

Penanda Aras Integriti Data Kewangan: Penemuan Corak Data Kewangan Indeks FBM KLCI Malaysia menggunakan Pendekatan Hukum Benford Pareto Diperluas (Financial Data Integrity Benchmark: Discovery of Patterns in Financial Data of Malaysia's FBM KLCI Index using an Extended Benford Pareto Law Approach)

SHAR NIZAM SHARIF, SAIFUL HAFIZAH JAAMAN* & SAIFUL IZZUAN HUSSAIN

Jabatan Sains Matematik, Fakulti Sains dan Teknologi, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia

Diserahkan: 29 Mei 2024/Diterima: 26 November 2024

ABSTRAK

Integriti data adalah penting dalam konteks akademik dan praktikal. Hukum Benford dengan prinsip Pareto menilai ketulenan data secara berkesan dengan memodelkan taburan digit pelopor signifikan. Hukum Benford menjangkakan corak taburan logaritma merentas set data yang pelbagai. Penyelidikan ini bertujuan untuk memperluas Hukum Benford Pareto dengan mengoptimumkan parameter bentuk taburan menggunakan kaedah simpleks, meningkatkan kebolehgunaannya sebagai alat forensik untuk mengesan manipulasi data dengan menganalisis penyelewengan daripada taburan digit yang dijangkakan. Kajian memanfaatkan Hukum Benford Pareto Diperluas pada data indeks FBM KLCI Malaysia menerusi metodologi penyelidikan dwi fasa. Pada mulanya, model dilatih menggunakan data dari 2010 hingga 2020 untuk menentukan taburan jangkaan bagi digit pelopor signifikan. Selepas itu, keberkesanan model diuji dengan set data baharu pada tahun 2020, 2021 dan 2022. Pengesahan model melibatkan ujian keakuratan sisihan min mutlak dan khi kuasa dua untuk menilai keakuratan kepada prinsip Hukum Benford dan mengesan anomali. Keputusan mengesahkan bahawa walaupun set data latihan akur kepada Hukum Benford Pareto Diperluas, sisihan min mutlak mengesan sisihan ketara dalam set data uji untuk tahun 2020 dan 2022 mencadangkan potensi manipulasi. Walaupun kajian kes ini memberi tumpuan kepada pasaran saham Malaysia, algoritma yang dibangunkan mempunyai potensi untuk aplikasi yang universal dalam pendekatan analisis forensik data di peringkat global.

Kata kunci: Digit pelopor signifikan; hukum Benford Pareto; pengoptimuman simpleks

ABSTRACT

Data integrity is crucial in academic and practical contexts. Benford's Law, rooted in the Pareto principle, effectively assesses data authenticity by modeling the distribution of significant leading digits. Benford's Law anticipates a logarithmic distribution pattern across diverse datasets. This study aims to extend the Benford Pareto Law by optimizing distribution shape parameters using the simplex method, enhancing its applicability as a forensic tool for detecting data manipulation by analyzing deviations from expected digit distributions. This study applied the Extended Benford Pareto Law to the FBM KLCI Malaysia index data, employing a dual-phase research methodology. Initially, the model was trained using data from 2010 to 2020 to determine the expected distribution of significant leading digits. Subsequently, the model's effectiveness was tested with new data sets in 2020, 2021, and 2022. The model evaluation involved absolute minimum deviation and chi-square tests to assess conformity to Benford's Law principle and detect anomalies. Results confirmed that while the training dataset conformed to Extended Pareto Benford's Law, the minimum absolute deviation test detected notable deviations in the test datasets for 2020 and 2022 suggesting potential manipulations. Although this case study focuses on the Malaysian stock market, the developed algorithm holds global potential for universal application in data forensic analysis approaches.

Keywords: Pareto Benford's Law; significant leading digit; simplex optimization

PENGENALAN

Dalam beberapa tahun kebelakangan ini, analisis corak taburan digit dalam set data kewangan telah muncul sebagai bidang penyelidikan yang kian mendapat perhatian dengan implikasi merangkumi pelbagai domain seperti

pengauditan, pengesanan pemalsuan dan penilaian integriti data. Teras kepada pendekatan analisis corak taburan data ialah Hukum Benford yang meramalkan corak taburan digit pelopor signifikan dalam set data fenomena semula jadi dengan tepat (Benford 1938). Digit pelopor signifikan

ialah kedudukan digit bukan sifar pertama integer paling kiri dalam suatu nombor (Kossovsky 2014). Contohnya, digit pelopor signifikan bagi nombor 576.43 ialah 5. Seterusnya, digit pelopor signifikan bagi nombor 0.0376 ialah 3 dengan mengabaikan digit 0. Pelopor signifikan untuk digit tunggal seperti digit 6 ialah 6. Bagi nombor negatif pula, tatatanda negatif dibuang. Oleh itu, digit pelopor signifikan untuk -63.79 ialah 6.

Kaedah lain untuk mentakrif digit pelopor signifikan bagi sebarang nombor adalah menerusi perwakilan tatatanda saintifik sebagai $A \times 10^N$ dengan N ialah nombor bulat dan A ialah nombor nyata dengan $1 \leq |A| < 10$ (Hill 1995). Untuk perwakilan nombor tatatanda saintifik, digit pelopor signifikan bagi bahagian kamilan A mengabaikan bahagian pecahan serta tatatanda positif dan negatif. Sebagai contoh, nombor 311.752 boleh ditulis dalam tatatanda saintifik sebagai 3.1175×10^2 dengan digit 3 sebagai digit pelopor signifikan.

Hukum Benford (1938) menegaskan bahawa digit pelopor signifikan dalam banyak set data tidak bertaburan secara seragam sebaliknya mengikuti corak taburan logaritma yang ditentukan. Nilai kebarangkalian nombor yang mempunyai digit 1 sebagai digit pelopor signifikan adalah dianggarkan 30% daripada keseluruhan data berangka. Kebarangkalian digit pelopor signifikan berkurangan apabila urutan digit pelopor signifikan meningkat.

