

Analisis Genre Film Berdasarkan Data Subtitle

Nathania Novenrodumetasa^{a1}, I Made Agus Dwi Suarjaya^{a2}, I Made Sunia Raharja^{b3}

^aProgram Studi Teknologi Informasi, Fakultas Teknik, Universitas Udayana, Bali, Indonesia
e-mail: ¹nathania@student.unud.ac.id, ²agussuarjaya@it.unud.ac.id,
³sunia.raharja@unud.ac.id

Abstrak

Film merupakan sekumpulan gambar yang mengalami banyak pemrosesan editing untuk mencapai hasil yang memuaskan. Film memiliki genre yang berfungsi untuk membedakan jenis film dan banyak film yang ternyata tidak sesuai dengan genre yang ada pada film tersebut. Berdasarkan masalah yang disebutkan tujuan dilakukan penelitian ini yaitu untuk menganalisis genre film berdasarkan data subtitle dari film. Tahap-tahap yang dilakukan dalam penelitian ini yaitu pengumpulan data subtitle, penggabungan kata menjadi kalimat, proses labelling, preprocessing, klasifikasi dan visualisasi. Bahasa pemrograman yang digunakan adalah Python, algoritma yang digunakan untuk klasifikasi adalah Naïve Bayes Classifier dan Random Forest dan visualisasi menggunakan Tableau. Dari penelitian ini diperoleh hasil akurasi menggunakan algoritma Random Forest lebih tinggi yaitu 0.841 dibandingkan dengan hasil akurasi Naïve Bayes yaitu 0.682. Salah satu hasil klasifikasi genre terbaik menggunakan Random Forest yaitu film *A Recipe for Seduction* yang memiliki klasifikasi genre *Mystery*, *Romance* dan *Thriller*.

Kata kunci: Film, Genre, Subtitle, Naïve Bayes Classifier, Random Forest

Abstract

Films are a collection of images that undergo a lot of editing processing to achieve satisfactory results. Films have a genre that serves to distinguish the types of films and many films that do not seem to fit the genre. Based on the problem mentioned, the purpose of this study was to analyze the film genre based on the subtitled data from the film. The steps taken in this study were the collection of subtitled data, combining words into sentences, labelling processes, preprocessing, classification and visualization. The programming language used is Python, the algorithm used for classification is Naveve Bayes Classifier and Random Forest and visualization using Tableau. From this study, the accuracy results using Random Forest algorithm were higher at 0.841 compared to the Naveve Bayes accuracy of 0.682. One of the best genre classifications using Random Forest is the film *A Recipe for Seduction* which has the genre classifications *Mystery*, *Romance* and *Thriller*.

Keywords : Film, Genre, Subtitle, Naïve Bayes Classifier, Random Forest

1. Introduction

Film merupakan salah satu media hiburan yang disenangi oleh masyarakat luas. Film dapat dinikmati oleh masyarakat dari berbagai cara misalkan dengan menontonnya di televisi, bioskop maupun internet. Film memiliki *genre* yang berfungsi untuk membedakan jenis film dan banyak film yang ternyata tidak sesuai dengan genre yang ada pada film tersebut. Film bisa memiliki lebih dari satu *genre* yaitu satu sebagai genre utama dan genre lain sebagai pendukung. Macam-macam genre yang diketahui pada umumnya yaitu *action*, *romance*, *fantasy*, *thriller*, *horror*, *comedy* dan masih banyak lagi.

Berdasarkan dari jurnal ilmiah yang berjudul *Comparative Analysis of The Competitiveness between Indonesian Movies against International Movie*, masyarakat Indonesia lebih banyak meminati film impor daripada film lokal, hal tersebut didukung dengan faktor-faktor yang mempengaruhi seseorang dalam memilih film yang akan ditonton [1]. Beberapa faktor yang

dapat mempengaruhi keputusan tersebut seperti genre, aktor atau aktris yang membintangi film tersebut dan rumah produksi.

