Analisis Produktivitas Sutradara pada Rekam Jejak Film dengan Metode K-Means

***Analysis of Director's Productivity on Film Track Records with the K-Means Method***

**Kelompok 5, Aryo Bagus K.T,Maulana Syarief H. , Rafli Alfian N.**

Teknik Informatika, Jurusan Teknologi Informasi, Politeknik Negeri Malang

Jl. Soekarno Hatta No. 9 Kota Malang, Indonesia

e-mail: [*aryobagus@polinema.ac.id*](mailto:aryobagus@polinema.ac.id)

(***received*:** ?, ***revised*:** ?, ***accepted*:** ? diisi oleh editor)

**Abstrak**

Analisis produktivitas sutradara pada rekam jejak film dengan metode K-Means adalah sebuah studi yang bertujuan untuk memahami dan mengelompokkan pola produktivitas sutradara berdasarkan jumlah film yang mereka hasilkan dalam periode waktu tertentu. Metode K-Means digunakan untuk mengelompokkan sutradara ke dalam beberapa kelompok berdasarkan kesamaan atribut produktivitas.

Dalam penelitian ini, data rekam jejak film sutradara dikumpulkan dan diproses menggunakan metode K-Means. Metode ini mengelompokkan sutradara berdasarkan tingkat produktivitas mereka, dengan sutradara yang memiliki jumlah film yang lebih besar dikelompokkan sebagai sutradara produktif dan sutradara dengan jumlah film yang lebih sedikit sebagai sutradara kurang produktif.

Hasil analisis ini memberikan wawasan tentang pola produktivitas sutradara dan membantu mengidentifikasi kelompok sutradara yang memiliki produktivitas tinggi atau rendah. Informasi ini dapat digunakan oleh praktisi industri film, peneliti, dan pengambil keputusan untuk mengoptimalkan potensi sutradara, memahami tren industri film, dan mempengaruhi pengambilan keputusan terkait produksi film.

Dengan menerapkan metode K-Means pada analisis produktivitas sutradara, penelitian ini memberikan pendekatan yang sistematis dan berbasis data untuk memahami dan mengklasifikasikan pola produktivitas sutradara dalam industri film.

**Kata kunci:** Sutradara, K-means, Film Indonesia

***Abstract***

*The analysis of director productivity in film track records using the K-Means method aims to understand and categorize patterns of director productivity based on the number of films they have produced within a specific time period. The K-Means method is utilized to cluster directors into groups based on similarities in productivity attributes.*

*In this study, director film track record data is collected and processed using the K-Means method. This method clusters directors based on their level of productivity, with directors who have a larger number of films being categorized as highly productive, while those with fewer films are classified as less productive directors.*

*The findings of this analysis provide insights into patterns of director productivity and aid in identifying groups of directors with high or low productivity levels. This information can be utilized by film industry practitioners, researchers, and decision-makers to optimize director potential, understand trends in the film industry, and inform decision-making related to film production.*

*By applying the K-Means method to analyze director productivity, this research offers a systematic and data-driven approach to comprehend and classify patterns of director productivity within the film industry.*

***Keywords:*** *Director, K-means, Indonesian Film*

# Pendahuluan (or Introduction)

Industri film merupakan salah satu industri yang terus berkembang dengan pesat di seluruh dunia. Di balik setiap film yang diproduksi, terdapat seorang sutradara yang bertanggung jawab dalam mengarahkan proses kreatif dan menghasilkan karya yang memikat penonton. Produktivitas sutradara, yang dapat diukur melalui jumlah film yang mereka hasilkan dalam periode waktu tertentu, menjadi faktor penting dalam menggambarkan karier dan kontribusi mereka dalam industri ini.

Dalam upaya untuk memahami pola produktivitas sutradara dan mengidentifikasi kelompok-kelompok yang ada berdasarkan rekam jejak film mereka, analisis data dapat menjadi alat yang sangat bermanfaat. Salah satu metode analisis yang digunakan secara luas adalah metode K-Means. Metode K-Means merupakan metode pengelompokan data yang mengelompokkan objek-objek ke dalam beberapa kelompok berdasarkan kesamaan atribut atau karakteristik yang dimiliki.

