Studi Komparasi Performa Algoritma *Supervised-Learning* dalam Klasifikasi Multidataset

I Gusti Ngurah Ryo Aditarta
Fakultas Ilmu Komputer
Universitas Brawijaya
Malang, Indonesia
ryoaditarta@student.ub.ac.id

2nd Dian Pandu Syahfitra
Fakultas Ilmu Komputer
Universitas Brawijaya
Malang, Indonesia
dianpandu@student.ub.ac.id

3rd Rayhan Egar Sadtya Nugraha
Fakultas Ilmu Komputer
Universitas Brawijaya
Malang, Indonesia
rayhanegar@student.ub.ac.id

4th Ahmad Zaki
Fakultas Ilmu Komputer
Universitas Brawijaya
Malang, Indonesia
ahmadzaki12@student.ub.ac.id

5th Muhammad Arya Ghifari
Fakultas Ilmu Komputer
Universitas Brawijaya
Malang, Indonesia
aryaghifary07@student.ub.ac.id

6th Arion Syemael Siahaan
Fakultas Ilmu Komputer
Universitas Brawijaya
Malang, Indonesia
siahaanarion@student.ub.ac.id

***Abstract*—This document is a model and instructions for LaTeX. This and the IEEEtran.cls file define the components of your paper [title, text, heads, etc.]. *CRITICAL: Do Not Use Symbols, Special Characters, Footnotes, or Math in Paper Title or Abstract.**

***Index Terms*—component, formatting, style, styling, insert**

I. Introduction

This document is a model and instructions for LaTeX. Please observe the conference page limits.

II. Algoritma Klasifikasi Supervised Learning

A. K-Nearest Neighbors (KNN)

Algoritma klasifikasi *nearest neighbors* seperti KNN merupakan salah satu algoritma dasar nonparametrik dalam pembelajaran mesin. Hal ini didasarkan pada *rationale* di mana fitur yang digunakan untuk mendeskripsikan label sebuah *domain point* memiliki relevansi dengan *domain point* lain dalam sisi *proximity* [4], [5]. Dalam implementasinya, KNN banyak digunakan dalam permasalahan klasifikasi dengan domain pengetahuan yang terdefinisi dan diketahui dengan baik, seperti klasifikasi tumor otak [6], klasifikasi tingkat keparahan Covid-19 [7], maupun prognosis kanker [8].

Secara prinsip, nilai parameter k yang merepresentasikan banyaknya *domain point* yang digunakan dalam penentuan label serta metode perhitungan jarak merupakan konsiderasi utama. Nilai k yang terlalu kecil memberikan model yang kompleks dan kurang mampu mengakomodasi *unseen data*, sedangkan nilai k yang terlalu besar memberikan model yang terlalu sederhana untuk secara akurat mengklasifikasikan suatu *domain point* [9]. Beberapa teknik, seperti *normalized class coherence*, *change-based KNN*, *variable selection* dan *weighting*, maupun normalisasi L1 dan LPP dapat digunakan untuk memberikan estimasi yang lebih baik [10]–[12]. Selain itu, pilihan perhitungan jarak antara domain point dalam proses pembelajaran model seperti Euclidean "(1)", Manhattan "(2)", maupun Minkowski "(3)" untuk pemetaan label kelas juga perlu untuk diperhatikan.

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(y_i - x_i)^2} \tag{1}$$

$$d(x,y) = \sum_{i=1}^{n}|x_i - y_i| \tag{2}$$

$$d(x,y) = (\sum_{i=1}^{n}|x_i - y_i|^p)^{\frac{1}{4}} \tag{3}$$

B. Ridge Classifier (RC)

Algoritma klasifikasi *Ridge Classifier* (RC) dikembangkan dari algoritma *Ridge Regression* yang mengkombinasikan *learning rule Regularized Learning Minimization* (RLM) dengan regresi *linear ordinary least squares* [1]. Algoritma RC melakukan *class labelling* berdasarkan tanda/*sign* dari suatu *data point* (positif atau negatif). Penggunaan fungsi regularisasi, seperti regularisasi L2/Tikhonov "(4)" "(5)" pada algoritma klasifikasi memberikan model yang lebih "stabil" dan mencegah terjadinya *overfitting* [4], [5]. Dibandingkan dengan regresi linear, algoritma RC akan meminimalisasi nilai koefisien w untuk masing-masing fitur, memberikan model dengan kemampuan generalisasi yang baik terhadap *unseen data*. Dengan demikian, fungsi yang digunakan dalam RC dapat didefinisikan secara formal pada "(6)".