Dikarenakan banyaknya ketertarikan masyarakat Indonesia pada film impor dibutuhkannya terjemahan agar masyarakat bisa mengetahui inti dari film tersebut. Ada dua macam terjemahan yang biasa digunakan yaitu *Subtitling* dan *Dubbing* [7]. *Dubbing* adalah terjemahan dalam bentuk audio dengan merekam ulang suara pada film dan menyesuaikannya sedangkan *Subtitling* adalah terjemahan berbentuk teks yang biasanya terdapat pada bagian bawah film. Dengan adanya terjemahan, penonton menjadi lebih mudah untuk mengerti alur cerita dengan membaca subtitle tersebut atau dengan mendengarkan percakapan pada film dengan Bahasa sesuai dengan tempat film tersebut ditayangkan. *Subtitle* film tidak hanya berfungsi untuk menerjemahkan Bahasa yang digunakan pada film, tetapi dapat membantu anak-anak berlatih membaca dan bermanfaat untuk penonton disabilitas tuli, autisme, hingga disleksia.

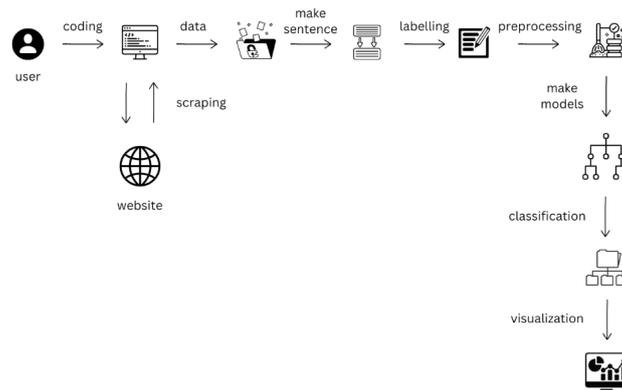
Metode yang digunakan untuk mengklasifikasikan *genre* pada penelitian ini adalah klasifikasi menggunakan *Naïve Bayes* dan *Random Forest*. Pada penelitian yang dilakukan oleh Nikhil Kumar Rajput dan Bhavya Ahuja Grover [6] mendeteksi *genre* berdasarkan subtitle menggunakan english subtitle. Penelitian ini menggunakan 964 film dari 6 genre yaitu action, fantasy, horror, romance, sports dan war dan menggunakan 6 model dan algoritma yang digunakan yaitu *Logistic Regression*, *Support Vector Machine*, *Naïve Bayes Classifier*, *Decision Tree*, *Neural Network* dan *K-Nearest Neighbor*. Hasil yang diperoleh dari penelitian ini yaitu hasil rata-rata yang didapat berada di antara 70-80% dengan menggunakan neural network, dan diperoleh hasil yang bagus dengan menggunakan logistic regression dan KNN dengan hasil 70-80% dan NN mempunyai hasil 77%.

Pada penelitian ini memiliki beberapa tahap yaitu pengumpulan *data*, pembuatan kalimat, *labelling*, *preprocessing*, pembuatan model, klasifikasi dan visualisasi. Pada tahap pengumpulan data menggunakan *web scraping data* yang diambil dari website berupa file subtitle berbahasa Inggris yang akan digunakan untuk penelitian. Tahap selanjutnya yaitu menggabungkan kata pada *file subtitle* menjadi kalimat agar dapat dilakukan proses *labelling*. Pada tahap berikutnya yaitu proses *labelling*, kalimat yang sudah ada diberikan label *genre* yang sesuai dengan kalimat tersebut. Selanjutnya data akan dilakukan *preprocessing* terlebih dahulu untuk membersihkan data yang akan digunakan. Setelah *preprocessing*, data yang sudah dilakukan *labelling* dan *preprocessing* akan digunakan sebagai dataset untuk pembuatan model menggunakan algoritma *Naïve Bayes* dan *Random Forest*. Tahap selanjutnya yaitu mengklasifikasikan data *subtitle* yang belum memiliki *genre* akan dilakukan klasifikasi menggunakan model yang sudah dibuat memakai algoritma *Naïve Bayes* dan *Random Forest*. Tahap terakhir yaitu visualisasi, data yang dipakai adalah data dari hasil klasifikasi yang sudah dilakukan.

