

PLANT-DISEASE RELATION MODEL USING BERT-BILSTM-CRF AND TRANSFER LEARNING APPROACH

SLAMET RIYANTO



COMPUTER SCIENCE STUDY PROGRAM
FACULTY OF MATHEMATICS AND NATURAL SCIENCES
IPB UNIVERSITY
BOGOR
2024



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

STATEMENT REGARDING DISSERTATION AND INFORMATION SOURCE AND COPYRIGHTS

At this moment, I declare that my dissertation is entitled *Plant-Disease Relation Model using BERT-BiLSTM-CRF and the Transfer Learning Approach*. This manuscript has not yet been submitted to any other university. All information sources from published and unpublished manuscripts are included in this dissertation and in the references at the end of the dissertation.

Therefore, I extend my dissertation's copyrights to IPB University in Bogor, Indonesia.

Bogor, July, 2024

Slamet Riyanto
G6601202026

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



RINGKASAN

SLAMET RIYANTO. Model *Plant-Disease Relation* Menggunakan Pendekatan BERT-BiLSTM-CRF dan *Transfer Learning*. Dibimbing oleh IMAS SUKAESIH SITANGGANG, TAUFIK DJATNA dan TIKA DEWI ATIKAH.

Plant-Disease Relations (PDR) merupakan subtugas ekstraksi relasi yang bertujuan menganalisis hubungan antara entitas penyakit manusia-tanaman dalam sebuah kalimat. Penelitian mengenai PDR pertama kali dikemukakan oleh Kim *et al.* pada tahun 2019. Mereka mengembangkan PDR sebagai korpus standar emas dengan memberi label manual pada entitas penyakit manusia-tanaman dengan bantuan para ahli. Adapun korpus tersebut memiliki empat kelas: pengobatan penyakit, penyebab penyakit, negatif, dan asosiasi. Selanjutnya, mereka mengusulkan metode *short dependency path-convolutional neural network* (SDP-CNN) untuk menganalisis hubungan entitas penyakit manusia-tanaman. Algoritma SDP mengidentifikasi jalur terpendek untuk menemukan entitas tanaman dan penyakit manusia, sementara CNN menganalisis hubungan antara dua entitas dalam klasifikasi. Hasil eksperimen menunjukkan kinerja model mencapai skor F1 sebesar 0,764 pada kelas tidak seimbang. Kombinasi SDP-CNN secara efektif mengekstrak fitur lokal dari jalur ketergantungan dan secara efisien menangkap informasi sintaksis dan semantik. Dengan berkonsentrasi pada bagian kalimat yang paling relevan, SDP-CNN dapat meningkatkan akurasi tugas ekstraksi relasi, khususnya dalam kalimat sederhana.

Di sisi lain, metode SDP-CNN memiliki beberapa keterbatasan: 1) tidak memperhitungkan keseimbangan kelas saat membangun model PDR; 2) metode SDP-CNN hanya optimal untuk kalimat pendek, namun tidak efektif untuk bahasa Inggris kompleks; dan 3) mengandalkan pakar saat menyusun data berlabel. Oleh karena itu, penelitian ini bertujuan untuk 1) memilih teknik terbaik melalui oversampling, undersampling, dan SMOTE untuk menangani kelas-kelas yang tidak seimbang; 2) meningkatkan skor F1 model hubungan penyakit manusia-tanaman yang diusulkan oleh Kim *et al.* (2019) dengan menggabungkan algoritma BERT-BiLSTM-CRF; dan 3) mengusulkan metode pembelajaran transfer multi-sumber untuk pelabelan entitas secara otomatis. Sedangkan kebaruan penelitian ini adalah sebagai berikut: 1) metode BERT-BiLSTM-CRF untuk memprediksi hubungan penyakit manusia-tanaman, dan 2) metode pembelajaran transfer multi-sumber untuk pelabelan otomatis.

Penelitian ini melakukan pra-pemrosesan dan menerapkan stop_word untuk mewujudkan tujuan pertama. Kemudian teks tersebut diubah menjadi teks numerik menggunakan Count Vectorizer dengan penambahan *n-gram*. Penelitian ini menggunakan rasio 80:20 dan *5-fold cross validation* untuk tahap pelatihan. Penelitian ini menggunakan pengklasifikasi pembelajaran mesin (Multinomial Naïve Bayes, K-Nearest Neighbour, Support Vector Machine, dan Random Forest) dan *Deep Learning* (*Long Short-Term Memory*). Selain itu, teknik *undersampling*, *oversampling*, dan *Synthetic Minority Oversampling Technique* (SMOTE) digunakan untuk mengatasi kelas yang tidak seimbang. Hasil eksperimen menunjukkan bahwa classifier LSTM dengan teknik *undersampling* lebih unggul dibandingkan semua *classifier machine learning*.

