BAB II

TINJAUAN PUSTAKA DAN DASAR TEORI

2.1. Tinjauan Pustaka

Dalam beberapa kasus tertentu, Auditor Internal membutuhkan penggunaan alat pendukung audit (*toolkit*) dalam mengevaluasi sebuah objek audit guna memastikan kesesuaian dengan standar, mempercepat proses audit serta pengambilan keputusan. Alat pendukung audit dimaksud merupakan penerapan berbagai metode dan teknologi informasi yang digunakan oleh organisasi audit untuk memfasilitasi pelaksanaan audit yang lebih efisien dan efektif (Dowling dan Leech, 2007). Alat pendukung audit dapat berupa *toolkit* kertas kerja elektonik yang dapat menghasilkan informasi bagi auditor dalam melakukan pemeriksaan.

Toolkit merujuk pada kumpulan alat atau perangkat yang digunakan untuk membangun, mengelola, atau mengoperasikan sistem informasi yang mencakup berbagai jenis perangkat lunak, perpustakaan, *framework*, atau bahkan perangkat keras yang digunakan untuk mengembangkan atau menjalankan sistem informasi. *Tools* (alat bantu) merupakan perangkat yang dapat digunakan untuk membantu dalam menerapkan teknik pelaksaan pengembangan sistem. *Tools* harus dapat mendukung (*support*) metode dan teknik yang digunakan dalam proses pengembangan sistem informasi (Repository Udinus, Ilmu Komputer, 2016).

Untuk memperjelas ruang lingkup penelitian ini, maka audit yang menjadi topik dalam penelitian adalah Audit Internal. Audit internal merupakan suatu aktivitas independen guna memberikan jaminan keyakinan serta konsultasi yang dirancang untuk memberikan suatu nilai tambah serta meningkatkan kegiatan operasi organisasi. *Audit internal* membantu organisasi dalam usaha mencapai tujuannya dengan suatu pendekatan disiplin yang sistematis untuk mengevaluasi dan meningkatkan keefektifan manajemen resiko, pengendalian dan proses pengaturan dan pengelolaan organisasi (The Institute of Internal Auditors, 2016).

Audit merupakan proses yang sistematis untuk mengevaluasi bukti secara objektif mengenai penyataan-pernyataan tentang kegiatan dan kejadian ekonomi dengan tujuan untuk menetapkan tingkat kesesuaian antara pernyataan-pernyataan

tersebut dengan kriteria yang telah ditetapkan serta penyampaian hasilnya kepada pemakai yang berkepentingan. Auditor secara profesional bertanggung jawab merencanakan dan melaksanakan audit untuk mancapai tujuan audit yang ditetapkan (The Institute of Internal Auditors, 2016).

Definisi kecurangan menurut Tuanakotta (2013:28) adalah "setiap tindakan ilegal yang ditandai dengan penipuan, penyembunyian, atau ancaman kepercayaan. Tindakan ini tidak tergantung pada penerapan ancaman kekerasan atau kekuatan fisik. Kecurangan dilakukan oleh individu, dan organisasi untuk mendapatkan uang, properti, atau layanan untuk menghindari pembayaran dengan maksud untuk mengamankan keuntungan bisnis pribadi" (Fahmi dan Syahputra, 2019). Dalam pengertian lain, definisi kecurangan sebagai "penyimpangan kebenaran yang disengaja untuk mempengaruhi pihak lain dengan sesuatu yang bernilai". Kecurangan menjadi semakin kompleks dan sulit dideteksi. Dalam organisasi, ada banyak area di mana kecurangan berpotensi terjadi, termasuk pada area sumber daya manusia, keuangan, dan manajemen rantai pasok (Sekar, 2022).

Mendeteksi kecurangan yang tersembunyi dalam laporan keuangan merupakan tugas yang sulit, dibutuhkan lebih dari sekedar penggunaan prosedur audit standar. Oleh karena itu, auditor memerlukan bantuan alat dan teknik yang efektif untuk menyederhanakan tugas audit guna membantu mendeteksi kemungkinan adanya laporan keuangan yang curang. Auditor internal dapat melakukan audit untuk mendeteksi kecurangan, tetapi seringkali terbatas pada pengujian berdasarkan teknik sampling atau penyaringan sederhana. Misalnya, kata kunci tertentu dicari dan disaring dalam transaksi pembayaran untuk menentukan apakah suatu pembayaran merupakan sebuah transaksi kecurangan (Sekar, 2022).

