

## PERFORMA ALGORITMA *K-MEANS* DAN *FUZZY C-MEANS* DALAM ANALISIS KLASTER PENDIDIKAN DI TINGKAT KECAMATAN JAKARTA

**Naramia Wijaya**

Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Tarumanagara,  
Jln. Letjen S. Parman No. 1, Jakarta, 11440, Indonesia  
E-mail: naramia1693@gmail.com

### ABSTRAK

Pendidikan yang berkualitas adalah factor penting dalam pembangunan sumber daya manusia yang unggul. Namun, kualitas pendidikan di Indonesia masih beragam, terutama di tingkat kecamatan. Penelitian ini bertujuan untuk mengeksplorasi dan membandingkan performa algoritma *K-Means* dan *Fuzzy C-Means* (FCM) dalam analisis klaster pendidikan di tingkat kecamatan Jakarta. Dalam penelitian ini, data yang digunakan diambil dari dataset pendidikan yang mencakup berbagai variabel terkait kualitas pendidikan di Jakarta. Metode yang digunakan adalah *K-Means Clustering* dan *Fuzzy C-Means Clustering*, dengan eksperimen dilakukan sebanyak 10 kali untuk berbagai jumlah klaster, mulai dari 2 hingga 7 klaster. Hasil eksperimen menunjukkan bahwa kedua algoritma menghasilkan klaster yang berbeda, dengan *Fuzzy C-Means* memberikan hasil yang lebih stabil dan sensitif terhadap perubahan data. Berdasarkan hasil perbandingan tersebut, *Fuzzy C-Means* lebih unggul dalam hal kualitas klaster dibandingkan dengan *K-Means*, meskipun keduanya memberikan informasi yang berguna dalam memahami pola pendidikan di tingkat kecamatan. Penelitian ini menyimpulkan bahwa penggunaan algoritma klasterisasi dapat membantu dalam menganalisis dan meningkatkan kualitas pendidikan di Indonesia, khususnya di Jakarta.

**Kata kunci:** *K-Means*, *Fuzzy C-Means*, Klasterisasi, Pendidikan, Jakarta.

### ABSTRACT

*Quality education is a crucial factor in human resource development. However, the quality of education in Indonesia varies significantly, especially at the district level. This study aims to explore and compare the performance of the K-Means and Fuzzy C-Means algorithms in clustering education data at the district level in Jakarta. The dataset used in this study includes various variables related to education quality in Jakarta. The methods employed are K-Means Clustering and Fuzzy C-Means providing more stable and sensitive results to data changes. Based on this comparison, Fuzzy C-Means outperforms K-Means in terms of cluster quality, although both provide valuable insights into understanding education patterns at the district level. This study concludes that the use of clustering algorithms can assist in analyzing and improving education quality in Indonesia, particularly in Jakarta.*

**Keywords:** *K-Means*, *Fuzzy C-Means*, Clustering, Education, Jakarta

## 1. PENDAHULUAN

Pendidikan merupakan salah satu faktor yang penting dalam menentukan pembangunan suatu negara, termasuk Indonesia [1]. Pendidikan yang berkualitas merupakan elemen kunci dalam pengembangan sumber daya manusia (SDM) di Indonesia [2], terutama untuk mengatasi masalah ketimpangan pembangunan antar daerah [3-4]. Berbagai faktor seperti akses terhadap fasilitas pendidikan, infrastruktur, dan tingkat perekonomian daerah mempengaruhi kualitas pendidikan di setiap kecamatan [5]. Dalam konteks ini, penting untuk menganalisis dan mengevaluasi distribusi serta kualitas pendidikan di berbagai wilayah menjadi hal yang penting [6-7]. Salah satu cara untuk menganalisis pola-pola tersebut adalah dengan menggunakan teknik klasterisasi yang mengelompokkan data berdasarkan karakteristik yang dimilikinya, yang memungkinkan untuk mengelompokkan daerah berdasarkan kriteria tertentu. Metode klasterisasi seperti *K-Means* dan *Fuzzy C-Means* banyak digunakan dalam berbagai bidang, termasuk untuk menganalisis data

pendidikan, kesehatan, dan ekonomi [8]. *K-Means* sering digunakan dalam analisis klusterisasi, namun memiliki keterbatasan pada data yang bersifat *fuzzy*, yaitu data tanpa batas yang tegas antara satu kluster dengan kluster lainnya [9]. Dalam penelitian ini, digunakan dua algoritma klustering yang populer yaitu *K-Means* dan *Fuzzy C-Means*, digunakan untuk menganalisis data pendidikan di Jakarta, dengan fokus pada kecamatan sebagai unit analisisnya [10].

