



UNIVERSITAS
ISLAM
INDONESIA

Pidato Ilmiah

Perkembangan Analisis Data Kategori dan Tantangan di Era *Big Data*

Prof. Dr. Jaka Nugraha, S.Si., M.Si.

Guru Besar Bidang Ilmu Statistika

Jurusan Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam UII

Pidato Pengukuhan Guru Besar
Rapat Terbuka Senat
Universitas Islam Indonesia

Yogyakarta, 25 Januari 2022/22 Jumadil Tsany 1443 H

Bismillahirrahmanirrahiim

Assalamu'alaikum Warahmatullahi Wabarakatuh.

Selamat pagi dan salam sejahtera bagi kita semua.

Yang saya hormati:

1. Kepala Lembaga Layanan Pendidikan Tinggi Wilayah V Daerah Istimewa Yogyakarta.
2. Ketua Pembina, Pengawas dan Pengurus Yayasan Badan Wakaf Universitas Islam Indonesia.
3. Ketua dan Anggota Senat Universitas Islam Indonesia.
4. Ketua dan Anggota Majelis Guru Besar Universitas Islam Indonesia.
5. Rektor dan Wakil Rektor Universitas Islam Indonesia.
6. Jajaran Pimpinan Fakultas : Dekan, Wakil Dekan, Ketua Jurusan dan Ketua Program Studi di lingkungan Universitas Islam Indonesia.
7. Segenap *Civitas Akademika* Universitas Islam Indonesia.
8. Para undangan tamu, keluarga dan sejawat, hadirin yang saya muliakan yang hadir secara luring maupun mengikuti secara daring melalui kanal Zoom dan Youtube.

Puji syukur saya panjatkan ke hadirat Allah Swt. yang telah melimpahkan rahmatnya kepada kita semua. Selawat dan salam semoga senantiasa tercurahkan kepada Nabi Muhammad saw., keluarga dan para sahabatnya, serta seluruh kaum muslimin yang senantiasa mengikuti petunjuknya. Pada kesempatan ini perkenankan saya menyampaikan pidato pengukuhan sebagai Guru Besar dengan judul: “Perkembangan Analisis Data Kategori dan Tantangan di Era *Big Data*”

Pendahuluan

Manusia sudah ditakdirkan mempunyai kemampuan merekam data/informasi di dalam otaknya dan mengolahnya data menjadi informasi. Mulai dari kemampuan yang paling dasar yaitu mengenali/mengidentifikasi, mengelompokan sampai dengan mencari pola hubungan dan memprediksi. Sebagaimana dalam Surah Al-Baqarah [2]: 31:

وَعَلَّمَ آدَمَ الْأَسْمَاءَ كُلَّهَا ثُمَّ عَرَضَهُمْ عَلَى الْمَلَائِكَةِ فَقَالَ أَنْبِئُونِي بِأَسْمَاءِ هَؤُلَاءِ إِنْ كُنْتُمْ صَادِقِينَ - ٣١

Dan Dia mengajarkan kepada Adam nama-nama (benda-benda) seluruhnya, kemudian mengemukakannya kepada para Malaikat lalu berfirman: "Sebutkanlah kepada-Ku nama benda-benda itu jika kamu memang orang-orang yang benar!"

Allah Swt. memerintahkan untuk berpikir yang termaktub dalam banyak ayat di Al-Qur'an, baik dalam kalimat perintah atau pun sindiran. Di antaranya QS. Yunus [10] : 101 ;

قُلْ انظُرُوا مَاذَا فِي السَّمَوَاتِ وَالْأَرْضِ وَمَا تُغْنِي الْآيَاتِ وَالنُّذُرُ عَنْ قَوْمٍ لَا يُؤْمِنُونَ - ١٠١

Katakanlah: **"Perhatikanlah apa yang ada di langit dan di bumi.** Tidaklah bermanfaat tanda kekuasaan Allah dan rasul-rasul yang memberi peringatan bagi orang-orang yang tidak beriman".

Dalam ayat lain,

أَوَلَمْ يَنْظُرُوا فِي مَلَكُوتِ السَّمَوَاتِ وَالْأَرْضِ وَمَا خَلَقَ اللَّهُ مِنْ شَيْءٍ وَأَنْ عَسَى أَنْ يَكُونَ قَدِ افْتَرَبَ
أَجْلُهُمْ قَبَائِي حَدِيثٌ بَعْدَهُ يُؤْمِنُونَ - ١٨٥

Dan apakah mereka tidak memperhatikan kerajaan langit dan bumi dan segala apa yang diciptakan Allah, dan kemungkinan telah dekatnya waktu (kebinasaan) mereka? Lalu berita mana lagi setelah ini yang akan mereka percayai? (QS. Al-A'raf [7]: 185).

Puncak dari kemampuan berpikir manusia dalam menghasilkan pengetahuan adalah mengenal Allah Swt., menjadi Ulil Albab yaitu menyatunya pikir dan zikir. Salah satu ayat yang sangat populer tentang karakter **Ulil Albab** yaitu (QS. Ali Imran: 190-191)

إِنَّ فِي خَلْقِ السَّمَوَاتِ وَالْأَرْضِ وَالاخْتِلَافِ اللَّيْلِ وَالنَّهَارِ لَآيَاتٍ لِأُولِي الْأَلْبَابِ ۗ - ١٩٠
الَّذِينَ يَذْكُرُونَ اللَّهَ قِيَامًا وَقُعُودًا وَعَلَى جُنُوبِهِمْ وَيَتَفَكَّرُونَ فِي خَلْقِ السَّمَوَاتِ وَالْأَرْضِ رَبَّنَا مَا
خَلَقْتَ هَذَا بَاطِلًا سُبْحَانَكَ فَقِنَا عَذَابَ النَّارِ - ١٩١

“Sesungguhnya, dalam penciptaan langit dan bumi, dan pergantian malam dan siang, terdapat tanda-tanda (kebesaran Allah) bagi orang yang berakal, (yaitu) orang-orang yang mengingat Allah sambil berdiri, duduk, atau dalam keadaan berbaring, dan mereka memikirkan tentang penciptaan langit dan bumi (seraya berkata), “Ya Tuhan kami, tidaklah Engkau menciptakan semua ini sia-sia; Maha Suci Engkau, lindungilah kami dari azab neraka.”

Beberapa ayat tersebut, semoga dapat menjadi pijakan untuk selalu belajar, mengembangkan pengetahuan khususnya bidang statistika.

Berpikir Kritis dan Statistika

Bapak/Ibu/Saudara hadirin Rapat Terbuka Senat yang dimuliakan Allah Swt.

Statistika menjadi bagian penting dalam berpikir kritis (*Critical thinking*). Michael Scriven dan Richard Paul memandang bahwa “*Critical thinking is the intellectually disciplined process of actively and skilfully conceptualizing, applying, analysing, synthesizing, and/or evaluating information gathered from, or generated by, observation, experience, reflection, reasoning, or communication, as a guide to belief and action* (Kennedy dan Jones, 2009). ”Berpikir kritis sebagai proses cerdas yang meliputi konseptualisasi, penerapan, analisis, sintesis dan/atau evaluasi informasi yang diperoleh dari hasil kumpulan observasi, pengalaman, refleksi, penalaran, atau komunikasi sebagai sebuah penuntun menuju kepercayaan/keyakinan dan tindakan. Woolfolk dkk. (2008) mengatakan “*Critical thinking is the evaluating conclusions by logically and systematically examining the problem, the evidence, and the solution*”.