Hukum Benford adalah sebuah hukum yang dapat memperkirakan frekuensi kemunculan sebuah angka dalam serangkaian data numerik. Jika data numerik tersebut dihasilkan tanpa ada unsur kesengajaan, maka frekuensi kemunculan angka tersebut akan sesuai dengan frekuensi harapan dalam Hukum Benford. Hal ini juga berarti jika ada unsur kesengajaan oleh manusia untuk menciptakan sebuah kombinasi angka dan dimasukkan dalam dataset tersebut maka hasil analisis Hukum Benford akan menunjukkan bahwa ada angka tertentu yang lebih banyak

muncul dari pada yang diperkirakan (Parlin, 2009). Frank Benford mengumpulkan data numerik untuk mendukung pengamatan bahwa meskipun nomor dipilih secara acak, angka mereka mengikuti distribusi probabilitas tertentu yang tidak cukup sesuai dengan intuisi manusia (secara intuitif angka harus memiliki probabilitas kejadian yang sama) (Bhattacharya, dkk., 2011).

Hukum Benford menjadi pilihan sebagai alat yang efektif untuk menargetkan area potensial yang menjadi perhatian dalam data akuntansi. Meskipun pendekatan ini bukanlah satu-satunya alat yang dapat digunakan untuk audit akuntansi, namun penerapan Hukum Benford dapat digunakan sebagai alat pelengkap untuk membantu pekerjaan audit. (Bhattacharya dkk., 2011).

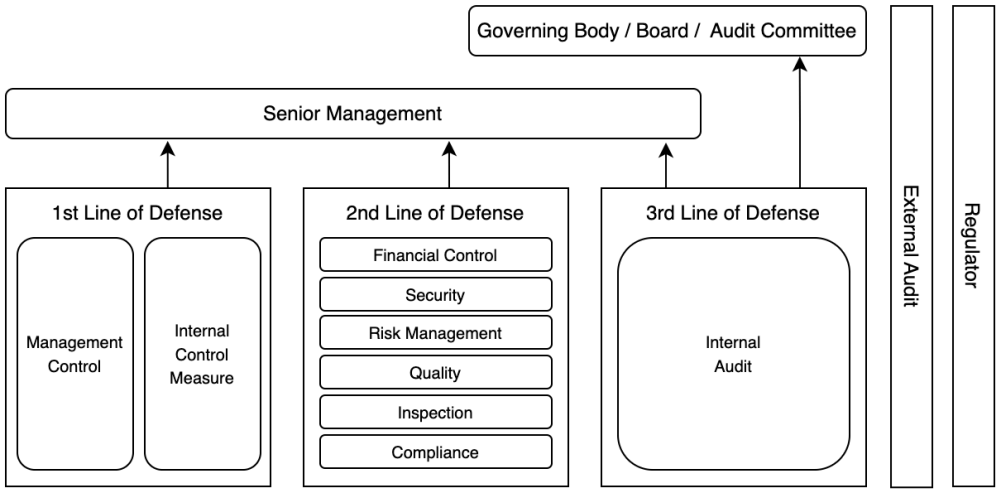
Karena landasan teori yang dikemukakan oleh Benford, akademisi telah menganggap analisis data sebagai sumber informasi utama dalam kegiatan audit (Pizzi dkk., 2021). Hukum Benford dianggap sebagai alat yang efektif dan sederhana di tangan auditor untuk mendeteksi kecurangan (Durtschi, 2004). Fakta bahwa sekitar 150 artikel tentang Hukum Benford ini diterbitkan selama lima dekade terakhir menunjukkan penerapannya pada kasus nyata. (Asllani dan Naco, 2014). Nigrini dan Mittermaier (1997) menggambarkan logika penggunaan Hukum Benford untuk membantu menemukan pola yang tidak biasa dalam aktivitas transaksi akuntansi. Sangat mungkin bahwa seseorang yang membuat entri kecurangan akan memasukkan jumlah yang sama atau jumlah yang serupa berkali-kali. Dalam kasus tersebut, variasi digit pertama yang dihasilkan dari distribusi probabilitas Hukum Benford dapat mengarahkan auditor untuk menemukan transaksi yang curang (Andari dan Ismatullah, 2019).

Sementara di sisi lain, teknik pendeteksian anomali menggunakan *machine learning* juga dapat diimplementasikan oleh auditor ke dalam perangkat teknologi pendukung audit untuk memperoleh wawasan tambahan tentang data dalam pelaksanaan audit. Pendekatan yang digunakan merupakan penerapan model *fraud and anomaly detection* menggunakan algoritma *K-Means*. *K-Means* menerapkan prinsip *clustering* yaitu membagi objek ke dalam kelompok yang berbeda, atau lebih tepatnya, mem-partisi dari kumpulan data menjadi subset (*cluster*), sehingga

data di setiap subset (idealnya) memiliki kemiripan sifat menurut beberapa ukuran jarak yang ditentukan (Sekar, 2022).