Algoritma *K-Means* adalah algoritma pembelajaran tidak terawasi yang membagi data ke dalam beberapa kluster berdasarkan kedekatannya terhadap pusat kluster (*centroid*) yang dihitung secara *iterative* [11]. Di sisi lain, *Fuzzy C Means* (FCM) merupakan variasi dari *K-Means* yang memungkinkan data untuk memiliki keanggotaan yang lebih dari satu kluster, memberikan fleksibilitas dalam penentuan keanggotaan kluster berdasarkan derajat kedekatannya. Kedua algoritma ini memiliki pendekatan yang berbeda dalam membentuk kluster yang dapat memengaruhi hasil analisis [12]. Penelitian ini bertujuan untuk membandingkan kinerja kedua algoritma tersebut dalam melakukan klustering data pendidikan pada tingkat kecamatan Jakarta, dengan menggunakan metrik *Silhouette Score* sebagai ukuran kualitas kluster. Selain itu, penelitian ini juga menganalisis pengaruh jumlah kluster terhadap performa kedua algoritma [13].

Dalam bidang analisis klustering, *K-Means* merupakan salah satu algoritma yang paling banyak digunakan karena kesederhanaannya dan efisiensinya dalam menangani data besar [14]. Beberapa penelitian yang relevan dengan penggunaan *K-Means* diantaranya adalah pengelompokan nilai siswa menggunakan metode *K-Means* yang dilakukan pada penelitian sebelumnya [10] yang menunjukkan hasil yang baik dalam mengelompokkan siswa berdasarkan nilai akademik mereka. Penelitian ini mengindikasikan bahwa *K-Means* dapat digunakan untuk tujuan klasifikasi dalam pendidikan untuk menentukan pola belajar siswa di berbagai wilayah.

Di sisi lain, *Fuzzy C-Means* (FCM) menawarkan pendekatan yang lebih fleksibel dengan memungkinkan data untuk berada dalam beberapa kluster sekaligus [15]. Penelitian yang pernah dilakukan Lee pada tahun 2019 tentang aplikasi FCM dalam deteksi kanker payudara menunjukkan bahwa FCM dapat memberikan hasil yang lebih akurat dalam pengelompokan data yang memiliki ketidakpastian atau ambiguitas tinggi [16]. FCM juga banyak digunakan dalam aplikasi medis lainnya seperti deteksi penyakit kulit oleh Wang yang memanfaatkan klusterisasi berbasis FCM untuk mengidentifikasi jenis kanker kulit [17].

## 2. METODE PENELITIAN

Data yang dipakai dalam penelitian adalah data pendidikan tahun 2022 yang mencakup berbagai indikator pendidikan di tingkat kecamatan Jakarta. Data ini mencakup informasi tentang periode data, kota atau kabupaten, kecamatan, tingkat pendidikan terakhir, jenis kelamin, dan jumlah penduduk di setiap kecamatan. Data yang digunakan terdiri dari beberapa variabel numerik yang mencerminkan kualitas pendidikan di Jakarta. Sebelum dilakukan klustering, data ini terlebih dahulu diproses dan dinormalisasi menggunakan *MinMaxScaler* untuk memastikan bahwa semua fitur memiliki rentang nilai yang sama.

### 2.1. Proses *Clustering*

Proses percobaan dilakukan dengan menggunakan dua algoritma klustering, yaitu *K-Means* dan *Fuzzy C-Means* (FCM). Algoritma *K-Means* membagi data ke dalam kluster dengan cara meminimalkan jarak total antara titik data dengan pusat kluster. Di sisi lain, FCM mengizinkan setiap titik data untuk mewakili keanggotaan terhadap beberapa kluster, dengan keanggotaan yang dihitung berdasarkan jarak relatif dari pusat kluster [18].

Pada percobaan ini kedua algoritma diuji dengan jumlah kluster yang bervariasi antara 2 hingga 7 kluster. Selain itu, eksperimen dilakukan sebanyak 10 iterasi untuk menguji stabilitas dan konsistensi hasil klustering. *Silhouette Score* digunakan untuk mengukur kualitas kluster yang terbentuk dengan nilai lebih tinggi menunjukkan kluster yang lebih baik.

