**Author Guidelines for the Preparation of Contributions to Springer Proceedings – FOCUS2024**

\*Author11, Author2 1, Author31, and Author42

1Address 1

1\*Email id

2 Address 2

2Email id

\* Corresponding Author

**Abstract –** The Human Resource (HR) department plays a significant role in managing workforce dynamics. This includes identifying potential employees for job promotion based on their abilities and performances. However, in the Malaysian job market, this process is often affected by manual intervention and biases persisting between people with different ethnicities, religions, and genders. This research aims to develop a classification model with Python to automate job promotion and enhance the transparency and efficiency of talent management processes. Additionally, a dashboard will be generated to facilitate visualization of the company's employee performance. By analyzing various factors such as employee performances and training records, the model will reduce the subjective human bias traditionally present in promotion decisions. The insights generated from this research offer valuable guidance for HR professionals seeking to adopt an automated merit-based advancement system for promotion practices within their organizations.

**Keywords –** Human resource, job promotion, automation, classification, dashboar

**1 Introduction**

The Human Resource (HR) department plays a crucial role in managing the workforce of an organization. Typically, HR is involved in various operations such as recruitment, employee training and development, fostering relationships with employees, and conducting performance appraisals. A key aspect of HR's mandate includes overseeing promotions within the company, as promotions and demotions can have a significant impact on morale, retention, and productivity. According to Indeed, an employee who has been promoted and changes jobs may expect an average pay raise between 10% and 20% (Bert, 2024).

Position promotion is an integral part of career development, as it rewards workers with more responsibilities and better pay. Based on organizational policies, these promotion-based decisions are taken on different aspects. These can be the length of service, experience, seniority, performance, etc (Barman, 2024). However, the job environment in Malaysia commonly faces bias and discrimination, which can impede the fairness of promotions. This bias can lead to negative outcomes, including a toxic culture, reduced employee satisfaction, and a high turnover rate. The repercussions of this prejudice are significant, as the issue not only affects the employees themselves but also impacts the entire enterprise.

**2 Problem Statement**

In a competitive business environment, the identification and promotion of talented employees are crucial for organizational success. However, the challenge of identifying high-potential employees based on their performance and abilities is exacerbated by the presence of biases in the Malaysian working environment. This scenario often results in discrimination and prejudice persisting between employees and higher management with different ethnicities, religions, skin color, and gender.

To further support this statement quantitatively, the State of Discrimination Survey Malaysia conducted by Architects of Diversity in 2023 revealed that discriminative experiences among Malaysians ranked second (30%) during the job search process and third, with 29% experiencing discrimination at work. These findings underscore the prevalence of discrimination and its detrimental impact on fostering toxic working environments in our multicultural country (Persatuan Pendidikan Diversiti, 2023).

Through this research, our aim is to enhance the efficiency and transparency of talent management processes by automating the promotion prediction process to identify employees most likely to be promoted. We hope to mitigate the impact of discrimination and biases in the Malaysian workforce, fostering a more inclusive and merit-based system for talent recognition and advancement.

**3 Objectives**

To solve the issue stated above, we had established several research objectives that we aim to achieve by the end of this project:

* Analyze the performance of employees through visualisation
* Develop a classification model to identify potential workers for position promotion based on their abilities and performances.
* Identify the most important factor that contributes to position promotion for the company.

**4 Methodology**

**4.1 Data Collection**

The dataset was retrieved from a public datahack contest held on the Analytics Vidhya website (**Link**). It contains various features related to employee details, performance, training evaluations and the state of whether the employee is recommended for promotion as the target variable. The dataset has 54808 rows and 13 columns, providing a large sample size for robust statistical analysis and model training.

**4.2 Data Cleaning**

The dataset was cleaned by checking for null values and any duplicated rows with Python. Two columns were identified with null values: "previous\_year\_rating" and "education." To handle these, the "previous\_year\_rating" column was filled with the value 0, indicating that the employee is working for the first year and does not have any previous year rating. For the "education" column, a clustering algorithm was employed. Specifically, Principal Component Analysis (PCA) was applied for dimensionality reduction on all numerical features and then k-means clustering was used to impute the missing values in the "education" column.