Pemerhatian terperinci digit pelopor signifikan oleh Benford (1938) melibatkan 20 set data berangka yang pelbagai. Skop pemerhatian Benford (1938) adalah meluas merangkumi sejumlah 20,229 titik data. Set data yang luas penting demi mewujudkan asas yang kukuh bagi Benford (1938) untuk menerbitkan persamaan taburan digit pelopor signifikan. Persamaan Hukum Benford (1938) diberikan menerusi Persamaan (1).

$$P(d_1) = \log\left(1 + \frac{1}{d_1}\right) \tag{1}$$

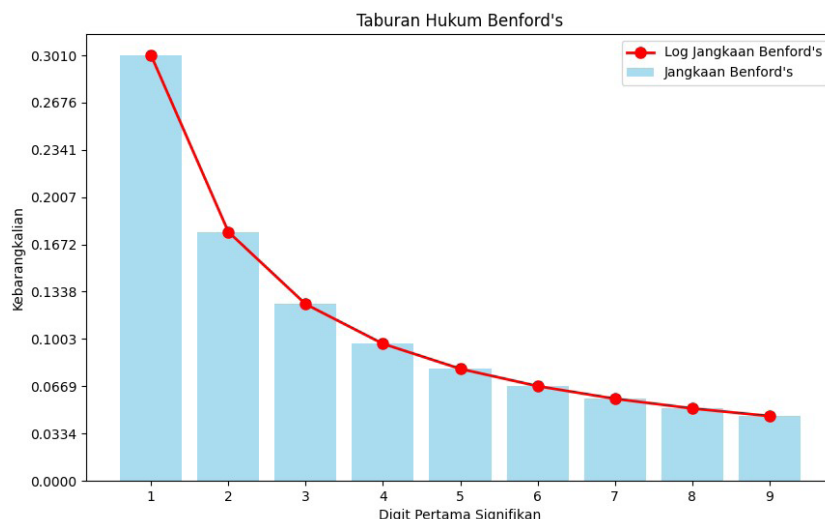
dengan P ialah kebarangkalian dan d_1 ialah digit pelopor signifikan dengan $d_1 \in \{1, 2, \dots, 9\}$. Menerusi Persamaan (1), nilai jangkaan digit pelopor signifikan Hukum Benford (1938) ditunjukkan dalam Jadual 1. Berdasarkan Jadual 1, pola taburan nilai jangkaan digit pelopor signifikan Hukum Benford (1938) ditunjukkan dalam Rajah 1.

Dalam dunia nyata, kekerapan digit pelopor signifikan dalam set data peristiwa sebenar mengikut corak taburan Rajah (1) dengan rapat, antaranya ialah nombor jujukan Fibonacci, kedalaman sungai, magnitud gempa bumi, luas atmosfera, populasi penduduk, audit cukai dan harga pasaran saham. Penyimpangan yang ketara corak taburan data peristiwa sebenar daripada corak taburan Hukum Benford menunjukkan penyelewengan data, mencadangkan manipulasi, penipuan atau ralat dalam data. Hal ini menjadikan Hukum Benford diguna pakai secara meluas dalam pelbagai bidang sebagai satu alat forensik yang berkesan untuk mengesan anomali dan isyarat penipuan serta manipulasi data (Nigrini 2012). Ley (1996) adalah pengkaji pertama yang mengkaji hubungan antara data indeks pasaran saham dengan Hukum Benford, mengamati taburan pulangan harian indeks perindustrian Dow Jones (DJIA) dari tahun 1900 hingga 1993 dan indeks S&P 500 dari tahun 1926 hingga 1993. Beliau membandingkan keputusan empirik dengan kekerapan taburan Hukum Benford menggunakan ujian keakuratan khi kuasa dua dan mendapati taburan data kedua-dua indeks tidak akur terhadap taburan Hukum Benford. Keputusan yang serupa diperoleh oleh Rodrigues (2004). Menurut Rodrigues (2004), kewujudan pengaruh manusia ke atas data merupakan punca utama taburan

JADUAL 1. Perkadaran digit pelopor signifikan Model Hukum Benford (1938)

d_1	p
1	0.3010
2	0.1761
3	0.1249
4	0.9690
5	0.7920
6	0.0669
7	0.0580
8	0.5120
9	0.0458
Total	1

Sumber: Benford (1938)



RAJAH 1. Perkadaran digit pertama Hukum Benford (1938)

data tidak akur kepada Hukum Benford, menyebabkan berlakunya penyelewengan data. Hairuddin dan Nur Firyal (2022) dalam kajian lain membuktikan data pasaran saham dipengaruhi oleh sentimen dan psikologi manusia terutamanya dalam sesuatu tempoh tertentu.

Krakar dan Žgela (2009) yang menggunakan Hukum Benford untuk mengkaji harga tutup saham dan pusing ganti saham syarikat yang disenarai dalam pasaran saham Zagreb mendapati data pusing ganti mematuhi Hukum Benford manakala data harga tutup saham tidak. Mereka mencadangkan kemungkinan wujud faktor psikologi mempengaruhi harga tutup harian saham yang memandu ke arah berlakunya penyelewengan data. Kajian oleh Corazza, Ellero dan Zorzi (2010) juga memberikan keputusan yang sama. Mereka menekankan kewujudan tingkah laku tidak rasional dalam pasaran saham. Dalam kajian lain, Zainol Abidin, Jaaman dan Bakar (2023) membuktikan tahap kepuasan pelabur terhadap sesuatu saham dipengaruhi oleh pelbagai faktor pasaran, memberi implikasi ke atas penjana portfolio.

Namun begitu, kajian oleh Shengmin dan Wenchao (2010) ke atas data pulangan harian indeks komposit pasaran saham Shanghai di China dari tahun 1990 hingga 2009 mendapati indeks komposit pasaran saham Shanghai akur terhadap Hukum Benford dan tidak menunjukkan sebarang gelagat yang boleh dikaitkan kepada berlakunya penyelewengan data. Lobão dan Pereira (2016) meneliti sembilan indeks pasaran saham Eropah untuk memperoleh sebarang petunjuk berkenaan halangan psikologi yang mungkin berlaku. Keputusan kajian mereka menunjukkan tiada penyimpangan antara corak data kajian dengan Hukum Benford, merumuskan tiada sebarang penyelewengan data berlaku terhadap pusing ganti harian, volum dagangan harian dan bilangan dagangan harian saham bagi enam indeks pasaran Eropah kecuali indeks pasaran saham Jerman, Belanda dan Finland. Keputusan

yang menunjukkan pulangan logaritma asli harian tidak menunjukkan keakuran terhadap taburan teori Hukum Benford memberi bayangan wujudnya penyelewengan data. Keputusan ini disokong oleh kajian Jayasree (2017) ke atas dagangan saham dan pulangan saham komponen Nifty Fifty bagi tempoh selama lima tahun.

Hukum Benford menawarkan cara mudah untuk mengesan penyelewengan data dengan meramalkan corak taburan digit pelopor signifikan data (Miller 2015). Sungguhpun begitu, pendekatan klasik Hukum Benford tidak teguh dalam mengesan corak taburan digit pelopor signifikan dalam pelbagai set data dunia nyata (Hürlimann 2015). Untuk mengatasi kekangan ini, Hürlimann (2015) telah mencadangkan Hukum Benford dengan prinsip Pareto yang menggabungkan dua parameter bentuk iaitu α dan β dalam memodelkan taburan digit pelopor signifikan. Namun, mereka tidak menyediakan panduan menyeluruh untuk menentukan nilai parameter bentuk α dan β , membataskan keberkesanan Hukum Benford Pareto merentas pelbagai set data dan disiplin. Oleh itu, objektif pertama kajian ini adalah untuk menentukan nilai parameter bentuk α dan β model.