## 2.2. *K-Means Clustering*

*K-Means* adalah sebuah algoritma yang sering dipergunakan untuk membagi data ke dalam beberapa kluster berdasarkan kedekatannya dengan pusat kluster atau yang seringkali dikenal dengan *centroid* [19]. Berikut adalah langkah-langkah dan rumus dasar dalam *K-Means*:

Langkah-langkah *K-Means*:

1. Inisialisasi: Menentukan jumlah kluster  $k$ , dan pilih secara acak  $k$  titik sebagai pusat kluster awal atau *centroid*.
2. Penugasan Data: Menentukan titik data yang masuk ke dalam kluster dengan menghitung jarak antara titik data dan setiap *centroid*. Titik data akan ditugaskan ke kluster dengan jarak terdekat.

$$\text{Jarak}(x_i, c_k) = \sqrt{\sum_{j=1}^d (x_{ij} - c_{kj})^2}$$

Di mana:

- $x_i$  adalah data titik ke -  $i$
- $c_k$  adalah *centroid* kluster ke- $k$
- $d$  adalah dimensi data

3. *Update Centroid*: Setelah penugasan titik data selesai, hitung ulang posisi *centroid* untuk setiap kluster. *Centroid* baru dihitung dengan mengambil rata-rata titik data dalam kluster tersebut.

$$c_k = \frac{1}{|C_k|} \sum_{x_i \in C_k} x_i$$

Di mana:

- $C_k$  adalah himpunan titik data dalam kluster ke- $k$
- $|C_k|$  adalah jumlah titik dalam kluster ke- $k$
- $x_i$  adalah titik data yang ada di kluster ke- $k$ .

4. Ulangi: Langkah 2 dan 3 diulang hingga *centroid* stabil, yaitu tidak ada perubahan signifikan dalam posisi *centroid*.

Tujuan dari *K-Means* adalah meminimalkan fungsi objektif yang dikenal sebagai *within-cluster sum of squares* (WCSS) atau inerti [20]:

$$J = \sum_{k=1}^K \sum_{x_i \in C_k} \|x_i - c_k\|^2$$

Di mana:

- $J$  adalah total jarak kuadrat dari semua titik data ke *centroid* kluster mereka.
- $C_k$  adalah kluster ke- $k$
- $x_i$  adalah titik data di kluster ke- $k$
- $c_k$  adalah *centroid* kluster ke- $k$

## 2.3. *Fuzzy C-Means Clustering*

*Fuzzy C-Means* (FCM) adalah metode klusterisasi yang lebih fleksibel karena memungkinkan data untuk menjadi anggota dari beberapa kluster dengan derajat keanggotaan tertentu (keanggotaan *fuzzy*) [12]. Berikut adalah rumus dasar dalam FCM:

Langkah-langkah *Fuzzy C-Means*:

1. Inisialisasi: Tentukan jumlah kluster  $c$  dan buat matriks keanggotaan *fuzzy*  $U = [u_{ij}]$ , di mana setiap  $u_{ij}$  adalah keanggotaan data  $x_i$  untuk kluster  $C_j$ . Nilai awal  $u_{ij}$  diambil secara acak dan memenuhi persamaan berikut:

$$\sum_{j=1}^c u_{ij} = 1 \text{ untuk semua } i$$

2. *Update centroid*: hitung posisi *centroid* untuk setiap kluster dengan mempertimbangkan keanggotaan *fuzzy*  $u_{ij}$ . *Centroid*  $u_{ij}$  dihitung dengan rumus:

$$c_j = \frac{\sum_{i=1}^n u_{ij}^m x_i}{\sum_{i=1}^n u_{ij}^m}$$

Di mana:

- $m$  adalah parameter *fuzziness* (biasanya  $m > 1$ ),
- $u_{ij}$  adalah keanggotaan  $x_i$  data pada kluster  $C_j$ ,
- $x_i$  adalah data titik ke- $i$
- $n$  adalah jumlah titik data,
- $c_j$  adalah *centroid* kluster ke- $j$

3. *Update keanggotaan*: hitung ulang keanggotaan *fuzzy*  $u_{ij}$  untuk setiap data  $u_{ij}$  dan kluster  $u_{ij}$  menggunakan rumus berikut:

$$U_{ij} = \frac{1}{\sum_{k=1}^c \left( \frac{\|x_i - C_j\|}{\|x_i - C_k\|} \right)^{\frac{2}{m-1}}}$$

Di mana:

- $x_i$  adalah data titik ke- $i$
- $c_j$  adalah *centroid* kluster ke- $j$
- $m$  adalah parameter *fuzziness* (biasanya  $m > 1$ ),
- $c_k$  adalah *centroid* kluster ke- $k$
- $u_{ij}$  adalah keanggotaan  $x_i$  data pada kluster  $C_j$ ,