**4.3 Data Visualisation**

We utilized PowerBI to visualize employee performance namely. This step included conducting descriptive analysis on summarized employee statistics, distribution of variables, and examining the relationship between features and the target variable "is\_promoted" to analyze employee performance within the company. For the generated content was visualized into a dashboard report for easier navigation and comprehensive analysis. Through visualization, we had a clearer understanding of the dataset's characteristics which allowed us to identify patterns and trends in employee performance and provide further insights for analytics and model training.

**4.4 Data Preprocessing**

Several preprocessing steps were carried out using Python Library to prepare the dataset for classification tasks. First, extreme outliers were removed using the z-score method to ensure the robustness of subsequent analyses. Following this, label encoding was applied to the "education" column to map categorical values to numerical representations, while one-hot encoding was implemented on the "recruitment\_channel" column to transform categorical data into binary vectors. Feature selection was conducted to enhance the predictive performance of the model by dropping columns deemed irrelevant for the classification task, which are "employee\_id", "region", "department", "gender" and "age".

To address the issue of class imbalance in the classification task, two separate dataframes were created to compare the performance of different approaches in handling this issue. Firstly, an oversampled dataframe (oversampled\_df) was generated by up-sampling the class with lower samples using the SMOTE (Syntheticers to Minority Over-sampling Technique) library, specifically by up-sampling class '1' to match the sample size of class '0' (where "0" ref not promoted and "1" refers to promoted). SMOTE works by selecting minority observations that are similar to each other and drawing a line between the examples in order to create new synthetic samples (Hoffman, 2021). Conversely, an undersampled dataframe (undersampled\_df) was created by down-sampling the class with higher samples using the Pandas library, in which class '0' was down-sampled to match the sample size of class '1'. These two approaches allow for a comparative analysis of handling class imbalance and its impact on classification performance. The dataset will then be split into 80:20 ratio for training and testing.

**4.5 Data Modelling**

For model training, we implemented 4 different classification algorithm models using Python for each dataframe which are "Logistic Regression", "Decision Tree", "Random Forest" and "K-nearest Neighbors". The objective was to evaluate the performance of different models on the same task and identify the most suitable model based on their performance metrics generalized on unseen testing data. Besides comparing internally between different models, we also compared the overall performance of classification models on the oversampled and undersampled data to identify any significant differences in the results.

**4.5 Model Evaluation**

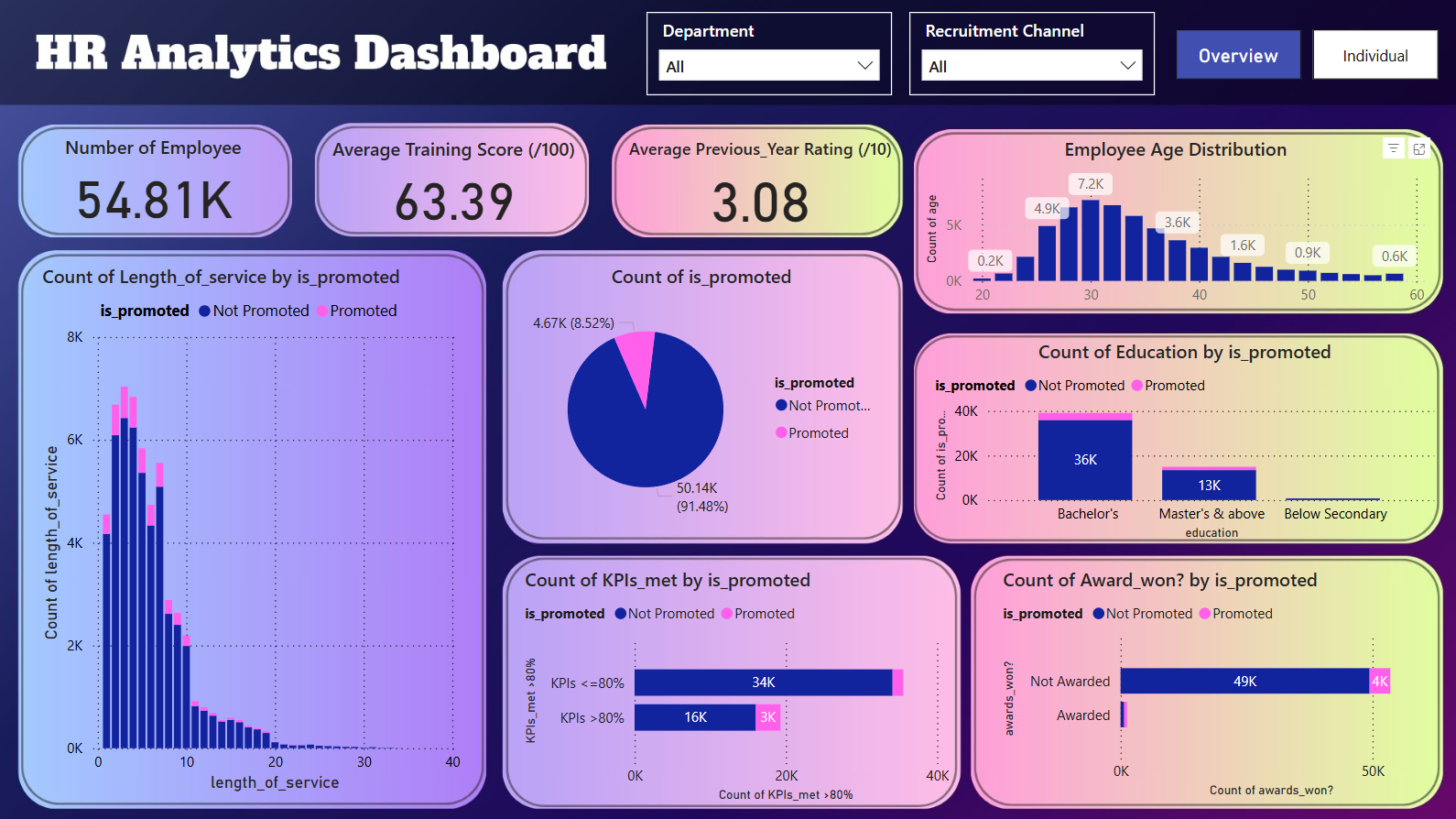
For each model trained on both dataframes, we evaluated their performance on testing data using key evaluation metrics such as accuracy, precision, recall, and F1-score. These metrics were organized into a table format to facilitate easy comparison across models and dataframes.