Sehingga kini, model Hukum Benford Pareto yang dicadangkan oleh Hürlimann (2015) belum banyak dikaji menyebabkan keberkesanan model tidak dapat untuk ditentusahkan membataskan potensi model untuk diadaptasi ke dalam pelbagai bidang. Dalam bidang kewangan, penyelidik lepas lazim menggunakan Hukum Benford (1938) dalam mengesan isyarat pemalsuan dan manipulasi dalam pelbagai jenis data, terutamanya data indeks pasaran saham. Hukum Benford Pareto masih belum diuji terhadap pelbagai set data kewangan khususnya dalam pasaran saham. Maka, objektif kedua kajian ini adalah menentukan keberkesanan model Hukum Benford Pareto dalam konteks data kewangan yang luas menggunakan data indeks pasaran saham.

Bagi objektif pertama menentukan nilai parameter bentuk α dan β model, kajian ini membina algoritma Hukum Benford Pareto Diperluas (HBPD) untuk menetapkan nilai parameter bentuk α dan β berdasarkan pendekatan kaedah pengoptimuman simpleks yang sistematik. Pendekatan yang dicadangkan menyediakan rangka kerja yang teguh dalam menganggar parameter bentuk taburan dengan ketepatan dan kebolehppercayaan yang dipertingkatkan. Bagi objektif kedua pula, kajian menekankan kebolehgunaan model HBPD kepada set data kewangan indeks pasaran saham. Berpandu atas dasar kepentingan analisis forensik data kewangan, penyelidikan ini mengkaji kebolehgunaan HBPD menganalisis corak taburan digit pelopor signifikan indeks FBM KLCI, iaitu indeks pasaran saham Malaysia. Oleh itu, kajian ini menyediakan bukti empirik bagi menyokong keteguhan model HBPD dalam konteks analisis data kewangan.

Sumbangan utama kajian ini ialah untuk mempertingkatkan pemahaman tentang corak taburan digit pelopor signifikan dalam set data pasaran kewangan. Pembinaan algoritma HBPD berupaya mengecam corak taburan digit pelopor signifikan, meluaskan potensi Hukum Benford dengan mengecam corak taburan digit pelopor signifikan sebagai penunjuk utama kelakuan data pasaran kewangan. Menurut Khalil dan Bakar (2023), kelakuan data pasaran kewangan adalah rumit, berubah-ubah dan sukar diramal. Di samping itu, kajian ini juga menawarkan satu alat analisis forensik bagi mengesan penyimpangan data yang lebih berdaya tahan yang sesuai dengan kerumitan pasaran kewangan. Sumbangan kajian penting kerana penyimpangan daripada corak taburan digit pelopor signifikan dalam data kewangan memperlihatkan kewujudan anomali data, mencadangkan isyarat manipulasi, pemalsuan atau ralat dalam mengekalkan integriti data.

DATA DAN METODOLOGI

Kajian ini menggunakan set data digit pelopor signifikan yang diperolehi daripada data lampau kadar pulangan harga tutup harian indeks FBM KLCI Malaysia. Pemilihan digit pelopor signifikan daripada harga tutup harian indeks FBM KLCI disebabkan oleh kepentingannya sebagai indeks penanda aras utama yang mencerminkan prestasi ekonomi dan pasaran saham Malaysia (Nurul Hanis, Ruzanna & Noriszura 2020).

Kajian ini menggunakan dua set data yang berbeza iaitu set data latihan dan set data uji seperti yang digunakan oleh Buvana dan Muthumayil (2021). Set data latihan bermula dari 1 Januari 2010 hingga 31 Disember 2020. Set data uji bermula dari 1 Januari hingga 31 Disember 2020, 1 Januari hingga 31 Disember 2021 dan 1 Januari hingga 31 Disember 2022. Dengan menggunakan data indeks pasaran FBM KLCI, kajian dapat menentukan corak taburan digit pelopor signifikan data sebagai penunjuk utama kelakuan dinamik pasaran Malaysia. Ini juga bagi

memastikan penemuan kajian adalah relevan dan boleh diguna pakai dalam konteks tempatan. Walaupun indeks pasaran saham FBM KLCI merupakan fokus dalam kajian ini, pendekatan tersusun rangka kerja algoritma yang dibina menyediakan kebolehgunaan meluas di luar konteks Malaysia. Metodologi rangka kerja algoritma baharu yang disaran boleh disesuaikan pada pelbagai set data pasaran kewangan global. Di bahagian metodologi dibentangkan pembinaan algoritma sebagai satu alat matematik forensik tersusun untuk menganalisis corak taburan digit pelopor signifikan dalam data berdasarkan prinsip Hukum Benford Pareto.

Pembinaan algoritma kajian dijalankan melalui dua peringkat iaitu peringkat pertama melatih model HBPD untuk menentukan corak taburan digit pelopor signifikan menggunakan data latihan. Seterusnya, peringkat kedua menguji model HBPD menggunakan data uji yang baharu. Pada peringkat pertama, langkah awal pembinaan algoritma melibatkan pengisihan digit pelopor signifikan daripada nilai bukan sifar kadar pulangan harga tutup saham data latihan. Penentuan kadar pulangan saham diberikan dalam Persamaan (2) (Jing & Samsudin 2018).

$$R = \frac{(V_t - V_o)}{V_o} \quad (2)$$

dengan R ialah kadar pulangan; V_t ialah nilai semasa harga tutup dan nilai awal harga tutup. Selepas itu, model sedia ada Hukum Benford dengan prinsip Pareto ditunjukkan menerusi persamaan (3) sebagai model asas untuk menganalisis taburan digit pelopor signifikan dalam set data latihan (Hürlimann 2015).

$$B_p(d; \alpha, \beta) = \frac{\alpha}{\alpha + \beta} \left\{ [\log(1+d)]^\beta - [\log(d)]^\beta \right\} + \frac{\beta}{\alpha + \beta} \cdot \sum_{k=1}^{\infty} \left\{ [k + \log(d)]^{-\alpha} - [k + \log(1+d)]^{-\alpha} \right\} \quad (3)$$

dengan $d \in \{1, \dots, 9\}$ dan parameter $\alpha, \beta > 0$.