$$||w|| = \sqrt{\sum_{i=1}^{d} w_i^2} \tag{4}$$

$$\arg min_w(L_s(w) + \lambda||w||^2) \tag{5}$$

$$\arg min_{w \in R^d}(\lambda||w||^2 + \frac{1}{m}\sum_{i=1}^{m}(< w, x_i > -y_i)^2) \quad (6)$$

Dalam melakukan proses pembelajaran model, parameter regularisasi menjadi konsiderasi utama. Semakin kecil nilai, nilai koefisien w akan semakin kecil, memberikan model dengan generalisasi yang baik namun dengan potensi *underfitting*. Nilai α yang semakin besar akan meminimalisasi efek regularisasi, memberikan model dengan kompleksitas yang lebih tinggi namun dengan potensi *overfitting* untuk *unseen data*. Selain dengan menggunakan *cross-validation* untuk *hyperparameter tuning*, teknik seperti *fractional ridge regression* [13] dengan memanfaatkan rasio L2-norm antara *regularized* dan *normal coefficients* dapat membantu menemukan nilai α yang optimal.

C. Logistic Regression (LR)

Logistic Regression (LR) merupakan algoritma klasifikasi yang didefinisikan secara formal sebagai komposisi fungsi sigmoid "(7)"terhadap suatu fungsi regresi linear untuk membuat kelas hipotesis "(8)" [4]. Algoritma LR, bersama dengan RC, digunakan dalam kasus klasifikasi dengan memperhatikan tanda/*sign* suatu *domain point* [9]. LR memanfaatkan regularisasi L2/*ridge* dengan koefisien regularisasi α, di mana semakin rendah nilai α, model yang dihasilkan akan lebih sederhana dengan koefisien *w* yang semakin kecil. Sebagai *weak learner*, *ensembling* dari model LR dengan menggunakan AdaBoost dapat membantu meningkatkan performa generalisasi model dengan proses iteratif. Studi [14] menunjukkan jika penggunaan metode *robust functional principal component analysis* (RFPCA) memberikan dampak positif pada performa klasifikasi model LR. Dengan demikian, *learning rule* algoritma klasifikasi LR dapat secara formal didefinisikan dalam bentuk "(9)"

$$\sigma_{sig}(z) = \frac{1}{1 + \exp(-z)} \quad (7)$$

$$H_{sig} = \sigma_{sig} \circ L_d \quad (8)$$

$$\arg min_{w \in R^d} \frac{1}{m}\sum_{i=1}^{m} \log(1 + \exp(-y_i < w, x_i >)) \quad (9)$$

D. Decision Tree (DT)

Decision Tree (DT) merupakan algoritma klasifikasi yang terdiri dari himpunan pertanyaan *if-else* sebagai *splitting criteria* yang dibangun secara iteratif dengan pendekatan *top-down* untuk memisahkan suatu kelas dari kelas lainnya. Setiap node yang terbentuk pada DT memiliki nilai entropi "(10)" yang merepresentasikan rata-rata informasi yang diperlukan untuk melakukan separasi label kelas [15]. Fitur yang dipilih sebagai *splitting criteria* sebuah node untuk menghasilkan *child node* didasarkan atas *gain measure* yang berbeda-beda untuk setiap algoritma DT [4]. Algoritma ID3 mengimplementasikan *Information Gain* (IG) "(11)" sebagai *gain measure*, sedangkan algoritma C4.5 mengimplementasikan Gain Ratio (GR) "(12)" [16]. Pemilihan hyperparameter DT yang optimal untuk suatu dataset, seperti *maximum depth* (MD), *maximum leaf nodes* dan *minimum samples* dalam pembentukan *leaf nodes* mampu memberikan model DT dengan performa generalisasi yang baik dan mencegah *overfitting* akibat *tree size* yang terlalu besar [17].

$$H(S) = \sum_{i=1}^{C} p_i \log_2(p_i) \quad (10)$$

$$IG(S, A) = H(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} H(S_v) \quad (11)$$

$$GR(S, A) = \frac{IG(S, A)}{SI(S, A)} \quad (12)$$

$$SI(S, A) = -\sum_{v \in Values(A)} \frac{|S_v|}{|S|} \log_2(\frac{|S_v|}{|S|}) \quad (13)$$