Statistika dikembangkan sebagai metodologi mengumpulkan data, menganalisis dan mengekstraksi data untuk mendapatkan informasi/pengetahuan. Statistika (*statistics*) adalah sekumpulan konsep dan metode yang digunakan untuk mengumpulkan dan menginterpretasikan data tentang bidang kegiatan tertentu dan mengambil kesimpulan dalam situasi di mana ada

ketidakpastian dan variasi. Pengumpulan data berkaitan dengan metode pengambilan *sampling*. Metode statistik dirancang untuk berkontribusi pada proses membuat penilaian ilmiah dalam menghadapi ketidakpastian dan variasi (Wallpole dkk, 2012).

Jadi, statistika dikembangkan sebagai alat bantu untuk dapat berpikir kritis dan Allah Swt. telah memerintahkan manusia untuk berpikir sebagaimana difirmankan dalam beberapa ayat yang telah saya sebutkan.

Data Kategori

Statistika adalah ilmu yang berkaitan dengan data sehingga statistika sangat identik dengan data. Sementara istilah 'data' dan 'statistik' sering digunakan secara bergantian, dalam penelitian ilmiah ada perbedaan penting di antara keduanya. Statistik (*Statistic*) adalah ukuran numerik dari sampel yang pada populasi disebut parameter. Data adalah informasi individu yang direkam dan digunakan untuk tujuan analisis. Data adalah informasi mentah sedangkan statistik adalah hasil analisis data atau hasil perhitungan.

Identifikasi data merupakan tahapan penting sebelum analisis dilakukan sebab metode analisis statistik sangat tergantung pada sifat data. Data yang berupa atau dinyatakan dalam bilangan disebut data kuantitatif sedangkan data bukan bilangan disebut data kualitatif. Data kuantitatif sering disebut dengan istilah data numerik yang mencakup data kontinu dan data diskrit. Data kontinu diperoleh dari hasil pengukuran sedangkan data diskrit diperoleh dari hasil menghitung atau membilang bukan mengukur. Dilihat dari skala

pengukurannya, data dapat diklasifikasikan menjadi empat tingkat yaitu nominal, ordinal, interval, dan rasio. Data numerik mencakup skala interval dan rasio sedangkan data kualitatif mencakup skala nominal dan ordinal. Pada kasus khusus, ketika hanya ada dua kemungkinan nilai maka disebut data biner. Data biner ini dapat dipandang sebagai data diskrit, atau ordinal, atau nominal tergantung interpretasi peneliti.

Agresti (2019) menggunakan istilah data kategori sebagai pengganti istilah data kualitatif. Power dan Xie (2008) mendefinisikan data kategori adalah hasil pengukuran variabel yang memiliki nilai sehingga memasukan data diskrit ke dalam data kategori. Pengembangan metode analisis untuk data kategori telah memberikan kontribusi yang sangat besar terhadap penelitian ilmiah di bidang berbagai bidang seperti biostatistik, ekonomi, psikologi, sosiologi dan kesehatan maupun transportasi. Pengembangan metode analisis data kategori pada masing-masing bidang ini menghasilkan metode maupun istilah yang spesifik.

Perkembangan metode analisis data kategori tidak secepat perkembangan metode analisis pada data numerik. Berbeda dengan data numerik, agar dapat dilakukan analisis maka data kategori direpresentasikan atau dikuantifikasi ke dalam nilai frekuensi dan proporsi. Analisis yang umum dilakukan pada data kategori mencakup (1) inferensi parameter proporsi, (2) asosiasi antar variabel, (3) pemodelan.

Inferensi Parameter Proporsi

Dalam kurun waktu tiga puluh tahun terakhir, telah banyak peneliti mengembangkan metode estimasi parameter proporsi untuk mendapatkan metode yang paling akurat (Pires dan Amado, 2008; Newcombe, 1998)). Inferensi parameter π secara garis besar dapat dikelompokkan ke dalam dua cara, yaitu metode eksak dan metode pendekatan distribusi normal. Metode eksak digunakan pada sampel kecil sedangkan metode pendekatan distribusi normal digunakan ketika jumlah sampel besar.

Estimasi interval untuk parameter π telah banyak dikembangkan oleh beberapa peneliti dengan menggunakan metode eksak seperti metode Clopper-Pearson (Clopper dan Pearson, 1934), *Central Exact interval* (Agresti dan Coull, 1998.) distribusi Beta($k, n-k-1$) dan distribusi F [García-Pérez :2005)]. Di antara metode eksak, metode Clopper-Pearson adalah metode yang paling baik dan perhitungannya lebih sederhana (Reed, 2007).

Disamping metode eksak, inferensi statistik juga dikembangkan dengan menggunakan metode pendekatan distribusi Normal seperti statistik *Wald*, *Wilson's Score*, *Likelihood based Methods* (Fleiss dkk: 2003). Pendekatan ini cocok pada dengan menggunakan aturan yang bersifat "*rule of thumb*" yaitu nilai harapan ($n\pi$) lebih dari 5. Ketika π mendekati 0.5 jelas distribusi binomial bersifat simetris yang relatif dekat dengan sifat distribusi normal. Tetapi π mendekati 0 (nol) atau mendekati 1 (satu) maka distribusinya menjadi tidak simetris khususnya pada jumlah sampel yang kecil. Pendekatan distribusi normal pada pada sampel besar

menghasilkan penaksir parameter proporsi yang lebih baik (Agresti dan Coull, 1998). Berdasarkan pengujian menggunakan data simulasi pada beberapa nilai π , sifat distribusi normal terpenuhi jika n lebih dari $20/\pi$. Ukuran sampel yang diperlukan pada Simple Asymptotic with *continuity correction* lebih kecil dibandingkan dengan *Wilson Score with continuity correction* (Nugraha dkk, 2021).

Pada inferensi selisih dua parameter proporsi, metode yang dikembangkan di antaranya adalah:

1. *Score* dari Farrington-Manning).
2. *Score* dari Miettinen-Nurminen).
3. *Score with Correction for Skewness* dari Gart-Nam.
4. *Score* dari Wilson.
5. *Score with Continuity Correction* dari Wilson.
6. *Chi-Square with Continuity Correction* dari Yates.
7. *Chi-Square* Pearson.
8. Beal's Haldane *method*.
9. Beal's Jeffreys-Perks *method*.
10. *Likelihood method*.
11. *Likelihood method based on 'exact'*.
12. *Likelihood method based on 'mid-p'*

Newcombe (1998b) melakukan evaluasi komparatif dari sebelas metode interval kepercayaan dan merekomendasikan bahwa metode *Score* Wilson yang dimodifikasi dapat menggantikan Pearson *Chi-Square* atau Yate's *Corrected Chi-Square*. Beal (1987) menemukan bahwa metode *Score* mempunyai performa yang sangat baik.

Asosiasi Antar Variabel Kategori

Hasil pengamatan yang variabelnya berupa kategori dapat disajikan dalam tabel kontingensi. Untuk menguji adanya asosiasi dalam tabel kontingensi, khususnya tabel 2x2 tidak ada kata sepakat dalam menetapkan uji yang paling sesuai. Perdebatan mengenai ukuran asosiasi dimulai pada awal abad 20 seperti Pearson (1913) dan Yule (1912). Pearson (1904, 1913) telah membahas dengan pendekatan distribusi kontinu bivariat. Pearson meyakini bahwa dapat dilukiskan asosiasi antar faktor dengan menggunakan ukuran seperti korelasi sebagaimana data kontinu. Ukuran tersebut dikenal dengan sebagai *Pearson's contingency coefficient* (uji Pearson Chi-square). Pearson berpendapat bahwa koefisien yang disarankan oleh Yule (1900, 1912) untuk tabel 2x2 yang disebut *odds ratio* tidaklah cocok.

