dari halaman web, menghasilkan informasi yang lebih cepat, akurat, dan dalam jumlah besar tanpa perlu dilakukan secara manual. Metode ini telah diterapkan dalam

penelitian akademik untuk meningkatkan efisiensi pengelolaan data. Kopravi dan Putra (2023) berhasil menerapkan *web scraping* pada *Google Scholar* untuk mengumpulkan data publikasi dosen dan menampilkan informasi sitasi secara *real-time*. Selain itu, penelitian yang dilakukan oleh Adila *et al.* (2022) menggunakan *web scraping* untuk mengambil data jurnal dari *SINTA*, menyimpannya dalam basis data lokal, serta mengatur pembaruan otomatis menggunakan *cron job*, sehingga data selalu terbaru tanpa campur tangan manual. Studi lain yang dilakukan oleh Dwicahyo & Ratnasari (2022) membandingkan berbagai metode *web scraping* seperti *Xpath Selector*, *CSS Selector*, *JSON Parsing*, dan *HTML Parsing*, menemukan bahwa masing-masing metode memiliki efektivitas yang berbeda tergantung pada struktur situs *web* yang digunakan.

Berdasarkan latar belakang tersebut, penelitian maka dibutuhkan sistem *web scraping* guna mengotomatisasi pengumpulan data luaran dari *Google Scholar* dan *SINTA*. Sistem ini dikembangkan dengan menggunakan metode *CRIPS-DM* dan harapannya dapat menghasilkan data yang mendukung proses pengambilan keputusan dalam evaluasi kinerja dosen terutama hasil luaran dosen seperti penelitian dan pengabdian masyarakat.

1.2 Rumusan Masalah

Berdasarkan latar belakang tersebut, rumusan masalah yang ditetapkan adalah sebagai berikut:

1. Bagaimana mengembangkan sistem *web scraping* yang dapat mengumpulkan data luaran (publikasi jurnal penelitian, pengabdian, dan HKI) dari platform *Google Scholar* dan *SINTA*?
2. Bagaimana mengembangkan dashboard membantu pihak pengambil keputusan (koordinator program studi) dalam mengevaluasi kinerja dosen khususnya untuk penelitian dan pengabdian.

1.3 Tujuan

Berdasarkan rumusan masalah yang ditetapkan, tujuan dari penelitian ini adalah sebagai berikut:

1. Mengembangkan sistem *web scraping* yang dapat secara otomatis mengumpulkan data (publikasi jurnal penelitian, pengabdian, dan HKI) dosen dari *Google Scholar* dan *SINTA*.
2. Menghasilkan sistem yang dapat menampilkan data publikasi dalam bentuk *dashboard* dan untuk memudahkan analisis lanjut pengambilan keputusan (koordinator program studi) di jurusan Teknologi Informasi Politeknik Negeri Jember.

1.4 Manfaat

Adapun manfaat yang diharapkan dalam penelitian ini adalah sebagai berikut:

1. Bagi Institusi Pendidikan:
 - a. Mempermudah proses pengumpulan data publikasi akademik dosen.
 - b. Mendukung evaluasi kinerja dosen berdasarkan data publikasi yang lebih terstruktur.
2. Bagi Pengembang Teknologi:
 - a. Memberikan wawasan mengenai penerapan *web scraping* dalam bidang akademik.
3. Bagi Peneliti dan Akademisi:
 - a. Mempermudah akademisi dalam mengumpulkan dan menganalisis jumlah publikasi serta sitasi dosen.
 - b. Mempercepat akses ke data publikasi dosen yang tersedia pada *Google Scholar* dan *SINTA* tanpa harus melakukan pencarian manual.

1.5 Batasan Masalah

Untuk menjaga fokus penelitian dan memastikan kelayakan, penelitian ini memiliki beberapa batasan masalah sebagai berikut:

1. Sistem *web scraping* akan dibuat dalam bentuk website.
2. Sistem *web scraping* yang dikembangkan hanya akan digunakan untuk mengumpulkan data luaran (publikasi jurnal penelitian, pengabdian, dan HKI) dari platform *SINTA* dan *Google Scholar*, yang kemudian diekstrak dalam bentuk data excel (.xlsx).

3. Pengguna dari sistem ini adalah koordinator program studi dari program studi Manajemen Informatika, Teknik Informatika dan Teknik Komputer dari jurusan Teknologi Informasi.
4. Elemen data yang dikumpulkan dalam proses *web scraping* terbatas pada informasi seperti judul penelitian, nama penulis, afiliasi institusi, tahun publikasi, jurnal tempat publikasi, jumlah sitasi, indeks akreditasi SINTA, informasi pendanaan, nomor pemohon untuk data HKI, link publikasi ke *Scopus* atau *DOI* jika tersedia, serta informasi *quartile* jurnal seperti Q1, Q2 dan seterusnya.
5. Penelitian yang dikumpulkan hanya mencakup publikasi dari tahun 2019 hingga 2024.
6. Pengembangan *dashboard* hasil *web scraping* menggunakan *framework FastAPI*.