Parameter α dan β mengambil kira nilai ekstrem kedua-dua spektrum kiri dan kanan ekor taburan. Parameter α memainkan peranan penting untuk menentukan bentuk ekor kiri taburan yang mewakili digit pelopor signifikan awalan. Parameter α mengukur kadar kecerunan ekor kiri taburan dan menentukan kadar kebarangkalian kekerapan digit pelopor signifikan awalan sehingga digit median. Nilai α yang lebih rendah menunjukkan ekor kiri yang lebih tebal. Maka, nilai α yang semakin rendah mencadangkan kekerapan yang semakin ekstrem bagi kebolehjadian digit pelopor signifikan awalan.

Sebaliknya, parameter β pula ialah parameter yang mengawal bahagian ekor kanan taburan yang diwakili oleh digit pelopor signifikan akhir. Parameter β menentukan kadar kebarangkalian kekerapan digit pelopor signifikan bermula daripada digit median hingga digit pelopor signifikan 9. Nilai β yang kecil menghasilkan

ekor kanan taburan yang tebal. Oleh itu, nilai β yang semakin rendah mencadangkan kekerapan yang semakin ekstrem bagi kebolehdijadian digit pelopor signifikan akhir. Parameter β penting kerana cirinya yang menyoroti serta mempertimbangkan kekerapan ekstrem bagi kebolehdijadian digit pelopor signifikan akhir yang mampu memberikan impak besar walaupun ia peristiwa yang jarang berlaku.

Dalam mengecam corak taburan, parameter α dan β memainkan peranan penting untuk mengamati set data yang mempunyai nilai kekerapan digit pelopor signifikan yang ekstrem dengan ketara. Oleh itu, parameter α dan β berupaya memberikan nilai yang luwes, membolehkan persamaan (3) disesuaikan terhadap pelbagai jenis data. Menurut Hürlimann (2015), pendekatan statistik seperti anggaran kebolehdijadian maksimum dan kaedah kuasa dua terkecil boleh digunakan untuk menentukan nilai α dan β namun Hürlimann (2015) tidak menyediakan metodologi secara khusus dan terperinci dalam penentuannya. Dalam kajian ini, HBP diperluaskan dengan menawarkan satu alat bersistematik dalam menentukan nilai parameter bentuk α dan β yang membawa kepada pembinaan algoritma HBPD.

Bagi memaksimumkan ketepatan model, kajian menggunakan teknik pengoptimuman simpleks yang terbukti berkesan dalam menentukan nilai optimum parameter bentuk taburan (Zulkarnain & Hadi 2024). Secara khusus, pembangunan algoritma mengaplikasikan kaedah pengoptimuman simpleks Nelder-Mead (NM) sebagai satu alat yang berkesan dalam pengoptimuman bebas terbitan untuk proses pembangunan algoritma HBPD (Zulkarnain & Hadi 2024). Secara lelaran, kaedah NM berusaha untuk meningkatkan kesesuaian nilai α dan β antara taburan digit yang dicerap dan dijangka. Algoritma pengoptimuman NM dimulakan dengan membentuk simpleks iaitu angka geometri yang terdiri daripada bucu dalam ruang parameter yang mewakili potensi penyelesaian (Ozaki, Yano & Onishi 2017). Simpleks ini dilaraskan secara dinamik semasa algoritma bergerak. Konfigurasi awal simpleks dilambangkan sebagai S , selalunya bermula dari titik asas x_{in} dalam ruang parameter R_n , berbentuk sama ada sebagai simpleks bersudut tegak berbanding dengan paksi koordinat atau sebagai simpleks biasa dengan panjang tepi yang sama (Ozaki, Yano & Onishi 2017). Semasa setiap lelaran, simpleks menjalani satu siri transformasi yang bertujuan untuk mencari penyelesaian yang optimum.

Proses pergerakan algoritma NM dimulakan dengan algoritma transformasi simpleks iaitu menyusun bucu simpleks dengan mengenal pasti nilai fungsi terbaik (l), kedua terbaik (s) dan terburuk (h) dalam simpleks semasa, S , mengikut nilai fungsi masing-masing $f_h = \max_j f_j$, $f_s = \max_{j \neq h} f_j$, $f_l = \min_{j \neq h} f_j$. Kedudukan susunan simpleks membimbing kepada peringkat seterusnya.

1. Pengiraan pusat bentuk. Pusat bentuk, c simpleks tanpa bucu terburuk x_h dikira dan berfungsi sebagai rujukan untuk transformasi selanjutnya (Persamaan 4):

$$c = \frac{1}{n} \sum_{j \neq h} x_j \quad (4)$$

2. Kemaskini simpleks dikawal melalui parameter pantulan, σ , pengembangan, γ , pengecutan, ω dan kerutan, δ . Parameter ini mesti memenuhi kekangan $\sigma > 0$, $0 < \omega < 1$, $\gamma > 1$, $\gamma > \sigma$, $0 < \delta < 1$. Bagi pantulan, σ , algoritma berusaha untuk menambah baik bucu terburuk dengan memantulkannya merentasi pusat bentuk dan menjana titik baru x_r (Persamaan 5):

$$x_r = c + \sigma(c - x_h) \quad (5)$$

Bagi pengembangan, γ , jika x_r memberikan hasil yang lebih baik daripada bucu semasa terbaik, simpleks dikembangkan lagi untuk meneroka lebih banyak ruang parameter (6):

$$x_e = c + \gamma(x_r - c) \quad (6)$$

Bagi pengecutan ω , jika x_r gagal menghasilkan penambahbaikan, simpleks mengecut sama ada ke arah x_r atau kembali ke arah c (Persamaan 7):

$$x_c = c + \omega(x_r - c) \text{ atau } x_c = c + \omega(x_h - c) \quad (7)$$

3. Operasi kerutan, δ . Jika tiada langkah pantulan, pengembangan dan pengecutan berjaya dalam menambah baik penyelesaian, simpleks mengerut ke arah bucu terbaik, x_l , mengurangkan semua dimensi simpleks untuk memperhalusi kawasan carian.

Melalui lelaran, kaedah NM melaraskan simpleks secara dinamik berdasarkan penilaian prestasi di bucu baharu. Penambahbaikan berterusan ini direka bentuk untuk menumpu secara berterusan ke arah penyelesaian yang paling tepat dengan meneroka dan mengeksploitasi ruang parameter dengan berkesan (Ozaki, Yano & Onishi 2017).

Dengan menyesuaikan teknik pengoptimuman simpleks, nilai optimum α dan β berjaya ditentukan. Secara tidak langsung, Persamaan (3) diperluas kebolehgunaannya kepada pelbagai data. Setelah nilai optimum α dan β dijana, algoritma HBPD diteruskan dengan mengira kekerapan jangkaan digit pelopor signifikan dalam set data latihan menggunakan Persamaan (3). Kekerapan jangkaan ini berfungsi sebagai penanda aras untuk menilai corak taburan yang dicerap dalam data sebenar.