K-Means Clustering merupakan algoritma yang membutuhkan parameter input sebanyak *k* dan membagi sekumpulan *n* objek kedalam *k* klaster sehingga tingkat kemiripan antar anggota dalam satu klaster tinggi sedangkan tingkat kemiripan dengan anggota pada klaster lain sangat rendah. Tujuan dari *K-Means* adalah unuk meminimalisir total dari jarak elemen-elemen antar klaster (jarak antara suatu elemen dalam sebuah kluster dengan nilai *centroid* kluster tersebut). Proses *clustering* bertujuan untuk meminimalkan terjadinya *objective function* yang diset dalam proses *clustering* yang pada umumnya memiliki karakteristik yang sama dikelompokkan dalam satu klaster yang sama dan data yang memiliki karakteristik berbeda dikelompokkan ke dalam kelompok lain (Sani, 2013).

Paradigma Maris Sekar berpandangan bahwa teknik pengelompokan data *K-Means* perlu bekerja dan diintegrasikan dalam model "*Three Lines*" agar dapat dimanfaatkan untuk mengelola risiko direksi secara optimal. Model Tiga Garis Pertahanan (*Three Lines of Defense*) adalah konsep yang berkembang dalam internal audit guna memperoleh hasil pengendalian yang efektif. Model ini diperkenalkan pada tahun 2013 oleh *The Institute of Internal Auditors (IIA)* untuk memperjelas peran dan tanggung jawab dalam mengelola dan mengendalikan risiko secara efektif. Tiga garis pertahanan didefinisikan pada Gambar 2.1. berikut:



Gambar 2.1. *The Three Lines of Defense Model*

Sumber: *Machine Learning for Auditors*, (Sekar, 2022, 5-6)

- **Garis Pertahanan ke-1:** entitas organisasi yang berperan dalam menilai, mengendalikan, dan memitigasi risiko dalam kerangka sistem pengendalian internal untuk dilaporkan kepada manajemen puncak.
- **Garis Pertahanan ke-2:** terdiri dari kontrol keuangan, keamanan (fisik dan informasi), manajemen risiko, kualitas, inspeksi, dan fungsi kepatuhan, beserta perannya dalam memastikan risiko dikelola dan dimitigasi secara efektif. Fungsi-fungsi ini juga dilaporkan kepada manajemen puncak.
- **Garis Pertahanan ke-3:** Auditor internal bertindak sebagai garis ke-3 dari pertahanan karena auditor secara independen menilai risiko organisasi, baik internal maupun eksternal, dan memastikan risiko dikelola secara efektif. Tim audit internal bekerja dengan manajemen puncak untuk memitigasi risiko dan melaporkan kepada komite audit/badan pengurus/dewan direksi.

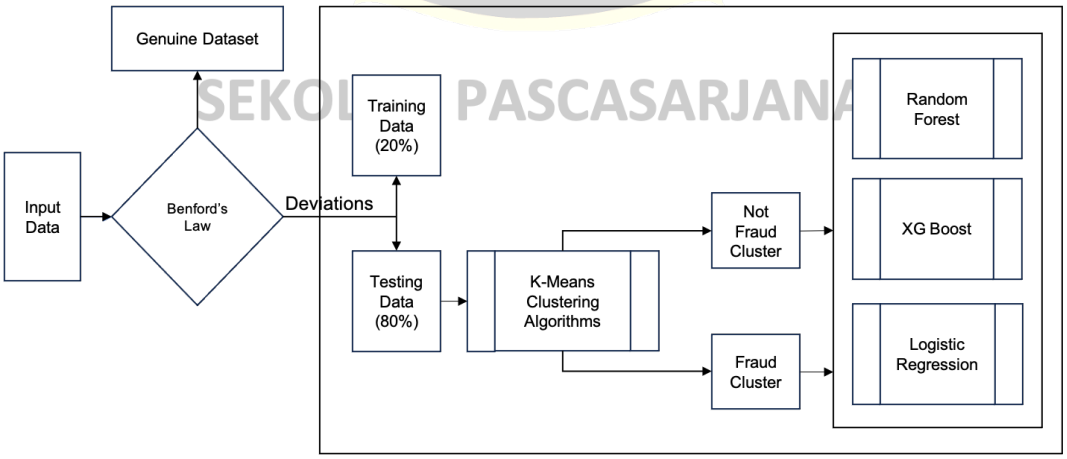
Pada gambar 2.1. dapat diartikan bahwa ketika sebuah objek audit telah melewati garis pertahanan ke-3, maka organisasi audit internal dianggap telah memberikan penjaminan kualitas (*quality assurance*) atas pelaksanaan program dan kegiatan secara efektif dan efisien. Namun apabila dalam pemeriksaan pihak Auditor Eksternal atau Aparat Penegak Hukum (APH) masih ditemukan adanya kecurangan pada transaksi keuangan, maka hal ini berpengaruh pada peran internal audit dianggap kurang profesional dan memadai dalam memitigasi risiko terjadinya kecurangan pada transaksi keuangan.