Untuk mencapai tujuan kedua, penelitian ini melakukan normalisasi korpus PDR. Setiap kalimat disusun dalam satu baris dan diakhiri satu titik. Selanjutnya, proses tokenisasi menggunakan BERT *Tokenizer*. Pada tahap pelatihan, data pelatihan dan pengujian dibagi dengan rasio 80:20. *Hyperparameter tuning* dilakukan menggunakan GridSearch untuk mendapatkan konfigurasi model yang optimal. Hasil kalkulasi GridSearch memperoleh parameter optimal sebagai berikut: epoch=40, dropout=0.3, batch_size=64, dan num_units=128. Untuk mendapatkan model yang optimal, percobaan dilakukan sebanyak sepuluh kali dengan parameter yang sama. Selain itu, eksperimen 11-14 dilakukan dengan memodifikasi nilai parameter epoch dan batch_size. Hasil percobaan ke 7 memperoleh skor F1 sebesar 0,78. Sedangkan pengurangan jumlah epoch pada percobaan ke-12 memperoleh skor F1 sebesar 0,79. Sementara itu, pengurangan jumlah batch_size pada percobaan ke-14 menghasilkan skor F1 sebesar 0,80. Selain itu, teknik *oversampling* digunakan untuk menyeimbangkan kelas dan kinerja lebih optimal dengan skor F1 0.91. Hal ini menunjukkan bahwa model BERT-BiLSTM-CRF memiliki kinerja lebih baik dibandingkan penelitian sebelumnya.

Untuk mencapai tujuan ketiga, pertama penelitian ini mengumpulkan data terkait biomedis dan botani. Kumpulan data ini digunakan sebagai kandidat *source domain* untuk pelatihan. Pada tahap *skimming*, seluruh dataset dipelajari untuk mengetahui relevansinya dengan *target domain* (korpus PDR). Untuk memastikan bahwa data sumber serupa dengan domain target, penelitian ini menggunakan metrik Jensen-Shenen Divergence (JSD). Setelah memperoleh data sumber, dilakukan *label mapping* pada dataset botani sesuai data target. Hal ini diperlukan karena data botani memuat label spesies, genus, family, subfamily, ordo, kelas, dan filum. Selanjutnya, algoritme berbasis aturan dibuat untuk membersihkan kumpulan data yang hanya berisi label "O". Data pelatihan, pengujian, dan evaluasi mengikuti rasio 80:10:10. Tahap pelatihan menggunakan *fine-tuning* dengan model BioBERT. Empat skema digunakan untuk mengembangkan model dari sumber dan data target. Hasil eksperimen menunjukkan bahwa pembelajaran transfer yang diusulkan memiliki kinerja tinggi dalam memberikan label secara akurat, dengan skor F1 sebesar 0,92. Kinerja ini dipengaruhi oleh penambahan algoritma *rule-based* pada tahap pra-premrosesan. Hasil ini menunjukkan bahwa metrik JSD berperan penting dalam memilih domain sumber yang berkualitas.

Berdasarkan hasil percobaan dapat disimpulkan bahwa 1) classifier LSTM lebih unggul dibandingkan classifier lainnya untuk tugas klasifikasi. Teknik *undersampling* digunakan untuk mengatasi ketidakseimbangan kelas pada data yang digunakan dalam penelitian ini. 2) Metode BERT-BiLSTM-CRF yang diusulkan lebih unggul dibandingkan yang digunakan pada penelitian sebelumnya. Kinerja ini dipengaruhi oleh lapisan BiLSTM yang beroperasi dua arah dan dapat menyimpan informasi dari dokumen panjang. 3) Metode pembelajaran transfer yang diusulkan bekerja dengan baik dalam memberikan label secara otomatis. hal ini dipengaruhi oleh kualitas sumber data yang digunakan sebagai data latih. Meskipun demikian, penelitian ini masih memiliki keterbatasan dan perlu dikembangkan lebih lanjut. Studi lebih lanjut tentang *co-reference resolution* diperlukan untuk meningkatkan kinerja model.