Dalam konteks ini, analisis produktivitas sutradara pada rekam jejak film dengan metode K-Means bertujuan untuk mengidentifikasi pola-pola produktivitas sutradara dan mengelompokkan mereka berdasarkan tingkat produktivitas yang dimiliki. Dengan menggunakan metode ini, kita dapat menggali wawasan yang lebih dalam tentang tren produktivitas sutradara dalam industri film, mengidentifikasi sutradara yang memiliki produktivitas tinggi atau rendah, serta memahami faktor-faktor yang mungkin mempengaruhi tingkat produktivitas sutradara.

Pada tulisan ini, kami akan menjelaskan lebih lanjut tentang analisis produktivitas sutradara dengan metode K-Means. Kami akan memperkenalkan konsep dasar K-Means dan bagaimana metode ini dapat diterapkan dalam mengelompokkan sutradara berdasarkan rekam jejak film mereka. Selanjutnya, kami akan membahas hasil dan kesimpulan yang dapat ditarik dari analisis ini, serta implikasinya terhadap industri film secara keseluruhan.

Dengan memahami produktivitas sutradara melalui analisis yang sistematis dan berbasis data, kita dapat memperoleh wawasan yang berharga tentang tren industri film, potensi sutradara yang menonjol, serta faktor-faktor yang dapat memengaruhi produktivitas dalam industri ini. Analisis ini dapat memberikan panduan bagi para praktisi industri, peneliti, dan pengambil keputusan untuk mengoptimalkan potensi sutradara dan mengembangkan industri film ke depan.

# Tinjauan Literatur (or Literature Review)

**K-means**

K-Means Clustering adalah metode untuk mengkategorikan atau pengelompokan sekelompok objek sesuai dengan atribut yang sama atau karakteristik ke dalam sejumlah groups (jumlah bilangan bulat positif). Ini mendefinisikan sebuah cluster oleh massa yang yang mewakili mean dari cluster.

**Kaggle**

Kaggle adalah kuis dan platform sumber daya yang disukai oleh ilmu data dan praktisi dan peneliti pembelajaran mesin. Kaggle menyediakan akses ke berbagai kumpulan data publik yang dapat digunakan untuk analisis, penelitian, dan pengembangan model prediktif. Platform Kaggle memungkinkan pengguna untuk berpartisipasi dalam kuis di mana mereka dapat bersaing dengan para profesional dan pakar di bidangnya untuk memecahkan tantangan analitik yang kompleks. Peserta dapat mengirimkan solusi mereka dalam bentuk kode dan model prediktif dan akan dinilai berdasarkan kinerja dan keakuratan solusi yang mereka ajukan.

**Cloudera**

Cloudera adalah perusahaan teknologi yang fokus pada solusi dan platform big data open source. Mereka menyediakan solusi untuk manajemen dan analisis data besar, keamanan informasi, dan operasi data besar untuk perusahaan di berbagai industri. Cloudera didasarkan pada platform Apache Hadoop, yang merupakan kerangka kerja sumber terbuka untuk memproses, menyimpan, dan menganalisis data besar. Mereka juga menawarkan distribusi komersial Hadoop yang dikenal sebagai Cloudera Distribution, yang mencakup komponen-komponen seperti Hadoop Distributed File System, MapReduce, Hive, Impala, HBase, dan lain-lain. Selain itu, Cloudera juga menyediakan produk dan layanan tambahan yang membantu organisasi mengelola dan menganalisis big data secara efektif. Mereka telah digunakan oleh perusahaan di berbagai industri, termasuk perbankan, telekomunikasi, kesehatan, e-commerce, dan lain-lain.

**Spark**

Apache Spark adalah kerangka kerja komputasi terdistribusi yang dirancang untuk komputasi yang cepat dan terukur. Spark mengatasi kinerja dan keterbatasan kerangka kerja komputasi lain seperti Apache Hadoop. Salah satu fitur utama Spark adalah kemampuannya untuk melakukan pemrosesan data di memori, yang menghasilkan pemrosesan yang lebih cepat daripada metode tradisional yang menggunakan disk sebagai sumber data. Spark mendistribusikan pemrosesan dan perhitungan data di beberapa node dalam sebuah cluster, memungkinkan pemrosesan dan penskalaan paralel untuk mudah dilakukan dengan menambahkan lebih banyak sumber daya. Spark menawarkan berbagai antarmuka pemrograman, seperti Scala, Python, Java, dan R, sehingga memudahkan pengembang untuk memanfaatkan kekuatan Spark dengan bahasa pemrograman pilihan mereka. Spark juga menyediakan berbagai modul dan pustaka yang mendukung tugas-tugas pemrosesan data, termasuk pemrosesan batch (Spark Core), streaming data real-time (Spark Streaming), analisis data interaktif (Spark SQL), pemrosesan pembelajaran mesin (Spark MLlib), dan pemrosesan grafik (GraphX). Spark telah menjadi pilihan populer untuk berbagai kasus komputasi, termasuk analisis data besar, pembelajaran mesin, pemrosesan aliran data real-time, dan pemrosesan grafik.