Oleh kerenanya, pengembangan alat bantu untuk pendeteksian kecurangan menjadi hal yang strategis untuk dilakukan organisasi internal audit guna mendukung garis pertahanan ke-3. Dalam hal ini, algoritma *K-Means Clustering* memungkinkan untuk dapat diterapkan dalam pengklasifikasian data-data transaksi keuangan sekaligus penerapan Hukum Benford untuk pendeteksian kemungkinan terjadinya praktik kecurangan pada transaksi keuangan. Tantangannya adalah bagaimana memperoleh data dan bukti yang akan digunakan dalam proses audit, karena kelengkapan dan keakuratan data sangat penting untuk memastikan bahwa kualitas audit tetap terjaga. Sebuah sistematika proses perlu dikembangkan untuk

memastikan bahwa auditor memperoleh data penting yang berkualitas untuk kelengkapan pemeriksaan.

Aspek yang menarik dalam penelitian ini adalah sebagian besar alat deteksi kecurangan banyak diasumsikan bekerja pada kumpulan data yang tidak terstruktur, namun memungkinkan juga bekerja pada data yang terstruktur. Sebagai contoh data laporan transaksi keuangan organisasi atau perusahaan menerapkan prinsip-prinsip akuntansi dalam format yang sesuai standar, seperti halnya pada organisasi sektor publik yang menjalankan sistem keuangan terintegrasi yang berlaku pada seluruh instansi pemerintahan. Data transaksi keuangan yang diperoleh dari sistem aplikasi keuangan dan perbendaharaan menjadi sumber dataset dalam penelitian ini. Melalui kumpulan data tersebut, diharapkan dapat dikembangkan model untuk mendeteksi anomali atau kemungkinan terjadinya kecurangan pada transaksi keuangan berdasarkan data yang dirilis dari sistem keuangan tersebut.

Jika ditinjau dari hasil penelitian sebelumnya tentang pendeteksian kecurangan (*fraud*) menggunakan Hukum Benford dalam kerangka *hybrid algorithms model* (Leena Kurien dan Chikkamannur, 2020), juga memanfaatkan algoritma *K-Means Clustering* sebagai teknik untuk memisahkan data kecurangan pada transaksi bisnis kartu kredit (*credit card*) pada platform *marketplace online* selama masa pandemi Covid-19. *Hybrid algorithms model* yang diterapkan mengacu pada prosedur sebagaimana terlihat pada Gambar 2.2. berikut:



Gambar 2.2. *Block Diagram Hybrid Model Algorithms*

Sumber: Kaithekuzhical Leena Kurien dan Chikkamannur (2020).

Implementasi *hybrid algorithms model* digunakan untuk mengetahui deviasi antara distribusi angka transaksi dengan distribusi angka teoritis Benford, jika tidak terdapat deviasi maka penerapan *hybrid algorithms model* dapat dihentikan (*exit the algorithms*), namun jika terdapat deviasi maka dataset dibagi ke dalam rasio 80:20%, dimana 80% untuk data *testing* dan 20% untuk data *training*. Algoritma *K-Means* dipergunakan untuk memisahkan transaksi yang dianggap *fraud* dan bukan *fraud* pada data *testing* dan selanjutnya di analisis menggunakan algoritma *random forest, XG Boost* dan *logistic regressions* untuk mengatahui tingkat performa dari *hybrid algorithms model*, sedangkan untuk data *training* tidak melalui pengelompokam data dan langsung di analisis menggunakan algoritma *random forest, XG Boost* dan *logistic regressions*.

Dari hasil penelitian tersebut, disimpulkan bawa penggunaan teknik klasterisasi *K-Means* dalam kerangka *hybrid algorithms model* telah memberikan hasil yang lebih baik dari penggunaan algoritma *random forest, XG Boost* dan *logistic regressions* secara terpisah atau yang lebih sederhana (Kaithekuzhical Leena Kurien da Chikkamannur, 2020). Hal ini memberikan gambaran yang cukup memadai dalam penelitian ini bahwa penggunakan algoritma *K-Means Clustering* untuk mendukung performa Hukum Benford memiliki hasil yang lebih baik dalam mendeteksi indikasi transaksi keuangan yang curang.