Kata kunci: *automatic labeling, conditional random field, long short-term memory, plant-disease relation, transfer learning*

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
- b. Pengutipan tidak mengikan kepentingan yang wajar IPB University.
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



SUMMARY

SLAMET RIYANTO. Plant-Disease Relation Model using BERT-BiLSTM-CRF and Transfer Learning Approach. Supervised by IMAS SUKAESIH SITANGGANG, TAUFIK DJATNA, and TIKA DEWI ATIKAH.

A Plant-Disease Relation (PDR) is a relation extraction sub-task that aims to reveal the relationship between plants and disease entities in a sentence. Research on the PDR was first proposed by Kim *et al.* in 2019. They developed PDR as a gold-standard corpus by manually labeling the plant and disease entities with help of the experts. The corpus has four classes: treatment of disease, cause of disease, negative, and associate. Subsequently, they proposed a short dependency path-convolutional neural network (SDP-CNN) method to analyze the relationship between plants and disease entities. The SDP algorithm identifies the shortest path to find plant and disease entities while CNN analyzes the relationship between the two entities in the classification. The experimental results show that the model performance achieved an F1 score of 0.764 for imbalanced class data. The SDP-CNN combination effectively extracts local features from dependency paths and efficiently captures syntactic and semantic information. By concentrating on the most relevant parts of the sentences, SDP-CNN can improve the accuracy of relation extraction tasks, particularly in simpler sentences.

On the other hand, the SDP-CNN method has several limitations: 1) Does not take class balance into account when building the PDR model; 2) the SDP-CNN method is only optimal for short sentences, but is not effective for complex English; and 3) rely on Experts when building ground truth datasets. Therefore, this research aims to 1) to select the best techniques in oversampling, undersampling, and SMOTE to deal with imbalance classes; 2) to improve the F1 score of plant-disease relation model proposed by Kim *et al.* (2019) by combine BERT-BiLSTM-CRF algorithm; and 3) to propose a multi-source transfer learning method for automatic labeling for replaced rely on the expert. Meanwhile, the novelty of this research is as follows: 1) BERT-BiLSTM-CRF method to predict plant-disease relation, and 2) multi-source transfer learning method for automatic labeling.

We performed preprocessing and applied stop_word to realize the first goal. Then, the text was converted to numeric text using a Count Vectorizer with the addition of n-grams. In both the training and testing stages, we utilized an 80:20 ratio and 5-fold cross validation. We used machine learning classifiers (multinomial naïve bayes, k-nearest neighbour, support vector machine, and random forest) and Deep Learning (LSTM). Additionally, we employed undersampling, oversampling, and the Synthetic Minority Oversampling Technique (SMOTE) to address with unbalanced classes. Overall, experiments were conducted on all classifiers using the eight schemes. The experimental results show that, the LSTM classifier with the undersampling technique is superior to all the machine learning classifiers.

To achieve the second goal, we normalized the PDR corpus. Each sentence was arranged in a single line and ended within a period. Next, the tokenisation process used the BERT Tokenizer. In the training stage, the training and testing data were divided in an 80:20 ratio. Hyperparameter tuning was performed using GridSearch to obtain the optimal model configuration. It took approximately 28

hours to obtain optimal parameters. The results of the GridSearch recommendation yielded the following parameters: epoch=40, dropout=0.3, batch_size=64, and num_units=128. To obtain the optimal model, we conducted the experiments ten times with the same parameters. In addition, we performed experiments 11–14 using four schemes by modifying the epoch and batch_size parameter values. The experimental results show that the 7th experiment obtained an F1 score of 0.78. We found that reducing the number of epochs in the 12th experiment resulted in an F1 score of 0.79. Furthermore, reducing the batch_size in the 14th experiment achieved an F1 score of 0.80. Apart from that, we also employed oversampling techniques to balance classes and achieving improved F1 score of 0.91. This demonstrated that the proposed BERT-BiLSTM-CRF model performs better than the previous research.