**Google Colab**

Google Colab adalah platform pengembangan dan eksekusi kode berbasis cloud yang disediakan oleh Google. Colab memungkinkan pengguna untuk menulis, menjalankan, dan membagikan kode Python secara interaktif melalui browser tanpa perlu mengatur lingkungan pengembangan lokal. Colab berjalan di infrastruktur Google Cloud dan menyediakan lingkungan yang kuat untuk pengembangan dan analisis data. Di Colab, pengguna dapat membuat dan menjalankan blok kode Python serta menggabungkannya dengan teks naratif, visualisasi, dan konten lain dalam format yang terstruktur.

# Metode Penelitian (or Research Method)

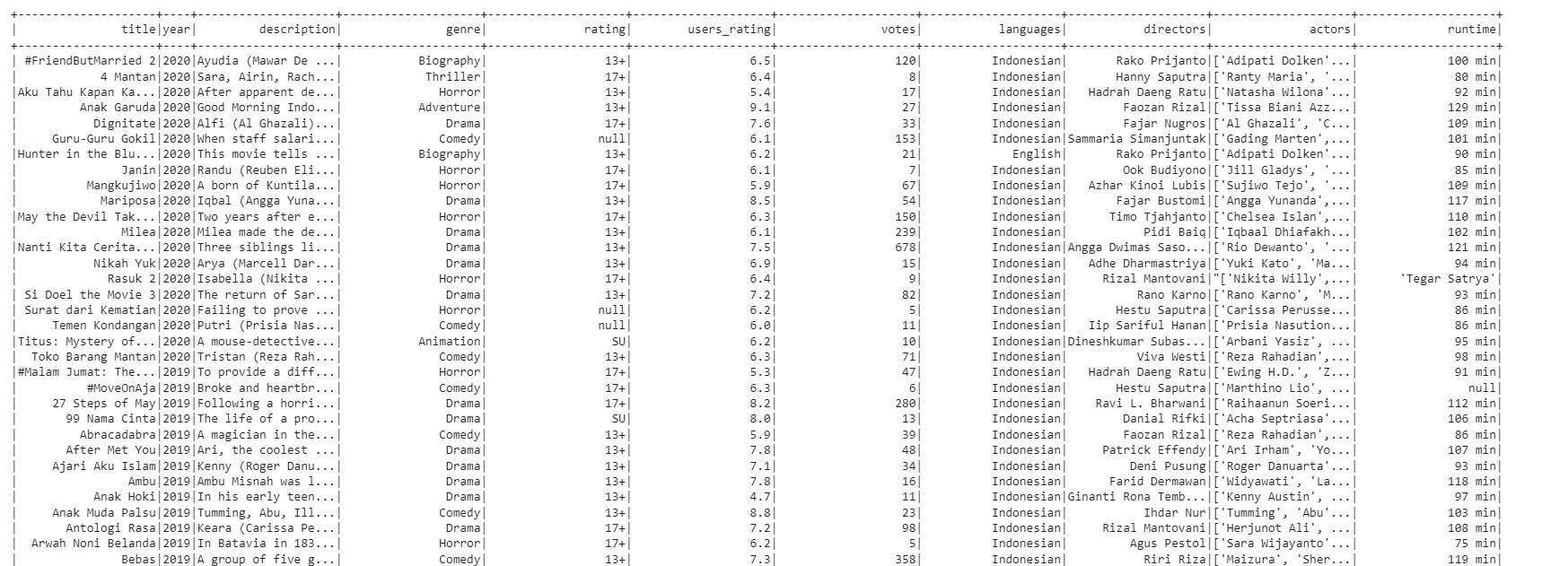
Metode klasifikasi untuk melihat sutradara dengan rekam jejak film pada penelitian ini menggunakan algoritma K-Means. Sedangkan tahapan penelitian ini menggunakan tahapan datamining, terdiri dari analisis kebutuhan pengumpulan data, preprocessing, dan visualisasi data.