E. Naive-Bayes (NB)

Algoritma Naive-Bayes adalah salah satu algoritma pembelajaran mesin yang populer untuk tugas klasifikasi. Algoritma ini didasarkan pada teorema Bayes dan beroperasi dengan asumsi "*naive*" bahwa semua fitur yang ada bersifat independen satu sama lain. Asumsi ini menyederhanakan perhitungan probabilitas, menjadikan algoritma ini sangat efisien meskipun dalam kenyataannya fitur-fitur tersebut mungkin tidak sepenuhnya independen [19]. Dalam konteks data fungsional, Naive-Bayes digunakan untuk mengklasifikasikan objek berdasarkan data pelatihan dengan menggunakan "*surrogate densities*" yang diturunkan dari skor *Functional Common Principal Component* (FCPC) [18].

Algoritma Naive-Bayes juga menghadapi tantangan dalam ruang berdimensi tinggi, di mana fungsi kepadatan probabilitas seringkali tidak ada sehingga pendekatan densitas klasik tidak dapat digunakan. Untuk mengatasi masalah ini, asumsi *naive* diterapkan pada skor FCPC yang memungkinkan definisi densitas dari data fungsional [18]. Studi simulasi dan aplikasi pada data nyata menunjukkan bahwa Naive-Bayes sering memberikan performa yang kompetitif dibandingkan dengan algoritma klasifikasi lainnya seperti regresi logistik multinomial, k-NN, analisis diskriminan linear, dan mesin vektor pendukung, terutama ketika jumlah komponen meningkat [18].

- Probabilitas posterior

$$P(C_k|x) = \frac{P(x|C_k) \cdot P(C_k)}{P(x)} \quad (14)$$

- Probabilitas kondisional

$$P(x|C_k) = \prod_{i=1}^{n} P(x_i|C_k) \quad (15)$$

- Estimasi probabilitas dengan distribusi normal (Gaussian Naive-Bayes)

$$C_k = \arg\max_{C \in C} P(C|x) = \arg\max_{C \in C} P(C) \cdot \prod_{i=1}^{n} P(x_i|C) \tag{16}$$

- Klasifikasi dengan Naive-Bayes

$$P(x_i|C_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} \cdot \exp\left(\frac{-(x_i - \mu_k)^2}{2\sigma_k^2}\right) \tag{17}$$

F. Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah algoritma pembelajaran mesin yang digunakan untuk klasifikasi dan regresi [21]. Algoritma ini bekerja dengan mencari *hyperplane* terbaik yang memisahkan data ke dalam kelas yang berbeda [20]. *Hyperplane* adalah batas keputusan yang memisahkan set data dengan label yang berbeda; dalam ruang dua dimensi, *hyperplane* adalah garis; dalam ruang tiga dimensi, *hyperplane* adalah bidang; dan dalam dimensi yang lebih tinggi, *hyperplane* adalah objek dengan dimensi lebih tinggi yang memisahkan data [22].

Pada dasarnya, SVM bertujuan untuk menemukan *hyperplane* yang memaksimalkan margin, yaitu jarak antara *hyperplane* dan titik data terdekat dari setiap kelas [21]. Untuk data yang tidak dapat dipisahkan secara linear, SVM menggunakan fungsi kernel untuk memetakan data ke dimensi yang lebih tinggi di mana data tersebut dapat dipisahkan secara linear [22]. Beberapa fungsi kernel yang umum digunakan adalah kernel linear, kernel polinomial, dan *Radial Basis Function* (RBF) [20].

Rumus penting dalam SVM melibatkan fungsi objektif yang bertujuan meminimalkan norma vektor bobot, dengan syarat bahwa data dapat dipisahkan dengan margin yang maksimal [21]. Untuk data yang tidak dapat dipisahkan secara sempurna, variabel *slack* digunakan untuk mengatasi kasus-kasus ini [22]. Implementasi SVM melibatkan pemilihan fungsi kernel yang sesuai dan penentuan parameter model yang optimal, serta memecahkan masalah optimasi untuk mendapatkan *hyperplane* yang dapat memprediksi label data baru [20].

