

BAB I

PENDAHULUAN

1.1 Latar Belakang

Saat ini, volume data yang sangat besar dihasilkan setiap hari dengan kecepatan yang belum pernah terjadi sebelumnya dari berbagai sumber (misalnya, kesehatan, pemerintah, jejaring sosial, pemasaran, keuangan). Ini karena banyaknya tren teknologi, termasuk Internet of Things (Oussous, 2018). Keberadaan internet membuat jumlah data yang terdapat di website – website semakin bertambah. Ini dapat menjadi peluang bagi data engineer untuk mengolah banyaknya data tersebut menjadi statistik yang bermanfaat bagi perusahaan atau peneliti. Namun dibutuhkan metode yang efisien untuk mengumpulkan banyaknya data yang ada di suatu website, dan web scraping termasuk salah satunya (Satriaajati et al., 2021).

Penelitian yang menerapkan teknik web scraping pernah dilakukan oleh Salim Satriaajati, Satria Bagus Panuntun, dan Setia Pramana pada tahun 2020 dengan melakukan web scraping pada situs berita yakni detik.com untuk mengumpulkan berita kriminal pada masa pandemi covid-19.

Web scraping sendiri adalah penggunaan alat teknologi untuk ekstraksi otomatis dan pengolahan data dari suatu web untuk tujuan analisis lebih lanjut dari data tersebut (Krotov, 2018; Satriaajati et al., 2021). Salah satu alat scraping yang terkenal adalah *beautifulsoup*, yaitu sebuah *open source* milik *python* yang berfungsi untuk mengurai dokumen HTML dan XML dari website dengan cepat dan sederhana (Hajba, 2018).

Sebagai target pengujian untuk implementasi web scraping, peneliti memilih website *myanimelist.net* yang merupakan komunitas dan database anime dan manga online paling aktif di dunia. Menurut *similiarweb.com* pada 1 Desember 2021, *myanimelist.net* menduduki peringkat 582 dunia dalam website dengan traffic tertinggi dan peringkat 10 dalam kategori *Arts & Entertainment*. *Myanimelist* memiliki banyak database seputar anime, mulai dari judul, tanggal rilis, status, pengisi suara hingga rating. Banyaknya data dalam website ini dapat kita ambil dan proses untuk menarik suatu kesimpulan.

Anime sendiri adalah animasi khas Jepang yang memiliki ciri pada gambar yang berwarna-warni yang menampilkan berbagai macam tokoh serta lokasi dari cerita yang ditunjukkan, anime dipengaruhi oleh gaya gambar manga komik khas Jepang (Purnamasari, 2018). Beriring perkembangan waktu, anime menjadi lebih terkenal karena perkembangan grafik yang lebih baik dan cerita yang beragam dari berbagai macam genre, contoh anime yang sedang naik daun sekarang adalah *Shingeki no Kyojin / Attack on Titan* dengan grafik yang sangat detail dan bergenre *action fantasy*, akan tetapi perkembangan anime tidak luput dari jasa seorang *seiyuu / Seiyū* (声優), bahasa jepang dari pengisi suara.

Seiyuu memiliki peran besar dalam industri anime karena beberapa dari mereka dapat menjiwai karakter yang diperankannya dengan baik, hal ini tentu dapat membuat penonton lebih nyaman dalam menikmati anime. *Seiyuu* berkontribusi ke pasar anime melalui mempromosikan karakter yang mereka perankan dan serial animenya. Penelitian tentang hubungan antara *seiyuu* dengan anime pernah dilakukan oleh Ruby Zelka Heinst pada tahun 2017 lewat tesisnya yang berjudul “*Seiyū: the art of voice acting the Japanese voice acting industry and its relation to the anime market*”. Ruby Zelka Heinst menyimpulkan bahwa beberapa *seiyuu* memiliki fanbase nya sendiri yang membuat anime yang diperankannya mendapat banyak penonton, karena banyak orang mendasarkan pilihan anime mereka pada *seiyuu* dan apakah *seiyuu* favorit mereka muncul di dalamnya atau tidak.

Pada penelitian kali ini, penulis akan mencoba mengimplementasikan metode web-scraping pada myanimelist untuk mengekstrak data seputar *seiyuu* dan anime yang diperankannya, lalu membuat uji korelasi untuk mencari tahu ada atau tidaknya hubungan antara data tersebut.

1.2 Rumusan Masalah

Dari latar belakang penelitian, peneliti ingin mencari tahu cara menggunakan metode *web-scraping* dan mengimplementasikannya pada situs myanimelist. Data-data yang diambil dari situs tersebut kemudian akan dianalisis menggunakan analisis korelasi. Oleh karena itu, peneliti merumuskan masalah pada penelitian ini berupa:

1. Apakah *scraping* dapat diimplementasikan dengan efektif di myanimelist.net?
2. Apakah terdapat hubungan korelasi antara popularitas *seiyuu* dengan anime?
3. Bagaimana kekuatan hubungan korelasi antara popularitas *seiyuu* dengan anime?

1.3 Tujuan Penelitian

Peneliti dalam penelitian ini bertujuan untuk mengembangkan metode *web-scraping* dengan mengimplementasikannya pada *website* myanimelist untuk mengambil data seputar *seiyuu* dan anime. Data kemudian dianalisis untuk mencari hubungan dari kedua data tersebut. Hasil analisis tersebut dapat menjadi acuan bagi peneliti untuk menarik kesimpulan dan menambah wawasan tentang *seiyuu* dan anime.

1.4 Manfaat Penelitian

Manfaat penelitian dari sudut pandang peneliti adalah peneliti dapat mempelajari *web-scraping* dan langsung mengimplementasikannya. Metode ini dapat berguna ketika peneliti ingin mengumpulkan banyak informasi dari suatu web secara otomatis. Sedangkan manfaat teoritis dari penelitian ini adalah memperkuat dan membuktikan bahwa *web-scraping* adalah metode pengumpulan data yang efektif.

1.5 Batasan Masalah

Penelitian ini memiliki beberapa batasan untuk menyederhanakan ruang lingkup penelitian.. Peneliti beranggapan bahwa objek penelitian berupa hubungan popularitas *seiyuu* dan anime yang dapat dibuktikan dengan data yang banyak dan metode uji statistik yang variatif. Oleh karena itu, peneliti membuat batasan masalah berupa:

1. Penelitian ini hanya berdasarkan data yang ada dalam myanimelist.net.

2. Penelitian ini mencakup 450 orang dalam 9 halaman yang ada dalam daftar orang populer di myanimelist.net.
3. Penelitian hanya menggunakan uji korelasi untuk mencari tahu hubungan antara *seiyuu* dan anime.