By comparing the results of each model across both dataframes, we aimed to identify the model with the best performance for further hyperparameter tuning. This comprehensive evaluation process enabled us to make informed decisions regarding the selection of the most effective classifier for the classification task at hand. Finally, we visualized the feature importance of the tuned model to identify the most significant factors contributing to position promotion within the company.

**5 Results and Discussion**

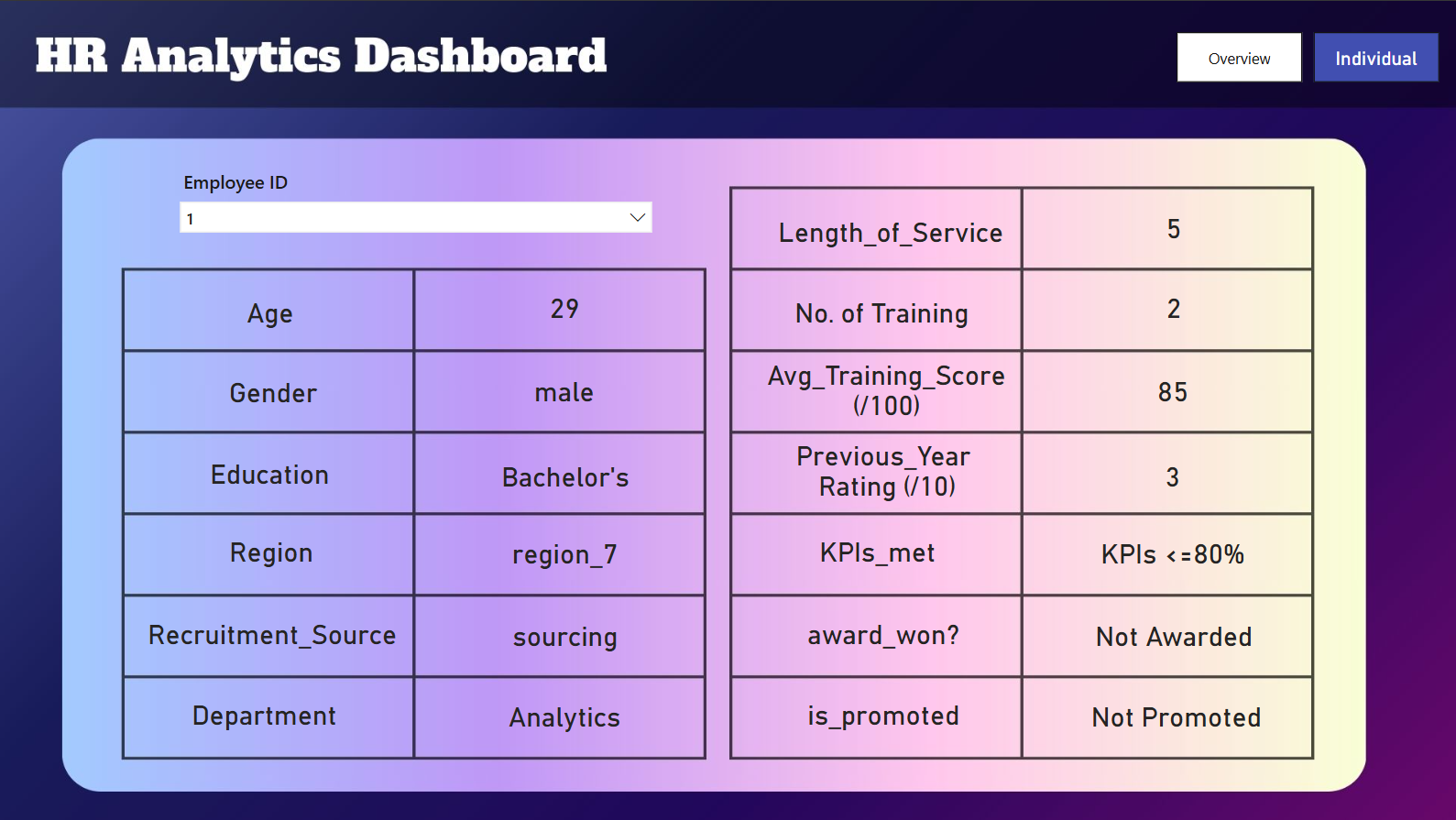
**5.1 Visualisation on Employee Performance**

In this project, PowerBI is ultilised for visualizing and analyzing employee performance. Its intuitive drag-and-drop interface made it accessible to users across varying technical skill levels to create insightful dashboards. Leveraging real-time data connections and dynamic visualizations, PowerBI provided immediate insights into workforce dynamics, enabling HR teams to monitor performance metrics continuously. Moreover, PowerBI offers the flexibility of cloud-based access, allowing multiple users to access and interact with the dashboard simultaneously. Additionally, PowerBI provides robust capabilities for data transformation and manipulation. With its comprehensive suite of tools, HR professionals can easily clean, filter, and format data to meet their specific requirements. This flexibility enables HR teams to tailor the dashboard to their unique needs.

****

**Fig. 1.** PowerBI Dashboard for Overview Performance

Based on the visualization in Fig. 1, it's evident that 8.52% of the 54.81K employees across all departments and channels have received promotions. The average training score is 63.39 out of 100, indicating moderate effectiveness of training programs for the company. The average previous year rating is 3.08 out of 10, which is relatively low, suggesting room for improvement compared to the recent training score. The majority of employees are in the 30-40 age range, with fewer employees in the 20-30 and 50+ age ranges. Most employees have a length of service between 0-10 years, and the count of promotions is also higher within this range. This infers that new employees are more easily promoted. Additionally, employees with higher education levels (Bachelor's and Master's) are more likely to be promoted as they likely possess more skills and knowledge in their respective fields. Moreover, it's very clear that employees who meet 80% or more of their KPIs have a higher chance of being promoted, highlighting the importance of KPI performance in promotion decisions. The ratio of awarded employees promoted to awarded employees not promoted is higher than the ratio of non-awarded employees promoted to non-awarded employees not promoted. This suggests that receiving awards is a strong indicator of promotion likelihood. The dashboard can also be filtered by department or recruitment channel using the slicer above.



**Fig. 2.** PowerBI Dashboard for Individual Performance

In addition to providing an overview of the performance of all employees, there is also a dashboard interface for HR to view the details and performance of an employee by searching using their unique employee ID, as shown in Fig. 2.