4. Ulangi: langkah 2 dan 3 diulang sampai konvergensi tercapai, yaitu keanggotaan dan *centroid* stabil.

Tujuan optimasi:

Tujuan dari *Fuzzy C-Means* adalah meminimalkan fungsi objektif yang disebut *sum of weighted squared errors* (SSE) [21]:

$$J = \sum_{i=1}^n \sum_{j=1}^c U_{ij}^m \|x_{ij} - C_j\|^2$$

Di mana:

- $J$  adalah total error
- $u_{ij}$  adalah keanggotaan titik data  $x_i$  pada kluster  $C_j$
- $c_j$  adalah *centroid* kluster ke- $j$ ,
- $m$  adalah parameter *fuzziness*.

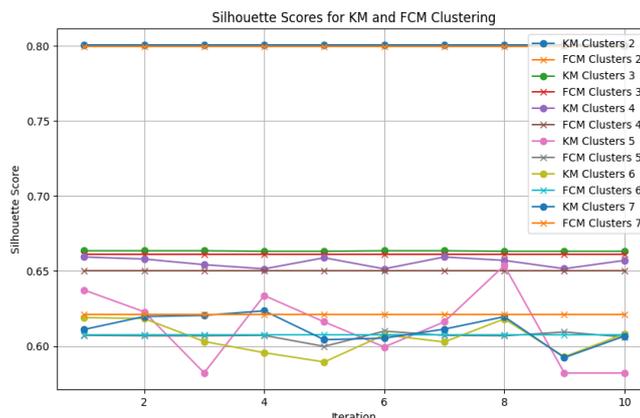
#### 2.4. Perbedaan antara *K-Means* dan *Fuzzy C-Means Clustering*

Perbedaan algoritma *K-Means* dan *Fuzzy C-Means* adalah *K-Means* mengelompokkan data ke dalam kluster-kluster yang tidak tumpang tindih (*hard clustering*), di mana setiap titik data hanya bisa menjadi anggota dari satu kluster, sedangkan *Fuzzy C-Means* memungkinkan data menjadi anggota dari lebih satu kluster dengan keanggotaan berbobot (*soft clustering*), yang memberi fleksibilitas lebih dalam menangani data yang memiliki ketidakpastian atau ambiguitas. Baik *K-Means* maupun

*Fuzzy C-Means* digunakan dalam klasterisasi, namun *Fuzzy C-Means* lebih fleksibel karena memungkinkan data untuk memiliki keanggotaan lebih dari satu klaster[22].

### 3. HASIL DAN PEMBAHASAN

Hasil klastering dari kedua algoritma dievaluasi menggunakan *Silhouette Score* yang merupakan ukuran yang menilai seberapa baik setiap titik data berada dalam klasternya. Nilai *silhouette score* berkisar antara -1 hingga +1, di mana nilai yang lebih tinggi menunjukkan bahwa titik data lebih cocok dengan klaster yang ada. Hasil ini akan digunakan untuk membandingkan performa kedua algoritma pada jumlah klaster yang berbeda [23].



Gambar 1 *Silhouette Score* Perbandingan Algoritma

Berikut adalah hasil rata-rata dari 10x eksperimen tiap klaster

Tabel 1 Hasil rata-rata tiap cluster

Algoritma	Jumlah Klaster	<i>Silhouette Score</i> (Rata-rata)
<i>K-Means</i>	2	0.800552
	3	0.663337
	4	0.655858
	5	0.612598
	6	0.605529
	7	0.611466
	<i>Fuzzy C-Means</i>	2
3		0.661171
4		0.649985
5		0.606766
6		0.607451
7		0.620929

Perbandingan tersebut mengungkapkan bahwa *Fuzzy C-Means* lebih unggul dalam mengatasi ketidakpastian dalam pengelompokan data pendidikan di Jakarta. *K-Means* dapat memberikan hasil yang baik pada jumlah klaster 2 dengan nilai rata-rata 0.800552 dibanding FCM dengan nilai rata-rata 0.799416, namun FCM memberikan fleksibilitas yang lebih tinggi dalam hal keanggotaan klaster dengan rentang 0.60766 hingga 0.799416. Dalam konteks pendidikan, di mana batas antara klaster-klaster seperti tingkat kualitas pendidikan mungkin tidak jelas, FCM memberikan keunggulan [24].