Hasil implementasi *Naïve Bayes Classifier* dan *Random Forest* pada proses klasifikasi *genre* memiliki hasil akurasi yang berbeda, algoritma *Random Forest* memiliki akurasi lebih tinggi dibandingkan dengan *Naïve Bayes*. Hasil akurasi pengujian algoritma *Random Forest* adalah 0.841, sedangkan hasil akurasi algoritma *Naïve Bayes* adalah 0.682. Dari hasil klasifikasi *genre* berdasarkan data *subtitle* didapatkan pembagian *genre* sesuai dengan kalimat yang sudah dibagi pada data *subtitle*. Dengan akurasi klasifikasi terbaik yang dilakukan dengan metode *Random Forest*, dapat diketahui lebih banyak *genre alternative* dari suatu film.

2. Research Method / Proposed Method

Penelitian ini dilakukan menggunakan dua algoritma yaitu *Naïve Bayes* dan *Random Forest* untuk mengklasifikasikan *genre* pada data *subtitle*. Langkah-langkah implementasi model pada penelitian adalah melakukan *scrapping data*, pembuatan kalimat dari data *subtitle*, melakukan *labelling* untuk data yang akan digunakan untuk pembuatan model, melakukan *preprocessing* pada data yang sudah dilabeli, pembuatan model menggunakan dua metode *Naïve Bayes* dan *Random Forest*, kemudian melakukan *training* dan *testing*.



Gambar 1. Gambaran Umum Penelitian

Gambar 1 merupakan gambaran umum pada penelitian ini. Pada tahap pengumpulan data menggunakan *web scraping* data yang diambil dari website berupa file subtitle berbahasa Inggris yang akan digunakan untuk penelitian. Data *subtitle* yang didapatkan dari website berisikan waktu penempatan *subtitle* dan teks *subtitle*. Tahap selanjutnya yaitu menggabungkan kata pada *file subtitle* menjadi kalimat agar dapat dilakukan proses *labelling*. Pada proses berikutnya yaitu proses *labelling*, kalimat yang sudah ada diberikan label *genre* yang sesuai dengan kalimat tersebut. Data *subtitle* yang sudah dilabeli akan dilakukan *preprocessing* terlebih dahulu untuk membersihkan data yang akan digunakan saat analisis data. Selanjutnya pada proses klasifikasi dilakukan menggunakan algoritma *Naïve Bayes Classifier* dan *Random Forest* untuk mengklasifikasikan *genre* pada data film. Tahap terakhir yaitu visualisasi hasil dari klasifikasi akan divisualisasikan untuk memudahkan orang awam membaca hasil dari klasifikasi yang telah dilakukan.

2.1. Web Scraping

Web Scraping merupakan salah satu teknik untuk mengumpulkan data dari halaman internet. *Website* yang akan dilakukan *scrape* (proses *web scraping*) adalah *website* yang berisikan data teks subtitle untuk sebuah film. Pada penelitian ini dilakukan *web scraping* pada *website* Podnapisi yang merupakan *web* yang berisikan kumpulan data *subtitle*. *Scraping* dilakukan pada *web* Podnapisi dikarenakan dapat lebih mudah diakses dan juga memiliki variasi Bahasa yang banyak pada data *subtitle* yang tersedia di *website* tersebut.

2.2. Penggabungan Kata Menjadi Kalimat

Pada tahap ini dilakukan penggabungan kata menjadi kalimat setelah data terkumpul. Data yang digunakan adalah data berformat *.srt* atau *.csv* yang berisikan teks *subtitle* dan waktu penempatan teks *subtitle*. Pada tahapan ini kata pada teks *subtitle* akan digabungkan menjadi kalimat menggunakan pemisah antar kalimat yang akan dibuat yaitu waktu antara teks *subtitle*.

2.3. Labelling

Pada tahap ini dilakukan proses *labelling* pada kalimat yang sudah dibentuk pada tahap sebelumnya. Proses *labelling genre* dilakukan secara manual pada setiap kalimat. Proses *labelling* dilakukan untuk mempersiapkan data *training* berisikan *subtitle* yang sudah memiliki label dan diketahui apa saja *genre* yang dimiliki.

2.4. Preprocessing

Pada tahap *preprocessing*, data *subtitle* akan melalui beberapa proses yaitu *case folding*, *cleansing*, *tokenizing* dan *stopwords removal*. *Case folding* dilakukan untuk mengubah semua huruf yang ada menjadi huruf kecil, *cleansing* dilakukan untuk membersihkan dan merapikan data, *tokenizing* dilakukan untuk memisahkan setiap kata yang ada pada setiap baris dan *stopwords removal* dilakukan untuk mengambil kata-kata penting dan membuang kata yang tidak mempunyai arti berdasarkan daftar *stopwords* yang ada.