- Fungsi Linear SVM

$$f(x) = w^T x + b \tag{18}$$

- Margin optimal

$$Margin = \frac{2}{||w||^2} \tag{19}$$

- Fungsi Objektif untuk SVM (data yang dapat dipisahkan secara linear)

$$\min_{w,b} \frac{1}{2}||w||^2 \tag{20}$$

- Fungsi Objektif untuk SVM (data yang tidak dapat dipisahkan secara linear)

$$\min_{w,b,\xi} \frac{1}{2}||w||^2 + C\sum_{i=1}^{n} \xi_i \tag{21}$$

- Fungsi Kernel

$$K(x_i, x_j) = x_i^T x_j \tag{22}$$

- Kernel Linear

$$K(x_i, x_j) = (x_i^T x_j + c)^d \tag{23}$$

- Kernel RBF

$$K(x_i, x_j) = \exp(-||x_i - x_j||^2/\gamma) \tag{24}$$

G. Random Forest (RF)

Algoritma *Random Forest* (RF) adalah metode pembelajaran *ensemble* yang beroperasi dengan membangun banyak pohon keputusan selama pelatihan dan menghasilkan mode dari kelas (klasifikasi) atau rata-rata prediksi (regresi) dari masing-masing pohon. Dalam konteks pembelajaran *semi-supervised*, algoritma *Co-Forest* memperluas pendekatan Random Forest dengan menggunakan beberapa *classifier* untuk menangani contoh-contoh yang tidak berlabel [23]. Algoritma ini secara iteratif menyempurnakan setiap *classifier* dengan contoh-contoh baru yang diberi label, yang dipilih berdasarkan kepercayaan *classifier* lain dalam *ensemble* [23]. Hasil eksperimen menunjukkan bahwa *Co-Forest* meningkatkan kinerja, terutama ketika proporsi data berlabel rendah, seperti pada dataset biologis [23]. Dengan menggunakan contoh yang tidak berlabel untuk meningkatkan pembelajaran dari sampel yang diberi label, *Co-Forest* menunjukkan peningkatan rata-rata sebesar 3,6% pada kondisi dengan 60% data tidak berlabel [23].

Pendekatan lain yang diperkenalkan adalah *Confidence weighted Random Forest* (CwRF), yang menambahkan skor kepercayaan untuk setiap node daun dalam pohon keputusan [24]. Skor kepercayaan ini digunakan untuk memberi bobot pada suara dari pohon-pohon tersebut, memberikan pengaruh lebih besar pada pohon yang membuat prediksi dengan lebih percaya diri [24]. Kepercayaan dihitung berdasarkan metrik *impurity* seperti entropi dan indeks Gini [24]. Skor kepercayaan ini kemudian digunakan untuk menimbang probabilitas kelas dari setiap pohon selama fase pengujian [24]. Hasil eksperimen menunjukkan bahwa CwRF secara konsisten mengungguli RF tradisional dan metode canggih lainnya pada berbagai dataset, menunjukkan efektivitasnya dalam meningkatkan proses pengambilan keputusan secara keseluruhan [24]. Algoritma ini terbukti efektif di berbagai aplikasi, memperkuat potensinya untuk diterapkan secara lebih luas [24].

- *GINI Impurity* (GI)

$$GI(S) = 1 - \sum_{i=1}^{C} \frac{|S_i|^2}{|S|} \tag{25}$$

- *Information Gain* (IG)

$$IG(S, A) = GI(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} GI(S_v) \tag{26}$$

- *Out-of-Bag Error*

$$OOB_Error = \frac{1}{n}\sum_{i=1}^{n} I(y_i \neq f_i(x_i)) \tag{27}$$

H. Extreme Gradient Boosting (XGBoost)

XGBoost (*eXtreme Gradient Boosting*) adalah algoritma pembelajaran mesin berbasis *boosting gradient* yang sangat skalabel [25]. Algoritma ini sering digunakan dalam berbagai kompetisi pembelajaran mesin karena kecepatan pelatihan dan kinerja generalisasinya yang unggul [26]. XGBoost bekerja dengan menggabungkan beberapa model pembelajaran lemah secara iteratif untuk membentuk model yang lebih kuat [26]. Algoritma ini membangun model *ensemble* dari pohon-pohon keputusan menggunakan fungsi aditif untuk meminimalkan fungsi objektif yang teratur [26]. Fungsi objektif dalam XGBoost menggabungkan fungsi *loss* dan penalti untuk menghindari *overfitting*, di mana fungsi *loss* dan regularisasi ini membantu mengontrol kompleksitas model [26]. Parameter penting dalam XGBoost termasuk laju pembelajaran, gamma, kedalaman maksimum, dan subsampling, yang semuanya digunakan untuk mengoptimalkan kinerja model [26].