Namun, *K-Means* masih menjadi pilihan yang sangat baik untuk klastering data pendidikan yang memiliki distribusi yang jelas dan tidak tumpang tindih khususnya pada jumlah klaster 2. Dalam

kasus Jakarta, di mana ada banyak variasi dalam kualitas pendidikan antar kecamatan, penggunaan FCM bisa menjadi pilihan yang lebih bijak untuk mendapatkan kluster yang lebih representatif [25].

#### 4. KESIMPULAN

Dari percobaan yang dilakukan, *Fuzzy C-Means* (FCM) secara keseluruhan memberikan hasil yang lebih baik dibandingkan *K-Means* dalam klustering data pendidikan di tingkat kecamatan Jakarta, berdasarkan nilai rata-rata *Silhouette Score* yang lebih stabil dengan rentang 0.06766 hingga 0.799416. FCM lebih mampu menangani ketidakpastian dalam pengelompokan data yang memiliki karakteristik tumpang tindih. Oleh karena itu, FCM dapat menjadi pilihan yang lebih tepat untuk analisis kluster pendidikan, terutama ketika data memiliki variasi yang kompleks.

Namun, meskipun FCM menunjukkan performa yang lebih baik, *K-Means* tetap merupakan algoritma yang kuat untuk aplikasi klustering, terutama pada jumlah kluster 2 dengan nilai rata-rata 0.800552 dibanding FCM dengan nilai rata-rata 0.799416, dalam kasus-kasus yang lebih sederhana atau ketika data memiliki batasan kluster yang lebih jelas.

#### UCAPAN TERIMA KASIH

Tiada kata-kata yang tepat selain mengucapkan terima kasih kepada Badan Pusat Statistik yang telah menyediakan data yang sangat berharga untuk penelitian ini. Selain itu, terima kasih kepada Universitas Tarumanagara khususnya program studi Teknik Informatika, atas dukungan fasilitas beserta dosen pengampu mata kuliah Machine Learning atas bimbingan selama perkuliahan. Penulis juga menghargai kontribusi dari rekan-rekan yang mau memberikan saran atau masukan dalam pelaksanaan penelitian ini.

#### DAFTAR PUSTAKA

- [1] J. Raya Palka NoKm *et al.*, “Analisis Peran Pendidikan Terhadap Kualitas Sumber Daya Manusia Guna Meningkatkan Produktivitas Masyarakat Di DKI Jakarta Deris Desmawan Universitas Sultan Ageng Tirtayasa,” *Jumek: Jurnal Manajemen dan Ekonomi Kreatif*, vol. 1, no. 2, 2023, [Online]. Available: <https://jakarta.bps.go.id/>
- [2] A. P. Yandika, “Hubungan Antara Tingkat Pendidikan dan Tingkat Pengetahuan Masyarakat Terhadap Kejadian Demam Berdarah,” *Jurnal Ilmiah Indonesia*, vol. 2022, no. 10, pp. 874–882, doi: 10.36418/cerdika.v2i10.458.
- [3] L. Luk, A. Mufida, and M. S. Nasir, “Analisis Dinamis Tingkat Pengangguran di Indonesia,” 2021. [Online]. Available: <https://economics.pubmedia.id/index.php/jmsd>
- [4] H. Sasono and Herlina, “Kolaboratif Perguruan Tinggi Dalam Pengembangan Umkm Di Kecamatan Pancoran, Jakarta Selatan”.
- [5] A. Anita And S. I. Astuti, “Digitalisasi Dan Ketimpangan Pendidikan: Studi Kasus Terhadap Guru Sekolah Dasar Di Kecamatan Baraka,” *Jurnal Pendidikan dan Kebudayaan*, vol. 7, no. 1, pp. 1–12, Jun. 2022, doi: 10.24832/jpnk.v7i1.2509.
- [6] P. A. Rizkianti, M. Asbari, N. P. Priambudi, S. Alhani, and J. Asri, “Pendidikan Indonesia Masih Buruk?,” *Journal Of Information Systems And Management*, vol. 03, no. 02, 2024, doi: 10.4444/jisma.v2i6.613.
- [7] A. S. Manurung, A. Halim, and A. Rosyid, “Pengaruh Kemampuan Berpikir Kreatif untuk meningkatkan Hasil Belajar Matematika di Sekolah Dasar,” *Jurnal Basicedu*, vol. 4, no. 4, pp. 1274–1290, Oct. 2020, doi: 10.31004/basicedu.v4i4.544.
- [8] D. Krasnov, D. Davis, K. Malott, Y. Chen, X. Shi, and A. Wong, “Fuzzy C-Means Clustering: A Review of Applications in Breast Cancer Detection,” Jul. 01, 2023, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/e25071021.
- [9] J. C. Bezdek, “FCM: The Fuzzy C-Means Clustering Algorithm,” 1984