Statistik uji yang lain yang diusulkan oleh Fisher (1922, 1925, 1936) dan Irwin (1949) adalah *Fisher's exact test*. Yates (1934) mengusulkan statistik *Pearson with Continuity Correction* yang memberikan *p-value* yang lebih baik dibandingkan pendekatan distribusi hipergeometrik dalam *Fisher's exact test*. Goodman dan Kruskal (1959) telah menjelaskan sejarah perkembangan ukuran asosiasi pada tabel kontingensi. Ukuran-ukuran asosiasi pada data ordinal yang didasarkan sifat konkordan dan diskordan diusulkan oleh Kendall (1945), Kruskal (1958), Goodman dan Kruskal (1979) dan Agresti (1984). Campbell (2007) telah memberikan bukti baru bahwa pada kondisi nilai harapan semua sel minimal sebesar 1, uji ' $N-1$ ' *Chi-square* lebih direkomendasikan dibandingkan dengan uji *Chi-square* Pearson, uji *Chi-*

square Pearson dengan “*Yate Correction*” maupun uji *Fisher-Irwin*.

Pada tahun 1970-an analisis tabel kontingensi berubah cukup dramatis dengan munculnya model log-linear oleh Goodman (1970, 1971). Model Log-linear merupakan model statistik yang paling populer dan penting untuk analisis data kategorial (Christensen, 1997). Model log-linear memiliki aplikasi dalam banyak bidang ilmiah, mulai dari ilmu sosial-ekonomi dan biologi, hingga masalah kedokteran, pemrosesan bahasa, dan genetika. Sekarang ini, aplikasi model linear sudah sangat luas dan banyak *software* yang sudah memuat analisis log-linear. Analisis log-linear membutuhkan dua langkah (a) mencari model yang sesuai, (b) menafsirkan model yang dipilih.

Model log-linear dapat digunakan untuk hubungan antara variabel-variabel kategori yang tidak membedakan antara variabel respon dan prediktornya. Dalam beberapa kasus seringkali terjadi kesimpulan yang kontradiksi ketika analisis hanya mempertimbangkan dua faktor (variabel) saja dibandingkan dengan analisis yang melibatkan tiga variabel. Hasil analisis dengan mempertimbangkan tiga variabel akan menghasilkan kesimpulan yang lebih akurat dibandingkan dengan hanya melibatkan dua faktor saja. Kontradiksi dua kesimpulan ini biasa disebut *Simpson Paradox*. Tiga asumsi dasar harus dipertimbangkan ketika menggunakan model loglinear (a) Pengamatan independen satu sama lain, (b) Jumlah pengamatan besar (Wickens, 1989).

Model Data Kategori Pendekatan GLM dan DCM

Terdapat dua filosofi yang berbeda dalam memandang data kategori. Kelompok pertama menganggap variabelnya pada hakikatnya hanya memiliki sifat kategori dan kelompok kedua mengasumsikan bahwa pada hakikatnya variabel bersifat kontinu tetapi dalam pengukuran atau observasi diperoleh nilai kategori. Dari dua filosofi yang berbeda ini menghasilkan beberapa metode analisis yang berbeda pada kasus/masalah yang sama.

Pengembangan model statistik pada kelompok pertama dilakukan dengan pendekatan transformasi yang biasa dikenal dengan istilah *link function* pada *Generalized Linear Model* (McCullagh dan Nelder: 1989; Rodriguez, 2001). Pada percobaan Binomial dan Multinomial dapat dilakukan transformasi logit (link logit) dengan rumus $\ln(\pi/(1-\pi))$ dan *link* probit yang merupakan *invers* dari fungsi distribusi Normal Standar. Pada percobaan Multinomial dan Poisson dapat menggunakan *natural logaritma function* (*log link*) atau transformasi *loglinear*.

Pemodelan untuk respon berupa variabel nominal, pendekatan paling sederhana adalah dengan menominasikan salah satu kategori respons sebagai *baseline* atau sel referensi untuk menyusun model logit (*log-odds*). Model ini dikenal dengan model logit Multinomial yang analog dengan model regresi logistik biner, kecuali bahwa distribusi probabilitas dari respon adalah multinomial bukan binomial.

Pengukuran variabel pada skala ordinal sudah tidak asing lagi yang merupakan pengembangan dari skala

nominal. Pengukuran ini banyak dijumpai pada penelitian konseling/psikologi (misalnya, peringkat dari 1 = sangat tidak mungkin sampai 5 = sangat mungkin) (Ness, 1995). Pada pengamatan kondisi klinis pasien setelah terapi dapat dicirikan sebagai memburuk (1), tidak berubah (2), atau membaik (3) (Grissom, 1994). Pemodelan pada data ordinal dapat dilihat sebagai pengembangan dari data biner, sehingga langkah-langkah analisisnya mengikuti model regresi logistik.

Terdapat beberapa metode untuk menganalisis data ordinal :

- a) *Model Cumulative Odds* (CO) (Agresti, 1996; Armstrong & Sloan, 1989; Long, 1997; McCullagh, 1980).
- b) *Model Continuation Ratio* (CR) (Armstrong dan Sloan, 1989; Cox, 1972; Greenland, 1994),
- c) *Model adjacent categories* (AC) (Agresti, 1989; Goodman, 1983).
- d) *Model partial proportional odds* (Peterson dan Harrell, 1990).
- e) *Model partial proportional hazards* atau *model unconstrained continuation ratio* (Bender & Benner, 2000; Cole dan Ananth, 2001).

Filosofi kedua yang memandang bahwa pada hakikatnya variabelnya kontinu, melakukan pengembangan metode analisis menggunakan pendekatan variabel laten. Variabel kategori dipandang sebagai *limited-dependent variable* (Maddala, 1983). Pendekatan ini menghasilkan metode analisis yang biasa disebut dengan *Discrete Choice Model* (DCM) yang banyak diaplikasikan pada data survei. DCM mengasumsikan bahwa pembuat keputusan (responden) menentukan

pilihan (membuat keputusan) di antara sekumpulan pilihan berdasarkan nilai utilitasnya. Utilitas inilah yang dipandang sebagai variabel laten yang bersifat kontinu dan tidak dapat diukur. Pilihan atau respon berupa data nominal dan dipilih salah satu yang mempunyai nilai utilitas tertinggi (maksimum) oleh pembuat keputusan. DCM dapat menggambarkan pembuat keputusan dalam menentukan pilihan. Pembuat keputusan atau responden di sini dapat berupa orang, keluarga, perusahaan atau unit pembuat keputusan yang lain. DCM mulai berkembang dan banyak diaplikasikan di bidang ekonomi dan transportasi setelah ditemukannya model *Multinomial Logit* (MNL) oleh Mcfadden pada tahun 1974 (McFadden, 1974). Model ini juga dikenal dengan nama *Conditional Logit Model* (Train, 2003). Dalam DCM, model yang telah dikembangkan dan diaplikasikan antara lain model Logit, model *Generalized Extreme Value* (GEV), model Probit dan Mixed Logit (McFadden dan Train, 2000; Bhat, 2001; Horowitz dan Savin, 2001).