**5.2 Evaluation on Classification Model**

The initial exploration of the dataset revealed a significant imbalance in the 'is\_promoted' column, with promoted employees (indicated by '1') being notably underrepresented. This imbalance necessitates the use of sampling techniques to address the disparity. Both oversampling and undersampling techniques have been applied to the dataset to manage the imbalance. Five machine learning algorithms—Logistic Regression, Decision Tree, Random Forest, k-Nearest Neighbors, and Gradient Boosting—have been evaluated on both the oversampled and undersampled datasets to determine the most effective method for improving prediction accuracy.

**Table 1.** Font sizes of headings. Table captions should always be positioned *above* the tables.

|  |  |  |
| --- | --- | --- |
| Heading level | Example | Font size and style |
| Title (centered) | **Lecture Notes** | 14 point, bold |
| 1st-level heading | **1 Introduction** | 12 point, bold |
| 2nd-level heading | **2.1 Printing Area** | 10 point, bold |
| 3rd-level heading | **Run-in Heading in Bold.** Text follows | 10 point, bold |
| 4th-level heading | *Lowest Level Heading.* Text follows | 10 point, italic |

Words joined by a hyphen are subject to a special rule. If the first word can stand alone, the second word should be capitalized.

Here are some examples of headings: “Criteria to Disprove Context-Freeness of Collage Languages”, “On Correcting the Intrusion of Tracing Non-deterministic Programs by Software”, “A User-Friendly and Extendable Data Distribution System”, “Multi-flip Networks: Parallelizing GenSAT”, “Self-determinations of Man”.

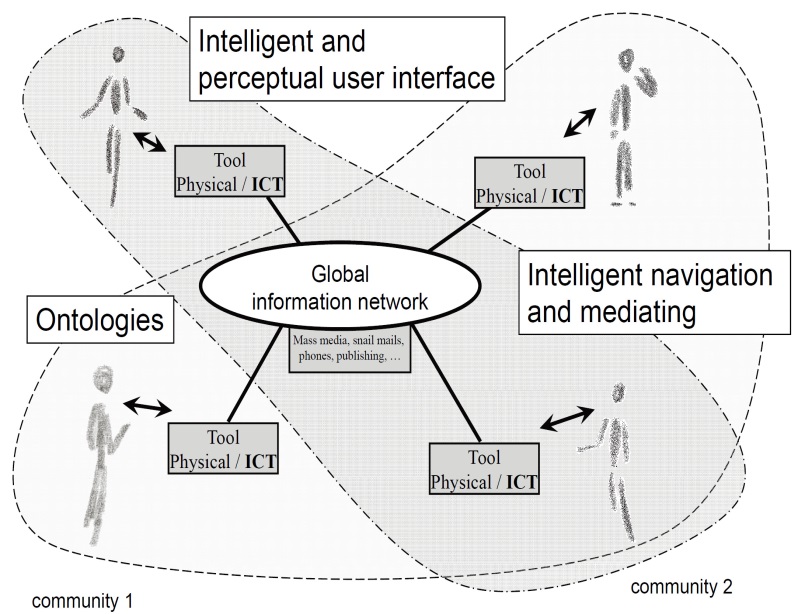
**Lemmas, Propositions, and Theorems.** The numbers accorded to lemmas, proposi­tions, and theorems, etc. should appear in consecutive order, starting with Lemma 1. Please do not include section counters in the numbering like “Theorem 1.1”.

**2.2 Page Numbering and Running Heads**

There is no need to include page numbers or running heads; this will be done at our end. If your paper title is too long to serve as a running head, it will be shortened. Your suggestion as to how to shorten it would be most welcome.

**2.3 Figures and Tables**

It is essential that all illustrations are as clear and as legible as possible. Vector graphics – instead of rasterized images – should be used for diagrams and schemas whenever possible. Please check that the lines in line drawings are not interrupted and have a constant width. Grids and details within the figures must be clearly legible and may not be written one on top of the other. Line drawings are to have a resolution of at least 800 dpi (preferably 1200 dpi). The lettering in figures should not use font sizes smaller than 6 pt (~ 2 mm character height). Figures are to be numbered and to have a caption which should always be positioned *under* the figures, in contrast to the caption belonging to a table, which should always appear *above* the table.



**Fig. 1.** Artifacts empowered by Artificial Intelligence (Source: LNCS 5640, p. 115)

Captions are set in 9-point type. If they are short, they are centered between the margins. Longer captions, covering more than one line, are justified. Captions that do not constitute a full sentence, do not have a period.