Seterusnya, algoritma HBPD dilengkapkan dengan ujian keakuran untuk menentu sahkan taburan digit pelopor signifikan yang dicerap dalam data latihan kepada taburan yang dijangkakan oleh HBPD. Ujian keakuran bermula dengan membuat hipotesis nol bahawa data latihan akur mengikut taburan HBPD. Dalam pelaksanaan dicerap HBPD, kajian menetapkan pengkodan mutlak dengan

menolak hipotesis nol jika nilai ujian keakuran melebihi nilai genting ketidakakuran.

Kajian menggunakan dua ujian keakuran iaitu ujian keakuran sisihan min mutlak (SMM) dan khi kuasa dua (KKD). SMM ialah ukuran keakuran HBPD yang mengabaikan bilangan rekod, N . Ujian keakuran SMM dilaksanakan menggunakan Persamaan (8) (Larsen 2017).

$$|M| = \frac{\sum_{i=1}^K |A - E|}{K} \quad (8)$$

dengan A menandakan kekerapan perkadaran digit pelopor signifikan sebenar yang diperoleh dalam data latihan; E menandakan jangkaan kekerapan perkadaran digit pelopor signifikan bagi data latihan dan K mewakili bilangan digit pelopor signifikan. Terdapat tiga bahagian kepada SMM. Sisihan dalam pengangka ialah perbezaan antara kekerapan perkadaran sebenar dengan jangkaan bagi setiap digit pelopor signifikan. Fungsi mutlak ($| \ |$) mewakili sisihan diberi tanda positif tanpa mengira sama ada sisihan tersebut positif atau negatif.

Dalam Persamaan (8), pengangka mewakili sisihan mutlak daripada penjumlahan digit pelopor signifikan manakala penyebut mewakili jumlah bilangan digit pelopor signifikan iaitu 9 (1,2,3,...,9). Secara matematik, Persamaan (8) ialah purata (min) sisihan mutlak. Semakin tinggi SMM, semakin besar perbezaan purata antara perkadaran sebenar dengan jangkaan, timbal balas. Jadual 2 memberi julat nilai genting SMM untuk digit pelopor signifikan (Nigrini 2012).

Ujian keakuran KKD pula dikira dengan menjumlahkan perbezaan kuasa dua antara frekuensi yang diperhatikan dan dijangkakan yang dibahagikan dengan frekuensi yang dijangkakan untuk menormalkan nilai. Ujian keakuran KKD diberi dalam Persamaan (9) (Rodriguez 2004).

$$|x^2| = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i} \quad (9)$$

dengan n ialah bilangan digit pelopor signifikan; O_i ialah nilai frekuensi yang diperhatikan dan E_i ialah nilai frekuensi yang dijangkakan masing-masing daripada digit pelopor signifikan ke- i . Ujian keakuran KKD dijalankan dengan 8 darjah kebebasan, dikira daripada jumlah bilangan digit pelopor signifikan tolak satu (9-1). Dengan aras keertian 0.05, nilai genting bagi taburan KKD pada 8 darjah kebebasan ialah 15.5. Jika nilai keakuran KKD yang dikira melebihi nilai genting ini, hipotesis nol akur kepada HBPD ditolak.

Pembinaan algoritma BPD seterusnya memasuki peringkat kedua iaitu menguji model HBPD yang dilatih menggunakan set data uji baharu. Sama seperti set data latihan, algoritma HBPD mengisih digit pelopor signifikan

bukan sifar dalam set data uji. Seterusnya, nilai optimum α dan β yang dijana daripada set data latihan dimanfaatkan dalam Persamaan (2) untuk memodelkan kekerapan jangkaan digit pelopor signifikan dalam set data uji. Pembinaan algoritma HBPD disempurnakan dengan ujian keakuran untuk menguji keberkesanan HBPD dalam memodelkan corak taburan digit pelopor signifikan yang diperhatikan dalam set data uji.

KEPUTUSAN DAN PERBINCANGAN

Sebagai rujukan, Jadual 3 memberikan tata tanda matematik dengan definisi. Hasil analisis memberikan nilai $\alpha = 0.072$ dan $\beta = 2.154$ sebagai nilai optimum parameter bentuk untuk taburan data latihan indeks pasaran FBM KLCI yang dijana menerusi pendekatan kaedah pengoptimuman NM yang diadaptasi dalam pembinaan algoritma model HBPD. Berdasarkan nilai optimum parameter bentuk, hasil analisis memberikan keputusan kekerapan perkadaran digit pelopor signifikan bagi data latihan indeks pasaran FBM KLCI (2010-2020) yang ditunjukkan dalam Jadual 4. Corak taburan Jadual 4 ditunjukkan dalam Rajah 2.

Hasil analisis data latihan menggambarkan fenomena tingkah laku indeks pasaran FBM KLCI yang sebenar berdasarkan nilai optimum parameter bentuk yang dijana. Kajian berjaya menggunakan nilai optimum parameter bentuk yang dijana untuk memperoleh keputusan kekerapan perkadaran untuk data uji yang baharu. Keputusan data uji yang baharu diberikan dalam dalam Jadual 5, 6 dan 7 serta corak taburannya masing-masing ditunjukkan dalam Rajah 3, 4 dan 5.

Seterusnya, keputusan ujian keakuran kepada taburan Hukum Benford Pareto Diperluas diberikan dalam Jadual 8. Keputusan kajian dapat dibincangkan melalui keputusan bagi set data latihan yang bermula dari tahun 2010 hingga 2020 dan keputusan bagi set data uji bagi tahun 2020, 2021 dan 2022. Berdasarkan keputusan set data latihan, kajian mendapati siri kebarangkalian bagi kebolehjadian digit pelopor signifikan FBM KLCI berbentuk taburan susutan logaritma seperti yang dijangkakan oleh teori HBPD. Berdasarkan keputusan set data latihan, kajian mendapati siri kebarangkalian bagi kebolehjadian digit pelopor signifikan FBM KLCI berbentuk taburan susutan logaritma seperti yang dijangkakan oleh teori HBPD. Keputusan ini dapat dilihat dalam Rajah 2. Merujuk kepada Jadual 8 pula, hasil nilai ujian keakuran SMM memberikan nilai 0.0065, menunjukkan keakuran boleh terima berpandukan Jadual 2. Sementara itu, nilai ujian keakuran KKD memberikan nilai 0.0078, kurang daripada nilai genting 15.5 pada 8 darjah kebebasan. Berdasarkan keputusan ujian keakuran SMM dan KKD, hipotesis nol akur kepada HBPD gagal ditolak, merumuskan bahawa data latihan akur mengikut taburan HBPD. Oleh itu, Jadual 8 mengesahkan bahawa taburan susutan logaritma digit pelopor signifikan set data latihan akur kepada prinsip taburan yang ditentukan