Untuk mengoptimalkan fungsi objektif, XGBoost menggunakan pendekatan orde kedua, di mana statistik gradien pertama dan kedua pada fungsi *loss* digunakan [26]. Skor untuk pemilihan split dihitung berdasarkan jumlah gradien pertama dan kedua, sementara bobot daun dioptimalkan untuk meminimalkan *loss* [26]. Dengan teknik ini, XGBoost mampu menangani dataset besar dengan efisiensi tinggi dan memberikan kinerja prediksi yang unggul [3]. Algoritma ini juga menggabungkan berbagai teknik seperti regularisasi, subsampling, dan optimasi berbasis cache untuk meningkatkan kecepatan pelatihan dan mengurangi *overfitting* [26]. Keseluruhan, XGBoost adalah algoritma yang sangat efektif dalam pembelajaran mesin dan telah terbukti unggul dalam berbagai tugas klasifikasi dan regresi [25]. XGBoost sangat diakui karena kemampuannya dalam menangani data dalam skala besar dan kompleks, serta memberikan hasil yang sangat akurat dalam berbagai kompetisi pembelajaran mesin [25] [26].

- *Objective function*

$$Obj(t) = \sum_{i=1}^{n} L(t,i) + f(t) \tag{28}$$

- *Regression loss*

$$L(t,i) = \frac{1}{2}\left(y_i - f_t(x_i)\right)^2 \tag{29}$$

- Penalti Regularisasi L1

$$f(t) = \lambda \sum_{j=1}^{K} |w_j| \tag{30}$$

- Penalti Regularisasi L2

$$f(t) = \lambda \sum_{j=1}^{K} w_j^2 \tag{31}$$

- *Gradient*

$$g_i = \partial_{\hat{y}_i} l(y_i, \hat{y}_i) \tag{32}$$

- *Hessian*

$$h_i = \partial^2_{\hat{y}_i} l(y_i, \hat{y}_i) \tag{33}$$

- Formula pembaruan bobot daun

$$w_j^* = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \tag{34}$$

I. Categorical Boosting (CatBoost)

Algoritma *machine learning* ini digunakan untuk menangani fitur kategorikal dan juga lebih cepat dibandingkan dengan algoritma penguat lainnya karena mengimplementasikan pohon simetris. CatBoost, yang merupakan implementasi dari Gradient Boosting on Decision Tree (GDBT), memiliki kombinasi gradient boosting dengan pohon keputusan yang memberikan hasil yang baik [28].

CatBoost dirancang untuk menangani data kategorikal secara efisien dengan mengkonversinya menjadi fitur numerik secara otomatis. Proses pelatihannya melibatkan pembuatan sejumlah model pohon keputusan secara berurutan, di mana setiap model baru berusaha untuk mengurangi kesalahan dari model sebelumnya. Teknik boosting ini memungkinkan CatBoost untuk memperbaiki kesalahan prediksi secara iteratif. Selain itu, CatBoost menggunakan pohon simetris yang mempercepat waktu prediksi dan pelatihan. Algoritma ini juga memanfaatkan regularisasi untuk menghindari overfitting dan meningkatkan generalisasi model [29].

J. Adaptive Boosting (AdaBoost)

K. Light Gradient Boosting Machine (LGBM)

L. Ensemble (Stacking)

III. Prepare Your Paper Before Styling

Before you begin to format your paper, first write and save the content as a separate text file. Complete all content and organizational editing before formatting. Please note sections III-A–III-E below for more information on proofreading, spelling and grammar.

Keep your text and graphic files separate until after the text has been formatted and styled. Do not number text heads—LaTeX will do that for you.

A. Abbreviations and Acronyms

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, ac, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

B. Units

- Use either SI (MKS) or CGS as primary units. (SI units are encouraged.) English units may be used as secondary units (in parentheses). An exception would be the use of English units as identifiers in trade, such as “3.5-inch disk drive”.
- Avoid combining SI and CGS units, such as current in amperes and magnetic field in oersteds. This often

leads to confusion because equations do not balance dimensionally. If you must use mixed units, clearly state the units for each quantity that you use in an equation.
- Do not mix complete spellings and abbreviations of units: "Wb/m^2" or "webers per square meter", not "webers/m^2". Spell out units when they appear in text: ". . . a few henries", not ". . . a few H".
- Use a zero before decimal points: "0.25", not ".25". Use "cm^3", not "cc".)