Text fragments of fewer than four lines should not appear at the tops or bottoms of pages, following a table or figure. In such cases, it is better to set the figures right at the top or right at the bottom of the page. A figure should never be placed in the middle of a paragraph.

If screenshots are necessary, please make sure that the essential content is clear to the reader.

*Remark 1.* In the printed volumes, illustrations are generally black and white (half­tones), and only in exceptional cases, and if the author or the conference organization is prepared to cover the extra costs involved, are colored pictures accepted. Colored pictures are welcome in the electronic version free of charge. If you send colored figures that are to be printed in black and white, please make sure that they really are also legible in black and white. Some colors show up very poorly when printed in black and white.

**2.4 Formulas**

Displayed equations or formulas are centered and set on a separate line (with an extra line or half line space above and below). Displayed expressions should be numbered for reference. The numbers should be consecutive within the contribution, with num­bers enclosed in parentheses and set on the right margin. Please do not include section counters in the numbering.

*x* + *y* = *z* (1)

Equations should be punctuated in the same way as ordinary text but with a small space before the end punctuation mark.

**2.5 Footnotes**

The superscript numeral used to refer to a footnote appears in the text either directly after the word to be discussed or – in relation to a phrase or a sentence – following the punctuation mark (comma, semicolon, or period).[[1]](#footnote-1)

For remarks pertaining to the title or the authors’ names, in the header of a paper, symbols should be used instead of a number. Please note that no footnotes may be included in the abstract. Font size for fotnotes should be 9 point.

**2.6 Citations and Bibliography**

For citations in the text, use square brackets and consecutive num­bers. Numbers should be grouped where appropriate. We would write [1-5] but [1], [3], [5], etc. The numbers in the bibliography section are without square brackets. Page numbers are required for quotations. Page numbers should be indicated in square bracets [5, p.26].

Please write all references using the latin alphabet. If the title of the book you are referring to is, e.g., in Russian or Chinese, then please write (in Russian) or (in Chi­nese) at the end of the transcript or translation of the title.

Please base your references on the examples given in the references section of these instructions. References that do not adhere to this style will need to be reformatted.

Please make sure that all your sources are correctly listed in the reference section. Do not include references to pieces of work that are not connected with your paper.

The references section at the end of this paper shows a sample reference list with entries for journal articles [1], an CCIS chapter [2], a book [3], proceedings without editors [4] and [5], as well as a URL [6]. Please note that Springer proceedings are cited with their publication acronyms and volume numbers.

Here is a list of the general rules to be followed:

* Italics or bold face are not to be used.
* Authors’ and editors’ names are to be structured as follows: “Surname, Initials
* of First Names”.
* A colon is to be inserted between the authors’ names and the title.
* The title should be followed by a period, unless page numbers are to be included, in which case the period follows the page numbers.
* All words except articles, prepositions, and conjunctions should be set with an initial capital.
* The abbreviations (ed.) or (eds.) should follow the Editors’ names.
* If a contribution from an edited volume is being cited, then the authors’ names and the title of the paper are given first. This is followed by “In:” and then the details of the book, as shown in reference list.
* The abbreviation pp. should be used for page numbers in books, whereas page numbers in journals should be given without pp.
* The volume number of a book in a series should be given as vol. xxx, where xxx is the volume number. It should be separated from the name/abbreviation of the series by a comma. The volume number of a journal should appear directly after the title of the journal. The issue number should appear directly next to it in brackets.
* The name and town of the publisher, separated by a comma, should appear before the copyright year.
* The copyright year should be given at the end, in brackets. If non-standard publications (e.g., technical reports) also require a month, then this should be included between the brackets, as should the appendages (to be published) or (to appear).

**2.7 Plagiarism**

Springer takes plagiarism seriously. If an author has copied from another author or has used parts of another author’s work, without his or her permission and a refer­ence, then the paper on SpringerLink will be given a “retracted” stamp, and an erra­tum explaining the reasons for the retraction will be included. In addition, the volume editors and the author’s academic supervisors will be informed.

**Acknowledgements.** This should always be a run-in heading and not a section or subsection heading. It should not be assigned a number. The acknowledgements may include reference to grants or supports received in relation to the work presented in the paper.