JADUAL 2. Nilai genting dan keputusan SMM

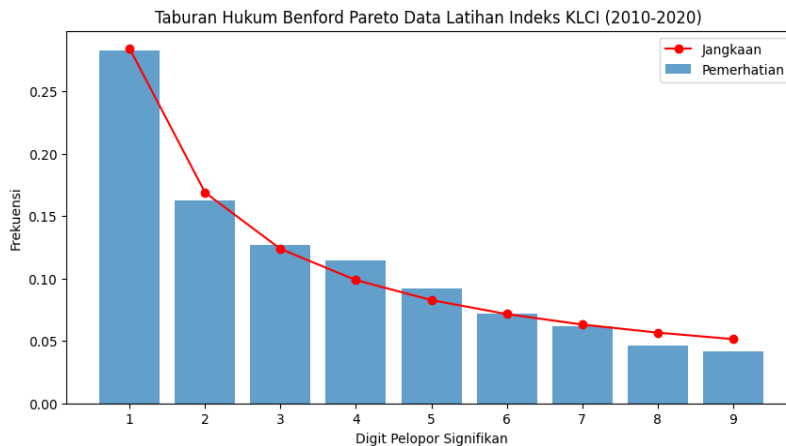
Julat	Keputusan
0.000 hingga 0.006	Keakuran rapat
0.006 hingga 0.012	Keakuran boleh terima
0.012 hingga 0.015	Keakuran boleh terima marginal
0.015 dan keatas	Ketidakkakuran

JADUAL 3. Definisi tata tanda matematik

Tata tanda	Definisi
d_1	Digit pelopor signifikan
$F_o(d_1)$	Kekerapan pemerhatian digit pelopor signifikan daripada data latihan
$F_t(d_1)$	Kekerapan jangkaan digit pelopor signifikan data latihan
$P_o(d_1)$	Nilai kebarangkalian sebenar digit pelopor signifikan data latihan
$P_t(d_1)$	Nilai kebarangkalian teori digit pelopor signifikan data latihan
$f_o(d_1)$	Kekerapan pemerhatian digit pelopor signifikan data uji
$f_t(d_1)$	Kekerapan jangkaan digit pelopor signifikan data uji
$p_o(d_1)$	Nilai kebarangkalian sebenar digit pelopor signifikan data uji
$p_t(d_1)$	Nilai kebarangkalian teori digit pelopor signifikan data uji
$ M $	Ujian keakuran SMM
χ^2	Ujian keakuran KKD
T	Jumlah bilangan

JADUAL 4. Kekerapan perkadaran data latihan indeks FBM KLCI (2010-2020)

d_1	$F_o(d_1)$	$F_t(d_1)$	$P_o(d_1)$	$P_t(d_1)$
1	737	740.8008	0.2829	0.2844
2	423	440.1297	0.1624	0.1690
3	330	322.3815	0.1267	0.1238
4	299	257.3923	0.1148	0.0988
5	239	215.6472	0.0917	0.0828
6	188	186.3544	0.0722	0.0715
7	161	164.5630	0.0618	0.0632
8	120	147.6627	0.0461	0.0567
9	108	134.1398	0.0415	0.0515
T	2605	2609	1	1



RAJAH 2. Taburan HPBD data latihan indeks FBM KLCI (2012-2020)

JADUAL 5. Kekerapan perkadaran data uji indeks FBM KLCI (2020)

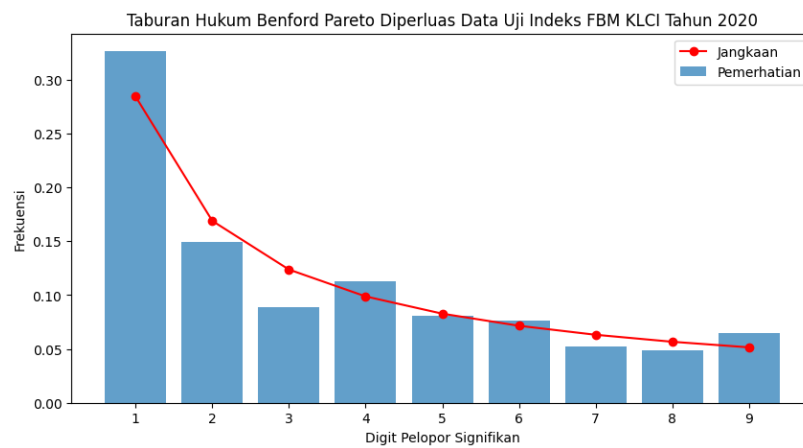
d_1	$f_o(d_1)$	$f_i(d_1)$	$p_o(d_1)$	$p_i(d_1)$
1	81	70.5254	0.3266	0.2844
2	37	41.9010	0.1492	0.1690
3	22	30.6912	0.0887	0.1238
4	28	24.5041	0.1129	0.0988
5	20	20.5299	0.0806	0.0828
6	19	17.7412	0.0766	0.0715
7	13	15.6666	0.0524	0.0632
8	12	14.0577	0.0484	0.0567
9	16	12.7703	0.0645	0.0515
T	248	248	1	1

JADUAL 6. Kekerapan perkadaran data uji indeks FBM KLCI (2021)

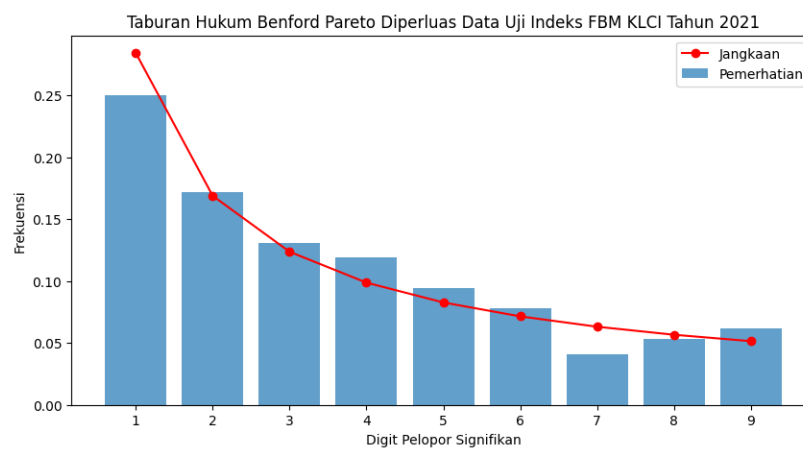
d_1	$f_o(d_1)$	$f_i(d_1)$	$p_o(d_1)$	$p_i(d_1)$
1	61	69.3879	0.2500	0.2844
2	42	41.2252	0.1721	0.1690
3	32	30.1962	0.1311	0.1238
4	29	24.1089	0.1189	0.0988
5	23	20.1988	0.0943	0.0828
6	19	17.4551	0.0779	0.0715
7	10	15.4140	0.0410	0.0632
8	13	13.8310	0.0533	0.0567
9	15	12.5643	0.0615	0.0515
T	244	244	1	1