Pada bidang DCM ini, saya di bawah bimbingan Prof. Dr. Suryo Guritno, M.Stat dan almarhumah Profesor Dr. Sri Haryatmi, M.Sc telah menulis disertasi pada tahun 2010 dengan judul “Pemodelan Pilihan Diskrit Menggunakan Model Probit dan Model *Mixed Logit* Pada Respon Multivariat”. Sebelumnya, pada tahun 1999 telah menulis tesis dengan judul “Penaksiran Parameter Regresi Logistik Bivariat”.

Data Kategori pada *Big Data*

Pada data survei yang memuat banyak variabel kategori, Kass (1980) telah mengusulkan penggunaan algoritma CHAID (*Chi-square Automatic Interaction Detector*). Metode ini digunakan untuk mendeteksi asosiasi antar variabel kategori yang dilakukan dengan cara pemisahan/pengelompokan populasi dan pengujian didasarkan pada statistik Pearson's *Chi-square*. Algoritmanya meminimumkan variasi variabel dependen dalam satu grup dan memaksimumkan variasi antar grup. Metode ini terus di berkembang menyesuaikan kompleksitas data. Pengembangan dari aspek algoritma seperti:

- a. Metode *Classification and Regression Trees* (CART) oleh Breiman dkk (1984),
- b. Metode *Iterative Dichotomiser 3* (ID3) oleh Quinlan (1986) dan
- c. Algoritma C4.5 oleh Quinlan (1993).

Dari sisi statistika, CHAID telah dikembangkan oleh:

- a. *Quick, Unbiased and Efficient Statistical Tree* (QUEST) (Loh dan Shih, 1997).
- b. *Classification Rule with Unbiased Interaction Selection and Estimation* (CRUISE) (Kim dan Loh, 2001).
- c. *Generalized, Unbiased, Interaction Detection and Estimation* (GUIDE) (Loh, 2009).
- d. *Conditional Inference Trees* (CTREE) (Hothorn dkk, 2006).

Analisis log-linier pada pendekatan klasik hanya mampu mencari asosiasi antara variabel kategori tidak lebih dari sepuluh variabel. Petitjean dkk. (2013) telah menunjukkan bahwa, analisis log-linear dapat diterapkan

pada kumpulan data dengan ratusan variabel. Petitjean dan Webb (2015) menggunakan *Prioritized Chordalylis* pada analisis log-linear yang diaplikasikan pada 2000 variabel menghasilkan proses komputasi menjadi cepat yaitu dalam hitungan detik. Pada pendekatan klasik, prosesnya membutuhkan waktu beberapa hari.

Seiring dengan kemajuan IT, data senantiasa berkembang dan semakin kompleks baik dari volume, cacah variabel dan kecepatan perubahan data yang kita kenal dengan *big data*. Dari hasil pengujian secara simulasi yang telah dilakukan oleh Petitjean, dkk. (2013, 2015), sangat mungkin untuk mengimplementasi model log-linear pada *big data*. Sedangkan implementasi model logistik pada *big data*, dengan menggunakan *Enhanced Logistic Regression* (ELR) menghasilkan akurasi yang lebih tinggi dibanding dengan model regresi logistik biasa (Dhamodharavadhani dan Rathipriya, 2019). Namun demikian dalam mengklasifikasi objek pada *big data*, metode Naïve Bayes menghasilkan akurasi yang lebih tinggi dibandingkan dengan regresi logistik (Jacob dan Vijayakumar, 2019). Penerapan regresi logistik nominal dan ordinal pada *big data*, masih menjadi permasalahan terutama dari aspek komputasinya.

Penutup

Analisis data tidak sebatas perhitungan matematis, penggunaan rumus dan komputasi tetapi juga merupakan seni untuk mendapatkan informasi yang relevan terhadap permasalahan yang akan diselesaikan sesuai dengan kaidah statistika. Analisis data merupakan perpaduan dari pemahaman masalah, pemilihan metode yang tepat dan proses komputasi yang akurat dan cepat. Analisis data kategori telah diaplikasikan dalam hampir semua bidang ilmiah, mulai dari ilmu sosial-ekonomi dan biologi, kedokteran, teknik, bahasa, dan genetika.

Analisis data kategori mayoritas pengembangan metodenya menggunakan pendekatan data kontinu dan masih ditemukan beberapa keterbatasan sehingga membutuhkan pengembangan lebih lanjut. Keterbatasan dari aspek interpretasi kuantitatif ditemukan pada ukuran asosiasi yang didasarkan pada statistik *Chi-square*. Interpretasi parameter λ yang dipakai untuk mengukur efek faktor maupun efek interaksi dalam model log-linear linear. Uji *goodness of fit* dan nilai pseudo *R-square* pada regresi logistik serta masalah komputasi ketika diterapkan pada *big data*.

Ucapan Syukur, Penghargaan, dan Ucapan Terima Kasih

Bapak/Ibu/Saudara hadirin Rapat Terbuka Senat yang dimuliakan Allah Swt.

Sebelum mengakhiri pidato ini, perkenankan saya kembali memanjatkan puji syukur kehadirat Allah Swt. atas rida, karunia, dan amanah sehingga saya berkesempatan untuk diangkat menjadi Guru Besar bidang Statistika di Universitas Islam Indonesia. Selanjutnya kami mengucapkan terima kasih kepada :

- 1) Kedua orang tua saya, Bapak Ngatiman Prpto Raharjo - Ibu Murini, Bapak dan Ibu mertua, Bapak Warsito - Ibu Zubaidah, yang senantiasa mendoakan untuk kebaikan putra-putrinya.
- 2) Istri tercinta Is Fatimah yang selalu mendampingi dalam suka maupun duka dan ketiga anak kami tercinta, Naufal Abdillah, Haikal Abdurahman dan Inaya Izaturrizqi sebagai penyemangat untuk terus berkarya.
- 3) Guru-guru kami di SDN Jurang-jero, SMPN 2 Karanganom dan SMAN Jatinom. Semoga ilmu yang telah diajarkan khususnya kepada kami tercatat sebagai amal jariyah.
- 4) Prof. Drs. Suryo Guritno, M.Stat., Ph.D, Almh. Prof. Dr. Sri Haryatmi, M.Sc., Almh Dra. Sri Pangesti SU., Drs. **Zulaela**, Dipl.Med.Stats., M.Si. yang telah membimbing skripsi, tesis dan disertasi, serta Prof. Drs. Subanar, PhD dan semua dosen di Jurusan Matematika-Statistika UGM.
- 5) Pengelola dan Pengasuh Pondok Pesantren Budi Mulia yang telah memberikan kesempatan kepada saya untuk banyak belajar tentang Islam dan Kebangsaan, serta

- teman-teman Santri Angkatan IV yang bersama-sama belajar dari tahun 1991-1993.
- 6) Kepala Lembaga Layanan Pendidikan Tinggi (LLDikti) Wilayah V Yogyakarta (2019- 2021), Prof. Dr. Didi Achjari, M.Com. Akt., beserta jajarannya yang telah mengusahakan penyelesaian pemberkasan dengan cepat di bulan Desember 2020.
 - 7) Ketua Umum Pengurus Yayasan Badan Wakaf, Rektor, Wakil Rektor Universitas Islam Indonesia, Ketua, Sekretaris, dan segenap anggota Senat Universitas Islam Indonesia, Ketua dan Anggota Majelis Guru Besar Universitas Islam Indonesia, Dekan dan Wakil Dekan Fakultas MIPA 2018-2022, Ketua dan Sekretaris Jurusan/Program Studi Statistika, serta segenap Bapak/Ibu Dosen dan Tenaga Kependidikan Fakultas MIPA Universitas Islam Indonesia dan semua pihak yang tak dapat saya sebutkan satu persatu.
 - 8) Tim Kerja DPA : Pak Agung Nugroho Adi, M.T., Abdur Rafik M.Si., Mutiara Nur Zulaikha, S.Psi., dan Aris Rusdan Perwira, Amd. yang telah berbagi tugas dengan sangat baik dalam menjalankan aktivitas di DPA, sehingga saya bisa mengurus kenaikan jabatan akademik Guru Besar ini.