2.2. Dasar Teori

2.2.1. Algoritma K-Means Clustering

Algoritma *K-Means* adalah teknik yang sederhana dan efektif, tetapi memakan waktu dalam mengelompokkan sejumlah data yang besar. Dalam algoritma *K-Means* tradisional, bagian paling signifikan yang memakan waktu adalah menetapkan datum ke klaster dengan jarak terpendek untuk menghitung jarak antara datum dan semua pusat klaster. Datum adalah informasi atau keterangan yang diperoleh dari suatu pengamatan yang berupa angka, simbol atau bahasa (sifat), sedangkan kumpulan beberapa datum disebut data.

K-Means Clustering menemukan grup data dalam kumpulan data yang memiliki fitur serupa dan mendekatkannya ke *centroid* klaster terdekat.

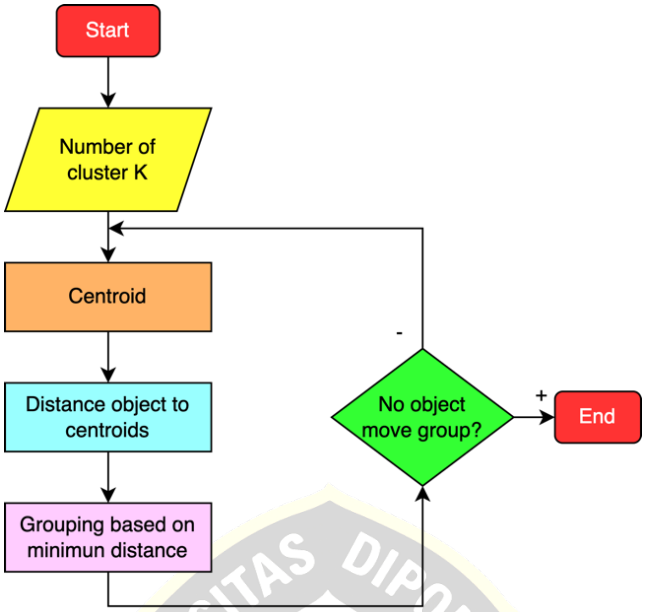
Centroid adalah rata-rata aritmatika dari semua titik dalam kumpulan data. Jumlah *centroid* ditentukan agar algoritma dapat melakukan segmentasi data. Segmentasi perilaku pelanggan, deteksi anomali, dan kategorisasi inventaris adalah beberapa kasus penggunaan *K-Means Clustering* yang umum (Sani, 2013). Algoritma *K-Means* mengelompokkan data berdasarkan kedekatan atau kemiripan data sehingga data-data yang memiliki karakteristik yang sama akan dimasukkan ke dalam klaster yang sama pula.

Prinsip kerja algoritma *K-Means* adalah menentukan jumlah *k* klaster terlebih dahulu kemudian menentukan titik *centroid* data secara random/acak. Setelah itu mengalokasikan data ke klaster terdekat dan proses tersebut akan diulang hingga menemukan *centroid* yang stabil. Prinsip dasar algoritma *K-Means* adalah sebagai berikut: (Orisa, 2022).

1. Menentukan *Centroid: Centroid* adalah pusat klaster. Tiap-tiap klaster memiliki *centroid*. Penentuan *centroid* awal dilakukan secara random.
2. Menentukan anggota *cluster:* melakukan pengelompokan data atau objek ke tiap-tiap klaster. Pengelompokan tersebut dihitung berdasarkan jarak antar data/objek ke *centroid* masing-masing klaster yang sudah ditentukan sejak awal. Data/objek yang paling dekat dengan *centroid* maka akan dikelompokkan menjadi satu *cluster* dengan *centroid* tersebut.

Setelah klaster diidentifikasi oleh algoritma *K-Means*, klaster dapat digabungkan dengan dataset asli dan diplot untuk melihat bagaimana klaster didistribusikan ke seluruh variabel. Kerangka algoritma *K-Means* dijelaskan pada Gambar 2.3.