2.5. Pembuatan Model

Pada tahap ini data subtitle yang sudah di lakukan labelling dan preprocessing akan digunakan sebagai dataset untuk pembuatan model menggunakan algoritma *Naïve Bayes* dan *Random Forest*.

2.6. Klasifikasi

Pada tahap ini data subtitle yang belum memiliki genre akan dilakukan klasifikasi menggunakan model yang sudah dibuat memakai algoritma *Naïve Bayes* dan *Random Forest*. Hasil dari proses klasifikasi akan berupa klasifikasi genre dari setiap baris pada data *subtitle*.

2.7. Visualisasi

Pada tahap ini hasil dari klasifikasi yang telah dilakukan akan dibuatkan visualisasi dan dibuat menggunakan *Tableau*. Visualisasi dengan *Tableau* adalah proses membuat grafik, diagram, dan tampilan visual lainnya yang menggambarkan data secara jelas dan menarik menggunakan perangkat lunak *Tableau*.

3. Kajian Pustaka

Kajian pustaka adalah proses menghimpun dan memahami berbagai teori yang relevan dengan permasalahan yang ingin diteliti. Beberapa topik yang dibahas dalam kajian pustaka ini meliputi *Web Scraping*, *Python*, *Machine Learning*, serta *Naive Bayes* dan *Random Forest*.

3.1. Web Scraping

Web Scraping merupakan teknik untuk mengekstraksi data dari sebuah web dan kemudian data yang sudah diekstraksi akan disimpan pada sebuah file atau sebuah basis data. *Web Scraping* dapat dilakukan otomatis menggunakan bot atau crawler web atau secara manual [3]. Melakukan *web scraping* secara manual memiliki beberapa cara yang dapat dilakukan yaitu menggunakan *Regular Expression*, *parsing DOM*, *parsing HTML* dan menggunakan *XPath*. Tujuan pengambilan data menggunakan Teknik web scraping yaitu bagaimana cara dalam mendapatkan data/informasi melalui penarikan dan ekstraksi data dengan *size* yang beragam yang diambil pada *website* dan digunakan bagi kepentingan lain [5].

3.2. Python

Bahasa pemrograman *Python* merupakan salah satu Bahasa pemrograman yang dapat melakukan eksekusi sejumlah instruksi secara langsung dengan metode orientasi objek. *Python* pertama kali dikenalkan kepada public pada tahun 1991 oleh programmer berkebangsaan dari Belanda yaitu Guido van Rossum. Keunggulan menggunakan Bahasa pemrograman *Python* menurut [4] yaitu mudah digunakan dalam pengembangan sebuah produk perangkat lunak, perangkat keras, *Internet of Things* (IoT), aplikasi *web*, maupun video game, kode mudah dipahami dan bahasa pemrograman ini memiliki library yang sangat banyak dan merupakan Bahasa yang mendukung ekosistem *Internet of Things* dengan sangat baik.

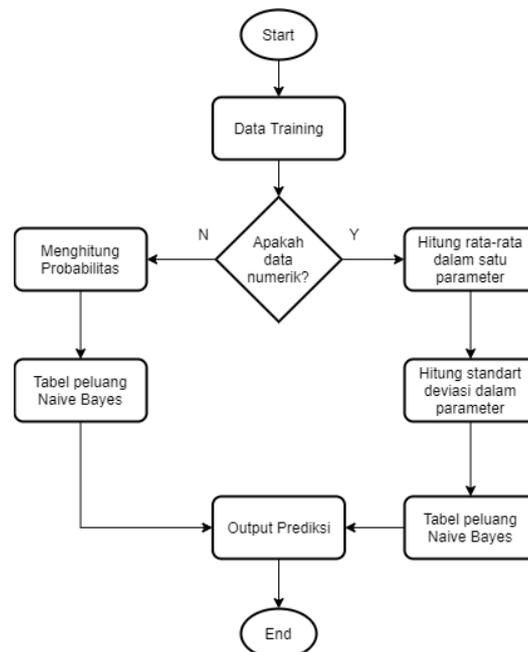
3.3. Machine Learning

Machine learning merupakan salah satu cabang dari *artificial intelligence* (AI) atau kecerdasan buatan. *Machine learning* adalah mesin pembelajaran yang dirancang untuk belajar secara otomatis tanpa instruksi pengguna. *Machine learning* didasarkan pada matematika, statistik, penambangan data, dan ilmu lainnya. *Machine learning* pertama kali dikemukakan oleh beberapa ahli matematika yaitu Adrien Marie Legendre, Thomas Bayes dan Andrey Markov di tahun 1920-an dengan mengemukakan dasar-dasar *machine learning* dan konsepnya. *Machine learning* memiliki dua algoritma yaitu *supervised learning* dan *unsupervised learning*.