Semoga Allah Swt. senantiasa melimpahkan rahmat, hidayah dan taufik kepada semua pihak yang telah membantu tercapainya jabatan akademik ini serta terlaksananya pidato pengukuhan hari ini. Semoga semua ini menjadi amal ibadah yang yang diridai Allah Swt.

*Billaahi taufik wal hidayah, Wassalamu'alaikum
Warohmatullaahi Wabarokatuh*

Daftar Pustaka

- Agresti A. (1984). *Analysis of Ordinal Categorical Data*. New York: Wiley.
- Agresti A. (1989). Tutorial on modeling ordered categorical response data. *Psychological Bulletin*, 105(2), 290–301.
- Agresti A. (2019). *An Introduction to Categorical Data Analysis, 3rd ed.*, Wiley.
- Agresti A. (1996). *An introduction to categorical data analysis*. New York: John Wiley & Sons.
- Agresti A.Z & Coull B. (1998). Approximate is better than “exact” for interval estimation of binomial proportions, *The American Statistician*, 52, 119–126.
- Armstrong B.G. & Sloan M. (1989). Ordinal regression models for epidemiological data. *American Journal of Epidemiology*, 129(1), 191–204.
- Beal S.L. (1987). Asymptotic confidence intervals for the difference between two binomial parameters for use with small samples, *Biometrics*, 43, 941—950.
- Bender R. & Benner A. (2000). Calculating ordinal regression models in SAS and S-Plus. *Biometrical Journal*, 42(6), 677–699.

- Bhat C.R. (2001). Quasi-random Maximum Simulated Likelihood Estimation of The Mixed Multinomial Logit Model, *Transportation Research*, 35B(7), 677-695.
- Breiman L. (2001). Random forest. *Machine Learning* 45, 5–32.
- Breiman L., Friedman J.H., Olshen R.A., & Stone C. J. (1984). *Classification And Regression Trees*. New York: Chapman and Hall.
- Campbell I. (2007), Chi-squared and Fisher-Irwin ests of two-by-two tables with small sample recommendations. *Statistics in Medicine*, 26, 3661-3675,
- Christensen R. (1997). *Log-linear Models and Logistic Regression* (2nd ed.). New York, NY: Springer. doi: 10.1007/b97647
- Clopper C.J. & Pearson E.S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, 26, 404-413.
- Cole S.R. & Ananth C.V. (2001). Regression models for unconstrained, partially or fully constrained

- continuation odds ratios. *International Journal of Epidemiology*, 30, 1379–1382.
- Cox D.R. (1972). Regression models and life tables [with discussion]. *Journal of the Royal Statistical Society B*, 74, 187–220.
- Dhamodharavadhani S. & Rathipriya R. (2019), *Enhanced Logistic Regression (ELR) Model for Big Data*, IGI Global publisher .
- Everitt B.S, (2006). *The Cambridge Dictionary Of Statistics Third Edition* , Cambridge University Press
- Fleiss J.L., Levin B. & Paik M.C. (2003). *Statistical Methods for Rates and Proportions*, Third Edition, Wiley, New York.
- Fisher R.A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 222 (594-604), 309-368.
doi:10.1098/rsta.1922.0009
- Fisher R.A. (1925). *Statistical Methods for Research Workers*. New York: Hafner Press.

- Fisher R.A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, **7**, 179–188.
- García-Pérez M.A. (2005). On the confidence interval for the binomial parameter, *Quality & Quantity*, **39**, 467–481.
- Greenland S. (1994). Alternative models for ordinal logistic regression. *Statistics in Medicine*, **13**, 1665–1677.
- Goodman L.A. (1970). The multivariate analysis of qualitative data: Interaction among multiple classifications. *Journal of the American Statistical Association*, **65**(329), 226–256. doi: 10.2307/2283589
- Goodman L.A. (1971a). The analysis of multidimensional contingency tables: Stepwise procedures and direct estimation methods for building models for multiple classifications. *Technometrics*, **13**(1), 31–66. doi: 10.2307/1267074
- Goodman L.A. (1971b). The partitioning of chi-square, the analysis of marginal contingency tables, and the estimation of expected frequencies in

- multidimensional contingency tables. *Journal of the American Statistical Association*, 66(334), 339-344.
doi: 10.2307/2283933
- Goodman L.A. & Kruskal W.H. (1979). *Measures of Association for Cross Classifications*. New York: Springer-Verlag
- Goodman L.A. (1983). The analysis of dependence in cross-classifications having ordered categories, using log-linear models for frequencies and log-linear models for odds. *Biometrics*, 39, 149–160.
- Horowitz J.L. & Savin N.E., (2001). Binary Response Models : Logits, Probits and Semiparametrics, *Journal of Economic Perspectives*, Volume 15, 4 pages 43-56
- Hothorn T.K, Hornik K. & Zeileis A. (2006). Unbiased recursive partitioning: a conditional inference framework. *Journal of Computational and Graphical Statistics*, 15, 651–674.
- Irwin, J.O. (1949). A note on the subdivision of χ^2 into components. *Biometrika*: 36, 130-134.
- Jacob S.S. & Vijayakumar R. (2019). Naïve Bayes & Logistic Regression on Big Data: A Performance

Analysis, *International Journal of Applied Engineering Research* Volume 14, Number 5, pp. 1102-1105

Kass G.V. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics* : 29 (2), 119–127.

Kendall M.G. (1945). The treatment of ties in rank problems. *Biometrika* :33, 239-251.

Kennedy M.L. & Jones R. (2009). *Critical thinking*. SLA 2009 Annual Meeting Washington DC.

Kim H. & Loh W.-Y. (2001). Classification trees with unbiased multiway splits. *J. Amer. Statist. Assoc.*, 96, 589–604

Kruskal W.H. (1958). Ordinal measures of association. *J. Amer. Statist. Assoc.* 53: 814-861.

Loh W.-Y. (2009). Improving the precision of classification trees. *Ann. Appl. Stat.*, 3, 1710–1737

Loh W.-Y. & Shih Y.-S. (1997). Split selection methods for classification trees. *Statistica Sinica* : Vol. 7, pp. 815–840.

- Maddala G.S. (1983), *Limited-Dependent and Qualitative Variables in Economics*, New York: Cambridge University Press
- Long J.S. (1997). *Regression models for categorical and limited dependent variables*. Thousand Oaks, CA: Sage
- McCullagh P. (1980). Regression models with ordinal data [with discussion]. *Journal of the Royal Statistical Society, B*, 42, 109–142.
- McCullough P. & Nelder J.A. (1989). *Generalized Linear Models*, 2nd. ed. Chapman and Hall, New York, NY
- McFadden D. (1974). Conditional Logit Analysis of Qualitative Choice Behavior, dalam Zarembka, ed., *Frontiers in Econometrics*, Academic Press, New York, pp. 105–142
- McFadden D. & Train K. (2000). Mixed MNL Models for Discrete Response, *Journal of Applied Econometrics* 15(5), 447-470
- Newcombe R.G. (1998). Interval Estimation For The Difference Between Independent Proportions: Comparison Of Eleven Methods, *Statistics in Medicine*, 17, 873–890.