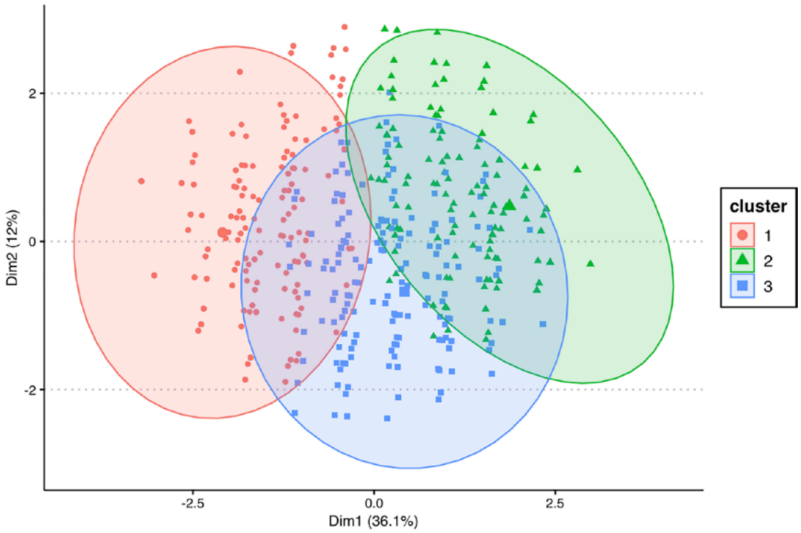
Keluaran algoritma *K-Means Clustering* pada Gambar 2.4 sangat bergantung pada penentuan jumlah klaster dan pemilihan *centroid* awal yang tepat. Permasalahan umum pada algoritma *K-Means* adalah penentuan *centroid* yang bukan merupakan pusat klaster yang sesungguhnya. Dalam prakteknya algoritma ini harus dijalankan berkali-kali dengan *centroid* awal yang berbeda-beda untuk mendapatkan *centroid* akhir yang dianggap paling baik (Orisa, 2022).



Gambar 2.3. Algoritma *K-Means Clustering*

Sumber: Repository Universitas Dian Nuswantoro (2013)

Untuk dapat lebih memahami bagaimana algoritma *K-Means Clustering* bekerja pada sekumpulan data, pada Gambar 2.4 memperlihatkan contoh grafik terdiri dari sekumpulan data yang terbagi dalam kelompok warna merah (*cluster* 1) dengan *centroid* berbentuk **noktah (dot)** berwarna merah, kelompok warna hijau (*cluster* 2) dengan *centroid* berbentuk **segitiga** berwarna hijau dan kelompok warna biru (*cluster* 3) dengan *centroid* berbentuk **segi empat** berwarna biru.



Gambar 2.4. Bentuk visualisasi *K-Means Clustering* (Sekar, 2022)

Metode *K-Means Clustering* menjalankan prosedur, sebagai berikut:

1. Pilih jumlah klaster *k* yang diinginkan.
2. Inisialisasi *k* pusat *cluster* (*centroid*) secara random/acak.
3. Tempatkan setiap data atau objek ke klaster terdekat. Kedekatan dua objek ditentukan berdasar jarak. Jarak yang dipakai pada algoritma *K-Means* diantranya menggunakan rumus *Euclidean distance* (*d*).

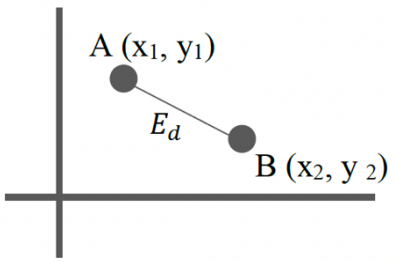
$$d_{Euclidean}(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (2.1)$$

Keterangan:

d (x,y)	= jarak antara data pada titik *x* dan *y*
x	= titik data *centroid* ($x_1, x_2, \ldots, x_n$)
y	= titik data objek ($y = y_1, y_2, \ldots, y_n$)
n	= jumlah atribut data
i=1	= merupakan indeks perhitungan ke-1 hingga ke-n

4. Hitung kembali pusat *cluster* dengan keanggotaan klaster yang sekarang.
5. Pusat klaster adalah rata-rata (*mean*) dari semua data atau objek dalam klaster tertentu.

Euclidean adalah salah satu dari banyak metode jarak yang dapat digunakan untuk mengukur tingkat kemiripan dari citra. Metode *Euclidean* adalah suatu metode pencarian kedekatan nilai jarak dari 2 buah variabel, selain mudah metode ini juga tidak memakan waktu, dan proses yang cepat. *Euclidean* adalah fungsi heuristik yang diperoleh berdasarkan jarak langsung bebas hambatan seperti untuk mendapatkan nilai dari panjang garis diagonal pada sebuah segitiga.



Keterangan:

A (x_1,y_1)	= titik x_1 dan y_1
B (x_2,y_2)	= titik x_2 dan y_2
E_d	= Jarak 2 buah objek

Gambar 2.5. Jarak *Euclidean* dua variable data (Afandi, dkk., 2020)

Rumus *Euclidean distance* (jarak *Euclidean*) digunakan untuk mengukur jarak antara dua titik dalam ruang *Euclidean*. Setiap kelompok atau *centroid* direpresentasikan oleh pusat gravitasi dari titik-titik yang ada di dalamnya. Jarak *Euclidean* digunakan untuk mengukur sejauh mana setiap titik data dari kelompok-kelompok tersebut dari *centroid* terkait.