C. Equations

Number equations consecutively. To make your equations more compact, you may use the solidus (/), the exp function, or appropriate exponents. Italicize Roman symbols for quantities and variables, but not Greek symbols. Use a long dash rather than a hyphen for a minus sign. Punctuate equations with commas or periods when they are part of a sentence, as in:

$$a + b = \gamma \tag{35}$$

Be sure that the symbols in your equation have been defined before or immediately following the equation. Use "(35)", not "Eq. (35)" or "equation (35)", except at the beginning of a sentence: "Equation (35) is . . ."

D. LaTeX-Specific Advice

Please use "soft" (e.g., `\eqref{Eq}`) cross references instead of "hard" references (e.g., `(1)`). That will make it possible to combine sections, add equations, or change the order of figures or citations without having to go through the file line by line.

Please don't use the `{eqnarray}` equation environment. Use `{align}` or `{IEEEeqnarray}` instead. The `{eqnarray}` environment leaves unsightly spaces around relation symbols.

Please note that the `{subequations}` environment in LaTeX will increment the main equation counter even when there are no equation numbers displayed. If you forget that, you might write an article in which the equation numbers skip from (17) to (20), causing the copy editors to wonder if you've discovered a new method of counting.

BibTeX does not work by magic. It doesn't get the bibliographic data from thin air but from .bib files. If you use BibTeX to produce a bibliography you must send the .bib files.

LaTeX can't read your mind. If you assign the same label to a subsubsection and a table, you might find that Table I has been cross referenced as Table IV-B3.

LaTeX does not have precognitive abilities. If you put a `\label` command before the command that updates the counter it's supposed to be using, the label will pick up the last counter to be cross referenced instead. In particular, a `\label` command should not go before the caption of a figure or a table.

Do not use `\nonumber` inside the `{array}` environment. It will not stop equation numbers inside `{array}` (there won't be any anyway) and it might stop a wanted equation number in the surrounding equation.

E. Some Common Mistakes

- The word "data" is plural, not singular.
- The subscript for the permeability of vacuum μ_0, and other common scientific constants, is zero with subscript formatting, not a lowercase letter "o".
- In American English, commas, semicolons, periods, question and exclamation marks are located within quotation marks only when a complete thought or name is cited, such as a title or full quotation. When quotation marks are used, instead of a bold or italic typeface, to highlight a word or phrase, punctuation should appear outside of the quotation marks. A parenthetical phrase or statement at the end of a sentence is punctuated outside of the closing parenthesis (like this). (A parenthetical sentence is punctuated within the parentheses.)
- A graph within a graph is an "inset", not an "insert". The word alternatively is preferred to the word "alternately" (unless you really mean something that alternates).
- Do not use the word "essentially" to mean "approximately" or "effectively".
- In your paper title, if the words "that uses" can accurately replace the word "using", capitalize the "u"; if not, keep using lower-cased.
- Be aware of the different meanings of the homophones "affect" and "effect", "complement" and "compliment", "discreet" and "discrete", "principal" and "principle".
- Do not confuse "imply" and "infer".
- The prefix "non" is not a word; it should be joined to the word it modifies, usually without a hyphen.
- There is no period after the "et" in the Latin abbreviation "et al.".
- The abbreviation "i.e." means "that is", and the abbreviation "e.g." means "for example".

An excellent style manual for science writers is [7].

F. Authors and Affiliations

The class file is designed for, but not limited to, six authors. A minimum of one author is required for all conference articles. Author names should be listed starting from left to right and then moving down to the next line. This is the author sequence that will be used in future citations and by indexing services. Names should not be listed in columns nor group by affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization).

G. Identify the Headings

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

Component heads identify the different components of your paper and are not topically subordinate to each other. Examples include Acknowledgments and References and, for these, the correct style to use is "Heading 5". Use "figure caption" for your Figure captions, and "table head" for your table title. Run-in heads, such as "Abstract", will require you

to apply a style (in this case, italic) in addition to the style provided by the drop down menu to differentiate the head from the text.