3.4. Naïve Bayes

Naive Bayes atau yang dikenal juga *Naive Bayes Classifier* merupakan metode pengklasifikasian yang menerapkan teknik *supervised* klasifikasi objek di masa depan dengan menetapkan label pada data menggunakan probabilitas bersyarat. Istilah *supervised* merujuk pada klasifikasi training data yang sudah diberi label. Manfaat dari menggunakan algoritma *Naive Bayes* adalah bahwa metode ini efisien dengan memerlukan jumlah data pelatihan yang relatif kecil untuk menentukan parameter estimasi yang dibutuhkan dalam proses klasifikasi.

Selain itu, *Naive Bayes* cenderung memberikan hasil yang lebih baik daripada yang diharapkan dalam berbagai situasi dunia nyata yang kompleks. Pada algoritma *Naive Bayes Classifier* terdapat dua macam yaitu data jenis kategori dan data jenis numerik, dalam tahapan perhitungan datanya dapat dilihat pada gambar berikut [2].



Gambar 2. Flowchart Tahapan Algoritma Naive Bayes

Gambar 2 merupakan *flowchart* tahapan menggunakan algoritma *Naive Bayes*. Langkah pertama yang dilakukan yaitu membaca data *training* yang telah disediakan kemudian akan dihitung jumlah dan probabilitas dari data tersebut. Jika data yang digunakan numerik maka dilakukan perhitungan nilai *mean* atau rata-rata dan standar deviasi dari masing-masing parameter.

3.5. Random Forest

Random Forest adalah metode klasifikasi yang merupakan bagian dari decision tree. Metode ini menggabungkan beberapa pohon keputusan (decision trees) yang dikenal sebagai "forest" untuk membuat prediksi yang lebih akurat dan stabil. Berikut adalah langkah-langkah utama yang digunakan dalam penggunaan metode Random Forest:

1. Pengumpulan Data: Mengumpulkan data yang akan digunakan untuk data training yang terdiri dari input fitur dan output target.
2. Pemilihan Sampel: Dalam setiap iterasi, secara acak ambil sejumlah sampel dari data training dengan penggantian (bootstrap sampling).
3. Membangun Pohon Keputusan: Untuk setiap sampel yang diambil, dibangun pohon keputusan (decision tree). Saat membangun tree, pada setiap langkah dipilih fitur terbaik untuk membagi data dan dilakukan secara acak. Hal ini mencegah satu fitur yang kuat mendominasi proses pembelajaran.
4. Membangun Forest: Ulangi langkah ke-3 dengan sejumlah tree yang ditentukan sebelumnya untuk membentuk forest, di mana setiap tree dibangun dengan sampel yang berbeda secara acak.
5. Prediksi: Pada proses klasifikasi, prediksi akhir didasarkan pada mayoritas suara dari semua pohon dalam forest.

4. Result and Discussion

Pada bagian ini akan membahas mengenai hasil dari penelitian analisis *genre* film berdasarkan data *subtitle* yang diambil dari *website* Podnapisi. Data diambil selanjutnya

dilakukan pembuatan kalimat kemudian disimpan di *google drive*. Analisis dilakukan di *Google Collab* dan Visualisasi dilakukan menggunakan *Tableau*.

4.1. Hasil Analisis

Hasil pengujian model analisis dengan metode *Naïve Bayes* dan *Random Forest* menggunakan 5089 baris data kalimat *subtitle* yang sudah dilabeli menghasilkan nilai akurasi tertinggi sebesar 0.84 pada metode *Random Forest*. Berikut table hasil pengujian system pada table 1.