**Gambar 1. Big data lifecycle**

**3.1 Pengumpulan data**

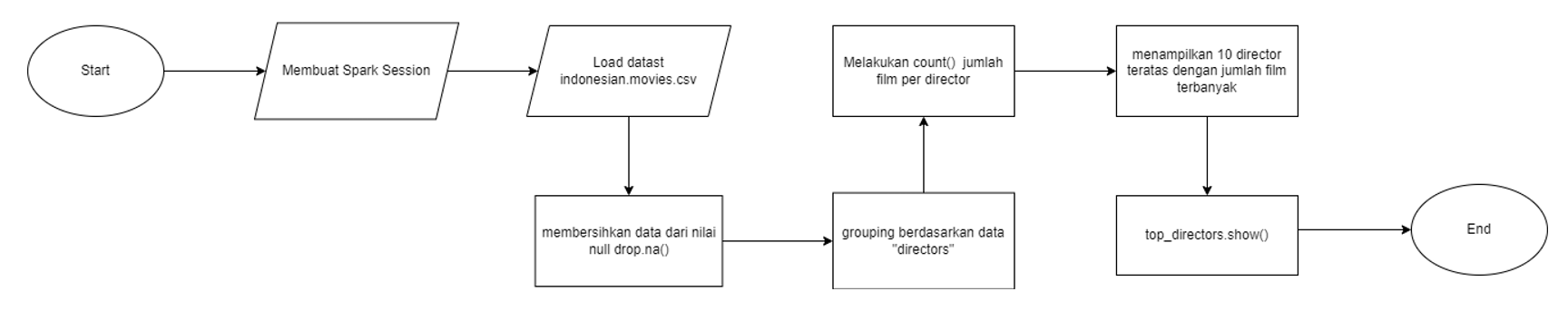
Kami mengambil dataset mengenai rating film indonesia.Langkah awal kami mengimport dataset dengan nama ‘indonesian\_movies.csv’. Sebagian data dapat dilihat pada gambar 1



**Gambar 2. Dataset**

**3.2 Preprocessing**

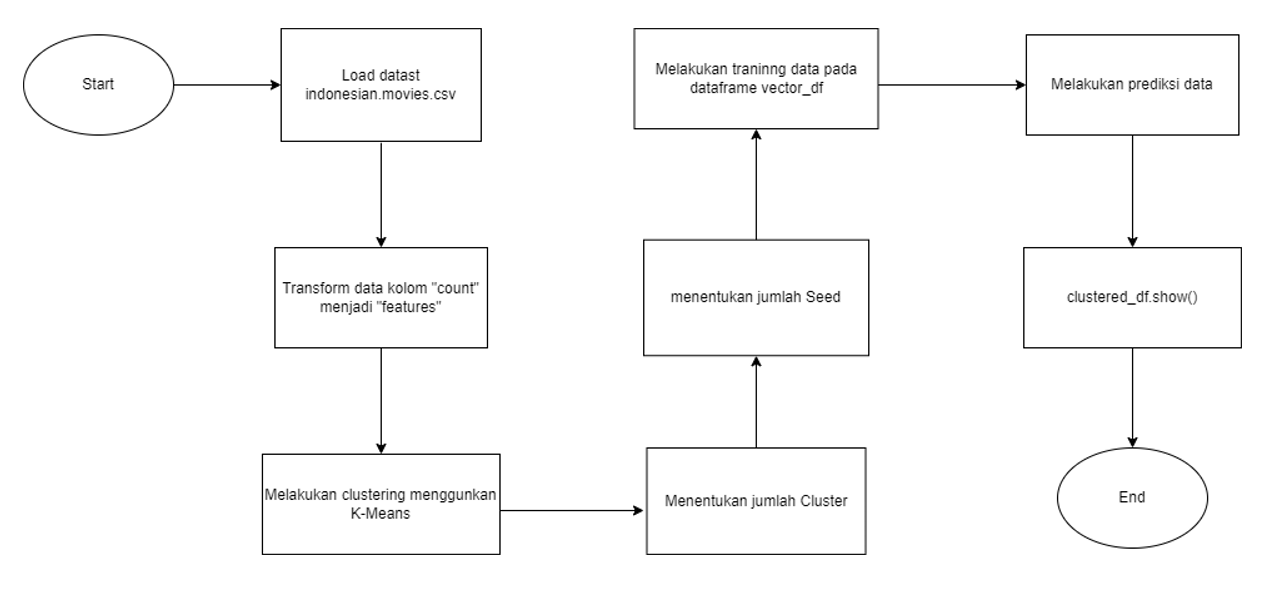
Pada tahap ini data akan dilakukan pembersihan, menghitung jumlah sutradara dalam data, menampilkan sutradara dengan film terbanyak, lalu mempersiapkan untuk clustering.