2.2.2. Hukum Benford

Hukum Benford mendefinisikan distribusi frekuensi digit dalam kumpulan data dari posisi pertama hingga posisi keempat dimulai dari kiri. (Sugiarto dkk., 2017). Menurut Newcomb (1881), Benford (1938), (Miller, 2015), (Said dan Mohammed, 2020) dan (Wang dkk., 2022), distribusi frekuensi dari angka nol sampai sembilan juga dikenal sebagai distribusi angka, mengikuti distribusi frekuensi logaritmik yang bergantung pada posisi digit tersebut. Namun Benford menegaskan khusus untuk digit pertama yang memiliki kemungkinan tidak sama bukan angka 0 melainkan salah satu dari sembilan kemungkinan, yaitu digit 1, 2, . . ., 9. Digit 1 memiliki prosentase lebih dari 30% sedangkan digit 9 kurang dari 5% (Arno Berger dan Theodore P. Hill, 2015). Frekuensi digit pertama s.d. keempat distribusi Benford secara berurutan tertuang pada Tabel 2.1. berikut ini:

Tabel 2.1. Probabilitas Digit Pertama s.d. Empat Hukum Benford (%)

d	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Prob(D_1=*d*)	0	30,10	17,60	12,49	9,69	7,91	6,69	5,79	5,11	4,57
Prob(D_2=*d*)	11,96	11,38	10,88	10,43	10,03	9,66	9,33	0,03	8,75	8,49
Prob(D_3=*d*)	10,17	10,13	10,09	10,05	10,01	9,97	9,94	9,90	9,86	9,82
Prob(D_3=*d*)	10,01	10,01	10,00	10,00	10,00	9,99	9,99	9,99	9,98	9,98

Sumber: *An Introdution to Benford's Law* (Arno Berger dan Theodore P. Hill, 2)

Berdasarkan pernyataan hukum Benford (*statement of Benford's Law*) yang dikemukakan oleh (Jasak dan Banjanović-Mehmedovic, 2008) menyebutkan bahwa himpunan bilangan yang memenuhi Hukum Benford untuk digit pertama dari *d* (urutan angka) 1-9 dapat dihitung dengan menggunakan persamaan:

$$\text{Prob}(D_1 = d) = \log_{10}\left(1 + \frac{1}{d}\right) \qquad for\ all\ d = 1,2,\ldots,9; \qquad (2.2)$$

Dimana:
Prob = probabilitas atau kemungkinan.
$D_1 = d$ = urutan angka atau digit ke-1

$Prob(D_1 = 1) = \log_{10} 2 = 0{,}3010$
$Prob(D_1 = 2) = \log_{10} \frac{3}{2} = 0{,}1760$
$Prob(D_1 = 3) = \log_{10} \frac{4}{3} = 0{,}1249$
$Prob(D_1 = 4) = \log_{10} \frac{5}{4} = 0{,}0969$
$Prob(D_1 = 5) = \log_{10} \frac{6}{5} = 0{,}0791$
$Prob(D_1 = 6) = \log_{10} \frac{7}{6} = 0{,}0669$
$Prob(D_1 = 7) = \log_{10} \frac{8}{7} = 0{,}0579$
$Prob(D_1 = 8) = \log_{10} \frac{9}{8} = 0{,}0511$
$Prob(D_1 = 9) = \log_{10} \frac{10}{9} = 0{,}0457$

Frekuensi kemunculan angka 1-9 dalam hukum Benford tersebut menjadi frekuensi harapan (*likelihood*) dalam suatu pengamatan, sebaliknya, frekuensi kemunculan nyata dari angka 1-9 dalam data transaksi keuangan menjadi frekuensi aktual. (Nigrini, 2011) menyebutkan bahwa proses utama di dalam analisis Hukum Benford adalah pemisahan digit-digit tertentu 1-9 dan menghitung frekuensinya. Pemisahan digit ini dilakukan menggunakan *excel worksheet* sebagai alat bantu.

Hukum Benford memerlukan pengujian kesesuaian untuk mengindikasikan ketidaknormalan data dengan membandingkan frekuensi harapan Hukum Benford dengan frekuensi aktual. Namun, harus diingat bahwa kesesuaian dengan frekuensi Hukum Benford tidak selalu mengindikasikan bahwa suatu data terbebas dari manipulasi, justru sebaliknya, kesesuaian frekuensi angka 1-9 dengan Hukum Benford selayaknya membuat kita lebih berhati-hati dalam memaknai data transaksi keuangan dalam sebuah penugasan audit. (Prasetyo dan Djufri, 2020).