JADUAL 7. Kekerapan perkadaran data uji indeks FBM KLCI (2022)

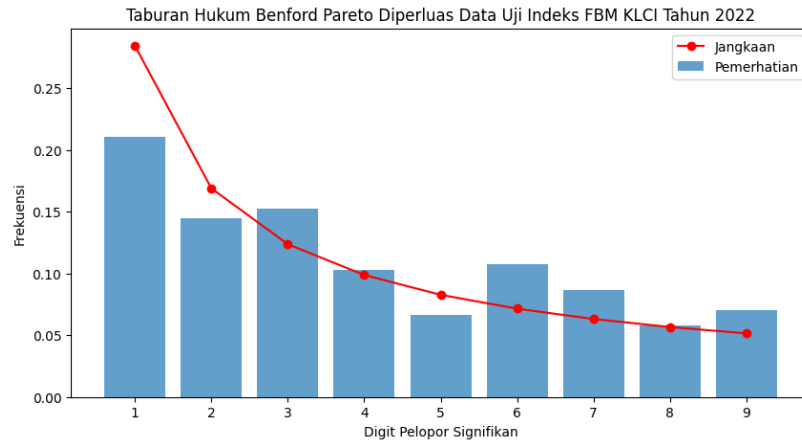
d_1	$f_o(d_1)$	$f_t(d_1)$	$p_o(d_1)$	$p_t(d_1)$
1	51	68.8191	0.2107	0.2844
2	35	40.8873	0.1446	0.1690
3	37	29.9487	0.1529	0.1238
4	25	23.9113	0.1033	0.0988
5	16	20.0333	0.0661	0.0828
6	26	17.3120	0.1074	0.0715
7	21	15.2876	0.0868	0.0632
8	14	13.7176	0.0579	0.0567
9	17	12.4614	0.0702	0.0515
T	242	242	1	1



RAJAH 3. Taburan HBPD data uji indeks FBM KLCI (2020)



RAJAH 4. Taburan HBPD data uji indeks FBM KLCI (2021)



RAJAH 5. Taburan HBPD data uji indeks FBM KLCI (2022)

JADUAL 8. Ujian keakuran

Ujian keakuran	Tahun (2010-2020)	Tahun 2020	Tahun 2021	Tahun 2022
$ M $	0.0065	0.02	0.01	0.03
χ^2	0.0078	0.03	0.02	0.07

Nilai lorek menunjukkan keputusan nilai ujian keakuran melebihi nilai genting ketidakakuran

oleh HBPD. Keputusan set data latihan dalam Jadual 8 memberikan keyakinan terhadap kebolehppercayaan algoritma HBPD dalam memahami fenomena taburan digit pelopor signifikan dalam set data, sekali gus menyokong kebolehgunaannya dalam mengesan isyarat manipulasi.

Bersandarkan kepada objektif kajian, algoritma HBPD telah berjaya menggambarkan tingkah laku senario sebenar pasaran saham Malaysia. Seterusnya, ketekalan algoritma yang telah dilatih, diuji menggunakan set data baharu. Merujuk kepada Rajah 3 hingga 5, kajian mendapati siri kebarangkalian bagi kebolehjadian digit pelopor signifikan FBM KLCI bagi tahun 2020, 2021 dan 2022 turut berbentuk taburan susutan logaritma sejajar dengan jangkaan Hukum Benford Pareto Diperluas. Keputusan bagi set data uji ini memberi keyakinan bahawa HBPD mampu menonjolkan tingkah laku fenomena sebenar peristiwa dalam pelbagai tempoh masa.

Seterusnya, nilai genting ujian keakuran dimanfaatkan untuk menentusahkan bentuk taburan susutan logaritma data baharu akur kepada HBPD. Bagi set data baharu, taburan digit pelopor signifikan yang menyeleweng daripada nilai genting menunjukkan anomali, mencadangkan kewujudan isyarat manipulasi dalam indeks pasaran FBM KLCI semasa dalam tempoh uji masing-masing. Merujuk kepada Jadual 8, ujian keakuran SMM mengesahkan bahawa taburan digit pelopor signifikan tidak akur kepada HBPD bagi tahun 2022 dan 2020 manakala akur bagi tahun 2021. Sementara itu, nilai statistik ujian keakuran KKD mengesahkan bahawa

taburan digit pelopor signifikan akur kepada HBPD bagi kesemua tahun. Oleh itu, kajian mendapati ujian keakuran SMM berjaya mengesan potensi isyarat manipulasi dalam data berbanding ujian KKD.

Keputusan ujian keakuran SMM boleh mendorong penganalisis untuk menjalankan kajian lanjut tentang proses transaksi urus niaga, pengumpulan data dan laporan penyampaian maklumat untuk memeriksa integriti dan ketepatan maklumat berkaitan indeks FBM KLCI yang mencerminkan pasaran saham Malaysia dalam tempoh tertentu. Berdasarkan analisis corak taburan digit set data uji, hasil ujian keakuran SMM memberikan nilai keakuran yang berbeza-beza kepada corak taburan yang dijangkakan oleh HBPD. Ini menunjukkan wujudnya kepelbagaian dalam corak taburan digit pelopor signifikan dalam tempoh tertentu. Dengan corak taburan digit pelopor signifikan sebagai penunjuk utama, keputusan kajian berupaya mencerminkan dinamik pasaran saham yang berubah-ubah atau wujud pelaporan yang bias pada tempoh tertentu. Oleh itu, pengujian corak taburan digit pelopor signifikan menggunakan pendekatan HBPD dari semasa ke semasa mampu memberikan pandangan berharga berkenaan tingkah laku indeks FBM KLCI secara berterusan, memastikan integriti dalam data. Pada masa yang sama, pihak berkepentingan boleh menyesuaikan strategi dan proses membuat keputusan mereka dengan sewajarnya.

Kefahaman menyeluruh tentang implikasi penyimpangan daripada Hukum Benford dapat membantu

organisasi meningkatkan strategi pengurusan risiko mereka. Sesebuah organisasi boleh menangani potensi penyelewengan dan penipuan maklumat laporan kewangan dengan mengesan potensi anomali pada peringkat awal. Ini dapat mengurangkan risiko kewangan, mengekalkan reputasi dan memastikan integriti organisasi tersebut.