Tabel 1 Hasil Pengujian Model

	Naïve Bayes	Random Forest
Accuracy	0.68	0.84
Precision	0.71	0.85
Recall	0.68	0.84
F1	0.62	0.83

Tabel 1 merupakan hasil pengujian model menggunakan metode *Naïve Bayes* dan *Random Forest*. Hasil pengujian menunjukkan model dengan metode *Random Forest* memiliki nilai akurasi lebih tinggi yaitu 0.84 dibandingkan dengan nilai akurasi dengan metode *Naïve Bayes* yaitu 0.68.

4.2. Hasil Klasifikasi

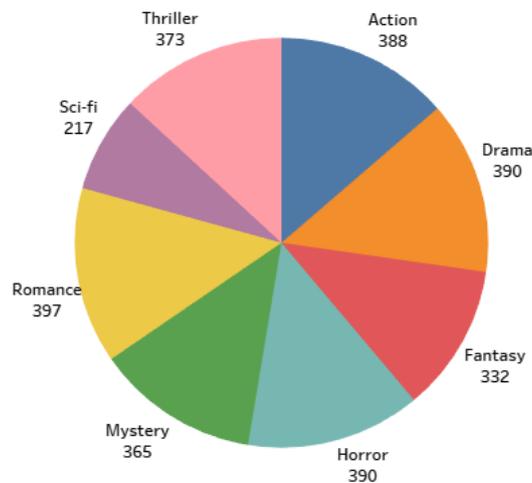
Hasil dari penelitian ini adalah mengklasifikasikan data *subtitle film* yang belum memiliki *genre* menggunakan metode *Naïve Bayes* dan *Random Forest*. Data *subtitle film* yang akan di klasifikasikan yaitu 458 *film* yang memiliki total 16728 baris.

	Judul	Tahun	Genre
0	1917.0	2019	[Action, Drama, Fantasy, Mystery, Romance, Thr...
1	12 Mighty Orphans	2021	[Action, Drama, Fantasy, Mystery, Romance, Thr...
2	15 Minutes Of War	2019	[Action, Drama, Fantasy, Horror, Romance, Thri...
3	400 Bullets	2021	[Action, Drama, Horror, Mystery, Romance, Sci-...
4	90 ML	2019	[Drama, Fantasy, Romance, Thriller]
...
454	Yalda A Night For Forgiveness	2019	[Drama, Fantasy, Romance, Thriller]
455	You People	2023	[Romance, Thriller]
456	Your Place Or Mine	2023	[Action, Drama, Romance, Thriller]
457	Yuko No Tenbin	2020	[Action, Drama, Fantasy, Romance, Sci-fi, Thri...
458	it's a Wonderful Binge	2022	[Romance, Thriller]

	Judul	Tahun	Genre
0	1917.0	2019	[Action, Drama, Fantasy, Horror, Mystery, Roma...
1	12 Mighty Orphans	2021	[Action, Drama, Fantasy, Horror, Mystery, Roma...
2	15 Minutes Of War	2019	[Action, Drama, Fantasy, Horror, Mystery, Roma...
3	400 Bullets	2021	[Action, Drama, Fantasy, Horror, Mystery, Roma...
4	90 ML	2019	[Action, Drama, Fantasy, Horror, Mystery, Roma...
...
454	Yalda A Night For Forgiveness	2019	[Action, Drama, Fantasy, Horror, Mystery, Roma...
455	You People	2023	[Action, Drama, Horror, Romance, Thriller]
456	Your Place Or Mine	2023	[Action, Drama, Fantasy, Horror, Romance]
457	Yuko No Tenbin	2020	[Action, Drama, Fantasy, Horror, Mystery, Roma...
458	it's a Wonderful Binge	2022	[Action, Drama, Horror, Romance, Thriller]