- Newcombe R.G. (1998). Two-sided confidence intervals for the single proportion: comparison of seven methods, *Statistics in Medicine*, 17, 857–872.
- Nugraha J. (1999). Estimasi Parameter Regresi Logistik Biner Bivariat. *Thesis*. FMIPA-UGM
- Nugraha J. (2010). Model Pemilihan Diskrit Menggunakan Model Probit Dan Model Mixed Logit Pada Respon Multivariat. *Disertasi*. FMIPA-UGM
- Nugraha J & Fauzan A. (2021). Identifikasi Sifat-Sifat Penaksir Parameter Proporsi Pada Distribusi Binomial. *Laporan Penelitian*. DPPM-UII.
- Pearson K. (1904). *Mathematical Contributions to the Theory of Evolution*. London, UK: Dulau and Co.
- Pearson K. (1913). On the probable error of a correlation coefficient as found from a fourfold table. *Biometrika* 9: 22-27.
- Peterson B. L. & Harrell F. E. (1990). Partial proportional odds models for ordinal response variables. *Applied Statistics*, 39(3), 205–217
- Petitjean F., Webb G.I. & Nicholson A.E. (2013). Scaling log-linear analysis to high-dimensional data, *IEEE Int. Conf. on Data Mining*, pp. 597–606.

- Petitjean F., & Webb G.I. (2015) , Scaling log-linear analysis to datasets with thousands of variables, *'Proceedings of the 2015 SIAM International Conference on Data Mining (SDM). Society for Industrial and Applied Mathematics.*
- Pires A.M., Amado C. (2008). Interval Estimators for a Binomial Proportion: Comparison of Twenty Methods. *REVSTAT - Statistical Journal*, 6, 165-197.
- Powers D. dan Yu Xie. (2008). *Statistical Methods for Categorical Data Analysis*, 2nd Edition. Emerald Group Publishing Limited.
- Quinlan J.R. (1986). Induction of decision trees. *Machine Learning* 1, 81–106.
- Quinlan J.R. (1993). C4.5: Programs for Machine Learning. *Sinica* 7, 815–840.
- Reed JF.III. (2007). Better Binomial Confidence Intervals. *Journal of Modern Applied Statistical Methods*: Vol. 6 ; No. 1, 153-161
- Ritschard G. (2006). Computing and using the deviance with classification trees. In A. Rizzi and M. Vichi (Eds.), *COMPSTAT 2006 - Proceedings in*

- Computational Statistics*, pp. 55–66. Berlin: Springer.
- Ritschard G. & Zighed D.A. (2003). Goodness-of-fit measures for induction trees. In N. Zhong, Z. Ras, S. Tsumo, and E. Suzuki (Eds.), *Foundations of Intelligent Systems, ISMIS03*, Volume LNAI 2871, pp. 57–64. Berlin: Springer.
- Rodriguez G. (2001). *Generalized Linear Models*, Princeton University
- Strobl C., Malley J. & Tutz G. (2009). An introduction to recursive partitioning. *Psychological Methods* 14(4), 323–348.
- Train K. (2003). *Discrete Choice Methods with Simulation*, UK Press, Cambridge
- Walpole RE., Mayers RH., Myers SL., Ye K. (2012). *Probability & statistics for engineers & scientists*. 9th ed. Prentice Hall.
- Wickens T.D. (1989). *Multiway Contingency Tables Analysis for the Social Sciences*. Hillsdale, NJ: Lawrence Erlbaum.
- Woolfolk A.H., Hughes M.M. & Walkup V. (2008). *Psychology in Education*. New York: Pearson.

- Yates F. (1984), Tests of significance for 2×2 contingency tables (with discussion). *Journal of the Royal Statistical Society Series A*: 147: 426-463.
- Yule G.U. (1900). On the association of attributes in statistics: With illustration from the material of the childhood society, etc. *Philosophical Transactions of the Royal Society, Series A*, 194, 257-319.
- Yule G.U. (1912). On the methods of measuring association between two attributes. *J. Roy. Statist. Soc.* 75: 579-642.

BIODATA

A. Identitas Diri

1	Nama lengkap	Prof. Dr. Jaka Nugraha, S.Si. M.Si
2	Jabatan Fungsional	Guru Besar Bidang Ilmu Statistika
4	NIP/NIK	956110102
5	NIDN	0524027101
6	Scopus ID/ h-Index	57193439320 / 2
7	Sinta ID	20784
8	Tempat dan Tanggal Lahir	Klaten, 24 Februari 1971
9	Alamat Rumah	Bulusan, RT 03 RW 39 Sardonoharjo, Ngaglik, Sleman, Yogyakarta
10	Nomor Telepon/HP	087738981118
11	Alamat Kantor	Jurusan Statistika, FMIPA, UII Jl. Kaliurang km 14,4 Besi Sleman Yogyakarta
12	Nomor Telepon/Faks	(0274) 898920
13	Alamat email	jnugraha@uii.ac.id

B. Riwayat Pendidikan

1. Sekolah Dasar : SD Negeri Jurang-jero (1977 – 1983)
2. SMP : SMP Negeri 2 Karanganom (1983 – 1986)
3. SMA : SMA Negeri Karanganom (ex. SMA Negeri Jatinom, 1986-1989)
4. Perguruan tinggi :

Jenjang	Sarjana	Magister	Doktor
Nama Perguruan Tinggi	UGM	UGM	UGM
Bidang Ilmu	Statistika	Statistika	Statistika
Tahun	1989-1994	1996-1999	2006-2010
Judul Skripsi/Tesis/Disertasi	Model Regresi Logistik	Estimasi Parameter Regresi Logistik Biner Bivariate	Model Pemilihan Diskrit Menggunakan Model Probit dan Model Mixed Logit Pada Respon Multivariat
Nama Pembimbing/Promotor	Drs. Zulaila, M.Si.; Dra. Sri Pangesti, SU.	Prof. Drs. Suryo Guritno, M.Stat. Ph.D.	Prof. Drs. Suryo Guritno, M.Stat. Ph.D.

C. Keluarga

1. Istri : Prof. Dr. Is Fatimah, S.Si., M.Si.
2. Anak : (1) Naufal Abdillah
(2) Haikal Abdurrahman
(3) Inaya Izaturrizqi

D. Pengalaman Jabatan Struktural

No	Jabatan	Tahun
1	Ketua Program Studi/Jurusan Statistika UII	1999 - 2001
2	Dekan FMIPA UII	2001 - 2006
3	Kepala Bidang Standar Mutu Akademik UII	2010 - 2016
4	Kepala Badan Pengembangan Akademik UII	2016 - 2018
5	Direktur Direktorat Pengembangan Akademik UII	2018 - sekarang

E. Perolehan Hibah Penelitian dalam 10 Tahun Terakhir

No	Tahun	Judul	Jenis
1	2013-2014	Model Indeks Kenyamanan Termal Termo Adaptif Psikologis PMVtap dengan Pendekatan SEM untuk Penetapan Standar Termal Ruang Ber AC Hemat Energi	Hibah Fundamental