- [10] A. Yudhistira and R. Andika, "Pengelompokan Data Nilai Siswa Menggunakan Metode K-Means Clustering," *Journal of Artificial Intelligence and Technology Information (JAITI)*, vol. 1, no. 1, pp. 20–28, Feb. 2023, doi: 10.58602/jaiti.v1i1.22.
- [11] I. Nozomi, "Penerapan Data Mining Untuk Peringatan Dini Banjir Menggunakan Metode Klastering K-Means (Studi Kasus Kota Padang)," Bulan Juni, 2023
- [12] M. Nawaz, R. Qureshi, M. A. Teevno, and A. R. Shahid, "Object detection and segmentation by composition of fast fuzzy C-mean clustering-based maps," *J Ambient Intell Humaniz Comput*, vol. 14, no. 6, pp. 7173–7188, Jun. 2023, doi: 10.1007/s12652-021-03570-6.
- [13] R. J. Woodman and A. A. Mangoni, "A comprehensive review of machine learning algorithms and their application in geriatric medicine: present and future," Nov. 01, 2023, *Springer Science and Business Media Deutschland GmbH*. doi: 10.1007/s40520-023-02552-2.
- [14] K. P. Sinaga and M. S. Yang, "Unsupervised K-means clustering algorithm," *IEEE Access*, vol. 8, pp. 80716–80727, 2020, doi: 10.1109/ACCESS.2020.2988796.
- [15] A. Jalal, A. Ahmed, A. A. Rafique, and K. Kim, "Scene Semantic Recognition Based on Modified Fuzzy C-Mean and Maximum Entropy Using Object-to-Object Relations," *IEEE Access*, vol. 9, pp. 27758–27772, 2021, doi: 10.1109/ACCESS.2021.3058986.
- [16] A. Abdelhafeez and H. K. Mohamed, "Skin cancer detection using neutrosophic c-means and fuzzy c-means clustering algorithms," *Journal of Intelligent Systems and Internet of Things*, vol. 8, no. 1, pp. 33–42, 2023, doi: 10.54216/JISIoT.080103.
- [17] S. M. Javidan, A. Banakar, K. A. Vakilian, and Y. Ampatzidis, "Diagnosis of grape leaf diseases using automatic K-means clustering and machine learning," *Smart Agricultural Technology*, vol. 3, Feb. 2023, doi: 10.1016/j.atech.2022.100081.
- [18] K. E. Setiawan, A. Kurniawan, A. Chowanda, and D. Suhartono, "Clustering models for hospitals in Jakarta using fuzzy c-means and k-means," in *Procedia Computer Science*, Elsevier B.V., 2022, pp. 356–363. doi: 10.1016/j.procs.2022.12.146.
- [19] D. Das, P. Kayal, and M. Maiti, "A K-means clustering model for analyzing the Bitcoin extreme value returns," *Decision Analytics Journal*, vol. 6, Mar. 2023, doi: 10.1016/j.dajour.2022.100152.
- [20] "Springer Theses Recognizing Outstanding Ph.D. Research." [Online]. Available: <http://www.springer.com/series/8790>
- [21] C. L. Chowdhary, M. Mittal, P. Kumaresan, P. A. Pattanaik, and Z. Marszalek, "An efficient segmentation and classification system in medical images using intuitionist possibilistic fuzzy C-mean clustering and fuzzy SVM algorithm," *Sensors (Switzerland)*, vol. 20, no. 14, pp. 1–20, Jul. 2020, doi: 10.3390/s20143903.
- [22] M. Shutaywi and N. N. Kachouie, "Silhouette analysis for performance evaluation in machine learning with applications to clustering," *Entropy*, vol. 23, no. 6, Jun. 2021, doi: 10.3390/e23060759.
- [23] A. C. Müller and S. Guido, "Introduction to Machine Learning with Python A GUIDE FOR DATA SCIENTISTS Introduction to Machine Learning with Python."
- [24] M. Jordan, J. Kleinberg, and B. Schölkopf, "Pattern Recognition and Machine Learning."
- [25] A. Smola and S. V. N. Vishwanathan, *Introduction to Machine Learning*. Cambridge University Press, 2008.