Gambar 3. Hasil Klasifikasi Genre Setiap Judul Film

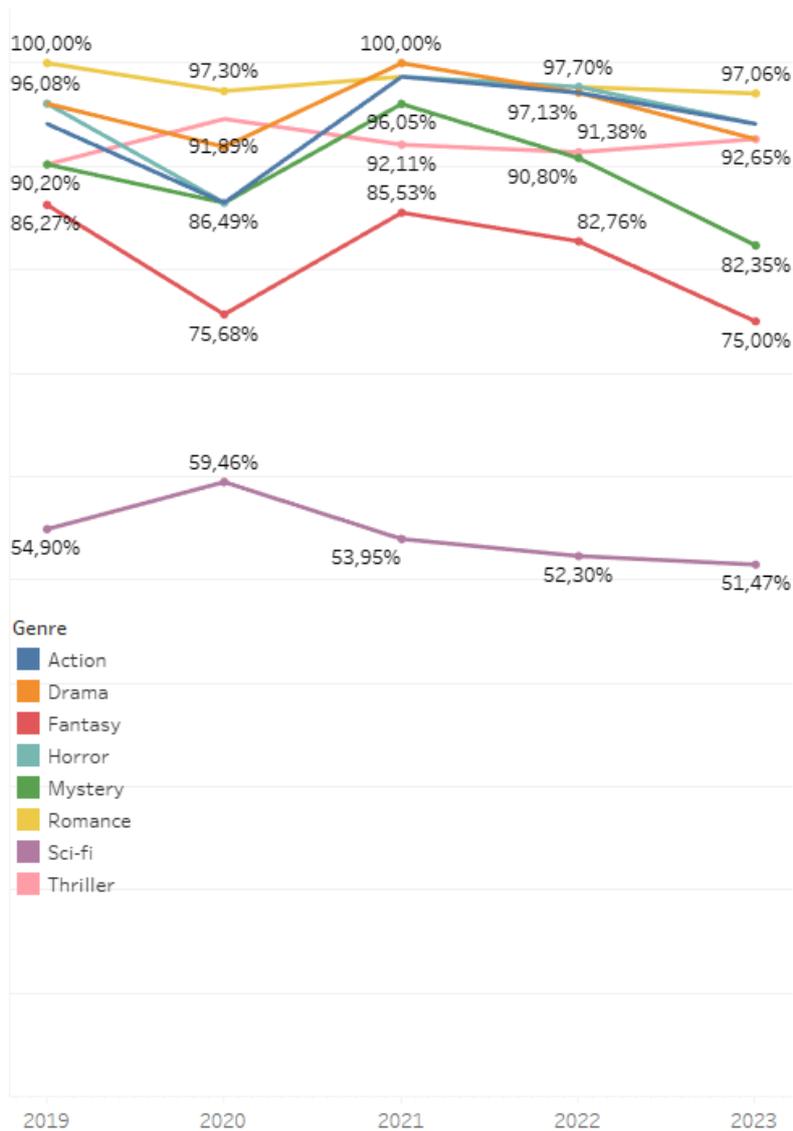
Gambar 3 merupakan hasil dari penggabungan klasifikasi genre kalimat subtitle sesuai judul film. Pada gambar 3 dapat dilihat hasil klasifikasi genre yang didapat pada setiap film. Gambar bagian atas merupakan hasil klasifikasi menggunakan metode *Naïve Bayes* dan gambar bagian bawah merupakan hasil klasifikasi menggunakan metode *Random Forest*.



Gambar 4. Visualisasi Total Jumlah Per Genre Tahun 2019-2023

Gambar 4 merupakan visualisasi total jumlah persebaran per-genre film pada tahun 2019-2023. Dari total 406 Film, terdapat 388 film mempunyai genre 'Action', 390 film mempunyai genre 'Drama', 332 film mempunyai genre 'Fantasy', 390 film mempunyai genre 'Horror', 365 film mempunyai genre 'Mystery', 397 film mempunyai genre 'Romance', 217 film mempunyai genre 'Sci-Fi' dan 373 film mempunyai genre 'Thriller'.

Grafik Persentase Genre per Tahun



Gambar 5. Grafik Persentase Genre Per Tahun 2019-2023

Gambar 4 merupakan visualisasi grafik persentase *genre* pada tahun 2019 hingga 2023. Pada grafik 4 dapat dilihat perbandingan *genre* yang tersebar pada setiap tahunnya yaitu pada tahun 2019 *genre* 'Romance' terdapat pada semua film yang ada pada dataset dan *genre* 'Sci-Fi' memiliki persentase paling rendah. Pada tahun 2020 terdeteksi *genre* 'Romance' mempunyai persentase paling tinggi dan *genre* 'Sci-Fi' memiliki persentase paling rendah. Pada tahun 2021 terdeteksi *genre* 'Drama' terdapat pada semua film yang ada pada dataset dan *genre* 'Sci-Fi' memiliki persentase paling rendah. Pada tahun 2022 terdeteksi *genre* 'Drama' dan 'Romance' mempunyai persentase paling tinggi dan *genre* 'Sci-Fi' memiliki persentase paling rendah. Pada tahun 2023 terdeteksi *genre* 'Romance' mempunyai persentase paling tinggi dan *genre* 'Sci-Fi' memiliki persentase paling rendah.