**Gambar 3. Flowchart preprocessing**

**3.3 Implementasi metode K-means**

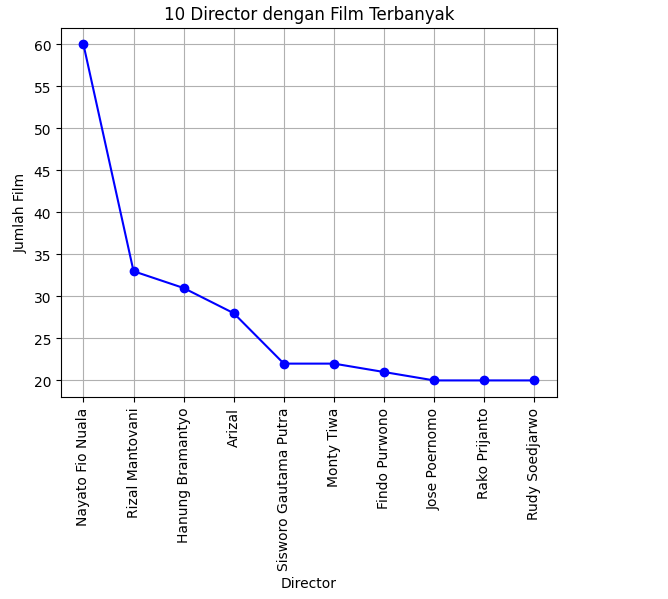
Pada tahap ini clustering data dengan menggunakan metode K-means dengan cara, melakukan pengelompokkan berdasarkan kolom directors dan genre, mengurutkan data berdasarkan urutan kolom genre dan deskripsi kolom count, Mengelompokkan dan agregat data berdasarkan genre, Memilih genre untuk menampilkan directory dengan film terbanyak.



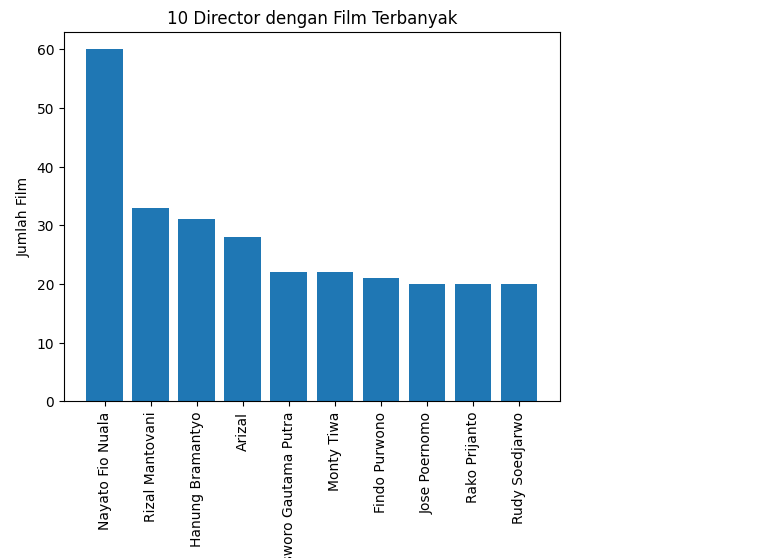
**Gambar 4. Flowchart implementasi metode K-means**