To achieve the third objective, we first collected datasets related to biomedicine and botany. These datasets were used as the candidate source domain for the training. During the skimming stage, all the datasets were studied to determine their relevance to the target domain (PDR corpus). The Jensen-Shenen Divergence (JSD) metric was used to ensure that the source data were similar to those of the target domain.. After obtaining the source data, we performed label mapping on the botanical dataset which refers to the target labels. This step is necessary since botanical datasets contain species, genus, family, subfamily, order, class and phylum labels. Subsequently, we constructed a rule-based algorithm to clean the dataset containing only “O” labels. The training, testing, and evaluating data followed a ratio of 80:10:10. GridSearch was employed to obtain the optimal parameters. During the training stage, we used fine-tuning with the BioBERT model. Four schemes were implemented to develop the models from the source and target data. The experimental results showed that the proposed transfer learning has high performance in providing labels accurately, with an F1 score of 0.92. This performance was influenced by the addition of rule-based algorithms at the preprocessing stage before the data were used for training. This result indicated that the JSD metric plays an important role in selecting quality source domains, thereby influencing model performance.

Based on the experimental results, it can be concluded that 1) the LSTM classifier is superior to other classifiers for classification tasks. Undersampling techniques were used to overcome imbalanced classes in the data used in this research. 2) The proposed BERT-BiLSTM-CRF method was superior to those used in previous studies. This performance is influenced by the BiLSTM layer, which operates in both ways and can store information from long documents. 3) The proposed transfer learning method worked well in providing labels automatically. it was influenced by the quality of the data source used as training data. Nevertheless, this research still has limitations and needs to be developed further. Further studies on co-reference resolution are needed to enhance model performance.

Keywords: automatic labeling, conditional random field, long short-term memory, plant-disease relation, transfer learning

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

**© Copyrights Belong to IPB University,
The Year 2024
The Constitution protects copyrights**

*Citing some or all of this manuscript without addressing its source is prohibited.
Citation is permitted for education, research, writing manuscripts, reports,
critiques, or review purposes and is not detrimental to IPB University.*

Publishing or copying any or all of this manuscript is prohibited.



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

PLANT-DISEASE RELATION MODEL USING BERT-BILSTM-CRF AND TRANSFER LEARNING APPROACH

SLAMET RIYANTO

Dissertation
As one of the requirements for obtaining a degree
Doctor on
Computer Science Study Program

**COMPUTER SCIENCE STUDY PROGRAM
FACULTY OF MATHEMATICS AND NATURAL SCIENCES
IPB UNIVERSITY
BOGOR
2024**



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Examiners on Closed Examination

1. Dr. Mushthofa, S.Kom., M.Sc
2. Dr. Esa Prakasa, MT

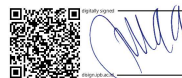
Examiners of the Promotion Examination

1. Dr. Mushthofa, S.Kom., M.Sc
2. Dr. Esa Prakasa, MT

Dissertation Title : Plant-Disease Relation Model using BERT-BiLSTM-CRF
and Transfer Learning Approach
Name : Slamet Riyanto
SIN : G6601202026

Approved by:

Supervisor
Prof. Dr. Imas Sukaesih Sitanggung, S.Si., M.Kom



Co-Supervisor
Prof. DrEng. Ir. Taufik Djatna, M.Si



Co-Supervisor
Dr. Tika Dewi Atikah, M.Si



Acknowledged by

Head of Doctoral Study Program in Computer Science:
Prof. Dr. Imas Sukaesih Sitanggung, S.Si., M.Kom.
NIP 197501301998022001



Dean of the Faculty of Mathematics and Natural Sciences
Dr. Berry Juliandi, S.Si, M.Si
NIP 197106141995121001



Date of Closed Examination: June, 20 2024
Date of Promotion Examination: July, 23 2024

Date of Graduation:



Dokumen ini ditandatangani secara elektronik menggunakan sertifikat dari BSrE, silahkan lakukan verifikasi pada dokumen elektronik yang dapat diunduh dengan melakukan scan QR Code



PREFACE

Alhamdulillah Rabbal'Alamiin. All praise to *Allah Subhanahu Wa Ta'ala* for all His gifts so that this dissertation was successfully completed. The theme chosen in the research carried out from February 2022 to November 2023, was the development of a machine learning algorithm to reveal the extraction of relationships between plant entities and human diseases that appear together in a sentence with the title "Plant-Disease Relation Model using BERT-BiLSTM-CRF and Transfer Learning Approach".