5. Conclusion

Hasil implementasi *Naïve Bayes Classifier* dan *Random Forest* pada proses klasifikasi *genre* memiliki hasil akurasi yang berbeda, algoritma *Random Forest* memiliki akurasi lebih tinggi dibandingkan dengan *Naïve Bayes*. Hasil akurasi pengujian algoritma *Random Forest* adalah 0.841, sedangkan hasil akurasi algoritma *Naïve Bayes* adalah 0.682. Dari hasil

klasifikasi genre berdasarkan data *subtitle* didapatkan pembagian *genre* sesuai dengan kalimat yang sudah dibagi pada data subtitle. Dengan akurasi klasifikasi terbaik yang dilakukan dengan metode Random Forest, dapat diketahui lebih banyak *genre alternative* dari suatu film seperti pada tiga film yang memiliki akurasi klasifikasi yang paling bagus pada film *A Recipe For Seduction* memiliki klasifikasi hasil *genre Mystery, Romance* dan *Thriller*, film *All The President's Men* memiliki klasifikasi hasil *genre Action, Drama, Mystery, Romance* dan *Thriller* dan film *Blood Bath In The Nameless River* memiliki klasifikasi hasil *genre Action, Drama, Horror, Mystery* dan *Thriller*.

References

- [1] M. F. F. Mardianto, D. R. Hastuti, D. A. Husada, and R. Andriawan, "Comparative Analysis of the Competitiveness Between Indonesian Movies Against International Movie, as a Reference in Developing Indonesia's Cinema and Curriculum About cinematography," *Int. J. Innov. Creat. Chang.*, vol. 5, no. 3, pp. 685–707, 2019, [Online]. Available: <https://scholar.unair.ac.id/en/publications/comparative-analysis-of-the-competitiveness-between-indonesian-mo>
 - [2] M. R. Effendi, "Akurasi Data Mining Untuk Menghasilkan Pola Kelulusan Mahasiswa dengan Metode NAÏVE BAYES," *J. Sist. Inf. Univ. Suryadarma*, vol. 3, no. 2, pp. 101–106, 2016, doi: <https://doi.org/10.35968/jsi.v3i2.66>.
 - [3] V. A. Flores, P. A. Permatasari, and L. Jasa, "Penerapan Web Scraping Sebagai Media Pencarian dan Menyimpan Artikel Ilmiah Secara Otomatis Berdasarkan Keyword," *Maj. Ilm. Teknol. Elektro*, vol. 19, no. 2, pp. 157–162, 2020, doi: <https://doi.org/10.24843/MITE.2020.v19i02.P06>.
 - [4] T. M. Kadarina and M. H. I. Fajar, "Pengenalan Bahasa Pemrograman Python Menggunakan Aplikasi Games Untuk Siswa/I Di Wilayah Kembangan Utara," *J. Abdi Masy.*, vol. 5, no. 1, pp. 11–16, 2019, doi: <http://dx.doi.org/10.22441/jam.2019.v5.i1.003>.
 - [5] D. D. Ayani, H. S. Pratiwi, and H. Muhandi, "Implementasi Web Scraping untuk Pengambilan Data pada Situs Marketplace," *J. Sist. dan Teknol. Inf.*, vol. 7, no. 4, pp. 257–262, 2019, doi: <http://dx.doi.org/10.26418/justin.v7i4.30930>.
 - [6] N. K. Rajput and B. A. Grover, "A Multi-label Movie Genre Classification Scheme Based on the Movie's Subtitle," *Multimed. Tools Appl.*, vol. 81, no. 22, pp. 32469–32490, 2022, doi: <https://doi.org/10.1007/s11042-022-12961-6>.
 - [7] D. Bordwell and K. Thompson, *Minding Movies: Observations on the Art, Craft, and Business of Filmmaking*. London: The University of Chicago Press, 2011.
-