**3 Additional Information Required from Authors**

**3.1 Copyright Form**

A prefilled copyright form will be provided by volume editors. Please send your signed copyright form to volume editors as a scanned PDF. One author may sign on behalf of all of the other authors of a particular pa­per. In this case, the author signs for and accepts responsibility for releasing this ma­terial on behalf of any and all co-authors. Digital signatures are not acceptable.

**3.2 Contact Author Information**

Kindly assure that, when you submit the final version of your paper, you also provide the name and e-mail address of the contact author for your paper. The contact author must be available to check the paper roughly before the book is due to leave the printing office. He or she will be contacted from the following e-mail address: typesetting@sps.co.in.

**3.3 Correct Representation of Author Names**

Authors’ names should be written out in full at the tops of the papers. They are shortened by us to “initials surname” in the running heads and will take the form “surname, given name” in the author index. If you or any of your co-authors have more than one family name, it should be made quite clear how your name is to be displayed in the running heads and the author index. Chinese authors should write their given names in front of their surnames at the tops of their papers. If you only have one (main) name, please make sure that this name is written out in full in the running heads, when you check your final PDF.

**4 Typesetting of Your Paper at Springer**

Please make sure that the paper you submit is final and complete, that any copyright issues have been resolved, that the authors listed at the top of the chapter really are the final authors, and that you have not omitted any references. Following publication, it is not possible to alter your paper on SpringerLink. Kindly note that we prefer the use of American English.

**4.1 What Will Be Done with Your Paper**

If the instructions have been followed closely, then only very minor alterations will be made to your paper. The production team at SPS checks the format of the paper, and if, for example, vertical spacing has been inserted or removed, making the paper unsightly or difficult to read, then this is remedied. In addition, running-heads, final page numbers and a copyright line are inserted, and the capitalization of the headings is checked and corrected if need be. Finally, the reference section is attuned to our specifications (see also Section 2.7).

**4.2 Proof Reading Stage**

Once the files have been worked upon, SPS sends a copy of the final PDF of each paper to its contact author. The contact author is asked to check through the final PDF to make sure that no errors have crept in during the transfer or preparation of the files. This should not be seen as an opportunity to update or copyedit the paper, which is not possible due to time constraints. Only errors introduced during the preparation of the files will be corrected. Particular attention should be paid to the references section.

If SPS does not receive a reply from a particular contact author, within the time-frame given (usually 72 hours), then it is presumed that the author has found no errors in the paper. The tight publication schedule of our proceedings series does not allow SPS to send reminders or search for alternative e-mail addresses on the Internet.

In some cases, it is the contact volume editor or the publication chair who checks all of the PDFs. In such cases, the authors are not involved in the checking phase.

The purpose of the proof is to check for typesetting or conversion errors and the completeness and accuracy of the text, tables and figures. Substantial changes in con­tent, e.g., new results, corrected values, title and authorship, are not possible and can­not be processed.

**5 Online Publication in SpringerLink**

All papers will be published in our digital library www.springerlink.com. Only subscribers to Springer’s eBook packages or to the electronic book series are able to access the full text PDFs and references of our online publications. Meta-data, abstracts and author e-mail addresses are freely available for all users.

**References**

1. Smith, T.F., Waterman, M.S.: Identification of Common Molecular Subsequences. J. Mol. Biol. 147, 195–197 (1981)
2. May, P., Ehrlich, H.C., Steinke, T.: ZIB Structure Prediction Pipeline: Composing a Complex Biological Workflow through Web Services. In: Nagel, W.E., Walter, W.V., Lehner, W. (eds.) Euro-Par 2006. LNCS, vol. 4128, pp. 1148–1158. Springer, Heidelberg (2006)
3. Foster, I., Kesselman, C.: The Grid: Blueprint for a New Computing Infrastructure. Mor­gan Kaufmann, San Francisco (1999)
4. Czajkowski, K., Fitzgerald, S., Foster, I., Kesselman, C.: Grid Information Services for Distributed Resource Sharing. In: 10th IEEE International Symposium on High Performance Distributed Computing, pp. 181–184. IEEE Press, New York (2001)
5. Foster, I., Kesselman, C., Nick, J., Tuecke, S.: The Physiology of the Grid: an Open Grid Services Architecture for Distributed Systems Integration. Technical report, Global Grid Forum (2002)
6. National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov

1. The footnote numeral is set flush left and the text follows with the usual word spacing. [↑](#footnote-ref-1)