The author would like to express gratitude to the supervisors, Prof. Dr. Imas Sukaesih Sitanggung, S.Si., M.Kom; Prof. DrEng. Ir. Taufik Djatna, M.Sc; and Dr. Tika Dewi Atikah, M.Si, who have guided and provided many suggestions during the doctoral research and completing this dissertation. Gratitude were also expressed to the promotion session moderator, Dr. Hendra Rahmawan, S.Komp., MT; external examiners, Dr. Mushthofa, S.Kom., M.Sc, and Dr. Esa Prakasa, MT. The author would like to thank to the Computer Science Postgraduate Study Program. The author would like to thank to National Research and Innovation Agency for doctoral scholarship through Degree by Research program. The author would like to thank to my best friends are Dr. Budi Nugroho, M.Si, Dr. Ekawati Marlina, MT, Dr. Al Hafiz Akbar M.S., Hendro Subagyo, M.Eng, Sjaeful Afandi, MT, and Rishadi, S.Pd. The author also would like to express love and gratitude to beloved parents, Mrs. Saniyem, Mrs. Ngatmi and Mr. Suratno, and beloved wife Sumiyati and children, Fascal, Izal and Alhazeen, for unlimited support, continous prayers, and unconditional love. The author also would like to express gratitude to all friends of the Doctoral Program of Computer Science 2020 for their support, and all of IPB University friends.

Hopefully, this dissertation will be valuable for stakeholders and the advancement of science.

Bogor, July, 2024

Slamet Riyanto

TABLE OF CONTENTS

LIST OF TABLES	xiii
LIST OF FIGURES	xv
I. INTRODUCTION	1
1.1 Background	1
1.2 Research Questions	2
1.3 Objectives	2
1.4 Advantages	2
1.5 Scope	3
1.6 Novelty	3
II. LITERATURE STUDY	4
2.1 Named Entity Recognition	4
2.2 Relation Extraction In Biomedical Field	5
2.3 Bidirectional Long Short-Term Memory-Conditional Random Field (BiLSTM-CRF)	8
2.4 Transfer Learning for Automatic Labeling	10
2.5 Imbalanced Class	14
III. MATERIALS AND METHOD	17
3.1 Research Stage	17
3.2 Materials	18
3.3 The Research Roadmap	18
3.3.1 Systematic Literature Review	19
3.3.2 Comparative Analysis for Imbalanced Class	21
3.3.3 Development of BERT-BiLSTM-CRF Architecture for Relation Extraction	29
3.3.4 Transfer Learning For Automatic Labeling	36
IV. RESULT AND DISCUSSION	45
4.1 Systematic Literature Review (SLR)	45
4.2 Comparative Analysis for Imbalance Class	49
4.3 Development Of BERT-BiLSTM-CRF Architecture for Relation Extraction	55
4.4 Transfer Learning for Automatic Labeling	59
V. CONCLUSIONS AND FUTURE WORK	65
5.1 Conclusions	65
5.2 Future Work	65
REFERENCE	66

LIST OF TABLES

1. Comparative performance of language modeling	8
2. Sampe of the dataset	21
3. Class of PDR dataset	22
4. Sample of PDR dataset	22
5. Hyperparameter tuning	29
6. Label Mapping from original label	41
7. Literature Review to find research gaps	45
8. Multinomial Naïve Bayes classifier results	50
9. K-Nearest Neighbors classifier results	51
10. Support Vector Machine classifier results	52
11. Random Forest Classifier Results	53
12. LSTM classifier results	53
13. Result of data normalization	55
14. BERT-BiLSTM-CRF model summary	56
15. Result of model performance	57
16. Result of model performance using hyper-parameter tuning	58
17. Comparison of previous research and our model	58
18. Data cleaning	60
19. Label Mapping from original label	60
20. PDNER Model and result of prediction on xexternal data	63

Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - Pengutipan tidak mengikis kepentingan yang wajar IPB University.
2. Dilarang mengutipkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

LIST OF FIGURES

1. Stages carried out in the research	17
2. Research roadmap	19
3. Systematic literature review process flow diagram	20
4. Shortest dependency parse tree	20
5. Flowchart to extract features	23
6. Undersampling and oversampling balancing classes algorithms	27
7. Hybrid resampling process	27
8. Flowcharts of methodology in this research	27
9. Research stages of this research	30
10. Tokenizing process	33
11. The architecture of BERT-BiLSTM-CRF for plant disease relation	35
12. NER task is divided into two sub-tasks	38
13. Transfer learning method for automatic labeling	40
14. Overall workflow for the proposed automatic labeling	44
15. Illustration of BERT data format	55
16. Source code snippet for model configuration	56
17. Training and validation accuracy	56
18. Taxonomy of domain-adaptation	61
19. Distribution label based on specificity and sensitivity metric	64

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.