2	2015-2016	Pengembangan Instrumen Indeks Kapasitas Masyarakat dalam Mitigasi Bencana	Penelitian Unggulan PT
3	2018	Perancangan Aplikasi Perpustakaan Berbasis Android dengan Metode <i>Rough Set</i>	Penelitian Madya UII
4	2019	Evaluasi Faktor yang Mempengaruhi Mahasiswa untuk Memilih Program Studi di FMIPA Beserta Tingkat Kepuasannya	Penelitian Institusi UII
5	2019	Modifikasi Model Log-linear untuk Analisis Tabel Kontingensi Multi-Arah pada <i>Big Data</i>	Penelitian Dasar Unggulan PT
6	2019	Kajian Preparasi dan Unjuk Kerja Nanostruktur SnO ₂ /Graphene-Like Materials Berbahan Dasar Biomassa untuk Aplikasi Sensor Gas	Penelitian Dasar
7	2020	Identifikasi Sifat Beberapa Penaksir Parameter Proporsi Pada Distribusi Binomial Satu Populasi	Penelitian Unggulan-UII

F. Publikasi Artikel Ilmiah

1. Internasional

No	Judul Artikel Ilmiah	Volume/No /Tahun	Nama Jurnal/Prosiding
1	Mixed Logit Model on Multivariate Binary Response Using Maximum Likelihood Estimator and Generalized Estimating Equations”,	Vol 4 (3) : 109-123. , 2011	Asian Journal of Mathematics and Statistics. DOI:10.3923/ajm s.2011
2	Application of Logit Model Toward The Study on The Effect of Catalyst Dosage to Methylene Blue Photodegradation	Volume: 11 No: 02,2011	International Journal of Electrical & Computer Sciences
3	Evaluation Of Photodegradation Efficiency On Semiconductor Immobilized Clay Photocatalyst By Using Probit Model Approximation	Volume 4, Issue 2, June 2013, Pages 125–130	International Journal of Chemical and Analytical Science. https://doi.org/10.1016/j.ijcas.2013.07.005
4	Comparison Of Logit Model And Probit Model On Multivariate Binary Response	IICMA 2013 Halaman 294-302	ISBN : 978-602-96426-2-9. Published by Indonesian Mathematical Society

No	Judul Artikel Ilmiah	Volume/No /Tahun	Nama Jurnal/Prosiding
5	Energy Consumption and Thermal Comfort favored by the Occupants in the Air Conditioned House	ICSBE 2014 page 76-83	The 3rd International Conference on Sustainable Built Environment ISBN:978-602-98397-4-6.
6	Random Effect Model and Generalized Estimating Equations for Binary Panel Responsel",	ICRIEMS 2014	Prosiding International Conference On Research, Implementation, And Education Of Mathematics And Sciences, Yogyakarta, Mei 2014
7	Modeling of Photocatalytic Activity of ZnO/AC By Using Linear Probability Model, Logit and Complementary Log Transformation	Volume 148 (2016) 1112 – 1120	Procedia Engineering https://doi.org/10.1016/j.proeng.2016.06.613
8	Control of Wastewater Using Multivariate Control Chart	1823,02012 6 (2017)	AIP Conference Proceedings https://doi.org/10.1063/1.4978199
9	Thermo-adaptive-Psychological Thermal Comfort Index of	Book Chapter p.237-250	Springer Nature Singapore Pte Ltd. 2018. Book

No	Judul Artikel Ilmiah	Volume/No /Tahun	Nama Jurnal/Prosiding
	PMVtapsem Development of a PMVtap Index Based on the SEM Approach .		Chapter : “Sustainable Future for Human Security :Society, Cities and Governance”, doi :10.1088/1742- 6596/909/1/0120 74 .
10	"Preparation, characterization, and modelling activity of potassium fluoride modified hydrotalcite for microwave assisted biodiesel conversion”,	Vol. 8 (June 2018) 63– 70	Sustainable Chemistry and Pharmacy https://doi.org/10.1016/j.scp.2018.02.004
11	Brown’s Weighted Exponential Moving Average (Bwema) With Levenberg- Marquardt Optimization To Forecasting Rate Of Return	Vol 8 (2018) p.1744- 1749	The Turkish Online Journal of Design, Art and Communication
12	Modelling on human immunodeficiency virus case using Poisson bivariate regression	2026, 020100 (2018)	AIP Conference Proceedings https://doi.org/10.1063/1.5065060

No	Judul Artikel Ilmiah	Volume/No /Tahun	Nama Jurnal/Prosiding
13	Pattern Analysis on Staff of Work Accident Handling Using Chi-Squared Automatic Interaction Detection and Log-Linear Models	2026, 020060 (2018);	AIP Conference Proceedings https://doi.org/10.1063/1.5065020
14	Sentiment Analysis on Mobile Banking Application Using Naive Bayes Classifier and Association Methods	7 (4.15) (2018) 244-247	International Journal of Engineering & Technology
15	Application of Geographically Weighted Regression Analysis for Modelling Pneumonia Toddlers in West Java Province	ICMM-2018	Conference Proceedings : Ahmad Dahlan International Conference on Mathematics and Mathematics (2018)
16	Clustering And Diversifying Student Spreads Based On Environmental Effects Assisted By Webgis Visualization	ICoSMEE-2019	Prosiding The 2nd International Conference on Science, Mathematics, Environment, and Education

No	Judul Artikel Ilmiah	Volume/No /Tahun	Nama Jurnal/Prosiding
17	Performance Analysis of Mixed Logit Models for Discrete Choice Models	Vol. 15 No. 3, (2019) 563-575	Pakistan Journal of Statistics and Operation Research. DOI: http://dx.doi.org/10.18187/pjsor.v15i3.2313
18	Physicochemical characteristics and photocatalytic performance of Tin oxide/montmorillonite nanocomposites at various Sn/montmorillonite molar to mass ratios	Vol. 193 (2020) 105671	Applied Clay Science https://doi.org/10.1016/j.clay.2020.105671
19	The Influence of Religiosity Toward Universitas Islam Indonesia Students Nationalism Using Structural Equation Modeling and Loglinear Model	Vol. 397. February 2020.	Advances in Social Science, Education and Humanities Research. Atlantis Press. DOI: https://doi.org/10.2991/assehr.k.200129.147 .
20	Designing Android-Based Library Application with Rough Set Method	Vol. 397. February 2020.	Advances in Social Science, Education and

No	Judul Artikel Ilmiah	Volume/No /Tahun	Nama Jurnal/Prosiding
			Humanities Research. Atlantis Press. DOI: https://doi.org/10.2991/assehr.k.200129.120
21	Linear regression model and spatial autoregressive model for modeling high school dropout",	Vol :2229, Issue 1. April 2020	AIP Conference Proceedings, https://doi.org/10.1063/5.0002420
22	Tobit Modeling And Central Composite Design For Optimization Of Specific Surface Area Of Clay By Activation	Vol: 1983 Issue (2020) 105671	Journal of Engineering Science and Technology, Scopus & Scimagojr, SJR= 0.24
23	Implementation of Naive Bayes Classification Method for Sentiment Analysis on Community Opinion to Indonesian Criminal Code Draft	Volume 474 2020	Advances in Social Science, Education and Humanities Research,