Text heads organize the topics on a relational, hierarchical basis. For example, the paper title is the primary text head because all subsequent material relates and elaborates on this one topic. If there are two or more sub-topics, the next level head (uppercase Roman numerals) should be used and, conversely, if there are not at least two sub-topics, then no subheads should be introduced.

H. Figures and Tables

a) Positioning Figures and Tables: Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation "Fig. 1", even at the beginning of a sentence.

TABLE I
TABLE TYPE STYLES

Table Head	**Table Column Head**		
	Table column subhead	***Subhead***	***Subhead***
copy	More table copy[a]		

[a]Sample of a Table footnote.

Fig. 1. Example of a figure caption.

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader. As an example, write the quantity "Magnetization", or "Magnetization, M", not just "M". If including units in the label, present them within parentheses. Do not label axes only with units. In the example, write "Magnetization (A/m)" or "Magnetization {A[m(1)]}", not just "A/m". Do not label axes with a ratio of quantities and units. For example, write "Temperature (K)", not "Temperature/K".

ACKNOWLEDGMENT

The preferred spelling of the word "acknowledgment" in America is without an "e" after the "g". Avoid the stilted expression "one of us (R. B. G.) thanks ...". Instead, try "R. B. G. thanks...". Put sponsor acknowledgments in the unnumbered footnote on the first page.

REFERENCES

Please number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use "Ref. [3]" or "reference [3]" except at the beginning of a sentence: "Reference [3] was the first ..."

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the abstract or reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors' names; do not use "et al.". Papers that have not been published, even if they have been submitted for publication, should be cited as "unpublished" [4]. Papers that have been accepted for publication should be cited as "in press" [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

REFERENCES

[1] Paper Arion
[2] Paper Arion 2
[3] Paper Arion 3
[4] Shai Shalev-Shwartz and Shai Ben-David, Understanding machine learning: From foundations to algorithms. Cambridge Etc: Cambridge University Press, 2014.
[5] F. Maymí and S. Lathrop, "AI in Cyberspace: Beyond the Hype," 2024.
[6] V. V. Putri Wibowo, Z. Rustam, and J. Pandelaki, "Classification of Brain Tumor Using K-Nearest Neighbor-Genetic Algorithm and Support Vector Machine-Genetic Algorithm Methods," in 2021 International Conference on Decision Aid Sciences and Application, Sakheer, Bahrain: IEEE, Dec. 2021, pp. 1077–1081. doi: 10.1109/DASA53625.2021.9682341.
[7] N. F. B. M. Noor, H. S. Sipail, N. Ahmad, and N. M. Noor, "Covid-19 Severity Classification Using Supervised Learning Approach," in 2021 IEEE National Biomedical Engineering Conference, Kuala Lumpur, Malaysia: IEEE, Nov. 2021, pp. 151–156. doi: 10.1109/NBEC53282.2021.9618747.
[8] A. P. Pawlovsky and H. Matsuhashi, "The use of a novel genetic algorithm in component selection for a kNN method for breast cancer prognosis," in 2017 Global Medical Engineering Physics Exchanges/Pan American Health Care Exchanges, Tuxtla-Gutierrez, Mexico: IEEE, Mar. 2017, pp. 1–5. doi: 10.1109/GMEPE-PAHCE.2017.7972084.
[9] A. C. Müller and S. Guido, Introduction to Machine Learning with Python: A Guide for Data Scientists. Beijing: O'reilly, 2017.
[10] K. Kim, "Normalized class coherence change-based k NN for classification of imbalanced data," Pattern Recognition, vol. 120, p. 108126, Dec. 2021, doi: 10.1016/j.patcog.2021.108126.
[11] K. Yuk Carrie Lin, "Optimizing variable selection and neighbourhood size in the K-nearest neighbour algorithm," Computers and Industrial Engineering, vol. 191, p. 110142, May 2024, doi: 10.1016/j.cie.2024.110142.
[12] S. Zhang, D. Cheng, Z. Deng, M. Zong, and X. Deng, "A novel k NN algorithm with data-driven k parameter computation," Pattern Recognition Letters, vol. 109, pp. 44–54, Jul. 2018, doi: 10.1016/j.patrec.2017.09.036.
[13] A. Rokem and K. Kay, "Fractional ridge regression: a fast, interpretable reparameterization of ridge regression," GigaScience, vol. 9, no. 12, p. giaa133, Nov. 2020, doi: 10.1093/gigascience/giaa133.
[14] B. Akturk, U. Beyaztas, H. L. Shang, and A. Mandal, "Robust functional logistic regression," Adv Data Anal Classif, Feb. 2024, doi: 10.1007/s11634-023-00577-z.
[15] M. Arifuzzaman, Md. R. Hasan, T. J. Toma, S. B. Hassan, and A. K. Paul, "An Advanced Decision Tree-Based Deep Neural Network in Nonlinear Data Classification," Technologies, vol. 11, no. 1, p. 24, Feb. 2023, doi: 10.3390/technologies11010024.