Selain itu, penyelewengan antara set data uji terhadap set data latihan menekankan kepentingan pengesanan model melalui ujian keteguhan. Sungguhpun HBPD berjaya memberikan cerapan bermakna tentang corak taburan digit dalam data kewangan, kebolehgunaannya mungkin berbeza-beza merentas set data dan tempoh masa yang berbeza. Oleh itu, pengesanan berterusan dan perincian algoritma dalam menentukan nilai parameter bentuk model adalah penting untuk memastikan ketepatan dan kebolehpercayaan HBPD. Walaupun kajian ini menggunakan data pasaran saham Malaysia, ia berupaya memberi implikasi tentang kebolehgunaan algoritma HBPD dalam memodelkan corak taburan digit pelopor signifikan pada pelbagai data pasaran saham yang lain. Akhirnya, pendekatan kaedah HBPD berpotensi menilai integriti data dengan berkesan.

KESIMPULAN

Kajian menyumbangkan rangka kerja algoritma baharu sebagai satu alat forensik untuk menganalisis corak taburan digit pelopor signifikan dalam data dengan memanfaatkan prinsip pareto. Melalui pembinaan dan pelaksanaan rangka kerja Alkhawarizmi baharu, kajian ini berjaya memberikan pemahaman asas mengenai fenomena tingkah laku sebenar corak taburan digit pelopor signifikan pasaran indeks FBM KLCI Malaysia dalam tempoh 10 tahun. Di samping itu, kajian berjaya menilai keakuratan taburan data kepada taburan fenomena semula jadi Hukum Benford serta implikasinya terhadap integriti data. Hasil kajian berjaya menonjolkan keberkesanan model HBPD dalam mengesan potensi isyarat manipulasi dalam set data yang menyeleweng. Sumbangan lain kajian ialah mampu memberikan pandangan berharga tentang tingkah laku corak taburan digit pelopor signifikan yang merentas pelbagai senario keadaan data dari semasa ke semasa. Ini ekoran daripada rangka kerja kajian yang membandingkan kekerapan perkadaran digit yang diperhatikan dan dijangka dalam pelbagai tempoh. Secara langsung, rangka kerja kajian menyumbang kepada perbincangan yang lebih mendalam dalam isu kebolehpercayaan data.

PENGHARGAAN

Terima kasih diucapkan kepada Jabatan Sains Matematik, Fakulti Sains dan Teknologi, Universiti Kebangsaan Malaysia kerana menyediakan fasiliti penyelidikan.

RUJUKAN

- Buvana, M. & Muthumayil, K. 2021. Prediction of COVID-19 patient using supervised machine learning algorithm. *Sains Malaysiana* 50(8): 2479-2497.
- Benford, F. 1938. The law of anomalous numbers. *Proceedings of the American Philosophical Society* 78(4): 551-572.
- Corazza, M., Ellero, A. & Zorzi, A. 2010. Checking financial markets via Benford's law: The S&P 500 case. In *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, disunting oleh Corazza, M. & Pizzi, C. Springer, Milano. pp. 93-103. <https://doi.org/10.1007/978-88-470-1481-7>
- Hairuddin, H.B.M. & Nur-Firyal, R. 2022. Hubungan antara emas, minyak dan pasaran saham Malaysia semasa perintah kawalan pergerakan COVID-19 menggunakan regresi kuantil dan siri masa. *Sains Malaysiana* 51(7): 2315-2327.
- Hill, T.P. 1995. A statistical derivation of the significant-digit law. *Statistical Science* 10(4): 354-363.
- Hürlimann, W. 2015. Benford's law in scientific research. *Int. J. Sci. Eng. Res.* 6(7): 143-148.
- Jayasree, M. 2017. Stock trading and stock returns: Understanding the distributional properties of the numbers - The evidence from India Nifty Fifty. *Jindal Journal of Business Research* 6(2): 171-185. <https://doi.org/10.1177/2278682117727209>
- Jing, P.K. & Samsudin, H.B. 2018. Performance of real estate investment trust (REITS) after global economy crises in Malaysia. *Journal of Quality Measurement and Analysis* 14(2): 35-44.
- Khalil, M.R.A. & Bakar, A.A. 2023. A comparative study of deep learning algorithms in univariate and multivariate forecasting of the Malaysian stock market. *Sains Malaysiana* 52(3): 993-1009.
- Kossovsky, A.E. 2014. *Benford's law: Theory, the General Law of Relative Quantities, and Forensic Fraud Detection Applications* (Vol. 3). World Scientific.
- Krakar, Z. & Žgela, M. 2009. Application of Benford's Law in payment systems auditing. *Journal of Information and Organizational Sciences* 33(1): 39-51.
- Larsen, J.E. 2017. Benford's Law and earnings management detection: The case of REITs. *Journal of Forensic & Investigative Accounting* 9(2): 779-790.
- Ley, E. 1996. On the peculiar distribution of the U.S. stock indexes' digits. *The American Statistician* 50(4): 311. <https://doi.org/10.2307/2684926>
- Lobão, J. & Pereira, C. 2016. Looking for psychological barriers in nine European stock market indices. *Dutch Journal of Finance and Management* 1(1): 1-14. <https://doi.org/10.20897/lectito.201639>

- Miller, S.J. 2015. A quick introduction to Benford's Law. *Benford's Law: Theory and Applications*. Princeton: Princeton University Press. pp. 3-22.
- Nigrini, M.J. 2012. *Benford's Law: Application for Forensic Accounting, Auditing, and Fraud Detection*. Hoboken, New Jersey: John Wiley & Sons, Inc.
- Nurul Hanis Aminuddin Jafry, Ruzanna Ab Razak & Noriszura Ismail. 2020. Ukuran kebersandaran bagi pulangan lima-minit berbanding pulangan harian menggunakan kopula statik dan dinamik. *Sains Malaysiana* 49(8): 2023-2034.
- Ozaki, Y., Yano, M. & Onishi, M. 2017. Effective hyperparameter optimization using Nelder-Mead method in deep learning. *IPSJ Transactions on Computer Vision and Applications* 9: 1-12.
- Rodriguez, R.J. 2004. Reducing false alarms in the detection of human influence on data. *Journal of Accounting, Auditing & Finance* 19(2): 141-158.
- Shengmin, Z. & Wenchao, W. 2010. Does Chinese stock indices agree with Benford's law? *2010 International Conference on Management and Service Science*. pp. 1-3. <https://doi.org/10.1109/ICMSS.2010.5575999>
- Zainol Abidin, S.N., Jaaman, S.H. & Bakar, A.S.A. 2023. The analysis level of optimism that influence investor's risk tolerance in asset allocation. *Sains Malaysiana* 52(10): 2999-3011.
- Zulkarnain, N. & Hadi, M.S.A. 2024. Modelling transmission dynamics of COVID-19 during pre-vaccination period in Malaysia: A GUI-based seird predictive model using streamlit. *Journal of Quality Measurement and Analysis* 20(1): 25-40.

*Pengarang untuk surat-menyurat; email: shj@ukm.edu.my