2. Nasional

No	Judul Artikel Ilmiah	Volume/No /Tahun	Nama Jurnal
1	Identifikasi Hasil Pirolisis Serbuk Kayu Jati Menggunakan Principal 7 Component Analysis	Vo. 6. No. 1, 2005,	<i>Jurnal Ilmu Dasar</i> , , FMIPA Universitas Jember. ISSN 1411-5735
2	Analisis Hubungan Kuantitatif Struktur dan Kelarutan Senyawa Aktif Pestisida Organofosfat : Pendekatan Model Linear dan Metode Kluster	Vo. 8. No. 1, 2007	<i>Jurnal Ilmu Dasar</i> , , FMIPA Universitas Jember. ISSN 1411-5735
3	Menghitung Nilai Probabilitas Pada Distribusi Normal <i>Multivariate</i>	Vol. 3, No. 2 (2007),	<i>Pythagoras, Jurnal Matematika</i> Jurusan Matematika UNY. ISSN : 1978-4538
4	Model Probit dan Model Logit pada Respon Biner	Vol 10. No. 1. (2008),	<i>Eksakta, Jurnal Ilmu-ilmu MIPA</i> FMIPA UII Yogyakarta,

No	Judul Artikel Ilmiah	Volume/No /Tahun	Nama Jurnal
			ISSN : 1411-1047
5	Model Respon Multinomial Saling Berkorelasi Menggunakan <i>Generalized Extreme Value</i> (GEV)	Vol. 11 No. 1 - June-2008	Jurnal Ilmu MIPA, Fak. Sains dan Teknologi-UNAIR
6	Uji <i>Goodness Of Fit</i> Pada Model Pemilihan Diskrit	Vol 11. No. 2 (2010)	<i>Eksakta, Jurnal Ilmu-ilmu MIPA</i> FMIPA UII Yogyakarta, ISSN : 1411-1047
7	Pengaruh Korelasi Antar Respon pada Model Multinomial Logit	Vol. 14 No. 3 - Agustus-2009	<i>Jurnal Matematika dan Sains (JMS)</i> FMIPA-ITB.
8	Estimasi Parameter Model Logit Pada Respons Biner Multivariat Menggunakan Metode MLE dan GEE	Vol. 10. No. 1 , 2009	<i>Jurnal Ilmu Dasar,</i> FMIPA Universitas Jember. ISSN: 1411-5735

No	Judul Artikel Ilmiah	Volume/No /Tahun	Nama Jurnal
9	Model Probit Pada Respons Biner Multivariat Menggunakan SMLE	Vol. 11. No. 1 ,2010	<i>Jurnal Ilmu Dasar</i> , , FMIPA Universitas Jember. ISSN: 1411-5735
10	Model Probit pada Respons Biner Multivariat Menggunakan <i>Simulated Maximum Likelihood Estimator</i>	Vol. 11 No. 1 Januari 2010, 70-75	<i>Eksakta, Jurnal Ilmu-ilmu MIPA</i> FMIPA UII Yogyakarta, ISSN : 1411-1047
11	<i>Parameter Estimation in Probit Model for Multivariate Multinomial Response Using SMLE</i>	Vol. 12 No. 1, 44-48 (2011)	<i>Eksakta, Jurnal Ilmu-ilmu MIPA</i> FMIPA UII Yogyakarta, ISSN : 1411-1047
12	Studi Simulasi Model Nested Logit dan Paired Combinatorial Logit pada Respon Multinomial	Vol. 13 No. 1-2. Agustus 2013, 63-71	<i>Eksakta, Jurnal Ilmu-ilmu MIPA</i> FMIPA UII Yogyakarta,

No	Judul Artikel Ilmiah	Volume/No /Tahun	Nama Jurnal
			ISSN : 1411-1047
13	Model Kapasitas Masyarakat Dalam Menghadapi Bencana Menggunakan Analisis Regresi Logistik Ordinal	Vol. 16 No. 1 Februari 2016	<i>Eksakta, Jurnal Ilmu-ilmu MIPA FMIPA UII Yogyakarta,</i> ISSN : 1411-10472018
14	Perbandingan Rough Set dan Algoritma Apriori untuk Sistem Rekomendasi Perpustakaan	4 (2), 25-31(2018)	<i>Unisda Journal of Mathematics and Computer Science (UJMC)</i>
15	Analysis of Factors Affecting the Human Development Index in West Kalimantan Province using Data Panel Data Regression	Vol. 18, ISSUE 2, August 2018.	<i>Jurnal Eksakta UII</i> DOI=10.20885/eksakta.voll8.iss2.art2
16	Study of Student Satisfaction Level in the Faculty Based on Performance	Vol 19 Issue 1. 83-97. Februari 2019	<i>EKSAKTA: Journal of Sciences and Data Analysis.</i>

No	Judul Artikel Ilmiah	Volume/No /Tahun	Nama Jurnal
	Assessment and Interest Level		DOI: 10.20885/e ksakta.vol1 9.iss1.art8
17	Analysis of Factors Influencing The Decision to Choose The Department in The Natural Science Campus	Vol. 1, ISSUE 1, Februari 2020. Hal 81-84	<i>EKSAKTA: Journal of Sciences and Data Analysis.</i> DOI: 10.20885/E KSAKTA.v ol1.iss1.art1 2

G. Penulisan Buku 10 Tahun Terakhir

No	Judul Buku	ISBN	Tahun
1	Pengantar Analisis Data Kategorik	978-602-280-095-8	2013
2	Metode Maksimum Likelihood dalam Model Pemilihan Diskrit	978-602-450-019-1	2017
3	Pemodelan Data Ordinal, Nominal dan Cacah	978-602-450-163-1	2017
4	Model Logit, Probit dan <i>Mixed</i> Logit pada Respon Multivariat	978-602-450-200-3	2017

No	Judul Buku	ISBN	Tahun
5	Implementasi metode kolaboratif dan bermain peran dalam pembelajaran di UII	978-602-450-322-2	2018
7	Book Chapter "Thermo-adaptive-Psychological Thermal Comfort Index of PMVtapsem Development of a PMVtap Index Based on the SEM Approach"	978-981-10-5432-7	2018
8	Analisis Tabel Kontingensi Menggunakan Model Log-linear	978-602-450-437-3	2019
9	Pengantar Peluang dan Distribusi	978-623-020-744-0	2020

H. Pengalaman Perolehan HKI

No	Judul/Tema HKI	Tahun	Jenis	Nomor P/ID
1	Buku dengan judul "Metode Maksimum Likelihood dalam Model Pemilihan Diskrit"	2018	Hak Cipta	EC000121516

No	Judul/Tema HKI	Tahun	Jenis	Nomor P/ID
2	Buku dengan judul "Pemodelan Data Ordinal, Nominal dan Cacah"	2018	Hak Cipta	EC000102419
3	Buku dengan judul "Model Logit, Probit dan Mixed Logit pada Respon Multivariat "	2018	Hak Cipta	EC00201850233
4	Buku dengan judul "Pengantar analisis Data Kategorik"	2018	Hak Cipta	EC00201850236
5	Buku dengan judul "Analisis Tabel Kontingensi Menggunakan Model Loglinear "	2020	Hak Cipta	EC00202039461
6	Buku dengan judul "Pengantar Peluang dan Distribusi "	2020	Hak Cipta	EC00202055264

I. Penghargaan yang Pernah Diraih dalam 10 tahun Terakhir (dari Pemerintah, Asosiasi atau Institusi Lainnya)

No	Jenis Penghargaan	Institusi Pemberi Penghargaan	Tahun
1	Dosen Berprestasi Peringkat 1	UII	2020

Demikian biodata ini saya buat dengan sebenarnya

Yogyakarta, 15 Januari 2022

Hormat Saya,

(Prof. Dr. Jaka Nugraha, S.Si, M.Si.)