Bagi data yang jumlahnya relatif sedikit, biasanya kesesuaian dengan frekuensi Hukum Benford dapat diestimasi secara visual menggunakan histogram. Namun, bagi data yang jumlahnya besar, hal ini relatif lebih kompleks karena hasilnya cenderung kurang akurat. Oleh karena itu, (Nigrini, 2011) menyatakan bahwa analisis secara visual sebaiknya dilengkapi dengan beberapa uji statistik, seperti uji z, *chi square*, atau *mean absolute deviation* (MAD). Berikut ini adalah

penjelasan mengenai beberapa parameter uji kesesuaian dalam perhitungan Hukum Benford (Said & Mohammed, 2020; Setyawan, 2020), antara lain sebagai berikut:

a. Uji *z Statistic*: merupakan tes kebenaran untuk menguji apakah frekuensi aktual untuk suatu digit berbeda secara signifikan dari frekuensi teoritis Hukum Benford. Adapun nilai Z dihitung dengan menggunakan persamaan:

$$z = \frac{|AP - EP| - (\frac{1}{2N})}{\sqrt{\frac{EP(1-EP)}{N}}} \quad (2.3)$$

Di mana:

AP = nilai probabilitas aktual kemunculan digit tertentu.
EP = nilai probabilitas yang diharapkan sesuai Hukum Benford.
N = jumlah data.

Pengujian *z* digunakan untuk menentukan perbedaan antara frekuensi suatu digit tertentu dengan frekuensi Hukum Benford secara statistika, dengan tingkat signifikasi sebesar 0,05 (5%) sehingga nilai *z* yang menjadi acuan adalah sebesar 1,96 (Nigrini, 2011). Hal ini bermakna jika nilai *z* yang dikalkulasi menggunakan persamaan (2.3) memiliki nilai lebih dari 1,96 dapat disimpulkan bahwa frekuensi digit tersebut berbeda secara signifikan dengan frekuensi Hukum Benford.

b. Uji *Chi Square*: digunakan untuk membandingkan suatu set hasil penelitian dengan hasil yang diharapkan dengan tingkat signifikansi yang digunakan adalah sebesar 0,05 (5%). Jika nilai *Chi Square* lebih dari ±0,05 berarti mengandung hipotesis bahwa sampel tidak sesuai atau menyimpang dari distribusi angka Hukum Benford. Nilai *Chi square* dihitung dengan menggunakan persamaan:

$$chi\sim square = \sum_{i=1}^{k} \frac{(AC-EC)^2}{EC} \quad (2.4)$$

Dimana:

AC = nilai perhitungan aktual kemunculan digit tertentu.
EC = nilai perhitungan harapan Hukum Benford.
k = jumlah kelas data (ada 9 untuk digit pertama dan kedua, dan 19 untuk uji kombinasi digit pertama dan kedua).

c. Uji *Mean Absolut Deviation* (*MAD*) adalah teknik untuk mengevaluasi metode peramalan menggunakan jumlah dari kesalahan-kesalahan yang absolut. *MAD* mengukur ketepatan ramalan dengan merata-rata dugaan nilai absolut dari masing-masing kesalahan. Nilai *MAD* dihitung dengan menggunakan persamaan:

$$\text{Mean Absolute Deviation} = \frac{\sum_{i=1}^{k} |AP - EP|}{K} \quad (2.5)$$

Dimana:

AP = nilai probabilitas riil kemunculan digit tertentu.

EP = nilai probabilitas yang diharapkan sesuai dengan Benford Law.

K = jumlah kelas data (ada 9 untuk digit pertama dan kedua, dan 19 untuk uji kombinasi digit pertama dan kedua).

Untuk mengetahui tingkat kesesuaian Hukum Benford maka nilai *MAD* hasil perhitungan dengan menggunakan persamaan di atas dibandingkan dengan nilai standar *MAD* pada tabel berikut ini:

Tabel 2.2. Uji Kesesuaian Hukum Benford Menggunakan *MAD*

Digit	Nilai	Kesimpulan
Digit Pertama	0,000 s/d 0,006	Sangat mirip dengan Hukum Benford
	0,006 s/d 0,012	Mirip dengan Hukum Benford
	0,012 s/d 0,015	Agak mirip dengan Hukum Benford
	diatas 0,015	Tidak mirip dengan Hukum Benford
Digit Kedua	0,000 s/d 0,008	Sangat mirip dengan Hukum Benford
	0,008 s/d 0,010	Mirip dengan Hukum Benford
	0,010 s/d 0,012	Agak mirip dengan Hukum Benford
	diatas 0,012	Tidak mirip dengan Hukum Benford
Kombinasi Digit Pertama dan kedua	0,0000 s/d 0,0012	Sangat mirip dengan Hukum Benford
	0,0012 s/d 0,0018	Mirip dengan Hukum Benford
	0,0018 s/d 0,0022	Agak mirip dengan Hukum Benford
	diatas 0,0022	Tidak mirip dengan Hukum Benford

Sumber: (Mark J. Nigrini, 2011: 160)