[16] F. Aaboub, H. Chamlal, and T. Ouaderhman, "Analysis of the prediction performance of decision tree-based algorithms," in 2023 International Conference on Decision Aid Sciences and Applications (DASA), Annaba, Algeria: IEEE, Sep. 2023, pp. 7–11. doi: 10.1109/DASA59624.2023.10286809.
[17] R. G. Mantovani, T. Horvath, R. Cerri, J. Vanschoren, and A. C. P. L. F. De Carvalho, "Hyper-Parameter Tuning of a Decision Tree Induction Algorithm," in 2016 5th Brazilian Conference on Intelligent Systems (BRACIS), Recife: IEEE, Oct. 2016, pp. 37–42. doi: 10.1109/BRACIS.2016.018.
[18] Y.-C. Zhang and L. Sakhanenko, "The naive Bayes classifier for functional data," Department of Statistics and Probability, Michigan State University, East Lansing, MI, USA, Accepted April 27, 2019.
[19] A. Khajenezhad, M. A. Bashiri, and H. Beigy, "A distributed density estimation algorithm and its application to naive Bayes classification," Sharif Intelligent Systems Laboratory, Department of Computer Engineering, Sharif University of Technology, Tehran, Iran, Accepted October 20, 2020.
[20] N. Guenther dan M. Schonlau, "Support vector machines," The Stata Journal, vol. 16, no. 4, pp. 917–937, 2016.
[21] E. Osuna, R. Freund, dan F. Girosi, "An improved training algorithm for support vector machines," in Proceedings of IEEE, 1997, pp. 252–1723.
[22] C. Cortes dan V. Vapnik, "Support-vector networks," Machine Learning, vol. 20, pp. 273-297, 1995.
[23] N. Settouti, M. El Habib Daho, M. E. Amine Lazouni, and M. A. Chikh, "Random forest in semi-supervised learning (Co-Forest)," in 2013 8th International Workshop on Systems, Signal Processing and their Applications (WoSSPA), Algiers, Algeria, 2013, pp. 326-329, doi: 10.1109/WoSSPA.2013.6602385.
[24] P. S. Akash, M. E. Kadir, A. A. Ali, M. N. Ahad Tawhid, and M. Shoyaib, "Introducing Confidence as a Weight in Random Forest," in 2019 International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST), Dhaka, Bangladesh, 2019, pp. 611-616, doi: 10.1109/ICREST.2019.8644396.
[25] C. Bentéjac, A. Csörgő, and G. Martínez-Muñoz, "A Comparative Analysis of XGBoost," Preprint, Nov. 2019.
[26] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16), New York, NY, USA, 2016, pp. 785–794, doi: 10.1145/2939672.2939785.
[27] P. Zhang, Y. Jia, and Y. Shang, "Research and application of XGBoost in imbalanced data," International Journal of Distributed Sensor Networks, vol. 18, no. 6, 2022, doi: 10.1177/15501329221106935.
[28] C. P. Ananda, "Machine Learning Untuk Prediksi Gaya Hidup Berdasarkan Socioeconomic Status Ses Menggunakan Algoritma Catboost Studi Kasus: Mahasiswa UIN Jakarta", Bachelor's thesis, Fakultas Sains dan Teknologi UIN Syarif Hidayatullah Jakarta
[29] R. Sanjeetha, A. Raj, K. Saivenu, M. I. Ahmed, B. Sathvik, A. Kanavalli, "Detection and mitigation of botnet based DDoS attacks using catboost machine learning algorithm in SDN environment," International Journal of Advanced Technology and Engineering Exploration, vol. 8, no. 76, p. 445., 2021.