# Hasil dan Pembahasan (or Results and Analysis)

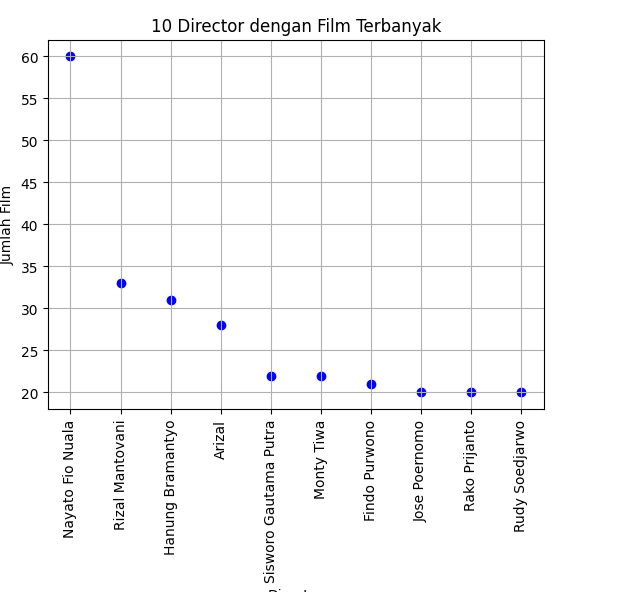
Hasil dari clustering menggunakan metode K-means pada penelitian ini menunjukan sutradara yang sudah rekam jejak film terbanyak. Berdasarkan Analisa data dan melakukan proses pengolahan data menggunakan metode K-means, dapat diamati dari tahun 2010 - 2020 yang diteliti ternyata terdapat 10 nama sutradara yang tercantum dengan rekam jejak film terbanyak.



**Gambar 5. Grafik linear peringkat**



**Gambar 6 Grafik Batang peringkat**

****

**Gambar 7 Grafik scatter plot**

# Kesimpulan (or Conclusion)

Metode K-Means digunakan untuk mengelompokkan sutradara berdasarkan pola produktivitas mereka dalam membuat film. Metode ini dapat membantu mengidentifikasi pola-pola yang muncul dalam rekam jejak film sutradara. Analisis produktivitas sutradara berfokus pada jumlah film yang mereka hasilkan dalam periode waktu tertentu. Dengan menggunakan metode K-Means, sutradara dapat dikelompokkan menjadi beberapa kelompok berdasarkan tingkat produktivitas mereka. Sutradara dengan produktivitas tinggi dapat diidentifikasi sebagai kelompok yang memiliki jumlah film yang lebih besar dibandingkan dengan kelompok lainnya. Mereka mungkin memiliki kemampuan untuk menghasilkan film secara efisien atau memiliki akses ke sumber daya yang memadai.

Sutradara dengan produktivitas rendah dapat termasuk dalam kelompok yang memiliki jumlah film yang relatif sedikit dalam rekam jejak mereka. Faktor-faktor seperti keterbatasan sumber daya, proyek-proyek yang lebih besar atau lebih kompleks, atau fokus pada kualitas daripada kuantitas film mungkin berperan dalam produktivitas yang lebih rendah. Analisis ini dapat memberikan wawasan tentang tren produktivitas sutradara dari waktu ke waktu. Misalnya, jika terdapat peningkatan jumlah sutradara yang produktif dalam beberapa tahun terakhir, hal ini dapat menunjukkan pertumbuhan industri film atau perubahan dalam preferensi penonton..

# Referensi (Reference)

Penulisan naskah dan sitasi yang diacu dalam naskah ini menggunakan type IEEE, disarankan menggunakan aplikasi manajemen referensi Mendeley. Artikel wajib merujuk minimal 15 referensi primer dan mutakhir (5 tahun terakhir) [Times New Roman, 11, normal], spasi 1. Perhatikan penulisan judul artikel tidak menggunakan huruf kapital semua (rujuk pada pedoman penulisan).

Contoh Penulisan

[1] E. Santucci, L. Didaci, G. Fumera, and F. Roli, “A Parameter Randomization Approach for Constructing Classifier Ensembles,” *Pattern Recognition.*, vol. 69, pp. 1–13, 2017.

[2] P. Panov and S. Džeroski, “Combining Bagging and Random Subspaces to Create Better Ensembles,” in *Proceedings of the 7th International Conference on Intelligent Data Analysis*, 2007, pp. 118–129.

[3] L. I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms: Second Edition*. New Jersey: John Wiley and Sons, 2004.

[4] R. M. Barts, “The Stub Loaded Helix: A Reduced Size Helical Antenna,” *Doctoral Dissertation*, 2003. [Online]. Available: http://hdl.handle.net/10919/29728.